Neural Machine Translation as a Novel Approach to Machine Translation

Lucia Benková, Ľubomír Benko

Department of Informatics, Constantine the Philosopher University in Nitra, Nitra, Slovakia
lucia.benkova@ukf.sk, lbenko@ukf.sk

Abstract

The aim of the paper is to present the most used machine translation- Statistical Machine Translation system and introduce a novel system- Neural Machine Translation. Neural Machine Translation structure is built on an encoder-decoder framework. The encoder transforms a source language sentence into continuous space representation through a recurrent neural network. Origin of neural networks was inspired by the understanding of the functioning of the human brain, or all connections between neurons. However, in contrast with the human brain, where neurons can freely interconnect, artificial neural networks consist of discrete layers, connections, and data dissemination. This paper deals with neural machine translation as a novel approach that is examined by many researchers that try to implement it into already used frameworks. The results show that neural machine translation offers an improvement of the translation output but still has to be evaluated in the future.

Keywords

Natural Language Processing. Machine Translation System. Neural Machine Translation. Statistic Machine Translation.

INTRODUCTION

In the beginning, machine translation was based mainly on Rule-based Machine Translation (RBMT), the idea being to create grammatical rules for the source and target language. Machine translation acted as a kind of translation between languages based on this set of rules. However, the problem was mainly the addition of new content, new language pairs, because maintaining and extending such a set of rules was too time-consuming and costly. Statistical Machine Translation (SMT) was created to overcome this problem (Koehn, 2010). SMT systems create statistical models by analyzing an aligned set of source and target language sentences (training set). It is then used to create a translation. The advantage of SMT is its automatic learning process and relatively easy adaptation. The disadvantage of SMT is the training itself, so it is necessary to create a usable tool and a large database of source and target language segments. The disadvantage is also difficult to work with grammatically more complicated languages. Neural Machine Translation (NMT) has recently started to be promoted for this reason. NMT looks at the sentence as a whole and can form associations between phrases even at greater distances in the sentence. The result should be improved by grammatical accuracy compared to SMT.

ISBN 978-80-7598-841-6 ISSN 2464-7470 (Print) ISSN 2464-7489 (On-line)

SMT and NMT operate on a statistical basis and use pairs of source and target language segments as a basis. In principle, SMT is Phrase-Based Statistical Machine Translation (PBSMT), which means that SMT divides source segments into phrases (Koehn, 2010). SMT creates a translation and language model during training. The translation model stores various phrase translations, and the language model stores the likelihood of a sequence of phrases on the target page. During the translation, the decoder selects a translation that works best based on these two models. In principle, SMT can produce very good results at the level of phrases, but the fluency and grammar of the translation are lagging behind several times. The paper describes the novel approach of neural machine translation and its usage by other researchers.

The rest of the paper is structured as follows. The next section describes the research background and introduces the main topics of Statistical Machine Translation and Neural Machine Translation. The third section summarizes the usage of the novel approach to Machine Translation. The last section provides the conclusion.

RESEARCH BACKGROUND

Statistical machine learning (SMT) is an approach to Machine Translation that is characterized by the usage of machine learning methods. SMT treats translation as a machine learning problem (Lopez, 2008). The basis of SMT is to create a system that can automatically discover translation rules of the large bilingual corpus, merge starting sentences of text (input data) with target sentences (output data) and "be taught" by the results of the statistical analysis of relevant data (Koehn, 2010). Statistical machine translation deals with the translation of text from one natural language to another. Its approach to machine translation is characterized by the usage of machine learning methods. This means that the learning algorithm is applied to a large group of the previously translated text, referred to as parallel corpus, parallel text, bitext or multi-text. This approach uses the power of computers to create sophisticated data models capable of translating text from one language to another. Basically, statistical machine translation systems use computer algorithms to create a translation that selects the best and most likely statistically output of the millions of possible permutations.

The advantage of statistical machine translation systems is the removal of manual translator work for each language pair. On the other hand, the disadvantage is the restriction to a single region of texts (domain), i.e. if the system is trained on one type of corpus (e.g. administrative), then it should be used to translate administrative texts, not e.g. technical texts. The quality of the translation would be unpublished in this case, and therefore it is important to train the system with the corpus, which is thematically similar to the starting text (Munková and Munk, 2016). As statistical machine translation has evolved over the years, its systems have evolved and improved too. In the very beginning, separate word translation was used, but progress in machine translation and in science itself was mainly in rapid development. New systems, larger collections of parallel corpora, and more powerful computers have continually improved the quality of statistical machine translation.

