Volume 3, No. 3, May-June 2012

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Better Game Design using Association Analysis

Cajetan Rodrigues
Department of Information Technology
St.Francis Institute of Technology
Mumbai-400103, India
r_cajetan@yahoo.com

Nazneen Ansari
Department of Information Technology
St.Francis Institute of Technology
Mumbai-400103, India
nazsfit@yahoo.com

Abstract: Decades of computer game playing has resulted in the generation of vast amounts of data in the form of players' preferences and habits of game-playing. This gives an opportunity to game designers to extract knowledge from such data and leverage that knowledge to design games that players would be extremely interested in. Better Game Design using Association Analysis is a new application of data mining, applied to the art of game design. This paper aims to determine whether Association Analysis applied to a player database would provide game designers with meaningful rules that would help improve the design of the game. A data set of game-players' habits and preferences is collected from St. Francis Institute of Technology, Mumbai, and subjected to the Apriori algorithm. The database contains various aspects of game design that users would like to experience in the games they play. The rules generated from association analysis would be of tremendous benefit to the Game Design industry, as they can then use them to design profitable games.

Keywords: Association analysis, Apriori algorithm, Game design, Strong rules.

I. INTRODUCTION

Data mining has been used to analyse large data sets and establish useful classification and patterns in the datasets. Data mining software applications includes various methodologies that have been developed by both commercial and research centres. These techniques have been used for industrial, commercial and scientific purposes [1].

Game design, a subset of game development, is the process of designing the content and rules of a game in the pre-production stage and design of game-play, environment, storyline, and characters during production stage. Better game design using Association Analysis is a new data mining approach that will aid game designers to know exactly what players want and expect from computer games. This knowledge in turn will enable them to take informed and strategic decisions while designing various aspects of the game. There is widespread consensus that games motivate players to spend time on task mastering the skills a game imparts. Nevertheless, the literature reveals that a number of distinct design elements, such as narrative context, rules, goals, rewards, multi-sensory cues and interactivity, seem necessary to stimulate players' interest in the game [2].

The standard approaches to solve the problem of finding and using the right ingredients to build up a great game, consist of one (or a mixture) of the following:

User polls and surveys
Gaming Forums
Market Research

And sometimes even wild guesses, with the hope that the game will turn out to be likeable.

This paper suggests a new approach using Association Analysis to study gamer activities, and identify frequent item-sets, such as popular weapons, popular roles, popular maps etc. Such item-sets can give us strong association rules. These rules can be translated by the game designer into knowledge, who can then use the same knowledge to enrich existing games, or create a new one that has a high chance of becoming a gamer favourite.

II. ASSOCIATION MINING

Association rule mining, one of the most important and well researched techniques of data mining, was first introduced by Agrawal, et al [3]. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories. Association rules are widely used in various areas such as telecommunication networks, market and risk management, inventory control etc.

Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a given database. The problem is usually decomposed into two sub-problems. One is to find those item-sets whose occurrences exceed a predefined threshold in the database; those item-sets are called frequent or large item-sets. The second problem is to generate association rules from those large item-sets with the constraints of minimal confidence. Suppose one of the large item-sets is Lk, Lk = {I1, I2…, Ik}, association rules with this item-set are generated in the following way:

The first rule is {I1, I2, … , Ik-1} ∩ {Ik}, by checking the confidence this rule can be determined as interesting or not. Then other rules are generated by deleting the last items in the antecedent and inserting it to the consequent. Further the confidences of the new rules are checked to determine the interestingness of them.

