Akademia Nauk Stosowanych w Nowym Sączu

Journal of Engineering, Energy and Informatics

No. 1/2024 (3)

Nowy Sącz, December 2024

Published by
Akademia Nauk Stosowanych w Nowym Sączu
ul. Staszica 1, 33-300 Nowy Sącz, tel.: 18 443 45 45, e-mail: sog@ans-ns.edu.pl

ISSN 2720-4162

Publisher's address
Wydawnictwo Naukowe Akademii Nauk Stosowanych w Nowym Sączu
ul. Staszica 1, 33-300 Nowy Sącz, tel.: +48 18 443 45 45, e-mail: wn@ans-ns.edu.pl, tbolanowska@ans-ns.edu.pl
wydawnictwo.ans-ns.edu.pl

Editorial Committee adress
ul. Zamenhofa 1A, 33-300 Nowy Sącz, tel.: +48 18 547 29 08, e-mail: jeei@ans-ns.edu.pl, http://jeei.ans-ns.edu.pl

Printed by
Wydawnictwo i drukarnia NOVA SANDEC s.c.
Mariusz Kałyniuk, Roman Kałyniuk
ul. Lwowska 14, 33-300 Nowy Sącz, tel.: 18 441 02 88, e-mail: biuro@novasandec.pl

Contents

Jerzy KOROSTIL[1], **Olesya AFANASYEWA**[2], **Olga KOROSTIL**[3]

[1] University of Applied Sciences in Nowy Sacz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, e-mail: jkorostil@ans-ns.edu.pl
[2] Pedagogical University of Krakow Faculty of Mathematics, Physics and Technical Sciences, Podchorążych 2, 30-084 Kraków, e-mail: olesia.afanasieva@up.krakow.pl
[3] PSChem Sp. z o.o., Świętokrzyska 18, 00-052 Warszawa, e-mail: jkorostil3@gmail.com

Analysis of forecasting parameters for technical objects

Abstract

The aim of the work is the development and research of parameters characterizing the process of functioning of a hybrid forecasting system oriented to use in technical object management systems. Parameters characterizing the process of its functioning have been developed and investigated in the paper.

Such parameters are relatedness and co-dimensionality, which characterize the consistency of the parameters of random events that negatively affect the technical object, with the parameters of counteracting this influence, which are formed within the framework of the forecasting system. A parameter has been developed that characterizes the degree of determinism of individual components of the system, which characterizes the amount of reduction in the degree of randomness of the obtained results of transformations implemented in the corresponding component. Also developed is the parameter of the measure of the countermeasure against the negative impact of a random process on a technical object, which is carried out by the corresponding component of the forecasting system, in the event of an external random event.

Thanks to the use of these parameters, it is possible to evaluate the key functions implemented by the forecasting system when it is used in management systems of technical objects and relevant technological processes.

Keywords: kinship, proportionality, consistency, determinism, counteracting negative influence.

Analiza parametrów prognozowania dla obiektów technicznych

Streszczenie

Celem pracy jest opracowanie i badanie parametrów charakteryzujących proces funkcjonowania hybrydowego systemu prognozowania zorientowanego do wykorzystania w systemach zarządzania obiektami technicznymi. W artykule opracowano i zbadano parametry charakteryzujące proces jego funkcjonowania.

Takimi parametrami są pokrewieństwo i współwymiarowość, które charakteryzują zgodność parametrów zdarzeń losowych negatywnie wpływających na obiekt techniczny z parametrami przeciwdziałania temu wpływowi, które kształtują się w ramach systemu prognozowania. Opracowano parametr charakteryzujący stopień determinizmu poszczególnych składowych układu, który charakteryzuje wielkość zmniejszenia stopnia losowości uzyskanych wyników przekształceń realizowanych w odpowiadającym im elemencie. Opracowano również parametr miary przeciwdziałania negatywnemu wpływowi procesu losowego na obiekt techniczny, realizowanego przez odpowiedni element systemu prognozowania, w przypadku wystąpienia zewnętrznego zdarzenia losowego.

Dzięki wykorzystaniu tych parametrów możliwa jest ocena kluczowych funkcji realizowanych przez system prognozowania, gdy jest on wykorzystywany w systemach zarządzania obiektami technicznymi i odpowiednimi procesami technologicznymi.

Słowa kluczowe: pokrewieństwo, proporcjonalność, spójność, determinizm, przeciwdziałanie negatywnemu wpływowi.

1. Introduction

In most cases, decisions about the need to use forecasting methods are made based on the analysis of various factors, which include data about the processes in relation to which they can be used, the nature of the problems for the solution of which a decision is made to use forecasting methods, and a number of other factors. In addition, there is a need to evaluate the effectiveness of the use of forecasting, which is mainly performed at the end of its current cycles. In general, methods of evaluating forecasting processes are based on the use of various analytical methods of analyzing the relevant processes. In order to implement such an assessment, in addition to the above approaches, it is advisable to analyze the capabilities of the selected forecasting system (SPG_i) solve the problems for which it is intended to be used. Such studies should be based on the results of the analysis of SPG_i parameters, their relationship with the features of the researched process and on the appropriate interpretation of SPG_i.

Forecasting in the classical interpretation is the process of determining the possibility of the occurrence of some event in accordance with the change in the value of the synchronizing parameter, which, in most cases, is time. We will call such forecasting synchronized. When it is necessary to predict the occurrence of some situation that may be caused by unforeseen events, then we can talk about situational forecasting. We will call the mechanisms of forecasting processes based on the use of extended methods of building hypotheses hypothetical.

Let's assume that forecasting is used to identify possible random events Vp_i that have a negative impact on the relevant technological processes TPr_i. Methods of using forecasting results Vp_i in control systems TPr_i allow different ways of their interpretation, examples of which can be the following descriptions.

1. Detection of the possibility of occurrence of unpredictable random events Vp_i and their causes.
2. By supplementing the forecasting system with means that provide expansion of its functional capabilities, the forecasting system turns into a hybrid forecasting system, which provides the possibility of expanding information about forecasted events.
3. The use of data obtained after the completion of a number of cycles of the forecasting process, for the implementation of the current cycle of the functioning of the forecasting system, allows solving problems related to the surrounding environment and technological processes.

The first description of the interpretation corresponds to the task of forecasting the selected event that may occur in the environment En_i in which TPr_i functions.

In the second case, it is possible to expand forecasting processes, which allows to connect it with solving other problems, for example, problems of protectioning TPr_i from dangerous events Vp_i, (Blanke, Kinnaert, Lunze, Staroswiecki, 2004). This approach to the interpretation of the prediction system SPG_i is quite common, as it allows expanding the capabilities of the SPG_i system itself (Resienkiewicz, 2019). This interpretation applies to hybrid forecasting systems, in which additional means of analysis can be used. The use of a hybrid approach to the creation of SPG_i makes it possible to more closely link the means of such a system and methods of forecasting with the features of the selected process TPr_i. This makes it possible to evaluate the parameters of the forecasting process and its effectiveness based on the data on the results of the impact on TPr_i of the processes on the part of SPG_i, which are aimed at counteracting the negative impact of Vp_i on TPr_i.

The third approach to the interpretation of SPG_i involves using the results of multiple prediction of events Vp_i with the aim of identifying patterns of their occurrence, as well as with the aim of using forecasting data to increase the accuracy of current forecasts.

In this work, the methods of analyzing the validity of the use of forecasting tools in the tasks of managing technical objects based on the determination of the values of the parameters characterizing them are investigated.

2. Parameters of the forecasting system

To determine the requirements for the SPG_i forecasting system, it is necessary to analyze the parameters that characterize it.

The characteristic that unites a group of two separate parameters, which are affinity and dimension, will be denoted by the abbreviation *AD*. The *AD* characteristic will determine the degree of consistency of the parameters used in SPG_i and TPr_i, which will be denoted by η_i. The given parameter types are described by the following definitions.

Definition 1. The affinity, which we will denote by η_i^a, means that the two compared parameters used in SPG_i and TPr_i are determined by the same natural factors, or have the same nature of interaction with the elements of $SUP_i(TPr_i)$, where SUP_i is the technological process control system TPr_i.

Let's assume that η_i^a can take the following limit values, which will be determined based on the analysis of factors affecting this value:

- $\eta_i^a = 0$ – means that the two compared parameters are not related, or do not have a common physical nature;
- 0<η_i^a <1 – means that the two compared parameters have differences in affinity;
- $\eta_i^a = 1$ means that the two compared parameters are related or have a common physical nature, the case when $\eta_i^a > 1$ is not considered.

There may be cases where the two selected parameters differ to some extent in their common physical nature. At the same time, they can interact, which depends on the magnitude of the degree of deviations of affinity between these two parameters (Gojda, 2013).

Definition 2. Proportionality, which will be denoted by η_i^d, means that the two compared parameters used in SPG_i and TPr_i are characterized by affinity and are measured by agreed measurement scales.

If two interacting parameters are related, but not commensurate, then in this case, an additional fragment of the technological process is designed, which would ensure commensurability of these parameters. There may be a case when such commensurability cannot be fully ensured for one reason or another. An example of such a reason can be undesirable redundancy in the implementation of TPr_i. In this case, partial redundancy is implemented in TPr_i, which is used only in case of occurrence of Vp_i. Such redundancy is activated by means of the system for counteracting the negative influence (SPW_i) of the event Vp_i on TPr_i, which is part of SPG_i. The range of values taken by η_i^d is determined similarly to the definition of η_i^a. The parameter η_i is a dimensionless value and is determined by the ratio:

$$\exists(\eta_i^\alpha = 0)[\eta_i = 0] \vee \exists(0 < \eta_i^a \leq 1)[\eta_i = \eta_i^a \cdot \eta_i^d] \vee \exists(\eta_i^a > 1)[\eta_i^a = \neg\eta_i].$$

The given characteristic can be used in various systems performing data analysis and transformation. The degree of inconsistency of the interacting parameters leads to errors in the analysis and calculation processes within the SPG_i system. If $\eta_i = 0$, it means that the calculations are performed with completely inconsistent systems SPG_i and TPr_i.

An important parameter for the SPG_i system is the degree of determinism of the processes implemented in the system components. The use of this parameter in forecasting systems is conditioned by the fact that in SPG_i one of the key components is the prediction component (SPB_i), which, by definition, cannot be considered deterministic (Billingslej, 2021). If we assume that the general purpose of using the SPG_i system is to achieve the necessary accuracy of predicting the values of the Vp_i parameters, then the degree of determinism of the components of the SPG_i system is a rather important parameter. We will denote this measure by $\lambda(\mathcal{K}_i)$, where $\mathcal{K}_i$ means the component of the SPG_i system, for example, the SPB_i component, the SWP_i component, or others. The measure λ takes into account the information component, which is a text description of the interpretation of the numerical component of the element $I_j\left(x_i^k\right) = j(x_i) * x_i$, where $j(x_i)$ is a text description of the interpretation of the variable x_i, which determines the numerical value of this component. The transformations $I_j\left(x_i^k\right)$ related to the determination of the value $\lambda(\mathcal{K}_i)$ concern not only the component x_i but also the component $j(x_i)$. The component $j(x_i)$ and the result of its transformations can change the interpretation of x_i and thus affect and determine the parameter λ (Li, Xia, Zong, Huang, 2009).

Let's assume that the determinism of some functional component will be determined by the amount of refinement of data or processes used in SPG_i. Such refinement is provided by the transformations used by the corresponding component and is also provided by the use of additional information by this component. The notion of determinism is quite general and, therefore, there is an opportunity to formulate various variants of interpretation of this concept. In this case, we will form an idea of the degree of determinism of the SPG_i component based on the following approach. The SPG_i system is focused on determining data about the random event Vp_i, on the basis of which it implements countermeasures against the negative impact of Vp_i on TPr_i. Therefore, determining the degree of determinism of the system SPG_i can be verified by the degree of elimination of the influence of Vp_i on TPr_i. In order to increase the value of λ, it is necessary to introduce additional components into the composition of SPG_i, which can provide such an increase. Consider the definition of determinism.

Definition 3. The degree of determinism $\lambda(\mathcal{K}_i)$ of some component $\mathcal{K}_i$ is determined by the amount of information expansion of the normalized text description of data interpretation, the amount of change in the value of this data and the amount of changes in other elements that are part of the information element $In_i(x_i)$, which, in sum, provides an increase in the degree of elimination of the influence of Vp_i on TPr_i. The value $\lambda(\mathcal{K}_i)$ can be described by the following relation:

$$\lambda(\mathcal{K}_{\mathrm{i}}) = \mathcal{K}_i^f\left[In(x_i^w) - In(x_i^v)\right], (1)$$

where $In(x_i^w)$ consists of the following components: the magnitude of the input data values $h(x_i)$, the normalized description of the textual interpretation of these data $j(x_i)$, the amount of data change and other elements with their interpretation used in $\mathcal{K}_i$, $In(x_i^v)$ – represents a structure similar to $In(x_i^w)$, but refers to the information component at the output of the component $\mathcal{K}_i$, f – means the function describing the method of information expansion. The subtraction operator used in (1) means the method of determining the amount by which $In\left(x_i^j\right)$ increases.

The next parameter of the SPG_i system is the measure of the impact of forecasting results on the control object μ. The possibility of using the impact measure μ as a characteristic of SPG_i is indirectly related to the parameters η_i, λ_i and factors determined on the basis of the analysis of the interpretation of the ways of realizing the impact of SPW_i on $SUP_i(TPr_i)$, to which include the technical conditions that determine the possibility of implementation by the SPW_i component of counteracting the influence of Vp_i on TPr_i, and the consistency of the relevant technical means of object protection (SZO) of the system $SUP_i(TPr_i)$ with SPW_i.

Let us assume that TPr_i is implemented by a technical object (TO_i), which is partially characterized by the presence of its own resources Rv_i, which determine the possibilities of using the elements of TO_i, for the implementation of TPr_i and external resources (Rz_i), which determine the external materials and means necessary for the functioning of the process TPr_i and the production of products $\mathcal{Pp}_i$. For the production of $\mathcal{Pp}_i$, the process TPr_i uses the specified resources, due to which they decrease, raw materials in this case, we will not consider. If the negative factor Vp_i acts on TPr_i and TPr_i is not protected, then TO_i loses resources and may stop functioning completely. Protection of TPr_i from Vp_i can be done in different ways:

1. Only by using the SZO system.
2. By using the SZO system together with the transition of TPr_i to such operating modes, for the duration of Vp_i, in which the parameters that are not protected by *SZO* do not require protection because in the selected modes, they are affected by the parameters agreed with Vp_i s not essential.
3. The *SZO* system is not used, and TPr_i switches to operating modes for which the action of the corresponding parameters does not affect TPr_i for the during the action of Vp_i.

Let us introduce the definition of μ. *Definition 4.* The parameter μ of the measure of the impact of countermeasures against threats for TPr_i caused by the occurrence of Vp_i is determined by the ratio of the number of countermeasures provided by the SZO system to the sum of these countermeasures and countermeasures provided by the corresponding change in the operation modes of TPr_i. According to the definition, μ is calculated by the following expression:

$$\mu(Vp_i) = \sum_{j=1}^{k} V\left[SZO\left(x_j^*\right)\right] / \{\sum_{j=1}^{k} V\left[SZO\left(x_j^*\right)\right] + \sum_{r=1}^{g} V[TPr\ (y_r)]\}, \quad (2)$$

where $V\left[SZO\left(x_j^*\right)\right]$ – the measure of the countermeasure of the parameter with Vp_i, which is analyzed and transmitted by the component SPW_i to the SZO, for the protection of TPr_i, $V[TPr(y_r)]$ is the measure of the effect on the parameters of TPr_i, which are associated with the corresponding parameters x_j^* of Vp_i, but x_j^* do not affect y_r due to the fact that in the changed operation mode TPr_i, the parameters y_r cannot be vulnerable to x_j^*.

Parameter μ is used throughout the entire period of use of SPG_i with the system $SUP_i(TPr_i)$. The methods described above can be much more, since for the selection of the modes of operation TPr_i, situations can be selected in which the permissible decrease of one or the other type of resource is taken into account, or a change in the number of products during the time interval of the action Vp_i.

An important, integral parameter of the SPG_i system is the forecasting efficiency, π. This parameter is determined by most factors related to the system and uses parameters η_i, λ_i and μ_i. To evaluate the factors used by the parameter π, the notion of the measure of semantic significance of information (SI-signification of information) is introduced, which is represented by the elements of the textual interpretation description $j(x_i)$. The SI property of some $j(x_i)$ is determined based on the semantic analysis data of the corresponding $j(x_i)$. The following factors can serve as an example of characteristics that are components of efficiency.

1. The average number of detected negative events Vp_i during a given period, which is determined by the number of operation cycles TPr_i.
2. Significance of information obtained as a result of prediction and used to implement countermeasures against the negative influence of Vp_i on TPr_i.
3. The significance of the information provided by additional means and used in the formation of the process of counteracting the negative influence of Vp_i on TPr_i.
4. The average number of successful processes of prejudice to the negative impact of events Vp_i, which is implemented by SPW_i.
5. Characteristics of the input data necessary to predict the occurrence of the event Vp_i.
6. The value of the forecasting time interval ΔT_i of the event Vp_i and a number of other factors that may be related to the process of forecasting Vp_i.

Depending on the features of TPr_i, the forecasting system can be expanded with additional components focused on the use of specific features of the corresponding TPr_i. An example of such a feature can be the need to form a time trend of occurrence of Vp_i over a given number of cycles of TPr_i functioning. The second example of such a feature may be the need to assess the degree of threat to TPr_i from the influence of Vp_i on TPr_i, which is implemented based on the use of selected hazard criteria and others.

3. Determining the values of the parameters of the forecasting system

The parameter of forecasting efficiency, due to its universality, needs a more precise definition of the components characterizing it. From the above factors that determine efficiency, it appears that it can be divided into two parts. The first part is a generalized component of π^u, which can be determined from the data obtained during a number of performed forecasting cycles. The second part is the operational component π^τ, which characterizes the efficiency during one forecasting cycle. The factors determining π^u are described in points 1 and 4, and the factors determining π^τ are described in points 2, 3, 5 and 6. Depending on the need, in each individual case of using π^u and π^τ, the number of factors determining the corresponding efficiency components can be changed. The determination of the total value of efficiency can be calculated based on the ratio:

$$\pi = \pi^u(\Delta t_i) + \sum_{i=1}^{m} \pi_i^\tau, \text{ where } m = (1,2,\dots,m).$$

The AD characteristic determines the degree of consistency of the parameters x_i and y_i interacting with each other, when the event Vp_i affects TPr_i. In many cases, there is a situation where the parameter x_i characterizing Vp_i indirectly affects TPr_i. Such mediation means that there are functional relationships between the parameters Vp_i and TPr_i, which ensure the necessary consistency according to the affinity parameter

η_i^α. Practically, the functional transformation $x_i^\alpha = f^\alpha(y_i)$ cannot provide complete affinity, where y_i is the parameter TPr_i, which interacts with the parameter x_i formed in SPG_i. To determine the value of the parameter η_i^α, the relation $\eta_i^\alpha = |y_i - x_i^\alpha|/y_i$ is used. The transformation $x_i^\alpha = f^\alpha(y_i)$ is preferably implemented within the framework of the technical object in which the corresponding process TPr_i functions and is formed on the basis of the physical interpretation of the corresponding parameter.

The next component of this group is the proportionality parameter η_i^d. It is defined by the relations: $[\aleph(y_i) = \aleph(x_i)] = (\eta_i^d = 1)$, where $\aleph$ is the scale in which y_i and x_i are measured. If there is a situation where $[\aleph(y_i) < \aleph(x_i)]$, then the scale alignment ratio $[\aleph(y_i) < \aleph(x_i)] \rightarrow [x_i^d = f^d(y_i)]$, where $f^d(y_i)$is the corresponding scale equalization function, which is set when designing protection means in TPr_i. Such alignment is carried out in the direction of a larger scale. This means that TPr_i should not receive the impact value x_i, which is significantly greater than the permissible impact value, which can lead to the destruction of TPr_i. Therefore, the protection system TPr_i activates additional means of protection to counteract the influence of x_i on TPr_i, the countermeasures of which are increased to the necessary extent. These tools are implemented within the framework of TPr_i. The value of the parameter η_i^d is determined by the relation: η_i^d= ($|y_i$-x_i^d $|$)/y_i. In the case where the relation: [$\aleph(x_i)$<$\aleph$(y_i)]→(η_i^d=0) is present, the value of x_i is not large enough to effectively influence y_i and consistency is not considered.

The next parameter of the components of the SPG_i system is the measure of their determinism $\lambda(\mathcal{K}_i)$. Let's accept the restriction that the information elements used to determine $\lambda(\mathcal{K}_i)$ consist of two parts. The first part is $j(x_i)$ – a textual description of the interpretation of the variable x_i. The second part is $j(x_i)$ is the value of the number x_i. Textual descriptions are implemented based on the use of ideas about semantic dictionaries Sc_i, which are formed for separate subject areas, which consist of TPr_i and En_i.

Text components are normalized texts that are formed only from phrases presented in Sc_i. For the formation of each segment of the text, defined grammar rules of the corresponding language are used. The component $j(x_i)$ describes the textual interpretation of the number x_i, which describes the possibilities of its use within the framework of the SPG_i system. An information element can be written in the form:

$$In(x_i^v) =< a_1, a_2, \dots, a_n >*< x_i >,$$

where a_i is the phrase of component $j(x_i)$, x_i is the value of the variable. Semantic parameters are used in order to be able to enter numerical estimates in the analysis of textual descriptions (Matsuo, Ishizuka, 2004; Liu, Yu, 2005). In Sc_i, each element, which is a separate phrase of the text, is assigned a semantic significance $\sigma_i^z(\alpha_i)$, which is determined based on the analysis of the subject area $Q = \{En_i, TPr_i\}$. Two adjacent phrases in $j(x_i)$, for example, $<\alpha_i * \alpha_j >$ are characterized by semantic consistency σ_i^u, which is determined by the relation $\sigma_i^u = \left|\sigma_i^z(\alpha_i) - \sigma_i^z(\alpha_j)\right|$. Based on the analysis of Q, the permissible ranges of values for σ_i^u are determined. Similarly, the limit values of σ_i^u for adjacent sentences are determined. The amount of information expansion is determined by the number of phrases with which $j(x_i)$ was expanded, taking into account the values of σ_i^u. The transformation of texts, which is implemented when using formula (1), is implemented by means of text transformations. The purpose of these transformations is to determine the magnitude of the expansion $In(x_i^v)$ (Korostil, Korostil, 2012).

The parameters of the event Vp_i are used to determine the degree of necessary counteraction to such an effect on TPr_i, which is carried out using SPW_i.

To calculate the value μ, the component SPW_i uses the simplified process model TPr_i, which is placed in SPR_i. This model takes into account the values of the following elements: the external resource Rz_i and the internal resource Rv_i, the amount of products $\mathcal{P}p_i$ produced by TPr_i. The quantity μ is dimensionless and is determined by the following limiting values. The maximum value of this parameter $\mu = 1$ corresponds to the situation when the corresponding parameters in TPr_i are selected for all parameters of the event Vp_i and the influence of Vp_i on TPr_i is counteracted using SZO protection tools. In this case, the influence of Vp_i on TPr_i is completely eliminated and it does not lead to additional changes in TPr_i resources. The minimum value $\mu = 0$ corresponds to the situation when the SZO system is not used. In this case, the influence of Vp_i on TPr_i is eliminated by transferring the technological process to a mode in which the parameters of Vp_i do not affect the technological process. This is possible due to the situation when the parameters of TPr_i, which is related to the parameters of Vp_i, are not used in TPr_i, or their values are much larger than the related parameters of Vp_i, which is provided by the selected operating mode TPr_i.

