Reflex Intellectual Text Processing Systems: Natural Language Text Addressing

Serhii Lenkov [0000-0001-7689-239X], Mykola Kubyavka [0000-0003-1745-7026],
Liubov Kubiavka [0000-0002-5141-9886], Yevhen Lenkov [0000-0002-3784-5790],
Valerii Shevchuk [0000-0002-6950-4827]

Military Institute of Kyiv National Taras Shevchenko University, Ukraine
Kyiv National Taras Shevchenko University, Ukraine
Central Research Institute of the Armed Forces, Ukraine
University of the State Fiscal Service, Ukraine

lenkov_s@ukr.net, nick.kub@ukr.net, klb_it@ukr.net,
torwer007@gmail.com, shevchuk764@ex.ua

Abstract . The article describes the method of automatic response to the content of the text of the message, which was based on a probabilistic-reflexive approach. The reflexive approach provided the choice of the most probable response to the set of input influences, with known probabilities of choosing the response for each input effect, as well as some combinations of input influences, and the method developed on its basis allowed to automatically determine the destination of the analyzed text.

Keywords : Information Technology; natural language text; reflexive approach; methods of information management.

1 Introduction

The ability to think is a true wealth, only the presence of strong feelings and thoughts proves you do not stand still, these thoughts are helping us to develop more than time and space, against those we are nothing. Thoughts, views and dreams are changing as we speak. It depends on our surrounding, mood and emotions. Leo Tolstoy said: "Any thought, expressed in words, is a force whose action is boundless" [1].

Our consciousness may not be limited to words, but we are able to convey thoughts, emotions and wishes to people taking in the form of words. Language captures our knowledge of the world and allows us to express our attitude towards it. It is the main means of human intercourse. Key words or simple patterns will not be able to cover the needs of modern man on natural language text information processing. As has been held by several scholars, a solution could be attributed to the use of artificial

intelligence. Thus, the information extraction (meaning of the text) or reading comprehension are one of the challenging tasks in the field of information technology and artificial intelligence. Although there are plenty studies on the construction of natural language processing systems [2-14], which arose by merging of computer linguistics and information retrieval systems [2]. Nevertheless, the implementation of problem-oriented intellectual systems that would be capable of selecting the necessary information from the natural language and applying this correctly was and remains a relevant problem of the present.

2 Research results analysis

Any information we undergo shape our entire lives and that could be expressed as text. This paper discusses the natural language texts processing and analysis using the influence of such information to us instead of basic lexicon-semantic approach. Different information affects differently, and this is due to a phenomenon such as its addressness. If the message is not correctly addressed it will not be understandable, and on the contrary, the address-oriented message will affect the recipient and will allow him to understand its meaning.

The information management models and methods [15] based on the results of the interaction theory in the context of means of processing natural linguistic information will be used to address the problem.

The reflex probabilistic approach was used to identify the most probable reaction to the incoming message, with known probabilities of choosing the reaction for each input effect, as well as some combinations of input effects [12]. On its basis, the reflex method of automatic response to the message content and the model for determining the most informative components of the natural language text are developed. Any natural language text is considered as a message perceived by the intellectual apparatus of a person.

Here are some definitions to take the point further.

Definition 1. Text addressness - is the semantic component of the information, which determines the information direction of the message content, has corrective influence on the awareness (world view) of the recipient.

Definition 2. The message recipient can be determined from the message itself.

Definition 3. The concept of text length is a quantitative indicator of information and conceptual segments in the message part (text).

Definition 4. According to the information disseminated (contractors), it should be understood that the notice is communicated and does not lead to changes in these consumers.

By notice, it will mean any information (telecast, newspaper articles, internet notes, etc.) that will be sent to consumers. It`s received receptions (can take) necessary decisions for us. For the entire flow of information (all messages) used and technologies that provide impact on the interaction contractors, it is necessary to distribute to the address. It has also been recognized as the classical interaction partners (consumer information), and which classical messages will be published in the names of the technologies. Please inform that the information about who to send the message is in the message itself.

