Clustering avatars behaviours from Virtual Worlds interactions

Gema Bello Orgaz
Computer Science Department
Universidad Autónoma de Madrid, Spain
gema.bello@uam.es

María D. R-Moreno
Departamento de Automática
Universidad de Alcalá, Spain
mdolores@aut.uah.es

David Camacho
Computer Science Department
Universidad Autónoma de Madrid, Spain
david.camacho@uam.es

David F. Barrero
Departamento de Automática
Universidad de Alcalá, Spain
david@aut.uah.es

ABSTRACT
Virtual Worlds (VWs) platforms and applications provide a practical implementation of the Metaverse concept. These applications, as highly inmersive and interactive 3D environments, have become very popular in social networks and games domains. The existence of a set of open platforms like OpenSim or OpenCobalt have played a major role in the popularization of this technology and they open new exciting research areas. One of these areas is behaviour analysis. In virtual world, the user (or avatar) can move and interact within an artificial world with a high degree of freedom. The movements and iterations of the avatar can be monitorized, and hence this information can be analysed to obtain interesting behavioural patterns. Usually, only the information related to the avatars conversations (textual chat logs) are directly available for processing. However, these open platforms allow to capture other kind of information like the exact position of an avatar in the VW, what they are looking at (eye-gazing) or which actions they perform inside these worlds. This paper studies how this information, can be extracted, processed and later used by clustering methods to detect behaviour or group formations in the world. To detect the behavioural patterns of the avatars considered, clustering techniques have been used. These techniques, using the correct data preprocessing and modelling, can be used to automatically detect hidden patterns from data.

Categories and Subject Descriptors
H.3.3 [**Information Search and Retrieval**]: Selection process; I.5.3 [**Pattern Recognition**]: Clustering

General Terms
Behavioral patterns, Data processing, Clustering Techniques

Keywords
Virtual Worlds, Hierarchical Clustering, graph and overlapping clustering

1. INTRODUCTION
A Metaverse has been defined as a digital, or electronic representation of our real world. In these representations people can interact freely using the metaphor of their real lives in a non-limited world by physics, age, sex, or other real world characteristics. Initially, the concept of a Metaverse [31] can be described as a collective on-line shared space, created joining some virtually enhanced physical reality with a physically persistent virtual space [9]. From previous definitions and concepts some new software applications have been developed such as Virtual Worlds or augmented reality in the last years.

Virtual Worlds (VWs) have become a very popular kind of software application that has been used in different fields, from games to simulation or education. They allow individuals to interact with others through their avatars and with objects in the environment. This interaction and its 3D appearance makes VWs a good environment to develop educational, training, and collaborative tasks. The attractive 3D graphical environments provided by these VW can be used to improve interaction and a sense of realism for the users who access them. In these environments it is not only possible to see, hear and touch virtual objects, but also to create, edit and manipulate them as if they were physical objects.

The paper shows how traditional Data Mining techniques such as clustering methods can be used to analyse the behaviour shown by avatars during their stay in the VW. The clustering techniques were designed to find hidden information or patterns in a dataset. They are based on a blind search in an unlabelled data collection, grouping the data with similar properties in clusters without the necessity of labelled data or human supervision. From the set of available clustering methods, it has been selected the K-means method [22, 18]. This method uses a fixed number of clusters to group or divide the data, so it provides a simple way to evaluate some behavioural patterns that can be simulated in a predefined experiment, so we can analyse how this technique is working. We have used a new VW platform, named VirtUAM (from Virtual Worlds at Universidad Autónoma de Madrid), and it is built on the top of the OpenSim platform.

The rest of the paper is structured as follows. Section 2 provides some basic definitions and concepts related to VWs and data preprocessing techniques that must be developed before the clustering technique could be applied. Section 3 describes the clustering technique used in this work. Section 4 shows the dataset and the experimental results. Finally, the last section provides the conclusions and future lines of work.

