特集論文　「Web インテリジェンスとインタラクションの新展開」

ラベル伝搬によるトレンドクエリのカテゴリ推定

Trend Query Classification using Label Propagation

吉田 光男
Mitsuo Yoshida
豊橋技術科学大学 情報・知能工学系
Department of Computer Science and Engineering, Toyohashi University of Technology
yoshida@cs.tut.ac.jp, http://www.cs.tut.ac.jp/~yoshida/

荒瀬 由紀 *1
Yuki Arase
マイクロソフトリサーチアジア
Microsoft Research Asia
arase@ist.osaka-u.ac.jp, http://www-bigdata.ist.osaka-u.ac.jp/arase_jp.html

keywords: query classification, graph construction

Summary

Query classification is an important technique for web search engines, allowing them to improve users' search experience. Specifically, query classification methods classify queries according to topical categories, such as celebrities and sports. Such category information is effective in improving web search results, online advertisements, and so on. Unlike previous studies, our research focuses on trend queries that have suddenly become popular and are extensively searched. Our aim is to classify such trend queries in a timely manner, i.e., classify the queries on the same day when they become popular, in order to provide a better search experience. To reduce the expensive manual annotation costs to train supervised learning methods, we focus on a label propagation method that belongs to the semi-supervised learning family. Specifically, the proposed method is based on our previous method that constructs a graph using a corpus, and propagates a small number of ground-truth categories of labeled queries in order to estimate the categories of unlabeled queries. We extend this method to cut ineffective edges to improve both classification accuracy and computational efficiency. Furthermore, we investigate in detail the effects of different corpora, i.e., web/blog/news search results, Tweets, and news pages, on the trend query classification task. Our experiments replicate the situation of an emerging trend query; the results show that web search results are the most effective for trend query classification, achieving a 50.1% F-score, which significantly outperforms the state-of-the-art method by 7.2 points. These results provide useful insights into selecting an appropriate dataset for query classification from the various types of data available.

1. は じ め に

入力された検索クエリに対して検索エンジンが関連度の高い付加情報*2を提示するために，検索クエリがどのようなカテゴリに属するのかを推定する研究が広く行われている [Carpineto 12, Jiang 13]．このようなクエリのカテゴリ推定に関する既存研究では，検索エンジンからランダムにサンプルされたクエリが使用されており，検索クエリの頻度変化には着目していなかったため，集中的に検索されるクエリ（トレンドクエリ）に注意が払われていなかった．集中的に検索されるクエリとは，図 1 に示すような検索頻度が一定期間に急上昇するクエリであり*3，突発的なイベントや新製品に関するクエリがしばしば該当する．ウェブ検索エンジンに入力される検索クエリの頻度はそのクエリ（キーワード）に対する人々の注目度をあらわしており，トレンドクエリは，まさに今，世間で注目されていることを示している．そのため，このようなクエリが属するカテゴリをタイムリに推定することは重要である．

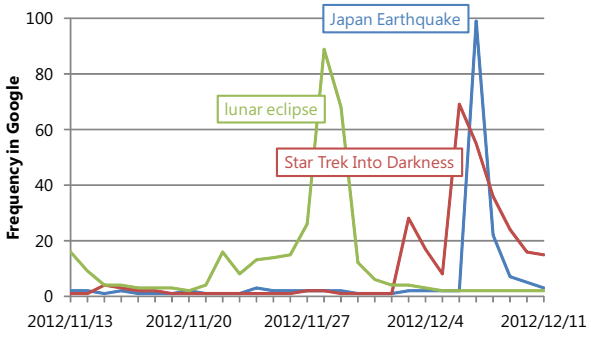

図 1　集中的に検索されるクエリ（トレンドクエリ）の例

ユーザが自由に入力することのできる検索クエリは多種多様であるため，カテゴリを推定するための辞書を作成し，ルールによる推定を行うのは困難であり，機械学習

*1　現在は，大阪大学大学院情報科学研究科に勤務．
*2　画像検索や商品検索などの専門検索のサマリや広告など．
*3　図 1 は Google トレンド（http://www.google.com/trends/）による調査．縦軸の値は，最大値を 100 とする相対値である．

が用いられる．しかし，このような検索クエリをカバーできるような学習データを集めるのは困難で，少量の学習データで推定できる手法が求められる．

本研究では，トレンドクエリの時事カテゴリをタイムリに推定する問題に取り組む．少量の学習データしか準備できない状況に対応するため，半教師あり学習手法を用いる．また，マスメディアに加え，ブログやソーシャルメディアなど，トレンドを即時反映する言語リソースが利用可能となってきているが，既存研究では学習データによる効果の系統的な検証は行われていなかった．そこで本論文では，トレンドクエリの分類に適した学習データを明らかにするために，検索エンジンの結果ページ，ソーシャルメディアでの言及，マスメディアによる報道など，様々な言語リソースを検討する．検索エンジンについては，通常のウェブ検索に加え，ニュース検索，ソーシャルメディアの 1 つであるブログ検索をそれぞれ利用する．集中的に検索された当日にカテゴリ推定を行うという，実環境に合わせた評価実験により，教師あり学習手法よりも，半教師あり学習手法の方が高精度に推定できることを示す．さらに，半教師あり学習手法に用いるグラフのエッジ数を制限することで，推定精度が向上することを示す．

2. 関 連 研 究

検索されるクエリは，恒常的に検索されるクエリ，周期的に検索されるクエリ，集中的に検索されるクエリに大別できる [Kulkarni 11]．恒常的に検索されるクエリとは，“Amazon” や “Twitter” のような著名なウェブサイト名など，日常的によく検索されるクエリである．周期的に検索されるクエリとは，“初詣” や “オリンピック” のような周期的に発生するイベント名など，周期を伴って検索されるクエリである．毎週放送されるようなアニメのタイトルも周期的に検索される．集中的に検索されるクエリとは，新製品の名称など，突発的な要因で急に検索頻度が高まるクエリである．これまでに行われてきた検索クエリのカテゴリ推定では，先の大別に着目せず，実質的に恒常的に検索されるクエリを対象とする場合が多い．例えば，検索クエリのカテゴリ推定タスクである KDD Cup 2005 では，MSN (Bing) の検索ログの中からランダムに選択されたクエリを対象としている [Li 05]．周期的に検索されるクエリを対象としたカテゴリ推定には，Jones らや Zhang らが取り組んでいるが [Jones 07, Zhang 10]，集中的に検索されるクエリを対象とした推定は Yoshida らの手法のみに限られる [Yoshida 12]．

検索クエリが属するカテゴリの推定では，そのタスクに応じてカテゴリがあらかじめ定められる．KDD Cup 2005 は，芸能人やサッカーなど 67 の時事カテゴリのいずれに属するかを推定するタスクであった．KDD Cup 2005 で利用された検索クエリおよびカテゴリの情報は公開されており[*4]，それらを利用した研究も多く見られる [Cao 09, Diemert 09, Khoury 11, AlemZadeh 12]．Broder らによれば，検索の意図は Navigational，Informational，Transactional に分けることができるとされ [Broder 02]，これらの 3 カテゴリに分類する問題に取り組んでいる研究もある [Yoon 10, Ji 11]．検索の意図を拡張し，商業的な情報に適したクエリか否かを推定したり [Pitler 09]，新鮮なウェブページを提示するのが望まれるクエリか否かを推定したりする研究もある [Cheng 13]．

