International Journal of Management, IT & Engineering
Vol. 8 Issue 5, May 2018,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

TEXT AND NETWORK BASED ANALYSIS OF TV SHOWS USING MINING TECHNIQUES

T.Maanasa*

Abstract

Television in India is becoming huge industry; it has thousands of programs in so many languages. Every year the small screen has been producing numerous celebrities. More than half of all Indian households own a TV. The television production industry has maintained its importance over the past few decades. The precise and timely prediction of program popularity is of great value added for content suppliers, broadcast TV operators and advertisers to get the financial gain. The predicted information can be beneficial for operators in TV program purchasing decisions and it can help advertisers to formulate reasonable advertisement investment plans. Besides this a technical matters, a precise program popularity prediction optimizes the whole broadcasting system, like the content delivery network strategy. The objective of this paper is to analyze the programmes popularity based on the people sentiments and interest for this data mining technology has been adopted. The

Keywords:

Data Minning;
Text Mining;
TV Show;
NAnalysis;

* **T. Maanasa, Assistant Prof., Chaitanya Bharthi Institute of Technolgoy, Gandipet, Hyderabad, India.**

Popularity.	concentrate is on points related to the user ratings of programmes, like discovering if sentimental, big-budget programmes are more popular than their low budget series of comedy, drama, games, news and sports. The analysis also predicted the relationship between whether any particular actors or actresses are likely to help a programme to succeed.

1. Introduction

Video watching in India has been evolving even though with limited infrastructure. The surveys showed in 2016, the country has a collection of over 857 channels of which 184 are pay channels including regional languages. India is the third largest online market with viewers of 50 million in addition, It is observed an another trend – today big brands promoting their presence on major social networks such as Facebook and Twitter, also they are making their presence felt on YouTube and bringing out creative too. The idiot box has not been left behind; it is also realized that there is a section of people that loves to use content online which leads for quite some time, major television networks created their presence on live streamings and have become huge with time. Thanks to recent advances as the world of measurement is changing in data collection, transfer, storage and analysis. Big or large data does not guarantee good data, and robust research methodologies are more vital than ever.

Several supervised learning models to predict the popularity of TV series using viewer ratings on IMDb as an indicator, and then used unsupervised learning to investigate the key features of the TV series. These key features obtained by unsupervised learning steps were then be used to improve our current prediction models [1]. In particular, the classification model with ratings divided into three subgroups provides the best outcome and is recommended for prediction, although the outcome falls in a relatively wide range. Nevertheless, linear regression using

selected features, either by using backward search or PCA, provides improved results compared to linear regression with all available features. If given more time, we will try using more sophisticated models to run on the data, for example, neuro network and kernel; we will also analyze the TV series ratings based on different time spans to see how the importance of features changes over time [2]. Evaluated using data collected from Jiangsu Cloud-media TV, which is one of the largest broadcast TV platforms in China. Several prediction models have been proposed based on video-on-demand (VOD) data from YouKu, YouTube, and Twitter. However, existing prediction methods usually require a large quantity of samples and long training time, and the prediction accuracy is poor for programs that experience a high peak or sharp decrease in popularity. This paper presents our improved prediction approach based on trend detection. First, a dynamic time warping-distance-based K-medoids algorithm is applied to group programs' popularity evolution into four trends. Then, four trend-specific prediction models are built separately using random forests regression [3]. CNN was the first news channel, which made an entry into India via satellite during the Gulf war, i.e. in early 1991. After that the Public television broadcasting system of India i.e. Doordarshan was challenged by 40 private channels in the 90's which included STAR-TV owned by Rupert Murdoch's news Corporation; SONY, owned by SONY Corporation of Japan and ZEE-TV owned by Subhash Chandra from India. Responding to the competition from STAR TV, Doordarshan supplemented its regional-language channels and the national network with five new satellite channels which provided programming similar to STAR TV. STAR had to use Hindi programming in its Zee Channel to capture the Indian audience [4]. Although a large number of research papers are focused on exploring the factors affecting movie ratings, very few studies have looked into TV series. While movies and TV series have similarities, TV series are different from movies in many aspects that worth investigating. The Internet Movie Database (IMDb) is a comprehensive online database that has a high degree of interactions with users, making it a fertile source of machine learning problems [5]. A machine learning approach has been provided for the prediction of movie popularity classification. This is a novel approach where the user rating decisions has been taken to purview along with inherent movie attributes to model the classification approach. An experimental insight has also been provided for the post release aspect of the movie that relates initial budget with each of the financial returns [6]. The Internet Movie Database (IMDb), a free, user-maintained, online resource of production details for over 390,000 movies, television series

