Yinghao LI, Heyan HUANG, Baojun WANG, Yang GAO, 2025. DRMSpell: dynamically reweighting multimodality for Chinese spelling correction. *Frontiers of Information Technology & Electronic Engineering*, 26(3):354-366
https://doi.org/10.1631/FITEE.2300816

DRMSpell: dynamically reweighting multimodality for Chinese spelling correction

Key words: Chinese spelling correction; Multimodality; Masking strategy

Yang GAO
E-mail: gyang@bit.edu.cn
ORCID: https://orcid.org/0000-0002-2422-0548
Yinghao LI
E-mail: yhli@bit.edu.cn
ORCID: https://orcid.org/0000-0002-9439-8544

Motivation

1. Current confusion set methods underutilize phonological and visual information, prompting the need for a more effective approach to character sequence correction.

2. Previous methods inadequately utilized multimodal information by simply summing or fusing modalities without considering their interdependencies, leading to suboptimal performance in Chinese spelling correction tasks.

Main idea

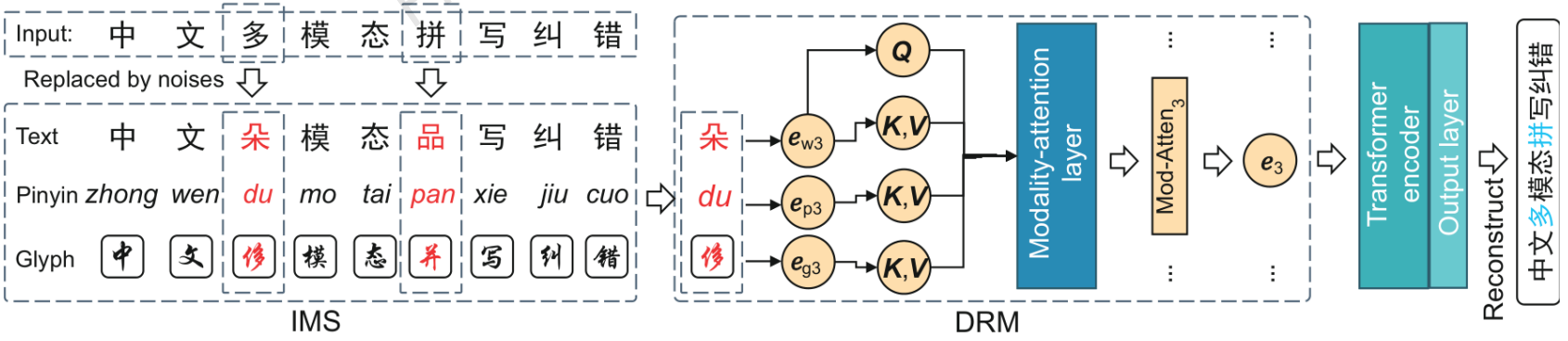
1. A novel pretrained language model called DRMSpell is proposed to build relationships among textual, phonological, and visual modalities, where DRM is specially designed to dynamically extract information by reweighting different modalities.

2. A masking strategy named IMS is designed for multimodal interactions, which helps DRM to model the relationships among the different modalities.

3. DRMSpell achieves the state-of-the-art performance on three benchmark datasets and an OCR dataset.

Method

1. The input sentence ``中文多模态拼写纠错" is transformed to text, pinyin and glyph modality by the input layer. Characters ``多" and ``拼" are chosen to be masked, of which the modalities are replaced independently by independent-modality masking strategy (IMS). Taking the character ``多" as an example, the DRM module dynamically reweights the three modalities. The Transformer encoder and output layer process the final input embedding to predict the correct sentence. The red denotes masking tokens and the blue denotes reconstructed tokens.



Method (Cont'd)

2. The character ``很(hen) '' is chosen to be masked. While confusion set based masking strategy (CMS) selects ``狠(hen) '' as the masking token for the three modalities, IMS selects the three characters independently for the three modalities.

