

推薦論文

実生活ツイートに対する局面推定の精度向上に関する検討

山本 修平^{1,a)} 佐藤 哲司^{1,b)}

受付日 2014年10月16日, 採録日 2015年2月4日

概要: 身近な出来事や関心事を投稿し共有する Twitter 上には、食事や交通、災害など、様々な生活の局面で有益なツイートが数多く投稿されている。著者らは、未知のツイートに適切な複数の局面を付与できる階層的推定法を提案している。階層的推定法は、教師なし学習として知られる LDA を用いて、大量ツイートからトピックを抽出する第 1 段階と、少量の訓練データを用いてトピックと局面の関連度を算出し、対応関係を構築する第 2 段階からなる。本論文では、対応付くトピックが競合する局面で推定精度が低下する問題を解決するため、新たな関連度の算出方法を提案する。関連度を各トピックで正規化することで、トピックが強く結び付く局面を同定し、局面に対応付くトピックの競合を防ぐ。トピック側で正規化した関連度を、さらに局面側で正規化することで、局面から見て強く結び付くトピックを同定する。収集した大量のツイートをを用いた評価実験を行った結果、これまで推定精度が低かった局面も適切に推定でき、すべての局面における F 値の平均も向上できることを明らかにした。

キーワード: Twitter, 実生活, LDA, 階層的推定法, マルチラベル分類

A Study on Upgrading Precision of Aspect Estimation for Real Life Tweets

SHUHEI YAMAMOTO^{1,a)} TETSUJI SATOH^{1,b)}

Received: October 16, 2014, Accepted: February 4, 2015

Abstract: Recently, lots of users share their current events and opinions by using the Twitter. Hence, some of these tweets are beneficial in several aspects of user's real life, i.e., eating, traffic, disaster, weather, and so on. We proposed hierarchical estimation framework to estimate appropriate several aspects for unknown tweets. Hierarchical estimation framework is composed of both unsupervised and supervised machine learnings. In the first phase, it extracts topics from a sea of tweets using latent Dirichlet allocation (LDA). In the second phase, it calculates the relevance between topics and aspects using a small set of labeled tweets to build associations among them. In this paper, we propose new relevance calculation method to conquer the problem of competitive associations among aspects, which estimating precisions decrease. We identify the aspects associated to topics by normalizing relevance in each topic and prevent association competitive. Moreover, we identify the topics strongly associated to aspects by besides normalizing in each aspect. Our sophisticated experimental evaluations with a large amount of actual tweets demonstrate the high efficiency of our proposed method in aspects that estimating precisions ever lowed. We clarify that our proposed method upgrades an average value of F-measure in all aspects.

Keywords: twitter, real life, LDA, hierarchical estimation framework, multi-label classification

1. はじめに

ツイートと呼ばれる短文記事を投稿する Twitter^{*1}は、最も広く普及しているマイクロブログの 1 つであり、2014

年 9 月末に 2 億 8,400 万人の月間アクティブユーザ数を記録している [13]. Twitter では、ユーザは自らの経験や日常生活でのイベントなど、身近な「今」を投稿しているため、他のユーザにとっても最新かつ有益なツイートが多い。た

¹ 筑波大学大学院図書館情報メディア研究科
Graduate School of Library, Information and Media Studies,
University of Tsukuba, Tsukuba, Ibaraki 305-8550, Japan

a) yamahei@ce.slis.tsukuba.ac.jp

b) satoh@ce.slis.tsukuba.ac.jp

^{*1} <https://twitter.com/>
本論文の内容は 2014 年 7 月のマルチメディア、分散、協調とモバイル (DICOMO2014) シンポジウムにて報告され、グループウェアとネットワークサービス研究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

表 1 実生活の局面
Table 1 Aspects of real life.

局面	典型的な単語
服飾	衣服, 服装, 着る, 装飾, 化粧, 理髪, 衣装 ...
交流	約束, 出会い, 招待, 友人, 誘い, 勧誘, 飲み会 ...
災害	洪水, 竜巻, 地震, 火事, 津波, 二次災害 ...
食事	料理, 外食, 食べ物, レストラン, ジャンクフード ...
行事	祭り, 冠婚葬祭, 日程, 開催日, 学園祭, 文化祭 ...
消費	購入, 買う, 注文, 安売り, 特売, ショッピング ...
健康	風邪, 体調, 怪我, 痛み, 健康法, 病気予防 ...
趣味	余暇, 娯楽, おもちゃ, 音楽, テレビ, ゲーム ...
居住	掃除, 家具, 洗濯, 住まい, 隣人, アパート ...
地域	観光, 地域情報, 地理情報 ...
学校	勉強, 宿題, 課題, 試験, テスト, 資格, 研究 ...
交通	電車, バス, 飛行機, 時刻表, 渋滞, 混雑, 遅延 ...
気象	天気, 気温, 湿度, 風, 花粉, 雨量, 空模様 ...
労働	アルバイト, 研修, 就職活動, 営業, 仕事 ...

例えば電車の遅延情報は交通機関を利用するユーザに役立つ、近所のスーパーマーケットの特売情報は買物に出かけるユーザを支援できる。これらのように地域性が高く新鮮かつ、他のユーザに有益なツイートを、著者らは「実生活ツイート」と呼び、実生活ツイートに対して表 1 に示すような生活の局面を付与することを試みてきている [18], [19].

例えば、電車の遅延に言及したツイートには「交通」の局面が付与され、スーパーマーケットの特売情報が記述されたツイートには「消費」の局面が付与されることが望ましい。ツイートによっては複数の局面を付与するのが適切な場合もある。例えば、「激しい雨のため、電車が遅れています」というツイートは、電車の遅延だけでなく、屋外の状況を確認できないユーザに対して、降雨を伝えることもできる。したがって、このツイートには「交通」だけでなく、「気象」を付与しておくことが望ましい。

著者らは、ツイートに対して複数の局面を付与するために、階層的推定法という新たなマルチラベル分類手法を提案している [19]. 階層的推定法は、大量のツイートから教師なし学習でトピックを抽出する第 1 段階と、少量の訓練データでトピックと局面の対応関係を構築する第 2 段階からなる。対応関係の構築については、複数のトピックで表現される局面や、特定のトピックで表現される局面など、局面によって様々な対応関係がある。このような対応関係を実現するため、局面とトピックの関連度を算出し、局面ごとに決定した閾値を超えた関連度を持つトピックを局面に対応付ける。関連度は、訓練データ中の単語の局面ごとの生起確率と、トピック中の単語の生起確率から算出する。局面を付与する際は、ツイートに出現した単語の各トピック中の生起確率と、トピックと局面の関連度からスコアを算出し、スコアが閾値を超えた局面をツイートに付与する。

このようにして実現する階層的推定法は、従来のマルチラベル分類手法に比べ平均的に高い推定精度を示した。一

方で、対応するトピックが競合する局面間では、推定精度が低くなるという問題が明らかになった。また、局面を付与する閾値を局面によらず一定としたため、過剰に付与される局面があることも明らかになった。

