Mitigating Idiom Inconsistency: A Multi-Semantic Contrastive Learning Method for Chinese Idiom Reading Comprehension

Mingmin Wu[1,2,3,4], **Yuxue Hu**[1,2,3,4], **Yongcheng Zhang**[1], **Zeng Zhi**[5], **Guixin Su**[1], **Ying Sha**[1,2,3,4]*

[1] College of Informatics, Huazhong Agricultural University, Wuhan, China
[2]Key Laboratory of Smart Farming for Agricultural Animals, Wuhan, China
[3]Hubei Engineering Technology Research Center of Agricultural Big Data, Wuhan, China
[4] Engineering Research Center of Intelligent Technology for Agriculture, Ministry of Education
[5] School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China
{wmm_nlp, zhyc, cometsue}@webmail.hzau.edu.cn, zhizeng@stu.xjtu.edu.cn, {hyx, shaying}@mail.hzau.edu.cn

Abstract

Chinese idioms pose a significant challenge for machine reading comprehension due to their metaphorical meanings often diverging from their literal counterparts, leading to metaphorical inconsistency. Furthermore, the same idiom can have different meanings in different contexts, resulting in contextual inconsistency. Although deep learning-based methods have achieved some success in idioms reading comprehension, existing approaches still struggle to accurately capture idiom representations due to metaphorical inconsistency and contextual inconsistency of idioms. To address these challenges, we propose a novel model, Multi-Semantic Contrastive Learning Method (MSCLM), which simultaneously addresses metaphorical inconsistency and contextual inconsistency of idioms. To mitigate metaphorical inconsistency, we propose a metaphor contrastive learning module based on the prompt method, bridging the semantic gap between literal and metaphorical meanings of idioms. To mitigate contextual inconsistency, we propose a multi-semantic cross-attention module to explore semantic features between different metaphors of the same idiom in various contexts. Our model has been compared with multiple current latest models (including GPT-3.5) on multiple Chinese idiom reading comprehension datasets, and the experimental results demonstrate that MSCLM outperforms state-of-the-art models.

Introduction

As a distinctive linguistic phenomenon in the Chinese language, Chinese idioms present a significant challenge for Chinese reading comprehension. Chinese idiom is a special kind of condensed form of language expression, typically consisting of four Chinese characters, which convey profound meanings that can differ greatly from their literal meanings, as shown in Table 1. Although large language models (LLMs) like GPT-3 have impressive performance in various reading comprehension tasks, they still struggle with understanding idioms (Bertolini, Weeds, and Weir 2022; Raunak et al. 2023). Given the prevalence of idioms in the Chinese language and the constantly emerging of new ones, a deeper comprehension of idioms is still critical for various natural language processing tasks (Shao et al. 2017; Liu, Pang, and Liu 2019). To address this challenge, Zheng et al. (2019) proposed a cloze task for idioms and built a widely used Chinese idiom cloze test dataset, ChID. This dataset contains passages with blank spaces accompanied by seven idiom options, as shown in Table 3. The task is to choose the idiom that best fits the context of the blank space, demonstrating the machine's comprehension of Chinese idioms.

Idiom	对牛弹琴
Literal Meaning	Play the piano to the cow.
Metaphorical Meaning	Talk to unreasonable people.

Table 1: An example of metaphorical inconsistency. The literal and metaphorical meanings of idioms are different.

Idiom	不翼而飞	平易近人
Literal	Fly away without wings.	Flatness facilitates approachability.
1. Metaphor	Things suddenly disappeared.	The person is kind and friendly.
2. Metaphor	News and words spread quickly.	The text is easy to understand.

Table 2: Examples of contextual inconsistency. The same idiom may have different meanings in different contexts.

The development of the pre-training models has led to the emergence of various Chinese idiom understanding methods based on BERT model (Devlin et al. 2018) and the Chinese idiom-oriented model (Tan, Jiang, and Dai 2021). CM model (Wang et al. 2020) uses BERT to encode the context sequence and the idiom's definition. SKER model (Long et al. 2020) constructs a synonym graph of idioms based on their definitions to create new idiom representations. However, these models do not account for the fact that idioms with similar semantic meanings can be applicable to differ-

*Corresponding author.

Passage	他的脚踝终于____，骨裂的结果让他提早告别了季后赛。 His ankle finally ____, and the fracture caused him to bid farewell to the playoffs early.		
Candidate Idioms	NO.1	劳苦功高	Work hard and do a lot of credit.
	NO.2	不治之症	A disease that cannot be cured.
	NO.3	积劳成疾	Long-term overwork and illness.
	NO.4	香消玉碎	A beautiful young woman dies.
	NO.5	急起直追	Act now and try to catch up.
	NO.6	生龙活虎	Lively and vigorous, full of vitality.
	NO.7	百发百中	Do things with full grasp.
Ground Truth	NO.3	积劳成疾	Long-term overwork and illness.

Table 3: An example of Chinese idiom cloze test. The translations of candidate idioms all refer to metaphorical meanings.

ent scenarios. A prompt-based projection method has been used to fuse the definitions of the idioms (Sha et al. 2023). Nevertheless, directly integrating the definitions of idioms into the model as external knowledge is not reasonable, as the meanings of idioms and their definitions are similar but not equivalent. Moreover, using the definition of the idiom as the representation of the idiom does not consider the different usage scenarios of the idiom in different contexts.

