Article

Not peer-reviewed version

Knowledge Graph-Based Question Answering via Enhanced Contrastive Learning

Madison Smith , Woods Ali , Tyler Johnson *

Posted Date: 30 August 2024

doi: 10.20944/preprints202408.2225.v1

Keywords: natural language processing; question answering; knowledge graphs; contrastive learning

Article

Knowledge Graph-Based Question Answering via Enhanced Contrastive Learning

Madison Smith, Woods Ali and Tyler Johnson *

University of Charleston
* Correspondence: t.johnson@gmail.com

Abstract: This study innovates upon the field of dialog-driven query resolution (DDQR) utilizing knowledge graphs (KGs). Traditional methods in DDQR predominantly depend on fully supervised signals that presuppose the existence of perfect logical forms for queries. This reliance on gold logical forms for answer extraction is impractical in diverse real-world applications. When such forms are absent, contemporary methods, which lean on weak supervisory signals or employ heuristic and reinforcement strategies, recast DDQR as a knowledge graph path optimization problem. Despite the non-availability of gold logical forms, the rich conversational context provided by comprehensive dialog histories and domain-specific knowledge can be leveraged to optimize path selection in KGs effectively. We introduce an advanced method, termed CONVEX (CONVersational EXploration), which utilizes contrastive learning for path ranking. CONVEX addresses critical challenges by enabling learning under weak supervision and integrating the conversational context to enhance the representation quality for more effective path discrimination. Extensive evaluations of CONVEX on established benchmarks demonstrate its superiority across various metrics over current methods, notably improving Mean Reciprocal Rank (MRR) and Hit@5 by up to 20 and 36 percentage points respectively.

Keywords: natural language processing, question answering, knowledge graphs, contrastive learning

1. Introduction

Dialog-driven query resolution (DDQR) using knowledge graphs (KGs) is a crucial, evolving field, increasingly relevant due to the rise of sophisticated personal assistant technologies like Siri and Google Assistant. Current literature in this domain generally diverges into two methodologies [12,43,45]: semantic parsing, which transforms user queries into executable logical forms on KGs, and information retrieval, which focuses on optimizing graph path selection based on query specifics [23,48].

Question Answering (QA) [57] systems are designed to retrieve information or provide answers to questions posed in natural language. The development of QA systems dates back to the early days of artificial intelligence, with initial systems focusing on structured data retrieval from curated databases. Over the years, the evolution of QA has been influenced by advancements in computational linguistics and machine learning, enabling these systems to parse complex queries and fetch relevant answers from large, unstructured datasets. Modern QA systems, such as those integrated into search engines and virtual assistants, employ sophisticated natural language processing techniques to understand the intent and context of a question, making them integral tools in information retrieval and decision support [60].

In the contemporary landscape, QA systems have diversified into various sub-fields, including factoid QA, list QA, definition QA, and conversational QA [57,58]. Each type caters to different user needs; for example, factoid QA focuses on providing concise, factual answers to specific questions, while conversational QA aims to engage the user in a dialogue, offering responses that require contextual understanding over multiple turns. This segmentation has prompted significant research in domain-specific QA systems, where knowledge bases are tailored to specific fields such as medicine, finance, or law. The integration of machine learning models, especially deep learning, has further propelled the capabilities of QA systems, enabling them to learn from vast amounts of data and

 doi:10.20944/preprints202408.2225.v1

improve their accuracy over time through techniques such as reinforcement learning and transfer learning.

The challenge with semantic parsing is its intensive demand for annotated training data [23], each query needing a precise logical counterpart. Conversely, retrieval-focused approaches merely require correct end entities or paths per query, simplifying data preparation but complicating the retrieval process.

Existing methods in DDQR often categorize the task as path optimization. Solutions range from heuristic-based models [10,67] to sophisticated reinforcement learning schemes [21]. However, these models either demand extensive manual rule crafting or heavily rely on user engagement for query refinement, which can burden users and lead to dissatisfaction if frequent adjustments are needed within short intervals [14].

This work introduces a novel framework, CONVEX (CONVersational EXploration), which leverages contrastive learning to refine KG path ranking by effectively incorporating dialog histories and domain insights into the learning process. The choice for contrastive learning stems from its capacity to differentiate between relevant and irrelevant paths by contrasting them against each other in a shared embedding space [34].

By integrating entire dialog histories and domain information, CONVEX aims to enhance the representational depth, improving the system's ability to discern and prioritize the most relevant KG paths. This approach addresses significant limitations in existing models by reducing dependency on manual rule creation and mitigating the cognitive load on users [69,70].

Key contributions of this paper are: 1) Introduction of CONVEX, a pioneering contrastive learning model for DDQR, designed to cohesively blend conversational context with KG path learning to elevate path ranking efficacy. 2) An extensive analysis of how additional contextual layers, such as dialog domain and response fluency, impact the performance of DDQR systems. Our findings indicate substantial improvements over traditional methods on several benchmark datasets. 3) Detailed experimental insights that validate the effectiveness of CONVEX, showcasing significant enhancements in ranking metrics across multiple scenarios.

The remainder of the paper is organized as follows: Section 2 reviews related work. Section 3 outlines the task definitions and notations. Section 4 describes the CONVEX framework. Section 5 discusses experimental setups, results, and both ablation studies and error analyses. We conclude our findings in Section 6.

2. Related Work

Question Answering over Knowledge Graphs (KGQA) represents a robust area of inquiry within the broader field of artificial intelligence. Comprehensive surveys on this topic have categorized various approaches, with specific details found in foundational works by Zaib et al. [43] and Diefenbach et al. [12], Fei et al. [74]. Our research intersects with these established methodologies but introduces novel elements that are particularly relevant to conversational contexts.

Historically, KGQA efforts have emphasized the development of semantic graph generation and subsequent re-ranking processes. Initial strategies involved comparing user queries against predefined query templates, as demonstrated by Bast and Haussmann [1], Fei et al. [76], who enriched these templates with potential relations to create a variety of query graph candidates. Further refinement was introduced by Yih et al. [41], who implemented a heuristic search algorithm staged to construct grounded query graphs, which were then evaluated using a neural ranking model to select the most accurate semantic graph. This paradigm was extended by Yu et al. [42] through the integration of a hierarchical representation of KG relations into a neural query graph ranking model, which was compared against a local sub-sequence alignment model employing cross-attention mechanisms [32]. Maheshwari et al. [28], Fei et al. [82] further explored this area by conducting a comprehensive empirical investigation of various neural query graph ranking models, culminating in the development of a self-attention-based slot matching model that leverages the intrinsic structure of query graphs for enhanced performance.

