Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 82 (2016) 72 – 79

Procedia Computer Science

Symposium on Data Mining Applications, SDMA2016, 30 March 2016, Riyadh, Saudi Arabia

Online Social Gaming and Social Networking Sites

Linah Aburahmah[a], Hajar AlRawi[b], Yamamah Izz[c], Liyakathunisa Syed[d]*

[a]*Software Engineering student at Prince Sultan University – College of Women, Riyadh, Saudi Arabia*
[b]*Computer Science student at Prince Sultan University – College of Women, Riyadh, Saudi Arabia*
[c]*Information Systems student at Prince Sultan University – College of Women, Riyadh, Saudi Arabia*
[d]*Assistant Professor at Prince Sultan University – College of Women, Riyadh, Saudi Arabia*

Abstract

Online social games are becoming a significant component in today's social media sites. The social networking sites environment has provided a platform for online games to develop and expand in the virtual medium. Users are now able to play games online, compare scores, and challenge each other among many other things. Due to the diverse user demographics of social media sites, various motivations to playing social games emerge. The need for this present research was to answer the question whether the integration of social games within social networking sites and apps have increased the likelihood of playing those games. Therefore, the main objective of this research is to predict from the fact that will be used to decide whether to include games (Adventure, Fighting, Design\Art, Virtual Life…etc.) in the social media sites that still have not been implemented (such as Twitter, Tumblr, etc.). Also, we want to discover whether the inclusion of social games has improved the services offered by social sites and whether social gaming has effects on human behavior with regards to socializing and interacting with others. The aim of this study was to examine the relation between social media apps and games, and whether the former has increased users participation in online games. The primary setting for the quantitative method of research study was the online gamers regardless their age, gender, and interests. The data were collected by distributing an online survey, and were analyzed using WEKA data mining tool. Two popular classification algorithms were used to predict the answer of this research question. The resulting data were compared and tested for their accuracy using different metrics.

Peer-review under responsibility of the Organizing Committee of SDMA2016

Keywords: Online games; Social sites; Data Mining; Random Forest; Neural Networks.

* E-mail address: linah.aburahmah@yahoo.com; hii93@hotmail.com; yamz_izz@hotmail.com; lsyed@psu.edu.sa

Peer-review under responsibility of the Organizing Committee of SDMA2016
doi:10.1016/j.procs.2016.04.011

1. Introduction

Social media sites have grown increasingly popular in the last few decades. Initially they were created as tools utilized in communication and connecting people, but in the recent years they have evolved enormously. Recently, a new service has emerged with the addition of social online games (Adventure, Design and Art, Fighting, Virtual Life, etc.) to these sites. Many research papers are aimed to find ways to improve online games[1, 2]. Researchers are interested in using data mining techniques on users of social media sites to find out ways to improve the games themselves and the site which hosts the game[3, 4]. In this paper, we want to uncover the following, whether people are more eager to playing games because of the social element in social media sites (comparing scores, challenging each other, etc.), and what motivates them to play.

The organization of the paper is as follows: Section 2 is a literature review that contains some related work that served as a reference for this study. In Section 3, we summarize the technique adapted to collect data necessary for this research. Section 4 presents the Data Mining Software Tool. In Section 5, we present the adapted Knowledge Discovery process. In section 6, we present the comparative analysis of different algorithms. Section 7 presents the Result Discussion and finally Section 8 presents the conclusion.

2. Literature Review

Few research papers were found that discuss the impact of social network sites on the behavior of the online gamers. Wender wrote a report entitled "Data Mining and Machine Learning with Computer Game Logs"[3], in this report, he explains the process of using machine learning algorithms to discover new data patterns of human behaviors in online games so game developers can use this knowledge derived to create artificial players that act in a way similar to the humans. Ansari, Talreja, and Desai[4] aimed in their research paper to analyze the online social games using association mining algorithms to provide the game designers with new data patterns that can help them in improving the design of online social games. In a research paper entitled "A Novel Approach for the Classification of Social Media Data using Decision Tree", the two authors Pillai and Oliver[5] used Naive Bayes algorithms to analyze the social network data for tweeter since it becomes one of the most popular social network sites and therefore, it generated a huge number of tweets every day. After applying the Naive Bayes, the authors applied a decision tree algorithm on the same data set to check the accuracy of the results generated from applying the Naive Bayes algorithm.

3. Dataset Description

The dataset used in this paper is generated from the answers to an online survey that was created in Google Forms and distributed to a sample of 118 online gamers through the Email, Facebook, and WhatsApp. 68 responders are females and the rest 50 are males who have various interests and backgrounds. There are multivariate attributes that were generated from the ten questions of the online survey as specified in Table 1 of Appendix.

The variables such as age and the time the respondent spent to play the games are of numeric type and their values are represented in ranges. The rest of the variables are of the type string.

4. Data Mining Software Tool

The mining tool that is used in this research is WEKA, which is a collection of machine learning algorithms that are used for mining the data[6]. This research depends on Weka tool for the following:

- Data pre-processing (data cleaning and reduction).
- Mining the data using classification algorithms.
- Evaluation of the model based on accuracy and the speed of the used algorithm.

5. Adapted Knowledge Discovery Process

5.1. Data Cleaning

Not all respondents responded to all of the survey questions, or provided inaccurate answers, especially in the questions that have the option 'other'. Some responders either left it blank or filled it with wrong value, which resulted in the dataset having many faults in attributes and missing values. To solve this issue and proceed with the research work flow, the missing values were replaced manually with the global constant “Unknown”, but the inaccurate values were replaced with the mode value (This value has been chosen by most of the responders). Another issue we faced in our data cleaning process was that WEKA did not accept all types of data, so we had to make our values nominal in the ARFF file in order to get more accurate results in WEKA

5.2. Classification Algorithms

Classification is one of the techniques that is used to predict categorical (discrete) data. In this research paper, two classification algorithms are applied to the data set which is the Random Forest and Neural Network algorithms. The class label is "TimeSpent" which indicates whether the time the users spent on social media sites increased, decreased, or remained the same with the addition of games on these sites which is the focus of this research.

5.2.1. Random Forest Algorithm

It is one of the most accurate classification algorithms that fall under the category of trees` algorithms. The main objective of applying this algorithm is to improve the prediction accuracy. It is called random forest because the idea behind this algorithm is to produce a group of de-correlated decision trees (de-correlation gives better accuracy because it is unusual for the same error to appear on different trees constructed from different samples). Each decision tree is constructed using different samples from original data set as shown in figure 1. [7]

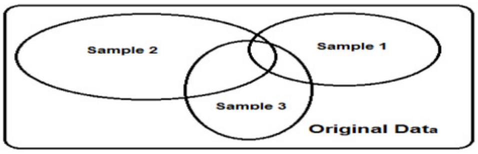

Fig. 1. How the samples are selected from the original dataset

5.2.2. Neural Network Algorithm

Also known as artificial neural network, neural networks algorithm is a highly connected information processing model inspired by the biological central nervous system. It consists of a large number of interconnected artificial neurons that exchange messages and process information in order to solve problems. Similar to our own nervous system, the neurons in a neural network are adaptive and engage in what is known as the learning phase. In this phase, neurons process internal and external information in the network, then change their structure or “learn” according to the obtained results.[8]

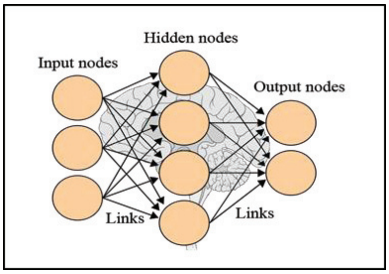

Fig. 2. Basic structure of the Neural Network. [9]

6. Comparative Analysis

6.1. Result of Random Forest

By applying this algorithm on the data set, Random forest of 100 trees were constructed and each considering 4 random features. The percentage of correctly classified instances indicated the level of accuracy for the applied algorithm. Since the percentage is 97%, it can be proved that the result is accurate. The classification results are as following:

Correctly Classified Instances	115	97.4576 %
Incorrectly Classified Instances	3	2.5424 %

Time taken to build the model is 0.06 s which indicated a high speed. The screenshot of the result below emphasized on the accuracy of each value of the class label:

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0.012	0.972	1	0.986	1	Increased
	0.941	0.01	0.941	0.941	0.941	0.999	Decreased
	0.97	0.019	0.985	0.97	0.977	0.999	Same
Weighted Avg.	0.975	0.016	0.975	0.975	0.975	0.999	

Fig. 3. Result from Applying Random Forest Algorithm

The result shows that the "Same" value of the class label is the most accurate one.

6.1.1. Evaluation methodology for Random Forest Algorithm

The random forest algorithm was initially selected to be applied on the current paper, after a wide comparison between it and other classification algorithms used in a research paper entitled "Comparison on Performance of Data Mining Algorithms in Classification of Social Network Data".[10] This research paper used social network data similar to the data used in our social gaming research. The classification algorithms that used are K-Nearest Neighbor (KNN) Algorithm, ID3 (Iterative Dichotomiser 3) Algorithm, C4.5 Algorithm, and Rnd Tree (Random Forest). The error rate after applying each algorithm was generated as shown in the table 2 below:

Table 1: Summary of the Results of Applying Different Decision tree Algorithms [10]

Classification Algorithm	Error rates	
	Face book	Twitter
C4.5	0.0860	0.1042
C-RT	0.1798	0.1976
CS-CRT	0.1798	0.1976
CS-MC4	0.1099	0.1107
ID3	0.1650	0.2084
KNN	0.1871	0.2097
Rnd Tree	**0.0004**	**0.0004**

According to the table, it is clear that random forest algorithm is the most accurate since it has the least error rate. Because the random forest algorithm has been proven for its accuracy, hence it has been chosen to be applied on our dataset.

6.2. Results for Neural Networks

Using the classifier "Multilayer Perceptron" under Functions tap and applying it to the training set, we obtain the following detailed results:

Correctly Classified Instances 115 97.4576 %
Incorrectly Classified Instances 3 2.5424 %

The percentage of correctly classified instances indicated the level of accuracy for the applied algorithm. Since the percentage is 97%, it can be proved that the result is accurate. The time taken to build the model is 2.26 sec, which indicates that it is not as fast as compared to random forest tree algorithm.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0.012	0.972	1	0.986	1	Increased
	0.941	0.01	0.941	0.941	0.941	0.949	Decreased
	0.97	0.019	0.985	0.97	0.977	0.986	Same
Weighted Avg.	0.975	0.016	0.975	0.975	0.975	0.985	

Fig. 4. Result from Applying Neural Network Algorithm

By examining the detailed screenshot above, we can conclude that the result with value "Same" generates the most accurate results.

6.2.1. Evaluation methodology for Neural Networks Algorithm

The Neural Networks were also considered and implemented due to their accurate results. This argument we are proposing is supported by a research paper entitled "The Combination and Comparison of Neural Networks with Decision Trees for Wine Classification" [11]. In this paper, the authors wanted to compare the performance of two different types of classification algorithms, J48 decision tree and neural networks, to see which one would yield the most accurate results. A summary of their findings is illustrated in Table 3.

Table 2. Summary of the Results of Applying NN and DT on the same Dataset [11]

Classification Approach	Training Performance	Generalization Performance	Training Time (Epochs)
Neural Networks	99.7±0.5%	98.7±1.2%	166±67 epochs
Knowledge extraction from trained networks	--	95.6±1.3%	--
Decision Trees	98.5±0.2%	96.8±1.7%	--

From the above results, we conclude that using neural networks for classification produces more accurate results than that of J48 decision tree algorithm. Hence we validated our results using Neural Networks and Random Forest Algorithm to obtain accurate results.

7. Results Discussion

After applying two of the most accurate classification algorithms which such as neural networks and the Random Forest algorithm on the same data set, we obtained exactly the same result which is 97%. The accuracy of these two algorithms have been proven to be higher than other classification algorithms when they were applied to datasets similar to this research dataset as it was shown in the sections 6.1.1 and 6.2.1. But there is a difference in the speed for building the model between the two algorithms. The neural network took more time to generate the result than the time taken by the decision tree. Hence, it is clearly evident that Random Forest algorithms provide more accurate results than Neural Network in terms of speed.

8. Conclusion

The main objective for this research was to study whether the social media sites have increased, decreased, or have no effect on the time the users spend playing online games. The data needed for this study were gathered by surveying people of different demographics. As stated above, Random Forest and Neural Networks which are regarded as two of the most famous and highly accurate classification algorithms were applied on the pre-processed, cleansed data set. The result we obtained was surprisingly contradictory to our initial assumption that social media increases the time spent playing online games. The actual result of analyzing the data showed that social media had no effect, be it positive or negative, on the time users spend playing online games.

