The Database Don

By Singapore Management University | Editorials
February 3, 2014

SMU Associate Professor Kyriakos Mouratidis aims to maximise the potential of databases to provide solutions to real-world problems.

AsianScientist (Feb. 3, 2014) – By David Tan – Professor Kyriakos Mouratidis likes things fast. He likes to be able to ask questions and get answers quickly.

"The main objective of my research is to provide fast answers," says the Associate Professor at the Singapore Management University (SMU) School of Information Systems, who specialises in spatial databases and is motivated by real-world needs and applications.

Professor Mouratidis is looking at voluminous data collections and ways to extract information from them efficiently.

"Database research studies ways to store, organise and browse massive amounts of data so that user queries can be answered quickly, without scanning the entire database," he explains.

In particular, he works with spatial databases, in which data records have locations (geographic coordinates) associated with them. With these, users can make queries such as a nearest neighbour query.

"If you want to find the ten nearest Chinese restaurants to where you are, this would be an example of a nearest neighbour query."

Traditional approaches allow for 'snapshot' queries on relatively static data. But with the proliferation of GPS-enabled devices such as smartphones and tablets, users can now ask more complex questions.

"This has given rise to advanced location-based services," he says.

Nearest neighbour queries and driving distance

"Research that I started during my PhD and that I am now continuing, deals with moving data points and moving user queries. These queries require continuous re-evaluation of dynamically changing results over a long period of time."

Such queries are known in the field as 'standing K nearest neighbour' (K-NN) queries, he says.

"A standing K-NN query would help me monitor which ten Facebook friends are closest to me at any time of the day, as my friends and I continuously move around. This query requires constant update of the 'closest ten' result in real time."

Professor Mouratidis specialises in developing algorithms and data organization methods that minimise the computational overhead to process multiple standing queries.

"Ideally, we want to only look into specific parts of the database in order to provide fast answers."

Another real-life application of this technology is in a smartphone app for booking a taxi. Taxi drivers and potential clients are continuously moving in the city. A taxi driver would want to know where and who the ten closest possible clients are. Users can update their location through the smartphone app, allowing the taxi driver to see a continuously refreshing list of the closest clients, thus enabling more efficient taxi services, says Professor Mouratidis.

While the aim appears simple, achieving it with a large volume of users is not.

"The issue is how to cope with scale. There could be hundreds of thousands of users that day, so how do you provide quick refreshes of the result lists?"

To aid this process, Professor Mouratidis is working on data structures and algorithms that allow for computation sharing and incremental evaluation of results.

As with most location-based queries, what matters is the travelling distance between two points, rather than the linear (also known as euclidean) distance between them, says Professor Mouratidis.

"In reality, we have to move along roads, we cannot pass through buildings, so our movement in the city is constrained by the road network. Therefore, we define proximity as the shortest driving distance from, say, a taxi to a potential client."

He has developed efficient methods for the continuous monitoring of K-NNs according to the driving distance between queries and data points in the city's road network. In addition to query and data movements, these algorithms can also handle network changes, such as varying traffic conditions.

Keeping things private

In addition to his research on standing spatial queries, Professor Mouratidis is also exploring location privacy. In an era where service providers are collecting ever more data about users and mining these data to thoroughly maximise their potential, the issue of privacy is a pressing one, he says. To address this, he is studying ways to mask database queries so that the user's privacy is protected.

"Looking at our previous examples, a very common location-based service is routing the user from a location in the city to another, which is essentially a shortest path query in the city's road network. Typically, the query is sent to a remote server, such as Google's server, which processes the query and returns the shortest route. The thing is that the server learns where the user is, where he is heading and what route he will most likely follow."

Some users simply do not wish to be profiled, says Professor Mouratidis.

"For instance, a route you're searching for takes you from home to the nearest supermarket, and you might not want to receive unsolicited advertisements from companies who've figured out your shopping habits and patterns."

"The situation becomes even more serious if the destination is of a sensitive nature, such as a health facility for cancer or HIV treatment. Leak or misuse of this information would be detrimental to the user," he adds.

Professor Mouratidis is working on a method by which users could still query a remote database, but prevent the database from learning anything about their locations, their destinations or the routes taken. The trade-off for privacy protection is slower query execution.

"The cost to the user is that he has to wait longer. He could choose an unprotected query to get a fast answer, or he could choose to mask the query and wait, say, five to ten times longer for the answer. The important thing is that he's given that option."

Professor Mouratidis acknowledges that location-based services might be reluctant to adopt a private query processing model.

"This option would disallow them gain of valuable information about their users and would incur both one-off and ongoing costs. For example, trusted hardware is expensive. The only incentive for location-based services to adopt a private processing option would be if the users strongly demanded it, or by fear that competition may start offering it. A general

impression is that this demand has started showing, but whether it is strong enough remains to be seen."

Fine-tuning preference queries

Professor Mouratidis also collaborates with Professor Pang Hwee Hwa of the SMU School of Information Systems on improving multi-criteria decision-making. Specifically, they are working on a mechanism by which users can fine-tune their preferences in search queries.

"For instance, users of tripadvisor.com who browse hotels, place different weights on key criteria such as price, cleanliness and service."

Based on a user's preferences, tripadvisor.com may produce an aggregate score for each hotel, and present him with the ten hotels with the highest scores. This is a typical top-K query, explains Professor Mouratidis.

By defining 'immutable regions' for each decision factor (e.g., price, cleanliness and service), which refer to the maximal deviation from the user-specified weights for which the query result (e.g., the top-ten list) remains the same, the two SMU professors are able to perform a sensitivity analysis to progressively fine-tune the query weights.

With its focus on real-world challenges and applications, it is clear that Professor Mouratidis' work will be able to improve and refine the way businesses interact with their customers. He is open to new collaborations within SMU or with external partners.

"Some aspects of my work, like spatial optimisation and multi-criteria decision making, might be interesting to colleagues in the business school but also to local authorities."

Asian Scientist Magazine is a media partner of the Singapore Management University Office of Research.

___