VOLUME 11, ISSUE 1, 2016

ISSN 2045-3345

FUZZY SETS IN UNEMPLOYMENT PROBLEM

Elena Rihova[I], Iva Peckova[II]
University of Economics, Prague, Czech Republic

ABSTRACT

Unemployment level moves in a cyclical manner. However, not only business cycle has an influence on unemployment levels, but also labour market policies and demographic developments may also influence the short and long-term evolution. Some unemployed are willing and able to work for pay currently available to work, and have actively searched for work, though. Unlike them, there is a group of unemployed, which are not are willing and able to change their position and find a job, "unhopefully" unemployed. For economic and social governance is key to define a group of "unhopefully" and "hopefully" unemployed. The purpose of this study was to find a statistical methodology, which help to define the group of "hopefully" unemployed. The study was held on unemployment from Czech Republic during 2014.

JEL CLASSIFICATION & KEYWORDS

■ C38 ■ C80 ■ E24 ■ UNEMPLOYMENT ■ CLUSTERING

INTRODUCTION

The problem of unemployment is still issue of the day. The International Labor Organization (ILO) definition of the unemployment rate is the most widely used indicator on labor market. Unemployment is not only economic indicator, but also affects mental and physical health (Pharr, 2012). Further risks associated with unemployment include de-skilling, degradation of physical and mental health, and a lasting reduction in life satisfaction. At the macroeconomic level, failing to integrate the generation implies a loss of production, productivity and probably innovation potential. In addition to the associated loss of gross domestic product (GDP), there is a fiscal cost of unemployment due to increased welfare payments and loss of tax revenues. In addition, from a societal perspective, the losses are probably even larger than the pure economic costs. The personal effects on health and social stability may translate into greater pressure on public services, such as health, welfare and integration services. From the foregoing it follows, that unemployment should be studied in more detail.

According published information from International Labour Organization (ILO) and by OECD, during last 3 years the unemployment rate in European Union is getting down and makes 9.1%. Among the EU states, the lowest unemployment rates are recorded in Germany (4.5%) and the Czech Republic (5%), and the highest in Greece (25.2%) and Spain (22.2%).

In those unemployment rates were included as long-term unemployed people, as also short-term unemployed (for instance, people, which wanted to change their work). From an economic perspective, unemployment may be viewed as unused labour capacity (Eurostat). While these numbers derive from a rather simple calculus supposing that one could fully integrate each European into the labour market, they still point to the quite substantial economic losses for societies faced with high unemployment.

Hence, the most important question comes up: how to define "hopefully" unemployed, i.e. unemployed, whom pay off to finance retraining courses and to invest in. There are many factors influence willing and able to work: beginning with personal motivation and finishing with age and the level of education. As is shown on Figure 1 there is almost no difference between in gender for unemployment rate in EU, but significant different in the age of the unemployed. Therefore, for the last year, the male unemployment rate is 10.1 and almost the same rate for female, 10.3. And completely different situation with the age of unemployed: the rate of unemployed, which are younger than 25 years is 22%, and the rate of unemployed between 25 and 74 years is markedly less (9.0%).

Despite that, other factors also affect the level of unemployment. In the economic reality, some of them are exactly known; their numerical values and functional relations are explicitly given. Some other quantities and relations are statistically estimated and they can be processed by means of the classical probabilistic methods (Makhalova, 2013). Nevertheless, there still exist economic variables, which are formulated in vague terms. Their vagueness does usually follow either from uncertain knowledge of exact values or from the tendentionally non-exact verbal expressions appearing optimal for their adequate representation. It concerns vague words like "long", "short", "adequate to situation", and additional relativization of numerical values like "approximately…", "rather more than…", "near to…" and others (Ribeiro, 1999). It is obvious, there are a lot of variables and relations which have to be specified, for instance, the level of education.

Detecting every feature value can be obtained by diagnostic tests of input attributes that have associated with integrated (financial and temporal) costs (Quinlan, 1986). An interesting problem here is to form such a method that would search for an optimal (or sub-optimal) sequence of tests to be undertaken, when a new subject is recognized, in order to minimize the cost of diagnostics. One approach for solving this problem is a decision with fuzzy logic help.

