



## EVALUATION OF COMPUTATIONAL INTELLIGENCE TECHNIQUES FOR DAILY PRODUCT SALES FORECASTING

Gediminas Žylius<sup>1)</sup>, Rimvydas Simutis<sup>2)</sup>, Vygandas Vaitkus<sup>3)</sup>

Department of Automation  
Faculty of Electrical and Electronics Engineering,  
Kaunas University of Technology  
Studentų g. 50-154, Kaunas, Lithuania  
E-mail: gediminas.zylius@ktu.edu, rimvydas.simutis@ktu.lt, vygandas.vaitkus@ktu.lt,

**Abstract:** Product sales forecasting is crucial task in inventory control and whole supply chain management. Accuracy of sales forecasting determines product logistics performance. In this paper we present study that aims to answer three questions: what input set is most informative for daily sales time series forecasting; do weather input features improve forecasting performance; what computational intelligence model is most appropriate for daily sales forecasting. In order to answer those questions we selected three computational intelligence models that are used for regression task together with various input sets for daily time series forecasting. Data collected consist of 89 real life product sales time series from various stores with historical period of 15 months. Results show that most useful input set is extracted from time series itself. Secondly, research results show that weather features do not improve forecasting performance. And finally, best forecasting results are achieved using support vector regression model. *Copyright © Research Institute for Intelligent Computer Systems, 2015. All rights reserved.*

**Keywords:** computational intelligence, regression, daily sales forecasting, data-based modeling.

### 1. INTRODUCTION

Accurately forecasting product sales would improve performance of product logistics and whole supply chain management (SCM): if one knows accurate amount of sales of each product for next period, product logistics optimization can be performed. So the performance of cash logistics optimization is decided upon forecasting accuracy. Sales time series are very complex processes (combination of mixed seasonality, non-linear trend and varying uncertainty) that differ between various sales products. In this research we perform daily sales forecasting of products whose demand may be sensitive to weather conditions and are perishable (such as milk, bread etc.) by data-based modeling using computational intelligence (CI) methods and study effects of various input features. Three different advanced CI models were selected that show most promising results for various regression tasks. The main purpose of using several models is to discover more robust input set that is the most appropriate for product sales modeling task despite the model used.

This research article is further divided into following sections: 1) literature review; 2) sales

forecasting models; 3) methodology; 4) experimental data; 5) results; 6) conclusions.

### 2. LITERATURE REVIEW

There were many researches done in the past related to data-based sales forecasting. Forecasting results might differ for different sales time series aggregation periods (because of forecasting model input features of daily, weekly or monthly product sales time series). In this research we are interested in daily forecasting research results.

The research area of daily product sales forecasting is not new. A model predictive control (MPC) combination with daily sales time series forecasting approach was taken by [1] for production – inventory control. It was shown that combination approach significantly improves classical MPC and GA-RBF neural network forecasting model incorporation significantly improved performance compared to linear AR (autoregressive) and Holt-Winters models. The details and daily fresh milk sales forecasting performance improvement of proposed adaptive RBF (radial basis function) model modification (called GA-RBF) when compared to other models is presented in [3]. Main idea is to use

specifically designed GA (genetic algorithm) optimization algorithm for variable selection and RBF network structure optimization when number of clusters parameter for fuzzy means clustering algorithm is optimized.

Authors in [2] combined SOM (self-organizing maps) and RBF neural network into enhanced cluster and forecast model (ECFM) in order to forecast daily sales of fresh foods at convenience stores in Taiwan. SOM clustering was used for RBF hidden layer node initialization. Results show improvement of forecasting precision compared to fuzzy neural network approach.

Hybrid of discrete wavelet transformation with SARIMA (seasonal autoregressive integrated moving average) called SW model was proposed by [4] and compared with pure SARIMA and classical seasonal decomposition with linear extrapolation with seasonal adjustment (CSD + LESA) methods for sales forecasting (including daily fashion sales time series). Authors used discrete wavelet decomposition and applied ARIMA model for components in approximate scale (trend) forecasting and SARIMA model for components in detailed scale (seasonal and stochastic) forecasting. After different component forecasting, inverse discrete wavelet transform was applied in order to get forecasting results. It was shown that for real-life highly volatile time series proposed model was the most appropriate.

