A Method of Movie Business Prediction Using Back-propagation Neural Network

Debaditya Barman
Department of Computer Science and Engineering Jadavpur University, Kolkata - 700032, India
E-mail: debadityabarman@gmail.com

Nirmalya Chowdhury
Department of Computer Science and Engineering Jadavpur University, Kolkata - 700032, India
E-mail: nirmalya_chowdhury@yahoo.com

Abstract— Film industry is the most important component of Entertainment industry. Profit and Loss both are very high for this business. Before release of a particular movie, if the Production House or distributors gets any type of prediction that how the film will do business, then it can be helpful to reduce the risk. In this paper we have proposed, back propagation neural network for prediction about the business of a movie. Note that, this method is successfully applied in the field of Stock Market Prediction, Weather Prediction and Image Processing.

Index Terms— Film Industry, Artificial Neural Network, Back- Propagation

I. Introduction

A movie [1], also called a film or motion picture, is a series of still or moving images. It is produced by recording photographic images with cameras, or by creating images using animation techniques or visual effects. The process of filmmaking has developed into an art form and has created an industry in itself.

Films are cultural artifacts created by specific cultures, which reflect those cultures, and, in turn, affect them. It is considered to be an important art form, a source of popular entertainment and a powerful method for educating or indoctrinating citizens. The visual elements of cinema give motion pictures a universal power of communication.

Film Industry is an important part of present-day mass media industry or entertainment industry (also informally known as show business or show biz). This industry [2] consists of the technological and commercial institutions of filmmaking: i.e. film production companies, film studios, cinematography, film production, screenwriting, pre-production, post production, film festivals, distribution; and actors, film directors and other film crew personnel.

The major business centers of film making are in the United States, India, Hong Kong and Nigeria. The average cost [3] of a world wide release of a Hollywood film or American film (including pre-production, film and post-production, but excluding distribution costs) is about $65 million. It can be stretched up to $300 million [4] (Pirates of the Caribbean: At World's End). Worldwide gross revenue [5] can be almost $2.8 billion (Avatar). Profit-loss is found to vary from a profit [6] of 2975.63 % (City Island) to a loss [7] of 1299.7 % (Zyzzyx Road). So it will be very useful if we can develop a prediction system which can predict about Film's business potential.

Many artificial neural network based methods are used to design for successful Stock Market Prediction[8], Weather Prediction[9], Image Processing [10], and Time Series Prediction [11], and Temperature Prediction system [12] etc. Here we have proposed a method based on back propagation neural network for prediction of profit/loss of a movie based on some pre-defined genres. The formulation of the problem is presented in the next section. Section III describes our proposed method. The method is presented in the form of an algorithm in section III-A. Experimental results on ten movies selected randomly from a given database can be found in section IV. Concluding remarks and scope for further work has been incorporated in section V.

II. Statement of the Problem

Every Film can be identified by certain film genres. In film theory, genre [13] refers to the method based on similarities in the narrative elements from which films are constructed. Most theories of film genre are borrowed from literary genre criticism. Some basic film genres are - action, adventure, animation, biography, comedy, crime, drama, family, fantasy, horror, mystery, romance, science-fiction, thriller, war etc. One film can belong to more than one genre. Like the movie titled "Avatar (2009)" belongs to [14] action, adventure, and fantasy genres.

Any film's success is highly dependent on its film genres. Other important factors are reputation of Film Studio or Production house and present popularity of casted actor/actress. We can consider these genres and the said factors as a Film's attributes. We can then collect these attribute's data of past Films. Based on these data we can predict about an upcoming Film's future business.

We have used 20 movie genres like Action, Adventure, Animation, Biography, Comedy, Crime, Documentary, Drama, Family, Fantasy, History, Horror, Musical, Mystery, Romance, Science Fiction, Sport, Thriller, War, and Western. Note that the factors such as the overall rating giving by the viewers, reputation of the film distributors and present popularity of actor/actress performed for the film, has been taken care of by the inclusion of the following 3 attributes- Distributor reputation, overall rating and Casting rating. All these rating are given in 10-scale rating. We have used all these 23 attribute's and normalized profit percentage value to train the neural network. After training we have used it to predict the normalized profit percentage of a given movie.

