

対象トピックへの関心度の高いブロガーコミュニティ判定方式

小倉 僚[†] 朝永 聖也[†] 中島 伸介[†] 張 建偉^{††} 稲垣 陽一^{†††}

中本 レン^{†††}

[†] 京都産業大学コンピュータ理工学部 〒603-8555 京都府京都市北区上賀茂本山

^{††} 筑波技術大学 産業技術学部 〒305-8520 茨城県つくば市天久保4-3-15

^{†††} 株式会社きざしカンパニー 〒103-0015 東京都中央区日本橋箱崎町20-14 日本橋巴ビル6F

E-mail: [†]{g1044228,g0946920,nakajima}@cse.kyoto-su.ac.jp, ^{††}zhangjw@a.tsukuba-tech.ac.jp,

^{†††}{inagaki,reyn}@kizasi.jp

あらまし 対象トピックの背景情報を把握するため、対象トピックに興味・関心を持つコミュニティについて調査することは有用である。そこで、対象トピックへの興味・関心の強いブロガーコミュニティを自動判定する方式を提案する。対象トピックに対する関心度の高いブロガーコミュニティを特定する方法としては、これまでの著者らの研究においてコミュニティ総人数に対する、そのコミュニティ内での発言ブロガー数の割合を調査する手法を提案している。しかしながら、対象トピックによっては判定できないケースが数多く存在するため、本稿では、対象トピックの共起語集合と、ブロガーコミュニティの関連語辞書との類似度に基づいて判定する手法を提案すると共に、従来手法との比較に基づく評価を行う。

キーワード ブログマイニング, ブロガーコミュニティ判定, トピック推定

Method for Identifying Blogger Communities with a High Degree of Interest for a Target Topic

Tsukasa OGURA[†], Seiya TOMONAGA[†], Shinsuke NAKAJIMA[†], Jianwei ZHANG^{††}, Yoichi

INAGAKI^{†††}, and Reyn NAKAMOTO^{†††}

[†] Faculty of Computer Science and Engineering, Kyoto Sangyo University Motoyama, Kamigamo, Kita-ku, Kyoto-City, Kyoto, 603-8555 Japan

^{††} Faculty of Industrial Technology, Tsukuba University of Technology Amakubo 4-3-15, Tsukuba-City, Ibaraki, 305-8520 Japan

^{†††} Kizasi Company, Inc. 6th floor Tomoe Nihonbashi Building, 20-14 Hakozaiki-Cho, Nihonbashi, Chuo-ku, Tokyo, 103-0015 Japan

E-mail: [†]{g1044228,g0946920,nakajima}@cse.kyoto-su.ac.jp, ^{††}zhangjw@a.tsukuba-tech.ac.jp,

^{†††}{inagaki,reyn}@kizasi.jp

Abstract In order to know about the target topic, it is effective to investigate about community with interests in the target topic. Therefore, we propose the method to determine the blogger community with interests in the target topic. As a method of determining the blogger community with interests in the target topic, there is a method based on the percentage of number of the blogger is saying by community for the total number of people by community, but the method may not be identified by the type of target topic. So, in this paper, we report investigates the method to determine the blogger community with interests in the target topic based on examining the overlap of word co-occurrence of target topic and word co-occurrence of community bloggers.

Key words blog mining, identifying blog community, topic detection

1. はじめに

近年、ブログサービスが普及し、一般ユーザでも手軽に自分の意見をインターネットを通じて公開することが可能となった。CGMと呼ばれる、このようなユーザ発信型コンテンツは、世論を強く反映したものであり、これら进行分析することで、ユーザの生の声を把握することも可能である。実際に自社製品に対するユーザの反応を調査する目的で、多くの企業がブログを対象とした評判分析を行っている。