Definition 5. Identification of the necessary relation to the received information - is the definition of the actual readiness of the recipient to take one or another decision in a given situation in the direction of the desired party, exercising influence.

Definition 6. Identification of forms, methods and content of information influence on counteragents of interaction - is determining the ways of providing information and the content of the message itself, which will provide targeted impact on the target audience in accordance with the objectives of influence. The recipient must also provide false information if he so affects him that he will take the necessary (appropriate for the purpose of the effect) decision.

Definition 7. Managing impulse awareness - is organizing, planning and verifying such information actions for the recipient who can form an adequate target for the impact of the relation to the actual and predicted situations. This, in turn, will increase the likelihood of making the right decisions.

And so, the first step, it is necessary to decompose the input text into clusters, among which there will usually be those that determine the necessary reaction under the given conditions. Since it is impossible to know in advance what these clusters are, we will create a series of combinatorial regularities of the input text and select the ones that most affect the recipient.

The acknowledgement that the informational influence of the natural language text leads to an increase or decrease in the probability of the desired reaction in some classes of recipients is the basis of the paper:

$$p_k(A_l) \neq p_k(A_l / I_j),$$

where A_i *is* the recipient reaction (behaviour); Ij *is* information influence of the message; $Pk(A_i)$ *is* absolute probability of the reaction (behaviour) A_i in the class of recipients Q_k, if this class does not get influenced; p_k *(A/lj) is* the reaction (behaviour) probability A_l in the class of recipients Q_k after the influence lj.

In addition to the determination of its impact on the recipient, it is necessary to select for each message such classes of recipients, which impact will be maximal. For the given lj and A_l we should choose a class of recipients Q^*, which impact $p_k(A/lj)$

will be maximal:

$$Q^* = \max_k \left(p_k (A_l / I_j) \right),$$

where $Q * is$ the class of recipients is chosen to implement the impact.

Thus, the function finding that determines the difference between the reaction probability R_t for the recipient class Q' without influence and with that Aj is a high-priority problem.

The information influences repetition is a way to solve this problem, but in practice there had been no such cases. Each effect is combinatorial and usually consists of many separate, sometimes insignificant elements. Then this problem will be solved in the following way. If message is informative and significantly increases the reaction probability, it means that the message parts (separate sentences, words, combination of letters) are the influence holders on the counterparty. When these parts arise, the reaction probability changes. However, as stated above, the messages are not actually repeated, but the parts definitely are. Therefore, the probability change of one or another reaction can be assessed not from the whole message, but from some of its part.

To obtain this information, the following method is used.
1. An expert is provided with a text message and a list of recipient classes with their properties.
2. The expert assesses the information and provides for such messages the most likely addressees (whose probability of the desired reaction will be changed).
3. The statistics accumulation is made possible through the involvement of many experts.
4. Based on the information received, recipients for new text messages are found.

A solution to the problem 4 requires the development of method by which the addressees for new messages are determined on the basis of an expert assessment. Existing methods of solving this problem are quite complex, require the use of linguistic analysis tools that are almost impossible to use for messages of different directions, or their creation will be too expensive, since linguistic systems are usually created for separate text classes.

One of the varieties of intellectual systems is reflex intellectual systems, which are nothing more than software or technical systems that form reactions to non-power influences, the basis of which is the reflex algorithm that operates on the principle of forming reaction-response (reflex) on the set input data (external influence). But this approach has not yet been used by anyone to identify the addressees of the message. Therefore, it is proposed to use the probabilistic-reflexive approach to solve the

problem. The basis of the reflexes is the following thesis: if it already was, and some reaction was positively reinforced, then it is necessary to do the same.

The reflexion approach provides the choice of the most probable reaction to the infinity of input influences, with known probabilities of choosing the reaction for each input effect, as well as some combinations of input influences.

