Bridging the Gap between Human Knowledge and Machine Learning

Juan C. ALVARADO-PÉREZ[a], Diego H. PELUFFO-ORDÓÑEZ[b], Roberto THERON[c]

jcalvarado@usal.es, diego.peluffo@campusucc.edu.co, theron@usal.es

[a] Universidad de Salamanca, Universidad Mariana

[b] Universidad Cooperativa de Colombia

[c] Universidad de Salamanca

KEYWORD	ABSTRACT
Data mining; visualization; machine learning	*Nowadays, great amount of data is being created by several sources from academic, scientific, business and industrial activities. Such data intrinsically contains meaningful information allowing for developing techniques, and have scientific validity to explore the information thereof. In this connection, the aim of artificial intelligence (AI) is getting new knowledge to make decisions properly. AI has taken an important place in scientific and technology development communities, and recently develops computer-based processing devices for modern machines. Under the premise, the premise that the feedback provided by human reasoning -which is holistic, flexible and parallel- may enhance the data analysis, the need for the integration of natural and artificial intelligence has emerged. Such an integration makes the process of knowledge discovery more effective, providing the ability to easily find hidden trends and patterns belonging to the database predictive model. As well, allowing for new observations and considerations from beforehand known data by using both data analysis methods and knowledge and skills from human reasoning. In this work, we review main basics and recent works on artificial and natural intelligence integration in order to introduce users and researchers on this emergent field. As well, key aspects to conceptually compare them are provided.*

1. Introduction

At present, the electronic and computer advances, such as sensors, satellites, GPS, technology web, cloud computing and social networking (Liu, 2014), (J. R. F. Díaz, 2011), among others, have contributed to the phenomenon of the generation of large volumes of data (Data Mountains) (Tufféry, 2011). These data mountains usually emerge and grow at a rapid pace, with nature and complex structure [4], and reach exorbitant dimensions, exceeding the capacity of conventional tools to capture, manage and process the data in a reasonable time. Recently, to encompass all aspects of this phenomenon in a unique concept, it has coined the term Big Data (Gantz & Reinsel, 2012). Additionally, the task of presenting and/or represent data in an understandable, intuitive and dynamic way is not a trivial task, since such data are often scattered, and come from various sources. Furthermore, when combined, data can become a new dataset, which may have an inconsistent and unpredictable structure (Dai & Hu, 2014).

Another aspect to take into account is the high number of variables representing data, here named dimensionality. For instance, complex systems (Ball, 2012) (i.e., social, biological physical and chemical systems (Turk-Browne, 2013) (Kutz, 2013)) generate data that normally have immediate use but have been stored for long time. Therefore, such data may present chaotic behavior (due to the nature of complex systems) and most of them have not been explored as a base to generate new knowledge (Ortigosa, Carro, & Quiroga, 2014). That said, one can say that available tools should evolve with the growth of data. Then, novel emergent techniques (Yongqing Wang & Li, 2014) (Dietterich, 1997) integrating the potential of computing resources (such as data mining (Butz et al., 2009)) and the qualities of human analysis and interpretation (such as visual analytics (Kim & Lee, 2014) (Theron & Fontanillo, 2015)). The synergetic integrations of artificial and natural intelligence represent suitable approaches to discover knowledge by devising a holistic view of the problem (Timaran, 2005) (Bertini & Lalanne, 2009). In this sense, interactive information visualization is a key tool that allows for access to and performing mental operations on large volumes of data, turning then the machine on a co-participant in the process of discovering patterns and unveiling of the emergent properties that are not readily anticipated a priori. The perception of a pattern can often be the basis of a new vision, so the data visualization (DV) facilitates understanding of the characteristics of both large scale and small scale of the data. A DV stage is often essential for designing a data analysis system when significant assumptions about the data need to made a priori (Bertini & Lalanne, 2009). Indeed, despite the existence of modern and powerful data processing tools, only a few can be used without a priori knowledge about the data (Y. Wang & Li, 2014), which is crucial to choose the best models, methods and algorithms to be tested on data. In this regard, DV can certainly help to build an initial idea about the structure of the data (outliers, clusters, local structures, etc.). In addition, visualizing data results useful for the users, especially for those who are not expert, may have visuals that allow intuitively analyze the results taking into account human perception system (Gibson, 2014).

In this paper, we describe two approaches for exploring data representing two independent learning forms, which use of computer processing and visual analytics that uses parallel thinking human being, namely the integration of artificial intelligence with the natural intelligence. Such an integration can be done by a Knowledge-Discovery-in-databases (KDD) process. This work briefly reviews approaches for artificial and natural intelligence integration, as well as presents an overview on data mining and information visualization.

