Volume 2, Issue 4, April 2017 ISSN (online): 2456-0006
International Journal of Science Technology Management and Research
Available online at: www.ijstmr.com

Clustering High Dimensional Game Behavior Data Based on Distance Clustering Algorithm

Monika Deshmukh[1]
Department of Computer Engg.,
K.K.Wagh Institute of Engg. Education & Research,
Nashik ,MH, India

Poonam Borhade[2]
Department of Computer Engg,
K.K.Wagh Institute of Engg. Education & Research,
Nashik, MH, India

Mayuri Bhamare [3]
Department of Computer Engg,
K.K.Wagh Institute of Engg. Education & Research,
Nashik, MH, India

Nikhil Tandel[4]
Department of Computer Engg,
K.K.Wagh Institute of Engg. Education & Research,
Nashik, MH, India

Abstract: ***The game industry is facing a surge of data, which results from increasingly available highly detailed information about the behavior of software and software users. The data can come from a variety of channels, e.g. behavioral telemetry, user testing, surveys, forums, be high-dimensional, time-dependent and potentially very large. The old adage of big data having volume, velocity, variety and volatility holds very true for behavioral telemetry from games. Our proposed system will make non specialist to run this cluster analysis. Our system will make the steps accurate before the actual clustering algorithm takes place on data. Behavioral data sets can be large, time-dependent and high-dimensional. Clustering offers a way to explore such data and to discover patterns that can reduce the overall complexity of the data. Clustering and other techniques for player profiling and play style analysis have, therefore, become popular in the nascent field of game analytics. The Proposed system is going to use clustering techniques to mine game behavioral data. Some clustering techniques such as k-means clustering, k- Medoids clustering and spectral clustering are used.***

Keywords : Behavior mining, Clustering, Game analytics.

I. INTRODUCTION

Since the first computer game, the behavior of players has been registered, and responses to these behaviors calculated in real-time by the game software. It was not long into the life of computer (or video) games that measures of player behavior began appearing in a form visual to the player, for example high score lists, which stem back to the earliest arcade games. With the advent of the massively multi-player online game (MMOG), e.g. Meridian 59 and Ever Quest, player behavior analysis became important to monitor the population of persistent virtual worlds, e.g. ensuring stable economies and detecting fraudulent behavior [22]. Contemporaneously, user oriented testing and research methods have been widely adopted by game development [9, 12,36]. Initially, laboratory based methods have been utilized to analyze the behavior of users of computer games and the resulting experience. Over the past decade, principles and practices from game user research and telemetry analysis have begun to merge, providing hitherto unprecedented analytical power to user research, e.g. via permitting the collection of behavioral data in "the wild", from user-game interaction, purchasing behavior, social behavior, etc., giving rise to a series of different forms of analytics enabling high-resolution and large-scale behavioral analysis [25,33,38]. The growing interest in user-oriented behavior analysis in computer games is in part driven by the emergence of the multi-player and MMOG and Free-to-Play (F2P) game forms, which can support populations in the millions, as well as millions of objects and AI/script-driven entities. These games form a rules-governed complex of potential or realized interactions, and can be highly complex in terms of the user

interactions they provide [2,5,12]. Given such complexity, obtaining insights that are meaningful and actionable for game developers can be challenging, and means that behavior analysis is difficult to perform without dimensionality reduction, e.g. clustering.

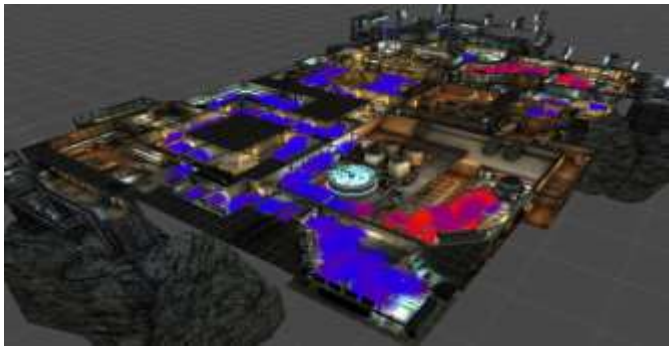

Fig. 1. Example of a 3D heat map (courtesy of Game Analytics) based on death events in the game Angry Bots.

