

多様な視点を用いた CGM 記事の可視化手法の提案

室田 邦雄[†] 佐藤 哲司^{††}

[†] 筑波大学図書館情報専門学群 〒305-8550 茨城県つくば市春日 1-2

^{††} 筑波大学院図書館情報メディア研究科 〒305-8550 茨城県つくば市春日 1-2

E-mail: ^{††}satoh@slis.tsukuba.ac.jp

あらまし 近年、個人の意見を発信するブログや生活の知恵を共有する質問回答サイトなど生活者情報が増えている。このような情報を理解するには、著者や対象読者が誰か、どのような状況で書かれたのかなど、記事が投稿された時のコンテキストを考慮する必要がある。本論文では、対象とする読者の集団であるコミュニティを視点として、投稿記事を配置・可視化する手法を提案する。提案法は、利用者が与える多様な視点の重みを用いて記事の配置を動的に再計算することで、特定のコミュニティに関連が強い、あるいはコミュニティの影響を強く受ける記事を把握できる。複数のカテゴリからなる質問回答サイトを対象に、個々のカテゴリに投稿された質問と回答記事をコミュニティに対応させて、複数のカテゴリを視点とした評価システムを実装し、利用者がインタラクティブに記事を再配置する提案法の有効性を確認した。

キーワード 可視化, 再配置, CGM

Interactive Visualization Method for Consumer Generated Media based on the user's view points

Kunio MUROTA[†] and Tetsuji SATOH^{††}

[†] School of Library and Information Science University of Tsukuba,
Kasuga 1-2, Tsukuba, Ibaraki, 305-8550 Japan

^{††} Graduate School of Library, Information and Media Studies, University of Tsukuba,
Kasuga 1-2, Tsukuba, Ibaraki, 305-8550 Japan

E-mail: ^{††}satoh@slis.tsukuba.ac.jp

Abstract Recently, consumer generated media such as Blogs and the question and answer sites has increased. To understand such CGM contents, it is necessary to consider the context i.e. when and who wrote the article, who can read it, and so on. In this paper, the community that is the potential readers of the article is assumed to be an aspect, and it proposes the method for arranging and visualizing the article. The proposal method can understand the article that the relation is strong in a specific community or the community strongly influences by dynamically calculating the arrangement of the article again by using the weights of various aspects that the user gives. For the question and answer site that consists of many categories, question and answer article contributed to individual category was made to correspond to the community, and the evaluation system in which two or more categories were assumed to be an aspect was mounted, and the effectiveness of the proposal method of the user's interactively relocating the article was confirmed.

Key words Visualize, relocation, CGM

1. はじめに

近年、個人の意見を発信するブログや生活の知恵を共有する質問回答サイト、匿名の掲示板などの生活者情報 (Consumer Generated Media, 以下 CGM と称す) が増えている。CGM

による情報共有・情報交換は、典型的な不具合の解決法を網羅したマニュアルでは解決できないような、個別的で複雑な問題の解決に極めて有効である。例えば、PC の不具合は CPU の型番や OS のバージョン、その他様々な要素によって起こる。このため、典型的な不具合への対処法を記したマニュアルでは

解決できないことも多く、自分と同じ環境で解決できた事例が参考となる場合が多い。このような生活の知恵を共有することを目的とした質問回答サイトに症例や実行環境などを併せて記事を書き込むことで、同じ不具合に遭遇した人や知識ある人が誰でも解決策を教えることができるようになっている。

CGM は既存のマスメディアの代替として大衆に向けての情報発信という役割だけでなく、同じ興味を持つなどの不特定少数を含む仲間内でのコミュニケーションツールとしての役割も持つ。そのため、すでに社会的役割が明確になっているマスメディアと違い、その内容や意図、対象などが複雑で、仲間内では通用しないような記事も頻繁に見受けられる。例えば「下げる^(注1)」という言葉が特別な意味で使われるなど、仲間内すなわちそのコミュニティに精通した人たち以外は理解が困難であることもよく見られる。

このため、CGM では従来からそのコミュニティに属していたのではない人が求める情報に到達するのは困難であり、有用な情報を見逃してしまう可能性も無視できない。したがって、CGM を従来からそのコミュニティに属していたのではない人が活用できるようにすることが CGM の利用価値を高めると確信している。

本研究は、様々な記事が混在している CGM を、閲覧者の視点に基づいて可視化することを目的とする。本論文では、Yahoo! 知恵袋^(注2)を対象にして、対象とする読者の集団と記事群であるコミュニティを視点として、投稿記事を配置・可視化する手法を提案する。提案法の特徴は、可視化した記事の位置をユーザの操作によって変化させるところである。ユーザが強調して閲覧したい視点、つまりコミュニティに対する重みを変えることで、記事の位置を変化させる。ユーザの操作によるインタラクティブな記事の再配置を行うことで、ユーザは自分の興味のある別のコミュニティに属した記事を見つけることができる。

本論文ではまず 2. で生活者情報の特徴について述べる。3. で先行研究について述べ、4. で提案手法についての概要を述べる。5. で研究に用いたデータ、および提案手法の評価方法を述べ、6. で提案手法の評価、および考察について述べる。最後に 7. で本研究における結論を述べる。

2. CGM 記事の特徴

インターネットの普及に伴って、大衆といわれる生活者が個人の意見や生活の知恵を発信・共有する CGM が台頭してきた。それにより、個人の意見を発信するブログや生活の知恵を共有する質問回答サイト、匿名のフォーラムなど多様な記事がテキストで流通するようになった。これらの CGM には、同じ興味を持った人の集まり (コミュニティ) が形成され、彼らが書き続ける記事の集合には、ある種の流れのようなもの (コンテキスト) があり、そこでの背景知識を持たないユーザにとって、そこに書かれている記事の内容を正しく理解することが困

難なことがままある。以下では、本論文で扱うコミュニティの性質を概観し、評価で使用する質問回答サイトの投稿記事との対応関係を述べる。

2.1 コミュニティの形成

CGM では匿名であっても記事の書き手が強く意識され、自然発生的にコミュニティが形成される。本論文では、コミュニティを以下のように定義する。

- ユーザの集まり
- 記事の集まり

ユーザの集まりとは、「ある事柄」に対して興味をもった人の集まりである。また、記事の集まりとは「ある事柄」を対象とした記事の集まりである。本論文では、特に断りがない場合は後者の記事の集まりをコミュニティと呼ぶ。

2.2 Yahoo!知恵袋の特徴

本研究では、典型的な CGM の一つである Yahoo!知恵袋 (以下では知恵袋という) を用いて実装・評価を行う。知恵袋は、生活の中で生まれた質問を投稿し、それに対して様々な人が回答する質問回答サイトである。2004 年に開設され、日々膨大な質問と回答記事が投稿^(注3)されている。

本研究では、CGM の題材として Nii より提供された Yahoo!知恵袋のデータを使用した。Yahoo!知恵袋とは生活の中で生まれた質問を投稿し、それに対して様々な人が回答する質問回答サイトである。

何らかの解決した問題や他の人の意見を聞きたいユーザは、質問形式にまとめて知恵袋に投稿する。その際、あらかじめ話題別に区分されたカテゴリのいずれかを選んで投稿する。自分の投稿する記事が、例えばスポーツに属していると思うならスポーツに類するカテゴリを、アウトドアに属していると思うならアウトドアのカテゴリを選ぶ。この際、同じ記事を複数のカテゴリに投稿する行為、マルチポストは暗黙の了解として自粛^(注4)されている。よって、質問の投稿者は複数のカテゴリに属しているような内容の質問でも、必ず 1 つのカテゴリを選んで投稿することになる。

2.3 Yahoo!知恵袋におけるコミュニティ

1 節で定義したコミュニティが Yahoo!知恵袋におけるカテゴリに対応させることができる。すなわち、パソコンや政治といったカテゴリを中心となる事柄として、そこに記事を投稿する、または記事閲覧するユーザ群と、ユーザが投稿した記事群をコミュニティとして見ることができる。

2.4 Yahoo!知恵袋の課題

閲覧者は自分の興味があること、知りたいことをもとにカテゴリを選んで記事を検索する。つまり両者の場合でも、カテゴリとはつまり「興味ある事柄」となる。その際、質問しようとする内容がすでに解決済みであり、知恵袋上に蓄積されている場合がある。しかしカテゴリには似たようなものが複数あり、またユーザの感覚によって投稿するカテゴリを選ぶため、閲覧

(注1): 掲示板などで使われる専門用語。通常、掲示板のスレッドに書き込むと、そのスレッドが掲示板のスレッド一覧の一番上になるが、コマンドによって一覧の順番を変えずに書き込むことができる。このコマンド動作を「下げる」と呼ぶ。

(注2): <http://chiebukuro.yahoo.co.jp/>

(注3): 2009 年 1 月現在、ユーザ数 2,871,900 人、質問総数 22,468,867 件、回答総数 64,726,980 件が投稿されている

(注4): ネットマナーとして浸透している。

者が考える「自分の興味があることが属するカテゴリ」と投稿者の考えるそれが違うことがある．言いかえると、投稿される記事にはそれが属するカテゴリ、つまりコミュニティのほかに別のコミュニティに属している内容のものがある．こういった記事の場合、その記事を望む閲覧者がその記事に到達できない可能性がある．また、この問題は Yahoo!知恵袋特有のものではなく、CGM の特徴の 1 つとして考えることができる．

2.5 コミュニティを視点とする記事の可視化

CGM の記事を正しく理解するためには、記事の投稿者や対象者、状況といった記事が投稿されたときのコンテキストを把握する必要がある．本論文におけるコンテキストとは、コミュニティに属する記事集合の全体を表す特徴である．本論文ではこのコンテキストを把握して記事を読むための可視化手法を提案する．本論文で提案する可視化手法はまず記事の特徴量から、記事を 2 次元上に配置する．これによりユーザは記事の特徴を俯瞰することができる．またユーザが視点を与えることで記事の重みを変化させ、特徴量を再計算する．本論文が提案する可視化手法の概念図を図 1 に示す．前述したとおり、主題としてはあるコミュニティに属する記事でも、別のコミュニティにも属するような内容の記事が CGM にはまみ見られる．この再配置によって記事の 2 次元上での位置を変化させる．つまり、別のコミュニティに視点を与えれば、そのコミュニティにより近い位置に配置されることになる．このとき記事が視点を与えたコミュニティの内容に近ければ近いほど、再配置時の位置がそのコミュニティの位置に近くなる．再配置された記事を見ることで、ユーザは自分の興味のある事柄に属し、かつ別のコミュニティに属する記事、もしくは別のコミュニティに近い内容の記事を見つけることができる．

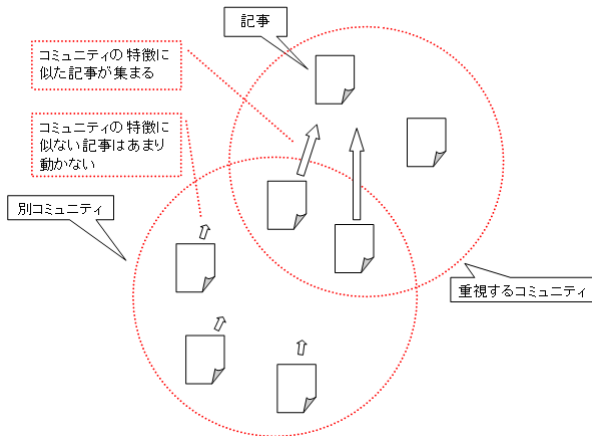


図 1 記事の再配置の概念図

3. 関連研究

Web における情報の配置・可視化にはいくつかの研究例がある．福地ら [1] は Web コミュニティ同士から関係を抽出して、その構造や形成過程を視覚化する方法を提案している．または津ら [2] は Web ページへの典型的な参照内容を表すキーワードを抽出する方法を提案している．手塚ら [3] は検索結果に表れ

る地物の座標からフィルタリングを行い、地図インタフェースに視覚化する方法を提案している．湯田ら [4] はソーシャルネットワークワーキングサイト mixi を対象に、mixi 内の人のつながりや集まりなどの人的ネットワークを解析し、高密度集団の特性を調べている．

これらの従来研究は、いずれも Web ページなどの、対象の位置を一意に求めて視覚化を実現している．それに対して、本研究はユーザの視点による相対的な記事の位置をユーザの操作によってインタラクティブに求めて視覚化を行う．

4. 多様な視点をを用いた可視化手法

記事の特徴ベクトルを求め、記事を x-y 平面上に配置する．また視点としてカテゴリの特徴ベクトルを作成し、これを記事の特徴ベクトルに反映することで視点によって記事の位置を変化させる．

4.1 記事の特徴量

記事の特徴量としてベクトル空間モデルを導入する．各記事から単語を抽出し、tf-idf によって各単語ごとの重みを計算して記事の特徴ベクトルを作成する．

まず記事から形態素解析エンジン MeCab^(注5)で単語を抽出する．このとき名詞のみを抽出し、またいくつかのストップワードを設定している．これは名詞が経験的にもっとも記事の特徴を現す語としての確実と考えられるためである．

記事の特徴量は次のようなベクトルで表現される．ここで d_j は記事の特徴ベクトル、 d_{ij} は記事に出現する単語 w_i の重み、 m は単語数である．

$$d_j = \begin{bmatrix} d_{1j} \\ d_{2j} \\ \vdots \\ d_{mj} \end{bmatrix} \quad (1)$$

また、記事集合全体の特徴量は次の行列によって表現される．

$$D = \begin{bmatrix} d_1 & d_2 & \cdots & d_n \end{bmatrix} \quad (2)$$

次に d_j の計算方法を示す．まず各単語における重み d_{ij} を求める．tf-idf によって重みを計算する際、tf 値を対数化している．これは tf-idf の値が単語の出現頻度数に大きく依存してしまうのを防ぐためである．対数化した tf 値 l_{ij} は次の式で求められる．ここで f_{ij} は単語 w_i の記事 d_j における出現頻度である．

$$l_{ij} = \log(1 + f_{ij}) \quad (3)$$

idf 値には文書頻度の逆数を用いた．ここで n は全文書数、 n_i は単語 w_i を含む文書数である．

$$g_j = \log \frac{n}{n_i} \quad (4)$$

各単語の重みを決める際、文書の長さによる影響を減らすために文書正規化を行う．文書正規化には文書に含まれる全ての単語の重みに 2 乗和を 1 にするコサイン正規化を用いた．

(注5): <http://mecab.sourceforge.net/>

$$n_j = \sqrt{\sum_{i=1}^m (l_{ij}g_j)^2} \quad (5)$$

式 (3)(4)(5) から，記事に出現する単語の重み d_{ij} は次の式で表される．

$$d_{ij} = \frac{l_{ij}g_j}{n_j} \quad (6)$$

式 (6) で求めた d_{ij} を式 (1) に代入し d_j を導出する．

4.2 視点の形成

ユーザが変える視点にはコミュニティ，すなわち Yahoo!知恵袋におけるカテゴリを用いる．各記事と同様にして，カテゴリの特徴量を求めて特徴ベクトルを作成する．このときカテゴリの記事には，そのカテゴリに含まれる記事すべてを1つの記事として扱う．

$$\mathbf{c}_k = \begin{bmatrix} d_{c_k1} \\ d_{c_k2} \\ \vdots \\ d_{c_km} \end{bmatrix} \quad (7)$$

\mathbf{c}_k はカテゴリの特徴ベクトル， d_{c_kl} はカテゴリに属する記事の単語である．式 (7) によりカテゴリの基準となるベクトルが作成され，2次元上でのコミュニティの位置が決まる．

