Optimisation Using Rearrangement of Order of BGP and TLBO Approach

Sahil Saharan\textsuperscript{a*}, J.S. Lather\textsuperscript{b} and R.Radhakrishnan\textsuperscript{c}

\textsuperscript{a} Department of Computer Applications, National Institute of Technology, Kurukshetra, India
\textsuperscript{b} Department of Electrical Engineering, National Institute of Technology, Kurukshetra, India
\textsuperscript{c} Department of Computer Science and Engineering, ABES Ghaziabad(Uttar Pradesh), India

Abstract

Semantic Web data is increasing day by day. The important issue of this web is to handle inordinate volume of information that has many other challenges like query processing and optimisation over widely distributed RDF data. The proposed approach use the reduced querying cost, and hereby optimising the execution time of the query. Optimization of query is one of the most popular problems existing in RDF data which is among the hardest combinatorial optimization problem. Our finding and experimental result concludes that TLBO(Teaching Learning Based Optimisation) outperforms in terms of execution time of the query when compare with ACO variants. In experiment, different types of queries are taken into account like chain, star to show the results of execution time taken by query. This approach centered on main-memory RDF data model.

© 2018 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 6th International Conference on Smart Computing and Communications.

Keywords: RDF query optimisation; Teaching Learning Based Optimisation(TLBO); SPARQL

1. Introduction

Internet data is ever-growing for sharing information among the people, not by machine. So a solution called Semantic web [1] that provide different facility to machine also like sharing, understanding and manipulation on information. And this linked data available on the semantic web is represented by Resource Description Framework (RDF) [2] which is a logical data model used to handle unstructured data, provide capability of machine-interpretability of data. Due to the machine- interpretability property of semantic data, complex query of a user can
be fulfilled easily and efficiently by semantic web instead of current web. Many challenges of this wide RDF data are query processing, ontology matching, query optimisation over billions of triples. Query optimisation is a wide area for research and different methods have been applied to query optimisation area on different database [3-9]. Fast RDF query engines find many problems in the process of optimisation when query execute through distinct part of the query processing and optimisation phase. These distinct parts with the help of some methods provide good query path of the query. The response time of the query depends on query paths. The query paths are proportional to the size of the query. Thus, to resolve these issues, various soft computing techniques have been introduced in the literature in Semantic web context are: two-phase optimisation (2PO) algorithm [10], genetic algorithm (GA) [11], ant colony optimisation (ACO) algorithm [12].

As these soft computing techniques used in the semantic web to RDF query optimisation are inspired by the already used techniques in different domains or traditional databases. But the Semantic web has a more complex and difficult environs to solve all the dependencies.

TLBO is a metaheuristic optimisation technique that is developed by Rao et al. [13,14]. Through this paper, we are introducing TLBO applicability to the RDF query optimisation in specific query forms over a single data source.

The paper structure has the following sections: section 2 introduced concept of RDF and SPARQL queries. In addition, we provide an illustration for BGP understanding and Jena1 API and the ARQ engine used are also introduced. In section 3, presents the description of problem and introduces our proposed optimisation technique used to optimised different forms of queries over RDF data to find the global optimal solution by arrangement of order of triple pattern. Through Section 4, we provide the performance of our technique by means of execution time, fitness value and solution quality. Finally in later sections, we collect discussion, conclusions and future direction of the proposed work.

2. RDF and SPARQL Queries

The Resource Description Framework(RDF) used to describe metadata i.e. data about data. Schema-free model property of RDF makes it a flexible and highly extendable mechanism. RDF can be taken as a graph data model, where each entity can be represented by vertex and relationship by the edge between them, in which each triple is, represented node-arc-node link [15] fashion.W3C official recommendation states about SPARQL [16] that SPARQL queries consist of triple pattern known as basic graph patterns (BGPs) [17]. These patterns are same as that of triples used in RDF except that they could be variable at all the three positions (subject, predicate, object) and also can be concrete(bound) or variable(unbound) [10].

Jena is an API used to store and manipulate RDF data. Query engine used by Jena is ARQ2 used for our experiment. The order changing technique of join operation of SQL queries enforces to study the rearrangement of order of triple pattern of SPARQL query that also provide greater impact on execution time of the query.

To explain the patterns order of a BGP, we are using a chain query. This Chain query(Table 1) is taken from LUBM3 dataset having six different pattern into it. Triple pattern is represented using sequence number 0 to 5 build in the where clause of the SPARQL pictured in Table 1. When executing this Chain query displayed in Table 1, it takes 108544ms whereas it takes 906ms for \{3, 4, 5, 0, 1, 2\} order and 318ms for \{2, 1, 0, 5, 4, 3\} order optimised Chain query. The reason for the higher execution time of unoptimised query in comparison with optimised query is already explained above.

<table>
<thead>
<tr>
<th>BGP of an example Chain query.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0                      ?advisor ub:teacherOf ?course.</td>
</tr>
<tr>
<td>1                      ?tAsst ub:teachingAssistantOf ?course.</td>
</tr>
<tr>
<td>2                      ?tAsst ub:name ?Name.</td>
</tr>
<tr>
<td>3                      ?std rdfs:subClassOf ub:Student.</td>
</tr>
<tr>
<td>5                      ?student ub:advisor ?advisor.</td>
</tr>
</tbody>
</table>

Main role of rearrangement of order is to provide an optimal query plan which have lower execution cost than the others. [18] suggests a solution to the problem of high execution time of a simple SPARQL query by mentioning

---

1 http://jena.apache.org
2 http://jena.apache.org/documentation/query
3 http://swat.cse.lehigh.edu/projects/lubm/

دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات