

特 集

博 士 論 文

◆ 機械学習・データマイニング

Query Refinement for Domain-Specific Web Search

(ドメイン指向 Web 検索のための質問の精練)

小山 聡

oyama@i.kyoto-u.ac.jp

京都大学大学院情報学研究所社会情報学専攻

指導教官：石田 亨

博士 (情報学), 2002年3月 取得



Keywords: Web 検索, データマイニング, 機械学習

概要: 本論文では, Web から特定のドメインに関係するページだけを検索できる専門検索エンジンを構築するための新しい手法を提案した. “検索隠し味”と呼ばれる, ドメインのページを分類するキーワードのブール式を利用者の質問に付加し, 汎用検索エンジンに転送することで, 専門検索エンジンを実質的に構築する. そのために, 検索隠し味を Web ページの訓練集合から抽出する機械学習のアルゴリズムを開発した. 決定木を用いて抽出された初期の複雑なブール式を, 検証集合に対する適合率と再現率の調和平均に基づいて単純化することで, 適合率と再現率のバランスのとれたブール式を, 汎用の検索エンジンに投入可能な小さなキーワード数で抽出できる. この手法により, 単純なシステム構成で, 応答性能の良いドメイン専門検索エンジンを, 専門的知識なしに構築することが可能になった.

さらに, ドメインの Web ページ集合内で利用者の要求を満たすページを検索する問題に, 連想ルールを適用する手法を提案した. ドメインに固有の検索対象の属性を考慮して連想ルールを適用することにより, 検索の制約条件の緩和や制約条件の追加といった質問の洗練を行うことを可能とした. そのために, ドメインの Web ページ集合から抽出された連想ルールを, キーワードのクラスタリングと統計的検定を用いて精選する手法を提案した. この手法により, ドメイン内での検索に有効な連想ルールの割合を向上させることが可能となった.

公表論文: S. Oyama, T. Kokubo, T. Ishida, T. Yamada and Y. Kitamura: Keyword Spices: A New Method for Building Domain-Specific Web Search Engines, International Joint Conference on Artificial Intelligence (IJCAI-01), pp. 1457-1463 (2001)

現職: 京都大学大学院情報学研究所社会情報学専攻助手

論文入手先: <http://www.dl.kuis.kyoto-u.ac.jp/~oyama/>

抱負: 今後も, 機械学習やデータマイニングの技術を基礎として, Web 検索や Web からの知識発見に関する研究を行っていきたくと考えている.

◆ 機械学習・データマイニング

Classification Learning and Rule Extraction Systems Using Extended Genetic Programming for Data Mining

(データマイニングを目的とした遺伝的プログラミングの拡張による分類学習システムとルール生成システム)

新美 礼彦

niimi@fun.ac.jp

桐蔭横浜大学大学院工学研究科制御システム工学専攻

指導教官：田崎栄一郎

博士 (工学), 2002年3月 取得



Keywords: データマイニング, 遺伝的プログラミング, 決定木学習, 相関ルール分析, 組合せ学習

概要: 本論文は, 遺伝的プログラミングを用いたデータマイニングについて論じたものであり, 7章からなる. 1章の序論に続き, 2章では, 提案する手法の核となる遺伝的プログラミングについて説明した. 3章では, 遺伝的プログラミングと決定木構築法を組み合わせる手法を提案した. 決定木構築法と遺伝的プログラミングを組み合わせることでデータマイニングに用いることにより, より精度の高い知識や意外な知識を抽出することを提案した. 4章では, 遺伝的プログラミングと相関ルール抽出法を組み合わせる手法を提案した. 遺伝的プログラミングにより相関ルール抽出法の出力から特徴を抽出すると同時に, 遺伝的操作により出力に含まれなかった知識の抽出も行うことができる. 5章では, 遺伝的プログラミングと論理関数を組み合わせる手法を提案した. これは, 遺伝的プログラミングの関数ノードに論理関数を定義することによって, 論理関数によるデータマイニングを行う手法である. 6章では, 遺伝的プログラミングに変更を加えることにより, 遺伝的プログラミングが大規模データマイニングにおいても十分に利用可能であることを明らかにした. 改善手法として, オブジェクト指向の組合せ学習, 無効ノード削除と連続値属性適応処理の二つの強化学習を組み合わせる方法, 交叉個体を複数指標により選ぶ方法の三つを提案した. 7章は, 本論文における提案手法を総括し, さらなる改善の可能性について検討した.

公表論文: A. Niimi and E. Tazaki: Combined Method of Genetic Programming and Association Rule Algorithm, Applied Artificial Intelligence, Vol. 15, No. 9, pp. 825-842 (2001)

現職: 公立はこだて未来大学システム情報科学部情報アーキテクチャ学科助手

論文入手先: niimi@fun.ac.jp

抱負: ネットワーク上に分散しているさまざまなデータベースを用いて, データマイニングを行う手法を研究していきたい. 遺伝的プログラミングの応用についても研究をすすめていきたい.