IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 December 2023, accepted 22 January 2024, date of publication 13 February 2024, date of current version 20 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365522

RESEARCH ARTICLE

SE-HCL: Schema Enhanced Hybrid Curriculum Learning for Multi-Turn Text-to-SQL

YIYUN ZHANG, SHENG'AN ZHOU, AND GENGSHENG HUANG

Institute of Electronic Information, Guangdong Vocational Institute of Public Administration, Guangzhou 510800, China

Corresponding author: Yiyun Zhang (zyyjinzhou@163.com)

This work was supported in part by Guangdong Vocational College under Grant X2021ZLGC5114 and Grant X2021ZLGC2211, and in part by the Department of Education of Guangdong Province under Grant 2018GkQNCX125.

ABSTRACT Existing multi-turn Text-to-SQL approaches, mainly use data in a randomized order when training the model, ignoring the rich structural information contained in the dialog and schema. In this paper, we propose to use curriculum learning (CL) to better leverage the curriculum structure of schema, query, and dialog for multi-turn question-query pairs. We design a model-agnostic framework named Schema Enhanced Hybrid Curriculum Learning (SE-HCL) for multi-turn Text-to-SQL to help the models gain a full contextual semantic understanding. Concretely, We measure the difficulty of the data from both a structural and model perspective. In terms of data structure, we mainly consider the turns of the question and the complexity of the schema and SQL query. Accordingly, we designed a data course module to dynamically adjust the difficulty of the data based on the convergence of the model and the schema enhancement method we designed. In terms of the model, we propose a scoring module that will judge the difficulty of a problem based on whether the model could solve the question effectively. Finally, we will consider both aspects and design a hybrid curriculum to determine the flow of model training. Our experiments show that our proposed method improves SQL-generated performance over previous state-of-the-art models on SparC and CoSQL, especially for hard and long-turn questions.

INDEX TERMS Natural language processing, semantic parsing, multi-turn text-to-SQL, curriculum learning.

I. INTRODUCTION

The Text-to-SQL task is a natural language processing (NLP) challenge that involves converting natural language questions into structured SQL (Structured Query Language) database queries. It aims to bridge the gap between human language and the language used to interact with relational databases. Datasets such as WikiSQL [1] and Spider [2] were constructed to explore SQL-generated algorithms. Spider is a challenging cross-domain text-to-SQL dataset where the database domains corresponding to the question in the test set do not intersect with the training set. Recent works on Spider [3], [4], [5], [6], [7] have shown that modeling relations between question and schema could effectively promote performance.

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

However, in real scenarios, as shown in Figure 1, in order to get answers, users need to conduct multiple turns of questions and answers with the dialogue system to comprehensively explore the data. The multi-turn text-to-SQL task is an extension of the traditional text-to-SQL task, designed to handle complex natural language interactions with a relational database across multiple conversational turns. In this task, the goal is to generate SQL queries that correctly and coherently respond to a series of conversational exchanges, where the database state and user's intent may evolve with each turn. The task of multi-turn text-to-SQL has more challenges and requires modeling not only the relational information between questions and schema but also the multi-turn conversation information. However, modeling that addresses multiple factors simultaneously tends to achieve sub-optimal performance.

In prior research in the field of multi-turn text-to-SQL, the primary emphasis has been on harnessing contextual

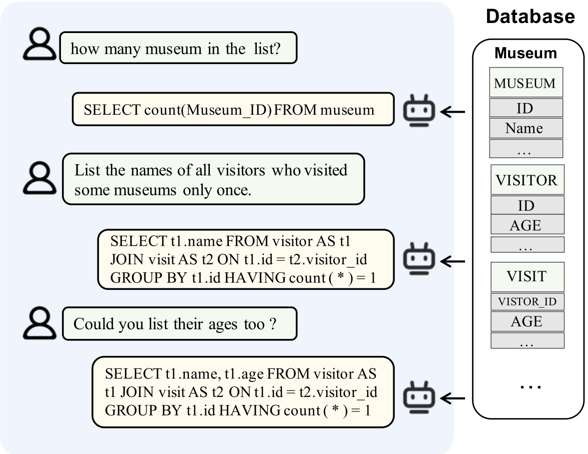


FIGURE 1. An example of the multi-turn text-to-SQL task. Given question, history context and database, the model needs to generate structured SQL query.

information [8], [9], [10]. In a multi-turn text-to-SQL task, models are confronted with the challenge of concurrently handling both relational modeling and contextual modeling. This entails the model's ability to effectively establish the entity mapping relationship between the user query and the database schema, while also comprehending the underlying intent of the current inquiry in the context provided. Several prior studies [9], [11], [12], [13] have employed neural network encoders that concatenate the current question, question context, and schema. Concurrently, a number of approaches have directly incorporated historically generated SQL queries [8], [14], [15], [16] to aid the model in SQL parsing for the present question. However, these methods have tended to overlook the exploration of the wealth of structural information embedded within the dataset.

In this paper, we propose a novel Schema Enhanced Hybrid Curriculum Learning framework to fully explore the structural information in the data and enhance the model's ability to understand structured information and generate structured queries. We designed a Schema Enhance Data Augmentation Module (SE-DAM), which contains three data enhancement strategies. Based on the data enhanced by SE-DAM, we propose a curriculum learning method with a hybrid update strategy. Furthermore, a curriculum judger is adapted to determine whether the model has completed curriculum learning.

Our main contributions can be summarized as follows:

- We propose a heuristic Schema Enhanced data augmentation Module (SE-DAM). We combined schema to fully explore the structural information in the data and proposed several data enhancement methods to enhance the model's ability to understand structured information and generate structured queries.
- We propose a curriculum learning framework (SE-HCL) with a hybrid update strategy. SE-HCL combines structure scores and model scores to determine the data sampling and order of course learning.
- We evaluate SC-HCL on two multi-turn text-to-SQL datasets SParC [17] and CoSQL [18]. We conduct a comprehensive evaluation of our training framework on multiple baseline methods, and our experimental results demonstrate the remarkable capabilities of the framework.

II. RELATED WORKS

A. TEXT-TO-SQL

The text-to-SQL task is centered around the objective of mapping natural language queries to SQL queries that are relevant to a database. Spider [2] stands as a well-recognized cross-domain single-turn dataset. A substantial body of research [3], [4], [5] has established the efficacy of modeling the relationship between the query and the database schema, particularly when applied to improving performance on the Spider dataset. Wang et al. [3] introduced the use of a relation-aware Transformer (RAT) [19] to encode the relational positions within sentence representations. This approach has found extensive adoption in text-to-SQL research, including works by Wang et al. [3], Lin et al. [4], Scholak et al. [20], and Yu et al. [21], for encoding the schema-linking relationships between natural language queries and the structured database schema. Cao et al. [5] have further advanced the modeling of relations through the application of line graph neural networks. Text-to-SQL in multi-turn dialogue scenarios requires solving complex contexts and complex structural references and links of schema at the same time, which is even more challenging [8], [15], [16]. Additionally, research by Cai and Wan [14] and Hui et al. [11] has leveraged graph neural networks to jointly encode multi-turn questions and schema information. Building upon the accomplishments of pre-trained models like T5, BERT, ALM, GanLM, and BART [22], [23], [24], [25], ScoRE [9] and Star [26] design pre-training framework which leverage contextual information to enrich natural language (NL) utterance and table schema representations for text-to-SQL conversations. Scholak et al. [12] have taken a more straightforward approach by imposing constraints on the auto-regressive decoders of super-large pre-trained language models, specifically T5-3B. Chen et al. [27] propose a dual learning method to generate rewritten question data with in-domain QR annotations and directly employ these rewritten questions for SQL query generation. RASAT [7] and QURG [28] introduce the co-reference relationship between dialogue histories in RAT to improve the model's understanding of dialogues. In this paper, we refer to RASAT and QURG, and continue to use RAT which introduces the co-reference relationship, innovatively design the curriculum learning method to the multi-turn text-to-SQL task and use schema enhancement to strengthen the dialogue understanding ability in multi-turn dialogue scenarios.

B. CURRICULUM LEARNING

Curriculum Learning (CL) constitutes a training strategy employed in deep learning, where a model is trained progressively from simpler to more complex data, mirroring the cognitive learning sequence found in human curricula. Serving as an accessible and adaptable tool, the CL strategy has showcased its formidable efficacy in enhancing the generalization capabilities and convergence speed of diverse models across a broad spectrum of domains, including but not limited to computer vision and natural language processing(NLP). The initial endeavor to introduce a curriculum-based approach to supervised learning can be traced back to Elman's work in the field of Natural Language Processing (NLP), specifically in the domain of grammar learning using recurrent neural networks [29]. Elman's work underscored the significance of the "starting small" principle, emphasizing the restriction of the scope of data exposure to neural networks during their initial training phases. This concept of gradually increasing the complexity of training data has also been revisited in subsequent research, as evident in the studies by Rohde [30] and Krueger [31].

A frequently explored application of Curriculum Learning (CL) is Neural Machine Translation, wherein the datasets exhibit significant variability in terms of quality, complexity, and noise, as discussed by Kumar et al. [32]. Correspondingly, CL has found utility in a variety of other NLP tasks characterized by noisy or heterogeneous data, such as natural language support [33], relationship extraction [34], reading comprehension [35], and more.

III. PRELIMINARIES

In this section, We first give a formal definition of the task, and then an introduction is given to the relation-aware transformer used in the method.

A. TASK FORMULATION

Given conversation $\mathcal{Q} = \{q_1, q_2, \cdot, q_t\}$, historical questions $q_{<t} = \{q_1, q_2, \ldots, q_{t-1}\}$, and SQL queries $y_{<t} = \{y_1, y_2, \ldots, y_{t-1}\}$, and schema $\mathcal{S} = \langle\mathcal{T}, \mathcal{C}\rangle$, which consists of a series of tables $T = \{t_1, \ldots, t_{|T|}\}$ and columns $C = \{c_1, \ldots, c_{|C|}\}$, the multi-turn text-to-SQL task map q_t to the SQL query y_t.

B. RELATION-AWARE TRANSFORMER (RAT)

The Relation-Aware Transformer (RAT) is a variant of the standard Transformer model [36]. RAT enhances the Transformer's capabilities by incorporating predefined relation features through the inclusion of relation embedding within the self-attention mechanism, as shown in Figure 3.

The standard Transformer model is an architectural framework comprising a series of multi-head self-attention layers. This architecture has found extensive application in tasks involving the processing of sequential inputs. For a given input embedding sequence $\boldsymbol{X} = \boldsymbol{x}ii = 1^n$, where $\boldsymbol{x}_i \in \mathbb{R}^{d_x}$, each Transformer layer transforms the input element $\boldsymbol{x}_i$ into $\boldsymbol{y}_i$ using H heads, as described below:

$$e_{ij}^{(h)} = \frac{\boldsymbol{x}_i \boldsymbol{W}_Q^{(h)} \left(\boldsymbol{x}_j \boldsymbol{W}_K^{(h)}\right)^\top}{\sqrt{d_z/H}} \tag{1}$$

$$\alpha_{ij}^{(h)} = \underset{j}{\text{Softmax}} \left(e_{ij}^{(h)}\right) \tag{2}$$

$$z_i^{(h)} = \sum_{j=1}^{n} \alpha_{ij}^{(h)} \left(\boldsymbol{x}_j \boldsymbol{W}_V^{(h)}\right) \tag{3}$$

$$z_i = \text{Concat}(z_i^{(1)}, \ldots, z_i^{(H)}) \tag{4}$$

$$\widetilde{\boldsymbol{y}}_i = \text{LayerNorm}(\boldsymbol{x}_i + z_i) \tag{5}$$

$$\boldsymbol{y}_i = \text{LayerNorm}(\widetilde{\boldsymbol{y}}_i + \text{FC}(\text{ReLU}(\widetilde{\boldsymbol{y}}_i))) \tag{6}$$

where h denotes the h-th head, $a_{ij}^{(h)}$ is the attention weights, Concat$(\cdot)$ is a concatenate operation, FC$(\cdot)$ is a full-connected layer, LayerNorm$(\cdot)$ is layer normalization, ReLU$(\cdot)$ is the activation function and $\boldsymbol{W}_Q^{(h)}$,$\boldsymbol{W}_K^{(h)}$,$\boldsymbol{W}_V^{(h)}$ are learnable projection parameters. Compared to the standard Transformer model, the RAT incorporates the utilization of learnable relation embeddings within the self-attention module as:

$$e_{ij}^{(h)} = \frac{\boldsymbol{x}_i \boldsymbol{W}_Q^{(h)} \left(\boldsymbol{x}_j \boldsymbol{W}_K^{(h)} + \boldsymbol{r}_{ij}^K\right)^\top}{\sqrt{d_z/H}} \tag{7}$$

$$z_i^{(h)} = \sum_{j=1}^{n} \alpha_{ij}^{(h)} \left(\boldsymbol{x}_j \boldsymbol{W}_V^{(h)} + \boldsymbol{r}_{ij}^V\right)^\top \tag{8}$$

where $\boldsymbol{r}_{ij}$ is the pre-defined relation embedding between input elements $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

