

検索新聞：「新聞形式型検索情報提示システムにおける マイクロブログを用いたユーザ適応」

Kensaku Shimbun : User adaptation using microblogs for news-paper-style retrieval system

関谷 英樹†
Hideki Sekiya

祖父江 翔†
Sho Sobue

田村 哲嗣‡
Satoshi Tamura

速水 悟‡
Satoru Hayamizu

1. はじめに

インターネットが普及するにつれ、様々な情報が Web 上にあふれ存在するようになった。その量は膨大であり、この中から有益な情報を得るためには、検索エンジンの活用は不可欠である。しかし、検索結果には不必要な情報が混濁していることも多々あり、ユーザが望む情報の取得と理解には多くの労力や時間を要してしまう。それゆえ、ユーザにとって有益な情報をどのように取得し提示するのが課題となる。そこで我々は、検索結果を新聞形式にまとめユーザに提示する検索新聞システム（以下本システム）を開発している。[1][2]

一方、今日では誰でも気軽に Web 上に文書などのコンテンツをアップロードできるようになった。ブログや SNS のような誰でもが手軽に情報を発信できるサービスがその一例である。近年、このような新しい Web の利用形態を対象とした研究が行われている。ブログから得られる嗜好情報を利用した協調フィルタリングによる情報推薦システムが提案されている[3]。また、ユーザのプロファイル（属性）を推定する研究としては、ブログの著者の性別を推定するものがある[4]。最近では短い文を投稿するマイクロブログが登場し、急速に広まりつつある。これを利用し、推薦のために投稿者の状況推定を行う研究[5]や、ユーザの関心度の推定を行う研究[6]、年齢、性別などの属性を推定したシステム[7]も発表されている。最近ではつぶやきのリンク先の内容を新聞形式で表示するサービスもある。[8]しかし、ブログを用いた Web 検索の研究は多く行われているが、マイクロブログを用いた Web 検索についての研究は未だ少ない。マイクロブログはユーザが端的かつリアルタイムに情報発信する特性があり、ユーザのプロファイルだけでなく、ユーザの最新の興味を抽出できる可能性がある。これを検索エンジンの結果表示において活用することができれば、ユーザが望む情報を効率的に提示できると考えられる。

そこで、本研究ではマイクロブログの1つである Twitter を用いて、Web 検索結果をユーザ適応することを目的とし、そのために Twitter の投稿を用いたユーザの嗜好推定を行った。これを、我々が構築した「検索新聞」と統合し、主観評価を行った。

本論文では、第2章で「検索新聞」の概要について述べ、第3章では Twitter 情報を用いたユーザの嗜好の推定手法、及びそれを用いた Web 検索実験について述べる。第4章では、Web からの文書選択の際の重要度のスコア付け手法、また、「検索新聞」と嗜好推定のスコアの統合、統合後のシステムの評価実験について述べる。

2. 「検索新聞」システム概要

検索新聞は、ユーザが検索キーワードを入力することにより Web 上から取得した文書を、複数の特徴量を用いて重要度の推定を行い、その結果を新聞形式でユーザに提示するシステムである。新聞形式でユーザへ提示する利点として、新聞に慣れ親しんでいるユーザに対し読みやすい結果の提示が可能なこと、まとまった形式であるため冗長な情報提示を避けられることなどが挙げられる。

ユーザにとって重要な文書を選択するため、3種類の重要度の評価尺度「専門性、可読性、新規性」により文書の重要度の推定を行う。また、Twitter の情報を用いることにより嗜好推定を行い、ユーザの嗜好に合わせた情報の提示を目指す。本システムの概要を図1に示す。

