Estimation of the Local and Global Coherence of Ukrainian Texts Using Transformer-Based, LSTM, and Graph Neural Networks

Artem Kramov, Sergiy Pogorilyy

Computer Engineering Department, Taras Shevchenko National University of Kyiv, 60 Volodymyrska Street, 01033 Kyiv, Ukraine

Abstract

In this paper, the different models for the estimation of both local and global coherence of Ukrainian-language texts have been considered. In order to evaluate the local coherence of a document, Transformer-based and LSTM neural networks have been proposed with further training on a Ukrainian-language news corpus. It has been shown that the LSTM-based approach outperforms the corresponding network based on the Transformer architecture according to the accuracy metrics while solving typical tasks on both test datasets. In order to investigate the connection between sentences revealed by the neural network, the Uniform Manifold Approximation and Projection dimension reduction technique has been utilized for the projection of sentences' embedding into 2D space. The clusters obtained may indicate the consideration of both the structure of a sentence and different types of connections between them by the designed model. In order to estimate the global coherence of a document, a model based on a graph convolutional neural network has been suggested. The appropriateness of taking into account the connection between all sentences despite their positions has been shown. The results obtained for the designed and trained global coherence estimation model may indicate the different aspects of the analysis of a text by the designed models that can lead to the usage of both local and global coherence estimation models according to an assigned task.

Keywords

local and global coherence of a document, Transformer-based neural network, sentence embedding, graph convolutional network, Ukrainian corpora

1. Introduction

A text's coherence implies the thematic integrity of its components. It may also be described as "the extent to which text makes sense by introducing, explaining and linking its concepts and ideas through a sequence of semantically and logically related units of discourse" [1]. Thus, the coherence of a text simplifies the perception of the thoughts of an author by a reader due to the availability of logical connections between the different parts of a document. Such a feature of a text can be achieved by the semantic consistency of lexical units, repeats, synonyms, antonyms, different phrases, etc. The availability of these text parameters is crucial while writing instructions, learning materials, articles, guides. Moreover, the lack of such thematic

ICTERI-2021, Vol II: Workshops, September 28 – October 2, 2021, Kherson, Ukraine
artemkramovphd@knu.ua (A. Kramov); sdp@univ.net.ua (S. Pogorilyy)
0000-0003-3631-1268 (A. Kramov); 0000-0002-6497-5056 (S. Pogorilyy)

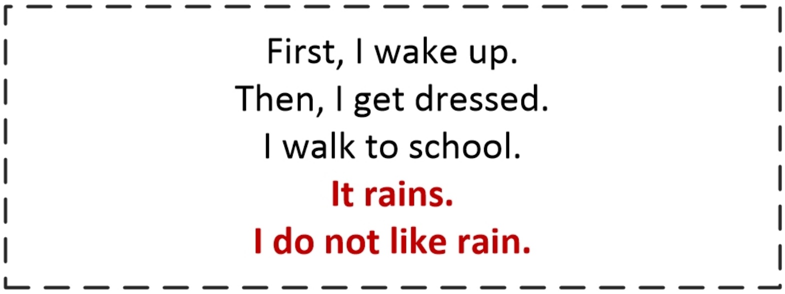

Figure 1: Example of the incoherent part of a text: corresponding sentences are highlighted in red

integrity within a sentence may indicate the disability to concentrate a person's thoughts around a certain topic that causes the usage of coherence estimation models for the prediction of schizophrenia symptoms [2].

Fig. 1 shows the simplified example of the incoherent part of a text (marked in red). As can be seen from this example, the common topic of the excerpt is devoted to some schedule. However, the semantic meaning of the last two sentences deviates from the general idea of the story that leads to the lack of a logical and semantic connection between sentences during the reading of the text. It makes the text less coherent that complicates its perception and understanding. Obviously, the perception of a text depends on the background knowledge of a reader and his own preferences; thus, the general mark of the coherence of a text may be considered as subjective judgment. However, the evaluation of a corresponding metric with further analysis of the relations between text's components may help to reveal the incoherent elements in order to either improve its thematic integrity or classify input documents according to an assigned task.

It should be mentioned that the coherence of a text can be considered at a local and global level [3]. Local coherence implies the availability of semantic similarity between neighbour discourses (phrases, sentences); in contrast, global coherence can be achieved by the logical connections between remote text spans within a document. These features should be both taken into account while estimating the coherence of a document in order to reveal different relations between discourses. Due to the complexity of the analysis of these connections by means of traditional algorithms, different machine learning methods, namely, deep learning models, are used in order to estimate both local and global coherence of an input document. In spite of the availability of state-of-the-art works devoted to the estimation of the coherence of English texts, the corresponding analysis of Ukrainian corpora is still at the initial stage. The purpose of the paper is the following:

- analysis of the state-of-the-art coherence evaluation methods for English and Ukrainian texts;
- designing and experimental verification of the effectiveness of Transformer-based and LSTM neural networks for the estimation of the local coherence of Ukrainian texts;
- investigation of the connection between sentences during the local coherence evaluation;
- designing and experimental verification of the effectiveness of a graph neural network in order to estimate the global coherence of Ukrainian documents.

