Regional and Business Studies (2015) Vol 7 No 1, 51-60
Kaposvár University, Faculty of Economic Science, Kaposvár

REGRESSION BASED MODELS AND EXPERT JUDGEMENT IN PREDICTIVE SITUATIONS

Gábor HÁMORI
Kaposvár University, Doctoral School of Management and Organizational Sciences
H-7400 Kaposvár, Guba S. u. 40

ABSTRACT

The main scope of this paper to give a short conceptual overview and comparison of regression based multivariate techniques and professional expert judgements in decision-making situations where the goal is to predict the value or the membership of a certain target variable. As a secondary research we overview the related literature. Within the framework of the survey, the most widespread CRISP-DM model development standard and the selection criterias of the appropriate regression based model are discussed as well.

Keywords: expert judgement, classification, regression, logit, CRISP-DM

INTRODUCTION

When browsing on an online bookshop, it is not unusual that we are automatically offered certain books which are very likely to fit our interest. We receive text messages from our mobile service provider, in which we are directly offered new products. When setting up a claim for a bank loan, our request is immediately judged by the loan-office. The above listed examples share that, in all cases, the calculations require predictive classification methods which forecast our expected behaviour.

In the last few decades, a noticeable technological revolution took place in the field of computer science. The sudden explosion of developments enabled us to do millions of operations on personal computers. Prior to that the application of complicated mathematical-statistical algorithms (henceforward: algorithms) in practice were only accessible to circles of academic researchers because of the process' significant time consuming nature and the limited access of the capacity of computers. However, the situation has drastically changed. A separate industry has been created in order to make the algorithms available to users. Special software finally became accessible that enables us to carry out in a few minutes analyses that formerly took weeks.

The quick pace of the expansion of computing possibilities and easy access have given a further push to the fundamental researches, and to the innovation of algorithms. New procedures and algorithms are regularly unveiled in professional scientific workshops. The most viable ones are immediately used in practice. The possibilities of new technologies in practice were first utilised by the corporate/ business sector. In those industries that are characterized by the fulfilment of massdemand and service, the process of business statistical analysis and modelling are evident. In situations of the claim for multiple decisions modern enterprises soon realized that, as a result of algorithms, predictive decision rules/formulas help them to make decisions faster and, more importantly, cheaper.

STATISTICAL ALGORTHMS VERSUS EXPERT EVALUATION

In contrast to the experts' judgement, these algorithm based decisions are free from cognitive distortions or biases. In general cognitive biases can be characterized as the tendency to make decisions and take action based on limited acquisition and/or processing of information or on self-interest, overconfidence, or attachment to past experience. Cognitive biases can result in perceptual blindness or distortion (seeing things that aren't really there), illogical interpretation, inaccurate judgments, irrationality (being out of touch with reality), and bad decisions. Cognitive biases can be broadly placed in two categories. Information biases include the use of heuristics, or information-processing shortcuts, that produce fast and efficient, though not necessarily accurate, decisions and not paying attention nor adequately thinking through relevant information. Ego biases include emotional motivations, such as fear, anger, or worry, and social influences such as peer pressure, the desire for acceptance, and doubt that other people can be wrong.

Decision algorithms are free from tiredness and exclude subjective elements. Another important difference is that algorithms are coherent in time, namely, they evaluate the same situation of a decision in two different times. A good example for that is the research (*Hoffman et al.*, 1968) in the course of which experienced radiologists were asked to evaluate chest radiographs into the categories of "normal" and "abnormal" in two different times, so those who participated in the research did not know that they were shown the same radiographs for the second time. In the 20% of the cases, the assessments were self-contradictory. A imilar per cent of inconsistency was observed in a research in which 101 auditors were asked to rate the reliability of the inside controls of the company (*Brown*, 1983).

In those situations, when the examined problem's complexity "size" is significant as a result of the decision point's complexness, algorithms are able to reveal such correspondences that a human mind would not be able to capture. Let's take the example of half a million clients and a loan-office possessing an accordingly vast amount of descriptive data. For instance, where they are looking for answers to these questions: How and what does someone's paying back of a loan depend?, or - Is it possible to give a probability estimate in the case of a new client who is likely to not be able to pay a loan back, or to decide who qualifies to receive a loan based on this data? With the help of suitable statistical analyses, these factors can be characteristically disclosed, which have an effect on the payment willingness. Together with the help of proper algorithm, a kind of formula can be produced that is able to estimate a specific case of a client's non-payment probabilit. Based on the revealed correspondences, such business decision making rules can be set up that run along with a target function, resulting in an optimal procedure, and are automatically executed during the application[1].

On the grounds of the above mentioned decision point's complexity, it might not be surprising that in these fields human (professional) decision/estimation performs worse than algorithmic analysis.

At the same time, it is surprising that in many other instances the same thing happens in the "small-sample" cases, namely, where the problem can only be described with far less data. *Paul* (1954), summed up 20 such research results that examined whether the qualified professionals' subjectivity-based evaluation on the clinical predictions were more accurate or the statistical predictions derived from given rules.

Based on the outcome of the research, the statistical prognoses are characteristically performed better than the ones made by professional estimations. Even more striking is that of *Kahneman*, who was honoured with the Nobel Prize in Economics in 2002 for his book *Thinking, Fast and Slow*, who wrote this:

"About 60% of the studies have shown significantly better accuracy for the algorithms. The other comparisons scored a draw in accuracy, but a tie is tantamount to a win for the statistical rules, which are normally much less expensive to use than expert judgment. No exception has been convincingly documented" (*Kahneman*, 2013).

The effectiveness and the relative power of statistical predictions can be observed in exotic areas, like *Ashenfelter*'s wine price forecast. In this study (*Ashenfelter*, 2007) a proper linear model was developed to predict the price of mature Bordeaux red wines. Surprisingly, the model performed better at auction than expert wine tasters did. This particular proper linear model has the following form

$$P = w_1(c_1) + w_2(c_2) + w_3(c_3) + \dots w_n(c_n) \tag{1}$$

The model calculates the summed result P, which aims to predict a target property such as wine price, on the basis of a series of variables. Above, c_n is the value of the nth variable, and w_n is the weight assigned to the nth predictor.

In the wine-predicting statistical model, c_1 reflects the age of the vintage, and other predictors reflect relevant climatic features where the grapes were grown. The weights for the cues were assigned on the basis of a comparison of these variables to a large set of data on past market prices for mature Bordeaux wines.

As a short list of studies here are some other examples (including small sample problems) of using statistical models successfully in contrast to expert judgement:

[1] Connected to the previously mentioned two questions, the bank-loan approval, or the so called scoring systems applied by great banks, is a good example. These systems work automatically based on the given data, with the exclusion of human agency, and they accomplish the estimation of the loan-applicants non-payment probability within a few seconds, and based on the results they take an offer whether to sign a contract or not if the non-payment probability is high.

- *Howard and Dawes* (1976) found they can reliably predict marital happiness with one of the simplest statistical model, using only two cues: P = [rate of lovemaking] - [rate of fighting]. The reliability of this model was confirmed by *Edwards and Edwards* (1977) and by *Thornton* (1979).
- *Wittman* (1941) constructed a statistical model that predicted the success of electroshock therapy for patients more reliably than the medical or psychological staff.
- *Carroll et. al.* (1988) developed a statistical model that predicts criminal recidivism better than expert criminologists.
- A predictive model constructed by *Goldberg* (1968) did a better job of diagnosing patients as neurotic or psychotic than did trained clinical psychologists.
- Statistical models regularly predict academic performance better than admissions officers, whether for medical schools (*DeVaul et al.,* 1957), law schools (*Swets et al.*, 2000), or graduate school in psychology (*Dawes*, 1971).
- Statistical models predict loan and credit risk better than bank officers (*Stillwell et. al.*, 1983).
- The prediction of newborns at risk for Sudden Infant Death Syndrome better with models than human experts do (*Lowry*, 1975; *Carpenter et al.*, 1977; *Golding et al.*, 1985).
- Statistical models are better at predicting who is prone to violence than are forensic psychologists (*Faust and Ziskin*, 1988).
- *Libby* (1976) found a simple model that predicted firm bankruptcy better than experienced loan officers.

According to some studies (*Leli and Filskov*, 1985; *Goldberg*, 1968) even when experts are given the results of statistical models, they still can't outperform the predictions of the models. As a conseqence Robyn M. *Dawes* (2002) drew out the normative implications of such studies:

„If a well-validated statistical model that is superior to professional judgment exists in a relevant decision making context, professionals should use it, totally absenting themselves from the prediction."