4.3 可視化のための2次元空間への射影

作成した特徴ベクトルは N 次元からなるため，そのままでは可視化が困難になる．ここでは N 次元を2次元に圧縮し x - y 平面に記事を配置する手法を示す．これには特異値分解によって左特異ベクトルを求め，このベクトルの張る空間への正射影を用いた．まず全ての記事のベクトル \mathbf{D} の特異値分解を行う．

$$\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (8)$$

\mathbf{U} 内の左特異ベクトルは左にあるものほど，変量の分散が大きくなるため記事の特徴を反映できる．よってこの \mathbf{U} の最初の2個の列ベクトルで構成される \mathbf{U}_2 を使い，特徴ベクトル \mathbf{d}_j を x - y 平面に配置する． x - y 平面での記事のベクトル \mathbf{d}_j^2 は次の式で表せる．

$$\mathbf{d}_j^2 = \begin{bmatrix} X_{d_j} \\ Y_{d_j} \end{bmatrix} = \mathbf{U}_2^T \mathbf{d}_j \quad (9)$$

これにより， N 次元の特徴ベクトルは2次元に圧縮される．また，同様にしてカテゴリの特徴ベクトル \mathbf{c}_k の x - y 平面での位置は次の式で表せる．

$$\mathbf{c}_k^2 = \mathbf{U}_2^T \mathbf{c}_k \quad (10)$$

4.4 視点を与えるユーザインタラクション

ユーザの視点を与えるインタラクションには適合性フィードバックを用いた．これによりユーザの重視するカテゴリに属する単語の重みを上げる．記事の特徴ベクトルに適合性フィードバックを適用することにより， x - y 平面に配置したときの位置が重視するカテゴリの位置に近づく．適合性フィードバックに

おける索引語，適応文書をそれぞれ記事の特徴ベクトル，カテゴリの特徴ベクトルに当てはめれば，適合性フィードバックを適応した記事の特徴ベクトル \mathbf{d}'_j は次の式で表せる．

$$\mathbf{d}'_j = \alpha \mathbf{d}_j + \beta \sum w_i \mathbf{c}_k \quad (11)$$

α と β は0以上の数である．また w_i はユーザが任意で変える視点の重みである． w_i を変えることにより，重視するカテゴリ，重視しないカテゴリを決める．

評価の際に使用する，適合性フィードバックを適応した記事の変化量 R_j は次の式から求められる． X_{d_j} と Y_{d_j} はそれぞれ変化前の記事の X と Y ， $X_{d'_j}$ と $Y_{d'_j}$ は変化後の記事の X と Y である．

$$R_j = \sqrt{(X_{d_j} - X_{d'_j})^2 + (Y_{d_j} - Y_{d'_j})^2} \quad (12)$$

5. システムの実装と評価方法

5.1 システム概要と評価データ

システムの実装には JAVA を用いた．tf-idf を求めるのには SlothLib^(注6)を用いた．特異値分解には JAMA^(注7)を用いた．

研究用のデータとして，提供された Yahoo!知恵袋のデータのうち2005年9月分を用いた．ここには241のカテゴリがあり，214,597の記事があった．241のカテゴリを小カテゴリとして再分類を行い15の大カテゴリを作成し，このうちの大カテゴリの1つ「インターネット，PCと家電」を用いて研究を行った．また「インターネット，PCと家電」には21の小カテゴリがあったが，このうち「インターネット」「オーディオ」「パソコン，周辺機器（以下パソコンと略す）」「家電，AV機器（以下AV機器と略す）」「携帯電話，モバイル（以下携帯電話と略す）」の5つの小カテゴリを用いて研究を行った．これは，この5つ以外の小カテゴリの記事数が極端に少なかったためである．

表1は「インターネット，PCと家電」内の小カテゴリの特徴ベクトルのコサイン類似度である．「インターネット」と「AV機器」の類似度は0.1816と小さく，「オーディオ」と「AV機器」は0.4659と大きいなど，同じ大カテゴリの中でも，相互の類似度に違いがあることが判る．

表1 カテゴリ間のコサイン類似度

	インターネット	オーディオ	AV 機器
インターネット	1	0.2113	0.1816
オーディオ	—	1	0.4659
AV 機器	—	—	1

5.2 提案法の評価方法

提案法の評価は次の項目によって行った．

- 記事の特徴ベクトルの評価
 - 記事の x - y 平面での位置

記事の特徴ベクトルの精度は記事の x - y 平面での位置に

(注6): Copyright (c) 2007, 京都大学情報学研究科社会情報学専攻 田中克己研究室

(注7): <http://math.nist.gov/javanumerics/jama/>

よって与えられる．同じ小カテゴリに属する記事は x - y 平面上において近くに配置される．また記事が属する小カテゴリの位置の近くに配置される．

- 再計算後の記事のベクトルの評価
 - － 再計算後の記事の x - y 平面での位置
再配置された記事の位置がコミュニティの位置の位置に近ければ近いほど精度が高いと判断できる．
 - － 記事とコミュニティの類似度
記事と記事が属していないコミュニティの類似度が高いほど再配置されたときの位置は大きく変化する．
 - － 再計算前と後との記事の位置の距離・変化量
記事の計算前と計算後との距離を求めることでシステムの精度を測ることができる．

