投稿に対するコメントとユーザプロファイルを用いた SNS への投稿の信憑性の推定手法の提案

彭　宇軒†　　前田　亮‡

†立命館大学情報理工学研究科　〒525-8577 滋賀県草津市野路東 1-1-1
‡立命館大学情報理工学部　〒525-8577 滋賀県草津市野路東 1-1-1
E-mail:　†gr0369pi@ed.ritsumei.ac.jp,　‡amaeda@is.ritsumei.ac.jp

あらまし　近年，SNS（ソーシャル・ネットワーキング・サービス）アプリの普及によりユーザが増加し，誰でも手軽に情報を獲得したり発信したりできる一方，その匿名性ゆえに，信憑性の低い情報が氾濫しやすいという問題が目立っている．信憑性が低い情報に対しては，時間が経つにつれてコメントとして疑問や反対の声が多くなる傾向がある．また，信用度の低いユーザの投稿は信憑性が低いと考えられる．そこで，本研究では，SNS への投稿とそれに対するコメントを収集し，コメント内容の感情傾向（支持，反対）を明らかにする．それとともに，ユーザのプロファイル情報に基づくユーザの信用度を評価することで，SNS への投稿の信憑性を推定する手法を提案する．

キーワード　SNS，情報信憑性，感情分析

1. はじめに

SNS（ソーシャル・ネットワーキング・サービス）の発展により，情報が広範囲かつ高速に伝播することが可能になっている．しかし，誰でも容易に利用できることで，情報の質を保証できず，大量の信憑性の低い情報も SNS を通して広まっている．特に社会的に重要な事象（自然災害，事件，経済危機など）に関連する偽情報は，伝播の速さはより一層高まり，それを信じる SNS 利用者にとって非常に悪い影響を与える可能性もある．したがって，SNS 投稿の信憑性を推定することは重要である．

SNS に投稿された情報の信憑性の判断は困難であるが，時間が経てば，投稿は多くのユーザに見られ，コメントされる．そのため，信憑性が低い情報に対しては，時間が経つにつれてコメントの中で疑問や反対の声が多くなる傾向がある．また，人の信用度はその人の発言の信頼性に繋がると考えられることから，SNS において信用度が低いユーザの投稿は信憑性が低いと思われる．

そこで，本稿では，SNS の投稿に対するコメントとユーザのプロファイルを収集し，コメントの感情傾向とユーザのプロファイルの完成度を活用した投稿の信憑性のすいすい手法を提案する．

2. 関連研究

Castillo ら[1]は，世界的に普及している SNS である Twitter の投稿の信憑性を自動的に評価する技術について述べている．彼らは，ユーザの投稿と転載行為，または投稿内容などの特徴を抽出して，決定木によりトレンドに関する Twitter の投稿が信頼できるかとうかを評価している．

Suzuki [2]は，信頼性が高い投稿内容が転載された時は，元の内容の保留率が高いのに対して，信頼性が低い投稿内容が転載された時は，転載者の意見や観点が付け加えられやすいことに注目し，転載投稿内容の保留率による信憑性推測の手法を提案した．

Ma ら[3]は Twitter と Weibo（中国版 Twitter）の投稿とそれに対するコメントを収集し，コメントに含まれている“not true”, “unconfirmed”などの単語をヒントにして，RNN（再帰的ニューラルネットワーク）を用いて投稿内容の信憑性推測を行った．

本稿は，Ma らが公開した Weibo のデータセット[4]を用いて，投稿に対するコメントの感情傾向とユーザの信用度を評価することで，投稿内容の信憑性の推測手法を提案する．

3. 提案手法

3.1 概要

図 1 に提案手法の概要を示す．

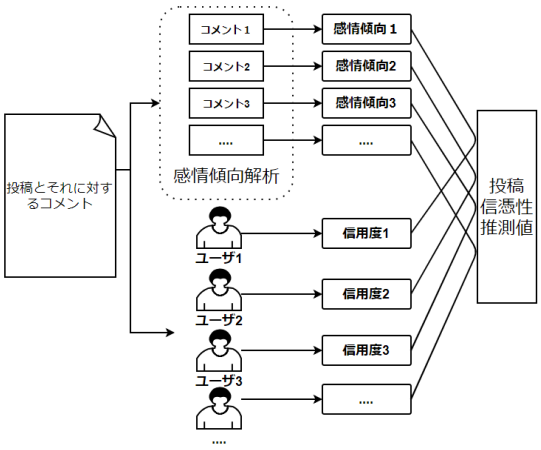

図 1　提案手法の概要

各ユーザの信用度 U とそのユーザの感情傾向値 P を求め，式(1)を用いて投稿の信憑性の推定値 S を算出する．式の n はコメントユーザの人数を表している．

$$S = \frac{\sum_{i=1}^{n} P_i U_i}{n} \quad (1)$$

3.2 感情傾向解析

今回利用するデータセットは中国の SNS である Weibo から収集したため，中国語を解析できるツール SnowNLP[1]を使用する．SnowNLP では，ナイーブベイズにより，式(2)を用いて中国語コンテンツの感情傾向を算出することができる．

$$P(C|F_1) = \frac{P(CF_1)}{P(F_1)} = \frac{P(C)\cdot P(F_1|C)}{P(F_1)} \quad (2)$$

式(2)は，カテゴリーC（ポジティブまたネガティブ）の中で特徴文書 F_1 の出現率 $P(C|F_1)$を計算する．最後の結果の値が 1 に近いほどコンテンツがよりポジティブな傾向（支持）になり，0 に近いほどコンテンツがよりネガティブな傾向（反対）になる．

