Demand Forecasting Based on the Set of Short Time Series

Arnis Kirshners, *Riga Technical University*, Galina Kuleshova, *Riga Technical University*, Arkady Borisov, *Riga Technical University*

Abstract **– This paper addresses the task of short historical time series and discrete descriptive parameters processing aimed at making demand forecast only on the basis of new product describing parameters. Several data mining methods are used for data processing including data extraction, pre-processing, cluster analysis and classification. Data preparation for data mining processes is made with user-defined parameters entered in the forecasting system. In the selected short historical time series the membership of an object in any class, which is a basis for creating prototypes, is determined using clustering. The k-means clustering algorithm is employed for finding the optimal number of clusters in the sample. The number of clusters is determined on the basis of the mean absolute error. As a result of classification, using inductive decision trees, a correlation between the prototype produced in the course of clustering and product describing parameters is determined. For new product demand clustering, a decision tree obtained as a result of classification is used. New product describing parameters are then projected on the tree, and a tree leave indicating the number of the prototype produced by means of clustering is found. The prototype curve structure depicts possible demand for a new product for the next period.**

Keywords **– short time series, data mining, clusterization, classification, decision tree**

I. Introduction

Commonly, demand forecasting is based on finding the regularities in time series related to the changes in economic indexes, for example, currency rate fluctuations, or also in technical systems, say, functioning conditions or environment modifications. It is assumed that in this kind of systems a time series is stationary, and the longer the observation length, the greater the probability that the regularities are determined effectively. (Armstrong et al. [1]).

However, in real life plenty of tasks exist in which the life cycle of time series is relatively short, and it is practically impossible to determine the regularities in time series like that. These tasks include goods life cycle analysis (Kirshners et al. [2]), e-service analysis during the initial stage of its introduction (Kirshners et al. [3]) and textile goods sales analysis at a wide choice of products and a short life cycle (Thomassey et al. [4]) etc. Here a forecasting task is formed that is based on a large amount of comparatively short time series where in each of the cases considered historical demand data are changed.

In this paper demand forecasting is accomplished for a small enterprise engaged in the retail trade of clothing. Demand forecasting is based on the forecasting system which determines an optimal order quantity in ordering new collection goods, which helps to avoid warehouse overstocking. The formation of this kind of system could help the enterprise to save and optimise circulating assets.

The system under consideration has to meet these requirements: easy exploitation and easy to understand representation of the results. This means that the system will be exploited by company's managers who deal with new collection ordering and who are not experts in the field of information technology but are able to work on a computer at the user level. Due to that, when making an order, the system has to ask the user possibly less questions so as to ensure a more effective use of manager time resources. Based on these requierements, a demand forecasting system aimed to perform the following actions has to be created:

- Data loading from a database according to the user's defined period (one year) on whose basis model training process will be carried out;
- Processing a large number of demands possessing different product life cycle duration;
- Preparing the selected data;
- Forecasting on the basis of historical demand data;
- Obtained results representation.

A forecast regarding a new product has to be made using only that product describing parameters like type, kind, price etc. Due to that, the forecast of that product is made using prototypes (demand period within the average value cluster) (Devisscher et al. [5]) that are produced in the course of historical demand data clustering (Kirshners et al. [6]). The correlation between the new product describing parameters and prototypes obtained as a result of clustering is determined on the basis of inductive decision trees. For that purpose, different existing technologies can be used on whose basis it is possible to combine continuous data (short time series) and discrete data (the product describing parameters) and, as a result, to acquire new knowledge (Kirshners et al. [6], Parshutin et al. [7], Symeonidis et al. [8]).

The prototypes obtained characterise the evaluation of the product historical demand in corresponding cluster that is combined with the parameters describing that product. Analysing the data set obtained by means of inductive decision trees, new knowledge is mined in the form of conditional rules where the left-hand part indicates demand data but the right-hand side indicates the parameters describing the product (Written et al. [9], Barsegyan et al. [10]). On the basis of the obtained inductive decision tree that is constructed using the product describing parameters, the parameters describing the new product are projected. As a result, the cluster number is determined that indicates the number of the prototype model derived in the course of clustering. The structure of the defined prototype curve indicates possible demand for the new product in the future. The constructed forecasting model has to ensure a high accuracy, due to that in the course of its construction clustering as well as classification errors have to be reduced.

II. Work objective, tasks and researches

The objective of this work is to develop a demand forecasting system aimed to optimise a trading company ordering process using the demand experience of previous years and new ordered product data that are characterised by these descriptive parameters: the price, type of goods, goods seasonality, collection lifecycle in months, colour, size and others. As a result, a forecast for possible product demand would be produced that would help manager to make a decision in ordering new clothing collection. To achieve the objective, these tasks have to be fulfilled:

- historical demand data summarisation according to the user's defined period;
- selected data preparation for data mining processes;
- pre-processing of the prepared data;
- finding optimal number of clusters and constructing prototype models;
- analysing the results of clustering;
- clustering results combination with product describing historical parameters, e.g., price, product type, product seasonality, colour, size etc.;
- new data classification based on inductive decision trees;
- analysing classification processes and results;
- testing the performance of the model developed;
- forecasting the demand for a new product;
- representation of the results.

