APPLICATION OF FUZZY CLUSTERING TO COMPARING ECONOMIC SITUATION OF INDUSTRIAL ENTERPRISES

Miroslawa Lasek
Faculty of Economic Sciences, Warsaw University
Dluga 44/50, PL - 00-241 Warsaw Poland
Phone: +48-22-831 32 01, Fax: +48-22-831 28 46
email: mlasek@wne.uw.edu.pl

ABSTRACT: A comparative study of the economic (financial and property) situation of enterprises was carried out with the aid of Fuzzy C-Means method (FCM). By the use of this method various degrees of membership of enterprises in clusters or classes of various economic situation were determined. The analysis of clusters enables the identification of numerous rules and connections based on financial statistics. As evaluation criteria for the state of business of enterprises, financial liquidity, indebtedness, profitability and operational efficiency were used. The situation of Polish listed companies from varied lines of trade were reviewed. The results have been used for the analysis of the economic state of companies which start to be listed on the stock exchange.

KEYWORDS: Economic analysis; Multicriteria Analysis; Comparison of economic situation of enterprises; Cluster Analysis; Fuzzy C-Means Clustering

1. INTRODUCTION

The purpose of this article is to present a new application area for fuzzy clustering, consisting in the analysis of economic (financial and property) situation of enterprises.

The need for carrying out analysis of economic situation of enterprises/companies is due to various reasons, e.g. (i) the management of a company is obliged to provide information on the economic state of a company to its owners who invest and put their own assets at risk, (ii) it is necessary that the managers (management board) of a company become familiar with the results of their own activity, (iii) potential investors, considering the opportunities for making investments in the company are interested in its state of business, (iv) it is necessary for the banks to evaluate a company's ability to meet its financial commitments, and (v) the customers of a company are interested in such information (e.g. in order to assess the stability of supplies).

The analysis of the economic situation, considering its multi-criteria approach, i.e. the need to take into account the varied points of view, such as liquidity, indebtedness, capacity, profitability, and relative nature of such evaluation criteria, requires making comparisons with the economic state of other enterprises.

Making comparisons in respect of economic situation is a problem which is difficult to algorithmisation. It requires expert's knowledge, experience and intuition. Such analyses are burdened with uncertainty and inaccuracy resulting from the conflicting nature among the evaluation criteria (e.g. a company with high liquidity does not need to be profitable, and vice versa a profitable company may show a low financial liquidity), the occurrence of varied values for deviations between the evaluations, and the appearance of incomparable companies, what may result, for instance, from a difference in the scope or method of production.

In such a situation, in order to be able to cope with the problem, we propose to apply the approach based on fuzzy clustering.

This article presents an attempt to apply one of the fuzzy clustering methods: Fuzzy C-Means method (FCM). The method enables the comparison of companies, offering the approach to uncertainty and inaccuracy which is characteristic to this problem, with the aid of the conceptual and mathematical apparatus of the fuzzy set theory.

For the needs of running calculations, and drawing up tables and diagrams, the DataEngine program of the German company MIT-Management Intelligenter Technologien GmbH of Aachen was used (see DataEngine (1997)). This program belongs to the group of programs suited for the data mining, and making use of Fuzzy C-Means Clustering (see e.g. DataEngine (1997), Practical (1997), Zimmermann (1996)).

2. EVALUATION CRITERIA FOR THE ECONOMIC SITUATION OF ENTERPRISES

Depending on the pre-conditions for the evaluation of enterprises (purpose of evaluation, line of trade and the kind of business activity, access to reliable data), the possibility of considering a varied combinations of the following criteria is assumed:

1. financial liquidity, or the ability to meet current commitments,
2. indebtedness, or the admissible level to which a company may be financed via debts,
3. profitability, or the ability to derive profit in the extent which enables funding the current business and development needs,
4. operational efficiency, or the skill of making use of the resources possessed.

The evaluation of a business in respect of each of the above stated criteria is carried out with the aid of various indicators – financial ratios as evaluation subcriteria, e.g. *liquidity* with the aid of current liquidity, quick liquidity, increased liquidity, *indebtedness* - rate of indebtedness, ability to pay interest, *profitability* - rate of return on equity, rate of return on assets, rate of gross profit, rate of operating profit, rate of net profit, *operational efficiency* - working capital, receivables rotation, stock rotation, payables rotation, efficiency of fixed assets, efficiency of assets.

