

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Ignacio Rojas Gonzalo Joya
Joan Gabestany (Eds.)

Advances in Computational Intelligence

12th International Work-Conference
on Artificial Neural Networks, IWANN 2013
Puerto de la Cruz, Tenerife, Spain, June 12-14, 2013
Proceedings, Part I



Springer

Volume Editors

Ignacio Rojas
University of Granada
Department of Computer Architecture
and Computer Technology
18071 Granada, Spain
E-mail: irojas@ugr.es

Gonzalo Joya
University of Malaga
Department of Electronics Technology
29071 Malaga, Spain
E-mail: gjoya@uma.es

Joan Gabestany
Universitat Politecnica de Catalunya
Department of Electronics Engineering
08034 Barcelona, Spain
E-mail: cabestany@aha-dee.upc.edu

ISSN 0302-9743

e-ISSN 1611-3349

ISBN 978-3-642-38678-7

e-ISBN 978-3-642-38679-4

DOI 10.1007/978-3-642-38679-4

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2013938983

CR Subject Classification (1998): J.3, I.2, I.5, C.2.4, H.3.4, D.1, D.2

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Table of Contents – Part I

Invited Talks to IWANN 2013

It's as Easy as ABC: Introducing Anthropology-Based Computing	1
<i>John N.A. Brown</i>	
Extreme Learning Machine: A Robust Modeling Technique? Yes!	17
<i> Amaury Lendasse, Anton Akusok, Olli Simula, Francesco Corona, Mark van Heeswijk, Emil Eirola, and Yoan Miche</i>	
A Novel Framework to Design Fuzzy Rule-Based Ensembles Using Diversity Induction and Evolutionary Algorithms-Based Classifier Selection and Fusion	36
<i>Oscar Cordón and Krzysztof Trawiński</i>	

Applications of Computational Intelligence

Using Nonlinear Dimensionality Reduction to Visualize Classifiers	59
<i>Alexander Schulz, Andrej Gisbrecht, and Barbara Hammer</i>	
Which Dissimilarity Is to Be Used When Extracting Typologies in Sequence Analysis? A Comparative Study	69
<i>Sébastien Massoni, Madalina Olteanu, and Nathalie Villa-Vialaneix</i>	
Implementation of the C-Mantec Neural Network Constructive Algorithm in an Arduino Uno Microcontroller	80
<i>Francisco Ortega-Zamorano, José Luis Subirats, José Manuel Jerez, Ignacio Molina, and Leonardo Franco</i>	
A Constructive Neural Network to Predict Pitting Corrosion Status of Stainless Steel	88
<i>Daniel Urda, Rafael Marcos Luque, Maria Jesus Jiménez, Ignacio Turias, Leonardo Franco, and José Manuel Jerez</i>	
Robust Sensor and Actuator Fault Diagnosis with GMDH Neural Networks	96
<i>Marcin Witczak, Marcin Mrugalski, and Józef Korbicz</i>	
Diffusion Methods for Wind Power Ramp Detection	106
<i>Ángela Fernández, Carlos M. Alaíz, Ana M. González, Julia Díaz, and José R. Dorronsoro</i>	

Computational Study Based on Supervised Neural Architectures for Fluorescence Detection of Fungicides	114
<i>Yeray Álvarez Romero, Patricio García Báez, and Carmen Paz Suárez Araujo</i>	
Study of Alternative Strategies to Selection of Peer in P2P Wireless Mesh Networks	124
<i>Lisette Valdés, Alfonso Ariza, Sira M. Allende, Rubén Parada, and Gonzalo Joya</i>	
A Cloud-Based Neural Network Simulation Environment	133
<i>Erich Schikuta and Erwin Mann</i>	
Performance Evaluation over Indoor Channels of an Unsupervised Decision-Aided Method for OSTBC Systems	144
<i>Paula M. Castro, Ismael Rozas-Ramallal, José A. García-Naya, and Adriana Dapena</i>	
A Decision-Making Model for Environmental Behavior in Agent-Based Modeling	152
<i>Noelia Sánchez-Maroto, Amparo Alonso-Betanzos, Óscar Fontenla-Romero, Miguel Rodríguez-García, Gary Polhill, and Tony Craig</i>	
Version of the New SHA Standard Applied to Manage Certificate Revocation in VANETs	161
<i>Francisco Martín-Fernández and Pino Caballero-Gil</i>	
System Identification of High Impact Resistant Structures	169
<i>Yeesock Kim, K. Sarp Arsava, and Tahar El-Korchi</i>	
Spikes Monitors for FPGAs, an Experimental Comparative Study	179
<i>Elena Cerezuola-Escudero, Manuel Jesus Domínguez-Morales, Ángel Jiménez-Fernández, Rafael Paz-Vicente, Alejandro Linares-Barranco, and Gabriel Jiménez-Moreno</i>	
On Second Language Tutoring through Womb Grammars	189
<i>Leonor Becerra Bonache, Veronica Dahl, and J. Emilio Miralles</i>	