To fulfil the tasks stated, plenty of researches have to be conducted that are related to data clustering and classification. Different solutions exist that implement the aforementioned tasks, for example, regression models, neural networks (Salam et al. [11]), inductive decision trees (Quinlan [12], Borisov et al. [13]) and associative analysis (Das et al. [14]). However, these models cannot be directly used for forecasting new orders in the clothing business since historical demand data change every year, the demand is rather high and a forecast has to be made based on the available information about a new product, say, the price, type, kind etc. Due to that, it is worth developing prototypes in historicala data and looking for associations among the demands in them. By combining these data structures a forecast can then be made.

Different techniques exist how to represent these prototype models in a data set and to extract new knowledge from it, e.g., with the help of decision trees, combination in clusters, neural networks or association rules. Out of these techniques, the one has to be chosen that would be easier to interpret and understand for the end user, so it would be more useful to employ cluster analysis determining prototype models in historical demand data. By analysing clustering results every object's class or prototype is determined which is combined with product describing parameters in a new data set. In the newly obtained data set, the correlation between prototype models and descriptive data is searched for.

It should be noted that using such different data type combination methods it is important to achieve that continuous data (e.g., product demand during a certain time period) are tightly connected with discrete data (demand describing attributes: the price, type, kind, colour, size etc.); in this case the correlations obtained as a result of classification will be true, which will lessen the error of the results to be forecasted.

III. System of demand forecasting

Demand forecasting system (see Fig. 1) is developed using the historical demand data (short time series) that are stored in the database management system (DBMS) *Microsoft Office Access 2007* (Sennov [15]). The user has to enter this information into the system:

- Historical demand period (year) which is sent to the DBMS with the help of language *PL/SQL* query (Andon et al. [16]);
- New product describing parameters (kind, type and the price) that are in their turn sent to the block *Forecasting Model* for new data classification.

The data selected as a result of the query, which contain historical information (i.e., real product sales data and three descriptive parameters of every product: the product type, price and kind) are passed to the data preparation block where the existing DBMS format with extension **.mdf* is transformed to the acceptable for the system format with extension **.csv*. The data set produced is divided into two subsets: one consisting of historical demand data (short time series) and objects identification numbers and the other consisting of historical data describing parameters (kind, type, price,) and objects identification numbers.

IV. Model of forecasting

Further, the data arrive at the forecasting system model and are pre-processed, i.e., they are cleaned from noise and improperly displayed data (Written et al. [9], Barsegyan et al. [10], Sukovs et al. [17]).

The data normalisation is based on the neutralisation of the dominant attribute values (Thomassey et al. [4], Kirshners et al. [6]) by formula (1):

$$y_i = \frac{x_i}{\sum_{j=1}^{n} x_j}, \tag{1}$$

where

y_i – the normalized attribute value at the time moment i;
x_i – the value of the time series x at the time moment i;
x_j – the value of the time series x in the period n;
n – the time series duration (number of periods).

