ISSN 2277-3061

Sentiment Analysis of Arabic Slang Comments on Facebook

Taysir H. A. Soliman[1], M. A. Elmasry[2], A. Hedar[3], and M. M. Doss[4]
[1] Information Systems Dept., Faculty of Computers & Information, Assiut University, Egypt
taysser.soliman@fci.au.edu.eg
[2]Information Systems Dept., Faculty of Computers & Information, Fayoum University, Egypt
mostafa_elmasry2006@yahoo.com[2]
[3]Computer Sciences Dept., Faculty of Computers & Information, Assiut University, Egypt
hedar@aun.edu.com
[4]Electrical Engineering Dept., Faculty of Engineering, Assiut University, Egypt
magdy@aun.edu.eg

ABSTRACT

Social networks have become one of our daily life activities not only in socializing but in e-commerce, e-learning, and politics. However, they have more effect on the youth generation all over the world, specifically in the Middle East. Arabic slang language is widely used on social networks more than classical Arabic since most of the users of social networks are young-mid age. However, Arabic slang language suffers from the new expressive (opinion) words and idioms as well as the unstructured format. Mining Arabic slang language requires efficient techniques to extract youth opinions on various issues, such as news websites. In this paper, we constructed a Slang Sentimental Words and Idioms Lexicon (SSWIL) of opinion words is built. In addition, we propose a Gaussian kernel SVM classifier for Arabic slang language to classify Arabic news' comments on Facebook. To test the performance of the proposed classifier, several Facebook news' comments are used, where 86.86% accuracy rate is obtained with precision 88.63 and recall 78.

Keywords: - Opinion mining, Social Network, sentiment analysis, support vector machines, Arabic slang comments, Facebook, slang sentimental words and idioms lexicon, microblogs.

INTRODUCTION

The revolution of current social networks is affecting our every daily lives, having a vast amount of information through microblogs, review sites, web forums and online discussions. Sentiment analysis of social networks has recently attracted researchers because of its effect on people’s behavior whether politically or on any other discipline. In addition, users are free to write whatever they want unlike previous sentiment analysis which targeted product reviews.

Opinion Mining (OM) includes several subtasks, such as subjectivity detection, polarity classification, review summarization, emotion classification, and sentiment analysis [1]. Opinion mining, which is called sentiment analysis, can be viewed as a classification process that aims to determine whether a certain text is written to express a positive or a negative Opinion about certain object (e.g., a topic, product, or person). Sentiment analysis is typically performed using one of two basic approaches: rule-based classifiers, in which rules derived from linguistic study of a language are applied to sentiment analysis and machine learning classifiers [2]. Currently, most of the systems built for sentiment analysis are tailored for the English language [3] but there has been some work on other languages. Two main characteristics shared between many social networking services are the short length of their update messages and language variations. For example, Facebook has a limit of 420 characters for status updates and Twitter has a 140-character limit. In addition, language variations give a large variety of short forms and irregular words, specifically for youth generations. These two characteristics induce significant data sparseness and thus affect the performance of typical sentiment classifiers learned from such noisy data.

In case of Arabic language, Arabic is divided into three types: Classical Arabic, Modern Arabic, and Colloquial Arabic [4]. As the official language of 22 countries, there are 49 million Arab users of Facebook [5]. Arabic language is a high complex language, which embeds five critical challenges for NLP tasks. First, Arabic is not a case-sensitive language; it has no capital letters. Second, Arabic is a high inflectional language; often a single word has more than one affix, such that it may be expressed as a combination of prefix(s), lemma, and suffix(s) [6]. Third, Arabic has some variants in spelling and typographic forms. Fourth, Arabic texts have different sorts of ambiguities (different meanings). For example, “ رجب ”/“Ragab” in Arabic may be used as a person name, month, or a fear verb. Fifth, Arabic resources, such as corpora, gazetteers, and NLP tools, are either rare or not free [7]. Existing Facebook sentiment analysis focus on the English language but very few focuses on Arabic slang comments.