Researchers in [5] proposed system that enables to handle promotion effects on sales using fuzzy neural network with weight elimination and then input it to artificial neural network together with time series inputs for daily sales of convenience store. Authors showed that combination of time series data with qualitative factors and proposed weight elimination scheme improves forecasting performance compared to conventional methods.

Automatic Box – Jenkins time series model identification and outlier adjustment methods were proposed by [6] for fast-food restaurant franchise daily time series data. Proposed method increased forecasting accuracy when compared to models without outlier adjustment.

A new algorithm based on support vector regression for daily perishable farm product sales forecasting was proposed by [7]. The authors applied fuzzy theory processing for exogenous variables (weather data, temperature data, week data, etc.) in order to form additional inputs to SVR model. Results showed that proposed model outperforms radial basis function neural network model.

Evolutionary neural network when network topology is optimized using genetic algorithm was used by [8] when forecasting daily apparel sales

time series. Results showed that performance most often increases compared to basic fully connected neural networks and SARIMA models.

Authors in [9] applied exponentially weighted quantile regression for daily supermarket sales forecasting. Directly forecasting quantiles avoids the need for assumptions regarding the spread and shape of the distribution. Robust point forecasting was obtained as weighted average of EWQR forecasts for several quantiles. Empirical results showed that proposed method performed well compared to traditional methods for various forecasting horizons.

Gray relational analysis and multilayer functional link network (GMFLN) model combination for daily forecasting of perishable food sales in convenience store was applied by [10]. As time series model neural network was used. Gray relational analysis selects more influential factors as inputs to neural network (and their logarithmic and exponential transforms) that are used for forecasting. Results showed that improvement in forecasting accuracy compared to moving average, ARIMA, GARCH and classical neural networks was achieved. Later same authors [12] proposed another model that integrates Gray relational analysis and extreme learning machines with Taguchi method for ELM optimization for daily retail sales time series forecasting. Results showed that smallest forecasting error is achieved when comparing with other models (GMFLN and GBPN).

A combination of ARIMA and neural networks to perform hybrid forecasting of daily supermarket sales time series was done in [11]. First, the original time series are modeled by SARIMAX (special case of ARIMA) process and the errors between ARIMA model and original time series are modeled using neural networks. Forecasted value is given by sum of both model forecasts. Researchers showed that hybrid methodology outperforms standalone methods.

### 3. SALES FORECASTING MODELS

In this section computational intelligence models used for product sales modeling for forecasting are described.

*v-Support Vector Regression (v-SVR)*. An application of support vector machines model (originally developed for classification problems) for regression tasks is called support vector regression (SVR). In this research *v*-support vector regression (*v*-SVR) model is used.

Given input vectors such that  $\mathbf{x}_i \in R^n$  (*i*-th observation *n*-dimensional vector) and targets  $y_i \in R^1$ , *v*-SVR optimization problem is formulated as follows [13]:

$$\min_{\mathbf{w}, \varepsilon, \xi_i, \xi_i^*, b} \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \left[ \nu \varepsilon + \frac{1}{l} \sum_{i=1}^l (\xi_i + \xi_i^*) \right] \right\} \quad \Delta \mathbf{w} = - \frac{\mathbf{J}^T e}{\mathbf{J}^T \mathbf{J} + \mu \mathbf{I}} \quad (2)$$

subject to 
$$\begin{cases} \left[ \mathbf{w}^T \phi(\mathbf{x}_i) + b \right] - y_i \leq \varepsilon + \xi_i, \\ y_i - \left[ \mathbf{w}^T \phi(\mathbf{x}_i) + b \right] \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \\ \varepsilon \geq 0. \end{cases} \quad (1)$$

where  $\phi(x_i)$  is kernel function that performs mapping of input space to high dimensional feature space (the space where linear regression is performed);  $\mathbf{w}$  is a parameter vector of n-dimensional hyperplane;  $b$  is hyperplane bias parameter;  $\xi_i, \xi_i^*$  are upper and lower training errors (slack variables) subject to  $\varepsilon$  – insensitive tube;  $C$  is a cost parameter, that controls the trade-off between allowing training errors and forcing rigid margins;  $\nu$  is regularization parameter that controls parameter number of support vectors;  $l$  – is number of data points (observations). In this research radial basis kernel function is used. The hyperparameters (that need to be optimized externally) are cost parameter  $C$ , radial basis kernel function width parameter and parameter  $\nu$ . The illustration of support vector regression is presented in Fig. 1. In this research v-SVR code that is implemented in *LIBSVM* library is used.

