

ブログユーザの行動・興味に関する時系列推移 3次元可視化システム

伊藤 正彦[†] 吉永 直樹[†] 豊田 正史[†] 喜連川 優[†]

[†] 東京大学生産技術研究所 〒153-8504 東京都目黒区駒場 4-6-1

E-mail: †{imash,ynaga,toyoda,kitsure}@tkl.iis.u-tokyo.ac.jp

あらまし ブログ上に記述される時間・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た。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、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。これにより、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。

キーワード 3次元情報可視化、インタラクティブシステム、ブログアーカイブ、時系列データ、係り受け解析

3D Visualization System for Time Sequential Changes in Blog Users' Activities and Interests

Masahiko ITOH[†], Naoki YOSHINAGA[†], Masashi TOYODA[†], and Masaru KITSUREGAWA[†]

[†] Institute of Industrial Science, University of Tokyo

Komaba 4-6-1, Meguro-ku, Tokyo, 153-8504 Japan

E-mail: †{imash,ynaga,toyoda,kitsure}@tkl.iis.u-tokyo.ac.jp

1. はじめに

ブログ等の CGM の普及にともない、ユーザは自身の興味、行動、主観的意見を即座にかつ簡単に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、製品、人物、政策等の評判調査など、マーケティング、社会分析の観点から重要なデータとなってきた。

これに伴い、注目キーワード抽出、それらに対する評判情報や時系列分布、および関連キーワードをブログから取得可能にする kizasi.jp^(注1) blogWatcher [1] 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。これらのサービスでは、話題のキーワードやそれらがいつ流行したのかを知ることは出来るが、ある話題に関する人々の行動、意見および興味が時間と共にどのように変化していくのかを可視化し探索するための仕組みは用意していない。

このような状況において、(i) 要素およびそれらの関係の時系列変化の可視化、(ii) 各要素の時系列頻度変化の可視化、および (iii) 詳細・関連情報探索のためのインタフェースを導入することでユーザの行動・興味の变化を俯瞰し興味深い領域をよ

り詳細に調査することが可能になると考えられる。

ウェブもしくはブログのリンク構造の様なネットワーク構造に関する時間変化を可視化する仕組みとしては、アニメーションを用いたもの [2] [3]、スプレッドシート風インタフェースを用いたもの [4] [5]、3次元空間の一軸を時間軸としてを用いたもの [6] 等様々な手法が提案されている。しかしながらこれらの手法では頻度変化を俯瞰する仕組みは用意されていない。一方、ThemeRiver [7] はキーワードごとの頻度変化を可視化可能にした。また、TIARA [8] は ThemeRiver に Tag-clouds [9] を組み合わせることにより、トピックを構成する要素の変化をも同時に可視化可能にした。これらの手法は変化の全体像の俯瞰に優れるが、構造変化の可視化やドリルダウンすることにより詳細レベルを表示し、それらの変化を同時に比較しながら探索することが困難である。CGM 分析結果に対する探索インタフェースとしては、ひとつなぎ^(注2)などの事例があるが、関係の時系列推移の可視化までは考慮していない。

提案するシステムでは、対象に関する行動（もしくは感想）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、(i) TimeSlices [10] を用いることで、対象と行動の

(注1): <http://kizasi.jp/>

(注2): <http://hitotsu.labs.goo.ne.jp/>

関係をツリー表示し、それらの時間推移（関係の変化 + 頻度の変化）を3次元空間を用いて可視化することが可能になる。(ii) TimeFluxes [13] を用いることで、ある行動に対する頻度変化を俯瞰することが可能になる。これにより、特定の行動の流行やその周期性などを容易に観測することが出来る。また、(iii) 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。TimeSlices は時間軸上に複数枚表示することが可能であり、これにより、複数時間の比較、キーワードごとの比較、男女ごとの比較、行動・感想の比較を実行でき、変化の詳細、影響等を探索可能なる。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. ブログ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び日本語ブログアーカイブの構築を進めてきている。この蓄積された全ブログデータに対して係り受け解析を行い、ブログユーザの行動・興味に関する記述を収集した。

解析に用いた J.DepP^(注3) では、複数の係り受け解析アルゴリズムが実装されており、今回は、素性列トライに基づく高速分類器 [11] を線形時間アルゴリズム [12] における係り受けの正誤分類に利用し、大規模ブログデータの解析を可能とした^(注4)。なお素性列トライは、組合せ素性に基づく線形分類器のための高速分類手法であり、実データからマイニングした基本分類問題の解を事前に求め、実際の分類の際に入力（分類問題）と類似する基本分類問題の解を部分結果として利用することで高速な分類を可能としている。

