

Received 7 March 2024, accepted 22 April 2024, date of publication 30 April 2024, date of current version 15 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3395330

RESEARCH ARTICLE

Core-View Contrastive Learning Network for Building Lightweight Cross-Domain Consultation System

JIABIN ZHENG¹, FANGYI XU², WEI CHEN¹, ZIHAO FANG³, AND JIAHUI YAO⁴

¹School of Computer Science, Peking University, Beijing 100871, China

²Institute of Communication Studies, Communication University of China, Beijing 100024, China

³School of Science, The Hong Kong University of Science and Technology, Hong Kong

⁴Institute of Social Science Survey, Peking University, Beijing 100871, China

Corresponding author: Fangyi Xu (xufangyi.yao@gmail.com)

This work was supported by the Art Project of the National Social Science Fund of China under Grant 2022CC02195.

ABSTRACT Cross-domain Consultation Systems have become essential in numerous critical applications, for instance, an online citizen complaint system. However, addressing complaints with distinct orality characteristics often necessitates retrieving and integrating knowledge from diverse professional domains. This scenario represents a typical cross-domain problem. Nevertheless, the prevailing approach of utilizing generative large language models to tackle this problem presents challenges including model scale and drawbacks like hallucination and limited interpretability. To address these challenges, we proposed a novel approach called the Core-View Contrastive Learning (CVCL) network. Leveraging contrastive learning techniques with an integrated core-adaptive augmentation module, the CVCL network achieves accuracy in cross-domain information matching. Our objective is to construct a lightweight, precise, and interpretable cross-domain consultation system, overcoming the limitations encountered with large language models in addressing such challenges. Empirical validation of our proposed method using real-world datasets demonstrates its effectiveness. Our experiments show that the proposed method achieves comparable performance to large language models in terms of accuracy in text-matching tasks and surpasses the best baseline model by over 24 percentage points in F1-score for classification tasks. Additionally, our lightweight model achieved a performance level of 96% compared to the full model, while utilizing only 6% of the parameters.

INDEX TERMS Cross-domain consultation system, cross-domain text matching, multi-view learning, contrastive learning, sentence semantic representation.

I. INTRODUCTION

Numerous crucial applications in today's world necessitate the utilization of consultation systems. Due to the rapid advancement of artificial intelligence, consultation systems have found widespread application across diverse industries. Ban et al. [1] design an earthquake emergency system for multi-user consultation simulation system based on GIS technology, computer technology, and modern communications technology. Xu and Ren [2] propose a new neural network

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Yu.

based method of information extraction for medical online consultation system with a coarse-granularity data annotation approach, which is more time-saving and robust compared with the traditional sequence labeling methods. Sha et al. [3] design an online psychological consultation expert system based on human-computer interaction. Chen et al. [4] propose two frameworks to support automatic medical consultation, namely doctor-patient dialogue understanding and task-oriented interaction. Nonetheless, a distinct category of consultation system in varying domains continues to confront substantial technical challenges. These systems entail consultations and responses characterized by notable

variations in text expressions. Moreover, each consultation often requires the retrieval and integration of knowledge from multiple domains. We refer to these systems as cross-domain consultation systems, which encompass typical domains such as citizen complaints [5] (as shown in Figure 1), medical consultation [4], and legal consultation [6].

Although the cross-domain consultation system is a newly defined concept, the handling of cross-domain information has been extensively discussed in various fields: Minmin et al. [7] proposed a cross-domain privacy enhancement scheme based on multi-blockchain to address privacy protection issues in edge computing, while Salehi et al. [8] combined traditional attribute access control methods with attribute-based group signatures to achieve dynamic access control in cross-domain environments. Yu et al. [9] present SPaC, a dataset for cross-domain Semantic Parsing in Context that consists of 4,298 coherent question sequences (12k+ individual questions annotated with SQL queries). Yuan et al. [10] empirically demonstrates on public datasets that the method achieves the best performance among several state-of-the-art alternative cross-domain recommendation models. Ojha et al. [11] propose to preserve the relative similarities and differences between instances in the source via a novel cross-domain distance consistency loss. The cross-domain consultation systems possess the following characteristics:

- **Variability in text expressions:** In the context of cross-domain consultation systems, user consultations often exhibit imprecision, lack of specialization, and reliance on spoken language, as well as being context-specific in nature. Consequently, the system must possess the capability to retrieve responses from standardized text in professional domains that align with user queries or requirements. To achieve this, the system necessitates robust semantic understanding, reasoning, and context transformation abilities in order to accurately map user expressions to relevant texts within professional domains.
- **Complexity of response mechanisms:** Consultations often entail the querying and integration of knowledge from diverse professional domains.

For such systems, common approaches include rule-based methods [12], generative methods [13], and retrieval-based methods [14]:

Rule-based methods utilize predefined rules and patterns to match and parse user consultations, enabling the system to provide corresponding responses. This approach often necessitates manual authoring and maintenance of an extensive set of rules and templates to cover various domains and types of questions.

Generative methods, on the other hand, learn statistical patterns and correlations between consultations and responses by analyzing large-scale corpora and training data. This approach often employs neural networks to model and predict the relationship between consultations and responses. However, it heavily relies on a significant amount of training

data and may struggle to generalize to new domains and types of consultations. For a cross-domain consultation system, which involves knowledge and semantics from multiple domains, traditional generative methods may not provide accurate and comprehensive responses.

Retrieval-based methods construct and maintain a matching knowledge base to match user consultations with corresponding responses, thus providing responses. The advantage of retrieval-based consultation systems lies in their ability to quickly retrieve relevant responses from large-scale text and exhibit high accuracy for common consultations. Compared to rule-based and generative methods, retrieval-based methods offer more flexibility and adaptability, making them suitable for different domains and types of consultations. However, when it comes to cross-domain scenarios, traditional retrieval matching algorithms such as BM25 [15] and text semantic similarity [16] algorithms may struggle to yield ideal results due to the variability in text expressions and even superficial semantics between the consultations and the ideal matching responses.

Based on the above issues, we proposed a new method called the Core-View Contrastive Learning (CVCL) network. CVCL incorporates multi-view contrastive learning techniques [17] with a core-adaptive augmentation module tailored to the consultation side, effectively capturing the latent semantic connections between consultations and responses (cross-domain-specific knowledge), thereby achieving accurate responses in cross-domain scenarios.

The core-adaptive augmentation module plays a pivotal role in adapting consultation representations. It allows us to transform consultation in a way that effectively narrows the gap between consultation texts and domain-specific knowledge, ensuring accurate matching across domains. Additionally, this module performs fine-grained extraction of core information representing different domains, facilitating the integration of interpretable cross-domain information. This sets CVCL apart from generative large language models, which often face challenges such as hallucinations and poor interpretability when handling such cross-domain problems.

By leveraging the power of the CVCL network, we have developed a lightweight response system that achieves an impressive 96% performance of the full-scale model while utilizing only 6% of its parameters. This significant achievement highlights the importance of our approach.

In this paper, We proposed a novel network architecture called CVCL for cross-domain consultation systems, specifically highlighting its key components and their crucial role in tackling the cross-domain text-matching problem. We conducted comprehensive experiments on real-world datasets to assess the performance of CVCL. The results clearly indicate that CVCL exhibits remarkable performance, surpassing other baseline models in various evaluation tasks. Furthermore, we emphasize the exceptional performance of the lightweight model, which achieves impressive results while utilizing minimal computational resources. The contributions of this article are summarized as follows:

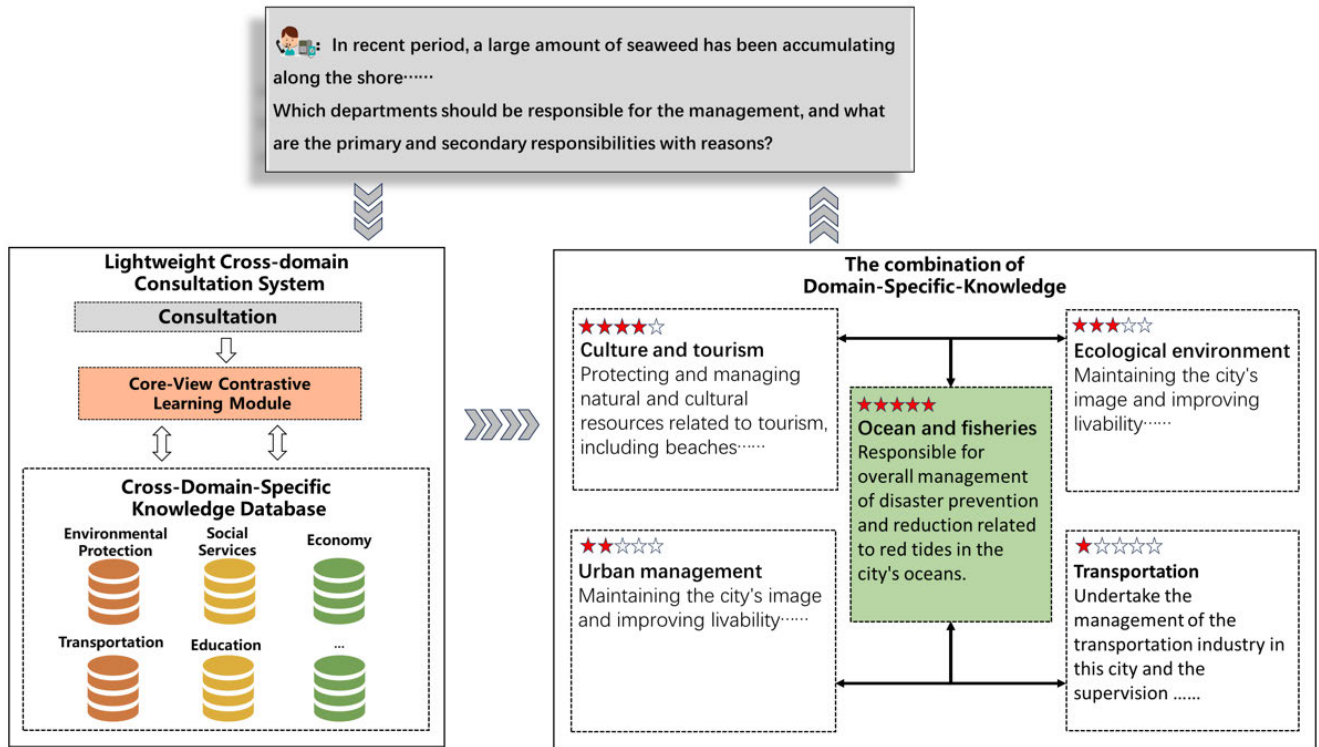


FIGURE 1. An illustration of cross-domain consultation system: For a complaint of citizen, the system returns to the user multiple responsible departments and their corresponding specific responsibilities, and explains the primary and secondary responsibilities.

- We proposed a novel network architecture called CVCL for cross-domain consultation systems, effectively mitigating the challenge of cross-domain text matching.
- By leveraging CVCL, we constructed a lightweight, precise, and interpretable cross-domain consultation system that overcomes the limitations encountered by large language models in addressing such challenges.
- The lightweight solution for a cross-domain consultation system achieved a performance level of 96% compared to the full model while utilizing only 6% of the parameters. c
- We conducted extensive experiments on real datasets to assess the performance of CVCL. The experimental results demonstrate that our proposed method achieves comparable performance of large language models in accuracy for text-matching tasks while surpassing them in terms of issues related to hallucination and interpretability. Additionally, in the classification task, our method outperforms the baseline model by over 24 percentage points in terms of F1-score.

II. RELATED WORK

A. LEARNING FOR SENTENCE EMBEDDING

The learning of universal sentence embedding has been extensively investigated in previous studies. Unsupervised methods such as Skip-Thought [18] leverage the sequential nature of text in books to train an encoder-decoder model that reconstructs the surrounding sentences of an encoded

paragraph. FastSent [19] compares models for learning distributed phrase and sentence representations, finding that the optimal approach depends on the application. Deeper models yield better results for supervised systems, while simpler linear models perform best for spatial distance decoding. FastSent also proposes two new unsupervised learning objectives.

Supervised methods include InferSent [20], which demonstrates the use of labeled data from the Stanford Natural Language Inference dataset to train universal sentence representations. Universal Sentence Encoder [21] proposes a sentence encoding model for transfer learning and showcases the efficiency and accuracy of these models on multiple NLP tasks, highlighting the advantages of sentence embedding in transfer learning. Sentence-BERT [22] introduces SBERT, a method that modifies the pre-trained BERT network structure to generate semantically meaningful sentence representations using siamese and triplet networks.

In recent years, unsupervised contrastive learning methods have emerged for learning sentence embedding. For example, SimCLR adapts the approach of creating semantically similar augmented samples for the same image to the task of learning sentence embedding by designing effective augmentation methods for natural language. ConSERT [23] fine-tunes BERT using unsupervised contrastive learning, addressing the collapse problem of BERT-derived sentence representations and making them more applicable to downstream tasks. DeCLUTR [24] reduces the performance

gap between unsupervised and supervised pre-training by incorporating self-supervised objectives into the pre-training of Transformer models. SimCSE [25] proposes a supervised method that introduces labeled pairs from natural language inference datasets into the contrastive learning framework, using “entailment” pairs as positive examples and “contradiction” pairs as difficult negatives. DiffCSE [26] combines contrastive loss for data augmentation of insensitive examples and replacement detection loss for sensitive examples, resulting in improved sentence embedding. CLSEP [27] introduces a novel data augmentation strategy called partial word embedding augmentation (PWVA) for text data, which enhances data in the word embedding space while preserving more semantic information, leading to better sentence embedding.

However, existing approaches that rely on pairs of semantically similar sentences often struggle to perform well in cross-domain response systems where the matching objects exhibit significant domain differences. General sentence embedding faces challenges when aligning their semantic embedding effectively in such scenarios.

B. CONTRASTIVE LEARNING

Contrastive learning has gained significant attention in the field of machine learning as a powerful technique for learning robust representations. It aims to maximize the similarity between positive samples while minimizing the similarity between negative samples. The concept of contrastive loss was first introduced by Chopra et al. [28], and SimCLR [23] was the pioneering work that applied contrastive loss to self-supervised image recognition tasks. The InfoNCE loss, proposed by Oord et al. [29], is a notable contrastive learning method that enables models to learn representations through negative sampling. This approach has shown promising results in various domains, including computer vision and natural language processing.

C. MULTI-VIEW LEARNING

Multi-view Learning (MVL) refers to a class of methods that leverage the diversity of different views in learning from multi-view data. These views can be derived from various sources or subsets of features [30]. By exploiting consistency and complementarity across multiple views, MVL aims to enhance the generalization capability of models. DICNet introduces an end-to-end framework for multi-view feature extraction, employing stacked autoencoders to learn view-specific representations and utilizing instance-level contrastive learning to extract consensus information from multiple views [31]. Similarly, CMC maximizes the interaction information between different views of the same or different scenes, enabling the learning of embedding representations for images [32]. CPC proposes a self-supervised learning approach that contrasts structural views of graphs to learn node and graph-level representations [33]. In this study, we focus on the concept

of multi-perspective learning, which effectively harnesses multi-view data to learn feature embedding resilient to viewpoint variations, as demonstrated by Shah et al. [34]. Building upon this, ERNIE-VIL 2.0 enhances the robustness and generalization of cross-modal representations by simultaneously modeling the correlations between internal and cross-modal views [17]. Our proposed method, Core-View Contrastive Learning, considers three different views in the consultation system: the original text of the query, the refined result of the query, and the matching response to the query. We perform multi-view learning through contrastive learning, aiming to align semantic embedding by learning the latent co-occurrence semantics of different texts. This approach enhances the robustness of text representations for cross-domain text-matching tasks.

III. PROPOSED METHOD: CORE-VIEW CONTRASTIVE LEARNING NETWORK

The overall architecture of our proposed solution is depicted in Figure 2. Our objective is to achieve precise alignment between consultation text and multiple domain-specific professional texts. To accomplish this, we aim to train an appropriate encoder that aligns consultations and domain-specific professional texts in the semantic embedding space. Traditional semantic similarity models often struggle to achieve desirable results in such cases. To overcome this challenge, we introduce the Core-View contrastive learning network, comprising the core-adaptive augmentation module, Con (Consultation) Encoder, CoA (Core-Adaptive Consultation) Encoder, DSK (Domain Specific Knowledge) Encoder, and multi-view contrastive learning technique. While initially incorporating a large language model in our overall architecture, we propose a lightweight approach that does not rely on the large model during inference, yet still achieves optimal results. In the following sections, we provide a detailed exposition of our proposed methodology.

A. PROBLEM FORMULATION

Our research aims to address the cross-domain consultation task, where, for each consultation, it is necessary to query the existing domain-specific knowledge database to retrieve all relevant professional knowledge and provide interpretable feedback. To tackle this problem, we proposed the CVCL network. The core-adaptive augmentation module in CVCL is designed to transform the consultation by effectively extracting the relevant information and appropriately expanding it, thereby reducing the distance between the textual representations of the consultation and the domain-specific knowledge. We refer to the module’s output as the core-view text. Additionally, we proposed three encoders: the Con Encoder, CoA Encoder, and DSK Encoder, to encode the semantic embedding of the consultation, core-view text, and specific knowledge text, respectively.