The remaining of this paper is organized as follows: Section 2 presents a brief overview on KDD and data mining. Section 3 refers to information visualization based on human perception. In section 4, the approaches for natural and artificial intelligence integration are described. Section 5 outlines the final remarks and conclusions.

2. Data processing - On the side of the machine

The machine learning (ML) essentially consists in synthesizing rules from data using statistical and mathematical methods(Agrawal, Srikant, & others, 1994) (Mitchell, 1997). Such methods are used as the basis to develop techniques aimed at extracting information and generating knowledge. In this connection, all these techniques are encompassed in the data mining (DM) (Fayyad, Piatetsky-Shapiro, Smyth, & Uthurusamy, 1996). The overall goal of DM is to "reconstruct a story" from the information contained in the data. Nonetheless, DM researchers has given different definitions to the term data mining, such as: "the process of extracting and refining useful knowledge from large databases "(Kerber, s. f.), "the process of extracting information previously unknown, valid and actionable from large databases for later use in decision making" (Hirji, 1999), and "the exploration and analysis, by automatic and semiautomatic means, large amounts of data in order to discover meaningful patterns and rules".

2.1. KDD process

KDD has been defined as an iterative, interactive and non-trivial process for identifying valid, novel, potentially useful and understandable patterns to discover knowledge (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). KDD can be applied in different domains, for example: Fraud detection (Phua, Lee, Smith, & Gayler, 2012), human and machine diagnosis (Huang, Chen, & Lee, 2007), psychological profiling (Mena, 2003), trading trend analysis (Ras, Tsumoto,

& Zighed, 2008), whether forecasting, among others. Common tasks in KDD are generation of rules of induction, classification, clustering, pattern recognition, predictive modeling (D. H. Peluffo-Ordóñez, Alvarado-Pérez, & Castro-Ospina, 2015), and variable correlation analysis, among others (Agrawal et al., 1994). In this process, different stages are considered. Firstly, the connection to the data set is established that can be strong, weak or moderately coupledb(Timarán Pereira, 2011) (Timaran, 2005). Once the connection is established, it is necessary to select the data of interest, by excluding data instances being not relevant. Then, a filtering process (also called data cleaning) must be applied to the data set for eliminating irrelevant and/or noisy data (i.e. outliers, missing data, erroneous values) (Alvarez, s. f.). Finally, the KDD core is performed to establish patterns that describe predictable recurring events or trends(Brachman, Khabaza, Hill, Kloesgen, & Augustin, 1996) (Chen, Han, & Yu, 1996) (López & Herrero, 2004) (Imielinski & Mannila, 1996) (Matheus, Chan, & Piatetsky-shapiro, 1993). Figure 1 shows graphically the stages of the KDD process.
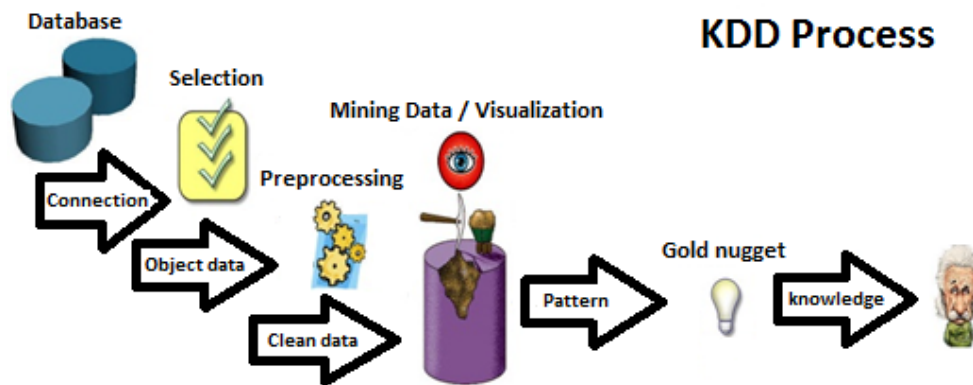
Figure 1: KDD Process. Connection, Selection, Preprocessing, Data Mining, Visualization, Discovery and Knowledge