Considering the method of counterparty selection, where the probability of the desired reaction is greater than the limit value. The method is based not on the probability calculation of addressing messages, but on the assessment of this probability for the deviation of conditional and unconditional probabilities of such addressing from parts of natural language text. The use of conceptual information and relevant content forms in the method as well as means of information influence are complementary, not alternative. This would affect the method of selecting the necessary objective-oriented information is used to calculate the amount of information action on the recipient.

Definition 7. The method of determining the message recipients - is a search, selection and presentation method of relevant information to a target audience, the application of which will lead to the necessary responses to those changes the awareness (world view) of the target audience.

The addressed information is the result produced by the intellectual apparatus of a person. The semantic component of information, which is the core and product of communication in the intellectual apparatus of a person, determines its addressness. The correct definition of the natural language text addressness is response to the message. The addressness implies that one should understand the person to whom this message is addressed.

Definition 8. The addressed information is the result produced by the intellectual apparatus of a person. The semantic component of information, which is the core and product of communication in the intellectual apparatus of a person, determines its addressness. The correct definition of the natural language text addressness is response to the message. The addressness implies that one should understand the person to whom this message is addressed.

The reaction caused in the process of the addressness determining may be wrong. And furthermore, the same result (assess of one or another addressness) may correspond to the reality or be disconnected with that at all.

When using the probabilistic-reflexive approach, it is necessary to determine what is the influence, and what are the reflexes in the means of information technology.

From definition 2, the influence is the natural language text (message), and from definition 4 it follows that the reaction is a class of recipients. Then the possibility of obtaining the correct addressee can be represented through a certain probability.

$$p_i \approx b_i / k_i, \tag{1}$$

where p_t *is* the probability that the produced will necessarily bring the desired result (correctly addressed message); b_i *is* tests conducted to correctly identify the information addressness; k_i *is* total number of tests.

Determining the addressness of text messages based on the evaluation of the results of past actions can be presented in the form below.

Table 1. Statistical data of the response development to determine the text addressee

Addressness	The number of times the targeting of text clusters was determined	The number of times the desired result was obtained
A_1	k_1	b_1
A_2	k_2	b_2
...	...	...
A_i	k_i	b_i
...	...	...
A_m	k_m	b_m

First, you need to determine what affects the addressness.

Definitions 1 and 2 provide an answer to this question. The message along with the natural language text fragments, by which that could be transmitted. Such fragments are called the elementary influence unit.

Definition 9. The elementary influence unit will be considered a separate symbol, letter combination, composition, word, phrase, sentence, text excerpts, image, statement, sound, etc., that smallest piece of information provided to the recipient for perception is perceived by him and leads to a complete change of his world view.

Through the fragments in messages addressed to one or another class of recipients, the probability of sending a message to these recipients can be estimated.

Numerical measure of influence on the characteristics of addressness text determining is used as a basis for solving the problem. The examination of various influences on the process of the information addressing identification will help to identify the dependence of the automatically determined information addressness from the addressnes determined by an expert.

An existing model for identifying the most informative components of the natural language text will be applied to assess the probability of a well-defined recipient.

The message is decomposed into numerous fragments:

$$I = \left\{ i_f \right\}, f = \overline{1, g}, \tag{2}$$

where I *is* information (message); if *is* the often-repeated fragment; g *is* the number of fragments.

Successfully identifying the necessary information addressing will facilitate the determination of approximate probability values of each of the addressing information classes subject to the availability of new addressing fragment.

$$\forall A_j \in A, i_f \in I : p(A_j / i_f) \approx n(A_j / i_f), \qquad (3)$$

where $n(A_j / i_f)$ *is* the frequency of the addressness determining A_g subject to the availability of the fragment; $p(A_j / i_f)$ *is* the addressnes determining probability A_g subject to the availability of the fragment i_f.

