ISSN : 2165-4069(Online)
ISSN : 2165-4050(Print)

IJARAI

International Journal of
Advanced Research in Artificial Intelligence

Volume 3 Issue 11

www.ijarai.thesai.org

A Publication of
The Science and Information Organization

IJARAI

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN ARTIFICIAL INTELLIGENCE

SAI
THE SCIENCE AND INFORMATION ORGANIZATION
www.thesai.org | info@thesai.org

Editorial Preface

From the Desk of Managing Editor…

"The question of whether computers can think is like the question of whether submarines can swim." ― Edsger W. Dijkstra, the quote explains the power of Artificial Intelligence in computers with the changing landscape. The renaissance stimulated by the field of Artificial Intelligence is generating multiple formats and channels of creativity and innovation.

This journal is a special track on Artificial Intelligence by The Science and Information Organization and aims to be a leading forum for engineers, researchers and practitioners throughout the world.

The journal reports results achieved; proposals for new ways of looking at AI problems and include demonstrations of effectiveness. Papers describing existing technologies or algorithms integrating multiple systems are welcomed. IJARAI also invites papers on real life applications, which should describe the current scenarios, proposed solution, emphasize its novelty, and present an in-depth evaluation of the AI techniques being exploited. IJARAI focusses on quality and relevance in its publications.

In addition, IJARAI recognizes the importance of international influences on Artificial Intelligence and seeks international input in all aspects of the journal, including content, authorship of papers, readership, paper reviewers, and Editorial Board membership.

The success of authors and the journal is interdependent. While the Journal is in its initial phase, it is not only the Editor whose work is crucial to producing the journal. The editorial board members , the peer reviewers, scholars around the world who assess submissions, students, and institutions who generously give their expertise in factors small and large— their constant encouragement has helped a lot in the progress of the journal and shall help in future to earn credibility amongst all the reader members.

I add a personal thanks to the whole team that has catalysed so much, and I wish everyone who has been connected with the Journal the very best for the future.

Thank you for Sharing Wisdom!

Editor-in-Chief
IJARAI
Volume 3 Issue 11 November 2014
ISSN: 2165-4069(Online)
ISSN: 2165-4050(Print)

Editorial Board

Reviewer Board Members

CONTENTS

Evaluating Sentiment Analysis Methods and Identifying Scope of Negation in Newspaper Articles

S Padmaja
Dept. of CSE, UCE
Osmania University
Hyderabad

Prof. S Sameen Fatima
Dept. of CSE, UCE
Osmania University
Hyderabad

Sasidhar Bandu
Dept. of English
Salman Bin Abdul-Aziz University
Saudi Arabia

Abstract*—Automatic detection of linguistic negation in free text is a demanding need for many text processing applications including Sentiment Analysis. Our system uses online news archives from two different resources namely NDTV and The Hindu. While dealing with news articles, we performed three subtasks namely identifying the target; separation of good and bad news content from the good and bad sentiment expressed on the target and analysis of clearly marked opinion that is expressed explicitly, not needing interpretation or the use of world knowledge. In this paper, our main focus was on evaluating and comparing three sentiment analysis methods (two machine learning based and one lexical based) and also identifying the scope of negation in news articles for two political parties namely BJP and UPA by using three existing methodologies. They were Rest of the Sentence (RoS), Fixed Window Length (FWL) and Dependency Analysis (DA). Among the sentiment methods the best F-measure was SVM with the values 0.688 and 0.657 for BJP and UPA respectively. On the other hand, the F measures for RoS, FWL and DA were 0.58, 0.69 and 0.75 respectively. We observed that DA was performing better than the other two. Among 1675 sentences in the corpus, according to annotator I, 1,137 were positive and 538 were negative whereas according to annotator II, 1,130 were positive and 545 were negative. Further we also identified the score of each sentence and calculated the accuracy on the basis of average score of both the annotators.

Keywords—Sentiment Analysis; Negation Identification; News Articles

I. INTRODUCTION

Access to popular communication platforms have given a way for the public to generate emotions, opinions, sentiments, evaluations, appraisals and attitudes towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. The volume of such information created every day is massive. Through such Online Social Networks (OSN) we share this information as part of our everyday lives for better understanding and integrating into our surrounding reality. With the effect of such networks on a daily basis, we start recoiling our valuable decisions and actions with certain predefined notions created by others through their reviews. Thus, it is both interesting and challenging to see what and how the trends in such OSNs change from time to time. Comprehensive investigations can lead to appropriate predictions. This gives an edge to Text Mining and explores the area of Sentiment Analysis (Opinion Mining) which is designated as an automatic processing of texts to detect opinions.

Authors of such OSNs express their opinions freely, but, this is not the case with news articles. The major difference between news and product reviews is that the target of the sentiment is less concrete and is expressed much less explicitly. Another major difference is Newspaper articles is that they give an impression of objectivity to refrain from using clearly positive or negative vocabulary. They resort to other means to express their opinion, like embedding statements in a more complex discourse, omitting facts that highlight some important people. For this reason sentiment analysis on news text is rather difficult compared to others. Moreover, the automatic detection of the scope of linguistic negation is a problem encountered in wide variety of documents like understanding tasks, medical data mining, general fact or relation extraction, question answering, sentiment analysis and many more.

The goal of this work is to evaluate and compare three sentiment analysis methods (two machine learning based and one lexical based) and also to identify the scope of negation in news articles for two political parties namely BJP and UPA by using three existing methodologies namely Rest of the Sentence (RoS), Fixed Window Length (FWL) and Dependency Analysis (DA) with respect to sentiment expressed in online news archives.

II. RELATED WORK

Sentiment Analysis of Natural Language texts is a broad and expanding field. A text may contain both Subjective and Objective sentiments. Wiebe (1994) [2] defines Subjective text as the "linguistic expression of somebody's opinions, sentiments, emotions, evaluations, beliefs and speculations". In her definition, the author was inspired by the work of the linguist Ann Ban field (1982) [3], who defines subjective as a sentence that takes a character's point of view and that present private states (that are not open to objective observation or verification), defined by Quirk (1985) [4], of an experiencer, holding an attitude, optionally towards an object. Bing Liu (2010) [5] defines Objective text as the facts that are expressed about entities, events and their properties. Esuli and Sebastiani (2006) [6] define Sentiment Analysis as a recent discipline at the crossroads of Information Retrieval and Computational Linguistics which is concerned not with the topic a document is about, but with the opinion it expresses.

While a wide range of human moods can be captured through Sentiment Analysis Hannak (2012) [7] says majority of studies focus on identifying the polarity of a given text—

that is to automatically identify if a message about a certain topic is positive or negative.

Polarity analysis has numerous applications especially on news articles using several methods. Pang and Lee (2002) [8] broadly classifies Sentiment Analysis methods into machine-learning-based and lexical-based. Machine learning methods often rely on supervised classification approaches, where sentiment detection is framed as a binary (i.e., positive or negative). This approach requires labeled data to train classifiers [8]. While one advantage of learning-based methods is their ability to adapt and create trained models for specific purposes and contexts, their drawback is the availability of labeled data and hence the low applicability of the method on new data. This is because labeling data might be costly or even prohibitive for some tasks. On the other hand, lexical-based methods make use of a predefined list of words, where each word is associated with a specific sentiment. The lexical methods vary according to the context in which they were created [8].

Sentiment Analysis work has been handled heavily in subjective text types where the target is clearly defined and unique across the text as the case in movie or product reviews. But when applying Sentiment Analysis to the News domain, Alexandra Balahur (2009) [9] says it is necessary to clearly define the scope of the tasks in three levels. They are definition of the target; separation of good and bad news content from the good and bad sentiment expressed on the target and analysis of clearly marked opinion that is expressed explicitly, not needing interpretation or the use of world knowledge. Thus, on the same lines we built a corpus of 689 political instances from three different news databases namely, The Hindu, Times of India and Economic Times for three distinct Indian parties namely United Progressive Alliance (UPA), Telugu Desam Party (TDP) and Telangana Rashtra Samithi (TRS) to analyze the choice of certain words used in political texts to influence the sentiments of public in polls. Following is the glimpse of work done using news articles by various authors either by adopting machine learning based (MLB) or Lexical based (LB) methods with their respective accuracies.

Also, Negation and its scope in the context of sentiment analysis has been studied in the past [10]. However, others have studied various forms of negation within the domain of sentiment analysis, including work on content negators, which typically are verbs such as “hampered”, “lacked”, “denied”, etc. [10] [11]. A recent study by Danescu- Niculescu-Mizil et al. looked at the problem of finding downward call for operators that include a wider range of lexical items, involving soft negators such as adverbs “rarely” and “hardly” [13]. With the absence of a general purpose corpus annotating the precise scope of negation in sentiment corpora, many studies incorporate negation terms through heuristics or soft-constraints in statistical models. In the work of Wilson et al., a supervised polarity classifier is trained with a set of negation features derived from a list of cue words and a small window around them in the text [12]. Choi and Cardie et al. combine different kinds of negators with lexical polarity items through various compositional semantic models, both heuristic and machine learned, to improve phrasal sentiment analysis [11]. In that work [11] the scope of negation was either left undefined or determined through surface level syntactic patterns similar to the syntactic patterns from Moilanen and Pulman [10].

III. DATA SETS

The work described in this paper was part of a larger research to improve the accuracy of sentiment analysis in the daily political news present in online news archives. For our study a sample of 513 Political news opinions, dating from January 1, 2014 to March 31, 2014, were obtained from two online Indian news archives namely The Hindu [1] and NDTV [2]. This sample of 513 political news opinions contained 1675 sentences in total.

We extracted only political news pertaining to 2014 General Elections for two leading parties explicitly “UPA” and “BJP”. Each sentence was manually annotated with the scope of negation by two annotators, after achieving inter-annotator agreement of 91% with a second annotator on a smaller subset of 20 sentences containing negation.

Among 1675 sentences in the corpus according to annotator I, 1,137 were positive and 538 were negative whereas according to annotator II, 1,130 were positive and 545 were negative sentences. Inter-annotator agreement was calculated using strict exact span criteria where both the existence and the left/right boundaries of a negation span were required to match.

The obtained political news corpus was annotated with a general principle to consider minimal span of a negation covering only the portion of the text being negated semantically and according to the following instructions:

A. Negation words

Words like “never”, “no”, or “not” in its various forms are not included in negation scope. For example, in the sentence, “It was not XYZ”, only “XYZ” is annotated as the negation span.

B. Noun phrases

Typically entire noun phrases are annotated as within the scope of negation if a noun within the phrase is negated. For example, in the sentence, “The consequence of the act was not due to Modi” the string “due to Modi” is annotated. This is also true for more complex noun phrases, e.g., “People did not expect Sonia to act in such a way” should be annotated with the span “expect Sonia to act in such a way”.

C. Adjectives in noun phrases

Do not annotate an entire noun phrase if an adjective is to be negated - consider the negation of each term separately. For instance, “Not top-drawer political party, but still wins. “top drawer” is negated, but “political party” may not, since it is still party, just not “top-drawer”.

D. Adverbs/Adjective phrases

a) Case 1: Adverbial comparatives like “very,” “really,” “less,” “more”, etc., annotate the entire adjective phrase, e.g., “It was not very good” should be annotated with the span “very good”.

b) Case 2: If only the adverb is directly negated, only

[1] http://www.thehindu.com/opinion/
[2] http://www.ndtv.com/article/list/opinion/

annotate the adverb itself. e.g., "Not only was it great", or "Not quite as great": in both cases the subject still is "great", so just "only" and "quite" should be annotated, respectively.

However, there are cases where the intended scope of adverbial negation is greater, e.g., the adverb phrase "just a small part" in "Modi was on stage for the entire speech. It was not just a small part".

c) Case 3: "as good as X". Try to identify the intended scope, but typically the entire phrase should be annotated, e.g., "It was not as good as I remember".

Note that Case 2 and 3 can be intermixed, e.g., "Not quite as good as I remember", in this case follow 2 and just annotate the adverb "quite", since it was still partly "as good as I remember", just not entirely.

E. Verb Phrases

If a verb is directly negated, annotate the entire verb phrase as negated, e.g., "appear to be fair" would be marked in "He did not appear to be fair". For the case of verbs (or adverbs), we made no special instructions on how to handle verbs that are content negators. For example, for the sentence "I can't deny it was good", the entire verb phrase "deny it was good" would be marked as the scope of "can't". Ideally annotators would also mark the scope of the verb "deny", effectively canceling the scope of negation entirely over the adjective "good". As mentioned previously, there are a wide variety of verbs and adverbs that play such a role and recent studies have investigated methods for identifying them [3] [4]. We leave the identification of the scope of such lexical items as future work.

One of the freely available resources for evaluating negation detection performance is the Bio-Scope corpus [3], which consists of annotated clinical radiology reports, biological full papers, and biological abstracts. Annotations in Bio-Scope consist of labeled negation and speculation cues along with the boundary of their associated text scopes. Each cue is associated with exactly one scope, and the cue itself is considered to be part of its own scope.

Traditionally, negation detection systems have encountered difficulty in parsing the full papers subcorpus, which contains nine papers and a total of 2670 sentences.

IV. Extracting Features

A. Bag-of-Words Features

Here each feature indicates the number of occurrences of a word in the document. The news for a given day is represented by a normalized unit length vector of counts, excluding common stop words and features that occur fewer than 20 times in our corpus [16].

B. Entity Features

As shown by Wiebe et al., it is important to know not only what is being said but about whom it is said [15]. The term "victorious" by itself is meaningless when discussing an election – meaning comes from the subject.

Similarly, the word "scandal" is bad for a candidate but good for the opponent. Subjects can often be determined by proximity. If the word "scandal" and "UPA" are mentioned in the same sentence, this is likely to be bad for "Sonia Gandhi". A small set of entities relevant to the party can be defined priori to give context to features. For example, the entities "Sonia Gandhi," "Rahul Gandhi", "Dr. Manmohan Singh", "UPA" and "Congress party" were known to be relevant before the general election. News is filtered for sentences that mention exactly one of these entities. Such sentences are likely about that entity, and the extracted features are conjunctions of the word and the entity. For example, the sentence "Sonia Gandhi is facing another scandal" produces the feature "Sonia Gandhi-scandal" instead of just "scandal." Two, Context disambiguation comes at a high cost: about 70% of all sentences do not contain any predefined entities and about 7% contain more than one entity [17].

These likely relevant sentences are unfortunately discarded, although future work could reduce the number of discarded sentences using co reference resolution.

C. Dependency Features

While entity features are helpful they cannot process multiple entity sentences. These sentences may be the most helpful since they indicate entity interactions [2]. Consider the following three example sentences:

- Narendra Modi defeated Rahul Gandhi in the debate.
- Rahul Gandhi defeated Narendra Modi in the debate.
- Narendra Modi, the president of BJP, defeated Sonia Gandhi in the last night's debate.

Obviously, the first two sentences have very different meanings for each candidate's campaign. However, representations considered so far do not differentiate between these sentences, nor would any heuristic using proximity to an entity. Three effective features rely on the proper identification of the subject and object of "defeated." Longer n-grams, which would be very sparse, would succeed for the first two sentences but not the third.

To capture these interactions, sentences were part of speech tagged, parsed with a dependency parser. The resulting parses encode dependencies for each sentence, where word relationships are expressed as parent-child links. The parse for the third sentence above indicates that "Narendra Modi" is the subject of "defeated," and "Sonia Gandhi" is the object. Features are extracted from parse trees containing the pre-defined entities (as mentioned in subsection 4.2). Note that they capture events and not opinions.

V. Implementation

In the pre-processing stage, the data is cleaned to hold only what is essential for the analysis. Steps like tokenization, stop word removal, lemmatization and pos tagging were performed using NLTK and Stanford POS tagger.

To bring a comparison between lexical based and machine learning methods, we implemented SentiWordNet (SWN) [10] which is based on English lexical dictionary called WordNet [5] and two machine learning based algorithms namely Naive Bayes (NB) and Support Vector Machine (SVM) using WEKA with 10 fold cross validation [7] respectively.

The scope of negation detection is limited to explicit rather than implied negations within a single sentence.

A. Dictionary Tagging

POS tagged sentences were given as an input to the Dictionary tagger. Dictionary tagger then tags each token of every sentence with tags like positive, negative, negation (inv). SentiWordNet values were taken to tag the tokens.

B. Negation Scope Determination

The scope of negation detection is limited to explicit rather than implied negations within a single sentence. A lexicon of negations was created to identify the presence of negation in the sentence. Using a statistics driven approach, Klima et al. was the first to identify negation words by analyzing word co-occurrence with n-grams that are cues for the presence of negation [6]. Klima's lexicon served as a starting point for the present work and was further refined through the manual inclusion of selected negation cues from the corpus. The final list of cues used for the evaluation is presented in Table 1.

TABLE I. LEXICON OF EXPLICIT NEGATIONS

Hardly	Lack	Neither	Nor
Never	No	Nobody	None
Nothing	Nowhere	Not	n't
cannot	Without	Bad	Uninspired
Expensive	Disappoint	Ditch	Misunderstand

The above list of lexicon serves as a reliable signal to detect the presence of explicit negations. It does not provide any means of inferring the scope of negation. To detect the scope of negation in the sentence three different approaches were implemented.

The first approach was Rest of the Sentence (RoS). In this method, all the words that follow the negation keyword are reversed in a sentence. For this, we used the negation tags given by the dictionary tagger. For every token containing the tag, if the previous tokens contains negation tag, we reverse the polarity of the current token, append negation token to the current token so that it gets propagated to the last word of the sentence and then add it to the score value.

Algorithm for Polarity Calculation

For a sentence score
If the negation tag is identified in a sentence
Return{reverse the values of the next tokens from the negation tag and take the sum of all scores in the sentence}
Else
Return {sum of all scores in the sentence}

The second was the Fixed Window Length (FWL) approach in which we considered a fixed length of 4 words followed by a negation keyword. Every word in a sentence was tagged as positive, negative or negation by the dictionary tagger as discussed in subsection 5.1. If the tagged sentence contains negation, then we started a counter equal to the window size to reverse the polarity of the tokens next to negation till the size is attained and then the resultant was added to the score value.

Algorithm for Polarity Calculation

For a sentence score
If the negation tag is identified in a sentence
Return {reverse the value for four consecutive scores from negation tag and then add the total scores in the sentence}
Else
Return {sum of all scores in the sentence}

The third approach was Dependency Analysis (DA). Only unigram features were employed, but each unigram feature vector is expanded to include bigram and trigram representations derived from the current token in conjunction with the prior and subsequent tokens. The distance measures can be explained as follows. Token-wise distance is simply the number of tokens from one token to another, in the order they appear in a sentence. Dependency distance was more involved, and was calculated as the minimum number of edges that must be traversed in a dependency tree to move from one node (or token) to another. Each edge was considered to be bidirectional. The number 0 implies that a token was, or was part of, an explicit negation cue. The numbers 1-4 encode step-wise distance from a negation cue, and the number 5 was used to jointly encode the concept as "not applicable". To get the parse tree of the sentences, we used Stanford parser. The reason for that was in the negation identification process, the kind of negation i.e. "No one likes his behavior", where 'no' is used to determine the behavior of one, is also identified. This rocess also takes care of the negation in conjunction sentences.

The output of which was given to the nltk parse function to get the Tree object of nltk, so that traversing through the parse tree was made possible.

Having determined the scope of negation cues, the sentiment scores associated with the words in the negation keywords' scope can be inverted. To this end, we introduce unigram sentiment modifiers, which are initialized at a value of 1, indicating that the sentiment score retrieved from the sentiment lexicon is considered to be the true sentiment score associated with that word in the considered context. In case a word is negated, the sentiment modifier may be multiplied with an inversion factor. Initially, we assume this factor to be equal to −1. Finally, when all word scores have been determined while accounting for negation, sentences can be classified as either positive or negative. To this end, we use a sentence scoring function. If the sum of word-level sentiment scores in a sentence produces a number smaller than 0, the sentence is classified as negative, else, the sentence is classified as a positive sentence. We ignored those sentences whose score is 0 as we are considering only two class problem.

Algorithm for Polarity Calculation

For a sentence score
If the negation word is identified in the sentence
Return {reverse the polarities of its parent nodes and then add the total scores in the sentence}

Else
Return {sum of all scores in the sentence}

VI. Results

In order to understand the advantages, disadvantages, and limitations of the various sentiment analysis methods and analyze the choice of words from news articles in winning a particular party, we present comparison results among them which are as shown below –

A. Prediction Performance

We illustrate a comparative performance evaluation of each method in terms of correctly predicted polarity. Here we depict the results for precision, recall, accuracy, and F-measure for the three previously described methods. Table 3 shows the performance of the results obtained for each labeled dataset. For the F-measure, a score of 1 is ideal and 0 is the worst possible. Among the methods the best F-measure was SVM with the values 0.688 and 0.657 for UPA and TRS respectively and NB with the value 0.723 for TDP. On par, the accuracy of SWN, NB and SVM with the values 0.742, 0.725 and 0.666 are the highest scores in UPA, TDP and TRS respectively.

TABLE II. Predicted Performance for Labeled Dataset

Metric	BJP			UPA		
	SVM	**NB**	**SWN**	**SVM**	**NB**	**SWN**
Recall	0.79	0.62	0.58	0.70	0.75	0.50
Precision	0.60	0.75	0.86	0.69	0.71	0.54
F-Measure	0.68	0.65	0.56	0.67	0.73	0.51
Accuracy	0.72	0.62	0.74	0.70	0.75	0.61

The evaluation metric for three different methodologies was calculated separately for both the parties. The results are as shown below:

TABLE III. Results to Detect Comparison of Negation Between Three Existing Methods

Metric	UPA			BJP		
	FWL	**RoS**	**DA**	**FWL**	**RoS**	**DA**
Recall	0.69	0.53	0.72	0.68	0.56	0.77
Precision	0.61	0.51	0.69	0.64	0.59	0.79
F-Measure	0.66	0.54	0.71	0.72	0.61	0.78

From the results we can observe that DA methodology was outperforming when compared to RoS and FWL.

TABLE IV. Observations Drawn from the Results after Applying DA Scope of Negation Detection

Negation	First word/ Phrase/ clause	Second word/ Phrase/clause	Result
Present	Positive	Positive	Negative
Absent	Positive	Positive	Positive
Present	Positive	Negative	Positive
Absent	Positive	Negative	Negative
Present	Negative	Positive	Positive
Absent	Negative	Positive	Negative
Present	Negative	Negative	Negative
Absent	Negative	Negative	Positive

VII. Conclusion

In this paper, we began the comparison of three representative sentiment analysis methods like Support Vector Machine, Naïve Bayes and SentiWordNet.

Our comparison study focused on detecting the polarity of content (i.e., positive and negative affects) from good or bad news for two different Indian political parties. Thus by extracting the average predicted performance we observed that the choice of certain words used in political text was influencing the Sentiments in favor of BJP which might be one of the causes for them be the winners in Elections 2014.

We also study the concept of scope of negation (t) identification which is precisely the sequence of words affected by t. Three sets of experiments were performed to bring a comparison in identifying the scope of negation in news articles for two political parties. Experimental results show that DA method outperforms better than other two.

Acknowledgment

I am indebted to the EMR Division under HRD Group of Council of Scientific & Industrial Research (CSIR) for providing me CSIR Research Fellowship (SRF).

References

[1] Wiberg, M., "The interaction Society: Theories Practice and Supportive Technologies." Idea Group Inc. 2004.
[2] Wiebe, J., "Tracking point of view in narrative." Computational Linguistics, 20, 1994.
[3] Banfield, A. "Unspeakable sentences: Narration and Representation in the Language of Fiction." Routledge and Kegan Paul, 1982.
[4] Quirk, R., " A Comprehensive Grammar of the English Language", Longman Publishing House, 1985.
[5] Bing Liu, "Sentiment Analysis and Subjectivity", Handbook of Natural Language Processing, Second Edition, 2010.

[6] Esuli, A. and F. Sebastiani, "SentiWordNet: A Publicly Available Resource for Opinion Mining." In Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2006, Italy, 2006.

[7] A. Hannak, E. Anderson, L. F. Barrett, S. Lehmann, A. Mislove, and M. Riedewald. Tweetin'in the rain: Exploring societal-scale effects of weather on mood. In Int'l AAAI Conference on Weblogs and Social Media (ICWSM), 2012.

[8] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine Learning techniques. In ACL Conference on Empirical Methods in Natural Language Processing, pages 79–86, 2002.

[9] Alexandra Balahur, Ralf Steinberger, "Rethinking Sentiment Analysis in the News: from Theory to Practice and back", European Commission-Joint Research Centre, University of Alicante, Department of Software and Computing Systems.WOMSA, pp.1-12, 2009.

[10] Karo Moilanen and Stephen Pulman 2007. Sentiment Composition. Proceedings of the Recent Advances in Natural Language Processing International Conference Borovets, Bulgaria.

[11] Yejin Choi and Claire Cardie. 2008. Learning with Compositional Semantics as Structural Inference For Subsentential Sentiment Analysis. Proceedings of the Conference on Empirical Methods on Natural Language Processing. ACL, Honolulu, HI.

[12] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann, 2005, Recognizing contextual polarity in phrase level sentiment analysis. Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing Vancouver, Canada.

[13] Cristian Danescu-Niculescu-Mizil, Lillian Lee, and Richard Ducott, 2008, Without a 'doubt'? Unsupervised discovery of downward-entailing operators. Proceedings of The 10th Annual Conference of the North American Chapter of the Association for Computational Linguistics. ACL, Boulder, CO.

[14] Kevin Lerman, Ari Gilder, Mark Dredze, 2008 Reading the Markets: Forecasting Public Opinion of Political Candidates by News Analysis.

[15] Gunnel Tottie. 1991. Negation in English Speech and Writing: A Study in Variation Academic, San Diego, CA.

[16] Penn_Treebank, 1992, The Penn Treebank Project [WWW]. Available from: http://www.cis.upenn.edu/~treebank/ [Accessed April 21, 2012]

[17] Insaac G. Councill, Ryan McDonald, 2010, What's Great and What's Not: Learning to Classify the Scope of Negation for Improved Sentiment Analysis, Negation and Speculation in Natural Language Processing (NeSp-NLP 2010), Proceedings of the Workshop, Uppnsala, Sweden, 10 July 2010.

[18] User Manual http://iasri.res.in/ebook/win_school_aa/notes/WEKA.pdf

[19] NlTK 3.0 documentation, http://nltk.org/

An Information-Theoretic Measure for Face Recognition: Comparison with Structural Similarity

Asmhan Flieh Hassan
PhD Candidate, Mathematics & Statistics
HuaZhong Normal University
152 Luoyu Road, Wuhan, China
Dept. of Maths, Univ. of Kufa, Najaf, Iraq

Zahir M. Hussain
Computer Science & Maths, University of Kufa
P.O.Box 21, Kufa, Najaf, Iraq
Professor (Adjunct), Engineering, Edith Cowan Univ., Australia

Dong Cai-lin
School of Mathematics & Statistics
HuaZhong Normal University
152 Luoyu Road, Wuhan, China 430079

***Abstract*—Automatic recognition of people faces is a challenging problem that has received significant attention from signal processing researchers in recent years. This is due to its several applications in different fields, including security and forensic analysis. Despite this attention, face recognition is still one among the most challenging problems. Up to this moment, there is no technique that provides a reliable solution to all situations. In this paper a novel technique for face recognition is presented. This technique, which is called ISSIM, is derived from our recently published information - theoretic similarity measure HSSIM, which was based on joint histogram. Face recognition with ISSIM is still based on joint histogram of a test image and a database images. Performance evaluation was performed on MATLAB using part of the well-known AT&T image database that consists of 49 face images, from which seven subjects are chosen, and for each subject seven views (poses) are chosen with different facial expressions. The goal of this paper is to present a simplified approach for face recognition that may work in real-time environments. Performance of our information - theoretic face recognition method (ISSIM) has been demonstrated experimentally and is shown to outperform the well-known, statistical-based method (SSIM).**

Keywords—Information Theoretic Similarity; Joint Histogram; Structural Similarity (SSIM); face recognition; Image Processing

I. Introduction

Face recognition is one of the fastest-growing topics within the image processing and pattern recognition. Applications of face recognition are manifold, including access control, security and video surveillance, credit card user identification, forensic analysis, entertainment, and automatic video indexing. These applications have made this topic a very popular research area in the last three decades; see [1, 2, 3] for surveys. In recent years, several approaches to face recognition have been developed. References [4, 5] give an overview of face recognition techniques.

Methods used for face recognition have improved significantly, thanks to the efforts of hundreds of researchers and funding bodies, causing its impact to spread rapidly to a variety of applications. In parallel with thousands of published papers on the topic, many useful datasets are created and used successfully for performance evaluation of various approaches.

Despite its huge literature, automatic face recognition is still a difficult problem. This difficulty is mainly due to artifacts that obscure the image features, like variations in illumination, facial expression, and head pose. Techniques that may offer efficient feature extraction with a high discrimination power and low computational complexity are crucial [6]. According to the approach of considering spectral analysis, face-recognition methods maybe divided into two main categories: visible (VIS) and infrared (IR).

Image similarity measurement could be a basic issue in each computer and human vision system, including many real-world applications. In image recognition applications, one usually has to realize the similarity between two images, i.e., a test image and a training database image. Image recognition has become a subject of great interest to researchers over the past decades because of its potential applications in several fields like Optical Character Recognition (OCR), identity authentication, human-computer interfacing, and surveillance. A variety of methods for image recognition, especially for face image recognition have been proposed [3]. In References [7-13], methods for image recognition have been classified as holistic methods, feature-based methods, and hybrid methods.

Several matching algorithms of face and object recognition systems have been designed based on image similarity measurement like SSIM [14].

This work presents a novel approach for face recognition based on information – theoretic approach rather than the statistical approach of SSIM. The new measure, which is called in this paper as ISSIM, is derived from our information-theoretic similarity measure (HSSIM) proposed in [15]. HSSIM has been modified to meet the requirements of feature extraction for the purpose of face recognition.

II. Background

Automatic face recognition systems attempt to recognize the identity of a given face image in comparison with images that they have in their memory. In most face recognition works, this memory of a face recognizer is represented by a training set saved as a database. In this work, our training set is the AT&T well known face database [16]. Thus, the task of the face recognizer is finding most similar feature vector among the training set to the feature vector for a given test image.

Simulation results showed that the proposed approach (ISSIM) presents higher performance than the standard SSIM in face recognition.

In [17], modified versions of SSIM are proposed in an effort to be utilized in recognition. In [18], instead of the usual approach for applying statistics or structural methods only, the Authors proposed a methodology that integrates higher-order illustration patterns extracted by Zernike moments with a modified version of SSIM (M-SSIM). Individual measurements and metrics resulted from mixed SSIM and Zernike-based approaches provide a powerful recognition tool with great results.

There are several similarity measures that are proposed and used for varied purposes, see [19-26]. Several methods was described and applied using entropy types to handle the face recognition and edge detection problems [27-31].

The rest of this paper is organized as follows: Section III presents the mathematical foundations of statistical and information theoretic measures. Section IV describes the design of novel information - theoretic measure ISSIM to be used in face recognition. Section V shows experimental results and discusses possible modifications and improvements to the system. Section VI presents concluding remarks.

III. SIMILARITY MEASURES

A similarity measure (between two sets of data points) is defined as the distance (based on a specific norm) between various data points.

The performance of any algorithm that detects similarity is heavily dependent upon choosing a good distance function over an input data set. While similarity is an amount that reflects the strength of relationship between two database items, dissimilarity deals with the measurement of divergence between two data items.

In this work an information - theoretic measure is proposed and its performance is analyzed in comparison with a well-known statistical measure.