Classical Arabic is the language used during the period before and during the Islam era and in which the holy Quran was subsequently written. It contains a rich vocabulary and sophisticated grammar. Modern Arabic is derived from the classical type, where it became the formal language for literature, media, and education. It has less sophisticated grammar. Finally, Colloquial (Slang) Arabic is the language used between people in every day communication, which is the focus of this paper. The young Arab generation use some slang words that they create themselves and communicate with each other even on Facebook and Twitter.

In this paper, a sentiment analysis methodology is proposed to classify Arabic slang comments on Facebook, based on Support Vector Machine (SVM). In addition, a Slang Sentimental Words and Idioms Lexicon (SSWIL) is developed, containing words and idioms used by the Arab youth generations. The paper is organized as follows: Section two clarifies previous work. Section three illustrates the opinion mining methodology for Arabic slang comments on Facebook. Section four discusses the results. Finally, section five concludes our work and introduces future work.

RELATED WORK

Examples of Arabic opinion mining include helmy and Daud [6], where they apply Bayes Point Machine (BPM) and SVM in classification task in trustworthy and untrustworthy classification of Islamic Hadith Narration. In addition, El-Halees [7] combine classification methods to classify Arabic documents because the accuracy of most methods is law. AbdelRahman *et al.* [8] present a novel solution for Arabic Named Entity Recognition (ANER) problem, which aims at boosting the identification of extracted named entities. Elarnaoty *et al.* [9] based their feature analysis on a semi-supervised pattern recognition technique to extract opinion holder in Arabic news. Elhawary and Elfeky [10] apply sentiment analysis on Arabic reviews to extract features by using Arabic lexicon words to identify reviews’ polarity (positive, negative or neutral). Zaidan and Callison-Burch [11] work on Arabic Commentary Dataset (AOC) to apply dialect labeling task. They partition news’ comments into dialectal sentences which is more accurate than Modern Standard Arabic sentences. Saad and Ashour [12] evaluate text preprocessing on Arabic text mining, stemming and pruning, document normalization and term weighting and enhance text representation. They study the impact of text-preprocessing and different term weighting schemes on Arabic text classification. In addition, they apply Boolean model, TF, IDF and TF-IDF for term weighting and apply C4.5 decision tree in classification task. Hijjawi and Bander [13] represent an approach of identification of opinions based on ontological exploration of texts.

We previously introduced this work but only three Facebook pages’ comments are taken [14]. Other previous work on Arabic social networks comments are rarely taken on political examples [15,16]. In case of English language, a large amount of work have been conducted on twitter sentiment analysis, such as Barbosa and Feng [17], Speriosu et al. [18], Bifet and Frank [19]. In addition, Tumasjan et al. [20] presented the political sentiment in microblogs, where they analysed 104,003 tweets before German federal election to predict election results. Diakopoulos and Shamma [21] tracked real-time sentiment analysis during US presidential election in 2008. Conover et al. [22] examined the retweet network of 250,000 political tweets during the six weeks prior to the 2010 U.S. midterm elections. Livne et al. [23] studied the use of Twitter by almost 700 political party candidates during the midterm 2010 elections in the U.S. Our research focuses on free text Arabic written by web users who comment on Facebook and Twitter or news websites. Comments on those networks are written with new sentiment words and idioms. So, they need new lexicon to facilitate feature extraction and sentiment analysis. In the next section, proposed sentiment analysis approach is explained.

PROPOSED SENTIMENT ANALYSIS APPROACH

In this section, the proposed sentiment analysis approach consists of three phases: data preparation (comments collection, XML View), data preprocessing (remove stop words, data auto correction, stemming) and data classification, as illustrated in Fig. 1. In last phase, the system classifies the data in three types: classifying comments based on classical opinion words lexicon, classifying comments based on classic lexicon and SSWIL and classifying using SSWIL only.

