Text coherence new method using word2vec sentence vectors and most likely n-grams

Mohamad Abdolahi
Kharazmi International Campus Shahrood University
Shahrood, Iran
mabdolahi512@yahoo.com

Morteza Zahedi
Kharazmi International Campus Shahrood University
Shahrood, Iran
zahedi@ganjineh.co.ir

***Abstract*— Discourse coherence modeling evaluation remains a challenge task in all Natural Language Processing subfields. Most proposed approaches focus on feature engineering, which accepts the sophisticated features to capture the logic, syntactic or semantic relationships between all sentences within a text. This paper investigates the automatic evaluation of text coherence. We introduce a fully-automatic rich statistical model of local and global coherence that uses word2vec approach to assess the coherence a document. Our modeling approach relies on numerical vectors derived from word2vec algorithm applied on a very large collection of texts. We successfully combined the word2vec vectors and most likely n-grams with cohesive LD-n-grams perplexity to assess the coherence and topic integrity of document. We present experimental results that assess the predictive power that it does not depend on the language and its semantic concepts. So it has the ability to apply on any language. Our model achieves state-of-the-art performance in coherence evaluation and order discrimination task on two datasets widely used in the previous methods.**

Keywords*—: *global text coherence; local text coherence; language models; word embeddings

I. INTRODUCTION

Text coherence is one of the most important and fundamental research fields of each generating or manufacturing text output. The use of automatic methods for evaluating or increasing the quality of coherence is considered one of the key problems in all text processing systems outputs such as statistical machine translation [1][2][3][4], text generation, mode detection, question answering, student essay scoring [5][6][7] and text summarization [8][9][10]. It is also quickly becoming mainstream and fundamental post processing in all NLP tasks. The targets of these methods have various dimensions like text quality, the correct position of words in a sentence, correct position of sentences in document, and semantic and conceptual continuity of consecutive sentences. It is also achieved through syntactical features such as the use of deictic, anaphoric and idiom elements or a logical tense structure. In the other word, coherence is a property of well-written texts that makes them easier to read and understand than a sequence of randomly consecutive sentences. Text coherence evaluation is divided into two main categories of local and global coherence evaluation. For sentences, local coherence means the well connectedness of adjacent sentences through lexical cohesion [11] or entity repetition [12] and global coherence is the discourse-level relation connecting remote sentences or adjacent paragraphs [13][14]. It is clear that the coherence of the text on both levels increases the comprehension of the text [15]. Up to now, much research has been done on local coherence, but less attention is done on research in the field of public evaluation. According to most studies, assess the integrity of the text is a semantic issue. But in this study, we are looking to evaluate the text coherence with statistical methods with both local and global coherence which captures text organization at the level of sentence to sentence and paragraph to paragraph transitions. This assessment is done regardless of the meaning of words and handcrafted rules.

Following reasons are the main causes of proposing simple and practical assessment of public coherence with simple algorithm and high accuracy:

- Firstly, a paragraph is a big part of each document and the subject integrity of each paragraph as a local cohesive unit is previously assessed.
- Secondly, the number of paragraphs in a text is much less than the number of its sentences, evaluating the subject dependency of few paragraphs is very simple operation than all sentences dependency in the document. It is global text coherence evaluation.

II. RELATED WORKS

The task of text coherence evaluation was first introduced by Foltz [16]. He bring to study text coherence as a function of semantic relatedness between two adjacent sentences within a text, and engaged a vector-based representation of lexical meaning to compute the semantic relatedness between sequence sentences. Since then, several supervised approaches such as entity-based model [3][4][17][18][19][20][21], discourse relation-based model [22], syntactic patterns-based model [23], co reference resolution-based model [24][25], content-based model via Hidden Markov Model [26][27] and cohesion-driven based model have been proposed. Broadly of these methods compute the relationship topics in adjacent sentences to compute the coherence in a supervised way. One of the most important approaches in coherence evolution is the models based on lexical chain. Lexical chains provide a representation of the lexical cohesion structure of a text. The

978-1-5386-4972-5/17/$31.00 ©2017 IEEE

words of a text can be presented by features Introduced in the previous section are causes the conceptual and thematic relationship between sentences in document. D. Xiong, Y. Ding are proposed a lexical chain model for evaluation of machine translation output [28]. S. Somasundaran and J. Burstein, Martin Chodorow are presents an investigation of lexical chaining methods for measuring discourse coherence quality in test-taker essays [18].