Red areas indicate high player death counts Heat maps like this can visualize aspects of behavioral data, are intuitive to understand and allow for, say, identifying flaws in game design. Yet, they do not reveal underlying causes or correlations among latent features that would explain observed frequency counts.

II. LITERATURE SURVEY

Within game-oriented AI research, agent modeling, adaptive game research, player experience modeling, adaptive game research and not the least game-user research, the use of telemetry data extracted from game play behavior has been in use for about a decade [12,25,38]; however it is only in recent years that the game industry (with a few exceptions) outside of the massively multi-player online (MMOG) segment has begun adopting player-derived telemetry data to evaluate player behavior in games, e.g. [12,22,31,34]. Early work in the area was championed by what is now known as Microsoft Studios Research, who gained international recognition for their user research in the Halo series of games [12, 36]. In the past five years, other major game development publishers such as Bioware, Blizzard, Bioware, Square Enix and EA Games have been collecting and analyzing massive-scale behavioral telemetry from their games [38], although the details of the methods used are kept confidential [16,34] outside the rare academic-industry partnerships [e.g. 1,33]. In recent years, the rise of the free-to-play (F2P) genre, e.g. on platforms such as Facebook and Google Play, has added to the industry´s focus on behavior analysis. In F2P games, which can be of a persistent nature similar to MMOGs, playing the game itself is free, and revenue dependent on the ability of the developer to convince a portion of the customer base to purchase virtual items for real money via microtransactions [16,37]. In order to be successful as a business model, these games require continued analysis of player behavior in order to be financially profitable [37]. Academic and industrial research has built considerable knowledge, but there is minimal knowledge exchange between the two. It is for example only recently that academic experts have gained access to commercial game datasets [38]. The recency of game telemetry as a research topic also means that most available work is case-based, e.g. application of a specific algorithm to behavioral data from a specific game.

III. RELATED WORK

We begin this paper by studying K-means clustering. Although the K-means algorithm is a popular and widely used baseline technique, experience shows that practitioners are often not familiar with its principles. That is to say that the algorithm operates on implicit assumptions which, if ignored, may lead to seemingly unreasonable results. We analyze these principles, discuss when and how to apply -means, and point out pitfalls in its use. We begin by considering data in Euclidean spaces. Hence, assume a data set X={x1,x2,……..,xn}(R^m) whose elements are dimensional real-valued vectors that are to be clustered. Whenever we set out to cluster such data, we are interested in regularities within the sample and must decide how to characterize and how to search for latent structures. The -means algorithm provides the arguably most popular approach to these problems. It represents structures in terms of Fig. 2. k-means clustering in action. In this didactic example, 150 data points were sampled from three bivariate Gaussians and the number of clusters to be found was set to k =3. Since the procedure is tailored towards Gaussian mixtures and since the initial choice of centroids was favorable, it took only five iterations to converge to the globally optimal solution.

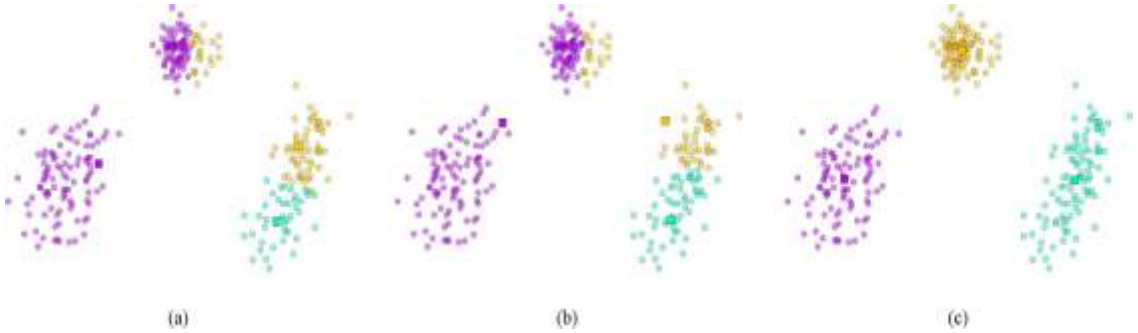

Fig.2. k-means clustering

Generally, however, there is no guarantee for k-means to behave like this. In practice, one should always run it several times and base any further analysis on the result that produced the minimal sum of distances. While -means clustering is guided by local properties of data (i.e., distances to centroids), spectral clustering assumes a global point of view. While -means clustering is clustering with m*n data matrices X=[x1,….xn] spectral clustering considers n*n similarity Matrice S whose Sij=s[Xi,Xj] entries indicate possibly abstract affinities between data objects. It is, therefore, related to the problem of graph partitioning, because affinity matrices can be seen as weighted graph adjacency matrices.

