

Sentiment Analysis of movie reviews using Microsoft Text Analytics and Google Cloud Natural Language API

S Jaswanth Sai Chowdary, Sk Waseem Farooq, Sk Abubakar Siddique, S Dinesh

Department of CSE, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India

ABSTRACT

Sentiment analysis is focused on the extraction of emotions and opinions of the people towards a particular topic from a structured, semi-structured or unstructured textual data. In this paper we focus our task of sentiment analysis on The Movie DB (TMDB) database which is an online database and contains information of movies, television shows and video games including cast, crew and reviews. We examine the sentiment expression to classify the polarity of the movie review on a scale of 1 (highly disliked) to 5 (highly liked) and perform feature extraction and ranking and use these features to train our multilabel classifier to classify the movie review into its correct label. We will collect the reviews for various movies and perform classification on them. The classification for the reviews are done based on score obtained for the text. We are using two approaches to generate scores for the reviews. These approaches are Microsoft cognitive services and Google cloud natural language API. Finally we are able to generate the scores and can classify them based on movie reviews.

Keywords : Sentiment Analysis, TMDB database, Microsoft cognitive services, Google cloud natural language API, classification

I. INTRODUCTION

Now a days Internet has become a huge Cyber Database which hosts gigantic amount of data which is created and consumed by the users. The database has been growing at an exponential rate giving rise to a new industry filled with it, in which users express their opinions across channels such as Rotten Tomatoes, Facebook, Twitter, and Foursquare. Opinions which are being expressed in the form of reviews provide an opportunity for new explorations to find collective likes and dislikes of cyber community. One such domain of reviews is the domain of movie reviews which affects everyone from writers, users, film critics to the production company. The movie reviews being posted on the

websites are not formal reviews but are rather very informal and are unstructured form of grammar. Opinions expressed in movie reviews give a very true reflection of the emotion that is being conveyed. The presence of such a great use of sentiment words to express the review inspired us to derive an approach to classify the polarity of the movie using these sentiment words. Sentiment Analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral. Sentiment Analysis is a technology that will be very important in the next few years. With opinion mining, we can distinguish poor content from high quality content. With

thetechnologies available we can know if a movie has more good opinions than bad opinions and find the reasons why those opinions are positive or negative. Most sentiment analysis studies are now focused on social media sources such as TMDb, IMDB [1], Twitter [2] and Facebook, requiring the approaches be tailored to serve the rising demand of opinions in the form of text. Furthermore, performing the phrase-level analysis of movie reviews proves to be a challenging task.

II. METHODS AND MATERIAL

Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be their judgment or evaluation, affective state which is the emotional state while writing, or the intended emotional communication which is the emotional effect the writer wishes to have on the reader. Sentiment is defined as any kind of emotion and in the context of sentiment analysis is the opinion that is being expressed in the form of text or speech.