Fuzzy logic allows the modelling of language related uncertainties, while providing a symbolic framework for knowledge comprehensibility (Makhalova, 2014). In fuzzy rule-based systems, the symbolic rules provide for ease of understanding and transfer of high-level knowledge, while the fuzzy sets, along with fuzzy logic and approximate reasoning methods, provide the ability to model fine knowledge details (Zadeh, 1975). One cannot deny that, fuzzy representation is becoming increasingly popular in dealing with problems of uncertainty, noise, inexact data and rapidly changing conditions. It has been successfully applied to problems in many various areas. The aim of this study with fuzzy logic helps to determine the group of

[I] elena.rihova@vse.cz

ISSN 2045-3345
RESEARCH JOURNAL OF ECONOMICS, BUSINESS AND ICT

Figure 1: Unemployment rates in EU, 2005-2014 (%)

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Male	8.4	7.6	6.6	6.6	9.0	9.7	9.6	10.4	10.8	10.1
Female	9.8	9.0	7.9	7.5	8.9	9.6	9.8	10.5	10.9	10.3
Less than 25 years	19.0	17.7	15.9	15.9	20.3	21.4	21.7	23.3	23.7	22.2
Between 25 and 74 years	7.7	7.0	6.1	5.9	7.6	8.3	8.3	9.1	9.5	9.0
Long-term unemployment rate	4.1	3.7	3.1	2.6	3.0	3.9	4.2	4.7	5.2	5.1
Male	3.8	3.5	2.9	2.4	2.9	3.9	4.2	4.7	5.2	5.1
Female	4.5	4.1	3.4	2.8	3.1	3.8	4.1	4.7	5.1	5.1
Very long-term unemployment rate	2.4	2.2	1.9	1.5	1.6	1.8	2.2	2.6	2.9	3.1

Source: Eurostat

"hopefully" unemployment, in other words, to apply the fuzzy logic on unemployed problem.

Fuzzy Sets and Fuzzy Clustering in Unemployment Problem

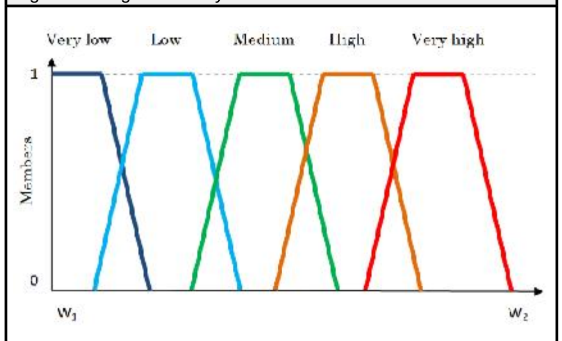
As a founder of fuzzy logic and fuzzy sets, L. Zadeh, said, much of the universality, elegance and power of classical mathematics derive from the assumption that real numbers can be characterized and manipulated with infinite precision. Indeed, without this assumption, it would be much less simple to define what is meant by the zero of a function, the rank of a matrix, the linearity of a transformation or the stationary of a stochastic process (Zadeh, 1975).

The human thinking is a combination of intuition and rigor. On the one hand, the human thinking considers the world as a whole, by analogy. On the other hand, the human thinking considers the world based on many facts, and therefore represents the fuzzy mechanism.

The laws of thoughts, which the computers based on, are formal and strictly; the laws of human thoughts are informal, fuzzy. The mathematical theory of fuzzy logic is a generalization of the classical set theory and classical formal logic. The American scientist Lofty Zadeh first proposed these concepts in 1965. The main reason for the emergence of a new theory was the presence of fuzzy and approximate reasoning in describing human processes, systems, facilities.

Fuzzy logic is an approach to computing based on "degrees of truth" rather than usual logic, which based on "true or false" (1 or 0). Fuzzy logic introduces a number of highly interesting concepts in dealing with the real world. In the real world, there is intrinsic uncertainty. Any calculation that we hold with real dataset, we must take this intrinsic uncertainty into account. Fuzzy thinking presents a contrast probabilistic thinking. That is why the statistics software based on fuzzy logic is much more convenient, faster and better suited to solving problems. Zadeh in his work in 1965 writes, "a fuzzy set U in X characterized by a membership function $f_U(x)$ which associates with each object in X a real number in the interval <0, 1>, with the value of $f_U(x)$ at x representing "the grade of membership" of x in U. Roughly, with the help of fuzzy sets theory we can use in machine learning such definition like "very low", "low", "medium", "high" and "very high", not like in traditional binary machine language, where we have only use "low" and "high". The significance of fuzzy variables is that they facilitate gradual transitions between states and, consequently, possess a natural capability to express and deal with observation and measurement uncertainties. Traditional variables, which we may refer to as crisp variables, do not have this capability. For instance, in Figure 2 wages within a range [W_1,W_2] is characterized as a fuzzy variable. In our research, we will unite the theory of fuzzy logic with one of the multivariate statistic method, with cluster analysis.