2. Related work

Taking into account the ambiguity and complexity of the estimation of different connections between text spans during the coherence evaluation process, corresponding state-of-the-art methods are based on the usage of neural network models. In the papers [4, 5] recurrent, recursive, and convolutional neural networks were designed in order to estimate the local coherence of an input document. Corresponding layers of networks are used for the vector representation of the groups of sentences with further binary classification (whether this text span is coherent or not) of each clique. The general coherence estimation value is calculated based on the coherence probability of all local sentence groups. In order to take into account the semantic consistency of all sentences simultaneously, the usage of a separate LSTM layer was suggested in the paper [6] for the additional consideration of an entire text while evaluating its coherence. The idea of taking into account both local and global consistency of sentences for the estimation of an output coherence value was extended in the papers [7, 8]. It was suggested to apply convolutional layers for either output or hidden states of LSTM layers in order to reveal the main components of sentences that represent the topic of a whole document. The results mentioned in the paper [8] showed that the corresponding model outperforms other methods while solving a sentence ordering task (verification if the coherence of the original version of a text is greater than the corresponding value of a document that is formed via permutation of its sentences). However, the appropriateness of the usage of this metric for the estimation of the effectiveness of coherence models was discussed in the paper [9]. It was shown that state-of-the-art models outperform each other while solving other considered tasks depending on the parameters of an input corpus (topic, sentence length, style of texts, etc.). Thus, the question of the comparison of coherence models due to their effectiveness remains open and requires further investigation.

In our previous works, we have performed the experimental verification of the models based on different neural networks [10, 11] for the Ukrainian corpora in order to estimate the local coherence of a text. According to the results obtained, a Transformer-based model outperformed other approaches while solving a typical task. However, taking into account the presence of a new context-based semantic embedding ELMo model trained on Ukrainian corpora, it is suggested to perform the additional verification of the effectiveness of considered networks for the estimation of local coherence with the further analysis of sentence representation produced by a model. Such an analysis may help to reveal the types of connections between neighbour text spans during the evaluation of the coherence of a document. Moreover, a graph neural network is proposed for the estimation of the global coherence of a text basing on the analysis of the connection between remote text spans.

3. Local coherence estimation

Firstly, let us define an input text T as the set of sentences $T = \{s_1, s_2, ..., s_N\}$. In order to estimate the coherence of a document according to the consistency of local spans, the text representation is transformed into the set of cliques - the ordered groups of the sentences of the fixed length $L = 3$:

$$T = \{\langle s_1, s_2, s_3\rangle, \langle s_2, s_3, s_4\rangle, ..., \langle s_{N-2}, s_{N-1}, s_N\rangle\} \quad (1)$$

Such a length was chosen due to the experimental selection of this parameter depending on the accuracy of a model [5]. The mentioned cliques are created by the iterative passing of an L-sized filter with an ordinary step across the input text. Similar to other local coherence estimation models, the output value is calculated as a product of the coherence probabilities of all cliques:

$$Coh(T) = \prod_{c \in T} p(y_c = 1), \quad (2)$$

where $p(y_c = 1)$ is the probability of the coherence of a clique $c \in T$. According to this equation 2, let us formulate a task for a model: it should be able to predict whether an input clique c is coherent or not. Thus, the model should be considered as a binary classifier; there are three inputs of the model (three sentences of a clique), an output value interprets the probability of the coherence of the input group.

3.1. Designing neural network models

Fig. 2 demonstrates the general architecture of the designed neural network. Firstly, each sentence is passed through a "Preprocessing" block that performs the formalization of an input sentence: tokenization and representation of each word in a vector space using ELMo semantic embedding model. The next step consists in the vector representation of the sentence by a "Sentence model". In order to perform such a representation two different approaches are utilized:

- passing of ordered word vectors through LSTM cells;
- applying of a set of Transformer-based encoders with a further aggregation of an output matrix.

In the case of the usage of LSTM cells, the processing is made in a word-by-word manner according to their positions within a sentence. However, the Ukrainian language falls into a category of inflected synthetic languages that has less dependency on word positions in comparison with the English language. Therefore, a Transformer-based architecture is also suggested while designing the "Sentence model". Such an architecture provides the usage of a self-attention mechanism and positional encoding of words for taking into account both connections between tokens and their order. The self-attention mechanism may help to reveal relations between text spans despite their sentence positions that can be important for the analysis of Ukrainian-language documents. A more detailed description of corresponding signal flow through layers can be found in our previous work [11]. As the output of encoders is represented as a matrix, the further aggregation is performed by a max-pooling layer in order to represent the whole sentence as a vector according to its most influential parts.

The next components of the network perform the concatenation of vector sentences with further binary classification by dense layers. The output of a sigmoid activation function represents the probability of the coherence of an input clique $p(y_c = 1)$.