本論文では、ツイートに対する局面の推定精度を向上させるため、新たな関連度の算出方法を提案する。関連度をトピック側で正規化することで、トピックがどの局面に対し強く結び付くかを同定し、局面間に対応付くトピックの競合を防ぐ。トピック側で正規化した関連度を、さらに局面側で正規化することで、局面から見て強く結び付くトピックを同定する。訓練データに付与されているラベルの分布から局面ごとに閾値を決定し、過剰な局面の付与を防ぐことで、推定精度の向上を狙う。

本論文の構成を以下に示す。2 章は、関連研究について述べる。3 章は、マルチラベル分類を実現する階層的推定法について述べた後、提案する改善手法について詳述する。4 章は、大量に収集した実際のツイートをを用いて提案手法の有効性を評価している。5 章で考察を行い、6 章で結論と今後の課題を述べる。

2. 関連研究

2.1 Twitter からの情報抽出に関する研究

Twitter から有益な情報を抽出する研究は、数多く行われている。Sakaki ら [11] は、Twitter ユーザをセンサとみなし、地震などの現実世界で起きるイベントを発見する手法を明らかにしている。Mathioudakis ら [8] は、収集したツイートからバーストキーワードを抽出し、キーワードの共起を用いてクラスタリングを行い、リアルタイムに変動するトレンドの発見を目指している。Zhao ら [17] は、Twitter に投稿された情報要求に関するツイートを抽出し、ユーザの情報要求を分析することで、現実世界のイベントやトレンドを発見できると報告している。Wang ら [14] は、過去のツイートからユーザの興味を推定し、ツイートと興味の近いユーザを推薦する手法を提案している。本論文は、未知のツイートに対して実生活の局面を推定することで、有益な実生活ツイートの抽出を目的としているため、これらの Twitter に関する研究とは異なる。

2.2 トピックモデルを利用した研究

トピックモデルに関する研究では、Blei ら [1] によって提案された潜在的ディリクレ配分法 (LDA) が広く知られている。LDA とは、1 つの文書に複数のトピックが存在すると仮定した確率的トピックモデルであり、それぞれのトピックがある確率を持って文書上に共起するという考えのもと、各トピックの確率分布を導出する教師なし学習モデルである。Riedl ら [10] は、LDA を用いて文書を話題ごとに分割する手法を述べている。LDA で得られた各トピック中の単語の生起確率から、文書中の単語をトピック ID に変

換し、文の境界の前後に一定の単語数の窓を設定し、各窓ごとにトピックの出現頻度の類似度を算出することで、話題の変換点を検出している。Zhang ら [16] は、LDA を用いてアーティストを推薦する手法を提案している。ユーザの嗜好アーティストと、そのアーティストのコミュニティに所属するユーザを特徴量として LDA で生成したトピック集合を用いて、アーティスト間の類似度、ユーザ間の類似度を算出し、精度だけでなく意外性のあるアイテムの推薦も目指している。本論文は、生成したトピックと局面の対応関係を構築し、トピックと局面の関連度とトピック中の単語の生起確率を用いて、未知の Tweet に対して複数ラベルを推定することに特徴がある。

2.3 マルチラベル分類に関する研究

マルチラベル分類手法には、SVM やナイーブベイズ分類器、LDA に基づく手法がある。SVM は、教師あり学習を行う識別手法の 1 つであり、高い分類性能と汎化能力を有している [4]。Chang ら [2] は LIBSVM という SVM 用のライブラリを公開している。LIBSVM は、ラベルの組合せを新たなクラスとしてモデルを構築し分類を行うことで、マルチラベル分類を実現している。

ナイーブベイズ分類器は、テキスト中に含まれる単語の生起が独立であるという仮定をおき、それらの単語が出現したときの、文書のクラスへの所属確率をベイズの定理から算出し、所属確率が最も高いクラスへ文書を分類する手法である [5]。Wei らは [15] は、ナイーブベイズ分類器で算出したクラス別の所属確率から平均値を求め、所属確率が平均値を超えたクラスを文書へ付与するマルチラベル分類手法を提案している。

Ramage ら [9] は、LDA を教師あり学習へ拡張した Labeled LDA (L-LDA) という、マルチラベル分類を目的としたモデルを提案している。L-LDA は文書にあらかじめ付与されているラベルを、その文書の内容を表すものとして利用するモデルである。

いずれの手法も、十分な訓練データを用いることで、ブログや新聞記事など比較的長い文書であれば、高い推定精度を示している。一方で、本研究で対象とするツイートは平均 45 文字と短い [20] ため、手がかりとできる語が少ない。また、本研究では生活の局面を推定対象としている。人間の生活は時間とともに変化していくため、最新に投稿されたツイートを訓練データとすることが望ましい。したがって、できる限り少量の訓練データで高い推定精度を得ることが求められる。以上の条件においては、従来手法は十分な性能が得られないことが課題となっていた [19]。

3. 局面の階層的推定法

3.1 階層的推定法の概要

3.1.1 トピックと局面の対応関係構築方法

著者らが先行研究 [19] で提案した階層的推定法を図 1 に示す。階層的推定法の第 1 段階では、大量のツイートから LDA でトピックを抽出する。抽出したトピックの中から、局面を表現するために必要となるトピックを対応付けるために、トピックと局面の関連の強さを表す関連度を算出する。局面 a とトピック t の関連度 $R(a, t)$ は、

$$R(a, t) = \sum_{w \in W} p(a, w) * p(w, t)^g \quad (1)$$

で算出する。ここで、 $p(w, t)$ は LDA を用いて抽出したトピック t における単語 w の生起確率、 $p(a, w)$ は訓練データから算出した局面 a における単語 w の生起確率である。 g は、トピック中の単語の生起確率の広がり調整するためのパラメータである。

関連度を 0 から 1 の範囲とするために正規化する。ここでは、以下の式 (2) に示す、各局面で正規化した関連度 $\hat{R}a(a, t)$ と、各トピックで正規化した関連度 $\hat{R}t(a, t)$ を用意する。

$$\hat{R}a(a, t) = \frac{R(a, t)}{\sum_{x \in T} R(a, x)}, \quad \hat{R}t(a, t) = \frac{R(a, t)}{\sum_{x \in A} R(x, t)} \quad (2)$$

ここで、 T は LDA で抽出した全トピック、 A は全局面である。 $\hat{R}a(a, t)$ は、局面 a がどのトピックから支持されているかを表す指標であり、 $\hat{R}t(a, t)$ は、トピック t がどの局面を支持しているかを表す指標である。

関連度 $\hat{R}a(a, t)$ を用いて、トピックと局面の対応関係を構築する。局面によっては、1 つのトピックで表現される対応関係や、複数のトピックで表現される対応関係などがある。このような対応関係を構築するために、局面ごとに閾値を決定し、関連度 $\hat{R}a(a, t)$ が閾値を超えたトピックと対応関係を構築する。局面 a と対応関係を構築するトピック集合 T_a は、パラメータ d を用いて、