3.3 ユーザ信用度

ユーザの信用度 U(n)はユーザのプロファイルに基づいて式(3)を用いて計算を行う．

$$U(n) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (3)$$

x はユーザのプロファイルの各項目の存在状況を表し，存在している場合は x= 1，存在していない場合は x= 0 とする．この結果の値が 1 に近いほど信用度がより高くなり，0 に近いほど信用度がより低くなる．

本稿で用いるプロファイル項目は，自己紹介，プロフィール画像，出身地，フォロー，フォロワー，友達（相互フォロー），個人認証，投稿件数とする．

4. 評価実験

4.1 データセット

Ma らが公開した Weibo のデータセットは，4,664 件の話題の投稿とそれに対するコメントを収集している．その中には，2,351 件の事実の投稿と 2313 件の偽情報の投稿が含まれている．詳細を表 1 に示す．

表 1 データセットの構成

ユーザ数	2,746,818
話題の投稿数	4,664
コメント合計	3,800,992
事実	2,351
偽情報	2,313

4.2 実験と考察

3 章で述べた提案手法を用いた評価実験を行う．まず，全ての事実の投稿の信憑性の推定値を計算し，得られた結果からさらに平均値を求める．その平均値を信憑性推定の基準にする．2,351 件の事実投稿の信憑性推定値の平均は 0.5817 であった．そこで，この 0.5817 を信憑性判定の基準とする．

次に，データセットにある 2,313 件の偽情報の投稿の信憑性推定値を計算する．図 2（横軸は投稿の番号，縦軸は推定値）は偽情報の信憑性推定値の分布状況を表している．

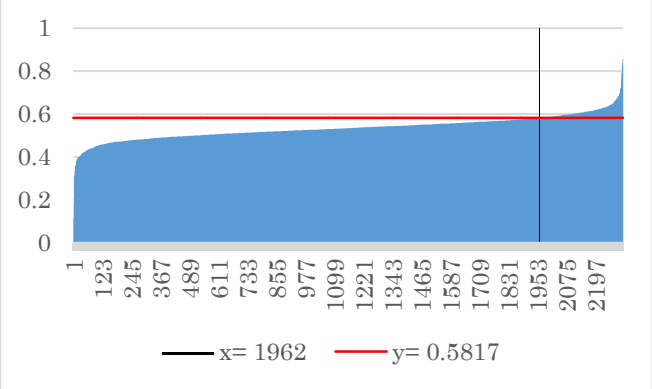

図 2　偽情報の推定値の分布

図 2 に示すように，判定基準を 0.5817 にした場合，2,313 件の偽情報のうち，1,962 件(約全体の 85%)が信憑性が低いと推定することができる．

事実情報の平均推定値を使用することで，データセット中の多くの偽情報が判別できることがわかった．しかし，その判定基準は事実の投稿の平均値であり，事実と偽情報が混在している SNS の環境では，偽情報を検出するにつれて，一部の推定値が平均値を満たさない真実の投稿も誤検出される．

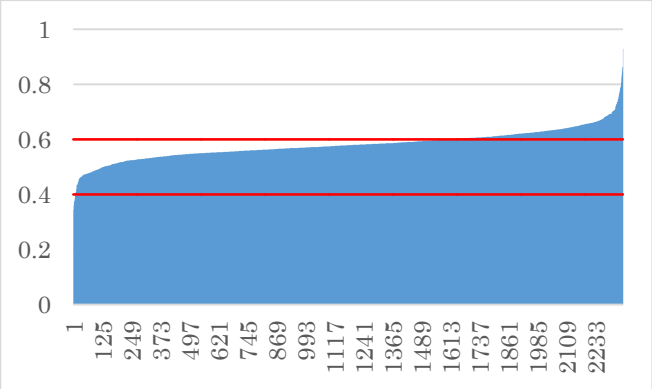

図 3　事実の推定値の分布

その原因は，図 3（横軸は投稿の番号，縦軸は推定値）に示したように，同じ範囲（0.4～0.6）に事実と偽情報の信憑性推定値のどちらも多く存在しており（事実は約 68%，偽情報は約 89%），推定値がこの範囲だと信憑性の推測は困難になる．そのような状況を避け

[1] https://github.com/isnowfy/snownlp

るため，ユーザの信用度計算と推定値計算において，より適当な重み付けを今後の課題として検討する必要がある．

5. おわりに

本研究では，SNS 投稿に対するコメントの感情傾向を解析し，ユーザのプロファイルに基づく信用度評価を加えて投稿内容の信憑性を推定する手法を提案した．今後の研究課題として，事実と偽情報の推定精度を向上するため，感情傾向の解析において，SNS で使われる略語，流行語など特有の表現を収集し，感情解析の訓練データに追加する．また，ユーザ信用度の計算において,プロファイル各項目の存在状況だけではなく，各項目の内容も詳しく解析し，重み付けの方法をさらに検討する．

本稿では中国語を主体とする Weibo のデータを使用して提案手法の実験を行ったが，今後は他の SNS および言語のデータも使用して提案手法の適用可能性を考察する．

参 考 文 献

[1] Castillo C, Mendoza M, Information credibility on twitter, Proceedings of the 20th International Conference on WWW. New York: ACM, pp.675－684, 2011.

[2] Suzuki Y, A credibility assessment for message streams on microblogs, In Proc. 3PGCIC, pp.527-530, 2010.

[3] Ma J, Gao W, Mitra P, et al. Detecting Rumors from Microblogs with Recurrent Neural Networks. Twenty-Fifth International Joint Conference on Artificial Intelligence, pp.3818-3824, 2016.

[4] Weibo データセット
https://sites.google.com/site/iswgao/ (参照 2019-1-9)