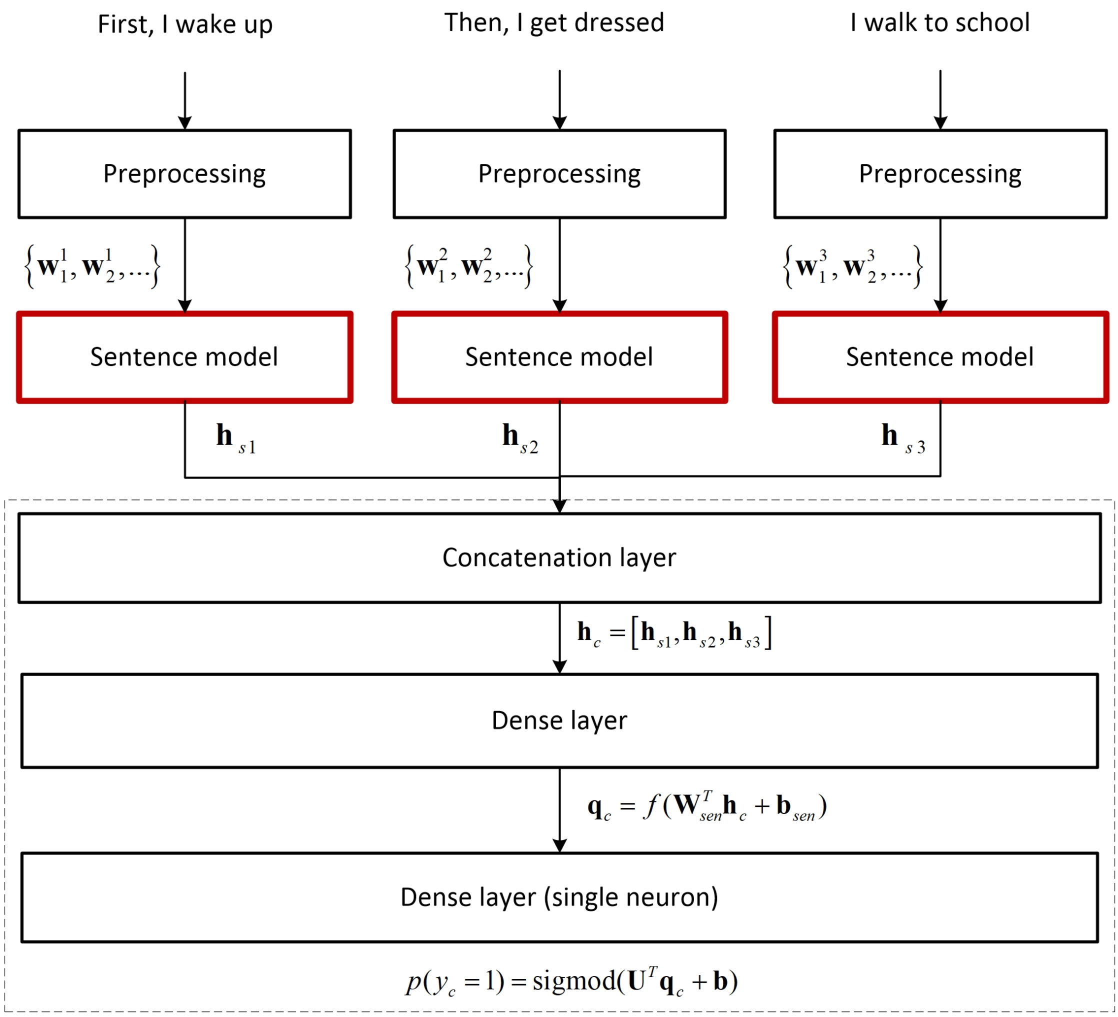

Figure 2: General architecture of a designed binary classifier for the discrimination of coherent and incoherent cliques; a variable component "Sentence model" is marked in red

3.2. Training and experimental verification

Training corpus[1] was formed from Ukrainian-language news taken by a web-crawler from a news portal "Ukrayinska Pravda" [12]. Coherent and incoherent cliques were formed from original papers and their permuted versions correspondingly. The parameters of the "Sentence model" were the following: 1024 LSTM cells and 1 encoder (8 attention heads); size of word vectors - 1024. In order to omit the overfitting of the networks, the accuracy of the sentence ordering task was calculated at the end of each epoch. The prepocessing of sentences, namely, their vector representation was made by a pre-trained ELMo embedding model [13].

Two test datasets were utilized for the experimental verification of the accuracy of the

[1]https://drive.google.com/drive/folders/1K_b5rFLyfWM80H5HI2xsAneOulq04Qoj?usp=sharing

Table 1
Accuracy of the solving of sentence ordering and insertion tasks for designed models on different datasets

Sentence model	Dataset	Sentence ordering, %	Insertion, %	Time per document, s
LSTM	#1	82.4	30.8	1.26
Transformer-based		70.2	23.0	0.43
LSTM	#2	77.7	31.0	0.71
Transformer-based		76.8	27.6	0.14

model. The first one[2] (Dataset #1) consisted of mentioned news formed by the web-crawler, an average sentence length - 10, mode - 9, a number of texts - 570. The second one[3] (Dataset #2) incorporated news from other sources, the articles of the National Academy of Sciences of Ukraine, blogs' posts, Ukrainian literature texts; an average sentence length - 8, mode - 6, a number of texts - 1783. Such a choice was made in order to estimate the accuracy of the models for the different types of documents. As has been mentioned earlier, there is still a discussion about the appropriateness of the sentence ordering task (discrimination of original papers from their permuted versions) as the metric of the accuracy of coherence models. Thus, in addition to the sentence ordering task, an insertion task was chosen for the experimental verification. This task correlates with the TOEFL sentence insertion task that requires an understanding of the topic of a text in order to select a correct sentence position. It consists in the detection of the original position of a taken sentence by comparing the corresponding coherence values of created documents with the coherence of the original document. A document is considered recognized if a sentence is inserted into its correct position (the coherence of the original document is higher than the coherence of other variants). The accuracy is calculated as the ratio of recognized documents to their total number. In our case, the longest sentence according to a count of words was chosen for each document assuming that it may contain enough information to detect its native position.

