

# Visualising Arabic Sentiments and Association Rules in Financial Text

Hamed AL-Rubaiee

Department of Computer Science  
and Technology  
University of Bedfordshire  
Bedfordshire, United Kingdom

Renxi Qiu

Department of Computer Science  
and Technology  
University of Bedfordshire  
Bedfordshire, United Kingdom

Dayou Li

Department of Computer Science  
and Technology  
University of Bedfordshire  
Bedfordshire, United Kingdom

**Abstract**—Text mining methods involve various techniques, such as text categorization, summarisation, information retrieval, document clustering, topic detection, and concept extraction. In addition, because of the difficulties involved in text mining, visualisation techniques can play a paramount role in the analysis and pre-processing of textual data. This paper will present two novel frameworks for the classification and extraction of the association rules and the visualisation of financial Arabic text in order to realize both the general structure and the sentiment within an accumulated corpus. However, mining unstructured data with natural language processing (NLP) and machine learning techniques can be arduous, especially where the Arabic language is concerned, because of limited research in this area. The results show that our frameworks can readily classify Arabic tweets. Furthermore, they can handle many antecedent text association rules for the positive class and the negative class.

**Keywords**—Opinion mining; Stock market; Twitter; Saudi Arabia; Association text rules; Data mining, Text Visualization

## I. INTRODUCTION

The most recent research studies have been particularly interested in efforts to visualize Arabic texts. For example, Hammo et al. developed a visualization system for analysing and visualising Arabic text (VistA) [1]. Their work was based on Obeid et al.'s study [2], which applied latent semantic indexing (LSI) as a dimensionality reduction technique that aimed to stem out data from Arabic documents. By contrast, this work will use a different approach based on a hybrid of natural language processing (NLP) and machine learning algorithms involving two modules. The first module will categorize tweets as positive or negative in accordance with their sentiment polarity. The second one will help to create models useful for sentiment visualization, referring to either the positive or the negative category. Thus, this research will consider two main areas: sentiment analysis in the Arabic language and text association rules.

Similarly, Kotsiantis and Kanellopoulos [4] have demonstrated the relationships between the objects in a transactional database. Their incidence in the database defined their relationships. However, the main limitation of association rule discovery lies in its manageability when the numbers of transactions start to increase. That results in the emergence of different sets of classified data, in which most occurring objects are assigned with others in all possible ways, some of which are irrelevant.

The application of machine learning to text mining has resulted in a number of tools that are now widely implemented in different areas of research. The strongest aspect of text mining is that it is capable of processing successfully unstructured data such as social networking, e-learning, bioinformatics, pattern matching, and sentiment analysis. It also searches for and identifies patterns in data. Text mining works successfully with PDF files, emails, and XML [5].

With the spread and the rise in the popularity of social media, sentiment analysis has become one of the core social media research techniques [6-8]. Recent research studies have applied sentiment analysis to the extraction of users' views on different topics, from politics to management. The technique works well at identifying whether sentiments are positive or negative and how they are expressed [9].

In their research, Wong, Whitney, and Thomas [10] stated that the association rule in data mining was based on the inclusion of the form  $X \rightarrow Y$ , where  $X$  was a set of preceding items and  $Y$  was the resultant item. It is, however, quite challenging to visualize associations in large sets of data, when over dozens of rules emerge.

This paper is organized as follows: Section 2 will discuss the sentiment analysis technique and text association rules. Section 3 will describe the methodology related to the process of the sentiment analysis of Arabic tweets and the extraction of text association rules. Section 4 will analyse the experimental findings and the visualization process. The final section will constitute the conclusion and recommendations for further work in this area.

## II. RELATED WORK

### A. Sentiment Analysis in the Arabic Language

There is scarce research on Arabic sentiment analysis, and that field is still in the initial stages of development. Arabic is very different from other languages. It has a unique structure and its own rules. For example, sentences are written from right to left, there are no capital letters, and there are a number of grammatical rules [11].

