



# **A Survey on Mitigating Relation Completion Problem in Database Application**

Suwarna Sanap, Amruta Ratnaparkhi, Madhuri Gaikwad, Prashant Kolekar

B. E Students, Department of Computer, GHRCEM, Savitribai Phule Pune University, Pune, India

**ABSTRACT:** In this project, we identify relation completion i.e. Relation Completion as one recurring problem central to the success of novel big data applications such as Entity Reconstruction and Data Affluence. Given relation  $R$ , relation completion occurs at linking entity pairs between two entity set under the relation  $R$ . In Pattern based search, for example, may produce results that are unspecific or ambiguous when taken out of context. To accomplish the RC goals, we propose to formulate the search queries for each entity  $x$  based on supernumerary information, so that detect its target entity  $y$  from the set retrieved documents. As an alternative, we propose the CoRE method because high quality patterns may decrease the likelihood of suitable target entities. CoRE method uses context terms learned ambient the expression of a relation supernumerary information in formulating queries. Our experimental study on large scale real world database indicates that context aware relation extraction method significantly predicts better performance than the pattern Based search method.

**KEYWORDS:** Context-aware relation extraction, pattern based search, relation completion, relation query expansion, big Data, rel term.

## **I. INTRODUCTION**

Now a days big data is giving development of new generation of application that attempts RC problems at linking related data from different sources. This data is typically unstructured and faults of any binding information (i.e. foreign keys). The Current data integration systems capabilities go beyond to link this big data. Relation completion(RC) as one of the recurring problem was occur where there are some information extraction task such as relation extraction, recognition of named entity which have been used to enable emerging application such as data enrichment and entity reconstruction. Pattern aware relation extraction (PaRE) method used pattern based search which uses extracted patterns as the auxiliary information in formulating search queries. In this paper relation completion task is the main focus of the work is presented which is occurred in Pattern aware relation extraction to overcome it we are using the context aware relation extraction( CoRE) method. In this research paper, author proposes that in big data application when we are accessing some data by entering small subset of query but there is occurring a Relation Completion (i.e. RC) problem in PaRE method. As an alternative, CoRE proposes a Context Aware Relation Extraction method to accomplish this Relation Completion(RC) problem [1]. CoRE proposes a Context Aware Relation Extraction method to extract the information by entering the small subset of query, Context-Aware Relation Extraction (CoRE) method which uses context terms learned surroundings the expression of a relation as the auxiliary information in formulating queries. The experimental results on several real-world web data application demonstrates that CoRE gives much higher accuracy than Pattern Aware Relation Extraction that is PaRE for the purpose of RC [1].

## **II. LITERATURE SURVEY**

Zhixu Li, Mohamed A. Sharaf, Laurianne Sitbon, Xiaoyong Du and Xiafang Zhou. "CoRE: Context Aware Relation Aware Extraction method for relation completion." In this research paper, author proposes that in big data application when we are accessing some data by entering small subset of query but there is occurring a Relation Completion (i.e. RC) problem in PaRE method. As an alternative, CoRE proposes a Context Aware Relation Extraction method to accomplish this Relation Completion(RC) problem. PaRE method uses high quality patterns and that patterns may decrease the probability of finding the suitable target entities, so that it detects its target entity from the retrieved documents. For that a pattern-based method uses extracted patterns as auxiliary information in formulating search queries. It proposes to formulate search queries for each query based on some auxiliary information that is what



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

is about we want to search (eg. Bob Brown joined in which university and faculty) ,for a given relation R, RC attempts at linking entity pairs between two entity (for eg. 'Bob Brown' has joined 'faculty' ) in this example Bob Brown and Faculty are two different entities and relation between that two entities is has 'joined' where.

CoRE proposes a Context Aware Relation Extraction method to extract the information by entering the small subset of query, Context-Aware Relation Extraction (CoRE) method which uses context terms learned surroundings the expression of a relation as the auxiliary information in formulating queries. The experimental results on several real-world web data application demonstrates that CoRE gives much higher accuracy than Pattern Aware Relation Extraction that is PaRE for the purpose of RC. Bhor Priyanka, Deolekar Shweta, Gaikwad Revati and Gonjare Shraddha. "Mitigating relation completion problem in big data application." In this research paper, in the proposed system three participants are involved to provide secure data storage and these participants are data multiple Rel queries are posed, each of which is based on the query entity  $x$  which we have entered to search in conjunction with the one of the pattern extracted by the Pattern Aware Relation Extraction (i.e. PaRE) method (e.g. Naina joined +in), (Naina Shrivastav works at), etc). Using pattern as auxiliary information will generate very strict Rel Queries and high quality patterns, which will return the least number of links, but most of which are RelDocs. Hence, if a query is used before in a webpage under the one of the Patterns, it will quickly matched with its correct target result entity, however, such assumptions is un-realistic for many query entities that used in very few web pages (i.e. long tail). For those entities, no web pages will be returned and will remain unmatched, this above formulation is orthogonal to the pattern-based, where Multiple Rel Queries are posed, each of which based on the query entity  $x$  and an target entity  $y$  from the target list. Hence, each of the retrieved documents is further processed to detect that any of the patterns extracted by the PaRE method to justify whether  $(x,y) \in R$ .

