

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 14 日現在

機関番号：14301

研究種目：基盤研究(B) (一般)

研究期間：2014～2016

課題番号：26280112

研究課題名(和文)大規模グラフストリームからのリアルタイム情報抽出基盤

研究課題名(英文)Infrastructure for real-time information extraction from large-scale graph streams

研究代表者

田島 敬史(Tajima, Keishi)

京都大学・情報学研究科・教授

研究者番号：60283876

交付決定額(研究期間全体)：(直接経費) 9,800,000円

研究成果の概要(和文)：本研究課題では、インターネット上のユーザの活動の時間的変化、ソーシャルネットワーク上の情報の流れの変化、交通システム上の交通流の時間的変化などの、グラフ構造を持つデータの時間発展をリアルタイムに解析するための基盤を確立することを目的として、これらのグラフデータの時間変化を高速かつ正確にモデル化する技術と、そのモデルに基づいて今後の変化を予測する技術を開発した。より具体的には、インターネット上での対立関係にある語の出現頻度の変化をモデル化し予測を行う技術や、ソーシャルネットワークのグラフ構造上に新規ノードが追加された時に、今後グラフ構造がどのように変化するかを予測する技術などを開発した。

研究成果の概要(英文)：The goal of this project is to establish the infrastructure for real-time analysis of temporal evolution of graph-structured data, such as the change of user activities on the Internet along the time, the temporal change of information flow on social networks, and the temporal change of traffic flow on transportation networks. To achieve this goal, we developed several methods for highly-efficient and accurate modelling of temporal evolution of these kinds of data, and also several methods for prediction of their future changes based on these modelling methods. For example, we developed a method of modelling and predicting the temporal change of frequencies of appearances of some words that are conflicting with each other, and a method of predicting structural changes of social networks that will happen after appearances of some new nodes.

研究分野：計算機科学

キーワード：Webマイニング グラフマイニング ソーシャルネットワーク 時間発展グラフ モデル化 予測技術

1. 研究開始当初の背景

近年、Facebook や Twitter などのソーシャルネットワーク上では大量の情報が流通し、かつ、その情報の流れは刻々と変化している。また、GPS 付き携帯端末や高度交通システムからは、巨大な交通ネットワーク上の人や物の流れに関する大規模データがリアルタイムに収集可能となっている。これらのデータは、時間とともに変化する時間発展グラフデータとして表現可能である。そして、これらの時間発展グラフデータのリアルタイム解析技術が社会的に重要となっている。例えば、災害時においては、Twitter 上の情報流の監視による、デマの大規模伝播の早期発見や、緊急性の高い情報の検出は重要な社会的課題である。また、交通ネットワーク上の情報流の解析技術も、交通の省エネルギー化や新型の位置依存型情報提供サービス産業の創造などの社会的影響力の大きい分野への応用が見込まれる。

2. 研究の目的

そこで、本研究では、大規模グラフストリームのリアルタイム解析処理のための基盤技術の開発を目的として研究を行った。特に、グラフストリームの時間的変化のモデル化技術、そのモデル化に基づく予測技術、および、それらの技術の様々な問題への応用手法に関する研究を行った。

3. 研究の方法

本研究の特徴は、グラフの時間的変化の情報に着目することである。従来より、グラフからの構造発見や、グラフ中の重要ノードの発見、グラフ構造の今後の変化を予測する研究などは数多くあったが、これまでの研究は基本的に現時点の瞬間のグラフの構造を用いており、グラフの時間的変化の様子の情報を用いていない。そこで、本研究ではそのような時間的変化の情報を利用することで、より精度の高い解析を実現することを目的として研究を行った。また、時間的変化の情報として、グラフ構造の変化だけでなく、グラフ上を流れる情報流の時間的変化も考慮することで、これまで実現されていなかったような新しい解析をも実現することも目的とした。

4. 研究成果

(1) グラフ上の情報の流れの時間的変化のモデル化に関する研究

本研究のもっとも重要な成果の一つは、グラフ上の情報の流れの時間的変化のモデルをリアルタイムに求める技術に関する一連の手法の開発である。

そのような技術に関する一つ目の成果は、

web 上の様々なページへの様々なユーザによるアクセスや、様々な感染症の同時期における症例数の増減などのように、複数のイベントストリームが同時進行で進んでいく様子を表すデータから、特徴的なパターンやパターンの変化、外れ値などを効率的に発見する手法群の開発である。様々な実データに基づく実験により、本研究で開発した手法が、既存手法よりもはるかに高効率で、かつ、より有効なパターンを発見できることを確認した。

また、グラフ上の情報流時間的変化モデル化技術に関する二つ目の成果として、Web 上で同時に並行して進行する多数のユーザの活動パターンを Web 上での注目を求めて争う行動であり、自然界で食物を求めて争う動物の生態系とよく似た競争原理によって支配されていると捕らえることで、各活動の周期性や複数の活動間の排反性を表現できる、非線形モデルを開発した。