An unexpected result of the researches, the superiority of the statistical predictions was even observable as the rules of the explanatory factors of simple linear combinations (*Kahneman*, 2013). Furthermore, Robyn M. *Dawes* presents evidence in the 1979 article *The robust beauty of improper linear models in decision making* that even such improper linear models are superior to clinical intuition when predicting a numerical criterion from numerical predictors. Improper linear models are those in which the weights of the predictor variables are obtained by some nonoptimal method; for example, they may be obtained on the basis of intuition, derived from simulating a clinical judge's predictions, or set to be equal.

The statistical predictions are based on models that are not only useful to forecast but also to reveal the predictions' explanatory factors, their correspondences with each other and with the required predicted quantity. Therefore, they are not just used as well-functioning "crystal balls" during the prediction but we can also gain knowledge, understanding the phenomenon of modelling from its inside mechanisms and its system of coherence. Schematized, we could have said that with the appropriate database a good statistical model is able to "learn" everything within a few minutes, as opposed to an expert of the field who might need decades to acquire this knowledge.

So far, we may conclude that in situations of prediction human wisdom and professional knowledge are completely unnecessary. However, "luckily" the statistics based algorithmic prognoses are not free from disadvantages. One instance of this is Paul *Meehl*'s "broken leg phenomenon". *Meehl*, in his mind-experiment, assumed that we possess a statistical algorithm that is able to predict based on previous experience that a specific professor is going to the cinema on Wednesday evening. The algorithm works perfectly until the time when the professor suddenly breaks his leg on a Tuesday, so he cannot go to the cinema on Wednesday. Thereby, algorithms malfunction in those situations where a previously never experienced, low probability and rare situation occurs, whose result is significant[2].

Another problem is the development of the particular algorithm, specifically its base, related to the structure of data. The predicting models can only perform good results in the world represented by the data. If there are some important data missing for some reason, the variables that highly affect the required prediction of quantity will worsen the quality of the prediction. In practice a very important problem is the quality of the data. Data quality essentially relates to the predictive power of the statistical models. If we run a model on inaccurate data, the explained contexts and prediction of the model will be misleading. The so-called GIGO rule sharply highlights this situation: "Garbage In Garbage Out", that is, if "garbage" is used in the modeling, the result will be "junk".

INDUSTRY STANDARD OF MODEL DEVELOPMENT

The correct practical development and application of predictive models is provided by the Cross Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM was conceived in 1996. In 1997 it got underway as a European Union project under the ESPRIT (European Strategic Program on Research in Information Technology) funding initiative. The project was led by five companies: SPSS, Teradata, Daimler AG, NCR Corporation and OHRA, an insurance company. The process diagram on *Figure 1* shows the relationship between the different phases of CRISP-DM.

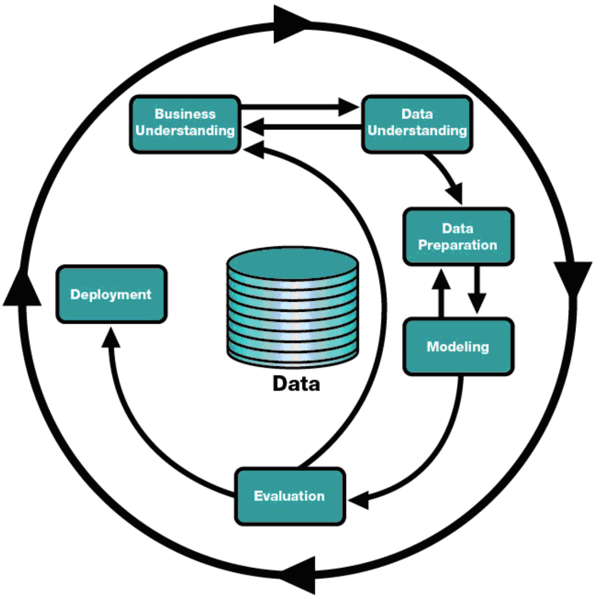
The sequence of the phases is not strict, and moving back and forth between different phases is always required. The arrows in the process diagram indicate the most important and frequent dependencies between phases. The next informative *Figure 2* shows a breakdown of CRISP-DM with tasks and roles for each iteration.

