

CLASSIFYING TEXT-BASED EMOTIONS USING LOGISTIC REGRESSION

FAHAD MAZAED ALOTAIBI

Faculty of Computing and Information Technology in Rabigh (FCITR) King Abdul Aziz University (KAU) Jeddah
 Saudi Arabia
fmmlotaibi@kau.edu.sa

Abstract: *Emotion detection textual content is getting popular among individuals and business companies to analyze user emotional reaction on the products they use. In this work, emotion detection from textual content is performed by using supervised learning-based Logistic Regression classifier. ISEAR dataset is used to training the classifier, while testing dataset is used to evaluate the prediction capability of the classifier for emotion classification. The prior works used rule-based techniques, supported by lexical resources. However, limited coverage of emotional clues, was the major issue, which resulted in poor performance of system. The proposed work overcomes this limitation by proposing supervised learning technique using Logistic Regression classifier. The results obtained are encouraging and show that the proposed system performed better than the similar methods.*

Keywords: Emotion detection, machine learning, Logistic Regression classifier, user reviews, ISEAR dataset

1. Introduction: Cognitive computation aims at designing computational models for different mental processes like sentiments, opinions, beliefs, perception and emotions. An emotion reflects status of human mind in terms of joy, fear, sadness, hate, disgust and others [1]. Emotion classification from text faces different challenges like multiple senses of a same word, domain dependency and lack of sufficient lexical resources. Therefore, computational linguistic experts are focusing more in this direction.

Prior studies on emotion detection from text have used lexical resources, which often leads to incorrect classification due to limited coverage of emotion clues [1, 2, 3]. However, we suggest a supervised machine learning approach based on Logistic Regression (LR) with publically available emotion dataset, namely ISEAR [4]. The proposed study works on 5 emotion signals and the classifier is trained on the well annotated emotion dataset. The LR algorithm works on training and testing dataset of emotions. Before, making input to the dataset, some preprocessing steps are applied on the input text. We used different Python libraries for the implantation of the system. Performance evaluation results are encouraging and shows the effectiveness of the proposed system.

Remained of the article is organized as follows: related work is presented in section 2; materials and methods are presented in section 3; result and analysis is shown in section 4; and finally conclusion and future work is shown section 5.

2. Literature Review: Asghar et al. [1] proposed a emotion detection system for emotion detection from social media text using hybrid classification scheme

classifiers which classify the emotions detected from text, emojis and slang terms. The system's performance can be improved by increasing the size of the dataset and further experiments required to be conducted using supervised learning techniques. Shaila and Vadivel [2] proposed a emotion detection system using supervised learning technique. For this purpose, artificial neural network system is designed to isolate +ive and -ive emotion signals from text. Experimental results show that uni-gram and bigram features have played significant role in the emotion detection. Kaewyong et al. [5] proposed lexicon-based technique for analyzing student feedback automatically. More than 1000 student responses were about teachers were collected from students to evaluate the performance of teaching faculty. Different scores using polarity lexicon. The experimental results have shown the performance improvement of the proposed system over similar methods.

Asghar et al. [3] proposed an emotion detection system from health reviews generated from patient feedback. A rule-based technique is implemented to classify emotion words from text. The system has effectively classified emotion expressed by patients in the input text. However, the system can be enhanced by incorporating emojis and slang terms with supervised learning techniques.

An affective computing technique is proposed by Cambria et al. [6] to detect sentiments and classify emotions from text. The sentiment are classified into +ive, -ive and neutral classes. The emotion recognition module classifies the text into different emotion classes. The system works in sequential and pipeline mode, i.e. firstly sentiment are detected and then the detected sentiments are categorized into different emotion types.

Kollias et al. [7] proposed a deep learning based Convolutional Neural Network model for emotion detection from facial expression in video content. The system has shown promising results when compared with machine learning models and other baseline methods. The system can be enhanced by incorporating more features of human computer interaction and performance of the system can be evaluated better in real life scenario.

Li and Ren [8] developed an emotion recognition system from Chinese text using Ren-CECPs (Corpus). The system gained performance improvement over similar methods and achieved an accuracy of more than 90%. The major limitation of their system is the low coverage of lexical entries, which if extended can improve the system performance.

Davletcharova et al. [9] have worked on the identification of emotions from human speech using speech classification system enriched with different speech features. They used 30 different subjects in the dataset preparation. The results obtained have shown significant improvement w.r.t accuracy over similar methods.

