

Sentiment Analysis based Music Ratification

Amit Menon¹ Forum Bhayani² Krishnanjali Shinde³

^{1,2}BE Student ³Professor

^{1,2,3}Department of Computer Engineering

^{1,2,3}Atharva College of Engineering, Mumbai, India

Abstract— The aim of this paper is to present an approach to music rating by performing sentiment analysis using supervised learning methods. The recent boost of use of web has resulted in large quantities of user generated data being created every day. People use the web now, to review any product from different sectors like food, electronics. Enterprises too have realized the value of online feedback for future sales. Though sentiment analysis has largely been employed in movie reviews and product recommendation systems, it has largely been kept away for the music portal. The current music charts largely rely on the number of downloads, number of air plays, and physical sales of the songs. This paper though limited to theoretical discussion, aims to create a new ranking system with more emphasis on user opinion, to see if there is a disparity in these two approaches and also reviews the related work done in this field.

Key words: Sentiment, Music, Analysis, Reviews, Ratings

I. INTRODUCTION

A. Introduction:

With the turn of the century, the Internet or specifically the Web 2.0 has seen a rapid spurt in the number of users. With this increase in the use of internet, there has also been a rapid increase in the online "presence" of brands, celebrities, politicians and people in general. "Online presence" here refers to the online activity of the concerned entity. Since a wide plethora of products are now available online for perusal, people have turned to internet itself to gauge the authenticity efficiency or in layman terms the goodness of the product. This tendency is not relatively new. Quite often the decision of people is based on what others think. Even before internet, we often relied on our friends to refer us to a good doctor, or a mechanic. Now the internet has made it easier for people with appropriate experience to be connected with people needing advice.

Pang and Lee [1] provide following statistics as evidence to the point that there is a reliance on the online media for evaluation of certain product.

- 1) 81% of internet users have done online research on a product at least once
- 2) On a random given day 20% users do a research on a product
- 3) Of all the readers of online reviews of hotels, restaurants, and various services, 73%-83% of them reported that their decision was significantly influenced by those reviews
- 4) Consumers are willing to spend more for a higher rated product
- 5) 32% have rated a product and 30% have posted and online comment about a commodity.

Thus it is evident that there is a growing interest in systems that can extract a sentiment from a substantial volume of data as it is beneficial for large number of online users as well as for brands and services in helping them assess their product from feedbacks.

B. Motivation:

Sentiment analysis though a relatively new field, has already experienced a steady undercurrent of interest. Hundreds of research papers have been published already since its recognition in late 90s. It has been extensively applied in various fields like movie reviews, politics, and search engines. The online music portal has largely refrained from adopting this new technique of assessment, and has depended on conventional methods of rating. This maybe because:

- 1) Music or song is not an essential commodity or service
- 2) No great resources are spent while listening to a song. Due to this, owners of various websites like Amazon, iTunes came up with star ratings which help users to give their opinion about a product. These star ratings come handy for other users while determining their likes and dislikes.

Various web applications for music available like Savan, Gaana, Billboard provide us with 'top 100' or 'weekly top' charts. These ranking of songs is done by using the number of downloads from iTunes, Amazon etc. which is often misleading. However, our argument is based on the fact that these ratings have certain shortcomings, that is, it is not subjective and doesn't explain a person's experience in a holistic manner.

To elaborate, we listen to a certain song sometimes just because it is suggested to us by a friend. It may so happen that we may not like the song after downloading and listening to it. Thus downloads become a partial criteria for appreciation of the song. Moreover since a song is not as essential commodity, a person still willing to comment (either positive or negative) shows a greater influence on the person. Hence that comment should be valued more.

There is also a considerable likelihood that the data is skewed. Of those who respond the vast majority will be users who will provide a positive rating; they're using it over competitive alternatives and want to justify their decision. Very few of the population will be non-users who have conducted an unbiased assessment of the available alternatives. Hence, having more weight-age on comment analysis provide help us in having a more precise way to decide a user's taste and likes.

C. Background:

Sentiment Analysis also known as opinion mining derives the opinion, or the attitude of a user. It is used to determine whether a piece of text is positive, negative or neutral. It is done by building a system which collects and categorize the opinions about a product. It gives a genuine inference that a particular user has liked or disliked something.

Sentiment Analysis has grown tremendously in a short span of time in terms of recognition in as an important area of research. Thus it has been well applied in different areas. [1] Pang and Lee state some applications as

- 1) Review related websites
- 2) Subcomponent technology in tools like search engines or recommendation systems.
- 3) Applications in business governance and intelligence

More importantly it can be used to analyse sales and product information by making sense of unstructured and different forms of data available on the internet.

A rating system based on Sentiment Analysis can be used in areas where descriptive reviews of system are beneficial and necessary. For example it can be used in areas of government bureaucracy, movie reviews. In a police system hierarchy, a DSP can use it to compare the functioning of police stations under him, which are rated by the citizens.

As seen above, consumers today rely heavily, on review related websites before purchasing a product. Music is also one, though not essential commodity. And thus the normal rules of economics also apply to it.

