

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 20 日現在

機関番号：24403

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23760358

研究課題名(和文)トレンド可視化法へのグラフ統合とルールマイニング適用による時系列因果関係の表出化

研究課題名(英文)Time-series Expression by Graph Integration and Rule mining for Trend Visualization Method

研究代表者

佐賀 亮介 (SAGA, Ryosuke)

大阪府立大学・工学(系)研究科(研究院)・准教授

研究者番号：10509178

交付決定額(研究期間全体)：(直接経費) 2,300,000円、(間接経費) 690,000円

研究成果の概要(和文)：本研究は、時系列データを対象とした情報可視化に関するものである。時系列データのトレンドとデータ間の関係性の同時可視化法として、FACT-Graphが開発されてきた。しかし、FACT-Graphにおいて時系列の変動を考慮した抽出や意味づけができていない。そこで、全期間から一度に可視化するのではなく、いくつかの期間ごとにグラフを抽出し、それらを統合することによりグラフを抽出し、また、ネットワークの関係性の明示化などを試みた。

本研究の結果、構造変化点を考慮したデータ分割法による時系列情報可視化法や、グラフ統合法の提案、またネットワーク関係性の明示化についての報告を行った。

研究成果の概要(英文)：This research is information visualization for time-series data. FACT-Graph has been developed for visualizing both trend visualization and relationships between data. However, there are problems that the visualization results don't express the trend change or the meanings of relationships clearly. Therefore, we propose the method to extract graphs from the separated time-series data and integrate each graph into one graph, and to attach meanings to the relationships.

As results of this research, we reported time-series visualization method by structural change point, proposal of 3 types graph integration methods, and network visualization with consideration of direction and position.

研究分野：工学

科研費の分科・細目：電気電子工学・システム工学

キーワード：システム情報(知識)処理 トレンド分析 時系列情報可視化

1. 研究開始当初の背景

企業や組織において情報化が進むにつれ、蓄えられた情報の活用が盛んに行われている。その情報は POS データから、行動履歴といったログデータ、クレーム記録などのテキストデータなど多くある。これらのデータからデータウェアハウスを作成し、新たな知識発見を試みるデータマイニングやテキストマイニングが多くなされている。そのテキストマイニングでは、キーワード抽出、要約、情報編纂、可視化と幅広く研究されている。

特に可視化において、キーワードをノードとした共起グラフ(以降、共起グラフ)を利用した可視化技術が開発されており(文献1,2)、申請者もまた時系列テキストデータを対象とした FACT-Graph と呼ぶトレンド可視化手法を開発している。FACT-Graph は、分析期間におけるテキストデータからキーワードを抽出し、キーワードのクラス遷移情報とキーワードのトレンドをノードに表し、さらに共起関係の推移を利用して大域的トレンドの可視化を行っている(図1)。この FACT-Graph を用いることで、トピック単位でのマクロなトレンドの発見や、頻度は少ないかもしれないが重要なキーワードなどを発見することができる。

FACT-Graph の事例として、政治の動向や犯罪の傾向推移において、キーワードの抽出やきっかけとなる事象の抽出を実現している。さらに、効率的な分析のために、分析対象の背景を参照可能にし、分析の試行錯誤を可能にするテキストマイニング支援環境を開発している(図2)。

しかしながら、この FACT-Graph を用いた分析において以下の問題点に気がついた。

問題1. 分析期間における時系列の変動を考慮して、キーワードやリンクが抽出できているとは限らない

問題2. 共起のみを取り扱っているため、リンクの解釈が困難である

2. 研究の目的

問題1は、FACT-Graph の生成に関与している。FACT-Graph はキーワードを抽出する際に、数ヶ月・数年という長い分析期間におけるデータベースから TF-IDF アルゴリズムを用いてキーワードを抽出している。しかしながら、この方法では時系列を考慮したキーワード抽出が為されているとは言い難い。特に、ある時期に現れる局所的なキーワードは、全期間において比べると頻度が相対的に少ない場合が多く、そのキーワードの抽出は TF-IDF アルゴリズムでは困難である。

