The 13th International Arab Conference on Information Technology ACIT'2012 Dec.10-13
ISSN : 1812-0857

OPINION MINING FROM ARABIC COMPARATIVE SENTENCES

Alaa El-Halees

Faculty of Information Technology, Islamic University of Gaza, Gaza, Palestine
alhalees@iugaza.edu.ps

ABSTRACT

This paper discuses the problem of identifying comparative opinion sentences in Arabic text. Most works in the field of opinion mining concentrate on extracting knowledge from direct opinions. Directly mining positive or negative opinions on a product review or its features is only one form of opinion mining; comparing a product review with some other competitive products is another form. Comparisons focus on mining opinions from comparative sentences, i.e., to determine which entities in products are preferred by its author. There are some work in this area in English language. This is the first in Arabic.

Mining from comparative text can be divided into three tasks,. The first task is to identify comparative statement from non-comparative ones. In this task, we used method that depends on linguistic classification, where we got f-measure of 63.73 %. Then we used three machine learning methods where we got better performance which is about 86.63% in the best case. Finally, for this task, we used combined approach of linguistic and machine learning where we got f-measure of 88.87%. In the second task we generated a set of rules to characterize three types of comparative statements. We left the third task for future work

KEYWORDS: Arabic comparative sentences, sentiment classification, Arabic Opinion Mining
.

1. INTRODUCTION

Nowadays with the availability of Web 2.0 that allows users to interact and collaborate with each others, public increasingly participate to express their opinions on the Internet. They can now post their views using Internet forums, discussion groups, product reviews, social networks, and blogs which are collectively called user-generated contents [1]. Manually extracting opinions from user-generated continents is time consuming. A better alternative is to use opinion mining. Opinion mining is a subtopic of text mining that it automatically extracts opinions, sentiments, and subjectivity from user-generated contents [2].

Most works in the field of opinion mining concentrate on extracting knowledge from direct opinions. Directly mining positive or negative opinions on an object or its features is only one form of opinion mining; comparing an object with some other similar objects is another form. Comparisons focus on mining opinions from comparative sentences, i.e., to determine which entities in a comparison are preferred by its author. A typical comparative sentence compares two or more entities. For example, the sentence, "the picture quality of Camera X is better than that of Camera Y", compares two entities "Camera X" and "Camera Y" with regard to their picture quality [3] [4].

Extracting comparative sentences from text is useful for many applications. For example, in the business environment, whenever a new product comes into market, the product manufacturer wants to know consumer opinions on the product, and how the product compares with those of its competitors [5]. Comparative opinion mining not only can be used in business, but also in many fields like education, politics, sports and others. For example, in education, it can use students' opinion to compare performance of different teachers teaching the same course.

In English, mining comparative opinions is a non-trivial task due to the large amount of reviews and their informal style. There are some publications in comparative opinions of English text such as work of [3] and [5]. Also, there are some other works in other languages such as in Chinese [6] and Korean [4]. In Arabic this is the first. Arabic, in general, is a challenging language because it has a very complex morphology as compare to English. This is due to the unique nature of Arabic morphological principles which is highly inflectional and derivational [7]. In case of mining from user-generated contains, it brings another complexity since most writer in the web express their opinion using local accent instead of standard Arabic language. So, we end up with many written accents instead of one formal language. Also, many times writers misspelled the words either by accident or deliberately (e.g. for short, repeating letters to insistent in some words, …etc).

Mining from comparative text can be divided to three task. First task is to identify comparative statement from non-comparative ones. This includes identifying subjective or objective comparative statements. The second task is to find the type of comparative statements, since there are at least four types. The third task is to extract comparative relations from the identified sentences [5]. This involves the extraction of entities and their features that are being compared, and the comparative keywords. This paper will concentrate on the first two tasks.

The rest of the paper is structured as follows: section two discusses related works, section three about identifying comparative sentences, section four about identifying comparative type, section five describes the conducted experiments, section six gives the results of experiments and section seven concludes the paper.

