

A LITERATURE REVIEW OF ABSTRACTIVE SUMMARIZATION METHODS

D.V. SHYPIK, P.I. BIDYUK

Abstract. The paper contains a literature review for automatic abstractive text summarization. The classification of abstractive text summarization methods was considered. Since the emergence of text summarization in the 1950s, techniques for summaries generation were constantly improving, but because the abstractive summarization require extensive language processing, the greatest progress was achieved only recently. Due to the current fast pace of development of both Natural Language Processing in general and Text Summarization in particular, it is essential to analyze the progress in these areas. The paper aims to give a general perspective on both the state-of-the-art and older approaches, while explaining the methods and approaches. Additionally, evaluation results of the research papers are presented.

Keywords: natural language processing, text summarization, abstractive text summarization, sequence to sequence models.

INTRODUCTION

Today there is a problem of processing large amounts of text information caused by a constantly growing volume of textual information. It is possible that this issue can be addressed by using natural language processing approaches, in particular, text summarization.

The main goal of Text Summarization (TS) is to create a summary — “a reductive transformation of source text to summary text through content reduction by selection and/or generalization on what is important in the source.”

In the future, TS may be essential for users to efficiently manage information, allowing saving time and resources, as well as to quickly find the specific information they are looking for within documents.

TS has experienced great development in recent years, and a wide range of techniques and paradigms have been proposed to tackle this research field. However, to produce a summary automatically is very challenging [2].

Extraction methods reached some serious progress [3], but there is an empirical limit intrinsic to pure extraction, as compared to abstraction [4]. Also, Laura Hasler claims that the technique humans practice is to copy and paste the same material present in the source documents [5]. However, some slight changes are applied in most of the cases, and two types of operations, atomic and complex, are identified, involving deletion, insertion, replacement, reordering or merging (the first two are atomic operations while the last three are complex). From the coherency evaluation standpoint, the results showed that 78% of the abstracts were more coherent than extracts.

GRAPH-BASED APPROACHES

The studies for abstractive sentence summarization used to be largely based on sentence compression [6, 7] and sentence fusion [8, 9]. Graph-based approaches also were very popular among older abstractive approaches, in particular, they were shown to be very successful for producing multi-document summaries [10].

Ganesan, Zhai, and Han in their work [11] proposed Opinosis – graph-based summarization framework, which generates abstractive summaries of opinions. The system considers a high redundancy of opinions. Opinosis employs shallow NLP. Firstly, the input text is represented as a textual directed graph. The work introduces lexical links usage in graph building, which should help with discovering new sentences or reinforcing existing ones. Then, candidate abstractive summaries are generated by choosing various sub-paths in the graph. They are analyzed and scored by using three unique properties of graphs (redundancy capture, gapped subsequence capture, collapsible structures), duplicated or extremely similar paths are excluded by using similarity measure. Authors created a dataset, consisting of reviews of various products. The evaluation of the created dataset shows that the summaries generated by the system have a higher correlation with human-made summaries than baseline extractive method. Evaluation conducted on dataset created by authors, resulting in recall for ROUGE-1 – 28,3; ROUGE-2 – 8,53; ROUGE-SU4 – 8,51; F-score for ROUGE-1 – 32,7; ROUGE-2 – 9,98; ROUGE-SU4 – 10,27. Unfortunately, as with all custom datasets, the results are not directly comparable to other algorithms. The authors also point out that this solution is more extractive than abstractive, as only the words from the original text can occur in the summary. It is abstractive in the sense that generated sentences are in general not from the original sentences set.

Lloret and Palomar [12] proposed to combine Graph-based abstractive approach with extractive approach (COMPENDIUM) in several ways. The graph-based approach is used to create new sentences by computing the shortest ‘valid’ path on word graph. The validity of resulting sentences is checked using several heuristics (sentence should have more than 3 words, one of the words is a verb, the sentence doesn’t end in an article). The performance is tested on the DUC 2002 test set, resulting in F-Measure ROUGE-1 – 20,85; ROUGE-2 – 6,68; ROUGE-SU4 – 7,04 for the best of proposed models. Authors state that even though the results are not very high, the approach is promising for future research.

