Recent Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics
Volume II: Applications

Editors

Krassimir T. Atanassov
Władysław Homenda
Olgierd Hryniewicz
Janusz Kacprzyk
Maciej Krawczak
Zbigniew Nahorski
Eulalia Szmidt
Sławomir Zadrożny

SRI PAS IBS PAN

Recent Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics Volume II: Applications

Systems Research Institute
Polish Academy of Sciences

Recent Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics
Volume II: Applications

Editors

Krassimir T. Atanassov
Władysław Homenda
Olgierd Hryniewicz
Janusz Kacprzyk
Maciej Krawczak
Zbigniew Nahorski
Eulalia Szmidt
Sławomir Zadrożny

IBS PAN **SRI PAS**

Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
www.ibspan.waw.pl

ISBN 9788389475367

Generalized net model of clustering with self-organizing map

E. Sotirova[1], T. Petkov[1], S. Surchev[1] and M. Krawczak[2]
[1] Prof. Asen Zlatarov University, Burgas-8010, Bulgaria
esotirova@btu.bg
[2] Systems Research Institute – Polish Academy of Sciences,
Wyzsza Szkola Informatyki Stosowanej i Zarzadzania, Warsaw, Poland
krawczak@ibspan.waw.pl

Abstract

The proposed generalized net model present the process of clustering with a self-organizing map (SOM). Here we use SOM as an effective software tool for the visualization of high-dimensional data. One the other hand, data mining is a process of discovering various models, summaries, and derived values from a given collection of data.

Keywords: generalized nets, self-organizing map, clustering.

1. Introduction

Data mining is a process of discovering various models, summaries, and derived values from a given collection of data [4]. Data mining techniques have been developed to turn data into useful task oriented knowledge. Associations reflect relationships among items in databases, and have been widely studied in the fields of knowledge discovery and data mining. Most algorithms for mining association rules identify relationships among transactions using binary values. Transactions with quantitative values and items are, however, commonly seen in real-world applications.

According to [5], the self-organizing map (SOM) is an effective software tool for visualization of high-dimensional data. It converts complex, nonlinear statistical relationships between high-dimensional data items into simple geometric relationships on a low-dimensional display. As it thereby compresses information while preserving the most important topological and metric relationships of the primary data items on the display, it may also be thought to produce some kind of abstractions. These two aspects, visualization and

Recent Advances in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Volume II: Applications *(K.T. Atanassow, W.Homenda, O.Hryniewicz, J.Kacprzyk, M.Krawczak, Z.Nahorski, E. Szmidt, S. Zadrożny, Eds.), IBS PAN - SRI PAS, Warsaw, 2010.*

abstraction, can be utilized in a number of ways in complex tasks such as process analysis, machine perception, control, and communication.

The SOM usually consists of a two-dimensional regular grid of nodes. A model of some observation is associated with each node.

Clustering is the process of organizing objects into groups whose members are similar in some way [4]. A cluster is therefore a collection of objects which are "similar" and are "dissimilar" to the objects belonging to other clusters. A SOM is a type of artificial intelligence that is trained using unsupervised learning to produce a low-dimensional representation of the input space of the training samples, called a map. The SOM is based on an issue of competitive learning. The net consists of a set *A* with *n* neurons, represented with weight vectors w_i. Furthermore, neurons are mutually interconnected and these bindings form some topological grid.

If we present a pattern *x* into this network, then exactly one neuron could be the winner and its weights are proportionally adapted to the pattern (the neuron is then closer). Neighborhood *N*(*c*) could be formally defined as a set of neurons that are topologically close to the winner. The winner of the competition is determined as the neuron with the minimum distance to the pattern. Then, adaptation of the weights proceeds. Weight vectors for the next iteration of the winner and neurons in the neighborhood are adapted in a way that current weights are modified (either added or subtracted) with a variance of current weight and input pattern. The weight vector of pattern is called template vector of that pattern. The SOM tries to adapt weights of neurons to cover the most dense regions and therefore SOM naturally finds data clusters [8, 9, 10] (Fig. 1).

