

ユーザの関心を含むマイクロブログのタイムライン情報をクエリ拡張に適用する手法の提案

岡本 里夏^{1,2,a)} ガルカビス, ビクトルス^{1,2} 大山 敬三^{2,1}

概要：マイクロブログサービスの一つである Twitter は、様々なユーザにコミュニケーション手段として利用されている。例えば、学術コミュニティにおいては学会開催の周知や研究者や学生同士論文の紹介、さらには Twitter ユーザ自身の考えなどのメモ代わりに使われる例もある。このように、マイクロブログのユーザは他のユーザをフォローする事によって、自身の関心のある情報の入手あるいはユーザ自身の情報発信に活用している。つまりマイクロブログユーザのタイムラインには、ユーザ自身にとって関心の高い情報が含まれている。

本報告では、ユーザ自身の関心と関連度の高いと考えられる、ユーザ自身及びフォロー先ユーザが発信した Twitter のタイムライン情報を用いたクエリ拡張の手法を提案する。また、その提案手法について、Text REtrieval Conference (TREC) の 2013 年の AdHoc Task で使用されたテストコレクションを利用し、疑似レlevanceフィードバックに適用した実験を行った。その実験の結果と考察についても報告する。

1. はじめに

研究者にとって大量な文献から自分の研究に関連度の高い適切な文献を検索してくることは以前と変わらず切実な問題である。しかし、かつて図書館やデータベースサービスで専門技術者や専門のサービスに文献検索を依頼していた時代とは違い、PC やインターネット接続の普及と高速化、および商用の Web 検索サービスの普及により、研究者自らが文献を検索する機会が格段に多くなった。また、論文やプロシーディングスも Web 上で書誌情報と関連づけられた PDF 形式のファイルで提供されたり、機関リポジトリに収録されたりすることが多くなり、こうした場面でも出版社の Web 上に設置された検索質問（以下、クエリ）の入力箇所（以下、検索窓）からキーワードを入力して論文を探すことが日常化している。

一方で、SNS (Social Network Services) を含む各種 Web サービスでは、様々な情報を交換をする機会が多くある。友人との会話や家族とのコミュニケーションを楽しむために参加する個人だけではなく、学術的な話題でも SNS 上で論文や関連する情報の発信が多く行われている。様々な情報に触れる機会が増えた一方で、大量の電子化されたテキ

ストがインターネット上に流通するようになった。そのため、大量の情報の中から自分に必要な文書や、紹介された情報の中から関連度の高いトピックをもれなく素早く探し出すためには、人力によるディレクトリ型の検索サービスでは追いつかず、クエリ入力による全文検索エンジンの力を借りる必要がある。しかし、多くの人が経験するように、デスクトップや Web において情報検索する際にはユーザが困難に直面する事がしばしばである。

昨今言われる情報爆発という言葉にあらわされる事態は、研究者に対する影響だけにとどまらない。なんらかの情報を PC や携帯端末で検索する必要がある人にとっても、日常の様々な場面でインターネット上の Web 情報や自分のデスクトップ上に溢れる各種ファイルの中から、自分にとって良い情報を得ることを困難にしている。この課題を解決するため、今日に至るまで多くの研究者によって検索機能の改善に向けた研究が進められている。ユーザが得られる検索結果の改善のための努力はしつづけられているが、検索システムが人の手を介さずに人々の興味関心を上手くとらえることは困難である。

現在に至るまで、そうした困難を改善するために数多くの提案がなされてきた。例えば、日々使用する PC やモバイル端末を使用するユーザのふるまいからユーザの興味関心を類推しそれを検索エンジンにフィードバックする仕組みもその一つである。しかし、ユーザは情報要求をすべてクエリとして表現できるわけではないし、ユーザの情報要

¹ 総合研究大学院大学複合科学研究科情報学専攻
〒101-8430 東京都千代田区一ツ橋 2-1-2

² 国立情報学研究所
National Institute of Informatics

a) lica@nii.ac.jp

求が直接ユーザのふるまいに現れる訳でもない。何故ならば、検索したい事とそれに関連して想定される自分の行動や思考を言語情報として簡潔に再現し入力すること自体が困難だからである。ユーザは何を検索したいのか具体的な検索のためのキーワードを思いつかずに、とりあえず脳裏に浮かんだ言葉を入力して検索することもあれば、そもそも「具体的な何か」を検索したいのではなく漠然とした「何か面白いもの」を探すために思いついた言葉を入力して検索することもある。

さらに、ユーザの情報要求や検索意図、等の情報は多岐に渡る。検索ユーザが検索結果を見て自分の検索したい内容を変えることもある。そのため、ユーザが自分の検索したいことを自分の情報要求を反映させてシステムに対して入力することは現状では実現困難である。

