

Web ニューストピックの時空間的偏在性の発見手法

山口 彰太[†] 是津 耕司[‡] 木俣 豊[‡] 田中 克己[†]

[†] 京都大学大学院情報学研究科社会情報学専攻 〒606-8501 京都府京都市左京区吉田本町

[‡] 情報通信機構 ユニバーサルコミュニケーション研究所 〒619-0289 京都府相楽郡精華町光台 3-5

E-mail: [†] {syamaguchi,tanaka }@dl.kuis.kyoto-u.ac.jp, [‡] {zetsu, kidawara}@nict.go.jp

あらまし Web ニュースは、時空間の選定方法により、その特徴が変化する。本稿では全 Web ニュースと選定した区間における情報の特徴が大きく異なることを情報に偏在性が存在すると定義した。さらに、Web ニュースの時空間モデリングを行い、偏在性を KL-divergence を用いる事で、検証する。さらに、偏在性が存在する任意の区間を求めるために、各区間における KL-divergence の値の関係について仮説をおき、連続区間、非連続区間における関係の利用方法と偏在性が存在する任意区間の算出方法、関係の分析手法について述べる。これらの関係を発見するために、2011年3月11日に発生した東北地震に関する世界各地の Web ニュースを対象として実験を行い、トピックや時空間において偏在性が存在することの検証を行う。

キーワード 時空間情報、情報の偏在性、確率分布、KL-divergence

1. はじめに

Web 上に存在する時空間情報を対象とした研究は数多く存在する。例えば、情報の即時性がある Twitter を用いた研究[1]では、地震に関する情報をソーシャルセンサーとして扱うことで、発信された時空間の軌跡を求めることで、地震の発生源を求めている。しかしながら、対象としている情報は時空間の選定方法により、どのような時空間に焦点を当てるかによって、情報が偏り、解釈や結果が大きく異なると考えられる。例えば、ユーザが 2011 年 3 月 11 日に発生した東北地震において全世界で話題となったトピックを知りたい場合を考える。この時、全世界で 2011 年 1 月 1 日から 12 月 31 日までに発信された全ての Web ニュースを対象とした場合、日本経済に与えた影響という話題が得られる。しかしながら、時間的な偏りとして、発生後 3 ヶ月程度の Web ニュースのみしか対象としない場合は、被災地の情報や原子力発電所といった、2011 年全体の話題とは異なる話題を得てしまう。また、空間的な偏りとして日本の Web ニュースのみであれば、被災地の情報や、原子力発電所などといった、時間的な偏りと同様にこちらも全世界の Web ニュースを対象とした場合とは異なる話題を得てしまうことが挙げられる。この時、ユーザが災害発生時から 3 ヶ月間に話題となったトピックや、日本において話題になったトピック等を対象としている場合は問題ないが、東北地震において 2011 年の世界全体で話題となったトピックを知りたい場合には異なる話題を得てしまうので、問題となる。

本稿では、時空間の選定方法により情報が偏ってしまうことを情報の偏在性が存在すると定義し、この現象を対象として、全 Web ニュースと選定された区間における Web ニュース間に偏在性が存在するかどうかを検出することと、偏在性が存在する区間を自動で発見することを目的としている。ユーザが利用している Web ニュースには全 Web ニュースと比べて情報の偏在性が存在している事をユーザに示すことで、ユーザは偏在性の存在する区間の情報ばかりを利用していないかを理解することができ、Web ニュースを正しく扱う事が出来ると考えられる。また、偏在性が存在する区間をユーザに示すことで、全 Web ニュースに存在する偏りがどのようなものであるかを分かりやすく伝えることが出来る。

本稿における技術的課題として、主に点挙げられる。1 点目として、Web ニュースの時空間モデリングである。これにより、Web ニュースを計算可能な形で表現する。このとき、各 Web ニュースが、いつ、どこで、何について書かれたものであるかを判定する必要がある。2 点目として、Web ニュース集合に存在する情報の偏在性の定義である。二つの Web ニュース集合がどのような状態である時偏在性が存在すると言えるのか指標を示し、計算可能な形で表現する必要がある。3 点目として、情報の偏在性が存在する任意の区間の発見である。全 Web ニュースから考えられる任意の区間における情報の偏在性を効率的に計算し、各区間における情報の偏りの値を比較する必要がある。