In the domain of Conversational Question Answering (ConvQA), seminal works have largely adopted semantic parsing techniques to tackle multi-turn dialogues [17,33,38,85]. The pioneering model by Saha et al. [36] combined elements of the Hierarchical Recurrent Encoder-Decoder (HRED) model [37] and the key-value memory network model [30], integrating these components to effectively compute contextual representations and facilitate dynamic answer generation. This approach was pivotal in establishing a framework that could accommodate the complexities inherent in conversational interactions, setting the stage for subsequent advancements. Follow-up studies have built on this foundation, exploring multi-task learning frameworks that further refine the handling of conversational nuances and query adjustments [17,33].

Christmann et al. [10] and Kaiser et al. [21] introduced innovative methods for maintaining conversational context and adapting to user interactions within a KG. These methodologies, while groundbreaking, exhibited limitations related to their reliance on rule-based systems or user-driven query reformulation, which can impose significant cognitive demands on users and are susceptible to biases and errors. Our work, named CONVEX, seeks to transcend these limitations by adopting a weak supervision approach that harnesses the full spectrum of conversational context and available KG paths to retrieve answers, eschewing the need for direct user input in the reformulation process.

Contrastive learning, initially introduced for facial recognition [9], has seen widespread application across various domains, including image caption identification [35], augmented image code computation [5], feature clustering [4], and dense information retrieval [15]. CONVEX adapts this framework to the domain of KGQA by developing a contrastive learning model that minimizes the embedding distance between similar items and maximizes it for dissimilar ones within the context of conversational interactions, fluent responses, and domain-specific information [93]. The subsequent sections will delve into the specifics of CONVEX's methodology, illustrating how it addresses the unique challenges of conversational question answering over knowledge graphs.

3. Preliminary

In the realm of Knowledge Graphs (KGs), a KG can be described as a triplet $\mathcal{K} = (\mathcal{E}, \mathcal{R}, \mathcal{T}^+)$, where $\mathcal{E}$ represents the set of entities, $\mathcal{R}$ denotes the relations, and $\mathcal{T}^+$ consists of triples indicating relational connections between entities. Specifically, a triple $\tau = (e_h, r_{h,t}, e_t) \in \mathcal{T}^+$ asserts a relationship $r_{h,t}$ linking a head entity e_h to a tail entity e_t. In the context of dialog-driven query answering, a conversation $\mathcal{C}$ spanning T turns is framed by sequences of questions $\mathcal{Q} = \{q^t\}$ and corresponding answers $\mathcal{A} = \{a^t\}$, structured as $\mathcal{C} = \langle (q^0, a^0), (q^1, a^1), ..., (q^{\mathrm{T}}, a^{\mathrm{T}}) \rangle$. Each question q^t and its fluent response v^t are token sequences, enhancing the conversational context and enabling richer interaction dynamics within the KG framework.

The conversation context is enriched by defining:

- **Context Entities** $\mathcal{E}_c$: A subset of $\mathcal{E}$, including entities mentioned throughout $\mathcal{C}$.
- **Context Paths** $\mathcal{P}_c$: Paths derived from $\mathcal{E}_c$, spanning up to three hops, thus encapsulating the breadth of relational connectivity relevant to the conversational queries.

Problem Formulation: The ConvQA task under our newly proposed CONVEX model aims to utilize the KG $\mathcal{K}$, integrating the conversational history $\mathcal{C}^t$ and context entities $\mathcal{E}_c^t$ to identify viable KG paths $\mathcal{P}_c^t$. This process is approached as an information retrieval (IR) challenge, scoring and ranking potential paths to ascertain those most likely to yield correct responses as specified by gold standard answers a^t. The CONVEX model employs contrastive representation learning to optimize this selection, ensuring that similar contexts yield proximal embeddings, while distinct scenarios are clearly separated in the embedding space.

Table 1. Notational Glossary for the CONVEX Framework.

Notation	Definition
$\mathcal{K}, \mathcal{E}, \mathcal{R}, \mathcal{T}^+$	Knowledge Graph, Entities, Relations, Triples
$\mathcal{C}, t$	Conversation, Turn Index
q^t, a^t	Query and its Answer at turn t
v^t	Verbatim response at turn t
τ^t	Tail entity at turn t
$\mathcal{C}^t$	Historical Context up to turn t
$\mathcal{E}_c, \mathcal{P}_c$	Contextual Entities, Derivable Paths
$\mathcal{D}^{t+}, \mathcal{D}^{t-}$	Positive/Negative Path Sets for q^t
s^t	Input sequence incorporating $\mathcal{C}^t$ and q^t
d	Dimensionality of the embedding space
$h^{(\cdot)}$	Contextual embeddings function
$\theta^{(\cdot)}$	Parameters subject to optimization
$\boldsymbol{W}^{(\cdot)}$	Weight matrices in neural layers
$\omega^{(\cdot)}$	Probability distributions over output space
ϕ^c, ϕ^p	Joint embeddings for conversation and path contexts

4. Methodology

Our methodology, designated as CONVEX, employs a contrastive learning framework to effectively rank knowledge graph (KG) paths based on the dialogue context of each conversational turn. The process, depicted is structured into three primary phases: 1) preprocessing to identify potential KG paths, 2) encoding the conversational context, and 3) embedding the dialogue and KG paths jointly for contrastive path ranking.

Preprocessing & Path Extraction: Initially, we process the input comprising questions q^t and answers a^t to identify relevant context entities and potential KG paths $\mathcal{P}_c^t$, similar to the approach used in [21]. These paths are preliminarily scored based on their alignment with the domain-specific characteristics of the query. The embeddings for these paths are prepared using sentence transformers that leverage BERT's [11] deep contextual representations.

Contextual Encoding: We further enrich the model by encoding the conversational history and integrating domain knowledge. A BART-based bidirectional encoder [25] captures the nuances of both the historical and current conversational context, generating embeddings that reflect the sequential interplay of dialogue exchanges. This step is crucial for aligning the conversational flow with the structured knowledge embedded in KG paths.

Joint Embedding and Ranking: The core of CONVEX lies in its ability to jointly embed the dialogue context and the KG paths within a shared embedding space. This shared space facilitates a contrastive ranking mechanism, where the goal is to minimize the distance between correct path-context pairs and maximize the distance for incorrect pairs. This mechanism is critical for enhancing the accuracy of the model in real-time conversational settings.