Artificial neural network [15] learning methods provide a robust approach to approximating real-valued, discrete-valued, and vector-valued target functions. For certain types of problems, such as learning to interpret complex real-world sensor data, artificial neural networks are among the most effective learning methods currently known.

Artificial neural network are applied in image recognition and classification[16], image processing [17], feature extraction from satellite images [18], cash forecasting for a bank branch [19] , stock market prediction [20], decision making [21], temperature forecasting [22], atomic mass prediction [23], Prediction of Thrombo-embolic Stroke [24], time series prediction [25], forecasting groundwater level [26].

Back-propagation is a common method of teaching artificial neural networks about how to perform a given task. It is a supervised learning method. It is most useful for feed-forward networks [27].

Back-propagation neural network is successfully applied in image compression [28], satellite image classification [29], irregular shapes classification [30], email classification [31], time series prediction [32], bankruptcy prediction [33], and weather forecasting [34].

III. Proposed Method

In this paper, we have proposed a method that uses a multilayer feed-forward neural network as shown in Fig 1. Note that, a multilayer feed-forward neural network consists of an input layer, one or more hidden layers, and an output layer.

Here the back-propagation algorithm [35] performs learning on the said multilayer feed-forward neural network. It iteratively learns a set of weights for prediction of the class label of instances. An example of a multilayer feed-forward
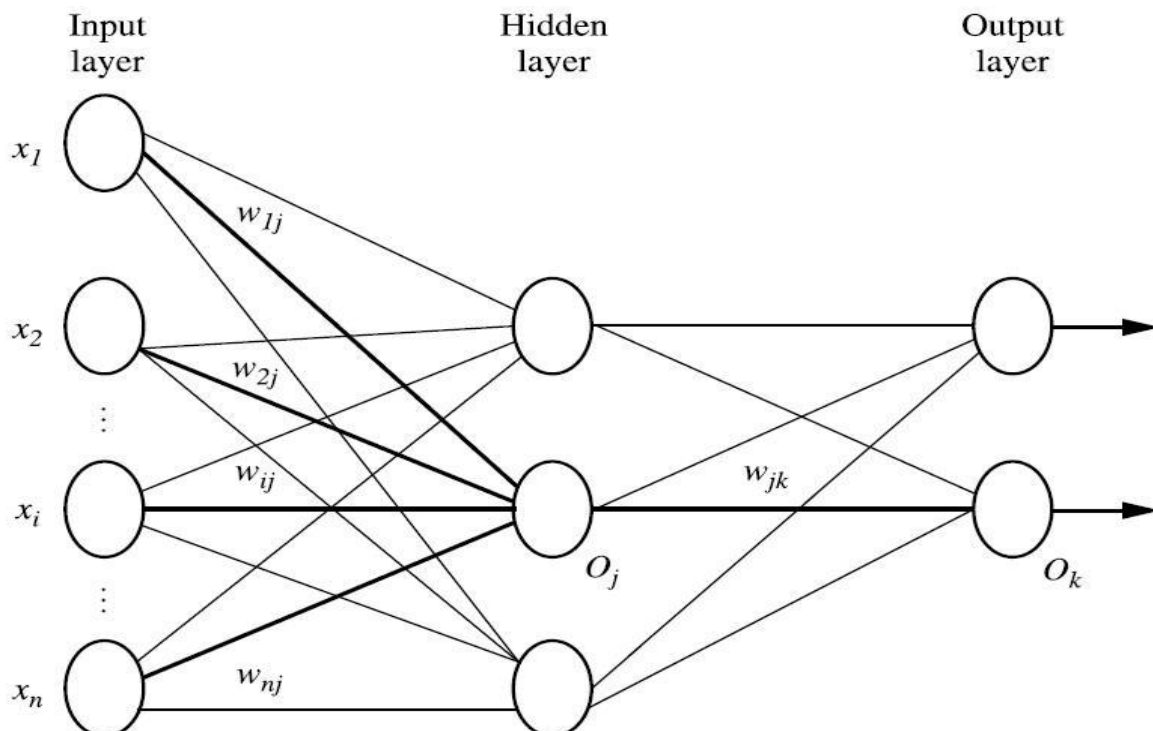
Network is shown in Fig 1.