A. SSIM Overview: A Structural Similarity Measure

Automatic similarity finding is an important approach in many image processing systems like those used for compression, enhancement, identity check, etc. An important achievement in image similarity was the proposal of a statistical image quality measure in 2004 [14]. The measure, well-known as SSIM, has been promising in many applications. It has been widely used for image quality assessment and many algorithms of image processing systems. The technique used in SSIM is based on using statistical measurements like mean and standard deviation to find a definition for a distance function that can measure the structural similarity between a test image and a training image. The measure has been put into the form:

$$\rho(x,y) = \frac{(2\mu_x\mu_{y+}C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \qquad (1)$$

where $\rho(x,y)$ is the SSIM measure of statistical distance (or similarity) between the image x (the test image or reference image) and (normally a corrupted version of x) image y. The statistical parameters μ_x, μ_y, σ_x^2, and σ_y^2 are the means and variances of pixel (illumination level) values in the two images x and y, while σ_{xy} represents the statistical covariance between images x and y. The constants C_1 and C_2 are given as follows: $C_1 = (K_1L)^2$ and $C_2 = (K_2L)^2$, with K_1 and K_2 are small constants, and $L = 255$ (the maximum pixel value). The Authors in [14] confirmed that these parameters (constants) have little effect on SSIM performance.

In [17], it has been shown that this approach gives a good level of similarity under noise free conditions, while its performance falls when noise increases. In addition, it may give a non-negligible similarity between two unrelated images. This is due to the fact that SSIM is totally dependent on the statistical features of the two images, which may have some hidden correlations. In [17] and [30], SSIM has been combined with edge detection filters (such as Canny's) to produce excellent results especially when the images are different from each other.

B. An Information-Theoretic Technique for Image Similarity

The use of information - theoretic analysis in image processing is possible if one imagines that images are 2D random variables.

The joint – histogram between two images is defined as the joint occurrence of intensity levels (pixel values) in the two images. Note that only considering a gray image is considered here, as grey version of an image contains most of the similarity information. For two images under similarity test A and B, assume that the two images have the same size and that the pixel intensity values are denoted by i and j, respectively. Note that both i and j range from 0 to $L = 255$. Now the joint histogram, denoted by $H_{ij}(A,B)$, represents the joint probability of each pixel (defined by its 2D location, as the two images have the same size) to take on the value i on the first image A and the value j on the second image B. Now every entry in the joint histogram will represent the number of times the intensity level i in any of the two images corresponds to the intensity level j in the other image.

In [15], a similarity error estimate between symmetrically – located entries of the joint histogram (around y=x relation) has been designed as follows:

$$E(x,y) = \sqrt{\frac{\sum_i \sum_j \left[(H_{ij}-H_{ji})\frac{1}{h_i+c}\right]^2}{2L^2}} \qquad (2)$$

where H_{ij} and H_{ji} are symmetrically- located entries of the joint histogram between two images, and h_i is the normal image histogram of the reference image histogram. The constant c is a very small positive constant, inserted principally to avoid division by zero, and $L = 255$ is the maximum pixel value.

It is easy to show the above relation is non-negative, that is:

$$E(x,y) \geq 0 \quad (3)$$

To be used as a similarity measure and compared with SSIM, a normalization process is necessary to ensure that the range of the measure is kept inside [0,1]. In [15], maximal error estimate $E_\infty(x,y)$ was used for normalization as follows:

$$e_o(x,y) = \frac{E(x,y)}{E_\infty(x,y)} \quad (4)$$

The normalization process will ensure that:

$$0 \leq e_o(x,y) \leq 1 \quad (5)$$

Based on the above error estimate an information - theoretic similarity measure (which was named as HSSIM) was proposed as follows:

$$\lambda_o(x,y) = 1 - e_o(x,y) \quad (6)$$

where:

$$0 \leq \lambda_o(x,y) \leq 1 \quad (7)$$

If one refers to the well-known SSIM by $\rho(x,y)$, then a similar inequality between 0 and 1 is obtained as follows:

$$0 \leq \rho(x,y) \leq 1 \quad (8)$$

Note that $\rho(x,y) = 1$ for totally similar (identical) images, while $\rho(x,y) = 0$ for totally dissimilar images. In practical applications for face and image similarity, SSIM and HSSIM adopt values above zero and less than one.

In this work the intention is to design a measure for face recognition based on HSSIM. However, a change in HSSIM structure is necessary as follows.

IV. A Face Recognition Approach Based on Joint Histogram

In [18], SSIM was utilized for face recognition between a reference face image x and images $\{y_{pr}\}$ for a number P of people, each person (whose database number is p such that $p \leq P$) has R face images with different poses (each pose is indexed by the variable r); hence a total of $N = R \cdot P$ face images in the database. One of the poses is chosen with a number $r = a$ that corresponds to the person p to serve as a reference image. Then two tests are performed: the first keeps the a^{th} pose in the database then starts recognition. This is an easy task, but helps in testing and comparing the performances of the proposed measures. The second test removes the a^{th} pose and starts recognition process, which is a hard task since different poses may be considered as different images.

This section describes a face recognition algorithm using the information - theoretic error estimate based on joint in (2) above.

Let us define a new error measure as follows. After running the test, define the maximum information – theoretic error for the whole test as E_m as follows:

$$E_m = \max_{p,r}\{E(x,y_{pr})\} \quad (9)$$

Now define the information – theoretic error measure $e(x,y)$ as follows:

$$e(x,y) = \frac{E(x,y)}{E_m} \quad (10)$$

This *database-dependent* (or, more accurately, *test-dependent*) normalization process will ensure that:

$$0 \leq e(x,y) \leq 1 \quad (11)$$

After normalization, an information - theoretic similarity measure for face recognition (which will be called henceforth as ISSIM) can be proposed as follows:

$$\lambda(x,y) = 1 - e(x,y) \quad (12)$$

where it is evident that:

$$0 \leq \lambda(x,y) \leq 1 \quad (13)$$

Hence, the new measure is well-defined.

For a comparison between a reference face x and a person p in the database, ISSIM is defined as follows:

$$\lambda(x,p) = \max_r\{\lambda(x,y_{p,r})\} \quad (14)$$

Here, the value of $\lambda(x,p)$ will indicate how much confidence one can put in the recognition process. Of course, a value of $\lambda(x,p) = 1$ is obtained only in the *ideal case* where the reference face pose exists in the database. Recognition is based on the following search:

$$p_i = \arg[\max_p\{\lambda(x,p)\}] \quad (15)$$

For SSIM-based statistical face recognition, one can define a similar process to find the best match as follows:

$$\rho(x,p) = \max_r\{\rho(x,y_{p,r})\} \quad (16)$$

with the recognition decision based on the following search:

$$p_s = \arg[\max_p\{\rho(x,p)\}] \quad (17)$$

Recognition Confidence

In most cases SSIM and ISSIM agree such that:

$p_i = p_s$.

However, the confidence in recognition varies significantly. By "confidence in recognition" it is meant that there is a "good" distance between the peaks defined in (15) and (17) and the next-in-height peaks who can confuse the decision when two peaks (using the same measure) are nearby in magnitude.

Now define the confidence in recognizing the test image x using a measure μ. First, find the second peak (maximum) in the above measure $\{\mu(x,p)|p = 1,\cdots,P$. Then, find the difference between the maximum and the second maximum for each measure, which is called here the MM – difference, denoted by $\partial(\mu,x)$, which can be $\partial(\rho,x)$ or $\partial(\lambda,x)$. The resulting quantity is the MM-difference for that measure.

Confidence of recognition of reference image x using a similarity measure μ is defined as:

$$\partial(\mu,x) = \max_p\{\lambda(x,p)\} - \max_p 2\{\lambda(x,p)\} \quad (18)$$

where $\max2_p\{\lambda(x,p)\}$ is the second maximum of the curve of similarity between the reference image x and other persons in database (note that all poses for each person are exhausted in $\mu(x,p)$ as per (14)). ISSIM outperforms SSIM as one has for most cases:

$$\partial(\lambda, x) > \partial(\mu, x) \quad \text{for most } x \tag{19}$$

If ∂ is extended to other peaks then the average of ISSIM distance is always larger:

$$\text{mean } \partial(\lambda, x) > \text{ mean } \partial(\mu, x) \quad \forall\, x \tag{20}$$

Hence, ISSIM gives a clearer decision in most cases. In easier cases, this difference can reach up to 0.5, equivalent to half the value of the total measure, which is 1. Of course, one can get even clearer decisions when the reference image is included in the database, although this case is not realistic.

V. Experimental Results and Performance

A. Image Database

A well-organized face image database was created by AT&T [16] and has been widely used for testing face recognition systems. This database is divided into two subgroups, for separate training and testing purposes. Throughout training, 49 images were used, containing seven subjects and every subject having 7 images with different facial expressions. A sample is shown in Figure (1) below. Note that in our work, the dimensions (size) of every face image are modified to be 92×92 pixels.

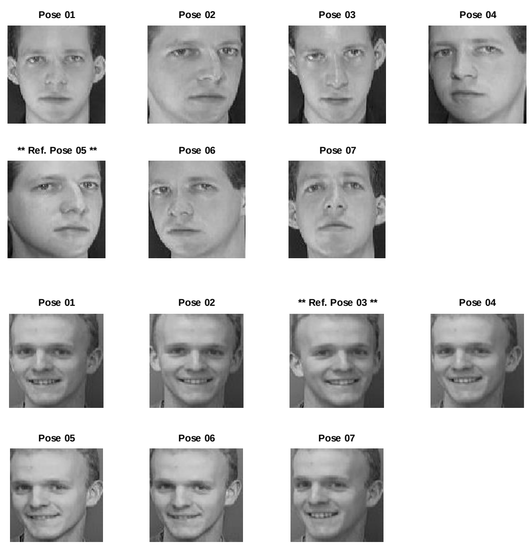


Fig. 1. Examples of the training and testing images taken from the well-known AT&T face image database

B. Testing

The face recognition system presented in this paper was developed, trained, and tested using MATLAB. The performance of the proposed information – theoretic algorithm ISSIM is tested and then compared with the well-known structural similarity measure SSIM. The results illustrate the efficacy of ISSIM for face recognition. The algorithm has been tested using two cases as follows:

Case 1- Test Image Included in the Database

If the test image (the reference pose) represents an image that belongs to the training dataset, then the program utilizes all poses related to the person under test.

In this testing case the recognition is high for both SSIM and ISSIM. Although less realistic than the next case, it can be used to compare the power of the proposed ISSIM as compared to SSIM in performance for face recognition. ISSIM outperforms SSIM by far.

Case 2- Test Image Excluded from the Database

If the test image is not included in the data set, the recognition is still very good and ISSIM still performs better than SSIM in recognizing the face despite the difference in poses.

To compare performance, the distance between he first and second maxima of the similarity curve are utilized for each recognition method. The more this distance the more confident is the decision of recognizing the person.

The Figures below show recognition results and performance analysis of the proposed ISSIM.

C. Discussion

Figures (2) and (3) present poses for two persons from the database, used for testing. The reference pose, considered as test image, is indicated.

A pose with no significant angle is chosen for the purpose of testing and comparison, although eye-glasses are considered in the second test (see Figure (3)), which increases the difficulty of recognition.

Figures (4) - (7) show results of similarity test between the reference image and other people in the database, where, for each person, the maximum similarity among all of his / her poses is considered.

Note that, for almost all cases, the proposed ISSIM keeps larger difference between the test image and other persons.

Figures (4) - (5) keeps the test pose as part of the database, giving an ideal maximum of 1 for both approaches.

Figures (6)-(7) are more realistic as the test excludes the image under test from the database. In this case the maximum similarity is less than 1 for both approaches, but ISSIM gives much better results.

Figures (8)-(9) show the performance measure for both approaches, where the proposed ISSIM keeps larger difference than SSIM. Hence, according to (18) and (19), better confidence in the recognition process is obtained.

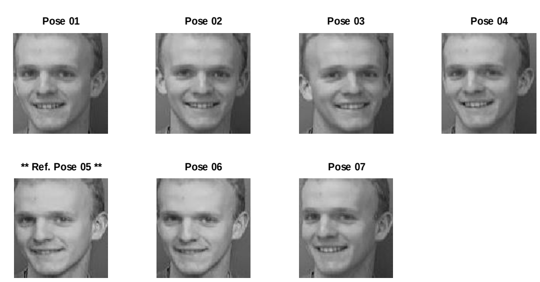


Fig. 2. Poses for Person no.5 in database. Reference Pose is indicated

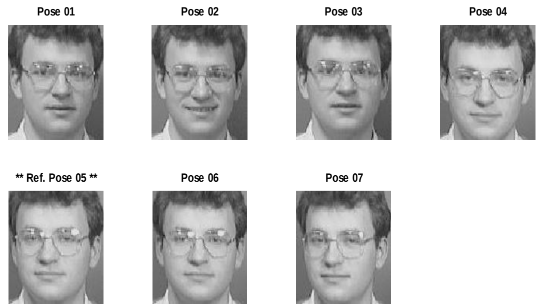


Fig. 3. Poses for Person no.6 in database. Reference pose is indicated

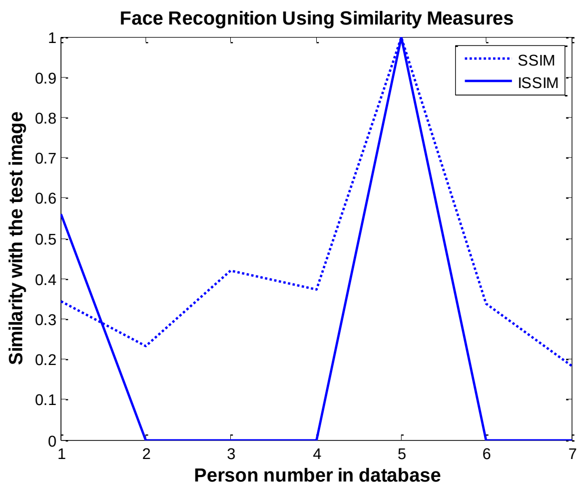


Fig. 4. Face recognition using the proposed ISSIM and the well-known SSIM with a test image as a pose belonging to person no. 5. Reference pose (as indicated in Figure (2)) is included in the database during the test

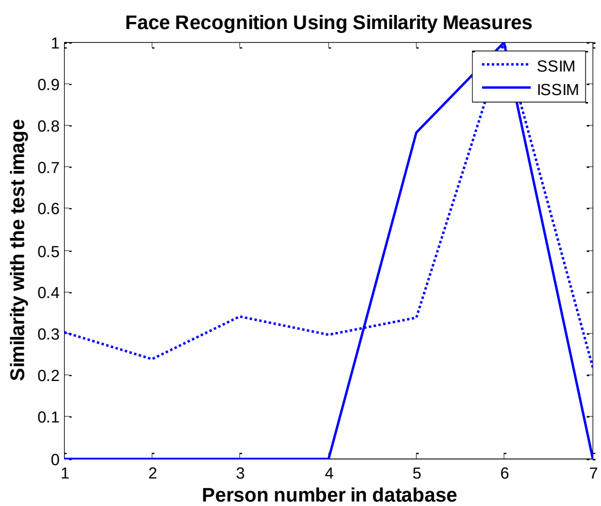


Fig. 5. Face recognition using the proposed ISSIM and SSIM with a test image as a pose belonging to person no. 6. Reference pose (as indicated in Figure (3)) is included in the database during the test

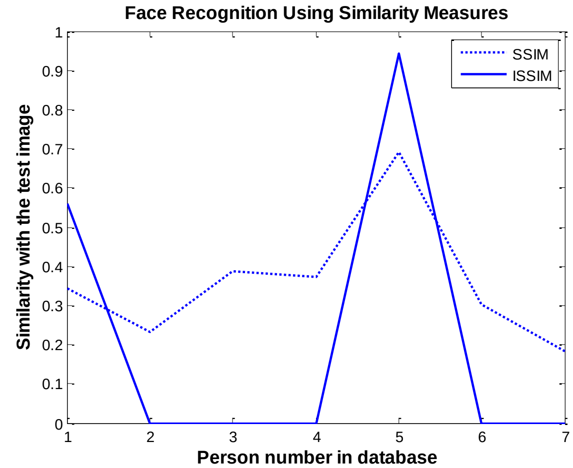


Fig. 6. Face recognition using the proposed ISSIM and SSIM with a test image as a pose belonging to person no.5. Reference pose (as indicated in Figure (2)) is excluded from the database in the test

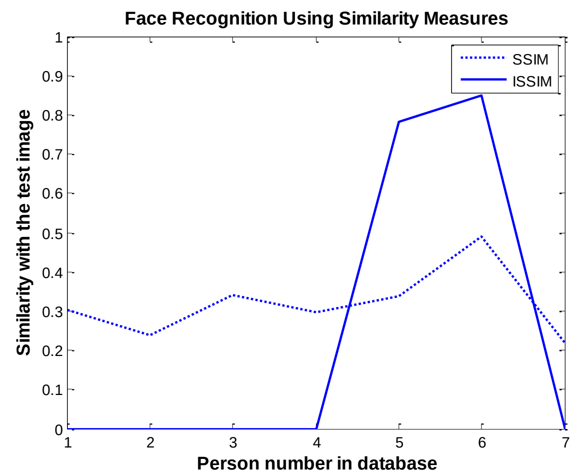


Fig. 7. Face recognition using the proposed ISSIM and SSIM with a test image as a pose belonging to person no. 6. Reference pose (as indicated in Figure (3)) is excluded from the database during the test

Fig. 8. Recognition confidence using the proposed ISSIM and SSIM with a test image as a pose belonging to person no. 5. Reference pose x (as in Figure

(2)) is excluded from the database during the test. It is clear that ISSIM $\lambda(x,p)$ gives more confidence than SSIM $\mu(x,p)$ as $\partial(\lambda,x) > \partial(\mu,x)$

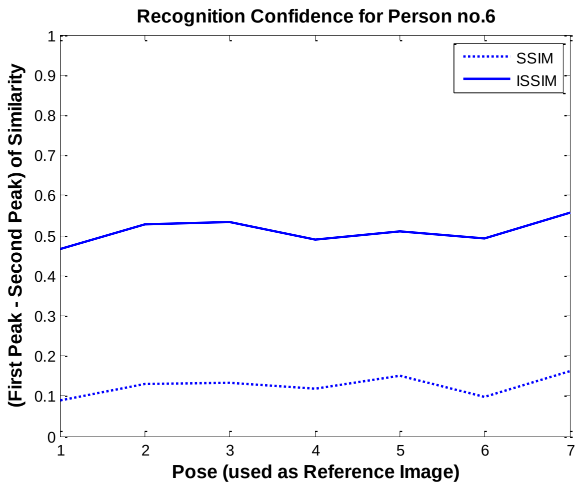


Fig. 9. Recognition confidence using the proposed ISSIM and SSIM with a test image as a pose belonging to person no. 6. Reference pose x (as in Figure (3)) is excluded from the database during the test. It is clear that ISSIM $\lambda(x,p)$ gives more confidence than SSIM $\mu(x,p)$ as $\partial(\lambda,x) > \partial(\mu,x)$.

VI. Conclusions

This paper presents a novel face recognition technique (called ISSIM) that uses information - theoretic similarity measure based on joint histogram. This measure utilizes the fact that similarity between two images would result in more symmetry along y=x curve in the joint histogram. Divergence from this symmetry is considered as a measure of dissimilarity.

The system was simulated numerically using on MATLAB using part of AT&T image database that consists of 49 face images, containing seven subjects with every subject having seven poses with different facial expressions. Results showed that ISSIM outperforms the well-known SSIM when used in face recognition.

The main performance measure that was considered here (and used for comparison purposes) is the distance between the maximum similarity found by the specific approach and the second maximum, which can be more confusing in the recognition task if it takes smaller values. As no more complications are used in the proposed system, it is evident that this system is well-suited for low-cost, real-time hardware or software implementation.

In this work, global face analysis is applied, where the whole image is treated at once. While good results are obtained under a standard database, difficulties may arise in practice. Local analysis for face images proved to play a significant role in improving face recognition.

The Authors intend to pursue this point in future works and extend their previous studies on local analysis to improve the performance of the above defined measure.

References

[1] W. A. Barrett, “A survey of face recognition algorithms and testing results,” *Systems and Computers*, 1998.

[2] T. Fromherz, P. Stucki, M. Bichsel, *A Survey of Face Recognition*, MML Technical Report, 1997.

[3] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, “Face recognition: a literature survey,” *ACM Computing Surveys*, 35(4), 2003.

[4] R. Chellappa, C. L. Wilson, and S. A. Sirohey, “Human and machine recognition of faces: a survey,” *Proceedings of IEEE*, 83(5), 1995.

[5] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, “Face recognition: a convolutional neural network approach,” *IEEE Transactions on Neural Networks*, 8(1), 1997.

[6] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” *IEEE Trans. Circuits Systems for Video Technology*, 14(1), 2004.

[7] M.Turk, A.Pentland, “Eigenfaces for recognition,” Journal of Cognitive Neuroscience, 3(1), 1991.

[8] W. Zhao, R. Chellappa, A. Krishnaswamy, “discriminant analysis of principal components for face recognition,” *Proceedings of the 3rd IEEE International Conference on face and Gesture Recognition*, FG’98, Nara, Japan, 1998.

[9] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: recognition using class specific linear projection,” IEEE Transactions on Pattern Analysis & Machine Intelligence (PAMI), Vol.19, 1997.

[10] T. Kanade, *Computer Recognition of Human Faces*, Birkhauser, Basel, Switzerland, and Stuttgart, Germany, 1977.

[11] A. V. Nefian, M. H. Hayes, “Hidden Markov models for face recognition,” *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol.5, 1998.

[12] L. Wiskott, J. M. Fellous, N. Kruger, C. Malsburg, “Face recognition by elastic bunch graph matching,” *IEEE Transactions on Pattern Analysis & Machine Intelligence (PAMI)*, Vol.19, 1997.

[13] J. Huang, B. Heisele, V. Blanz, “Component-based face recognition with 3d morphable models,” *Proceedings of the 4th International Conference on Audio and Video Based Biometric Person Authentication (AVBPA)*, Guildford, UK, 2003.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, 13 (4), 2004.

[15] A. F. Hassen, Z. M. Hussain, D. Cai-Lin, “An information - theoretic image quality measure: comparison with statistical similarity," *Journal of Computer Science*, in press, 2014.

[16] AT&T Laboratories, *The Database of Faces*, Cambridge [online], ©2002 [accessed 10/09/2014]. Available from: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[17] A. N. Hashim and Z. M. Hussain, “Novel image dependent quality assessment measures,” *Journal of Computer Science*, 10 (9): 1548-1560, 2014.

[18] A. N. Hashim and Z. M. Hussain, “Local and semi-global feature-correlative techniques for face recognition,” *International Journal of Advanced Computer Science and Applications*, 5(8), 2014.

[19] F. G. Ashby, “towards a unified theory of similarity and recognition,” *Psychological Review*, 95(1), 1988.

[20] R. Jain, R., S. N. J. Murthy, P. L-J., Chen, S. Chatterjee, “Similarity measures for image databases,” *IEEE Int. Conference on Fuzzy Systems*, V.3, 1995.

[21] S. Santini, R. Jain, “Similarity is a geometer,” *Multimedia Tools and Applications*, 5(3), 1997.

[22] S. Santini, R. Jain, “Similarity measures,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9), 1999.

[23] J. Williams, N. Steele, N. Deference, “Distance and similarity ASA basis for fuzzy decision support based on prototypical decision classes,” *Fuzzy Sets and Systems*, V.131, 2002.

[24] A. K. Patidar, J. Agrawal, N. Mishra, “Analysis of different similarity measure functions and their impacts on shared nearest neighbor clustering approach,” *International Journal of Computer Applications* , 40(16), 2012.

[25] A. Huang, “Similarity measures for text document clustering,” *NZCSRSC 2008*, Christchurch, New Zealand, April 2008.

[26] K. Taghva and R. Veni, “Effects of Similarity Metrics on Document Clustering,” *Seventh International Conference on Information Technology*, 2010.

[27] Y. D. Zhang and L. N. Wu, “Pattern recognition via PCNN and Tsallis entropy,” *Sensors Journal*, 8, 2008.

[28] B. H. Shekar, M. S. Kumari, L. M. Mestetskiy, N. F. Dyshkant, “Face recognition using kernel entropy component analysis,” *Neurocomputing Journal*, 74(6), 2011.

[29] M. A. El-Sayed, S. Abdel-Khalek, E. Abdel-Aziz, “Study of efficient technique based on 2D Tsallis entropy for image thresholding,” *International Journal on Computer Science And Engineering (IJCSE),* 3(9), 2011.

[30] M. A. El-Sayed, “A new algorithm based entropic threshold for edge detection in images,” *International Journal of Computer Science*, 8(1), 2011.

[31] M. A. El-Sayed, T. Abd-El Hafeez, “New edge detection technique based on the Shannon entropy in Gray level images,” *International Journal on Computer Science and Engineering (IJCSE),* 3(6), 2011.

A Multistage Feature Selection Model for Document Classification Using Information Gain and Rough Set

Mrs. Leena. H. Patil
Research Scholar, Dept. of Computer Sci. and Engg.
Sant Gadge Baba Amravati University
Amravati, India

Dr. Mohammed Atique
Associate Professor, Dept.of Computer Sci.and Engg.
Sant Gadge Baba Amravati University
Amravati, India

***Abstract*—Huge number of documents are increasing rapidly, therefore, to organize it in digitized form text categorization becomes an challenging issue. A major issue for text categorization is its large number of features. Most of the features are noisy, irrelevant and redundant, which may mislead the classifier. Hence, it is most important to reduce dimensionality of data to get smaller subset and provide the most gain in information. Feature selection techniques reduce the dimensionality of feature space. It also improves the overall accuracy and performance. Hence, to overcome the issues of text categorization feature selection is considered as an efficient technique. Therefore, we proposed a multistage feature selection model to improve the overall accuracy and performance of classification. In the first stage document preprocessing part is performed. Secondly, each term within the documents are ranked according to their importance for classification using the information gain. Thirdly rough set technique is applied to the terms which are ranked importantly and feature reduction is carried out. Finally a document classification is performed on the core features using Naive Bayes and KNN classifier. Experiments are carried out on three UCI datasets, Reuters 21578, Classic 04 and Newsgroup 20. Results show the better accuracy and performance of the proposed model.**

Keywords—Introduction; Document Preprocessing; Information Gain; Rough Set; Classifiers

I. INTRODUCTION

In the field of data mining, it has been observed that the data grow rapidly. With the rapid growth of data and the availability an increasing number of electronic documents, the task of classification becomes a key method [1]. Document preprocessing is an important parameter and feature selection is a common problem used in preprocessing for machine learning, data mining and pattern recognition [1][2]. Text categorization has always been a hot topic due to explosive growth of digital documents available. Due to huge development information acquirement and storage, tens, hundreds and even thousands of features [16] are acquired and stored in real world databases. Storing and processing relevant or irrelevant attributes becomes computationally very expensive and impractical [16]. A major problem of text categorization is its high dimensionality of features, due to which it misleads to the classifier [8]. The computational complexity of machine learning methods used for text categorization be increased and may bring about inefficient and results of low accuracy due to redundant or irrelevant terms in the feature space [6][14]. Mostly it is important to reduce dimensionality of the data to smaller set of features and relevant information for decreasing the cost in storing and reduction in the processing time [6], [13]. To overcome this, few attributes can be omitted, which will not seriously affect on classification accuracy. Many techniques in feature selection have been categorized namely filter and wrapper. The former employs to select attributes according to some significance measures such as consistency [4], information gain [3], distance [5], dependency [6] and others, later employs a learning algorithm to evaluate the attribute subsets. Rough set theory proposed [13] as a tool to organize conceptualize and analyze various types of data from knowledge discovery. It is useful in dealing with uncertainty and vague, knowledge information system. Rough set theory with attribute reduction offers a systematic framework for [15] distance based measures which attempt to retain the ability of original features for the objects from the universe. In a wide range of text categorization many feature selection methods are used [19]. Information gain is considered as the most effective method compared to other methods such as term strength, mutual information, X2 statistic, document frequency [24].

The task of text categorization is to classify the documents into predefined categories based on the contents of document [24]. Many methods have been applied to text categorization task on machine learning, such as KNN, Naive Bayes, C4.5 and SVM[14][15]. Several dimension reduction techniques like PCA, GA, IG [19] are carried out; still the problem of time complexity and text categorization[16][24] can be improved. Hence, in this we proposed multistage approaches: document preprocessing, feature selection and reduction technique which are used to reduce the high dimensionality of feature space. It removes the redundant and irrelevant attributes and thereby decreases the computational complexity of the machine learning process and increases the performance of classification. In the first stage documents are preprocessed with various steps. In the second stage, information gain is used to rank the importance of the features. In third stage Rough set approach is used to reduce the attributes. Finally, to evaluate the effectiveness of dimension reduction methods, experiments are conducted on Reuters-21,578, Classic 04 and

Newsgroup 20 dataset collection. For overall accuracy and performance the different classifiers like KNN and Naive Bayes are used. The results show that the proposed model is able to achieve high categorization effectiveness as measured by precision, recall and F-measure.

II. Proposed Multistage Model

Figure 1 shows the outline of proposed multistage model for document classification. Preceding sections describes the different stages of proposed multistage model.

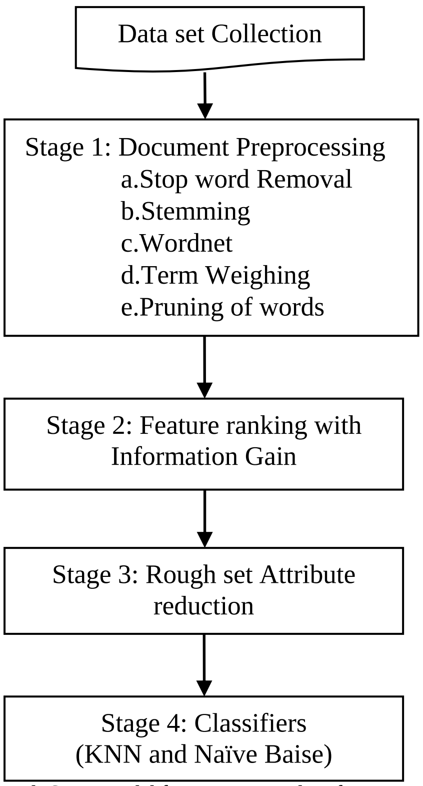


Fig. 1. Proposed MultiStage Model for Document classification

A. Document Preprocessing

To browse thousands of documents easily, document preprocessing becomes a most important tread. It articulates the required transformation processing to obtain the selected representation of documents. Thousands of words are present in a document set therefore; the aim of this is to reduce dimensionality to have better accuracy for classification [20]. Document preprocessing is divided into following stages:

- Stop words removal: Stop word List is used that contains the words to be expelled. The Stop word list is applied to remove terms that have a special meaning, but do not separate for topics.
- Word Stemming: Stemming algorithm such as porters is used to shrink a word to its stem or root form.
- WordNet: WordNet Senses Disambiguation is applied as an English Database.
- Global Unique words and frequent word set gets generated.

1) Stop-Word Removing

Stop-words remove the non-information behavior words from the text documents and reduce noisy data. To categorize large amount of word documents, stop word removal pay the similar advantages. Firstly, it could save an enormous amount of space. Secondly, it reduces the noise and keeps the core words, and makes later processing more effective and efficient.

2) Word Stemming

This process is used for transforming the words into their stem. In many languages the various syntactic forms of words are used and explicate. The most important technique called stemming is used for the reduction of words into their root. Many words from the English language can be reduced to their base form or stem, e.g. searches, searching belong to root stem search. An algorithm Porter Stemmer is apposite to stem documents. It is really a relatively accurate.

3) Wordnet

WordNet is a linguistic English Database developed in cognitive science laboratory [20][22]of Princeton University. It organizes words into a group called sysnets. Each of these contains a collection of synonymous words and corresponds to a concept. Therefore, WordNet is considered to be an English thesaurus database which maps their concepts. WordNet has four category noun, verbs, adjectives and adverbs. Using lexical database the WordNet approach measures the relatedness of terms from the words. WordNet as a dictionary covers some specific terms from every subject related to their terms. WordNet as a lexical database record all the stemmed words from the standard documents into their specific lexical categories.