Hybrid Intelligent Systems

Simulated Annealing for Real-Time Vertical-Handoff in Wireless Networks	198
<i>María D. Jaráiz-Simon, Juan A. Gómez-Pulido, Miguel A. Vega-Rodríguez, and Juan M. Sánchez-Pérez</i>	
A Fuzzy Tabu Search Approach to Solve a Vehicle Routing Problem	210
<i>Kaj-Mikael Björk and József Mezei</i>	

Improved Particle Swarm Optimization Method in Inverse Design Problems	218
<i>Y. Volkan Pehlivanoglu</i>	
Solving the Unknown Complexity Formula Problem with Genetic Programming	232
<i>Rayco Batista, Eduardo Segredo, Carlos Segura, Coromoto León, and Casiano Rodríguez</i>	
Three Alternatives for Parallel GPU-Based Implementations of High Performance Particle Swarm Optimization	241
<i>Rogério M. Calazan, Nadia Nedjah, and Luiza de Macedo Mourelle</i>	
A Particle-Swarm-Optimized Fuzzy Classifier Used for Investment Decision Support	253
<i>Lars Krueger and Matthias Walter</i>	
Ant Colony Optimization Inspired Algorithm for 3D Object Segmentation	262
<i>Rafael Arnay and Leopoldo Acosta</i>	

Kernel Methods and SVM

Kernelizing the Proportional Odds Model through the Empirical Kernel Mapping	270
<i>María Pérez-Ortiz, Pedro Antonio Gutiérrez, Manuel Cruz-Ramírez, Javier Sánchez-Monedero, and César Hervás-Martínez</i>	
Parallel Support Vector Data Description	280
<i>Phuoc Nguyen, Dat Tran, Xu Huang, and Wanli Ma</i>	
Antinoise Texture Retrieval Based on PCNN and One-Class SVM	291
<i>Le Tian, Yi-De Ma, Li Liu, and Kun Zhan</i>	

Learning and Adaptation

A FPGA Spike-Based Robot Controlled with Neuro-inspired VITE	299
<i>Fernando Perez-Peña, Arturo Morgado-Estevez, Alejandro Linares-Barranco, Angel Jiménez-Fernández, Juan Lopez-Coronado, and Jose Luis Muñoz-Lozano</i>	
A Cognitive Approach for Robots' Autonomous Learning	309
<i>Dominik M. Ramík, Kurosh Madani, and Christophe Sabourin</i>	
Self-Organizing Incremental Neural Network (SOINN) as a Mechanism for Motor Babbling and Sensory-Motor Learning in Developmental Robotics	321
<i>Tarek Najjar and Osamu Hasegawa</i>	

Alternative OVA Proposals for Cooperative Competitive RBFN Design in Classification Tasks	331
<i>Francisco Charte Ojeda, Antonio Jesús Rivera Rivas, María Dolores Pérez-Godoy, and María Jose del Jesus</i>	
Committee C-Mantec: A Probabilistic Constructive Neural Network	339
<i>José Luis Subirats, Rafael Marcos Luque-Baena, Daniel Urda, Francisco Ortega-Zamorano, José Manuel Jerez, and Leonardo Franco</i>	

**Mathematical and Theoretical Methods in
Computational Intelligence**

Secure Semi-supervised Vector Quantization for Dissimilarity Data	347
<i>Xibin Zhu, Frank-Michael Schleif, and Barbara Hammer</i>	
Border-Sensitive Learning in Kernelized Learning Vector Quantization	357
<i>Marika Kästner, Martin Riedel, Marc Strickert, Wieland Hermann, and Thomas Villmann</i>	
Smoothed Emphasis for Boosting Ensembles	367
<i>Anas Ahachad, Adil Omari, and Aníbal R. Figueiras-Vidal</i>	
F-Measure as the Error Function to Train Neural Networks	376
<i>Joan Pastor-Pellicer, Francisco Zamora-Martínez, Salvador España-Boquera, and María José Castro-Bleda</i>	
Isomorphisms of Fuzzy Sets and Cut Systems	385
<i>Jiří Močkoř</i>	
Hierarchical Modified Regularized Least Squares Fuzzy Support Vector Regression through Multiscale Approach	393
<i>Arindam Chaudhuri</i>	
Minimal Learning Machine: A New Distance-Based Method for Supervised Learning	408
<i>Amauri Holanda de Souza Junior, Francesco Corona, Yoan Miche, Amaury Lendasse, Guilherme A. Barreto, and Olli Simula</i>	
Extending Extreme Learning Machine with Combination Layer	417
<i>Dušan Sovilj, Amaury Lendasse, and Olli Simula</i>	

**Data Mining with Evolutionary Computation and
ANN**

Texture Classification Using Kernel-Based Techniques	427
<i>Carlos Fernandez-Lozano, Jose A. Seoane, Marcos Gestal, Tom R. Gaunt, and Colin Campbell</i>	