これらの指標から，提案法の有効性を評価する．

6. 提案法の評価

提案法の評価には各小カテゴリから 30 件づつ，計 150 件の記事を用いた．

6.1 カテゴリの x - y 平面での位置

式 (10) によって導出された小カテゴリの x - y 平面での位置を表 2 および図 2 に示す．ここで求められた位置がカテゴリ，つまりコミュニティの位置となる．「オーディオ」を主点にして表 1 を見ると「AV 機器」が最も類似度が高くなっており，それ以外にも「パソコン」「携帯電話」「インターネット」との順で類似度が高くなっていた．図 2 を見ても「オーディオ」の位置に最も近いのは「AV 機器」で「パソコン」「携帯電話」「インターネット」と続いている．それ以外のカテゴリでも，概ね類似度が高いものは近くに，低いものは遠くに配置されている．しかし図 2 では「インターネット」の位置に最も近いのは「携帯電話」であるが，コサイン類似度が最も高かったのは「パソコン」であり，結果に多少のズレがある．これについてはさらなる検討が必要である．

表 2 カテゴリの位置

カテゴリ	X	Y
インターネット	0.4016	-0.2330
オーディオ	0.4854	0.2830
パソコン	0.4232	0.0193
AV 機器	0.4506	0.3670
携帯電話	0.4314	-0.2258

6.2 記事の x - y 平面での位置

式 (9) によって導出された小カテゴリごとの記事の x - y 平面での位置を図に示す．いずれの場合でも，記事は X 軸上においてはカテゴリの値以下となり， Y 軸上においてはカテゴリの前後に配置される．しかし概ねカテゴリの位置に近く配置されており，記事の位置としては適当である．

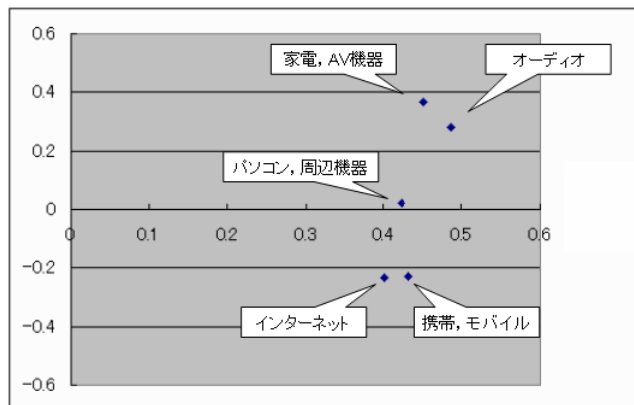


図 2 カテゴリの位置

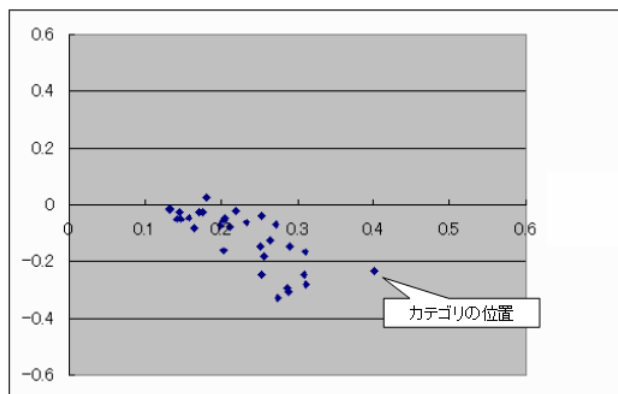


図 3 「インターネット」内の記事の位置

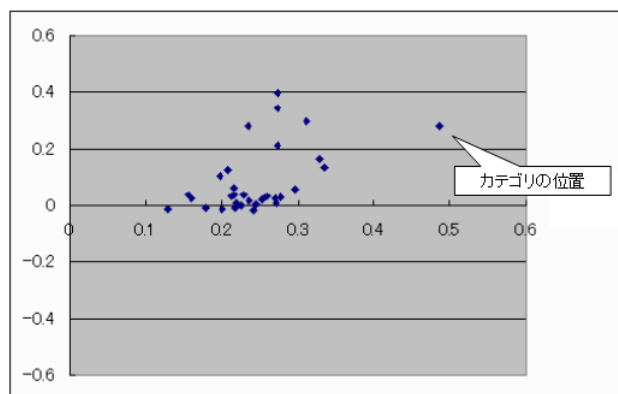


図 4 「オーディオ」内の記事の位置

6.3 適合性フィードバックを適応した記事の位置

式 (9) によって導出された記事の x - y 平面での位置を図 8 に，式 (11) を適応した記事の x - y 平面での位置を図 9, 10, 11, 12, 13 に示す．また計算前の全記事と「インターネット」を重視した記事の位置の比較を図 14 示す．いずれの場合でも，変化後の記事の位置はカテゴリの位置に近づいている．

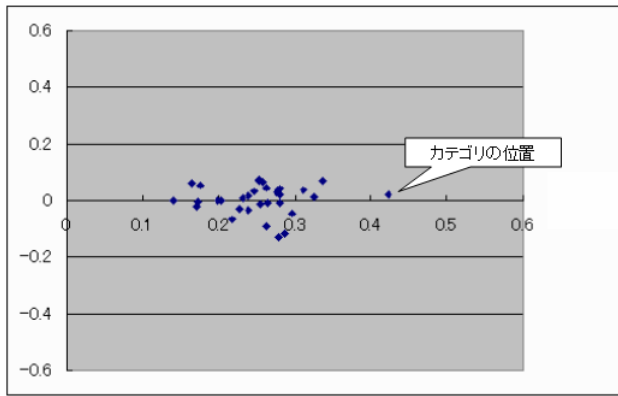


図 5 「パソコン, 周辺機器」内の記事の位置

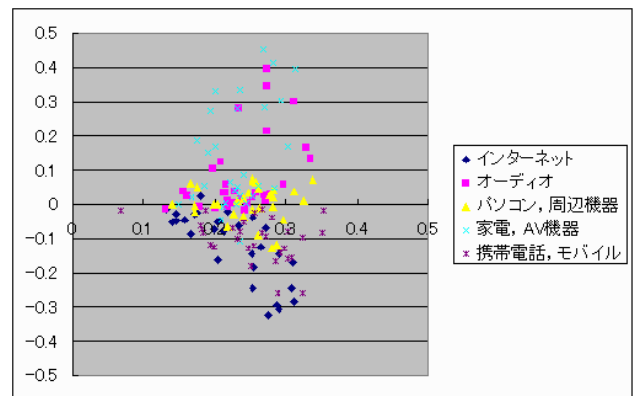


図 8 変化前の全記事の位置

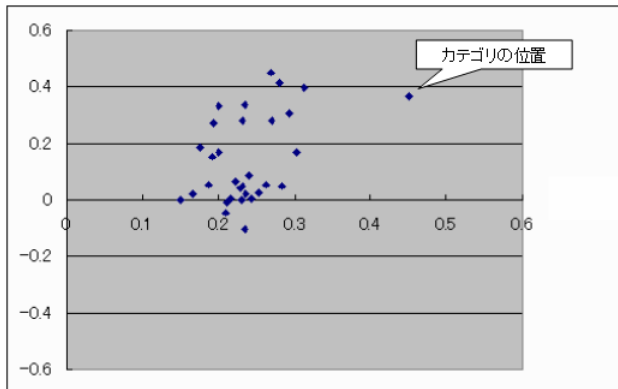


図 6 「家電, AV 機器」内の記事の位置

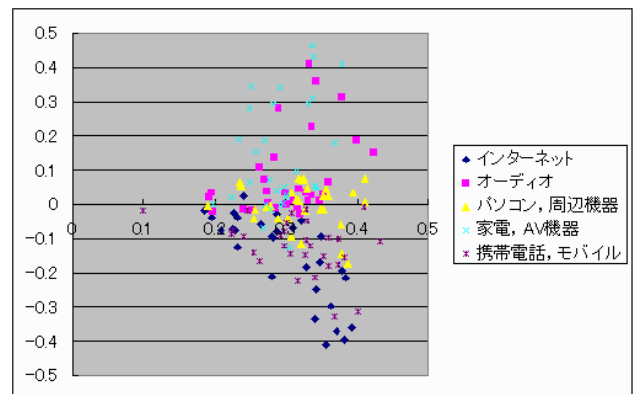


図 9 「インターネット」を重視した記事の位置

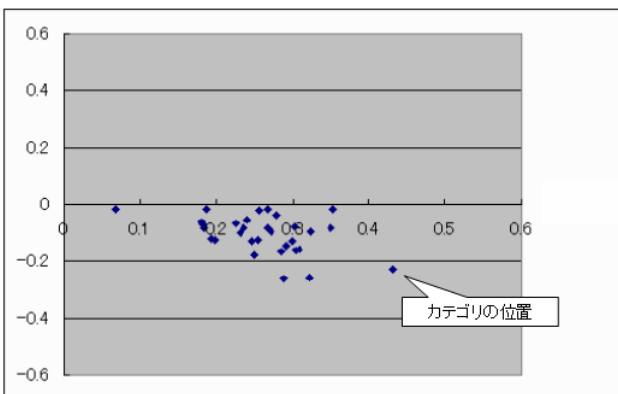


図 7 「携帯電話, モバイル」内の記事の位置

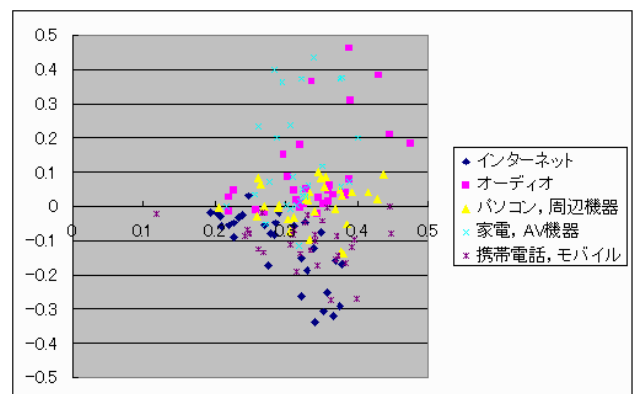


図 10 「オーディオ」を重視した記事の位置

6.4 適合性フィードバックを適応した記事の変化量

「インターネット」を重視したときの変化量が多かった他カテゴリの上位 10 件および下位 10 件の記事を表 3, 表 4 にそれぞれ示す。ここで類似度とは個々の記事と「インターネット」とのコサイン類似度である。

「パソコン」「携帯電話」の記事がほぼ上位を占めており、「AV 機器」の記事は上位 10 件に含まれていない。つまり「インターネット」と類似度の高い「パソコン」「携帯電話」の変化量は大きく、類似度の低い「AV 機器」の変化量は小さくなっている。下位 10 件には「AV 機器」の記事が多く含まれている。また「インターネット」と類似度の高い「携帯電話」や「パソコン」の記事も多く含まれている。しかし、これらはどれも記