推定手法そのものに焦点を当てれば，教師あり機械学習手法，半教師あり学習手法に大別できる．Baeza-Yates らは検索クエリを訪問先のウェブページで拡張し，そのウェブページのテキストを素性として教師あり学習手法を適用する手法を提案している [Baeza-Yates 06]．検索クエリは高々数単語のテキストであるものの，ウェブページのテキストを追加することで，単語を素性とする教師あり学習を適用しやすくする．この手法は，検索クエリと訪問先にページが密接に関係しているという仮説に基づいている．一方で，ユーザの行動ログを利用せず，検索結果の上位には，検索クエリと関係するウェブページが提示されるという経験則をもとに，検索結果のスニペット（サマリ）を利用する手法が提案されている．このような手法は，[Shen 06, Broder 07, Diemert 09] で提案されており，検索クエリに検索スニペットのテキストを追加し，教師あり学習手法を適用する．

Li らは検索クエリと訪問先のウェブサイトとで 2 部グラフを構築し，そのグラフに半教師あり学習手法を適用することで検索クエリを分類する手法を提案している [Li 08]．ある検索者が，任意の検索クエリの結果に含まれるサイトを訪問したか否かでグラフのエッジを生成しており，同じウェブサイトを訪問されるのに利用された検索クエリは，同じカテゴリに属すという仮説に基づいている．Hu らは Wikipedia 内に存在するグラフ構造に着目し，検索クエリを分類する手法を提案している [Hu 09]．Wikipedia のデータには，見出し語（ページ）間のリンク情報，見出し語とカテゴリのリンク情報，カテゴリ間のリンク情報などが含まれており，見出し語，カテゴリをノードとするグラフを構築することができる．見出し語を検索クエリと見なすことで，半教師あり学習手法を適用することができる．このような手法は，[Khoury 11, AlemZadeh 12] でも提案されている．Yoshida らはトレンドクエリを分類するために，即時性の高い外部リソースとしてソーシャルメディア（Twitter[*5]）の投稿に着目している [Yoshida 12]．カテゴリが既知な検索クエリを頻繁に言及する傾向のあるユーザはそのカテゴリの情報をフォローする傾向がある，また，カテゴリが既知な検索クエリと頻繁に共起するハッシュタグはそのカテ

*4 http://www.sigkdd.org/kdd-cup-2005-internet-user-search-query-categorization
*5 https://twitter.com/

ゴリの他の検索クエリとも共起しやすい，という関係をグラフで表現し，半教師あり学習手法を適用して分類している．

本研究では，Yoshida らが取り組んだ問題 [Yoshida 12] と同様に，トレンドクエリの時事カテゴリをタイムリに推定する問題に取り組む．Yoshida らによる外部リソースの検証は Twitter およびニュース記事にとどまっていたが，本論文では，これまでの研究で頻繁に利用されている検索結果のスニペットを外部リソースとして利用する．また，Yoshida らと同様にグラフベースの半教師あり学習手法を用いるが，グラフのエッジ数を制限することで，推定精度が大きく向上することを実データを用いた実験を通じて示す．

3. 提 案 手 法

3·1 問 題 定 義

本研究は，n個の検索クエリ集合 $Q=\{q_1,\ldots,q_n\}$ が与えられたとき，それぞれの検索クエリ q_i に，あらかじめ定めた時事カテゴリ集合 $C=\{c_1,\ldots,c_m\}$ のいずれが当てはまるかを推定する問題に取り組む．ある検索クエリに複数のカテゴリが当てはまる事を考慮し，マルチラベル分類問題として取り扱う．以下，時事カテゴリをラベルと呼ぶ．

ウェブ検索エンジンで検索される検索クエリの種類は膨大であり，全ての検索クエリに手動でラベルを割り当てる事は困難である．今回，存在する検索クエリの数に対し，正解ラベルの数が少ない事が予想されるため，半教師あり学習手法である Yoshida らの手法 [Yoshida 12] を用い，グラフのエッジ選択により精度向上を図る．

3·2 グラフの構築

本研究では，トレンドクエリのカテゴリ推定に適したグラフ $G=\{V,E,W\}$ を構築し，グラフベースの半教師あり学習手法を適用する．ここで V は n 個のクエリノードで構成されるものとする．E はノード間のエッジの有無（1 または 0）を表し，W はエッジの重み（スコア）をあらわす行列（$n\times n$）である．もし，ノード v_i から v_j の方向にエッジがないならば $E_{ij}=0$ および $W_{ij}=0$ となる．

検索クエリ q_i から q_j への重み，つまりクエリノード v_i から v_j への重み W_{ij} を，それぞれの検索クエリをもとに生成した特徴ベクトル間の類似度 $sim(q_i,q_j)$ をもとに算出する．本研究では，この特徴ベクトルの生成に，検索エンジンの結果ページ，Twitter，ニュースのデータを用いる．次節では，検索エンジンの結果ページを用いる場合を例にして説明する．

3·3 エッジ の 生 成

§1 特徴ベクトルの生成

検索クエリ q_i を含む検索結果ページ集合 P_i を準備し，検索結果ページ集合群 $P_{all}=\{P_1,\ldots,P_n\}$ を構築する．

P_{all} に含まれるページから単語を抽出し，その頻度順に並べた単語集合 $T_{all}=\{t_1,\ldots,t_m\}$ を生成する（m は抽出された単語のタイプ数）．T_{all} には高頻度の単語から低頻度の単語まで様々な単語が含まれており，この単語集合をもとに特徴ベクトルを生成すると，低頻度の単語については素性が発火せず，類似度計算に寄与しない．そのため，使用する単語数に上限 r を設け，使用する単語集合 $T_{use}=\{t_1,\ldots,t_r\}$ を生成する（$r\leq m, T_{use}\subseteq T_{all}$）．

q_i の特徴ベクトル $\vec{b_i}$ は，素性をT_{use}の各単語，素性値を各単語のP_iにおける出現回数とするベクトルとする．表 1 に，各データを使用した場合のP_iとして収集する対象を示す．Twitter のデータを利用する場合は，検索クエリ q_i を含むツイート集合を P_i とし，ニュースのデータを利用する場合は，検索クエリ q_i を含む記事集合を P_i とする．Twitter の場合，言及したユーザを利用してベクトルを生成することもできる．この場合，検索クエリ q_i を含むツイートを行ったユーザ集合を P_i，それらのユーザ ID 集合をT_{all}と見なすことで計算できる．

