

A Noisy Context Optimization Approach for Chinese Spelling Correction

Guangwei Zhang^{1,3}, Yongping Xiong¹, Ruifan Li^{2,3*}

¹School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China

²School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China

E-mail: {gwzhang, ypxiong, rli}@bupt.edu.cn

Abstract—The task of Chinese Spelling Correction (CSC) aims to detect and correct Chinese spelling errors. Recently, BERT-based models have dominated the research on CSC. These methods suffer from unsatisfactory performance when dealing with noisy contexts since incorrect characters are often influenced by other typos. In this paper, we propose A Noisy Context Optimization approach for Chinese Spelling Correction (NCO-Spell). Specifically, first, in the pre-training stage, the multi-character masking strategy is proposed to construct multi-typo texts, enabling the model more robust in noisy environments. Furthermore, we increase the size of the confusion set by dynamically updating. Second, in the inference stage, an iterative algorithm is incorporated to correct the wrong characters one by one, and each iteration will gradually reduce the noise. Extensive experiments and detailed analyses on a widely used benchmark demonstrate that NCO-Spell is effective.

I. INTRODUCTION

Chinese contains more than 10,000 characters, of which 3,500 are common characters [1]. Usually, spelling errors could occur. These spelling errors are typically caused by human writing, automatic speech recognition, or optical character recognition systems. As shown in Table I, the first case is that the sentence contains only one typo, and its semantics are similar. The second case contains multiple consecutive different typos, and they are pronounced similarly.

Chinese spelling correction (CSC) task has been proposed to address this problem. One of the mainstream methods is based on language models [2], [3]. Recently, BERT [4] has been proven to be effective in a wide range of applications and has become the fundamental model for CSC tasks, and achieved the best performance [5]–[8]. However, BERT is a masking model that masks characters based on a specific probability. For multi-typo texts, the context of each character contains at least one typo, which results in noisy information. BERT could not effectively deal with this situation.

Basically, Chinese spelling errors are mainly caused by the misuse of phonetically or visually similar characters. Therefore, it is of importance to incorporate the similarity knowledge of characters into the model. To this end, some works utilize confusion sets to incorporate this knowledge [9], [10]. However, confusion sets are usually generated by heuristic rules or human annotations, which have limited coverage and the similarity knowledge of characters is not fully exploited.

*Corresponding author.

TABLE I

EXAMPLES OF CHINESE SPELLING ERRORS. MISSPELLING CHARACTERS ARE MARKED IN RED, AND THE CORRECT CHARACTERS ARE MARKED IN BLUE.

Type	Input	Translation
Wrong	他没学过汉字。	He has never learned a man .
Correct	他没学过汉字。	He has never learned Chinese characters.
Wrong	下次我们一起持分号码?	Shall we hold the number together next time?
Correct	下次我们一起吃饭好吗?	Shall we have dinner together next time?

In this paper, we propose NCO-Spell, a Noisy Context Optimization Approach for CSC task. Firstly, we propose a multi-character-based masking strategy. This strategy simulates multiple adjacent misspellings in the dataset and increases the masking of dynamic confusion sets. We update the uncorrected characters in the training process into the confusion set and obtain a dynamic confusion set. Secondly, we propose a method of iterative inference in the inference stage. Only the wrong character with the highest probability will be corrected each time. The corrected result is used for the next iteration to gradually reduce the influence of noise. Thirdly, we conduct experiments on the widely used benchmark dataset, i.e., SIGHAN15 [3]. The results show that our approach outperforms all compared methods.

The major contributions are highlighted as follows. 1) We propose a masking strategy based on multi-characters, which makes the model more robust to noisy contexts. 2) We propose an iterative inference method to gradually reduce the impact of noise. 3) We conduct extensive experiments to show our method outperforms state-of-the-art baselines.