Entity based models are the most famous approaches proposed for evaluating the local coherence. The model analyses the grammatical role of words in adjacent sentences, and extracted patterns from them to assess local coherence [32]. The model proposed by R. Barzilay, M. Lapata for the first time [17][19][30], but some novel approaches such as neural network models [33] and original bipartite graph [34] are proposed in recent years. In 2013, Strube and Guinaudeau are proposed an approach that offers a combination of entity grade and graph-based model. They overcome the limitation ability of entity grade to detect consistency in just neighbor's sentences [35]. The other novel method is proposed by Petersen and Simonsen [36]. Their model is a combination of graph theory and entropy method for assessing the consistency of document sentences. The novelty of their model is that by increasing more nouns in the document, more peripheral information is participating in the context. With an increase in adverse information in text focusing on main issue has fallen down and led to lowering the global coherence. M. Mesgar and M. Strube introduce novel graph-based coherence features based on frequent subgraphs and compare their ability to assess the readability of Wall Street Journal articles [37]. To achieve this goal, they use the entity graph coherence model by Guinaudeau and Strube. The main idea of the model is that the coherence texts are consistent of particular patterns in their extracted subgraphs. Methods based on statistical machine translation algorithms are one of the most popular methods in text coherence evaluation [29][38]. In the mentioned models EM and IBM algorithm in statistical machine translation are used. The main idea of these models is the meaning of each word in the target language introduced several words. Therefore each word lead to link into multiple sentences and the algorithm chooses the most likely sentences.

In order to overcome the limitation of semantic features, modern approaches try to use neural network to extract the syntactic representation of discourse coherence [39]. Li et al. [40] proposed neural deep model. L. Logeswaran and H. Lee offered an end-to-end neural approach based method to address the sentence ordering problem. Their approach tried to model coherence recent successes in capturing semantics using distributed representations and using RNNs for sequence modeling tasks [41].

III. Text Preprocessing

The first subject matter to consider any text processing approach is preparing input text to apply text processing algorithms. The words appear in documents are often have many structural variants. Therefore, before applying the processing algorithm on any text, the structure of its words and sentences must be the same format and standard. On the other hand, text preprocessing is different according to type of text, language and type of text processing algorithm. So choosing the preprocessing type and the percentage of exploiting rate of it to text has a huge impact on the accuracy and speed of the final processing algorithm. The most important preprocessing methods are Tokenization, Stop word removal, Stemming and POS tagging. But their biggest challenge is their restriction to a particular filed, having no ability to apply and extend to all areas and languages.

A. Text preprocessing required in the proposed method

In our proposed method, all sentence components are needed. Therefore, the preprocessing is different from other previous methods. In this approach text is converted to separate sentences and sentence's matrix is created by word2vec word vectors and normalized by n-grams model. So the following preprocessing operations are not performed:

- Stop words: the sentence matrix is including all word vectors. Any words have an important impact on other words and it should not be deleted.
- Stemming: There are differences between each word and its stem vectors. For example the vectors of "study" and "studying" are different. Of course, it is quite logical. In addition to its spelling, every word vector created by the word2vec algorithm determines its position in the sentence and its grammatical rule.
- POS tagging: our method is purely statistical. POS tagging is a semantic operation and no need to perform on words.

In this study some basic preprocessing will be done on input text to ready for separating sentences, extracting word vectors, creating numerical matrix of each sentence, diagnosing most likely n-grams, and comparing matrices.

- Ignoring and removing spacing between words and punctuation marks.
- Removing extra spaces characters between words.
- Do not convert uppercase to lowercase characters.
- Unification of accented characters.

B. combining n-grams model and word2vec to sentences matrix size normalization

All texts are not equal in sentences size and their created matrices are having different size. Word2Vec approach only calculates word vectors and there is not any comment on phrase vectors or n-gram vectors. To reduce the matrix rows for sentences with words more than threshold, n-grams vectors can be used. Bi-grams to 5-grams vectors can be obtained by combining and averaging uni-grams vectors. We employ more than one n-grams vector to normalize the larger sentences matrix. Table (2) and (3) are using most likely tri-grams vector to normalize two sentences matrix size.

"The cat sat on the hat. The dog ate the cat and the hat."

