Application of a neuro fuzzy network in prediction of absenteeism at work

Conference Paper - January 2012

CITATIONS
36

READS
3,863

4 authors:

Andréa Martiniano
Nove de Julho University
6 PUBLICATIONS **129** CITATIONS
SEE PROFILE

Ricardo Pinto Ferreira
Universidade Nove de Julho
22 PUBLICATIONS **167** CITATIONS
SEE PROFILE

Renato Jose Sassi
Nove de Julho University
127 PUBLICATIONS **386** CITATIONS
SEE PROFILE

Carlos Affonso
São Paulo State University
13 PUBLICATIONS **146** CITATIONS
SEE PROFILE

Application of a neuro fuzzy network in prediction of absenteeism at work

A. Martiniano, R. P. Ferreira
Post-graduation in Production Engineering
Universidade Nove de Julho - Uninove
São Paulo - Brazil
andrea.martiniano@gmail.com, log.kasparov@gmail.com

R. J. Sassi, C. Affonso
Post-graduation in Production Engineering
Universidade Nove de Julho - Uninove
São Paulo - Brazil
sassi@uninove.br, carlos.affonso@uninove.com.br

Abstract - **Competitiveness, market share, professional development and personal support to community action, health, culture, education and sport, are linked to a promising new market. Coupled with the development of organizations, the pressure to achieve goals more audacious, employees increasingly overwhelmed, they end up buying some disturbance in the health-related type of labor activity. The objective of this paper is to apply a neuro fuzzy network in the prediction of absenteeism at work. To make the database were collected records of absenteeism from work during the period of July/07 to July/2010 in a Courier company. Absences certified with the International Classification of Diseases were stratified into 21 categories, the data were tabulated and filtered in MS Excel software. A neuro fuzzy network was developed using an artificial neural network architecture multilayer perceptron with the error backpropagation algorithm. This paper presents partial results of the use of neuro fuzzy network to predict absenteeism at work.**

Keywords - Absenteeism, International Classification of Diseases; Prediction of Absenteeism; Neuro Fuzzy Network.

I. INTRODUCTION

The evolution of society has highlighted the importance of the relationship between men, between different cultures and between markets. In these circumstances, human work has become more complex, supplanting the traditional meaning that identifies it as a means of satisfying needs [1].

Absenteeism represents a loss of productive and profitable capacity for the company. A service is no longer provided due to the absence of an employee, reducing the company's revenue. Thus, the knowledge of absenteeism-disease, in its multiple determinants, becomes fundamental for the organization and planning of actions of the services of attention to the health of the worker and the determination of its profile allows the identification of excesses of occurrences of cases of certain morbidities [2].

The cost of absenteeism due to sickness is, in many organizations, greater than the salary paid to the sick employee who fails to attend the workplace and, with no substitutes available, the implications for productivity are far greater [3].

Neuro *fuzzy* networks have emerged as a promising tool, as they bring together the benefits of Neural Networks and *Fuzzy* Logic. Thus learning and computational power of Neural Networks, the representation capacity and reasoning of *Fuzzy* Logic are combined [4].

Artificial Intelligence (AI) Techniques can be combined to generate architectures as is the case of the MLP neuro *fuzzy* network used in this work [4].

The great advantage of this type of system is due to the action of simultaneous effort or cooperation obtained by combining two or more techniques. This cooperation reflects in obtaining a more robust system with fewer deficiencies [5].

The rest of the paper is organized as follows: in section II is presented the theoretical foundation of Absenteeism, in section III is presented the Methodology, in section IV is presented the neuro *fuzzy* network in the prediction of absenteeism at work, in section V are described the results of the research. The work is closed in section VI with the conclusion.

II. ABSENTEEISM

Absenteeism in general is defined as not showing up at work as scheduled. There is historically a long research, as this phenomenon, in part, generates a high cost for companies in addition to its *status* as an unfavorable indicator [6].