and video games, which contains information such as title, genre, box-office taking, cast credits and user's ratings [7]. We demonstrate our broadcast TV data mining system with two end-user applications that utilize rich text content in time-continuous video. Novelty concept extraction produces a semantic content description that facilitates finding new and relevant information from dynamically updating TV program index in a content-rich manner. We demonstrate the applicability of our data mining system with new end-user services: Catch-up TV Guide for browsing recently aired programs and Novelty Cloud for quick overviews of broadcast news topics. Novelty word summaries were the most popular way to examine program content in our Catchup TV Guide service [8].

2. Telivision sector in Indian economy

The share of Indian television sector in economy has been presented in the following as shown in Figure 1.

- The television pie is continues to be the most dominated segment in the entertainment industry, with the share accounting of 44.24 % revenue in 2016, which is expected to increase further to 48.18 % by 2021.
- A share of 79.54 % which includes television, print and films together for marketshare in 2016, in terms of value.
- Print media occupied 2nd largest sector in the overall entertainment industry in India, following which sectors of Out of Home (OOH) and Radio are predicted to contribute almost 2 % each to the entire industry by 2021.
- Numbers shows that Indian print media industry generated revenues worth US$ 4.51 billion in FY2017 (till December 2016).

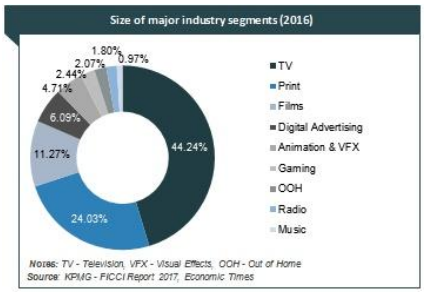

Figure 1. Pie chart showing share of print and media

The typical Indian soap opera the most common genre on Indian TV. Crime, thriller, drama and fiction shows are extremely popular among audiences of Indian origion compared to science & technology and news, as they reflects real family issues depict in a melodramatic fashion.

There are thousands of TV programs in India, all ranging in length, air time, genre and language. The Hindi television industry is by far the biggest. However, some have much greater influence on the audiences, like Zee, Colors, Star Plus, Sony etc. the organizations which perform the study on audience interst and give ratings are:

1. Doordarshan Audience Research Team-DART (During the days of the single channel Doordarsha)
2. Indian National Television Audience Measurement-TAM (backed by AC Nielsen) & INTAM (introduced by vested commercial interests)
3. Audience Measurement Analytics Limited -aMap (funded by American NRI investors)
4. Broadcast Audience Research Council- real-time audience metrics system

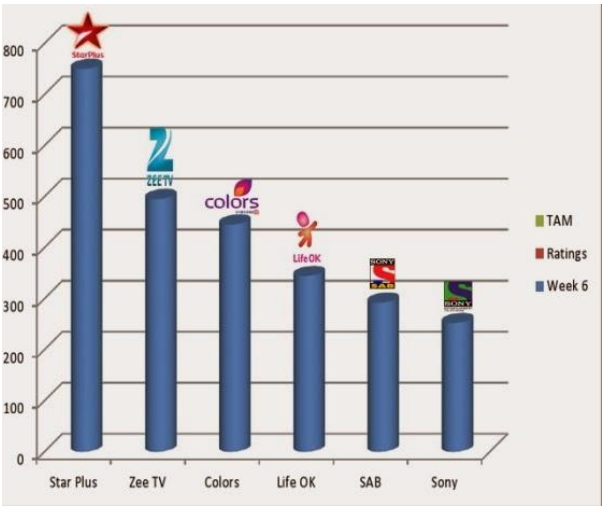

Figure 2. Ratings (courtacy: tellytrp.in)

