

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

Emotion Recognition Between POMS and K-NN Classifier Algorithm Using Twitter API

Vaishnavi Burkhande¹, Gauri Lipane¹, Kajal Raut¹, Shivani Jawalkar¹, P.V.Ambekar²

B.E. Students, Department of Computer Engineering, Sinhgad Institute of Technology and Science, Narhe, Pune, India¹

Professor, Department of Computer Engineering, Sinhgad Institute of Technology and Science, Narhe, Pune, India²

ABSTRACT: The analysis of social networks is a very challenging research area while a fundamental aspect concerns the detection of user communities. The existing work of emotion recognition on Twitter specifically depends on the use of lexicons and simple classifiers on bag-of words models. The vital question of our observation is whether or not we will enhance their overall performance using deep learning algorithms. After that, we take the advantage of hashtags to create three huge emotion-categorized datasets corresponding to classifications of various emotions. We then compare the overall performance of several phrase and character-based recurrent and convolutional neural networks with the performance on bag-of-words and latent semantic indexing models. Additionally inspect the transferability of the final hidden state representations among various classifications of emotions, and whether or not it is possible to build a unison

model for predicting all of them using a shared representation. The contribution work is to apply machine learning algorithm for emotion classification, it gives less time consumption without interfere human labeling. We show that recurrent neural networks, especially character-based approaches can increase over bag-of-words and latent semantic indexing models. While the transfer capabilities of these models are poor, the novel proposed training searching produces a unison model with overall performance comparable to that of the three single models.

KEYWORDS: Emotion Recognition, Text Mining, Twitter, Recurrent Neural Networks, Convolutional Neural Networks, K-NN Classifier

I.INTRODUCTION

Emotions can be defined as conscious affect attitudes, which constitute the display of a feeling. In recent years, a large number of studies have focused on emotion detection using opinion mining on social media. Due to some intrinsic characteristics of the texts produced on social media sites, such as the limited length and casual expression, emotion recognition on them is a challenging task. Previous studies mainly focus on lexicon-based and machine learning based methods. The performance of lexicon-based methods relies heavily on the quality of emotion lexicon and the performance of machine learning methods relies heavily on the features. Therefore, we work with three classifications that are the most popular, and have also been used before by the researchers from computational linguistics and natural language processing (NLP). Paul Ekman defined six basic emotions by studying facial expressions. Robert Plutchik extended Ekman's categorization with two additional emotions and presented his categorization in a wheel of emotions.

Finally, Profile of Mood States (POMS) is a psychological instrument that defines a six-dimensional mood state representation using text mining. Previous work generally studied only one emotion classification. Working with multiple classifications simultaneously not only enables performance comparisons between different emotion categorizations on the same type of data, but also allows us to develop a single model for predicting multiple classifications at the same time.

1.1 Motivation

The system developed based on our proposed approach would be able to automatically detect what people feel about

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

their lives from twitter messages. For example, the system can recognize:

- percentage of people expressing higher levels of life satisfaction in one group versus another group,
- percentage of people who feel happy and cheerful,
- percentage of people who feel calm and peaceful, and
- percentage of people expressing higher levels of anxiety or depression.

II. RELATED WORK

The paper [1] focuses on studying two fundamental NLP tasks, Discourse Parsing and Sentiment Analysis. The development of three independent recursive neural nets: two for the key sub-tasks of discourse parsing, namely structure prediction and relation prediction; the third net for sentiment prediction. Advantages are: The latent Discourse features can help boost the performance of a neural sentiment analyzer. Pre-training and the individual models are an order of magnitude faster than the Multi-tasking model. Disadvantages are: Difficult predictions to multi-sentential text.

In [2] paper analyze electoral tweets for more subtly expressed information such as sentiment (positive or negative), the emotion (joy, sadness, anger, etc.), the purpose or intent behind the tweet (to point out a mistake, to support, to ridicule, etc.), and the style of the tweet (simple statement, sarcasm, hyperbole, etc.). There are two sections: on annotating text for sentiment, emotion, style, and categories such as purpose, and on automatic classifiers for detecting these categories. Advantages are: Using a multitude of custom engineered features like those concerning emoticons, punctuation, elongated words and negation along with unigrams, bigrams and emotion lexicons features, the SVM classifier achieved a higher accuracy. Automatically classify tweets into eleven categories of emotions. Disadvantages are: Does not summarize tweets. Does not automatically identifying other semantic roles of emotions such as degree, reason, and empathy target.

In [3] paper we i) demonstrate how large amounts of social media data can be used for large-scale open-vocabulary personality detection; ii) analyze which features are predictive of which personality dimension; and iii) present a novel corpus of 1.2M English tweets (1,500 authors) annotated for gender and MBTI. Advantages are: The personality distinctions, namely INTROVERT-EXTROVERT (I-E) and THINKING-FEELING (T-F), can be predicted from social media data with high reliability. The large-scale, open-vocabulary analysis of user attributes can help improve classification accuracy.

