УДК 81’243’366
DOI 10.31494/2412-9208-2022-1-3-379-388

ПРОБЛЕМА СПІВВІДНОШЕННЯ РУЧНОГО ТА АВТОМАТИЗОВАНОГО ОЦІНЮВАННЯ МАШИННОГО ПЕРЕКЛАДУ

THE PROBLEM OF CORRELATION BETWEEN MANUAL AND AUTOMATIZED ASSESSMENT OF MACHINE TRANSLATION

Irina SUIMA,
Candidate of Philological Sciences,
Associate Professor

Ірина СУЇМА,
кандидат філологічних наук,
доцент

irinasuima2017@gmail.com
https://orcid.org/0000-0002-2209-8614

Oles Honchar Dnipro national university,
✉72, Gagarin avenue, Dnipro, 49010, Ukraine

Дніпровський національний університет імені Олеся Гончара,
✉пр. Гагаріна, 72, м. Дніпро, 49010, Україна

Original manuscript received: October 26, 2022
Revised manuscript accepted: November 16, 2022

ABSTRACT

The article under review outlines the problems of development and assessment of machine translation that can greatly facilitate global communication, despite the imperfect quality of the source text. Most often the results of online tools require post-editing and can only be effectively used by those who already speak the target language to some extent. The need for a competent translation is growing every year. Today, the search for an algorithm to deliver this quality of translation is one of the most important questions in computer science and linguistics, therefore informing the scientific relevance of this work. It is analyzed different approaches to the machine translation systems, their characteristics, efficacy and the quality of their output. Different approaches to the machine translation systems, their characteristics, efficacy and the quality of their output are analyzed in the article. The main problems we see arising from such translations goes from the fact that the systems depend on a large amount of high-quality data sets (i.e., corpora of texts for specific language pairs). The quality of these sets directly influences the quality of the output, which in our case is the quality of the target language text. It can be seen by comparing the average quality of translation between Google’s and Microsoft’s systems. The former one makes less mistakes on average and does not have as many issues in regards to identifying a contextual meaning of a polysemantic lexeme.

It is underlined in the article, that this issue can be fixed to a certain extent one of two ways: hiring professional translators and linguists to compile those parallel corpora or create a possibility for every person to contribute to this process even on a small scale. The first approach would be very time and labor consuming, but would ultimately provide us with a higher quality data set, which may lead to further improvements in MT. The second is already being deployed by all three major NMT systems but may lead slower progression due to lack of quality control and oversight.

The potential prospect of this research is seen in widening the subject area of texts chosen to reflect the variety of writing styles in use on the Internet right now. Inclusion of texts from confessional, business, and other styles may allow us to highlight more lacunae in the neural network models and to suggest further means of improvement.

Key words: *machine translation, target language, source language, improvement, contextual meaning, communication.*

Introduction. A cursory examination of sources on computer technology developed for translation and word processing shows that the problems of machine translation and pattern recognition are closely linked to the problems of artificial intelligence and cybernetics. The problems of creating an artificial resemblance of the human mind to solve complex problems and model mental activity have been studied for a long time.

Machine translation quickly became not just a theoretical discipline but a cornerstone of scientific cooperation, sitting on the crossroads of computer science, engineering and linguistics [1; 2; 3; 4; 5]. The first generation of machine translation systems was based on sequential translation algorithms, that could on translate word-by-word, phrase-by-phrase. The capabilities of such systems were determined by the available vocabulary sizes, which directly depended on the amount of addressable computer memory. Translation of the text was carried out in separate sentences with meaningful connections between them not being taken into account. Such systems are called direct translation systems. Later on, they were replaced by subsequent systems, in which the translation from language to language was performed at the level of syntactic structures. The translation algorithms used a set of logical operations, with the following steps [6]: Translation system consists of three stages: a translation model, a language model, and a decoder [1 : 143].

The first generation of machine translation systems was based on sequential translation algorithms, that could on translate word-by-word, phrase-by-phrase. The capabilities of such systems were determined by the available vocabulary sizes, which directly depended on the amount of addressable computer memory. Translation of the text was carried out in separate sentences with meaningful connections between them not being taken into account. Such systems are called direct translation systems. Later on, they were replaced by subsequent systems, in which the translation from language to language was performed at the level of syntactic structures. The translation algorithms used a set of logical operations, with the following steps [7]: 1) analyzing the translation sentence; 2) constructing its syntactic structure according to the rules of grammar of the source language; 3) transforming it into a syntactic structure of the original sentence according to the target language grammar; 4) synthesizing the original sentence, substituting the right words from the dictionary. Such systems are called T-systems (from the word «transfer»).

