Computers and Artificial Intelligence；Vol.3, No.3; 2026
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.26v3n3.7483
Published by: Zeus Press

Research on Cross-lingual Causal Representation Learning Based on Multi-dimensional Semantic Disentanglement

Wenwen Zhao*

School of Computer and Artificial Intelligence, Xinjiang Hetian College, Hetian, Xinjiang, China

**Corresponding author: Wenwen Zhao.*

Abstract

This paper focuses on several key challenges in cross-lingual natural language processing, including unstable semantic transfer, the entanglement of language-specific features, and insufficient interpretability of learned representations. To address these issues, the study introduces the necessity of combining multi-dimensional semantic disentanglement with causal representation learning. A cross-lingual causal representation learning framework is proposed, covering semantic dimension decomposition, causal factor modeling, language-invariant representation extraction, semantic intervention mechanisms, and cross-lingual transfer optimization. Finally, the paper summarizes the experimental findings and discusses the theoretical and practical significance of the proposed approach.

Keywords

semantic disentanglement, causal representation learning, cross-lingual transfer, language-invariant representation, natural language processing

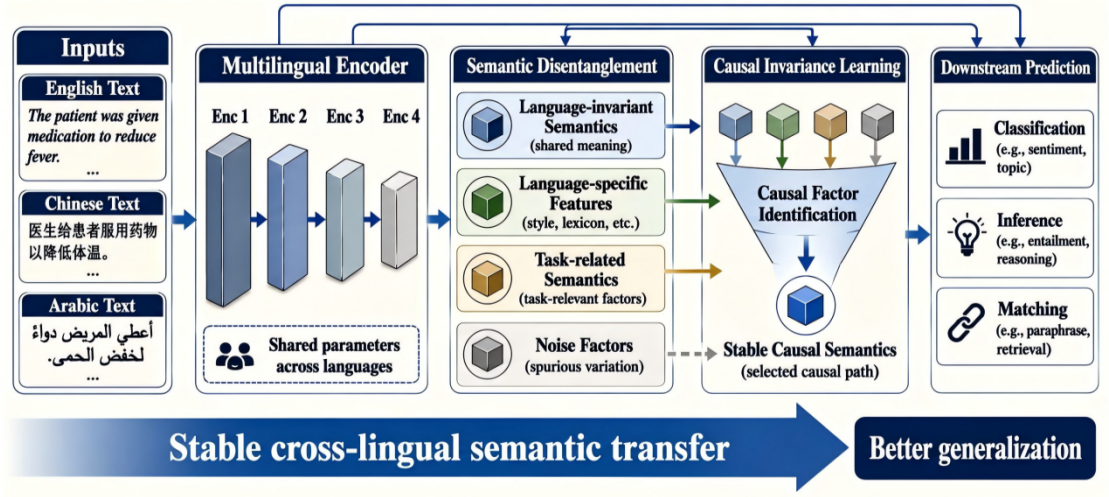
1. Introduction

Cross-lingual natural language processing plays an important role in tasks such as machine translation, cross-lingual text classification, cross-lingual sentiment analysis, and low-resource language understanding. Existing cross-lingual representation learning methods can obtain a unified semantic space through multilingual pre-trained models. However, representations learned across different languages are still easily affected by differences in word order, grammatical rules, cultural context, and data distribution. As a result, semantic information, language-specific features, and task-related noise may be mixed within the same representation space. Moreover, methods based only on correlation learning are often unable to explain why cross-lingual transfer fails in certain cases. Therefore, it is necessary to introduce causal representation learning so that stable semantic factors that truly influence task outputs can be identified from the perspective of causal mechanisms. Based on multi-dimensional semantic disentanglement, this paper divides cross-lingual text representations into language-invariant semantic factors, language-specific expression factors, task-causal factors, and non-causal interfering factors, thereby constructing a more stable, interpretable, and transferable cross-lingual causal representation learning method.