FORMAL AND SLANG ARABIC LANGUAGE

NEW TOOLS EXILES ANALYSIS

Tools exiles in formal Arabic language are “لا, لن, لم, ما, ليس, ليست”, which are “la, ln, lm, ma, lays, laysat” mean “No, Not” and negative suffixes and prefixes, which are exiles in English. In slang Arabic language, there is an exile tool “ش, شي” added as a suffix of the Arabic verbs like "يحكم" which is “Govern” to get the opposite meaning, which is “يحكمش” or “يحكمشي”, where it is “لا يحكم” in formal Arabic language that means “Not Govern” in English. Sometimes, Egyptians use this exile tool another way, they add character “م” at the beginning of the verb (replace the formal exile tool “ما”) then they add the previous suffix, the previous example will be “ميحكمش” or “ميحكمشي” that give the same meaning. To clarify the previous exile tool, for example “ميحكمش مصر” means “not govern Egypt” since the word “ميحكمش” contains prefix “م” and suffix “ش”. Another exile tool is a suffix “مش” which means “لا” as formal Arabic exile tool and mean “No” in English.

PROPOSED SLANG SENTIMENTAL WORDS AND IDIOMS LEXICON (SSWIL)

Annotated corpora for training and testing are very important to make sentiment analysis possible. A number of research groups have developed Arabic sentiment analysis corpora [1,24,25]. Furthermore, Abdul-Mageed and Diab [26] use a machine translation procedure to translate available English lexicons into Arabic, retrieving 229,452 entries, including expressions commonly used in social media. In addition, Rushdi-Saleh *et al.* [27] have developed Opinion Corpus for Arabic (OCA), including a parallel English version called EVOCA. Comments perform a binary classification problem, having people “satisfied” and “dissatisfied” with particular news. After surveying comments of 1846 comments of famous Arabic Facebook pages, as shown in Table 1, new 43 words and new 27 idioms are collected, having a “Satisfied” class, and about new 91 words and new 31 idioms means, having “dissatisfied” which do not exist in Arabic lexicon. This way of writing has new opinion words that describe the satisfaction or dissatisfaction of comments like “ حلو, كويس, تظبيط, جامد” , which are “Helw, kowyes, tazbeet, gamed” which means “Satisfy”. “وحش, فاكس, تعبان, زفت” which are “wehesh, fakes, ta’ban, zeft” which means “Dissatisfy” are also added, as shown in Table 2. There are also new idioms in this way of writing that describe the satisfaction or dissatisfaction of comments like “ روش طحن, زي الفل” which are “Rewesh tahn, zay elfol” which means “Satisfy” and “يخبط فالحلل, هبل فالجبل” which are “yehkabat felhelal, habal felgabal”, which mean “Dissatisfy”.

Table 1. Sample of Arabic Slang Facebook comments

English Translation	Comment in Arabic
Willing god, Founding Committee of the Constitution write a good Constitution	انشاء الله اللجنة التأسيسية الموجودة حاليا تحط دستور كويس
You know you people and more welcome comes guest	انت عارف شعبك كويس وكثر الترحيب بيجيب الضيف المش كويس
Bad elections and the country in a bad state	انتخبات فكسانه والبلد فحاله تقرف

Table 2. Examples of SSWIL

Examples of SSWIL for satisfaction words	جامد ,حـ لو ,كـويـ س ,مـ بروك, دلـ ع , مظ بوط , صح , شغال , دلـ ع , روش ,كـ فاءه ,عـ سل , سـ كر, جمـ يل , نـ ضـ يف, حـ بوب
Examples of SSWIL for satisfaction idioms	الله ,الله عليك ,اخر حلاوه ,الحمد لله ,ماشاء الله ,ربنا معاك ,زي الفول يكرمك جامد جدي ,من الاخر ,جـ نـ يه دهب , زي الـ عـ سل , روش طحن
Examples of SSWIL for dissatisfaction words	فـ اكـ س ,مجرم ,لـ ئـ يم ,اتـ م ,تـ عـ بان , فـ اكـ سان, قـ رف, بـ ايـ ظ, فـ كك ,بـ هدلـ ه ,عـ كـ نن , صدع ,بـ ؤس ,حـ شو ,خـ نـ يق ,لـ لا سف
Examples of SSWIL for dissatisfaction idioms	كـ بر دماغك , كـ لام فـ الـ هجايـ ص, حاجه تـ قرف , فـ اكـ س موت مـ يه من تـ حت تـ بن , يـ خـ بط فـ ي الـ حـ لل

DATA PREPARATION

The objective of this phase is to collect Arabic slang comments from different sites. In addition, since thse comments these comments are into XML format.

