REPUTATION MANAGER

Mohmmad Aslam (IIT Gandhinagar)

I. ABSTRACT

With the rise of internet in the recent times, the need of managing the image of an organization in this domain has become a crucial practice. This web application is a tool to extract data from various popular social platforms and use sentimental analysis to classify them into positive and negative feedback which will be very useful for the company. This paper highlights the methods of extracting useful data from Twitter, YouTube comments and online news platforms and implementation of sentimental analysis on them using techniques of artificial intelligence and deep learning.

II. INTRODUCTION

The importance of reputation to businesses can never be stressed enough. Building a good image leads to new customers, increased sales and growth and likewise, negative ones have the potential to tear a business down. Reputation management is a taxing undertaking, which led to efforts to automate such process.

With the internet being a massive weapon of both construction and destruction, our Reputation Management is designed to monitor users' online reviews of products and services. It helps spread positive customer experiences through review platforms and help tackle the negative reviews and comments.

Fig1 - Visual Representation

Reputation Manager uses the relevant information to increase a company's sales by giving it insights on how consumers view the business. In the event, the software captures a negative review, the company can contact that person and assist him/her in getting a richer experience, which could alter what he/she thinks of the organization. In essence, the software helps not only customers but companies as well, by helping them improve.

Reputation Manager provides:

- A secure online environment: Corporate teams that access and exchange restricted data online should do so confidently without the threat of privacy breaches. This is why roles, privileges and access must be assigned in a secure setting.
- Single platform/window: We offer a single dashboard through which users can hand out information. It serves as the single destination for information gathered from Twitter, YouTube comments, and different news sources.
- Search and monitoring functionalities: We track emerging stories and conversations and their corresponding origins.
- Cost-effective and practical: Conventional methods of monitoring millions of conversations about a single brand is not only tiresome, but very expensive as well. Our tool makes sure that it accomplishes in a cost effective and practical manner.

III. RELATED WORK

- Benefits of online reputation management for organizations operating in various industries by Lukáš Vartiak from University of Žilina, Faculty of Operation and Economics of Transport and Communications, Department of Communications.
- Corporate reputation in management research: a review of the feedback, reviews and assessment of the concept by Annika Veh, Markus Göbel and Rick Vogel.
- Large-scale Sentiment Analysis for Reputation Management by Georgios Petasis, Nikos Tsirakis, Dimitris Spiliotopoulos, P.Tsantilas

IV. METHODOLOGY

A. Extraction of Tweets

Tweets are extracted using the Twitter API offered by Twitter's developer platform. Endpoints included in the "Tweets and Users" preview allow developers to request Tweet and user information from the Twitter API. The GET/users endpoint provides developers with information about a Twitter user. The information provided by the endpoint is in the form of hydrated user objects. The hydrated user object contains public Twitter account metadata such as name, description, location, and more, which is returned in the payload. A payload is the data that is returned from a request to the API.

The extracted information is then saved into a dataframe using Pandas for further processing and analysis.

B. Extraction of YouTube comments

YouTube is the world's largest video-sharing site with about 1.9 billion monthly active users. People use it to share info, teach, entertain, advertise and much more.

Thus, YouTube comments have some pivotal data that one can utilize to carry out research and analysis of reviews and feedback of consumers on various companies, services and products. YouTube comments are extracted using YouTube API, which enables one to search for videos matching specific search criteria.

The extracted comments are stored in a CSV file with various columns extracted like username, description, time stamp, etc.

C. Extraction of News from Online Platforms

News circulated by various journalists and published by different newspapers provides vital first-hand information, which is extracted with the help of News API, which is a simple HTTP REST API for searching and retrieving live articles from all over the web.

One can search for articles with any combination of the following criteria:

- Keyword or phrase
- Date published
- Source name
- Source domain name
- Language

One can sort the results in the following orders:

- Date published
- Relevancy to search keyword
- Popularity of source

D. Sentimental Analysis using SIA VADER

VADER (**Valence Aware Dictionary and Sentiment Reasoner**) is a lexicon and rule-based sentiment analysis tool that is ***specifically attuned to sentiments expressed in social media.***

VADER uses a combination of a sentiment lexicon is a list of lexical features (e.g. words) which are generally labelled according to their semantic orientation as either positive or negative.

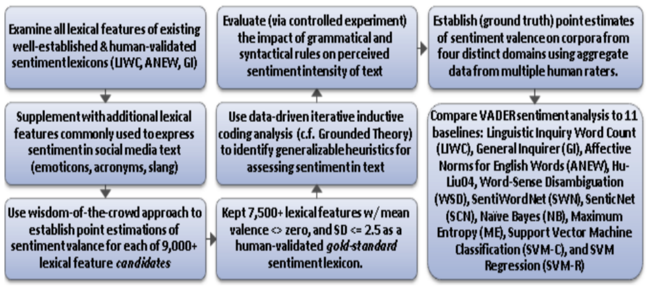

Fig2 - Flow chart of SIA VADER

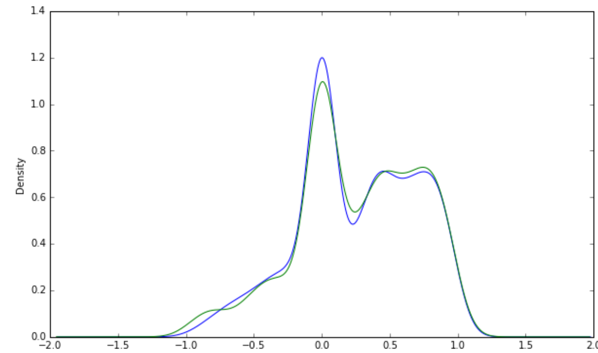

Fig3- Sentimental Analysis using SIA VADER

VADER has been found to be quite successful since VADER not only tells about the Positivity and Negativity of a sentence but also tells us about **how positive or negative a sentiment is** with a **Compound Score.** The positive, neutral and negative scores represent the proportion of the text that falls inside these categories whilst the compound score represents the sum of all the lexicon ratings where +1 represents most positive and -1 represents most negative.

If this score is between -0.2 and 0.2 then the sentiment is neutral. If it is lower than -0.2 than the sentiment is negative and if the compound score is bigger than 0.2 it's positive.

The VADER sentiment analysis library allows us to easily implement sentiment analysis that operate at almost real-time speed.

VADER has a lot of advantages over traditional methods of Sentiment Analysis, including:

- It works exceedingly well on social media type text, yet readily generalizes to multiple domains.
- It doesn't require any training data but is constructed from a generalizable, valence-based, human-curated gold standard sentiment lexicon.
- It is fast enough to be used online with streaming data.
- It does not severely suffer from a speed-performance trade-off.

To refine the results further, the scores have been normalized considering factors like the impact of number of followers on the spread of the comment and review on the internet.

E. Sentiment Analysis using LSTM

LSTM, which stands for ***Long Short Term Memory*** is a powerful architecture of Recurrent Neural Networks. Using such architecture, the model was created and trained.

The data is cleaned by removing the following:
URLs, @mentions, and punctuations. Stop words may or may not be removed. Here, they were kept.

The model is trained on pre-processed labelled twitter data (tweets) using a pre-trained embedding. Pre-trained embeddings are gloVe. It has embeddings for 4,00,000 words which were directly imported and embedding layer was set to non-trainable.

Below is the detailed explanation of the model architecture:

```
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         (None, 45)                0
_________________________________________________________________
embedding_1 (Embedding)      (None, 45, 50)            20000050
_________________________________________________________________
lstm_1 (LSTM)                (None, 45, 128)           91648
_________________________________________________________________
dropout_1 (Dropout)          (None, 45, 128)           0
_________________________________________________________________
lstm_2 (LSTM)                (None, 128)               131584
_________________________________________________________________
dropout_2 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 129
_________________________________________________________________
activation_1 (Activation)    (None, 1)                 0
=================================================================
Total params: 20,223,411
Trainable params: 223,361
Non-trainable params: 20,000,050
```

Fig4 - LSTM Model architecture

Layers:

1. **Input Layer** - Each tweet is padded and made equal to the maximum length of all tweets, which is 45.
2. **Embedding Layer** - Each input is converted into embeddings, which are pre-trained glove embeddings with 50 features per word and hence the output shape from embedding layer is (None, 45, 50).
3. **Two LSTM layers each with 128 units** are added and a **dropout layer** has been added to each of them to prevent overfitting. The dropout probability has been set to 0.5.
4. Finally, a **hidden layer (of 1 unit) with sigmoid activation** has been added for the classification problem.

Model was trained on optimizer "adam" and loss=" binary_crossentropy".

Fitting is done over 50,000 training, for example, in batch sizes of 32 for 50 epochs and an accuracy of 76% was obtained on test set.

The saved model is then used for analysis of the extracted data.

Finally, we used a combined score from both the models. Also, we laid some rules. Few of them are:

1) If user searches for a word that is a bi-gram (for example - Apple Macbook), then information containing the bi-gram (both the words together will be displayed first) and then those information where one of the words, or both the words appear separately will be displayed.
2) Score is calculated by normalizing the follower count (in case of twitter) or normalizing the like count (in case of youtube) and then multiplying it with the sentiment value returned by the value, (called the compound score).
   For example, in case of twitter:
   score = normalized-follower * compound score;
   where
   normalized-follower = follower - min(follower) ÷(max(follower) - min(follower)_

• V. Experimental Results

- With the assimilation of the extracted data from diverse social platforms and sentimental analysis of it using various techniques, we integrate all of it to make a wholesome product in form of a web application for the convenient usage by the consumer.
- Our web application behaves a tool for companies and organizations, both small and big for reviewing the feedback running across the internet and classify them as positive and negative with a compound score, gauging their impact on the image and reputation of the company. This will help the company to counter and act upon the feedback or comment.

• VI. Conclusion

Internet is a place where consumers exchange their experience. This process contributes to shaping online reputation of organizations. Positive online reputation is desired by most organizations. But only few organizations know that it can be changed from positive to negative per day. To avoid inconveniences, organizations need to manage their online reputation. The purpose of this study and development of a Reputation Manager is to demonstrate the benefits of online reputation management based on analysis of related case studies.

Some of the other features of our Reputation Manager is that it:

- Helps in quickly resolving customer complaints.
- Helps the company show concern over customer service.
- Helps in earning customer trust.
- Creates positive word-of-mouth about your organization.
- Increases transparency.
- Identifies business opportunities.
- Aids in growing presence in digital channels.
- Promotes online endorsements.
- Aids the company in addressing negative reviews and comments.

VII. Acknowledgment

This study and development of the project owes a major part of itself to the guidance of Dr. Vipul Mishra, our mentor from Bennett University, who supported and helped us at every step of the development.

References

[1] YouTube Data API Documentation: https://developers.google.com/youtube/v3.

[2] Twitter API Documentation: https://developer.twitter.com/en/docs/labs/tweets-and-users/quick-starts/get-tweets

[3] News API Documentation: https://newsapi.org/docs

[4] Usage of NLTK-SIA VADER Documentation:

[5] Online Reputation Management Author: Robbin ten Dolle University of Twente P.O. Box 217, 7500AE Enschede The Netherlands