Prediction of movies box office performance using social media

Conference Paper · August 2013

DOI: 10.1145/2492517.2500232

6 authors, including:

Merin Jose
Kristu Jayanti College

Chien-Chung Chan
University of Akron

Kathy Liszka
University of Akron

Federico de Gregorio
University of Akron

Prediction of Movies Box Office Performance Using Social Media

Krushikanth R. Apala, Merin Jose, Supreme Motnam, C.-C. Chan, Kathy J. Liszka, and Federico de Gregorio[1]
Department of Computer Science
Department of Marketing[1]
The University of Akron
Akron, OH 44325
{chan, liszka}@uakron.edu

***Abstract*—In this study, we apply data mining tools to generate interesting patterns for predicting box office performance of movies using data collected from multiple social media and web sources including Twitter, YouTube and the IMDb movie database. The prediction is based on decision factors derived from a historical movie database, followers count from Twitter, and sentiment analysis of YouTube viewers' comments. We label the prediction in three classes, Hit, Neutral and Flop, using Weka's K-Means clustering tool. Interesting patterns for prediction are generated by Weka's J48. Since our prediction is for movies yet to be released in summer 2013, the performance of the final results will be validated by a follow-up study.**

Keywords—Sentiment analysis; Data mining; YouTube; Twitter; IMDb; Movie Trailer

I. INTRODUCTION

Social media such as Twitter and YouTube have been used for sharing contents and comments on all types of subjects by millions of people on a daily basis. It is clear that businesses have a strong interest in tapping into these huge data sources to extract information that might improve their decision making process. For example, predictive models derived from social media for successful movies may facilitate filmmakers making more profitable decisions.

The topic of movies is of considerable interest in the social media user community [1]. Research has been done to generate models for predicting revenues of movies. Most of them derived results from single data source. In [1], a linear regression model based on the chatter of Twitter [12] has been used to predict box-office revenues of movies. Tweets were used to study the hype factor of the gross collection of movies [2]. Some used Wikipedia activities to predict box-office performance [3]. The Internet Movie Database (IMDb) is another popular data source for these types of studies [4, 5]. In [15], the prediction of movies performance were based on the film critics' reviews where as in [16 - 19] predictions were done based on regression or stochastic models on IMDb data. In [20], a web mining approach was introduced to predict the movie success and academy awards in 2006. The approach combines social network analysis and automatic sentiment analysis. Trends and real world events in movie business were predicted by weighing the forum posts on IMDb. Reference [21] describes a movie rating approach based on data mining of 240 movies from IMDb where Weka and J48 were used to create the prototype model. In [22], methods to predict the popularity of movies were discussed to evolve as a guiding strategy for Content Distribution/ Delivery Network (CDD). Actor and director popularity were considered as base criteria for predicting the popularity of a movie. Data mining techniques to find the correlation between the quality of the movie and the budget of the movie along with other analysis to find the contributing factors of a successful movie were discussed in [4]. Another interesting work applies image processing and speech data processing to computerized systems [23]. The proposed system rates movies by considering factors such as violence and language used in the movie. However the accuracy of the system was around 66.67%. Bipolar classification of movie reviews was discussed in [24] where the reviews were obtained from two huge data corpus and were classified as positive and negative. YouTube is one of the most popular social media sites. Millions of people use YouTube to upload, share and talk about videos on each day. In a recent study on YouTube, it is estimated that 72 hours of video are uploaded every minute [6].

In this work, we generate a predictive model consisting of genre of successful movies ranked by user ratings of the IMDb movie database, popularity of the director, leading actor and leading actress represented by Twitter's followers count, popularity of movies represented by the number of comments and views of official movie trailers accessible by YouTube, and sentiment toward a movie derived from YouTube viewers' comments. We also consider if a movie is a sequel or not. Together, we have used eight attributes to represent the assessment of pre-released movies derived from social media sources. All attribute values are mapped into an integral ranking of 1 to 10 by a simple min-max normalization process. Then, the K-Means clustering tool of Weka [14] is applied to generate three classes of movies: Hit, Neutral, or Flop. Finally, we use Weka's J48 decision tree classifier to generate a predictive model.

In our experiments, we collected social media data during February and March of 2013 for 35 movies to be released in May, June and July 2013 in the US. It is interesting to note that only four out of the eight attributes are identified by J48 as relevant in predicting the success of a movie, namely, the popularity of actor and actress, genre and sequel status. The final results will be validated when the movies are released

ASONAM'13, August 25-29, 2013, Niagara, Ontario, CAN

this summer and we will present the updated results in the workshop.

