Intelligent information systems

UDC 004.855 doi: 10.20998/2522-9052.2020.4.14

Svitlana Krepych, Iryna Spivak

West Ukrainian National University, Ternopil, Ukraine

FORECASTING SYSTEM OF UTILITIES SERVICE COSTS BASED ON NEURAL NETWORK

Abstract. The work is devoted to the problem of excessive spending of people's funds on utilities, especially in winter, when these costs can amount to more than 25% of the family budget. The question of the possibility of saving at least part of these costs by monitoring their possible value and reducing this indicator is an urgent task. Hence, the development of a software system for forecasting utility costs is an urgent practical task. To solve this problem, the authors propose to use a neural network, because it is advisable to use it in situations where there is predetermined known information and on its basis the user needs to get the predicted new information. The method for forecasting utility costs based on the use of a neural network takes into account user's data of utility service costs entered manually or obtained from the EPS system, which is convenient because you can immediately get a large set of input data to more accurately predict future costs. Another type of input data is data obtained from weather forecast sites to determine forecast indicators for correct the training of neural network. Based on these data, the network studies and builds a separate model for forecasting utility costs for each user. Considering that the data on utility service costs entered by users into the system each month may not match the date, it is proposed to take into account this inaccuracy, to given the input data for forecasting as an interval corridor of values which containing the minimum and maximum forecast limits. The developed software system and the method of forecasting utility service costs were tested on the example of a real user of the EPS system.

Keywords: utilities cost; forecasting system; neural network; interval data; estimation accuracy.

Introduction

Nowadays there is an acute problem in the country of rising prices for utilities, especially for heating. Various studies and statistical samples show that the average citizen has to spend a large percentage of income to cover utility costs, which especially in winter, often is about 30-50% of family income. Together with the dynamics of rising prices for food, fuel and other goods, this factor greatly affects the standard of living in the country [1]. And it is possible only approximately predict the amount of costs, based on the season and previous costs. Therefore, the actuality task is to forecasting the most accurate amounts of utility service costs for the current or next month, based on weather indicators, to save and planning family budgets and expenses [2]. The aim of the work is to develop a software system for forecasting utility service costs using a neural network. To achieve the goal of the research it is necessary to solve the following tasks: to determine the main factors of utility service costs, to group them according to the indicators on which they have a certain impact; to explore existing tools for the development of neural networks; to develop a neural network that will take into account factors and predict the final result of utility costs; to test the network on a randomly generated data set, and to compare the result with real data; to develop a software system for forecasting utility service costs using a neural network.

Statement of task

As mentioned above, the percentage of utility costs from the total family income is a significant part (Fig. 1) [3], which in winter is at least a quarter or even a third of family earnings. For many families, especially those who rent an apartment, the question of the possibility of saving part of the budget is an urgent task. The monthly expenses for the family's utilities service consist of paying for the consumption of water, gas, electricity, as well as paying for cable TV and the maintenance of the adjacent territory.

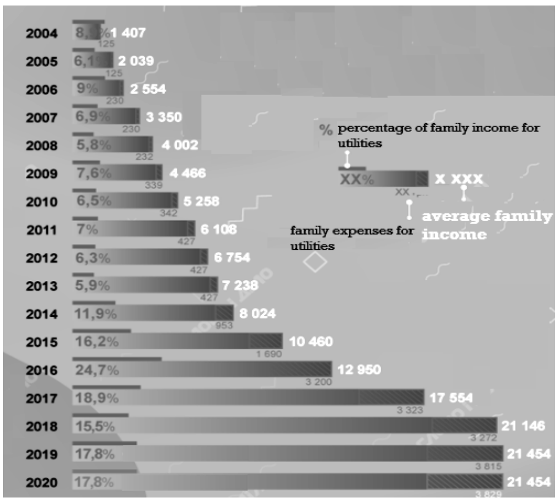

Fig. 1. Average utility costs for a family of 2 adults and 1 child for a 2-room apartment of 20 m^2.

The last two types of payments are stable and relatively unchanged, so they will not be considered in the system of forecasting utility costs.

Estimating the main sources of costs for heating, we get, that the most important factor influencing these indicators is weather conditions and season, and to understand the main sources of electricity costs, we need to analyze the main costs of electricity of appliances (Table 1).