A treebank corpus is used by Socher et al. [10] in their experiments conducted on the classification of user sentiment using recursive deep learning models. The sentiment are analyzed are fine-grained level. The dataset is split into training and testing chunks and an accuracy of more than 80% is achieved, showing improvement over other methods.

3. Material and Methods

The proposed method is based on Logistic Regression (LR), an efficient machine learning technique [11, 12, 13]. In first phase, LR receives data from the ISEAR emotion dataset, and then emotion classification is performed by applying LR classifier.

3.1 Data Acquisition

A public dataset, namely ISEAR [4] is used to conduct the experiments. A partial listing of training and testing datasets are presented in Table 1 and Table 2 respectively.

Table 1 A Sample list of Emotion Training dataset obtained from SEAR Dataset

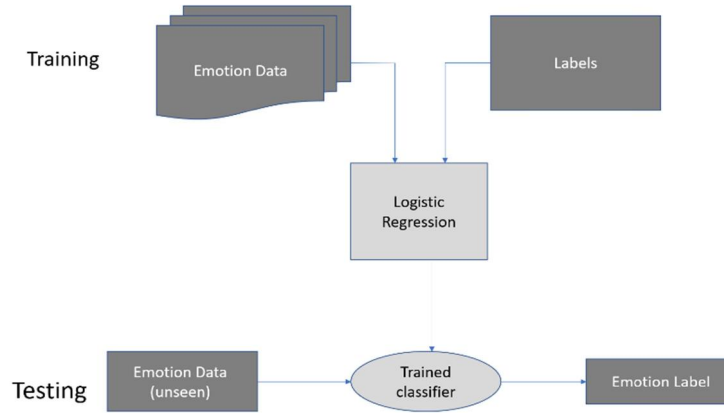
Review Text	Emotion Type
When my husband discovered I was having an affair with another man	Guilt
I was biking with my boyfriend, the sun shining after a long period of bad beather, afterwards we spent the evening and night together.	Joy
Once in a public ladies room, I opened the door of a cubicle that was taken, although it was not locked	Shame
When the relationship with my girlfriend broke up	Sadness
Some weeks ago I experienced a day that everything went wrong I broke several things I felt very ashamed for I am not a solvent	Shame
Sleeping alone in a tent, while hearing several noises such as rustling etc, I am afraid that there are people who are up to no good	Fear
When I am in an auditorium with people and I ask a silly question so that everyone starts to laugh	Shame
A friend came to tell me that her father invited me to go on holiday with them in Benidorm otherwise I could not afford to go on holiday	Joy

Table 2 A Sample list of Emotion Test dataset obtained from SEAR Dataset

Review Text	Emotion Type
At an important moment in a volleyball game we made a point and by this we won a set from a much stronger competitor	Joy
I fell down in the bus when it braked suddenly and everybody started looking at me	Shame
My father was very angry with me, and I locked me up in the shower. Then my father knocked his fist through the door, I was scared to death	Fear
When my grandmother died, her life was heavy and her last hours were terrible.	Sadness
When my father (physically handicapped) asks me to do something and I raise several objections. Often I can fix it in few moments	Guilt

3.2 Classification of Emotions from Text Using Logistic Regression

The Logistic Regression (LR) classifier is applied to classify emotions from the input text. Diagrammatic representation is shown in Fig. 1 that LR performs emotion classification using supervised machine learning approach [14, 15, 16, 17].

**Fig. 1** Text-based Emotion Classification Using Logistic Regression

Text-based emotions are detected and classified using Logistic Regression (LR) performs by taking training and testing data [18, 19, 20]. Mathematical formulation of LR algorithm (Algorithm 1) is presented as follows:

$$\text{Data} = \{(A, b_i)\}$$

$$(\text{Predict}) = \frac{1}{1 + e^{-(a_0 + a_j X)}} \quad (1)$$

a_0 = intercept(bias),
 X = input vector, and A = input coefficient

Algorithm 1. Working steps of Logistic Regression classifier

```

Suppose X denotes Explanatory vector
Suppose  $B_{i,j}$  = regression Coefficient associated with ith explanatory vector and jth emotion type
# jth outcome for jth Observation
 $P(j,i) \leftarrow B_i \cdot X_j$ 
 $i=1$ 
While ( $i \leq 4$ )
{