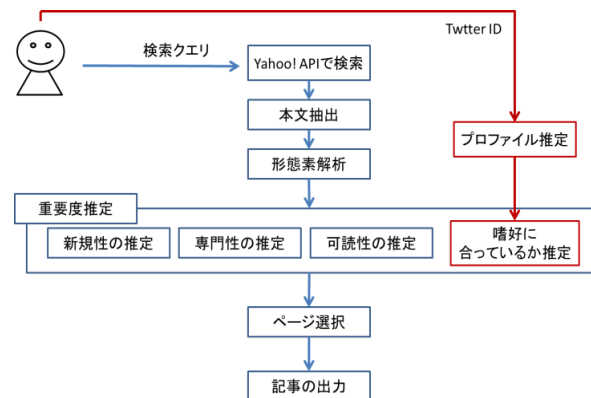


図1 本システムの概要

3. マイクロブログを用いたユーザ適応

本章では、マイクロブログの特徴、また、Twitter 情報を用いたユーザの嗜好の推定について述べる。

マイクロブログは、ミニブログやつぶやきブログとも呼ばれ、ユーザは居場所やテレビの感想、他のユーザの投稿を引用した意見など、自由な内容を書き込むことができる。また、書き込んだ内容に対しての返信など、ユーザ間でコミュニケーションを取ることもできる。

マイクロブログはブログの一種であるが、従来のブログと異なり、次の特徴がある。第一に、インターフェースが単純で直観的であることが挙げられる。インターフェースは、テキストボックスと投稿ボタンのみで構成されており、容易にかつ気軽に投稿を行える。第二として、投稿内容が短いことが挙げられる。多くのマイクロブログには、投稿

† 岐阜大学大学院工学研究科

‡ 岐阜大学工学部

文字制限がある。このため、ユーザは端的な情報を投稿する傾向があり、冗長でない、ユーザの最も言いたい発言を取得しやすい。第三として、他のユーザの投稿を同じ画面上で一度に見ることができる。そのため、ブログと違いあらゆる情報を、手間をかけずに見ることができる。第四として、情報が広がりやすく、更新が速いことが挙げられる。リツイートを選択するだけで、他のユーザの投稿内容を他の人に広げることができる。そのため、ユーザは他に広げたいと思う内容について、気軽に情報発信者になることができる。また、ユーザはブログ以上の速さと頻度で投稿をする傾向がある。そのため、ブログ以上にリアルタイムな内容が投稿される。

以上のことから、マイクロブログはブログよりもユーザが気軽にリアルタイムな内容を投稿しており、速報性が高いと言える。これを利用することで最新のユーザの嗜好を推定し、検索へ反映できる可能性がある。そこで本研究では、速報性の高いマイクロブログである Twitter からユーザの特性を推定し、ユーザのリアルタイムな嗜好に合わせた情報提供することを目標にする。その取り組みとして、投稿内容のジャンル推定をしてユーザの嗜好の推定を行い、その結果を Web 検索へ適応させることを試みた。

ユーザの投稿は、Twitter API[9]を利用し取得する。

3.1 Twitter 特有の用語

Twitter には、サービス特有の表現や用語が存在する。その一部の定義を表 1 に示す。

表 1 Twitter 特有の用語

Twitterの用語	定義
タイムライン(TL: Time Line)	ユーザ自身や他のユーザを含めた時系列順に並ぶ投稿ログのことを指す。
ツイート(つぶやき/Tweet)	ユーザの投稿データのこと。もしくは、投稿することを指す。
フォロー(follow)	特定のユーザの投稿をユーザ自身のタイムラインに表示させるようにする仕組みのことを指す(閲覧ユーザの登録)。非公開にされていない限りフォローすることは、相手に事前に許可を取る必要がなく、ユーザ自身が自由に行うことができる。
フォロワー(follower)	あるユーザをフォローしているユーザを指す。
リスト(list)	好みのテーマに沿った複数のアカウントの投稿だけをタイムラインに表示させる仕組みのことを指す。大量のフォローをしていても、これを用いて必要な投稿だけを取り出して閲覧することができる。
リツイート(ReTweet/RT)	他のユーザの投稿内容について、興味を持ったリ上げたりしたいと思うものを再投稿することを指す。ブログやSNSなどで、他人の記事を引用して紹介する行為に近い。