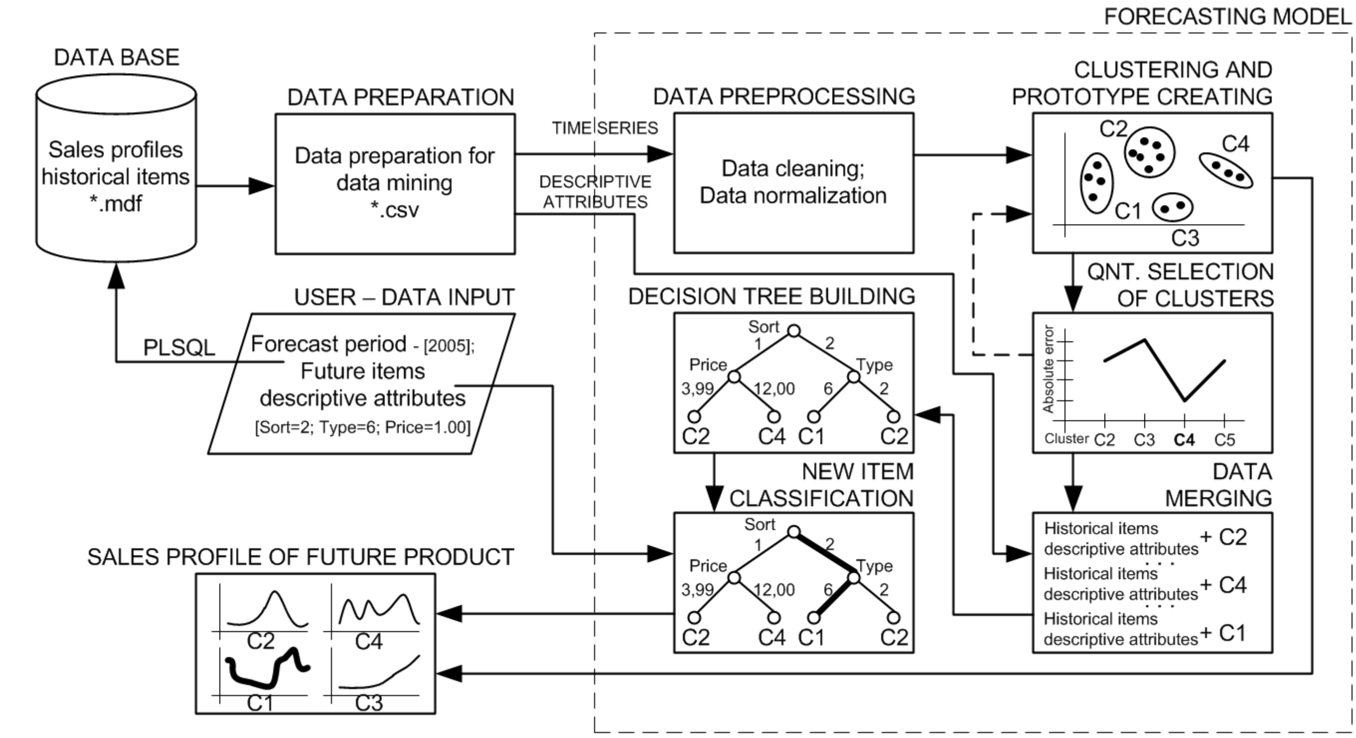

Fig. 1. System of demand forecasting

Within this study, for grouping similar short time series the *k-means* algorithm is employed (Barsegyan et al. [10], Sukovs et al. [17]), but for classification - inductive decision trees (Written et al. [9], Barsegyan et al. [10]) with the algorithm *C4.5* (Quinlan [12]) are used.

A. Clustering analysis

The demand values are grouped into clusters so that those historical demand values that have similar curve structure, are in one group. The closeness between the demand values (time series) is determined using a Euclidean measure with the help of the modified k-means algorithm. The clustering has to be multiply repeated since in the *k-means* algorithm the initial centroids (cluster centres) are chosen randomly (see Fig. 2) By forming this group combination according to the mean values, a prototype model (see Fig. 2 , bottom) is obtained that represents the average demand value of this group in each period. In this way, group sets are determined for all clusters

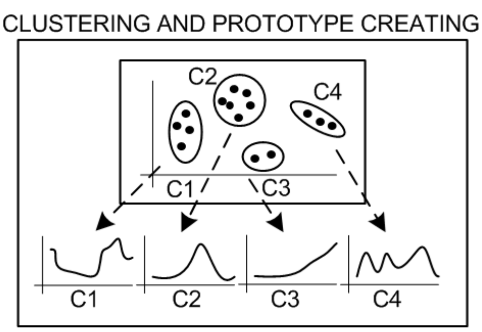

Fig. 2. Clustering and creating prototypes

within the interval from two to the maximum calculated by formula (2) (Thomassey et al. [4], Kirshners et al. [6]):

$$Max_{count\ of\ clusters} = \sqrt{n}, \tag{2}$$

where

$Max_{count\ of\ clusters}$ – the maximum count of clusters;
n – the number of records in the dataset.

The maximum calculated has to be large enough to accurately perform the clustering and at the same time not too large to help prevent noise that would affect the cluster ditribution.

B. Count selections of clusters

The number of clusters in the data set is determined taking into account the mean clustering absolute error as follows. First, the distance matrix (see Table I) is obtained using the *k-means* algorithm. Then, modifying the algorithm (Kirshners et al. [6]) the mean absolute error (Thomassey et al. [4], Montgomery et al. [18]) for each cluster is calculated. For that purpose, the distances to the centroid are summed over each cluster; the obtained sum is then divided by the number of records in the cluster by formula (3):

$$AE = \frac{d_1+d_2+\cdots+d_n}{c_n}, \tag{3}$$

where

d_1, d_2, d_n – the distances from the corresponding record to the centroid;
c_n – the number of records in a cluster;
AE – the mean absolute error in a cluster.

TABLE I

OBJECT DISTANCES TO THE CENTROIDS USED FOR CALCULATING THE MEAN CLUSTERING ABSOLUTE ERROR

NUMBER OF OBJECT	PERIODS				DISTANCE MEASURE
C	T1	T2	…	T12	D
C_1					D_1
C_2					D_2
…					…
C_N					D_N

Further, in each cluster AE_n, the obtained values of the mean absolute error are summed up and divided by the number of clusters C_n, and the mean clustering absolute error *MeanAE* (Montgomery et al. [18]) is calculated by formula (4):

$$MeanAE = \frac{AE_1 + AE_2 + \cdots + AE_n}{C_n}, \tag{4}$$

The graph of the mean clustering absolute error as a function of the number of clusters (see Fig. 3) illustrates that the first noticeable minimum is achieved at 10 clusters. Each point in Fig. 3 is found by the summation of the mean absolute errors in a cluster from Table II and the division of this sum by the number of clusters under consideration.

Since the variations observable on the graph after the point 10 are minimal, the authors have chosen this count of clusters for further experiments. The prototypes obtained at the number of clusters equal to 10, are optionally displayed for the second and tenth clusters (see Fig. 4).