Building machine translation systems based on obtaining some meaningful representation of the input sentence through its semantic analysis is considered to be the aim of machine translation. It should then be followed by a synthesis of the sentence in the target language according to the obtained meaningful representation. Such systems are called I-systems (from the word «interlingua»). It is generally believed that the next generations of machine translation systems will belong to the class of I-systems [8].

ICV 2021: 85.25
DOI 10.31494/2412-9208-2022-1-3

At the present stage of research, there are two main incentives for the development of machine translation. The first is purely scientific, it is determined by the complexity and intricacy of machine translation models. As a type of linguistic activity, translation affects all levels of language – from grapheme recognition to conveying the content of individual sentences and text as a whole. There is a need to accelerate the process of and increase the volume of translation, thus increasing the requirements for translation as an industrially applicable product.

The second incentive is social. It is driven predominantly by the growing role of translation in the modern world as a prerequisite for the provision of interlingual communication, the volume of which is increasing every year.

Methods and methodology of research. Creating an optimal and effective methodology for translation quality assessment (TQA) is a problem for various reasons. In particular, in order to decide which functions actually constitute a good translation, a number of other factors have to be taken into account, such as, for example, the genre of the text being translated or the purpose of the translation. Thus, in the process of evaluating the translation of a legal document, more attention should be paid to the aspect of accuracy that does not apply to literary translation. Another problem is the purpose of the translation, i.e., if the translation is made only for internal review and usage or for publishing.

Finally, the purpose of the evaluation itself plays an important role, whether it is the evaluation of the work of a professional translator for a monetary compensation, the quality check of translation within a particular translation service provider, the comparison between MT systems, measurement of the development of MT systems over time, etc. All these factors create difficulties for the development of a single unique indicator that would be suitable for any purpose and condition. Researchers in translation studies address this problem by defining what is a good translation and how it should be evaluated when covering the subject of translation as a phenomenon. Each of the approaches focuses on different aspects of translation quality. For example, House's functional model [9 : 113] depends on the situational features of the source text and translation and their comparison from a functional standpoint.

Results and discussions. The main evaluation parameter for this approach is the functional equivalence of the two texts or, in other words, how well the goal of the translation coincides with the purpose of the original text. However, this approach has some drawbacks, as the translation is not always done with the same purpose as the original text.

Some approaches have already been criticized for focusing too much on the purpose of the text, as they do not have an accurate quantitative model for translation quality assessment and can offer only a generalized idea whether a translation is good or not. Thus, theoretical approaches to QA provide strong arguments for those aspects to be taken into account in order to assess translation, but, on the other hand, they do not create a practical tool that can meet the needs of the industry and be used in on a daily.

ICV 2021: 85.25
DOI 10.31494/2412-9208-2022-1-3

Recent papers often offer a method based on the error calculation scale. They have much in common with existing industry standards, and essentially consist of a classification of errors.

Evaluation methods are required to be accurate and quantitative to be able to register even the smallest differences. Since the evaluation of MTs should be conducted on a regular basis, the procedure should be clear and relatively quick and simple. In contrast, we see that many methods of evaluating human (authored) translation only provide a theoretical basis for translation quality, leaving the rating procedure unclear.

In this section, we will review MT-specific QA methods, both automatic and manual, and discuss their advantages and disadvantages, as well as their similarity to human translation evaluation. Keep in mind that this paper mainly aims to depict the general long-term changes in FAMT systems and will be relying on manual human evaluation of examples rather than one of the quantitative automatic methods below.

The fastest, most affordable, and easiest way to measure MT quality is by employing an automated method of testing. The general idea is that a good automatic translation is one that is close to human translation.

The segments of the automatically translated text are compared to the one or more segments of reference human translations. This can be done word-by-word or phrase-by-phrase.