3.2 ジャンル推定とプロフィールベクトル

嗜好に合わせた Web 検索を行うため、投稿内容をジャンル推定し、Web ページをランキングする。この Web 検索の概要を図 2 に示す。

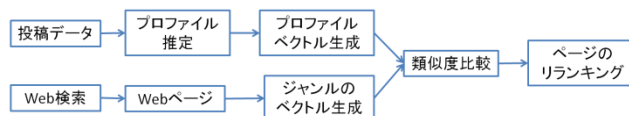


図 2 投稿内容を用いた Web 検索概要

ユーザが Twitter へ投稿したデータをジャンル推定し、プロフィールベクトルを作成する。Web から取得した文章にも同様にジャンル推定を行い、2つの推定結果の類似度を比較することで Web ページのランキングを行う。

マイクロブログでは容易に書き込みができるため、体調など嗜好を推定する上で重要でない投稿も多く含まれる。それらを取り除くため Twitter 状況推定を参考に、場所や感情を含まない投稿と、他のユーザの投稿を引用した投稿のみを取り出すフィルタを作成した。具体的には、Twitter で特徴的な表現である語尾に「なう」をつける投稿や他の投稿を引用して投稿する引用 RT(リツイート)を使用する。「なう」については場所や感情表現を持たないものを対象にするため、投稿者の状況推定を行ってからジャンル推定を行う。

ジャンル推定は入力形態素に対して、ラベル(ジャンル)を付与する問題として扱い、推定モデルには識別モデルのひとつである CRF(Conditional Random Fields)を用いる[10]。CRF は条件付き確率に基づく推定モデルで、入力特徴量に制約がなく定義しやすい点と推定確率を直接推定できるという特徴がある。入力特徴量を x 、ラベルを y 、推定ラベルを \hat{y} 、入力 x におけるラベル y となる条件付き確率を $P(y|x)$ としたとき、以下の式で定義される。

$$\hat{y} = \operatorname{argmax}_{y \in Y} P(y|x) \quad (1)$$

本研究における CRF は、入力特徴量を入力形態素の情報(対象の形態素より前後 2 形態素の表層形および品詞名)、出力ラベルを推定ジャンル(文化、経済、娯楽、家庭、国際、科学、社会、スポーツ)とする。

取得した投稿データをフィルタリングし、推定モデルに基づき推定したジャンルの推定確率をベクトル化する。これをプロフィールベクトルと呼ぶことにする。その概要を図 3 に示す。このプロフィールベクトルを生成することで、ひとつのジャンルだけを嗜好とせず、複数のジャンルにまたがる嗜好も捉えることができる。

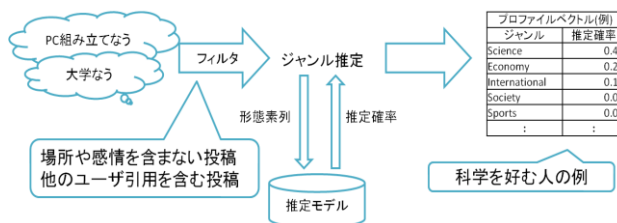


図 3 プロフィールベクトルの生成

プロフィールベクトルに基づいた Web 検索結果のランキングを図 4 に示す。プロフィールベクトルと同様に、プロフィールベクトル生成時に用いたジャンル推定モデルを用いることで、Web ページに対してもジャンル推定を行い、Web ページのベクトルを作成する。その後、プロフィールベクトルと Web ページのベクトル間の類似度を測り、類似

しているベクトルの Web ページから上位に表示していく。類似度の推定にはコサイン距離を用いている。これにより、プロフィールベクトルと同じような内容が上位に検索結果として表示される。