問題2は、FACT-Graph の分析時に関与する事項である。共起関係により表されるリンクを見ることで、慣用語・専門語のような語句の組合せやトピックの存在、キーワード間の依存関係などを把握することが可能である。特にトピック間をつなぐ“かけはし”と

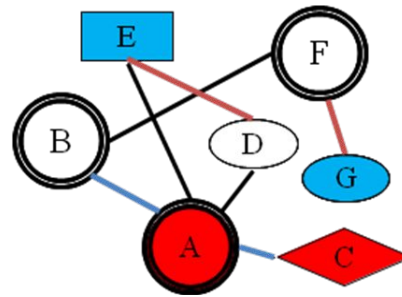


図1 FACT-Graph

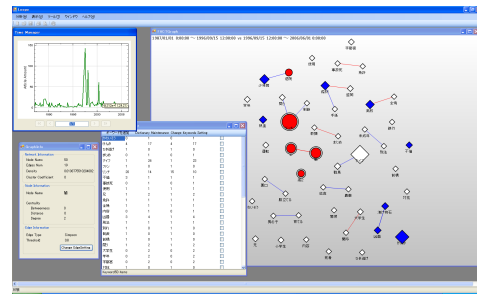


図2. FACT-Graph 分析ツール Loopo

なる語句が重要なキーワードになる可能性があるため、そのキーワードを把握するためにも、共起関係の識別は FACT-Graph の解釈において重要な意味を持つ。しかし、出力された FACT-Graph によっては、リンクの解釈に多大なコストを強いられる一方、FACT-Graph の解釈を誤ることも考えられる。

この問題をうけ、次のことを達成することを本研究での目的と設定した。

目的1. 時系列に沿って共起グラフを生成し、そのグラフを元にキーワードとリンクの抽出を行う

目的2. 共起グラフに方向性を付加することで、FACT-Graph を有向グラフに拡張し、キーワード間の関係表出化を行う

3. 研究の方法

まず、目的1: グラフをベースとした重要キーワードとリンクの抽出を遂行する。この目的を達成するために、時系列データのある単位期間(1日、1週間、1ヶ月など)ごとに区切り、共起グラフを生成する。このグラフは、各期に特化したものを抽出したものになる。続いて、そして各共起グラフについて合成方法は様々な方式が考えられるが、今回は、各共起グラフのキーワード群を集合とみなし、各共起グラフにおいて和集合・積演算により合成し、リンクは検定に基づいて合成を行う。具体的に、リンク間の類似度がある閾値を超えるかどうかという帰無仮説を立て、この帰無仮説を採択するか、棄却するかでリンクの有無を決定する。これにより、時

系列ごとに有効なキーワードやリンクのみを抽出することができる。最後に従来のクラス遷移分析により大域的なトレンドを付与する。

目的2: FACT-Graphの有向グラフによるキーワード間の方向性付与について遂行する。この目的を達成するために、目的1で作成した単位期間ごとの共起グラフのリンクを、Jaccard係数やSimpson係数ではなく、アソシエーションルールや系列データマイニングなどのルールマイニングにより有向性を保持したまま抽出する。これにより無向グラフであったFACT-Graphに有向性を付与できる。続いて、目的1と同様に各期におけるグラフを合成することを試みる。これらの有向グラフを、2者間における推移の判定を用いることで、総合的に統合する。この方式を用いることで、リンクの推移を統計的に尤もらしく統合することができる。

4. 研究成果

申請者は、まず分割によりキーワード抽出結果がどのように変わるかどうかに関して議論を行った。それにより、全体から上位n件抜き出したものと、分割したもとのキーワードを抽出した場合を比べた結果、分割したものが、より詳細な結果が得られ、時系列の分割による抽出の有用性が示された(研究成果1,)。このことから、時系列データの分割法として、構造変化点検出法である逐次確率比検定(図2)による、データを分割法を提案した(研究成果2, など)。また、分割したデータを可視化し、それらのグラフ結果をまとめる際に、クラス分け可能な可視化を行い、分析し、新たな知見を得ている(研究成果3, 4, 6)。リンクの表出化に関しては、ウェブアクセスログの結果を用いて図5のように表出化ができた。これによりアクセスの流れの時間的推移が把握できるようになった(研究成果 など)。