Fig. 1: a multilayer feed-forward neural network

Since we have 23 attributes to be considered for this method, we need 23 nodes in the input layer, we have taken 10 nodes in the hidden layer and 1 node in the output layer.

Note that input nodes have received real numbers that represents the values of the individual genres. A positive (negative) real number is generated at the output node which indicates the predicted profit (loss) of the movie of under consideration.

We have used a database of released movie in the year 2011 [36], 2010 [37] and 2009 [38] for training of the said network. After the network has been successfully trained it can be used for prediction of profit/loss of new movie to be released. In our experiment we have taken some of the released movie (not taken in the training set) of 2010 for evaluating the efficiency of the trained back-propagation network. The experimental results are presented at the section IV.

3.1 Algorithm

Back-propagation Neural network learning for prediction, using the back-propagation algorithm.

Input

- *D*, a data set consisting of the genres of the movies and their actual values in percentage of their profit or loss associated target values(profit or loss percentage);

- *l*, the learning rate;

Output

- *A trained neural network, which can predict profit percentage.*

Method

$Step1.$ Initialize all weights and biases in *network*;

$Step2.$ While terminating condition is not satisfied {

$Step3.$ For each training instance X in D {

// propagate the inputs forward:

$Step4.$ For each input layer unit j {

$Step5.\ O_j = I_j$ // output of an input unit is its actual input value

$Step6.$ For each hidden or output layer unit j {

$Step7.\ I_j = \sum_i W_{i,j} O_i + \theta_j;$ //compute the net input of unit j with respect to the

Previous layer, i

$Step8.\ O_j = \frac{1}{1+e^{-I_j}}\ ;\}$ // compute the output of each unit j

// Back propagate the errors:

$Step9.$ For each unit j in the output layer

$Step10.\ Err_j = O_j(1 - O_j)(T_j - O_j);$ // compute the error

$Step11.$ For each unit j in the hidden layers, from the last to the first hidden layer

$Step12.\ Err_j = O_j(1 - O_j)\sum_k Err_k\ w_{jk}\ ;$ / compute the error with respect to the next higher

Layer, k

$Step13.$ For each weight w_{ij} in network {

$Step14.\ \Delta w_{ij} = (l)Err_j\ O_i\ ;$ // weight increment

$Step15.\ w_{ij} = w_{ij} + \Delta w_{ij}\}\ ;$ // weight update

$Step16.$ For each bias θ_j in *network* {

$Step17.\ \Delta\theta_j = (l)Err_j\ ;$ // bias increment

$Step18.\ \theta_j = \theta_j + \Delta\theta_j\ ;$ // bias update

$Step19.\ \}\}$

$Step20.$ Stop

At first, the weights in the network are initialized to small random numbers [39], bias associated with each unit also initialized to small random numbers.

The training instance from movie database is fed to the input layer. Next, the net input and output of each unit in the hidden and output layers are computed. A hidden layer or output layer unit is shown in Fig 2.
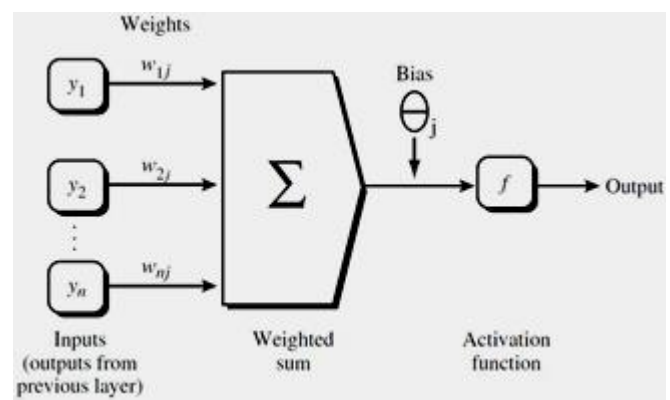

Fig. 2: A hidden or output layer unit j

The net input to unit j is

$$I_j = \sum_i w_{ij}\ O_i + \theta_j \qquad (1)$$

Where w_{ij} is the connection weight from unit i, in the previous layer to unit j; O_i is the output of unit i from the previous layer; and θ_j is the bias of the unit.

