

Overview of existing algorithms for emotion classification. Uncertainties in evaluations of accuracies.

H Avetisyan, O Bruna, J Holub

Department of Measurement, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, Prague 6, 166 27, Czech Republic

E-mail: avetihak@fel.cvut.cz

Abstract. A numerous techniques and algorithms are dedicated to extract emotions from input data. In our investigation it was stated that emotion-detection approaches can be classified into 3 following types: Keyword based / lexical-based, learning based, and hybrid. The most commonly used techniques, such as keyword-spotting method, Support Vector Machines, Naïve Bayes Classifier, Hidden Markov Model and hybrid algorithms, have impressive results in this sphere and can reach more than 90% determining accuracy.

1. Introduction

Sentiment analysis (SA) is a method of extracting, analyzing, and determining emotions and sentiments of either text or speech information. It is widely used in e-shop or reviews of services to know customers' opinion about specific products or services taking into account specific criteria. It also can be used in social networks to analyze users' attitude in different situations (sadness, happiness, irony, etc.). Existing many approaches and algorithms trying to deal with SA. All of them have their advantages and disadvantages and different accuracies.

In this paper, information about some of commonly used current algorithms with their related works with obtained results has been collected. It was also paid attention to that fact that most of techniques were not evaluated the same way so problems during comparisons of accuracies of individual techniques may occur.

2. Used algorithms and their assessment

2.1. *Keyword based approach*

2.1.1. *Description*

One of the most used technique is keyword based approach. It detects emotions in the text at the basic word level and it is highly reliable mostly for analyzing emotion-bearing words and for simple sentences with clearly expressed emotions.

2.2. *Learning based approaches*

2.2.1. *Descriptions*



Learning based approaches are trained from a training set. Then the classifier determines either emotion directly for a word or a more complex structure of classifiers. Several types of learning based algorithms exist.

2.2.2. Support Vector Machines

2.2.2.1. Description

One of learning based approaches is the Support Vector Machines (SVM) binary classification technique. This algorithm uses training examples which contain information about their category. SVM creates a model which linearly classifies input data in already existing categories [1].

2.2.2.2. Related works and results

Table 1. Related works using SVM algorithm and obtained results

Paper name	Year	Algorithm	F-Score, %	Accuracy, %
[2]	2002	SVM with features based on unigrams	N/A	82.9
[3]	2007	SVM with General Inquirer, WordNet Affect, and other features	N/A	73.89
[4]	2009	SVM with Information Gain feature extraction	92.86 (POS) 88.28 (NEG)	91.15
[5]	2011	SVM with a linear kernel	N/A	80.29
[6]	2015	Speaker-dependent SVM with thresholding fusion	N/A	75.67

2.2.3. Naïve Bayes Classifier

2.2.3.1. Description

Another learning based approach is the Naïve Bayes Classifier (NBC). Different types of NBC exist depending on representation of input text. However, multinomial Naïve Bayes model has better results and is more accurate than other event models [7]. In this model, the frequencies of occurrences of specific words have been counted and demonstrated as a vector [8].

2.2.3.2. Related works and results

Table 2. Works related to Naïve Bayes Classifier and obtained results

Paper name	Year	Algorithm	F-Score, %	Accuracy, %
[2]	2002	NBC with features based on unigrams	N/A	78.7
[9]	2012	NBC and Naïve Search	N/A	~85
[10]	2013	NBC with Facebook Query Language query	72	N/A
[11]	2014	ERR-based NBC	84	N/A

[12]	2014	Multinomial NBC with features	91.2	N/A
------	------	----------------------------------	------	-----

2.2.4. Hidden Markov Model

2.2.4.1. Description

The Hidden Markov Model (HMM) is also simple and frequently used learning based algorithm for emotion detection. Basically, HMM is a technique which is able to distribute classes over sequence of observations.

2.2.4.2. Related works and results

Table 3. Works related to Hidden Markov Model and obtained results

Paper name	Year	Algorithm	F-Score, %	Accuracy, %
[13]	2003	Continuous HMM	N/A	77.8
[14]	2012	A High-order HMM with Viterbi algorithm	35.3	N/A
[15]	2013	3 states HMM	N/A	82.95

2.3. Hybrid approaches

2.3.1. Description

This type of approaches is combination of already mentioned approach types. It is able to detect emotions based on detected keywords, learned patterns, and other additional information from various dictionaries and thesauri [16].

