

Predicting Quality of Answer in Collaborative Question Answer Learning

Kohei ARAI

Graduate School of Science and Engineering,
Saga University
Saga Japan¹

ANIK Nur Handayani

Electrical and Information Technology,
State University of Malang
Malang, Indonesia

Abstract— Studies over the years shown that students had actively and more interactively involved in a classroom discussion to gain their knowledge. By posting questions for other participants to answer, students could obtain several answers to their question. The problem is sometimes the answer chosen by student as the best answer is not necessarily the best quality answer. Therefore, an automatic recommender system based on student activity, may improve these situations as it will choose the best answer objectively. On the other side, in the implementation of collaborative learning, in addition to sharing information, sometimes students also need a reference or domain knowledge which relevant with the topic. In this paper, we proposed answer quality predictor in collaborative question answer (CQA) learning, to predict the quality of answer either from recommender system based on users activity or domain knowledge as reference information.

Keywords— collaborative question answer learning; domain knowledge; answer quality predictor; recommender.

I. INTRODUCTION

The concept of Collaborative Learning is two or more people learn or attempt to learn something together than independent. Different with individual learning, in collaborative learning student can exploit and share their resources and skills by asking, evaluating, monitoring one another's information and idea, etc [1]. Collaborative Learning is a model that knowledge can be created by sharing experiences within a population where members actively interact [2] [3]. Including both directly with face-to-face conversations [5] or using computer discussions (online forums, chat rooms, etc.) [6].

In [3] authors indicate that when they found some problem, students learn better when they learn together more frequently than working individually as members in a group. Indeed, the effectiveness of collaborative learning on the internet has been identified by various studies. Interaction among students is fostered as communication over the internet is unpretentious and convenient when addressing to a single user or multiple users. By posting questions for other participants to answer, students could obtain several answers to their question. The problem is sometimes the answer chosen by student as the best answer is not necessarily the best quality answer. The decision of an asker is influenced by subjective reasoning such as the relations between students, the asker's own point of view, his lack on the subject and others [7]. Therefore, an automatic recommender system may improve

these situations as it will choose the best answer objectively. On the other side, in the implementation of collaborative learning, in addition to sharing information, sometimes students also need a reference or domain knowledge which relevant with the topic. The function of domain knowledge is used as knowledge about the environment in which the target information operates as a reference. In [11], we had developed collaborative question answer (CQA) using domain knowledge and answer quality predictor. Besides providing answer quality predictor as a recommender, the system also provides an answer that is taken from the domain knowledge as a reference.

In this paper, we proposed answer quality predictor in collaborative question answer (CQA) learning, to predict the quality of answer either from recommender system based on users activity or domain knowledge as reference information. With the proposed system right after collaborative answer, then answer quality predictor will give recommendation from the entire student's answer. And in the same time QA tools will extract answer from domain knowledge. The information from domain knowledge and answer quality predictor will be reprocess in the recommender system to predict as a bad, medium, or good answer. The paper is organized as follows. First is introduction for the question and answering system. Section 2 presents the proposed method. Section 3 explains implementation and result. Finally section 4 is summary and conclusion of this paper.

II. PROPOSED METHOD

Our proposed method in this paper consists of four parts. There are data collection, annotator, feature extraction, and coefficient correlation with answers. We explored Decision Trees classifier to get high precision on the target class (Weka framework used in this study [15]). Figure 1 shows the architecture of the system.

A. Data Collection

There are two kind of data, first is data that derived from Indonesian Yahoo! Answers (<http://id.answers.yahoo.com/>) and choosing the internet and computer category. This data had been processed through the answer quality predictor [9] [10]. For the domain knowledge we used id.wikipedia [8], there were over 100.000 articles in the Indonesian Wikipedia project. We collected 556 data from answer yahoo that could be processed in the QA tools to extract answer from domain knowledge [11].