Sentiment analysis or opinion mining has been successfully applied to social media. It uses a combination of NLP and text mining to classify sentiments as positive or negative. For example, Duncan's and Zhang's research [15] specifically touched on Twitter sentiment analysis, which the nature of

Twitter and tweets reinforced. For example, spelling mistakes and the use of slang are very typical for Twitter. Also, a tweet has length restrictions; overall, it should not exceed 140 characters. Thus, the use of classification in this case was very challenging. The findings show that the level of accuracy of the neural networks in one of the experiments was relatively low. In fact, sentiment analysis applied to social media is quite different from traditional text mining. The traditional text analysis technique is based on using initially pre-defined classes to form categories in a document, and, thus, it forces the data into already existing themes [13].

The objective of sentiment analysis is entirely different. It seeks to develop new categories with regard to participants' opinions and views. The strength of sentiment analysis lies in its capacity to measure scores of sentiments by comparing that with a dictionary. Despite the uniqueness of the method, sentiment analysis has focused mainly on English text. Using the same sentiment lexicons on any other language would result in adaptation errors [12].

Machine learning (ML), which is also known as a corpus-based technique, is a supervised method in which data sets are labelled positive or negative and represented in feature vectors. These vectors, in turn, are used as training data to identify and categorize specific features in a certain class [14].

### B. Text Association Rules

Chen et al. have described an association rule as an implicative insinuation of the form  $A \Rightarrow B$ , where A and B are frequent item sets in a transaction database and  $A \cap B = \emptyset$ . In practical usage, the rule is  $A \Rightarrow B$  [16].

Text mining is withal defined as text data mining or erudition revelation from textual databases. Sodality rules are engendered by analysing data for frequent if/then patterns and utilizing the criteria support and confidence to identify the most consequential relationships. Support is a denouement of how frequently the items appear in the database. Confidence indicates the number of times the if/then representations have been found to be true. An integrated framework called associative classification was proposed that purposed to discover a set of rules that satisfied user-specified minimum support and minimum confidence as a classifier. This was done by fixating on a special subset of association rules, whose right-hand-side was restricted to the classification class attribute. The frequent if/then patterns were mined utilizing methods such as the Apriori algorithm, the Classification Based on Associations (CBA) algorithm, and the FP-Growth algorithm [8, 9].

Lopes et al. [5] expounded the quandary of mining association rules from text. They commenced by representing the text as bag of words: Let  $I = \{i_1, i_2, \dots, i_m\}$ . and Let D are a set of transactions, where each transaction T is a set of items that represent the document so  $T \subseteq I$ . An association rule is an involvement of the form  $X \Rightarrow Y$  where  $X \subset I$  and  $Y \subset I$ , and  $X \cap Y = \emptyset$ . The rule  $X \Rightarrow Y$  holds confidence if the document D contains X and Y, and support if the document contains  $X \cup Y$ . The left of the rule is the head of the rule and the set of residual words is the rule body.

Tan, Kumar, and Srivastava [7] described several key properties that should be considered to quantify the correlation between data attributes. For instance, they described the sensitivity of quantification to the row and the column scaling operation. They reported that metrics such as support, confidence, lift, correlation, and collective strength caused conflict regarding the interestingness of the pattern, and the correct metric to be used was seldom recognized.

Association rule mining can be divided into two phases. In the first phase, frequent patterns are mined with regard to the threshold minimum support. In the second phase, association rules are created with regard to the confidence threshold and minimum confidence [11].

## III. METHODOLOGY

The major target of this paper is the extraction of the association rule and the visualization of positive and negative sentiments in financial Arabic text. In general, our frameworks will commence with the following:

- The Arabic sentiment analysis model,
- Pre-processing of the Arabic text,
- Classification as positive and negative sentiments,
- Model evaluation.

After the text classification, the second framework will proceed as follows:

- The creation of the Arabic text association rule model,
- The pre-processing of the Arabic text,
- Finding the frequent item set,
- Creating and visualizing the Arabic association rules for each class.

Figure 1 summarizes the first model, which was the process of opinion mining Arabic tweets. In trading strategies on the Saudi stock market, Twitter was chosen as a platform for opinion mining to illustrate the association text rules involved in Modern Standard Arabic (MSA).