事自体とカテゴリとの類似度が低くなっている。次に、図 15 に記事 5922976 の適合性フィードバックを適応する前の記事の位置と適応した後の記事の位置を示す。図のように、記事は「インターネット」の位置により近く再配置されている。また「インターネット」以外のカテゴリを重視した際の記事の再配置でも、同様に類似度の高いカテゴリの記事、記事自体の類似度の高いものの変化量が大きく、より重視するカテゴリーの近くに配置された。

6.5 考 察

本研究では、様々な記事が混在している CGM をコミュニティの視点に基づいて可視化することを目的に、記事を可視化

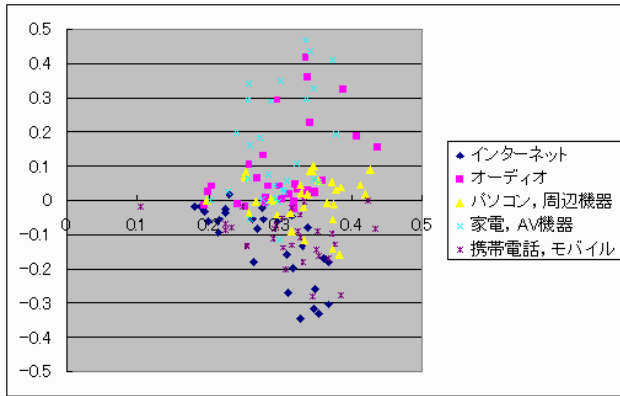


図 11 「パソコン，周辺機器」を重視した記事の位置

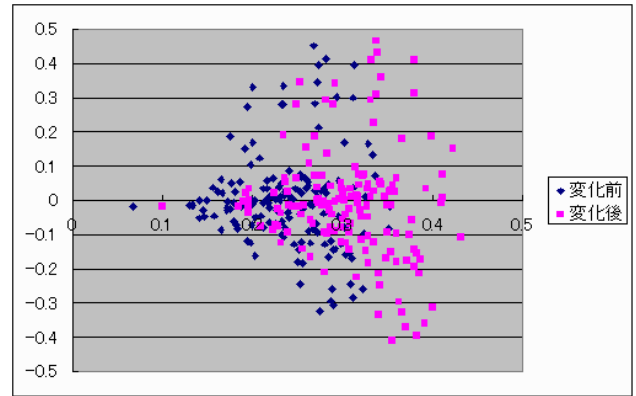


図 14 変化前の記事と「インターネット」を重視した記事の位置の比較

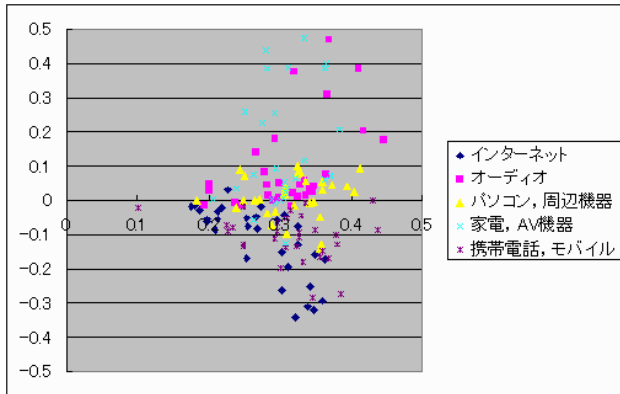


図 12 「家電，AV機器」を重視した記事の位置

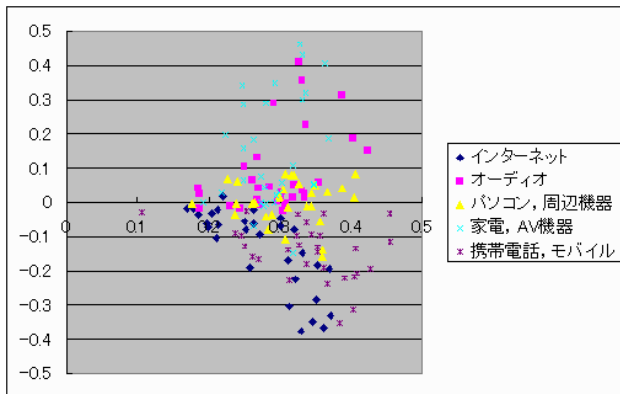


図 13 「携帯電話，モバイル」を重視した記事の位置

表 3 「インターネット」を重視した時の変化量の多い記事

記事番号	カテゴリ	変化量	類似度
5922976	パソコン	0.1154	0.1316
5887377	携帯電話	0.1031	0.1350
6112112	パソコン	0.0970	0.1921
5969916	携帯電話	0.0965	0.1439
6151672	オーディオ	0.0946	0.0891
6052523	パソコン	0.0931	0.1518
5914006	オーディオ	0.0904	0.0749
5924770	パソコン	0.0901	0.1085
5991800	オーディオ	0.0894	0.0666
6210269	オーディオ	0.0877	0.0766