Table 1 shows the results obtained for designed models on different datasets. Despite the expected increase of accuracy while using the self-attention mechanism, the LSTM approach outperforms the Transformer-based sentence model for both tasks for the Dataset #1 and #2. Such a result may indicate the necessity to lay emphasis on the word positions during the sentence representation by a network that underlines the significance of the structural consistency of text spans while evaluating the coherence of Ukrainian-language documents. It should be mentioned that the Transformer-based network shows the higher accuracy for both tasks on the Dataset #2 in comparison with the Dataset #1 in spite of the different topics of the documents of the Dataset #2 and the training corpus. Moreover, a "Time per document" metric shows the speedup of execution into 3-5 times for the Transformer-based model in comparison with the LSTM network. Thus, it can be concluded that the Transformer-based network can be utilized for the training and integration of the coherence estimation model into a general-purpose system.

Taking into account the average number of the sentences of the Dataset #1 and #2, there is an

[2]https://drive.google.com/drive/folders/1NGMjuhMIY1HWOJSd96vleiDH07NTwmYI?usp=sharing
[3]https://drive.google.com/drive/folders/1KNp53vFgghP28zftF6Ej_KASeHKvMNmp?usp=sharing

improvement from 11.1% (no skill model) to 30.8% and from 12.5% to 31.0% for the insertion task correspondingly. Moreover, despite the different types of training documents (news only) and of the texts of the Dataset #2, the metrics for this corpus are lower just for the sentence ordering task while using the LSTM network. Such results of the sentence ordering and insertion task for this dataset can indicate the possibility to apply the trained models for the related issues on different corpora (e.g. the discrimination of human-written or machine-generated texts, automatic extraction of texts from heterogeneous resources, etc.).

In order to understand the coherence evaluation process of neural networks, namely, a way how sentences are grouped and connected with each other, it is suggested to analyze sentences' embedding of the Transformer-based model. The next subsection is devoted to this analysis.

3.3. Analysis of the sentence representation of the Transformer-based model

It is suggested to analyze sentence representation using the output of the "Sentence model". Thus, each sentence of training documents was passed through this part of the network with a further saving of a relation "vector-sentence". The dimension of vectors equals 1024. In order to visualize and group sentences into clusters, the Uniform Manifold Approximation and Projection (UMAP) [14] dimension reduction technique was applied for the formed set of vectors and corresponding labels. Thus, all vectors were projected into a 2D space. According to the retrieved projection, five different clusters of sentences were formed. Fig. 3 shows the graphical representation of retrieved clusters.

Let us analyze the central points of each cluster for understanding the representation of each group. Table 2 shows the correspondence of a number of a cluster to its 2 sentences (translated from the Ukrainian language) that represent central points. The cluster #1 incorporates sentences with direct speech (full quote) with the mention of a corresponding author. Taking into account the type of training documents (news), the sentences of the cluster #2 describe references to the source of information. The cluster #3 contains the biggest number of sentences; this group consists of the sentences that represent different topics: politics, sport, science, etc. The clusters #4 and #5 consists of sentences that are used at the start and finish of a quote correspondingly. Thus, the formed group of sentences may indicate the extraction of a structure of a document (quotes, reference, start and finish of a quote) while analyzing the coherence of a text by the designed model at the first steps. As has been mentioned, the cluster #3 incorporates the sentences of different topics. In order to analyze relations between these sentences, let us select 5 random points from this cluster and their nearest neighbours according to cosine distance. Table 3 shows retrieved results.

According to the relations of sentences and their nearest neighbours, the conclusions about the availability of following different connections between sentences during the analysis of an input clique can be drawn:

- thematic connection (the content of the neighbour sentences of points 1 and 3 refer to a common topic - Moon investigation and weather forecasting correspondingly);
- titles of the paragraphs of articles (point 2);
- common beginning of phrases (point 4);
- first-person narrative (point 5).

