Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2007 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2007

Using Neural Networks to Forecast Box Office Success

Mike Henry
Institute for Research in Information Systems

Ramesh Sharda
Institute for Research in Information Systems

Dursun Delen
Oklahoma State University

USING NEURAL NETWORKS TO FORECAST BOX OFFICE SUCCESS

Mike Henry, Ramesh Sharda, Dursun Delen[1]
Institute for Research in Information Systems (IRIS)
William S. Spears School of Business
Oklahoma State University
dursun.delen@okstate.edu

Abstract

Predicting box office receipts of a particular movie has intrigued many researchers, domain experts and industry leaders as a challenging problem. In this paper, we report on the current status of a prediction system being built at the Institute for Research in Information Systems (IRIS) at Oklahoma State University since 1998. In our model, the forecasting problem is converted into a classification problem, that is, rather than forecasting the pinpoint estimate of box office receipts, a movie is classified into one of nine financial success categories, ranging from a "flop" to a "blockbuster." The prediction results of different datasets representing different time windows and different combination of predictors are presented using average percent hit rate of bingo and one-away predictions. In the latest tests the prediction results of artificial neural networks improved to almost 50% on "bingo" and close to 90% on "one-away".

Keywords: Prediction, Classification, Forecasting, Motion Pictures, Box-office Receipts, Data Mining, Neural Networks, Performance Measures

INTRODUCTION

Predicting success of a movie at the box office is a major challenge for a high stakes decision. One of the greatest examples of how predicting a movie's success in this "land of hunch and guess" is evident in the movie "Pirates of the Caribbean". According to www.wikipedia.com, "when production for the film was first announced, movie fans and critics were skeptical of its chances of success". In fact, much of the skepticism was due to the fact that the movie was based on a popular Disney park theme ride. However, Pirates of the Caribbean became a monumental success. In fact, according to the-movie-times.com, it was so successful that it was the 3rd highest grossing film in 2006, the 22nd highest grossing film of all time, and had such staying power that the sequel, Pirates of the Caribbean: Dead Man's Chest, became the 6th highest grossing film of all time. The initial predictions that Pirates of the Caribbean would become a flop, proved to be incorrect. Thus the need to develop a more concrete, accurate forecasting technique is evident.

Despite the difficulty associated with the unpredictable nature of the problem domain, several researchers have attempted to develop models for forecasting the financial success of motion pictures, primarily using statistics-based forecasting

[1] Corresponding Author: dursun.delen@okstate.edu; (918) 594-8283 (voice); (918) 594-8281 (fax)

approaches. Most analysts have tried to predict the total box-office receipt of motion pictures after a movie's initial theatrical release. However, most (Litman 1983; Sawhney and Eliashberg 1996) did not get sufficiently accurate results for decision support. Litman and Ahn (1998) summarized and compared some of the major studies on predicting financial success of motion pictures. Yet, these previous studies leave us with an unsatisfied need for a more accurate forecasting method, especially prior to a movie's theatrical release. Most studies indicate that box-office receipts tend to tail-off after the opening week. Research shows that 25 percent of total revenue of a motion picture comes from the first two weeks of receipts (Litman Ahn 1998). Thus, once the first week of box-office receipts are determined, the total box-office receipts of a particular movie can be forecasted with very high accuracy (Sawhney and Eliashberg 1996). Therefore, the accurate estimate of the box office receipts of motion pictures before its theatrical release is a more difficult problem to the industry.

Sharda and Delen (2006) have been investigating the use of neural networks to take on this important problem. They converted the forecasting problem into a classification problem and used neural networks to place box-office receipts between nine categories, ranging from a 'flop' to a 'blockbuster'. Using the data from 1998 to 2002, they trained a neural network model that could predict the movie class within one category about 75% of the time. The objective of this paper is to present an update on this project by describing results of neural networks trained using recent data as well as additional variables.