2. BASICS ABOUT VIRTUAL WORLDS

Although VWs have been used in different domains such as Economy or E-Commerce [32], the most popular are related to massively-multiplayer on-line games, such as World of Warcraft or Doom [4]. VWs can be used as a new powerful instrument for instruction and education, where characteristics like persistence allow continuous and growing social interactions, which can be the basis for collaborative education [26, 28]. However, there are some characteristics that have not been fully explored in this kind of environments. Due to the immersive characteristics of VWs and the possibility to acquire automatically data from users. The information that can be collected from the avatars involves their current position, eye-gazing data, and other kind of interactions such as conversations, or operations performed in the virtual world (building objects, results of script compiling, etc...). It is possible to design experiments where the behaviour of a set of avatars can be analysed. Some of the previously described features could be extracted and analysed in a real environment, but with a very high complexity.

Currently, there exists an important number of available VWs platforms that can be used to design an implement virtual spaces, some of the most popular ones are, Active Worlds [1], Second Life (SL) [30], OpenCobalt [24] and OpenSim [25]. OpenSim is an open source multi-platform, multi-user 3D application server. It can be used to create VWs which can be accessed through a variety of clients and on multiple protocols. OpenSim can be used to simulate virtual environments similar to Second Life, given that it supports the Second Life messaging protocol and the Linden Scripting Language used to program in the VW. As a consequence, any SC client can be used with OpenSim, among others like RealXtend or HippoViewer.

In order to use OpenSim as an E-Learning and a Virtual Lab platform we have designed, on top of the OpenSim platform, a VW platform named VirtUAM. VirtUAM is comprised of four different modules: a grid of computers hosting the VW, a Web portal to provide users with access to information and data, several backoffice databases and finally a statistical module. In our platform users can build their own virtual space with an almost unlimited number of objects to interact with. Additionally, the platform can be accessed only by registered users. In this way we can prevent external users from accessing the VW-learning environment and interfering in the avatars tasks [27].

The reasons why the OpenSim platform seemed to be the most appropriate for our purposes can be summarized as follows:

- OpenSim is, unlike other virtual platforms like SL, open-source software, which allows administrators as well as other users to modify the program whenever the need arises. Such modifications might aim at the storage of player's behaviour within a database system.
- Information related to avatar behaviours, interactions, or chat records can be easily retrieved from external VW applications. In this way we are able to analyse data on avatar behaviour.
- Unlike other virtual worlds platforms like Active Worlds or SL, OpenSim is not a public virtual place. Instead it can be fully configured by its administrators to provide a virtual environment to limited users. This means that other external non-authorized users would not be able to visit our private buildings and possibly interfere in a non desired way.
- The virtual space which can be built by its users as well as the number of prims created in this platform are both unlimited.

Unlike real life, collecting information about the avatars behaviour in VWs is an easy task, however there are a number of issues that make it a non-trivial task. The platform monitor the avatars' activity, and log it in a set of log files. In order to use this data in the clustering subsystem, it is neccesary to preprocess it before clustering it.

3. AVATARS BEHAVIOUR ANALYSIS

Basically, we can distinguish two phases in our approach to VW behaviour analysis. The first one deals with the data preprocessing, i.e., the set of activities needed in order to feed the clustering algorithm with the data, and the behaviour analysis, which is done in this case with a classical clustering algorithm. In the following we describe these steps in more detail.

3.1 Data Preprocessing

The information preprocessing tasks [21] consist on clean and prepare the data for a subsequent analysis. The datasets usually contain outliers, missing values and categorical features that cannot be processed by most of the Data Mining techniques. These datasets also contain a high number of variables or features, which introduces noise and additional complexity to the analysis. These features are usually reduced through the analysis of their distribution and the correlation between them. Once the variables have been preprocessed, if the dimensionality of the problem is high, some techniques such as *projections* are used to reduce the complexity. However, projections lose the original information of the variables, which becomes unrecoverable.

There are several techniques which reduce the feature sets avoiding projections. These methods apply a guided search among the different attributes looking for the most useful variables for the analysis. These methods are usually known as *feature selection methods* [19]. Curiel *et al.* [11] applied Genetic Algorithms [16] to simplify prognosis of endocarditis using a codification where each individual of the population is a feature set. Blum and Langley [7] showed some examples of relevant features selection in datasets and applied them to several Machine Learning techniques. They defined different degrees of relevant features such as strong or weak relevant features. They also study some methodologies such as heuristic search, filters and wrapper approaches which are automatic feature selection methods usually validated by classification techniques. Some of these techniques usually introduce *over-fitting* to the model and are computationally expensive. Roth and Lange [29] applied these techniques to the clustering problem. This work presents a new feature and simple selection methodology whose goal is to make an intensive reduction of the attributes dimension oriented to clustering analysis. It applies clustering methods to validate the chosen feature set.

