



ARTIFICIAL LEARNING APPROACHES FOR THE NEXT GENERATION WEB: PART I

MÉTODOS DE APRENDIZAJE ARTIFICIAL PARA LA SIGUIENTE GENERACIÓN DEL WEB: PARTE I

J.R. Gutiérrez-Pulido¹, E.M. Ramos-Michel², M.E. Cabello-Espinosa³, S. Legrand⁴ and D. Elliman⁵
¹Faculty of Telematics, University of Colima, México, ²Faculty of Telematics, University of Colima, México, ³Info Systems and Computing Dept, University of Valencia, Spain, ⁴Dept of Computer Science, University of Jyväskylä, Finland, ⁵Computer Science and IT School, University of Nottingham, UK
E-mails: jrgp@ucol.mx, ramem@ucol.mx, mcabello@dsic.upv.es, stelegra@cc.jyu.fi, dge@cs.nott.ac.uk

(Recibido: junio de 2005; aceptado: octubre de 2006)

Abstract

In this paper we present a review of learning approaches that have been used by the research community to carry out clustering and pattern recognition tasks. Artificial neural networks are then introduced by presenting existing topologies, learning algorithms, and recall approaches. Finally, the relation of these techniques with the semantic web ontology creation process, as we envision it, is introduced. In part II of this paper, an artificial learning approach based on Self-Organizing Maps (SOM) that we have proposed as an ontology learning tool for assembling and visualizing ontology components from a specific domain for the semantic web is introduced.

Key words: *Ontology creation process, semantic Web approach, clustering, pattern recognition, artificial neural networks.*

Resumen

En este artículo presentamos una reseña de las técnicas de aprendizaje artificial que han sido utilizadas por la comunidad científica para llevar a cabo tareas de agrupamiento y reconocimiento de patrones. Se hace una introducción a las redes neurales artificiales y se presentan las topologías existentes, así como algoritmos de aprendizaje y técnicas de recuerdo. Finalmente, se presenta la relación entre estas técnicas y el proceso de creación de ontologías para Web semántico desde la perspectiva en que la entendemos. En la segunda parte de este artículo se introducen los Mapas Auto-Organizados o por sus siglas en inglés SOM (Self-Organizing Maps), que se han propuesto como una herramienta de aprendizaje artificial de ontologías para agrupar y visualizar componentes de conocimiento de dominio específico para el Web semántico.

Descriptores: Proceso de creación de ontologías, Web semántico, agrupamiento, reconocimiento de patrones, redes neuronales.

Clustering

Clustering is the unsupervised process of grouping patterns, observations, data items, or feature vec-

tors (Haykin, 1999). This problem has been addressed in different contexts and by researchers since the 60's in many disciplines, reflecting its broad appeal and usefulness as one of the steps in

exploratory data analysis. As a task, clustering is subjective in nature. The same dataset, for instance, may need to be partitioned in various ways for different purposes. This subjectivity may be integrated with the cluster criterion by incorporating domain knowledge in one or more phases of clustering. Every clustering algorithm uses some type of knowledge either implicitly or explicitly. A pattern set can be denoted as $S = \{d_1, \dots, d_m\}$. The i^{th} pattern in S is denoted as $d_i = \{a_{i1}, \dots, a_{in}\}$, $d_i \in \mathcal{R}^n$.

This pattern set is viewed as an $m \times n$ matrix. The individual scalar components a_{ik} are called features or patterns. Some classic approaches to the problem include partitional methods (Quan *et al.*, 2004), hierarchical agglomerative clustering (Stolz *et al.*, 2004), and unsupervised Bayesian clustering (Picton, 1994). Figure 1 shows a taxonomy of some clustering methods. A widely used partitional procedure is the k-means algorithm (Jain *et al.*, 2000). A problem with this procedure is the selection of k a priori. An alternative to these methods is SOM which does not make any assumptions about the number of clusters a priori, the probability distributions of the variables, or the independence between variables. Four important issues about clustering arise (Paulus and Hornegger, 2003):

- The method of the data preparation.
- The choice of the proximity measure.
- The choice of the classification method.
- The selection of the quality criteria.

Some of these issues will be discussed later in this paper.

