Opinion Mining: Analysis of Comments Written in Arabic Colloquial

Ahmed Y. Al-Obaidi, Venus W. Samawi

***Abstract*— In Arab nations, people used to express their opinions using colloquial dialects depending on the country to which they belong to. Analyzing reviews written in various Arabic dialects is a challenging problem. This is because some words could have different meanings in various dialects. Furthermore, dialects could contain words that do not belong to classical Arabic language. This research tackles the problem of sentiment analysis of reviews and comments written in colloquial dialects of Arabic language, at which the ability of different machine learning algorithms and features are examined in polarity determination. In this work, people's reviews (written in different dialects) are classified into positive or negative opinions. Each dialect comes with its own stop-words list. Consequently, a list of stop-words that suits different dialects in addition to modern standard Arabic (MSA) is suggested. In this paper, a light stemmer that suits dialects is developed. Two feature sets are utilized (bag of words (BoW), and N-gram of words) to investigate their effectiveness in sentiment analysis. Finally, Naïve-Bayes, Support vector machine (SVM), and Maximum Entropy machine learning algorithms are applied to study their performance in opinion mining. F1-measure is used to evaluate the performance of these machine learning algorithms. To train and test the suggested system performance, we built a corpus[1] of reviews by collecting reviews written in two dialects (Saudi dialect and Jordanian dialect). The testing results show that Maximum Entropy outperforms the other two machine learning algorithms. Using N-gram (with N=3) as features set improves the performance of the three machine learning algorithms.**

***Index Terms*— Arabic Colloquial Dialects, Opinion Mining, Sentiment Analysis, Machine Learning, Natural Language processing.**

I. Introduction

ANALYZING sentimental contents is a gold mine for individuals and companies to track their reputation and get timely feedback about their products and actions. Sentiment analysis offers these organizations the ability to monitor different social media sites in real time and act accordingly. Marketing managers, campaign managers, politicians, and even equity investors and online shoppers are the direct beneficiaries of sentiment analysis technology. Social networks and hundreds of other sites receive every day huge number of sentimental contents generated by internet users about every single aspect of life. Most users write their sentiments in their colloquial variant of their language.

To accomplish and improve sentiment analysis, various linguistic features and machine learning algorithms are used. Pang et al [1] investigate the use of different features and machine learning approaches to determine the polarity. N-gram approach and part of speech (POS) tagging are used as features. Naive Bayes Classification, Support Vector Machines (SVM), and Maximum Entropy are trained and tested with three folds cross-validation. Maximum accuracy achieved (82.9%) when SVM used with unigrams presence approach. Turney [2] used semantic orientation scores of the constituent adjectives to assess the sentiment orientation of customer reviews. Co-occurrence frequency of adjectives on the Web (with several positive or negative seed adjectives) was used to measure the orientation of adjectives. In (2004), Kim and Hovy [3] used WordNet distance (from positive and negative seed words) to determine polarity scores to a large list of words. In [4] Hiroshi et al extract sentiment scores for all words in the documents using deep language analysis for machine translation. Kennedy and Inkpen [5] specify the sentiment of customer reviews by counting positive and negative terms, taking into consideration contextual valence shifters (e.g. negations and intensifiers). In [6], Blitzer et al examined domain adaptation for sentiment classifiers. They used online reviews for different products. In [7], Andreevskaia and Bergler combined a lexicon-based classifier and a corpus-based classifier, using precision based weighting, to improve classification performance. Tkachenko and Lauw [8] developed a generative model for comparative text by extracting statements of comparing products from review comments and generate a gold standard of product quality for specific predefined characteristics. In [9], Kessler et al predict products ranking using gold rankings sources. Dictionary-based, machine learning, and comparison-based are used as opinion mining methods to perform product ranking.

In this study, we are interested in opinion mining for sentiments written by Arabic users. Although Arabic sentimental comments are plentiful on Internet, there are few attempts to build opinion mining systems for Arabic Language (which is a morphologically-Rich Languages (MRL)) [10].

Regarding opinion mining for Arabic language, Ahmad et

Manuscript received June 17, 2016; revised August 8, 2016.

Venus W. Samawi: Department of Computer Multimedia Systems, Faculty of Information Technology, Isra University, Amman, Jordan. P.O. Box 22, 33 (11622); Email: venus.samawi@iu.edu.jo or drvenus2004@gmail.com).

