

マイクロブログにおける投稿パターンを考慮した コミュニケーションスキルの分析

石川 将吾 †

武田 直人 ††

関 洋平 †††

† 筑波大学 情報学群 知識情報・図書館学類

†† 筑波大学大学院 図書館情報メディア研究科

††† 筑波大学 図書館情報メディア系

1 はじめに

イベントや企画の開催にあたっては、ソーシャルメディアを利用することができれば、多くの参加者を集めることが可能となる。一方、知らない人と繋がるためには、その人のコミュニケーションスキルを判定する必要がある。本研究は、マイクロブログへの投稿時間帯や他のユーザとの交流活動を手がかりとして、コミュニケーションスキルを分類する手法を提案する。

2 関連研究

Twitterの投稿パターンを利用した研究について説明する。田中ら [1] は、Twitterに投稿されるツイートから、特定の時間帯のユーザの行動を推定している。一方、Takedaら [2] や加藤ら [3] は、教師あり学習を用いて、ユーザの属性を推定する手法を提案した。双方とも、素性としては、ツイートの内容、投稿時間帯を利用しており、Takedaらは、投稿場所の情報も利用している。また、伊集ら [4] は、投稿時間帯とその頻度から、ユーザの年齢層別にその時間にどのような投稿をしているのかを示すモデルを構築した。

本研究では、教師あり学習を用いて、投稿時間帯やツイートの種類を素性として、ユーザのコミュニケーションスキルを推定する。

3 投稿パターンを考慮したユーザの適性分析

本研究では、ユーザの発信したツイートの内容を参照して作成した訓練データを用いて、特定の時間帯の投稿頻度や、ツイートの種類を素性として、教師あり学習に基づきコミュニケーションスキルを分類する。

3.1 訓練データの作成

イベントや企画の開催にあたっては、開催地の近隣の地域の住民に声をかける必要がある。本研究の実験では、つくば市のユーザを対象とした。具体的には、つくば市の広告や宣伝のためのアカウントやbotを除外し

た966ユーザを対象とした。また、各ユーザについて、最大で1,000ツイートを新しいものから順に収集した。さらに、収集したツイートの内容を参照して、人手により分類をした結果、コミュニケーションスキルの高いユーザは177人、低いユーザは147人、どちらでもないユーザは642人となった。コミュニケーションスキルの高いユーザの判定は、挨拶やお礼、礼賛、心配、応援などの他人への気遣い、頻繁な活動報告を手がかりとし、低いユーザの判定は、犯罪・性的発言など現実社会で見ない発言や、社会情勢への不満が多い、情緒の不安定さを手がかりとした。

3.2 投稿パターンの抽出と素性の選択

本研究では、ユーザのTwitter上の活動およびその時間帯の傾向を明らかにして、コミュニケーションスキルの分析に利用する。ユーザの投稿パターンは、以下の4つの素性を利用してベクトルを構築する。

- 通常ツイートの時間帯別の投稿数の割合
- メンションの時間帯別の投稿数の割合
- リツイートの時間帯別の投稿数の割合
- ユーザの1日あたりの投稿数の平均
- 投稿時間帯についての投稿数の分散

投稿時間帯は、1時間単位で定義した。また、すべての時間帯別の投稿数の割合について、相互情報量を用いて、コミュニケーションスキルの高いユーザ、低いユーザ、どちらでもないユーザについてそれぞれ特徴のある上位 N 件*の投稿時間帯を、素性として選択した。さらに、ユーザがツイートを投稿する時間帯の頻度を分散を用いて表し、活動が同一の時間帯に集中しているか、様々な時間帯におよんでいるか、等の情報を反映した指標として利用する。したがって、投稿パターンのベクトルの次元数は、 $3 \times N + 2 - M^\dagger$ となる。

3.3 教師あり学習によるユーザの分類

前節で説明した訓練データと素性を利用して、教師あり学習により、コミュニケーションスキルの3つのクラスを分類する。機械学習の分類器は、Random Forestを用いた‡。また、実装にはscikit-learn§を利用した。

*4節の実験では、 $N = 12$ とした。

† M は3つのクラスで重複して選択した素性の数

‡なお、ロジスティック回帰およびパラメータ調整したSVMによる分類も行ったが、Random Forestを使用した結果が、最良の値となったため、本稿ではRandom Forestの結果のみを示す。

§<http://scikit-learn.org/stable/>

Communication Skill Analysis Focusing on Posting Patterns in Microblog

†Shogo ISHIKAWA ††Naoto TAKEDA †††Yohei SEKI

†College of Knowledge and Library Science, School of Informations, University of Tsukuba

††Graduate School of Library, Information and Media Studies, University of Tsukuba

†††Faculty of Library, Information and Media Science, University of Tsukuba

4 実験：コミュニケーションスキルの分類

4.1 目的

提案手法の有効性と素性の効果を明らかにするために、次の2つの手法との比較実験を行う。(1) ベースライン：ランダムに正解ラベルを付与する手法；(2) 比較手法：提案手法から1日あたりの投稿数の平均と投稿時間帯についての投稿数の分散を素性から除いた手法。

