Movie Business Trend Prediction using Market Basket Analysis

Debaditya Barman
Assistant Professor, Department of Computer Science
University of Gour Banga
Malda -732103, West Bengal, India

Nirmalya Chowdhury,Ph.D
Associate Professor, Department of Computer Science and Engineering
Jadavpur University
Kolkata - 700032, India

ABSTRACT
Trend analysis can be defined as a process of comparing past business data to identify any significant, consistent results or trends. It is very useful method to understand any business's performance. Successful business trend analysis can take business to the right direction. Film industry is the most important component of Entertainment industry. Profit and loss both are very high for this business. A large amount of money invested in this high risk industry. Before release of a particular movie, if the Production House gets any type of business trend prediction about how the film will do business, it can be helpful to reduce the risk. It is a known fact that viewer's taste in movies can change time to time. For instance in 2009, most profitable movie in USA was New Moon (1319.422%) which belongs to Adventure, Drama and Fantasy genres. So, profitable trend determination by movie genre analysis can be helpful for making business decisions like marketing strategy, advertising strategy etc. In this work we have tried to determine and predict the trend in movie business by analyzing movie genres. Film genres find both academic and practical applications as films can be categorized by genre at every stage of their existence, from the initial approach the screenwriter takes, to where they end up on the shelves of our local store, to how their impact on cultural history is assessed. There is also a lot of commercial interest in the way people classify and choose to watch movies — this is very important for the initial marketing of a movie, and for companies like Netflix or LoveFilm, who rely on genre categories to help their customers make their picks.

General Terms
Machine intelligence, Pattern Recognition

Keywords
Film industry, Film genre, Trend prediction, Market Basket Analysis, Apriori rules

1. 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.

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.

The process of filmmaking has developed into an art form and has created an industry in itself. 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).

It is a known fact that viewer's taste in movies can change time to time. Following examples can be helpful to illustrate this fact. In 2009 most profitable movie in USA was New Moon (1319.422%) which belongs to Adventure, Drama and Fantasy genres. Paranormal Activity 2 released in 2010 was most profitable movie in USA(5791.452%) a Horror movie. In 2011 most profitable movie in USA was Insidious (5606.421%) which belongs to Horror and Thriller genres. So it is pretty clear that profitable trend determination and prediction can be helpful for making business decisions like marketing strategy, advertising strategy etc. Target audience also can be determined by using the MPAA (Motion Pictures Association of America) rating.

Association rule mining [8] finds interesting associations and/or correlation among large set of data items. Association rules show attributes value conditions that occur frequently together in a given dataset. Association rules also provide information of this type in the form of "if-then" statements. These rules are computed from the data and unlike the if-then rules of logic; association rules are probabilistic in nature. In addition to the antecedent (the "if" part) and the consequent (the "then" part), an association rule has generally two measurements (Support and Confidence) that express the interestingness about the rule.

Association rule mining has been successfully applied in Retail marketing [9], Medical Diagnosis [10], Analysis of Protein Sequences [11], Analysis of telephone calling patterns [12], Fraud detection [13], Analysis of Census Data [14] etc.

The formulation of the problem is presented in the next section. Section III describes our proposed method. The proposed method is presented in the form of an algorithm in section III-A. Experimental results can be found in section IV. Concluding remarks and scope for further work have been incorporated in section V.

2. STATEMENT OF THE PROBLEM

Every Film can be identified by certain film genres. In film theory, genre [15] 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. it One film can belong to more than one genre. As an example the movie titled "Alice in Wonderland (2010)" belongs to [16] action, adventure, and fantasy genres. Viewer's selection of a movie is often based on the film's genres. More viewers certainly can lead a movie to achieve more success. So we can say that any film's success is highly dependent on its film genres.

Inspired from "Market Basket Analysis"- a typical example of frequent item set mining, we have used it to predict movie trend. Viewer's choice of selection in movies has been analyzed by finding associations between different genres.

Each genre is represented by a Boolean variable which represents presence or absence of that genre in a particular movie. Thus each Movie or film can be represented by a Boolean vector of values assigned to these Boolean variables. These vectors can be analyzed for extracting the genres that are frequently associated or found together. For example in a certain month it may be found that those movies make profits which are having nonzero values for romantic and comedy genres. This type of association between genres and profit can be represented by following rule.

$$Romantic\ (Yes) \wedge Comedy\ (Yes) \Rightarrow Profit\ (Yes)$$

$$[support = 50\%, confidence = 95\%]$$

Support and confidence are two measures of any rule's significant. They respectively reflect usefulness and certainty of discovered rules. A support of 50% means 50% of the entire movies under the analysis are romantic comedy and they made profit. A confidence of 95% means that 95% of all romantic comedy movies made profit. Typically, association rules can be considered as interesting if they satisfy both a minimum support and confidence threshold. We will discuss about these threshold values later.