The paper [4] develops a multi-task DNN for learning representations across multiple tasks, not only leveraging large amounts of cross-task data, but also benefiting from a regularization effect that leads to more general representations to help tasks in new domains. A multi-task deep neural network for representation learning, in particular focusing on semantic classification (query classification) and semantic information retrieval (ranking for web search) tasks. Demonstrate strong results on query classification and web search. Advantages are: The MT-DNN robustly outperforms strong baselines across all web search and query classification tasks. Multi-task DNN model successfully combines tasks as disparate as classification and ranking. Disadvantages are: The query classification incorporated either as classification or ranking tasks not comprehensive exploration work.

In [5] paper explored an application of deep recurrent neural networks to the task of sentence-level opinion expression extraction. DSEs (direct subjective expressions) consist of explicit mentions of private states or speech events expressing private states; and ESEs (expressive subjective expressions) consist of expressions that indicate sentiment, emotion, etc., without explicitly conveying them. Advantages are: Deep RNNs outperformed previous (semi)CRF baselines; achieving new state-of-the-art results for fine-grained on opinion expression extraction. Disadvantages are: RNNs do not have access to any features other than word vectors.

III. OPEN ISSUES

The ability of the human face to communicate emotional states via facial expressions is well known, and past research has established the importance and universality of emotional facial expressions. However, recent evidence has revealed that facial expressions of emotion are most accurately recognized when the perceiver and expresser are from the same

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

cultural in group. Paul Ekman explains facial expressions to define a set of six universally recognizable basic emotions: anger, disgust, fear, joy, sadness and surprise. Robert Plutchik defined a wheel-like diagram with a set of eight basic, pairwise contrasting emotions; joy –sadness, trust – disgust, fear – anger and surprise – anticipation. Consider each of these emotions as a separate category, and disregard different levels of intensities that Plutchik defines in his wheel of emotions.

Disadvantages:

A. Ekman's Facial expressions limitations:

1. Image quality

Image quality affects how well facial-recognition algorithms work. The image quality of scanning video is quite low compared with that of a digital camera.

2. Image size

When a face-detection algorithm finds a face in an image or in a still from a video capture, the relative size of that face compared with the enrolled image size affects how well the face will be recognized.

3. Face angle

The relative angle of the target's face influences the recognition score profoundly. When a face is enrolled in the recognition software, usually multiple angles are used (profile, frontal and 45-degree are common).

4. Processing and storage

Even though high-definition video is quite low in resolution when compared with digital camera images, it still occupies significant amounts of disk space. Processing every frame of video is an enormous undertaking, so usually only a fraction (10 percent to 25 percent) is actually run through a recognition system.

B. Plutchik's algorithm limitations:

1. The FPGA Kit uses hardware that is expensive. Thus, making this approach a cost ineffective technological solution.
2. Also, there is an additional dimension which involves a lot of tedious calculations.

IV. PROPOSED METHODOLOGY

Profile of Mood States is a psychological instrument for assessing the individual's mood state. It defines 65 adjectives that are rated by the subject on the five-point scale. Each adjective contributes to one of the six categories. For example, feeling annoyed will positively contribute to the anger category. The higher the score for the adjective, the more it contributes to the overall score for its category, except for relaxed and efficient whose contributions to their respective categories are negative. POMS combines these ratings into a six-dimensional mood state representation consisting of categories: anger, depression, fatigue, vigour, tension and confusion. Comparing to the original structure, we discarded the adjective blue, since it only rarely corresponds to an emotion and not a color, and word-sense disambiguation tools were unsuccessful at distinguishing between the two meanings. We also removed adjectives relaxed and efficient, which have negative contributions, since the tweets containing them would represent counter-examples for their corresponding category.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