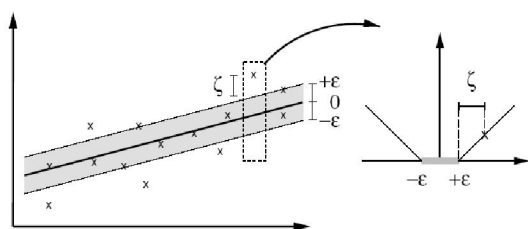


Fig. 1 – An illustration of soft margin loss function, the linear v-SVR case [14].

*Feed Forward Neural Networks (FFNN)*. In this research one hidden layer feed-forward neural network model is used. Logarithmic sigmoid transfer functions are used in the hidden layer and linear transfer function in the output layer. Levenberg – Marquardt backpropagation [15] training algorithm is selected. This training algorithm is frequently used as the most effective training algorithm for function approximation (regression) problems and uses mean-squared-error (MSE) cost function. It employs Jacobian matrix ( $\mathbf{J}$ ) w.r.t. network weights. An update of network weights using network output error ( $e$ ) is calculated by formula:

The main idea of this algorithm is to interpolate between Gauss – Newton and gradient descend algorithms by controlling scalar value  $\mu$ .

Levenberg – Marquardt backpropagation uses separate validation set as stopping criterion in order to effectively speed up training process. As parameter we use number of neurons in hidden layer. Neuron number is selected as percentage of number of observations in training set. For forecasted value estimation we use average ensemble of neural networks, because of random weight initialization during training. In this research we use neural network formation software from *MATLAB Neural Network Toolbox*.

*Gradient Boosting Regression (GBR)*. Boosting is kind of ensemble learning that combines simple (weak) learners into one strong ones. The idea of this model is to boost predictive performance by iteratively adding predictions of several elementary models. It is common to use decision trees as elementary models. At first iteration a simple model  $h(\mathbf{x})_1$  is fitted to training set  $\mathbf{y}$  (for example, as average constant at first step  $h(\mathbf{x})_1 = \frac{1}{N} \sum_{j=1}^N y_j$ ). In the next step, simple model is fitted to residuals  $e(\mathbf{x}) = \mathbf{y} - h(\mathbf{x})_1$  and so on. It learns how to improve performance by fitting models to residuals that improve performance compared to previous step model. The final model is obtained as linear combination of models fitted to residuals and a constant (average model):

$$f(\mathbf{x}) = \sum_{i=1}^M w_i h(\mathbf{x})_i + \text{const.} \quad (3)$$

A gradient boosting [16] is formulated as follows. Using square loss function for regression,

$$L(\mathbf{y}, f(\mathbf{x})) = \sum_{j=1}^N \frac{(y_j - f(x_j))^2}{2}, \quad (4)$$

gradient boosting assumes residuals to be gradients of error surface:

$$\frac{\partial L(y_j, f(x_j))}{\partial x_j} = y_j - f(x_j), \quad (5)$$

and using steepest descend, update is performed as

$$f(\mathbf{x})_i = f(\mathbf{x})_{i-1} - w_i \sum_{j=1}^N \frac{\partial L(y_j, f(x_j)_{i-1})}{\partial x_j} = f(\mathbf{x})_{i-1} - w_i \sum_{j=1}^N [y_j - f(x_j)_{i-1}], \quad (6)$$