A Genetic Algorithms-Based Approach for Optimizing Similarity Aggregation in Ontology Matching	435
<i>Marcos Martínez-Romero, José Manuel Vázquez-Naya, Francisco Javier Nóvoa, Guillermo Vázquez, and Javier Pereira</i>	
Automatic Fish Segmentation on Vertical Slot Fishways Using SOM Neural Networks	445
<i>Álvaro Rodríguez, Juan R. Rabuñal, María Bermúdez, and Alejandro Pazos</i>	
Clustering of Gene Expression Profiles Applied to Marine Research	453
<i>Vanessa Aguiar-Pulido, Victoria Suárez-Ulloa, Daniel Rivero, José M. Eirín-López, and Julián Dorado</i>	
Genetic Programming to Improvement FIB Model: Bond and Anchorage of Reinforcing Steel in Structural Concrete	463
<i>Juan Luis Pérez, Ismael Vieito, Juan Rabuñal, and Fernando Martínez-Abella</i>	
Rainfall Forecasting Based on Ensemble Empirical Mode Decomposition and Neural Networks	471
<i>Juan Beltrán-Castro, Juliana Valencia-Aguirre, Mauricio Orozco-Alzate, Germán Castellanos-Domínguez, and Carlos M. Travieso-González</i>	
Self Organizing Network	
Self-regulating Neurons in the Sensorimotor Loop	481
<i>Frank Pasemann</i>	
Comparison of Two Memristor Based Neural Network Learning Schemes for Crossbar Architecture	492
<i>Janusz A. Starzyk and Basawaraj</i>	
Geometrical Complexity of Data Approximators	500
<i>Evgeny M. Mirkes, Andrei Zinovyev, and Alexander N. Gorban</i>	
Self-Organization Process in Large Spiking Neural Networks Leading to Formation of Working Memory Mechanism	510
<i>Mikhail Kiselev</i>	
Self-organized Learning by Self-Enforcing Networks	518
<i>Christina Klüver and Jürgen Klüver</i>	
Network Anomaly Detection with Bayesian Self-Organizing Maps	530
<i>Emiro de la Hoz Franco, Andrés Ortiz García, Julio Ortega Lopera, Eduardo de la Hoz Correa, and Alberto Prieto Espinosa</i>	

Advances in Computational Intelligence

A Novel Neural Network Parallel Adder	538
<i>Fangyue Chen, Guangyi Wang, Guanrong Chen, and Qinbin He</i>	
Improved Swap Heuristic for the Multiple Knapsack Problem	547
<i>Yacine Laalaoui</i>	
Maximum Margin Clustering for State Decomposition of Metastable Systems	556
<i>Hao Wu</i>	
Hybrid Approach for 2D Strip Packing Problem Using Genetic Algorithm	566
<i>Jaya Thomas and Narendra S. Chaudhari</i>	
Sea Clutter Neural Network Classifier: Feature Selection and MLP Design	575
<i>Jose Luis Bárcena-Humanes, David Mata-Moya, María Pilar Jarabo-Amores, Nerea del-Rey-Maestre, and Jaime Martín-de-Nicolás</i>	
SONN and MLP Based Solutions for Detecting Fluctuating Targets with Unknown Doppler Shift in Gaussian Interference	584
<i>David Mata-Moya, María Pilar Jarabo-Amores, Nerea del-Rey-Maestre, Jose Luis Bárcena-Humanes, and Jaime Martín-de-Nicolás</i>	
An Ensemble of Computational Intelligence Models for Software Maintenance Effort Prediction	592
<i>Hamoud Aljamaan, Mahmoud O. Elish, and Irfan Ahmad</i>	
Sleep Stage Classification Using Advanced Intelligent Methods	604
<i>José Manuel Sánchez Pascualvaca, Carlos Fernandes, Alberto Guillén, Antonio M. Mora, Rogerio Largo, Agostinho C. Rosa, and Luis Javier Herrera</i>	
An n -Spheres Based Synthetic Data Generator for Supervised Classification	613
<i>Javier Sánchez-Monedero, Pedro Antonio Gutiérrez, María Pérez-Ortiz, and César Hervás-Martínez</i>	
Improving the Classification Performance of Optimal Linear Associative Memory in the Presence of Outliers	622
<i>Ana Luíza Bessa de Paula Barros and Guilherme A. Barreto</i>	
SMBSRP: A Search Mechanism Based on Interest Similarity, Query Relevance and Distance Prediction	633
<i>Fen Wang, Changsheng Xie, Hong Liang, and Xiaotao Huang</i>	

An Unfolding-Based Preprocess for Reinforcing Thresholds in Fuzzy
Tabulation 647
*Pascual Julián-Iranzo, Jesús Medina-Moreno, P.J. Morcillo,
Ginés Moreno, and Manuel Ojeda-Aciego*

Author Index 657

Clustering of Gene Expression Profiles Applied to Marine Research

Vanessa Aguiar-Pulido^{1,*}, Victoria Suárez-Ulloa^{2,*}, Daniel Rivero¹,
José M. Eirín-López², and Julián Dorado¹