Priya Pujari, Arti Waghmare, "Improved search result of keyword query using data imputation approach", In this research paper, Keyword query is used to search any web documents, queries offers simple access to data over databases, but this frequently suffer from low quality of ranking of data. It would be beneficial to categorize and separate queries that are likely to have the low ranking quality of data to improve the user satisfaction and confidence. For example, the system may recommend to the user alternate and simple queries for such difficult queries. Imputation is the process of supplementing missing information with the substituted values. Numerous current modern and exploration of large data sets contain missing values. Investigating the information and inclination coming about because of difference between missing and complete data. Issues connected with missing values are loss of productivity, entanglements in taking care of missing and complete data. Recently, some retrieved based methods are proposed to retrieve missing values from outer resource such as World Wide Web (WWW) but fails to respond to large no. Of search queries. The approach is built called interactive Inferring-Retrieving data Imputation approach (TRIP approach), which is the interaction between the retrieving based methods and inferring based methods.

Fabio Petroni, Luciano Del Corp and Rainer Gemulla, "CoRE: Context Aware Relation Extraction method with Factorization machines". In this paper author proposed CORE, a matrix factorization model for open RE that incorporates contextual information and their model is based on factorization machines and the open-world assumption, integrates various forms of contextual information, and is extensible. The open RE task is also related to open information extraction (open IE) which extracts large amounts of surface relations and their arguments from natural language text; e.g., "criticizes" ("Kevin", "eye care hospital"). Targeted IE methods aim to extract from natural-language text new instances of a set of predefined relations, taken from a KB. Most existing methods make use of distant supervision, i.e., they start with a set of instances (i.e. pairs of entities) for the relations of interest, search for these instances in text, learn a relation extractor from the so so-obtained training data. Their experimental study on a large and real-world dataset indicates that CoRE has significantly better prediction performance than state-of-the-art approaches when contextual information is available. Their experimental study suggests that exploiting context can significantly improve prediction performance.

### III. PROPOSED SYSTEM

In this project, CoRE proposes, a Context-Aware Relation Extraction Method, which is particularly designed for relation completion (RC) task. This project proposes an integrated model to learn high-quality relation-context terms for CoRE. This model incorporates and expands methods that are based on positional proximity, terms frequency and discrimination information. This project proposes a tree based query formulation method, which selects a small subset of search queries to be issued as well as schedules the order of issuing queries. This project uses a confidence-aware method that be schedules as well issued as the order of issuing queries. This project propose a confidence-aware

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

method that estimates the confidence that a candidate target entity is the correct one. This enables CoRE to reduce the number of issued search queries by terminating the search whenever it extracts a high-confidence target entity. CoRE proposes a Context Aware Relation Extraction method to accomplish this Relation Completion(RC) problem. PaRE method uses high quality patterns and that patterns may decrease the probability of finding the suitable target entities, so that it detects its target entity from the retrieved documents. For that a pattern-based method uses extracted patterns as auxiliary information in formulating search queries. It proposes to formulate search queries for each query based on some auxiliary information that is what is about we want to search (eg. Bob Brown joined in which university and faculty) ,for a given relation R, RC attempts at linking entity pairs between two entity(for eg. 'Bob Brown' has joined 'faculty' ) in this example Bob Brown and Faculty are two different entities and relation between that two entities is has 'joined' where. CoRE proposes a Context Aware Relation Extraction method to extract the information by entering the small subset of query, Context-Aware Relation Extraction (CoRE) method which uses context terms learned surroundings the expression of a relation as the auxiliary information in formulating queries [1].

