From Public Polls to Tweets: Developing an Algorithm for Classifying Sentiment from Twitter Based on Computing with Words

Saad M. Darwish*, Magda M. Madbouly, Mohamed A. Hassan
Department of Information Technology, Institute of Graduate Studies and Research, Alexandria University, 163 Horreya Avenue, El Shatby 21526. P.O. Box 832, Alexandria, Egypt.

* Corresponding author. Tel.: +201222632369; email: saad.darwish@alex-igsr.edu.eg
Manuscript submitted May 25, 2015; accepted August 12, 2015.
doi: 10.17706/jcp.11.3.238-246

Abstract: Uncertainty is an intrinsic part of sentiment analysis, especially when dealing with social media (Twitter data) that known as noisy texts. Although there are many researches have been done in sentiment analysis, accuracy in identifying sentiment from Twitter data is still far from satisfactory. In this paper, we address this limitation by investigating computing with words (CWW) and granule computing (GC) approaches that all contribute towards improving sentiment classification on Twitter. CWW can provide a solid basis for the computational theory of perceptions under the environments of imprecision, uncertainty, and partial truth. CWW technique is employed to translate propositions expressed in a natural language into what is called generalized constraint language (GCL) with possibilistic type and applying rules of fuzzy constraint propagation. GC is engaged to infer an answer to a query expressed in a natural language "which granule does the proposition belongs to?" The experimental results show that it is feasible to use CWW to classify sentiment.

Key words: Computing with words, opinion mining, sentiment analysis, NLP.

1. Introduction

Social network sites such as Facebook, Twitter and other microblogging services provide an opportunity for public to give opinions about some issues of interest. Twitter is an ideal platform for users to spread not only information in general but also political opinions, whereas Facebook provides the capability for direct dialogs [1]. Sentiment analysis helps the people to evaluate opinions, emotions, attitude and behavior of others, which is used to make decisions based on the user preference [2]. Opinion Mining (sentiment analysis) is a new and exciting field of research concerned with extracting opinion related information from textual data sources. It has the potential for a number of interesting applications both in commerce and academic areas, and poses novel intellectual challenges, which continues to attract considerable research interest [3]. Opinion mining is a topic in text mining, natural language processing (NLP), and Web mining discipline.

Opinions are fuzzy in nature and dealing with the semantic part of the expressed sentiments possesses many challenges and require effective techniques to properly extract and summarize people's views [4]. The first challenges is that opinion word is considered to be positive in one condition may be considered negative in another condition. A second challenge is that people do not always state the opinions in a same way [5].

Further challenges includes dealing with negation expressions; produce a summary of opinions based on product features, complexity of the sentence/document, handling of hidden product features, identify sarcasm (the use of irony to mock or convey contempt), and spelling and grammatical mistakes.

Given an object and a collection of reviews on it, the tasks in the opinion mining process usually consists of the following: (1) Identify and extract object features that have been commented on in each review. (2) Determine the orientation and strength whether the opinion is positive, negative, or neutral. (3) Provide summary of opinion in textual or in a visualization way [4]. In the literature, opinions can be classified at different levels: document, sentence (phrase) and feature [4], [5]. Analyzing sentiment at document (classifying an opinionated document as expressing a positive, negative, or neutral opinion) may not give accurate results as for example when using the term-frequency/presence method does not reflect the relationship between a feature and its related opinions; it only analyzes presence of used opinion terms in the document. The sentence level opinion mining is associated with two tasks. First one is to identify whether the given sentence is subjective (opinionated) or objective. The second one is to find opinion of an opinionated sentence as positive, negative or neutral. The assumption is taken at the sentence level is that a sentence contains only one opinion. The task of opinion mining at feature level is to extract the features of the commented object and after that determine the opinion of the object, i.e. positive or negative and then group the feature synonyms and produce the summary report.

