

Design of Opinion Mining Based Movie Rating System

Abhishek Deshmukh¹, Atul Karande²,
Intkhab Alam Shaikh³, Shagufta Rajguru⁴

¹²³⁴Department of Computer Engineering, Fr.C.R.I.T. Vashi, Navi Mumbai, India

Abstract: The opinion of people about a product or a service is considered to determine its quality. This design paper implements the same concept to analyze and evaluate user reviews about movie and predicts the opinion of people in terms of sentiment rating associated with each movie. It gives brief description about the steps involved in sentiment analysis and its application on automated movie rating system to rate movies based on reviews provided by people.

Keywords: sentiment analysis, opinion mining, natural language processing, movie rating, user reviews.

I. INTRODUCTION

Once the products are reviewed by customers and critics, text analysis field is useful. It helps in decision making and prediction procedures. Also for a movie, users and critics add their reviews about a particular movie, which helps other users in deciding whether to watch the movie or not. Today there is tremendous data available on social sites, blogs etc. which can help to predict public opinion, OBMRS will be useful for such users. Similar work is seen in IMDB and rotten tomatoes. IMDB as name specifies is Internet Movie Database. IMDB calculates overall score based on users rating from 1 to 10. The TomatoMeter is another such online movie analyzer that rates a movie as fresh if it has a percentage above 60 and rotten otherwise.

OBMRS does not just rely on numbers, the system analyses user comment, check for sentimental keywords and divides it into positive, neutral or negative. System consists of a sentiment library designed for English sentiment analysis. Opinion mining typically occurs in two or three stages (1) The input text is split into sections, such as sentences, and each section tested to see if it contains any sentiment: if it is subjective or objective. (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 gathers all comments for a particular movie and then calculates an average rating to score it. This score is generated for every movie in the system. The system also sorts and displays top rating movies as per analysis and calculates a top ten list automatically. This provides an automated movie rating system based on sentiment analysis. Important step in analysis process is tokenization. Tokenization is the process of breaking a stream of text into meaningful words (stems), phrases or symbols. The tokens can be used further for parsing (syntactic analysis) or text mining. Tokenization is generally considered easy relative to other tasks in text mining and also one of the uninteresting phases.

A. Organization of the paper

The organization of different sections of this design paper is explained below:

Section 1: Includes the abstract and introduction of the topic “opinion mining based movie rating system.

Section 2 A: Illustrates block diagram and contains brief explanation of different modules of the system.

Section 2 B: Describes the workflow of the system with help of flow diagram

Section 3: It concludes and summarizes the entire paper.

Section 4: Future enhancements that can be made to this project are listed.

Section 5: Lists the references.

II. PROPOSED SYSTEM

The opinion mining based movie rating system is basically an online automated movie rating system where information related to movies will be stored and displayed. In addition to this, users who are registered with the system can add their reviews of any particular movie. The system will calculate the rating by analyzing these user reviews using the concepts of Sentiment Analysis or also called Opinion Mining. The system mainly contains 7 modules as shown in the following block diagram of system.

Following section will describe all the modules of the system.

A. User Accounts

1. Admin Account

Administrator account will manage and maintain the databases of the system and authenticate and approve the user request in the registration process. Admin will also be responsible for calculating and displaying the rating with each movie present in the system. Database activities include adding new entries into database, updating those entries, deleting the wrong entries from those databases, etc.