IV. PRAPOSED SYSTEM

To point out peculiarities of cluster analysis for game behavioral data. We discussed and reviewed pitfalls, common and uncommon problems, and theoretical foundations of popular algorithms. In particular, we discussed K-means clustering, matrix factorization, and spectral clustering and exposed the principles they work on. This was motivated by several years of experience obtained in the game industry and in games research.

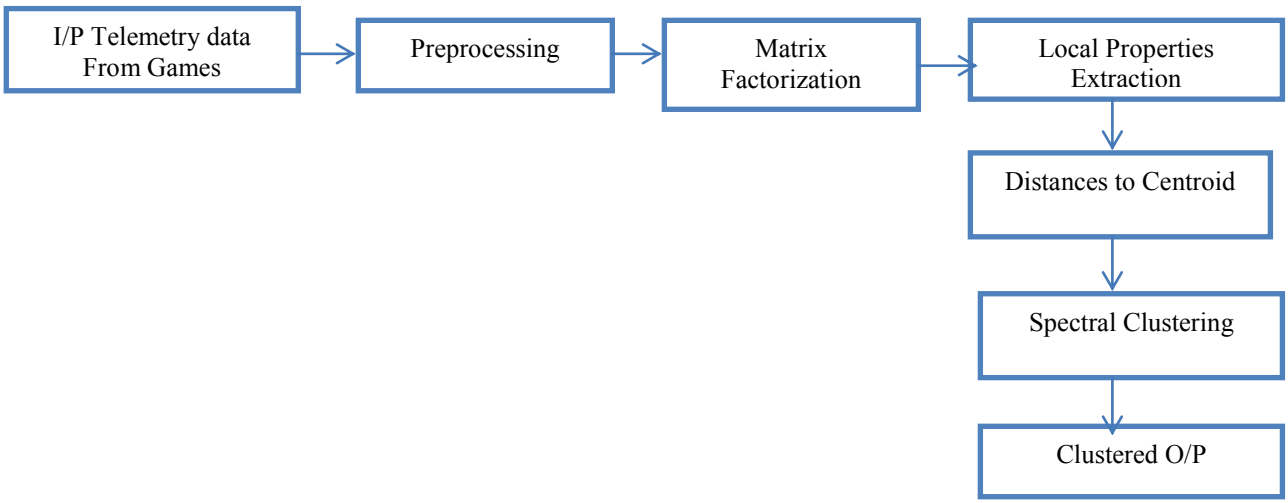

Fig3.Praposed System

Preprocessing:

Without knowledge of the game under examination and its mechanics, choices made during analysis ranging from feature selection and data preprocessing to visualization and interpretation run the risk of leading to flawed or useless results. Proper preprocessing enhances the chances for the means algorithm to identify structures that are not necessarily Gaussian.

Matrix factorization:

The task of K-means clustering can be viewed as a *matrix factorization problem*. In data mining and pattern recognition, factorization methods are frequently used for dimensionality reduction or latent component detection. To see how this relates to -means

clustering, we consider a data matrix, a matrix of centroid vectors, and a matrix of indicator variables. Given the above notation, we then find that, upon convergence of the K-means algorithm.

$$\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \approx \begin{bmatrix} \mu_{11} & \cdots & \mu_{1k} \\ \vdots & & \vdots \\ \mu_{m1} & \cdots & \mu_{mk} \end{bmatrix} \begin{bmatrix} z_{11} & \cdots & z_{1n} \\ \vdots & & \vdots \\ z_{k1} & \cdots & z_{kn} \end{bmatrix}$$

Local property extraction:

For Clustering we have to extract local properties of data (i.e., distances to centroids) using which feature we will do clustering.