¹ Artificial Neural Networks and Adaptive Systems Laboratory (RNASA-IMEDIR),
Information and Communication Technologies Department, Faculty of Informatics,
University of A Coruña, Campus de Elviña, 15071 A Coruña, Spain

² Chromatin Structure and Evolution (CHROMEVOL) Group,
Department of Biological Sciences, Florida International University,
33181 North Miami, FL, USA

{vaguiaar,v.ulloa,drivero,jeirin,julian}@udc.es

Abstract. This work presents the results of applying two clustering techniques to gene expression data from the mussel *Mytilus galloprovincialis*. The objective of the study presented in this paper was to cluster the different genes involved in the experiment, in order to find those most closely related based on their expression patterns. A self-organising map (SOM) and the k-means algorithm were used, partitioning the input data into nine clusters. The resulting clusters were then analysed using Gene Ontology (GO) data, obtaining results that suggest that SOM clusters could be more homogeneous than those obtained by the k-means technique.

Keywords: clustering, microarray, neural networks, data mining, bioinformatics, gene ontology.

1 Introduction

Gene expression can be defined as the process by which information from a gene is used in the synthesis of a functional gene product, which is often a protein. Measuring this activity or expression for thousands of genes at once enables creating a global picture of cellular function, which is known as gene expression profiling.

Microarrays [1, 2] are tools widely used to analyse gene expression profiles of a large number of genes simultaneously. This method is an approach to the quantitative analysis of the proteins being produced under given environmental and physiological circumstances, assuming that each gene would produce one single type of protein, which is not absolutely accurate but is widely accepted as an approximation [3].

The application of clustering techniques to this kind of data allows identifying non-obvious relationships between genes such as co-expression phenomena [4, 5]. This approach represents a valuable contribution to marine research since molecular data

* The first two authors contributed equally to this paper.

remains scarce for this kind of organisms, despite their environmental relevance, especially regarding seawater pollution monitoring.

The mussel *Mytilus galloprovincialis* is considered an excellent sentinel organism in coastal environmental control given its sessile condition, ubiquity and extremely high seawater filtering rate [6-8]. Although much effort is being placed on the sequencing of the genomes of mussels and other molluscs, there is still an important gap in the knowledge of these marine invertebrates [9].

The present work contributes to this goal by covering technical aspects of molecular data management which are relevant for marine biology research. Several helpful and widely known techniques are used here to get an insight into non obvious biological patterns, constituting the basis to many other more sophisticated methods of gene expression analysis that very often rely not just in the experimental quantitative data, but also in qualitative metadata by using Gene Ontology (GO) annotation statistics.

2 Materials and Methods

2.1 *Mytilus Galloprovincialis* Data

For this study, a dataset previously published by Banni et al. [10] was used. This data was obtained as a result of an expression profiling experiment by array and it represents temporal expression analysis of female digestive gland tissue from the mussel *M. galloprovincialis*. It contains 11 gene expression records from 295 genes. The data was retrieved from the Gene Expression Omnibus (GEO) database and it can be accessed at the following link: <http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE23052>.

2.2 Clustering Techniques

In this study, the performance of two clustering techniques was compared: a *Self-Organising Map (SOM)* [11, 12] and the *k-means* [13] algorithm. These techniques have been applied over the time to solve a variety of problems in many different environments, obtaining good results [14-27]. Both techniques were implemented using Matlab and several configurations were tested to achieve the results shown in this paper.

2.2.1 Self-Organising Map (SOM)

A SOM is a type of Artificial Neural Network (ANN) which uses unsupervised learning to group instances taken as input, projecting these onto a regular, usually two-dimensional grid called map. In this technique, an instance will be mapped into the node which is nearer to it using some metric. Unlike other ANNs, a neighbourhood function is used in order to preserve the topological properties of the input space.

2.2.2 K-Means

K-means is a method designed for cluster analysis which partitions the instances taken as input into k clusters in such a way that each instance will belong to the cluster with the nearest mean.

2.3 Ontologies

An ontology [28] is a formal representation of knowledge, involving a set of concepts within a domain, and the relationships between pairs of concepts. It can be used to model a domain and support reasoning about entities. Ontologies can be graphically represented as graphs (nodes = concepts; edges = relationships) or trees (nodes and leaves = concepts; branches = relationships, including hierarchical relationships). Ontologies have been widely used, especially in fields related to biomedicine, gaining a lot of attention in the past few years [29-31].

2.3.1 Gene Ontology (GO)

Reported knowledge about genomic data and their products is wide and heterogeneous in nature, however, big efforts have been carried out by specialized consortiums in order to standardize this knowledge. This has been done by defining specific terms and relationships among them, so the gene attributes are then described in a more machine-like manner. This approach allows for the application of Knowledge Discovery in Databases techniques, very useful in functional analysis of massive genomic data. Furthermore, Evidence Codes are used to account for the reliability of the annotations, and weights are established for analysis automatization.