The calculation methods and interpretation of financial ratios have been presented in detail in economic and business literature. Therefore, they will not be discussed in this article.

3. DATA MATERIAL USED IN ANALYSIS

The state of business of 72 enterprises - Polish listed companies from various lines of trade were reviewed, such as:

- foodstuffs (10 companies),
- consumer goods (8 companies),
- wood and furniture (2 companies),
- papermaking and publishing (4 companies),
- chemicals and mineral fibres (12 companies),
- metal processing (1 company),
- electrical engineering (6 companies),
- computers (2 companies),
- construction engineering (13 companies),
- trade (14 companies).

The subject of analysis were the companies for which complete data were possible to be obtained. There were analysed the economic ratios (indicators of financial and property standing) of such companies for the year 1996 (see Financial results (1997)).

For the economic evaluation of enterprises, 15 ratios were reviewed, namely the indicators for: working capital, current liquidity, quick liquidity, increased liquidity, ability to pay interest, receivables rotation by days, stock rotation by days, payables rotation by days, efficiency of fixed assets, efficiency of assets, rate of return on equity capital, rate (margin) of gross profit, rate of operating profit, rate of pre-tax profit, rate of net profit on sales. For these ratios a correlation analysis was made.

Some of the ratios were strongly correlated, e.g. rate of net profit and of pre-tax profit (0.9644; 0.000); rate of increased liquidity and of current liquidity (0.9479; 0.000), rate of quick liquidity and of increased liquidity (0.8798; 0.000); in brackets there were, successively, stated: a correlation coefficient and a significance level (the significance level equal to zero shows a significant correlation).

In order to avoid the effect of „replication of characteristics", a set of the eight least correlated, i.e. the most independent ratios was selected for further analysis:

- working capital,
- rate of quick liquidity,

- receivables rotation by days,
- stock rotation by days,
- payables rotation by days,
- rate of return on equity capital,
- rate (margin) of gross profit,
- rate of net profit.

Basic descriptive statistics of the reviewed ratios have been calculated. Considering the various measurement scales (ranges of values) of the indicators under review (e.g. working capital: - 35 326,8 – 268 498,5 vs. rate of quick liquidity: 0,278 – 7,426), it was necessary to make the normalization of variables. This procedure enabled avoiding a different weighting of variables resulting from the differences in the measurement scales (see Sharma (1996)). The scaling of variables was carried out in order to ensure that all values are within the interval between 0 and 1, i.e. in case if the *i* value of *X* indicator is marked with x_i, according to the formula: $\frac{x_i - \min\limits_i(x_i)}{\max\limits_i(x_i) - \min\limits_i(x_i)}$.

4. FUZZY CLUSTERING OF INDUSTRIAL ENTERPRISES WITH THE FUZZY C-MEANS ALGORITHM

Comparative study of the economic situation of enterprises was carried out with the aid of Fuzzy C-Means method (FCM), an iterative method of clustering. In this method, cluster centres, i.e. objects which „best" represent clusters, are being obtained in the process of iteration.

The number of clusters should be given before commencing procedure of iterative clustering. The number of clusters considered must be higher or equal to 2, and lower or equal to the number of objects for which such clusters are built up. We must also state the so-called *m* exponent; $m > 1.0$. The *m* exponent shows the extent of fuzziness of the results. In case if $m \to 1$, the results start to be more and more closer to the results obtained with the aid of a „non-fuzzy" method. In case if $m \to \infty$, the values of membership of objects to clusters tend to take values more and more closer to the inverse of a number of *c* classes, i.e. $\frac{1}{c}$.

The Fuzzy C-Means algorithm was presented in DataEngine (1997), Hirota and Pedrycz (1996), Weber (1996), Zimmermann (1996).

The algorithm determines the cluster centres and the values of degrees of membership for particular objects to clusters. Since euclidean distance was assumed as a measure of distance, clusters with a spherical structure are possible to be detected. The „stretching" clusters are not identified by the algorithm (see e.g. DataEngine (1997)).

