A Neural Local Coherence Analysis Model for Text Clarity Scoring

Panitan Muangkammuen[1], **Sheng Xu**[1,2], **Fumiyo Fukumoto**[1],
Kanda Runapongsa Saikaew[3], **Jiyi Li**[1]
[1]University of Yamanashi
{g19tk021,g18tka07,fukumoto,jyli}@yamanashi.ac.jp
[2]Hangzhou Dianzi University
[3]Khon Kaen University
krunapon@kku.ac.th

Abstract

Local coherence relation between two phrases/sentences, such as cause-effect, and contrast gives a strong influence on whether a text is well-structured or not. This paper follows the assumption and presents a method for scoring text clarity by utilizing local coherence between adjacent sentences. We hypothesize that the coherence knowledge learned from the different domain data is beneficial for capturing a well-structured text and thus helpful for scoring text clarity. We propose a text clarity scoring method that utilizes local coherence analysis with an out-domain setting, i.e., the training data for the source and target tasks are different from each other. The method based on the pre-trained language model BERT firstly trains the local coherence model as an auxiliary manner and then re-trains it together with the text clarity scoring model. The experimental results by using the PeerRead benchmark dataset show the improvement compared with a single model, scoring text clarity model[1].

1 Introduction

Text clarity scoring can be defined as a task that assesses well-structured in a text by grading. It is beneficial for not only authors/reviewers of scholarly papers but also students to improve their writing skills. Among several properties such as spelling, grammar, and word choice, local coherence, which captures text relatedness at the level of sentence-to-sentence transitions (Barzilay and Lapata, 2008), is one of the main properties to identify whether a text is well-structured or not. Well-known early attempts for modeling coherence are lexical coherence models (Halliday and Hasan, 1976), psychological models of discourse (Foltz et al., 1998), rhetorical structure theory (Mann and Thompson, 1998), lexical chains (Morris and Hirst, 1991), and entity grid model (Barzilay and Lapata, 2008). More recently, the coherence model based on deep learning techniques has been intensively studied. These attempts include recursive and recurrent neural networks (Li and Hovy, 2014; Li and Jurafsky, 2017), a combination of LSTM and CNN (Mesgar and Strube, 2018), a deep coherence model based on CNN (Cui et al., 2017), SKIPFLOW LSTM (Tay et al., 2018), and a pre-trained generative model (Xu et al., 2019). It enables to encode patterns of semantic changes in a text. Despite some successes, techniques explored so far mainly rely on word sequence within a sentence. Farag and Yannakoudakis (2019) attempted to encode information about the types of grammatical roles, such as clausal modifiers of nouns and coordinating conjunction in a sentence obtained by using the Stanford Dependency Parser (Chen and Manning, 2014). for coherence modeling. The authors utilize a hierarchical neural network model trained two tasks, grammatical roles, and coherence in a multi-task manner.

In this paper, we propose a method for text clarity scoring by utilizing a coherence model as an auxiliary manner. Instead of training grammatical roles and coherence tasks with the same data, our method explicitly trains the coherence model by using an existing coherence relation training data and utilizes it for scoring text clarity on the target of scholarly papers. Several ideas were proposed for

[1]Our source code is available at https://github.com/panitan-m/local_coh

Proceedings of the 28th International Conference on Computational Linguistics, pages 2138–2143
Barcelona, Spain (Online), December 8-13, 2020

(a) There is a Eurocity train on Platform 1.
(b) Its destination is Rome.
(c) There is another Eurocity on Platform 2.
(d) Its destination is Zürich.

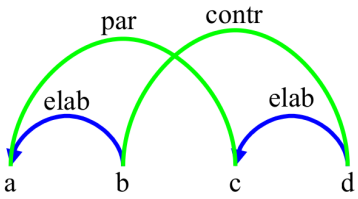

Figure 1: Coherence relation example from Wolf et al. (2003). Sentence (b) elaborates on sentence (a), sentence (a) is parallel with sentence (c). Furthermore, sentence (b) is in contrast with sentence (d).

categorizing coherence relations (Hobbs, 1979; Hovy, 1991; Sanders et al., 1992; Knott and Dale, 1994). We hypothesize that coherence relation knowledge learned from out-domain data is also possible to help to discriminate well-structured of the target text and thus help to score the text clarity. The method based on the language model pre-training BERT (Devlin et al., 2019) firstly trains the local coherence model by using BERT sentential encodings as an auxiliary manner and then re-trains it together with the text clarity scoring model, adapted from the delay multi-task approach (Shimura et al., 2019). In such a way, the coherence relation can also help the model to learn well-structured of a text. The main contributions of our work can be summarized: (1) We propose a method for scoring text clarity based on local coherence relation between two sentences/segments, (2) We introduce a learning framework that firstly trains the local coherence model as an auxiliary manner and then re-trains it together with the text clarity scoring model, (3) We show that the coherence relations learned from out-domain data are also possible to help for capturing the well-structured target text and thus beneficial for scoring the text clarity.

