ISSN: 2424-8460(O)
2251-2608(P)

Original Research Article

Pinyin-enhanced BERT for Chinese spelling correction with continuous pinyin features and user-specific lexicon adaptation

Fan Cui*, Dan Wang, Zundong Mao
Chizhou University, Chizhou, Anhui, 247100, China

Abstract: Chinese spelling correction (CSC) is an important task in natural language processing (NLP), with applications in text generation, intelligent input methods, and other scenarios. Although many advanced models have improved general-purpose correction performance, two major challenges remain. First, under the dominance of pinyin-based input methods, phonetic similarity errors have become prevalent, yet existing models still struggle to correct them effectively. Second, the presence of low-frequency domain-specific terms (e.g., in medicine or law) makes it difficult for general models to adapt to out-of-domain texts. To address these issues, we propose Pinpin-BERT, a CSC approach that enhances BERT with continuous pinyin features. Leveraging the property that consecutive Chinese characters often have stable pinyin patterns, the model employs a convolutional neural network to extract continuous pinyin features and integrates them with BERT's semantic representations for more accurate correction. In addition, a user-defined lexicon is incorporated to improve domain adaptability. Experimental results on three official benchmark datasets and a self-constructed medical dataset demonstrate that the proposed method achieves consistently strong performance across all evaluations.
Keywords: pinyin; convolutional neural network; Chinese spelling correction

1. Introduction

The Chinese pinyin input method is built around the core logic of mapping pinyin to Chinese characters. Compared with traditional input methods such as handwriting or Wubi typing, it offers higher input efficiency and lower error rates. These advantages are particularly evident on small-screen devices like smartphones, where it provides greater ease of use and a smoother user experience. Moreover, the pinyin input method has a low learning curve and aligns well with general user habits—Its core workflow simply requires typing the pinyin of a character and selecting the desired word from a list of candidates. However, this "Pinyin Input - Candidate Word Selection" interaction model has inherent limitations: users often select incorrect candidates due to inattentive operations, leading to spelling errors. Consequently, in computer-generated Chinese text, most spelling mistakes originate from the misuse of characters with similar or identical pinyin representations.

It is worth noting that mainstream pretrained language models, such as BERT, primarily focus on capturing semantic information in the Chinese Spelling Correction (CSC) task \cite{b1}. As a result, the candidate characters they generate are largely filtered based on semantic similarity, making them less effective at handling spelling errors caused by pinyin similarity. Liu et al. \cite{b2} further highlighted the significance of this issue, showing that approximately 83\% of Chinese spelling errors are related to phonetic (pinyin) similarity, while only 48\% are associated with visual (glyph) similarity—Underscoring the crucial role of phonetic features in CSC. To address this limitation, recent studies have incorporated additional network modules into pretrained language model architectures to integrate phonetic and visual information, aiming to enrich candidate generation

ISSN: 2424-8460(O)
2251-2608(P)

and overcome the semantic bias. For example, SpellGCN \cite{b3} initializes character node representations with BERT and employs two independent Graph Convolutional Networks (GCNs) to model glyph and pinyin similarities within a confusion set. ReaLise \cite{b4} introduces supplementary Gated Recurrent Unit (GRU) and Convolutional Neural Network (CNN) modules to extract phonetic and glyph features, respectively. PHMOSpell \cite{b5} leverages a VGG19 network for glyph feature extraction and a Neural Text-to-Speech (TTS) network for accurate phonetic representation. Although these approaches recognize and exploit the complementary value of glyph and phonetic features in CSC, they still exhibit a key technical shortcoming: existing models lack targeted optimization for pinyin-similar errors, particularly overlooking the modeling and utilization of continuous pinyin features, which are essential for capturing realistic phonetic error patterns in Chinese text.