Banerjee and Sugiyama [13] described new multi-document abstractive summarizer. Authors state that documents are not equal by information quantity about the topic, so initially the system evaluates the most important document from the set by using Lex Rank, Pairwise Cosine Similarity, and Overall Document Collection Similarity. Then each sentence from the selected document is used to generate separate clusters or appended to existing. Each cluster consists of a word-graph structure. Clusters are further constructed by including sentences from other documents that have high similarity with them. For each one, K-shortest paths are selected, which are then used to construct the sentences by using a proposed integer linear programming problem that maximizes information content and linguistic quality and reduces redundancy in the final summary. The system outperforms the best extractive summarizer by ROUGE scores on both DUC 2004 and DUC 2005 datasets. For DUC-2004 best proposed system reached

recall for ROUGE-2 – 11,99; ROUGE-SU4 – 14,76. Other metrics were used for DUC-2005 – recall for ROUGE-L – 35,77; ROUGE-SU4 – 12,41. Also, the work included manual evaluation of Informativeness and Language Quality by 10 evaluators – the proposed system reached 4,1 / 5 in Informativeness, 3,63 / 5 in Language Quality on randomized set of summaries.

HEURISTIC BASED METHODS

Some researchers use heuristic methods for sentence generation, for example, the article of **Ganest and Lapalme** [14]. The work introduces the concept of Information Items (INIT). INIT is “the smallest element of coherent information in the text or a sentence”. The framework consists of the following steps:

1) INIT retrieval. Proposed: As the definition of INIT is intentionally vague and leaves out implementation details, authors proposed two candidates for this step: Semantic Role Labeling (SRL) and predicate-logic analysis. Implemented: extraction of subject–verb–object triples, with tags for date and location;

2) INIT selection. Proposed and implemented: frequency-based models (as in extractive summarization);

3) Sentence generation. Proposed: text generation patterns or heuristic rules. Implemented: heuristic engine (implementing linguistic rules) for sentence generation (SimpleNLG) is used.

Evaluation results on the TAC 2010 dataset are – Pyramid Score – 0,315; Linguistic Quality – 2,17 and overall responsiveness – 2,30.

It is also interesting to note, that in fact, part of the implementation of Khan, Salim, and Kumar [15] were close to some of the proposed ideas of Ganest and Lapalme [14] (former proposed SRL usage, sentence generation through using heuristics on the last step).

The proposed framework of **Khan, Salim, and Kumar** [15] uses heuristics for sentence generation, genetic algorithm for information selection.

Novelty:

1) The first work to implement semantic role labeling (SRL) in multi-document abstractive summarization. Rather than simply selecting sentences from source documents, semantic representation is used to represent source documents;

2) Proposed clustering of the semantically similar PAses (Predicate Argument Structures) by utilizing semantic similarity measure;

3) Ranking PAS based on the features weighted and optimized by genetic algorithm; since text features are sensitive to the quality of the generated summary.

Algorithm:

1) the document is split into sentences;

2) PAses are selected from each sentence in the document collection using semantic role labeler (SENNA);

3) semantic similarity matrix of PAS is computed;

4) semantically similar PAses are clustered using ‘Agglomerative hierarchical clustering’ (HAC) algorithm based on the average linkage method;

5) the PASs in each cluster are scored based on features, weighted and optimized by genetic algorithm. Highest ranked PAses are selected from each cluster;

6) heuristic engine for generation sentences in English (SimpleNLG) is used to generate sentences from argument structures.

Content selection is conducted by ranking PASEs based on optimized features (step 4, 5). In step 6 language generation is used to generate sentences from PASEs.

On DUC 2002 dataset evaluation of Pyramid Score (0,5) and Average Precision (0,7) showed that authors' approach outperforms comparison models. Human-written summaries reached 0,69 for Pyramid Score, 0,85 for Average Precision.