EARLIER NEURAL NETWORK MODEL

As described by Sharda and Delen (2006), the neural network model for this problem is initially based upon the following parameters. Seven categories were used for input variables. The first variable is the rating system from the Motion Picture Association of America which includes the ratings G, PG, PG-13, R, and Not Rated values. The rating of a movie can be a strong contributing factor to a movie's success. In fact, there have been some indications that ratings of G and PG actually increase box office revenue because these films appeal to a larger audience.

The second variable is the competition at the time of a movie's release. Movies compete with each other throughout the year for the top box office spot. Based upon the analysis of the data, it was discovered that June and November were high competition months; May, July, and December were medium competition months, and the remaining months were recorded as low competition months.

The third variable is the star value that each movie possesses. A superstar actor or actress can be defined as someone who contributes significantly to the upfront sale of the movie regardless of the script, costars, or the director. The value of the star variable is derived from the researcher's knowledge of movies as well as by averaging the actor's past history of movie making success. The star power of an individual is placed into three categories, high, medium, and insignificant. An example of a high powered star would be Tom Hanks for his roles in movies such as Forrest Gump and Castaway, which were monumental successes. A medium powered star would be someone like Bruce Campbell, star of the Army of Darkness and the Evil Dead series, a very well known actor for his roles in cult classic movies, but may not have had the lead role in a super blockbuster movie. An insignificant star would be anyone who may be new to the industry or hasn't quite starred in any movies that have gained them a fair amount of recognition.

The fourth variable is the genre of each movie. Included are 10 independent genres, where each movie could be assigned to more than one genre at the same time. For example, in Austin Powers III: Gold Member, we classified the movie as being both a comedy and an action flick. The fifth variable is the amount of technical effects that a movie contains. Recorded are three levels of technical effects: high, medium, and low. Movies with high technical content, as well as animation films, are given a rating of high. Movies with moderate special effects were given a medium rating, and movies with little or no special effects were given a low technical effects rating. The decision of which rating a movie receives is based upon the researcher's knowledge of movies and their opinion of special effects.

The sixth variable is whether or not a movie is a sequel. We've included this variable because sequels are normally created based upon the financial success of the prequel movie. The seventh variable is the number of screens the movie was shown on during its initial release. It was found that there is a close correlation between a movies' financial success and the number of screens on which the movie was shown during its initial launch.

For dependent variables, the revenue that the movie generated was categorized into 9 different classes. These ranged from a class one, also known as a flop (those movies that made less than 1 million dollars) to a class 9, also known as a blockbuster, (those movies that made more than 200 million dollars).

The statistical software Statistica (www.Statsoft.com) was used to create the neural network models. Neural networks are known to be biologically inspired, highly sophisticated analytical techniques, capable of modeling extremely complex non-linear functions. For many years linear modeling has been the commonly used technique in capturing and representing functional relationships between dependent and independent variables, largely because of its well-known statistically explainable optimization strategies. In the problem scenarios where the linear approximation of a function was not valid (which was frequently the case), the models suffered accordingly. Now, such cases can easily be modeled with neural networks. Simply put, *neural networks* are analytic techniques modeled after the processes of learning in the cognitive system and the neurological functions of the brain, capable of predicting new observations (on specific variables) from other observations (on the same or other variables) after executing a process of so-called learning from existing data (Haykin 1998). Applications of neural networks have been reported in many diverse fields addressing problems in areas such as prediction, classification, and clustering. A classical reference for the fundamentals of neural networks is Rumelhart and McClelland (1986).

Multi layer perceptron (MLP) neural network architecture is known to be a strong function approximator for prediction and classification problems. It has been shown (Hornik et al. 1990) that given the right size and structure, MLP is capable of learning arbitrarily complex nonlinear functions to an arbitrary accuracy level. MLP is essentially the collection of nonlinear neurons (perceptrons) organized and connected to each other in a multi-layer structure.

Figure 1 graphically illustrates the neural network model employed in this study. Based on the results of an extensive experimental study, a two-hidden layered, multi-layered, feed-forwards neural network structure with sigmoid activation function is found to be the best neural network model for this dataset.

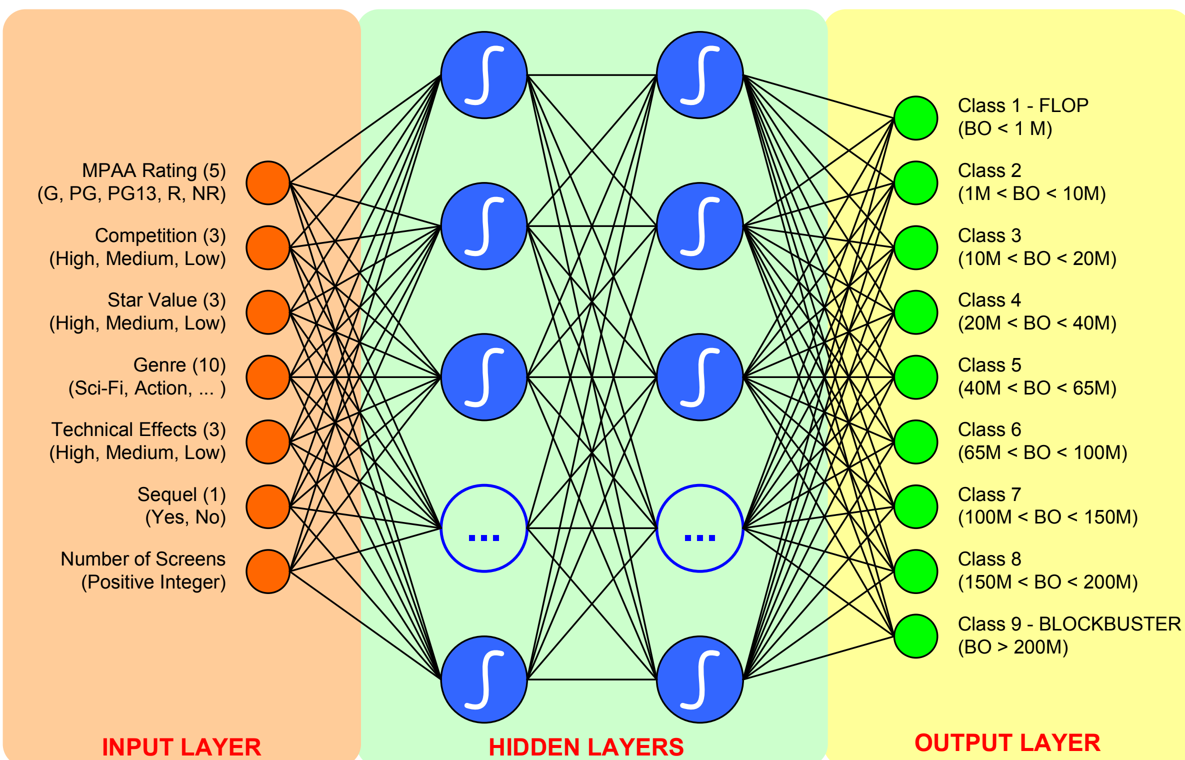

Figure 1. Graphical representation of our MLP neural network model

PERFRMANCE MEASURES

In classification problems, the average percent hit rate indicates the rate at which the testing data samples are classified into the correct classes. In our case, we have two different hit rates: the exact (bingo) hit rate (only counts the correct classifications to the exact same class) and the within 1 class (1-Away) hit rate. The hit rate measures the average accurate classification rate of the neural network prediction and the desired output. Algebraically, APHR can be formulated as follows:

$$APHR = Average\left(\frac{Number\ of\ samples\ correctly\ classified}{Total\ number\ of\ samples}\right)$$

$$APHR_{Bingo} = \frac{1}{g}\sum_{i=1}^{g}\frac{p_i}{n_i} \quad (1)$$

$$APHR_{1-Away} = \frac{1}{g}\sum_{i=1}^{g}\frac{p_{i-1}+p_i+p_{i+1}}{n_i} \quad (2)$$

where, g is the total number of classes, n_i is the total number of samples in class i, and p_i is the total number of samples correctly classified in class i.