Ahmed Y. Al-Obaidi: Lead software engineer, with Internet Brands, Inc, LA, 909 North Sepulveda, BLvd ElSegundo, CA 90245 USA. (E-mail: ahmad.alobaidy@gmail.com).

[1] *The corpus is made publicly available at https://code.google.com/p/omcca/*

al [11] a local grammar approach for three languages: Arabic, Chinese and English is suggested. They selected and compared the distribution of words in a domain-specific document to the distribution of words in a general corpus. Abbasi et al [12] performed a study for sentiment classification on Arabic and English inappropriate content. They applied their approach on a U.S. supremacist forum for English and Middle Eastern extremist group for Arabic language. Saleh et al. [13] showed that using automatic machine translation to translate Arabic text to English and then perform analysis will, slightly, lower results compared with direct analysis of Arabic text. Muhammad Abdul-Mageed et al. [14] constructed a sentence-level sentiment analysis system for MSA contrasting language independent only features and combining language independent and language-specific feature sets, namely morphological features specific to Arabic. They find that the stem lemmatization setting outperforms both surface and lemma settings. They also empirically illustrated that adding language specific features improves performance. Misbah et al [15] suggested an optimized approach for mining opinions in Arabic Religious Decrees via an improved "Semantic Orientation using Point wise Mutual Information" algorithm. Both supervised and unsupervised learning algorithms are trained. The achieved accuracy rate was 73.08%. In [16], Al-Kabi et al develop a tool to analyze Arabic opinions written in colloquial Arabic and/or Modern Standard Arabic (MSA), and determine the opinion polarity based on manually constructed lexicon. Also opinion features are collected manually. Up to our knowledge, none of these few researches has investigated the effect of features and machine learning algorithms on determining polarity when colloquial varieties of Arabic are used. This study tries to fill this gap.

As mentioned before, the problem of classifying reviews written in Arabic dialects is challenging problem. This is due to the facts: (1) Arabic dialects lack grammatical case. (2) Dialects have a more complex cliticization system than MSA, where circumfix negation is allowed, and the attached pronouns could act as indirect objects. (3) The same word in different Arabic dialects may mean different thing, and it may has different synonyms in each dialect. (4) Stop words varies from Arabic dialect to another, and also may differ MSA. In this work, we suggested a stop words list that suits all dialect and differ than the MSA stop words. We also, developed a modified light stemmer to suit the reviews written in Arabic dialects.

The purpose of this study is to determine the effect of linguistic features and the used machine learning algorithm on the performance of sentiment analysis for a colloquial variety of Arabic language. Linguistic features will be extracted from (bag of words, and phrases of N-gram (N=1, 2, or 3). The resulted features will be used to train three different machine learning algorithms (Naïve Bays, SVM, and Maximum Entropy). To test and evaluate the system performance, five-folds cross validation with f1-measure are used. Finally, the linguistic feature that best suits sentiment analysis problem will be specified. The most proper machine learning approach will be determined.

This paper is organized as follows. Section 2, illustrates how the corpus is constructed. The main ideas underlying opinion mining, and tackles the construction of the suggested sentiment analysis model is exemplifies in section 3. In section 4, the evaluation of the proposed sentiment analyzer behavior is discussed, and the recommended classifier, along with the proper sentiment features is determined. Finally, we concluded in section 5.

II. Data Collection

Since there is no data publicly available which is suitable for researches concerning opinion mining in colloquial varieties of Arabic language, we had to collect our own. The data has been collected from Jeeran web site [17]. Jeeran is a reviewing platform for the Arab world launched in 2010. It provides a platform for users to add their reviews regarding various kinds of public places, such as hotels, shops, restaurants and libraries. In Jeeran web site, to write a review, user should provide a textual opinion about the place which is to be reviewed, in addition to a numerical rate for the place. The rate is between 1 and 5, where 1 represents an extremely negative opinion, and 5 is extremely positive. Most opinions are written in the dialect of the country in which the place is located. The usable data are for places in Hashemite Kingdom of Jordan (HKJ), and the Kingdom of Saudi Arabia (KSA). Reviewers from other countries are few compared with HKJ and the KSA reviewers, which makes them unusable for machine learning process.

To collect the data and construct a dataset suitable for opinion mining for different Arabic dialects, three steps are implemented. At first, perform web crawling, then filter the reviews (keep only the reviews of Hashemite Kingdom of Jordan and the Kingdom of Saudi Arabia reviewers). Finally, extract some attributes from each review. The collected dataset consists of 28,576 reviews, which represents the opinions of 5,422 different reviewers. It covers 27 different categories. Table I shows the details statistics of the extracted reviews.