ところで、SNSは昨今電子化された言語資源としても注目されている。そうしたSNSの一つであるマイクロブログサービスのTwitter^{*1}は、様々なユーザによって多岐にわたるコミュニケーション手段として利用されている。例えば、学術コミュニティにおいては学会開催の周知や研究者や学生同士論文の紹介、さらにはTwitterユーザ自身の考えなどのメモ代わりに使われる例もある。マイクロブログのユーザは他のユーザをフォローする事によって、自身の関心のある情報の入手あるいはユーザ自身の情報発信に活用している。つまり、マイクロブログユーザのタイムラインにはユーザ自身にとって関心の高い情報が含まれている。これらの情報から各ユーザの潜在的な情報要求を推測することができれば、検索の結果をユーザの情報要求により近づけることができる。

我々は検索ユーザの興味関心をフィードバックするものとして、マイクロブログでやりとりされる情報に注目している。それらを用い、ユーザの検索体験の向上を目的として、デスクトップ、モバイル端末、ソーシャルネットワークサービス(SNS)等で活動するユーザのふるまいを記録し、そのログの解析結果から暗黙的レlevance・フィードバックを行うシステムを試作する。

本報告では、ユーザ自身の関心と関連度の高いと考えられるユーザ自身及びフォロー先ユーザが発信したTwitterのタイムライン情報を用いたクエリ拡張の手法を提案する。また、その提案手法について、Text REtrieval Conference (TREC)の2013年のAdHoc Taskで使用されたテストコレクションを利用し疑似レlevanceフィードバックを適用した実験を行ったので、その結果について考察する。

2. 関連研究

本研究においては、ユーザのSNSにおける閲覧行動や他ユーザのフォローなどのふるまいやSNS上のコンテン

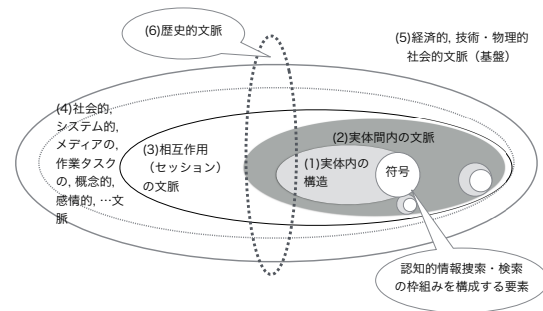


図1 情報検索・検索に関する文脈階層の入れ子モデル
 「情報検索の認知的転回—情報検索と情報検索の統合」[6]
 p.182, 図 6.7 より再構成)

ツを検索ユーザのコンテキストを構成するものと捉える。ここでは、検索のコンテキストを情報検索に利用するための関連研究について概観する。

2.1 検索コンテキスト

検索コンテキストをあらわすものについて、IngwersenとJärvelinはネスト状のモデルを提案している[図1 情報検索・検索に関する文脈階層の入れ子モデル][5][6].

- (1) Intra object structures (実体内の構造)
 例: 探索者の心にある様々な認知構造, 一連のプログラム, 等
- (2) Inter-object contexts (実体間内の文脈)
 例: 文書属性 (他の情報物への参照または発リンク, 引用または被リンク)
- (3) Interaction (session) context (相互作用 (セッション) の文脈)
 例: 探索者をモデルの核とする場合→社会的相互作用や相互作用型の情報検索活動, インタフェースをモデルの核とする場合→セッション (検索式の拡張過程のようなシステム内相互作用を含む)
- (4) Social, Systemic, Media, Worktask, Conceptual, Emotional ... contexts (社会的, システム的, メディアの, 作業タスクの, 概念的, 感情的...文脈)
 例: 内側の輪→セッションの枠組みのITや情報物と概念的・感情的特色を持つ探索者, 外側の輪→より遠隔の社会組織的文脈
- (5) Economic techno-physical- and societal contexts (Infra-structures) (経済的, 技術・物理的社会的文脈 (基盤))
- (6) Historical Context (歴史的文脈)

このモデルを今回の実験について当てはめると、「実体間内の文脈」と「相互作用 (セッション) の文脈」の一部に対応していると理解できる。

*1 Twitter, <https://twitter.com/> (accessed 2014-10-20.)

2.2 検索コンテキストについての取組み

検索のコンテキストについては、SIGIR 2011 で開催されたワークショップ “entertain me” [2] において検索コンテキストについて話し合われた。

また、2012年2月にオーストラリアの Lorne で開催された The Second Strategic Workshop on Information Retrieval in Lorne (以下 SWIRL 2012) という今後の情報検索研究の方向性を戦略的に議論するためのワークショップにおいて、Capturing context として、主題の一つにあげられている [1]。

Text REtrieval Conference (以下、TREC とする) *2 においても検索コンテキストの重要性が認識され、2012年に最初の Context suggestion Track[3] が実施された。タスク参加者にはコンテキストのサンプルが与えられ、それを用いて Web からユーザのコンテキストに沿った 50 件のリンク付きのサジェスションを作ることが求められた。

2.3 ユーザ・コンテキストの探索と利用

Dumais は 2012 年の SIGIR の講演で、ユーザのコンテキストの利用とコンテキストを探すという二つの側面について言及している [4]。その中で、コンテキストはクエリを改善するものとして言及している一方で、ユーザのコンテキストを利用する際の課題についても述べている。Dumais はユーザ側の課題として、