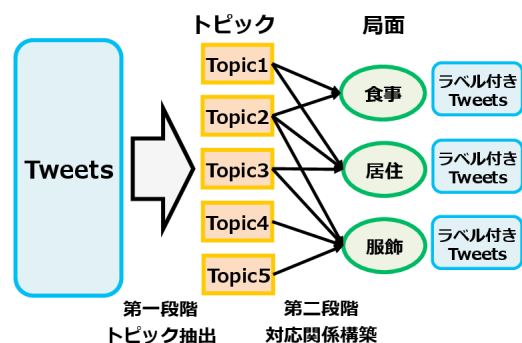


図 1 階層的推定法

Fig. 1 Hierarchical estimation framework.

$$T_a = \{t | \hat{R}a(a, t) > \max_{x \in T} (\hat{R}a(a, x)) - \sigma(\hat{R}a(a, T)) * d\} \quad (3)$$

とする。ここで、 $\sigma(\hat{R}a(a, T))$ は、局面 a の全トピック T に対する関連度の標準偏差である。 d はトピックとの対応関係を調整するためのパラメータであり、 d を大きくすることで、より多くのトピックが局面に関連付けられる。

3.1.2 トピック数の最適化方法

LDA では、生成するトピック数をパラメータとして与える。関連度に基づいてトピックと局面の対応関係が構築されることから、生成するトピック数によって、局面と結びつくトピックが変動する。

最適なトピック数を決定するため、JS Divergence を用いて、ある1つの局面と他の局面との類似度を計算する。2つの局面の確率分布が同じである場合、JS Divergence は0となり、より異なっている場合、JS Divergence は1に近づく。本論文の場合は、局面間の確率分布 $\hat{R}a(a, t)$ が異なっている方が望ましい。そのため、各局面間の JS Divergence の合計値を最大とするトピック数を最適であるとした。JS Divergence の合計値 JS_{sum} は、以下の式で求められる。

$$JS_{sum} = \sum_{(p, q) \in A} D_{JS}(\hat{R}a(p, *) || \hat{R}a(q, *)) \quad (4)$$

$$D_{JS}(P || Q) = \frac{1}{2} \left(\sum_{t \in T} P(t) \log \frac{P(t)}{R(t)} + \sum_{t \in T} Q(t) \log \frac{Q(t)}{R(t)} \right)$$

ここで、 $R(t)$ は確率分布 $P(t)$ と $Q(t)$ の平均であり、 $R(t) = \frac{P(t)+Q(t)}{2}$ で与えられる。

3.1.3 ツイートに対する局面の推定方法

未知のツイートに局面を付与する際は、ツイート中に出現した単語のトピック中の生起確率と、トピックと局面の関連度 $\hat{R}a(a, t)$ と $\hat{R}t(a, t)$ を用いて、局面ごとにスコアを算出する。ツイート tw と各局面 a とのスコア $S(tw, a)$ は、以下の式で算出する。

$$S(tw, a) = \sum_{t \in T} \sum_{w \in W_{tw}} p(w, t)^g * \hat{R}a(a, t) * \hat{R}t(a, t) \quad (5)$$

スコアの平均値 $E(S(tw, A))$ と標準偏差 $\sigma(S(tw, A))$ を用いて各スコアを正規化し、閾値 r を超えたスコアを持つ局面を、ツイートに対して付与する。未知のツイート tw に対して推定する局面集合 A_{tw} は、

$$A_{tw} = \left\{ a \mid \frac{S(tw, a) - \mu(S(tw, A))}{\sigma(S(tw, A))} > r \right\} \quad (6)$$

とする。各スコアについて、平均値との差を標準偏差で除すことで、平均を0、標準偏差を1とした値に正規化する。

3.2 階層的推定法の改善方法

3.2.1 階層的推定法の課題

このようにして実現する階層的推定法は、平均して高い推定精度が得られている一方で、他の局面に比べ推定精度

が低くなる局面があることも明らかとなった。この原因として、以下の4課題があげられる。

課題1 式(2)で得られる関連度 $\hat{R}a(a, t)$ では、いずれの局面にも強く結びつくトピックが出現し、推定精度を低下させる。

たとえば、著者らの先行研究[19]では動詞が集まるトピックを確認している。動詞は局面によらず、多くのツイートに出現することから、このようなトピックはいずれの局面とも高い関連度 $\hat{R}a(a, t)$ で結びつくだけでなく、局面を表現するために特徴的なトピックの関連度を相対的に低下させている。この結果、ツイート中の動詞の出現頻度に応じて、スコア $S(tw, a)$ が高くなるため、ツイートが言及している局面に関係のないスコアが算出され、局面の推定精度が低下する。

課題2 複数のトピックが集まって表現される局面では、課題1に示すトピックの関連度が高くなり、適切な対応関係を構築できない。

たとえば、災害の局面を表現する典型的なトピックはないが、台風や地震など複数のトピックが集まって適切に表現される場合に、これらのトピックと災害の関連度に比べ、課題1で示したトピックと災害の関連度が高くなる。この結果、ツイート中の災害に特徴的な単語に比べて、他の特徴的でない単語が強く考慮されたスコアが算出されるため、推定精度が低下する。

課題3 課題1に示す多くの局面に強く結びつくトピックであるが、局面によっては、自身を表現するために必要なトピックである場合がある。

たとえば、先行研究では地名が集まるトピックを確認している。実生活ツイートは、頻繁に実世界の出来事に言及するため、様々な局面で地名を含めたツイートが多い。この結果、地名が集まったトピックは課題1で示したように様々な局面と高い関連度で結び付き、他の特徴的なトピックの関連度を低下させる。しかし、課題1の例で示したトピックに比べ、地名が集まるトピックは地域の局面を表現するためには重要である。

課題4 スコアから局面を付与する式(6)では、閾値 r を局面によらずに固定としている。 r を低くしたとき、頻繁に出現する局面ではツイートに対して十分に局面が付与されるが、めったに出現しない局面ではツイートに対して過剰に局面が付与されるため、高い再現率が得られる代わりに適合率が低下する。対して、 r を高くした場合では、高い適合率が得られるが再現率が低下するという問題がある。このような問題を解決するため、各局面によって適切な閾値 r を決定する必要がある。

課題1と2については、ストップワードをあらかじめ定義し該当する単語を除去する方法が考えられるが、移り変わりの速いメディアとして知られる Twitter で、ストップ

ワードとなる語を手で列挙するのは現実的ではない。

また、LDA はストップワードを高い生起確率で集めたトピックを生成できるため [12]、ストップワードを多く含むトピックを除去する方法が考えられる。しかし、ほとんどの局面には不要であっても、課題 3 のように特定の局面に対しては必要なトピックも考えられ、一概にそのようなトピックを除去することが難しいと考えられる。

