World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(REVIEW ARTICLE)

Check for updates

Improvement of feature engineering on emotion detection from textual data

Bikram Sarkar * and Joydeep Mukherjee

School of Education Technology, Jadavpur University, Kolkata, India.

World Journal of Advanced Engineering Technology and Sciences, 2024, 12(01), 073–076

Publication history: Received on 25 March 2024; revised on 06 May 2024; accepted on 09 May 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.12.1.0171

Abstract

This paper introduces a new method for selecting terms in the field of emotion recognition from text. Instead of focusing solely on very common or very rare terms, this approach considers moderately frequent terms as well. The idea is that these moderately frequent terms might also contain important information for distinguishing between emotions. Compared to traditional methods like Chi-Square and Gini-Text, this new approach performs better in many cases. To represent documents, the bag-of-words approach is used, where each document is represented by a vector. In this vector, each selected term is given a weight of 1 if it appears in the document, and 0 if it does not. Importantly, this new method includes terms that are not selected by Chi-Square and Gini-Text. Experiments conducted on a standard dataset demonstrate that including moderately frequent terms improves the accuracy of emotion recognition. This improvement is evident in terms of accuracy scores.

Keywords: Text categorization; Emotion Detection; Feature selection; Machine learning

1. Introduction

Recognizing emotions in natural language has become a challenging task due to the vast amount of data available on the Internet. Various methods have been developed and applied across different mediums such as images, speech, videos, and text to tackle this challenge. Understanding emotions is particularly crucial for improving human-robot interactions, as emotions play a significant role in such interactions. While recognizing emotions in images and videos is computationally expensive and difficult, text-based emotion recognition has gained prominence due to the exponential growth of textual data on the Internet. With the rise of social networks and internet-based applications, people often express their emotions through text. Text classification, the process of categorizing documents into predefined categories based on their content, plays a pivotal role in this domain. Text classification essentially mimics human abilities to categorize documents, enabling fast and accurate classification of vast amounts of information. Emotions expressed in text can be categorized into various classes such as anger, joy, disgust, sadness, fear, and surprise, among others. These emotions can further be categorized as primary, secondary, or tertiary forms, with some being considered desirable (e.g., joy) and others undesirable (e.g., anger). Psychological research has identified fundamental emotions, which are influenced by biological and social factors. These basic emotions include enjoyment/joy, interest/excitement, surprise, anger/rage, disgust, distress/anguish, fear, and shame/humiliation. Emotion recognition from text often focuses on a set of basic emotion categories, such as anger, disgust, fear, guilt, joy, shame, and sadness. Various methods exist for emotion recognition from text. One approach is lexicon-based, where emotions are identified based on predefined emotion lexicons. Another approach involves machine learning-based classifiers trained on labeled data to recognize emotions in new text inputs. Both approaches have been explored in recent research to improve emotion recognition accuracy and efficiency.

* Corresponding author: Bikram Sarkar

2. Literature Survey

Before text classification can be performed, several preprocessing steps are necessary. Firstly, the bag-of-words (BOW) representation is employed, treating each unique word in the document set as a distinct feature or term. Secondly, term scoring is crucially applied in two main areas: term selection and term weighting. Term selection involves choosing a subset of terms from the initial document collection to represent documents, discarding irrelevant terms based on various approaches. This is essential for managing the high dimensionality of the feature space and avoiding the inclusion of non-informative words, which can negatively impact automatic categorization systems. After selecting a subset of terms, the relative influence of these terms is quantified through term weighting. In the realm of emotion recognition from text, the BOW representation often leads to extremely sparse vectors due to the brevity of text documents, resulting in what is known as the feature sparseness problem. This issue hampers classifier performance, as sparse features can lead to less effective learning. To address the feature sparseness problem, previous research has explored various techniques. Some studies have introduced novel term weighting methods that consider feature similarities, while others have enriched the BOW representation by incorporating semantic and syntactic relations between words. These enhancements aim to improve classifier recall by mitigating sparseness issues. Additionally, research has highlighted the importance of rare words, demonstrating that they can significantly contribute to classifier accuracy, especially in domains with a high number of training samples. For instance, in patent classification studies, the inclusion of rare technical terms has been shown to enhance classification performance. This underscores the value of considering both frequent and rare terms in text classification tasks.

