

# SUPPLIER SHORT TERM LOAD FORECASTING USING SUPPORT VECTOR REGRESSION AND EXOGENOUS INPUT

Marin MATIJAŠ<sup>\*</sup> — Milan VUKIĆEVIĆ<sup>\*\*</sup> — Slavko KRAJCAR<sup>\*</sup>

In power systems, task of load forecasting is important for keeping equilibrium between production and consumption. With liberalization of electricity markets, task of load forecasting changed because each market participant has to forecast their own load. Consumption of end-consumers is stochastic in nature. Due to competition, suppliers are not in a position to transfer their costs to end-consumers; therefore it is essential to keep forecasting error as low as possible. Numerous papers are investigating load forecasting from the perspective of the grid or production planning. We research forecasting models from the perspective of a supplier. In this paper, we investigate different combinations of exogenous input on the simulated supplier loads and show that using points of delivery as a feature for Support Vector Regression leads to lower forecasting error, while adding customer number in different datasets does the opposite.

**Key words:** supplier, exogenous input, short term load forecasting, support vector regression, electricity market

## 1 INTRODUCTION

Load forecasting is a fundamental task in power systems and electric utilities operation from the development of first electricity grids. Due to importance of keeping equilibrium between production and consumption, many approaches have been developed throughout its history. A good survey of papers dating as far as 1966 is given in [1]. Researches devoted to load forecasting have a common goal of reducing load forecasting error as low as possible. Decrease of load forecasting error directly decreases risk of blackout and suppliers operational costs. Bunn and Farmer showed the importance of good load forecasting by reporting that increase in load forecasting error of 1% resulted in 10 million Pounds increase in operating costs [2]. Decrease of load forecasting error is also important for power system control, supply security, power system planning and liberalized electricity market operation.

In liberalized electricity markets, participants pay the difference between their forecasted and realized load at the price of balancing energy. Accordingly, producers and suppliers' expenditure is linked to their load forecasting error through balancing energy cost. Due to competition, market participants are not able to transfer their load forecasting costs to end-consumers. This forces them to improve their forecasting in order to remain profitable. Different parties in the electricity market need different types of load forecasts for their business processes. For that purpose, load forecasting is most frequently divided in four types from the perspective of time period:

VSTLF very short term load forecast, up to 24 hours in the future;

STLF short term load forecast, between 24 hours and 86 hours in the future;

MTLF medium term load forecast, between 86 hours and 6 months in the future;

LTLF long term load forecast, 6 months and more in the future.

Although all types of load forecasting are important in liberalized electricity market operation, the most important is Short Term Load Forecasting (STLF). This is because electricity is most frequently scheduled on a daily basis and that time period corresponds to short term forecasting. Proportion of STLF and its dominance over other types of load forecasting can be found in surveys on load forecasting in [3–7].

Depending on the algorithms that dominate in a load forecasting model, these can be separated in the following groups: autoregressive models, moving average, exponential smoothing models and machine learning based models.

STLF is different for production and consumption because it depends on different parameters. In production, parameters affecting load forecast are chosen depending on the type of power plant, while in consumption, parameters are chosen depending on the type of consumption and vary accordingly [8]. If we consider market participant that plans consumption of the whole grid, it can make forecasts and plans according to the measurements made on the highest voltage level of the grid. These are often real-time measurements on a high voltage grid that can be obtained and aggregated in less than one hour. If we consider market participants in a liberalized market, they do not plan the whole grid load which has few important effects for both forecasting of production and consumption. Considering consumption forecasting, difference with the introduction of liberalized markets is in the supplier's perspective because:

---

<sup>\*</sup> Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, marin.matijas@fer.hr, <sup>\*\*</sup> Faculty of Organizational Sciences, University of Belgrade, Belgrade, Serbia

- Metering of end-consumers, especially households is not done with same time granularity or accuracy as metering on upper voltage levels. Substantial number of customers are metered monthly, quarterly, annually or biannually depending on the market;
- Meteorological data is hard to link with consumption, especially in situation in which GIS system or database which links points of delivery with geographical locations is not in place;
- Customers can leave suppliers through a supplier switch;
- New end-customers join suppliers with no historical consumption or without data available in the forecasting system.

