Varazdin Development and Entrepreneurship Agency
in cooperation with
Megatrend University, Serbia
University North, Croatia
Faculty of Management University of Warsaw, Poland
Faculty of Law, Economics and Social Sciences Sale - Mohammed V University in Rabat

Economic and Social Development

30th International Scientific Conference on Economic and Social Development

Editors:
Khalid Hammes, Igor Klopotan, Milica Nestorovic

ISSN 1849-7535

Book of Proceedings

Belgrade, 25-26 May 2018

Varazdin Development and Entrepreneurship Agency
in cooperation with
Megatrend University, Serbia
University North, Croatia
Faculty of Management University of Warsaw, Poland
Faculty of Law, Economics and Social Sciences Sale - Mohammed V University in Rabat

Editors:
Khalid Hammes, Igor Klopotan, Milica Nestorovic

Economic and Social Development

30th International Scientific Conference on Economic and Social Development

Book of Proceedings

Belgrade, 25-26 May 2018

RISK ASSESSMENT BASED ON INTEGRATED FUZZY MEP METHODOLOGY

Marija Kerkez
Faculty of Business Studies, Serbia
majap@rcub.bg.ac.rs

Nebojsa M. Ralevic
Faculty of Technical Sciences, Serbia
nralevic@uns.ac.rs

Tanja Todorovic
Faculty of Technical Sciences, Serbia
ttodorovic@uns.ac.rs

Boris Zezelj
Faculty of Technical Sciences, Serbia
zezelj_boris@yahoo.com

ABSTRACT

In the effort to overcome the limitations of some mathematical models to deal with real-world problems, evolutionary algorithms were developed as alternative optimization technique. Genetic Programming (GP) represents a branch of evolutionary programming, where encoding is performed by using an evolutionary algorithm and resulting in solutions that consist of computer programs. So far, genetic programming has been successfully implemented in various optimization, search and model approximation problems. As a subset of machine learning paradigm, GP, developed by Koza, uses genetic algorithms (GA) to automatically generate computer programs. Multi Expression Programming (MEP), a Genetic Programming variant, is used in developing models for characterization of system behavior by directly extracting knowledge from data. MEP is considered an efficient technique for solving complex problems, having a distinctive feature to store multiple solutions in a single chromosome. However, the decoding process remains at the same complexity level. MEP has a potentially wide range of applications. Real systems that exist in socio-economic environment are characterized by dynamic structure, reflected in nonlinearity, uncertainty and other inherent aspects. Thus, standard mathematical approach, relying on precise mathematical relations, has certain limitations in modeling complex systems. As an alternative, fuzzy mathematics is applied when modelling vague and complex relations and systems. The application of fuzzy systems theory is recommended in situations where data values and relations are uncertain and imprecise and their estimation relies on incomplete expert judgment. The principles of fuzzy mathematics have been extensively used in risk assessment. This paper proposes an integrated methodology for risk assessment that combines MEP and fuzzy mathematics.

Keywords: *Credit scoring model, Genetic Programming, Risk Assessment*

1. INTRODUCTION

Banks industry are exposed to challenges in finding new ways of doing business with less risk, but efficient and profitable. Credit risk is a critical component of any banking organisation and loans are the largest source of credit risk. Credit scoring models have been used by banks to determine if applicant belong to a good or a bad applicant group. It is shown that these models reduce the cost of credit analysis, enable a faster credit decision, and reduce the potential risk (Lee et al, 2002).

Credit risk assessment was mostly based on statistical and operational methods. Most recently different neural network, expert systems, fuzzy systems, genetic algorithm and genetic programming techniques have been proposed. Hybrid systems such as Genetic Fuzzy Systems (Cordón, 2011; Koshiyama, 2015, Prasad et al, 2008, Grosan and Abraham, 2006) have been widely employed to solve classification and regression problems. They have capability of extracting knowledge from datasets and state it in the linguistic labels with reasonable accuracy. In this paper we proposed integrated method of MEP and Fuzzy Inference System (FIS). Multi Expression Programming is a Genetic Programming variant. MEP techniques are substantially helpful for determine empirical models and to indicate behavior by directly extracting experimental information (Kerkez et al, 2017). One of the advantages of this tool is that, unlike traditional methods, the MEP does not require assumptions that simplify the development of the model. Evaluation of the applicant's is done using fuzzy logic and Fuzzy Inference System. Defining applicants not only in good and bad group, but in „somewhere between" provides the possibility of reconsidering a particular client using an individual approach as a proactive model of cooperation.

2. ABOUT MULTI EXPRESSION PROGRAMMING

The MEP uses the following steps to develop the best program until a termination condition is achieved (Oltean and Grosşan, 2003; Alavietal, 2010; Alaviand and Gandomi, 2011):

a) two parents are selected using the tournament procedure and they are again combined with the fixed crossover probability,
b) obtaining two off spring on the basis of a re-combination of two parents,
c) mutation of the off spring and the replacement of the worst individual in the current population with the best of them.

In the MEP scheme, the first symbol of chromosomes must be a terminal symbol. In this way, only syntactically correct programs are obtained. MEP chromosomes, using a set of functions $F = \{+, *\}$ and a set of terminals $T = \{a, b, c, d\}$, is represented as follows:

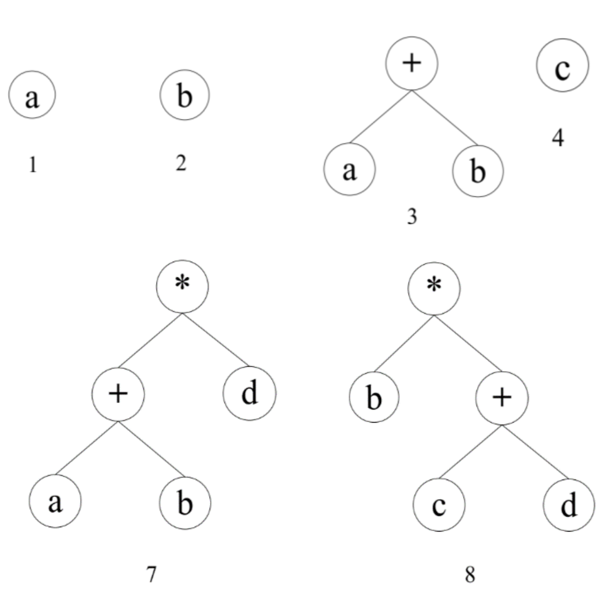
1: a
2: b
3: +1,2
4: c
5: d
6: +4,5
7: *3,5
8: *2,6

The algorithm starts with a randomly selected population of individuals. There are certain limitations for creating a valid MEP individual (more about that in Oltean and Dumitrescu, 2002). The value of these expressions can be calculated by reading the chromosomes from the top to the bottom. Partial results are calculated using dynamic programming and are placed in a conventional manner. The chromosome described above encodes the following expressions:

E1	a
E2	b
E3	a+b
E4	c
E5	d
E6	c+d

E7 (a+b)*d
E8 b*(c+d)

Figure 1: MEP chromosome represented as trees (adapted from Oltean, 2006)



Due to their multiple expressions, each MEP chromosome should be viewed as a forest of trees (Figure 1), not as one tree, as is the case with genetic programming. As MEP chromosomes encode more than one problem solution, it's interesting to see how fitness is assigned. The best expression is chosen after controlling the fitness of all expressions in MEP chromosomes using the following equation (Oltean and Groşan, 2003):

$$ff = \min_{i=l,g}\left\{\sum_{j=1}^{n}\left|E(k)-O_k^i\right|\right\}$$

where n is the number of fitness cases; E(k) expected value, O_k^i is the rate comes for the k-th case by the i-th expression encoded in the current chromosome, g is the number of genes.

The standard MEP algorithm uses the evolutionary model as a stable state, as its basic mechanism. The MEP algorithm begins with the creation of a random population of individuals. The standard MEP algorithm looks like this (Oltean et al, 2009):

%Randomly create the initial population P(0)
while not stop condition do
% Select two individuals p_1 and p_2 from the current population
% Crossover the parents $p1$ and $p2$, obtaining the o_1 and o_2
% Mutate the o_1 and o_2
if fitness (the best offspring) is better than fitness (the worst individual) then
% Replace the worst individual with the best offspring
end if
end while
% Output *S* as the best solution (individual) found

The MEP accepts a set of data relating to our research. In order to achieve a consistent, data are devided, several combinations of training and testing sessions are considered. The model with the best performance of data sets for learning and testing was finally selected as a result of the work. Operators are selected in relation to the problem and the data of interesting. Problems of regression type, binary classification and classification of several classes can be done in the MEP. For the regression type model, in order to evaluate the performance of the derived models, the correlation coefficient (R), the mean square error (RMSE) and the mean percentage error (MAE) are used.

