Multi-label Emotion Classification using Content-Based Features in Twitter

Iqra Ameer\(^1\), Noman Ashraf\(^1\), Grigori Sidorov\(^1\), Helena Gómez Adorno\(^2\)

\(^1\) Instituto Politécnico Nacional, Centro de Investigación en Computación, Mexico
\(^2\) Universidad Nacional Autónoma de México, Instituto de Investigación en Matemáticas Aplicadas y en Sistemas, Mexico

iqraameer133@gmail.com, nomanashraf@sagitario.cic.ipn.mx, sidorov@cic.ipn.mx, helena.gomez@iimas.unam.mx

Abstract. Multi-label Emotion Classification is a supervised classification problem that aims to classify multiple emotion labels from a given text. Recently, Multi-label Emotion Classification has appealed to the research community due to possible applications in E-learning, marketing, education, and health care, etc. We applied content-based methods (words and character n-grams) on tweets to show how our purposed content-based method can be used for the development and evaluation of the Multi-label Emotion Classification task. The results achieved after our extensive experimentation demonstrate that content-based word unigram surpassed other content-based features (Multi-label Accuracy = 0.452, Micro\(F_1\) = 0.573, Macro\(F_1\) = 0.559, Exact Match = 0.141, Hamming Loss = 0.179).

Keywords. Multi-label emotion classification, content-based methods, twitter.

1 Introduction

Multi-label Emotion Classification (MEC) is a supervised classification task to determine the presence of multiple emotions in a given piece of text. Single-label Emotion Classification captures only one emotion in text whereas MEC captures all present emotions which best represent the mental state of the writer. MEC has captured the intention of the research community due to its potential applications in various domains including, E-learning, marketing, education, health care [15], etc.

Due to the current boom in data creation and the usage of smartphones by people which results in them expressing their thoughts over public social media communication forums, such as Facebook, Twitter, YouTube, etc. There is no doubt that humans are emotional creatures, emotions are an important part of human life, and affect our choices as well as our psychological and physical health. In this case, the implementation of MEC systems is necessary. There are two types of automatic Emotion Classification: (1) Single-label Automatic Emotion Classification — associate a single label to the given instance from the finite set of pre-defined labels, which best describes the emotion in the given instance, and (2) Multi-label Automatic Emotion Classification — associate multiple labels to the given instance from the finite set of pre-defined labels, which best describe the mental state of the author.

MEC on Twitter is a challenging task as only 280 characters are allowed and users have to use informal language structure such as a short form of texts, emojis’ etc. Another reason is a tweet may contain more than one emotion label in a given content.

This study describes our development process of the automatic MEC model for tweets. We used content-based...
methods particularly word n-grams and character n-grams as features along with their combinations. We used MEKA\(^2\) implementation of four multi-label and 11 single-label classifiers. Methods were evaluated by using the tweets dataset of SemEval-2018\(^3\) competition. The remaining article is structured as follows: Section 2 represents the related work. Section 3 briefs the task description and corpus provided by the SemEval-2018. Section 4 describes the evaluation methodology. Section 5 concludes the article and proposes future work.

2 Related Work

Lately, scientists have shown significant attention to MEC in the textual content. In this section, we explore the prior work of this domain. The A-A is an unsupervised emotion classification model\(^1\) that characterizes the content using unigrams with a multi-label classifier such as Label Powerset (LP) to detect emotions from suicide notes\(^10\) and 15 emotions detected such as anger, joy, guilt, and love, etc. To detect emotions from Brazilian Portuguese short texts, a lot of multi-label classifiers used for example RAkEL and HOMER\(^1\).

The Semantic Evaluation series (SemEval-2018)\(^4\) played an important role in the emotion Classification task. In task 4 SemEval-2007 competition, the organizers provided the news headlines and asked to classify the polarity and emotions\(^19\). The rule-based system UPAR7 that uses dependency graphs performed best out of all three participants\(^7\). In task 5 SemEval-2018 contest, the Maximum number of researchers used lexica, word n-grams or word embeddings along with Deep Learning (DL) based models such as Convolution Neural Network (CNN), Recurrent Neural Network (RNN) or Long-Short Term Memory Network (LSTM) architectures to classify multiple emotions from the text. The best performing team (NTUA-SLP) implemented Bidirectional LSTM (Bi-LSTM) with a multilayer self-attention mechanism\(^12\), as Bi-LSTM used to perform well in classification task\(^\,\,\,2\).

