# Predicting box-office success of motion pictures with neural networks

**ARTICLE** *in* EXPERT SYSTEMS WITH APPLICATIONS · FEBRUARY 2006

Impact Factor: 2.24 · DOI: 10.1016/j.eswa.2005.07.018 · Source: DBLP



**2 AUTHORS:**


Ramesh Sharda
Oklahoma State University - Stillwater



Dursun Delen
Oklahoma State University - Stillwater

ELSEVIER

Expert Systems with Applications 30 (2006) 243–254

Expert Systems with Applications

www.elsevier.com/locate/eswa

# Predicting box-office success of motion pictures with neural networks

Ramesh Sharda*, Dursun Delen

*Department of Management Science and Information Systems, William S. Spears School of Business, Oklahoma State University, Stillwater, OK 74078, USA*

**Abstract**

Predicting box-office receipts of a particular motion picture has intrigued many scholars and industry leaders as a difficult and challenging problem. In this study, the use of neural networks in predicting the financial performance of a movie at the box-office before its theatrical release is explored. In our model, the forecasting problem is converted into a classification problem-rather than forecasting the point estimate of box-office receipts, a movie based on its box-office receipts in one of nine categories is classified, ranging from a 'flop' to a 'blockbuster.' Because our model is designed to predict the expected revenue range of a movie before its theatrical release, it can be used as a powerful decision aid by studios, distributors, and exhibitors. Our prediction results is presented using two performance measures: average percent success rate of classifying a movie's success exactly, or within one class of its actual performance. Comparison of our neural network to models proposed in the recent literature as well as other statistical techniques using a 10-fold cross validation methodology shows that the neural networks do a much better job of predicting in this setting.


*Keywords:* Forecasting; Prediction; Motion pictures; Box-office receipts; Neural networks; Logistic regression; CART; Sensitivity analysis

## 1. Introduction

Forecasting box-office receipts of a particular motion picture has intrigued many scholars and industry leaders as a difficult and challenging problem. To some analysts, 'Hollywood is the land of hunch and the wild guess' (Litman & Ahn, 1998) due largely to the difficulty and uncertainty associated with predicting the product demand. Such unpredictability of the product demand makes the movie business one of the riskiest endeavors for investors to take in today's competitive world. In support of such observations, Jack Valenti, president and CEO of the Motion Picture Association of America, once mentioned that '… No one can tell you how a movie is going to do in the marketplace… not until the film opens in darkened theatre and sparks fly up between the screen and the audience' (Valenti, 1978). Trade journals and magazines of the motion picture industry have been full of examples, statements, and experiences that support such a claim.

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 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 & Eliashberg, 1996) did not get sufficiently accurate results for decision support. Litman and Ahn (1998) summarize and compare some of the major studies on predicting financial success of motion pictures. Most studies indicate that box-office receipts tend to tail-off after the opening week. Research shows that 25% of total revenue of a motion picture comes from the first 2 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 & Eliashberg, 1996). Therefore, the accurate estimate of the box-office receipts of motion pictures especially before its theatrical release is a more difficult problem for the industry.

In our study, we explore the use of neural networks in forecasting the financial performance of a movie at the box-office before its theatrical release. Here, we convert the forecasting problem into a classification problem, i.e. rather

* Corresponding author.
*E-mail address:* sharda@okstate.edu (R. Sharda).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.07.018

than forecasting the point estimate of box-office receipts, we classify a movie based on its box-office receipts in one of nine categories, ranging from ‘flop’ to ‘blockbuster’.

Neural networks (NN) are known to be biologically inspired 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. Applications of neural networks have been reported in many diverse fields addressing problems in areas such as prediction, classification, and clustering. Many application bibliographies exist (Sharda, 1994). However, none of these include an application in forecasting box-office success of theatrical movies. This study is one of the first to attempt the use of neural networks for addressing this challenging problem that has drawn the attention of many researchers in such areas of decision support systems and management science.

The remainder of this paper is organized as follows. Section 2 briefly reviews the literature on forecasting the box-office success of theatrical movies. Section 3 gives the details of our methodology by specifically talking about the data, the neural network model, the experiment methodology and the performance measures used in this study. Next, the experimental results along with a comparative study of the neural network model to those of other statistical models are shown and explained. Finally, the Section 5 of the paper discusses the overall contribution of this study, along with its limitations and further research directions.

## 2. Literature review

Literature on forecasting financial success of new motion pictures can be classified based on the type of forecasting model employed: (i) econometric/quantitative models-those that explore factors that influence the box-office receipts of newly released movies (Litman, 1983; Litman & Kohl, 1989; Litman & Ahn, 1998; Neelamegham & Chintagunta, 1999; Ravid, 1999; Elberse & Eliashberg, 2002; Sochay, 1994) and (ii) behavioral models-those that primarily focus on the individual’s decision-making process with respect to selecting a specific movie from a vast array of entertainment alternatives (De Silva, 1998; Eliashberg & Sawhney, 1994; Eliashberg et al., 2000; Sawhney & Eliashberg, 1996; Zufryden, 1996). These behavioral models usually employ a hierarchical framework, where behavioral traits of consumers are combined (mostly in a sequential process) with the econometric factors in developing the forecasting models. Another classification is based on the timing of the forecast: (i) before the initial release-i.e. forecasting the financial success of the movies before their initial theatrical release (De Silva, 1998; Eliashberg et al., 2000; Litman, 1983; Litman & Kohl, 1989; Sochay, 1994; Zufryden, 1996), and (ii) after the initial release-i.e. forecasting the financial success of the movies after their initial theatrical release, when the first week of receipts are known (Neelamegham & Chintagunta, 1999; Ravid, 1999; Sawhney & Eliashberg, 1996). Forecasting models that fall into the category of ‘after the initial release’ tend to generate more accurate forecasting results due to the fact that those models have more explanatory variables including box-office receipts from the first week of viewership, movie critic reviews, and word-of-mouth effects. Our study falls into the category of *quantitative models* for model type classification, and into the category of *before the initial release* in timing of the forecast classification.

We differentiate our study from the others as follows. First, there is no reported study on using neural networks to predict box-office receipts of new motion pictures. Our study seems to be the first attempt of its kind in this problem domain. Another distinguishing feature of our study comes from its longitudinal nature. Our study is based on five consecutive years of data that covers movies released between 1998 and 2002. Our study also compares the difference between those of individual years and the combined data set of all 5 years. Finally, we have provided comparisons of our model with the models proposed in the literature as well as other statistical techniques. The comparison results suggest that our neural networks model performs better than the one ones reported in the literature.

## 3. Method

### 3.1. Neural networks and statistical modeling

Though commonly known as black box approach or heuristic method, in the last decade, artificial neural networks have been studied by statisticians in order to understand their prediction power from a statistical perspective (Cheng & Titterington, 1994; White, 1989a, b, 1990). These studies indicate that there are a large number of theoretical commonalities between the traditional statistical methods, such as discriminant analysis, logistic regression, and multiple linear regression, and their counterparts in neural networks, such as multi-layered perceptron, recurrent networks, and associative memory networks.