2.3.2. Related works and results

Table 4. Works related to hybrid approaches and obtained results

Paper name	Year	Algorithm	F-Score, %	Accuracy, %
[17]	2004	Hybrid SVM (PMI/Osgood and Lemmas), 100 folds	N/A	89
[18]	2013	Keyword-spotting method and rule- based method	76.97	N/A
[19]	2013	Multinomial NBC with greedy search	N/A	85
[20]	2014	NBC and SVM using Information Gain and Chi- Square methods	N/A	71
[21]	2015	SVM and CRF with applied rules	N/A	91

3. Assessment of algorithms

3.1. Definition of the problem

It is important to mention that the researchers use different methodologies to assess the algorithms but use the same terminology, which may cause confusion for a reader or other researchers. In Tables 1-4, columns F-score and Accuracy are mentioned. The most common way of evaluating result of the AI is called F-score. Some authors calculated them in a different way than others and named obtained results as “accuracy”. This can cause a problem while comparing obtained results of different algorithms and approaches. For example, authors mostly assessed their algorithms using F-score [4], [12], [14], etc.

In some cases authors calculated the percentage of prediction of individual emotions and presented their results [6], sometimes with calculation of average result [13]. Note that first mentioned article calculates 6 types of emotions and the second mentioned article calculates 7 types. This also makes comparison of two algorithms impossible.

Other authors presented average results of their approaches without mentioning any specific information about steps or formulas by which they got their results [17].

These are few examples of fact that the term “Accuracy” is highly figurative and can’t be used as a standard term for comparing various algorithms.

3.1.1. Assessment using F-score

The first type is F-score which is calculated using 2 parameters called precision and recall. Precision and recall are measured using 4 parameters:

TP – the number of true positive samples

TN – the number of true negative samples

FP – the number of false positive samples

FN – the number of false negative samples

Precision (equation (3.1)) is the number of true positive samples divided by the number of predicted positive samples.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.1)$$

Recall (equation (3.2)) is the number of true positive samples divided by the actual positive samples.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.2)$$

The harmonic mean of precision and recall is called F-score and determined by equation (3.3). The F-score unit is percent.

$$F = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.3)$$

3.1.2. Assessment using term “Accuracy”

The second type of result is called Accuracy. This parameter presents the level of accuracy of specific algorithms. In this paper, simple average of all emotions accuracies were counted using the equation (3.4):

$$\text{Average} = \frac{1}{n} \cdot \sum_{i=1}^n \text{Emotion}_i \quad (3.4)$$

where n – the number of calculated emotions

$Emotion_i$ – the value of each individual emotion accuracy

The unit of measurement of Accuracy is percent.

3.2. Idea for solution of a problem

To resolve this incompatibility there is an idea to make standardized testing system for emotion classification algorithms and obtain results for them on the same conditions and criteria, and using the same linguistic resources [22]. This way is possible to prevent false accuracies and bring all the results to the same unit of measurement.

At this time there are no specific recommendations about approaches for algorithms testing standardization or make above mentioned results comparable. Nevertheless, this article should help other researchers to take into account above mentioned problem in their future works.

4. Conclusions

In this paper it was observed existing algorithms for emotion classification. Most commonly used and most efficient approaches have been taken into consideration and tables with information about them have been made. However, it has been noticed that between results of different approaches incomparableness can occur. To prevent this problem, it was recommended to make standardized testing system for emotion classification algorithms to be sure that those techniques are obtaining same type of results.

