Semantic Feature Based Arabic Opinion Mining Using Ontology

Abdullah M. Alkadri[1], Abeer M. ElKorany[2]
Computer Science Department
Faculty of Computers and Information,Cairo University
Cairo, Egypt

***Abstract*—with the increase of opinionated reviews on the web, automatically analyzing and extracting knowledge from those reviews is very important. However, it is a challenging task to be done manually. Opinion mining is a text mining discipline that automatically performs such a task. Most researches done in this field were focused on English texts with very limited researches on Arabic language. This scarcity is because there are a lot of obstacles in Arabic. The aim of this paper is to develop a novel semantic feature-based opinion mining framework for Arabic reviews. This framework utilizes the semantic of ontologies and lexicons in the identification of opinion features and their polarity. Experiments showed that the proposed framework achieved a good level of performance compared with manually collected test data.**

***Keywords*—Opinion Mining; Sentimental Analysis; Ontology; Feature extraction; Polarity identification;**

I. Introduction

As a result of dramatically increase of using the internet in the recent years; huge information of people opinions was produced on the web, people can post their views using Internet forums, discussion groups, product reviews and blogs. Analyzing this information manually is time consuming and maybe impossible mission. For example, if we wanted to judge the article or product positively or negatively according to the comments of the people, it is difficult to read all comments and classify them manually, so we need an automated technique to do such a task. Opinion mining is the appropriate way which automatically extracts knowledge from people comments.

Opinion mining (also called sentiment analysis, sentiment mining, sentiment classification, subjectivity analysis, and review mining or appraisal extraction) is a subtopic of text mining that it automatically extracts opinions, sentiments, and subjectivity from user-generated reviews [1]. Basic task in opinion mining is to determine the subjectivity, polarity (positive or negative) of a piece of text in other words: What is the opinion of the writer. Opinion mining has a wide range of applications from different domains such as commercial, governmental, political, educational and others [2].

Nowadays, there are three levels of Sentiment Classification in Opinion mining (document, sentence and feature). According to [3], the sentence and document level analyses do not discover what exactly people liked or not. However, studying the opinion text, especially feature level, is extremely challenging. For the ordinary user, it is too complex to analyze opinions about object and object features in the online review sites on the Web. To do such analysis it is necessary to perform some kind of opinion mining, feature-based opinion mining so as to identify the features in the review and classify the sentiments of the opinion for each of these features [4]. The feature-based opinion mining of object reviews is a difficult task, owing to both the high semantic variability of the opinions expressed, and the diversity of the features and sub- features that describe the products and the polarity of opinion words used to depict them [5]. In the last few years, new approaches based on both semantic web technologies and domain-dependent corpora for feature-based opinion mining have appeared [6]. In [7] Isidro et el. Believe that the already mature Semantic Web technology could be a valuable addition to traditional opinion mining approaches. More concretely, ontologies constitute the standard knowledge representation mechanism for the Semantic Web and can be used to structure information. The formal semantics underlying ontology languages enables the automatic processing of the information in ontologies and allows the use of semantic reasoners to infer new knowledge.

In the proposed work, an ontology is viewed as a formal and explicit specification of a shared conceptualization [8]. Ontologies provide a structured knowledge representation and a common vocabulary for a domain (e.g. hotel domain). In this work, the Web Ontology Language (OWL), the W3C standard used to represent ontologies in the Semantic Web, has been used to represent the concepts and features of the application domain (in our case hotel domain). The main contribution of the proposed framework is how to classify Arabic views of people about an entity (Object) in a specific domain to positive or negative opinions. We need a point of view about an entity through extracting the view about its features (attributes). For example, if the entity is a hotel its features will be a room, bar, lunch and so on. Most of researches done in this field were focused on English texts with very limited researches in an Arabic language. Limitation of research work in this area is due to the following reasons:

- Content found on Forums and Blogs is written in many forms of Arabic Dialect which makes the task of using a semantic approach for mining opinions very challenging. Moreover, the majority of the available preprocessing tools are mainly built for the modern standard Arabic.
- The limitation in availability of appropriate datasets, no opinion-related (or sentiment) Arabic Lexicon is present to assist in the task of measuring the polarity

