Demand Prediction with Multi-Stage Neural Processing

Maciej Grzenda and Bohdan Macukow

Warsaw University of Technology, Faculty of Mathematics and Information Science,
Pl. Politechniki 1, 00-661 Warszawa, POLAND
grzendam@mini.pw.edu.pl,macukow@mini.pw.edu.pl
WWW home page: http://www.mini.pw.edu.pl/~ grzendam
http://www.mini.pw.edu.pl/~ macukow

Abstract. In many technical issues, the processes of interest could be precisely modelled if only all the relevant information were available. On the other hand, detailed modelling is frequently not feasible due to the cost of acquiring appropriate data. The paper discusses the way self-organising maps and multilayer perceptrons can be used to develop two-stage algorithm for autonomous construction of prediction models. The problem used as a case study is the problem of heat demand prediction in a district heating company. Additionally, because of non-standard evaluation of prediction models, evolutionary construction of multilayer perceptrons has been applied.

1 Introduction

The results of modelling strongly depend on the quality and accuracy of data used to devise the models. At the same time, only limited number of data patterns can be acquired due to technical and financial limitations.

The paper proposes the way existing data can be used to devise prediction models. Heat consumption by the customers of a district heating company is used as a case study of the process that could be precisely analysed for each consumer separately. However, this approach would require a separate project to be performed for each out of hundreds or thousands of consumers separately. Obviously, it remains unfeasible to carry out detailed engineering calculations for each of them. Among the reasons lack of detailed data and significant cost of acquiring it is not of least importance. Before actual solutions proposed to tackle this issue are described a brief overview of district heating is presented.

Modern district heating [13] companies transmitting heat through the pipeline system from the heat sources to heat consumers face new challenges. The development towards more competitive energy markets means that many consumers may decide to resign from district heating services based on central heat sources and replace them with individual boiler systems. Therefore, utility companies have to improve service quality and minimise its price [1]. At the same time, the most significant part of district heating cost is usually the cost of producing the heat. To reach these goals means to optimise heat supply. However, the latter objective can not be fulfilled without detailed understanding of consumers' demand and the way it changes over the time. Moreover, because of limited speed of water flow, on-site monitoring of heat consumers is not enough to control heat production appropriately. Thus, there is a growing need for demand prediction.

When district heating systems were developed, demand for heat significantly depended on ambient temperature. However, improved thermal properties of modern buildings result in increased share of hot tap water in overall heat consumption. This part of demand is not directly related to ambient temperature. Still, many buildings constructed using traditional technologies continue to consume majority of heat for space heating. Therefore, due to inherently different nature of processes involved, results of works regarding load prediction in electrical power systems can not be directly applied for district heating systems.

Although detailed calculations are possible to carry out for each consumer, it would take too much resources to acquire all necessary properties of each building to perform them. At the same time, it remains unclear whether all the consumers, in spite of diverse thermal properties can be approximated by a single model. The remainder of this paper aims to propose an alternative approach that takes into account different profiles of consumers, but resigns from performing exhaustive engineering calculations.

The work investigates yearly heat demand profiles and continues previous research on power demand prediction [4, 5]. The primary objective is to develop a set of prediction models capable of forecasting heat demand in different groups of consumers. To achieve this goal, groups of consumers with similar demand profile are identified first, so as to build prediction models for these groups next.

The first part of the paper presents the way self-organising maps [7, 8] can provide valuable insight into typical heat demand profiles. The information available in sales database system has been retrieved and transformed into core data patterns. Thus, time series of sales volume for each consumer has been obtained. Self-organising maps have been applied to build a network of neurons representing typical sales profiles and enable visualisation of consumers needs.

The second part of the paper discusses the use of multilayer perceptrons to predict detailed heat demand. These models are built separately for different groups of consumers identified by self-organising map. In order to support non-standard error measures and address automatic architecture selection, evolutionary construction of MLP has been applied. Thus, the way heterogeneous solution built of self-organising maps, multilayer perceptrons and evolutionary programming can be applied has been proposed.

The remainder of this paper is organised as follows:

- section 2 describes the data set and the problem itself,
- neural networks-based approach is outlined in section 3, it contains two subsections describing the phases of the algorithm proposed, namely SOM-based identification of consumer groups, and construction of MLP-based prediction models for each of the groups,
- section 4 summarises the whole work.

2 Data set and problem description

The problem analysis is based on the data set from one of the Polish district heating utilities. Like other utilities of this type, it provides heat consumers with heat required for space heating and hot tap water purposes. The challenges faced by this company include proper reaction to substantial changes in heat demand. Significant period of time is required to distribute heat from heat source to distant consumers, in some cases it may reach 3 hours. During this time, demand for hot tap water may largely change, as it reaches its peak in early morning and late evening. Therefore to ensure adequate heat supply means to predict demand for heat. However, this demand can be expected to change for different consumers' not equally. For instance, the heat demand of a bank may depend on ambient temperature only, while the needs of a hotel are bound to be significantly increased at peak hours every day.

To analyse the problem the monthly sales information from the billing system for the period of last four years has been used. Unlike other data sets describing consumers, the billing data provides complete information for virtually every consumer $c_i \in C, C = \{c_i : i = 1, ..., N\}$.

For each consumer average monthly heat consumption has been calculated first. After detailed analysis $N = 1109$ sales profiles have been obtained. Every sales profile of a consumer $c_i, i = 1, ..., N$ is represented by a vector $S_i = (s_{i,1}, ..., s_{i,12})$, where $s_{i,m}$ denotes the average heat sale to consumer i during month m. Months are indexed in a traditional way i.e. $i = 1$ stands for January.

Furthermore, each sales profile has been normalised so as to obtain average demand profile $D_i = (d_{i,1}, d_{i,2}, ..., d_{i,12})$ out of original consumer profile S_i as follows:

$$d_{i,k} = \frac{s_{i,k}}{\sum_{m=1}^{m=12} s_{i,m}}, k = 1, ..., 12 \tag{1}$$

Thus, $N = 1109$ normalised demand profiles have been obtained. These correspond to the average yearly consumption of heat of each consumer. Some of demand profiles are depicted on fig. 1.

While, in general, demand profiles reflect space heating needs resulting from average ambient temperatures, one can observe that minimal demand is reached by the consumers in the period between May and August. Similarly, peak heat consumption is attained between December and March. In other words, the exact time location of minimal and maximal heat consumption significantly varies among consumers. What is of outstanding importance even normalised monthly heat sales can differ as much as 300% among sample consumers randomly selected from the whole data set.

3 Neural networks and demand profile identification

While a separate demand profile D_i for each consumer could be devised, the problem remains to identify common profile groupings. The latter are necessary for further market analysis, for instance to evaluate the impact of new consumers

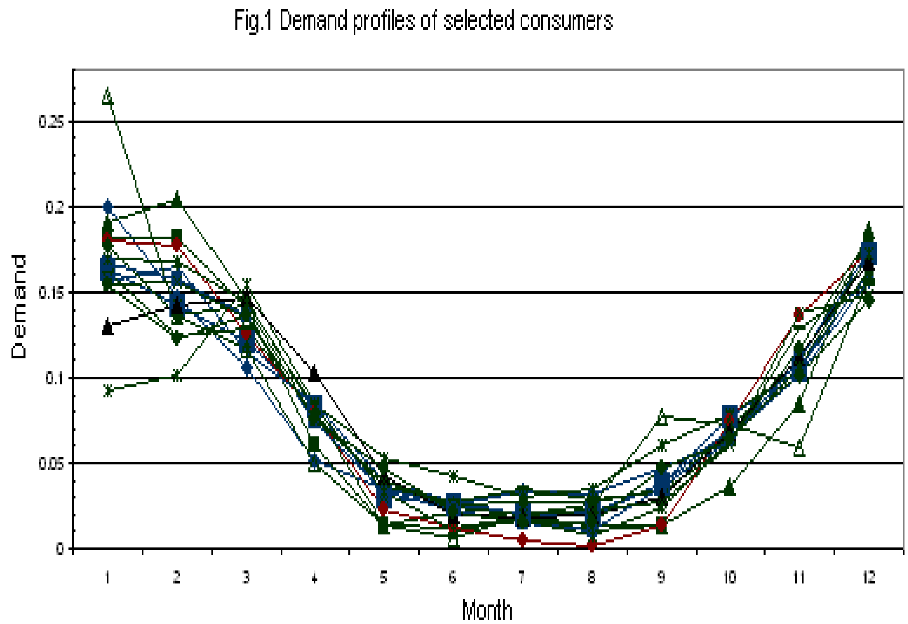

Fig. 1. Sample demand profiles D_i of different consumers

on the district heating system in view of prospective heat demand. Furthermore, there is a need to predict heat demand of a consumer to optimise heat supply. Separate models for each consumer are not feasible, however. Still, such prediction models can be devised for groups of consumers with similar behaviour regarding heat sales profile. Therefore, the question is whether the observed variety of demand profiles can be explained by the existence of a number of typical demand profiles among consumers. In order to answer this question groupings of demand profiles are located first. Prediction models are constructed using multilayer perceptrons then.

3.1 Demand profiles and SOM-based approach

Algorithm overview Different approaches could be proposed to determine possible groupings of demand profiles. Among them numerous clustering algorithms can be listed. Both agglomerative and partitioning algorithms like complete linkage or k-means algorithm could be applied [10]. However, significant drawbacks of this approach would be the lack of straightforward visualisation for multidimensional data groupings and limited potential for the identification of natural modality of the set. Another approach proposes the use of cooperative agents in each consumer to tackle the optimisation of heat supply [1]. Unfortunately, the latter method requires each consumer to be supplied with heat exchange substation capable of running such software agents. This would inevitably result in significant expenditures of a company.

Therefore, in order to answer the problem of demand profiles identification, self-organising neural networks have been applied. The purpose of using self-organised neural networks, namely self-organised maps (SOMs) [8], is to provide measure of locating demand pattern groupings i.e. typical profiles $\hat{D}_k, k << N$ that would explain diversity in the population of $D_i, i = 1, .., N$ demand profiles. At the same time, spatial distribution of typical profiles and their relation to existing marketing consumer categories are to be considered.

The following solutions have been applied:

- two-dimensional square lattice of J neurons,
- demand profiles $D_i, i = 1, .., N$ were used as input patterns.

Therefore, each neuron j located in the lattice is represented by a weight vector

$$w_j = [w_{j,1}, w_{j,2}, .., w_{j,12}]^T, j = 1, 2, ..., J. \tag{2}$$

During the learning process neurons are tuned to the input patterns. Moreover, the weight update algorithm tunes the weights of the winning neurons and some of its spatially close neighbour neurons. Thus, topographical map of input space is being created [7]. In other words, as a result of this cooperative process, both input pattern groupings and their relation to each other can be identified.

Results A number of computation series have been performed, using different weight update algorithms. Not only standard Winner-Takes-All (WTA) [7, 8], but also Winner Takes All with Conscience (CWTA) [7] and Neural Gas (NGAS) [11] algorithms have been used to tune neuron weights. The two latter algorithms aim to overcome deficiencies of standard WTA algorithm. Among these deficiencies of standard WTA weight update algorithm, overrepresentation of regions with low input density is not of least importance [7].

To illustrate the results of computations, the computations performed using NGAS method have been selected. In all cases, for each neuron $j, j = 1, ..., J$ in the lattice, at the end of tuning phase, the winning count Y_j has been calculated in the following manner:

$$Y_j = \sum_{i=1}^{N} eval(i, j) \tag{3}$$

while

$$eval(i, j) = \begin{cases} 1 & \text{if} \parallel D_i - w_j \parallel = min_{k=1,...,J} \parallel D_i - w_k \parallel \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Thus $Y_j : j = 1, ..., J$ denotes the number of input patterns that are the most similar to the weight vector w_j in terms of Euclidean distance. Fig. 1 depicts weight vectors w_j such that $Y_j > 20$, after computations performed for a lattice of $J = 100$ neurons, using NGAS method, $N = 616$. Each profile corresponds to a single neuron. Its (x, y) location on the lattice is provided in the figure legend.

