

ソーシャルブックマークにおけるタグの時系列的な依存関係の解析

川中翔¹, 佐藤周行²

¹ 東京大学基盤情報学専攻

² 東京大学情報基盤センター

我々は新しいものや概念が新たに創られた場合それを創造と呼ぶが、実際にはそれ以前に既に成立していた概念が下地にあって創られる。これらは解析すべき重要な依存関係であるが、既存の関係性解析の研究は時系列的に静的なものが多くを占めている。本研究では、ソーシャルブックマークにおける概念を記述するタグを解析することで、時系列的な概念の依存関係を抽出する手法を提案する。この手法は、共起率が高いタグ同士は依存関係を持ち、出現順が早い方が依存関係の親となるタグである、との仮定に基づいている。さらに本研究では実験によって手法の評価を行ない、直観的に有用な結果を抽出することに成功した。

Analysis of Chronological Tag Dependency in Social Bookmark

Sho Kawanaka¹, Hiroyuki Sato²

¹Department of Frontier Informatics, the University of Tokyo

²Information Technology Center, the University of Tokyo

“Creation” can be defined as making something exist that has not existed before. However, in many cases, something “new” is derived from something “old”. These are essential to analyze, but previous work mainly concentrate on chronologically static relations. In this paper, we propose a method of extracting these chronological concept dependency by analyzing tags that describe concepts in Social Bookmark. This method, is based on the hypo high coocurrent tags have dependency and the first emergence time of tags determines order of dependency. Furthermore, we evaluate this method by experiments, which show fair success.

1 はじめに

近年様々な技術の進歩により、情報の伝播性が高まり、また情報の複製が容易に可能となっている。この傾向をふまえ、今後は、新しい情報がどのように創られ、どのように後の情報に影響を与えているかなどの、時系列を考慮した情報の関係評価が、重要になることが予想される。我々は従来存在しなかった物や概念が新たに創り出されたとき、それを「創造」と呼ぶが、多くの場合、実際にはそれ以前のものや概念が下地にあって創られる。一例として、プログラミング言語「Ruby」は、Perl, Javaなどの既存のプログラミング言語の特徴を組み合わせで誕生した言語であり、後にはRubyを元にしたフレームワーク「Ruby in Rails」が誕生した。このような時系列を考慮したものや概念の関係の解析は極めて重要である。

この種の研究をおこなうにあたり、解析のためのデータの取得が困難であるとの問題があり、これまでの時系列的な解析は自己相関分析などプリミティブなものが多くを占

める。しかしながら近年登場した「Consumer Generated Media(CGM)」と呼ばれる、エンドユーザがコンテンツを生成するメディアでは属性情報が定型化されており、時間情報が取得しやすくなっている。またCGMなどのWeb上のメディアにおいては、コンテンツ同士が互いに影響し、次々と新たなコンテンツが発生するマッシュアップと呼ばれる現象が一般的にみられ、そこで知識を抽出する場合に時間情報は特に解析すべき本質的な情報であるといえる。マッシュアップとは、Fig.1に示すような、複数の異なるサービスや概念を組み合わせることで、新しいサービスや概念を形作ることである。Web上におけるサービスや、動画共有サイトやソーシャルネットワーク上の概念のマッシュアップなどが代表的なものであり、この背景を元として様々な研究の可能性が考えられる。Semantic Webの視点から、マッシュアップを促進させるようなアーキテクチャの設計や、Web Miningの視点からWeb上における既存の概念の依存関係の抽出などが代表的なものである。本研究では後者の視点に立ち、CGMを解析すること

で新たな時系列的解析について試みたので報告する。

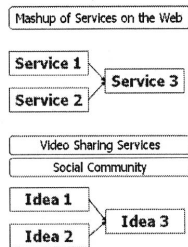


Fig. 1: Mashup

本研究では、解析対象として CGM の一つであるソーシャルブックマークを選択し、利用者によるログデータを分析することで、時系列的な概念の関係について議論を行う。ソーシャルブックマークでは多くのエンドユーザによって意味付けがなされており、またタイムスタンプが自動的に保存される。このようなソーシャルブックマークの時間の取得しやすさとその集散的な知識を利用し、統計的に大量のデータを解析することで時系列的な概念間の関係の抽出を試みている。

背景となる従来手法は次に示す通りである。既存の語の関係解析は次のものが主に存在する。

- クラスタリングによる語の分類、階層化 [7]
- 共起度などによる語の結びつきの強さ [5]
- 同義語、反義語の検出 [6]
- 同位語の検出 [8]

一方時系列的な分析としては、一つの語の期間的な盛り上がり (burst[4]) に着目した自己相関分析などが存在する。本研究では、語の間の関係解析と時間情報を利用した解析の両方のアプローチを元に、依存関係の解析を行う。

本研究では、時系列的な概念間の関係について、問題設定を行い、解析のための仮説を立て、実際のデータに手法を適用した実験を行い、仮説に関する評価を行う。副次的な意義は次の通りである。応用先としては新たな重要度評価として検索などへのサービスの適用が考えられる。また別の視点からの意義として、集散的な知識への理解の深化が挙げられる。将来的には、情報の新規発生について定量的に分析することで、創造性の評価が可能となり、創発的な社会の実現に貢献することが大きな目標である。

本稿の構成は次の通りである。2 章では解くべき問題とアプローチについて説明し、3 章では実験について報告す

る。4 章では関連研究について説明し、5 章ではまとめと今後の課題を述べて締めくくる。

2 ソーシャルブックマークにおける時系列性の解析

本章では解くべき問題とアプローチについて詳しく記述する。

2.1 時系列的な依存関係

本研究で設定する問題は次の通りである。ある概念があるとき、その概念の成立に影響を与えた他の概念があるとする。本研究では、この種の依存関係の抽出手法を提案し、その妥当性および依存関係の存在性についての検証を行う。

本研究で用いる用語を次のように定義する。まず「時系列的な依存関係」を次のように定義する。ある概念 $(A_1, A_2, \dots)$ が成立したあとに、その影響を受けて、他の概念 B が成立したとする。この場合、各 $(A_1, A_2, \dots)$ を B の親概念と呼び、 B を各 $(A_1, A_2, \dots)$ にとっての子概念と呼ぶとする。これらの関係を時系列的な依存関係と呼ぶこととする。これらの関係は Fig.2 のように表される。

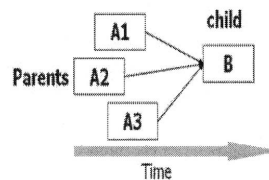


Fig. 2: Definition

2.2 ソーシャルブックマーク

本研究では Web 上のソーシャルブックマークサービス (SBM) を解析することで研究を行なう。SBM とは Web 上のブックマーク管理・共有サービスのことある。サービス内でユーザは Web 上の様々なドキュメントについてタグを付加することで意味付けを行なっている。(タグとはユーザが任意に決められるフリーワードである。) ユーザはタグを利用することで、ドキュメントが含む概念を記述しており、タグは概念と同一視して考えることができる。また SBM は世の中の流れを反映していると考えられ、タ

グ付けが行なわれた時刻が自動的に保存される。これらの性質を元に、タグ付けを大量に収集、統計的に解析することで概念間の時系列的な依存関係の抽出をおこなう。

SBMを解析する場合の基本性質について次に示す。SBMにおける基本情報は、ある時刻 t_i において、あるユーザ u_j が、あるタグ c_k を、あるドキュメント d_l に付ける現象で、これは記号的には $Annotation(u_j, c_k, d_l, t_i)$ と表される。我々は特に時刻について着目し、大量に収集し・統計的に分析することで依存関係の抽出をおこなう。

2.3 ソーシャルブックマークにおけるタグの時系列的な依存関係の例

SBM「del.icio.us」において、プログラミング言語「Ruby」を表すタグ「ruby」と、Rubyを利用したフレームワーク「Ruby on rails」を表すタグ「rubyonrails」が存在する。（「Ruby on Rails」は「Ruby」を元に出現した概念であり、これらは明らかに時系列的な依存関係を持つ。）タグ「ruby」と「rubyonrails」は意味的に近いため、SBM上でのタグの利用において文書間、ユーザ間において高い共起性を有する。（予備実験において指標AEMIを利用し、del.icio.usでの両タグのユーザ間共起度を測ったところ、お互いに相手タグが最も共起度が高いタグであった。AEMIについては3.1節で後述する。）また予備実験において、「ruby」というタグが初めて使われた360日後に、「rubyonrails」というタグが初めて使われており、この前後関係は現実社会のものと一致している。（日時の差は現実のものと必ずしも一致しないが、データの量が増えるにつれ近づく傾向があると考えられる。）このように現実上の時系列的な依存関係を持つ概念同士は、SBM上においてもタグとして、各タグの発生時期の前後という性質と、各タグの利用における共起性の存在が一般に考えられる。我々は、SBMにおけるタグの利用の方を解析することで、現実上の概念依存関係の抽出を目指す。

2.4 時系列的な依存関係抽出手法の提案

まず依存関係抽出のための考え方を示す。背景として、依存関係の根拠となる、概念成立過程における実際の証拠を得るのは、困難であると考えられる。しかしながら、人々による概念の利用という状況証拠からの推定は可能であり、少なくとも傍観者の立場からみた概念の依存関係の抽出は可能であるとの仮定を元に解析を行なう。

続いて依存関係抽出のための仮説を説明する。タグ付けの空間がある時、次の2つの仮説を軸に抽出を行なう。

1. 共起度の高いタグペアは関係が深い。

2. 関係が深いペア (a, b) があるとき、一方 a がもう一方 b の後から出現したなら、 a は b から派生した概念である。

上記仮説を元にした依存関係抽出のための手法を示す。ある入力タグ A があるとき次の手順によって抽出をおこなう。

1. A と共起度の高いタグを N 件取得する。（ N は依存関係のあるタグの数を示す閾値である。）
2. A と抽出された各タグの初登場時期（初めて使われた日時）を比較し、早い方を親概念、遅い方を子概念とする。
3. A 以外の抽出されたタグ間についても同様に、共起度が高い（一方から N 件以内）場合は前後関係を取得し、依存関係を定める。
4. 結果として A と A と依存関係のあるタグは、共起度による関係の深さと、前後関係の2つを属性に持つ有向リンクにより表されるグラフ関係が抽出できる。

2.5 結果の表示

前節手法によって抽出されたタグ間の依存関係をユーザに分かり易く示すインターフェイスを示す。予め各タグについて関係の深いタグとその前後関係を抽出し、データベースに格納する。ユーザから任意のタグについて問い合わせがあると、Fig.3に示されているように依存関係のあるタグを示す。またFig.4に示すように、サブ入力として、表示するタグ、リンクの密度を変更、リンクの特徴の表示、また親概念のみ、もしくは子概念のみに抽出も可能とすることで柔軟な表示への対応を可能とする。

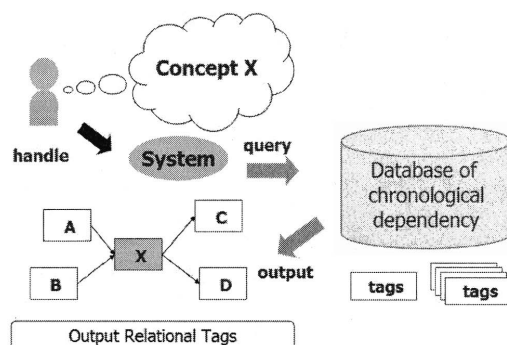


Fig. 3: Interface

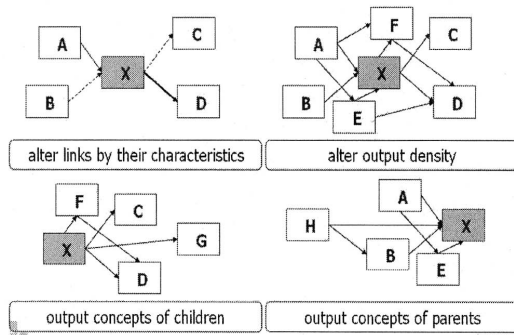


Fig. 4: Interface2

3 実験

2章で述べた仮説がどういう状況で成立するか、提案インターフェイスが効果的であるかについて実験によって検証を行なう。今回は2つの見方から実験を行う。

1. 今回取得する *Annotation* の集合空間を我々の社会と独立した空間と捉え、その空間内における関係を抽出、関係性を描画する閲覧インターフェイスの実装をおこなう。
2. 今回取得する *Annotation* の集合空間を我々の社会の実空間の部分空間と捉え、そこから我々の空間で通用する知識の抽出を試みる。
 1. の場合は必ずしもサービス空間内におけるタグの認識が我々の実空間でのタグに関する認識と一致しているとは限らない。サービス内におけるナビゲーションとして用いることができる。
 2. の場合はこの場合、タグ同士の関係について人間の感覚からの正解を与え、評価を行うことができる。

3.1 データセット

前節の手法について実際のデータによって実験をおこなった。今回用いたのは2007年11月中に取得した、SBMdel.icio.usのデータである。

del.icio.usでは各ページについてタグ付けの情報がRSS形式で提供されており過去に遡ってデータの取得が可能であり、本研究に適しているため選択した。またdel.icio.us

Table. 1: detail of analysis data

オブジェクト数	3177
ユーザ数	270932
タグ数	82393
アノテーション数	3649162

は世界で最も利用されているSBMで2007年11月にはユーザ数が100万人に達している。del.icio.usは、典型的な外部リンク参照型のサービスであり、サービス内における概念の関係と外部における関係の共通性は比較的高いと考えられる。ユーザの多くは体系的に情報を扱うことを志向し、del.icio.usにおいてタグ付けを行なうときには、既に付けられているタグのうち件数の多いものが表示されるため、他の人の影響があると考えられる。タグ付けについての制限などは特にない。

取得したデータの詳細はTable.1に示す通りである。

なお、今回解析対象として扱ったタグは登場回数が上位2000件のものに限った。計算量を少なくすることと、ノイズの除去を意図している。

また共起度については指標 *AEMI* を利用して求めた。*AEMI* は確率を考慮した精細な共起度を測るための指標で以下のように表される。

$$AEMI(a, b) = MI(a, b) + MI(\bar{a}, \bar{b}) \quad (1)$$

$$-MI(a, \bar{b}) - MI(\bar{a}, b) \quad (2)$$

$$MI(a, b) = P(a, b) \log \frac{P(a, b)}{P(a)P(b)} \quad (3)$$

この場合 $P(a)$ はタグ付けにおいてタグ a が用いられる確率であり、 $P(a, b)$ はタグ付けにおいてタグ a と b の両方が用いられる確率である。さらに $P(\bar{a})$ はタグ a が投稿において用いられない確率を表す。 MI は共起率を評価するための一つの指標であり、 $AEMI$ は MI を組み合わせることで、スケールを考慮した確率的な共起度の高さを測ることができる。

3.2 インターフェイスの実装

2.3節に示した手法に基づき、3.1節に示したデータセットにおけるタグについて実際に依存関係を可視化した。Fig.5とFig.7はそれぞれタグweb2.0とタグflockについて依存関係を可視化したものである。Fig.6は同じくweb2.0の子概念のものだけを密度を変えて表示したものである。また出現時期が違いのものについては実線とし、近いものについては破線としている。

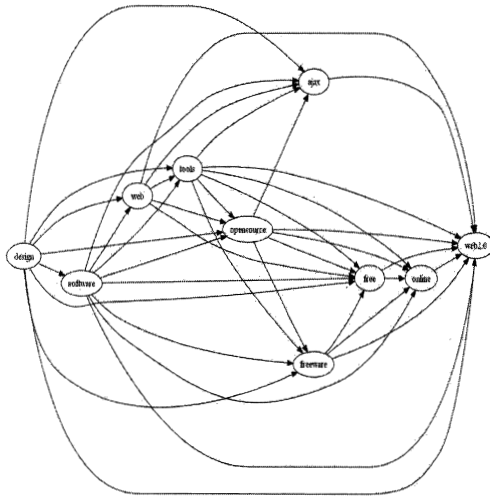


Fig. 5: web2.0

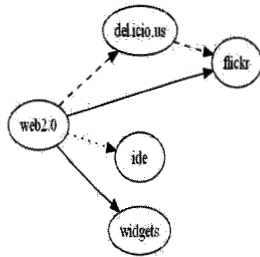


Fig. 6: web2.0

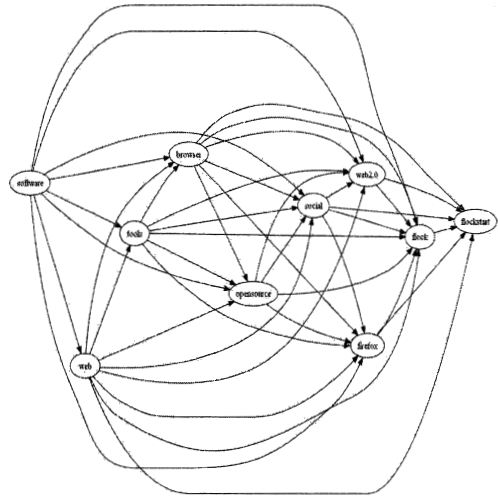


Fig. 7: flock

3.3 知識抽出

続いて *del.icio.us* を我々の実空間の部分空間として捉え、そこからの知識抽出についての実験を行なったので報告する。概要は次の通りである。2.3 節の手法によって抽出された依存関係と、人間の選ぶ依存関係 (正解) との一致度を比較し、手法の妥当性を検証する。

人間による正解の取得方法と検証方法は次の通りである。まず、各タグに対して、共起度上位 M 件 ($M > N$) の中から、依存関係があるとするタグを N 件選んで貰う。 (M は候補タグ数を示す閾値である。) これと手法によって抽出された N 件との一致度を比較し、高いほど再現率・精度が高くなると評価する。また人間が選んだタグ間の前後関係について、どのような場合に精度が高くなるかについて検証を行なう。なお、正解タグの選定については、2 章の時系列的な依存関係に基づき、筆者があらかじめ与え、それらは結果の評価時に示すものとする。

3.3.1 問題点

本研究ではタグの初登場時期に着目しているが、解析対象の *del.icio.us* のサービス開始は 2000 年に入ってからである。すなわち、それ以前に実空間に既に出ていたタグについてはサービス内における初登場時期にあまり意味はなく、当然それらのタグについては前後関係の精度が悪くなると予想される。ここでは検証するにあたりサービス

開始以前から存在していたタグと、そうでないタグについて区別して評価する。

各タグについて、サービス内において何番目に初出現したかを表す関数 $first(tag)$ を定義する。このとき、タグがサービス内において何番目に出現したか ($first(tag)$) と、タグの使用回数を示したのグラフが Fig.8 である。(四角記号、縦軸: 各タグの使用回数、横軸: $first(tag)$) 約 80000 のタグのうち、使用頻度 2000 番以上のもののなかで 1783 個までが 10000 番目までに含まれており、939 個までが 2000 番目までに含まれていた。Fig.8 の十字記号部分は、横軸を区間ごとに分け平均を取ったものであり、一定の相関性が見られることがわかる。

これは次の 2 つの理由から説明がつく。

- 登場時期が早いタグほど、以後の投稿時に選ばれる機会が増え、登場回数が多くなる傾向がある。
- 登場頻度が高い傾向があるタグほど、初回登場時も早くなる傾向がある。

このとき、上記の傾向に比して、登場時期が遅いにも関わらず出現回数が増えるタグの存在が考えられ、これらは途中から発生した可能性が高いタグ、もしくは途中から出現した重要なタグと考えられる。

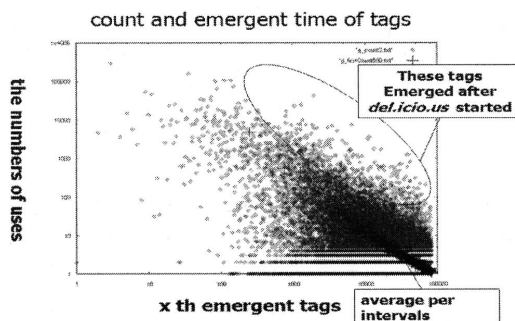


Fig. 8: count and first used timing of tags in log scale

次の式 (4) のスコアが高いタグから上位 50 件を抽出した一覧を次に示す。これらのタグについては、本手法が効果的に適用されるタグと位置づけ、評価時の比較に用いる。

$$\log(\text{出現回数}) * \log(\text{何番目に出現したか}) \quad (4)$$

Recently emergent significant tags

web2.0, imported, flock, citation, jquery, iphone, ajax, songbird, usb, beryl, rails, flockstart, flickr, firefox, s3, speedtest, rubyonrails, powershell, portable, widgets, del.icio.us, storage, ide, password, ubuntu, hdr, online, trac, bibliographic, dojo, 2.0, lists, lifehacker, whois, oreilly, photos, ie7, taskbar, gtd, conversion, adserver, update, firefox:bookmarks, google, gmail, cheatsheet, seo, vista, sandbox, translator

3.3.2 検証

検証方法の詳細と結果を示す。

各タグについて依存関係のあるタグ N 件があるとする。これらに対して、人間による正解と比較することで手法の妥当性を検証する。

人間による正解の取得方法と検証方法は次の通りである。まず、各タグに対して、共起度上位 M 件 ($M > N$) の中から、依存関係があるとするタグを N 件選んで貰う。(M は候補タグ数を示す閾値である。) これと手法によって抽出された N 件との一致度を比較し、高いほど再現率・精度が高くなると評価する。また人間が選んだタグ間の前後関係について、どのような場合に精度が高くなるかについて検証を行なう。なお、正解タグの選定については、2 章の時系列的な依存関係に基付き、筆者があらかじめ与え、それらは結果の評価時に示すものとする。また前章で抽出した途中から出現したと考えられるタグと、単純に出現回数が多いタグをそれぞれ入力タグ群として各タグに正解を与えた場合の精度について比較を行なう。

検証に用いたタグは次の 10 のタグである。出現回数が上位のもの、および途中から出現したと考えられる傾向が高いものの中から、被験者が詳しいものとして選択した。なお、($n = 5, m = 50$) として検証をおこなった。

- 共起回数が高いタグのサンプル (A 群)
software, web, webdesign, programming, shopping
- 途中から出現したとされるタグのサンプル (B 群)
web2.0, iphone, vista, ajax, rails

以下に具体例を示す。次のタグはタグ web2.0 の共起度上位 50 件のもので、このうちの上位 5 件が手法により抽出された候補タグである。

Relation tags of “web2.0”

ajax, tools, online, web, collaboration, social, office, javascript, webdesign, photo, editor, productivity, blog, design, browser, internet, photography, framework, photos, storage, community, technology, business, bookmarks, tagging, search, del.icio.us, flickr, sharing, development, widgets, free, socialsoftware, tags, webdev, flock, cool, bookmarking, blogs, wiki, powerpoint, resources, presentation, awards, blogging, tips, dojo, todo, css, tool