Opinion detection at sentence level can produce better results on opinion extraction and orientation rather than opinion detection at document level. Feature-based approach makes more sense as usually people may like some features and dislike some others. However, some people give their opinions on a few features that are of their interest and hence don't comment on other features. In addition, existing approaches to automatic identification and extraction of opinions from text can be grouped into three main categories [4]: (1) keyword spotting, in which text is classified into categories based on the presence of fairly unambiguous affect words, (2) lexical affinity, which assigns arbitrary words as probabilistic affinity for a particular opinion, and (3) statistical methods, which consist in calculating the valence of keywords, punctuation and word co-occurrence frequencies based on a large training corpus.

Existing literature on opinion mining (OM) presents a varied range of techniques, tools and methods to achieve the required objectives. These techniques can be grouped as NLP, machine learning, rule-based techniques, semantic using semantic web techniques and resources like ontology, and statistical methods using techniques like K-nearest. Although these methods showed some good performance; however, results are far from satisfactory. This is because the opinion text does not clearly show or indicate which polarity classes they belong to. Sometimes this might be due to the fact that subjective text is very vague and it is very difficult to make a clear boundary between positive and negative sentiments. This shows that we need more effective tools and techniques in addressing and better understanding such unclear (fuzzy) texts. Recent work somewhat related is found in [4], [5].

Fuzzy logic is an intelligent control technique which relies on human-like expert knowledge using IF-THEN reasoning rules. Such rules are based on sets that have flexible membership functions rather than just the normal crisp binary logic [3]. Fuzzy approach will drastically enhance and improve the extraction, summarization and presentation of opinions with their weight-ages and strengths; this will help to increase the accuracy of classifications [4]. Research on Fuzzy-based opinion systems is still in its infancy. There are few researches in this field. Many of these fuzzy-based systems have either not addressed many essential challenges of opinion mining or have not utilized the powerfulness of Fuzzy logic technique. For example, most fuzzy-based systems used only adjectives and adverbs as opinion words. Nouns and verbs can also express opinions. Furthermore, fuzzy-based systems do not address properly problems like negations, vague words and ambiguous words.

Recently, many papers have highlighted the simplicity and effectiveness of computing with words (CWW) in the implementation of target selection and pattern recognition [6-8]. CWW is a methodology in which the objects of computation are words and propositions drawn from a natural language. A basic assumption in CWW is that information is conveyed by constraining the values of variables. Furthermore, information is assumed to consist of a collection of propositions expressed in natural or synthetic language. Typically, such propositions play the role of linguistic characterization of perceptions. Explicitation and constraint propagation play pivotal roles in CWW throw translation of propositions expressed in a natural language into what might be called the generalized constraint language (GCL) and applying rules of constraint propagation to expressions in this language [6].

1.1. Motivation and Contribution

Albeit, sentiment analysis is highly utilized in different applications, it is still an unsolved problem mainly due to difficulties coming with that. In many cases, opinions are hidden in Twitter data that has common characteristics such as very short text length, spelling variation, special words, topic variation, and variety in writing style. It is very difficult for a human reader to extract pertinent sentences and organize them into usable forms. An automated opinion mining system is thus needed. Calculating opinion strength more accurately is an important area that needs more research. Traditional methods provide scores for positive and negative dimension of a subjective word. However, how to combine these values to arrive to more realistic value reflecting the strength of an opinion word is still needs to be improved.

Facing this challenge, we propose an innovative method to overcome the flaws of fuzzy-based sentiment analysis. The novelty of the proposed semantic orientation (SO) identification for English text is that it is done based on CWW concept, which provides a foundation for a computational theory of perceptions [6]. Explanation of the tolerance for imprecision is an issue of central importance in CWW. Thus the analyst is able to provide a much more expressive language for knowledge representation. To the best of our knowledge (based on Google scholar), this is the first work that brings the concept of CWW in the opinion mining domain. Experiments reveal that the improved the hybridization of CWW, sentiment analysis not only could decline the training subset size significantly but also, the accuracy of classification is enhanced.

This paper is organized as follows. In the following section, a brief overview on previous work is introduced. The proposed computing with words sentiment analysis (CWW-SA) is described in Section 3. The results are discussed in Section 4. The conclusion follows in Section 5.

