Formation of an indicator of the information message reliability level by fuzzy logic toolkit*

Iryna Pikh[1,†], Vsevolod Senkivskyy[1,†], Alona Kudriashova[1,*,†], Lyubov Tupychak[1,†], and Roman Andriiv[1,†]

[1] *Lviv Polytechnic National University, Stepan Bandera Str., 12, Lviv, 79013, Ukraine*

Abstract

In the conditions of the growing information flow and the presence of a large amount of inaccurate or questionable data, an important task is to identify the messages veracity. The proposed research is devoted to the development of models and approaches for assessing the information messages reliability level using fuzzy logic tools. Fuzzy logic-based prognostic assessing allows one to take into account the data fuzziness and ambiguity, creating more flexible and adaptive models for analysis. Linguistic variables are singled out and grouped by types and functional characteristics – generalized factors of influence on the veracity of the message content, which serve as the basic information basis of the studied issues. A universal term-set of values of linguistic variables and their corresponding linguistic terms containing a descriptive identification of the importance level of the variable in the separation quanta of the values set is designed. A model of logical inference is developed, which reflects the hierarchical dependency of the information messages veracity degree on the values of linguistic terms of factors, and serves as a basis for prognostic assessment of the news reliability level. The membership functions of the linguistic variables at the separation points of the term-sets of values are calculated based on the results of the processing of matrices of pairwise comparisons of the factors ranks, and their visual graphic representation is performed. Knowledge matrices are constructed and the general form of fuzzy logic equations is designed for linguistic variables that determine the integral indicator of the information messages reliability level.

Keywords

reliability level of a message, fuzzy logic, linguistic variable, membership function, knowledge matrix

1. Introduction

In today's information space, the issue of the message's reliability is becoming more and more relevant. Today's world is in the conditions of an information revolution, when the amount of data generated every day exceeds the capabilities of traditional analysis methods. With the technology development, the emergence of new media and the increase in the speed of information dissemination, users are increasingly faced with the need to verify its veracity. This problem is especially acute in the era of social networks, where the lack of strict moderation and control over authenticity can lead to the spread of fake news, manipulation, and facts distortion. In the conditions of information oversaturation of modern society, the problem of information messages reliability becomes extremely relevant.

It is advisable to focus the research of the information messages reliability level on the development of methods, models and tools for prognostic assessment of the probability of the false information appearance arising as a result of the news filtering and distorting, the appearance of fake messages, manipulation of data accuracy, and the use of sources that had a good reputation. It requires a comprehensive approach that integrates the efforts of various areas of computer science, information technology, linguistics, sociology, and other sciences to create reliable methods of

AdvAIT-2024: 1st International Workshop on Advanced Applied Information Technologies, December 5, 2024, Khmelnytskyi, Ukraine - Zilina, Slovakia

* Corresponding author.
† These authors contributed equally.
iryna.v.pikh@lpnu.ua (I. Pikh); vsevolod.m.senkivskyi@lpnu.ua (V. Senkivskyy); alona.v.kudriashova@lpnu.ua (A. Kudriashova); lyubov.l.tupychak@lpnu.ua (L. Tupychak); roman.andriiv@icloud.com (R. Andriiv)
0000-0002-9909-8444 (I. Pikh); 0000-0002-4510-540X (V. Senkivskyy); 0000-0002-0496-1381 (A. Kudriashova); 0000-0002-0963-3360 (L. Tupychak); 0009-0001-2450-5439 (R. Andriiv)

analysing the factors influencing the process of the information messages appearance for the purpose of prognostic assessment of their reliability level.

Known analysis methods often turn out to be insufficient to solve this task due to the complexity and ambiguity of the data provided. Based on traditional logic, they are not always able to adequately work with the information containing inaccuracies or uncertainties. In this connection, there is a need to create effective tools that would prognostically assess the veracity and reliability of information messages.

One of the effective ways of solving the stated problem and a promising approach for solving such tasks is the use of fuzzy logic, which allows modelling processes with uncertainties and partial truths and opens new perspectives in solving the problem of assessing the information messages reliability. Thanks to the ability of fuzzy systems to process heterogeneous and imprecise data, it is possible to create prognostic models that increase the accuracy and reliability of conclusions about the veracity of this or that information.

In view of what has been said, it is considered expedient to use the toolkit of the multifactorial analysis theory, modelling theory, and the fuzzy set theory in order to identify the linguistic factors influencing the veracity of the news sites context, the processing of which by fuzzy logic toolkit will enable the creation of a mathematical apparatus for fuzzy modelling, which reflects the direct dependency of the information messages veracity degree on the values of the linguistic terms of the factors, and will become the basis for the prognostic formation of the indicator of the news reliability level.

2. Literature review

A significant increase in the amount of information in social networks has led to a new important area of research, which is actively developing, related to the detection of the degree of information messages veracity and reliability. This is confirmed by the analysis of literary sources on this topic, which testifies to the growing interest of the scientific community and workers in the information services area regarding the problem of establishing and ensuring the messages content veracity.

