VFAST Transactions on Computer Sciences
http://vfast.org/journals/index.php/VTCS@ 2018 ISSN(e): 2308-8168; ISSN(p): 2411-6335
Volume 6, Number 1, January-December, 2018

pp. 93-97

PERFORMING SUBJECTIVITY CLASSIFICATION IN TEXT USING SUPPORT VECTOR MACHINE

ALAA OMRAN ALMAGRABI
Department of Information Systems, Faculty of Computing and Information Technology (FCIT) King Abdul Aziz University (KAU) Jeddah, Kingdom of Saudi Arabia
aalmagrabi3@kau.edu.sa

Abstract:*In this paper, I address the problem of the subjectivity classification in text. The subjective text is opinion bearing, whereas the objective text is text without expressing opinions. The supervised learning technique namely, Support Vector Machine (SVM) is used to classify the text as subjective and objective. A publically available dataset of drug reviews is used to conduct the experiments using WEKA platform. The experimental results show that the proposed SVM classifier performed better than the other classifiers.*
Keywords:-subjectivity classification, sentiment analysis, support vector machine, supervised machine learning

1. Introduction: Opinion mining, also called sentiment analysis deals with the acquisition, analysis and summarization of user reviews from different social media forums [1, 2]. The input text in form of user reviews or tIets can be categorized as subjective or objective. The subjective text contains opinion (sentiment) expressed by users and the objective text contains information without any opinions.

The previous studies [1,2, 3, 4] on the subjectivity classification used list of opinion words with other lexical resources, such as SentiWordNet [3,4]. The major drawback of such approach is the limited coverage of opinion words, i.e. if an opinion word is not present on the lexicon, then the text can't be classified correctly as subjective or objective.

Aforementioned problem can be solved by using a supervised learning-based subjectivity classification approach. Different machine learning classifiers, such as Support Vector Machine Naïve Bayes (N.B), K-Nearest Neighbor (KNN) and others can be used for the efficient classification of text [5].

In this work, a supervised learning-based technique using SVM, is proposed for the efficient classification of text as subjective or objective. The experiments are conducted using Weka platform with drug review dataset.

Rest of the paper is organized as follows: in section 2, related work is presented; proposed method is shown in section 3; section 4 shows experimental results and final section concludes the work with future directions.

2. Related Work: In this section, I present some of the selected studies performed on sentiment analysis with subjectivity analysis paradigm. The machine learning approaches can be applied to address the problem of subjectivity analysis in text [6] presents a technique using a trained machine learning classifier, namely SVM. They used a data set of topic-specific articles from Wikipedia (objective documents) and review social media sites (subjective documents).

Wiebe J [7] in his work, proposed a technique to identify subjective terms in writing. For this purpose, two interpretations of private state sentences Ire differentiated using subjective features. After the recognition of characters, a mechanism is proposed to choose subjective words.

Pang et al. [8] proposed a system to classify user feedback as +ive or -ive. The classifier is devised on the basis of an aggregated sentiment of the reviews. The performance of the proposed method w.r.t other machine learning algorithms, is evaluated on movie review dataset.

Strapparava and Mihalcea [9] performed the task of emotions annotation in text automatically. They used six basic emotions, namely: Anger, Disgust, Fear, Joy, Sadness and Surprise to construct a large dataset. A dataset of 1000 news headlines is used to conduct the experiments. However, their approach lacks in consideration of lexical structure of emotions to deal with deeper understanding of semantic text processing.

Prem et al. [10] proposed a hybrid technique by combining the lexical resources and the text classification methods to detect and classify opinion sentences in political, products and movie domain. They suggest "a pooling

multinomial classifier, which provides a platform where composite Naive Bayes classifier can be launched which, took background knowledge and training examples under consideration". The system is capable of acquire and apply knowledge from different sources. The text is classified using two different repositories, including lexical resources and the labelled training set.

A unique machine learning-based approach, namely "thumbs up and thumbs down" was proposed by applying different machine learning classifiers, such as NB, Maximum Entropy, and SVM. Experimental results show that SVM performed better than other approaches.

Das and Bandyopadhyay [14] developed an emotion extraction and tagging system by using Ekman's six basic emotions at three levels of intensities i.e. medium, low and high.

They achieved promising results on a gold standard dataset. However, metaphors and their impact on detecting emotion at sentence level, is not considered.

3. Methods: The proposed supervised machine learning algorithm-based emotion classification system. Firstly, supervised machine learning is introduced as follows:

3.1 Supervised Machine Learning

The supervised machine learning algorithms `are based on training and testing datasets. The training dataset is used to train the classifier and testing dataset is used to test the prediction capability of the trained classifier [5]. There are different supervised machine learning algorithms, such SVM, NB, KNN, Logistic Regression etc. In this work SVM is selected as the target classifier for subjectivity classification.