II. RELATED WORK

CSC is an important and challenging task in natural language processing. It mainly needs to detect typos based on the judgment of the semantics and correct these typos with a full understanding of the context. Most of the early works used statistics and rules for detection and correction and then adopted the credibility of the language model for judgment [3], [11]. These error correction techniques calculate the sentence composition probabilities through local context information and failed to consider the global context information.

Recently, BERT [4] has been proposed, which is a bidirectional language model based on transformer encoders. More and more works use BERT-like models to directly map each

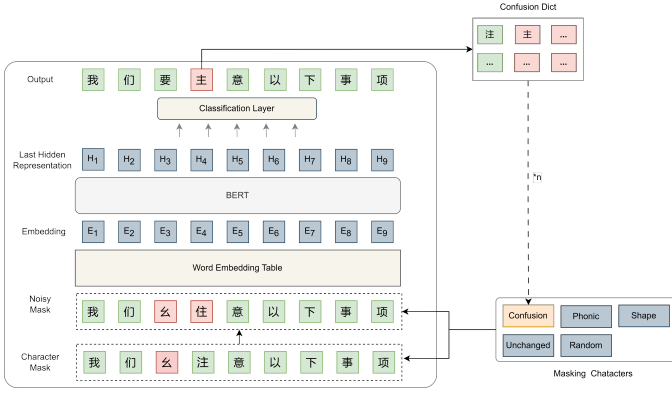


Fig. 1. Framework of NCO-Spell. Take the sentence “我们要注意以下事项” as an example. On the left, this component presents the masking strategy, which forces the model to construct a noisy context for the character “么”. On the right, this component presents the update module for the confusion set, which illustrates how our approach adds new error-prone characters “注-主” into the confusion set.

character in a sentence to the correct character for CSC task. For example, considering that Chinese typos are mostly similar in pronunciation or shape, Hong et al. [12] improved the masking method and added a confidence-similarity filter in the decoding stage. Cheng et al. [5] constructed graphs for the relationship between the phonetic and grapheme of characters to participate in character classification to fuse similar information of characters. Liu et al. [6] used similar characters for masking in the pre-training stage, integrates phonetic and stroke information into the output, and learns more character information. Huang et al. [13] fused phonetics, glyphs, and character information through a gating system in the fine-tuning stage to obtain more character information. Furthermore, Li et al. [14] refined the knowledge representations of pre-trained language models, and guides the model to avoid predicting these common characters through an error-driven way. Li et al. [15] proposed the LEAD framework, which renders the CSC model to learn heterogeneous knowledge from the dictionary in terms of phonetics, vision, and meaning. However, only relying on the previous random masking, BERT could not effectively deal with this situation.

In addition, some works are proposed based on confusion set [3], [5], [6], [10]. Wang et al. [10] proposed a pointer network model guided by a confusion set for Chinese spelling correction. Cheng et al. [5] used a confusion set in the final character classification. Liu et al. [6] used a fixed-size phoneme and grapheme confusion set to obtain character information. The performance of these models heavily relies on the quality of the confusion set.

III. METHODOLOGY

In this section, we describe the proposed approach in detail.

A. Task Formulation

Chinese Spelling Correction (CSC) task can be formalized as a sequence labeling problem. Given a text sequence of n Chinese characters $X = (x_1, x_2, \dots, x_n)$, the goal is to output

TABLE II
PROBABILITY DISTRIBUTION OF DIFFERENT MASKING STRATEGIES.

Strategy	Random	Unchanged	Phonic	Shape	Confusion
PLOME	10%	15%	60%	15%	-
NCO-Spell	10%	15%	50%	15%	10%

$Y = (y_1, y_2, \dots, y_n)$, where X represents the original text containing some error characters, and Y represents the correct text after correction.

