

研究報告書

「高精度でスケーラブルな多項関係予測の実現」

研究タイプ: 通常型

研究期間: 平成 22 年 10 月～平成 26 年 3 月

研究者: 鹿島 久嗣

1. 研究のねらい

データに潜む規則性を発見し、これを新たな知見の創出や意思決定に役立てるためのデータ解析技術は今後の情報処理技術の要であり、その研究は様々な研究分野において精力的に進められてきた。そのなかでも「これから何が起こるのか」を予測する予測的なデータ解析はより直接的に競争力のある意思決定に結び付くため、とりわけその重要性は高い。従来、予測的なデータ解析は個々のデータを対象とした予測をその主な対象としてきたが、近年ではより高度な予測を目指したデータ間の高次関係の予測の重要性が認識され始めてきている。しかしながら、これを高精度かつ効率良く実現するデータ解析が実現しているとは言い難いのが現状である。そこで本研究では、汎用、高精度かつスケーラブルなデータ間の高次関係予測機構の実現を目指す。また開発した技術を様々な実問題に応用することで実社会へのインパクトを目指す。

2. 研究成果

(1) 概要

多項関係の解析手法として広く用いられている多次元配列のテンソル分解に着目し、テンソル分解を用いた高精度でスケーラブルな多項関係の予測を実現するための3つの課題、すなわち、予測性能の悪化等を引き起こすデータの疎性の課題と定式化の準最適性の課題ならびに予測時の非効率性という課題を解決するための様々な技術を開発し、学術的にも高い評価を受けた。また、企業の販路開拓や連携をサポートするための企業間の取引ネットワークを対象とした関係予測や、実際のデータに合わせたモデル探索を実現するためのデータ解析コンペティションの開催といった試みを行い、実世界応用への手ごたえを得た。

(2) 詳細

研究テーマ A 「データ間の高次関係予測のための基礎技術の開発」

多項関係の解析手法として広く用いられている多次元配列のテンソル分解に着目し、テンソル分解を用いた高精度でスケーラブルな多項関係の予測を実現するための3つの課題である (i) データの疎性: 疎データに対する予測性能の悪化、(ii) 準最適性: 解の最適性の欠如による予測性能の悪化、(iii) スケーラビリティ: 予測時の非効率性の問題 を解決するための技術を開発した。

(i) データの疎性を解決する情報統合予測技術

実世界の関係データは極めて疎であり、とくに3項以上の関係の場合には深刻な予測精度の低下を引き起こす。データ過疎による関係予測精度の低下に対処するため、関係データの他にデータ間の類似度などの利用可能な補助情報を用いて予測精度を補完するためのテン

ソル分解の新たな枠組みを開発した(文献 1)。この研究は機械学習・データマイニングの国際会議 ECML PKDD 2011 において(共著者が学生であったため)最優秀学生論文賞(Best Student Paper Award)を受賞し、トップ論文誌への掲載を推薦されるなどの高い評価を受けた。

(ii) 最適性を保証する定式化

多くのテンソル分解の定式化は最適性が保証されないものであり、多数存在する局所解に陥ることで予測精度が低下してしまうという問題がある。この問題に対処するためテンソル分解を最適解が保証される凸最適化問題として定式化するとともにその予測性能に関する理論的解析を行った(文献 2)。2項関係よりも高次の関係の予測における予測精度はこれまでほとんど何もわかっていない問題であり、この問題に対する理論的な見通しを与えた意義は大きい。さらに、高次関係を高精度で効率よく予測できる新しい方法として、低次元埋め込みを利用した方法を新たに開発した。この方法は固有値問題を繰り返し解くことによって(準)最適解を得る従来法とは異なり、固有値問題をただ一回解くだけで最適解が求めることのできる非常に効率的な方法である(文献 3)。

(iii) 関係予測時の高速化技術

高次関係のモデル化の効率化に加えて、高次関係予測のスケーラビリティに関する新たな課題として、テンソル分解に基づく予測時のスケーラビリティについての課題を提起した。データは疎であっても、予測時には補完された密な多次元配列から所望の要素を効率よく特定する必要があるが、素朴な探索では非常に膨大な時間を要してしまう。この問題意識は応用上きわめて重要でありながら、これまで気づかれていなかった課題であり、我々はオブジェクトのクラスタリングを利用してこれを極めて効率的に行う方法(現在投稿中)を提案した。この研究によって、テンソル分解に基づくモデル化から予測までの一連の流れの高精度高効率化が実現したことの意義は大きい。

(iv) その他(モデリング技術)

時間とともに構造が変化するネットワークにおける、時間をまたいだ関係の予測という、関係予測の新しい問題を提案し、これを効率的に解くことのできる解法を示した(文献 4)。他、様々なオブジェクトの参加する複数種類の関係を同時に予測する手法や、複数の人間によって作成された関係データセットから関係予測モデルを構築するための手法等、現実に表れる様々な問題設定に対応したモデリング手法を開発した。

研究テーマ B 「関係予測技術の実世界応用」

関係予測の技術を用いた先進的な実世界応用を行った。特に (i) 企業間の取引ネットワークを対象とした予測や、(ii) 関係予測問題自体をデータ解析コンペティションの枠組みを利用して解く試みを行った。

