Volume 3, Issue 6, June 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Predicting Movie Success: Review of Existing Literature

Arundeep Kaur, AP Gurpinder Kaur
Department of Computer Science",
Swami ViVekanand Institute of Engineering
& Technology, Punjab Technical University, Jalandhar, India

Abstract- Indian Hindi Cinema industry popularly known as Bollywood has reached staggering proportions in terms of volume of business (184.3 billion), manpower employment (over 6 million workers), movies produced (more than 100 in a year) and its reach (exported to more than 100 countries worldwide). With so much at stake and highly uncertain nature of returns, it is of commercial interest to develop a model which can predict success of a movie. This however, is not an easy work, since movies have been described as experience goods with very less shelf life; it is difficult to forecast demand for a movie. There are number of parameters that may influence success of a movie like – time of its release, marketing gimmicks, lead actor, lead actress, director, producer, writer, music director – being some of the factors. The present study aims to develop a model based upon neural networks that may help in predicting the success of a movie in advance thereby reducing certain level of uncertainty.

Keywords – Movie, Neural Networks, ncertainty,

I. Introduction

Today, the trouble is that the more things change, the more they stay in the same horizons. However, this may not be time for the movie industry, as it can break completely free of the cycles which had marked its history for hundreds of years , and it will be in fact , a departure from reality , It's not predicting the future success of movie is problematical, it's the realization that you have to relive the past again and again and still make highly intelligent guess about the success and failure of the movie. An attempt is made to predict the past as well as the future of movie for the purpose of business certainty or simply a theoretical condition in which decision making [the success of the movie] is without risk, because the decision maker [movie makers and stake holders] has all the information about the exact outcome of the decision, before he or she makes the decision [release of the movie]. With over two million spectators a day and films exported to over 100 countries, the impact of Bollywood film industry is formidable. From the first Indian film "RajaHarishchandra by Dhundhiraj Govind (Dadasaheb) Phalke in 1913 to 1981, India produced over 15000 feature films. Since then it has produced, at least another 15000 at a rate of more than 1000 films a year (1091 in 2006, 1146 in 2007 and 1325 in 2008) in 26 languages [1]. The industry is world's largest in terms on number of movies produced and also in terms of number of cinema goers. Bollywood produces as many films as the next three largest producers – US, Japan and China- combined. In terms of money it is second only to Hollywood [2]. Now, film making in India is a multimillion dollar industry employing over 6 million workers and reaching millions of people worldwide. In 2008 industry was valued at 107.1 billion rupees. Pricewaterhouse Coopers [3] predict that industry will be 184.3 billion in 2013. With such a fortune and employment of so many people at stake every Friday, it will be of immense interest to producers to know the probability of success or failure of a movie. However, due to their definition as experience goods with short product life time cycles, it is difficult to forecast the demand for motion pictures. Nevertheless, producers and distributors of new movies need to forecast box-office results in an attempt to reduce the uncertainty in the motion picture business and as a stake holder in the movie industry , one needs to know then the minimum sum of money a he/she can accept to forgo the opportunity to participate in an event [make/distribute etc ., movie] for which the outcome [success or failure of movie], and therefore his or her receipt of a reward, is uncertain [success of the movie].

II. Related Work

Neural Networks have been extensively used in forecasting and prediction studies, and so it has been also employed for predicting the success and failure the movies also. However, Hadavandi *et al [*4] have used integration of genetic fuzzy systems and artificial neural networks for stock price forecasting .Stock market prediction is regarded as a challenging task in financial time-series forecasting. The central idea to successful stock market prediction is achieving

best results using minimum required input data and the least complex stock market model. They used stepwise regression analysis (SRA) to determine factors which have most influence on stock prices. In the next stage divided raw data into k clusters by means of self-organizing map (SOM) neural networks. Finally, all clusters were fed into independent GFS models with the ability of rule base extraction and data base tuning. They evaluated capability of the proposed approach by applying it on stock price data gathered from IT and Airlines sectors, and compare the outcomes with previous stock price forecasting methods using mean absolute percentage error (MAPE). Results show that the proposed approach outperforms all previous methods, so it can be considered as a suitable tool for stock price forecasting problems. An organization has to make the right decisions in time depending on demand information to enhance the commercial competitive advantage in a constantly fluctuating business environment. Therefore, estimating the demand quant i try for the next period most likely appears to be crucial. Efendigil et al [5] have suggested a decision support system for demand forecasting wit h artificial neural network and Neuro fuzzy logic. Neural networks have been used in predicting the traffic flow. Chan et al [6] have deployed neural networks based upon exponential smoothing method for predicting the traffic flow. Reuter and Muller [7] have developed artificial neural network for forecasting of fuzzy time series. Forecasting box office revenue of a movie before its theatrical release is a difficult and challenging problem. In a study conducted by Zhang et al, [8] a multi-layer BP neural network (MLBP) with multi -input and multi output was employed to build the prediction model. All the movies were divided into six categories ranged from "blob" to "bomb" according to their box office incomes, and the purpose is to predict a film into the right class. The selections of the input variables were based on market survey and their weight values were determined by using statistical method. As to the design of the neural network structure, theoretical guidance and plentiful experiments were combined to optimize the hidden layers" parameters which include the number of hidden layers and their node numbers. Then a classifier with dynamic thresholds was used to standardize the output for the first time, and to improve the robustness of the model to a high level. Finally, a 6-fold cross-validation experiment methodology was used to measure the performance of the prediction model. The comparison results with the MLP method showed that the MLBP prediction model achieves more satisfactory results, and it is more reliable and effective to solve the problem. A Bayesian belief network (BBN), which is known as a causal belief network, was constructed [9] to investigate the causal relationship among various movie attributes in the performance prediction of box-office success. Subsequently, sensitivity analysis was conducted to determine those attributes most critically related to box- office performance. Finally, the probability of a movie's box-office success was computed using the BBN model based on the domain knowledge from the value chain of theoretical motion pictures. The results confirmed the improved forecasting accuracy of the BBN model compared to artificial neural network and decision tree.