Multi layer perceptron (MLP) neural network architecture is known to be a strong function approximator for prediction and classification problems. It has been shown that given the right size and structure, MLP is capable of learning arbitrarily complex non-linear functions to an

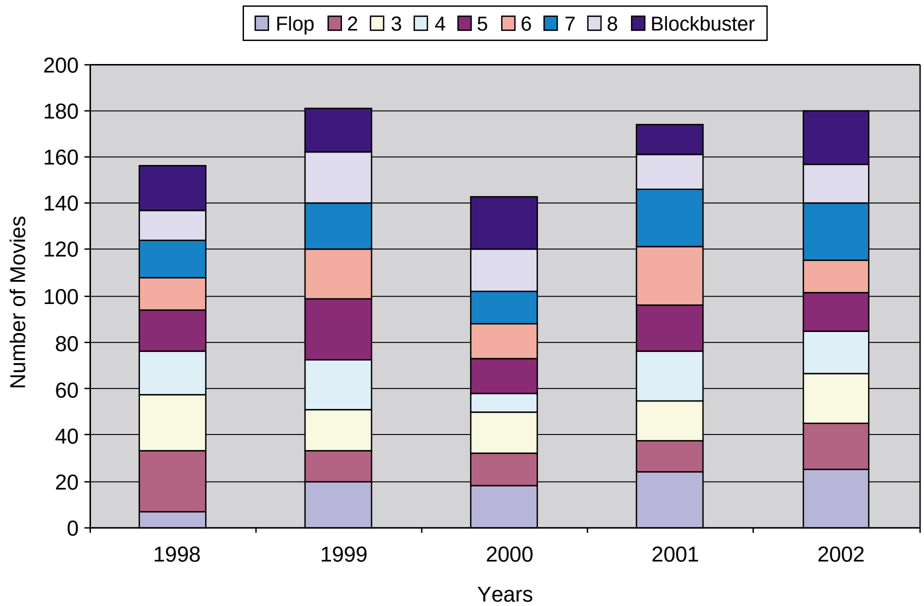

Fig. 1. Summary statistics of the complete dataset

arbitrary accuracy level (Hornik, Stinchcombe, & White, 1990). Thus, it is a likely candidate for exploring the rather difficult problem of mapping movie performance to the underlying characteristics.

### 3.2. Data and variable definitions

The initial launch of our study was in 1998. For each year since 1998, the available data have been collected, organized, and analyzed on a year-by-year basis. An early performance was reported in Sharda, Amato, and Meany (2000). Now, enough analysis results exist to report on this project. In our study, 834 movies released were used between 1998 and 2002. The sample data was drawn (purchased) from ShowBiz Data, Inc. (ShowBiz Data, 2002). The summary statistics of the complete dataset is presented in Fig. 1 in the form of a stacked bar chart. Good representations of each of the nine classes are present in each of the 5 years.

The variable of interest in our study is box-office gross revenues. It does not include auxiliary revenues such as video rentals, international market revenues, toy and soundtrack sales, etc. Another important difference between our study and previous efforts is that the forecasting problem is converted into a classification problem. As mentioned earlier, a movie based on its box-office receipts is classified in one of nine categories, ranging from a 'flop' to a 'blockbuster.' This process of converting a continuous variable in a limited number of classes is commonly called 'discretization' in neural network literature. Discrete values are a limited number of intervals in a continuous spectrum, whereas continuous values can be infinitely many. Many prefer using discrete values as opposed to continuous ones in developing prediction models because (i) discrete values are closer to a knowledge-level representation (Simon, 1981); (ii) data can be reduced and simplified through discretization (Liu et al., 2002), (iii) for both users and experts, discrete features are easier to understand, use, and explain (Liu et al., 2002); and finally, (iv) discretization makes many learning algorithms faster and more accurate (Dougherty, Kohavi, & Sahami, 1995).

In our study, the dependent variable was discretized into nine classes using the following breakpoints:

| Class No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Range (in Millions) | $<1$ (Flop) | $>1$ $<10$ | $>10$ $<20$ | $>20$ $<40$ | $>40$ $<65$ | $>65$ $<100$ | $>100$ $<150$ | $>150$ $<200$ | $>200$ (Blockbuster) |

Seven different types of independent variables were used. Our choice of independent variables is based on the feedback that have been received from the industry experts and previous studies conducted in this field. Each categorical-independent variable (except the *Genre* variable) is converted into a 1-of-*N* binary representation, which created a number of pseudo representations that increased the independent variable count from 7 to 26. A neural network treats these pseudo variables as different mutually exclusive information channels. In the process of value assignment, all pseudo representations of a categorical variable are given the value of 0, except the one that holds true for the current case, which is given the value of 1. For instance, in 1-of-*N* binary representation, the Star Value variable would be represented with three pseudo variables denoting *high*, *medium* and *low* and for instance for the movie 'Terminator' because of the superstar Arnold Schwarzenegger, the value of *high* would be set to 1 and the values of medium and low would both be set to 0.

We used most of the variables mentioned in previous studies (Elberse & Eliashberg, 2002; Litman, 1983; Litman & Kohl, 1989; Neelamegham & Chintagunta, 1999; Ravid, 1999; Sawhney & Eliashberg, 1996; Sochay, 1994). Following is a short description of these variables with their respective representation schemas.

#### *3.2.1. MPAA rating*

A commonly used variable in predicting the financial success of a movie is the rating assigned by the Motion Picture Association of America (MPAA). There are five possible rating categories, each represented with a binary variable: G, PG, PG-13, R, and NR. These ratings help to assess the degree of sexual content, violence, and adult language for any movie before its theatrical release. Though earlier studies that used regression-based forecasting models showed no significant contribution of MPAA rating to predict box-office receipts (Litman, 1983; Litman & Kohl, 1989; Sochay, 1994), others indicated a high level of significance (Ravid, 1999). Litman and Ahn (1998) and Ravid (1999) indicated that some ratings (e.g. G and PG) led to increased box-office revenues than others when all other influencing factors are kept equal, largely due to the fact that these films appeal to a larger potential crowd. For the sake of completeness, in this study we decided to include MPAA ratings by using five binary variables to represent the MPAA ratings.

#### *3.2.2. Competition*

Each movie competes for the same pool of entertainment dollars against movies released at the same time or movies released some time ago that are still playing in theatres. Many studies have found the time of release, as an indication of the level of competition, to be a significant contributor to a movie's box-office receipts (Krider & Weinberg, 1998; Litman & Kohl, 1989; Radas & Shugan, 1998; Sochay, 1994), i.e. the financial success of a movie is expected to be highly correlated with the time-dependent competitive forces in the marketplace. These forces are expected to negatively influence the success of a movie. Ironically though, most blockbuster movies were released during the high competition season. In our study, we represent the competition using three binary pseudo variables (High Competition, Medium Competition, and Low Competition). Based on our profile investigation of 834 movies, we assigned 'High Competition' to the movies that are released in the months of June and November; 'Medium Competition' to the movies released in the months of May, July and December, and 'Low Competition' to the movies released in the rest of the months.

