ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 8, Issue, 06, pp.33087-33089, June, 2016

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

EFFECTS OF DATA PREPROCESSING ON THE EFFICIENCY OF TAXONOMIC NEURAL NETWORKS

*Dragan Popovic, Evagelia Boli, Ankica Hosek-Momirovic, Milos Popovic, Vladimir Savic and Jasna Popovic

Kosanciceva 17, Krusevac, Serbia, Zip code, 37000

ARTICLE INFO

Article History:
Received 07th March, 2016
Received in revised form
23rd April, 2016
Accepted 14th May, 2016
Published online 30th June, 2016

Key words:
Neural Networks,
Classification,
Nonlinear Models.

ABSTRACT

Objectives: For the classification of respondents drawn from different populations and described by different sets of variables, six taxonomic neural networks designed in a similar manner, but supplied with different preprocessors for input data transformation, were applied.

Methods: The standard method used to classify objects was implemented by an Intruder neural network variation (Momirovic, Hosek. Popovic & Boli, 2002) codenamed Intruds. This network only standardizes input data, but does not perform any additional transformations over them.

Results: Each of these networks consisted of Lebart`s adaptive perceptrons (Lebart, 1998) which classified objects by means of a backward error propagation technique at the output end. The initial classification in all the networks was carried out by transformation of a fuzzy classification, defined by the objects` position on the periphery of the hyperellipsoid formed by a semiorthogonal transformation of principal components of variables subjected to a monotonic transformation, into a hard classification. The final classification efficiency was checked by each network based on the outcome of a posteriori classification performed by Fisher`s linear classifiers in the total space of eventually transformed variables.

The neural networks under the basic code name Invader perform nonlinear monotonic transformations over these data. Invader 1 transforms variables into sigmoid (0,1) form, Invader 2 transforms variables into sigmoid (-1,1) form, and Invader 3 – into staircase (-1, 0, 1) form.

The neural network codenamed Bibltax reduces to zero all standardized values which differ from the expected value in the sample of objects by one half of the standardized deviation in order to eliminate noise produced by poorly differentiated objects; and the Exnetg neural network, a variation of the Exelnet neural network (Momirovic, 2002), transforms variables into standardized image form to eliminate noise produced by the unique variance of measurement instruments.

Conclusions: Comparative analysis of the effectiveness of these taxonomic networks clearly shows that, in the all cases, better taxonomic effectiveness, or surer breaking up of the analyzed set of objects, is achieved by means of data preprocessing

Citation: Dragan Popovic, Evagelia Boli, Ankica Hosek-Momirovic, Milos Popovic, Vladimir Savic and Jasna Popovic, 2016. "Effects of data preprocessing on the efficiency of taxonomic neural networks", *International Journal of Current Research*, 8, (06), 33087-33089.

INTRODUCTION

The nature of taxonomic neural networks, and often the nature of the data to be subjected to taxonomic operations, sometimes require appropriate transformation of these data before running a neural network. These are usually cases when distributions of variables which describe objects are incorrect not because the sample of objects is drawn from a heterogeneous population, but because of the characteristics of measurement instruments, including cases when measurement instruments are burdened with higher error variance.

Corresponding author: Dragan Popovic,
Kosanciceva 17, Krusevac, Serbia, Zip code, 37000.

However, often transformation, usually a nonlinear one, is also required because relationships among variables are genuinely nonlinear, as is the case with psychological and anthropological characteristics in general. This situation is particularly frequent in criminological research because of the nature of samples of criminals or potential criminals and because the variables that describe the sample members are derived from measurement instruments whose reliability is insufficient and homogeneity is low, and also because psychic characteristics, a dominant factor of criminal behavior, are connected by complex nonlinear relationships. Therefore, the aim of this research was to compare, on several criminologically relevant examples, the efficiency of taxonomic neural networks supplied with different preprocessors, as well as to compare their efficiency with the efficiency of a standard neural network proved to be more efficient than common taxonomic algorithms used in psychological and criminological research.