Furthermore, to construct a lightweight system, we leveraged multi-view contrastive learning techniques to align the semantic embeddings of the three components. Ultimately,

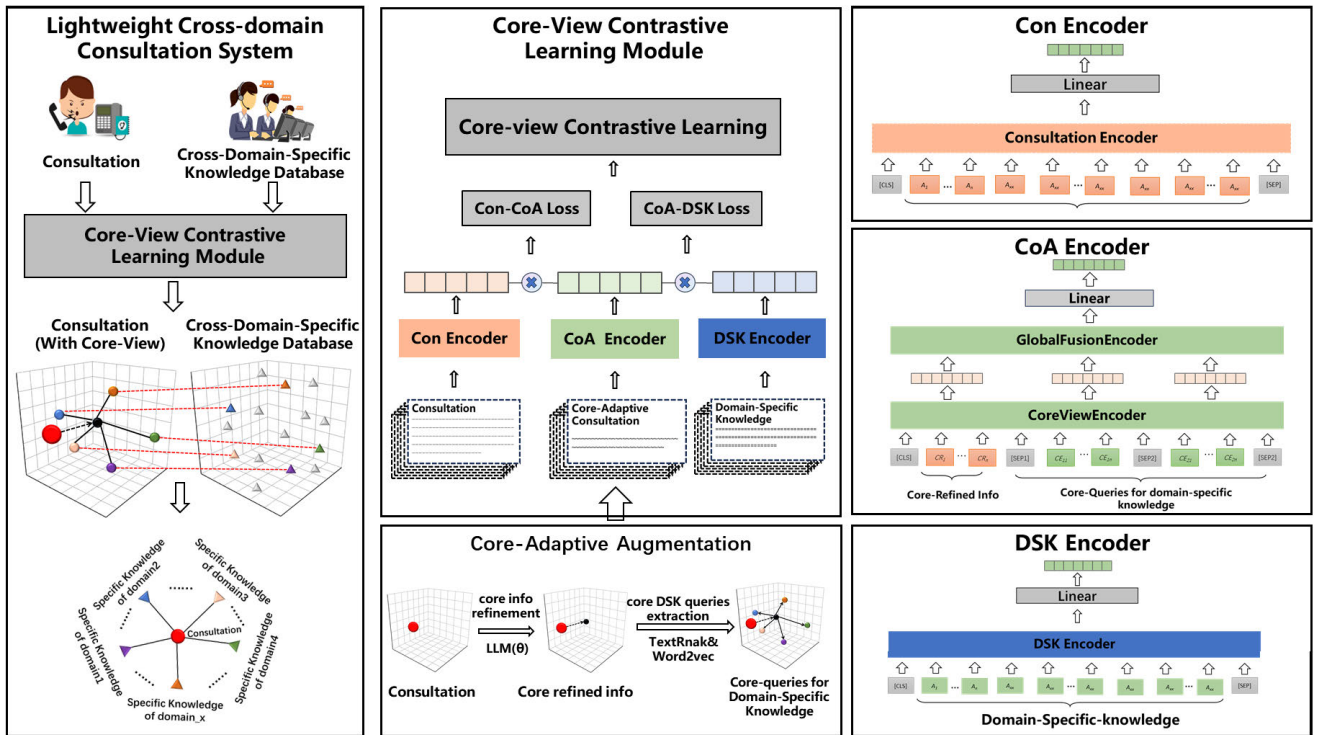


FIGURE 2. The overall architecture of the CVCL Network. The CVCL module extracts query keys from consultations that correspond to the various domain-specific knowledge. This enables precise matching with the content in the domain-specific knowledge database. Additionally, the integration of multi-view learning techniques allows the model to directly construct semantic embedding from consultations that contain diverse interfering information. These embeddings are then precisely aligned with the professional knowledge in different domains, achieving equivalent and accurate matching. As a result, a lightweight system is established.

a retrieval architecture based on the similarity of embeddings was proposed: all domain-specific knowledge was transformed into embeddings, and an embedding retrieval index was constructed using the Approximate Nearest Neighbor Search algorithm [35]. The semantic embedding of the consultation text was used as the query, and the domain-specific knowledge list was ranked based on the similarity of the matches.

B. CORE-VIEW CONTRASTIVE LEARNING MODULE

Our proposed Core-view Contrastive Learning Module comprised the core-adaptive augmentation Module and three encoders: the Con (Consultation) encoder, the CoA (Core-Adaptive Consultation) encoder, and the DSK (Domain-Specific Knowledge) encoder. The core-adaptive augmentation module served the purpose of eliminating irrelevant content based on the original consultation text, extracting more precise and concise core information, and incorporating additional relevant content to minimize the textual domain difference between the user’s consultation content and the matching object from a textual perspective. Ultimately, the multi-view contrastive learning technique is employed to align the semantic embedding of the aforementioned three encoders in the embedding space.

1) CORE-ADAPTIVE AUGMENTATION

The cross-domain response scenario presents two challenges:

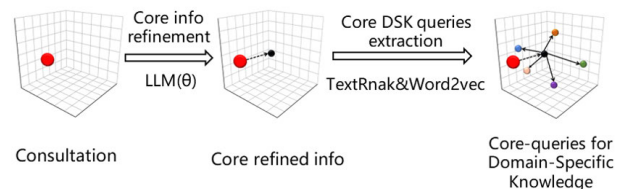


FIGURE 3. Workflow of core-adaptive augmentation, where DSK represents domain-specific knowledge.

- User consultations primarily utilize flexible, colloquial, and contextual language, whereas domain-specific knowledge is usually conveyed through professional and standardized text. This significant difference in text domains creates a challenge when it comes to aligning their semantics.
 - A single user consultation may involve retrieving and integrating specific knowledge from multiple domains.
- As shown in Figure 3, we have developed an adaptive text augmentation method tailored to the characteristics of consultation content. The primary objective of this method is to minimize interference from personal privacy information and irrelevant content while transforming the consultation content into a format that aligns more effectively with domain-specific knowledge. Our proposed approach is as follows:
- Part 1: Core information refinement
In order to address the formidable challenge of removing personal privacy, location information, time, and other

irrelevant details from the consultation while retaining the essential core content necessary for accurate matching with domain-specific knowledge.

To address the challenges posed by the complexity of the task and the limited availability of annotated data, we employ sophisticated techniques that harness the capabilities of a large language model and incorporate meticulously designed prompts. Through the utilization of these techniques, we successfully extract the pivotal elements from the original consultation content and refine them into what we term the Core-Refined (CR) Information. This meticulous process guarantees the retention of essential information required for precise alignment with relevant domain-specific knowledge.

- Part2: Domain specific knowledge queries extraction Our approach includes a crucial step of extracting information modules from the refined core information obtained through core information refinement. These information modules are specifically designed to match domain-specific knowledge.

To identify the essential elements representing the essence of the consultation, we extract keywords and key phrases from the core content. Techniques such as TextRank [36] are employed to assign importance scores to words and phrases based on their co-occurrence patterns. We then generate synonymous expressions for these keywords and key phrases, expanding the range of compatible information modules. This expansion is achieved by leveraging techniques like Approximate Nearest Neighbor (ANN) indexing with word2vec embeddings [37]. These information modules are carefully crafted to align with the specific knowledge and terminology of different domains.

In summary, the core information refinement step focuses on distilling the consultation to its essential core, while the domain-specific knowledge queries extraction step tailors queries for accurate matching with domain-specific knowledge. These steps offer significant value by addressing the challenges of removing irrelevant information, ensuring precise alignment with different domains, and providing interpretable results that highlight the relevance and correspondence between the consultation and specific domains.

2) CON ENCODER & DSK ENCODER

The Con Encoder and DSK Encoder are used to obtain the semantic embeddings for the consultation and domain-specific knowledge, respectively. We utilize the RoBERTa architecture and parameters for both the Con Encoder and DSK Encoder in this work, RoBERTa shares the same underlying architecture as BERT [38], but it further optimizes key hyperparameters and benefits from pretraining on a larger dataset. As a result, RoBERTa has emerged as the state-of-the-art approach for various natural language processing tasks. following the approach proposed in RoBERTa [39]. By leveraging the RoBERTa model, we extracted sentence embedding by retrieving the output of the CLS token and

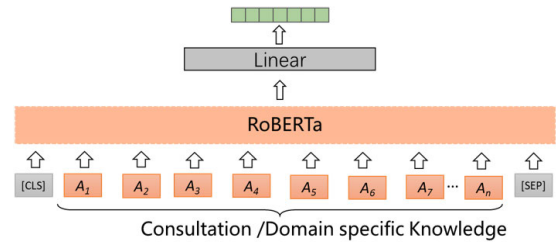


FIGURE 4. The architecture of the Con Encoder and DSK Encoder.