IV. METHODS

A. MODEL OVERVIEW

In Figure 2, our proposed framework consists of the structural scorer, the curriculum training loop, and the curriculum judger. Specifically, given train data, we first use the data augmentation module to structurally augment the data and then the structural scorer scores the augmented data. Then we will use the scored data to train the model according to the strategy of curriculum learning. In the training loop, we set the model scorer to score the data according to the model's confidence in the generated sentences and whether the generated sentences are correct, and then we will mix the scores of the updated data again for the next round of training. At the same time, our course judges decide whether to end the curriculum training according to the degree of convergence of the model.

B. STRUCTURAL AUGMENTATION MODULE

This module consists of the data augmentation module and the structural scorer. First of all, we weaken or enhance the data (collectively referred to as data enhancement). The enhancement of data mainly includes three aspects: enhancement of dialogue rounds, enhancement of Schema, and query statement Coarse-to-fine. The detailed enhanced method is shown in Table 2.

TABLE 1. **All relations used in our experiment. *Q* stands for question, *T* stands for table and *C* stands for column in table.**

Relations	Description
Q-Q-Distance-d Q-Q-Identity Q-Q-Generic	The distance between the question item q_a and the question item q_b in the input question is d Question item q_a is question item q_b itself Question item q_a and question item q_b has no pre-defined relation
Q-*-Generic	Question item q_a and schema item s_b has no pre-defined relation
Q-T-Exactmatch Q-T-Partialmatch Q-T-Nomatch	Question item q_a is spelled exactly/partially/not the same as table item t_b
Q-C-Exactmatch Q-C-Partialmatch Q-C-Nomatch	Question item q_a is spelled exactly/partially/not the same as column item c_b
Q-C-Valuematch	Question item q_a is spelled exactly the same as a value in column item c_b
*-Q-Generic	Schema item s_a and question item q_b has no pre-defined relation
--Identity	Schema item s_a is schema item s_b itself
*-T-Generic	Schema item s_a and table item t_b has no pre-defined relation
*-C-Generic	Schema item s_a and column item c_b has no pre-defined relation
T-Q-Exactmatch T-Q-Partialmatch T-Q-Nomatch	Table item t_a is spelled exactly/partially/not the same as question item q_b
T-*-Generic	Table item t_a and schema item s_b has no pre-defined relation
T-T-Generic	Table item t_a and table item t_b has no pre-defined relation
T-T-Identity	Table item t_a is table item t_b itself
T-T-Foreign-Forward T-T-Foreign-Backward T-T-Foreign-Both	One or more columns in table item t_a is a foreign key for certain column in table item t_b One or more columns in table item t_b is a foreign key for certain column in table item t_a Table item t_a and t_b satisfy both "T-T-Foreign-Forward" and "T-T-Foreign-Backward" relations
T-C-Primary T-C-Has	Column item c_a is the primary key for table item t_b Column item c_a belongs to table item t_b
T-C-Generic	Table item t_a and column item c_b has no pre-defined relation
C-Q-Exactmatch C-Q-Partialmatch C-Q-Nomatch	Column item c_a is spelled exactly/partially/not the same as table item t_b
C-Q-Valuematch	Column item c_a is spelled exactly the same as a value in question item q_b
C-*-Generic	Column item c_a and schema item s_b has no pre-defined relation
C-T-Pk C-T-Has C-T-Generic	Column item c_a is the primary key for table item t_b Column item c_a belongs to table item t_b Column item c_a and table item t_b has no pre-defined relation
C-C-Identity	Column item c_a is column item c_b itself
C-C-Sametable	Column item c_a and column item c_b are in the same table
C-C-Foreign-Forward C-C-Foreign-Backward	Column item c_a has a forward/reverse foreign key constraint relation with Column item c_b
C-C-Generic	Column item c_a and column item c_b has no pre-defined relation
No-Relation	Item a and item b has no relation

For the structural scorer, we mainly use human prior knowledge to design a heuristic scoring mechanism based on structural factors such as the number of turns and the complexity of the SQL query. Specifically, it mainly includes: 1) the current dialogue turn t, whether the current question is a continuation of the previous question $Bool_{follow}$, whether there is omission and reference $hasRef$ in the current question. 2) the number of table or column used ($NumS_{use}$) and not used ($NumS_{unuse}$). 3) The complexity of the SQL query($Score_{complex}$), the number of nesting(Num_{nest}), the number of table joins (Num_{join}) and the number of conditions(Num_{cond}).

$$Score_{struct} = Score_{turn} + Score_{schema} + Score_{query} \tag{9}$$

C. TEXT-TO-SQL MODEL

In our work, we use RASAT [7] as our text-to-SQL model. RASAT follows the architectural framework of T5,