Gene Ontology covers three key aspects in gene description: the Biological Process it takes part in, Molecular Function of its corresponding gene product and Cellular compartment referring to the specific cellular location where it mainly displays its action. The GO terms belonging to any of these three categories have ancestor (lower level) - descendent (higher level) relationships between them, becoming more specific and informative the higher the term's level is. This ontology structure complies with the general build up of ontologies, with GO terms being represented as nodes and relationships as branches. A thumb rule in ontology, that also applies to GO, is that if a gene is annotated with a specific GO term, the correspondence with all its ancestor terms is automatically inferred. This has critical implications for the functional analysis of datasets that analysis tools, such as those embedded in the Blast2GO suite, have taken into account.

2.3.2 Blast2GO

Blast2GO [32-35] is a software suite designed for functional annotation of genomic sequences using GO terminology and for the analysis of such annotation data. In this paper, genes were annotated specifically with terms belonging to the Biological Process type, and the resulting clusters were analysed using the statistical tools provided by the Blast2GO software in order to obtain those terms that are more representative of each clustered gene set.

Blast2GO ranks the GO terms related to the sequences in each set based on scores. Scores are calculated out of the number of sequences annotated with a given term and the distance (number of intermediate nodes) from the GO term directly assigned to the sequence to that one that is being scored. This way, the fact that more general GO terms are more likely to get high scores as they add up all sequences annotated by descendent nodes, is compensated by the fact of getting penalized by the distance to the actual term reported as gene annotation. Therefore, the score is calculated according to the following formula:

$$score = \sum_{gos} seq \times \alpha^{dist}$$

where *seq* is the number of different sequences annotated at a child GO term, *dist* the distance to the node of the child and α is a constant parameter set to default value 0.6.

3 Results and Discussion

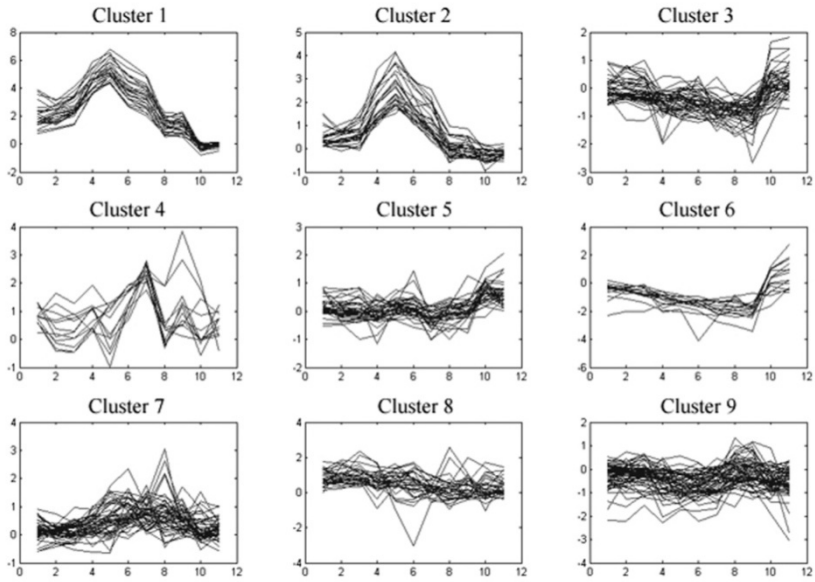
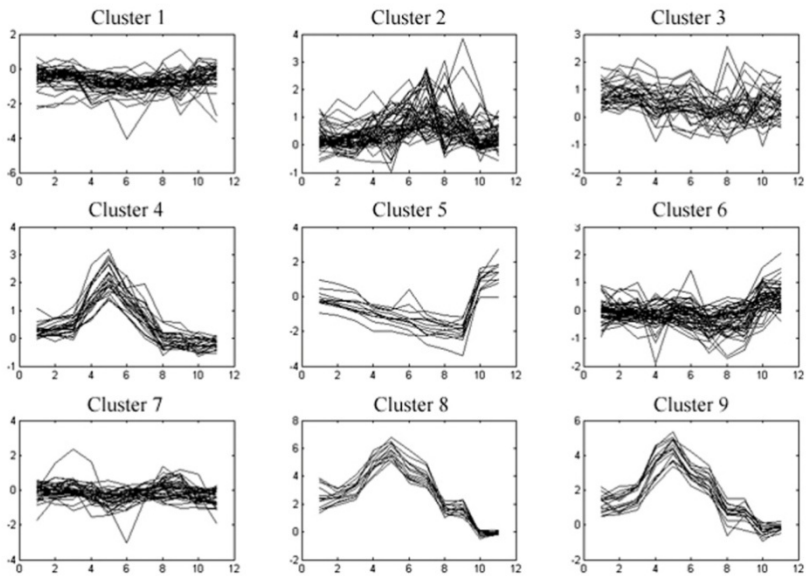
Many tests were run until obtaining the best configuration parameters of the two techniques used in this paper. The same distance metric was used for both methods so that results could be fairly compared. The metric that was finally used was the Euclidean distance. In the case of the SOM technique, different architectures were tested in order to choose the one that obtained the best clusters, that is, an architecture involving 9 neurons. As for the k-means technique, several cluster numbers were tested and, finally, k=9 seemed to obtain the best partitions.