The fundamental work [1] describes the methodology of automated detection of fake news using the method of probabilistic analysis of the main components to determine the characteristics that may indicate the possibility of information falsification or unreliability. In the publication [2], the detection of fake news and classification of unreliable information is carried out with the help of extended text representation with recognition of multi-EDU structures. The work [3] recommends implementing the process of detecting fake news campaigns using convolutional networks in the form of graphs. The publication [4] is focused on identifying disinformation in social networks using a framework capable of recognizing the information and classifying it by type: true text data; fake data. In [5], an alternative approach to the manual verification of the facts objectivity is considered, the essence of which is the use of machine learning algorithms to automate the process of detecting fake news. At the same time, the necessity of obtaining a large set of training data and selecting appropriate functions that can best detect deception is indicated. An interesting study [6] proposes a fact-checking model that takes into account users' literacy and trust in the news in order to research their behaviour in a social network environment prone to disinformation. Survey data collected from social media users is analysed using structural equations. The publication [7] contains a methodology for surveying teachers, students, and specialists in the media industry to find advanced practices for training fact-checking and verification skills, since one of the primary requirements of the mass media market is to master modern methods of verifying the facts veracity and establishing the information messages reliability by future journalists. In work [8], it is noted that the use of various methods of detecting fake news, in particular with the use of fuzzy logic and artificial intelligence, consists in the development of mechanisms capable of performing an automated check of the content authenticity provided in libraries. The developed model can be integrated into library digital data processing systems, which, if necessary, will mark certain news content as potentially fake, preventing the uncontrolled dissemination of untruth through libraries. The research [9] proposes an innovative hybrid model based on fuzzy logic to improve the performance of fake news detection. The model uses a combination of news articles and textual and numerical contextual information. The model assessment results indicate that combining fuzzy logic with deep learning can improve fake news detection and become a reliable tool to combat disinformation. In [10], the authors assessed the methods based on the results of modelling for information about a sensitive area built on fuzzy logic. Possible security vulnerabilities of an IoT environment insensitive to data transfer conditions are explored. An algorithm is proposed that will guarantee detection and protection methods aimed at protecting extremely sensitive areas, that is, where the attack probability is maximum. The publication [11] substantiates the feasibility of using fuzzy logic, the technique of fuzzy logic control and the corresponding functions of intelligent control systems, which can be used as an alternative to traditional methods of process control. The benefits of using fuzzy components by distributed systems are provided by high-level functions such as fusion and decision making. The problem of methods selecting for increasing the solutions reliability is considered in the study [12]. Effective factors that can significantly influence the achievement of decision-making goals in the environment of fuzzy processes development are identified. The factors reproduced in the classical form in natural language are singled out, which significantly increases the confidence in the decisions made by determining the maximum and minimum values of the membership functions. The effectiveness of the proposed solutions is assessed using the procedure of fuzzy inference and Zadeh-Mamdani approach. The authors of [13] note that fake news has become a phenomenon that can be considered a targeted disinformation policy used to spread false data or distorted messages. This problem needs to be studied from both technical and social aspects, since the use of technical means to disseminate information is aimed at people. In the publication [14], it is recommended to assess the data reliability in the information space with the help of signs that do not require a certain level of users' qualification, which include the dubiousness of the presented facts, emotionality, tonality and sensationalism of the content. The work [15] is focused on the criteria for monitoring the text data reliability in the media environment, the essence of which involves the information balance, the separation of facts from general considerations, the definition of accuracy, objectivity and completeness of information messages.

As a result of the analysis of literary sources, it is noted that the methods of detecting the credibility and veracity level of data disseminated by modern media sources, despite their originality and novelty, do not fully use the fuzzy logic tools to determine the generalized indicator of the information messages reliability level. At the same time, the need for a wider application of the fuzzy set theory tools is confirmed in order to create effective mechanisms for assessing the information reliability in the media space, thanks to which it becomes possible to examine information resources for the presence of the influence of negative factors on them, identify information security problems in a timely manner, recommend proven solutions for use on critical objects. It is worth emphasizing that the research presented by these authors is aimed at the use of fuzzy logic components for modelling processes that recommend a reasonable and effective preliminary analysis of the factors of news formation through the use of original approaches and models of prognostic formation of the indicator of the information messages reliability level.

Considering the above, it is believed that the subject area research in the direction of prognostic assessment of the information messages reliability level is timely and relevant.

3. Material and methods

3.1. Model of logical inference

In the general interpretation, technological or production processes largely depend on the established norms or characteristics that predictably determine the quality of their implementation. The situation is more complicated with social processes, the factors of which often have little formalized interpretation and a subjective, often unclear way of expression, which creates problems for their research. Processes aimed at determining the information messages reliability level can be attributed to them with a high probability degree. The above once again emphasizes the need to use an adequate theoretical apparatus for the study of processes described by fuzzy or descriptive (linguistic) characteristics, which ultimately proves in favour of the use of the fuzzy set theory and its defining direction – fuzzy logic, the effectiveness of which is confirmed by a number of publications [16-18].

The starting position defining the research concept, the essence of which is to determine the degree or parts of the influence of factors related to the formation of an integral indicator of the reliability level of information messages (IM) has the following interpretation [17].

Let $D=\{d_1,d_2,...,d_m\}$ be a generalized set of some information messages; $M=\{x_{1_m},x_{2_m},...,x_{n_m}\}$ be a number of factors influencing the reliability level of the m-th message, where n_m is a number of factors of the m-th message. Each of the messages can be characterized by a certain function $A_m(x)$, that determines the reliability level of its information content. Summarizing the initial conditions formulated, it can be stated that:

$$A_m(x) \equiv \bigcup_{j=1}^{n} \omega\left(x_{j_k}\right), \quad (k=1,2,...,m), \tag{1}$$

where: $A_m(x)$ is a numerical indicator of the reliability level of the m-th message; $\omega\left(x_{j_k}\right)$ is a numerical measure of the indicator of the reliability level introduced by the j-th factor in the k-th message.

Then the preliminary statement is formulated as follows: among information messages there will be at least one for all factors of which the condition (1) is fulfilled:

$$(\exists d)\ (\forall x)\ A_m(x);\ d\in D;\ x\in M\ . \tag{2}$$

In the process of research, the statements formulated by the founder of the fuzzy set theory Zadeh and developed and supplemented in works [17-19] are used.

The basic information unit of the fuzzy set theory is considered a linguistic variable (LV), which, unlike a mathematical variable, is described by words of ordinary language. The set of possible values of a linguistic variable forms a term-set, the elements of which are terms. An important characteristic of LV is the membership function (MF), formed on the basis of the term-set of values and linguistic terms (LT), characteristic for the factors of a random information process. A term-set or universal set contains numerical or conventional (in the absence of numerical) characteristics of the established boundaries of linguistic variables, linguistic terms describe discrete states of LV at certain separation points of the existence domain, or the basic scale of values of linguistic variables, using verbal qualitative formulations [20].