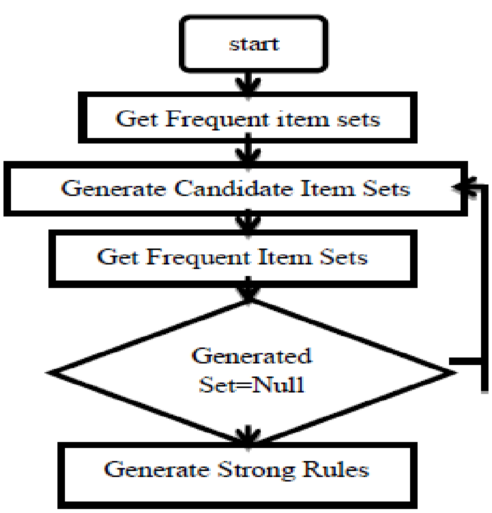
Those processes are iterated until the antecedent becomes empty. Since the second sub-problem is quite straightforward, most of the researches focus on the first sub-problem. The first sub-problem can be further divided into two sub-problems: candidate large item-sets generation process and frequent item-sets generation process. We call those item-sets whose support exceeds the support threshold as large or frequent item-sets. Those item-sets that are expected or have the hope to be large or frequent are called candidate item-sets. In many cases, the algorithms generate an extremely large number of association rules, often in thousands or even millions. Further, the association rules are sometimes very large. It is nearly impossible for the end users to comprehend or validate such large number of complex association rules, thereby limiting the usefulness of the data mining results. Several strategies have been proposed to reduce the number of association rules, such as generating only "interesting" rules, generating only "non-redundant" rules, or generating only those rules satisfying certain other criteria such as coverage, leverage, lift or strength [4].

III. APRIORI

Apriori uses a complete, bottom-up search with a horizontal layout and enumerates all frequent item sets [5]. An iterative algorithm, Apriori counts item sets of a specific length in a given database pass. The main property of Apriori algorithm is that all non-empty subsets of a frequent item set must also be frequent.

```
C_k: Candidate item set of size k
L_k : frequent item set of size k
L_1 = {frequent items};
for (k = 1; L_k !=∅; k++) do begin
    C_k+1 = candidates generated from L_k;
   for each transaction t in database do
  increment the count of all candidates in C_k+1 that
are contained in t
   L_k+1 = candidates in C_k+1 with min_support
   end
return ∪_k L_k;
```

Algorithm [5]



Flowchart [5]

IV. ASSOCIATION MINING FOR DESIGNING GAMES

A survey was conducted in the college and we collected data from 100 students. The survey questions were based on different types of games played by them, different platforms used, game preferences, preferred game facilities/features, etc. A dataset (game.arff) was created in WEKA tool which consisted of 18 attributes and 100 records.

We loaded the data set into WEKA, performed a series of operations using WEKA's attribute and discretisation filters, and then performed association rule mining on the resulting data set.

WEKA allows the resulting rules to be sorted according to different metrics such as confidence, leverage, and lift. In this example, we have selected lift as the criteria. Furthermore, we have entered 1.5 as the minimum value for lift (or improvement). Lift is computed as the confidence of the rule divided by the support of the right-hand-side (RHS). In a simplified form, given a rule L => R, lift is the ratio of the probability that L and R occur together to the multiple of the two individual probabilities for L and R, i.e.,

lift = Pr(L,R) / Pr(L).Pr(R).

If this value is 1, then L and R are independent. The higher this value, the more likely that the existence of L and R together in a transaction is not just a random occurrence, but due to some relationship between them. Here we also change the default value of rules (10) to be 100; this indicates that the program will report no more than the top 100 rules (in this case sorted according to their lift values). The upper bound for minimum support is set to 1.0 (100%) and the lower bound to 0.1 (10%). Apriori in WEKA starts with the upper bound support and incrementally decreases support (by delta increments which by default are set to 0.05 or 5%). The algorithm halts when either the specified number of rules are generated, or the lower bound for minimum support is reached.

Results of association analysis of "game.arff" by WEKA tool:

```
=== Run information ===
 Scheme:       weka.associations.Apriori -N 20 -T 1 -C 1.5 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:      gamesurvey11-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-R2-weka.filters.unsupervised.attribute.Remove-R9,10
Instances:    100
Attributes:   16
              Gender
              Start_age
              Platform preference
              Game preference
              Hrs/week
              Skill Level
              Frequency Increase
              Most Exp. Game
              Online Play
              Story Influence
              Customization
```


CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

Chat/ Message
Social Interaction
Game Play Feature
Boss Preference
Weapon Preference

