pakdd View Reviews

Paper ID
423

Paper Title
Rel2Graph: Automated Mapping From Relational Databases to a Unified Property Knowledge Graph

Track Name
Research Track

Reviewer #1

Questions

1. Brief summary of the paper
This paper proposes Rel2Graph, a novel approach for Knowledge Graph Construction from Relational Databases. In particular, it supports mapping SQL queries into Cypher queries. The approach is evaluated on 2 datasets.

2. List three, or more, strong aspects of this paper. Please number each point.
+ The topic is interesting, worth investigating, and relevant for the conference audience.

+ The results of the empirical evaluation seem promising.

+ Data repair is also addressed.

3. List three, or more, weak aspects of this paper. Please number each point.
- No motivation to compute Cypher queries raher than other queries over knowledge graphs.

- No empirical comparison with other previous approaches.

- Related work is incomplete.

- The full translation between SQL and Cypher is not included.

4. Detailed comments to the authors.
This paper proposes Rel2Graph, a novel approach for Knowledge Graph Construction from Relational Databases. In particular, it supports mapping SQL queries into Cypher queries. The approach is evaluated on 2 datasets.

The topic is interesting, worth investigating, and relevant for the conference audience. The results of the empirical evaluation seem promising. Data repair is also addressed, which is an interesting point.

However, there are several weaknesses. Firstly, there is no motivation to compute Cypher queries raher than other queries over knowledge graphs.

Furthermore, there is no empirical comparison with other previous approaches.

Related work is also incomplete, and some recent (2023) approaches such as https://doi.org/10.1201/9781003313267-8 and https://doi.org/10.1007/978-981-99-7019-3_42 are not considered.

From a technical point of view, the full translation between SQL and Cypher is not included. It is not even crystal clear if the system is complete, i.e,. if it covers full SQL expressivity.

Minor comment: Rel2Graph does not use a special calligraphical font in the abstract.

5. Overall Recommendation.
Weak Reject

Reviewer #2

Questions

1. Brief summary of the paper
The paper proposes a migration strategy from relational data to graphs. This allows to leverage the schema-level information available in relational data in semantic-level tasks.

2. List three, or more, strong aspects of this paper. Please number each point.
- The problems of database and query mapping are particularly relevant to solve several downstream tasks.
- The proposed methodology is sound and well explained.
- The consistency accuracies are quite promising.

3. List three, or more, weak aspects of this paper. Please number each point.
- The empirical evaluation is quite limited as focused on a particular scenario, based on accuracy only.
- The results fail to provide readers with comprehensive and exhaustive analysis of the results. The concluding remarks "enhance the inherent structure of relational data through graph representation, resulting in optimized query operations when addressing semantic-level tasks, e.g. semantic parsing" seem to be not adequately supported by empirical evidence.
- The related work section needs improvements. It fails to clearly specifiy the position of the work in the existing literature as some of the cited papers are weakly related to the submitted work.

4. Detailed comments to the authors.
The paper addresses an interesting research problem, i.e., the migration of data from relational schema to knowledge graph. Solving the task is instrumental for effectively and efficiently addressing several downstream tasks such as semantic parsing.

I appreciated the explanation of the motivation examples and the the formalization of the problem. Conversely, I have some concerns on the empirical evaluation which seems to be not broad enough to support the introductory and concluding statements of the authors. Some comments in the

above sections are a little bit overestimating what really comes out from the experiments. I see results achieved in particular cases, with a limited set of datasets and metrics, without sufficient justifications.

I also suggest the revise the related work section. It is not well to-the-point. It includes weakly related citations and fails to highlight pros and cons of the proposal compared to existing approaches.

5. Overall Recommendation.
Weak Reject

Reviewer #3

Questions

1. Brief summary of the paper
The paper tackles the lack of systematic evaluations in converting relational databases to graphs and proposes Rel2Graph, an automatic knowledge graph construction (KGC) approach.