By addressing *Ai* under its unconditional probability $p(Aj)$ and partial conditional probabilities $p(A_j / i_{f_1}),..., p(A_j / i_{f_i}),..., p(A_j / i_{f_g})$ the compatible conditional probability can be estimated:

$$p(A_j / i_{f_1},...,i_{f_i},...,i_{f_g}) = p(A_j / I), \qquad (4)$$

The problem of finding a compatible conditional probability of a correctly determined information addressing cannot be solved by using the probability theory methods, taking into account the partial conditional and unconditional probabilities. The first step is to define the issue clearly in order to estimate a compatible conditional probability by partial. We should apply the evaluation method for compatible conditional probability by partial that will enable us to select the addressing that would have been determined by the expert. For instance:

$$\forall A_g \in A \exists A_g \, p(\eta_{Mo}(A_g / I) \geq \eta_{Mo}(A_g / I) / p(A_g / I) \geq p(A_g / I)) \approx 1, \qquad (5)$$

where $\eta Mo(A_g / I)$ *is* assessment of the compatible conditional addressing probability A_g of information $I,$ which was obtained by the evaluation method for compatible conditional probability by partial; Mo *is* the evaluation method for compatible conditional probability by partial; $p(\eta_{Mo}(A_g / I) \geq \eta_{Mo}(A_g / I) / p(A_g / I) \geq p(A_g / I))$ *is* the conditional probability whereby conditional probability of information addressnes determining A_g reaches maximum, then the compatible conditional probability evaluation of this addressing A_g is maximal.

Definition (7) shows that if the compatible conditional probability of the recipient determining Ag reaches maximum, then almost always its evaluation is maximal. An optimal Mo method always gives the highest evaluation of the greatest $p(A_g / I)$, $p(A_j / i_{f_1}),..., p(A_j / i_{f_i}),..., p(A_j / i_{f_g})$, can be obtained compatible conditional

probability. To assess the effectiveness of this method the experimentally.

The deviation of this equation from the unit is a criterion for the effectiveness of the evaluation method for compatible conditional probabilities by partial:

$$p(\eta_{\text{Mo}}(A_g/I) \geq \eta_{\text{Mo}}(A_g/I)/p(A_g/I) \geq p(A_g/I)) \rightarrow \max, \text{, with restrictions:}$$

• the data collection I is decomposed into numerous fragments I_f,

• there is information with the highest evaluation of addressing information influence Ai;

$$\forall A_g \in A : p(A_g/i_{f_1}),..., p(A_g/i_{f_i}),..., p(A_g/i_{f_g});$$

$$\forall A_g \in A \exists A_g \in A : p(A_g/I) \geq p(A_g/I);$$

$$\forall E_g \in A : p(A_g/i_{f_1}),..., p(A_g/i_{f_i}),..., p(A_g/i_{f_g}) \xrightarrow{Mo} \eta_{\text{Mo}}(A_g/I); \qquad (6)$$

A mathematical model presented in the paper [16] will be applied not in relation to influence on the counterparty, but to the addressee information identifying.

Using values from Table 1 in the formula (1) of work [16] we obtain a new formula for assessing the influence of some fragment if on the message addressing Ai:

$$1. \text{ If } k_i > b_i \wedge b_i > 0 : \ w(A_i/i_f) = \frac{b_i \cdot (1 - p(A_i))}{(k_i - b_i) \cdot p(A_i)}; \qquad (7)$$

$$2. \text{ If } k_i = b_i \vee b_i = 0 : \ w(A_i/i_f) = \frac{(b_i + 1) \cdot (1 - p(A_i))}{(k_i - b_i + 1) \cdot p(A_i)}, \qquad (8)$$

where $w(A_i/i_f)$ is influence assessment of the message fragment i_f on the response (the addressee choice) A_i; b_i is number of tests, when the necessary information addressing was correctly determined; k_i is total number of tests.