RESULTS

For all tests conducted, a 10-fold cross validation was used to obtain the best results. In 10-fold cross-validation, the entire data set is divided into 10 mutually exclusive subsets (or folds) with approximately the same class distribution as the original data set (stratified). Each fold is used once to test the performance of the classifier that is generated from the combined data of the remaining nine folds, leading to 10 independent performance estimates (Sharda and Delen 2006).

Four experiments were conducted to improve the accuracy of the neural network model. The first experiment was to verify and retest the movie data from 1998 to 2002 that was originally collected. The second experiment was to collect and test new data from 2003 to 2005. The third experiment was to test data from all years combined. The forth experiment was to add new input variables to the original model configuration to see if better prediction results could be achieved. Through these experiments, the accuracy of predicting to within one category increased to 86%.

The prediction results are presented in two ways, the average of predicting movie success exactly and the average of predicting to within one class of the actual result. A confusion matrix is created from the results to display the number of movies that were classified correctly as a bingo hit rate and those that were within one class of the actual prediction as a 1-away hit rate.

Table 1 shows the confusion matrix for the 1998 to 2002 model created during the Sharda and Delen (2006) analysis. Highlighted in yellow is the number of movies that were predicted correctly. In that project, the bingo hit rate was 36.9% and the 1-away hit rate was 75.2%.

Table 1. Confusion matrix for the aggregated 10-fold neural network classification results

		Actual Categories									Avg.
		1	2	3	4	5	6	7	8	9	
Predicted Categories	1	37	35	5	4	0	0	0	1	2	
	2	33	37	13	14	0	1	0	1	1	
	3	5	13	28	21	1	4	8	7	4	
	4	15	3	16	38	0	2	3	4	9	
	5	0	0	6	13	55	30	7	3	2	
	6	0	1	2	3	31	26	19	13	4	
	7	0	0	8	5	5	12	24	21	10	
	8	0	0	5	2	3	7	24	20	16	
	9	0	0	9	1	2	7	8	22	43	
	BINGO	0.411	0.416	0.304	0.376	0.567	0.292	0.258	0.217	0.473	**0.369**
	1-Away	0.778	0.955	0.620	0.713	0.887	0.764	0.720	0.685	0.648	**0.752**

Source: Sharda and Delen (2006)

The first experiment was to verify and correct any data inconsistencies that were found within the original data collected in the Sharda and Delen (2006) research. Examples of the types of errors that were fixed were found in "The Bourne Identity" and "Spiderman". In "The Bourne Identity", the original data had the star value of insignificant. However, by the time "The Bourne Identity" was released, Matt Damon had starred in many popular movies such as "Ocean's Eleven", "Saving Private Ryan", and "Good Will Hunting".

As another example, for "Spiderman", the technical effects were recorded as medium quality whereas the graphics for that movie were extremely high. After editing the original data, a neural network model was retrained. For this experiment, the size of the original model consisted of 849 movies and the new model consisted of 1,380 movies. Table 2 shows the additional movies that were added to the model sorted by classification.

Table 2. Movies added to 1998 – 2002 neural network model

Class 1	Class 2	Class 3	Class 4	Class5	Class 6	Class 7	Class 8	Class 9
263	137	62	34	12	14	4	4	1
							Total	531

The new changes to the data created more accurate results. The results are presented as a column plot in Figure 2. The bingo hit rate improved from 36.9% to 56.06% and the 1-away rate improved from 75.20% to 88.61%.