2. Related Work and Theoretical Basis

Cross-lingual representation learning is an important foundation of cross-lingual natural language processing. Its core objective is to map texts in different languages into a shared semantic space, so that a model trained on a source language can be transferred to target-language tasks. Existing studies mainly rely on multilingual pre-trained models, cross-lingual word vector alignment, and semantic contrastive learning [1]. These methods reduce representation gaps across languages through shared parameters, shared vocabularies, or parallel corpora. As shown in Figure 1, multilingual texts are first encoded by a unified multilingual encoder, and samples from different languages are transformed into initial semantic representations under shared parameters. This provides the basis for cross-lingual transfer. However, such methods are still largely driven by correlation learning. They tend to mix language expression habits, lexical styles, syntactic structures, task semantics, and noise factors in the same representation space. When the source language and the target language differ significantly in grammatical structure, word order, or cultural expression, the model may rely on language-related surface features for prediction, which weakens cross-lingual generalization [2].

Figure 1: Diagram of the relationship between cross-language semantic decoupling and causal representation learning



Semantic disentanglement learning provides an effective way to address these problems. Its basic idea is to decompose mixed semantic representations into several latent factors with clear meanings, including language-invariant semantics, language-specific features, task-related semantics, and noise factors. As shown in the Semantic Disentanglement module in Figure 1, the model separates shared semantics from language-specific expressions through multi-dimensional disentanglement, enabling cross-lingual tasks to focus more on stable semantic information. Furthermore, causal representation learning emphasizes the identification of stable causal factors that truly affect task outputs, rather than merely capturing surface correlations in the training data [3]. In cross-lingual scenarios, language type and expression form may influence observed text, but the factors that actually determine classification, inference, or matching results should be the core semantics shared across languages. Therefore, this paper combines semantic disentanglement with causal invariance learning. Stable causal semantics are selected through causal factor identification and then used for downstream prediction tasks, thereby improving cross-lingual transferability, robustness, and interpretability [4].

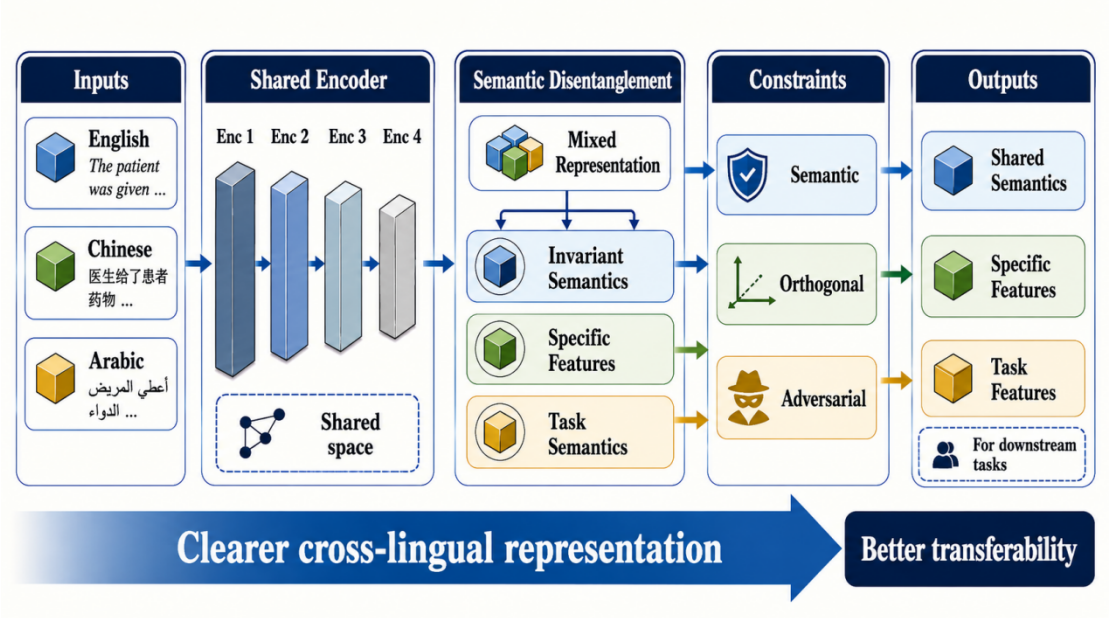
3. Methodology

3.1 Multi-dimensional Semantic Disentanglement Model

To address the entanglement of language features, shared semantics, and task-discriminative information in cross-lingual text representations, this paper constructs a multi-dimensional semantic disentanglement model [5]. Although texts in different languages vary in lexical form, word order, and expression habits, their deep semantic meanings may remain consistent. Therefore, the model should not directly use mixed representations for downstream prediction. Instead, it first decomposes text representations into several semantic subspaces and then reduces information interference among them through constraint mechanisms.

As shown in Figure 2, the proposed model consists of an input layer, a shared encoder, a semantic disentanglement module, a constraint mechanism module, and a disentangled output module [6].

Figure 2: Multi-dimensional Semantic Disentanglement Model for Cross-lingual Text



Specifically, let the cross-lingual text sample be x_i^l, where iii denotes the sample index and lll denotes the language type. Texts in different languages are first fed into the shared encoder to obtain a mixed semantic representation in a unified representation space. This mixed representation contains language-invariant semantics, language-specific expressions, and task-related semantics at the same time [7]. The semantic disentanglement module then decomposes this mixed representation into three subspaces, corresponding to Invariant Semantics, Specific Features, and Task Semantics in Figure 2. This process can be expressed as shown in Formula 1:

$$h_i = E(x_i^l), (z_i^{inv}, z_i^{spe}, z_i^{task}) = D(h_i) \quad (1)$$

where $E(\cdot)$ denotes the shared encoder, which maps texts in different languages into the same representation space; $D(\cdot)$ denotes the semantic disentanglement function; z_i^{inv} represents the language-invariant semantic factor, which mainly preserves the core meaning shared across languages; z_i^{spe} represents the language-specific expression factor, which mainly describes differences in vocabulary, syntax, word order, and language style; and z_i^{task} represents the task-related semantic factor, which mainly preserves discriminative information directly related to classification, inference, or matching tasks. Through this decomposition, the model avoids compressing all information into a single vector space, thereby reducing incorrect associations caused by language differences during cross-lingual transfer [8].

In Figure 2, the split from Mixed Representation into three semantic branches reflects the key design of the proposed model. The language-invariant semantic branch is used to learn stable cross-lingual semantics. For example, although texts in different languages that express “the patient took medication to reduce body temperature” differ in surface form, their core meaning remains consistent. The language-specific expression branch preserves the unique expression patterns of each language, such as Chinese word order, English prepositional structures, and Arabic morphological variation. The task semantic branch focuses on information more directly related to label prediction, such as sentiment words in sentiment classification, logical relations in natural language inference, or semantic correspondence in sentence matching. After these three types of information are modeled separately, the subsequent model can more clearly determine which factors should be used for cross-lingual transfer and which factors should only be retained as language-specific background information [9].

To ensure that the disentangled subspaces have clear boundaries, this paper further introduces semantic constraints, orthogonal constraints, and adversarial learning mechanisms, as shown in the Constraints module

in Figure 2. The semantic constraint requires semantically equivalent cross-lingual samples to remain close in the language-invariant semantic space, thereby enhancing the consistency of shared semantics [10]. The orthogonal constraint encourages different semantic subspaces to be as independent as possible, preventing language-related information from leaking into the shared semantic space. The adversarial learning mechanism weakens the residual language-type information in z_i^{inv} through a language discriminator, making the shared semantic representation more language-invariant. By integrating these constraints, the joint optimization objective of the model can be written as shown in Formula 2:

$$L = L_{task} + \lambda_1 \sum_{(i,j)\in P}(z_i^{inv} - z_j^{inv})_2^2 + \lambda_2((z_i^{inv})^T z_i^{spe})_2^2 + ((z_i^{inv})^T z_i^{task})_2^2 + ((z_i^{spe})^T z_i^{task})_2^2) - \lambda_3 L_{adv} \quad (2)$$

where L_{task}denotes the downstream task prediction loss, which ensures that the disentangled representations still retain task-discriminative ability. PPP denotes a set of semantically equivalent cross-lingual sample pairs, such as parallel sentences, translated sentences, or semantically similar text pairs. The second term is the semantic consistency constraint, which reduces the distance between equivalent samples in the language-invariant semantic space. The third term is the orthogonal constraint, which reduces overlap among language-invariant semantics, language-specific expressions, and task semantics. L_{adv} denotes the language adversarial loss. By maximizing the difficulty of language discrimination, the model suppresses residual language information in the shared semantic space. λ_1, λ_2, and λ_3 are weight coefficients used to balance semantic preservation, subspace separation, and language debiasing.