§2 重みの計算

検索クエリ q_i と q_j との類似度 $sim(q_i,q_j)$ は，次の式のように計算されるコサイン類似度とする[*6]．

$$sim(q_i,q_j)=\frac{\vec{b_i}\cdot\vec{b_j}}{|\vec{b_i}||\vec{b_j}|}$$

グラフベースの半教師あり学習手法を用いる場合，グラフのエッジ選択を行うことで，性能が向上することが知られている [Zhu 09]．Yoshida らの手法 [Yoshida 12] ではエッジ選択は類似度に対する閾値によって行われていたが，エッジの重みの分布は検索クエリごとに大きく異なるため，閾値の設定に性能が敏感になるという問題があった．そこで本研究では，ノード v_i を起点とするエッジ数を k 以下にするという制約を設ける．各ノード v_j（$j\neq i$）を走査して v_i との類似度の大きい上位 k 件を選択し，それらを $E_{ij}=1$ に，選択されなかったものを $E_{ij}=0$ とする．この際，クエリペア q_i，q_j 間のエッジが起点によって $E_{ij}=1, E_{ji}=0$ のように異なる場合があるため，G は有向グラフとなる．

最終的な重みは，上のように k によるエッジ選択を行った後，類似度 $sim(q_i,q_j)$ を正規化する以下の式で計算される．

$$W_{ij}=\frac{sim(q_i,q_j)}{\sum_{q_l\in\{q_l|E_{il}=1\in E\}}sim(q_i,q_l)}$$

パラメータ k の影響は，5·5 節において調査する．

*6 コサイン類似度のほかに，ジャッカード係数，ダイス係数，シンプソン係数なども試したが，推定精度はコサイン類似度を利用したときが最も高かった．

表 1 P_i の生成方法

データ	収集対象
検索エンジン	q_i を含む検索結果ページ集合
Twitter（単語）	q_i を含むツイート集合
Twitter（ユーザ）	q_i を含むツイートをポストしたユーザ集合
ニュース	q_i を含むニュース記事集合

なお，先に述べた Yoshida らの手法は半教師あり学習手法に Twitter のデータを適用する手法である．提案手法においても Twitter のデータを利用するが，パラメータ r および k の導入，ユーザノードの省略において Yoshida らの手法と異なる．

3·4 ラベル伝搬アルゴリズム

構築した有向グラフ G 上でラベル伝搬アルゴリズムにより，クエリのカテゴリを推定する．グラフベースの半教師あり学習の枠組みでは，あらかじめ n_0 個のクエリノード集合 V_0 に，既定のラベル集合 C に含まれるラベルが割り当てられているものとする．残りの n_1 個のクエリノード集合 V_1 にはラベルが割り当てられておらず（$n = n_0 + n_1$），V_0 から V_1 にラベル情報を伝搬させることで，ラベルの推定を行う．

このようなグラフを利用した半教師あり学習手法は，ラベル伝搬（Label Propagation）アルゴリズムとして提案されている [Zhu 03, Zhou 04]．今回，ラベル伝搬アルゴリズムとして Talukdar らによって提案された手法 [Talukdar 08, Talukdar 09, Talukdar 10] を利用する．この手法では，入力として無向グラフを受け付けるため，G を以下のようにして，無向グラフに変換する．

$$W'_{ij} = \frac{W_{ij} + W_{ji}}{2}$$

4. 実 験 設 定

4·1 実験対象クエリ

評価実験に使用する検索クエリは，Google トレンド[*7]，Yahoo!急上昇ワード[*8]，MSN JAPAN 気になる言葉[*9]を毎時に巡回し，2012 年 6 月 1 日から 7 月 5 日までの間に出現したキーワードである．この期間に，2,510 キーワードを収集できた．日々の収集キーワード数を図 2 に示す．Yahoo!急上昇ワードは平日にのみ提供されており，休日は収集数が 0 になる（グラフでは空白）．また，Google トレンドからの収集キーワードが最も多かった．

評価実験のために収集した検索クエリ（キーワード）は，検索ログ等における頻度変化の情報をもとに収集しておらず，集中的に検索されている時期に収集できているとは限らない．そこで，Google トレンドの検索頻度データと照合することにより，適切な時期に収集できているか確認する．収集した 2,510 キーワードのうち，392 キーワードは収集日前後 7 日（計 15 日間）の検索頻度データを取得することができた[*10]．検索頻度が最大であった日の分布を図 3 に示す．取得できたキーワードの 57.1%は当日以降 8 日間の中に検索頻度が最大になる日が存在した．

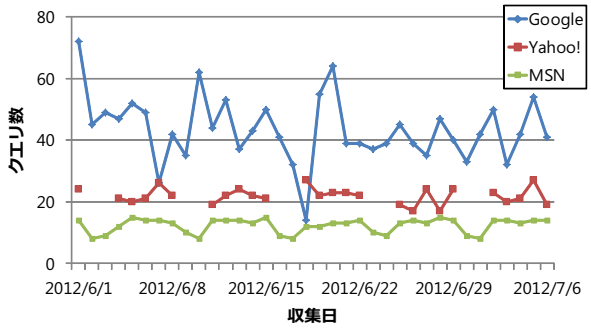

図 2 日別の収集キーワード数

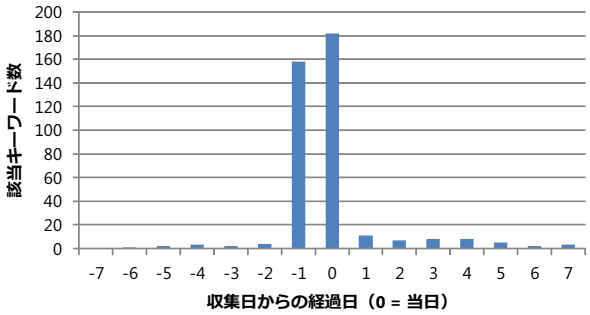

図 3 検索頻度が最大になる日の分布

収集した検索クエリの中からランダムに 1,114 キーワードを選択し，情報工学を専攻する 20 代の男女 3 名[*11]によるアノテーションを行った．アノテートするラベル（時事カテゴリ）は KDD Cup 2005 で使用されたラベルを日本語訳したものとし，最大 5 ラベルの制約の中で，アノテータはそれぞれ任意の数のラベルを付与した．つまり，クエリごとに 1 から 5 のラベルが正解データとして付与されている．この際，集中的に検索された時点（キーワードを収集した時点）のラベルを付与するように指示した．

*7 http://www.google.co.jp/trends/
*8 http://searchranking.yahoo.co.jp/burst_ranking/
*9 http://jp.msn.com/