提案システムでは、J.DepP が出力した係り受け解析結果のうち、名詞と動詞の間の係り受けを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）を名詞に付随する格助詞で分類しデータベース化した。分類に用いた格助詞としては、高頻度で、かつ動作主、経験者、道具、対象、場所、時間などを示すのに主として用いられる {を, と, に, が, で} のみを考慮した。

現在、約150万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、係り元の名詞もしくは係り先の動詞を用いる。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（約1万8千種類）を検索対象および検索キーワードに用いることにより、対象に対する感想の時系列変化を調べることも可能である。

現在、図1の様に結果をテーブル形式で可視化するシステムも用意している。ここでは、単語をクリックすることで付帯情

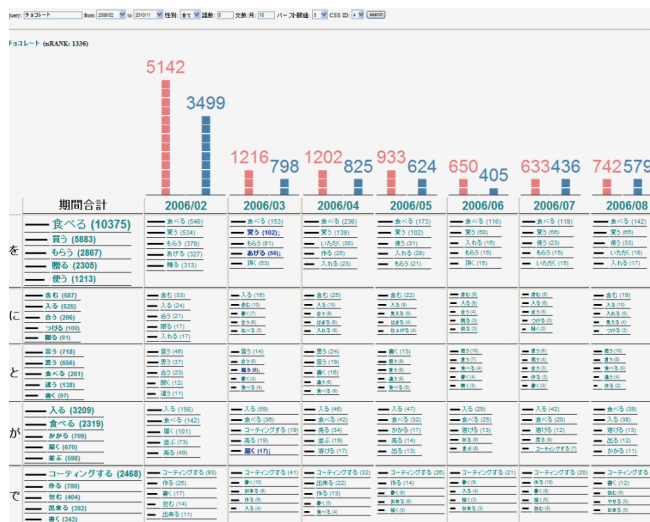


図1 係り受け解析に対する検索結果のテーブル表示

報（同一の動詞に係る他の名詞の格助詞による分類）の詳細を探索可能である。しかしながら、時間軸方向の一覧性、詳細を見たときにその前の状態を同時に見るのが困難、他のトピックとの時間軸および付帯情報表示の同期をとりながらの比較が難しい等の制約がある。

3. 3次元可視化システム

3.1 TimeSlices による時系列関係変化の可視化

検索結果としてのイベント群は、入力キーワードを中心とし、その下に格助詞、各格助詞の下に動詞が配置されたツリー表現で TimeSlices [10] 上に可視化される。図2の例では、入力キーワード「チョコレート」を中心として、その周りに格助詞、動詞がツリー表示されている。見せたい格助詞のみを表示するフィルタリング機能も用意している（図2では {を, に, が} のみ表示）。イベントは各月ごとに集約されており、各月の出現頻度上位 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。ノードサイズは、巨大なるのを防ぐため $r\sqrt[3]{c_i(t)}$ で計算している (r: 定数, $c_i(t)$: イベント i の月 t における出現頻度)。ノードの色は要素ごとに任意の色を設定可能であるが、動詞を表すノードは TimeSlice ごとの差異を認識しやすくするために、TimeSlice と同じ色になるように設定出来る。

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、時間軸 (Timeline) 上に任意の数の TimeSlice を追加することが可能である。これにより、異なる複数時期による人々の動向の違いの詳細を比較することが可能となる (図2)。

複数時間におけるイベント群の類似・差異を理解しやすくするために、隣り合う TimeSlice の間では同じイベントが同じ位置に配置される。この際、文献 [10] では全時間の全ノードお

(注3): <http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/jdepp/>

(注4): ブログデータの解析速度は 3.2Ghz CPU のサーバで 1 秒辺り約 21000 文であった。

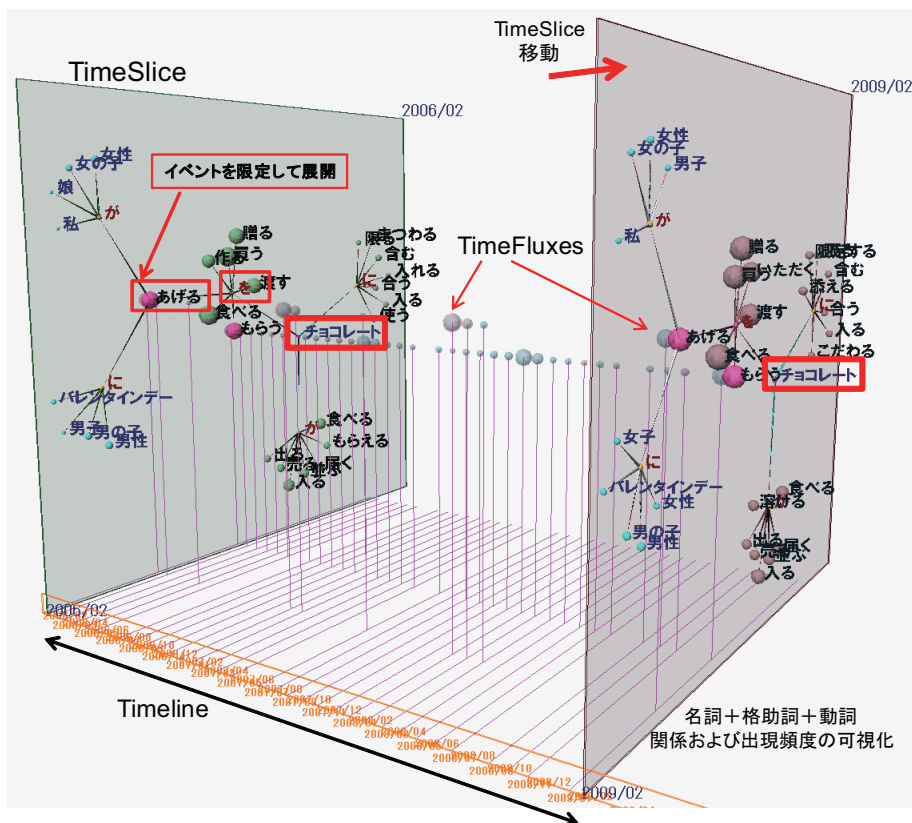


図 2 TimeSlices によるブログ上の話題変化の可視化・探索システム

よびエッジからなるスーパーグラフからグラフのレイアウトを計算する Diehl [14] らの手法を応用したが、本システムではこれに加え、表示されている全ノードおよびエッジのみからなる部分スーパーグラフを作成しレイアウト計算を行う手法を用意した。これにより、効率の良い可視化空間の利用が行える。各 TimeSlice に表示されているグラフは力学的アプローチ [15] に基づいたレイアウトを行っている。ユーザは対話的にノードをドラッグしノードの位置を変更することができる。その際に、各 TimeSlice 間でノードの位置が自動的に同期される。

3.1.1 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 3) [10] 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。この際、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（もしくは形容詞）は各 TimeSlice 上で同じ位置に配置される。1つの TimeSlice に複数キーワードからなる Tree を集約表示する Aggregate view [10] も利用可能である。また同様の仕組みを用いることでキーワード比較の他に、男女比較、動詞（行動）・形容詞（感想）比較を行うことも可能である。

3.1.2 タイル表示

複数の TimeSlices をより詳細に比較するために、タイル表示を可能にした (図 4)。タイル表示は 3次元空間で実現され、ユーザはシームレスに通常の 3次元表示とタイル表示を切り替えることが可能である。表示方式を自由に切り替え可能にすることで、ユーザは 3次元表示で全体を俯瞰し、興味のある

領域を表示した状態でタイル表示に切り替え、その領域に関する詳細を探索する、といった操作が可能になる。具体的には、TimeSlices を奥行き方向にスライドさせた状態で、ユーザ視点を時間軸方向から TimeSlices を見るように移動し、パース無しモードで描画することで実現した。パース無しで描画することで、3次元表示特有のパースがかかることにより重なり合うノードの位置がずれる現象が防止される。

また、同様の仕組みを用い、奥行き方向のスライドをしないことにより、複数時間の状態を重畳表示して比較することも可能である。

3.2 TimeFluxes による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、流行の周期性等が直感的に理解可能になる (図 2 および図 3)。この球の列を TimeFluxes [13] と呼ぶ。各月ごと以外にも、数ヶ月ごとに集約したものを一つの球として可視化することも可能である。各ブログは語尾などの分析により簡単な男女判定を行っており、色違いの 2種の TimeFluxes を用いることにより男女ごとの記述頻度の変化を同時に可視化し比較することが可能になる。図の例では男性に青、女性に赤を割り当てている。これにより男女の割合の逆転などの現象を容易に観測出来る。

また、前の月に比較して変化量の大きな箇所をハイライト表示することにより、パース部分を強調することも可能にした。図 5(a) では、選択したイベントに関してパース箇所をハイライト表示している。図 5(b) では、表示している全イベント

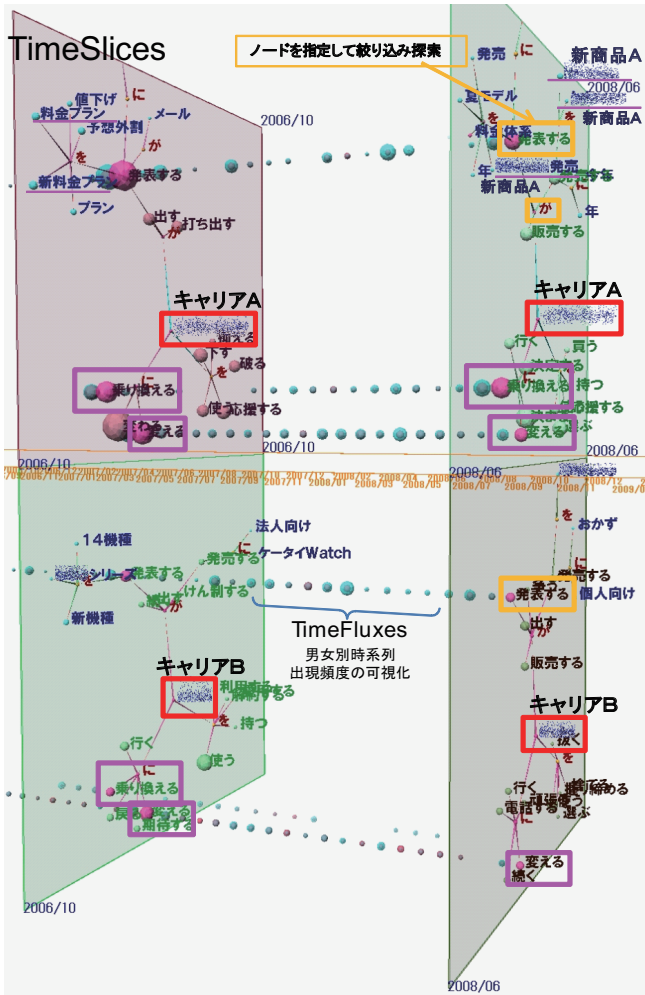


図3 複数 TimeSlices による複数トピックの比較

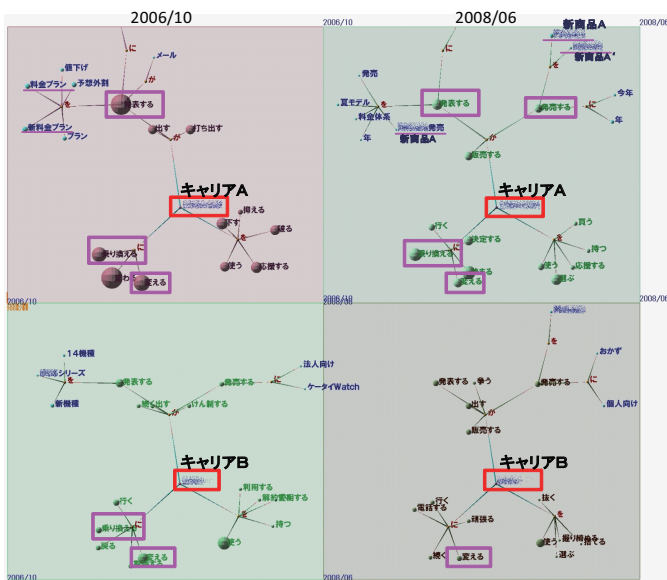
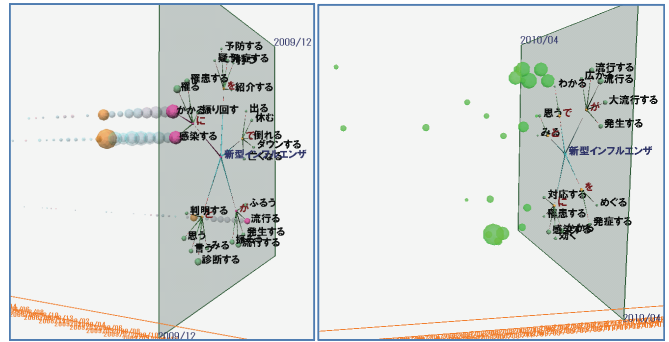


図4 タイル表示を用いた比較インターフェイス

に関してバーストしている部分のみをハイライト表示している。これらの機能は、特に注目すべき時期およびイベントの発見および探索を容易にする。



(a) 選択イベントに対するバースト強調表示 (b) 全表示イベントに関するバースト箇所表示

図5 バースト箇所のハイライト

3.3 探索インターフェイス

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イスを提供する。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を行う。これにより、選択したイベントに関してさらに詳細な「誰が、どこで、何を」などの情報を一覧出来る。図6の例では、「感染する」というノードを選択することで、「新型インフルエンザに感染する」というイベントを選択し、付帯情報として、例えば「高校生が」「神戸で」などを検索し「感染する」の子ノードとして可視化している。図1の場合と異なり、本手法では、詳細を展開した際にも直前までのコンテキストを可視化した状態で保持しているため、他の状況との関係を比較しながら探索することが可能である。

複数の TimeSlices 間で同じ動詞のノードがある際は展開操作の同期をとることにより同じ状況に対する比較を行うことを可能にしている。図3の場合、一つの TimeSlice に対して展開操作（例えばイベント「キャリアAが発表する」を展開）を行うことで、異なる時間を表す TimeSlice でも同じイベント（「キャリアAが発表する」）に対して展開操作が自動的に行われる。同様に、異なるキーワードを表す TimeSlices 間では同等のイベント（例えば「キャリアBが発表する」）に対する展開操作が自動的に行われる。

他の探索パターンとしては、動詞（形容詞）や他の名詞に中心を移動させる、もしくは他の3つ組みに焦点を移動するなどの操作を可能にする。これにより、探索しながらそこから派生した他の探索コンテキストに次々と移動していくことが可能になる。例えば、図6の例で「高校生が感染する」という3つ組みを選択することで、高校生が感染したのものとして他にどのようなものが、いつ頃流行したのか等の情報を探索できる。

4. 可視化例

図2は、「チョコレートをあげる」ことに関する世の中の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。左の TimeSlice は2006年2月の状態を表しており、右の TimeSlice は2009年2月の状態を表している。その間の TimeFlux を見てみると男女比が「女

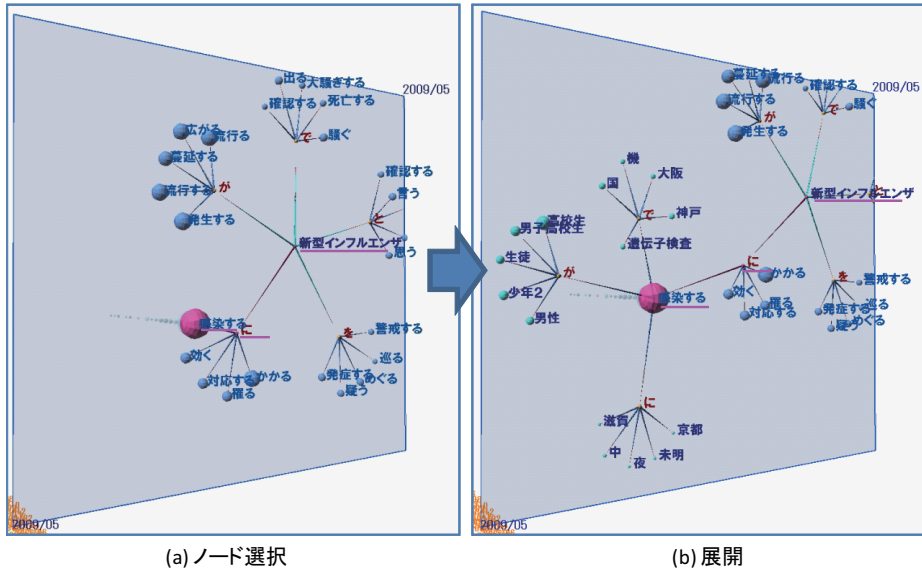


図6 ノードの選択・展開による付帯情報可視化

性>男性」から「女性=男性」もしくは「女性<男性」へと変化してきていることが分かる。次に、そのイベントに関して展開し詳細を見ると、2006年2月には「女性が」「男性に」「チョコレートあげる」例しか存在しないが、2009年には「男子が」「女子に」「チョコレートあげる」例が上位に出現していることが確認出来る。