本稿では、各 Web ニュースが発信された時刻と発信

されたメディアから時空間情報を割り当てた。さらに、Web ニュースの内容をトピックとして扱うことで、各 Web ニュースの内容を割り当てた。次に、Web ニュースの総記事数から、各時空間で各トピックについて書かれている記事数を元に確率分布を用いて表現することで、Web ニュースの時空間モデリングを行った。次に、情報の偏在性を、全 Web ニュースを用いた確率分布と任意の区間に存在する Web ニュースを用いた確率分布間の違いを示す KL-divergence という指標を用いて定義した。また、偏在性の存在する任意の区間を発見するために、各区間における KL-divergence の値の関係について仮説を用いて説明し、連続区間における関係の利用と非連続区間における関係の利用について述べる。さらに、KL-divergence の値の関係を分析するための手法について述べる。これらの関係を検証するために、2011 年 3 月 11 日に発生した東北地震に関する Web ニュースを対象として実験を行い、区間の関係について考察した。

本稿の構成は以下の通りである。まず 2 節では関連研究の紹介と本稿との違いを述べる。3 節では、本稿の基本となる概念について定義と説明を述べる。4 節では、任意の区間における情報の偏在性の関係を連続区間と非連続区間に分けて考え、これらの関係を分析する手法について述べる。5 節では、2011 年 3 月 11 日に発生した東北地震をトピックとした Web ニュースを対象に行なった区間の関係を検出する実験の方法と結果について述べ、6 節で実験結果の考察と今後の課題について述べ、最後に 7 節で本稿のまとめを述べる。

2. 関連研究

2.1. 時空間情報を対象とした研究

時空間情報を対象とした研究として、主に情報の時空間による分析を行う研究が挙げられる。[1]では、マイクロブログの中でも情報の即時性が高い Twitter を用いたリアルタイムイベント検出を行なっている。この研究では、地震や台風に関する情報のみを対象として、各情報を時空間に配置することで、発信された時空間の変移を検出し、地震の発生源や台風の軌跡を自動で検出することを目的としている。この手法の利点として、リアルタイムで発生したイベントのみを対象としているため、イベントが発生した時空間に情報が集中しやすく、各単位時間に存在する情報の空間上の平均値を求めることで、ある程度の精度が見込める点である。この手法の欠点としてイベントそのものの検出ではなく、イベントが与えた社会的な影響などといった情報を扱くと、様々な時空間に情報が分散してしまうため、精度が極端に落ちてしまう事が考えられる。

[2]では検索エンジンのクエリログを用いて、検索ク

エリの時空間分析を行う事で、クエリの時空間的特徴とその変移を検出することを目的としている。この研究でも[1]と同様に、台風というクエリを検索した時間と IP アドレスから割り出された空間を時空間上に配置し、その変移を検出することで、台風の軌跡を検出している。また、野球というクエリを空間で分析し、クエリ数の違いと検索された野球チームの違いから、検索クエリを用いても時空間的特徴を検出する事が可能であることを示している。この手法の利点として、州ごとの特徴を検出したい場合には、州ごとにクエリ分析を行えば、容易に特徴を検出し、比較することができる点である。この手法の欠点として、州ではなく、複数の空間で同じ特徴が検出された場合、それらを統合して扱うことが出来ない点である。

[3]では、ブログを用いた時空間分析を行い、トピックの時空間的特徴とその変移を検出することを目的としている。この研究では、各トピックの出現頻度を用いるのではなく、出現確率から確率分布を算出することによって分析を行なっている。これにより、各空間のトピックの変移や、各時間によるトピック数の違いを知ることが出来る。この手法の利点として、各時間や各空間におけるトピックの出現確率の変移やその値から特徴が容易に検出できる点である。この手法の欠点として検出した様々な特徴の意味について、個々の時空間ではわかるものの、複数の時空間と比較した際の特徴の意味については分からない点である。

[4]では、2次元の地図に時間軸を追加した3次元上に情報を配置し、その分布から情報をグループ化し、その形を分類することにより、特徴を抽出することを目的としている。この手法の利点として時空間上に配置されているグループがどのような形をしているかにより、時空間的な特徴を視覚的に検出できる点である。この手法の欠点として情報の出現頻度のみに依存するため、潜在的に偏っているようなグループ化は行えない点である。