2 Neural Clarity Learning

2.1 Learning Local Coherence with Sentential Encoders

To learn a coherence model, we utilize an existing coherence relation training data provided by Discourse Graphbank 1.0[2] (Wolf and Gibson, 2005). We used eleven coherence relations such as *elaboration* and *parallel* between adjacent sentences as "has coherence relation" and *none* relation between them as "does not have coherence relation". An example passage with coherence relations is shown in Figure 1. Figure 2 illustrates our neural clarity learning framework. The left-hand side of the framework shows a clarity score prediction model, and the right-hand side indicates the local coherence model. The contextualized sentence representation that we used is BERT, a Bidirectional Transformer model (Devlin et al., 2019). A transformer encoder computes the representation of each token through an attention mechanism concerning the surrounding tokens. As shown in the right-hand side of Figure 2, the input of the BERT is a pair (S_i,S_j) of two adjacent sentences $S_i = \{w_1, \cdots, w_m\}$ and $S_j = \{w_1, \cdots, w_n\}$ that are concatenated by a special token [SEP]. It consists of tokens that are segmented by BERT tokenizer using WordPiece embeddings vocabulary (Wu et al., 2016). The representation of each token is the sum of the corresponding token, segment, and position embeddings. The first token of every input is the special token [CLS], and the final hidden state corresponding to this [CLS] token is regarded as an aggregated representation of the input sentence pair. We used this aggregated representation as our sentential embedding of a sentence pair. Let $\mathbf{w} \in \mathbb{R}^{d \times (m+n)}$ be a sentential embedding of a two-sentence pair (S_i,S_j). The sentential embedding is passed to the fully connected layer $\mathbf{FC}_{co}$ and applies the softmax function to obtain probabilities of two predicted labels, "has coherence relation" or "does not have coherence relation" in the output layer. The network is trained with the objective that minimizes the binary cross-entropy loss between the predicted distributions and the actual distributions (one-hot vectors).

2.2 Learning Clarity Score with Paragraph Encoders

Like the coherence model, which takes a pair of two sentences as input and coherence labels as output, the clarity scoring model accepts a paragraph and clarity score as a training instance shown in the left-hand side of Figure 2. The final hidden state of [CLS] token is used as the aggregate representation of

[2] catalog.ldc.upenn.edu/LDC2005T08

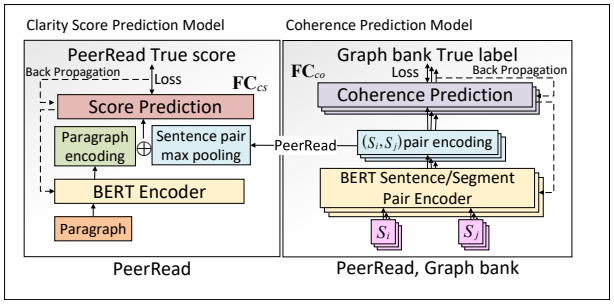

Figure 2: Neural Clarity Learning Framework

the paragraph. Then it is passed through an applied dropout (Hinton et al., 2012) fully connected layer, $\mathbf{FC}_{cs}$. The dropout randomly sets values in the layer to 0. Finally, we obtain the score from linear regression. The network is trained with the objective that minimizes the mean-square error (MSE).

2.3 Clarity Score Prediction

We hypothesize that the coherence relation information learned from out-domain data also helps assess the clarity. For the sentence pair encoding results obtained from the coherence model, we apply a max-pooling operation over the sentence pair timeline. These features from the pooling layer are concatenated with the paragraph encoding obtained from the score prediction model, then passed to a fully connected layer $\mathbf{FC}_{cs}$. The two models' parameters are trained to minimize the MSE.