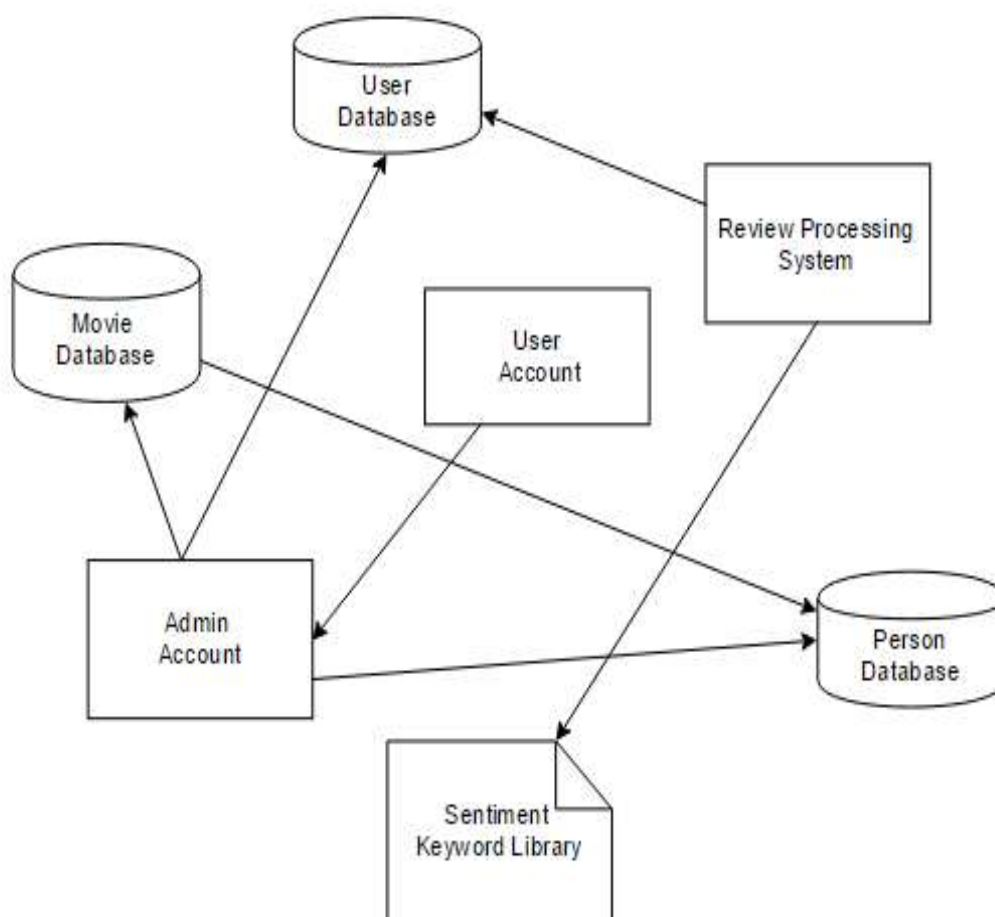


Fig. 1. Block Diagram

2. Registered User Account

The user can search a movie or person i.e. actor, director, singer, writer, etc. by name. The user does not have to be registered to just check the information, view the movie/person page, see rating to any particular movie, 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 add reviews to any movie.

B. Databases

1. Movie Database

This is the central module of the system. It will contain name, cast, year, plot, rating, reviews and other movie information. The cast and review section will link to person database and review processing system respectively.

The rating which will be displayed with each movie will come from the review processing module. As mentioned earlier each and every entry in the database will be made by the admin only.

2. User Database

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

3. Person Database

In this module information of people involved or related to movie will be stored. These people can be actors of the movie, director, singer, screenwriter, etc. The information will be in the form of name, work, age, biography, nationality, photos, 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 will be 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

The previously discussed module implements the concept of sentiment analysis on user reviews to generate a sentiment score, but the algorithm that module uses will include a library of sentimental keywords from where it will compare the keywords found in reviews. So, this library will contain possible sets of positive, negative and neutral keywords with respective sentiment scores. In this project, this library will be implemented using the SentiWordNet which is an online lexical resource for opinion mining.

Now, the flow chart of the system is described in the following section:

1. **Registration:** The very first stage is to register with the system. 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. Afterwards user can change some the username and update his/her profile and set profile picture.
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 can go through the biography of cast.
4. **Review/Comment:** User can post reviews for any movie and see the reviews posted by others. Based on other user's reviews, user can now make a decision for that movie.
5. **Logout:** At the end user can simply logout just by clicking logout button provided on the page.

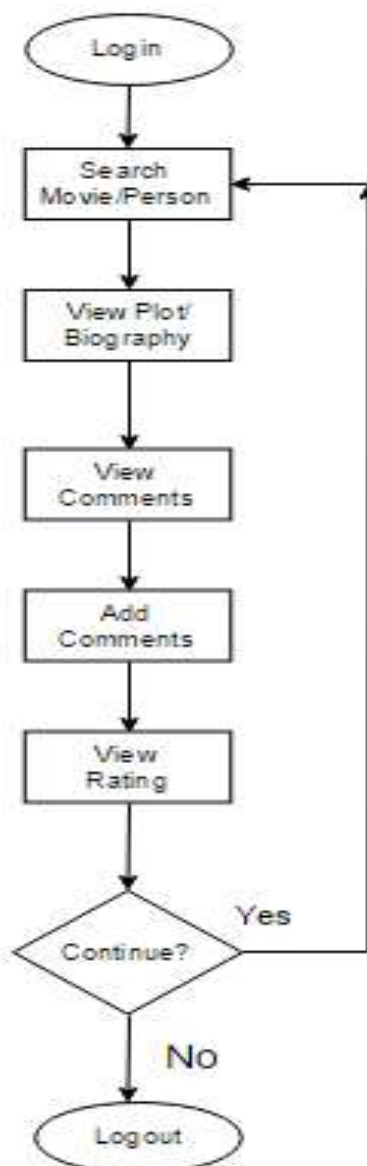


Fig. 2. Flow Chart

Conclusion

So, basically this project aims to achieve the development of an online automated movie rating system using the concepts of sentiment analysis and data mining. Based on the user's reviews for movies stored in database, the system will provide the effective rating to the movie on the sentiment scale of 1-10. This rating will help other users in getting the correct information regarding the movie and effective review about it. The user will be able to access every information related to movies and persons involved in the movie from the system.

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 paper includes sentimental keyword library for English language only, in future same library for other language can be created.
3. Here, in this paper 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.

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.
- [6] 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.