RELEVANT RULE DERIVATION FOR SEMANTIC QUERY OPTIMIZATION

Junping Sun†, Nittaya Kerdprasop‡, and Kittisak Kerdprasop∗
School of Computer and Information Sciences, Nova Southeastern University
Fort Lauderdale, Florida 33314, USA†
Department of Mathematics, Chulalongkorn University
Bangkok 10330, THAILAND‡
School of Computer Engineering, Suranaree University of Technology
Nakorn Ratchasima, THAILAND∗

ABSTRACT

Semantic query optimization in database systems has many advantages over the conventional query optimization. The success of semantic query optimization will depend on the set of relevant semantic rules available for semantic query optimizer. The semantic query optimizer utilizes a set of available semantic rules to further explore extra query optimization plans for conventional query optimizer to choose. Semantic rules represent the dynamic database state at an instantaneous time point. Finding such set of relevant semantic rules can be very beneficial to support both semantic and conventional query optimizations. In this paper, we will show how to use inductive logic programming approach to derive relevant rules from the data in a database. Language bias as heuristic is used to reduce the search space as well as the costs in the process of inductive rule derivation. Effectiveness and efficiency of our bias generator algorithm are evaluated and their evaluation results are presented in this paper.

1 INTRODUCTION

Query optimization in relational database systems has been an active research subject for many years. The techniques used for relational query optimization can be classified into conventional query optimization and semantic query optimization. The conventional query optimization utilizes syntactic and logic equivalence transformation of a query expression, and various statistical estimation profiles of a database. The capabilities of conventional query optimization are limited to syntactic information in a query expression, however. In comparison, semantic query optimization uses not only syntactic transformations, but also semantic knowledge expressed by semantic integrity constraints or semantic rules to aid the query transformation (semantic equivalence transformation).

Early research work of semantic query optimization can be traced back to knowledge-based query optimization [11], QUIST (QUery Improvement through Semantic Transformation) [14] [15], deductive logic-based approach [3] [5], and others [4] [23] [24]. In QUIST [15], King used a set of integrity constraints as a knowledge base to transform or simplify a query into an optimized one by semantic reasoning. Heuristics such as *index introduction, restriction introduction, join elimination, contradiction detection, etc.* are used to explore more alternatives for conventional query optimizer. Further, if there is contradiction detected between a query and integrity constraints, then it implies that the query cannot be satisfied. So the empty answer to the query can be given directly from the semantic transformation. Of course, the query is semantically optimized since the query can be answered without accessing the physical database. Charkravarthy *et al.* [3] proposed a deductive database-based logic model, which consists of extensional database (EDB), intensional database (IDB), and integrity constraints (ICs), to formalize the semantic query optimization; and resolution refutation method was used to perform semantic query transformation. Siegel *et al.* [21] [22] used automatic rule derivation method based on deduction to learn new rules applicable to semantic query transformation. With the set of user-defined integrity constraints and new derivable rules associated with EDB and IDB predicates, more opportunities can be explored for both semantic query optimization and conventional query optimization. All of these methods mentioned above [3] [5] [11] [14] [15] [20] [21] [22] conduct semantic transformation based on a set of integrity constraints, a set of intensional database predicates(IDB), and a set of extensional database predicates (EDB). But the rich semantics embedded in database instances is overlooked in terms of semantic rule derivation [21] [22] and semantic query optimization. In this paper, we will present a heuristic-based approach to induce a set of relevant semantic rules from not only the set of ICs, IDB and EDB predicates, but also the database instances for semantic query optimization. From this point of view, more semantic rules will provide more alternatives for semantic query optimization.