Due to the mentioned reasons, load forecasting error is generally higher when forecasting load of a supplier. Exercising of demand side management on a larger scale and scheduling optimization need more accurate forecasts. Introduction of efficient demand side management is a step towards more efficient electricity grid. Scheduling optimization is needed in order to manage scheduling of customers on lower level as a step towards more efficient electricity markets.

Our approach is assessing supplier load forecasting in the electricity market. Besides previously mentioned exogenous data related to STLF, suppliers are exposed to different market events such as customer migration. In order to include effects on load like supplier switch, we use exogenous data related to customer migration, namely number of points of delivery (PoDs), to reduce load forecasting error of a supplier.

### 1.1 Related work

Exogenous input has been opted in or out, throughout history of load forecasting with different conclusions. Hor *et al* [9] used gross domestic product and population growth as input for regression based monthly forecast for England and Wales along with various metrological variables. Comparison of three models implementing social and economic variables showed that variation of metrological data implementing cooling degree days and heating degree days performed the best, resulting with mean average percentage error (MAPE) of 1.98%. Mohamed and Bodger [10] forecasted LTLF using electricity price, gross domestic product and population as exogenous output for multiple linear regression. Since dependency of considered exogenous inputs and electricity load is non-linear, same dataset can give better results with another type of regression that considers non-linearity. In their implementation of VSTLF and STLF for a large electricity distribution company, Souza *et al*[11] used standard Holt-Winters model. On top of it, they implemented complex holidays rule based system in order to substantially minimize forecasting error. Some approaches showed that holiday or even temperature used as an exogenous input can increase the forecasting error in estimating future load. Interesting example is winning entry of EUNITE competition by Chang *et al*[12] which observed

that high uncertainty of temperature for the forecasted period can lead to increase of error. The same approach excluded holiday as an input variable as it led to worse forecasting performance. In recent years approaches using spot price as an exogenous input for load forecasting have been proposed. In a broad range of papers different approaches and variables such as weather, social and economic exogenous input have been used, including customer related inputs [13]. Karsaz *et al* [14] implemented estimation of electricity load and price together via recursive co-evolutionary approach. By recursively estimating Neural Network (NN) weights for electricity load and price forecasting, they got 5–20% improvement over other approaches in three experiments they did.

Models that use machine learning techniques have been developed with the introduction of neural networks in 1990s. In SCOPUS database for 2010 year, 70 out of 118 papers devoted to STLF were still utilizing NNs [15] on 29<sup>th</sup> January, 2011. Since algorithm implementing Support Vector Machines (SVM) won EUNITE load forecasting competition in 2001 [12], more and more models have been implementing algorithms based on SVM and their number is 18 published papers in 2010 [15]. A recent survey on Time Series prediction using Support Vector Machines [7] showed that most of the algorithms in a survey were implemented in financial market prediction and electric utility forecasting. Pai and Hong [16] proposed Recurrent Support Vector Machine that used Genetic Algorithm for optimization of SVM. Li *et al* [17] introduced combination of “similar day method” with Support Vector Regression (SVR) that gives lower error as “similar day method” positively influences the estimation of output. Niu *et al* [18] proposed implementation of chaotic time series and SVM for STLF that can effectively increase accuracy as it gives general models of non-linearity.

## 2 METHODOLOGY

Based on implementations of support vector machines for solution of load forecasting problem, we are selecting one instance of the algorithm for the problem of supplier short term load forecasting.