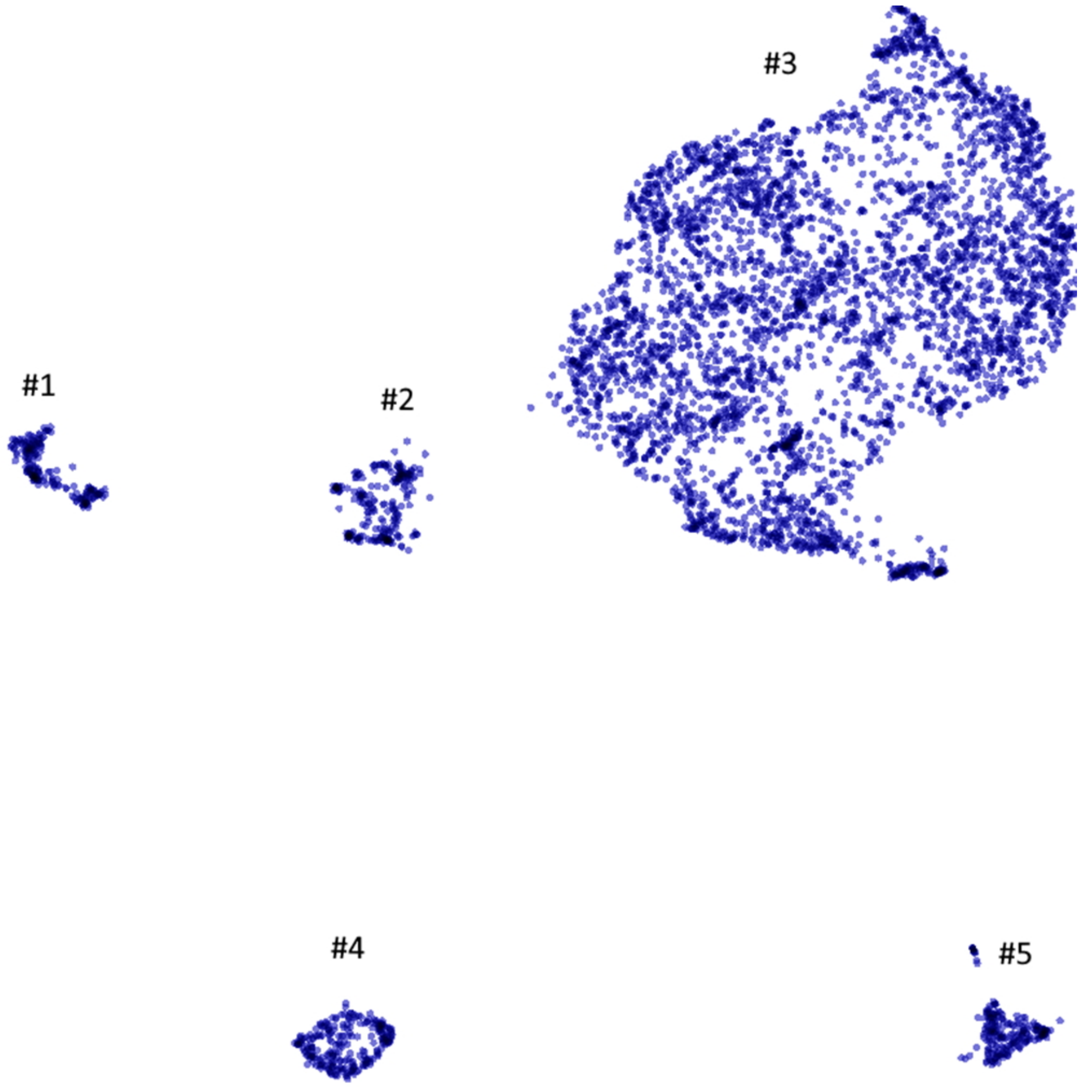

Figure 3: Visualization of the formed clusters of sentences using the UMAP technique

Thus, besides the semantic similarity measure of sentences, there are many other types of connections based on the structure of text spans (quotes, introductory phrases), common pronouns, dates, etc. It underlines the necessity of the analysis of different components for the coherence evaluation of documents. The further usage of the suggested model on other corpora may help to reveal corresponding connections between sentences in order to understand which combination of text components is typical for the creation of the coherent speech of a defined topic.

Table 2
Sentences that correspond to the central points of clusters

Cluster	Sentences
#1	"It changed the whole history of the Middle East", - said historian Joseph Manning, one of the authors. "Our first assumption after studying these petroglyphs is that they were created about 10 thousand years before our era", - commented TEAS Garj, director of the Maharashtra State Department of Archeology.
#2	This was reported by StockWorld with reference to the EIB. This was reported by DW with reference to the AP.
#3	Almost 80 percent of all this waste is plastic products. The thickness of this thinnest layer is only 15 microns (the entire cornea is about 500 microns).
#4	"It happens absolutely every day. "Let's see what the results will be this time.
#5	For them, the animal is part of the family", - says Popiel. It depends on the technology", - Babiy emphasizes.

Table 3
Example of the sentences of the cluster #3 and their nearest neighbours according to cosine distance

#	Sentence	Neighbour sentences
1	Previously, successful experiments on the cultivation of plant seeds have been repeatedly conducted at the International Space Station.	(1) In particular, during 2018, Chinese companies lost about $ 2.4 trillion, as well as American companies Apple, Ford and others. (2) The station is scheduled to be launched into lunar orbit in 2025.
2	Artificial Intelligence.	(1) Consciousness switch. (2) A new layer of the human eye.
3	In Western Europe and in the south - almost April, +12 ... +20 degrees.	(1) Ice on the roads in places. (2) Fog in most central, southern and eastern regions.
4	According to him, after that he decided to leave Macedonia.	(1) According to him, after the fight against the Puerto Rican, he became a little closer to the desired goal. (2) According to her, at first the researchers thought that in front of them were the remains of ancient bread.
5	This time I originally planned to move from Buenos Aires to Tierra del Fuego.	(1) We thought about them when we made this film. (2) And on Sunday I finished this work.

4. Global coherence estimation

As has been mentioned earlier, the global coherence of a text implies the consistency of its remote spans within an entire document. In order to formalize such a connection, let us represent an input text T as a directed graph $G(V, E)$, where V is a set of vertices (each vertex corresponds to a certain sentence), E denotes edges that represent relations between sentences. Directed edges should be established according to the availability of the relation between vertices; the weight of an edge should represent the consistency measure of sentences. Then, based on established edges and corresponding weights, the coherence value of the text T may be estimated. Such a choice, i.e. the representation of the text T in a graph form, is conditioned by the possibility of this structure to formalize long-distant relations between sentences and to visualize corresponding connections according to their consistency measure. The visualization of the coherence estimation process may help to reveal either weak or strong connections between text spans for the further analysis of the impact of the global thematic consistency of a text on its coherence evaluation.

Thus, there are 3 consequent steps left that need to be implemented for the usage of the suggested graph structure for the evaluation the coherence of a document:

- establishment of the edges of the directed graph G;
- calculation of the weights of edges;
- coherence estimation based on a graph built.