In addition, Chinese input methods exhibit a remarkable degree of intelligent personalization. When users from different professions input the same pinyin, the input method often suggests distinct candidate words that reflect their professional background—For instance, medical professionals are more likely to see suggestions related to medical instruments and pharmaceuticals, while lawyers receive candidates dominated by legal terminology. The underlying mechanism is that users in different fields have varying domain-specific vocabulary needs. By learning from each user's input history and typing habits, the input method constructs a personalized lexicon that prioritizes vocabulary consistent with the user's professional domain. Similarly, the issue of domain adaptation is prevalent across many natural language processing (NLP) tasks \cite{b6}, and the Chinese Spelling Correction (CSC) task is no exception. Most existing CSC approaches follow a supervised learning paradigm, which limits their generalization ability. Consequently, when the domain of the input text differs from that of the training corpus, correction accuracy tends to decline significantly. To address this problem, this study draws inspiration from the personalization mechanism of input methods and introduces a user-specific lexicon to assist CSC models. Taking a medical professional as an example, when the doctor enters the error-containing text " 需要开具头包拉定分散片 "(need to prescribe Cefradine Dispersible Tablets), the Chinese Spelling Correction (CSC) model trained on general corpora cannot directly correct the error due to its limited generalization ability. At this time, the model will filter the top K candidates for characters with confidence below the threshold. such as the candidate characters " 胞 , 孢 , 苞 , 雹 "(cell,spore,bud,hail) for the wrong character " 包 "(bag), generate multiple groups of candidate sentences, and then determine the candidate word as " 孢 "(spore) based on the medical terminology " 头孢拉定 "(cefradine) in the user's personalized dictionary.

To address these challenges, we propose a pinyin-enhanced BERT language model and introduce a user-specific lexicon. Our main contributions are summarized as follows:

We propose Pinyin-BERT, an enhanced BERT model that integrates continuous pinyin features. A convolutional neural network (CNN) module is employed to extract continuous pinyin representations from Chinese character sequences, which are then fused with BERT's semantic features. This design strengthens the model's ability to correct phonetic-similarity errors and compensates for the limitations of relying solely on semantic information.

We design a domain-personalized lexicon–assisted strategy. When the model's confidence in a predicted character falls below a threshold, it generates the top-K candidate characters and corresponding candidate sentences. By matching these candidates against a user's domain-specific lexicon (e.g., a medical dictionary), the model selects the sentence containing the largest number of domain-relevant terms as the final correction result, thereby improving domain adaptability

ISSN: 2424-8460(O)
2251-2608(P)

2. Related work

In recent years, the task of Chinese Spelling Correction (CSC) has received widespread attention in the research community. It is important to note that CSC differs significantly from Chinese Grammatical Error Correction (GEC) [7]: CSC focuses exclusively on character-level detection and correction, with the core objective being the replacement of incorrect characters, whereas GEC addresses a broader range of errors, including sentence-level issues such as deleting redundant characters or inserting missing ones. Although the scope of CSC is more narrowly defined, focusing primarily on character replacement, designing methods that are both efficient and accurate remains a significant challenge. The development of Chinese Spelling Correction (CSC) techniques can be divided into several stages. Early research primarily relied on rule-based unsupervised methods, where researchers designed various customized rules to handle different types of spelling errors [8,9], typically supported by n-gram language models. With the advancement of neural network technologies, CSC modeling gradually shifted toward data-driven approaches. On one hand, some studies formulated CSC as a sequence labeling task [3], employing bidirectional LSTMs as the core framework to learn contextual dependencies within character sequences for correction. On the other hand, the "copy mechanism"from sequence-to-sequence (Seq2Seq) frameworks has also been applied to CSC [10]. This mechanism constrains the model to copy candidate corrections only from a predefined"error-correct character confusion set," thereby reducing invalid predictions. With the breakthroughs of large-scale pretrained language models such as BERT in natural language processing [1], numerous BERT-based CSC models have been proposed, significantly improving task performance. For example, FASpell [11] by Hong et al. uses a language model as a candidate character generator and selects the optimal correction candidates based on a "confidence-similarity curve." Soft-Masked-BERT [12] adopts a two-stage architecture, first detecting errors using a GRU-based module and then correcting them with a BERT module. However, it is important to note that the original pretrained BERT model focuses solely on semantic features of characters and neglects the glyph and phonetic features of Chinese characters—Both of which are key factors contributing to spelling errors. To address this limitation, subsequent studies have designed specialized network architectures to incorporate glyph and phonetic features into pretrained models, further enhancing correction accuracy and specificity [3-5]. Recently, CSC research has begun to focus on the issue of "noise interference caused by incorrect characters"and has proposed corresponding optimization strategies. Wang et al. [13] introduced the Dynamic Connection Network (DCN), which reduces the disruption of noise on the coherence of character sequences by learning dependencies between adjacent Chinese characters, thereby preventing the model from generating incoherent sentences. Liu et al. [14] proposed a Chinese spelling correction method based on the Rewriting Language Model (ReLM), reframing the correction task from traditional sequence labeling to sentence-level rewriting. Wang et al. [15] presented a self-supervised CSC approach, training solely on error-free data and incorporating a denoising decoding strategy (D^2C) to enhance correction accuracy. Although the aforementioned methods improve CSC model performance from different perspectives, two key limitations remain. First, in current input scenarios dominated by pinyin-based input methods, most spelling errors arise from the misuse of characters with similar pinyin. Second, models trained on general-domain corpora have limited generalization ability and cannot effectively handle correction needs in specific domains, such as medical or legal texts.