*10 2014 年 2 月 4 日に取得した．検索頻度の低いキーワードに関しては，検索頻度データを返さないようである．
*11 著者は含まれない．

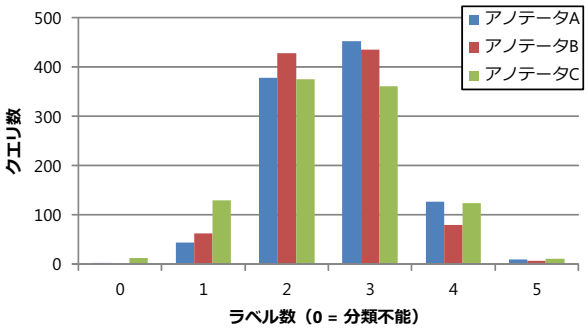

図 4 アノテータによって付与されたラベル数の分布

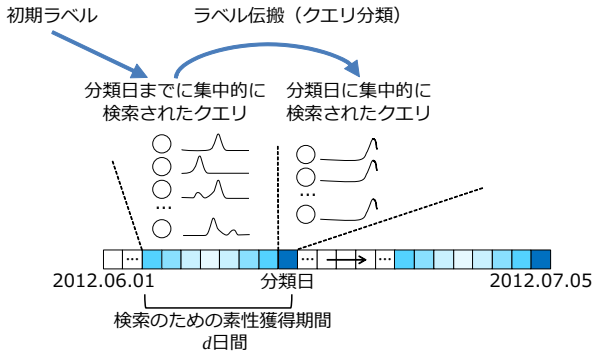

図 5 実験方法

図 4 は各アノテータが付与したラベルの数を示す．横軸の 0 から 5 は付与されたラベル数を表し，縦軸はクエリの数をあらわす．ラベル数が 0 のクエリ数は分類不能であると判断されたクエリの数である．アノテータごとにばらつきがあるものの，1 つのクエリに対して 2 または 3 のラベルが付与されている．

4·2 実 験 手 順

集中的に検索された当日に適切に推定できているかを評価するため，図 5 のように d 日間のウィンドウを動かしながら評価する．ウィンドウの初日から $d-1$ 日目までのクエリに初期ラベルを割り当て，ラベル伝搬アルゴリズムによって d 日目（テスト日）が集中的に検索された日であるクエリのカテゴリを推定する．使用するクエリに合わせ，クエリの特徴ベクトルを生成するためのデータもウィンドウ内に収集したものに限る．この環境により，集中的に検索された当日以降のデータの混入，およびデータの準備を無期限に行うといった非現実的な状況を排除できる．

今回，曜日の影響を排除するため，$d=7$（$=1$ 週間）として評価実験を行う．本環境では 29 のウィンドウで評価を行うことになり，最初のテスト日は 2012 年 6 月 7 日である．最初の 14 ウィンドウをデベロップメント区間としてパラメータの獲得を行う．残りの 15 ウィンドウをテスト区間として性能を評価する．

評価指標は，検索クエリのカテゴリ推定タスクとして著名な KDD Cup 2005 と同様の指標を用いる [Li 05]．まず，各アノテータによって作成された正解データに対する適合率（Precision），再現率（Recall），F 値（F-measure）計算する．そして，それぞれのマクロ平均を計算する．

$$Precision = \frac{\sum_i \text{ラベル } c_i \text{として正解したクエリの数}}{\sum_i \text{ラベル } c_i \text{として推定されたクエリの数}}$$

$$Recall = \frac{\sum_i \text{ラベル } c_i \text{として正解したクエリの数}}{\sum_i \text{ラベル } c_i \text{に属する正解クエリの数}}$$

$$F\text{-}measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

4·3 比較手法および利用データ

提案手法の有効性を示すために，ベースラインとの比較を行う．ベースラインは，KDD Cup 2005 で最高性能を達成した Shen らの手法 [Shen 06] と時期を限定した推定問題に取り組んでいる Yoshida らの手法 [Yoshida 12] とした．Shen らの手法は教師あり学習手法（SVM）に検索エンジンのデータを適用する手法である．さらに検索エンジンから取得したウェブページ中にクエリとカテゴリ名が共起する際には，そのカテゴリをクエリのカテゴリとするシンプルな分類器を用意する．そして適合率または F 値を最大化するように SVM による分類結果と混合する．

検索エンジンのデータは，集中的に検索されるクエリを収集した時点で，ウェブ検索，ブログ検索，ニュース検索を利用して Google，Yahoo! JAPAN，Bing からそれぞれ上位 10 件を取得した．ただし，Bing はブログ検索を提供していないため，ブログ検索には Bing の結果は含まれない．また，ウェブ検索は適合順，ブログ検索およびニュース検索は日付順で取得した．Twitter のデータは Twitter Streaming APIs (Public streams)[*12]を利用して収集し，ニュースデータはニュースポータルサイト Ceek.jp News[*13]から収集した．

各データにおけるカバー率および準備データ量を表 2 に示す．カバー率とは，検索クエリを含むデータをどの程度準備できたかの指標であり，テスト日において，特徴ベクトルを生成するための P_i を準備できた（$|P_i|>0$）検索クエリの割合をあらわす．準備データ量とは，テスト日における P_i の累計ページ数または累計ツイート数をあらわす．P_i を準備できない場合，クエリのカテゴリを推定できなくなるため，カバー率の低いニュース検索，Twitter，ニュースは再現率に上限ができ，推定精度が低下する．

なお，ニュース検索とニュースのカバー率の差は，収集元の異なりもあるが，データ収集方法の異なりによる影響が大きいと考えられる．ニュース検索は集中的に検

*12 https://dev.twitter.com/streaming/public
*13 http://news.ceek.jp/

表 2 実験対象クエリに対するデータのカバー率

	カバー率（%）	準備データ量
ウェブ検索	100.0	24,655
ブログ検索	100.0	15,404
ニュース検索	92.9	14,259
Twitter	93.1	427,761
ニュース	62.4	30,251

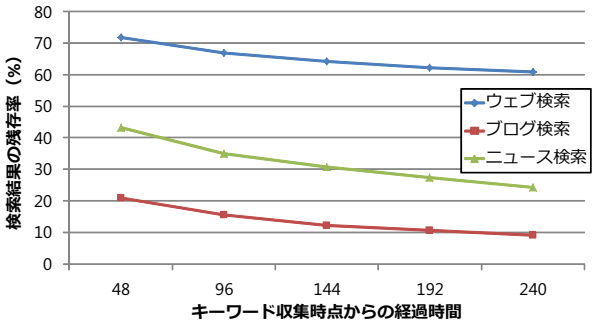

図 6 時間経過に伴う検索結果の残存率

索された時点で検索した結果を利用しているものの，その結果ページに現れるニュースは実験区間ウィンドウ外のニュース（例えば 2011 年のニュースなど非常に古いもの）が含まれる場合があり得る．一方，ニュースの方はウィンドウ内に報道された記事を収集した上で，検索クエリとの照合を行っている．これらの差が，カバー率に影響を与えたものと考えられる．

検索エンジンのデータは検索結果の上位 10 件を取得しており，カバー率や準備データ量の，トレンドクエリ発生時からの時間経過に対する変動は微小である．ここでは，検索結果のページ（URL）の残存率を調べることで，集中的に検索されるクエリに対する検索エンジンの時間的追従性を明らかにする．キーワード収集時点の検索結果ページが，48 時間後，96 時間後，…，240 時間後の検索結果にどれだけ残っていたかを図 6 に示す．ブログ検索，ニュース検索は日付順に取得しているため，残存率が低く，240 時間経過したブログ検索における残存率は 9.2%であった．ウェブ検索の残存率は比較的高いものの，48 時間の経過で 29.3%の結果ページが検索結果上位 10 件から除かれている．

Twitter およびニュースのデータの即時性を明らかにするために，それらのカバー率および準備データ量の時間的推移を調べた．本実験では，収集日（テスト日）を末尾とする 7 日間のウィンドウ内で素性の獲得を行うが，ここでは，そのウィンドウを通常の 7 日前から 7 日後まで移動させ，その間のカバー率および準備データ量を計算する．図 7 は Twitter の，図 8 はニュースのウィンドウを移動させた場合のカバー率および準備データ量の変化である．Twitter およびニュースともに，キーワード収集日当日にカバー率がほぼ最大値に達しており，即時性の

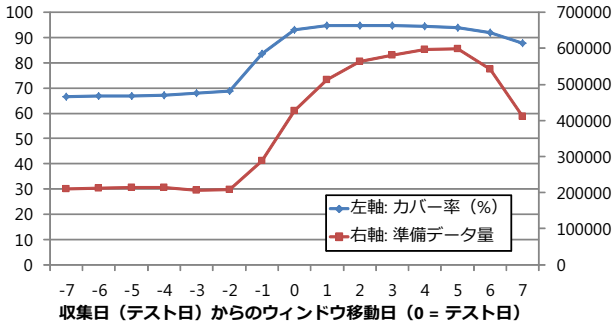

図 7 カバー率および準備データ量の変化（Twitter）

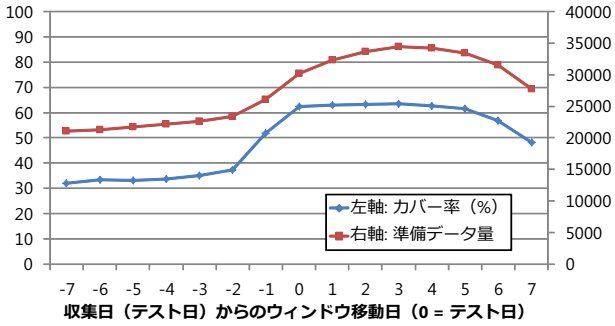

図 8 カバー率および準備データ量の変化（ニュース）

高さが表れている．Twitter に関しては，当日以降も準備データ量が大幅に増加しており，トレンドクエリに関する話題が継続していることが示唆される．

4·4 パラメータの獲得

デベロップメント区間で実験を行い，推定性能（F 値）を最大化するようなパラメータを獲得する．推定性能は分類器が取り出したカテゴリの数に依存するため，分類器が取り出したカテゴリの数が 3 の時の性能を用いてパラメータの獲得を試みた．提案手法に関し，獲得したパラメータを表 3 に示す．Twitter のデータにおいて，類似度の計算に言及ユーザを使う際はユーザの足きりを行わない場合が推定性能最大となったため，今回，パラメータ r を未設定とした．

ベースラインにおいては，データ準備に関するパラメータが明示されていなかったものの，パラメータ r を獲得することで性能向上が認められたため，本実験で比較する際もパラメータ r を導入する．Shen らの手法で SVM を利用する際，単語の足きりは $r = 9000$，カーネルは線形カーネル（$C = 0.035$）を使用したときに推定性能が最大化した．また，Yoshida らの手法において，単語の足きりは $r = 8000$，グラフ混合比率は $\alpha = 0.6$ を使用したとき（クエリ間類似度よりもユーザを優先する）に推定性能が最大化したため，テスト区間でそれらの設定を使う．

表 3 デベロップメント区間で獲得したパラメータ（提案手法）

データ（手法）	k	r
ウェブ検索	10	8,000
ブログ検索	10	9,000
ニュース検索	20	2,000
Twitter（単語）	50	9,000
Twitter（ユーザ）	30	-
ニュース	20	3,000

5. 実 験 結 果

5·1 ベースラインとの比較

ベースラインを含めた実験結果を表 4 および表 5 に示す（太字は最高性能を示す．以下同じ.）．横軸の 1 から 5 の数値は，分類器が取り出したカテゴリの数である．アノテータによる平均ラベル数は 2.57 であることから，3 の時に高い性能を示す傾向がある．実験結果より，ウェブ検索を利用した提案手法が最高性能を示すことがわかる．KDD Cup 2005 で最高性能を達成した Shen らの手法と比べても，3 つのカテゴリを推定した場合で 7.2 ポイントも高い F 値を達成している．なお，符合検定により，提案手法（ウェブ検索）とそのほかとの手法の間に，有意水準 1%で有意差を確認した．

Shen らの手法において，SVM による分類器との混合方法について，EDP（適合率最大化），EDF（F 値最大化）の双方を試したところ，EDF の方が良い性能を示した．Yoshida らは，混合したグラフが最高性能を示すと報告していたものの，本環境においてはクエリ間の類似度を使う場合と比較し，有意水準 5%でも有意差を確認できなかった．

なお，アノテータ A のラベル情報を正解データ，アノテータ B のラベルデータを分類器による分類結果と見立てた場合など，アノテータ間の一致を調べたところ，F 値の平均は 63.5%であった．この値は人間がカテゴリ推定したときの期待値であり，提案手法は期待値の 8 割にまで迫っていることがわかる．

5·2 グラフ構造および語彙の差

テスト区間における提案手法に関するグラフ構造の異なりを表 6 に示す．エッジ数は各ノードの持つ平均エッジ数である．なお，この値はパラメータ k に大きく依存する．また，提案手法におけるパラメータ k は有向エッジにおける上限であるため，正規化した後の無向エッジにおける上限はその 2 倍になる．パス長は任意の 2 ノード間の距離の平均であり，クラスタ係数は任意の 2 ノードを選択したときその 2 ノードが隣接している割合である．隣接ラベル率はあるテストノードの隣接ノードが正解ラベルを持つ初期ノードの割合であり，隣接ノード率はあるテストノードに少なくとも 1 つの正解ラベルを持つ初期ノードが存在する割合である．エッジ数，パス長，クラスタ係数，隣接ノード率に関して，それぞれのデータ間で違いがあるものの，推定性能との関係性は確認できない．一方，隣接ラベル率はその大小関係が表 4 に示す分類性能の大小関係と一致しており，隣接ラベル率を高めることは分類性能の向上につながると考えられる．このことは，ラベル伝搬アルゴリズムの特性を考えれば妥当である．ウェブ検索の検索結果ページ集合には多様性があり，それが隣接ラベル率の向上に影響を与えた可能性があるものの，その詳細は明らかにできておらず，今後の課題とする．