A set of factors (linguistic variables) that have an impact on the information messages reliability is considered the initial information base of the proposed research. The analysis of literary sources has made it possible to identify a number of factors related to the process of forming the indicator of the news reliability level. According to the results of the expert survey, a number of the most significant among them are singled out namely: fact checking, availability of multiple publications, the author's professionalism, the author's objectivity, the information source, informativeness of the message context, presence of refutation and criticism, social trust. As a result, the linguistic variable identifying the integral indicator of the information messages reliability level is presented in the form of a function:

$$D=F_D(B,T,L), \tag{3}$$

the arguments of which are linguistic variables of the second level B,T,L

B is focused on factors of organizational orientation; the linguistic variable T defines a function-procedure, the arguments of which are LVs relating to the author and the context of the message; the linguistic variable L characterizes the part of the indicator of IM reliability, which is assessed by the attitude (loyalty) of users in relation to the received news.

The essence of linguistic variables of the second level is represented by formalized expressions describing their functional dependency on the LVs of the third level of the hierarchy.

$$B = F_B\left(b_1, b_2, b_3\right), \quad (4)$$

where: b_1 is LV "information source"; b_2 is LV "fact checking"; b_3 is LV "multiple publication" (verification through multi-post search).

$$T = F_T\left(t_1, t_2, t_3\right), \quad (5)$$

where: t_1 is LV "author's professionalism"; t_2 is LV "author's objectivity"; t_3 is LV "context informativeness".

$$L = F_L\left(l_1, l_2\right), \quad (6)$$

where: l_1 is LV "refutation and criticism"; l_2 is LV "social trust".

Based on the completed structuring by functional groups and levels of the factors importance in the process of forming an indicator of the information messages reliability level, a graphic model of logical inference is designed (Figure 1), which represents a multi-level hierarchy of connections between the components of the initial database and defines an algorithm for calculating the values of the membership functions of linguistic variables taking into account linguistic terms.

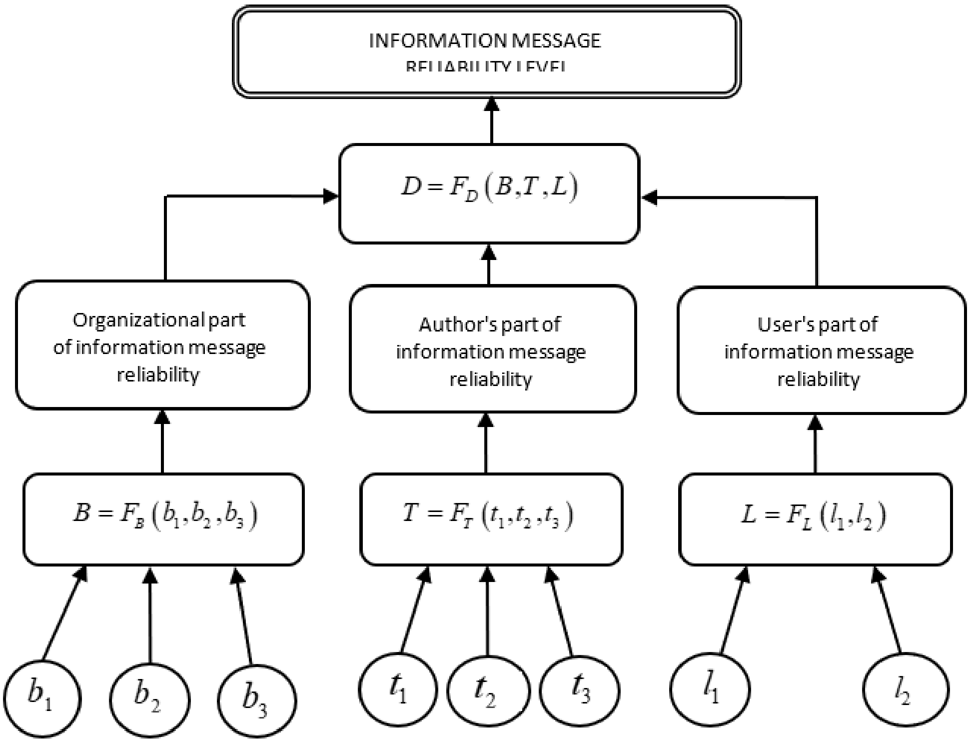

Figure 1: Model of logical inference – formation of an indicator of the information message reliability level.

The multi-level model conveys the logic of the algorithm for the formation of the IM reliability indicator according to the "bottom-up" principle, taking into account the partial indicators of the message reliability obtained at each level of the hierarchy. The parts of influence of the third-level LV are determined through the universal term-set of values and linguistic terms of variables.

3.2. Term-set of values and membership functions of linguistic variables

The research model of the textbook adaptation process and the above structuring of the initial database determine the transition to the following tasks, the essence of which is the formation of a universal base or term-set of values and functions of linguistic variables. It should be noted that a term-set represents the real (most often numerical), or conditional, in their absence, boundaries of action or influence on the process by a specific linguistic variable. Linguistic terms provide the information in a descriptive form about the importance of LV, attributed to the separation points of the universal set U. For the convenience of representing the specified actions, Table 1 is designed, which represents a mathematical symbol of a linguistic variable, its descriptive essence, a universal set of LV values, and a special set of linguistic terms grouped into three-dimensional scales to reproduce the qualitative states of linguistic variables in the quanta of the set U.