- Privacy (プライバシー)
- Transparency and control (透明性と制御)
- Consistency (一貫性)
- Serendipity (偶有性)

を挙げている。一方、サーバ側の課題として、

- System optimization (Storage, run-time, caching, etc.) (システム最適化：記憶装置、実行時、キャッシュ等)
- Evaluation (評価)

を挙げている。

一般に、商用検索エンジンでは、検索結果の改善のために、多くのユーザのサーバ側のログを使用している。しかし、特定のユーザが複数のアクセス先を探索したログは統合できない、商用サーチエンジン企業外部では利用できない、プライバシーの問題、等が存在する。

2.4 マイクロブログの検索

Miyaniishi らはマイクロブログを対象とした検索についての研究を行った。彼らはツイートを選択した結果を元に、疑似レlevance・フィードバックによってクエリ拡張手法を Twitter タイムラインの検索結果の改善に適用することを提案している [7]。

2.5 ユーザのふるまいの利用

土岐らは Twitter タイムラインの閲覧時のふるまいを利用してユーザ・プロフィールを構成し、そのプロフィールとユーザが見落とした Twitter タイムラインの情報からユーザが興味のあるツイートをマッチングによって抽出しユーザに推薦する手法を提案している [8]。

3. 実験システム

本章では、今回実験用に試作したシステムについて述べる。今回の実験で対象とするマイクロブログサービスは Twitter とした。

3.1 検索エンジン

検索エンジンは Lemur Project の Indri 5.6^{*3*}4 を使用した。

3.2 Twitter 閲覧用モバイル端末向けソフトウェア

今回の実験のために、Twitter のユーザ・タイムラインを閲覧した時のユーザのふるまい (閲覧履歴、操作履歴、等) を取得するためのアプリケーション・ソフトウェア (以下、Twitter ブラウザ、とする) を Objective-C を用いて実装した。Twitter ブラウザは Apple Inc. が iPhone や iPad 用に提供しているオペレーティングシステムである iOS の ver. 7.1.1 上で動作する。

3.3 デスクトップ用 Twitter タイムライン取得用ソフトウェア

Twitter Inc. が提供する Streaming API を用いたユーザ・タイムライン取得のためのソフトウェア (以下、Streaming API アプリ、とする) を Objective-C で実装した。Streaming API アプリは Apple Inc. の OS X Mavericks(10.9.3) 上で動作する。今回の実験では、この Streaming API アプリを利用して、Twitter のユーザタイムラインを取得した。

4. 実験

今回の実験に際しては、マイクロブログを閲覧するときのユーザの文脈をあらわす素性として、Twitter のユーザ・タイムラインに注目した。

4.1 目的

今回の実験の目的は、次の3点である。

- Twitter ユーザ・タイムラインを利用して疑似レlevance・フィードバックを行った場合と、疑似レlevance・フィードバックを行わない場合を比較する。
- Twitter ユーザ・タイムラインを利用して疑似レlevance

*2 Text REtrieval Conference (TREC), <http://trec.nist.gov/> (accessed 2014-10-20.)

*3 <http://www.lemurproject.org/indri.php> (accessed 2014-10-20.)

*4 Indri の 2014-10-20 現在の最新バージョンは 5.7 である。

| | |
|------------------------------|-------------|
| Size compressed: | 5.54 TB |
| Size uncompressed: | 27.3 TB |
| Number of WARC files: | 33,447 |
| Number of documents: | 733,019,372 |

図 2 ClueWeb12 の統計データ

ンス・フィードバックを行った場合と、ClueWeb12 をフィードバックを行う対象の文書とした場合を比較する。

- 実験での検索結果について、検索が成功した場合と失敗した場合の内容を調査する。

4.2 手順

実験は次の手順で行われた。

- (1) Twitter を閲覧する時のユーザのふるまいを記録するソフトウェアを作成する。
- (2) 用意したペルソナに適合する Twitter ユーザのフォローを行う
- (3) Twitter のタイムラインを別に用意したソフトウェア (Twitter Streaming API を使用) で取得する。
- (4) 取得したタイムラインを Indri が読み込める形式 (TREC TEXT 形式) に加工する。
- (5) 加工したタイムラインを IndriBuildIndex でインデックスを作成し、TREC Web Task のクエリを使用して検索する。
- (6) 別途作成したプログラムを用いて、加工したタイムラインの検索結果から上位 5 件を選び、擬似的なレバンスフィードバック文書として Indri のクエリパラメータファイルを作成する。
- (7) 作成したクエリパラメータファイルを用いて ClueWeb12 のデータセットを検索する。
- (8) 出力された検索結果とクエリ拡張の結果ログを保存する。
- (9) 検索結果は t 検定を行い、クエリ拡張のログの分析を行う。