1) *Comments collection:* To collect comments, the dataset is collected from news websites like: Aljazera[1], bbcarabic[2], Alyoum Alsabe[3] and Alarabia[4], Constitution Facebook Page, People's Opinion Facebook page, where these portals are very popular in the Arabic countries. We collected 1846 comments from previous websites, where the number of comments taken from each website is shown in Table 3.

Table 3. Comments Datasets and its sources

News Website	Num. of comments
Aljazera.net	370
Alarabia.com	472
BBCArabic.com	23
Youm7.com	490
Constitution Facebook Page	197
People's Opinion Facebook page	294

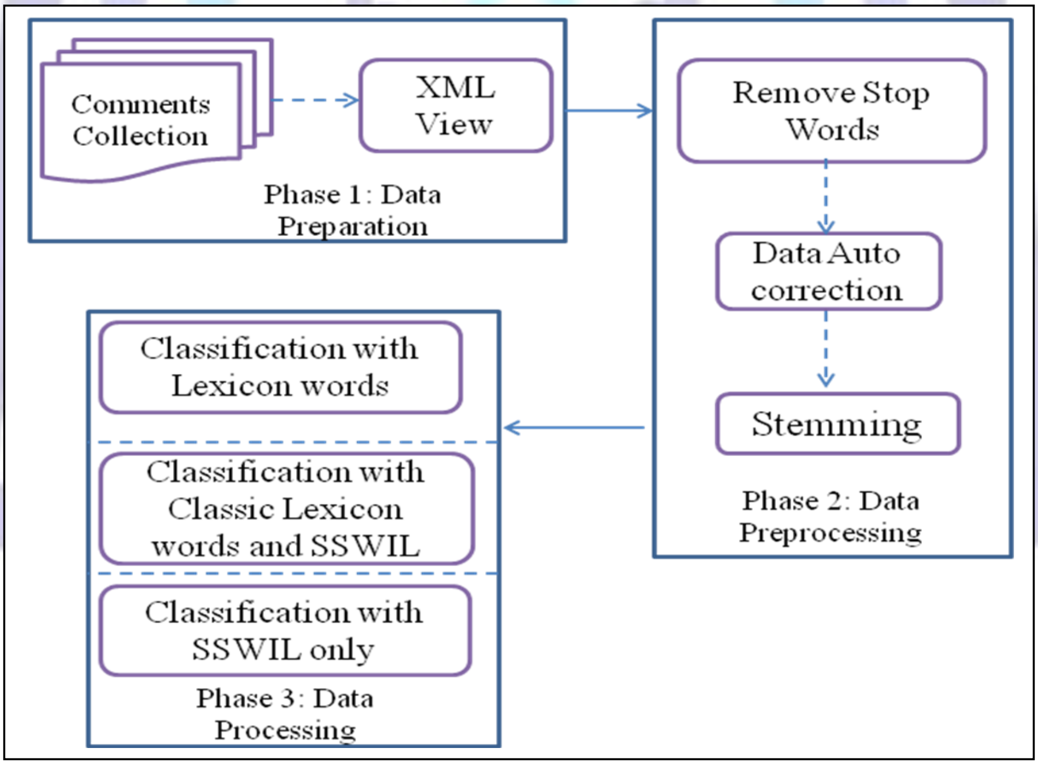

Comments Collection
XML View
Phase 1: Data Preparation
Remove Stop Words
Data Auto correction
Stemming
Phase 2: Data Preprocessing
Classification with Lexicon words
Classification with Classic Lexicon words and SSWIL
Classification with SSWIL only
Phase 3: Data Processing

Fig.1 Proposed system phases

The comments are written in Arabic language with free text format. Different point of views about the news from the Middle East is written, showing satisfaction or dissatisfaction. Taking Egyptian presidential elections as an example, the percent of satisfaction and dissatisfaction on Egyptian presidential election in 2012 are taken. Taking the comments of

[1] http://www.aljazeera.net/portal

[2] http://www.facebook.com/bbcarabicnews?ref=ts

[3] www.youm7.com,

[4] *ww.alarabiya.net/*

same news from various news websites. The comments are written freely, without grammar rules, unstructured and with slang language that have non lexicon words and idioms. Example is:

يعم وطي صوتك حرام عليك

This example indicates that there are new idioms not included in Arabic lexicon, such as “حرام عليك” which shows that the commenter is not satisfied with what is happening. The comments are written in many forms.

