DEVELOPMENT OF THE ALGORITHM OF DETERMINING THE STATE OF EVAPORATION STATION USING NEURAL NETWORKS

A. Ladanyuk
Doctor of Technical Sciences,
Professor, Head of Department*
E-mail: ladanyuk@ukr.net

V. Kyshenko
PhD, Professor*
E-mail: vdk.nuft@gmail.com

O. Shkolna
Postgraduate Student*
E-mail: evlens@ukr.net

M. Sych
Postgraduate Student*
E-mail: a.d.111@bk.ru

*Department of automation and
intelligent control systems
National University of Food Technologies
Vladimirska str., 68, Kyiv, Ukraine, 01033

Для розв'язання важливої задачі автоматичної оптимізації розподілу теплових ресурсів між технологічними ділянками цукрового заводу розроблений алгоритм оперативної оцінки стану випарної установки. Алгоритм включає самоорганізаційні карти Кохонена, метод оцінки якості кластеризації та метод нечіткої класифікації на основі нейронних мереж. В якості вхідних даних використовують часові ряди технологічних змінних цукрового виробництва. Алгоритм доцільно використовувати в автоматизованих системах керування випарною установкою цукрового заводу

Ключові слова: випарна установка, нейронні мережі, самоорганізаційні карти Кохонена, кластеризація, класифікація

Для решения важной задачи автоматической оптимизации распределения тепловых ресурсов между технологическими участками сахарного завода разработан алгоритм оперативной оценки состояния выпарной установки. Алгоритм включает самоорганизационные карты Кохонена, метод оценки качества кластеризации и метод нечеткой классификации на основе нейронных сетей. В качестве входных данных используют временные ряды технологических переменных сахарного производства. Алгоритм целесообразно использовать в автоматизированных системах управления испарительной установкой сахарного завода

Ключевые слова: испарительная установка, нейронные сети, самоорганизационные карты Кохонена, кластеризация, классификация

1. Introduction

Modern level of development of hardware and software tools enables the maintenance of databases of operative information at all levels of management. The widespread use of databases, in particular in industry, led to accumulation of large volumes of heterogeneous data that potentially contain useful analytical information. It can help to reveal hidden trends, build the strategy of development and find new solutions. As a result, there is a growing interest and a wide use of various methods of data analysis at different levels of production management in order to detect hidden analytical information.

Many methods of data analysis include intelligent data analysis (Data Mining), which, according to [1, 2], gives the possibility to convert data into information and then information into knowledge. This knowledge can be used, in particular, to improve automated control systems of the production processes. Data mining includes a set of functional modules for tasks such as associative and correlation analysis, problems of classification, forecasting, cluster analysis, emissions analysis, etc. [3].

The use of existing methods of data mining in decision support systems, which make a part of the automated control system of evaporation station (ES) of a sugar factory, will increase the efficiency of the ES operation. Given the fact that ES is a subsystem of the technological complex of a sugar factory [4], this will improve the performance and energy efficiency of the factory as a whole.

Under conditions of growing capacity of enterprises of sugar industry, the development of effective systems of resource saving control of technological objects with the use of the latest information technologies, including knowledge engineering, is a relevant task [5, 6].

Data mining methods are increasingly often used in various spheres of life but these methods are only starting to be used in automated control systems. Therefore, the development of the algorithm of determining the state of evaporation stations on the basis of intellectual methods is an important problem.

2. Literature review and problem statement

The key to the efficient operation of a sugar factory is the use of modern information technologies at the various levels of management [8] and constant improvement of automated

control systems [9]. The effective work of automated control system may not only ensure a proper course of technological process in certain areas of production but also lead to saving energy and material resources. It is in the context of resource saving that a special importance of evaporation stations among the production sections of a sugar factory is highlighted [10], that is why the work of evaporation stations is enhanced both through the improvement of the apparatus part [10, 11] and through modernization of the automated control system [13–15]. A decision support system is often distinguished as a part of an automated control system [16–18], which expands capacities of an automated system and increases the efficiency of its work.

When creating decision support systems for automated systems of controlling complex dynamic objects, there emerges a task of determining their state [7]. Change in the state of an object can be caused by both a change in the external environment and the change in the parameters of the object. Dynamic models are used to describe the state of dynamic control systems, both determinate and stochastic. The methods of analysis of dynamics of complex dynamic objects are probabilistic, statistical, determinate, fuzzy and neural models [14]. To determine the current state of an object, it is necessary to define in advance a set of possible states of an object. This can be done using heuristic methods, based on the experience and intuition of a developer or an expert, but during the analysis of complex multidimensional nonlinear objects, an expert may not always accurately analyze large volumes of information and track all the hidden patterns that lead to the need of generalization, simplification and other causes of the precision loss of a model. Given this, it is advisable to use data mining to determine the state of a control object taking into consideration the information hidden in data, which has not been done up to now.

Statistical methods are quite efficient and accessible but their use requires a significant amount of experimental data, which, taken together, will describe an object exactly enough, which is quite difficult for multidimensional objects.

The methods of fuzzy logic and neural methods offer the best opportunities for obtaining patterns with necessary accuracy. In particular, there is a variety of clustering methods [2, 3, 19], which include the Kohonen self-organizing maps (SOM), with the help of which it is possible to carry out automatic unsupervised clusterization on the basis of patterns hidden in existing data [4]. Due to their efficiency, these methods are used increasingly often in various areas of activity, therefore, it is worthwhile using them to determine the state of an object.

3. Aim and tasks of the study

The aim of the work is to develop an algorithm of determining the state of evaporation station of a sugar factory as a control object using the data mining methods and to consider the possibility of automation of its work.

To accomplish the aim, it is necessary to solve such tasks as:

– to conduct a preliminary analysis of time series of the evaporation station of a sugar factory and to define a set of parameters that will be used for further determination of the state of an object;

– to determine a set of possible states of the control object based on the time series of the selected parameters of the ES;

– to determine current state of an object from the set of the possible ones based on the current values of selected parameters of the ES.