3.2.2 関連度正規化の改善方法

本論文では、課題 1 と課題 2 を解決することを目的に、新たな関連度正規化式を提案する。式 (2) に示す関連度算出式では、いずれの局面とも強く結びつくトピックが出現し (課題 1)、このようなトピックがノイズとなり、局面を適切に表現できるトピックの関連度 $\hat{R}a(a, t)$ を低くしている (課題 2)。一方で、各トピックで正規化した関連度 $\hat{R}t(a, t)$ は、課題 1 に示すようなノイズとなるトピックでは、いずれの局面にも同等に低い関連度を与えるため、課題 1 の解決ができる。しかし、 $\hat{R}t(a, t)$ で対応関係を構築する場合、いずれのトピックとも強く結び付かない局面の出現が考えられるため、課題 2 の解決には至らない。そこで、各トピックで正規化した関連度 $\hat{R}t(a, t)$ をさらに各局面で正規化することにより、局面から見て関連が強いトピックを同定する。新たな関連度の正規化式を以下に示す。

$$\hat{R}t(a, t) = \frac{R(a, t)}{\sum_{x \in A} R(a, x)}, \quad \hat{R}a(a, t) = \frac{\hat{R}t(a, t)}{\sum_{x \in T} \hat{R}t(a, x)} \quad (7)$$

改善前の式 (2) では、関連度 $R(a, t)$ から $\hat{R}a(a, t)$ と $\hat{R}t(a, t)$ の両方を算出していたが、新たな正規化式では、関連度 $R(a, t)$ から $\hat{R}t(a, t)$ を算出した後、 $\hat{R}t(a, t)$ を各局面で正規化した $\hat{R}a(a, t)$ を算出する。

3.2.3 局面を付与する閾値の決定手法

本論文では、課題 4 を解決できる閾値の決定手法を提案する。ナイーブベイズ分類器 [5] では、クラス別の所属確率を算出する際に、文書中に出現する単語だけでなく、訓練データに現れたラベルの数も手がかりにしている。ラベルの数は、事前確率としてナイーブベイズ分類器の式中に導入されており、訓練データ中に多く出現するラベルには高い確率を与え、あまり出現しないラベルには低い確率を与える。本論文では、訓練データ中に出現する局面の数の分布を求め、スコアの正規化式と同様に、平均が 0 で標準偏差が 1 となるように正規化したものを閾値とする。局面 a の閾値 $r(a)$ は、

$$r(a) = \frac{E(L(A)) - L(a)}{\sigma(L(A))} \quad (8)$$

とする。ここで、 $L(a)$ は局面 a のラベル数であり、 $E(L(A))$ はラベル数の平均、 $\sigma(L(A))$ はラベル数の標準偏差である。平均値から $L(a)$ の差を求めることにより、局面 a が訓練データ中に多いほど閾値 $r(a)$ が低くなるため、局面 a は

付与されやすい局面とできる。逆に、局面 a が訓練データ中に少ないほど、局面 a は付与されにくい局面とできる。閾値 $r(a)$ を、式 (6) の r とすることで、局面ごとに最適な閾値を決定する。

4. 評価実験

改善後の階層的推定法を用いて、適合率、再現率、F 値を評価する。比較対象として、改善前の階層的推定法と、SVM、LDA を教師あり学習へと拡張した L-LDA、ナイーブベイズ分類器をマルチラベリングへ拡張した NBML を用いる。また、LDA で抽出するトピック数を変更した場合の提案手法の推定精度を比較することで、3.1.2 項で示したトピック数選択方法の有効性を明らかにする。

4.1 データセットとパラメータ設定

4.1.1 データセット：トピック抽出に用いるツイート

提案手法の第 1 段階における LDA を用いたトピック抽出のために、2012 年 4 月 15 日から 2012 年 8 月 14 日の間に、日本語で Twitter に投稿されたツイートを Search API*2 を用いて収集した。その中から、ツイートのロケーション情報に「京都」あるいは「Kyoto」と入力されているツイートを使用した。以上の条件により評価に使用するツイート数は、2,390,553 件となった。

4.1.2 データセット：実生活ツイート

トピックと局面の対応関係を構築するため、人手によってツイートに局面をラベル付けした。1,500 件のツイートに対して、第 1 著者 (実験参加者 A) と他 2 名 (実験参加者 B および C) の合計 3 名で、各ツイートに対して適切な局面を付与する人手判定を行った。実験参加者にはガイドラインとして、表 1 に示す各局面に含まれる典型的な単語と、その局面に分類されるツイートの例 (各局面 1 件ずつ) と、それが分類された理由を提示した。人手判定では、各ツイートに対して適切な局面として第 1 位、第 2 位、第 3 位まで付与することとした。いずれの局面にも適さないと判断した場合は、「非実生活」を付与することとした。第 3 位以内に該当しなかった局面はすべて第 4 位とした。なお、用意した 1,500 件のツイートは、いずれもロケーション情報に「京都」あるいは「Kyoto」と表記されたものであり、上記のトピック抽出に使用するツイートと重複はない。また、3 名の実験参加者はいずれも「つくば市」在住の大学生であり、京都で生活したことはない。

人手判定の結果、第 1 位に分類された局面について、実験参加者間の一致度を κ 値 [3] によって評価した。実験参加者 A と実験参加者 B の κ 値は 0.687、実験参加者 A と実験参加者 C の κ 値は 0.595、実験参加者 B と実験参加者 C の κ 値は 0.576 となった。 κ 値の平均は 0.619 であり、高

*2 <https://dev.twitter.com/docs/api/1/get/search>

表 2 人手判定の結果，正解ラベルとして付与された局面の数

Table 2 Number of correct-label aspects.

局面	服飾	交流	災害	食事	行事	消費	健康	趣味	居住	地域	学校	交通	気象	労働	非実	合計
ラベル数	181	379	86	287	311	435	177	348	213	432	195	169	226	262	1,391	5,092
$r(a)$	0.53	-0.13	0.85	0.18	0.10	-0.32	0.54	-0.03	0.42	-0.31	0.49	0.57	0.38	0.26	-3.53	

い一致 (substantial) であった。

各ツイートに対して複数の局面をラベル付けするため，3名の手判定の結果を用いる．ツイート tw に対して正解となる局面集合 AC_{tw} は，

$$AC_{tw} = \{a | U_{score}(tw, a) \leq 10\} \quad (9)$$

とする．ここで， $U_{score}(tw, a)$ は，実験参加者がツイート tw に対して，局面 a を第何位に選択したかを合計した値であり，以下の式で求められる。

$$U_{score}(tw, a) = \sum_{u \in U} rank(tw, a, u) \quad (10)$$

ここで， U はすべての実験参加者を表し， $rank(tw, a, u)$ は，ツイート tw に対して実験参加者 u が局面 a を分類した順位である．実験参加者がいずれの順位にも入れなかった局面は，値を 4 として計算する． $U_{score}(tw, a)$ の最小値は，実験参加者 3 名が同じ局面を第 1 位に選択した場合であり， $rank(tw, a, u) = 1$ となるため， $U_{score}(tw, a) = 3$ となる．最大値は，実験参加者の全員が特定の局面を第 3 位以内に選択しなかった場合，すなわち， $rank(tw, a, u) = 4$ の場合で，そのとき $U_{score}(tw, a) = 12$ となる。