3. GENETIC FUZZY SYSTEMS APPLICATION

In this part of the paper we will present the possibility of applying an integrated approach to fuzzy mathematics and multi expression programming in personal credit risk assessment. For the purpose of this work we use German data[1] set with 15 attributes and 520 past applicants for credit. A large set of data, with high dimension and sometimes with irrelevant attributes in the training dataset, can lead to less accurate results in analysis. Therefore, it is important to extract the most significant features to increase predictive accuracy, speed and scalability. In addition, the data are slightly adjusted by the experts due to the easier presentation of the proposed methodology. Data conclude the following three groups: (a) The personal condition (age, sex, marriage, education etc.), (b) The financial condition (profession, job and title, years of services, income, and annual family revenue), (c) Social condition (bad credit record if any, insurance). In classification problems we used methodology described in (Koshiyama, 2015). The main source of information consists of a dataset containing n patterns x_i. Each patterns are described by values of m characteristics $X_j\left(i=1,\ldots,n \text{ and } j=1,\ldots,m\right)$.

In the fuzzification process we have membership functions:

$$A_{ij}=\left\{\left(x_{ij},A_{ij}\left(x_{ij}\right)\right)\mid x_{ij}\in X_j\right\},\ \text{where}\ A_{ij}:X_j\to[0,\ 1].$$

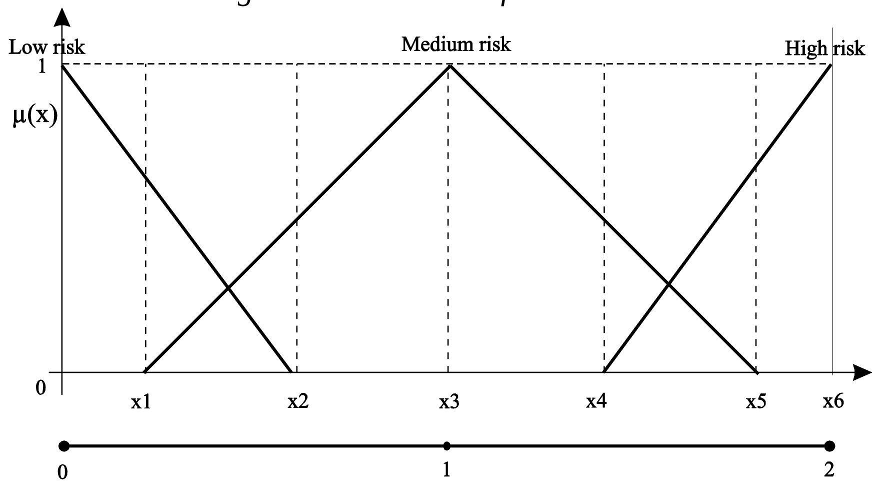
Fuzzy numbers are symmetrical triangular fuzzy number and membership functions are defined by the expert. Input data are membership degree of fuzzy set as result of mapping crisp data set. In the next step Fuzzy Inference System accepts data. Fuzzy rules are defined as follows

R: IF X_1 is A_{i1} and X_2 is A_{i1} and ….and X_j is A_{i1} THEN x_i in score s

The rule base is a set of linguistic terms in the form of IF–THEN rules with antecedents and consequents connected with AND operator. Aggregation operator used in this work is maximum. The output of the fuzzy system are in the interval [0, 2] (see figure 2). Each risk element is described by fuzzy sets as low risk, medium risk, and high risk; values x_1, x_2, x_3, x_4, x_5, and x_6 are for left and right borders of triangular fuzzy numbers. Those values represent value of total evaluation. More detailed information about this procedure can be found in (Gajovic, Kerkez and Kocovic, 2017). Several combinations of training and testing sets are considered, where data are randomly divided into learning and testing subsets and in different percentage of data representation (60:40, 80:20, 70:30, 50:50). The basic operators and mathematical functions are used with uniform crossover type. Crossover probability is set on 0.9 and number of iterations is 100/500. Based on fuzzy rules, each new applicant can be apprise with selected variables and based on the predictor values, be classified in a high credit risk (score 2), medium credit risk (1) and low credit risk (score 0).

[1] This data set originally consist of 30 variables for 1000 past applicants for credit.