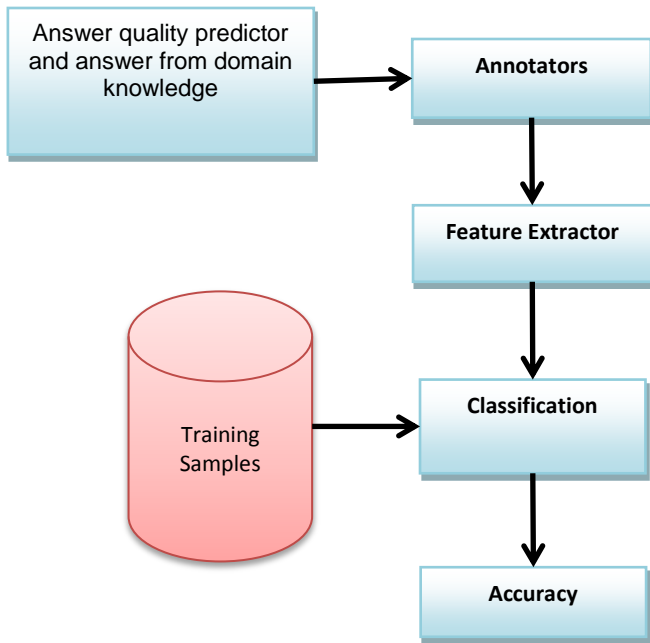


Figure1. Architecture of the System

TABLE 1. DATA COLLECTION

Category	Internet and Computer
Domain Knowledge	id.wikipedia
QA pairs	id.answeryahoo
Question	350
Answer	350
Data Training	250
Data Testing	100

The quality of a Q&A depends on the question part and answer part. For the question part we use most popular resolved question. Users could not get any useful information from bad questions. The reality bad questions always lead to bad quality answers. Therefore we decide to estimate the good answer by using annotators, and for all we got 350 Q&A pair. Table 1 shows the data collection of the research proposed.

B. Annotator

The quality of a question answer depends on both the question part and the answer part. It is often impossible to gather evaluative data about answers from the askers themselves, and that was the case here. The information of collaborative question answer is typically complex and subjective. We use annotators for manual judgment of answer quality and relevance. General, good answers tend to be relevant, information, objective, sincere and readable. We may separately measure these individual factors and combine scores to calculate overall the quality of the answer. Our annotators read answers, consider all of the above factors and specify the quality of answers in three levels: Bad, Medium and Good. Consider factors are as follows,

1. This answer provides enough information for the question. (informative)
2. This answer is polite (not offending). (polite)
3. This answer completely answers the whole question. (complete)
4. This is an easy to read answer. (readable)

5. This answer is relevant to the question. (relevant)
6. The answer contains enough detail. (detailed)
7. This answer is useful or helpful to address the question. (helpful)

C. Feature Extraction

A number of features have been identified in the literature for predicting the quality of answer. In this research, we used text feature and readability feature to predict the quality of answer. The selection of these two kinds of features based on the first Q&A pairs is already processed in the answer quality predictor; second there is internal reviewer from in term of domain knowledge system.

1) *Text features are those extracted from the textual content of the articles used by [12] [13] [14]. The general intuition behind them is that a mature and good quality text is probably neither too short, which could indicate incomplete topic coverage, nor excessively long, which could indicate verbose content. We use the following features :*

- a) *Character length: Number of characters for the answer.*
- b) *World length: Number of words for the answer.*
- c) *Sentences length: Number of sentences for the answer.*

2) *Readability Feature*

These features, first used in [15], are intended to estimate the age or US grade level necessary to comprehend a text. They comprise several metrics combining counts of words and sentences. The intuition behind these features is that good information should be well written, understandable, and free of unnecessary complexity. The features are;

a) *Automatic readability index (t_{rari}) : This metric was proposed in [16] and consists of using the average of word per sentence and the average of characters per words to estimate the readability.*

$$t_{rari} = 4.71 \frac{\text{characters}}{\text{words}} + 0.5 \frac{\text{words}}{\text{sentences}} - 21.43 \dots (1)$$

b) *Coleman Liau (t_{rci}): This metric was proposed in [84] and consists of the average of characters per word and the number of sentences in a fragment of 100 words (w_f).*

$$t_{rci} = 5.89 \frac{\text{characters}}{\text{words}} + 0.5 w_f - 15.48 \dots (2)$$

c) *The compound is a compound word or a combination of all of the basic morpheme that existed as a word that has a specifically pattern of phonological, grammatical, and semantic according to the rules of the language. The specific pattern distinguishes as a phrase or combination of words. For example, in Indonesian, kamar mandi is a compound word, while the baju hijau is the phrase while in English; the blackbird is a compound word, while the black bird is the phrase. [17]*

d) *Loan word is a word that derived from foreign languages that has been integrated into an Indonesian and generally accepted to be used. Indonesian has absorbing many words from other languages, especially those that have direct*

contact with the community, either through tarding (Sanskrit, Chinese, Arabic), or colonialism (Portuguese, Dutch, Japanese), as well as the development of science (English). [17]

e) Abstract noun is a type of noun (to explain the names of objects) which the existence could not be captured using human eyes and can only be imagined. The examples of abstract nouns are science, dreams, ideas, inspiration, happiness and others. [17]

f) Conjunction: In grammar, conjunction (abbreviated conj or cnj) is a part of speech that connects two words, sentences, phrases or clauses.