```

$$\text{Prob(emoi)} = \frac{e^{B_{..}Xi}}{1 + \sum e^{B_{ei}Xi}}$$

Emotion_category=choose emoi with highest probability

}

While end

3.3 Dataset Loading

In this phase, dataset is loaded, implementation detail is presented in Fig. 3

```
| #Loading Data
import pandas as pd
cat =["joy","fear","sadness","shame","guilt"]
#data = pd.read_excel("f:\\msdocuments\\isear-emotions.xlsx")
data = pd.read_excel("f:\\msdocuments\\Five_emotion.xlsx")
array = data.values
X = array[:,1]
y = array[:,0]
# Loading Stopwords
fd = open('stopwordlist.txt','r')
swlist = fd.read().lower()
#relig_prac = data[data.Relig_Prac==2]
#len(relig_prac)
len(data)
```

Fig. 3 Dataset Loading

3.4 Loading and Training Logistic Regression Classifier

We load Logistic Regression classifier using python and the implementation detail is given in Fig. 3.

```
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.linear_model import LogisticRegression

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.pipeline import Pipeline
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=42)# Building a Pipeline for Classifier on

#text_clf = Pipeline([('vect', CountVectorizer()),('tfidf', TfidfTransformer()),('clf',LogisticRegression())])

# Training a Classifier
text_clf = text_clf.fit(X_train,y_train)
```

Fig. 4 Implementation Code of LR Technique

A partial list of dataset listing is given as follows:

sadness	When my friends did not ask me to go to a New Year's party with them.
shame	When I could not remember what to say about a presentation task at an accounts meeting.
guilt	When my uncle and my neighbour came home under police escort.
joy	[On days when I feel close to my partner and other friends. When I feel at peace with myself and also experience a close contact with people whom I regard greatly.]
fear	Every time I imagine that someone I love or I could contract a serious illness, even death.
sadness	When I think about the short time that we live and relate it to the periods of my life when I think that I did not use this short time.
shame	When I realized that I was directing the feelings of discontent with myself at my partner and this way was trying to put the blame on him instead of sorting out my own feelings.
guilt	I feel guilty when when I realize that I consider material things more important than caring for my relatives. I feel very self-centered.
joy	After my girlfriend had taken her exam we went to her parent's place.
fear	When, for the first time I realized the meaning of death.
sadness	When I recently thought about the hard work it takes to study, and how one wants to try something else. When I read a theoretical book in English that I did not understand.
shame	When I was tired and unmotivated, I shouted at my girlfriend and brought up negative sides of her character which are actually not so important.