Through this design, the model preserves stable shared semantics, isolates language-specific expression factors, and highlights task-related semantic information. Finally, it outputs Shared Semantics, Specific Features, and Task Features, as shown in Figure 2. This structure transforms cross-lingual text representation from a mixed representation into an interpretable multi-dimensional semantic representation, providing a foundation for subsequent causal factor identification and stable cross-lingual transfer learning [11].

3.2 Cross-lingual Causal Representation Learning Method

After multi-dimensional semantic disentanglement, the model obtains three types of features: Shared, Specific, and Task. However, these features are not automatically equivalent to causal semantic factors. To identify the semantic information that has a stable influence on task outputs, this paper introduces a cross-lingual causal representation learning method based on disentangled features. The basic assumption is that texts in different languages may differ in vocabulary, grammar, and expression form, but if they convey the same core meaning, their effect on downstream prediction should remain stable. Therefore, the model should reduce its reliance on language-specific surface features and learn causal semantic representations with consistent predictive effects across languages. As shown in the figure above, Shared and Task features are mainly used to construct stable causal factors, while Specific features are weakened or constrained to avoid interference from language-specific information [12].

Specifically, let the language-invariant semantics, language-specific features, and task-related features produced by the semantic disentanglement model be z_i^{sh}, and z_i^{ta}, respectively. This paper uses a causal selection function $C(\cdot)$ to fuse shared semantics and task semantics while suppressing the direct influence of language-specific features on prediction results. The stable cross-lingual causal representation is defined as shown in Formula 3:

$$r_i^c = C(z_i^{sh}, z_i^{ta}) = \sigma(W_c[z_i^{sh}, z_i^{ta}] + b_c) \odot [z_i^{sh}, z_i^{ta}] \quad (3)$$

where r_i^c represents the causal representation of the i-th sample; $[z_i^{sh}, z_i^{ta}]$ denotes the concatenation of shared semantic features and task semantic features; W_c and b_c are learnable parameters; $\sigma(\cdot)$ denotes the gating function; and $\odot$ denotes element-wise weighting. This formula corresponds to the funnel structure in the Causal Learning module shown in the figure. It selects Causal Factors that can stably influence task outputs from multiple types of disentangled features and eventually forms a Stable Representation. Through the gating mechanism, the model strengthens its attention to cross-lingual shared semantics and task-discriminative semantics, while reducing the weight of language-specific expressions in the causal path [13].

To ensure that the learned causal representation remains stable across different language environments, this paper further designs a cross-lingual causal consistency optimization objective. This objective consists of task

prediction loss, cross-lingual invariance loss, counterfactual consistency loss, and language alignment loss, corresponding to the Invariant, Counterfactual, and Alignment mechanism modules in as shown in Formula 4:

$$L_{causal} = L_{task}(f(r_i^c), y_i) + \alpha \sum_{(i,j)\in P} \| r_i^c - r_j^c \|_2^2 + \beta \| f(r_i^c) - f(\tilde{r}_j^c) \|_2^2 + \gamma L_{align} \quad (4)$$

where L_{task} denotes the downstream task loss, ensuring that the causal representation still supports classification, inference, or matching tasks. P denotes a set of semantically equivalent cross-lingual sample pairs, such as an original sentence and its translation, parallel corpora, or semantically similar texts. The second term is the cross-lingual invariance constraint, which reduces the distance between equivalent texts from different languages in the causal representation space. $\tilde{r}_j^c$ denotes the counterfactual causal representation generated by replacing language-specific features, perturbing expression style, or adjusting syntactic form. The third term requires the predictions of the original sample and the counterfactual sample to remain consistent, thereby reducing the model's dependence on language surface forms. L_{align} denotes the language alignment loss, which further constrains the stable semantic distributions across languages. α, β, and γ are weight coefficients used to balance different optimization objectives.

Through this method, the model extracts causally stable semantic factors from disentangled features and strengthens cross-lingual consistency through invariance constraints, counterfactual enhancement, and language alignment. Compared with correlation-based cross-lingual representation learning, this approach focuses more on the semantic factors that truly affect task outputs, thereby reducing language bias and improving robustness, interpretability, and low-resource transfer performance [14].