3. The pinyin-BERT model

With the widespread use of pinyin input methods, modern document spelling errors are increasingly caused

ISSN: 2424-8460(O)
2251-2608(P)

by the misuse of characters with similar pinyin. Therefore, in pinyin-based input scenarios, it is necessary to design a model capable of more accurately generating candidate characters with similar pronunciations. In this paper, we propose a pinyin-enhanced BERT approach. As shown in Table 1, ignoring tones, we collected pinyin information for 7,331 commonly used Chinese characters. It can be observed that a single pinyin typically corresponds to multiple characters—For example, the pinyin "ji"corresponds to as many as 121 different characters, and on average, each pinyin maps to more than 18 characters. Clearly, converting a single pinyin to the correct Chinese character is highly ambiguous, making accurate conversion challenging.

Table 1. Statistics on the correspondence between Chinese characters and pinyin.

	Number	Max	Min	Avg
Pinyin	402	121	1	18.2

However, when multiple consecutive pinyin syllables are available, the model can be more confident in converting them into the correct Chinese characters. For instance, in the word " 鼻严 "(" 严 " being a typo), it is difficult to correct " 严 " (strict) based solely on its pinyin, since many characters share the pronunciation "yan", such as " 研 , 炎 , 眼 , 艳 , 烟 , 演 " (research,inflammation,eye,gorgeous,smoke, perform), and so on. However, if the two syllables are considered together as "bi yan", the candidate words are limited to " 鼻炎 " (rhinitis) or " 闭眼 " (close eye). Moreover, the longer the consecutive pinyin sequence considered, the fewer candidate words remain, allowing the model to correct errors with higher probability. Motivated by this characteristic of Chinese characters, we propose a pinyin-enhanced BERT-based CSC model. The framework of the model is illustrated in **Figure 1**.

3.1. Task definition

The objective of the Chinese Spelling Correction (CSC) task can be formally described as follows: given an input text sequence $X=\{x_1, x_2, ..., x_n\}$, the goal of CSC is to automatically correct misspelled characters and generate the correct target sequence $Y=\{y_1, y_2, ..., y_n\}$ where x_i and y_i $(1 \le i \le N)$ represent individual Chinese characters, and N denotes the total number of characters in the sequence.