4. Materials and methods for determining the state of evaporation station of a sugar factory as a control object

The work examines a four-case evaporation station with a concentrator, the typical scheme of which is shown in Fig. 1.
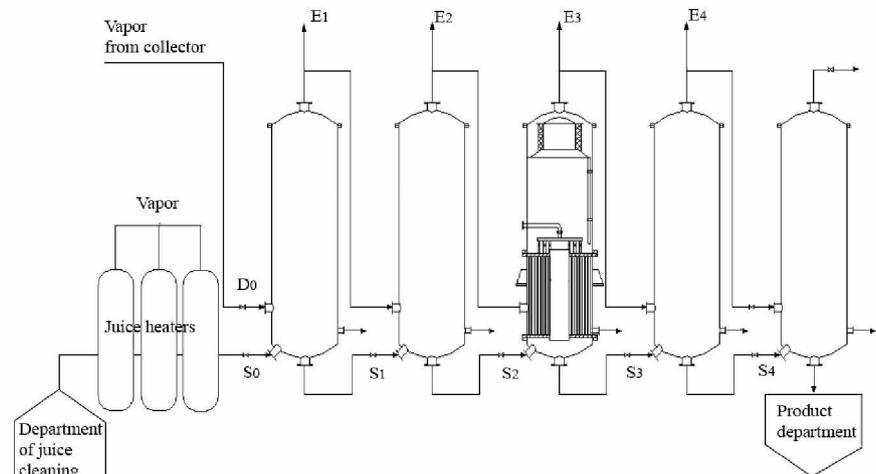

Fig. 1. Typical scheme of five-case evaporation station of sugar factory

The Kohonen self-organizing map is a neural network without feedback, in which unsupervised training algorithm is used [2]. As a result of self-organizing, SOM forms a topological representation of the source data of the neurons, obtained at the output. SOM can be trained to learn or find relationships between the inputs and outputs or organize data in such a way that will make it possible to detect previously unknown patterns or structures in them [3].

Self-organization algorithm of Kohonen provides the reflection of topology of space of great dimensionality on the neural maps that usually form a two-dimensional grid. Thus, the reflection of space of great dimensionality is formed on the plane. The property of preserving topology means that SOM distributes the similar vectors of the input data by neurons, i. e., the points located close to each other in the space of input, are displayed on the map on the neurons that are closely located. Thus, SOM may be used both as a means of clustering and as a means of visual representation of the data of large dimensionality [4].

4. 1. Preparation and data pre-processing to determine the set of states of the evaporation station with the use of the Kohonen self-organizing algorithm (SOM)

At modern sugar factories, in all sectors of production, the value of current controlled parameters from sensors and

controllers is gathered and stored in the programs of the SCADA types or in real-time archives, such as Proficy Historian. Changes in the values of parameters are stored in the form of time series. In this work, we used the data from the real-time archives Proficy Historian that preserves values of the controlled parameters of the evaporation station at a sugar factory with the capacity of 2500 tons of sugar beet per day. The evaporation station consists of four cases, which are evaporation stations of the Robert type, and a concentrator. The second case consists of two sequentially installed evaporation stations.

The parameters that were chosen for the analysis: juice consumption in the ES, syrup consumption at the outlet of the ES, the level in the juice collector, the level in case 1, the level in case 2A, the level in case 2B, the level in case 3A, the level in case 4, the level in case 3B, the level of the concentrator, the level in the syrup collector, pressure of the secondary steam in case 1, pressure of the secondary steam of case 2, pressure of return steam, temperature of the secondary steam of case 1, dilution in the concentrator, syrup density at the outlet of the ES, temperature of the secondary steam of case 2, temperature of the secondary steam of case 3, temperature of secondary steam of case 4, temperature of the secondary vapor of the concentrator, temperature of return vapor, juice temperature before ES, syrup temperature after the ES.

As it is known, the data, on the basis on which the training of neural network will be carried out, require pre-processing [7], which includes the following steps:

1. Encoding of input vectors for supplying the data, which contain only numeric values, to the input of neural network. Within the set task, all of these parameters have numeric values.

2. Data normalization. The vectors of input data may have a different scale. It is proposed to conduct normalization by the scale graded from zero to one but in this case it will be difficult to analyze results of clustering.

3. Data pre-processing. Removing obvious regularities from data makes it easier for the neural network to detect nontrivial patterns. Taking into consideration that it is not usually known in advance how useful these or other variables (components) that describe the input vectors may appear, the researcher may be tempted to increase the number of input parameters in hope that the network itself will determine which ones are the most important, but with an increase in dimensionality of the input vector, there is a decrease in the accuracy of forecasts, so we will perform a correlation analysis of the input data. We will use the method of searching for the maximum of inter-correlation function, which allows defining a linear dependence between the two processes, which, unlike the Pearson correlation, occur with a certain time lag (shift). Fig. 2 presents one of the results, which shows that the parameter "Temperature of the secondary vapor in the concentrator" correlates with the "Dilution in the concentrator" with the negative value of 0.965.

As a result of the performed correlation analysis, we can see that all the temperatures of secondary vapor in the cases of ES strongly correlate with the respective values of pressure of secondary vapor. To reduce dimensionality of the input vector, we will delete the parameters of secondary vapor temperature.