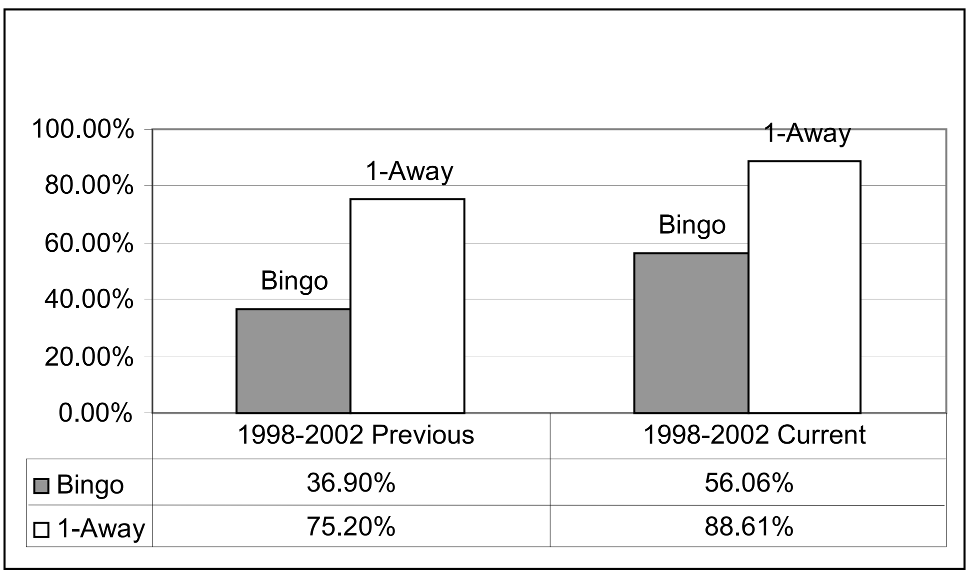

Figure 2. The hit rate statistics for 1998-2002 previous model and 1998-2002 current model

The second experiment was to collect data from 2003 to 2005 and to compare the test results with those that were originally collected in the Sharda and Delen research. For this experiment, the original 1998 to 2002 model contained 849 movies and the 2003 to 2005 model contained 920 movies.

According to Figure 3, the bingo hit rate improved from 36.9% to 63.07% and the 1-away rate improved from 75.20% to 89.43%. From these results, it is reasonable to conclude that the more data used for training allows for better prediction results.

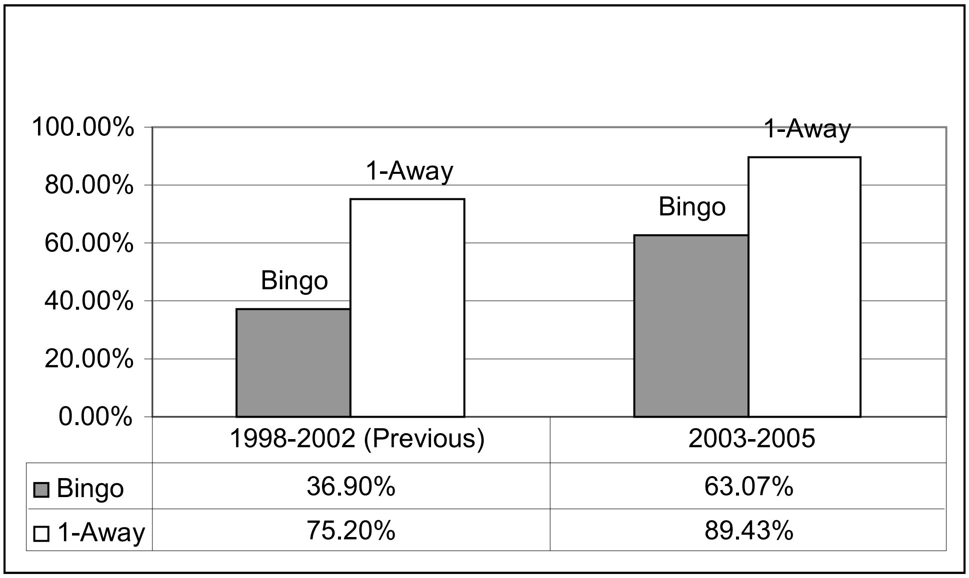

Figure 3. The hit rate statistics for 1998-2002 previous model and 2003-2005 model

In experiment three, a neural network model was created from 1998 to 2005. This model was then compared with the results from the current 1998 to 2002 model and the 2003 to 2005 model to test how more data will affect forecast accuracy. For this experiment, the current 1998 to 2002 model contains 1,380 movies, the 2003 to 2005 model contains 921 movies, and the 1998 to 2005 model contains 2,301 movies.

