

Title	Judgement assistance of search engine results by using the Batch-Learning Self-Organizing Map algorithm
Author(s)	中本, 修
Citation	
Issue Date	2003-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/452
Rights	
Description	Supervisor:Ho Tu Bao, 知識科学研究科, 修士

Judgment assistance of search engine results by using the Batch-Learning Self-Organizing Map algorithm

Osamu Nakamoto

School of Knowledge Science,
Japan Advanced Institute of Science and Technology
March 2003

Keywords: Batch-Learning Self-Organizing Map algorithm, search engine,

Objective

It proposes the system which becomes the judgment assistance of the result of the search engine in the this thesis.

The document of the prodigious number exists by the spread of the Internet in Web. As a result, it came to be able to obtain a lot of information at once. However, it is difficult to search out information, which is necessary. The search engine exists as a tool helps this.

The search engine can be classified into the robot type and the directory type. Present, especially use is a robot type search engine. It can be said that it is difficult to look for information is necessary from the result of the robot type search engine. The reason is that the retrieval result is displayed without being made to the classification.

There is two timing that the retrieval result classifies. One is a method of the classification before, and another one is a method of the classification after. The classification before is done by the directory type search engine.

It thinks the classification after it retrieves it. There is a classification according to an absolute standard. There is a classification relatively done. A relative classification will be done in this thesis. There is Self-Organization Map(SOM) as a means to classify, and to output from the feature of input information. However, SOM has the feature depends in the order of inputting data. I want to treat any data equally this time. Then, Batch-Learning Self-Organization Map (BLSOM) which removes the input order dependence

from SOM is applied.

Methodology

The word was paid to attention to request the feature of the document. The document is formed with the word. To guess what feature the document has from the word, the frequency of the word and the location, etc.

Frequency and the position of words have stored the search engine index file as a score. The document is a set which consists of the word and the score. Document set is the vector document in another expression. The file of the vector of the document, which describes the word and the score from the index file is made. This file of the vector of the document is used to compare features of the document besides a certain document.

Evaluation

The system was evaluated by the recall ratio and the precision ratio. The recall ratio is a word composition rate in the group. The precision ratio is a use rate of the word in the group. The document of the document, which was adjacent by the map felt comparatively similar was abundant. Therefore, it succeeded as a system. However, it takes necessary time to obtain an enough classification result. It is necessary to devise it a little more in the practicality.