TABLE I
EXTRACTED REVIEWS

City	# Negative Reviews < 3	# Neutral Reviews = 3	# Positive Reviews >3	Total
Damma	119	275	937	1,331
Khobar	114	298	1321	1,733
Jeddah	481	645	5284	6,410
Riyadh	664	893	7251	8,808
Amman	1246	1554	7494	10,294
Total	2,624	3,665	22,287	28,576

A. Web Crawling

A script has been written to download the pages of the web site. Since Jeeran's web server limits the number of requests from single client per seconds, we had to pause 1 second between requests. The crawling took one week to finish the web site. At the end, 1 GB of data is collected. The collected data represent reviews about 5,043 different places such as shops, restaurants, and malls.

B. *Reviews Filtering*

The downloaded data was in HTML format. Fig. 1 shows a screenshot of sample reviews and how useful data is rendered in a web browser.

Fig.. 1. Screenshot of sample review

A Java program was written to parse the HTML pages, and extract the useful data from them.

C. *Attribute Extraction*

The program extracted the following attributes from each review:

1) Review ID: identification-number of the review, which is unique in the entire corpus.
2) Subject: name of the place which is been reviewed.
3) Category: shops, hotels, bookstores, etc.
4) Rate: numerical representation of the opinion, from 1 to 5 (1 represents an extremely negative opinion, 5 extremely positive).
5) Body: textual opinion of the place in natural language.
6) Useful: the number of people who find the review helpful.
7) Author: name of the opinion holder.
8) Number of Reviews for the author: number of all the reviews published by this author
9) City: the city in which the place is located.
10) Date: the date in which the review was written.

Although, we have extracted all the above attributes, we have used only two of them. The rest could be used in a future works. We used the Body as textual sentiment, the rate as human provided class to the review (negative or positive).

D. *Data Balancing*

From Table 1, one can observe that the number of positive reviews is much more than the negative reviews. This leads to produce unbalanced training set. Using unbalanced data will often cause misclassification of documents which in reality belong to the rare classes. That would mean the positive-sentiments will dominate the negative sentiments, causing misclassification of the negative sentiments. Dataset could be artificially rebalanced by down-sampling positive reviews, or up-sampling negative reviews. In down-sampling, randomly select positive reviews equal to number of negative reviews. In up-sampling technique, negative reviews are duplicated, and the number of randomly selected positive reviews is equal to the number of the resulted negative reviews [18], [19].

III. OPINION MINING

Textual information mainly could be categorized into two types: facts and opinions. Facts can be supported by objective evidence and empirically proven to be true. They are objective expressions about entities, events and their properties. On the other hand, opinions are subjective expressions that depend on person feelings, which vary from one person to another based on, persons prospective, emotions, and understandings. So, opinions are beliefs that describe people’s sentiments, judgments or feelings toward entities, events and their properties. In this work, we only focus on opinion expressions that express people’s positive or negative sentiments. Therefore, opinion mining is considered as text classification problem. In this work, sentiment analysis model for a colloquial variety of Arabic language (SAMCAL) is suggested. The textual sentiment is to be classified into one of two categories, either positive or negative. Supervised machine learning is used to design the textual sentiment classifier. The rate field has been considered as the human provided class; rate of 4 and more are considered positive, and 2 or less are considered negative. The suggested model mainly consists of three stages, preprocessing stage (normalization, stop word removal, and modified version of light stemming), document representation and features extraction (bag of words, and word’s, and N-gram phrases), and classification stage. For training and testing the system, cross validation 5-fold is used. Dataset is partitioned into two sets; training set which consists of 80% of total set and testing set which consists of the remaining 20%.

A. *Preprocessing Stage*

Each document, whether training or testing document, is passed through three preprocessing steps: normalization, stop word removal, and modified version of light stemming.

Normalization

At first, each Arabic word is normalized as follows:

1) Remove punctuation
2) Remove diacritics (primarily weak vowels). Some dictionary entries contained weak vowels. Removal made everything consistent.
3) Remove Kashida or elongation character “ـ” which make, for example, جمـــيـــــــــل same as جميل (Jameel, beautiful).
4) Remove non letters (numbers, non-Arabic letters, punctuation marks, etc).
5) Replace (alef mad'da) آ , (alef hamza)أ , and (alef kasra) إ with (alef) ا
6) Replace final (alef maksura) ى with (ya'a) ي
7) Replace final (ta’a marbuta) ة with (ha’a marbuta) ه

Removing Stop Words

Stop words should be considered very carefully not to contain any negation or adjective word since these kinds of words are essential in opinion text. Some of the stop words are dialect specific, which is not used normally in standard Arabic text classification systems. Table II illustrates the suggested stop words list that suits both Jordanian and Saudi Arabia dialects.