These associations of genres can be treated as viewer's trend or movie business's trend. These business trends can help Film Distributors to develop business strategies that we have discussed earlier. For instance if in a certain month viewers are preferring romantic- comedy films, the distributors should launch their romantic-comedy films at that month. If previous month's trend was adventure-action film, then distributors should go for extensive advertising or marketing for their upcoming adventure-action films.

We have considered only successful movies (with a profit of more than 0%) for our experiments. Due to lack of published materials, we cannot get any idea about the duration or lifetime of any such movie business trend. So for trend determination we have assumed that minimum lifetime of a movie business trend is one month. The lifetime of the trend is increased until we found another trend. We have generated a set of association rules (minimum support = 15%, minimum confidence = 70%) for each month from 2009 to 2011, using the genres of the movies released within a month. The genres associated with profit with highest support have been tagged as a profitable trend or trends. Note that we have achieved 100% confidence for 31 cases out of 36 cases and lowest confidence is found to be 80%.

We have considered a five-month time frame as a maximum lifetime or duration of a movie trend. We have used first three month's data to compute three best trends and validated these trends by next two month's trend. At first the entire data set have been divided into 12partitions. Each partition consists with the data of 3 consecutive months. Three best association rules according to their support have been derived from the first 3 month's data, and then these rules are validated by the next two month's business trend. Note that As you can see, the results from our previous experiments (Business trend determination) have been used here to validate our forecasted result.

In data mining, association rule learning is a popular and well-explored method for discovering interesting relations between variables in large databases [17]. It is intended to identify "strong rules" in databases using different measures of interestingness [18].By "strong rules" we mean the rules that satisfy both a minimum support and minimum confidence value. Based on the concept of strong rules, Rakesh Agrawal et al.[19] introduced association rules for discovering relations between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule $\{Onions, Potatoes\} \Longrightarrow \{Burger\}$ found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy hamburger meat. Such information can be used as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements [20]. In addition to the above example from market basket analysis, association rules are employed today in many application areas including Web usage mining [21, 22, 23 and 24], intrusion detection [25, 26, 27, 28, and 29], Continuous production [30] and bioinformatics [31, 32, 33, 34, and 35].

We have used Apriori algorithm [36] to determine association rules, which extract general trends in our movie database. It is a classic algorithm for frequent item set mining and association rule learning over transactional databases. In our case it proceeds by identifying the frequent individual profit making genres in the database and extending them to larger and larger genre sets as long as those genre sets appear sufficiently often in the database.

3. PROPOSED METHOD

As stated above we have used association rule mining approach to find profitable business trend in our movie database. Association rules are basically if/ then statement. We have used these rules to uncover relationships between movie genres. An association rule has two parts, an antecedent (if) and a consequent (then). An antecedent is an item found in the data [37]. A consequent is an item that is found in combination with the antecedent. In our case, combination of movie genres is an antecedent and the profit the movie makes is a consequent. As we have mentioned earlier we have measured the rule's interestingness according to its support and confidence.

We have applied apriori algorithm on each month's data to generate a set of rules. The rule having highest support and confidence value is issued to identify the profitable business trend for each month.

Let $\mathcal{I} = \{I_1, I_2, \dots, I_m\}$ be a set of items that are actually movie attributes (movie genres and profit) [38]. Let D, the task relevant data, be a set of database transactions where each transaction T is a set of items such that $T \subseteq \mathcal{I}$. Note that, in this case a movie is a transaction and D is the movie database. Let A be a set of items. A transaction T is said to contain A if and only if $A \subseteq T$. An association rule is an implication of the form $A \Rightarrow B$ where $A \subset \mathcal{I}, B \subset \mathcal{I}$, and $A \cap B = \phi$. In this case A is the genres and B is the profit. The rule $A \Rightarrow B$ holds in the movie database D with support s, where s is the percentage of movies that contain $A \cup B$. This is taken to be the probability $P(A \cup B)$. The rule $A \Rightarrow B$ has confidence c in D, where c is the percentage of movies in D containing A that also contain B. This is taken to be the conditional probability $P(B|A)$. That is,

$$support(A \Rightarrow B) = P(A \cup B) \quad (3.1)$$

$$confidence\ (A \Rightarrow B) = P(B|A) \quad (3.2)$$

Rules that satisfy both a minimum support threshold (*min_sup*) and a minimum confidence threshold (*min_conf*) are called "strong rules".

