Emotion Analysis On Text

Bhagyashree

Directed by: Prof. U. S. Tiwary

Indian Institute of Information Technology, Allahabad

May 16, 2019

Overview

- Introduction
- Problem Definition and Objective
- Methodology
- Implementation
- Future Scope of Project
- References

What are Emotions?

It is difficult to answer the question "How many emotions we have?" But Some of the common emotions we express in our daily lives include Joy, Happiness, Fear, Sadness, Surprise, Anger and so on...which combine to form other emotions.

For example:

Excitement = Joy and Surprise

Disgust = Anger and Sadness

Emotional state of a person can be recognized by:

- Speech
- Facial expressions
- Text

Problem statement and Objective

Problem statement

I aim to make machines understand and detect emotions from any textual input.

- Text based emotion recognition system
- To give an overall sentiment of the sentence/paragraph in terms of positivity/negativity and neutrality
- Determine the overall emotion(Happy, Sad etc)

Objective

Given any textual input such as a word/phrase/sentence the machine should be able to give a percentage of various emotions that constitutes the input.

Emotion Detection Methods

Methods

There are three basic methods to detect emotions from text:

1. Keyword based detection
2. Learning-based detection
3. Hybrid detection

Tools

For syntactic structure of the word, Continuous Bag of Words(CBOW) and skip-gram architecture is being used. Concept of Word2vec(Developed by Google) is used.

Keyword based detection

Procedure

1. nltk(natural language toolkit) package is used 2. Assign Emtion labels(positive or negative) 3. Removal of unnecessary words(stop words) 4. Stemming and creation of data frame

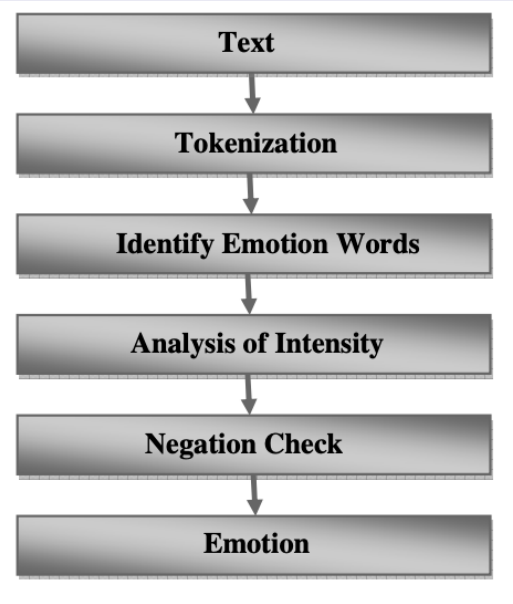

Figure: Keyword Spotting

Learning based detection

Procedure

1. Detect emotions based on a previously trained classifier
2. Word2vec tool is used.
3. To get vectors of words, Continuous Bag of Words and skip-gram architecture is used.
4. Machine learning algorithms such as Naive Bayes classifier, support vector machines and k means clustering is applied

Hybrid Detection

Procedure

1. Combination of 2 methods such as Keyword based and learning based
2. Emotions are detected based on the combination of detected keywords, learned patterns and other required information.

Removal of Stop words and pos tagging

Removal of stop words

1. Stop words are a set of commonly used words in any language.
2. These are few unnecessary words in the sentence, which needs to be removed.
3. Reduces memory overhead
4. Can potentially improve power of prediction

Pos tagging

1. Not all words contribute to the level emotion in a sentence.
2. Certain parts of speech specifically contribute to the emotion level in a sentence.
3. It tokenizes as well as indicate various parts of speech.

Algorithm(SVM)

In [5]:

```
#comprehensive cleaning
def cleaning(text):
    txt = str(text)
    txt = re.sub(r"http\S+", "", txt)
    if len(txt) == 0:
        return 'no text'
    else:
        txt = txt.split()
        index = 0
        for j in range(len(txt)):
            if txt[j][0] == '@':
                index = j
        txt = np.delete(txt, index)
        if len(txt) == 0:
            return 'no text'
        else:
            words = txt[0]
            for k in range(len(txt)-1):
                words+= " " + txt[k+1]
            txt = words
            txt = re.sub(r'[^\w]', ' ', txt)
            if len(txt) == 0:
                return 'no text'
            else:
                txt = ''.join(''.join(s)[:2] for _, s in itertools.groupby(txt))
                txt = txt.replace("'", "")
                txt = nltk.tokenize.word_tokenize(txt)
                #data.content[i] = [w for w in data.content[i] if not w in stopset]
                for j in range(len(txt)):
                    txt[j] = lem.lemmatize(txt[j], "v")
                if len(txt) == 0:
                    return 'no text'
                else:
                    return txt
```

Figure: Data pre-processing

Algorithm(SVM)

In [72]:

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import classification_report
from sklearn import svm
from sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test = train_test_split(data.content, data.sentiment, test_size=0.25, random_state=0)

x_train = x_train.reset_index(drop = True)
x_test = x_test.reset_index(drop = True)

y_train = y_train.reset_index(drop = True)
y_test = y_test.reset_index(drop = True)

vectorizer = TfidfVectorizer(min_df=3, max_df=0.9)

train_vectors = vectorizer.fit_transform(x_train)
test_vectors = vectorizer.transform(x_test)

model = svm.SVC(kernel='linear')
model.fit(train_vectors, y_train)
predicted_sentiment = model.predict(test_vectors)

print(classification_report(y_test, predicted_sentiment))
predicted_sentiments = []
for s in range(len(predicted_sentiment)):
    predicted_sentiments.append(predicted_sentiment[s])

prediction_df = pd.DataFrame({'Content':x_test, 'Emotion_predicted':predicted_sentiment, 'Emotion_actual': y_test})
prediction_df.to_csv('emotion_recognizer_svm.csv', index = False)

elapsed_time = time.time() - start_time
print ("processing time:", elapsed_time, "seconds")
```

Figure: Training Model

Accuracy

```
In [73]:
1 from sklearn.metrics import accuracy_score, confusion_matrix
2
3 cm = confusion_matrix(predicted_sentiment, y_test)
4 print(cm)
5
6 print(accuracy_score(y_test, predicted_sentiment))
```

```
[[  0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   1   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   1   0   0]
 [  0   0   1   0   0   0   6   0   3   0   6   1   5]
 [  0   0   0   0   0   1   0   0   0   0   0   1   1]
 [  1   4   7   3   4  17  12   3  67   6  46  15  64]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   1   4   0   4  13  14  10  40   4  87  11  71]
 [  0   0   0   0   0   0   0   0   0   0   0   0   1]
 [  3   3   6  13  11  24  39  22 145  17 149  37 245]]
0.3248
```

Figure: Result

Future scope of project

Future work

1. Emotion detection is a growing research field. A lot of work has already been done on sentiment analysis, facial recognition and speech recognition.
2. My project aims to recognize emotions in a word/sentence or paragraph.
3. Determining overall impact of sentence can help in various fields specially in marketing. Consumers state of mind can be read by analysing what they write.
4. More features(emotions) can be added to detect separately like recognizing sarcasm.

Chopade, C. R. (2013).
Text based emotion recognition: A survey.

Haji Binali, C. w. (2010).
Computational approaches for emotion detection in text.

V V Ramalingam1, A. P. (2018).
Emotion detection from text.