

# QUANTUM COMPUTING FOR RANGE QUERIES AND PREFERENCE QUERIES

by

XIAOTIAN YOU

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

## ABSTRACT

Quantum computing is a popular topic in computer science. In recent years, many new studies came out in different areas such as machine learning, network and cryptography. However, the topics in the database area seem long neglected. There is an open problem in the database area: Can we make an improvement on existing data structures or data mining algorithms by quantum techniques?

Assume that there is a dataset of cars, where each car has its price and horsepower. Consider two different applications. One application is that a user gives acceptable ranges of the price and horsepower, and then we could find all cars within the ranges. Another application is that a user gives his or her preference (e.g., a cheap car with a high horsepower), and then we could find a set of cars satisfying this preference. We study the problem called the *quantum range query* problem for the first application and the problem called the *quantum preference query* problem for the second application. In this thesis, we introduce these two quantum problems in the database area, and discuss how to solve them with quantum algorithms to obtain better efficiency than classical algorithms.

The first quantum problem is the quantum range query problem. Consider a dataset

of key-record pairs. Given an interval as a query range, a B+ tree can report all the records with keys within this interval, which is called a *range query*. A classical B+ tree answers a range query in  $O(\log_B N + k)$  time, where  $B$  is the branching factor of the B+ tree,  $N$  is the total number of records, and the output size  $k$  is the number of records in the interval. It is asymptotically optimal in a classical computer but not efficient enough in a quantum computer, because it is expected that the execution time is linear to the output size. Motivated by the different scenarios in real-world applications, we study the quantum range query problem in three different cases. The first case is the *static quantum range query*, where the dataset is immutable. The second case is the *dynamic quantum range query*, where insertions and deletions are supported. The third case is the *high-dimensional static quantum range query*, where the keys are high-dimensional points. Firstly, we propose the static quantum B+ tree that answers a static quantum range query in  $O(\log_B N)$  time, which is asymptotically optimal in quantum computers. Since the execution time does not depend on output size (i.e.,  $k$ ) and  $k = O(N)$ , it is exponentially faster than the classical data structure. To achieve this significant improvement on range queries, we design a hybrid quantum-classical algorithm to do the range search on the static quantum B+ tree. Secondly, we extend it to a dynamic quantum B+ tree. The dynamic quantum B+ tree performs an insertion or a deletion in  $O(\log_B N)$  time and answers a dynamic quantum range query in  $O(\log_B^2 N)$  time. Thirdly, based on the static quantum B+ tree, we propose the static quantum range tree which answers a  $d$ -dimensional static quantum range query in  $O(\log_B^d N)$  time, which cannot be achieved by any classical data structure.

The second problem is the quantum preference query problem. The preference query problem, which is to find the most preferred tuples from a dataset, is widely discussed in the database area. In this problem, a utility function is given by the user to evaluate to what extent the user prefers the tuple. However, considering a dataset consisting of  $N$  tuples, the existing algorithms either need  $O(N)$  time to answer a query or need  $O(N)$  time for a cold start to answer a query. The reason is that in a classical computer, a linear time is needed to evaluate the utilities by the utility function for  $N$  tuples. In this thesis, we discuss the quantum preference query (QPQ) problem. In this problem, the dataset is given in a quantum memory, and we use a quantum computer to return the answers. Taking the advantage of quantum parallelism, the quantum algorithm can theoretically perform better than their classical competitors. To better cover all the possible study directions,

we discuss this problem in different kinds of input and output. In the QPQ problem, the input can be either a number  $k$  or a threshold  $\theta$ . Given  $k$ , the problem is to return  $k$  tuples with the highest utilities. Given  $\theta$ , the problem is to return all the tuples with utilities at least  $\theta$ . Also, in the QPQ problem, the output can be classical (i.e., a list of tuples) or in a quantum state (i.e., a superposition in quantum bits). Based on amplitude amplification and post-selection, we propose four quantum algorithms to solve the problems in the above four scenarios. We give an accurate analysis of the number of memory accesses needed for each quantum algorithm, which shows that the proposed quantum algorithms are at least quadratically faster than their classical competitors.

For the quantum range query problem, in the experiment, we did simulations to show that to answer a range query, the three quantum data structures are up to  $1000\times$  faster than their classical competitors. To the best of our knowledge, the quantum B+ tree is the first tree-like quantum data structure that achieves a better complexity than classical data structures. For the QPQ problem, the experimental results also show that the QPQ quantum algorithms are up to  $1000\times$  faster than their classical competitors, which proved that QPQ problem could be a future direction of the study of preference query problems.