4.3 データセット

実験に使用するデータセットは次に挙げる 2 種類のデータセットを使用した。

4.3.1 ClueWeb12

ClueWeb12 は Lemur Project が提供する Web ページのコレクションで、2012 年 2 月 10 日から同年 5 月 10 日までに集められた英語の Web ページ 733,019,372 件を含む [図 2 ClueWeb12 の統計データ *5]。ClueWeb12 は 2013

*5 Lemur Project, The ClueWeb12 Dataset,

年の TREC Web Track, AdHoc Task で用いられた。今回そのテストコレクションを用いて実験を行うため、このデータセットを検索対象のデータセットとした。

4.3.2 Twitter とユーザ・タイムライン

今回の実験にあたって、検索するユーザが Twitter のタイムラインを情報収集のために読んでいることを想定している。このユーザ・タイムラインと呼ばれる Twitter ユーザが送信した文の集合を取得するために、Twitter Inc. 提供の API を用いた。

また、自分以外のユーザが送信した文のみを素性として使うことを目的としていることから、mention と呼ばれるユーザ間の公開通信がタイムライン内に混入することを避けるため、使用したアカウントを非公開 (Web 上にユーザの発言が公開されない設定) とした。実験では、Twitter のユーザ・タイムラインは、疑似レバンス・フィードバックをするための文書集合として用いた。以下に取得したタイムラインの属性について記す。

- Twitter ID: _twi_twi
- フォローしている Twitter アカウント数: 642 人 (2014 年 7 月 13 日現在)
- ツイートの取得期間: 2014 年 5 月 26 日 ~ 6 月 25 日
- 取得した英語ツイートの件数:
- データセット 1 ... 2991 件
 - 2014-05-26 ~ 2014-06-01 の期間で取得。
- データセット 2 ... 9100 件
 - 2014-05-26 ~ 2014-06-25 の期間で取得。
 - データセット 2 はデータセット 1 を含む。重複したツイートは Twitter タイムラインをマージする際に排除される。
- 文書の形式: Twitter API を使用して取得した JSON 形式のデータを、プログラムにより TRECTEXT 形式 (XML) に変換して保存。

4.4 テストコレクション

実験に使うクエリとその正解データについては、TREC Web Track 2013, AdHoc Task*6 のテストコレクションを使用した。

4.5 実験システムの種類

今回の実験用のシステムには、次の 3 つのシステムを用意した。

4.5.1 ベースライン・システム (baseline)

Indri 5.6 デフォルトの設定で IndriRunQuery という検索クエリを実行するプログラムを動かしたものをベースラ

Dataset Details, Summary Statistics より引用。
<http://lemurproject.org/clueweb12/specs.php> (accessed 2014-10-20.)

*6 TREC Web Track 2013, <http://research.microsoft.com/en-us/projects/trec-web-2013/> (accessed 2014-10-20.)

インとした。主なデフォルト値は次の通りである。

- 取得件数：1,000 件
- 疑似レlevance・フィードバック：無し
- クエリ言語：Indri Query Language
- スムージング: Dirichlet smoothing
($\mu = 2500$, μ はスムージングパラメータ。
2500 は Indri のデフォルト値)
- 結果フォーマット：TREC

4.5.2 実験システム1 (ClueWeb12-PRFB)

Indri 5.6 の疑似レlevance・フィードバックを使ったシステムである。フィードバックする文書は ClueWeb12 である。Twitter のユーザ・タイムラインをフィードバックする文書として検索した場合の比較対照の為のシステムである。

この実験のために、疑似レlevance・フィードバックの値は、5つの文書から10語の単語を選ぶように設定し、それ以外はベースライン・システムと同様とした。この実験システム1については、クエリ発行の際にストップワードなどの処理は行わなかった。

4.5.3 実験システム2 (Twitter-PRFB)

Twitter のユーザ・タイムラインをフィードバックする文書として検索するシステムである。疑似レlevance・フィードバックを行うために、まずテストコレクションのクエリを用いて Twitter のユーザ・タイムライン (ユーザがフォローする Twitter ユーザが送信した「ツイート(tweet)」と呼ばれる140文字以内の文の集合) から作成した検索用データセットを対象にクエリを用いて検索する。ツイートの検索結果上位5位までを疑似レlevance・フィードバックを行うための文書とし、実験システム1と同様に、5文書から10語選ぶように設定を行い、それ以外はベースライン・システムと同様とした。実験システム2についても、クエリ発行の際にストップワード等の処理は行わなかった。

4.6 Twitter 文書の疑似レlevance・フィードバックの検証

先に述べた3つのシステム以外に、Twitter 文書を疑似レlevance・フィードバックする時のクエリ拡張に使う文書数や語の数を変えた場合、検索結果が異なるかを調べた。この検証のためにツイートされたそのままの文を1文書とした「URL有り」の場合と、<http://t.co/> など、ツイートの含まれる URL を削除した「URL無し」の場合についてクエリ拡張される語について確認を行った。実行単位ごとに、少なくとも1文書あたり1語より多く、文書数の2倍までの語が展開されるように設定した。[表1 疑似レlevance・フィードバック検証用パターン]^{*7} に示したこの検証での独立変数は、疑似レlevance・フィードバックに使う