As a result of preparation and pre-processing of data, we will receive a specified list of the parameters selected for analysis (Table 1).
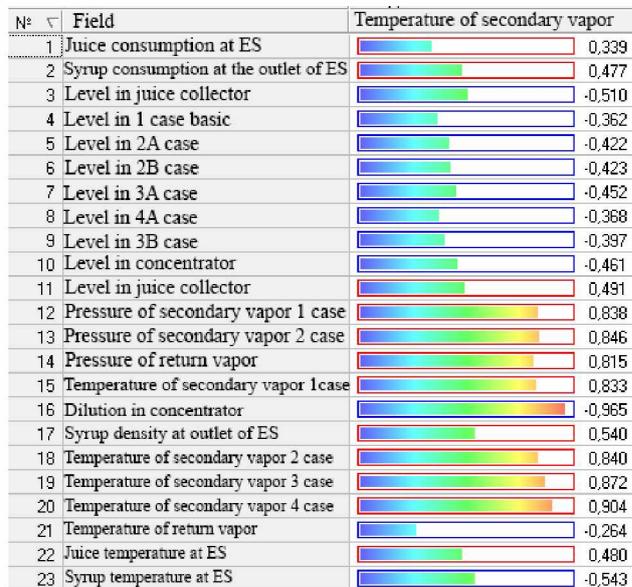
№	Field	Temperature of secondary vapor	
1	Juice consumption at ES		0,339
2	Syrup consumption at the outlet of ES		0,477
3	Level in juice collector		-0,510
4	Level in 1 case basic		-0,362
5	Level in 2A case		-0,422
6	Level in 2B case		-0,423
7	Level in 3A case		-0,452
8	Level in 4A case		-0,368
9	Level in 3B case		-0,397
10	Level in concentrator		-0,461
11	Level in juice collector		0,491
12	Pressure of secondary vapor 1 case		0,838
13	Pressure of secondary vapor 2 case		0,846
14	Pressure of return vapor		0,815
15	Temperature of secondary vapor 1case		0,833
16	Dilution in concentrator		-0,965
17	Syrup density at outlet of ES		0,540
18	Temperature of secondary vapor 2 case		0,840
19	Temperature of secondary vapor 3 case		0,872
20	Temperature of secondary vapor 4 case		0,904
21	Temperature of return vapor		-0,264
22	Juice temperature at ES		0,480
23	Syrup temperature at ES		-0,543

Fig. 2. Result of search for the maximum of inter-correlation function for the parameter "Temperature of secondary vapor of concentrator»

Table 1

Specified list of parameters selected for analysis

No.	Parameters	Average value	Units of measurement
1	Juice consumption in ES	152,9	m^3/hour
2	Syrup consumption at the outlet of ES	36,9	m^3/hour
3	Level in juice collector	49,1	%
4	Level in case 1	39,6	%
5	Level in case 2A	44,1	%
6	Level in case 2B	44,8	%
7	Level in case 3A	67,3	%
8	Level in case 4	57,0	%
9	Level in case 3B	49,6	%
10	Level in concentrator	19,9	%
11	Level in syrup collector	35,3	%
12	Pressure of secondary vapor of case 1	124,1	kPa
13	Pressure of secondary vapor of case 2	93,07	kPa
14	Pressure of return vapor	166,8	kPa
15	Delusion in concentrator	39,4	%
16	Syrup density at the outlet of ES	65,1	%
17	Juice temperature in ES	124,01	°C
18	Syrup temperature after ES	85,9	°C

Before starting a neural network analysis, it is worthwhile clearing the time series of input data from emissions in the real time archives Proficy Historian, where each value is compared to the permissible range of each parameter.

4. 2. Clustering with the use of SOM of the prepared sets of input data

Clustering of the selected data with the use of SOM in the software product Deductor Studio was performed. The number of entries in the data array is 32000. Each entry

was linked to time, the interval between entries is 1 second, generally such data array covers a 9-hour working period of the evaporation station.

Clustering with a training period of 500 epochs was conducted, with the grid size of output neurons of 16×12 and automatic determination of the number of clusters. During training, one can observe the change in the value of quantization error (Fig. 3):

From the start of training and up to the 180th epoch, values of the maximum and mean errors were constantly changing, and starting with the 200th epoch, their value was almost unchanged. This means that for training the network, 200 epochs are enough.

As a result, we obtain a trained neural network with 6 clusters (Fig. 4, a, b).

The boundaries of clusters passed mostly on the cells, the distance from which to their neighbors is the longest (Fig. 4, a). As it was already noted, the neighboring values of the input data space fall into the nearby cells of two-dimensional grid but the distance between them is reflected only on the matrix of distances, where the value of distance from the smallest to the largest corresponds to the color of cells from blue to red.

A 10-fold clustering with the change in setting parameters of training was carried out. The number of training epochs changed from 200 to 500 and the grid size of the output layer of neurons ranged from 16×12 (192 cells) to 25×22 (550 cells). The number of clusters was determined automat-

ically. The largest number of clusters (20) was obtained with the grid size of 25×22 (Fig. 5, a, b).

As a result, 10 variants of splitting a set of data into clusters were obtained. Afterwards, it is necessary to determine one best clustering option.
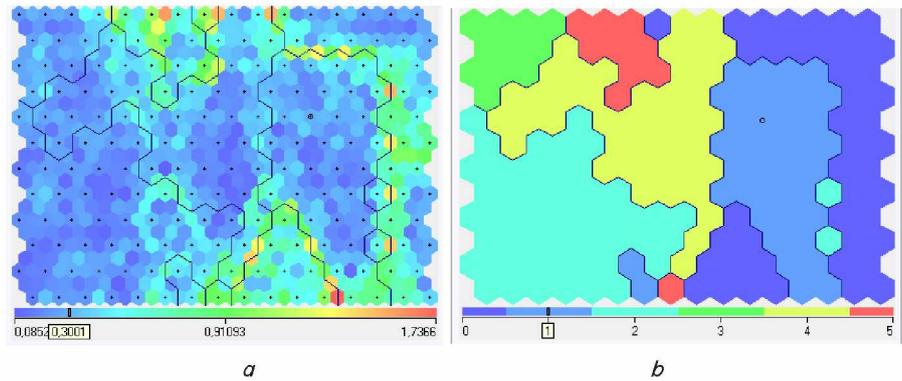