グラフの構築にはクエリ間の類似度を利用しており，この類似度はクエリと共起する語をもとに計算される．本研究で利用した外部リソースは，ウェブ検索，ブログ検索，ニュース検索，Twitter，ニュースの 5 種であり，それぞれのグラフ構造に異なる特性があるならば，それぞれの語彙にも異なりが確認できると考えられる．テスト区間における各データで利用される語彙間のジャッカード係数をまとめた結果を表 7 に示す．ウェブ検索ブログ検索の語彙間のジャッカード係数が最も高いものの，0.6 に満たなかった．語彙間のジャッカード係数が低いことから，それぞれのデータで，異なる特徴を持つグラフが生成できていることが示唆される．

5·3 カテゴリ別の性能

提案手法におけるテスト区間でのカテゴリ別の推定性能（F 値）を表 8 に示す．横軸の 2 および 3 の数値は，分類器が取り出したカテゴリの数である．縦軸は，時事カテゴリを表しており，アノテータによる利用頻度の高い上位 5 件を抽出した．表 4 に示した全体の性能評価では，ウェブが最も高性能であったが，カテゴリ別にみると,「テレビ」の最高性能はブログ検索の結果を利用したものであるなど，異なる傾向がある．このことから，検索エンジンごとに得意とする推定カテゴリが異なることが確認できた．

隣接ラベル率を確認すると,「テレビ」においてはウェブ検索が 25.5%であるのに対し，ブログ検索は 31.6%であった．特定のカテゴリに関する隣接ラベル率の差が，推定性能に影響を与えたものと考えられる．さらに,「テレビ」カテゴリに属するキーワード[*14]の量は 312 キーワードであり，そのうち 106 キーワードが Google トレンドの検索頻度データと照合可能であった．つまり，比較的検索頻度の高いキーワード群であることが示唆される．また，前日が検索頻度最大になるキーワードが 54.7%存在し，ブログ検索の結果を日付順に取得していることからも，検索ピーク直後のウェブコンテンツ（ブログ）を使用できたものと推察される．

*14 アノテータのうち少なくとも 1 名が「テレビ」のラベルを割り当てたキーワード.

表 4 ベースラインとの比較（F 値）

		F 値（%）				
		1	2	3	4	5
提案手法	ウェブ検索	**36.1**	**47.9**	**49.1**	**47.0**	**43.6**
	ブログ検索	31.8	43.1	45.5	44.7	41.9
	ニュース検索	30.8	41.2	43.9	42.5	40.2
	Twitter（単語）	27.3	36.9	39.2	37.7	36.1
	Twitter（ユーザ）	21.4	30.2	32.0	31.5	30.5
	ニュース	22.4	32.5	36.3	36.1	34.9
Shen	SVM	27.0	37.4	37.8	36.7	35.0
	EDP	23.7	29.4	34.5	34.7	36.1
	EDF	29.3	39.3	41.9	40.9	40.1
Yoshida	Query	26.7	36.5	38.7	37.8	36.0
	User	23.0	31.5	33.6	33.3	31.8
	QueryTwitter	26.1	36.5	38.2	37.3	36.2

表 5 ベースラインとの比較（適合率，再現率）

		適合率（%）					再現率（%）				
		1	2	3	4	5	1	2	3	4	5
提案手法	ウェブ検索	**64.6**	**54.9**	**45.7**	**38.7**	**33.1**	**25.0**	**42.5**	**53.1**	**59.9**	**64.1**
	ブログ検索	56.9	49.4	42.3	36.8	31.8	22.0	38.3	49.2	57.0	61.6
	ニュース検索	58.6	49.4	42.4	36.1	31.4	20.9	35.3	45.5	51.7	56.1
	Twitter（単語）	52.0	44.4	37.9	32.1	28.1	18.5	31.6	40.6	45.7	50.2
	Twitter（ユーザ）	42.7	37.7	32.0	27.5	24.3	14.3	25.2	32.1	36.8	40.7
	ニュース	57.3	49.8	43.2	36.7	31.9	13.9	24.2	31.4	35.6	38.6
Shen	SVM	48.3	42.9	35.1	30.2	26.6	18.7	33.2	40.9	46.8	51.5
	EDP	42.4	33.6	32.1	28.5	27.4	16.5	26.1	37.4	44.3	53.2
	EDF	52.4	45.0	38.9	33.6	30.4	20.4	35.0	45.4	52.2	59.1
Yoshida	Query	50.9	43.9	37.5	32.1	28.1	18.1	31.3	40.1	45.8	50.1
	User	46.0	39.3	33.5	29.1	25.4	15.4	26.3	33.6	38.9	42.5
	QueryTwitter	49.5	43.8	36.9	31.7	28.3	17.7	31.3	39.6	45.3	50.5

5·4 グラフの混合による性能の変化

提案手法では，検索エンジン，Twitter，ニュースをそれぞれ単独で利用し，グラフを構築している．本節では，グラフの混合によって性能が向上するかを検証する．4·4 節のパラメータを維持した状態で，ウェブ検索のグラフに他のグラフを混合する．混合の際は，グラフの混合比率 β を設定する．β は 0 から 1 までの値をとり，$\beta = 1$ の時，グラフ混合は行われずにウェブ検索の結果のみを利用したことを示す．

表 9 にデベロップメント区間で獲得したパラメータ β とテスト区間における推定性能（F 値）およびグラフ構造（隣接ラベル率，隣接ノード率）を示す．ブログ検索のグラフを混合した場合に最も性能が向上することを確認した．この結果に関し，符号検定により，有意水準 1%で有意差を確認した．一方，Twitter のグラフを混合する場合，単語およびユーザのいずれのグラフを混合した場合でも，性能が低下した．隣接ラベル率を確認すると，いずれも低下しているものの，隣接ノード率が向上している．性能が低下した Twitter は隣接ラベル率が大きく低下しており，性能を向上させるためには隣接ラベル率をある程度の値以上にとどめておく必要があると考えられる．これらのことから，特性の異なるグラフを混合することで推定性能向上の可能性はあるものの，単純な混合のみでは，特性の違いを十分にとらえられず，性能向上が難しいことがわかる．

5·5 許容するエッジ数と性能の関係

本研究では，グラフを構築する際に，パラメータ k を導入することにより，ノードが持ちうるエッジ数に制約を設けている．本節では，テスト区間において，許容するエッジ数を変化させ，性能にどのように影響を与えるかを調べた．図 9 は，5·1 節で示した検索エンジン（ウェブ）の条件のうち，パラメータ k を変動させた実験結果である（分類器が取り出すラベルの数は 3 に固定）．

エッジの削減処理を行わない場合は適合率 32.3%，再現率 37.6%，F 値 34.7%であり，エッジを削減するごと

表 6 テスト区間におけるグラフ構造（平均値）

	エッジ数	パス長	クラスタ係数	隣接ラベル率（%）	隣接ノード率（%）
ウェブ検索	15.4	2.3	0.39	23.8	83.6
ブログ検索	16.2	2.2	0.38	21.0	81.1
ニュース検索	28.2	1.9	0.55	20.5	80.2
Twitter（単語）	69.6	1.6	0.51	15.6	85.6
Twitter（ユーザ）	36.9	1.9	0.49	15.6	68.2
ニュース	28.2	1.8	0.49	20.0	54.3

表 7 各データで利用される語彙間のジャッカード係数

	ウェブ検索	ブログ検索	ニュース検索	Twitter（単語）
ブログ検索	0.598			
ニュース検索	0.241	0.210		
Twitter（単語）	0.447	0.474	0.184	
ニュース	0.295	0.270	0.358	0.247

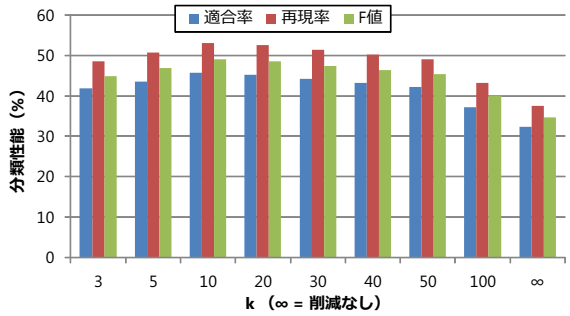

図 9 検索結果（ウェブ）ページを利用した推定性能に対するパラメータ k の影響

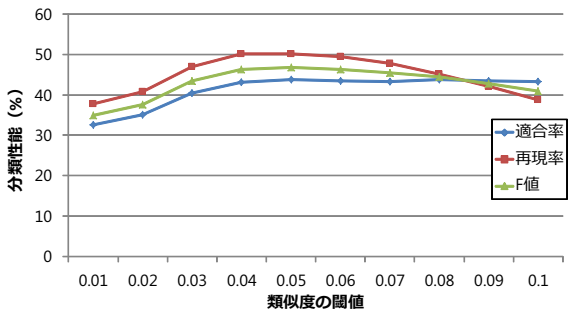

図 10 検索結果（ウェブ）ページを利用した推定性能に対する類似度閾値の影響

に性能が向上し，実際にデベロップメント区間で獲得した $k = 10$ のときに最高性能を示したことから，エッジの削減処理が有用であることが確認できた．このことから，ラベル伝搬アルゴリズムの効果を最大化するためには，ノイズの少ないグラフを構築する必要があることがわかる．

ノイズの少ないグラフを構築する別の方法として，3·3·2 節で述べた検索クエリ間の類似度 $sim(q_i, q_j)$ の値に閾値を設け，ある値未満のエッジを除去する方法が挙げられる [Yoshida 12]．図 10 は，5·1 節で示した検索エンジン（ウェブ）の条件をもとに，類似度の閾値を変動させた実験結果である（分類器が取り出すラベルの数は 3 に固定）．閾値 0.05 の時に性能が最大になるものの*15，性能（F 値）は 46.8%にとどまっており，提案手法を上回ることはなかった．この結果に関し，符号検定により，有意水準 1%で有意差を確認した．このことから，ノイズの少ないグラフを構築する手法としては，提案手法の方が優れていることがわかる．

*15 閾値 0.2 の時の F 値は 20.2%であり，閾値を 0.1 よりも大きくすると，性能は著しく悪化することを確認した．

5·6 使用するページ数と性能の関係

本研究では，検索エンジンのデータを利用する際，標準で取得できる最大量である上位 10 件を利用した．本節では，テスト区間において，使用量を変化させ，性能にどのような影響を与えるか調べた．図 11 は，5·1 節で示した検索エンジン（ウェブ）の条件のうち，使用するページ数のみを変化させた実験結果である（分類器が取り出すラベルの数は 3 に固定）．

上位 1 件を利用した場合は適合率 36.9%，再現率 42.8%，F 値 39.7%であり，上位 7 件の時点で適合率 45.5%，再現率 52.9%，F 値 48.9%と性能の頭打ちが発生しており，必ずしも大量のページを準備する必要のないことが示唆される．

6. おわりに

本論文では，集中的に検索される検索クエリ（トレンドクエリ）の時事カテゴリをタイムリに推定する問題に適した素性を明らかにするために，検索エンジンの結果ページ，ソーシャルメディアでの言及，マスメディアによる報道など，様々な言語リソースを検討した．その結果，検索エンジンの結果ページを特徴量に変換し，グラ

表 8 カテゴリ別の性能評価（F 値）

	2			3		
	ウェブ検索	ブログ検索	ニュース検索	ウェブ検索	ブログ検索	ニュース検索
有名人・著名人・芸能人	**75.2**	70.4	73.6	**73.8**	70.3	72.1
テレビ	55.4	**60.2**	52.2	55.2	**57.7**	52.9
人物検索	**69.7**	56.0	63.1	**70.2**	65.4	65.4
地域・地方	58.2	50.3	**58.4**	**55.8**	50.0	55.0
企業・産業	**54.7**	51.6	40.8	**51.9**	49.7	44.2

表 9 グラフ混合による性能（F 値）およびグラフ構造

		F 値（%）					グラフ構造（%）	
混合対象	β	1	2	3	4	5	隣接ラベル率	隣接ノード率
ブログ検索	0.8	**36.5**	**48.2**	**50.1**	47.8	44.5	21.0	89.7
ニュース検索	0.6	36.1	47.2	49.5	**48.0**	**44.8**	20.4	91.3
Twitter（単語）	0.6	35.7	47.8	48.9	47.2	44.1	16.1	93.6
Twitter（ユーザ）	0.8	35.3	47.1	48.2	46.6	43.5	18.5	89.4
ニュース	0.5	35.8	47.1	49.6	47.9	44.9	20.7	88.4
混合無し	1.0	36.1	47.9	49.1	47.0	43.6	23.8	83.6

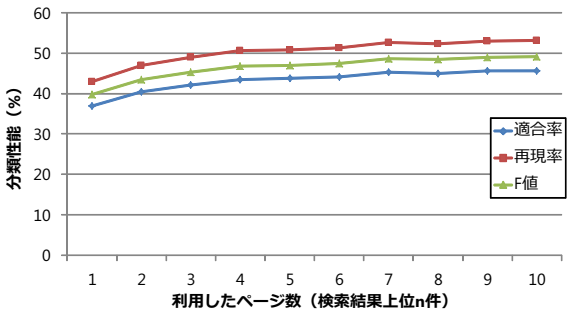

図 11 検索結果（ウェブ）ページを利用した推定性能に対する使用ページの影響

フベースの半教師あり学習手法を適用する手法が最良であることが明らかになった．

半教師あり学習手法のためのグラフ構築に関し，エッジ数の制約が性能向上に大きく寄与していることが確認できた．さらに，使用する検索結果のページ数を変化させた実験により，大規模なデータを必ずしも準備する必要のないことが示唆された．また，言語リソースを混合した場合の性能を検証したところ，ウェブ検索エンジンとブログ検索エンジンの結果ページを混合した場合に推定性能が最大化することが明らかになった．

今後，分類器の構築に関し，オンライン学習などを取り入れた効率的な手法を検討する．また，提案手法を実在の検索エンジンに適用し，ユーザの満足度を向上させるような取り組みを検討していきたい．そして，本研究で作成したデータセットを公開し，本研究課題に広く取り組める基盤を構築していきたいと考えている．