- Direct Comments: are comments that are related with the topic and it is useful in our study, where they are written directly with expressive words. ل لصـ بر حدود))) ... ل كن ال شعب صرىالم صـ بور

Patience has limiations but Egyptians have patience

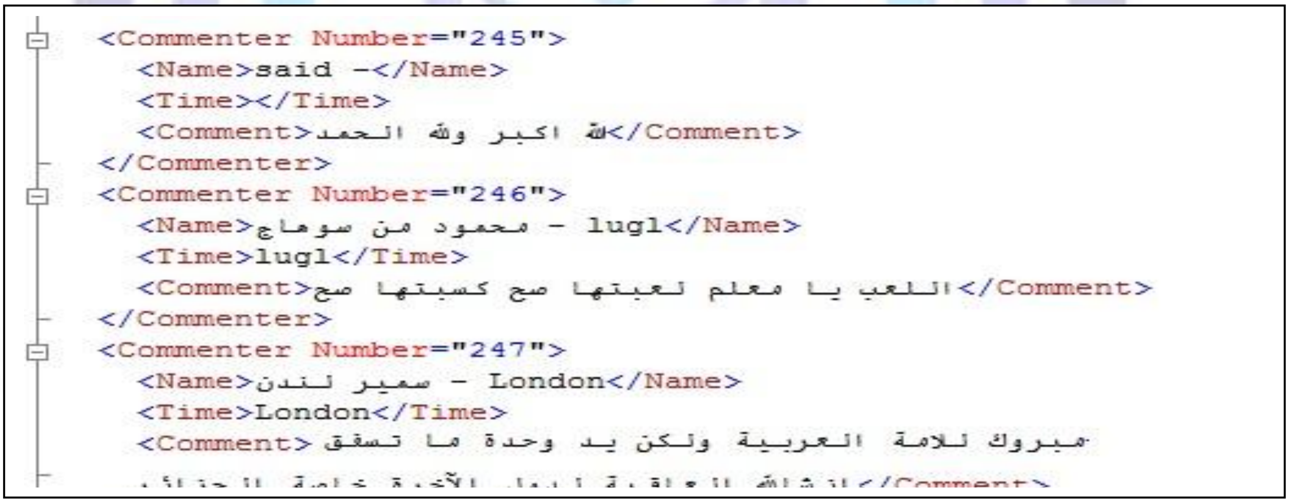
- *Direct Modern Comments:* Some commenter may use non lexicon idioms and words to create their comments; they write comments in slang Arabic language so the sentiment of the sentence will be understand by human analysis. For instance, instead of comment an item, the user may replay to another commenter, another comments may written sentimentally and subjectively in Arabic but with English characters, for instance the following comment: “*da ragel koyes ya gedaan w nsebo yakhod forsetoh*” which means “*This is a good man give him a chance*”. Comments are written sentimentally and subjectively in Arabic but with English characters and numbers, which is called Franco-Arab Language. Many Egyptians’ web users use numbers instead of Arabic characters like “3” to replace “*ع*” and “7” instead of “*ح*”. For instance, comment may be like this: *“mosh hayenfa3 yo7ko” which means “cannot govern”.* The system focuses on the comments written in Arabic language and ignores comments written in Franco-Arab language.

- *Indirect Comments:* Many comments in different web pages are not related to the topic. People attempt to comment on anything, even with unrelated words or nonsense. For example,

" نفس التمثلية حصلت مع الاخ المهندس عصام شرف و كانوا فرحنين بيه و هو بيحلف اليمن"