Table 1
Term-sets of values of linguistic variables

Variable	Linguistic description of a variable	Universal base of values (term-set U)	Linguistic terms (set H)
b_1	Information source (reliability)	(1-5) c. u.	Low, medium, high
b_2	Fact checking	(1-5) c. u.	Infrequent, frequent, constant
b_3	Multiple publication (number)	(1-5) c. u.	Small, medium, big
t_1	Author's professionalism	(1-5) c. u.	Low, medium, high
t_2	Author's objectivity	(1-5) c. u.	Low, medium, high
t_3	Informativeness of the message context	(1-5) c. u.	Low, medium, high
l_1	Refutation and criticism (frequency)	(1-5) c. u.	Small, medium, significant
l_2	Social trust	(1-5) c. u.	Low, medium, high

Processing of linguistic variables according to the rules of fuzzy logic is performed at discrete separation points u_i of the term-set U. Additionally, at these points, the ranks $r_q(u_i)$ related to the linguistic terms of LV are specified. Thus, in the generalized notation, the universal term-set $U = \{u_1, u_2, ..., u_n\}$ and the preference ranks $r_q(u_i)$ of linguistic terms in the separation quanta u_i $(i = 1, ..., n)$ are considered as initial data.

Taking into account the expressed warnings, the linguistic term "an indicator of the information messages reliability level" appears as a fuzzy set [17]:

$$D_F = \left\{ \frac{\mu_d(u_1)}{u_1}, \frac{\mu_d(u_2)}{u_2}, ..., \frac{\mu_d(u_n)}{u_n} \right\}, \qquad (7)$$

where: $D_F \subset U$; $\mu_q(u_i)$ is a membership function of the element $u_i \in U$ to the set D_F.

In the expression (7) $\mu_d(u_i)$ mean the normalized values of membership functions used in the process of forming and solving fuzzy logic equations. The relationship between the membership functions of linguistic variables and the corresponding ranks of the linguistic terms of the specified variables is presented by the following relations:

$$\frac{\mu_1}{r_1} = \frac{\mu_2}{r_2} = \ldots = \frac{\mu_n}{r_n}, \tag{8}$$

where: $\mu_i = \mu_q(u_i)$; $r_i = r_q(u_i)$ for all $i = 1,\ldots,n$.

Membership functions must satisfy the normalization condition, i.e.:

$$\mu_1 + \mu_2 + \ldots + \mu_n = 1.$$

The ranks of factors (linguistic variables) are usually determined by the methods of ranking or modelling of hierarchies, which makes it possible to calculate the membership function at five points of the universal term set according to the following relations [17]:

$$\left.\begin{aligned}
\mu_1 &= \left(1 + \frac{r_2}{r_1} + \frac{r_3}{r_1} + \ldots + \frac{r_n}{r_1}\right)^{-1}; \\
\mu_2 &= \left(\frac{r_1}{r_2} + 1 + \frac{r_3}{r_2} + \ldots + \frac{r_n}{r_2}\right)^{-1}; \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\mu_n &= \left(\frac{r_1}{r_n} + \frac{r_2}{r_n} + \frac{r_3}{r_n} + \ldots + 1\right)^{-1}.
\end{aligned}\right\} \tag{9}$$

The specified number of control nodes determines the appropriate informativeness of the visual graphic representation of the numerical characteristics of the membership functions.

Taking into account the specified prerequisites, the research task is formulated and voiced as follows: to ensure the achievement of the maximum of the function that determines the information messages reliability level with positive values of the linguistic terms of the universal base U and the maximum values of the membership functions at the separation points of the term-set [17]. The mathematical representation of the task is given in the following interpretation:

$$\left.\begin{aligned}
&D_F = F(b_k, t_j, l_p) \to \max,\ k = \overline{1,3};\ j = \overline{1,3}; \\
&b_k > 0,\ t_j > 0,\ l_p > 0,\ p = \overline{1,2}; \\
&\mu_d(u_i) \to \max,\ u_i \in U,\ D_F \subset U,\ i = \overline{1,5}.
\end{aligned}\right\}. \tag{10}$$

The formalized representation of the task (10) contains the components of the universal term-set of values b_k, t_j, l_p, in relation to the linguistic variables of the process.

Finally, square inversely symmetric matrices $A = a_{ij}$ ($a_{ij} = r_i/r_j$; $i, j = 1,\ldots,5$) in representation (11) are constructed for each linguistic variable and its corresponding three linguistic terms of the set H. Ranks are selected from the set of integers from 1 to 9 [20].

$$A = \begin{bmatrix}
1 & \frac{r_2}{r_1} & \frac{r_3}{r_1} & \frac{r_4}{r_1} & \frac{r_5}{r_1} \\
\frac{r_1}{r_2} & 1 & \frac{r_3}{r_2} & \frac{r_4}{r_2} & \frac{r_5}{r_2} \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{r_1}{r_5} & \frac{r_2}{r_5} & \frac{r_3}{r_5} & \frac{r_4}{r_5} & 1
\end{bmatrix}. \tag{11}$$

For example, the experimental calculation of the membership functions of one of the linguistic variables, characterized by three linguistic terms of the set H at five separation points u_1, u_2, u_3, u_4, u_5 of the term-set U is presented below.

4. Experiment, results

A random linguistic variable is selected, for example, b_1 "information source", the universal base of values of which is in the interval (1-5) of conditional units at the points $U(b_1)=[1;2;3;4;5]$, for which linguistic terms determining the information source reliability are set in Table 1 – $H(b_1)$= <low, medium, high>. For the term "low", the variable rank on the interval increases. The lowest rank of LV at the point u_1 is set by the number 1. First, an inversely symmetric matrix is created for the specified variable, which displays the value of the ranks in relation to the term "low". According to theoretical principles, the elements of only the fifth row are specified, since the rest of the matrix elements a_{ji} are obtained based on the dependency: $a_{ji}=a_{5j}/a_{5i}$; $i,j=\overline{1,5}$.