B. Multi-character Mask

The framework of our proposed NCO-Spell is shown in Fig. 1. The input of this module is the sequence of embeddings $E = e_1, e_2, \dots, e_n$, where e_i denotes the embedding of character x_i in a given text $X = (x_1, x_2, \dots, x_n)$. Then, E is fed into the transformer encoder and the encoder generates hidden representation matrix $H = (h_1, h_2, \dots, h_n)$ for H , where $h_i \in R^{768}$ is the representation of x_i . Finally, the final output of this module is fed into a one-layer classification network with softmax normalization.

In order to construct the noisy context, after masking the selected characters, we mask the other characters with a certain probability in the sentence to simulate the situation of multi-typo text. When selecting the source of mask characters, we added dynamic confusion set to enhance the model’s learning of error-prone characters. In the following subsections, we will introduce the masking strategy and the dynamic confusion set in the masked content.

Masking Strategy. To make the model robust to the noisy context, we propose a multi-character masking strategy for the pre-training corpus. This method contains two key points:

- 1) **Masking Characters.** Because the frequency of phonetic errors is about twice that of visual errors [16], so in the masking strategy, the proportion of characters that are phonetically related to the original character is large, and we increase the selection of the dynamic confusion set. Finally set the corresponding selection probability of each strategy as shown in Table II.
- 2) **Masking Position.** The contextual disturbance mainly occurs in noisy environments around misspelled characters, so the distance between these masked characters and the original typo is defined as the “masking distance”. In Fig. 1, the masking distance is 1. If there are no characters to mask in the original sentence, the original text will not be masked in any positions.

Dynamic Confusion Set. The initial confusion set is constructed by Wang et al. [9]. When using a confusion set, the performance of the model generally depends heavily on the quality of the confusion set, and it is difficult to find the latest obfuscated characters that are beyond the existing field. As the training iterations progress, the model will continue to generate new errors, and characters will be associated with more error-prone characters, so the size of the confusion set also needs to be continuously updated. The dynamic update of the confusion set includes two points:

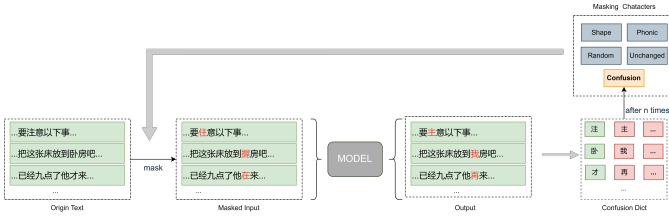


Fig. 2. Illustration of Updating Strategy on the Confusion Set. These are some sentences where the correction fails. For example, “注” is masked to “住” and as the input. After being corrected by the approach, “主” is obtained, which is the uncorrected character. Then our approach temporarily stores correct and uncorrected in the confusion dictionary and updates them into the confusion set after n times of training.

TABLE III
EXAMPLES OF NCO-SPELL RESULTS FOR MULTI-TYPOS.

Type	Text	Translation
Origin	他说「请座, 请座」。	He said "Please take a seat, please take a seat".
Result	他说「请座, 请坐」。	He said "Please take a seat, please sit down".
Origin	有吗吗、弟弟还有我。	Is there any, My brother and me.
Result	有妈妈、弟弟还有我。	Do you have a mother, brother and me?
Origin	下次我们一起持分号码?	Shall we hold the number together next time?
Result	下次我们一起吃分号码?	Shall we have a semicolon next time?

- 1) Selection of Samples. During the training process of the model, there will be correction failures, including changing correct characters to other characters or incorrect characters not being corrected successfully. The result is characters that the model believes fit the context but are actually incorrect, they are the model's confusing characters. Such characters will be updated into the confusion set as new confusion set samples.
- 2) Dynamic Update Method. After obtaining new samples of the confusion set, temporarily store the new samples of the confusion set for several training times in a dictionary, and update the data in the dictionary into the confusion set after this training process. In order to ensure the quality of the confusion set, the confusion character pairs to be updated will be added to the confusion set only after the loss is reduced to a certain threshold. In the subsequent training process, the corresponding characters in the confusion set will be randomly used for masking. The confusion set update strategy is shown in Fig. 2.