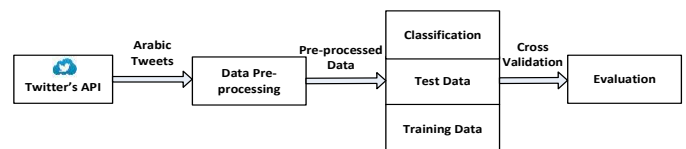


Fig. 1. The Process of Opinion Mining Arabic tweets

### A. Data collection

The tweets were obtained from Mubasher firm's Twitter Account. Mubasher is high-ranking stock analysis software in the Kingdom of Saudi Arabia (KSA) and the Middle East. The tweets were gathered over a one-month period from March 1, 2016, to April 1, 2016. The data set includes 2,590 tweets, which cover most of the quota sectors of the Saudi stock market. A selection of over 100 terms and expressions in Arabic from the emotion corpus (for instance, increase, magnification, decline, fall, elevate, cash dividends, distribution of bonus shares, not to distribute) was then divided

Identify applicable sponsor/s here. If no sponsors, delete this text box (sponsors).

between two classes: positive and negative. The most prevalent MSA words (including the following) fell into these the two classes:

(انخفاض Growth), (تراجع Decline), (ارتفاع Rise), (انخفاض fall), (ازدهار Prosperity), (تقلص Shrink), (ربح Profit), (ارتفاع خسائر Loss), (توزيع ارباح Dividend), and (ارتفاع خسائر High losses) .

To accumulate Arabic tweets in the corpus of data, a desktop application was developed utilizing C# and Twitter's official developers' API. The tweets utilized in the study did not involve hashtags, links, or special characters. Tweets that were duplicates or retweets were eliminated.

Three Mubasher workers who have experience with Saudi Stock Shares annotated the data manually. Negative tweets were given the label '-1', while positive tweets were given the label '1'. Neutral tweets were ignored, and the impertinent tweets were expunged from the data set.

Table 1 illustrates the number of tweets in the data set: In total, 2,590 Arabic tweets were accumulated, 934 marked tweets were utilized for the training dataset, and 1,656 extraneous tweets were erased from the data set.

TABLE I. NUMBER OF TWEETS

Positive	Negative	Total
467	467	934

### B. Data Pre-processing

Social media channels commonly contain words with unclear meanings; opinion mining predicated on social media is still under development. This is categorically true in situations with spelling mistakes, the utilization of emoticons and other characters that express special denouement, or the utilization of English pronunciation in association with Arabic characters. Modern Standard Arabic (MSA) will be used to gratify validation requisites for this study. These kinds of tweets consist of independent, semantic-oriented Arabic lexica, which confound the research even further. To address the challenge, the ontology of an incipient keywords process model will be established to ameliorate the text mining.

Data pre-processing includes cleaning and acclimating text for classification. The pre-processing stage has several steps, for instance, online text cleaning, white space abstraction, abbreviation expansion, stemming, stop-word abstraction, negation handling, and feature selection. The final step is called filtering, while the rest are called transformations [17].

After the data labelling was completed, Rapidminer<sup>1</sup> was used to replace some Arabic letters that had different shapes. For example, (أ-آ-إ-أ-أ-أ-ؤ-ة-ي) were replaced with (-أ-ه-و-ي) to remove the diacritical marks. The five pre-

processing steps below were then performed using Rapidminer:

- 1) Tokenization: divided each tweet into multiple token-based whitespace characters;
- 2) Stop-word removal process: removed the Arabic stop words;
- 3) Light stemming: removed the suffixes and prefixes from each token;
- 4) Filtering token by length: abstracted worthless terms and was set to three;
- 5) Setting N-gram to two: N-gram was a series of n tokens from a given text [18].

Then SVM was applied with the weighting scheme TF-IDF (Term Frequency-Inverse Document Frequency) to build a classification model that could classify tweets into positive and negative classes according to their sentiment polarity. Determinately, the evaluation was carried out using the accuracy, precision, and recall methods.

### C. Classification method

SVM's rudimentary conception involves finding a hyper-plane, which vector  $\vec{w}$  represents, that disunites the document vectors in one class from those in other documents. SVM has been applied successfully in many opinion mining tasks. It has outperformed other machine learning techniques due to the associated advantage. For instance, the powerful in high-dimensional spaces [19].

### D. Evaluation:

The widely known performance metrics that were utilized to evaluate the classification results were precision, recall, and accuracy [19, 20].