Masking strategy	Masking token	Textual modality	Phonological modality	Visual modality
CMS	狠(hen)	狠	hen	狠
IMS	狠(hen) 跟(gen) 很(hen)	狠	gen	很

Method (Cont'd)

3. Self attention computes attention only once among all the characters in a sequence. The formulation of self-attention on an *n*-length sequence with only word embedding is defined as follows:

$$\text{Self-Atten}_X = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}$$

$$\begin{cases} \boldsymbol{Q} = \boldsymbol{E}_{\mathrm{w}}\boldsymbol{W}_{\boldsymbol{Q}} \in \mathbb{R}^{n\times d_k}, \\ \boldsymbol{K} = \boldsymbol{E}_{\mathrm{w}}\boldsymbol{W}_{\boldsymbol{K}} \in \mathbb{R}^{n\times d_k}, \\ \boldsymbol{V} = \boldsymbol{E}_{\mathrm{w}}\boldsymbol{W}_{\boldsymbol{V}} \in \mathbb{R}^{n\times d_k}, \end{cases}$$

Method (Cont'd)

4. DRM computes the attention called modality-attention (Mod-Atten) among the three modalities for each character in *X*. Modality-attention will be applied *n* times in a sequence with a length of *n*. The modality-attention of the *n*-length sequence *X* with input embeddings {$\boldsymbol{E}_{\mathrm{w}}$; $\boldsymbol{E}_{\mathrm{g}}$; $\boldsymbol{E}_{\mathrm{p}}$} is defined as follows:

$$\begin{aligned}\text{Mod-Atten}_X = (&\text{Mod-Atten}_1; \\ &\text{Mod-Atten}_2; ...; \text{Mod-Atten}_n),\end{aligned}$$

$$\text{Mod-Atten}_i = \text{softmax}\left(\frac{\boldsymbol{Q}_i^{\mathrm{m}}\left(\boldsymbol{K}_i^{\mathrm{m}}\right)^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}_i^{\mathrm{m}}$$

$$\begin{cases}\boldsymbol{Q}_i^{\mathrm{m}} = \boldsymbol{E}_i^{\mathrm{m}}\boldsymbol{W}_{\boldsymbol{Q}_i}^{\mathrm{m}} \in \mathbb{R}^{3\times d_k}, \\ \boldsymbol{K}_i^{\mathrm{m}} = \boldsymbol{E}_i^{\mathrm{m}}\boldsymbol{W}_{\boldsymbol{K}_i}^{\mathrm{m}} \in \mathbb{R}^{3\times d_k}, \\ \boldsymbol{V}_i^{\mathrm{m}} = \boldsymbol{E}_i^{\mathrm{m}}\boldsymbol{W}_{\boldsymbol{V}_i}^{\mathrm{m}} \in \mathbb{R}^{3\times d_k},\end{cases}$$

$$\boldsymbol{E}_i^{\mathrm{m}} = (\boldsymbol{e}_{\mathrm{w}_i}^{\mathrm{T}}; \boldsymbol{e}_{\mathrm{g}_i}^{\mathrm{T}}; \boldsymbol{e}_{\mathrm{p}_i}^{\mathrm{T}}) \in \mathbb{R}^{3\times d_k},$$

Major results

Performance of our method, SOTA, and baseline models on the SIGHAN15 test dataset (%)

Model	Character level						Sentence level							
	D-P	D-R	D-F1	C-P	C-R	C-F1	D-Acc	D-P	D-R	D-F1	C-Acc	C-P	C-R	C-F1
PLOME (Liu et al., 2021)	94.5	87.4	90.8	97.2	84.3	90.3	–	77.4	81.5	79.4	–	75.3	79.3	77.2
MLM-phonetics (Zhang RQ et al., 2021)	–	–	–	–	–	–	–	77.5	83.1	80.2	–	74.9	80.2	77.5
ReaLiSe (Xu et al., 2021)	–	–	–	–	–	–	84.7	77.3	81.3	79.3	84.0	75.9	79.9	77.8
MDCSpell (Zhu et al., 2022)	–	–	–	–	–	–	–	80.8	80.6	80.7	–	78.4	78.2	78.3
CoSPA (Yang SJ and Yu, 2022)	**95.9**	88.6	92.1	**98.5**	85.3	91.4	–	79.0	82.4	80.7	–	76.7	80.0	78.3
ECOPO (Li YH et al., 2022)	–	–	–	–	–	–	85.0	77.5	82.6	80.0	84.2	76.1	81.2	78.5
NM (Yang HY, 2023)	–	–	–	–	–	–	–	78.3	82.1	80.1	–	77.3	81.0	79.1
CL (Zhang D et al., 2023)	–	–	–	–	–	–	80.9	**85.8**	75.4	80.3	79.3	**84.7**	73.0	78.4
ECSpell (Lv et al., 2023)	–	–	–	–	–	–	86.9	82.2	80.2	81.2	86.1	80.5	78.6	79.5
DORM (Liang et al., 2023)	–	–	–	–	–	–	–	77.9	**84.3**	81.0	–	76.6	**82.8**	79.6
BERT*	92.5	85.3	88.8	96.0	81.9	88.4	83.5	74.3	79.2	76.7	82.1	71.5	76.3	73.8
ChineseBert*	93.2	88.2	90.6	95.6	84.2	89.6	84.1	75.0	81.3	78.0	83.5	72.4	78.6	75.4
DRMSpell	93.9	**90.5**	**92.2**	97.3	**88.0**	**92.4**	**87.2**	79.0	83.7	**81.3**	**86.6**	79.3	81.9	**80.6**

Major results (Cont'd)

Overall performance of models on the OCR dataset

Model	Character level						Sentence level							
	D-P	D-R	D-F1	C-P	C-R	C-F1	D-Acc	D-P	D-R	D-F1	C-Acc	C-P	C-R	C-F1
FASPell (Hong et al., 2019)	–	–	–	–	–	–	18.6	78.5	18.6	30.1	17.4	**73.4**	17.4	28.1
BERT*	92.6	84.4	88.3	75.3	63.5	68.9	77.5	84.9	76.6	80.6	58.9	63.3	57.2	60.1
ChineseBert*	**95.1**	79.9	86.8	**79.6**	63.5	70.6	74.8	**87.4**	73.9	80.1	59.7	68.8	58.1	63.0
DRMSpell	92.4	**85.7**	**88.9**	79.3	**68.0**	**73.2**	**79.5**	86.3	**78.9**	**82.4**	**63.9**	68.5	**62.6**	**65.4**

Major results (Cont'd)

Ablation study on DRM and different masking strategies at the sentence level

Setting	D-P	D-R	D-F1	C-P	C-R	C-F1
DRMSpell (DRM + IMS)	79.0	83.7	81.3	79.3	81.9	80.6
DRMSpell (IMS)	75.8	80.1	77.9	74.8	80.6	77.6
DRMSpell (DRM + CMS)	77.2	82.9	79.9	75.9	80.2	78.0
DRMSpell (CMS)	75.9	79.2	77.5	75.0	79.8	77.3

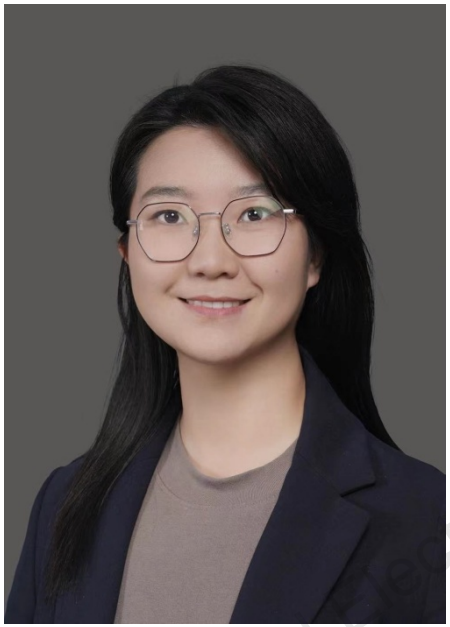
Conclusions

1. IMS enhances the model by allowing independent masking of different modalities, facilitating richer interactions and improving learning during training.

2. DRM dynamically adjusts the significance of each modality, optimizing the model's focus on relevant information for error correction, particularly in complex cases.

3. DRMSpell achieves state-of-the-art performance in CSC tasks across various benchmarks, demonstrating the effectiveness of its multimodal approach and robustness in correcting Chinese character errors.

Yinghao Li, born in Shanxi in 1994 and is currently a doctoral student at the School of Computer Science and Technology, Beijing Institute of Technology. His research interests include Chinese spelling error correction, grammatical error correction, and in-context learning. He has published several papers in conferences such as ACL and EMNLP.



Yang Gao, Doctoral Supervisor, primarily engages in large language model training, automatic text generation technologies, and the application and transformation of these technologies. She has published over 60 high-level papers in international journals and conferences, including ACL, AAAI, WWW, IJCAI, EMNLP, TKDE and others. She has served as a chair in the text generation area for conferences like EMNLP and COLING, as an editorial board member for journals like *Web Intelligence* and *Natural Language Processing Journal*, as a program committee member for international conferences such as AAAI, ACL, EMNLP, NAACL, ICDM, and as a reviewer for journals including *TNNLS* and *Computing Surveys*.