我々もこれまでに、ブログ分析に基づき流行語の候補語の早期発見を目指した研究に取り組んでいる [1], [2]。この研究では、対象トピックへの関心度の高いブロガーコミュニティの判定を精度良く行う必要があるが、現状では十分とはいえない。この対象トピックに対する関心度の高いブロガーコミュニティを特定する方法としては、あるコミュニティ総人数に対する、そのコミュニティ内の発言ブロガー数の割合を分析する手法を採用している。すなわち、コミュニティ内の多くのブロガーがその対象トピックについて言及しているのであれば、このブロガーコミュニティは対象トピックへの関心度が高いと判定するものである。しかしながら、対象トピックによっては精度良く判定できないケースが数多く存在するため、本稿では、対象トピックの共起語集合と、ブロガーコミュニティの関連語集合の類似度に基づいて、判定する手法を提案する。

あるトピックへの関心度の高いブロガーコミュニティを判定する手法は、我々が取り組んでいる流行語候補の早期発見手法に役立つだけでなく、例えば、あるトピックに関して関心が高いブロガーコミュニティを把握することができれば、どのようなコミュニティに商品を訴求すればよいかということにも応用することができる。以上を踏まえ、本稿では対象トピックへの関心度が高いブロガーコミュニティの判定方式を提案すると共に、その有効性について評価する。

以下、2章にて関連研究について述べる。3章にてブロガーコミュニティの特定方法について述べる。4章にて対象トピックへの興味・関心の強いブロガーコミュニティの判定方式について述べる。5章では従来手法と提案手法の比較評価について述べる。最後に6章で、まとめと今後の課題について述べる。

2. 関連研究

ブログ等の分析に基づくコミュニティ判定に関する関連研究として以下が挙げられる。

横山ら [4] は、潜在的ディリクレ文法を用いて投稿記事のトピック推定を行い、正確な情報伝播をとらえたネットワークを抽出する枠組みを提案している。松永ら [5] は、ブログ記事から「興味の対象」を示すキーワードを抽出し、クラスタリングによりブロガーコミュニティを抽出している。谷口ら [6] は、PageRank と Betweenness Clustering によりコミュニティの抽出と分析を行っている。内田ら [7] は、Trackback を辿ることで Weblog 記事のネットワークを取得し、クラスタリングにより各記事の本文の特徴語を抽出する手法を行なっている。大塚ら [8] は、HITS アルゴリズムでは抽出しにくいページをオー

ソリティ、ハブに加え、中間ノードを導入した3階層以上の多階層のリンク構造を持つウェブコミュニティ抽出法を提案している。

以上の通り、コミュニティの抽出や、コミュニティの分析に関する研究は行われているが、コミュニティの抽出と分析に基づく対象トピックに対するコミュニティの関心度を算出する研究は行われていない。

3. 潜在的ブロガーコミュニティの特定

本研究で扱うブロガーコミュニティは、ブロガーの意識にかかわらず、同じトピックについて興味や関心を持つブロガー集合を指す。したがって、SNS 等でユーザが明示的に登録するようなコミュニティとは異なり、潜在的なブロガーコミュニティを特定する必要がある。また、提案手法では、対象トピックの共起語集合とブロガーコミュニティの関連語辞書の比較から、関連の深さを調べるため、コミュニティ内のブロガーが記事に書かれたキーワードを重要視し、ブロガー同士の面識等は考慮しない。よって、過去に投稿したブログ記事の履歴から、ブロガー毎の興味度の強さを分析することで、潜在的ブロガーコミュニティを特定する。なお、このコミュニティ特定手法は、著者らが過去に行った技術「ブロガーの熟知度分析 [3]」に基づいている。以下にこの技術について説明する。

3.1 潜在的ブロガーコミュニティ名および関連語辞書の作成

潜在的なコミュニティ名は、自動抽出と拡充した独自シソーラスに基づき、ブログでよく言及されるトピックを用いて作成する。作成した各潜在的コミュニティ名との共起度が高い400語の語句を抽出する。なお、共起語の選定には自らの生活体験を表すような語句を優先的に採用し、不適切な語句を排除することにより、実体験に即したブログ記事を記述するブロガーを分類できる精度を上げている。