2. RELATED WORK

The first work in this area was that of Jindal and Liu in [8]. They studied the problem of identifying comparative sentences in English text. They classified comparative sentences into different categories based on linguistic perspective. Then they proposed an approach based on pattern discovery and supervised learning to identify comparative sentences. Their technique based on keyword strategy to achieve a high recall, and then build a machine learning model to automatically classify each sentence into one of the two classes, "comparative" and "non-comparative", based on the filtered data to improve the precision. Similar work was done by Yang and Ko in [9] but for Korean text instead of English.

Jinal and Liu in [5] continued their work in mining comparative sentences. They extracted comparative relations from the identified sentences. It involved the extraction of entities and their features that are being compared, and comparative keywords. The relation is expressed with (<relationWord>, <features>, <entityS1>, <entityS2>), where: relationWord: The comparative keyword used to express a comparative relation in a sentence; features: a set of features being compared; entityS1 and entityS2: Sets of entities being compared. Entities in entity S1 appear to the left of the relation word and entities in entityS2 appear to the right of the relation word. For this task they used new type of rules called label sequential rules.

However, the most recent work in this area is the work published by Xu et.al. in [9]. They used comparative opinion mining in business intelligence. They proposed a novel graphical model to extract and visualize comparative relations between products from customer reviews, with the interdependencies among relations taken into consideration, to help enterprises discover potential risks and further design new products and marketing strategies.

3. IDENTIFYING COMPARATIVE SENTENCES

As defined by [8], a comparative sentence is a sentence that expresses a relation based on similarities or differences of more than one object. Therefore; the first task in mining comparative statements is to identify comparative statement from non-comparative ones. To do that we used two approaches, linguistic approach and machine learning approach. .

3.1 LINGUISTIC CLASSIFICATION

In linguistic classification, the sentence is classified based linguistic properties . In this research we used that most popular one with is Part of Speech (POS) of the words contain in the sentence . In this case, if the sentence has word of POS tag JJR (Adjective Comparative), JJS (Adjective Superlative), RBR (Adverb Comparative), or RBS (Adverb Superlative), then it is usually express direct comparisons. The disadvantage of this approach is that some statements are comparative but they do not use the comparative words such as " الكمبيوتر المكتبي يتحمل شغل عن اللاب توب " means (Desktop bear service than laptop) which means "Desktop has better service than laptop". On the other hand, some statements have a comparative words but they are not comparative statements such as " سيارة أ أعلى من سيارة ب بمتر" (Car A is higher than Car B by one meter). That does not imply that Car A is better or worse that Car B.

3.2 MACHINE LEARNING APPROACH

To overcome the limitations of linguistic classification, we used machine learning classification methods with linguistic properties such as normalization, stemming and POS . We used three well known supervised learning methods which are: Naïve Bayes, K-nearest neighbors and Support Vector Machine to predict comparative statements given a trained data which contain comparative and non-comparative statements. The following is the description of the methods:

Naïve Bayes classifiers are widely used because of their simplicity and computational efficiency. It uses training methods consisting of relative-frequency estimation of words in a document as words probabilities and uses these probabilities to assign a category to the document. To estimate the term $P(r \mid c)$ where r is the review and c is the class (comparative and non-comparative). Naïve Bayes decomposes it by assuming the features are conditionally independent [10].

k-Nearest Neighbor is a method to classify documents. In the training phase, documents have to be indexed and converted to vector representation. To classify new review r; the similarly of its review vector to each review vector in the training set has to be computed. Then its k nearest neighbor is determined by measuring similarity which may be measured by, for example, the Euclidean distance [11].

Support Vector Machine is a learning algorithm proposed by [12]. In its simplest linear form, it is a hyperplane that separates a set of positive examples from a set of negative examples with maximum margin. Test reviews are classified according to their positions with respect to the hyperplanes.

4. IDENTIFYING COMPARATIVE TYPES

The second task in this work is to identify comparative sentence type. As described by [3], comparative statements can be classified into four main types:

1- *Non-equal gradable comparisons:* Relations of the type greater or less than that express an ordering of objects with regard to some of their features. In Arabic the word in balance of "أفعل", if its verb consist of three letters such as " دراسة أوراكل أعمق من مايكروسوفت" (Studding Oracle is deeper than Microsoft). If the verb contain more than three letters the statement will contain the word "اقل او اكثر" (more or less) such as " سعيد أكثر اجتهادًا من اخيه" (Sai'd is more assiduity than his brother).

