

科学研究費助成事業 研究成果報告書

平成 26 年 5 月 20 日現在

機関番号：32663

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500302

研究課題名(和文)メディア横断型パーソナル・パスファインダー構築のための基礎的研究

研究課題名(英文)Basic research for construction of cross-media personal pathfinder

研究代表者

栗山 和子 (KURIYAMA, Kazuko)

東洋大学・社会学部・教授

研究者番号：10349144

交付決定額(研究期間全体)：(直接経費) 1,500,000円、(間接経費) 450,000円

研究成果の概要(和文)：WWWの利用者の目的に合った情報アクセスツールを提示するための基礎的研究として、他者を意識したWWW上での情報行動という観点から、コミュニティ型サイトの利用に着目し、投稿の表現や形式が利用者の利用行動にどのような影響を与えているかについて調査するため、コミュニティ型質問回答サイト(CQA)「Yahoo!知恵袋」、「OK Wave」、および、マイクロブログ「Twitter」の投稿データを用いて、投稿のタイプやトピックについて分析し、その有用性について考察した。その結果、CQAにおける良質な回答の特徴とヘビーユーザの特徴、Twitterにおける大学図書館の投稿タイプと利用方法を明らかにした。

研究成果の概要(英文)：The purpose of this study is to investigate the information behaviour of the WWW users for recommendation of an information access tool in accord with the user's purpose. To consider the information behaviour with consciousness of others, I examined how they used the community sites. I collected the data which were posted by users on the community question answering (CQA) sites "Yahoo! Chiebukuro" and "OKWave", and microblog "Twitter", analyzed them on the types and the topics, and considered the usefulness. As a result, I showed some characteristics of the high quality answer in CQA and the behaviour of the heavy users, and the tweet types and the use of the university libraries on Twitter.

研究分野：総合領域

科研費の分科・細目：情報学・図書館情報学・人文社会情報学

キーワード：コミュニティサイト CQA ソーシャルメディア SNS 図書館 情報アクセスツール パスファインダ

1. 研究開始当初の背景

- (1) 情報アクセスツールとしてのコミュニティサイトの利用：近年、質問回答サイト(CQA)や SNS などのコミュニティサイトの利用者が増加している。コミュニティサイトは、登録した利用者が(匿名で)投稿することでその仕組みが成り立っているため、従来のサーチエンジンのような情報アクセス支援ツールとしてではなく、コミュニケーションや知識の共有を目的としたものとして捉えられる。
- (2) 利用目的とコミュニティサイト選択のミスマッチの問題：コミュニティサイトの利用者の年齢・性別・職業・利用目的等は様々であり、サイトに投稿される情報・知識の内容や表現は多様である。質問・回答を目的としているはずの CQA においてコミュニケーションを目的とした利用が数多く散見される一方、SNS を過去の投稿から情報や知識を抽出する情報検索ツールとして利用している例も多く見られる。利用者は自分の目的に合わせて使用するサイトを選択しているが、実際には、自分や周囲の人の知識や経験などからサイトを選ぶ傾向にあり、必ずしも本当に目的にあったものを選んでいるとは限らない。そこで、本研究では、コミュニティサイトに投稿された記述から、利用者の求めているコミュニケーションの種類を分析し、数多くのコミュニティサイトや情報アクセスツールの中から利用者の目的に合ったサイトを提示することができないだろうかと考えた。

2. 研究の目的

本研究の目的は、WWW の利用者が必要な情報を探したり、他の利用者とのコミュニケーションをとったり、自分の意見や小説・映像などの芸術的創作物を発表したりするとき、どのように情報アクセスツール(サイト)を使い分けているのかを明らかにし、利用者の目的や内容に合った情報アクセスツールを選択し提示することである。本研究ではコミュニティ型質問回答サイト(CQA)「Yahoo! 知恵袋」、「OK Wave」、および、マイクロブログ「Twitter」の投稿データを用いて、広い意味でのコミュニティサイトの投稿データを多面的に分析し、その表現や構造が利用者の情報アクセスツールの利用行動にどのような影響を与えているかについて検証を行い、様々な WWW 上のメディアから利用者の目的に合うメディアを提示する、メディア横断型パーソナル・パスファインダーの構築の要件について考察する。

3. 研究の方法

- (1) 「Yahoo! 知恵袋」に投稿された回答を対

象に、まず、人間の判定者二名がベストアンサーに相応しい回答を選出する。次に、その結果と「ベストアンサー」情報とを併せて、良質回答の特徴を分析する。そして、その特徴に基づいて良質回答自動予測システムを機械学習によって構築し、そのシステムの評価結果を人間(判定者)と比較する。