Vapnik introduced linear classifiers and optimal separating hyperplanes in 1960s. After introduction of SVM for classification in 1995, it was recognized as a good solution for different tasks. Importance of the algorithm increased with SVR, a version of SVM for regression. SVR was proposed in 1996 by Vapnik, Drucker, Burges, Kaufman and Smola [19]. It was soon successfully applied to variety of time series problems one of which was load forecasting. Its biggest advantage over other solutions, (*eg* NNs) is that it is based on structural risk minimization principle. Instead of finding empirical errors, SVR aims to minimize upper bound of the generalization error, therefore lowering chance of over fitting. Here we give an overview of SVR algorithm proposed in [20] and used in this paper.

For a given set of examples  $(x_i, y_i)$  and a capacity constant  $C$ , the following optimization problem has to be solved:

Minimize

$$\phi(\omega, \xi, \xi^*) = \frac{1}{2}(\omega^\top \omega) + C\left(\sum_{i=1}^l \xi + \sum_{i=1}^l \xi^*\right) \quad (1)$$

with respect to

$$(\omega^\top x_i) \geq y_i - \varepsilon - \xi_i^*, \quad i = 1, \dots, n, \quad (2)$$

$$(\omega^\top x_i) \leq y_i + \varepsilon + \xi_i^*, \quad i = 1, \dots, n, \quad (3)$$

$$\xi_i^* \geq 0, \quad i = 1, 2, \dots, n, \quad (4)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, n. \quad (5)$$

Because of computational complexity, instead of minimization, a dual formulation of the problem can be solved:

Maximize

$$W(\alpha, \alpha^*) = -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \langle x_i x_j \rangle + \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) - \sum_{i=1}^l \varepsilon (\alpha_i^* + \alpha_i) \quad (6)$$

with respect to

$$0 \leq \alpha_i, \alpha_i^* \leq C, \quad (7)$$

$$\sum_{i=1}^l \alpha_i^* = \sum_{i=1}^l \alpha_i. \quad (8)$$

Lagrange multipliers  $\alpha_i$  and  $\alpha_i^*$  act as forces pushing the predictions towards target value. Instead of being solved directly, dual problem is iteratively decomposed into a small working set  $S$  and minimized on the working set only, keeping the other variables fixed. A working set is computed on the basis of feasible directions. For each feasible direction  $i$  if either  $\alpha_i$  or  $\alpha_i^*$  is non-zero, other Lagrange multiplier is zero and variable complies with Karush-Kuhn-Tucker's conditions. Non-zero coefficient will be equal to or larger than  $\varepsilon$ . Because errors below  $\varepsilon$  are tolerated, training data inside the so-called “ $\varepsilon$ -tube”, do not contribute to the problem solution. First  $m$  training data outside of “ $\varepsilon$ -tube” are support vectors. Generally, the larger the  $\varepsilon$  value, the fewer the number of support vectors, and thus the sparser the representation of the solution. Besides these, important feature of SVR is called “kernel trick”. For nonlinear cases, the dot product  $\langle x_i, x_j \rangle$  in (6) becomes a kernel function  $\langle \phi(x_i), \phi(x_j) \rangle = K(x_i, x_j)$ . The idea is introduced by function  $\phi: R_d \rightarrow F$ , which maps the input space to the high dimensional feature space. If any function can satisfy Mercer's condition, it can be used as a kernel function.

Which kernel is most suitable depends on the inner product in the data that the kernel will be used with. The most used kernels are: linear, radial basis function (RBF), polynomial and Fourier kernels.

### 3 EXPERIMENT

Experiment consists of support vector regression based short term load forecasting on one control and two different datasets in which input data has been varied in order to get the combination that leads to lowest forecasting error.

#### 3.1 Dataset preparation

The experiment is conducted on three datasets representing three different types of suppliers. Dataset A represents a consumption of wider grid area with a mix of industrial and household end-consumers. Dataset A is a control group as it does not include data regarding customers and PoDs of a particular supplier, but has aggregated consumption of one transmission system. Data of a real electricity system for the period between January 1, 2009 and December 31, 2009 is analyzed.