The most established automatic metric in the field of translation is the BLEU (Bilingual Evaluation Understudy). This was one of the first metrics that had a high correlation with human evaluation methodology. The BLEU score is based on one or more reference translations. They are compared to the source text by segments, usually sentences, and the scores for all segments are averaged for the whole body to obtain the total translation quality score, which is always a value in the range from 0 to 1, with 1 meaning that the output is identical to the reference translation. Because even human translation is almost never going to be identical to the reference translation, it is virtually impossible to reach the value of 1. By using several reference translations, one can increase the BLEU score, as there is a greater possibility of compliance with the reference [8].

BLEU metrics are based on calculation of accuracy. However, accuracy here simply counts the number of unigrams (words) in a possible translation that occur in any reference translation, and then divides by the total number of these word variants in the possible translation. At the same time, the results show that MT systems can generate grammatically incorrect expressions that will receive a high score and therefore machine translation in this example will get an exact score of 1. Thus, BLEU uses modified accuracy algorithm, which is calculated by first counting the maximum number of times a word appears in any reference translation.

The idea of BLEU evaluation is based on two concepts related to translation quality: accuracy and flexibility. A translation that uses the same words (unigrams) is considered accurate, while one that uses similar

ICV 2021: 85.25
DOI 10.31494/2412-9208-2022-1-3

structures (n-grams) is more freeform and flexible. Therefore, the greater the number of n-grams, the higher the flexibility of translation.

It has been widely acknowledged that the BLEU metric has certain flaws, one of which is the behavior with rule-based systems: although statistical systems are usually highly correlated with people's decisions when evaluating their translations, this does not apply to RBMT. Other major weakness of the metric is that it does not take into account synonyms or paraphrases.

Moreover, words have the same weight in the context of evaluation, so whether the system is omitting words that contain a specific necessary context or a simple article is irrelevant. Despite all these shortcomings, the BLEU metric is useful for estimating small differences in the same system and is still the most popular metric in the MT community.

The notion of accuracy and volume in the context of translation evaluation can be explained as follows: accuracy is the percentage of correctly translated words, and volume is the percentage of all translated words.

METEOR metric was created as an alternative to BLEU, but unlike the previous one, the key idea of the algorithm is to focus on the volume of translation rather than accuracy. In addition, it evaluates unigrams only, not taking into account n-grams [9 : 223].

The proposed translation is aligned to the reference text according to the algorithm that matches the lexemes in both SL and TL texts. Only two texts can be aligned, so if there is more than one reference text, alignment is done for each of them. Afterwards, the texts are compared, and the metric is calculated on the basis of their similarity.

A completely different but also interesting method is the Word Error Rate (WER) metric, which calculates how many replacements, deletions and insertions are needed to convert the MT text into a reference translation. This indicator has been used in various works, but has a significant disadvantage displaying different results depending on specific reference translation used in the process.

TER (Translation Edit Rate) measures the quantity of edits that a person must make in editing, so the original products of the MT exactly correspond to the reference translation. Based on their experiment, its authors claim that only one reference for the test with the TER methodology gives the same correlation with human-based evaluation as BLEU will have with four reference translations.

Other automatic metrics include PER, ROUGE, and although they provide a reliable way to evaluate MTs quickly and cheaply to observe improvements in MT system diagnostics or MT system comparisons, none are capable of achieving high enough quality to replace human judgment completely.

However, because they take into account only the length of the segments of the sentence, they do not focus on such properties of the text as intralingual references, style of the text or grammar peculiarities. Finally, automated metrics are comparative benchmarks, i.e., they are based on the idea of referencing one or more authored human translations, but they cannot take into consideration all the synonymous structures and paraphrases in the text.

ICV 2021: 85.25
DOI 10.31494/2412-9208-2022-1-3

A more accurate, but also more expensive and time-consuming method of quality assessment is done with the help of a human translator. Most automated metrics generally measure two quality characteristics of a translation – accuracy and speed. For example, they can be evaluated on a scale from 1 to 5, where each point is associated with a single characteristic. In addition, each measurement is sometimes evaluated by slightly different characteristics.

Manual quality assessment is relatively easy and simple to perform, but, on the other hand, it suffers from high subjectivity. Experts often disagree on estimates, and several evaluations are needed for a more accurate assessment. In addition, bilingual evaluators are not always available, so one would need to resort to reference-based evaluation, which is also one of the disadvantages of automated metrics.