4) Term weighing

Words are converted into terms, and thereby have to be measured the appearance of terms. Such processing is defined as term weighing. Therefore, each document depends on the term vector form they contained. The document vector is in the following format:

$$d = \{ w_1, \ldots, w_i \ldots\ldots, w_{|T|} \},$$

Where w_i is the weight of the term with number i in the document, T is the term set and |T| is the cardinality of T.

The $tf - df$ is used for its weighing scheme to obtain the term vector T. $tf - df$ (term frequency x document frequency) is represented by $tfdf_{ij}$ and evaluated for the value calculated by dividing the term frequency (TF) by the document frequency (DF), where TF is the number of times a term t_j appears in a document d_i divided by the total frequencies of all terms in d_i , and DF is used to determine the number of documents containing term t_j divided by the total number of documents in the document set D.

$$tfdf_{ij} = \frac{TF}{DF} ,$$

where $TF = \frac{f_{ij}}{\sum_{j=1}^{m} f_{ij}}$

and $DF = \frac{|\{d_i\, t_j \in d_i \in D\}|}{|D|}$

5) Pruning of words

The pruning process basically filters less frequent features in a document collection. The term vector is very high-dimensional and sparse. Also, it's seen that a number of elements in the term vector is ' '0' ' . Hence, it is therefore required to prune the words those appear less than two times in the documents. This procedure shrinks the term vector dimension further.

B. Feature Selection

Feature selection (FS) is a term given to the problem of selecting input attributes which are most predictive of a given outcome. The main objective of feature selection is to find the minimal feature subset from the problem domain which retain the suitable features, representing with high accuracy. Feature selection techniques are categorized into two forms: filter and wrapper. The filter method is used to select attributes according to their significance measures such as consistency [4], information gain [3], distance, dependency and others. The wrapper method employs a learning algorithm to evaluate the attribute subsets. The Significance measures are categorized into two: consistency based measure and distance based measure**.** From the original set of attributes the feature selection process chooses the subset of attributes. The main aim of feature selection recognizes the relevant features and [14][17] abolishes the irrelevant of dispensable features. The feature selection molded by two steps: First one rank the feature according to their importance and secondly, reduce the attributes. Hence, thereby decreases the computational time and increases the accuracy.

1) Feature Ranking with Information Gain

Many feature selection methods are successfully used for text categorization. [24] has compared five different feature selection methods like information gain, X2 statistic document frequency, term strength, and mutual information. They reported that information gain is the most effective method as compared with the other feature selection methods. [16] has presented that information has become one of the most popular approaches employed as a term importance criteria in the text document. [19] has presented that the information gain is based on information theory. [17] has proposed that before attribute reduction each term within the text are ranked depending on their importance for the classification. The terms are arranged in decreasing order using information gain. With this process for classification term of less importance are removed and terms of highest importance are identified, where attribute reduction methods are applied.

[9] has presented that a major problem of text classification is the high dimensionality of feature space and redundant terms. Therefore it is desirable to find some methods which can reduce attributes for improving the overall performance of classification. To solve such issue Information Gain was proposed which defines the expected reduction in entropy caused by partitioning the text according to the term. The Information Gain of term t is defined as:

$$Information\ Gain\ (t) = \sum_{i=1}^{|c|} P(C_i) log\ P(\ C_i) + P(t) \sum_{i=1}^{|c|} P(C_i|t) log P(C_i|t) + P(\bar{t}) * \sum_{i=1}^{|c|} P(C_i|\bar{t})\ log\ P(\ C_i|\bar{t})$$

Where C_i represents the ith category. $P(C_i)$ is the probability of the i[th] category. $P(t)$ and $P(\bar{t})$ are the probabilities that the term t appears or not in the documents respectively $P(C_i|t)$ is the conditional probability of the ith category given that term t appeared and $P(C_i|\bar{t})$ is the conditional probability of the ith category given that the term t does not appear. Before attribute reduction each term within the text are ranked depending on their importance for the classification. The terms are arranged in decreasing order using information gain. With this process for classification term of less importance are removed and terms of highest importance are identified, where the attribute reduction methods are applied. Feature selection is the dimensionality reduction technique where the dimension space is reduced by selecting the best features which represents the document and inputting it to the classifier

2) Feature Selection using Rough set

After document preprocessing the IG method is applied where the terms of high importance in document are acquired. Through IG the number of terms in the document is reduced but still the problem is high dimensionality of feature space for text categorization. Hence to reduce the dimensionality and time complexity used for text categorization and to increase the performance, the attribute reduction based on rough set is carried out. The purpose of this method is to minimize the information loss and maximize the reduction in dimensions. In 1982, Pawlak introduced the concept of Rough set theory [13][14]. The theory initially developed for a finite universe of discussion in which the knowledge base is a partition, obtained by any equivalence relation. In rough sets theory, the data is organized in a table called decision table. Rows of the decision table correspond to objects, and columns correspond to attributes. In the data set, a class label to indicate the class to which each row belongs. The class label is called as decision attribute, the rest of the attributes are the condition attributes. [13][14] has developed a mathematical tool of rough set theory.

Definition 1: (Decision table). A decision table is an ordered tuple $S = < U, A, V, f >$, where $U = \{x_1, x_2, \ldots, x_n\}$ is a finite set of objects; $A = C \cup D$ is a finite set of attributes, where C is a set of condition attributes, D = {d} represents the decision attribute (or class label), $C \cap D = \emptyset$; $V = \bigcup_{a \in A} V_a$, where V_a denotes the domain of attribute a; $f: U \times A \to V$ is an information function which associates a unique value of each attribute with every object belonging to U, such that for any $x \in U\ and\ a \in A\ , f(x, a) \in V_a$.

Definition 2: (Indiscernibility relation). Given a decision table $S =< U, C \cup D, V, f >$, and an attribute set $B \subseteq$

$(C \cup D)$, B determines an indiscernibility relation IND(B) on U as follows:

$$IND(B) = \{ (x,y) \epsilon\, U \times U : for\ all\ a\, \epsilon\, B\, , f(x,a) = f(y,a) \}$$

The equivalence relation IND(B) partitions the set U into disjoint subsets, which is denoted by U/IND(B) (or U/B), where an element from IND(B) is called an equivalence class. For every object $x \in U$, let $[x]_B$ denote the equivalence class of relation IND(B) that contains element x, called the equivalence class of x under relation IND(B).

Definition 3:(Lower and Upper approximation). For the given S a subset of attribute $A \subseteq Q$ determines the approximation space. $AS = (U, IND(A))$ in S. For given $A \subseteq Q$ and $X \subseteq U$, the A-lower approximation $\underline{A}X$ of the set X in AS and the A-upper approximation $\overline{A}$X of the set X in AS are defined as follows:

$$\underline{A}X = \{ x \in U : [x]_A \subseteq X \} = \cup \{ Y \in A^{\star} : Y \subseteq X \}$$

$$\overline{A}X = \{ x \in U : [x]_A \cap X \neq \emptyset \} = \cup \{ Y \in A^{\star} : Y \cap X \neq \emptyset$$

Rough set provides the concept to determines for a given information system the most important attributes. The main idea of the reduct is fundamental for rough set theory. An essential part of an information system is a reduct which is related to a subset of attributes. Another important part is a core. The reduct and core is an important concept of rough set theory which is generally used for feature selection and attribute reduction. Rough set theory determines the significance measures, degree of attributes and dependency.

Definition 4: (Positive Region). For the given information system $S = < U, Q, V, f >$ with the condition and decision attribute. $Q = C \cup D$ A ⊂ C can be defined as A positive region $POS_A(D)$ in the relation $IND(D)$ as

$$POS_A(D) = \cup \{ \underline{A}X : X \in IND(D) \}$$

$POS_A(D)$ contains all the objects in U. A positive region for any two subsets of attributes A, $B \in Q$ in the information system S. The subset of attributes $B \in Q$ defines the indiscrenibility relation IND (B) which defines the classification $B * (U/IND(B))$ with respect to subset A. Positive region of B is defined as

$$POS_A(B) = \cup_{X \in B} \underline{A}\ X$$

Definition 5: (Dependency): Positive region of B contains the entire object. The cardinality of positive region B defines a measure $\gamma_A(B)$ of dependency of the set of attributes B on A

$$\gamma_A (B) = \frac{Card\ (POS_A\ (B))}{Card\ (U)}$$

From the information system S a set of all attributes B depends on A in S, which is denoted as A→ B; iff satisfies the equivalence relation $IND(A) \subseteq IND(B)$. Two sets A and B are independent of S if neither $A \rightarrow B$ nor $B \rightarrow A$ hold. The dependency of set B to degree K to the set A in S is denoted as

$$A \overset{k}{\rightarrow} B, \qquad 0 \leq k \leq 1 if\ k = \gamma_A(B)$$

Definition 6: (Significance) Rough set defines a measure of significance or coefficient of significance of the attribute $a\, \epsilon\, A$ from set A with respect to classification $B *$ $(U/IND(B))$ generated by set B.

$$\mu_{A,B}(a) = \frac{card\ (\, POS_A(B)) - \ Card\ (\, POS_{A-\{a\}}(B))}{Card\ U}$$

A significance of attribute a in the set $A \subset Q$ can be computed with respect to original classification Q*.

Quick Reduct Algorithm is the most well-known algorithm for feature selection using Rough sets [12][13]. This is an incremental procedure, where it starts with an empty set and in each step a feature is added to the Reduct, in such way that dependency measure increases. The procedure stops when the dependency measure of the set of features being considered is equal to the dependency measure using all the conditional features. The algorithm attempts to calculate a reduct without exhaustively generating all possible subsets [13]. Its pseudo-code algorithm is given below:

QuickReduct(C,D)
C, the set of all conditional features;
D, the set of decision features.
(1) $R \leftarrow \{\ \}$
(2) do
(3) $T \leftarrow R$
(4) $\forall\, x \in (C - R)$
(5) if $\gamma_{R\, \cup \{x\}}(D) > \gamma_T(D)$
(6) $T \leftarrow R \cup \{x\}$
(7) $R \leftarrow T$
(8) Until $\gamma_R(D) = \gamma_C(D)$
(9) return R

The QUICKREDUCT algorithm attempts to calculate a reduct without fully generating all possible subsets. It starts off with an empty set and adds in turn, one at a time, those attributes that result in the greatest increase in the rough set dependency metric, until this produces its maximum possible value for the dataset.

III. Classifier

The size of information grows rapidly, the problem arises of handling the data. It is infeasible to classify the data manually so automatic methods have been approached to reduce the time and effort for classification. Many document classifications have been built to categories the document according to their content. To improve the accuracy of classifier, researchers have worked on many ranking methods which select the term such as term frequency, chi squared, mutual information, and information gain. Still the problem arises is redundancy in the selected term. Redundant terms are equivalent to noise which causes a reduction in the accuracy of classifier. The classification accuracy changes according to the features being input to the classifier. If the features are of less redundant then the accuracy increases else it decreases. Feature selection algorithm with redundancy reduction for text classification, algorithm helps to decrease in redundant which improves the efficiency of the classifier.

A. Naive Bayes

Most Widely used classifier is the naive bayes. This classifier built the concept of probabilistic Classification where the probability is calculated for each document. it shows the belonging to the categories specified [10][12][13]. Many approaches using naive bayes classifier, multinomial naive bayes is used where the probability $P(C_j \mid d_i)$ of a document d_i belongs to the category C_j . C_j is calculated through the following equation :

$$P(C_j|d_i) = \frac{P(C_j)\prod_{k=1}^{|d|} P(w_{di,k}|C_j)}{\sum_{r=1}^{|c|} P(C_r)\prod_{k=1}^{|dj|} P(w_{d,k}|C_r|)}$$

where $|C|$ is the number of categories. $|d_i|$ be the length of document. $P(C_j)$ is probability of category is calculated according to the equation.

$$P(C_j) = \frac{1 + \sum_{i=1}^{|D|} P(C_j|d_i)}{|C| + |D|}$$

The probability of word gives that the category occurred $P\ (w_i \mid c_j)$ is calculated through the equation.c

$$P(w_i|c_j) = \frac{1 + \sum_{i=1}^{|D|} N(w_i, d_i)\ P\ (\ y_i = c_j|d_i)}{|v| + \sum_{i=1}^{|D|}\ \sum_{j=1}^{|D|} N(\ w_{i,} d_i)\ P(y_i = c_j|d_i)}$$

where $|D|$ is the number of documents in the training set, $|v|$ is the number of words in the training set.

B. K-Nearest Neighbor

The KNN [3][17] algorithm is a well-known instance-based approach that has been widely applied to text categorization due to its simplicity and accuracy. To categorize an unknown document, the KNN classifier ranks the document' s neighbors among the training documents and uses the class labels of the k most similar neighbors. Similarity between two documents may be measured by the Euclidean distance, cosine measure, etc. The similarity score of each nearest neighbor document to the test document is used as the weight of the classes of the neighbor document. If a specific category is shared by more than one of the k-nearest neighbors, then the sum of the similarity scores of those neighbors is obtained from the weight of that particular shared category [2]. When classification is done by means of the KNN, the most important parameter affecting classification is k-nearest neighbor number. Usually, the optimal value of k is empirically determined. k value is determined so that it would give the least classification error.

IV. Experimental Evaluation

1) Performance Analysis

To evaluate the accuracy of text categorization results f-measure, precision and recall are used. These significance measures are mostly used to evaluate the accuracy of the result of classifiers for text categorization. The f-measure shows the combination of both precision and recall used in information retrieval. Precision is the proportion of correctly proposed document to the proposed document. Recall is the proportion of the correctly proposed documents to the test data that have to be proposed. In this paper F-measure, Precision, and Recall are not separated, they are computed for each class and average values of measures are used. Precision P and Recall R of each class are defined in equation below

$$\boldsymbol{P} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}}$$

$$\boldsymbol{R} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}}$$

$$\boldsymbol{F} = \frac{\boldsymbol{2*P*R}}{\boldsymbol{P+R}}$$

Where TP, FP, and FN are true positive, false positive and false negative.

2) Results

The data used in the experiments are outlined in Table I, where the three datasets are used and downloaded from UCI machine learning databases. In the first stage preprocessing performed in four steps. Firstly stop words are removed, those which are useless for classification and may not be longer used. Stop words are removed according to the stop word list of 571 words. After stop word removal porter stemming algorithm is applied for stemming which reduce a words to its stem or root form. In the third step a wordNet, English thesaurus database is applied to have sense word. Lastly, the document vectors with tfidf weighing scheme is applied and the terms are extracted. All these process runs on Personal Computer with Windows XP and Intel® Core™ i7 CPU 2.66 GHz, 8.00 GB memory. The software used is MATLAB R2010b. The detail description about the preprocessed data is shown in Table II.

TABLE I. Data Set Description

Sr.No	Data Set	No. of Documents	Features	Classes
01	Reuters 21578	212	6539	04
02	Classic 4	54	1625	06
03	Newsgroup 20	52	1454	04

TABLE II. Preprocessed Document

Sr.No	Data Set	No. of Documents	Features Extracted
01	Reuters 21578	212	5677
02	Classic 4	54	1411
03	Newsgroup 20	52	976

The Classification KNN and Naive Bayes are applied on the whole dataset Reuters 21578, Classic04 and Newsgroup 20 and overall performances are examined. The results using KNN and Naive Bayes are summarized in Table III.

TABLE III. PERFORMANCE ANALYSIS ON THREE DATASET USING KNN AND NAIVE BAYES CLASSIFIER

Data set	No. of Features	KNN			Naive Bayes		
		Precision	Recall	F-measure	Precision	Recall	F-measure
Reuters 21578	5677	82.79	93.45	87.79	79.78	83.90	81.78
Classic 04	1411	84.67	89.76	87.14	78.89	83.45	81.10
News group 20	976	81.98	91.89	86.64	80.97	87.56	84.13

The result shows that better accuracy is obtained by using KNN as compared to Naive Bayes Classifier.

In the Second Stage after the preprocessing, Information Gain is applied where the features are ranked and it reduces the dimension of feature space. In this the features are ranked individually and the classification is performed. The Classifier performance is examined with IG method and the results are shown in Table IV.

TABLE IV. PERFORMANCE ANALYSIS WITH IG USING KNN AND NAIVE BAYES CLASSIFIER

Date set	No. of Features	KNN			Naive Bayes		
		Precision	Recall	F-Measure	Precision	Recall	F-Measure
Reuters 21578	497	90.67	94.55	92.56	87.67	88.78	88.22
Classic 04	112	90.86	92.67	91.75	89.56	93.67	91.56
Newsgroup 20	43	88.97	96.78	92.71	81.78	92.76	86.92

In third stage after Information gain dimension reduction using rough set is examined. Using Rough set the attributes are reduct due to which dimension gets reduce and the feature space also decreases. Table V shows the classification performance of the feature ranking and the feature reduction performed by IG-RS method.

TABLE V. ERFORMANCE ANALYSIS WITH IG-RS USING KNN AND NAIVE BAYES CLASSIFIER

Date set	Features	KNN			Naive Bayes		
		Precision	Recall	F-Measure	Precision	Recall	F-Measure
Reuters 21578	3576	91.75	95.58	93.63	88.87	91.46	90.15
Classic 04	867	92.46	96.33	94.38	91.67	95.89	93.73
Newsgroup 20	292	90.64	97.91	94.13	83.08	94.25	88.32

The results show the better accuracy for KNN classifier as compared to Naive Bayes in terms of performance measures of classifier Precision, Recall and F-measure. With respect to the classifiers' performances, KNN Classifier shows higher performance than the Naive Bayes Classifier. Consequently, it is seen that a higher classifier performance is acquired with fewer features through hybrid methods.

V. CONCLUSION

A multistage feature selection model is proposed for document classification using information Gain and Rough set. Firstly, document preprocessing is carried out where the features are obtained through different steps like stopword removal, stemming, Wordnet, term weighing and pruning. On the original preprocessed document the classifier KNN and Naive Bayes are applied without dimension reduction and the classification performance are observed in terms of recall, precision and F-measures. Secondly, feature selection method Information Gain is applied in which features are ranked depending on their importance. Thirdly, rough set feature selection and attribute reduction is performed. Hence in proposed model features of less importance are ignored due to which dimensionality of feature space is reduced. Again computational time and complexity of the method are also reduced. At each stage classifiers performance are evaluated in term of precision, recall and f-measures. To analyze the effectiveness and accuracy of proposed model, experiments are performed using KNN and Naive Bayes classifier on Reuters 21578, Classic 04 and News Group 20.

From the experimental results, hence it is concluded that A Multistage feature selection model for document classification using Information Gain and Rough set is efficient to reduce the dimensionality of feature space.

Future scope of the work for text categorization is to reduce the dimensionality, computational time and complexity by developing different reduction algorithm. Also the classification performance can be improved by developing different hybrid model which will be more useful for document clustering.

REFERENCES

[1] Chun-Ling Chen, Frank S.C. Tseng, Tyne Liang, "An integration of WordNet and fuzzy association rule mining for multi-label document clustering", Data & Knowledge Engineering 69, 1208–1226, 2010.

[2] F. Beil, M. Ester, X. Xu, "Frequent term-based text clustering", Proc. of Int'l Conf. on knowledge Discovery and Data Mining KDD '02 pp. 436–442, 2002

[3] C.K. Lee, G.G. Lee, "Information gain and divergence-based feature selection for machine learning-based text categorization", information Processing Manage 42 155–165, 2006

[4] M. Dash, H. Liu, "Consistency-based search in feature selection", Artificial Intelligence 151 155–176, 2003.

[5] R. Jensen, Q. Shen, "Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches", IEEE transactions on Knowledge and Data Engineering 16 (12) 1457– 471, 2004.

[6] Qinghua Hu, Daren Yu, Jinfu Liu, Congxin Wu, "Neighborhood rough set based heterogeneous feature subset selection" Information Sciences 178, 3577–3594, 2008.

[7] W. Pedrycz, G. Vukovich, "Feature analysis through information granulation and fuzzy sets", Pattern Recognition 35 825–834, 2002

[8] Q. Shen, R. Jensen, "Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring", Pattern Recognition 37, 1351–1363, 2004

[9] R.B. Bhatt, M. Gopal, "On fuzzy-rough sets approach to feature selection", Pattern Recognition Letters 26, 965–975, 2005

[10] R.B. Bhatt, M. Gopal, "On the compact computational domain of fuzzy-rough sets, Pattern Recognition Letters 26 1632–1640, 2005.

[11] D.G. Chen, C.Z. Wang, Q.H. Hu, "A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets", Information Sciences 177, 3500–3518, 2007

[12] Z. Pawlak, "Rough Sets, Theoretical Aspects of Reasoning About Data", Kluwer Academic Publishers, Dordrecht, 1991.

[13] Z. Pawlak, A. Skowron, "Rough Sets: Some Extensions", Information Sciences 177, 28–40, 2007

[14] Yuhua Qian, Jiye Liang, Witold Pedrycz, Chuangyin Dang, Positive approximation: An accelerator for attribute reduction in rough set theory", Artificial Intelligence 174, 597–618, 2010.

[15] Yuhua Qiana, Jiye Lianga, Witold Pedrycz, Chuangyin Dang, "An efficient accelerator for attribute reduction from incomplete data in rough set framework", Recognition 44, 1658–1670. 2011

[16] C.S. Yang, L. Shu, "Attribute reduction algorithm of incomplete decision table based on tolerance relation", Computer Technology and Development 16 (9), 68–69 72, 2006

[17] Y.H. Qian, J.Y. Liang, F. Wang, "A new method for measuring the uncertainty in incomplete information systems", Fuzziness and Knowledge-Based Systems 17 (6), 855–880, 2009.

[18] J.Y. Liang, K.S. Chin, C.Y. Dang, C.M. Yam Richid, "A new method for measuring uncertainty and fuzziness in rough set theory", International Journal of General Systems 31 (4) 331–342, 2002

[19] G.Y. Wang, H. Yu, D.C. Yang, "Decision table reduction based on conditional information entropy", Chinese Journal of Computer 25 (7) 759–66, 2002.

[20] C.L.Chen, F.S.C. Tseng, T. Liang, "An integration of fuzzy association rules and WordNet for document clustering", Proc. of the 3th Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 147–159, 2009.

[21] Porter, M. F. "An algorithm for suffix stripping. Program", 14(3), 130 - 137, 1980.

[22] George A. Miller –"WordNet: A Lexical Database for English".

[23] H. Wang, "Nearest neighbors by neighborhood counting", IEEE Transactions on PAMI 28 942–953, 2006

[24] Harun Uguz, "A Two stage Feature selection method for text categorization by using Information gain, Principal Component analysis and Genetic Algorithm", Knowledge Based System 24,1024-1032, 2011.

Double Competition for Information-Theoretic SOM

Ryotaro Kamimura
IT Education Center and Graduate School of Science and Technology
Tokai University, 1117 Kitakaname, Hiratsuka, Kanagawa, Japan

***Abstract*—In this paper, we propose a new type of information-theoretic method for the self-organizing maps (SOM), taking into account competition between competitive (output) neurons as well as input neurons. The method is called ”double competition”, as it considers competition between outputs as well as input neurons. By increasing information in input neurons, we expect to obtain more detailed information on input patterns through the information-theoretic method. We applied the information-theoretic methods to two well-known data sets from the machine learning database, namely, the glass and dermatology data sets. We found that the information-theoretic method with double competition explicitly separated the different classes. On the other hand, without considering input neurons, class boundaries could not be explicitly identified. In addition, without considering input neurons, quantization and topographic errors were inversely related. This means that when the quantization errors decreased, topographic errors inversely increased. However, with double competition, this inverse relation between quantization and topographic errors was neutralized. Experimental results show that by incorporating information in input neurons, class structure could be clearly identified without degrading the map quality to severely.**

***Keywords*—double competition, self-organizing maps, mutual information, class structure**

I. Introduction

A. Goal of the Present Paper

The present paper aims to show that the concept of competition among components in neural networks should be extended to all components of neural networks. Many methods have been developed to realize competition in neural networks. However, we think that they are only related to one aspect of competition. For example, competitive learning is in particular specialized in the competition between output neurons. In standard competitive learning, output neurons compete with each other to represent input patterns. If a neuron wins the competition, it tries to represent input patterns as efficiently as possible. A number of variants to overcome the problems such as dead neurons, the number of neurons, and initial conditions have been developed [1], [2], [3], [4], [5] ,[6], [7], [8], [9]. However, the focus in competitive learning is on competition between output neurons. We have mentioned that competition can be realized in any component of neural networks. Then, in addition to output neurons, we can consider input neurons in competitive neural networks. We can imagine a case where output as well as input neurons compete with other to represent input patterns. The goal of the present paper is to show that the extension of competition into input neurons can improve the performance of neural networks.

B. Information-Theoretic SOM

We apply the information-theoretic method to SOM (information-theoretic SOM), which is based on competition between neurons. The self-organizing map is one of the most important techniques in neural networks [10], [11] and has been used to visualize complex and highly structured data. In SOM, much attention has been paid in particular to topological preservation, and many methods to measure topological consistency have been proposed [12], [13], [14], [15], [16], [17], [18]. In addition, many visualization methods have also been developed to interpret the SOM knowledge obtained by learning [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]. However, in spite of having a good reputation for visualization, SOM has faced difficulty in visualizing results obtained by learning. In the SOM, competition and cooperation between neurons are simultaneously performed in learning. In particular, cooperation processes need extensive fine tuning to maintain topological preservation. However, as more focus is placed on cooperation processes, it becomes more difficult to visualize class structure or class boundaries, since cooperation processes have roles to diminish discontinuity between neurons related to class boundaries. Though several methods have been developed to measure and extract discontinuity on the output space [30], [31], it is still difficult to extract clear class structure.

To overcome this shortcoming of SOM, we have introduced several information-theoretic methods to realize SOM [32], [33]. Information-theoretic methods are numerous in neural learning [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47]. From the information-theoretic point of view, learning in neural networks lies in the acquisition of information content on input patterns. Though we need expensive computations to measure the information content or mutual information, there are a number of information theoretic methods available to do this. In particular, we have introduced similarity between competitive learning and mutual information maximization. When mutual information is defined between input patterns and output neurons and is maximized, just one neuron fires, while all the others ceases to do. Thus, mutual information maximization corresponds to the competitive processes of competitive learning. One of the main merits of this information-theoretic method is that it is easy to control the process of competition and cooperation. Depending on the information obtained by the information-theoretic method, we can control final connection weights and corresponding outputs. For example, when information obtained in learning is larger, competition between neurons becomes more intense, and more severe competition processes are realized. On the other hand, when obtained information is smaller, competition between neurons becomes weaker and all neurons tend to fire equally. This means that just by adjusting the information to be obtained by learning, we can control the competition processes. In addition, the method is not winner-take-all, and many neurons can participate in competition and cooperation. By controlling the information content in neural networks, we can easily control its performance.

C. Necessity of Double Competition

The information-theoretic SOM has shown good performance in clarifying class structure. However, the method cannot always detect clear class structure, in particular when the problems are complex. To resolve this, we introduce the concept of competition into input neurons, as mentioned above. In our framework, the input neurons must compete with each other to represent input patterns. In addition, if an output neuron fires at the same time as an input neuron, the corresponding connection weights between the input and output neuron should be stronger. We call this competition "double competition", because input and output neurons compete with each other to represent input patterns.

We have so far tried to introduce competition in input neurons, which is called "information enhancement". In information enhancement, we tried to enhance competition between neurons by focusing on specific input neurons [48], [49]. On the other hand, we have combined information maximization in input neurons and output neurons, which are separately defined [50]. Those methods have shown improved performance for several problems. However, they are not always effective for taking into account the combined effect of input and output neurons. In this double competition, we suppose two types of actions, namely, competition and mutual interaction. In the competition, input as well as output neurons compete with each other. In addition, we suppose some interaction between input and output neurons. Concretely, when an input and output pattern fire in the same way, the interaction between them becomes stronger.

D. Outline

In Section 2, we first explain the correspondence between information maximization and competitive learning. We explain the concept of double competition to include input and output neurons. Then, we try to present the information-theoretic learning method to realize double competition by using the free energy. Finally, we explain how to estimate the firing probabilities of input neurons. In Section 3, we present two experimental results from the well-known machine learning database. Using a principal component analysis, we try to show that class structure can be clarified by using the present method. However, we point out that topological preservation may be sacrificed for this better visualization. Thus, it is important to more closely examine the relations between better visualization and topological preservation.

II. Theory and Computational Methods

A. Double Competition

Competitive learning has been considered to be one of the most important learning methods in neural networks [51], [2], [52], [53], [4], [5], [3], [54], [55], [1], [3], [56], [7], [57], [58]. In particular, we have introduced information-theoretic competitive learning [59], [60], [61]. Contrary to the computational methods so far developed, we have supposed that competitive learning is a realization of mutual information maximization between output neurons and input neurons. In competitive learning, attention has been mainly to output neurons. However, we can imagine that any components in a neural network compute with each other and we try to apply the concept of competition to input neurons. In the input neurons, we focus on the importance of input neurons. Because input neurons correspond to input variables, the importance of input variables should be taken into account.

Let us explain how to produce self-organizing maps by using a network architecture in Figure 1. The sth input pattern can be represented by $\mathbf{x}^s = [x_1^s, x_2^s, \cdots, x_L^s]^T,\ s = 1, 2, \cdots, S$. Connection weights into the jth competitive neuron are denoted by $\mathbf{w}_j = [w_{1j}, w_{2j}, \cdots, w_{Lj}]^T,\ j = 1, 2, \ldots, M$. Supposing that the firing rate $p(k \mid s)$ of the kth input neuron for the sth input pattern can be computed, then the distance between input patterns and connection weights can be computed by

$$\| \mathbf{x}^s - \mathbf{w}_j \|^2 = \sum_{k=1}^{L} p(k \mid s)(x_k^s - w_{kj})^2. \tag{1}$$

The firing rate $p(k \mid s)$ is considered to be the importance of the kth input neuron for the sth input pattern. The output from an output neuron is computed by

$$v_j^s = \exp\left(-\frac{\| \mathbf{x}^s - \mathbf{w}_j \|^2}{2\sigma^2}\right), \tag{2}$$

where σ denotes a spread parameter and defined by

$$\sigma = \frac{1}{\beta}, \tag{3}$$

where β is larger than zero. By normalizing the output, we have the firing rate

$$p(j \mid s) = \frac{\exp\left(-\frac{\|\mathbf{x}^s - \mathbf{w}_j\|^2}{2\sigma^2}\right)}{\sum_{m=1}^{M} \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{w}_m\|^2}{2\sigma^2}\right)}. \tag{4}$$

We should also compute the collective output from the neuron. In the self-organizing maps, the collective outputs are determined by distances from the winner. The winner c_1 is determined by

$$c_1 = \mathrm{argmax}_j\ v_j^s. \tag{5}$$

Following the formulation of SOM, we compute the distance between the winner and the other neurons by

$$\phi_{jc_1} = \exp\left(-\frac{\| \mathbf{r}_j - \mathbf{r}_{c_1} \|^2}{2\sigma_{ngh}^2}\right), \tag{6}$$

where $\mathbf{r}_j$ denotes the position of the jth neuron on the output map and σ_{ngh} is the spread parameter. Thus, the expected output is approximated by this function

$$q(j \mid s) = \frac{\phi_{jc_1}}{\sum_{m=1}^{M} \phi_{mc_1}}. \tag{7}$$

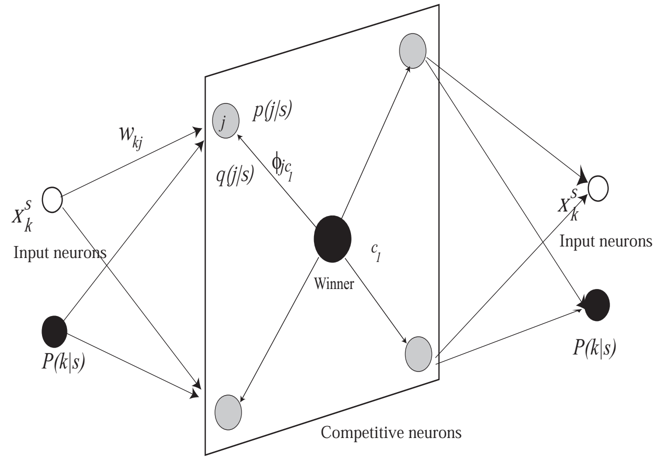


Fig. 1. Network architecture for the information-theoretic self-organizing maps where only connection weights from a winner at the center are plotted.