Dataset B is created by simulating supplier behaviour. Actual consumption of customers from a large database is pre-processed with linear transformation

$$E = \alpha * e + a \quad (9)$$

where  $\alpha$  is a real number from an interval  $[0.7, 1.3]$ ,  $e$  is the original consumption and  $a$  number between  $\frac{-0.1}{n} \sum_{k=1}^n e_k$  and  $\frac{0.1}{n} \sum_{k=1}^n e_k$ , where  $e_k$  is energy in the  $k$ -th interval for a given customer and  $n$  is the total number of time intervals. Constrained linear transformation has been used in order to preserve similar load patterns at similar consumption levels. This pre-processing is made before normalization and other changes in order to ensure that no private data is used in the analysis.

Switching suppliers is simulated randomly on dataset created in described fashion with a condition that, at each moment, at least 80% of PoDs supplied by the supplier have to be the same throughout the whole year. This condition is added as in this stage of electricity markets development, only a small fraction of customers change their suppliers. Another condition for switching suppliers is that a switched PoD cannot switch back. This is in order to include cases in which supplier switch has been started by a supplier and customer cannot start a new contract with a supplier. This is in accordance to present market practice in retail electricity markets where long periods of supplier switch or penalties imposed by suppliers force end- customers to switch once a year or less. Average load of Dataset A is three times bigger than average load of Dataset B. Dataset B is a simulation of a medium to large sized supplier of commercial and industrial end-consumers. They used the same period as for Dataset A. Dataset C is created as a simulation of supplier behaviour in a same manner as Dataset B. It simulates behaviour of a small supplier for commercial

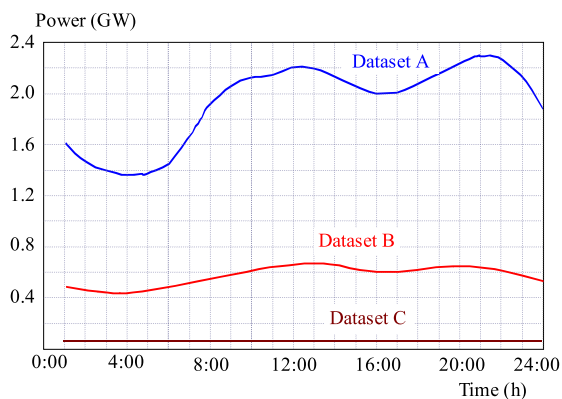


Fig. 1. Absolute relation between average hourly load for Datasets A, B and C

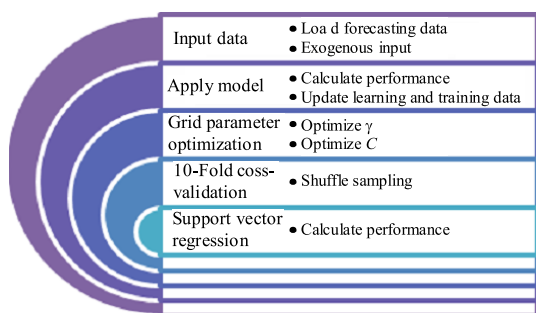


Fig. 2. Five layers of forecasting model from the top layer in which data is fed, to the last layer where core algorithm is calculating the predictions

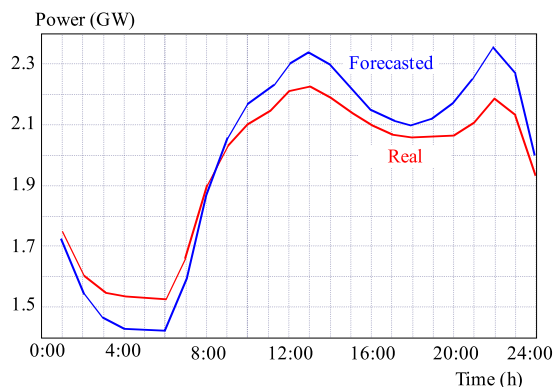


Fig. 3. Average difference between real and forecasted values without holidays for year 2010

and industrial customers. Period tested is the same as for other two datasets, and average load of Dataset A is 31 times bigger than average load of Dataset C. Volatility in Dataset C is higher than in other two datasets due to a small number of customers, and additionally because supplier switch is taken into account. In Fig. 1 we can see the difference of average daily load between the three datasets.