Pattern recognition

Pattern Recognition (PR) is the study of how machines can observe their environment, distinguish patterns from their background, and make decisions about the categories of the patterns that are being studied. A pattern can be defined as an Entity that could be given a name. Some examples of such entities are *fingerprint images, handwritten words, human faces, speech signals, text documents*, and the like. The most common methods to deal with PR tasks are (Jain *et al.*, 2000):

- Template matching. One of the simplest and earliest approaches to PR. It determines the similarity between two Entities of the same

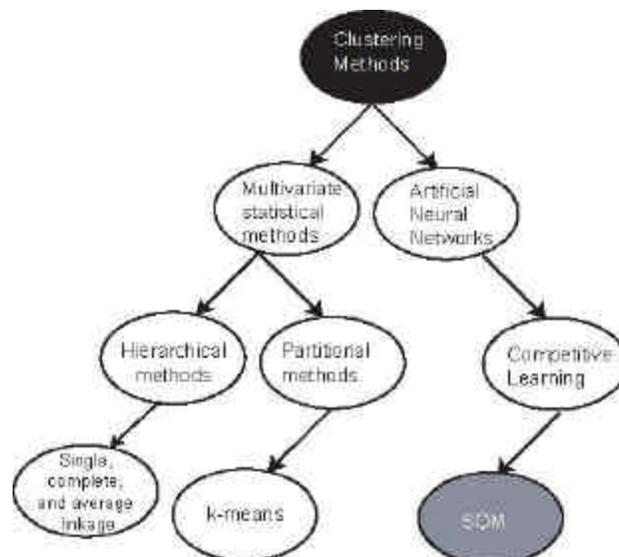


Figure 1. Clustering methods

type. A proto type of the pattern to be recognized must be available.

- Statistical classification. Each pattern is represented in terms of a number of features, which is viewed as a point in a multidimensional space.
- Syntactic matching. It provides a description of how the given pattern is constructed from primitives. Primitives are the simplest sub-patterns to be recognized. A more complex pattern can be defined in terms of the interrelations between these primitives.
- Artificial neural networks. They have the ability to learn complex non-linear input-output relationships and adapt themselves to the data.

Pattern recognition systems and the like, eg knowledge management, usually involve the following stages (Simpson, 1990); (Ritter and Kohonen, 1989); (Apte and Damerau, 1994):

Data acquisition: The acquisition and collection of measurements of data that have to be converted into a numerical form. Although this stage seems trivial, it is really important because the subsequent steps depend on the captured data.

Data processing: A priori knowledge is sometimes desirable at this stage which involves data removal, data scaling, data transformation, and if required, partitioning the data into subdatasets. This is a necessary step before training. It may also involve some kind of labeling of the dataset.

Feature extraction: The transformation of data into feature vectors. The latter must describe important phenomena from the data, distinguishing categories from one another. The dimensionality of the dataset is reduced at this stage.

Learning: If artificial neural networks are involved in the process, then this stage is needed. An issue at this point is the scaling of the feature

vectors. The variables with larger variance tend to dominate over variables with smaller variance. In order to avoid this effect, they are usually normalized.

Visualization and interpretation: These are two key issues in data analysis. These usually include cluster analysis and novelty detection. It usually requires iteration and an evaluation process for choosing the best solution to the problem of visualization and clustering based on minimizing the classification error.

In the 80's, Artificial Intelligence created a lot of excitement. The most significant and widespread outcome of it was the development of knowledge based expert systems. The knowledge base of expert systems is static, and such systems do not exhibit any automatic learning capability (Adeli and Hung, 1995). Artificial neural networks, on the other hand, do exhibit learning capabilities as will be described in the following section.

Artificial neural networks

Artificial Neural Networks (ANN) have become conventional tools for solving a large variety of pattern recognition and clustering tasks, frequently as an alternative to conventional statistical techniques. They are inspired by studies of biological nervous systems and are an attempt at creating machines that work in a similar way to the human brain (Petersohn, 1998).

Biological neural networks

The basic building block of the nervous system is a nerve cell called neuron. Neurons are one of the most important properties of animals. Plants, for instance, do not have nerve cells.