Using the data in the table we can to calculate the approximate consumption of electricity of household appliances and to analyze the average cost of electricity:

Fluorescent lamp 4 pcs · 100 W · 8 h · 30 d = 96 kW/m;
laptop 1 pcs · 50 W · 5 h · 30 d= 7.5 kW/m;
TV 1 pcs · 100 W · 4 h · 30 d = 12 kW/m;
Microwave oven 1 pcs * 600 W * 15 h/m = 9 kW/m;
Iron 1 pcs · 1000 W · 3 h/m = 3 kW;
Refrigerator 1 pcs · 30 W · 24 h · 30 d = 21.6 kW;
Washing machine 1 pcs · 2000 W · 10 h/m = 20 kW;
Electric kettle 1 pcs · 1500 W · 8 h/m = 12 kW.

Therefore, as we can see from the estimated data, the overall average value for the month only for these appliances is already amount 181.1 kW.

Table 1 – **Electricity consumption by appliances**

Appliance	Power, W	Appliance	Power, W	Appliance	Power, W
Coffee grinder	200	Razor	15	Blender	300
Coffee maker	800	laptop	20-80	Toaster	800-1500
Fluorescent lamp 60 W	16	PC	100-500	Fluorescent lamp 75 W	20
Fluorescent lamp 40 W	11	Inkjet printer	100	Fluorescent lamp 100 W	30
Microwave oven	600-2000	Electric stove	4000-6000	Electric clock	10
Automatic washing machine	2000	TV 25”	150	Gas drying for things	300
Hand washing machine	300	TV 19”	70	TV 12”	30
Vacuum cleaner	700-2000	Pump	75-400	Iron	1000-2000
Handheld vacuum cleaner	100	Satellite dish	5	Laser printer	600
Shaver	100	Electric dryer	400-2000	Electric oven	3000
Halogen lamp 40 W	40	Compact fluorescent lamp 25 W	28	Compact fluorescent lamp 20W	22
Large refrigerator with freezer	540	Medium refrigerator with freezer	475	Electric bell	10
Electric blanket	200-500	Drying	1000	Hair dryer	1200-2000

The class of tasks that can be solved using a neural network is determined by the actual operation of the network and its training [4]. Thus, it is advisable to use the network in situations where there is predetermined known information and on its basis the user needs to get the predicted new information. Taking into account that the information on utility service costs for a certain period is available and it is enough to train the neural network, it is advisable to use this data to design a software system that will predict the cost of utilities for the next few months.

Main part

The input parameters of the system are data of weather factors, as well as user data. Based on this, a complete set of data for analysis is formed. This information is transmitted to the system and processed by the neural network, which is a controlling element of the system. Additional data coming into the system is the weather history for a given locality, the history of payments that have already been processed by the neural network. This information is needed to obtain the most accurate prediction of the costs that the system returns at the output. The diagram of variants of software system using is presented in Fig. 2.

To take into account the inaccuracy of the user's submission of monthly data on a basis with strictly the same calendar day, it is advisable to submit the input indicator of a particular utility service in an interval form [5-7]:

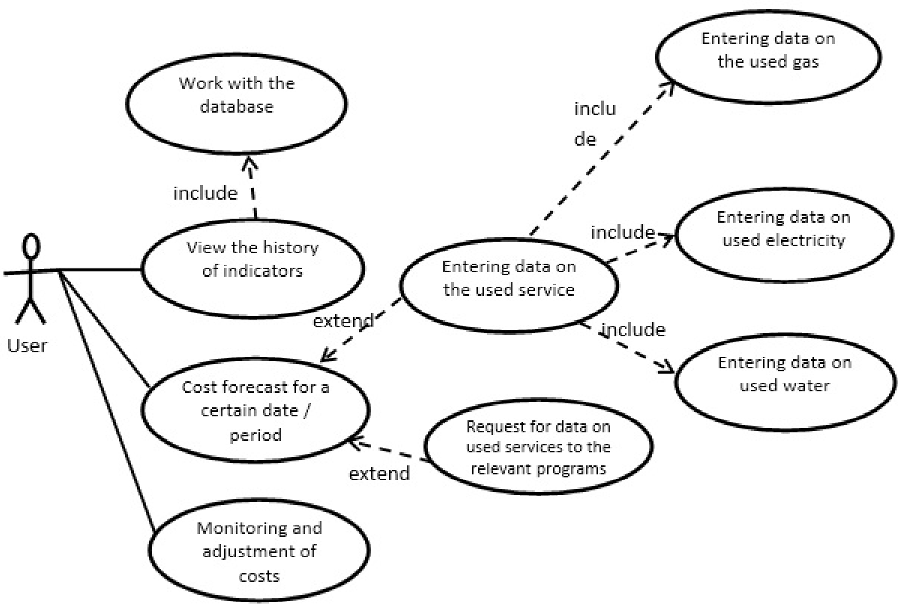

Fig. 2. Use–case system diagram

$$\left\{ I_g \in [I_g^-; I_g^+];\ I_l \in [I_l^-; I_l^+];\ I_w \in [I_w^-; I_w^+], \right. \quad (1)$$