3 Experiments

3.1 Experimental Setup

We chose the Discourse Graphbank dataset to learn the coherence model. We obtained 8,101 pairs of two adjacent segments from 135 annotated texts 49% of them have coherence relation, and 51% do not. The coherence relations we used consist of elaboration (1,831), attribution (662), parallel (507), cause-effect (340), contrast (189), temporal sequence (135), condition (108), others (183), and none (4,146) relation where the value marked with brackets shows the number of relations in our data. "None" denotes non-coherence relation. We divide these pairs into training, validation, and test data.

The PeerRead dataset (Kang et al., 2018) was selected as a benchmark. We chose ACL 2017 and ICLR 2017, having clarity scores and merged them. We used the Introduction section in our experiments. As a result, we obtained 212 papers, 1,081 paragraphs, which has more than one sentence. We also divided these paragraphs into three folds. Tables 1 and 2 show the statistics of the datasets.

	# of pairs
Training	5,467
Valid	608
Test	2,026
Total	8,101

Table 1: Graphbank Data Statistics: The # of pairs refers to the number of two adjacent discourse segments defined by Foltz et al. (Foltz et al., 1998).

	# of Paragraphs	# of Sentences
Training	691	3,150
Valid	173	820
Test	217	993
Total	1,081	4,963

Table 2: PeerRead Data Statistics: The datasets indicate the collection from ACL and ICLR 2017.

The coherence model adopted BERT_base model as a pre-training then fine-tuned on the Discourse Graphbank dataset. The model takes a segment pair as input and learns to predict whether it has a coherence relation or not. We tuned model hyperparameters, batch size, and learning rate (Adam). We obtained the model with 86% accuracy. When we trained the combined model, the model hyperparameters were fixed, i.e., batch size 8, learning rate 3e-5. Likewise to other baselines, these values were fixed, except the models from PeerRead, we followed the original. We performed the test with ten random initializations of the weight matrices and data splits.

3.2 Results

We compared our model to the baselines: (i) Three models from PeerRead (Kang et al., 2018), (ii) The same three models with BERT_base as word embedding, (iii) Single model which is text clarity scoring based on BERT_base but without coherence identification, (iv) Combined-no-DG is a method that utilizes sentence pair encodings obtained by BERT and does not train a model by using Discourse Graphbank dataset, and (v) Combined method, which utilizes the coherence model pre-trained on Discourse Graphbank dataset, then fine-tunes along with trained single model. The results are shown in Table 3.

Model	MSE
PeerRead-CNN	.0382
PeerRead-RNN	.0346
PeerRead-DAN	.0334
BERT-CNN	.0322
BERT-RNN	.0329
BERT-DAN	<u>.0284</u>
Single	.0288
Combined-no-DG	.0294
Combined	**.0273**

Table 3: Main results: Bold font and underline show the best and the second results, respectively. "PeerRead-X" refers to the result obtained by GloVe840B embeddings which are used in (Kang et al., 2018), and "BERT-X" indicates the result obtained by BERT_base model as embedding.

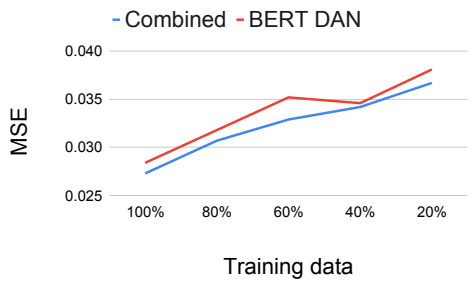

Figure 3: Performance against the percentage of the training data: We ran ten times for each volume of training data size except for 100% and obtained the average MSE.

We can see from Table 3 that the result obtained by "Combined" is the best among "Combined-no-DG" and "Single". Moreover, it is better than the three models from PeerRead and the same models with BERT_base. The result obtained by "Combined-no-DG" is worse than that obtained by "Single" indicates that paragraph encoding is advantageous in predicting text clarity score, but sentence pair obtained by BERT sentence encoding is not, without coherence relation knowledge.

We also examined how the percentage of training data affects overall performance. Figure 3 shows an MSE against the percentage of the PeerRead training data by "Combined" and "BERT-DAN", which are the best and the second result. We ran ten times for each volume of training data size except for 100% and obtained the averaged MSE. Overall, the curves show that more training data helps the performance, while the curves obtained by our model "Combined" increase slowly compared to BERT-DAN, which indicates that local coherence relation between two adjacent phrases/sentences can help to learn a better model for scoring text clarity.