ISBN 978-80-7598-841-6 ISSN 2464-7470 (Print) ISSN 2464-7489 (On-line)

The first systems for statistical machine translation were based on the translation of individual words. Although this system is no longer widely used, many of its principles and methods are still up to date.

The smallest units in this system are words that can be translated, inserted, omitted, or their order in the sentence changed. These systems are based solely on lexical translation - the translation of isolated words. It requires dictionaries that map the translation of words from one language to another (Koehn, 2010). Looking at a common vocabulary, we find that a word can have more meanings in a foreign language. Some of them are used more, some less. As an example, a translation of the German word *Haus* into English can be used. In this case, the English word *house* will in most cases be considered the correct translation. Options such as *building* or *home* are also common, while others are used only in certain specific circumstances, e.g. the word *shell* that can refer to the slug home.

The correct translation or the most probable possibility of translation is then selected using parallel corpora. Let's say that in the hypothetical text the word *Haus* would appear 10 000 times. Of which 8 000 would be translated as *house*, 1 600 times as *building*, 200 times as *home*, etc. Based on these calculations, it can be estimated the likelihood of a lexical translation. Formally speaking, the aim is to find a function

$$p_f : e \rightarrow p_f(e),$$

which will help in translating another German text to determine what translation of *Haus* is most likely. This function returns the foreign word *f* (in this case the word *Haus*) and the probability for each of the possible translations *e*. This will tell how likely it is to have the correct translation.

A machine translation system based on the translation of individual words was already mentioned. Words like the smallest translation units, may not be the best choice. Sometimes one word in a foreign language is translated into two English words or vice versa. Word-based models often diverge and differ in these cases. In more advanced statistical machine translation, the basic unit of translation is expanded from words to phrases of potentially unlimited length and may not be defined as phrases from a syntactic point of view (Chiang, 2007).

At present, one of the best systems for statistical machine translation is considered on phrase-based models - systems that translate a small sequence of words at once. In phrase models, any sequence of contiguous words can be considered a phrase. Each input phrase is non-empty and is translated exactly to one non-empty output phrase. However, phrases are not required to have the same length, so this model can produce translations of varying length (Lopez, 2008).

Phrase translation systems work by dividing the input sentence into segments - phrases (polyword units). Each of these segments is translated into the target language and the phrases are finally sorted. However, the number of phrases at the input and language targets may not match.

One of the basic elements in any statistical machine translation is a language model that measures the likelihood of a given word sequence that will be actually used by English-speaking person. It goes without saying that it is required of the machine translation system not only to produce output words that are correct with respect to the original text but also to put them in the right string (Koehn, 2010). The language model, however, usually does

ISBN 978-80-7598-841-6 ISSN 2464-7470 (Print) ISSN 2464-7489 (On-line)

much more than just allows smooth output. It supports difficult decisions on word order and word translation. For example, the probabilistic language model p_{LM} should prefer the correct word order instead of the wrong word order:

$$p_{LM}(the\ house\ is\ small) > p_{LM}(small\ the\ is\ house).$$

Formally speaking, a language model is a function that takes an English sentence and returns the probability that the sentence was created by an English-speaking person. Based on the example above, it is more likely that an English-speaking person would rather say a sentence *the house is small* than *small the is house*. Therefore, a good language model of p_{LM} assigns a higher probability of the first sentence.

This advantage of the language model helps statistical machine translation systems to find the right word order. Another area where the language model helps translation is the choice of words. If a foreign word (for example, German *Haus*) has several translations (*house*, *home*, ...), the more common translation (in this case *house*) will be favoured based on the likelihood of a lexical translation. However, other translations may be appropriate in certain specific contexts. Here the language model that gives a higher likelihood of a more natural choice of words in the context, is applied again. For example:

$$p_{LM}(I\ am\ going\ home) > p_{LM}(I\ am\ going\ house).$$

One of the main methods in the language model is the N-gram language model. N-gram is a term commonly used with language models for speech recognition. It can give the probability of the next word, based on the previous sequence of words from the training corpus (Yamamoto et al., 2003). The principle of modelling language using n-grams is that the model divides the sentence into several fragments (words/phrases) that often occur in the corpus and carry language information and determine the probability of individual fragments. If the fragments of the sentence are in the correct order, then the sentence should have a high probability (Munková and Munk, 2016). Returning to the example, after analyzing a great deal of text, it was identified that *going* is followed by *home* more often than *house*.