According to the stated conditions, the inversely symmetric matrix takes the following form:

$$A_{low}(b_1)=\begin{bmatrix} 1 & 6/9 & 4/9 & 3/9 & 1/9 \\ 9/6 & 1 & 4/6 & 3/6 & 1/6 \\ 9/4 & 6/4 & 1 & 3/4 & 1/4 \\ 9/3 & 6/3 & 4/3 & 1 & 1/3 \\ 9 & 6 & 4 & 3 & 1 \end{bmatrix}. \qquad (12)$$

Taking into account the elements of the matrix (12) and the expressions (9), the values of the membership functions are calculated for the term "low" at five points of the universal term set:

$\mu_{low}(u_1)=0,39$; $\mu_{low}(u_2)=0,24$; $\mu_{low}(u_3)=0,16$; $\mu_{low}(u_4)=0,12$; $\mu_{low}(u_5)=0,04$.

For the term "medium", the following inversely symmetric matrix is formed:

$$A_{medium}(b_1)=\begin{bmatrix} 1 & 5 & 9 & 4 & 1 \\ 1/5 & 1 & 9/5 & 4 & 1/5 \\ 1/9 & 5/9 & 1 & 4/9 & 1/9 \\ 1/4 & 5/4 & 9/4 & 1 & 1/4 \\ 1 & 5 & 9 & 4 & 1 \end{bmatrix}. \qquad (13)$$

Similarly to the previous steps, the corresponding membership functions are obtained:

$\mu_{medium}(u_1)=0,05$; $\mu_{medium}(u_2)=0,25$; $\mu_{medium}(u_3)=0,45$; $\mu_{medium}(u_4)=0,20$; $\mu_{medium}(u_5)=0,05$.

The matrix for the linguistic term "high" is presented in a similar form:

$$A_{high}(b_1)=\begin{bmatrix} 1 & 4 & 5 & 8 & 9 \\ 1/3 & 1 & 5/3 & 8/3 & 9/3 \\ 1/5 & 4/5 & 1 & 8/5 & 9/5 \\ 1/8 & 4/8 & 5/8 & 1 & 9/8 \\ 1/9 & 4/9 & 5/9 & 8/9 & 1 \end{bmatrix}. \qquad (14)$$

The following membership functions are obtained:

$\mu_{high}(u_1)=0,037$; $\mu_{high}(u_2)=0,148$; $\mu_{high}(u_3)=0,185$; $\mu_{high}(u_4)=0,296$; $\mu_{high}(u_5)=0,33$.

Normalization of membership functions is performed by multiplying their values by a coefficient:

$$k_r=1/\max\mu_r(u_i),\ \ (i=1,...,5),$$

where r means ”low”, ”medium”, ”high”.

Based on the results of using the formula $\mu_{r_n}(d_i) = k_r \times \mu_r(u_i)$ for the linguistic variable "information source" (the source reliability) with the specified terms "low", "medium", "high", the following normalized values of the membership functions are obtained:

$$\mu_{low_n}(u_1)=1;\ \mu_{low_n}(u_2)=0,61;\ \mu_{low_n}(u_3)=0,41;\ \mu_{low_n}(u_4)=0,31;\ \mu_{low_n}(u_5)=0,10;$$

$$\mu_{medium_n}(u_1)=0,11;\ \mu_{medium_n}(u_2)=0,55;\ \mu_{medium_n}(u_3)=1;\ \mu_{medium_n}(u_4)=0,55;\ \mu_{medium_n}(u_5)=0,11;$$

$$\mu_{high_n}(u_1)=0,11;\ \mu_{high_n}(u_2)=0,49;\ \mu_{high_n}(u_3)=0,56;\ \mu_{high_n}(u_4)=0,90;\ \mu_{high_n}(u_5)=1.$$

The normalized numerical priorities of the membership functions of LV " information source" (reliability) relative to the terms "low", "medium", "high" are presented by the expression (7) in fuzzy sets.

$$\begin{aligned} low\,reliability &= \left\{\frac{1}{1};\frac{0,61}{2};\frac{0,41}{3};\frac{0,31}{4};\frac{0,10}{5}\right\} \text{ c. u.;} \\ medium\ reliability &= \left\{\frac{0,11}{1};\frac{0,55}{2};\frac{1}{3};\frac{0,55}{4};\frac{0,11}{5}\right\} \text{ c. u.;} \\ high\ reliability &= \left\{\frac{0,11}{1};\frac{0,49}{2};\frac{0,56}{3};\frac{0,90}{4};\frac{1}{5}\right\} \text{ c. u.} \end{aligned} \quad (15)$$

An important step in the application of fuzzy logic tools for the study of processes containing uncertainties is the visualization of membership functions [18]. The purpose of visualizing the values of the membership functions for the LV "information source" (reliability) according to the terms "low", "medium", "high" is to graphically represent the source involvement in a certain reliability level depending on the specified term. The visualization helps to show that reliability is not a well-defined quantity (for example, "reliable" or "unreliable"), but varies gradually. The values of the membership functions indicate the degree of "attachment" of the LV to each term and determine to what extent the information source corresponds to a certain reliability level.

One of the key points in the practical application of the visualization is that the membership functions can overlap, that is, the same source can have some degree of membership in several terms at the same time. This provides flexibility in reliability classification and avoids rigid boundaries between terms.

The presented paper uses Gaussian membership functions, which, unlike triangular ones, have a smoother version, in which the membership changes more smoothly, which is useful for simulated processes where changes in reliability occur more gradually.

A visual graphic representation of fuzzy sets (15) of the linguistic variable "information source" by the terms "low", "medium", "high" is presented in Fig. 2.

Similarly, in a more complete interpretation, a graphical representation of the membership functions of the remaining linguistic variables is carried out, the absence of which in this paper does not change the essence of the general approach to the research problem.

The next step in the practical implementation of the task concerns the formation of a fuzzy knowledge base and the design of fuzzy logic equations. The formation of a fuzzy knowledge base ensures the construction of effective models for complex systems and processes where uncertainty and inaccuracy are inherent characteristics, providing flexibility and accuracy in decision-making.

The basis for the formation of a fuzzy knowledge base is the model of logical inference in Figure 1, which determines the sequence of actions regarding the process of establishing an indicator of the information messages reliability level. Fuzzy logical inference is a process in which, on the basis of fuzzy rules and fuzzy data (linguistic variables), results are calculated that may also be fuzzy.