図3および図4は、複数の携帯キャリアのマーケティング効果を可視化、比較した例である。図において、上部の TimeSlices はキャリア A に関する話題を可視化している、下部の TimeSlices はキャリア B に関する話題を可視化している。キャリア B に比べてキャリア A では「乗り換える」「変える」の記述が目立つ。TimeFluxes を表示すると、キャリア A 側では何度か「乗り換える」ピークが来ていることが確認出来る。図3, 4における左側の TimeSlices は一度目のピークが来たタイミング(2006年10月)に合わせてある。右側の TimeSlices は二度目のピークが来た2008年6月に合わせてある。これらのタイミングでキャリア A で何らかの発表もしくは発売をしていることが、ノードのサイズから伺える。「発表する」「発売する」に関して、詳細を展開表示することで、一度目のピークでは「新料金プラン」、二度目のピークでは「新商品 A」を発表している事が分かる。一方、キャリア B も何度も「発表する」の頻度が大きくなっていくにもかかわらず、それが「乗り換える」や「変える」という行動にあまり影響を与えていないことが伺える。

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3次元可視化し、探索するためのシステムを提案した。提案したシステムは汎用的な時系列情報の構造および頻度変化の可視化フレームワークとして基盤を構築しており、様々なデータに対する時系列探索の為のアプリケーション構築に適用可能である。インタラクティブな3次元アプリケーション構築環境として、IntelligentBox [16] を用い

ている。現在、Twitter や TV キャプションといった他の情報リソースに対する比較インターフェースも構築中である。

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。また上記のような大量キーワードが入力として与えられた際に、膨大な量の情報から自動的に興味深い領域を見つけ、それらを俯瞰出来る仕組みが必要とされる。

文献

- [1] T. Nanno, T. Fujiki, Y. Suzuki, and M. Okumura, Automatically Collecting, Monitoring, and Mining Japanese Weblogs. In *Proceedings of the 13th International World Wide Web Conference*, pp. 320-321, 2004.
- [2] M. Kitsuregawa, T. Tamura, M. Toyoda, and N. Kaji, SocioSense: A System for Analysing the Societal Behavior from Long Term Web Archive, In *APWeb; LNCS, Vol. 4976*, pp. 1-8, 2008.
- [3] N. Nakazono, M. Kazuo, and J. Tanaka, NeL2: Network Drawing Tool for Handling Layered Structured Network Diagram. In *Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation*, pp. 109-115, 2006.
- [4] E. H. Chi and S. K. Card, Sensemaking of Evolving Web Sites Using Visualization Spreadsheets, In *Proceedings of the 1999 IEEE Symposium on Information Visualization*, pp. 18-25, 1999.
- [5] M. Toyoda, and M. Kitsuregawa, A System for Visualizing and Analyzing the Evolution of the Web with a Time Series of Graphs, In *Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia*, pp. 151-160, 2005.
- [6] E. H. Chi et al. Visualizing the Evolution of Web Ecologies, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 400-407, 1998.
- [7] S. Havre, E. Hetzler, P. Whitney, and L. Nowell, The-

- meRiver: Visualizing Thematic Changes in Large Document Collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1),9-20, 2002.
- [8] S. Liu, M. X. Zhou, S. Pan, W. Qian, W. Cai, and X. Lian, Interactive, Topic-based Visual Text Summarization and Analysis, In *CIKM '09*, , pp. 543-552, 2009.
- [9] F. B. Viégas and M. Wattenberg, Tag Clouds and the Case for Vernacular Visualization, *Interactions*, 15(4), pp. 49-52, 2008.
- [10] 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.
- [11] N. Yoshinaga and M. Kitsuregawa, Polynomial to Linear: efficient Classification with Conjunctive Features, In *Proceedings of the the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 (EMNLP '09)*, pp. 1542-1551, 2009.
- [12] M. Sassano, Linear-Time Dependency Analysis for Japanese, In *Proceedings of the 20th international conference on Computational Linguistics (COLING '04)*, pp. 8-14, 2004.
- [13] M. Itoh, 3D Techniques for Visualizing Users' Activities on Microblogs, In *Proceedings of the IET International Conference on Frontier Computing - Theory, Technologies and Applications*, ISBN: 978-1-84919-208-8, 2010
- [14] S. Diehl and C. Görg, Graphs, They are Changing. In *The 10th Symposium on Graph Drawing*, pp. 23-30, 2002.
- [15] T. M. J. Fruchterman and E. M. Reingold, Graph Drawing by Force-directed Placement, *Software Practice and Experience*, 21 (11), pp. 1129-1164, 1991.
- [16] Y. Okada and Y. Tanaka, IntelligentBox: A Constructive Visual Software Development System for Interactive 3D Graphic Applications, In *Proceedings of Computer Animation '95*, pp. 114-125, 1994.