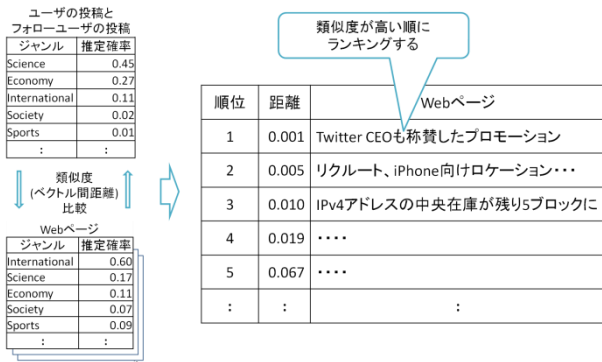


図4 Web 検索結果のリランキング

3.3 予備実験

実際にプロフィールベクトルに合った Web ページを取得できるか予備実験を行った。実験条件を表2に示す。テストデータには3名のデータを用いた。

表2 実験条件

推定モデル	CRF(Conditional Random Fields)
使用ジャンル	毎日新聞記事の8ジャンル (文化,経済,エンタメ,家庭,国際,科学,社会,スポーツ)
学習データ	毎日新聞記事2002~2007年 (各ジャンル3000記事)
テストユーザ	理系大学生3人
テストデータ	Web検索結果に出てくるWebページ(300件)
本文抽出	ExtractContentモジュール
検索クエリ	キャンプ
ベクトル類似度	コサイン距離
評価	主観による判断

投稿データを用いたプロフィールベクトルの生成では、3人のユーザは以下の表3のような特徴であった。ユーザAの投稿およびフォローしているユーザの投稿は、パンやコーヒーなど料理に関する内容が多く見られ、ユーザBは情報に関する内容が多くありながら、IT業界について触れた投稿が含まれていたことにより、次に経済が高い値を取ったと考えられる。ユーザCは、情報系や、スポーツに興味のあるユーザをフォローしていることから、科学とスポーツが高い値となったと考えられる。これより、ユーザの投稿をジャンル推定し嗜好をプロフィールベクトルの形で表すことは有用であると考えられる。

表3 実験に用いたユーザの特徴

ユーザA	家庭及び文化の嗜好 フォローしているユーザに食べ物に関する投稿
ユーザB	科学と経済の嗜好 IT業界についての投稿
ユーザC	科学とスポーツの嗜好 フォローユーザーが情報系大学生

各ユーザについて、検索クエリ「キャンプ」で検索を行った結果を表4に示す。

表4 各ユーザの実験結果

検索クエリ:キャンプ	
ユーザAの検索結果	
1位	歌オオートキャンプ場ルビックオフィシャルサイト(137位)
2位	2011年度春季キャンプガイド(69位)
3位	岐阜県中津川市 福岡ローマン渓谷 オートキャンプ場(97位)
4位	富士すそ野ファミリーキャンプ場(9位)
5位	休暇村 佐渡 - 佐渡にあるオートキャンプ場の休暇村 釣りや海水浴で...(193位)

検索クエリ:キャンプ	
ユーザBの検索結果	
1位	Error page(209位)
2位	エンバカデバロ・デベロッパークャンプ(147位)
3位	土佐西南大規模公園 オートキャンプ場 とまろっと ~ 清流・四万十川の...(42位)
4位	キッズ・マーケット・キャンプ2011について(169位)
5位	ハヶ岳のキャンプ場 - ハヶ岳オートキャンプ場 - 山梨県北杜市小淵沢町...(167位)

検索クエリ:キャンプ	
ユーザCの検索結果	
1位	みどりの村キャンプ場 更新のお知らせ(144位)
2位	中村憲剛オフィシャルサイト:@niffy - メッセージ: キャンプ(120位)
3位	山梨のキャンプ場 山梨県 ウェストリバーオートキャンプ場HP 関東...(189位)
4位	小菅村 東部森林公園キャンプ場 ほうれん坊(200位)
5位	セキュリティ&プログラミングキャンプ2009、開講 - @IT(112位)

各ページの後に括弧で書かれた数値は、Yahoo!検索における検索結果の順位である。ユーザAについては、結果上位に「アウトドアにおけるキャンプ」、ユーザBは2位に「エンジニアのためのキャンプ」、4位に「経済を学ぶキャンプ」、ユーザCは2位に「サッカー選手のキャンプ」、5位に「プログラミングのキャンプ」となった。これより、ユーザの嗜好に合わせて検索結果が変化したことを確認した。もとの検索結果で下位であったページが上位に表示されることにより、ユーザの嗜好に合い、かつ検索の手間が省ける可能性があると考えられる。

4. 重要度スコア

Web 検索で取得した文書に対して重要度の推定を行う。重要度の尺度としてはTF*IDFが多く用いられている。これは単語の出現頻度をもとに重要度を算出する手法である。しかし、重要度はユーザによって異なる指標であり、常に変化する不確定な尺度である。これを単一で定めることは困難であり、多角的に判断する必要がある。

本システムでは、先行研究として3種類の重要度の尺度「専門性」、「可読性」、「新規性」を用い、Web から取得した文書について重要度推定を行っているが、本研究では、新たに「ユーザの嗜好」についてのスコアを追加した。本システムのインターフェースを図5に示す。

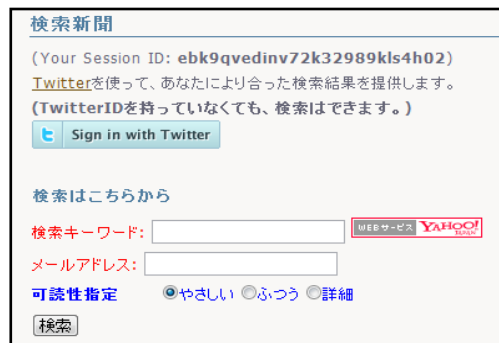


図5 検索新聞インターフェース

Web 情報の取得は Yahoo!API[11]により行う。Twitter へログインすることで、3章の方法を利用しユーザの嗜好をスコアに統合することができる。Twitter 情報を用いない場合は「専門性」, 「可読性」, 「新規性」のスコアのみを用いる。

4.1 専門性

利用者が情報を重要と位置付ける要素の1つとして、その情報がどの程度専門的な内容を含んでいるか、という指標を考え、その尺度を本論文では「専門性」と呼ぶことにする。ある単語が一般的ではなく、一部の専門分野でのみ使われることが多い場合や、あるジャンルに偏って多く出現する場合に専門的であると考えられる。

本論文では、専門性のスコア付けとして「名詞」, 「動詞」に着目しジャンルの推定を行うことで、入力文書のジャンル、及びスコアを決定する。毎日新聞で定められたジャンルの中から8ジャンル(文化、経済、娯楽、国際、科学、社会、スポーツ、家庭)に対してCRFを使用し、ジャンルごとに言語モデルを作成する。モデルの作成には、毎日新聞記事データ集[12]を学習データとして用い、オープンソースのCRFツールキットCRF++[13]により作成した。

4.2 可読性

情報がどの程度複雑な内容を含んでいるか、どの程度読みやすいかという指標を考え、その尺度を「可読性」と呼ぶ。キーワードを入力する際に、ユーザが可読性を「やさしい」, 「ふつう」, 「詳細」の3種類から選択できる。この選択により文書のスコア付けの際の重みが増減し、新聞紙面に提示される文書が変わる。

文書のスコア付けには、教科書コーパスを用いたテキストの難易度推定[14]を利用した。これは、テキストの難易度を表す区分として学年区分を使用するものである。テキストの難易度は、小学1年から高校3年までの12学年に、大学を加えた計13学年の段階区分で表わされる。教科書コーパスは、小・中・高の英語を除く全教科の教科書、大学は、教養課程の講義で使用されているものを用いる。難易度が既知である教科書を利用することで、小学校から高校までの幅広い学年において、漢字や語彙だけでなく言語表現も考慮した難易度推定が可能である。

それぞれの区分の教科書をコーパスとして作成された言語モデルを用い、入力された日本語テキストに対して、小学1年から大学までの13の難易度クラスの中から難易度クラスが1つ決定される。本研究ではここで得られた最大の尤度を各文書のスコアとする。

この手法を利用することによりテキストの難易度を求めることができる。しかし、本システムの検索は一般のニュースやWeb記事を対象としているため、易しいと推定される文書が少なく、小学生などの低い難易度の文書は稀にしか取得できない問題がある。そこで、小・中学生のためのポータルサイトであるYahoo!きっず[15]の「こどもニュース」, 「用語集」(話題になった用語の解説)の文書を用いる。これは、ユーザが可読性「やさしい」を選択した場合のみ提示し、ユーザが理解しやすい文書の提示を行う。

また、入力キーワードに対するユーザの理解がより深まることを期待し、ニュース検索などでは取得できない、キ

ーワードそのものに対する説明をWikipediaから取得する[16]。キーワードの説明として、「第一文」, 「概要」, 「関連項目」の3つに注目した。これらを提示することにより、ユーザの興味を広げ、関連語を次の検索へ役立てることができる。と考える。

4.3 新規性

その情報がどの程度流行を反映しているか、という指標を考え、その尺度を「新規性」と呼ぶ。これは、「ユーザが新しい情報を望む傾向にある」という考えに基づいている。取得したWebページがどの程度流行を反映しているか推定するため、ブログ内の出現単語をランキングにしたkizasi.jp[17]のツールであるkizAPI[18]を利用し、上位30キーワードと、共起の関連語情報を使用する。kizasi.jpは、ブログに出現する単語の出現情報を見ることにより、注目キーワードの動向を確認することの出来るサイトである。

キーワードのランキング順位でスコアを調節し、共起の関連語にはそのキーワードの10分の1のスコアを与えることでスコア付けを行う。

4.4 重要度スコアの統合

ユーザの嗜好のスコア $Int(T)$ の算出には以下の式を用いる。

$$Int(T) = 1 - |d_c(T_{vec}, P_{vec})| \quad (2)$$

ここで、 $d_c(T, P)$ は、Web ページ T のベクトル T_{vec} とユーザのプロファイルベクトル P_{vec} のコサイン距離である。ユーザがTwitterへのログインを行った場合、次式を用いて入力テキスト T の統合スコア $S(T)$ を算出する。

$$S(T) = \alpha * Exp(T) + \beta * Nov(T) + \gamma * Rea(T) + \delta * Int(T) \quad (3)$$

ここで、 $S(T)$ は入力されたテキスト T の重要度の総合スコア、 $Exp(T)$, $Nov(T)$, $Rea(T)$ はそれぞれ「専門性」, 「新規性」, 「可読性」の正規化済みスコア、 α , β , γ , δ は各重要度の重み付けに用いる値である。本研究では、可読性の選択、ユーザの嗜好のスコアによる提示文書の変化を調査し、適度な変化を持たせるため、重みの値を $\alpha = 1$, $\beta = 1$, $\gamma = 1.5$, $\delta = 1.5$ のように設定した。また、重要な文は文書の初めの方に存在し易いという考えに基づき、文書が長い場合は「。」までを1区切りとし、400文字以内に納まるよう残りを切り捨てる。

本システムで作成されるプロファイルベクトルを図6に示す。それぞれのプロファイルベクトルを用いてユーザに提示される新聞画面の例を図7, 図8に示す。検索キーワードには「地震」を用いた。紙面の右下部にはキーワードと、kizasi.jpから取得した、現在の流行ワードが書き出される。

economy -> 0.374450	science -> 0.393657
home -> 0.272396	home -> 0.209683
culture -> 0.087305	sports -> 0.177148
science -> 0.068587	society -> 0.064899
sports -> 0.064828	culture -> 0.056446
international -> 0.048819	entertainment -> 0.047299
society -> 0.044045	international -> 0.029771
entertainment -> 0.039569	economy -> 0.021097

図6 ユーザX(左)とユーザY(右)のプロファイルベクトル

図7 ユーザXの提示結果の例

図6のプロファイルベクトルより、ユーザXは科学と家庭、ユーザYは経済と家庭の数値が大きいことが分かる。また、提示結果を見てみると、ユーザXの結果では地震のメカニズムやプレートの仕組みの話など、科学寄りの話題が上位に提示されている。しかし、ユーザYの結果では保険の制度や費用などの経済寄りの話題が上位に提示されている。このことから、ユーザのプロファイルベクトルを推定することで、それに合った検索結果の提示を行えていることが示された。