a   *b*

Fig. 4. Result of clustering (number of epochs — 500; size of the output grid of neurons — 16×12; automatic determination of the number of clusters): *a* — matrix of distances where the range of changes in color from blue to red corresponds to the value of the Euclidean distance from cell to its nearest neighbors (0.085 — blue, 1.736 — red); *b* — map of splitting into 6 clusters with numbers from 0 to 5, each of which is highlighted with a certain color
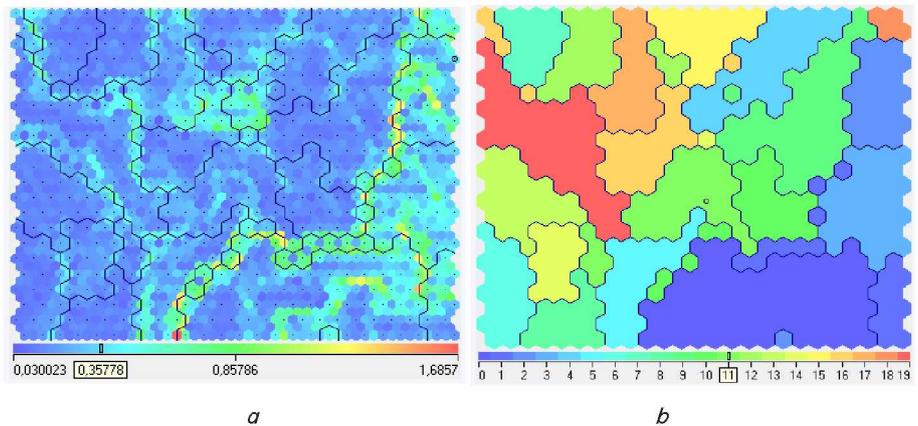

a   *b*

Fig. 5. Result of clustering (number of epochs — 200; size of the output grid of neurons — 25×22; automatic determination of the number of clusters): *a* — matrix of distances; *b* — map of splitting into clusters
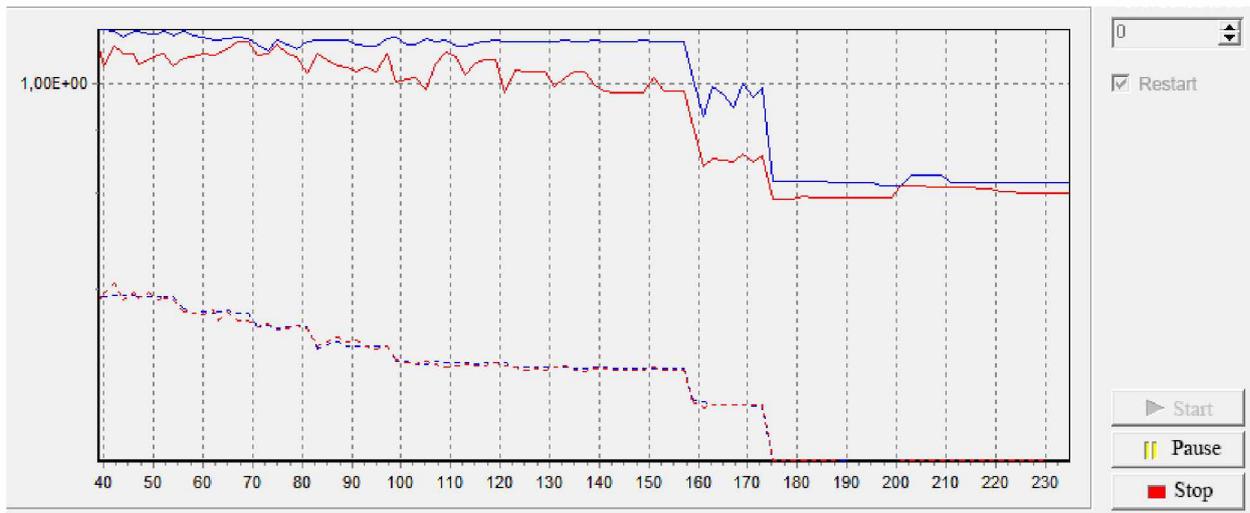

Fig. 3. Diagram of change in the maximum and mean errors in the training of the testing and learning sets

4. 3. Determination of the best clustering option

Having obtained 10 clustering options with a different number and form of clusters, it is necessary to choose the best option. To do this, we will use one of the methods of assessment of quality of clear clustering, namely, the Silhouette index [8].

For the element x_j, which belongs to cluster c_p, the mean value of the distance from it to the elements of the same cluster is a_{pj}, and the mean distance from it to the elements of another cluster c_q will be denoted as d_{qj}, then the minimum value among all d_{qj} will be denoted as b_{pj}. Then the "silhouette" of every single item is defined as:

$$S_{xj} = \frac{b_{pj} - a_{pj}}{\max(a_{pj}, b_{pj})}. \qquad (1)$$

Thus, a high value of s_{xj} characterizes the "best" belonging of the element x_j to the cluster c_p. Assessment of the cluster structure is achieved by the mean value of the silhouette by each of the elements:

$$SWC = \frac{1}{N} \sum_{j=1}^{N} S_{xj}, \qquad (2)$$

where N is the number of elements.

Having determined a silhouette index for each of 10 clustering options, we choose the option with the highest value of the index. This is the clustering variant that was obtained as a result of training the network with the following parameters: the number of epochs – 200; grid size of the output neurons – 18×14; automatic determination of the number of clusters. Visualization of the clustering result is shown in Fig. 6, a–d.