Distance to centroid:

We are dealing with data which are not additive so that the notion of a mean is ill defined. An example related to game mining is the problem of clustering player names. As such, names, i.e., strings of characters, do not allow for computing averages. Nevertheless, we can compute, say, the edit distance between strings and the fact that the -means algorithm can be kernelized indicates that it is applicable to situations like these. An even simpler solution is to consider the use of the K-medoids algorithm.

As a contribution we are using Euclidean distance for clustering. Euclidean distance computes the root of square difference between co-ordinates of pair of objects.

$$Dist_{XY} = \sqrt{\sum_{k=1}^{m} (X_{ik} - X_{jk})^2}$$

Spectral clustering:

We discuss *spectral clustering*, another approach that applies matrix factorization. While K-means clustering is guided by local properties of data (i.e., distances to centroids), spectral clustering assumes a global point of view. Spectral clustering, therefore, relies on algebraic graph theory and its name derives from the fact that it clusters according to spectral properties, i.e., eigenvectors, of the Laplacian matrix. Unfortunately, the related machine learning literature is often vague as to why the Laplacian emerges in this context and thus obscures implicit assumptions of this method.

Algorithm K-means: Basic Euclidean distance metric

Let X = {x1,x2,x3,…….,xn} be the set of data points and V = {v1,v2,…….,vc} be the set of centers.

1. Select 'c' cluster centers randomly.

2. Calculate the distance between each data point and cluster centers using the Euclidean distance metric as follows

$$Dist_{XY} = \sqrt{\sum_{k=1}^{m} (X_{ik} - X_{jk})^2}$$

3. Data point is assigned to the cluster center whose distance from the cluster center is minimum of all the cluster centers.

4. New cluster center is calculated using:

$$V_i = \left(\frac{1}{Ci}\right) \sum_{1}^{ci} x_i$$

where, 'ci' denotes the number of data points in ith cluster.

5. The distance between each data point and new obtained cluster centers is recalculated.

6. If no data point was reassigned then stop, otherwise repeat steps from 3 to 5.

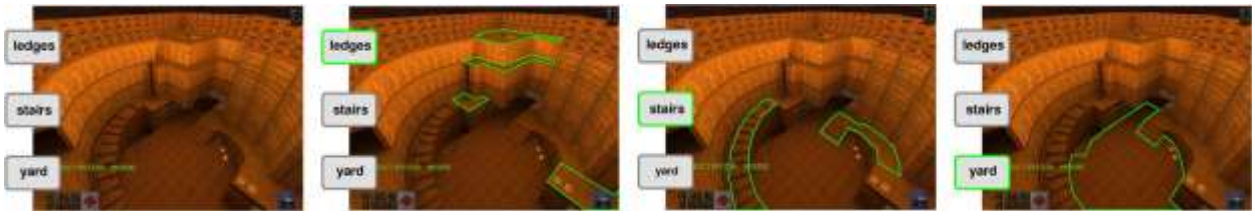

Fig.4.Screenshots showing a prominent area of the popular *Quake II* map *q2dm1*.

During a match, players moving in any of the highlighted locations may have to behave according to constraints or tactics that apply to these locations. That is, they will have to behave differently, depending on where they are located. The highlighted locations, therefore, form semantically distinct parts of the map and the question is if these can be determined through clustering.

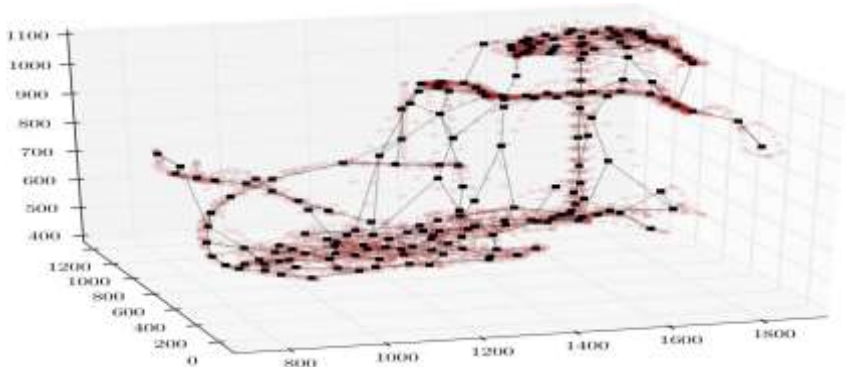

Fig. 5. Waypoint map of 200 nodes resulting from k-means clustering of player trajectories recorded on the map.