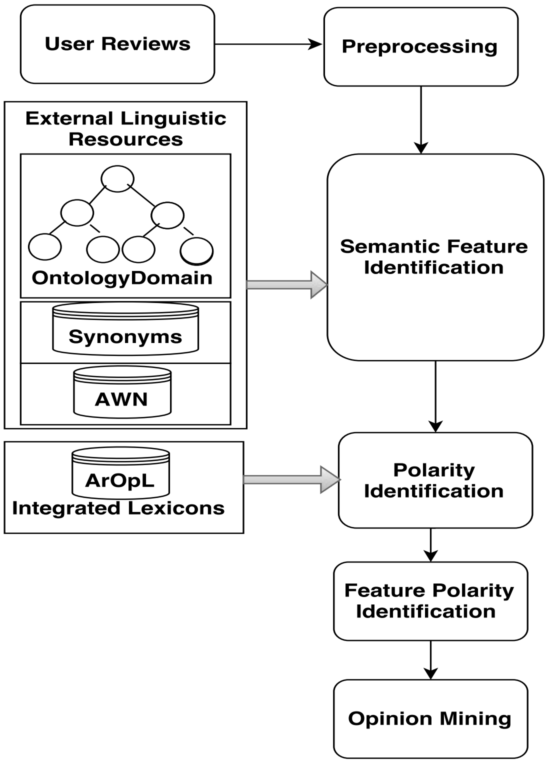

Fig. 2: Proposed system architecture.

- Third, compare the nouns of reviews with the synonyms in AWN [35]. If the noun does not exist, we ignore it.

Table I provides an example of feature identification from a review. According to the example in this table, the identified features within the given review will be (شامبو، فوطه، الغرفه، دورة المياه، السرير، التلفزيون ، الانترنت). More concretely, "الغرفه"، "شامبو"، and "السرير ، التلفزيون" are identified as subclasses in the ontology classes "مرافق الغرفة"، "منتجات نظافة"، "غرفة الفندق" respectively. "فوطه" and "دورة المياه" are identified as features because they are synonymous of the "مناشف" and "حمام" classes from synonyms dictionary and AWN respectively.

C. Polarity Identification

For polarity identification, a list of opinion words is essential, i.e., an opinion lexicon. Opinion words are words that express positive or negative sentiments. Thus, we used the developed new large lexicon (ArOpL). During this phase we ignore the selected features which we already extracted in section IV-B. The polarity is determined by aggregating the polarity of the extracted words in reviews based on our new dictionary. In other words, for each review our method assigns the scores +1 and -1 to the positive and negative words respectively.

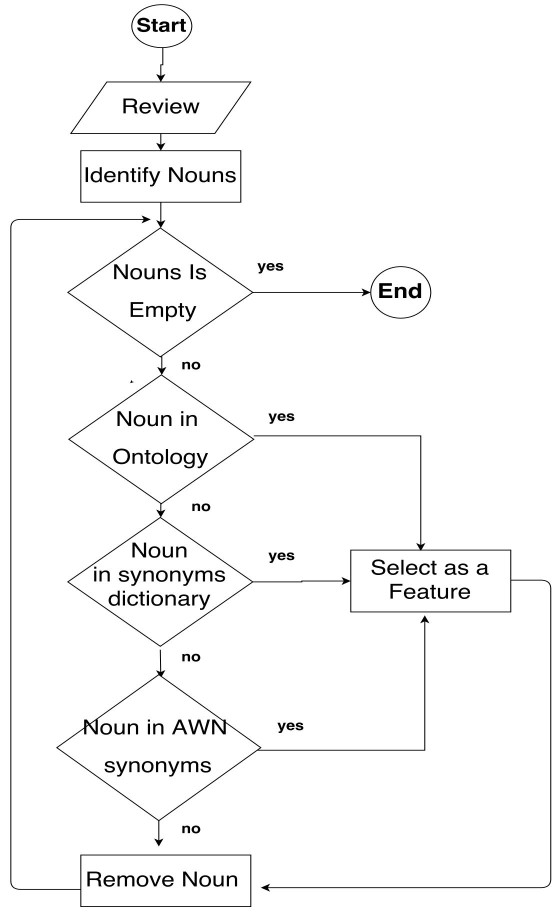

Fig. 3: Semantic Feature Identification Process.

TABLE I: Example of Feature Identification

D. Feature Polarity Identification (Tuple Generation)

Using the extracted features and the lists of positive and negative words generated by previous phases, we identify the opinion orientation expressed for each feature. This step generates a set of tuples containing features and their polarities. In order to generate these tuples, it is necessary to obtain the words from around the feature. The words that are close to the feature can be obtained in different ways. The following three methods have been implemented to evaluate our solution:

- N-GRAM Before: this method obtains the N-GRAM words before the feature in the users review.
- N-GRAM After: this method obtains the N-GRAM words after the feature in the users review.
- N-GRAM Around: this method obtains the N-GRAM words before the feature and the N-GRAM words after the feature in the users review.