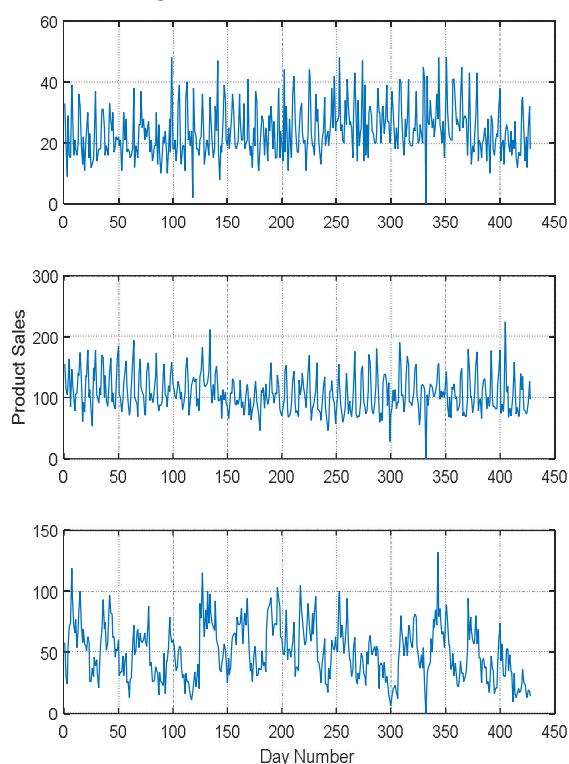
so that the iterative model fitting to residuals is considered as gradient descend minimization problem. This gradient view of residuals lets to choose any loss function  $L(\mathbf{y}, f(\mathbf{x}))$  that needs to be minimized. Algorithm selects functions in a greedy fashion. The multipliers  $w_i$  are selected using line search method.

In order to regularize model training, there are few parameters that need to be optimized externally using validation set: iteration number (from gradient descend perspective) or number of ensemble weak learners (from boosting perspective)  $M$ ; shrinkage parameter that controls gradient descend step parameter  $w$ ; weak learner parameters. Because we choose decision trees as a weak learner, the parameter is number of leafs in the tree.

A *MATLAB* code implementation of gradient boosted regression trees can be found in [17].

#### 4. METHODOLOGY AND EXPERIMENTAL DATA

As shortly mentioned in introduction section, for this research we selected data of 89 daily product sales time series with historical period of 15 months. Data were collected from stores in various regions that belong for one worldwide known retail company. Few time series examples showing complexity of daily product sales time series are illustrated in Fig. 2.



**Fig. 2 – Examples of product sales time series data used in this research.**

As data provider mentioned, this set consists of historical product sales that may be sensitive to weather events, so together with historical daily sales, weather data containing 17 different climatological indicators were provided from weather stations for corresponding regions nearby to particular store. Further we describe climatological inputs: maximum temperature; minimum temperature; average temperature; departure from normal temperature; average dew point; average wet bulb; heating (departure from  $\sim 18$  degree Celsius base); cooling (departure from  $\sim 18$  degree Celsius base); sunrise time; sunset time; snowfall (in centimeters); total precipitation (in centimeters); average station pressure; average sea level pressure; resultant wind speed; resultant wind direction; average wind speed. This weather information is further used as inputs to forecasting models.

Despite weather inputs other inputs that carry seasonal and trend information are extracted from time series together with calendar effect inputs. Further we describe those inputs: weekday; month number; day of month;  $i-1$  (previous) day sales;  $i-7$  day sales;  $i-14$  day sales;  $i-21$  day sales;  $i-28$  day sales; mean of previous 5 day sales. Those features were selected as most promising after preliminary experimentation and investigation of product sales process factors.

There are also some important local events that may significantly affect sales, such as school days, promotion days and other marketing events. However, data provider did not supplied this information so only features that were explained above are used in this research. Despite that, some holiday events that are stationary, such as New Year are captured by calendar inputs, other effects of events may be sensitive to region and even though might help in one dataset, may be redundant in other (for example, celebration days may differ from region to region, their effect may be controversial). Because we seek to obtain robust input feature set, results with event features may not well generalize to the future observations. Dynamic behavior that is significant and may exist because of some local events may be obtained using time series features explained above.

In order to assess effects of every possible elementary input set would take assessment of combination (using binomial formula) equal to  $2^D - 1 = 2^{26} - 1$  input sets (because there are total possible 26 inputs) we performed categorization of elementary inputs into total of 14 groups +/- weather inputs, (see Table 1).

As mentioned above for every time series there are 15 month historical data. In order to assess forecasting performance data were divided into training set (first part of 1 year length data) and test set (rest of data). For model parameter calibration 5 fold cross-validation method was selected (using training data set). Forecasting on test set was performed with each calibrated model (SVR, GBM and FFNN) and with different input sets +/- weather inputs.

In order to evaluate forecasting accuracy, a popular forecasting accuracy metric called MAPE (mean absolute percentage error) was used which is described as:

$$MAPE = \frac{1}{N} \sum_{j=1}^N \frac{|y_j - \hat{y}_j|}{(1/N) \sum_{k=1}^N y_k}, \quad (7)$$

where  $y$  is real sales value and  $\hat{y}$  is forecasted sales value.

**Table 1. Description of input feature sets.**

Feature Set Number	Description of input set
1	only weather features
2	weekday; month number; day of month
3	$i-7$ day; $i-14$ day; $i-21$ day; $i-28$ day sales
4	weekday; month number; day of month; $i-1$ day sales
5	weekday; month number; day of month; $i-7$ day; $i-14$ day; $i-21$ day; $i-28$ day sales
6	weekday; month number; day of month; mean of previous 5 day sales
7	$i-1$ day; $i-7$ day; $i-14$ day; $i-21$ day; $i-28$ day sales
8	$i-1$ day sales; mean of previous 5 day sales
9	$i-7$ day; $i-14$ day; $i-21$ day; $i-28$ day sales; mean of previous 5 day sales
10	weekday; month number; day of month; $i-1$ day; $i-7$ day; $i-14$ day; $i-21$ day; $i-28$ day sales
11	weekday; month number; day of month; $i-1$ day; mean of previous 5 day sales
12	weekday; month number; day of month; $i-7$ day; $i-14$ day; $i-21$ day; $i-28$ day sales; mean of previous 5 day sales
13	$i-1$ day; $i-7$ day; $i-14$ day; $i-21$ day; $i-28$ day sales; mean of previous 5 day sales
14	weekday; month number; day of month; $i-1$ day; $i-7$ day; $i-14$ day; $i-21$ day; $i-28$ day sales; mean of previous 5 day sales

## 5. RESULTS

In this section we present forecasting results. Table 2 summarizes all the results. First column in the table describe feature set number (same as in Table 1) which was used as input with particular forecasting model. Even though there are presented 14 input sets, 13 of them are used with and without weather input features, in order to assess importance of weather input feature set when they are added to particular input feature set for daily product sales forecasting. So for example feature set number 2 is weekday, month number and day of month feature set (total of 3 dimensions) which is used alone or together with weather input features (3+17=20 dimensions). Results in each column represent median MAPE value (median of 89 time series) when forecasted with each model separately (SVR (support vector regression), GBR (gradient boosted regression), FFNN (feedforward neural networks)) and last set of columns present results averaged over all three models (median of 3\*89=267 forecasting MAPE values). Together with median MAPE

columns we present additional columns with statistical hypothesis testing in order to evaluate effect of adding weather input features for particular input set. This test is called sign test, which tests whether difference between sets (which is itself a random variable) has median equal to 0. A small  $p$ -value (usually set to be less than 0.05) rejects hypothesis that random variable (the difference) has median equal to 0. In this case small  $p$ -value shows that distribution of difference between MAPEs with added weather input set and without added weather input set for particular input set has nonzero median, which means it is biased and input set with smaller median MAPE is significantly better compared to other set. So for example, considering feature set number 2 we see that  $p$ -value is 0.00, together with information that median MAPE with weather features is 40.81 and without is 37.05, we can conclude that weather inputs significantly worsens forecasting accuracy when added to particular feature set. At first it would seem reasonable to do, for example, Students  $t$ -test for both sets (MAPE distributions with and without weather features),

however we cannot assume that those distributions are mutually independent (in fact dependency is high, when MAPE is high for particular time series it might remain high despite the forecasting model and inputs used because of high signal-to-noise ratio, so MAPE values depend on particular time series), however, the independence of difference between MAPEs (MAPE difference between all 89 time series forecasted with added weather features

and 89 time series forecasted without added weather features with particular model) can be considered random variable. High  $p$ -value means that there is no significant difference of forecasting performance when weather features are added or not. From the results in Table 2 we can see that almost all the time adding weather features only worsens forecasting performance. This means that weather features do not capture useful information for forecasting.

Table 2. Daily sales forecasting results.