This cell is in charge of transmitting information to and from all over the human body. Basically, a neuron consists of three (Figure 2) sections:

- The body cell.

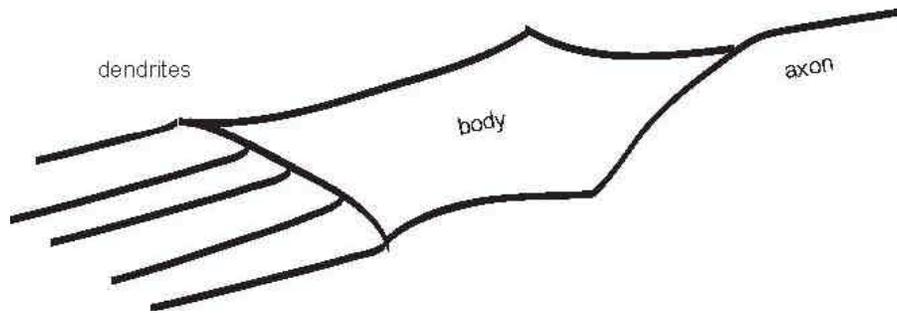


Figure 2. Biological neuron components

- Dendrites (lots of them).
- An axon.

Dendrites receive signals from other cells. These signals are passed on to the body cell where they are processed. Depending on the result of this process, the cell produces a pulse along its axon, which is passed on to succeeding cells.

The nervous system is composed of billions of neurons with the axon from one neuron branching out and connecting as many as 10,000 other neurons (Schroeder and Noy, 2001). All the neurons interconnected by axons and dendrites that carry signals that are regulated by synapses constitute a neural network. This is the same principle used in artificial neural networks.

Artificial neural networks

Artificial neural networks are a class of flexible semiparametric models for which learning algorithms have been developed over the years. They are based on how is thought the brain might process information. They resemble the brain in two respects (Quan *et al.*, 2004):

- Knowledge is acquired by the network through a learning process.
- Interneuron connection strength is used to store the knowledge.

There exist a number of different ANN approaches. They share, however, some of the following characteristics (Haykin, 1999); (Peterson, 1998); (Yom, 2004); (Masters, 1993):

- They are fault tolerant and degrade gracefully. Resistance to hardware failure is therefore much greater than in conventional computers.
- They are able to learn from the inputs they are given. They can even make decisions based on incomplete data.
- They are able to infer general properties from the different classes of patterns they have been shown and give, to some extent, the correct response even to examples that they have not been shown.
- They are inherently parallel and distributed processing architectures. It is desirable therefore to run these models on hardware that supports parallel processing.
- They learn interconnection weights adaptively by undergoing the modification of many processing elements at once.
- They process numerical vectors and so require patterns to be represented using quantitative features only.

ANN can be described as a graph with a set of nodes and edges connecting nodes. Such connections have weights which are changed using specific algorithms (Figure 3). This change is known as learning. During learning, the nodes that are close up to a certain geometric distance will activate each other and learn something from the same input. In a broad sense ANN consist of three elements (Schroeder and Noy, 2001):

- A Topology. How the processing elements are inter connected.
- A Learning method. How the process units are to store the data.
- A Recalling method. How the data is to be retrieved.

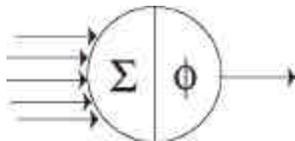


Figure 3. Artificial neuron components

These aspects will be briefly discussed in the following sections.

Topologies

Traditionally the following topologies are regarded as *pure* artificial neural networks (Ripley, 1996); (Masters, 1993); (Yom, 2004):

Signal-transfer networks

These networks are designed for signal transformation. The output signal values depend only on input signals. The mapping is parametric and the basic functions can be fitted to data by algebraic computation. These networks are also known as *feed forward networks*. These are faster than feedback networks as a solution can be reached with only one pass. Signal transfer networks also guarantee reach stability. They can not do some tasks that feedback networks can.

These are characterized by the lack of feedback. Single-layer forward networks are common.

State-transfer networks

These networks are also known as feedback networks or recurrent Networks. They can be created from feedforward networks by connecting the output of the neurons to the inputs. Feedback networks allow outputs to be input to preceding layers and lateral connections. They also are allowed to send inputs to other nodes in the same layer. These networks must iterate over many cycles until the system stabilizes. Feedback loops allow trainability and adaptability.

Competitive learning networks

The cells of these networks receive identical inputs on which they compete. By means of lateral interactions one of the cells becomes the *winner*. For different input signals the winner may alternate. These networks represent and classify samples that lie in a neighbourhood on the output space. This operation is commonly called clustering in pattern recognition. These networks are also called *winner-take-all* networks.

Learning

In order to learn, ANN have to be trained. In traditional programming, input-output relationships are established beforehand by the analyst. In contrast, ANN do not require instructions or rules about how to process the data. They determine, in most of the cases, relationships by looking at examples of *input-output* pairs. There are two types of learning (Pulido *et al.*, 2006):

Supervised

This kind of learning involves a teacher and requires input vectors to be paired with target vectors in a training set.

Unsupervised

This learning does not need a teacher and extracts features from the inputs themselves. No classes need to be defined a priori.

After learning, either supervised or unsupervised, the system is ready to use. From the wide range of proposed architectures of ANN, unsupervised models are regarded as well suited for text clustering. In a supervised environment, for instance, one would have to define input-output mappings anew every time the archive has changed. This would be highly labour-intensive and it is exactly this toil that unsupervised learning is designed to render unnecessary. In unsupervised environments, on the other hand, it remains the task of the ANN to uncover the structure of the document archive.

The knowledge in such ANN is not stored in specific memory locations as in conventional computing. Knowledge instead is distributed over the nodes and their connecting links, and is realised as a combination of the dynamic response to the inputs and the given network architecture (Masters, 1993).

Recall

The computation of an output for a given input performed by ANN is known as Recall. The objective of it is to retrieve information. In other words, it corresponds to the decoding of the stored knowledge from the network. There are two modes of recall (Yom, 2004); (Petersohn, 1998):

Heteroassociative

Memories where the stored output patterns are different from the corresponding input patterns are called heteroassociative. They produce an output that can be different from the input vector by providing a link between two distinct datasets.

Autoassociative

Autoassociative memories produce an output vector that is similar to the input vector. They are content-addressable and global (Picton, 1994). Content-addressable means that recall is done not through the address location but through content. During information retrieval no address is used except the input data. Global means that all the weights contain the knowledge in a distributed fashion and the weights are shared by all the memories in the system.

Discussion

For the semantic web to become a reality, a number of frameworks have to be built to support the artificial learning activities involved in the process. These activities (PR), as we envision this process, are as follow: *Gathering, Extraction, Organization, Merging, Refinement, and Retrieval.*

Learning techniques maybe applied by the knowledge engineer for *extraction* tasks (Principe, 2000); (Ehrig and Staab, 2004); (Maedche and Staab, 2001); (Legrand and Pulido, 2004). We have in particular used clustering (Johnson and Wichern, 1998) and artificial neural network techniques (Elliman and Pulido, 2002) and have obtained encouraging results (Pulido *et al.*, 2006). Semantic retrieval is the ultimate semantic web goal and it will take a while yet before we see smart software applications, but when the semantic web is populated, then those applications, eg semantic robots, agents, will traverse the web looking for data for us in a knowledge-based fashion (Pulido *et al.*, 2006). In the meanwhile, we still have to wait for those frameworks to mature.

In part II of this paper, an artificial learning approach based on Self-Organizing Maps (SOM) that we have proposed as an ontology learning tool for assembling and visualizing ontology

components from a specific domain for the semantic web is introduced.

References

- Adeli H. and Hung S. *Machine learning: neural networks, genetic algorithms, and fuzzy systems*. USA. Wiley & sons. Inc. 1995.
- Apte Ch. and Damerau F. Automated learning of decision rules for text categorization. *ACM Trans. on information systems*, 12(3):233–251. 1994.
- Ehrig M. and Staab S. QOM-Quick ontology mapping. In: The Semantic Web -ISWC 2004. Third international semantic Web conference (3rd, 2004). McIlraith S. et al., editors, Vol. 3298 of LNCS, Springer 2004, pp. 683–697.
- Pulido J.R.G. et al. Identifying ontology components from digital archives for the semantic web. In: IASTED advances in computer science and technology (ACST), CD edition, 2006.
- Elliman D. and Pulido J.R.G. Visualizing ontology components through selforganizing maps. In: International conference on information visualization (IV02) (6th, 2002, London, UK). Williams D, editor, pp. 434–438.
- Haykin S. *Neural networks: a comprehensive foundation*. 2nd edition. Prentice Hall. 1999.
- Jain A. et al. Data clustering: A review. *ACM computing surveys*, 31(3):264–323. 1999.
- Jain A. et al. Statistical pattern recognition: A review. *IEEE Trans. on analysis and machine intelligence*, 22(1):4–37. 2000.
- Johnson R. and Wichern D *Applied multivariate statistical analysis*. 4th edition. New Jersey. Prentice Hall. 1998.
- Legrand S and Pulido J.R.G. A hybrid approach to word sense disambiguation: Neural clustering with class labeling. In: Workshop on knowledge discovery and ontologies. European conference on machine learning (ECML) (15th, 2004, Pisa, Italy). Buitelaar P et al., editors, september 2004, pp. 127–132.
- Maedche A. and Staab S. Ontology learning for the semantic Web. *IEEE intelligent system*, 16(2): 72–79. 2001.
- Masters T. *Practical neural network recipes in C++*. Morgan Kaufman. 1993.
- Paulus D. and Hornegger J. *Applied pattern recognition*. Vieweg. 4th edition. 2003.
- Petersohn H. Assessment of cluster analysis and self-organizing maps. *Int. J. Uncertainty, Fuzziness and knowledge-based systems*, 6(2):139–149. 1998.
- Picton P. *Introduction to neural networks*, chapter 1 and 7. Macmillan, UK. 1994.
- Principe J. *Neural and adaptive systems, fundamentals through simulations*, chapter 7. Wiley. USA. 2000.
- Quan T. et al. Automatic generation of ontology of scholarly semantic Web. In: The Semantic Web ISWC 2004: International Semantic Web Conference, (3rd, 2004). McIlraith S. et al., editors, Vol. 3298 of LNCS, springer 2004, pp. 726–740.
- Ripley B. *Pattern Recognition and Neural Networks*, chapter 1,9. University Press. Cambridge. 1996.
- Ritter H. and Kohonen T. Self-organizing semantic maps. *Biological cybernetics*, (61):241–254. 1989.
- Schroeder M. and Noy P. Multi-agent visualisation based on multivariate data. In: International conference on autonomous agents, ACM (5th, 2001, New York). pp. 85–91.
- Simpson P. *Artificial neural system, chapter 2,3. Neural networks: Research and applications*. Pergamon Press. USA. 1990.
- Stolz C. et al. Measuring semantic relations of web sites by clustering of local context. In: Web engineering: International conference, ICWE 2004 (4th, 2004). Koch N. et al., editors, Vol. 3140 of LNCS, springer 2004, pp. 182–186.
- Pulido J.R.G. et al. Ontology languages for the semantic Web: A never completely updated review. *Knowledge-based systems*. Elsevier 19 (7):489–497. 2006.
- Yom E. An introduction to pattern classification. In: Machine Learning 2003, (Bousquet O. et al., editors), Vol. 3176 of LNCS, springer 2004, pp. 1–20.

Suggesting biography

- Bolodeau M. and Brenner D. *Theory of multivariate statistics*. Chapter 10. Springer Texts in Statistics. New York. Springer-Verlag. 1999.
- Ding C. and He X. Cluster structure of K-means clustering via principal component analysis. In: *Advances in knowledge discovery and data mining: Conference, PAKDD 2004 (8th, Pacific-Asia, 2004)*. Dai H. et al., editors, Vol. 3056 of LNCS, springer 2004, pp. 414–418.
- Elliman D. and Pulido J.R.G. Automatic derivation of on-line document ontologies. In: *Int. workshop on mechanisms for enterprise integration: From objects to ontologies (MERIT 2001)*, the European conference on object oriented programming (15th, june 2001, Budapest, Hungary).
- Elliman D. and Pulido J.R.G. Self-organizing maps for detecting ontology components. In: *The 2003 International conference on artificial intelligence (IC-AI) (2003, Las Vegas, USA)*. Arabnia H. et al. editors, CSREA Press, 2003, pp. 650–653.
- Ghahramani Z. Unsupervised learning. In: *Machine learning 2003*, (Bousquet O. et al., editors), Vol. 3176 of LNCS, springer 2004, pp. 72–112.
- Hatala M. and Richards G. Value-added meta-tagging: Ontology and rule based methods for smarter metadata. In: *Rules and rule markup languages for the semantic Web: International work shop, rule ML 2003 (2nd, 2003)*. Schroeder M. and Wagner G., editors, Vol. 2876 of LNCS, springer 2003, pp. 65–80.
- Maedche A. and Zacharias V. Clustering ontology-based metadata in the semantic web. In: *Principles of Data Mining and Knowledge Discovery: European Conference, PKDD (6th, 2002)*. Elomaa T. et al., editors, Vol. 2431 of LNCS, springer 2002, pp. 348–360.
- Nelson M. and Illingworth W. *A practical guide to neural nets*. Addison Wesley. 1990.
- Park L et al. A novel web text mining method using the discrete cosine transform. In: *Principles of Data Mining and Knowledge Discovery: European Conference, PKDD (6th, 2002)*. Elomaa T. et al., editors, Vol. 2431 of LNCS, springer 2002, pp. 385–397.
- Simula O. et al. Self-organizing maps in analysis of large-scale industrial systems. In: *Kohonen maps (1999, Berlin)*. Oja E. and Kaski S. editors, springer verlag, pp. 375–387.
- Tanaka H. et al. An efficient document clustering algorithm and its application to a document browser. *Information processing and management*, 35: 541–557. 1999.
- Wasserman P. *Neural computing, theory and practice*, chapter 1, app. A. New York. Van Nostrand Reinhold. 1989.
- Zaruda J. *Introduction to artificial neural systems*, chapter 1, 2, 7. USA. West Publishing Company. 1992.

About the authors

- Jorge Rafael Gutiérrez-Pulido*. Information Systems Laboratory, Faculty of Telematics, Colima University, México. He holds a PhD in Computer Science (2004) from The University of Nottingham, UK, a MSc in Telematics (1999), and a BA in Informatics (1995), both from The University of Colima, México. Dr. Pulido has written a number of journal papers, conference papers, and books on the following topics: The Semantic Web, Ontologies, Ontology Learning, Intelligent Data Mining, Knowledge discovery, Self-Organizing Maps, Internet Computing, Semantic Information Retrieval, and Data Visualization.
- Erika Margarita Ramos-Michel*. Faculty of Telematics, Colima University, México. She holds a MSc in Telematics (1998), and a BSc in Computer Systems (1995), both from The University of Colima, México. She is currently doing a PhD in Computer Science at CICESE, México. Research interests: Pattern recognition and image processing.
- María Eugenia Cabello-Espinosa*. University Politechnic of Valencia, Valencia, Spain. She holds a BA in Physics from UNAM (1994), and MSc in Computer Science from Arturo Rosenbleuth Foundation (2000), both from Mexico. Currently she is doing a PhD in Declarative Programming and Programming Engineering at The University Politecnico of Valencia in Spain. Research interests: Knowledge-Based Systems, Diagnostic Systems, and Software architecture based in components and aspects oriented.
- Steve Legrand*. Computer and Mathematical Sciences (CoMaS), Jyväskylä University, Finland. He holds an MSc from Jyväskylä University, Finland, a Post-Graduate Diploma in Applied Computing from University of Central Queensland, Australia, and a BSc from Griffith University, Australia. Currently he is doing a PhD in Jyväskylä University, Finland. Research interests: The semantic web, disambiguation, word sense discrimination, ontology, and natural language processing.
- Dave Elliman*. Image Processing & Interpretation Research Group Leader, School of Computer Science and IT (CSiT), Nottingham University, UK. Research interests: Pattern Recognition, Machine Vision, Agent-based Architectures for Problem Solving and Searching, Neural Networks, Constraint-based Reasoning, Steganography and Security especially relating to document servers, and Cognitive Architectures and Emotion.