

ブロガーアーカイブに対する係り受け解析を用いた3次元話題探索システム

3D Visualization System for Exploring Topic Transition Based on Dependency Analysis for a Japanese Blog Archive

伊藤 正彦 吉永 直樹 豊田 正史 喜連川 優*

Summary. This paper proposes a 3D system for visualizing time sequential changes in topics of Blog contents that are archived and analyzed by dependency analysis for Japanese. We use TimeSlices to visualize the results of analysis that consists of set of targets and actions, and are summarized for each month. It enables us to explore changes in users' activity, interest, and thoughts for specified keywords.

1 はじめに

ブログ等のCGMの普及にともない、ユーザは簡単に自身の興味、行動、意見をウェブ上に反映することが可能になってきている。これら、時間・社会状況とともに変化するユーザの生の声は、マーケティング等の社会分析の観点から重要なデータとなってきている。これに伴い、特定のキーワードに対する評判情報、関連キーワードの時系列分布をブログから取得可能にする kizasi.jp¹ 等のCGM分析サービスが近年注目されるようになってきている。

提案するシステムでは、対象に関する行動の記述(例: 新型インフルエンザに感染する)の時系列推移を、月単位の粒度で分析した結果を検索し可視化を行う。可視化環境として、TimeSlice [1] を用いることで、対象と行動の関係をツリー表示し、それらの時間推移を3次元空間を用いて可視化することが可能になる。また、ノードの選択、展開といった操作を行うことで、誰が、どこで、何を、などの行動の詳細を探索可能にした。さらに、各ブログは簡易的に男女の性別を判定しており、ブログに記述された行動に関する男女比の推移が可視化可能になる。これらにより、ブログに記述された対象に対する行動、興味および感想の時系列推移を可視化・探索する話題探索システムを実現した。

2 ブロガーアーカイブに対する係り受け解析

著者等の研究室では、過去5年間、3億5000万エントリに及ぶ日本語ブロガーアーカイブの構築を進めてきている。さらに、蓄積された全ブログデータに対して日本語係り受け解析を行った²。提案シス

テムでは、係り受け解析結果のうち、名詞が動詞に係っているもののみを1つのイベント(対象+行動)として抽出し、その係り受けの意味的關係(行為の動作主や、対象など)を名詞に付く格助詞{を,と,に,が,で}で分類しデータベース化した。現在、約160万種類の名詞に対する約11万種類の動詞の記述をイベントとして利用可能である。データベースに対する検索キーワードとしては、名詞もしくは動詞を用いる。検索結果は、名詞+格助詞+動詞の3つ組みの集合で表される。動詞の代わりに形容詞を検索対象および検索キーワードに用いることも可能である。

3 ブログ上の話題探索システム

検索結果は、入力キーワードを中心としたツリー表現で TimeSlices [1] 上に可視化される(図1)。イベントは各月ごとに集約されており、出現頻度上位n個(nは任意)のみを表示可能である。また、出現頻度に応じてノードサイズを変更している。

各検索結果は年月情報を持っており、TimeSlice 上には指定した年月の結果のみが表示される。ユーザは TimeSlice を時間軸に沿って移動させることにより、出現頻度上位n個のイベントおよびそれぞれの頻度の時間変化をアニメーション表示させることが出来る。また、TimeSlice を追加することにより、時期による人々の動向の違いを比較出来る(図1(a))。