Once the data has been preprocessed, the second step is related to normalization. It allows to compare data features with different kind or range of values. Z-Score [8] and Min-Max [15] normalization methods are commonly used for preprocessing the data. Both normalization algorithms take the attribute records and they try to find a standard range for them. Min-Max has a fixed range, $[0, 1]$ (it is sensitive to outliers), while Z-Score depends of mean and standard deviation (it tries to approximate the distribution to a normal distribution, it is usually used to avoid outliers). Min-Max computes maximum and minimum values of the attributes to set the range while Z-Score applies recentralization.

Another popular normalization method is the *Normalized Compression Distance (NCD)* [10], which we used to compute the similarity distance between conversations in the VW. The NCD provides a measure of similarity between by using compressors. Given two objects x and y, their similarity is given by

$$NCD(x, y) = \frac{max\{C(xy) - C(x), C(yx) - C(y)\}}{max\{C(x), C(y)\}} \quad (1)$$

Where C is a compression algorithm, the function $C(x)$ is the size of the C-compressed version of x, and xy denotes the concatenation of objects x and y. NCD generates a number $0 \leq NCD(x, y) \leq 1$, and values near 0 denote similar objects while, on the contrary, values near 1 denote disimilar objects.

3.2 Data extraction and processing

The platform generates a large amount of different kinds of raw data. All of these data is automatically extracted and stored in several databases from each avatar connected to the VW. Data is kept using a predefined format that later can be easily exported and analysed by external statistical, or Data Mining, modules. All this information is stored by the platform in log files, so it first has to be extracted.

Figure 1: Two avatars maintaining a public conversation, the rest of the avatars in a radius fewer than 20 meters from them could ear their conversation.

In order to perform this task, regular expressions [13] have been used. In this way data like the messages sent between avatars, or the exact position of an event in the world, can be used [14].

Currently, three different data types are captured and later preprocessed. These can be briefly summarized as follows:

- *BourneChannel*. This is the general chat (see Figure 1) used in most of VW platforms. Following a short example of part of the text chat is shown.

```
[2011/12/16 11:46:41.00]
3230 Say <132.436, 48.58641, 23.92175>
'9b72c3b6-9aff-4d63-a9f6-234753936202':
what do you think about facebook?

[2011/12/16 11:46:41.00]
3220 Say <132.436, 48.58641, 23.92175>
'9b72c3b6-9aff-4d63-a9f6-234753936202':
I think it is a cool application

[2011/12/16 11:46:41.00]
400 Shout <132.436, 48.58641, 23.92175>
'9b72c3b6-9aff-4d63-a9f6-234753936202':
yes, but what do you think about the privacy?

[2011/12/16 11:47:58.00]
3230 Say <132.436, 48.58641, 23.92175>
'9b72c3b6-9aff-4d63-a9f6-234753936202':
well I'm not really worried about that
...
```

  The first field represents the time, followed by the ID and the command used to send the message. `Say` or `Shout` differs from the distance in meters that these messages could be ear by other avatars. The next field is the position of the avatar inside the world and it is represented as a vector. Following the position, an unique identifier (UUID) for the avatar is used. In the last position of each line is the exact message interchanged between avatars.

- *BourneIM*. It stores the private messages between two or more avatars (see Figure 2). This kind of information is currently used to analyse the behaviour of

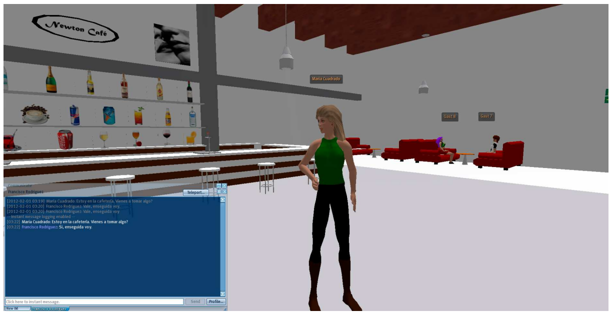

Figure 2: An avatar is sending a private message to a friend or other known avatar, the surrounding avatars cannot receive these messages.

