Movies Emotional Analysis Using Textual Contents

Amir Kazem Kayhani[1], Farid Meziane[1(✉)], and Raja Chiky[2]

[1] School of Science, Engineering and Environment,
University of Salford, Salford M5 4WT, UK
A.K.Kayhani@edu.salford.ac.uk, F.Meziane@salford.ac.uk
[2] ISEP-LISITE, Paris, France
raja.chiky@isep.fr

Abstract. In this paper, we use movies and series subtitles and applied text mining and Natural Language Processing methods to evaluate emotions in videos. Three different word lexicons were used and one of the outcomes of this research is the generation of a secondary dataset with more than 3658 records which can be used for other data analysis and data mining research. We used our secondary dataset to find and display correlations between different emotions on the videos and the correlation between emotions on the movies and users' scores on IMDb using the Pearson correlation method and found some statistically significant correlations.

Keywords: Emotional analysis of movies · Text mining

1 Introduction

Contextual analysis of videos (in this paper, this will encompass movies and TV series) is very important for media companies as it helps them define standard measures and a better understanding of the huge number of video contents without watching them. It also allows them to predict viewers interest, classify and cluster millions of videos based on their contents for different age groups and smart recommendation systems.

Research in emotional analysis and clustering of movies is traditionally based on users' interests and profiles, general characteristics of the movies such as country of production, genre, production year, language and duration or based on linking the emotions and sentiments on users reviews on social media. In this paper we are adopting a completely different approach for emotional analysis of movies by using textual analysis based on their subtitles. The dataset used in this study is freely available and downloaded from Opensubtitles website [1]. To our knowledge, this is the first study on emotional analysis on movies that is based on textual contents. The methodology used in this research is composed of two phases.

---

[1] Opensubtitles.org

In the first phase, after data preparation and cleansing we performed sentiment analysis on more than 3650 subtitle files with three different lexicons and calculated the percentage of words with different emotions (trust, joy, fear, positive, etc.) on every SubRip Subtitle file. A SubRip file is the file associated with the subtitle (with the .srt extension). The structure of a subtitle file contains "the section of subtitles number", "The time the subtitle is displayed begins", "The time the subtitle is displayed ends", and the "Subtitle". This phase also includes movies' scoring based on their adult contents (violence and sexual content) that can be used as a source for age and parental ratings and guidance.

In the second phase, scores normalisation is performed. The emotional scores are normalised to values between 0 and 100 and these are assigned to every video. Although the emotional rating of the videos is useful for data analysis on its own, the outcome of this scoring is also used as a new dataset with more than 3650 items and 34 features. This new dataset will be used for other data mining applications such as recommendation systems and predicting viewers interest and score to movies. This last aspect is not covered in this paper.

The remaining of this paper is organised as follows. In Sect. 2 we review some related works. The data cleansing and preparation phase is described in Sect. 3. The emotions analysis is performed in Sect. 4 with the lexicons used in this study and their developments described in Sect. 5. The correlations between emotions and the IMDb scores analysis is given in Sect. 6. Finally, Sect. 7 discusses the results of this research, draw some conclusions and provides some insights into future works.

2   Literature Review

Plutchick [5] developed his emotions model based on eight human emotions including acceptance, surprise, sadness, anger, joy, fear, anticipation and disgust. He also defined some emotions as compounds of two other emotions (for example love is defined as a compound of joy and trust) and he also defined levels of intensity for each of the eight emotions. His model became the most popular for displaying human emotions. Plutchik's wheel of emotions is utilized in many researches in psychology and interdisciplinary fields such as NLP.

Alsheikh, Shaalan and Meziane [1] used the polarity, intensity and combinational concepts in Plutchik's wheel of emotions with text mining and sentiment analysis methods to evaluate trust as an emotion between sellers and buyers in the Customer to Customer marketplace. They have used text mining methods to find correlations between emotions on the hosts' description of their facilities (accommodation in this case as they have used Airbnb as a case study) and negative sentiments by guests through their reviews. They have also used the combinational concepts of emotions based on Plutchick and Ekman emotional model for calculating trust.

Cambria, Livingstone, and Hussain [2] proposed a new model for human emotions which they named "The Hourglass of Emotions". Their model is a reinterpretation of the Plutchik's wheel of emotions and is specifically designed

for applications such as sentiment analysis, social data mining and NLP. They used polarity and intensity levels of emotions in Plutchik's wheel of emotions and defined 4 main emotions (instead of the 8 emotions in the Plutchik's model).

Topal and Ozsoyoglu [7] proposed a model for emotional analysis and classification of movies based on the viewers reviews on the IMDb website arguing that there is a close link between users' reviews on IMDb and emotions on the movies. In addition, they have also found correlations between high level of emotionality in movies and high scores (7 or more) to movies by IMDb users. In their research, Topal and Ozsoyoglu [7] used the Hour Glass of Emotions [2] to find emotional scores for each movie based on the reviews and used the K-means algorithm for clustering movies based on their emotional scores according to users' reviews on IMDb. A recommendation system based on the movies emotional score by their method and users' emotional preferences was then proposed.

Li, Liao and Qin [3] stated that most of the works on clustering and recommendations systems for movies are based on users' profiles and interests in movies and/or based on users' social media interactions. However, this approach is not so accurate as it is not based on movies contents, features and characteristics but are focused on users. They proposed clustering movies based on their characteristics such as year and country of production, director, movie type, language, publishing company, casts and duration. They used Jaccard distance for calculating the similarities between movies. Although they have used movies characteristics for clustering, they have combined the results of clustering with users ratings for improving their movie recommendation system [3].

3   Data Understanding and Preparation

In the original dataset, there was a large number of subtitle files in different languages. About 3650 of them have been selected as our sample where 119 are for movies (mostly from top 200 IMDb movies) and the rest are for video series. Two layers of data cleansing were performed on the subtitle files to prepare them for text mining and analysis. The first layer is removing special characters, numbers and extra spaces. In the second layer stop words were removed. In addition, all the letters in all words are converted to lower case and then converted to their root forms (stemming). The output of the data cleansing and preparation phase is a single text file where the tab separator character is used to separate the contents of the different files. Since text mining was performed on a large number of text files and the performance is also important for us, this method of buffering all the files can also help to increase our code performance and execution speed.

Furthermore, we performed an analysis on the number of words and their grammatical roles (part of speech tagging) before and after data cleansing. Although the number of words with all grammatical roles has decreased; as expected, the biggest change occurred among numbers and verbs which have fallen from more than 5.8 million and 3.6 million to 36 and 26,324 respectively.

4   Emotional Analysis

The output of the data cleansing phase is used as the input for emotional analysis. Our method for emotional analysis is similar to the most common methods used for sentiment analysis and is based on Term Frequency (TF) such the work of Rafferty for emotional analysis of the Harry Potter's books [6]. However, our work has additional complexities since it has been defined for emotional analysis on videos (instead of books) and some additional features such as the calculation of 'in-between' emotions, enhancements in data cleansing and normalising the results for a better visualisation and preparation for machine learning tasks. We used three lexicons namely, NRC, AFINN and Ero (these lexicons will be described in Sect. 5) with thousands of words associated to different emotions and we have counted the number of words with each emotion on movies to find the ratio of words with each emotion on every movie (or episode). $E(m, e)$ is the percentage of emotion e in movie m and is calculated using Eq. 1 where $Num_{m,e}$ represents the number of words with emotion e (ex. joy) in movie m (ex. Titanic) and W_m represent the number of all words in movie m.

$$E(m, e) = \frac{num_{m,e}}{W_m} * 100 \tag{1}$$

We have also calculated the total percentage of emotionality in movies by counting the number of all the words in each movie which are also in our NRC Emotion Lexicon. We called this value Emotional Expression $EE(m, e)$ and is calculated using Eq. 2, where Num_m is the number of words in movie M which are also in the NRC lexicon and W_m is the number of all words in movie m.

$$EE(m, e) = \frac{Num_m}{W_m} * 100 \tag{2}$$

In addition, the difference between positivity and negativity for each movie m or episode is calculated using Eq. 3.

$$E(Pos - Neg, m) = \frac{Num_{pos,m} - Num_{neg,m}}{W_m} * 100 \tag{3}$$

The values for 8 in-between emotions which are not in our lexicons were calculated, but they are combinations of other emotions based on Plutchik's wheel of emotions. Calculating the in-between emotions was performed using two methods. The first method is based on the average value of the related emotions. For instance, from the psychological point of view and based on Plutchik's wheel of emotions, love is the combination of joy and trust and its value in movie m is the average of these two emotions as given in Eq. 4, where $P_{joy,m}$ and $P_{trust,m}$ are the percentages of joy and trust in movie m respectively.

$$L(i) = \frac{P_{joy,i} + P_{trust,i}}{2} \tag{4}$$

The second method is based on expanding the NRC Emotion Lexicon according to Plutchik's wheel of emotions (this will be described in Sect. 5). Finally, we
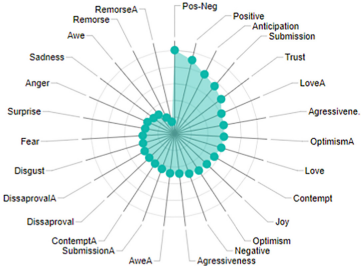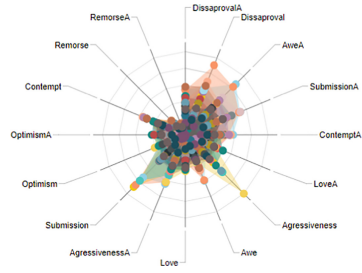
Fig. 1. Emotions in Titanic.

Fig. 2. Emotions in Fargo (all episodes).

have normalised emotional percentage results in order to have a standard score from 0 to 100 for all emotions in our dataset. The outcome of this part of the research is the production of two datasets (one normalised and the other not) in the form of two CSV files with 34 columns. After converting the unstructured text data of the initial sample dataset to a structured dataset, we analysed the results for finding the statistical characteristics of the new dataset. Radar charts in Figs. 1 and 2 show the results of emotional analysis for Titanic and Fargo (all episodes) respectively.

5   Lexicons Developments

NRC Lexicon is used as the main Lexicon in this research. The original NRC lexicon [4] consisted of 14,183 rows (words) and 10 columns (8 emotions and 2 sentiments) which is available in over 40 languages and show the association between each word and each emotion in a 0 and 1 matrix where 1 represents the existence of an association and 0 the non-existence. In this research we used the English version of NRC lexicon [4]. The NRC Lexicon consists of the following columns: Trust, Joy, Anticipation, Anger, Disgust, Sadness, Surprise, Fear, Positive and Negative. We expanded the NRC lexicon based on the Plutchik's wheel of emotions. We considered that some emotions like Love, Submission, Optimism or Awe are in fact combination of two other emotions. Figure 3 shows the frequency of association of words with each emotion in the expanded NRC lexicon.

AFINN is the second used lexicon [8] and includes 1477 ratings from −5 (extremely negative) to +5 (extremely positive). We have only used the combination of very negative words in AFINN dataset with −4 and −5 scores for detecting offensive words which may also be associated with sex and/or violence [8]. For simplicity and better understanding, in this research we refer to this secondary lexicon as AFINNVN (AFINN Very Negative). Using this method was our first attempt for scoring movies and series based on adult contents (violence, sex, drugs) which can help us in defining a parental score and finding some interesting correlations and increasing our accuracy for clustering the movies and series.
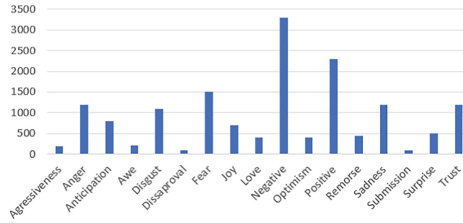
Fig. 3. Frequency of association of words with each emotion in NRC lexicon and expanded NRC lexicon

Although using very negative words in AFINN helped us to detect some offensive, violence and sex related words, it was not enough for detecting many of the sex related words in the movies and after searching, it seems that there is no available and open source lexicon for detecting such words and contents. Hence, there was a need to develop such a lexicon that we named Ero. Ero is constructed from 1422 words which are mostly copied from the "Dirty Sex Dictionary" by filtering bold words on the mentioned HTML page. However, some preparations and modification were needed as many of the words on the list have commonly non-sexual meanings. This lexicon significantly enhanced our work for scoring adult contents.
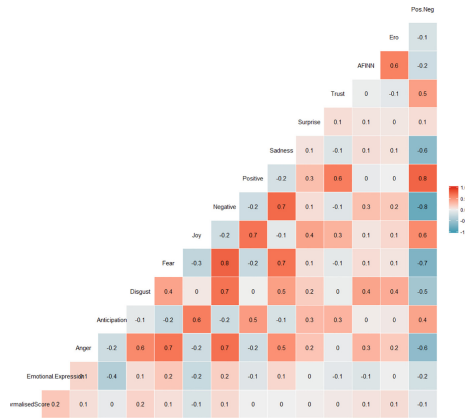
After finding the amount and percentage of emotions and normalising the results, the secondary dataset has been prepared for further analysis. One of the interesting analysis was finding correlations between emotions in movies and the scores given by the users as provided by the IMDb website. While the range of scores in our sample dataset was between 4.8 and 9.5, we normalised the scores to the range 0 to 100 as the emotions were scored in the range 0–100. Based on the size of our dataset and since our data is normalised, we used Pearson method in R for finding the correlations between the two variables X and Y as given in Eq. 5.

$$r_{X,Y} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{5}$$

6   Correlation Between Emotions and IMDb Scores

The correlogram shown in Fig. 4 visualises the correlations between emotions (excluding in-between emotions), sentiments (positive and negative) and users' scores to movies on the IMDb database using the Pearson correlation method. As expected, all the emotions have considerable correlation with either negative or positive sentiments and there is a negative correlation between negative and positive sentiments. However, it is interesting to note that correlations between negative and positive sentiments are not very high (+0.2). On the other hand, we can see that there is a high correlation (+0.6) between usage of offensive words (AFINNVN) and sexual words (Ero). The main reason could be common

usage of sexual words as offensive words. One of the interesting correlations is the very high correlation (+0.7) between anger and fear (probably since anger can cause fear). The only emotion with positive correlation with both positive (0.3) and negative (0.1) sentiments is surprise as surprising happen in both positive and negative ways. In addition, based on the Pearson correlation results (Fig. 4), we can see that positive sentiment and joy, trust and anticipation emotions have negative correlation with Emotional expression which is the ratio of emotional words to all words in a movie or episode. It was completely against our expectations since words with positive sentiment and joy, trust and anticipation emotions will be counted as emotional expression. Based on these results we can conclude that an increase in each of the positive sentiments of joy, trust or anticipation emotions usually will lead to decreasing other emotions.

	Pos.Neg	Ero	AFINN	Trust	Surprise	Sadness	Positive	Negative	Joy	Fear	Disgust	Anticipation	Anger
Ero	-0.1												
AFINN	-0.2	0.6											
Trust	0.5	-0.1	0										
Surprise	0.1	0	0.1	0.1									
Sadness	-0.6	0.1	0.1	-0.1	0.1								
Positive	0.8	0	0	0.6	0.3	-0.2							
Negative	-0.8	0.2	0.3	-0.1	0.1	0.7	-0.2						
Joy	0.8	0.1	0.1	0.3	0.4	-0.1	0.7	-0.2					
Fear	-0.7	0.1	0.1	-0.1	0.1	0.7	-0.2	0.8	-0.3				
Disgust	-0.6	0.4	0.4	0	0.2	0.5	0	0.7	0	0.4			
Anticipation	0.4	0	0	0.3	0.3	-0.1	0.5	-0.2	0.6	-0.2	-0.1		
Anger	-0.6	0.2	0.3	0	0.2	0.5	-0.2	0.7	-0.2	0.7	0.6	-0.2	
Emotional Expression	-0.2	0	-0.1	-0.1	0	0.1	-0.1	0.2	-0.2	0.2	0.1	-0.4	1
normalisedScore	-0.1	0.1	0.1	0	0	0	0.1	-0.1	0.1	0.2	0	0.1	0.2

Fig. 4. Correlations Heatmap. Blue colour indicates negative correlations, red indicate positive correlations and white indicate not having correlations (Color figure online)

Computing correlations between users' scores in IMDb and sentiments and emotions on movies is of great interest. Indeed, it can help in predicting users' scores and can help media companies to invest in movies which can be more interesting and attractive to users and also can help them to increase the quality of their movies and series based on users emotional interests. Generally, in statistics using two alpha p-values for accepting Pearson correlation results as statistically significant; Alpha <0.01 and alpha <0.05 are two common alpha p-values used in statistical studies. According to the results shown in Fig. 4 we can clearly see that there are lots of considerable and statistically significant correlations between emotions on the movies and users' scores on IMDb. Although there are cases where there is no statistically significant correlation between some emotions and users score on IMDb, most of the emotions have statistically significant correlation with users scores on IMDb and most of the correlations are statistically significant at p-value alpha level <0.01. Most effective parameter in users' scores to movies is 'Emotional Expression' on movies with +0.22 Pearson

correlation which is also statistically significant at the level of alpha <0.01. It means that in general movies with rich words and higher level of emotionality are more likely to be interesting to viewers and get higher score in IMDb. However, this conclusion is not true for all emotions as some emotions have negative Pearson correlation with users' scores on IMDb and it is more likely to negatively affect users' scores to movies. Consequently, for more accurate analysis we should consider each emotion separately.

Some emotions such as joy, love and anticipation have negative correlation with users' scores and it is more likely to affect users score to movies in a negative way but on the other hand most of the emotions such as anger, fear and contempt have positive effect on users score and satisfaction. Furthermore, we can see that words in AFINNVN and Ero have positive correlation with users' scores on IMDb and their correlation is statistically significant at the alpha level <0.01. Although many of the correlations are statistically significant according to their p-value, Pearson correlation coefficient with values between −0.3 and 0.3 usually indicate a weak correlation.

7   Discussions, Conclusions and Future Works

We have performed the emotional analysis on a relatively large data set with more than 3650 subtitle files and the results of this analysis created a secondary dataset which can be used for further research and analysis. We are particularly aiming at using the results of this research to improve on recommendation systems.

References

1. Alsheikh, S.S., Shaalan, K., Meziane, F.: Exploring the effects of consumers' trust: a predictive model for satisfying buyers' expectations based on sellers' behavior in the marketplace. IEEE Access **7**, 73357–73372 (2019)
2. Cambria, E., Livingstone, A., Hussain, A.: The hourglass of emotions. In: Esposito, A., Esposito, A.M., Vinciarelli, A., Hoffmann, R., Müller, V.C. (eds.) Cognitive Behavioural Systems. LNCS, vol. 7403, pp. 144–157. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34584-5_11
3. Li, B., Liao, Y., Qin, Z.: Precomputed clustering for movie recommendation system in real time. J. Appl. Math. 1–9 (2014)
4. Mohammad, S., Turney, P.: Crowdsourcing a word-emotion association lexicon. Comput. Intell. **29**(3), 436–465 (2013)
5. Plutchik, R., Kellerman, H.: Emotion theory, research and experience. Theories of emotion. Psychol. Med. **1**(1), 207–207 (1981)
6. Rafferty, G.: Harry potter NLP (2019). https://github.com/raffg/harry_potter_nlp/blob/master/sentiment_analysis.ipynb. Accessed 10 Oct 2019
7. Topal, K., Gultekin, O.: Emotional classification and visualization of movies based on their IMDb reviews. Inf. Disc. Deliv. **45**(3), 149–158 (2017)
8. Nielsen, F.Å.: A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In: Rowe, M., Stankovic, M., Dadzie, A., Hardey, M. (eds.) Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big Things Come in Small Packages, vol. 718, pp. 93–98, May 2011