IV. PROPOSED APPROACH

The purely linguistic elements that make a text coherent are subsumed under the cohesion term. However, the mentioned text-based features that guarantied cohesion do not necessarily help achieve coherence. They also do not always contribute to the meaningfulness of a text. It has been stated that a text is coherent when its components have a logical link. We express the probability of a text made up of sentences relatedness. In the proposed method, sentences are accepted as a smallest unit of a coherent text and to assess the integrity of document, topic dependency of sentences is assessed. Then, according to the word2vec vector of each word, sentences matrix are formed. The local coherence in this research is the integrity paragraph. So for the first, sentence dependency of each paragraph is evaluated as local coherence. The task of predicting the next sentence is dependent on its n–i previous sentences.

$$P(t) = P(S_1 .. S_n) = P(S_1)\, P(S_2|S_1)\, P(S_3|S_1, S_2) \dots P(S_n|S^1 \dots S_{n-1}) = \prod_{i=1}^{n} P(S_n|S_1 \dots S_{n-i}) \qquad (1)$$

This is clearly a simplest view of text coherence. But our model has some notion of sentences type that typically go together. It is unlikely to find the exact same sentence repeated several times in a corpus. What we can find and count is the number of times a given structure or word appears in the corpus. We will therefore estimate P(Si | Si-1) from features that express its structure and content.

$$P(S_i \mid S_{i-1}) = P((a_{(i-1)}, a_{(i-2)}, \dots a_{(i-n)}) \mid a_{(i-1,1)}, a_{(i-1,2)}, \dots, a_{(i-1,m)})) \qquad (2)$$

Where $((a_{(i-1)}, a_{(i-2)}, \dots a_{(i-n)})$ are features relevant for sentence S_i and $a_{(i-1,1)}, a_{(i-1,2)}, \dots, a_{(i-1,m)}$ for sentence S_{i-1}.

We use LD bigrams proposed by R. Rosenfeld [29]. He proved the information contained in the history of words is reduced by increasing the distance between them. But the information available will remain constant by the amount of more than five words. The theory is true for semantic integration of all components of text especially the consecutive sentences. So dependencies of sentences with a distance of more than five are almost constant. Given this idea, the value of (n) in formula (1) is equal to five. Our proposed algorithm makes a vector with 12 elements (fig 1):

1. Find topic sentence of paragraph (first sentence).
2. Applied the method to determine the dependency between the first sentence and other sentences in paragraph.
   2-1. topic sentence with second sentence.
   2-2. topic sentence with third sentence.
   … ..
   2-5. topic sentence with sixth sentence.
   2-6. generates first vector.
3. Determined the second sentence as topic sentence and apply the 2th step.
4. Apply N-5 times steps 1 to 3 (N is the number of paragraph sentences)
5. Generate the matrix of paragraph

Fig. 1. Example of a figure caption.

$$\{A_1, \dots, A_5\} = \text{dist}_{i=2..6}(S_1, S_i)$$
$$\{A_6, \dots, A_{10}\} = \text{dist}_{i=1..5}(S_i - S_{i+1})$$
$$\{A_{11}\} = \text{mean}(A_1, \dots, A_5)$$
$$\{A_{12}\} = \text{mean}(A_6, \dots, A_{10})$$

Fig. 2. Calculating the initial vector

$$\{A_{1,k}, \dots, A_{5,k}\}_{k=2..N-5} = (\text{dist}_{i=2..6,k}(S_1, S_i))_{k=2..N-5}$$
$$\{A_{6,k}, \dots, A_{10,k}\} = (\text{dist}_{i=1..5}(S_i - S_{i+1}))_{k=2..N-5}$$
$$\{A_{11,k}\} = \text{mean}(A_{1,k}, \dots, A_{5,k})_{k=2..N-5}$$
$$\{A_{12,k}\} = \text{mean}(A_{6,k}, \dots, A_{10,k})_{k=2..N-5}$$

Fig. 3. Calculating the matrix of paragraph

TABLE I. OBTAINED PERPLEXITY FROM LONG-DISTANCE BIGRAMS, TRAINING ON BROWN DATA SET TEXT WITH A MILLION WORDS [29]

Distance	1	2	3	4	5	6	7	8	9	10	1000
Perplexity	83	119	124	135	139	138	138	139	139	139	141

The five first elements are including coherence amount of topic sentence with the next five sentences. Other five elements are including the difference between the five previous amounts respectively. Eleventh element is including the average of five primary values and the twelfth element is including the average of five secondary values. The reason for determining and applying the difference between the obtained values is the importance of the process of change and reduce the amount gained. Then, the first sentence in paragraph is logically removed and the above algorithm is done on the new paragraph with one less sentence.