A. Architecture

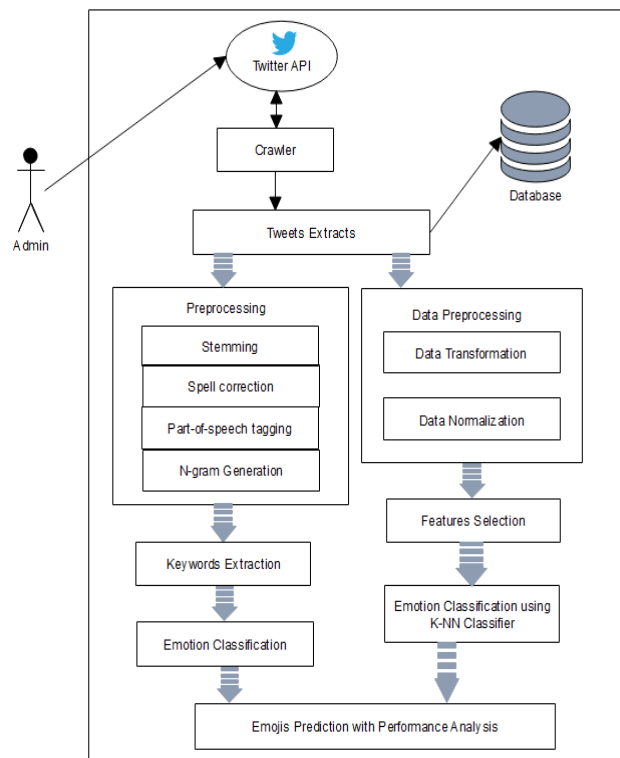


Fig. 1 System Architecture

Contribution of this paper is to implement machine learning algorithm gives less time consumption without interfere human labeling. The k-NN classifier works on testing dataset with help of huge amount of training dataset. It gives same result as POMS tagging methods. The contribution work is prediction of Emojis for emotion recognition using tweet contents.

Advantages are:

- Increases human-computer interactions
- Low-cost
- Fast emotion recognition system
- Scalable
- Comparable quality to experts

B. Algorithms

1: Indexing By Latent Dirichlet Allocation (LDA) Algorithm

First and foremost, LDA provides a generative model that describes how the documents in a dataset were created. In this context, a dataset is a collection of D documents. Document is a collection of words. So our generative model describes how each document obtains its words. Initially, let's assume we know K topic distributions for our dataset, meaning K multinomials containing V elements each, where V is the number of terms in our corpus. Let β_i represent the multinomial for the i th topic, where the size of β_i is V : $|\beta_i| = V$. Given these distributions, the LDA generative process is as follows:

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

Steps:

1. For each document:

(a) Randomly choose a distribution over topics (a multinomial of length K)

(b) for each word in the document:

(i) Probabilistically draw one of the K topics from the distribution over topics obtained in (a), say topic_{β_j}

(ii) Probabilistically draw one of the V words from β_j

2. Sentiment Analysis using Sentiwordnet Dictionary

```
polarizedTokensList ← newList()
```

```
while tokenizedTicket.hasNext() do
```

```
token ← tokenizedTicket.next()
```

```
lemma ← token.lemma
```

```
polarityScore ← null
```

```
if DomainDictionary.contains(lemma, pos) then
```

```
if SentiWordNet.contains(lemma, pos) and
```

```
SentiWordNet.getPolarity(lemma, pos) != 0 then
```

```
polarityScore ← SentiWordNet.getPolarity(lemma, pos)
```

```
else
```

```
domainDicToken ← DomainDictionary.getToken(lemma, pos)
```

```
if domainDicToken.PolarityOrientation == "POSITIVE" then
```

```
polarityScore ← DefaultPolarity.positive
```

```
else
```

```
polarityScore ← DefaultPolarity.negative
```

```
end if
```

```
end if
```

```
polarizedTokensList.add(token, polarityScore)
```

```
end if
```

```
end while
```

```
return polarizedTokensList
```

3. K-NN Classifier Algorithm

Steps:

1. Load the data

2. Initialize the value of k

3. For getting the predicted class, iterate from 1 to total number of training data points

4. Calculate the distance between test data and each row of training data. Here we will use Euclidean distance as our distance metric since it's the most popular method. The other metrics that can be used are Chebyshev, cosine, etc.

5. Sort the calculated distances in ascending order based on distance values

6. Get top k rows from the sorted array

7. Get the most frequent class of these rows

8. Return the predicted class

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

V.RESULT AND DISCUSSIONS

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GBmemory, Windows 7, MySQL 5.1 backend database and Jdk1.8. The application is web application used tool for designcode in Eclipse and execute on Tomcat server. Some functionsused in the algorithm are provided by list of jars like Twitter-coreand Twitter-stream jars etc.

Tweets are retrieved in a streaming way, and Twitter providesthe Streaming API for developers and researchers toaccess public tweets in real time. The aim of this paper isto bridge the gap by carrying out a performance evaluation,which was from two different aspects of NLP and machinelearning algorithms. The Unison model is the combination ofEkman’s, Plutchik’s and POMS emotion categories and the k-NN Classifier algorithm uses for emotionrecognition performance. Fig. 2 shows total emotion identified using unison model and fig. 3 showing the accuracy as well as execution time for performance of Unison Model.