一方で、テキスト文書からリンクやノードを抽出し、要因モデルを作成する方法に関して、提案し、実験している(研究成果)。これにより、レビューコメントのようなテキスト文書を利用した要因モデルのグラフかと、そのグラフから要因モデルを作成可能となった。また、それらの要因モデルをグラフとみなし、その合成方法として、和演算的合成法、確率的合成法などを提案し、実験を行っている(研究成果)。また、その合成方法において、様々な合成方法に気付き、対比させることによる合成などを用いた比較分析法も提案している(研究成果5,)。これらの技術要素を組み合わせたグラフ合成法について、現在執筆中である。

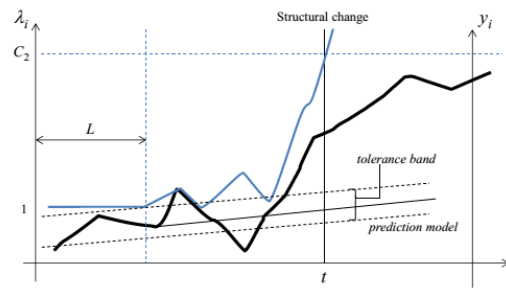


図3. 逐次確率比検定法による構造変化点検出

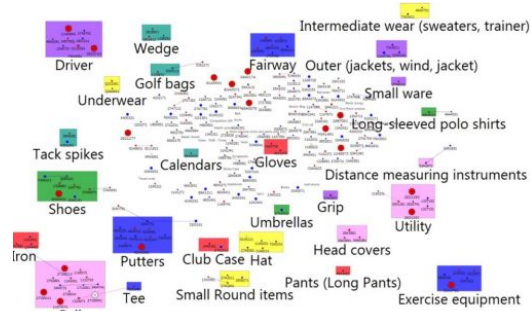


図4 クラス表示したFACT-Graph

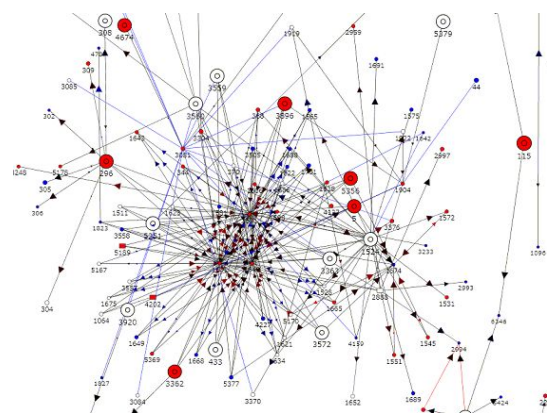


図5 関連ルールをベースにした可視化

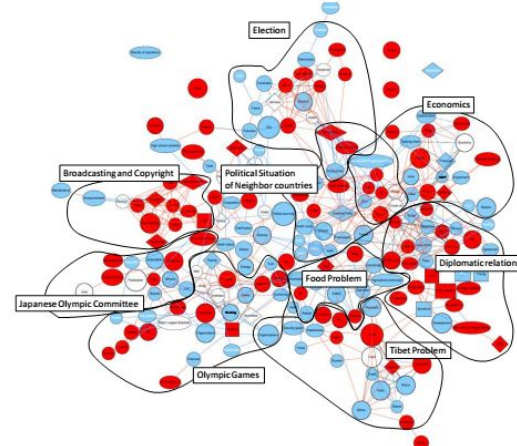


図6 新聞記事における比較合成可視化法

5. 主な発表論文等

〔雑誌論文〕(計6件)