Figure 2: Wages as fuzzy variable



Source: Authors

Cluster analysis is one of the major techniques in pattern recognition. It is approach to unsupervised learning. The hard clustering methods restrict that each object of the data set belongs to exactly one cluster, which in does not not happen in real life. Fuzzy set theory proposed by L. Zadeh in 1965 gave an idea of uncertainty of belonging, which was described by a membership function. Application of fuzzy set theory in cluster analysis was early proposed in the work of Belman, Kalaba and Zadeh and Ruspini.

These algorithms are based on objective functions J, which are mathematical criteria that quantify the goodness of cluster models that comprise prototypes and data partition. Objective functions serve as cost functions that have to be minimized to obtain optimal cluster solutions (Řezanková, 2011). Thus, for each of the following cluster models the respective objective function expresses desired properties of what should be regarded as ''best'' results of the cluster algorithm.

The main advantage of fuzzy C-means clustering, it allows gradual memberships of data points to clusters measured as degrees in <0, 1>. This gives the flexibility to express that data points can belong to more than one cluster.

Fuzzy partitioning is carried out through an iterative optimization of the objective function; with the update of membership and the cluster centres. The larger membership values indicate higher confidence in the assignment of the pattern to the cluster. The probabilistic membership degree depends not only on the distance of the datum x_j to cluster c_i, but also on the distance between this data point and other clusters (Oliviera, 2007). Therefore, this enables to divide unemployed into groups more optimal, taking into account all the factors affecting unemployment.

Case Study

The main aim of cluster analysis is to classify the similar objects to groups are similar as possible and the objects in different groups are dissimilar as possible. One of the principal question of classify task is to found the optimal number of clusters for current data set.

Very often a researcher does not have a priory information about the number of clusters in data set. Consequently, finding the optimal number of clusters is an important problem. The problem of finding an optimal number of clusters is usually called cluster validity problem. In order to solve the cluster validity problem, validity indices must enclose, take into account, some specific areas that enable to solve this problem successfully. Those areas are compactness, separation, noise and overlap. As Wanga and Zhang stayed (Wanga and Zhang, 2007), the most reliable index for validity fuzzy C-means clustering results is XB index (Xie and Beni, 1991).

XB index (Xie and Beni, 1991) includes two components: compactness in the numerator and separation - formula (1). Small value of compactness is evidence of a good partition, and high value of separation is evidence of a good partition. The minimum value of this index corresponds with optimal number of clusters, in other words, with the best clustering performance for the data set. Unfortunately, this index has tendency to monotonically decrease with increasing number of clusters.

$$XB = \frac{J_m(u,v)/n}{Sep(v)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^{m} \left\| x_i - x_j \right\|^2}{n.\min_{i,j} \left\| v_i - v_j \right\|^2} \qquad (1)$$

Where n is the number of objects, *c* is the number of clusters, u_{ij} is the fuzzy membership, $x_i x_j$ are objects, v_i and v_j are the fuzzy cluster centres, and *m* is fuzzifier.

The next reliable index for validity fuzzy C-means clustering results is *E* index, formula (2) (Makhalova, 2014).

$$E = \frac{2}{\dfrac{1}{\dfrac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^{2}} + \dfrac{1}{1 - \dfrac{\sum_{i=1}^{n} d_{c,\min}}{d_{1,\min}}}} \qquad (2)$$

where d_{cmin} is the minimal distance between objects in case of *c* clusters and d_{1min} is the minimal distance between objects in case when all objects belong to one cluster.

One of the components of *E* index is based on fuzzy clustering theory and the other one is based on hard clustering theory. The theory of fuzzy clustering is based on the assumption that each object belongs to each cluster with a membership degree. The hard clustering theory is based on the assumption that each object belongs to one cluster, the average distance from the cluster centre and points of this cluster should be minimal. The optimal number of clusters *c** for the dataset *X* can be found by solving $min_{2 \le c \le n-1} E$.

For fuzzy C-means clustering was selected data contains information about unemployment in Czech Republic in 2014. The numerical variables, which affect up on unemployment were selected. On chosen data set was applied fuzzy C-means clustering with Euclidean distance. As Bezdek stated (Bezdek, 1981), selected measure of distance in fuzzy C-means clustering do not affect the accuracy of the results. All calculation were providing in MATLAB R2008a.

Obtained results of validity indices for fuzzy C-means clustering are shown in Table 1.

To sum up the results shown in Table 1 can be chose the optional number of clusters is two.

Table 1: Cluster validity indices, the results

The number of clusters	XB	E
2	**2,344**	**1,173**
3	3,210	2,564
4	4,431	2,645
5	4,875	3,501
6	5,001	4,780

Source: Authors

To give a detail account of obtained results can be ascertain, which objects are assigned to every of two clusters. The first cluster can be called ‘hopeless unemployment’. In this group were assigned young people without high school education (variable ‘education’ to show up by oneself), which have no work more, then 18 months, and older people (more than 45 years old), which are unemployment more, that 18 months too. The second cluster can be called ‘hopeful unemployment’. In this group were assigned young people with high school education, which have no work less than 18 months. Notable is, that the variable ‘Education’, like a qualitative variable not be drawn into clustering procedure, but shown up independently.

CONCLUSION

The fuzzy sets in unemployment problem presented in this paper have been good results on the test case. Using fuzzy C-means clustering algorithm provides the following benefits:

1. the high classification accuracy, achieved through a combination of advantages of fuzzy logic;
2. the learning process is fast, and the result is easy to interpret;
3. the fuzzy techniques can model human knowledge and experience enables the generation of the proposed system. Another good feature of fuzzy clustering in unemployment problem is that the algorithm is capable to produce degree of membership for every unemployed. Those degree of membership allows to determine the degree to which unemployment should be include in “hopeful” unemployed, hence it simplicity economic decision making.

The deficiencies on this phase of the research issue not found. In the future research issue we plan to increase the number of input attributes and increase datasets. Finally, a number of empirical studies are currently under way.

ACKNOWLEDGMENT

This article was created with the help of the Internal Grant Agency of University of Economics in Prague No. IG410045 under the title “The period of unemployment after the crisis”.

REFERENCES

Bezdek, J. C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

Eurostat. Unemployment statistics: http://ec.europa.eu/eurostat/statistics-explained/index.php/Unemployment_statistics

Makhalova, E., & Pecáková, I. (2013). Application fuzzy decision trees for creditworthiness evaluation. In International Scientific Conference "Whither Our Economies". Vilnius, 2013 pp. 26-34, ISSN 2029-8501. URL: http://woe.mruni.eu/wp-content/uploads/2013/11/WOE-2013-ONLINE.pdf

Makhalova, E., Cséfalvaiová, K., & Pecáková, I. (2014). Estimating the Correct Number of Clusters for Overlapping Data. Forum Statisticum Slovacum. 2014, sv. 10/6: 90-96., ISSN 1336-7420 URL:http://ssds.sk/casopis/archiv/2014/fss0614.pdf

Oliviera, J. W. (2007). Advances in Fuzzy Clustering and its Applications, John Wiley & Sons, London, 2007.

Pharr, J. (2012). The Impact of Unemployment on Mental and Physical Health, Access to Health Care and Health Risk Behaviors. In International Scholarly Research Network ISRN Public Health, 2012.

Quinlan, J. R. (1986). Induction on decision trees, in Machine Learning, 1986, 1:81-106.

Řezanková, H., & Löster, T. (2011). Analysis of the Dependence of the Housing Characteristics on the Household Type in the Czech Republic. In: Aplimat [CD-ROM]. Bratislava, 01.02.2011-04.02.2011. Bratislava: Slovak University of Technology, 2011, pp. 1629-1636, ISBN 978-80-89313-52-5.

Riberiro, R. A. (1999). Soft computing in financial engineering, Heidelbarg ,1999, pp. 3-10.

Wanga, W., & Zhang, Y. (2007). On fuzzy cluster validity indices, in Fuzzy sets and systems. 2007, 158: 2095-211.

Xie, X. L., & Beni, G. (1991). Avalidity measure for fuzzy clustering, ieee trans. In Pattern anal. Mach. Intelligence, 1991, 13: 841-847.

Zadeh, L. A. (1975). Fuzzy logic and approximate reasoning. In Synthese ,1975, 407-428.