B. Free Energy Minimization

Learning should be performed to reduce the difference between actual and expected outputs. We represent this difference by using the Kullback-Leibler divergence

$$C = \sum_{s=1}^{S} p(s) \sum_{j=1}^{M} p(j \mid s) \log \frac{p(j \mid s)}{q(j \mid s)}. \tag{8}$$

In addition to this KL divergence, we have other errors which must be minimized, namely quantization errors between connection weights and input patterns

$$Q = \sum_{s=1}^{S} p(s) \sum_{j=1}^{M} p(j \mid s) \|\mathbf{x}^s - \mathbf{w}_j\|^2. \tag{9}$$

Fixing this quantization errors and minimizing the KL-divergence, we have the optimal firing rates

$$p^*(j \mid s) = \frac{q(j \mid s) \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{w}_j\|^2}{2\sigma^2}\right)}{\sum_{m=1}^{M} q(m \mid s) \exp\left(-\frac{\|\mathbf{x}^s - \mathbf{w}_m\|^2}{2\sigma^2}\right)}. \tag{10}$$

We have the following equation called "free energy"

$$F = 2\sigma^2 \sum_{s=1}^{S} \log \sum_{j=1}^{M} q(j \mid s) \exp\left(-\frac{\| \mathbf{x}^s - \mathbf{w}_j \|^2}{2\sigma^2}\right). \tag{11}$$

Finally, by differentiating the free energy, we can obtain the re-estimation formula

$$\mathbf{w}_j = \frac{\sum_{s=1}^{S} p^*(j \mid s) p(k \mid s) \mathbf{x}^s}{\sum_{s=1}^{S} p^*(j \mid s) p(k \mid s)}. \tag{12}$$

As shown in the equation (10), connection weights are modified to make the actual outputs closer to expected outputs.

C. Estimating Firing Rates of Input Neurons

To obtain connection weights w_{kj}, we must estimate the firing rates of input neurons $p(k \mid s)$ in the equation (1). For this purpose, we first compute the outputs from the jth neuron by

$$\hat{v}_j^s = \exp\left(-\frac{\sum_{k=1}^{L}(x_k^s - w_{kj})^2}{2\sigma^2}\right), \tag{13}$$

Normalizing this output, we have the estimated firing rates

$$\hat{p}(j \mid s) = \frac{\hat{v}_j^s}{\sum_{m=1}^{M} \hat{v}_j^s}. \tag{14}$$

By using $\hat{p}(j \mid s)$, we have the output from the kth input neuron

$$\hat{v}_k^s = \exp\left(-\frac{\sum_{j=1}^{M} \hat{p}(j \mid s)(x_k^s - w_{kj})^2}{2\sigma^2}\right), \tag{15}$$

Then, we have the firing rate of the kth input neuron

$$\hat{p}(k \mid s) = \frac{\hat{v}_k^s}{\sum_{l=1}^{L} \hat{v}_l^s}. \tag{16}$$

Putting this firing rate $\hat{p}(k \mid s)$ in the equation (1), (2), we have the distance considering input neurons,

$$\| \mathbf{x}^s - \mathbf{w}_j \|^2 = \sum_{k=1}^{L} \hat{p}(k \mid s)(x_k^s - w_{kj})^2. \tag{17}$$

III. Results and Discussion

A. Experiment Outline

We applied the method two data sets, namely, the glass and dermatology data. Both were taken from the well-known machine learning database [62]. The number of input neurons and patterns for the glass data were 214 and 10, respectively. The number of input neurons and patterns for the dermatology data were 366 and 34, respectively. All the data were normalized to range between zero and one. For quantitative evaluation, we used the well-known quantization and topographic errors. There have been many attempts [12], [13], [14], [15], [16], [17], [18] to measure map quality quantitatively. Among them, both errors are very simple and easy to implement. For visual evaluation, we used the principal component analysis (PCA) to summarize connection weights. As mentioned in the introduction section, there is difficulty in interpreting SOM knowledge, a number of methods have been developed to clarify the knowledge [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]. In this study, we used the PCA for clarification, in particular for simplifying the knowledge. It is easy to demonstrate the performance by using the techniques specific to the SOM, such as the U-matrix. However, we used the PCA so that the present results could be widely interpreted and reproduced.

B. Glass Data

1) Firing Rates of Input Neurons: First, we examine how the firing rates could be changed by increasing the parameter β or by increasing information content in input neurons. When the parameter β was increased from one in Figure 2(a) to three in Figure 2(c), little change in the firing rates could be seen. When the parameter β was increased from four in Figure 2(d) to eight in Figure 2(f), the firing rates became gradually differentiated. When the parameter β was increased from ten in Figure 2(g) to 14 in Figure 2(i), higher and lower firing rates becomes clearer. Finally, when the parameter β was increased from 16 in Figure 2(j) to 20 in Figure 2(l), the clearest firing rates could be seen. Input neurons No.6 and No.8 had the highest firing rates, while input neuron No.5 had the lowest firing rate. The results show that when the parameter β was increased, the firing rates became gradually clearer.

2) Results of PCA for Connection Weights: Figure 3 shows the results of the PCA for connection weights by the SOM and the information-theoretic method with double competition. Figure 3(a) shows the results of PCA by using the conventional SOM. We can see that a condensed group could be seen on the right hand side of the map, and the remaining connection weights were scattered widely on the left hand side of the map. When the parameter β was two and three in Figures 3(b) and (c), connection weights were close to those by the conventional SOM in Figure 3(a). When the parameter β was increased from four in Figure 3(d) to eight in Figure 3(f), a group on the right side began to separate from the others. When the parameter β was increased from 10 in Figure 3(g) to 14 in Figure 3(i), two explicit groups of connection weights on the left and right hand sides began to form. Finally, when the parameter β was increased from 16 in Figure 3(j) to 20 in Figure 3(l), connection weights were separated into two groups on the left hand side and right hand side. In addition, another group could be seen in the middle of the map.

Figure 4 shows the results of PCA by the information-theoretic methods without considering input neurons. When the parameter β was increased from two in Figure 4(a) to 20 in Figure 4(c), a condensed group on the right hand side remained the same, but connection weights on the left hand side became more scattered. These results show that when the parameter β was increased, input patterns were separated into explicit groups by using the information on input neurons. On the other hand, without considering input neurons, explicit groups could not be expected.

3) Quantization and Topographic Errors: We have seen that class structure is clearer by using the information-theoretic method with double competition. The next step is to quantify the map quality obtained using this method. Figure 5(a) shows the quantization errors by the information-theoretic method with double competition in red, without considering input neurons in blue, and SOM in black. The information-theoretic method without considering input neurons showed a sharp decrease in quantization errors, while the quantization errors by SOM and the method with double competition had relatively higher errors. Topographic errors using the information-theoretic method without considering input neurons were quite large. On the other hand, topographic errors did not increase when using the information-theoretic method with double competition. The decrease in quantization errors and increase in the topographic errors by the information-theoretic method without considering input neurons can be inferred from free energy equation (18) (for more detailed discussion, see the discussion section). On the other hand, quantization and topographic errors did not change excessively with the double competition information-theoretic method, and were close to the errors obtained by the conventional SOM. Thus, it can be said that the introduction of information on input neurons attenuated the operation of the free energy.

C. Dermatology Data

1) Firing Rates of Input Neurons: We applied the information-theoretic method to the well-known data set of the dermatology from the machine learning database. Figure 6 shows the firing rates of input neurons when the parameter β was increased from one (a) to 15 (i). Even if we increased the parameter β beyond this point, little change could be seen in the firing rates. When the parameter β was one in Figure 6(a), the firing rates were almost uniform. When the parameter β was two and three in Figure 6(b) and (c), small changes in the firing rates appeared. When the parameter β was increased from five in Figure 6(d) to nine in Figure 6(f), differences between higher and lower rates became larger. When the parameter β was increased from 11 in Figure 6(g) to 15 in Figure 6(i), higher and lower firing rates were at their largest.

2) Results of the PCA for Connection Weights: Figure 7 shows the results of the PCA for connection weights by the SOM (a) and the information-theoretic method with double competition when the parameter β was increased from one (b) to 15 (i). By using the SOM, as shown in Figure 7(a), connection weights seemed to be divided into three groups with weak boundaries. When the parameter β was one and three in Figures 7(b) and (c), the results of the PCA were almost equivalent to that by the SOM in Figure 7(a). When the parameter β was increased from five in Figure 7(d) to nine in Figure 7(f), a distinct group became separated on the right hand side of the map. When the parameter β was further increased from 11 in Figure 7(g) to 15 in Figure 7(i), three groups were clearly separated.

Figure 8 shows the results of the PCA by the information-theoretic method without considering input neurons. When the parameter β was increased from one in Figure 8(a) to nine in Figure 8(b), three groups became more apparent. Then, even when the parameter β was increased from nine in Figure 8(b) to 15 in Figure 8(c), the results of the PCA remained almost the same. The results of the PCA by the information-theoretic method without considering input neurons were inferior to those by the information-theoretic method with double competition in terms of class structure. This shows that the information of input neurons is critical in clarifying class structure.

3) Quantization and Topographic Errors: Figure 9 shows the quantization and topographic errors when the parameter β was increased from one to 15. Figure 9(a) shows quantization errors by the SOM in black, the information-theoretic method with double competition in red, and without considering input neurons in blue. By using the information-theoretic method without considering input neurons, the quantization errors decreased sharply from the beginning onwards. On the other hand, by using the information-theoretic method with double competition, quantization errors increased and became larger than that by the conventional SOM. Figure 9(b) shows topographic errors when the parameter β was increased from one to 15. By using the information-theoretic method without considering input neurons, the topographic error increased sharply and eventually became much larger than the error obtained by the conventional SOM. On the other hand, by using the information-theoretic method with double competition, the

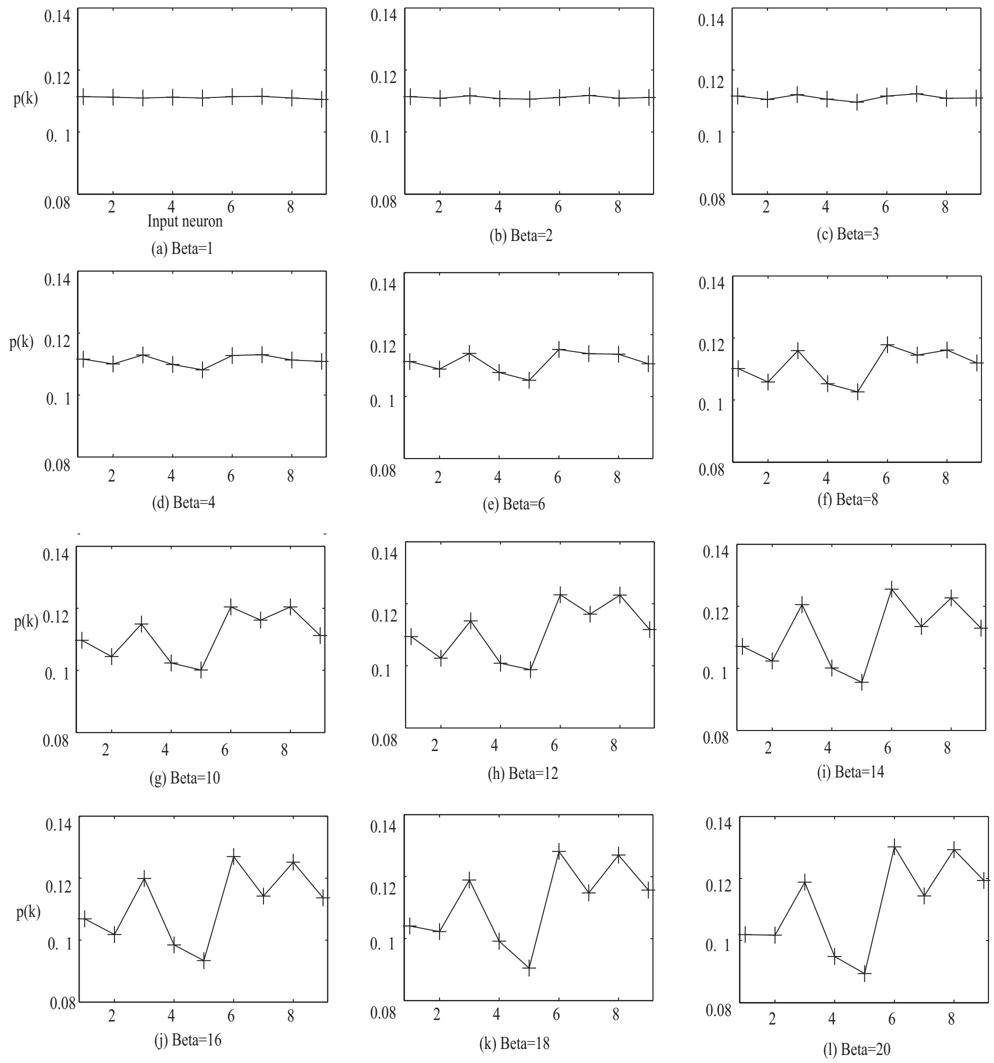


Fig. 2. Firing rates of input neurons by the information-theoretic method with double competition when the parameter β was increased from one (a) to 20 (t) for the glass data.

topographic error increased less than by using the information-theoretic method without double competition. The behavior of the information-theoretic method without considering input neurons can be inferred from the free energy equation (18). By introducing the firing rates of input neurons, this tendency was attenuated. When using the information-theoretic method with double competition, the quantization and topographic errors did not increase or decrease to the extent observed when using the information-theoretic method without considering input neurons.

D. Discussion

1) Validity of Methods and Experimental Results: In this paper, we have proposed a new type of information-theoretic method which takes into account the firing rates of input neurons. We have so far shown that competitive learning as well as self-organizing maps aim to maximize mutual information between input patterns and output neurons [59], [60], [61]. However, little attention has been paid to information content in input neurons. In particular, we have not fully used any information on input neurons in learning processes. Thus, we have introduced the firing rates of input neurons in the learning procedure of the self-organizing maps. We succeeded in determining the re-estimation formula for connection weights. We applied the method to two well-known data sets from the machine learning database, namely, the glass and dermatology data. In both data sets, we succeeded in extracting clearer class structure, particularly by detecting clear class boundaries for the both data sets.

In addition, we could see that quantization and topographic errors were inversely related when we used the method without considering input neurons. This inverse relation can be predicted by examining the free energy equation. The free energy equation in its expanded form appears as the following

$$
\begin{aligned}
F &= \sum_{s=1}^{S} p(s) \sum_{j=1}^{M} p^*(j \mid s) \|\mathbf{x}^s - \mathbf{w}_j\|^2 \\
&\quad + 2\sigma^2 \sum_{s=1}^{S} p(s) \sum_{j=1}^{M} p(j \mid s) \log \frac{p^*(j \mid s)}{q(j \mid s)}, \quad (18)
\end{aligned}
$$

recalling that the spread parameter σ is defined by using the

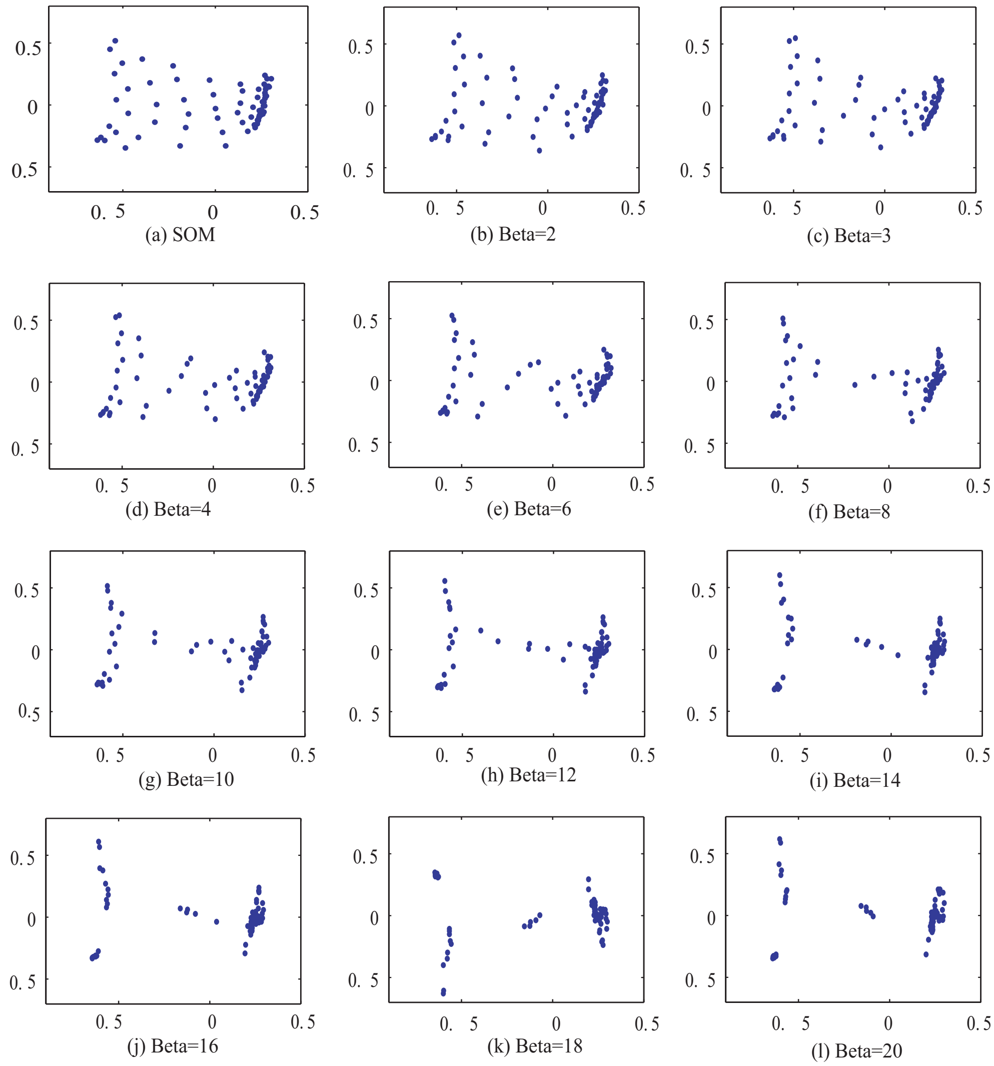


Fig. 3. Results by PCA for connection weights by double competition for the glass data.

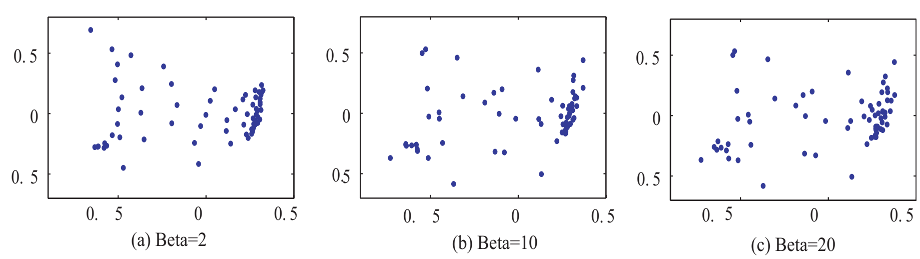


Fig. 4. Results by PCA for connection weights without considering input neurons for the glass data.

other parameter β.

$$\sigma = \frac{1}{\beta}. \tag{19}$$

When the parameter β was increased, and the spread parameter σ was decreased, the first term of the free energy became more effective. This means that quantization errors decreased, as shown in Figures 5 (a) and 9(a). On the other hand, when the parameter β is decreased and the spread parameter σ is increased, the effect of the second term of the free energy becomes dominant. The second term is the KL divergence is used to imitate the collective behavior of output neurons. Thus, when the parameter β is decreased, the topological errors should decreased as well. This is shown in Figure 5(b) and 9(b). The introduction of input neuron firing rates in the learning processes attenuated this tendency.

2) Problems of the Method: There are two problems of this information-theoretic method, namely, the estimation of firing rates of input neurons and degradation in terms of quantization and topographic errors.

First, there is a problem with estimating the firing rates of input neurons, which must be computed in order to realize competitive processes. However, in the computation of competitive neurons, we must insert the firing rates of input neurons

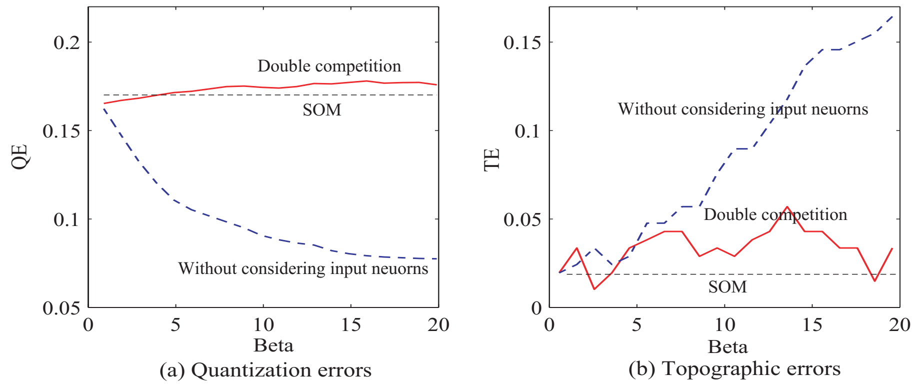


Fig. 5. Quantization and topographic errors by SOM in black, information-theoretic with double competition (in red) and without considering input neurons (in blue) for the glass data.

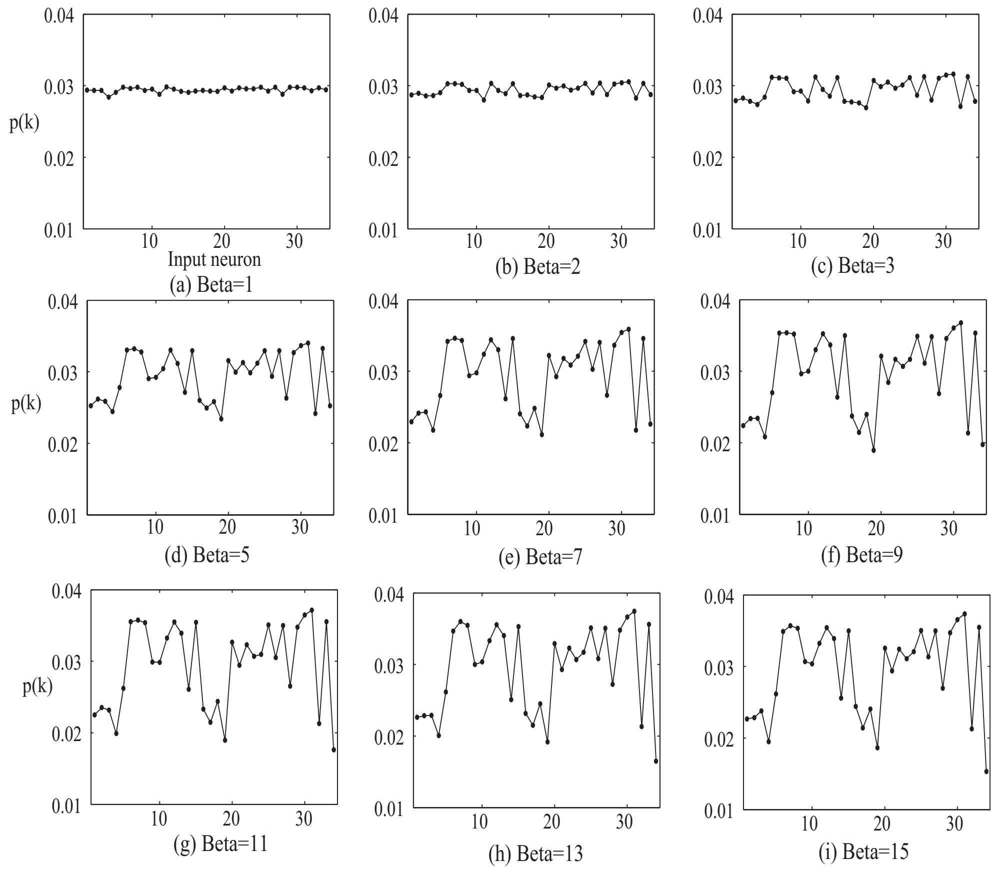


Fig. 6. Firing rates of input neurons when the parameter β was increased from one to 20 for the dermatology data.

into the equation (1). In Section II.C, we briefly presented how to estimate the firing rates of input neurons. However, in the estimation of the firing rates, we must insert the firing rates of competitive neurons into the equation (2). We should thus more carefully examine whether the firing rates of input neurons can be stabilized for the precise computation of the information content, and for producing stable self-organizing maps.

Second, we have a problem of degradation in terms of quantization and topographic errors. In Figures 5 and 9, quantization and topographic errors increased, though they did not reach extreme values as was the case with the method without considering input neurons. We must explain why and how the degradation occurred and try to improve quantization and topographic errors.

3) Possibilities of the Method: The method presented in this paper can be considered as a new input variable selection in SOM, and opens up the possibility of having competition in all components of neural networks. First, this method is an extension of the self-organizing maps which takes into account the importance of input neurons or input variables. The competition between input neurons can be considered as the introduction of the importance of input variables in the self-organizing maps. As is well known, variable se-

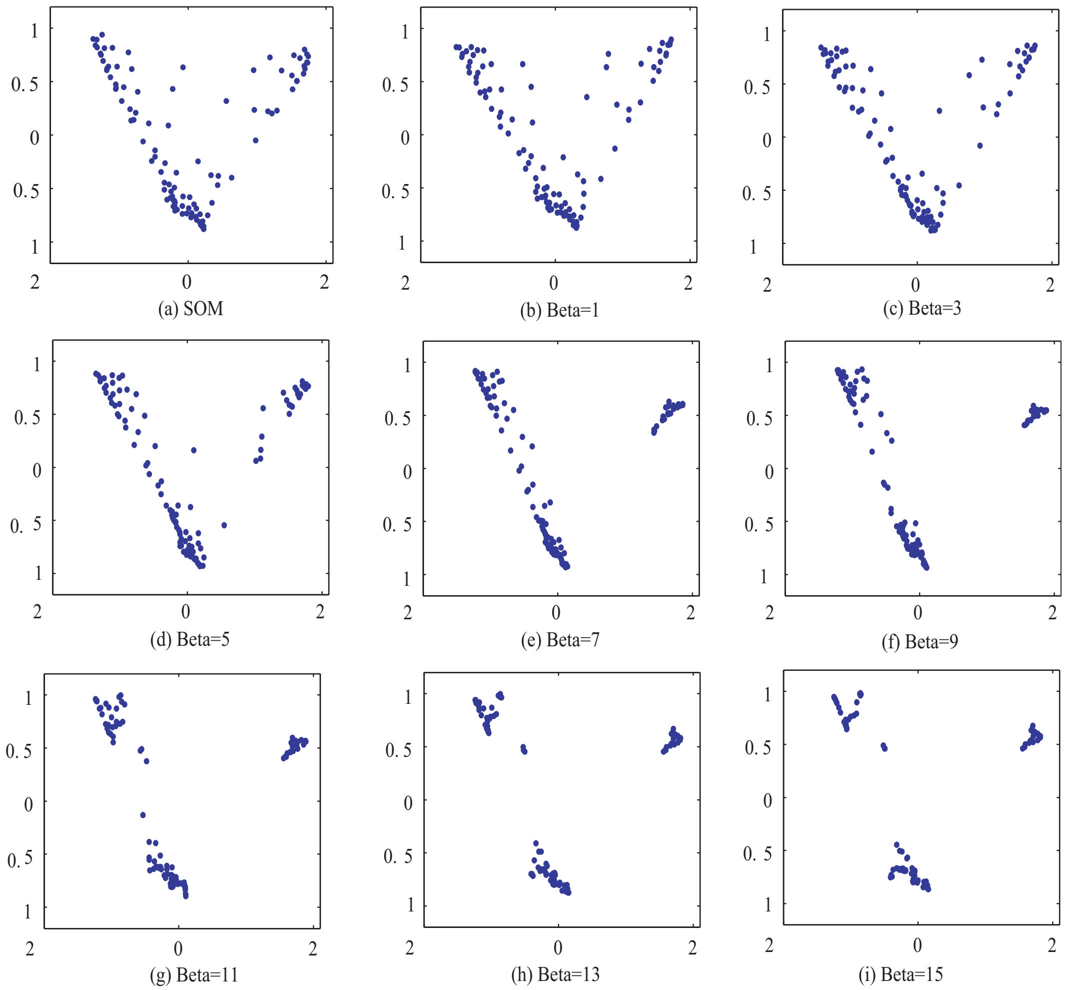


Fig. 7. Results by PCA for connection weights by double competition for the dermatology data.

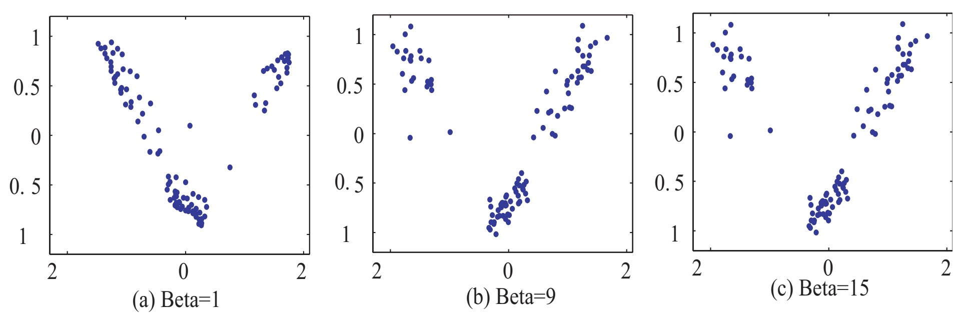
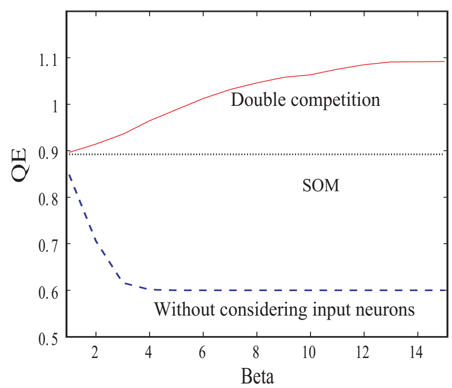


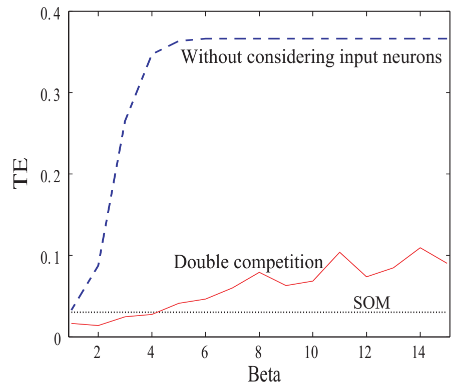
Fig. 8. Results of the PCA for connection weights without considering input neurons for the dermatology data.

lection has played important roles in learning, in particular in supervised learning [63], [64]. In unsupervised learning, such as SOM, the criteria to choose important variables have not been determined. However, in the information-theoretic method, the criteria to measure the importance of neurons is naturally introduced: the importance is measured in terms of information content in neurons. When this information increases, the importance of the neurons becomes larger. We use the importance of input neurons to visualize input patterns by SOM, as it plays an important role in this regard. Thus, it is important to examine relations between the importance of input neurons and the visualization of SOM.

Second, there is the possibility of having competition among all components in neural networks. In the present model of a neural network, in addition to input and output neurons, there are connection weights from the input neurons to output neurons. If it is possible to take into account the competition between all these connection weights, much better performance of a network can be expected. This means that in a neural network, every component competes with each other to most efficiently process outside stimuli.



(a) Quantization errors



(b) Topographic errors

Fig. 9. Quantization and topographic errors by the SOM, and the information-theoretic method with double competition and without considering input neurons for the dermatology data.

IV. Conclusion

In this paper, we have introduced an information-theoretic method considering information in input neurons to realize competitive learning as well as the self-organizing maps. When mutual information is maximized between neurons and input patterns, just one neuron wins the competition. Namely, mutual information maximization corresponds to competitive learning. However, we can imagine that any component in a neural network should contain information on input patterns. Thus, we tried to take into account input neurons in addition to the output or competitive neurons usually used in competitive learning. We applied the information-theoretic method to the self-organizing maps by adding cooperation processes to competitive learning. Then, we applied the information-theoretic methods to two well-known data sets, namely, glass and dermatology data sets from the machine learning database. We found that by increasing information in input neurons, connection weights tended to be divided into clear groups. In addition, the inverse relation between quantization and topographic errors which was observed in the information-theoretic competitive learning without considering input neurons, was neutralized by considering these input neurons. However, quantization and topographic errors tended to degrade map quality when using the information-theoretic method. Thus, we should examine how and why this deterioration occurred in terms of quantization and topographic errors to realize the information-theoretic method with better quantization and visualization performance.

References

[1] A. Luk and S. Lien, "Properties of the generalized lotto-type competitive learning," in *Proceedings of International conference on neural information processing*, (San Mateo: CA), pp. 1180–1185, Morgan Kaufmann Publishers, 2000.

[2] M. Van Hulle, "Topographic map formation by maximizing unconditional entropy: a plausible strategy for 'on-line' unsupervised competitive learning and nonparametric density estimation," *IEEE Transactions on Neural Networks*, vol. 7, no. 5, pp. 1299–1305, 1996.

[3] M. M. V. Hulle, "The formation of topographic maps that maximize the average mutual information of the output responses to noiseless input signals," *Neural Computation*, vol. 9, no. 3, pp. 595–606, 1997.

[4] M. M. Van Hulle, "Faithful representations with topographic maps," *Neural Networks*, vol. 12, no. 6, pp. 803–823, 1999.

[5] M. M. Van Hulle, "Entropy-based kernel modeling for topographic map formation," *IEEE Transactions on Neural Networks*, vol. 15, no. 4, pp. 850–858, 2004.

[6] I. Katsavounidis, C. C. J. Kuo, and Z. Zhang, "A new initialization technique for generalized lloyd iteration," *IEEE Signal Processing Letters*, vol. 1.

[7] H. Xiong, M. N. S. Swamy, and M. O. Ahmad, "Competitive splitting for codebook initialization," *IEEE Signal Processing Letters*, vol. 11, pp. 474–477, 2004.

[8] H. Xiong and M. N. S. Swamy, "Branching competitive learning network: a novel self-creating model," *IEEE Transactions on Neural Networks*, vol. 15, no. 2, pp. 417–429, 2004.

[9] S. Wu, W. Liew, H. Yuan, and M. Yang, "Cluster analysis of gene expression data based on self-splitting and merging competitive learning," *IEEE transactions on information technology in biomedicine*, vol. 8, no. 1, pp. 5–15, 2004.

[10] T. Kohonen, *Self-Organization and Associative Memory*. New York: Springer-Verlag, 1988.

[11] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag, 1995.

[12] K. Kiviluoto, "Topology preservation in self-organizing maps," in *In Proceedings of the IEEE International Conference on Neural Networks*, pp. 294–299, 1996.

[13] S. Kaski, J. Nikkila, M. Oja, J. Venna, P. Toronen, and E. Castren, "Trustworthiness and metrics in visualizing similarity of gene expression," *BMC Bioinformatics*, vol. 4, no. 48, 2003.

[14] J. Venna and S. Kaski, "Neighborhood preservation in nonlinear projection methods: an experimental study," in *Lecture Notes in Computer Science*, vol. 2130, pp. 485–491, 2001.

[15] G. Polzlbauer, "Survey and comparison of quality measures for self-organizing maps," in *Proceedings of the fifth workshop on Data Analysis (WDA04)*, pp. 67–82, 2004.

[16] J. A. Lee and M. Verleysen, "Quality assessment of nonlinear dimensionality reduction based on K-ary neighborhoods," in *JMLR: Workshop and conference proceedings*, vol. 4, pp. 21–35, 2008.

[17] T. Villmann, R. D. M. Herrmann, and T. Martinez, "Topology preservation in self-organizing feature maps: exact definition and measurement," *IEEE Transactions on Neural Networks*, vol. 8, no. 2, pp. 256–266, 1997.

[18] H.-U. Bauer and K. Pawelzik, "Quantifying the neighborhood preservation of self-organizing maps," *IEEE Transactions on Neural Networks*, vol. 3, no. 4, pp. 570–578, 1992.

[19] J. Vesanto, "SOM-based data visualization methods," *Intelligent Data Analysis*, vol. 3, pp. 111–126, 1999.

[20] S. Kaski, J. Nikkila, and T. Kohonen, "Methods for interpreting a self-organized map in data analysis," in *Proceedings of European Symposium on Artificial Neural Networks*, (Bruges, Belgium), 1998.

[21] I. Mao and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection," *IEEE Transactions on Neural Networks*, vol. 6, no. 2, pp. 296–317, 1995.

[22] C. De Runz, E. Desjardin, and M. Herbin, "Unsupervised visual data mining using self-organizing maps and a data-driven color mapping,"

in *Information Visualisation (IV), 2012 16th International Conference on*, pp. 241–245, IEEE, 2012.

[23] S. Shieh and I. Liao, "A new approach for data clustering and visualization using self-organizing maps," *Expert Systems with Applications*, 2012.

[24] H. Yin, "ViSOM-a novel method for multivariate data projection and structure visualization," *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 237–243, 2002.

[25] M.-C. Su and H.-T. Chang, "A new model of self-organizing neural networks and its application in data projection," *IEEE Transactions on Neural Networks*, vol. 123, no. 1, pp. 153–158, 2001.

[26] S. Wu and T. Chow, "Prsom: A new visualization method by hybridizing multidimensional scaling and self-organizing map," *Neural Networks, IEEE Transactions on*, vol. 16, no. 6, pp. 1362–1380, 2005.

[27] L. Xu, Y. Xu, and T. W. Chow, "PolSOM-a new method for multidimentional data visualization," *Pattern Recognition*, vol. 43, pp. 1668–1675, 2010.

[28] Y. Xu, L. Xu, and T. Chow, "Pposom: A new variant of polsom by using probabilistic assignment for multidimensional data visualization," *Neurocomputing*, vol. 74, no. 11, pp. 2018–2027, 2011.

[29] L. Xu and T. Chow, "Multivariate data classification using polsom," in *Prognostics and System Health Management Conference (PHM-Shenzhen), 2011*, pp. 1–4, IEEE, 2011.

[30] E. Merényi, K. Tasdemir, and L. Zhang, "Learning highly structured manifolds: harnessing the power of soms," in *Similarity-Based Clustering*, pp. 138–168, Springer, 2009.

[31] K. Tasdemir and E. Merényi, "Exploiting data topology in visualization and clustering of self-organizing maps," *Neural Networks, IEEE Transactions on*, vol. 20, no. 4, pp. 549–562, 2009.

[32] R. Kamimura, "Self-enhancement learning: target-creating learning and its application to self-organizing maps," *Biological cybernetics*, pp. 1–34, 2011.

[33] R. Kamimura, "Constrained information maximization by free energy minimization," *International Journal of General Systems*, vol. 40, no. 7, pp. 701–725, 2011.

[34] R. Linsker, "Self-organization in a perceptual network," *Computer*, vol. 21, pp. 105–117, 1988.

[35] R. Linsker, "How to generate ordered maps by maximizing the mutual information between input and output," *Neural Computation*, vol. 1, pp. 402–411, 1989.

[36] R. Linsker, "Local synaptic rules suffice to maximize mutual information in a linear network," *Neural Computation*, vol. 4, pp. 691–702, 1992.

[37] R. Linsker, "Improved local learning rule for information maximization and related applications," *Neural Networks*, vol. 18, pp. 261–265, 2005.

[38] H. B. Barlow, "Unsupervised learning," *Neural Computation*, vol. 1, pp. 295–311, 1989.

[39] H. B. Barlow, T. P. Kaushal, and G. J. Mitchison, "Finding minimum entropy code," *Neural Computation*, vol. 1, pp. 412–423, 1989.

[40] G. Deco, W. Finnof, and H. G. Zimmermann, "Unsupervised mutual information criterion for elimination of overtraining in supervised multiplayer networks," *Neural Computation*, vol. 7, pp. 86–107, 1995.

[41] G. Deco and d. Obradovic, "Linear redundancy reduction learning," *Neural Networks*, vol. 8, no. 5, pp. 751–755, 1995.

[42] G. Deco and L. Parra, "Non-feature extraction by redundancy reduction in an unsupervised stochastic neural networks," *Neural Networks*, vol. 10, no. 4, pp. 683–691, 1997.

[43] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on Neural Networks*, vol. 5, no. 4, pp. 537–550, 1994.

[44] J. W. Fisher and J. C. Principe, "A methodology for information theoretic feature extraction," in *IEEE World congress on computational intelligence*, pp. 1712–1716, 1998.

[45] K. Torkkola, "Nonlinear feature transform using maximum mutual information," in *Proceedings of International Joint Conference on Neural Networks*, pp. 2756–2761, 2001.

[46] K. Torkkola, "Feature extraction by non-parametric mutual information maximization," *Journal of Machine Learning Research*, vol. 3, pp. 1415–1438, 2003.

[47] V. Sindhwani, S. Raskshit, D. Deodhare, D. Erdogmus, and J. C. Principe, "Feature selection in mlps and svms based on maximum output information," *IEEE Transactions on Neural Networks*, vol. 15, no. 4, pp. 937–948, 2004.

[48] R. Kamimura, "Information-theoretic enhancement learning and its application to visualization of self-organizing maps," *Neurocomputing*, vol. 73, no. 13-15, pp. 2642–2664, 2010.

[49] R. Kamimura, "Double enhancement learning for explicit internal representations: unifying self-enhancement and information enhancement to incorporate information on input variables," *Applied Intelligence*, pp. 1–23, 2011.

[50] R. Kamimura, "Dependent input neuron selection in contradiction resolution," in *Computational Intelligence & Applications (IWCIA), 2013 IEEE Sixth International Workshop on*, pp. 45–50, IEEE, 2013.

[51] D. E. Rumelhart and D. Zipser, "Feature discovery by competitive learning," *Cognitive Science*, vol. 9, pp. 75–112, 1985.

[52] M. M. Van Hulle, "The formation of topographic maps that maximize the average mutual information of the output responses to noiseless input signals," *Neural Computation*, vol. 9, no. 3, pp. 595–606, 1997.

[53] M. M. Van Hulle, "Topology-preserving map formation achieved with a purely local unsupervised competitive learning rule," *Neural Networks*, vol. 10, no. 3, pp. 431–446, 1997.

[54] S. C. Ahalt, A. K. Krishnamurthy, P. Chen, and D. E. Melton, "Competitive learning algorithms for vector quantization," *Neural Networks*, vol. 3, pp. 277–290, 1990.

[55] L. Xu, "Rival penalized competitive learning for clustering analysis, RBF net, and curve detection," *IEEE Transaction on Neural Networks*, vol. 4, no. 4, pp. 636–649, 1993.

[56] Y. J. Zhang and Z. Q. Liu, "Self-splitting competitive learning: a new on-line clustering paradigm," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 369–380, 2002.

[57] J. C. Yen, J. I. Guo, and H. C. Chen, "A new k-winners-take-all neural networks and its array architecture," *IEEE Transactions on Neural Networks*, vol. 9, no. 5, pp. 901–912, 1998.

[58] S. Ridella, S. Rovetta, and R. Zunino, "K-winner machines for pattern classification," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 371–385, 2001.

[59] R. Kamimura, T. Kamimura, and T. R. Shultz, "Information theoretic competitive learning and linguistic rule acquisition," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 16, no. 2, pp. 287–298, 2001.

[60] R. Kamimura, T. Kamimura, and O. Uchida, "Flexible feature discovery and structural information control," *Connection Science*, vol. 13, no. 4, pp. 323–347, 2001.

[61] R. Kamimura, "Information-theoretic competitive learning with inverse Euclidean distance output units," *Neural Processing Letters*, vol. 18, pp. 163–184, 2003.

[62] K. Bache and M. Lichman, "UCI machine learning repository," 2013.

[63] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[64] E. P. Xing and R. M. Karp, "Cliff: clustering of high-dimensional microarray data via iterative feature filtering using normalized cuts," *Bioinformatics*, vol. 17, no. suppl 1, pp. S306–S315, 2001.

AUTHOR PROFILE

Ryotaro Kamimura is currently a professor of IT Education Center and Graduate School of Science and Technology of Tokai University in Japan. His main research interest is information-theoretic approaches to neural computing.

Mobile Learning-system usage: Scale development and empirical tests

An integrated framework to measure students' behavioural intention

Saleh Alharbi
School of ICT, Griffith University
Gold Coast, Australia
Shaqra University, Saudi Arabia

Steve Drew
School of ICT, Griffith University
Gold Coast, Australia

***Abstract*—Mobile technologies have changed the shape of learning for learners, society, and education providers. Consequently, mobile learning has become a core component in modern education. Nevertheless, introducing mobile learning systems does not automatically guarantee that learners will develop a positive behavioural intention to use it and therefore use it. Thus, acceptance-of-technology and system-success studies have increased. As yet, however, much of the research regarding understanding students' behavioural intention to use mobile learning systems seems to suffer from several shortcomings. On top of that, there is no common cognitive theoretical foundation. This study introduces a theoretical framework that combines the Unified Theory of Acceptance and Use of Technology (UTAUT) and Information System (IS) Success Model. This integration resulted in three success measures and two acceptance constructs. The success measures included the following: a) information quality, b) system quality, and c) user satisfaction; whilst the following were the acceptance measures: a) effort expectancy, b) performance expectancy, and c) social influence. Further, this study introduces lecture attitude as a new construct that is believed to moderate students' behavioural intention. The relationships between the different factors form the research hypotheses.**

***Keywords*—Mobile learning; Mobile learning; Higher education; UTAUT; IS Success**

I. Introduction

Knowledge acquisition is no longer restricted to a certain place and time. In fact, there is a rapid change taking place to traditional learning methods[1]. Learning in the 21st century, or the digital age, is affected by the rapid development of information and communication technologies and the availability of low-cost mobile devices[2] (mobile laptops, tablets, smart phones, PDAs, etc.), and this has resulted in mobile devices becoming more pervasive. Mobile learning is not yet well defined in the literature due to the argument regarding whether to focus on the mobility of learners or devices. Further, it is argued that mobile learning is defined from a technical perspective instead of through the consideration of pedagogical elements. Generally, mobile learning is defined as the conducting of educational activities using a mobile device and wireless service in which both learner and device are mobile[3].

For learners, a mobile-learning environment assists in accessing content quicker, allowing collaborative learning, improving communication between learners, and allowing learners to conduct study-related activities from different locations[4]. For education providers, there have been various initiatives investigating the proliferation and role of the mobility of devices and learners. Therefore, the acceptance and success of mobile Learning-systems, as they are Information Systems in nature, have drawn researchers' attention.

The main purpose of this paper is to develop a framework that assists in understanding students' behavioural intention to use mobile Learning-systems in a higher-education setting. The rest of this paper is structured as follows: First, literature reviews about previous models and theories that have been used to understand the intention and acceptance of an IS are discussed. Second, the two models used in this paper are presented, namely the Unified Theory of Acceptance and Use of Technology[5] and the DeLone and McLean model(D&M henceforth)[6, 7].

Third, the research model and hypotheses development are described. The methodology section provides comprehensive details about the research instruments, constructs validation, sampling and the outline for the research method, data collection, and analysis tools are elaborated. The Data analysis and the discussion follow the methodology section where the research hypotheses were examined, and the results were discussed. This paper hopes to contribute to the work in developing a framework that can be used with students' intention to use mobile Learning-systems.

II. Acceptance, theories and model

Reviewing the relevant literature reveals that investigating Information-System (IS) acceptance has received great attention during the last three decades. Among these models, research such as [8] cited eight models that explain human behaviour and predict IS acceptance: the Theory of Reasoned Action (TRA) [9]; then, based on TRA, Davis [10] introduced the technology acceptance model (TAM); the theory of planned behaviour (TPB) [11]; the motivational model (MM) [12]; the social cognitive theory (SCT) [13, 14]; a combination of TAM and TPB (C-TAM-TPB) [15]; the model of PC utilisation (MPCU)[16, 17]; and the innovation diffusion theory (IDT) [18, 19].

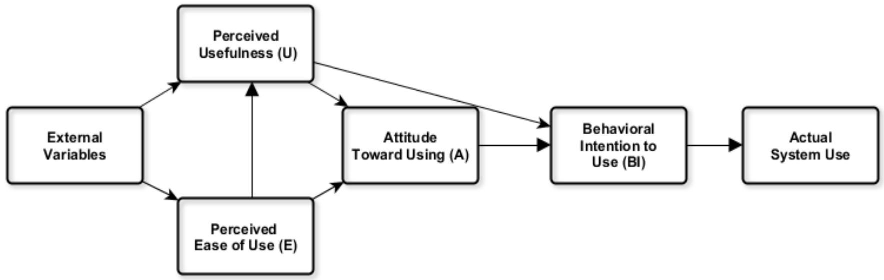


Fig.1. Technology Acceptance Model (TAM) [10]

TRA is suggested to be a fundamental theory in understanding human behaviour. In TRA, behaviour and intention are influenced by two main constructs: attitude about behaviour and subjective norms[9]. Following TRA, TAM was introduced to help understand users' acceptance and usage of a given IS[10]. In TAM, perceived ease of use and perceived usefulness are the core constructs that affect users' attitude and intention, and therefore their use of IS.

Based on a research conducted by Davis [10] the extended TAM, known as the unified theory of acceptance and use of technology (UTAUT), was introduced. UTAUT constructs are derived from the eight models mentioned above Wang, et al. [8].

In terms of measuring IS success, In their research, Wang and Shee [20] cited that the D&M model on IS success [6, 7] appears frequently in system-success studies[21-23].

In this paper, the IS-success model and UTAUT are combined to provide the research-model construction and hypothesis formulation. Our research has two objectives. First, we suggest a framework that can be used to measure students behavioural intention to use mobile-learning systems. The second objective is to examine the relationship between the various variables and students' behavioural intention to use such systems.

In the following section, both the UTAUT and IS-success models are introduced in more detail.

A. *Unified Theory of Acceptance and Use of Technology*

The UTAUT [5] attempted to unify previous theories, as there was an argument about similarities in variables that predicted IS acceptance introduced within these models in different terminologies[24]. UTAUT, as shown in Fig.2, suggests that four core constructs, namely performance expectancy, effort expectancy, social influence, and facilitating conditions affect users' behavioural intention and use behaviour. It also incorporate four other variables: gender, age, experience, and voluntariness of use that [5] highlight to moderate users' adoption of an IS.

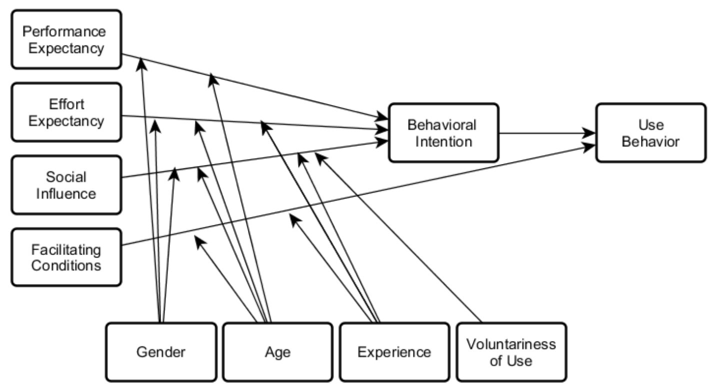


Fig.2. Unified Theory of Acceptance and Use of Technology (UTAUT)[5]

Using these eight determinants in UTAUT, it is evident from the literature that UTAUT is able to explain approximately 70% of technology acceptance behaviour [5, 25, 26]. Further, UTAUT has received researchers' attention to empirically validate the model, and it has been successfully tested in the realm of mobile-technology adoption, which is similar to the scope of this study[27] [28] [24] [8, 26]. As shown in Fig.1. it is clear that TAM[10] provides the basis for UTAUT. The original TAM suggests that the acceptance or rejection of an IS can be measured based upon two beliefs: perceived usefulness and perceived ease of use. Perceived usefulness (PU) is defined as "the degree to which a person believes using a particular system would enhance his or her job performance" [10], and the other belief is "perceived ease of use" (PEOU), which is defined as "the degree to which a person believes that using a particular system would be free of effort" [10].

Within UTAUT, the two prominent beliefs in TAM are similar to performance expectancy and effort expectancy, respectively. The other constructs are 1) social influence, which directly affects behavioural intention to use the IS and 2) facilitating conditions, which directly impacts use behaviour. Within the current research interest and focus, the direct determinants of behavioural intention are used to avoid incorrect inference.

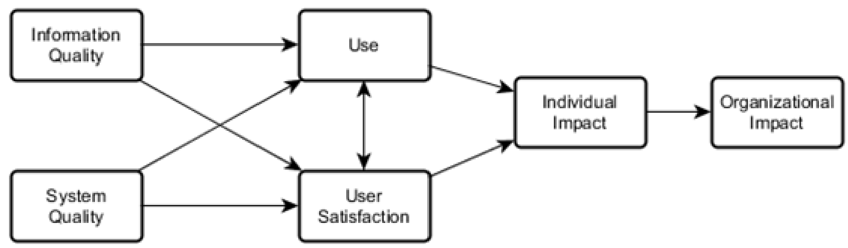


Fig.3. D&M IS Success Model[6]

Thus, facilitating condition was eliminated from the proposed model, as it is not a direct determinate on behavioural intention to use[5]. Further, age and gender are also removed for simplicity, and the other two variables, experience and voluntariness of use, suggested by UTAUT are omitted because experience moderates user behaviour, and the current study investigates mobile learning in a voluntary-usage environment. Moreover, because the research goal is to measure students' behavioural intention to use mobile Learning-systems, the use behaviour in UTAUT[5] and use in the D&M[6, 7] model are also eliminated.

B. IS Success model

D&M [6] proposed a model for measuring IS success. After a comprehensive review of relevant literature regarding IS success measures, D&M concluded that IS success can be measured using a multidimensional model that adopts six different success categories: system quality, information quality, use, user satisfaction, individual impact, and organizational impact (see Fig.3).

System quality and information quality affect use and user satisfaction. Further, user satisfaction can be affected by the amount of use and vice versa. Use and user satisfaction jointly and separately have a direct association with individual impact.

Finally, individual impact is a direct antecedent of organisational impact. Hence, the D&M model essentially provides a multitude of IS-success measures and proposes temporal and causal interdependencies between quality characteristics (system quality); IS-output quality (information quality); output consumption (use); users' response (user satisfaction); behavioural effects of the IS on users (individual impact); and, lastly, IS effects on organisational performance (organisational impact)[29, 30]. The relationship between the six categories has been empirically investigated by many researchers (e.g., [29-32]).

In response to suggestions from the literature and evidence from empirical studies, an updated IS-success model was proposed [7]. In the updated IS-success model, DeLone and Mclean [7] introduced "service quality" as a new measurement, and both individual and organisational impacts were grouped into a new category called "net benefits" (see Fig.4).

In this research, the categories adopted from the updated IS Success model [7] are explained in the research-model section.

III. Research Model

Various types of models have been applied to the context of mobile learning in order to understand and explain students' use of mobile learning and their satisfaction about mobile Learning-systems. In a mobile-learning context, however, there is a gap in the literature with regard to providing a theoretical framework in which empirical research can be grounded[33, 34]. In addition, Sun and Zhang [35] highlight that previous theories can be further improved. Most importantly, in their research to validate D&M model (Rai, Lang, & Welker, [36] recommended integrating theories and developing a multi-constructs model that considers beliefs, attitude, and behaviour in addition to IS-success measures.

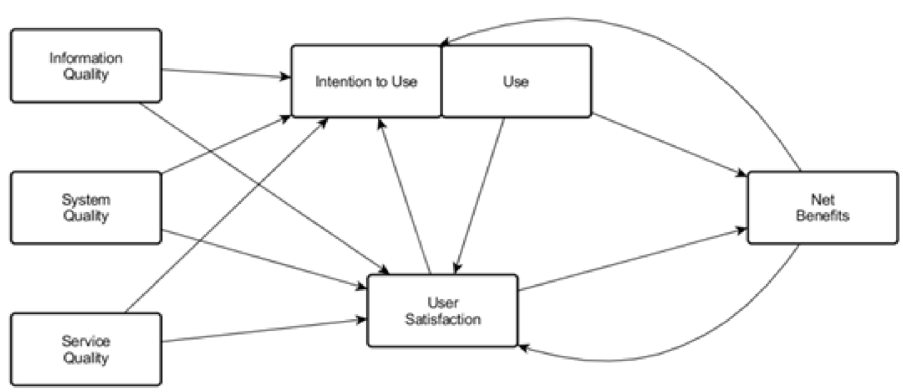


Fig.4. Updated D&M IS Success Model[7]

Therefore, the research model, as shown in Fig.5, in this research combines constructs from UTAUT [5] and success categories from the D&M model[7]. And it also introduces a new moderator found in the literature of mobile and eLearning: lecturer attitude. The following subsections provide a comprehensive look at the theoretical groundwork provided by prior studies in order to formulate relevant hypotheses for this research.

A. The relationship between UTAUT constructs and behavioural intention

As discussed earlier, and in accordance with the current study objectives, the three core constructs in UTAUT have been adopted in this study. These constructs include performance expectancy, effort expectancy, and social influence. This is because they directly impact behavioural intention. However, the fourth construct, which is facilitating conditions, is eliminated from the current study due to the absence of its effect on behavioural intention[5]. Therefore, in relation to UTAUT variables, three hypotheses were introduced in this study.

1) Performance expectancy

First, performance expectancy replaced determinants found in other models (Table I). In this study, performance expectancy is defined as the "degree to which a student believes that using mobile learning systems is helpful, useful and helps him/her to do tasks quickly, and attain gain in learning outcomes". In addition, performance acceptance is a direct determinant of a user's behavioural intention to use an IS, thus it can be validated[5]. Therefore, the following is hypothesised:

a) H1: Performance expectancy would positively affect students' behaviour intention to use mobile Learning-systems.

2) Effort expectancy

Second, effort expectancy, which is also proposed in UTAUT, combines other variables (Table I). Within this study, effort expectancy is referred to as "the degree of ease associated with the use of mobile Learning-systems: the ease of using the systems, the flexibility of interaction, and interaction with mobile Learning-systems is clear and understandable". Effort expectancy is already validated to have a direct impact on a user's behavioural intention to use IS[5]. Therefore, hypotheses on the relationship between effort expectancy and behavioural intention are as follows:

a) H2: Effort expectancy would positively affect students' behaviour intention to use mobile Learning-systems.

3) Social influence

Further, the linkage between the third construct, social influence, and behavioural intention is examined. Considering the current study context, social influence is defined as the "degree to which a student perceives the importance of others believe he or she should use mobile Learning-system". Similar to the previous constructs, social influence is empirically tested to be used as a direct determinate of a user's intention to use an IS[5]. Therefore, the following is the hypotheses on the relationship between social influences and behavioural intention:

a) H3: Social influence would positively affect students' behaviour intention to use mobile Learning-systems.

TABLE.I. ADAPTED FROM[5], CITED IN[25]

UTAUT Constructs	**The Sub-Constructs**	**The source theory/ies**
Performance Expectancy	Perceived Usefulness	TAM/TAM2/C-TAM-TPB
	Extrinsic Motivation	MM
	Job-Fit	MPCU
	Relative Advantage	IDT
	Outcome Expectations	SCT
Effort Expectancy	Perceived Ease of Use	TAM/TAM2
	Complexity	MPCU
	Ease of Use	IDT
Social Influence	Subjective Norm	TRA, TAM2, TPB/DPTB, C-TAM/TPB
	Social Factors	MPCU
	Image	IDT

B. Success measures

Success measures vary from one IS to another. Stockdale and Borovicka [37] states that success measures are influenced by the type of system being evaluated. Thus, it is important to relate the context of the IS to the appropriate success measure[38].In this study, information and system quality are adapted from DeLone and McLean [7]. In addition, findings from Wixom and Todd [39] is discussed.

According to DeLone and McLean [6, 7], information quality is the quality of the output of the IS. It considers the completeness and whether the IS provides all relevant information. Further, information quality is measured by the format and information presentation. Accuracy and correctness of information are also included in information quality measure. Accuracy concerns data correctness; currency assess whether the information is up to date.

The other success measure in the D&M model, system quality, measures the functionality and performance of the IS [7]. System quality considers various dimensions of the IS, such as reliability, flexibility, accessibility, and usefulness.

It has been found in the literature that validates the D&M model[7] that information quality and system quality jointly or separately affect user satisfaction—the user's response to the IS[40-42]. Consequently, user satisfaction also affect the user's intention to use the IS[6, 7].

Therefore, based on the discussion above, the following is hypothesised:

a) H4: Information quality would positively affect students' satisfaction about mobile Learning-systems.

b) H5: System quality would positively affect students' satisfaction about mobile Learning-systems.

c) H6: Students' satisfaction would positively affect students' intention to use mobile Learning-systems.

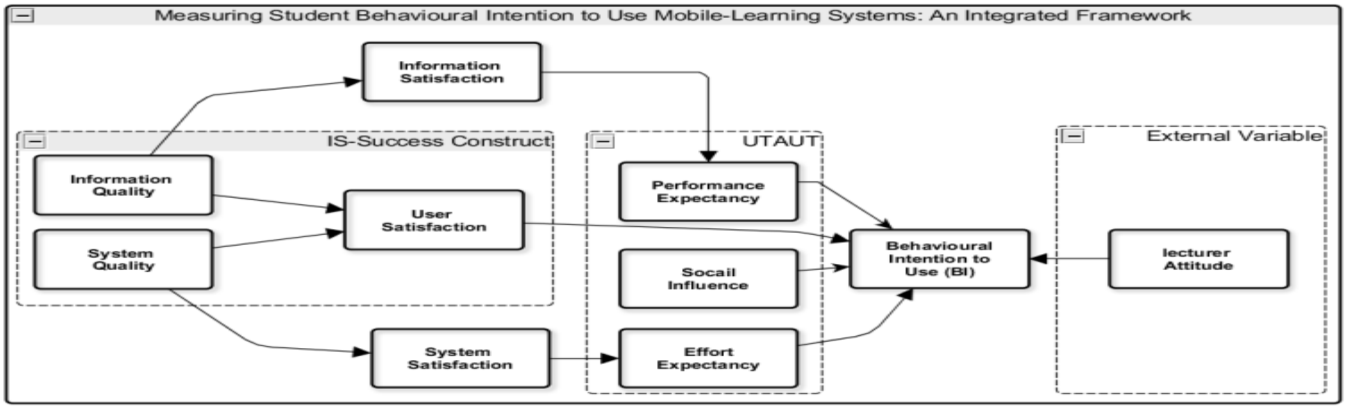


Fig.5. The research model

Further, Seddon and Kiew[40] revised the D&M model and replaced use with usefulness. The authors concluded that system usefulness positively impacts the actual use. However, not using the system does not automatically mean it is not useful. In addition, in a research on theoretical integration of user satisfaction and technology acceptance, Wixom and Todd [39] introduced two measures: information satisfaction and system satisfaction. The former measures the satisfaction with information produced by the system. The latter addresses the degree of favourableness with regard to the system and interaction mechanism. In their conclusion, the authors highlight that information and system satisfaction are directly affected by information and system quality, respectively. In addition, the more information satisfaction, the more likely one will find the IS useful. In the same vein, the more system satisfaction, the more likely one will find an IS easy to use. It is noteworthy that usefulness and ease of use are the main constructs in TAM. However, as UTAUT is employed in this study instead of TAM, the performance expectancy and effort expectancy are used. They capture usefulness and ease of use, respectively[5].