A discourse connective is a conjunction joining sentences. This definition may overlap with that of other parts of speech, so what constitutes a "conjunction" must be defined for each language. In general, a conjunction is an invariable grammatical particle, and it may or may not stand between the items it conjoins.[17]

D. Correlation Coefficient

The function of the correlation coefficient is to know how closely one variable is related to another variable [18], in this case the correlation between individual features and the annotators scores (good answers have higher scores: Bad = 0, Medium = 1, Good = 2). Table 2 shows coefficient correlation with the quality of answer. Surprisingly, all of feature has the strongest correlation with the quality of the answer, except for Auto read index and Coleman liau index.

From the calculation of Corr, we can see that text feature and readability feature affects the quality of the answers. In this study (Computer and Internet) using a lot of loan word (eg. Computer → komputer, Processor → prosesor, etc) , abstract noun word (eg. Principle → prinsip, definition → definisi, etc), and compound word (database, how it work → carakerja, etc). Auto read index and Coleman liau index is a feature used to calculate English language readability parameter. Several intelligent of United States of America used this parameter to measure readability of electronic letter. This convinced us that the character of one language is different one another.

TABLE 2. COEFFICIENT CORRELATION

Features	Correlation
Number of loan word	0.948
Number of Abstract noun word	0.999
Number of conjunction	0.828
Number of compound word	0.861
Number of char	0.971
Number of word	0.961
Number of sentence	0.928
Auto read index	0.0088
Coleman liau index	0.0002

The formula for Pearson’s Correlation Coefficient:

$$r_{xy} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{[\sum X^2 - \frac{(\sum X)^2}{n_x}][\sum Y^2 - \frac{(\sum Y)^2}{n_y}]}} \quad (3)$$

III. IMPLEMENT AND RESULT

We will implement the proposed methods to the Q&A pair of data. There are four kind data for the classification, first is data that acquired from the entire feature, data with high correlation (> 0.1 and > -0.1), text feature data, and readability feature data. We build the predictor using 250 training data and 100 testing data.

TABLE 3. ACCURACY OF ALL FEATURE AND CORR FEATURE

Feature	Training Set	Data Test	CV		
			5	10	15
All Feature	93.6	72.8	88.4	88.4	88.4
Corr Feature	93.2	71.2	88	88.4	88.4
Text Feature	90.4	70.4	84.4	84	83.2
Readability Feature	93.2	73.6	87.6	86.4	87.6

Table 3 shows prediction accuracy for the different implementation of answer quality, in particular comparing the choice in classifier algorithm, feature sets (using all feature, Correlation feature, text feature, and readability feature) and test option. By using C4.5 results, the best performance of the entire variant feature is all features with 93.6 of accuracy slightly adrift of 0.4 with Corr feature. We can conclude that text feature is a part of readability feature because some text feature parameter are count character, word, and sentence. For the count word parameter, in the computer and internet subject mostly users used Loan word. Another interesting result from the table 3 we could see that the differences between all features, Correlation feature and readability feature, is not too significant for accuracy it is about 0.4. This indicates that feature which does not have high correlation is not too pretty significant impact for classification results. Figure.2 shows the result classification on training data using Weka framework.

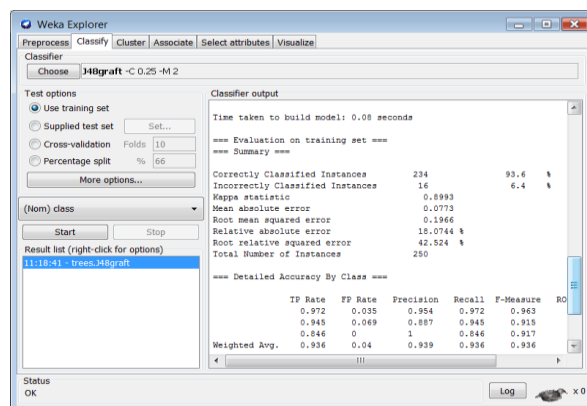


Figure 2. Result of Classification on Training Data

IV. CONCLUSION

In this paper we presented our knowledge to quantify and predict quality of answer in collaborative question answer (CQA) learning, especially for Indonesian. Beyond developing

models to select best answer and evaluate the quality of answers, there are several important lessons to learn here for measuring content quality in CQA. We find that domain knowledge information sometime isn't providing good answer. On the other side students answer better than domain knowledge answer.