2.2. 空間を分割する手法

空間を分割する手法として、ボロノイ図を用いた手法が挙げられる。ボロノイ図を用いる場合、空間を2次元のユークリッド平面の地図としてみなし、各情報を発信された座標に配置することで、情報間の距離を計算し、複数の領域に空間を分割する事ができる。この手法の利点として、連続している空間における情報の偏りを各領域の範囲を参照することで、視覚的に検出することが容易であることが挙げられる。また、[5]では、発信された情報の空間の分布を示す手法として、Quad-Tree を用いている。Quad-Tree とは、ボロノイ図と同様に空間をユークリッド平面としてみなし、領域内に配置されている情報の数が閾値以上の場合は、領

域が等しくなるように格子状に4分割し、閾値未満の場合は分割を止めることで、各領域内に配置されている情報の数が全て閾値未満になるような分割手法である。この手法の利点として、ポロノイ図に比べ、各情報の距離を算出する必要がないため計算量が少なく済む点と、各領域間の範囲が分割した回数により定まるため、比較が容易である点が挙げられる。欠点として、格子状に分割するため、空間の特徴が離散している場合の検出には向かない点が挙げられる。

3. 基本概念

3.1. モデル化のためのパラメータの定義

本稿では、情報の時空間的偏在性を発見するために、時空間情報が付与されている文書のみを対象とする。この時、対象とする文書集合 D を $D = \{d_1, d_2, \dots, d_{|D|}\}$ として定義する。この集合に加え、各文書が書かれた時刻の集合である T を $T = \{t_1, t_2, \dots, t_{|T|}\}$ とし、各文書が書かれた場所の集合である L を $L = \{l_1, l_2, \dots, l_{|L|}\}$ として定義する。一つの文書 d_x には t_x と l_x がそれぞれひとつずつ割り当てられるものとする。さらに、各文書の内容を示すトピックの集合 Θ を $\Theta = \{\theta_1, \theta_2, \dots, \theta_{|\Theta|}\}$ として定義する。一つの文書 d_x には一つ以上の θ が割り当てられるものとする。これらの集合が存在する時、文書集合 D のそれぞれの要素について時刻や場所、トピックをラベル付けした集合 C は $C = \{\{d_1, t_1, l_1, \{\theta_1 \dots\}\}, \{d_2, t_2, l_2, \{\theta_2 \dots\}\}, \dots, \{d_n, t_n, l_n, \{\theta_k \dots\}\}, \}$ として表される。

3.2. Web ニュースのモデル化

本稿では、[3]の研究で用いられている Theme Lifecycle と Theme Snapshot という手法で、Web ニュースのモデル化を行なった。このモデルでは、時空間情報を出現頻度ではなく、各時間や各空間における出現確率の確率分布を用いている。出現確率を用いる利点として時空間ごとに異なる文書数にあまり影響されないため、複数の地域を比較する際には、情報の数に依存しない特徴検出が可能になる点である。欠点として出現頻度の少ない時空間の特徴は、個々の情報に強く影響されてしまう点である。

任意の時間、空間、トピックに関する確率分布は(1)式のように示される。

$$P(l, t, \theta) \quad (1)$$

(1)式から、各空間における確率分布を示した Theme Lifecycle というモデル化手法は(2)式のように示される。この式は、ある空間が一意に決められた場合に各トピックの時間による出現確率の変移を示した確率分布である。

$$P(t, \theta | l) \quad (2)$$

(1)式から、各空間における確率分布を示した Theme

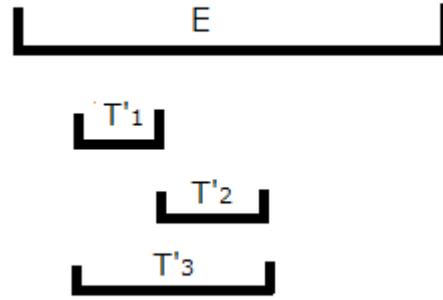


図 1: E, T'1, T'2, T'3 における関係

Snapshot というモデル化手法は(3)式のように示される。この式は、ある時間が一意に決められた場合に各空間による各トピックの出現確率の違いを示した確率分布である。

$$P(l, \theta | t) \quad (3)$$

3.3. KL-divergence

本稿では、二つの確率分布の違いを表す指標として、KL-divergence を用いた。P と Q をそれぞれ確率分布とし、P(i)を P の確率分布に従った i が出現する確率、Q(i)を Q の確率分布に従った i が出現する確率としたときの KL-divergence(D_{KL})は(4)式で求められる。

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (4)$$

KL-divergence の値は、二つの確率分布が違うほど大きな値を取り、類似しているほど小さな値を取ることが知られている。すなわち、任意区間において偏在が存在するときは、KL-divergence の値が大きくなる。

4. 任意の区間における関係の発見

本節では、任意の区間における関係の発見の必要性と、分析手法について述べる。図 1 に示されるような $E=[t_1, t_{|T|}]$, $T'_1=[t_i, t_j]$, $T'_2=[t_j, t_k]$, $T'_3=T'_1+T'_2=[t_i, t_k]$ という時間区間を考える。このとき、 T'_1 と E, T'_2 と E のそれぞれの KL-divergence の値と T'_3 の KL-divergence の値には、一般的に関係を得られない事が考えられる。例えば、E に A と B という二つの話題が均等に出現し、 T'_1 では A の話題のみ、 T'_2 では B の話題のみが出現した時を考える。このとき、E と T'_1 , E と T'_2 の確率分布は異なり、KL-divergence の値も大きくなる事が推定できるが、A と B の話題が存在する T'_3 の区間と E の確率分布には違いが現れないため、KL-divergence の値が小さくなる可能性がある。このように、 T'_1 と T'_2 からなる T'_3 であっても、 T'_1 と E, T'_2 と E の違いには必ずしも影響を与えない場合が存在する事が推定される。

しかしながら、特定の条件下において、 T'_1 と T'_2 の KL-divergence の値と T'_3 の KL-divergence の値には関係が存在する場合がある。例えば、E, T'_1 , T'_2 それぞれの時間区間全てに A と B という話題が均等に出現した

場合を考える。この時、 E と T'_1 、 E と T'_2 の確率分布が類似している時、それぞれの KL-divergence の値は小さくなる。また、 T'_1 と T'_2 から、 T'_3 であっても A と B の話題が均等に出現すると推定できるため、このときの E と T'_3 の KL-divergence の値も小さくなると推定できる。本稿では、特定の条件に依存する関係、及び依存しない関係の発見を行う。

4.1. 連続区間における関係の利用

本節では、連続区間における関係について述べる。仮に任意の区間における関係が存在しない場合を考える。このとき、時間区間集合 T から考えられる全ての時間区間は $[t_1, t_1]$, $[t_1, t_2]$, \dots , $[t_1, t_{|T|}]$, $[t_2, t_2]$, \dots , $[t_{|T|}, t_{|T|}]$ となる。すなわち、 $\sum_{n=1}^{|T|} n$ 個、 $\frac{1}{2}|T|(|T|+1)$ 個の確率分布を求め、それぞれに対する KL-divergence を計算する必要がある。一方、任意の区間における関係が存在する場合を考える。例えば、図 1 で示される E, T'_1, T'_2, T'_3 が $E=[t_1, t_{|T|}]$, $T'_1=[t_i, t_i]$, $T'_2=[t_j, t_j]$, $T'_3=T'_1+T'_2=[t_i, t_j]$ である時を考える。 T'_1 と T'_2 は単位時間区間として表現されている。このとき、 T'_3 の KL-divergence の値が T'_1 と T'_2 の値と関係が存在する時、新たに T'_3 の確率分布を計算すること無く、 T'_3 の KL-divergence の値を推測することが可能になる。仮に、全ての時間区間においてある関係が得られた場合、単位時間区間数 $|T|$ 個の確率分布と KL-divergence の値を求めるだけで、任意時間区間数 $\frac{1}{2}|T|(|T|+1)$ 個の KL-divergence の値を推測する事が可能になる。

4.2. 非連続区間における関係の利用

次に非連続区間における関係の利用について述べる。連続区間では、時間区間の場合、3月11日のみ、3月11日から4月11日までのような区間で示されたが、非連続区間の場合は、3月11日と4月11日のみを用いた確率分布を用いて KL-divergence を計算する必要がある。周期的に出現する話題に対しては、これらの非連続時間区間を求める必要がある。例えば、災害などに関するニュースの場合、発生一ヶ月後、二ヶ月後、...といったように、周期的にある話題が出現する。また、花粉症といったような話題も毎年春に多く出現するという周期性が存在する。また、地域区間の場合を考える。例えば、イランとイラクは地学上では連続した区間であるが、各地域で話題となる Web ニュースに関しては、イラン・イラク戦争などローカルに影響を与えた話題であれば共通して得られると推定できるが、世界経済の情勢などグローバルな話題に対しては、連続区間による関係はあまり存在しないと考えられる。また、日本とアメリカのように地学上距離のある地域であっても、共通の話題が出現するといった、非連続