Formally speaking, in language models, it is anticipated to calculate the probability of a string:

$$W = w_1, w_2 \ldots w_n.$$

Simplistically, $p(W)$ is the probability of randomly selecting a sequence of English words (whether in a book or a magazine) and getting W. To calculate $p(W)$, it is needed to collect a large amount of text, where is calculated how often W is present. Most of the long word sequences, however, will not be found in the text at all. Therefore, it is required to analyze the calculation of $p(W)$ into smaller steps for which can be collected sufficient statistics and further divide the probability estimate.

Dealing with the limited amount of data that limits us in gathering enough statistics to reliably estimate the probability of distribution is a major problem in language models.

On the other hand, the Neural Machine Translation (NMT) structure is built on an encoder-decoder framework. The encoder transforms a source language sentence into continuous space representation through a recurrent neural network (RNN) from which the decoder generates a target language sentence using another RNN (Cheng, 2019). NMT uses deep learning, which in principle is represented by the neural network (Bessenyei, 2017;

ISBN 978-80-7598-841-6 ISSN 2464-7470 (Print) ISSN 2464-7489 (On-line)

Stencl and Stastny, 2010). Machine learning could simply be understood by algorithms that process data from which they can learn, and on that basis, they can make decisions or predict solutions to certain problems. Origin of neural networks was inspired by the understanding of the functioning of the human brain, or all connections between neurons. However, in contrast with the human brain, where neurons can freely interconnect, artificial neural networks consist of discrete layers, connections, and data dissemination. Neural networks use distributed, parallel information processing to perform calculations. Knowledge is stored primarily through the strength of links between individual neurons. The basic feature of neural networks is learning. The neuron receives signals from the environment from other neurons, processes them and sends them as input signals for the neurons in its surroundings. Multilayer neural networks consist of three layers (Fig. 1):

1. Input layer - the input is from the external world and the output is another neuron,
2. Hidden layer - input is from the external world or from other neurons, the output is another neuron,
3. Output layer - the input is similar to the hidden layer and the output is directed to the external world.

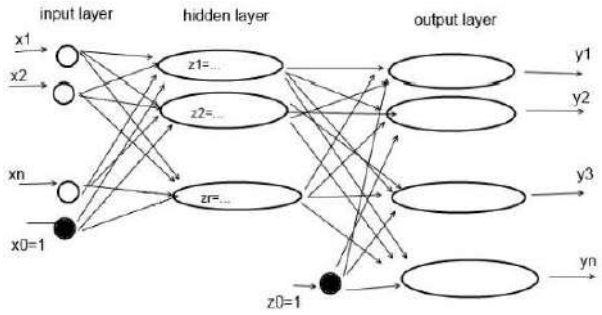

Figure1 *Multilayer neural networks* (Cheng, 2019)

Each neuron assigns some weight to its input. The weight represents the degree of fulfilment of the task being performed, the higher weight means the better solution. The final output of the neural network is thus affected by the total sum of the weights. The essence of machine translation is the different length of input $X = (x_1, x_2, \dots, x_T)$ and output $Y = (y_1, y_2, \dots, y_{T'})$. In other words, T and T' may not be the same. For this reason, it is necessary to use a special type of neural networks - recurrent neural network. The RNN retains its internal state as long as it reads the sequence of inputs, in this case, a sequence of words, and can process the input of different lengths. The goal of RNN is to compact the sequence of input symbols into a fixed vector by recursion. Recursion in simplicity means defining a function or method by itself. The overall architecture is based on the encoder-decoder principle (Kalchbrenner a Blunsom, 2013; Sutskever et al., 2014; Bahdanau et al., 2016; Cho et al., 2014).

The encoder is a straightforward RNN application based on sequential summaries, i.e. an activation function is recursively applied to an input sequence or sentence until the last input state of RNN is a summary of the entire input sentence. First, each word of the source sentence is represented as a so-called 1-of-K encoded vector. Words are equidistant from each other, which means that there is no relationship between words. A hierarchical approach is used to extract a sentence representation (a vector that summarizes an input sentence). In principle, the network will learn from data. The encoder then linearly projects

ISBN 978-80-7598-841-6 ISSN 2464-7470 (Print) ISSN 2464-7489 (On-line)

the 1-of-K encoded vector using an E matrix that has as many columns as there are words in the source dictionary and as many lines as the programmer chooses (typically 100-500). The projection results in a continuous vector for each source word. Each vector is later updated to improve compiler performance.