As can be seen, building up and applying a model is a very complex multistep process that requires knowledge and experts from different specializations. In the further section we focus on the modelling step, namely how an appropriate predictive method can be selected.

[2] In these short-term economic predictions, such an atypical event is the turn of the trend or crises situation

Figure 1

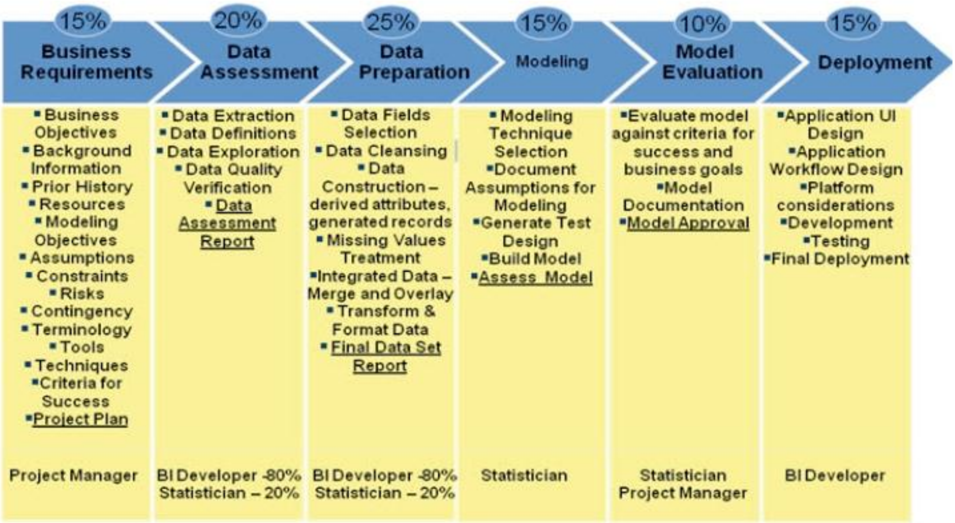
Process diagram of CRISP-DM



Source: *http://www.infobuild.co.za/wp-content/uploads/2012/10/www_predictive_analysis_11.jpg*

Figure 2

BreakdownofCRISP-DM



15%	20%	25%	15%	10%	15%
Business Requirements	Data Assessment	Data Preparation	Modeling	Model Evaluation	Deployment
▪ Business Objectives ▪ Background Information ▪ Prior History ▪ Resources ▪ Modeling Objectives ▪ Assumptions ▪ Constraints ▪ Risks ▪ Contingency ▪ Terminology ▪ Tools ▪ Techniques ▪Criteria for Success ▪Project Plan	▪ Data Extraction ▪ Data Definitions ▪ Data Exploration ▪ Data Quality Verification ▪ Data Assessment Report	▪ Data Fields Selection ▪ Data Cleansing ▪ Data Construction – derived attributes, generated records ▪ Missing Values Treatment ▪Integrated Data – Merge and Overlay ▪ Transform & Format Data ▪ Final Data Set Report	▪ Modeling Technique Selection ▪Document Assumptions for Modeling ▪Generate Test Design ▪Build Model ▪Assess Model	▪Evaluate model against criteria for success and business goals ▪Model Documentation ▪Model Approval	▪Application UI Design ▪Application Workflow Design ▪Platform considerations ▪Development ▪Testing ▪Final Deployment
Project Manager	BI Developer -80% Statistician – 20%	BI Developer -80% Statistician – 20%	Statistician	Statistician Project Manager	BI Developer

Source: *http://www.infobuild.co.za/wp-content/uploads/2012/10/www_predictive_analysis_21.jpg*

SELECTION OF THE PREDICTIVE MODEL

The most frequent prediction tasks can be separated into two sections from a statistical point of view. One of them is when the predicted result is defined on a quantity scale. Typical examples are the expected tax incomes or the prediction of given products' selling figures. The most popular technique of these type predictions is multivariate linear regression. The other predicting situation takes place when the forecast refers to a kind of group where the prediction belongs, namely the target variable's outcome is a category. These situations are called classifications. In practice, classification exercises could be enumerated at length. Its importance in economics is the customer non-payment predictions (scoring) of the big enterprises, first of all banks, clients' non-payment probability. In these cases, classification refers to the prediction of which group a client belongs to: "paying" or "not paying". Another, mainly in the case of telecommunication companies, the prediction of the churn of the clients is a highly important task. Based on the predictable future behaviour of the clients, we can distinguish two categories: "stays" and "churns". It is also a classification task. Statistical classification is used to reveal cheats. The tax authorities use statistical algorithms to state a given client's probability of evading tax, or rather to pay correctly. In healthcare researches and applications, it is frequently used to determine chance a particular patient has in the occurrence of a given disease or state.

The listed examples show the great influence binary classifications have among predicting tasks these days. Generally, the aim of the listed predicting situations is twofold. On the one hand, it is highly important to have good classification efficiency, so-called "good predicting ability". It is especially important, in the case of profit orientated enterprises. On the other hand, with the help of modelling we can understand what factors lead to the occurrence or avoidance of a particular event. Several statistical methods can be used in order to reach these goals.

In practice, binary classifications out of statistical algorithms are the most popular, and the most often used traditional technique is the dichotomy logistic regression[3]. Its popularity comes from many factors. First, it has hardly any restrictions in the processing of data, so it is seen as a robust method. Adequate predicting efficiency is paired with the good interpretability of the model; in the case of sufficient requirements, the factors on the occurrence of the event and their prediction's weight is identified with the model. Its spread relies highly on its accessibility in every statistical/data mining software available on the market. Connected to this, in higher education the topic of advanced, multivariate statistics has been taught for years as the basis of the curriculum. The interest in this method indicates that it is a

[3] As a reminder: In the case of binary logistic regression, the natural logarithm of the oddity of estimated classification probability of the two group membership can be described as a linear function of explanatory variables:

$$\log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)=\beta_0+\beta_1 X_1+\ldots+\beta_k X_k=\beta' x \qquad (2)$$

very well researched field, with many publications about the different technical details of logistical regression in the last few decades.

CONCLUSIONS

The quick changing, atypical, low level of structure and data with missing the decision point situations proves that the human wisdom still seems to be inevitable. In this field the analysis of the effectiveness of combining statistical models with expert judgements should be the subject of further researches.

However, the priority of statistical predictions is unquestionable in those environments where the mass-like decisions and the information are structured and are expansively available[4]. Nevertheless, the predictive power of a statistical model depends largely on the availability of data and the data quality, as well as the applied methodology for model development. In most cases of binary decision situation, multivariate logistic regression is one of the most favourable modelling techniques.

REFERENCES

Ashenfelter, O. (2007): Predicting the quality and prices of bordeaux wines. American Association of Wine Economists working paper, No. 4 [online] <URL: http://www.wine-economics.org/workingpapers/AAWE_WP04.pdf>

Bloom R.F., Brundage E.G. (1947): Predictions of success in elementary school for enlisted personnel. In: Stuit D.B. (ed.): Personnel research and test development in the natural bureau of personnel, 233-261. p. Princeton: Princeton University Press.

Brown P.R. (1983): Independent auditor judgment in the evaluation of internal audit functions. In: Journal of accounting research, 21. 444-455. p.

Carpenter, R.G., Gardner, A., McWeeny, P.M., Emery J.L. (1977): Multistage scoring system for identifying infants at risk of unexpected death. In: Arch Dis Child. 52. 8. 606–612. p.

Carroll, J.S., Wiener, R., Coates, D., Galegher, J., Alibrio, J.J. (1988): Evaluation, diagnosis, and prediction in parole decision-making. In: Law and Society Review, 17. 199-228. p.

Dawes, R.M. (1971): A Case study of graduate admissions: Applications of three principles of human decision-making. In: American Psychologist, 26. 180-188. p.

Dawes, R.M. (1979): The robust beauty of improper linear models in decision making. In: American Psychologist, 34. 7. 571-582. p.

Dawes, R.M. (2002): The ethics of using or not using statistical prediction rules in psychological practice and related consulting activities. In: Philosophy of Science, 69. S178-S184.

[4] Nowadays, such a typical environment is an enterprise (the most characteristic are banks/insurance companies, the pharmaceutical industry, telecommunications, trade and mass-production) Further examples could be the government system, for instance, the health care, retirement system or the authorities of taxation.