For example, the quality of translation can be measured by clarity, conciseness, readability and accuracy. All this can be measured with the help of bilingual experts on a scale from 1 to 5, where 5 is given to translations when all information is stored with the source text, and a score of 1 for translation with virtually no information from the source text. When evaluators are bilingual, a good professional translation is necessary to serve as a quality standard for assessing the adequacy of MT translation.

Exploring the existing methods of assessing the quality of machine translation the equivalence of translation by levels should be taken into account: **translation of words > translation of phrases > translation of sentences > translation of text**. MT can be fairly accurately assessed at each of these levels, and even at the level of morphemes. The quality of the information conveyed via a translation is often measured by a task-based assessment where the translated text is used to perform a specific task, for example, to answer questions with multiple choices about the contents of the text or to extract specific information from automatically translated text.

The percentage of correct answers gives us a certain estimate of how well a MT manages to carry over the meaning and content of the original text. Another approach is to measure the time it takes for human evaluators to read segments of the translated text with some of the words replaced by spaces or underscores.

The number of correctly identified words is usually correlated with the readability of the text. In addition, there are metrics for translated texts MT based on editing effort. Editing score is measured by the number of corrected words, the amount of time spent or the number of keystrokes the editor must perform. Such metrics may require even more time and expense than other human assessment methods. In addition, they require a more clear procedure to calculate the final estimate based on the data obtained.

In the MT rating method, the systems are compared between each other, i.e. the user is provided with the original text and its translations, and they must order them according to perceived quality. These metrics are best at comparing different systems, but they do not say how effectively the «best» system actually works. In addition, they are also subjective, as it is not clear what exactly constitutes a best translation.

ICV 2021: 85.25
DOI 10.31494/2412-9208-2022-1-3

The use of error rating scales to assess MTs has certain advantages and disadvantages and requires special training, as evaluators need to study different types of errors. This requires a lot of time and well-trained bilingual evaluators; the significance of errors needs to be adjusted in the metric in accordance with the purpose of evaluation and characteristics of the texts. However, this method, due to the quantitative nature, allows for more accurate and gradual assessments compared to other methods, where quality is assessed from the point of view of quite abstract and nebulous concepts of good and bad translations.

Correlation can be defined as a measure of the similarity between human and automated MT quality assessments. Correlation is usually checked at two levels: at the sentence level, where scores are calculated by metrics for translated sentences, and then correlated with human scores for those same sentences. And at the corpus level, where sentence scores are added together for human judgment and judgment metrics, these scores are summed to indicate the ratio. Data on sentence-level correlation is rarely reported, although correlation figures that were given, show that at the sentence level the correlation is significantly worse than at the corpus level.

In practice, when machine translation is evaluated according to the aforementioned metrics, there are many serious problems, including problems of correlation of automatic evaluation with human judgments (expert evaluation) in regards to the quality of translation and results. It is clear that human evaluation of translation quality is the gold standard, but such methods are more expensive and time-consuming, so they are not always readily available. When evaluating a translation, one must take care to maintain objectivity. Almost all popular metrics (BLEU, METEOR etc.) are based on human judgments in some way, but do not always give an adequate assessment of the quality of translation and can only show correlation compared to professional translation.

Firstly, the problem of correlation between automatic and expert evaluation is closely related to the quality of the MT software with which the translation was performed, i.e., the accuracy and precision of the original machine translation, which we want to evaluate with the reference text. It is very unlikely that the metric will give the machine translation a score of 1, i.e., full correspondence with the reference text, because even with high-quality machine translation it is almost impossible to fully preserve the sentence structure that was in the reference translation. The point is, there can be multiple valid approaches for translating a sentence, each of them having varying grammatical structures and lexical contents. So, translations of the same text can have different word counts, i.e., different sentence constructions and lengths, which may affect the evaluation of the translation using automatic metrics. This suggests that the metric works better at the level of large corpora of texts than at the level of individual sentences, therefore the larger the volume of text is estimated, the higher the correlation.

Secondly, as a rule of thumb, there may be many options for correct translation of a long sentence and most of them will differ significantly in vocabulary, namely the correspondence of the vocabulary of the MT to the

ICV 2021: 85.25
DOI 10.31494/2412-9208-2022-1-3

standard translation and is the basis of all automatic metrics. Expert assessment is also quite subjective. Therefore, when studying the problem of correlation of automatic assessments with expert ones, as a rule, the former use several reference translations made by different people, and several expert assessments, from which the average value is taken.