文書数、展開する語数、URLの有無である。なお、この検証に際しては、表中の Twitter-PRFB システムに分類されたものはストップワード処理を行い、拡張されたクエリを表示できるようにクエリ発行時に指定している。

表1 疑似レlevance・フィードバック検証用パターン

| RunID | システムの 種類 | フィード バックす る文書数 | 展開す る語数 | ツイートのURL の有無 |
|--------------------------|--------------------|----------------------|------------|-----------------|
| baseline | ベースライン | なし | N/A | N/A |
| indri-prf * | ClueWeb12- PRFB | 5 | 10 | N/A |
| twitter-prfb-1d2t | Twitter-PRFB | 1 | 2 | 有 |
| twitter-prfb-2d3t | Twitter-PRFB | 2 | 3 | 有 |
| twitter-prfb-2d4t | Twitter-PRFB | 2 | 4 | 有 |
| twitter-prfb-3d4t | Twitter-PRFB | 3 | 4 | 有 |
| twitter-prfb-3d5t | Twitter-PRFB | 3 | 5 | 有 |
| twitter-prfb-3d6t | Twitter-PRFB | 3 | 6 | 有 |
| twitter-prfb-4d5t | Twitter-PRFB | 4 | 5 | 有 |
| twitter-prfb-4d6t | Twitter-PRFB | 4 | 6 | 有 |
| twitter-prfb-4d7t | Twitter-PRFB | 4 | 7 | 有 |
| twitter-prfb-4d7t | Twitter-PRFB | 4 | 8 | 有 |
| twitter-prfb-5d6t | Twitter-PRFB | 5 | 6 | 有 |
| twitter-prfb-5d7t | Twitter-PRFB | 5 | 7 | 有 |
| twitter-prfb-5d8t | Twitter-PRFB | 5 | 8 | 有 |
| twitter-prfb-5d9t | Twitter-PRFB | 5 | 9 | 有 |
| twitter-prfb-5d10t | Twitter-PRFB | 5 | 10 | 有 |
| twitter-prfb-noURL-5d6t | Twitter-PRFB | 5 | 6 | 無 |
| twitter-prfb-noURL-5d7t | Twitter-PRFB | 5 | 7 | 無 |
| twitter-prfb-noURL-5d8t | Twitter-PRFB | 5 | 8 | 無 |
| twitter-prfb-noURL-5d9t | Twitter-PRFB | 5 | 9 | 無 |
| twitter-prfb-noURL-5d10t | Twitter-PRFB | 5 | 10 | 無 |

5. 実験結果

この章では、Twitter のタイムライン閲覧時のツイートを疑似レlevance・フィードバック用の文書を用いて行った実験結果を示す。なお、統計解析には、R version 3.1.0 (Apple Macintosh Mavericks 用)^{*8*9} を使用した。

5.1 実験システムの比較

TREC テストコレクション用いて実験を行った結果は以下の通りである。

5.1.1 Twitter 文書を用いた疑似レlevance・フィードバックの状況

以下に Twitter 文書「データセット2」から疑似レlevance・フィードバックに用いた、上位5文書までの文書数をクエリ番号毎に示す [表2 疑似レlevance・フィードバックしたツイート文書数 (単位: 件)]。表中の灰色の部分は、疑似レlevance・フィードバックするための文書を

^{*7} ※ RunID が indri-prf は、ClueWeb12-PRFB のシステムと同じである。

^{*8} The R Project for Statistical Computing, <http://www.r-project.org/> (accessed 2014-10-20.)

^{*9} 2014年10月20日現在の最新バージョンは3.1.2である。

表 2 疑似レlevance・フィードバックしたツイート文書数 (単位: 件)

| | | | | | | | | | | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| クエリ番号 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 |
| URL有り | 0 | 4 | 5 | 5 | 3 | 5 | 3 | 1 | 5 | 2 | 5 | 5 | 3 | 5 | 5 |
| URL無し | 0 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| クエリ番号 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 | 226 | 227 | 228 | 229 | 230 |
| URL有り | 4 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 1 | 5 |
| URL無し | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| クエリ番号 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 |
| URL有り | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 3 | 2 | 5 | 5 |
| URL無し | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 1 | 4 | 5 | 5 | 5 |
| クエリ番号 | 246 | 247 | 248 | 249 | 250 | | | | | | | | | | |
| URL有り | 5 | 5 | 5 | 0 | 5 | | | | | | | | | | |
| URL無し | 5 | 5 | 5 | 0 | 5 | | | | | | | | | | |