The set of data was split into 8 clusters. Quantization error is negligible. Basically, the boundaries of clusters pass through the cells with significant distances between the values that got in the adjacent cells. Fig. 5 shows that the 3rd cluster basically defined the parameter "Syrup density", since the lowest values of this parameter got in this cluster. The parameter "Pressure of secondary vapor" influenced distinguishing cluster number five, as it is here that the lowest values are focused. The map "Juice consumption" shows that its value varied mainly in the range of 100–166 m^3/h, which corresponds to colors from green to orange, but there is one red cell numbered 124, where there are entries with the juice consumption value from 166 m^3/h to 202 m^3/h. This cell contains 33 entries. Blue cell contains 82 entries, the range of change in the parameter "Juice consumption" is from 39 m^3/h to 100 m^3/h.
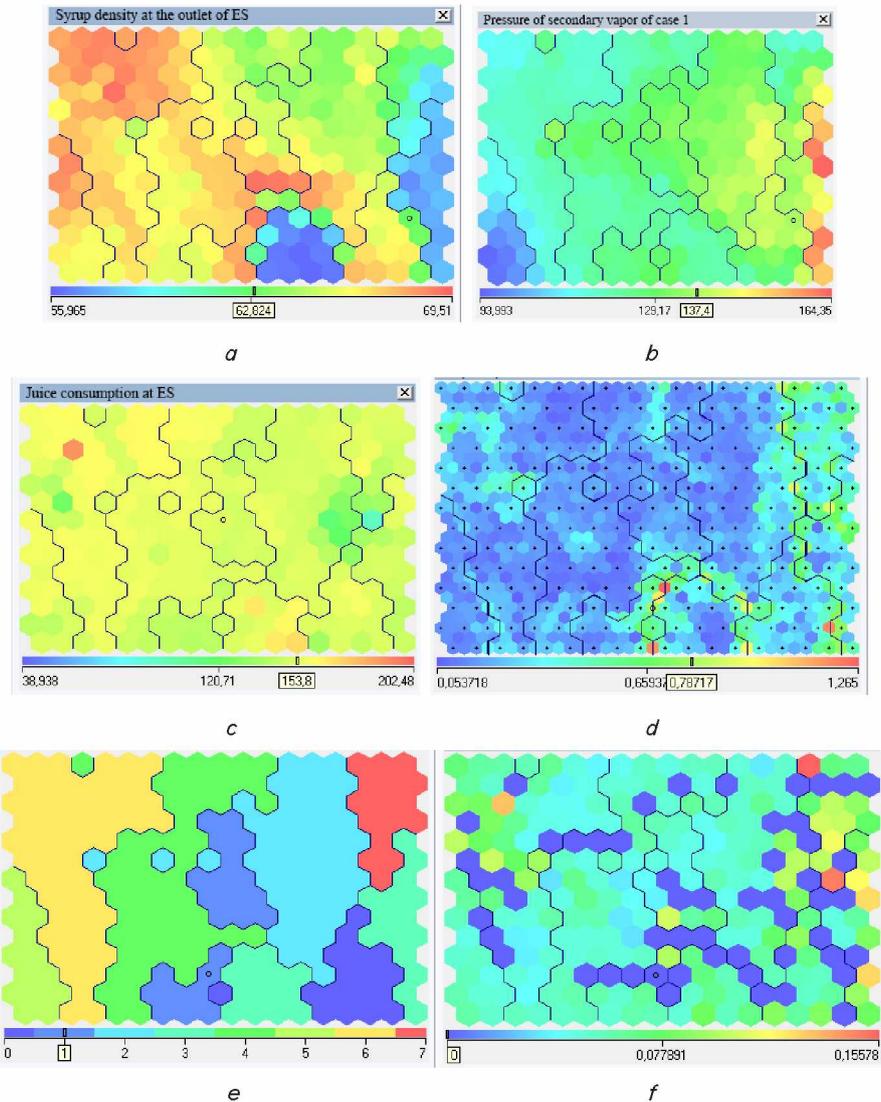

Fig. 6. The best result of clustering by silhouette index value: a — map of distribution of the parameter "Syrup density at the outlet of ES; b — map of distribution of the parameter "Pressure of secondary vapor of case 1 of ES; c — map of distribution of the parameter "Juice consumption in ES"; d — matrix of distances; e — map of splitting into clusters; f — matrix of quantization errors

Let us consider a fragment of the table of clusters profiles (Fig. 7).

The number of entries that each of the clusters contains in a numeric value and in percentage of the total number of input entries can be seen in Fig. 6. The largest is the fourth cluster, which contains 7100 entries, which corresponds to 22.3 %. The smallest is the seventh cluster of 1768 entries. The level of significance of six parameters displayed in the table of parameters in each of the clusters is 100 %.

Let us consider the diagram of location, which shows the dependence of values of one field on the other two. It allows assessing visually the dependence, which is represented in the form of points in a multi-dimensional space. The color and size of dots are additionally informative (Fig. 8).

The dependence of values of juice consumption on syrup density and pressure of secondary vapor of case 1 is shown in the diagram of location (Fig. 8). The colors of the points correspond to the numbers of clusters.

	4	6	2	3	1	5	0	7	Total
– + Field ▼	7100 (22,3%)	6694 (21,0%)	5141 (16,2%)	3416 (10,7%)	3364 (10,6%)	2476 (7,8%)	1855 (5,8%)	1768 (5,6%)	
⊞ **9.0** Level in 2B case %	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
⊞ **9.0** Level in 2A case %	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
⊞ **9.0** Level in 1 case basic %	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
⊞ **9.0** Level in juice collector %	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
⊞ **9.0** Syrup consumption at outlet of ES	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
⊞ **9.0** Juice consumption at ES	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