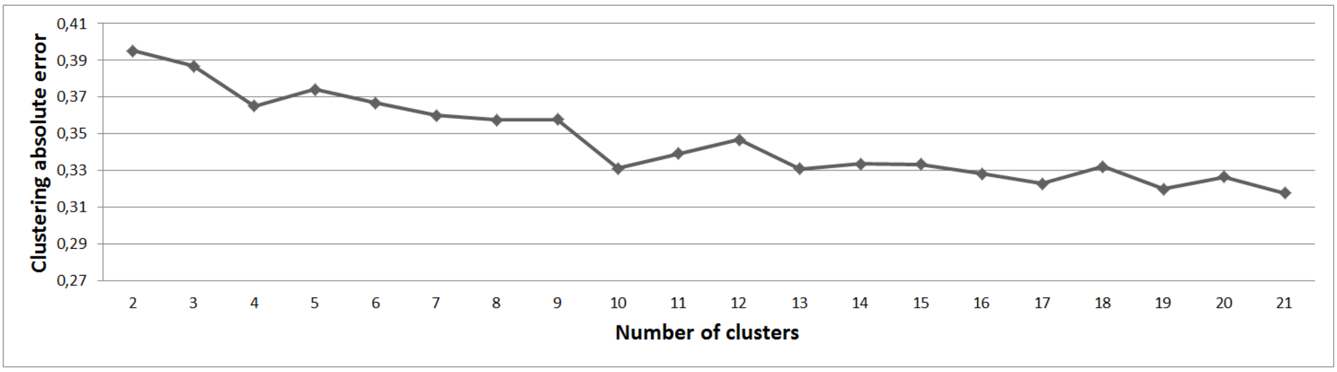

Fig. 3. *MeanAE* at different counts of clusters

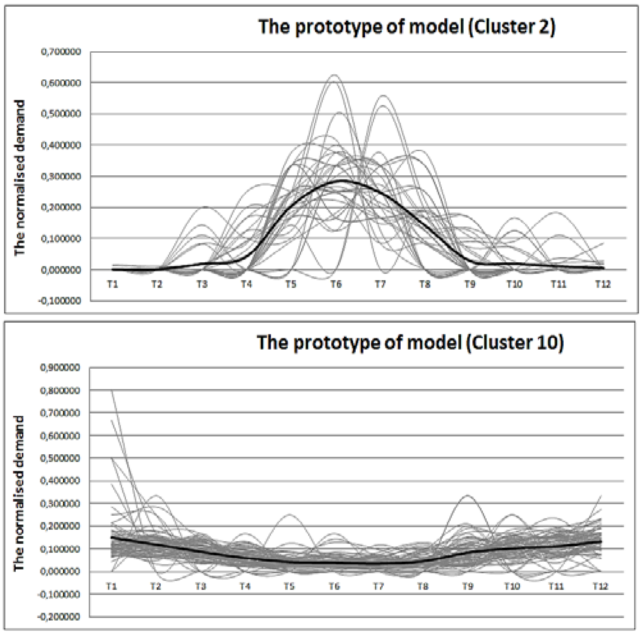

Fig. 4. Clusters and their links with the prototype, notation: the prototype in a cluster (bold line) and the demand curves (thin lines).

A. Different data types combination

Data combination procedure is executed for discrete product-describing parameters (kind, type and price), adding the class or prototype model number produced by clustering. The records are combined according to their object identification numbers. The data set obtained is then employed for constructing an inductive decision tree, searching for the links between prototype models and product-describing parameters.

B. Classification by decision tree

To determine the correlation between the prototype models and the product-describing parameters, inductive decision trees (Quinlan [12]) were chosen as the classification rules obtained with their help are less complex and easier to interpret for the end user as compared to those produced with the use of a neural network (Craven et al. [19]).

The decision tree's leaves display one of the classes, i.e. one of the prototype models, obtained by means of clustering. Each internal decision tree node contains one of the values that describe the product's discrete parameters in the data set.

TABLE II

MEAN ABSOLUTE ERROR IN A CLUSTER (CLUSTERING WITH 10 CLUSTERS)

	Number of cluster (prototype)									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Number of records	81	36	20	14	105	36	23	18	18	72
The mean clustering absolute error	0.217	0.372	0.314	0.29	0.205	0.424	0.348	0.467	0.416	0.248

To obtain a more compact decision tree and effective classification rule set, pruning of the branches of minor importance with the post-pruning method (Quinlan [12]) is performed that is commonly used after the inductive decision tree is constructed. The *C4.5* algorithm (Quinlan [12]) applied in classification is able to process different data types and uses one of the post-pruning techniques for reducing the tree size. The inductive decision tree was constructed with the *Weka 3.6.1* program using 10-fold cross-validation which ensured a a more clear determination of the classifier accuracy (Kohavi [20]). The constructed decision tree (see Fig. 5) depicts the links between the prototype models and the