Although deep learning-based methods have achieved some success in idioms reading comprehension, existing approaches still struggle to accurately capture idiom representations. Most Chinese idioms have their roots in ancient myths and historical allusions. And, as time passed, the meanings and usage of these idioms have evolved, leading to inconsistencies in their expressions and meanings (Wang and Luo 2021; Abdessamad 2023). It can be primarily attributed to two factors: metaphorical inconsistency and contextual inconsistency. Table 1 provides an example of the metaphorical inconsistency of idioms, where the idiom's literal meaning is completely different from its metaphorical meaning. Table 2 provides several examples of contextual inconsistency, where the meaning of an idiom may vary depending on the context and subject. Therefore, it is essential to understand the metaphorical and contextual inconsistencies of idioms to better comprehend their meanings and usage in different situations.

To address the above challenges, we propose a model named Multi-Semantic Contrastive Learning Method (MSCLM). We use a metaphor contrastive learning module to select metaphorical static representations as positive examples for contextual dynamic representations, and representations of other idioms as negative examples. This method bridges the semantic similarity between literal and metaphorical meanings of idioms, and enables the model to learn metaphorical meanings that can distinguish idioms in their semantic feature space. Thus, this module effectively mitigates the issue of metaphorical inconsistency. Inspired by fusing convolutional neural networks (CNNs) and attention mechanisms (Carion et al. 2020; Li et al. 2021, 2022), we adopt a double fusion of convolution and cross-attention mechanism, and propose a multi-semantic cross-attention module. Specifically, 2D convolutions are applied to capture local relationships in text sequences for natural language processing, thereby showing high efficiency in phrase analysis, idiom formation, and modeling local dependencies in language (Liu et al. 2018; Wang and Gang 2018). Meanwhile, a cross-attention mechanism is applied to capture global contextual information, which helps the model better understand the semantics of the entire text sequence (Lee et al. 2018; Hou et al. 2019). This module uses different convolution kernels and cross-attention mechanisms to capture the multi-semantic features of idioms and effectively mitigate the issue of contextual inconsistency.

Experimental results on several benchmark datasets demonstrate that our method outperforms current state-of-the-art methods. In addition, in order to further verify the model's ability to understand the actual meaning of the idiom, we constructed a new test set CIDT based on the definition of the idiom and the candidate set of synonyms for the idiom. The model needs to select the most matching idiom according to the definition of the idiom. Our MSCLM model also achieves state-of-the-art performance on this additional test set. The main contributions of this paper are as follows:

- A metaphor contrastive learning module is proposed, which enables the model to learn metaphorical feature representations of idioms in the semantic feature space, effectively mitigating the metaphorical inconsistency of idioms.
- A multi-semantic cross-attention module is proposed, which allows the model to learn idiom representations from multiple semantic features in different contexts, effectively mitigating the contextual inconsistency of idioms in different contexts.
- We propose a general framework for cloze-style reading comprehension tasks and a new Chinese idiom cloze test dataset CIDT, enabling a multidimensional understanding of idiom semantic representations.

Related Work

Although the comprehension of idioms is important for many natural language processing tasks (Shao et al. 2017; Liu, Pang, and Liu 2019; Qin et al. 2021), it has only been gaining attention since 2019. Zheng et al. (2019) proposed the cloze test task and created the dataset ChID to assess machines' understanding of Chinese idioms. Early work used