4. Experimental Design and Results Analysis

4.1 Experimental Setup

To verify the effectiveness of the proposed cross-lingual causal representation learning method based on multi-dimensional semantic disentanglement, the experiments are conducted on four typical cross-lingual tasks: natural language inference, semantic matching, text classification, and sentiment analysis [15]. Representative public cross-lingual datasets are selected, including XNLI, PAWS-X, MLDoc, and the Multilingual Sentiment Dataset. These datasets are chosen because they cover different task types, language ranges, and levels of semantic difficulty, making it possible to evaluate the proposed model more comprehensively in terms of cross-lingual transfer, semantic alignment, and causally stable representation learning. The detailed dataset settings are shown in Table 1.

Table 1: Experimental datasets and task settings

Dataset	Task Type	Main Language Coverage	Experimental Objective	Evaluation Metrics
XNLI	Natural language inference	English, Chinese, Arabic, French, German, etc.	To determine whether a sentence pair expresses entailment, neutrality, or contradiction	Accuracy
PAWS-X	Cross-lingual semantic matching	English, Chinese, Japanese, Korean, French, etc.	To determine whether two sentences have the same meaning	Accuracy / F1
MLDoc	Cross-lingual text classification	English, German, French, Spanish, etc.	To classify news texts by topic	Accuracy / Macro-F1
Multilingual Sentiment Dataset	Cross-lingual sentiment analysis	English, Chinese, Arabic, Spanish, etc.	To identify the sentiment polarity of a text	Accuracy / Macro-F1

As shown in Table 1, XNLI is mainly used to examine the model's cross-lingual transfer ability in complex semantic reasoning scenarios, because natural language inference requires the model not only to understand the meaning of individual sentences but also to identify logical relations between sentence pairs. PAWS-X is more suitable for evaluating the model's ability to capture fine-grained semantic differences, especially when two sentences share similar lexical forms but express different meanings. It can therefore test whether the model has learned stable semantics rather than simply relying on surface-level word overlap. MLDoc is used for cross-lingual text classification and helps evaluate the model's transfer performance in long-text topic recognition. The cross-lingual sentiment analysis dataset is used to test the model's ability to identify

sentiment-related semantic factors, and is particularly useful for analyzing how differences in language style influence model predictions. The experiments adopt a source-language training and target-language testing setting. Specifically, English is mainly used as the source language for model training, and the trained model is then directly transferred to other target languages for zero-shot or few-shot testing. This setting better reflects practical low-resource language scenarios in cross-lingual natural language processing. To ensure the comparability of experimental results, this paper selects mBERT, XLM-R, a standard cross-lingual contrastive learning model, a semantic disentanglement model without causal constraints, and a causal representation model without semantic disentanglement as baseline models. Among them, mBERT and XLM-R represent typical multilingual pre-trained models; the standard cross-lingual contrastive learning model is used to compare semantic alignment performance; and the remaining variants are used to analyze the respective contributions of semantic disentanglement and causal constraints to the final results. In terms of evaluation metrics, classification tasks mainly use Accuracy and Macro-F1. Accuracy measures the overall prediction correctness, while Macro-F1 better reflects balanced performance across different categories, especially when the class distribution is not fully consistent across languages. For semantic matching and inference tasks, this paper mainly focuses on Accuracy, while also using F1 to analyze the stability of the model in distinguishing positive and negative samples. Through the above experimental design, the proposed method can be systematically evaluated from three perspectives: task performance, cross-lingual transfer ability, and model robustness.

4.2 Results Analysis

To comprehensively evaluate the effectiveness of the proposed method, this section analyzes the results from four perspectives: overall performance, cross-lingual transfer ability, ablation study, and interpretability. The experimental results show that the model based on multi-dimensional semantic disentanglement and causal representation learning performs well across different tasks. The improvement is particularly clear in low-resource languages and in languages with large structural differences. This suggests that relying only on multilingual pre-trained models for representation alignment still has certain limitations. By separating language-invariant semantics, language-specific expressions, and task-related semantics through semantic disentanglement, and then selecting stable semantic factors through causal representation learning, the proposed method can effectively reduce the interference of language bias in model prediction.