$I_g^- = I_g - \delta \cdot I_g; I_g^+ = I_g + \delta \cdot I_g$; $I_l^- = I_l - \delta \cdot I_l; I_l^+ = I_l + \delta \cdot I_l$; $I_w^- = I_w - \delta \cdot I_w; I_w^+ = I_w + \delta \cdot I_w$ – minimum and maximum value of utility service costs entered by the user into the EPS system, relatively, for gas, electricity and water; δ - the percentage of deviation from the user-specified indicator. For adequate and correct adjustment of network weights it is advisable that the indicator δ put within such limits - $0{,}1 < \delta < 0$.

The condition of concordance of the forecasted utility service costs calculated by the system and the specified interval corridor of possible costs [7] are:

$$I_g^{\exp} \subset [I_g^-; I_g^+];\ I_l^{\exp} \subset [I_l^-; I_l^+];\ I_w^{\exp} \subset [I_w^-; I_w^+], \quad (2)$$

$I_g^{\exp}$, $I_l^{\exp}$, $I_w^{\exp}$ – forecasted utility service costs.

The indicator of concordance of forecast and real data depends on the size of the input sample on which the network is trained. As the larger the input sample, as the higher the concordance rate.

Preparation of data for further processing in the neural network takes place in several stages. First, the availability of the necessary weather information is checked in the local database. If it is not present, it is got from the weather site [8], also in the same way getting the information about the time of sunrise and sunset, its aggregated into average values for the month and recorded in the local database for future use (Listing 1).

```
    public async Task<WeatherEntity>
ParseWetherByDate(DateTime date, int cityCode, string
cityName)
      {
         var link =
$"{_weatherSearchUrl}/{cityCode}/{cityName}/{date.ToStrin
g("yyyy-MM-dd")}";
         var content = await _httpClient.GetStringAsync(link);
         var htmlDocument = new HtmlDocument();
         htmlDocument.LoadHtml(content);
         var document = htmlDocument.DocumentNode;
         var tempTable =
document.QuerySelectorAll(".at_temp").Select(p =>
p.InnerText);
         var avgTemp = (int)tempTable.Select(p =>
p.Split("&deg;C").First()).Average(p => Convert.ToInt16(p));
         var wetTable =
document.QuerySelectorAll(".at_r>.at_cnt>.vl_parent>.vl_ch
ild").Select(p => p.InnerText);
         var avgWet = wetTable.Select(p =>
Convert.ToDouble(p) / 100).Average();
         var windTable =
document.QuerySelectorAll(".at_winter>.at_cnt>.vl_parent>.
vl_child").Select(p => p.InnerText);
         var avgWind = windTable.Select(p =>
Convert.ToDecimal(p.Replace('.', ','))).Average();
         var weather = new WeatherEntity
         {
            Date = date,
            Temperature = avgTemp,
            Wetness = avgWet,
            Wind = avgWind
         };
         return weather;
      }
```

Listing 1

The next step is to download ready-made datasets for the current user from the database (Listing 2).

```
private void LoadModelFromFile(MLContext mlContext)
      {
using (var fileStream = new FileStream(_modelPath,
FileMode.Open, FileAccess.Read, FileShare.Read))
         {
var model = mlContext.Model.Load(fileStream);
}
}
```

Listing 2

Next, a set of data is formed from the current query and the information obtained in the first step. Using the Predict method, the predicted result is obtained ((Listing 3).

```
privatePredictionPredict(MLContextmlContext)
{
         ITransformer loadedModel;
using (var stream = new FileStream(_modelPath,
FileMode.Open, FileAccess.Read, FileShare.Read))
         {
loadedModel = mlContext.Model.Load(stream);
         }
var predictionFunction =
loadedModel.MakePredictionFunction<DataSet,
Prediction>(mlContext);
var set = new DataSet();
var prediction = predictionFunction.Predict(set);
return prediction;
      }
```

Listing 3

The server and, in particular, the neural network are implemented in the C # programming language using the ML.NET library. A regression three-level network model was used. The "back-side" approach was used as a corrective function for train the network.

The network step by step learns on each request and data set and as a result it learns to adjust weights correctly that the coefficient of dependence of these or those factors was the closest to real, and the more precisely adjusted weights, the more exact result is.