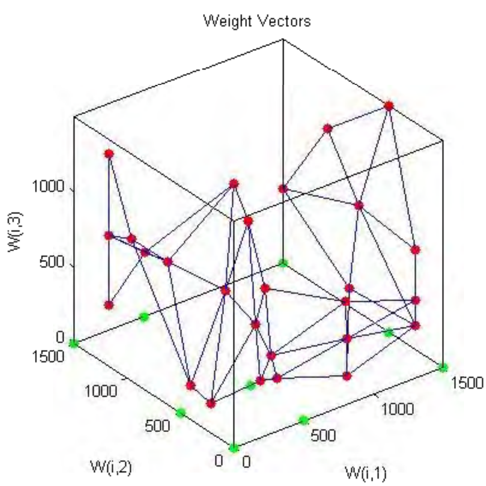

Figure 1

Here we design a Generalized Net model (GN, [1,2]) that presents the work of the process of the SOM clustering [3, 4]. All definitions related to the concept SOM are taken from [6, 7].

2 A GN-model

Initially the following tokens enter in the GN:

- in place L_5 – one token with initial characteristic "Database with SOM structures'';
- in place L_6 – one token with initial characteristic "Number of iterations for learning of the SOM";
- in place L_7 – one token with initial characteristic "Preliminary condition for learning";

The GN is presented on Fig. 2. by a set of transitions:

$$A = \{Z_0, Z_1, Z_2, Z_3, Z_4, Z_5\}.$$

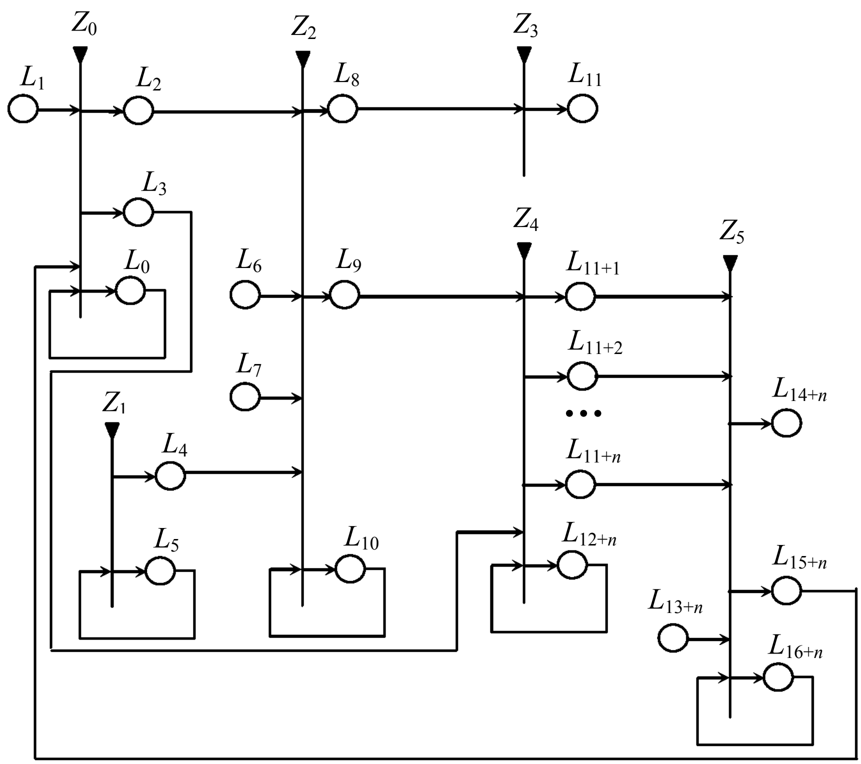

Figure 2: GN model of the clustering with a self-organizing map

These transitions describe the following processes:

- Z_0 = Forming the input data
- Z_1 = Selecting of SOM structure
- Z_2 = Learning of the SOM
- Z_3 = Data visualization
- Z_4 = Testing of the results
- Z_5 = Determination of the correct results

In details, the transitions of the GN-model have the following forms.

$$Z_0 = <\{L_1, L_{15+n}, L_0\}, \{L_2, L_3, L_0\}, R_0, \vee (L_1, L_{15+n}, L_0)>,$$

where

$$R_0 = \begin{array}{c|ccc} & L_2 & L_3 & L_0 \\ \hline L_1 & False & False & True \\ L_{15+n} & False & False & True \\ L_0 & W_{0,2} & W_{0,3} & True \end{array},$$

where

$W_{0,2} = W_{0,3}$ = "The data are normalizing".

Tokens that enter places L_2 and L_3 obtain the characteristics "Normalizing data".

$$Z_1 = <\{L_5\}, \{L_4, L_5\}, R_1, \vee (L_5)>$$

where

$$R_1 = \begin{array}{c|cc} & L_4 & L_5 \\ \hline L_5 & True & W_{4,5} \end{array},$$

where

$W_{4,5}$ = "The structure is chosen".

Tokens that enters place L_5 obtain the characteristic "Selected structure of the SOM".

$$Z_2 = <\{L_2, L_4, L_7, L_4, L_{10}\}, \{L_8, L_9, L_{10}\}, R_2, \vee(L_2, L_4, L_7, L_4, L_{10})>$$

where

$$R_2 = \begin{array}{c|ccc} & L_8 & L_9 & L_{10} \\ \hline L_2 & False & False & True \\ L_6 & False & False & True \\ L_7 & False & False & True \\ L_4 & False & False & True \\ L_{10} & W_{10,8} & W_{10,9} & True \end{array}.$$

where
$W_{10,8} = W_{10,9}$ = "The SOM is learned".

Tokens that enter in places L_8 and L_9 obtain characteristics "Learned SOM".

$$Z_3 = <\{L_8\},\{L_{11}\},R_3,\vee(L_8)>$$

where

$$R_3 = \begin{array}{c|c} & L_{11} \\ \hline L_8 & True \end{array}$$

Tokens that enters position L_{11} obtain the characteristic "Data visualization of the results of the SOM".

$$Z_4=<\{L_9, L_3\},\{L_{11+1}, L_{11+2}, ...,L_{12+n}\}, R_4, \wedge(L_9, L_3)>$$

where

$$R_4 = \begin{array}{c|ccccc} & L_{11+1} & L_{11+2} & \ldots & L_{11+n} & L_{12+n} \\ \hline L_9 & False & False & \ldots & False & True \\ L_3 & False & False & \ldots & False & True \\ L_{12+n} & W_{12+n,11+1} & W_{12+n,11+2} & \ldots & W_{12+n,11+n} & True \end{array}$$

where
$W_{12+n,11+1}$ = "There is a vector that hits the cluster 1";
$W_{12+n,11+2}$ = "There is a vector that hits the cluster 2";
…………………………………………………
$W_{12+n,11+x}$ = "There is a vector that hits the cluster n".

Tokens that enter in positions L_{11+1}, L_{11+2},…, L_{11+n} obtain characteristics "There is a vector that hits the cluster".

$$Z_5 = <\{L_{11+1}, L_{11+2}, \ldots, L_{11+n}, L_{13+n}, L_{16+n}\}, \{L_{14+1}, L_{15+n}, L_{16+n}\}, R_5, \vee(L_{11+1}, L_{11+2},\ldots, L_{11+n}, L_{13+n}, L_{16+n})>$$

where

$$R_5 = \begin{array}{c|ccc} & L_{14+n} & L_{15+n} & L_{16+n} \\ \hline L_{11+1} & False & False & True \\ L_{11+2} & False & False & True \\ \ldots & \ldots & \ldots & \ldots \\ L_{11+n} & False & False & True \\ L_{13+n} & False & False & True \\ L_{16+n} & W_{16+n,14+n} & W_{16+n,15+n} & True \end{array}$$

where
$W_{16+n,14+n}$ = "The SOM is verified and ready for clustering ";
$W_{12+n,11+2}$ = "The SOM is verified, but it is not ready for clustering";

Tokens that enter places L_{14+n} and L_{15+n} obtain characteristics “There is a vector that hits the cluster”.

