

■ Takeaki Uno

Associate Professor, Principles of Informatics Research Division

[TITLE]

Getting ordinary PCs to solve huge scale problems

[MAIN BODY]

A computer program has to have a good blueprint for how to proceed huge scale operations. Such a program has to be coded with well-designed algorithms. My goal is to get ordinary PCs to solve huge scale problems that only high-performance computers can solve right now by developing new algorithms.

There are two basic targets in algorithm research: theory and practical application. In the first kind of research, researchers tend to assume worst-case conditions for a problem to be solved and to design a method based on that assumption. But conditions in real world are rarely so dire, and using a method designed to solve the worst case problems can result in executing lots of unnecessary procedures. In my research, I design the most efficient ways for solving real problems while trying to fully understand the theoretical research results.

Improving data-mining efficiency

In the realm of databases, there's a field called data mining. For example, to formulate a sales strategy for convenience stores, systems look for combinations of products purchased from past data, based on various conditions like "people purchasing sandwiches also tend to buy tea." Some online-bookstores apply this technique, too. They do a search based on the idea (collaborative filtering) that people who buy a certain book are likely to buy another certain book, then select certain books and post signs saying: Recommended Books.

Checking lots of these combinations takes a lot of time. If the search results are numerous, the information becomes hard to use. It's more efficient to look for meaningful combinations. To that end, I've developed several algorithms so that the computation is done in quite short

time. It is a development by theoretical approaches, without using any help by hardware.

Potential of algorithm research

The results of algorithm research are often applied by researchers in other fields. I'm pretty active in helping users apply the results of algorithm research. In return, I receive problems from users that pollinate my research activities.

Right now, I'm helping create algorithms to reduce the calculations needed to analyze and study the genome data. To analyze the arrangements of the various bases that constitute the genomes, we use restriction enzymes to cut the genome into several hundred base units before examination. But a genome can't be cut sequentially from beginning to end, because the genome is shredded at the section affected by the enzyme's action, leading to the mixture of numerous fragments. We have to confirm the arrangement of the fragments first, then look at the ends of each fragment and determine what fragment connects to what fragment to reconstruct the original genome. This generally requires several hundred computers running at full capacity for days. After reviewing the algorithms used for this task, I managed to reduce the computing time needed to several hours by a new method: grouping and comparing similar fragments.

We can achieve technical breakthroughs by reviewing data processing from an algorithmic perspective. My interest is in tackling various problems based on this approach.

(Interviewed and summarized by Atsushi Saito)