HYBRID METHODS

TOPIARY, the system described by Zajic and Dorr [16], combines sentence compression algorithm by means of linguistically-motivated heuristics (modified Hedge Trimmer) and Unsupervised Topic Discovery (UTD) – statistical method, which generates a set of topics from document corpus. Hedge Trimmer algorithm is modified in order to take a list of topics with relevance scores as additional input, dynamically change compression rate in order to include highest scoring topic if it is missed. The algorithm won a prize on Document Understanding Conference 2004 Workshop as best performing by ROUGE-1, ROUGE-2 and ROUGE-L measures, scoring 24,9; 6,45; 19,95 respectively.

Compress (Clarke and Lapata, 2008) [17] – uses integer linear programming (ILP) to infer globally optimal compressions in the presence of linguistically motivated constraints. The authors introduced the usage of global constraints, designed their system to use less local syntactic knowledge. Three models are presented and compared – unsupervised, semi-supervised, and fully supervised approaches. The results have shown that semi-supervised model with the proposed constraints performing the best in terms of human-evaluated Grammaticality and Importance (information content of summary). Evaluation conducted on data-set created by author. The best scores are 3,76 for Grammaticality, 3,53 for Importance.

Woodsend and others proposed a novel model [18], which consists of three components. Content selection was performed by an SVM, which gave a salience score for each phrase. To generate compressions and paraphrases the model used Quasi-Synchronous Grammar (QG) rules. The third component, an Integer Linear Programming ILP model combined the output of previously mentioned components into an output summary by optimizing content selection and surface realization preferences jointly. Similar to Clarke and Lapata, the ILP model includes global constraints relating to sentence length, overall summary length, grammaticality and topics inclusion. The work is evaluated by humans on DUC-2004 headlines dataset, by means of Grammaticality and Importance, reaching scores 5,36 and 4,94 respectively. In comparison conducted by authors, proposed model outperformed TOPIARY by 2,33 points in Grammaticality, 1,49 points in Importance.

Bing and others [19] proposed an abstraction-based multi-document summarization system that creates new sentences by employing the proposed integer linear optimization model. The system operates syntactic units like a noun or verb phrases instead of whole sentences. In the first step, a noun or verb phrases are

extracted from the documents via constituency trees. Then, for each phrase, a saliency score is calculated using concept-based weight incorporating position information. New sentences are created by selecting and merging phrases, insuring their validity by solving a proposed linear optimization model. Performance evaluations are carried out on TAC 2011 dataset. The system reached recall for ROUGE-2 – 11,7; ROUGE-SU4 – 14,7; F-score for ROUGE-2 – 11,7; ROUGE-SU4 – 14,8. The average Linguistic Quality assessed for resulting summaries was 3,43.

METHODS BASED ON SEQUENCE TO SEQUENCE MODELS

The task of abstractive sentence summarization i.e. generating a shorter sentence version while trying to preserve its original meaning is increasingly intriguing for researchers, especially with the development of sequence-to-sequence framework [20].

Creation of Encoder-Decoder Recurrent Neural Networks and their utilization for encoding variable sentence length with further decoding into variable sentence length [21, 22] lead to significant progress in machine translation. Together with previously proposed Bidirectional Neural [23], it led to the invention of Attentional Encoder-Decoder RNNs [24] and their usage in the field of abstractive sentence summarization [25]. Also, creation of Gigaword dataset [26] had a huge role in providing the previously inaccessible amount of training data.

Rush, Chopra, and Weston [25] proposed a fully data-driven approach to abstractive sentence summarization (which authors call ABS – Attention-Based Summarization). The method utilizes a local attention-based model that generates each word of the summary conditioned on the input sentence. The work tries several encoder architectures:

1. Bag-of-words encoder ignores properties of original order and relationships between neighboring words. This model can capture the importance of words, potentially can learn to combine words, but is inherently limited in representing continuous phrases.

2. The convolutional encoder allows capturing local interactions between words. Standard TDNN is used. Minuses: it produces a single representation (vector) of the entire sentence, ignoring length and other differences.