5. References

- [1] H. Binali, C. Wu, and V. Potdar, "Computational Approaches for Emotion Detection in Text," *4th IEEE Int. Conf. Digit. Ecosyst. Technol. - Conf. Proc. IEEE-DEST 2010, DEST 2010*, vol. 37, no. 5, pp. 498–527, 2010.
- [2] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," *Proc. Conf. Empir. Methods Nat. Lang. Process.*, pp. 79–86, 2002.
- [3] S. Aman and S. Szpakowicz, "Identifying Expressions of Emotion in Text," in *Text, Speech and Dialogue*, vol. 4629, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 196–205.
- [4] W. Zheng and Q. Ye, "Sentiment classification of Chinese traveler reviews by support vector machine algorithm," *3rd Int. Symp. Intell. Inf. Technol. Appl. IITA 2009*, vol. 3, pp. 335–338, 2009.
- [5] R. Burget, J. Karásek, and Z. Smékal, "Recognition of emotions in Czech newspaper headlines," *Radioengineering*, vol. 20, no. 1, pp. 39–47, 2011.
- [6] S. Gupta, A. Mehra, and Vinay, "Speech emotion recognition using SVM with thresholding fusion," in *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)*, 2015, pp. 570–574.
- [7] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," *AAAI/ICML-98 Work. Learn. Text Categ.*, pp. 41–48, 1998.
- [8] K.-M. Schneider, "On Word Frequency Information and Negative Evidence in Naive Bayes Text Classification," *Adv. Nat. Lang. Process.*, pp. 474–485, 2004.
- [9] M. M. Itani, R. N. Zantout, L. Hamandi, and I. Elkabani, "Classifying sentiment in arabic social networks: Naïve search versus Naïve bayes," *2012 2nd Int. Conf. Adv. Comput. Tools Eng. Appl.*, pp. 192–197, 2012.
- [10] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro, "Sentiment analysis of Facebook statuses using Naive Bayes Classifier for language learning," in *IISA 2013 - 4th International Conference on Information, Intelligence, Systems and Applications*, 2013, pp. 198–205.
- [11] S. Shaheen, W. El-Hajj, H. Hajj, and S. Elbassuoni, "Emotion Recognition from Text Based on Automatically Generated Rules," in *2014 IEEE International Conference on Data Mining*

- Workshop*, 2014, pp. 383–392.
- [12] S. Yoshida, J. Kitazono, S. Ozawa, T. Sugawara, T. Haga, and S. Nakamura, “Sentiment analysis for various SNS media using Naïve Bayes classifier and its application to flaming detection,” in *2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD)*, 2014, pp. 1–6.
 - [13] B. Schuller, G. Rigoll, and M. Lang, “Hidden Markov model-based speech emotion recognition,” in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, 2003, vol. 2, pp. II–1–4.
 - [14] D. T. Ho and T. H. Cao, “A High-Order Hidden Markov Model for Emotion Detection from Textual Data,” in *Knowledge Management and Acquisition for Intelligent Systems*, D. Richards and B. H. Kang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 94–105.
 - [15] S. Rustamov, E. Mustafayev, and M. a. Clements, “Sentiment analysis using Neuro-Fuzzy and Hidden Markov models of text,” *2013 Proc. IEEE Southeastcon*, pp. 1–6, 2013.
 - [16] E. C.-C. Kao, C.-C. Liu, T.-H. Yang, C.-T. Hsieh, and V.-W. Soo, “Towards Text-based Emotion Detection A Survey and Possible Improvements,” in *2009 International Conference on Information Management and Engineering*, 2009, pp. 70–74.
 - [17] T. Mullen and N. Collier, “Sentiment Analysis using Support Vector Machines with Diverse Information Sources,” *Proc. 2004 Conf. Empir. Methods Nat. Lang. Process. (EMNLP 2004)*, pp. 412–418, 2004.
 - [18] U. Krcadinac, P. Pasquier, J. Jovanovic, and V. Devedzic, “Synesketch: An Open Source Library for Sentence-Based Emotion Recognition,” *IEEE Trans. Affect. Comput.*, vol. 4, no. 3, pp. 312–325, 2013.
 - [19] N. Chirawichitchai, “Sentiment classification by a hybrid method of greedy search and multinomial naïve bayes algorithm,” in *2013 Eleventh International Conference on ICT and Knowledge Engineering*, 2013, pp. 1–4.
 - [20] X. Sun and C. Li, “Hybrid model based sentiment classification of Chinese micro-blog,” in *2014 International Conference on Audio, Language and Image Processing*, 2014, pp. 358–361.
 - [21] D. S. Nair, J. P. Jayan, Rajeev R.R, and E. Sherly, “Sentiment Analysis of Malayalam film review using machine learning techniques,” in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015, pp. 2381–2384.
 - [22] C. Strapparava and A. Valitutti, “WordNet-Affect: an affective extension of WordNet,” *Proc. 4th Int. Conf. Lang. Resour. Eval.*, pp. 1083–1086, 2004.

6. Acknowledgements

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS16/095/OHK3/1T/13