# Predicting Movie Success Using Neural Network

[1]Arundeep Kaur, [2]AP Nidhi

Department of computer Science, Swami Vivekanand Institute of Engineering & Technology,
Punjab Technical University, Jalandhar, India

**Abstract:** ***In this research work we have developed a mathematical model for predicting the success class [flop , hit , super hit] of the Indian movies, for doing this we have develop a methodology in which the historical data of each component [e. G actor , actress, director, music ]that influences the success or failure of a movie is given is due weightage and then based on multiple thresholds calculated on the basis of descriptive statistics of dataset of each component it is given class [flop , hit, super hit] label. This dataset is then subjected to neural network [LM] based learning algorithm for automating the process and results in terms of match between actual class labels and predicted labels are evaluated. Results show that our strategy of identifying the class of success is highly effective and accurate which apparent from the classification matrix also.***

**Keywords:** Movie prediction, neural network, weights of variables

## 1. 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 "Raja Harishchandra 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].

## 2. Research Gap

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 [2] 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* [1] in their study concluded that the MLBP prediction model [1] 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.

### 2.1 Rationale of the Study

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.

## 3. Proposed Work

For developing a model that can help to predict whether the movie flop, hit, or superhot, we propose that we need to create the historical data set relating to parameters that influence movie success and to develop an algorithm to assign weights and develop a mathematical model to automate and predict movie success and finally evaluate the performance of the algorithm to know how good or bad our movie prediction system is.

## 4. Implementation Steps

The entire study was conducted in a following manner as listed below [7]:

1) Collection of data pertaining to parameters under study (Actor, Actress, Producer, Director, Writer, Music Director Time of Release and Marketing Budget)
2) Data processing for assigning weights and calculating thresholds
3) Input & Target pattern formation
4) Architecture design and Neural Network Learning
5) Performance Analysis

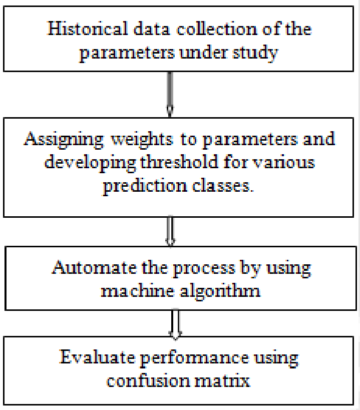


**Figure 4.1:** Basic Flow of the thesis

**Step 1:** Develop/collect a dataset based on following attributes [7]

**Table4.1:** Dataset characteristics of Neural Network input

| *S. No* | *Dataset Characteristics :* | *Multivariate* |
|---|---|---|
| a | Attribute Characteristic | Real Valued |
| b | Missing Values | None |
| c | Number of Instances of Observations: | 111 |
| d | Number of Attributes : | 07 |
| e | Parameter that influence characters | |

**Step 2:** Based on the above characteristics, assign weights to each features row consisting of following parameters [7]

**Table 4.2:** Attributes/Parameters under Study of Texture Based Observations

| 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 | $\frac{\sum A_m}{10} = A_p$ |
| 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 | $\frac{\sum MB}{10} = MB_w$ |
| 8 | Time of Release | | Release during holiday season =0.9<br>Release during other time =0.7 |

**Step 3:** Design neural network classifier with table (4.2) as input dataset for building learning validation and testing phases

**Step 4:** Design of output/target layer

| Target | Classes | Target Pattern |
|---|---|---|
| Class A | Flop | 1 0 0 0 0 0 |
| Class B | Hit | 0 1 0 0 0 0 |
| Class C | Superhit | 0 0 1 0 0 0 |

**Step 5:** Run neural network based on LM algorithm [3] having different configuration of hidden layers to finally find I-H-O architectures combination that produces best results in terms of true positive rate and that can be visualized in confusion matrix.

## 5. Results

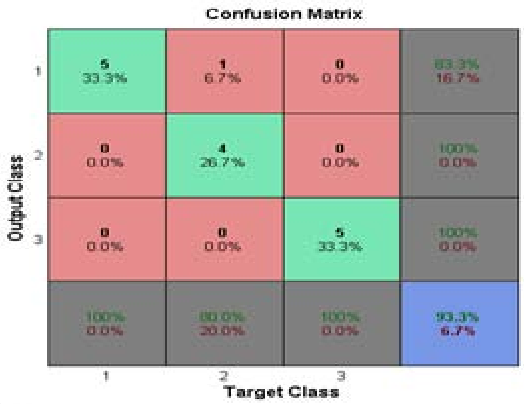


**Figure 5.1:** Confusion Matrix for predicting movie success

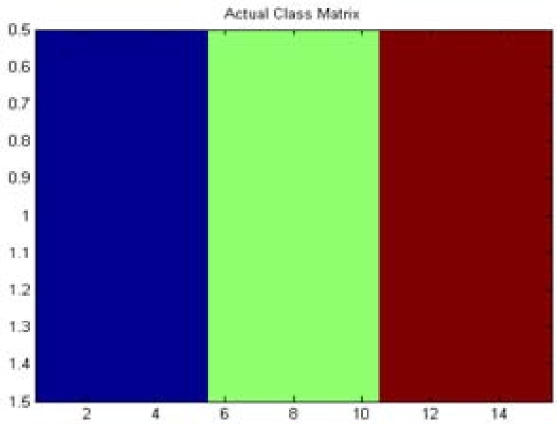


**Figure 5.2:** Actual Class matrix (Actual movie success)

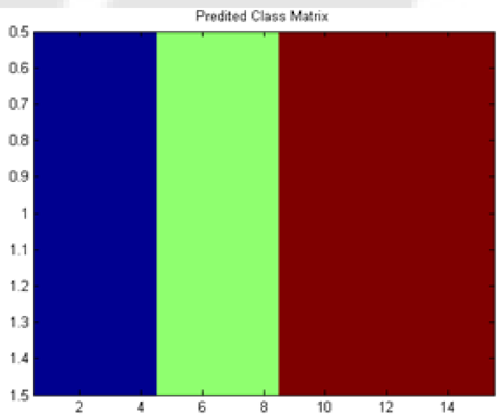


**Figure 5.3:** Predicted Class Matrix (Movie success as predicted by the designed algorithm)

### 5.1 Interpretation of Results

The strength of our algorithm is that it identifies predictors of movie success as well as their quantities so that the round truth is properly matched with the dedicated results, once this prediction framework work is put into practice. Therefore after designing multiple classifiers with various possible parameters of input observations, hidden layers and fixed number of output classes. We have tried to build a low computational resource intensive as well as less time consuming framework to predict movie success and it is apparent from the confusion matrix the accuracy is quiet high (93.3%).

The selection of parameters for the design of classifier has been meticulously and empirically found after many experiments. The appropriate selection of initial weights for the learning function was found by analysis of historical data. If initial weights are too small then net input to hidden or output unit will approach 0 which would have led to slow learning but if weight were too large the initial input signal to each hidden or output unit would fall in the saturation region where the derivative of the activation function (sigmoid) would have very small value 0.

The selection of learning rate was also done keeping in mind changes in weight factor must be small in order to reduce oscillations or any deviation. For deciding the training and testing patterns. We developed disjoint sets of training and testing datasets and got these validated using K4 cross validation method. In all the various deigns of classifiers the major focus was also to identify the no of hidden units care was taken that no unnecessary additional computational resource usage comes into play for each additional hidden layers and finally we can see from figure (5.2 & 5.3) then the size, length and color of actual and predicted class matrix are similar, due to high accuracy.

## 6. Conclusion

The data pertaining to parameters under study were collected from the leading Bollywood websites.

- Data Normalization must be used to reduce the number of samples and the complexity of the neural network and the computation time of the neural network.
- For the classification schemes, it was found that training the model with a large number of test data and with fast training algorithm would greatly enhance the accuracy and hence the reliability of the system.
- The design of our classifier was done by running the neural network with different number of hidden layers and it was apparent from the graphs that it affected the accuracy.
- It was found that as we increase the number of hidden layers there was also an increase in computation time but high order of accuracy is also achieved until we have reached the maximum of hidden layers, therefore, we need an optimal combination of parameters to achieve 93.3% accuracy

## 7. Future Scope

We can explore more unsupervised machine learning algorithms which would offer more versatile method of predicting movie success. These methods may be based on some computational clustering technique and which can be evaluated on the basis of recall and precision values.

- We can explore more algorithms and techniques for the feature extraction and classification of parameters influencing movie success to further improve the accuracy of the defect identification system.
- We can further improve the system by reducing the complexity. The main objective could be to find the best algorithms which optimize the performance and complexity this can be done by changing normalization of input data or by changing sample methods with other possible learning rate parameters etc.
- The accuracy of classifier can also be enhanced by using more and equal number of training patterns.

## 8. Acknowledgement

I am thankful to A P Nidhi, Assistant Professor, Swami Vivekanand Institute of Engineering and Technology, Banur, for providing constant guidance and encouragement for this research work.

## References

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

[2] 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.

[3] The Levenberg-Marquardt Algorithm , Ananth Ranganathan, 8th June 2004

[4] 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.

[5] K.Y. Chan, T.S.Dhillon, 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.

[6] 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.

[7] Arundeep Kaur and AP Gurpinder Kaur, Predicting Movie Success: Review of Existing Literature , International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, June 2013