また、グラフ上の情報の流れの時間的変化を予測する技術の応用の一つとして、ソーシャルネットワークサービス上での情報転送サービスの最適化を行うシステムを開発した。ソーシャルネットワーク上には、自身が新たな情報を発信するのではなく、他の情報源から大量の情報を収集し、その中から有用と思われるもののみを選別し、これらを転送することをサービスとしているノードが存在する。このようなサービスにおいては、より有用な情報を、その情報の到着からできるだけ早く転送し、かつ、情報過多を防ぐため、一定時間内に転送する情報量を一定量以内に抑える必要がある。

単純に、総量を一定量以内に抑えながら、より有用な情報を提供するのであれば、多くの情報を見てから転送する情報を選択する（例えば、一日の間の全ての情報を見てから、それらの上位から総量の上限に達するまで情報を選択していく）方が、より良い情報を選択できる。しかし、それでは、各情報の到着から転送までの時間は長くなってしまふ。反対に、情報到着後すぐに転送をするようにすると、後により良い情報が来ても情報量の制限のために転送できないという場合が起きうるといふ、質と早さのトレードオフが生じる。そこで、過去のメッセージ流の時間変化の情報から今後のメッセージの到着数を予測し、この予測に基づいて適切な情報転送を行うシステムを開発した。

(2) グラフの時間的変化のモデル化に関する研究

本研究のもう一つの重要な成果は、グラフ構造自身の時間的変化のモデル化技術に関する研究である。

そのような技術に関する一つ目の成果として、グラフの時間発展を解析するためのテンソル解析技術を確立した。グラフデータはテンソルによって表現することができ

る。そこで、本研究ではグラフデータをテンソルにより表現し、非線形テンソル解析手法を用いて解析する手法を考案した。この手法では、そのグラフ構造の時間発展を非線形微分方程式によってモデル学習し、予測することを可能とする。

また、二つ目の成果として、グラフストリームの今後の時間変化をリアルタイムに予測する手法を開発した。本手法は大規模なデータストリームの中から重要な特徴や潜在的なトレンドをリアルタイムに検出し、長期的かつ継続的に将来のデータ予測を行う。

三つ目の成果としては、グラフストリームの中でも、特にソーシャルネットワークのグラフに関して、グラフ構造の時間発展を予測するための手法を開発した。ソーシャルネットワークグラフの時間的変化の予測においては、グラフへの有用なノードの新規出現を検出し、そのような新規有用ノードに、今後、どのようなリンクがはらされるかを予測することが重要となる。本手法では、与えられたソーシャルネットワークグラフの過去の有用ノード新規出現において重要な役割を果たしていたノードを発見することによって、そのような変化を予測する。

本研究では、そのような重要な役割を果たすノードとして、二種類のノードを考えた。一つ目は、ソーシャルネットワーク上において、そのノードがある新しいノードへリンクをはると多くのノードがそれを真似して同じノードへリンクをはるという傾向があるようなノードである。このようなノードは、あるノードへ向けてリンクをはるという行為がグラフ上を伝搬していく際のハブの役割を果たしている。よって、このようなノードを発見することにより、新しく追加されたノードに対して、今後どのようなリンクがはられるかを予測することができる。

ソーシャルグラフの時間発展に重要な役割を果たす二種類目のノードは、他のノードへの強い影響力は持たず、リンク伝搬のハブとなるわけではないが、有用な新規ノードの早期発見には優れており、新しい有用なノードにいち早くリンクをはることが多く、そのため、多くのノードからリンクされるノードへ早期にリンクする傾向を持つノードである。このようなノードを発見し、グラフ上に新たに追加されたノードがこのようなノードからリンクされているかを見ることで、その新しいノードへ、今後、多くのリンクが張られるかを予想することができる。

本研究では、まず、これら二種類のノードの各々を用いてソーシャルネットワークのグラフ構造の今後の変化を予測する手法を開発した。さらに、これら二種類のノードに基づく予測は相補的な性質を持っていることから、これら二種類の両者を組み合わせる

ことで、より予測精度を高める手法を開発した。

5. 主な発表論文等

[雑誌論文] (計 7 件)

1. Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, Lei Li, Christos Faloutsos, “Non-linear Dynamics of Information Diffusion in Social Networks”, ACM Transactions on the Web (TWEB), Vol.11, No.2, pp.11:1-11:40, 2017, 査読有
2. Yasuko Matsubara, Yasushi Sakurai, Christos Faloutsos, “Ecosystem on the Web: Non-linear Mining and Forecasting of Co-evolving Online Activities”, World Wide Web, Vol.20, No.3, pp.439-465, Springer, 2017, 査読有
3. 松原靖子, 櫻井保志, Christos Faloutsos, “生態系モデルに基づくオンライン活動データの非線形解析”, 電気情報通信学会論文誌 D, Vol.J100-D, No.4, pp.457-471, 2017, 査読有
4. 松原靖子, 櫻井保志, “大規模データストリームのリアルタイム予測”, 情報処理学会論文誌：データベース, Vol.9, No.4, pp.32-45, 2016, 査読有
5. 松原靖子, 櫻井保志, Willem G. van Panhuis, Christos Faloutsos, “大規模疫病データのための非線形モデル解析”, 情報処理学会論文誌：データベース, Vol.9, No.4, pp.17-31, 2016, 査読有
6. Thinh Minh Do, Yasuko Matsubara, Yasushi Sakurai, “Non-linear Time-series Analysis of Social Influence”, Journal of Information Processing, Vol.24, No.6, pp.937-945, 2016, 査読有
7. 本田 崇人, 松原 靖子, 根山 亮, 櫻井 保志, “車両走行センサデータからの自動パターン検出”, 情報処理学会論文誌：データベース, Vol.9, No.3, pp.1-13, 2016, 査読有

[学会発表] (計 12 件)

1. Yasushi Sakurai, Yasuko Matsubara, Christos Faloutsos,

- “Smart Analytics for Big Time-series Data”,
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Tutorial, 2017年8月13日, Halifax (Canada), 査読有
2. Tinh Minh Do, Yasuko Matsubara, Yasushi Sakurai,
“Automatic and Effective Mining of Coevolving Online Activities”,
Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), pp.233-246, 2017年05月23日, Jeju (South Korea), 査読有
 3. Daichi Imamori, Keishi Tajima,
“Predicting Popularity of Twitter Accounts through the Discovery of Link-Propagating Early Adopters”,
ACM Conference on Information and Knowledge Management (CIKM), pp.639-648, 2016年10月25日, Indianapolis (USA), 査読有
 4. Yasuko Matsubara, Yasushi Sakurai,
“Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences”,
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp.1045-1054, 2016年08月17日, San Francisco (USA), 査読有
 5. Yasuko Matsubara, Yasushi Sakurai,
Christos Faloutsos,
“Non-Linear Mining of Competing Local Activities”,
International World Wide Web Conference (WWW), pp.737-747, 2016年04月14日, Montreal (Canada), 査読有
 6. Yasushi Sakurai, Yasuko Matsubara,
Christos Faloutsos,
“Mining Big Time-series Data on the Web”,
International World Wide Web Conference (WWW), Tutorial, pp.1029-1032, 2016年04月12日, Montreal (Canada), 査読有
 7. Yasuko Matsubara, Yasushi Sakurai,
Christos Faloutsos,
“Mining and Forecasting of Big Time-series Data”,
ACM SIGMOD International Conference on Management of Data pp.919-922, 2015年05月31日, Melbourne (Australia), 査読有
 8. Yasuko Matsubara, Yasushi Sakurai,
Christos Faloutsos,
“The Web as a Jungle: Non-Linear Dynamical Systems for Co-evolving Online Activities”,
International World Wide Web Conference (WWW), pp.721-731, 2015年05月18日, Florence (Italy), 査読有
 9. Yasuko Matsubara, Yasushi Sakurai,
Naonori Ueda, Masatoshi Yoshikawa,
“Fast and Exact Monitoring of Co-evolving Data Streams”,
IEEE International Conference on Data Mining (ICDM), pp.390-399, 2014年12月14日, Shenzhen (China), 査読有
 10. Yasuko Matsubara, Yasushi Sakurai,
Willem G. van Panhuis, Christos Faloutsos,
“FUNNEL: Automatic Mining of Spatially Coevolving Epidemics”,
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp.105-114, 2014年08月24日, New York City (USA), 査読有
 11. Xiaoqi Zhao, Keishi Tajima,
“Online Retweet Recommendation with Item Count Limits”,
IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp.282-289, 2014年08月11日, Warsaw (Poland), 査読有
 12. Yasuko Matsubara, Yasushi Sakurai,
Christos Faloutsos,
“AutoPlait: Automatic Mining of Co-evolving Time Sequences”,
ACM SIGMOD Conference on Management of Data (SIGMOD), pp.193-204, 2014年06月22日, Snowbird (USA), 査読有

[産業財産権]

○出願状況 (計1件)

名称：時系列データストリームのための情報予測装置、方法、及びプログラム

発明者：松原 靖子, 櫻井 保志

権利者：同上

種類：特許

番号：特願 2016-138075

出願年月日：2016年07月12日

国内外の別：国内

6. 研究組織

(1) 研究代表者

田島 敬史 (T A J I M A, K e i s h i)

京都大学・情報学研究科・教授
研究者番号：60283876

(2) 研究分担者

櫻井 保志 (SAKURAI, Yasushi)

熊本大学・先端科学研究部(工)・教授
研究者番号：30466411