To build the basis of the neural network, an algorithm for processing the data set is designed, which is then used in the calculation of network weights. The basic formulas used for this:

- MapValueToKey – uses the indicator as a certain key of the data set, in this case used to classify the months, because from them mainly depend the weather.
- ReplaceMissingValues – views all variants of the indicator and inserts missed cases.
- OneHotEncoding – converts data into a sorted vector.
- Normalize – converts data into a certain range, in this case used to normalize additional weather indicators.

Fig. 3 shows schematically the method of calculating the projected costs for the consumed gas on the specified date, and Fig. 4 - for the consumed electricity. An analysis of the dependence of the accuracy of the prediction of the result on the number of data sets was performed. The system was trained using a number of random data sets and tested for accuracy 10 times. As the accuracy assessment took the average value of 10 attempts. Since the system is customized for a specific user, it does not require a lot of data to work quite accurately. As can be seen from Fig. 5, even with 10 data sets, it gives an error of 27%, and with 100 sets of only 7%.

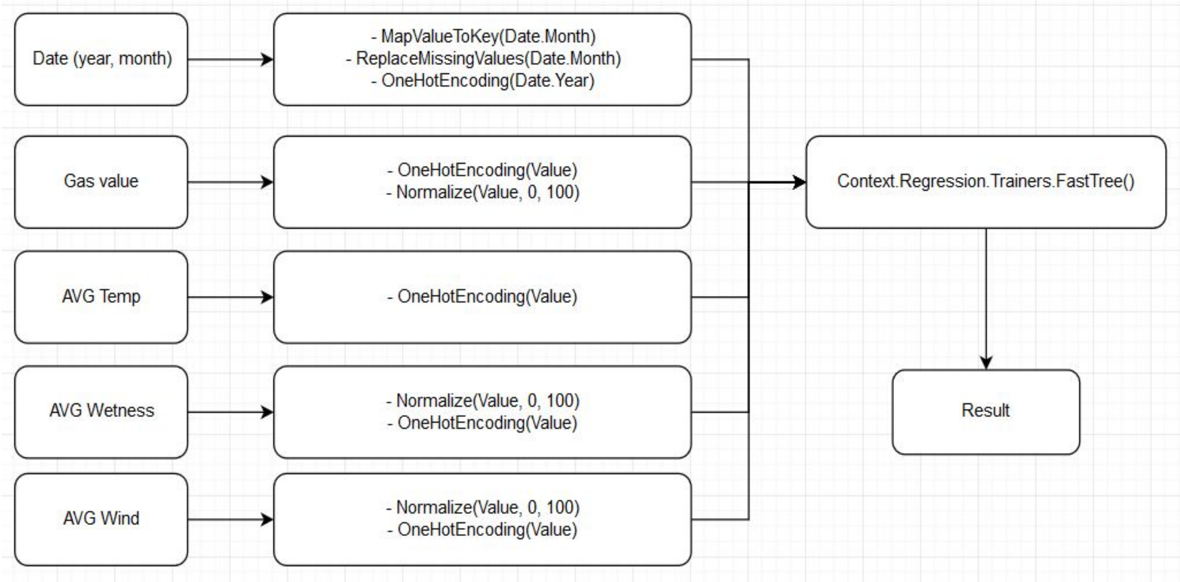

Fig. 3. Neural network diagram for gas cost calculations

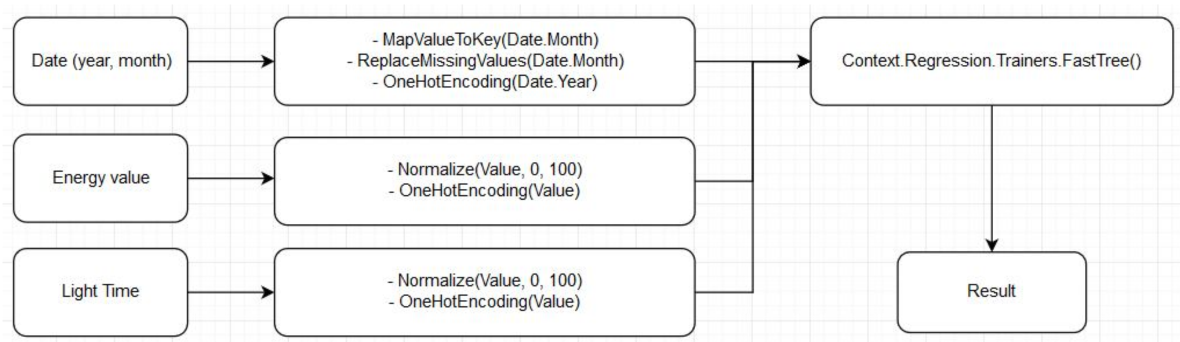

Fig. 4. Neural network diagram for calculating the cost of electricity

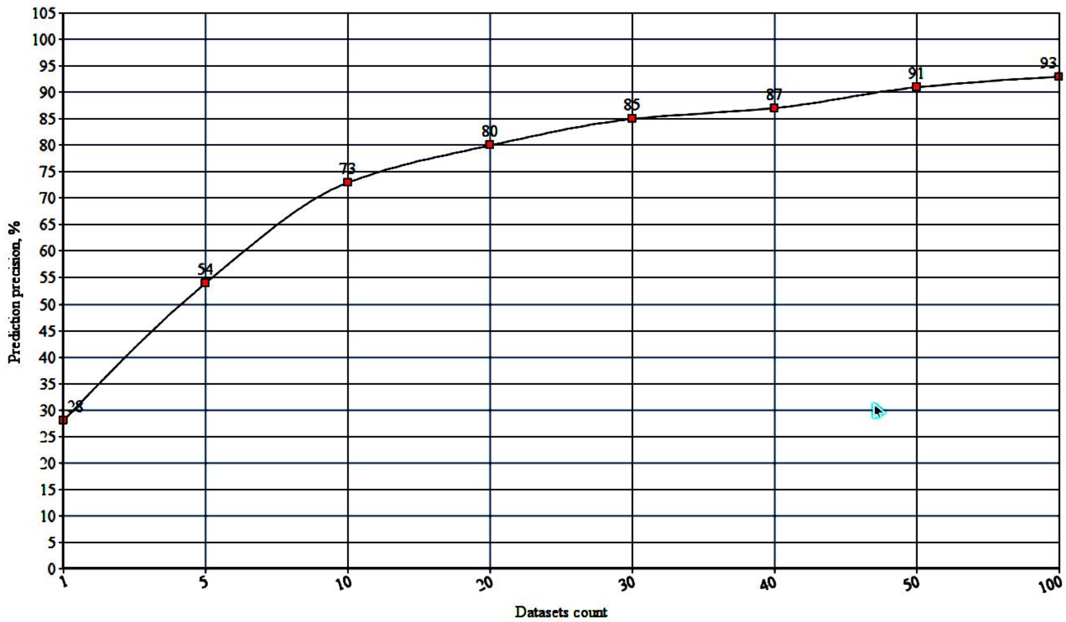

Fig. 5. Graph of the dependence of the accuracy of cost forecasting on the amount of user input

Consider the example of neural network training and the results of forecasting utility service costs. The input data presented in Table 2 are obtained from the EPS system for quick adaptation of the software system to any user, because we can immediately get the minimum required set of input data to obtain an accuracy of more than 90%.