(i) 企業間ネットワーク解析による経営意思決定サポート

中小企業の販路開拓や企業間連携をサポートすることを目的とした「企業間つながり検索システム(SMEET)」を開発する東京大学 坂田一郎教授らのグループと協力し、関係予測の技術を利用することで企業間の「うまくいく」取引を予測することで取引先を推薦する試みを行った。ベンチマークデータを用いた実験では、関東甲信越における繊維、製材、出版など 14 業

種 3 万社、約 3.5 万の取引関係から約 8 割の予測正解率を実現した(文献 5)。

(ii) コンペティションを利用した関係予測モデリング

実際のデータ解析においては、しばしば結果を大きく左右するのはデータ固有の性質にあった手法の選択とデータの特徴をうまく捉えたヒューリスティクスであり、必ずしも洗練された手法が全ての場合でうまくいくことが保証されないため、データに合ったモデルを人手で広範囲に探索する必要がある。この「モデリングの労働集約性」を解決するための試みとしてクラウドソーシングを利用した関係予測の実験に取り組んだ。我々はデータ解析のクラウドソーシングの形式として現在のところ最有力であるデータ解析コンペティションに着目し、関係予測モデル構築を題材としたコンペティションを開催した。結果として得られたモデルの性能は我々の初期分析結果を大きく上回るものであり、また参加者から得られる複数の予測を機械学習によって上位統合することでさらなる予測精度向上が見込めるなどの数々の知見を得た。

3. 今後の展開

クラウドソーシングを利用して多くのデータサイエンティストを巻き込んだ関係データの解析は、本研究で開発した汎用的な方法論と実際のデータ解析をつなぐための枠組みとして有望視しており、その仕組みづくり等の研究を行っていく予定である。

4. 評価

(1) 自己評価

当初掲げていた研究のねらいのうち、疎なデータから多数のオブジェクト間関係を高精度・高速に予測するという要素技術的な課題については概ね解決できた。その成果は権威ある国際会議における採択やいくつかの受賞という形で高い学術的評価を得た。

一方で、開発した要素技術の統合によるツール化といったところまでには至らなかった。また応用面については様々な可能性を模索し、いくつかの実応用に取り組んだものの大きなインパクトをもたらすまでには至らなかった。しかしながら、応用への試行錯誤のなか、汎用的な方法論と個別の実世界データをつなぐための枠組みとして、多数のデータサイエンティストを構成要素とする新たなデータ解析の方法論の着想に到達した。

(2) 研究総括評価(本研究課題について、研究期間中に実施された、年2回の領域会議での評価フィードバックを踏まえつつ、以下の通り、事後評価を行った)。

高精度でスケーラブルな高次関係予測手法の開発と応用の研究である。

さきがけ期間を通じ、「関係データ」に焦点を当て、主にテンソル分解と呼ばれる関係のモデリング技術に着目することで、大規模な関係データに対して高速かつ高精度で予測を行うことのできる手法を開発している。

当初目標のうち、要素技術の統合によるツール化、実応用において大きなインパクトをもたらすまでには至っていないが、疎なデータから多数のオブジェクト間関係を高精度・高速に予測するという要素技術的な課題についてはほぼ解決でき、国際会議における採択や受賞という形で高い学術的評価を得ている。この点については、評価できる。

5. 主な研究成果リスト



(1)論文(原著論文)発表

1. Atsuhiko Narita, Kohei Hayashi, Ryota Tomioka, and Hisashi Kashima. Tensor Factorization Using Auxiliary Information. *Data Mining and Knowledge Discovery*. 2012, Vol.25, No.2, pp.298-324.
2. Ryota Tomioka, Taiji Suzuki, Kohei Hayashi, and Hisashi Kashima. Statistical Performance of Convex Tensor Decomposition. *Advances in Neural Information Processing Systems (NIPS)*. 2011, pp.972-980..
3. Nozomi Nori, Danushka Bollegala, and Hisashi Kashima. Multinomial Relation Prediction in Social Data: A Dimension Reduction Approach. *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI)*. 2012, pp.115-121.
4. Satoshi Oyama, Kohei Hayashi, and Hisashi Kashima. Cross-temporal Link Prediction. *Proceedings of the 11th IEEE International Conference on Data Mining (ICDM)*, 2011, pp.1188-1193.
5. Junichiro Mori, Yuya Kajikawa, Hisashi Kashima, and Ichiro Sakata. Machine Learning Approach for Finding Business Partners and Building Reciprocal Relationships. *Expert Systems with Applications*. 2012, Vol.39, No.12, pp.10402-10407

(2)特許出願

研究期間累積件数:0件

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

- 2011 情報処理学会 平成 23 年度 山下記念研究賞
- 2011 ECML PKDD Best Student Paper Award (指導および共著)
- 2012 マイクロソフトリサーチ日本情報学研究賞
- 2013 船井情報科学振興財団 船井学術賞
- 2013 PAKDD Best Student Paper Runner Up Award (指導および共著)
- 2013 第 5 回データ工学と情報マネジメントに関するフォーラム(DEIM2013)優秀論文賞
- 2013 人工知能学会論文賞