



Title	A Study on Machine Learning Algorithms Using Feature Interactions [an abstract of dissertation and a summary of dissertation review]
Author(s)	新, 恭兵
Citation	北海道大学. 博士(情報科学) 甲第14581号
Issue Date	2021-03-25
Doc URL	http://hdl.handle.net/2115/81146
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	theses (doctoral - abstract and summary of review)
Additional Information	There are other files related to this item in HUSCAP. Check the above URL.
File Information	Kyohei_Atarashi_abstract.pdf (論文内容の要旨)



[Instructions for use](#)

学位論文内容の要旨

博士の専攻分野の名称 博士（情報科学） 氏名 新 恭兵

学位論文題名

A Study on Machine Learning Algorithms Using Feature Interactions

(特徴の組合せを扱う機械学習アルゴリズムに関する研究)

大量のデータから自動的に規則を学習する機械学習の技術が多くの分野で成果をあげている。近年は、単に高い精度で予測をするだけでなく、できるだけ少ない計算リソースで高速に学習することや、機械学習手法に対する解釈性・説明性等が求められるようになってきている。一般に、機械学習手法の予測性能と、解釈性や説明性、要求する計算リソースはトレードオフの関係にあり、高性能で高速に動作しさらに説明性も高い手法を構築することは困難である。

様々な機械学習手法の中でも、特徴の組合せを扱う手法は古くから様々な問題に活用されていたが、予測性能と解釈性・透明性の両方を高いレベルで両立させていることから、再び注目を集めている。しかし、特徴の組合せを用いる手法は、特徴の数とデータの数のどちらも大きい場合は計算量の観点からプログラムを動かすことが困難である、データが大規模な場合であっても利用することができる方法もあるが最適化問題が非凸最適化問題であり解くことが難しく結果的に時間がかかる・良い解が得られるとは限らない、といった解決すべき問題が多く残っている。そこで本研究では、組合せを扱う手法の主要な応用先の一つであるリンク予測のためのより高精度・高速な方法の提案と、特徴の組合せを用いる手法をデータが大規模な場合にも高速に学習するアルゴリズムの提案を行う。

本論文は以下のような構成となっている。

1章では、研究の背景と目的、本研究の貢献の概要、そして本論文の構成について述べる。

2章では、線形モデルやカーネル法、Factorization Machines といった特徴の組合せを用いる主要なモデルとその拡張、そしてその最適化アルゴリズムといった、基本となる既存の技術・事項について述べる。

3章では、特徴ベースのリンク予測のための、特徴の組合せを用いる新たな手法を提案する。リンク予測とは、与えられた二つのオブジェクトの間にリンクがあるか否かを推定する問題である。リンク予測の問題の形式は様々あるが、特徴ベースのリンク予測は其中最も汎用的なものである。特徴ベースのリンク予測として定式化されるいくつかの問題において、二つのオブジェクトの間の特徴の組合せのみを用いる手法が有効であることが知られている。そのような問題においてより高精度な予測を行うために、二例間の高次の組合せを用いるカーネル関数を導出し、その効率の良い計算アルゴリズムとそのカーネル関数を用いたモデルを提案する。提案法と既存の研究の関連について述べ、そして提案法の本質的な部分を取り入れた深層ニューラルネットを提案する。最後にいくつかの実データを用いた実験で提案法の有効性を示す。

4章では、Itemset カーネルに対するランダム特徴を提案する。ランダム特徴とは、特徴ベクトルを「カーネル関数を内積の意味で近似するような高次元空間」にランダムな行列やベクトルを用いて写像することでカーネル法を近似する手法の総称である。Itemset カーネルは特徴の組合せを扱う様々なカーネル関数の一般化となっており、既存の様々な手法を Itemset カーネルを用いて記

述することができる。そのため、提案法を用いた線形モデルは、特徴の組合せを用いることができ、そして高速に学習が可能である。また、既存のランダム特徴の高速化・省メモリ化技術を取り入れた、より高速かつ省メモリなアルゴリズムを提案する。さらに、提案法の近似誤差や、アルゴリズム内で用いることのできる確率分布に関する理論的な結果を述べる。そしてその結果から、より高速で省メモリに動作するアルゴリズムの提案も行う。また、より広いクラスのカーネル関数である重み付き Item-multiset カーネルへの拡張も行う。最後に、提案法が既存手法とほとんど劣らない精度でより高速に動作することを実データ・人工データを用いた実験で確認し、提案法の有用性を示す。

5章では本論文のまとめと今後の展望について述べる。