

半構造化データからの学習に関する研究



教授 松本 哲志

Prof.

Satoshi MATSUMOTO

Efficient learning from semi-structured data

Keyword : 計算論的学習理論, 質問学習,
半構造化データ
Computational Learning Theory,
Query Learning,
Semi-structured data

計算論的学習理論では、いくつかの学習モデルが提案されています。その中でも、Angluinによって提案された質問学習モデルに着目しています。この学習モデルでは、学習者が教師に対していくつかの質問を繰り返し行うことにより、学習の対象となる言語族を見つけ出します。

インターネット上に存在するデータは、HTML形式やXML形式などの形式を用いて記述されている場合が大多数です。このようなHTMLファイルやXMLファイルなどのWeb文書が半構造化を持つことに注目し、質問学習モデルの枠組みの中で、半構造化データに共通して現れる特徴を抽出する手法について研究を行っています。

In computational learning theory, some learning models are proposed. In Angluin's query learning model, a learning algorithm collects information about a target language to be learned by accessing oracles that answer specific types of queries. Finally, the learning algorithm outputs a representation of the target language.

Due to the rapid growth of Internet usage, semi-structured documents such as XML/HTML files have been rapidly increasing. We aim to establish methods for extracting structural features common to such semi-structured data in the framework of Angluin's query learning model.