Feature Set Number	SVR Median MAPE (%) with weather features	SVR Median MAPE (%) without weather features	SVR Sign test p-value	GBR Median MAPE (%) with weather features	GBR Median MAPE (%) with weather features	GBR Sign test p-value	FFNN Median MAPE (%) with weather features	FFNN Median MAPE (%) without weather features	FFNN Sign test p-value	All Median MAPE (%) with weather features	All Median MAPE (%) without weather features	All Sign test p-value
1	39.85	-	-	42.22	-	-	43.34	-	-	41.16	-	-
2	40.81	37.05	0.00	40.92	39.85	0.67	40.82	38.83	0.00	40.82	39.37	0.00
3	35.42	33.49	0.01	34.86	34.34	0.40	37.25	34.68	0.00	35.47	34.34	0.00
4	38.84	36.33	0.00	38.53	38.67	0.40	37.89	36.53	0.00	38.53	36.93	0.00
5	35.79	36.07	0.53	36.83	37.07	0.53	37.82	36.52	0.02	36.84	36.52	0.03
6	38.71	34.32	0.00	37.38	37.06	0.29	38.35	35.90	0.00	37.94	35.11	0.00
7	33.19	32.72	0.14	33.97	33.44	0.01	35.03	33.47	0.00	33.86	33.17	0.00
8	39.04	36.51	0.02	38.46	36.34	0.20	40.75	37.10	0.14	39.14	36.55	0.00
9	32.66	31.96	0.00	34.56	33.70	0.14	34.60	32.91	0.00	34.16	33.60	0.00
10	35.92	33.99	0.03	35.29	35.81	0.03	36.78	34.39	0.02	36.01	34.61	0.00
11	37.84	34.78	0.00	36.90	36.45	0.40	38.04	35.47	0.00	37.52	35.60	0.00
12	35.46	34.18	0.02	36.43	35.40	0.29	36.78	34.67	0.02	36.10	34.67	0.00
13	32.82	<b>31.89</b>	0.03	34.17	33.42	0.06	34.69	32.67	0.00	34.05	<b>32.87</b>	0.00
14	35.86	32.46	0.03	36.16	35.29	0.03	36.90	34.44	0.00	36.16	33.96	0.00

The graphical illustration of the results is depicted in Fig. 3. From Fig. 3 it is seen that average of all 3 models when forecasted with added weather features worsens forecasting performance. Also we see that SVR model performs best (without added weather features case) for all feature sets when considering separate model forecasts. Also we see few best feature sets that have smallest median MAPE (averaged over all models): input feature number 13, 7, 9 and 14 without added weather features. The number 13 has smallest median MAPE, but in order to test whether it has significantly different median, we again perform sign test of MAPE difference between all pairs of sets with set number 13: 13-7;13-9;13-14. We got following  $p$ -values accordingly: 0.0009; 0.00005; 0.00009. So this means that feature set number 13 (without added weather features) is most informative considering all models. Also it is most informative considering all forecasting models separately: SVR with feature number 13 (without added weather

features) achieves the most accurate forecasting performance.

Fig. 4 shows an example of best sales forecasting results (best according average MAPE between all 3 forecasting models).

Additionally, in order to evaluate information gain of every individual feature, we added individual feature importance graph (Fig. 5) using RReliefF filter feature selection method [18] that is widely used in practice. Following numbers in the graph show importance of following features: 1) month number; 2) day of month; 3) weekday; 4) value of  $i-1$  day; 5) value of  $i-7$  day; 6) value of  $i-14$  day; 7) value of  $i-21$  day; 8) value of  $i-28$  day; 9) mean of previous 5 day sales; 10) maximum temperature; 11) minimum temperature; 12) average temperature; 13) departure from normal temperature; 14) average dew point; 15) average wet bulb; 16) cooling (departure from  $\sim 18$  degree Celsius base); 17) heating (departure from  $\sim 18$  degree Celsius base); 18) sunrise time; 19) sunset time; 20) snowfall (in centimeters); 21) total precipitation

(in centimeters); 22) average station pressure; 23) average sea level pressure; 24) resultant wind speed; 25) resultant wind direction; 26) average wind speed. However it is worth to mention that some weather features were not available for some stores. So estimation variance in some weather features might be bigger and overestimation or underestimation may occur.