- Higher precision meant fewer false positives.  
- Precision =  $tp/(tp+fp)$
- Higher recall meant fewer false negatives.  
- Recall =  $tp/(tp+fn)$
- Accuracy involved calculating the ratio of true results (positives and negatives)  
- Accuracy =  $(tp+tn)/(tp+fp+fn+tn)$

Figure 2 illustrates the second model, which was the process of engendering association rules in Arabic tweets. After the implementation of the previous classification model, a text association rules framework was employed to differentiate between the text rules for each class (positive, negative).

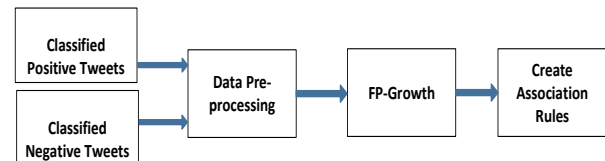


Fig. 2. The process of creating association rules for Arabic tweets

<sup>1</sup> <https://rapidminer.com>

E. Data Pre-processing

The same corpus classified as either positive or negative was used in this stage; and the same data pre-processing procedure was carried out separately for the positive class and the negative class.

F. Frequent-Pattern Method

An important algorithm in the data-mining field is a Frequent-Pattern tree algorithm. FP-growth is an approach that does not require candidate generation. It stores relevant item set information and allows an efficient novel structure to discover the frequent item sets. FP-growth has a way of decomposing the mining process into small tasks on a conditional FP-tree. First, it looks into the data set to find the frequent items at level-1 by computing the support for frequent items. Those frequent 1-item sets are stored in descending order of their supports. In the next step, the data set is scanned again to build an FP-tree using the head table with a null label root. The database scanning process continues for each transaction T to re-sort the frequent items in the header table according to the frequency of their occurrences and insert them in the FP-tree [21].

G. Create Association Rules

Association rules are if/then statements that help to expose relationships between seemingly unrelated data. An association rule has two components, an antecedent (if) and a consequent (then). An antecedent is an item set found in the data set, and a consequent is an item set found in incorporation with the antecedent [16].

IV. THE EXPERIMENT RESULTS

A. Experiment

SVM and weighting schemes (TF-IDF) were performed to explore the polarity of a given text, and to generate the word vectors. Table 2 shows the precision and recall for the SVM classifier.

TABLE II. SVM PRECISION AND RECALL

Classifier	Accuracy	Recall	Precision
SVM	82.31	82.31	82.75

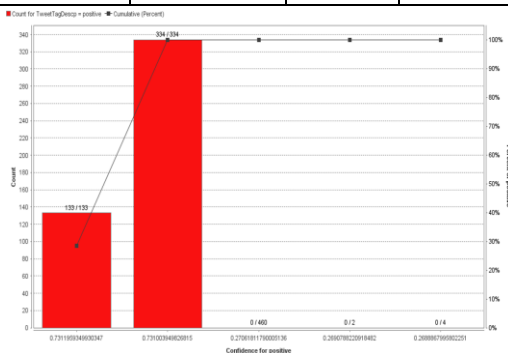


Fig. 3. SVM lift chart for the positive class

The lift chart is a way of evaluating the performance of data mining model and the predictive accuracy of one model against another [22]. The lift chart is a discrete version of representing

and visualizing the classifier performance. The highest confidence numbers are shown first. As is evident, the confidence numbers decrease at some point. For example, Figure 3 and Figure 4 show the lift charts for the positive and negative classes, respectively, for an SVM classifier.

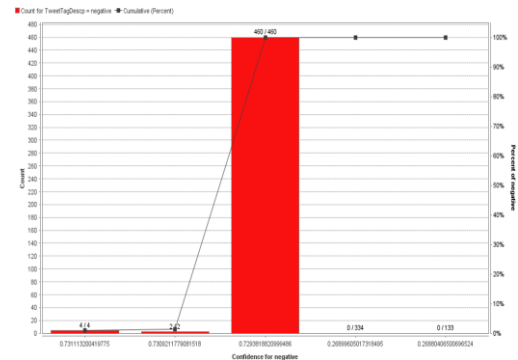


Fig. 4. SVM lift chart for the negative class

Eventually, the best precision that SVM achieved was 82.31%. On the other hand, there was virtually a 20% misclassification in our corpus. Table 3 shows the misclassification that occurred during the experiment.