In the following section, we will describe the details of data collection. Section III presents the experiments performed. Results and discussion are given in Section IV, followed by our conclusion.

II. DATA COLLECTION

We gathered data from official trailers of movies with pre-release dates in months of May, June and July 2013 on the US market. We identified these movies from the website http://www.comingsoon.net [7]. After identifying the movies, we tracked the trailers released for these movies from the http://themoviebox.net website [8]. From the trailer data, we can identify the video id number of these movie trailers on YouTube. The total list of 35 movies considered with Trailer name and release date is shown in Table I.

A developers guide for YouTube was written for different languages such as Java, .NET, PHP, Python and JavaScript. For interacting with the YouTube API, we wrote a code snippet in PHP which was run using the Wamp server. All the videos in YouTube have corresponding video IDs. We extracted the comments by passing the video ID to the code. Once the code is executed it uses the YouTube API to provide comment which we stored as a data file in text format. We used the open source Zend framework for developing web applications and services using PHP (http:// framework, zend.com). The Zend Framework uses 100% object-oriented code and utilizes most of the new features of PHP, namely namespaces, late static binding, lambda functions and closures [9].

YouTube data collection was done over a period of one month from February 28, 2013 to March 28, 2013. We extracted the view count and total number of comments daily. The YouTube API has a restriction where we can obtain the most recent 1000 comments only at a time [9]. Our tally for the total number of view count for 35 movies was over 77 million. The total number of comments was over 98K.

From Twitter, the followers count for directors was approximately 137K, for actors was 7.6 million, and for actresses was 20.8 million.

Table I. Movie Data

Movie Title	Release Date	Trailer Version
Iron Man 3	5/3/2013	jobloMovienetworks
The Iceman	5/3/2013	movieclipsTRAILERS
The Great Gatsby	5/10/2013	Cieon Movies
Peeples	5/10/2013	Official Trailer #1
Star Trek Into Darkness	5/15/2013	jobloMovienetworks
Black Rock	5/17/2013	movieclipsTRAILERS
Frances Ha	5/17/2013	Trailer
Fast & Furious 6	5/24/2013	Trailer
Epic	5/24/2013	Cieon Movies
The Hangover Part III	5/24/2013	Official Teaser Trailer
The East	5/31/2013	Official Trailer
After Earth	5/31/2013	Official Trailer #1
Now You See Me	5/31/2013	Official Trailer #1 [HD]
The Purge	6/7/2013	Official trailer
The Internship	6/7/2013	Official Trailer (HD)
Man of Steel	6/14/2013	Official Trailer #2 [HD]
This is the End	6/14/2013	Official Trailer
Monsters University	6/21/2013	Official Trailer #2 (HD)
Unfinished Song	6/21/2013	Official Trailer #1
World War Z	6/21/2013	Official Trailer #1
The Heat	6/28/2013	OFFICIAL TRAILER #2 HD
I'm So Excited	6/28/2013	Official Trailer (2013)
Kick-Ass 2	6/28/2013	Official Trailer
Despicable me 2	7/3/2013	Trailer (HD)
The Lone Ranger	7/3/2013	Official Trailer #2
Kevin Hart: Let Me Explain	7/4/2013	TEASER HD trailer
Grown Ups 2	7/12/2013	Official Trailer
Pacific Rim	7/12/2013	Official CES Trailer (HD)
The Conjuring	7/19/2013	Official Trailer (HD)
Girl Most Likely	7/19/2013	Trailer
Populaire	7/19/2013	Trailer
R.I.P.D	7/19/2013	Trailer
Turbo	7/19/2013	Official Trailer
The Wolverine	7/26/2013	Official Trailer
The Smurfs 2	7/31/2013	Official Trailer

Next we describe how the data for the attributes were collected.

A. Genre of the movie

The IMDb movie database is one the most comprehensive resources containing detailed information about almost any film ever made. It has a vast amount of data containing valuable information about general trends in films [10]. Movies are classified under the category genre as action, adventure, animation, comedy, drama, fiction, romance and thriller. We took the 50 top listed movies for the years 2010, 2011 and 2012 based on the ratings from IMDb. Table II shows the frequency weight distribution of the genres from the top 50 movies, which will be used as a measurement of genre popularity in our model.