Looking at the results in Fig. 5 it is seen that largest weights are given to calendar (except month number) and time series features. All individual weather features have less average weights, meaning they contain very little useful information for forecasting.

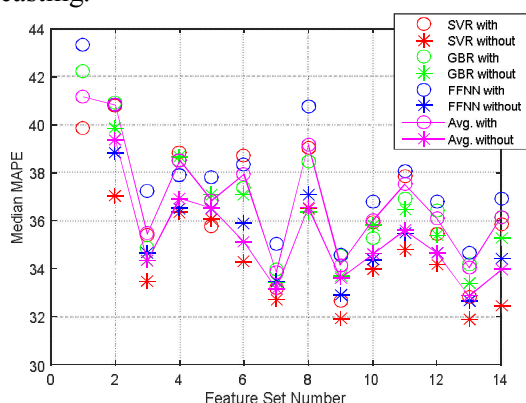


Fig. 3 –A graphical illustration of results in Table 2.

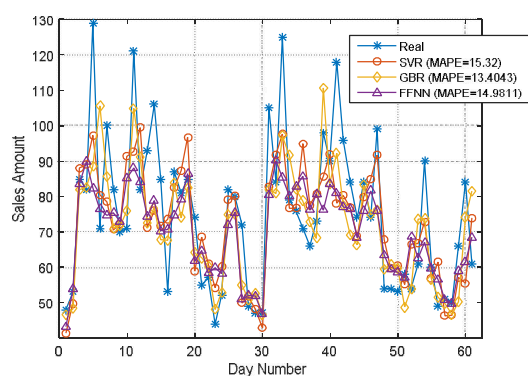


Fig. 4 –Example of best sales time series (according to average MAPE between models) forecasting results.

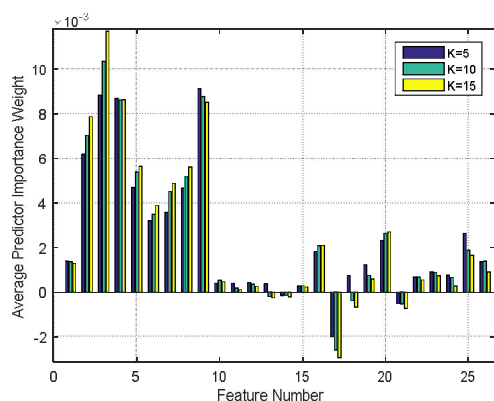


Fig. 5 –Individual feature importance weights using RRELIEFF feature ranking method with various K parameters. Results are relatively consistent with respect to parameter K.

## 6. CONCLUSIONS

In this paper we presented research related to product sales forecasting when various computational intelligence models and input features are used for daily sales time series forecasting.

Results showed that weather input features alone or added with other features do not improve performance when compared to other features extracted from time series itself (value of  $i-1$ ,  $i-7$ ,  $i-14$ ,  $i-21$ ,  $i-28$  day and mean of previous 5 day sales) or calendar effect input features (month number, day of month, weekday). Results also showed that features extracted from time series (feature set number 13 in Table 1) let achieve best forecasting accuracy when averaged over all models and overall (using only with v- support vector regression model). Another finding is that v- support vector regression model seems to be the most appropriate forecasting model between all three models achieving smallest median mean absolute percentage error for all input features (without weather inputs).

The limitation of this research methodology is that forecasting techniques used in this research are classical standalone computational intelligence methods. However in practice there exist variations of those methods and various combinations (partly discussed in literature review section). This research main purpose is to evaluate what input features are best for sales forecasting independently of what computational intelligence techniques are used for regression. The feature sets investigated are universal. Even though there might be other information available (like promotion days), it is not universal and may not be effective or available for different areas. In order to efficiently narrow research scope, input feature subsets were selected after prior knowledge and experimentation.