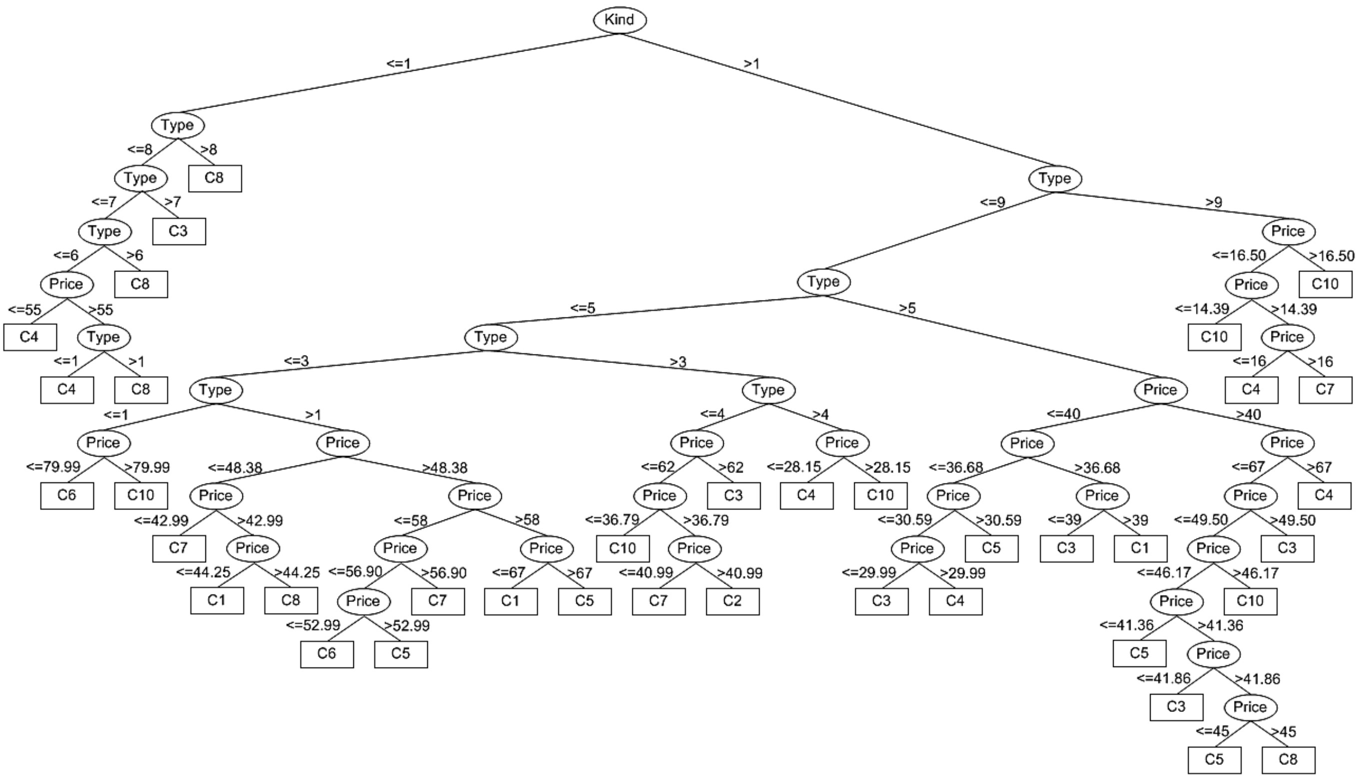

Fig. 5. Decision tree

the product-describing parameters. The parameters of the decision tree constructed are given in Table III.

TABLE III

PARAMETERS OF THE CONSTRUCTED DECISION TREE

Number of classes	Number of records	Number of attributes	Number of leaves	Number of nodes
10	423	4	38	75

The leaves of the decision tree display one of the classes, i.e., one of the prototype models; the nodes contain the product-describing attributes and the figures above the branches are splitted values.

In case if the same dataset is used both for decision tree training and testing, a problem appears that is related to classifier adjusting to the learning set data. Due to that, in this study a separate dataset was used for testing. The evaluation of the constructed forecasting model as well as calculation of the clustering and classification errors using the test set are described in more detail in (Kirshners et al. [6]).

C. New product forecasting

Since the product lifecycle in the task under consideration is comparatively small, the constructed forecasting system model, when analysing historical demand data also develops prototype models based on the time interval which is 12 months, or demand periods. When a forecast for a new product is made, the decision maker develops a prototype model using the experience of the year which in his opinion is close or similar to the year being forecasted.

New product demand forecasting is based on the constructed inductive decision tree (see Fig. 5) projecting the on it new product describing parameters, which, in particular, include these data:

1. Kind of product (fixed-demand - 2);
2. Type of product (pants - 3);
3. Price (56.69).

As a result, a decision tree is constructed (see Fig. 6) on which the bold line depicts the tree projection path till the tree's leave is reached which indicates the number of the demand prototype model, which is 5 in this case, and the graph of the prototype model (see Fig. 7) corresponds to the prototype model of the fifth cluster.

Looking at the curve obtained, the decision maker can draw a conclusion that possible demand for a new product will be topical all the year with an up-going trend in spring and summer months.