Table II. Genre Frequency Distribution

Genre	Frequency
Action	0.52
Adventure	0.08
Animation	0.1
Comedy	0.02
Drama	0.06
Fiction	0.06
Romance	0.12
Thriller	0.04

B. Director, Actor, Actress Popularity

The popularity of director, leading actor and leading actress of a movie is represented by their followers count on Twitter [12]. The counts were collected during the same time period that we collected the YouTube data. Table III shows the followers counts of the directors, actors and actresses of the 35 movies in our data set. Note that one actor and one actress have no Twitter accounts, so the entries are set to zero. This may put those movies at a disadvantage.

C. View and Comment Count

We represent the popularity of a pre-released movie by the number of views and comments present for its official trailer. We collected the view and comment counts from the YouTube trailers in the same time period.

D. Sequel Movie

A movie is a sequel if it is a successor of an already released movie. It can have two values "Yes" and "No". The sequel list is obtained from IMDb [10]. While calculating the decision score, since a the successor of an already released movie has more popularity than a new movie, we have assigned numerical value 10 to "Yes" and 1 to "No".

E. Sentiment Analysis

Sentiment is analyzed on the whole set of comments extracted from YouTube. We collected a word list for identifying sentiment from [13]. The dataset was processed using a set of positive and negative words and then classified as positive, negative or neutral. The processing is carried out in the following manner. Firstly, the dataset is divided into a group of single comments. Each comment is then divided into tokens using the annotator property "tokenize". After tokenizing, each token is compared with the words present in positive and negative word files. Sentiment is based on frequency count of the number of words matched. Neutral sentiment indicates an equal count of positive and negative matches.

Table III. Director, Actor and Actress Followers Count.

Movie Title	Director	Actor	Actress
Iron Man 3	7943	530173	1618985
The Iceman	1836	4638	8553
The Great Gatsby	8303	568136	4520
Peeples	714	218448	814802
Star Trek Into Darkness	657	44600	4456331
Black Rock	7800	No Account	71805
Frances Ha	315	1042	740
Fast & Furious 6	6368	616,965	1,940,211
Epic	760	930313	6654
The Hangover Part III	212	6301	175
The East	2546	5817	11433
After Earth	7632	257326	148647
Now You See Me	37	67661	2513
The Purge	1374	27689	5402
The Internship	6930	6683	840
Man of Steel	8996	913187	308
This is the End	7863	395617	6749351
Monsters University	2927	9319	1074
Unfinished Song	2032	674	3527
World War Z	1140	102397	54
The Heat	9282	341953	42720
I'm So Excited	6021	40791	139
Kick-Ass 2	6194	447658	4957232
Despicable me 2	59	12389	767
THE LONE RANGER	302	517433	2429
Kevin Hart: Let Me Explain	3566	368115	No Account
Grown Ups 2	368	697261	473
Pacific Rim	209	23609	28270
The Conjuring	3154	14380	442
Girl Most Likely	4096	118	364
Populaire	5972	658	11571
RIPD	9681	14521	507
Turbo	2843	14521	910
The Wolverine	9010	400935	684
The Smurfs 2	30	22130	836

III. EXPERIMENT

There are three steps in our experiments: normalizing the training data, applying K-Means clustering and generating a predictive model.

For normalization, we applied the following simple min-max method [11], followed by applying a round function to map each attribute value into an integral rank of 1 to 10.

$$v_i' = \frac{v_i - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A$$

where v_i is the value to be mapped and v_i' is the new value obtained after normalization. Also max_A and min_A are the actual maximum and the minimum values. The new maximum and minimum values are new_max_A and new_min_A respectively.

For clustering, we used the K-Means tool from Weka. The movies were grouped into three classes: Hit, Neutral, and Flop. Before grouping, we considered two possibilities for assigning weights to the attributes. One way is to consider all attributes being equal and using a uniform weight assignment, the other is to give a higher weight to the sentiment attribute and treat the rest of attributes as equal. The results of clustering are shown in Table IV where we see that only one movie, "World War Z", is clustered differently. It is clustered as Neutral by the uniform weighted approach, and as Hit by the approach put more weight on Sentiment. By probing further, the movie does have high number of view and comment counts and positive sentiment, even though its actress has low number of followers count.

The clustering result is used to label each movie's decision value for creating a training set to generate a predictive model. We used Weka's J48 and Naïve Bayesian algorithms to create two models validated by 3-fold cross-validation. Results are presented in the following section.

Table IV. Clustering results

Movie Title	Uniform Weights	Non-Uniform Weights
Iron Man 3	Hit	Hit
The Iceman	Flop	Flop
The Great Gatsby	Neutral	Neutral
Peeples	Flop	Flop
Star Trek Into Darkness	Hit	Hit
Black Rock	Flop	Flop
Frances Ha	Flop	Flop
Fast & Furious 6	Hit	Hit
Epic	Neutral	Neutral
The Hangover Part III	Neutral	Neutral
The East	Neutral	Neutral
After Earth	Neutral	Neutral
Now You See Me	Flop	Flop
The Purge	Flop	Flop

The Internship	Flop	Flop
Man of Steel	Hit	Hit
This is the End	Hit	Hit
Monsters University	Neutral	Neutral
Unfinished Song	Flop	Flop
World War Z	Neutral	Hit
The Heat	Neutral	Neutral
I'm So Excited	Flop	Flop
Kick-Ass 2	Hit	Hit
Despicable me 2	Neutral	Neutral
THE LONE RANGER	Neutral	Neutral
Kevin Hart: Let Me Explain	Neutral	Neutral
Grown Ups 2	Neutral	Neutral
Pacific Rim	Neutral	Neutral
The Conjuring	Neutral	Neutral
Girl Most Likely	Flop	Flop
Populaire	Flop	Flop
RIPD	Neutral	Neutral
Turbo	Flop	Flop
The Wolverine	Hit	Hit
The Smurfs 2	Neutral	Neutral

IV. RESULTS AND DISCUSSION

A. Uniform Weighted Approach

In this approach equal weights are given are to all the attributes. Out of the 35 movies, 7 movies were clustered as a Hit, 12 movies were labeled as a Flop and 16 movies were grouped under Neutral, meaning prediction is not certain. Class labels are based on a total value score of 80, i.e, 10 for each of the eight attributes. The cut-off value range for each class is as follows: (1) Hit: in the range of 42 to 70, (2) Neutral: in the range of 21 to 33, and (3) Flop: in the range of 9 to 20.

The labeled dataset is then tested using Weka's Naïve Bayesian and J48 classifiers to generate predictive models. The resultant confusion matrices are shown in Tables V and VI. The decision tree generated by J48 is shown in Figure 1. From Figure 1, we can observe that the attributes present in the decision tree are Actress, Sentiment, Genre and Director. We can also observe that Actress is the root node which is the most relevant factor according to the J-48 Classifier.

Table V: Uniform Weighted Approach:
Confusion Matrix using Naïve Bayesian Classifier

Predicted / Labeled	Hit	Flop	Neutral
Hit	6	0	1
Flop	0	10	2
Neutral	1	0	15

Table VI: Uniform Weighted Approach:
Confusion Matrix using J-48 Classifier

Predicted / Labeled	Hit	Flop	Neutral
Hit	7	0	0
Flop	0	12	0
Neutral	0	0	16

```
J48 pruned tree
------------------

Actress  <= 2
|  Sequel <= 1
|  |  Genre <= 2
|  |  |  Actor <= 3: Box Office Flop (12.0)
|  |  |  Actor > 3: Neutral (2.0)
|  |  Genre > 2: Neutral (7.0)
|  Sequel > 1
|  |  Genre <= 2: Neutral (7.0)
|  |  Genre > 2: Box Office Hit (2.0)
Actress  > 2: Box Office Hit (5.0)
```

Fig. 1. J48 tree for Uniform Weighted Approach

B. Non-Uniform Weighted Approach

In this approach, more weight is given to the sentiment attribute. Out of the 35 movies, 8 are clustered as Hits, 12 as a Flop and 15 as Neutral. Class labels are based on a total value score of 100, i.e, 30 for the Sentiment attribute, and 10 for each of the rest seven attributes. The cut-off value range for each class is as follows: (1) Hit: in the range of 42 to 66, (2) Neutral: in the range of 22 to 35, and (3) Flop: in the range of 13 to 19.

The results of classification and testing using Weka are shown in Table VII and VIII. The decision tree generated using J48 is shown in Figure 2. From Figure 2, we can observe that the attributes present in the decision tree are Genre, Sequel and Actor. We can also observe that Genre is regarded the most relevant attribute according to J-48.