N-GRAM indicates the number of words near the feature that are to be selected in the polarity identification process. Also a Negative Words have been handled, the negation word usually reverses the polarity of the word in the sentence. Our proposed technique recognizes Negation words such as ”غير ، ليس ، لا” and then reverses the opinion orientation. For example, the sentence, ”الفندق غير جميل” conforms to the negation word ”غير” then it is assigned the negative orientation, although ”جميل” is a positive word.

E. Opinion Mining

The global polarity of a review is obtained by determining the majority of polarized features which our system already identified. If the major features are polarized as positive, then the global polarity is considered positive. Likewise, if the major features are polarized as negative, then it considered as negative. Otherwise the global polarity is considered as neutral.

V. EXPERIMENT AND DISCUSSION

A. Experiment setup

Since there was no tagged data in Arabic hotel domain, we collected the test reviews manually from a variety of related websites which have relevant to hotel domain. We crawled these reviews from different countries and three websites (www.tripadvisor.com), (http://www.booking.com) and (http://www.agoda.com) to find data related with the hotel domain. The total numbers of reviews that have been used are 890 reviews, 690 reviews were used for the sake of ontology extension and the rest 200 reviews, half of them are negative and the rest are positive were used for experiments. In order to measure the effectivnesss of the proposed model.

We manually tagged the reviews to collect the baseline results used to evaluate our proposed method. The details of the manual tagging are described as follows: The reviews have been shown to three educated annotators. They read the reviews and identified all features and associated polarities. According to features polarities they are classified into three categories: positive (1), negative (-1) and neutral (0). An example of the manual tagging is shown in Table II. Finally, the manual results and the output produced by our system are compared with each other. The following experiments have been conducted:

TABLE II: Example of Manual Tagging

Review	Manual Tagging
الفندق هادي ونظيف والاضاءة كانت جميلة	الفندق#هادي#1 الفندق#نظيف#1 الاضاءة#جميلة#1
الانترنت غير مجاني الفطور سئ	الانترنت #غير مجاني#-1 الفطور# سئ#-1

TABLE III: Feature identification accuracy

reviews	manually tagging	Proposed System Output	accuracy
positive	882	763	0.865
negative	757	655	0.866

1) Experiment1 measuring the feature identification accuracy: The aim of this experiment is to measure the accuracy of the correctly identified features (feature accuracy) using the manually tagged reviews. We take the labeled features as baseline in contrast to the results from the proposed method which described in section IV-B to obtain the number of features correctly identified feature accuracy. The results of this experiment are shown in Table III.

2) Experiment 2 measuring the Feature Polarity Identification and global polarity accuracy: During this experiment, two levels of accuracy measurements were tackled. On the first level, feature polarity identification where we applied different value of N-GRAM to identify correct polarity to correct feature. While the second level aims to measure accuracy of complete review. The three N-GRAM methods, which explained before, are used to compare the manual results and the output produced to obtain the number of features correctly classified (feature polarity identification accuracy) and the review global polarity (opinion mining accuracy). Different values for the N-GRAM parameter (between 1 and 4) have been used to discover the best result as presented in Table IV.

B. Discussion

The results are shown in Table IV have been divided into two different categories: the average accuracy of both feature polarity identification and opinion mining classification of the entire document for both positive and negative reviews. The results of the N-GRAM Around method for the opinion mining are shown in Table IV. As shown in the table, the best average success rate for the feature polarity identification process is obtained with N-GRAM = 4 with an accuracy of 72% in

TABLE IV: Feature Polarity Identification and Opinion Mining accuracy

		average feature polarity identification accuracy			opinion mining accuracy		
Method	Ngram	Positive Reviews	Negative Reviews	average	Positive Reviews	Negative Reviews	average
After	1	0.45	0.38	0.415	0.76	0.3	0.53
	2	0.56	0.48	0.52	0.94	0.71	0.825
	3	0.63	0.53	0.58	0.98	0.82	0.9
	4	0.66	0.56	0.61	0.99	0.85	0.92
Around	1	0.54	0.45	0.495	0.86	0.55	0.705
	2	0.65	0.57	0.61	0.99	0.88	0.935
	3	0.70	0.61	0.655	0.98	0.92	0.95
	4	**0.72**	**0.63**	**0.675**	**1**	**0.91**	**0.955**
Before	1	0.29	0.23	0.26	0.22	0.09	0.155
	2	0.38	0.33	0.355	0.55	0.22	0.385
	3	0.45	0.40	0.425	0.76	0.37	0.565
	4	0.50	0.44	0.47	0.86	0.59	0.725

the Positive Reviews and 63% in the Negative Reviews. This means that the feature based polarity calculated using 4 words before the feature and 4 words after the feature in the users review has achieved good accuracy. In fact, the worst results are obtained with N-GRAM = 1 with an accuracy of 54 in the Positive Reviews and 45% in the Negative Reviews. The best results for Opinion mining are obtained with N-GRAM=4 with an accuracy of 100% in the Positive Reviews and 91% in the Negative Reviews.

Table IV also shows the results obtained when using the N-GRAM After method. At first sight it will be noted that these results are worse than those obtained with the N-GRAM Around method. Here, the best average success rate for the feature polarity identification process is also obtained with N-GRAM = 4 with an accuracy of 66% in the Positive Reviews and 59% in the Negative Reviews. This means that the feature based polarity calculated using the next 4 words of the feature identified obtains good results. The best results for Opinion mining are obtained with N-GRAM=4 with an accuracy of 99% in the Positive Reviews and 85% in the Negative Reviews.

The results obtained with N-GRAM Before method are worse than those obtained with the ”N-GRAM After method. More concretely, the best average success rate for the feature polarity identification process is also obtained with N-GRAM = 4 with a maximum accuracy of 50% in the Positive Reviews and 44% in the Negative Reviews. This means that the feature based polarity calculated using 4 words before of the feature identified obtains good results. The best results for Opinion mining are obtained with N-GRAM=4 with an accuracy of 86% in the Positive Reviews and 59% in the Negative Reviews.

Of the three proposed methods, the N-GRAM Around method is that which achieves the best results for both the feature polarity identification process and the Opining mining of users opinions in the Arabic language, obtaining accuracies of 67.5% and 95.5% in all Reviews, respectively.

VI. CONCLUSION AND FUTURE WORK

Arabic opinion mining is a challenging problem. It is concerned with analyzing the opinions that appear in users reviews, and determine whether these opinions are positive or negative. In this paper, a new methodology is proposed for feature-based Arabic opinion Mining. This approach is going through five different stages: Ontology and lexicon Development, Semantic Feature Identification, Polarity Identification, Feature Polarity Identification and Opinion Mining.

The main contributions of this work are: First, an ontology and lexicon development. Second, ontology-based feature identification, finally, three different configurable N-GRAM methods for feature polarity identification are proposed. These methods can be configured with different parameters to obtain the best polarity identification approach.

In spite of all the advantages and possibilities of the proposed method, it has several limitations that could be improved in the future. First, the proposed approach can be improved by incorporating opinion mining techniques based on machine learning. Second, since the current ontology is static and knowledge represented in it is not enough, it would be interesting to construct a semi-automatic ontology based on ontology learning techniques from the users reviews. Finally, we plan to apply the proposed approach in another domain such as product reviews.

REFERENCES

[1] A. Harb, M. Plantié, G. Dray, M. Roche, F. Trousset, and P. Poncelet, “Web opinion mining: How to extract opinions from blogs?” in *Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology*. ACM, 2008, pp. 211–217.

[2] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and trends in information retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.

[3] B. Liu, “Sentiment analysis and subjectivity.” *Handbook of natural language processing*, vol. 2, pp. 627–666, 2010.

[4] R. Feldman, “Techniques and applications for sentiment analysis,” *Communications of the ACM*, vol. 56, no. 4, pp. 82–89, 2013.

[5] A. Balahur and A. Montoyo, “Semantic approaches to fine and coarse-grained feature-based opinion mining,” in *Natural language processing and information systems*. Springer, 2009, pp. 142–153.

[6] E. Cambria, B. Schuller, B. Liu, H. Wang, and C. Havasi, “Knowledge-based approaches to concept-level sentiment analysis,” *IEEE Intelligent Systems*, no. 2, pp. 12–14, 2013.

[7] I. Peñalver-Martinez, F. Garcia-Sanchez, R. Valencia-Garcia, M. Á. Rodríguez-García, V. Moreno, A. Fraga, and J. L. Sánchez-Cervantes, “Feature-based opinion mining through ontologies,” *Expert Systems with Applications*, vol. 41, no. 13, pp. 5995–6008, 2014.

[8] R. Studer, V. R. Benjamins, and D. Fensel, “Knowledge engineering: principles and methods,” *Data & knowledge engineering*, vol. 25, no. 1, pp. 161–197, 1998.

[9] S. A. Morsy, “Recognizing contextual valence shifters in document-level sentiment classification,” 2011.

[10] A. M. Azmi and S. M. Alzanin, “Aara–a system for mining the polarity of saudi public opinion through e-newspaper comments,” *Journal of Information Science*, vol. 40, no. 3, pp. 398–410, 2014.