In more detail, the preprocessing phase involves identifying and encoding potential KG paths using domain-specific transformers that prepare the path representations for subsequent ranking tasks. These representations are tuned to capture the essential features of the paths that are most likely to answer the given query in a conversation.

The contextual encoding phase leverages a sophisticated encoder to integrate the conversation history, which includes both the questions and the corresponding fluent responses. This integrated approach helps in understanding the progression of the conversation, thus providing a richer context for the subsequent ranking of KG paths.

The joint embedding and ranking phase is designed to operate in an end-to-end fashion post-preprocessing. It intricately combines the conversation's encoded context with the candidate KG paths, employing a contrastive learning strategy to distinguish between the most and least relevant paths effectively. This phase is pivotal in ensuring that the ranked paths are not only contextually relevant but also dynamically aligned with the ongoing conversation.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 August 2024 doi:10.20944/preprints202408.2225.v1

As illustrated in Table 2, each step of the CONVEX methodology utilizes cutting-edge AI techniques to enhance the interaction between conversational context and KG path ranking, ensuring a robust and adaptive QA system. The implications of such an approach are profound, allowing for more nuanced and accurate responses within conversational AI applications.

Table 2. Summary of the CONVEX Methodology Steps

Step	Process	Technique Used
1	Path Extraction	Domain-Specific Transformers
2	Contextual Encoding	BART-based Bidirectional Encoder
3	Joint Embedding	Contrastive Ranking Mechanism

4.1. Joint Contrastive Learning

CONVEX integrates three core trainable modules, each associated with a specific loss function. These modules include the encoder, domain identification pointer, and contrastive ranking module, each critical for the joint learning process. The losses from these modules are combined in a weighted average to optimize the model holistically:

$$L = \lambda_1 L^{dm} + \lambda_2 L^{rk} + \lambda_3 L^{dec}, \tag{1}$$

where $\lambda_1, \lambda_2, \lambda_3$ are the weighting factors for each loss component. The loss components are defined as:

$$\begin{aligned}
L^{dm} &= -\sum_{j=1}^{m} \log p(y_j^{(dm)}|x), \\
L^{rk} &= \begin{cases} 1 - \cos(\phi^c, \phi^p), & \text{if } y^{(rk)} = 1 \\ max(0, \cos(\phi^c, \phi^p) - \alpha), & \text{if } y^{(rk)} = -1 \end{cases}, \\
L^{dec} &= -\sum_{l=1}^{n} \log p(y_l^{(dec)}|x),
\end{aligned} \tag{2}$$

These definitions reflect the integration of various learning signals to improve the model's effectiveness across different tasks. The complete learning algorithm for CONVEX is outlined in Algorithm 1.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 August 2024 doi:10.20944/preprints202408.2225.v1

Algorithm 1: CONVEX Algorithmic Implementation

Input: Training set $S_{train} = \{(q^t, C^t, v^t, \tau^t, \mathcal{D}_c^{t+}, \mathcal{D}_c^{t-})\}$

for $S_{batch} \in S_{train}$ **do**
- $q_b \leftarrow extractQuestions(S_{batch})$
- $C_b \leftarrow extractConvHistory(S_{batch})$
- $v_b \leftarrow extractFluentResponses(S_{batch})$
- $\tau_b \leftarrow identifyDomains(S_{batch})$
- $y_b^{(rk)} : y_b^{(rk)} \in \mathbb{R}^{\{1,-1\} \times b}$
- **for** $y_i^{(rk)} \in y_b^{(rk)}$ **do**
  - **if** $y_i^{(rk)} = 1$ **then**
    - $\mathcal{D}^+ \leftarrow fetchPositivePaths(S_{batch})$
    - $p_i \sim uniformlySample(\mathcal{D}^+)$
  - **else**
    - $\mathcal{D}^- \leftarrow fetchNegativePaths(S_{batch})$
    - $p_i \sim uniformlySample(\mathcal{D}^-)$
  - **end**
- **end**
- $h_b^{(p)} \leftarrow encodePaths(p_b)$
- $h_b^{(dm)} \leftarrow encodeDomains(\tau_b)$
- **begin** CONVEX forward
  - $h_b^{(enc)} \leftarrow CONVEX.encoder(q_b, C_b)$
  - $\omega_b^{(dec)} \leftarrow CONVEX.decoder(h_b^{(enc)}, v_b)$
  - $\omega_b^{(dm)} \leftarrow CONVEX.domainPointer(h_b^{(enc)})$
  - $\phi_b^c, \phi_b^p \leftarrow CONVEX.ranking(h_b^{(enc)}, h_b^{(dm)}, h_b^{(p)})$
- **end**
- $L_b^{dm} = \frac{1}{b}\sum_{i=1}^{b} - \sum_{j=1}^{m} \log p(y_j^{(dm)} | \omega_i^{(dm)})$
- $L_b^{dec} = \frac{1}{b}\sum_{i=1}^{b} - \sum_{k=1}^{n} \log p(y_l^{(dec)} | \omega_i^{(dec)})$
- $L_b^{rk} = \frac{1}{b}\sum_{i=1}^{b} \begin{cases} 1 - cos(\phi_i^c, \phi_i^p), & \text{if } y_i^{(rk)} = 1 \\ max(0, cos(\phi_i^c, \phi_i^p) - \alpha), & \text{if } y_i^{(rk)} = -1 \end{cases}$
- Update CONVEX weights w.r.t. $\lambda_1 L_b^{dm} + \lambda_2 L_b^{rk} + \lambda_3 L_b^{dec}$

end

4.2. Inference Process

Post-training, CONVEX's inference involves several steps to utilize the learned models effectively. Initially, the context entities $\mathcal{E}^c$ are identified within the input question and conversational history. Subsequently, relevant context paths $\mathcal{P}^c$ are extracted, and the conversation is encoded to identify the domain and score the candidate paths. The highest-ranked path determines the response, which is then used to generate a fluent answer via the decoder module.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 August 2024 doi:10.20944/preprints202408.2225.v1

5. Experiments

5.1. Settings

Model Implementations. Table 3 summarizes the hyperparameters used in the experiments. The CONVEX framework is implemented with a uniform space dimension of $d = 768$ across all components. The encoder and decoder are built upon the BART (base) architecture [25]. Training parameters include a batch size of 32, a learning rate of $1e-4$, over 120 epochs, with model checkpoints saved periodically. Optimization is conducted using the AdamW algorithm, with a specific adjustment for weight decay [27]. We apply a residual dropout of 0.1 across various modules, including the domain pointer and the ranking module, to mitigate overfitting. Initial embeddings for domains and KG paths are generated using a pre-trained BERT model. The maximum token limit for the input sequence combining conversational history and the current question ($C^t + q^t$) is set to 150. We apply weighting factors $\lambda_1 = 0.25$ and $\lambda_2 = 0.25$ for the domain pointer and decoder losses, respectively, and $\lambda_3 = 1.0$ with a margin $\alpha = 0.1$ for the path ranking cosine embedding loss.