In Fig. 5. Stairs, yard, upper ledges, and elevator leading to the latter are recognizable. Expert maneuvers such as strafe jumping from the ledges caused several waypoints to occur "in midair".

CONCLUSION AND FUTURE WORK

Cluster analysis allows for finding latent patterns in game. Idea of cluster analysis is straightforward. It point out the peculiarities of cluster analysis for game behavioral data. This system is based on matrix factorization, and spectral clustering and exposed to the principles to be worked on. Also this proposed system has the methods, algorithms, and tools that clearly address the challenges of game behavior. In future work we will solve issues related to making sure that analysis results can be delivered to and acted upon by stakeholders.

REFERENCES

1. I. Dhillon, Y. Guan, and B. Kulis,'Kernel k-means, spectral clustering and normalized cuts' in Proc. KDD, 2004. P. Hall,J. Marron and A. Neeman,'Geometric reprsentations of high dimension, low sample size data' J. Royal Statist. Soc. B, vol. 67, no.3,2005.
2. R. Thawonmas and K. Iizuka,'Visualization of online-game players based on their action behaviors' Int. J. Comp. Games Technol., vol. 2008
3. F. Murtagh,'The remarkable simplicity of very high dimensional data: Applications of model-based clustering' J. Classif., vol. 26, no. 3, 2009.
4. A. Drachen, A. Canossa, and G. Yannakakis,'Player modeling using selforganization in tomb raider: Underworld' in Proc. CIG, 2009.
5. G. Yannakakis and J. Hallam,'Real-time game adaptation for optimizing player satisfaction' IEEE Trans. Computat. Intel. AI in Games, vol. 1, no. 2, 2009.
6. J. Baek, G McLachlan, and L. Flack,'Mixtures of Factor Analyzers with Common Factor Loadings: Applications to the Clustering and Visualization of High-Dimensional Data' IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 32, NO. 7, JULY 2010
7. A. Drachen, R. Sifa, C. Bauckhage, and C. Thurau,'Guns, swords, data: Clustering of player behavior in computer games in the wild' in Proc. CIG, 2012.
8. A. Drachen, C. Thurau, R. Sifa, and C. Bauckhage,'A comparison of methods for player clustering via behavioral telemetry' in Proc. FDG, 2013.
9. S. Wang, J. Yang, S. Chen and W. Kuo,'The Clusters of Gaming Behavior in MMORPG: A Case Study in Taiwan' IIAI International Conference on Advanced Applied Informatics,2015
10. A. Drachen, G. N. Yannakakis, A. Canossa and J. Togelius. Player Modeling using Self-Organization in Tomb Raider: Underworld. In Proc. of IEEE Computational Intelligence in Games, 2009.

11. J. Bohannon. Game-Miners Grapple With Massive Data. Science, 330(6000):30-31, 2010.
12. C.Thurau and C. Bauckhage. Analyzing the evolution of social groups in world of warcraft. In Proc. of IEEE Comp. Intelligence in Games, 2010.
13. A. Cutler and L. Breiman. Archetypal Analysis. Technometrics, 36(4):338-347, 1994.
14. A. Drachen and A. Canossa. Evaluating motion. Spatial user behavior in virtual environments. Int. Journal of Arts and Technology, v. 4 N3, 2011.
15. N. Ducheneaut and R. J. Moore. The Social Side of Gaming: A study of interaction patterns in a Massively Multiplayer Online Game. In Proc. of the 2004 ACM Conf. on Computer supported cooperative work, 2004.
16. L. Finesso and P. Spreij. Approximate Nonnegative Matrix Factorization via Alternating Minimization. In Proc. 16th Int. Symposium on Mathematical Theory of Networks and Systems, 2004.
17. G. Golub and J. van Loan. Matrix Computations. Johns Hopkins University Press, 3rd edition, 1996.
18. K. Isbister and N. Schafer. Game Usability. Morgan Kaufman, 2008.
19. B. J. Jansen. Understanding User-Web Interactions via Web Analytics. Morgan & Claypool Publishers, 2009.
20. I. Jollie. Principal Component Analysis. Springer, 1986.
21. J. H. Kim, D. V. Gunn, E. Schuh, B. C. Phillips, R. J. Pagulayan, and D. Wixon. Tracking real-time user experience (true): A comprehensive instrumentation solution for complex systems. In Proc. of CHI, 2008.
22. D. King and S. Chen. Metrics for Social Games. Presentation at the Social Games Summit, 2009.
23. D. D. Lee and H. S. Seung. Learning the Parts of Objects by Nonnegative Matrix Factorization. Nature, 401(6755):788-799, 1999.
24. T. Mahlman, A. Drachen, A. Canossa, J. Togelius, and G. N. Yannakakis. Predicting Player Behavior in Tomb Raider: Underworld. In Proceedings of IEEE Computational Intelligence in Games, 2010.
25. L. Mellon. Applying metrics driven development to MMO costs and risks. Versant Corporation, 2009.
26. O. Missura and T. Gärtner. Player modeling for intelligent difficulty adjustment. In Proc. of the ECML-09 Workshop From Local Patterns to Global Models, 2009.
27. P. Paatero and U. Tapper. Positive Matrix Factorization: A Non-negative Factor Model with Optimal Utilization of Error Estimates of Data Values. Environmetrics, 5(2):111-126, 1994.
28. R. Pagulayan, K. Keeker, D. Wixon, R. L. Romero, and T. Fuller. Usercentered design in games. In The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, pages 883-903. L. Erlbaum Associates, 2003.
29. J.-K. L. R. Thawonmas, K. Yoshida and K.-T. Chen. Analysis of revisitations in online games. Journal of Entertainment Comp., 2011.
30. R. Thawonmas and K. Iizuka. Visualization of online game players based on their action behaviors. Int. Journal of Computer Games Technology, 2008.
31. C. Thurau, K. Kersting, and C. Bauckhage. Convex Non-Negative Matrix Factorization in the Wild. In Proc. IEEE Int. Conf. on Data Mining, 2009.
32. C. Thurau, K. Kersting, M. Wahabzada, and C. Bauckhage. Descriptive matrix factorization for sustainability: Adopting the principle of opposites. Journal of Data Mining and Knowledge Discovery, 2011.
33. B. Weber and M. Mateas. A Data Mining Approach to Strategy Prediction. In IEEE Symposium on Computational Intelligence in Games, 2009.
34. G. N. Yannakakis and J. Hallam. Real-time Game Adaptation for Optimizing Player Satisfaction. IEEE Transactions on Computational Intelligence and AI in Games, 1(2):121-133, 2009.
35. K. Kersting, M. Wahabzada, C. Thurau, and C. Bauckhage. Hierarchical Convex NMF for Clustering Massive Data. Proc. of ACML, 2010.
36. G. Ostrouchov. On FastMap and the convex hull of multivariate data: toward fast and robust dimension reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8): 1340-1343, 2010.
37. C. Fraley and A.E. Raftery. How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. The Computer Journal, 41(8), 1998.
38. Y. Zheng and X. Zhou: Computing with Spatial Trajectories. Springer, 2011.
39. J. Miller, J. Crowcroft. Avatar Movement in World of Warcraft Battlegrounds. In Proceedings of IEEE Netgames, 2009.
40. F. Southey, G. Xiao, R. C. Holte, M. Trommelen and J. Buchanan. Semi-Automated Gameplay Analysis by Machine Learning. In proceedings of AIIDE, 2005.
41. T. Marsh, S. P. Smith, K. Yang and C. Shahabi. Continous and Unobtrusive Capture of User-Player Behavior and Experience to Assess and Inform Game Design and Development. In Proceedings of Fun and Games, 76-86, 2006.
42. B. G. Weber, M. John, M. Mateas and A. Jhala. Modeling Player Retention in Madden NFL 11. In Proceedings of IAAI, 2011.
43. G. Zoeller. Game Development Telemetry. In Proceedings of the Game Developers Conference, 2011.
44. J. Han J. and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2006
45. C. Thompson. Halo 3: How microsoft labs invented a new science of play. Wired Magazine, 15(9). T. Fields and B. Cotton. Social Game Design: Monetization Methods and Mechanics. Morgan Kauffman Publishers, 2011.
46. G. N. Yannakakis. Game AI Revisited. In Proceedings of ACM Computing Frontiers Conference, 2012.