## A. ARCHITECTURE OF CoRE SYSTEM:

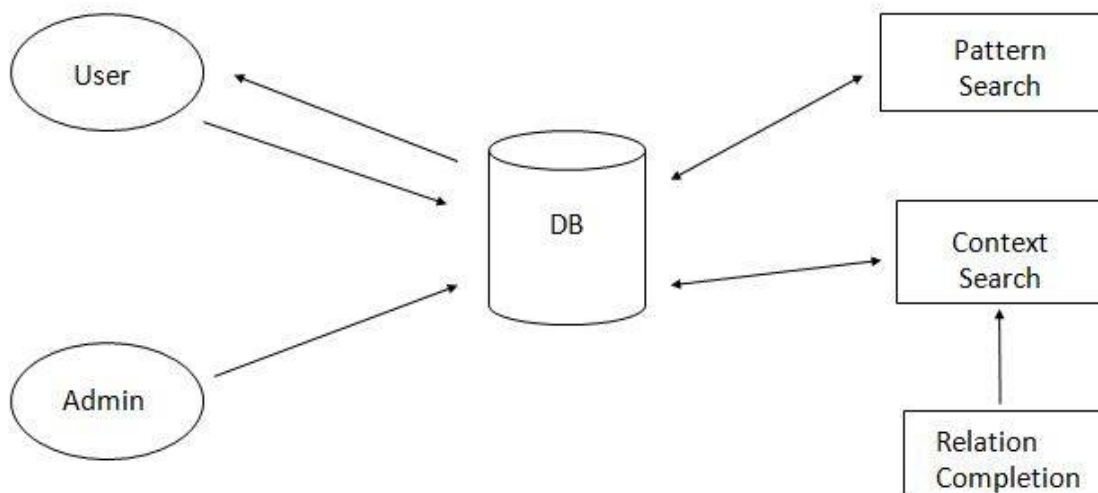


Fig 1: Architecture of CoRE system

Relation Completion is fastly becoming one of the fundamental task in many of emerging application that is we are using now a days. In above diagram some techniques are used which are given below:

### a) Pattern Based Search:

Multiple RelQueries are posed, each of which is based on the query entity a in conjunction with one of the patterns extracted by the PaRE method (e.g., ("Bob Brown joined" ∪ "in"), ("Bob Brown works at"), etc.). Using patterns as auxiliary information will generate very strict RelQueries, which will return the least number of web documents, but most of which are RelDocs. Hence, if a query entity a happened to appear in a webpage under one of the used patterns, it will be quickly matched with its correct target entity. However, such assumption is unrealistic for many query entities that appear in very few webpages (i.e., long tail). For those entities, no webpages will be returned and will remain unmatched. This formulation is orthogonal to the pattern-based one above, where Multiple RelQueries are posed, each of which is based on the query entity a and an entity bc from the target list. Hence, each of the retrieved documents is processed to detect any of the patterns extracted by the PaRE method to justify whether δa; bc∩ 2 R. Obviously, this



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

formulation incurs a large overhead as it requires posing a large number of RelQueries for each query entity as well as processing the documents retrieved by those queries.

## a) Rel Term Search:

CoRE utilizes the existing set of linked pairs towards learning Relation Expansion Terms (i.e., RelTerms) for any relation R. This task involves two main challenges:

- (i) learning a set of high-quality candidate RelTerms from each existing linked pair
- (ii) Consolidating and pruning those individual candidate sets into a minimal global set of RelTerms that are used in the formulation of RelQueries.

CoRE formulates and issues a set of Relation Queries (i.e., RelQueries) for each query entity a based on the set of learned RelTerms. However, there are many possible formulations, each of which is based on a and a conjunction of RelTerms. Clearly, formulating and issuing all those queries will incur a large overhead, which is impractical. Hence, one major challenge is to minimize the number of issued RelQueries while at the same time maintaining high-accuracy for the RC task. Towards achieving that goal, we propose two orthogonal techniques:

1) a confidence-aware termination (CA-Term) condition, which estimates the confidence that a candidate target entity is the correct one a tree-based query formulation method, which selects a small subset of RelQueries to be issued as well as schedules the order of issuing those RelQueries.

## b) Context aware relation extraction:

This formulation is based on our proposed CoRE, in which multiple RelQueries are posed, each of which is based on the query entity a in conjunction with several RelTerms extracted by the CoRE method for e.g., (“Bob Brown” | “Department”), (“Bob Brown” | “Faculty”), etc.). By using RelTerms, a limited number of documents are retrieved, among which some are RelDocs that contain the correct target entity. Context-based formulation tries to strike a fine balance between a very strict RelQuery formulation (i.e., pattern based) and a very relaxed one (i.e., query-based). Towards this, CoRE exploits Rel Term towards a flexible query formulation in which a RelQuery is formulated based on the query entity a in conjunction with one or more RelTerms.

