8th Int. Workshop on IFSs, Banská Bystrica, 9 Oct. 2012
Notes on Intuitionistic Fuzzy Sets
Vol. 18, 2012, No. 4, 26–31

Publication's assessment with intuitionistic fuzzy estimations

Evdokia Sotirova[1], Sotir Sotirov[1], Ivelina Vardeva[1] and Beloslav Riečan[2]

[1] Prof. Asen Zlatarov University
1 "Prof. Yakimov" Blvd, Burgas–8000, Bulgaria,
e-mails: {esotirova, ssotirov}@btu.bg

[2] Faculty of Natural Sciences, Matej Bel University
Tajovského 40, SK–974 01 Banská Bystrica, Slovakia
Mathematical Institute of Slovak Acad. of Sciences
Štefánikova 49, SK–81473 Bratislava, Slovakia
e-mails: riecan@fpv.umb.sk, riecan@mat.savba.sk

Abstract: In the paper we investigate how to apply some data mining techniques for clustering and classification the assessment of the different publications and articles. For this aim we propose to use neural network and decision tree to analyze given collection of data. We use the Intuitionistic fuzzy estimation as an input vector for the self organizing map that gives us 6 clusters. To predict the next data we must have the rules that can be obtained from the decision tree.

Keywords: Learning system, Data mining, Decision tree, Neural network.

AMS Classification: 03E72.

1 Introduction

The reviewing and rating process of the publications and articles is very important for assessment of its quality. The reviewers have to prepare anonymous feedback. They give clear responds of the different questions like: well-written, grammatical errors, objective, complete, originality of contents, significance for theory and others.

The idea in the present paper is to combine data mining techniques and publication's assessment in intuitionistic fuzzy form to use the advantages of both different tools [3]. It is a continuation of the series researches of the some of the authors connected with different processes of functioning of universities, collated in [13], [14].

Some of the most popular data mining techniques are association, classification, clustering and prediction. The neural networks are one of the basic tools for realization of data mining techniques.

A neural network is presented in [10] and it determines the degree of assimilation, non assimilation and uncertainty in the students answers' evaluation based on defined set of criterions. The network could be used for the evaluation of the students' answers in the closed tests in training systems. A multilayer perceptron is used in [11] for obtaining lecturers' evaluation with intuitionistic fuzzy estimations. It determines the degree of approve, non-approve and uncertainty in the lecturers based on the students answers' investigations. The evaluation is based on defined set of criterions. A self-organizing map (SOM) is used to data mine the students' answers and facilitate their analysis by grouping the similar answers in [12], i.e. clustering.

In the present paper is used a decision tree for classifying the publication's assessments obtained in intuitionistic fuzzy form in the outputs of the SOM. Decision trees are a simple, but powerful form of multiple variable analyses. They provide capabilities to supplement a variety of data mining tools and techniques such as neural networks.

The publications are evaluated using different criteria. Let we have n publications and m criteria. The assessment for the degree of the acceptance (μ) and the non-acceptance (ν) of the publication is formed on the basis of these criteria. For publication i, $i = 1, \ldots, n$:

$$\mu_i = \frac{\sum_{j=1}^{m} \varepsilon_i^j}{m}, \quad \nu_i = \frac{\sum_{j=1}^{m} \delta_i^j}{m}$$

where:

- ε_i^j is the degree of the acceptance of the publication i according to criterion j,
- δ_i^j is the degree of the non-acceptance of the publication i according to criterion j,
- m is the total number of criteria.

These assessments, which estimate the degree of the acceptance (μ) and the non-acceptance (ν) of the publication i, are represented by ordered pairs $\langle \mu_i, \nu_i \rangle$ of real numbers from the set $[0, 1]\times[0, 1]$.

The degree of uncertainty $\pi_i = 1 - \mu_i - \nu_i$ represents those cases where the reviewer is currently unable to answer the asked question and needs additional information. Everywhere the ordered pairs have been defined in the sense of intuitionistic fuzzy sets [1, 2].

2 Realization

A SOM is a neural network that uses unsupervised learning. It produces high dimensional data to the low dimensional space (called map) [4, 5, 6, 7, 8, 9]. The main thing to use SOM is representing data. It discretizes representation of the input space of the training samples in the sense that they use a neighborhood function to preserve the topological properties of the input space.

The weight vector of the unit is closest to the current object becomes the winning or active unit. During the training stage, the values for the input variables are gradually adjusted in an attempt to preserve neighborhood relationships that exist within the input data set. As it gets closer to the input object, the weights of the winning unit are adjusted as well as its neighbors.

With SOM, clustering is performed by having several units compete for the current object. Once the data have been entered into the system, the network of artificial neurons is trained by providing information about inputs.