一方ランダムに並べられたこの 50 のタグから、被験者が選択したタグは次の 5 つであった。

- 親概念候補
web, blog, collaboration, development
- 子概念
ajax

このうち手法により共起度上位 5 件以内に入っているものは, web, collaboration, ajax の 3 つであり, 上位 20 件以内に入っているものは, web, collaboration, ajax の 4 つであった。

また人間の与えた前後関係と手法による前後関係が一致していたのは, web, blog, collaboration, web の 4 つであった。

同様に他のタグについても、選ばれたタグの一致度と、前後関係の判定結果を示す表が Table.4 と Table.5 である。一致度は正解として与えられたタグが、手法によって抽出されたタグの上位 5 件以内なら○、20 件以内なら△としている。タグ名に米印がついている方のタグが人間が親概念と判定したタグである。

例として、入力:software*, 正解:internet、一致度×、前後関係○と示されているタグのペアは人間がタグ software について、一つの子概念となるというタグ internet を与え、それは手法によっては上位 20 件以内に抽出されていないが、前後関係は手法によって与えられるものと一致していることを示している。

全体の傾向は Table.2 と Table.3 のようになった。Table.2 に示されているように B 群の方が一致度は高く、前後関係の精度も高い。Table.3 では各タグペアの前後関係の精度について、各タグの登場時期の差毎の統計を表している。 $first_d(a, b)$ はタグペア (a, b) がある時 $first_d(a, b) = |first(a) - first(b)|$ と定義される、初登場時期の差を表す関数である。Table.3 に示されているように初登場時期が離れているほど、精度が高くなる傾向がみられた。

Table. 2: results

	5 件一致度	20 件一致度	前後関係精度
A 群	8/25	16/25	19/25
B 群	15/25	23/25	22/25

Table. 3: accuracy of orders

$first_d$	1~10	10~100	100~1000	1000~
精度	2/5	5/8	6/6	26/27

3.3.3 結論

検証をした結果、実験の範囲内において、手法に一定の妥当さがみられた。また特に、サービス開始以後に出現したと考えられるサービスにういて高い精度を示すことが分かった。すなわち、データ取得期間が長いほど本手法はより効果的となる。またタグペアの前後関係の精度について、出現時期が離れているものほど精度が高くなるということが分かった。

しかしながら、現状の実験では不十分であり、今後は、正解の与え方などの条件を変えることや、データを増やすなどの変更をしてさらに十分な実験を行っていく必要がある。

4 関連研究

Web 上における行動を分析するにあたり時間情報の考慮は効果的である。Adar ら [1] はブログ上での情報の流れについて、テキスト類似度、リンク、時間情報を元に解析するモデルを提案した。これは Web 上における情報の流れについて定量的に分析した数少ない研究である。Google Trends[10] は期間ごとの検索語の使用回数の変化を表示している。Dobunko ら [1] は画像共有サービス Flickr において、期間ごとの重要なタグを抽出し、またそれらについて可視化をおこなった。彼らは入力となる期間を柔軟することができる経済的なアルゴリズムを提案している。Kleinberg[4] はドキュメントストリームにおける盛り上がりを検出する手法を提案した。

SBM は近年出現した重要な知識プラットフォームとして注目を集め、複数の視点から研究がおこなわれている。Golder ら [3] は SBM における、ユーザ、タグ、ブックマークの性質について分析をおこなった。彼らはユーザの行動、タグの種類、ブックマークの人気に関する興味深い法則を発見している。丹羽ら [6] は SBM におけるユーザベース

Table. 4: Relations of frequent tags

入力	正解	一致度	前後関係
software	computer*	△	×
software*	opensource	○	○
software*	windows	○	○
software*	internet	×	○
software*	freeware	○	○
web	technology*	×	○
web	software*	△	×
web*	web2.0	○	○
web*	webdesign	○	×
web*	blog	×	○
webdesign	*web	○	○
webdesign	*graphics	△	×
webdesign	*design	○	○
webdesign	*html	△	○
*webdesign	web2.0	△	×
programming	*design	×	○
programming	*web	△	○
*programming	java	△	×
*programming	c	×	○
*programming	database	×	○
*shopping	amazon	○	○
*shopping	ecommerce	△	○
*shopping	online	×	○
*shopping	wishlist	×	○
*shopping	bargains	×	○

Table. 5: Relations of recently emergent significant tags

入力	正解	一致度	前後関係
web2.0	*blog	△	○
web2.0	*collaboration	○	○
web2.0	*web	○	○
web2.0	*development	△	○
*web2.0	ajax	○	×
iphone	*phone	×	○
iphone	*apple	○	○
iphone	*mobile	○	○
iphone	*browser	△	○
iphone	*ipod	×	○
vista	*windows	○	○
vista	*microsoft	○	○
vista	*xp	△	○
vista	*virtual	△	○
vista	*activation	○	×
ajax	*javascript	○	○
ajax	*web	○	○
ajax	*web2.0	○	×
ajax	*design	△	○
ajax	*html	○	○
rails	*web2.0	△	○
rails	*ruby	○	○
rails	*development	△	○
rails	*framework	○	○
rails	*agile	○	○

の共起度とドキュメントベースの共起度を比較することで、Synonymと呼ばれる同じ意味で用いられる語と高い精度で発見する手法を提案した。Mika[5]はSBMにおけるユーザとタグとドキュメントの関係を3部グラフの一種として定義し、そこからオントロジーを構築する可能性を示した。

5 おわりに

本稿では、時系列的な依存関係を定義し、その依存関係抽出手法を提案した。また手法によって抽出された依存関係を可視化するインターフェイスの提案および部分的な実装を行ない、直観的に有用な結果を抽出することに成功した。また依存関係の評価を行い、本手法の精度が高くなる状況についての知見が得られた。今後は、さらなる様々な条件下での実験、取得された依存関係の分類などを行っていく必要がある。

参考文献

- [1] Adar, E. Adamic, L. A. "Tracking Information Epidemics in Blogspace," In Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence, pp.207-214 (2005)
- [2] Dubinko, M. Kumar, Magnani, J. Novak, J. Raghavan P, Tomkins, A "Visualizing Tags over time," In Proceedings of the 15th International Conference on World Wide Web, pp.193-202 (2007)
- [3] Golder, S.A. HUBerman, B.A. "The Structure of Collaborative Tagging System," Journal of Information Science, 32(2), pp.198-208 (2005)
- [4] Kleinberg, J. "Bursty and Hierarchical Structure in Streams," In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Volume 7, Number 4, pp.373-397(25) (2002)
- [5] Mika, P. "Ontologies are us: A unified model of social networks and semantics," 4th International Semantic Web Conference, pp.522-536 (2005)
- [6] 丹羽智史, 土肥拓生, 本位田真一, "Folksonomy の 3 部グラフ構造を利用したタグクラスタリング," JAWS2006 (2006)
- [7] 丹羽智史, 土肥拓生, 本位田真一, "Folksonomy マイニングに基づく Web ページ推薦システム," 情報処理学会論文誌 Vol.47, No.5, pp.1382-1392 (2006)
- [8] Wang, R.C. Cohen, W.W. "Language-Independent Set Expansion of Named Entities using the Web," In Proceedings of the IEEE International Conference on Data Mining, pp.342-350 (2007)
- [9] del.cio.us. <http://del.cio.us/>
- [10] Google Trends <http://www.google.co.jp/trends>