次に、潜在的ブロガーコミュニティ毎に共起語で構成された関連語辞書を作成する。各潜在的ブロガーコミュニティに対して、一定期間内のブログ記事を対象とし、そのコミュニティ名であるキーワードとの共起度が高い400語の語句を抽出する。なお、共起語の選定は半自動で行って、共起語候補を自動的に抽出した後に、最終的にはすべての辞書に対して人間の目で確認し、共起語として不適切な語句を排除しているため、関連語辞書の選定精度は100%と考えて良い。

共起語の選定に関しては、自らの生活体験を表すような語句を優先的に採用することで、実体験に即したブログ記事を記述するブロガーを分類しやすくしている。具体的には、商用ブログの多い特定のカテゴリに対して、生活者の実体験を示す語彙を増やすために、共起条件を手動で追加した。例えば、「温泉」のカテゴリに対して、「温泉+行った」「温泉+宿泊した」などの共起条件を用いることで、実際に温泉に行ったブロガーが書いた記事から実体験に即する語彙を抽出した。手動で設定した共起条件の数は、カテゴリごとに大きく異なるが、商用ブログやスパムブログの多い「温泉」カテゴリのケースでは、50個の共起条件を設定している。

3.2 ブLOGGER興味度スコアに基づくコミュニティの特定

興味度スコアの算出方法を説明する。基本的な考え方としては、対象となる潜在的BLOGGERコミュニティに関連するキーワードを含んだ記事の投稿数に基いて算出する。なお、各BLOGGERは潜在的BLOGGERコミュニティごとに異なる複数の興味度スコアを有する。つまり、あるBLOGGERが「経済」と「政治」に関する潜在的BLOGGERコミュニティに属する場合、このBLOGGERは「経済」に対する興味度スコアと「政治」に関する興味度スコアを別々に有することになる。ここで、対象コミュニティ g_i に対する、あるブログ記事 e_k の関連度スコアを $relevance_{g_i}(e_k)$ とすると、以下のように表すことができる。

$$relevance_{g_i}(e_k) = \sum_{j=1}^n \alpha_{ij} \cdot \beta_{ji} \cdot \gamma_{ij} \quad (1)$$

ただし、 n はこの潜在的BLOGGERコミュニティ g_i の共起語数であり、今回は $n = 400$ である。 α_{ij} は潜在的BLOGGERコミュニティ g_i の共起度順位 j 番目の共起語 ω_{ij} の重みであり、 $\alpha_{ij} = (n - j + 1)/n$ で表される。これは、各共起語の共起度以上に、共起度順位の高い語句の重みを大きくするためのものである。共起度順位1位の重みは $400/400$ 、2位の重みは $399/400$ となり、400位の重みは $1/400$ となる。 β_{ij} は潜在的BLOGGERコミュニティ g_i の j 番目の共起語 ω_{ij} の共起度である。そして、 γ_{ij} は順位 j 番目の共起語 ω_{ij} が該当記事 e_k 内に存在するかどうかを表現する変数であり、存在する場合1、存在しない場合0の値をとる。

共起度の算出法としては、単純頻度、 t スコア、 MI スコア、 $LogLog$ スコアなど多くの尺度が提案されている。単純頻度では、常識的なものを抽出するのに対して、特徴的な語を上位におく t スコアや MI スコアでは、納得できるごがなくなる傾向がこれまでに行った実験で見られた。そのため、本手法では、それらの中間の尺度 $LogLog$ スコアを採用している。ブログ記事の相互数を N とし、キーワード x と周辺語 y の出現回数をそれぞれ N_x と N_y とする。 x と y の共起回数を N_{xy} とすると、 $LogLog$ スコアの算出式は下記である。

$$LogLog \text{ score} = \log \frac{N_{xy} \cdot N}{N_x \cdot N_y} \cdot \log N_{xy} \quad (2)$$

次に、対象潜在的BLOGGERコミュニティ g_i に対するBLOGGER b の興味度スコアを $knowledge_{g_i}(b)$ とすると、以下のように表すことができる。

$$knowledge_{g_i}(b) = \frac{l}{n} \cdot \frac{\log(m)}{m} \cdot \sum_{k=1}^m relevance_{g_i}(e_k) \quad (3)$$