TABLE III. SVM MISCLASSIFICATION

Accuracy: 82.31% +/- 2.93 % (mikro: 82.32%)			
	True negative	True positive	Class precision
Pred.negative	403	102	79.80%
Pred.positive	63	365	85.28%
Class recall	86.48%	78.16%	

This study aimed to continue to extract and visualize the association rules for each class because the authors believe that correlation between terms in our corpus enhanced the understanding of the text structure and clarified the sentiments expressed. This may also have improved the accuracy of our classification model.

After the separate implementation of the aforementioned association rules model for each class, the default values for most of the parameters were utilized to provide frequent items and to produce association rules that were more accurate for the positive corpus. For example, minimum support=0.01, minimum confidence=0.8, max items=2.

Experts in data mining argue that some terms or words occur with higher frequency in the dataset, while others rarely appear. In this case, the values of the minimum support will control the rule discovery. If the minimum support is set at a high value, rules that infrequently occur will not be found. Otherwise, if the minimum support is set at a low value, rules that frequently occur will be engendered. This will cause a problem called “rare item”; as a result of good rules with high confidence, it may be ignored simply because good rules have very little support [15, 23].

The frequent items produced for the positive class were 1,223. For example, Table 4 shows the frequent term that had the highest support value in the positive class with an item size equal to 1.

TABLE IV. THE FREQUENT POSITIVE TERM SIZE EQUAL TO 1

Support	Arabic Term	Term in English
0.152	الاول	First
0.131	بالربع	Quarter
0.114	تراجع	Decline
0.105	شركة	Company
0.105	تعلن	Announces

Table 5 shows some of the frequent terms or item sets that had the highest support value in the negative class with an item size equal to 2.

TABLE V. THE FREQUENT POSITIVE TERM SIZE EQUAL TO 2

Support	Arabic Term1	Term1 in English	Arabic Term2	Term2 in English
0.178	ارباح	Earnings	توزيع	Sharing out
0.155	ارباح	Earnings	تعلن	Announces
0.150	ارباح	Earnings	شركة	Company
0.124	ارباح	Earnings	زيادة	Increment
0.122	ارباح	Earnings	عموميه	generality

As is evident in Figure 5, the term “earnings” (ارباح) correlated with the other terms that appeared in the premises column with the highest support values.

No.	Premises	Conclusion	Support	Confidence
45	وتوزيع	ارباح	0.011	0.833
46	شركة استت	ارباح	0.011	0.833
208	المساهمين	ارباح	0.015	0.875
287	توزيع ارباح	ارباح	0.064	1
288	ارتفاع ارباح	ارباح	0.028	1
289	ارباح المنصف	ارباح	0.015	1
290	زيادة المساهمين	ارباح	0.015	1
291	وتوزيع ارباح	ارباح	0.011	1
292	القائي العام	ارباح	0.011	1
293	ارباح استت	ارباح	0.011	1

Fig. 5. The term “earnings” (ارباح) and its correlations

Figure 6 shows the association between terms that were related to the positive sentiment “earnings” (ارباح). For example, the most important rules for the term “earnings” (ارباح) were [ارباح] --> [وتوزيع] (support: 0.011 confidence: 0.833), the term’s meaning entailed sharing out the profits of some company in the Saudi stock market. So the term (sharing out) was arranged together with the term “earnings” to compose positive phrases and sentences such as the following:

البنك السعودي الهولندي يوافق على زيادة رأس المال وتوزيع أرباح

Saudi Hollande Bank approves of a capital increase and dividends.

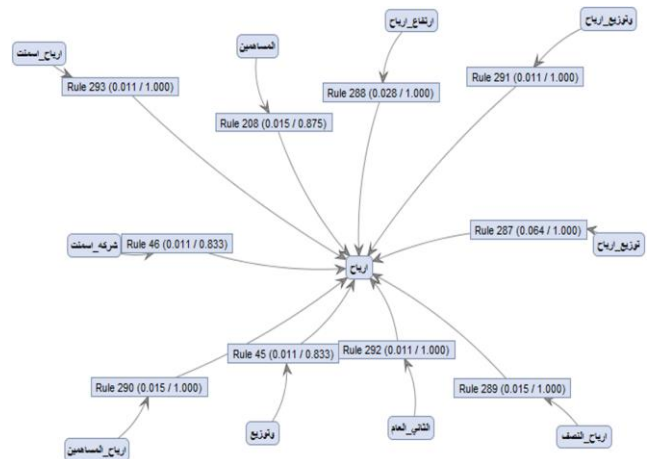


Fig. 6. Visualize the association rules term “earnings” (ارباح)

Finally, the aforementioned steps were followed to visualize the association rules for the negative class in our corpus.