Load of all three datasets is of hourly granularity. Along with label that is presented by hourly consumptions, datasets are extended with empirically found set

of features presented by 36 dummy variables 4 of which correspond to a season, 7 correspond to days in a week, 24 correspond to hours in a day and one additional feature marks holidays. Exogenous data are represented by number of customers, number of PoDs and square root of their product.

MAPE is used as a measure of forecasting error because it is its most widely used measure [6, 7]. MAPE can be calculated as

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{f_i - a_i}{a_i} \right|. \quad (10)$$

Where  $a_i$  stands for actual consumption in the given period and  $f_i$  stands for forecasted value.

### 3.2 Experiment steps

In the pre-processing, first all non-dummy variables in datasets are normalized on interval  $[0, 1]$  in order to minimize higher weight of some features on selection of support vectors.

Forecasting is done as a simulation of forecasting in a real-world. Because forecasting is done for a day in advance, learning set has been created up to the data of two days before the forecasted day. This is called D-1 forecasting. That means, if we on Tuesday forecast consumption of Wednesday, the last used day for learning is Monday (two days prior). This delay can be longer depending on time periods in which meters are read. Load forecasting error increases with the increase of meter reading periods. Therefore, our forecasting error is higher than in the case when forecasting is done with the most recent meter readings available for the forecasted period.

Layer diagram of the forecasting model with five layers can be seen in Fig. 2. The first layer consists of algorithm for SVR. In order to select optimum parameters performance is measured and logged along with SVR. We log SVR parameters, performance and loop time for later analysis.

In order to decrease bias, learning has been done with 10-fold stratified cross validation for each selection of parameters (second layer). For each learning set, an algorithm has been optimized through Grid Parameter Optimization as described in [21] (third layer). After optimum parameters have been found, SVR model is applied to a testing set and load forecast has been made. Afterwards, learning test is extended for 24 hours, testing set is moved forward for 24 hours and whole process is repeated (fourth layer). Fifth layer consists of data input which, besides load features, take exogenous input for estimation of future load.

## 4 DISCUSSION

Experiments have been conducted using open source program RapidMiner 5.0 on a personal computer with a

**Table 1.** MAPE of experiments with different combination of exogenous input. Best results is shown in bold letters

Type/Size	Dataset B (%)	Dataset C (%)
None	9.12	8.07
Cust	10.97	10.16
PoD	<b>6.19</b>	<b>7.06</b>
Custpod	12.11	7.07
Sqrt (cust, pod)	19.63	16.04

2.83 GHz Core(TM) 2 Duo E8300 CPU with 2 GB memory under Microsoft Windows XP Professional with Service Pack 3 with an average runtime of 21:37 minutes per forecasted day. Runtime highly depends on how far the day in the forecasted timeline is from the moment of regression (*eg* it takes more historical data for forecasting farther days). Taking more historical data increases the number of support vectors thus affecting performance. Increase of cost parameter  $C$  also increases runtime, as it increases the margin which, in turn, returns more support vectors. More support vectors result in higher dimensional spaces used for calculation.

Experiments were conducted on the three datasets both with 7 combinations of exogenous input and without them. Comparison of average realized and forecasted values for a typical dataset can be seen in Fig. 3.

We have obtained similar result regarding holidays as in [12]. Using holiday as a parameter increased forecasting error from 5.47% to 6.23% for Dataset A. Our findings suggest that most easily obtainable exogenous input, the number of customers, is not suitable. It increases forecasting error either used solely or in combination with other exogenous inputs as can be seen in Table 1.