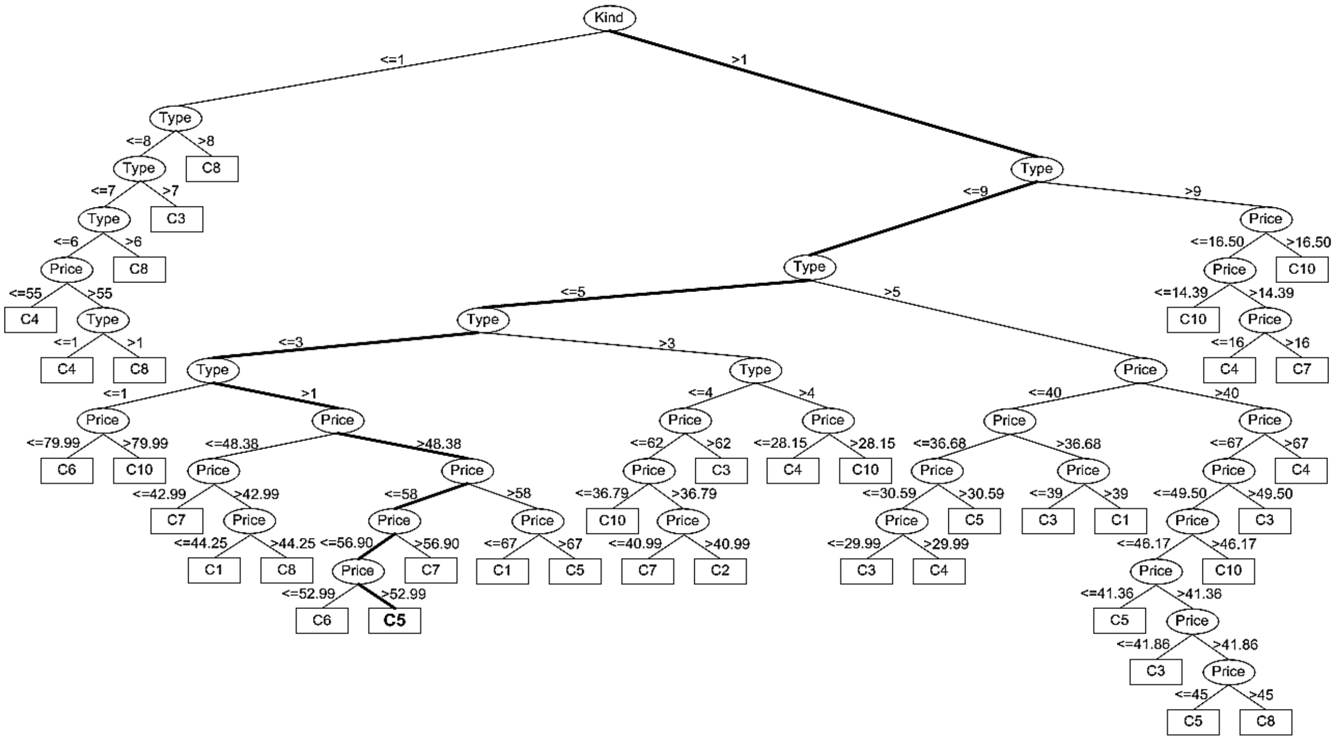

Fig. 6. Decision tree projected with new product-describing parameters

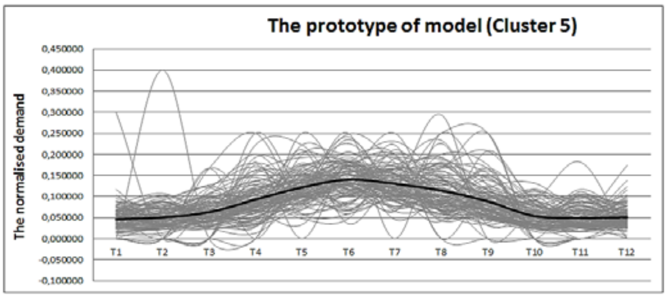

Fig. 7. Possible demand for a new product for next 12 months (bold line)

Once a forecast is prepared, the decision maker can determine new product ordering quantities and delivery times more accurately.

V. USED DATA

In this study the demand data of a retail clothing company for year 2005 were used that contained 423 records after the data cleaning process which were applied in model training and 149 records for year 2006 used for model testing that were selected randomly and contained 35% of the number of records of the training set as well as characterize all product kind and types. In order to enable the use of data for comparison with the constructed prototype models, their lifecycle was normalised to 12 months. The training set also contained the product-describing attributes (kind, type and price) about each object. The product kind was characterised as follows: seasonable (1) and fixed-demand (2) products, where their discretised values are given in parentheses. In turn, product types can be as follows: jackets (1); vests (2); pants (3); shirts (4); jumpers (5); bags (6); belts (7); shorts (8); t-shirts (9), shoes (10) and others (11). Both these attributes contained discrete values but the attribute *price* contained the continuous values.

VI. RESULTS OF EXPERIMENTS

The constructed demand forecasting system was trained using the training set for year 2005 and tested using the test data set for year 2006; in the course of these processes, 10-fold cross-validation was used.

From the clustering results (see Fig 4 and Fig 7), it can be concluded that effective object partitioning by classes is achieved. Based on the effective cluster count obtained for the given training data set (see Table II) and forecasting model evaluation with the test set (Kirshners et al. [6]) it is possible to state that the constructed system has a sufficiently high accuracy.

The obtained inductive decision tree structure (see Table III) denotes the complex tree structure; at the same time, it is easy to interpret. The product-describing attribute *price* is located on the lowest tree structure levels, which indicates that this attribute has a low informativity level when choosing a prototype model. Post-pruning performed after the tree construction enables a significant tree size reduction without loosing the link between the attributes at the same time.

VII. Conclusions

Using the technique discussed, a demand forecasting system is developed that implements a new product forecast for the next 12 months based on short historical time series and product-describing discrete parameters. A system with the integrated forecasting module is worked out that enables the decision maker to obtain possible forecast for a new product within a short period of time based on these data entered into the system: the historical period (year) and new product describing parameters (kind, type and price).

The developed demand forecasting system enables, according to the user entered parameters, automatic data selection and pre-processing from the database, sending the obtained data sets to the forecasting model.

The modification to the *k-means* clustering algorithm is offered that enables the determinationof the count of clusters on the basis of the mean clustering absolute error.