3.1 探索インターフェース

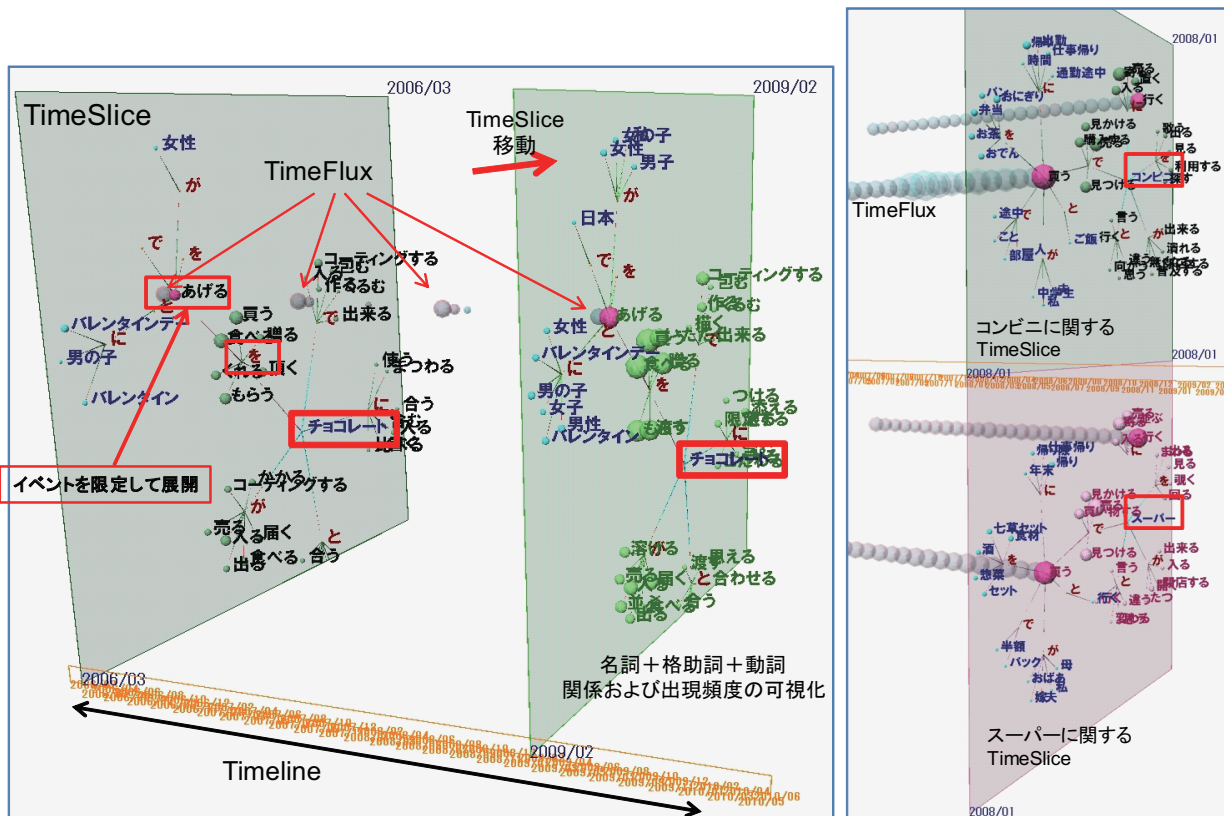
可視化結果から新たな検索要求を作成し、次々と情報探索を行うインターフェースを提供する。図1の例では、ノードを選択することで、イベントを限定し、それに対する付帯情報を検索し展開表示している。同様に、中心を動詞(形容詞)、他の名詞、もしくは他の3つ組みに移動するなどの操作を可能にする。

Copyright is held by the author(s).

* Masahiko ITOH, Naoki YOSHINAGA, Masashi TOYODA, and Masaru KITSUREGAWA, 東京大学生産技術研究所

¹ <http://kizasi.jp/>

² <http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/jdepp/>



(a) 「チョコレートをあげる」に関する傾向の変化

(b) 「コンビニ」と「スーパー」の比較

図 1. TimeSlices によるブログ上の話題変化の可視化・探索システム

3.2 時系列頻度変化の可視化

選択したノードに関して月ごとの頻度を球の列として可視化することで、どの時期にそのイベントが話題になったか、他のイベントとの傾向の違い、等が直感的に理解可能になる。この球の列を TimeFluxes と呼ぶ。色違いの 2 種の TimeFluxes を用いることにより男女ごとの変化を同時に可視化し比較することが可能になる

3.3 複数キーワード比較

複数のキーワードに対する検索結果を上下の TimeSlices に分けて表示する Split view (図 1(b)) を用いることで、複数キーワードに対する比較を行うことが可能になる。中心となる名詞、同じ格助詞、および同じラベルのついた動詞 (形容詞) は同じ位置に配置される。1 つの TimeSlice に集約表示する Aggregate view [1] も利用可能である。

4 可視化例

図 1(a) は、「チョコレートをあげる」ことに関する傾向の変化を可視化した例である。始めに「チョコレート」を入力し可視化した。結果から「あげる」を選択し TimeFlux を表示すると、毎年バレンタインデーの時期にのみ集中してイベントが発生して

いることが分かる。また、男女比が「女性 > 男性」から、ほぼ「女性 = 男性」へと変化してきていることも分かる。次に、そのイベントに関して展開し詳細を見ると、2009 年には「男子が」「女子に」「チョコレートをあげる」例が出現していることが確認出来る。図 1(b) の例では「コンビニ」と「スーパー」に行くユーザの男女の違い、購入物の傾向の違い、季節ごとの傾向の違い等を可視化している。

5 まとめと今後の課題

本稿では、人々の対象に対する行動、興味および意見の時系列変化をブログから抽出および可視化し、探索するためのシステムを提案した。今後の課題としては、より大量のキーワード (例: 全議員名) に対するブログ上の意見を比較、あるいはグループごとに集約して見せる可視化手法を検討したい。

参考文献

[1] M. Itoh, M. Toyoda, and M. Kitsuregawa. An Interactive Framework for Visualizing Time-series of Web Graphs in a 3D Environment. In *Proceedings of the 14th International Conference on Information Visualization (IV2010)*, pp. 54-60, IEEE, 2010.