4 Conclusion

We proposed a method for scoring text clarity by utilizing local coherence between adjacent sentences. The experimental results using the PeerRead benchmark dataset show the improvement compared with a single model, scoring text clarity model. Future work will include (i) incorporating the global coherence model into our framework to improve the overall performance of scoring text clarity, (ii) extending our model to capture other criteria such as word choice and grammar for modeling clarity, and (iii) applying our model to the Essay Scoring task.

Acknowledgements

We are grateful to the anonymous reviewers for their insightful comments and suggestions. This work was supported by the Grant-in-aid for JSPS, Grant Number 17K00299, Support Center for Advanced Telecommunications Technology Research, Foundation, and KDDI Foundation Research Grant Program.

References

Regina Barzilay and Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.

Danqi Chen and Christopher Manning. 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 740–750. Association for Computational Linguistics.

Baiyun Cui, Yingming Li, Yaqing Zhang, and Zhongfei Zhang. 2017. Text coherence analysis based on deep neural network. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, page 2027–2030. Association for Computing Machinery.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Youmna Farag and Helen Yannakoudakis. 2019. Multi-task learning for coherence modeling. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 629–639. Association for Computational Linguistics.

Peter W. Foltz, Walter Kintsch, and Thomas K Landauer. 1998. The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, 25(2-3):285–307.

M. A. K. Halliday and R. Hasan. 1976. *Cohesion in English*. Longman, London.

G. E. Hinton, N. Srivastava, A. Krizhevsky, H. Sutskever, and R. R. Salakhutdinov. 2012. Improving Neural Networks by Preventing Coadaptation of Feature Detectors. In *arXiv preprint arXiv:1207.0580*.

Jerry R Hobbs. 1979. Coherence and coreference. *Cognitive science*, 3(1):67–90.

Eduard H Hovy. 1991. Approaches to the planning of coherent text. In *Natural language generation in artificial intelligence and computational linguistics*, pages 83–102. Springer.

Dongyeop Kang, Waleed Ammar, Bhavana Dalvi, Madeleine van Zuylen, Sebastian Kohlmeier, Eduard Hovy, and Roy Schwartz. 2018. A dataset of peer reviews (PeerRead): Collection, insights and NLP applications. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1647–1661. Association for Computational Linguistics.

Alistair Knott and Robert Dale. 1994. Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18(1):35–62.

Jiwei Li and Eduard Hovy. 2014. A model of coherence based on distributed sentence representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2039–2048. Association for Computational Linguistics.

Jiwei Li and Dan Jurafsky. 2017. Neural net models of open-domain discourse coherence. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 198–209. Association for Computational Linguistics.

William C Mann and Sandra A Thompson. 1998. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Mohsen Mesgar and Michael Strube. 2018. A neural local coherence model for text quality assessment. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4328–4339. Association for Computational Linguistics.

Jane Morris and Graeme Hirst. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1):21–48.

Ted J. M. Sanders, Wilbert P. M. Spooren, and Leo G. M. Noordman. 1992. Toward a taxonomy of coherence relations. *Discourse Processes*, 15(1):1–35.

Kazuya Shimura, Jiyi Li, and Fumiyo Fukumoto. 2019. Text categorization by learning predominant sense of words as auxiliary task. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1109–1119. Association for Computational Linguistics.

Yi Tay, Minh C. Phan, Luu Anh Tuan, and Siu Cheung Hui. 2018. Skipflow: Incorporating neural coherence features for end-to-end automatic text scoring. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18)*, pages 5948–5955. AAAI Press.

Florian Wolf and Edward Gibson. 2005. Representing discourse coherence: A corpus-based study. *Comput. Linguist.*, 31(2):249–288.

Florian Wolf, Edward Gibson, Amy Fisher, and Meredith Knight. 2003. A procedure for collecting a database of texts annotated with coherence relations. *Database documentation*.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, and Mohammad Norouzi. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. In *arXiv preprint ArXiv: 1609.08144*.

Peng Xu, Hamidreza Saghir, Jin Sung Kang, Teng Long, Avishek Joey Bose, Yanshuai Cao, and Jackie Chi Kit Cheung. 2019. A cross-domain transferable neural coherence model. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 678–687. Association for Computational Linguistics.