Table 3. Hyperparameters for CONVEX.

Hyperparameters	Value
epochs	120
batch size	32
learning rate	$1e-4$
dropout ratio	0.1
optimizer	AdamW
model dimension	768
v^t max length	50
$C^t + q^t$ max length	150
domain pre-trained embeddings	BERT
KG paths pre-trained embeddings	BERT
$\lambda_1, \lambda_2, \lambda_3$	$0.25, 1.0, 0.25$
margin α	0.1

Datasets and Models for Comparison. For evaluating the performance of CONVEX, we utilize the ConvQuestions [10] and ConvRef [21] datasets, incorporating fluent responses [19]. Baselines include the original CONVEX model [10] and CONQUER [21], an RL-based conversational QA method. We also compare against recent semantic parsing-based model OAT [29] and Focal Entity [24].
Evaluation Metrics. The effectiveness of CONVEX is assessed using Precision at the top rank (P@1), Mean Reciprocal Rank (MRR), and Hit at 5 (H@5), along with precision, recall, and F1-score for domain identification. BLEU-4 and METEOR scores are utilized for evaluating fluent response generation.

5.2. Results

Research Questions: The primary research question, **RQ**, asks: What is the efficacy of the contrastive learning approach in CONVEX for ranking KG paths? This question is further subdivided into: **RQ1.1:** What impact does conversational context have on CONVEX's efficiency? **RQ1.2:** How do individual tasks within CONVEX perform, such as domain identification and fluent response generation?
Overall Performance on ConvQA datasets. Table 4 shows that CONVEX surpasses all previous baselines on the ConvQuestions dataset in metrics such as P@1, H@5, and MRR. Similarly, for ConvRef, CONVEX outperforms existing models, particularly in metrics that evaluate deeper conversational understanding and context utilization.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 August 2024 doi:10.20944/preprints202408.2225.v1

Table 4. Comparative performance of CONVEX on employed datasets, showing improvements across all evaluated metrics.

Dataset	ConvQuestions			ConvRef		
Model	P@1	H@5	MRR	P@1	H@5	MRR
Previous CONVEX	0.184	0.219	0.200	0.225	0.257	0.241
CONQUER	0.240	0.343	0.279	0.353	0.429	0.387
OAT	0.250	-	0.260	-	-	-
Focal Entity	0.248	-	0.248	-	-	-
CONVEX	**0.292**	**0.529**	**0.398**	**0.335**	**0.599**	**0.441**

5.3. Ablation Study

We perform an ablation study to delineate the contributions of different components in CONVEX. Removing the full conversational context, domain information, or fluent responses each detrimentally impacts the performance, underscoring their collective importance in our model's architecture.

Table 5. Ablation study results showing the impact of different components and training strategies on the performance of CONVEX.

Configuration	ConvQuestions			ConvRef		
Full Model	0.292	0.529	0.398	0.335	0.599	0.441
w/o Full Conv. History	0.214	0.375	0.299	0.247	0.449	0.324
w/o Domain Info	0.247	0.436	0.296	0.266	0.472	0.356
w/o Fluent Responses	0.265	0.441	0.324	0.279	0.503	0.397
Train Separately	0.255	0.413	0.328	0.304	0.529	0.408

5.4. Error Analysis

Our detailed error analysis, focusing on 250 randomly sampled incorrect predictions, highlights two main types of errors: Incorrect ranking of paths with semantically similar KG relations and the absence of complete gold KG paths in the training data, underscoring areas for potential improvement in future iterations of CONVEX.

Table 6. Summary of performance improvements achieved by CONVEX over the baseline.

Metric	Baseline	CONVEX	Improvement
Hit@5	0.343	0.529	+54.2%
Precision@1	0.240	0.292	+21.7%
MRR	0.279	0.398	+42.7%

5.4.1. Incorrect Ranking of Semantically Similar KG Relations

CONVEX occasionally misranks paths when they contain KG relations that are semantically similar but contextually distinct. Consider the query, "*What kind of book is it?*" set within a rich conversational history:

- q^1: "*Who wrote The Secret Garden?*"
- v^1: "*Frances Eliza Hodgson Burnett.*"
- q^2: "*Where is the story set?*"
- v^2: "*In Yorkshire.*"
- q^3: "*When was it published?*"
- v^3: "*In 1910.*"

The system needs to identify the KG path containing the relation "main subject (P921)," pointing correctly to "*adventure (Q1436734)*." However, CONVEX often prioritizes paths featuring the "genre

 doi:10.20944/preprints202408.2225.v1

(P136)" relation due to its prevalent usage across training instances. This example underscores a key challenge: distinguishing between closely related semantic relations like "main subject" and "genre," which often leads to suboptimal ranking outcomes.

5.4.2. Absence of Gold KG Paths

A significant challenge in our dataset is the absence of gold-standard KG paths, affecting over 25% of the training examples and 19% of the test cases. This absence has a direct impact on CONVEX's learning efficacy, as instances lacking a gold path are invariably treated as errors, thus degrading overall performance metrics. Enhanced annotation efforts to include comprehensive gold KG paths could potentially ameliorate this issue, leading to more accurate learning and prediction.

Table 7. Summary of major error types identified in CONVEX's performance, highlighting the proportion of instances affected and the impact level on the results.

Error Type	Instances	Impact on Results
Semantically Similar Relations	35%	High
Absence of Gold Paths	25%	Critical

5.4.3. Improvement

Based on these findings, we propose several avenues for future research:

1. **Enhanced Relation Disambiguation:** Developing more sophisticated techniques for relation extraction and disambiguation could help in more accurately distinguishing between similar relations. Incorporating additional contextual cues from the KG, such as entity types or relation aliases, may provide further discriminative power.
2. **Richer Training Data:** Expanding the dataset to include more comprehensive gold KG paths could reduce the incidence of training and testing without adequate ground truth, thus improving the reliability of the learning process.
3. **Advanced Contextual Modeling:** Leveraging newer models that better capture the nuances of conversational context may also help in more effectively discerning the relevance of different KG paths based on the dialogue history.

These strategies represent promising research directions that could substantially refine the efficacy of KG path ranking methodologies in conversational question-answering systems.