- (2) 有用な情報源の一つである図書館および図書館資料に利用者を誘導するにはどのような方法が効果的であるか検討するために、大学図書館のソーシャルメディア利用について考える。具体的には、日本の国立大学の付属図書館が Twitter をどのように利用しているのか、簡単な実態調査を行い、大学図書館の広報活動および図書館サービスのツールとしての Twitter の有効性について考察する。
- (3) 「OKWave」に投稿された質問を収集し、その質問を分析することにより、質問の投稿数の多いヘビーユーザの特徴について考察する。また、ヘビーユーザの利用が最も多い「アンケート」カテゴリに注目し、質問投稿数の少ないライトユーザと質問投稿数の多いヘビーユーザとでは、質問行動や質問の内容に違いがあるのかどうかについても簡単な分析を行い、ヘビーユーザがどのような回答を求めているのかについて考察する。

4. 研究成果

- (1) 「Yahoo! 知恵袋」から「恋愛相談」「パソコン」「一般教養」「政治」のカテゴリを対象に、判定者によるベストアンサー判定実験を行い、得られた結果とベストアンサー情報に基づいて、良質回答の特徴を分析した。

まず、判定者によるベストアンサー判定実験の後、良質回答の特徴として考えられるものを判定者間で議論し、以下のようにリストアップした。

- a. 【分かりやすさ】記述が分かりやすいもの、読みやすいもの。
- b. 【詳しさ】詳しい説明があるもの。詳しいけれど、説明がよくわからないものは BA に選択していない。
- c. 【根拠】情報の根拠となる出展(URL)があるもの。あるいは自分や他人の体験に基づいているもの(自分の体験 > 他人の体験 > 体験なし、推察)
- d. 【丁寧さ】日本語が丁寧であるもの。日本語が丁寧な回答を BA に選択するケースが多い。
- e. 【ポジティブ】質問者にポジティブなコメントを含むもの。例えば、「がんばってください」などの表現があるもの。
- f. 【的確さ】質問に対して的確で、説得力があるもの(ただし根拠は示されていない)

ない)。

次に、良質回答の特徴リストとその集計結果に基づいて、サポートベクターマシン(SVM: Support Vector Machine)による機械学習に与える素性を定義し、特に多かった特徴と計算が比較的容易な特徴として、b.【詳しさ】、c.【根拠】、d.【丁寧さ】の3つの特徴を用いて機械学習を行った。機械学習システムが判定者らの一致率を上回った件数は、「パソコン」では7件(14%)と9件(18%)、「一般教養」では5件(10%)と4件(8%)である。これらを取り出して機械学習が上回った理由を調査した。例えば詳細に記述されている回答と、詳細には記述されていないが根拠(となるURL)が示されている回答とがあった場合(ただし回答の内容はほぼ同じ)、機械学習システムと判定者は選択した回答が異なっていた。機械学習システムは前者の詳細な回答の方を選択し、判定者らの方は後者の根拠がある回答の方を選択する傾向にあった。今回の評価データでは、このようなケースは前者(詳細にかかれていないが根拠が示されていない)がベストアンサーとなる傾向にある。機械学習システムは、訓練データ(500件)を学習する上で、その傾向を学習したため、その結果が反映して、一致率が向上したと考えられる。

各カテゴリに対する有効な素性を、特徴としてまとめると、「パソコン」と「一般教養」のような特定の事実を聞く情報検索型タイプのカテゴリで【丁寧さ】の重要性が確認された。一方、「恋愛相談」と「政治」のような他人の助言を広く聞く社会調査型タイプのカテゴリで【詳しさ】が求められている。このように、情報検索型タイプで【丁寧さ】が、社会調査タイプで【詳しさ】が重要であるという新たな知見が得られた。また、【詳しさ】が有効に働かないカテゴリ(一般教養)があることも見出された。

以上の結果から、今後、機械学習した計算機によって膨大なCQAアーカイブから良質な回答を自動的に発見できる可能性が示された。また本研究では、訓練データの使用に適したカテゴリについても分析し、「恋愛相談」カテゴリで学習したモデルは、他のカテゴリで学習したモデルよりも適応能力が高く、さらに全カテゴリから学習したモデルともほぼ同程度の評価結果が示された。この知見は、CQAアーカイブからの良質回答自動発見システムを構築する際、訓練データセットの最適化に関する有効な手がかりとなると考えられる。

- (2) 国立大学の付属図書館・図書室のTwitterアカウントを同定し、公式アカ

ウントのツイートの内容について分類を行った。

国立大学86校のうち、附属大学図書館・図書室もしくはそれに類する部署に関連するTwitterアカウントは65であった。1つの大学が複数のアカウントを使用している場合があるので、アカウントを運用している大学は37校である。そのうち、自己紹介などから公式であると確認できたのは47で、30校である。公式アカウントの内訳は、公式(試行)9、公式(自動)2(3)、公式(学生)3である。フォロー数、フレンド数、ツイート数は、利用期間が長くなるほど多くなっているように見えるが、これも平均をとったためである。いずれもアカウントごとのばらつきが大きく、一定の傾向はないと考えられる。各アカウントについて、フォローとフレンドが重複している割合を、フォローとフレンドの重複している数/フレンド数として計算すると、重複の割合の平均は47.9%であり、フレンドの約半数がフォローと重複していることがわかった。これは、あるアカウントからフォローされた場合にそのアカウントをフレンドに登録する、または、あるアカウントをフレンドとして登録するとそのアカウントからフォローされる、相互的なフォローがTwitter上では多いためであると考えられる。