以上の処理によって，人手判定した 1,500 件のツイートに対して，ラベル付けされた局面数を集計した結果を表 2 に示す．服飾の局面は，1,500 件のツイートの中で合計で 181 件ラベル付けされている．1,500 件のツイートに対するすべてのラベル数は 5,092 件となっており，1 つのツイートに対して平均 3.39 件のラベルが付与されている．評価実験では，いずれの局面にも属さない「非実生活」についても 1 つのクラスとしてトピックと対応関係を構築し，非実生活を推定できることを評価する。

正解データとして使用する 1,500 件のツイートについて，各ツイートに付与された局面の数を集計した結果を表 5 に示す．最も多いラベル数は 3 で，820 件のツイートが存在する．ラベル数が 6 あるツイートは 11 件存在する。

4.1.3 提案手法のパラメータ設定

LDA は，事前いくつかのパラメータを設定する必要がある．関連研究 [6] を参考に，ハイパーパラメータである α は $\frac{50}{|T|}$ ， β は 0.1 とした． $|T|$ は LDA で生成するトピック数である．イテレーション回数は，予備実験の結果から安定した値が得られる 100 とした．また，提案手法の関連度算出式 (1) やスコア算出式 (5) のパラメータ g は，予備実験の結果から $g = 0.3$ とした。

4.2 評価尺度

提案手法の有効性を議論するには，推定した局面がどれだけ正解しているかという正確性と，すべての正解のうちどれだけ提案手法で局面を推定できたかという網羅性の，2 つの観点から評価するために，本論文では，正確性を適合率 (Precision)，網羅性を再現率 (Recall)，および適合率と再現率の調和平均である F 値 (F-measure) によって提案手法の推定精度を評価する．実験では，1,500 件の正解データについて，10 分割交差検定によってトピックと局面の対応関係を構築，および推定精度の評価を行う。

4.3 比較手法

本節では，提案手法の比較対象となる各手法の設定について説明する．各手法に与える素性は共通して，提案法と同様に，ツイートを MeCab [7] により形態素解析した結果得られた名詞，動詞，形容詞の Bag of Words とした。

4.3.1 L-LDA

L-LDA (Labeled LDA) は，マルチラベルが付与されたデータからモデルを学習し，未知のデータに対するマルチラベル分類が可能である．L-LDA は Web ページの分類において，SVM を上回る推定性能を示している [9]．L-LDA のイテレーション回数は安定した値が得られる 100 とし，ハイパーパラメータである α と β は，ともに 0.1 とした。

4.3.2 SVM

本論文では，SVM の実装として，マルチラベル分類も可能である LIBSVM [2] を比較対象として使用する．カーネルは線形カーネルを選択し，パラメータは LIBSVM のツールでグリッドサーチ^{*3}を実行し得られた $C = 1.0$ とした。

4.3.3 NBML

比較対象の 1 つとして，NBML (Naive Bayes Multi-label Classification) を用いる．NBML は Wei らが，Naive Bayes 分類器をマルチラベル分類へ拡張した手法である [15]．NBML はクラス別の所属確率から平均値を求め，平均値を超える所属確率を持つクラスへ文書を分類することで，マルチラベル分類を実現している。

4.4 実験結果

4.4.1 トピック数ごとの JS_{sum} 値と推定精度の比較

LDA で抽出するトピック数を，50, 100, 200, 500, 1,000

^{*3} LIBSVM Tools: Grid parameter search for regression
http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/#grid_parameter_search_for_regression

表 4 各手法の適合率, 再現率, F 値

Table 4 Precision, Recall, F-measure of each method.

局面	適合率					再現率					F 値				
	改善前	改善後	L-LDA	SVM	NBML	改善前	改善後	L-LDA	SVM	NBML	改善前	改善後	L-LDA	SVM	NBML
服飾	0.72	0.65	0.43	0.64	0.83	0.57	0.67	0.69	0.28	0.3	0.63	0.65	0.52	0.38	0.51
交流	0.37	0.43	0.43	0.41	0.53	0.68	0.58	0.68	0.35	0.54	0.48	0.49	0.53	0.37	0.53
災害	0.34	0.67	0.67	0.44	0.76	0.59	0.49	0.49	0.44	0.21	0.43	0.55	0.54	0.44	0.33
食事	0.77	0.78	0.41	0.51	0.73	0.75	0.80	0.77	0.64	0.51	0.76	0.79	0.53	0.57	0.60
行事	0.42	0.36	0.47	0.56	0.56	0.57	0.73	0.51	0.20	0.45	0.49	0.47	0.48	0.29	0.49
消費	0.52	0.41	0.64	0.43	0.52	0.55	0.75	0.40	0.45	0.46	0.53	0.53	0.49	0.43	0.49
健康	0.37	0.54	0.43	0.48	0.76	0.63	0.54	0.55	0.28	0.38	0.47	0.54	0.48	0.35	0.50
趣味	0.38	0.56	0.44	0.43	0.57	0.77	0.48	0.62	0.54	0.44	0.51	0.51	0.51	0.48	0.49
居住	0.58	0.59	0.38	0.64	0.71	0.57	0.57	0.62	0.34	0.41	0.57	0.58	0.46	0.44	0.51
地域	0.81	0.45	0.62	0.62	0.62	0.58	0.90	0.73	0.54	0.65	0.67	0.60	0.67	0.57	0.63
学校	0.73	0.73	0.37	0.88	0.81	0.65	0.68	0.81	0.36	0.52	0.68	0.70	0.51	0.49	0.63
交通	0.43	0.64	0.33	0.71	0.82	0.82	0.76	0.82	0.44	0.50	0.56	0.69	0.47	0.54	0.62
気象	0.69	0.91	0.25	0.47	0.81	0.70	0.58	0.84	0.63	0.58	0.70	0.71	0.38	0.53	0.67
労働	0.44	0.64	0.64	0.52	0.56	0.60	0.43	0.50	0.19	0.35	0.51	0.50	0.55	0.28	0.43
非実	0.93	0.93	0.94	0.93	0.93	0.93	0.99	0.51	0.99	0.93	0.93	0.96	0.66	0.96	0.93
平均	0.57	0.62	0.50	0.58	0.70	0.66	0.66	0.63	0.44	0.49	0.59	0.64	0.52	0.47	0.56

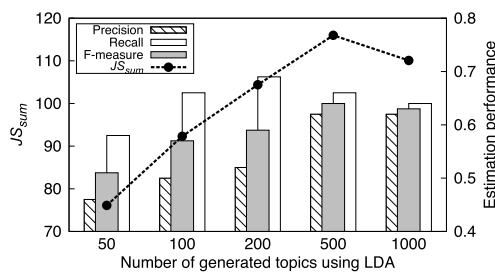


図 2 各トピック数における JS_{sum} 値とその推定精度

Fig. 2 JS_{sum} and estimation performance in each topic.