Therefore, the discussion above led to the following hypothesis:

d) H7: Information quality would positively affect information satisfaction of mobile Learning-systems.

e) H8: System quality would positively affect system satisfaction of mobile Learning-systems.

f)H9: Information satisfaction would positively affect performance expectancy.

g) H10: System satisfaction would positively affect effort expectancy.

C. The relationship between the introduced construct and behavioural intention

In a study of acceptance of mobile learning, Wang, et al. [8] highlights that the mobile-learning context is not necessarily similar to other IS, and therefore UTAUT core constructs may not be sufficient in determining a user's behavioural intention. Further, Pedersen and Ling [43] ,as cited in Wang, et al.,[8], suggest to modify existing models in order to apply them to mobile Internet services, including mobile learning. Therefore, an additional construct was incorporated in this study: lecturer attitude.

1) Lecturer's attitude

Very little research focuses on addressing the impact of instructors' opinions on students' behavioural intention to use mobile devices in learning. Researchers such as Brubaker [44] investigated instructors' attitudes towards using laptop devices during lectures; the result reveals that a majority of respondents emphasise that laptops distract students. A recent study on students' perceptions confirms the finding. The recent qualitative study by Gikas and Grant [45] reflects that students are frustrated because of anti-technology instructors who are unwilling to incorporate technology into their courses. By contrast, Alsaggaf, et al. [46] studied faculty perception in using mobile devices in their classes, and the result showed that lecturers may have a positive believe on students using mobile devices. Therefore, from the discussion above, researchers believe that lecturers' attitudes could affect students' behavioural intention to use mobile Learning-systems. Hence, the following is hypothesised:

a) H11: Lecturers' attitude toward using mobile devices would positively affect students' behavioural intention to use mobile Learning-systems.

IV. METHODOLOGY

A quantitative empirical method is used to validate the research model. From a methodological point of view, a survey is used within this research to accomplish the study objectives[47, 48].First, this study is based on well-tested and validated research instruments in previous similar researches. Further, this study objectively investigates the relationships between various constructs, therefore using survey as method for data collection enables testing the research hypotheses. The necessary data for the model validation is collected using an online survey. Online surveys provide researchers with various benefits[49], including saving researchers time and reducing expenses by overcoming geographic distance. Further, online surveys enable recruiting unique subjects.

Further explanation and verification of the model constructs will be undertaken. The development of the scale will be based on previously-validated scales available from relevant literature. Specifically, the questionnaire will be constructed from the original UTAUT model[5] and IS success model[6, 7]. Further, for other measures proposed by authors, experts from the mobile-learning field were contacted to ensure content validity. The participants in this research will be undergraduate and postgraduate students from different faculties and disciplines. Participants will be recruited by emailing the URL to the questionnaire. A probability-sampling technique, particularly random sampling, is utilized in this study to achieve the sample frame. Random sampling is used when each unity in the population has the chance to participate[50]. SPSS software package is used to accomplish proper statistical processing and therefore determine significant relationships between the different variables within the research model.

A. Survey population

Participants in this project were any person enrolled in any undergraduate or postgraduate degree at Griffith University, Australia. The potential participant pool includes students from any level of study and including on-campus and off-campus students. That includes those who are currently doing their English course at Griffith English Language Centre. Participants were recruited by word of mouth, and via email during which official calls for participation were issued.

B. Instrument development

To ensure content validity, the questionnaire used in this study was adapted from the original measurement scales used in UTAUT model[5], IS success model[6, 7], Modified IS Success[40], and on the basis of literature review, the lectures' attitude is added as a new construct. The necessary modifications and wording changes and validation was made to fit the context of mobile learning context. To avoid issues that can occur in wordings, measurement and ambiguities, the questionnaire was pre-tested by two native English speakers. Sekaran and Bougie [51] highlight that such pre-test is essential because wording problems significantly influence accuracy[52].

The research instrument consists of five main sections. The first section incorporates a nominal scale to identify respondents' demographic information. The second section to the fifth section uses 7-point Likert response scale where 7: Strongly agree, 6: Moderately agree, 5: Slightly agree, 4: Neutral, 3: Slightly disagree, 2: Moderately disagree, and 1: Strongly disagree.

The second section concerns UTAUT constructs. IS Success items are presented in the third section. The fourth section consists of the Modified IS Success variables. Finally, the introduced variables, lecturer's attitude is included in the fifth section. The sections from two to five are presented in the Table 2 below with the subsections for each model. The full questioner, including the demographics information is available in Appendix A.

TABLE.II. RESEARCH INSTRUMENTS

Mobile Learning-system usage: An integrated framework to measure students' behavioural intention	
Scales and items	
2. UTAUT(adapted from Venkatesh, Morris, Davis, & Davis (2003)**)**	
Section I	**Performance Expectancy**
PEE1	I feel that mobile learning is useful.
PEE2	Mobile learning improves my study efficiency.
PEE3	Mobile learning improves my study convenience.
PEE4	Mobile learning lets me do study related tasks more quickly.
Section I	**Effort Expectancy**
EFE1	Skilfully using mobile learning is easy for me.
EFE2	I find that using mobile learning is easy.
EFE3	Learning how to use mobile learning is easy for me.
EFE4	My interaction with mobile learning is clear and understandable.
Section III	**Social Influence**
SOI1	Those people that influence my behaviour think that I should use mobile learning
SOI2	Those people that are important to me think that I should use mobile learning
Section IV	**Behavioural Intention to Use**
BI1	I intend to use the mobile learning system in the future
BI2	I predict I would use the mobile learning system in the future
BI3	I plan to use the mobile learning system in the future
3. IS Success (adapted from DeLone & McLean (1992,2003)**)**	
Section I	**Information Quality**
IQ1	The mobile learning system provides information that is exactly what you need (Content Accuracy)
IQ2	The mobile learning system provides information you need at the right time (Availability)
IQ3	The mobile learning system provides information that is relevant to your course (Usability, relevance)
IQ4	The mobile learning system provides sufficient information for your purposes (Quantity of information)
IQ5	The mobile learning system provides information that is easy to understand (Understandability)
IQ6	The mobile learning system provides up-to-date information (Currency)
IQ7	The mobile learning system provides information that appears readable, clear and well formatted (User interface)
IQ8	The mobile learning system provides required information on time. (Timeliness)
IQ9	The mobile learning system provides information that is suitably concise.
Section II	**System Quality**
SQ1	The mobile learning system allows a high level of customization for different courses
SQ2	The mobile learning system provides for personalized information presentation
SQ3	The mobile learning system is easy to use
SQ4	The mobile learning system is user-friendly (Easy to learn)
SQ5	The mobile learning system provides a high of availability (Access)
SQ6	The mobile learning system provides an appropriate level of on-line assistance and explanation (User requirements)
SQ7	The mobile learning system provides interactive features for an effective user experience

SQ8	The mobile learning system provides satisfactory support to users of the system (Help and training)
SQ9	The mobile learning system has features that support the needs of a range of different courses (Flexibility)
SQ10	The mobile learning system has a high level of reliability
SQ11	The mobile learning system provides high-speed information access (Efficiency)
4. Modified IS Success(Adapted from Seddon and Kiew(2007)**)**	
Section I	**User Satisfaction**
US1	Mobile learning systems is effective
US2	Mobile learning systems is efficient
US3	Overall, I am satisfied with mobile learning systems
Section II	**Information Satisfaction**
IS1	Overall, the information I get from mobile learning system is very satisfying
IS2	I am very satisfied with the information I receive from mobile learning system
Section III	**System Satisfaction**
SS1	All things considered, I am very satisfied with mobile learning system
SS2	Overall, my interaction with mobile learning system is very satisfying
5. Lecturers' attitude (New scale**)**	
Section I	**Lecturers' attitude**
LT1	I can use my mobile device in a formal learning environment e.g. searching resources in lectures
LT2	Lectures encourage me to use mobile devices device in a formal learning environment e.g. searching resources in lectures
LT3	Lecturers say that mobile devices sometimes can be very distracting.

C. Data Collection

The questionnaire was made available at the first semester of the academic year 2014. The survey was distributed online by emailing the potential population the URL to the survey. At this time, 204 responses were recorded. Of that, only 124 responses yielded valid responses that were used for analysis.

D. Reliability

Reliability assessment was done using Cornbach Alpha[53]. Reliability concerns internal consistency between multiple measurements of variables, and Cornbach Alpha is commonly used to measure it[54]. As per many studies(i.e.,[55, 56], constructs are considered to have internal consistency reliability when the Cronbach Alpha value exceeds 0.07.

In this study, the reliability assessment was done using Statistical Package for Social Sciences (SPSS) version 22. All measures in this study show a high level of reliability, ranging from 0.924 to 0.981. All scales exceeded 0.70, and therefore the survey is considered reliable. However, the new introduced scale, lecturer attitude shows a low reliability score of .63 which suggest that this construct needs further revision. Further, according to De Vaus [57] reliability score might be attributed to the smaller number of items. The table below (Table III) summaries the reliability analysis for all constructs. The overall reliability for all scales exceeded 0.70, and therefore the survey is considered reliable.

TABLE.III. RELIABILITY ANALYSIS

Scale	Number of Items	Cronbach Alpha
Performance Expectancy (PEE)	4	0.924
Effort expectancy (EFE)	6	0.960
Social Influence (SOI)	2	0.958
Behavioural intention to use (BIU)	3	0.973
Information Quality(IQ)	9	0.958
System Quality(SQ)	11	0.963
User Satisfaction (US)	3	.958
Information Satisfaction (IS)	2	.960
System Satisfaction (SS)	2	.981
Lecturer Attitude (LT)	3	.63
Overall reliability	43	0.98

E. Ethics

This research is being conducted in accordance with the ethics requirements by the relevant research ethics committee. Prior to the commencement of the data collection stage, ethical approval was obtained. Before commencing the survey, a full disclosure of the research title, purpose, expected benefits, and the ethical conducts of the research was provided to all participants. Further, participants were made aware of the voluntary participation in which they do not have to answer every question unless they wish do so, and they may withdraw at any stage of the questionnaire. In addition, data was collected anonymously and no personal information about the subjects were collected. The confidentiality of the data collected was assured to all participants. Finally, participants were provided with the researchers' information and contact details, and the research ethics committee contact details for any inquiry.

V. DATA ANALYSIS

A. Demographics

Most of the participants were female, 77 % females and 32 females. The majority of participants were between 18 and 24 years, with 50.81 % from 18 to 24, 24.19% from 25 to 34, 13.71% from 35 to 44, and 11.29% range from 17 to 18, and above 44. The rest of demographic information regarding the level of education, device types, the various use of mobile devices, and the use of Griffith mobile app are presented in the figure presented in the next page (Fig. 6).

B. Statistical analysis and hypotheses testing

In line with the study objective, correlation analysis was conducted to examine the relationship between the variables used within this study, and therefore to empirically decide whether or not to accept or reject the null hypotheses. The strength of correlation coefficients is determined based on the categorisation proposed by Dancey and Reidy [58] as follows: a)perfect correlation(1), b) Strong (0.7-0.9), c) Moderate(0.4-0.6), d) Weak(0.1-0.3), e) Zero(0).

Hypotheses on the relationship between UTAU constructs and behavioural intention are presented first.

1) The relationship between UTAUT constructs and behavioural intention

a) H1: Performance expectancy would positively affect students' behaviour intention to use mobile Learning-systems.

The correlation analysis result in Table IV below shows that there is a strong positive relationship between PE and BIU and this correlation is significant, r(124) =.828, p < .005. This correlation suggests that when performance expectancy increases, students' behavioural intention to use mobile-learning systems will increase. Hence, H1 is supported.

TABLE.IV. PE AND BIU CORRELATIONS

Correlations		
Factors		BIU
PE	r-value	.828**
	p-value	.000
	N	124

PE: Performance expectancy; BIU: Behavioural intention to use

b) H2: Effort expectancy would positively affect students' behaviour intention to use mobile Learning-systems.

The correlation analysis result in Table V below shows that there is a fairly strong positive and significant relationship between EF and BIU, r(124) =.664, p < .005. This correlation suggests that when effort expectancy increases, students' behavioural intention to use mobile-learning systems will increase. Hence, H2 is supported.

TABLE.V. EF AND BIU CORRELATIONS

Correlations		
Factors		BIU
EF	r-value	.664**
	p-value	.000
	N	124

EF: Effort expectancy; BIU: Behavioural intention to use

c) Social influence would positively affect students' behaviour intention to use mobile Learning-systems.

The correlation analysis result in Table VI below shows that there is a fairly a weak positive relationship between SOI and BIU, r(124) =.323, p < .005. Since the correlation is significant, H3 is statistically supported.

TABLE.VI. SOI AND BIU CORRELATIONS

Correlations		
Factors		BIU
SOI	r-value	.323**
	p-value	.000
	N	124

SOI: Social influence; BIU: Behavioural intention to use

2) The relationship between Success measures constructs and behavioural intention

a) H4: Information quality would positively affect students' satisfaction about mobile Learning-systems.

The correlation analysis result in Table VII below shows that there is a fairly strong positive and significant relationship between IQ and SS, r(124) =.870, p < .005. This correlation suggests that when information quality increases, students' satisfaction about mobile-learning systems will increase. Hence, H4 is supported.

TABLE.VII. IQ AND SS CORRELATIONS

Correlations		
Factors		SS
IQ	r-value	.870**
	p-value	.000
	N	124

IQ: Information Quality; SS: System Satisfaction

3) The relationship between Success measures constructs and behavioural intention

a) H4: Information quality would positively affect students' satisfaction about mobile Learning-systems.

The correlation analysis result in Table VII below shows that there is a fairly strong positive and significant relationship between IQ and SS, r(124) =.870, p < .005. This correlation suggests that when information quality increases, students' satisfaction about mobile-learning systems will increase. Hence, H4 is supported.

TABLE.VIII. IQ AND SS CORRELATIONS

Correlations		
Factors		SS
IQ	r-value	.870**
	p-value	.000
	N	124

IQ: Information Quality; SS: System Satisfaction

b) H5: System quality would positively affect students' satisfaction about mobile Learning-systems.

The correlation analysis result in Table VIII below shows that there is a strong positive and significant relationship between IQ and US, r(124) =.825, p < .005. This correlation suggests that when system quality increases, students' satisfaction about mobile-learning systems will increase. Hence, H5 is supported.

TABLE.IX. SQ AND BIU CORRELATIONS

Correlations		
Factors		US
SQ	r-value	.825**
	p-value	.000
	N	124

SQ: System Quality; US: User Satisfaction

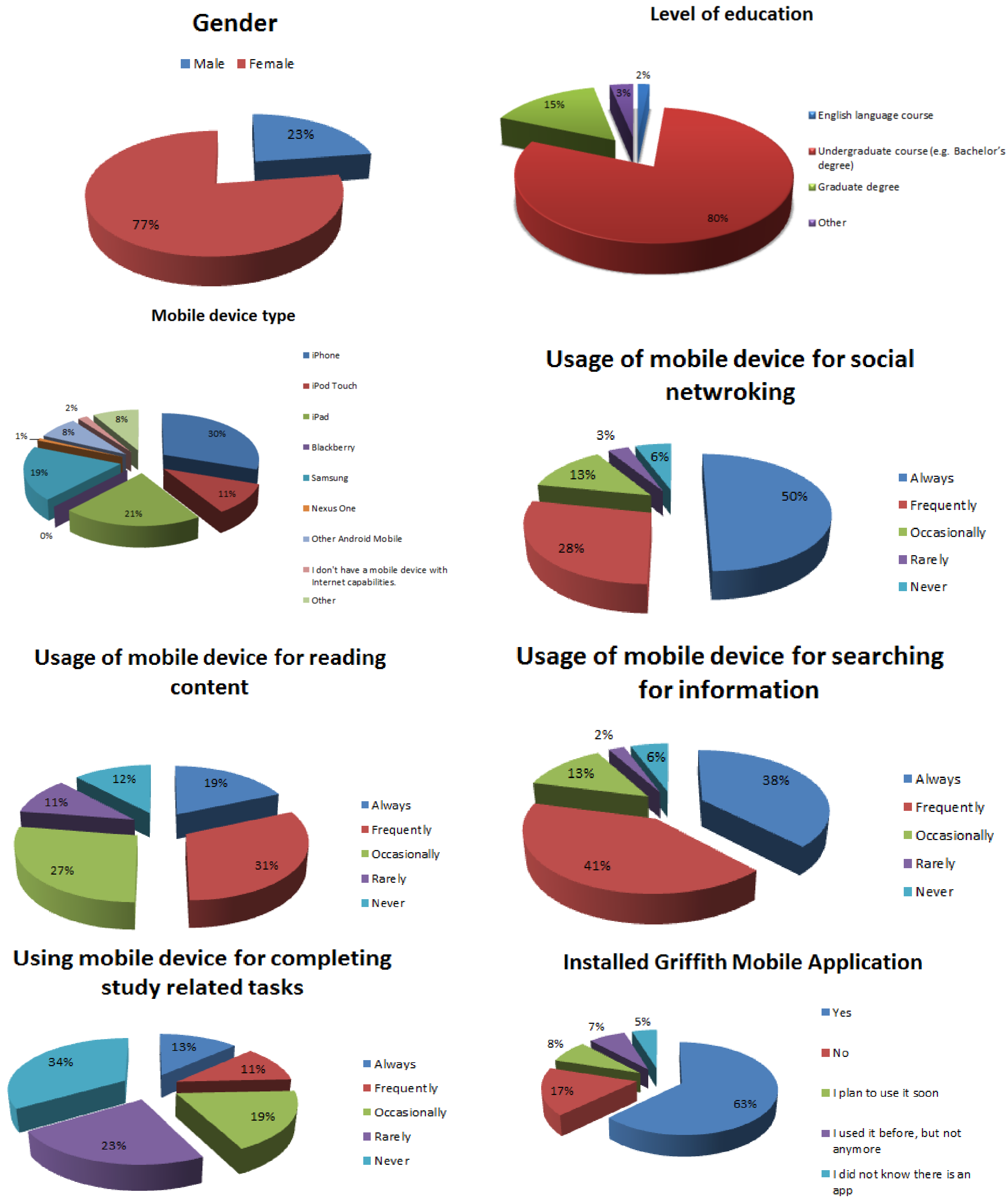


Fig.6. Respondants demographics

c) H6: Students' satisfaction would positively affect students' intention to use mobile Learning-systems.

The correlation analysis result in Table IX below shows that there is a fairly strong positive and significant relationship between SS and BIU, r(124) =.686, p < .005. This correlation indicates that students' satisfaction about mobile-learning systems will increase students' behavioural intention to use mobile-learning systems. Hence, H6 is supported.

TABLE.X. SS AND BIU CORRELATIONS

Correlations		
Factors		BIU
SS	r-value	$.686^{**}$
	p-value	.000
	N	124
SS: System Satisfaction; BIU: Behavioural intention to use		

d) H7: Information quality would positively affect information satisfaction of mobile Learning-systems.

The correlation analysis result in Table X below shows that there is a fairly strong positive and significant relationship between IQ and IS, r(124) =.847, p < .005. This correlation suggests that when information quality increases, information satisfaction of mobile-learning systems will increase. Hence, H7 is supported.

TABLE.XI. IQ AND IS CORRELATIONS

Correlations		
Factors		IS
IQ	r-value	$.847^{**}$
	p-value	.000
	N	124
IQ: Information Quality; IS: Information Satisfaction		

e) H8: System quality would positively affect system satisfaction of mobile Learning-systems.

The correlation analysis result in Table XI below shows that there is a strong positive and significant relationship between SQ and SS, r(124) =.835, p < .005. This correlation suggests that when system quality increases, students' satisfaction of mobile-learning systems will increase. Hence, H8 is supported.

TABLE.XII. SQ AND SS CORRELATIONS

Correlations		
Factors		SS
SQ	r-value	$.835^{**}$
	p-value	.000
	N	124
SQ: System Quality; SS: System Satisfaction		

f)H9: Information satisfaction would positively affect performance expectancy.

The correlation analysis result in Table XII below shows that there is a strong positive and significant relationship between IS and PE, r(124) =.745, p < .005. This correlation suggests that when information satisfaction increases, students' performance expectancy will increase. Hence, H9 is supported.

TABLE.XIII. IS AND PE CORRELATIONS

Correlations		
Factors		PE
IS	r-value	$.745^{**}$
	p-value	.000
	N	124
IS: Information Satisfaction; PE: Performance expectancy		

g) H10: System satisfaction would positively affect effort expectancy.

The correlation analysis result in Table XIII shows that there is a strong positive and significant relationship between SS and EF, r(124) =.745, p < .005. This correlation suggests that when system satisfaction increases, students' effort expectancy will increase. Hence, H10 is supported.

TABLE.XIV. SS AND EF CORRELATIONS

Correlations		
Factors		EF
SS	r-value	$.708^{**}$
	p-value	.000
	N	124
SS: System Satisfaction; EF: Effort expectancy		

4) The relationship between lecturer attitude constructs and behavioural intention

a) H11: Lecturers' attitude toward using mobile devices would positively or negatively affect students' behavioural intention to use mobile Learning-systems.

The correlation analysis result in Table XIV below shows that there is a fairly a weak positive relationship between LT and BIU, r(124) =.312, p < .005. Since the correlation is significant, H11 is statistically supported.

TABLE.XV. LT AND BIU CORRELATIONS

Correlations		
Factors		BIU
LT	r-value	$.323^{*}$
	p-value	.000
	N	124
LT: Lecturer attitude; BIU: Behavioural intention to use		

The table below summarise the hypothesis after the testing was done.

TABLE.XVI. HYPOTHESIS SUMMARY

No.	Statement	Result
H1	Performance expectancy would positively affect students' behaviour intention to use mobile Learning-systems	Supported
H2	Effort expectancy would positively affect students' behaviour intention to use mobile Learning-systems	Supported
H3	Social influence would positively affect students' behaviour intention to use mobile Learning-systems	Supported

H4	System quality would positively affect students' satisfaction about mobile Learning-systems	Supported
H5	Information quality would positively affect students' satisfaction about mobile Learning-systems	Supported
H6	Students' satisfaction would positively affect students' intention to use mobile Learning-systems	Supported
H7	Information quality would positively affect information satisfaction of mobile Learning-systems	Supported
H8	System quality would positively affect system satisfaction of mobile Learning-systems	Supported
H9	Information satisfaction would positively affect performance expectancy	Supported
H10	System satisfaction would positively affect effort expectancy	Supported
H11	Lecturers' attitude toward using mobile devices would positively affect students' behavioral intention to use mobile Learning-systems	Supported

VI. Discussion

The current study combines well-known theories that have been used in similar researches. Research model employ constructs found in UTAUT, IS Success, Modified IS Success, and other relevant literature.

In general, the statistical analysis shows that the findings of the current study are consistent with the original theories findings[5, 7, 40, 44-46]. All constructs within this study were proven to have positive correlations that are statistically significant. Overall, the analysis shows that students behavioural intention to use a mobile learning system is greatly affected by their effort expectancy and performance eexpectancy, information and system satisfaction, information and system quality. Additionally, with less effectiveness, lecturer attitude and social influences are less likely to influence one's behavioural intention. The findings suggest that all previously mentioned variables can positively influence students' behavioural intention to use mobile-learning systems. Noticeably, the relationship between performance expectancy and behavioural intention to use is stronger than the relationship between effort expectancy and behavioural intention. It is also noteworthy to mention that a large percentage of respondents were female. Hence further investigation on the gender effect would lead to further findings.

In Summary, the statistical analysis proves the ability of the proposed research model to measure the behavioural intention of students to use mobile-learning systems. Additionally, revision and further testing is required to validate the effect of lecturers' attitudes on students' behavioural intention.

VII. Conclusion

This study has explored acceptance theories and success models and their usage in mobile-learning context in higher-education. Despite the wide spread of mobile Learning-systems adoption, It has been noticed that there is a lack in investigating student behavioural intention to use such systems. Therefore this study proposes an integrated framework to measure student behavioural intention to use mobile Learning-systems. This framework combines an acceptance theory (UTAUT), and an IS-Success model (D&M). Constructs adapted from UTAUT are: 1) performance expectancy, 2) effort expectancy, and 3) social influences. Further, constructs adapted from D&M model are: information quality, 2) system quality, and 3) system satisfaction. Moreover, two additional constructs were found in the literature, namely, information satisfaction and system satisfaction. In addition, lecturers' attitude is introduced in this research. The research model was validated using a questioner distributed to university students via online survey. The necessary steps were undertaking to ensure content validity and reliability of the research instruments. The data were collected and analysis using SPSS to investigate the relationships proposed in the research hypotheses. The overall results confirms the findings found in similar literature, and shows a strong and positive correlations between the various study constructs and students' behavioural intention to use mobile-learning systems. Overall, students tend to develop a positive behavioural intention to use mobile-learning systems. Students believe that behavioural intention to use mobile-learning systems is greatly affected by the perception of its ease of use and usefulness. Additionally, Information and system quality are also important factors that improve students' behavioural intention by increasing students' satisfaction about information and system quality. In contrast, the results show that social influence and lecturers' attitude toward using mobile devices during lectures are less likely to hinder students from developing a positive behavioural intention.

The research findings are valuable for paving the future of assessing students' behavioural intention to use mobile-learning systems. However, the limitation of the current study should be noted. The following subsection describes some of the limitations and provides suggestions for future improvements.

A. Research limitations and future work

There are various limitations to this study. First, is the limited ability to generalise the findings. Online survey was employed in this study, and online surveys are not free of limitations[59]. The lack of personal contact with respondent may affect the response rate in web-based surveys more than in other type of surveys[60]. In addition, a higher sample size would lead to make the conclusion more general.

Further research may investigate the role of other variables, including users' characteristics, and adding more variables to the original constructs found in the models used for this research. A systematic research may also extend this exploratory study.

In addition, several other statistical tests including factor analysis, multiple regressions, and structural equation modelling, etc. could be conducted to confirm variables' validity. Those approaches were beyond the current study scope; however it remains an area of interest for a future research.

Finally, the research model is subject to further modification. The preliminary analysis shows that further validation and investigation may reveal other factors in the context of mobile-learning systems.

REFERENCES

[1] K. A. Marold, G. Larsen, and A. Moreno, "Web-based learning: is it working? A comparison of student performance and achievement in web-based courses and their in-classroom counterparts," in *IRMA Conference*, 2000, pp. 350-353.

[2] R. S. Cobcroft, S. J. Towers, J. E. Smith, and A. Bruns, "Mobile learning in review: Opportunities and challenges for learners, teachers, and institutions," 2006.

[3] M. O. M. El-Hussein and J. C. Cronje, "Defining Mobile Learning in the Higher Education Landscape," *Educational Technology & Society,* vol. 13, pp. 12-21, 2010.

[4] Y.-M. Huang, Y.-L. Jeng, and T.-C. Huang, "An Educational Mobile Blogging System for Supporting Collaborative Learning," *Educational Technology & Society,* vol. 12, pp. 163-175, 2009.

[5] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS quarterly,* pp. 425-478, 2003.

[6] W. H. DeLone and E. R. McLean, "Information systems success: the quest for the dependent variable," *Information systems research,* vol. 3, pp. 60-95, 1992.

[7] W. H. DeLone and E. R. McLean, "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update," *Journal of management information systems,* vol. 19, pp. 9-30, Spring2003 2003.

[8] Y. S. Wang, M. C. Wu, and H. Y. Wang, "Investigating the determinants and age and gender differences in the acceptance of mobile learning," *British Journal of Educational Technology,* vol. 40, pp. 92-118, 2009.

[9] M. Fishbein and I. Ajzen, *Belief, attitude, intention and behavior: An introduction to theory and research*, 1975.

[10] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly,* vol. 13, pp. 319-340, 1989.

[11] I. Ajzen, "The theory of planned behavior," *Organizational behavior and human decision processes,* vol. 50, pp. 179-211, 1991.

[12] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "Extrinsic and intrinsic motivation to use computers in the workplace1," *Journal of applied social psychology,* vol. 22, pp. 1111-1132, 1992.

[13] A. Bandura, "Social foundations of thought and action," *Englewood Cliffs, NJ,* 1986.

[14] D. R. Compeau and C. A. Higgins, "Computer self-efficacy: Development of a measure and initial test," *MIS quarterly,* pp. 189-211, 1995.

[15] S. Taylor and P. Todd, "Assessing IT Usage: The Role of Prior Experience," *MIS quarterly,* vol. 19, pp. 561-570, 1995.

[16] H. Traindis, "Interpersonal behaviour," *Brooks/Cole, Monterey, CA,* 1977.

[17] R. L. Thompson, C. A. Higgins, and J. M. Howell, "Personal computing: toward a conceptual model of utilization," *MIS quarterly,* pp. 125-143, 1991.

[18] G. C. Moore and I. Benbasat, "Development of an instrument to measure the perceptions of adopting an information technology innovation," *Information systems research,* vol. 2, pp. 192-222, 1991.

[19] E. M. Rogers, *Diffusion of innovations*: Simon and Schuster, 2010.

[20] Y.-S. Wang, H.-Y. Wang, and D. Y. Shee, "Measuring e-learning systems success in an organizational context: Scale development and validation," *Computers in Human Behavior,* vol. 23, pp. 1792-1808, 2007.

[21] J. Heo and I. Han, "Performance measure of information systems (IS) in evolving computing environments: an empirical investigation," *Information & Management,* vol. 40, pp. 243-256, 2003.

[22] G. G. Gable, D. Sedera, and T. Chan, "Enterprise systems success: a measurement model," 2003.

[23] B. L. Myers, L. A. Kappelman, and V. R. Prybutok, "A comprehensive model for assessing the quality and productivity of the information systems function: toward a theory for information systems assessment," *Information Resources Management Journal (IRMJ),* vol. 10, pp. 6-26, 1997.

[24] Q. Min, S. Ji, and G. Qu, "Mobile commerce user acceptance study in China: a revised UTAUT model," *Tsinghua Science & Technology,* vol. 13, pp. 257-264, 2008.

[25] Y.-L. Wu, Y.-H. Tao, and P.-C. YangJ, "Using UTAUT to explore the behavior of 3G mobile communication users," in *Industrial Engineering and Engineering Management, 2007 IEEE International Conference on*, 2007, pp. 199-203.

[26] K. Jairak, P. Praneetpolgrang, and K. Mekhabunchakij, "An acceptance of mobile learning for higher education students in Thailand," in *Sixth International Conference on eLearning for Knowledge-Based Society, Thailand*, 2009, pp. 17-18.

[27] C. Carlsson, J. Carlsson, K. Hyvonen, J. Puhakainen, and P. Walden, "Adoption of mobile devices/services—searching for answers with the UTAUT," in *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on*, 2006, pp. 132a-132a.

[28] J. Park, S. Yang, and X. Lehto, "Adoption of mobile technologies for Chinese consumers," *Journal of Electronic Commerce Research,* vol. 8, pp. 196-206, 2007.

[29] T. McGill, V. Hobbs, and J. Klobas, "User developed applications and information systems success: A test of DeLone and McLean's model," *Information Resources Management Journal (IRMJ),* vol. 16, pp. 24-45, 2003.

[30] P. B. Seddon, "A respecification and extension of the DeLone and McLean model of IS success," *Information systems research,* vol. 8, pp. 240-253, 1997.

[31] J. Etezadi-Amoli and A. F. Farhoomand, "A structural model of end user computing satisfaction and user performance," *Information & Management,* vol. 30, pp. 65-73, 1996.

[32] D. L. Goodhue and R. L. Thompson, "Task-technology fit and individual performance," *MIS quarterly,* pp. 213-236, 1995.

[33] A. Barker, G. Krull, and B. Mallinson, "A proposed theoretical model for m-learning adoption in developing countries," in *Proceedings of mLearn*, 2005, p. 4th.

[34] M. Sharples, J. Taylor, and G. Vavoula, "Towards a theory of mobile learning," *Proceedings of mLearn 2005,* vol. 1, pp. 1-9, 2005.