3.2. Model architecture

As shown in Figure 2, the model mainly consists of a BERT encoder and a convolutional neural network (CNN). First, semantic features of each character, denoted as , are extracted using the BERT language model.

$$h_i=BERT(e_i) \tag{1}$$

Here, e_i represents the input vector to the BERT model. Subsequently, character-level pinyin features c_i are obtained through a convolutional neural network, which can be expressed as follows:

$$c_i=W^t \bullet Concat\ (Conv(p_{i-1}, p_i, p_{i+1})) \tag{2}$$

Here, Conv denotes the convolution operation, p_i represents the pinyin vector of the i-th character, Concat indicates the concatenation operation, and W_t is a learnable linear layer that transforms the dimensionality of the concatenated convolutional features to match that of the semantic features, facilitating feature fusion. Finally, the pinyin features and semantic features are summed and normalized to obtain the final character representation O_i, which can be expressed as follows:

$$O_i=LayerNorm\ (c_{i\,+}\, h_i) \tag{3}$$

Layer norm denotes the layer normalization operation. As mentioned previously, model parameters trained on general corpora often fail to adapt well to domain-specific data. However, we observe that, although the model may not directly correct errors, the correct result often appears among the top few predicted characters. As shown in the left part of Figure 1, if the model confidence P (i.e., the probability of the model predicting the

ISSN: 2424-8460(O)
2251-2608(P)

first candidate character) exceeds a threshold η (set to 0.85 in our experiments), only the first candidate character is retained. If $P < \eta$, the model retains the top K most probable candidate characters. The final predicted result of the model can be expressed as follows:

$$\hat{y}_i = \begin{cases} arg\max\left(W^o \cdot O_i\right), & \text{if} P > \eta \\ \text{Top}_k\left(\text{softmax}\left(W^o \cdot O_i\right)\right), & \text{if} P \le \eta \end{cases} \tag{4}$$

To enable the convolutional network to learn Pinyin features more efficiently, the model's final optimization loss function is composed of two components:

$$\mathcal{L}_c = -\sum_{i=1}^{N} y_i \log \hat{y}_i \tag{5}$$

$$\mathcal{L}_p = -\sum_{i=1}^{N} y_i \log\left(W^p \cdot c_i\right) \tag{6}$$

$$\mathcal{L} = \mathcal{L}_c + \lambda \mathcal{L}_p \tag{7}$$

Here, W_p denotes the output layer for pinyin, $\mathcal{L}_i$ represents the model's final prediction loss, and $\mathcal{L}_p$ represents the pinyin prediction loss. The overall loss of the model, $\mathcal{L}$ is obtained by summing these two components, with λ as a hyperparameter.

3.3. User dictionary filtering

If a sentence contains n ($0 \le n \le N$) characters with prediction confidence lower than P, the model will generate Kn candidate sentences. Suppose there exists a user dictionary $M \in \{t_1, t_2, ..., t_T\}$, where t_i ($0 \le i \le T$) denotes a Chinese word. The user dictionary filtering works as follows: each of the Kn candidate sentences is segmented, and the number of words that appear in the dictionary is counted. The sentence containing the largest number of dictionary words is selected as the final prediction. If multiple sentences contain the same maximum number of dictionary words, the one with the highest ranking among them is chosen as the final prediction.

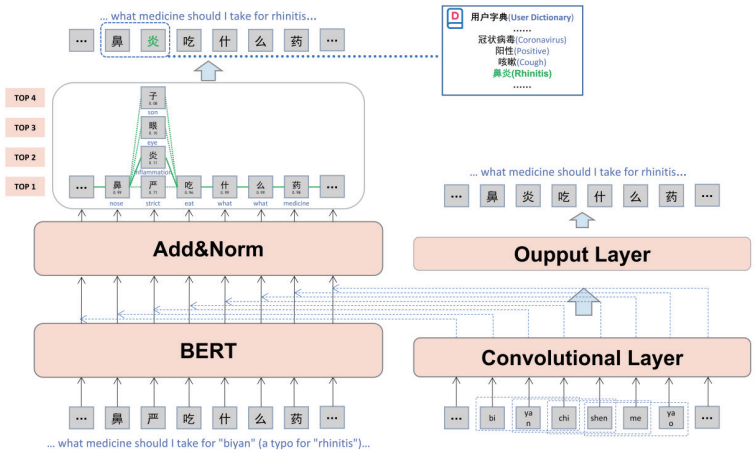

Figure 1. The continuous pinyin feature-enhanced BERT model structure.

ISSN: 2424-8460(O)
2251-2608(P)

4. Experiments

4.1. Evaluation metrics

The model is evaluated at the sentence level using Accuracy (Acc), Recall, Precision, and F1-Score, assessing both detection and correction aspects [3].

4.2. Dataset

The official SIGHAN training data and the pseudo data generated by Wang et al. [10] is used as the training and test sets. In addition, to evaluate the model's domain adaptability, we collected and constructed a medical-domain test set of 1,000 sentences, reflecting the consecutive errors commonly caused by users when using pinyin input. The statistics of the dataset are presented in **Table 2**.

Table 2. Medical field test set statistics.

Statistic	Number of sentences
Min sentence length	4
Max sentence length	36
Avg sentence length	10.79
Error/Truth	1000/1000

4.3. Comparison methods

RoBERTa [1]: This model fine-tunes the RoBERTa-base on the CSC training data.

REALISE [4]: This model uses a GRU network to extract phonetic features and a convolutional network to capture character shape features. Finally, it predicts the output through adaptive fusion based on the Transformer.

DCN [13]: This model employs a unique dynamic connection network that generates K paths (where K represents the number of candidate words and n denotes the sentence length) during the model output phase. It then scores the paths using the dynamic connection network and selects the optimal one.

4.4. Results

The experimental results of the proposed pinyin-enhanced BERT method, pinyinBERT, on three official test sets are shown in Table 3. Compared to directly fine-tuning the RoBERTa model, pinyinBERT achieves an F1 score improvement of 4.7%, 8.4%, and 5.5% at the detection level, respectively. At the correction level, the F1 score increases by 2.8%, 3%, and 1.5%, respectively. Although some errors in the official datasets are due to the misuse of visually similar characters, most spelling errors are caused by the misuse of similar pinyin, a common characteristic of Chinese characters. There are also cases where both factors contribute, such as "郎" and "朗," which are similar in both shape and pinyin. Misuse of characters that are only similar in shape accounts for a small portion of the errors. Therefore, this method incorporates pinyin features into the character model through a convolutional network, aiding in correction and enhancing the model's prediction capability. Especially in character-level evaluation metrics, pinyinBERT significantly outperforms the baseline RoBERTa model. Although the official dataset emphasizes shape-related errors in handwriting scenarios, the pinyin-enhanced design of pinyinBERT brings its overall performance close to the latest models in the Chinese spelling correction field. This indicates that the model can specifically address phonetic errors in pinyin input scenarios while effectively adapting to shape-related errors, demonstrating the generality and effectiveness of the method.

To further explore the performance of the proposed method on texts from other domains, we constructed a medical test set with 1,000 entries based on the characteristics of spelling errors caused by input methods. The medical professional dictionary was selected from the "National Health and Family Planning Commission's

ISSN: 2424-8460(O)
2251-2608(P)

Common Clinical Medical Terms, Pages 1-108, 2019" document. The experimental results are shown in Table 4. As can be seen, compared to the direct 0utput of the PinyinBERT model, after incorporating the specialized vocabulary list, the model's performance improved by approximately 2% across all metrics. This demonstrates that the simple approach of adding a user-defined dictionary can effectively enhance the model's generalization ability.

When the model's confidence score P for predicting a character is below a threshold, the K most probable candidate words before that character are saved. If the sentence contains n low-confidence characters, the candidate words are combined to generate Kn possible sentences. Therefore, as K increases, the number of candidate sentences grows. We set K = 4 consistently across the three official test sets and the medical test set. This setting strikes a balance between error correction and inference efficiency: it ensures high-probability candidates are covered to avoid missing the correct character, while also preventing an excessive increase in candidate sentences when K is too large (e.g., K = 5 or K = 6), which would escalate computational costs and slow down inference speed. Furthermore, the confidence threshold also significantly impacts the number of candidate sentences. As a result, we conducted a study on the effect of threshold size on sentence generation quantity, with the results shown in Table 5. From the table, it can be observed that the number of candidate sentences for the medical test set is significantly higher than that of the official test sets. This is because the model, trained on general data, tends to have lower confidence in character predictions for out-of-domain medical text, leading to a larger number of candidate sentences generated under the same threshold.

To further explore the impact of the confidence threshold P on model performance, this study uses the Medical test set as the experimental object and adopts the detection-level F1 score (D-F1) and correction-level F1 score (C-F1) as the core evaluation metrics. A controlled variable experiment was conducted, and the results are shown in Figure 2. As the confidence threshold decreases, both D-F1 and C-F1 show an overall upward trend. However, there are significant stagebased differences in performance improvement: when the threshold is above 0.85, each 0.05 decrease in the threshold (e.g., from 0.95 to 0.9, and from 0.9 to 0.85) leads to a noticeable increase in both D-F1 and C-F1 (with an average improvement of 1.2-1.5 percentage points). However, when the threshold falls below 0.85, the performance gain diminishes significantly, and the marginal benefit continues to decrease, with performance becoming relatively flat or even fluctuating. This indicates that lowering the

Table 3. The performance of our model and all baseline models on SIGHAN test sets.

	Character Level						Sentence Level							
Method	Detection Level			Correction Level				Detection Level				Correction Level		
SIGHAN2013	D-P	D-R	D-F	C-P	C-R	C-F	Acc	D-P	D-R	D-F	Acc	C-P	C-R	C-F
RoBERT	80.5	88.0	84.1	98.0	86.5	91.9	77.3	85.1	76.9	80.8	75.6	83.6	76.0	79.6
REALISE	-	-	-	-	-	-	82.7	88.6	82.5	85.4	81.4	87.2	81.2	84.1
DCN	-	-	-	-	-	-	-	86.8	79.6	83.0	-	84.7	77.7	81.0
PinyinBERT	78.6	92.3	84.9	97.9	95.4	96.6	82.5	86.6	82.2	84.3	80.7	84.6	80.3	82.4
SIGHAN2014	D-P	D-R	D-F	C-P	C-R	C-F	Acc	D-P	D-R	D-F	Acc	C-P	C-R	C-F
RoBERT	82.6	78.0	80.2	96.9	75.9	85.1	74.1	61.2	67.3	64.1	73.6	60.3	66.4	63.2
REALISE	-	-	-	-	-	-	78.4	67.8	71.5	69.6	77.7	66.3	70.0	68.1
DCN	-	-	-	-	-	-	-	67.4	70.4	68.9	-	65.8	68.7	67.2
PinyinBERT	79.3	79.3	79.3	98.5	88.9	93.5	77.0	64.9	69.6	67.2	76.5	64.0	68.7	66.2
SIGHAN2015	D-P	D-R	D-F	C-P	C-R	C-F	Acc	D-P	D-R	D-F	Acc	C-P	C-R	C-F
RoBERT	**86.9**	87.3	87.1	95.1	82.0	88.1	82.9	73.2	80.4	76.7	81.7	71.0	78.0	74.5
REALISE	-	-	-	-	-	-	**84.7**	**77.3**	**81.3**	**79.3**	**84.0**	**75.9**	**79.9**	77.8
DCN	-	-	-	-	-	-	-	77.1	80.9	79.0	-	74.5	78.2	76.3
PinyinBERT	84.2	**87.2**	85.7	**96.4**	**90.9**	**93.6**	84.6	76.8	81.2	78.9	83.2	74.0	78.2	76.0

ISSN: 2424-8460(O)
2251-2608(P)

threshold further no longer provides meaningful performance gains. Based on a comprehensive trade-off between "performance and efficiency," the confidence threshold is ultimately set to 0.85. At this threshold, the D-F1 and C-F1 values are close to their peak performance, while also avoiding the inference efficiency loss caused by lower thresholds, effectively balancing model accuracy and real-time requirements in practical applications.

Table 4. Experimental results of pinyinBERT on medical test set.

Method	Detection Level				Correction Level			
	Acc	D-P	D-R	D-F	Acc	C-P	C-R	C-F
RoBERT	30.5	26.3	33.7	29.5	21.0	19.2	25.3	21.8
pinyinBERT	40.0	36.1	41.2	38.5	38.1	34.6	39.9	36.1
PinyinBERT +userDict	42.1 (+2.1)	38.3 (+2.2)	43.8 (+2.6)	40.9 (+2.4)	40.0 (+1.9)	36.2 (+1.6)	41.5 (+1.6)	38.7 (+1.6)

Table 5. Placement confidence P and average number of generated candidate sentences for small and large models.

P	SIGHAN2013	SIGHAN2014	SIGHAN2015	Medical
0.95	2.7	3.5	5.3	7.5
0.9	3.3	4.1	6.3	10.1
0.85	3.6	5.2	7.0	18.3
0.8	5.9	6.9	8.1	29.8
0.75	8.2	10.2	13.9	41.9
0.7	10.3	13.2	17.3	66.4
0.65	15.3	17.0	21.6	106.3
0.6	20.1	23.4	33.4	180.7

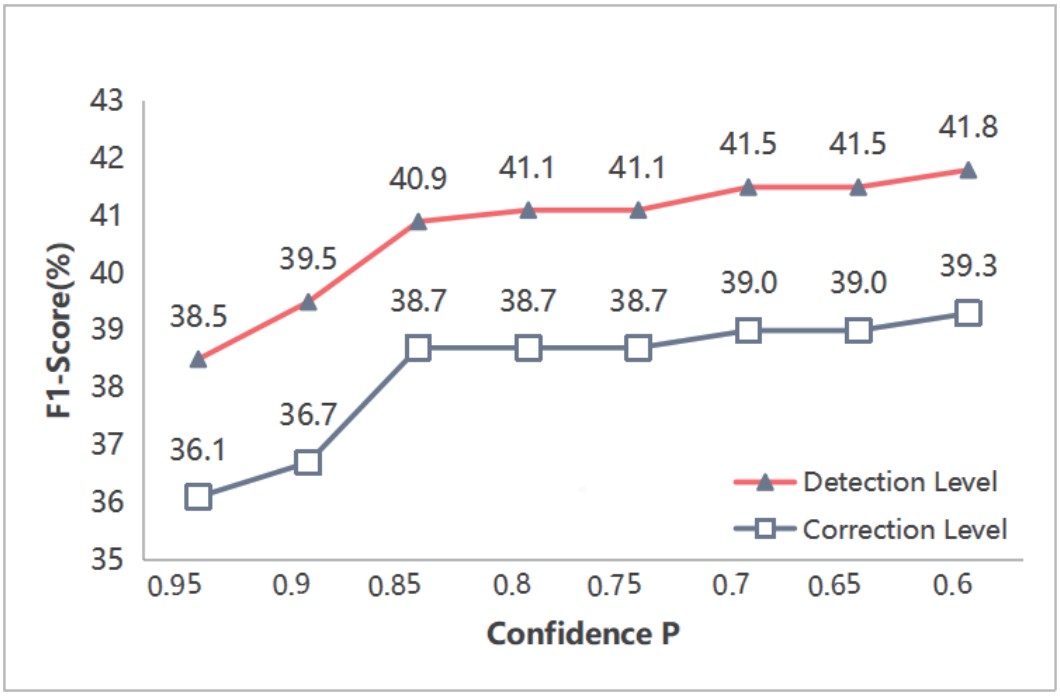

Figure 2. The impact of the confidence threshold P on model performance.