TABLE II
SAMPLE STOP WORDS.

Iza	ازا	Kaan	كان	lia'an'na	لان
Min	من	Kuent	كنت	lia'an'naha	لانها
Eilla	الى	Kaanat	كانت	wikem'n	وكمان
Fee	في	ind'da	عند	taba'an	طبعا
Wa	و	sur'et	صرت	BisAraha	بصراحة
Ao	او	Webbess	وبس	Alsaraaha	الصراحة
Ana	انا	Huna	هذا	ind'ahom	عندهم
an'na	ان	Wemin	ومن	a'anhha	عنها
Lao	لو	a'ad'a	عاد	thoum'ma	ثم
na'am	نعم	an'nahu	انه	katha'alieka	كذلك
Sart	سرت	Alheen	الحين	aid'daan	ايضا
Laha	لها	kem'n	كمان	beln'nisbah	بالنسبة
Haik	هيك	min'ni	مني	lah'huwa	وله
		Wahuna	وهذا	lea'an'nahu	لانه
		Feehe	فيه	kuen'na	كنا

Stemming

In this work, slightly modified light stemmer is developed to suit the task on opinion mining and user generated content. The main steps of the light stemmers are:

1) *Prefix Remova*l: In this stage, the stemmer removes prefixes listed in Table III if it leaves 3 or more characters.
2) *Suffix Removing*: Table III also shows a list of Arabic suffixes. In this stage, the stemmer removes the suffixes if it leaves 3 or more characters.

The modification we contributed to the stemmer was to remove any repetition in letters. The reason why this step is essential for user generated content is because users sometimes repeat letters as indication of their enthusiasm. For example, user may use “greaaaaat” instead of “great” or “جميييييل” (Jameeeeeeeeel) instead of “جميل” (Jameel) to indicate her or his enthusiasm.

TABLE III.
AFFIXES OF LIGHT STEMMER.

Prefix		Suffix	
Waal	وال	Ha	ها
Bal	بال	An’na	ان
Faal	فال	At	ات
Kaal	كال	W n	ون
Al	ال	Yen	ين
		Yh	يه
		Ha’a	ه

B. *Document Representation Stage*

Different kinds of features are used by researches concerning sentiment classification. In any machine learning application, the main task is to find a suitable set of features that improves the classification results. Terms and their frequency, part of speech tags, opinion words and phrases, syntactic dependency, and negation [20] are the most popular features that are used in sentiment analysis. Before extracting features, it is important to choose proper document representation approaches (e.g. single words, stemmed single words, phrases, and N-grams). In this study, the documents are represented as bag of words (stemmed single words), and N-grams phrases (the phrase could be of length 1, 2, and 3 stemmed words).

C. *Classification Stage*

Opinion words and phrases are used for training a sentiment classifier (supervised, or unsupervised). In this work, we are interested in supervised classification, which can be formulated as a supervised learning process with two class labels (positive and negative). Three different classifiers are applied to sentiment classification. The classifiers are trained using bag of words and 3-gram phrases. The used classifiers are:

– Support vector machines (SVM).
– Naïve Bayes (multinomial Naïve Bayes)
– Maximum Entropy.

IV. EVALUATION

Text classification systems are typically evaluated using same performance measures of information retrieval systems. These metrics include recall, precision, accuracy and error rate and F1-measure. Given a test set of N documents, these values could be easily computed [21]:

1) *True Positive (TP):* the number of items correctly labeled as belonging to the positive class.
2) *True Negative (TN):* the number of items correctly labeled as belonging to the negative class.
3) *False Positive (FP):* the number of items incorrectly labeled as belonging to the positive class.
4) *False Negative (FN):* the number of items incorrectly labeled as belonging to the negative class.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1-measure = 2.\frac{Precision \cdot Recall}{Precision + Recall} \tag{3}$$

$$Error = \frac{FP + FN}{N} \tag{4}$$

Most often, large imbalance between the numbers of positive versus negative examples will cause TN or TP to dominate the accuracy and error rate of a system. This will cause misinterpretation of the system results. For example, in case negative examples of a category constitute 95% of the test set, a trivial classifier which makes negative predictions of all documents has an accuracy of 95% (or 5% error rate). Though, such system is useless. Therefore, recall, precision, and F1-measure are more commonly used in text categorization evaluations (instead of accuracy and error rate). In this work, F1-measure is used as performance metric since it gives more complete measure.

In this work, three supervised classifiers are trained and tested. We conducted many experiments, and analyzed the result in order to better understand the behavior of each classifier. The performance of each classifier is evaluated using F1-measure. Their results are compared to find out the classifier that most suits this application.

A. Experiment-1: Assessments of Classifiers

In the first experiment, we transformed the raw text into a bag of words. Bag of words represents unordered collection of sentiment words, disregarding the grammar (Arabic dialects lack grammatical case), and even word order. The vectors are feed to the classifier. All tests in this study are done using programs we developed using Java and with help of APIs from WEKA. Table IV shows the results of each tested algorithm. The table shows that Maximum Entropy has the best performance measure. Table IV shows that Maximum Entropy has slight advantage over Naïve Bayes. SVM has noticeably performed less than both. This should be expected since the real advantage of discriminative model such as SVM and Maximum Entropy appear usually with larger training sets. On the other hand, Naïve Bayes is known for performing well with relatively smaller sets.

TABLE IV.
PERFORMANCE OF THE 3 MACHINE LEARNING ALGORITHMS

Algorithm	F1-measure
SVM	0.8194
Naïve Bayes	0.8329
Maximum Entropy	0.8383

B. Experiment-2: Assessments of Feature Extraction

Instead of using bag of words, we have applied N-gram of words to represents sentiments. N-gram is a contiguous sequence of N or less words that appear document. For example for N=2 or less words, from the phrase “The food was not good”, we get the following features:

“The”, “food”, “was”, “not”, “good”, “The food”, “food was”, “was not”, “not good”

As we can see, this will give a better handle for negated adjectives. For example if “مش” which means “not” came before “كويس” which means good, it will negate the adjective and it is much better to handle two words as single one. We used N=3 to set the maximum sequence to 3 words. N with larger value has been tried but didn’t improve the performance, as illustrated in Table V. Using N-gram as sentiment feature has improved the performance of all classifiers. We think this happened because N-gram concatenates the negation with subject followed, which is very important in sentiment analysis.

TABLE V.
F1-MEASURE USING N-GRAM, AND BAG OF WORDS.

Classifier	Features	
	BoW	N-gram
SVM	0.8194	0.8343
Naïve Bayes	0.8329	0.8646
Maximum Entropy	0.8383	0.8675

From Table V, it is clearly seen that using N-gram (as feature to train the classifiers) has improved the performance of all classifiers. We think this happened because N-gram concatenates the negation with subject followed, which is very important in sentiment analysis.

V. CONCLUSION

In this work, opinion mining for colloquial varieties of Arabic language has been studied. The problem has been approached as text classification using supervised machine learning algorithm. Multiple machine learning algorithms were tested. Results have showed that Maximum Entropy has slight advantage over Naïve Bayes. SVM has noticeably performed less than both. This should be expected since the real advantage of discriminative model such as SVM and Maximum Entropy appear usually with larger training sets. On the other hand, Naïve Bayes is known for performing well with relatively smaller set. Most reviews available on the internet seem to be positive. Using training set with mostly positive reviews lead to unbalanced classifier. Thus, the data need to be balanced. We used Down-sampling the positive reviews to balance out the negative ones. We also tried oversampling by duplicating the negative reviews. Oversampling hasn’t shown any improvement in performance. It increases the overall true positives, but increases the false positives too. This will keep the overall effectiveness almost the same. Using N-gram has improved the overall performance of classification. We attribute that to the better handling of the negation of adjective which is very important in sentiment analysis.