1. Ryosuke Saga and Hiroshi Tsuji, Improved keyword extraction by separation into multiple document sets according to time series, Communications in Computer and Information Science, Vol. 374, 2013, pp.450-453,10.1007/978-3-642-39476-8_91 (査読有)
2. Ryosuke Saga, Naoki Kaisaku, and Hiroshi Tsuji, Relearning Process for SPRT in Structural Change Detection of Time-Series Data, Annals of Information Systems, Special Issue on Real World Data Mining Applications (To Appear)(査読有)
3. Ryosuke Saga, Mauricio Letelier, Naoki Kaisaku, Yukihiro Takayama, Hiroshi Tsuji, Visualization analysis of e-commerce site from web access log by FACT-Graph and sequential probability ratio test, Vol. 17, No. 2, pp. 145-155, 2013 (査読有)
4. Ryosuke Saga, Hiroshi Tsuji, Comparison Analysis for Text Data by Integrating Two FACT-Graphs, Intelligent Interactive Multimedia: Systems and Services Smart Innovation, Systems and Technologies, Vol. 14, pp 143-151, 2012 (査読有)
5. Ryosuke Saga, Tetsuya Fujita, Kodai Kitami, Kazunori Matsumoto: Improvement of Factor Model with Text Information Based on Factor Model Construction Process. Vol. 254, Intelligent Interactive Multimedia Systems and Services, pp. 222-230 (査読有)
6. Ryosuke Saga, Mauricio Letelier, Naoki Kaisaku, Yukihiro Takayama, and Hiroshi Tsuji: Knowledge Discovery in Web Access Log of E-commerce Site with FACT-Graph and Sequential Probability Ratio Test, Advances in Knowledge-Based and Intelligent Information and Engineering Systems, Vol. 243, pp. 2080-2089, 2012 (査読有)

〔学会発表〕(計10件)

Hiroshi Kobayashi, Ryosuke Saga, Finding Division Points for Time-Series Corpus Based on Topic Change, 16th International Conference on Human-Computer Interaction, 2014年6月22日~2014年6月27日, Crete, Greece

開作直樹, 佐賀亮介, 時系列構造変化点検出のための SPRT における重回帰モデルの再学習方法, 平成 26 年電気学会全国大会, 2014 年 3 月 18 日 ~ 2014 年 3 月 20 日, 愛媛県松山市
小林寛史, 佐賀亮介, 時系列データのキ

ーワード抽出とその評価手法の提案, 平成 26 年電気学会全国大会, 2014 年 3 月 18 日 ~ 2014 年 3 月 20 日, 愛媛県松山市
Ryosuke Saga and Rikuto Kunimoto, Integration of Path Models for Structure Equation Modeling by Use of Union Operation for Merging Multiple Viewpoints, 2013 International Symposium on Computational and Business Intelligence, 2013 年 8 月 24 日 ~ 2013 年 8 月 25 日, New Delhi, India

Rikuto Kunimoto and Ryosuke Saga, Path Model Integration Method for Structure Equation Modeling by OR and Probability Concepts, IEEE SMC 2013, 2013 年 10 月 13 日 ~ 2013 年 10 月 16 日, Manchester, UK.

Yukihiro Takayama, Ryosuke Saga, Takao Miyamoto, Detecting Outlier in Graph Structure Data Using Centrality, Asian Conference on Information Systems 2013, 2013 年 10 月 31 日 ~ 2013 年 11 月 2 日, Phuket, Thailand.

恵明, 佐賀亮介, 辻洋, 日中カスタマーレビューに基づいたブランド認知の差異発見, 平成 25 年電気学会情報システム研究会, 2013 年 9 月 10 日 ~ 2013 年 9 月 11 日, 福岡県福岡市

Ryosuke Saga, Proposal of Directed FACT-Graph Based on Association Rule, The first Asian conference on Information Systems, 2012 年 12 月 06 日 ~ 2012 年 12 月 8 日, Siem Reap, Cambodia

6. 研究組織

(1) 研究代表者

佐賀 亮介 (SAGA, Ryosuke)

大阪府立大学・工学研究科・准教授

研究者番号: 10509178