#### *3.2.3. Star value*

As others before us, we included another independent variable to measure the financial success of a movie depending upon the presence or absence of any box office superstars in the cast (e.g. actors, actresses, and directors). Ravid (1999) found conflicting results with respect to the contribution of superstars to the financial success of a movie. A superstar actor/actress can be defined as one who contributes significantly to the up-front sale of the movie regardless of the script, the co-stars, and the director. The value of a star/superstar is determined by averaging his/her recent past history of movie-making prices. We used three independent binary variables to represent the degree of star value of actors/actresses in our model: A+/A (high) star value, B (medium) star value, and C (insignificant) star value. We did not use a variable to signify the importance (or star value) of the director due to the fact that the most earlier studies that included star value of the movie director did not find it to be a significant contributor (Litman, 1983; Litman & Kohl, 1989; Sochay, 1994). In fact recent studies do not even include it into their prediction models (Neelamegham & Chintagunta, 1999; Sawhney & Eliashberg, 1996).

#### *3.2.4. Content category (genre)*

One commonly used, yet rarely found to be a significant contributor, is the content category variable (Litman, 1983; Litman & Kohl, 1989; Sochay, 1994). In our study, we followed the convention and included content category as one of the independent variables. Specifically, we placed each film into as many of the content categories as the movie can be classified in, using 10 binary-independent variables, i.e. a movie can be assigned to more than one content category at the same time. For instance, 'Austin Powers III: Gold member', a blockbuster of 2002, is classified as both a 'comedy' and an 'action flick'. The categories included are Sci-Fi, Historic Epic Drama, Modern Drama, Politically Related, Thriller, Horror, Comedy, Cartoon, Action, and Documentary.

#### *3.2.5. Technical effects*

We also used a variable to capture the technical merit (special technical effects) of a movie. Following the classification on the ShowBiz dataset, we used three levels of technical effects categories, which are represented by three binary-independent variables. Movies with high technical content and special effects, such as animations and science fiction movies, are given the highest rating. Movies with moderate special effects are given a medium rating, and movies with little or no special effects are given the low technical effects ratings.

#### *3.2.6. Sequel*

Similar to other studies (De Silva, 1998; Ravid, 1999), we also included a binary variable to specify whether a movie is a sequel (value of 1) or not (value of 0). Intuitively,

Table 1
Summary of independent variables

| Independent variable name | No. of values | Possible values |
|---|---|---|
| MPAA Rating | 5 | G, PG, PG-13, R, NR |
| Competition | 3 | High, medium, low |
| Star value | 3 | High, medium, low |
| Genre | 10 | Sci-Fi, historic epic drama, modern drama, politically related, thriller, horror, comedy, cartoon, action, documentary |
| Special effects | 3 | High, medium, low |
| Sequel | 1 | Yes, no |
| Number of screens | 1 | Positive integer |

one would expect sequels to be positively correlated with financial success of a movie since they build on the success of the previous versions. So far, the empirical evidences as they were published in the related literature fall short on collectively supporting such a claim.

*3.2.7. Number of screens*

Previous research efforts showed close correlations between a movies' financial success and the number of screens on which the movie was shown during its initial launch (Jones & Ritz, 1991; Neelamegham & Chintagunta, 1999). In our model, we represented the number of screens a movie is scheduled to be shown at its opening with a continuous variable. Since neural networks can handle a mix of continuous and discrete values in both input and output layers, we chose to use this variable as the only continuous variable in the model. Our experimental runs indicated that the continuous values of this variable for NN model generated better prediction results than did the discretized values.

A summary of the above-mentioned and briefly defined decision variables is given in Table 1. In total, 26 decision (representing seven categories) and 9 output variables are used in this study.

The neural network models were developed using a commercial software product called NeuroSolutions (NeuroDimension, 2004). Fig. 2 depicts our Multi Layer Perceptron (MLP) neural network model with two hidden layers. The input layer is condensed to show only the seven aggregated independent variables (PEs), as opposed to all 26, in order not to unnecessarily clutter the picture.

We used a MLP neural network architecture with two hidden layers, and assigned 18 and 16 processing elements (PE) to them, respectively. Our preliminary experiments showed that for this problem domain, two hidden layered MLP architectures consistently generated better prediction results than single hidden layered architectures. In both hidden layers, sigmoid transfer functions were utilized. These parameters were selected on the basis of trial runs of many different neural network configurations and training parameters.