After evaluation of local coherence in paragraph, public coherence will be assessed by creating virtual paragraph. Virtual paragraphs consist of title of document and topic sentences of each paragraph. The proposed method applies to virtual paragraph and evaluates the global coherence.

TABLE II. FIRST SENTENCE MATRIX, GENERATING BY COMBINING WORD2VEC AND N-GRAMS MODEL

word		vector				
1	the	-0.16441	0.03211	0.08352	0.07601	…
2	cat	0.27043-	0.23799	0.03023-	0.26455-	…
3	sat	0.32563	0.02780	0.12020-	0.36229	…
4	on	0.13098	0.19419-	0.04385	0.19398	…
5	the	0.16441-	0.03211	0.08352	0.07601	…
6	hat	0.37875	0.06620	-0.02849	0.06544	…
7	.	0.15080	-0.11711	-0.13287	0.18885	…

TABLE III. SECOND SENTENCE MATRIX, GENERATING BY COMBINING WORD2VEC AND N-GRAMS MODEL

word		Vector				
1	the	-0.16441	0.03211	0.08352	0.07601	..
2	dog	-0.14440	0.11054	-0.01605	-0.20122	..
3	ate	0.05521	-0.01441	0.04162	-0.05403	..
4	the	-0.16441	0.03211	0.08352	0.07601	..
5	cat	-0.27043	0.23799	-0.03023	-0.26455	..
6	and the hat	0.07641	0.01832	0.03570	0.04898	..
7	.	0.150808	-0.11711	-0.13287	0.18885	..

V. DATABASE AND EVALUATION METHOD

To test our proposed method, we first select five standard documents. By changing the position of the text sentences, we randomly construct 9 new texts with displaced sentences for each text. The generated texts respectively has 10%, 20%, …, 90% displacement. Thus, we have a small database with fifty texts in five categories of 10 identical texts but different levels of coherence. Then apply the algorithm to each text and compare the results together. If the reduction of coherence is approximately equal to the percentage of displacement of sentences, the proposed method can assess the coherence of input text with high precision.

VI. THE INNOVATIONS OF PROPOSED METHOD COMPARED TO PREVIOUS METHODS

Our proposed method is tried to analyzing problems and shortcomings of previous methods, using new technologies like deep learning, transforming words into numerical vectors and using statistical methods to assess the coherence of texts.

Most of previous approaches based on entity based method [17][19][30][31][32]. In entity based method the integrity of a sentence is evaluate due to the transfer of grammatical status of a noun or noun phrase with next and previous sentences. The disadvantage of these methods is that these transfers are only examined in successive sentences. Each transfer accepted with an error rate and determines the coherence with its next sentence. But with advances in text and in text with large number of sentences and combining these small errors together, the last sentences may not have any thematic relation with the beginning sentences, but the document assumes coherent. More importantly, this is often not the same two consecutive sentences have any common entity. In this case, the coherence of these two sentences is assumed very poorly, that may not be the case.

But in our proposed method, only consecutive sentences are not compared. Rather, all the sentences are compared with the topic sentence. Meanwhile, in an iterative algorithm each sentence is considered topic sentence for its next sentences. Finally, the coherence between all sentences is evaluated simultaneously.

In entity-based approaches, two adjacent sentences are only used certain words like names and name phrases to distinguish and evaluate the relationship. The method presented in this article not dependent on a particular word or its grammatical status. Typically, all words in sentences with each grammatical position participate in the recognition of coherence.

Methods that emphasize the assessment of global coherence often use graph theory in combination with entity-based approaches. These methods, in addition to having problems with traditional entity based methods, they also have problems with graphs theory. The most important problem comes with text enlargement, increasing the number of sentences and their entities which causes too large and complex graphs. Of course, some solutions such as dimensionality reduction, extracting subgraphs and bipartite graphs have been proposed for this problem. Although these methods lead to simplify the processing algorithms, but in any case, they reduce accuracy.

Extraction of entities such as nouns and noun phrases is more in the semantic field which makes the design system involved with semantic concepts in the field of linguistics. Even in graph-based methods, which are computational sections of graphs theories, entities extracting are semantic. But the proposed approach is not involved with semantic concepts at all.

The most important innovation of the proposed approach is statistical pre-processing operations, matrices comparison and coherence evaluation. Statistical approaches in text processing led to get rid of engaging with meaning of words, language, scope and size of text. It is also able to apply on combined texts more than one language.