3. Attention-based encoder based on the article of Bahdanau, Cho, and Bengio [24], removes the need for single representation. The author proposes to think of this as of “soft alignment” between input and output summary.

Generation model uses a feed-forward neural network – Neural Network Language Model (NNLM). The encoder and the generation model are trained jointly on the sentence summarization task. Expectedly, the Attention-based encoder significantly outperforms other proposed encoders and baseline.

Several other articles significantly influenced the further route of research:

1. Gated Recurring Unit (GRU) was proposed [27], as an alternative to Long Short-Term Memory (LSTM) gated unit in RNN. The article of Chung and others [28] has shown that proposed units are comparable or better than LSTM in the task of sequence modeling. Also, the additive nature of both models has advantages over classical tanh units:

- the unit can remember existence of a specific feature in the input stream for a long sequence of steps. Forget gate of the LSTM unit or the update gate of the GRU should specifically decide when the feature is forgotten;
- the memory of these units effectively creates shortcut paths that bypass multiple temporal steps. This property allows to back-propagate the error easily, as it vanishes longer.

2. Large Vocabulary Trick (LVT) was proposed [29] for machine translation. The approach is an approximate training algorithm based on (biased) importance sampling that allows training neural models with a much larger target vocabulary. The algorithm effectively keeps the computational complexity during training at the level of using only a small subset of the full vocabulary. Authors claim that the proposed approach allows us to efficiently use a fast computing device with limited memory, such as a GPU, to train a neural machine translation model with a much larger target vocabulary.

In “**sequence-to-sequence RNNs for text summarization**” Nallapati, Xiang and Zhou [30] demonstrated that the sequence-to-sequence models are extremely promising for summarization. Full seq2seq model is used, as previously proposed [24], with encoder Bidirectional GRU-RNN as and Unidirectional Attentional GRU-RNN as the decoder. Their experiments showed that LVT-trick significantly improves training speed without sacrificing performance. Also, more classic features, like PoS tags, named-entity tags, TF-, IDF that are encoded together with words provided an additional performance improvement. Results of this works have shown that the proposed model outperforms the previous state-of-art [25] on Gigaword, DUC 2003, DUC 2004 corpora.

During further research Nallapati and others in their work “**Abstractive text summarization using sequence-to-sequence RNNs and beyond**” [31] proposed Switching Pointer-Generator to avoid the generation of “UNK” token (a token that is generated by most summarization systems, which try to generate a word that is out of their training dictionary). Authors show, that even though the model learns to use pointers very accurately not only for named entities but also for multi-word phrases, the performance improvement of the overall model is not significant. It is proposed, that model impact may be clearer in other document sets, where the tail distribution of rare words is heavier. Hierarchical attention model is proposed (attention not only on the word level, but also on sentence level), but it didn't show a significant difference on all datasets. Also the work introduced a new large scale dataset “CNN/Daily Mail”, which is very important due to the lack of the former.

The next year, same authors published **new work** [32] which proposed new extractive approach and compared the results with their previous model on the CNN/Daily Mail dataset using ROUGE metric. The comparison showed that extractive model *significantly outperformed* abstractive on the longer texts.

See, Liu, and Manning pointed [33], that even though the system proposed in “Abstractive text summarization using sequence-to-sequence RNNs and beyond” [31] reached a new level of accuracy, it still has typical drawbacks of sequence to sequence models – it inaccurately reproduces factual details, has high level of repeating themselves and not always deals with out-of-vocabulary (OOV) words. The work tries to address this issue through a proposed novel variant of coverage vector and modified pointer-generator network. In contrast to the

abovementioned work where pointer-generator network is activated only for OOV words or named entities, the model is allowed to learn when to apply the network. The proposed model is evaluated to have higher both ROUGE and METEOR scores than the predecessor, was shown to reduce inaccuracies and repetition.