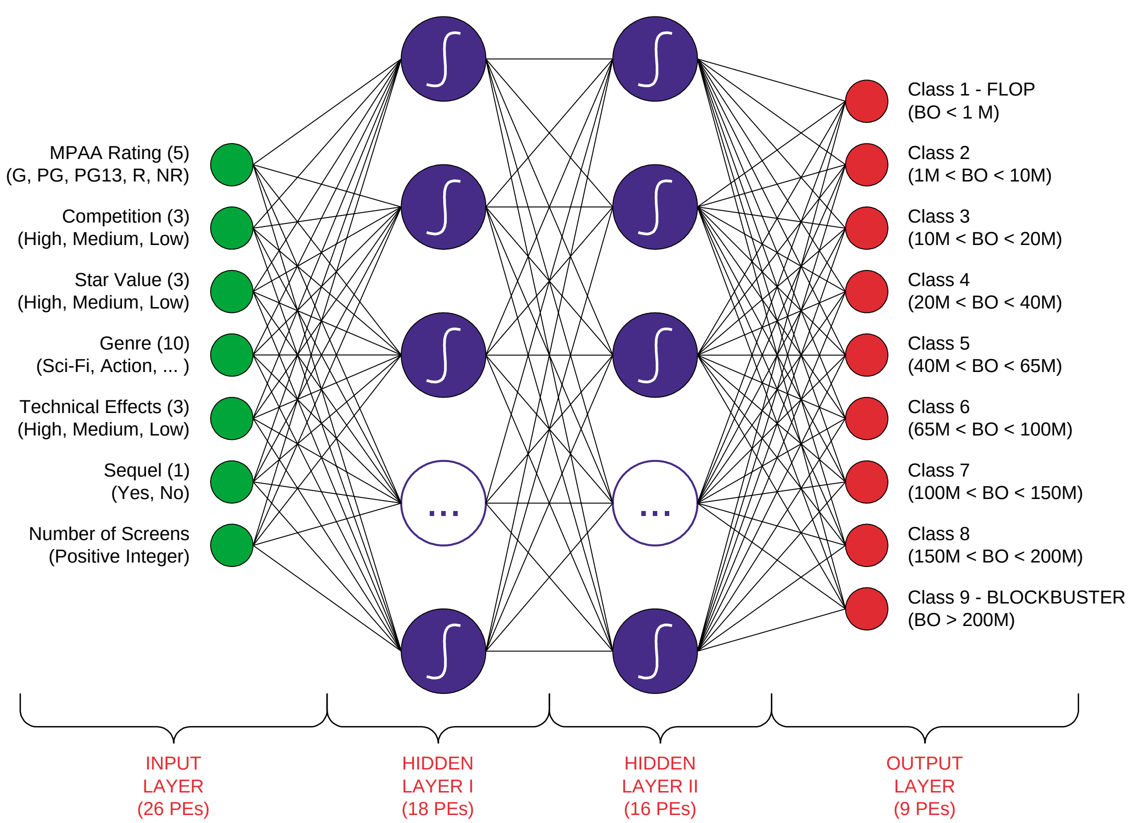


Fig. 2. Graphical representation of our MLP neural network model.

### 3.3. Experiment methodology

Many researchers in the past few years have studied the performance of neural networks in predicting a variety of classification problems over a wide range of different business settings. Many of these studies, however, were based on a single experiment, and/or the method of selecting the training and testing samples was not clear. We believe that because of the stochastic nature of the neural network training, better experimental design methods are necessary to develop the objective performance measures of neural networks. As opposed to using a single neural network experiment to base our results upon, we chose to follow a more statistically sound experimental design methodology, called *k*-fold cross-validation. In *k*-fold cross-validation, also called rotation estimation, the complete dataset ($D$) is randomly split into $k$ mutually exclusive subsets (the folds: $D_1, D_2, \ldots, D_k$) of approximately equal size. The classification model is trained and tested $k$ times. Each time ($t \in \{1,2,\ldots, k\}$), it is trained on all but one folds ($D \backslash D_t$) and tested on the remaining single fold ($D_t$). The cross-validation estimate of the overall accuracy is calculated as simply the average of the $k$ individual accuracy measures.

Since the cross-validation accuracy would depend on the random assignment of the individual cases into $k$ distinct folds, a common practice is to stratify the folds themselves. In stratified *k*-fold cross-validation, the folds are created in a way that they contain approximately the same proportion of predictor labels as the original dataset. Empirical studies showed that stratified cross validation tend to generate comparison results with lower bias and lower variance when compared to regular *k*-fold cross-validation (Kohavi, 1995).

In this study, to estimate the performance of neural network classifier a stratified 10-fold cross-validation approach is used. Empirical studies showed that 10 seems to be a 'magic' number of folds (that optimizes the tradeoff between time it takes to complete the experiment and minimizing the bias and variance associated with the validation process) (Breiman, Friedman, Olshen, & Stone, 1984; Kohavi, 1995). 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.

### 3.4. Performance metrics

We used percent success rate to measure the predictive performance of our neural network approach. The percent success rate (also known as average percent hit rate (APHR)) is arguably the most intuitive measure of discrimination for predictive accuracy of classification problems. It is the ratio of total correct classifications to total number of samples, averaged for all classes in the classification problem. Intuitively, the bigger values of APHR should indicate better classification performance. One should, however, judge the magnitude of this percentage in relation to the expected percentage of correct classification if the assignment was made randomly. Therefore, the prediction accuracy should be judged with respect to the number of classes presented in the classification problem. A relatively smaller value of APHR may indicate a reasonably good performance when the number of classes is large, and vice versa.

In classification problems, the APHR 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 in Eqs. (1)–(3)

$$\mathrm{APHR} = \frac{\text{Number of samples correctly classified}}{\text{Total number of samples}} \tag{1}$$

$$\mathrm{APHR}_{\mathrm{Bingo}} = \frac{1}{n}\sum_{i=1}^{g} p_i \tag{2}$$

$$\mathrm{APHR}_{1\text{-Away}} = \frac{1}{n}\left((p_1 + p_2) + \sum_{i=2}^{g-1} p_{i-1} + p_i + p_{i+1} + (p_{g-1} + p_g)\right) \tag{3}$$

where $g$ is the total number of classes ($=9$), $n$ is the total number of samples ($=834$), and $p_i$ is the total number of samples classified as class $i$.

## 4. Results

### 4.1. Neural network performance

Recall that our neural network model aims to categorize a film in one of the nine categories. Accuracy of the model is measured using two metrics. The first metric is the percent correct classification rate, which we have dubbed 'bingo'. This metric is a reasonably conservative measure for accuracy in this scenario. In reality, a movie producer might be glad to predict within one (or maybe even two categories) on either side. We also report the 1-away correct classification rate. Table 2 presents our aggregated 10-fold cross-validation neural network results in a confusion matrix. A confusion matrix is a commonly used tabular representation of classification results. The columns in a confusion matrix represent the actual classes and the rows represent the predicted classes. The intersection cells of the same classes represent the correct classification of the samples for that class. For instance, the upper left corner cell

Table 2
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** |

denotes the number of samples classifies as *class-1* (i.e. 'flop') while in fact they belong to *class-1*, whereas the lower right corner cell denotes the number of samples classifies as *class-9* (i.e. 'blockbuster') while in fact they belong to *class-9*. In summary, the diagonal cells from upper left corner to lower right corner (shaded in Table 2) shows the correct classifications (bingo) while others represents the misclassifications. The immediate neighboring cells shows the misclassified instances by a single class, which were used in calculating 1-away classification accuracies. Table 2 also shows the summary statistics on the prediction accuracy (both bingo and 1-away) of each class separately as well as the overall prediction accuracy.

In this experiment, we combined the data for all 5 years, implicitly making the assumption that the predictors for financial success of the movies have not changed significantly between the years 1998 and 2002. In order to test the validity of this assumption, we developed classification models for each of the 5 years using the exact same experimental design as the one used for the combined dataset. The detailed results of these experiments are presented in Table 3.

Fig. 3 presents the summary statistics for all 5 years, both individually and combined, using standard bar charts. Notice that the mean, standard deviation, and the median calculated for each of the 5 years are not significantly different from the ones calculated for all 5 years. This suggests that the predictable behavior of the model does not change over time for the variables included in the model and for the time period used in this study.

### 4.2. Comparison to other models

We compared our model's performance with other popular classification methods used for similar problem scenarios. Specifically, we used traditional statistical classification methods, such as discriminant analysis and multiple logistic regression, and a popular decision tree method called classification and regression trees (CART). Short descriptions for each of these three classification methods follow.

