







### 5. 標準化判別係数による考察

訂正情報の判別における特徴量は、先行研究の特徴量と新特徴量をすべて統合したものが効果的であるという結果になった。そこで、8種類の特徴量それぞれの判別への影響について検討する。ここでは標準化判別係数に着目して、影響の度合いやそれぞれの特徴量の役割について明らかにする。その結果を表 5に示す。

表 5 各特徴量の標準化判別係数

各特徴量	標準化判別係数
リプライの有無	-0.0188
いいねの有無	0.0100
引用RTの有無	0.3699
RTの有無	0.1905
URLの有無	0.4098
形態素数	-0.1169
訂正特徴語の異なり語数	-0.6315
非訂正特徴語の異なり語数	0.3794

この結果の標準化判別係数における絶対値から読み取れるように、引用RTの有無、URLの有無、各特徴語の異なり語数が訂正情報へ大きく影響し重要であることがわかる。また、訂正情報群の重心は負の値、非訂正情報群の重心は正の値となっているため、標準化判別係数が相対的に大きな負の値となっている訂正特徴語の異なり語数は、語数が多いほど訂正情報と判別される可能性が高くなることが示されている。一方、相対的に大きな正の値の係数に着目すると、引用RTの有無、URLの有無、非訂正特徴語の異なり語数が含まれており、それぞれの有無の割合および非訂正特徴語の異なり語数が多いツイートは、非訂正情報と判別される可能性が高くなることが同じく示されている。

以上のことから、先行研究と自分の特徴量を合わせた特徴量が、訂正情報を判別する上でもっとも有効であると結論づけられる。

### 6. むすび

本研究では、Twitterにおける新型コロナウイルス関連の、7種類の流言に関するツイートの収集を行い、訂正情報群と非訂正情報群それぞれのツイートに含まれる語に着目して特徴量を構成した。また、それらの特徴量を用いた場合の訂正情報か否かの判別精度について検証した。その結果、語の出現頻度を考慮していない従来の特徴量よりも、本研究の特徴量を用いた場合のほうが高い精度を実現できることを確認できた。また、先行研究の特徴量と自分の特徴量を組み合わせることで、さらに精度が向上する

ことも判明した。このように、訂正情報か否かの判別精度が向上することで流言検出力も高まり、最終的な流言拡散防止につながると予想される。

課題としては、非訂正情報を訂正情報と誤判別してしまう適合率の改善が挙げられる。誤判別の原因としては、非訂正情報が流言を発信しているものから疑問視しているものまでと表現が多様であることや、今回の分析において訂正情報と比較して非訂正特徴語の異なり語数が十分に得られなかったことなどが考えられる。よって、今後は分析対象のデータ数を増やすことで、様々な種類の非訂正情報に共通して出現する語の傾向を明らかにし、判別精度の向上を目指す。

### 参考文献

- [1] 2020 年度 SNS 利用動向に関する調査, ICT 総研, <https://ictr.co.jp/report/20200729.html/> (参照 2022-02-07)
- [2] 鳥海不二夫ほか, ”ソーシャルメディアを用いたデマ判定システムの判定制度評価”, デジタルプラクティス, 3 巻 3 号 pp-201-208, 2012.
- [3] 西村涼太ほか, ”情報の信頼性の関心を高める流言注意喚起ボットの開発”, ワークショップ 2020, pp-43-50, 2020.
- [4] 梅本美月ほか, ”Web ページに含まれる流行情報への気づきを与える提示方法の検討”, 情報処理学会論文誌, 62 巻 1 号 pp-183-192, 2021.
- [5] 池田圭祐ほか, ”ロコミに着目した情報拡散モデルの提案及びデマ情報拡散抑制手法の検証”, 情報処理学会論文誌数理モデル化と応用 (TOM), 11 巻 1 号, pp-21-36, 2018.
- [6] 宮部真衣ほか, ”人間による訂正情報に着目した流言防止拡散サービスの構築”, 情報処理学会論文誌, 55 巻 1 号, pp-563-573, 2014.
- [7] 末吉美喜, テキストマイニング入門: Excel と KH Coder で分かるデータ分析, オーム社, 2019.
- [8] 小林雄一郎, ことばのデータサイエンス, 朝倉書店, 2020.
- [9] 金明哲, テキストアナリティクスの基礎と実践, 岩波書店, 2021.
- [10] 高木章光, 鈴木英太, 最新データサイエンスがよ〜くわかる本, 秀和システム, 2019.
- [11] 菅民郎, 例題と Excel 演習で学ぶ多変量解析一回帰分析・判別分析・コンジョイント分析編一, オーム社, 2016.
- [12] 菅民郎, Excel で学ぶ多変量解析入門, オーム社, 2013.