と変化させ、 JS_{sum} を算出するとともに、各トピック数で実際にツイートに局面を推定した際の、適合率, 再現率, F 値を図 2 に示す. 図では、横軸にトピック数、左の縦軸に JS_{sum} 値、右の縦軸に推定精度を示している. 推定精度については、各トピック数ごとの差異が明らかとなるように、0.4 から 0.8 までの範囲としている. それぞれ、10 分割交差検定により F 値を最大とするときのパラメータ d を選択した. d はトピックと局面の対応関係を構築する際に変化させることで、最適な値を得られる. JS_{sum} が最大となったトピック数は 500 であり、トピック数 500 で適合率と F 値が最大値を示した. したがって、以降の実験では LDA で抽出するトピック数を 500 とし、トピックと局面の対応関係を構築する.

4.4.2 改善前と改善後の階層的推定法の推定精度の比較

本論文で提案した改善手法を階層的推定法に適用した場合と、適用しなかった場合で推定精度を評価した結果を表 3 に示す. 改善手法を適用しない場合は、式 (6) におけるパラメータ r を決定する必要があることから、0.0, 0.5, 1.0 と r を変化させ、それぞれの適合率, 再現率, F 値を評価した. パラメータ d は 4.4.1 項と同様に、F 値を最大とする点を選択した. 適合率では、改善前の $r = 1.0$, 再現

表 3 各閾値を用いた階層的推定法の推定精度の比較

Table 3 Compared to estimation performance of hierarchical estimation framework using each threshold value.

手法	Precision	Recall	F-measure
改善前, $r = 0.0$	0.57	0.66	0.59
改善前, $r = 0.5$	0.61	0.57	0.58
改善前, $r = 1.0$	0.68	0.54	0.58
改善後	0.62	0.66	0.64

率では改善前の $r = 0.0$ と改善後の手法, F 値では改善後の手法が最大となった.

4.4.3 各手法の推定精度の比較

各手法の適合率, 再現率, F 値を表 4 に示す. 改善前の提案手法については、F 値で最大値を示した $r = 0.0$ とパラメータ d を選択した. 太字となっている数値は、すべての手法で比較したときの最大値である. 服飾の局面では、適合率の最大値は 0.83 で NBML, 再現率の最大値は 0.69 で L-LDA, F 値の最大値は 0.65 で改善後の提案法となっている. 災害の局面では、改善前の提案法は F 値が 0.43 であるのに対し、改善後の提案法は 0.55 と 0.12 高くなっている. 同様に、交通の局面でも改善後の提案法が 0.13 高い値を示している. すべての局面についてマクロ平均をとった結果、各評価値で最大値を示したのは、適合率で NBML, 再現率で改善前と改善後の提案法, F 値で改善後の提案法であった.

提案手法と比較手法について、各ツイートに対して推定したラベル数を集計した結果を表 5 に示す. ラベル数の最頻値は改善前と改善後の提案法で 5, SVM では 3, NBML では 2 であった. ツイートあたりの平均ラベル付与数は、改善後の提案法が 5.15 で最大となった. 人手判定の結果に最も近い平均ラベル付与数を示したのは SVM であった.

表 6 各局面に対して高い関連度 $\hat{R}a$ で結び付くトピック

Table 6 Topics with high relevance $\hat{R}a$ associated to each aspect.

局面	$\hat{R}a$ 1 位			$\hat{R}a$ 2 位			$\hat{R}a$ 3 位			$\hat{R}a$ 4 位		
	topic	$\hat{R}a$	$\hat{R}t$	topic	$\hat{R}a$	$\hat{R}t$	topic	$\hat{R}a$	$\hat{R}t$	topic	$\hat{R}a$	$\hat{R}t$
改善前：服飾	#119	0.023	0.667	#474	0.009	0.399	#240	0.009	0.385	#454	0.007	0.170
改善後：服飾	#119	0.024	0.667	#474	0.014	0.399	#240	0.014	0.385	#222	0.008	0.228
改善前：災害	#178	0.020	0.234	#469	0.014	0.205	#277	0.013	0.404	#380	0.013	0.253
改善後：災害	#277	0.019	0.404	#391	0.019	0.395	#380	0.012	0.253	#150	0.011	0.239
改善前：交通	#60	0.032	0.369	#201	0.025	0.582	#149	0.020	0.455	#42	0.016	0.440
改善後：交通	#201	0.023	0.582	#149	0.018	0.455	#42	0.017	0.440	#60	0.014	0.369
改善前：地域	#60	0.019	0.279	#314	0.015	0.258	#44	0.010	0.182	#486	0.008	0.147
改善後：地域	#60	0.009	0.279	#314	0.008	0.258	#28	0.008	0.237	#125	0.007	0.209

表 5 各手法のラベル数別のツイート数

Table 5 Number of labeling by each method.

ラベル	改善前	改善後	L-LDA	SVM	NBML	人手判定
1	20	0	101	0	165	1
2	23	0	154	137	531	111
3	83	22	259	1,250	442	820
4	277	389	369	80	243	442
5	504	574	307	33	90	115
6	425	382	182	0	23	11
7	143	118	80	0	6	0
8	25	15	37	0	0	0
9	0	0	9	0	0	0
10	0	0	2	0	0	0
平均	5.13	5.15	4.16	3.00	2.75	3.39

表 7 交通の局面に結び付くトピックの上位語

Table 7 High occurrence probability terms in topics associated to Traffic aspect.

トピック	上位語
Topic60	京都, 交通, 河原町, 四条, 三条 地下鉄, 烏丸, 便利, 嵐山, 案内
Topic201	電車, 乗る, 阪急, 降りる, 遅れる 特急, 快速, 車両, 乗車, 列車
Topic149	バス, 乗る, 夜行, 運転, 高速 勤務, 渋滞, タクシー, 新幹線, 事故
Topic42	止まる, 京阪, 事故, でる, 全部 全国, 阪神, タイヤ, 電車, 処理

5. 考察

5.1 階層的推定法の改善手法に関する考察

図 2 より, JS_{sum} 値はトピック数が 500 のときに最大となった. トピック数ごとの推定精度を見ると, 適合率と F 値でもトピック数が 500 のときに最大値を示している. このことから, JS_{sum} 値によるトピック数の選択手法は, 階層的推定法において有効であると考えられる. また 500 の近辺でいずれの値もほぼ安定していることから, ある程度のトピック数とすることで, 安定した性能が得られるといえる.

図 3 より, 改善前と改善後の階層的推定法の推定精度を見ると, 適合率では改善前の提案法で $r = 1.0$ としたときが最大であるが, 再現率と F 値では改善後の提案法が最大となった. 改善後の閾値決定手法は, 閾値 r は訓練データのラベルの分布から局面ごとに算出するため, 評価対象のデータに適した閾値が決定できる. このため, 改善後の提案法は再現率を低下させず適合率を上昇できたと考えられる.