Obtained results show that a number of distinctive demand profiles providing centroids of input consumer's profiles have been identified. This suggests that observed diversity in the input patterns can be explained by underlying typical demand profiles. In addition, some 40% of neurones remain virtually inactive i.e. $Y_j \leq 2$, in spite of using NGAS algorithm that promotes diversity among neurones. In other words, when dealing with heat demand prediction, it is enough to concentrate on a limited number of consumers demand patterns.

The profile groupings identified in the form of weight vectors reflect the diversity of consumers' demand profiles. In particular, one may notice that peak consumption is reached in January, February or December, depending on resulting $\hat{D}_k$ profile considered.

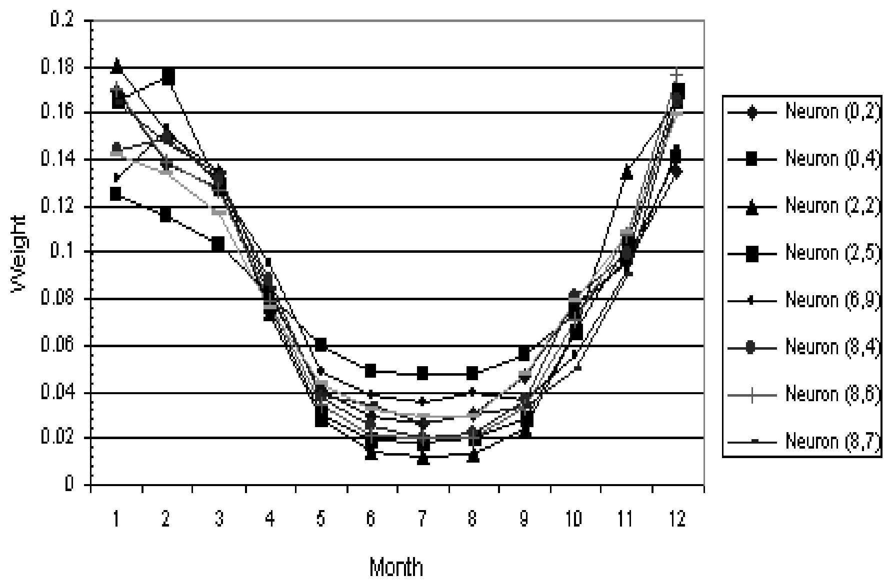

Fig. 2. Demand profiles represented by weight vectors of the winning neurons.

Additionally, one of previous works [6] investigates the relation between marketing consumer category and its location on the lattice. The categories discussed were FH standing for family houses, C standing for commercial buildings, S standing for schools, H standing for hotels and HS for hospitals. It is worth emphasising that detailed analysis of contextual map [7] built on the SOM network has shown limited impact of consumer's category on its demand profile. In general, some consumer categories are represented by a consistent, relatively limited part of contextual map. In other words, consumers of these categories share a set of similar demand profiles. These include hotels. On the other hand, family houses demands can be related to the largest number of neurons. In fact, no more than 10% of FH consumers are represented by a single neuron on the SOM lattice. As a consequence, it remains impossible to create a single prediction model for each marketing consumer category, as the diversity of demand profiles may be too large for the consumers of some categories.

To sum up, previous investigation of the relation between consumer's category and demand profiles of consumers of the category shows that similarity in demand profiles is to some extent related to consumer's category. On the other hand, diverse typical demand profiles have been identified by the SOM approach, as depicted on fig.2. Therefore, separate prediction models should be constructed for different groups of consumers as defined by the proximity to weight vectors of SOM neurones [6]. This approach allows to address by a single prediction model a group of consumers displaying similar demand profile. This similarity of a demand profile, characteristic for all the consumers related to a neuron, may result from real consumer's category, but also thermal properties of buildings and the share of space heating needs in overall demand.

Therefore, the next step is to construct a separate prediction model for each group of consumers C_i. More formally:

$$C_i = \begin{cases} \{c_n : \| D_n - w_i \| = min_{j=1,\ldots,J;j:Y_j \geq 10} \| D_n - w_j \|\} \; when \; Y_i \geq 10 \\ \emptyset \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{5}$$

One can easily see that $C = C_1 \cup C_2 \cup ... \cup C_J$ and $C_i \cap C_j = \emptyset, i \neq j; i, j = 1, ..., J$. For each of the so defined $C_i : i = 1, ..., J \wedge C_i \neq \emptyset$ sets of consumers a separate prediction model based on multilayer perceptron is constructed in the next stage. The role of each prediction model is to predict heat consumption for consumers $c_i \in C_k$ from its group. As the consumers classified to each set C_i share similar demand profile, their overall thermal properties and proportion of heat consumption for hot tap water are similar as well. Therefore, thanks to SOM approach groups of consumers suitable for being represented by a single prediction model have been established.

3.2 Multilayer perceptrons to predict heat demand

Algorithm overview The problem of electrical power demand prediction has been analysed for over 40 years. Different methods based on pattern recognition, knowledge-based systems, statistic methods and artificial neural networks have been proposed and applied during this period [2, 4, 5, 9, 12]. For an overview of the subject see [9]. Some of the best results have been obtained using neural networks [2]. Heat demand prediction may seem to be a similar problem. However, it depends on its own set of factors, including thermal properties of a building and temperature required to ensure human comfort which results from the purposes the building is used for. Thus, unfortunately the prediction models devised for power demand prediction can not be applied for heat demand prediction.

In our work, in order to construct heat prediction models, MLP networks have been applied. For each non-empty group of consumers $C_j, j = 1, ..., J$ a single MLP network is constructed. To train the neural networks, evolutionary construction of multilayer perceptrons $ECoMLP$ has been applied [4, 5].

The purpose of the algorithm is to minimise the prediction error measured on a learning set and obtain networks with generalisation abilities, thus making it possible to use the final networks for the purpose of load prediction. In our case, it will be used to predict heat demand at selected consumers belonging to the same group C_j. Detailed time series available for selected consumers of the C_j group has been used to construct both learning and testing set. This contains heat consumption at each hour of selected months combined with average ambient temperature at this hour. The primary objective it to train multilayer perceptron to predict heat demand at each hour depending on ambient temperature, time of the day and day of week. Additionally selected hours are marked as peek hours in terms of hot tap water demands.

The motivation for using evolutionary method is as follows:

- The problem requires numerous multilayer perceptrons to be constructed, as there is a need for a separate MLP network for each group C_i,
- Because of the number of networks involved, no trial-and-error architecture selection can be accepted. Similarly architecture selection can not rely on human expert intervention
- The objective function is not a standard mean square error. The problem may require non-differentiable and even non-continuous error measure for the reasons listed below.

Before sample results are presented an overview of evolutionary algorithm used to construct the multilayer perceptrons is presented. The algorithm is a result of previous works [4, 5] and is based on the concepts of evolutionary programming (EP)[3] and self-adaptive control of algorithm settings.

Evolutionary construction of multilayer perceptrons $ECoMLP$

- $t := 0$,
- create initial population of MLPs
  $\Phi(0) = \{\phi_i : i = 1, ..., N\}$ satisfying the following assumptions:
  - fixed number of input layer nodes $card(L_0)$ and output layer neurones $card(L_M)$ defined by the problem,
  - random number of hidden neurones,
  - randomly chosen connection weights,
- in a sequence of generations construct consecutive populations $\Phi(t)$ by:
  - applying standard EP selection and promoting the best $\frac{N}{2}$ networks to the next population, duplicating them and producing child subpopulation thereafter; the networks are evaluated using fitness function $F()$,
  - affecting child networks with mutation operator composed of:
    - hidden node mutation (add or delete neuron with p_a and p_d probability),
    - weight mutation in accordance with the learning rule of each individual (see below),

  * mutation of learning rule with probability of p_{lr},
- $t := t + 1$.