=== Associator model (full training set) ===
Apriori
=======
Minimum support: 0.2 (20 instances)
Minimum metric <lift>: 1.5
Number of cycles performed: 16
Generated sets of large item-sets:
Size of set of large item-sets L(1): 29
Size of set of large item-sets L(2): 119
Size of set of large item-sets L(3): 150
Size of set of large item-sets L(4): 115
Size of set of large item-sets L(5): 35
Size of set of large item-sets L(6): 4
Best rules found:

a. Gender=m Story Influence=yes Customization=yes 44 ==> Chat/ Message=yes Weapon Preference=gun 20 conf:(0.45) < lift:(1.75)> lev:(0.09) [8] conv:(1.3)
b. Chat/ Message=yes Weapon Preference=gun 26 ==> Gender=m Story Influence=yes Customization=yes 20 conf:(0.77) < lift:(1.75)> lev:(0.09) [8] conv:(2.08)
c. Story Influence=yes Customization=yes 56 ==> Gender=m Chat/ Message=yes Weapon Preference=gun 20 conf:(0.36) < lift:(1.62)> lev:(0.08) [7] conv:(1.18)
d. Gender=m Chat/ Message=yes Weapon Preference=gun 22 ==> Story Influence=yes Customization=yes 20 conf:(0.91) < lift:(1.62)> lev:(0.08) [7] conv:(3.23)
e. Platform preference=desktop Chat/ Message=yes 35 ==> Gender=m Customization=yes Social Interaction=yes 21 conf:(0.6) < lift:(1.62)> lev:(0.08) [8] conv:(1.47)
f. Gender=m Customization=yes Social Interaction=yes 37 ==> Platform preference=desktop Chat/ Message=yes 21 conf:(0.57) < lift:(1.62)> lev:(0.08) [8] conv:(1.41)
g. Gender=m Platform preference=desktop Story Influence=yes 46 ==> Chat/ Message=yes Social Interaction=yes 23 conf:(0.5) < lift:(1.61)> lev:(0.09) [8] conv:(1.32)
h. Chat/ Message=yes Social Interaction=yes 31 ==> Gender=m Platform preference=desktop Story Influence=yes 23 conf:(0.74) < lift:(1.61)> lev:(0.09) [8] conv:(1.86)
i. Platform preference=desktop Frequency Increase=yes 29 ==> Gender=m Story Influence=yes Social Interaction=yes 20 conf:(0.69) < lift:(1.6)> lev:(0.08) [7] conv:(1.65)
j. Frequency Increase=yes Story Influence=yes 29 ==> Gender=m Platform preference=desktop Social Interaction=yes 20 conf:(0.69) < lift:(1.6)> lev:(0.08) [7] conv:(1.65)
k. Gender=m Platform preference=desktop Social Interaction=yes 43 ==> Frequency Increase=yes Story Influence=yes 20 conf:(0.47) < lift:(1.6)> lev:(0.08) [7] conv:(1.27)
l. Gender=m Story Influence=yes Social Interaction=yes 43 ==> Platform preference=desktop Frequency Increase=yes 20 conf:(0.47) < lift:(1.6)> lev:(0.08) [7] conv:(1.27)
m. Platform preference=desktop Frequency Increase=yes Story Influence=yes 24 ==> Gender=m Social Interaction=yes 20 conf:(0.83) < lift:(1.6)> lev:(0.08) [7] conv:(2.3)
n. Gender=m Social Interaction=yes 52 ==> Platform preference=desktop Frequency Increase=yes Story Influence=yes 20 conf:(0.38) < lift:(1.6)> lev:(0.08) [7] conv:(1.2)
o. Chat/ Message=no 57 ==> Skill Level=novice Weapon Preference=gun 21 conf:(0.37) < lift:(1.6)> lev:(0.08) [7] conv:(1.19)
p. Skill Level=novice Weapon Preference=gun 23 ==> Chat/ Message=no 21 conf:(0.91) < lift:(1.6)> lev:(0.08) [7] conv:(3.3)
q. Chat/ Message=yes 40 ==> Gender=m Platform preference=desktop Story Influence=yes Social Interaction=yes 23 conf:(0.57) < lift:(1.6)> lev:(0.09) [8] conv:(1.42)
r. Gender=m Platform preference=desktop Story Influence=yes Social Interaction=yes 36 ==> Chat/ Message=yes 23 conf:(0.64) < lift:(1.6)> lev:(0.09) [8] conv:(1.54)
s. Chat/ Message=yes 40 ==> Gender=m Platform preference=desktop Customization=yes Social Interaction=yes 21 conf:(0.53) < lift:(1.59)> lev:(0.08) [7] conv:(1.34)
t. Gender=m Platform preference=desktop Customization=yes Social Interaction=yes 33 ==> Chat/ Message=yes 21 conf:(0.64) < lift:(1.59)> lev:(0.08) [7] conv:(1.52)