The frequent items produced for the negative class were 534. For example, Table 6 shows the frequent terms that had the highest support values with an item size equal to 1.

TABLE VI. THE FREQUENT NEGATIVE TERM SIZE EQUAL TO 1

Support	Arabic Term	Term in English
0.178	ارباح	Earnings
0.155	توزيع	Sharing out
0.150	تعلن	Announces
0.124	شركة	Company
0.122	زيادة	Increment

Table 7 shows some of the frequent terms or item sets that had the highest support value with an item size equal to 2.

TABLE VII. THE FREQUENT NEGATIVE TERMS SIZE EQUAL TO 2

Support	Arabic Term1	Term1 in English	Arabic Term2	Term2 in English
0.129	الاول	First	بالربع	Quarter
0.129	الاول	First	بالربع الاول	First-Quarter
0.058	الاول	First	تراجع	Decline
0.015	الاول	First	تعلن	Announces
0.062	الاول	First	ارباح	Earnings

As is evident in Figure 7, the term “first” (الاول) correlated with the other terms that appeared in the premises column with the highest support values.

No.	Premises	Conclusion	Support	Confidence
25	تراجع_ربح	الأول	0.013	0.857
100	ربح	الأول	0.129	0.984
102	ربح_الأول	الأول	0.129	1
103	تراجع_ربح	الأول	0.049	1
104	ربح_الأول	الأول	0.017	1
105	ربح	الأول	0.017	1
106	ارتفاع_خسائر	الأول	0.015	1
107	الأول_عدم	الأول	0.013	1

Fig. 7. The correlations of the term “earnings” (ارباح)

Figure 8 shows the association between terms that were related to negative sentiments, for instance, “decline” (تراجع) and “higher losses” (ارتفاع خسائر). For example, the most important rule for the term “first” (الأول) was [تراجع] --> [الأول] (support: 0.013 confidence: 0.857). The term’s meaning was dividends are decline for some company in the Saudi stock market, so the term “decline” (تراجع) was arranged together with the term “first” (الأول) to compose negative phrases and sentences such as the following:

شركه ميردنتتراجع 71 % بالربع الأول من هذا العام

(Mubarrad Company down 71% for the first quarter of this year)

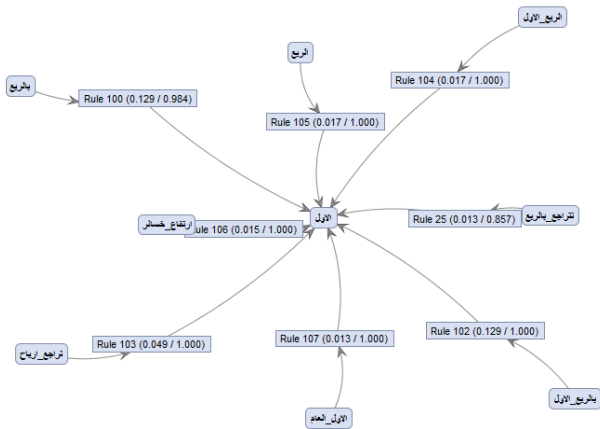


Fig. 8. Visualization of the association rules term “first” (الأول)

The experiments above were conducted on a Modern Standard Arabic (MSA) corpus and could be summarised as having used two important measures (support and confidence) to extract and visualize association rules that could expose the sentiments behind Arabic financial texts.