As far as the number of hidden neurones is concerned, it is randomly chosen from a predefined range. Then it can be changed by the mutation operator within the same range. The weight mutation has been strictly based on the form of the *learning rule* controlling the behaviour of the mutation operator.

Learning rule and weight mutation

- let us define for each individual in the population $\phi_i \in \Phi(t)$ learning rule (Met_i, Par_i) where:
  - $Met_i \in \{0, 1\}$ stands for mutation type, discrete or real,
  - Par_i is defined by $Par_i = (p_m)$ for discrete type and $Par_i = (p_m, \mu)$ for the real mutation type, respectively,
- mutate the weights of the network in accordance with the learning rule of the individual, using uniform distribution over a specified range of weights, in case of discrete mutation, or Gaussian distribution, otherwise; thus Par is used to store the distribution settings,
- mutate learning rules as well.

Every network can be mutated using a different learning rule. As a consequence, numerous search strategies are applied and verified in each generation. The solution belongs to the class of self-adaptive control methods. For detailed presentation and discussion of the method see [5] or [4].

Results In order to discuss the results of constructing prediction models, one of the consumer groups C_j has been selected. This group contains $card(C_j) = 66$ consumers. The best representative of the group in terms of proximity to the weight vector of a neuron w_j is a popular tourist hostel. Detailed time series are available for three consumers in the group. All the three consumers are hotels of different sizes. Out of all data patterns in detailed time series described above, learning set L and testing set V have been created, containing 1918 and 959 data patterns respectively.

In order to construct prediction model for the group $ECoMLP$ method has been used. As mutation procedures are based on random processes, 50 runs of the algorithm have been executed. The fitness function took into account desired form of predicted heat demand profile. A priori knowledge regarding acceptable error rate from the point of view prospective heat supply optimisation has been applied. Therefore, the following fitness function has been devised to drive evolutionary process:

$$F(\phi_i^g) = \frac{\sum_{j=1}^{card(L)} \theta(\phi_i^g, j)}{card(L)} \tag{6}$$

$\theta(\phi_i^g, j)$ stands for an error rate of prediction of pattern j by neural network ϕ_i^g i.e. multilayer perceptron i in generation g. The error rate takes into account the following information regarding heat demand prediction:

- detailed prediction of peak demand remains virtually impossible due to partly random character of hot tap water demands,
- it is acceptable to some extent to underestimate demand for heat during a single hour assuming the prediction for both preceding and succeeding hour is not underestimated.

Therefore,

$$\xi(x) = \begin{cases} 0.9 \; when \; x > 0.9 \\ x \quad otherwise \end{cases} \tag{7}$$

$$\theta(\phi_i^g, j) = \begin{cases} 0 & when \; (\xi(phi_i^g(l_j)) \geq 0.85 \times \xi(d_j)) \\ & \wedge(\xi(phi_i^g(l_j)) \leq \xi(d_j)) \\ & \wedge(\phi_i^g(l_{j-1}) \geq (d_{j-1})) \wedge (\phi_i^g(l_{j+1}) \geq (d_{j+1})) \\ \mid \phi_i^g(l_j) - d_j \mid & otherwise \end{cases} \tag{8}$$

For clarity purposes $d_0 = d_{card(L)+1} = \phi_i^g(l_0) = \phi_i^g(l_{card(L)+1}) = 0$ for $i = 1, ..., card(\Phi)$, l_j denotes learning pattern j. It is obvious that by using this or similar fitness function $F()$ as defined in formulas 6, 7 and 8, expectations regarding suitable heat demand prediction can be expressed. Moreover, different approaches considering heat demand can be defined and compared in terms of predicted demand profiles. In case of fitness function defined above, an average value of fitness function $F()$ after $G = 1000$ generations for the best individual ϕ_b^G was equal to 0.1122. This value is an average over 50 algorithm runs. Standard deviation of the error rate was equal to 0.001. An average value of fitness function and standard deviation of its value on the testing set were equal to 0.1186 and 0.0019 respectively. Thus, proper generalisation has been achieved by the networks.

Typical architecture of the optimal in terms of fitness function multilayer perceptron is based on the layers composed of 6 input nodes, and on average 2 neurones in each hidden layer. One neuron is used in the output layer to produce predicted heat demand. Results show that both network architecture and weight values have been set by the learning processes. What is important, the whole process does not involve human expert intervention, thus it may be used for training of numerous networks for their groups of consumers C_i. This is due to the self-adaptive procedure responsible for autonomous selection of algorithm settings. Considering the fact this process has to be repeated for all $C_i : C_i \neq \emptyset$ no other solution could be accepted for real-life solutions. One can also notice that thermal properties of buildings and hot tap water consumption may change over the time, thus both stages of the algorithm would have to be repeated every few months.

4 Conclusions

Numerous real-life technical problems could be solved using detailed computation. However, detailed input data is frequently not available, thus other approaches need to be considered. In our work two-stage algorithm for building diverse prediction models is proposed. A real-life problem of heat demand prediction in district heating systems is used as a case study.

In the first part, the paper shows the way SOM neural networks can provide for time-series data analysis. Results of the computations suggest that the diversity of heat demand profiles can be explained by a number of base demand profiles. Such profiles have been obtained in the form of weight vectors of the neurons with the highest winning rate. The results show that a limited number of typical demand profiles provides a valid representation of all consumers.

In the second part of the work, evolutionary algorithm is applied to develop multilayer perceptrons predicting heat demand for their respective groups identified before by the SOM algorithm. Sample computations results show the way the algorithm can be used to adopt a priori human expert knowledge regarding acceptable error rates. This knowledge is used to search for an optimal heat supply volume that would ensure compromise between consumers' demands and cost of heat production.

References

1. Davidsson, P.,Wernstedt, F., Embedded Agents for District Heating Management, Third International Joint Conference on Autonomous Agents and Multi-Agent Systems, 2004, July 19-23, 2004, New York
2. Doveh, E. et al, Experience with FNN Models for Medium Term Power Demand Predictions *IEEE Transactions on Power Systems*, 538-546, 1999
3. Fogel, D.B., An Overview of Evolutionary Programming. *The IMA Volumes in Mathematics and its Applications*, **111**, 89-109, 1999
4. Grzenda, M., Macukow, B., Evolutionary Model for Short Term Load Forecasting. Proc. of MENDEL 2001 conference. Brno, pp. 119-124, 2001
5. Grzenda, M., Macukow, B., Evolutionary Neural Network-Based Optimisation for Short-Term Load Forecasting, Control and Cybernetics, **vol. 31**, 2/2002, pp. 371-382, 2002
6. Grzenda, M., The Use of Neural Networks to Control Data Acquisition Process, Proc. of ICAISC 2006, Zakopane, *to be published*
7. Haykin, S., Neural Networks: a Comprehensive Foundation, Prentice-Hall Inc., 1999
8. Jang, J.-S.R.,Sun, C.-T.,Mizutani E., Neuro-Fuzzy and Soft Computing. A Computational Approach to Learning and Machine Intelligence, Prentice Hall, 1997
9. Loi, L.L., *Intelligent System Applications in Power Engineering. Evolutionary Programming and Neural Networks*, John Wiley and Sons, 1998
10. Lattin, J., Carroll, J., Green, P.: Analyzing Multivariate Data, Thomson Learning, 2003
11. Martinetz M., Berkovich S., Schulten K., "Neural-gas" Network for Vector Quantization and Its Application to Time Series Prediction, IEEE Trans. on Neural Networks, **vol.4**, pp. 558-569, 1993
12. Osowski, S. and Siwek, K., Selforganizing Neural Networks for Short Term Load Forecasting in Power System. *Engineering Applications of Neural Networks (EANN)*, 235-268, 1998
13. Pierce, M. A., Environmental Benefits from District Heating in the Nordic Capitals, 16th Congress of the World Energy Council, Tokyo, 1995