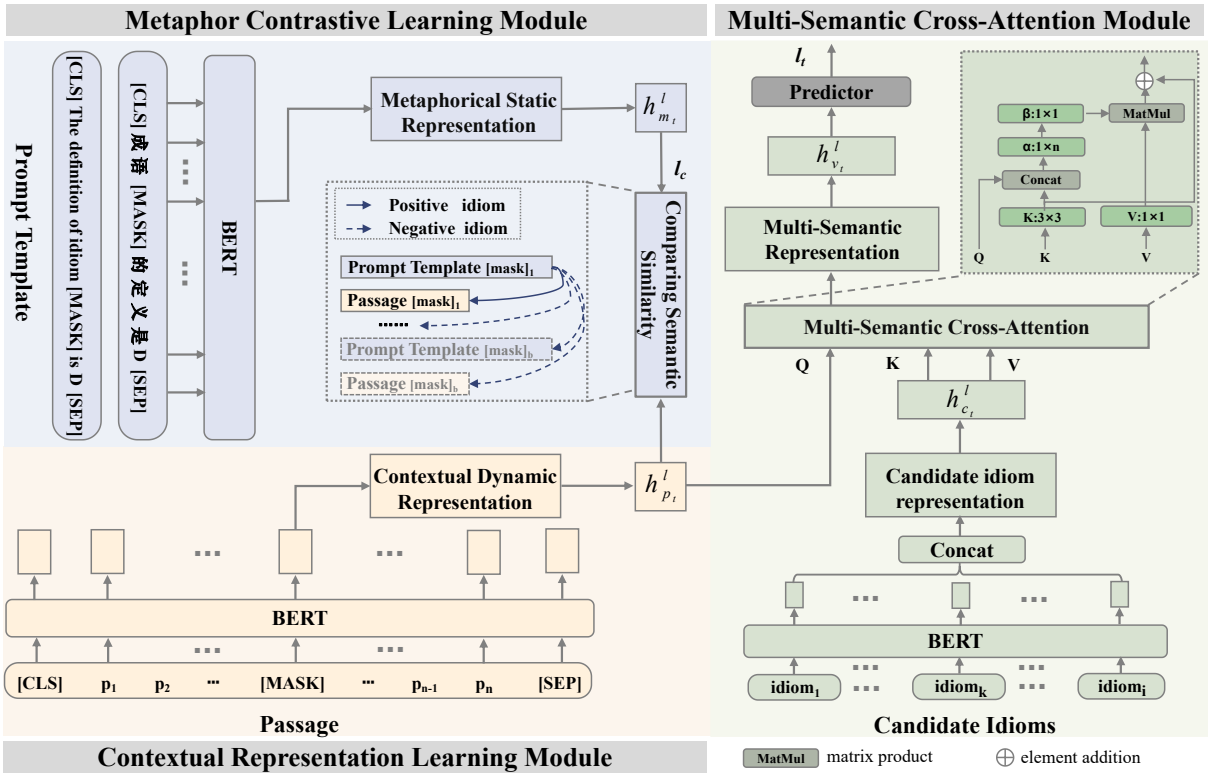


Figure 1: The architecture of our model. D represents the definition of the idiom. The batch size is b. The size of i is the number of idioms in the idiom candidate set, which is 7 for the ChID dataset and 4 for the CCT dataset.

BiLSTM (Xu et al. 2019) to encode the input and obtain the hidden state of the blank space, which is used to select idioms. BERT model (Lan et al. 2019) exhibits enhanced feature extraction capabilities compared to LSTM, leading to significant improvement in the accuracy of idiom reading comprehension tasks. The first popular BERT-based idiom cloze model is the dual embedding pre-training model (Tan and Jiang 2020), which encodes the context and learns the dual embedding of the idiom to predict the idiom. This model is pre-trained on a large Chinese corpus in two stages (Tan, Jiang, and Dai 2021). However, the basic BERT model does not learn the correspondence between the literal and actual metaphorical meanings of idioms.

Research has confirmed that synonyms of idioms can facilitate the model's understanding of idioms (Long et al. 2020). Researchers constructed a synonym graph based on the idiom's definition and encoded the graph to obtain the idiom representation. However, this method did not consider the multiple meanings of idioms in different contexts. To obtain more accurate idiom representation, external knowledge, such as idiom definitions, was introduced (Wang et al. 2020; Sha et al. 2023). However, using only the definitions of idioms and concatenating their characters into the input sequence cannot be completely equivalent to the representation of idioms in different contexts.

Large language models (LLMs) have made significant strides in NLP tasks such as reading comprehension (Chakrabarty et al. 2022). These LLMs have expanded their capabilities by incorporating richer corpus information and utilizing sophisticated pre-training techniques (Zhang et al. 2021; Wang et al. 2023b). As a result, their performance on the idiom cloze dataset ChID has shown considerable improvements. However, it is important to note that despite these advancements, the unique linguistic characteristics of Chinese idiomatic expressions still pose challenges in terms of inconsistency. Even the strongest ChatGPT model exhibits limitations in comprehending sentences that contain idiomatic phrases (Raunak et al. 2023).

To address the challenges of metaphorical and contextual inconsistency, we propose the metaphor contrastive learning module which can more accurately capture the metaphorical features of idioms by using contrastive learning (Wang et al. 2022, 2023a) and the multi-semantic cross-attention module which can learn idioms from multiple semantic features in different contexts and better capture the relationship between idioms and context.

MSCLM Model

To address the metaphorical and contextual inconsistency, we propose the MSCLM model, which comprises the contextual representation learning module, the metaphor contrastive learning module, the multi-semantic cross-attention module, and the idiom prediction module, as shown in Figure 1. During training, each idiom is presented with various

contexts, and the contextual representation learning module captures the dynamic contextual representation of each idiom. The metaphor contrastive learning module employs a contrastive learning method to learn accurate metaphorical static representations, mitigating metaphorical inconsistency. The multi-semantic cross-attention module utilizes cross-attention to capture the relationship between idioms and contexts, and learns multi semantic features that help mitigate contextual inconsistency. The idiom prediction module predicts the correct idiom.

Problem Formulation

The Chinese idiom cloze test is a task that involves filling in a blank space in a passage with an appropriate idiom. The passage $P_t = \{\mathrm{p}_1, \mathrm{p}_2, \cdots, [\mathrm{MASK}], \cdots, \mathrm{p}_\mathrm{n}\}$ is given with each character $\mathrm{p_i}$ representing a Chinese character, and the blank space is marked by [MASK]. The objective is to select the most appropriate idiom from the t-th candidate set $C_t = \{\mathrm{c}_1, \mathrm{c}_2, \cdots, \mathrm{c}_i\}$, which contains the possible idioms that can complete the passage.

Contextual Representation Learning Module

The contextual representation learning module is used to capture the dynamic contextual representation of each idiom. We set the [CLS] and [SEP] characters at the beginning and end of each sentence in the passage collection P. These serve as markers for sentence sequence representation and sentence boundary. The blank space that needs to be filled is marked as [MASK]. The context of the idiom that fills the blank space changes with different passage inputs, even for the same idiom. Therefore, the learned representation of the [MASK] is a dynamic representation of the idiom.