## REFERENCES

- [1] P. Doganis, E. Aggelogiannaki, H. Sarimveis, A combined model predictive control and time series forecasting framework for production inventory systems, *International Journal of Production Research*, (46) (2008), pp. 6841-6853.
- [2] W. I-Lee, B.Y. Shih, C.-Y. Chen, A hybrid artificial intelligence sales-forecasting system in the convenience store industry, *Human Factors and Ergonomics in Manufacturing & Service Industries*, (22) (2012), pp. 188-196.
- [3] P. Doganis, A. Alexandridis, P. Patrinos, H. Sarimveis, Time series sales forecastig for short shelf-life food products based on artificial neural networks and evolutionary computing,

- Journal of Food Engineering*, (75) (2006), pp. 196-204.
- [4] T.-M. Choi, Y. Yu, K.-F. Au, A hybrid SARIMA wavelet transform method for sales forecasting, *Decision Support Systems*, (51) (2011), pp. 130-140.
- [5] R. J. Kuo, P. Wu, C. P. Wang, An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination, *Neural Networks*, (15) (2002), pp. 909-925.
- [6] L.-M. Liu, S. Bhattacharyya, S. L. Sclove, R. Chen, W. J. Lattyak, Data mining on time series: an illustration using fast-food restaurant franchise data, *Computational Statistics & Data Analysis*, (37) (2001), pp. 455-476.
- [7] X. F. Du, S. C. H. Leung, J. L. Zhang, K. K. Lai, Demand forecasting of perishable farm products using support vector machine, *International Journal of Systems Science*, (44) (2013), pp. 556-567.
- [8] K.-F. Au, T.-M. Choi, Y. Yu, Fashion retail forecasting by evolutionary neural networks, *International Journal of Production Economics*, (114) (2008), pp. 615-630.
- [9] J. W. Taylor, Forecasting daily supermarket sales using exponentially weighted quantile regression, *European Journal of Operational Research*, (178) (2007), pp. 154-167.
- [10] F. L. Chen, T. Y. Ou, Gray relation analysis and multilayer functional link network sales forecasting model for perishable food in convenience store, *Expert Systems with Applications*, (36) (2009), pp. 7054-7063.
- [11] L. Aburto, R. Weber, Improved supply chain management based on hybrid demand forecasts, *Applied Soft Computing*, (7) (2007), pp. 136-144.
- [12] F. L. Chen, T. Y. Ou, Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry, *Expert Systems with Applications*, (38) (2011), pp. 1336-1345.
- [13] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology (TIST)*, (2) 27 (2011), pp. 1-27.
- [14] B. Scholkopf, A. J. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2002, 18 p.
- [15] M. T. Hagan, M. Menhaj, Training feedforward networks with the Marquardt algorithm, *IEEE Transactions on Neural Networks*, (5) (1994), pp. 989-993.
- [16] J. H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, *The Annals of Statistics*, (29) 5 (2001), pp. 1189-1232.
- [17] K. Hara and R. Chellappa, Computationally Efficient Regression on a Dependency Graph for Human Pose Estimation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, OR, USA (June 23-28, 2013), pp. 3390-3397. MATLAB code available at: <http://www.mathworks.com/matlabcentral/fileexchange/42130-boosted-binary-regression-trees>, accessed September 2015.
- [18] M. R. Šikonja and I. Kononenko, An adaptation of Relief for attribute estimation in regression, in *Proceedings of the International Conference on Machine Learning (ICML 97)*, 1997, pp. 296-304.



**Gediminas Žylius** is an informatics engineering PhD student at Department of Automation, Kaunas University of Technology, Lithuania (supervisor prof. habil. dr. Rimvydas Simutis). He obtained his M. Sc. diploma with honors in Control Technologies (Intelligent Control Systems) in year of 2015. PhD research topic is development and investigation of specialized forecasting and optimization methods for financial services sector. His research interests are (but not limited to) machine learning, time series forecasting, intelligent control systems.



**Rimvydas Simutis** is professor of automation and control system engineering at Department of Automation, Kaunas University of Technology, Lithuania. He obtained his PhD in Control Engineering at Kaunas University of Technology and worked several years as Alexander von Humboldt grant holder and post-doctoral researcher at Leibniz Universität Hannover. His research interest includes advanced modeling, optimization and control system techniques with application in Technical and Financial systems. He participated in various research projects with international automation companies as senior researcher or supervisor.



**Vygandas Vaitkus** is associate professor of automation and control system engineering at Department of Automation, Kaunas University of Technology, Lithuania. He obtained his PhD in Control Engineering at Kaunas University of Technology. His research interest includes process identification and optimization with application in Technical and Financial systems. He participated in various research projects as supervisor.