In previous approaches, local coherence is examined at the level of several consecutive sentences. Therefore, sections with a greater distance in the text may have a weakly related topic or not have any relation. But in the proposed approach, local coherence is raised at the level of a paragraph and a coherent paragraph is assumed a local coherent section. Detecting local consistency on a scale of one paragraph in addition to being simple includes larger text area. The proposed method evaluates public coherence by creating a small virtual paragraph, combined topic sentences of each paragraph in the text. Then it assesses the coherence of virtual paragraph. The action leads to public coherence evaluation, which does not have much computational perplexity.

CONCLUSION

Modeling text coherence is an area in text processing that has received a lot of interest with a surge of automatic methods in two last decades. In this paper, we present a wordembeding model for English discourse coherence modeling. We successfully combined the word2vec vectors and most likely n-grams with cohesive LD-n-grams perplexity to assess the coherence and topic integrity of document. Experiments show that our model is robust among language and domains. Our proposed method has some advantages over the previous models. It will not involve semantic concepts of words and has much more easier text pre-processing algorithm. The method has a strong math foundation due to numeric matrices and matrix operations performing. Our model does not suffer from the computational complexity and the problem of data fragmentation .Provided word vectors for other language, it will be also easily applicable on other languages. Finally, these promising results on local and global coherence modeling make us believe that our vector space based representation can be used without much modification for other tasks such as extractive summarization or topic segmentation. Evaluation results on sentence reordering documents show the effectiveness of the proposed model. Our future work is to apply the proposed method on Persian documents with other coherence evaluation task.

REFERENCES

[1] Z. Lin, C. Liu, H. T. Ng, and M. Y. Kan. "Combining coherence models and machine translation evaluation metrics for summarization evaluation". In ACL (1), pp. 1006–1014, 2012.

[2] D. Xiong, Y. Ding, M. Zhang, and C. L. Tan. "Lexical chain based cohesion models for document-level statistical machine translation". In EMNLP, pp. 1563–1573, 2013.

[3] D. Xiong, M. Zhang, and X. Wang, "Topic-based coherence modeling for statistical machine translation", Trans. Audio, Speech and Lang", Proc., vol. 23, pp. 483–493, 2015.

[4] C. Petersen, C. Lioma, J. G. Simonsen, and B. Larsen. "Entropy and graph based modelling of document coherence using discourse entities: an application to IR", In ICTIR, pp. 191–200, 2015.

[5] H. Yannakoudakis, T. Briscoe,"Modeling coherence in ESOL learner texts", In Proceedings of ACL, pp. 33-43, 2012.

[6] J. Burstein, J. Tetreault and S. Andreyev,"Using entity-based features to model coherence in student essays", In proeedings of NAACL-HLT, pp. 681-684, 2010.

[7] D. Higgins, J. Burstin, D. Marcu and C.Gentile,"Evaluating multiple aspects of coherence in student essays", In proceedings of NAACL-HLT, pp. 185-192, 2004.

[8] A. C¸elikyilmaz, D. Hakkani, "Discovery of topically coherent sentences for extractive summarization". In ACL, pp. 491–499, 2011.

[9] D. Parveen and M. Strube. "Integrating importance, non-redundancy and coherence in graph-based extractive summarization". In IJCAI, pp. 1298–1304, 2015.

[10] R. Zhang. "Sentence ordering driven by local and global coherence for summary generation". In ACL, pp. 6–11, 2011.

[11] M. A. K. Halliday, R. Hasan, "Cohesion in english", London: Longman, 1976.

[12] B. J. Grosz, A. K. Joshi., and S. Weinstein, "Centering: a framework for modeling the local coherence of discourse", computational linguistics; vol. 21, pp. 203–225, 1995.

[13] W. C. Mann, S.Thompson, "Rhetorical structure theory: toward a functional theory of text organization:, 1988.

[14] A. Kehler, "Coherence, reference, and the theory of grammar", Stanford, California: CSLI Publications, 2002.

[15] I. Tapiero, "Situation models and levels of coherence: Towards a definition of comprehension", Mahwah, New Jersey: Lawrence Erlbaum Associates, 2007.

[16] P. W. Foltz, W. Kintsch and T. K. Landauer, "The measurement of textual coherence with latent semantic analysis",Discourse Processes, pp. 285-307, 1998.

[17] R. Barzilay, M. Lapata, "Modeling local coherence: An entity-based approach", Proceeding ACL '05 the 43rd annual meeting on association for computational linguistics, pp. 141-148, 2005.