Which means “this is the same episode which happened with Eng. Essam Sharaf “

2) XML View: Comments are collected manually from different news websites, which are opinions of a specific topic (in this case: Egyptian presidential election in 2012). The system extracts the comments from the web pages and organizes the data in XML view to facilitate the manual analysis and system process. The system builds an XML, as shown in Fig. 2, schema for the data.



```
<Commenter Number="245">
  <Name>said -</Name>
  <Time></Time>
  <Comment>الحمد لله وﷲ اكبر</Comment>
</Commenter>
<Commenter Number="246">
  <Name>محمود من سوهاج - lugl</Name>
  <Time>lugl</Time>
  <Comment>اللعب يا معلم لعبتها صح كسبتها صح</Comment>
</Commenter>
<Commenter Number="247">
  <Name>سمير لندن - London</Name>
  <Time>London</Time>
  <Comment> مبروك للامة العربية ولكن يد وحدة ما تسقق
  انشالله العاقبة لدول الأخرة خاصة الجزائر</Comment>
```

Fig. 2. Samples of collected comments in XML view

DATA PREPROCESSING

REMOVE STOP WORDS:

A list of Arabic stop words is used to remove unwanted words to facilitate the data processing, where 613 words are used for removing unneeded words (يا ,من, الى, عن, فوق, ال). Removing unwanted words is a basic operation when mining unstructured data. For example, after removing stop words, the comment “ ربنا يقدرك يا محمد مرسي على خدمة مصر والف مبرووووك” will be “ربنا يقدرك محمد مرسي خدمة مصر والف مبرووووك”.

DATA AUTO CORRECTION:

As the web users write comments in free text, ungrammatically and unstructured Arabic language, comments always contains many syntax errors that make the mining process very difficult. As example of errors, web users may repeat a character in a word like “مبرووووووووك” instead of “مبروك”, which means “Congratulation”. Our system solves these problems to repack words to its correct syntax. After making data auto correction on the previous example, it will be “ربنا يقدرك محمد مرسي خدمة مصر والف مبروك”.

STEMMING:

Arabic stemmer is used to stem the yield words of each comment to get the words' root. We use stemmer of [28] to stem the words to return its root to extract features and identify satisfaction classification. The system tokenizes the comments into words and stems the words. The stemmer does not stem unknown words which may be new modern words or not stems the words not written well in syntax. After stemming the previous example it will be “ رب قدر محمد مرسي خدم مصر والف مبروك”.

DATA CLASSIFICATION:

Support Vector Machines is a classification technique that has been used in a variety of applications; a detailed explanation of SVM can be found in [33]; in the classification phase, three techniques are performed: classifying comments without applying SSWIL, classifying comments after the creation of SSWIL, and classifying comments using SSWIL only. All three are based on SVM classifier.

Comments classification Using Classical Lexicon without SSWIL: The system classifies the comments using SVM technique into two classes: satisfaction and dissatisfaction classes. The system uses a list of 613 satisfaction words to classify the comments as a satisfaction comments and a list of 700 dissatisfaction words to classify the comments as dissatisfaction comments. The comments of six news websites, as described previously and in Table 3, are used as a dataset to test the classification method. All the words in the two lists are lexicon words, which are formal Arabic language words without SSWIL. Table 4 illustrates the results of satisfaction, dissatisfaction, and outliers when applying the proposed sentiment analysis approach using classical lexicon. For example, having Youm7 news comments source, number of collected comments are 490 on Egyptian presidential election: 240 of them are satisfied, 81 are dissatisfied, 169 are outliers .

Table 3. Classic Classification Results

Comments Source	#of comments	Satisfaction	dissatisfaction	Outliers
Alarabia.com	472	298	92	82
Algazira.net	370	252	39	79
Youm7.com	490	240	81	169
Bbcarabic.com	23	15	4	4
Constitution Page	197	83	19	95
People’s Opinion page	294	187	41	66

Comments’ Classification using Classic Lexicon and SSWIL: The number of outliers’ comments is more than any of the two target classes as the web users use modern sentimental words, which does not belong to Arabic lexicon. After using the SSWIL as lists of sentimental words with SVM, we get better results than the first classification process, as shown in Table 5.

TABLE 5. sswiL with classical lexicon classification results

Comments Source	#of comments	Satisfaction	Dissatisfaction	Outliers
Alarabia.com	472	312	86	74
Algazira.net	370	324	19	27
Youm7.com	490	343	72	75
Bbcarabic.com	23	17	4	2
Constitution Page	197	97	18	82
People’s Opinion page	294	198	40	56

Classification Using SSWIL only: Applying SSWIL only in the classification process, it is noticed that the SSWIL affects positively the classification process, as shown in Table 6.