4.5 評価実験

提示結果を用いて、Twitter 情報を用いたユーザの嗜好推定についての評価を行った。被験者 10 人に対して以下の内容で評価を行った。

- 作成されたプロファイルベクトルについて
 - (1) 推定結果は妥当であるか
 - (2) 好みに合った推定結果か
- 提示された検索結果について
 - (3) プロファイルベクトルに合った結果はどちらか
 - (4) 好みに合った結果はどちらか

被験者には個々のプロファイルベクトルとともに、実際に本システムで出力される、①Twitter 情報のスコアを統合した結果、②Twitter 情報のスコアを統合していない結果(専門性、可読性、新規性のみ)の2種類を提示し、評価してもらった。

図8 ユーザYの提示結果の例

プロファイルベクトルについての評価では、普段のツイートの内容を考慮して妥当な推定結果が出ているかを(1)で、また、ツイートの内容にかかわらず、好みに合った推定結果であるかを(2)で評価する。

提示結果についての評価では、プロファイルベクトルを反映した提示結果はどちらかを(3)で、また、プロファイルベクトルにかかわらず好みに合った結果はどちらかを(4)で評価する。

プロファイルベクトルについての評価結果を図9に、提示結果についての評価結果を図10に示す。

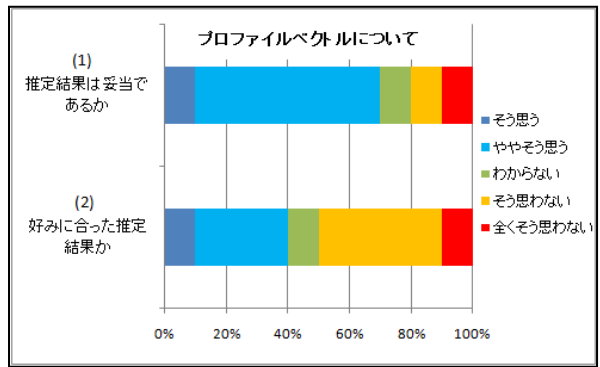


図9 評価実験：プロファイルベクトルについて

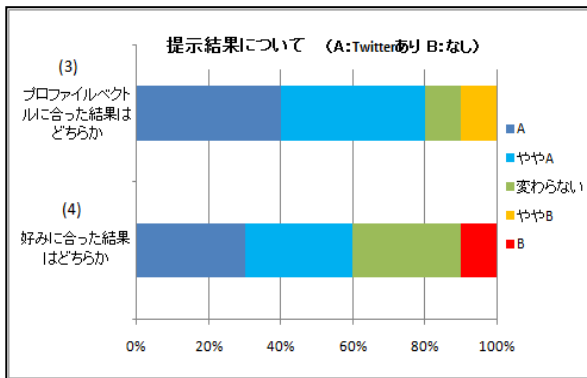


図10 評価実験：提示結果について

図9の(1)の結果より、「そう思う」、「ややそう思う」の結果が多く選択されていることが分かる。これは、プロフィールベクトルの推定によりユーザのツイートをよく推定できていることを示している。また、(2)の結果では評価が分かれてしまった。これは、ユーザは必ずしも興味のあることばかりをつぶやいているわけではなく、比較的最新のツイートしか取得できていないことが原因であると考えられる。また、「なう」や投稿の引用ではすべての嗜好を取得できていない。よって、ユーザの嗜好を表すつぶやきの取得方法を新たに検討する必要がある。

図10の(3)、(4)の結果より、Twitterを用いた結果が多く選択されていることが分かる。これより、ユーザの嗜好推定を本システムに統合した際も、プロフィールベクトルに合った検索結果を提示できることが示せたとともに、ユーザの嗜好に合わせた結果を提示できていることを示すことができた。

5. まとめ

本研究では、マイクロブログの投稿からユーザの嗜好を推定し、Web検索結果を反映させた検索新聞を作成する枠組みを提案した。嗜好を推定するため、Twitterで特徴的な表現である「～なう」やユーザの投稿の引用を含むものを用いて、ジャンル推定を行った。ジャンル推定結果をベクトル化することで、プロフィールベクトルを生成し、嗜好推定に用いた。また、検索新聞で用いられる3種類のスコアとの統合により、検索新聞のシステムでユーザの嗜好に合わせた結果の提示を試みた。

評価実験では、プロフィールベクトルの妥当性、検索新聞での提示結果の有用性についての評価を行った。その結果、本手法が有効であり、Twitterのユーザの投稿から嗜好を推定し、ユーザに合ったWebページが選択できることを示した。また、検索新聞との統合によりユーザの嗜好に合った情報を提示できていることが評価できた。しかし、「なう」や投稿の引用がユーザの嗜好をすべて表しているとは言えない。よって今後は、それ以外の情報からユーザの嗜好を表したつぶやきを取得する方法、また、ジャンル推定以外にユーザの嗜好を推定する方法を検討していきたいと考えている。今回は10人の被験者で実験を行ったが、今後は人数を増やし、年代層、性別なども考慮した被験者実験を行うことで、本システムの有用性を確かめる必要がある。

また、本研究をサービスとして実用する場合、著作権に関する問題が発生することが考えられる。サーバで解析した結果をもとに、ユーザ側で情報取得を行うこと等により問題を回避できる可能性があるが、これに関しては今後調査を行っていく必要があると考えられる。

参考文献

- [1] 祖父江 翔, 瀬谷 将士, 山本 孝二, 田村 哲嗣, 速水 悟. “検索新聞：新聞形式による検索情報要約システムの提案.” 人工知能学会 2010 年全国大会 (JSAI2010), 3D1-2, 2010
- [2] 関谷 英樹, 祖父江 翔, 田村 哲嗣, 速水 悟. “検索新聞：可読性に着目した検索支援システム.” 人工知能学会 2011 年全国大会 (JSAI2011), 発表予定
- [3] 小原 恭介, 山田 剛一, 絹川 博之, 中川 裕志. “Blogger の嗜好を利用した協調フィルタリングによる Web 情報推薦システム.” 人工知能学会 2005 年全国大会 (JSAI2005), 2C2-02, 2005.
- [4] 池田 大介, 南野 朋之, 奥村 学. “blog の著者の性別推定.” 言語処理学会 第 12 回年次大会 (NLP2006), pp.582-585, C2-3, 2006.
- [5] 白木 敦夫, 矢野 幹樹, 酒井 佑太, 小澤 俊介, 杉木 健二, 松原 茂樹, 河口 信夫. “モバイルアプリケーション推薦のための Twitter 発言者の状況の推定.” マルチメディアの分散, 強調とモバイル (DICOMO2010), pp.251-257, 2010.
- [6] Jianshu Weng, Ee-peng Lim, Qi He, and Cane Wing-Ki Leung. “What Do People Want in Microblogs? Measuring Interestingness of Hashtags in Twitter.” In Proceeding of the IEEE International Conference on Data Mining, 2010.
- [7] KDDI 研究所. “「つぶやき」から投稿者のプロフィールを自動推定する技術の開発に成功.” http://www.kddilabs.jp/press/detail_167.html.
- [8] Twitter 新聞発行サービス, paper.li. <http://paper.li/>.
- [9] Twitter developers. <http://dev.twitter.com/>.
- [10] John Lafferty, McCallum Andrew, and Pereira Fernando. “Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data.” In Proceeding of the 18th International Conference on Machine Learning, 2001.
- [11] Yahoo! デベロッパーネットワーク. <http://developer.yahoo.co.jp/>.
- [12] 毎日新聞記事データ集. <http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html>.
- [13] 工藤 拓. CRF++: Yet Another CRF toolkit. <http://crfpp.sourceforge.net/>.
- [14] 近藤 陽介, 松吉 俊, 佐藤 理史, “教科書コーパスを用いた日本語テキストの難易度推定” 言語処理学会 第 14 回年次大会 (NLP2008), pp.1113-1116, 2008.
- [15] Yahoo!きっず. <http://kids.yahoo.co.jp/>.
- [16] Wikipedia. <http://www.wikipedia.org/>.
- [17] kizasi.jp, <http://kizasi.jp>
- [18] きざしツール: kizAPI. <http://kizasi.jp/tool/kizapi.html>.