## V. CONCLUSION AND FUTURE WORKS

In the present study, the authors designed and implemented Arabic text classifications regarding Saudi stock market opinions through the SVM algorithm and the extraction of the association rules presented. Moreover, they visualised financial Arabic text to understand the sentences’ structure and the sentiments behind them. The results of the study show that text pre-processing is an essential factor in opinion mining

classification and in the extraction and visualisation of the association rules for Arabic text. Moreover, visualisation can help to sort out the misclassification that is possible with the Arabic language because of the size and ratio of the vocabulary and because of how it is characterised. In addition, as humans were involved in labelling the data, it is possible that human error occurred; for this reason, the visualisation of the text shows the importance of the correlation between terms that involved in the textual structured contents. The current study should be repeated to compare and address other metrics, for instance, lift, correlation, and collective strength. These metrics are typically used to extract and search for interesting association patterns from textual data. Several key properties should be considered in the examination of the correlations between textual data attributes in order to select the right measures for an Arabic financial domain.

## ACKNOWLEDGMENT

This work was sponsored by Mubasher Company. We would like to thank the Managing Director of Mubasher KSA, Dr Walid Rasheed Al-Ballah, for his supplementary comments and support.

## REFERENCES

- [1] B. Hammo, N. Obeid, and I. Huzayyen, "ViStA: a visualization system for exploring Arabic text," International Journal of Speech Technology, vol. 19, pp. 237-247, 2016.
- [2] Obeid, I. Huzayyen, and B. Hammo, "Experimenting with Arabic text visualizing," in Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on, 2013, pp. 1-6.
- [3] R. Elmasri and S. Navathe, "Fundamentals of Database Systems, chapter Appendix C: An Overview of the Network Data Model," ed: Addison-Wesley, 2000.
- [4] S. Kotsiantis and D. Kanellopoulos, "Association rules mining: A recent overview," GESTS International Transactions on Computer Science and Engineering, vol. 32, pp. 71-82, 2006.
- [5] F. Gharehchopogh and Z. Khalilfeli, "Analysis and evaluation of unstructured data: text mining versus natural language processing," in Application of Information and Communication Technologies (AICT), 2011 5th International Conference on, 2011, pp. 1-4.
- [6] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in LREC, 2010, pp. 1320-1326.
- [7] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in Proceedings of the Workshop on Languages in Social Media, 2011, pp. 30-38.
- [8] Z.-H. Deng, K.-H. Luo, and H.-L. Yu, "A study of supervised term weighting scheme for sentiment analysis," Expert Systems with Applications, vol. 41, pp. 3506-3513, 2014.
- [9] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in Proceedings of the 2nd international conference on Knowledge capture, 2003, pp. 70-77.
- [10] P. C. Wong, P. Whitney, and J. Thomas, "Visualizing association rules for text mining," in Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on, 1999, pp. 120-123, 152.
- [11] K. Ahmad and Y. Almas, "Visualising sentiments in financial texts?," in Ninth International Conference on Information Visualisation (IV'05), 2005, pp. 363-368.
- [12] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," IEEE Intelligent Systems, pp. 15-21, 2013.
- [13] G. Fu and X. Wang, "Chinese sentence-level sentiment classification based on fuzzy sets," in Proceedings of the 23rd International Conference on Computational Linguistics: Posters, 2010, pp. 312-319.