Absenteeism is considered to be the absence or absence of an employee in their work environment and can also be defined as: temporary or permanent incapacity of this absence [7]. It is also known by the expression absenteeism or absenteeism, an expression used to designate the absence of the employee from his workplace. Even if it is not motivated by prolonged illness or legal leave [8].

The lack of a benefits policy is directly linked to the causes of absenteeism. With the employee frustrated, he starts not to fulfill his full working day, is motivated to constant delays or becomes slow in fulfilling his duties [8].

Absenteeism can be attributed to known causes and to unknown causes. Among the known causes are all those that are covered by law and are therefore justified to the employer by requesting permission to be absent. This is the case for vacations, marriages, births, deaths and changes of domicile. Ignored absences are generally justified by

health problems of the worker and/or their dependents or random factors of the most diverse kind [9].

These problems can be avoided by recognizing that the employee may not be available to perform their work as scheduled. This often means that the work is done less efficiently by another employee or not at all [10].

Regarding the prediction of absenteeism, the works [11, 12, 13] can be highlighted for their contributions.

III. METHODOLOGY

The absenteeism records attested with the International Code of Diseases (ICD) were stratified into 21 categories [25], in order to obtain the impact of these absences, these parameters were converted through *Fuzzy* sets. The hours of leave were tabulated by category in a spreadsheet in MS Excel *software* and transformed into a table with zeros and ones, 0 (zero) for non-absence and 1 (one) for absence. This table was processed by the neuro *fuzzy* network. The free *software* SCILAB 5.1 was used as a programming language, due to the advantages pointed out in (http://www.scilab.org/.) As a metric for verifying the validity of the network, the mean absolute error was chosen.

To model the pertinence functions corresponding to the categories by ICD, Gaussian functions were used, as shown in equation (1) below:

$$\mu^{\alpha}(x) = e^{\frac{-1}{\sigma}(\square\square\square\alpha)^2} \quad (1)$$

Where:

c_{α}: center of the Gaussian function

σ_{α}: dispersions of the function

α:linguistic label (high, medium, low).

The training algorithm used in the MLP was the *error backpropagation* which basically consists of determining the variations in the synaptic weights of the network, aiming to minimize the error obtained in the output by learning the training vector. For this, the algorithm is based on the descending gradient method [14,22]. The *error backpropagation* works as follows: a pattern is presented to the input layer of the network, this pattern is processed, layer by layer, until the output provides the desired response, the f_{MLP} is calculated as shown in equation (2).

$$f_{MLP}(x) = \varphi\left(\sum_{1}^{Non} v_l \cdot \varphi\left(\sum w_{lj}x_l + b_{l0}\right) + b_0\right) \quad (2)$$

Where v_l and w_{lj} are synaptic weights; b_{l0} and b_0 are the *biases*; and φ the activation function, commonly specified as the sigmoid function, as shown in fig. 1 [15].

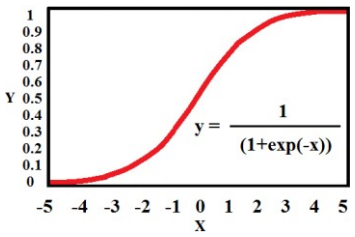

Figure 1. Sigmoid function. Adapted from [16].

The parameters used for the development of the MLP RNF were: number of input neurons equal to 21, number of hidden layers equal to 3, number of neurons in the hidden layers equal to 10, constant learning rate equal to 0.1, constant moment factor equal to 0.7, stopping criterion was the maximum number of epochs equal to 65.

Table I presents the International Code of Diseases (ICD) by categories that were used as inputs in the neuro *fuzzy* network.