As a result, the t-th input sequence to the BERT model can be represented as $P_t = \{[\mathrm{CLS}], \mathrm{p}_1, \mathrm{p}_2, \cdots, [\mathrm{MASK}], \cdots, \mathrm{p}_\mathrm{n}, [\mathrm{SEP}]\}$. For instance, the example in Table 3 can be rephrased as "[CLS] His ankle finally [MASK], and the fracture caused him to bid farewell to the playoffs early. [SEP]". Next, the transformer blocks of successive layers in BERT perform feature extraction:

$$h^l_{p_t} = \text{Transformer}\,(P_t) \tag{1}$$

The last layer (l-th layer) of the model obtains the hidden layer representation of [MASK] as the contextual dynamic representation $h^l_{p_t}$ for the idiom. While $h^l_{p_t}$ is the basic representation for the blank space and can be directly used for idiom prediction, its performance is not optimal. This is because a simple contextual representation is not accurate enough, and only using the contextual dynamic representation cannot enable the model to learn the corresponding relationship between the literal meaning and the actual metaphorical meaning of the idiom. There is a metaphorical inconsistency between the literal meaning and the actual metaphorical meaning of the idiom.

Metaphor Contrastive Learning Module

To address metaphorical inconsistency, we use a metaphor contrastive learning module based on the prompt method to semantically narrow the gap between idioms' literal and metaphorical representations. Specifically, we first use a Chinese whole word mask BERT model (Cui et al. 2021) to obtain the corresponding metaphorical static representation. Then, we construct positive and negative samples, and use the SimCSE model (Gao, Yao, and Chen 2021) to make the semantic distance of positive samples closer in the semantic feature space of idioms.

For the t-th input sequence P_t, we create the corresponding prompt template $M_t = \{$[CLS]成语[MASK]的定义是D[SEP]$\}$ (meaning: [CLS] The definition of idiom [MASK] is D [SEP]]), where D represents the definition of the idiom. For example, the prompt template corresponding to the idiom "积劳成疾(Long-term overwork and illness)" in Table 3 is "[CLS] The definition of idiom [MASK] is long-term overwork and illness [SEP]". We also use the BERT model to process template sentences generated by the prompt method.

$$h^l_{m_t} = \text{Transformer}\,(M_t) \tag{2}$$

The final layer (l-th layer) of the model obtains the hidden layer representation of [MASK] as the metaphorical static representation $h^l_{m_t}$ of the idiom.

Based on the SimCSE model, we select the contextual dynamic representation $h^l_{p_t}$ and metaphorical static representation $h^l_{m_t}$ as positive samples, and the representations of other contexts and template sentences within a mini-batch as negative samples for $h^l_{p_t}$ and $h^l_{m_t}$. The metaphor contrastive learning module mitigates metaphorical inconsistency by maximizing the consistency of positive samples in the idiom semantic space. The training objective for ($h^l_{p_t}$, $h^l_{m_t}$) with a mini-batch of N pairs is:

$$l_c = -\log \frac{e^{sim\left(h^l_{p_t}, h^l_{m_t}\right)/\tau}}{\sum_{t'=1}^{N} e^{sim\left(h^l_{p_t}, h^l_{m_{t'}}\right)/\tau}} \tag{3}$$

where τ is a temperature hyperparameter and $sim(h_1, h_2)$ is the cosine similarity $\frac{h_1^T h_2}{\|h_1\| \cdot \|h_2\|}$. The metaphor contrastive learning process is trained simultaneously with the idiom cloze test. Therefore, l_c is an additional loss value obtained through contrastive learning, in addition to the basic idiom cloze test loss.

Multi-Semantic Cross-Attention Module

To effectively tackle contextual inconsistency, we propose a multi-semantic cross-attention module that integrates semantic features of idioms to enhance the model's ability to comprehend idioms across diverse contexts.

We employ a BERT model to encode each idiom in the candidate set. The idiom is represented as a sequence with [CLS] and [SEP] denoting the beginning and end of the idiom, respectively. As an idiom usually contains four Chinese characters, the k-th idiom $c_k = \{[\mathrm{CLS}], \mathrm{w}_1, \mathrm{w}_2, \mathrm{w}_3, \mathrm{w}_4, [\mathrm{SEP}]\}$ in the t-th candidate set C_t is input to the BERT model. For example, the candidate idiom "劳苦功高 (Work hard and do a lot of credit.)" in Table 3 can be rewritten as "$\{[\mathrm{CLS}], 劳, 苦, 功, 高, [\mathrm{SEP}]\}$".

The idiom c_k is then transformed into an idiom representation $h^l_{c_k} \in \mathbb{R}^{N\times d}$ using a continuous l-layer Transformer, where N is the maximum sequence length and d is the hidden layer's dimension.

$$h^l_{c_k} = \text{Transformer}\left(c_k\right) \tag{4}$$

The final layer (l-th layer) of the model obtains the hidden layer representation of [CLS] as the idiom representation h^l_c. Then, the idiom representations of i idioms in the t-th candidate set C_t are concatenated to form the candidate idiom representation $h^l_{c_t}$,

$$h^l_{c_t} = \text{Concat}\left(h^l_{c_1}, \ldots, h^l_{c_k} \ldots, h^l_{c_i}\right) \tag{5}$$

where the Concat function is the concatenation operation.