Prototype models are constructed in clusters on the basis of the clustering results of the training set on whose basis a forecast for a new product was made. The constructed prototype models that are visually understandable even for non-specialists characterise the demand structure of the class.

A combination of data mining techniques is developed that implements the operation of the forecasting system model, which is based on the modification of the *k-means* clustering algorithm used for processing short time series with demand and the *C4.5* classification algorithm.

VIII. Further research

Future work could employ more complicated data pre-processing techniques aimed to improve the results achieved as well as to choose a data set with more records, which would lessen the clustering and classification errors even more.

Acknowledgments

This work has been supported by the European Social Fund within the project «Support for the implementation of doctoral studies at Riga Technical University».

References

[1] **J. S. Armstrong**, **F. Collopy**, **J. T. Yokum**, "Decomposition by causal forces: A procedure for forecasting complex time series," International Journal of Forecasting, 21, 2005, pp. 25-36.
[2] **A. Kirshners**, **A. Sukov**, "Rule induction for forecasting transition points in product life cycle data," *Scientific proceedings of Riga Technical University, Information Technology and Management Science,* Issue 5, Vol.36, RTU, Riga, 2008, pp. 170-177.
[3] **A. Kirshners**, **Y. Kornienko**, "Time-series data mining for e-service application analysis," *Scientific Proceedings of Riga Technical University, Information Technology and Management Science*, Issue 5, Vol. 40, RTU, Riga, 2009, pp. 94-100.
[4] **S. Thomassey**, **A. Fiordaliso**, "A hybrid sales forecasting system based on clustering and decision trees," Decision Support Systems, Volume 42, Issue 1, 2006, pp. 408-421.
[5] **M. Devisscher**, **B. De Baets**, **I. Nopens**, *Pattern discovery in intensive care data through sequence alignment of qualitative trends: proof of concept on a diuresis dataset. Appearing in the Proceedings of the ICML/UAI/COLT 2008 Workshop on Machine Learning for Health-Care Applications,* Helsinki, Finland, 2008.
[6] **A. Kirshners**, **S. Parshutin**, **A. Borisov**, "Combining clustering and a decision tree classifier in a forecasting task," Automatic Control and Computer Sciences, Vol. 44, N 3, 2010, pp. 124-132.
[7] **S. Parshutin**, **A. Borisov**, "Data mining driven decision support. Polish Journal of Environmental Studies," Vol. 18, N4A, 2009, pp. 8-11.
[8] **A. L. Symeonidis** et al., "Data mining for agent reasoning: A synergy for training intelligent agents," Engineering Applications of Artificial Intelligence, 20 (8), 2007, pp. 1097-1111.
[9] **I. H. Written**, **E. Frank**, *Data mining: practical machine learning tools and techniques - 2nd edition*. Amsterdam etc.: Morgan Kaufman, 2005.
[10] **A. Barsegyan**, **M. Kupriyanov**, **V. Stepanenko**, **I. Holod**, *Technologies of Data Analysis: Data Mining, Visual Mining, Text Mining, OLAP*. St.Petersburg, 2007, 384. p. (In Russian).
[11] **N. A. Salam**, **A. M. Zakaria**, **M. S. Samir**, "On the application of artificial neural network in analyzing and studying daily loads of Jordan power system plant," *Computer Science and Information Systems,* Vol.5, Issue 1, 2008, pp. 127-136.
[12] **J. R. Quinlan**, *C4.5: Programs for Machine Learning.* UK: Morgan Kaufmann publishers, 1993, 302 p.
[13] **A. Borisov**, **Y. Kornienko**, "A study of methods of classifier construction and updating," Automatic control and computer sciences, Vol.42, N6, 2008, pp. 300-305.
[14] **G. Das**, **K. Lin**, **H. Manilla**, **G. Renganathan**, **P. Smyth**, *Rule Discovery from Time Series. In Proceedings of the 3rd International Conference of Knowledge Discovery and Data Mining,* 1998, pp. 16-22.
[15] **A. Sennov**, *Access 2003. Practical working out of databases. The Training course.* St.Petersburg, 2005, 256 p. (In Russian).
[16] **F. Andon**, **V. Reznichenko**, *Language of inquiries SQL. The Training course.* St.Petersburg, Kiev, 2006, 416. p. (In Russian).
[17] **A. Sukovs**, **L. Aleksejeva**, **K. Makejeva** et al., *Data Mining: Fundamentals*. Riga: Riga Technical University, 2007, 130 p. (In Latvian).
[18] **D. C. Montgomery**, **C. L. Jennings**, **M. Kulachi**, Introduction to time series analysis and forecasting. Wiley-interscience, 2008, 472 p.
[19] **M. W. Craven**, **J. W. Shavlik**, "Extracting tree-structured representations of trained networks," *Advances in Neural Information Processing Systems,* Vol.8, MIT Press, 1996, pp. 24-30.
[20] **R. Kohavi**, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proceedings of the 14th International Conference on Artificial Intelligence (IJCAI-95),* San Mateo, CA: Morgan Kaufman, 1995, pp. 1137-1143.