The proposed system (Fig. 1) works as follows:

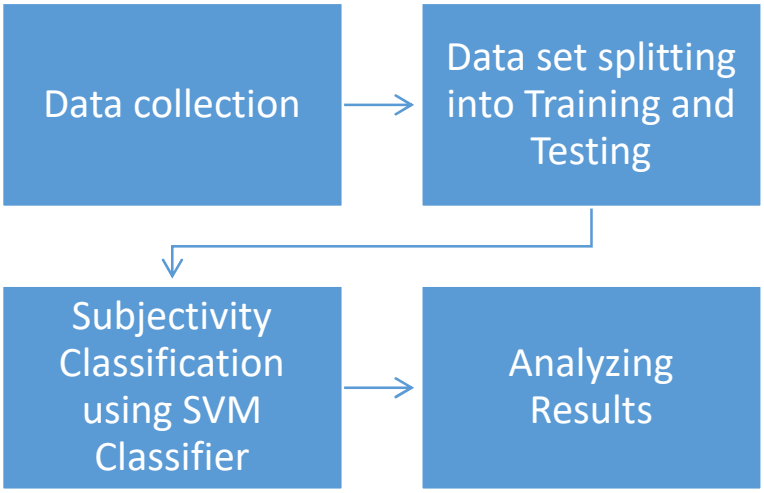

Fig. 1 Proposed System

3.2 Data Collection: I used a publically available dataset on drug reviews [15]. I used Weka [16] platform for performing subjecting classification using SVM.

3.3 Data set splitting into Training and Testing

The acquired dataset is split into training (80%) and testing (20%) in Weka Platform. The classifier is trained on the training dataset and its prediction is performed on testing dataset. Table 1 shows a sample set of reviews from the dataset.

Table-1: Sample Reviews from Drug Data Set

Text-id	Review Text	Subjectivity Label
1	I was dizzy nauseous, and exhausted for all seven days of treatment	Sub
2	It worked and got rid of the smell and infection	Sub
3	The pill started to dissolve as soon as I placed it on my tongue and tastes horrible	Sub
4	The first pill I took made my face look like I layed in a tanning bed for an hour or two. I just thought it was fever from the infection and went to bed.	Sub

5	Compared to other reviews, I haven't really had any of the common reactions, but when I have experienced them they seemed to go away on their own.	Sub
6	This is the second time I have used this drug	Obj
7	I didn't come off the couch for 2 days	Obj
8	I've taken this drug twice before, and am on my third time using it.	Obj
9	The taste is awful, easier to take with yogurt or pudding.	Sub
10	The taste makes me shiver and lasts for a few seconds.	Sub

3.4 Subjectivity Classification Using SVM

The subjectivity detection classification module aims at classifying the text as subjective or objective. For example, the sentence: "*This laptop is beautiful*", when passed through the SVM classifier, is classified as "*subjective*". Table 2 shows a sample dataset entries, where the sentences are classified as *subjective* and *objective* by the Support Vector Machine.

3.5 Support Vector Machine

Support Vector Machine (SVM) is a non-probabilistic binary linear classifier that constructs a hyperplane or set of hyperplanes in an excessive or countless dimensional space, which can be used for classification, regression, or different tasks. The major concept underlying SVM, for subjectivity classification is to discover a hyper plane which divides the documents, or in our case, objective and subjective sentences as per the comment/sentiment [6].

$$D = \{(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)\}$$

Table-2: A sample output of SVM classifier for subjectivity classification

Input Sentence	TIets	Subjective term (s)	Subjective/Objective
1	Did not help with pain	Help	Subjective
2	I took it for 4 days	Nil	Objective
3	I am on pills, awesome**!** Superb, love them**!**	Superb, love,	Subjective
4	The result is neck pain and sciatic leg pain	Pain	Subjective

4. Implementation and Results

4.1. Weka Platform

The Waikato Environment for Knowledge Analysis (Weka) platform [16] is used for the implementation of the proposed system. For this purpose, data files of subjectivity classification are transformed into Comma Separated Value (CSV) format. For making the files w.r.t to validation, the CSV files are transformed into attribute-relation file format (arff). To build a model, SVM is applied on the training set for analysis and the testing data set is evaluated.

4.1. Confusion Matrix

The confusion matrix is the matric representation of correct and incorrect prediction labels. Different measures like precision, recall and accuracy, are applied. The correctly specified labels are represented by true positive, negative, neutral and false positive, negative, neutral are incorrectly specified labels.

Following metrics are used to evaluate the performance of the proposed system [17].

$$Precision\ (p) = \frac{tp}{tp + fp} \qquad (i)$$

$$Recall\ (r) = \frac{TP}{tp + fn} \qquad (ii)$$

$$F - measure = \frac{2(p)(r)}{p + r} \qquad (iii)$$

$$\text{Accuracy} = \frac{tp + tn}{tp\ +\ fp + tn + fn} \qquad (iv)$$

Where tp is the number of true +ive classifications, fp is the number false +ive classifications, tn is the number of true -ive classifications, and fn is the number of false -ive classifications.