2.2 DM techniques

The DM techniques are classified into two groups: Supervised (Weiss & Kulikowski, 1991) and unsupervised techniques (Riquelme, Ruiz, & Gilbert, 2006). In supervised learning, often also called directed data mining, the variables under investigation can be split into two groups: explanatory variables and one (or more) dependent variables. The target of the analysis is to specify a relationship between the explanatory variables and the dependent variable as it is done in regression analysis. To apply directed data mining techniques the values of the dependent variable must be known for a sufficiently large part of the data set. Some well-known approaches are decision trees, Bayesian classifier, neural networks, and for classification tasks. On the other hand, unsupervised techniques do not require performing training stages (no labeled data are needed) and the data processing is exploratory. Clustering and association are some examples of unsupervised approaches. Figure 2 depicts the classification of supervised and unsupervised methods.
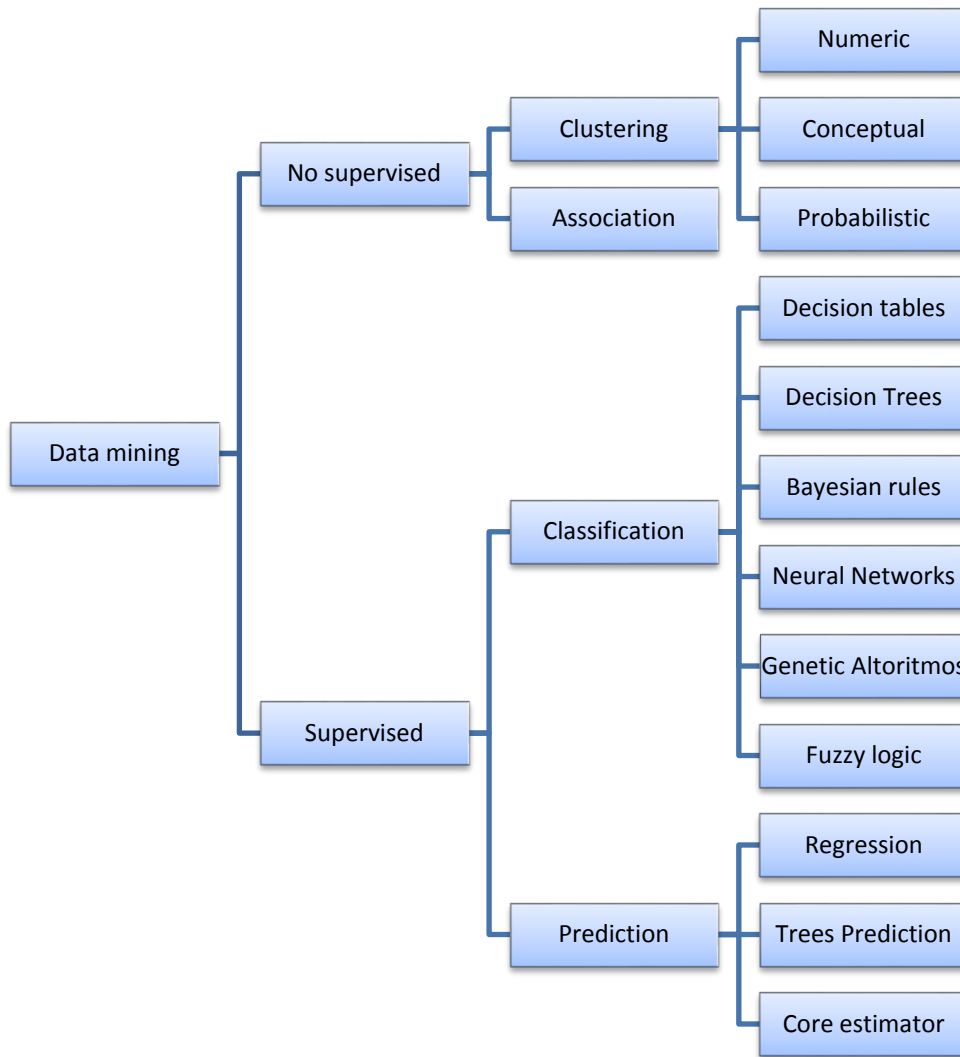
Figure 2: Classification of DM techniques.

3. Information visualization – On the side of the human

The human being is intuitive and deductive as well as has a convergent and divergent thought, having then the virtue of interpreting the world possible in a holistic, panoramic, parallel, flexible and contextual way, allowing for a quickly integrated sight of problems and taking out general concept for doing mind maps and establishing dynamic relationships among objects. All aforementioned is provided by imagination, creativity and spontaneity being characteristics that make a person able to give different meanings to the same term (i.e. the word love may be mean something different regarding the point of view) (Peluffo-Ordónez, Alvarado-Pérez, Lee, & Verleysen, 2015). Also, such a virtue makes a human being able to interpret different and ambiguous feelings such as irony, moods and sarcasm (N. V. Díaz & García, 2002), which cannot be interpreted by machine as feelings. This is due to the data interpretation by

a machine that is deterministic. Nevertheless, there exists inherent subjectivism associated with the analyst experience, which may bias the outcomes when analyzing data. To deal with this drawback, objective processes become necessary for supporting the process of decision making (Tufféry, 2011). A human being is rounded by different aroma, tastes, smelling, feelings, sounds and colors that are perceived and internalize through senses, who build a subjective concept from them. However, visual stimulus are the ones that give the most information to the brain so this process makes the sight the widest bandwidth sense, because it permits to perceive a lot of varied information from the environment (Pethuru, 2014). Broadly speaking, it means that "a picture is worth a thousand words" so if representative, such picture will facilitate human to build a mind map for concepts or problems learning.