表 4 より, 改善前に比べ, 災害と交通の局面では F 値が大きく上昇し, 地域の局面は F 値が大きく低下している. 改善手法の効果を確認するため, これら 3 つに加えて服飾の局面における, 改善前と改善後のトピックとの対応関係を表 6 に示す. 服飾の局面は, 改善前と改善後とともに推定

精度が高いことから, 模範となる局面として選定した. 表では, 局面ごとに $\hat{R}a(a, t)$ の高い上位 4 トピックを抽出し, そのトピック番号と関連度 $\hat{R}a(a, t)$ と $\hat{R}t(a, t)$ を示している. 改善前の災害 (Dis.) の局面に対して最も強く結び付くトピックは, Topic178 であり, $\hat{R}a(Dis., Topic178) = 0.020$, $\hat{R}t(Dis., Topic178) = 0.234$ である.

服飾の局面は, 改善前と改善後の上位 3 トピックは同じであり, $\hat{R}a$ 1 位のトピックである Topic119 は, $\hat{R}t(a, t)$ の値が他のトピックに比べて高い. 服飾の局面は, Topic119 という局面を表現する典型的なトピックが結び付いたため, 高い推定精度を示せたと考えられる. 災害や交通の局面では, 改善前に比べ改善後の上位トピックに $\hat{R}t(a, t)$ の値が高いトピックが結び付いており, 服飾のような推定精度が高い局面の対応関係に近づいていることが分かる.

交通の局面では改善前に $\hat{R}a$ 1 位であった Topic60 が, 改善後に $\hat{R}a$ 4 位となり, 他のトピックは 1 つずつ順位を上げている. 交通の局面に結び付くトピックを詳しく分析するため, 各トピックの上位語を表 7 に示す. Topic60 以外のトピックでは, 交通の局面に関連する単語を数多く確認できる. これに対して Topic60 では, 「交通」や「地下鉄」など交通の局面に関連する単語もあるが, 「京都」や「河原町」など地名に関する単語が多い. 地名は他の局面でも出現しやすい単語であるため, Topic60 は他の局面とも強く結び付くことが考えられる. 3.2 節の課題 1 と課題 2

で言及したように、このようなトピックが原因で、交通の局面の推定精度を低下させていたが、関連度の正規化手法の改善により、Topic60よりも交通を適切に表現できるトピックの関連度を高められ、推定精度が大きく向上したと考えられる。

一方で、地域の局面は推定精度が改善前に比べ低くなった。表6より、地域の局面に $\hat{R}a$ 1位と2位で結び付くトピックに変化はないが、関連度 $\hat{R}a(a, t)$ の値が改善前に比べ約半分に低下した。 $\hat{R}a$ 1位で結び付くTopic60は上記で述べたとおり、地名が集まったトピックであるために他の局面とも強く結び付くと考えられる。このようなトピックは、他の局面では推定精度を低下させる要因となるが、地域の局面を表現するためには必要なトピックである。本論文で提案した関連度の正規化手法は、Topic60のように多数の局面に強く結び付くトピックには低い関連度を与えるため、地域の局面とTopic60の関連度 $\hat{R}a(a, t)$ が低下し、地域の局面の推定精度を低下させたと考えられる。Topic60と地域の局面のような対応関係は、3.2節で述べた課題3の代表であり、解決にはさらなる手法の精緻化が必要である。

5.2 各手法の推定性能に関する考察

表4より、F値では改善後の提案法（本節では提案法とする）、NBML、L-LDA、SVMの順に高い値を示した。提案法に比べ、適合率で高い値を示したのはNBMLであった。表5より、提案法はすべての手法の中で最も多い平均ラベル付与数を示し、一方でNBMLはすべての手法の中で最も少ない平均ラベル付与数であった。このことから、適合率ではNBML、再現率では提案法が高い値を示したと考えられる。L-LDAは平均ラベル付与数が提案法に比べ少ないにもかかわらず、適合率と再現率の両評価値で提案法が高い値を示している。このことから、L-LDAに比べ提案法がツイートに相応しい局面のスコアを高精度に算出できていることが考えられる。提案法の平均ラベル付与数は約5件であり、正解データの平均ラベル付与数に比べ、2件多い結果となった。提案法が対象データセットに応じた数のラベルを付与するためには、訓練データに付与されているラベル数の分布からパラメータを算出し、提案法の式(6)に組み込むことなどが考えられる。

提案法とNBMLが、実際にツイートに局面を付与した例を表8と表9に示す。表8は、人手判定によって得られた正解の局面を、提案法が完全に推定することに成功した例である。ツイートは宇治市の住居と天気に関する情報を記述しており、人手判定で得られた局面はその内容に一致している。NBMLが地域の局面を推定できなかった原因は、「宇治橋」という単語が、訓練データ中に一度も出現していなかったために、「地域」の局面との尤度を適切に計算できていなかったためである。提案法は、「宇治」が高

表8 提案法が有効に機能したツイート例
Table 8 Complete estimated aspects for tweet by HEF.

正解局面 (人手)	居住, 地域, 気象
NBML が推定した局面	居住, 気象
提案法が推定した局面	居住, 地域, 気象
ツイート本文	春の観光シーズンで宇治橋周辺は お天気もいいので賑わっています。 近鉄小倉駅の新築も残り2室です。

表9 提案法が過剰に局面を推定したツイート例
Table 9 Estimated extra aspects for tweet by HEF.

正解局面 (人手)	消費, 地域
NBML が推定した局面	消費, 地域, 食事
提案法が推定した局面	消費, 地域, 食事
ツイート本文	四条河原町のマックの前で ホットコーヒー無料配布しとる! マックのコーヒーの推し半端ない!

い生起確率で所属するトピックと地域の局面との関連度を手がかりに、地域の局面に対して高いスコアを算出したため、完全に正解局面を推定することができた。

表9は、人手判定によって得られた正解の局面に対して、提案法が過剰な局面を推定した例である。正解の局面は消費と地域であり、ツイートの内容と一致している。提案法とNBMLは、この2つの局面に加えて食事の局面を推定している。NBMLについては、「コーヒー」という単語が食事の局面に対して高い尤度を持っていたためである。提案法については、「コーヒー」という単語が高い生起確率で所属するトピックのいずれもが、食事の局面と高い関連度で結び付いていたためである。このような場合に、ツイートに過剰な局面を推定しないための改善案としては、1つの単語だけでなく、ツイート中に出現する複数の単語を組合せたスコアの算出方法が考えられる。

6. 結論

本論文では、局面の推定精度を向上するために、階層的推定法の改善手法を提案した。関連度の正規化の改善手法では、局面間では対応付くトピックが競合することを防ぐため、各トピックで正規化した関連度をさらに各局面で正規化することで、局面に強く結び付くトピックを同定する。閾値の自動決定手法では、訓練データに付与されているラベルの分布を用いて、頻出する局面は閾値を低くし、あまり現れない局面は閾値を高くすることで、局面ごとにラベルの付与されやすさを調整する。

