

Implementation of Opinion Mining Based Movie Rating System

Intkhab Alam Shaikh¹, Abhishek L. Deshmukh², Atul S. Karande³,
Shagufta Rajguru⁴

^{1,2,3,4} Fr.C.R.I.T. Vashi, Navi Mumbai, Department of Computer Engineering

ABSTRACT

This paper presents an experimental work on a document-level sentiment analysis of movie reviews. A linguistic feature oriented scheme that analyses the textual reviews of a movie review and assign it a sentiment label on each sentimental keyword has been devised. The scores from multiple reviews are then collected and a net sentiment profile of the movie is generated. A keyword-label based scheme with two different semantic feature selections containing adjectives, adverbs and n-gram feature mining is used. Also this keyword-label scheme has been used to compute the document-level sentiment for each movie reviewed. This paper explains the steps involved in sentiment analysis and its application on automated movie rating system to rate movies based on reviews quoted by people.

Keywords: Sentiment analysis; opinion mining; document-level sentiment analysis; keyword-label scheme; movie rating; user reviews

I. INTRODUCTION

Automated text analysis is used in fields where products are reviewed by customers and critics. It helps in decision making and prediction procedures. Same is the case with movie rating system, user and critics add their reviews about particular movie, which helps other users in deciding whether to watch the movie or not. In today's age of rapid growth of the World Wide Web there is huge data available on social sites, posts etc. which can help to predict public opinion. Due to the large number of reviews, it is difficult for potential customer to get resourceful review details, it is a challenging task. Also instead of referring the newspapers or websites that only contains a five star based rating which are usually not accurate, here, OBMRS is useful for users.

OBMRS does not just rely on numbers or predefined scale of 1-5 or 1-10 star/rating instead the system analyses user comment, checks for sentimental keywords and divides it into positive, neutral or negative.

System contains a sentiment library designed for English sentiment analysis. Opinion mining typically occurs in two or three stages: (1) the input text is split into pieces, such as sentences, and each piece tested to see if it contains any sentiment. (2) The subjective sentences are analyzed to detect their sentiment polarity. (3) The object about which the opinion is expressed may be extracted. The system now collects all comments for a given movie and then computes an average rating for it. This score is generated for every movie in the system. This provides an automated movie rating system based on keyword-label scheme of sentiment analysis. Important step in analysis process is tokenization. Tokenization is the process of categorizing a collection of text into meaningful words, phrases or symbols. The tokens can be used further for parsing (syntactic analysis) or text mining.

Some of the relevant past works on sentiment classification can be found in [1], [2], [3], [4], [5], [6], [7] and [8].

II. DESIGN

The opinion mining based movie rating system is basically an online automated movie rating system where information related to movies is stored and displayed. In addition to this, users who are registered with the system can add their reviews of any particular movie. The system then calculates the rating by analyzing these user reviews using the concepts of Sentiment Analysis. The system mainly contains 6 modules as shown in the following block diagram of system.

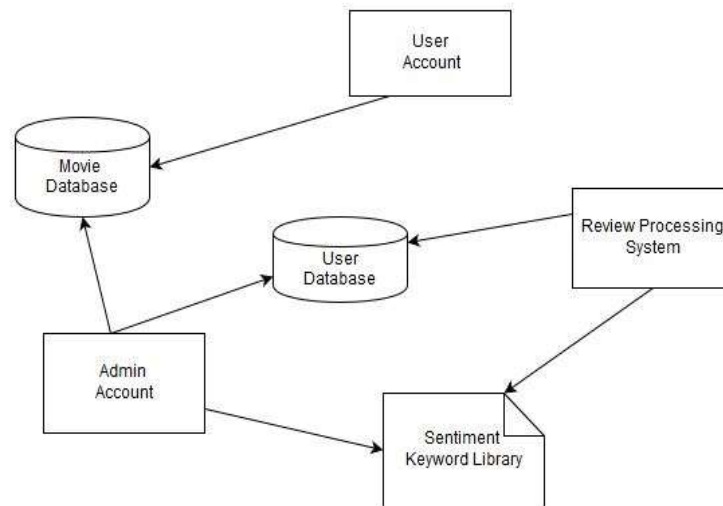


Fig. 1: Block Diagram

Following section describes functionality of all the modules:

A. User Accounts

1. Admin Account

Administrator account manages and maintains the databases of the system and authenticates and approves the user request in the registration process. Admin handles the Review Processing Module which calculates and displays the rating with each movie present in the system. Database activities include adding new movies into database, information of new users are stored into the database, deleting the wrong entries from those databases, etc.

2. Registered User Account

The user can search a movie by name. The user need not be registered to just check the information, view the movie page, see rating to any particular movie and see user's reviews for that movie. However, to comment on any movie or to add review user has to go through registration process. For registration the user has to submit email-id and password. The admin will authenticate the user by verifying information provided. Once the user is registered with the system, can submit reviews to any movie.