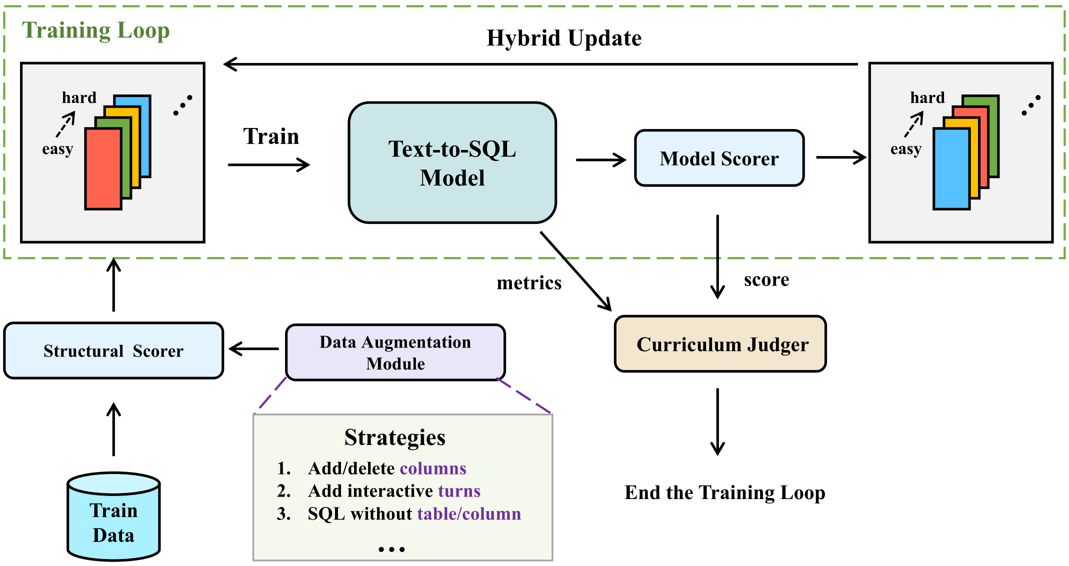


FIGURE 2. Overview of SE-HCL, including Data Augmentation Module (DAM) and Training Loop. In DAM, the training data is first augmented and then the augmented data is scored by a structural scorer. In the training loop, the data is first sorted according to its score (combined structure score and model score), and then the top K difficulty data is selected according to the order and sent to the text-to-SQL model for training. The trained model rescores the data, and then the score will be combined with the structure score to update.

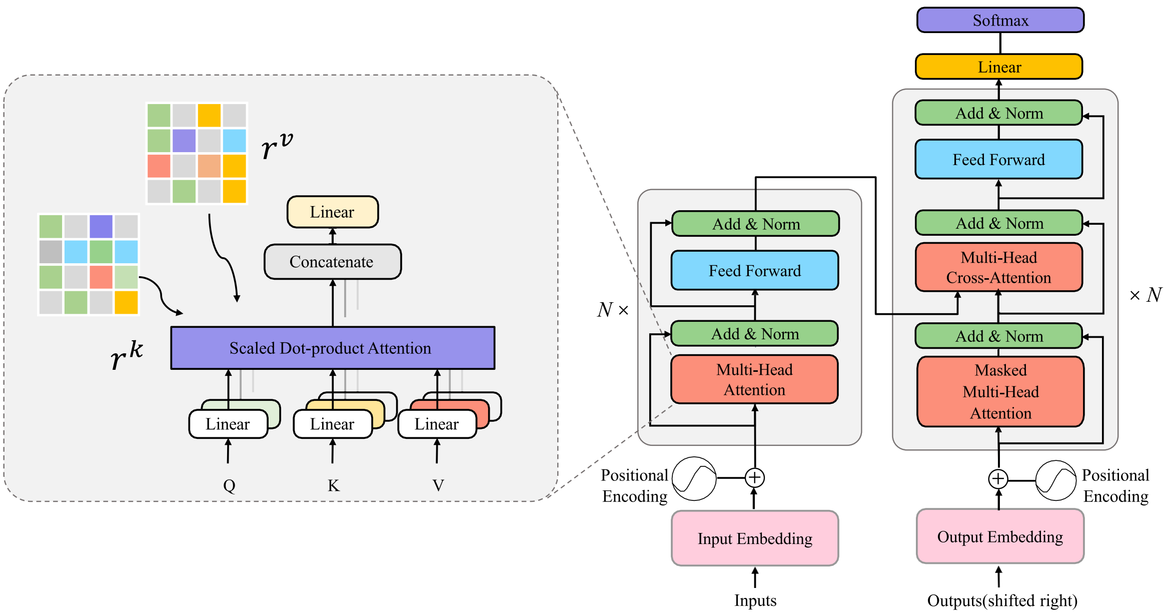


FIGURE 3. The illustration of Relation-Aware Transformer (RAT).

which adopts a sequence-to-sequence (seq2seq) structure comprising N layers of both encoders and decoders. Notably, RASAT makes a significant modification to the standard self-attention mechanism within the encoder by replacing it with relation-aware self-attention. This modification introduces two supplementary relation embeddings into the model's architecture. The utilization of relation-aware self-attention is a pivotal aspect of RASAT, enhancing its ability to capture and represent complex relationships within the input data. This adaptation allows for more sophisticated information processing and context comprehension. For more details about RASAT please refer to [7]. The relations used in our text-to-SQL model are shown in Table 1.

D. CURRICULUM TRAINING LOOP

In the training loop, we first arrange the data in order from easy to difficult according to the structure score and model

TABLE 2. Description of enhanced examples for each enhanced method.

Data Enhance Method	Example	Example Description
Enhance Dialogue Turn Delete or add dialog history unrelated to the current question	**Turn 1:** How many museum in the list? ~~**Turn 2:** List the names of all visitors who visited some museums only once.~~ **Add Turn:** List the names of all visitors under the age of 30. **Turn3:** Could you list their ages too?	•Add new turn in dialog history •Delete Turn2
Enhance Schema Remove or add tables or columns in tables that are not relevant to the current question	**Database:** College **Table:** classroom **Columns:** ~~building~~ | room_number | capacity | location ~~**Table:** course~~ ~~**Columns:** course_id | title | dept_name | credits~~ **SQL:** SELECT room_number FROM classroom WHERE capacity > 10	•Delete "building" and add "location" in table "classroom" •Delete table "course"
SQL Query Coarse-to-Fine Mask specific column names and condition values in the SQL query early in the course	**Before Enhance:** SELECT Starting_Year FROM technician WHERE Team = "CLE" INTERSECT SELECT Starting_Year FROM technician WHERE Team = "CWS" **After Enhance:** SELECT [COL1] FROM technician WHERE [COL2] = [VALUE1] INTERSECT SELECT [COL3] FROM technician WHERE [COL4] = [VALUE2]	•Mask column name and value in SQL

score according to the corresponding weights (in the first training loop, due to the lack of model scores, the weights corresponding to the model scores are reset is 0). After sorting the data according to order, select more difficult data according to a certain proportion to train the text-to-SQL model. At the same time, we calculate the model's score for the data based on the perplexity of the query statement generated by the model. This score will be used to determine the end of the training cycle and to update the initial data score for the next round of training. The specific update method uses momentum update to ensure the stability of the score, as shown in the formula 10.

$$Score_t = \beta Score_{t-1} + (1-\beta) Score_{model} \tag{10}$$

$$P = 1 - \alpha t \tag{11}$$

$$\mathrm{PPL}(X) = \exp\left\{-\frac{1}{t}\sum_{i}^{t} \log p_\theta\left(x_i \mid x_{<i}\right)\right\} \tag{12}$$

E. CURRICULUM JUDGE

In order to judge whether the model has converged, we set up a course discriminator to judge whether the model has completed the course learning based on the indicators and perplexity of the model in the past t rounds, thus ending the training cycle. The complete algorithm process is shown in Algorithm 1.