To leverage the mutual semantic relationship information between idioms, we integrate convolution and cross-attention to facilitate the learning of contextual dynamic representations of idioms in different contexts. Specifically, the candidate idiom representation $h^l_{c_t}$ is first subjected to group convolution processing for contextualizing each key representation.

$$h_k = f_{K_{3\times3}}\left(h^l_{c_t}\right) \tag{6}$$

where $f_{K_{3\times3}}$ is the convolution operation of 3×3. The learned contextualized key h_k contains mutual information between idioms. Then concatenate contextual dynamic representation $h^l_{p_t}$ and contextualized key h_k. The concatenated vectors get the attention matrix h_A through a $1 \times n$ convolution $f_{\alpha_{1\times n}}$ and a 1×1 convolution $f_{\beta_{1\times1}}$:

$$h_A = f_{\beta_{1\times1}}\left(f_{\alpha_{1\times n}}\left(\text{Concat}\left(h_k, h^l_{p_t}\right)\right)\right) \tag{7}$$

where n is the length of the sequence after concatenation. 3×3 convolution can capture local features and idiom structure. 1×1 convolution can help the model learn the correlation information between different channels so that the model can better understand the semantics and context of idioms. The attention matrix h_A is learned based on the relationship between the idiom contextual dynamic representation and the idiom representation in the candidate set, instead of obtaining each idiom representation in isolation. Next, we aggregate the candidate idiom representation based on the attention matrix h_A to calculate the multi-semantic representation $h^l_{v_t}$ of the idiom.

$$h^l_{v_t} = f_{V_{1\times1}}\left(h^l_{c_t}\right) \otimes h_A \oplus h_k \tag{8}$$

where $f_{V_{1\times1}}$ represents 1×1 convolution, $\otimes$ denotes the matrix multiplication operation, $\oplus$ represents the addition fusion operation. The multi-semantic representation $h^l_{v_t}$ makes full use of the semantic information between idioms in the candidate set to guide the semantic learning of idioms in different contexts.

Idiom Prediction Module

The final idiom prediction module scores each candidate idiom in the candidate idiom set C_t based on the multi-semantic representation $h^l_{v_t}$. After applying the softmax function, it computes the probability of selecting idiom c_j ($c_j \in C_t$) given the context sequence P_t,

$$\mathrm{P}\left(c_j \mid P_t\right) = \frac{\exp\left(w \cdot \left(h_{c_i} \otimes h^l_{v_t}\right) + b\right)}{\sum_{c'_j \in C} \exp\left(w \cdot \left(h_{c'_j} \otimes h^l_{v_t}\right) + b\right)} \tag{9}$$

where $w \in \mathbb{R}^d$, $b \in \mathbb{R}$ are parameters mapping the representation into a score, $\otimes$ is the element-wise multiplication and h_{c_i} denotes the embedding vector of each candidate idiom.

The prediction loss l_t for the idiom cloze test is the minimization of the cross-entropy loss between the correct idiom and the predicted idiom,

$$l_t = -\sum_{j=1}^{i} \mathrm{c}_g \log \mathrm{P}\left(\mathrm{c}_j \mid P_t\right) \tag{10}$$

The candidate set of idioms consists of i idioms, and c_g is the one-hot label distribution for the correct idiom. The prediction loss l_t and the contrastive learning loss l_c are summed to form our overall training objective. The final loss function l_f can be expressed as:

$$l_f = l_c + l_t \tag{11}$$

Idiom prediction and contrastive learning are trained simultaneously, and we fine-tune all parameters using the total training objective l_f.

Experiments

Data and Experimental Setup

Datasets The datasets used for our experiments are listed below.

- **ChID** (Zheng, Huang, and Sun 2019), a large-scale Chinese idiom reading comprehension dataset, comprises diverse data types such as news, novels, and essays sourced from the internet. The dataset is segmented into two categories: in-domain and out-of-domain data. The in-domain data includes a training set, a validation set (**Dev**), and a test set (**Test**), while the out-of-domain data comprises the test set (**Out**). The dataset also has a **Sim** test set which has the same passages as the **Test** set. But the candidate idioms in **Sim** are similar to the correct answer, making **Sim** more challenging than **Test** set.
- **ChID-Competition** [1] is an online dataset built upon the ChID dataset. The dataset consists of a collection of passages, each containing multiple blanks that share a set of candidate idioms. One must choose an idiom from a fixed-length set of candidates, with each option allowed only once.
- **CCT** (Jiang et al. 2018) dataset is more challenging. It consists of 108,987 sentences and 7,395 idioms, with 7,071 and 508 idioms in the training and testing sets, respectively. The candidate idioms set includes the correct answer and three other randomly selected idioms.

[1] https://github.com/chujiezheng/ChID-Dataset/tree/master

Model	Dev	Test	Sim	Out
Human	–	87.1	82.2	86.2
LM	71.8	71.5	65.6	61.5
AR	72.7	72.4	66.2	62.9
SAR	71.7	71.5	64.9	61.7
BERT-WWM	75.4	75.7	70.2	66.1
SKER	76.0	76.3	68.8	68.3
BTSM	81.9	81.8	74.1	72.0
CM	83.0	83.1	76.1	77.6
BERT-IDM	–	83.2	–	67.5
PRIEM	95.8	95.7	97.9	92.6
GPT-3.5	–	39.8	32.1	29.6
MSCLM-C	81.7	81.8	75.3	74.7
MSCLM-M	87.1	86.8	91.4	80.5
MSCLM-(C+M)	**97.1**	**96.8**	**98.8**	**95.2**

Table 4: Performance of accuracy on the dataset ChID.