な区間であっても関係が存在する場合もある。そのため、本稿では、地学上隣接している地域であっても、その地域を非連続区間として扱うこととする。

非連続区間に関係が存在しない時、時間区間集合 T から考えられる、連続区間も含めた任意の時間区間の数は、 $\sum_{i=1}^{|T|} |T|C_i$ 個となる。 $|T|=10$ である場合でも、連続区間における任意の時間区間数が 55 個であるのに対し、非連続区間においては、1013 個と、非常に多くの計算が必要になる。しかしながら、非連続区間に関係が存在するときは、この計算回数を減らすことが出来る。仮に、全ての時間区間においてある関係が得られた場合、単位時間区間数 10 個の確率分布と KL-divergence の値を求めるだけで非連続区間も含めた任意時間区間における KL-divergence の値を推定することが可能となる。

4.3. 任意の連続区間における関係の分析手法

現在分析手法に関しては、実験を行いながら検証を行なっている。そのため、分析手法は、最終論文にて掲載する予定である。

5. 実験

複数の時間区間における KL-divergence の値の関係を検出するために、本稿では 2011 年 3 月 11 日に発生した東北地震に関する世界中の Web ニュースを対象として実験を行なった。

5.1. 実験データ

本稿では、2011 年 3 月 11 日に発生した東北地震に関する Web ニュースを収集し、実験を行なった。まず、世界 18 地域と各地域を代表する Web ニュースメディアを手動で選定し、2011 年 3 月 11 日から 2011 年 7 月 1 日までに発信されたニュースを対象とした。また、各ニュースサイトで、東北地震をクエリとし、英語で書かれている記事のみをクロールした。対象とした地域、ニュースメディアの URL、取得記事数を表 1 に示す。さらに、Web ニュースに書かれている内容を推定するために、[6]の研究で挙げられているサブトピックの概念を用いて、東北地震の内容を示す代表的なサブトピックを手動で 8 つ選定した。以下に示す。

- 1 原子力発電所
- 2 復興への動き
- 3 経済への影響
- 4 被災者支援
- 5 計画停電
- 6 風評被害
- 7 被災規模
- 8 その他

これらのサブトピックを 3 月 11 日から 3 月 31 日までの記事 1566 件に対し、筆者が手動で割り当てた。この方法によって各文書に割り当てたサブトピックから、

表 1: 実験データの地域, URL, 取得記事数

地域名	ニュースメディア URL	取得記事数
New York	http://www.nytimes.com/	990
Los Angeles	http://www.latimes.com/	341
China	http://www.xinhuanet.com/english/	1020
Russia	http://www.itar-tass.com/en/	90
India	http://www.hindustantimes.com/	126
Japan	http://mdn.mainichi.jp/	440
Canada	http://www.vancouversun.com/	100
Germany	http://www.expatica.com/de/	132
France	http://www.expatica.com/fr/	200
Australia	http://www.theaustralian.com.au/	2955
Chile	http://www.ilovechile.cl/	19
Taiwan	http://www.chinapost.com.tw/	725
Iraq	http://www.npr.org/sections/iraq/	632
Korea	http://english.chosun.com/	220
Italy	http://www.agi.it/	217
New Zealand	http://www.stuff.co.nz/sunday-star-times/	912
Singapore	http://www.asiaone.com/A1Home/A1Home.html	881
South Africa	http://www.citypress.co.za/	95

各文書が複数のクラスに属する事を許したクラスを 8 個作成し, これを学習データとして, 4 月 1 日から 7 月 1 日までに生成された Web ニュースに対してクラスタリングを行うことで, 全文書にサブトピックの割り当てを自動で行なった. また, 各 Web ニュースメディアの記事を HTML タグによる分析を行うことで, 発信日時を抽出した. この発信日時を記事が生成された時間とし, 発信されたニュースメディアに対応する地域を生成された空間として, 各文書に時空間情報を割り当てた.