Table 2: Overall performance comparison of different models on cross-lingual tasks

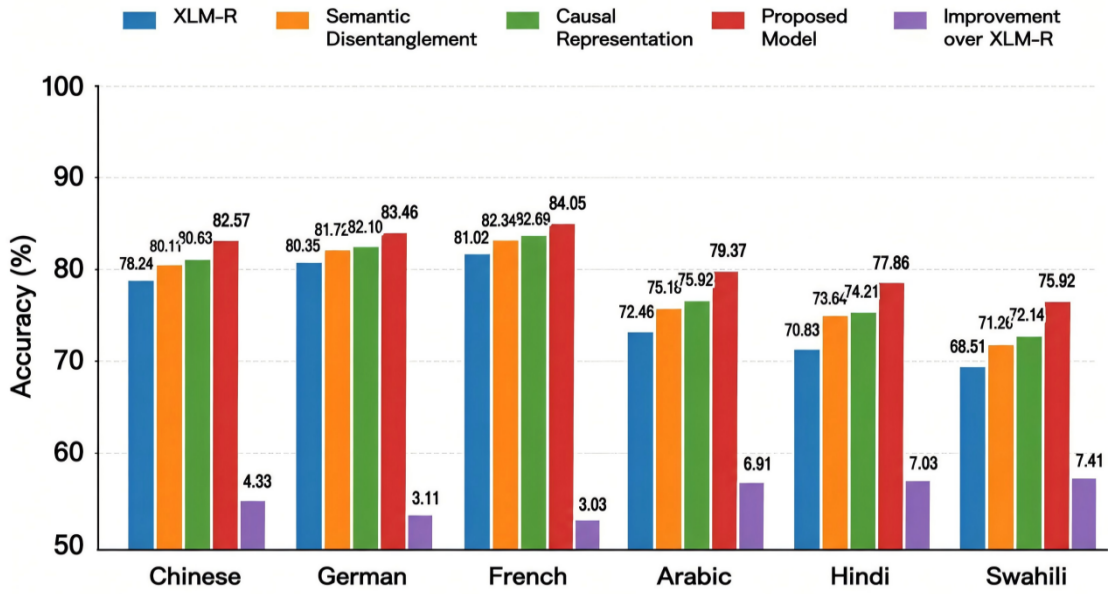
Model	XNLI Accuracy / %	PAWS-X F1 / %	MLDoc Macro-F1 / %	Sentiment Macro-F1 / %	Average / %
mBERT	71.84	78.26	82.15	79.43	77.92
XLM-R	75.62	81.37	85.48	82.76	81.31
Cross-lingual Contrastive Model	76.38	82.04	86.12	83.31	81.96
Semantic Disentanglement Model	77.46	83.19	87.04	84.52	83.05
Causal Representation Model	78.03	83.76	87.45	85.08	83.58
Proposed Model	80.41	86.25	89.36	87.42	85.86

As shown in Table 2, the proposed method outperforms all baseline models across the four cross-lingual tasks. Compared with mBERT and XLM-R, the average performance of the proposed model improves by 7.94 and 4.55 percentage points, respectively. This indicates that although traditional multilingual pre-trained models can provide basic cross-lingual representations, their representation spaces still suffer from semantic entanglement. Compared with the standard cross-lingual contrastive learning model, the proposed method further improves the average result by 3.90 percentage points, showing that simply reducing the vector distance between samples from different languages is not sufficient to guarantee stable transfer. The proposed model also performs better than methods using only semantic disentanglement or only causal representation learning, suggesting that the two components are complementary. Semantic disentanglement separates the mixed semantic space, while causal representation learning further identifies key semantic factors that have a stable influence on task outputs.

In terms of cross-lingual transfer ability, this paper further compares model performance across different target languages. Because languages differ substantially in grammatical structure, lexical morphology, and expression patterns, transfer performance often varies greatly between high-resource and low-resource

languages. Figure 3 presents the results obtained after training on English as the source language and transferring the model to different target languages.

Figure 3: Cross-lingual transfer performance across different target languages



As shown in Figure 3, the proposed model achieves clear improvements in all target languages. The gains are especially prominent for Arabic, Hindi, and Swahili, which are either low-resource languages or languages with larger structural differences from English. This shows that the proposed method does not merely improve the model's fitting ability on high-resource languages. Instead, it improves stability in complex transfer settings through cross-lingual causal semantic learning. For German and French, which are relatively closer to English in terms of resources and linguistic transfer conditions, XLM-R already provides a strong transfer foundation, so the improvement is comparatively smaller. For Arabic, Hindi, and Swahili, however, differences in linguistic structure, morphological variation, and available training resources are more significant. Traditional models are more likely to rely on language-level surface patterns, which leads to weaker transfer performance. By reducing the interference of language-specific expression factors, the proposed method encourages the model to focus more on shared cross-lingual semantics and therefore performs better on low-resource language tasks.