MATERIALS AND METHODS

For the classification of respondents drawn from different populations and described by different sets of variables, six taxonomic neural networks designed in a similar manner, but supplied with different preprocessors for input data transformation, were applied. Each of these networks consisted of Lebart`s adaptive perceptrons (Lebart, 1998) which classified objects by a backward error propagation technique at the exit. The initial classification in all the networks was carried out by transformation of a fuzzy classification, defined by the objects` position on the periphery of hyperellipsoid formed by semiorthogonal transformation of principal components of variables subjected to a monotonic transformation, into a hard classification. The final classification efficiency was checked by each network based on the outcome of a posteriori classification performed by Fisher`s linear classifiers in the total space of eventually transformed variables. The standard method used to classify objects was implemented by an Intruder neural network variation (Momirovic, Hosek, Popovic & Boli, 2002) codenamed Intruds. This network only standardizes input data, but does not perform any additional transformations over them. The neural networks under the basic code name Invader (Momirovic, 2001) perform nonlinear monotonic transformations over these data. Invader l transforms variables into sigmoid (0, 1) form, Invader 2 transforms variables into sigmoid (-1, 1) form, and Invader 3 – into staircase (-1.0, 1) form. The neural network codenamed Bibltax (Momirovic, 2001) reduces to zero all standardized values which differ from the expected value in the sample of objects by one half of the standard deviation in order to eliminate noise produced by poorly differentiated objects; and the Exnetg neural network, a variation of the Exelnet neuronal network (Momirovic, 2002), transforms variables into standardized image form to eliminate noise produced by the unique variance of measurement instruments. These networks were competitively used in six numerical examples. In the first example (Crim314), a sample of 314 serious criminals was described by 22 variables for assessing criminal behavior, sociological and microsociological characteristics, as well as characteristics of cognitive processors and conative regulators. In the second example (M666), a sample of 666 respondents from a formally normal population was described by 17 variables for assessing effectiveness of cognitive processors and conative regulators, socialization level as well as modality and intensity of aberrant behavior. In the third example (Ag495), a sample of 495 delinquents was described using 30 Likert scales for assessment of basic aggressiveness. In the fourth example (Mor318), a sample of 318 respondents from a formally normal population was described by 29 variables for assessing efficiency of cognitive processors and conative regulators and by variables for evaluation of disturbances of moral judgment and behavior. In the fifth example (P623), a sample of 623 respondents from a formally normal population was described by 9 variables for assessing personality traits implied by current theories about the structure of conative factors. In the sixth example (Psy204), a sample of 204 psychopathic criminals was described by 18 variables for assessment of cognitive and conative factors assumed by cybernetic theories of cognitive and conative functioning. The effectiveness of each network applied in each example was assessed by a fraction of correctly recognized entities on the basis of the outcomes of regular Fisher linear classifiers.

RESULTS

The results obtained in the conducted experiment are presented in the following Table 1. Comparative analysis of the effectiveness of these taxonomic networks clearly shows that, in the all cases, better taxonomic effectiveness, or surer breaking up of the analyzed set of objects, is achieved by means of data preprocessing.

Table 1. Efficiency coefficients of taxonomic networks

Network	Crim314	M666	Ag495	Mor318	P623	Psy204
Intruds	.955	.983	.859	.912	.974	.990
Invader l	.990	.994	1.000	.997	.998	.951
Invader 2	.936	.992	.998	.994	.995	.995
Invader 3	.968	1.000	.998	.940	.995	.995
Bibltax	.994	.983	.996	.953	.987	.995
Exnetg	.962	.986	.994	1.000	1.000	.995

However, this considerably depends on the data preprocessing technique which is dependent on the nature of the analyzed data, so it cannot be determined with certainty which processing techniques leads to the highest taxonomic efficiency; even an exotic operation built into the Bibltax network yields the best result in one of the analyzed cases. Generally, the Exnetg network which transforms input data into standardized image form behaves somewhat better than the other networks, but this network also behaves relatively poorly in a certain case, when elimination of the effects of badly differentiated objects performed by Bibltax delivers relatively best results. Although more than satisfactory results were obtained by each of the networks, application of several different data preprocessing techniques seems to be the most reasonable strategy if it is not clear from the nature of the data which method will provide the best breaking up of the set of objects.

Conclusion

These networks were competitively used in six numerical examples. Comparative analysis of the effectiveness of these taxonomic networks clearly shows that, in all the cases, better taxonomic efficiency, or surer breaking up of the analyzed set of objects, is achieved by means of data preprocessing. However, this considerably depends on the data preprocessing technique which is dependent on the nature of the analyzed data, so it cannot be determined with certainty which processing technique leads to the highest taxonomic efficiency; even an exotic operation built into the Bibltax network yields the best result in one of the analyzed cases. Generally, the Exnetg network that transforms input data into standardized image form behaves somewhat better than the other networks, but this network also behaves relatively poorly in a certain case, when elimination of the effects of badly differentiated objects performed by Bibltax delivers relatively best results.

REFERENCES

Lebart, L. 1998. Correspondence analysis, discrimination and neural networks. In C. Hayashi et al. Data Science, Classification, and Related Methods, pp. 423-430. Tokyo: Springer.

Momirovic, K. 2001. An extravagant algorithm for automatic classification by neural networks. Technical Report, Institute of Criminological and Sociological Research, Belgrade.

Momirovic, K. 2001. Automatic classification by neural networks with input data preprocessing. Technical Report, Institute of Criminological and Sociological Research, Belgrade.

Momirovic, K. 2002. Neural network for taxonomic analysis of objects described by a set of standardized image variables. Technical Report, Institute of Criminological and Sociological Research, Belgrade.

Momirovic, K. and Popovic, D. 2003. Construction and application of taxonomic neural networks. (Monography) Multidisciplinary Research Center, Faculty of Sport and Physical Education, Pristina.

Momirovic, K., Hosek, A., Popovic, D, A., Boli, E. 2002. Cluster analysis by neural networks. *Proceedings of the l0th International Congress on Physical Education and Sport.* Komotini: Bemocritus University of Trace. Greece.

Popovic, D. 1993. *Programs and subprograms for the analysis of quantitative modifications* (monography). University of Priština, Faculty of Physical Education, Multidisciplinary Research Center, Pristina.

****