Table 6. SSWIL only classification results

Comments Source	#of comments	satisfaction	dissatisfaction	Outliers
Alarabia.com	472	68	0	404
Algazira.net	370	231	4	135
Youm7.com	490	261	11	218
Bbcarabic.com	23	8	0	15
Constitution Page	197	20	35	142
People's Opinion page	294	35	9	250

EXPERIMENTAL EVALUATION

Evaluating results, 1355 random comments are taken, applying the three types of classification; as seen in the three types of classification, web users write the comments with a new syntax, which affect the results of classification processes. Extraction techniques fail to extract the opinion words at the first classification type but it performs well at the second classification type after adding the SSWIL. Applying the proposed mining in all comments, the percent of classified comments in the first type (using classic lexicon) produce 75.35%, accuracy rate while the second classification type (using SSWIL with classic lexicon) produce 86.86%, as illustrated in Table 7. In addition, applying the system using SSWIL only, it gives 43.02% as a percent of comments classification and 56.98% not classified. The results are enhanced in the second type after applying SSWIL lists. In addition, in the first case, precision and recall give the highest rate in case of classifying comments with SSWIL, which are 88.63% and 78%, respectively, having F-measure of 82.9% , sensitivity of 78%, and specificity of 54.54%. This is due to the reason that microblogs users use both slang language mixed with classical Arabic.

Table 7. Evaluation results

Classification Type	Precision	Recall	F-Measure	Sensitivity	Specificity	Accuracy Rate
Classic Classification	82.4	59.33	68.98	59.33	68.85	75.35
SSWIL with Classic Lexicon Classification	88.63	78	82.97	78	54.54	86.86
SSWIL only Classification	83.07	36	50.23	36	88.54	43.02

CONCLUSION AND FUTURE WORK

In this paper, a sentiment analysis approach was proposed to mine unstructured and ungrammatical customers' Arabic slang comments based in new Arabic Slang Sentiment Words and Idioms Lexicon (SSWIL). The new lexicon was collected manually from micro blogs websites. In addition, SVM technique was applied with SSWIL to classify comments as satisfy or dissatisfy comments. In a future study, we will improve classification accuracy results of web users Arabic slang comments. In addition, SSWIL will be updated with new sentiment words and idioms and working on Franco-Arab comments.

REFERENCES

[1] Abdul-Mageed, M. and Diab, M. (2012) AWATIF: A Multi-Genre corpus for modern standard Arabic subjectivity and sentiment analysis, Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), European Language Resources Association (ELRA), pp. 3907-3914.

[2] Danah, M. and Ellison, N. B. (2007) Social network sites: definition, history, and scholarship, *Journal of Computer-Mediated Communication*, Vol. 13, Issue 1, pp. 210–230.

[3] Korayem, M., Crandall, D. and Abdul-Mageed M. (2012) Subjectivity and sentiment analysis of Arabic: a survey, Advanced Machine Learning Technologies and Applications, Communications in Computer and Information Science series 322, (Springer), AMLTA 2012.

[4] R.S, M. and Teresa, M. (2011) Bilingual experiments with an Arabic-english corpus for opinion mining, Proceedings of Recent Advances in Natural Language Processing, pp. 740–745, Bulgaria.

[5] http://visual.ly/facebook-users-arab-world-2013

[6] Helmy, T. and Daud, A. (2010) Intelligent agent for information extraction from Arabic text without machine translation, [C3LSW2010] Workshop on Cross-Cultural and Cross-Lingual Aspects of the Semantic Web Shanghai, China.

[7] El-Halees, A. (2011) Arabic opinion mining using combined classification approach, International Arab Conference on Information Technology (ACIT'2011), Riyadh, Saudi Arabia, 2011.

[8] AbdelRahman, S., Elarnaoty, M., Magdy, M. and Fahmy, A. (2010) Integrated machine learning techniques for Arabic named entity recognition, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No. 3.