As shown in the Fig 2, each unit in the hidden and output layers takes its net input and then applies an activation function to it. The function (sigmoid) symbolizes the activation of the neuron represent by the unit. Given the net input I_j to unit j, then O_j, the output of unit j computed as

$$O_j = \frac{1}{1+e^{-I_j}} \tag{2}$$

The error of each unit is computed and propagated backward. For a unit j in the output layer the error Err_j is computed by

$$Err_j = O_j(1 - O_j)(T_j - O_j) \tag{3}$$

where O_j is the actual output of unit j, and T_j is the known target value of the given training instance. The error of a hidden layer unit j is

$$Err_j = O_j(1 - O_j)\sum_k Err_k \, w_{jk} \tag{4}$$

where w_{jk} is the weight of the connection from unit j to a unit k in the next higher layer, and Err_k is the error of unit k.

The weights and biases are updated to reflect the propagated errors. Weights are updated

By the following equations, where Δw_{ij} is the change in weight w_{ij}

$$\Delta w_{ij} = (l)Err_j \, O_i \tag{5}$$

$$w_{ij} = w_{ij} + \Delta w_{ij} \tag{6}$$

l is the learning rate. In our experiment it is 0.1.

Biases are also updated, if $\Delta \theta_j$ is the change in θ_j then

$$\Delta \theta_j = (l)Err_j \tag{7}$$

$$\theta_j = \theta_j + \Delta \theta_j \tag{8}$$

The weight and bias increments could be accumulated in variables, so that the weights and biases are updated after all of the instances in the training set have been presented. In our experiment we have used this strategy named epoch updating, where one iteration through the training set is an epoch.

The training stops when

- All Δw_{ij} in the previous epoch were so small as to be below some specified threshold, or

- The percentage of instance misclassified in the previous epoch is below some threshold,

Or

- A pre specified number of epochs have expired.

In our experiment we have specified the number of epochs as 1000.

IV. Experimental Result

We have carried out our experiments on a movie database containing 395 American films with 23 attributes, released in 2011, 2010 and 2009. We have used 385 movies out of the total 395 movies to train the neural network. The attributes of the remaining 10 movies are used as input of the trained network to predict the 10-scale profit percentage.

We have used MATLAB 7.10.0 f(R2010a) or our experiment. We have used its inbuilt NNTOOL (neural network tool) package. In our experiment, training Algorithm (TRAINLM) was Levenberg-Marquardt algorithm, adaptive learning function (LEARNGDM) was the gradient descent with momentum weight and bias learning function, and performance function (MSE) was mean squared error. Transfer function (TANSIG) was hyperbolic tangent sigmoid transfer function. In our experiment the value of minimum gradient magnitude (MIN_GRAD) was $1e^{-010}$. Initial value for parameter mu (μ) was 0.001. This value was multiplied by mu_dec (0.1) whenever the performance function is reduced by a step. It is multiplied by mu_inc (10) whenever a step would increase the performance function. If mu becomes larger than mu_max (10000000000), the algorithm terminates. The parameter mem_reduc (1) is used to control the amount of memory used by the algorithm. Maximum numbers of training epochs or iterations were 1000.

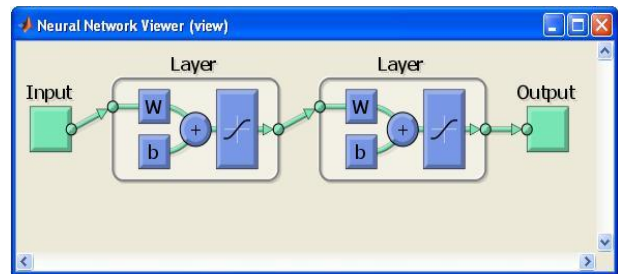Neural network, we have used in our experiment similar to following Fig 3.