The SOM used for the clustering of the intuitionistic fuzzy assessments has 6 neurons in one layer. The vector for the learning the SOM has 162 different points that describes the working area of the neural network. The number of neurons determines the number of the clusters, in which hit corresponding points. In the two dimensional SOM is used 3×2 neurons, i.e. 6 neurons.

As a test vector is used the estimations for the degree of the acceptance (μ) and the non-acceptance (ν) of the publication. The test vectors enter in the inputs of the SOM. Every test vector hit the cluster that represents typical publications' acceptance. The general view of the 55 test vectors is shown below:

$$\begin{bmatrix} \mu_1 & \nu_1 & \pi_1 \\ \mu_2 & \nu_2 & \pi_2 \\ \vdots & \vdots & \vdots \\ \mu_{55} & \nu_{55} & \pi_{55} \end{bmatrix}$$

The SOM gives in which cluster hit the test vectors, but does not give the rules. A possibility to obtain the rules is to use a classification tree.

The assessments of the publication obtained in the output of the neural network were divided into training set and the validation set. Partitioning the data into training and validation sets is done randomly according to predetermined proportions (60 % for the training set – 33 rows, 40% for the validation set – 22 rows). The training partition contains the data used to build the model. The validation partition is used to assess the performance of the model.

The input variables are: μ, ν and π and the output variable is *Number of cluster*. The classification tree consists of 5 decision nodes and 6 terminal nodes and is shown on Fig. 1. The classification tree was done by XLMiner for Excel [15].

From the tree we obtain six classification rules:

1) IF $\pi \leq 0.3$ AND $\mu > 0.5$ THEN *Number of cluster* = 1;
   This reflects the cases when the publications have a high (maximal or up the middle) assessment for acceptance (μ) and low assessment for the degrees of the non-acceptance (ν) and non-answered questions (π).
2) IF $0.3 < \pi \leq 0.5$ AND $\nu \leq 0.3$ THEN *Number of cluster* = 2;
   This reflects the cases when the publications have a less than average assessment for μ, minimal assessment for ν and under or around the middle assessment for π.
3) IF $\pi > 0.5$ AND $\nu \leq 0.3$ THEN *Number of cluster* = 3;
   This reflects the cases when the publications have greater than average assessment for π and low assessments for μ and ν.

4) IF $\pi \leq 0.3$ AND $\mu \leq 0.5$ THEN *Number of cluster* = 4;
This reflects the cases when the publications have a low and approximately equal assessments for μ and ν and low assessment for π.

5) IF $\pi > 0.3$ AND $\nu > 0.3$ AND $\mu > 0.1$ THEN *Number of cluster* = 5;
This reflects the cases when the publications have approximately equal assessments for ν and π and low assessment for μ.

6) IF $\pi > 0.3$ AND $\nu > 0.3$ AND $\mu \leq 0.1$ THEN *Number of cluster* = 6;
This reflects the cases when the publications have approximately equal assessments for ν and π and very low assessment for μ.

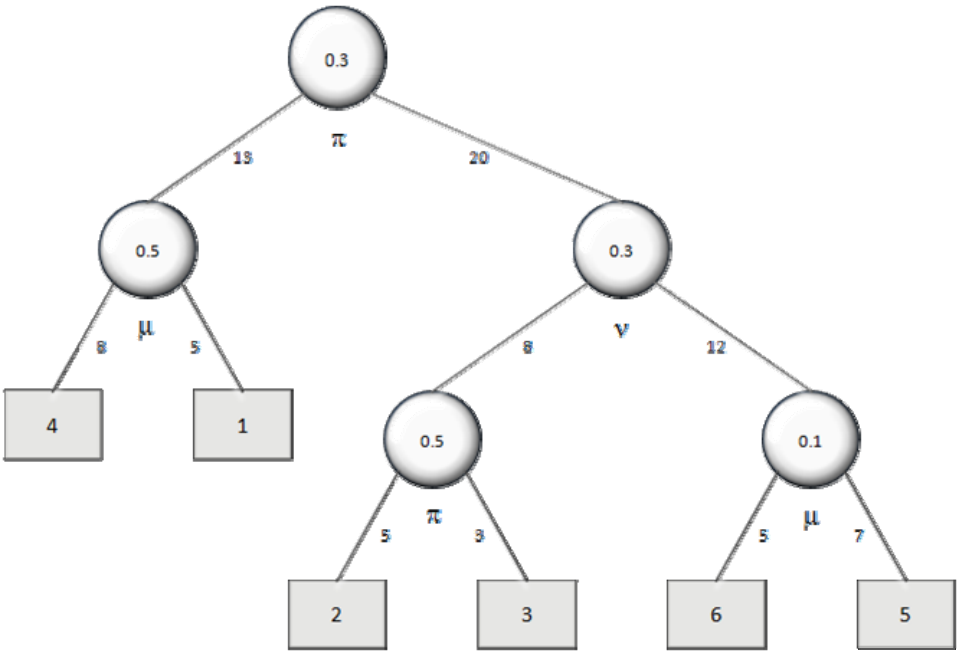

Figure 1: Graphical representation of the classification tree (using training data)