[35] H. Sun and P. Zhang, "The role of moderating factors in user technology acceptance," *International Journal of Human-Computer Studies,* vol. 64, pp. 53-78, 2006.

[36] A. Rai, S. S. Lang, and R. B. Welker, "Assessing the validity of IS success models: An empirical test and theoretical analysis," *Information Systems Research,* vol. 13, pp. 50-69, 2002.

[37] R. Stockdale and M. Borovicka, "Using Quality Dimensions in the Evaluation of Websites," *Information and Communication Technologies in Tourism 2006,* pp. 344-344, 2006.

[38] I. Chang, Y.-C. Li, W.-F. Hung, and H.-G. Hwang, "An empirical study on the impact of quality antecedents on tax payers' acceptance of Internet tax-filing systems," *Government Information Quarterly,* vol. 22, pp. 389-410, 2005.

[39] B. H. Wixom and P. A. Todd, "A theoretical integration of user satisfaction and technology acceptance," *Information systems research,* vol. 16, pp. 85-102, 2005.

[40] P. Seddon and M.-Y. Kiew, "A partial test and development of DeLone and McLean's model of IS success," *Australasian Journal of Information Systems,* vol. 4, 2007.

[41] J. Roldán and A. Millán, "Analysis of the information systems success dimensions interdependence: An adaptation of the DeLone & McLean's model in the Spanish EIS field," in *BITWorld 2000. Conference Proceedings,* 2000.

[42] S. Rivard, G. Poirier, L. Raymond, and F. Bergeron, "Development of a measure to assess the quality of user-developed applications," *ACM SIGMIS Database,* vol. 28, pp. 44-58, 1997.

[43] P. E. Pedersen and R. Ling, "Modifying adoption research for mobile Internet service adoption: Cross-disciplinary interactions," in *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*, 2003, p. 10 pp.

[44] A. T. Brubaker, "Faculty perceptions of the impact of student laptop use in a wireless internet environment on the classroom learning environment and teaching," Master's thesis, School of Information and Library Science, University of North Carolina, Chapel Hill, NC, 2006.

[45] J. Gikas and M. M. Grant, "Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media," *The Internet and Higher Education,* vol. 19, pp. 18-26, 2013.

[46] W. Alsaggaf, M. Hamilton, and J. Harland, "Mobile devices in computer programming lectures: Are CS lecturers prepared for mobile learning?," in *Computer Science & Education (ICCSE), 2012 7th International Conference on*, 2012, pp. 1369-1374.

[47] A. Bryman and E. Bell, *Business Research Methods*. New York: Oxford University Press, 2011.

[48] J. Collis and R. Hussey, *Business research: a practical guide for undergraduate and postgraduate students*: Basingstoke : Palgrave Macmillan, 2009.

[49] K. B. Wright, "Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services," *Journal of Computer-Mediated Communication,* vol. 10, pp. 00-00, 2005.

[50] A. Bryman, *Social Research Methods*. New York: Oxford University Press, 2012.

[51] U. Sekaran and R. Bougie, *Research Methods for Business: A Skill Building Approach*: John Wiley & Sons, 2010.

[52] W. G. Zikmund, J. C. Carr, and M. Griffin, *Business research methods*: CengageBrain. com, 2012.

[53] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika,* vol. 16, pp. 297-334, 1951.

[54] J. Hair, W. Black, B. Babin, R. Anderson, and R. Tatham, "Multivariate Data Analysis: Pearson Education," *New Jersey: Hoboken,* 2006.

[55] J. C. Nunnally, *Psychometric Theory*: New York: McGraw-Hill, 1967.

[56] U. Sekaran, *Research methods for business: A skill building approach*: New York, USA: John Wiley & Sons., 2006.

[57] D. De Vaus, *Surveys in Social Research*: Taylor & Francis, 2002.

[58] C. P. Dancey and J. Reidy, *Statistics Without Maths for Psychology*: Prentice Hall, 2004.

[59] F. J. Fowler, *Survey Research Methods*, 4th ed.: SAGE Publications, 2009.

[60] T. Shih and X. Fan, "Comparing response rates from web and mail surveys: A meta-analysis," *Field methods,* vol. 20, pp. 249-271, 2008.

Appendices

A. Research instruments

Mobile Learning-system usage: An integrated framework to measure students' behavioural intention ***Scales and items***						
Demographic Information						
1. **Section I**		**Demographic Characteristics Information**				
Q	**Variable**	**Value**				
SUR	Are you taking this survey on a mobile device?	SUR1: ☐ Yes SUR2: ☐ No				
AGE	In which category is your age?	AGE1: ☐18-24 years AGE2: ☐25-34 years AGE3: ☐35-44 years AGE4: ☐other, please specify:_______				
GEN	Please specify your gender	GEN1: ☐Male GEN2: ☐Female				
EDU	Level of education(Current course)	EDU1: ☐English language course EDU2: ☐Undergraduate course (e.g. Bachelor's degree) EDU3: ☐Graduate degree(Please specify:_________) EDU4: ☐Other, please specify:__________				
OWN	Please indicate the electronic equipment you currently own or plan to buy in the next three months. (Select all that apply)	OWN1: ☐ Netbook OWN2: ☐ Desktop OWN3: ☐ Laptop OWN4: ☐ Mobile phone (NOT Internet-capable) OWN5: ☐ Internet-enabled mobile device (e.g., smartphone, tablet, etc.) OWN6: ☐ Dedicated e-book device (e.g., Kindle, Nook, Sony Reader, etc.) OWN7: ☐ MP3 Player OWN8: ☐ Other ,please specify:_______________				
DTPE	Which of the following Internet-enabled mobile devices do you currently use? (Select all that apply.)	DTPE1: ☐ iPhone DTPE2: ☐ iPod Touch DTPE3: ☐ iPad DTPE4: ☐ Blackberry DTPE5: ☐ Samsung DTPE6: ☐ Other Android Mobile (Please specify) DTPE7: ☐ Nexus One DTPE8: ☐ I don't have a mobile device with Internet capabilities. DTPE9: ☐ Other ,please specify:_______________				
ACC	What library/academic information or resources have you tried to access using your mobile device? (Select all that apply).	ACC1: ☐ View library hours ACC2: ☐ Ask a question ACC3: ☐ Using the directory to view contact information ACC4: ☐ Search library catalogue / databases ACC5: ☐ Request an item through interlibrary loan ACC6: ☐ Find out about labs ACC7: ☐ View campus news ACC8: ☐ View locations on the map ACC9: ☐ View your timetable ACC10: ☐ Renew library items ACC11: ☐ None of these ACC12: ☐ Other, please specify:_______________				
USE	***To what degree do you use your Internet-enabled mobile device for the following activities?***	*Always*	*Frequently*	*Occasionally*	*Rarely*	*Never*
USE1	Social networking	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
USE2	Reading content (e.g., e-books, articles, etc.)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
USE3	Getting news alerts	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐

USE4	Accessing email	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
USE5	Text messaging	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
USE6	Searching for information	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
USE7	Getting directions	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
USE8	Uploading content	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
USE9	Playing games	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
USE10	Listening to music or watching videos	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
Continued: Mobile Learning-system usage: An integrated framework to measure students' behavioural intention Scales and items						
USE11	Completing coursework or participating in lectures	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
GAPP	Have you used Griffith University Application for mobile devices?	GAPP1: ☐ Yes GAPP2: ☐ No GAPP3: ☐ I plan to use it soon GAPP4: ☐ I used it before, but not anymore GAPP5: ☐ I did not know there is an app				

UTAUT(adapted from Venkatesh, Morris, Davis, & Davis (2003))

2. **Section II**		**Performance Expectancy**						
		Strongly Disagree			*Neutral*			*Strongly Agree*
PEE1	I feel that mobile learning is useful.	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
PEE2	Mobile learning improves my study efficiency.	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
PEE3	Mobile learning improves my study convenience.	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
PEE4	Mobile learning lets me do study related tasks more quickly.	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

3. **Section III**		**Effort Expectancy**						
		Strongly Disagree			*Neutral*			*Strongly Agree*
EFE1	Skilfully using mobile learning is easy for me.	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
EFE2	I find that using mobile learning is easy.	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
EFE3	Learning how to use mobile learning is easy for me.	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
EFE4	My interaction with mobile learning is clear and understandable.	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

4. **Section IV**		**Social Influence**						
		Strongly Disagree			*Neutral*			*Strongly Agree*
SOI1	Those people that influence my behaviour think that I should use mobile learning	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SOI2	Those people that are important to me think that I should use mobile learning	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

5. **Section X**		**Behavioural Intention to Use**						
		Strongly Disagree			*Neutral*			*Strongly Agree*

BI1	I intend to use the mobile learning system in the future	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
BI2	I predict I would use the mobile learning system in the future	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
BI3	I plan to use the mobile learning system in the future	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

IS Success (adapted from DeLone & McLean (1992,2003))

6.	**Section V**	**Information Quality**						
		Strongly Disagree			*Neutral*			*Strongly Agree*
IQ1	The mobile learning system provides information that is exactly what you need (Content Accuracy)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
IQ2	The mobile learning system provides information you need at the right time (Availability)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
IQ3	The mobile learning system provides information that is relevant to your course (Usability, relevance)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
IQ4	The mobile learning system provides sufficient information for your purposes (Quantity of information)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
IQ5	The mobile learning system provides information that is easy to understand (Understandability)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
IQ6	The mobile learning system provides up-to-date information (Currency)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
IQ7	The mobile learning system provides information that appears readable, clear and well formatted (User interface)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
IQ8	The mobile learning system provides required information on time. (Timeliness)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
IQ9	The mobile learning system provides information that is suitably concise.	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

Continued: Mobile Learning-system usage: An integrated framework to measure students' behavioural intention
Scales and items

7.	**Section VI**	**System Quality**						
		Strongly Disagree			*Neutral*			*Strongly Agree*
SQ1	The mobile learning system allows a high level of customization for different courses	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SQ2	The mobile learning system provides for personalized information presentation	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SQ3	The mobile learning system is easy to use	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SQ4	The mobile learning system is user-friendly (Easy to learn)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SQ5	The mobile learning system provides a high of availability (Access)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SQ6	The mobile learning system provides an appropriate level of on-line assistance and explanation (User requirements)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SQ7	The mobile learning system provides interactive features for an effective user experience	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SQ8	The mobile learning system provides satisfactory support to users of the system (Help and training)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SQ9	The mobile learning system has features that support the needs of a range of different courses (Flexibility)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SQ10	The mobile learning system has a high level of reliability	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SQ11	The mobile learning system provides high-speed information access (Efficiency)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

Modified IS Success(Adapted from Seddon and Kiew(2007))

8. Section VII		User Satisfaction						
		Strongly Disagree			*Neutral*			*Strongly Agree*
US1	Mobile learning systems is effective	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
US2	Mobile learning systems is efficient	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
US3	Overall, I am satisfied with mobile learning systems	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

9. Section VIII		Information Satisfaction						
		Strongly Disagree			*Neutral*			*Strongly Agree*
IS1	Overall, the information I get from mobile learning system is very satisfying	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
IS2	I am very satisfied with the information I receive from mobile learning system	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

10. Section IX		System Satisfaction						
		Strongly Disagree			*Neutral*			*Strongly Agree*
SS1	All things considered, I am very satisfied with mobile learning system	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
SS2	Overall, my interaction with mobile learning system is very satisfying	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

Lecturers' attitude (New scale)

11. Section X		Lecturers' attitude						
		Strongly Disagree			*Neutral*			*Strongly Agree*
LT1	I can use my mobile device in a formal learning environment e.g. searching resources in lectures	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
LT2	Lectures encourage me to use mobile devices device in a formal learning environment e.g. searching resources in lectures	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
LT3	Lecturers say that mobile devices sometimes can be very distracting.	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

Realising Dynamism in MediaSense Publish/Subscribe Model for Logical-Clustering in Crowdsourcing

Hasibur Rahman*, Rahim Rahmani, Theo Kanter
Department of Computer and Systems Sciences (DSV)
Stockholm University
Nod Buildning, SE-164 07 Kista, Sweden

***Abstract*—The upsurge of social networks, mobile devices, Internet or Web-enabled services have enabled unprecedented level of human participation in pervasive computing which is coined as crowdsourcing. The pervasiveness of computing devices leads to a fast varying computing where it is imperative to have a model for catering the dynamic environment. The challenge of efficiently distributing context information in logical-clustering in crowdsourcing scenarios can be countered by the scalable MediaSense PubSub model. MeidaSense is a proven scalable PubSub model for static environment. However, the scalability of MediaSense as PubSub model is further challenged by its viability to adjust to the dynamic nature of crowdsourcing. Crowdsourcing does not only involve fast varying pervasive devices but also dynamic distributed and heterogeneous context information. In light of this, the paper extends the current MediaSense PubSub model which can handle dynamic logical-clustering in crowdsourcing. The results suggest that the extended MediaSense is viable for catering the dynamism nature of crowdsourcing, moreover, it is possible to predict the near-optimal subscription matching time and predict the time it takes to update (insert or delete) context-IDs along with existing published context-IDs. Furthermore, it is possible to foretell the memory usage in MediaSense PubSub model.**

Keywords*—Internet; crowdsourcing; pervasive computing; context information; dynamism; context-ID; logical-clustering; Publish/Subscribe; MediaSense

I. Introduction

The penetration of pervasive devices is escalating and the rate of proliferation is always on the rise. This pervasiveness of computing devices has paved the way where any situation can be sensed and analyzed anywhere for anything. This directly corresponds to the distributed dissemination and acquisition of context information from physical objects. This has become possible by and large due to spontaneous human participation from online community which is most popularly known as *crowdsourcing*. This trend of crowdsourcing has been facilitated by the deployment of pervasive devices along with increasing popularity of Internet-enabled services and the trend is expected to upsurge. For instance, billions of mobile devices are already in use today and Ericsson predicts that 50-500 billion mobile devices will be in use by 2020 [1]. This coupled with increased deployment of sensors in the Internet-of-Things (IoT) will empower human to spontaneously participate in the crowdsourcing. Social-networks are anticipated to contribute to this cause as well, for example, a tweet feed can be considered as sensor data [11]. This surge of social networks, mobile devices, Internet or Web-enabled services have enabled unprecedented level of human participation in crowdsourcing which has been branded as “human-in-the-loop-sensing” or citizen sensor networks [12, 13]. This phenomenon has allowed us to encounter vast amount of real-time crowd-sourced data from distributed context sources. Ericsson envisions a world which is connected in real-time with people using things around us to create new innovative ideas- which is known as the Networked Society [1]. This Networked Society can be viewed as another way of defining the crowdsourcing. In a nut-shell, the followings are the properties and requirements for crowdsourcing:

- People
- Pervasive devices
- Internet or Web-enabled services
- Surrounding things
- Context Information

Although crowdsourcing is gaining popularity very fast and this, however, brings forth many challenges in the real-time distributed systems communication. Sharing heterogeneous context information obtained from distributed sources is one of them [4, 5, 11]. Publish/Subscribe (*PubSub*) model has perhaps emerged as most popular and efficient form of communication system to sharing ubiquitous context information. PubSub is an enabler for real-time context information sharing and providing means of notification for distributed devices [4, 5, 6, 7, 11]. By leveraging the PubSub in the crowdsourcing model can unravel the challenge of sharing context information in real-time [18].

Research in pervasive computing has resulted in MediaSense and was originally developed by the research group called *Immersive Networking* as context sharing platform in the Internet-of-Things domain based on peer-to-peer (p2p) technologies [2, 3]. MediaSense can run on any platform that runs JAVA. However, the promise and potential of

MediaSense makes it a good candidate to utilize it beyond the mentioned scope. It has the potential to be utilized in crowdsourcing domain. MediaSense is an open source platform which can be used for real-time and scalable seamless context sharing [2, 3].

In response to the challenge of sharing context information in crowdsourcing, our previous paper presented the scalable MediaSense platform as the PubSub model [18]. Results suggested that MediaSense platform is very fast, efficient and capable of supporting large-scale system. However, as crowdsourcing evolve around pervasive devices and pervasive computing is always changing and this dynamic nature of pervasive computing further challenges the scalability of PubSub model. A PubSub model must cope with the fast varying anytime, anywhere computing i.e. crowdsourcing. The distributed objects with heterogeneous context sources demand scalable computing when detecting changes and adjusting accordingly. The changes could be anything such as network connectivity, bandwidth, insertion and deletion of PubSub items, etc. Moreover, since logical-clustering involves physically distributed but logically synchronized sinks, hence it is mandated that we investigate its stability in case of failure of one of the sinks. The natural question arises what happens one of the sinks is down? Will the system be stable? Can MediaSense still be able to synchronize without failed sink(s)? Therefore, this mandates that we further examine MediaSense's scalability in dynamic environment. The aim is to enable real-time response to the fast varying nature of crowdsourcing. The massive scale of context information in crowdsourcing requires adjusting to the dynamic environment along with efficient and scalable acquisition, dissemination, and management. This paper particularly enlightens MediaSense's impact as PubSub model for dynamic crowdsourcing environment.

The rest of the paper is organized as follows: section II shows the related work, section III outlines the motivation of the work, section IV draws the approach while section V demonstrates the evaluation of the work, finally section VI concludes the paper and briefly hints at the future work.

II. Related Work

Related work in the aforementioned scenario focused on feasibility of using Publish/Subscribe model for mobile systems [4] where they focused on scalability and mobility issues; for mobile crowdsensing which focused on real-time data delivery and saving energy [5]. And others have proposed different methods to implement PubSub, for example, Le Subscribe proposed web based publish/subscribe system [6, 7], the Toronto Publish/Subscribe System (ToPSS) utilized DBMS-based matching algorithm [8] and PARDES implemented rule-based matching algorithm [9] for PubSub model. None of the above mentioned model alone offers the advantages that MediaSense offers as highlighted before.

Franco in [10] portrayed that spontaneous human participation i.e. crowdsourcing is pivotal for future pervasive computing. The human engagement in distributed collaboration would enrich the urban networks which will implement the idea of sensing, actuating and computing anything anywhere and anytime. Human participation in real-time crowdsourcing is further highlighted in [12, 13]. Demirbas et al. in [11] also illustrated crowd-sourced sensing and they showed Twitter as an example of achieving this. Ericsson [1] predicts that in future people will be connected along with things and will produce innovative ideas through the Networked Society. All these researches show that heterogeneous context will be generated from distributed sources in real-time. In light of this, one of our previous papers proposed the idea of logical-clustering based on context similarity [14] and we further demonstrated its performance in [15]. The definition of context by Dey AK (2001, [17]) is widely accepted, based on this our definition of context is: "*Sensor's flow packets that describe the current situation of the sensor*". Although our initial proposal concentrated generally on wireless sensor networks scenario and flow-sensors, however, our approach has the ingredients to suit the crowdsourcing platform as well. Similar context is the basis for logical-clustering. Context similarity is calculated based on similar flow of context of flow- sensors [14, 15]. Our proposal implies that heterogeneous context generated from distributed sources would be logically clustered based on context-similarity. The main goal of our research was to provide a mean for managing huge context information in a proficient manner. The challenge of sharing clustering identification has been addressed in our previous paper [18] by employing a PubSub model in MediaSense. This opens up the floodgate for sharing the clustering identification. This PubSub would act like a driving wheel for logical-clustering concept.

Zaslavsky in [19] portrayed that key to efficient pervasive computing i.e. crowdsourcing is to adjust applications' behavior and functionality. This underpins the need for applications' capability to cope with the dynamic environments. An application cannot be called scalable if it fails to address the aforementioned scenario. This was further discussed in [20] that it is inconvenient if pervasive system is static i.e. if not dynamic.

III. Motivation

The unprecedented power and promise of pervasive devices capitalized by human will lead the future pervasive environment. Huge amount of heterogeneous data i.e. context information generated from crowdsourcing necessitates proper management; and logical-clustering of context is one of the techniques to manage context information proficiently and share resources remotely thus enabling heterogeneous interoperability [14]. This approach can even be applied to the Networked Society concept where similar ideas from connected people can be categorized into a cluster meaning that clustering will be done based on similar context i.e. ideas. However, solution to the PubSub of context-IDs was missing in the existing proposal. Therefore, the primary motivation of this work is to address the PubSub issue of the proposed logical-clustering concept. In logical-clustering, each cluster is identified as context-ID and published on the Internet so that other interested entities can subscribe to the context-ID. The idea of logical-sink was utilized to control the enormous number of entities in a small-scale network. Logical-sink implies that sinks will be physically distributed but logically synchronized. PubSub is the enabler for accomplishing logical-sink. In our previous paper [18], we adopted MediaSense as PubSub enabler in logical-clustering. This approach solved the PubSub issue for both fronts i.e. for dissemination (publishing) of context-IDs in the Internet and for logical-sink synchronization. Fig. 1 (elaborated further in next section) shows the incorporation of MediaSense into the logical-clustering concept. Diversity and heterogeneity are not only related to the context information but also to the environment itself. Our previous paper dealt with the static scenario where only regular publish/subscribe items have been addressed. The paper did not take into consideration of dynamic situation where it might require to alter or update the context-IDs along with regular publish/subscribe. This motivated us to investigate further the MediaSense credibility whether it can match the demand of crowdsourcing dynamism. In addition, it has been observed that MediaSense initially takes some time to match a subscription compared with other distributed system such as PARDES system, therefore, another goal of this paper is to identify the reason behind this delay and propose a potential solution to the problem. With the ever increasing smartdevices and increasing popularity of intelligent systems, it is desirable to have a model which can predict the outcome in some capacity. And this paper will also explore if it is possible to predict the PubSub messages per second and the memory consumption for which the MediaSense was evaluated in the previous paper. Finally, it is unknown what happens when one of the physical sinks down and logical-sink synchronization, stability will further be evaluated.

IV. Approach

Firstly, this section briefly discusses how MediaSense works and follows by modifications made to the current MediaSense implementation to adjust to the approach i.e. using MediaSense as PubSub model is highlighted.

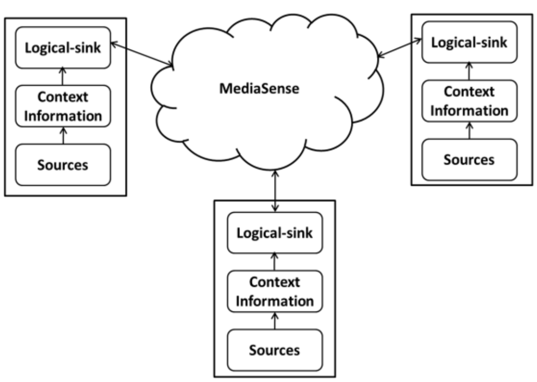


Fig. 1. MediaSense as PubSub model in logical-clustering

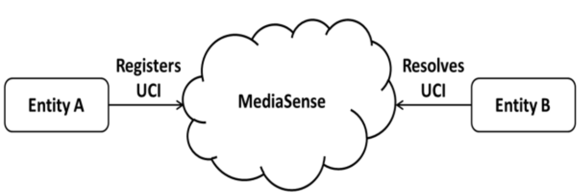


Fig. 2. MediaSense registering and resolving UCI

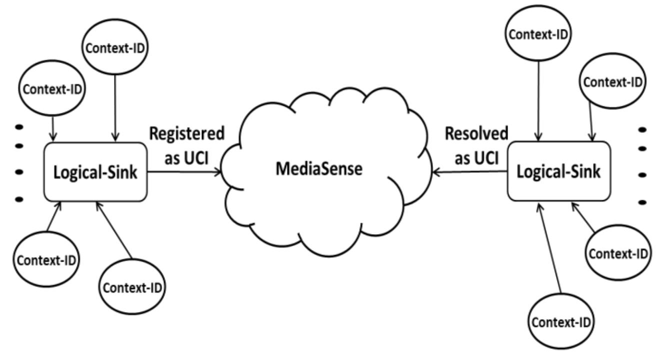


Fig. 3. Our approach to utilize MediaSense

A. MediaSense

MediaSense uses a p2p infrastructure and implemented in JAVA. Distributed Context eXchange Protocol (DCXP) is used to disseminate information between all the entities that are using the platform. MediaSense can offer real-time context sharing, and context entity is referred to as Universal Context Identifier (UCI) in MediaSense. An entity requires to resolving this UCI in order to fetch context information, but before an entity can fetch context information the entity that holds the context information needs to be registered. Fig. 2 gives an idea of how this mechanism works. Entity A registers a UCI in MediaSense using the Registrator class and entity B resolves the UCI by using the Resolver class to fetch context information associated with the resolved UCI. An entity can register more than one UCI. However, the only drawback with MediaSense is that an entity needs to know the UCI prior to resolving.

B. MediaSense as PubSub in logical-clustering

The contribution of this paper begins with adoption of MediaSense into logical-clustering concept. This sub-section describes the approach and modifications made to the MediaSense platform to fit into the proposal. Currently, an entity registers the host ID and hash key along with the UCIs. Host ID and hash key remain unchanged for a particular entity. The idea is that a logical-sink registers itself as UCI and the context-IDs associated with the logical-sink as UCI's data. Other logical-sink resided remotely resolves the UCI and fetches the context-IDs. This is shown in fig. 3. Logical-sink collects data i.e. context information from distributed sources e.g. sensors, mobile devices and other physical objects that produce context information, and is responsible for creating the context-IDs based on the context similarity (see fig. 1). Logical-sink needs to be synchronized as well i.e. changes in a physical sink should be synchronized with other physical sink(s). This synchronization could be achieved by the MediaSense PubSub model too. Fig. 4 illustrates this. In this later case, a physical sink would be registered as UCI and changes inside the sink would be shared with other physical sinks over MediaSense. Therefore, our approach would be evaluated for both these purposes.

However, the current MediaSense implementation does not support the registration of context information along with the UCI at the same time. Rather it collects context information and this is sent over MediaSense as a message. This method would incur delay in our approach as there might be millions of context-IDs to be published and subscribed. Hence, the MediaSense platform has been modified in a manner that the context information can be registered at the same time as UCI. Therefore, whenever a logical-sink is registered, its context information is also registered in parallel. This will further enable faster and real-time synchronization of context information. And, changes in the logical-sink can be updated using the MediaSense Updater class. Fig. 5 & 6 show the algorithms for UCI and context information registration and resolve. Algorithm for registration first begins with initializing MediaSense platform and starting the MediaSense bootstrap. MediaSense bootstrap needs to be initiated only once inside a network. As we assume that MediaSense entities are already up and running, so time to set MediaSense up is not included in the evaluation. The algorithm next checks if the UCI is registered. UCI is updated with new and old context information- if UCI is already registered. Otherwise, UCI is registered along with its context information. The registered UCI can be deleted and a logical-sink in essence can register multiple UCIs at the same time. This gives us flexibility; for example, an entity acting as both physical sink (part of logical-sink) and logical-sink (while communicating other logical-sinks) can communicate with other entities using different UCIs. The registered UCIs are saved on the MediaSense platform which means the context information is never lost, as long as the UCI is not deleted, when an entity dies or fails. This guarantees no central point of failure.

Fig. 6 shows the algorithm for resolving UCI. The algorithm first resolves the context information from the UCI if it exists. The algorithm then fetches context-IDs until the list is empty. The context-ID that is to be subscribed is then checked against the fetched context-IDs and a notification message can be sent to the subscription requestor when match is found. If the UCI is being requested to be resolved is nonexistent then a message notifies that UCI does not exist.

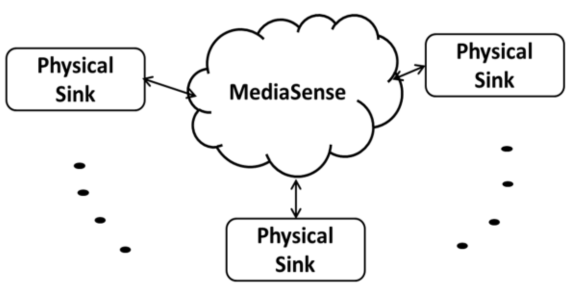


Fig. 4. MediaSense as PubSub for logical-sink synchronization

```
Algorithm UCIRegistration

Initialize MediaSense platform
Run the MediaSense bootstrap
// measurement starts from here
if UCI is not registered
       Invoke Registrator class
       Initialize registration and add UCI
invoking MediaSensePlatform's registerUCI method
       Add context information
else if
       Invoke Update class
       Initialize Updating and update UCI
invoking MediaSensePlatform's update method
       Update context information
end if

end UCIRegistration
```

Fig. 5. Algorithm for UCI and context information registration

```
Algorithm UCIResolve

Initialize MediaSense platform
// measurement starts from here
if UCI exists
       Invoke Resolver class
       Initialize resolve and resolve UCI
invoking MediaSensePlatform's resolveUCI method
       Resolve context information
       while context-ID list is not empty
             get context-ID
             if list contains context-ID
                    subscription matched
              end if
       end while
else if
       UCI does not exist
end if

end UCIResolve
```

Fig. 6. Algorithm for UCI and context information resolve

V. EVALUATION

This section first begins with highlighting the need for modification and then exhibits the evaluation of MediaSense as a PubSub model.

The evaluation can be divided into three parts: (i) PubSub for the context-IDs sharing in logical-clustering for which each published context-ID is matched for subscription, and (ii) PubSub for logical-sink synchronization for which all the changes are published to the other physical-sinks, and (iii) dynamic behavior of MediaSense.

TABLE I. REQUIRED TIME FOR PUBLISHING

# of published context-IDs	Current MediaSense	Modified MediaSense	% improvement
1000	7.34 ms	4.17 ms	76
10000	8.93 ms	5.37 ms	66
100000	10.74 ms	6.23 ms	72
200000	11.65 ms	6.69 ms	74

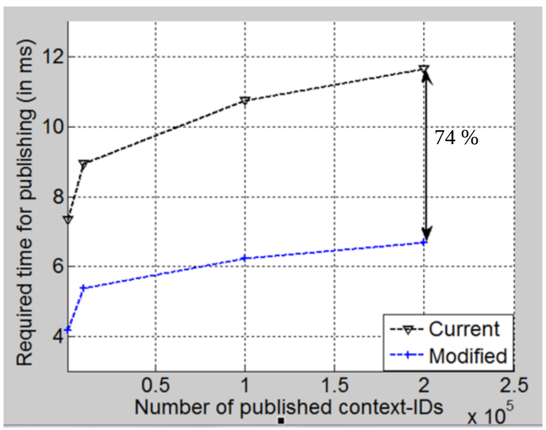


Fig. 7. Publishing time difference in MediaSense (current vs. modified)

A. Current vs. Modified MediaSense

In current MediaSense, if we want to share context-IDs then each context-ID would need to be registered as UCI. This will sustain delay. Table II summarizes the time required to publish items i.e. context-IDs on current and modified MediaSense platform. It can be clearly seen that current MediaSense takes longer time compared to the modified MediaSense- if we publish context-IDs as UCIs. Hence, it is efficient to register context-IDs as context information and sink as UCI. This way we can achieve nearly 74 % improvement. Fig. 7 further illustrates this.

B. MediaSense for logical-clustering

The PubSub model that we proposed initially for logical-clustering could send maximum 1000 messages/sec for PubSub events. However, we have achieved better result with MediaSense. It can support as high as 3537 messages/sec. This result has been obtained by running the PubSub for 1 second and result is the average for multiple simulations. This gives an increase of 254 % which outperforms our former idea. It clearly shows that MediaSense can be an efficient PubSub model. The rest of this sub-section will demonstrate performance of MediaSense for various scenarios and under assumption that all the MediaSense entities are already up and running. In order to evaluate its performance we have used three PCs with one PC acting as host sink and remaining two as recipient sinks. All three PCs have similar RAM size but the recipient sinks have different processors. The results have been obtained by simulating multiple times and the average results have been presented. Subscription matching time is shown in logarithmic scale and in milliseconds (ms).