TABLE I - International Code of Diseases

I	Some infectious and parasitic diseases (A00-B99)
II	Neoplasms [tumors] (C00-D48)
III	Diseases of the blood and hematopoietic organs and some immune disorders (D50-D89
IV	Endocrine, nutritional and metabolic disorders (E00-E90)
V	Mental and behavioral disorders (F00- F99)
VI	Diseases of the nervous system (G00-G99)
VII	Diseases of the eye and appendages (H00-H59)
VIII	Diseases of the ear and mastoid apophysis (H60- H95)
IX	Diseases of the circulatory system (I00-I99)
X	Diseases of the respiratory system (J00-J99)
XI	Diseases of the digestive system (K00-K93)
XII	Diseases of the skin and subcutaneous tissue (L00- L99)
XIII	Diseases of the musculoskeletal system and connective tissue (M00-M99)
XIV	Diseases of the genitourinary system (N00-N99)
XV	Pregnancy, childbirth and puerperium (O00-O99)
XVI	Some conditions originating in the period perinatal (P00-P96)
XVII	Congenital malformations, deformities and chromosomal anomalies (Q00-Q99)
XVIII	Symptoms, signs and abnormal clinical and laboratory findings, not classified under other party (R00-R99)
XIX	Injuries, poisoning and some other consequences of external causes (S00-T98)
XX	External causes of morbidity and mortality (V01-Y98)
XXI	Factors influencing health status and contact with health services (Z00-Z99)

Source: Adapted from DATASUS [17].

The other causes of absenteeism at work such as: unjustified absences, dental certificates, medical appointments, companion certificates, blood donation,

physiotherapy and examinations were quantified separately and will not be presented in this paper.

IV. THE *FUZZY* NEURO-NETWORK IN THE PREDICTION OF ABSENTEEISM AT WORK

Fuzzy Logic is a mathematical theory developed since 1965 [18], its main objective is to allow the modeling of the approximate mode of human reasoning, imitating the human ability to make decisions in environments of uncertainty, imprecision and vagueness expressed by a set of linguistic variables [19].

Fuzzy logic provides a high-level and user-friendly interface for developing programs, helping designers to focus on functional goals rather than mathematical details [20].

The *fuzzy* set is the bridge that connects the imprecise concept to its numerical modeling, assigning to each element of the universe a value between 0 and 1, which represents the degree of pertinence of this individual to the *fuzzy* set [18, 21, 22].

Artificial Neural Networks are models inspired by the structure of the brain aiming to simulate human behavior in processes such as: learning, adaptation, association, fault tolerance, generalization and abstraction [14, 20]. These models consist of simple processing units, called artificial neurons, which calculate mathematical functions. Artificial Neural Networks are convenient for massive handling of numerical data [20].

The Artificial Neural Network combined with *Fuzzy* Logic forms the neuro *fuzzy* network. Fig. 2 shows the flow in the MLP neuro *fuzzy* network having as output the prediction of absenteeism at work in hours.

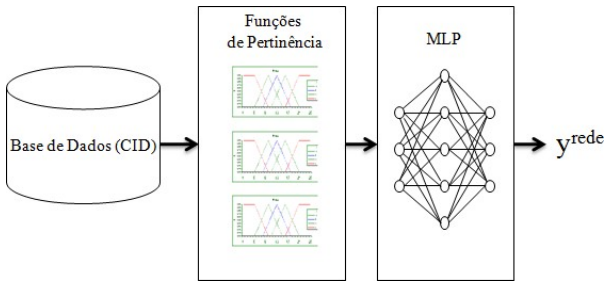

Figure 2: Flow in the MLP neuro *fuzzy* network. Adapted from [23,24].

Fig. 2 shows:
- The RNF input the causes of absenteeism (ICD);
- The Relevance Functions;
- The *Multilayer Perceptron* (MLP);
- The output of the neuro *fuzzy* network (y).[rede]

V. RESULTS

Fig. 3 shows the actual results of absenteeism (Yreal), and the prediction of absenteeism at work, in hours, calculated by the neuro *fuzzy* network (Yrede - training) and (Yrede - test).

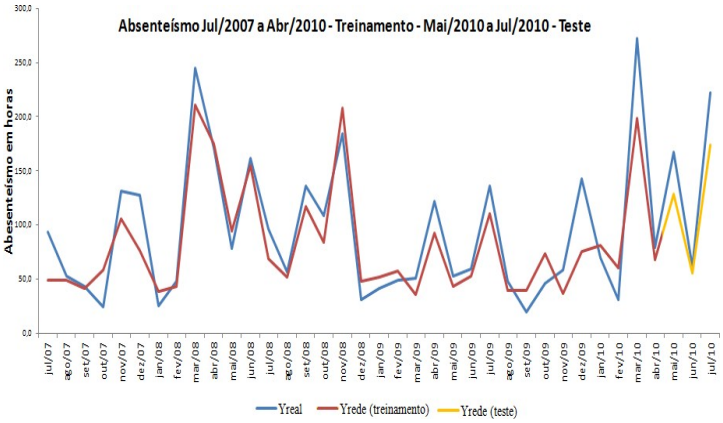

Figure 3. Actual and predicted Absenteeism in hours.

With the forecast of absenteeism it is possible to estimate what portion of productivity may be affected, for example, in the month of March 2008 there are 250 hours of absenteeism, which may represent for the company 10%, 15% or 20% of productive capacity, this early information helps in decision making, in relation to:

- vacation planning;
- call for overtime;
- weekend or holiday work;
- hiring temporary labor, etc.

Aiming to reduce the impacts of absenteeism at work.

In this context, the prediction of absenteeism can be of great help in decision making avoiding loss of productivity and profitability.

Fig. 4 shows the six categories with the highest representation in the database, together corresponding to 78.65% of absences attested with ICD. It is observed that category XIX, injuries, poisoning and some other consequences of external causes (S00-T98), category XII, diseases of the skin and subcutaneous tissue (L00-L99) and category XIII, diseases of the musculoskeletal system and connective tissue (M00-M99), are the diseases that most cause absenteeism in the company surveyed, for this reason they should be better investigated aiming at preventive actions of occupational health.

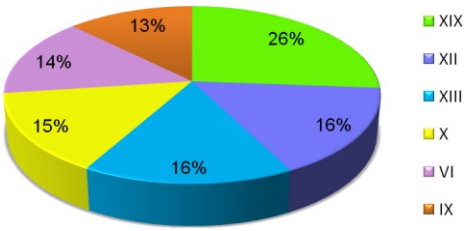

Figure 4 Distribution by International Classification of Diseases (ICD) with
higher incidence.

VI. CONCLUSION

It is concluded that the neuro *fuzzy* network for predicting absenteeism at work can be an excellent tool to assist decision making, since production can be maintained with planned measures such as vacation distribution and in production maintenance measures such as calling for overtime and/or work on the weekend or holiday, hiring temporary labor, etc. With the advantage of the timely planning of the call-up of employees in advance or temporary hiring.

It is intended to continue this initial work, using a larger database, adding all the causes of absenteeism, which were not presented. It is also intended to use Data Mining to extract useful knowledge in the database. Finally, to conduct a *survey* in the same *Courier* company, in order to identify patterns of behavior of absentee and presentee employees.

ACKNOWLEDGEMENT

To the Nove de Julho University for the scholarship in the Graduate Program in Production Engineering.

REFERENCES

[1] G. Lumare Júnior, Customer economic value: a theory of orders. São Paulo: *Financial Times* - Prentice Hall*, 2007.*

[2] L. S. Silva, T. M. M. Pinheiro, S. Sakurai, Profile of absenteeism in a state bank in Minas Gerais: analysis in the period from 1998 to 2003. Available at: < http://redalyc.uaemex.mx/pdf/630/63009609.pdf> Accessed: Jan. 20, 2011.

[3] M. Pauly, Multiplier effect: financial consequences of absenteeism - 2006. Available at: <http://www.wharton.universia.net/index.cfm?fa=viewfeature&id=1075 &language=portuguese> Accessed: 04 Jan. 2011.

[4] F. Gomide, M. Figueiredo, W. Pedrycz, *A neural Fuzzy network: Structure and learning, Fuzzy Logic and Its Applications, Information Sciences and Intelligent Systems, Bien, Z. and Min, K., Kluwer Academic Publishers, Netherlands,* 1998, pp. 177-186.

[5] R. J. Sassi, A Hybrid Architecture for Knowledge Discovery in Databases: *Rough Sets* Theory and Artificial Neural Networks Self-Organizing Maps. 2006. Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia Elétrica, São Paulo. 169 p.