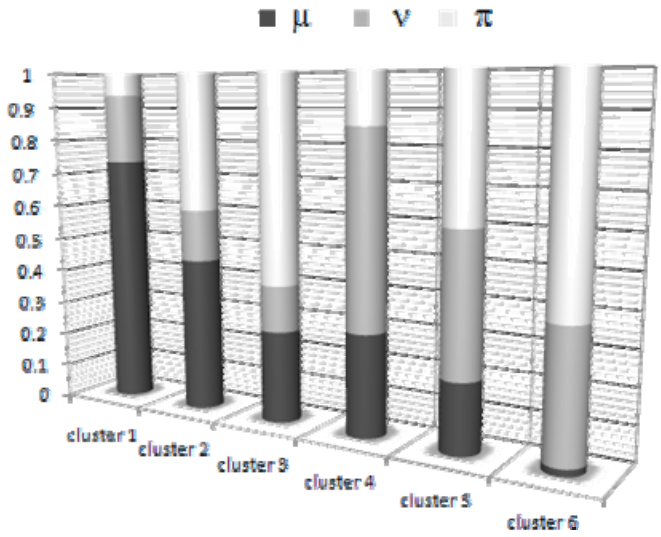

Figure 2: Graphical representation of the average values of the μ, ν and π in clusters (using training data)

3 Conclusion

In the present paper we use self-organizing map for clustering and decision tree classification the assessment of the different publications and articles. The suggested evaluation methods and procedures are intended to make publication's assessments as objective as possible.

From the classification tree we obtain six decision rules for prediction the values of new or unseen observations.

Due to the specificity of creating new materials for publishing the subjective assessment cannot be avoid but should be made objective. That could be achieved by science-based quantitative methods using the instruments of subjective statistics.

Acknowledgment

The authors are grateful for the support provided by the National Science Fund of Bulgaria under grant DID-02-29/2009 for the purchase of the software product *Excel XLMiner*.

References

[1] Atanassov, K., *Intuitionistic Fuzzy Sets*, Springer Physica-Verlag, Berlin, 1999.

[2] Atanassov, K., Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, Vol. 20, 1986, 87–96.

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

[4] Reljin, I., B. Reljin, G. Jovanović, Clustering and Mapping Spatial-Temporal Datasets Using SOM Neural Networks, *Journal of Automatic Control*, University of Belgrade; Vol. 13, 2003, No. 1, 55–60

[5] Kohonen, T., *Self-Organizing Maps. Series in Information Sciences*, Vol. 30. Springer, Heidelberg. Second ed., 1997.

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

[7] Kaedi, M., M. Nematbakhsh, N. Ghasem-Aghaee, Fuzzy Association Rule Reduction Using Clustering in SOM Neural Network, *IADIS, European Conference Data Mining* 2008, 139–143.

[8] 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*, Vol. 93, 1996, Issue 2, 402–417.

[9] Wasserman, P. D., *Neural Computing: Theory and Practice*. Van Nostrand Reinhold, New York, 1989

[10] Sotirov, S., D. Orozova, E. Sotirova, Neural network for defining intuitionistic fuzzy sets in e-learning, *Notes on Intuitionistic Fuzzy Sets*, Vol. 15, 2009, No. 2, 33–36.

[11] Sotirov, S., E. Sotirova, M. Krawczak, Application of data mining in digital university: multilayer perceptron for lecturer's evaluation with intuitionistic fuzzy estimations, *Issues in Intuitionistic Fuzzy Sets and Generalized Nets*, Vol. 8, Warsaw, Poland, 2010, 102–107.

[12] Sotirova, E., Classification of the students' intuitionistic fuzzy estimations by a 3-dimensional self organizing map, *Notes on Intuitionistic Fuzzy Sets*, Vol. 17, 2011, No. 4, 64–68.

[13] Shannon, A., Atanassov K., Orozova D., Krawczak M., Sotirova E., Melo-Pinto P., Petrounias I., Kim T. *Generalized Nets and Information Flow within a University*, Warsaw, Warsaw School of Information Technology, 2007.

[14] Shannon, A., D. Langova-Orozova, E. Sotirova, I. Petrounias, K. Atanassov, M. Krawczak, P. Melo-Pinto, T. Kim. *Generalized Net Modelling of University Processes*. North Sydney: KvB Visual Concepts Pty Ltd, Monograph, No. 7, 2005.

[15] XLMiner Data Mining Add-in For Excel, `www.xlminer.com`.