Through dynamically updating the confusion set, we obtain more error-prone characters. The initial confusion set contains 4,922 characters, and each character corresponds to 7.8 confusion characters on average. In contrast, the dynamic confusion set contains 6,685 characters, and each character corresponds to 116 confusion characters on average.

C. Iterative Inference

In order to further illustrate the problem of the current model in multi-typo text, we use NCO-Spell to conduct experiments on some sentences containing multi-typoes, and the correction results are shown in Table III.

It can be found that affected by the noise of the context, it is often impossible to completely correct all typos, especially when there are continuous errors, the errors in the middle position are affected by the noise, so they cannot be detected. There are a total of 1100 sentences in the evaluation set of SIGHAN15, of which 123 sentences contain multi-typoes, accounting for about 11%. Therefore, it is necessary to optimize the error correction for multi-typo texts.

The iterative inference method aims to reduce the noise effect in the sentence by correcting only the character with the highest probability in each iteration, putting the corrected sentence as input into the next iteration for correction, and so on until there is no need to correct characters. In this way, on the one hand, the influence of the noisy context can be gradually reduced, on the other hand, the recall rate can be improved, all wrong characters can be corrected as much as possible, and each corrected character is the character with the highest confidence, and each iteration obtains an optimal solution of the sentence. Iterative Inference is model-agnostic and it can be used in other CSC methods to achieve further improvements. The flow chart of the inference algorithm is shown in Fig. 3.

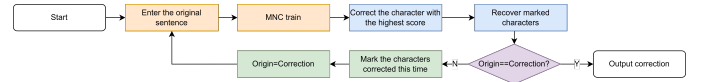


Fig. 3. Flowchart of Iterative Inference. A character will be corrected only once, and only the character with the highest score in the sentence will be corrected every time.

IV. EXPERIMENTS

In this section, we present the details for pre-training NCO-Spell and the fine-tuning results on the most widely used benchmark dataset.

A. Pre-training Settings

- **Dataset.** We use ChineseNlpCorpus¹² as the original pre-training data. We decompose the information into sentences using periods, question marks, exclamation points, and other punctuation marks indicating the end, and after the screening, 110.7 million effective sentences are formed.
- **Parameter Settings.** The composition of the transformer encoder is the same as BERT, with a learning rate of 5e-5. These parameters are set empirically because pre-training is expensive. This article does not pre-train from scratch, but trains based on the pre-trained model released by the PLOME model. The threshold of the loss is set to 4.0, and only when the loss is lower than the threshold, the wrong character will be added to the confusion dictionary, and the confusion set is updated every 5 training epochs.

¹<https://github.com/SophonPlus/ChineseNlpCorpus>

²https://github.com/brightmart/nlp_chinese_corpus

B. Fine-tuning Settings

- **Training Data.** Following Liu et al. [6], the training data is composed of 10K manually annotated samples from SIGHAN [2], [3], [17] and 271K automatically generated samples from Wang et al. [9].
- **Evaluation Data.** We use the latest SIGHAN test dataset [3] as in PLOME to evaluate the proposed approach.
- **Parameter Settings.** Following Liu et al. [6], we set the maximum sentence length to 180, batch size to 32, and the learning rate to $5e-5$. Take the most final training results with the best model in 10 rounds of evaluation effects.
- **Evaluation Metrics.** Following previous work [5], [6], we use the precision, recall, and F1 scores as the evaluation metrics. Besides character-level evaluation, we also report sentence-level metrics on the detection and correction sub-tasks.

C. Baseline Models

We use the following methods for experimental evaluation.