3. Data variables evaluated for predictions and discussion

The following parameters have been observed to analyse the propability ofTV show ratings

Table 1. Data variables used

	Details	Data Example
Program features	Assessing elemnts	Drama, comedy, social cause, sports, games, science and technology
Program performance	Rating	History of TRP on week/month/year basis
Promotional aid	Investment on promotional trailer in time intervals , social media	Spending on air promos
Audience session	Interest of audience	Brand, cast, time of show etc.
Social/on-line behavior	Media informaiton	Fcebook, Twitter, YouTube etc.

1. **Sentiments and Drama**: Most of the Indian families relay on middle class attitude therefore melodrama seies are most popular examples like Yeh hi mohabathein, Diya aur bathei hum, kumkum bhagya etc.

2. **Scocial Issues** : the need to rethink on print and media's investment in women's issues arose out of the understanding, that the nation could not improve and socioeconomic development would remain disfigure as long as women were left behind as the lesser half of the society. The main aspects of programmes for women are to create awareness of their role and responsibility as social beings along with men and specific interest of women and their role in the social and family structure.

Case study: Balika Vadu on colors TV is one of such kind which represented the women and social issue. It represented the story of rural Rajasthan which revolved around the life of a child bride from childhood to womanhood. This helped to re think on chaild marriages in India and it holded TRP rating of 4.

3. **Cultural/Religion**:

Ramayan and Mahabharat are the Indian epics which are not based on any social develomnetal issues but on hindu mythologies and are targeted at the middle-class hindu families which constituted a large section of the viewers and it depicted as cultural rather than religious containing universalistic values applicable to all human kind.

Hara hara mahadev, Siyaki Ram, Ganesh, Shani mahimas are such TV shows that rated on top for months.

4. Reality shows: shows like kon banegha karor pathi which is hosted by the legendary Indian cinema star Mr.A mithab Bachan and Bollywood Badsha Mr.Sharuk Khan which turned into a big hit on small screen which aired repeatedly based on odians poll and trp ratings. Another example is Big Boss done by Mr.Salman Khan was also a viewers favourate show.

5. Music/sports/games/news: these attributes are specially ment for particular gender and age groups.

The method adopted and architexture in this work is shown in Figure 3. By this it is capable of different topics using traditional text-based methods as well as a novel technique which uses channels specific network information. Text-based approach consists of textual data such as topic definition, viewer opinion and pre-assigned labels form the text document. This Document is then transformed into tokens along with their frequencies to infer data for learning. The framework then uses traditional classification kernels to predict the probability of class. In the network-based approach, the channel class is predicted based on the similarity between two topics. It is defined using viewer similarity metric which assumes, if there is significant overlap among influential users generating content on two programmes, then there is close relationship between these two topics, therefore, the network based model can give the class of the famous channels using the categories of its similar parameters.