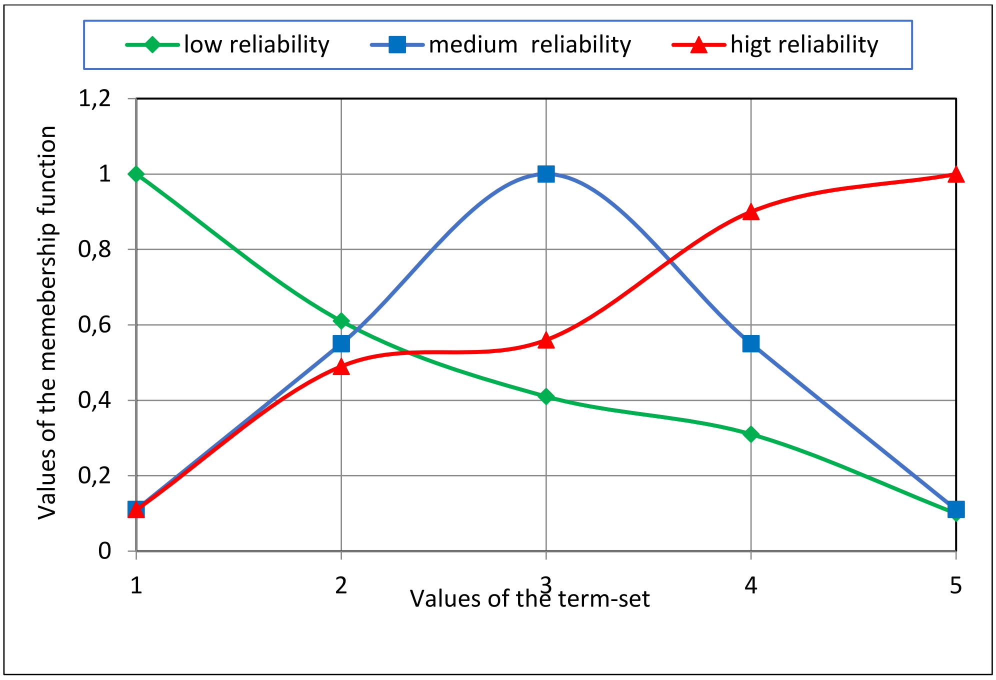

Figure 2: Visualization of membership functions of the linguistic variable "information source".

This allows for decision-making and analysis in situations where the information is incomplete, uncertain or unclear. According to the formalized interpretation of the task expressed by the entry (10), the indicator of the information reliability level is identified by the authors as the linguistic variable of the highest level D, for which a set of linguistic terms $H(D)$ = <low, medium, high> is formed. According to the model hierarchy, the basic sets of linguistic variables of the next hierarchy level relate to the author's, organizational, and user's parts of the IM reliability level, for which similar fuzzy sets receive the following expression:

$H(B)$ = <low, medium, high> – the organizational part of the IM reliability;

$H(T)$ = <low, medium, high> – the author's part of the IM reliability;

$H(L)$ = <low, medium, high> – the user's part of the IM reliability.

The basis of a fuzzy knowledge base is a set of fuzzy rules, which have the form: "if A, then C", where A and C are fuzzy sets. Such rules define relationships between linguistic variables. Therefore, the fuzzy knowledge base for the linguistic variable, which determines the generalized level of reliability of the information message, takes the following form:

IF (B = low) AND (B = medium) AND (B = high)
AND (T = low) AND (T =medium) AND (T =high)
AND (L = low) AND (L =medium) AND (L =high),
THEN (D = low) AND (D =medium) AND (D = high).

The knowledge matrix based on the fuzzy knowledge base is presented in Table 2.

Table 2
Knowledge matrix for the linguistic variable D

Organizational part of the IM reliability (*B*)	Author's part of the IM reliability (*T*)	User's part of the IM reliability (*L*)	Generalized IM reliability level (*D*)
low	low	medium	low
low	medium	low	
low	high	medium	medium
medium	low	high	

high	medium	high	high
high	high	medium	

The knowledge matrix became the basis of fuzzy logical equations, the solution of which led to obtaining membership functions for the specified terms of the linguistic variable D:

$$\mu_{low}(D) = \mu_{low}(B) \wedge \mu_{low}(T) \wedge \mu_{medium}(L) \vee \mu_{low}(B) \wedge \mu_{medium}(T) \wedge \mu_{low}(L);$$

$$\mu_{medium}(D) = \mu_{low}(B) \wedge \mu_{high}(T) \wedge \mu_{medium}(L) \vee \mu_{medium}(B) \wedge \mu_{low}(T) \wedge \mu_{high}(L);$$

$$\mu_{high}(D) = \mu_{high}(B) \wedge \mu_{medium}(T) \wedge \mu_{high}(L) \vee \mu_{high}(B) \wedge \mu_{high}(T) \wedge \mu_{medium}(L).$$

According to the logical inference model, similar procedures are performed for the next-level MF, i.e. $B = F_B(b_1, b_2, b_3)$, $T = F_T(t_1, t_2, t_3)$, $L = F_L(l_1, l_2)$, for which corresponding logical statements, knowledge matrices, and fuzzy logical equations are formed alternately.

Taking into account that the applied aspects of the research are carried out for the linguistic variable b_1 "information source", the following representation is focused on the linguistic variable $B = F_B(b_1, b_2, b_3)$.

A fuzzy knowledge base for qualitative assessment of the linguistic variable (B) is presented below.

IF (b_1) = (low, medium, high)

AND (b_2) = (infrequent, frequent, constant)

AND (b_3) = (small, medium, big)

THEN (B) = (low, medium, high).

The following data regarding the knowledge matrix are presented in Table, as above.