- 1) BERT [4]: The word embedding is used as the input of BERT, and the correct character is predicted based on the mask model.
- 2) PLOME [6]: This method adopts a pre-training masking strategy with misspelling knowledge, and utilizes GRU networks to model such knowledge based on characters' phonics and strokes.
- 3) REALISE [18]: This method captures the semantic, phonetic, and graphic information of the text, and mixes the information in these modalities to predict the output.
- 4) ECOPO [14]: This method refines the knowledge representation of the pre-trained language model, it can be combined with existing CSC methods
- 5) LEAD [15]: This method constructs positive and negative samples, and uses a contrastive learning-based training scheme to improve the representation.
- 6) CoSPA [19]: This method proposes a variable replication mechanism and uses ResNet to mine the glyph information of characters to enhance the visual representation.
- 7) PGBERT [20]: This method proposes a phonology and grapheme augmentation pre-training method.

In addition, we implement PLOME(cfs), which added dynamic confusion set based on PLOME. PLOME(iter) represents the result of inference based on the iteration of PLOME. NCO-Spell(iter) represents the result of inference based on the iteration of NCO-Spell.

V. RESULTS AND ANALYSIS

A. Main Results

The main results can be found in Table IV. From the table, we observe that:

- 1) At the sentence level, PLOME(cfs) has significantly improved compared to PLOME, and the accuracy rate has been greatly improved, indicating that the dynamic confusion set allows the model to have more semantic

TABLE IV

THE PERFORMANCE OF OUR APPROACH AND BASELINE MODELS ON SIGHAN15. THE RESULTS AT THE BOTTOM FOUR ROWS ARE OUR IMPLEMENTATION. ECOPO INDICATES THE IMPROVEMENT OF THE MODEL FOR REALISE. THE SYMBOL '-' MEANS NO RESULTS ARE REPORTED IN THE CORRESPONDING PAPERS.

Method	Character-level (%)						Sentence-level (%)					
	Detection			Correction			Detection			Correction		
	P	R	F	P	R	F	P	R	F	P	R	F
BERT [4]	90.9	84.9	87.8	95.6	81.2	87.8	68.4	77.6	72.7	66.0	74.9	70.2
PLOME [6]	94.5	87.4	90.8	97.2	84.3	90.3	77.4	81.5	79.4	75.3	79.3	77.2
REALISE [18]	-	-	-	-	-	-	77.3	81.3	79.3	75.9	79.9	77.8
ECOPO [14]	-	-	-	-	-	-	77.5	82.6	80.0	76.1	81.2	78.5
LEAD [15]	-	-	-	-	-	-	79.2	82.8	80.9	77.6	81.2	79.3
CoSPA [19]	95.9	88.6	92.1	98.5	85.3	91.4	79.0	82.4	80.7	76.7	80.0	78.3
PGBERT [20]	95.2	87.5	91.2	97.2	85.1	90.8	78.1	82.8	80.4	76.0	80.6	78.2
PLOME(cfs)	87.3	87.0	87.1	97.4	84.7	90.7	79.3	82.2	80.7	77.4	80.2	78.7
NCO-Spell	89.7	86.8	88.2	98.2	85.3	91.3	81.5	82.7	82.1	79.9	81.1	80.5
PLOME(iter)	88.6	88.1	88.3	98.3	86.5	92.0	82.1	83.3	82.7	80.5	81.6	81.0
NCO-Spell(iter)	88.7	88.1	88.4	98.3	86.5	92.0	82.3	83.5	82.9	80.6	81.8	81.2

knowledge and the correct rate of corrected sentences is improved.

- 2) Compared with the PLOME(cfs), the performance of NCO-Spell has improved in most indicators, and the detection and correction indicators at the sentence level have increased by 1.4% and 1.8% respectively. This shows that the pre-training data constructed by the multi-character masking strategy better fit the real data, which is not constructed by the previous masking strategy.
- 3) ECOPO has added an error-driven probability comparison optimization module, REALISE has integrated multi-channel information of characters, but their performance is still not as good as the performance of NCO-Spell, which shows the effectiveness of the multi-character masking strategy.
- 4) Compared with NCO-Spell, the performance of most indicators of NCO-Spell(iter) has been improved, and other indicators have also achieved comparable performance. Compared with PLOME, the performance of most indicators of PLOME(iter) has been improved. Shows the effectiveness of iterative inference.
- 5) The performance of PLOME(iter) and NCO-Spell(iter) is similar, because after iterating the algorithm in the inference stage, the result is less affected by noise, and the enhancement effect of the model on the noisy environment is covered to a certain extent.

