Sentiment Analysis for Arabic Reviews in Social Networks Using Machine Learning

Mustafa Hammad and Mouhammd Al-awadi

Abstract In this emerging age of social media, social networks become growing resources of user-generated material on the internet. These types of information resources, which are an expansive platform of humans' emotions, opinions, feedback, and reviews, are considered powerful informants for big industries, markets, news, and many more. The great importance of these platforms, in conjunction with the increasingly high number of users generating contents in Arabic language, makes maiming the Arabic reviews in social networks necessary. This paper applies four automatic classification techniques; these techniques are Support vector Machine (SVM) and Back-Propagation Neural Networks (BPNN), Naïve Bayes, and Decision Tree. The main goal of this paper is to find a lightweight sentiment analysis approach for social networks' reviews written in Arabic language. Results show that the SVM classifier achieves the highest accuracy rate, with 96.06% compared with other classifiers.

Keywords Polarity · Sentiment analysis · Text classification · Data mining · Arabic language · Social media

1 Introduction

Finding an automatic way to analyze and classify users' reviews in social networks is very important. This is because it is the most empirical way to get the direct feedback or information from people. For example, finding which reviews are giving positive or negative polarities helps in dealing with customers' behaviors. The process of classifying texts or documents according to their polarity is known as Sentiment Analysis (SA) [1,2].

M. Hammad(✉) · M. Al-awadi
Department of Information Technology, Mutah University, Al Karak, Jordan
e-mail: hammad@mutah.edu.jo

© Springer International Publishing Switzerland 2016

S. Latifi (ed.), *Information Technology New Generations*,
Advances in Intelligent Systems and Computing 448,
DOI: 10.1007/978-3-319-32467-8_13

Sentiment analysis, which is part of Natural Language Processing (NLP), is set to extract the meanings from a text in a way to find the polarity of the text. During sentiment analysis process, it is important to try to maintain data accuracy despite the need for features selection methods. These features are then classified to which polarity this text belongs.

A study done on 2014 [3] estimated that the percentage of websites containing Arabic language contents constitutes around 0.8 % of all the contents on the internet. Users of social networks on the web create around 30% of the Arabic content. However, Arabic language has fuzzy and vague semantics, which makes the process of analyzing Arabic texts hard to be developed in a systematic way. Moreover, Arabic users use informal Arabic language, which is deferent than the formal Modern Standard Arabic (MSA), and can be vary from one to another. Informal Arabic language is the most using language of speaking and communication on social media outlets. Moreover, this spoken Arabic varies, in terms of vocabulary and sentence structure, by country. Unlike English language, the research interest on processing Arabic language texts is started recently.

This paper performs SA on Arabic reviews and comments by using four machine learning algorithms, which are Support Vector Machine (SVM) [4], Back-Propagation Neural Networks (BPNN) [5], Naïve Bayes [6], and Decision Tree [7]. We collected 2000 Arabic reviews from social media in order to evaluate the different machine learning algorithms based on sentiment analysis. In this paper, a particular attention is paid on the preprocessing and features selection for Arabic reviews to optimize the classification process. Our motivation was to evaluate the strengths and weaknesses of these commonly used machine learning methodologies in the field of informal Arabic text sentiment analysis.

The rest of this paper is organized as follows; related works are presented in Section 2. Section 3 describes the used machine learning algorithms. The proposed methodology is presented in Section 4. Section 6 presents the experimental results, followed by the conclusion.

2 Related Work

Sentiment analysis is a research space in which a lot of difficult problems are to be tackled. A number of different approaches have been taken to sentiment analysis. In general, these approaches have relied on one of two techniques: either supervised or unsupervised machine learning. These techniques for sentiment analysis have been principally focused on Indo-European languages, especially English. However, to date, there are few approaches to sentiment analysis social media's texts that have focused on Arabic language.

Santidhanyaroj et al. [8] presented a system that uses analysis of social network data to evaluate public sentiment of issues and actions affecting the general social conscience. They made use of two algorithms: Naïve Bayes and SVM classifiers. Within the model for sentiment analysis, a pre-categorized data set was compiled from Twitter. The SVM outperformed the Naïve Bayes classifier in analysis and provided more consistent, reliable results.

A sentimental analysis for Arabic language using two popular tools for social analysis: Senti-Strength and Social Mention was proposed by Khasawneh et al. [9]. The data set for this study was taken from Facebook and Twitter posts in Arabic. Based on the contents of the posts, the tools were used to identify polarity of the comments. A comparison of the two tools in ability to guess polarity revealed that the Naïve Bayes gave 91.83% for the Social Mention tool and 95.59% accuracy for Senti-Strength.

Kamal et al. [10] merged the rule-based and machine-learning methods to propose a new sentiment analysis method for identifying polarity. This approach's novelty stems from its use of NLP features and statistical sets, which allow for taxonomy of sentiment at the word-level.

In [11], Abdul-Mageed and Diab exhibited AWATIF a multi-type corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis (MSA SSA). This approach is based on by looking to demonstrate how annotation examines inside subjectivity and assessment dissection (SSA) can both be roused by existing phonetic hypothesis and cater for type subtleties.

Other works, such as [12], used three different datasets pulled from data on Twitter to classify the sentiment of Semitic features using three different methods of analysis: replacement, augmentation, and interpolation. Alaa El-Halees [13] presented another approach for extracting opinions from Arabic text. His combined approach was tri-fold. It consisted of a lexicon-based method for text classification, a training set of the classified texts for Maximum Entropy model, and finally a K-nearest neighbor for classification of the remaining texts.

M. Rushdi-Saleh et al. presented an Opinion Corpus for Arabic (OCA) in [14]. The corpus consists of Arabic text pulled from web pages specifically focused on movies and films. They produced different classifiers using SVM and Naïve Bayes algorithms. A. Al-Subaihin et al. [15] proposed an Arabic language sentiment analysis tool that relies on human-based computing. The tool is beneficial as it aids in the construction, dynamic development, and maintenance of the lexicon.

3 Machine Learning Algorithms

In sentiment analysis, the appropriate type of machine learning to use is text classification, which is known as "supervised learning". This section briefly describes back-propagation neural network, naïve bayes, decision tree, and support vector machine classification methodologies, which are employed in this paper.

3.1 Back-Propagation Neural Networks

Artificial Neural Networks (ANN) simulates human neurons based on building three layers; input layer, hidden layers and output layer. Input layer consists of vectors after applying feature dimensionality reduction, hidden layers is the activity region in categorization. Output layer represents the categories. There are two types of training the neural networks, which are feed-forward and feedback. Feedback algorithm is designed to reduce the mean square error between the real output of a networks and the desired output. Within the back-propagation algorithm, there is both a forward pass, which gets the activation value, and a backward pass, which adjusts weights biases relative to desired and actual outputs. Until the network converges, the backward and forward passes iterate continuously.

3.2 Naïve Bayes

Naïve Bayes classifier is frequently used for text classification problems, which is a simple probabilistic on the so-called Bayesian theorem. Bayes' classifier is particularly suited when the dimensionality of the inputs is high. It assumes that a document's features are conditionally independent of one another. Given a document T_i; the probability that this text is belong to a class C_j, noted as $P(C_j| T_i)$, is calculated as: is calculated as:

$$P(Cj|Ti) = \frac{P(\{Ti|Cj\}) * P(Cj)}{P(Ti)} \quad (1)$$

where $P(T_i |C_j)$, $P(C_j)$, and $P(T_i)$ is the probability of text Ti is in class C_j, class C_j occurrence, text Ti occurrence respectively in the training set. Bayes' classifier classifies the target text to the class that has the highest probability value among all others classes.