Every message consists of numerous fragments. Then the total effect will be determined from the formula (1) [7]:

$$w(A_i/I) = \sum_{f=1}^{g} \sqrt{\frac{(w(A_i/i_f) - 1)^2}{w(A_i/i_f)}} = \sum_{f=1}^{g} \frac{w(A_i/i_f) - 1}{\sqrt{w(A_i/i_f)}}, \qquad (9)$$

where $w(A_i/I)$ is evaluation all influences on the addressee choice.

Using these formulas, estimation of the probability of choice we can proceed the assessment of the new probability of addressee choice A_i:

$$p(A_i \ / \ I) = \frac{w(A_i \ / \ I) \cdot p(A_i)}{1 + p(A_i) \cdot (w(A_i \ / \ I) - 1)},$$ (10)

where $p(A_i \ / \ I)$ is the evaluation probability of the addressee choice A_i.

Since there are many recipients, the probability is estimated for each of them. The recipients whose estimated probability is more than 0.5 are chosen.

The expert information is needed on how the message classes are addressed to the recipients (recipient classes). Let's consider this question.

To determine the quantitative measure of the message classes (forms) influence on recipient classes, expert analysis of processes accompanied decisionmaking on the message implementation in various influence components and the various functional roles performance has been undertaken. Experts were asked to identify the subjective probability of selecting one or another form of text messages in various influence components and the functional roles implementation.

Definition 10. Subjective probability - is an expert assessment of the probability of a particular event occurrence.

A sociological survey of school-aged children, preschool-age children and psychologists working in these educational institutions was conducted, an analysis of social surveys on the Internet and the media on these issues, a survey of students and elderly, as well as experts in the field of sociological research, the integrated expert assessment of the subjective probability of implementing one of the model attributes was formulated. The subjective expert relation to the implementation of various message forms in different conditions could be considered in the context of probabilities.

In the process of analysis, subjective-probabilistic performance indicators were identified:

- the class of messages for children;
- the class of messages for working youth;
- the class of messages for young people;
- the class of messages for entrepreneurs;
- the class of messages for employees;
- the class of messages for servicemen;
- the class of messages for pensioners.

The subjective-probabilistic implementation indicators of the message class were found to be applicable to recipient classes [16], and the data obtained were included in the expert tables.

A recent analysis is a basis for assessing the information actions importance in the

process of influence management. This contributed to identifying those forms of text messages that can maximally effectively ensure the implementation of these actions on the recipients. The information provided in the expert tables has become the basis for determining the necessary in formation of optimal, rational or appropriate informational influences on recipients by means of communication. The effective implementation of message tools requires the methods development for identifying information actions and the required forms of communication based on the received subjective-probabilistic characteristics of the scope of application.

The expected outcome of this task is a summary table, in a situation or if there is a problem in the influence management process, evaluating one of the influence components and knowing the information type (functional role), the most suitable and informative message (natural language text) can be easily identified. To find a text that will form the necessary relation to the reality (necessary knowledge) of the recipient would be easier with those summary tables. The use of the above method helps to identify the recipient and the text defining to be sent to the specified addressee.

This method is not about "understanding" of a limited natural language, but system that perceives information representations in a natural language without restrictions.

3   Conclusion

In summary, the authors consider it is necessary to mention such remarkable features of human intelligence as its extraordinary flexibility and mobility. Indeed, as soon as a person learns something, the field of his knowledge immediately expands. One of the main aspects in human learning is not the more he learns the better he can solve a new task, but rather to be able to match, join and combine with all the previous tasks in order to make a decision. The so-called reflex intellectual systems are based on this peculiarity of human consciousness, which are nothing more than software or technical systems that form reactions to influences, based on the reflex algorithm built on introformational methods [12] and works on the principle of forming reaction-response (reflex) on the set of input data (external influence).