## c) Relation Completion:

Rel Terms from each of the existing individual linked pair, CoRE selects a set of general RelTerms from those candidates. The goal is to select a set of high-quality RelTerms for effective query formulation, and in turn accurate relation completion (i.e., finding target entities). In CoRE, this task takes place in two steps: in the first step, CoRE uses a local pruning strategy to eliminate the least effective RelTerms, and in the second step, CoRE uses a global selection strategy to choose the most effective RelTerms. During the local pruning step, CoRE verifies the effectiveness of each RelTerm in extracting the target entity for the linked pair from which it was learned. In particular, in the verification of a linked pair such as considered as a seed RelQuery without auxiliary information and each learned RelTerm used as a candidate to such seed query with auxiliary information

## IV. CONCLUSION AND FUTURE WORK

In this paper, we conclude that result will get from Pattern Aware Relation Extraction (PaRE) method is some time not accurate because it uses particular extracted pattern search for extracting the information from web big database application. We identify the Relation completion is one recurring problem that is the central to the big database application. As an alternative of PaRE we are proposing the CoRE. So we conclude that CoRE will effective and efficient method which uses Rel Terms instead of Pattern-Based search for extraction of information from web big data application. Thus from this survey we conclude that CoRE will give much more accuracy than PaRE, also th Relation Completion (RC) task will be solved by the CoRE method efficiently and accurate than Pattern Aware Relation Extraction method. We also demonstrate the effectiveness and efficiancypof our proposed techniques in learning relation terms and formulating queries.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

## REFERENCES

- [1] Zhixu Li, Mohamed A. Sharaf, Laurianne Sitbon, Xiyong Du and Xiofong Zhou, "CoRE: Context Aware Relation Aware Extraction method for relation completion", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, vol. 26, no 4, April 2014
- [2] Bhor Priyanka, Deolekar Shweta., Gaikwad Revati and Gonjare Shraddha "Mitigating relation completion problem in big data application" IJARIE, vol. 2 issue 2, pp. 378-385, 2016.
- [3] Priya Pujari, Arti Waghmare, "Improved search result of keyword query using data imputation approach", IRJET, vol.2, issue 2, pp. 1691-1695, Nov.2015
- [4] Fabio Petroni, Luciano Del Corro and Rainer Gemulla, "CoRE: Context Aware Relation Extraction with Factorization Machines", Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp.1763-1773, September 2015.
- [5] Veena Talikoti and M. Velammal, "Context Based Document Retrieval using Taxonomic Tree Structure", Proceedings of 23rd IRF International Conference, 5 April 2015, Chennai, India
- [6] B.Sivaranjani and Meena Selvaraj, "Context Aware Relation Extraction Method for Relation Completion", IJIRCCCE Vol. 3, Issue 3, pp.1927-193, March 2015
- [7] Yankaj Lin, Zhiyuan Liu, Maosong Suni, Yang Liu and Xuan Zhu, "Learning Entity and Relation Embeddings for Knowledge Graph Completion" In proceedings of AAAI'15 conference, feb 2015
- [8] Robert West, Evgeniy Gabrilovich, Kevin Murphy, Shaohua Sun, Rahul Gupta and Dekang Lin, "Knowledge Base Completion via Search-Based Question Answering", International World Wide Web Conference Committee (IW3C2), pp. 515-525, April 7-11, 2014, Seoul, Korea
- [9] Ravi Gummadi, Anupam Khulbe, Arvind Kalavagattu, Sunil Salvi and Subbarao Kambhampati, "SMARTINT: Using Mined Attribute Dependencies to Integrate Fragmented web databases", ACM 978-1-4503-0632-4/11/03., 28 March 2011.
- [10] Nguyen Bach and Badaskar, "A survey on Relation Extraction", Language Technologies Inst., Carnegie Mellon Univ., 2007.
- [11] Michela Banko, Michael J Cafarella, Stephan Soderland, Matt Broadhead and Oren Etzioni, "Open Information Extraction Form the Web", IJCLIA, vol. 7, pp. 2670-2676, 2007.
- [12] Eugene Agichtein and Luis Gravano "Snowball: Extracting Relations from Large Plain-Text Collection", Proc. Fifth ACM Conf. Digital Libraries (ACMDL), pp. 85-94, 2000.
- [13] Natalia Konstantinova, "Review of Relation Extraction Methods: What Is New Out There?", Springer International Publishing, Switzerland 2014, pp. 15-28, 2014.