Before the grouping of companies is commenced, the number of clusters into which we want to subdivide them has to be indicated. In order to state the number of clusters, quality indicators: *pc* (*partition coefficient*) and *pe* (*partition entropy*) were used (see DataEngine (1997), Zimmermann (1996)):

a) *pc* (*partition coefficient*)

$$pc = \sum_{k=1}^{K} \sum_{i=1}^{c} \frac{\left(m_{ik}\right)^2}{K}$$

b) *pe* (*partition entropy*)

$$pe = -\frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{c} m_{ik} \ln\left(m_{ik}\right)$$

where: *K* is the number of objects subdivided in *c* classes, m_{ik} is a degree of membership of *k* object to *i* cluster ($k = 1, ..., K; i = 1, ... c$).

For the above presented measures, there occurs:

$$\frac{1}{c} \le pc \le 1,$$

$$0 \le pe \le \ln(c).$$

The higher is *pc* and the lower is *pe*, the result of clustering is „less fuzzy". When $pc = 1$ and $pe = 0$, we have to do with „non-fuzzy" clusters. The values of *pc* and *pe* indicators for different number of clusters: from 2 to 16 were examined. The more strongly the objects concentrate around the cluster centres, the better is the structure of clusters,

and thus the quality of results is higher. The „best" number of clusters is the number where the value of the *pe* entropy indicator is below a growing trend (together with the number of clusters), and of the *pc* indicator is above a falling trend (together with the number of clusters) (see DataEngine (1997)). In our case, as shown in Figure 1, a possible solution is 6 clusters.

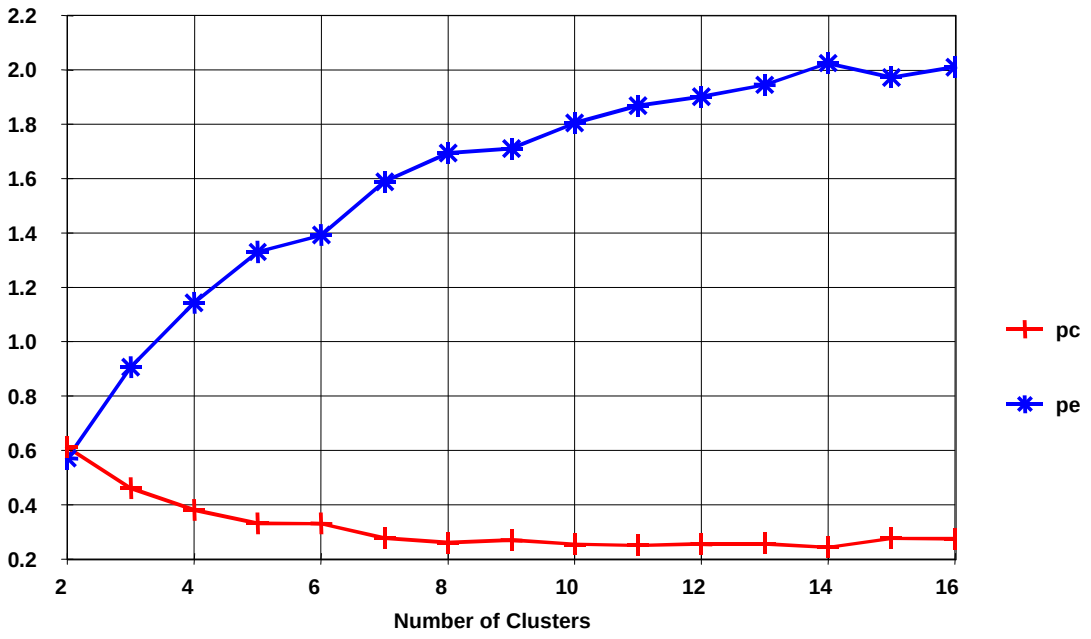

Figure 1. Diagram of *pc* and *pe* quality indicators for cluster analysis of enterprises
Source: The results as obtained from *DataEngine* program

Assuming the eight criteria for the evaluation of enterprises (working capital, rate of quick liquidity, receivables rotation by days, stock rotation by days, payables rotation by days, rate of return on equity capital, rate (margin) of gross profit, rate of net profit) and the six clusters, cluster centres and degrees of membership of companies to clusters were generated.

5. INTERPRETATION OF THE RESULTS

In Figure 2 a chart of the profiles of six cluster centres was shown (marked *A*-*F*).

The *A* cluster includes companies with a relatively high working capital, high rate of gross profit, and with a slow stock rotation. Among the first ten companies with the highest degree of membership in the cluster, there are as many as five companies of chemicals and mineral fibres industries and two companies of electrical engineering industry.

The *B* cluster represents companies with a small working capital and not high liquidity, characterised by a relatively slow rotation of stocks and payables. Among the first ten companies with the highest degree of membership to the *B* cluster there are four trading companies, two companies of electrical engineering industry and two companies of consumer goods industry.

The *C* cluster represents companies with a high working capital, characterised by slow rotation of receivables and payables, with not so high rate of gross profit and high level of quick liquidity. Among the first ten companies with the highest degree of membership to the *C* cluster there are three construction engineering companies, three trading companies, and two companies of chemicals and mineral fibres industries.

Within the *D* cluster there are companies of quick rotation of payables and receivables, characterised by a high rate of gross profit and with very high liquidity. Among the first ten companies with the highest degree of membership to the *D* cluster there happened to be as many as five companies of chemicals and mineral fibres industries, and three companies of consumer goods industry.

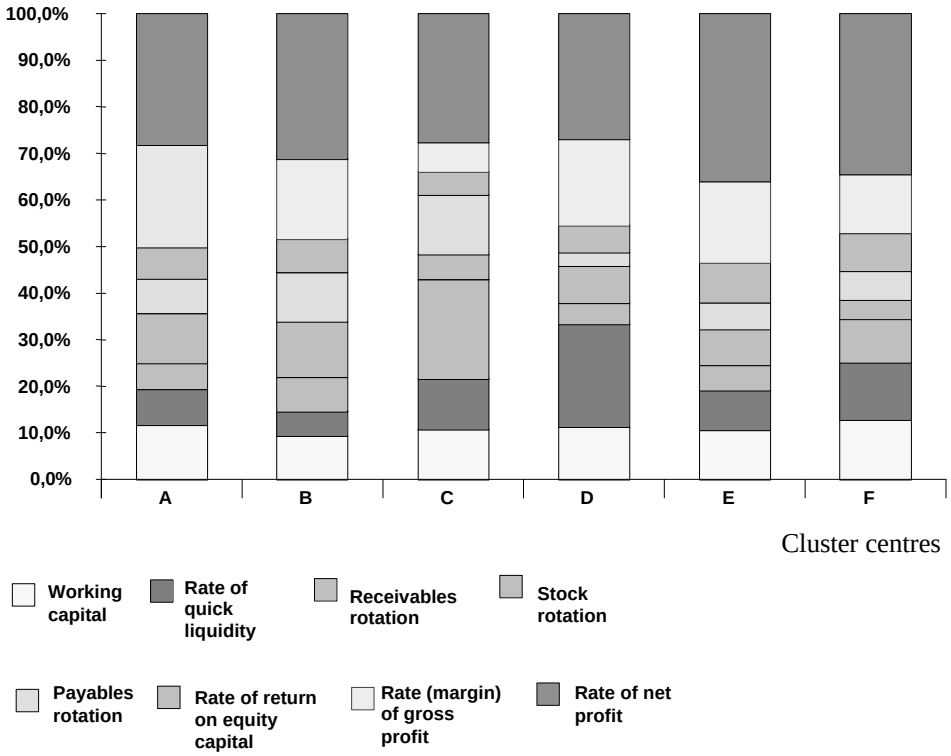

Figure 2. Bar chart of the profiles of cluster centres *A*-*F* of companies
Source: The results as obtained from *DataEngine* program

The *E* cluster represents companies with a rather small working capital, with a relatively high rate of return on equity capital, and with quick rotation of receivables and payables. Among the first ten companies with the highest degree of membership belonging to the *E* cluster, four companies of food industry, and two companies of papermaking and publishing industries were found.

The *F* cluster confines companies with a high working capital, quick rotation of stocks and relatively quick rotation of payables, but with slow rotation of receivables. Among the first ten companies with the highest degree of membership to the *F* cluster as many as eight of them are construction engineering companies.

The analysis of clusters enables the identification of numerous rules and connections based on financial statistics. Let us review for example the difference in the affiliation of companies to clusters depending on industrial branch.