In “**Selective encoding for abstractive sentence summarization**” [34] it is pointed, that unlike machine translation, where alignment between all parts of input and output is required, there is no explicit alignment between input sentences in sentence summarization. The challenge is not to infer the alignment, but to select the highlights and filter secondary information. To solve this task, it is proposed to extend seq2seq framework with additional selective encoding model. The result consists of sentence encoder (bidirectional GRU), and attention equipped decoder (attentional GRU). The selective gate network constructs a second level sentence representation by controlling the information flow from the encoder to the decoder. Also, the proposed layer was shown to perform as expected – it has highlighted the representation of important words from the input sentence. The model is evaluated on the Gigaword, DUC 2004 and MSR abstractive sentence summarization datasets. The proposed selective encoding model outperformed the state-of-art baseline models.

As it was shown, seq2seq framework performance quickly deteriorates with a length of the generated sequence [27]. So, in “Retrieve, rerank and rewrite: Soft template based neural summarization.” authors point that similar sentences should hold similar summary patterns, so existing summaries can be used as “soft templates” to guide the seq2seq model [20]. As the name of the article hints, the model has three steps:

1. Retrieve step: Popular IR platform (Lucene) is used to retrieve candidate templates. The system finds out analogies of the given sentence in the corpus and picks their summaries as candidate templates. Recurrent Neural Network (RNN) encoder is applied to convert the input sentence and each candidate template into hidden states, encoder output is shared by “rerank” and “rewrite” steps.

2. Rerank: In Retrieve, the template candidates are ranked according to the text similarity between the corresponding indexed sentences and the input sentence. However, for the summarization task, the soft template is expected to resemble the actual summary as much as possible. So, this step tries to choose the “closest” template to the processed sentence.

3. Rewrite: summary generation. The summary is generated according to the hidden states of both the sentence and template. Concatenation function is applied to combine the hidden states of the sentence and template.

By comparison to other strategies of choosing soft templates, the authors show that proposed “Rerank” step has a room for improvement. Also, the quality of the summaries depends on the quality of the imported external summaries, which shows that soft templates themselves, not other architecture changes have great importance. In total, the proposed model significantly outperforms the state-of-art seq2seq models, and even soft templates themselves demonstrate high competitiveness.

Inspired by results of “Retrieve, rerank and rewrite” [20] with soft templates **Wang and others** [35] proposed a new model called BiSET (Bi-directional Selective Encoding with Template for Abstractive Summarization) to enhance soft template usage in text summarization. The work introduces:

1. A novel bi-directional selective mechanism with two gates to mutually select important information from both article and template to assist with summary generation. The mechanism is inspired by the research of “bidirectional attention flow mechanism” in machine reading comprehension [36] and the selective mechanism described earlier [34].

2. Fast Rerank method to automatically select high-quality templates from the training corpus. The method is based on Convolutional Encoder, Similarity Matrix, and Pooling layer.

F-1 scores for ROUGE-1, ROUGE-2, ROUGE-L were used as metrics. Removal of template-to-article attention or article-to-template attention lead to reduction in all used metrics, showing that every part of proposed bi-directional selective mechanism is improving performance. Comparison with simple ‘Concatenation’ approach to soft templates showed that proposed selective mechanism significantly outperformed in all metrics.

Results of the works, presented in this section are shown in Tables 1–4.

Table 1. Comparison of methods’ performance on Gigaword dataset, full-length ROUGE F1

Model	ROUGE-1	ROUGE-2	ROUGE-L
ABS (Rush 2015) [25]	29,55	11,32	26,42
ABS+ (Rush 2015) [25]	29,78	11,89	26,97
CA2s2 (Chopra 2016)	33,78	15,97	31,15
FeatSeq2Seq (Nallapati 2016) [30]	32,67	15,59	30,64
SEASS (Zhou 2017) [34]	36,15	17,54	33,63
R ³ Sum (Cao 2018) [20]	37,04	19,03	34,46
BiSET (Wang 2019) [35]	39,11	19,78	36,87

Table 2. Comparison of methods’ performance on Gigaword dataset, ROUGE Recall (output capped at 75 bytes)

Model	ROUGE-1	ROUGE-2	ROUGE-L
Compress (Clarke and Lapata, 2008) * [17]	19,63	5,13	18,28
ABS (Rush 2015) [25]	30,88	12,22	27,77
ABS+ (Rush 2015) [25]	31,00	12,65	28,34

* provided in the work of Rush and others [25], not in original work

Table 3. Comparison of methods’ performance on DUC-2004 dataset, ROUGE Recall (output capped at 75 bytes)

Model	ROUGE-1	ROUGE-2	ROUGE-L
Compress (Clarke and Lapata) * [17]	19,77	4,02	17,3
W&L (Woodsend) * [18]	22	6	17
TOPIARY (Zajic) * [19]	25,12	6,46	20,12
ABS (Rush) [25]	26,55	7,06	22,05
ABS+ (Rush) [25]	28,18	8,49	23,81
FeatSeq2Seq (Nallapati 2016) [30]	28,35	9,46	24,59
SEASS (Zhou 2017) [34]	29,21	9,56	25,51

* provided in the work of Rush and others [25], not in original work.

Table 4. Comparison of methods' performance on CNN/Daily Mail dataset, ROUGE Recall (output capped at 75 bytes)

Model	ROUGE-1	ROUGE-2	ROUGE-L
Nallapati 2016 [30]	35,46	13,30	32,65
Nallapati 2017 (extractive) [32]	39,6	16,2	35,3
Nallapati 2017 (abstractive) [32]	37,5	14,5	33,4
See 2017 ** [33]	39,53	17,28	36,38

** uses non-anonymized dataset, so not directly comparable.

CONCLUSIONS

This paper introduces some important information concerning both state-of-art and older approaches of abstractive text summarization. This review could serve as a starting point for novice researchers to get familiar with the field. Modern data-driven methods had more focus, as they are mostly left out from the topic literature reviews, and also as they tend to give better results and have pretty substantial differences from the older approaches, that were more structure-based or linguistic-focused. However, in contrary to more classical models, sequence to sequence-based ones tend to “lose control” on the long text samples, and also they require much bigger datasets. As we can see, Gigaword dataset is dominating in training new models, as older ‘DUC’ and ‘TAC’ datasets do not provide the desired sample quantity.

To summarize, further research seems to be aligned with improving the quality of the sequence to sequence models on the long texts or proposing an alternative machine learning method.

REFERENCES

1. Jones S.K. Automatic summarizing: factors and directions [Электронный ресурс] / S.K. Jones // MIT Press. — 1999. — Available at: <https://www.cl.cam.ac.uk/archive/ksj21/ksjdigipapers/summbok99.pdf>.
2. Multi-document summarization by sentence extraction [Электронный ресурс] / J.S. Goldstein, V. Mittal, J.G. Carbonell, M. Kantrowitz. — 2000. — Available at: http://scholar.google.com.ua/scholar_url?url=https://kilthub.cmu.edu/articles/Multi-Document_Summarization_By_Sentence_Extraction/6624470/files/12121496.pdf&hl=uk&sa=X&scisig=AAGBfm3dUni3D9yq1qbG7bN3z4ow9ChpyA&nossl=1&oi=scholar.
3. Lloret E. Text summarisation in progress: a literature review / E. Lloret, M. Palomar // Artificial Intelligence Review. — 2011. — N 37. — P. 1–41.
4. Genest P. HEXTAC: the Creation of a Manual Extractive Run [Электронный ресурс] / P. Genest, G. Lapalme, M. Yousfi-Monod. — 2009. — Available at: <http://www.mymcorner.net/files/Genest-Lapalme-Yousfi-Monod-09.pdf>.
5. Hasler L. From extracts to abstracts: human summary production operations for computer-aided summarisation [Электронный ресурс] / L. Hasler. — 2007. — Available at: <http://rgcl.wlv.ac.uk/events/CALP07/papers/10.pdf>.
6. Erkan G. Lexrank: graph-based lexical centrality as salience in text summarization / G. Erkan, D. Radev // Journal of Artificial Intelligence Research. — 2004. — Vol. 22. — P. 457–479.
7. A perspective-based approach for solving textual entailment recognition [Электронный ресурс] / O.Ferrandez, D. Micol, R. Munoz, M. Palomar. —

2007. — Available at: https://www.researchgate.net/profile/Joao_Cordeiro11/publication/234803426_Biology_Based_Alignments_of_Paraphrases_for_Sentence_Compression/links/560fa81b08ae0fc513ef311e/Biology-Based-Alignments-of-Paraphrases-for-Sentence-Compression.pdf#page=80.
8. *Filippova K.* Multi-Sentence Compression: Finding Shortest Paths in Word Graphs [Електронний ресурс] / К. Filippova. — 2010. — Available at: <https://www.aclweb.org/anthology/C10-1037.pdf>.
 9. *Gardner J.* An integrated framework for de-identifying unstructured medical data [Електронний ресурс] / J. Gardner, L. Xiong // Elsevier. — 2009. — Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.157.7185&rep=rep1&type=pdf>.
 10. *Hliaoutakis A.* The AMTE_x approach in the medical document indexing and retrieval application / A. Hliaoutakis, K. Zervanou, E. Petrakis // *Data & Knowledge Engineering*. — 2009. — N 68. — P. 380–392.
 11. *Ganesan K.* Opinosis: A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions [Електронний ресурс] / К. Ganesan, C. Zhai, H. Jiawei. — 2010. — Available at: <https://www.aclweb.org/anthology/C10-1039.pdf>.
 12. *Lloret E.* Analyzing the Use of Word Graphs for Abstractive Text Summarization [Електронний ресурс] / E. Lloret, M. Palomar. — 2011. — Available at: <https://pdfs.semanticscholar.org/7d25/67d7eefc772865992e93996c1cd7f6ba6319.pdf>.
 13. *Banerjee S.* Multi-document abstractive summarization using ILP based multisentence compression [Електронний ресурс] / S. Banerjee, P. Mitra, K. Sugiyama. — 2015. — Available at: <https://www.ijcai.org/Proceedings/15/Papers/174.pdf>.
 14. *Ganest P.* Framework for Abstractive Summarization — Available at: <https://pdfs.semanticscholar.org/fdf9/e7d06bf21093e29923742d2040b0e495bc1d.pdf>.
 15. *Khan A.* A framework for multi-document abstractive summarization based on semantic role labelling / A. Khan, N. Salim, Jaya Kumar Y. // *Appl Soft Comput* 30:737–747. — 2015. — doi:10.1016/j.asoc.2015.01.070
 16. *Zajic D.* BBN/UMD at DUC-2004: Topiary [Електронний ресурс] / D. Zajic, B. Dorr, R. Schwartz. — 2004. — Available at: <http://users.umiacs.umd.edu/~bonnie/Publications/Attic/DUC2004-HEADLINE.pdf>.
 17. *Clarke J.* Global inference for sentence compression an integer linear programming approach / J. Clarke, M. Lapata // *Journal of Artificial Intelligence Research*. — 2008. — N 31. — P. 399–429.
 18. *Woodsend K.* Title Generation with Quasi-Synchronous Grammar [Електронний ресурс] / K. Woodsend, Y. Feng, M. Lapata. — 2010. — Available at: https://www.research.ed.ac.uk/portal/files/23634327/2010_Woodsend_Feng_ET_AL_Title_Generation_with_Quasi_Synchronous_Grammar.pdf.
 19. *Abstractive Multi-Document Summarization via Phrase Selection and Merging* [Електронний ресурс] / L. Bing, P. Li, Y. Liao et al. — 2015. — Available at: <https://www.cs.cmu.edu/~lbing/pub/acl2015-bing.pdf>.
 20. *Retrieve, Rerank and Rewrite: Soft Template Based Neural Summarization* [Електронний ресурс] / Z. Cao, W. Li, F. Wei, S. Li. — 2018. — Available at: <https://pdfs.semanticscholar.org/c93b/8518204ef722f4c749628023c6d5d061a5fa.pdf>.
 21. *Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation* [Електронний ресурс] / К. Cho, В. Merriënboer, С. Gulcehre et al. — 2014. — Available at: <https://www.aclweb.org/anthology/D14-1179.pdf>.
 22. *Sutskever I.* Sequence to Sequence Learning with Neural Networks [Електронний ресурс] / I. Sutskever, O. Vinyals, Q.V. Le. — 2014. — Available at:

- <https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf>.
23. *Schuster M.* Bidirectional recurrent neural networks / M. Schuster, K. Paliwal // IEEE Transactions on Signal Processing. — 1997. — N 45. — P. 2673–2681.
 24. *Bahdanau D.* Neural machine translation by jointly learning to align and translate [Электронный ресурс] / D. Bahdanau, K. Cho, Y. Bengio. — 2015. — Available at: <https://arxiv.org/pdf/1409.0473.pdf>.
 25. *Rush A.M.* A Neural Attention Model for Sentence Summarization [Электронный ресурс] / A.M. Rush, S. Chopra, J. Weston. — 2015. — Available at: <https://www.aclweb.org/anthology/D15-1044.pdf>.
 26. *Napoles C.* Annotated Gigaword [Электронный ресурс] / C. Napoles, M. Gormley, B. Van Durme. — 2012. — Available at: <https://www.cs.cmu.edu/~mgormley/papers/napoles+gormley+van-durme.naaclw.2012.pdf>.
 27. *On the Properties of Neural Machine Translation: Encoder–Decoder Approaches* [Электронный ресурс] / K.Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio. — 2014. — Available at: <https://arxiv.org/pdf/1409.1259.pdf>.
 28. *Empirical evaluation of gated recurrent neural networks on sequence modeling* [Электронный ресурс] / J.Chung, Ç. Gülçehre, K. Hyun Cho, Y. Bengio. — 2014. — Available at: <https://arxiv.org/pdf/1412.3555.pdf>.
 29. *On Using Very Large Target Vocabulary for Neural Machine Translation* [Электронный ресурс] / S. Jean, K. Cho, R. Memisevic, Y. Bengio. — 2014. — Available at: <https://www.aclweb.org/anthology/P15-1001.pdf>.
 30. *Nallapati R.* Sequence-to-sequence RNNs for text summarization [Электронный ресурс] / R. Nallapati, B. Xiang, B. Zhou. — 2016. — Available at: <https://pdfs.semanticscholar.org/033b/c4febf590f6e011e9b0f497cadfe6a4c292d.pdf>.
 31. *Abstractive text summarization using sequence-to-sequence RNNs and beyond* [Электронный ресурс] / R. Nallapati, B. Zhou, C. Santos et al. — 2016. — Available at: <https://arxiv.org/pdf/1602.06023.pdf>.
 32. *Nallapati R.* SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents [Электронный ресурс] / R. Nallapati, F. Zhai, B. Zhou. — 2017. — Available at: <https://arxiv.org/pdf/1611.04230>.
 33. *See A.* Get To The Point: Summarization with Pointer-Generator Networks [Электронный ресурс] / A. See, P.J. Liu, C.D. Manning. — 2017. — Available at: <https://arxiv.org/pdf/1704.04368>.
 34. *Selective Encoding for Abstractive Sentence Summarization* [Электронный ресурс] / Q. Zhou, N. Yang, F. Wei, M. Zhou. — 2017. — Available at: <https://www.aclweb.org/anthology/P17-1101.pdf>.
 35. *Wang K.* BiSET: Bi-directional Selective Encoding with Template for Abstractive Summarization [Электронный ресурс] / K. Wang, X. Quan, R. Wang. — 2017. — Available at: <https://www.aclweb.org/anthology/P19-1207.pdf>.
 36. *Bidirectional attention flow for machine comprehension* [Электронный ресурс] / M. Seo, A. Kembhavi, A. Farhadi, H. Hajishirzi. — 2017. — Available at: <https://arxiv.org/pdf/1611.01603.pdf>.

Received 25.09.2019

From the Editorial Board: the article corresponds completely to submitted manuscript.