5. Conclusion

This paper proposes a Pinyin-enhanced BERT-based CSC model, tailored to address spelling errors caused by input methods in real-world scenarios. The model utilizes a convolutional network to extract continuous

ISSN: 2424-8460(O)
2251-2608(P)

Pinyin features, which assist the model in performing corrections. Experimental results demonstrate that incorporating Pinyin features into the language model significantly improves its performance. Additionally, due to the domain adaptation challenges in CSC tasks, inspired by the use of user dictionaries in input methods, the model's prediction results are adjusted by adding a specialized dictionary at the model's output. The experimental results show that this simple approach enhances the CSC model's domain adaptation capabilities to a certain extent.

About the author

Fan Cui (1997-), male, Han Chinese, from Chizhou, Anhui Province, holds a master's degree. His research focuses on artificial intelligence, natural language processing, and Chinese spelling correction.fcui@czu.edu.cn.

References

[1] Y. M. Cui, W. X. Che, T. Liu, et al., Pre-training with whole word masking for Chinese BERT, IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 3504–3514, 2021.

[2] C. L. Liu, M. H. Lai, K. W. Tien, et al., Visually and phonologically similar characters in incorrect Chinese words: Analyses, identification, and applications, ACM Transactions on Asian Language Information Processing, vol. 10, no. 2, p. 10, 2011.

[3] X. Cheng, W. Xu, K. Chen, et al., SpellGCN: Incorporating phonological and visual similarities into language models for Chinese spelling check, arXiv preprint arXiv:2004.14166, 2020.

[4] H. D. Xu, Z. Li, Q. Zhou, et al., 'Read, listen, and see: Leveraging multimodal information helps Chinese spell checking, arXiv preprint arXiv:2105.12306, 2021.

[5] L. Huang, J. Li, W. Jiang, et al., PHMOSpell: Phonological and morphological knowledge guided Chinese spelling check, in Proc. 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int. Joint Conf. on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 5958–5967.

[6] F. Boudin, A. Aizawa, Unsupervised domain adaptation for keyphrase generation using citation contexts, arXiv preprint arXiv:2409.13266, 2024.

[7] M. Qiu, Q. Gao, L. Yang, et al., Chinese Grammatical Error Correction: A Survey, arXiv preprint arXiv:2504.00977, 2025.

[8] S. H. Wu, C. L. Liu, L. H. Lee, Chinese spelling check evaluation at SIGHAN bake-off 2013, in Proc. Seventh SIGHAN Workshop on Chinese Language Processing, 2013, pp. 35–42.

[9] L. C. Yu, L. H. Lee, Y. H. Tseng, et al., Overview of SIGHAN 2014 bake-off for Chinese spelling check, in Proc. Third CIPS-SIGHAN Joint Conf. on Chinese Language Processing, 2014, pp. 126–132. [10] D. Wang, Y. Song, J. Li, et al., A hybrid approach to automatic corpus generation for Chinese spelling check, in Proc. 2018 Conf. on Empirical Methods in Natural Language Processing, 2018, pp. 2517–2527.

[10] Y. Hong, X. Yu, N. He, et al., FASPell: A fast, adaptable, simple, powerful Chinese spell checker based on DAE-decoder paradigm, in Proc. 5th Workshop on Noisy User-generated Text (W-NUT 2019), 2019, pp. 160–169.

[11] S. Zhang, H. Huang, J. Liu, et al., Spelling error correction with softmasked BERT, arXiv preprint arXiv:2005.07421, 2020.

[12] B. Wang, W. Che, D. Wu, et al., Dynamic connected networks for Chinese spelling check, in Findings of the

ISSN: 2424-8460(O)
2251-2608(P)

Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, pp. 2437–2446.

[13] L. Liu, H. Wu, H. Zhao, Chinese spelling correction as rephrasing language model, in Proc. AAAI Conf. on Artificial Intelligence, vol. 38, no. 17, pp. 18662–18670, 2024.

[14] L. Jiang, H. Wu, H. Zhao, et al., Chinese spelling corrector is just a language learner, in Findings of the Association for Computational Linguistics ACL 2024, 2024, pp. 6933–6943.