Table 3
Knowledge matrix for the linguistic variable B

Information source (reliability) (b_1)	Fact checking (b_2)	Multiple publication (b_3)	Organizational part of the reliability of IM (B)
low	infrequent	medium	low
medium	infrequent	small	
high	infrequent	medium	medium
low	frequent	big	
medium	constant	big	high
high	constant	medium	

Table 3 provides the construction of a system of fuzzy logical equations for calculating the values of membership functions of the linguistic variable B for the terms "low, medium, high":

$$\mu_{low}(B) = \mu_{low}(b_1) \wedge \mu_{infrequent}(b_2) \wedge \mu_{medium}(b_3) \vee \mu_{medium}(b_1) \wedge \mu_{infrequent}(b_2) \wedge \mu_{small}(b_3);$$

$$\mu_{medium}(B) = \mu_{high}(b_1) \wedge \mu_{infrequent}(b_2) \wedge \mu_{medium}(b_3) \vee \mu_{low}(b_1) \wedge \mu_{frequent}(b_2) \wedge \mu_{big}(b_3); \quad (16)$$

$$\mu_{high}(B) = \mu_{medium}(b_1) \wedge \mu_{const}(b_2) \wedge \mu_{big}(b_3) \vee \mu_{high}(b_1) \wedge \mu_{const}(b_2) \wedge \mu_{medium}(b_3).$$

To solve the system of fuzzy equations (16) in order to obtain the numerical values of the membership functions of the variable B

B, i.e. b_2 and b_3, at the points of the universal base of values *U,* which determines the areas of their existence, are calculated similarly to the above for the variable b_1.

Omitting similar procedures for linguistic variables T and L, the final stage concerns the linguistic variable D, which determines the integral value of the indicator of the prognostic reliability level of information messages. The fuzzy set of the variable D is described by membership functions with three fuzzy terms "low", "medium", "high". Numerical values D are formed at three points of the basic universal set. On the basis of (7), the value of the variable at the separation points of the set U is obtained from the expression:

$$D(B,T,L) = \left\{ \frac{\mu_{low}(D)}{d_1}, \frac{\mu_{medium}(D)}{d_2}, \frac{\mu_{high}(D)}{d_3} \right\}, \tag{17}$$

where d_1, d_2, d_3 are quantitative values of the variable D regarding the terms "low", "medium", "high". The formation of a fuzzy logical conclusion, knowledge matrix and fuzzy logical equations in relation to the linguistic variable D is carried out according to actions similarly applied to the linguistic variable B. Under the condition of obtaining the values of the membership functions $\mu_{low}(D)$, $\mu_{medium}(D)$ and $\mu_{high}(D)$, of the linguistic variable D, defuzzification of the fuzzy set (17), i.e., the calculation of the indicator of the prognostic reliability level of information messages, is carried out according to the expression, which is considered the basis of the method of the centre of mass of a flat figure, limited by the abscissa axis and the graph of the membership function of the linguistic variable D [17]:

$$P = \frac{\sum_{i=1}^{m} \left[\underline{D} + (i-1)\frac{\overline{D} - \underline{D}}{m-1} \right] \mu_i(D)}{\sum_{i=1}^{m} \mu_i(D)}, \tag{18}$$

where: P is a numerical indicator of the reliability level of information messages; $\underline{D}, \overline{D}$ are respectively, the lower and upper quantitative value of the variable D; m is a number of qualitative terms of assessing the variable D. When performing the calculations according to formula (17), the following initial data are recommended: $m = 3$; $\mu_1(D) = \mu_{low}(D)$, $\mu_2(D) = \mu_{medium}(D)$, $\mu_3(D) = \mu_{high}(D)$; quantitative values of the variable D regarding the terms "low", "medium", "high" – (1, 50, 100); lower and upper values D – $\underline{D} = 1$, $\overline{D} = 100$. The value of the indicator P is obtained as a percentage.

To confirm the effectiveness of the application of the expression (18), the message reliability level indicator is calculated. In addition to the above-mentioned initial data, the values of the membership functions of the linguistic variable D are set: $\mu_1(D) = \mu_{low}(D) = 0{,}10$; $\mu_2(D) = \mu_{medium}(D) = 0{,}45$; $\mu_3(D) = \mu_{high}(D) = 0{,}65$. Substituting them into formula (18), the following expression is obtained:

$$P = \frac{1 \cdot 0{,}10 + 50 \cdot 0{,}45 + 100 \cdot 0{,}65}{0{,}10 + 0{,}45 + 0{,}65} = 72{,}92\%.$$

The result of the calculation shows that with the specified initial data of the membership functions of the linguistic variables, significantly related to the process of prognostic assessment of the information data veracity, the indicator of the reliability level of media and printed messages is quite high. Provided there is a simulation model for calculating this indicator, the possibility of predicting the messages objectivity level through a change in the initial parameters, the essence of which is determined by the linguistic terms of the universal base of values, becomes real.

At the same time, the software gives an opportunity to provide feedback, the essence of which is that the user enters the value of the indicator acceptable to him into the system and receives at the output the characteristics of linguistic variables (the process factors) that determine the desired information reliability level.

5. Conclusions

Summarizing the above, it is claimed that as a result of the growth of information flows, there is a need to apply a scientific approach to assessing the reliability of data, which are generally characterized by fuzziness and ambiguity. This toolkit includes the tools of the fuzzy set theory and its applied direction – fuzzy logic, which ensures the creation of more flexible and adaptive models for data analysis and practical application. As a result of the review of literary sources and the general problems of modern society in determining the information messages veracity and objectivity, linguistic variables – generalized factors of influence on the data content veracity, which serve as the basic information basis of the subject under study – are singled out and grouped by type and functional characteristics. These include linguistic variables focused on factors of organizational orientation (information source, fact checking, multiple publication); concerning the author's professionalism, objectivity and the context informativeness; which express the attitude (loyalty) of users (refutation and criticism, social trust) in relation to the received news. A universal term-set of values of linguistic variables and their corresponding linguistic terms containing a descriptive identification of the importance level of the variable in the separation quanta of the set of values is designed. The term-set is considered a key concept in fuzzy systems, which allows working with values that do not have distinctive boundaries. A graphic multi-level model of logical inference is developed, which reflects the hierarchical dependency of the information messages veracity degree on the values of linguistic terms of linguistic variables, and becomes the basis for calculating the indicator of prognostic assessment of the news reliability level.