Figure 3 presents the affiliation degrees of the three food industry companies (*p1*, *p2* and *p3*) and of the three electrical engineering companies (*p4*, *p5* and *p6*) to six clusters (*A*-*F*). The food industry companies, with the highest membership degrees, belong to the *E* cluster (degrees of membership of companies to considered cluster are: 0.40, 0.52, 0.55). Two electrical engineering companies, with the highest degrees of membership, belong to the *B* cluster (membership degrees: 0.78 and 0.61), and one company – to the *A* cluster (degree of membership: 0.34). The latter company has also a high degree of membership to the *B* (0.27). The food industry companies most weakly belong to the *C* cluster, with the following membership degrees: 0.02, 0.01 and 0.01 and the electrical engineering companies belong to the discussed cluster with very low membership degrees: 0.01, 0.03 and 0.06. The companies of the both industrial branches belong weakly also to the *D* cluster (see Figure 3).

In Figure 4, the three companies with the highest rate of net profit (namely *z1*, *z2* and *z3*), and the three companies of the lowest rate of net profit (*z4*, *z5* and *z6*) are presented. The two companies of the highest rate of net profit belong to a trading branch, and one - to construction engineering industry. Among the companies of the lowest rate of net profit is a trading company, wood and furniture company and a construction engineering company. Within a group of both the highest and the lowest rate of net profit, construction engineering companies were found. The companies with the highest rate of net profit, with absolutely the highest membership degree, belong to the *C* cluster (membership degree of each of those companies to *C* is: 0.64, 0.76 and 0.35). The affiliation of the companies of the lowest rate of net profit to clusters is rather controversial. If we take into consideration their degrees of membership to clusters, they mostly belong to the *B* cluster, but their membership degrees to the discussed cluster (0.34, 0.22 and 0.19) are not so much higher than those to other clusters (for example, the degree of membership of *z4* company to the *F*

cluster is 0.22, and of *z5* company to the same cluster is 0.21, whereas in the case of *z6* company it is equal to its affiliation to the *B* cluster (0.19)).

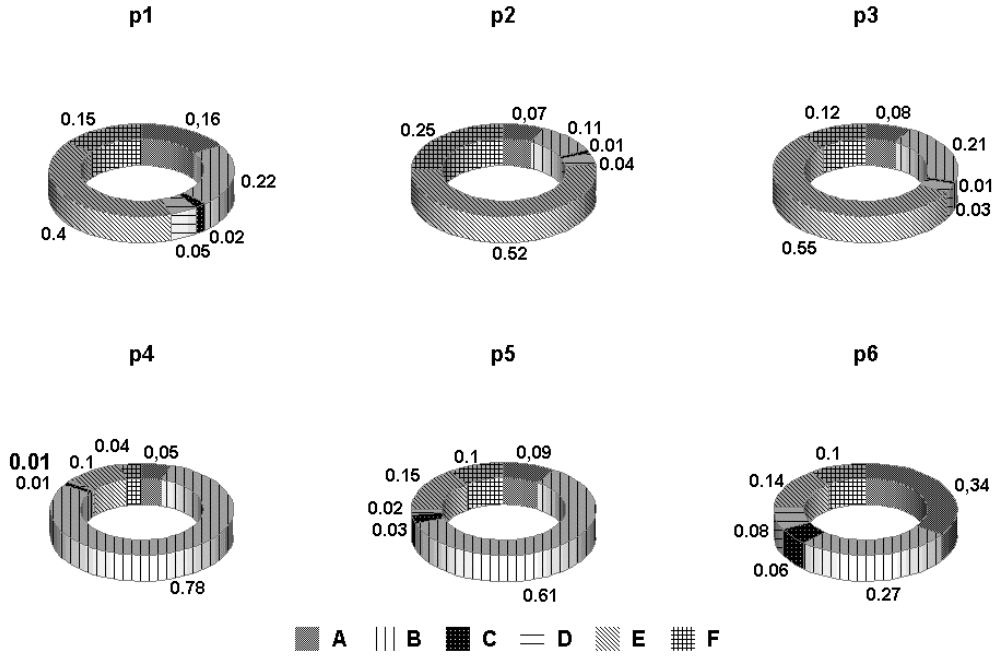

Figure 3. Membership degrees of *p1*, *p2*, *p3* food industry companies and *p4*, *p5*, *p6* electrical engineering companies to clusters
Source: The results as obtained from *DataEngine* program