4.2 実験方法

評価尺度は、コミュニケーションスキルの段階（高い・低い・それ以外）ごとの精度・再現率・F値、および、スキルを横断したF値のマクロ平均と評価データ全体に対する正解率を使用する。評価方法は、Leave-One-Out法を採用する。実験データとして、コミュニケーションスキルの高いユーザ、低いユーザ、どちらでもないユーザの3つの正解ラベルを付与した各147ユーザを対象とした。また、ソーシャルメディア上でイベントの参加者を募る状況を想定して、以下の3つのタスクを設定した。

タスク1：サッカーイベントを開催

タスク2：プログラミング勉強会を開催

タスク3：お店のアルバイトの募集

タスク評価は、(1) それぞれのタスクについて、ユーザの過去のツイートからそのタスクに興味があると判断できるユーザを抽出し、(2) 提案手法による分類の結果から、タスクに興味があり、かつ、コミュニケーションスキルの高いとされるユーザを、タスクに適性のあるユーザと判定し、分類実験と同じ尺度で評価する[¶]。

4.3 実験結果

提案手法、ベースライン、比較手法によるコミュニケーションスキルの分類の結果を表1に示す。また、タスク別での評価を表2に示す。

表1: コミュニケーションスキルの分類実験の評価

手法	スキル	精度	再現率	F 値		正解率
				平均		
提案手法	高い	0.49	0.67	0.57	0.50 ^{**}	0.51
	低い	0.51	0.48	0.50		
	その他	0.52	0.38	0.44		
ベースライン	高い	0.33	0.30	0.31	0.31	0.31
	低い	0.31	0.29	0.30		
	その他	0.30	0.34	0.32		
比較手法	高い	0.41	0.51	0.46	0.38	0.39
	低い	0.39	0.36	0.37		
	その他	0.36	0.29	0.32		

^{**} ベースラインと比較手法に対して t-検定で有意差あり

4.4 考察

分類実験の結果から、提案手法は、ベースライン、比較手法のそれぞれに対して、コミュニケーションスキルの高いユーザ、低いユーザの分類の精度が高いことが分かる。なお、提案手法のF値の平均は、他の2つの手法に対して、t-検定（両側検定、有意水準5%）に

[¶]ただし、精度と再現率は紙数の制約で割愛する。

表2: 提案手法に対する3つのタスク別の評価

タスク	ユーザ数	スキル	F 値		正解率
			平均		
タスク1	240	高い 低い その他	0.51 0.37 0.31	0.40	0.41
タスク2	83	高い 低い その他	0.38 0.48 0.28	0.38	0.38
タスク3	87	高い 低い その他	0.35 0.33 0.47	0.39	0.39

において有意差を確認できる。結果について調査したところ、コミュニケーションスキルの低いユーザの分散が他のユーザよりも大きくなっていった。このことから、コミュニケーションスキルの低いユーザは、ツイートの投稿時間帯にばらつきがあると分かる。また、投稿時間帯についての投稿数の分散が、分類に有効なことを示唆している。タスク別の評価では、コミュニケーションスキルの高いユーザのF値がタスク1で最大となり、タスク3で最小となった。タスク1でコミュニケーションスキルが高いと分類されたユーザを調べると、ほぼ全ての時間帯でメンションを投稿するユーザが多く存在した。これによりツイートの投稿時間帯のばらつきが減り、コミュニケーションスキルの高いユーザの特徴が得られたと考える。タスク3では、コミュニケーションスキルが高い、または低いユーザの多くが、その他のユーザに分類されていた。このような場合には、それぞれのクラスに分類されるようにコストを調整する必要がある。

5 おわりに

本研究では、投稿パターンを用いたコミュニケーションスキルの分類の有効性が確認できた。また、提案手法を用いて、設定したタスク別の評価を行った。今後は、コミュニケーションスキルの低いユーザのツイートには、人を中傷する言葉や、性的な意味の言葉が現れることから、コミュニケーションスキルごとの特徴語を素性とするものの効果について検討する予定である。

謝辞

本研究の一部は、科学研究費補助金基盤研究B（課題番号16H02913）の助成を受けて遂行された

参考文献

- [1] 田中成典, 中村健二, 寺口敏生, 中本聖也, 加藤諒. マイクロブログから抽出したユーザの習慣に基づく行動推定に関する研究. 情報処理学会論文誌データベース, Vol. 6, No. 3, pp. 73-89, 2013.
- [2] Naoto Takeda and Yohei Seki. Twitter User Classification with Posting Locations. In *Digital Libraries: Knowledge, Information, and Data in an Open Access Society*, Vol. 10075 of *Lecture Notes in Computer Science*, pp. 297-310, Tsukuba, Japan, 2016. Springer.
- [3] 加藤諒, 中村健二, 山本雄平, 田中成典, 坂本一磨. マイクロブログにおけるユーザの属性と習慣行動の推定に関する研究. 情報処理学会論文誌, Vol. 57, No. 5, pp. 1424-1435, 2016.
- [4] 伊集竜之, 遠藤聡志, 山田孝治. 投稿時間のクラスター分析によるTwitterユーザの年齢層推定. 人工知能学会全国大会論文集, Vol. 29, , 1M4-5, 2015.