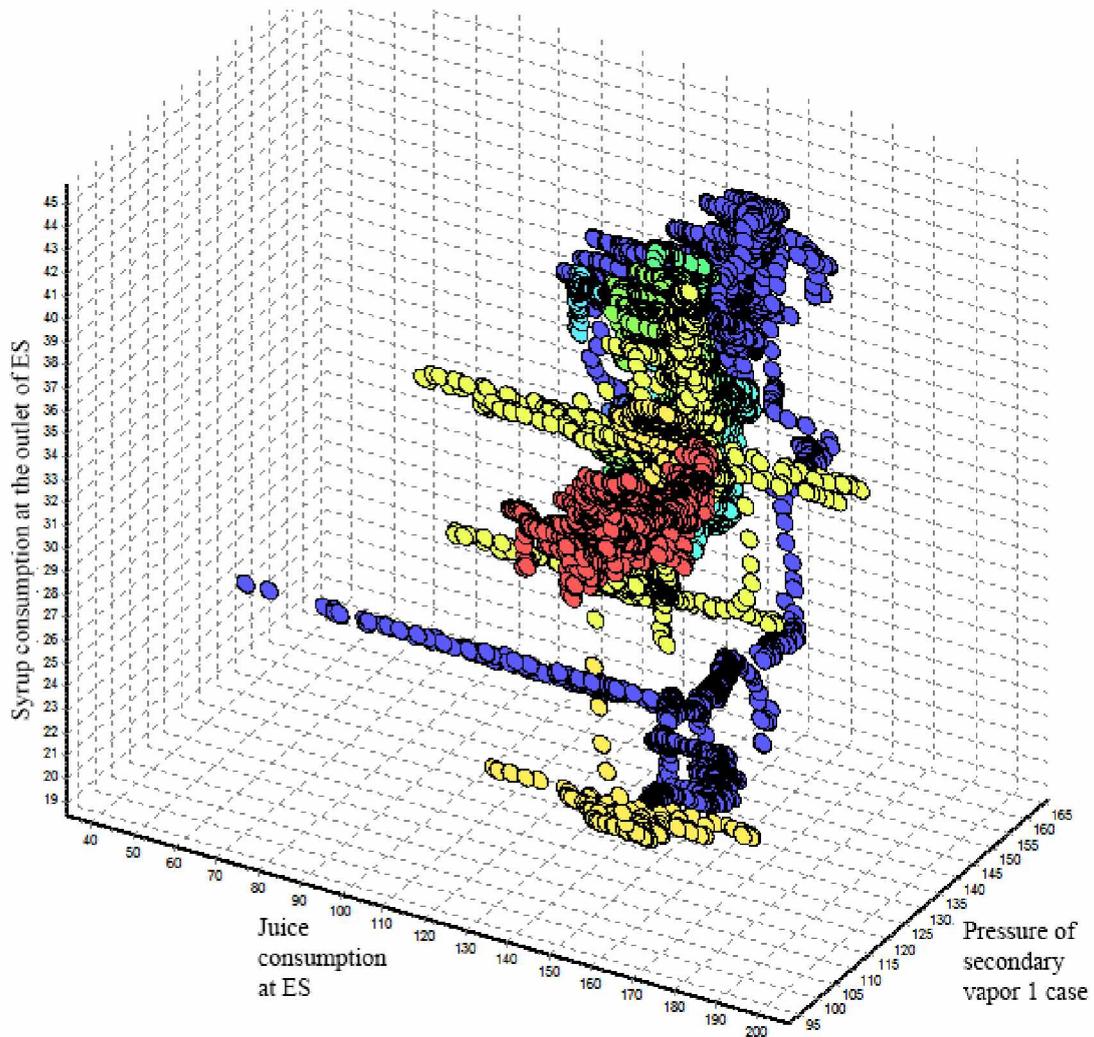Fig 7. Fragment of table of clusters profiles

Fig. 8. Diagram of location of entries in three-dimensional space
(entries that belong to one cluster are highlighted by the same color)

4. 4. Classification of the current state of evaporation plant

As a result of the performed clustering, a set of clusters, each of which corresponds to a certain state of a control object, was obtained. To determine the current state of the object, it is necessary to define to which cluster the entry, which is defined by a set of the current parameters listed in Table 1, belongs. Within the framework of the set task, it is necessary to perform fuzzy classification using neural networks. It is necessary to train the neural network, in which the input continuous data will be the values of parameters in Table 1, and categorical target values will be the numbers of clusters.

As a result of training the neural network, we will obtain the table of probabilities of belonging of each observation to a specific class (Fig. 9).

We can see that each of the observations has a numeric value of probability of belonging to a particular class. In the given example, the observations from 1958 to 1970 with an average probability of 0.9 belong to class number three.

Let us consider the table of sensitivity analysis, i. e., significance of each of the variables, Fig. 10.

In this case we can see that the parameters "Dilution in the concentrator", "Syrup density at the outlet", "Pressure of secondary vapor in case 2" are the most important. So their values are more significant during classification.

Using a 3D diagram, the scale of which is graded according to the values of probability of belonging to the fourth, second and sixth classes, let us consider distribution of the observations by different probabilities of belonging to a particular class (Fig. 11).
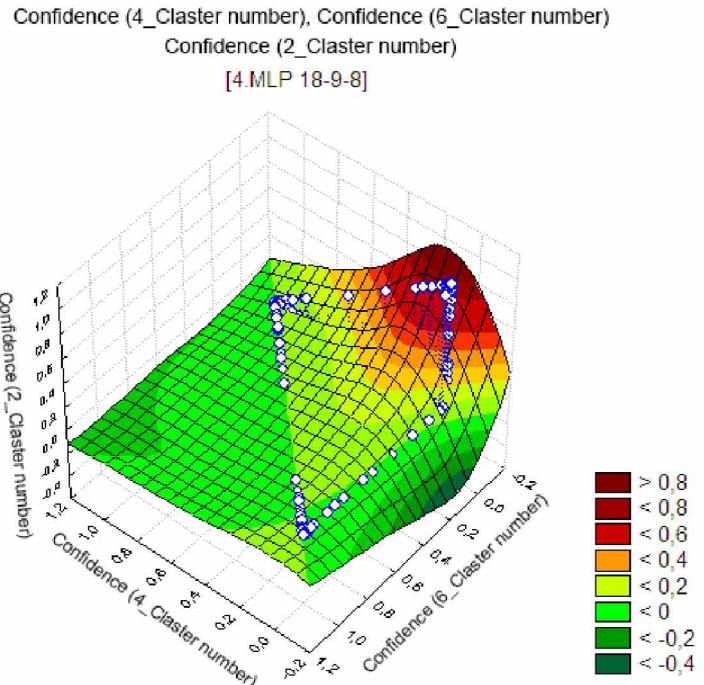

Fig. 11. Distribution of observations by different probabilities of belonging to each of the three largest classes (4th, 6th and 2nd classes)

The probability of belonging of each observation to each of the three largest classes (4th, 6th and 2nd classes) may be seen from the diagram where points correspond to the observations and the scales are graded from – 0.4 to 1.2 of probability of belonging of the observation to a specific class.