ただし、 e_k はBLOGGER b が投稿した記事である。 m はBLOGGER b が対象期間内に投稿した記事数である。 l はBLOGGER b が対象期間内に投稿した記事に出現した共起語数である($l \leq n$)。したがって、 l/n はBLOGGER b が使用した共起語の全共起語に対する網羅率である。 $\log(m)/m$ では、関連性の低い記事を大量に投稿した場合に、そのBLOGGERの興味度が高くなってしまいう問題に対して、記事数の増加の影響を緩和させている。最終的に対象となる潜在的BLOGGERコミュニティに対する興味

度スコアが、設定した閾値を超えれば、その潜在的BLOGGERコミュニティはそのBLOGGERが属するものと判定する。

4. 対象トピックへの興味・関心の強いBLOGGERコミュニティの判定方式

本方式では、対象トピックの共起語集合とBLOGGERコミュニティの関連語辞書の重なりに基いて対象トピックと関連の深いBLOGGERコミュニティを判定する。以下、4.1節にて、BLOGGERコミュニティの中領域グループの再構成、4.2節にて、対象トピックに対する関心度の算出について述べる。

4.1 BLOGGERコミュニティの中領域グループの再構成

「BLOGGERの熟知度分析 [3]」で作成されたBLOGGERコミュニティは約13,000件の小領域グループが存在するが、各グループのトピックが細かすぎるため、小領域グループをさらにグループ化することで、中領域グループを形成し、この中領域グループレベルでのBLOGGERコミュニティ関連語辞書を作成する。なお、どのように小領域グループを中領域グループへ対応を行うのかは、「ブログの体験熟知度に基づくブログランキングシステム [3]」で使用した分類を用いる。

関心度を求めるパターンとしては、対象トピックの共起語集合と小領域グループ時点でのBLOGGERコミュニティの関連語辞書の関心度を算出し、中領域グループごとに集計した平均値を、その中領域グループレベルでの関心度とする場合と、対象トピックの共起語集合と作成した中領域グループのBLOGGERコミュニティの関連語辞書で関心度を算出する場合を検討する。

中領域グループのBLOGGERコミュニティの関連語辞書の作成手法は、以下の4パターンを検討する。

関連語辞書作成パターン 1

中領域グループレベルごとに、対象小領域グループのBLOGGERコミュニティの関連語辞書のキーワードの出現回数を求め、出現回数の多い上位400語をその中領域グループでのBLOGGERコミュニティの関連語辞書とする。

関連語辞書作成パターン 2

中領域グループレベルごとに、対象小領域グループのBLOGGERコミュニティの関連語辞書のキーワードの共起度を求め、共起度の高い上位400語をその中領域グループでのBLOGGERコミュニティの関連語辞書とする。

関連語辞書作成パターン 3

中領域グループレベルごとに、対象小領域グループのBLOGGERコミュニティの関連語辞書のキーワードの共起度を求め、各対象小領域グループの共起度順位の高い順に、その中領域グループに所属する対象小領域グループの数を考慮して選出した400語を、その中領域グループでのBLOGGERコミュニティの関連語辞書とする。

関連語辞書作成パターン 4

関連語辞書作成パターン1および関連語辞書作成パターン2で作成した各400語のうち、それぞれの上位200語を抽出して作成した関連語辞書を、中領域グループでのBLOGGERコミュニティの関連語辞書とする。

4.2 ブロッガーコミュニティの対象トピックに対する関心度の算出

本方式では、対象トピックの共起語集合とブロッガーコミュニティの関連語辞書の重なりに基づいて関心度を算出する。具体的には、2つのキーワード群がどの程度類似しているかを関心度として算出する。例えば「AKB48」トピックの共起語として「アイドル」や「歌手」が得られた場合、関連の深いブロッガーコミュニティ（中領域グループ）である「アーティスト」や「女性スター」の関連語辞書内にも「アイドル」や「歌手」といったキーワードを含んでいる可能性が高いと考えられるためである。対象トピックの共起語集合、ブロッガーコミュニティの関連語辞書における共起度ランキングにはそれぞれ表1に示すような順位点を用いた。なお、関心度の高さは、以下で表現されるコサイン類似度を使用する。

$$Sim(\vec{T}, \vec{C}) = \frac{\sum_{i=1}^N (t_i \cdot c_i)}{|\vec{T}| |\vec{C}|} \quad (4)$$