The mathematical apparatus of the interaction theory [12] is quite simple and convenient for the implementation of processing natural language text systems. During the implementation of the above method, existing models for identifying the most informative components of the natural language text from the position of automatic addressing of these messages to different classes of recipients have been improved. The model innovativeness is to use a probabilistic-reflex approach to determine the natural-language text addressee. The proposed model differs from the existing in the difference identification of the conditional probabilities of the recipients occurrence as

the information influence extent of the natural language text fragments, which automatically identified the most likely addressees of this text.

The developed technology meets the requirements for simplicity, is informatively clear and minimal costly. Its foundation is based on the models and methods of influence management [15, 16], and the results are getting closer and closer to mankind to create a complete artificial intelligence.

4   References

1. Templeton J.: World laws of life, http://www.mirradosti.com.ua/?p=1669, last access 2019/05/27
2. Popov E.: Communication with a computer in a natural language. Nauka, Moskva (1982).
3. Bolshakov E., Lande D., Noskov A., Klyshynskyy E., Peskova A., Yahunova E.: Naturally texts automatic processing on language and computer linguistic. MYEM Publ, Moskva (2011).
4. Brill E., Mooney R. J.: An Overview of Empirical Natural Language Processing. In: AI magazine, vol. 18(4), pp. 13-24 (1997).
5. Pasichnik V., Vysotska V., Shestakevych T., Shcherbina Y.: Mathematical linguistics. New World Publ., Lvov, Ukraine (2012).
6. Fedorchuk A.: Content Monitoring of information flows. In: Nat. Acad. Science Problems. Functioning, Trends of development, vol. 3(12) (2005).
7. Grishman R.: Computational Linguistics: An Introduction. Cambridge etc.: Cambridge University Press (1986).
8. Bondarenko M.: Osnovy teoriyi syntezu nadshvydkodiyuchykh struktur movnykh system shtuchnoho intelektu. Kiev, Ukraine (1997).
9. Belonogov G.: Komp'yuternaya lingvistika i perspektivnyye informatsionnyye tekhnologii. In: Nauchno-tekhnicheskaya informatsiya, vol. 2(8), pp. 30-43 (2004).
10. Teslya Yu., Kosenkova O.: Zastosuvannya teoriyi informatsiynoyi vzayemodiyi do pobudovy system smyslovoho poshuku informatsiyi. In: Radioelektronika i informatika. vol. 1, pp.103-105 (2000).
11. Fayn V.: Raspoznavaniye obrazov i mashinnoye ponimaniye yestestvennogo yazika. Nauka, Moskva (1987).
12. Teslya Yu.: Vvedeniye v informatiku prirody. Maklaut, Kiev, Ukraine (2010).
13. Teslya Yu., Timchenko A.: Opyt razrabotki i primeneniya v stroitel'stve instrumental'nykh programmnykh sredstv yestestvenno-yazykovogo obshcheniya. In: Tezisy dokladov na II mezhdunarodnom nauchno-tekhnicheskom seminare "Teoreticheskiye i prikladnyye problemy modelirovaniya predmetnykh oblastey sistemakh baz dannykh i znaniy". MGP "Tirazh". Tuapse, pp. 226-228 (1993).
14. Cunningham H., Humphreys K., Gaizauskas R., Wilks Y.: Software Infrastructure for Natural Language Processing. In: Proc. of the 5th Conference on Applied Natural Language Processing. Washington, pp. 237-244 (1997).
15. Kubiavka M.: Modeli ta metody upravlinnya informatsiynym suprovodzhennyam v umovakh hibrydnoyi viyny: synopsis for Ph.D. thesis in Engineering Science: 05.13.06, Kyiv. Taras Shevchenko nat. uni. of Kyiv (2017).
16. Kubiavka M.: The control of information support of the processes of influence on the opponent. Zbirnyk naukovykh prats Viiskovoho instytutu Kyivskoho nats. Univers. imeni Tarasa Shevchenka. vol. 54, pp. 174-181 (2017).