4.1. Establishment of edges

In order to find out which edges should be established between vertices, it is suggested to investigate the dependency of the accuracy of a semantic similarity graph method [15] (MSV approach) on its regulative parameter. The MSV approach implies the representation of a text as a graph $G^M(V^M, E^M)$ where each vertex $v_i^M \in V^M$ represents a certain sentence; edges $e_{ij}^M \in E^M$ are established between vertices if a corresponding weight exceeds a pre-defined threshold value: regulative parameter $\theta \in [0; 0.9]$. The weight is calculated according to cosine distance between sentences' vectors $\mathbf{s}_i, \mathbf{s}_j$ and their positions (i, j) within a text:

$$weight(e_{ij}) = \frac{\cos(\mathbf{s}_i, \mathbf{s}_j)}{|i - j|} \tag{3}$$

The coherence of a text is calculated as the average edges weights' value of the graph $G^M(V^M, E^M)$. Thus, the investigation of the dependency of the method's accuracy on the value of the regulative parameter θ may help to reveal the appropriate topology of the graph according to an assigned task. As the weight value depends on the positions of sentences, only connections that represent non-remote sentences are left during the increase of θ. Thus, the analysis of the retrieved most effective graph structure can provide information about a maximum distance between sentences that should be taken into account during the global coherence estimation.

In order to represent sentence vectors, the "Sentence model" of the previously considered Transformer-based neural network was utilized. The sentence ordering task was chosen as an accuracy metric. Fig. 4 demonstrates the dependence of the accuracy of the method (MSV approach) on the value of the regulative parameter θ (step equals 0.1) while solving the sentence ordering task (SOT).

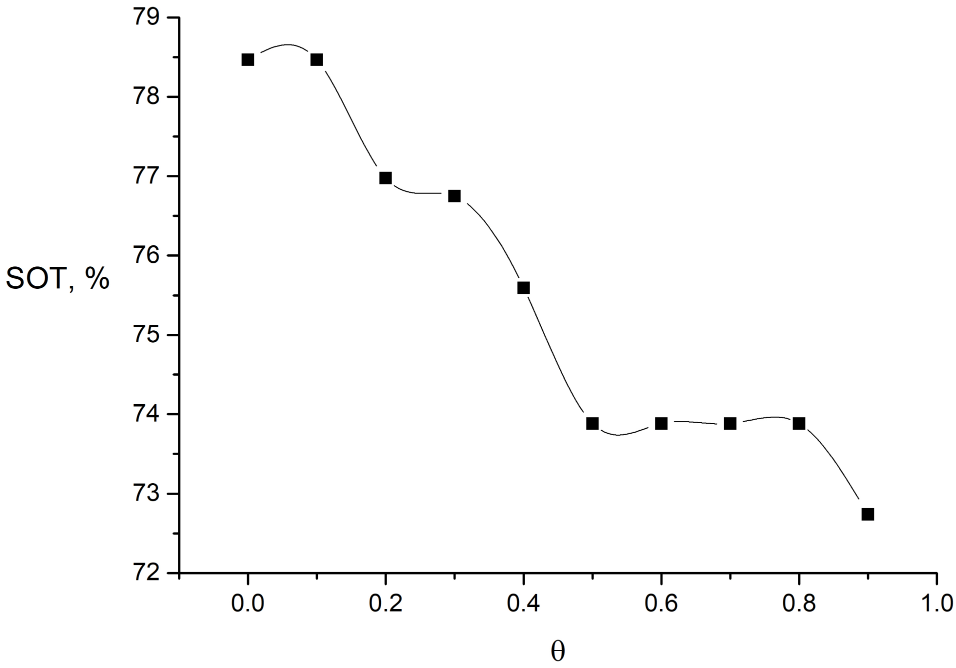

Figure 4: Dependence of the accuracy of the MSV approach on the value of the regulative parameter θ while solving the sentence ordering task

As can be seen from this plot, the highest accuracy was achieved with values $\theta = \{0.0, 0.1\}$; the further increase of θ leads to the decrease of the accuracy. Such a result indicates the advisability to establish edges between all sentences despite their positions within a document. Thus, a graph $G(V, E)$ that represents an input text should be considered as complete and acyclic. The next steps consist in the evaluation of edge weights with further coherence estimation. In order to implement these steps, it is suggested to design and train a graph neural network.

4.2. Graph neural network

Graph neural networks [16] are widely utilized for the solving of different tasks (classification, embedding projection, etc.) connected with the processing of graph structures. Such models are based on capturing the dependence of graphs via message passing between their vertices. According to the previously assigned tasks, the input of the network is represented as a complete acyclic unweighted graph $G(V, E)$; the output of the network should interpret the coherence measure of the input structure. Each vertex $v_i \in V$ is represented by a sentence vector. Fig. 5 demonstrates the general architecture of the designed neural network.

As can be seen from the figure, the neural network incorporates three main components: initialization of nodes vectors, a message-passing mechanism [17] and a classification model.

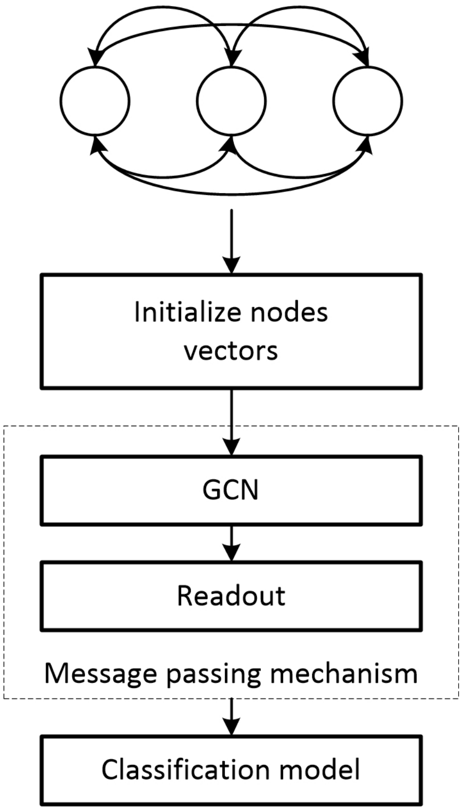

Figure 5: General architecture of the designed neural network for the global coherence estimation