References

1. Global social networks by users 2015 | Statistic. (2015, August 1). Retrieved October 24, 2015, from http://www.statista.com/statistics/272014/global-social-networks-ranked-by-numberof-users/
2. Yee, N. (2006). Motivations for Play in Online Games. CyberPsychology& Behavior,9(6), 772-775. Retrieved October 24, 2015, from http://www.nickyee.com/pubs/Yee%20%20Motivations%20(2006).pdf
3. Rossi, L. (2010). Playing Your Network: Gaming in Social Network Sites. SSRN Electronic Journal SSRN Journal. Retrieved October 24, 2015, from http://ssrn.com/abstract=1722185

4. Jacobs, M., &Sihvonen, T. (2011). In Perpetual Beta? On the ParticipatoryDesign of Facebook Games. Retrieved October 5, 2015, from http://www.digra.org/dl/db/11312.19220.pdf
5. Pillai, D., & Oliver, J. (2015). A Novel Approach for the Classification of Social Media Data using Decision Tree. International Journal of Innovative Research in Computer and Communication Engineering, 3(6). Retrieved January 6, 2016, from http://www.ijircce.com/upload/2015/june/157_IJIRCCE_divya.pdf
6. Weka 3: Data Mining Software in Java. (n.d.). Retrieved November 18, 2015, from http://www.cs.waikato.ac.nz/ml/weka/
7. Drachen, A., Thurau, C., Togelius, J., Yannakakis, G., &Bauckhage, C. (n.d.). Chapter 12: Game Data Mining. Retrieved October 22, 2015, from https://andersdrachen.files.wordpress.com/2012/08/gamedatamining_intro.pdf
8. SINGH, Y., & CHAUHAN, A. (n.d.). NEURAL NETWORKS IN DATA MINING. Journal of Theoretical and Applied Information Technology, 37-42. Retrieved November 22, 2015, from http://jatit.org/volumes/research-papers/Vol5No1/1Vol5No6.pdf
9. Tadiou, K. (2013, May 26). Introduction to Artificial Intelligence - The Future of Human Evolution. Retrieved November 21, 2015, from http://futurehumanevolution.com/artificialintelligence-future-human-evolution/introduction-to-artificial-intelligence
10. Nancy, P., & Ramani, R. (2011). A Comparison on Performance of Data Mining Algorithms in Classification of Social Network Data. International Journal of Computer Applications, 32(8). Retrieved November 15, 2015, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.259.3668&rep=rep1&type=pdf
11. Chandra, R., Chaudhary, K., & Kumar, A. (n.d.). The Combination and Comparison of Neural Networks with Decision Trees for Wine Classification. 10-17. Retrieved November 21, 2015, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.384.3681&rep=rep1&type=pdf
12. Wender, S. (n.d.). Data Mining and Machine Learning with Computer Game Logs. Retrieved October 22, 2015, from https://www.cs.auckland.ac.nz/research/gameai/projects/StefanWender Data Mining Game Logs.pdf
13. Ansari, N., Talreja, M., & Desai, V. (2012). Data Mining in Online Social Games. Retrieved October 20, 2015, from http://link.springer.com/chapter/10.1007/978-81-322-0740-5_95
14. Shin, D., & Shin, Y. (2011). Why do people play social network games? Computers in Human Behavior, 852-861. Retrieved October 24, 2015, from http://www.sciencedirect.com/science/article/pii/S0747563210003535

Appendix A. Appendix A. Online Survey Dataset

Table 3. Results of the Online Survey

Sl.No	Question	Attribute Name	Attribute Values	Number of responders
1	You are	Gender	Male	50
			Female	68
2	Age	Age	<12 years old	2
			12 - 19 years old	45
			20 - 29 years old	67
			>29 years old	4
3	What platform do you use to play video games?	Platform	Video game consoles (PS, Xbox, Wii...etc.)	60
			PC	31
			Social media sites	24
			Other	3
4	What type of games do you play?	GameType*	Adventure games	49
			Virtual life games	22
			Design/Art games	19

			Fighting/War games	39
			Trivia games	14
			Other	12
5	How many hours per day do you spend playing?	HoursPlayed	<1 hour	51
			1 - 2 hours	31
			2 - 3 hours	16
			>3 hours	16
6	Do you think social media sites played a role in spreading these games?	SM-RoleInGames	Yes	62
			No	10
			To some extent	46
7	What social media sites did you use to find out about these games?	SocialMediaSite *	Facebook	60
			Twitter	32
			Instagram	34
			Tumblr	8
			Snapchat	17
			Youtube	61
			Other	7
8	Do you solely use social media sites to play games?	IsSoleUse	Yes	30
			No	80
9	Has the time you spend on social media sites increased/decreased with the addition of games?	TimeSpent	Increased	35
			Decreased	17
			Stayed the same	66
10	What motivated you to start playing these games?	MotivatedToPlay	Friends/Family	93
			Ads	13
			Celebrities	5
			Other	7

In Table 3, * indicates the responder may choose more than one value for this attribute.