Table 2 – **Input data for neural network training**

Date	Elec-tricity	Gas	Water (1 counter)	Water (2 counter)
01.11.2016	575	1150	7	1
27.11.2016	722	1473	12	1
30.12.2016	1030	1880	18	2
04.01.2017	1204	1927	21	2
28.01.2017	1386	2235	25	3
…	…	…	…	…
04.09.2017	2564	3042	76	13
03.10.2017	2888	3111	82	14
02.11.2017	3138	3348	89	15
30.11.2017	3492	3594	95	16
31.12.2017	3887	3971	104	18
30.01.2017	4067	4279	111	18
…	…	…	…	…
31.07.2018	5122	5032	153	27
05.09.2018	5291	5059	160	29
01.10.2018	5448	5123	166	30
31.10.2018	5609	5297	173	31
02.12.2018	5787	5596	179	33
31.12.2018	5958	5869	186	34
…	…	…	…	…
28.05.2019	6741	6940	218	41
01.07.2019	6975	6894	223	44
01.08.2019	7062	6960	229	45
03.09.2019	7146	7011	233	46
30.09.2019	7279	7070	238	48

Fig. 6 shows a graph comparing the input data set, in particular for gas, and forecast data.

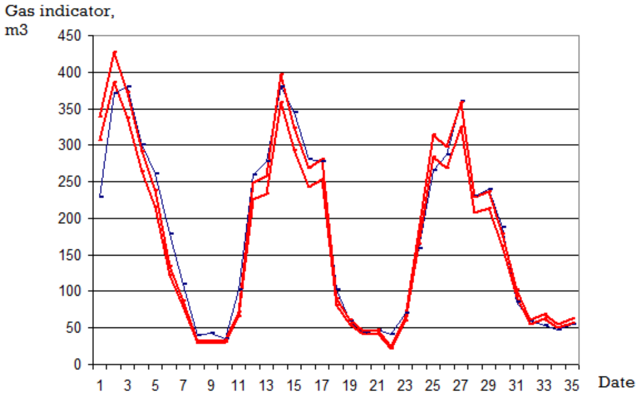

Fig. 6. Graph of comparison of gas consumption indicators on real and forecast data, where - interval corridor of input values with an error of 5%; - projected gas utilization rate calculated by the system

Another part of the experiment was division of quantity of input data for data only for neural network training (70%) and prediction (30%). As can be seen from Fig. 7, the neural network predicts light consumption within 8% accuracy, which set an interval corridor, easily adapts to insignificant fluctuations of indicators, but it does not immediately respond to significant deviations, the cause of which may be different. Therefore, one of the real reasons is failure of the EPS system or their incorrect input into the system by the users.