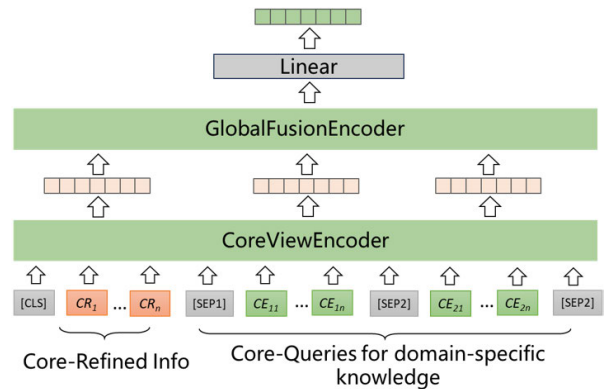


FIGURE 5. The architecture core-view encoder.

passing it through a linear layer. The network structure is illustrated in Figure 4.

3) COA ENCODER

As shown in Figure 5, the CoA Encoder plays a crucial role in incorporating Core-Refined information and domain-specific knowledge queries (Core queries) to generate comprehensive embeddings of the aforementioned content, thereby obtaining a global semantic embedding. To further enhance the accuracy and interpretability of the entire system, we independently encode each domain-specific knowledge query (Core-Query) as a matching query tailored to a specific domain. Subsequently, we combine the results obtained from both types of embeddings to produce a combination of domain knowledge, which serves as the matching result for the query sentence.

We proposed a novel approach, the Two-stage BERT Encoder, to tackle the challenges mentioned above. In the input stage, we employed a separator cls from the start of each segment to distinguish the Core-Refined information and domain-specific knowledge queries, followed by the first-stage encoding. We extract the embedding of the CLS token, representing the embedding representation of that particular segment. These embeddings are then passed to the subsequent stage of the BERT model to generate the global embedding representation. This hierarchical encoder comprises two key components: the CoreViewEncoder and the GlobalFusionEncoder.

The proposed Two-stage BERT Encoder addresses the need for a comprehensive encoding strategy that captures

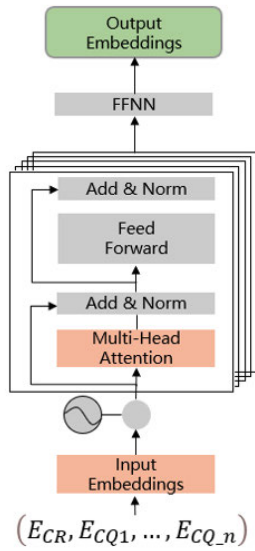


FIGURE 6. The architecture BERT encoder.

both Core-Refined information and domain-specific knowledge queries. By separating and encoding these components separately, we can effectively capture their respective semantics and later fuse them for a more holistic representation. The specific architecture of the CoA Encoder is as follows:

- CoreViewEncoder: As the first-stage encoder, the role of the CoreViewEncoder is to accept the results of the Core-Adaptive Augmentation, which include the Core-Refined information representing the core information of the consultation and the Core-Queries for domain-specific knowledge representing the matching queries for specific domain knowledge. The Core-ViewEncoder then outputs the semantic embeddings of these components. The output can be represented as:

$$D = (E_{CR}, E_{CQ1}, \dots, E_{CQn}) \quad (1)$$

Here, E represents the generated embedding of each sub-element. E_{CR} represents the embedding of the Core-Refined information, and E_{CQ_i} represents the embedding of the i-th group of Core-Queries.

- GlobalFusionEncoder: As shown in Figure 5, the second-stage encoder is used to combine and extract information from the multiple sentence embedding mentioned above, serving as the overall representation of the core information. In the second stage, the input consists of the encoded embedding from the first stage, which is then passed through a linear layer to obtain the final global representation, which we call as core-view embedding.

The CoA Encoder consists of two modules, both of which are built around the BERT architecture. Here are the details of the BERT:

As shown in Figure 6, the core component of the CoA Encoder's CoreViewEncoder module is the Multi-Head Attention mechanism. The Multi-Head Attention layer is

calculated as follows:

$$\text{BertAtt}(D) = \text{LayerNorm}(D + \text{MultiHead}(D)) \quad (2)$$

$$\text{MultiHead}(D) = \text{Concat}(\text{head}_1, \dots, \text{head}_n) \times W^O \quad (3)$$

where h is the number of heads for Bert self-attention, LayerNorm refers to layer normalization, $\times W^O$ is the weight matrix for dimensional transformation, and each head_h is computed as follows:

$$\text{head}_i = \text{Softmax} \left(Q_i \times K_i^T / \sqrt{d_e} \right) \times V_i \quad (4)$$

$$Q_i = D \times W_i^Q \quad K_i = D \times W_i^K \quad V_i = D \times W_i^V \quad (5)$$

where W_i^Q , W_i^K , W_i^V are the weight matrices of the i-th header; and $\text{Softmax} \left(Q_i \times K_i^T / \sqrt{d_e} \right)$ is an $m \times m$ matrix, where the entries in the a-th row and jth column denote the attentional weights of the a-th sentence to the jth sentence. Here, V_i contains the information passed by the sentences to the subsequent layers, while the attention weight matrix $\text{Softmax} \left(Q_i \times K_i^T / \sqrt{d_e} \right)$ acts as a gating mechanism, controlling the amount of information passed (i.e., after multiplication, it is difficult to pass information to those sentences with low attention scores). Next, the output of BertAtt is passed through a standard feed-forward neural network using a residual mechanism and layer normalization computed as follows:

$$D_1 = \text{MultiHead}(D) \quad (6)$$

$$D' = \text{LayerNorm} \left(D_1 + \text{Relu} \left(D_1 \times W^r \right) \times W^S \right) \quad (7)$$

The calculation of the final global embedding S, based on the intermediate document representations Z formed by multiple identical BERT layers, is as follows:

$$S_1 = \lambda_1 \text{FFNN} \left(\text{Avg}(z) \times W^t \right) \quad (8)$$

$$S_2 = \lambda_2 E_{CR} + \lambda_3 E_{CQ1} + \dots + \lambda_3 E_{CQ_n} \quad (9)$$

$$S = S_1 + S_2 \quad (10)$$

where W^t is a weight matrix used for linear transformation, Z is the output of the final BERT layer, Avg represents the average pooling layer, and S represents the global semantic representation of the sentence-level BERT encoder. Finally, we compute the weighted sum of the global semantic representation and all Core-view representations(Core-Refined embedding and Core-Queries embedding) as the final embedding S. λ_1 and λ_2, λ_3 represent the weights for these parts. We do this to enhance the semantic representation capability of the Core-Query representations enabling them to match domain-specific knowledge. This serves as an interpretable basis for reasoning, demonstrating the correspondence between the various components in the consultation and different domain-specific knowledge.

4) MULTI-VIEW CONTRASTIVE LEARNING

To learn the relationship between Consultation, Core-Adaptive Consultation, and Domain Specific Knowledge, we employ a multi-view contrastive learning technique to

align them in the spatial embedding space. During each iteration of the dataset S , we construct two sets of data pairs:

$$S = \{(Con, CAC), (CAC, DSK)\}$$

In the provided context, “Con-CAC” represents the data pairs of Consultation and Core-Adaptive Consultation, and “CAC-DSK” represents the data pairs of Core-Adaptive Consultation and Domain Specific Knowledge. Concerning the positive examples in contrastive learning, we utilize the core-adaptive augmentation module to obtain the Con-CAC data pairs. For the CAC-DSK data pairs, we employ the Uncertainty-Guided Data Annotation method [40].¹ Negative examples are generated by randomly selecting other data within the same batch. The three sets of data are then passed through their respective encoders to obtain embedding representations, which are used to learn the relationships between the data pairs. Following the principles of InfoNCE [30], we define the loss as follows:

$$L_{(x,y)} = -\frac{1}{N} \sum_i \log \frac{\exp(h_x^{i\top} h_y^i / \tau)}{\sum_{j=1}^N \exp(h_x^{i\top} h_y^j / \tau)} \quad (11)$$

$$L_{\text{multi-view}} = \sum_{s \in S} \lambda_s L_s \quad (12)$$

In the provided context, x and y represent a pair of texts from each view in S , h represents the embedding encoded by the respective encoders, τ is the temperature used to scale the logits, and N represents the number of data pairs in a batch. We aim to minimize the overall contrastive loss across all view pairs by scaling the different losses with the parameter λ . Here, s belongs to S and represents a type of view pair within S .

5) THE LIGHTWEIGHT CROSS-DOMAIN CONSULTATION SYSTEM

By utilizing the aforementioned modules, we can construct a complete cross-domain consultation system. The construction of the final system involves two core steps: recall and ranking. Firstly, we utilize the DSK Encoder to build an embedding index for cross-domain-specific knowledge. In the recall phase, we perform retrieval using core-view embedding (the output of the CoA Encoder). In the ranking phase, we calculate the embedding similarity [41] between the core-query embeddings² and the specific knowledge embedding retrieved in the recall phase. This process yields the final matching results.