*Multinomial logistic regression (logit)* is a generalization of linear regression. It is used primarily for predicting binary or multi-class dependent variables. Because the response variable is discrete, it cannot be modeled directly by linear regression. Therefore, rather than predicting the point estimate of the event itself, it builds the model to predict the odds of its occurrence. In a two-class problem, odds greater than 50% would mean that the case is assigned to the class designated as '1' and '0' otherwise. While logistic regression is a very powerful modeling tool, it assumes that the response variable is linear in the coefficients of the predictor variables. Furthermore, the modeler, based on his or her experience with the data and data analysis, must choose the right inputs and specify their functional relationship to the response variable (Coskunoglu et al., 1985).

Discriminant analysis is one of the oldest statistical classification techniques, first introduced by Fisher (1936). Using the historic data, it finds hyperplanes (e.g. lines in two dimensions, planes in three, etc.) that separate the classes from each other. The resultant model is very easy to interpret because it simply determines on which side of the line (or hyper-plane) a point falls. Training is simple and scalable. Despite its scalability and simplicity, discriminant analysis is not a popular technique in data mining for two main reasons: (1) it assumes that all of the predictor variables are normally distributed (i.e. their histograms look like bell-shaped curves), which may not be the case, and (2) the boundaries that separate the classes are all linear forms (such as lines or planes). However, sometimes the data just cannot be separated that way.

*Classification and regression trees (CART)*, first introduced by Breiman et al. (1984), is a popular classification tree induction technique in the data-mining arena. It is based on a computational-statistical algorithm, which is a special case of general recursive partitioning algorithms. As the name implies, this technique separates observations in binary and sequential branches of a tree for the purpose of improving prediction accuracy. In doing so, CART uses a

Table 3
Confusion matrixes for each of the 5 years representing aggregated 10-fold neural network classification results

| Year | | | Actual Categories | | | | | | | | | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| 2002 | Predicted Categories | 1 | 9 | 4 | 2 | 3 | 1 | 2 | 0 | 0 | 0 | |
| | | 2 | 8 | 6 | 4 | 5 | 0 | 1 | 1 | 1 | 0 | |
| | | 3 | 2 | 5 | 9 | 1 | 2 | 0 | 2 | 2 | 1 | |
| | | 4 | 2 | 1 | 1 | 3 | 3 | 0 | 1 | 1 | 1 | |
| | | 5 | 2 | 0 | 1 | 4 | 8 | 4 | 0 | 0 | 0 | |
| | | 6 | 1 | 1 | 1 | 0 | 2 | 5 | 4 | 0 | 0 | |
| | | 7 | 1 | 2 | 2 | 2 | 0 | 2 | 11 | 7 | 3 | |
| | | 8 | 0 | 0 | 1 | 1 | 0 | 0 | 4 | 2 | 5 | |
| | | 9 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 4 | 13 | |
| | | BINGO | 0.360 | 0.300 | 0.429 | 0.158 | 0.500 | 0.357 | 0.440 | 0.118 | 0.565 | **0.367** |
| | | 1-Away | 0.680 | 0.750 | 0.667 | 0.421 | 0.813 | 0.786 | 0.760 | 0.765 | 0.783 | **0.711** |
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| 2001 | Predicted Categories | 1 | 12 | 3 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | |
| | | 2 | 9 | 8 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | |
| | | 3 | 1 | 2 | 5 | 5 | 0 | 0 | 2 | 3 | 0 | |
| | | 4 | 0 | 0 | 5 | 3 | 5 | 3 | 2 | 3 | 1 | |
| | | 5 | 0 | 0 | 0 | 9 | 10 | 6 | 1 | 0 | 1 | |
| | | 6 | 0 | 0 | 0 | 2 | 3 | 8 | 3 | 1 | 1 | |
| | | 7 | 1 | 0 | 2 | 1 | 2 | 5 | 8 | 2 | 4 | |
| | | 8 | 0 | 0 | 1 | 0 | 0 | 2 | 6 | 5 | 2 | |
| | | 9 | 1 | 0 | 0 | 1 | 0 | 1 | 2 | 1 | 2 | |
| | | BINGO | 0.500 | 0.615 | 0.278 | 0.143 | 0.500 | 0.320 | 0.320 | 0.333 | 0.154 | **0.351** |
| | | 1-Away | 0.875 | 1.000 | 0.778 | 0.810 | 0.900 | 0.760 | 0.680 | 0.533 | 0.308 | **0.753** |
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| 2000 | Predicted Categories | 1 | 11 | 4 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | |
| | | 2 | 2 | 8 | 3 | 0 | 0 | 2 | 1 | 0 | 0 | |
| | | 3 | 0 | 1 | 4 | 3 | 0 | 1 | 2 | 1 | 2 | |
| | | 4 | 1 | 0 | 3 | 1 | 3 | 1 | 0 | 0 | 1 | |
| | | 5 | 1 | 0 | 1 | 2 | 7 | 5 | 1 | 1 | 0 | |
| | | 6 | 1 | 0 | 1 | 0 | 1 | 1 | 3 | 0 | 0 | |
| | | 7 | 1 | 1 | 2 | 0 | 2 | 5 | 3 | 0 | 0 | |
| | | 8 | 0 | 0 | 1 | 1 | 1 | 0 | 2 | 6 | 9 | |
| | | 9 | 1 | 0 | 2 | 1 | 0 | 0 | 0 | 10 | 11 | |
| | | BINGO | 0.611 | 0.571 | 0.222 | 0.125 | 0.467 | 0.067 | 0.214 | 0.333 | 0.478 | **0.364** |
| | | 1-Away | 0.722 | 0.929 | 0.556 | 0.750 | 0.733 | 0.733 | 0.571 | 0.889 | 0.870 | **0.755** |
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| 1999 | Predicted Categories | 1 | 9 | 1 | 0 | 1 | 2 | 2 | 0 | 0 | 0 | |
| | | 2 | 7 | 6 | 2 | 1 | 1 | 0 | 0 | 2 | 1 | |
| | | 3 | 0 | 3 | 9 | 2 | 2 | 0 | 2 | 0 | 0 | |
| | | 4 | 2 | 1 | 5 | 5 | 7 | 2 | 0 | 2 | 1 | |
| | | 5 | 0 | 1 | 0 | 3 | 7 | 5 | 1 | 0 | 1 | |
| | | 6 | 1 | 0 | 1 | 2 | 6 | 6 | 6 | 0 | 1 | |
| | | 7 | 0 | 0 | 1 | 3 | 2 | 5 | 11 | 2 | 0 | |
| | | 8 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 8 | 6 | |
| | | 9 | 0 | 1 | 0 | 3 | 0 | 1 | 0 | 8 | 9 | |
| | | BINGO | 0.450 | 0.462 | 0.500 | 0.238 | 0.259 | 0.286 | 0.550 | 0.364 | 0.474 | **0.387** |
| | | 1-Away | 0.800 | 0.769 | 0.889 | 0.476 | 0.741 | 0.762 | 0.850 | 0.818 | 0.789 | **0.762** |
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| 1998 | Predicted Categories | 1 | 5 | 3 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | |
| | | 2 | 1 | 16 | 3 | 2 | 0 | 1 | 0 | 0 | 0 | |
| | | 3 | 0 | 5 | 5 | 8 | 0 | 0 | 0 | 1 | 1 | |
| | | 4 | 1 | 2 | 9 | 3 | 5 | 1 | 1 | 0 | 2 | |
| | | 5 | 0 | 0 | 2 | 2 | 7 | 5 | 1 | 0 | 0 | |
| | | 6 | 0 | 0 | 1 | 0 | 3 | 3 | 2 | 1 | 2 | |
| | | 7 | 0 | 0 | 2 | 0 | 2 | 4 | 4 | 2 | 4 | |
| | | 8 | 0 | 0 | 1 | 1 | 1 | 0 | 4 | 7 | 5 | |
| | | 9 | 0 | 0 | 1 | 1 | 0 | 0 | 3 | 2 | 5 | |
| | | BINGO | 0.714 | 0.615 | 0.208 | 0.158 | 0.389 | 0.214 | 0.250 | 0.538 | 0.263 | **0.353** |
| | | 1-Away | 0.857 | 0.923 | 0.708 | 0.684 | 0.833 | 0.857 | 0.625 | 0.846 | 0.526 | **0.756** |