内訳としては、開館・閉館時間/施設利用、図書館イベント、新着図書・雑誌紹介に関するツイートが多く、貸出・閲覧・複写、図書館システムの保守・管理、(レファレンスサービスが少ない傾向にあることがわかる。開館・閉館時間/施設利用が多くなった理由は、ツイートを収集したのが8月15日だったため、最新の100件の中では、期末試験期間中および夏季休暇期間中の開館・閉館時間の変更や館内施設・設備の工事についてのツイートが多くなったと考えられる。ほとんどの大学図書館では公式サイトのトップページに開館日時を掲載しており、トップページに掲載していない場合でもワンクリック先のページに掲載しているため、開館・閉館時間を頻繁にツイートする必要性は低いと考えられる。図書館イベントは、来館する予定のない利用者にもイベントを知らせる手段としてツイートが利用されていると考えられる。「拡散希望」のように、積極的にリツイートを勧める表現を含むものもあり、Twitterを広報ツールとして活用しているといえる。新着図書・雑誌紹介のツイートが1番目に多かったアカウントは12であるが、そのうち8アカウントは、ツイートの半数以上が新着図書・雑誌紹介であり、1つは、RSSフィ

ードから Twitter に自動的に投稿をしている。貸出・閲覧・複写、図書館システムの保守・管理については、貸出・閲覧・複写や図書館システムの保守・管理についてのお知らせがなければツイートされないため、少ないと考えられる。レファレンスサービスについては、Twitter の自己紹介やツイートの中で、ツイートの返信をされても返信はできないことを明記している図書館もあり、Twitter を対話的に利用する意図がない場合も多く、返信(Reply)の利用も少ない。以上のことから、図書館の利用やイベント・展示に関するお知らせとしては積極的に利用されているものの、Twitter のソーシャルメディアとしての性質を利用して、対話的なコミュニケーションを行っている国立大学の大学図書館はほとんどないことがわかった。

結果として、ツイートの内容としては、開館・閉館時間や施設・設備の利用、図書館の行事・イベント、新着図書・雑誌紹介に関するツイートが多く、フォローやフレンドとのやりとりはほとんどないことがわかった。Twitter や Facebook などのソーシャルメディアの利点は、リアルタイム性や双方向の対話性にあるが、大学図書館の Twitter 利用では、まだ、その利点を十分活用できていないと考えられる。大学図書館の主な利用者は、学生と教職員であり、特に、学生は、ソーシャルメディアを積極的に利用している年代であるので、新着資料の紹介にとどまらない、学生のニーズや興味に合わせた速報的なサービスを行うことが望ましいと思われる。

- (3) 「OKWave」に投稿された質問を収集し、その質問を分析した結果、ヘビーユーザは、質問回数が少ないライトユーザに比べて、回答に対するお礼コメントを多く投稿しており、回答者とのコミュニケーションや CQA のマナーを重視していることがわかった。また、「OKWave」全体では、「恋愛・人生相談」のカテゴリに対する質問が多いが、ヘビーユーザは、「社会問題」と「アンケート」のカテゴリに多く投稿しており、自分自身の個人的な相談よりも、社会的な出来事や他人の意見・体験に興味を持っていると考えられる。
(本研究結果は、論文として投稿中であるため、研究成果は概略に留める。)

5 . 主な発表論文等

〔雑誌論文〕(計2件)

石川 大介、酒井 哲也、関 洋平、栗山 和子、神門 紀子、複数判定者によるコミュニティ QA の良質回答判定、情報知識

学会、査読無、Vol.21、No.2、2011、pp.169-177

DOI: 10.2964/jsik.21_08

石川 大介、酒井 哲也、関 洋平、栗山 和子、神門 紀子、コミュニティ QA における良質回答の自動予測、情報知識学会、査読有、Vol.21、No.3、2011、pp.362-382

DOI: 10.2964/jsik.21-041

〔学会発表〕(計1件)

栗山 和子、大学図書館の Twitter 利用、情報学基礎研究会報告、筑波大学東京キャンパス文京校舎、2013年9月26日、Vol.2013-IFAT-112、No.3、2013、pp.1-8

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

6 . 研究組織

(1)研究代表者

栗山 和子 (KURIYAMA, Kazuko)

東洋大学・社会学部・教授

研究者番号: 10349144

(2)研究分担者(0)

(3)連携研究者(0)