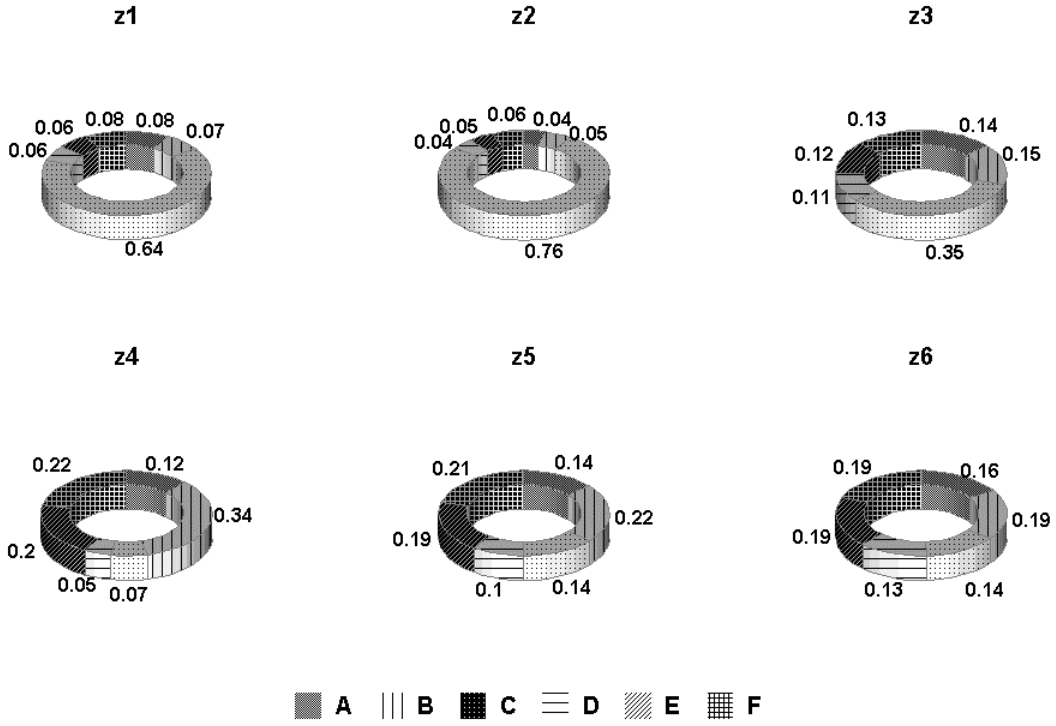

Figure 4. Membership degrees of *z1*, *z2* and *z3* companies of the highest rate of net profit (respectively from: trade, construction engineering industry and trade), and of *z4*, *z5* and *z6* of the lowest rate of net profit (respectively from: trade, wood and furniture industry, and construction engineering industry)
Source: The results as obtained from *DataEngine* program

The histograms presented in Figure 5 illustrate the number of companies belonging to the clusters reviewed, with various degrees of affiliation. They show that small number of companies „decidedly" belongs to particular clusters, i.e. with a very high degree of membership. This indicates considerable „fuzziness" of the problem.

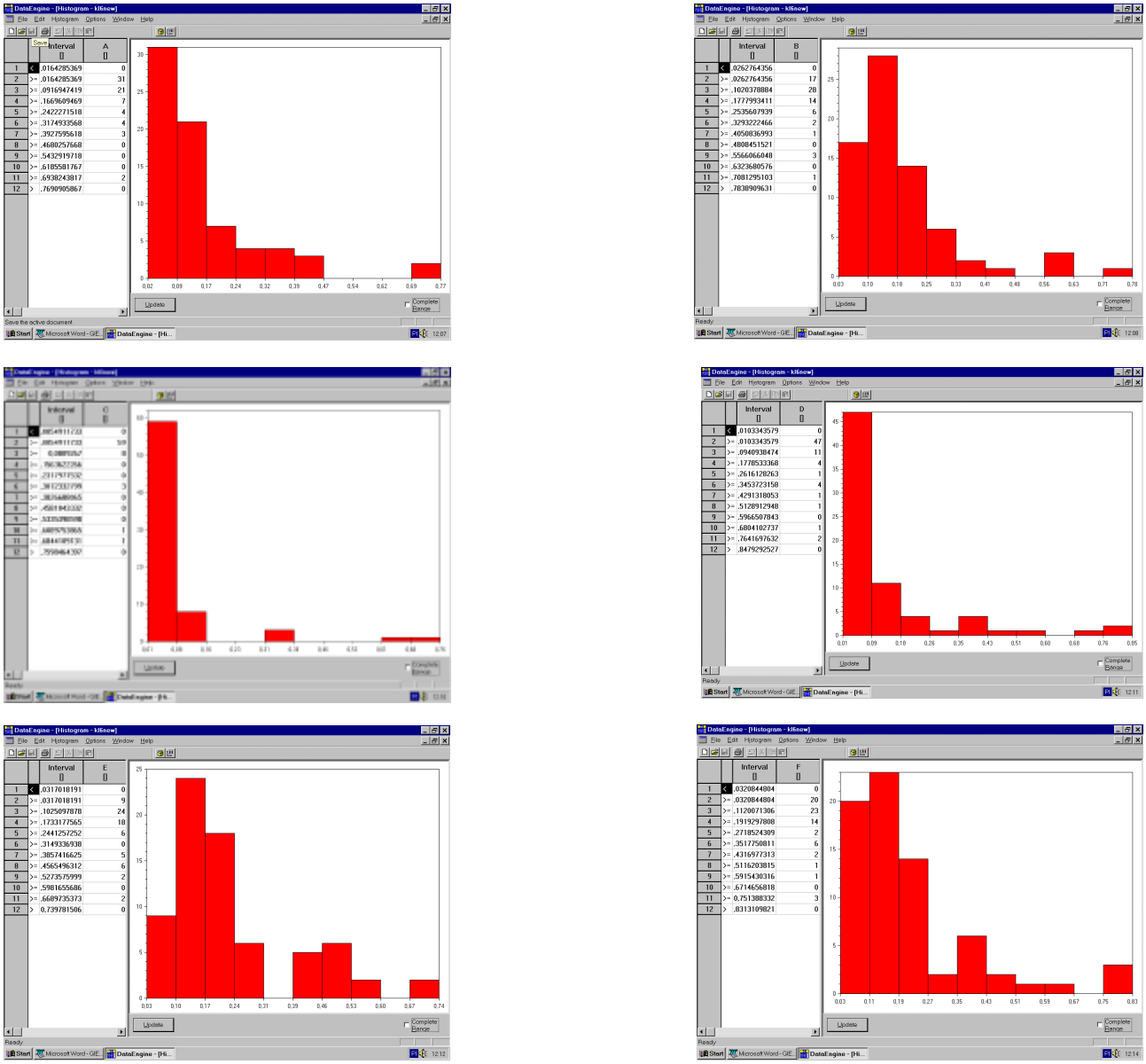

Figure 5. Histograms for membership degrees of companies to individual clusters (*A* – *F*)
Source: The results as obtained from *DataEngine* program

The results of the subdivision of companies into clusters have been used for the analysis of the state of business of companies which start to be listed on the stock exchange. A starting point for such analysis was the positioning of such companies in relation to cluster centres, as shown in Figure 6, in respect of the three criteria: quick liquidity, rotation of payables by days, and rate of net profit.

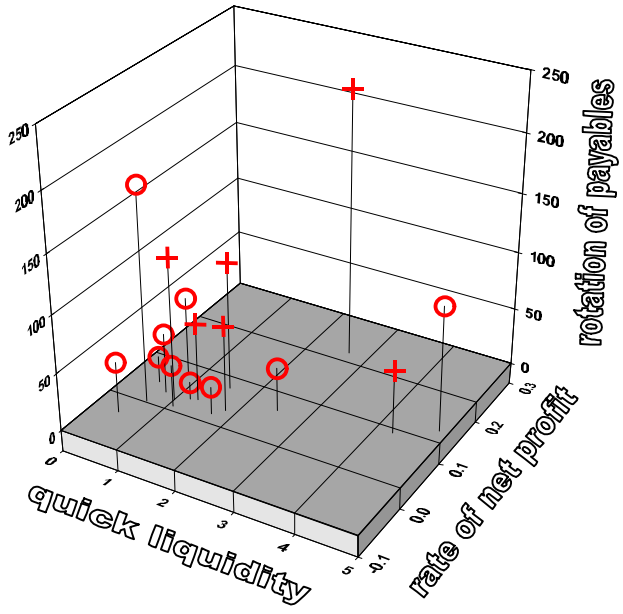

Figure 6. Cluster centres (crosses) and companies (circles) in respect of three characteristics: quick liquidity, rotation of payables and rate of net profit
Source: The results as obtained from *DataEngine* program

6. CONCLUSIONS

The comparative study of the economic situation of enterprises is one of the new practical applications of Fuzzy Clustering. This method seems to be suitable for comparing the economic state of industrial enterprises. In all the test cases considering so far we have obtained reasonable results, i.e. reasonable clusters have been identified on the test data.