Banni et al. [10] also used the k-means algorithm for the computation of different gene expression trends, obtaining similar results to those presented here. The authors of this paper obtained ten clusters but concluded that two of those clusters could be merged into one.

Fig. 1 and Fig. 2 show the differences between the performances of the proposed models. The horizontal axis represents time (in months) and the vertical axis represents the expression level. Although there are resemblances in the results obtained by both techniques, some genes are clustered into different partitions. Observing these figures, we can study the behaviour of the different genes in terms of gene expression over the time for each cluster and for each technique.

Analysing these graphics and the genes contained in each cluster, we found the following:

- the k-means technique divided cluster 1 obtained by the SOM technique into two different clusters (clusters 8 and 9)
- most of the elements contained in cluster 2 obtained by the SOM, were part of k-means' cluster 4
- clusters 3 and 4 of SOM corresponded to cluster 6 of k-means
- clusters 4 and 7 of SOM corresponded to cluster 2 of k-means
- cluster 6 of SOM and cluster 5 of k-means were very similar, the same happens for cluster 8 of SOM and cluster 3 of k-means.

**Fig. 1.** SOM clustering**Fig. 2.** K-means clustering

Results are mainly consistent between both techniques regarding the biological sense of the obtained clusters. Table 1 and Fig. 3 show the most relevant results of the GO term analysis for each cluster, presenting the levels they belong to as a way of measuring how specific the terms are and how informative they get. It is worth highlighting that in the case of k-means clusters, there are two gene sets that have not obtained any representative GO term by failing to achieve the minimum score threshold set by default in Blast2GO analysis tool, while this happens for only one of the clusters obtained by the SOM technique.

Table 1. TOP 3 Highest scored GO-terms level

Cluster #	K-means	SOM
1	10, 9, 8	N/A
2	6,5,2	6,5,1
3	5,4,3	10,8,8
4	6,5,1	7,6,5
5	6,5,4	8,3,2
6	2,3,8	6,5,4
7	6,5,4	6,5,2
8	N/A	4,5,3
9	N/A	6,3,1

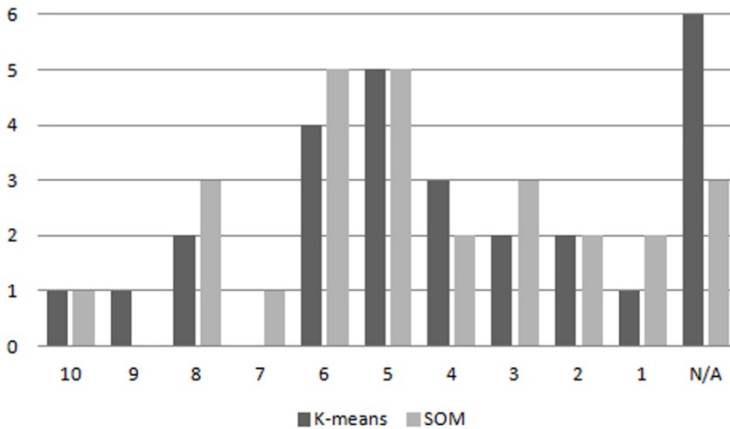


Fig. 3. GO-term level occurrence

Cases such as the SOM cluster 2 and k-means cluster 4, that are similar in terms of included genes and profile shape, obtain very similar GO terms statistics, having the same GO term "translation" (GO:0006412) as the most representative of the group. As an example of clusters displaying different ontological representation, we find the case of SOM clusters 6 and 3, and k-means clusters 5 and 6, all alike in the profile shapes. While SOM cluster 6 throws the same ontological results as k-means cluster number 5, there exist significant differences between SOM cluster 3 and k-means cluster 6. For the first one (SOM cluster 3), the most representative GO terms are terms of high specificity, belonging to the GO levels 10-8 (being the higher the GO level, the more specific the term), while for the latter (k-means cluster 6), the GO terms that obtained the highest representation belong to the levels 2-3, meaning that these are rather general terms not so informative of the biological meaning of this group of genes. This can be understood as being SOM clusters more homogeneous than those obtained by the k-means technique, since more specific GO terms are obtained meaning that more sequences are directly annotated by terms with a closer relationship. However, ontological analysis has the drawback of low statistical significance due to the general lack of functional information for these sequences.

4 Conclusions and Future Work

This work presents a study of gene expression analysis using data from a mussel species obtained from a year-long experiment. Two techniques, a self-organising map (SOM) and the k-means algorithm, were used in order to partition 295 genes with 11 gene expression records over the time into nine clusters. These clusters were then annotated and analysed, obtaining results that suggest that SOM clusters could be more homogeneous than those obtained by the k-means technique.