3. Method

Each aspect of the proposed framework

3.1. Chi-Square

Chi-Square is a statistical measure commonly used in feature selection to assess the independence between a term's occurrence and its class label. It is calculated using Equation (1), which considers the frequencies of term occurrences in both positive and negative class documents.

3.2. Gini-Text

Originally developed to measure impurity in decision trees, the Gini Index is adapted for feature selection as Gini-Text. This measure evaluates the impurity of a term based on its distribution across positive and negative class documents. Lower values indicate less impurity, making the term more informative for classification.

3.3. Relevance Frequency

Relevance Frequency focuses on the ratio of a term's occurrences in positive and negative class documents. It assumes that terms with similar ratios contribute equally to classification, regardless of their absolute frequencies. Equation (3) computes the relevance frequency, emphasizing terms more prevalent in the positive class.

3.4. Binary Term Weighting

Binary term weighting simplifies term representation by assigning a binary value (1 or 0) to each term based on its presence or absence in a document. This straightforward approach is computationally efficient and particularly suitable for short documents where term re-occurrence is rare.

3.5. Proposed Scheme

The proposed scheme introduces a novel approach that combines the relevance frequency factor with a delta (Δ) factor to compute relevance scores for each term. Equation (4) outlines this calculation, which considers both the term's class distribution and its discriminative power.

3.6. Construction of Training and Test Vectors

Training and test vectors are constructed by sorting terms based on their relevance scores and selecting a predetermined number of top-ranking terms. These binary-valued feature vectors serve as input for training and testing the classification model.

The framework's performance is evaluated using the ISEAR dataset, which includes seven basic emotion categories. Accuracy, calculated from the model's predictions, serves as the primary evaluation metric. The discussion highlights the effectiveness of each feature selection and weighting scheme, as well as the proposed scheme's contributions to improving classification accuracy.

This comprehensive framework offers a systematic approach to emotion recognition from text, leveraging various feature selection and weighting techniques to optimize classification performance.

4. Result and Discussion

The experimental results depicted in Table 1 showcase the performance of different feature selection schemes in emotion recognition from text. Across various emotion categories in the ISEAR dataset, Relevance Score consistently outperforms other schemes in terms of accuracy. Notably, Delta also achieves top accuracies in several categories, while Chi-Square exhibits superior performance in fewer categories. Conversely, Gini-Text fails to demonstrate significant improvements across any category. Analyzing the impact of the number of selected features reveals interesting insights. In categories like disgust, fear, joy, and sadness, increasing the number of features leads to improved accuracy. However, this trend is less pronounced in categories such as anger, guilt, and shame, indicating a limited number of discriminative terms within these categories. Notably, Gini-Text consistently underperforms across all categories as the number of selected terms decreases, while Relevance Score maintains strong performance even with a smaller number of features. Further comparison of the best and second-best accuracy scores obtained using different feature selection approaches highlights the effectiveness of Relevance Score in achieving top performances across multiple categories. However, the optimal number of features varies across categories. For example, while Chi-Square achieves best scores with 800 and 1000 features in joy and disgust categories, Relevance Score outperforms with just 50 and 200 features in anger and shame categories, indicating its efficiency in selecting fewer but more informative features. Analyzing the average number of terms per document in the ISEAR dataset reveals the prevalence of shorter emotion documents, with many terms occurring only once. This underscores the suitability of binary term weighting due to the low recurrence of terms in documents. Further examination of feature selection schemes' behaviors reveals that Chi-Square tends to select positive class indicative terms, while Gini-Text and Relevance Score prioritize negative class indicative terms. Notably, Chi-Square's selection heavily favors positive class terms, potentially leading to poor performance in extreme filtering scenarios due to the exclusion of negative class indicative terms. Conversely, the inclusion of comparatively rare features, particularly those asymmetrically distributed between positive and negative classes, contributes to Chi-Square's superior performance in certain categories. In summary, while Chi-Square favors positive class terms, Gini-Text and Relevance Score prioritize negative class terms. Relevance Score, in particular, selects terms indicating negative class membership to a greater extent, leading to improved performance. These findings underscore the importance of selecting features that effectively capture class distributions, with Relevance Score demonstrating notable efficacy in achieving this goal.