ただし、 $\vec{T}$ は対象トピックの共起語集合の順位点を値とした特徴ベクトル、 $\vec{C}$ はブロッガーコミュニティの関連語辞書の順位点を値とした特徴ベクトルである。 t_i はベクトル $\vec{T}$ における i 番目の要素、 c_i は $\vec{C}$ における i 番目の要素である。 N は $\vec{T}$ 及び $\vec{C}$ の共通するキーワード数であり、今回は 0~400 である。

表1 関心度で用いる共起度ランキングに対する順位点算出方法

共起度順位	順位点
1	400
2	399
3	398
⋮	⋮
399	2
400	1

5. 対象トピックと関連度の高いコミュニティ判定方式に関する評価実験

評価実験として、「対象トピックの共起語集合とブロッガーコミュニティの関連語辞書の重なりを調査する手法」（提案手法）と、「コミュニティ総人数に対する、そのコミュニティ内での発言ブロッガー数の割合を調査する手法」（従来手法）を用いて、6つのトピック（女子会/AKB48/K-POP/スマートフォン/Android/Facebook）に対する関連の深いブロッガーコミュニティの判定結果（ランキング）を比較する。なお、提案手法では、対象トピックの共起語集合と小領域グループ時点でのブロッガーコミュニティの関連語辞書の関心度を算出し、中領域グループごとに集計した平均値によってランキングを算出する。

比較評価尺度として、対象トピックと関連が深いブロッガーコミュニティの理想ランキングの上位 10 件をどの程度含めることができたかを示す適合率と、 DCG 及び $nDCG$ を算出する。 DCG は減損累積利得と呼ばれ、順位を含めて正解データのランキングをどれだけ再現できるのかを評価する。すなわち、単純に理想ランキングに含まれるだけではなく、順位も合致し

ている程高く評価できる指標である。なお、 $nDCG$ とは、正規化減損累積利得であり、理想ランキングの DCG の値を 1 として正規化したときの値である。また、 DCG では表1に示すような利得スコアを用いた。 DCG および $nDCG$ の算出式を以下に示す。

$$DCG_p = rel_1 + \sum_{i=2}^p \frac{rel_i}{\log_2(i+1)} \quad (5)$$

$$nDCG_p = \frac{DCG_p}{IDCG_p} \quad (6)$$

ただし、 rel_i は、ランキング i 番目のアイテムの利得スコアである。また、 p は、ランキングを考慮する件数を示しており、今回の実験では $p = 10$ である。なお、 $IDCG_p$ は、理想的な DCG_p の値を示す。すなわち、ランキング上位であれば、高い重みでスコアを加算することができるため、スコアの高いランキング上位のアイテムを上位にランキングできれば DCG の値は高くなる。

表2 DCG で用いる利得スコア算出方法

順位	利得スコア
1	10
2	9
3	8
⋮	⋮
9	2
10	1

5.1 アンケート調査に基づく理想ランキングの作成

対象トピックと関連が深いブロッガーコミュニティの理想ランキングを作成するために、被験者 13 人のアンケート調査を行った。アンケートでは、従来手法と提案手法で得られた各ブロッガーコミュニティ 30 件、計 60 件の中から重複を除いたブロッガーコミュニティより、対象トピックと関連が深いと考えられるブロッガーコミュニティを 1 位から 10 位まで選出してもらう。ここで、被験者 13 人によって選出されたブロッガーコミュニティの順位の平均が高い順に集計し、その上位 10 件を実験で使用する理想ランキングとした。