2. List three, or more, strong aspects of this paper. Please number each point.
1. The approach facilitates the conversion of conjunctive SQL queries into pattern-based NoSQL queries, highlighting the critical role of query mapping.
2. The evaluation is conducted on Spider and KaggleDBQA benchmarks for semantic parsing, utilizing the execution accuracy (EA) metric to quantify the results.

3. List three, or more, weak aspects of this paper. Please number each point.
1. It lacks a comparative analysis with existing approaches.
2.The paper uses a simple "Repair" scenario in order to check mapping consistency, that might not cover all complex database transformation challenges.

4. Detailed comments to the authors.
This paper presents an automated approach for mapping relational databases into knowledge graph. It is capable of identifying different SQL clauses and mapping them into pattern-based queries. Experiments use Evaluation Accuracy to study the system performance under two databases of different number of SQL queries.

The studied problem is well-established problem and there should exist numerous relevant literatures. The paper lacks a comparative analysis with existing approaches.

5. Overall Recommendation.
Weak Reject

Reviewer #4

Questions

1. Brief summary of the paper
The paper explores the domain of automated Knowledge Graph Construction (KGC), a research field focusing on representing knowledge from diverse data source in a unified manner to facilitate efficient quering. The primary contributions of the paper are twofold:

1. Rel2Graph Creation: The authors introduce the concept of Rel2Graph, a unified property knowledge graph derived from relational databases. In this, rows from tables identified as entity tables are transformed into graph nodes and rows from tables identified as linking tables are transformed into graph edges. Apart from this an edge is constructed between two entity tables with format “_HAS_”. Apart from this, authors addresses various challenges during KG construction, including issues like "No primary key" or "No/Incorrect Foreign Key."

2. Translation of SQL Queries to Cypher Queries: The paper presents a methodology for translating conjunctive SQL queries into Cypher queries. The process involves parsing the SQL query into a JSON-izable parse tree, followed by pattern-based token matching. The authors handle sub-queries and group queries by utilizing the result as the starting point of the parent query.

The evaluation of the constructed Knowledge Graph and Cypher queries was performed on Spider and KaggleDBQA datasets, using execution accuracy (EA) as the metric for quantifying the results.

2. List three, or more, strong aspects of this paper. Please number each point.
S1. The concept of construction of Property Knowledge Graph by finding entity tables and linking tables.
S2. The authors demonstrate a comprehensive understanding of errors that may arise during Knowledge Graph construction and have implemented a data repair pipeline to rectify such issues.
S3. The use of Spider and KaggleDBQA is quite interesting and useful.

3. List three, or more, weak aspects of this paper. Please number each point.
W1. Algorithm 1 pseudocode (Line 13-17) does not align with the code explanation, specifically concerning Line 13, where the code iterates over each foreign key in the table, contrary to the mentioned condition of "neither of the two arguments being met ..."
W2. Clarifications are needed for Lines 10, 11, 13, and 14 in Algorithm 2's pseudocode, outlining how nested sub-queries are handled during the translation from SQL to Cypher.
W3. How execution accuracy (EA) is a good metric to quantify the KG quality. (especially considering potential variations in results with the addition of more data to the KG)
W4. A table stating which types of queries are prone to faliure would enhance the comprehensiveness of the paper.
W5. Missing important citations for some of the initial work like
(a) Roberto De Virgilio, Antonio Maccioni, and Riccardo Torlone. 2013. Converting Relational to Graph Databases. In Proc. Of GRADES (colocated with SIGMOD), (b) Roberto De Virgilio, Antonio Maccioni, and Riccardo Torlone. 2014. R2G: a Tool for Migrating Relations to Graphs. In Proc. Of EDBT. (c) Konstantinos Xirogiannopoulos, Virinchi Srinivas, and Amol Deshpande. 2017.

GraphGen: Adaptive Graph Processing using Relational Databases. In Proceedings of the Fifth International Workshop on Graph Data-management Experiences & Systems (GRADES'17). etc.
W5. The overall approach appears to lack the novelty required to address the presented problem statement.

