
表 題 : gBoost: a mathematical programming approach to graph classification and regression

雑誌名 : *Machine Learning*, **75**, 1(2009), 69-89.

著 者 : Hiroto Saigo, Sebastian Nowozin, Tadashi Kadowaki, Taku Kudo, Koji Tsuda

日 時 : 2019 年 7 月 22 日 (月) 15 時 00 分

場 所 : 10-21

発表者 : 白川 稜

概 要 : グラフマイニング手法は、分類または回帰問題の特徴として利用することができる頻出な部分グラフパターンを列挙する。しかしながら、頻出な部分グラフパターンは必ずしも学習に有用であるとは限らない。ここでは、逐次的に有用なパターンを収集する数理計画ブースティング手法 (gBoost) を提案する。既存のブースティング手法である AdaBoost と比較して、gBoost はより少ないイテレーションで予測ルールを構築することができる。ブースティング手法をグラフデータに適用するため、分枝限定法を用いたパターン探索アルゴリズムを DFS コード木に基づき設計する。構築された探索空間は計算時間を最小化するため、後のイテレーションで再利用される。出力ラベルは探索空間を枝刈りするための情報源として利用されるため、本手法では頻出部分構造マイニングによる単純な方法よりも効率的な学習が可能である。加えて、数理計画問題を設計することで、パターン探索アルゴリズムの修正なしに広い範囲の機械学習の問題を解くことができる。

— Seven questions to be answered —

- Q.1 この論文（研究）の扱っているテーマは何か？
部分グラフ支持子を特徴量としたグラフ分類回帰問題。
- Q.2 何故、この論文（研究）を取り上げたか？ また、自分の研究との関連についても述べよ。
自身の研究のベースとなる手法であるため。
- Q.3 これまでこのテーマに関する方法論の問題点は何か？
全部分グラフの総数はグラフサイズに対して指数関数的に増加するため、取り扱いが困難である。
- Q.4 提案する方法論の独自性は何処にあり、どの点で有利と著者らは言っているか（と思うか）？ また、どの点は不十分あるいは劣っているか？
グラフ分類回帰問題を線形計画問題として定式化し、列生成法のアイデアを元にブースティング手法を構築する。加えて特徴探索において、bound の取り方を与え branch&bound 法を用いることで、従来では扱うことのできない数の特徴を考慮することが可能となる。
- Q.5 発表者の視点でこの論文を評価した場合、どこに利点があると思うか？
従来では扱うことのできない数の特徴を考慮できる点とそれによる精度の向上。
- Q.6 発表者の視点でこの論文を評価した場合、どういう点が不十分であると思うか？
- Q.7 この論文（研究）を発展させる方向はどの辺にあるか？（できれば具体的なアイデアを述べよ）
探索の効率化.branch&bound 法を用いた上でも探索空間が膨大であるため、厳密に探索を行うには相応のコストを要する。従って、MCTS や A*を用いた低コスト探索アルゴリズムの応用が有用であると考え。

1. はじめに

2. 実験

2.1. 結果

3. 考察

4. 今後の課題

文献

- [1] T. Cazenave and N. Jouandeau, “A parallel Monte-Carlo tree search algorithm,” in Proc. Comput. Games , Beijing, China, 2008, pp. 72 - 80
- [2] M. Enzenberger and M. Müller, “A lock-free multithreaded Monte- Carlo tree search algorithm,” in Proc. Adv. Comput. Games , Pamplona, Spain, 2010, vol. 6048, pp. 14 - 20.