2. Literature Review

Sentiment analysis has attracted a lot of interest in the recent past. An increasing number of scientists are tackling the task of automated sentiment detection and classification. This has resulted in a computing domain called effective computing [1]. Many techniques are used to identify the SO of a given text like lexicons (e.g. WordNet and SentiWordNet), and statistical techniques, which looks at occurrence of a word compared to other words with known polarity [4]. Other techniques use training documents, labeled or unlabeled as a source of knowledge. However, these opinion mining approaches don't allow classifying reviews granularity in order to determine the strength of each opinion. There is a need to increase the classifications of opinions and assign weightages for different opinion words. Fuzzy logic can add such a dimension to properly analyze opinions and classify them at different strengths.

In the past few years, researchers have done lots of work on fuzzy based sentiment analysis and they work well on their own datasets. Guohong *et al.* [9] presented a fuzzy set theory based on framework for Chinese sentence-level sentiment classification. The fuzzy set theory provides a straightforward way to model the intrinsic fuzziness between sentiment polarity classes. They firstly recommend a strategy to estimate sentence sentiment intensity. Then, they defined three fuzzy sets to signify the particular sentiment polarity, namely +ve, -ve and neutral sentiments. Based on sentiment intensities, they auxiliary built a membership functions to indicate the scale of an opinionated sentence in different fuzzy sets. Finally, they verified sentence level polarity using maximum membership principle. This method achieved better results when compared with other Chinese opinion mining systems. However, the achieved precision, recall and F-score are not high and are within mid-range scores. The author did not show how to address many problems like negations, vague and ambiguous words, etc. In addition, the system could not properly aggregate multiple granularity polarity within opinionated text and hence more tailored techniques need to be developed.

Animesh *et al.* [10] proposed a supervised opinion mining systems called Fuzzy Opinion Miner (FOM). FOM executes the following tasks: extract product features on which customers has commented, identify opinion sentences in each review and extract opinion phrases, and measure the strength of opinion phrases and summarize the results. This system has few drawbacks. FOM does not focus on all features mentioned in the review. Moreover, FOM focuses on adjectives and adverbs as opinion words. Verbs can also be opinions. Also FOM does not use full Fuzzy features like fuzzy sets, rules and defuzzification process. It only uses Fuzzy weights which are assigned to opinion words. In addition to the above, the system was not compared to other system to show its performance and advantages. Moreover, precision, recall and F-score measures are not calculated to present system performance.

Samaneh *et al.* [11], proposed a fuzzy logic system that (FLS) performs sentiment classification of customer reviews. Here customer reviews were classified into various subclasses using holistic lexicon approach. FLS used adjectives, adverbs, verbs and Nouns as opinion words. Special degree for each opinion words were assigned by human experts. Based on these fuzzy sets, fuzzy rules were designed to address each case and accordingly find the orientation when a condition is met. The output is computed by using the Mamdani's defuzification function (center of gravity). The authors do not report any results. Precision, recall and F-score were not calculated to see the performance of their system.

Pranali *et al*. [2] used a rule based and fuzzy logic for the categorization of subjective and objective sentences. From subjective sentences the opinion expressions are extracted and their semantic scores are checked using the SentiWordNet directory. The final score of each individual sentence is calculated after considering the whole sentence structure, and contextual information. The system computes the intensity of each word using a fuzzy intensity finder algorithm, which calculates the weight of the extracted opinionated phrases as the weights of individual words in the phrases. One drawback of this algorithm and an NLP based approach in general is that it would likely perform very poorly when used on grammatically incorrect text. As much of the sentiment content available on the Internet could fall into that category, methods to find out and possibly correct bad English would be necessary before use on a larger scale.

Shweta Rana [3] suggested a method that is based on combinations of opinion words around each product feature in a review sentence. This methodology extracts the sentiment from Hindi text and determines the strength of opinion orientation on the product feature using fuzzy logic technique. Another related work in [12] where the authors described a methodology that shows the feasibility of type-2 fuzzy logic representation of Turkish emotion-related words. Their system has the ability to capture emotions word uncertainty since the emotion is described by its membership function with footprint of uncertainty (FOU).