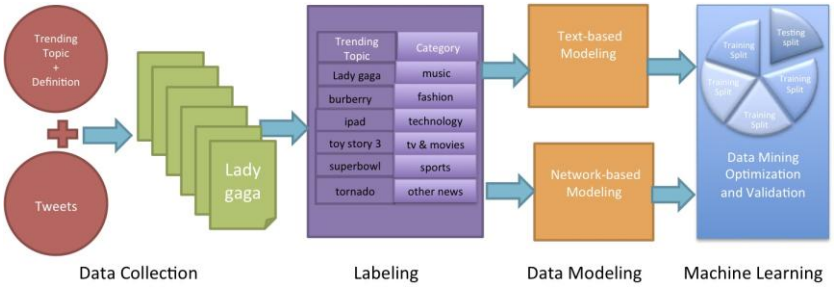

Figure 3. System Architecture

The goal of this framework is to classify trending topics (news or event for which messages are in high volume) on Twitter into known categories such as sports, politics etc. The framework is capable of classifying topics using traditional text-based techniques as well as a novel technique which uses Twitter specific network information (network-based). In text-based approach textual data such as topic definition, tweets and pre-assigned labels form the text document. Document is transformed into word tokens along with their frequencies to infer data for learning. The framework then uses traditional classification kernels (e.g. Naive Bayes, SVM) to predict the topic class based on the previously trained examples. In the network-based approach, the topic class is predicted based on the class of topics similar to it. Similarity between two topics is defined using User Similarity metric which assumes that if there is significant overlap among influential users generating content on two topics, then there is close relationship between these two topics. Therefore, the network based model can predict the class of the target topic using the categories of its similar topics. Since the network model is dependent on finding similar topics, our next step would be to integrate the two models in the framework so that text based model can be employed when the network based model cannot be employed.

Based on viewrs constraints and interests the TV channels share is shown in Figure 4.

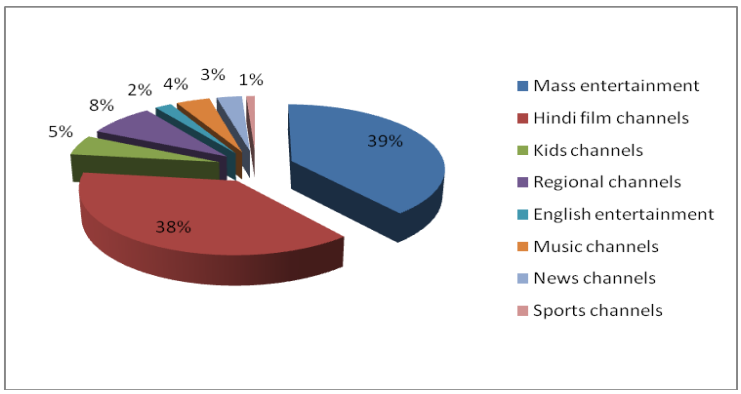

Figure 3. Telivision Viewrs share in India (Source:TAM)

4. Conclusion

From the study it is concluded that the TV is the largest entertainment medium which has a major economy share compared to movies, radio and print. It was observed that DD, startplus, sony, colors, zee, life ok are the key channels that are viewed by the users. DD channel one of the largest network that reaches to remote users. Zee entertainment has more of 22 channels available in all regional languages ranging from melodrama series to news but lagged behind starplus on other hand star plus entered as mass entertainment, sports, music, news, life style, movies etc. cand continuing to go rapidly with extended seervices in asiapacific countries honking, thailan, india, Pakistan, chaina, tiwan, korea, dubai, Singapore, japan, phillipines, it is also extended their markets by taking over regional language channels like maa in south india. Sony channel specialized in exlusive sports, movies both hindi and english.

References

[1] Yushu Chai, Yiwen Xu, Zihui Liu, “Behind the TV Shows: Top-Rated Series Characterization and Audience Rating Prediction”.

[2] Scott Sereday and Jingsong Cui, Data Science, Nielsen , “using machine learning to predict future tv ratings”.

[3] Chengang Zhu ; Guang Cheng ; Kun Wang, “Big Data Analytics for Program Popularity Prediction in Broadcast TV Industries”, *IEEE Vol.5, EISSN:2169-3536, 27 October 2017*

[4] Indian Television: From Edutainment to Entertainment

[5] Augustine, Achal, and Manas Pathak, "User rating prediction for movies", *Technical report, University of Texas at Austin, 2008.*

[6] Khalid Ibnal Asad , Tanvir Ahmed , Md. Saiedur Rahman, "Movie Popularity Classification based on Inherent Movie Attributes using C4.5, PART and Correlation Coefficient", *IEEE/OSA/IAPR International Conference on Informatics, Electronics & Vision.*

[7] M. Saraee, S. White & J. Eccleston, "A Data Mining Approach To Analysis And Prediction Of Movie Ratings", *University of Salford, England, 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9.*

[8] Mika Rautiainen, Jouni Sarvanko , Arto Heikkinen, Mika Ylianttila, Vassilis Kostakos , "An Online System with End-User Services: Mining Novelty Concepts from TV Broadcast Subtitles", *ACM 978-1-4503-2174-7/13/08.*