ChID-Competition				CCT	
Model	**Dev**	**Test**	**Out**	**Model**	**Test**
AR	65.4	65.6	55.6	**Human**	70.0
BERT	72.7	72.4	64.7	**BiLSTM**	89.5
BTSM	92.4	92.0	90.2	**BTSM**	93.7
GPT-3.5	–	23.7	23.5	**GPT-3.5**	73.2
MSCLM	**96.4**	**96.5**	**95.3**	**MSCLM**	**96.4**

Table 5: Results on ChID-Competition and CCT.

Model	BERT	PRIEM	GPT-3.5	MSCLM-C	MSCLM
Accuracy	24.7	74.6	84.8	94.6	**98.8**

Table 6: Experimental results of accuracy on CIDT.

- **CIDT.** To further assess the model's capability in comprehending the metaphorical meaning of idioms, we created a Chinese idiom definition test dataset, CIDT. After the model training is completed, CIDT is used as an additional test set. As shown in Table 7, given the explanation of an idiom, the model is required to choose the corresponding correct idiom. The other six idioms in the candidate set are all synonyms of the correct answer, and 3600 idioms are used as test data.

Evaluation Metrics The evaluation metric is accuracy, which is defined as the proportion of test samples where the model's chosen idiom corresponds to the correct idiom.

Experimental Settings We trained our model using the Chinese pre-trained BERT (Cui et al. 2021) model, which is based on a whole-word mask. We set the maximum input sequence length to 128 and the batch size to 32. The model was run on an NVIDIA 3080Ti GPU. For training, we used Transformers 3.1.0 (Wolf et al. 2020) and the total number of epochs was 5. The initial learning rate was set to 5e-5, and we employed the warm-up linear schedule strategy with 1000 warm-up steps. Our code and dataset are available.

Results and Discussion

Methods to Compare We use the following methods to compare with our model.

Language Model (LM) (Zhou et al. 2016): This method uses a standard BiLSTM model to encode the given sequence and compare its hidden state with the representation of each candidate idiom to select the idiom.

Attentive Reader (AR) (Luong et al. 2015): This method enhances the BiLSTM model using attention mechanism.

Standard Attentive Reader (SAR) (Hermann et al. 2015): An improved version of AR using a bilinear function as a matching function for more flexible computations.

BERT-WWM (Cui et al. 2021): An upgraded version of the BERT model that uses a whole word mask to mask the word.

Synonym Knowledge Enhanced Reader (SKER) (Long et al. 2020): This method constructs a synonym graph based on cosine similarity of idiom embeddings and encodes the graph to replace the original idiom representation.

BERT-based Two-stage Model (BTSM) (Tan, Jiang, and Dai 2021): This model is pre-trained on a large Chinese corpus and fine-tuned to predict idioms.

Correcting the Misuse (CM) (Wang et al. 2020): This method introduces idiom definitions and uses an attribute attention mechanism to balance the weights of different attributes among different representations of idioms.

BERT-IDM (Dai et al. 2023): Using the definition of idioms to re-train the BERT, and then uses multi-granularity integrated attention to capture multi-dimensional information.

PRIEM (Sha et al. 2023): Fusing the definitions of idioms through the prompt method and then using orthogonal projection to distinguish idioms' representations.

GPT-3.5 (OpenAI 2023): A large language model developed by OpenAI, has achieved superior performance in various NLP tasks.

MSCLM-C: Our method only uses the metaphor contrastive learning module.

MSCLM-M: Our method only uses the multi-semantic cross-attention module.

MSCLM-(C+M): Our method both uses the metaphor contrastive learning module and the multi-semantic cross-attention module.

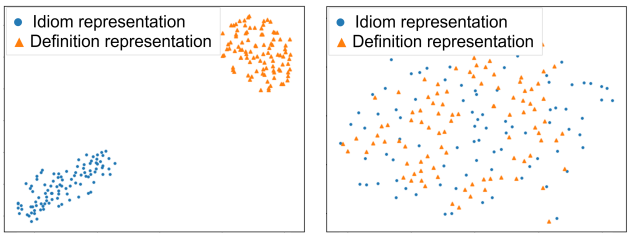
Results and Analysis Table 4 displays the accuracy of all methods on the ChID dataset. Table 5 and 6 show the experimental results of the ChID Competition, CCT, and CIDT datasets. The performance of humans, LM, AR, and SAR is directly obtained from ChID (Zheng, Huang, and Sun 2019). The results in Tables 4, 5 and 6 indicate that our MSCLM model outperforms other methods. Compared with the best results in the past, our model has an improvement of 1%-5% on each data set, especially on the ChID's **Sim** and the CIDT datasets, reaching 98.8% accuracy.

Previous methods based on the BERT model do not fully consider both the metaphorical inconsistency and the contextual inconsistency of idioms. The GPT-3.5 model has shown remarkable proficiency in handling various reading comprehension tasks. While results in Tables 4 and 5 reveal that there is still significant scope for enhancing its reading comprehension ability concerning Chinese idioms.