Fig. 8 shows MediaSense's performance for different number of published context-IDs. This result is obtained for both published and subscribed duration. Context-IDs have been generated randomly using UUID in JAVA. For this particular scenario, each of the published context-ID is matched for subscription on the recipient sinks. It can be seen that both sinks give almost similar results. No significant fluctuation in terms of performance. MediaSense provides PubSub messages per second of around 2911, 1789, and 931 for context-IDs size of 10K, 50K, and 100K respectively. Although it is apparent that the performance reduces with the increase size of context-ID, but PubSub lowers only by one-third while the magnitude of the context-ID increased by ten-fold. This is due to the fact that time for resolving UCI increases when we want to publish and subscribe larger size. Moreover, subscription matching always vitiates when published item increases as can be seen from previous examples of PubSub [6, 7, 8, 9]. This can be understood from the fact that with the increased size of published item, the matching takes longer time.

Fig. 9 shows the subscription matching for context-IDs in MediaSense. Again almost identical performance for both sinks. Subscription matching duration understandably increases with the size of context-IDs. The result suggests that for hundred-fold increase in the context-ID size, matching duration increases only by 86 %.

Fig. 10 shows subscription matching time for a single context-ID. The ^{i}th context-ID is matched from i-size of the context-ID. Surprisingly, sinks have slightly different result for this scenario. The difference largely can be seen at the beginning (for 100K) and for 1 million. The one-millionth context-ID took 8.76 ms to match with the published context-IDs. While most of the PubSub systems are centralized and do not scale well in the distributed computing, the PARDES large-scale PubSub system in [9] is a distributed PubSub system which showed that one publication can be matched in 4.25 ms for 200K subscriptions, although for our approach we are matching subscription against published items and result illustrates that it takes 7.71 ms to match 200,000th item for 200,000 published items in real-time. This increase perhaps due to time required to resolve UCI with large context-IDs (see further fig. 14).

However, if we analyze fig. 11 it can be observed that the increase rate for subscription matching is much higher in PARDES compared to MediaSense. The matching rate increases nominally for MediaSense. It increases by merely 7% when context-IDs increase from 500K to 1 million and from 1 million to 2 million. As for PARDES, we see that it increases by 54%, 89%, and 125% when subscriptions increase from 25K to 50K, 50K to 100K, and 100K to 200K respectively. Since PARDES did not show its results beyond 200K and if we take the minimum increase rate which is 54% and plot them, then we see that PARDES overtakes MediaSense from 500K and beyond. MediaSense shows 99% improvement compared to PARDES for 2 billion context-IDs matching. This result signifies that our approach is easily suitable for large-scale PubSub scenarios and scales very efficiently with nominal increase in matching duration in a distributed large-scale scenario. The scalability efficiency can further be seen from table II and III. It is mentioned earlier that for all the PubSub systems, PubSub messages/sec decreases with the increase in published items. Le Subscribe system is a very efficient and fast PubSub system as outlined in [6, 7], but our approach has outperformed its counting algorithm as table II and III confirm. MediaSense achieves as high as 2058% increase in subscription matching and 1200% increase in PubSub messages/sec. Although Le Subscribe has other algorithms which performs better compared to its counting algorithm, but the other algorithms eliminate a portion of subscriptions to achieve this. This contradicts our approach and we do not eliminate any context-ID (i.e. subscription), hence other algorithms were not considered for comparison. And we have shown that our approach performs better compared to other approaches.

The above scenarios have been evaluated on the same network and with same Internet speed. In order to verify whether Internet speed plays a significant role in the MediaSense performance, we have tested our approach in a different network with one-third slower Internet speed. Fig. 12 illustrates this case. The result demonstrates that Internet does play a role in determining the performance. Interestingly, the fluctuation mostly varies between 5K and 20K. As for 50K and 100K, the fluctuation is insignificant. For example, for the size of 10K, network-2 (with low speed) shows 31 % performance reductions while for the 100K size, the decrease is merely 3 %. This indicates that although with low Internet speed MediaSense demonstrates slight performance reduction, however, the decrease rate is marginal.

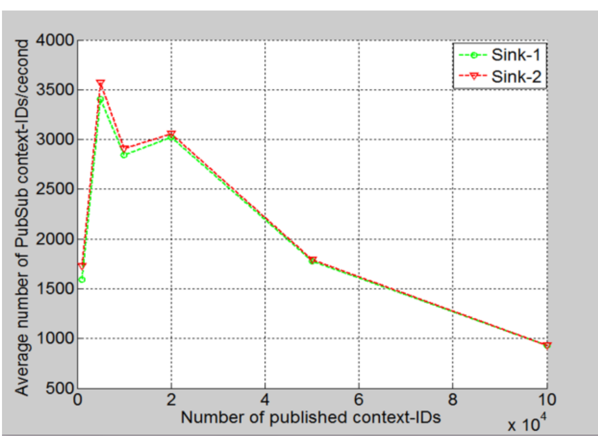


Fig. 8. MediaSense PubSub messages per second

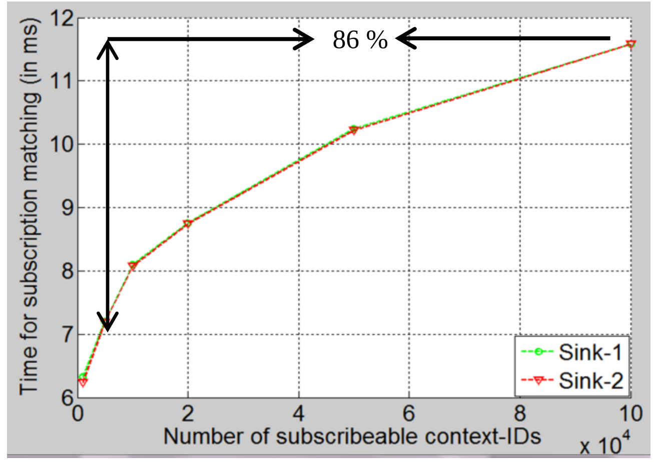


Fig. 9. MediaSense subscription matching

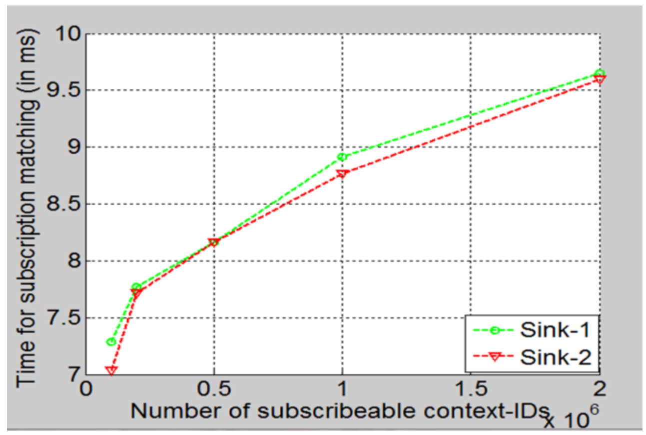


Fig. 10. MediaSense subscription matching for ith item

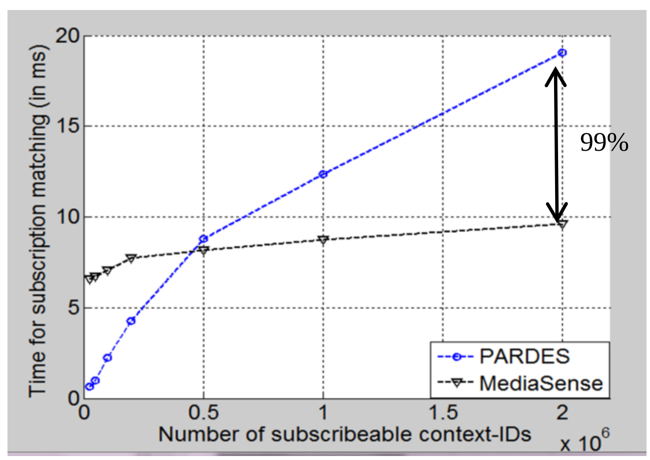


Fig. 11. Subscription macthing time comparisons

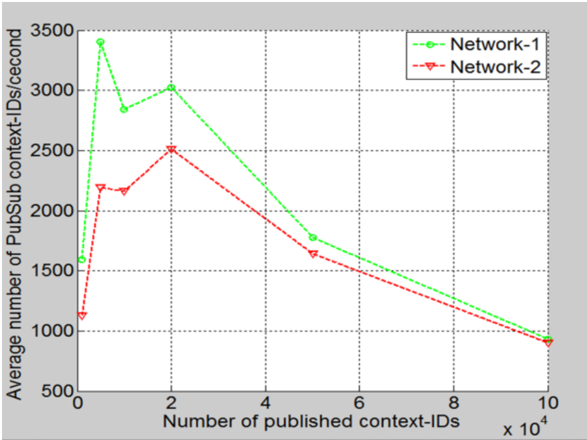


Fig. 12. MediaSense PubSub messages per second in different Internet speed

TABLE II. SUBSCRIPTION MATCHING

# of context-IDs	Le Subscribe (Counting)	MediaSense	% improvement
500 K	85 ms	14.76 ms	476
1 million	350 ms	16.22 ms	2058

TABLE III. PUBSUB MESSAGES/SEC

# of context-IDs	Le Subscribe (Counting)	MediaSense	% improvement
15 K	621	3151	407
1 million	7	91	1200

C. MediaSense for logical-sink

As for logical-sink i.e. synchronization of physical sinks, matching for published items is not required. In order to synchronize each physical sink, only the changes need to be retrieved in other sinks. And, depending on the nature of changes and need, each physical sink would decide whether to save the changes in a file or as UCI on the MediaSense. And, since no matching operation required in this case, MediaSense can provide as high as 9032 event changes per second. This is a further improvement by factor of nearly 3 compared to PubSub messages per second. This overwhelming number makes MediaSense a very competent and efficient tool for PubSub model in crowdsourcing- especially for the purpose of logical-clustering.

D. MediaSense memory usage

Memory usage plays an important part in the PubSub model evaluation as highlighted by earlier researches [7, 8, 9]. MediaSense is very efficient in terms of memory usage as well. Fig. 13 confirms this. Memory usage grows linearly. 37 MB of memory is required in order to store 1 million context-IDs. ToPSS PubSub prototype in [8] and Le Subscribe prototype (the counting algorithm was described in [6] and its memory usage was shown in [7]) required very large memory sizes, for example, ToPSS occupied minimum of 4400 KB memory to store 1000 subscriptions, and in our approach it is possible to store 1000 subscription with 39 KB of memory. This gives an 11216 % improvement in terms of memory usage for this particular scenario. However, this is not always the case as illustrated in table IV. The table further shows the comparison between these three PubSub models. MediaSense and Le Subscribe grow linearly. Table IV also reflects this where MediaSense's % improvement compared to ToPSS varies and the comparison is stable with Le Subscribe in terms of memory requirements. MediaSense betters Le Subscribe and ToPSS respectively by 163% and minimum by 451%.

TABLE IV. MEMORY USAGE

# of context-IDs	MediaSense	ToPSS (Kdb)	Le Subscribe (Counting)	% improvement
1000	0.038 MB	4.3 MB	-	11216 / -
1 million	37.1 MB	381.46 MB	97.66 MB	928 / 163
2 million	74.38 MB	762.94 MB	195.31 MB	926 / 163
5 million	185.97 MB	1024 MB	488.28 MB	451 / 163

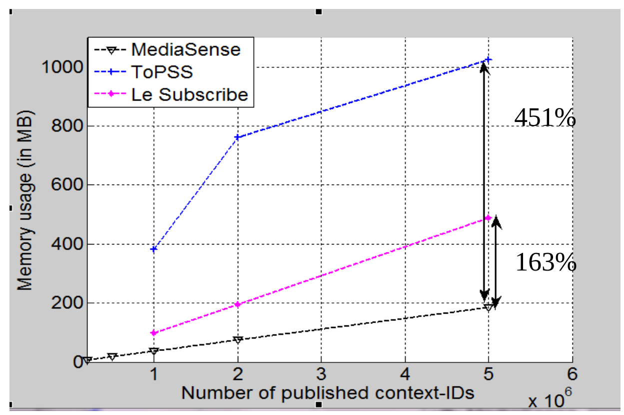


Fig. 13. MediaSense memory usage

E. UCI resolved delay analysis

We have seen in fig. 10 & 11 that context-ID matching takes bit long time initially and we further assumed that this could be due to the time that sink takes to resolves UCI. We have seen from fig. 10 & 11 that subscription matching grows linearly but initially takes some time. If the time required to resolve UCI can be ignored then this could result in faster subscription matching which is desirable in real-time computing. The following figures (fig. 14 & 15) further discuss the issue. First, fig. 14 shows the comparison for subscription matching between UCI resolved and without UCI resolved. The result in this particular figure has been simulated for context-ID matching for every published context-ID. The result is out of the blue for us, we did not expect this result. Our assumption was that without UCI resolved would result in faster context-ID subscription matching. However, MediaSense demonstrated almost identical performance for both scenarios.

For example, MediaSense demonstrated only 23% increased subscription matching time for UCI resolved compared to without UCI resolved for 5K published context-IDs. Moreover, this subscription matching time reduces to almost 0% if the published context-ID is increased to 100K. This could be understood from the fact that as we are matching for each published context-ID and time for subscription is matching is short (measured in ms) as well as for UCI resolving. Therefore, with the increase of published context-ID, the resulting subscription matching is independent of time required for UCI resolving. Nonetheless, if we now examine fig. 15 we can see the significance of discarding required time for UCI resolving.

Fig. 15 shows the subscription matching required for ^{i}th context-ID from i-size of the context-ID. Fig. 11 also showed the result for this scenario. Fig. 15 clearly shows the difference. Since pervasive computing is a dynamic environment and more often than not it is desirable to match a context-ID as fast as possible with minimal delay. This motivated us to look into a solution for finding a faster approach for context-ID matching.

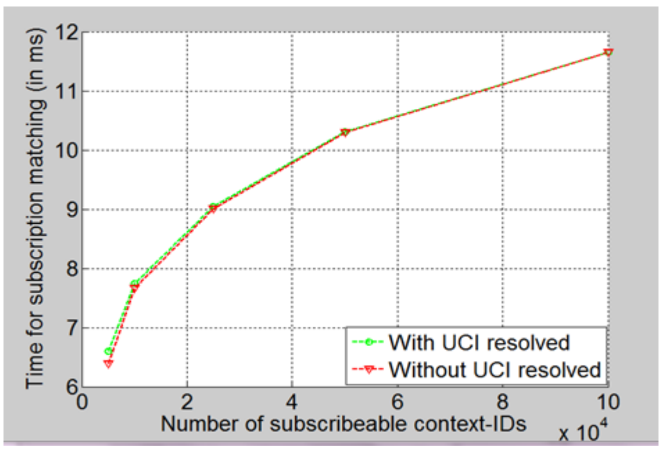


Fig. 14. Subscription matching with and without resolved UCI

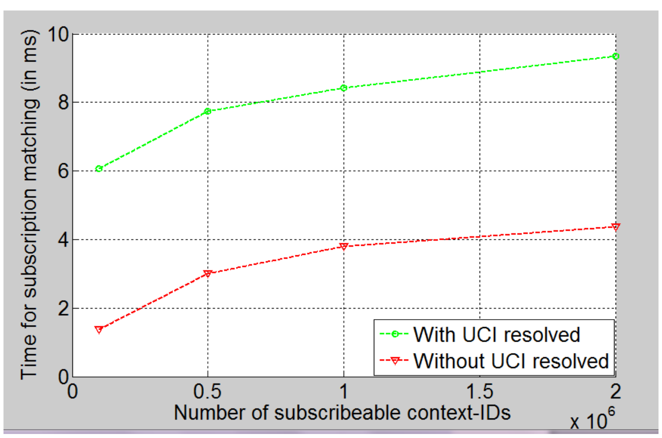


Fig. 15. ^{i}th Subscription matching with and without resolved UCI

Fig. 15 exhibits this. The figure shows the subscription matching from 100K to 2 million. Both results i.e. for both with and without UCI resolved qualitatively reveals similar performance.

However, without UCI resolved clearly outperforms other approach. The improvement percentage is significant. It betters the UCI resolving by 338% and 114% respectively for 100K and 2 million context-IDs. However, it leads to another research question if we ignore the UCI resolving then how do other sinks resolve the context-IDs? This could be done by employing adaptability and awareness in MediaSense which is part of our future work.

F. Dynamic MediaSense PubSub

The previous evaluations have been explored for static scenario which means it did not consider the dynamic environment. This sub-section will examine if MediaSense can fulfill the demand of crowdsourcing dynamism. The current MediaSense allows a UCI to be updated and deleted, however, since the MediaSense had been modified to fit into logical-clustering concept, therefore, the MediaSense has been further extended to adapt to crowdsourcing dynamism. The extended MediaSense now can be used to insert and delete any context-ID anytime. The remainder of this sub-section examines the MediaSense platform's performance for context-IDs insertion and deletion scenarios.

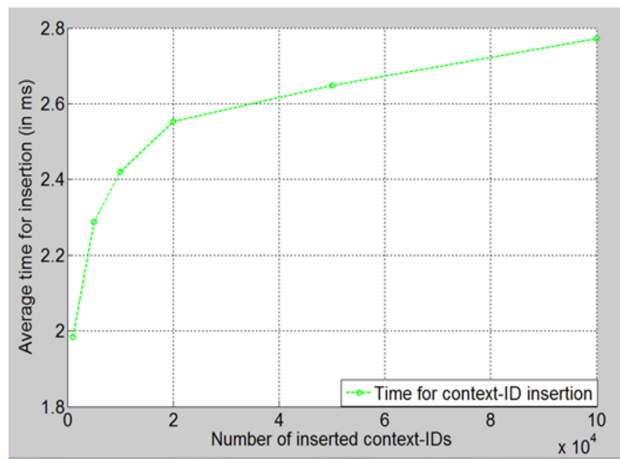


Fig. 16. Average time for context-ID insertion (I)

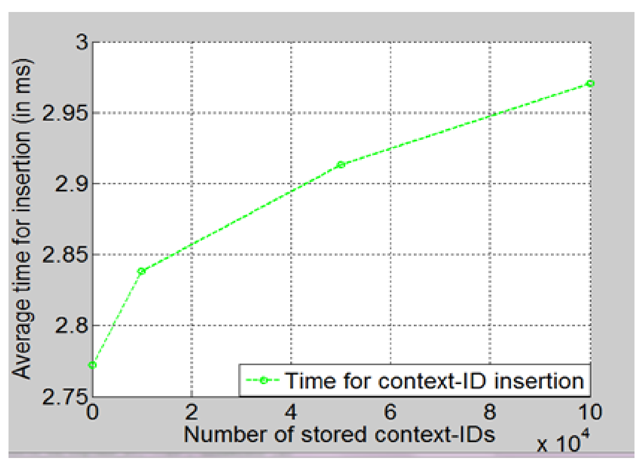


Fig. 17. Average time for context-ID insertion (II)

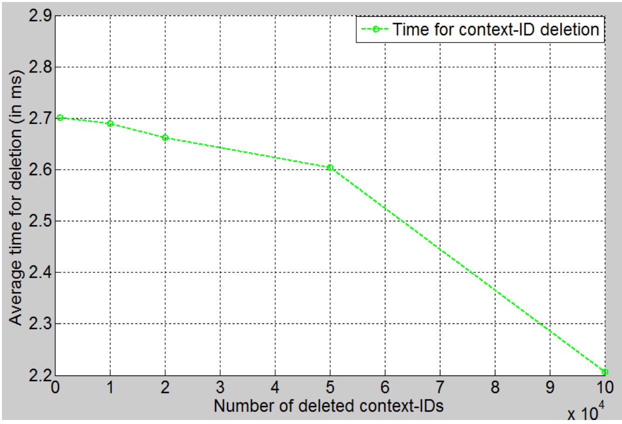


Fig. 18. Average time for context-ID deletion

Fig. 16 shows the context-ID insertion scenario for an already resolved UCI. As expected, the time for insertion increases with the increased number of context-ID. When the number of context-ID is increased from 1K to, average time for context-ID insertion is increased by 40%. The increase is not substantial compared to increase in number which is a 9900% upsurge. More importantly and perhaps significantly, this context-ID insertion follows a specific pattern for most cases.

For example: when the number of context-ID is increased from 5K to 10K the average time for insertion increases by 6%. The same goes true for 10K to 20K increases and for 50K to 100K. Therefore, we can conclude that a 100% increase in context-ID insertion would employ about 5% increases in time (see table V). This phenomenon could be very significant given that in dynamic real-time crowdsourcing it is always of great advantageous to predict the outcome beforehand. Therefore, with this pattern we can always predict the time required for context-ID insertions. Fig. 16 has been evaluated with very small stored context-ID, and in fig. 17 we further investigate if already stored context-ID for a UCI has any impact on average context-ID insertion. Thus we increase the number of stored context-ID in a UCI from 1 to 100K and the average time for context-ID insertion varies merely by around 3% and varies by just 7% when number of stored context-ID in a UCI increased from 1 to 100K. These numbers are very minimal compared to the increase in stored context-ID and does not offer a bottleneck for context-ID insertion.

Fig. 18 shows the context-ID deletion. This result is very surprising for us and it was totally unexpected. Our assumption was that average time for deletion of context-ID would grow with the increase of number of context-ID. Surprisingly, the average time decreases when number of context-ID to be deleted increases. However, if we closely investigate and look at the fig. 18 then we find out the time decrease is very minimal. The decrease is almost negligible when context-ID to be deleted increased from 1K to 50K (only 4%) and the rate is just 22% when context-ID to be deleted increased from 1K to 100K. This assures that MediaSense does not consume too much time to delete context-IDs.

This result is indeed beneficial for dynamic crowdsourcing as we want to acquire outcome faster in real-time.

TABLE V. INSERTION TIME % INCREASE

# of context-IDs increase	1K to 5K	10K to 20K	20K to 50K	50K to 100K	1K to 100K
% increase in average time for insertion	15	6	4	5	40

G. Prediction in MediaSense evaluation

In the above results, it has been observed in many scenarios that the results tend to follow a specific pattern. For example, it has been revealed by fig. 11, 14 & 15 that subscription matching grows linearly and so does the memory growth as observed by fig. 13 and table IV.

Therefore, the objective of this sub-section is to examine and propose some formulas where it can be possible to predict the outcome of the result. Since the real-time crowdsourcing is dynamic and it is imperative that the system is able to pre-determine the outcome. This intelligence in the MediaSense system would give us flexibility in terms of predicting such as time for subscription matching, memory occupation, etc. Table VI portrays the published time percentage increase when the number of context-IDs is increased. The observation indicates that published time increases between 4% – 6% for a 100% increase in the context-IDs size. And if we further analyze table VII we observe that this increase for published time follow a specific pattern. For example, for each 100% increase published time increases by about 5±1%. Even when we have 400% increases then MediaSense demonstrates around 16% - 18% increase. Hence, analyzing the above results the following formula for MediaSense published time increase can be written:

$$P\_T_i = \left((5 \pm 1)\ .\ P\_I_f\right)\% \ldots \ldots \ldots (1)$$

Where P_T_i is the published time increase and P_I_f is the percentage increase factor (for example, for a 100% increase P_I_f would be 1 and for a 400% increase P_I_f would be 4). Although by using eq. 1, it might not be always possible to predict exact published time increase, however, we can at least predict nearest value. As for subscription matching table VIII indicates that it varies always. This is understandable from the fact that while subscribing for a context-ID, MediaSense battles with bandwidth while resolving UCI, and it might not provide any stable equation. Nevertheless, we can at least provide an equation which can provide us a near optimal value for subscription matching. The equation can be written as:

$$S\_M_i = (15 \pm 5)\% \qquad \ldots \ldots \ldots \quad (2)$$

S_M_i is the subscription matching increase. Eqn. 2 is true only when each published context-ID is matched, but as for ^{i}th context-ID subscription macthing from i-size of the context-ID, the subscription matching increases by about 10% in most cases as indicated by table IX.

TABLE VI. PUBLISHED TIME % INCREASE (I)

# of context-IDs increase	1K to 2K	5K to 10K	10K to 20K	20K to 30K	25K to 50K	50K to 100K	100K to 200K
% increase in published time	6	6	5	4	4	4	6

TABLE VII. PUBLISHED TIME % INCREASE (II)

# of context-IDs increase	1K to 5K	2K to 10K	10K to 50K	25K to 100K
% increase in published time	18	17	16	12

TABLE VIII. SUBSCRIPTION MATCHING % INCREASE (I)

# of context-IDs increase	1K to 5K	2K to 5K	5K to 10K	10K to 25K	25K to 50K	50K to 100K
% increase in subscription matching	15	20	18	19	15	14

TABLE IX. SUBSCRIPTION MATCHING % INCREASE (II)

# of context-IDs increase	100K to 200K	250K to 500K	500K to 1 m	1 m to 2 m
% increase in subscription matching	10	9	9	11

It is also possible to predict the memory usage in MediaSense. This can be seen from the fig. 13 and table IV. The memory usage grows linearly and minimally. MediaSense memory usage corresponds to the following equation:

$$M_u = 0.0381 \, . \, N_{C\_id} \text{ (KB) where } N_{C\_id} \geq 5000 \ \dots \dots \dots \ (3)$$

Where, M_u is the memory usage and N_{C_id} is the total number of context-ID to be published.

As mentioned earlier that one of the objectives of this paper is to examine if MediaSense remains stable when one of the physical sinks down, according to our finding it does remain stable (the results are not shown here due to page limitation).

From the above results, it is clear MediaSense can adjust to the dynamic nature of crowdsourcing environment and fulfill the mentioned demand without any performance degradation. Moreover, it is also possible to predict the outcome of MediaSense PubSub result which makes MediaSense more attractive as a PubSub model.

VI. CONCLUSION

The growing popularity of crowdsourcing in pervasive computing gives rise to many challenges. Sharing context information in real-time is one of them for example in logical-clustering scenario. The challenge of sharing context information is unraveled by employing MediaSense as PubSub model. MediaSense demonstrated very efficient performance for the PubSub purpose and it performs better than existing PubSub models and requires only 9.59 ms to match two-millionth published context-ID, furthermore the memory requirement is very low. However, the results are analyzed only for static environment. The contribution of this extended paper begins with extending MediaSense to counter the dynamic nature of logical-clustering for crowdsourcing. The paper first proposes a solution for reducing the delay to subscription matching. The solution works very well for ^{i}th item subscription matching, however, when each published item is subscribed then the solution does not offer any improvement. Nevertheless, for the ^{i}th item case the new solution improves by 114% for two-millionth published context-ID which could be hugely significant in dynamic crowdsourcing. However, this solution brings forth a new research question: if we ignore the UCI resolving then how do other sinks resolve the context-IDs? This could be countered by employing adaptability and awareness in MediaSense which is part of our future work.

As for updating published context-IDs i.e. inserting or deleting context-IDs from an existing UCI. The result shows average time for insertion is just 5% for 100% increase in context-IDs. The deletion of context-ID demonstrated a surprising behavior, while deletion time was expected to rise with the escalation of context-ID but the result indicated the opposite. In addition, based on the acquired results few formulas have been presented to predict the outcome for publish and subscribe context-IDs time and for memory usage. The formulas could be very significant in dynamic logical-clustering since it would help to regulate the outcome beforehand.

Although MediaSense did live up to its expectation as scalable PubSub model for both static and dynamic environments but its viability can be further examined. For example: adaptability and awareness in MediaSense; to have prior knowledge of UCI before resolving; and how it will perform on devices with limited computational capabilities. Crowdsourcing heavily involves mobile devices; therefore MediaSense's performance on mobile devices will also be explored. Thus the mobility, energy (e.g. on android devices) issues of MediaSense along with performance in devices with limited computational capabilities (such as on raspberry pi) can be examined.

ACKNOWLEDGMENT

The work is partially supported by funding from the European Union FP7 MobiS project. We would also like to thank Mr. Jamie Walters and Mr. Johan Eliasson for their feedback about extending MediaSense.

REFERENCES

[1] [online] 5G Radio Access, Research and Vision: http://www.ericsson.com/res/docs/whitepapers/wp-5g.pdf [Last Accessed: 08-February-2014]

[2] T. Kanter, S. Forsström, V. Kardeby, J. Walters, U. Jennehag and P. Österberg., "MediaSense – an Internet of Things Platform for Scalable and Decentralized Context Sharing and Control," In: ICDT 2012,, The Seventh International Conference on Digital Telecommunications, pp. 27-32, April 2012.

[3] [online] MediaSense | The Internet of Things Platform, http://www.mediasense.se/ [Last Accessed: 08-February-2014]

[4] G. Cugola and H. Jacobsen, "Using Publish/Subscribe Middleware for Mobile Systems", In ACM SIGMOBILE Mobile Computing and Communications Review, vol. 6, pp.25–33, October 2002.

[5] I. P. Zarko, A. Antonic and K. Pripužic, "Publish/subscribe middleware for energy-efficient mobile crowdsensing". In Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication (UbiComp '13 Adjunct). Zurich, pp. 1099-1110, September 2013.

[6] J. Pereira, F. Fabret, _F. Llirbat, and D. Shasha, "Efficient matching for web-based publish/subscribe systems," 7th International Conference, CoopIS 2000, Eilat, Israel, September 6-8, 2000.

[7] F. Fabret, H. A. Jacobsen, F. Llirbat, J. Pereira, K A. Ross, and D. Shasha, "Filtering algorithms and implementation for very fast publish/subscribe systems", In Proceedings of the 2001 ACM SIGMOD international conference on Management of data (SIGMOD '01), 2001

[8] Ashayer, G.; Leung, H.K.Y.; Jacobsen, H.-A., "Predicate matching and subscription matching in Publish/Subscribe systems," Distributed Computing Systems Workshops, 2002. Proceedings. 22nd International Conference on , vol., no., pp.539,546, 2002

[9] E. Fidler. PADRES: A Distributed Content-Based Publish/Subscribe System. PhD thesis, University of Toronto, 2006.

[10] F. Zambonelli, "Pervasive urban crowdsourcing: Visions and Challenges", 2011 IEEE (PERCOM Workshops), pp. 578-583, March 2011.

[11] M. Demirbas, M.A. Bayir, C.G. Akcora, Y.S. Yilmaz, H. Ferhatosmanoglu, "Crowd-sourced sensing and collaboration using twitter," World of Wireless Mobile and Multimedia Networks (WoWMoM), 2010 IEEE International Symposium on a , vol., no., pp.1-9, June 2010.

[12] M.K. Boulos, B. Resch, D.N. Crowley, J.G. Breslin, G. Sohn, R. Burtner, W.A, Pike, E. Jezierski, K.Y.S Chuang, "Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples". International journal of health geographics, 10(1), 67, 2011.

[13] Sheth, A., "Citizen Sensing, Social Signals, and Enriching Human Experience," Internet Computing, IEEE , vol.13, no.4, pp.87,92, July-Aug. 2009.

[14] R. Rahmani, H. Rahman, and T. Kanter, "Context-Based Logical Clustering of Flow-Sensors - Exploiting HyperFlow and Hierarchical DHTs", In Proceeding(s) of 4th International Conference on Next Generation Information Technology, 2013 ICNIT, June 2013.

[15] R. Rahmani, H. Rahman, and T. Kanter, "On Performance of Logical-Clustering of Flow-Sensors", The International Journal of Computer Science Issues (IJCSI), Vol. 10, Issue 5, No 2, September 2013.

[16] A. Tootoonchian, Y. Ganjali, "HyperFlow: A Distributed Control Plane for OpenFlow", Proceedings of the 2010 Internet Network, 2010.

[17] Dey AK (2001), "Understanding and using context. Personal and Ubiquitous Computing", 5: 20–24.

[18] H. Rahman , R. Rahmani, and T. Kanter, "Enabling Scalable Publish/Subscribe for Logical-Clustering in Crowdsourcing via MediaSense", IEEE Science and Information Conference 2014, August 27-29, 2014, London, UK

[19] A. Zaslavsky, "Adaptibility and Interfaces: Key to Efficient Pervasive Computing", NSF Workshop series on Context-AwareMobile Database Management, Brown University, Providence, 24-25 January, 2002

[20] M. Miraoui, C. Tadj and C. b. Amar, "Dynamic Context-Aware Service Adaptation in a Pervasive Computing System", IEEE Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, 2009