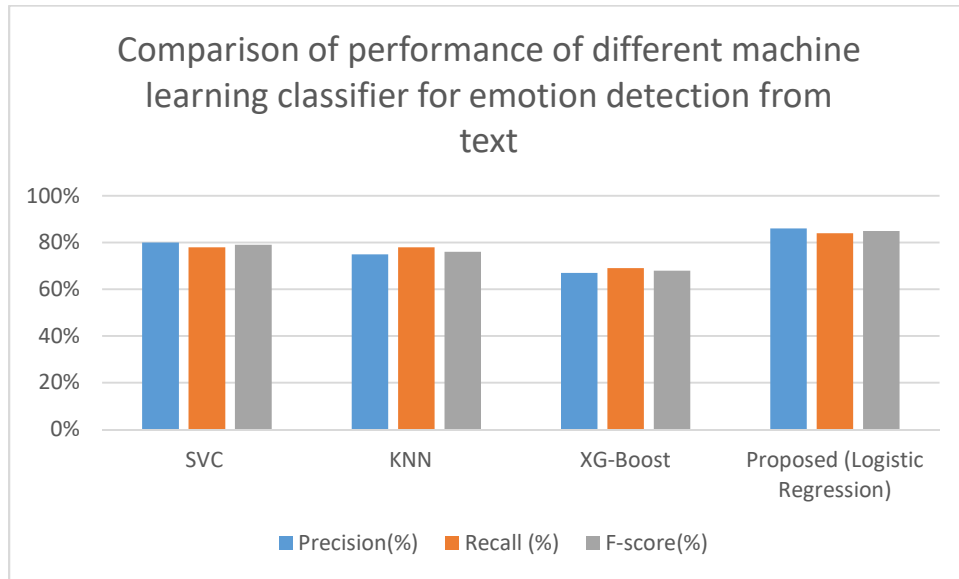


Fig. 6 Performance comparison Logistic Regression with other methods

5. Conclusion and Future Work: In this study, Logistic Regression classifier based supervised learning approach is proposed to classify text into different emotions classes. We applied ISEAR dataset in the proposed experimental setup by splitting the dataset into training and testing splits. Preprocessing was applied to clean the text prior to its feeding to the classifier. Five emotion classes were used for text classification into emotions. The proposed system produced promising results when compared with other methods. The main drawback of the system is that it needs further exploration with deep learning models, as such technique are now becoming popular in data science community due their automatic feature engineering.

REFERENCES

- [1] Asghar, M. Z., Khan, A., Bibi, A., Kundi, F. M., & Ahmad, H. (2017). Sentence-level emotion detection framework using rule-based classification. *Cognitive Computation*, 9(6), 868-894.
- [2] Shaila, S. G., & Vadivel, A. (2015). Cognitive based sentence level emotion estimation through emotional expressions. In *Progress in Systems Engineering* (pp. 707-713). Springer, Cham.
- [3] Asghar, M. Z., Khan, A., Khan, K., Ahmad, H., & Khan, I. A. (2017). COGEMO: Cognitive-Based Emotion Detection from patient generated health reviews. *Journal of Medical Imaging and Health Informatics*, 7(6), 1436-1444.
- [4] ISEAR Dataset, available at: <http://emotionresearch.net/toolbox/toolboxdatabase> Last accessed 10-nov-2018.
- [5] Kaewyong, P., Sukprasert, A., Salim, N., & Phang, A. (2015). The possibility of students' comments automatic interpret using lexicon based sentiment analysis to teacher evaluation. In *3rd International Conference on Artificial Intelligence and Computer Science (AICS2015)* (pp. 179-189).
- [6] Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2), 102-107.
- [7] Kollias, D., Tagaris, A., & Stafylopatis, A. (2016, December). On line emotion detection using retrainable deep neural networks. In *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on* (pp. 1-8). IEEE.
- [8] Li, J., & Ren, F. (2011, September). Creating a Chinese emotion lexicon based on corpus Ren-CECps. In *Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE International Conference on* (pp. 80-84). IEEE.
- [9] Davletcharova, A., Sugathan, S., Abraham, B., & James, A. P. (2015). Detection and analysis of emotion from speech signals. *Procedia Computer Science*, 58, 91-96.
- [10] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1631-1642).

- [11] Babajide Mustapha, I., & Saeed, F. (2016). Bioactive molecule prediction using extreme gradient boosting. *Molecules*, 21(8), 983.
- [12] Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2), 102-107.
- [13] Rodriguez-Pomeda, J., Casani, F., & Alonso-Almeida, M. D. M. (2017). Emotions' management within the Real Madrid football club business model. *Soccer & Society*, 18(4), 431-444.
- [14] Asghar, M. Z., Rahman, F., Kundi, F. M., & Ahmad, S. (2019). Development of stock market trend prediction system using multiple regression. *Computational and Mathematical Organization Theory*, 1-31.
- [15] Kratzwald B, Ilić S, Kraus M, Feuerriegel S, & Prendinger H (2018) Deep learning for affective computing: Text-based emotion recognition in decision support, *DECIS SUPPORT SYST*, 115, 24-35.
- [16] Madhala P, Jussila, J, Aramo-Immonen H, Suominen A (2018) Systematic Literature Review on Customer Emotions in Social Media. In *Proc. of ECSM 2018 5th European Conference on Social Media* (p. 154). Academic Conferences and publishing limited.
- [17] Singh P, Sawhney R S, Kahlon K S (2018) Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government. *ICT Express*, 4(3): 124-129.
- [18] Varathan, K. D., Giachanou, A., & Crestani, F. (2017). Comparative opinion mining: a review. *Journal of the Association for Information Science and Technology*, 68(4), 811-829.
- [19] Zhou P, Qi Z, Zheng S, Xu J, Bao H, & Xu B (2016). Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. *arXiv preprint arXiv:1611.06639*.
- [20] Syaifudin Y W PuspitasariD (2017) Twitter Data Mining for Sentiment Analysis on Peoples Feedback Against Government Public Policy. *MATTER: International Journal of Science and Technology*, 3(1):110-122.