This work address the limitation of existing fuzzy based sentiment analysis system when dealing with vague and ambiguous words by investigating CWW approach that contribute towards reducing classification errors. A key aspect of CWW is that it involves a fusion of natural languages and computation with fuzzy variables to deal with a proposition described in natural language without the need to numerical annotated lexicons (e.g. SentiWordNet). The proposed system uses both of adjectives, adverbs, nouns, and verbs as opinion words as opposed to existing systems that uses only adjectives and adverbs as opinion words; this may result in strong sentiment calculations.

3. Proposed CWW Sentiment Analysis

This section describes a novel semi-unsupervised cluster approach to sentiment analysis based on computing with words, called Computing with words sentiment analysis (CWW-SA). The proposed approach operates on a sentence-level that effectively encodes the characteristics of tweets and combines corpus and dictionary-driven methods to successfully acquire emotion related words for emotion classification experiments, and exploits the words strength in subjectivity lexicon to calculate the sentiment weight using the membership functions. The system has the ability to deal with tweets that have multiple opinions and has the flexibility to construct the suitable protoforms to describe the tweet features in natural language. Our approach as illustrated in Fig. 1 consists of three modules: preprocessing, natural language processing (NLP), and CWW module.

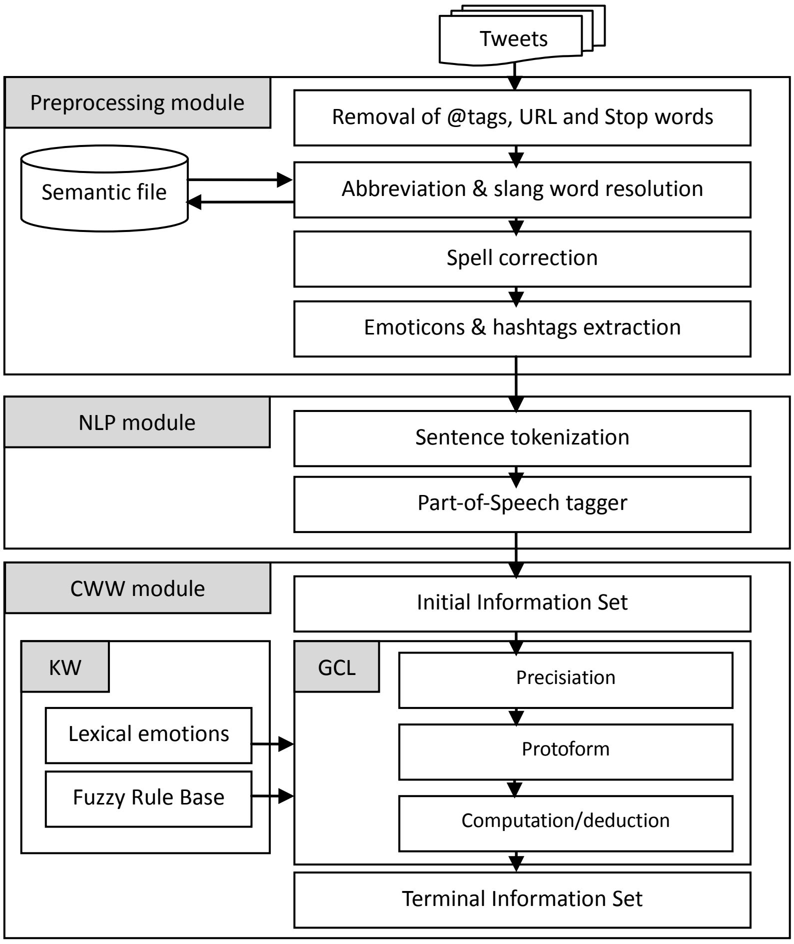

Fig. 1. The proposed CWW-SA approach.