At present, modern technology is making that images tell histories from big data, taking them into a multimedia and multidimensional world where human being is able of manipulating and interacting with data to build a complex reality representation in a graphic fashion, as well as identifying under-lying patterns and/or models (Cook, Earnshaw, & Stasko, 2007). Our world perception has been encoded by different symbols such as: letters, numbers, diagrams, general data, etc., which should be intelligible for human beings. Thus, some visualization approaches are used that provide a real semiotic ability for human interpretation, so it uses different types of resources such as colors, forms, textures and geometrical representations based on psychological phenomena like color interpretation, pre-attentive process, context and reference interpretation (Keim et al., 2008). There several tools for intelligent visualization (Koh, Slingsby, Dykes, & Kam, 2011), providing a semiotic relationship with explored data –by telling the data's story in a graphic, suitable and efficient way.

4. Artificial and natural intelligence integration

The modern computer and its emergent systems have implicitly generated a scenario where a human being and machine competition takes place, in which the latter has gradually increased its involvement into exclusively human beings activities. However, a machine might not be seen as a human being replacement since they both have special characteristics (as mentioned above) making them essentially different from each other. In this connection, it is worth to underline that under a proper integration of artificial –by machines- and natural –by human beings- intelligence may perform better and more robust processes. This means that an integrated system can be formed so that improved outcomes are reached. For instance, computers are deterministic, objective, tireless and untiring, as well as have a high processing ability (Alvarado-Pérez & Murillo, s. f.). Meanwhile, human beings have flexible, parallel and dynamic thought within a holistic context viewpoint. In this sense, by joining the characteristics from both approaches, versatile and suitable systems able to generate solutions benefitting society can be accomplished. Even though the two studied approaches are essentially different, they may have similar properties to be exploited to improve the system performance. On one hand, computers can be flexible (just like human beings) by adding to them some autonomous learning techniques such as fuzzy logic. On the other hand, human brain is divided into two hemispheres, being the right one in charge of creativity and feeling, meanwhile the left one provides the logic (Roselli, 2011), which is closer to the way how machines process the information. Indeed, many learning and processing systems for computers emulate somehow the brain behavior as it is the case of neural networks (Huang, Tsou, & Lee, 2006).

By doing an adequate synergetic integration between human beings skills and computer abilities involving the best properties from each one, artificial intelligence along with natural intelligence may reach both accurate calculation ability as well as holistic and flexible learning processes. Then, aimed at providing a robust and proper data exploration, an integrated system may discover innovative and useful knowledge in a more efficient way in comparison with traditional approaches (Kononenko & Kukar, 2007). Within integration frameworks, Data Mining (DM) algorithms are able to exhaustively explore and analyze data by taking advantage of information visualization (IV) techniques that facilitates human interpretation. In fact, visualization processes are aimed at providing users with an efficient way to interpret data making the DM outcomes more intelligible and facilitating mind map building in a graphic way (Torres Ponjuán, 2009) in this way the holistic and flexible human being thought is involved. In this sense, automatic learning and DM is on the computer side while the IV on the human interpretation side as explained in Figure 3. Integrated systems work iteratively and interactively demanding often a high computational cost as well as a strongly collaborative work between analyst and machine(Alonso, Martínez, Pérez, & Valente, 2012).
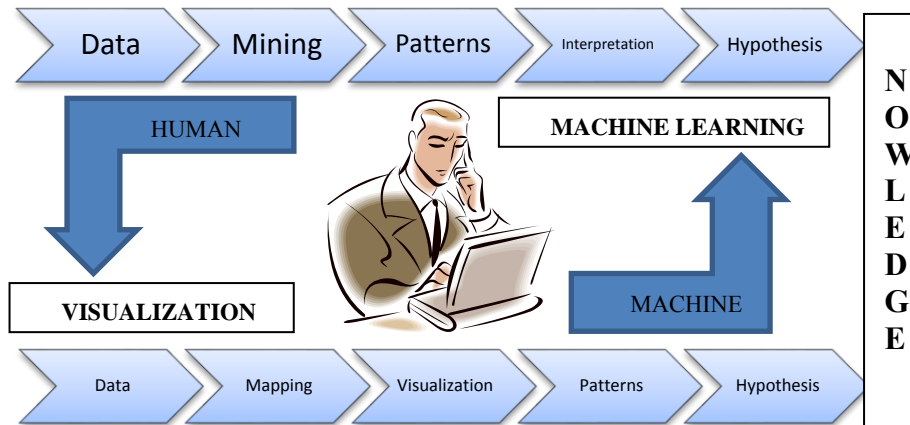
Figure 3: Integrating human beings in machine, through knowledge discovery process

4.1. Ways to integrate Information Visualization and Machine Learning

Las características ausentes en un método podrían encontrarse en otros, y por tanto, una apropiada integración de ellos lograría potenciar sus propiedades y generar una mejor representación de los datos.