Algorithm 1 Curriculum Training

Input: Training data D_{train}
Output: Trained model θ_t

```
Sort the train data D_train with structrual score Score_struct
Train epoch t = 1
End training flag f
while end is False do
    According to the score, select the data with a percentage of P as D_t from difficult to easy.
    Train model θ_t with D_t from θ_{t-1}.
    Use θ_t to score D_t based on metric PPL
    According to the metrics and PPL of the model on the validation set, update the value of f
    Update D_t score.
    t+ = 1
end while
return θ_t
```

V. EXPERIMENTS

In this section, we describe the experimental setups and evaluate the effectiveness of our proposed framework. Since our training framework is model-agnostic, we combine SE-HCL with different models to verify the effectiveness of our

TABLE 3. Detailed statistics for SParC dataset [17] and CoSQL dataset [18].

Dataset	Number of Questions	Train / Dev	Database / Domain	User Questions	Average Turn	System Response
SParC	4,298	3,034 / 422	200 / 138	15,598	3.0	✗
CoSQL	3,007	2,164 / 293	200 / 138	12,726	5.2	✓

approach and conduct several ablation experiments. We also compare our method with others in terms of conversation turns and SQL difficulty, demonstrating the advantages of our method on multi-turn and difficult questions.

A. EXPERIMENTAL SETUP

a: DATASETS

We train our SE-HCL on two large-scale cross-domain context-dependent text-to-SQL datasets, SparC [17] and CoSQL [18]. The details of those datasets are organized in Table 3.

b: EVALUATION METRICS

We evaluate from two aspects: the structural accuracy of the SQL and the execution accuracy of the SQL. We utilize the official assessment criteria: Exact Match accuracy (EM) and Execution accuracy (EX). EM evaluates whether the entire predicted sequence matches the ground truth SQL query (excluding values), while EX assesses whether the predicted executable SQL queries (including values) yield the same results as the corresponding gold-standard SQL queries. In the case of SParC and CoSQL, which encompass multi-turn dialogues, both EM and EX can be computed at both the question and interaction levels. Consequently, there are four evaluation metrics for these two datasets, specifically Question-level Exact Match (QEM), Interaction-level Exact Match (IEM), Question-level Execution accuracy (QEX), and Interaction-level Execution accuracy (IEX). For IEM and IEX, if all the predicted SQL in interaction is correct, the interaction match score is 1.0, otherwise, the score is 0.0.

c: IMPLEMENTATION DETAILS

We set the learning rate to $1e$-4, batch size to 2048, and the maximum gradient norm to 10. During inference, we set the beam size to 5 for SQL parsing. Models are trained with 4 NVIDIA A100-80GB GPU cards. Our code is provided in the supplementary material.

d: EXPERIMENTAL RESULTS

As shown in Table 4, we combine HCL with RASAT and compare it with previous works on SParC and CoSQL datasets. RASAT achieves comparable performance to previous state-of-the-art methods, including HIE-SQL [8], UNIFIEDSKG [37] and RASAT [7]. RASAT combined with our method can achieve better performance. It emphasizes the importance of our curriculum training strategy for multi-turn text-to-SQL tasks.

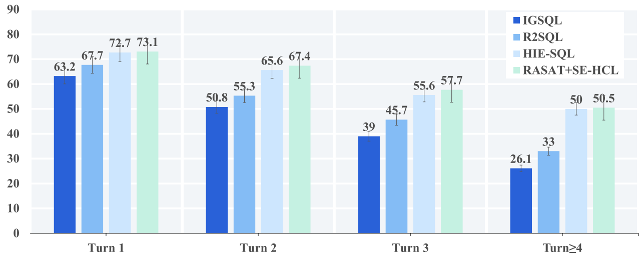


FIGURE 4. Performances of previous works and RASAT+SE-HCL in different turns on SparC.

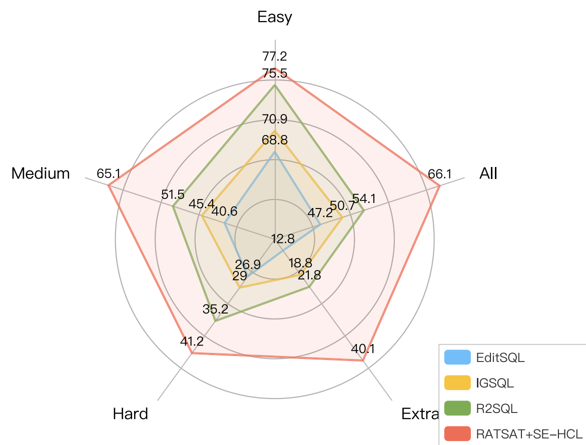


FIGURE 5. Performances of previous works and RASAT+HCL in different difficulty levels on SparC.

In order to delve deeper into the investigation of the benefits offered by SE-HCL in the realm of contextual comprehension, we conducted an assessment of its performance using various question turns on the SparC dataset, as depicted in Figure 4. This evaluation involved a comparative analysis between SE-HCL and previously established robust methods. It's important to note that as the number of turns in the questions increased, the inherent complexity of the task also escalated. This is primarily due to the fact that models are required to handle co-reference and omission with longer dependencies, making the task more challenging.

Besides, SE-HCL can achieve more improvements as the interaction turn increases. Furthermore, we evaluate the performance of SE-HCL on the different difficulty levels of target SQL as shown in Figure 5, and we observe that SE-HCL surpass previous works.

VI. ANALYSIS

A. ABLATION STUDY

In order to assess the impact of our proposed structural augmentation and curriculum training strategies, we undertake

TABLE 4. Performances on the SParC and CoSQL. HCL combined with RASAT outperforms the performance of previous methods.