4. Detailed comments to the authors.
Typos:
• Page 1 – Abstract – pattern-based NoSQL queries -> pattern-based Cypher queries
• Page 2 – Section 2 – KCG from Text Data -> KGC from text data
• Page 3 – Section 3 – consisting of relation databases -> consisting of relational databases
• Page 4 – Algorithm 1 – There is an indentation issue. All the lines under RelDB2GraphBuilder should have one tab indentation.
• Page 5 – section 3.1 – This particular scenario is elaborated upon from line 6 to line 9 in Algorithm 1 -> This particular scenario is elaborated upon from line 4 to line 7 in Algorithm 1
• Page 5 – section 3.1 – pseudocode outlined from line 10 to line 13 -> pseudocode outlined from line 8 to line 11
• Page 5 – Fig.1 -
• Page 8 – section 3.2 – illustrated next to Algorithm 1 -> illustrated next to Algorithm 2
• Page 8 – section 3.2 – We allign FROM with MATCH -> We align FROM with MATCH
• Page 8 – section 3.2 – Line 9 in Algorithm 1 -> Line 9 in Algorithm 2
• Page 10 – section 4.3 – including the number of domains (#DB), the count -> (In table 3 (#DB) column is missing.)

5. Overall Recommendation.
Reject

Reviewer #5

Questions

1. Brief summary of the paper
The paper introduces Rel2Graph, an approach to automatically construct a property knowledge graph from relational databases, facilitating efficient querying and supporting the mapping of SQL queries to NoSQL queries. Rel2Graph is evaluated using the Spider and KaggleDBQA benchmarks for semantic parsing, employing the execution accuracy (EA) metric to quantify results. The approach includes a systematic process for database mapping and query translation, ensuring data consistency at the semantic query level. The constructed knowledge graph aims to optimize query operations for semantic-level tasks such as semantic parsing, natural language interface, and question answering.

2. List three, or more, strong aspects of this paper. Please number each point.
S1: The paper introduces Rel2Graph, an automatic knowledge graph construction method from relational databases, which also supports mapping SQL queries to NoSQL queries2.

S2: It evaluates the approach using two widely used benchmarks, Spider and KaggleDBQA, and employs the execution accuracy metric for quantifying results.

S3: The proposed method facilitates efficient querying and could significantly benefit downstream NLP tasks like semantic parsing and question answering.

3. List three, or more, weak aspects of this paper. Please number each point.
W1: While the paper discusses the evaluation of the proposed approach on two datasets, it may not be sufficient to establish the generalizability and robustness of the method across a wider variety of databases.

W2: The paper mentions data repair during the database mapping process but does not provide detailed solutions for handling complex data inconsistencies, which could be critical for the practical application of the approach.

W3: The paper introduces an automated approach, but it does not discuss the complexity of the implementation or its usability for non-experts, which are important factors for practical adoption.

W4: The paper uses Execution Accuracy (EA) as the sole metric for evaluation. Including additional performance metrics could provide a more comprehensive assessment.

4. Detailed comments to the authors.
Innovative Approach: The Rel2Graph method presents an innovative approach to Knowledge Graph Construction (KGC) from relational databases2. It is commendable that the paper addresses the challenge of query mapping by supporting the translation of conjunctive SQL queries into pattern-based NoSQL queries3. This could significantly contribute to the field by enabling more efficient querying and data management.

Evaluation Metrics: The use of the Execution Accuracy (EA) metric to quantify the results of the proposed approach is a sound choice4. However, it would be beneficial to include additional metrics that could provide a more comprehensive evaluation of the data consistency and the quality of the query mappings, such as precision, recall, and F1-score.

Future Work: The paper could expand on the potential applications of the Rel2Graph approach in real-world scenarios. Discussing how this method could be integrated into existing systems and the possible challenges that might arise during such integrations would provide valuable insights for practitioners in the field. Additionally, exploring the scalability of the approach and its performance on larger datasets would be a worthwhile direction for future research.

5. Overall Recommendation.
Weak Accept