DM and IV are fields that have been independently created and developed, being the former able to process data outputting if-then rules and patterns while the latter represents data in an easy-to-interpret fashion without using additional algorithm support. By construction, these methods were no intend-ed to interact with each other. Nonetheless, approaches to integrate DM and IV have been proposed (Bertini & Lalanne, 2009). Following are described approaches to integrate DM and IV, namely: Computer visualization, visually improved data mining.

4.2. Visualization supported in machine learning.

The V++ objective is to make a data representation allowing for detecting and drawing patterns through computationally optimal views. This approach is mainly a visual technique that takes advantage of computer algorithms such as those based on DM. Visualization attempts to reveal and visualize the most relevant patterns, that is to say those ones being more informative regarding some criteria. Thus, intended to facilitate human interpretation, visualization allow for low-dimensional data representation (2 or 3 dimensions). Nonetheless, when dealing with big data, the original data space embedding space into a lower dimension space is not a trivial task. For instance, just look at a 2D conventional screen, which itself is inherently limited by its available rectangular space. Indeed, the most common (being feasible and real) representation relies on a 2D plotting whether n-dimensional data (n >3) or even one-dimensional data (normally an extra corresponding axis is used, i.e. the axis of ordinal numbers being the simplest one). On this regard, available 2D representation must be optimally utilized for visualization purposes, and then an intrinsic bi-dimensional space for properly representing of data should be determined. To this end, there are different approaches such as: data projection, dimensionality reduction, multidimensional scaling, and data clustering(Diego H. Peluffo-Ordóñez, Lee, & Verleysen, 2014). Properly set, visualization approaches along with human interaction may reach the best 2D representation according to the user's needs. In Figure 4, the V++ process is shown. Notice that the symbol ++ means that visualization is improved by DM techniques.
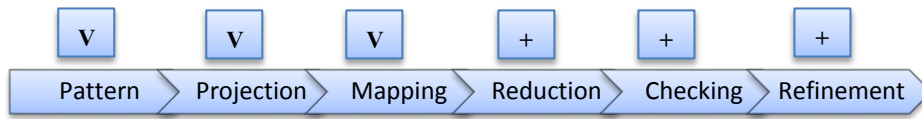
Figure 4: V++ Process

4.3. Machine learning supported in Visualization.

The core of this type of combination relies on the DM techniques. Mean-while, VI is an additional stage to visualize and validate the DM outcomes. This is done through an interactive interface allowing for not only interpret-ing the extracted model by the DM, but also visually verifying, assessing and refining the model outputted by the DM techniques. In this way, a clear interaction between analyst and model takes place, which allows a deeper comprehension of the underlying data information (Aguilar, Guerrero, & Penalvo, 2010). In other words, DM is in charge of innovative and useful patterns discovery, while IV only generate a visual representation of them. Figure 5. shows graphically the stages of the M++ process. Symbol ++ denotes DM is supported by IV.
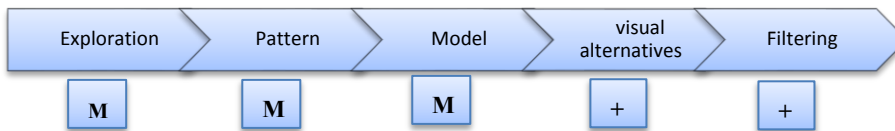

Figure 5. Process in M++

## 4.4. Synergistic cooperation between visualization and machine learning.

VM techniques involve IV and DM, in which both fields are equally im-portant. In fact, they combine in a synergic way the parallel thought from the human being with the processing capability of computers (Alvarado-Pérez & Peluffo-Ordóñez, 2015). Doing so, the best from the machine and human world converge to a meaningful data analysis and interpretation(Kai Puolamäki et al., 2010). Two types of VM approaches are explained below.

4.4.1. Integration black box with feedback.

In this approach, algorithms are black boxes, that is, the analyst cannot observe the way the DM algorithm internally works. Nevertheless, user may modify parameters whose effect is automatically reflected in the IV process, yielding then an adaptive approach. In other words, VM-IBBF creates a feedback loop be-tween IV and DM reaching more meaningful alternatives and models, ac-cording to the user's criterion.

4.4.2. Integration white box

This type of integration requires a strong joint effort between analyst and machine. Hence, every single stage of the DM algorithms can be observed so that model building process is easily accessed and modified. So, VM-IMB provides better understanding and allows users customize the algorithms. The interactive and iterative process of VM-IMB is shown Figure 6.
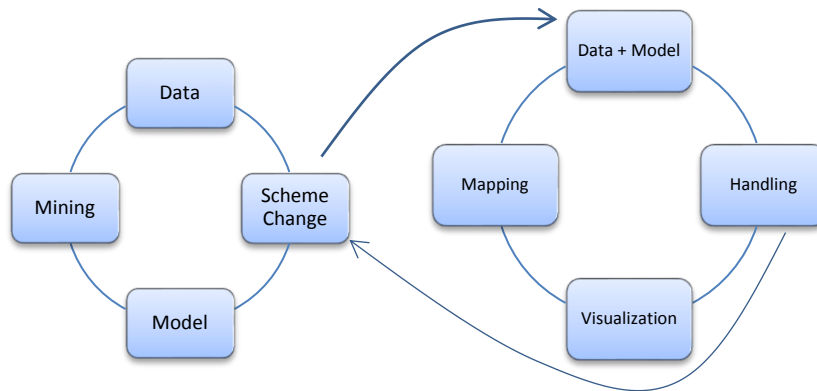
Figure 6. Interactive and Iterative process in the Integration White Box: VM

5. Conclusions and final remarks

In this work, we present a brief review of the main approaches for artificial and natural intelligence integration. Particularly, artificial intelligence is associated with DM (how to computer explore and analyze data). Correspondingly, natural intelligence refers to IV which allows for an intelligible human interpretation. We emphasize the benefit of the field of IV to describe, explore and analyze data in order to discover underling knowledge from data. The processes for knowledge extraction and schematic representation of reality by visual models are transversal topics very useful in several fields involving data. Nonetheless, such processes performance is strongly dependent on the nature of data as well as the right use of the data processing techniques. Given this, the synergic integration of artificial and natural intelligence represents a good alternative. Indeed, by properly mixing the holistic, flexible and parallel interpretation from human beings and the high computing storing capabilities of machines it is possible to discover relevant (even unexpected) new patterns within big data sets in an effective way. It is important to remark that due to the high demand of computational load, integration approaches are often implemented on parallel and distributed architectures such as computer clusters and grids, which in turn contribute to significantly improve the performance in terms of time processing and accuracy.

6. References

Agrawal, R., Srikant, R., & others. (1994). Fast algorithms for mining association rules. En *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487–499).

Aguilar, D. A. G., Guerrero, C. S., & Penalvo, R. T. S. and F. G. (2010). Visual Analytics to Support E-learning. http://doi.org/10.5772/7932

Alonso, F., Martínez, L., Pérez, A., & Valente, J. P. (2012). Cooperation between expert knowledge and data mining discovered knowledge: Lessons learned. *Expert Systems with Applications*, *39*(8), 7524-7535. http://doi.org/10.1016/j.eswa.2012.01.133

Alvarado-Pérez, J. C., & Murillo, S. (s. f.). Knowledge discovery in databases from a perspective of intelligent information visualization.

Alvarado-Pérez, J. C., & Peluffo-Ordóńez, D. H. (2015). Artificial and Natural Intelligence Integration. En S. Omatu, Q. M. Malluhi, S. R. González, G. Bocewicz, E. Bucciarelli, G. Giulioni, & F. Iqba (Eds.), *Distributed Computing and Artificial Intelligence, 12th International Conference* (pp. 167-173). Springer International

Publishing. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-319-19638-1_19. http://dx.doi.org/10.1007/978-3-319-19638-1_19

Alvarez, L. A. (s. f.). Aplicación de minería de datos a estudios históricos prosopográficos, 38.

Ball, P. (2012). *Why Society is a Complex Matter: Meeting Twenty-first Century Challenges with a New Kind of Science* (2012.ª ed.). Springer. http://dx.doi.org/10.1007/978-3-642-29000-8

Bertini, E., & Lalanne, D. (2009). Surveying the complementary role of automatic data analysis and visualization in knowledge discovery. En *Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration* (pp. 12–20). http://dx.doi.org/10.1145/1562849.1562851

Brachman, R., Khabaza, T., Hill, M., Kloesgen, W., & Augustin, S. (1996). An Overview of Issues in Developing and Knowledge Discovery Industrial Data Mining Applications, 89-95.

Butz, A., Fisher, B., Christie, M., Krüger, A., Olivier, P., & Therón, R. (2009). *Smart Graphics*. Proc. of 10th International Salamanca, España: Springer. Recuperado a partir de http://www.amazon.ca/Smart-Graphics-International-Symposium-Proceedings/dp/364202114X. http://dx.doi.org/10.1007/978-3-642-02115-2

Chen, M.-S., Han, J., & Yu, P. S. (1996). Data mining: an overview from a database perspective. *Knowledge and data Engineering, IEEE Transactions on*, *8*(6), 866–883. http://dx.doi.org/10.1109/69.553155

Cook, K., Earnshaw, R., & Stasko, J. (2007). Guest Editors' Introduction: Discovering the Unexpected. *IEEE Computer Graphics and Applications*, *27*(5), 15-19. http://doi.org/10.1109/MCG.2007.126

Dai, W., & Hu, P. (2014). Research on personalized behaviors recommendation System based on cloud computing. *Telkomnika*, *12*(2), 1480-1486. http://doi.org/10.11591/telkomnika.v12i2.3443

Díaz, J. R. F. (2011). El impacto de las redes sociales en la propiedad intelectual. *REDES*.

Díaz, N. V., & García, I. S. (2002). ?` Pensabas que emocionarse era sencillo? Las emociones como fenómenos biológicos, cognoscitivos y sociales. *Revista Puertorriqueña de Psicología*, *13*(1), 1.

Dietterich, T. G. (1997). Machine-learning research: Four current directions. *AI Magazine*, *18*(4), 97-136.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Commun. ACM*, *39*(11), 27–34. http://doi.org/10.1145/240455.240464

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*. The MIT Press. Recuperado a partir de http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0262560976

Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the Future*, *2007*, 1–16.

Gibson, J. J. (2014). *The Ecological Approach to Visual Perception: Classic Edition*. Psychology Press.

Hirji, K. K. (1999). Discovering data mining: from concept to implementation. *SIGKDD Explor. Newsl.*, *1*(1), 44–45. http://doi.org/10.1145/846170.846181

Huang, M.-J., Chen, M.-Y., & Lee, S.-C. (2007). Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. *Expert Systems with Applications*, *32*(3), 856-867. http://doi.org/10.1016/j.eswa.2006.01.038

Huang, M.-J., Tsou, Y.-L., & Lee, S.-C. (2006). Integrating fuzzy data mining and fuzzy artificial neural networks for discovering implicit knowledge. *Knowledge-Based Systems*, *19*(6), 396-403. http://doi.org/10.1016/j.knosys.2006.04.003

Imielinski, T., & Mannila, H. (1996). A database perspective on knowledge discovery. *Communications of the ACM*, *39*(11), 58–64. http://dx.doi.org/10.1145/240455.240472

Kai Puolamäki, Alessio Bertone, Roberto Therón, Otto Huisman, Jimmy Johansson, Silvia Miksch, … Salvo Rinzivillo. (2010). Chapter 4 in Mastering The Information Age – Solving Problems with Visual Analytics. En *Mastering the Information Age Solving Problems with Visual Analytics* (Daniel Keim, Jörn Kohlhammer, Geoffrey Ellis and Florian Mansmann). Germany.

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual Analytics: Definition, Process, and Challenges. En A. Kerren, J. T. Stasko, J.-D. Fekete, & C. North (Eds.), *Information Visualization* (pp. 154-175). Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-540-70956-5_7

Regular Issue
Vol.4 n.1
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

62

Kerber, E. S. B. L. R. (s. f.). Integrating Inductive and Deductive Database Mining.

Kim, K., & Lee, J. (2014). Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. *Pattern Recognition*, *47*(2), 758-768. http://doi.org/10.1016/j.patcog.2013.07.022

Koh, L. C., Slingsby, A., Dykes, J., & Kam, T. S. (2011). Developing and Applying a User-Centered Model for the Design and Implementation of Information Visualization Tools. En *2011 15th International Conference on Information Visualisation (IV)* (pp. 90-95). http://doi.org/10.1109/IV.2011.32

Kononenko, I., & Kukar, M. (2007). *Machine Learning and Data Mining*. Elsevier.

Kutz, J. (2013). *Data-driven modeling and scientific computing: Methods for Integrating Dynamics of Complex Systems and Big Data*. Oxford University Press.

Liu, Y. (2014). A knowledge discovery method based on Web information retrieval. En *WIT Transactions on Information and Communication Technologies* (Vol. 46 VOLUME 1, pp. 537-544). http://doi.org/10.2495/ISME20130701

López, J. M. ., & Herrero, J. G. (2004). Técnicas de Análisis de Datos. *Universidad Carlos III, Madrid*.