From Figure 4, it appears the most accurate model was the 2003 to 2005 model and the worst was the 1998 to 2005 model. However, the 2003 to 2005 model has less data that it has to train and test itself on, far less than the amount of data present in the 1998 to 2005 model.

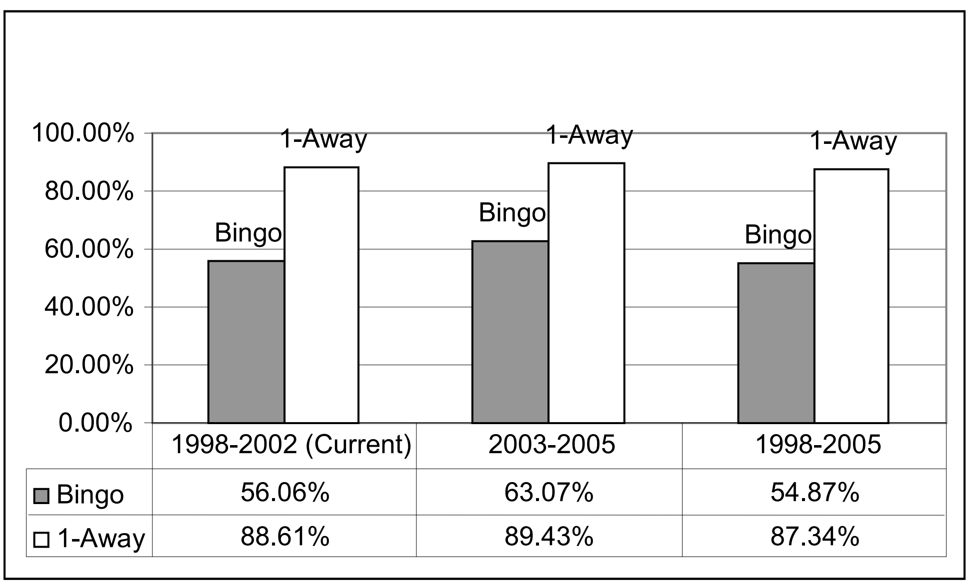

Figure 4. The hit rate statistics for 1998-2002 current model, 2003-2005 model, and 1998-2005 model.

In order to accurately compare performance, each model is tested on the same set of data. We performed a pure test of the models built thus far on 2006 data. This ensures that the models are tested outside the range of data that the models were originally trained on.

First, nine movies were chosen from 2006, one movie for each box office performance class. Table 3 displays the nine movies, their actual box office performance, and the prediction by each model. As seen in Table 3, the 1998 to 2005 model was the most accurate model. It was able to predict four movies correctly while the 2003 to 2005 model predicted three, and the 1998 to 2002 predicted two correctly.

Table 3. Summary of model performance on nine movies from 2006 in each class.

	Movie	Actual	1998-2002	2003-2005	1998-2005
1.	Pirates of Caribbean 2	**9**	7	9	7
2.	Ice Age: Meltdown	**8**	6	8	7
3.	The Break Up	**7**	6	6	7
4.	Inside Man	**6**	5	5	6
5.	Underworld 2	**5**	6	6	7
6.	The Grudge 2	**4**	5	4	7
7.	Ultraviolet	**3**	4	4	4
8.	BloodRayne	**2**	2	4	2
9.	Brick	**1**	1	2	1

Because the nine movies analyzed above do not provide any statistically valid evidence, each model was then tested on a much larger set of movies that were released in 2006. For this experiment, the 2006 dataset consisted of 347 movies.

In Figure 5, the 1998 to 2005 model had the best prediction results, while the 2003 to 2005 model had the worst prediction results. Similar to the findings in Table 3, it is reasonable to conclude that the more data that the model is trained on, the better it will do when forecasting movie success. Table 4 displays the predictions made by the 1998 to 2005 model.