表 4 「インターネット」を重視した時の変化量の少ない記事

記事番号	カテゴリ	変化量	類似度
6204045	携帯電話	0.0318	0.0117
6262655	オーディオ	0.0319	0.0083
6307340	携帯電話	0.0397	0.0417
6232654	オーディオ	0.0403	0.0330
6054749	AV 機器	0.0451	0.0463
5973351	AV 機器	0.0472	0.0262
6220013	AV 機器	0.0480	0.0557
6191320	携帯電話	0.0481	0.0544
6127955	パソコン	0.0489	0.0333
5906858	パソコン	0.0489	0.0698

し，さらにユーザの操作によってインタラクティブに記事を再配置する手法を提案している．

提案法における記事を可視化する部分については，記事の内容から記事の特徴ベクトルを求め， N 次元からなる記事の特徴ベクトルを x - y 平面に射影することで実装した．図 3～7 が示すように，記事の位置がその記事が属しているカテゴリの位置の近くに配置されており，記事の特徴を反映した座標の導出を十分に達成できていると考えられる．

またユーザが与える視点によって記事を再配置する部分については，適合性フィードバックの記事の特徴ベクトルに適應することで実装した．図 8～14 が示すように，適合性フィード

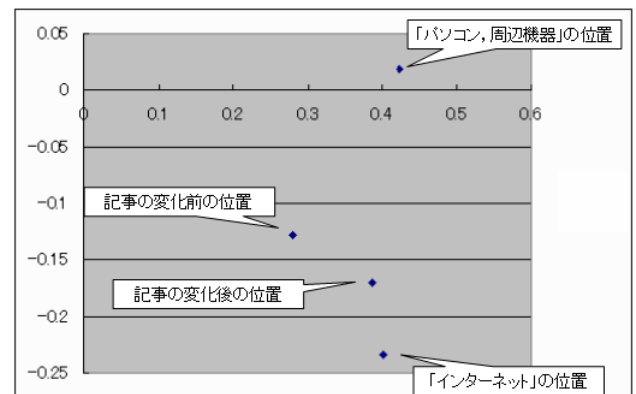


図 15 記事 5922976 の変化前と変化後の位置

バックにより重視するカテゴリーの特徴の重みが上がり、記事全体が重視するカテゴリーに近づいている。また、図 15 および表 3, 4 が示すように一律に変化するのではなく、記事の特徴がカテゴリーの特徴に似ているものは変化が大きく、似ていないものは変化が小さくなっており、ユーザの視点を反映した記事の再配置が十分に達成できていると考えられる。

以上のことより本研究における目的を達成できているが、まだ残された課題がある。前述したとおり、カテゴリーの位置が記事の位置よりも X 軸のみ大きな値になってしまった。これはカテゴリーと記事に含まれる単語数の差が原因だと考えられる。カテゴリーに含まれる単語は、そこに属する全ての記事の単語である。また本研究で特異値分解で求めた U の 1 列目は全て正の数であった。カテゴリーには重要となる単語を多く含まれるため、 U によって x - y 平面に射影する際、 x の値が大きくなったのだと考えられる。この問題の解決が今後の課題の 1 つである。また本研究では視点、つまり重視するカテゴリーは 1 つだけとして研究、実装を行った。よって、複数のカテゴリーを視点として記事を再配置することが今後の課題の 1 つである。

7. 結 果

本研究では、様々な記事が混在している CGM をコミュニティの視点に基づいて可視化することを目的として、記事を可視化し、さらにユーザの操作によってインタラクティブに記事を再配置する手法を提案した。

記事の可視化は、記事の内容から記事の特徴ベクトルを求め、特異値分解を用いて x - y 平面に射影することで実装した。またコミュニティの特徴ベクトルも記事と同様に求め、特異値分解を用いて x - y 平面に射影した。これについては、コミュニティの位置の近くに、そのコミュニティに属する記事が配置されたことで、提案手法が十分有効であることが確認できた。また、ユーザの操作によるインタラクティブな記事の再配置は、適合性フィードバックの記事の特徴ベクトルに適應することで実装した。これらについては、視点としたコミュニティに類似度の高い記事がより変化し、逆に類似度の低い記事があまり変化しなかったことで、十分に反映した記事の再配置が行われたことが確認できた。

しかし、コミュニティの位置の X 座標が大きくなりすぎることや、複数の視点に対応していないなどの課題が残った。そのため、これらの解決が今後の課題だと考えられる。

謝 辞

本研究の実装・評価に際し、大学共同利用機関法人 国立情報学研究所から提供を受けた、Yahoo!知恵袋のデータを利用している。ここに記して謝意を示す。

文 献

- [1] 福地健太郎, 豊田正史, 喜連川優. Web community browser: 大規模 web コミュニティチャートの可視化. 第 13 回データ工学ワークショップ (DEWS2002) 論文集, A1-4, 2002.
- [2] 是津耕司, 木俣豊, 田中克己. Web ページのアスペクトの発見. 日本データベース学会論文誌 (DBSJ Letters), Vol. 2, No. 4, pp. 15-18, 2004.

- [3] 手塚太郎, 近藤浩之, 田中克己. 混合ガウス分布を用いたウェブコンテンツの地域性推定とオブジェクトレベルローカルサーチ. 情報処理学会論文誌 データベース, Vol. 1, No. 1, pp. 13-25, 2008.
- [4] 湯田聡夫, 小野直亮, 藤原義久. ソーシャル・ネットワーキング・サービスにおける人的ネットワークの構造. 情報処理学会論文誌, Vol. 43, No. 3, pp. 865-874, 2006.