From the rules, the designer gets a fair picture of what the gamer really wants. In order to create a game that not only will capture the attention of the gamer, but also will sell comsiderably well, the game designer must base his design on these generated rules. These association rules guide the designer in making games specially for the target gaming community. For example, rule number 17 says:

Chat/ Message=yes ==> Gender=m Platform preference=desktop Story Influence=yes Social Interaction=yes

This means that if you provide an in-game chat facility to communicate with other players, the game will mostly be preferred to be played on a desktop, by males, and they would be looking for a good story-line in the game, as well as the ability to work in a team. Therefore to make more profits out of the game, the designer must include those particular features in the game. Gamers who find these desirable features would remain interested in it, and the "Stickiness" factor of the game may also increase.

Rule number 4 says:

Gender=m Chat/ Message=yes Weapon Preference=gun
==> Story Influence=yes Customization=yes

Thus if the designer gives an in-game chat facility to males who prefer guns, they would also be interested in the story and would want the facility to customize their own characters in the game. This would make their gaming experience much

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

more enjoyable, and the game designer can expect them to remain loyal to the game, in turn increasing his/her profits.

Such rules are subjective, since they depend on the target group of gamers from whom the data is collected. Thus, in China gamers may prefer "Dragon" monsters, while in India, the focus may not be as much on the Bosses of the game as on the variety of weapons available. Therefore, the designer or any person who collects data must first decide the correct data to collect, and from whom. The correct framing of the survey questions (or any other method employed for data-collection) is very important. They must reflect those aspects of the game that will maximize the fun factor and also the profits for the creators of the game. Depending on all these, the knowledge reaped from different groups of people from different parts of the world will enable the designers to customize the game accordingly, in turn making profits for themselves, since the game will then have a large group of players.

V. CONCLUSION

This paper shows how Association Analysis could be used to extract knowledge from a gamer dataset to create strong rules that can guide the game design process. Using the strong rules in the design phase will enable the designers to make a successful game and tap into a wide pool of gamers, thereby generating commensurate revenue.

VI. ACKNOWLEDGMENT

Cajetan Rodrigues thanks all his professors at SFIT for their encouragement and help in writing technical papers. Both authors thank the students of BE-IT for their participation in the survey for this paper.

VII. REFERENCES

[1]. Ramesh Vamanan and K. Ramar, "Classification of Agricultural Land Soils: A Data Mining Approach", International Journal on Computer Science and Engineering, ISSN : 0975-3397 Vol. 3 No. 1, pp 379-384, Jan 2011.

[2]. M. Dodlinger, "Educational Video Game Design: A Survey of the Literature", Journal of Applied Educational Technology, Volume 4, Number 1, pp 21-31, Spring/Summer 2007.

[3]. Agrawal, R., Imielinski, T., and Swami, A. N., "Mining association rules between sets of items in large databases",Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp 207-216.

[4]. Sotiris Kotsiantis, Dimitris Kanellopoulos," Association Rules Mining: A Recent Overview", GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), pp. 71-82, 2006.

[5]. T.Anu Radha, P. Lavanya, "Recent Trends in Parallel and Distributed Apriori Algorithm", International Journal of Engineering Research and Applications, Vol. 1, Issue 4, pp.1820-1822