Models	SParC QEM	SParC IEM	SParC QEX	SParC IEX	CoSQL QEM	CoSQL IEM	CoSQL QEX	CoSQL IEX
EditSQL [16]	47.2	29.5	-	-	39.9	12.3	-	-
GAZP [38]	48.9	29.7	47.8	-	42.0	12.3	38.8	-
TreeSQL v2 [39]	52.6	34.4	50.4	29.4	-	-	-	-
IGSQL [14]	50.7	32.5	-	-	44.1	15.8	-	-
RichContext [40]	52.6	29.9	-	-	41.0	14.0	-	-
IST-SQL [15]	47.6	29.9	-	-	44.4	14.7	-	-
R^2SQL [11]	54.1	35.2	-	-	45.7	19.5	-	-
DELTA [27]	58.6	35.6	-	-	51.7	21.5	-	-
SCoRE [9]	62.2	42.5	-	-	52.1	22.0	-	-
RAT-SQL+TC [13]	64.1	44.1	-	-	-	-	-	-
HIE-SQL [8]	64.7	45.0	-	-	56.4	28.7	-	-
UNIFIEDSKG [37]	61.5	41.9	67.3	46.4	54.1	22.8	62.2	26.2
RASAT [7]	65.0	45.7	69.9	50.7	56.2	25.9	63.8	34.8
RASAT+SE-HCL	**67.2**	**48.5**	**71.5**	**53.3**	**57.2**	**28.1**	**66.3**	**37.2**

TABLE 5. Ablation study of our method on the SParC and CoSQL, where RASAT is baseline method, ID ⑦ is RASAT+SE-HCL.

ID	Method	SParC QEM	SParC IEM	SParC QEX	SParC IEX	CoSQL QEM	CoSQL IEM	CoSQL QEX	CoSQL IEX
①	RASAT	65.0	45.7	69.9	50.7	56.2	25.9	63.8	34.8
②	① + Enhance dialogue turn	64.8	46.1	69.5	51.3	56.8	27.2	64.3	35.6
③	② + Enhance Schema	66.1	45.4	**72.1**	53.3	55.8	27.2	65.7	34.3
④	③ + SQL Coarse-to-Fine	66.6	46.5	70.4	52.8	56.2	28.5	65.2	36.6
⑤	④ + Structural scorer	65.5	47.7	70.7	51.3	55.3	**28.9**	64.3	34.1
⑥	⑤ + Model Scorer	63.2	47.1	70.4	50.7	56.2	27.6	65.5	35.7
⑦	⑥ + Judger	**67.2**	**48.5**	71.5	**53.3**	**57.2**	28.1	**66.3**	**37.2**

TABLE 6. SE-HCL with different baseline models on SparC.

Model	With SE-HCL	SParC QEM	SParC IEM	SParC QEX	SParC IEX
RAT-SQL	✗	60.4	42.7	-	-
	✓	63.2 (+2.8)	44.5 (+1.8)	-	-
LGE-SQL	✗	61.7	45.8	-	-
	✓	66.2 (+4.5)	**48.8** (+3.0)	-	-
T5-3B+PICARD	✗	65.5	44.3	67.2	52.1
	✓	67.0 (+1.5)	47.7 (+3.4)	66.7 (-0.5)	51.3 (-0.8)
RASAT	✗	65.0	45.7	69.9	50.7
	✓	**67.2** (+2.2)	48.5 (+2.8)	**71.5** (+1.6)	**53.3** (+2.6)

an ablation study of each component within our approach, as summarized in Table 5.

Experiment ②, ③ and ④ verify the effectiveness of the DAM module. Furthermore, experiment ⑤ and ⑥ shows that hybrid score update can effectively improve the effectiveness of curriculum learning, thereby improving the performance of the model. RASAT with the complete curriculum learning method task obtains the best performance, curriculum judeger can effectively prevent overfitting of the model.

In order to further verify the versatility of our method, we conducted experiments on our training framework on 4 methods, as shown in Table 6 and Table 7. We conducted experiments on the SparC and CoSQL data sets. The experimental results show that our training framework has

TABLE 7. SE-HCL with different baseline models on CoSQL.

Model	With SE-HCL	CoSQL QEM	IEM	QEX	IEX
RAT-SQL	✗	53.0	21.7	-	-
	✓	54.1 (+1.1)	21.0 (-0.7)	-	-
LGE-SQL	✗	55.7	21.9	-	-
	✓	56.3 (+0.6)	23.5 (+1.6)	-	-
T5-3B+PICARD	✗	56.5	24.5	62.1	35.5
	✓	56.9 (+0.4)	25.5 (+1.0)	65.9 (+3.8)	**37.8** (+2.3)
RASAT	✗	56.2	25.9	63.8	34.8
	✓	**57.2** (+1.0)	**28.1** (+2.2)	**66.3** (+2.5)	37.2 (+2.4)

TABLE 8. Case study on the SParC. RASAT+SE-HCL generates correct SQL in these cases while RASAT model fails.

Description Final Question	A database about dog kennels. **List the last name of the owner owning the youngest dog.**
Question #1 RASAT RASAT+SE-HCL	Which one is the oldest dog? List the information of it. **SELECT** * **FROM** Dogs **ORDER BY** age **ASC LIMIT** 1 **SELECT** * **FROM** Dogs **ORDER BY** age **ASC LIMIT** 1
Question #2 RASAT RASAT+SE-HCL	What are the last names of the oldest dogs? **SELECT** T1.last_name **FROM** Owners **AS** T1 **JOIN** Dogs **AS** T2 **ON** T1.owner_id = T2.owner_id **WHERE** T2.age = (**SELECT** max(age) **FROM** Dogs) **SELECT** T1.last_name **FROM** Owners **AS** T1 **JOIN** Dogs **AS** T2 **ON** T1.owner_id = T2.owner_id **WHERE** T2.age = (**SELECT** max(age) **FROM** Dogs)
Question #3 RASAT	What about the owner last name of the youngest dog? **SELECT** T1.last_name **FROM** Owners **AS** T1 **JOIN** Dogs **AS** T2 **ON** T1.owner_id = T2.owner_id **WHERE** T2.age = (**SELECT** max(age) **FROM** Dogs)
Description Final Question	A database about museum and visitor. **What are the id, name and membership level of visitors who have spent the largest amount of money in total in all museum tickets?**
Question #1 RASAT RASAT+SE-HCL	What is the total spent on all visits? **SELECT** sum(Total_spent) **FROM** visit **SELECT** sum(Total_spent) **FROM** visit
Question #2 RASAT RASAT+SE-HCL	Find the name of the visitor who has spent the most money for his or her visits. **SELECT** t1.name **FROM** visitor **AS** t1 **JOIN** visit **AS** t2 **ON** t1.id = t2.visitor_id **GROUP BY** t2.visitor_id **ORDER BY** t2.Total_spent **DESC LIMIT** 1 **SELECT** t1.name **FROM** visitor **AS** t1 **JOIN** visit **AS** t2 **ON** t1.id = t2.visitor_id **GROUP BY** t2.visitor_id **ORDER BY** sum(t2.Total_spent) **DESC LIMIT** 1
Question #3 RASAT RASAT + SE-HCL	What are his id and membership level? **SELECT** t1.name , t1.Level_of_membership **FROM** visitor **AS** t1 **JOIN** visit **AS** t2 **ON** t1.id = t2.visitor_id **GROUP BY** t2.visitor_id **ORDER BY** t2.Total_spent **DESC LIMIT** 1 **SELECT** t2.visitor_id , t1.name , t1.Level_of_membership **FROM** visitor **AS** t1 **JOIN** visit **AS** t2 **ON** t1.id = t2.visitor_id **GROUP BY** t2.visitor_id **ORDER BY** sum(t2.Total_spent) **DESC LIMIT** 1

good versatility and has achieved significant performance improvements on four different models.

B. CASE STUDY

In Table 8, we demonstrate the enhanced precision of SE-HCL in guiding the model to produce more accurate SQL structures. This is exemplified through two instances of question-SQL pairs extracted from the SParC dataset. We present a comparative analysis between the predictions generated by RASAT and RASAT+SE-HCL. In the first scenario, RASAT fails to consider the "youngest dog" condition when responding to Question #3. However, when augmented with SE-HCL, RASAT+SE-HCL accurately predicts this condition by distinguishing between the "oldest" information from the dialogue history and the "youngest" aspect within the current question.

In the second case, where the database schema is more intricate, the RASAT model fails to aggregate "Total_spent." Additionally, "visitor_id" is absent from the select clause. However, with RASAT integrated with SE-HCL, it correctly generates the select clause and sums "Total_spent" in the order clause.

VII. CONCLUSION

We propose SE-HCL, a novel text-to-SQL training framework that utilizes curriculum learning to better leverage structural information. We measure the difficulty of the data from both a structural and modeling perspective. We designed the data course module which first simplifies the data and then gradually increases the difficulty of the data. Furthermore, we propose a scoring module that judges the difficulty of a question. Finally, a curriculum judger is designed to make a decision whether to end the training based on model performance. Our experiments show that HCL effectively improves the performance of multi-turn text-to-SQL on SparC and CoSQL, especially for difficult and long-turn questions.