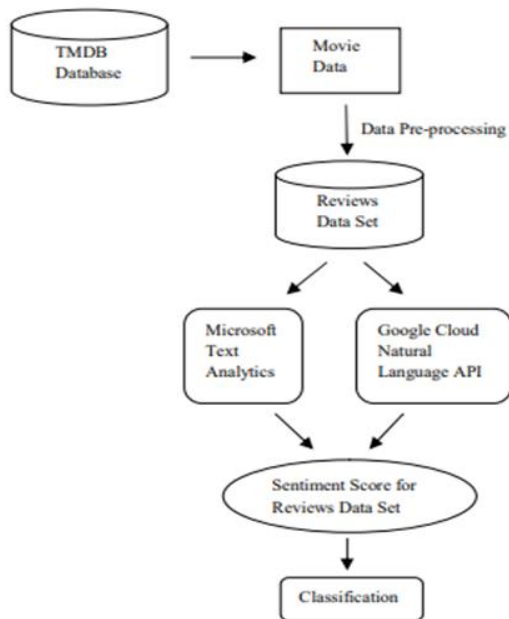


Figure 1 : Process Model Diagram

A. Feature Selection

Most researchers apply standard feature selection in their approach to improve computational performance with a handful using more sophisticated approaches. Papers focusing entirely on feature selection to improve sentiment analysis are few. We fetched the review data based on content keyword by eliminating unwanted fields from TMDb [3] database. Based on the movie id data related movie is generating which is shown in below figure.

```

Array
(
  [0] => Array
  (
    [author] => Gialy
    [content] => The reviews for _Black Panther_ are all essentially saying the same thing: This is a great movie because it's so different from anything we've seen before, not j
    _final rating:**** - I really liked it. Would strongly recommend you give it your time...
    [id] => 5a844ef3a3862a180c2f
    [url] => https://www.themoviedb.org/review/5a844ef3a3862a180c2f
  )
  [1] => Array
  (
    [author] => hankster3000
    [content] => Overrated and overhyped. Definitely avoid. Watch on redbox if you must.
    [id] => 5a87855a6eb2a33a047d32
    [url] => https://www.themoviedb.org/review/5a87855a6eb2a33a047d32
  )
  [2] => Array
  (
    [author] => Hvide Queen43
    [content] => This is definitely one of Marvel's best because the story is compelling and the characters are well developed. Black Panther and his friends and family must defe
    [id] => 5a88949a6eb2a33a047d32
    [url] => https://www.themoviedb.org/review/5a88949a6eb2a33a047d32
  )
  [3] => Array
  (
    [author] => Cressor
    [content] => Really good movie. Good story, lots of fun. This is what DC is missing.
    [id] => 5a8af4c70eb2a33a047d32
    [url] => https://www.themoviedb.org/review/5a8af4c70eb2a33a047d32
  )
  [4] => Array
  (
    [author] => Per Gunnar Jonsson
    [content] => It was with some hesitation that I went to see this movie. I was afraid that it would be burdened with too much political and social preaching. Especially know
  )
)
  
```

Figure 2 : Movie data send as response by TMDb database

Here details about the author of the movie, id of user and other unnecessary fields are eliminated. The review given by user is with key content. The final reviews are gathered from users by eliminating the unwanted fields which is shown in the below figure.

```

Array
(
  [reviews] => Array
  (
    [0] => The reviews for _Black Panther_ are all essentially saying the same thing: This is a great movie because it's so different from anything we've seen before, not j
    _final rating:**** - I really liked it. Would strongly recommend you give it your time...
    [1] => Overrated and overhyped. Definitely avoid. Watch on redbox if you must.
    [2] => This is definitely one of Marvel's best because the story is compelling and the characters are well developed. Black Panther and his friends and family must defe
    [3] => Really good movie. Good story, lots of fun. This is what DC is missing.
    [4] => It was with some hesitation that I went to see this movie. I was afraid that it would be burdened with too much political and social preaching. Especially know
    I actually found the movie fairly entertaining. It is far from great, certainly very far from the insane hype surrounding it, but pretty okay. It is more or less a standard issue :
    The story itself is a traditional revenge story with some black power stuff. Nothing to write home about but it works as a vehicle for the action and special effects which is what
    The characters are a wee bit disappointing as far as the main characters are concerned. I felt that the main protagonist has little in terms of charisma. Killmonger, the main bad j
    Personally the character I liked best was actually Klawe which was pretty cool, insane but cool. The female warriors, which were both kick-ass and smart, and Shuri is also fairly j
    Special effects! This is of course where this movie shines. Overall I think the special effects were good. The design of the various gadgets, ships etc were very nice. The rampagin
    On the whole Black Panther is a quite enjoyable special effects and action movie. Nothing more, nothing less. Some social preaching nonsense but less than I feared, huh no story. j
    [5] => I was really surprised how much I liked this movie. There was so much going on; immigration concerns; the fight between saving the world and protecting your ou
    This movie isn't just explosions and one-liners in an exotic location. This is a great movie in general. It's always interesting. There are great relationships between all the j
    *****
    I also have to add that; you get to actually see Andy Serkis! When I saw his name I thought we was going to be some kind of CGI monster. I'm just happy to see him get actual scr
    [6] => The Black Panther is a fun romp, but it definitely has some issues. First, the plot.
    The action is fun, the emotions hit home, and the story isn't dirt-poor.
    While the CGI gets ripped on from time to time, I thought it was solid enough for a Marvel movie. The fight scenes were relatively well-choreographed, as well. The ritual combat s
    In fact, that's another appeal, at least to me. The secret technology base and the fun action scenes were a great callback to James Bond-type movies. I do think the movie would b
    The story, well, it's a mixed bag. Everything seemed crammed together with major emotional keys being rushed through or somewhat ignored. The concept is fine though. A Long lost r
  )
)
  
```

Figure 3 : After eliminating unwanted fields reviews of each movie

B. Microsoft Text Analytics

The Text Analytics is a cloud-based service that provides advanced natural language processing over raw text, and includes four main functions: sentiment analysis, key phrase extraction, language detection, and entity linking. The API is backed by resources in Microsoft Cognitive Services[4], a collection of machine learning and AI algorithms in the cloud, readily consumable in your development projects.

C. Google Cloud Natural Language API

Google Cloud Natural Language [5] reveals the structure and meaning of text both through powerful pre trained machine learning models in an easy to use REST API and through custom models that are easy to build with Auto ML Natural Language. You can use Cloud Natural Language to extract information about people, places, events and much more mentioned in text documents, news articles, or blog posts. You can use it to understand sentiment about your product on social media or parse intent from customer conversations happening in a call centre or a messaging app. You can analyse text uploaded in your request or integrate with your document storage on Google Cloud Storage.

D. Classification

We can see that the SentiScore(Sentiment Score) is calculated from the two features(Microsoft Text Analytics and Google Cloud Natural Language API). The weight of SentiScore is in the range of 0 to 1. Now, once we get the SentiScore[6], we determine the label using the following algorithm:

Algorithm 1. Proposed Algorithm for determining the label

```

Initialize SentiScore = 0
SentiScore = value from any of the feature
If SentiScore >= 0.8 && <=1.0, then:
    SentiLabel = 5
Elseif,SentiScore>=0.6&&<0.8,then:

```

```

    SentiLabel = 4

```

```

Else if SentiScore >=0.4 &&<0.6,then:

```

```

    SentiLabel = 3

```

```

Else if SentiScore >= 0.2 &&<0.4,then:

```

```

    SentiLabel = 2

```

```

Else

```

```

    SentiLabel = 1

```

III.RESULTS AND DISCUSSION

Both the features analyse the reviews data set and generate the SentiScore. Based on the algorithm SentiLabel is determined. Finally the class of the movie is obtained from the SentiLabel. We use this scale for defining the class label of the movie reviews which is divided as follows:

1. Sentiment Label 1 belong to class 1
2. Sentiment Label 2 belong to class 2
3. Sentiment Label 3 belong to class 3
4. Sentiment Label 4 belong to class 4
5. Sentiment Label 5 belong to class 5

For any movie finally the classification is done based on the reviews from scale of class 1(highly disliked) to class 5(highly liked).

IV.CONCLUSION

In this work, we extracted new features that have a strong impact on determining the polarity of the movie reviews and applied computation linguistic methods for the calculation of sentiment score for the reviews dataset. We then performed the classification for the movies by using an algorithm for SentiLabel generation.

In future, we would like to evaluate the effectiveness of the proposed sentiment classification features and techniques for other tasks, such as sentiment classification. We would like to apply in-depth concepts of Natural Language Processing for better

prediction of the polarity of the document. We would also like to extend this technique on other domains of opinion mining like television shows, political discussion forums etc.

V. REFERENCES

- [1] Sahu, T. P., & Ahuja, S. (2016). Sentiment analysis of movie reviews: A study on feature selection & classification algorithms. 2016 International Conference on Microelectronics, Computing and Communications (MicroCom). DOI:10.1109/microcom.2016.7522583
- [2] Magdum, S. S., & Megha, J. V. (2017). Mining online reviews and tweets for predicting sales performance and success of movies. 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). DOI:10.1109/iccons.2017.8250738
- [3] <https://developers.themoviedb.org/3/movies/get-movie-reviews>
- [4] <https://azure.microsoft.com/en-in/services/cognitive-services/text-analytics/>
- [5] <https://cloud.google.com/natural-language/>
- [6] Topal, K., & Ozsoyoglu, G. (2016). Movie review analysis: Emotion analysis of IMDb movie reviews. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). DOI:10.1109/asonam.2016.7752387

Cite this article as : S Jaswanth Sai Chowdary, Sk Waseem Farooq, Sk Abubakar Siddique, S Dinesh, "Sentiment Analysis of movie reviews using Microsoft Text Analytics and Google Cloud Natural Language API", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 1, pp. 147-150, January-February 2019. Available at doi : <https://doi.org/10.32628/CSEIT195130>
Journal URL : <http://ijsrcseit.com/CSEIT195130>