

## 論文の内容の要旨

A Study on Efficient Algorithms for Machine Learning from Large-scale Data  
(大規模データからの機械学習のための効率的なアルゴリズムに関する研究)

氏名 松島 慎

(本文)

Machine learning has been developed greatly as a tool for extraction of useful knowledge from large amounts of data.

As processing power and network speed have not kept pace with our ability to collect large amounts of data, development of more efficient machine learning algorithms is a real and pressing need.

In this thesis we focus on efficient algorithms for supervised learning. First we focus on online algorithms that are based on batch learning model assuming that a fixed number of examples are drawn from a certain distribution. Second we look at online algorithms that are based on online learning model assuming that an unfixed number of examples are coming one by one without any probabilistic assumption.

In both models, we claim online algorithms are efficient and then suitable to treat with large-scale data.

The main contribution of the thesis is two-fold. First, we consider efficient batch learning to exploit the traits of a single computer. Especially we focus on the exploitation of memory hierarchy with a claim of necessity of a systems-aware algorithm and we derive efficient online algorithms in terms of wall clock time.

We give an architecture called dual cached loops to deal with the problem which occurs when data size exceeds the main memory capacity.

We demonstrate that this can be applied for various problems in supervised learning.

Finally, focusing on the online learning model, we give efficient online algorithms in multiclass classification and demonstrate that they can reduce learning time.