In [2] Godes and Mayzlin state that in spite of various available means of reviews the most common and effective tool of spreading awareness of a product is still word of mouth (WOM). This is exceptionally true for a song or music in general as our inclination to listen to a song has a direct correlation with what people around us listen to. However our argument is based on the fact that it does not guarantee that the song is well received.

II. REVIEW OF LITERATURE

[5] Gives a brief introduction to the supervised training methods in machine learning. It begins with the terms and processes involved in supervised learning, applications such as classification and regression, and then moves on to the challenges associated with it such as computational complexity, errors and scalability.

[1]Pang and Lee delve deep into the intricacies of sentiment analysis. They begin with the early history of the field introducing the reader about what exactly sentiment analysis is, moving on to the difficulties and applications. They also talk about classification problems and interestingly discuss about analysing the degree of positivity and negativity in a statement, which parallels our own discussion. They also discuss problems of multiclass classification and agreement detection to a certain topic. They closely analyse the relation between class viewpoints and perspectives in sentiment analysis.

Next the features important to this classification are discussed, in which the effectiveness of bigrams and unigrams are compared; parts of speech, syntax and negation are discussed, which we found helpful for our classification of comments. The paper also discusses unsupervised methods for sentiment analysis, economic impact of the same.

[3]Lak and Turekten provide a detailed comparison between star ratings and sentiment analysis as a tool for rating. Star rating is a quick measure of opinion in online world. It is brief, opinionated and always polarized. They act as a summary for a long review. However they fail when a detailed explanation is needed. They analyse sentiment analysis as a tool for rating in areas where information is unstructured and not extremely polarized. In their paper they conclude that there is a certain discrepancy between these two approaches as sentiment analysis is a limited in its ability to detect extremeness. This is because natural language is more neutral in nature. They also conclude that sentiment analysis is a better tool to capture general sentiment of review. They state that these metrics can be used in a complementary sense to each other and not as an alternative.

Pagin and Wimmers [4] discuss various ways for incorporating weights in different scoring metrics. They concern themselves only with the task on weight incorporation and not with the actual scoring techniques. In their paper they discuss a general rubric to use a weighted scoring system which is universally applicable to any scoring rule. They visualize a scoring rule as a tuple where each entry maybe a search term or a number. The important question being addressed is how to incorporate weights in such a way that it gives more weightage to certain attribute. For instance in our case, we wish to device a scoring system, that gives more weightage to the score from users comments when assigning a final rating to the song.

III. IMPLEMENTATION DETAILS

We aim to create a music web application that gives a platform to the user to give their opinions about particular song. This web app will contain a database of users and songs (limited in number) with a provision for users to comment about the song. This expression is then mined using sentiment analysis and each comment is scored according to the degree of positivity and negativity. Thus each song is assigned a final score based on the normalized scores from this analysis, likes and downloads. Our System has the following components:

- 1) Web app: This is the main platform. It will be built in python using Django framework. It will have pages for authentication, charts and comments.
- 2) Analyser: This analyses comments to give them a score.
- 3) Charts: The final ranking of songs will be displayed on this page which will be a result of the cumulative score.

The web app will have basic set of pages namely a landing page, Authentication (login/signup) page, Home page. Apart from that it will have pages for displaying charts for weekly and yearly top songs separately for English and Hindi songs.

Fig.1. Shows a lower level data flow diagram for the proposed web application. This gives a broad level idea about the transfer of data from different modules.

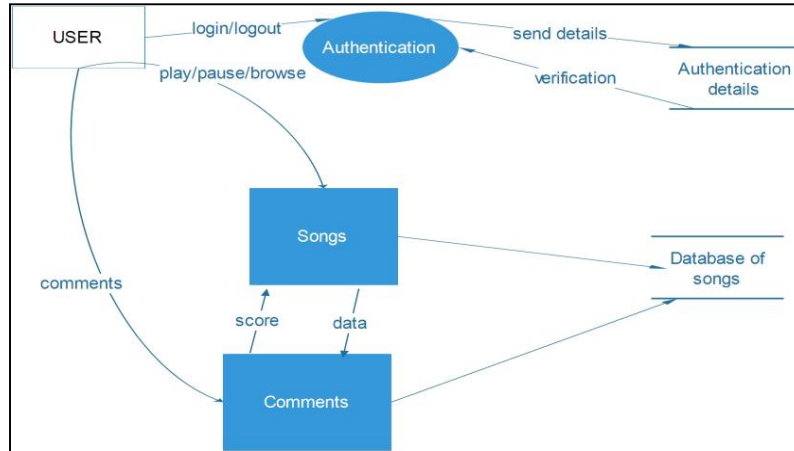


Fig. 1: Data flow Diagram

A. Technical Details

1) Software Details

a) Django:

The web app will be built on Django. Django is a high-level Python Web framework that is used for rapid development and clean, pragmatic design. It is built by experienced developers and handles many of the hassles of Web development independently. It's free and open source. It is also known for its security features

b) Python NLTK:

We also make use of this library which provides us with a plethora of NLP tool. Besides it has more than 50 corpora and lexical resources. It makes the task of pre-processing, classification tokenization, stemming, parsing and semantic reasoning.

c) My SQL for Database:

My SQL database is used to maintain datasets that application would use to verify authentication and details of the user.

d) Bootstrap:

Bootstrap is the most easy to use HTML, CSS, and JS framework for developing responsive, mobile first projects on the web.