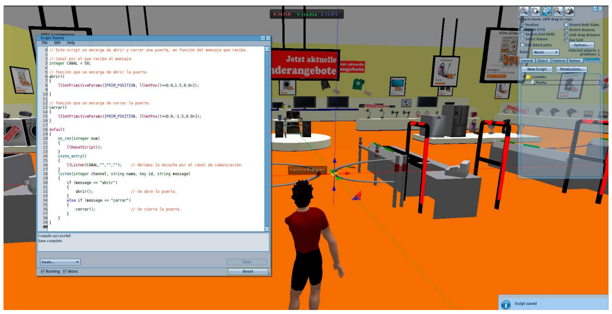

Figure 3: An avatar is programming using the Linden Scripting Language to generate some functionality in a particular object.

avatars that use the VW to learn foreign languages [5], in our case, German. There are several differences with previous (public) chat, one is related to the messages commands, now they are not necessary because the message is directly send to the owner. Finally, there exists an identification of the sender (From) and the receiver (To) using an UUID (unique identifier). An example of log containing this kind of data follows.

```
[2011/12/16 11:37:01.00]
<88.22602, 75.07764, 21.35308>
From '6cc7aa52-1d0d-499b-8096-7672c13008fc'
To 'b293be15-4fdd-4b5a-9b43-bbb39a54b4ea':
Hi, tell me something about your weekend

[2011/12/16 11:37:08.00]
<88.22602, 75.07764, 21.35308>
From '6cc7aa52-1d0d-499b-8096-7672c13008fc'
To '3d7167a4-fed4-4ea7-b14d-b49811ab14aa':
John I've asked to mary about her last weekend...

[2011/12/16 11:37:01.00]
<88.22602, 72.07764, 21.67529>
From  'b293be15-4fdd-4b5a-9b43-bbb39a54b4ea'
To '6cc7aa52-1d0d-499b-8096-7672c13008fc':
Hi David, I went to a party
...
```

- *Events*. From the whole set of available events, we have used the position (track) of the avatars, the compilation of scripts developed by the avatars, a set of "touch" events (i.e. touch an object). An example of scripting code which has been generated by an avatar is shown in Figure 3. The format of the information stored related to events is more complex than previous ones, to some extent because events need more attributes to be correctly described. Briefly, we can enumerate the position of the event, actions carried over objects in the VW, different identifiers (UUID) about the avatar, objects or even interaction ("touch", "run", "fly") made. An example of how information about events is stored follows.

```
[2011/12/16 11:54:40.00]
AbsolutePosition:
<217.4741, 43.02611, 23.63022>
CreatorID:
25144443-9113-4a5d-ac52-c8f695c448c4
Description:
LastOwnerID:
25144443-9113-4a5d-ac52-c8f695c448c4
LinkNum: 2
Material: 3
Name: flecha_abajo
OwnerID:
25144443-9113-4a5d-ac52-c8f695c448c4
UUID:
5823aa1f-481d-46d6-8968-f20bca09add7
```

After preprocessing, and once data has been normalized, data collected in the VW is prepared to be clusterized. We have used a classical K-means algorithm.

3.3 The K-Means Algorithm

Using a cluster algorithm with the data collected in the VW it is possible to identify those users with similar behaviour, which is a powerful tool for futher analysis. It is particularly interesting to perform a pedagogical analysis.

Clustering is a well known technique in Data Mining with an extense literature. There are three main techniques to deal with the clustering problem [17]: overlapping [6] (or non-exclusive), partitional [23] and hierarchical [20]. *Partitional clustering* consists in a disjoint division of the data where each element belongs only to a single cluster; *overlapping clustering* allows to each element belonging to multiple clusters and finally *hierarchical clustering* nests the clusters formed through a partitional clustering method creating bigger partitions.

In this work we applied partitional clustering because we wanted to classify the avatar behaviour, and therefore need to group all the elements of the dataset in no overlapped clusters. This need comes from the educational domain where our work is placed as a request made by teachers.

In order to validate the feature selection, we have applied clustering twice, one with the whole variable set extracted from the preprocessing phase, and another using the resulting variable set of the selected features phase. In this way, these two resulting groups can be compared to validate that the clusters obtained using the selected features are similar to the the clusters obtained using all the preprocessing variables. Therefore, the clustering technique is used as a tool to measure how good is working the selection feature strategy.