4.2. Accuracy-based Evaluation of Subjectivity Classification

An experiment is conducted to evaluate the performance of the proposed system in terms how accurately it predicts the subjective and objective sentences. Results obtained (Table 3)show that about 84% accuracy is obtained for subjective and 79% for objective sentences.

Table-3: Accuracy subjective and objective Sentences

Dataset	Polarity Tag	Total no. sentences	Accuracy (%)
Drug Reviews	subjective	1413	0.841
	objective	656	0.792

4.3. Comparison with other Classifiers

In this experiment, I compare the performance of the prosed system with other classifiers in terms different evaluation measures like NB and KNN. The results presented in Table 4 show that the proposed system (SVM) for subjectivity classification has outperformed the other classifiers.

Table-4: Performance Evaluation Results

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Naïve Bayes	59	59	59	57
K-Nearest Neighbour (KNN)	70	74	71	65
Proposed (Support Vector Machine)	**93**	**94**	**94**	**94**

5. Conclusion and Future Work

In this paper subjectivity classification problem is addressed and supervised machine learning technique, namely Support Vector Machine. Drug reviews available in the bench mark dataset are used in the experiments under Weka platform. The results show that SVM performed better than the other classifiers. In future, further experiments with other classifiers using extended datasets in multiple domains can be carried out.

REFERENCES

[1] Khan, A., Asghar, M. Z., Ahmad, H., Kundi, F. M., & Ismail, S. (2017). A rule-based sentiment classification framework for health reviews on mobile social media. Journal of Medical Imaging and Health Informatics, 7(6), 1445-1453.

[2] Kundi, F. M., Ahmad, S., Khan, A., & Asghar, M. Z. (2014). Detection and scoring of internet slangs for sentiment analysis using SentiWordNet. Life Science Journal, 11(9), 66-72.

[3] Kundi, F. M., Khan, A., Ahmad, S., & Asghar, M. Z. (2014). Lexicon-based sentiment analysis in the social Ib. Journal of Basic and Applied Scientific Research, 4(6), 238-48.

[4] Ahmad, S., Kundi, F. M., Tareen, I., & Asghar, M. Z. (2016). Lexical Based Semantic Orientation of Online Customer Reviews and Blogs. arXiv preprint arXiv:1607.02355.

[5] Asghar, M. Z., Khan, A., Khan, F., & Kundi, F. M. (2018). RIFT: A Rule Induction Framework for Twitter Sentiment Analysis. Arabian Journal for Science and Engineering, 43(2), 857-877.

[6] Zhang J, Yu CT, Meng W (2007) Opinion retrieval from blogs. In: Silva MJ, Laender AHF, Baeza-Yates RA, McGuinness DL, Olstad B, Olsen ØH, Falcão AO (eds) CIKM, ACM, pp 831–840.

[7] J. M. Wiebe, “Identifying subjective characters in narrative,” Proceedings of the 13th conference on Computational linguistics, Association for Computational Linguistics, Vol. 2, pp. 401-406, Aug. 1990.

[8] W. Zhang, H. Xu, and W. Wan, “Iakness Finder: Find Product Iakness from Chinese Reviews by Using Aspects Based Sentiment Analysis, ” Expert Systems with Applications, vol. 39, no. 11, pp. 10283-10291, 2012.

[9] Strapparava C, Mihalcea R, "Learning to identify emotions in text”, Proceedings of the 2008 ACM symposium on Applied computing, ACM.

[10] Melville, Prem, Wojciech Gryc, and Richard D. Lawrence. "Sentiment analysis of blogs by combining lexical knowledge with text classification." Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009.

[11] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002.

[12] Dave K, Lawrence S, Pennock D (2003) Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th international conference on WorldWideIb,ACM, New York, NY, USA, WWW’03, pp 519–528. doi:10.1145/775152.775226.

[13] Pang B, Lee L (2004) A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd annual meeting on Association for Computational Linguistics, pp 271–278.

[14] Das D, Bandyopadhyay S (2010) Identifying Emotional Expressions, Intensities and Sentence Level Emotion Tags Using a Supervised Framework. PACLIC. Vol. 24.

[15] Asghar, M. Z. (2017). Data sets for User Reviews on Drugs.

[16] Holmes, G., Donkin, A., & Witten, I. H. (1994, November). Weka: A machine learning workbench. In Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on (pp. 357-361). IEEE.

[17] Kundi, F. M., Khan, A., Asghar, M. Z., & Ahamd, S. (2014). Context-aware spelling corrector for sentiment analysis. MGT Res Rep, 2(5), 1-10.