B. Databases

1. Movie Database

This is the central module of the system. It contains name, cast, year, plot, rating, reviews and other movie information. The rating which is displayed with each movie comes from the review processing module. As mentioned earlier each and every entry in the database is handled by the admin only.

2. User Database

This database module contains all the user information. The users are the same discussed in registered user account. The information is in the form of user name, email-id, reviews/comments, etc.

C. Other Modules

1. Review Processing Module

This module of the system deals with the review processing and implements sentiment analysis concept. The reviews or comments added by the user are first sent to admin and admin then forward them to this module, the module then processes the reviews, calculates the sentiment score and returns this score to the admin. The admin then displays this rating on movie page.

2. Sentiment Keyword Library

This module is a library of sentimental keywords from where the Review Processing module compares the keywords found in reviews. The library contains possible sets of positive and negative keywords with corresponding weight age scores. Each keyword in the database is assigned a score between 1 and 10 based on its polarity.

The flow chart of the system is described in the following section:

1. Registration: While using the system the very first stage is to register. The user submits email-id and password, which is forwarded to admin for verification and authentication. If the user information is correct the admin will update the user database by adding the user information into database.
2. Log-in: Once the user is registered, he/she can login into system using username and password. The username initially is the email-id of the user provided at the time of registration process.
3. Search/ View: The user can search or view the movie or person using search space given at the top of the page. Here user can view the description of movie i.e. plot of the movie and other movie information.
4. Review/Comment: User can add reviews for any movie and see the reviews posted by others.
5. Logout: At the end user can simply logout just by clicking logout button provided on the page.

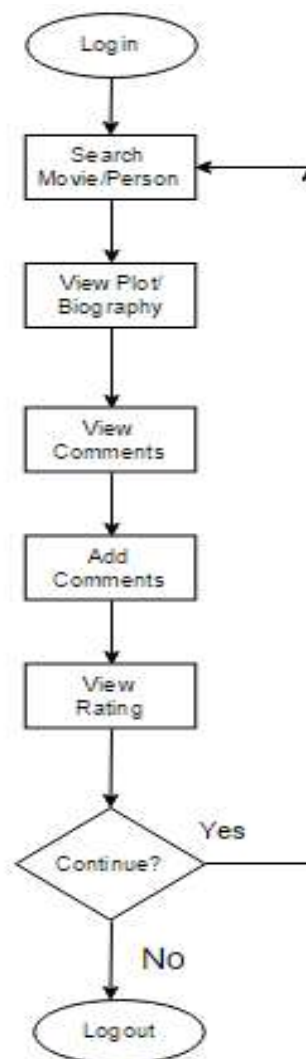


Fig.2: Flow Chart

III. IMPLEMENTATION DETAILS

The computational method in this paper are based on the document-level sentiment analysis[7]. The document-level sentiment analysis approach involves obtaining sentiment score for each selected opinion containing term of the text by a lookup in its library. In this lexical resource or keyword library each term t occurring in library is associated to two numerical scores $positive(t)$ and $negative(t)$, describing the positive and negative polarities of the term, respectively. These two scores are determined based on their use to describe the sentiment measure.

Table 1. Sentiment Keyword Library Example

Sr No.	Positive Keyword	Score	Negative Keyword	Score
1	Awesome	10	Boring	2
2	Funny	7	Irritating	1
3	Intelligent	8	Poor	3
4	Better	6	Unbearable	4

To get the the sentiment score we need to first extract relevant opinionated terms and then lookup for their scores in the library. These terms are what we are calling the sentiment keywords in this paper. This keyword library was created based on the linguistic features to be used, determining how much weight is to be given to each keyword and the collection method for consolidating sentiment scores.

This paper explored different linguistic features and scoring schemes for calculating the sentiment score. The research in the field of Computational Linguists shows that adjectives are good markers of opinions. For instance, if a review sentence says “The movie was brilliant”, then use of adjective ‘brilliant’ explains that the movie was appreciated by the reviewer and possibly he had a delightful experience watching it. Sometimes, Adverbs further modify the opinion conveyed in review sentences. For example, the sentence “The movie was extremely good” articulates an extra positive opinion about the movie than the sentence “the movie was good”. A related previous work has also concluded that ‘Adverb+Adjective’ combine produces better results than using adjectives alone[8]. Hence we preferred the ‘adverb+adjective combination over extracting ‘adjective’ alone. The adverbs are usually used as complements or modifiers. Some more instances of adverb usage are: he ran slowly, he behaved badly, very risky journey, very politely, hardly corrupt, rarely good etc. In all these examples adverbs modify the adjectives. Though adverbs are of various kinds, but for sentiment classification only adjectives of degree seem useful.

IV. DATASET AND PERFROMANCE MEASURE

The experimental evaluation was conducted on an average sized datase. To evaluate the performance of document-level sentiment analysis approach implementation, standard information retrieval performance assessment have been used.

A. Collecting Datasets

10 reviews each for 20 movies from the popular movie review database website www.metacritic.com [9] were taken. All these reviews were manually evaluated and then results were compared with performance of our algorithmic formulations.

V. RESULTS

As mentioned earlier different linguistic feature selection, weighing and aggregation schemes have been used in this paper for document-level sentiment classification. The results obtained were than compared with Sentigem [10] sentiment analysis tool results. Out of 200 movie reviews collected, 112 are labeled positive and 88 are labeled as negative reviews.

Table I presents the values of performance measures obtained for this paper’s implementations and Sentigem API.

Table 2. Performance measure value comparison

Method	Performance measure Value	
	Count	Percentage
Keyword-label Scheme	168	84%
Sentigem API	164	82%

The table II presents a comparison of the sentiment label assignments by this algorithmic formulations and the Sentigem API with manually labeled data.

Table 3. Total percentage of ‘positive’ and ‘negative’ labels

Method	Movie Review Dataset	
		Percentage (%)
Keyword-label Scheme	Positive	56
	Negative	44
Sentigem API	Positive	54
	Negative	46

Thus, document-level sentiment classification results obtained by keyword-label scheme algorithmic formulation are not only realistically accurate as compared to actual manual sentiment labels, but are also comparable to the results obtained by the Sentigem API.

CONCLUSION

This experimental work explores the use of ‘Adverb+ Adjective’ combination for document-level sentiment classification of a review. The document-level schemes implemented in this paper include use of ‘Adverb + Adjective’ combination only. This is done to analyze the opinionated value of different linguistic features of a review as in English language text document Adjectives are used to express the sentiment of opinion. As a result, many of the sentiment calculation were highly affected by the assumption that a review refer only the opinionated information i.e. Adjectives or sentimental keyword in a review together to generate a document level sentiment summary. Thus, this project provides an online automated movie rating system using the concept of keyword-label scheme based document-level based technique of sentiment analysis using.

FUTURE SCOPE

Some of the enhancements that can be made in the system in near future are listed below:

- [1] Right now the reviews that are posted in the system are only considered, but in future user’s reviews from other social networking sites such as facebook and twitter can also be collected and used to calculate the sentiment score.
- [2] This project includes sentimental keyword library for English language only, in future same library for other language can be created.
- [3] Here, in this project sentiment analysis concept is used to get reviews on movies, same concept can be extended for books, cars, songs, political leaders during elections to get the tendency of voters, in e-commerce websites, etc. as it is based on the document-level based technique of sentiment analysis.
- [4] The project lacks robustness since aspect-level sentiment analysis which is more domain dependent is not taken into consideration.

REFERENCES

- [1]. K. Dave, S. Lawrence & D. Pennock, “Mining the Peanut Gallery-Opinion Extraction and Semantic Classification of Product Reviews”, Proceedings of the 12th International World Wide Web Conference, pp. 519-528, 2003.
- [2]. P. Turney, “Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews”, Proceedings of ACL-02, 40th Annual Meeting of the Association for Computational Linguistics, pp. 417-424, Philadelphia, US, 2002
- [3]. A. Esuli & F. Sebastiani, “Determining the Semantic Orientation of terms through gloss analysis”, Proceedings of CIKM-05, 14th ACM International Conference on Information and Knowledge Management, pp. 617-624, Bremen, DE, 2005
- [4]. B. Pang, L. Lee & S. Vaithyanathan, “Thumbs up? Sentiment classification using machine learning techniques”, Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 79-86, Philadelphia, US, 2002.
- [5]. M. K. Jiawei Han, Data Mining: Concepts and Techniques. 500 Sansome Street, Suite 400, San Francisco, CA 94111: Diane Cerra, 2006.
- [7]. V. K. Singh, M. Mukherjee and G. K. Mehta, “Sentiment and Mood Analysis of Weblogs using POS Tagging based Approach”, In S. Aluru et al. (Eds.): IC3 2011, CCIS 168, pp. 313-324, Springer-Verlag, Berlin Heidelberg, 2011.
- [8]. V. K. Singh, P. Waila, Marisha, R. Piryani & A. Uddin, “Sentiment Analysis of Textual Reviews: Evaluating Machine Learning, Unsupervised and SentiWordNet Approaches”, In Proceedings of 5th International Conference of Knowledge and Smart Technologies, Burapha University, Thailand, Jan. 2013.
- [9]. F. Benamara, C. Cesarano & D. Reforigiato, “Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone”, Proceedings of ICWSM 2006, CO USA, 2006.
- [10]. metacritic <http://www.metacritic.com/>
- [11]. Sentiment Analysis API : <https://sentigem.com/>