As future work, we plan to apply more techniques to this type of data, such as biclustering.

Acknowledgements. Vanessa Aguiar-Pulido acknowledges the funding support for a research position by the "Plan I2C" Program from Xunta de Galicia (Spain), partially funded by the European Social Fund (ESF). José M. Eirín-López was awarded with grants by the Spanish Ministry of Economy and Competitiveness (CGL2011-24812 & Ramon y Cajal Subprogramme) and by the Xunta de Galicia (10-PXIB-103-077-PR). Finally, the following projects also supported the work presented: "Galician Network for Colorectal Cancer Research" (REGICC, Ref. 2009/58) from the General Directorate of Research, Development and Innovation of Xunta de Galicia, "Ibero-American Network of the Nano-Bio-Info-Cogno Convergent Technologies", Ibero-NBIC Network (209RT-0366) funded by CYTED (Spain), grant Ref. PIO52048, RD07/0067/0005 funded by the Carlos III Health Institute.

References

1. Schena, M., Shalon, D., Davis, R.W., Brown, P.O.: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470 (1995)
2. Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O., Davis, R.W.: Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. U. S. A.* 94, 13057–13062 (1997)
3. Gibson, G.: Microarray Analysis. *PLoS Biol.* 1, e15 (2003)
4. Lee, H.K., Hsu, A.K., Sajdak, J., Qin, J., Pavlidis, P.: Coexpression analysis of human genes across many microarray data sets. *Genome Res.* 14, 1085–1094 (2004)
5. Chou, J.W., Zhou, T., Kaufmann, W.K., Paules, R.S., Bushel, P.R.: Extracting gene expression patterns and identifying co-expressed genes from microarray data reveals biologically responsive processes. *BMC Bioinformatics* 8, 427 (2007)
6. Kock, W.C.: Monitoring bio-available marine contaminants with mussels (*Mytilus edulis* L) in the Netherlands. *Environ. Monit. Assess.* 7, 209–220 (1986)
7. Laffon, B., Rabade, T., Pasaro, E., Mendez, J.: Monitoring of the impact of Prestige oil spill on *Mytilus galloprovincialis* from Galician coast. *Environ. Int.* 32, 342–348 (2006)
8. Scarpato, A., Romanelli, G., Galgani, F., Andral, B., Amici, M., Giordano, P., Caixach, J., Calvo, M., Campillo, J.A., Albadalejo, J.B., Cento, A., BenBrahim, S., Sammari, C., Deudero, S., Boulahdid, M., Giovanardi, F.: Western Mediterranean coastal waters—monitoring PCBs and pesticides accumulation in *Mytilus galloprovincialis* by active mussel watching: the *Mytilos* project. *J. Environ. Monit.* 12, 924–935 (2010)
9. Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong, Z., Que, H., Xie, Y., Holland, P.W., Paps, J., Zhu, Y., Wu, F., Chen, Y., Wang, J., Peng, C., Meng, J., Yang, L., Liu, J., Wen, B., Zhang, N., Huang, Z., Zhu, Q., Feng, Y., Mount, A., Hedgecock, D., Xu, Z., Liu, Y., Domazet-Loso, T., Du, Y., Sun, X., Zhang, S., Liu, B., Cheng, P., Jiang, X., Li, J., Fan, D., Wang, W., Fu, W., Wang, T., Wang, B., Zhang, J., Peng, Z., Li, Y., Li, N., Chen, M., He, Y., Tan, F., Song, X., Zheng, Q., Huang, R., Yang, H., Du, X., Chen, L., Yang, M., Gaffney, P.M., Wang, S., Luo, L., She, Z., Ming, Y., Huang, W., Huang, B., Zhang, Y., Qu, T., Ni, P., Miao, G., Wang, Q., Steinberg, C.E., Wang, H., Qian, L., Liu, X., Yin, Y.: The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490, 49–54 (2012)
10. Banni, M., Negri, A., Mignone, F., Boussetta, H., Viarengo, A., Dondero, F.: Gene expression rhythms in the mussel *Mytilus galloprovincialis* (Lam.) across an annual cycle. *PLoS One* 6, e18904 (2011)
11. Kohonen, T.: Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 43, 59–69 (1982)
12. Kohonen, T.: Essentials of the self-organizing map. *Neural Netw.* 37, 52–65 (2013)
13. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.*, pp. 281–297. Univ. of Calif. Press (1965)
14. Yan, A., Hu, X., Wang, K., Sun, J.: Discriminating of ATP competitive Src kinase inhibitors and decoys using self-organizing map and support vector machine. *Mol Divers* 17, 75–83 (2013)
15. Wang, L., Wang, M., Yan, A., Dai, B.: Using self-organizing map (SOM) and support vector machine (SVM) for classification of selectivity of ACAT inhibitors. *Mol Divers* 17, 85–96 (2013)