Thirdly, it is important to understand that the absolute value of the metric is not important. For example, the BLEU result of 0.6 may be worse than the result of METEOR of 0.37. More informative is the relative difference in the results of one metric under different conditions. Therefore, more often automatic metrics are used to compare the work of different MT algorithms, different versions of one MT, the work of one MT with texts of different types etc. For a more transparent and accurate result, one can simultaneously assess the quality of the MT on several metrics and the quality of the final assessment using a combined value, such as the weighted sum of individual estimates. This may have considerable research value or provide an opportunity to consider possible translation adjustment methods.

It was also reported that the correlation depends on the type and subject matter of the translated text. Correlation is given a lot of attention when it comes to evaluating translations of texts that are not very similar in origin and sentence structure (for example, translation from English to Arabic).

Even if the metric correlates well with human judgments in one study in one corpus, this successful correlation cannot be transferred to another corpus. Good metric scores, by text type or area, are important for reusing metrics. A metric that works only for text in a specific area is useful, but less useful than one that works in many areas, because creating new metrics for each new score is undesirable due to data disparity.

Recent studies have shown that metrics can be used effectively to evaluate small texts. METEOR showed a very high result: up to 0.964 correlation with human judgments at the corpus level, compared to BLEU of 0.817 on the same data set. At the sentence level, the maximum correlation with the human score was reached for BLEU and is currently set at 0.403.

Conclusions. From our analysis we can conclude that the most capable NMT system for English-to-Ukrainian language pair is Yandex. Translator due to its enhaced ability to reconstruct the text in the target language with minimal grammatical losses. Google's algorithm follows closely but lacks the same understanding of TL's inherent grammatical categories, while Microsoft's system still largely relies of statistics for its output, and as such, cannot match both of them in the level of accuracy.

It is evident from our analysis that the most common mistakes include those regarding the grammatical structure of the sentence in the target language (inflections, cases, grammatical gender etc.). We strongly believe that these issues with sentence reconstruction stem from two main problems. They are: 1) authors of the original network model are mainly English-centric, and therefore are not aware of the differences during the process of training the model. This mostly applies to Google's and Microsoft's systems, as they are large multinational

ICV 2021: 85.25
DOI 10.31494/2412-9208-2022-1-3

corporations with based in English-speaking countries. Our hypothesis is confirmed by the fact that Yandex's (which is mostly a Russian-speaking company) system handles English-to-Ukrainian translation significantly better that either of the former implementations of NMT algorithms. 2) the corpora of the parallel texts for the English-Ukrainian language pair does not seem to be particularly large. This issue affects all three implementations in equal measure, as some of the concepts slipped past the training process, and as a consequence were not rendered properly and accurately in the target language.

As far as this research is concerned, this problem cannot be mitigated easily. One might suggest employing a number of translators specifically to gather a large corpus of parallel texts translated by professionals, on the basis of which a system may be trained. We believe this to be an unfeasible task, due to time and financial investments required. There is a potential in crowdsourced approach to the training of neural networks, but that may prove unfeasible as well due to inability to verify whether one's translation would be an accurate one.

To sum up: the remaining issues regarding the translation of sentence's contents require both more cognitive layers inside the neural network and much larger data sets for its training. And whereas the former problem will be solved eventually, it is the latter problem which may prove too difficult to overcome for many more years.

Література

1. Вознюк М. Ю. Критерії оцінювання перекладу. *Вісник ЛНУ імені Тараса Шевченка*. 2011. № 9 (220). С. 143-149.
2. Chan S. Routledge Encyclopedia of Translation Technology. Oxon: Routledge, 2015. 285 p.
3. Darwish A. Transmetrics: A Formative Approach to Translator Competence Assessment and Translation Quality Evaluation for the New Millennium. 2001. URL : http://www.translocutions.com/translation/transmetrics_2001_revision.pdf
4. Forcada M. L. Making sense of neural machine translation. *Translation Spaces,* 2017. Iss. 6. P. 291-309.
5. Gordin M. D. Scientific Babel: How Science Was Done Before and After Global English. Chicago, Illinois: University of Chicago Press, 2015. 233 p.
6. Hearne M. Statistical Machine Translation: A Guide for Linguists and Translators. *Language and Linguistics Compass 5(5).* 2001. P. 205-226.
7. Hutchins W. J. Machine translation: past, present, future. New York City, 1986. 236 p.
8. Hutchins W. J. Machine translation: a concise history. New York, 2005. URL : https://pdfs.semanticscholar.org/e97a/40cc28ce17a17ce9b73d77e69ffa1210fa25.pdf.
9. Jurafsky D. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2009. 2nd edition. Prentice Hall. 400 p.