2- *Equative comparisons*: Relations of the type equal to that state two objects are equal with respect to some of their features. For instance, in Arabic "الجامعتين نفس المستوى في التعليم" (The two universities are in the same level of education).

3- *Superlative comparisons*: Relations of the type greater or less than all others that rank one object over all others, e.g..in Arabic it adds the "ال" to the comparison word " الاهلي المصري الافضل في التاريخ" (Egyptian Al-Ahly is the best through history).

4- *Non-gradable comparisons*: Sentences that compare features of two or more objects, but do not grade them. There are three main types : Object A is similar to or different from object B with regard to some features, e.g., " تدريس الدكتور رشدي يختلف عن تدريس الدكتور عيسى". (Teaching of Dr. Roshdy is different than of Dr. Issa). Object A has feature f1, and object B has feature f2 (f1 and f2 are usually substitutable), e.g.,." الكمبيوتر المكتبي يستخدم سماعات خارجيه اما اللاب توب يستخدم سماعات داخلية" (desktop PCs use external speakers but laptops use internal speakers). Object A has feature f, but object B does not have, e.g., " جوال أ يستخدم سماعات اذن و جوال ب لا يستخدم". (phone A has an earphone, but cell phone B does not have) .

In this paper we will try to generate rules that can cover each comparative type.

5.0 EXPERIMENTAL SETUP

To evaluate our approaches, a set of experiments was designed and conducted. In this section we describe experiments' design including the corpus, the preprocessing stage and evaluation metrics.

5.1 CORPUS

We collected documents related to opinions expressed in Arabic from three different domains: "Education", "Technology" and "Sports". In each domain, the documents were manually grouped into two types comparative statements and non-comparative statements.

As depicted in table 5.1, we used total of 1048 posts contain 435 comparative statement and 613 non-comparative statements.

	Number of statements	
Domain	**comparative**	**Non-comparative**
Education	**104**	**186**
Technology	**115**	**202**
Sports	**216**	**225**
Total	**435**	**613**

Table 5.1: Description Of Corpus Used In The Experiments

5.2 PREPROCESSING

After we collected the data associated with the three domains, we striped out the HTML tags and non-textual contents. Then, we separated the documents into posts and converted each post into a single file. For Arabic scripts, some alphabets have been normalized (e.g. the letters which have more than one form) and some repeated letters have been cancelled (that happens in discussion when the user wants to insist on some words), some of the wrong spelling words are corrected. After that, the sentences are tokenized, stop words removed and Arabic light stemmer applied. We used Stanford POS tagger for Arabic language [13] to give part of speech tag for each word in the text. Then, we obtained vector representations for the terms from their textual representations by performing TFIDF (Term Frequency–Inverse Document Frequency) weight which is a well known weight presentation of terms often used in text mining [14]. We also removed some terms with a low frequency of occurrence.

5.3 EVALUATION METRICS

There are various methods to determine effectiveness; however, accuracy, precision and recall are the most common in this field. Accuracy measures the percentage of the test set that the classifier has labeled correctly. Furthermore, the precision and recall are calculated. Precision is the percentage of predicted documents class that are correctly classified. Recall is the percentage of the total documents for the given class that are correctly classified. We also computed the F-measure, a combined metric that takes both precision and recall into consideration [15]. To apply the test, we used 10 cross validation testing

6.0 RESULTS

Following are the results for the two tasks.

6.1 RESULTS FOR IDENTIFYING COMPARATIVE SENTENCES

We tested three approaches to identify comparative sentences. The first approach is to use only POS. In this approach we considered a statement as comparative statements if it has one of the comparative tags which are: JJR, JJS, RBR or RBS. Table 6.1 gives the results.

Domain	Precision	Recall	F-Measure
Education	31.3	79.3	44.88
Sports	73.8	88.57	80.51
Technology	59.4	73.8	65.81
Average	54.83	80.55	63.73

Table 6.1: Identifying Arabic comparative sentences using POS tag

From the table we notice that in general there is an acceptable recall for identifying comparative statements using only POS. However, the precision is low. This result, also, applied in English text as recorded by [8]. In general the f-measure can be improved.

Using only machine learning, we got figure 6.1 for three classifiers K-nearest neighbors , Support Vector Machine and Naïve Bayes.

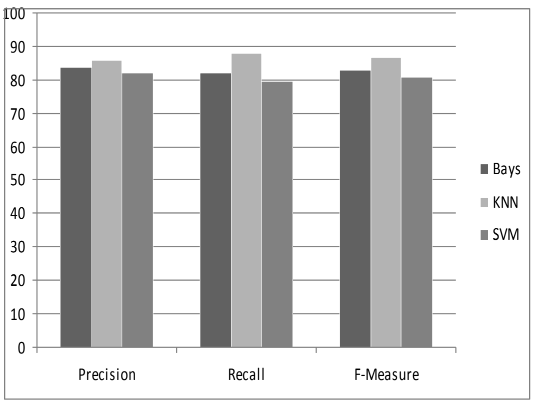

Figure 6.1: Identifying Arabic comparative statements using common machine learning methods.

From the figure we can find that much improvement done for precision for all classifiers comparing with using only POS. But, for recall the POS method got equal to machine learning classifiers, and some time less such as in case of SVM.

Using machine learning with POS tagging we got results as in figure 6.2

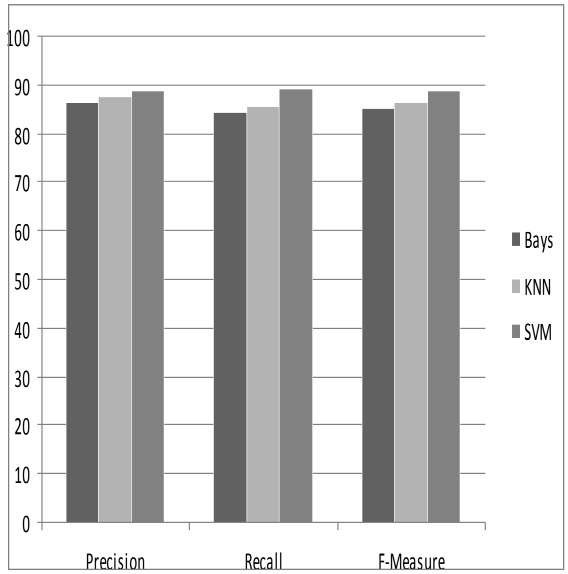

Figure 6.2: Identifying Arabic comparative statements using machine learning methods and POS tags.

From the graph we can conclude that performance of naïve Bayes and SVM slightly increase when we use POS with the classifier to identify comparative statements. However; that is not the case if we use KNN where the performance decrease when use POS with the classifier comparing to using only the classifier.

6.2 EXAMPLES OF COMPARATIVE TYPES

To identify the type of Arabic comparative statement, basically, we generate a set of rules that can cover many types as follows:

1) The most rule that identify the first type which is *Non-equal gradable comparisons* are:

a) JJR (adjective comparative) or RBR (adverb comparative) followed by IN (Preposition) such as " ويندوز 7 أفضل من ويندوز فيستا في السرعة" (Windows 7 better than Vista in term of speed).

b) JJR or RBR followed by NNP then IN " عاطف أفضل بكثير من محمد" (Attaf much better than Muhammed).

c) Replace JJR or RBR with "متفوق" or "متفوقة" " such as "أحمد متفوق على حسن" (Ahmed superior of Hassn)

2) Rules that covers the second type (Equative comparisons) are:

a) Using the " نفس الشيىء"(same) as" الجامعتين أ و ب نفس الشئ بالتعليم " (Both University A and B are the Same in Education).

b) Using the "متعادلين "(even) as "متعادلين الزمالك والاهلي " (Ahly and Zamalk are even).

3) The third Type is superlative comparisons has the following rules:

a) JJR followed by NN and without IN " ويندوز 7 افضل نظام تشغيل" (Windows 7 the best operating System)

b) Using JJS as "السامسونغ هو الاقوى في تصوير " (Samsung is the best in picturing) .

c) Has words such as "لا مجل للمنافسة" (No one can compete with it) such as " الجامعة أ هي رائدة الجامعت ولا مجل للمنافسة في ذلك", (University A is the best and has no competitive).

4) In this work we did not get a discrete rules for part four (Non-gradable comparisons) it needs more complex method to get.

7. CONCLUTION

This paper discussed a way to mine comparative sentences from Arabic text. For the first task, which is identifying comparative statements from no-comparative ones, we used three approach: firstly we used POS tags, in which we got f-measure of 63.73 % on average with low precision. Secondly we used three machine learning methods classifiers, K-nearest neighbors , Support Vector Machine and BayesNaive. Using these methods the best f-measure we got was 86.63% by using KNN. We can conclude that using machine learning methods is much better than using POS. Thirdly, we combined the two approaches where we got f-measure of %88.87 using SVM and POS which is a little improvement comparing to only use machine learning.

For the second task we generated a set of rules to characterize the three types of comparative statements which are: Non-equal gradable comparisons, equative comparisons and Superlative comparisons. For the fourth type, which is Non-gradable comparisons we could get any discrete rules.

In the future work, for the first task, we can improve the performance of identifying comparative statements by using more linguistic features other than POS such as (word semantics, named entity, ..etc). For the second task, we can use methods that can automatically defined the comparative type. For the third task, we look for methods to extract comparative relations from the identified sentences

[1] B. Liu. "Searching Opinions in User-Generated Contents" . Invited talk at the Sixth Annual Emerging Information Technology Conference (EITC-06), Aug 10-12,, Dallas, Texas, 2006.

[2] A. Harb and M. Dray. "Web opinion mining: how to extract opinions from blogs?" Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology. ACM New York, NY, USA. 2008

[3] M. Ganapathibhotla and B. Liu . " Mining opinions in comparative sentences". In the Proceedings of the 22nd International Conference on Computational Linguistics 2008. Stroudsburg, PA, USA.

[4] S.Yang and Y. Ko. "Extracting Comparative Sentences from Korean Text Documents Using Comparative Lexical Patterns and Machine Learning Techniques". In the proceedings of the ACL-IJCNLP2009.Suntec,Singapore,4 August 2008.

[5] N. Jindal and B. Liu. "Mining Comparative Sentences and Relations" . In the proceedin*gs of the 21st Conference on Artificial Intelligence AAAI-06*, AAAI Press

[6] F. Hou and G Li. "Mining Chinese comparative sentences by semantic role labeling ". International Conference on Machine Learning and Cybernetics, 12-15 July 2008

[7] S. Elkateb, W. Black, , P. Vossen, D. Farwell , A. Pease, and C. Fellbaum . "Arabic WordNet and the Challenges of Arabic" . Proceedings of the International Conference of Arabic NLP/MT, London.

[8] [N. Jindal, and **B. Liu**. ."Identifying Comparative Sentences in Text Documents." In: Proceedings of SIGIR 2006 Conference.

[9] K. Xu, S. S. Liao, J. Li and Y. Song. "Mining comparative opinions from customer reviews for Competitive Intelligence" Decision Support Systems. 50 (2011) 743–754.

[10] R. Du; R. Safavi-Naini, and W. Susilo. "Web filtering using text classification" . *The 11th IEEE International Conference on* In Networks, 2003. ICON2003. The 11th IEEE International Conference on (2003), pp. 325-330.

[11] B. Dasarathy. "Nearest neighbor (NN) norms : NN pattern classification techniques". IEEE Computer Society Press, 1991.

[12] C. Cortes and V. Vapnik. Support-Vector Networks. Machine Learning, 20, 1995

[13] K. Toutanova and C. D. Manning. "Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger". In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very*

Large Corpora (EMNLP/VLC-2000), pp. 63-70.

[14] G. Salton and C. Buckley. "Term-weighting approaches in automatic text retrieval ". Information Processing & Management 24 (5): 513–523. 1988.

[15] J. Makhoul, F. Kubala, R.Schwartz, R. Weischedel. "Performance measures for information extraction". In: Proceedings of DARPA Broadcast News Workshop, Herndon, VA, February 1999