B. Results for Multi-typo Subset

To better illustrate the role of NCO-Spell in multi-typo texts, we test on a test set containing multi-typos, which is a subset of SIGHAN15 and contains 123 test data. Table VI presents the results. We observe that after adding the multi-character masking strategy, the sentence-level performance is significantly improved compared to the baseline model PLOME, which demonstrates the effectiveness of the strategy. Compared with NCO-Spell direct inference, after adding iterative inference, the sentence-level detection and correction indicators increased by 6.9% and 6.9% respectively; Compared with PLOME direct inference, after adding iterative inference, the sentence-level detection and correction indicators increased by 8.5% and 6.7% respectively. This verifies the effectiveness and model-agnostic of iterative inference.

TABLE V

THE PERFORMANCE OF OUR APPROACH AND BASELINE MODEL PLOME ON THE MULTI-TYPO SUBSET OF SIGHAN15. ALL THE EXPERIMENTAL RESULTS ARE USING OUR IMPLEMENTATION.

Method	Character-level (%)						Sentence-level (%)					
	Detection-level			Correction-level			Detection-level			Correction-level		
	P	R	F	P	R	F	P	R	F	P	R	F
PLOME	97.1	79.2	87.2	97.0	76.8	85.7	63.7	58.5	61.0	61.1	56.1	58.5
PLOME(iter)	97.1	81.0	88.3	97.1	78.6	86.8	73.6	65.9	69.5	69.1	61.8	65.2
NCO-Spell	98.3	77.5	86.6	96.9	75.1	84.6	67.3	60.2	63.5	62.7	56.1	59.2
NCO-Spell(iter)	97.5	81.0	88.5	97.1	78.6	86.8	74.5	66.7	70.4	70.0	62.6	66.1

TABLE VI

THE PERFORMANCE OF OUR APPROACH AND BASELINE MODEL PLOME ON THE MULTI-TYPO SUBSET OF SIGHAN15. ALL THE EXPERIMENTAL RESULTS ARE USING OUR IMPLEMENTATION.

Method	Character-level (%)						Sentence-level (%)					
	Detection			Correction			Detection			Correction		
	P	R	F	P	R	F	P	R	F	P	R	F
PLOME	97.1	79.2	87.2	97.0	76.8	85.7	63.7	58.5	61.0	61.1	56.1	58.5
PLOME(iter)	97.1	81.0	88.3	97.1	78.6	86.8	73.6	65.9	69.5	69.1	61.8	65.2
NCO-Spell	98.3	77.5	86.6	96.9	75.1	84.6	67.3	60.2	63.5	62.7	56.1	59.2
NCO-Spell(iter)	97.5	81.0	88.5	97.1	78.6	86.8	74.5	66.7	70.4	70.0	62.6	66.1

C. Effects of Different Masking Distance

We suppose that when the noise distribution is denser, it has a greater impact on character correction. Therefore, when we perform sampling replacement for misspelled positions, the masking distance is set to 1/2/3 respectively. The results are shown in Table VII. We observe that the model achieves the best performance when the masking distance is 1. Therefore, it is used as the final selected hyperparameter. This result also validates our hypothesis in Section V-B that previous CSC models perform poorly on multi-typo texts due to the noise around typos.

D. Results of Iterative Inference for Continuous typos

Since the effect of sentences with multi-typos is mainly affected by noise, the data set is divided into data containing continuous and discontinuous errors according to the noise distribution for experiments. The number of sentences is 59 and 64, respectively. The experimental results are shown in Fig. 4.