Initialization of sentence vectors is performed using the previously considered "Sentence model" based on LSTM cells; corresponding parameters are set as trainable. As for the message passing mechanism, its purpose consists in the vector representation of vertices due to the relation between them (graph convolutional network, GCN) with the further aggregation of the retrieved matrix into a common vector (readout). The update of the vector representation of vertices according to the GCN is described by the following equations:

$$\mathbf{M}_i = \max_{j \neq i} e_{ij} \mathbf{v}_j \tag{4}$$

$$\mathbf{v}'_i = (1 - \eta_i)\mathbf{M}_i + \eta_i \mathbf{v}_i, \tag{5}$$

where $\mathbf{M}_i$ represents the impact of other vertices on the current one, e_{ij} is a weight of edge between vertices i and j (calculated using separate sequence of dense layers), $\mathbf{v}$ is a vector representation of a node, η_i is a free parameter that denotes a forget threshold during the update of the vector of a node $\mathbf{v}'_i$. After the update of all vertices, the readout function is applied in order to aggregate nodes' vectors into a common one. It was decided to utilize an extra LSTM layer for the aggregation of vertices according to the positions of sentences:

Table 4
Accuracy of the solving of sentence ordering and insertion tasks for the designed global coherence estimation model (GCN) in comparison with the LSTM-based model

Model	Dataset	Sentence ordering, %	Insertion, %	Overlapping, %
LSTM	#1	82.4	30.8	37.0
GCN		75.4	27.0	
LSTM	#2	77.7	31.0	32.8
GCN		76.5	23.6	

$$\mathbf{h}_G = \mathrm{LSTM}(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{|V|}) \tag{6}$$

The next classification model implies the recognition of the input vector $\mathbf{h}_G$, namely, whether it represents coherent information or not.

The training process was performed on the same dataset as for local coherence estimation models with the calculation of the accuracy of the sentence ordering task at the end of each epoch. All pre-trained models are available on Google Colab[4] for further testing and usage in other related researches. Table 4 shows the accuracy of the solving of sentence ordering and insertion tasks on different datasets for the designed global coherence estimation model.

As can be seen, the LSTM-based local coherence estimation model outperforms the designed GCN-based model while solving the tasks on both datasets. The last column shows the overlapping percentage of the documents recognized by the global coherence estimation model and the LSTM-based model while solving the insertion task. The corresponding percentage does not exceed 37% that indicates the different types of documents that were processed by models in the right manner. Let us calculate the average number of the sentences of texts that were recognized by either the global coherence model or the LSTM-based model. According to the results obtained, the average sentence number is greater for the LSTM-based model (9 and 8 sentences for the Dataset #1 and #2 correspondingly) in comparison with the global coherence estimation model (7 sentences for both datasets). Such results may indicate the dependency of the effectiveness of models on the text size while solving typical tasks. In order to verify such dependency, it is suggested to decrease the number of sentences of texts for both datasets; thus, the corresponding number belongs to a range $[3; 6]$. Table 5 shows the accuracy of the solving of sentence ordering and insertion tasks on the created Dataset #3 (144 documents) and Dataset #4 (981 documents) using the mentioned models.

As can be seen, the global coherence model outperforms the LSTM-based model on new datasets while solving both tasks (except for the insertion task on the Dataset #4). Thus, it can be concluded that there are different directions for the usage of the designed models while analyzing the coherence of a document. The local coherence model may be utilized for the tracking of the consequent connection between sentences within a whole document; however, the building of the dependency of the coherence of local groups on the sequence number of each clique can help to reveal weak thematic connections just between neighbour text spans. As for the global coherence model, it may be suitable for the analysis of the consistency of the

[4]https://colab.research.google.com/drive/1iF6t9UIN8xtahE_e3fNzMea6un6VR5UP?usp=sharing

Table 5
Accuracy of the solving of sentence ordering and insertion tasks for the designed global coherence estimation model (GCN) and the LSTM-based local coherence model on new Dataset #3 and #4

Model	Dataset	Sentence ordering, %	Insertion, %
LSTM	#3	72.2	44.4
GCN		74.3	48.6
LSTM	#4	72.7	35.0
GCN		74.7	32.4

fragment of a text (e.g., paragraph, excerpt) for more detailed consideration of the connection between all sentences. The increase in the size of an input excerpt can lead to the complication of such analysis due to the lower impact of remote text spans on each other and the complexity of the appropriate visualization of a graph that represents the whole document.

