学位（修士）論文

タイトル：　Subgraph-Feature Search for Learning Classifiers and Regressors under Fixed Budget Constraint

著者：　白川　稜

指導教官：　中村　篤祥

作成日：2019年2月12日

Abstract

Recently, in various fields such as computational chemistry and bioinformatics, attention has been paid to classification and regression of data whose inputs are graphs of arbitrary size and shape, sometimes called graph classification/regression. In these problems, whether a certain subgraph is included or not is a fundamental feature for graphs that have no direct descriptors available. The simplest solution, enumerating subgraph patterns and use them as features, is so expensive since the number of possible subgraph patterns are intractably large due to the combinatorial explosion. Currently, discriminative pattern mining has been studied and given a good solution to this problem. The previous search method for such discriminative subgraph patterns tries to find the best one by the depth-first search with some pruning rules. Such an exact search, however, is costly for heavily repeated use, which is necessary to obtain the number of features enough for high performance classifiers or regressors. To overcome this computational cost issue, we propose two approximate algorithms based on (i) Best- first Search, (ii) Monte Carlo Tree Search(MCTS). We demonstrate effectiveness of our proposed methods by comparing performance in real problems using QSAR datasets.

概要

近年、計算化学やバイオインフォマティクスなどのさまざまな分野では、入力が任意のサイズと形状のグラフデータに対する分類・回帰問題が注目されている。これらの問題において、入力グラフに直接的な記述子が存在しない場合、特定の部分グラフを含むか否かという記述子は最も基本的な特徴量として利用される。しかし、部分グラフを全列挙し特徴として利用する最も単純な手法は、グラフのサイズに対して部分グラフの総数が組み合わせ爆発を起こし膨大になるため、非常に大きなコストを要する。この問題点に対して現在では識別パターンマイニングという手法が研究され、優れた解決方法として利用されている。既存の識別部分グラフパターンの探索手法は、いくつかの枝刈り規則を利用した深さ優先厳密探索によって最適な特徴を発見する。しかし、分類・回帰モデルの学習に利用する場合には性能向上のために反復的に探索を行う必要があるため、厳密探索は依然として大きなコストを要する。この探索における計算コストの問題を解消するために、本研究では(i)最良優先探索、(ii)モンテカルロ木探索に基づく２種類の近似探索手法を提案する。また、実データセットのベンチマークであるQSARを利用して提案手法の有効性を実験的に評価する。

学会発表

学会名：　The 14th International Conference on Mining and Learning with Graphs (MLG 2018)

発表年月日：　2018年8月20日

タイトル：　Jointly Learning Relevant Subgraph Patterns and Nonlinear Models of Their Indicators.

著者：　Shirakawa R, Yokoyama Y, Okazaki F, Takigawa I

内容：　非線形グラフ分類・回帰モデルの学習及びを特徴探索を同

時に行う学習アルゴリズムの提案。特徴探索に関して学習

モデル基準の枝刈り規則を設計することで大幅な探索コス

トの削減を実現。グラフ分類・回帰問題における非線形性

の影響を人口データセット及び実データセットにおける性

能比較より評価。

学会発表

学会名：　第21回情報論的学習理論ワークショップ(IBIS2018)

発表年月日：　2018年11月5日

タイトル：　適応的な部分グラフ指示子の探索・選択に基づく非線形グラフ分類回帰.

著者：　白川 稜、横山 侑政、岡崎 文哉、瀧川 一学

内容：　非線形グラフ分類・回帰モデルの学習及びを特徴探索を同

時に行う学習アルゴリズムの提案。特徴探索に関して学習

モデル基準の枝刈り規則を設計することで大幅な探索コス

トの削減を実現。グラフ分類・回帰問題における非線形性

の影響を人口データセット及び実データセットにおける性

能比較より評価。

学会発表

学会名：　2018 International Workshop on New Frontiers in Convergence Science and Technology

発表年月日：　2018年11月9日

タイトル：　Jointly Learning Relevant Subgraph Patterns and Nonlinear Models of Their Indicators.

著者：　Shirakawa R, Yokoyama Y, Okazaki F, Takigawa I

内容：　生物・化学分野に対する情報科学技術の応用方法の提案。

　　　　化合物やRNA二次構造などグラフ構造で表現できるデー

タに対して、情報科学のグラフ分類・回帰問題を適応する

ことによる物性、活性予測を利用したスクリーニング応用

の期待。

学会発表

学会名：　ACML 2019 Workshop on Statistics & Machine Learning Researchers in Japan, (2019)

発表年月日：　2019年11月17日

タイトル：　Learning a Nonlinear Model of Subgraph Features Using Monte Carlo Tree Search.

著者：　Shirakawa R, Nakamura A, Kudo M

内容：　部分グラフ指示子を特徴量としたグラフ分類・回帰モデル

の学習における、識別部分グラフ探索の近似探索手法の提

案。既存探索手法に対してモンテカルロ木探索を利用した

近似探索手法を適用することで、探索コストの大幅な削減

を達成。加えて、汎化能力の向上による学習精度の向上も

達成。

学会発表

学会名：　第22回情報論的学習理論ワークショップ(IBIS2019)

発表年月日：　2019年11月21日

タイトル：　モンテカルロ木特徴探索に基づく非線形グラフ分類回帰.

著者：　白川　稜、中村　篤祥、工藤　峰一

内容：　部分グラフ指示子を特徴量としたグラフ分類・回帰モデル

の学習における、識別部分グラフ探索の近似探索手法の提

案。既存探索手法に対してモンテカルロ木探索を利用した

近似探索手法を適用することで、探索コストの大幅な削減

を達成。加えて、汎化能力の向上による学習精度の向上も

達成。

補助業務

＜ティーチャーアシスタント＞

授業：　情報学I―オブジェクト指向プログラミング

情報学II―パターン認識と機械学習

期間：　2018.4.6〜2018.4.18

　　　　2018.12.21〜2019.1.7

内容：　コーディング等の技術指導

　　　　グループワークまとめ及び技術指導

社会貢献活動

＜小学校の先生及び生徒に向けたプログラミング

教育の補助活動＞

内容：　来年度から始まるプログラミング教育に向けて小学生及び小学校教諭を対象とした補助活動。

　　　　・プログラミング的思考とプログラミング言語（スクラッチ）を授業する動画の作成・配信

　　　　・小学校教諭への意識調査アンケート、セミナーの実施

社会貢献活動

＜Fintech分野の大会への参加＞

大会：　Fintech Data Championship

主催：　株式会社POL

共催：　株式会社MILIZE

協力：　日本経済新聞社クロスメディア営業局

データ・教材協力：　True Data社、Aidemy社

協賛：　アセットマネジメントOne、ニッセイアセットマネジメン

ト・日本生命、三菱UFJ信託銀行

監修：　東京大学　松尾　豊　特任准教授

登壇：　京都大学　加藤 康之　特任教授

内容：　ITの金融分野への応用に関する金融業界及び経済の専門家

の方々との意見交換、ワークショップ、データ分析コンペ

ティション