2) Hardware details

- OS: Linux/Windows
- Ram: Minimum 4 GB
- Processor: 3rd Gen i3
- Hard disk: 200 GB

B. Analyser:

The analyser is the main component of our system. The main task of the analyser is to analyse each comment made on a particular song according to the positivity and negativity of the comment. This is then used in the scoring metric to impart an overall score to the song. This score is then used to rank and display the song accordingly.

Fig. 2 shows the internal working of the analyser i.e. the steps associated with the process.

Since already classified comments for songs are not easily available we make use of a combination of corpora from the sites of Cornell University and corpora available from the NLTK library. The Cornell university corpus contains classified movie reviews and the NLTK library contains general positive and negative statements. This combination is used to obtain a better performance from the analyser.

We rate each comment based on the number of positively identified words and negatively identified words. The final score is summation of positive words and summation of negative words. This is done for every comment and a final score is given to each song which gives more weightage to the score from comments.

To implement a full scale music web app with all is quite a task so we reduce the scope of this app by putting the following constraints.

- 1) Only 100 songs (50 English and 50 Hindi) will be used
- 2) Comments will be limited to English language only.
- 3) Each user is only allowed to comment once
- 4) No URL's are allowed
- 5) Each comment limited to 100 words

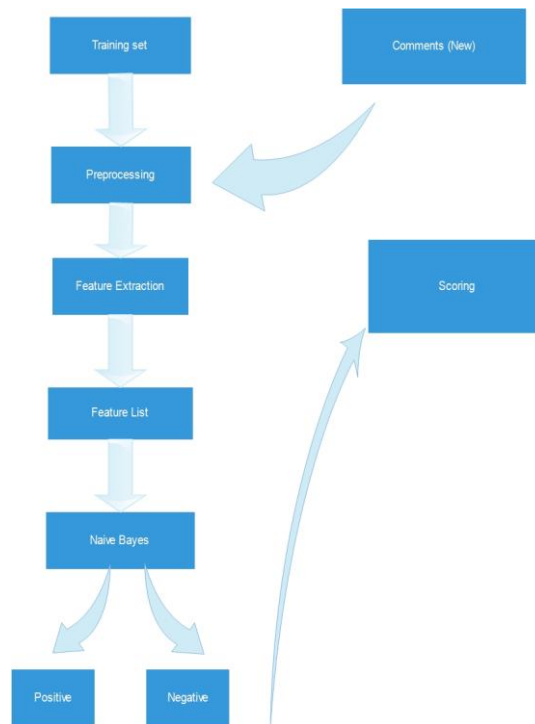


Fig. 2: Working of the analyser

IV. CONCLUSION

Music or a song is not a required commodity. It does not take a large number of resources to listen to a song. Star ratings based on physical factors provide a quick measure of the goodness for products belong to different sectors like, movies, electronics, food, hotels, etc. For these products user comments just provide an additional measure or an in depth review of the product. But since music is not a required commodity, a person explicitly stating his opinion for a song is an indication of genuinely liking or disliking the song. A rating system of songs which gives more weightage to the comments is bound to produce a substantial deviation from normal results. Though this may be true for other products too, it gives a much more holistic approach in the music portal and hence is choice for our project.

ACKNOWLEDGMENT

We would like to thank our project guide Professor Krishnajali Shinde for her enormous co-operation and guidance. We sincerely thank her for wholeheartedly supporting the project and gave freely of her valuable time while making this project. All the inputs given by her have found a place in the project. The technical guidance provided by her was more than useful and made the project successful. She has always been a source of inspiration for us. It was memorable experience learning under such a highly innovative, enthusiastic and hardworking teacher. We are also thankful to our Principal Dr. S. P. Kallurkar, our HOD Professor Mahendra Patil, our Project co-ordinator Professor Deepali Maste and all the staff members of the Computers department who have provided us various facilities and guided us to develop a very good project idea. Finally, we would also like to thank teachers of our college and friends who guided and helped us while working on the project

REFERENCES

- [1] Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis" Foundations and Trends in Information Retrieval Vol. 2, No 1-2 (2008) 1–135, 2008
- [2] D. Godes, and D. Mayzlin, "Using online conversations to study word-of-mouth communication" Marketing Science, 2004
- [3] Parisa Lak, and Ozgur Turekten "Ratings versus Sentiment Analysis- A Comparison of Explicit and Implicit Measures of Opinions", 47th Hawaii International Conference on System Science, 2014
- [4] Ronald Fagin, Edward L. Wimmers, A formula for incorporating weights into scoring rule", Theoretical Computer Science 239, 309–338, 2000
- [5] Oded Maimon , and Lior Rokach, "Introduction to supervised methods", Data mining and knowledge discovery handbook, 2007