3 Task Description and Dataset

MEC on Twitter at SemEval-2018 had corpora for the Arabic, English, and Spanish languages. However, for this study, we have chosen only the English language.

3.1 Task Description

The MEC task: A tweet is given, determine the tweet as “neutral or no emotion” or one, or more from a set of 12 following emotion labels that best show the emotional condition of the author:

- Anger (also frustration and wrath),
- Happy (also love, joy, joyfulness, delight, glad, pleased, satisfied, elated),
- Sadness (also grief, sorrow, pity, melancholy, despair, depth of feeling),
- Shock (also surprise),
- Fears (also anxiety, fear, terror, panic),
- Resentment (also contempt, disdain, irritation, anger, anger at, irritation at, annoyance, annoyance at),
- Pride (also glory, pride, self-esteem, self-respect, self-confidence, self-reliance, self-importance, self-resilience),
- Despondency (also despondency, despondent, dejected, depressed, despair, discouraged, dispirited, disheartened, discouraged, dispirited, disheartened),
- Hope (also hope, hopeful, expectant, expectant of, looking forward to, looking forward to, hopeful of, hopeful of),
- Love (also love, love for, affectionate, loveable, dear, endearing, endearing, appealing, lovable, lovable),
- Interest (also interest in, interest of, interesting, interesting in, interesting in, fascinating, fascinating in, fascinating in, captivating, captivating in),
- Relaxed (also calm, peaceful, tranquil, serene, tranquil, serene, peaceful, calm, at peace, at peace, at ease).

With emotions, word n-grams, and elongated words for the emotion detection task.

Table 1. Example of Multi-label Tweet Dataset

<table>
<thead>
<tr>
<th>No.</th>
<th>Tweet</th>
<th>Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>@NHLexpertpicks @usahockey USA was embarrassing to watch. When was the last time you guys won a game..? #horrible #joke</td>
<td>anger, disgust</td>
</tr>
<tr>
<td>02</td>
<td>My sister is graduated with 3.9 CGPA</td>
<td>happy, love, surprise</td>
</tr>
</tbody>
</table>

\(^{1}\)http://waikato.github.io/meka/

\(^{2}\)https://competitions.codalab.org/competitions/17751

\(^{3}\)https://semeval.github.io/
Table 2. Percentage of Tweets that were annotated with a given emotion

<table>
<thead>
<tr>
<th>Language</th>
<th>anger</th>
<th>anti.</th>
<th>disg.</th>
<th>fear</th>
<th>joy</th>
<th>love</th>
<th>optim.</th>
<th>pessi.</th>
<th>sadn.</th>
<th>surp.</th>
<th>trust</th>
<th>neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eng</td>
<td>36.1</td>
<td>13.9</td>
<td>36.6</td>
<td>16.8</td>
<td>39.3</td>
<td>12.3</td>
<td>31.3</td>
<td>11.6</td>
<td>29.4</td>
<td>5.2</td>
<td>5.0</td>
<td>2.7</td>
</tr>
</tbody>
</table>

— Anticipation (also concern and attention),
— Disgust (also disregard and dislike),
— Fear (also nervousness, worry, and panic),
— Joy (also peacefulness and delight),
— Love (also affection),
— Optimism (also hope and self-assurance),
— Pessimism (also distrust, no assurance),
— Sadness (also sorrow, unhappiness),
— Surprise (also shock, amazement),
— Trust (also liking, approval, and acceptance).
— Neutral or no emotion.

**4.1 Pre-processing**

Before feature extraction process, we applied following pre-processing steps \(^3\)\(^18\) on dataset of SemEval-2018:

— Lower-cased the tweets,
— Punctuation marks are removed,
— Stop words are removed,
— Normalized the elongated words.

**4.2 Evaluation Measures**

To evaluate this study, we used official evaluation measures of SemEval-2018 task: (i) Accuracy (multi-label), (ii) Micro-averaged F\(_1\) score, (iii) Macro-averaged F\(_1\) score\(^5\) along with official measure, we also reported results with:

— **Exact Match**: Calculates the percentage of samples whose predicted labels are the same as their gold labels.
— **Hamming Loss**: Calculates the average of how many times a label of an instance is classified incorrectly.

**4.3 Features**

The pre-processed text was used to generate the features for single-label and multi-label ML algorithms. In this study, we applied Content Based Method, which aims to classify the author’s emotions by analysis content of the message. We applied 10 content based features – 3 word-based features and 7 character-based features.

We also performed experiments with combinations of word n-grams and combination for character n-grams as they supposed to improve the results in classification problems \(^4\)\(^17\). \(N\) varies from 1-3 for word-based features and 3-9 for character-based features. We weighted the features by well-known TF-IDF scores by using Scikit-learn\(^6\) implementation.

---

\(^3\) https://competitions.codalab.org/competitions/17751

\(^4\) https://scikit-learn.org/

---
Table 3. Best results obtained using content based methods

<table>
<thead>
<tr>
<th>Features</th>
<th>MLC</th>
<th>SLC</th>
<th>Acc.</th>
<th>EM</th>
<th>HL</th>
<th>Mi-F1</th>
<th>Ma-F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word 1-gram</td>
<td>BR</td>
<td>RF</td>
<td>0.452</td>
<td>0.141</td>
<td>0.179</td>
<td>0.573</td>
<td>0.559</td>
</tr>
<tr>
<td>Word 2-gram</td>
<td>BR</td>
<td>DT</td>
<td>0.366</td>
<td>0.053</td>
<td>0.200</td>
<td>0.515</td>
<td>0.496</td>
</tr>
<tr>
<td>Word 3-gram</td>
<td>LC</td>
<td>SMO</td>
<td>0.308</td>
<td>0.136</td>
<td>0.237</td>
<td>0.373</td>
<td>0.363</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Character N-grams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Char 3-gram</td>
</tr>
<tr>
<td>Char 4-gram</td>
</tr>
<tr>
<td>Char 5-gram</td>
</tr>
<tr>
<td>Char 6-gram</td>
</tr>
<tr>
<td>Char 7-gram</td>
</tr>
<tr>
<td>Char 8-gram</td>
</tr>
<tr>
<td>Char 9-gram</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Combination of Word N-grams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word 1-3-gram</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Combination of Character N-grams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Char 3-9-gram</td>
</tr>
</tbody>
</table>

4.4 Single-label and Multi-label Machine Learning Algorithms

We have an MEC task, and the aim is to classify multiple emotions among 12 emotions. In this paper, we tried MEKA implementation of various machine learning multi-label along with single-label classifiers however, we are reporting results with only best-performing algorithms. Multi-label classifiers including Binary Relevance (BR), BPNN, Classifier Chain (CC), and Label Combination (LC) and single-label classifiers involving BayesNet, SGD, SMO, Voted Perceptron, AdaBoostM1, AttributeSelectedClassifier(ASC), Bagging, FilteredClassifier(FC), DecisionTable(DT), J48, and RandomForest(RF).

5 Results and Analysis

In Table 3 only the best scores are represented. From Table 3, we can notice that the best results are achieved on word-unigram feature by using multi-label Binary Relevance and single-label Random Forest classifiers (Multi-label Accuracy = 0.452, MicroF1 = 0.573, MacroF1 = 0.559, Exact Match = 0.141, Hamming Loss = 0.179). The combination of word N-grams (n = 1-3) also performing the same as word unigrams with a slight difference of 0.001.

Character n-grams are not performing well to classify emotions. Therefore we can say that, for all considered evaluation measures, word 1-gram (content-based feature), together with Binary Relevance and Random Forest classifiers is helpful in the MEC problem.

For multi-label classifiers, Binary Relevance is performing better as compared to the other four multi-label classifiers. The behavior of a single-label classifier is fluctuating with the combination of different multi-label classifiers. This indicates that the performance of the Binary Relevance is dependent on a single-label machine learning classifier or vice versa.

6 Conclusion and Future Work

In this article, we tackled MEC as a supervised classification problem. We considered 12 emotion labels as in SemEval-2018 Task-1 (see section 3.1). We implemented content-based methods (words and character n-grams) and performed extensive experimentation. Results show that word unigrams are better as a feature to classify multiple emotions from a given tweet (Multi-label Accuracy = 0.452, MicroF1 = 0.573, MacroF1 = 0.559, Exact Match = 0.141, Hamming Loss = 0.179).

To handle the MEC problem, the combination of multi-label Binary Relevance with a single-label Random Forest classifier is beneficial.
Forest classifier is best. The possible future work is as follows: we plan to explore neutral deep learning for the emotion classification task, for example, Bi-LSTM, attention mechanism, self-attention.

We will also see how stylometry-based methods behave on an emotion classification problem. We can also increase the dataset by using data augmentation techniques to see the behaviors of different classification methods on the MEC problem [5].

Acknowledgements

This work is done with partial support of CONACYT, SNI, CONACYT project A1-S-47854, SIP projects 20200797 and 20200859.

References