In sentences containing continuous typos, the performance of the iterative inference method is significantly better than that of the direct inference method. At the third iteration, the performance achieved by the iterative inference method is close to that achieved by the direct inference method and continuously improves performance in constant iterations. Analyzing the reasons for this phenomenon, we argue that in sentences containing continuous typos, continuous typos are

TABLE VII

ABLATION EXPERIMENT RESULTS (SENTENCE-LEVEL METRICS) FOR DIFFERENT MASKING DISTANCES IN MULTIPLE CHARACTER MASKS.

Masking Distance	Whole Set						Multi-typo Set					
	Detection			Correction			Detection			Correction		
	P	R	F	P	R	F	P	R	F	P	R	F
0	79.3	82.2	80.7	77.4	80.2	78.7	63.7	58.5	61.0	61.1	56.1	58.5
1	81.5	82.7	82.1	79.9	81.1	80.5	67.3	60.2	63.5	62.7	56.1	59.2
2	80.0	80.9	80.5	77.9	78.7	78.3	63.4	57.7	60.4	61.6	56.1	58.7
3	80.8	80.9	80.8	78.6	78.7	78.7	64.9	58.5	61.5	60.4	54.5	57.3

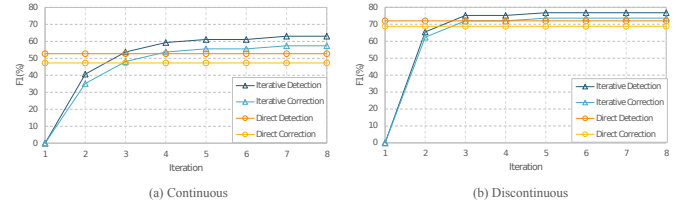


Fig. 4. Iterative inference experiments with sentences containing continuous or discontinuous errors. Our method consistently achieves better performance.

greatly affected by neighboring noise, and it is difficult for direct inference methods to fully correct them.

In sentences containing discontinuous typos, although the performance of iterative inference is improving as the number of iterations increases, the improvement is not obvious compared with the performance of the direct inference method. In a sentence only containing discontinuous typos, the adjacent characters of the wrong character are all correct characters, and the adjacent context noise is less, which is less affected, so the direct inference method can also correct more characters at one time.

At the same time, we can find that sentences containing discontinuous typos have an overall higher score than sentences containing continuous typos, because they are less affected by noise, which also explains that in our multi-character masking strategy, why setting the masking distance to 1.

VI. CONCLUSIONS

We propose NCO-Spell for CSC task in noisy contexts. We construct noisy context through a multi-character masking strategy with dynamic confusion sets in the pre-training stage. Furthermore, in order to improve the performance of character correction in noisy environments, we introduce an iterative inference method. Experimental results on a widely used benchmark show that our NCO-Spell outperforms all the compared baseline models. In the future, we will explore the large language models (LLMs) to enhance the performance of our method.

ACKNOWLEDGMENT

The authors would like to thank the editors and anonymous reviewers for their valuable comments on improving the final version of this article. This work was supported by High Performance Computing Platform of BUPT.

REFERENCES

- [1] H. Wang, B. Wang, J. Duan, and J. Zhang, "Chinese spelling error detection using a fusion lattice lstm," *Transactions on Asian and Low-Resource Language Information Processing*, pp. 1–11, 2021.
- [2] J. Yu and Z. Li, "Chinese spelling error detection and correction based on language model, pronunciation, and shape," in *Proceedings of The Third CIPS-SIGHAN Joint Conference on Chinese Language Processing*, 2014, pp. 220–223.

