

06

グラフを用いてデータの関係を素早く発見

～ L_1 -グラフ構成法のための高速化アルゴリズム～



どんな研究

ICTの進歩に伴い大規模なデータが利用可能になりつつありますが、データ間の疎なつながりを表現できるグラフ構造の研究が盛んに行われています。本研究では代表的なグラフ構造である L_1 -グラフを大規模な多次元データから高速に計算するためのアルゴリズムの開発に取り組みました。

どこが凄い

グラフにおいてノードがエッジを持つ条件を解析することにより、高速なアルゴリズムを確立しました。従来のアルゴリズムよりも大規模なデータを短時間で処理することが可能となったため、データに潜んだ関係を素早く見つけ出すことができるようになりました。

めざす未来

本技術は従来技術では考えられなかった規模のデータが分析可能であり、ビッグデータに基づいた推薦・予測・理解などに活用できます。IoTにより将来的に世界の至る所からデータが得られるようになりますが、それらを瞬時に解析・利用できる社会を実現することができます。

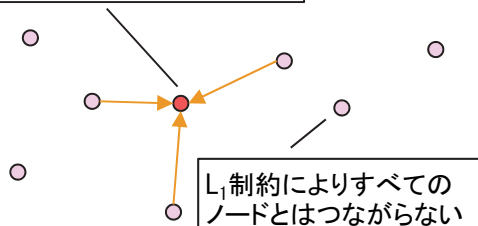
背景： L_1 -グラフ

- L_1 制約によりグラフ構造を計算しデータ間の疎な関係を適切に表現

目的関数を最小化する回帰分析からエッジの重みを計算

$$\min_{w_p} \frac{1}{2M} \|x_p - w_p X\|_2^2 + \lambda \|w_p\|_1$$

w_p : エッジの重み
 M : データの次元数
 x_p : データポイント
 X : データ行列
 λ : 正則化パラメータ



課題：高い計算コスト

- 回帰分析の係数をすべてのノードに対して計算しなければならない

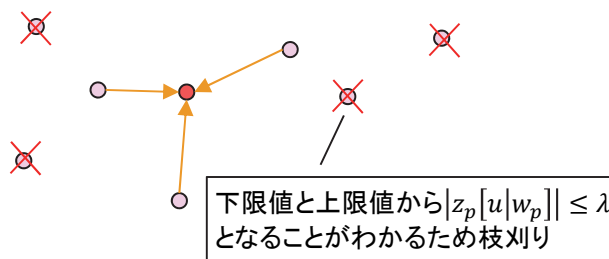
$$w_p[u] \leftarrow \begin{cases} z_p[u|w_p] - \lambda & z_p[u|w_p] > 0 \text{ かつ } |z_p[u|w_p]| > \lambda \\ z_p[u|w_p] + \lambda & z_p[u|w_p] < 0 \text{ かつ } |z_p[u|w_p]| > \lambda \\ 0 & |z_p[u|w_p]| \leq \lambda \end{cases}$$

ここで $z_p[u|w_p] = \frac{1}{M} \sum_{i=1}^M x_u[i](x_p[i] - \bar{x}_p[i]u)$

重み $w_p[u]$ を計算するためにはパラメータ $z_p[u|w_p]$ が必要になる

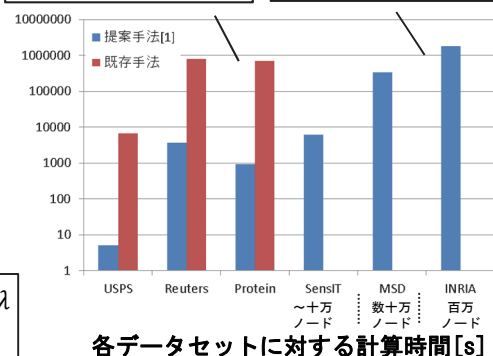
手法：ノードの枝刈り

- 特異値分解からパラメータ $z_p[u|w_p]$ の下限値 $\underline{z}_p[u|w_p]$ と上限値 $\bar{z}_p[u|w_p]$ を高速に計算



枝刈りしない手法より最大1300倍高速

提案手法は大規模データへ対応可能



関連文献

[1] Y. Fujiwara, Y. Ida, J. Arai, M. Nishimura, S. Iwamura, "Fast Algorithm for the Lasso based L_1 -Graph Construction," *PVLDB*, 10(3), pp. 229-240, 2016

連絡先

藤原 靖宏 (Yasuhiro Fujiwara) ソフトウェアイノベーションセンター
 E-mail: fujiwara.yasuhiro(at)lab.ntt.co.jp