Matheus, C. J., Chan, P. K., & Piatetsky-shapiro, G. (1993). Systems for Knowledge Discovery in Databases. *IEEE Transactions On Knowledge And Data Engineering*, *5*, 903–913. http://dx.doi.org/10.1109/69.250073

Mena, J. (2003). *Investigative Data Mining for Security and Criminal Detection*. Butterworth-Heinemann.

Mitchell, T. M. (1997). *Machine Learning* (1.ª ed.). McGraw-Hill Science/Engineering/Math.

Ortigosa, A., Carro, R. M., & Quiroga, J. I. (2014). Predicting user personality by mining social interactions in Facebook. *Journal of Computer and System Sciences*, *80*(1), 57-71. http://doi.org/10.1016/j.jcss.2013.03.008

Peluffo-Ordónez, D. H., Alvarado-Pérez, J. C., Lee, J. A., & Verleysen, M. (2015). Geometrical homotopy for data visualization. En *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*.

Peluffo-Ordóñez, D. H., Alvarado-Pérez, J. C., & Castro-Ospina, A. E. (2015). On the Spectral Clustering for Dynamic Data. En J. M. F. Vicente, J. R. Álvarez-Sánchez, F. de la P. López, F. J. Toledo-Moreo, & H. Adeli (Eds.), *Bioinspired Computation in Artificial Systems* (pp. 148-155). Springer International Publishing. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-319-18833-1_16

Peluffo-Ordóñez, D. H., Lee, J. A., & Verleysen, M. (2014). Short Review of Dimensionality Reduction Methods Based on Stochastic Neighbour Embedding. En T. Villmann, F.-M. Schleif, M. Kaden, & M. Lange (Eds.), *Advances in Self-Organizing Maps and Learning Vector Quantization* (pp. 65-74). Springer International Publishing. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-319-07695-9_6

Pethuru, R. (2014). Data Visualization: Creating Mind´s Eye. En *Handbook of Research on Cloud Infrastructures for Big Data Analytics*. IGI Global.

Phua, C., Lee, V., Smith, K., & Gayler, R. (2012). A Comprehensive Survey of Data Mining-based Fraud Detection Research. *Computers in Human Behavior*, *28*(3), 1002-1013. http://doi.org/10.1016/j.chb.2012.01.002

Ras, Z. W., Tsumoto, S., & Zighed, D. A. (2008). *Mining Complex Data: ECML/PKDD 2007 Third International Workshop, MDC 2007, Warsaw, Poland, September 17-21, 2007, Revised Selected Papers*. Springer Science & Business Media.

Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Mineria de datos: Conceptos y tendencias. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, (029), 11–18.

Roselli, M. (2011). Maduración cerebral y desarrollo cognoscitivo. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, *1*(1). Recuperado a partir de http://revistaumanizales.cinde.org.co/index.php/Revista-Latinoamericana/article/view/336

Theron, R., & Fontanillo, L. (2015). Diachronic-information visualization in historical dictionaries. *Information Visualization*, *14*(2), 111–136. http://dx.doi.org/10.1177/1473871613495844

Timarán Pereira, R. (2011). Arquitecturas de Integración del Proceso de Descubrimiento de Conocimiento con Sistemas de Gestión de bases de datos: un Estado del Arte.

Timaran, R. (2005). *Nuevas primitivas sql para el descubrimiento de conocimiento en arquitecturas fuertemente acopladas con un sistema de gestion de bases de datos* (Doctoral). Universidad del Valle, Santiago de Cali, Colombia.

Torres Ponjuán, D. (2009). Aproximaciones a la visualización como disciplina científica. *ACIMED*, *20*(6), 161–174.

Regular Issue
Vol.4 n.1
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

63

Tufféry, S. (2011). *Data Mining and Statistics for Decision Making*. John Wiley & Sons. http://dx.doi.org/10.1002/9780470979174

Turk-Browne, N. B. (2013). Functional interactions as big data in the human brain. *Science*, *342*(6158), 580–584. http://dx.doi.org/10.1126/science.1238409

Wang, Y., & Li, Q. (2014). Review on studies and advances of machine learning approaches. *Telkomnika*, *12*(2), 1487-1494. http://doi.org/10.11591/telkomnika.v12i2.3635

Wang, Y., & Li, Q. (2014). Review on the Studies and Advances of Machine Learning Approaches. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, *12*(2), 1487–1494. http://dx.doi.org/10.11591/telkomnika.v12i2.3635

Weiss, S. M., & Kulikowski, C. A. (1991). *Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems*. M. Kaufmann Publishers.

Regular Issue
Vol.4 n.1
http://adcaij.usal.es

Advances in Distributed
Computing and Artificial
Intelligence Journal

64