[6] G. Johns, *Presenteeism in the workplace: A review and research agenda. Journal of Organizational Behavior.* vol. 31, p. 519 - 542, 2010.

[7] A. M. Silva, R. P. Ferreira, R. J. Sassi, *Control and monitoring of the indexes of absenteeism and presenteeism whith aid of the technology of the information.* CONTECSI - 2010, ISBN: 978-85-99693-06-3, 2010.

[8] I. Chiavenato, Recursos humanos na empresa. 3. ed. São Paulo: Atlas, 1994.

[9] I. Penatti, J. S. Zago, O. Quelhas, Absenteeism: the consequences in people management. III SEGeT - Symposium of Excellence in Management and Technology. Resende, 2006.

[10] W. Cascio, J. Boudreau, *Investing in People: Financial Impact of Human Resource Initiatives, Pearson Education, Inc. Publishing as FT Press,* New Jersey, 2008.

[11] J. M. Ivancevich. *Predicting Absenteeism from Prior Absence and Work Attitudes. The Academy of Management Journal*. Vol. 28, no. 1, pp. 219-228, Mar. 1985.

[12] J. H. Morris, J. D. Sherman, R. A. Snyder. *Prediction of Absenteeism from Attitudes, Prior Absenteeism, and Performance, Personnel Review,* Vol. 18 no. 1, p.16 - 22, 1989.

[13] LT. Yen , DW. Edington, P. Witting . *Prediction of prospective medical claims and absenteeism costs for 1284 hourly workers from a manufacturing company. Journal of Occupational Medicine*, vol. 34, no.4, p.428-435, ISSN: 0096-1736, 1992.

[14] S. Haykin, *Neural Networks: A Comprehensive Foundation. New York: Willey & Sons,* 1999.

[15] R. P. Ferreira, C. O. Affonso, R. J. Sassi, Application of a Neuro *Fuzzy* Network in the prediction of urban traffic behavior in the City of São Paulo. In: 5th *Iberian Conference on Information Systems and Technologies.* Santiago de Compostela. Sistemas *y Tecnologías de Información* CISTI, vol. II. p. 82-85, 2010, ISBN: 978- 989-96247-3-3,2010.

[16] R. J. Sassi, L. A. Silva, D. M. E. Hernandez, *A Methodology using Neural Networks to Cluster Validity Discovered from a Marketing Database. In: 10th Brazilian Symposium on Neural Networks* (SBRN), 2008, Salvador. IEEE *Proceedings of* SBRN. vol. 1. p. 03-08, 2008.

[17] Datasus, Version 1.6c in Portuguese, 2000.

[18] L. A. Zadeh, *Fuzzy Sets, Information and Control*, vol. 8. pp. 338-353, 1965.

[19] K. M. Passino, S. Yurkovich, *Fuzzy Control*, Addison Wesley Longman, Inc. 1998.

[20] M. G. Simões, I. S. Shaw, Control and *fuzzy* modeling. São Paulo: Blucher: FAPESP, 2007.

[21] M. C. Nicoletti, H. A. Camargo, Fundamentals of Set Theory *Fuzzy*, Edusfscar, 2004.

[22] M. A. C. Pacheco, M. M. B. R. Vellasco (org), Intelligent decision support systems: economic analysis of oil field development projects under uncertainty. Rio de Janeiro: Ed. PUC-Rio: Ed. Interciência, 2007.

[23] C. O. Affonso, Application of Neuro *Fuzzy* Networks to Polymer Processing in the Automotive Industry. 2010. Dissertation (Master) - Nove de Julho University, Production Engineering, São Paulo. 110 p.

[24] R. P. Ferreira, Combination of Artificial Intelligence Techniques for Prediction of Urban Vehicular Traffic Behavior in the City of São Paulo, 2011. Dissertation (Master) - Nove de Julho University, Production Engineering, São Paulo. 107 p.

[25] Absenteeism at Work Data set available in: (http://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work).