When a fixed sentence representation of a source sentence is created using the encoder and RNN, we use the decoder with RNN to create a translation. Starting from RNN, the internal state of RNN is calculated based on the source sentence summation vector, the preceding word, and the previous internal state. Using the internal hidden state it is possible to score each target word based on how likely it will follow all previously translated words based on the source sentence. This is possible by assigning probabilities to each word. The difference between score and probability is that the sum of the probabilities of all possible words equals 1, but the score does not need to be 1. Based on the score, the next step is to calculate the probability that serves to select a word by choosing from a multinomial distribution. After selecting the i-th word, it returns to the first step, calculating the hidden state of the decoder, evaluating and normalizing the target word, and selecting the next $(i + 1)$ word. The procedure is repeated until the end of the sentence (called <eos>) is reached. By using a neural network, translation performance can be maximized (Bahdanau et al., 2016).

Corpus is needed to train a neural network and the maximum log-likelihood estimation method is used (Cheng, 2019). Each corpus element is a pair of source and target sentences. Each sentence is a sequence of numerical indexes corresponding to words, which is equivalent to binary vectors (one element vector is set to 1). During the training process, the NMT system attempts to set the neural network weights parameter based on the reference values (translation from target to source language). Taking any pair from the corpus, the NMT system can calculate the conditional log-probability of the target sentence from the source sentence. The result is a neural network that can process source segments and transform them into target segments, with NMT passing through whole sentences, not just phrases. The advantage of this approach is precisely the appropriate context of the translation, which also improves the fluency of the translation. But the accuracy of the terminology can sometimes be insufficient.

MACHINE TRANSLATION APPLICATIONS

SMT is used for many years to produce the output for various language pairs. Authors in (Munková et al., 2014, 2013) focused on the preparation of text where it depends on the data sources used. The aim of this work was to determine to what extent it is necessary to carry out the time-consuming data pre-processing in the process of discovering sequential patterns in e-documents. Munkova et al. (Munková et al., 2020) focused on the evaluation of translation quality of sentences of the MT output and post-edited MT output. The authors' used metrics of automatic MT evaluation for a language pair Slovak-German. The MT translation was done using an SMT system. Munk and Munkova (Munk and Munkova, 2018; Munkova and Munk, 2015) introduced an exploratory data technique representing an instrument to evaluate and improve MT systems. The authors used residual analysis to identify the differences between an SMT system output and post-edited MT regarding human translation. Using residual analysis, the authors identified sentences that contained significant differences for the scores of automatic metrics between MT output and post-

ISBN 978-80-7598-841-6 ISSN 2464-7470 (Print) ISSN 2464-7489 (On-line)

edited MT output from Slovak into English. A system for post-editing the SMT system output was presented by (Munková et al., 2016). The aim of the study (Munkova et al., 2019) is to compare translation quality and effectiveness (translator productivity) using measures of the automatic evaluation of machine translation output. The examined translation(s) was a legal text, translated from Slovak (mother tongue) into German. We distinguish human translation (HT), machine translation (MT) and post-edited MT (PEMT). For the evaluation we used our own tool, wherein were implemented the metrics of automatic MT evaluation.

Many authors analyze neural machine translation as a novel approach and try to implement it into already used frameworks. Banik et al. (Banik et al., 2019) analyzed the a statistical approach to combine the outputs of various machine translation systems. The authors have selected only the best phrases among the multiple systems' outputs and merged them into the final translation. The used NMT systems were Google Translate (Wu et al., 2016) and Bing Microsoft Translate (Dolan et al., 2002) and the authors used a Hierarchical system. The experiment was done using 8 different language pairs and the results were evaluated based on a fuzzy-based MT evaluation metric LeBleu (Virpioja and Grönroos, 2015). The results of the experiment showed that the outputs obtained by Google and Bing are very similar most of the times. The output from the SMT system may have different word orders or synonyms. The system combination model produces translations matching with those of Google and Bing Microsoft Translate. Bentivogli et al. (Bentivogli et al., 2018) compared the NMT with the phrase-based MT system on an English-German and English-French dataset. The analysis was done thoughtfully. Not only did the authors evaluate the translation quality using TER and HTER metrics but also based on morphological analysis. The morphological analysis consisted of identifying lexical errors, morphology errors and word order errors. The lexical analysis has shown various results. The NMT results for proper nouns were worse than for the phrase-based MT in the case of English-French language pair. On the other hand, the NMT showed better handling of complex sentences in the case of English-German language pair. The morphological error identification showed that NMT makes considerably less morphology errors in both language pairs. The word ordering errors analysis showed that this issue is language specific. The NMT has done well for the English-German language pair where it was successful at generating well-formed sentences. Also in the case of language pair with less complex reordering phenomena the NMT performed better than phrase-based MT. Bentivogli et al. (Bentivogli et al., 2018) showed that NMT is superior to phrase-based MT but identified also some shortcomings of NMT. NMT has issues with the translation of proper nouns and with the reordering of particular linguistic constituents. Xia (Xia, 2019) introduced a statistical machine translation system based on deep neural network. The focus of the article is oriented to the word alignment and pre-ordering in SMT. The word alignment model was created by a combination of multi-layer neural network and undirected probability graph model. The linearly ordered pre-ordering model was created using the multi-layer neural network to vocabulary the representation. Both of these models were combined in the same deep neural network framework named DNNAPM. The framework was tested on a sample of 100 000 sentence pairs. Accuracy was used as the evaluation metric for the field of word segmentation. Marzouk and Hansen-Schirra (Marzouk and Hansen-Schirra, 2019) analyzed the application of controlled language to improve the machine translation output. The authors compared the impact of nine controlled language rules to the quality of NMT output and compared the results for other MT systems: rule-based, statistical and hybrid MT. The experiment was done with the English-German language pair using texts of the technical

ISBN 978-80-7598-841-6 ISSN 2464-7470 (Print) ISSN 2464-7489 (On-line)

domain. The results of the experiment showed that NMT behaves differently when controlled language is applied. The quality of the NMT output is higher without the application of controlled language. This is in contradiction with other MT (rule-based, statistical, and hybrid) where the application of controlled language improves the output of the translation. The NMT system obtained the best results between all of the MT systems regardless of the application of the controlled language. The limitation of the experiment was in the use of only one language pair and the experiment should be done for other language pairs. The application of the controlled language did not bring any expected results. Based on the results of the experiment it seems that the application of controlled language could become obsolete for the novel MT system. Pinnis et al. (Pinnis et al., 2018b) presented an integration of NMT systems into document workflow translation of a cloud-based translation system and introduced examples of formatting-rich document translation. Pinnis et al. (Pinnis et al., 2018a) validated the NMT application to more difficult language pairs with less resources available for the NMT training. The authors compared the SMT and NMT systems for highly inflected languages (Estonian, Latvian and Russian). The authors also compared the results of the SMT and NMT systems output for a broad data domain and narrow data domain. The MT output was evaluated using automated (BLEU, NIST, and ChrF2) and manual methods (system comparative evaluation and error analysis of translations). The results of the evaluation showed that NMT system achieved better results for 83 % language pairs of broad domain. On the other hand, the narrow domain results showed that the SMT system produced significantly better translations than NMT system.

CONCLUSION

In this paper was described the novel approach of machine translation- neural machine translation system and the most used statistical machine translation system. Both of the machine translation systems were introduced and described in detail. The SMT as the most used system is getting replaced by the NMT as a novel approach. This paper presented the use of NMT by other researchers. Other authors start to use it and analyze its potential for the specific language pairs. The results of the described experiments show a potential improvement of the translation output of NMT in comparison with the SMT output. Despite that, there are some shortcomings of the novel NMT system where the SMT still offers better results. The future work would be focused on a detailed analysis of the Neural Machine Translation system output for a flective language such as the Slovak language. The research would compare the difference of output of SMT and NMT systems. Also, it would be interesting to compare Google Translate before it changed to NMT and observe whether the change improved the translation quality also for flective languages.

ACKNOWLEDGEMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-18-0473 and Scientific Grant Agency of the Ministry of Education of the Slovak Republic (ME SR) and of Slovak Academy of Sciences (SAS) under the contracts No. VEGA-1/0809/18.

REFERENCES

Bahdanau, D., Cho, K., Bengio, Y., 2016. Neural machine translation by jointly learning to align and translate. Proc. Int. Conf. Learn. Represent. 15.

Banik, D., Ekbal, A., Bhattacharyya, P., Bhattacharyya, S., Platos, J., 2019. Statistical-based system combination approach to gain advantages over different machine translation systems. Heliyon 5, e02504.

Bentivogli, L., Bisazza, A., Cettolo, M., Federico, M., 2018. Neural versus phrase-based MT quality: An in-depth analysis on English–German and English–French. Comput. Speech Lang. 49, 52–70.

Bessenyei, G., 2017. Neural Machine Translation: The Rising Star [WWW Document]. Memsource. URL https://www.memsource.com/blog/2017/09/19/neural-machine-translation-the-rising-star/?utm_source=mailchimp&utm_medium=email&utm_content=blog_article (accessed 10.4.17).

Cheng, Y., 2019. Neural Machine Translation. In: Joint Training for Neural Machine Translation. Springer, Singapore, pp. 1–10.

Chiang, D., 2007. Hierarchical phrase-based translation. Comput. Linguist. 33, 201–228.

Cho, K., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Doha, Qatar, pp. 1724–1734.

Dolan, W.B., Pinkham, J., Richardson, S.D., 2002. MSR-MT: The Microsoft research machine translation system. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 237–239.

Kalchbrenner, N., Blunsom, P., 2013. Recurrent Continous Translation Models. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Seattle, USA, pp. 1700–1709.

Koehn, P., 2010. Statistical Machine Translation. Cambridge University Press.

Lopez, A., 2008. Statistical machine translation. ACM Comput. Surv. 40, 1–49.

Marzouk, S., Hansen-Schirra, S., 2019. Evaluation of the impact of controlled language on neural machine translation compared to other MT architectures. Mach. Transl. 33, 179–203.

Munk, M., Munkova, D., 2018. Detecting errors in machine translation using residuals and metrics of automatic evaluation. J. Intell. Fuzzy Syst. 34, 3211–3223.

Munková, D., Kapusta, J., Drlík, M., 2016. System for Post-Editing and Automatic Error Classification of Machine Translation. In: DIVAI 2016 : 11th International Scientific Conference on Distance Learning in Applied Informatics, Sturovo, May 2-4, 2016. Wolters Kluwer, ISSN 2464-7489, Sturovo, pp. 571–579.

Munkova, D., Munk, M., 2015. Automatic Evaluation of Machine Translation Through the Residual Analysis. In: Huang, DS and Han, K (Ed.), ADVANCED INTELLIGENT COMPUTING THEORIES AND APPLICATIONS, ICIC 2015, PT III, Lecture Notes in Artificial Intelligence. pp. 481–490.

ISBN 978-80-7598-841-6 ISSN 2464-7470 (Print) ISSN 2464-7489 (On-line)

Munková, D., Munk, M., 2016. Evalvácia strojového prekladu. Univerzita Konštantína Filozofa v Nitre, Nitra.

Munková, D., Munk, M., Skalka, J., Kasaš, K., 2020. Automatic Evaluation of MT Output and Post-edited MT Output for Genealogically Related Languages. In: Innovation in Information Systems and Technologies to Support Learning Research. EMENA-ISTL 2019. Springer, Cham, pp. 416–425.

Munková, D., Munk, M., Vozár, M., 2013. Data Pre-processing Evaluation for Text Mining: Transaction/Sequence Model. Procedia Comput. Sci. 18, 1198–1207.

Munková, D., Munk, M., Vozár, M., 2014. Influence of stop-words removal on sequence patterns identification within comparable corpora, Advances in Intelligent Systems and Computing.

Munkova, D., Wrede, O., Absolon, J., 2019. Comparison of human, machine and Post-Editing Translation from Slovak into German by means of automatic Evaluation. ZEITSCHRIFT FUR SLAWISTIK 64, 231–261.

Pinnis, M., Krišlauks, R., Deksne, D., Miks, T., 2018a. Evaluation of neural machine translation for highly inflected and small languages. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 445–456.

Pinnis, M., Skadiņš, R., Šics, V., Miks, T., 2018b. Integration of neural machine translation systems for formatting-rich document translation. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 494–497.

Stencl, M., Stastny, J., 2010. Neural network learning algorithms comparison on numerical prediction of real data. In: 16th International Conference on Soft Computing MENDEL 2010, Brno. Brno, pp. 280–285.

Sutskever, I., Vinyals, O., Le, Q. V., 2014. Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. 3104–3112.

Virpioja, S., Grönroos, S.-A., 2015. LeBLEU: N-gram-based Translation Evaluation Score for Morphologically Complex Languages. In: Proceedings of the Tenth Workshop on Statistical Machine Translation. Association for Computational Linguistics (ACL), pp. 411–416.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., Dean, J., 2016. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.

Xia, Y., 2019. Research on statistical machine translation model based on deep neural network. Computing 1–19.

Yamamoto, H., Isogai, S., Sagisaka, Y., 2003. Multi-class composite N-gram language model. Speech Commun. 41, 369–379.

ISBN 978-80-7598-841-6 ISSN 2464-7470 (Print) ISSN 2464-7489 (On-line)