3.1. Preprocessing Module

Tweets are noisy texts and preprocessing tasks are required to reduce data sparseness and improve the generalization capability of the proposed algorithm. All the tweets were preprocessed as follows: (1) Remove all targets (e.g. @username), URLs (e.g. www.url.com) and stop words (e.g. the), (2) Replace all abbreviation and slang words to original words using a semantic file, which has list of abbreviation words used in Twitter and its original words (e.g. *Gr8* replaced to *great*, *LOL* replaced to *laugh out loud*), (3) Correct spellings, a sequence of repeated characters are tagged by a weight (e.g. niiiiiice become nice and consider the weight of repeat). We do this to differentiate between the regular usage and emphasized usage of a word, and (4)Extract emoticons (e. g. :) ,=) ,:-) :D :(:'(:|) and hashtags, which are used in next steps.

3.2. Natural Language Processing (NLP) Module

Natural language processing tasks could be implemented using a variety of tools; in this module we used Stanford CoreNLP toolkits (an integrated suite of natural language processing tools for English) [13] to perform the following tasks: (1) Sentence tokenization, which is a process of breaking a sentence into a list of tokens (words), and (2) Part-of-speech tagging (linguistic tagging), which is the process of marking up a tokenized word as adjectives, adverbs, verbs etc. By using part-of-speech tagging (feature); each token has its corresponding part of speech (e.g. token-POS) that is used to weight emotions.

3.3. Computing with Words (CWW) Module

This module represents the main contribution of the proposed system for sentiment analysis to calculate the sentiment of tweets with the vague or uncertain knowledge. Computing with words module consists of two sub modules, world knowledge (WK) and generalized constraint language (GCL) sub modules to deal with qualitative aspects that are presented in qualitative terms by means of linguistic variables. In the CWW words are employed as variables and perception as propositions.

3.3.1. World knowledge (WK) sub module

The world knowledge (WK) sub module consists of a lexical emotions list, and IF-THEN rules; by modifying the WK sub module a system can scale on different domains. The lexical emotions list was built by using the emotion annotation and representation language (EARL) that classifies 48 emotions [14], and enhanced by adding each emotion synonym from WordNet dictionary [15]. Furthermore, each emotion in the list is combined with its strength through employing existing subjectivity lexical corpus in [16]; finally all the emotions are classified into three fuzzy granules, positive P, negative N, and neutral U. The proposed system builds if-then-rules based on the existing assumption that if the tweet has emoticon or hashtag; often the polarity of the tweet is the same like the polarity of the emoticon or hashtag [17].

3.3.2. Generalized constraint language (GCL) sub module

Given a collection of propositions expressed in a natural language that constitute the initial information set P, the goal of the CWW is to induce the perception from p to the terminal information set q using the generalized constraint language (GCL). The problem is to compute an answer to q given p, $ans(q|p)$. GCL deduction of $ans(q|p)$ involves three steps: the first step precisiation of p and q, resulting in precisiands p^* and q^*, respectively where p^* determines POS to each token and q^* determines the granules (positive P, negative N, and neutral U) which the proposition should belong to. The second step involves construction of the protoforms of p^* and q^* where p^{**} {Sentiment(Str.POS.token/tokens)} and q^{**} {Granules ($P/N/U$) does (tokens) belongs to}. In the third step, p^{**} and q^{**} are applied to computation/deduction that conveys $ans(q|p)$ [18]. In order to define the relation between the proposition p and its granules, the system uses possibilistic GCL constraint that takes the expression "*X is R*" where R playing the role of the possibility distribution of X.

In the deduction phase, we construct three membership functions $\mu_{positive}$, $\mu_{negative}$,and $\mu_{neutral}$ to determine the fuzzy weight of each token depending on POS and strength of each token(See Eq.1 for $\mu_{positive}$ where t denotes the fuzzy weight of POS of the token and Str is the strength of the token). In our case, adjectives and adverbs are the most important parts of a sentence to decide the opinion; so we assign weight equal 0.7 to them. Also verbs sometimes carrying the opinion (e.g. love, hate, and like), we assign 0.6 to verbs and 0.5 to the rest of POS. Then we calculate the average (positive/negative/neutral) values extracted from associated membership function of all tokens for the tweet. Now, we have three polarities (averages) to the three granules (P, N, U); the final result should be the maximum polarity that the preposition's granule likely to belong.