Figure 2: Score classification



The German credit dataset composed of 700 instances of creditworthy applicants and 300 instances of bad credit applicants. Result of our model shows 0,6444 creditable applicant with 0.003 standard deviation. Percent of applicant that need to be reconsider is 6,80%. The rest are the bad credit applicant. The proposed methodology provides flexibility regarding the specificity of the problem and can easily be adapted to different classification problems with different types of attributes. The model is characterized by interpretability and ability to present a possible classifier in different ways.

4. CONCLUSION

Real systems such as bank borrowing activities are characterized by dynamic structure, reflected in nonlinearity, uncertainty and other inherent aspects. Credit risk assessments is crucial for banks to help determine probable risk and to make right decision regarding credit approval, but and other connected factors, such is interest rate, repayment, grace period etc. In the process of the personal credit risk analysis credit scoring is evaluating technique that helps banks to determine customer's credit risk. In this paper we used empirical credit scoring model, based on past data of relevant attributes to identify creditable or bad credit applicants. Even more, methods can be used in identifying creditworthiness of new applicants. Thus, standard mathematical approach, relying on precise mathematical relations, has certain limitations in modelling complex systems. As an alternative, integrated techniques are gaining in greater significance in risk assessments. Presented method using integrated techniques of genetic programming and fuzzy mathematics. The model helps in the faster applicant selection, but it also require credit professionals to make decisions in the definition of attributes, and in the evaluation process.

LITERATURE:

1. Alavi, A.H., Gandomi, A.H., Sahab, M.G. and Gandomi, M. (2010) Multi Expression Programming: A New Approach to Formulation of Soil Classification. *Engineering with Computers*, 26 (2), pp. 111-118.
2. Cordón, O. (2011). A historical review of evolutionary learning methods for Mamdanitype fuzzy rule-based systems: Designing interpretable genetic fuzzy systems? *Int. J. Approx. Reason.*, 52 (6), pp. 894–913.
3. Gajovic, V., Kerkez, M. and Kocovic, J. (2017). Modeling and simulation of logistic processes: risk assessment with a fuzzy logic technique. *Simulation. Transactions of the Society for Modeling and Simulation International*, pp. 1-12.

4. Gandomi H., Alavi, A.H. and Yun, G.J. (2011). Formulation of uplift capacity of suction Caissons using multi expression programming. *KSCE CivilEng.*, 15 (2), pp. 363–373.
5. Grosan, C. and Abraham, A., Stock Market Modeling Using Genetic Programming Ensembles. In: *Genetic Systems Programming. N. (Nedjah et al. eds.), Studies in Fuzziness and Soft Computing*, pp. 131-146. Springer Verlag, Germany 2006.
6. Kerkez, M., Ralević, N., Milutinović, O., Vojinović, Ž. and Mladenović-Vojinović, B. (2018). Integrated Fuzzy System and Multi-Expression Programming Techniques for Supplier Selection. *Technical Gazette*, 26 (1) (in press).
7. Koshiyama, A.S., Marley, Vellasco, M.B.R. and Tanscheit, R. (2015). GPFIS-CLASS: A Genetic Fuzzy System based on Genetic Programming for classification problems. *Applied Soft Computing*, 37, pp. 561–571.
8. Lee, T. S., Chiu, C. C., Lu, C. J., and Chen, I. F. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, 23(3), pp. 245–254.
9. Oltean, M. (2006). Multi Expression Programming, Technical instructions. http://www.mep.cs.ubbcluj.ro/
10. Oltean, M. and Dumitrescu, D., Multi expression programming. Technical report, Babes-Bolyai University, Cluj-Napoca, Romania, 2002.
11. Oltean, M. and Grosan, C., (2003). A comparison of several linear genetic programming techniques. *Complex Syst*., 14 (4), pp. 1–29.
12. Oltean, M., Groşan, C., Dioşan, L. and Mihăilă, C. (2009). Genetic programming with linear representation: a Survey. *International Journal on Artificial Intelligence Tools*, 18 (02), pp. 197-238.
13. Piramuthu, S. (1999). Financial credit-risk evaluation with neural and neurofuzzy systems. *European Journal of Operational Research*, 112(2), 310–321.
14. Prasad, G.V.S.N.R.V., Dhanalakshmi, Y., Kumar, V. and Ramesh Babu, I. (2008). Modeling An Intrusion Detection System Using Data Mining And Genetic Algorithms Based On Fuzzy Logic. *International Journal of Computer Science and Network Security*, 8 (7), pp. 319-325.