5.2 各ランキングの評価

理想的なランキング、提案手法によるランキング、従来手法によるランキングのそれぞれを表3~8に示す。下線は、理想ランキングに適合していないブロッガーコミュニティを示す。また、6つの各トピック（女子会/AKB48/K-POP/スマートフォン/Android/Facebook）に対する比較評価結果を表9,10に示す。

表9, 10より、「スマートフォン」トピック、および「Android」トピックでは従来手法よりも、提案手法の方が良好な結果が得られた。また、「AKB48」トピック、「K-POP」トピック、および「Facebook」トピックでは、提案手法の適合率は高くないが、 $nDCG$ の値が高く、これは、理想ランキングの上位がうまく適合していることが分かる。しかし、「女子会」トピックでは、どちらの手法にも良い結果が得られなかった。これは、「女子

会」トピックと関連が深いと考えられる「恋愛・結婚」コミュニティや「美容」コミュニティに対し、「魚介」コミュニティや「アウトドア」コミュニティなど、関連が深いと考え難いプロガーコミュニティが出現しているため、「女子会」トピックについて知っているプロガーに、コミュニティの偏りが少なかった可能性、または、アンケートに基づく理想ランキングの作成において、被験者に偏りがあった可能性が考えられる。

表 3 「女子会」トピックにおける各ランキング結果

順位	理想ランキング	従来手法	提案手法
1	恋愛・結婚	魚介	酒
2	美容	買い物	家族
3	スイーツ	紅茶	グルメ
4	酒	コーヒー	顔文字
5	仕事・バイト	アウトドア	中華料理・韓国料理
6	ファッション	写真	恋愛・結婚
7	グルメ	顔文字	アウトドア
8	学校生活	キッズ	コーヒー
9	趣味	手芸・工芸	日本の地域
10	ダイエット	フルーツ	紅茶

表 4 「AKB48」トピックにおける各ランキング結果

順位	理想ランキング	従来手法	提案手法
1	女性スター	お笑い	女性スター
2	音楽	女性スター	アーティスト
3	アーティスト	日本の業界	音楽
4	TV・ラジオ	男性スター	男性スター
5	インターネット	買い物	韓国スター
6	著名人	TV・ラジオ	TV・ラジオ
7	日本の業界	アウトドア	お笑い
8	ファッションモデル	アーティスト	舞台
9	写真	光学機器	映画
10	ファッション	趣味	アニメ・コミック

表 5 「K-POP」トピックにおける各ランキング結果

順位	理想ランキング	従来手法	提案手法
1	韓国スター	韓国スター	韓国スター
2	アーティスト	お笑い	音楽
3	音楽	語学	アーティスト
4	女性スター	男性スター	女性スター
5	男性スター	光学機器	TV・ラジオ
6	TV・ラジオ	女性スター	世界の地域
7	著名人	アーティスト	舞台
8	エンタメ	日本の業界	お笑い
9	インターネット	買い物	男性スター
10	ファッション	映画	映画

6. おわりに

本論文では、対象トピックへの関心度の高いプロガーコミュニティ判定方式として、対象トピックの共起語集合 400 語と、プロガーコミュニティの関連語辞書の共起度の上位 400 語の重なりを調査する手法を提案した。また、コミュニティ総人数に対する、そのコミュニティ内での発言プロガー数の割合を調査

表 6 「スマートフォン」トピックにおける各ランキング結果

順位	理想ランキング	従来手法	提案手法
1	携帯電話	買い物	携帯電話
2	パソコン	アウトドア	パソコン
3	インターネット	自動車	インターネット
4	カメラ	魚介	カメラ
5	コンピュータ	DIY	AV・情報家電
6	写真	写真	コンピュータ
7	AV・情報家電	仕事・バイト	ゲーム
8	ゲーム	光学機器	ゲーム機
9	生活家電	顔文字	写真
10	音楽	日本の業界	生活家電