Table VII: Non-Uniform Weighted Approach:
Confusion Matrix using Naïve Bayesian Classifier

Predicted / Labeled	Hit	Flop	Neutral
Hit	7	0	1
Flop	0	10	2
Neutral	1	1	13

Table VIII: Non-Uniform Weighted Approach:
Confusion Matrix using J-48 Classifier

Predicted / Labeled	Hit	Flop	Neutral
Hit	6	0	2
Flop	0	12	0
Neutral	0	0	15

```
J48 pruned tree
------------------
Genre <= 2
|  Sequel <= 1
|  |  Actor <= 4: Box Office Flop (12.0)
|  |  Actor > 4: Neutral (2.0)
|  Sequel > 1: Neutral (7.0)
Genre > 2
|  Sequel <= 1: Neutral (8.0/2.0)
|  Sequel > 1: Box Office Hit (6.0)
```

Fig. 2. J48 tree for Non-Uniform Weighted Approach

The re-substitution error estimates of Uniform and Non-Uniform Weight approaches generated by Weka are shown in Table IX. The resulting clustering results are shown in Table X.

Table IX: Comparisons of Uniform and Non-Uniform Weighted Approaches

Approach	Accuracy (Naïve Bayesian)	Accuracy(J48)
Uniform Weighted	89	100
Non Uniform Weighted	86	94

Table X: Category of movies in Uniform and Non-Uniform Weighted Approaches

	Hit	Neutral	Flop
Uniform	7	16	12
Non-Uniform	8	15	12

C. *Preliminary Results*

Among the 35 movies on our list, 15 of them have been released up to the weekend of June 7, 2013. We have tracked movie gross collections from the IMBbPro Movie Website http://pro.imdb.com and http://www.boxofficemojo.com. Two released movies, “Black Rock” and “Frances Ha”, were not listed; therefore, they are excluded from the list of our preliminary results. Table XI shows the title of movies that have gross collections listed at the websites. For each movie, we tracked the budget of the movie and weekly world wide release and USA release gross collections separately. From the budget and gross income, we calculated the net profits and profit ratios. Table XII shows the preliminary results for the movies listed on Table XI. The WK column gives the number of weeks a movie has been released up to June 7, 2013. The Pred. column indicates the predicted results from our model where N denoting Neutral. The WP column is worldwide release profit, and WPR is the worldwide profit ratio. Similarly, the USP is USA release profit and USPR is the profit ratio of the US release. The profits are shown in millions with negative profits denoted in parenthesis.

Table XII shows only one misclassification by our model at this point, Movie #10 “The Purge”. Among the five movies classified as Neutral by our model, one seems to be moving toward a Hit, two tend to be a Flop, and two have mixed gross profits from the worldwide and US releases.

Table XI: Partial list of movie titles

Movie	Title
1	Iron Man 3
2	The Iceman
3	The Great Gatsby
4	Peeples
5	Star Trek Into Darkness
6	Fast & Furious 6
7	Epic
8	The Hangover Part III
9	The East
10	After Earth
11	Now You See Me
12	The Purge
13	The Internship

Table XII: Preliminary box office performance

Movie	WK	Pred.	WP	WPR	USP	USPR
1	6	Hit	$990.00	5.95	$194.00	1.97
2	6	Flop	($8.15)	0.19	($8.15)	0.19
3	5	N	$174.00	2.66	$31.00	1.30
4	5	Flop	($5.88)	0.61	($5.88)	0.61
5	4	Hit	$186.00	1.98	$10.00	1.05
6	3	Hit	$425.00	3.66	$43.00	1.27
7	3	N	$89.00	1.89	($16.00)	0.84
8	3	N	$170.00	2.65	($1.00)	0.99
9	2	N	($6.16)	0.05	($6.16)	0.05
10	2	N	($35.30)	0.73	($84.00)	0.35
11	2	Flop	($14.10)	0.81	($14.10)	0.81
12	1	**Flop**	$31.10	11.37	$31.10	11.37
13	1	Flop	($40.70)	0.30	($40.70)	0.30

V. CONCLUSION

In this paper, predictive models for the box office performance of the movies was represented by factors derived from social media and IMDb. According to our models, we have identified the following patterns: (1) the popularity of leading actress is crucial to the success of a movie, (2) the combination of past successful genre and a sequel movie is another pattern for success, (3) a new movie in the not popular genre and an actor with low popularity could be a pattern for a Flop. It is surprising that sentiment score and view and comment counts were not identified as relevant in our experiments. We believe it is related to how weights are assigned to each attribute. Further studies to determine different weighting methods will be beneficial. In addition, our prediction is for movies yet to be released. The preliminary result of tracking 13 of the movies shows a good prediction performance from our model. A follow-up study on the final performance of our models will be validated and presented once all of the movies are released. Future work to improve our models will include further refinement of the Neutral class and characterization of movie box office performance in terms of net profits and profit ratios.