mathematical algorithm to identify a variable and corresponding threshold for the variable that splits the input observation into two subgroups: one comprising observations with variable values lower than the threshold, and the other comprising observations with variable values higher than the threshold. It does this at each leaf node until the complete tree is constructed. The objective of the splitting algorithm is to find a variable-threshold pair that maximizes the homogeneity (order) of the resulting two subgroups of samples. The most commonly used mathematical algorithm for splitting is called gini index (Breiman, 1996).

In the following section, we present the results of our method compared to the three classification methods described above. We used exactly the same training and testing data set generated using stratified 10-fold cross-validation for all three models and the neural network model. The aggregated results are shown in Table 4.

As the results indicate, on an average, neural network model generated significantly better classification accuracy than all the other methods. Out of the remaining three, CART generated the best results for both 'bingo' and '1-away' followed by Logistic Regression and finally Discriminant Analysis. Fig. 4 illustrates these results graphically.

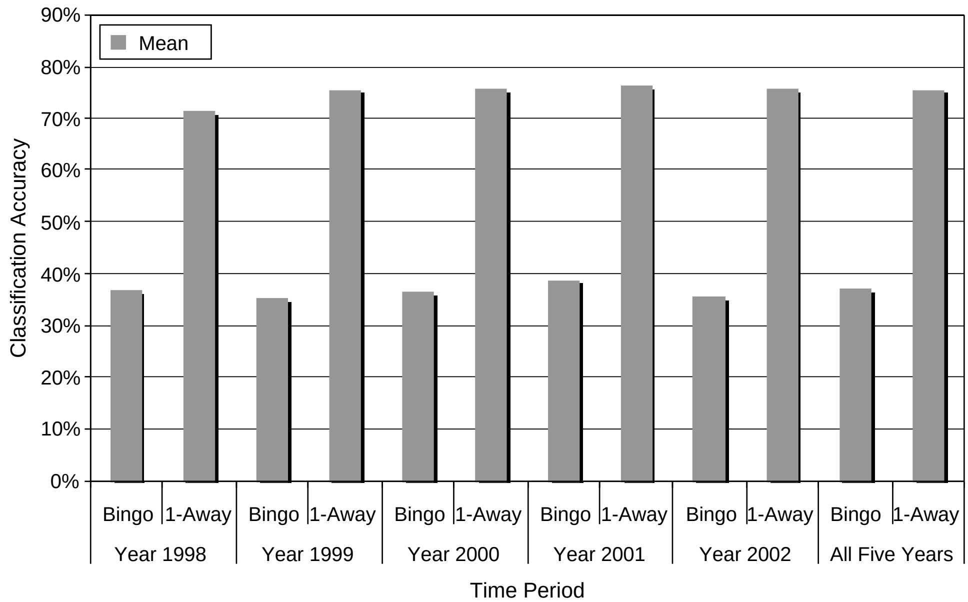


Fig. 3. Graphical representation of the hit rate summary statistics.

### 4.3. Sensitivity analysis on neural network output

As has been noted by statisticians and investigators on neural network applications in marketing, neural networks may offer better predictive ability, but not much explanatory value. This criticism is generally true, but sensitivity analysis can be performed to generate some insight into the problem.

Sensitivity analysis is a method for extracting the cause and effect relationship between the inputs and outputs of a neural network model. It is a commonly used method in neural network studies for identifying the degree at which each input channel (independent variables) contributes to the identification of each output channel (dependent variables). In the process of performing sensitivity analysis, the neural network learning is disabled so that the network weights are not affected. The basic idea is that the inputs to the network are perturbed slightly, and the corresponding change in the output is reported as a percentage change in the output (Principe, Euliano, & Lefebre, 2000). The first input varied between its mean plus (or minus) a user-defined number of standard deviations, while all other inputs are fixed at their respective means. The network output is computed and recorded as the percent change above and below the mean of that output channel. This process is repeated for each and every input variable. As an outcome of this process, a report (usually a column plot) is generated, which summarizes the variation of each output with respect to the variation in each input.

Our sensitivity analysis results are based on the combined data set of all 5 years. After the training, the network weights are frozen (testing stage) and the cause and effect relationship between the independent variables and

Table 4
Performance comparison of forecasting techniques: Average Percent Hit Rates

| | Logistic Regression | | Discriminant Analysis | | Classification and Regression Tree | | Neural Networks | |
|---|---|---|---|---|---|---|---|---|
| Folds | Bingo | 1-Away | Bingo | 1-Away | Bingo | 1-Away | Bingo | 1-Away |
| 1 | 31.02% | 67.38% | 30.77% | 67.06% | 31.04% | 71.35% | 37.25% | 75.23% |
| 2 | 30.91% | 69.81% | 29.30% | 68.54% | 32.16% | 70.06% | 35.61% | 73.19% |
| 3 | 28.87% | 70.62% | 28.38% | 67.88% | 29.26% | 73.13% | 33.91% | 76.50% |
| 4 | 31.07% | 68.05% | 29.75% | 66.31% | 32.31% | 69.99% | 35.39% | 74.79% |
| 5 | 29.56% | 69.40% | 28.53% | 68.21% | 30.74% | 70.79% | 39.51% | 76.14% |
| 6 | 29.11% | 67.13% | 28.27% | 66.86% | 29.96% | 69.96% | 36.63% | 73.28% |
| 7 | 32.43% | 71.27% | 32.01% | 69.30% | 34.06% | 72.92% | 36.94% | 76.38% |
| 8 | 30.01% | 71.95% | 29.98% | 68.78% | 30.37% | 71.97% | 38.19% | 75.09% |
| 9 | 28.94% | 70.86% | 27.67% | 67.67% | 30.81% | 72.29% | 36.82% | 77.01% |
| 10 | 29.77% | 69.52% | 27.83% | 66.28% | 31.13% | 68.27% | 38.75% | 74.34% |
| **Mean** | **30.17%** | **69.60%** | **29.25%** | **67.69%** | **31.18%** | **71.07%** | **36.90%** | **75.20%** |
| **St. Dev.** | **1.16%** | **1.65%** | **1.39%** | **1.04%** | **1.36%** | **1.54%** | **1.67%** | **1.33%** |
| **Median** | **29.89%** | **69.67%** | **28.92%** | **67.77%** | **30.93%** | **71.07%** | **36.88%** | **75.16%** |