DeVaul, R.A., Jervey, F., Chappell, J.A., Carver, P., Short, B., and O'Keefe S. (1957): Medical school performance of initially rejected students. In: Journal of the American Medical Association, 257. 47-51. p.

Faust, D., Ziskin, J. (1988): The expert witness in psychology and psychiatry. In: Science, 241. 1143−1144. p.

Goldberg L.R.(1968): Simple models of simple process? Some research on clinical judgments. In: American Psychologist, 23. 483-496. p.

Golding, J., Limerick, S., MacFarlane A. (1985): Sudden Infant Death: Patterns, Puzzles and Problems. Somerset: Open Books.

Edwards, D., Edwards, J. (1977): Marriage: Direct and Continuous Measurement. In: Bulletin of the Psychonomic Society, 10. 187-188. p.

Hand D.J. (1997): Construction and Assessment of Classification Rules. Chichster: John Wiley and Sons 232 p.

Hoffman, P.J., Slovic, P., Rorer, L.G. (1968): An Analysis-of-Variance Model for the Assessment of Configural Cue Utilization in Clinical Judgment. In: Psychological Bulletin 69. 338-339. p.

Hosmer, D.W., Lemeshow, S.(2000): Applied logistic regression. Chichster: John Wiley and Sons, 373. p.

Howard, J.W., Dawes, R.M. (1976): Linear prediction of marital happiness. In: Personality and Social Psychology Bulletin, 2. 478-480. p.

Hunter, J.E., Hunter R.F. (1984): Validity and Utility of Alternate Predictors of Job Performance, In: Psychological Bulletin, 96. 1. 72-98. p.

Kahneman, D. (2013): Thinking, fast an slow (in Hung.). Budapest: HVG Kiadó 257-270 p.

Leli, D.A., Filskov S.V. (1984): Clinical Detection of Intellectual Deterioration Associated With Brain Damage. In: Journal of Clinical Psychology, 40. 6. 1435-1441. p.

Libby, R. (1976): Man versus Model of Man: Some Conflicting Evidence. In: Organizational Behavior and Human Performance, 16. 1-12.

Lowry C. (1975): The identification of infants at high risk of early death. MSc dissertation. Dept of Medical Statistics, London School of Hygiene and Tropical Medicine.

Milstein, R.M., Wilkinson, L., Burrow, G.N., Kessen, W. (1981): Admission decisions and performance during Medical School. In: Journal of Medical Education, 56. 2. 77-82. p.

Nagar, Y., Malone T.W. (2011): Combining Human and Machine Intelligence for Making Predictions. In: MIT center for collective intelligence working paper 2. 1-6. p.

Nagar, Y., Malone, T.W. (2010): Combining Human and Machine Intelligence for Making Predictions. Proceedings of the Workshop on Computational Social Science and the Wisdom of Crowds (Neural Information Processing Systems NIPS 2010 Conference), Whistler, Canada, December 10, 2010.

Oskamp S. (1965): Overconfidence in Case Study Judgments. In: Journal of Consulting Psychology, 63. 81-97. p.

Meehl, P.E. (1954): Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence. Minneapolis: University of Minnesota Press,149 p.

Sawyer, J. (1966): Measurement and prediction, clinical and statistical. In: Psychological Bulletin, 66. 3. 178-200. p.

Stillwell, W.G., Barron, F.H., Edwards W. (1983): Evaluating credit applications: A validation of multiattribute utility weight elicitation techniques. In: Organizational Behavior and Human Performance, 32. 1. 87-108. p.

Swets, J.A., Dawes, R.M, Monahan, J. (2000): Psychological science can improve diagnostic decisions. In: Psychological Science in the Public Interest, 1. 1. 1-26. p.

Thornton, B. (1977): Linear prediction of marital Happiness: A replication. In: Personality and Social Psychology Bulletin, 3. 674-676. p.

Wiesner, W.H. Cronshaw, S.F. (1988): A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. In: Journal of Applied Psychology, 61. 275-290. p.

Wittman M.P. (1941): A Scale for Measuring Prognosis in Schizophrenic Patients. In: Elgin State Hospital Papers, 4. 20-33. p.

Corresponding author:

Gábor HÁMORI
Kaposvár University,
Doctoral School of Management and Organizational Sciences
H-7400 Kaposvár, Guba S. u. 40
e-mail: hamorimail@gmail.com