The MSCLM-C method introduces the definition knowl-

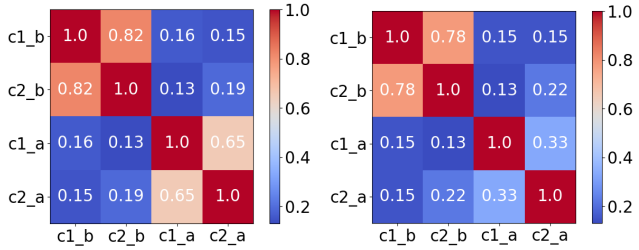
Content	“成语#idiom#的解释是形容变化极多。” “The explanation of #idiom# is to describe something that is subject to a wide range of changes.”
Candidate Idioms	#千变万化#，#千姿百态#，#包罗万象#，#变化无常#，#千奇百怪#，#瞬息万变#，#天翻地覆# #Ever-changing#，#Diverse-varying#, #All-inclusive#，#Unpredictable#，#Bizarre#，#Constantly-changing#，#Earth-shaking#
Ground Truth	#千变万化# #Ever-changing#

Table 7: An example of our constructed Chinese idiom definition test dataset CIDT.



(a) Before applying metaphor contrastive learning module. (b) After applying metaphor contrastive learning module.

Figure 2: The representation distribution of 100 idioms and definitions before and after metaphor contrastive learning. The blue origin is the idiom representation and the orange triangle is the idiom's definition representation.



(a) The similarity of the idiom “不翼而飞” in two contexts before and after the multi-semantic cross-attention module. (b) The similarity of the idiom “平易近人” in two contexts before and after the multi-semantic cross-attention module.

Figure 3: The heat map of the similarity of two idioms in two contexts before and after multi-semantic cross-attention. The c1 and c2 in the figure represent two different contexts, _b represents before training, and _a represents after training.

edge of idioms through the prompt method, and then uses contrastive learning to narrow the semantic relationship between the literal and metaphorical meanings of idioms, thus mitigating metaphorical inconsistency of idioms. With a multi-semantic cross-attention module, learning the multiple semantic features of idioms in different contexts helps to mitigate the contextual inconsistency of idioms. The model learns the correlation between idiom representations and different contextual representations, and the MSCLM-M method performs better.

The CIDT dataset requires the model to correctly understand the metaphorical meaning of idioms, and to distinguish idioms with similar meanings when selecting idioms. Since the sentences in CIDT are relatively short and straightforward, the performance of GPT-3.5 is much higher than it is on ChID and CCT. Our MSCLM model achieves the best performance on CIDT, further verifying that our model achieves a deep understanding of idiom semantics.

Visualization and Analysis of MSCLM Model

To better illustrate the impact of the proposed modules, we visualize and analyze the idiom representations before and after applying metaphor contrastive learning and multi-semantic cross-attention.

We randomly selected 100 idiom representations and their definition representations, and used the t-SNE algorithm to draw the 2D distribution of representations before and after metaphor contrastive learning. As depicted in Figure 2, the distance between the idiom's representation and its definition representation becomes closer after training. In the semantic feature space of the idiom, the literal and metaphorical meanings of idioms are more consistent.

We also randomly selected 20 idioms with dual metaphorical meanings, and selected two corresponding contexts for the metaphorical meanings of the idioms. We drew the similarity heat map of the idioms' representation in the two selected contexts before and after the multi-semantic cross-attention, as shown in Figure 3. The similarity of the same idiom in the two contexts before multi-semantic cross-attention is very high, around 0.8. However, after the multi-semantic cross-attention, the similarity has significantly reduced. This shows that the semantic distinction of the same idiom in different contexts is more obvious, and the multi-semantic cross-attention module can effectively mitigate the contextual inconsistency.

Conclusion

In this paper, we propose an MSCLM model for Chinese idiom reading comprehension, which utilizes a metaphor contrastive learning module to reduce the semantic distance between the idiom representation and its actual metaphorical meaning. By learning the metaphorical features of idioms in the semantic feature space, our model mitigates the metaphorical inconsistency of idioms. Next, we employ a multi-semantic cross-attention module to help the model learn the semantics of idioms in multiple contexts to further mitigate the contextual inconsistency of idioms. Our MSCLM model outperforms state-of-the-art models on previous Chinese idiom comprehension datasets. In the future, we plan to combine our module with some large language models for reading comprehension research on English slang and other metaphorical languages.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities 2662021JC008, the National Natural Science Foundation of China (No. 62272188), and the Inner Mongolia Autonomous Region Major Science and Technology Project 2021ZD0046. Thank the experimental teaching center of College of informatics in Huazhong Agricultural University for providing the experimental environment and computing resources.

References

Abdessamad, L. 2023. A Study on English Translation of Chinese Idioms. *Open Journal of Modern Linguistics*, 13(3): 373–381.

Bertolini, L.; Weeds, J.; and Weir, D. 2022. Testing Large Language Models on Compositionality and Inference with Phrase-Level Adjective-Noun Entailment. In *Proceedings of the 29th International Conference on Computational Linguistics*, 4084–4100. Gyeongju, Republic of Korea: International Committee on Computational Linguistics.

Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; and Zagoruyko, S. 2020. End-to-end object detection with transformers. In *European conference on computer vision*, 213–229. Springer.

Chakrabarty, T.; Saakyan, A.; Ghosh, D.; and Muresan, S. 2022. Flute: Figurative language understanding through textual explanations. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 7139–7159.

Cui, Y.; Che, W.; Liu, T.; Qin, B.; and Yang, Z. 2021. Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29: 3504–3514.