Type of exogenous input in Table 1 is as follows:

None – no exogenous input;

Cust – number of customers;

PoD – number of PoDs;

Custpod – both number of customers and number of PoDs;

Sqrt (cust, pod) – square root of product between number of customers and number of PoDs.

Number of customers performed worse on the dataset B. We expect this to be due to discrete change of number of customers in the case of dataset B. Dataset C which had more frequent changes of customers with non-discrete intervals returned better performance results using number of customers as a parameter. Best performance was obtained in spring time and worst performance on holidays, *eg* for Dataset B with PoD, day with lowest MAPE is 28<sup>th</sup> April, 2009 with 0.91% and day with highest MAPE is 25<sup>th</sup> December, 2009 with MAPE of 39.03%. More complex combinations of exogenous input did not increase performance compared to using only PoD.

## 5 CONCLUSION

In this paper, we research forecasting models from the perspective of a supplier. We simulate supplier load in

two different scenarios and show how different exogenous input affect supplier load forecasting using support vector regression.

We show that number of PoDs increase performance, compared to forecasting without it as a parameter. Therefore we propose it as a suitable input parameter for SVR forecasting of supplier load. Furthermore, it is identified that the number of customers has negative impact on forecasting performance in different combinations of exogenous inputs.

Since forecasting models highly depend on geographical location, market role and forecasting horizon, our future work will be focused on different load forecasting techniques using these inputs as parameters for achieving more accurate load forecasting models.

## REFERENCES

- [1] GROSS, G.—GALIANA, F. D.: Short-Term Load Forecasting, Proceedings of the IEEE **75** (1987).
- [2] BUNN, D. W.—FARMER, E. D. E.: Comparative Models for Electrical Load Forecasting, International Journal of Forecasting **2** (1986), 241–242.
- [3] HIPPERT, H. S.—PEDREIRA, C. E.—SOUZA, R. C.: Neural Networks for Short-Term Load Forecasting: a Review and Evaluation, IEEE Transactions on Power Systems **16** (2001), 44–55.
- [4] TZAFESTAS, S.—TZAFESTAS, E.: Computational Intelligence Techniques for Short-Term Electric Load Forecasting, Journal of Intelligent and Robotic Systems **31** (2001), 7–68.
- [5] KYRIAKIDES, E.—POLYCARPOU, M.: Short Term Electric Load Forecasting: a Tutorial, Computer Engineering **418** (2007), 391–418.
- [6] HAHN, H.—MEYER-NIEBERG, S.—PICKL, S.: Electric Load Forecasting Methods: Tools for Decision Making, European Journal of Operational Research **199** (2009), 902–907.
- [7] SANKAR, R.—SAPANKEVYCH, N. I.: Time Series Prediction Using Support Vector Machines: A Survey, IEEE Computational Intelligence Magazine **4** (2009), 24–38.
- [8] SOLIMAN, S. A.—AL-KANDARI, A. M.: Electrical Load Forecasting: Modeling and Model Construction, Elsevier Inc., 2010.
- [9] HOR, C.-L.—WATSON, J. S.—MAJITHIA, S.: Analyzing the Impact of Weather Variables on Monthly Electricity Demand, IEEE Transactions on Power Systems **22**(10) (Nov 2005), 987–2085.
- [10] MOHAMED, Z.—BODGER, P.: Forecasting Electricity Consumption in New Zealand using Economic and Demographic Variables, Energy **30** (2005), 1833–1843.
- [11] SOUZA, R. C.—BARROS, M.—MIRANDA, C. V. C.: Short Term Load Forecasting using Double Seasonal Exponential Smoothing and Interventions to Account for Holidays and Temperature Effects, Acapulco, Mexico, 2007.
- [12] CHANG, M.-W.—CHEN, B.-J.—LIN, C.-J.: EUNITE Network Competition: Electricity Load Forecasting, 2001.
- [13] FEINBERG, E. A.—GENETHLIOU, D.: Load Forecasting, Applied Mathematics for Restructured Electric Power Systems: Optimization, Control and Computational Intelligence, Power Electronics and Power Systems (2005), 269–285, Springer.
- [14] KARSAZ, A.—MASHHADI, H. R.—MIRSALEHI, M. M.: Electrical Power and Energy Systems Market clearing Price and Load Forecasting using Cooperative Co-Evolutionary Approach, International Journal of Electrical Power and Energy Systems **32** (2010), 408–415.