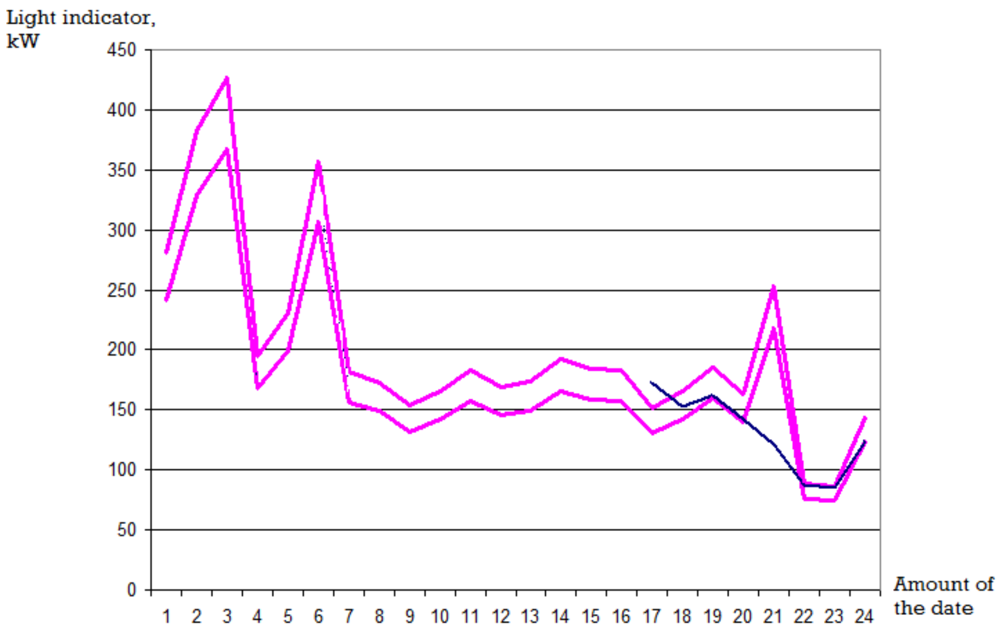

Fig. 7. Graph of comparison of input and forecast data by the system based on training the network with the part of the data, where - the interval corridor of input values with an error of 5%, - the forecast indicator of electricity use, calculated by the system

Therefore, based on the results of forecasting the system of electricity costs in accordance with formulas (1) and (2), was build system of coordination of indicators, I_l, that is:

$$\begin{array}{ll} 172 \not\subset [130,2;151,2]; & 121 \not\subset [217,62;252,72]; \\ 153 \subset [142,29;165,24]; & 87 \subset [76,26;88,56]; \\ 162 \subset [159,96;185,76]; & 85 \subset [74,4;86,4]; \\ 142 \subset [140,43;163,08]; & 124 \subset [123.,69;143,64]. \end{array} \quad (3)$$

Based on the results of training and forecasting of utility service costs by the system, in particular of the electricity, from expression (3) we can conclude that the system makes an adequate forecast within the established accuracy and small sample of input data.

Over time and increase the amount of input data, the network can learn to minimize these differences, but it is not insured against accidental changes in the data due to various factors.

There are two variants for solving this problem in further development:

1. Identify and increase the number of factors influencing the network.

2Ignore data jumps, when the indicator is much higher or lower than the diapason of previous indicators.

The main screen forms of the software system for forecasting utility service costs using a neural network are shown in Fig. 8 – 10.