The application of Fuzzy C-Means Clustering enabled to avoid making a judgement whether a given object belongs or does not belong to a certain cluster or class (as in a traditional cluster analysis: see e.g. Michaud (1997), Sharma (1996), Zait and Messatfa (1997)). With the aid of the method discussed, various degrees of membership of enterprise to clusters or classes of various economic condition are determined. Such an approach reflects the reality much more since it enables to avoid assigning „forcedly" each enterprise to only one cluster or class, what results in losing a considerable amount of information in the process of clustering.

Fuzzy C-Means method does not require any assumptions to be made as far as the nature of input data are concerned. Nevertheless, it requires certain, often arbitrarily determined, parameters. These include measure of distance and number of clusters into which enterprises are subdivided. However, it should be admitted that if we modify the parameters within certain limits, some basic conclusions remain essentially unaltered.

The main disadvantage of the method is the fact that we can obtain many different clustering results depending on the choice of criteria, weighing them differently, and by fixing the number of clusters or measure of distance. It often requires some experimental search, and depending on the context of the application, we may then be able to produce the „best" clusters for the ultimate purpose.

Another disadvantage is the fact that Fuzzy Clustering may represent only spherical clusters.

A precondition for the application of the method, as described in this article in the comparative analysis of the economic situation of enterprises, is the possession of a data set enabling the formation of clusters and the correct selection of criteria for the evaluation of enterprises. In the selection of criteria there may be helpful a statistical analysis with the aid of traditional methods (descriptive statistics, correlation analysis).

The implementation of the method is only possible owing to suitable software: a very convenient tool appears to be DataEngine program of the German company Management Intelligenter Technologien GmbH from Aachen.

Although our results show the suitability of the method, there still remain some problems requiring further research and solution. Among them, there is the above mentioned problem of parameters selection. Also, the choice of evaluation criteria for enterprises is still largely heuristic in nature. It is also related to the question of assigning weights

to criteria. All these problems depend on a clustering task, e.g. forecasting, competition analysis, evaluation of credit ability. It would be also worthwhile to work out helpful indications or even methods for the interpretation of results, e.g. confidence intervals and hypothesis verification, as being proposed by traditional statistical tests.

When referring to the comparative analysis of the situation of industrial enterprises, there also seems promising the application of fuzzy methods other than Fuzzy Clustering as described in this article, for example such as Fuzzy Kohonen Networks. It seems realistic that fuzzy rule expertise base is built up with the aid of WinRosa program of the Management Intelligenter Technologien GmbH. The efforts aimed at using Fuzzy Kohonen Networks and building up a rule base for the analysis of the economic situation of industrial enterprises have already been made by our team within the Faculty of Economic Sciences of the Warsaw University.

7. REFERENCES

[1] DataEngine. From Data to Information. (1) Tutorials and Theory, (2) Overview and User Manual, 1997, MIT GmbH, Aachen.

[2] Financial results of publicly listed companies (in Poland), 1997, Notoria Serwis, Warsaw, March 1997.

[3] Hirota, K.; Pedrycz, W., 1996, “Directional fuzzy clustering and its application to fuzzy modelling”, Fuzzy Sets and Systems 80 (1996) 315-326.

[4] Michaud, P., 1997, “Clustering techniques, Future Generation Computer Systems” 13 (1997) 135-147.

[5] Practical Applications of Intelligent Data Analysis, 1997, Proceedings, First International Data Analysis Symposium, September 12, 1997, Aachen, Germany.

[6] Sharma, S., 1996, “Applied Multivariate Techniques”, John Willey & Sons, New York.

[7] Weber, R., 1996, “Customer Segmentation for Banks and Insurance Groups with Fuzzy Clustering Techniques”, in: J.F. Baldwin (Ed.), Fuzzy Logic, John Wiley & Sons, pp. 187-196.

[8] Zait, M.; Messatfa, H., 1997, “A comparative study of clustering methods”, Future Generation Computer Systems 13 (1997) 149-159.

[9] Zimmermann, H.-J., 1996, “Datenanalyse, Anwendung von Dataengine mit Fuzzy Technologien und Neuronalen Netzen”, VDI-Verlag, Düsseldorf.