Asur and Huberman [10] in their paper have demonstrated how social media content can be used to predict real -world out comes. In particular, they have used the chatter from Twitter.com to forecast box-office revenues for movies. They have shown that a simple model built from the rate at which tweets are created about particular topics can outperform market-based predictors. Some studies have also been carried out and have been reported in ne was median and social media (YouTube), but no academic record for these studies is available. One such study was carried out by researchers of Indian Institute of Management Ahmadabad and other one was carried in China [11].The present study will use neural network based machine learning algorithm for predicting movie success.

III. METHODOLOGY

Experimental set up for research work: Following tools and frameworks have been used for the implementation of the proposed model for predicting the success or failure of a motion picture.

1. Statistics Toolbox of MATLAB 2011b provides algorithms and tools for organizing, analyzing, and modeling data. Statistics Toolbox includes specialized data types for organizing and accessing heterogeneous data as well using it for classification machine algorithms. Dataset arrays store numeric data, text, and metadata in a single data container. This tool has been used to analyze computations and to make result graphs.
2. Neural Network Toolbox of Mat lab 2011b provides functions and apps for modeling complex nonlinear systems that are not easily modeled with a closed-form equation. Neural Network Toolbox supports supervised learning with feed forward, radial basis, and dynamic networks. It also supports unsupervised learning with self-organizing maps and competitive layers. We used the toolbox to design, train, visualize, and simulate neural networks as well as for applications such as data fitting, pattern recognition, clustering, time-series prediction, and dynamic system modeling and control.

Implementation Steps

The entire study was conducted in a sequential manner as listed below:

i. Collection of data pertaining to parameters under study (Actor, Actress, Producer, Director, Writer, Music

Director , Time of Release and Marketing Budget)

ii. Data processing for assigning weights and calculating thresholds

iii. Input & Target pattern formation

iv. Architecture design and Neural Network Learning

v. Performance Analysis

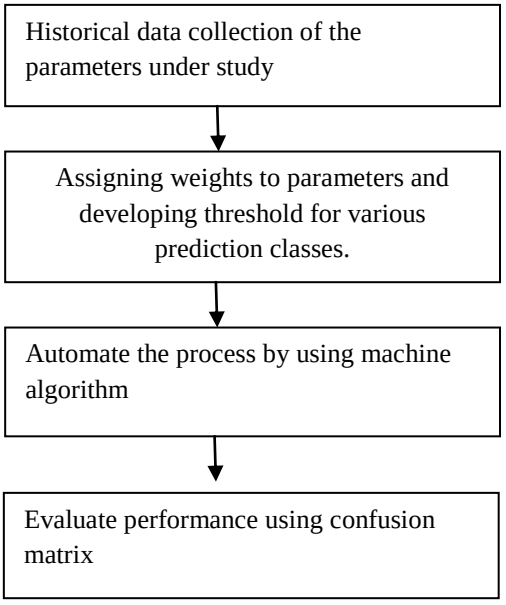

Fig1. Basic Flow of the research process

Step – 1 Since our research work will be driven mainly based on the observations from the historical data pertaining to the parameters under study it is anticipated that dataset will be non linear in nature. And will have data point's distribution in such a manner that it will be difficult to find the hyper plane. So, will be using SVM for feature extraction.

Step 2 Initialization of weight

Weights to the parameters under study were assigned as per the formulas given:

$$\frac{\sum Ah}{10} = Aw$$

Where A_W is weight assigned to the actor.

A_h is number of hit movies delivered out of the last 10 released movies of the concerned actor

$$\frac{\sum Ash}{10} = Asw$$

Where As_W is weight assigned to the actress.

A_h is number of hit movies delivered out of the last 10 released movies of the concerned actress

Similarly weights will be assigned to other factors (Director, Producer, Music Director, and Writer). Only two types of weights will be assigned for the Time of Release. Movies released during holiday season (Diwali, Dushera, Id, New Year, Christmas etc.) will be assigned a weight of 0.9. While the movies released during other time will be assigned weight 0.7.