$$\mu_{\text{positive}}(t)=\begin{cases}0 & \text{if} \quad t \text{ is neutral}\\ t^{1/2} & \text{if} \quad t \text{ is positive and Str is strong}\\ t^{2} & \text{if} \quad t \text{ is positive and Str is weak}\\ 0 & \text{if} \quad t \text{ is negative}\end{cases} \tag{1}$$

4. Performance Evaluation

In this section, we evaluate the whole system and present results for predicting the semantic orientations on Twitter. The datasets used in our experiments is the same test set for Task 10: Sentiment Analysis in Twitter at SemEval-2015[19], which consists of 353 tweets labeled as positive, 183 tweets labeled as negative, and 319 tweets labeled as neutral. The performance measure used to evaluate the algorithm was F-measure [20]. First, the performance of the system is evaluated using the data in the confusion matrix as shown in Table 1 for three granules classifier. The overall classification accuracy was 0.6607, 0.4843, and 0.4907 for each granule respectively with the use of this unstructured contest dataset that contains most common challenges such as negation, sarcasm, contrasting opinions, slang words, and grammatical mistakes. The obtained test results exhibit accurate classification with low false alarms with this test set.

The results show that the classification performance of negative and neutral granules is less than the positive granule. One of the possible explanations is that the number of positive emotions and emoticons in the used lexicon list is greater than the negative and neutral ones. In general, the performance of our system could be improved when more IF-THEN rules are available to the WK sub module, which represent the expertise knowledge and using large lexicon list. According to experimental results in table 2, the proposed approach significantly improves sentiment classification compared with the DIEGOLab approach[21] that employs a support vector machine classifier trained using a number of features including n-grams, dependency parses, synset expansions, word prior polarities, and embedding clusters. However, this system classifies positive and negative granules only whereas the suggested system that classifies 3 granules.

Table 1. Matching Matrix Results

		Predicted			
		Positive (P)	Negative (N)	Neutral (U)	Accuracy
Correct	Positive (P)	**259 (TP)**	40 (FN)	54 (FU)	0.6607
	Negative (N)	54 (FP)	**93 (TN)**	36 (FU)	0.4843
	Neutral (U)	118 (FP)	68 (FN)	**133 (TU)**	0.4907

Table 2. Comparison of Our System with DIEGOLab System

	F-measure	Classification classes		
		positive	negative	neutral
CWW-SA (our system)	**0.5725**	√	√	√
DIEGOLab system	0.5672	√	√	

5. Conclusion

This paper proposes a methodology for classifying sentiment from Twitter using Computing with Words. Our system exploit the capability of computing with words to operate on information described in natural language. The system uses part of speech tagging as feature, and depends on the capabilities of generalized constraints language to infer an answer to a query expressed in a natural language “which granule does the proposition belongs to?” There are several key points that why our technique is able to achieve a significant improved performance: (1) CWW is a powerful tool and it is built to resolve the problems of extraction and presentation of opinions with their weight-ages and strength. (2) The system addresses properly problems like negations, vague words and ambiguous words. (3) CWW provides a much more expressive language for knowledge representation and much more versatile machinery for reasoning and computation. Experimental results show that the proposed approach achieves good performance in sentiment analysis along different tweets from the contest dataset. As a further study, we plan to enhance the lexical list and if-then rules that are used as world knowledge in CWW module to improve the accuracy.

References

[1] Lawrence, N., Nicolas, J., & Herman, A. (2014). Importance of the neutral category in fuzzy clustering of sentiments. *International Journal of Fuzzy Logic Systems*, *4(2),* 1-6.

[2] Pranali, T., Ashish, S., & Sachin, J. (2014). Opinion mining in natural language processing using SentiWordNet and fuzzy. *Journal of Emerging Trends & Technology in Computer Science*, *3(3),* 153-158.

[3] Shweta, R. (2014). Sentiment analysis for Hindi text using fuzzy logic. *Indian Journal of Applied Research*, *4(8),* 437-440.

