

機関番号：32714

研究種目：若手研究（B）

研究期間：2009～2010

課題番号：21760305

研究課題名（和文） 時系列データ可視化技術の有用性検証と分析方法論の開発

研究課題名（英文） Verification of the availability of visualization method for time-series data and development of analysis methodology

研究代表者

佐賀 亮介 (SAGA RYOUSUKE)

神奈川工科大学・情報学部・助教

研究者番号：10509178

研究成果の概要（和文）：

本研究では時系列データの可視化技術である FACT-Graph に対して有用性を検証し、またその分析方法の策定を目指している。FACT-Graph はトレンドの変化と関係性を組み合わせることでそのトレンドの背景にある情報を把握するための可視化手法である。この有用性を検証するために、さらに Web アクセスログデータや新聞記事に適用することで有用性を検証し、また新たな比較分析への発展を見せた。その際に閾値の設定法や分析法に対して知見を得て、分析に応用している。

研究成果の概要（英文）：

This research aims to verify the availability of FACT-Graph, which is one of visualization methods for time-series data, and find the effective analysis methodology. FCT-Graph makes visualization to recognize the background by combining the change of trends with relationships. To verify the availability, we analyze the trend of Web access log and newspaper articles by the analysis tools and learned a lot about the configuration of thresholds and the analysis methodology.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	2,700,000	810,000	3,510,000
2010年度	800,000	240,000	1,040,000
総計	3,500,000	1,050,000	4,550,000

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

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。テキストマイニングの分野では、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。

その中で、テキストデータを扱った場合、キーワード抽出は必須の技術であり、またそれらの関係性も昨今注目を浴びている重要な要素である。このテキストデータの情報可視化において、重要単語と関係性を考慮することは共起グラフなどで表現可能である。しかしながら、時系列テキストデータをどのように可視化するには、それにさらにトレンド情報を付与する必要がある。即ち、変動の可視化と集合の関係を時系列変化へと推し進

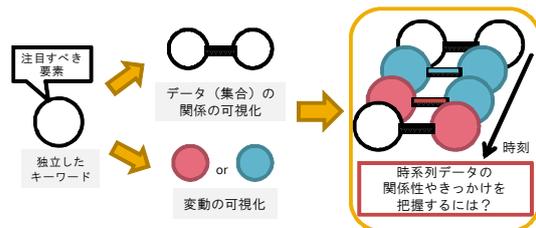


図1 時系列データ可視化における関係を考慮した時の課題

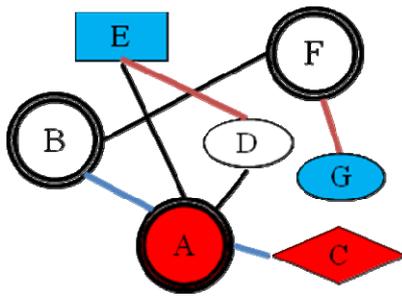


図2 FACT-Graph

め、その背景に何があるのかを可視化により把握することが求められている (図1)。

この背景のもと、時系列テキストデータにおいて、申請者は頻度と鮮度を元にしたトレンド分析手法を開発した。これを申請者はFACT-Graph (Frequency And Concurrence based Trend Graph)と呼んでいる(図2)。FACT-Graphでは、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、連続した期間におけるキーワードのクラス遷移情報と共起関係を利用して可視化を行っている。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語句や複数の語で構成される話題などを効率的に分析することができる。

しかしながら、FACT-Graphには、以下のような問題点がある。

(1) FACT-Graph を分析するのに十分な機能やインターフェイスを提供していない。ユーザの気づきや不要語の設定を反映させることが困難である

(2) 活用事例が少なく、有用性が十分でない。また、テキストデータだけを対象としているため、一般データへの有用性が確認されていない

(3) FACT-Graph の分析結果は分析者に依存する。FACT-Graph 生成に限っても分析期間やTF(Term Frequency)の閾値に関して分析者の主観的な経験に依存する

これらを解決する必要性が本研究の背景としてあった。

2. 研究の目的

上記のを受け、本研究では以下の目的を設定した。

(1) FACT-Graph 自体を直感的に操作し、ユーザの意図を FACT-Graph 自体に反映することが可能なインターフェイスを持つ、システムを作成する

(2) 時系列情報をもつ様々なテキストデータやログデータを FACT-Graph に出力し、有用性を検証する。また新たな活用事例を模索する

(3) FACT-Graph 出力の際に必要な閾値などの閾値の設定法、分析方法論を目的(1)、目的(2)などを経て知見を得る

この目的を達成することで分析の効率性を挙げる事が可能である。また FACT-Graph の有用性を様々な時系列データから検証することにより、手法の利用価値向上が得られると考えられる。また、分析ソフトウェアの提供と分析から得られた知見を再利用することによって、目的としている FACT-Graph をより幅広く利用可能にし、トレンド分析の間口を広くし、時系列分析が比較的容易にできることを期待している。

3. 研究の方法

まず、2009年度は目的(1)ユーザの意図をFACT-Graph 自体に反映することが可能なインターフェイスを持つ、システムの作成について遂行した。この基本アイデアに関しては2008年度の研究論文("頻度と鮮度に基づいた共起グラフの推移情報表示環境", 2008年)で提示しており、それを改良し、ソフトウェアとして完成させる。

続いてその後、そのソフトウェアなどを用いて、目的2: 時系列情報をもつ様々なテキストデータやログデータを用いたACT-Graphの有用性の検証を行う。アクセスログなどテキストデータではないものを分析するために、本来FACT-Graph で表示する。キーワードを、ログに記録している属性やメタデータに置き換える。その際、FACT-Graph の分析プロセスやFACT-Graph を出力するためのパラメータ(TF-IDFなどの閾値、ランキング)を記録しておく。

これらの目的2によって、得られた経験に基づき、パラメータの設定などの目的3: 分析方法論の策定を行う。

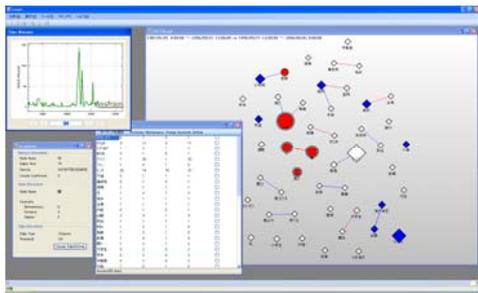


図 3. FACT-Graph 分析ツール Loopo

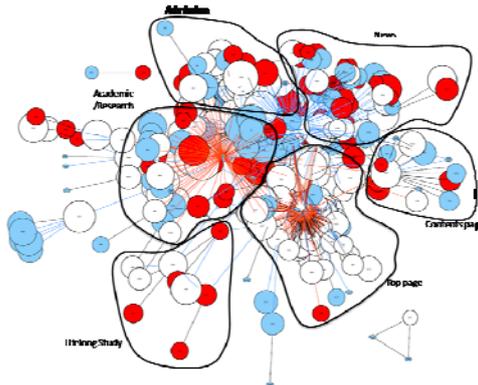


図 4 FACT-Graph を用いた Web アクセスログの可視化グラフ

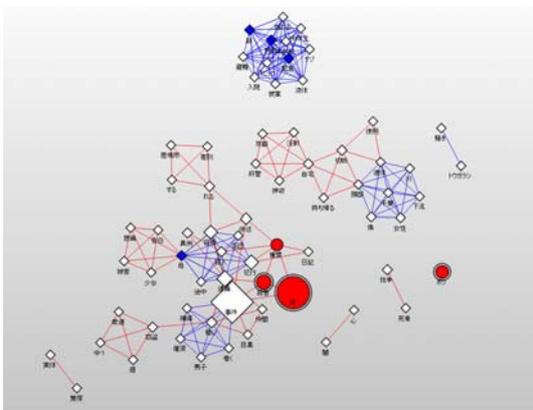


図 5 FACT-Graph を用いた犯罪記事の可視化グラフ

4. 研究成果

申請者は、まず Loopo と呼ばれる FACT-Graph を分析するためのソフトウェアを開発した。このソフトウェアの開発による効果は分析の効率性を挙げることである。このソフトウェアを開発する際、従来のテキストマイニングシステムと FACT-Graph の分析方法とを比較したとき、1) 分析者の気づきを反映しづらい、2) 情報の参照性が乏しい、3) 対話型インターフェイスの欠如といった3つの問題点があった。これらを解決するために、

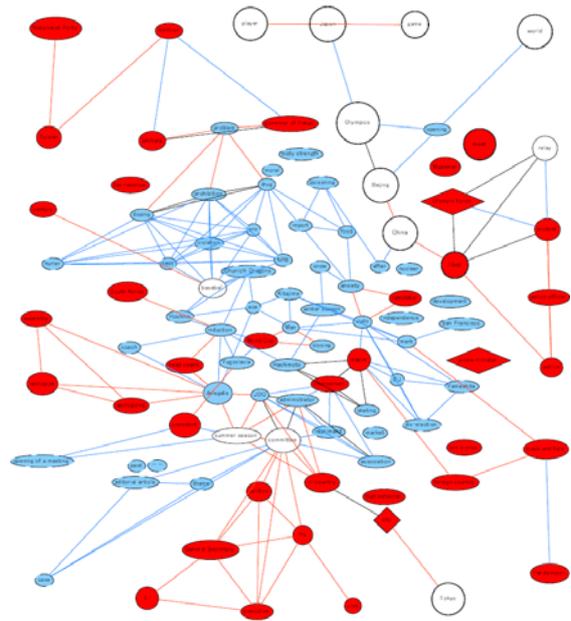


図 6 FACT-Graph による 2 者の社説の差異の可視化

FACT-Graph View、Time Manager、Keyword Manager、Graph Info と呼ばれる 4 つの GUI ベースのコンポーネントを実装した (図 3)。

続いて、Web アクセスログや新聞記事を用いた FACT-Graph の有用性検証を行った。Web アクセスログは大阪府立大学から貸与されたデータを用いて分析を行った。その結果、図 4 に示されるような結果が得られ、時期により Web サイトのアクセス状況がどのように変化しているか、またそれと関連しているページがどのようなものかを可視化することに成功した。また、新聞記事への適用は 20 年間における犯罪記事を対象にした新聞記事の見出しを分析した (図 5)。(その結果は開発ツールのケーススタディとして Artificial life and Robotics 2010 (AROB '10) に報告済みであり、Best Paper Awards を受賞した) また、その新聞記事分析において、各社の新聞の書き方や内容に差があることに気がついた。そこで、新聞記事間の比較分析ができないかと考え、クラス遷移分析を比較分析へと転用することで、FACT-Graph を比較分析へと応用することができた (図 6)。

これらの分析において、閾値の設定法、特にクラス遷移分析で用いる Term Frequency (TF) と Document Frequency (DF) の閾値を再考する際、表示対象の頻度に注目した。ここで、テーマに関連した単語や Web ページのトップページなど、何かしら関係が集中しやすい対象が存在するとき、Zipf の法則が成り立つことに気がついた。この Zipf の法則はパ

レートの法則とも言われ、20:80 の法則とも言われる。この 20:80 という値から、上位 20% を閾値としてキーワードのクラス遷移に用いる方法を考案した。また、一つに集中しにくい場合においては、適当な間隔にて分散することが考えられる。このことから、TF、DF という値それぞれに対して偏差平方和などを用いる方法で 2 値への離散化を行う方法で妥当な閾値を設定する方法を考案した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 6 件)

1. Ryosuke Saga, Seiko Takamizawa, Kodai Kitami, Hiroshi Tsuji, Kazunori Matsumoto, Comparison Analysis for Text Data by Using FACT-Graph, Lecture Notes in Computer Science, 査読有, Vol. 6772, 2011, pp.75-83
2. Ryosuke Saga, Hiroshi Tsuji, Takao Miyamoto, Kuniaki Tabata, Development and Case Study of Trend Analysis Software Based on FACT-Graph, Journal of Artificial Life and Robotics, 査読有, Vol. 15, 2010, pp. 234-238.
3. Ryosuke Saga, Hiroshi Tsuji, Kuniaki Tabata, Loopo: Integrated Text Miner for FactGraph-based Trend Analysis, Lecture Notes on Computer Science, 査読有, Vol. 5618, 2009, pp. 192-200

[学会発表] (計 12 件)

1. Ryosuke Saga, T. Miyamoto, Hiroshi Tsuji, Kazunori Matsumoto, Visualization of WebSite Usage with Consideration of Trend and Relationships, IADIS International Conference Information Systems 2011, 2011年3月12日, Hotel Palacio De Los Velada (スペイン)
2. Ryosuke Saga, S. Takamizawa, Hiroshi Tsuji, Kazunori Matsumoto, Application for Text-based comparison analysis using Frequency and Co-occurrence Trend-graph, IADIS International Conference Information Systems 2011, 2011年3月12日, Hotel Palacio De Los Velada (スペイン)
3. 高見沢 聖子, 佐賀 亮介, 辻 洋, 松本 一教, FACT-Graphを用いた社説の比較分析, 第73回情報処理学会全国大会, 2011年3月4日. 東京工業大学 (東京都)
4. Y. Hayashi, F. Masruri, Ryosuke Saga, Hiroshi Tsuji, Enhanced Visualization on

Preference Transition for Sales Records, IEEE International Conference on Industrial Informatics, 2010年7月14日, Osaka University Nakanoshima Center (Osaka)

5. 沓澤 将, 佐賀 亮介, 北見 孝大, ネットワーク分析を用いたゲームソフトの購入要因分析, 2010年電子情報通信学会総合大会, 2010年3月17日, 東北大学 (仙台市)
6. Zhongqi Sheng, Marie Sano, Yoshihiro Hayashi, Hiroshi Tsuji, Ryosuke Saga, Visualization study of the relationships between responses in choice-type questionnaire, 2009 International Conference on Information Management, Innovation Management and Industrial Engineering (ICII2009), 2009年12月27日, Xi'an, China
7. Yoshihiro Hayashi, Ryosuke Saga, Hiroshi Tsuji, Visualizing Method based on Item Sales Records and its Experimentation, IEEE International Conference on Systems, Man & Cybernetics (SMC2009), 2009年10月11日, San Antonio, USA
8. 佐賀 亮介, 辻 洋, 田畑邦晃, FACT-Graphを用いたトレンド分析支援ツールの開発, 第8回情報科学技術フォーラム (FIT2009), 2009年9月4日, 東北工業大学 (仙台市)
9. Ryosuke Saga, Hiroshi Tsuji, Kuniaki Tabata, Loopo: Integrated Text Miner for FactGraph-based Trend Analysis, HCI International 2009, 2009年7月22日, San Diego, USA
10. Yoshihiro Hayashi, Ryosuke Saga, Hiroshi Tsuji, Competition state Visualization Technique for Sales and its Application, The 22th International Conference on Industrial Engineering & Other Applications of Applied Intelligent Systems IEA/AIE conference, 2009年6月27日, Tainan, Taiwan

6. 研究組織

(1)研究代表者

佐賀 亮介 (SAGA RYOUSUKE)
神奈川工科大学・情報学部・助教
研究者番号: 10509178