The model "works" according to the "bottom-up" principle. Inversely symmetric factor rank matrices are constructed for the linguistic variables of the process of formation of the IM reliability level indicator and the linguistic terms assigned to them. Based on the results of the processing of the matrices, the values of the membership functions of the linguistic variables at the separation points of the term-sets are calculated. One of the essential tools of fuzzy logic in the study of information processes is the visualization of membership functions, which determines the graphical representation of the involvement of a linguistic variable to a certain reliability level depending on the specified term. A graph of the LV "information source" is constructed, which clearly demonstrates how an information source is assessed by the reliability level in conditions of uncertainty, using fuzzy logic. The practical implementation of the task concerns the design based on the fuzzy logical inference of the fuzzy knowledge base. Knowledge matrices are constructed and the general form of fuzzy logic equations is designed for linguistic variables, which determine the numerical indicator of the information message's reliability level. The expression and initial data for calculating the indicator of the prognostic reliability level of information messages are presented.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] D. K. Dixit, A. Bhagat, D. Dangi, Automating fake news detection using PPCA and levy flight-based LSTM. Soft Comput., num. 26, num. 22 (2022) 12545–12557.

[2] Y. Wang, L. Wang, Y. Yang, Y. Zhang, Detecting fake news by enhanced text representation with multi-EDU-structure awareness. Expert Syst. Appl., num. 206 (2022).

[3] D. Michail, N. Kanakaris, I. Varlamis, Detection of fake news campaigns using graph convolutional networks. Int. J. Inf. Manag. Data Insights, num. 2 (2022).

[4] S. Rastogi, D. Bansal, Disinformation detection on social media: An integrated approach. Multimed. Tools Appl., num. 81, num. 28 (2022) 40675–40707.

[5] S. Rastogi, D. Bansal, A review on fake news detection 3T's: typology, time of detection, taxonomies. Int. J. Inf. Secur. 22 (2023) 177–212. doi:10.1007/s10207-022-00625-3.

[6] I. Kožuh, P. Čakš, Social Media Fact-Checking: The Effects of News Literacy and News Trust on the Intent to Verify Health-Related Information October Healthcare, 11(20):2796 (2023)
[7] A. López-Meri, et al, Digital Competencies in Verifying Fake News: Assessing the Knowledge and Abilities of Journalism Students. Societies *14*(5),66 (2024).
[8] U. Mertoğlu, & B. Genç Automated Fake News Detection in the Age of Digital Libraries. Information Technology and Libraries. 39(4) (2020).
[9] C. Xu, M.-T. Kechadi, Fuzzy Deep Hybrid Network for Fake News Detection. In The 12th International Symposium on Information and Communication Technology (SOICT 2023), December 07–08 (2023), Ho Chi Minh, Vietnam. ACM, New York, NY, USA 8 Pages.
[10] I. Memon, R. A. Shaikh, M. K. Hasan, R. Hassan, A. U. Haq, K. A. Zainol, Protect mobile travelers information in sensitive region based on fuzzy logic in IoT technology. Security and Communication Networks, (1), (2020) 8897098. doi:10.1155/2020/8897098.
[11] C. Dumitrescu, P. Ciotirnae, C. Vizitiu, Fuzzy logic for intelligent control system using soft computing applications. Sensors, 21(8), (2021) 2617. doi:10.3390/s21082617.
[12] J. Serrano-Guerrero, F. P. Romero, J. A. Olivas, Fuzzy logic applied to opinion mining: a review. Knowledge-Based Systems, 222: 107018 (2021).
[13] P. Nordberg, J. Kävrestad, M. Nohlberg, Automatic detection of fake news. In: 6th International Workshop on Socio-Technical Perspective in IS Development, virtual conference in Grenoble, France, June 8-9, 2020. CEUR-WS, (2020) 168-179.
[14] Z. Wu, S. Shen, X. Lian, X. Su, E. Chen, A dummy-based user privacy protection approach for text information retrieval. Knowledge-Based Systems 195 (2020): 105679.
[15] D. Choudhury, T. Acharjee, A novel approach to fake news detection in social networks using genetic algorithm applying machine learning classifiers. Multimed Tools Appl 82, (2023) 9029–9045. doi:10.1007/s11042-022-12788-1.
[16] V. Senkivskyy, I. Pikh, S. Babichev, A. Kudriashova, N. Senkivska, Modeling of Alternatives and Defining the Best Options for Websites Design. 2nd International Workshop on Intelligent Information Technologies and Systems of Information Security, March 24–26, 2021, Khmelnytskyi, Ukraine. Pp. 259–270.
[17] V. Senkivskyy, I. Pikh, A. Kudriashova, N. Senkivska, L. Tupychak, Models of Factors of the Design Process of Reference and Encyclopedic Book Editions. Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2021. Lecture Notes on Data Engineering and Communications Technologies. Springer, 77 (2022) 217–229.
[18] B. Durnyak, P. Shepita, L. Tupychak, Y. Petriv, J. Shepita, Post-press product quality assessment models for the IIoT system CEUR Workshop Proceedings 3736 (2024) 73–85.
[19] S. Kavita, R. K. Srivastava, Predicting software bugs of newly and large datasets through a unified neuro-fuzzy approach: Reliability perspective. Advances in Mathematics: Scientific Journal 10.1 (2021) 543-555. doi:10.37418/amsj.10.1.54
[20] M. Sharkadi, Methodology for the development of information technology for security-oriented management of socio-economic systems based on Fuzzy Logic. Computer Systems and Information Technologies , (3) (2024) 86-91. doi:10.31891/csit-2024-3-11.