Fig. 3: Neural network used for the experiment

It may be noted that the data about investment and earnings of all the movies are obtained from Wikipedia (http://en.wikipedia.org). And rating and genres of the movies are obtained from imdb (*http://www.imdb.com/*). Results obtained are presented in the following table (Table 1).

Table1: Experimental Result

No.	Name of the Film	Actual profit percentage (10 scale)	Predicted profit Percentage (10 Scale)
1	Tangled	1.272007446	4.0252
2	The Tourist	1.92902012	1.7805
3	Toy Story 3	4.315859555	4.9917
4	Tron: Legacy	1.353310371	2.6042
5	Twelve	-0.4866566	-0.11574
6	Unstoppable	0.766373326	0.21165
7	Wall Street: Money Never Sleeps	0.924971729	0.73692
8	The Wolfman	-0.04910428	0.0089708
9	Yogi Bear	1.519801763	1.7784
10	You Will Meet a Tall Dark Stranger	1.2850658	1.7083

We have defined a threshold value Δ (± 0.75) to measure performance of our proposed method. So, in general we are allowing up to 15% of error in our prediction. Let the 10 scale actual profit percentage be α. We have defined that the business of a movie is successfully predicted if the result is in the range of $\alpha + \Delta$ and $\alpha - \Delta$.

We have plotted (Fig 4.) our actual profit percentage α (10 scale), predicted profit percentage (10 scale), upper limit ($\alpha + \Delta$) and lower limit ($\alpha - \Delta$).
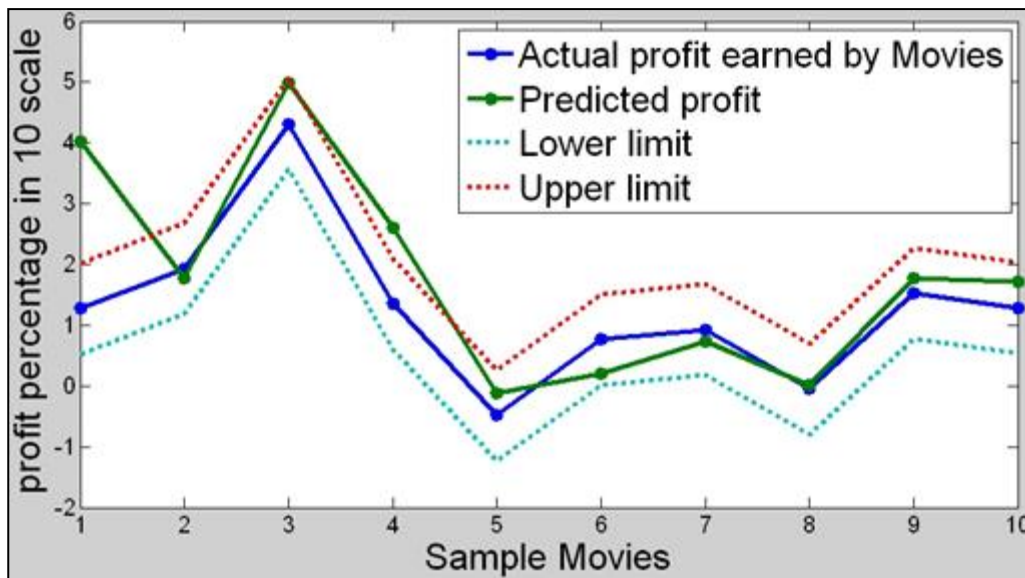

Fig. 4: Graph obtained from Experimental Result

From the above graph it can be found that 8 out of 10 given movie's business are successfully predicted by our proposed method. Thus the rate of successful prediction by the back-propagation network is 80%.