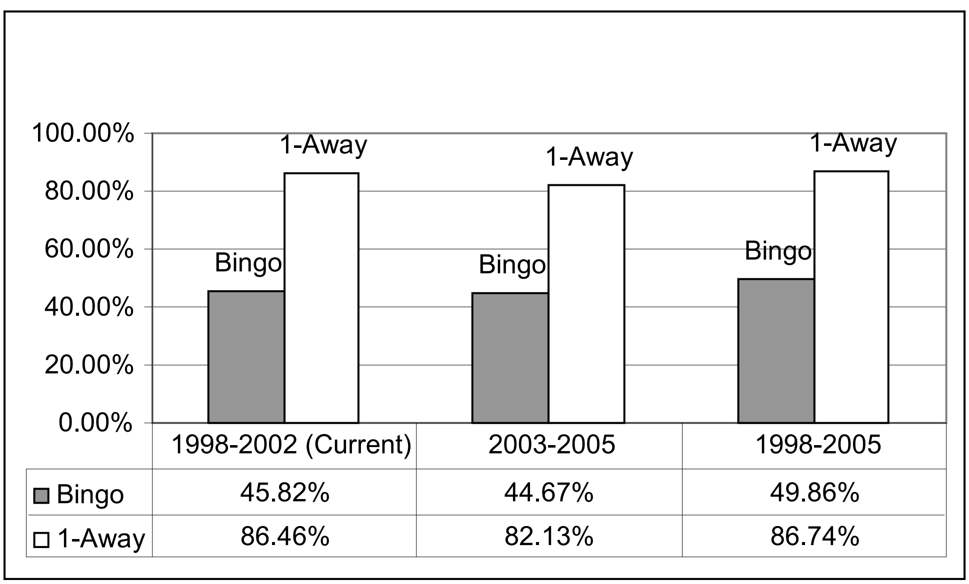

Figure 5. The hit rate statistics for 1998-2002 current model, 2003-2005 model, and 1998-2005 model.

Another test was conducted to see if new additional variables will increase the model's forecast accuracy. The dataset for this experiment consisted of 347 movie entries from 2006. In experiment four, eight input variables were added to the original model's configuration. These eight variables included:

- **Genre** – crime, family, romance, adventure, sports, foreign/subtitled films
- **Star value** – director
- **Star value** – international stars

Table 4. Confusion matrix for the 1998-2005 model on 2006 dataset

Actual Categories (columns); Predicted Categories (rows)

	1	2	3	4	5	6	7	8	9
1	94	25	2	2	0	0	0	0	0
2	20	31	9	4	0	0	0	0	0
3	0	14	19	10	4	0	0	0	0
4	0	5	9	19	21	4	3	0	0
5	0	0	0	1	0	1	0	0	0
6	0	0	0	1	2	5	0	0	0
7	0	0	1	5	9	11	5	5	6
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0

Bingo	49.86%
1-Away	86.74%

For this experiment, a neural network model was built using a subset of 2006 data and tested on another subset of 2006 data. In Figure 6, the 2006 model with new variables slightly improved the 1-away rate of the model; however, the original 2006 model without the new variables had a better bingo rate.

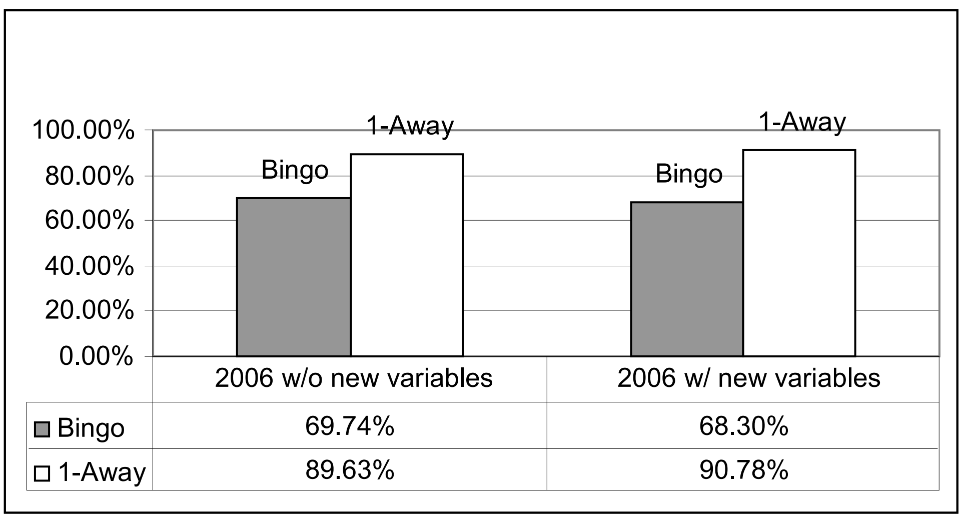

Figure 6. The hit rate statistics for 2006 model w/o new variables, 2006 model w/ new variables.

DISCUSSION AND CONCLUSION

The accuracy of the neural network model presented in this study can be improved by adding some of the other determinant variables such as production budget and advertising budget, which are known to be industry trade secrets and are not publicly released. Just like many other stochastic modeling techniques, neural networks start from a random set of weights. By utilizing the architectural parameters such as learning algorithm, learning rate, number of PEs in the hidden layers, etc., it adjusts those weights to create a map between the input and the output vectors. Correct choice of those architectural parameters plays a great role in developing better neural network models. There is no close form solution to what those architectural parameters should be for a given data set of a given problem domain. Modelers use their experiences, hunches, and rules of thumb, along with trial and error procedures to better configure those architectural parameters.

Lately, researchers have been developing hybrid architectures in which they apply genetic algorithms and other intelligent search techniques to optimize the architectural parameters of neural networks. Reported results are promising. Application of such hybrid architecture can improve the results we have obtained in this study. In addition to MLP, some other neural network architectures can be used to improve the accuracy of the neural network models. Future tests will determine which variables improve the model's accuracy.

Another method to improve the predictive accuracy of a system is through Meta modeling (combining multiple classifiers into a single predictive model). Experience has shown that combining the predictions from multiple methods often yields more accurate predictions than can be obtained with any one method alone. In our case one could use other models (e.g., decision trees, multiple logistic regression, discriminant analysis, etc.) along with neural networks combined through a Meta model to generate better predictions.

In conclusion, neural networks can handle complex forecasting problems in difficult business situations. Re-evaluating the data to ensure consistency and accuracy improves the model's prediction accuracy. A model trained on more data will yield better prediction results. Finally, adding new variables to the original model configuration only slightly improves the prediction results.

ACKNOWLEDGEMENTS

Support from National Science Foundation (NSF) and Oklahoma Louis Stokes Alliance for Minority Participation (OKAMP) to Michael Henry is greatly appreciated. The movie forecasting project began with Edith Bodnar (then Edith Meany) working on an exploratory effort to apply neural networks for this problem while she was a student in Professor Ramesh Sharda's class on information systems at the University of Wisconsin-Madison. Many students have assisted with data collection and modeling. They are recognized in Sharda and Delen (2006).

REFERENCES

Delen, D., Sharda, R. and Kumar, P. "Movie Forecast Guru: A Web-based DSS for Hollywood Managers", to appear in Decision Support Systems, 2006.

Hornik, K., Stinchcombe, M. and White, H. "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward network" Neural Networks, 3, 1990, pp. 359-366.

Hykin, S.S. Neural Networks: A Comprehensive Foundation. New Jersey: Prentice Hall, 1998.

Litman B.R. and Ahn, H. "Predicting Financial Success of Motion Pictures" appeared in The Motion Picture Mega-Industry by B.R. Litman. Allyn & Bacon Publishing, Inc.: Boston, MA., 1998.

Litman, B.R. "Predicting Success of Theatrical Movies: An Empirical Study," *Journal of Popular Culture*, Vol. 16, No. 9, 1983, pp. 159-175.

Rumelhart, D.E., and McCelland, J.L. Parallel Distributed Processing, Vol. 1, MIT Press, 1986.

Sawhney, M.S. and Eliashberg, J. "A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures," Marketing Science, 15(2), 1996, pp. 113-131.

Sharda, R. and Delen, D. "Predicting box-office success of motion pictures with neural networks" Expert Systems with Applications, 30, 2006, 243-254.