Fig. 8. Selecting date or period for forecast

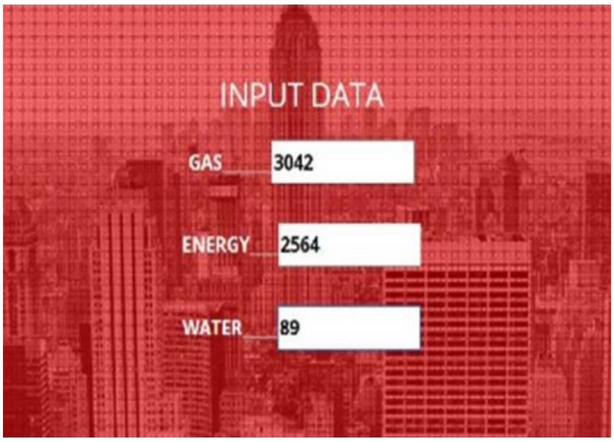

Fig. 9. Entering user data

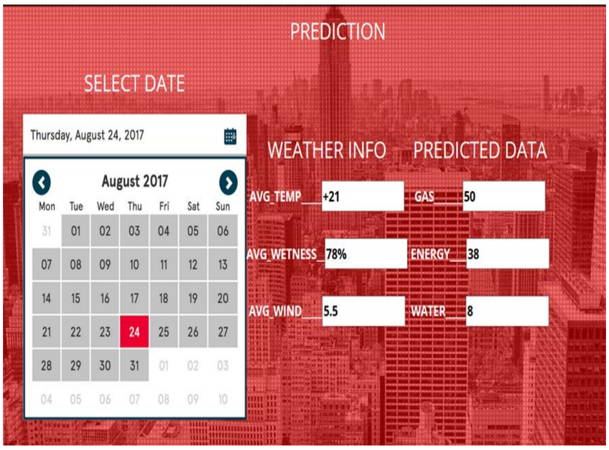

Fig. 10. The result of the forecasting system

Fig. 8 shows the calendar for selecting the date when the user wants to make a forecast of utility costs.

Fig. 9 shows the form in which the user enters the data of the counter used services, if he does not have access to the system of entering EPS indicators, and Fig. 10 shows the result of the software system, which shows the estimated cost of services for the selected date.

Conclusions

Research on forecasting utility service costs is an urgent task in recent years, especially in Ukraine, because in to pay utility costs you need to spend a significant part of salary, which in winter period often exceeds 25% of earnings. Therefore, the possibility of reducing this percentage, in particular by forecasting possible costs in accordance with the forecasted weather conditions and reducing them by adjusting certain indicators is appropriate and necessary.

Today the authors have designed and partially implemented a system for forecasting utility costs using a neural network. However, this problem is multifaceted and contains a large number of different factors that need to be considered in future research. Take into account that all factors have a certain randomness of influence and error of measurements and calculations; it is advisable in the future to apply the methods of analysis of interval data not only to concordance the data, but also to build a forecast model.

REFERENCES

1. (2017), *Communal Genocide 2016/17*, available to: https://bastion.tv/komunalnij-genocid-201617_n15658
2. Krepych, S.Ya., Spivak, I.Ya. and Kapush, M.V. (2018), "System of forecasting the cost of utilities using neural networks", *Proceedings of the seminar Computer System and Information Technologies CSIT'2018*, June 2, 2018, Ternopil, pp. 37-38.
3. (2019), *Utility payments: how Ukrainian expenditures on utilities have changed over the last 15 years*, available to:: https://www.slovoidilo.ua/2019/12/16/infografika/suspilstvo/platezhi-komunalku-yak-zminyuvalysya-vytraty-ukrayincziv-komunalni-posluhy-ostanni-15-rokiv
4. (2020), *How neural networks work: about a complex system in simple words*, available to: https://robo-hunter.com/news/kak-rabotayt-neironnie-seti-o-slojnoi-sisteme-prostimi-slovami14200
5. Bobalo, Y., Dyvak, M., Krepych, S. and Stakhiv, P. (2014), "Evaluation of functional device suitability considering both random technological deviations of its parameters from their nominal values and the process of components' aging", *Przegland Elektrotechniczny*, Warszawa, Poland, 90(4), pp. 224-228.

6. Krepych, S., Stakhiv, P. and Spivak, I. (2013), "Analysis of the tolerance area parameters REC based on technological area scattering", *12th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM'2013*, pp.179-180.
7. Krepych, S., Spivak, I., Spivak, S. and Krepych, R. (2019), "Modelling of the Process of the Small Hydroelectric Power Station on the Basis of Interval Input and Output Data Analysis", *9th International Conference on Advanced Computer Information Technologies, ACIT 2019*, pp.44-47.
8. (2020), *Weather forecast*, available to: https://meteo.ua/

Received (Надійшла) 29.09.2020
Accepted for publication (Прийнята до друку) 11.11.2020

ВІДОМОСТІ ПРО АВТОРІВ/ ABOUT THE AUTHORS

Крепич Світлана Ярославівна – кандидат технічних наук, доцент, доцент кафедри комп'ютерних наук, Західноукраїнський національний університет, Тернопіль, Україна;
Svitlana Krepych – Candidate of Technical Sciences, Associate Professor, Associate Professor of Computer Science Department, West Ukrainian National University, Ternopil, Ukraine:
e-mail: msya220189@gmail.com; ORCID ID: https://orcid.org/0000-0001-7700-8367.