Case name	Sample	Cluster number - 0 9.RBF 18-8-8	Cluster number - 1 9.RBF 18-8-8	Cluster number - 2 9.RBF 18-8-8	Cluster number - 3 9.RBF 18-8-8	Cluster number - 4 9.RBF 18-8-8	Cluster number - 5 9.RBF 18-8-8	Cluster number - 6 9.RBF 18-8-8	Cluster number - 7 9.RBF 18-8-8
1958	Train	0,036916	0,000003	0,000002	0,910320	0,000000	0,000031	0,000030	0,052698
1959	Train	0,036916	0,000003	0,000002	0,910320	0,000000	0,000031	0,000030	0,052698
1960	Train	0,036850	0,000003	0,000002	0,910325	0,000000	0,000031	0,000030	0,052759
1961	Train	0,040233	0,000004	0,000003	0,905941	0,000000	0,000035	0,000035	0,053749
1962	Train	0,042177	0,000004	0,000003	0,902995	0,000000	0,000039	0,000038	0,054745
1963	Test	0,039171	0,000004	0,000002	0,907044	0,000000	0,000034	0,000033	0,053712
1964	Train	0,039725	0,000004	0,000002	0,906607	0,000000	0,000035	0,000034	0,053593
1965	Train	0,038910	0,000004	0,000002	0,908067	0,000000	0,000033	0,000033	0,052951
1966	Train	0,040135	0,000004	0,000002	0,906303	0,000000	0,000035	0,000034	0,053487
1967	Test	0,036569	0,000003	0,000002	0,911011	0,000000	0,000030	0,000029	0,052355
1968	Train	0,034821	0,000003	0,000002	0,913494	0,000000	0,000027	0,000026	0,051626
1969	Train	0,036625	0,000003	0,000002	0,910471	0,000000	0,000030	0,000029	0,052840
1970	Train	0,038777	0,000003	0,000002	0,908175	0,000000	0,000033	0,000032	0,052977

Confidence levels (export)
Samples: Train, Test

Fig. 9. Fragment of the table of probabilities of belonging of each observation to each class

Sensitivity analysis for **Cluster number** (export)
Samples: Train, Test

	Level in 2A case %	Level in 2B case %	Level in 3A case %	Level in 4 case %	Level in 3B case %	Level in concent-rator %	Level in juice collector %	Pressure of secondary vapor 1 case kPa	Pressure of secondary vapor 2 case kPa	Pressure of return vapor kPa	Dilution in con-centrator %	Syrup density at the outlet of ES %	Juice tempera-ture at ES C	Syrup tempera-ture after ES C
Netwo 9.RBF	1,219243	1,098496	1,001939	0,994350	0,982997	6,658526	1,152653	1,507903	2,036989	1,214410	1,096555	2,085137	1,363699	1,498933

Fig. 10. Values of sensitivity of parameters

Given the fact that we have more classes than are presented in the diagram, some points do not belong to the distribution plane.

Using the trained neural network, it is possible to classify the current state of the control object. To do this, it is necessary to set the current values of the input parameters and obtain probability of belonging of this observation to one of the classes.

5. Results of development of an algorithm of determining the state of evaporation station of a sugar factory

Using the above mentioned methods, the possibility to determine the state of an object by the following algorithm was shown:

1. Preparation and pre-processing the data of ES at a sugar factory.

2. Clustering with the use of SOM in order to determine possible options of the states of a control object.

3. Determination of the best splitting option by the silhouette index.

4. Based on the best clustering option, to train a neural network with the aim of fuzzy classification of possible states of the object.

5. Classification of the current state of an object based on the values of the input parameters and trained neural network.

6. Discussion of results of development of an algorithm of determining the state of evaporation plant at a sugar factory

To use such methods in DSS, it is necessary to develop the software that will ensure implementation of this algorithm and provide a user-friendly interface for displaying the results of determining current state of an object for a person who makes decisions (PMD). With this purpose, it is possible to use any programming language but then the implementation of algorithms of neural data analysis will require large amounts of time and significantly increase complexity of the task. Therefore, we propose to use the R programming language for data mining.

R is a programming language for statistical data processing and work with graphics, as well as free software environment with an open source code. R-scripts are simple to use in automation and to integrate in the industrial systems. R is supported by such software packages for statistic data processing as: Mathematica, MATLAB, STATISTICA, Oracle R Enterprise and SQL Server. Using the Python programming language, it is possible to provide an access to the R functions using the RPy package. Therefore, it is better to implement DSS with the help of the programming language Python with integrated R to solve the problems of data mining.

It should be noted that clustering and training a neural network for classification in DSS (2–4 steps of the algorithm) are defined once in a certain period, and the result, i. e., the trained neural network for fuzzy classification will be used repeatedly to determine the current state of an object (step 5). If the current state of an object is impossible to classify, as it does not belong to any of the classes, one must repeat steps 2–4, adding unrecognized observations to the array of input data.

As a result of the conducted research, we obtained the algorithm of determining the current state of a control object, taking the evaporation station of the sugar factory as an example. The set of possible states of evaporation station of a sugar factory as a control object was determined by using the Kohonen self-organizing algorithm, besides, the neural fuzzy method of classification of current state of the object was used according to the obtained clustering results. With the help of the programming languages Python and R, the work of the developed algorithm of determining the set of states and the current state of a control object was automated.

7. Conclusions

As a result of the conducted research, the algorithm of defining the state of the evaporation station at sugar factory as a control object on the basis of neural networks was developed and the following problems were solved.