REFERENCES

[1] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in proceedings of the of the ACL-02 conference on empirical methods in natural language processing (EMNLP), vol. 10, 2002, pp. 79–86
[2] Peter D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in the proceedings of 40th Annual Meeting on Association for Computational Linguistics (ACL '02), Philadelphia, Pennsylvania, 2002, pp. 417-424
[3] Soo-Min Kim and E. Hovy, "Determining the sentiment of opinions," in proceedings of the 20th international conference on Computational Linguistics (COLING '04), Stroudsburg, PA, USA, 2004.
[4] K. Hiroshi, N. Tetsuya, and W. Hideo, "Deeper sentiment analysis using machine translation technology," in The 20th International Conference on Computational Linguistics (COLING-04), Stroudsburg, PA, USA, 2004.
[5] A. Kennedy, and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters," journal of Computational Intelligence, Vol. 22, Issue 2, 2006, pp. 110-125
[6] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, boom-boxes and blenders: domain adaptation for sentiment classification," in proceedings of the Association for Computational Linguistics (ACL), 2007., 440-447
[7] A. Andreevskaia and S. Bergler, "When specialists and generalists work together: overcoming domain dependence in sentiment tagging," in proceedings of Association for Computational Linguistics (ACL-08), HTL, Columbus, Ohio, USA, 2008, pp. 290-298.
[8] Maksim Tkachenko and Hady W. Lauw. “Generative modeling of entity comparisons in text.” Proceeding of CIKM '14, Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management Pages, ACM New York, NY, USA, 2014, pp 859-868
[9] Wiltrud Kessler, Roman Klinger, and Jonas Kuhn, Towards Opinion Mining from Reviews for the Prediction of Product Rankings, Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2015), Lisboa, Portugal, 17 September, 2015, pp. 51–57.

[10] Mohamed Elarnaoty, Samir AbdelRahman, and Aly Fahmy, "A Machine Learning Approach for Opinion Holder Extraction in Arabic Language," International Journal of Artificial Intelligence & Applications (IJAIA), Vol3, No. 2, 2012., pp. 45-63
[11] K. Ahmad, D. Cheng Y. Almas, "Multi-lingual Sentiment Analysis of Financial News Streams". In Proceedings of the First International Conference on Grids in Finance, Palermo, Italy, 2006
[12] A., Chen, H., & Salem, A. Abbasi, "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums," ACM Transactions on Information Systems, Vol. 26, issue 3, 2008, pp.1-34
[13] .Rushdi-Saleh M., Martín-Valdivia, Alfonso Ureña-López L., Perea-Ortega J., "Bilingual Experiments with an Arabic-English Corpus for Opinion Mining," in proceedings of Recent Advances in Natural Language Processing (RANLP), Hissar, Bulgaria, 2011, pp. 740-745
[14] Mona T. Diab , Mohammed Korayem, Muhammad Abdul-Mageed, "Subjectivity and sentiment analysis of modern standard Arabic," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, Oregon, 2011, pp.587-591.
[15] A.M. Misbah and I.F. Imam, "Mining opinions in Arabic text using an improved "Semantic Orientation using Pointwise Mutual Information" Algorithm," in In Proceedings of the 8th International Conference on Informatics and Systems (INFOS), Cairo, 2012, pp. SE-61 - SE-69
[16] Mohammed N. Al-Kabi, Izzat M. Alsmadi, Amal H. Gigieh, Heider A. Wahsheh, Mohamad M. Haidar, "Opinion Mining and Analysis for Arabic Language", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 5, 2014, pp181-195
[17] (2013, Jan.) Jeeran. [Online]. http://jo.jeeran.com/
[18] L.O., Hall, "Data Mining from Extreme Data Sets: Very Large and or/ Very Skewed Data Sets," in Systems, Man, and Cybernetics, 2001 IEEE International Conference, Vol. 4, Tucson, Arizona, 2001.
[19] I. Chairi, S. Alaoui, and A. Lyhyaoui, "Learning from imbalanced data using methods of sample selection," in Proceeding of Multimedia Computing and Systems (ICMCS), Tangier, 2012.,IEEE, pp 254 - 257
[20] R. McDonald, K., Hannan, T. Neylon, M. Wells, , and J.Reynar, "Structured models for fine-to-coarse sentiment analysis," in Proceedings of ACL-2007, the 45th Annual Meeting of the Association of Computational Linguistics, 2007, pp 432–439
[21] David D. Lewis, "Evaluating and optimizing autonomous text classification systems," in 18th annual international ACM SIGIR conference on Research and development in information retrieval, Seattle, Washington, 1995. pp246-254

Date of modification: 2/11/2016
The changes are:
Page 1: The email address of prof. Venus W. Samawi
Page 3: At the end of the second column, point 4 the word punctuation
Page 5: Tables number (IV, and V) was wrong
Page 4: Eq. 3
Page 6: Refine some references

ISBN: 978-988-14047-1-8 (revised on 30 November 2016) WCECS 2016
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)