IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105315 |

SQL Query Generation from Natural Language Using NLP

Abhishek Kharade, Parth Kasurde, Samadhan Sable, Shubham Karadkar, Prof. Rupali Waghmode

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

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 [1].

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[2].

To validate NE, NEA, and QI, the text cosine similarity measure is employed. The rule-based technique is computationally efficient, provides accurate translation, and does not necessitate extensive training on the dataset[3].

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[4].

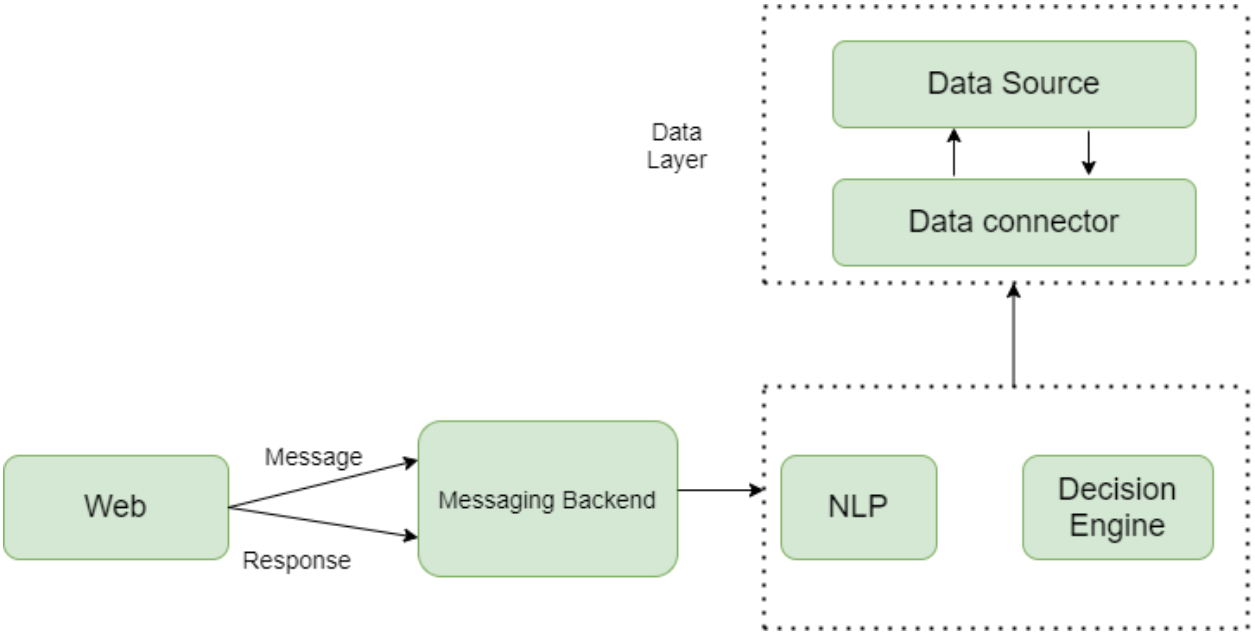
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105315 |

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[5].

IV. PROPOSED SYSTEM



Accept the message from the user.
Now, perform the following to the accepted message:
• Firstly, check each word whether it is spelled correctly or not. If not, perform operation based upon selected option. – If Yes is selected, perform the following operation using the suggested string. – Otherwise, perform the operations using the user's previous string.
• Split the message into words.
• Execute an SQL using Regular Expression to check the words are available in the database.
• Store the words present in the database into an array called "important words".
• Execute SQL query using the above array words.
• If the result of above step produces single row then display the answer to the user.
• If multiple rows are produced, then display options to the user with the help of "title" column in the table.
• If no result is produce, then check all the keywords in "Answer" columns. Do the following: – If a match is found, store the user's question and found answer into another table. So, if same question is asked, the system can provide answer. – Otherwise, display sorry message to user.
• Based on options selected or another message entered, Go to step 2.

V. RESULT AND DISCUSSION

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and Jdk 1.8. The application is web application used tool for design code in Eclipse and execute on Tomcat server.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105315 |

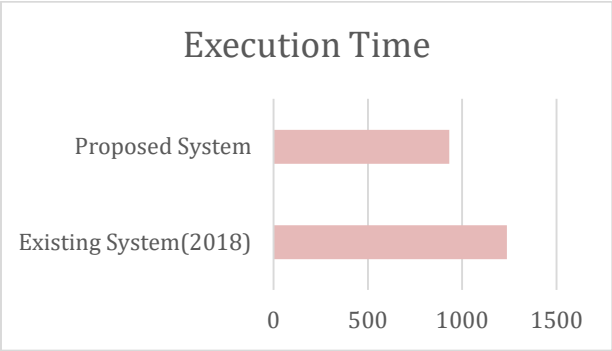


Figure 2: overall system execution graph

Existing System (2018)	Proposed System
1236	932

Table 1: overall system execution table

VI. 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: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details