- [15] SCOPUS, <http://www.scopus.com/>, Online Accessed: 29th January 2011.
- [16] PAI, P. F.—HONG, W. C.: Forecasting Regional Electricity Load based on Recurrent Support Vector Machines with Genetic Algorithms, *Electric Power Systems Research* **74** (2005), 417–425.
- [17] LI, X.—SUN, C.—GONG, D.: Application of Support Vector Machine and Similar Day Method for Load Forecasting, *Advances in Natural Computation* **3611** (2005), 602–609.
- [18] NIU, D.—WANG, Y.—DUAN, C.—XING, M.: A New Short-Term Power Load Forecasting Model Based on Chaotic Time Series and SVM, *Journal of Universal Computer Science* **15** (2009), 2726–2745.
- [19] DRUCKER, H.—BURGES, C. J. C.—KAUFMAN, L.—SMOLA, A. J.—VAPNIK, V.: Support Vector Regression Machines, *Advances in Neural Information Processing Systems* **9** (1997), 155–161.
- [20] SMOLA, A. J.—SCHÖLKOPF, B.: A Tutorial on Support Vector Regression, *Learning*, 1998.
- [21] HSU, C.-W.—CHANG, C.-C.—LIN, C.-J.: A Practical Guide to Support Vector Classification, *Bioinformatics* **1** (2010), 1–16.

Received 30 January 2011

**Marin MATIJAŠ** was born in Split, Croatia on December 18, 1984. He received his M.Eng EE degree from the Faculty of Electrical Engineering and Computing, University of

Zagreb, Croatia in 2008. Since 2009 he has been a PhD candidate at the same Faculty. From 2008 he is with HEP Supply Llc, an electricity supplier in Croatia. His areas of interest include time series forecasting, risk management, data mining, strategic planning and energy data management.

**Milan VUKIĆEVIĆ** was born in Belgrade, Serbia on June 12, 1983. He received his M.Sc. degree from the Faculty of Organizational Sciences, University of Belgrade, Serbia in 2008. Since 2009 he has been a PhD candidate at the same Faculty. From 2008 he has been teaching and research assistant at the Faculty of Organizational Sciences, University of Belgrade, Serbia within the Centre for Business Decision Making. He is currently a PhD student. His main research interests include clustering and classification algorithm design, data mining, decision support systems and meta-learning.

**Slavko KRAJCAR** graduated in 1973 from the Faculty of Electrical Engineering and Computing, University of Zagreb. He received his MSc degree in 1980 and a PhD degree in 1988 in the field of planning the distribution networks at the same university. He has been working in the Faculty of Electrical Engineering and Computing since 1973, first as a teaching assistant and afterwards as a professor. He wrote more than 100 R&D papers, the majority of which are implemented in practice. He was the head of different national and international projects. He is still leading a lot of different S&T projects at Faculty.



**EXPORT - IMPORT**  
of periodicals and of non-periodically  
*printed matters, books and CD-ROMs*

Krupinská 4 PO BOX 152, 852 99 Bratislava 5, Slovakia  
tel: ++421 2 638 39 472-3, fax: ++421 2 63 839 485  
[info@slovart-gtg.sk](mailto:info@slovart-gtg.sk) <http://www.slovart-gtg.sk>

