MASTER'S THESIS

Incremental maintenance of minimal and minimum bisimulation of cyclic graphs

Deng, Jintian

Date of Award:
2011

Incremental Maintenance of Minimal and Minimum Bisimulation of Cyclic Graphs

DENG Jintian

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Philosophy

Principal Supervisor: Assistant Professor CHOI Koon Kau

Hong Kong Baptist University

September 2011

Abstract

Graph databases have numerous recent applications, such as the Semantic Web, ontology representation, social networks and XML *etc.* A fundamental structural index for data graphs, namely *bisimulation*, has been reported useful in efficient path query processing and optimizations including selectivity estimation and graph compression, among many others. It is desirable to adapt a *minimal* or *minimum* bisimulation for query efficiency of their applications. In this thesis, we devise two incremental maintenance algorithms of *bisimulation* for a possibly *cyclic* data graph. Cyclic graphs are ubiquitous among the data on the Web, however, previous work on the maintenance of minimum bisimulation mostly focuses on *acyclic* graphs. In practice, graphs are subject to update. Therefore *bisimulation* has to be updated accordingly.

Firstly, in the context of database applications, it is natural to compute *minimal* bisimulation with merging algorithms. We propose an incremental maintenance algorithm for a *minimal* bisimulation of *cyclic* graphs, in the style of merging. In order to prune the computation on non-bisimilar `SCC`s, we propose a feature-based optimization. The features are designed to be constructed and used more efficiently than bisimulation minimization. We also conduct an experimental study that verifies the effectiveness and efficiency of our algorithm. For illustration purpose our features-based optimization can prune 50% (on average) unnecessary bisimulation computation on our synthetic data. Our experiment verifies that we extend the current state-of-the-art with a capability on handling *cyclic* graphs in maintenance of *minimal* bisimulation.

Secondly, while previous algorithm focuses on optimization issues and does not guarantee *minimum* bisimulation, we propose a hybrid incremental algorithm which guarantees *minimum* bisimulation even for a *cyclic* graph. To study the problem in the context of cyclic graphs, we first show that the two classes of existing minimization algorithms – merging algorithm and partition refinement – have their strengths and weaknesses. Second, we propose a novel hybrid algorithm and its analytical model. To the best of our knowledge, this is the first maintenance algorithm that guarantees *minimum* bisimulation of *cyclic* graphs. Third, while small updates on a graph may affect the *minimum* bisimulation arbitrarily, our preliminary experiments show that in practice, this is often rare. We propose to partially reuse the *minimum* bisimulation before an update to optimize maintenance performance. Finally, we present an experimental study on both synthetic and real-life datasets that verifies the efficiency and effectiveness of our algorithms.

Thirdly, while for path query answering, many indexes are *bisimulation*-based, there have been a large number of ad-hoc indexes for reachability query in graph databases. The structures and performances of these indexes on different graphs have known to be very different. In this thesis, we propose a hierarchical prediction framework to select the optimal index for a graph database.

Keywords: graph analysis, graph algorithms, evolving graphs, cyclic graphs, minimal bisimulation, minimum bisimulation, reachability query, path query.

Table of Contents