VI. REFERENCES

[1] Sitaram Asur and Bernardo A. Huberman, “Predicting the Future with Social Media,” http://arxiv.org/abs/1003.5699, March 2010.

[2] Ajay Siva Santosh Reddy & Pratik Kasat, "Box-Office Opening Prediction of Movies based on Hype Analysis through Data Mining", *International Journal of Computer Applications* 56(1):1-5, October 2012. Published by Foundation of Computer Science, New York, USA. DOI: 10.5120/8852-2794.

[3] Márton Mestyán, Taha Yasseri, János Kertész, "Early Prediction of Movie Box Office Success based on Wikipedia Activity Big Data," http://arxiv.org/abs/1211.0970.

[4] M. Saraee, S. White & J. Eccleston, "A data mining approach to analysis and prediction of movie ratings", University of Salford, England, The Fifth International Conference on Data Mining, Text Mining and their Business Applications, 15-17 September 2004, pp 344-352.

[5] W. Zhang and S. Skiena, "Improving Movie Gross Prediction Through News Analysis," 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, pp. 301 – 304.

[6] "YouTube", youtube.com, Inc., [Online]. Available: http://www.youtube.com/yt/press/statistics.html

[7] "Coming soon movies", Inc, . [Online]. Available: http://www.comingsoon.net/movies.php

[8] "Movie box", Inc, . [Online] Available: http://themoviebox.net/view/trailers

[9] "Google Developers", Inc,. [Online] Available: https://developers.google.com/youtube/2.0/developers_guide_protocol

[10] "Internet movie database," IMDb.com, Inc. , [Online]. Available: http://imdb.com. [Accessed 26 April 2013].

[11] Han, Jiawei and Micheline Kamber, Data Mining: Concepts and Techniques, 3rd Edition, Morgan Kaufmann Publishers, 2011, ISBN: 978-0-12-381479-1, pp. 112-114

[12] "Twitter," Twitter.com, Inc. , [Online]. Available: http://twitter.com. [Accessed 16 April 2013].

[13] "Sentimental Analysis", Inc. [Online]. Available: http://www.cs.uic.edu/~liub/FBS/sentiment-analysis [Accessed 23 March 2013].

[14] M. Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, Volume 11, Issue 1.

[15] M. Joshi, D. Das, K. Gimpel, and N. A. Smith. Movie reviews and revenues: An experiment in text regression. In Proceedings of NAACL-HLT, 2010.

[16] A. Chen, "Forecasting gross revenues at the movie box office," Working paper, University of Washington Seattle, WA, June 2002.

[17] J. S. Simonoff and I. R. Sparrow, "Predicting movie grosses: Winners and losers, blockbusters and sleepers," Chance, vol. 13(3), pp. 15–24, 2000.

[18] M. S. Sawhney and J. Eliashberg, "A parsimonious model for forecasting gross box-office revenues of motion pictures," Marketing Science, vol. Vol. 15, No. 2, pp. 113–131, 1996.

[19] R. Sharda and E. Meany, "Forecasting gate receipts using neural network and rough sets," in Proceedings of the International DSI Conference, 2000, pp. 1–5.

[20] Krauss, Jonas; Nann, Stefan; Simon, Daniel; Fischbach, Kai," Predicting movie success and academy awards through sentiment and social network analysis," University of Cologne.

[21] S. Kabinsingha, S. Chindasorn, C. Chantrapornchai, "Movie Rating Approach and Application Based on Data Mining", International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 1, July 2012.

[22] L. Zhu, M. Zhu and S. Yao, "The popularity of movies predict system based on data mining technology for CDN," in IEEE International Conference on the 3rd Computer Science and Information Technology, , 2010.

[23] P. Changkaew and R. Kongkachandra, "Automatic Movie Rating Using Visual and Linguistic Information," in NCCIT, Bangkok, Thailand, 2010.

[24] P. Chaovalit and L. Zhou, "Movie Review Mining: a Comparison between Supervised and Unsupervised," in the 38th Hawaii International Conference on System Sciences, Hawaii, USA, 2005.