Співак Ірина Ярославівна – кандидат технічних наук, доцент, доцент кафедри комп'ютерних наук, Західноукраїнський національний університет, Тернопіль, Україна;
Iryna Spivak – Candidate of Technical Sciences, Associate Professor, Associate Professor of Computer Science Department, West Ukrainian National University, Ternopil, Ukraine;
e-mail: spivak.iruna@gmail.com; ORCID ID: https://orcid.org/0000-0003-4831-0780.

Система прогнозування витрат на комунальні послуги на основі нейронної мережі

С. Я. Крепич, І. Я. Співак

Анотація. Робота присвячена проблемі надмірної витрати коштів громадян на оплату комунальних послуг, особливо у зимовий період, коли вказані витрати можуть становити понад 25% бюджету сім'ї. Питання можливості економії хоча б частини вказаних витрат шляхом моніторингу їх можливої величини та зменшення цього показника на сьогодні постає актуальною задачею. Звідси розробка програмної системи прогнозування витрат на комунальні послуги є актуальним практичним завданням. Для рішення цього завдання авторами пропонується використання нейронної мережі, адже її доцільно застосовувати в ситуаціях, коли є наперед визначена відома інформація і на її основі користувачеві необхідно отримати прогнозовану нову інформацію. Метод прогнозування витрат на комунальні послуги базований на використанні нейронної мережі приймає на вхід дані користувачів про витрати на послуги введені вручну або отримані із системи EPS, що є зручним, адже можна одразу отримати великий набір вхідних даних для більш точнішого прогнозування майбутніх витрат. Ще одним типом вхідних даних є дані отримані із сайтів прогнозу погоди для визначення прогнозних показників для коректного навчання мережі. На вказаних даних мережа навчається та будує для кожного користувача окрему модель прогнозування витрат на комунальні послуги. Враховуючи, що внесені користувачами у систему дані про витрати на комунальні послуги кожного місяця можуть не співпадати по даті, то для врахування цієї неточності пропонується вхідні дані для прогнозування подавати у вигляді інтервального коридору значень, що містить мінімальну та максимальну межу прогнозу. Розроблена програмна система та, відповідно, метод прогнозування витрат на комунальні послуги було апробовано на прикладі реального користувача системи EPS.
Ключові слова: комунальні послуги; система прогнозування витрат; нейронна мережа; інтервальні дані; точність оцінювання.

Система прогнозирования расходов на коммунальные услуги на основе нейронных сетей

С. Я. Крепыч, И. Я. Спивак

Аннотация. Работа посвящена проблеме чрезмерного расхода средств граждан на оплату коммунальных услуг, особенно в зимний период, когда указанные расходы могут составить более 25% бюджета семьи. Вопрос возможности экономии хотя бы части указанных расходов путем мониторинга их возможной величины и уменьшение этого показателя на сегодня стоит актуальной задачей. Отсюда разработка программной системы прогнозирования расходов на коммунальные услуги является актуальной практичной задачей. Для решения этой задачи авторами предлагается использование нейронной сети, ведь ее целесообразно применять в ситуациях, когда предопределена известная информация и на ее основе пользователю необходимо получить прогнозируемую новую информацию. Метод прогнозирования расходов на коммунальные услуги основанный на использовании нейронной сети принимает на вход данные пользователей о расходах на услуги введены вручную или полученные из системы EPS, что является удобным, ведь можно сразу получить большой набор входных данных для более точного прогнозирования будущих расходов. Еще одним типом входных данных есть данные получены с сайтов прогноза погоды для определения прогнозных показателей для корректного обучения сети. На указанных данных сеть учится и строит для каждого пользователя отдельную модель прогнозирования расходов на коммунальные услуги. Учитывая, что внесенные пользователями в систему данные о расходах на коммунальные услуги каждый месяц могут не совпадать по дате, то для учета этой неточности предлагается входные данные для прогнозирования подавать в виде интервального коридора значений, который содержит минимальную и максимальную границу прогноза. Разработана программная система и, соответственно, метод прогнозирования расходов на коммунальные услуги были апробированы на примере реального пользователя системы EPS.

Ключевые слова: коммунальные услуги; система прогнозирования расходов; нейронная сеть; интервальные данные; точность оценки.