



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



SQL Query Generation from Natural Language Using NLP

Abhishek Kharade, Parth Kasurde, Samadhan Sable, Shubham Karadkar

Department of Computer Science, Sinhgad Institute of Technology and Science, Pune, India

ABSTRACT: Most of existing studies on parsing natural language (NL) for constructing structured query language (SQL) do not consider the complex structure of database schema and the gap between NL and SQL query. In this paper, we propose a schema-aware neural network with decomposing architecture, namely HSRNet, which aims to address the complex and cross-domain Text-to-SQL generation task. The HSRNet models the relationship of the database schema with a hierarchical schema graph and employs a graph network to encode the information into sentence representation. Instead of end-to-end generation, the HSRNet decomposes the generation process into three phases. Given an input question and schema, we first choose the column candidates and generate the sketch grammar of the SQL query. Then, a detail completion module fills the details based on the column candidates and the corresponding sketch.

KEYWORDS: Semantic parsing, SQL generation, deep learning, neural network, graph encoder, natural language process

I. INTRODUCTION

An extremely flexible and functional database query language, SQL is used to access and manage database systems. Data sets have evolved gradually from one domain to the opposite domain and from one table mode to several table modes. The selection of tables is required for SQL generation in the multi-table mode. The data set is additionally split into simple problem and complex problem modes based on the problem's complexity. The number of keywords, nesting depths, and the number of phrases included in SQL query statements all affect how complicated the data set is. Currently, most data sets require one round to produce finished SQL statements, which has However, for non-technical users, there are greater challenges in learning SQL.

II. MOTIVATION

To enable users query huge amounts of data from the database, text-to-SQL technology will translate problems expressed in natural language into appropriate SQL commands. This paper discusses the building of a text to SQL data set, the analysis of several job model technologies, and the technology-based application areas. For large-scale annotation data sets like wiki sql, spider, and others, it is discovered that the existing model still has a lot of space for development, but there is limited study on Chinese text to SQL technology.



III. LITERATURE REVIEW

Sr no	Project title	Authors	Year	Advantages	Disadvantages
1	A Comprehensive Exploration on Spider with Fuzzy Decision Text-to-SQL Model	Qing Li, Lili Li, Qi Li, and Jiang Zhong	2020	The problem of the complex and cross-domain text-to-SQL generation task.	To improve the model choice, we utilised the attention mechanism and a fuzzy decision mechanism to successfully overcome the order problem in the conventional model.
2	A Multiple-Integration Encoder for Multi-Turn Text-to-SQL Semantic Parsing	Run-Ze Wang, Zhen-Hua Ling, Jing-Bo Zhou, and Yu Hu	2021	The SPaRC dataset demonstrates that our solution outperformed the most sophisticated benchmarks in terms of multi-turn text-to-SQL generation accuracy.	a module for improving utterance representation is constructed to include historical utterance information into the representation of each token in a given speech by careful selection. Second, a module for improving schema discrepancies is created to incorporate previously predicted SQL queries into the representation of schema items.
3	Automatic Text-to-SQL Machine Translation for Scholarly Publication Database Search	SulochanaDeshmukh, Marwan Bikdash	2020	It is a domain-specific Text-to-SQL converter that is being suggested. As a result, the range of named entities is constrained to	To validate NE, NEA, and QI, the text cosine similarity measure is employed. The rule-based technique is computationally



				database values. The database itself serves as our knowledge base as a result.	efficient, provides accurate translation, and does not necessitate extensive training on the dataset.
4	GuideSQL: Utilizing Tables to Guide the Prediction of Columns for Text-to-SQL Generation	Huajie Wang, Lei Chen, Mei Li, Mengnan Chen	2020	We employ past query attention to obtain context-dependent SQL query information on the difficult text-to-SQL benchmark SParC.	In order to prevent mistakes brought on by incorrect predictions of table-column dependencies, we describe a unique solution dubbed GuideSQL that predicts tables first and utilises a pruning mechanism to remove the columns that don't belong to the anticipated tables.
5	Hierarchical Schema Representation for Text-to-SQL Parsing with Decomposing Decoding	MEINA SONG , ZECHENG ZHAN , AND HAIHONG E	2018	HSRNet, a schema-aware neural network with a deconstructing architecture, intends to handle the difficult, cross-domain Text-to-SQL generation task.	The data into several baselines, we show the value of our hierarchical schema representation. We also demonstrate how the decomposing architecture greatly enhances the functionality of our model.



Open Issues:

With the rapid advancement of computer technology, information from numerous industries in real life, such as sales information from various businesses, school selection information, grade information, and firm accounting information, is being saved in an increasing number of databases. SQL is a database query language that provides tremendous flexibility and sophisticated functionality for accessing and operating database systems.

IV.CONCLUSION

The work of converting text to SQL is currently being employed more and more in the realm of practical application, which has caught the interest of academia. The model's decoding techniques, which include intermediate representation, tree decoding, graph network modelling, etc., have produced some results, but the model's impact on complicated data is not immediately apparent. At the same time, we should take into account the standardisation of database tables, the utilisation of outside information, and the incorporation of progressive discussion in practical implementation.

REFERENCES

- 1) Qing Li , Member, IEEE, Lili Li , Qi Li, and Jiang Zhong “A Comprehensive Exploration on Spider with Fuzzy Decision Text-to-SQL Model” Proces. 2020
- 2) Run-Ze Wang , Zhen-Hua Ling , Senior Member, IEEE, Jing-Bo Zhou, and Yu Hu “A Multiple-Integration Encoder for Multi-Turn Text-to-SQL Semantic Parsing”, VOL. 29, 2021
- 3) SulochanaDeshmukh, MarwanBikdash “Automatic Text-to-SQL Machine Translation for Scholarly Publication Database Search”, Greensboro, NC. 27411
- 4) Huajie Wang , Lei Chen†, Mei Li , Mengnan Chen “GuideSQL: Utilizing Tables to Guide the Prediction of Columns for Text-to-SQL Generation”
- 5) MEINA SONG1 , (Member, IEEE), ZECHENG ZHAN1 , AND HAIHONG E. “Hierarchical Schema Representation for Text-to-SQL Parsing with Decomposing Decoding ” 10.1109/ACCESS.2017.DOI



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details