V. Conclusion and Scope for the Further Work

Note that, it is very difficult even for the human domain expert to predict the possible profit or loss of a new movie to be released. It seems that the genres of the movie play a significant role in the profit of the movie but it is very difficult to analytically establish the relation of the value of the genres of a given movie with the profit that it makes. In this paper we are attempted to develop a heuristic method using back-propagation neural network to solve this problem. And it feels that a success rate of 80% provided by the proposed method is significant considering the highly unpredictable nature of the movie business world.

Our proposed method fails to predict the actual business for the following movies titled "Tangled" and "Tron: Legacy". The possible reasons for this failure are stated below.

In case of the movie titled "Tangled" it was released in November 24, 2010 [40]. On this day other films which were released in the US box office were "Burlesque", "Faster", "Love and Other Drugs", "The Nutcracker" and "Tangled". The films- "Burlesque" had an overall rating of 6.1 [41] out of 10, "Faster" , "Love and Other Drugs" and "The Nutcracker" had an overall rating 6.5 [42], 6.6 [43], 4.3 [44] out of 10 respectively,

whereas "Tangled" had overall rating of 7.9 [45] out of 10. It is clear from the above mentioned rating that the movie titled "Tangled" was very good compared to other movies released on the same date. So it had done quite well business than expected. We believed that if some of the movies released on that day would have similar or better rating than the movie named "Tangled" then the said movie could not have made such a huge profit.

"Tron: Legacy", was a sequel to the 1982 film "Tron" [46]. In 1982 it was a box office hit; it had earned a profit of 94.11765% [47]. Although the movie titled "Tron: Legacy" had overall rating of 7.0 [48], but due to the fact that it was a sequel of a box office hit movie it had done quiet good business than anticipated.

Further research work can be conducted in the search for more genres and/or division of an existing genre into subgenres that may led to a higher success rate of prediction.

Acknowledgement

References

[1]   http://en.wikipedia.org/wiki/Film.

[2]   http://en.wikipedia.org/wiki/Film_industry

[3]   http://www.the-numbers.com/glossary.php

[4]   http://en.wikipedia.org/wiki/List_of_most_expensive_films

[5]   http://en.wikipedia.org/wiki/List_of_highest-grossing_films

[6]   http://en.wikipedia.org/wiki/City_Island_%28film%29

[7]   http://en.wikipedia.org/wiki/Zyzzyx_Road

[8]   "Stock Market Prediction with Back propagation Networks" by Bernd Freisleben Published in: IEA/AIE '92 Proceedings of the 5th international conference on Industrial and engineering applications of artificial intelligence and expert systems

[9]   "Training back propagation neural networks with genetic algorithm for weather forecasting" by Gill, J.; Singh, B.; Singh, S.; This paper appears in: Intelligent Systems and Informatics (SISY), 2010 8th International Symposium

[10] "Multispectral image-processing with a three-layer back propagation network" by McClellan, G.E.; DeWitt, R.N.; Hemmer, T.H.; Matheson, L.N.; Moe, G.O.; Pacific-Sierra Res. Corp., Arlington, VA This paper appears in: Neural Networks, 1989. IJCNN., International Joint Conference

[11] Time Series Prediction and Neural Networks R.J.Frank, N.Davey, S.P.Hunt Department of Computer Science, University of Hertfordshire, Hatfield, UK.

[12] An Efficient Weather Forecasting System using Artificial Neural Network Dr. S. Santhosh Baboo and I.Kadar Shereef

[13] http://en.wikipedia.org/wiki/Film_genre

[14] http://www.imdb.com/title/tt0499549/

[15] "Machine Learning" by Tom Mitchell page 81

[16] Application of artificial neural networks in image recognition and classification of crop and weeds" by C.C. YANG, S.O. PRASHER, J.A. LANDRY, H.S. RAMASWAMY and A. DITOMMASO

[17] "Applications of Artificial Neural Networks to Facial Image Processing " by Thai Hoang Le

[18] "The application of neural networks, image processing and CAD-based environments facilities in automatic road extraction and vectorization from high resolution satellite images " by F. Farnood Ahmadia, M. J. Valadan Zoeja, H. Ebadia, M. Mokhtarzadea

[19] Cash Forecasting: An Application of Artificial Neural Networks in Finance " by PremChand Kumar, Ekta Walia

[20] "Stock Market Prediction Using Artificial Neural Networks " by Birgul Egeli, Meltem Ozturan, Bertan Badur

[21] "Artificial neural network models for forecasting and decision making" by Tim Hill, Leorey Marquez, Marcus O'Connor, William Remus

[22] "Application of Artificial Neural Networks for Temperature Forecasting " by Mohsen Hayati, and Zahra Mohebi

[23] "Atomic Mass Prediction with Articial Neural Networks " by xuru.org

[24] "Designing an Artificial Neural Network Model for the Prediction of Thrombo-embolic Stroke " by D.Shanthi, Dr.G.Sahoo , Dr.N.Saravanan

[25] "Time Series Prediction and Neural Networks " by R.J.Frank, N.Davey, S.P.Hunt

[26] "Forecasting groundwater level using artificial neural networks " by P. D. Sreekanth, N. Geethanjali, P. D. Sreedevi, Shakeel Ahmed, N. Ravi Kumar and P. D. Kamala Jayanthi

[27] http://en.wikipedia.org/wiki/Backpropagation

[28] "Image Compression with Back-Propagation Neural Network using Cumulative Distribution Function " by S. Anna Durai, and E. Anna Saro

[29] "Satellite Image Classification using the Back Propagation Algorithm of Artificial Neural Network. " by Mrs. Ashwini T. Sapkal, Mr. Chandraprakash Bokhare and Mr. N. Z. Tarapore

[30] "Irregular shapes classification by back-propagation neural networks " by Shih-Wei Lin, Shuo-Yan Chou and Shih-Chieh Chen

[31] " Email Classification Using Back Propagation Technique " by Taiwo Ayodele, Shikun Zhou, Rinat Khusainov

[32] "Parallel back-propagation for the prediction of time series " by Frank M. Thiesing, Ulrich Middelberg and Oliver Vornberger

[33] "Applying back propagation neural networks to bankruptcy prediction " by Yi-Chung Hu and Fang-Mei Tseng

[34] "An Efficient Weather Forecasting System using Artificial Neural Network " by Dr. S. Santhosh Baboo and I.Kadar Shereef

[35] Data Mining: Concepts and Techniques, 2nd ed.Page 328 by Jiawei Han and Micheline Kamber

[36] http://en.wikipedia.org/wiki/List_of_2011_box_office_number-one_films_in_the_United_States

[37] http://en.wikipedia.org/wiki/List_of_American_films_of_2010

[38] http://en.wikipedia.org/wiki/List_of_American_films_of_2009

[39] Data Mining: Concepts and Techniques, 2nd ed.Page 328 by Jiawei Han and Micheline Kamber

[40] http://www.film-releases.com/film-release-schedule-2010.php

[41] http://www.imdb.com/title/tt1126591/

[42] http://www.imdb.com/title/tt1433108/

[43] http://www.imdb.com/title/tt0758752/

[44] http://www.imdb.com/title/tt1041804/

[45] http://www.imdb.com/title/tt0398286/

[46] http://en.wikipedia.org/wiki/Tron:_Legacy

[47] http://en.wikipedia.org/wiki/Tron_%28film%29

[48] http://www.imdb.com/title/tt1104001/

Debaditya Barman: Debaditya Barman is currently working on the project titled "In search of suitable methods for Clustering and Data mining" in Jadavpur University under Nirmalya Chowdhury. He has obtained his M.E. and B.E. (Computer Sc. And Engg.) from Jadavpur University, India. His area of interests are Machine Learning, Business intelligence, Data mining and Opinion mining.

Nirmalya Chowdhury: Nirmalya Chowdhury did his Ph.D. in Engineering from Jadavpur University, India in 1997. At present, he is working as Associate Professor in the department of Computer Science and Engineering, Jadavpur University, India. His fields of research include Pattern Recognition, Soft Computing, Natural Language Processing and Bioinformatics.