5. Conclusions

According to the performed investigation of the effectiveness and principal of work of the suggested models based on neural network for the local and global coherence estimation of Ukrainian-language texts, the following conclusions can be drawn:

- Suggested local coherence estimation model based on LSTM cells outperformed corresponding suggested Transformer-based approach while solving typical tasks on different Ukrainian-language test datasets. Such results may indicate the appropriateness of taking into account word positions within a text while analyzing text spans for the coherence evaluation of Ukrainian corpora. However, the usage of the Transformer-based architecture allowed to achieve a document processing speedup 3-5 times in comparison with the LSTM network that can underline the possibility of the usage of this model in general-purpose systems.
- According to the analysis of sentence representation in a vector space performed by the Transformer-based model, five different clusters were detected that had been formed according to the structure of sentences (quote, reference, etc.). Thus, it is possible to draw a conclusion about the structural consideration of sentences by the coherence estimation model at first. More detailed analysis of the biggest cluster revealed the availability of the different type of connections (thematic relation, common start of phrases, titles, first-person narrative) between sentences according to the common stylistic design of training documents (news) that may indicate the possibility of the usage of the suggested sentence model for the analysis of the combination of text spans that are typical for the generation of a coherent speech in a considered area.
- The negative dynamic of the change of the accuracy of the semantic similarity graph method during the increase of a regulative parameter may indicate the availability of the connection between all sentences during the global coherence estimation of a Ukrainian-language text.
- Despite the lower accuracy of the designed global coherence model on Dataset #1 and #2 in comparison with the LSTM-based model, the first one outperformed the second one on

texts with the smaller average number of texts' sentences. Such results may indicate the advisability of the usage of the designed models for different purposes while analyzing the coherence of Ukrainian documents. While the local coherence model may be suitable for the analysis of the connection between neighbour sentences within a whole text, the global coherence model can be appropriate for a more detailed analysis of the consistency of text spans within a single short excerpt.

Acknowledgments

We thank Andriy Rysin for providing us with the Ukrainian corpus formed by texts from different sources.

References

[1] C. Petersen, C. Lioma, J. G. Simonsen, B. Larsen, Entropy and graph based modelling of document coherence using discourse entities: An application to ir, in: Proceedings of the 2015 International Conference on The Theory of Information Retrieval, ICTIR '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 191–200. doi:10.1145/2808194.2809458.

[2] D. Iter, J. Yoon, D. Jurafsky, Automatic detection of incoherent speech for diagnosing schizophrenia, in: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, Association for Computational Linguistics, New Orleans, LA, 2018, pp. 136–146. URL: https://www.aclweb.org/anthology/W18-0615. doi:10.18653/v1/W18-0615.

[3] T. Enos, Encyclopedia of rhetoric and composition, 1 ed., Routledge, 2010.

[4] J. Li, E. Hovy, A model of coherence based on distributed sentence representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 2014, pp. 2039–2048. URL: https://www.aclweb.org/anthology/D14-1218. doi:10.3115/v1/D14-1218.

[5] B. Cui, Y. Li, Y. Zhang, Z. Zhang, Text coherence analysis based on deep neural network, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 2027–2030. doi:10.1145/3132847.3133047.

[6] A. Lai, J. Tetreault, Discourse coherence in the wild: A dataset, evaluation and methods, in: Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue, Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 214–223. URL: https://www.aclweb.org/anthology/W18-5023. doi:10.18653/v1/W18-5023.

[7] M. Mesgar, M. Strube, A neural local coherence model for text quality assessment, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 4328–4339. URL: https://www.aclweb.org/anthology/D18-1464. doi:10.18653/v1/D18-1464.

[8] H. C. Moon, T. Mohiuddin, S. Joty, C. Xu, A unified neural coherence model, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the

9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 2262–2272. URL: https://www.aclweb.org/anthology/D19-1231. doi:10.18653/v1/D19-1231.

[9] L. Pishdad, F. Fancellu, R. Zhang, A. Fazly, How coherent are neural models of coherence?, Proceedings of the 28th International Conference on Computational Linguistics (2020) 6126–6138. doi:10.18653/v1/2020.coling-main.539.

[10] A. Kramov, S. Pogorilyy, Automated methods of coherence evaluation of ukrainian texts using machine learning techniques, PROBLEMS IN PROGRAMMING 2-3 (2020) 295–303. doi:10.15407/pp2020.02-03.295.

[11] A. Kramov, S. Pogorilyy, Evaluation of the coherence of ukrainian texts using a transformer architecture, 2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT) (2020) 296–301. doi:10.1109/atit50783.2020.9349355.

[12] O. Prytula, Ukrayinska pravda - online news about ukraine, 2020. URL: https://www.pravda.com.ua/eng/.

[13] M. Fares, A. Kutuzov, S. Oepen, E. Velldal, Word vectors, reuse, and replicability: Towards a community repository of large-text resources, in: Proceedings of the 21st Nordic Conference on Computational Linguistics, Association for Computational Linguistics, Gothenburg, Sweden, 2017, pp. 271–276. URL: https://www.aclweb.org/anthology/W17-0237.

[14] L. McInnes, J. Healy, N. Saul, L. Großberger, Umap: Uniform manifold approximation and projection, Journal of Open Source Software 3 (2018) 861. doi:10.21105/joss.00861.

[15] S. Tan, Z. Zhou, Z. Xu, P. Li, On efficient retrieval of top similarity vectors, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 5236–5246. URL: https://www.aclweb.org/anthology/D19-1527. doi:10.18653/v1/D19-1527.

[16] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, Graph neural networks: A review of methods and applications, AI Open 1 (2020) 57–81. URL: https://www.sciencedirect.com/science/article/pii/S2666651021000012. doi:https://doi.org/10.1016/j.aiopen.2021.01.001.

[17] G. Nikolentzos, A. Tixier, M. Vazirgiannis, Message passing attention networks for document understanding, Proceedings of the AAAI Conference on Artificial Intelligence 34 (2020) 8544–8551. URL: https://ojs.aaai.org/index.php/AAAI/article/view/6376. doi:10.1609/aaai.v34i05.6376.