The partitional clustering algorithm that has been chosen is the K-means algorithm. K-means [22] is a popular and well known clustering algorithm. It is a straightforward clustering guided method, usually by a heuristic or directly by a human, to try to classify data. Given a fixed number of clusters (k), K-means tries to find a division of the dataset [23] based on a set of common features given by distances or metrics that are used to determine what elements belong to each cluster. K-means has also been improved though different techniques, like Genetic Algorithms [3].

K-means takes a set of n observations $\{x_1, \ldots, x_n\}$, which are represented as d-dimensional real vectors, partitioning this set in k collections. $S = \{S_1, S_2, \ldots, S_k\}, k \leq n$. The algorithm minimizes the distance between the elements of the clusters and the respective cluster centroids. It is represented by the following formula:

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x_j \in S_i} ||x_j - \mu_i||^2 \qquad (2)$$

Where μ_i is the position of the centroid i defined by the mean of the elements of S_i. The distance applied is the Euclidean distance, defined by

$$||\mathbf{p} - \mathbf{q}|| = ||\mathbf{q} - \mathbf{p}|| = \sqrt{\sum_{i=1}^{d} (q_i - p_i)^2} \qquad (3)$$

with $\mathbf{p} = (p_1, \ldots, p_d)$ and $\mathbf{q} = (q_1, \ldots, q_d)$. This algorithm is applied to the data gathered in the VW, which type of data is available in the platform is described in the next section.

4. INITIAL RESULTS USING CLUSTERING

In order to validate the preprocessing and clustering techniques that we have described above, we have carried out some experiments in the context of teaching foreign languages [5]. The goal of these experiments is double: Firstly, we want to verify if the algorithm is able to identify which students are performing which tasks. Secondly, we tried to check out how the feature selection affects the results.

To this end, we have introduced a set of controlled experiments where students had some clear, and different, tasks to develop. We have prepared four tests with nine different avatars in each session. For each experiment, thirty minutes was given to the users, and the avatars had to complete one of the following goals:

- Design and develop a particular set of objects described by a manual. For example, we asked avatars to design and build a supermarket. It includes programming the necessary functionality in the objects, such as open automatically the door when an avatar is approaching (see Figure 3).
- Find some people and go to a virtual room to talk about a predefined topic. For instance, a short discussion about social networks and how they can influence the current human social interactions (see Figure 1).
- Find a group of friends and send them private messages related to personal issues, like hobbies or other personal questions clearly different from the previous ones (see Figure 2).

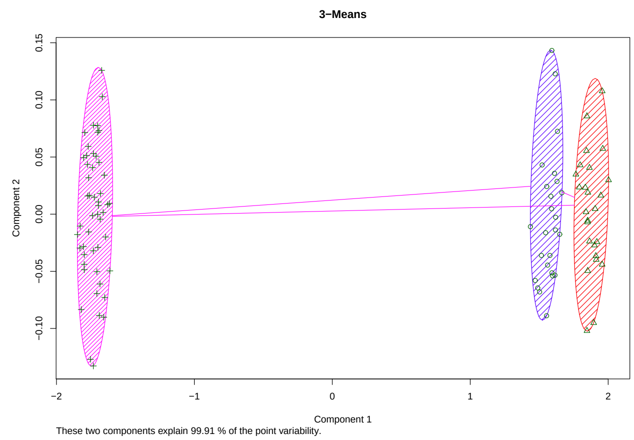

Figure 4: Projection of the data into the three clusters provided by the K-means algorithm using two components.

Using the proposed experimental design, a set of 30 sentences were extracted from both the `BourneChannel` and the `BourneIM` databases. The first 10 messages were taken from the initial part of the experiment, the following 10 from the middle, and the last 10 from the last phase of the test. For the `Event` database all the data was took into account because the information extracted from this source is mainly numerical. It is well known that for numerical data the clustering algorithm need more data to perform correctly. In order to process those sentences in natural language, we have used a classical statistical approach in Information Retrieval named TF-IDF [2], it extracts the relevant keywords and assigns them a weight that measures its relevance. These sets of keywords were given to the clustering algorithm.

The output of the K-means algorithm is shown in Figure 4. The figure has been plotted once the data has been preprocessed and projected into the two principal components. Figure 4 shows how the algorithm is able to discriminate clearly the three different behaviours from avatars in the experiments.

Taking into account the number of avatars, and how they were trained, it is possible to discriminate individually their behaviour. Figure 5 represents 20 avatars, one column for users and the three clusters detected using K-means. In this mosaic we can see how each avatar can be included in one of the clusters. Brown color represents those avatars dedicated to programming and design tasks. Purple color is used to represent those avatars that were involved in the public conversations, and finally the green color is used to classify the

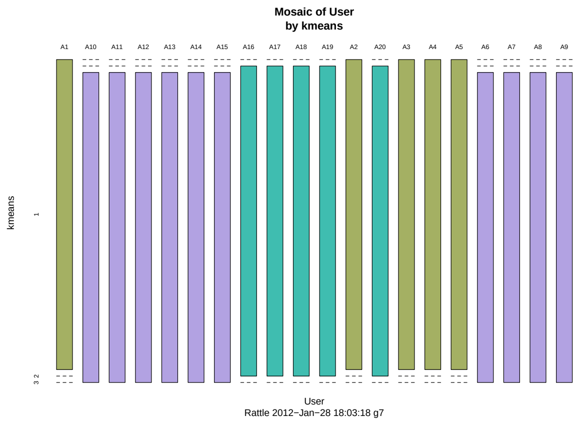

Figure 5: Distribution of avatars among the clusters generated by K-means. Clusters identified by K-means are represented with colours: Brown (programming and design tasks), purple (public chats) and green (private chats).

avatars who employed their time in private communications.

In this synthetic experiment the clustering is clearly excellent. Nonetheless, this outstanding performance can be explained because these users were previously trained and they follow the suggested goals of the experiment. In a more realistic scenario this excelent results are hardly repeatable. However, these results show how using the appropriate data processing and normalization techniques can be employed by traditional clustering algorithm to classify and follow the behaviour of human users in VW.

5. CONCLUSIONS AND FUTURE WORK

VWs have been used for a variety of purposes, from games to business or education. Some of the main work carried out in this field is directly related to the utilization of this immersive environment to provide spaces for communication and interaction. Other areas are related to the study of social behaviour and how cooperative attitudes can be achieved by humans. However, no much effort has been carried out to analyse and integrate the available information such as the avatar position, the interactions of them in the world or the textual information exchanged in chat conversations. Data Mining provides some promising tools such as clustering that can be used in this context.

This paper has shown an initial result in this direction. We have shown how raw data is extracted from a VW platform, preprocessed and later given to a clustering algorithm to automatically cluster the behaviour of a set of avatars. The initial tests showed some promising results and look to be useful to classify avatars into several groups. The use of a K-means algorithm over the text exchanged through the chat, and the VW interactions of avatars, reveals that these techniques can be used with a high degree of accuracy to classify simple behavioural patterns.

The future work will be focused on the feature selection and the analysis of different tools such as wrappers, heuristics and filter methods to extract and reuse data in VW in a more effective way. Different partitional clustering techniques can be also combined: Expectation Maximization and Spectral Clustering can be used to study the effects of the variable selection. These methods will be applied to other more complex and realistic datasets in an educational context to assess the validity of this strategy.

6. ACKNOWLEDGEMENTS

This work has been partly supported by: Spanish Ministry of Science and Education under the project TIN2010-19872. We would like to acknowledge Héctor D. Menéndez for his valuables suggestions.

7. REFERENCES

[1] ActiveWorlds. http://www.activeworlds.com/, february 2012.

[2] R. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.

[3] G. Bello, H. Menéndez, and D. Camacho. Using the clustering coefficient to guide a genetic-based communities finding algorithm. In H. Yin, W. Wang, and V. Rayward-Smith, editors, *Intelligent Data Engineering and Automated Learning - IDEAL 2011*, volume 6936 of *Lecture Notes in Computer Science*, pages 160–169. Springer Berlin / Heidelberg, 2011.

[4] F. Bellotti, R. Berta, A. D. Gloria, and V. Zappi. Exploring gaming mechanisms to enhance knowledge acquisition in virtual worlds. In *DIMEA '08: Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts*, pages 77–84. ACM, 2008.

[5] A. Berns, A. Gonzalez-Pardo, and D. Camacho. Designing videogames for foreign language learning. In *4th International Conference ICT for Language Learning, Florence (Italy)*, 2011.

[6] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. Pal. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing (The Handbooks of Fuzzy Sets)*. Springer, 1 edition, Mar. 2005.

[7] A. L. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artif. Intell.*, 97:245–271, December 1997.

[8] S. R. Carroll and D. J. Carroll. *Statistics Made Simple for School Leaders*. Rowman & Littlefield, 2002.

[9] E. Castonova. *Synthetic Worlds: The Business and Culture of Online Games*. University of Chicago, 2005.

[10] R. Cilibrasi and P. Vitanyi. Clustering by compression. *IEEE Transactions on Information Theory*, 51(4):1523–1545, 2005.

[11] L. Curiel, B. Baruque, C. Dueñas, E. Corchado, and C. Pérez-Tárrago. Genetic algorithms to simplify prognosis of endocarditis. In *Proceedings of the 12th international conference on Intelligent data engineering and automated learning*, IDEAL'11, pages 454–462, Berlin, Heidelberg, 2011. Springer-Verlag.

[12] K. Delac, M. Grgic, and S. Grgic. Independent comparative study of PCA, ICA, and LDA on the FERET data set. *International Journal of Imaging Systems and Technology*, 15(5):252–260, 2005.

[13] J. Friedl. *Mastering Regular Expressions*. O'Reilly Media, Inc., 2006.

[14] A. Gonzalez-Pardo, D. F. Barrero, D. Camacho, and M. D. R-Moreno. A case study on grammatical-based representation for regular expression evolution. In Y. Demazeau, F. Dignum, J. Corchado, J. Bajo, R. Corchuelo, E. Corchado, F. Fernández-Riverola, V. Julián, P. Pawlewski, and A. Campbell, editors, *Trends in Practical Applications of Agents and Multiagent Systems*, volume 71 of *Advances in Soft Computing*, pages 379–386. Springer Berlin / Heidelberg, 2010.

[15] J. Han and M. Kamber. *Data mining: concepts and techniques*. Morgan Kaufmann, 2006.

[16] J. H. Holland. *Adaptation in natural and artificial systems*. MIT Press, Cambridge, MA, USA, 1992.

[17] E. Hruschka, R. Campello, A. Freitas, and A. de Carvalho. A survey of evolutionary algorithms for clustering. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 39(2):133 –155, march 2009.

[18] G. Jimenez-Diaz, H. D. Menendez, D. Camacho, and P. A. Gonzalez-Calero. Predicting performance in team games. In I. I. for systems, C. Technologies of Information, and Communication, editors, *ICAART 2011 - Proceedings of the 3ed International Conference on Agents and Artificial Intelligence*, volume 1, pages 401 – 406, 2011.

[19] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artif. Intell.*, 97:273–324, December 1997.

[20] G. N. Lance and W. T. Williams. A General Theory of Classificatory Sorting Strategies: 1. Hierarchical Systems. *The Computer Journal*, 9(4):373–380, Feb. 1967.

[21] D. T. Larose. *Discovering Knowledge in Data*. John Wiley & Sons, 2005.

[22] D. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[23] J. B. Macqueen. Some methods of classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[24] OpenCobalt. http://www.opencobalt.org/, february 2012.

[25] OpenSimulator. http://opensimulator.org, february 2012.

[26] M. Rico, D. Camacho, and E. Pulido. Automatic guidance tools for enhancing the educative experience in non-immersive virtual worlds. In *Proc. of the IEEE Engineering Education Conference-EDUCON 2010*, volume 1. IEEE Press, 2010.

[27] M. Rico, G. Martínez, X. Alamán, D. Camacho, and E. Pulido. Improving the programming experience of high school students by means of virtual worlds. *International Journal of Engineering Education*, 27(1):52–60, 2011.

[28] T. Ritzema and B. Harris. The use of second life for distance education. *J. Comput. Small Coll.*, 23(6):110–116, 2008.

[29] V. Roth and T. Lange. Feature selection in clustering problems. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[30] SecondLife. http://secondlife.com/, february 2012.

[31] N. Stephenson. *Snow Crash*. Bantam Books, 1992.

[32] D. Talbot. The fleecing of the avatars. *Technology Review*, 111(1):58–62, 2008.