The parameter μ, according to the formula (2), determines the ratio of the number of pre-emptive countermeasures $V\left[SZO\left(x_j^*\right)\right]$ to the sum of $V\left[SZO\left(x_j^*\right)\right]$ and quantities $V[TPr(y_r)]$, which we will also call operative counteraction. Of course, preemptive counteracting of attacks is much more effective compared to operative counteracting, which is implemented in the real-time operation mode $TPr(y_r)$. The SPW_i system implements all algorithms counteracting the implementation of $V\left[SZO\left(x_j^*\right)\right]$, and in the case of $V[TPr(y_r)]$ takes part in counteracting the negative impact of Vp_i together with the systems controlling the technological processes of functioning TO_i. From this it follows that the dependence of copper on the parameters μ and the SPW_i system will not lead to a decrease in the measure of counteracting the negative influence of Vp_i in an uncontrolled manner.

The magnitude of influence μ, which characterizes the component SPW_i acting on TPr_i, also depends on the value and number of parameters Vp_i, to counteract which it is necessary to use means of countering the negative influence of Vp_i on TPr_i. Thanks to the use of the μ parameter, it is possible to determine the required number of countermeasures against events Vp_k, which may be similar to Vp_i on successive cycles of functioning $SUP_i(TPr_i)$. By increasing the rate of change of μ, it is possible to determine, for example, the possibility of dangerous situations in $SUP_i(TPr_i)$.

Information about the prediction time period Δt of the occurrence of Vp_i is needed by the SPW_i component. This information allows you to monitor the possibility of fulfilling the following conditions:

– the amount of time for predicting the moment of occurrence of Vp_i and the process of implementing counteraction $SPW_i(x_i^*)$ of the influence of Vp_i on $TPr_i(y_i)$ should not go beyond the prediction interval ΔT_i;
– within the time limit ΔT_i component SPW_i should prepare the processes of biasing the impact of Vp_i on TPr_i, if it turns out that it is possible, and the implementation of a number of other processes that may be associated with the specific features of TPr_i and the requirements of the forecasting goal.

Additional means used within the SPG_i system can be divided into the following types:

- means that can be used to implement the processes of counteracting the influence of Vp_i on TPr_i;
- means that implement additional functionality to expand information that is necessary for the formation of prejudice processes or processes of counteracting the influence of Vp_i on TPr_i.

Consider additional means of the second type, an example of which is a component or a decision-making system (SPR_i). This component is intended to implement the corresponding information extension of the data received from SPB_i by the SPW_i component. When using advanced information, the SPW_i component can implement a more effective countermeasure against the influence of Vp_i on TPr_i. The capabilities of the SPR_i component, in each individual case, are determined by different requirements. Known representations of components of type SPR_i are quite developed and have different interpretations (Chakravorty, Ghosh, 2009; Turban, 1995). Examples of the functionality of the SPR_i component can be the following functions:

- if necessary, the SPR_i component can calculate the data received from SPW_i, the values of which it can use to form recommendations for the implementation of countermeasures;
- if the choice of a solution for SPW_i depends on the parameters formed in SPB_i and on the factors characterizing $SUP_i(TPr_i)$, then the necessary dependencies are formed in SPR_i, based on the use of which corresponding recommendations are formed, which are transferred to SPW_i and others.

The SPR_i component can contain information about $SUP_i(TPr_i)$ and about the means and processes of countering the influence of possible Vp_i, if they are provided in TPr_i. The SPW_i component determines countermeasure processes taking into account the information received from SPB_i. Thanks to the use of SPR_i, the SPG_i system becomes more versatile, and its adaptation to a separate TPr_i consists in implementing additional functionality and data in the SPR_i component.

4. Organization of the process of functioning of system components prognostication

The organization of the forecasting process SPG_i can be implemented in various variants, which differ from each other using additional components, taking into account individual event parameters Vp_i and others. Let's consider the possible organization of SPG_i work using the given parameters and tools. We will not describe in detail how to use all parameters and factors. Let's consider the organization of the functioning of individual components of the system.

The input data needed to implement the prediction of the random event Vp_i must be adapted to the component SPB_i. The event Vp_i is such a deviation of the value of the parameter of the external factor, which is Vp_i, which has a negative effect on TPr_i. The event Vp_i is generated by the process $Pr_i[En(Vp_i)]$, which is implemented in the En environment in which TPr_i functions. Such an event means that an unacceptable deviation of the parameter that characterizes Vp_i in $Pr_i[En(Vp_i)]$ has occurred. Such a deviation occurs as a result of an accelerated change in the value of the corresponding parameter. Therefore, the initial data $\{\xi_0, \xi_1, \dots, \xi_n\}$ about such a process must be ordered in relation to the selected synchronizing parameter. Such a parameter is time intervals written in the form $\Delta t_i \in \Delta T$. Since the occurrence of Vp_i is determined by the acceleration with which the values of ξ_1 change, the input data must be replaced by their differences between adjacent parameters: $x_i = |\xi_{i-1} - \xi_i|$. Then, you can write the input data in the form: $\{x_1, \dots, x_n\}$ and, to implement the prediction, we will use them as variable values of the parameter characterizing the process $Pr_i[Vp_i(x_i)]$. Consider the following statement.

Statement 1. If the set of variable quantities $\{\xi_1, \dots, \xi_n\}$, which are the initial data used for forecasting, have a common feature β, based on which they are combined into the corresponding set Ξ_k, then there exists a function of the increments of these quantities, which can be represented in the form:

$$\vartheta = \chi(x_1, \dots, x_n),$$

where $\{x_1, \dots, x_n\}$ are independent variables for which $x_i \in X_k$, t_i is a synchronization parameter, ϑ is a dependent variable of a function that can be extrapolated.

The function ϑ is formed based on the use of variables $x_i \in X_k$. Synchronized forecasting provides the possibility of ordering the values $\{x_1, \dots, x_n\}$ in such a way that they are associated with the parameter t_i. If we assume that the function $\vartheta = \chi(x_1, \dots, x_n)$, is discrete, then it exists, by definition, since $\{x_1, \dots, x_n\} \in X_k$ and the place $x_i = |\xi_i - \xi_{i+1}|$, where $\{\xi_1, \dots, \xi_n\}$ are the initial data. If the ordered set ξ_i represents a function with discontinuities, then the discontinuity points are constant and therefore predictable. Discontinuity points in $\phi(\xi_i)$ can be located only within intervals of type $\xi_i - \xi_{i+1}$. Since ξ_i and ξ_{i+1} are numbers obtained on the basis of their registration or from other sources, the function $\vartheta = \chi(x_1, \dots, x_n)$ can be interpreted as pseudo-continuous. The breakpoints in $\phi(\xi_i)$ are constant and known, if the function values at these points cause a negative impact on TPr_i, then in TPr_i protection against their influence is assumed. The function, which is built on the variables determined by the differences between successive values of the initial data $\{\xi_1, \dots, \xi_n\}$, is a function $\vartheta = \chi(x_1, \dots, x_n)$ that can be extrapolated and used to implement prediction processes.

In the process of preliminary analysis of the initial data, the prediction time is determined. The information about the expected event includes, at a minimum, the following data: the values of the parameters characterizing Vp_i, the time interval during which the expected event may occur.

Let's accept the hypothesis that the predicted event Vp_i must be known to some extent by $SUP_i(TPr_i)$ and, as a result, by SPG_i. A characteristic feature of SPG_i, for cases when they are oriented to service $SUP_i(TPr_i)$, is their functional orientation to the corresponding object (Tetlock, Gardner, 2017). Prediction results from SPB_i are transferred to component SPW_i. The initial information about Vp_i, which is known, is used by the SPB_i component and placed in the SPR_i component. Since the component SPW_i provides the possibility of using data about the expected event, it implements the management of the reaction process of the SPG_i system to the possible impact of Vp_i on $SUP_i(TPr_i)$. This part of the process can be of two types: the first type of process is related to the implementation of preventing the influence of Vp_i on the object, which will be denoted by the symbols $Pr_i(PWz_i)$ and the second type of process is related to the implementation of counteracting the influence of Vp_i we will mark the object as - $Pr_i(PWd_i)$ (Cedro, Wilczkowski, 2018).

The process $Pr_i(PWz_i)$ enables the protection system of the SZO object to activate the processes of prejudging the possibility of the influence of Vp_i on $SUP_i(TPr_i)$. The process $Pr_i(PWd_i)$ enables the SZO system to activate the processes of counteracting the possibility of the influence of Vp_i on $SUP_i(TPr_i)$. The use of the first or second type of reaction processes for predicting the occurrence of the event Vp_i depends on the following factors. The first factor is determined by the time interval required for the preparation of protective equipment in the SZO system, which is part of the SUP_i system of the corresponding TPr_i. During the time indicated by the process $Pr_i(PWz_i)$ and determined by the time interval $\Delta\tau_i$, the SZO system performs the processes of preparing the SZO system to counteract the influence of Vp_i. This time interval depends on the time required by the SPW_i component to perform data preparation procedures for their transmission to the SZO system. This is due to the fact that SPW_i may require additional data or calculations performed in the SPR_i component, and from the prediction period, which is determined by the SPB_i component and depends on the complexity of the prediction process. This interval also depends on the transition speed of the anomaly in En_i into a random event Vp_i. From the SZO side, after its activation by the process $Pr_i(PWz_i)$, it is necessary to set the values of the parameters of the corresponding protection means to those that correspond to the values of the parameters of the impact on them of the corresponding event Vp_i.

In case of activation of the process $Pr_i(PWd_i)$, the SZO system in real time, determined by the speed of implementation of the impact process Vp_i, provides the necessary value of the countermeasure against the corresponding impact. The case when it is necessary to activate the countermeasure of Vp_i on $SUP_i(TPr_i)$ is less desirable. To avoid this, it is possible to increase the amount of information about possible events Vp_i, which would reduce the time for calculations performed in the component SPR_i. This problem can be solved by two approaches that complement each other. Such approaches are based on the following hypothesis.

Hypothesis 1. Types of different events Vp_i in environment Q can be repeated and there is a possibility to determine the probability of such repetitions.

The first approach is to accumulate data about different Vp_i in the component SPR_i. When recognizing another event Vp_i, which has already occurred in the past, or is sufficiently similar to it, the SPW_i component can use ready-made data that have been calculated or used in appropriate cases. This approach requires recognition of the current event Vp_i by the system SPW_i (Lin, 2007).

The second approach is more effective and consists of the following. The SPB_i component is extended by additional functions that implement the prediction of the possibility of occurrence of Vp_i events of a certain type in the next case. In this case, we are not talking about the complete identity of the two events, since they may differ in the values of the parameters characterizing them. The similarity of two different events is determined in the following way.

Definition 4. Two random events Vp_i are similar if, to counteract their impact on the object $SUP_i(TPr_i)$, the same type of countermeasures are used in the SZO system.

The next component of the SPG_i system is the SPR_i component. The SPR_i component contains data and function implementation algorithms necessary for the successful implementation of $SUP_i(TPr_i)$ protection against the negative influence of Vp_i. These include the following data, functional components and other information:

1. Information about defenses aimed at countering threats that Vp_i can form.
2. Algorithms for determining current values of parameters characterizing the forecasting system.
3. Information on previous forecasting cycles and data on implemented protection processes TPr_i from Vp_i, which can be used in current cycles of forecasting process implementation.
4. Information about the environment in which TPr_i functions and is the immediate environment of the protection object En_i.
5. Components necessary for the analysis and transformations of textual descriptions of the interpretation of system elements, for example $j(x_i)$, and other additional textual information necessary for the implementation of counteraction processes.
6. Additional functional tools necessary for the implementation of forecasting processes.
7. Means of storage and formation of data about the parameters of the SPG_i system and the threats that were foreseen by the forecasting processes.

The SPR_i component is used within the framework of the SPG_i system as an auxiliary tool in the implementation of the process of forecasting and countermeasures against the influence of Vp_i on TPr_i. The SPR_i system is one of the components, the use of which allows increasing the value of the parameter of the degree of determinism of the SPW_i component and the SPB_i component. This leads to an increase in the value of the influence measure parameter SPW_i on TPr_i. The conditional scheme of the influence of the system SPW_i on TPr_i with the aim of counteracting the negative influence of $Vp_i(x_i)$ on TPr_i can be written in the form:

$$Vp_i(x_i) \to SWP_i(x_i^*), \text{ where } x_i^* = \psi[x_i, x_i^r, x_i^p(\lambda, \eta)],$$

where x_i^r – information obtained from SPR_i, x_i^p is a variable that characterizes the influence of parameters λ, η on x_i. The SPW_i system largely performs intermediate functions between the SPB_i component and $SUP_i(TPr_i)$ using the SPR_i component. Therefore, the autonomy of its functioning, in comparison with the SPB_i component, is significantly smaller.

A certain degree of affinity between the parameters Vp_i and TPr_i is ensured at the design stage of TPr_i. Since TPr_i functions in the environment of En_i, the process of $Pr_i(TPr_i)$ must involve the interaction of its own characteristics with the parameter's characteristic of En_i, where the processes of $Pr_i(Vp_i)$.

An important characteristic that determines the degree of agreement between Vp_i and TPr_i is the AD characteristic consisting of two parameters: affinity and co-dimensionality. The affinity parameter can change if fragments of other environments appear in the surrounding environment. This case requires separate research. Therefore, we will assume that the affinity parameter is within acceptable limits. The parameter of commensurability is analyzed in cases where the commensurability of two parameters is the same and in the case when there is a relationship: $[\aleph(x_i) > \aleph(y_i)] \to [x_i^d = f^d(y_i)]$. The possibility of using such a ratio in relation to a separate parameter is provided in the design process TPr_i. The implementation of a change in the value of this component can change the component of the matching parameter.

The measure of determinism λ is more complicated, as it is determined within the framework of transformation processes performed by one or another component. This parameter, unlike the η parameter, is not a constant characteristic, but depends on the conditions of the transformation process and on the data itself, which are subjected to such transformations. For example, the SPR_i component can consist of a number of transformation algorithms determined by the corresponding problems that need to be solved. Each individual algorithm with SPR_i, depending on the input data and the requirements for the method of their transformations, can form different values of the value of λ. The transformations implemented by the SPR_i component consist of two parts. The first part carries out mathematical transformation of numerical values (functional, logical, combinational or other). The second part performs the transformation of the textual descriptions of the interpretation of the corresponding data [$j^d(x_i)$] using the textual descriptions of the interpretation of the general principle of transformations performed by this algorithm [$j^a(l_i^x)$]. As a result of semantic transformations of these interpretations, we get a textual description of the interpretation of the obtained result, which, by definition, consists of a larger number of semantic elements, which are phrases. Formally, such a transformation can be described by the relation:

$$j_i^R(x_i^* * l_i^x) = \mathcal{F}\{j^d(x_i), j^a(l_i^x)\}, \quad (3)$$

where $j_i^R(x_i^* * l_i^x)$ description of the textual interpretation of the result of semantic analysis and transformation of two components, which are $j^d(x_i)$ and $j^a(l_i^x)$. The description of the transformation process (3) in an explicit form requires the use of a textual semantic analyzer (Jurafsky, Martin, 2009).

For each transformation of numerical values of variables $x_i \to x_i^*$, transformations of their textual interpretations describing the corresponding variables are performed. These text transformations are implemented in accordance with the rules that provide changes in the size of the text description of the interpretation of the obtained result. Such changes lead to an increase in the size of textual descriptions of the result of transformations. The corresponding increase in the final version of the text description $j_i^R(x_i^* * l_i^x)$ in relation to $j^d(x_i)$ and $j^a(l_i^x)$ occurs due to the following factors. The increase occurs due to the use of additional data related to the forecasting object and located in SPR_i and due to the expansion of the semantic value $j^d(x_i) \to j^a(l_i^x)$, which occurred due to the use of the functional transformation $x_i \to x_i^*$.

5. Conclusions

The work examines the parameters characterizing the processes that occur in the hybrid forecasting system. The use of the proposed parameters makes it possible to evaluate the effectiveness of the forecasting process based on the obtained results, which ensures the protection of technological process from the influence of random non-negative events. When using the proposed parameters, it becomes possible to increase the efficiency of using forecasting results. The proposed parameters are related to the reconciliation of variables that determine random events random non-negative events, which have a negative impact on the technological process technological process with variables that oppose the influence of random non-negative events and characterize the processes in object protection system. The importance of such consistency is determined by the need to ensure effective counteraction on the part of technological process to the possible influence of random non-negative events. New parameters are determined, which are a measure of the determinism of the transformations measure of determinism components of the forecasting system performed by individual components in the forecasting system and a measure of the influence of a system to counteract the negative impact of a random event on a technical facility component on the protection system TPr_i technological process, which activates the protection means TPr_i technological process by transmitting the relevant information to of the SZO object protection system included in the management system of the corresponding object $SUP_i(TPr_i)$ technological process control system.

The parameter of the degree of determinism measure of determinism components of the forecasting system of individual components or individual algorithms used in forecasting system is in a certain sense similar to the widely used parameter of prediction accuracy and, as a consequence, the accuracy of transformations implemented directly with the results of predictions. Due to the introduction of this parameter, it became necessary to use textual interpretations of the corresponding parameters, which made it possible to more closely associate this parameter with the subject area at the level of textual descriptions of the interpretation of the corresponding variables.

The use of the parameter measures of the impact of countering threats to the technological process, which determines the degree of influence of a system to counteract the negative impact of a random event on a technical facility on technological process control system, made it possible to expand the idea of ways to protect technological process control system and made it possible to determine the necessary degree of counteraction to the negative influence of random non- negative events. This accuracy value is determined by the degree of elimination of the influence of random non-negative events on technological process control system due to its compensation by the influence of a system to counteract the negative impact of a random event on a technical facility. This allows you to set the limits of the required accuracy of forecasting and transformations implemented in forecasting system.

References

Billingslej, P. (2021). *Prawdopodobieństwo i miara*. Warszawa: Wydawnictwo Naukowe PWN.

Blanke, M., Kinnaert, M., Lunze, J., Staroswiecki, M. (2004). *Diagnosis and Fault-Tolerant Control*. Berlin Heidelberg: Springer-Verlag.

Cedro, M., Wilczkowski, D. (2018). *Pomiary elektryczne i elektroniczne*. Warszawa: Wydawnictwo PKŁ.

Chakravorty, S., Ghosh, S. (2009). Power Distribution Planning Using Multi-Criteria Decision Making Method, Int. *Journal of Computer and Electrical Engg, 1*(5), 1793-8163.

Gojda, J. (2013). *Pomiary parametrów ruchu drogowego*. Warszawa: Wydawnictwo: PWN.

Jurafsky, D., Martin, J.H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Prentice-Hall.

Korostil, J., Korostil, O. (2012). Analysis and interpretation of text model. *Journal of Konbin, 24*(1), 51-56.

Li, S., Xia, R., Zong, C., Huang, C. R. (2009). A framework of feature selection methods for text categorization. [Suntec, Singapore]: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the *4th International Joint Conference on Natural Language Processing of the AFNLP, 2*.

Lin, H.-D. (2007). Automated visual inspection of ripple defects using wavelet characteristic based multivariate statistical approach. *Image and Vision Computing*, *25*, 1785-1801.

Liu, H., Yu, L. (2005). Toward Integrating Feature Selection Algorithms for Classification and Clustering. *IEEE Transactions on Knowledge and Data Engineering*, *17*, 491-502.

Matsuo, Y., Ishizuka, M. (2004). Keyword Extraction from a Single Document Using Word Co-occurrence Statistical Information. *International Journal of Artificial Intelligence Tools*, *13*(01), 157-169.

Resienkiewicz, M. (2019). *Hybrydowe modele prognozowania w produkcji i metodyka oceny ich efektywności*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej.

Tetlock, P.E., Gardner, D. (2017). *Superprognozowanie sztuka i nauka prognozowania*. Warszawa: Wydawnictwo: CeDeWu.

Turban, E. (1995). *Decision Support and Expert Systems*. London: Prentice-Hall.

Przemysław KOZAK[1], **Grzegorz PRZYDATEK**[2]

[1] University of Applied Sciences in Nowy Sącz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, e-mail: przemekkozak@onet.pl
[2] University of Applied Sciences in Nowy Sącz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, e-mail: gprzydatek@ans-ns.edu.pl

Recognition of the efficiency of the structure of an active municipal waste landfill based on the quality of the water environment

Abstract

A landfill site is an engineered facility intended for depositing waste. Regardless of their type, such facilities pose a potential threat to the water and ground environment, as well as to human health and life, because leachate water in contact with groundwater deteriorates their quality. To prevent such a phenomenon, it is necessary to periodically test the composition of water within the landfill. The aim of the research was to analyse the efficiency of the structure of an operated municipal waste landfill based on the analysis of the composition of leachate and groundwater, including the operation and post-operation phases. Based on the physico-chemical analysis of leachate and groundwater in 2016-2020, increased average TOC (Total Organic Carbon) values exceeding 10 mg·dm^{-3} were shown. Although the concentrations of other indicators, including heavy metals, were within permissible limits, the test results showed limited efficiency of the barrier protecting against the impact of stored waste on the aquatic environment. For this reason, there was a need to expand the scope of research, in particular to include toxicity tests of both leachates and water.

Keywords: waste, sealing, monitoring, groundwater, leachate.

Rozpoznanie sprawności konstrukcji czynnego składowiska odpadów komunalnych na podstawie jakości środowiska wodnego

Streszczenie

Składowisko odpadów jest urządzonym obiektem inżynieryjnym przeznaczonym do deponowania odpadów. Niezależnie od ich rodzaju, obiekty takie stwarzają potencjalne zagrożenie dla środowiska wodno-gruntowego, a także dla zdrowia i życia ludzi, gdyż wody odciekowe w kontakcie z wodami podziemnymi wpływają na pogorszenie ich jakości. Aby zapobiec takiemu zjawisku konieczne jest okresowe przeprowadzanie badań składu wód występujących w obrębie składowiska. Celem badań była analiza sprawności konstrukcji eksploatowanego składowiska odpadów komunalnych na podstawie analizy składu wód odciekowych i podziemnych, obejmującej fazę eksploatacji i poeksploatacyjną. Na podstawie analizy fizyczno-chemicznej wód odciekowych i podziemnych w latach 2016-2020 wykazano podwyższone średnie wartości OWO (Ogólny Węgiel Organiczny) przekraczające 10 mg·dm^{-3}. Pomimo, iż stężenia pozostałych wskaźników, w tym metali ciężkich mieściły się w granicach dopuszczalnych, wyniki badań wykazały ograniczoną sprawność przesłony chroniącej przed oddziaływaniem składowanych odpadów na środowisko wodne. Z tego względu wskazano na potrzebę poszerzenia zakresu badań w szczególności o testy toksyczności zarówno odcieków, jak i wód.

Słowa kluczowe: odpady, uszczelnienie, monitoring, wody podziemne, wody odciekowe.

1. Introduction

There are surface and underground waters within the landfill site, which, in contact with leachate water, pose a potential threat to human health and life, as well as the natural environment due to the potential ease of migration of pollutants. Surface waters are mainly fed by atmospheric precipitation and are reservoirs, rivers or lakes, as well as transitional and coastal waters. In turn, groundwater is supplied indirectly through the seepage of groundwater and rainwater. This is related to the conditions of rainwater infiltration (Bujakowski et al., 2014; Przydatek, 2012).

In accordance with the guidelines of the Regulation (2013), monitoring of landfills includes research in three phases, i.e. pre-exploitation, operational, and post-exploitation. During the first phase, the input condition is analysed and lasts until a building permit is issued, and during the second phase, it lasts from the moment of obtaining approval for development until consent is granted to close the landfill. However, the last one took into account 30 years since the closure of the facility (Mizerna, Bożem, 2015).

Groundwater contamination usually occurs because of damage to the landfill insulation or improper functioning of the landfill drainage system. An important aspect is the selection of the location of the landfill, taking into account topography, environmental and climatic conditions (Górecka, Koda, 2010; Klimek et al., 2010). The efficiency of the landfill structure, including sealing, which includes a natural and artificial barrier, remains significant (Wysocka, 2015).

2. Research methodology

The aim of the study was to analyse the efficiency of the structure of the operated municipal waste landfill based on the analysis of the composition of leachate and groundwater, and the hypothesis was that the design of the tested landfill minimizes the impact of stored waste on the quality of groundwater. The research facility is a municipal waste landfill located in south-eastern Małopolska. The facility has been operating since 1985 and non-hazardous and neutral waste is deposited there, with particular emphasis on fractions after secondary segregation.

Groundwater research in 2016-2020 included the analysis of the results from the piezometer P11 (20 tests – sector VI) in the inflow, as well as P13 (20 tests – sector VI) and P5 for the exploitation phase (20 tests – sector VI) and P6 for the phase post-exploitation (10 tests – sectors I-V) at the outflow.

Figure 1. Location of groundwater and leachate testing points in the area of the landfill (northeastern Małopolska, southern Poland)

Legend: P5 – groundwater outflow, P6 – outflow, P11 – inflow, P13 – outflow, SR3 – expansion well, PO1 – lower leachate tank, PO2 – upper leachate tank

The study of leachate water in the years 2016-2020 was carried out in the operational phase (20 measurements – closed sectors IV-V and active VI), and post--exploitation phase (10 measurements – closed sectors I-III) – in the PO1 leachate tank (lower) and the PO2 tank (upper) (closed sectors IV-V and open sectors VI) and the SR3 expansion well located 30 [m] before the first sewage well (Figure 1).

The scope of the research included the determination of physical and chemical parameters (10): pH, EC (electrolytic conductivity), TOC (total organic carbon), PAHs (polycyclic aromatic hydrocarbons) and Cr (chromium), Cu (copper), Zn (zinc), Cd (cadmium), Hg (mercury), Pb (lead). The collected samples of groundwater and leachate were sent to an accredited laboratory, and the test results were subjected to comparative analysis with the permissible limit values specified in the water law permit for leachate water (Rozkrut, 2021), Regulation (2019b) for sewage and leachate and in Regulation (2019a) for groundwater. Research including the identification of the composition of groundwater and leachates in the area of the landfill site, taking into account the indicated indicators, was conducted by Przydatek (2021).

3. Construction and operation of a landfill

From 1985 to the end of 2020, 1,302,634 [Mg] of waste was deposited in the landfill site consisting of six sectors. Sectors I-IV were closed after they were full until 2009. In turn, in 2017, in the closed sector V, on the contrary, from 2017 to 2020, 122,125 [Mg] of waste was deposited in the active sector VI (area 2.2 [ha]).

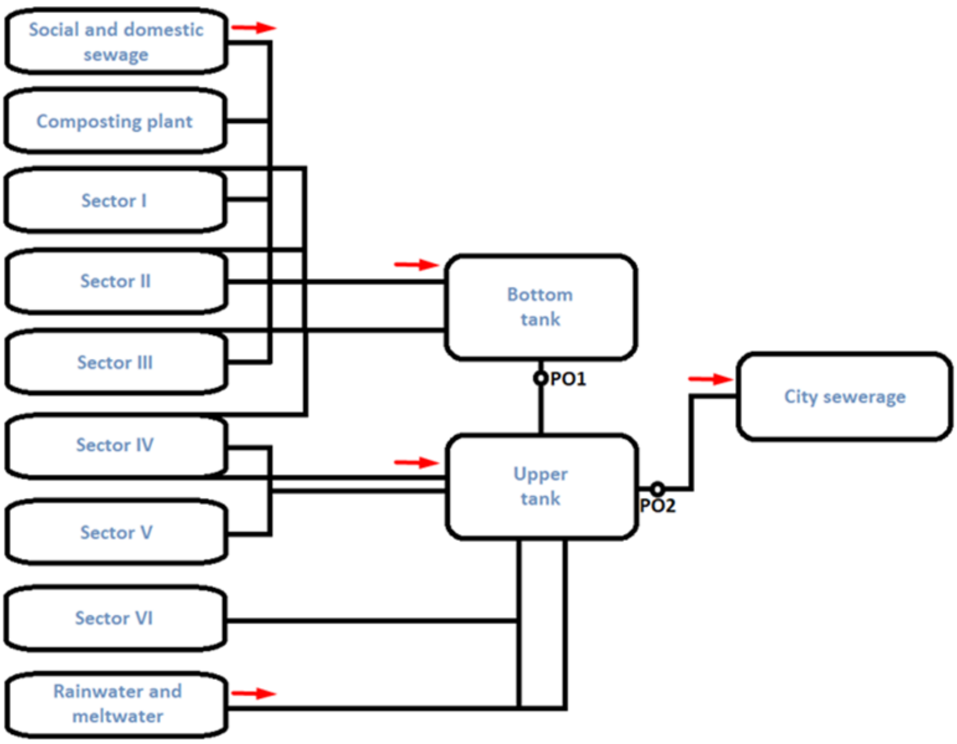

Figure 2. Scheme of water drainage from the landfill site

The analysed landfill has a seal, bottom, and slopes. The insulation was made of a polyethylene geomembrane with a thickness of 2 [mm] and a density of 0.942 [$g \cdot cm^{-3}$] and a geotextile with a thickness of 8 [mm] and a density of 1200 [$g \cdot m^{-3}$], as well as a layer of clay with a thickness of 1 [m].

After the closure process, the landfill is undergoing a recultivation process. After filling a given sector, it is covered with a reclamation layer composed of two parts: a 0.4 [m] thick sealing layer (mineral) of low-permeability soil and a 0.2 [m] thick soil-forming layer (mineral-humus). The thickness of the biological layer is up to 1 [m] for sowing and planting.

Drainage includes the intake of leachate water, its pre-treatment and discharge into the municipal sewage system. The sources of sewage are rainwater from closed sectors I-V, open sector VI, rainwater and meltwater (from roads and squares, composting plants and animal shelters), as well as social and domestic sewage from administrative and social buildings (Figure 2). The drainage system in the landfill consists of open and closed channels:

- drainage discharging leachate water into tanks;
- a perimeter ditch running along the dike of the entire perimeter of sectors V and VI;
- two leachate water reservoirs with pumping stations (the lower one for reclaimed sectors I-III and the upper one for closed sectors IV and V and active VI);
- control and expansion well.

4. Qualitative analysis of the composition of groundwater and leachate

Groundwater

The pH value in groundwater in the inflow was on average 7.4 [pH] and the results were in the range of 6.6-8.1 [pH]. However, in the outflow in the observation wells, the results were at the level of 6.7-8.2 [pH], with an average of 7.3-7.5 [pH].

In the piezometric well P11, the PEW concentration was in the range of 290-1,000 [µS · cm-1] with an average of 674 [$\mu S \cdot cm^{-1}$]. At the outflow in P13, it ranged from 500 to 1,100 [$\mu S \cdot cm^{-1}$] with an average value of 758 [$\mu S \cdot cm^{-1}$]. For the exploitation phase, the average value was 1,114 [$\mu S \cdot cm^{-1}$] with the results ranging from 600 to 1,621 [$\mu S \cdot cm^{-1}$], and for the post-exploitation phase it was 1,178 [$\mu S \cdot cm^{-1}$] in the range of 900-1,700 [$\mu S \cdot cm^{-1}$].

However, the concentration of PAHs in all observation wells was 0.000002 [$mg \cdot dm^{-3}$] throughout the entire research period. The situation was similar to the chromium concentration, which was recorded at 0.01 [$mg \cdot dm^{-3}$].

The minimum value of zinc in all piezometers was 0.02 [$mg \cdot dm^{-3}$]. However, the maximum in P11 was recorded at the level of 0.14 [$mg \cdot dm^{-3}$], and in P13 it was 0.25 [$mg \cdot dm^{-3}$]. However, the maximum value of zinc for the exploitation phase was 0.17 [$mg \cdot dm^{-3}$], and for the post-exploitation phase it was 0.13 [$mg \cdot dm^{-3}$]. The average in all observation wells was in the range of 0.04-0.06 [$mg \cdot dm^{-3}$], including 0.04 [$mg \cdot dm^{-3}$] in the inflow.

The lowest copper concentration in all piezometers was 0.005 [$mg \cdot dm^{-3}$]. However, in the inflow, a maximum of 0.016 [$mg \cdot dm^{-3}$] was recorded with an average of 0.006 [$mg \cdot dm^{-3}$]. In the piezometric well P13, the average value was 0.006 [$mg \cdot dm^{-3}$], while the maximum value was 0.023 [$mg \cdot dm^{-3}$]. The average in P5 was 0.007 [$mg \cdot dm^{-3}$] with a maximum value of 0.021 [$mg \cdot dm^{-3}$]. In turn, the maximum result in P6 was recorded at 0.045 [$mg \cdot dm^{-3}$], and the average result was 0.018 [$mg \cdot dm^{-3}$].

The minimum value of cadmium in all piezometers was 0.0002 [$mg \cdot dm^{-3}$], and the maximum in P11 and P5 - 0.001 [$mg \cdot dm^{-3}$], in P13 - 0.005 [$mg \cdot dm^{-3}$], and in P6 - 0.01 [$mg \cdot dm^{-3}$]. The average at the inflow was 0.0006 [$mg \cdot dm^{-3}$], at the outflow it was from 0.0004 to 0.0015 [$mg \cdot dm^{-3}$].

The minimum lead concentration in all samples was 0.002 [mg · dm^{-3}], and the maximum was 0.01 [mg · dm^{-3}]. The average value over 5 years was 0.005-0.006 [mg · dm^{-3}]. The situation was similar with the mercury concentration, which amounted to 0.0001 [mg · dm^{-3}] in all observation wells. The average included concentrations in the range of 0.0006-0.0007 [mg · dm^{-3}].

Table 1
Composition of groundwater at the landfill in 2016-2020

Parameter	**Value**	**Inflow**	**Outflow**			**Poor chemical condition according to the Regulation (2019a)**
		P11	**P13**	**P5**	**P6**	
Reaction	Min [pH]	6.6	6.9	6.8	6.7	<6.5 and >9.5
	Max [pH]	8.1	8.1	8.2	8.0	
	Average [pH]	7.4	7.5	7.4	7.3	
EC	Min [μS · cm^{-3}]	290	500	600	900	< 3000
	Max [μS · cm^{-3}]	1,000	1,100	1,621	1,700	
	Average [μS · cm^{-3}]	674	758	1 114	1 178	
TOC	Min [mg · dm^{-3}]	2.55	4.81	11.9	10.8	**< 20**
	Max [mg · dm^{-3}]	30.8	19.6	48.3	40.6	
	Average [mg · dm^{-3}]	10.8	10.9	**22.1**	**19.0**	
PAHs	Min [mg · dm^{-3}]	0.000002	0.000002	0.000002	0.000002	< 0.0005
	Max [mg · dm^{-3}]	0.000002	0.000002	0.000002	0.000002	
	Average [mg · dm^{-3}]	0.000002	0.000002	0.000002	0.000002	
Cr	Min [mg · dm^{-3}]	0.01	0.01	0.01	0.01	< 0.1
	Max [mg · dm^{-3}]	0.01	0.01	0.01	0.01	
	Average [mg · dm^{-3}]	0.01	0.01	0.01	0.01	
Zn	Min [mg · dm^{-3}]	0.02	0.02	0.02	0.02	< 2
	Max [mg · dm^{-3}]	0.14	0.25	0.17	0.13	
	Average [mg · dm^{-3}]	0.04	0.06	0.04	0.05	
Cu	Min [mg · dm^{-3}]	0.005	0.005	0.005	0.005	< 0.5
	Max [mg · dm^{-3}]	0.016	0.023	0.021	0.045	
	Average [mg · dm^{-3}]	0.006	0.006	0.007	0.018	
Cd	Min [mg · dm^{-3}]	0.0002	0.0002	0.0002	0.0002	< 0.01
	Max [mg · dm^{-3}]	0.001	0.005	0.001	0.01	
	Average [mg · dm^{-3}]	0.0006	0.0008	0.0004	0.0015	
Pb	Min [mg · dm^{-3}]	0.002	0.002	0.002	0.002	< 0.1
	Max [mg · dm^{-3}]	0.01	0.01	0.01	0.01	
	Average [mg · dm^{-3}]	0.006	0.006	0.005	0.006	
Hg	Min [mg · dm^{-3}]	0.0001	0.0001	0.0001	0.0001	< 0.005
	Max [mg · dm^{-3}]	0.001	0.001	0.001	0.001	
	Average [mg · dm^{-3}]	0.0007	0.0007	0.0006	00007	

Leachate

The pH in the upper reservoir ranged from 7.5 to 8.6 [pH] with an average of 8.1 [pH], while in the lower reservoir, the range of results was 7.7-8.4 [pH] with an average value of 8 .0 [pH]. At the sewage intake point, i.e. in SR3, the results ranged from 7.7 [pH] to 9.1 [pH] with an average of 8.4 [pH].

In the PO1 tank, the PEW value was from 653 to 12,000 [$\mu S \cdot cm^{-3}$], in PO2 from 907 to 11,390 [$\mu S \cdot cm^{-1}$], and in SR3 from 601 to 12,000 [$\mu S \cdot cm^{-1}$]. The average over the 5-years was: 6,575, 7,220, and 7,122 [$\mu S \cdot cm^{-1}$], respectively, which means comparable results for leachates and sewage.

However, TOC results in samples from the upper reservoir included values in the range of 185-920 [$mg \cdot dm^{-3}$] with an average value of 489.1 [$mg \cdot dm^{-3}$], and in the lower reservoir from 178 to 853 [$mg \cdot dm^{-3}$] at an average concentration of 391.3 [$mg \cdot dm^{-3}$]. The average TOC concentration in the expansion well samples was 532 [$mg \cdot dm^{-3}$] with values ranging from 262-1,627 [$mg \cdot dm^{-3}$]. It should be emphasized that the average values of this indicator exceed the concentration permitted for wastewater – 30 [$mg \cdot dm^{-3}$] specified in the Regulation (2019b).

The minimum value of PAHs in three measurement places was 0.000002 [$mg \cdot dm^{-3}$]. However, the maximum value in PO2 was 0.00042 [$mg \cdot dm^{-3}$], in PO1 it was 0.00037 and in SR3 it was 0.00083 [$mg \cdot dm^{-3}$]. The average ranged from 0.00011 [$mg \cdot dm^{-3}$] in the lower leachate tank to 0.00016 [$mg \cdot dm^{-3}$] in the lower one and the expansion well.

The minimum chromium content in the places where leachates and sewage samples were taken was 0.01 [$mg \cdot dm^{-3}$], and the maximum in PO2 was recorded at 0.45 [$mg \cdot dm^{-3}$] with an average of 0.12 [$mg \cdot dm^{-3}$]. However, the lowest average result in PO1 was 0.07 [mg · dm-3] and the highest was 0.15 [$mg \cdot dm^{-3}$] with an average of 0.07 [$mg \cdot dm^{-3}$]. In turn, in the SR3 expansion well, the maximum result was recorded at the level of 0.40 [$mg \cdot dm^{-3}$] with the highest average of 0.14 [$mg \cdot dm^{-3}$]. Average values at points PO2 and SR3 showed that the permissible values were exceeded (Rozkrut, 2021).

The zinc concentration results in the upper reservoir were in the range of 0.05-0.37 [$mg \cdot dm^{-3}$] with an average of 0.11 [$mg \cdot dm^{-3}$]. In the lower reservoir, an average result of 0.09 [mg · dm-3] was obtained, with results ranging from 0.02-0.32 [mg · dm-3]. In SR3, the zinc content was in the range of 0.05-0.81 [$mg \cdot dm^{-3}$] with an average of 0.14 [$mg \cdot dm^{-3}$].

The minimum copper content in leachate and sewage samples was 0.0005 [$mg \cdot dm^{-3}$], while the maximum in tanks was 0.051-0.052 [$mg \cdot dm^{-3}$], and in the expansion well the maximum was 0.067 [$mg \cdot dm^{-3}$]. The highest average occurred at point SR3 - 0.022 [$mg \cdot dm^{-3}$].

Throughout the entire research period, the content of cadmium and lead was 0.01 [mg · dm-3].

However, the mercury concentration in three measurement places was recorded from 0.0005 to 0.001 [$mg \cdot dm^{-3}$]. In the upper reservoir, the average concentration was 0.0008 [$mg \cdot dm^{-3}$], in the lower reservoir 0.0009 [$mg \cdot dm^{-3}$], and the lowest in SR3 – 0.0006 [$mg \cdot dm^{-3}$].

Table 2
Composition of leachate at the tested landfill in 2016-2020

Parameter	**Value**	**Collection point leachate**		**Collection point wastewater**	**Permissible values according to Rozkrut (2021), Regulation (2019b)**
		PO2	**PO1**	**SR3**	
Reaction	Min [pH]	7. 5	7. 7	7. 7	-
	Max [pH]	8. 6	8. 4	9. 1	
	Average [pH]	8. 1	8. 0	8. 4	
EC	Min [$\mu S \cdot cm^{-3}$]	653	907	601	-
	Max [$\mu S \cdot cm^{-3}$]	12,000	11,390	12,000	
	Average [$\mu S \cdot cm^{-3}$]	6,575	7,220	7,122	
TOC	Min [$mg \cdot dm^{-3}$]	185.0	178.0	262.0	< **30**
	Max [$mg \cdot dm^{-3}$]	920.0	853.0	1,627.0	
	Average [$mg \cdot dm^{-3}$]	**489.1**	**391.3**	**532.0**	
PAHs	Min [$mg \cdot dm^{-3}$]	0.000002	0.000002	0.000002	< 0.0005
	Max [$mg \cdot dm^{-3}$]	0.00042	0.00037	0.00083	
	Average [$mg \cdot dm^{-3}$]	0.00011	0.00016	0.00016	
Cr	Min [$mg \cdot dm^{-3}$]	0.01	0.01	0.01	< 0.1
	Max [$mg \cdot dm^{-3}$]	0.45	0.15	0.40	
	Average [$mg \cdot dm^{-3}$]	**0.12**	0.07	**0.14**	
Zn	Min [$mg \cdot dm^{-3}$]	0.05	0.02	0.05	< 2.0
	Max [$mg \cdot dm^{-3}$]	0.37	0.32	0.81	
	Average [$mg \cdot dm^{-3}$]	0.11	0.09	0.14	
Cu	Min [$mg \cdot dm^{-3}$]	0.005	0.005	0.005	< 0.5
	Max [$mg \cdot dm^{-3}$]	0.052	0.051	0.067	
	Average [$mg \cdot dm^{-3}$]	0.021	0.014	0.022	
Cd	Min [$mg \cdot dm^{-3}$]	0.01	0.01	0.01	< 0.4
	Max [$mg \cdot dm^{-3}$]	0.01	0.01	0.01	
	Average [$mg \cdot dm^{-3}$]	0.01	0.01	0.01	
Pb	Min [$mg \cdot dm^{-3}$]	0.01	0.01	0.01	< 1.0/0.5
	Max [$mg \cdot dm^{-3}$]	0.01	0.01	0.01	
	Average [$mg \cdot dm^{-3}$]	0.01	0.01	0.01	
Hg	Min [$mg \cdot dm^{-3}$]	0.0005	0.0005	0.0005	< 0.06
	Max [$mg \cdot dm^{-3}$]	0.001	0.001	0.001	
	Average [$mg \cdot dm^{-3}$]	0.0008	0.0009	0.0006	

5. Discussion

The average value of the reaction in groundwater at the inlet and outlet was up to 7 [pH], which indicated a neutral reaction. Daniszewski and Draszawka-Bołzan (2012) similarly showed the neutral reaction of water and at the same time indicated its alkaline character in the post-exploitation phase of the landfill. The average PEW value was lower than in leachates, ranging from 674 [μS · cm^{-1}] at the inflow to 1,178 [μS · cm^{-1}] at the outflow. Grygorczuk-Petersons and Wiater (2012) showed, in similar conditions, a higher PEW 2340 result [μS · cm^{-1}]. Analysis of the composition of groundwater showed increased TOC concentrations in groundwater below the landfill. At the outflow, the average TOC concentration in the exploitation phase was higher by 11.3 [mg · dm^{-3}], and in the post-exploitation phase by 8.2 [mg · dm^{-3}] about to the one recorded at their inflow, which indicated the poor chemical condition of the water. Differently, Przydatek (2013) showed a lower average concentration of TOC at the outflow of water from a landfill in Masovia, which amounted to 5.4 [mg · dm^{-3}]. It should be emphasized that the microelement test results did not indicate that the permissible level was exceeded, at both the inflow and outflow. In turn, some researchers (Grygorczuk--Petersons, Wiater, 2012) detected increased mercury and cadmium content in water in the area of the landfill.

The analysis of the composition of sewage and leachate confirmed that in the case of pH, the results ranged from 6.7 to 8.2 [pH], which indicated the occurrence of a neutral and alkaline reaction (Koc-Jurczyk, Różak, 2011). The latter result according to Jagiełło (2003) indicates that the landfill has been in use for over 10 years. Other researchers Przydatek (2012) and Żelezik (2015) obtained the reaction in leachates with the values of; 7.3-7.6 [pH] and 7.6-7.7 [pH]. The results of specific electrolytic conductivity ranged from 6,575 to 7,220 [μS · cm^{-3}], which may be due to the dilution of leachates with rainwater (Żelezik, 2015). However, the average TOC content in the tested leachate samples ranged from 391.3 to 532 [mg · dm^{-3}], which exceeded the permissible level (Regulation, 2019b). Melnyk et al. (2014) showed lower TOC concentrations in a landfill. The increased concentration of chromium recorded in two leachate test points confirms that the source of this contamination may be pigment or leather waste, which is a carcinogenic factor (Kim et al., 2015). The concentration of other heavy metals Zn, Cu, Cd, Pb and Hg in the tested leachates was low. According to Tałałaj (2014) low concentration of heavy metals in leachates may result from the formation of stable organic complexes by the metals. Some researchers (Melnyk et al., 2014) used toxicity tests to identify the impact of stored waste on the aquatic environment, which they considered an excellent complement to the monitoring system based mainly on chemical indicators.

6. Summary and conclusions

The analysis of physical and chemical indicators of groundwater, sewage, and leachate showed that in the vast majority of results, the permissible level was not exceeded. However, qualitative analysis of leachate water showed increased TOC and chromium concentrations. The average TOC content in the tested leachate samples ranged from 391.3 to 532 [mg · dm^{-3}], showing exceedances. Similarly, the analysis of the composition of groundwater confirmed that TOC concentrations in groundwater downstream of the landfill in the exploitation and post-exploitation phases were higher by 11.3 [mg · dm^{-3}] and 8.2 [mg · dm^{-3}], respectively, than in the inflow. It should be noted that the test results for microelements in leachate and water did not indicate that the permissible level was exceeded.

To sum up, the results of the analysis indicate the possibility of limited effectiveness of the barrier sealing the bottom of the landfill site, which confirms the impact of leachate on the aquatic environment; it is advisable to expand the scope of research to include toxicity tests for both leachates and groundwater.

References

Bujakowski, F., Ostrowski, P., Sopel, Ł., Złotoszewska-Niedziałek, H. (2014). Glacitektonika krawędziowa w dolinie Wisły a dynamika wód podziemnych. *Landform Analysis, 26*, 61-69 (in Polish).

Daniszewski, P., Draszawka-Bołzan, B. (2012). Wpływ składowiska odpadów na środowisko naturalne w Międzyzdrojach. *International Letters of Chemistry, Physics and Astronomy, 3*, 67-72 (in Polish).