3 Conclusion

Here we use self-organizing maps as an effective software tool for visualization of high-dimensional data and as a tool for data mining – process of discovering various models, summaries, and derived values from a given collection of data. The proposed GN model presents the process of clustering with SOM.

References

[1] Atanassov, K. Generalized nets, World Scientific, Singapore, New Jersey, London, 1991.

[2] Atanassov, K. On Generalized Nets Theory, “Prof. M. Drinov” Academic Publishing House, Sofia, 2007.

[3] Freeman, A., Neural Network, Algorithms Applications and Programming Techniques. Addison Wesley, Reading, MA, 1991.

[4] Glenn, J., Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining. John Wiley, 2006.

[5] Irini Reljin, Branimir Reljin, Gordana Jovanović, Clustering and Mapping Spatial-Temporal Datasets Using SOM Neural Networks, Journal of automatic control, University of Belgrade; vol. 13(1), 2003, 55–60.

[6] Kohonen, T. Self-Organizing Maps. Series in Information Sciences, Vol. 30. Springer, Heidelberg. 1995, second ed. 1997.

[7] Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J. SOM_PAK: The self-organizing map program package. Report A31. Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland, 1996. http://www.cis.hut.fi/research/som_lvq_pak.shtml.

[8] Marjan Kaedi, Mohammadali Nematbakhsh, Nasser Ghasem-Aghaee, Fuzzy Association Rule Reduction Using Clustering In SOM Neural Network, IADIS, European Conference Data Mining, 2008, 139–143

[9] Paul Mangiameli, Shaw K. Chen, David West**, A** comparison of SOM neural network and hierarchical clustering methods, Neural Networks and Operations Research/Management Science, European Journal of Operational Research, Volume 93, Issue 2, 1996, 402–417.

[10] Wasserman, P. D., Neural computing, Theory and Practice. Van Nostrand Reinhold, New York, 1989.

The papers presented in this Volume 2 constitute a collection of contributions, both of a foundational and applied type, by both well-known experts and young researchers in various fields of broadly perceived intelligent systems.
It may be viewed as a result of fruitful discussions held during the Ninth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets (IWIFSGN-2010) organized in Warsaw on October 8, 2010 by the Systems Research Institute, Polish Academy of Sciences, in Warsaw, Poland, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences in Sofia, Bulgaria, and WIT - Warsaw School of Information Technology in Warsaw, Poland, and co-organized by: the Matej Bel University, Banska Bystrica, Slovakia, Universidad Publica de Navarra, Pamplona, Spain, Universidade de Tras-Os-Montes e Alto Douro, Vila Real, Portugal, and the University of Westminster, Harrow, UK:

Http://www.ibspan.waw.pl/ifs2010

The consecutive International Workshops on Intuitionistic Fuzzy Sets and Generalized Nets (IWIFSGNs) have been meant to provide a forum for the presentation of new results and for scientific discussion on new developments in foundations and applications of intuitionistic fuzzy sets and generalized nets pioneered by Professor Krassimir T. Atanassov. Other topics related to broadly perceived representation and processing of uncertain and imprecise information and intelligent systems have also been included. The Ninth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets (IWIFSGN-2010) is a continuation of this undertaking, and provides many new ideas and results in the areas concerned.

We hope that a collection of main contributions presented at the Workshop, completed with many papers by leading experts who have not been able to participate, will provide a source of much needed information on recent trends in the topics considered.

ISBN-13 9788389475367
ISBN 838947536-7
9 788389 475367