6. Conclusions and Future Work

Our research has critically addressed the utility of contrastive representation learning within the domain of conversational question answering over knowledge graphs. We posited the challenge as a KG path ranking problem and enriched the model's context handling by incorporating comprehensive dialogue histories, fluent responses, and domain-specific information. The implemented model, named CONVEX, substantiates the premise that contrastive learning can markedly enhance performance in this intricate task.

Conclusions: The empirical results obtained from the CONVEX framework have demonstrably surpassed those of traditional baseline approaches, establishing a robust case for the efficacy of our method. The joint embedding of conversational context and KG paths within a unified representation space has been shown to significantly influence the outcome of ranking metrics positively. Moreover, detailed ablation studies have shed light on the individual contributions of various conversational elements, affirming that each component—complete dialog history, fluent responses, and domain information—plays a vital role in the model's success.

Future Directions: Despite the advancements presented, our work unveils several avenues for future exploration and improvement: 1) *Relation Extraction Enhancement:* The error analysis highlighted challenges in accurately identifying and linking the correct KG relations. Future work could enhance relation extraction techniques by integrating richer KG contexts, akin to approaches seen in RECON [2]. Improving this aspect could refine the accuracy of the path ranking process. 2) *Precision Enhancement:* The disparity between the high Hit@5 scores and lower Precision@1 indicates potential improvements in the model's precision. Innovating new methodologies or refining existing algorithms to better discriminate between closely ranked KG paths could help elevate the Precision@1 scores. 3) *Exploring Additional Contextual Data:* Integrating external knowledge sources or more expansive KG embeddings could provide deeper insights and more accurate context capturing, potentially leading to better model performance. 4) *Multi-turn Dialogue Handling:* Further research into handling multi-turn dialogues more effectively, perhaps by leveraging recent advancements in transformer models or recurrent neural strategies, could improve how context is maintained and utilized across longer conversations.

We believe that the insights garnered from the CONVEX framework will catalyze further research into this relatively underexplored area of IR within the ConvQA domain. By highlighting both the strengths and limitations of our approach, we hope to inspire future studies to build on our foundational work, pushing the boundaries of what is possible in conversational AI.

References

1. Hannah Bast and Elmar Haussmann. 2015. More Accurate Question Answering on Freebase. In *CIKM*. Association for Computing Machinery, 1431–1440.
2. Anson Bastos, Abhishek Nadgeri, Kuldeep Singh, Isaiah Onando Mulang, Saeedeh Shekarpour, Johannes Hoffart, and Manohar Kaul. 2021. RECON: Relation Extraction using Knowledge Graph Context in a Graph Neural Network. In *Proceedings of the Web Conference 2021*. 1673–1685.
3. Debanjali Biswas, Mohnish Dubey, Md Rashad Al Hasan Rony, and Jens Lehmann. 2020. VANiLLa: Verbalised ANswers in Natural Language at Large scale. (2020).
4. Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. 2018. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European conference on computer vision (ECCV)*. 132–149.
5. Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. 2020. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in Neural Information Processing Systems* 33 (2020), 9912–9924.
6. Jia Chen, Yiqun Liu, Jiaxin Mao, Fan Zhang, Tetsuya Sakai, Weizhi Ma, Min Zhang, and Shaoping Ma. 2021. Incorporating Query Reformulating Behavior into Web Search Evaluation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 171–180.
7. Jia Chen, Jiaxin Mao, Yiqun Liu, Fan Zhang, Min Zhang, and Shaoping Ma. 2021. Towards a better understanding of query reformulation behavior in web search. In *Proceedings of the Web Conference 2021*. 743–755.
8. Laura Chiticariu, Yunyao Li, and Frederick Reiss. 2013. Rule-based information extraction is dead! long live rule-based information extraction systems!. In *Proceedings of the 2013 conference on empirical methods in natural language processing*. 827–832.
9. Sumit Chopra, Raia Hadsell, and Yann LeCun. 2005. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1. IEEE, 539–546.
10. Philipp Christmann, Rishiraj Saha Roy, Abdalghani Abujabal, Jyotsna Singh, and Gerhard Weikum. 2019. Look before You Hop: Conversational Question Answering over Knowledge Graphs Using Judicious Context Expansion. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management CIKM*. 729–738.
11. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, 4171–4186.
12. Dennis Diefenbach, Vanessa Lopez, Kamal Singh, and Pierre Maret. 2018. Core techniques of question answering systems over knowledge bases: a survey. *Knowledge and Information systems* 55, 3 (2018), 529–569.

 doi:10.20944/preprints202408.2225.v1

13. Timothy F Doyle. 2007. *The role of context in meaning and understanding*. Ph. D. Dissertation. Universitaet Potsdam.
14. Ahmed Hassan, Xiaolin Shi, Nick Craswell, and Bill Ramsey. 2013. Beyond clicks: query reformulation as a predictor of search satisfaction. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 2019–2028.
15. Gautier Izacard, Mathilde Caron, Lucas Hosseini, Sebastian Riedel, Piotr Bojanowski, Armand Joulin, and Edouard Grave. 2021. Towards Unsupervised Dense Information Retrieval with Contrastive Learning. *arXiv preprint arXiv:2112.09118* (2021).
16. Endri Kacupaj, Barshana Banerjee, Kuldeep Singh, and Jens Lehmann. 2021. ParaQA: A Question Answering Dataset with Paraphrase Responses for Single-Turn Conversation. In *The Semantic Web*. Springer International Publishing, Cham, 598–613.
17. Endri Kacupaj, Joan Plepi, Kuldeep Singh, Harsh Thakkar, Jens Lehmann, and Maria Maleshkova. 2021. Conversational Question Answering over Knowledge Graphs with Transformer and Graph Attention Networks. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 850–862.
18. Endri Kacupaj, Shyamnath Premnadh, Kuldeep Singh, Jens Lehmann, and Maria Maleshkova. 2021. VOGUE: Answer Verbalization Through Multi-Task Learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 563–579.
19. Endri Kacupaj, Kuldeep Singh, Maria Maleshkova, and Jens Lehmann. 2022. An Answer Verbalization Dataset for Conversational Question Answerings over Knowledge Graphs. *arXiv preprint arXiv:2208.06734* (2022).
20. Endri Kacupaj, Hamid Zafar, Jens Lehmann, and Maria Maleshkova. 2020. Vquanda: Verbalization question answering dataset. In *European Semantic Web Conference*. Springer, 531–547.
21. Magdalena Kaiser, Rishiraj Saha Roy, and Gerhard Weikum. 2021. Reinforcement Learning from Reformulations In Conversational Question Answering over Knowledge Graphs. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 459–469.
22. Johannes Kiesel, Xiaoni Cai, Roxanne El Baff, Benno Stein, and Matthias Hagen. 2021. Toward Conversational Query Reformulation. In *DESIRES*.
23. Yunshi Lan, Gaole He, Jinhao Jiang, Jing Jiang, Wayne Xin Zhao, and Ji-Rong Wen. 2021. A Survey on Complex Knowledge Base Question Answering: Methods, Challenges and Solutions. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. International Joint Conferences on Artificial Intelligence Organization, 4483–4491. Survey Track.
24. Yunshi Lan and Jing Jiang. 2021. Modeling transitions of focal entities for conversational knowledge base question answering. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*.
25. Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 7871–7880.
26. Jimmy J Lin, Dennis Quan, Vineet Sinha, Karun Bakshi, David Huynh, Boris Katz, and David R Karger. 2003. What Makes a Good Answer? The Role of Context in Question Answering.. In *INTERACT*, Vol. 3. Citeseer, 25–32.
27. Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*.
28. Gaurav Maheshwari, Priyansh Trivedi, Denis Lukovnikov, Nilesh Chakraborty, Asja Fischer, and Jens Lehmann. 2019. Learning to Rank Query Graphs for Complex Question Answering over Knowledge Graphs. In *The Semantic Web – ISWC 2019*. Springer International Publishing, Cham, 487–504.
29. Pierre Marion, Paweł Krzysztof Nowak, and Francesco Piccinno. 2021. Structured Context and High-Coverage Grammar for Conversational Question Answering over Knowledge Graphs. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2021).
30. Alexander Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. 2016. Key-Value Memory Networks for Directly Reading Documents. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 1400–1409.

31. Kyoungsik Na. 2021. The effects of cognitive load on query reformulation: mental demand, temporal demand and frustration. *Aslib Journal of Information Management* (2021).
32. Ankur Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit. 2016. A Decomposable Attention Model for Natural Language Inference. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2249–2255.
33. Joan Plepi, Endri Kacupaj, Kuldeep Singh, Harsh Thakkar, and Jens Lehmann. 2021. Context Transformer with Stacked Pointer Networks for Conversational Question Answering over Knowledge Graphs. In *The Semantic Web*. Springer International Publishing, 356–371.
34. Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*. PMLR, 8748–8763.
35. Alec Radford, Jeffrey Wu, Dario Amodei, Daniela Amodei, Jack Clark, Miles Brundage, and Ilya Sutskever. 2019. Better language models and their implications. *OpenAI Blog https://openai. com/blog/better-language-models* 1 (2019), 2.
36. Amrita Saha, Vardaan Pahuja, Mitesh M Khapra, Karthik Sankaranarayanan, and Sarath Chandar. 2018. Complex sequential question answering: Towards learning to converse over linked question answer pairs with a knowledge graph. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
37. Iulian V. Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. 2016. Building End-to-End Dialogue Systems Using Generative Hierarchical Neural Network Models. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16)*. AAAI Press, 3776–3783.
38. Tao Shen, Xiubo Geng, Tao Qin, Daya Guo, Duyu Tang, Nan Duan, Guodong Long, and Daxin Jiang. 2019. Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base. In *EMNLP)*. 2442–2451.
39. Kuldeep Singh, Arun Sethupat Radhakrishna, Andreas Both, Saeedeh Shekarpour, Ioanna Lytra, Ricardo Usbeck, Akhilesh Vyas, Akmal Khikmatullaev, Dharmen Punjani, Christoph Lange, et al. 2018. Why reinvent the wheel: Let's build question answering systems together. In *Proceedings of the 2018 World Wide Web Conference*. 1247–1256.
40. Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer Networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*. MIT Press, 2692–2700.
41. Wen-tau Yih, Ming-Wei Chang, Xiaodong He, and Jianfeng Gao. 2015. Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base. In *ACL*. Association for Computational Linguistics, 1321–1331.
42. Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cicero dos Santos, Bing Xiang, and Bowen Zhou. 2017. Improved Neural Relation Detection for Knowledge Base Question Answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 571–581.
43. Munazza Zaib, Wei Emma Zhang, Quan Z Sheng, Adnan Mahmood, and Yang Zhang. 2021. Conversational Question Answering: A Survey. *arXiv preprint arXiv:2106.00874* (2021).
44. Matthew J Smith. Getting value from artificial intelligence in agriculture. *Animal Production Science*, 2018.
45. Hao Fei, Shengqiong Wu, Meishan Zhang, Min Zhang, Tat-Seng Chua, and Shuicheng Yan. Enhancing video-language representations with structural spatio-temporal alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
46. Jim Virgona, Geoff Daniel, et al. Evidence-based agriculture-can we get there? *Agricultural Science*, 23(1):19, 2011.
47. Manlio Bacco, Paolo Barsocchi, Erina Ferro, Alberto Gotta, and Massimiliano Ruggeri. The digitisation of agriculture: a survey of research activities on smart farming. *Array*, 3:100009, 2019.
48. Hao Fei, Yafeng Ren, and Donghong Ji. Retrofitting structure-aware transformer language model for end tasks. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 2151–2161, 2020.
49. Mohit Jain, Pratyush Kumar, Ishita Bhansali, Q. Vera Liao, Khai Truong, and Shwetak Patel. Farmchat: A conversational agent to answer farmer queries. *ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(4), 2018.
50. Kweku Opoku-Agyemang, Bhaumik Shah, and Tapan S. Parikh. Scaling up peer education with farmers in india. In *Information and Communication Technologies and Development*, ICTD '17, pages 15:1–15:10. ACM, 2017. ISBN 978-1-4503-5277-2.

51. Shengqiong Wu, Hao Fei, Fei Li, Meishan Zhang, Yijiang Liu, Chong Teng, and Donghong Ji. Mastering the explicit opinion-role interaction: Syntax-aided neural transition system for unified opinion role labeling. In *Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence*, pages 11513–11521, 2022.
52. Wenxuan Shi, Fei Li, Jingye Li, Hao Fei, and Donghong Ji. Effective token graph modeling using a novel labeling strategy for structured sentiment analysis. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4232–4241, 2022.
53. Hao Fei, Yue Zhang, Yafeng Ren, and Donghong Ji. Latent emotion memory for multi-label emotion classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7692–7699, 2020.
54. Fengqi Wang, Fei Li, Hao Fei, Jingye Li, Shengqiong Wu, Fangfang Su, Wenxuan Shi, Donghong Ji, and Bo Cai. Entity-centered cross-document relation extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9871–9881, 2022.
55. Ling Zhuang, Hao Fei, and Po Hu. Knowledge-enhanced event relation extraction via event ontology prompt. *Inf. Fusion*, 100:101919, 2023.
56. Rob Lokers, Rob Knapen, Sander Janssen, Yke van Randen, and Jacques Jansen. Analysis of big data technologies for use in agro-environmental science. *Environmental modelling & software*, 84:494–504, 2016.
57. Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V Le. Qanet: Combining local convolution with global self-attention for reading comprehension. *arXiv preprint arXiv:1804.09541*, 2018.
58. Shengqiong Wu, Hao Fei, Yixin Cao, Lidong Bing, and Tat-Seng Chua. Information screening whilst exploiting! multimodal relation extraction with feature denoising and multimodal topic modeling. *arXiv preprint arXiv:2305.11719*, 2023.
59. Jundong Xu, Hao Fei, Liangming Pan, Qian Liu, Mong-Li Lee, and Wynne Hsu. Faithful logical reasoning via symbolic chain-of-thought. *arXiv preprint arXiv:2405.18357*, 2024.
60. Matthew Dunn, Levent Sagun, Mike Higgins, V Ugur Guney, Volkan Cirik, and Kyunghyun Cho. SearchQA: A new Q&A dataset augmented with context from a search engine. *arXiv preprint arXiv:1704.05179*, 2017.
61. Hao Fei, Shengqiong Wu, Jingye Li, Bobo Li, Fei Li, Libo Qin, Meishan Zhang, Min Zhang, and Tat-Seng Chua. Lasuie: Unifying information extraction with latent adaptive structure-aware generative language model. In *Proceedings of the Advances in Neural Information Processing Systems, NeurIPS 2022*, pages 15460–15475, 2022.
62. Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27, 2011.
63. Hao Fei, Yafeng Ren, Yue Zhang, Donghong Ji, and Xiaohui Liang. Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings in Bioinformatics*, 22(3), 2021.
64. Shengqiong Wu, Hao Fei, Wei Ji, and Tat-Seng Chua. Cross2StrA: Unpaired cross-lingual image captioning with cross-lingual cross-modal structure-pivoted alignment. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2593–2608, 2023.
65. Z. Sun, P.K. Sarma, W. Sethares, and E.P. Bucy. Multi-modal sentiment analysis using deep canonical correlation analysis. *Proc. Interspeech 2019*, pages 1323–1327, 2019.
66. Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*, 2016.
67. Hao Fei, Bobo Li, Qian Liu, Lidong Bing, Fei Li, and Tat-Seng Chua. Reasoning implicit sentiment with chain-of-thought prompting. *arXiv preprint arXiv:2305.11255*, 2023.
68. Aniruddha Tammewar, Alessandra Cervone, and Giuseppe Riccardi. Emotion carrier recognition from personal narratives. *Accepted for publication at INTERSPEECH*, 2021. URL https://arxiv.org/abs/2008.07481.
69. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423.
70. Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. Next-gpt: Any-to-any multimodal llm. *CoRR*, abs/2309.05519, 2023.
71. Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

 doi:10.20944/preprints202408.2225.v1

72. Jane Mills, Matthew Reed, Kamilla Skaalsveen, and Julie Ingram. The use of twitter for knowledge exchange on sustainable soil management. *Soil Use and Management*, 35(1):195–203, 2019.
73. Sarah Van Dalsem. An iphone in a haystack: the uses and gratifications behind farmers using twitter. Master's thesis, University of Nebraska, 2011.
74. Hao Fei, Shengqiong Wu, Wei Ji, Hanwang Zhang, Meishan Zhang, Mong-Li Lee, and Wynne Hsu. Video-of-thought: Step-by-step video reasoning from perception to cognition. In *Proceedings of the International Conference on Machine Learning*, 2024.
75. Naman Jain, Pranjali Jain, Pratik Kayal, Jayakrishna Sahit, Soham Pachpande, Jayesh Choudhari, et al. Agribot: agriculture-specific question answer system. *IndiaRxiv*, 2019.
76. Hao Fei, Shengqiong Wu, Wei Ji, Hanwang Zhang, and Tat-Seng Chua. Dysen-vdm: Empowering dynamics-aware text-to-video diffusion with llms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7641–7653, 2024.
77. Mihir Momaya, Anjnya Khanna, Jessica Sadavarte, and Manoj Sankhe. Krushi–the farmer chatbot. In *2021 International Conference on Communication information and Computing Technology (ICCICT)*, pages 1–6. IEEE, 2021.
78. Hao Fei, Fei Li, Chenliang Li, Shengqiong Wu, Jingye Li, and Donghong Ji. Inheriting the wisdom of predecessors: A multiplex cascade framework for unified aspect-based sentiment analysis. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI*, pages 4096–4103, 2022.
79. Shengqiong Wu, Hao Fei, Yafeng Ren, Donghong Ji, and Jingye Li. Learn from syntax: Improving pair-wise aspect and opinion terms extraction with rich syntactic knowledge. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pages 3957–3963, 2021.
80. Bobo Li, Hao Fei, Lizi Liao, Yu Zhao, Chong Teng, Tat-Seng Chua, Donghong Ji, and Fei Li. Revisiting disentanglement and fusion on modality and context in conversational multimodal emotion recognition. In *Proceedings of the 31st ACM International Conference on Multimedia, MM*, pages 5923–5934, 2023.
81. Hao Fei, Qian Liu, Meishan Zhang, Min Zhang, and Tat-Seng Chua. Scene graph as pivoting: Inference-time image-free unsupervised multimodal machine translation with visual scene hallucination. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5980–5994, 2023.
82. Hao Fei, Shengqiong Wu, Hanwang Zhang, Tat-Seng Chua, and Shuicheng Yan. Vitron: A unified pixel-level vision llm for understanding, generating, segmenting, editing. 2024.
83. Sanjeev Arora, Yingyu Liang, and Tengyu Ma. A simple but tough-to-beat baseline for sentence embeddings. In *ICLR*, 2017.
84. Abbott Chen and Chai Liu. Intelligent commerce facilitates education technology: The platform and chatbot for the taiwan agriculture service. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 11:1–10, 01 2021.
85. Shengqiong Wu, Hao Fei, Xiangtai Li, Jiayi Ji, Hanwang Zhang, Tat-Seng Chua, and Shuicheng Yan. Towards semantic equivalence of tokenization in multimodal llm. *arXiv preprint arXiv:2406.05127*, 2024.
86. Jingye Li, Kang Xu, Fei Li, Hao Fei, Yafeng Ren, and Donghong Ji. MRN: A locally and globally mention-based reasoning network for document-level relation extraction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1359–1370, 2021.
87. Hao Fei, Shengqiong Wu, Yafeng Ren, and Meishan Zhang. Matching structure for dual learning. In *Proceedings of the International Conference on Machine Learning, ICML*, pages 6373–6391, 2022.
88. Hu Cao, Jingye Li, Fangfang Su, Fei Li, Hao Fei, Shengqiong Wu, Bobo Li, Liang Zhao, and Donghong Ji. OneEE: A one-stage framework for fast overlapping and nested event extraction. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1953–1964, 2022.
89. Isakwisa Gaddy Tende, Kentaro Aburada, Hisaaki Yamaba, Tetsuro Katayama, and Naonobu Okazaki. Proposal for a crop protection information system for rural farmers in tanzania. *Agronomy*, 11(12):2411, 2021.
90. Hao Fei, Yafeng Ren, and Donghong Ji. Boundaries and edges rethinking: An end-to-end neural model for overlapping entity relation extraction. *Information Processing & Management*, 57(6):102311, 2020.
91. Jingye Li, Hao Fei, Jiang Liu, Shengqiong Wu, Meishan Zhang, Chong Teng, Donghong Ji, and Fei Li. Unified named entity recognition as word-word relation classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10965–10973, 2022.

92. Mohit Jain, Pratyush Kumar, Ishita Bhansali, Q Vera Liao, Khai Truong, and Shwetak Patel. Farmchat: a conversational agent to answer farmer queries. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–22, 2018.
93. Shengqiong Wu, Hao Fei, Hanwang Zhang, and Tat-Seng Chua. Imagine that! abstract-to-intricate text-to-image synthesis with scene graph hallucination diffusion. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pages 79240–79259, 2023.
94. Hao Fei, Tat-Seng Chua, Chenliang Li, Donghong Ji, Meishan Zhang, and Yafeng Ren. On the robustness of aspect-based sentiment analysis: Rethinking model, data, and training. *ACM Transactions on Information Systems*, 41(2):50:1–50:32, 2023.
95. Yu Zhao, Hao Fei, Yixin Cao, Bobo Li, Meishan Zhang, Jianguo Wei, Min Zhang, and Tat-Seng Chua. Constructing holistic spatio-temporal scene graph for video semantic role labeling. In *Proceedings of the 31st ACM International Conference on Multimedia, MM*, pages 5281–5291, 2023.
96. Shengqiong Wu, Hao Fei, Yixin Cao, Lidong Bing, and Tat-Seng Chua. Information screening whilst exploiting! multimodal relation extraction with feature denoising and multimodal topic modeling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14734–14751, 2023.
97. Hao Fei, Yafeng Ren, Yue Zhang, and Donghong Ji. Nonautoregressive encoder-decoder neural framework for end-to-end aspect-based sentiment triplet extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9):5544–5556, 2023.
98. Yu Zhao, Hao Fei, Wei Ji, Jianguo Wei, Meishan Zhang, Min Zhang, and Tat-Seng Chua. Generating visual spatial description via holistic 3D scene understanding. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7960–7977, 2023.
99. Nikulsinh M Chauhan et al. Information hungers of the rice growers. *Agriculture update*, 7(1/2):72–75, 2012.
100. Ellen M Voorhees, Donna K Harman, et al. *TREC: Experiment and evaluation in information retrieval*, volume 63. Citeseer, 2005.
101. Rodrigo Nogueira, Wei Yang, Kyunghyun Cho, and Jimmy Lin. Multi-stage document ranking with BERT. *arXiv preprint arXiv:1910.14424*, 2019.
102. Shengyao Zhuang and Guido Zuccon. Fast passage re-ranking with contextualized exact term matching and efficient passage expansion. *CoRR*, abs/2108.08513, 2021. URL https://arxiv.org/abs/2108.08513.
103. Michael Bendersky and W Bruce Croft. Discovering key concepts in verbose queries. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 491–498, 2008.
104. Hamed Zamani and Nick Craswell. Macaw: An extensible conversational information seeking platform. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2193–2196, 2020.
105. Marcin Kaszkiel and Justin Zobel. Passage retrieval revisited. In *ACM SIGIR Forum*, volume 31, pages 178–185. ACM New York, NY, USA, 1997.
106. Xiaoyong Liu and W Bruce Croft. Passage retrieval based on language models. In *Proceedings of the eleventh international conference on Information and knowledge management*, pages 375–382, 2002.
107. Nick Craswell, Bhaskar Mitra, Emine Yilmaz, and Daniel Campos. Overview of the trec 2020 deep learning track. *arXiv preprint arXiv:2102.07662*, 2021.
108. Alistair Moffat, Falk Scholer, Paul Thomas, and Peter Bailey. Pooled evaluation over query variations: Users are as diverse as systems. In *proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1759–1762, 2015.
109. Peter Bailey, Alistair Moffat, Falk Scholer, and Paul Thomas. Uqv100: A test collection with query variability. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 725–728, 2016.
110. Chao-Chun Hsu, Eric Lind, Luca Soldaini, and Alessandro Moschitti. Answer generation for retrieval-based question answering systems. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4276–4282, 2021.
111. Hannah Bast and Claudius Korzen. A benchmark and evaluation for text extraction from pdf. In *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 1–10. IEEE, 2017.

doi:10.20944/preprints202408.2225.v1

112. Michail Salampasis, Norbert Fuhr, Allan Hanbury, Mihai Lupu, Birger Larsen, and Henrik Strindberg. Integrating ir technologies for professional search. In *European Conference on Information Retrieval*, pages 882–885. Springer, 2013.
113. John I Tait. An introduction to professional search. In *Professional search in the modern world*, pages 1–5. Springer, 2014.
114. Suzan Verberne, Jiyin He, Udo Kruschwitz, Gineke Wiggers, Birger Larsen, Tony Russell-Rose, and Arjen P de Vries. First international workshop on professional search. In *ACM SIGIR Forum*, volume 52, pages 153–162. ACM New York, NY, USA, 2019.
115. Nandan Thakur, Nils Reimers, Andreas Rücklé, Abhishek Srivastava, and Iryna Gurevych. Beir: A heterogenous benchmark for zero-shot evaluation of information retrieval models. *arXiv preprint arXiv:2104.08663*, 2021.