REFERENCES

[1] V. Zhong, C. Xiong, and R. Socher, "Seq2SQL: Generating structured queries from natural language using reinforcement learning," 2017, *arXiv:1709.00103*.
[2] T. Yu, R. Zhang, K. Yang, M. Yasunaga, D. Wang, Z. Li, J. Ma, I. Li, Q. Yao, S. Roman, Z. Zhang, and D. Radev, "Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-SQL task," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 1–11.
[3] B. Wang, R. Shin, X. Liu, O. Polozov, and M. Richardson, "RAT-SQL: Relation-aware schema encoding and linking for text-to-SQL parsers," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 7567–7578.
[4] X. V. Lin, R. Socher, and C. Xiong, "Bridging textual and tabular data for cross-domain text-to-SQL semantic parsing," in *Proc. Findings Assoc. Comput. Linguistics: EMNLP*, 2020, pp. 1–23.
[5] R. Cao, L. Chen, Z. Chen, Y. Zhao, S. Zhu, and K. Yu, "LGESQL: Line graph enhanced text-to-SQL model with mixed local and non-local relations," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang. Process.*, 2021, pp. 1–15.
[6] J. Li, B. Hui, R. Cheng, B. Qin, C. Ma, N. Huo, F. Huang, W. Du, L. Si, and Y. Li, "Graphix-t5: Mixing pre-trained transformers with graph-aware layers for text-to-SQL parsing," 2023, *arXiv:2301.07507*.
[7] J. Qi, J. Tang, Z. He, X. Wan, Y. Cheng, C. Zhou, X. Wang, Q. Zhang, and Z. Lin, "RASAT: Integrating relational structures into pretrained Seq2Seq model for text-to-SQL," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2022, pp. 3215–3229.
[8] Y. Zheng, H. Wang, B. Dong, X. Wang, and C. Li, "HIE-SQL: History information enhanced network for context-dependent text-to-SQL semantic parsing," in *Proc. Findings Assoc. Comput. Linguistics, ACL*, 2022, pp. 1–11.
[9] T. Yu, R. Zhang, A. Polozov, C. Meek, and A. H. Awadallah, "Score: Pre-training for context representation in conversational semantic parsing," in *Proc. ICLR*, 2021, pp. 1–16.
[10] Z. Cai, X. Li, B. Hui, M. Yang, B. Li, B. Li, Z. Cao, W. Li, F. Huang, L. Si, and Y. Li, "STAR: SQL guided pre-training for context-dependent text-to-SQL parsing," 2022, *arXiv:2210.11888*.
[11] B. Hui, R. Geng, Q. Ren, B. Li, Y. Li, J. Sun, F. Huang, L. Si, P. Zhu, and X. Zhu, "Dynamic hybrid relation exploration network for cross-domain context-dependent semantic parsing," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 13116–13124.
[12] T. Scholak, N. Schucher, and D. Bahdanau, "PICARD: Parsing incrementally for constrained auto-regressive decoding from language models," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2021, pp. 1–7.
[13] Y. Li, H. Zhang, Y. Li, S. Wang, W. Wu, and Y. Zhang, "Pay more attention to history: A context modelling strategy for conversational text-to-SQL," 2021, *arXiv:2112.08735*.
[14] Y. Cai and X. Wan, "IGSQL: Database schema interaction graph based neural model for context-dependent text-to-SQL generation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 1–10.
[15] R.-Z. Wang, Z.-H. Ling, J. Zhou, and Y. Hu, "Tracking interaction states for multi-turn text-to-sql semantic parsing," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 13979–13987.
[16] R. Zhang, T. Yu, H. Er, S. Shim, E. Xue, X. V. Lin, T. Shi, C. Xiong, R. Socher, and D. Radev, "Editing-based SQL query generation for cross-domain context-dependent questions," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 1–12.
[17] T. Yu, R. Zhang, M. Yasunaga, Y. C. Tan, X. V. 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, "SParC: Cross-domain semantic parsing in context," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 1–13.
[18] T. Yu et al., "CoSQL: A conversational text-to-SQL challenge towards cross-domain natural language interfaces to databases," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 1–18.
[19] P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2018, pp. 1–5.
[20] T. Scholak, R. Li, D. Bahdanau, H. de Vries, and C. Pal, "DuoRAT: Towards simpler text-to-SQL models," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2021, pp. 1–9.
[21] T. Yu, C.-S. Wu, X. V. Lin, B. Wang, Y. C. Tan, X. Yang, D. Radev, R. Socher, and C. Xiong, "GraPPa: Grammar-augmented pre-training for table semantic parsing," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–14.
[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.* Minneapolis, MI, USA: Association for Computational Linguistics, vol. 1, Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423
[23] J. Yang, S. Ma, D. Zhang, S. Wu, Z. Li, and M. Zhou, "Alternating language modeling for cross-lingual pre-training," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 1–8.
[24] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *J. Mach. Learn. Res.*, vol. 21, no. 140, pp. 1–67, 2020.
[25] J. Yang, S. Ma, L. Dong, S. Huang, H. Huang, Y. Yin, D. Zhang, L. Yang, F. Wei, and Z. Li, "GanLM: Encoder–decoder pre-training with an auxiliary discriminator," in *Proc. 61st Annu. Meeting Assoc. Comput. Linguistics*, Jul. 2023, pp. 9394–9412.
[26] Z. Cai, X. Li, B. Hui, M. Yang, B. Li, B. Li, Z. Cao, W. Li, F. Huang, L. Si, and Y. Li, "STAR: SQL guided pre-training for context-dependent text-to-SQL parsing," in *Proc. Findings Assoc. Comput. Linguistics, (EMNLP)*. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022, pp. 1235–1247.
[27] Z. Chen, L. Chen, H. Li, R. Cao, D. Ma, M. Wu, and K. Yu, "Decoupled dialogue modeling and semantic parsing for multi-turn text-to-SQL," in *Proc. Findings Assoc. Comput. Linguistics*, 2021, pp. 1–12.
[28] L. Chai, D. Xiao, Z. Yan, J. Yang, L. Yang, Q.-W. Zhang, Y. Cao, and Z. Li, "QURG: Question rewriting guided context-dependent text-to-SQL semantic parsing," in *PRICAI 2023: Trends in Artificial Intelligence*. Cham, Switzerland: Springer, 2023, pp. 275–286.
[29] J. L. Elman, "Learning and development in neural networks: The importance of starting small," *Cognition*, vol. 48, no. 1, pp. 71–99, Jul. 1993.
[30] D. L. T. Rohde and D. C. Plaut, "Language acquisition in the absence of explicit negative evidence: How important is starting small?" *Cognition*, vol. 72, no. 1, pp. 67–109, Aug. 1999.
[31] K. A. Krueger and P. Dayan, "Flexible shaping: How learning in small steps helps," *Cognition*, vol. 110, no. 3, pp. 380–394, Mar. 2009.
[32] G. Kumar, G. Foster, C. Cherry, and M. Krikun, "Reinforcement learning based curriculum optimization for neural machine translation," in *Proc. Conf. North*, 2019, pp. 2054–2061.
[33] B. Xu, L. Zhang, Z. Mao, Q. Wang, H. Xie, and Y. Zhang, "Curriculum learning for natural language understanding," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 6095–6104.
[34] Y. Huang and J. Du, "Self-attention enhanced CNNs and collaborative curriculum learning for distantly supervised relation extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 389–398.

[35] Y. Tay, S. Wang, A. T. Luu, J. Fu, M. C. Phan, X. Yuan, J. Rao, S. C. Hui, and A. Zhang, "Simple and effective curriculum pointer-generator networks for reading comprehension over long narratives," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 4922–4931.

[36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. U. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

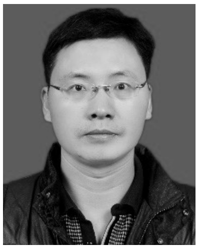
[37] T. Xie et al., "UnifiedSKG: Unifying and multi-tasking structured knowledge grounding with text-to-text language models," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022, pp. 602–631.

[38] V. Zhong, M. Lewis, S. I. Wang, and L. Zettlemoyer, "Grounded adaptation for zero-shot executable semantic parsing," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 6869–6882.

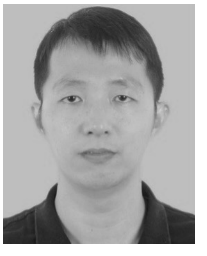
[39] X. Wang, S. Wu, L. Shou, and K. Chen, "An interactive NL2SQL approach with reuse strategy," in *Proc. Int. Conf. Database Syst. for Adv. Appl.*, 2021, pp. 280–288.

[40] Q. Liu, B. Chen, J. Guo, J.-G. Lou, B. Zhou, and D. Zhang, "How far are we from effective context modeling? an exploratory study on semantic parsing in context," in *Proc. 29th Int. Conf. Int. Joint Artif. Intell.*, 2020, pp. 3580–3586.

YIYUN ZHANG received the B.S. degree in computer and communication from Lanzhou University of Technology. She is currently a Lecturer with the Institute of Electronic Information, Guangdong Vocational College. Her current research interests include artificial intelligence and digital media technology.



SHENG'AN ZHOU received the B.S. and M.S. degrees in computer science from South China University of Technology. He is currently a Professor with the Institute of Electronic Information, Guangdong Vocational College. His current research interests include artificial intelligence and higher vocational education.



GENGSHENG HUANG received the B.S. and M.S. degrees in computer science from Beijing University of Posts and Telecommunications. He is currently an Associate Professor with the Institute of Electronic Information, Guangdong Vocational College. His research interests include artificial intelligence and vocational education.