To further verify the function of each module, an ablation study is conducted. The semantic disentanglement module, causal invariance constraint, counterfactual enhancement mechanism, language alignment loss, and orthogonal constraint are removed separately, and the resulting performance changes are observed. The ablation results are shown in Table 3.

Table 3: Ablation study of different model components

Model Variant	XNLI Accuracy / %	PAWS-X F1 / %	MLDoc Macro-F1 / %	Sentiment Macro-F1 / %	Average / %
Proposed Model	80.41	86.25	89.36	87.42	85.86
w/o Semantic Disentanglement	77.92	83.41	86.72	84.93	83.25
w/o Causal Invariance	78.31	83.96	87.08	85.17	83.63
w/o Counterfactual Enhancement	79.02	84.62	87.85	85.94	84.36
w/o Language Alignment Loss	79.18	85.03	88.14	86.21	84.64
w/o Orthogonal Constraint	78.64	84.11	87.36	85.48	83.90

Table 3 shows that removing any key module leads to a decline in performance. The largest average decline occurs when the semantic disentanglement module is removed, indicating that if the mixed semantic space is not decomposed, the model is still affected by the entanglement of language features and task features. When the causal invariance constraint is removed, the decline is more obvious on XNLI and PAWS-X, suggesting that natural language inference and semantic matching rely more heavily on stable semantic relations rather

than surface lexical overlap. The counterfactual enhancement mechanism mainly contributes to robustness. When the expression form changes but the core meaning remains unchanged, the model can still maintain relatively stable predictions. Although language alignment loss and orthogonal constraint bring slightly smaller individual improvements, both are important for reducing language bias and subspace overlap, and they are necessary for effective disentangled representation learning.

In addition to task performance, this paper also evaluates representation quality from the perspectives of interpretability and robustness. By comparing language identification accuracy, semantic clustering purity, and counterfactual consistency across different models, it is possible to judge whether the model truly reduces language bias and strengthens stable semantic expression. Ideally, language identification accuracy in the language-invariant semantic space should be low, indicating that the model no longer retains obvious language-type information. Semantic clustering purity and counterfactual consistency should be high, showing that the model organizes representations around core semantics rather than surface expressions.

Table 4: Interpretability and robustness analysis of learned representations

Model	Language Identification Accuracy ↓ / %	Semantic Clustering Purity ↑ / %	Counterfactual Consistency ↑ / %	Low-resource Language Average Performance ↑ / %
mBERT	84.37	72.18	76.45	70.62
XLM-R	79.52	76.84	80.13	73.93
Semantic Disentanglement Model	68.26	81.47	83.56	75.36
Causal Representation Model	65.74	82.19	85.08	76.14
Proposed Model	54.31	87.62	89.74	77.72

Table 4 further demonstrates the representational advantages of the proposed method. The proposed model achieves the lowest language identification accuracy, which indicates that language-specific information is effectively weakened in the shared semantic space and that the model no longer relies excessively on the language identity of a text for prediction. At the same time, it achieves the highest semantic clustering purity and counterfactual consistency. This means that texts expressing the same or similar meanings in different languages are more likely to be clustered in the same semantic region, and the model can maintain stable predictions even when language expression, sentence structure, or surface wording is changed. This result is consistent with the theoretical design of the proposed method: multi-dimensional semantic disentanglement reduces feature entanglement, while causal representation learning selects stable semantic factors, thereby improving the reliability of cross-lingual transfer.

Overall, the proposed method performs well in overall task performance, low-resource language transfer, module effectiveness, and representation interpretability. Multi-dimensional semantic disentanglement enables the model to distinguish between cross-lingual shared meaning and language-specific expression, while causal representation learning further helps identify stable factors that truly affect task outputs. When these two components are combined, the model not only improves prediction accuracy on cross-lingual tasks, but also reduces language bias, enhances counterfactual robustness, and provides a clearer, more stable, and more interpretable semantic foundation for downstream cross-lingual classification, inference, and matching tasks.