1. Preliminary analysis of time series of evaporation station of sugar factory by removing obvious patterns and emissions was carried out. From a set of monitored parameters of the evaporation station, the most important ones were selected and the "redundant" settings were removed. When identifying the pairs of parameters strongly correlated to each other, i.e., having the value of inter-correlation function close to one, one of them was removed from the set. As a result, a list of the most important parameters was obtained.

2. By using the Kohonen self-organizing maps, several variants of splitting into clusters were obtained and the best clustering option was defined with the help of the silhouette index. As a result of clustering, we obtained a set of possible states of ES, where each cluster corresponds to a specific possible state of the object and is characterized by a specific range of change in the values of each of the parameters.

3. The possibility of determining the state of an object in the current moment of time was implemented by using the method of fuzzy classification with a trainer on the basis of neural networks. The current values of the meaningful parameters and the set of possible states of ES obtained by clustering were used as the input data.

References

1. Pivnyak, G. G. Information Technology Glossary [Text] / G. G. Pivnyak, B. S. Busygin, M. N. Divizinyuk. – Dnipropetrovsk, 2010. – 600 p.

2. Mohammed, J. Z. Data Mining And Analysis [Text] / J. Z. Mohammed, M. Jr. Wagner. – New York: USA, 2014. – 607 p.

3. Han, J. Data mining: Concepts and techniques [Text] / J. Han, B. Kamber. – San Francisco:USA, 2006. – 743 p.

4. Karim, Md. E. Fuzzy Clustering Analysis [Text] / Md. E. Karim, F. Yun. – Karlskrona: Sweden, 2010. – 63 p.

5. Pryadko, M. O. Heat Technology Fundamental of Sugar Production [Text] / M. O. Pryadko, M. O. Maslikov, V. P. Petrenko, V. I. Pavlenko, V. M. Filonenko. – Kyiv, 2007. – 296 p.

6. Medida, S. Pocket Guide on Industrial Automation For Engineers and Technicians [Text] / S. Medida. – Austin: USA, 2007. – 172 p.

7. Ladanyuk, A. P. System Analysis of Complex Control Systems [Text] / A. P. Ladanyuk, Y. V. Smityuh, L. A. Vlasenko, N. A. Zaets, I. V. Elperin. – Kyiv, 2013. – 274 p.

8. Yadav, U. J. Problems and Prospects of IT Implementation in Sugar Factory [Text] / U. J. Yadav, B. S. Sawant // International Journal of Advanced Research in Computer Science and Software Engineering. – 2012. – Vol. 2, Issue 8. – P. 453–466.

9. Langhans, B. Crystallization – a central competence, the key to success [Text] / B. Langhans // International Sugar Journal. – 2004. – Vol. 106, Issue 1265. – P. 266–268.

10. Adriano, V. E. Design of Evaporation Systems and Heaters Networks in Sugar Cane Factories Using a Thermoeconomic Optimization Procedure [Text] / V. E. Adriano, A. N. Silvia // Int. J. of Thermodynamics. – 2007. – Vol. 10, Issue 3. – P. 97–105.

11. Lehnberger, A. Falling-film evaporator plant for a cane sugar factory: Presentation of the concept and operating results [Electronic resource] / A. Lehnberger, F. Brahim, S. S. Mallikarjun. – United Kingdom, 2014. – Available at: https://www.bma-worldwide.com/fileadmin/_migrated/content_uploads/ISJ_2014_BMA-Evaporator.pdf

12. Mushiri, T. To Design and Implement a Reliable Sugar Evaporation Control System that will Work in an Energy Saving Way [Text] / T. Mushiri, Ch. Mbohwa // Proceedings of the World Congress on Engineering, 2015. – P. 326–331.

13. Vlasenko, L. O. Improving the Efficiency of the Evaporation Station Sugar Factory Through the Use of Statistical Methods Diagnosis [Text] / L. O. Vlasenko, M. A. Sych // Proceedings of International Scientific Conference "New Ideas in Food Science – New Products of Food Industry", 2014. – 259 p.

14. Ladanyuk, A. P. Control of Evaporation Station Under Uncertainty: Intellectualisation of Application Functions [Text] / A. P. Ladanyuk, V. D. Kyshenko, O. V. Shkolna // Proceedings of the National University of Food Technologies. – 2015. – Vol. 6. – P. 7–14.

15. Zaets, N. A. Modeling evaporation process for the synthesis of automatic control system [Text] / N. A. Zaets, N. M. Lutska // Scientific Journal of the National University of Life and Environmental Sciences of Ukraine. – 2011. – Vol. 161. – P. 180–186.

16. Lyne, P. W. L. Decision Support Systems For Sugarcane Production Managers [Text] / P. W. L. Lyne // Proc S Afr Sug Technol Ass. – 2012. – Vol. 85. – P. 206–220.

17. Rozman, C., The Development of Sugar Beet Production and Processing Simulation Model – a System Dynamics Approach to Support Decision-Making Processes [Text] / C. Rozman, A. Skraba, K. Pazek, M. Kljajic // Organizacija. – 2014. – Vol. 47, Issue 2. – P. 99–105. doi: 10.2478/orga-2014-0011

18. Barons, M. J. Dynamic Bayesian Networks for decision support and sugar food security [Electronic resource] / M. J. Barons, X. Zhong, J. Q. Smith // United Kingdom. – 2014. – Available at: http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/paper14-18/14-18w.pdf

19. Rodriguez, A. Clustering by fast search and find of density peaks [Text] / A. Rodriguez, A. Laio // Science. – 2014. – Vol. 344, Issue 6191. – P. 1492–1496. doi: 10.1126/science.1242072

20. Manzhula, V. G. Kohonen Neural Network and Fuzzy Neural Network in the Data Intelligent Analysis [Text] / V. G. Manzhula, V. G. Fedyashov // Basic Research. – 2011. – Vol. 4. – P. 108–114.

21. Sivogolovko, E. V. Assessing Quality Methods of the Distinct Clustering [Text] / E. V. Sivogolovko // Computer Tools in Education. – 2011. – Vol. 4. – P. 14–31.