5.2. 各時間区間の KL-divergence の計算手法

本稿では, 任意の時間区間における KL-divergence の値を算出するために, 各時間区間における確率分布を全て計算し, 全時間区間である [2011/3/11, 2011/7/1] の確率分布を用いた. 本実験では, 最初に最小の時間区間である単位時間区間を 1 日単位として, [2011/3/11, 2011/7/1] として表される全体時間区間の 112 日間の中で考えられる全ての任意時間区間 6328 区間を求めた. 次に, 任意時間区間のそれぞれに対する確率分布を求め, 全体時間区間の確率分布を用いることで KL-divergence の値を算出した.

5.3. 実験結果

本節では, 実験により得られた各任意区間における KL-divergence の値の一部を表に示す. 実験結果は, 最終論文にて掲載する予定である.

6. 考察と今後の課題

本節では, 実験の結果, それぞれの時間区間につい

て, どのような関係が得られたかを簡潔に述べる. また, その関係の中でも, ある条件でのみ得られる関係について簡潔に述べる. さらに, これらの結果が得られた理由についてそれぞれ考察する. こちらも最終論文にて掲載する予定である.

本稿では, KL-divergence を用いることで, 各時間区間の関係の検出を行なったが, JS-divergence や各地域とサブトピックにおける出現確率の差の平均を用いるなどによる関係の検出を行うことも考えられる.

また, このように情報の偏りが発生している時間区間を求めることで, 情報の信憑性の検証と, 局所性の検出に用いる事ができると考えられる. 情報の信憑性の検証として, ユーザが閲覧や分析を行なっている Web ニュースの区間が全 Web ニュースと比べて, 偏りを発生していないかを確認し, 偏りが発生していれば発生している事をユーザに伝える事が挙げられる. ユーザが利用している Web ニュースは Web ニュース全体と比べて偏っているという本来ユーザが知り得なかった偏りを知ることで, 自身の利用している Web ニュースは全体の傾向と異なる事を理解した上で, 情報を正しく扱えるようになると考えられる. また, 局所性の検出として, 一般的なトピックであるのにも関わらず, 全体の Web ニュース集合と大きく異なるような区間を検出することで, その地域特有のトピックや, そのトピックが強く影響を与えている区間を知ることができる.

7. まとめ

本稿では, 時空間情報を扱う際に, 時空間の選定方法により発生する偏りを時空間的偏在性として定義した. この時, 時空間的偏在性が存在する任意の区間を求めるために, その関係について仮説をおき, 説明した. そのまた, この関係を分析するために, 2011 年 3 月 11 日に発生した東北地震に関する世界中の Web ニュース記事を用いて実験を行い, 関係の考察と, 時空間的偏在性の応用例について述べた.

謝 辞

本研究の一部は, グローバル COE 拠点形成プログラム「知識循環社会のための情報学教育研究拠点」(研究代表者: 田中克己) によるものです. ここに記して謝意を表します.

参 考 文 献

- [1] Takeshi Sakai, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World Wide Web (WWW '10). ACM., New York, NY, USA, 851-860.
- [2] Lars Backstrom, Jon Kleinberg, Ravi Kumar, and

- Jasmine Novak. 2008. Spatial variation in search queries. In Proceeding of the 17th international conference on World Wide Web (WWW'08). ACM, New York, NY, USA, 357-366.
- [3] Qiaozhu Mei, Chao Liu, Hang Su, and ChengXiang Zhai. 2006. A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In Proceeding of the 15th international conference on World Wide Web (WWW'06). ACM, New York, NY, USA, 533-542.
- [4] Kyoung-Sook Kim, Koji Zettsu, Yutaka Kidawara, Yasushi Kiyoki. 2010. StickViz: A New Visualization Tool for Phenomenon-Based k-Neighbors Searches in Geosocial Networking Services. In Proceedings of the 2010 12th International Asia-Pacific Web Conference (APWEB'10).IEEE Computer Society, Washington, DC, USA, 22-28.
- [5] Tatsuya Fujisaka, Ryong Lee, and Kazutoshi Sumiya, 2009. Exploring Refional Characteristics Using the Movement History of Mass Mobile Microbloggers. 情報処理学会研究報告. データベース・システム研究会報告, 1-8.
- [6] Yoonjung Choi, Yuchul Jung, and Sung-Hyon Myaeng. 2010. Identifying Controversial Issues and their Sub-topics in News Articles. In Proceedings of Pacific Asia Workshop on Intelligence and Security Informatics 2010, 140-153.