- [3] Y.-H. Tseng, L.-H. Lee, L.-P. Chang, and H.-H. Chen, "Introduction to sighthan 2015 bake-off for chinese spelling check," in *Proceedings of the Eighth SIGHAN Workshop on Chinese Language Processing*, 2015, pp. 32–37.
- [4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "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)*, 2019, pp. 4171–4186.
- [5] X. Cheng, W. Xu, K. Chen, *et al.*, "SpellGCN: Incorporating phonological and visual similarities into language models for Chinese spelling check," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 871–881.
- [6] S. Liu, T. Yang, T. Yue, F. Zhang, and D. Wang, "Plome: Pre-training with misspelled knowledge for chinese spelling correction," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2021, pp. 2991–3000.
- [7] B. Wang, W. Che, D. Wu, S. Wang, G. Hu, and T. Liu, "Dynamic connected networks for chinese spelling check," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2021, pp. 2437–2446.
- [8] H. Wu, S. Zhang, Y. Zhang, and H. Zhao, "Rethinking masked language modeling for Chinese spelling correction," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, A. Rogers, J. Boyd-Graber, and N. Okazaki, Eds., Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 10743–10756. DOI: 10.18653/v1/2023.acl-long.600. [Online]. Available: <https://aclanthology.org/2023.acl-long.600>.
- [9] D. Wang, Y. Song, J. Li, J. Han, and H. Zhang, "A hybrid approach to automatic corpus generation for chinese spelling check," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 2517–2527.
- [10] D. Wang, Y. Tay, and L. Zhong, "Confusionset-guided pointer networks for chinese spelling check," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 5780–5785.
- [11] J.-F. Yeh, S.-F. Li, M.-R. Wu, W.-Y. Chen, and M.-C. Su, "Chinese word spelling correction based on n-gram ranked inverted index list," in *Proceedings of the Seventh SIGHAN Workshop on Chinese Language Processing*, 2013, pp. 43–48.
- [12] Y. Hong, X. Yu, N. He, N. Liu, and J. Liu, "Faspell: A fast, adaptable, simple, powerful chinese spell checker based on dae-decoder paradigm," in *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, 2019, pp. 160–169.
- [13] L. Huang, J. Li, W. Jiang, *et al.*, "Phmospell: Phonological and morphological knowledge guided chinese spelling check," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2021, pp. 5958–5967.
- [14] Y. Li, Q. Zhou, Y. Li, *et al.*, "The past mistake is the future wisdom: Error-driven contrastive probability optimization for chinese spell checking," in *Findings of the Association for Computational Linguistics: ACL 2022, Dublin, Ireland, May 22-27, 2022*, 2022, pp. 3202–3213.
- [15] Y. Li, S. Ma, Q. Zhou, *et al.*, "Learning from the dictionary: Heterogeneous knowledge guided fine-tuning for chinese spell checking," in *Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, 2022, pp. 238–249.
- [16] C. L. Liu, M. H. Lai, Y. H. Chuang, and C. Y. Lee, "Visually and phonologically similar characters in incorrect simplified chinese words," in *COLING 2010, 23rd International Conference on Computational Linguistics, Posters Volume, 23-27 August 2010, Beijing, China*, 2010.
- [17] S.-H. Wu, C.-L. Liu, and L.-H. Lee, "Chinese spelling check evaluation at sighthan bake-off 2013," in *Proceedings of the Seventh SIGHAN Workshop on Chinese Language Processing*, 2013, pp. 35–42.
- [18] H. Xu, Z. Li, Q. Zhou, *et al.*, "Read, listen, and see: Leveraging multimodal information helps chinese spell checking," in *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021*, 2021, pp. 716–728.
- [19] S. Yang and L. Yu, "Cospa: An improved masked language model with copy mechanism for chinese spelling correction," in *Uncertainty in Artificial Intelligence, Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence, UAI 2022, 1-5 August 2022, Eindhoven, The Netherlands*, 2022, pp. 2225–2234.
- [20] L. Bao, X. Chen, J. Ren, Y. Liu, and C. Qi, "Pgbert: Phonology and glyph enhanced pre-training for chinese spelling correction," in *Natural Language Processing and Chinese Computing*, 2022, pp. 16–28.