Table 1 Performance of different feature selection schemes

	Gini-Text	**Relevance Score**
Average(A)	140	56.54
Average(C)	153.43	67
Average A/C	0.56	0.45

5. Conclusion

A novel feature selection scheme is introduced to investigate its impact on emotion recognition from text. Through experimentation on an emotion dataset, various feature reduction techniques such as Chi-Square, Gini-Text, Delta, and Relevance Score selection schemes are compared. The One Against All (OAA) approach, employing linear Support Vector Machine (SVM) as the classifier, is utilized for binary classification. Results demonstrate that the Relevance Score selection scheme significantly enhances classification performance across multiple emotion categories. Relevance Score strikes a better balance between term occurrence frequencies and class distribution, leading to improved feature selection. However, there remain areas warranting further exploration. Alternative term weighting schemes could replace binary term weighting, potentially enhancing performance. Additionally, other feature selection methods may offer opportunities for further improvement. Furthermore, the proposed Relevance Scoring method, which computes scores based on the multiplication of relevance and delta factors, could be combined with other selection schemes for even better results. Each selection scheme has its strengths and weaknesses, particularly when dealing with varying numbers of filtered terms. By leveraging the strengths of multiple schemes, more informative subsets of terms can be obtained, thereby improving overall results.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Kim, J.-H.; Kim, B.-G.; Roy, P.P.; Jeong, D.-M.; Kima, B.-G. Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure. IEEE Access 2019, 7, 41273–41285. [CrossRef]

[2] Li, S.; Deng, W. Deep Facial Expression Recognition: A Survey. IEEE Trans. Affect. Comput. 2020. [CrossRef]

[3] Nguyen, D.H.; Kim, S.; Lee, G.-S.; Yang, H.-J.; Na, I.-S.; Kim, S.H. Facial Expression Recognition Using a Temporal Ensemble of Multi-level Convolutional Neural Networks. IEEE Trans. Affect. Comput. 2019. [CrossRef]

[4] Ekman, P. An argument for basic emotions. Cogn. Emot. 1992, 6, 169–200. [CrossRef]

[5] Cavallo, F.; Semeraro, F.; Fiorini, L.; Magyar, G.; Sincak, P.; Dario, P. Emotion Modelling for Social Robotics Applications: A Review. J. Bionic Eng. 2018, 15, 185–203. [CrossRef]

[6] De Diego, I.M.; Fernandez-Isabel, A.; Ortega, F.; Moguerza, J.M. A visual framework for dynamic emotional web analysis. Knowl. Based Syst. 2018, 145, 264–273. [CrossRef]

[7] Saldias, F.B.; Picard, R.W. Tweet Moodifier: Towards giving emotional awareness to Twitter users. In Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, UK, 3–6 September 2019; pp. 1–7. Appl. Sci. 2020, 10, 5351 11 of 13

[8] Franzoni, V.; Milani, A.; Gervasi, O.; Murgante, B.; Misra, S.; Rocha, A.M.A.; Torre, C.M.; Taniar, D.; Apduhan, B.O.; Stankova, E.; et al. A Semantic Comparison of Clustering Algorithms for the Evaluation of Web-Based Similarity Measures; Springer: Cham, Switzerland, 2016; pp. 438–452.

[9] Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep learning-based text classification: A comprehensive review. arXiv 2020, arXiv:2004.03705.

[10] Talanov, M.; Vallverdú, J.; Distefano, S.; Mazzara, M.; Delhibabu, R. Neuromodulating Cognitive Architecture: Towards Biomimetic Emotional AI. In Proceedings of the 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, Gwangiu, Korea, 24–27 March 2015; pp. 587–592.

[11] Shivhare, S.N.; Khethawat, S. Emotion Detection from Text. arXiv 2012, arXiv:1205.4944.

[12] Ishizuka, M.; Neviarouskaya, A.; Shaikh, M.A.M. Textual Affect Sensing and Affective Communication. Int. J. Cogn. Inform. Nat. Intell. 2012, 6, 81–102. [CrossRef] 13. Lövheim, H. A new three-dimensional model for emotions and monoamine neurotransmitters. Med. Hypotheses 2012, 78, 341