表 7 「Android」トピックにおける各ランキング結果

順位	理想ランキング	従来手法	提案手法
1	携帯電話	コンピュータ	携帯電話
2	インターネット	AV・情報家電	パソコン
3	パソコン	日本の業界	コンピュータ
4	コンピュータ	携帯電話	インターネット
5	カメラ	ゲーム	AV・情報家電
6	AV・情報家電	自動車	ゲーム機
7	写真	お笑い	カメラ
8	ゲーム	買い物	著名人
9	音楽	パソコン	ゲーム
10	生活家電	ゲーム機	日本の業界

表 8 「Facebook」トピックにおける各ランキング結果

順位	理想ランキング	従来手法	提案手法
1	インターネット	日本の業界	インターネット
2	写真	買い物	携帯電話
3	パソコン	魚介	パソコン
4	携帯電話	アウトドア	写真
5	世界の地域	写真	顔文字
6	学校生活	語学	恋愛・結婚
7	仕事・バイト	仕事・バイト	語学
8	カメラ	家計	職業・職種
9	社会	芸術	家計
10	行事・季節	動物	著名人

表 9 理想ランキングに対する従来手法と提案手法の適合率

検証トピック	理想ランキング	従来手法	提案手法
女子会	1	0	0.3
AKB48	1	0.4	0.4
K-POP	1	0.4	0.6
スマートフォン	1	0.1	0.9
Android	1	0.5	0.7
Facebook	1	0.2	0.4
平均	1	0.267	0.55

する手法との比較によって、提案手法の有用性を示した。

本方式では、プロガーコミュニティの関連語辞書に含まれるキーワードによって、精度が変わるため、今後は、未検証の関連語辞書作成パターンで実験を行う予定である。

7. 謝 辞

本研究の一部は、文部科学省科学研究費助成事業 (学術研究

表 10 理想ランキングに対する従来手法と提案手法の $nDCG$

検証トピック	理想ランキング	従来手法	提案手法
女子会	1	0	0.378
AKB48	1	0.505	0.745
K-POP	1	0.534	0.881
スマートフォン	1	0.055	0.984
Android	1	0.588	0.908
Facebook	1	0.150	0.749
平均	1	0.305	0.774

助成基金助成金) 基盤研究 (C)(課題番号:#23500140) による.
ここに記して謝意を表します.

文 献

- [1] Shinsuke Nakajima, Jianwei Zhang, Yoichi Inagaki and Reyn Nakamoto. Early Detection of Buzzwords Based on Large-scale Time-Series Analysis of Blog Entries, 23rd ACM Conference on Hypertext and Social Media (ACM Hypertext 2012), pp.275-284, June 2012.
- [2] 中島伸介, 張建偉, 稲垣陽一, 中本レン, 大規模なブログ記事時系列分析に基づく流行語候補の早期発見手法, 情報処理学会論文誌:データベース (TOD56), 2013 年.
- [3] 稲垣陽一, 中島伸介, 張建偉, 中本レン, 桑原雄, ブロガーの体験熟知度に基づくブログランキングシステムの開発および評価, 情報処理学会論文誌:データベース, Vol. 3, No. 3(TOD47), pp. 123-134, 2010 年.
- [4] 横山 正太郎, 江口 浩二, 大川 剛直, 潜在トピックを用いたブログ空間からの情報伝搬ネットワーク抽出 (情報検索 Web 情報システム,[特集] データ工学論文), 電子情報通信学会論文誌. D, 情報・システム J93-D(3), 180-188.
- [5] 松永拓, 平手勇宇, 山名早人, キーワードの出現に基づくブログコミュニティ抽出とオピニオンリーダーの発見, DEWS2007, C3-7
- [6] 谷口 智哉, 松尾 豊, 石塚 満, Blog コミュニティの抽出と分析, 第6回 セマンティックウェブとオントロジー研究会, 人工知能学会研究会資料, SIG-SWO-A401-08
- [7] 内田 誠, 柴田尚樹, ブログ記事ネットワークからの emerging topic の抽出と可視化, 人工知能学会全国大会 2006, 3D2-03
- [8] 大塚 浩司, 大町 真一郎, 阿曾 弘具, リンク構造を用いたウェブコミュニティ抽出法, IEICE technical Report, DE2006-104(2006-7)