

マイクロブログにおけるユーザ属性を考慮した 情報伝播の解析に関する研究

寺口敏生[†] 田中成典[‡] 中村健二^{‡†} 中本聖也[†] 加藤諒[‡]

関西大学大学院総合情報学研究科[†] 関西大学総合情報学部[‡] 立命館大学情報理工学部^{‡†}

1. はじめに

インターネットを介した風評被害の拡大が問題となっている。これは、リアルタイム情報が高速に伝播する性質をもつマイクロブログの普及[1]が一因であると考えられる。そこで、著者らは、マイクロブログにおける風評被害の拡大を防止するための手法について研究を行っている。風評被害の拡大の防止には、風評の把握と風評に反応するユーザ層を把握し、各ユーザに特化した情報戦略を実施することが効果的であると考えられる。マイクロブログから風評を抽出するための研究として、トピック抽出手法[2][3]が提案されており、一定の精度で風評に関わる話題を抽出可能であると考えられる。また、風評に反応するユーザ層の把握には、ブログの投稿内容から属性を推定する研究[4][5]が適用できると考えられる。これらの手法に加えて、リアルタイムに情報が発信されるマイクロブログの特性に基づき、投稿時間の傾向からユーザの生活サイクルを考慮し、ユーザの属性を推定する手法を考案した。本研究では、以上の手法を組み合わせ、風評に反応したユーザの属性を推定する。そして、入力した風評に対し、どのようなユーザ層が反応しているかなどの情報伝播状況を提示するシステムを開発する。

2. 研究の概要

本研究では、マイクロブログの投稿内容を解析することで、風評に反応するユーザ層を特定し、マイクロブログ上での風評の情報伝播状況を提示するシステムを開発する。本システムの

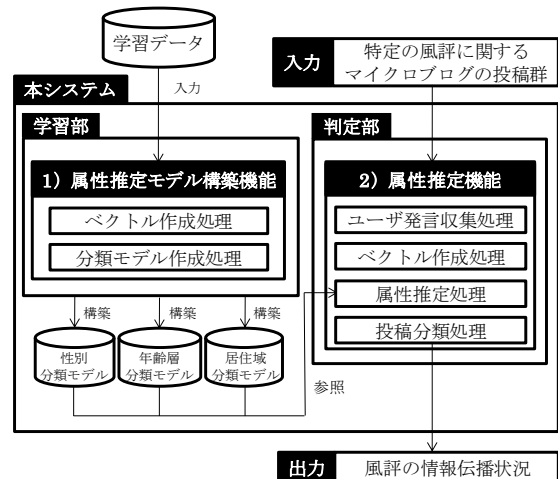


図1 本システムの概要

概要を図1に示す。入力データは、特定の風評に関するマイクロブログの投稿群とし、出力データは、風評の情報伝播状況とする。本システムは、1) 属性推定モデル構築機能、2) 属性推定機能により構成される。

2.1 属性推定モデル構築機能

本機能では、ユーザの性別、年齢層と居住域を推定するための SVM (Support Vector Machine) の分類モデルを構築する。各モデルに用いたカテゴリは、既存研究[4]に倣い、性別、年齢層 (10代以下, 20代, 30代, 40代以上)、居住域 (北海道・東北, 関東, 中部, 近畿, 中国・四国, 九州・沖縄) とした。ベクトル作成処理では、まず、学習データからノイズを除外した投稿群に対し形態素解析を行い、形態素ごとの χ^2 値を算出する。次に、事前実験から、性別の分類には χ^2 値の上位 512 件、年齢層と居住域には上位 1024 件の形態素を属性推定に有用な語として選択する。そして、それらの語を素性とする特徴ベクトルを各ユーザの投稿群から作成する。なお、年齢層に関しては、各年齢層の生活時間の傾向を考慮するため、利用時間ごとの投稿数を特徴ベクトルに追加する。分類モデル作成処理では、作成したユーザ毎の特徴ベクトルを SVM に入力し、各分類モデルを構築する。

Research for Analyzing Information Diffusion Considering User Attribute of Microblog

[†] Toshio Teraguchi, Seiya Nakamoto

Graduate School of Informatics, Kansai University, 2-1-1 Ryozenji-cho, Takatsuki City, Osaka 569-1095, Japan

[‡] Shigenori Tanaka, Ryo Kato

Faculty of Informatics, Kansai University, 2-1-1 Ryozenji-cho, Takatsuki City, Osaka 569-1095, Japan

^{‡†} Kenji Nakamura

College of Information Science and Engineering, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu City, Shiga 525-8577, Japan

2. 2 属性推定機能

本機能では、特定の風評に関するマイクロブログの投稿群から、ユーザの属性を推定する。ユーザ発言収集処理では、入力された特定の風評に関する投稿群を解析し、それらの投稿を行ったマイクロブログユーザを特定する。そして、各マイクロブログユーザの過去の投稿ログを取得する。ベクトル作成処理では、属性推定モデル構築機能のベクトル作成機能と同様にして、各ユーザの特徴ベクトルを作成する。属性推定処理では、各ユーザの特徴ベクトルとユーザ属性の分類モデルを用いて、SVMにてマイクロブログユーザの属性を推定する。投稿分類処理では、属性推定処理にて推定した結果を用いて、Google Maps上に各ユーザをマッピングし、入力した風評の情報伝播状況を提示する。

3. 実証実験と考察

本システムの有用性を実証するため、マイクロブログのTwitterを解析対象とし、マイクロブログユーザの属性を推定する実験と、特定の風評に関する情報伝播状況を可視化して分析する。

3. 1 属性推定の実証実験

本実験では、マイクロブログユーザの属性を正しく推定可能であるかを検証するため、属性毎の推定精度を評価した。なお、本実験では、各属性が判明しているTwitterのユーザのうち性別毎に100人、年齢層毎に50人、居住域毎に50人の投稿群を実験データとして用いる。

属性推定の実験結果を表1に示す。結果から、提案手法がブログを対象とした既存手法と同等の精度で属性が推定できていることがわかる。このことから、マイクロブログでも同様に属性推定が可能であることが確認できた。年齢層に関しては、特に20代の推定精度が低かった。これは、20代前半の大学生が「授業」や「学校」などの単語を頻繁に使用するため、10代と誤判定されたと考えられる。また、利用時間ごとの投稿数を特徴ベクトルに追加したことで、20代以下と30代以上の判定精度が向上した。

3. 2 情報伝播状況の可視化

風評被害が拡大した情報伝播状況を可視化した結果を図2に示す。実際にマイクロブログを介して風評被害が拡大した「福島県の農家が熊本産として米を産地偽装」というトピックについて発信しているマイクロブログユーザをクローリングして取得し、情報伝播状況を分析した。

表1 属性推定の結果

	性別	年齢層	居住域
適合率	0.82	0.61	0.69
再現率	0.82	0.50	0.65
F値	0.82	0.55	0.67



図2 情報伝播状況を可視化した結果

結果として、北海道・東北のユーザには風評の拡大を防止させるような投稿が多く見られた。しかし、その他の居住域のユーザには、図2のような風評被害を拡大させる投稿が各性別や各居住域に関わらず伝播していることが分かった。そのため、現状のままでは、風評被害がさらに拡大することが予測されるため、全国で一様に配信できるマスメディアを活用した早急な対策をとることが必要であると考えられる。

4. おわりに

本研究では、マイクロブログの投稿内容を解析することで、風評に反応するユーザ層を特定し、マイクロブログ上での風評の情報伝播状況を提示するシステムを開発した。そして、分析の結果から、特定の風評における情報伝播状況を把握できたことから、本システムの有用性を実証した。今後は、投稿時間傾向のようなマイクロブログ特有の特性を追加して属性を推定することで、より精細な情報伝播の解析を目指す。

参考文献

- [1] 総務省：平成23年度版情報通信白書，ぎょうせい，2011.
- [2] Du, Y., He, Y., Tian, Y. Chen, Q. and Lin, L. : Microblog Bursty Topic Detection based on User Relationship, *IEEE Joint International Information Technology and Artificial Intelligence Conference*, IEEE, pp.260-263, 2011.
- [3] Yamaguchi, Y., Amagasa, T. and Kitagawa, H. : Tag-based User Topic Discovery using Twitter Lists, *International Conference on Advances in Social Networks Analysis and Mining*, IEEE, pp.13-20, 2011.
- [4] 大倉務，清水伸幸，中川祐志：スケラブルで汎用的なブログ著者属性推定手法，自然言語処理研究会報告，情報処理学会，Vol.2007，No.94，pp.1-5，2007.
- [5] Ikeda, D., Takamura, H. and Okumura, M. : Semi-Supervised Learning for Blog Classification, *Proceedings of the 23th AAAI Conference on Artificial Intelligence*, AAAI, pp.1156-1161, 2008.