For assigning weights to the parameter marketing budget, a base value of Rs.10 crores will be taken. Weights will then be assigned as per the following formula:

$$\frac{\sum MB}{10} = MBw$$

Where MB_W is weight assigned to the Marketing budget.

$\sum$ MB is the sum total of all the expenses incurred on promotion of the movie (TV advertisements, events, launch on you tube etc.)

However for our research work, all the weights and biases are set to small real random values between 0 &1

Table – 1 Attributes/Parameters under study

S.No	Attribute	Description	Mathematical Expressions
1	Actor	Leading Actor and status of his last 10 movies	$\frac{\sum A_h}{10} = A_w$
2	Actress	Leading Actress and status of his last 10 movies	$\frac{\sum A_{sh}}{10} = A_{Sw}$
3	Director	Director and status of his last 10 movies	$\frac{\sum A_d}{10} = A_d$
4	Producer	Producer and status of his last 10 movies	$\frac{\sum A_p}{10} = A_p$
5	Music Director	Music Director and status of his last 10 movies	$\frac{\sum A_m}{10} = A_m$
6	Writer	Writer and status of his last 10 movies	$\frac{\sum A_w}{10} = A_w$
7	Marketing Budget	Base value of Rs.10.00 crores	$\frac{\sum MB}{10} = MBw$
8	Time of Release		Release during holiday season =0.9 Release during other time =0.7

Step 4:- Selection of classifier for the above dataset: Dataset in our case is non-stationary as well as non-linear, so we will be using neural network for understanding the pattern and predicting the success of movie. Movie business in India is a billion dollar industry, employing thousands of people as such success or failure of a movie can have profound effect on the stake holders. It is therefore of prime interest of the stake holders to know how the movie will fare. The present study aims to develop a model for the same. Many machine learning algorithms are available; however for in this study we will be using Artificial Neural networks utilizing supervised learning.

Step – 5 Evaluation of the Results: Model so developed will be tested for actual prediction of the movies and efficacy will be accordingly established.

IV. Conclusion

As movies are defined as experience goods with short product lifetime cycles, it is difficult to forecast the demand for motion pictures. Nevertheless, producers and distributors of new movies need to forecast box-office results in an attempt to reduce the uncertainty in the motion picture business. The study intends to develop a model to predict the financial success of a movie.

V. Future Scope

Literature survey has revealed only two studies which have attempted to predict the success of movies. While one study uses Bayesian belief network to predict the success, the other one uses neural network for the same. Lee and Chang in their study using Bayesian Belief Network for predicting box office performance concluded that Bayesian Belief Networks were better in predicting the success as compared to neural networks. However, Zhang *et al* in their study concluded that the MLBP prediction model achieves more satisfactory results as compared with MLP method, and it is more reliable and effective to solve the problem. Since, not much work has been carried out in this area we intend to develop a model which can predict the financial success of the movie.

Acknowledgement

I am thankful to AP GURPINDER KAUR, ASST.PROF, SWAMI VIVEKANAND INSTITUTE OF ENGG. TECHNOLOGY, BANUR, for providing constent guidance and encouragement for this research work.

References

[1] Olivier Berlet. "*Bollywood/Africa: A Divorce*"? Black Camera 2010, vol. 2 (1), page 126-143.

[2] S.Camaeron. *"Handbook on Economics of Leisure"*. Edward Elgar Publishing, 01-Jan- 2011, page 299-311. [3] http://www.pwc.in/en_IN/in/assets/pdfs/PwC- Indian-Entertainment-and-Media-Outlook- 2009.pdf.

[4] E.Hadavandi, H.Shavandi, A.Ghanbari. *"Integration of Genetic Fuzzy Systems and Artificial Neural Networks for Stock Price Forecasting"*. Knowledge Based Systems 2010, vol. 23 (8), page 800-808.

[5] T. Efendigil, S.Onut, C.Kahraman. *"A Decision Support System for Demand Forecasting with Artificial Neural Network and Neuro Fuzzy Models: A Comparative Analysis"*. Expert Systems with Applications 2009, vol. 36 (3), part 2, 5697-5707.

[6] K.Y. Chan, T. S. Dhilon, J. Singh, E. Chang. *"Traffic Flow Forecasting Neural Network Based on Exponential Smoothing Method"*. 6th IEEE Conference on Industrial Electronics and Application 21-23 June 2011, page 376- 381.

[7] U.Reuter, B.Moller. *"Artificial Neural Network for Forecasting of Fuzzy Time Series"*. Computer Aided Civil and Infrastructure Engineering 2010, vol.25 (5), page 363-374.

[8] L.Zhang, J. Luo, S. Yanng. *"Forecasting Box Office Revenues of Movies with BP Neural Networks"*. Expert Systems with Applications 2009, vol. 36 (3) part 2, page 6580-6587.

[9] K.J.Lee, W. Chang. *"Bayesian Belief Network for Box Office Performance: A Case Study of Korean Movies"*. Expert Systems with Applications, 2009, vol. 36 (1), page 280-291.

[10] http://arxiv.org/abs/1003.5699 Assessed on February 18, 2013

[11] http://www.psfk.com/2013/01/movie-success- algorithm.html