改善手法の有効性を評価するため、実際のツイートを用いて評価実験を行った結果、すべての局面におけるF値の平均も改善前の手法より向上できることを明らかにした。推定精度が大きく上昇した局面では、改善前に局面間で競合していたトピックの関連度を低下させることにより、局

面を適切に表現するトピックの関連度を相対的に高め、これまで適切に推定できなかった局面についても、推定精度を向上できることを明らかにした。

今後の課題は、関連度や対応関係の構築手法を精緻化することにより、局面を表現するのに不可欠なトピックを抽出する手法を明らかにすること、実験参加者に上限なしでツイートに局面を付与してもらい、本論文とは異なる特性を持った正解データを用いて評価すること、および階層的推定法を別のデータセットに適用し、提案法の有効性を検証することである。

謝辞 本研究の一部は、JSPS 科研費 25280110 の助成を受けたものです。ここに記して謝意を示します。

参考文献

- [1] Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent Dirichlet Allocation, *The Journal of Machine Learning Research*, Vol.3, pp.993–1022 (2003).
- [2] Chang, C. and Lin, C.: LIBSVM: A Library for Support Vector Machines, *ACM Trans. Intell. Syst. Technol.*, Vol.2, No.3, pp.1–27 (2011).
- [3] Cohen, J.: A Coefficient of Agreement for Nominal Scales, *Educational and Psychological Measurement*, Vol.20, No.1, pp.37–46 (1960).
- [4] Cortes, C. and Vapnik, V.: Support-Vector Networks, *Mach. Learn.*, Vol.20, No.3, pp.273–297 (1995).
- [5] Domingos, P. and Pazzani, M.: On the Optimality of the Simple Bayesian Classifier Under Zero-One Loss, *The Journal of Machine Learning Research*, Vol.29, No.2-3, pp.103–130 (1997).
- [6] Griffiths, T.L. and Steyvers, M.: Finding scientific topics, *The National Academy of Science*, Vol.101, pp.5228–5235 (2004).
- [7] Kudo, T.: Yet Another Part-of-Speech and Morphological Analyzer (2005), available from (<http://mecab.sourceforge.net/>).
- [8] Mathioudakis, M. and Koudas, N.: TwitterMonitor: Trend Detection over the Twitter Stream, *Proc. 2010 Int'l Conf. Management of Data, SIGMOD '10*, pp.1155–1158 (2010).
- [9] Ramage, D., Hall, D., Nallapati, R. and Manning, C.D.: Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-labeled Corpora, *Proc. 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP '09*, pp.248–256 (2009).
- [10] Riedl, M. and Biemann, C.: TopicTiling: A Text Segmentation Algorithm Based on LDA, *Proc. ACL 2012 Student Research Workshop, ACL '12*, pp.37–42 (2012).
- [11] Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, *Proc. 19th Int'l Conf. World Wide Web, WWW '10*, pp.851–860 (2010).
- [12] Suresh, V., Krishnamurthy, A., Badrinath, R. and Veni Madhavan, C.: *A Stylographic Study and Assessment of Machine Translators*, Lecture Notes in Computer Science, Vol.7014, pp.364–375 (2011).
- [13] Twitter, Inc.: Twitter Reports Third Quarter 2014 Results (2014), available from (<https://investor.twitterinc.com/releasedetail.cfm?ReleaseID=878170>).
- [14] Wang, B., Wang, C., Bu, J., Chen, C., Zhang, W.V., Cai, D. and He, X.: Whom to Mention: Expand the Diffusion of Tweets by @ Recommendation on Microblogging Systems, *Proc. 22nd Int'l Conf. World Wide Web, WWW '13*, pp.1331–1340 (2013).
- [15] Wei, Z., Zhang, H., Zhang, Z., Li, W. and Miao, D.: A Naive Bayesian Multi-label Classification Algorithm with Application to Visualize Text Search Results, *International Journal of Advanced Intelligence*, Vol.3, No.2, pp.173–188 (2011).
- [16] Zhang, Y.C., Séaghdha, D.O., Quercia, D. and Jambor, T.: Auralist: Introducing Serendipity into Music Recommendation, *Proc. 5th ACM Int'l Conf. Web Search and Data Mining, WSDM '12*, pp.13–22 (2012).
- [17] Zhao, Z. and Mei, Q.: Questions About Questions: An Empirical Analysis of Information Needs on Twitter, *Proc. 22nd Int'l Conf. World Wide Web, WWW '13*, pp.1545–1556 (2013).
- [18] 山本修平, 佐藤哲司: Twitter からの実生活情報の抽出法の提案, 第 4 回データ工学と情報マネジメントに関するフォーラム DEIM2012 論文集, F3-4 (2012).
- [19] 山本修平, 佐藤哲司: トピックと局面の対応関係に基づく実生活ツイートのマルチラベル分類, 情報処理学会論文誌データベース (TOD), Vol.7, No.2, pp.24–36 (2014).
- [20] 水沼友宏, 池内 淳, 山本修平, 山口裕太郎, 佐藤哲司, 島田 諭: Twitter におけるバーストの生起要因と類型化に関する分析, 情報社会学会誌, Vol.6, No.2, pp.69–84 (2013).

推薦文

本論文は、Twitter 上のツイートに対し、食事や交通、災害などの局面を推定して分類する方法の提案に関するものである。著者らは、教師なし学習として知られる LDA を用いて、大量ツイートからトピックを抽出する第 1 段階と、少量の訓練データを用いてトピックと局面の関連度を算出し、対応関係を構築する第 2 段階からなる階層的推定法を提案している。本論文では、対応付くトピックが競合する局面で推定精度が低下する問題を解決するため、関連度を各トピックで正規化することでトピックが強く結び付く局面を同定し、対応付くトピックの競合を防ぐ、新たな関連度の算出方法を提案。収集した実データを用いた評価実験を行い、その有効性を実証している。提案内容、評価ともに適切に論述されており、また、提案手法が効果的でない事例なども示して有用な情報を提供している点が高く評価できる優れた論文であるため、推薦する。

(グループウェアとネットワークサービス研究会主査
市村 哲)



山本 修平

2014 年筑波大学大学院図書館情報メディア研究科博士前期課程修了。修士(情報学)。現在、同研究科博士後期課程に在籍。情報検索、データマイニング等に興味を持つ。日本データベース学会学生会員。



佐藤 哲司 (正会員)

1980年山梨大学工学部電子工学科卒業。同年日本電信電話公社(現NTT)武蔵野電気通信研究所に入所。以来、データベースマシン、マルチメディアデータベース、情報検索・情報共有の高次化等に関する研究・開発に従事。

1994年工学博士(大阪大学)取得。2007年4月より筑波大学大学院図書館情報メディア研究科教授、現在に至る。情報検索・知識発見、社会ネットワーク分析、社会インタラクションに興味を持つ。電子情報通信学会、日本データベース学会各会員。