A set of items be called an itemset. If an itemset contains *k items* then it is a *k-itemset.* Like, the set *{Romance, comedy, Profit}* is a three item set. The occurrence frequency of an itemset is the number of transactions that contain the particular itemset. This is also referred to as the frequency, support count, or count of the itemset. Note that the itemset's support defined in Equation (3.1) is sometimes referred to as relative support, whereas the occurrence frequency is called the absolute support. If the relative support of an itemset *I* satisfies a prespecified minimum support threshold (i.e., the absolute support of *I* satisfies the corresponding minimum support count threshold), then *I* is a frequent itemset.3 The set of frequent *k*-itemsets is commonly denoted by L_k.

From equation (3.2), we have

$$confidence\ (A \Rightarrow B) = P(B|A) = \frac{support\ (A \cup B)}{support\ (A)} = \frac{support\_count\ (A \cup B)}{support\_count\ (A)} \quad (3.3)$$

Equation (3.3) shows that the confidence of the rule $A \Rightarrow B$ can be easily derived from the support counts of *A* and $A \cup B$.

***Algorithm 1**: To generate the set of Association rules*

Input:

- *D*, the movie database;
- *min_sup*, the minimum support count threshold.

Output:

- *L*, frequent itemsets in *D*.

Method:

1. L_1= find_frequent_1-itemsets (*D*);
2. for $(k = 2; L_{k-1} \neq \emptyset; k++)${
3. C_k = apriori_gen(L_{k-1}) ;
4. for each movie $t \in D$ {// scan *D* for Counts
5. C_t = subset (C_k, t); // get the subsets of *t* that are candidates
6. for each candidate $c \in C_t$
7. $c.count++$;
8. }
9. $L_k = \{ c \in C_k | c.count \geq min\_sup \}$
10. }
11. return $L = \cup_k L_k$;

Procedure apriori_gen (L_{k-1}: frequent$(k-1)$- itemsets)

1. for each itemset $l_1 \in L_{k-1}$
2. for each itemset $l_2 \in L_{k-1}$
3. if $(l_1[1] = l_2[1]) \wedge (l_1[2] = l_2[2]) \wedge \ldots \wedge (l_1[k-2] =$ *l2k−2⋀l1k−1 <l2k−1* then {
4. $c = l_1 \bowtie l_2$;//join step: generate candidates
5. *if* has_infrequent_subset (c, L_{k-1}) then
6. *delete c*;
7. *else add c to* C_k;
8. }
9. return C_k ;

Procedure has_infrequent_subset (c: *candidate k − genreset*);

L_{k-1}: *frequent* $(k-1)-$ *itemsets*);//use prior knowledge

1. for each (k-1)- subset *s* of *c*
2. if $s \notin L_{k-1}$ then
3. return TRUE;
4. return FALSE;

Apriori algorithm used *level-wise* iterative search approach. Here *k*-itemsets are used to explore *(k+1)* itemsets. At first the set of frequent-1 itemsets, L_1 is found by scanning the movie database. Next L_1 is used to find L_2 , the set of frequent 2 itemsets, which is used to find L_3, and so on, until no more frequent k-itemsets can be found[77]. If the itemsets *I* does not satisfy the minimum support threshold, *min_sup,* then *I* is not frequent; that is $P(I) < min\_sup$. If a genre *A* is added to itemsets *I,* then the resulting itemsets (i.e.$I \cup A$) cannot occur more frequently than *I.* Therefore, $I \cup A$ cannot be frequent either. That is, $P(I \cup A) < min\_sup$.

A two step process Joining and Pruning are used in the apriori algorithm.

The Joining step:

To find L_k , a set of candidate *k* – itemsets is created by joining L_{k-1} with itself [39]. This set of candidates is C_k. Let l_1, l_2 be itemsets in L_{k-1}.The join $L_{k-1} \bowtie L_{k-1}$ is performed where members of L_{k-1} are joinable if their first *(k-2)* items are in common. That is, members l_1 and l_2 of L_{k-1} are joined if $(l_1[1] = l_2[1]) \wedge (l_1[2] = l_2[2] \wedge .. \wedge (l_1[k-2] = l_2[k-2]) \wedge (l_1[k-1] = l_2[k-1]))$. The condition $l_1[k-1] < l_2[k-1]$ simply ensures that no duplicates are generated. The resulting itemset formed by joining l_1, l_2 is $l_1[1], l_1[2], \ldots, l_1[k-2], l_1[k-1], l_2[k-1]$.

The Pruning Step:

C_k is a superset of L_k, that means its members may not be frequent, but all of the frequent *k* – itemsets are included in C_k [39]. A scan of the movie database to determine the count of each candidate in C_k would result in the determination of L_k . To reduce the huge size of C_k apriori used as follows. Any *(k-1)* subset that is not frequent cannot be a subset of a frequent *k*-itemset . Hence, if any *(k-1)* subset of a candidate k-itemset is not in L_{k-1}, then the candidate cannot be frequent either and so can be removed from C_k.

Strong association rules (association rules that satisfy both minimum support and minimum confidence criteria) can be found from the frequent itemset generated from the movie database, D. Confidence can be found by the Equation (3.3).

Based on this equation, association rules can be generated as follows:

- For each frequent itemset *l*, generate all nonempty subsets of *l*.
- For every nonempty subset *s* of *l*, output the rule "$s \Rightarrow (l-s)$" *if* $\frac{support\_count\ (l)}{support\_count\ (s)} \geq min\_conf$, where min_conf is the minimum confidence threshold.

Note that the rules are generated from frequent itemsets so each one automatically satisfies minimum support.

Algorithm 2: *To predict profitable business trends for a particular month*

Input:

- $D_0, D_1, D_2, D_3, \ldots, D_{10}$, Film's data base containing 3 month's data.
- min_sup, the minimum support count threshold.

Output:

- R the business trends forecast

Method:

1. $L_i =$ association rules;
2. for $(i = 0; i < 11; i++)\{$
3. $L_i = association\_rule(D_i, \min\_sup)$;
4. $R_i = best\_3\_rule(L_i)$;
5. }
6. return $R = \cup_i R_i$;

Procedure $association\_rule\ (D, \min\_sup)$

1. for each dataset $D\{$
2. apriori $(D, \min\_sup)$;
3. }

Procedure $best\_3\_rule(L)$;

1. $l_i \in L$;
2. $descending\_sort\_support\_and\_confidence(L)$;
3. Select first 3 rules l_1, l_2, l_3;

First of all we have extracted 3 best association rules from the first three month's data. These rules are then used to predict the business trend for the next two months. In this way we have predicted the trend for the eleven data partitions, which means twenty-two month's business trend. We have validated these results by the results of previous experiment (Algorithm 1).

4. EXPERIMENTAL RESULT

We have used 395 movie's data released from 2009 to 2011. The data attributes are USA Release date, 20 movie genres (like Action, Adventure, Animation, Biography, Comedy, Crime, Documentary, Drama, Family, Fantasy, History, Horror, Musical, Mystery, Romance, Science Fiction, Sport, Thriller, War, Western) and Profit percentage. Except US Release date and profit percentage, all attributes' values are either "y" or "n". Movies with less than 0% profit value, which is denoted by "n", are removed from the database. USA release date helps us to group the movies, which are released in same month. After month-wise grouping, we have removed this attribute. After preprocessing the database it is looked like as shown in Table 1.

Table 1. Snapshot of the data used for trend determination

Action	Adventure	Fantasy	.	.	.	Western	Profit
y	y	n	.	.	.	n	y
y	n	n	.	.	.	n	y
n	n	y	.	.	.	y	n
n	y	y	.	.	.	n	y

We have carried out our experiment using SPSS Clementine 11.1 Software. Minimum support and confidence are set to 15% and 70 % respectively. Maximum number of antecedents is set to three, because we have not found any film that belongs to more than three genres.

As we have mentioned in Section II, at first we have generated a set of association rules (minimum support = 15%, minimum confidence: 70%) for each month from 2009 to 2011, using the genres of the movies released within a month. The genres with highest support have been tagged as a profitable trend or trends. We have provided the results in Table 2.One can see from the table that, we have achieved 100% confidence level for 31 cases out of 36 cases and lowest confidence level is 80%.

Table 2. Month wise trend for the successful movies

Month	Trend		
	Antecedent	**Support**	**Confidence**
Jan 09	Comedy=Y	57.14	100.00
Feb 09	Action = Y	40.00	100.00
Mar 09	Action = Y	57.14	100.00
Apr 09	Drama = Y	80.00	100.00
May 09	Action = Y	66.67	100.00
Jun 09	Drama = Y	41.67	80.00
Jul 09	Adventure = Y	40.00	100.00
Aug 09	Thriller/Drama = Y	27.27	100.00
Sep 09	Comedy=Y	38.46	100.00
Oct 09	Thriller = Y	20.00	100.00
Nov 09	Comedy=Y	35.29	100.00
Dec 09	Action/ Adventure/Fantasy = Y	21.43	100.00
Jan 10	Action = Y	60.00	100.00
Feb 10	Action = Y	37.50	100.00
Mar 10	Comedy = Y	58.33	100.00
Apr 10	Comedy = Y	62.50	100.00
May 10	Action = Y	66.67	100.00
Jun 10	Comedy = Y	45.45	100.00
Jul 10	Action/ Adventure = Y	40.00	100.00

Aug 10	Crime = Y	30.00	100.00
Sep 10	Comedy = Y	38.46	100.00
Oct 10	Action = Y	16.67	100.00
Nov 10	Drama = Y	81.81	100.00
Dec 10	Drama = Y	58.33	85.71
Jan 11	Comedy = Y	42.86	100.00
Feb 11	Drama = Y	36.36	100.00
Mar 11	Science Fiction = Y	33.33	100.00
Apr 11	Thriller = Y	21.05	100.00
May 11	Comedy = Y	41.67	80.00
Jun 11	Comedy = Y	54.55	83.33
Jul 11	Comedy = Y	53.85	100.00
Aug 11	Drama = Y	42.86	100.00
Sep 11	Comedy = Y	16.67	100.00
Oct 11	Action = Y	21.43	100.00
Nov 11	Comedy = Y	70.00	100.00
Dec 11	Comedy = Y	54.55	83.330

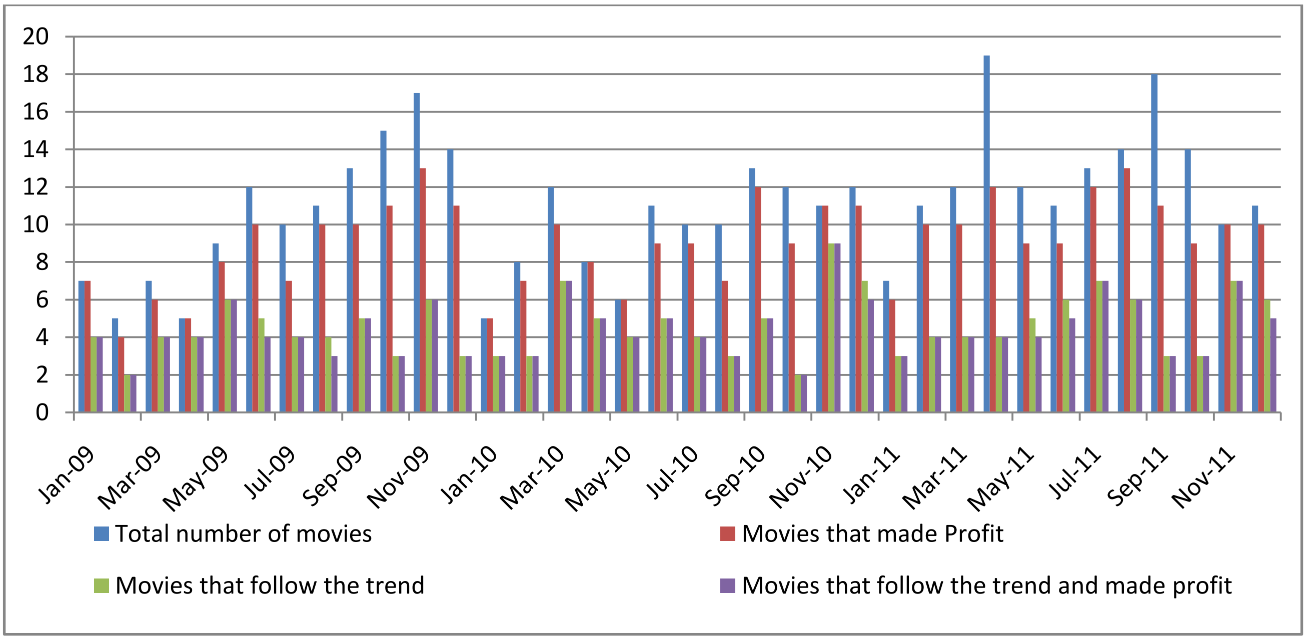

Fig 1: Total number of movies, trend following movies and trend following successful movies

Let us consider a scenario where twenty movies released in a particular month and fifteen out of twenty movies made profit. Among these twenty movies ten movies belong to romance-comedy genres, five movies belong to action genre and remaining five movies belong to thriller genre.Nineout of ten romance-comedy movies made profit. Clearly in this month viewers preferred to watch romantic comedy movies. Alternatively we can say that this month's trend was romance-comedy.

Let's relate this scenario to our chart shown in figure 1. Values of the four parameters we have used in this charts are like following: Total number of movies: 20

Movies that made Profit: 15

Movies that follow the trend: 10

Movies that follow the trend and made profit: 9

We have defined two ratios to validate our findings about the profit making movie business trends. *Ratio1* is the ratio between Movies that made profit to Total number of movies. *Ratio2* is the ratio between Movies that follow the trend and made profit to Movies that follow the trend. For the above case Ratio1 is .75 and Ratio2 is 0.9.

We have computed these two ratios from 36 months data, which we have analyzed earlier and plotted the frequency distribution, cumulative percentages in figure 2 and figure 3 respectively.

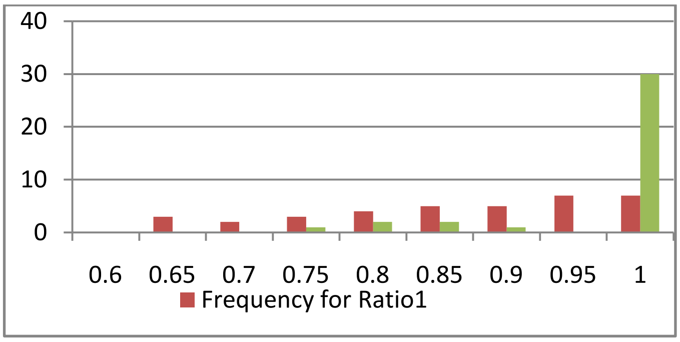

Fig 2: Frequency Distribution

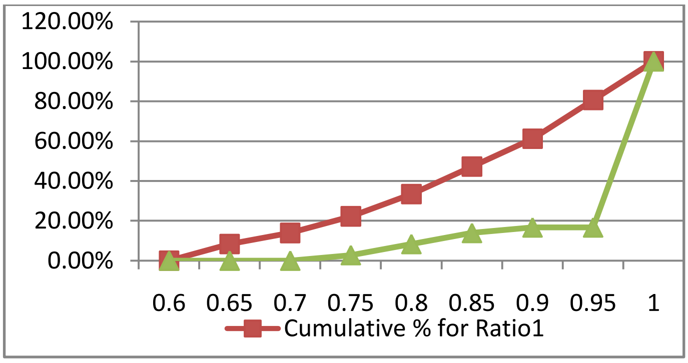

Fig3: Cumulative percentage

As we have mentioned in Section II, we have assumed 5 month as a maximum span or lifetime of the movie business's trend or trends. First three months data has been used to compute three best trends and validated these trends by next two month's trends.

As you can see from the table 3, we have successfully predicted trend/ trends for 19 cases out of 22 cases. This leads us to a success rate of 86.36%.It can be seen from the result of our previous experiment shown in table 1, how frequently business trend changes.

Considering such a volatile nature of movie business trend a success rate of 86.36% is seems to be a reasonably good result. It is obvious that such a prediction should be very helpful to Distributors, Producers and filmmakers. If a film follows the trend then it will be safe to say that it has a very good chance of being successful.

5. CONCLUSION AND SCOPE FOR FURTHER WORK

Note that, it is very difficult even for a human 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. But we can certainly try to predict the possible movie trend based on the genres. These trends for future months can be treated as guidelines for the production houses to drive this business towards more profitability. In this paper we have attempted to develop a method using association rule mining to find out the most preferable and popular genres which can be represented as movie business's trend. In our experiment we have achieved a success rate of 86.36% which is a reasonably good result considering the very high unpredictability of movie business.

Further research work can be conducted in the following areas. We can search for more genres and/or division of an existing genre into subgenres that may led to a higher success rate of prediction. Also we can conduct research for finding features of profitable movie trend in terms of some specific genres possibly using a psychological analysis of peoples' likings or interests in some specific kinds of movies.

6. REFERENCES

[1] Information about the films from Wikipedia, http://en.wikipedia.org/wiki/Film [Accessed 9 January 2013]

[2] Information about the film industry from Wikipedia, http://en.wikipedia.org/wiki/Film_industry [Accessed 9 January 2013]

[3] Nash Information Services, The numbers, "Box office data, movie stars, idle speculation", http://www.the-numbers.com/glossary.php [Accessed 25 January 2013]

[4] Information about most expensive films from Wikipedia, http://en.wikipedia.org/wiki/List_of_most_expensive_films [Accessed 9 January 2013]

[5] Information about highest grossing films from Wikipedia, http://en.wikipedia.org/wiki/List_of_highest-grossing_films [Accessed 9 January 2013]

[6] Information about the film "City Island" http://en.wikipedia.org/wiki/City_Island_%28film%29 [Accessed 10 January 2013]

[7] Information about the film "Zyzzyx Road" http://en.wikipedia.org/wiki/Zyzzyx_Road [Accessed 10 January 2013]

[8] "Association Rule Mining: Applications in Various Areas" by AkashRajak and Mahendra Kumar Gupta, Proceedings of International Conference on Data Management (ISBN: 0230-63469-9, Macmillan India Ltd., New Delhi), Ghaziabad, India, pp. 3-7, February 25-26, 2008.

[9] "The Application of Association Rules in Retail Marketing Mix " by Hongwei Liu, Bin Su and Bixi Zhang Proceedings of IEEE International Conference on Automation and Logistics, 2007, page No. 2514 - 2517.

[10] G. Serban, I. G. Czibula, and A. Campan, "A Programming Interface For Medical diagnosis Prediction", StudiaUniversitatis, "Babes-Bolyai", Informatica, LI(1), pages 21-30, 2006.

[11] N. Gupta, N. Mangal, K. Tiwari and P. Mitra, "Mining Quantitative Association Rules in Protein Sequences", In Proceedings of Australasian Conference on Knowledge Discovery and Data Mining – AUSDM, 2006

[12] "Searching customer patterns of mobile service using clustering and quantitative association rule" by So Young Sohn and Yoonseong Kim published in Expert Systems with Applications Volume 34 (2008), Page No. 1070–1077

[13] "Association rules applied to credit card fraud detection" by D. Sa´nchez , M.A. Vila, L. Cerda and J.M. Serrano. Published in Expert Systems with Applications Volume 36 (2009), Page No. 3630–3640

[14] "Mining spatial association rules in census data" by D. Malerba, F. Esposito and F.A. Lisi, In Proceedings of Joint Conf. on "New Techniques and Technologies for Statistcs and Exchange of Technology and Know-how", 2001.

[15] Information about the film genre from Wikipedia, http://en.wikipedia.org/wiki/Film_genre [Accessed 10 January 2013]

[16] Information about the film "Avtar" from IMDB http://www.imdb.com/title/tt0499549/ [Accessed 9 January 2013]

[17] Definitions of Association rule from Wikipedia http://en.wikipedia.org/wiki/Association_rule_learning [Accessed 12 January 2013]

[18] Piatetsky-Shapiro, Gregory (1991), *Discovery, analysis, and presentation of strong rules*, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds.,*Knowledge Discovery in Databases*, AAAI/MIT Press, Cambridge, MA.

[19] Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*. pp. 207.

[20] IstiSurjandaridanAnnury Citra Seruni, Design of product placement layout in retail shop using market basket analysis, MAKARA, TEKNOLOGI, VOL. 9, NO. 2, NOvEMBER 2005, pp 43-47

[21] BamshadMobasher, Honghua Dai, Tao Luo and Miki Nakagawa , " Personalization based on association rule discovery from web usage data ", WIDM '01 Proceedings of the 3rd international workshop on Web information and data management, pp 9-15

[22] Raymond Kosala, HendrikBlockeel, " Web mining research: a survey ", ACM SIGKDD Explorations Newsletter, Volume 2 Issue 1, June, 2000, pp 1-15

[23] JaideepSrivastava, Robert Cooley, MukundDeshpande and Pang-Ning Tan, "Web usage mining: discovery and applications of usage patterns from Web data", ACM SIGKDD Explorations Newsletter,Volume 1 Issue 2, January 2000, pp 12-23.

[24] R Cooley, "Web mining: information and pattern discovery on the World Wide Web ", Proceedings of Ninth IEEE International Conference on Tools with Artificial Intelligence, 1997. pp 558-567.

[25] Lee, Wenke, and Salvatore J. Stolfo. *Data mining approaches for intrusion detection*. Defense Technical Information Center, 2000.

[26] Lee, Wenke, Salvatore J. Stolfo, and Kui W. Mok. "Mining audit data to build intrusion detection models." In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining*, pp. 66-72. 1998.

[27] Tajbakhsh, Arman, Mohammad Rahmati, and AbdolrezaMirzaei. "Intrusion detection using fuzzy association rules." *Applied Soft Computing* 9, no. 2 (2009): 462-469.

[28] Lee, Wenke, Salvatore J. Stolfo, and Kui W. Mok. "Adaptive intrusion detection: A data mining approach." *Artificial Intelligence Review* 14, no. 6 (2000): 533-567.

[29] Ertoz, Levent, Eric Eilertson, AleksandarLazarevic, Pang-Ning Tan, Vipin Kumar, JaideepSrivastava, and Paul Dokas. "Minds-minnesota intrusion detection system." *Next Generation Data Mining* (2004): 199-218.

[30] Raddatz, M., M. Schlüter, S. C. Brandt, M. Jarke, T. Grimbach, and M. Weck. "Identification and reuse of experience knowledge in continuous production processes." In *Proceedings of the 9th IFAC Symposium on automated systems based on human skill and knowledge, France: Nancy*. 2006.

[31] Creighton, Chad, and Samir Hanash. "Mining gene expression databases for association rules." *Bioinformatics* 19, no. 1 (2003): 79-86.

[32] Georgii, Elisabeth, Lothar Richter, Ulrich Rückert, and Stefan Kramer. "Analyzing microarray data using quantitative association rules." *Bioinformatics* 21, no. suppl 2 (2005): ii123-ii129.

[33] Carmona-Saez, Pedro, Monica Chagoyen, Andres Rodriguez, OswaldoTrelles, Jose Carazo, and Alberto Pascual-Montano. "Integrated analysis of gene expression by association rules discovery." *BMC bioinformatics* 7, no. 1 (2006): 54.

[34] Ordonez, Carlos, Norberto Ezquerra, and Cesar A. Santana. "Constraining and summarizing association rules in medical data." *Knowledge and Information Systems* 9, no. 3 (2006): 1-2.

[35] Ordonez, Carlos, Norberto Ezquerra, and Cesar A. Santana. "Constraining and summarizing association rules in medical data." *Knowledge and Information Systems* 9, no. 3 (2006): 1-2.

[36] RakeshAgrawal and RamakrishnanSrikant, "Fast algorithms for mining association rules in large databases". Proceedings of the *20th International Conference on Very Large Data Bases, VLDB*, pages 487-499, Santiago, Chile, September 1994.

[37] Association rules in Data Mining http://searchbusinessanalytics.techtarget.com/definition/association-rules-in-data-mining [Accessed on 25 January 2013]

[38] Han, Jiawei, and Micheline Kamber. "Data mining: concepts and techniques (the Morgan Kaufmann Series in data management systems)." Second edition, pp- 230-240.

[39] Han, Jiawei, and Micheline Kamber. "Data mining: concepts and techniques (the Morgan Kaufmann Series in data management systems)." Second edition, pp- 235

Table 3. Predicted Movie business trend

Training Data Set	Best 3 Trends		No.	Test Month	Best Trend of the Test Month		Result
	Antecedent	Support (%)			Antecedent	Support (%)	
Jan – Mar 09	Comedy= Y	35.71	1	Apr 09	Drama = Y	80.00	Success
	Drama = Y	21.43	2	May 09	Action = Y	66.67	Success
	Action = Y	21.43					
Apr – Jun 09	Action = Y	50.00	3	Jul 09	Adventure = Y	40.00	Success
	Drama = Y	35.00	4	Aug 09	Thriller = Y/ Drama = Y	27.27	Success
	Adventure = Y	35.00					
Jul – Sep 09	Comedy = Y	28.00	5	Oct 09	Thriller = Y	20.00	Success
	Thriller = Y	28.00	6	Nov 09	Comedy = Y	35.29	Success
	Drama = Y	28.00					
Oct – Dec 09	Drama = Y	42.86	7	Jan 10	Action = Y	50.00	Failure
	Comedy = Y	32.14	8	Feb 10	Action = Y	37.50	Failure
	Adventure = Y	25.00					
Jan – Mar 10	Comedy = Y	45.00	9	Apr 10	Comedy = Y	55.56	Success
	Action = Y	30.00					

	Romance = Y	30.00
Apr – Jun 10	Comedy = Y	45.00
	Action = Y	40.00
	Adventure = Y	40.00
Jul – Sep 10	Comedy = Y	34.62
	Drama = Y	30.77
	Action = Y	30.77
Oct – Dec 10	Drama = Y	57.14
	Comedy = Y	39.29
	Adventure = Y	25.00
Jan – Mar 11	Comedy = Y	39.13
	Thriller = Y	30.44
	Drama = Y	26.09
Apr – Jun 11	Comedy = Y	44.44
	Action = Y	29.63
	Drama = Y	22.22
Jul – Sep 11	Comedy = Y	36.67
	Romance = Y	23.33
	Action = Y	16.67

10	May 10	Action = Y	57.14	Success
11	Jul 10	Action = y/ Adventure = Y	40.00	Success
12	Aug 10	Crime = Y	30.00	Failure
13	Oct 10	Action = Y	16.67	Success
14	Nov 10	Drama = Y	75.00	Success
15	Jan 11	Comedy = Y	42.86	Success
16	Feb 11	Drama = Y	36.36	Success
17	Apr 11	Thriller = Y	21.05	Success
18	May 11	Comedy = Y	41.67	Success
19	Jul 11	Comedy = Y	53.85	Success
20	Aug 11	Drama = Y	42.86	Success
21	Oct 11	Action = Y	21.43	Success
22	Nov 11	Comedy = Y	70.00	Success