[18] S. Somasundaran, J. Burstein, M. Chodorow, "Lexical chaining for measuring discourse coherence quality in test-taker essays", Proceedings of COLING 2014, the 25th iInternational conference on computational linguistics: Technical papers, pp. 950–961, 2014.

[19] R. Barzilay, M. Lapata, "Modeling local coherence: An entity-based approach", Computational Linguistics, Vol 34, pp. 1-34, March 2008.

[20] V. W. Feng, G. Hirst, "Extending the entity-based coherence model with multiple ranks", In Proceedings of EACL, pp. 315-324, 2012.

[21] M. Zhang, V. W. Feng, B. Qin, G. Hirst, T. Liu, and J. Huang, "Encoding world knowledge in the evaluation of local coherence", In NAACL HLT, pp. 1087–1096, 2015.

[22] Z. H. Lin, H. T. Ng and M. Y. Kan, "Automatically evaluating text coherence using discourse relations", In Proceedings of ACL, pp. 997-1006, 2011.

[23] A. Louis, A. Nenkova,"A coherence model based on syntactic patterns", In Proceedings of EMNLP-CNLL, pp.1157-1168, 2012.

[24] R. Iida, T. Tokunaga,"A metric for evaluating discourse coherence based on coreference resolution", In Proceedings of COLING, Poster, pp.483-494, 2012.

[25] M. Elsner, E. Charniak,"Coreference-inspired coherence modeling", In Proceedings of ACL, Short Papers, pp.41-44, 2008.

[26] M. Elsner, J. Austerweil and E Charniak,"A unified local and global model for discourse coherence", In Proceedings of NAACL, pp.436-443, 2007.

[27] F. Xu, Q. Zhu, G. Zhou and M. Wang,"Cohesion-driven discourse coherence modeling", Journal of Chinese Information Processing, vol. 28, pp. 11-21, 2014.

[28] D. Xiong, Y. Ding, "Lexical chain based cohesion models for document-level statistical machine translation", Proceedings of the 2013 conference on empirical methods in Natural Language Processing, pp. 1563–1573,Seattle, Washington, 2013.

[29] R. Rosenfeld, "A maximum entropy approach to adaptive statistical language modeling", Computer Speech & Language, pp. 187–228, 1996.

[30] M. Lapata, R. Barzilay, "Automatic evaluation of text coherence: models and representations", Proceeding in IJCAI'05 Proceedings of the 19th international joint conference on artificial intelligence, pp. 1085-1090, Edinburgh, Scotland, 2005.

[31] J. Burstein, J. Tetreault, S. Andreyev, " Using entity-based features to model coherence in student essays", Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL. Los Angeles, CA: Association for Computational Linguistics, pp.681-684, 2010.

[32] K. Filippova, M. Strube, "Extending the entity-grid coherence model to semantically related entities", Proceeding in ENLG '07 the eleventh european workshop on Natural Language Generation, pp. 139-142, Germany, June 17 - 20, 2007.

[33] C. Lioma, F. Tarissan, "Exploiting the bipartite structure of entity grids for document coherence and retrieval", ICTIR" ,16, September 2016 ACM. ISBN 978-1-4503-4497-5/16/09. $15.00, 2016.

[34] F. Xu, S. Du, "An entity-driven recursive neural network model for chinese discourse choherence modeling", International Journal of Artificial Intelligence and Applications (IJAIA), Vol.8, No.2, 2017.

[35] C. Guinaudeau, M. Strube, "Graph-based local coherence modeling", Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pp. 93–103, 2013.

[36] C. Petersen , J. G. Simonsen "Entropy and graph based modeling of document coherence using discourse entities: an application to IR", Proceedings of the 2015 International Conference on The Theory of Information Retrieval, pp. 191-200, 2015.

[37] M. Mesgar, M. Strube "Graph-based coherence modeling for assessing readability", Proceedings of the fourth joint conference on lexical and computational semantics, pp. 309–318, 2015.

[38] R. Soricut, D. Marcu, "Discourse generation using utility-trained coherence models", Proceedings of the COLING/ACL on Main conference poster sessions, pp. 803-810, 2006.

[39] J. Li, E. Hovy, "A Model of coherence based on distributed sentence representation", Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.2039-2048, 2014.

[40] J. Li, E. Hovy, "A model of coherence based on distributed sentence representation", In Proceedings of EMNLP, pages2039-2048, 2014.

[41] L. Logeswaran, H. Lee, D. Radev, "Sentence ordering using recurrent neural networks, Cite as:arXiv:1611.02654 [cs.CL], 8 Nov 2016.