With appropriate features, we could build models that could have significantly higher probability of identifying the best answer class than classifying a non-best answer. From the entire system from 9 features, we conclude as following:

1) *From the four existing feature, the highest accuracy exist on all feature set (comparing with correlation coefficient set, text feature set and readability feature set).*

2) *The best performance for all feature set by using C4.5 classifier, with averaged accuracy 93.6 for training set, 72.8 data test and 88.4 for cross validation.*

In the future our models and predictions could be useful for predictor quality information as a recommender system to complete collaborative question answer learning.

ACKNOWLEDGMENT

The authors wish to thank to the students of Electrical and Information Technology State University of Malang Indonesia, who contributed and supported to the experiments.

REFERENCES

- [1] Dillenbourg, P. (1999). Collaborative Learning: Cognitive and Computational Approaches. Advances in Learning and Instruction Series. New York, NY: Elsevier Science, Inc.
- [2] Chiu, M. M. (2000). Group problem solving processes: Social interactions and individual actions. *Journal for the Theory of Social Behavior*, 30, 1, 27-50.600-631.
- [3] Chiu, M. M. (2008). Flowing toward correct contributions during groups' mathematics problem solving: A statistical discourse analysis. *Journal of the Learning Sciences*, 17 (3), 415 – 463
- [4] Mitnik, R., Recabarren, M., Nussbaum, M., & Soto, A. (2009). Collaborative Robotic Instruction: A Graph Teaching Experience. *Computers & Education*, 53(2), 330-342.
- [5] Chiu, M. M. (2008). Effects of argumentation on group micro-creativity. *Contemporary Educational Psychology*, 33, 383 – 402.
- [6] Chen, G., & Chiu, M. M. (2008). Online discussion processes. *Computers and Education*, 50, 678 – 692
- [7] Wang, C.C., Hung J.C., Yang C.Y., Shih T.K. (2006). An Application of Question Answering System for Collaborative Learning. *IEEE Conference on ICDCSW'06*

- [8] Kohei Arai, Anik Nur Handayani, Question Answering System for an Effective Collaborative Learning. (IJACSA) Vol 3. No.1, 2012. Page 60-64.
- [9] Kohei Arai, Anik Nur Handayani, Predicting Quality of Answer in Collaborative Q/A Community. (IJARAI), Vol 2. No.3, 2013. Page 21-25.
- [10] Kohei Arai, Anik Nur Handayani, Question Answering for Collaborative Learning with Answer Quality Predictor. (IJMECS), Vol 5. No.5, 2013. Page 12-17.
- [11] Kohei Arai, Anik Nur Handayani, Collaborative Question Answering System Using Domain Knowledge and Answer Quality Predictor (IJMECS), Vol. 5 Number 11, 2013. Page 21-27
- [12] Watzlawick, P. (1967) *Pragmatics of Human Communications: A Study of Interactional Patterns. Pathologies and Paradoxes.* W.W. Norton, New York.
- [13] Smith, B. L., & MacGregor, J. T. (1992). "What Is Collaborative Learning?". National Center on Postsecondary Teaching, Learning, and Assessment at Pennsylvania State University
- [14] Chiu, M. M. (2004). Adapting teacher interventions to student needs during cooperative learning. *American Educational Research Journal*, 41, 365-399.
- [15] L. Rassbach, T. Pincock, and B. Mingus. Exploring the feasibility of automatically rating online article quality. <http://upload.wikimedia.org/wikipedia/wikimania2007/d/d3/RassbachPincockMingus07.pdf>.
- [16] E. A. Smith and R. J. Senter. Automated readability index. 1967.
- [17] http://id.wiktionary.org/wiki/Kategori:Kata_bahasa_Indonesia
- [18] Chiu, M. M. (2000). Group problem solving processes: Social interactions and individual actions. *for the Theory of Social Behavior*, 30, 1, 27-50.600-631.

AUTHORS PROFILE

Kohei Arai received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 and also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission A of ICSU/COSPAR since 2008. He wrote 26 books and published 500 journal papers.

Anik Nur Handayani received the B.E. degree in electronics engineering from Brawijaya University, and the M.S. degree in Electrical Engineering, from Institute of Technology Sepuluh Nopember, Surabaya, Indonesia, in 2004 and 2008, respectively. She is currently a PhD Student at Information Science in Saga University, Japan.