16. Zhu, D., Huang, H., Yang, S.X.: Dynamic Task Assignment and Path Planning of Multi-AUV System Based on an Improved Self-Organizing Map and Velocity Synthesis Method in Three-Dimensional Underwater Workspace. *IEEE Trans. Syst. Man Cybern. B Cybern.* (2012)
17. Piastra, M.: Self-organizing adaptive map: Autonomous learning of curves and surfaces from point samples. *Neural Netw.* (2012)
18. Marique, T., Allard, O., Spanoghe, M.: Use of Self-Organizing Map to Analyze Images of Fungi Colonies Grown from *Triticum aestivum* Seeds Disinfected by Ozone Treatment. *Int. J. Microbiol.*, 865175 (2012)
19. Bae, M.J., Kim, J.S., Park, Y.S.: Evaluation of changes in effluent quality from industrial complexes on the Korean nationwide scale using a self-organizing map. *Int. J. Environ. Res. Public Health* 9, 1182–1200 (2012)
20. Wiggins, J.L., Peltier, S.J., Ashinoff, S., Weng, S.J., Carrasco, M., Welsh, R.C., Lord, C., Monk, C.S.: Using a self-organizing map algorithm to detect age-related changes in functional connectivity during rest in autism spectrum disorders. *Brain Res.* 1380, 187–197 (2011)
21. Yang, Z., Wu, Z., Yin, Z., Quan, T., Sun, H.: Hybrid Radar Emitter Recognition Based on Rough k-Means Classifier and Relevance Vector Machine. *Sensors (Basel)* 13, 848–864 (2013)
22. Sun, G., Hakozaiki, Y., Abe, S., Vinh, N.Q., Matsui, T.: A novel infection screening method using a neural network and k-means clustering algorithm which can be applied for screening of unknown or unexpected infectious diseases. *J. Infect.* 65, 591–592 (2012)
23. Armstrong, J.J., Zhu, M., Hirdes, J.P., Stolee, P.: K-means cluster analysis of rehabilitation service users in the Home Health Care System of Ontario: examining the heterogeneity of a complex geriatric population. *Arch. Phys. Med. Rehabil.* 93, 2198–2205 (2012)
24. Stricker, M.D., Onland-Moret, N.C., Boer, J.M., van der Schouw, Y.T., Verschuren, W.M., May, A.M., Peeters, P.H., Beulens, J.W.: Dietary patterns derived from principal component- and k-means cluster analysis: Long-term association with coronary heart disease and stroke. *Nutr. Metab. Cardiovasc. Dis.* (2012)
25. Konicek, A.R., Lefman, J., Szakal, C.: Automated correlation and classification of secondary ion mass spectrometry images using a k-means cluster method. *Analyst* 137, 3479–3487 (2012)
26. Chang, N.B., Wimberly, B., Xuan, Z.: Identification of spatiotemporal nutrient patterns in a coastal bay via an integrated k-means clustering and gravity model. *J. Environ. Monit.* 14, 992–1005 (2012)
27. Zhang, S., Jin, W., Huang, Y., Su, W., Yang, J., Feng, Z.: Profiling a *Caenorhabditis elegans* behavioral parametric dataset with a supervised K-means clustering algorithm identifies genetic networks regulating locomotion. *J. Neurosci. Methods* 197, 315–323 (2011)
28. Gruber, T.R.: A translation approach to portable ontology specifications. *Knowl. Acquis.* 5, 199–220 (1993)
29. Warita, K., Mitsuhashi, T., Tabuchi, Y., Ohta, K., Suzuki, S., Hoshi, N., Miki, T., Takeuchi, Y.: Microarray and gene ontology analyses reveal downregulation of DNA repair and apoptotic pathways in diethylstilbestrol-exposed testicular Leydig cells. *J. Toxicol. Sci.* 37, 287–295 (2012)
30. Schaid, D.J., Sinnwell, J.P., Jenkins, G.D., McDonnell, S.K., Ingle, J.N., Kubo, M., Goss, P.E., Costantino, J.P., Wickerham, D.L., Weinshilboum, R.M.: Using the gene ontology to scan multilevel gene sets for associations in genome wide association studies. *Genet. Epidemiol.* 36, 3–16 (2012)

31. Ma, N., Zhang, Z.G.: Evaluation of clustering algorithms for gene expression data using gene ontology annotations. *Chin. Med. J (Engl.)* 125, 3048–3052 (2012)
32. Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M.: Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676 (2005)
33. Conesa, A., Gotz, S.: Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics*, 619832 (2008)
34. Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., Conesa, A.: High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435 (2008)
35. Gotz, S., Arnold, R., Sebastian-Leon, P., Martin-Rodriguez, S., Tischler, P., Jehl, M.A., Dopazo, J., Rattei, T., Conesa, A.: B2G-FAR, a species-centered GO annotation repository. *Bioinformatics* 27, 919–924 (2011)