References

1. Vozniuk, M. Yu. (2011). *Cryterii otsiniuvannia perekladu* [Criteria of translation assessment]. Visnyk LNU imeni Tarasa Shevchenko – Herald of Taras Shevchenko LNU, № 9 (220), Vol. II, 143–149 [In Ukrainian].
2. Chan, S. (2015). Routledge Encyclopedia of Translation Technology. Oxon : Routledge. [in English].

ICV 2021: 85.25
DOI 10.31494/2412-9208-2022-1-3

3. Darwish, A. (2001). Transmetrics: A Formative Approach to Translator Competence Assessment and Translation Quality Evaluation for the New Millennium. URL : http://www.translocutions.com/translation/transmetrics_2001_revision.pdf [in English].

4. Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces,* Iss. 6, 291–309 [in English].

5. Gordin, M. D. (2015). Scientific Babel: How Science Was Done Before and After Global English. Chicago, Illinois: University of Chicago Press. [in English].

6. Hearne, M. (2011). Statistical Machine Translation: A Guide for Linguists and Translators. *Language and Linguistics Compass,* 5(5), 205–226 [in English].

7. Hutchins, W. J. (1986). Machine translation: past, present, future. New York City. [in English].

8. Hutchins, W. J. (2005). Machine translation: a concise history. New York. URL : https://pdfs.semanticscholar.org/e97a/40cc28ce17a17ce9b73d77e69ffa1210fa25.pdf. [in English].

9. Jurafsky, D. (2009). Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2nd edition. Prentice Hall. [in English].

АНОТАЦІЯ

Стаття присвячена розглядові особливостей розвитку та оцінки машинного перекладу, що може значно покращити глобальні комунікації, незважаючи на недосконалу якість вихідного тексту. Найчастіше результати онлайн-інструментів вимагають постредагування і можуть ефективно використовуватись лише тими, хто певною мірою вже говорить мовою перекладу. Потреба в якісному перекладі зростає з кожним роком. Сьогодні пошук алгоритму для забезпечення такої якості перекладу є одним із найважливіших питань інформатики та лінгвістики, що вказує на наукову новизну цієї роботи. У статті здійснено аналіз різних підходів до проєктування систем машинного перекладу, їхніх характеристик, ефективності та якості вихідного тексту. Основні проблеми, які ми бачимо в таких перекладах, пов'язані з тим, що системи залежать від великої кількості високоякісних наборів даних (тобто корпусів текстів для певних мовних пар). Якість цих наборів безпосередньо впливає на якість виводу, у нашому випадку це якість тексту цільової мови. Це можна побачити, порівнявши якість перекладу між системами Google і Microsoft. Перший у середньому робить менше помилок і не має стільки проблем щодо визначення контекстного значення полісемантичної лексеми. У статті підкреслено, що цю проблему певною мірою можна вирішити одним із двох способів: використати знання професійних перекладачів і лінгвістів для складання паралельних корпусів або створити можливість для кожної людини зробити внесок у цей процес навіть у невеликому масштабі. Перший підхід забирає багато часу та праці, але в підсумку надає нам більш якісний набір даних, що може призвести до подальшого покращення якості перекладу. Друга вже впроваджується всіма трьома основними системами наукового машинного перекладу, але може призвести до сповільнення прогресу через відсутність контролю за якістю. Потенційна перспектива цього дослідження полягає в розширенні предметної галузі текстів, обраних для відображення різноманітності стилів письма, що використовуються в Інтернеті зараз. Включення текстів конфесійного, ділового та інших стилів може дозволити нам виділити більше лакун у моделях нейронних мереж та запропонувати подальші шляхи вдосконалення.

Ключові слова: *машинний переклад, мова перекладу, мова оригіналу, покращення, контекстуальне значення, спілкування.*

ICV 2021: 85.25
DOI 10.31494/2412-9208-2022-1-3