[9] Elarnaoty, M., AbdelRahman, S. and Fahmy, A. (2012) A machine learning approach for opinion holder extraction in Arabic language, International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.2.

[10] Elhawary, M. and Elfeky, M. (2010) Mining Arabic business reviews, 2010 IEEE International Conference on Data Mining Workshops DOI 10.1109/ICDMW.2010.24

[11] Zaidan, O. and Callison-Burch, C. (2011) The Arabic Online Commentary Datasets: an Annotated dataset of informal arabic with high dialectal content, ACL (Short Papers), pp.37-41.

[12] Saad, M. and Ashour, W. (2010) Arabic Text Classification Using Decision Trees, Workshop on computer science and information technologies, Moscow – Saint-Petersburg, Russia.

[13] Hijjawi, M. and Bander, Z. (2011) An Arabic Stemming approach using machine learning with Arabic dialogue system, ICGST AIML-11 Conference, Dubai, UAE.

[14] Soliman, T. H.A., Mahmoud, M., A., Hedar, A. and Doss, M., M. (2013) Mining social networks' Arabic slang comments, Proceedings of IADIS European Conference on Data Mining (ECDM'13), Prague, Czech Republic, 2013.

[15] Shokry, A. M. (2013) Arabic sentence level sentiment analysis, M.Sc. Thesis, Computer science and engineering department, American University in Cairo.

[16] Hamouda, A. A. and El-taher, F. E. (2013) Sentiment Analyzer for Arabic Comments System, International Journal of Advanced Computer Science and Applications, Vol. 4, No.3, pp. 100-103.

[17] Barbosa, L., Feng, J. and Robust, J. (2010) Sentiment detection on twitter from biased and noisy data, Proceedings of COLING, pp. 36–44, 2010.

[18] Speriosu, M., Sudan, N., Upadhyay, S., and Baldridge, J., (2011) Twitter polarity classification with label propagation over lexical links and the follower graph. *Proceedings of the EMNLP First workshop on Unsupervised Learning in NLP,* pp.53–63.

[19] Bifet, A. and Frank, E. (2010) Sentiment knowledge discovery in twitter streaming data, Discovery Science (2010), Springer, pp. 1–15.

[20] Tumasjan, A., Sprenger, A., Sandner, T. and Welpe, I. (2010) Predicting elections with twitter: What 140 characters reveal about political sentiment, Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, pp. 178–185.

[21] Diakopoulos, N. and Shamma, D. (2010) Characterizing debate performance via aggregated twitter sentiment. Proceedings of the 28th international conference on Human factors in computing systems, ACM, pp. 1195–1198, 2010.

[22] Conover, M. et al. (2011) Political polarization on twitter, Proc. 5th intl. conference on weblogs and social media, 2011.

[23] Livne, A., Simmons, M., Adar, E. and Adamic, L. (2011) The party is over here: Structure and content in the 2010 election, Fifth International AAAI Conference on Weblogs and Social Media.

[24] Abdul-Mageed, M., Kuebler, S. and Diab, M. (2012) SAMAR: A System for Subjectivity and Sentiment Analysis of Arabic Social Media, Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), Republic of Korea.

[25] Abdul-Mageed, M. and Diab, M. (2011) Subjectivity and sentiment annotation of modern standard Arabic newswire, Proceedings of the 5th Linguistic Annotation Workshop, pp. 110–118.

[26] Abdul-Mageed, M. and Diab, M. (2012) Toward building a large-scale Arabic sentiment lexicon, Proceedings of the 6th International Global Word-Net Conference, Matsue, Japan.

[27] Rushdi-Saleh, M., Martın-Valdivia, M., Urena-L´opez, L. and Perea-Ortega, J.. (2011) Oca: Opinion corpus for Arabic, Journal of the American Society for Information Science and Technology, 62(10): pp.2045–2054.

[28] Elbeltagy, S., R. and Reafea, A. (2011) An accuracy-enhanced light stemmer for Arabic text, ACM Transactions on Speech and Language Processing (TSLP), Vol. 7, Issue 2, No. 2.

[29] Yu, H. and Kim, S. (2012) SVM Tutorial: Classification, Regression and Ranking, Handbook of Natural Computing, Springer Link, 2012, pp. 479-506.