Górecka, A., Koda, E. (2010). Analiza możliwości ograniczenia zagrożeń środowiska wodno--gruntowego, wynikających z eksploatacji modernizowanego składowiska odpadów komunalnych. *Inżynieria i Kształtowanie Środowiska, 19*(3), 48-62 (in Polish).

Grygorczuk-Petersons, E.H., Wiater, J. (2012). Wpływ składowiska odpadów komunalnych na jakość wód podziemnych. *Inżynieria Ekologiczna, 31*, 38-46 (in Polish).

Jagiełło, E. (2003). Metale ciężkie w wodach powierzchniowych w pobliżu składowiska odpadów komunalnych. *Zeszyty Problemowe Postępów Nauk Rolniczych, 492*, 101-108 (in Polish).

Kim, H.S., Kim, Y.J., Seo, Y.R. (2015). An overview of carcinogenic heavy metal: molecular toxicity mechanism and prevention. *J Cancer Prev, 20*(4), 232-240.

Klimek, A., Wysokiński, L., Zawadzka-Kos, M., Osęka, M., Chrząszcz, J. (2010). *Poradnik metodyczny w zakresie PRTR dla składowisk odpadów komunalnych* (in Polish).

Koc-Jurczyk, J., Różak, J. (2011). Skład odcieków pochodzących z rekultywowanego składowiska odpadów komunalnych. *Inżynieria Ekologiczna, 27*, 72-80 (in Polish).

Melnyk, A., Kuklińska, K., Wolska, L., Namieśnik, J. (2014). Chemical pollution and toxicity of water samples from stream receiving leachate from controlled municipal solid waste (MSW) landfill. *Environmental Research, 135*, 253-261.

Mizerna, K., Bożym, M. (2015). Monitoring jakości wód na składowiskach odpadów komunalnych. *Archiwum Gospodarki Odpadami i Ochrony Środowiska, 17*(4), 51-58 (in Polish).

Przydatek, G. (2012). Ocena sprawności systemu odwodnienia eksploatowanego składowiska odpadów komunalnych. *Infrastruktura i ekologia terenów wiejskich, 2*(3), 217-226 (in Polish).

Przydatek, G. (2013). Jakość wód podziemnych w rejonie użytkowanego składowiska odpadów. *Infrastruktura i ekologia terenów wiejskich, 1*(4), 59-70 (in Polish).

Przydatek, G. (2021). Using advanced statistical tools to assess the impact of a small landfill site on the aquatic environment. *Environ Monit Assess, 193*, 71.

Rozkrut, P. (2021). *Operat wodnoprawny na wprowadzanie ścieków przemysłowych zawierających substancje szczególnie szkodliwe dla środowiska wodnego do urządzeń kanalizacyjnych Tarnowskich Wodciągów Sp. z o.o.* (in Polish).

Rozporządzenie Ministra Gospodarki Morskiej i Żeglugi Śródlądowej z dnia 11 października 2019 r. w sprawie kryteriów i sposobu oceny stanu jednolitych części wód podziemnych (Dz.U. z 2019 r., poz. 2148). (2019a). (in Polish).

Rozporządzenie Ministra Gospodarki Morskiej i Żeglugi Śródlądowej z dnia 12 lipca 2019 r. w sprawie substancji szczególnie szkodliwych dla środowiska wodnego oraz warunków, jakie należy spełnić przy wprowadzaniu do wód lub do ziemi ścieków, a także przy odprowadzaniu wód opadowych lub roztopowych do wód lub do urządzeń wodnych (Dz.U. z 2019 r. poz. 1311). (2019b). (in Polish).

Rozporządzenie Ministra Środowiska z dnia 30 kwietnia 2013 r. w sprawie składowisk odpadów (Dz.U. z 2013 r., poz. 523) (in Polish).

Tałałaj, I.A. (2014). Uwalnianie metali ciężkich na wybranym składowisku odpadów komunalnych w trakcie roku kalendarzowego. *Rocznik Ochrona Środowiska, 16*, 404-420 (in Polish).

Wysocka, M.E. (2015). Wpływ lokalizacji składowisk odpadów na jakość wód podziemnych. *Rocznik Ochrona Środowiska, 17*(2), 1074-1093 (in Polish).

Żelezik, M. (2015). Analiza monitoringu składowiska odpadów komunalnych w Promniku w latach 2011-2012. *Proceedings of ECOpole, 9*(1), 365-372 (in Polish).

Jerzy KOROSTIL[1], **Olesya AFANASYEWA**[2], **Olga KOROSTIL**[3]

[1] University of Applied Sciences in Nowy Sacz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, e-mail: jkorostil@ans-ns.edu.pl
[2] Pedagogical University of Krakow Faculty of Mathematics, Physics and Technical Sciences, Podchorążych 2, 30-084 Kraków, e-mail: olesia.afanasieva@up.krakow.pl
[3] PSChem Sp. z o.o., Świętokrzyska 18, 00-052 Warszawa, e-mail: jkorostil3@gmail.com

Designing a forecasting system using its parameters

Abstract

The work solves the problem of calculating parameter values and their use in building a system for forecasting external, random events that negatively affect the functioning process implemented by a technical object. To clarify the functioning process, a number of conditions defining the boundaries within which the parameters are studied are proposed. The design of the system is performed in stages, and therefore the investigated parameters are analyzed at each stage separately.

The first stage of the process of designing a forecasting system is the formation of a hypothesis that determines the general purpose of using the system. The work offers an interpretation of the hypothesis that is adapted to the solved problem. A description of the process of preliminary analysis of input data is provided, and the processes related to the requirements for the components of the forecasting system, which expand its functionality, are considered. The peculiarities of the functioning of the component, which predicts a random event, are considered. The statement is proved, which defines the type of functions on the basis of which the prediction is realized. At the second stage, the processes of determining the values of the countermeasures against the impact of a random event on the functioning of the technological process are implemented. As part of these processes, textual descriptions of data interpretation are used to expand their functionality related to the implementation of the corresponding countermeasure. In this regard, the statement about the non-contradiction of the system of algorithms, which is modified based on the use of information from textual descriptions of the interpretation of data used in the forecasting system, is proved.

Keywords: forecasting, parameters, random event, determinism, hypothesis.

Projektowanie systemu prognozowania z wykorzystaniem jego parametrów

Streszczenie

Praca rozwiązuje problem obliczania wartości parametrów i ich wykorzystania przy budowie systemu prognozowania zewnętrznych, losowych zdarzeń, które negatywnie wpływają na proces funkcjonowania realizowany przez obiekt techniczny. Dla sprecyzowania procesu funkcjonowania zaproponowano szereg warunków określających granice, w których badane są parametry. Projektowanie systemu odbywa się etapami, w związku z czym badane parametry są analizowane na każdym etapie z osobna.

Pierwszym etapem procesu projektowania systemu prognozowania jest formułowanie hipotezy, która określa ogólny cel wykorzystania systemu. W pracy zaproponowano interpretację hipotezy dostosowaną do rozwiązywanego problemu. Przedstawiono opis procesu wstępnej analizy danych wejściowych oraz rozpatrzono procesy związane z wymaganiami dla elementów systemu prognozowania rozszerzających jego funkcjonalność. Uwzględniono specyfikę działania komponentu, który przewiduje zdarzenie losowe. Udowodniono stwierdzenie, które określa rodzaj funkcji, na podstawie których realizowana jest predykcja. Na drugim etapie realizacji procesu wyznaczania znaczeń parametrów przeciwdziałania wpływu zdarzenia losowego na funkcjonowanie technologicznego procesu wykorzystywani tekstowe opisy interpretacji danych do rozszerzenia ich możliwości funkcjonalnych związanych z realizacją odpowiedniego przeciwdziałania. W tym zakresie udowodnione jest stwierdzenie o spójności systemu algorytmów, który jest modyfikowany w oparciu o wykorzystanie informacji z tekstowych opisów interpretacji danych wykorzystywanych w systemie prognostycznym.

Słowa kluczowe: prognozowanie, parametry, zdarzenie losowe, determinizm, hipoteza.

1. Introduction

Forecasting, within the $SU(Pr_i)$ technological process control system, is used for the anticipating detection of a random event Vp_i with the aim of activating countermeasures against the influence of Vp_i on the technological process $Pr_i(TO_i)$. Let's assume that forecasting is used to service the technical object TO_i that functions in the environment (Ep_i) and activates the corresponding functionalization process $Pr_i(TO_i)$. We limit ourselves to predicting events that negatively affect $Pr_i(TO_i)$ that occur in Ep_i. To implement forecasting methods, initial data Df_i is formed on the basis of data on test results, input data Di_i is obtained as a result of monitoring Ep_i and operational data Dw_i, which are formed on the basis of using the results of preliminary analysis of input data and are directly used to implement forecasting processes. Let us assume that the data Di_i are closely related to the processes that generate events Vp_i in the environment Ep_i and their determination can be carried out in the following ways. The first method consists in increasing the amount of information about processes $Pr_i(Vp_i)$, which leads to the occurrence of Vp_i in Ep_i. This approach consists in conducting extended studies of the processes occurring in Ep_i. The second way is to use additional data about $Pr_i(TO_i)$, which can be used when predicting the event Vp_i. This approach is more related to $Pr_i(TO_i)$ and is therefore more relevant to the researched problem.

The implementation of forecasting requires the formulation of a hypothesis Hp_i regarding the events Vp_i, which are planned to be predicted. The hypothesis contains a description of the prediction goal Cp_i. In addition to the target description, Hp_i may contain some input Di_i. The prediction goal is dominant and describes Vp_i and its relationship with $Pr_i(TO_i)$ to a certain extent. The description Cp_i may represent the identifier Vp_i and some information $Iv_i(x_i)$ regarding the event and its relationship with $Pr_i(TO_i)$. The above Hp_i components are supplemented by textual descriptions of their interpretation, which are formed on the basis of data relating to the various stages of the TO_i tests. To build textual descriptions of the interpretation, specialists in the subject area form a semantic dictionary Sc_i, for the components of the forecasting system SPG_i, for anomalies An_i and events Vp_i that appear in Ep_i. Text descriptions related to SPG_i data and components are presented in the form: $j(x_{i,k}), j(y_i)$ and are used to implement the processes of their semantic analysis and transformations (Korostil, Afanasyewa, Korostil, 2021). In the process of conducting tests, modeling of possible Vp_i is used, which allows for establishing the corresponding threats to $Pr_i(TO_i)$. Based on the information obtained during the tests, the initial versions of the input data Df_i are formed. In the process of testing, not only possible attacks on $Pr_i(TO_i)$, but also means of countering the corresponding attacks are investigated.

Let's accept the following conditions.

Condition 1. The event Vp_i occurs in Ep_i, within the set time interval ΔTp_i, which is determined for Ep_i and coordinated with the process $Pr_i(TO_i)$.

Condition 2. The event Vp_i is not one-time and can be repeated within time intervals, the value of which may change during the functioning of $Pr_i(TO_i)$.

Condition 3. If Vp_i of different types occur in Ep_i, it means that Ep_i can be divided into components of the environment Ep_i^k, each of generates the corresponding types Vp_i^k.

The event Vp_i, which can occur as a result of the functioning of processes $Pr_i(Ep_i)$, is characterized by parameters related to the parameters $Pr_i(TO_i)$. The process of occurrence of Vp_i is characterized by the following factors:

- at the moment τ_i of the occurrence of Vp_i within the prediction interval ΔTp_i,
- by the values of the parameters y_i, which characterizes Vp_i and are commensurate with the parameters x_i, which counteract the negative influence of y_i on the technological process control system $SU(Pr_i)$.

Let's assume that Vp_i is determined by the occurrence of anomalies An_i in the Ep_i environment. Anomalies arise as a result of the functioning of processes $Pr_i(An_i)$. Anomalies An_i are characterized by deviations of the values of parameters y_i, which are eliminated during the operation of the control system of the corresponding process $SU(Pr_i)$. If the value of the parameters An_i increases and $SU(Pr_i)$ cannot eliminate An_i, then Vp_i occurs. From the first and second condition, it follows that the event Vp_i in relation to the intervals ΔTp_i, in general, is not random, because Vp_i within the framework of a separate interval ΔTp_i will necessarily occur. The sizes of consecutive forecasting intervals ΔTp_i during the process of functioning $Pr_i(TO_i)$ may be different, which is connected with the probability of occurrence of Vp_i during the established period $\mathfrak{D}(\Delta Tp_i)$, which may represent a separate cycle of functioning $Pr_i(TO_i)$. The need to use the period $\mathfrak{D}(\Delta Tp_i)$ is determined by the fact that establishing the probability of occurrence of Vp_i during $\mathfrak{D}(\Delta Tp_i)$ will allow to selection time intervals in which the occurrence of Vp_i has a sufficiently low probability. Such information can be used to form an optimal strategy for the functioning of $Pr_i(TO_i)$ during $\mathfrak{D}(\Delta Tp_i)$, due to the use of different modes of functioning of $Pr_i(TO_i)$ in these intervals.

2. Implementation of the stages of functioning of the hybrid forecasting system

The first stage of the implementation of the SPG_i forecasting system is the formation of the hypothesis $Hp_i(Vp_i)$, about the possibility of predicting the occurrence of a random event Vp_i, which negatively affects $Pr_i(TO_i)$ and implementing countermeasures against such negative impact (Korostil, Afanasyewa, Korostil, 2022).

Let's consider a possible version of the definition of the hypothesis.

Definition 1. We will call the hypothesis Hp_i a relationship in which, based on the use of information about the input data and on the basis of possible hypothetical transformations, the possibility of predicting the occurrence of Vp_i and the possibility of counteracting the negative influence of Vp_i on the corresponding $Pr_i(TO_i)$ are asserted.

Using the notion of An_i occurring in Ep_i and the possibility of occurrence of An_i event Vp_i, it is possible to directly associate An_i with Vp_i, which is written in the form: $An_i \rightarrow Vp_i$. In accordance with modeling concepts, the object model SPG_i in general, consists of components using the following elements (Zarubin, 2006):

- input data Di_i used for calculations and transformations,
- which is implemented by the forecasting system model $MO_i(SPG_i)$,
- algorithms of calculations and transformations $Al_i(SPG_i)$ reflecting the processes that occur in the system SPG_i,
- output data Dv_i or events $\mathcal{F}(\wp)$ characterizing the results of functioning $MO_i(SPG_i)$.

The $MO_i(SPG_i)$ model can describe the processes in the modeled object with varying degrees of approximation of the reproduction of the corresponding processes in relation to the processes that may take place in the case of the functioning of the real object. Therefore, it is appropriate to talk about the degree of closeness *(χ)* of the $MO_i(SPG_i)$ model to the real processes being modeled, which is formally described by the ratio: $NMO_i = \chi(MO_i)$, where NMO_i approximate model. The degree of closeness of an arbitrary model can also be determined by the degree of correspondence of the obtained modeling results to the real results of the functioning of the modeled objects. Based on definition (1), it can be assumed that the hypothesis Hp_i is an approximate model of $MO_i(SPG_i)$. The degree of its closeness to the model is minimal or $\chi[MO_i(SPG_i)] = min$, since it does not use, but only declares possible algorithms for transformations of Al_i input data Di_i. Thanks to this approach to the interpretation of the hypothesis Hp_i, it can be described in the form of $NMO_i(SPG_i)$, which significantly distinguishes Hp_i from the ideas about hypotheses used in probability theory (Billigslej, 2021; Gajek, Kałuszka, 2000). Based on the above, definition (1) can be presented in a form that corresponds to the ideas about NMO_i.

Definition 2. The hypothesis Hp_i represents the model MO_i of the system SPG_i, for which the relation $\chi[MO_i(SPG_i)] = min$ is fulfilled.

The given definition allows for a more constructive approach to the formation of the hypothesis Hp_i, which constitutes the first stage of the construction of SPG_i.

The next stage of implementation of the SPG_i system is the *SE* stage, or the stage of stipulation the event Vp_i. As part of this stage, processes are implemented, thanks to which the reliability of predicting the occurrence of Vp_i can increase to a certain extent. This is facilitated by preliminary analysis and transformations of input data Di_i into operational data Dw_i. The degree of reliability of the prediction process can be determined by the degree of correspondence of the prediction result in relation to the real occurrence in the current time interval ΔTp_i of the event Vp_i. Preliminary analysis methods can use different algorithms, depending on the overall assessment of the nature of the data and the chosen prediction algorithms. An example of preliminary analysis can be the process of filtering input data Di_i. An example of the prediction process is the use of the prediction function extrapolation method. Such a function describes the change in the increment value of the parameter $P_j(Vp_i)$ characterizing Vp_i per unit of time δt_i. The unit of time δt_i calculated based on data on the speed of the technological process $Pr_i(TO_i)$, which determines the real-time mode of operation of $Pr_i(TO_i)$. As part of this stage, the threshold value of $P_j(Vp_i)$ is also determined, within which the control system $SU_i(Pr_i)$ can counteract the negative influence of Vp_i on the process $Pr_i(TO_i)$. The $SU_i(Pr_i)$ system eliminates the unacceptable influence of the value of the parameter An_i, which has a negative effect on $Pr_i(TO_i)$, if the value of $P_j(An_i)$ does not exceed the set threshold. Data on the amount of change $\Delta P_j(Vp_i)$ can be used to construct the development process function of the corresponding An_i. This function describes the variation of $P_j(Vp_i)$ and is synchronized with the time intervals allocated to the extrapolation function used by the prediction process.

The *CE* event conditioning stage is intended to specify the requirements for the following functionality of the SPG_i system:

- prediction procedures;
- possibilities of counteracting the influence of Vp_i on $Pr_i(TO_i)$;
- a number of aspects characterizing the process of functioning of SPG_i;
- which relate to various components that can be part of the system.

The *CE* stage concerns most of the components involved in the functioning of the SPG_i system. Since the forecasting process is directly related to the components of SPG_i, the goal of forecasting $Cp_i \subset Hp_i(SPG_i)$ must determine certain requirements for the functioning of these components. Such requirements should extend the functionality of algorithms $Al_i(SPG_i)$. The use of NMO_i provides an opportunity to more constructively implement changes aimed at expanding the protection capabilities of $Pr_i(TO_i)$, from Vp_i. During the functioning of SPG_i, individual components of SPG_i may be modified. The requirements for such a modification can have a direct form in relation to the corresponding component, or it can take an indirect form due to the fact that the components of SPG_i are connected to each other by the corresponding structure.

The $SU_i(Pr_i)$ system implements the interaction management processes between individual components of the SPG_i systems. Such interaction can be foreseen in the algorithms of the corresponding components, which can also lead to the implementation of the modification of the goal, the description of which is placed in Hp_i. An important function of the $SU_i(Pr_i)$ system is solving the problems of modifying individual components of the SPG_i systems, which allows adapting the forecasting system to changes that may be caused by the environment in which the SPG_i functions. Fulfillment of such requirements leads to an increase in the degree of approximation of NMO_i to SPG_i. The SPG_i system provides the following possibilities for the implementation of $SU_i(Pr_i)$ protection processes:

- biased opposition to the possible negative influence of Vp_i on $SU_i(Pr_i)$;
- countering the negative influence of Vp_i on $SU_i(Pr_i)$ in the real-time mode of the functioning of $Pr_i(TO_i)$;
- formation of recommendations to avoid the negative influence of Vp_i on $SU_i(Pr_i)$ by changing the operating modes of $SU_i(Pr_i)$.

3. Functioning components of the prediction system prognostication

The *EF* stage is implemented by the prediction system SPR_i of the event Vp_i. The main processes of this stage include the following:

- determining the probability and intensity of the occurrence of Vp_i;
- implementation of prediction function extrapolation;
- definition of the prediction interval ΔTp_i.

To determine the probability $Pb_i(Vp_i)$ and the intensity $It_i(Vp_i)$ of the occurrence of the event Vp_i, at the initial stage of the operation of the process $Pr_i(TO_i)$, data obtained in the process are used conducting $Pr_i(TO_i)$ tests (Bidiuk, Romanenko, Tymoszczyk, 2013; Dittmann, 2008). During the operation of $Pr_i(TO_i)$, events Vp_i are monitored, data from which are used to modify the current value of the parameters $Pb_i(Vp_i)$ and $It_i(Vp_i)$ which characterize the occurrence of event Vp_i. The values of these parameters may change during the operation of $Pr_i(TO_i)$. Consider the following statement.

Statement 1. Random processes $Pr_i(An_i)$ in the environment Ep_i, which lead to the occurrence of anomalies An_i and the rise of random events Vp_i based on them, are described by exponential functions.

This statement is based on the ideas about the events of An_i and Vp_i. The speeds of the processes $Pr_i(TO_i)$, $Pr_i(An_i)$ and $Pr_i(Vp_i)$ are compared based on the determination of the change in the values of the co-dimensional parameters of the processes $Pr_i(TO_i)$ and $Pr_i(Vp_i)$. Such changes in the value of the parameters that characterize them can be considered the speeds of the corresponding processes, which will be denoted by $Sp_i(TO_i)$ and $Sp_i(Vp_i)$. Since it is assumed that the event Vp_i is random and negatively affects $Pr_i(TO_i)$ during its operation, it means that $Sp_i(Vp_i) \gg Sp_i(TO_i)$. Let's choose some conditional interval of the functioning of processes $Pr_i(TO_i)$ and $Pr_i(Vp_i)$, which we will denote by $\Delta t_i(TO_i)$and $\Delta\tau_i(Vp_i)$. Since the relation $Sp_i(Vp_i) \gg Sp_i(TO_i)$ holds, it can be assumed that the interval $\Delta t_i(TO_i)$ is described by a linear function, for example: $Pr_i(TO_i)= k + \gamma t$, where *k* and *γ* are coefficients and *t* is an independent time variable of the functioning process. In order for the condition $Sp_i(Vp_i) \gg Sp_i(TO_i)$ to be fulfilled, it is necessary that the projection of the interval $\Delta\tau_i(Vp_i)$ on $\Delta t_i(TO_i)$, while ensuring the necessary increase in $Pj_i(Vp_i)$ was shorter than the interval $\Delta t_i(TO_i)$. Otherwise, the necessary requirements for the relevant processes will not be fulfilled. To prevent this from happening, it is necessary that the trajectory of $Pr_i(Vp_i)$ in relation to the trajectory of $Pr_i(TO_i)$ at the selected time interval is described by an exponential type function, for example: $\varphi e = \alpha e^{\beta t}$, where α and β are coefficients. The coefficients are selected when forming the description of the interpretation of the processes $Pr_i(TO_i)$ and $Pr_i(Vp_i)$, *t* is an independent variable of the exponential function that determines the time of the functioning process. The independent variable t_i changes in accordance with the operation time of the corresponding process and cannot take negative values. The coefficients $\alpha \geq 0$ and $\beta \geq 1$ of the function $Pr_i(Vp_i)$ varies from 0 to ∞, which corresponds to the assumption that the event Vp_i within the selected interal $\Delta t_i(TO_i)$ is necessarily will come. The exponential function, in accordance with the given conditions, takes exclusively positive values, which corresponds to the semantics of the occurrence of Vp_i and condition (1). This proves the statement that to predict the occurrence of Vp_i by the method of extrapolation, you have to choose an exponential function.

The definition of the prediction interval ΔTp_i is based on the analysis of its components, which include subintervals.

1. The subinterval determined by the time of development of the next anomaly $\Delta\tau_i(An_i)$, which can turn into Vp_i.
2. The subinterval of direct prediction described by the extrapolation function $\varphi e(t) = \alpha e^{\beta t}$.