- [14] A. Shoukry and A. Rafea, "Sentence-level Arabic sentiment analysis," in Collaboration Technologies and Systems (CTS), 2012 International Conference on, 2012, pp. 546-550.
- [15] B. Duncan and Y. Zhang, "Neural networks for sentiment analysis on Twitter," in Cognitive Informatics & Cognitive Computing (ICCI\* CC), 2015 IEEE 14th International Conference on, 2015, pp. 275-278.
- [16] M.-S. Chen, J. Han, and P. S. Yu, "Data mining: an overview from a database perspective," IEEE Transactions on Knowledge and data Engineering, vol. 8, pp. 866-883, 1996.
- [17] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," Procedia Computer Science, vol. 17, pp. 26-32, 2013.
- [18] G. T. a. N. S., "Feature Selection and Classification Approach for Sentiment Analysis," Machine Learning and Applications: An International Journal (MLAIJ), vol. 2, 2015.
- [19] M. Rushdi - Saleh, M. T. Martín - Valdivia, L. A. Ureña - López, and J. M. Perea - Ortega, "OCA: Opinion corpus for Arabic," Journal of the American Society for Information Science and Technology, vol. 62, pp. 2045-2054, 2011.
- [20] F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme," Decision Support Systems, vol. 57, pp. 245-257, 2014.
- [21] Z. Qiankun and S. S. Bhowmick, "Association Rule Mining: A Survey," Technical Repo~, CAIS, Nanyang Technological University, Singapore, p. 1, 2003.
- [22] T. Jaffery and S. Liu, "Measuring Campaign Performance by Using Cumulative Gain and Lift Chart," in SAS Global Forum, 2009.
- [23] B. Liu, Y. Ma, and C. K. Wong, "Improving an association rule based classifier," in European Conference on Principles of Data Mining and Knowledge Discovery, 2000, pp. 504-509.
- [24] N. Obeid, I. Huzayyen, and B. Hammo, "Experimenting with Arabic text visualizing," in Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on, 2013, pp. 1-6.
- [25] R. Elmasri and S. Navathe, "Fundamentals of Database Systems, chapter Appendix C: An Overview of the Network Data Model," ed: Addison-Wesley, 2000.
- [26] S. Kotsiantis and D. Kanellopoulos, "Association rules mining: A recent overview," GESTS International Transactions on Computer Science and Engineering, vol. 32, pp. 71-82, 2006.
- [27] F. Gharehchopogh and Z. Khalifelu, "Analysis and evaluation of unstructured data: text mining versus natural language processing," in Application of Information and Communication Technologies (AICT), 2011 5th International Conference on, 2011, pp. 1-4.
- [28] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in LREC, 2010, pp. 1320-1326.
- [29] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in Proceedings of the Workshop on Languages in Social Media, 2011, pp. 30-38.
- [30] Z.-H. Deng, K.-H. Luo, and H.-L. Yu, "A study of supervised term weighting scheme for sentiment analysis," Expert Systems with Applications, vol. 41, pp. 3506-3513, 2014.
- [31] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in Proceedings of the 2nd international conference on Knowledge capture, 2003, pp. 70-77.
- [32] P. C. Wong, P. Whitney, and J. Thomas, "Visualizing association rules for text mining," in Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on, 1999, pp. 120-123, 152.
- [33] K. Ahmad and Y. Almas, "Visualising sentiments in financial texts?," in Ninth International Conference on Information Visualisation (IV'05), 2005, pp. 363-368.
- [34] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," IEEE Intelligent Systems, pp. 15-21, 2013.
- [35] G. Fu and X. Wang, "Chinese sentence-level sentiment classification based on fuzzy sets," in Proceedings of the 23rd International Conference on Computational Linguistics: Posters, 2010, pp. 312-319.
- [36] A. Shoukry and A. Rafea, "Sentence-level Arabic sentiment analysis," in Collaboration Technologies and Systems (CTS), 2012 International Conference on, 2012, pp. 546-550.
- [37] B. Duncan and Y. Zhang, "Neural networks for sentiment analysis on Twitter," in Cognitive Informatics & Cognitive Computing (ICCI\* CC), 2015 IEEE 14th International Conference on, 2015, pp. 275-278.
- [38] M.-S. Chen, J. Han, and P. S. Yu, "Data mining: an overview from a database perspective," IEEE Transactions on Knowledge and data Engineering, vol. 8, pp. 866-883, 1996.
- [39] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," Procedia Computer Science, vol. 17, pp. 26-32, 2013.
- [40] G. T. a. N. S., "Feature Selection and Classification Approach for Sentiment Analysis," Machine Learning and Applications: An International Journal (MLAIJ), vol. 2, 2015.
- [41] M. Rushdi - Saleh, M. T. Martín - Valdivia, L. A. Ureña - López, and J. M. Perea - Ortega, "OCA: Opinion corpus for Arabic," Journal of the American Society for Information Science and Technology, vol. 62, pp. 2045-2054, 2011.
- [42] F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme," Decision Support Systems, vol. 57, pp. 245-257, 2014.
- [43] Z. Qiankun and S. S. Bhowmick, "Association Rule Mining: A Survey," Technical Repo~, CAIS, Nanyang Technological University, Singapore, p. 1, 2003.
- [44] T. Jaffery and S. Liu, "Measuring Campaign Performance by Using Cumulative Gain and Lift Chart," in SAS Global Forum, 2009.
- [45] B. Liu, Y. Ma, and C. K. Wong, "Improving an association rule based classifier," in European Conference on Principles of Data Mining and Knowledge Discovery, 2000, pp. 504-509.