However, the core information refinement component of the core-adaptive augmentation heavily relies on a large language model. In order to build a lightweight system, we proposed a model based on a multi-view contrastive learning technique. This technique enables the alignment of the core-view embedding with domain-specific knowledge and

¹Detailed descriptions are provided in the subsequent “Complaint and Responsibility” section.

²encoding results of the Core-View Encoder in the CoA Encoder.

also aligns the consultation with the core-view embedding. As a result, we can directly utilize the Con Encoder to obtain the encoding embedding, which serves as the global semantic embedding for recall. The ranking phase remains unchanged. Through this approach, we achieve a lightweight version of the cross-domain consultation system.

IV. EXPERIMENTS

A. CITIZEN COMPLAINT CONSULTATION EXPERIMENTS

To validate the effectiveness and assess the practical application of this approach, this study received data support from a large city in China with a population exceeding 10 million. Ultimately, in the context of citizen complaints, we verified the feasibility of the proposed method based on real data. In a citizen complaint consultation system, a complaint may involve multiple departments with different responsibilities in various domains. Some departments bear primary responsibilities, while others have secondary responsibilities. The citizen complaint consultation system needs to identify the corresponding responsible departments based on the complaint content and provide interpretable justifications.

1) CITIZEN COMPLAINT DATA

The citizen complaint data used in the study is derived from real-world scenarios and has undergone anonymization for privacy protection. Each citizen complaint entry includes fields such as complaint ID, original complaint text, and primary responsible department. After undergoing data cleansing, a total of 24,000 complaint records remain for analysis.

2) GOVERNMENT RESPONSIBILITY DATA

The government responsibility data is obtained from legal documents, organizational regulations, normative documents, and other sources. A total of 120,000 records were extracted and annotated with responsible departments. The dataset covers 1,051 departments, including but not limited to the Ecological Environment Bureau, Urban Management Bureau, Market Supervision Bureau, and other departments. The responsibilities span various fields such as the economy, environment, education, and more.

3) COMPLAINT AND RESPONSIBILITY PAIR DATA

Since we only have access to customer complaints and their corresponding primary responsible departments, this results in a lack of matching between complaints and specific responsibilities. we adopt a combination of a large model and manual annotation to obtain training data. The core objective is to identify one or more government responsibilities that match user complaints. We first construct initial complaint-responsibility data pairs for discrimination. Then, we use the large model to determine their consistency, and for a small portion where the model’s performance is inadequate, we resort to manual annotation.

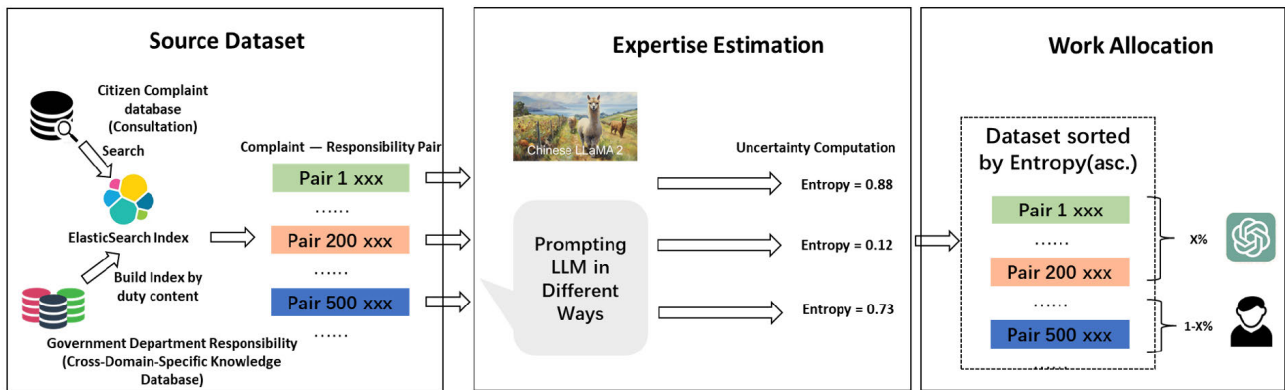


FIGURE 7. The Pipeline of Uncertainty-Guided Data Annotation.

As shown in Figure 7, we employ the CoAnnotating method [40] for annotation. We use a set of prompts $P_i = P_{i1}, P_{i2}, \dots, P_{ik}$, where each prompt guides the large model³ to perform annotations in different ways. We calculate the uncertainty u_i of the large language model (LLM) to guide the work allocation process. We employ two easily implementable and previously proven effective methods to compute u_i . In each case, by prompting the LLM k times with different prompts p_{ij} , we obtain k annotations $A_i = a_{i1}, a_{i2}, \dots, a_{ik}$ for each instance. We calculate the uncertainty of each prediction by computing the entropy, which measures the impurity within a set of data and can be used to quantify the uncertainty related to class labels. The higher the entropy value, the greater the corresponding uncertainty. We can utilize this measure to estimate the level of uncertainty:

$$u_i = - \sum_{j=1}^k P_{\theta}(a_{ij} | p_{ij}) \ln P_{\theta}(a_{ij} | p_{ij}) \quad (13)$$

where $P_{\theta}(a_{ij} | p_{ij})$ represents the frequency of a specific prediction among all predictions. Ultimately, we obtained 100,000 training data pairs, with only 2% of the data being manually annotated. We conducted random spot checks on the final dataset and found that the qualification rate was 98%.

In the training phase, we randomly select 95,000 records from the 100,000 available in the dataset, with the remaining 5,000 used for model evaluation.

B. RESULTS ANALYSIS

1) TASK1: CORE-ADAPTIVE AUGMENTATION

We conducted a preliminary analysis of the effectiveness of Core-Adaptive augmentation. Firstly, we calculated the text similarity of the complaint and responsibility pairs in our experiment to verify the significant differences in consultation and domain-specific knowledge mentioned earlier. Additionally, we compared the traditional text similarity training data to further demonstrate the initial differences in

³the Chinese-Llama-2-7b model.

TABLE 1. Result of core-adaptive augmentation: Comparison of similarity metrics.

| Dataset | Jaccard Coeff. | TF-IDF | Lev. Distance | Bert Score |
|-------------------|----------------|---------------|---------------|---------------|
| STS-B-3 | 0.4535 | 0.4258 | 0.6170 | 0.4563 |
| STS-B-4 | 0.5117 | 0.4963 | 0.6797 | 0.7742 |
| STS-B-5 | 0.6139 | 0.6068 | 0.7766 | 0.8637 |
| w/o Core-Adaptive | 0.0722 | 0.0454 | 0.0666 | 0.0762 |
| w/o Core refine | 0.1269 | 0.0930 | 0.2580 | 0.1832 |
| w/o core-queries | 0.0938 | 0.0835 | 0.2353 | 0.2167 |
| Our method | 0.1825 | 0.1978 | 0.3269 | 0.1363 |

the text domain. Finally, we compared the similarity between the Core-Adaptive augmented core-adaptive consultation data and domain-specific knowledge. In the comparative analysis, we selected the STS-B [42] dataset as the benchmark. We divided the dataset into three categories: STS-B-3, STS-B-4, and STS-B-5, based on the scores of 3 or above, 4 or above, and 5, respectively.

Based on the results of task 1 shown in Table 1, we have the following conclusions:

- The matching data in the cross-domain consultation system exhibits significant differences in the textual domain compared to traditional text similarity data, which serves as a reason for not validating our model's performance on datasets like STS-B. Additionally, pretrained models struggle to capture their shared semantic features.
- Core-Adaptive Augmentation directly helps improve the similarity of matching data in the textual domain of the cross-domain consultation system. However, there is still a fundamental difference in similarity compared to open-source text similarity data. Further supervised training is required to learn new representations and achieve semantic alignment between consultations in the cross-domain system and their corresponding domain-specific knowledge.

2) TASK2: COMPLAINT AND RESPONSIBILITY MATCHING

We conducted an experiment to determine the matching between citizen complaints and department responsibility.

We calculated the cosine distance between the semantic embedding of the two and set a threshold⁴. If the distance is above the threshold, it is considered a match; otherwise, it is not. We used accuracy [43] as the evaluation metric:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (14)$$

The baseline models for the comparative experiments are a series of unsupervised text similarity models: ESimCSE [25], DCLR [44], DiffCSE [26], GSInfoNCE [45], InfoCSE [46], InforMinCL [47], and MixCSE [48].

- ESimCSE (Contrastive SimCSE): It constructs positive and negative example pairs by modifying positive and negative sentences to weaken the sentence length information implicitly encoded in the SimCSE method using BERT for sentence representation.
- DCLR (Debiased Contrastive Learning of unsupervised sentence Representations): It uses instance weighting to penalize incorrect negative examples and generates noise-based negative examples to address biases caused by within-batch negatives or random negatives, ensuring consistency in the representation space.
- DiffCSE (Difference-based Contrastive Learning for Sentence embedding): It randomly masks original sentences and samples edited sentences from a masked language model, then learns the subtle differences between the original and edited sentences. DiffCSE utilizes equivariant contrastive learning, learning sentence embedding representations that are insensitive to certain types of data augmentation and sensitive to other “harmful” data augmentations.
- GSInfoNCE (Gaussian Smoothed InfoNCE): It is an improvement over InfoNCE. Due to the existence of negative pairs with false negatives, performance may degrade when the batch size exceeds a threshold. Therefore, GSInfoNCE introduces Gaussian noise on top of InfoNCE, expanding the negative pairs by adding random Gaussian noise without increasing the batch size.
- InfoCSE (Information-aggregated Contrastive Learning of Sentence embedding): It is an information-aggregated contrastive learning framework for learning unsupervised sentence embedding. InfoCSE introduces an additional masked language model task and a specialized network to make the [CLS] position of the model contain more densely aggregated sentence information.
- InforMinCL (Information Minimization based Contrastive Learning): It is a contrastive learning model for sentence embedding based on information minimization. It maximizes mutual information between positive instances and minimizes information entropy between positive instances to preserve useful information and reduce redundant information in sentence embedding caused by maximizing mutual information between positive instances in contrastive learning.

⁴In our experiments, the threshold used is 0.57.

TABLE 2. Result of complaint and responsibility Matching Task(Baseline models).

| Model | Consultation | Core-Adaptive Consultation |
|------------|--------------|----------------------------|
| ESimCSE | 0.59274 | 0.61506 |
| DCLR | 0.44334 | 0.49374 |
| DiffCSE | 0.62694 | 0.66096 |
| GSInfoNCE | 0.52326 | 0.57114 |
| InfoCSE | 0.7056 | 0.819 |
| InforMinCL | 0.36846 | 0.45774 |
| MixCSE | 0.60066 | 0.67194 |

TABLE 3. Result of consultation and responsibility Matching Task(Our model), CIR means Core-view info refinement, CDQE means Core-DSK queries extracton.

| Model | Consultation | Core-Adaptive Consultation |
|---------------|--------------|----------------------------|
| w/o CIR | 0.72446 | 0.79574 |
| CVCL w/o CDQE | 0.54186 | 0.60346 |
| CVCL | 0.86086 | 0.90354 |
| LLM | 0.87353 | 0.91624 |

- MixCSE (contrastive learning with mixing negatives): It improves model performance by constructing challenging negative examples, overcoming the issue of randomly sampled negatives having less impact on sentence representations compared to challenging negatives.

From the results in Table 2, it can be observed that, among all the baseline models, incorporating the core-adaptive augmentation led to significant improvements in the final results.

To validate the effectiveness of our proposed approach, we conducted a series of ablation experiments. The results, as shown in Table 3, demonstrate that both our core information refinement and core DSK query extraction approaches have a positive impact. Specifically, the core information refinement approach plays a significant role, and the evaluation of core queries for the Domain-Specific Knowledge Matching Task confirms the system’s good interpretability. Our system not only identifies multiple domain-specific knowledge items for consultation but also directly finds the matching sub-elements between the consultation and each domain-specific knowledge item.

Furthermore, we also compared the performance of our system with a large language model⁵ on this task. The results demonstrate that our system performs remarkably well, achieving a performance level of only 1.4 percentage points lower than the large language model.

The model was trained using PyTorch on a Linux system, with computations accelerated by an NVIDIA GeForce RTX 3090 GPU. The training process spanned 5 epochs and took a total of 50 minutes. It utilized a batch size of 64 and a maximum sequence length of 256.

⁵the Chinese-Lama-2-7b.

TABLE 4. Result of Classification of responsible department Task(CIR means Core-view info refinement, CDQE means Core-DSK queries extraction).

| Model | Pre with Consultation | | | Pre with Core-Adaptive Consultation | | |
|---------------|-----------------------|---------------|---------------|-------------------------------------|---------------|---------------|
| | Precision | Recall | F1 | Precision | Recall | F1 |
| ESimCSE | 0.2957 | 0.3820 | 0.3334 | 0.3327 | 0.4135 | 0.3687 |
| DCLR | 0.2463 | 0.2220 | 0.2335 | 0.2773 | 0.2758 | 0.2765 |
| DiffCSE | 0.3150 | 0.1820 | 0.2307 | 0.3417 | 0.2365 | 0.2795 |
| GSInfoNCE | 0.3240 | 0.2340 | 0.2717 | 0.3750 | 0.2653 | 0.3108 |
| InfoCSE | 0.3253 | 0.3860 | 0.3531 | 0.4520 | 0.3920 | 0.4199 |
| InforMinCL | 0.1963 | 0.1890 | 0.1926 | 0.2550 | 0.2284 | 0.2410 |
| MixCSE | 0.4003 | 0.1750 | 0.2435 | 0.4733 | 0.3391 | 0.3951 |
| CVCL w/o CIR | 0.6036 | 0.3640 | 0.4541 | 0.7545 | 0.4553 | 0.5679 |
| CVCL w/o CDQE | 0.5058 | 0.3882 | 0.4393 | 0.6323 | 0.4857 | 0.5494 |
| CVCL | 0.7471 | 0.5634 | 0.6423 | 0.7713 | 0.5917 | 0.6697 |

3) TASK3: CLASSIFICATION OF RESPONSIBLE DEPARTMENT

Our goal in this task is to classify the responsible department for every complaint. Firstly, we obtained sentence embedding for citizen complaints and constructed an embedding index for all department duties. Based on this index, we retrieved the top-k most similar duties and selected the department with the highest frequency as the final classification result. For evaluation, we used Macro-Precision [43], Macro-Recall [43], and Macro-F1 score as the metrics:

$$\text{Precision}_M = \frac{\sum_i i = 1^l \frac{TP_i}{TP_i + FP_i}}{l} \tag{15}$$

$$\text{Recall}_M = \frac{\sum_i i = 1^l \frac{TP_i}{TP_i + FN_i}}{l} \tag{16}$$

$$\text{F1score}_M = \frac{2 * \text{Precision}_M * \text{Recall}_M}{\text{Precision}_M + \text{Recall}_M} \tag{17}$$

The results from Table 4 highlight the effectiveness of our core-adaptive augmentation module in improving the outcomes of Task 3. Initially, we employed the same set of baseline models as in Task 2 to complete Task 3, demonstrating the effectiveness of the core perspective enhancement. Incorporating this enhancement led to significant improvements among all the baseline models. Furthermore, through ablation experiments using our model, we showcased the positive effects of both the core-view information refinement and core-DSK queries extraction modules within our core-adaptive augmentation module. These components collectively contributed to a notable improvement in overall performance. improvements in the final outcomes.

We also conducted an analysis of successful and unsuccessful cases. In Table 5, we listed successful cases, including instances where the baseline model made incorrect predictions but CVCL made correct ones. Through our analysis, we found that our model outperforms CVCL, particularly in cases where complaints are too disorganized and contain excessive irrelevant information. Table 6, on the other hand, presents unsuccessful cases. From Table 6, it can be observed that some complaints involve multiple

departments, each bearing varying degrees of responsibility. When calculating accuracy, we primarily focused on the top-1 accuracy (classification of the main responsible department). These cases represent the challenging aspect of this task and are where our model currently shows poorer performance.⁶

4) TASK4: CORE-QUERIES FOR DOMAIN-SPECIFIC KNOWLEDGE MATCHING

We aimed to demonstrate the model’s performance in terms of interpretability. We utilized the Core-queries for Domain-Specific Knowledge from the core-adaptive augmentation module as domain query embedding. These embedding were then compared with the responsibility embedding obtained from the system. The closest distance was considered the final explanation result. To evaluate this matching process, since there was a lack of supervised data, we employed a manual evaluation method. For Task4, it is not directly comparable to other tasks as it was designed to demonstrate the system’s performance in interpretability. Therefore, no other baselines or ablation experiments were conducted for this task. Instead, we manually evaluated 500 inference results and achieved an accuracy of 92% in matching.

5) ANALYSIS OF THE PERFORMANCE OF THE LIGHTWEIGHT MODEL

According to the comprehensive comparison shown in Figure 8, our lightweight model, utilizing only 6% of the parameters of the full model, achieves performance equivalent to 96% of the full model’s performance in both Task 2 and Task 3. Compared to the best-performing baseline model, the lightweight model exhibits a performance improvement of 53%. This result highlights the advantages of the lightweight system, as it achieves near-full model performance while maintaining a significantly smaller parameter size. Regarding the training and testing time, since both the full-scale model

⁶The dataset used in this study was originally in Chinese. To facilitate international readability of the paper, all data examples and tables have been translated into English. The translation was carried out by the author of this paper, Jiabin Zheng.

TABLE 5. The successful cases of the proposed model and other comparison algorithms.

| consultation | Classification of Responsible Department | | |
|---|--|---------------------------------------|---------------------------------------|
| | Baseline | Our model | Label |
| Construction work is conducted daily nearxx School on xx Road, North District of the city. The noise is very loud, severely disturbing residents. I am dissatisfied with this situation and request to file a complaint. I urge the relevant department to promptly coordinate with the construction site to reduce the noise. | Environmental Protection Bureau | Municipal Urban Management Bureau | Municipal Urban Management Bureau |
| On February 6, 2021, I was assaulted by someone and reported the incident to the XXX police station. However, the police station has yet to resolve the issue. Furthermore, police officers at the station insulted and assaulted me. Additionally, staff in the duty room of Xifu Town Police Station assaulted me at the entrance. I am dissatisfied with this situation and wish to file a complaint. I request the police station to provide a copy of my statement, verify whether the police officer insulted me, and provide a reasonable explanation regarding the alleged detention by the head of Xifu Town Police Station. | Bureau of Justice | Public Security Bureau | Public Security Bureau |
| On December 31, 2021, I booked a room at XX Hotel, located near XX Road, Pingdu City, through the XX platform using my mobile phone number xxxxxx. At 17:00, due to changes in my schedule, I requested a refund, which the hotel agreed to. However, when I submitted the request on Meituan, the hotel refused and instructed me to contact Ctrip. I made the booking through Meituan and I am dissatisfied with this. I request the department to coordinate the refund as soon as possible | Bureau of Culture and Tourism | Market Supervision Administration | Market Supervision Administration |
| I am a resident of X Village, X Street. My family owns three properties. After a unified census by the village committee in 2021, the information was recorded in the Housing Management Bureau. However, we have not yet received the property ownership certificates. As a result, my wife is unable to register as a resident. I am dissatisfied with this situation and request that the relevant department coordinate the issuance of the property ownership certificates. I hope the relevant authorities will address this matter and provide a response." | Development and Reform Commission | Natural Resources and Planning Bureau | Natural Resources and Planning Bureau |

TABLE 6. The failed cases of the proposed model and other comparison algorithms.

| consultation | Classification of Responsible Department | | |
|--|--|--|-----------------------------|
| | Baseline | Our model | Label |
| My child is studying at XXX School, and I am dissatisfied with the poor quality and insufficient quantity of food provided in the school cafeteria. Additionally, I am concerned that leftover food is being reused. I am calling to file a complaint and request that the relevant department address and respond to this matter. | Market Supervision Administration | Health Commission | Education Bureau |
| I have established a company in Shibe District, but I have not registered for tax. I inquired about the procedures with the relevant department, asked about the required documents, and received clarification. | Market Supervision Administration | Market Supervision Administration | Tax Bureau |
| I reside in XX Town, XX District. On April 29, 2022, I paid 3850 yuan for driving lessons through XX company’s platform. I have already passed the theoretical exam (subject one), but I have not been scheduled for the practical training (subject two) so far. After contacting the company to request a refund, they refused, which I believe is unreasonable. I am calling to request a prompt refund. | Finance Bureau | Development and Reform Commission | Transportation Bureau |
| Mr. X complained about an employee named X at the legal aid center on June 14th. The employee did not have a work permit. Mr. X reported a fraud issue, but the lawyer informed him that it was a regular loan. He believes that the employee’s response does not align with national legal policies and suggests that the employee might not be well-versed in the new Civil Code. | Ecological Environment Bureau | Human Resources and Social Security Bureau | Municipal Bureau of Justice |
| Mr. X reported: His boat, placed at the entrance of XXX, was confiscated by fishery administration personnel. The XX Police Station told him to bring the necessary documents to sail at XX. Upon arrival at the scene, personnel from multiple departments passed the buck to each other, telling him that he needed to complete the procedures first and could not go back. After reflecting on the matter, he received a reply from the fishery administration stating that the boat number was unclear, and a fine of 50,000 yuan was required, with five boats confiscated. | Transportation Bureau | Transportation Bureau | Marine Development Bureau |

and the lightweight model were trained using multi-view contrastive learning techniques simultaneously, it can be assumed that the training time is the same for both models.

However, during the testing phase, the lightweight model completes inference using only 6% of the time required by the full model, with a marginal performance decrease of just 4%.

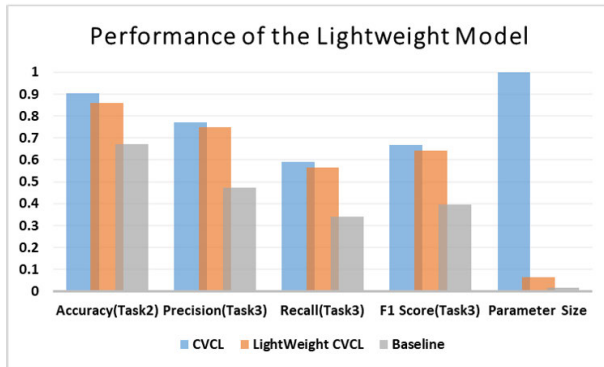


FIGURE 8. Excellent results of our Lightweight Model.

By employing careful optimization strategies, our lightweight model successfully strikes a balance between model size and performance. It offers significant advantages in resource-constrained environments, such as mobile devices or embedded systems. The model not only provides efficient computing and storage requirements but also enables fast deployment and execution, offering a more flexible and feasible solution for real-world application scenarios.

6) ANALYSIS & DISCUSSION

This paper aims to introduce the features and challenges of cross-domain consultation systems and proposes an innovative approach called the Core-View Contrastive Learning (CVCL) network to address these challenges. The core-adaptive augmentation module in the CVCL network bridges the gap between different domains by aligning the semantics of consultations with the target domain, enabling accurate retrieval and integration of cross-domain information.

In addition, we have devised a lightweight model utilizing core-view contrastive learning techniques. Remarkably, despite utilizing only 6% of the parameters of the full model, our lightweight model achieves performance equivalent to 96% of the full model's performance. When compared to the best-performing baseline model, our lightweight model demonstrates a performance improvement of 53%. This underscores the benefits of the lightweight system, which effectively balances smaller parameter sizes with nearly full model performance.

Furthermore, our CVCL network addresses the issues of large-scale models in handling cross-domain problems, such as their excessive size and limited interpretability. While large language models often have a massive number of parameters, our lightweight model, built on core-view contrastive learning, not only approaches the performance of the full model but also offers higher efficiency and interpretability, making it suitable for various real-world applications.

In conclusion, our research presents an innovative CVCL network that demonstrates significant performance

advantages in cross-domain consultation systems. By reducing parameter size, improving efficiency, and enhancing interpretability, our lightweight model provides a flexible and feasible solution for addressing cross-domain problems.

REFERENCES

- [1] Y. Ban, H. Liu, H. Liu, and J. Yuan, "Design and development of an earthquake emergency coordinating consultation system based on GIS," in *Proc. 2nd Int. Conf. Saf. Produce Informatization (IICSPI)*, Nov. 2019, pp. 231–234.
- [2] B. Xu and W. Ren, "Understand online medical consultation based on dialogue state tracking," in *Proc. 4th Int. Conf. Comput. Sci. Artif. Intell.*, 2020, pp. 168–174.
- [3] Y. Sha, T. Feng, X. Xiong, and T. Yang, "Designing online psychological consultation expert system using human-computer interaction," *Mobile Inf. Syst.*, vol. 2021, 2021, Art. no. 6458924.
- [4] W. Chen, Z. Li, H. Fang, Q. Yao, C. Zhong, J. Hao, Q. Zhang, X. Huang, J. Peng, and Z. Wei, "A benchmark for automatic medical consultation system: Frameworks, tasks and datasets," *Bioinformatics*, vol. 39, no. 1, 2023, Art. no. btac817.
- [5] X. Peng, Y. Li, Y. Si, L. Xu, X. Liu, D. Li, and Y. Liu, "A social sensing approach for everyday urban problem-handling with the 12345-complaint hotline data," *Comput., Environ. Urban Syst.*, vol. 94, Jun. 2022, Art. no. 101790.
- [6] N. Zhang, Y.-F. Pu, S.-Q. Yang, J.-L. Zhou, and J.-K. Gao, "An ontological Chinese legal consultation system," *IEEE Access*, vol. 5, pp. 18250–18261, 2017.
- [7] H. Minmin, Y. Lingyun, P. Xue, and Z. Chuan, "Trusted edge and cross-domain privacy enhancement model under multi-blockchain," *Comput. Netw.*, vol. 234, Oct. 2023, Art. no. 109881.
- [8] A. Salehi S., R. Han, C. Rudolph, and M. Grobler, "DACP: Enforcing a dynamic access control policy in cross-domain environments," *Comput. Netw.*, vol. 237, Dec. 2023, Art. no. 110049.
- [9] T. Yu, R. Zhang, M. Yasunaga, Y. Chern Tan, X. Victoria Lin, S. Li, H. Er, I. Li, B. Pang, T. Chen, E. Ji, S. Dixit, D. Proctor, S. Shim, J. Kraft, V. Zhang, C. Xiong, R. Socher, and D. Radev, "SPaC: Cross-domain semantic parsing in context," 2019, *arXiv:1906.02285*.
- [10] F. Yuan, L. Yao, and B. Benatallah, "DARec: Deep domain adaptation for cross-domain recommendation via transferring rating patterns," 2019, *arXiv:1905.10760*.
- [11] U. Ojha, Y. Li, J. Lu, A. A. Efros, Y. Jae Lee, E. Shechtman, and R. Zhang, "Few-shot image generation via cross-domain correspondence," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10743–10752.
- [12] B. P. Mc Cune, R. M. Tong, J. S. Dean, and D. G. Shapiro, "RUBRIC: A system for rule-based information retrieval," *IEEE Trans. Softw. Eng.*, vols. SE-11, no. 9, pp. 939–945, Sep. 1985.
- [13] J. Yin, X. Jiang, Z. Lu, L. Shang, H. Li, and X. Li, "Neural generative question answering," 2015, *arXiv:1512.01337*.
- [14] Y. Kim, M. Kim, A. Cattle, J. Otmakhova, S. Park, and H. Shin, "Applying graph-based keyword extraction to document retrieval," in *Proc. 6th Int. Joint Conf. Natural Lang. Process.*, 2013, pp. 864–868.
- [15] S. E. Robertson and S. Walker, "Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval," in *Proc. 17th Annu. Int. ACM-SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*. Springer, 1994, pp. 232–241.
- [16] C. D. Corley and R. Mihalcea, "Measuring the semantic similarity of texts," in *Proc. ACL Workshop Empirical Modeling Semantic Equivalence Entailment*, 2005, pp. 13–18.
- [17] B. Shan, W. Yin, Y. Sun, H. Tian, H. Wu, and H. Wang, "ERNIE-ViL 2.0: Multi-view contrastive learning for image-text pre-training," 2022, *arXiv:2209.15270*.
- [18] R. Kirov, Y. Zhu, R. R. Salakhutdinov, R. Zemel, R. Urtasun, A. Torralba, and S. Fidler, "Skip-thought vectors," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.
- [19] F. Hill, K. Cho, and A. Korhonen, "Learning distributed representations of sentences from unlabelled data," 2016, *arXiv:1602.03483*.
- [20] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes, "Supervised learning of universal sentence representations from natural language inference data," 2017, *arXiv:1705.02364*.

- [21] D. Cer, Y. Yang, S. Kong, N. Hua, N. Limtiaco, R. St. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Y.-H. Sung, B. Strope, and R. Kurzweil, “Universal sentence encoder,” 2018, *arXiv:1803.11175*.
- [22] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” 2019, *arXiv:1908.10084*.
- [23] Y. Yan, R. Li, S. Wang, F. Zhang, W. Wu, and W. Xu, “ConSERT: A contrastive framework for self-supervised sentence representation transfer,” 2021, *arXiv:2105.11741*.
- [24] J. Giorgi, O. Nitski, B. Wang, and G. Bader, “DeCLUTR: Deep contrastive learning for unsupervised textual representations,” 2020, *arXiv:2006.03659*.
- [25] T. Gao, X. Yao, and D. Chen, “SimCSE: Simple contrastive learning of sentence embeddings,” 2021, *arXiv:2104.08821*.
- [26] Y.-S. Chuang, R. Dangovski, H. Luo, Y. Zhang, S. Chang, M. Soljačić, S.-W. Li, W.-T. Yih, Y. Kim, and J. Glass, “DiffCSE: Difference-based contrastive learning for sentence embeddings,” 2022, *arXiv:2204.10298*.
- [27] Q. Wang, W. Zhang, T. Lei, Y. Cao, D. Peng, and X. Wang, “CLSEP: Contrastive learning of sentence embedding with prompt,” *Knowl.-Based Syst.*, vol. 266, Apr. 2023, Art. no. 110381.
- [28] S. Chopra, R. Hadsell, and Y. LeCun, “Learning a similarity metric discriminatively, with application to face verification,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 539–546.
- [29] A. van den Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” 2018, *arXiv:1807.03748*.
- [30] C. Xu, D. Tao, and C. Xu, “A survey on multi-view learning,” 2013, *arXiv:1304.5634*.
- [31] C. Liu, J. Wen, X. Luo, C. Huang, Z. Wu, and Y. Xu, “DICNet: Deep instance-level contrastive network for double incomplete multi-view multi-label classification,” 2023, *arXiv:2303.08358*.
- [32] Y. Tian, D. Krishnan, and P. Isola, “Contrastive multiview coding,” in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*. Glasgow, U.K.: Springer, 2020, pp. 776–794.
- [33] K. Hassani and A. H. Khasahmadi, “Contrastive multi-view representation learning on graphs,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 4116–4126.
- [34] K. Shah, A. Shah, C. P. Lau, C. M. de Melo, and R. Chellapp, “Multi-view action recognition using contrastive learning,” in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 3381–3391.
- [35] W. Li, Y. Zhang, Y. Sun, W. Wang, M. Li, W. Zhang, and X. Lin, “Approximate nearest neighbor search on high dimensional data—Experiments, analyses, and improvement,” *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1475–1488, Aug. 2020.
- [36] R. Mihalcea and P. Tarau, “Textrank: Bringing order into text,” in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Jul. 2004, pp. 404–411.
- [37] Y. Song, S. Shi, J. Li, and H. Zhang, “Directional skip-gram: Explicitly distinguishing left and right context for word embeddings,” in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, vol. 2, 2018, pp. 175–180.
- [38] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” 2018, *arXiv:1810.04805*.
- [39] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “RoBERTa: A robustly optimized BERT pretraining approach,” 2019, *arXiv:1907.11692*.
- [40] M. Li, T. Shi, C. Ziems, M.-Y. Kan, N. F. Chen, Z. Liu, and D. Yang, “CoAnnotating: Uncertainty-guided work allocation between human and large language models for data annotation,” 2023, *arXiv:2310.15638*.
- [41] F. Rahutomo, T. Kitasuka, and M. Aritsugi, “Semantic cosine similarity,” in *Proc. 7th Int. Student Conf. Adv. Sci. Technol. (ICAST)*, 2012, vol. 4, no. 1, p. 1.
- [42] D. Cer, M. Diab, E. Agirre, I. Lopez-Gazpio, and L. Specia, “SemEval-2017 task 1: Semantic textual similarity—multilingual and cross-lingual focused evaluation,” 2017, *arXiv:1708.00055*.
- [43] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
- [44] K. Zhou, B. Zhang, W. Xin Zhao, and J.-R. Wen, “Debiased contrastive learning of unsupervised sentence representations,” 2022, *arXiv:2205.00656*.
- [45] X. Wu, C. Gao, Y. Su, J. Han, Z. Wang, and S. Hu, “Smoothed contrastive learning for unsupervised sentence embedding,” 2021, *arXiv:2109.04321*.
- [46] X. Wu, C. Gao, Z. Lin, J. Han, Z. Wang, and S. Hu, “InfoCSE: Information-aggregated contrastive learning of sentence embeddings,” 2022, *arXiv:2210.06432*.
- [47] S. Chen, J. Zhou, Y. Sun, and L. He, “An information minimization based contrastive learning model for unsupervised sentence embeddings learning,” 2022, *arXiv:2209.10951*.
- [48] Y. Zhang, R. Zhang, S. Mensah, X. Liu, and Y. Mao, “Unsupervised sentence representation via contrastive learning with mixing negatives,” in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 10, pp. 11730–11738.



JIABIN ZHENG is currently an Artificial Intelligence Algorithm Engineer with the School of Computer Science, Peking University. His research interests include information retrieval, text matching, and dialog systems.



FANGYI XU received the B.A. degree from Peking University, the M.F.A. degree from the University of Southern California, and the Ph.D. degree from Peking University. She was a Post-doctoral Researcher with Tsinghua University. She is currently an Associate Professor with the Communication University of China. Her research interest includes communication.



WEI CHEN received the Ph.D. degree from Peking University. She is currently an Associate Researcher with the School of Computer Science, Peking University. Her research interests include unstructured data management, cloud computing, integration of unstructured data extraction, and data mining.



ZIHAO FANG is currently pursuing the bachelor's degree with the School of Science, The Hong Kong University of Science and Technology.



JIAHUI YAO is currently an Engineer with the Institute of Social Science Survey, Peking University. Her research interests include big data and survey data and computational social science.