3.3 Decision Trees

Decision tree [23] is one of the most common techniques for classification. The decision tree is composed of nodes that are connected with branches. The leaf nodes are the final decision that categorized the target input to a specific class. All inner nodes are set to split the classes into subsets of classes according to a set of selected features checks.

Designing the decision tree is done by analyzing the training data to find the discriminatory features. The most significant features are placed at the top of the tree, and recursively add the next features checks in the following nodes levels until we reach the tree's leafs, which labels the target text to a category.

3.4 Support Vector Machine (SVM)

SVM is based on finding the best hyper plane that divide the data into two categories with the largest possible margin. In this linear classification, the hyper plane $y = w f(x) + b$ Where w is the weights vector, b is the bias, and f(x) denotes the mapping function from input feature space with maximum margin. SVM can be used to create boundaries among categories in case of nonlinear classifications. In this case, the data is mapping into another space, which SVMs perform the linear algorithm over it.

4 Sentiment Analysis Model

The proposed sentiment analysis model consists of two basic phases; training the classifier with a set of reviews, and then testing it with another set. However, data sets in both phases need to be preprocessed before the classification process. The preprocessing step, as in any text classification, has a huge impact in the classification results. Figure 1 describes the proposed approach.

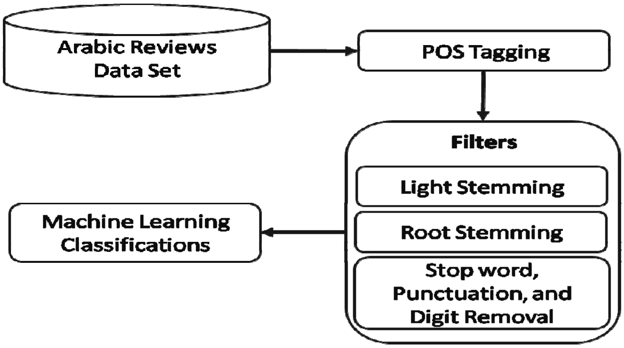

Fig. 1 Overview of the proposed sentiment analysis approach.

The first step is to categorize a review's words to recognize its structure. This is known as parts of speech (POS) tagging. The main goal for POS tagging is to identify all the structural features of a sentence, such as verbs, nouns, adjectives, and adverbs. We used well defined Java code for tagging the various parts of speech in Arabic reviews.

The second step in the proposed model is set to filter unimportant features in the data. Stemming is a morphological analysis technique for text processing. Light stemming removes some common affixes, which include definite article prefix and pluralizing suffix. On the other hand, root stemming reduces the words to only their stems.

In the final step, before applying machine learning algorithms, stop words that do not affect the polarity of a text are removed. In addition to that, punctuations and digits are removed by using their Unicode representations.

5 Experimental Results

To evaluate Machine learning algorithms on Arabic reviews sentiment analysis we used the RapidMiner tool[1], which is an open source java-based environment for machine learning and data mining. RapidMiner creates XML files that contain the user-defined operations to be applied to the data. There are many operators on the RapidMiner tool for different machine learning processes, such as preprocessing

[1] http://rapidminer.com

and visualizing data. We found that the RapidMiner tool can accommodate the Arabic language easily.

To evaluate the proposed approach, the experiments is built based on dividing the labeled data into two groups; the first group is set for tanning the machine learning algorithms, which is 70% of the collected data set, and the second group, which is the remaining 30% of the data set, is set to evaluate and test the classifiers.

5.1 Collected Data Set

In order to apply machine learning classifications, a set Arabic reviews should be collected to train and test the classifiers. The proposed data set is collected from Jordanian hotels' customers' reviews on the internet. The collected data set is a combination of Arabic reviews and comments from Facebook, Twitter, and YouTube. The total number of collected reviews is 2,000. Table 1 shows the total number of reviews for each category of the three polarities; positive, negative, and neutral. The collected data set is has a lot of informal Arabic and vernaculars.

Table 1 Number of reviews by polarity.

Category	Number of Arabic Reviews
Positive	1003
Negative	593
Neutral	404

5.2 Performance Measures and Results

In our experiments, we have used the most common performance measures that include recall (R), precision (P), and F-measure (F). Given a test set of reviews S1 that are labeled to the polarity k, and a prediction set S2 that is labeled with polarity k by the machine learning algorithm, the recall (R) and precision (P) measures are defined by:

$$R = \frac{S1 \cap S2}{S1} \tag{2}$$

and

$$P = \frac{S1 \cap S2}{S2} \tag{3}$$

respectively [16]. To compare the different algorithms with a single rate, we used the F-measure, which combine recall and precision. F-measure that is used in our experiments is defined as:

$$F - measure = \frac{2RP}{R+P} \tag{4}$$

The experiments' results for applying sentiment analysis for the proposed data set using SVM, BPNN, Naïve Bayes, and Decision Tree classifications algorithms are shown in Table 2 in term in terms of Precision, Recall and F-measure.

As shown in Table 2, SVM achieved the highest accuracy, which is 96.06%, followed by Naïve Bayes with average accuracy of 88.38%, and Decision Tree with average accuracy of 85.82%. BPNN was the lowest average accuracy of 69.77%. Moreover, SVM has the highest precision and recall values.

Table 2 Performance results for each classifier.

Classifier	Precision	Recall	F-measure
SVM	95.80%	96.40%	96.06%
BPNN	72.14%	67.61%	69.77%
Naïve Bayes	92.62%	84.99%	88.38%
Decision Tree	83.98%	87.99%	85.82%

In order to test the learning abilities we setup five different experiments with 300, 600, 900, 1200, and 1500 training reviews set. Figure 2 shows the F-measure value for each classifier with these five training sets. As shown in Figure 2, the accuracy of SVM, Decision tree, and Naïve Bayes classifiers increase as long as we increase the training data. However, BPNN does not learn reasonably well. We believe this is because the ambiguity of the data set, which affect the behavior of the neural network.

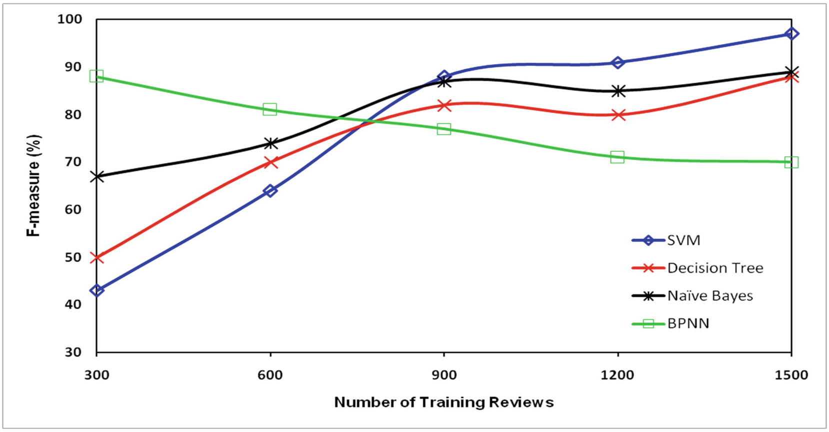

Fig. 2 F-measure value versus number of training reviews.

Figure 3 shows the average training time in seconds for each classifier. As shown in the figure, Decision tree classifier needs the largest amount of training time compared to other three classifiers. SVM has the smallest training time, which is 6.45 seconds. Naïve Bayes has also a small training time with 11.41 seconds.

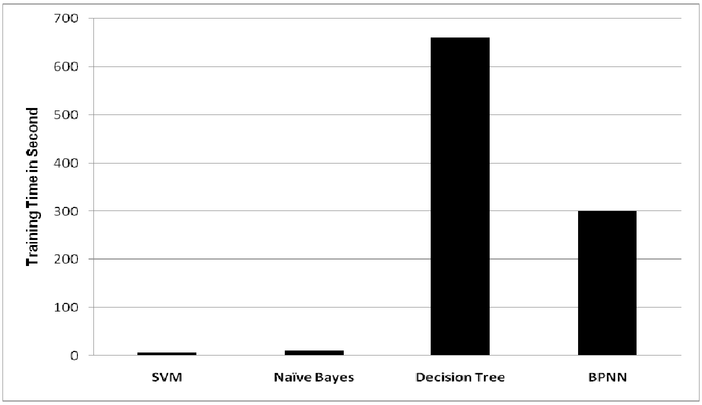

Fig. 3 Average training time for each classifier.

6 Conclusion

This paper applies sentiment analysis on Arabic reviews, which are collected from social media using machine learning algorithms. Experimental results showed that the proposed preprocessing technique for Arabic reviews allow different machine learning algorithms to be applied in the extracted feature for the target data set. In this paper, we used four different classification algorithms to analyze the polarity of sentiments for Arabic reviews. Results showed that the most accurate machine-learning algorithm, based on our approach, was SVM with F-measure equal to 96.06%.

Sentiment analysis of Arabic language requires a lot more research especially in the preprocessing phase. As future works, larger and more diverse data sets can be used. It is also important to increase the size of the dictionaries and remedy the memory problem in RapidMiner.

References

1. Nasukawa T., Jeonghee, Y.: Sentiment analysis: capturing favorability using natural language processing. In: Proceedings of the 2nd International Conference on Knowledge Capture. ACM (2003)
2. Dave, K., Lawrence, S., Pennock, D.M.: Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th International Conference on World Wide Web. ACM (2003)
3. Usage of content languages for websites. http://w3techs.com/technologies/overview/content_language/all
4. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning **20**, 273–297 (1995)
5. Goh, A.T.C.: Back-propagation neural networks for modeling complex systems. Artificial Intelligence in Engineering **9**(3), 143–151 (1995)

6. Chen, J., Huang, H., Tian, S., Qu, Y.: Feature selection for text classification with naïve bayes. Expert Systems with Applications **36**, 5432–5435 (2009)
7. Quinlan, J.R.: Induction of decision trees. Machine Learning **1**, 81–106 (1986)
8. Santidhanyaroj, P., Khan, T., Gelowitz, C.M., Benedicenti, L.: A sentiment analysis prototype system for social network data, pp. 1–5 (2014)
9. Khasawneh, R.T., Wahsheh, H.A., AI-Kabi, M.N., Aismadi, I.M.: Sentiment Analysis of Arabic Social Media Content: A Comparative Study, pp. 101–106 (2013)
10. Kamal, A., Abulaish, M.: SMIEEE.: Statistical Features Identification for Sentiment Analysis using Machine Learning Techniques, pp. 178–181 (2013)
11. Abdul-Mageed, M., Diab, M.: AWATIF: a multi-genre corpus for modern standard arabic subjectivity and sentiment analysis. In: Proceedings of LREC, Istanbul, Turkey (2012)
12. Elhawary, M., Elfeky, M.: Mining arabic business reviews. In: 2010 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 1108–1113. IEEE (2010)
13. El-Halees, A.: Arabic opinion mining using combined classification approach. In: Proceedings of the International Arab Conference on Information Technology, ACIT (2011)
14. Rushdi-Saleh, M., Martín-Valdivia, M., Ureña-López, L., Perea-Ortega, J.M.: Bilingual Experiments with an Arabic-English Corpus for Opinion Mining (2011)
15. Al-Subaihin, A., Al-Khalifa, H., Al-Salman, A-M.: A proposed sentiment analysis tool for modern arabic using human-based computing. In: Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services. ACM (2011)
16. Rijsbergen, C.J.V.: Information Retrieval: Butterworth-Heinemann (1979)