5件取得できなかった部分である。

この表にもあるとおり、クエリ番号 201 と 249 は Twitter 文書に対する検索結果がなく、疑似レlevance・フィードバックに使える文書がなかった。疑似レlevance・フィードバック用に指定された文書がない場合、Indri は ClueWeb12 の文書を使って疑似レlevance・フィードバックを行う。そのため、この2つのクエリに関しては、5文書10語でクエリ拡張を行った場合に、実験システム2 (ClueWeb12-PRFB) と同じシステムとなる。この表の「URL有り」について、取得できた文書数の割合を文書数別にヒストグラムにしたものを示す。[図 3 疑似レlevance・フィードバックによってフィードバックされた文書数 (割合) (URL有り)]

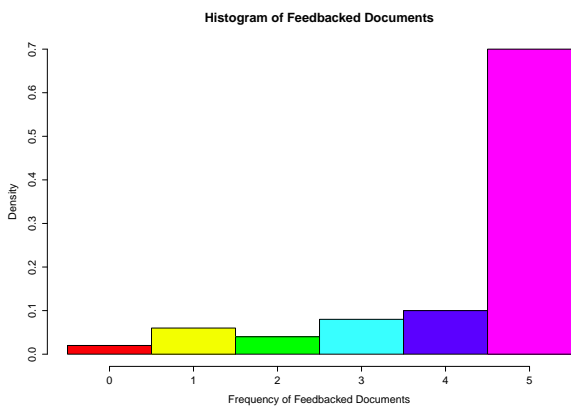


図 3 疑似レlevance・フィードバックによってフィードバックされた文書数 (割合) (URL有り)

また、「URL無し」について同様にヒストグラムを示す。[図 4 疑似レlevance・フィードバックによってフィードバックされた文書数 (割合) (URL無し)]

5.1.2 TREC 評価ツールによる検索結果の評価

この節では、TREC 評価ツールによる検索結果の評価について述べる。検索結果の評価のために、TREC Web Track 2013 の AdHoc Task において使用された評価ツール ndeval と、qrel (正解文書セット) を利用した。たとえ

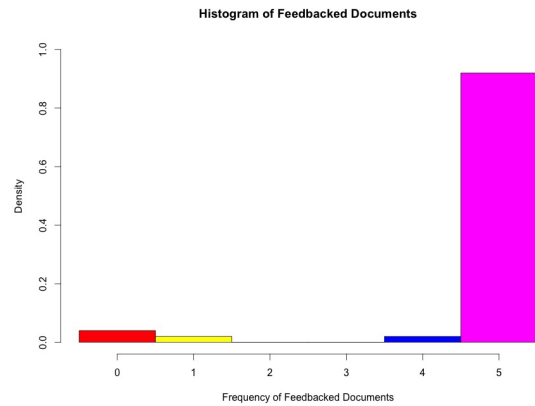


図 4 疑似レlevance・フィードバックによってフィードバックされた文書数 (割合) (URL無し)

ば次のようなコマンドを投入して、検索結果より評価指標の一覧を得る。

```
$ ./ndeval -c -traditional qrels.txt trec-format-run-to-evaluate.txt >normal-nd-evaluation
```

今回の実験では、3つのシステムについて t 検定 (95%信頼区間) を行った。2つのシステム毎に対応のあるデータ (各クエリによる検索結果から算出した評価指標) を3つのシステムについてのすべての組み合わせで比較するため、ボンフェローニ補正を適用した ($p < 0.0167$)。次に各評価指標の t 検定の結果を示す。それぞれの帰無仮説、対立仮説は、

帰無仮説「システムごとの評価指標の平均の差は 0 である」

対立仮説「システムごとの評価指標の平均の差は 0 ではない」

である。[表 3 TREC 評価ツールによる検索結果の t 検定結果]

今回の実験の t 検定に用いた評価指標は次のとおりである。

- ERR-IA@5
- nERR-IA@5
- α -DCG@5
- α -nDCG@5
- MAP-IA
- P-IA@5

TREC では@k の部分、@5、@10、@20 についても評価指標の値を計算するが、今回は @5 のみを対象とすることにした。これらの指標について結果をまとめたものを [表 3] に示す。なお、t の値がマイナスを示しているものは比較時にペアとした実験システムの左側の値が小さいことを示す。

得られた結果について、MAP-IA@5 で比較した場合を箱ひげ図 [図 5 MAP-IA@5 で比較した場合] で表す。

表 3 TREC 評価ツールによる検索結果の t 検定結果

| 評価指標 | Paired t-test | t | df | Mean of the difference | p-value | Significance (p < 0.0167) |
|-----------|---------------------------------|---------|----|------------------------|------------------|---------------------------|
| ERR-IA@5 | ClueWeb12-PRFB vs. baseline | -2.7088 | 49 | -0.07704378 | 0.009276 | 有 |
| | Twitter-PRFB vs. baseline | -1.2757 | 49 | -0.05281326 | 0.2081 | 無 |
| | Twitter-PRFB vs. ClueWeb12-PRFB | 0.5992 | 49 | 0.02423052 | 0.5518 | 無 |
| nERR-IA@5 | ClueWeb12-PRFB vs. baseline | -2.8804 | 49 | -0.08657212 | 0.005876 | 有 |
| | Twitter-PRFB vs. baseline | -1.3909 | 49 | -0.05864096 | 0.1706 | 無 |
| | Twitter-PRFB vs. ClueWeb12-PRFB | 0.6847 | 49 | 0.02793116 | 0.4968 | 無 |
| α-DCG@5 | ClueWeb12-PRFB vs. baseline | -3.0473 | 49 | -0.07573202 | 0.003714 | 有 |
| | Twitter-PRFB vs. baseline | -1.7015 | 49 | -0.06032626 | 0.09519 | 無 |
| | Twitter-PRFB vs. ClueWeb12-PRFB | 0.4235 | 49 | 0.01540576 | 0.6738 | 無 |
| α-nDCG@5 | ClueWeb12-PRFB vs. baseline | -3.2002 | 49 | -0.08345998 | 0.002411 | 有 |
| | Twitter-PRFB vs. baseline | -1.8022 | 49 | -0.06478952 | 0.07767 | 無 |
| | Twitter-PRFB vs. ClueWeb12-PRFB | 0.5084 | 49 | 0.01867046 | 0.6134 | 無 |
| MAP-IA | ClueWeb12-PRFB vs. baseline | -5.1975 | 49 | -0.0282914 | 3.916e-06 | 有 |
| | Twitter-PRFB vs. baseline | -3.6913 | 49 | -0.02393806 | 0.0005605 | 有 |
| | Twitter-PRFB vs. ClueWeb12-PRFB | 0.6621 | 49 | 0.00435334 | 0.511 | 無 |
| P-IA@5 | ClueWeb12-PRFB vs. baseline | -2.8615 | 49 | -0.0494762 | 0.006184 | 有 |
| | Twitter-PRFB vs. baseline | -1.8787 | 49 | -0.5378574 | 0.06624 | 無 |
| | Twitter-PRFB vs. ClueWeb12-PRFB | -0.1612 | 49 | -0.00430954 | 0.8726 | 無 |

6. 考察

6.1 Twitter 文書を用いた疑似レlevance・フィードバック

6.1.1 疑似レlevance・フィードバックに使用できた文書数

結果より、Twitter のツイートに含まれる URL 文字列を取り除いた場合は、残した場合よりも、指定した文書数を取得できたものが多く存在した。フィードバックできた文書が 5 件未満だったクエリの件数は、URL 文字列を残した場合 14 件、削除した場合は 4 件であった。

Twitter のツイートは送信一回あたり 140 文字という制限がある。そのため、疑似レlevance・フィードバックをするための十分な情報量がなく、クエリを構成する語に対応したクエリ拡張のための語を抽出できなかったことが考えられる。

6.2 検索結果の評価指標の比較

今回の実験では、提案手法である実験システム 2 (Twitter-PRFB) とベースライン・システム、および実験システム 1 (ClueWeb12-PRFB) の比較において、t 検定による有意差は認められなかった。ベースライン・システムと実験システム 1 の間には有意差が認められたが、MAP-IA 以外はベースライン・システムと実験システム 2 の比較においては大きな差がなかった。検定で有意差が認められなかった理由について以下の仮説を考える。

- ベースラインと提案手法は同じ検索エンジンを用いており、本質的に差がない。
- ベースラインと提案手法は、差がないわけではないが、提案手法はクエリによって得意・不得意があり、値がばらついて全体として差を示せなかった。
- 今回の実験方法が適切ではなく、本来得るべき正しい値が出せていなかった。

今回、Twitter の文書をユーザのコンテキストと見なして疑似レlevance・フィードバックすることを提案手法とした。しかし、ClueWeb12 と Twitter に含まれる語の数と分布は明らかに違う。現段階でははっきりしたことは言えないが、恐らくインデックスされた語の出現頻度や分布が疑似レlevanceフィードバックやクエリ拡張のプロセスに大きく影響していると考えられる。

6.3 疑似レlevance・フィードバック

Twitter のツイートには略語や短縮 URL などの多くのノイズが含まれる。今回はそれを極力さけるため、ニュースサイトや官公庁のアカウントなどツイート文の表現上のノイズが少ないと考えられるユーザを積極的にフォローした。

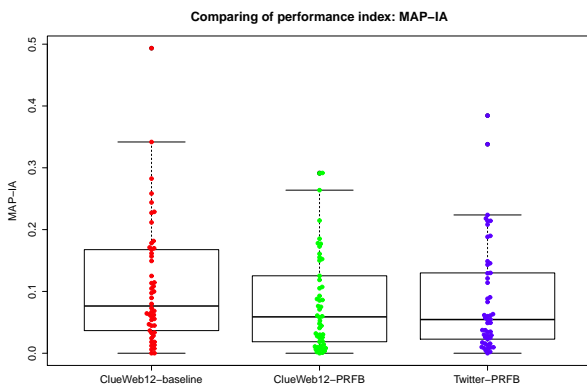


図 5 MAP-IA@5 で比較した場合

5.2 疑似レlevance・フィードバックの検証結果

Twitter 文書を用いてどのような疑似レlevance・フィードバックが行われたか検証するため、[表 1] のパターンを用いて検索実験を行った。疑似レlevance・フィードバックの比較は、nERR-IA@5 を用いて行った。なお、nERR-IA@5 の値がすべて 0 であったクエリ番号は、203, 215, 219, 224, 241, 247, 249 の計 7 件であった。