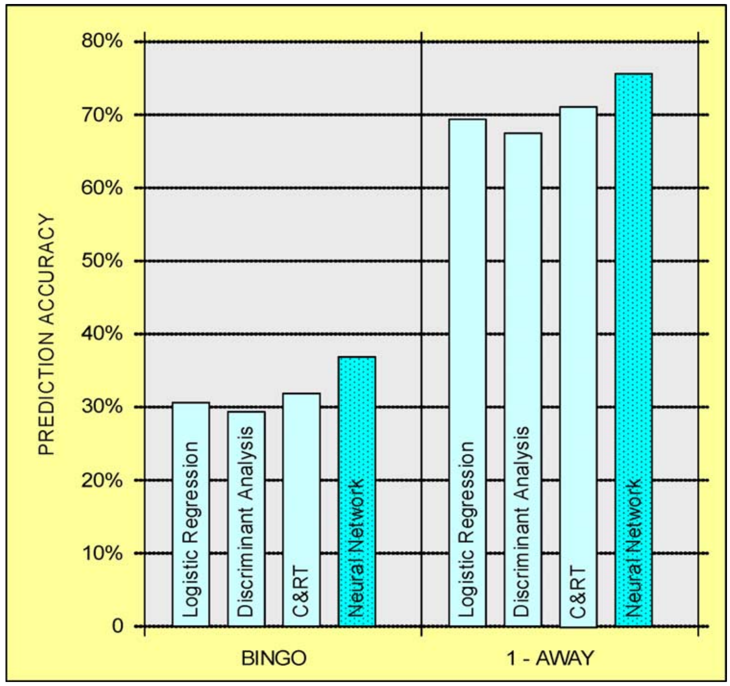


Fig. 4. Summary of the model comparison results presented in a bar chart.

the dependent variables are investigated as per the above-mentioned procedure. The results are summarized and presented as a column plot in Fig. 5. The $x$-axis represents the input variables and the $y$-axis represents the percent change on the output variables, while the input variables (one at a time) are perturbed gradually around their mean with the magnitude of $\pm 1$ standard deviation.

According to the sensitivity results, the major contributors to the prediction of the success of a motion picture are *number of screens, high technical effects* and *high star value*. Contrary to some of the earlier studies, the variables such as *competition, MPAA ratings*, and most of the *genre types* do not seem to have a significant contribution to the prediction of motion picture success.

## 5. Discussion and conclusion

The results show that the neural networks employed in this study can predict the success category of a motion picture before its theatrical release with pinpoint accuracy (i.e. Bingo) with 36.9% and within one category with 75.2% accuracy. Compared to the limited number of previously reported studies, these prediction accuracies are

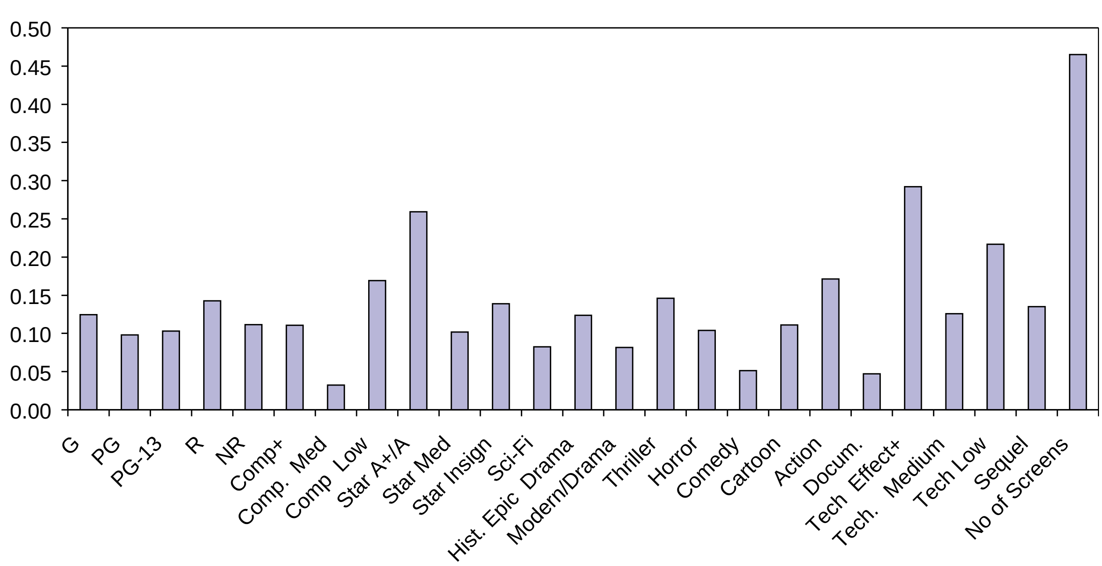


Fig. 5. Sensitivity analysis results for all independent and dependent variables.

significantly better. Compared to the other model types (i.e. logistic regression, discriminant analysis and classification and regression trees), using the exact same experimental conditions, as reported in Section 4, neural networks performed significantly better. Because the neural network models built as part of this study are designed to predict the financial success of a movie before its theatrical release, they can be used as a powerful decision aid by studios, distributors, and exhibitors. In any case, the results of this study are very intriguing, and once again, prove the continued value of neural networks in addressing difficult prediction problems.

Other leading prediction models reported in this application domain such as Neelamegham and Chintagunta (1999) and Sawhney and Eliashberg (1996) appear to do a good job of predicting the box office performance based upon initial viewer feedback and/or based on the actual initial box-office receipts. They do not, however, either address or do a good job of predicting the initial financial figures. Thus, it is conceivable that our model can be used to develop the initial forecasts that are then refined using the models mentioned above for a complete multi-stage life-cycle success of a movie. Use of different forecasting techniques at different stages of forecasting may provide a better overall performance than application of the same model for all stages. Multiple forecasting techniques may be used even at a single stage to allow for verification and validation of forecasts generated by incumbent techniques. In that sense, the neural network modeling offers a competitive forecasting technique in this context.