When designing any $Pr_i(TO_i)$, the technical requirements for the corresponding object are determined (Bidiuk, 2004; Dechter, 2003). For a certain type of $Pr_i(TO_i)$, in addition to the requirements for the reliability of the functioning of $Pr_i(TO_i)$, requirements for stability in relation to the selected Vp_i are formed. Such requirements provide initial data on the values of the probability of occurrence of Vp_i and the possibility of their influence on $Pr_i(TO_i)$, which should be countered by the SZO_i, protection system. They are formed based on the use of information obtained in the conditions of simulation of the impact on TO_i of possible Vp_i. Such data are the probability of occurrence $Pb(Vp_i)$ and their intensity $It(Vp_i)$. Using $Pb(Vp_i)$, it is possible to determine the time intervals during which the occurrence of Vp_i has certain probabilities. Such intervals in the theory of probability are called confidence intervals, the definition of which requires the distribution function $Pb(Vp_i)$. The value of the corresponding confidence interval is determined as a percentage in relation to the interval on which the distribution function $Pb(Vp_i)$ is determined and based on the data obtained during the $Pr_i(TO_i)$ and SPG_i tests. Data on the intensity of occurrence of events Vp_i, are used to determine the starting points of the forecast intervals, (Chung, AitSahlia, 2003).

The interval of direct prediction of the event Vp_i is determined by the time interval $\Delta\tau(\varphi e)$, during which this process is described by a function φe. For convenience, let's assume that the function φe is active during the time interval $\Delta\tau_i$ *(φe)*. The operation process φe is synchronized with the moments of time, which corresponds to the initial moment of the prediction interval ΔTp_i. Such synchronization is necessary because anomalies An_i develops on the interval $\Delta\tau_i(An_i)$, which can lead to the occurrence of Vp_i. The time interval $\Delta\tau_i(An_i)$ begins at the moment when the activity of An_i increases. Such activity is determined by the value of the current value of the key parameter An_i and is confirmed by the increase in the growth rate of the value of this parameter. The moment of transition from $\Delta\tau_i(An_i)$ to $\Delta\tau_i(\varphi e)$ is determined by the change in the value of the threshold $Pj(An_i)$, after which $SU_i(Pr_i)$ cannot eliminate the corresponding anomaly An_i. The ratio for determining ΔTp_i can be written in the form: $\Delta Tp_i = \Delta\tau_i(An_i) + \Delta\tau_i(\varphi e)$.

4. Parameters for counteracting the negative impact of random events

The main parameters characterizing the effectiveness of using the hybrid forecasting system for the protection of $Pr_i(TO_i)$ include:

- parameter of the degree of determinism of the processes (λ) implemented in the system SPG_i;
- parameter μ, measures of the influence of countermeasures against threats, for $Pr_i(TO_i)$, conditioned by the occurrence of Vp_i;
- parameter of the prediction performance measure (π) provided by SPG_i.

The degree of determinism λ of the SPG_i system depends on a number of factors that are determined to one degree or another by various components of the forecasting system. Therefore, let's divide such factors into the following groups:

- the first group of factors represents numerical data, the interpretation of which is to some extent adequate for the system;
- the second group of factors is usage results data of algorithms used in SPG_i;
- the third group of factors refers to textual descriptions of numerical data and algorithms of the SPG_i system.

Initial data obtained from known sources do not require expansion of semantic components due to the fact that they are used only once when activating the process of system functioning.

The determinism measure parameter λ, in most cases, depends on the factors of the third group, the elements of which are textual descriptions of data interpretation and system components.

The transformation of data Dj_i at the stage of input analysis is carried out within the framework of their mathematical interpretation, which can determine the clarification of data with the aim of possibly increasing the efficiency of their use. In this case, the efficiency of data use leads to an increase in the degree of determinism. All other types of numerical data used in the system correspond to the description of the interpretation, the elements of which are presented in the semantic dictionary Sc_i. Let's assume that the expansion of their semantics can be determined by the situation when, as a result of transformations, additional data appear, which determines the need to expand their semantics.

The textual descriptions of the interpretation of the algorithms used within SPG_i can also be modified during their analysis and transformations.

Modification and expansion of textual descriptions of the interpretation of system components is the main factor in changing the value of the semantic parameters of the elements to which they refer (Hobson, Cole, Hannes, 2022; Thomas, 2020). Such changes may occur as a result of their formal transformations and expansions by elements existing in the dictionary Sc_i. Each element Sc_i has its own semantic significance $\sigma^Z(x_i)$. Text descriptions of interpretation TOI_i consist of informational elements $In_i(x_i)$ that describe data and elements that describe algorithmic means $Al_i(x_i)$. Fragments of the text description of the elements of the algorithmic tools are used to describe the processes of implementation of the functioning of SPG_i. Thus, the text descriptions of SPG_i interpretation not only describe the numerical data of the system, but also describe the algorithmic capabilities of SPG_i functioning processes.

The expediency of using TOI_i in technical systems is determined by the fact that such descriptions have wide possibilities in displaying the processes occurring in the system and the data it uses. The range of such possibilities is determined by the requirements for methods of normalization of texts representing TOI_i. For ease of use of TOI_i, let's assume that the text description $Al_i(x_i)$ represents the text model $TMO_i(Al_i)$ of the corresponding $Al_i(x_i)$. Any TO_i and corresponding processes $Pr_i(TO_i)$ can be described with varying degrees of completeness. The completeness of the description in TMO_i of the process $Pr_i(TO_i)$ is determined by a more or less detailed description of the functionality of individual components in TO_i, which may consist of individual elements. Such elements $\hbar_{ij}[Pr_i(TO_i)]$ can be replaced, eliminated, or added.

We will show that the transition from TOI_i to TOI_i^*, which is determined by the result of the analysis and transformations of text descriptions in TMO_i if the corresponding logical model $L(Al_i)$, will not lead to the emergence of a contradiction in the logical model $L_i^*(Al_i^*)$ and, as a result, in the system of means implementing the process of functioning of the SPG_i system.

Consider the following statement.

Statement 2. The extension $L(Al_i) \to L_i^*(Al_i^*)$ as a consequence of the modification $TMO_i(Al_i) \to TMO_i^*(Al_i^*)$, will not lead to a contradiction within the model $L_i^*(Al_i^*)$.

To prove the statement, consider the transformation of the model $TMO_i^*(Al_i^*)$, into the corresponding logical model $L_i^*(Al_i^*)$ (Korostil, 2012). Let us show that $L_i^*(Al_i^*)$ is not contradictory. The transformation $TMO_i(al_i) \to TMO_i^*(al_i^*)$, is implemented based on the use of methods of analysis of the corresponding text descriptions of individual elements $al_i \in Al_i$ and their groups. In this analysis, the semantic parameters of the textual description characterizing individual elements $TOI_i(al_{ij}) = j(al_i)$ are used. The unary type semantic parameter is the semantic significance parameter σ^Z for $j(al_i)$. The values of unary semantic parameters are determined on the basis of data on the subject area in which $Pr_i(TO_i)$ functions, and are recorded in Sc_i. The structure of text descriptions consists of descriptions of elements $j(al_i)$ that correspond to the syntactic rules of the grammar of the selected language $\Gamma(Sy)$, which provide the types of text structure accepted in it. To ensure the necessary semantics of the texts, in the grammar of type Γ(Sy) binary semantic parameters are used, for example, the parameters of semantic consistency σ^U, semantic contradiction σ^S and others. Therefore, the textual description of the interpretation $TMO_i(Al_i)$ corresponds to the syntactically permissible structures Γ(Sy) of the selected grammar and provides an appropriate way of forming the text structure adopted in it. The use of semantic parameters to form $TMO^*_i{}_i(Al^*_i)$ allows to ensure the semantic consistency of the description of the interpretation Al_i. The textual description of the logic interpretation Al_i corresponds to the algorithms that represent the model $TMO_i(Al_i)$, which describes the corresponding processes $Pr_i(TO_i)$. Thus, the transformation $TMO_i(Al_i) \rightarrow TMO^*_i{}_i(Al^*_i)$ does not lead to contradictions in $TMO^*_i{}_i(Al^*_i)$, since the expansion of the text description of the interpretation meets the semantic requirements of their interpretation and the syntactic requirements of Γ(Sy). In addition, the extension elements are semantically consistent within the framework of $TMO^*_i{}_i(Al^*_i)$ due to the use of the corresponding semantic parameters σ^P and do not go beyond the scope of their interpretation. Since $TMO^*_i(Al^*_i)$ is not contradictory, when transforming $TMO^*_i(Al^*_i) \rightarrow L^*_i(Al^*_i)$ we get the model $L^*_i(Al^*_i)$, which is non-contradictory, since it represents a non-contradictory model $TMO^*_i(Al_i)$, presented in a shortened form, which uses semantically consistent element notations, which proves the statement.

The use of the μ parameter, which characterizes the degree of counteraction to the negative influence of Vp_i on $Pr_i(TO_i)$ is aimed at determining the practical degree of implementation of counteraction to the influence of Vp_i on $Pr_i(TO_i)$. The use of this type of parameter in TO_i control systems is mandatory, because within the framework of the SPG_i system it can be considered an analogue of feedback parameters, which in deductive control systems enable the system to adapt to changes occurring in $Pr_i(TO_i)$. The μ parameter also allows for the analysis of the level of security of the $Pr_i(TO_i)$ functioning process in case of external threats. The parameter μ is dimensionless and takes values in the range [0,1]. When $\mu \rightarrow 1$ it means that for all threats from Vp_i the protection system SZO_i provides the necessary protection. If $\mu = 0$, it means that in relation to threats from Vp_i, the system SZO_i does not provide protection for $Pr_i(TO_i)$. In this case, the corresponding $Pr_i(TO_i)$ may go into a catastrophic state. If in the process of functioning of $Pr_i(TO_i)$, the value of μ decreases, it means that new Vp_i arise, the detection and countermeasures of which require the expansion of the capabilities of protection means, or the use of additional protection means.

The determinism parameter λ of the $\mathcal{K}_i$ component of the SPG_i system determines the extent to which the distribution range of the initial numerical data decreases in relation to the distribution value of the input numerical data, which can be interpreted as the values of their mathematical expectations $M_i(D_v)$ and $M_i(D_w)$. But such distribution, in this approach, is expanded and determined by two components of the information element $In_i(x^k_i)$. The first component is the value by which the mathematical expectation $M(D^{\mathcal{K}}_{w,v})$ of the numerical data obtained at the output of the components $\mathcal{K}_i$ is refined. The second component is the amount of increase in the semantic significance of the text model $RTM_i(\sum^r_j TOI_j)$, where TOI_j is a textual description of the elements of the source data $D^{\mathcal{K}}_v$, or other transformation results implemented by the component $\mathcal{K}_i$. This component represents the amount of change in the sum of parameters of semantic significance $\sigma^Z[j(x_i)]$ of all elements $TOI_i \in Sc_i$, which make up the text model of the transformation result. Such a change is carried out by the $\mathcal{K}_i$ component in relation to the values of the input data values, which leads to a semantic expansion of the interpretation of these data. Changing the value of the sum of parameters $\sigma^Z[j(x_i)]$ can be done by replacing elements in TOI_i, for example, from $\sigma_i^Z[j(x_i)]$ to $\sigma_k^Z[j(x_k)]$, for which $\sigma_k^Z > \sigma_i^Z$, and by adding new elements from Sc_i to TOI_i , which agree semantically with the elements in TOI_j within its syntactic structure. For the next steps of analysis and transformations implemented in the $\mathcal{K}_i$ component, not only the numbers Dv_j, but also descriptions of modifications or changes in interpretations that occurred with the data Dv_j are transmitted. Models of such textual descriptions $TMO_j(Al_k)$, allow modification of data transformation in $\mathcal{K}_i$ components with the aim of increasing their degree of determinism.

If $\lambda(\mathcal{K}_i) = 0$, then the component $\mathcal{K}_i$ does not implement determinism changes in the component $\mathcal{K}_j$. If $\lambda(\mathcal{K}_i) > 0$, then transformation $\mathcal{K}_i$ increases the amount of determinism due to refinement of Dv_j, or replacement or addition of elements $\mathcal{k}_{ij}(Pr_i)$ in accordance with the following ratio:

$$\lambda(\mathcal{K}_i) = \sum_q^m \sigma^{Z*}[j(x_i)] + \Delta MO_i(D_v).$$

The quantity $\lambda(\mathcal{K}_i)$ is a dimensionless numerical quantity, $\sigma^Z[j(x_i)]$ and $\Delta MO_i(D_v)$ are dimensionless quantities. The component $\Delta MO_i(D_v)$ describes the change in the number of text elements in $MO_i^*(D_v)$, if the text description is expanded.

The efficiency measure π is an integral parameter that combines all the factors taken into account in each individual case of system design SPG_i. For example, the proportionality parameter η, which characterizes the possibility of interaction between parameters of negative influence and parameters of counteraction to this influence, can be taken into account only at the testing stage. Then the determination of the value of the parameter π can be limited only to the values of the parameters $\lambda(\mathcal{K}_i)$ and μ, which can be written in the form: $\pi(SPG_i) = \sum_{i=1}^m \lambda(\mathcal{K}_i) + \sum_{j=1}^k \mu(y_j, x_j)$, where $\mathcal{K}_i$ is the component that realizes the change of determinism in the process of functioning of SPG_i, x_j – is a parameter that characterizes the process of counteracting the corresponding parameter y_j, which characterizes Vp_i and exerts a negative influence on SPG_i.

5. Conclusions

In the work, in accordance with the stages of designing a forecasting system, an analysis of parameters and methods of their determination is carried out. At the first stage of design, a hypothesis is formed about the possibility of predicting the occurrence of a random event that negatively affects the process of managing a technical object. The proposed and described interpretation of the hypothesis provides the possibility of its more effective use. Such an interpretation defines a hypothesis as a description of a forecasting model characterized by a minimal degree of its closeness to the modeled system.

At the stage of stipulation, the occurrence of a random event, the processes of preliminary data analysis are implemented, which allows to increase the efficiency of their use. At this stage, it is possible to determine the magnitudes of the covariance values of the random event parameters with the parameters that characterize the resistance of the technical object to their negative impact. At the stage of condition, the methods of specifying the goal of forecasting are considered, which allows taking into account the features of countermeasures against the negative impact of a random event on a technical object.

At the stage of predicting the occurrence of a random event, the method of implementing this process is investigated. In this regard, the statement about the need to use extrapolation methods, which are implemented on the basis of the use of exponential functions, is proved.

The paper analyzes the method of using parameters related to counteracting the negative impact of random events on the process of functioning of a technical object. One of the key parameters characterizing the degree of, which is the description of its textual interpretation. Changes in the text description make it possible to modify the algorithms of data transformations, which are implemented in the corresponding component. The paper proves the statement that the modification of text descriptions within the framework of the described system does not lead to contradictions in the modified algorithms used in the forecasting system.

References

Bidiuk, P.I. (2004). Modelowanie i prognozowanie geteroskedestycznych procesów. *Badania systemowe i informacyjni technologii, N1*, 115-124.

Bidiuk, P.I., Romanenko, B.D., Tymoszczyk, O.L. (2013). *Analiza szeregów czasowych*. Kijów: NTUU „KPI".

Billingsley, P. (2021). *Prawdopodobieństwo i miara*. Warszawa: Wydawnictwo Naukowe PWN.

Chung, K.L., AitSahlia, F. (2003). *Elementary Probability Theory*. New York: Springer.

Dechter, R. (2003). *Constraint processing*. San Francisko, CA: Morgan Kaufmann Publishers.

Dittmann, P. (2008). *Prognozowanie w przedsiębiorstwie. Metody i ich zastosowanie*. Kraków: Oficyna Wolters Kluwer Business.

Gajek, L., Kałuszka, M. (2000). *Wnioskowanie statystyczne, modeli i metody*. Warszawa: Wydawnictwa Naukowo-Techniczne.

Hobson, L., Cole, H., Hannes, H. (2022). *Przetwarzanie języka naturalnego w akcji*. Warszawa: Wydawnictwo Naukowe PWN.

Korostil, J., Afanasyewa, O., Korostil, O. (2021). Basic concepts of text models and their possible usage. *Journal of Engineering, Energy and informatics, 1*(1), 105-115.

Korostil, J., Afanasyewa, O., Korostil, O. (2022). Forecasting in technical facility control systems. *Journal of Engineering, Energy and Informatics, 1*(2), 63-75.

Korostil, O.J. (2012). Analysis of methods of interpretation of text models. *Collection of scientific works of the IPME named after G. E. Pukhova, National Academy of Sciences of Ukraine. Kyiv, 62*, 81-90.

Thomas, T. (2020). *Natural Language Processing with Spark Nlp: Learning to Understand Text at Scale*. O'REILY MEDIA, Wydawnictwo Helion.

Zarubin, V.S. (2006). *Mathematical modeling in engineering*. Moskwa: MSU Publishing House, Bauman.

Daniel DROZD[1], **Zenon JABŁOŃSKI**[2], **Grzegorz LITAWA**[3]

[1] University of Applied Sciences in Nowy Sącz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, e-mail:ddrozd@ans-ns.edu.pl
[2] University of Applied Sciences in Nowy Sącz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, email: zjablonski@ans-ns.edu.pl
[3] University of Applied Sciences in Nowy Sącz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, email: glitawa@ans-ns.edu.pl

Recognition of people in an IoT system with limited resources

Abstract

The Internet of Things (IoT) is currently a rapidly growing field of science. It is often the case that devices working in IoT cannot exceed predetermined dimensions. The small size does not allow the use of large computing power. For this reason, we are dealing with the use of Internet of Things in resource-limited environments. In this paper, a study was conducted to test the accuracy of face recognition. By analyzing the results obtained in this paper, the reliability of the solutions ob-tained with the library used can be checked.

Keywords: face recognition; IoT; microcontroller; IoT system with limited resources; OpenCV; recognition accuracy; cybersecurity.

Rozpoznanie osób w systemie IoT z ograniczonymi zasobami

Streszczenie

Internet rzeczy (IoT) to obecnie dynamicznie rozwijająca się dziedzina nauki. Często zdarza się, że urządzenia działające w ramach IoT nie mogą przekraczać wcześniej określonych rozmiarów. Niewielkie rozmiary nie pozwalają na użycie dużych zasobów obliczeniowych. Z tego powodu mamy do czynienia z wykorzystaniem Internetu rzeczy w środowiskach o ograniczonych zasobach. W niniejszym artykule przeprowadzono badanie w celu sprawdzenia dokładności rozpoznawania twarzy. Analizując wyniki uzyskane w tym artykule, można zweryfikować niezawodność rozwiązań uzyskanych przy użyciu używanej biblioteki.

Słowa kluczowe: rozpoznawanie twarzy; Internet rzeczy (IoT); mikrokontroler; system IoT z ograniczonymi zasobami; OpenCV; dokładność rozpoznawania; cyberbezpieczeństwo.

1. Introduction

Devices that are part of the "Internet of Things" are increasingly becoming an integral part of our live. They allow us to save our time, automate processes or increase our safety. Smart homes are becoming something normal and are more and more often taken into account when planning investments. The growing number of companies producing IoT equipment, and the increasing affordability of this type of solutions, resulted in the dynamic growth of the industry. The aim of the article was to check the operation of the automated facial recognition system and recording using devices with limited performance. Checking the operation of the system on the most popular Raspberry Pi microcomputers. Integration with the Xbee module and the STM32 microcontroller, through which you can integrate the proposed system with IoT devices operating in sensor networks. Additionally, a database of 15 actors was prepared, 4 people were selected from them, and the behavior of the system was examined in more detail.

2. Materials about IoT – Internet of Things

But what IoT really is? IoT, or Internet of Things, is a network of interconnected things. Put simply, it will be a network of interconnected "smart" devices. They can be equipped with many sensors used for monitoring the environment, which permit for in-dependent operation but their full potential is revealed when they are connected in a net-work. It is estimated that the current number of IoT devices is around 20 billion world-wide. This is a great amount that gives us to understand, how commonly IoT is used (Arora et al., 2021; Alam et al., 2020).

Sample and the most common places of IoT application will be devices from the "smart home" family. Everything, starting of intelligent lighting, automatic heating system, air conditioning control, and so on, adds up to so-called smart home. The possibility of creating "scenes" gives almost unlimited possibilities of personalizing all scenarios.

There are many off-the-shelf platforms that allow us to easily build IoT hardware. The Arduino platform is undoubtedly the most popular for its ease of use. Numerous libraries and community support mean that when we encounter a problem, we can solve it quickly with the help of the Internet. Another popular platform is ESP32, equipped with Wi-Fi and Bluetooth, which is perfect for creating Internet of Things devices. Other plat-form Raspberry Pi with high computing power finds its application in places requiring a large number of resources. It is no different in the case of facial recognition – it is a complicated process that requires a camera and a strong computing unit. The next of the most popular platforms is the STM32. These are 32-bit processors on the ARM core. Depending on the needs, we meet high-power microcontrollers, with a large number of peripherals and energy-saving. Producer shares configurator in which everyone will surely find something that will meet his expectations.

As in the case of microcontrollers, there are also many solutions to facilitate communication between individual devices. The obvious ones will be Wi-Fi and Bluetooth, but in the case of IoT, other communication interfaces have gained the greatest popularity. The most common communication standards in the case of the Internet of Things are: ZigBee, Z-Wave, Threed, standard cellular network, Sigfox or LoRa (Saravanan et al., 2021; Hancke et al., 2016; Ziegler, 2019).

ZigBee is the most widespread standard in the area of smart home devices. Based upon the IEEE 802.15.4 standard communication on the 2.4 GHz band is characterized by a mesh topology that guarantees stability and a long connection distance (up to 100 m). Thanks to the securing with a 128-bit cryptographic key, communication is adequately protected against unauthorized access and error detection allows for possible correction. ZigBee Green Power is an energy-saving version, designed for battery-powered devices (Yasuura et al., 2017; Ahmed, 2018; Suganthi, 2021).

Z-Wave is a standard developed in 1999 by the Danish company Zensys. It operates at frequencies below 1 GHz, which makes it resistant to interference from other devices communicating via Bluetooth or Wi-Fi. One network can be composed of at most 232 devices. The mesh topology used in this solution significantly increases the range, which in communication between two devices is up to 100 meters in an open space (Veneri et al., 2018).

3. Research methodology

The main objective of this research is to recognize people in IoT systems with limited resources. IoT devices are typically characterized by low computational power, which can repeatedly generate limitations. By using wireless connectivity and additional hardware to support higher computation, this can be prevented.

4. Recognition of people in an IoT system with limited resources

In order to run on limited resources, a team of devices consisting of a raspberry pi 4 responsible for image capture and STM32 responsible for data transmission and storage was created. In our research, we used Python language to implement the face recognition system. It is a high-level programming language and an open source project managed by the Python Software Foundation.

OpenCV is one of the most popular libraries for real-time image analysis. It was initiated by Intel, based on open source. The big advantage is multiplatform, which allows easy implementation regardless of the operating system. OpenCV can be used in:

- face recognition;
- objects detection;
- set of 2D and 3D tools;
- evaluation of emotions;
- understanding of movement;
- stereo vision from two cameras, taking perception into account;
- motion tracking;
- segmentation and recognition;
- augmented reality;
- gesture recognition;
- human-computer interaction (HCI);
- mobile robotics.

Face recognition is a library that allows us to quickly implement face recognition. It includes many useful functions such as tagging faces in images, recognizing faces and describing who a person is – based on previously learned patterns (Chen et al., 2020; Li et al., 2011).

Facial recognition is an involuntary activity performed by a human. To our brain, this seems like a simple and effortless activity. In the case of computer and electronics, it looks much more complicated. The program needs to be provided with images for analysis, for which a camera is necessary. All of them must be analyzed and then properly interpreted. The project used the prototype kit STM32, RaspberryPi and XBee modules, which helped in communication between individual devices. It is presented in the diagram below (Figure 1).

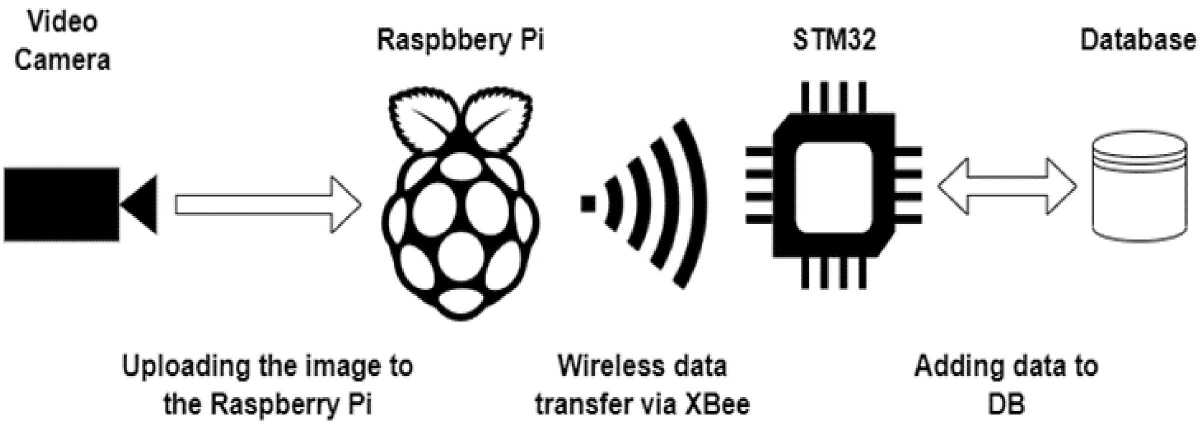

Figure 1. Diagram showing communication

The first step in the facial recognition is finding the face in the image. For this purpose, a histogram of oriented gradients, called HOG in brief, is created. The illustration below shows the traditional photo and the visualization of the histogram (Figure 2).

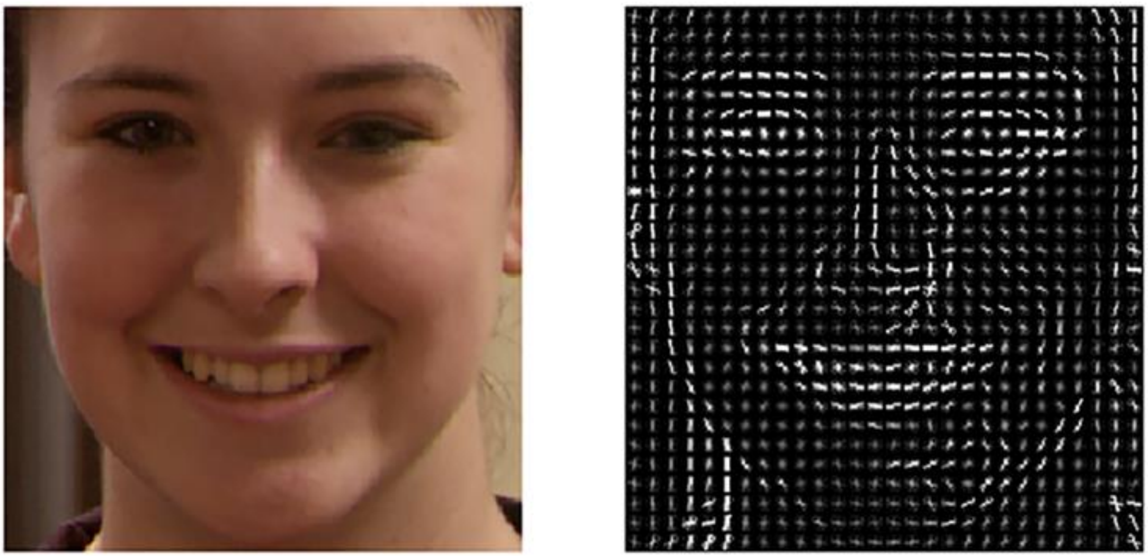

Figure 2. Photography and histogram representation of oriented gradients
(source: https://peerj.com/articles/cs-236/)

Each photo is a set of pixels. When a HOG is created, each pixel is analyzed in the following way:

1. We "look" at each pixel together with the adjacent pixels.
2. "We draw an arrow" to the darkest adjacent pixel.

The biggest advantage of a gradient is that regardless of the brightness of the image, it does not change. Depending on the system performance, the image is analyzed in larger or smaller blocks. The image processed in this way will be used in the face recognition and dlib libraries in order to find the face.

The next step is to create 128 values with the help of a convolutional neural network. This model was previously trained on a group of test faces, so that the newly "delivered" face should always return almost the same vector. The vectors encoded in this way are saved to the database and then the appropriate person is assigned to them (Kumar et al., 2020; Chantzis et al., 2021) (Rohlf, 2011; Bolle et al., 2008; Ślot, 2008).

The last step in recognition of the person's identity is comparing previously saved vectors. If the distances between them are not significant, we can determine with a high degree of confidence who the person currently being analyzed is.

The information collected in this way is encoded and then sent using the XBee module to the server, which saves it in the database. The data has been formatted in such a way to provide as much information as possible in one string. Individual parameters are separated by a previously defined separator (Li et al., 2011; Wechsler, 2009; Silhavy, 2020). The data we provide is:

1. Identifier of the recognized person.
2. The distance that separates it from the entries assigned to it.
3. Identifier of the second closest recognized person.
4. The distance that separates it from the entries assigned to it.
5. Number of entries in the database that were correctly assigned before the incorrect categorization occurred,
6. (In the case of working on photos) Name of the tested file.

All these parameters allow for even more accurate recognition and efficient testing of new entries, people, etc. The following algorithms are a very good illustration of how the program works (Figure 3).

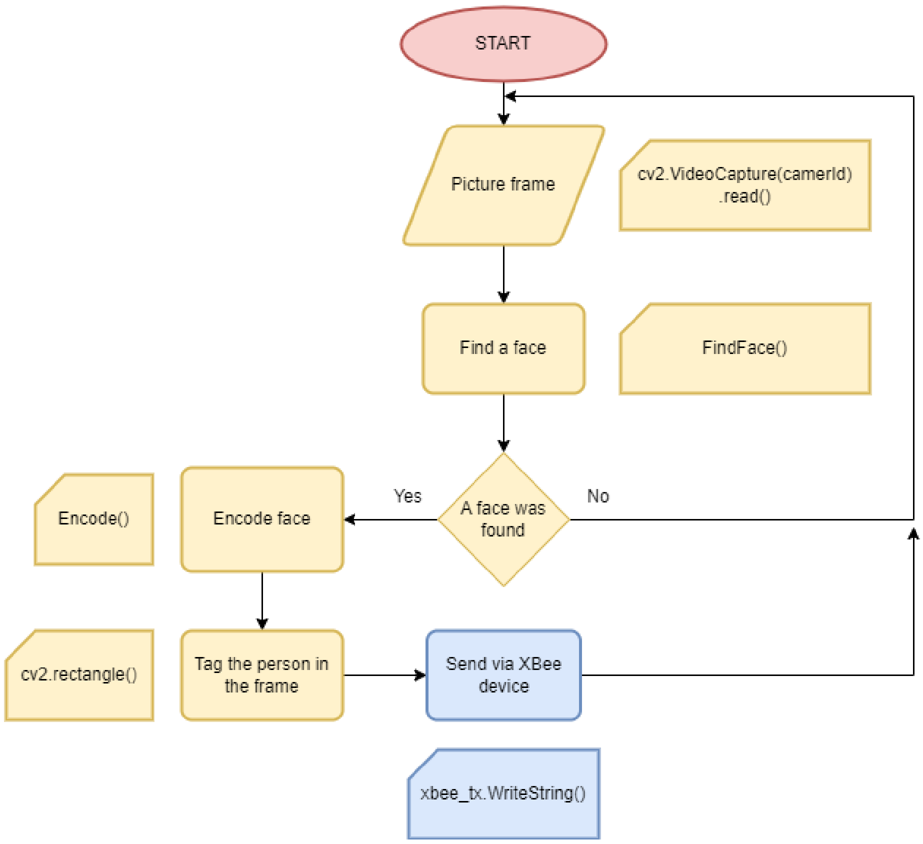

Figure 3. The general operating algorithm

At the moment of receiving the information, STM32 checks what time interval has occurred between the successive identical entries. If it is greater than a minute, information is saved to the database, otherwise it goes to the beginning of the program (Figure 4).

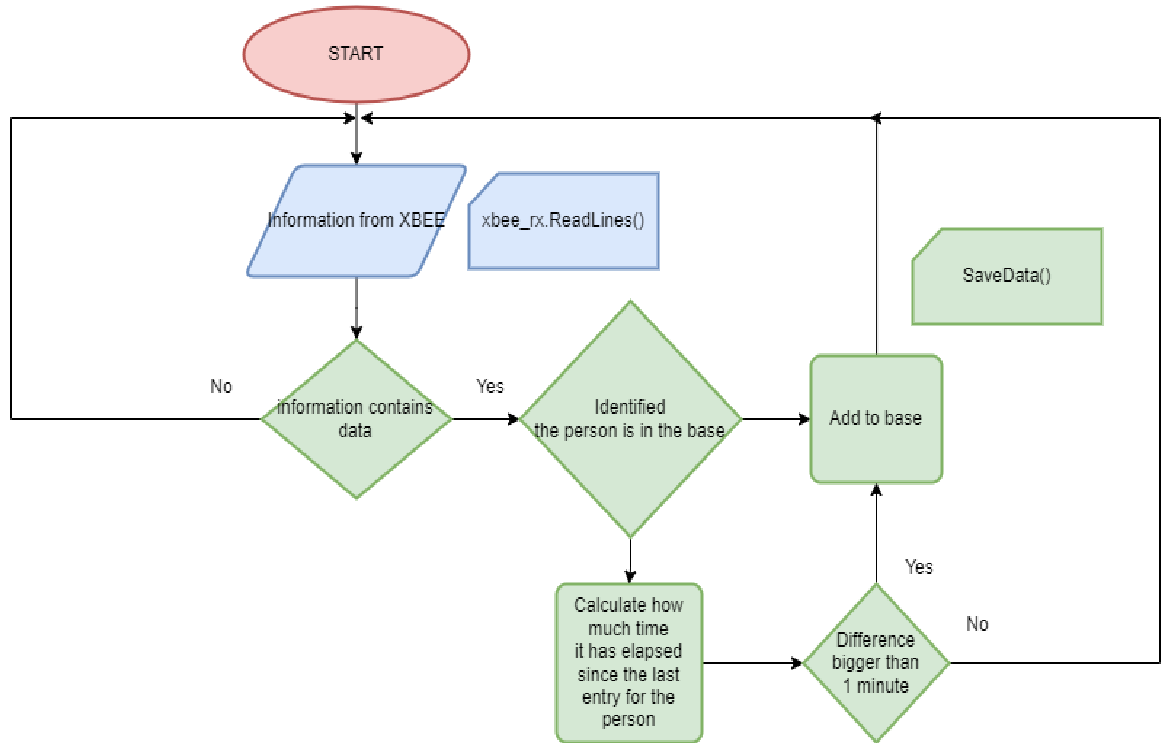

Figure 4. The algorithm of the program that puts information into the database

5. Analysis and test of the developed system

The system created for demonstration purposes provides a basis for creating interesting new functionalities. The example scenario will illustrate many possibilities of its application. For example, when a person saved in the database enters the house, we can adjust the temperature inside to individual preferences of the person entering, and when the person leaves the building, an economic heating or cooling plan will be activated. When it comes to inventing new scenarios, the only limit is the creativity of the creators.

In order to create a database of people to be used for research purposes, photographs of famous people downloaded from the Internet were used. The data of ten people were collected. Each entry in the database was generated from the ten photographs.

The tests were conducted with the help of two to five photos. To make it easier to test the application's capabilities, it was adapted to work with photos. This allows both the creation of a larger base of people and faster testing. The program has several modes and depending on the needs, one can run it with different parameters. Two prepared web interfaces present the test results. The first view shows the records from the database. The information is stored in five columns. The first one gives the date and time when a person was spotted. If someone is in the field of view of the camera for less than 60 seconds, the program will register his/her presence only once. Otherwise, their presence will be registered every minute. During this time the program can easily record the presence of other, non-repeating people. The next column stores the file name in case the application is testing with photo files. This facilitates later analysis of the collected information. The third column stores information about who was recognized, in this case in the photo. The next value is the distance. As described earlier, the distance tells us how similar the person is to the person to whom it was assigned. The last column is the rating. It displays the percentage of certainty in a given analysis and the previously described ratio of photos identified before the appearance of the most similar person (Table 1).

Table 1
Web interface showing entries from the database

File	Who	Distance	Rate
target1.jpeg	Jessica Alba	0.1654938	93.333% (14/15)
target2.jpeg	Angelina Jolie	0.3052380	90.0% (9/10)
target3. jpeg	Jcssica Alba	0.32S9543	100.0% (15/15)
target11.jpg	Jessica Alba	0.3727370	93.333% (14/15)
target12.jpeg	Jessica Alba	0.3812532	93.333% (14/15)
target13.jpeg	Jessica Alba	0.4219757	93.333% (14/15)
target14.jpeg	Jessica Alba	0.3664828	93.333% (14/15)
target20.jpeg	Unknow Person	0.6071871	0% (0/n)
target31.jpeg	Angelina Jole	0.3454402	90.0% (9/10)
target32.jpeg	Angelina Jole	0.4112590	100.0% (10/10)
target4l.png	Unknow Person	0.5779084	0% (0/n)
target42.jpeg	Unknow Person	0.5631111	0% (0/n)
target43.jpeg	Ellie Sattler	0.1669919	100.0% (10/10)
target44.jpeg	Ellie Sattler	0.3519967	100.0% (10/10)
target5l.jpeg	Claire Dearing	0.3827268	100.0% (9/9)
target52.jpeg	Claire Dearing	0.4382213	77.778% (7/9)

The second feature of the web application is the recognition statistics view. This view presents the minimum, average, and maximum distance for each person's entries in the form of a bar graph. The range of values is from 0 to 1, where 0 means the distance calculated from the photo or image frame equal to the one stored in the database, which may suggest that a photo that is already in the database was used. The given value indicates the distance between the analyzed image and the nearest corresponding record. Based on it, several conclusions can be made.

The first one is the discrepancy between the minimum and maximum value. This value suggests that a wide variety of photos/frames were given to the test. The system operated on relatively easy (low minimum value) and difficult (high maximum value) images. The average itself allows to determine the specific coefficient of difficulty level. The closer its value is to zero, the easier the data were to recognize.

The second one is as follows. If a person has maximum values close to the threshold level indicated by the yellow line in the graph, it means that there are entries in the database that are on the verge of being marked as unknown. The recognition threshold is fixed and based on testing, but it is certainly not perfect. These two facts suggest that an entry should not be trusted 100 percent. On the other hand, low minimum values suggest that there are cases that are easy to classify. One can assess whether the database is well learned by checking whether the entry "Unknown Person" appears in the graph and how large the bars are. The smaller the minimum, maximum, and average values, the better the learned base is because the distances to known images are small (Figure 5).

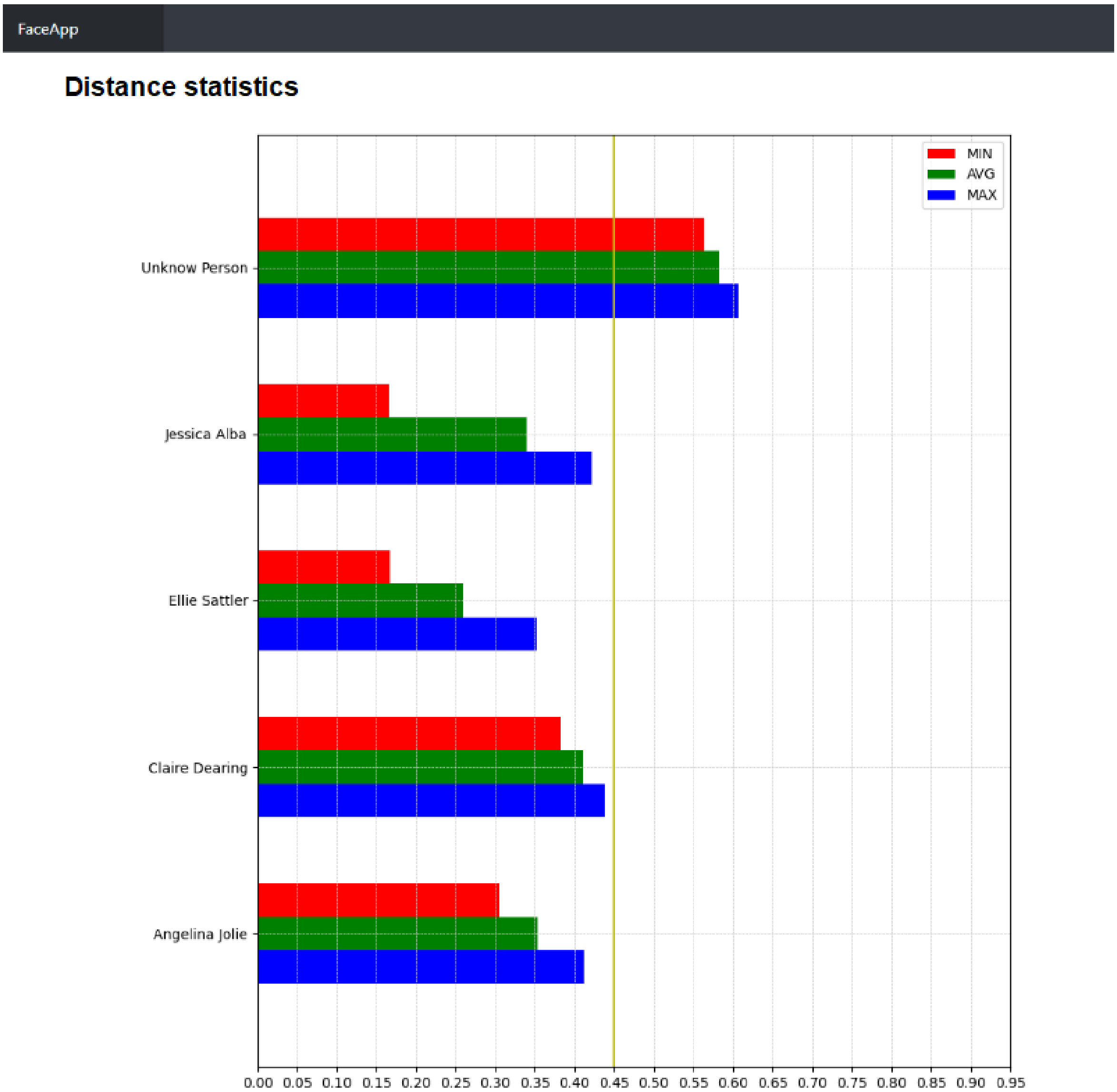

Figure 5. Distance statistics

6. Conclusions

Due to the conducted research, we can conclude that the effectiveness of the program is quite high. The created database of famous people allows to recognize them easily. It could be observed that in some samples the evaluation coefficient had the value of 14/15, which gave the fact that one photo from the group of 15 known people allowed for wrong assignment of a given person.

As can be easily seen, the real-time face recognition requires adequate performance. Raspberry Pi 4 in the version with 2 GB of RAM at the most optimal settings, provides the ability to analyze approximately one frame per second. Although this value is not very high, however, it proves to be sufficient in many applications. Attempts to run the system on other Raspberry microcomputers were made, but their recognition performance was unacceptable. The same system was also tested on Raspberry Pi Zero W, Raspberry pi 3, Raspberry Pi 4. In the case of the Raspberry Pi Zero W, the frame refresh rate was significantly below 1 frames per second [FPS]. For the Raspberry Pi 3, the image refresh rate with the recognition script was around 2-3 FPS, which was still unsatisfactory. The Raspberry Pi 4 approached the value of 5 FPS, which can be considered the absolute minimum if we want to properly carry out the face recognition process and we don't want to miss any information. In this case, due to the use of a popular single-board computer, the simplicity of the solution in relation to the achieved effect is very satisfactory. IoT is something that was once portrayed as science fiction in literature. Full automation of cyclical processes (and more) is slowly becoming our everyday life. The electronics we interact with almost all the time allow us to have constant control over our other devices, homes, cars, etc. The Internet of Things is with us in our homes, workplaces and other places. The ability to control lighting and heating with a smartphone is becoming the norm.

References

Ahmed, B. (2018). *Secure and Smart Internet of Things (IoT)*. Denmark: River Publishers, Gistrup.

Alam, M., Ara, K., Khan, S. (2020). *Internet of Things (IoT)*. Springer Nature.

Arora, S. Bhatia, M.P.S. (2021). Challenges and opportunities in biometric security: A survey. *Information Security Journal: A Global Perspective, 30*, 1-21.

Bolle, R.M., Connell, J.H., Pankanti, S. (2008). *Biometria*. Warszawa: WNT Scientific and Technical Publishers (in Polish).

Chantzis, F., Stais, I., Calderon, P.; Deirmentzoglou, E., Woods, B. (2021). *Practical IoT Hacking*. San Francisco: No Starch Press.

Chen, X., Yan, H., Yan, Q., Zhang, X. (2020). *Machine Learning for Cyber Security*. Springer Nature.

Hancke, P.G., Markantonakis, K. (2016). *Radio Frequency Identification and IoT Security*. 12th International Workshop, RFIDSec Hong Kong, China, November 30 -- December 2, 2016.

Kumar, S., Bhushan, B., Narayan, C.N. (2020). *IoT Security Paradigms and Applications*. Abingdon: Taylor & Francis Group.

Li, Z.S., Jain, K.A. (2011). *Handbook of Face Recognition*. Springer Science & Business Media.

Rohlf, J.F. (2011). *Biometry*. W.H. Freeman, Gordonsville.

Saravanan, V., Anpalagan, A., Poongodi, T., Khan, F. (2021). *Securing IoT and Big Data*. Abingdon: Taylor & Francis Group.

Silhavy, R. (2020). *Intelligent Algorithms in Software Engineering*. Springer Nature.

Suganthi, K. Machine (2021). *Learning and Deep Learning Techniques in Wireless and Mobile Networking Systems*. Abingdon: Taylor & Francis Group.

Ślot, K. (2008). *Wybrane zagadnienia biometrii*. Warszawa: WKŁ Transport and Communication Publishers (in Polish).

Veneri, G., Capasso, A. (2018). *Hands-On Industrial Internet of Things*. Birmingham: Packt Publishing.

Wechsler, H. (2009). *Reliable Face Recognition Methods*. Springer Science & Business Media.

Yasuura, H., Kyung, C., Liu, Y., Lin, Y. (2017). *Smart Sensors at the IoT Frontier*. Springer International Publishing.

Ziegler, S. (2019). *Internet of Things Security and Data Protection*. Springer, Cham.

Grzegorz LITAWA[1], **Mateusz ŻELASKO**[2], **Julia LITAWA**[3]

[1] University of Applied Sciences in Nowy Sącz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, e-mail: glitawa@ans-ns.edu.pl
[2] University of Applied Sciences in Nowy Sącz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, email: mat.zelasko@gmail.com
[3] University of Applied Sciences in Nowy Sącz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, email: jula.litawa@gmail.com

Detection and analysis of network security threats using artificial intelligence

Abstract

This article describes ways of detecting and analysing major security threats in network. Our research is based on a neural network created in Python language with the use of TensorFlow. With the help of artificial intelligence, we can analyse language for the presence of disinformation and propaganda. We have used BERT (Bidirectional Encoder Representations from Transformers) model, in order to catch fake, malicious and suspicious content. Basing on the BERT model, other models were created, for example HerBERT, which was trained to understand Polish. Furthermore, we have availed ourselves of DNS analysis, so we can localize IP addresses of Internet domains. In this work, we will discuss these methods of protecting information space.

Keywords: machine learning, language analysis, protecting information space, LSTM, Keras. BERT, DNS, NLP.

Wykrywanie i analiza zagrożeń bezpieczeństwa sieci przy użyciu sztucznej inteligencji

Streszczenie

W artykule opisano metody wykrywania i analizy głównych zagrożeń dla bezpieczeństwa sieci. Nasze badania opierają się na sieci neuronowej stworzonej w języku Python przy użyciu biblioteki TensorFlow. Dzięki sztucznej inteligencji jesteśmy w stanie analizować język pod kątem obecności dezinformacji i propagandy. Wykorzystaliśmy model BERT (Bidirectional Encoder Representations from Transformers), aby wykrywać fałszywe, złośliwe i podejrzane treści. Na bazie modelu BERT stworzono inne modele, takie jak HerBERT, który został przeszkolony do rozumienia języka polskiego. Ponadto skorzystaliśmy z analizy DNS, dzięki czemu jesteśmy w stanie zlokalizować adresy IP domen internetowych. W pracy tej omówimy te metody ochrony przestrzeni informacyjnej.

Słowa kluczowe: uczenie maszynowe, analiza języka, ochrony przestrzeni informacyjnej, LSTM, Keras, BERT, DNS, NLP.

1. Introduction

Natural Language Processing (NLP) has become a key field in the area of artificial intelligence, and the Bidirectional Encoder Representations from Transformers (BERT) model has played a significant role in its development. Natural language analysis is the field concerned with understanding, processing, and generating human language by computers. This is important due to the increasing amount of text data available online, such as articles, reviews, social media and much more. Artificial intelligence, especially NLP models, helps in the automatic understanding and processing of these texts (Géron, 2020; Brownlee, 2020).

The BERT model developed by Google is one of the most important innovations in the field of NLP. This based on the Transformers architecture model has gained huge popularity thanks to its ability to understand context and relationships between words in sentences. Its "bidirectionality" means it is able to analyze both the previous and next words in a sentence, leading to more precise natural language processing (Srebrovic et al., 2020; Bezliudnyi et al., 2023).

The BERT model is used in various fields such as sentiment analysis. It helps in understanding whether a given text is positive, negative or neutral, which is useful in monitoring content transmitted on networks. Search engines improve the relevancy of search results by taking into account the meaning of keywords in context. Machine translation enables more precise translations between different languages. Fake news detection helpfuls in identifying disinformation and fake news on social media and the internet. Natural language analysis using artificial intelligence and the BERT model has a huge impact on the way we process and reason about texts. With this tool, we are able to extract valuable information from the text data ocean, which is applicable in many fields, from business to scientific research. The development of NLP and models like BERT promises even more exciting applications in the future.

2. Language analysis

Nowadays, we have to face threats connected with information war, because they affect our world in many ways. We have to give end-user a chance to protect from it. Artificial intelligence models can be used in tasks in the range of natural language processing – including search, chatbots, sentiment analysis, and autocomplete. We have utilized it in detecting propaganda. The 2016 US presidential elections resulted in collecting a big dataset of Twitter posts marked as Russian propaganda, which we have used in model training. In 2023 there exist a lot of pre-trained models suitable for tasks like this. Propaganda comes in many forms, but it can be recognized by its persuasive function, sizable target audience, the representation of a specific group's agenda, and the use of faulty reasoning and/or emotional appeals (Miller, 1939). Since propaganda is conveyed through the use of a number of techniques, their detection allows for a deeper analysis at the paragraph and the sentence level that goes beyond a single document-level judgment on whether a text is propagandistic. In our research, we have chosen BERT model, because it can be fine-tuned in order to fit the given task of propaganda recognition (e.g. repetition, red herring) (Da San Martino et al., 2019; Bezliudnyi et al., 2023).

3. Python program for Propaganda classification using TensorFlow

Splitting data into training, validation, and testing sets is a common practice in machine learning. Its purpose is to evaluate and optimize the performance of a model. The training set is used to train the model by adjusting its parameters based on the examples in the set. It aims to help the model learn the underlying patterns in the data. The validation set is used to evaluate the performance of the model during training and to tune the hyperparameters. Hyperparameters are parameters that cannot be learned from the data. They are set before training (e.g. learning rate, number of layers and number of neurons). This set is used to determine the optimal values for these hyperparameters. The purpose of using the testing set is to evaluate the final performance of the model after it has been trained and optimized on the training and validation sets. It helps to estimate the generalization error of the model, which is the error rate on new, unseen examples. The testing set is essential to assess the model's ability to generalize to new data and to compare the performance of different models (Albon, 2018).

```
[ ]: import random
random_state = random.randint(0, 9999)

train, test = train_test_split(dataset, test_size=0.1,␣
 ↪random_state=random_state)
train, val = train_test_split(train, test_size=0.2, random_state=random_state)
```

Figure 1. Split the data into training, validation, and test sets

```
[ ]: train['label'] = train['label'].astype('int32')
val['label'] = val['label'].astype('int32')
test['label'] = test['label'].astype('int32')
```

Figure 2. Convert label column to integers

4. Define the BERT layer

We load Keras layer from TensorFlow Hub for preprocessing the input text data. This layer is based on the BERT architecture and is specifically designed for English language text data that is uncased. Second layer loads another Keras layer from TensorFlow Hub that contains a pre-trained BERT model for English language text data. The trainable=True argument makes the BERT layer trainable during the fine-tuning process (Brownlee, 2020).

```
[ ]: bert_preprocess_layer = hub.KerasLayer("https://tfhub.dev/tensorflow/
 ↪bert_en_uncased_preprocess/3")
bert_layer = hub.KerasLayer('https://tfhub.dev/tensorflow/
 ↪bert_en_uncased_L-24_H-1024_A-16/4', trainable=True)
# CHANGE IT FOR SOMETHING MORE SUFFICIENT !!!
```

Figure 3. Define the BERT layer

An LSTM cell can be considered in two ways. The first is to treat it as a black box. In this case, the state of the cell is divided into two vectors: $h_{(t)}$ – corresponds to the short-term state of the cell and $C_{(t)}$ – corresponds to the long-term state of the cell. The second way is to define the long-term state of $C_{(t-1)}$ starting on the left side of the cell and ending on the right side of it. Long-term state, it passes the forget gate and changes its value, then sums up with the entry gate and is passed on without any modification. The long-term state $C_{(t)}$ is copied and passed through the *tanh* function (hyperbolic tangent), the result of this operation is filtered by the output gate and the state $h_{(t)}$ is created – short-term. This state is equal to the output of the cell for this time step (Brownlee, 2020; Essien et al., 2020).

5. Add an LSTM layer with 64 units

We use the LSTM layer to further process the contextualized word embeddings generated by a BERT model. LSTM model is a special type of RNN architecture that is capable of learning long-term dependencies because the memory, which is implemented by the classical RNN model, is short-lived – at each training step, the information in the memory is combined with new information and is completely overwritten after a few iterations (Géron, 2020; Brownlee, 2020; Essien et al., 2020). Mixing LSTM and BERT layers gave us the even better performance of this state-of-art model.

```
[ ]: lstm_output = tf.keras.layers.LSTM(64)(bert_output['sequence_output'])
```

Figure 4. Define the LSTM layer with 64 units

Then, we need to batch the datasets. Batching is the process of splitting a dataset into smaller subsets, called batches or mini-batches, to be processed and trained on a model in parallel. Dataset batching is a technique used in deep learning and other machine learning applications to train models more efficiently. Instead of training the model on the entire dataset at once, the dataset is divided into smaller batches or mini-batches, and the model is trained on each batch separately. There are several benefits of batching. Model is processing smaller batches of data at a time. Regardful to it model can be trained faster. Furthermore, this process reduces amount of memory required to train the model. In addition, batching is helpful in improving the generalization of the model, because it prevents overfitting on individual examples (Loukas, 2020).

This model after training can be connected to the end user interface as for example browser extension which would analyse text and mark for the user probably harmful information. If detected propaganda text contains a hyperlink, it could be collected and then used for online training of the DNS analyser described below.

6. DNS domain classification

Domain Name System is a protocol, which associates IP address with more user-friendly domain names. It is composed of two or more parts, separated by dots. These parts are also called labels or segments. This system is hierarchical. The rightmost label represents the top-level domain (TLD) while the leftmost label represents the specific domain. It is very important for Internet because it is used by many internet services, even those malicious. DNS plays a significant role in identifying and blocking access to domains, which might be associated with malicious activity. We have created a neural network model which distinguish between benign and malicious DNS domain. As the DNS domains are not very complicated the preprocessing and the model used are not sophisticated, just character-level text tokenization and the basic Deep Neural Network model (Demertzis et al., 2019).

7. Network traffic analisys

The TextVectorization layer in TensorFlow is a preprocessing layer that can be used to vectorize text data. Vectorizing is the process of converting data into a numerical vector or array format that can be processed by a machine learning algorithm. In machine learning, the model input data needs to be in a numeric format, so it can be processed by the algorithms that perform computations on the data (Géron, 2020).

On the lower level of internet communications techniques lays the flow of IP packets between internet hosts. In modern times most of the network traffic is encrypted by protocols like HTTPS or SSH, so it is pointless to look for information related to propaganda or misinformation by itself in it, but it can be used to detect anomalies in network traffic which may indicate a compromised host. It is important for information space as it may allow threat actors to acquire credentials to social media accounts which could be used then for the purpose of spreading propaganda and misinformation. Also, if compromised man has access to some sensitive data, it can be used as a fuel for psychological warfare. This scenario occurred in Poland in 2021 when the head of the chancellery of Poland's prime minister had his email hacked, which led to the leak of sensitive documents and emails online. For training and testing there was used USTC--TFC2016 dataset which consists of samples of benign and malicious network traffic from different sources. The dataset was also pre-processed by mapping IP addresses into random, so the model would not learn specific addresses as malicious or benign, which can vary. The important thing to notice is the fact that cybersecurity is a constant arms race – threats are constantly changing, so it is important to ensure that the model deployed in the wild would be trained using the most actual samples.

Analysing of network traffic is a difficult task that can be solved in many ways. We're proposing a combination of two different approaches which differs in data preprocessing but are using the same neural network model.

We concentrate on Transmission Control Protocol and User Datagram Protocol flows, splitting them and analysing only two first packets. This naive approach bases on the fact that in network flow most of datagram fields do not change, so only a small probe should be sufficient to classify flow as benign or malicious. As the most significant advantage, it is relatively small computational complexity of preprocessing data, so this approach could be utilized in real-time network traffic analysis. We are splitting network traffic into different flows using IP addresses and ports. IP addresses are randomized so the model will not be dependent on specific subnetwork. Then the flows are converted into 24-bit RGB bitmaps with resolution 24x42 which is suitable for two Ethernet frames. If (and this scenario happens in most situations) packets are not maximum size, the rest of bitmap is padded with zeros, resulting in black pixels. Then two first packets of every flow are converted into 24-bit RGB bitmaps with resolution 24x42 (Demertzis et al., 2019).

8. DNS domain classification

Data preprocessing incorporates converting network traffic data into images in order to make use of the methods used in image analysis – probably the most studied case of using artificial intelligence. Convolutional neural networks (CNNs) are a commonly used type of deep learning algorithm in image analysis, and they have been found to be particularly effective in processing images created from network traffic data. This is because CNNs are designed to identify and extract meaningful patterns and features from images through a series of convolutional and pooling layers. By converting network traffic data into images, data preprocessing can enable the use of CNNs for analysing and classifying network traffic. The rescaling layer in TensorFlow Keras can be used to scale the input data to be between 0 and 1 by dividing each pixel value by the maximum possible value, which is 2^24 – 1. By applying this rescaling operation to the input data, the pixel values are normalized, and the model is more capable to learn the important features and patterns in the data. Additionally, normalization helps to prevent the model from being overly sensitive to the range and magnitude of the input data, which can improve the model's accuracy and generalization ability (Livieris, 2020).

The presented models can be used for comprehensive user protection, for example, in conjunction with a web browser interface. They can also be used to detect and analyse domains from which potentially harmful content is being sent, in order to take further steps to protect users, messengers, websites, or entire computer networks (Chollet, 2019).

9. Conclusions

During our research, we have created programs with use of artificial intelligence, which can help use face cyber security threats. The first one, which has been described in more specific way, is analysing text. It can be helpful in protecting people from disinformation. We haven't concentrated on the code of second program, because of it obvious drawbacks, that impact its accuracy. In further research, we can improve it, by for example, giving the model more examples or we can fit it better. As a result model would be more adjusted to protect our devices from malware and malicious domains and protect users, messengers, websites, or entire computer networks.

References

Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D., Chen, W., Ju, Y. (2019). A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. *Neural Comput Appl, 31*(5), 1665-1667.

Albon, C. (2018). *Uczenie maszynowe w Pythonie. Receptury*. Gliwice: Helion.

Bezliudnyi, Y., Shymkovych, V., Kravets, P., Novatsky, A., Shymkovych, L. (2023). *Pro-Russian propaganda recognition and analytics system based on text classification model and statistical data processing methods*. Retrieved from: https://ela.kpi.ua/bitstream/123456789/55707/1/278923-643037-1-10-20230510.pdf.

Brownlee, J. (2020). *How to Develop LSTM Models for Time Series Forecasting*. Retrieved from: https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/.

Brownlee, J. (2020). *Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras*. Retrieved from: https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/.

Chollet, F. (2019). *Deep Learning. Praca z językiem Python i biblioteką Keras*. Gliwice: Helion.

Da San Martino, G., Seunghak, Y., Barron-Cedeno, A., Petrov, R., Nakov, P. (2019). *Fine--Grained Analysis of Propaganda in News Articles*. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (pp. 5636-5646). Hong Kong.

Demertzis, K. Iliadis, L. Bougoudis, I. (2019). *Gryphon: a semi-supervised anomaly detection system based on one-class evolving spiking neural network*. Neural Comput Appl.

Essien, A. Giannettic C. (2020). A Deep Learning Model for Smart Manufacturing Using Convolutional LSTM Neural Network Autoencoders. *IEEE Transactions on Industrial Informatics PP, 99*, 6069-6078.

Géron, A. (2020). *Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow*. Gliwice: Helion.

Livieris, I.E. (2020). A CNN–LSTM model for gold price time-series forecasting. *Neural Computing and Applications, 32*, 17351-17360.

Loukas, S. (2020). *Time-Series Forecasting: Predicting Stock Prices Using An LSTM Model*. Retrieved from: https://towardsdatascience.com/lstm-time-series-forecasting-predicting-stock-prices-using-an-lstm-model-6223e9644a2f.

Srebrovic, R., Yonamine, J. (2020). *Leveraging the BERT algorithm for Patents with TensorFlow and BigQuery*. Retrieved from: https://services.google.com/fh/files/blogs/bert_for_patents_white_paper.pdf.

Anna KOCHANEK[1], Karol PLATA[2]

[1] University of Applied Sciences in Nowy Sącz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, e-mail: akochanek@ans-ns.edu.pl
[2] University of Applied Sciences in Nowy Sącz, Faculty of Engineering Sciences, Zamenhofa 1a, 33-300 Nowy Sącz, e-mail: karoleuro5@interia.pl

The efficiency of vehicle fleet management using the example of a company transporting a T60 medium

Abstract

The transport of hazardous substances is an integral part of today's society. Nearly 90% of dangerous goods are transported by road in Poland. Due to the lack of a nationwide monitoring system, based on the electronic transmission of data related to the transport of this cargo, it is extremely important to comply with transport safety rules right from loading. In order to improve the efficiency of the fleet, it is important to implement monitoring at each stage of the transport of hazardous materials. In the case of the transport of dangerous goods, monitoring is one of the elements that minimise the consequences of accidents by, among other things, pinpointing the exact location of the vehicle, checking the vehicle's route and the driver's compliance with traffic regulations and working time limits. The aim of this study is to present the concept of effective truck fleet management. The hypothesis of the thesis identifies areas through which the efficiency of fleet management can be improved, using the example of a company transporting a medium T60.

Keywords: dangerous goods, transport, efficiency, vehicle fleet, medium T60.

Efektywność zarządzania flotą pojazdów na przykładzie firmy transportującej medium T60

Streszczenie

Transport substancji niebezpiecznych jest integralną częścią dzisiejszego społeczeństwa. Blisko 90% towarów niebezpiecznych przewożonych jest w Polsce transportem drogowym. Ze względu na brak ogólnokrajowego systemu monitoringu, opartego na elektronicznej transmisji danych związanych z przewozem tego ładunku, niezwykle istotne jest przestrzeganie zasad bezpieczeństwa transportu już od załadunku. W celu poprawy efektywności floty istotne jest wdrażanie monitoringu na każdym etapie przewozu materiałów niebezpiecznych. W przypadku transportu towarów niebezpiecznych, monitoring jest jednym z elementów minimalizujących skutki wypadków, m.in. poprzez dokładne określenie lokalizacji pojazdu, sprawdzenie trasy przejazdu pojazdu oraz przestrzegania przez kierowcę przepisów ruchu drogowego i czasu pracy. Celem pracy jest przedstawienie koncepcji efektywnego zarządzania flotą samochodów ciężarowych. Postawiona hipoteza w pracy określa obszary dzięki, którym można zwiększyć efektywności zarządzania flotą pojazdów na przykładzie firmy transportującej medium T60.

Słowa kluczowe: ładunki niebezpieczne, transport, efektywność, flota pojazdów, medium T60.

1. Introduction

The selected transport company operates internationally and transports Class 3 and Class 9 hazardous materials. It uses high-class tankers for the transport of these materials (Janusz, 2017). In addition to transporting hazardous substances, it also provides transport services for tarpaulin combinations. Over the past 25 years, the company has also offered other forms of transport for the carriage of construction materials, right up to sensitive loads such as electronic devices (Ciuła et al., 2023). Since 1998, the company has been fulfilling orders on the basis of the ADR convention, observing international guidelines with regard to the marking of hazardous materials and their protection (Brodzik R. 2020). In addition, the company conducts training to minimise the likelihood of accidents involving tankers that transport bituminous masses (Rogalski et al., 2018). The fleet of vehicles used meets the Euro5 standard. The vehicles are of the Mercedes-Benz make and carry the classification code ADR: FL. The fleet of vehicles meets certain design features for vehicles carrying flammable liquids and gases. The company has single--compartment insulated vehicles with code L4BH, designed for the transportation of hazardous materials (Generowicz et al., 2023). The tankers it owns are made of high-grade stainless steel with capacities ranging from 30,000 [dm3] to 31,000 [dm3]. The company handles the transportation of the hazardous substance T60 for a company that uses the medium substance T60 as a binding agent in the production process of carbon electrodes (http://mptransport.pl/index.php/flota, download date: 15.01.2024).

The vehicle fleet complies with the technical standards and legal requirements that allow the vehicles to operate in Poland and within the European Union (Mężyk, 2016). Figure 1 shows the classification of legal requirements for the selection of vehicles transporting hazardous substances (Pusty, 2009).

Figure 1. General classifications of legal requirements for the selection of vehicles transporting dangerous substances

(source: own compilation based on Pusty, 2009)

The tankers used have skeletonless cylinder-shaped tanks with a wall thickness of 3 [mm] and a stainless steel density of 8,000 [kg · m^{-3}] (Janczak, 2014). The tankers used by the company, when moving in traffic, must have:

– pneumatic connections for the brake system;
– a control panel for the braking system;
– safety devices for lateral entry of the vehicle into another traffic lane;
– a safety valve to be used against overpressure support;
– a loading hatch;
– a main rear unloading valve and a hermetically sealed upper valve;
– a closing flap;
– a clamping ring with a safety handle;
– a pump used for unloading and loading the substance;
– a wedge;
– spare wheel holders;
– a fire extinguisher and a first aid kit;
– a toolbox (Nowacki et al., 2017).

The above-mentioned components of the tank car's structure must have continuous technical inspections so that the vehicle meets safety requirements and does not present a danger on the road (https://www.tdt.gov.pl/dzialalnosc, download date: 27.01.2024). Figure 2 shows the tanker of the company described.

Figure 2. Tanker used for the transport of hazardous substances
(source: http://mptransport.pl/index.php/flota)

The essence of the distribution of the T60 medium is not the primary purpose of the manufacturer's business, but is a business development phase acting to achieve success in the sale of a product (substance). The substance transported is a component of another product. Distribution decisions are not only limited to the transport organisation, but to:

– the management of the mutual relationships of the actors involved;
– controlling the flow of information;
– obtaining financial resources;
– taking care of market development;
– compliance with certain requirements and standards (Nowakowska-Grunt, Starostki-Patyk, 2017).

The composition of the hazardous substance medium T60 is set out in detail in the safety data sheet, drawn up in accordance with European Parliament and Council Regulation 2006 No. 1907/2006/EC. The drawing up of a certified study confirming the identification of the substance sold and transported is required mainly for reasons of operational safety and recognition of possible risks (Urban et al., 2014). This document is mandatory and is checked by the Road Transport Inspectorate (Kowalski et al., 2022). According to the law, every manufacturer, supplier as well as user should be able to read the safety data sheet. The description of the substance shall include the following information:

– the identification of the mixture;
– hazard identification;
– first aid measures;
– fire-fighting measures;
– environmental release measures;
– handling in storage;
– control of personal protective equipment;

– physico-chemical properties;
– toxicological information;
– ecological information;
– transport information;
– waste treatment;
– regulatory information (https://www.laboratoria, download date: 27.01.2024).

Medium T60 is a binding agent in electrode manufacture, identified in the safety data sheet as the binder pitch SP 84 in article P11 (Jancza, 2014). The mixture contains heterocyclic hydrocarbons and aromatic hydrocarbons, which may cause, among other things:

– allergic skin reaction,
– severe burns,
– mucous membrane irritation,
– genetic defects,
– tumours,
– infertility,
– harmful effects on the health of the child in pregnant women,
– harmful effects on aquatic organisms (Szaniawska et al., 2015).

This substance is for specialist use only. In order to transport this substance, the carrier must have the appropriate authorisation. The characteristics of the substance, which are described in the table, are of particular relevance to transport and storage 1 (Grzegorczyk, Buchcar, 2021).

Table 1
Chemical composition of medium T60

Identification of the substance CAS number	EINECS	Name of the substance	Hazard class	Substance content [%]
65995-93-2	266-028-2	Coal tar, high temperature pitch	Rakotw. Kat. 2; T; R45	>50
50-32-8	200-028-5	Benzopyrene	Rakotw. Kat. 2, Muta. 1B	<2,5

(source: own compilation based on: RUTGERS, Charecteristics sheet)

Storage of the T60 medium substance requires a suitably adapted warehouse, workstation and specific storage conditions, i.e. sealed, dry and durable high-temperature silos (Ciuła et al., 2023). When changing the location of the hazardous substance in the unloading room, adequate ventilation should be provided. An important requirement according to the RUTGERS safety data sheet for persons working in substance relocation is to have certified protective clothing (overalls), gloves (CE marking, cat III), oxygen mask (with P3 filter) and sealed goggles and appropriate footwear (Mężyk, 2015).

Efficient fleet management requires constant attention to quality, maintaining high standards of vehicle durability and efficiency, reducing operating costs and meeting production targets (Banach, 2015). Efficient and safe management refers to a long-term process of planning, taking decisive action and anticipating the consequences of actions taken. To this end, a transport company employs a fleet management employee to:

- ensure that tyres are changed on time (responsible for purchasing the right tyres, scheduling visits to the vulcaniser);
- monitor the quality of the cars and control the quality of the technical condition of the vehicle (monitors the service activities) and send the vehicle for periodic check-ups and technical inspections;
- create analyses and reports and describe the conclusions of the obtained results;
- control the costs of vehicle operation (e.g. payment of insurance, finding the most cost- and time-efficient repair shop, monitoring of vehicle combustion, fuel prices and available budget);
- monitor fuel prices when paying with fuel cards valid in the European Union (https://e100.eu/pl/blog/business, download date: 01.02.2024).

The components of an effective truck fleet management system include the introduction of, among other things, appropriate processes and systems for monitoring the location and condition of vehicles. This allows decision-makers to properly manage the fleet in order to reduce risks, improve safety and check the current location of equipment and vehicles (Jurecki, 2014). The introduction of GPS systems also makes it possible to monitor fuel consumption, manage required maintenance and communicate freely in remote areas where mobile network signals are unavailable. The overarching goal of these activities is to increase productivity and efficiency and equivalently improve the safety of employees driving trucks and carrying out transport orders (https://www.verizonconnect.com, download date: 01.02.2024).

2. Objective and research methodology

The aim of the study is to find an element of HGV fleet management that can be improved and managed more effectively. The methodology used in the thesis is mainly focused on the results of my own research – a survey and analysis of the elements of HGV fleet management. The subject of the study in the thesis, consists of many elements, such as:

- employees of a company transporting T60 medium;
- employees managing the truck fleet;
- data available from the transport company: number of employees, scope of activities of the position, fleet management logic, main elements of fleet management, description of the company, description of the T60 medium.

The subjects of this study are the employees in charge and fleet management of a company transporting hazardous material. The research carried out is intended to highlight which of the main elements of truck fleet management are the least effective.

The main research problem posed in this thesis is as follows: “Can the efficiency of vehicle fleet management be increased using the example of a company transporting a T60 medium?”. The specific research problems, on the other hand, took the form of the following questions:

1. Can knowledge of and compliance with legislation related to the transport of hazardous materials increase the efficiency of truck fleet management?
2. Do in-house drivers contribute to the efficiency of fleet management?
3. How can managers contribute to increasing the efficiency of fleet management?
4. How can the company's fleet management process be improved?

Given the above research questions, the research hypothesis in this thesis is as follows: "There are areas for increasing the efficiency of vehicle fleet management using the example of a company transporting a T60 medium".

The techniques used in the study are subordinate to the research method – analysis. The research techniques used in the study are:

- Survey;
- observation;
- document analysis;
- content analysis;
- statistical techniques.

In this study, research tools such as a survey questionnaire and an EXCEL spreadsheet were used.

3. Description of own research

3.1. Analysis of own research – questionnaire

For the research part of the study, a questionnaire was prepared on drivers involved in the transport of hazardous materials. The group surveyed was 50 people. The survey was conducted in the months of January to May 2023. The first part of the questionnaire contained a metric, the next part referred to questions directed towards the drivers working in the described company.

Knowledge of legislation, standards and compliance with all regulations by drivers transporting hazardous materials is a very important aspect in the effective management of a fleet of vehicles. Transporting, loading and unloading hazardous materials in compliance with standards and legislation is designed to protect the company from the potential risks of cargo unsealing. Unsealing a load during improper loading, unloading or transport can lead to numerous financial penalties, environmental hazards and road accidents. On this basis, six questions were selected to be answered by drivers transporting hazardous materials. Then, on the basis of an analysis of the answers obtained, the company can select training courses for drivers transporting hazardous materials accordingly.

100% of the people surveyed in the study were male. Seventeen respondents were between the ages of 26 and 35. Thirteen respondents indicated that they were aged between 36 and 45. Twelve people indicated that they were over 45 years old. Eight respondents indicated that their age was between 18 and 25. Additionally, the survey asked about the length of service of the respondents. Most respondents (22 persons) have seniority between 4 and 6 years and 18 respondents have seniority between 7 and 10 years. Respectively, 5 respondents each indicated that their seniority is within 1 to 3 years and above 10 years. Another aspect examined in the survey was the education of the respondents. Vocational education is held by 24 respondents, secondary education is held by 14 respondents, primary education is held by 10 respondents and tertiary education is held by two respondents. In addition, all survey respondents are involved in the transport of hazardous materials.

The first question posed was as follows: "How often do you come into contact with the transport of hazardous materials?". The first question related to the frequency of contact of the respondents with the transport of hazardous materials. Thirty-seven respondents have frequent contact with the transport of hazardous materials. Eight have infrequent contact with hazardous materials and five have daily contact with hazardous materials. All employed drivers in the company are licensed to transport hazardous materials. However, some of the drivers also transport food materials in the form of fruit. Based on the question posed, it can be concluded that some of the drivers have very little contact with the transport of hazardous materials.

Another question posed in the survey was as follows: "Have you been made well acquainted with the law concerning the transport of hazardous materials in Poland?". Forty-five respondents answered that they were well acquainted with the law concerning the transport of hazardous materials. In contrast, only 5 respondents indicated that they were not well acquainted with the law related to the transport of hazardous materials in the country. Analysing the above responses, it can be concluded that adequate training should be provided to make the employees more familiar with the current laws on the transport of hazardous materials.

The third question posed is as follows: "Do you comply with the legislation on the transport of hazardous materials?". Figure 3 shows the individual responses to the question posed.

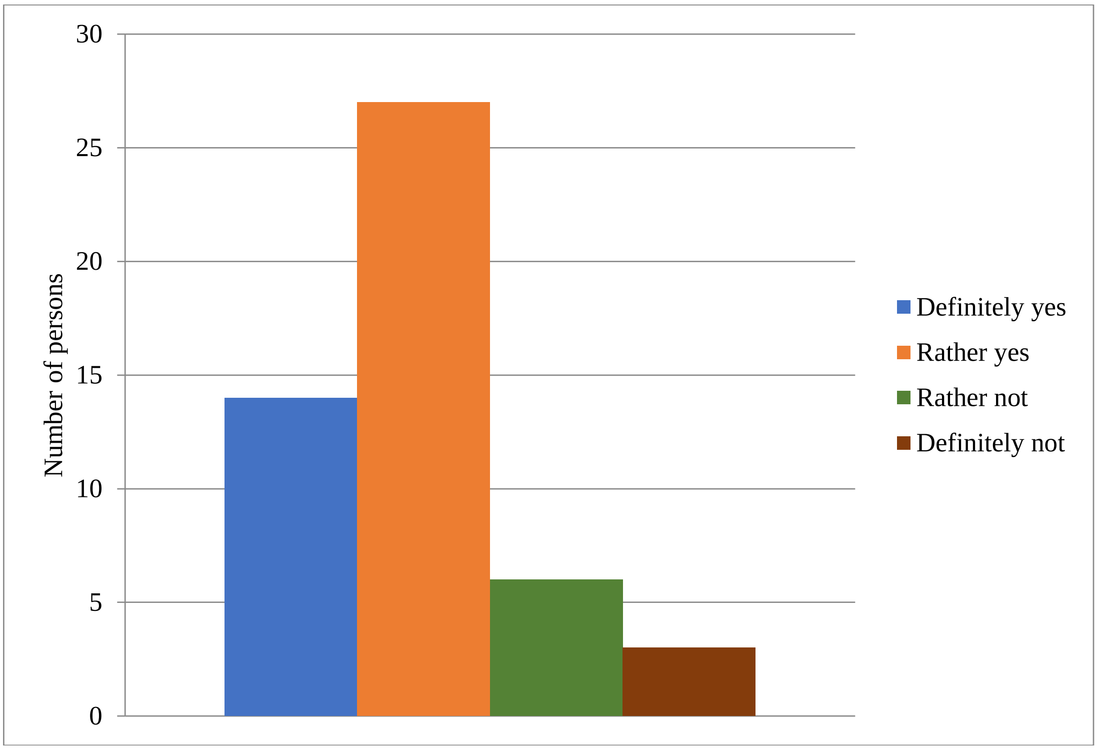

Figure 3. Compliance with the law relating to the transport of hazardous materials
(source: own elaboration)

In the third question, respondents were asked to indicate whether they comply with the law on the transport of hazardous materials. Twenty-seven indicated that they rather comply with the regulations. Fourteen respondents indicated that they definitely comply with the regulations related to the topic described. Six people indicated that they tended not to comply with them and 3 people indicated that they definitely did not comply with these regulations. Comparing question 2 with question 3, it can be concluded that some drivers, despite knowing the law on the transport of hazardous materials, do not comply with it at work.

The fourth question posed in the survey was as follows: "Are you aware of the danger caused by the transport of hazardous materials?". The majority of those surveyed, as many as 90%, are aware of the danger caused by the transport of hazardous materials. On the other hand, 10% of the respondents have no such awareness related to this risk. Analysing the above answers to the question asked, it can be concluded that the company should organise training. This training should be aimed at making employees aware of the dangers of not complying with legislation on the transport of hazardous materials.

The fifth question concerned the proper parking of a tanker truck with hazardous substances. The question was as follows: "When transporting hazardous materials, is parking done in designated areas?". Respondents 54% indicated that parking does not take place in designated areas. In contrast, the remaining 46% of respondents indicated that parking takes place in designated places. The parking of a combination of vehicles transporting hazardous substances is a very important aspect. Leaving a vehicle in an inappropriate place is punishable by a fine of 30 000 [PLN]. Vehicles transporting hazardous goods may park in a car park which is suitably prepared for this purpose in accordance with the regulation of the Minister of Internal Affairs and Administration on the technical conditions of car parks. Parking spaces for vehicles transporting dangerous goods must be at least 3.5 [m] wide and at least 20 [m] long. According to the new regulations, these sites must be suitably located, i.e. no closer than 300 metres from residential buildings and water intakes. In the event that such a parking space is separated by a wall with adequate fire resistance, which is at least four metres high, the minimum distance from houses and water intakes will be halved. Parking stalls for vehicles carrying hazardous loads must be at least 30 metres from the stalls of other cars. Where there is a forest of at least 3 [ha] nearby, such a site must be set back 40 metres from the tree line. The same distance must be maintained if there are railway tracks or gas pipelines in the vicinity.

The last question was about vehicle monitoring and read as follows: "Is the vehicle equipped with monitoring?". Respondents at 58% indicated that their vehicle is not equipped with monitoring. In contrast, 42% of respondents indicated that their cars have monitoring on board. In this case, vehicle monitoring is based on better fuel management. Fleet fuel management for trucks in companies should be based on an efficient and intelligent monitoring system that supports the recording and control of the vehicle refuelling process (Wieteska et al., 2016). The GPS tool should facilitate the management of fuel management, the staff team and the vehicle fleet. GPS Online provides transport companies with a comprehensive service, control and equipment system containing information on:

- fuel management system based on GPS technology;
- fuel distributors;
- ON fuel tanks;
- terminals for access to tanks;
- access to the distribution system;
- parts and accessories;
- modules for automatic drive and time accounting;
- notifications of urgent and worrying events;
- automatic report on control of consumption and exceeding of fuel standards, as shown in Figure 4 (http://www.gpsonline.com.pl, download date: 02.02.2024).

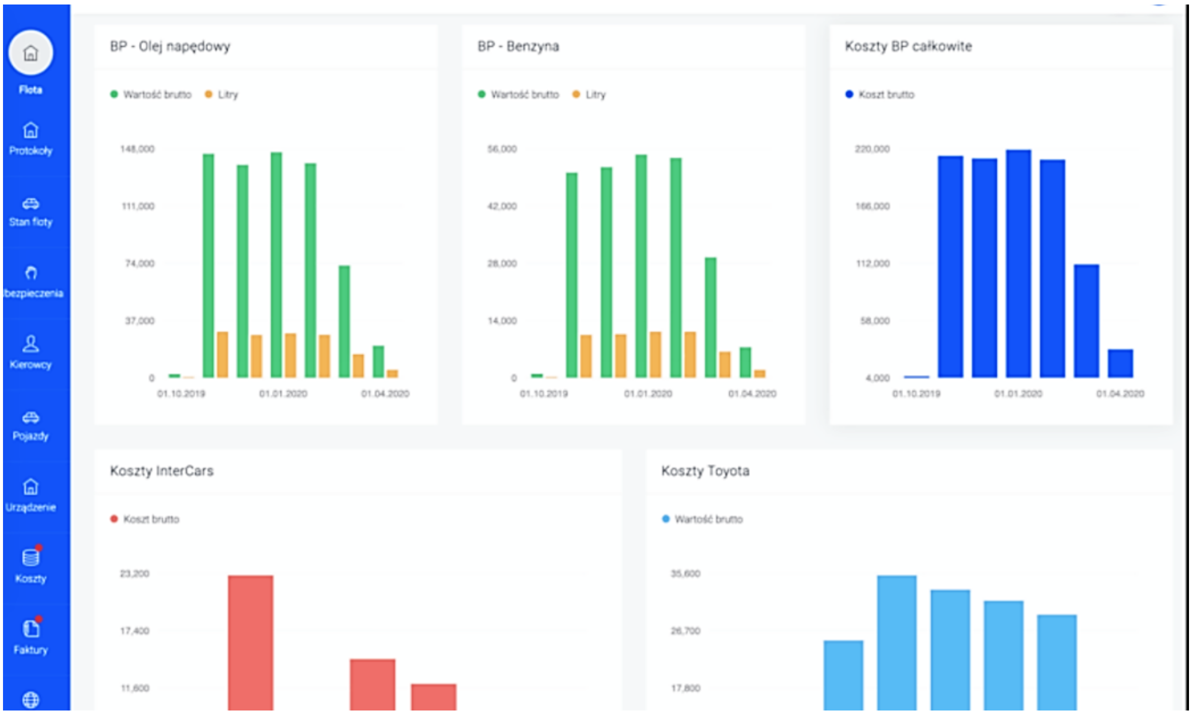

Figure 4. Mobile fuel fleet management

(source: own elaboration)

Truck assemblies have been continuously improved over the years. They are currently characterised by low fuel consumption due to the vehicle configuration together with its bodywork, driving style and maintenance (Redmer et al., 2014). These factors have a decisive impact on lower fuel consumption, which is influenced by vehicle features such as:

– traditional adaptive cruise control and the Eco option;
– adaptive cruise control function that guarantees constant vehicle speed (minimum fuel consumption) and anticipates changes in road conditions, in particular curves, gradients and intersections;
– function of switching off the running engine when the vehicle is idling, thus saving approximately 1.5 [l] less fuel when the truck is stationary;
– fitting of alloy wheels reduces weight, increases payload and reduces fuel consumption;
– a system that informs when a drop in tyre pressure is detected helps maintain the correct tyre pressure and saves fuel;
– the driver assistance function, DPA, allows you to correct and maintain an efficient driving style (https://www.daftrucks.pl, download date: 02.02.2024).

3.2. Efficiency of fleet management based on a described transport company

The area of truck fleet management is undergoing dynamic development, from a company perspective. For this reason, a comprehensive truck fleet management is an essential aspect. In order to effectively manage a truck fleet, elements such as:

– monitoring;
– regular servicing and maintenance;
– route optimisation;

– cost control;
– safety;
– data analysis (Nivette, 2022).

Figure 5 shows the developed plan for the effective management of the truck fleet.

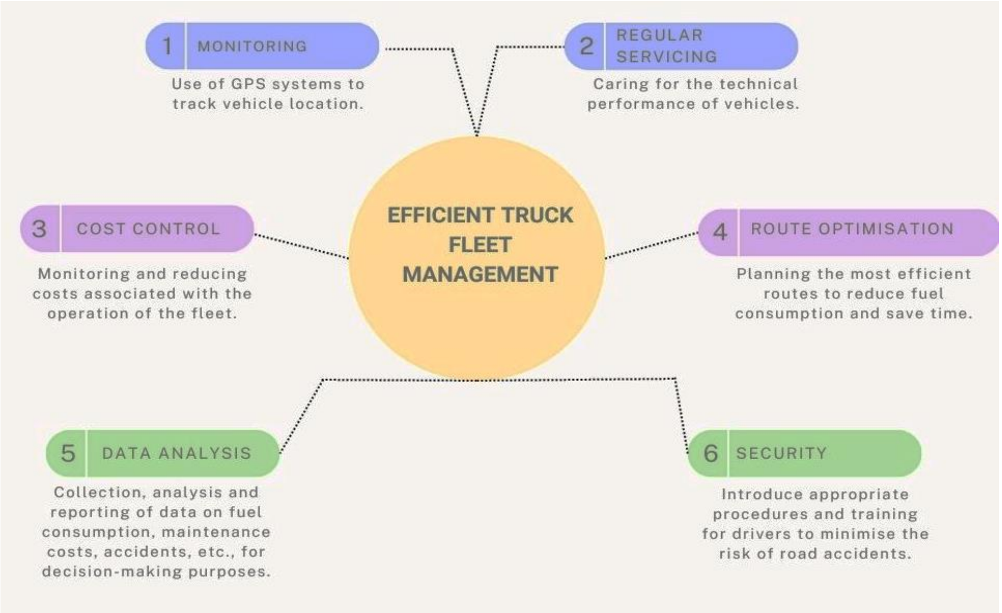

Figure 5. Plan for the effective management of the truck fleet

(source: own elaboration)

All the elements listed in the figure above have already been implemented by the company in its truck fleet management system for several years. Table 2 shows the most important criteria and functions.

Table 2

Key criteria and functions for truck fleet management

Criteria / Functions	SCALE	GPS system	Service repair	Cost monitoring	Route planning	Data reading	Driver training
		8	7	9	7	5	4
Monitoring	5	40	35	45	35	25	20
Regular servicing	3	24	21	27	21	15	12
Cost control	4	32	28	36	28	20	16
Route optimisation	4	32	28	36	28	20	16
Data analysis	3	24	21	27	21	15	12
Security	2	16	14	18	14	10	8

(source: own elaboration)

In the table above, employees in the transport company have adopted criteria and functions that are classified as follows:

A) functions:
– monitoring,
– regular maintenance
– cost control,

- route optimisation,
- data analysis,
- security.

B) Criteria:
- GPS system,
- service repair,
- cost monitoring,
- route planning,
- data reading,
- driver training.

A scale of 1 to 5 was adopted for each function, while a scale of 1 to 10 was adopted for each criterion. Company staff jointly assigned a scale for each function and criterion. Each function was then multiplied by the criteria. The resulting scores were added together and the sum of the individual functions was obtained (Table 3).

Table 3
The sum of the most important functions for the management of heavy goods vehicle fleets

Functions \ Criteria	Σ
Monitoring	200
Regular servicing	120
Cost control	160
Route optimisation	160
Data analysis	120
Security	80

(source: own elaboration)

4. Conclusions

Based on the result, it can be concluded that the most important function for the company is vehicle monitoring. The second most important function for the company is cost control and route optimisation. The next most important functions are regular maintenance, data analysis and safety, which is the least important function according to the employees. The results presented above coincide with a survey of employees who transport hazardous materials. The company should take decisive steps in training employees on the law and on hazardous materials standards. Failure to comply with legislation can lead to a wide variety of accidents and increase the resulting damage. Underestimating the problems that can occur as a result of non-compliance and disregarding basic safety rules not only puts health or life at risk, but also causes irreversible environmental pollution.

Following the survey, it can be seen that there is no systematic training in the transport of dangerous goods for drivers. Any change in the regulations should lead to the organisation of training for all persons involved in the process of transporting dangerous goods in the company. Companies also do not use precise and clear instructions for their employees on the problem described. As a result, when a problem arises, drivers do not always know how to react, which often leads to losses for the company. Appropriate training could help in this regard. The hypothesis put forward in the paper that there are areas for increasing the efficiency of vehicle fleet management using the example of a company transporting a T60 medium is a correct hypothesis, as the above conclusions show.

References

Banach, Z. (2015). *Vehicle fleet management systems – application in three-tier architecture*. Warszawa: Warsaw University of Technology.

Brodzik R. (2020). Hazards in road transport of hazardous materials. *Material Economy and Logistics Journal, 4*. DOI: https://doi.org/10.33226/1231-2037.2020.4.3.

Ciuła, J., Kowalski, S., Wiewiórska, I. (2023). Pollution Indicator of a Megawatt Hour Produced in Cogeneration – the Efficiency of Biogas Purification Process as an Energy Source for Wastewater Treatment Plants. *Journal of Ecological Engineering*. DOI: https://doi.org/10.12911/22998993/158562.

Ciuła, J., Wiewiórska, I., Banaś, M., Pająk, T., Szewczyk, P. (2023). Balance and Energy Use of Biogas in Poland: Prospects and Directions of Development for the Circular Economy. *Energies, 16*, 3910. DOI: https://doi.org/10.3390/en16093910.

Generowicz, A., Gronba-Chyła, A., Kulczycka, J., Harazin, P., Gaska, K., Ciuła, J., Ocłoń, P. (2023). Life Cycle Assessment for the environmental impact assessment of a city' cleaning system. The case of Cracow (Poland). *Journal of Cleaner Production, 382*, 135184. DOI: https://doi.org/10.1016/j.jclepro.2022.135184.

Grzegorczyk, K., Buchcar, R. (2009). *Dangerous goods: transport in practice*. Kraków: ADR.

Janczak, A. (2014). *ADR transport – classification of dangerous goods*. Retrieved from: http://www.bhp.abc.com.pl/czytaj/-/artykul/transport-adr-klasyfikacja-towarow-nie bezpiecznych.

Janczak, A. (2014). *Carriage of dangerous goods*. Retrieved from: http://www.bhp.abc.com.pl/czytaj/-/artykul/przewoz-towarow-niebezpiecznych.

Janusz, M. (2017). The carriage of dangerous goods by road with particular reference to liquid fuels. *OAFL – Student online journal, 2*(2).

Jurecki, R., Chaba, R. (2014). *A modern tool for fleet management in a company to improve driving economy and driver safety*. Conference materials of the IX International Science Technical Conference, Automotive Safety 2014, Rajecké Teplice, 8-10 April 2014, 160-165.

Kowalski, S., Opoka, K., Ciuła, J. (2022). Analysis of the end-of-life the front suspension beam of a vehicle. *Eksploatacja i Niezawodnosc – Maintenance and Reliability, 24*(3), 446-454. DOI: http://doi.org/10.17531/ein.2022.3.6.

Mężyk, A. (2015). Education of vehicle drivers as a method of improving road safety. *Logistics, 4*.

Mężyk, A. (2016). Transport of hazardous cargo in urban areas. *Buses – Technika. Eksploatacja. Transport systems, 12*.

Nivette, M. (2022). *Supporting vehicle fleet management in an enterprise with mobility plans and minimising maintenance costs*. Warszawa: Warsaw University of Technology.

Nowacki, G., Chmieliński, M. (2017). Analysis of the transport of dangerous goods in the European Union Member States. *Safety and Ecology, 9*.

Nowakowska-Grunt, J., Starostki-Patyk, M. (2017). *Distribution logistics*. Warszawa: WKiŁ.

Pusty, T. (2009). *Carriage of dangerous goods. Poradnik kierowcy*. Warszawa: WKŁ.

Redmer, A., Kiciński, M., Rybak, R. (2014). Management of the truck fleet – methods. *Materials management and logistics, 4*, 11-18.

Regulation of the European Parliament and of the Council numbered 1907/2006/EC. (2006).

Regulation of the Minister of the Interior of 13 November 2012 on the technical conditions of car parks to which vehicles transporting dangerous goods are removed.

Rogalski, G., Pyza, D. (2018). *Hazards in road transport of dangerous goods*. Warszawa: Scientific Papers of the Warsaw University of Technology, Department of Transport.

Szaniawska, D., Ćwirko, K. (2015). Chemical safety education in the transport of hazardous materials. *General and Professional Education, 2*.

Urban, J., Szylar, K. (2014). Safety in the transport of dangerous goods. In: *TransLogistics 2014*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej.

Wieteska, S., Piechota, A. (2016). Modern methods of managing truck and van fleets in Poland. *ASO, 8*(2), 105-119.

http://mptransport.pl/index.php/flota.

http://www.gpsonline.com.pl.

https://e100.eu.

https://www.daftrucks.pl.

https://www.laboratoria.

https://www.tdt.gov.pl/dzialalnosc.

https://www.verizonconnect.com.