◇ 参 考 文 献 ◇

[AlemZadeh 12] AlemZadeh, M., Khoury, R., and Karray, F.: Query Classification using Wikipedia's Category Graph, *Journal of Emerging Technologies in Web Intelligence*, Vol. 4, No. 3, pp. 207–220 (2012)

[Baeza-Yates 06] Baeza-Yates, R., Calderón-Benavides, L., and González-Caro, C.: The Intention Behind Web Queries, in *String Processing and Information Retrieval*, Vol. 4209, pp. 98–109 (2006)

[Broder 02] Broder, A.: A taxonomy of web search, *SIGIR Forum*, Vol. 36, No. 2, pp. 3–10 (2002)

[Broder 07] Broder, A. Z., Fontoura, M., Gabrilovich, E., Joshi, A., Josifovski, V., and Zhang, T.: Robust Classification of Rare Queries Using Web Knowledge, in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*, pp. 231–238 (2007)

[Cao 09] Cao, H., Hu, D. H., Shen, D., Jiang, D., Sun, J.-T., Chen, E., and Yang, Q.: Context-Aware Query Classification, in *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*, pp. 3–10 (2009)

[Carpineto 12] Carpineto, C. and Romano, G.: A Survey of Automatic Query Expansion in Information Retrieval, *ACM Computing Surveys*, Vol. 44, No. 1, pp. 1:1–1:50 (2012)

[Cheng 13] Cheng, S., Arvanitis, A., and Hristidis, V.: How Fresh Do You Want Your Search Results?, in *Proceedings of the 22nd ACM International Conference on Conference on Information and Knowledge Management*, pp. 1271–1280 (2013)

[Diemert 09] Diemert, E. and Vandelle, G.: Unsupervised Query Categorization using Automatically-Built Concept Graphs, in *Proceedings of the 18th International Conference on World Wide Web*, pp. 461–470 (2009)

[Hu 09] Hu, J., Wang, G., Lochovsky, F., Sun, J.-t., and Chen, Z.: Understanding User's Query Intent with Wikipedia, in *Proceedings of the 18th International Conference on World Wide Web*, pp. 471–480 (2009)

[Ji 11] Ji, M., Yan, J., Gu, S., Han, J., He, X., Zhang, W. V., and Chen, Z.: Learning Search Tasks in Queries and Web Pages via Graph Regularization, in *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*, pp. 55–64 (2011)

[Jiang 13] Jiang, D., Pei, J., and Li, H.: Mining Search and Browse Logs for Web Search: A Survey, *ACM Transactions on Intelligent Systems and Technology*, Vol. 4, No. 4, pp. 57:1–57:37 (2013)

[Jones 07] Jones, R. and Diaz, F.: Temporal Profiles of Queries, *ACM*

Transactions on Information Systems, Vol. 25, No. 3 (2007)
[Khoury 11] Khoury, R.: Query classification using Wikipedia, *International Journal of Intelligent Information and Database Systems*, Vol. 5, No. 2, pp. 143–163 (2011)
[Kulkarni 11] Kulkarni, A., Teevan, J., Svore, K. M., and Dumais, S. T.: Understanding Temporal Query Dynamics, in *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*, pp. 167–176 (2011)
[Li 05] Li, Y., Zheng, Z., and Dai, H. K.: KDD CUP-2005 Report: Facing a Great Challenge, *ACM SIGKDD Explorations Newsletter*, Vol. 7, No. 2, pp. 91–99 (2005)
[Li 08] Li, X., Wang, Y.-Y., and Acero, A.: Learning Query Intent from Regularized Click Graphs, in *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*, pp. 339–346 (2008)
[Pitler 09] Pitler, E. and Church, K.: Using Word-Sense Disambiguation Methods to Classify Web Queries by Intent, in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 1428–1436 (2009)
[Shen 06] Shen, D., Sun, J.-T., Yang, Q., and Chen, Z.: Building Bridges for Web Query Classification, in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*, pp. 131–138 (2006)
[Talukdar 08] Talukdar, P. P., Reisinger, J., Pasca, M., Ravichandran, D., Bhagat, R., and Pereira, F.: Weakly-Supervised Acquisition of Labeled Class Instances using Graph Random Walks, in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 582–590 (2008)
[Talukdar 09] Talukdar, P. and Crammer, K.: New Regularized Algorithms for Transductive Learning, in *Machine Learning and Knowledge Discovery in Databases*, Vol. 5782, pp. 442–457 (2009)
[Talukdar 10] Talukdar, P. P. and Pereira, F.: Experiments in Graph-Based Semi-Supervised Learning Methods for Class-Instance Acquisition, in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1473–1481 (2010)
[Yoon 10] Yoon, S., Jatowt, A., and Tanaka, K.: Intent Feature Discovery using Q&A Corpus and Web Data, in *Proceedings of the 4th International Conference on Uniquitous Information Management and Communication*, pp. 40:1–40:7 (2010)
[Yoshida 12] Yoshida, M. and Arase, Y.: Exploiting Twitter for Spiking Query Classification, in *Information Retrieval Technology*, Vol. 7675, pp. 138–149 (2012)
[Zhang 10] Zhang, R., Konda, Y., Dong, A., Kolari, P., Chang, Y., and Zheng, Z.: Learning Recurrent Event Queries for Web Search, in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 1129–1139 (2010)
[Zhou 04] Zhou, D., Bousquet, O., Lal, T. N., Weston, J., and Schölkopf, B.: Learning with Local and Global Consistency, in *Advances in Neural Information Processing Systems 16* (2004)
[Zhu 03] Zhu, X., Ghahramani, Z., and Lafferty, J.: Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions, in *Proceedings of the 20th International Conference on Machine Learning*, pp. 912–919 (2003)
[Zhu 09] Zhu, X. and Goldberg, A. B.: Introduction to Semi-Supervised Learning, *Synthesis Lectures on Artificial Intelligence and Machine Learning*, Vol. 3, No. 1, pp. 1–130 (2009)

〔担当委員：大向 一輝〕

2014 年 5 月 8 日　受理

著　者　紹　介

吉田　光男（正会員）

2009 年筑波大学第三学群情報学類卒業．2011 年同大学院システム情報工学研究科博士前期課程修了，2014 年同博士後期課程修了．博士（工学）．同年より豊橋技術科学大学大学院工学研究科（情報・知能工学系）助教．ウェブ工学，自然言語処理，情報検索に関する研究に従事．言語処理学会，情報処理学会の各会員．

荒瀬　由紀

2006 年大阪大学工学部電子情報エネルギー工学科卒業．2007 年同大学院情報科学研究科博士前期課程修了，2010 年同博士後期課程修了．博士（情報科学）．同年，Microsoft Research Asia に入社し，Natural Language Computing グループ研究員となる．2014 年より大阪大学大学院情報科学研究科准教授．言い換え表現抽出，統計的機械翻訳，ウェブデータマイニングに関する研究に従事．ACL，言語処理学会，情報処理学会，日本データベース学会の各会員．