Beyond the accuracy of our results in predicting box-office success, these neural network models could also be adapted to forecast the success rates of other media products. The particular parameters used within the model of a movie or other media products could be altered using the already trained neural network model, in order to better understand the implications of different parameters on the end results-box-office receipts. During this alternative experimentation process, the manager of a given entertainment firm could find out, with a fairly high accuracy level, how much a specific actor, a specific release date, or the addition of more technical effects, could mean to the financial success of a film, or a television program.

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 these 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.

From an application perspective, once developed to a production system, such a neural network model can be made available (via a web server or as an application service provider) to industrial decision makers, where individual users can plug in their own movie parameters to forecast the potential success of a motion picture before its theatrical release. A neural network model can be designed in a way such that it can calibrate its weights (continuous self learning) by taking into account new samples (movies that are released and determined box-office receipts) as they become available. Much additional work, in terms of modeling extensions, further experimentation for testing the performance, and applications to other media product demand forecasting, remains to be done.

## Acknowledgements

Several graduate assistants have participated in this project as the first author attempted to continue the study for several years. These include, in chronological order, Edith Meany, Niketu Mithani, and Prajeeb Kumar. Their contributions for data collection and analysis are much appreciated. We also gratefully acknowledge Professors Joshua Eliashberg, Mohanbir Sawhney, Pradeep Chintagunta, and Ramya Neelamegham's assistance in replicating models proposed by them.

## References

Breiman, L. M. (1996). Some properties of splitting criteria. *Machine Learning*, *24*(1), 41–47. Coskunoglu, O., Hansotia, B., & Muzaffar, S. (1996). A New Logit model for decision making and its applications. *The Journal of the Operations Research Society*, *36*(1), 35–41.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

Cheng, B., & Titterington, D. M. (1994). Titterington Neural networks: A review from a statistical perspective. *Statistical Science*, *9*(1), 2–54.

Coskunoglu, O., Hansotia, B., & Muzaffar, S. (1985). A New Logit model for decision making and its applications. *The Journal of the Operations Research Society*, *36*(1), 35–41.

De Silva, I. (1998). Consumer Selection of Motion Pictures, appeared in The Motion Picture Mega-Industry by B. Litman. Allyn & Bacon Publishing, Inc.: Boston, MA.

Dougherty, J.,Kohavi, R. & Sahami, M. (1995). Supervised and Unsupervised Discretization of Continuous Features, In Proc. Twelfth

International Conference on Machine Learning, Los Altos, CA. Morgan Kaufmann, pp. 194–202.

Elberse, A. & J. Eliashberg (2002). The Drivers of Motion Picture Performance: The Need to Consider Dynamics, Endogeneity and Simultaneity, to appear in the *Proceedings of the Business and Economic Scholars Workshop in Motion Picture Industry Studies*, Florida Atlantic University, pp. 1–15.

Eliashberg, J., & Sawhney, M. S. (1994). Modeling goes to hollywood: Predicting individual differences in movie enjoyment. *Management Science*, *40*(9), 1151–1173.

Eliashberg, J., Junker, J. J., Sawhney, M. S., & Wierenga, B. (2000). MOVIEMOD: An implementable decision support system for prerelease market evaluation of motion pictures. *Marketing Science*, *19*(3), 226–243.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Eugen*, *7*, 179–188.

Hornik, K., Stinchcombe, M., & White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward network. *Neural Networks*, *3*, 359–366.

Jones, M., & Ritz, J. C. (1991). Incorporating distribution into new product diffusion models. *International Journal of Research in Marketing*, *8*, 91–112.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In S. Wermter, E. Riloff, & G. Scheler (Eds.), *The fourteenth international joint conference on artificial intelligence (IJCAI), Montreal, Quebec, Canada* (pp. 1137–1145). San Francisco, CA: Morgan Kaufman Publishing.

Krider, R. E., & Weinberg, C. B. (1998). Competitive dynamics and the introduction of new products: The motion picture timing game. *Journal of Marketing Research*, *35*(1), 1–15.

Litman, B. R. (1983). Predicting success of theatrical movies: An empirical study. *Journal of Popular Culture*, *16*(9), 159–175.

Litman, B. R., & Ahn, H. (1998). Predicting financial success of motion pictures. In B. R. Litman (Ed.), *The motion picture mega-industry*. Boston, MA: Allyn & Bacon Publishing, Inc..

Litman, B. R., & Kohl, A. (1989). Predicting financial success of motion pictures: The 80's experience. *The Journal of Media Economics*, *2*(1), 35–50.

H. Liu, F. Hussain, T.L. Chew, M. Dash (2002). Discretization An Enabling Technique, Data Mining and Knowledge Discovery, Vol. 6, pp. 393–423.

Neelamegham, R., & Chintagunta, P. (1999). A Bayesian Model to forecast new product performance in domestic and international markets. *Marketing Science*, *18*(2), 115–136.

NeuroDimension, Inc. (2004). Developers of NeuroSolutions v4.01: Neural Network Simulator. The World Wide Web address is www.nd.com, Gainesville, FL.

Principe, J. C., Euliano, N. R., & Lefebre, W. C. (2000). *Neural and adaptive systems: Fundamentals through simulations*. New York: John Wiley and Sons.

Radas, S., & Shugan, S. M. (1998). Seasonal marketing and timing new product introductions. *Journal of Marketing Research*, *35*(3), 296–315.

Ravid, S. A. (1999). Information, blockbusters, and stars: A study of the film industry. *Journal of Business*, *72*(4), 463–492.

Sawhney, M. S., & Eliashberg, J. (1996). A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science*, *15*(2), 113–131.

Sharda, R. (1994). Neural networks for the OR/MS analyst: An application bibliography. *Interfaces*, *24*(2), 116–130.

R. Sharda., H. Amato & E. Meany (2000). Forecasting Gate Receipts Using Neural Networks and Rough Sets, the *proceedings of the International DSI Conference*, Athens, Greece, pp. 1–5.

ShowBiz Data, Inc. (2002). The World Wide Web address is www.showbizdata.com, Los Angeles, CA.

Simon, H.A. (1981). The sciences of the Artificial, 2nd ed., Cambridge, MA.

Sochay, S. (1994). Predicting the performance of motion pictures. *The Journal of Media Economics*, *7*(4), 1–20.

Valenti, J. (1978). Motion Pictures and Their Impact on Society in the Year 2000, speech given at the Midwest Research Institute, Kansas City, April 25, p. 7.

White, H. (1989a). Learning in artificial neural networks: A statistical perspective. *Neural Computation*, *1*, 425–464.

White, H. (1989b). Some asymptotic results for learning in single hidden-layer feedforward network models. *Journal of American Statistical Association*, *84*(408), 1003–1013.

White, H. (1990). Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings. *Neural Networks*, *3*, 335–349.

Zufryden, F. S. (1996). Linking advertising to box office performance of new film releases: A marketing planning model. *Journal of Advertising Research*, *July–August*, 29–41.