疑似レlevance・フィードバックによってどのようなクエリ拡張がなされたかは解析中であるが、現段階で展開されたクエリを見てわかることは、Twitter の短縮 URL などのノイズと見なせる語が多く展開されている、ということである。

こうしたサイトは URL をツイートに含み、そこから自分のサイトへの誘導をはかる。しかしその結果、メッセージには Twitter システムに文字数節約のために変換された短縮 URL が含まれてしまう。短縮されない URL の場合はドメイン名にクエリに関係するキーワードを含む場合も考えられるので、URL 自体がノイズであるとは言い切れないが、Twitter の短縮 URL は t.co のドメイン名で始まりランダムな文字列が識別子として展開される。短縮 URL のままではそれ自体に意味はなくノイズであると考えられる。これを本来の意味のある情報にするためには、短縮 URL に紐付けられた元の URL にアクセスし、そのページをもレバンスな文書としてフィードバックする方法が考えられる。

また、提案手法において、疑似レバンスフィードバックを行うときに Indri のパラメータファイルに Twitter 文書の検索結果上位 5 件の文書を記載したが、このことが Indri のクエリ拡張プロセスに大きく影響し、t.co などのフラグメンテーション情報が展開されたと考えられる。原因を特定するため、Twitter 文書からのクエリ拡張を別に行う等、他の方法を用いて引き続き検証を行いたい。

7. おわりに

本報告では、ユーザの SNS 利用時のふるまいを利用したワークスペース検索システムの構築に向け行った、Web 検索時におけるマイクロブログの素性を用いて疑似レバンス・フィードバックの実験を行った。

Twitter のツイートを疑似レバンス・フィードバックの文書データとして利用することは、ノイズの影響による検索結果の不安定さを招くこともあるが、フォローするユーザのツイート次第では、疑似レバンス・フィードバックの素性として有効な場合があることを実験によって示した。一方、提案手法が安定して良好な検索結果を得られるというわけではなく、疑似レバンス・フィードバックに用いるツイート由来のノイズやクエリ拡張の失敗、不適切な重み付け等の原因によって、不安定さが顕著になる場合もあることが明らかになった。

本研究では、個人が直面する検索体験の改善を目標としている。人間が自分自身で積極的に検索をよくするためにできることをシステムに取り入れ、ユーザによるアノテーションなども視野に入れた機能の実装を今後も行っていく。今回は閲覧したツイートを保存する機能だけを使ったが、開発した Twitter ブラウザでは自分の操作履歴を保存しておくことができるので、今後はアノテーションなどでもできるようにしたい。

謝辞 本稿を執筆するにあたり石川哲朗氏（東京工業大学大学院博士課程所属）に統計検定手法について様々なアドバイスをいただいたことに深謝する。

参考文献

- [1] Allan, J., Croft, B., Moffat, A. and Sanderson, M.: Frontiers, challenges, and opportunities for information retrieval, *ACM SIGIR Forum*, Vol. 46, No. 1, pp. 2–32 (online), DOI: 10.1145/2215676.2215678 (2012).
- [2] Belkin, N. J., Clarke, C. L., Gao, N., Kamps, J. and Karlgren, J.: Report on the SIGIR workshop on "entertainment", *ACM SIGIR Forum*, Vol. 45, No. 2, p. 51 (online), DOI: 10.1145/2093346.2093354 (2012).
- [3] Dean-hall, A., Clarke, C. L. A., Thomas, P., Voorhees, E. and Kamps, J.: Overview of the TREC 2012 Contextual Suggestion Track (2012).
- [4] Dumais, S. T.: Putting context into search and search into context, *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '12*, ACM Press, p. 1021 (online), DOI: 10.1145/2348283.2348447 (2012).
- [5] Ingwersen, P. and Järvelin, K.: *The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)*, Springer-Verlag New York, Inc. (2005).
- [6] Ingwersen, P., Järvelin, K., 公男細野, 信之緑川, 和明岸田: 情報検索の認知的転回: 情報検索と情報検索の統合, 丸善 (2008).
- [7] Miyanishi, T., Seki, K. and Uehara, K.: Improving pseudo-relevance feedback via tweet selection, *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management - CIKM '13*, ACM Press, pp. 439–448 (online), DOI: 10.1145/2505515.2505701 (2013).
- [8] 土岐真里奈, 牛尼剛聡: ソーシャルストリーム閲覧時の振舞いを利用したユーザプロフィール構成手法, 情報処理学会論文誌. データベース, Vol. 6, No. 4, pp. 35–45 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/110009610561/>) (2013).