Dai, Y.; Liu, Y.; Yang, L.; and Fu, Y. 2023. An Idiom Reading Comprehension Model Based on Multi-Granularity Reasoning and Paraphrase Expansion. *Applied Sciences*, 13(9): 5777.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Gao, T.; Yao, X.; and Chen, D. 2021. Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821*.

Hermann, K. M.; Kocisky, T.; Grefenstette, E.; Espeholt, L.; Kay, W.; Suleyman, M.; and Blunsom, P. 2015. Teaching machines to read and comprehend. *Advances in neural information processing systems*, 28.

Hou, R.; Chang, H.; Ma, B.; Shan, S.; and Chen, X. 2019. Cross attention network for few-shot classification. *Advances in neural information processing systems*, 32.

Jiang, Z.; Zhang, B.; Huang, L.; and Ji, H. 2018. Chengyu Cloze Test. In *BEA@ NAACL-HLT*, 154–158.

Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; and Soricut, R. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Lee, K.-H.; Chen, X.; Hua, G.; Hu, H.; and He, X. 2018. Stacked cross attention for image-text matching. In *Proceedings of the European conference on computer vision (ECCV)*, 201–216.

Li, Y.; Pan, Y.; Yao, T.; Chen, J.; and Mei, T. 2021. Scheduled Sampling in Vision-Language Pretraining with Decoupled Encoder-Decoder Network. arXiv:2101.11562.

Li, Y.; Yao, T.; Pan, Y.; and Mei, T. 2022. Contextual transformer networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 1489–1500.

Liu, Y.; Liu, B.; Shan, L.; and Wang, X. 2018. Modelling context with neural networks for recommending idioms in essay writing. *Neurocomputing*, 275: 2287–2293.

Liu, Y.; Pang, B.; and Liu, B. 2019. Neural-based Chinese idiom recommendation for enhancing elegance in essay writing. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 5522–5526.

Long, S.; Wang, R.; Tao, K.; Zeng, J.; and Dai, X.-Y. 2020. Synonym knowledge enhanced reader for Chinese idiom reading comprehension. *arXiv preprint arXiv:2011.04499*.

OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774.

Qin, R.; Luo, H.; Fan, Z.; and Ren, Z. 2021. IBERT: Idiom Cloze-style reading comprehension with Attention. *arXiv preprint arXiv:2112.02994*.

Raunak, V.; Menezes, A.; Post, M.; and Awadallah, H. H. 2023. Do GPTs Produce Less Literal Translations? *arXiv preprint arXiv:2305.16806*.

Sha, Y.; Wu, M.; Zeng, Z.; Ge, X.; Huang, Z.; and Wang, H. 2023. A Prompt-Based Representation Individual Enhancement Method for Chinese Idiom Reading Comprehension. In *International Conference on Database Systems for Advanced Applications*, 682–698. Springer.

Shao, Y.; Sennrich, R.; Webber, B.; and Fancellu, F. 2017. Evaluating machine translation performance on chinese idioms with a blacklist method. *arXiv preprint arXiv:1711.07646*.

Tan, M.; and Jiang, J. 2020. A BERT-based dual embedding model for Chinese idiom prediction. *arXiv preprint arXiv:2011.02378*.

Tan, M.; Jiang, J.; and Dai, B. T. 2021. A BERT-based two-stage model for Chinese chengyu recommendation. *Transactions on Asian and Low-Resource Language Information Processing*, 20(6): 1–18.

Wang, F.; Wang, Y.; Li, D.; Gu, H.; Lu, T.; Zhang, P.; and Gu, N. 2023a. CL4CTR: A Contrastive Learning Framework for CTR Prediction. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 805–813.

Wang, S.; and Luo, H. 2021. Exploring the meanings and grammatical functions of idioms in teaching Chinese as a second language. *International Journal of Applied Linguistics*, 31(2): 283–300.

Wang, W.; and Gang, J. 2018. Application of convolutional neural network in natural language processing. In *2018*

international conference on information Systems and computer aided education (ICISCAE), 64–70. IEEE.

Wang, X.; Zhao, H.; Yang, T.; and Wang, H. 2020. Correcting the misuse: A method for the Chinese idiom cloze test. In *Proceedings of Deep Learning Inside Out (DeeLIO): The First Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, 1–10.

Wang, Y.; Wang, J.; Zhao, D.; and Zheng, Z. 2023b. Shuo Wen Jie Zi: Rethinking Dictionaries and Glyphs for Chinese Language Pre-training. *arXiv preprint arXiv:2305.18760*.

Wang, Z.; Wang, P.; Huang, L.; Sun, X.; and Wang, H. 2022. Incorporating hierarchy into text encoder: a contrastive learning approach for hierarchical text classification. *arXiv preprint arXiv:2203.03825*.

Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, 38–45.

Xu, G.; Meng, Y.; Qiu, X.; Yu, Z.; and Wu, X. 2019. Sentiment analysis of comment texts based on BiLSTM. *Ieee Access*, 7: 51522–51532.

Zhang, Z.; Han, X.; Zhou, H.; Ke, P.; Gu, Y.; Ye, D.; Qin, Y.; Su, Y.; Ji, H.; Guan, J.; et al. 2021. CPM: A large-scale generative Chinese pre-trained language model. *AI Open*, 2: 93–99.

Zheng, C.; Huang, M.; and Sun, A. 2019. ChID: A large-scale Chinese IDiom dataset for cloze test. *arXiv preprint arXiv:1906.01265*.

Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.; and Xu, B. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)*, 207–212.