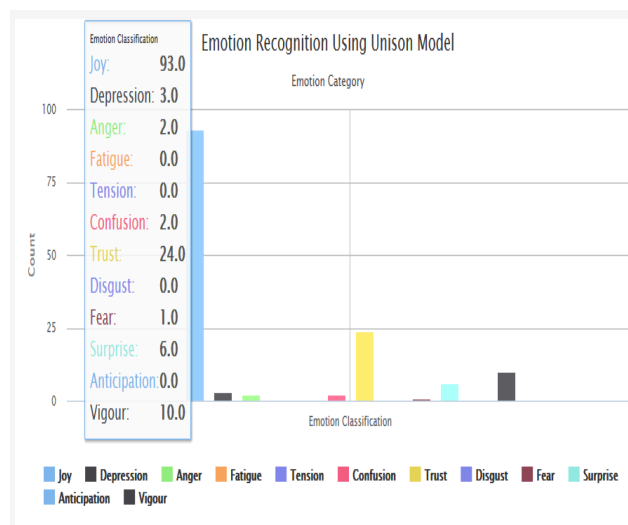


Fig. 2 Emotion detection using Unison model

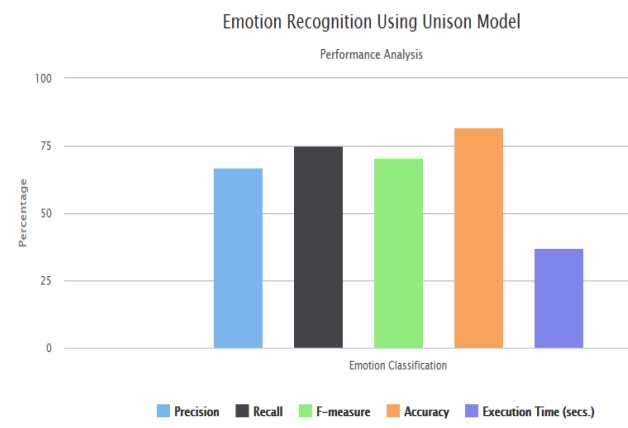


Fig. 3 Performance of emotion detection

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

Fig. 4 shows the emotion classification with the help of k-NN algorithm. The performance of emotion classification of k-NN classifier algorithm shown in fig. 5.

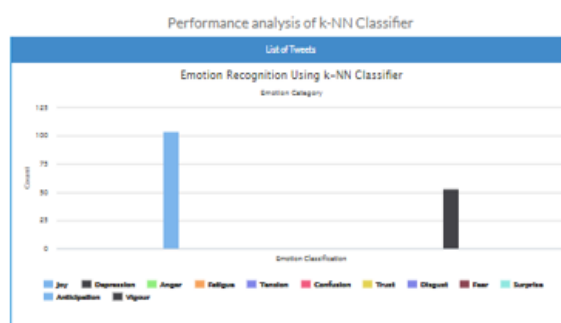


Fig. 4 Emotion recognition using k-NN Algorithm



Fig. 5 Performance Analysis of k-NN Algorithm

As compared to unison model and k-NN classifier algorithm, it gives results better than unison model within short time period.

VI. CONCLUSION

This paper as a first study working on predicting POMS's categories, we believe they are as predictable as Ekman's and Plutchik's. We also showed that searching for tweets containing POMS adjectives and later grouping them according to POMS factor structure yields a coherent data set whose labels can be predicted with the same accuracy as other classifications. We proposed an alternative training strategy that samples training instances based on the difference between train and validation accuracy and showed that it improves over alternating strategy. This project shows result of three Ekman's, Plutchik's and POMS Algorithm of unison model of emotion detection. As well as perform the machine learning K-NN classifier algorithm for emotion detection. Finally showing the accuracy is as compared to unison model of K-NN classifier algorithm. And it gives result better than unison model within short time period.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

REFERENCES

- [1] B. Nejat, G. Carenini, and R. Ng, “Exploring Joint Neural Model for Sentence Level Discourse Parsing and Sentiment Analysis,” Proc. of the SIGDIAL 2017 Conf., no. August, pp. 289–298, 2017.
- [2] S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, “Sentiment, emotion, purpose, and style in electoral tweets,” Information Processing and Management, vol. 51, no. 4, pp. 480–499, 2015.
- [3] B. Plank and D. Hovy, “Personality Traits on Twitter —or— How to Get 1,500 Personality Tests in a Week,” in Proc. of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2015, pp. 92–98.
- [4] X. Liu, J. Gao, X. He, L. Deng, K. Duh, and Y.-Y. Wang, “Representation Learning Using Multi-Task Deep Neural Networks for Semantic Classification and Information Retrieval,” Proc. of the 2015 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 912–921, 2015.
- [5] O. Irsoy and C. Cardie, “Opinion Mining with Deep Recurrent Neural Networks,” in Proc. of the Conf. on Empirical Methods in Natural Language Processing. ACL, 2014, pp. 720–728.