Arnis Kirshners, Mg. sc. ing., Ph. D. Student and Senior Laboratory Assistant, Faculty of Computer Science and Information Technology, Riga Technical University, Department of Modelling and Simulation. He received his diploma from the Department of Modelling and Simulation at the Technical University of Riga. His research interests include data mining and knowledge extraction, bioinformatics, intelligent systems, programming, database and evolutionary computing. Previous publications: A. Kirshners, A. Sukov, "Rule induction for forecasting transition points in product life cycle data," *Scientific proceedings of Riga Technical University, Information Technology and Management Science,* Issue 5, Vol.36, RTU, Riga, 2008, pp. 170-177. A. Kirshners, Y. Kornienko, "Time-series data mining for e-service application analysis," *Scientific Proceedings of Riga Technical University, Information Technology and Management Science*, Issue 5, Vol. 40, RTU, Riga, 2009, pp. 94-100. A. Kirshners, S. Parshutin, A. Borisov, "Combining clustering and a decision tree classifier in a forecasting task," Automatic Control and Computer Sciences, Vol. 44, N 3, 2010, pp. 124-132. Contact information: Institute of Information Technology, Riga Technical University, 1 Kalku Street, Riga, LV-1658, Latvia, phone: +371 67089530, e-mail: arnis.kirsners@rtu.lv.

Galina Kuleshova is Research Scientist in the Faculty of Computer Science and Information Technology at Riga Technical University (Latvia). She received her M.Sc. degree in *Decision Support Systems* from Riga Technical University. Current research interests include artificial neural networks, classification methods and bioinformatics. Contact information: Institute of Information Technology, Riga Technical University, 1 Kalku Street, Riga, LV-1658, Latvia, phone: +371 67089530, e-mail: galina.kulesova@cs.rtu.lv.

Arkady Borisov is Professor of Computer Science in the Faculty of Computer Science and Information Technology at Riga Technical University. He holds a Doctor of Technical Sciences degree in Control in Technical Systems and the Dr. habil. sc. comp. degree. His research interests include fuzzy sets, fuzzy logic, computational intelligence and bioinformatics. He has 210 publications in the area. Contact information: Institute of Information Technology, Riga Technical University, 1 Kalku Street 1 Kalku Street, Riga, LV-1658, Latvia, phone: +371 67089530, e-mail: arkadijs.borisovs@cs.rtu.lv.

A. Kiršners, G. Kuļešova, A. Borisovs. Pieprasījumu prognozēšana balstoties uz īsām laika rindu kopām
Rakstā tiek risināts uzdevums, kas saistīts ar īsu vēsturisku laika rindu un diskrētu aprakstošo parametru apstrādi, ar mērķi veikt pieprasījuma prognozi, balstoties tikai uz jaunā produkta aprakstošajiem parametriem. Datu apstrādei tiek pielietotas tādas datu ieguves metodes, kā datu apkopošana, pirmapstrāde, klasteru analīze un klasifikācija. Datu sagatavošana datu ieguves procesiem tiek veikta pēc lietotāja definētiem un prognozēšanas sistēmā ievadītiem parametriem. Atlasītajās vēsturiskajās laika rindās ar klasterizācijas palīdzību nosaka to piederību kādai no klasēm, uz kuru bāzes tiek veidoti paraugmodeļi. Klasterizācija tiek veikta, izmantojot modificētu *k-vidējo* algoritmu, atrodot optimālo klasteru skaitu datu kopā, kas nepieciešams datu kopas klasterizācijai, vadoties pēc vidējās absolūtās klasterizācijas kļūdas. Klasifikācijas rezultātā, ar induktīvo lēmuma koku palīdzību, tiek noteikta saikne starp klasterizācijā iegūto paraugmodeli un preces diskrētajiem aprakstošajiem parametriem. Jauna produkta pieprasījuma prognozei, izmanto klasifikācijas rezultātā izveidoto lēmuma koku, uz kura projicējot jaunā produkta aprakstošos parametrus, atrod koka lapu, kas norāda uz klasterizācijas rezultātā iegūtā paraugmodeļa numuru. Šī paraugmodeļa līknes struktūra norāda uz iespējamo jaunā produkta pieprasījumu nākamajam periodam.

А. Киршнерс, Г. Кулешова, А. Борисов. Прогнозирование спроса на базе множества коротких временных рядов
В статье решается задача, связанная с обработкой коротких временных рядов и описывающих товар дискретных параметров с целью произвести прогноз спроса, основываясь только на описывающие новый продукт параметры. Для обработки используются такие методы добычи данных, как сбор и предобработка данных, кластерный анализ и классификация. Подготовка данных для процессов добычи данных происходит с учетом параметров, которые пользователь ввел в систему. С помощью кластерного анализа в отобранных коротких исторических временных рядах определяют принадлежность объекта к одному из классов, на основе которых создаются образы. Кластерный анализ произведен на базе модифицированного алгоритма *к-средних*, определяя в выборке данных оптимальное количество кластеров на основе средней абсолютной ошибки кластеризации. В результате классификации, используя индуктивные деревья решения, образуется связь между образом, полученным в процессе кластеризации, и дискретными параметрами, описывающими товар. При прогнозировании спроса на новый продукт используется полученное в процессе классификации индуктивее дерево решений, на котором проецируются параметры, описывающие новый продукт. В результате, получают лист дерева, который указывает на номер спроса, полученный в процессе кластеризации. Структура кривой полученного спроса указывает на возможный спрос на новый продукт в будущем.