[4] Maqbool, A., Noamie, S., & Ahmed, A. (2014). Semantic and fuzzy aspects of opinion mining. *Journal of Theoretical and Applied Information Technology*, *63(2),* 330-342.

[5] Nidhi, M., & Jha, K. (2012). Classification of opinion mining techniques. *International Journal of Computer Applications*, *56(13),* 1-6.

[6] Lotfi, A., & Zadeh. (2002). From computing with numbers to computing with words — from manipulation of measurements to manipulation of perceptions. *International Journal of Applied Mathematics and Computer Science*, *12(3),* 307-324.

[7] Pakizar, S., & Atsushi, I. (2012). Computing with words for direct marketing support system. *Proceedings* of *the Midwest Artificial Intelligence and Cognitive Science Conference* (pp. 19-26).

[8] Zhao, D., Gao, Z., & Bie, R. (2007). Computing with words in data mining and pattern recognition. *Advanced Intelligent Computing Theories and Applications. With Aspects of Contemporary Intelligent Computing Techniques*, 285-293. Springer Berlin Heidelberg.

[9] Guohong, F., & Xin, W. (2010). Chinese sentence-level sentiment classification based on fuzzy sets. *Proceedings of the 23rd International Conference on Computational Linguistics* (pp. 312-319).

[10] Animesh, K., & Deba, M. (2010). Finding opinion strength using fuzzy logic on web reviews. *International Journal of Engineering and Industries, 2(1),* 37-43.

[11] Samaneh, N., Masrah, M., & Rabiah, A. (2010). Sentiment classification of customer reviews based on fuzzy logic. *Proceedings of International Symposium Information Technology*: *Vol. 2* (pp. 1037-1040).

[12] Cakmak, O., Kazemzadeh, A., Yildirim, S. & Narayanan, S. (2012). Using interval type-2 fuzzy logic to analyze Turkish emotion words. *Proceedings of Information Processing Association Annual Conference* (pp. 1-4).

[13] Stanford EDU. From http://nlp.stanford.edu/software/corenlp.shtml

[14] Ellen, D., *et al*. (2007). The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data. *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction* (pp. 488-500).

[15] George, A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38(11),* 39-41.

[16] PITT EDU. From http://mpqa.cs.pitt.edu/lexicons/subj_lexicon

[17] Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *Proceedings of the 7th Conference on International Language Resources and Evaluation* (pp. 1320-1326).

[18] Lotfi, A., & Zadeh. (2006). Generalized theory of uncertainty (GTU)-principal concepts and ideas.

Computational Statistics & Data Analysis, *51(1),* 15-46.

[19] International Workshop on Semantic Evaluation 2015. From http://alt.qcri.org/semeval2015

[20] Lima, A., & de Castro, L. (2012). Automatic sentiment analysis of twitter messages. *Proceedings of 4th International Conference on Computational Aspects of Social Networks* (pp. 52-57).

[21] Abeed, S., Azadeh, N., & Davy, W. (2015). DIEGOLab: An approach for message-level sentiment classification in Twitter. *Proceedings of the 9th International Workshop on Semantic Evaluation* (pp. 510-514).

Saad M. Darwish received his Ph.D. degree from the Alexandria University, Egypt. His research and professional interests include image processing, optimization techniques, security technologies, and machine learning. He has published in journals and conferences and severed as TPC of many international conferences. Since Feb. 2012, he has been an associate professor in the Department of Information Technology, IGSR, Egypt.

Magda M. Madbouly received her Ph.D. degree from the Alexandria University, Egypt. Her research and professional interests include artificial intelligence, cloud computing, neural networks and machine learning. She is an assistant professor in the Department of Information Technology, Institute of Graduate Studies and Research (IGSR), Egypt.

Mohamed A. Hassan received his B.Sc. degree in commerce from the Faculty of Commerce, Alexandria University, Egypt in 2005. Currently he is a M.Sc. student in the Department of Information Technology, Institute of Graduate Studies and Research (IGSR), Alexandria University, Egypt. His research and professional interests include text classification, sentiment analysis, fuzzy logic and machine learning.