[11] J. B. Salamah and A. Elkhlifi, “Microblogging opinion mining approach for kuwaiti dialect,” in *The International Conference on Computing Technology and Information Management (ICCTIM2014)*. The Society of Digital Information and Wireless Communication, 2014, pp. 388–396.

[12] S. B. Hamouda and J. Akaichi, “Social networks text mining for sentiment classification: The case of facebookstatuses updates in the arabic springera,” *International Journal Application or Innovation in Engineering and Management*, vol. 2, no. 5, pp. 470–478, 2013.

[13] M. Abdul-Mageed, M. Diab, and S. Kübler, “Samar: Subjectivity and sentiment analysis for arabic social media,” *Computer Speech & Language*, vol. 28, no. 1, pp. 20–37, 2014.

[14] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub, “Arabic sentiment analysis: Lexicon-based and corpus-based,” in *Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on*. IEEE, 2013, pp. 1–6.

[15] M. Elhawary and M. Elfeky, “Mining arabic business reviews,” in *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1108–1113.

[16] A. El-Halees, “Arabic opinion mining using combined classification approach,” 2011.

[17] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004, pp. 168–177.

[18] M. N. Al-Kabi, I. M. Alsmadi, A. H. Gigieh, H. A. Wahsheh, and M. M. Haidar, “Opinion mining and analysis for arabic language,” *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 5, no. 5, pp. 181–195, 2014.

[19] X. Ding, B. Liu, and P. S. Yu, “A holistic lexicon-based approach to opinion mining,” in *Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM, 2008, pp. 231–240.

[20] A. Muangon, S. Thammaboosadee, and C. Haruechaiyasak, “A lexiconizing framework of feature-based opinion mining in tourism industry,” in *Digital Information and Communication Technology and it’s Applications (DICTAP), 2014 Fourth International Conference on*. IEEE, 2014, pp. 169–173.

[21] E. Bjørkelund, T. H. Burnett, and K. Nørvåg, “A study of opinion mining and visualization of hotel reviews,” in *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services*. ACM, 2012, pp. 229–238.

[22] L. Zhao and C. Li, *Ontology based opinion mining for movie reviews*. Springer, 2009.

[23] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, “Ontology-based sentiment analysis of twitter posts,” *Expert systems with applications*, vol. 40, no. 10, pp. 4065–4074, 2013.

[24] J. M. Ruiz-Martínez, R. Valencia-García, and F. García-Sánchez, “Semantic-based sentiment analysis in financial news,” in *International Workshop on Finance and Economics on the Semantic Web (FEOSW 2012)*, 2012, p. 38.

[25] L. A. Freitas and R. Vieira, “Ontology based feature level opinion mining for portuguese reviews,” in *Proceedings of the 22nd international conference on World Wide Web companion*. International World Wide Web Conferences Steering Committee, 2013, pp. 367–370.

[26] M. S. Chaves, L. A. de Freitas, and R. Vieira, “Hontology: A multilingual ontology for the accommodation sector in the tourism industry.” in *KEOD*, J. Filipe and J. L. G. Dietz, Eds. SciTePress, 2012, pp. 149–154.

[27] H. Knublauch, R. W. Fergerson, N. F. Noy, and M. A. Musen, “The protégé owl plugin: An open development environment for semantic web applications,” in *The Semantic Web–ISWC 2004*. Springer, 2004, pp. 229–243.

[28] S. Baccianella, A. Esuli, and F. Sebastiani, “Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining.” in *LREC*, vol. 10, 2010, pp. 2200–2204.

[29] R. Mihalcea, C. Banea, and J. M. Wiebe, “Learning multilingual subjective language via cross-lingual projections,” 2007.

[30] M. Abdul-Mageed, M. T. Diab, and M. Korayem, “Subjectivity and sentiment analysis of modern standard arabic,” in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, 2011, pp. 587–591.

[31] A. Mourad and K. Darwish, “Subjectivity and sentiment analysis of modern standard arabic and arabic microblogs,” in *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*, 2013, pp. 55–64.

[32] A. M. Shoukry, “Arabic sentence level sentiment analysis,” Ph.D. dissertation, The American University in Cairo, 2013.

[33] M. Althobaiti, U. Kruschwitz, and M. Poesio, “Automatic Creation of Arabic Named Entity Annotated Corpus Using Wikipedia,” in *Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Gothenburg, 2014, pp. 106–115.

[34] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, “Feature-rich part-of-speech tagging with a cyclic dependency network,” in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 2003, pp. 173–180.

[35] L. Abouenour, K. Bouzoubaa, and P. Rosso, “On the evaluation and improvement of arabic wordnet coverage and usability,” *Language resources and evaluation*, vol. 47, no. 3, pp. 891–917, 2013.