5. Conclusion

This paper proposes a cross-lingual causal representation learning method based on multi-dimensional semantic disentanglement to address semantic entanglement, language bias, and unstable transfer in cross-lingual NLP. The method first uses a shared encoder to obtain multilingual representations, and then decomposes them into language-invariant semantics, language-specific expressions, and task-related semantics. On this basis, causal representation learning is introduced through invariance constraints, counterfactual enhancement, and language alignment loss to identify stable causal semantic factors and reduce reliance on surface language features. Experimental results show that the proposed method performs well in natural language inference, semantic matching, text classification, and sentiment analysis. It also improves low-resource language transfer, robustness, and interpretability, providing a feasible approach for stable cross-lingual understanding.

References

[1] Ki, Dayeon, Park, Cheonbok, and Kim, Hyunjoong. "Mitigating semantic leakage in cross-lingual embeddings via orthogonality constraint." Proceedings of the 9th Workshop on Representation Learning for NLP (RepL4NLP-2024) (2024): 256-273.

[2] Fukushima, Keita, Kajiwara, Tomoyuki, and Ninomiya, Takashi. "Reversible disentanglement of meaning and language representations from multilingual sentence encoders." Proceedings of the 5th Workshop on Multilingual Representation Learning (MRL 2025) (2025): 265-270.

[3] Hua, Tianze, Yun, Tian, and Pavlick, Ellie. "mOthello: When do cross-lingual representation alignment and cross-lingual transfer emerge in multilingual models?" Findings of the Association for Computational Linguistics: NAACL 2024 (2024): 1585-1598.

[4] Hämmerl, Katharina, Libovický, Jindřich, and Fraser, Alexander. "Understanding cross-lingual alignment—A survey." Findings of the Association for Computational Linguistics: ACL 2024 (2024): 10922-10943.

[5] Li, Jiahuan, Huang, Shujian, Ching, Aarron, et al. "PreAlign: Boosting cross-lingual transfer by early establishment of multilingual alignment." Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (2024): 10246-10257.

[6] Wang, Hetong, Minervini, Pasquale, and Ponti, Edoardo. "Probing the emergence of cross-lingual alignment during LLM training." Findings of the Association for Computational Linguistics: ACL 2024 (2024): 12159-12173.

[7] Liu, Danni, and Niehues, Jan. "Middle-layer representation alignment for cross-lingual transfer in fine-tuned LLMs." Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (2025): 15979-15996.

[8] Jung, Haeji, Oh, Changdae, Kang, Jooeon, et al. "Mitigating the linguistic gap with phonemic representations for robust cross-lingual transfer." Proceedings of the Fourth Workshop on Multilingual Representation Learning (MRL 2024) (2024): 200-211.

[9] Gui, Anchun, and Xiao, Han. "Multi-level multilingual semantic alignment for zero-shot cross-lingual transfer learning." Neural Networks 173 (2024): 106217.

[10] Ji, Shaoxiong, Mickus, Timothee, Segonne, Vincent, et al. "Can machine translation bridge multilingual pretraining and cross-lingual transfer learning?" Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (2024): 2809-2818.

[11] Wang, Sicheng, Wu, Wenyi, and Zhang, Zibo. "NeighXLM: Enhancing cross-lingual transfer in low-resource languages via neighbor-augmented contrastive pretraining." Findings of the Association for Computational Linguistics: EMNLP 2025 (2025): 3019-3030.

[12] He, Zhimin, Zhang, Meishan, Zhang, Yue, et al. "Zero-shot cross-lingual document-level event causality identification with heterogeneous graph contrastive transfer learning." Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (2024): 17835-17852.

[13] Hudi, Febri, Mino, Hideya, Imamura, Kenji, et al. "Disentangling pretrained representation to leverage low-resource languages in multilingual machine translation." Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (2024): 4797-4808.

[14] Nguyen, Hoang, Le, Phong, and van Genabith, Josef. "CORI: CJKV benchmark with romanization integration—A step towards cross-lingual transfer beyond textual scripts." Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (2024): 4056-4068.

[15] Zhu, Zhaowei, Li, Xiaoyu, Zhang, Min, et al. “Code-switching can be better aligners: Advancing cross-lingual transfer through representation-level phrase exchange.” Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (2024): 153-160.

Funding

This research was funded by Key Project of Xinjiang Hetian College (No. 2026ZR005).

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights