

平成 31 年度卒業論文

画像付きフェイクニュースとジョークニュースの
検出・分類に向けた機械学習モデルの検討

電気通信大学 情報理工学部 総合情報学科
メディア情報学コース

学籍番号 : 1510151

氏名 : 柳 裕太

主任指導教員 : 田原 康之 准教授

指導教員 : 大須賀 昭彦 教授

指導教員 : 清 雄一 准教授

提出年月日 : 平成 31 年 2 月 8 日 (金)

概要

SNS の発展によりあらゆる情報入手が容易になった反面，人を欺くために故意に作成された虚偽の情報であるフェイクニュースが社会問題になっている．特に画像と併せて発信されたものは，テキストのみならず画像と併せた分析アプローチが有効である．虚偽の情報としては，もう 1 つジョークニュースというものもある．これは人を欺くためではなく，社会風刺や皮肉のために作られた情報という特徴がある．しばしばこの 2 カテゴリーが混同され，ジョークニュースが批判に晒されることがあることも問題となっている．

既にテキスト・画像をニューラルネットワークの一種である CNN によって分析して真偽を判定する自動判別モデルが提案されている．実際に真実・フェイクとのカテゴリ分類において優秀な成績を収めているものの，ジョークとしての嘘情報と人を欺くための嘘情報が区別されていない．

本研究では，正しい情報・ジョークニュース・フェイクニュースの 3 カテゴリーを分類することで，より画像つきフェイクニュースの検出精度を向上させることを目指した．

実際に SNS から収集した画像つきのデータセットを対象にカテゴリ分類を行った結果，3 カテゴリーでもマクロ F 値が約 0.93 と良好な結果を示した．

目次

| | |
|-----------------------|---|
| 概要 | i |
| 第 1 章 序論 | 1 |
| 1.1 背景 | 1 |
| 1.2 先行研究 | 1 |
| 1.3 課題 | 2 |
| 第 2 章 提案手法 | 4 |
| 2.1 モデル概観 | 4 |
| 2.2 特徴生成器 | 4 |
| 2.3 ニュース分類器 | 4 |
| 第 3 章 評価実験 | 6 |
| 3.1 データセット | 6 |
| 3.2 比較対象手法 | 6 |
| 3.3 実験条件 | 6 |
| 3.4 実験結果 | 6 |
| 第 4 章 評価 | 7 |
| 4.1 考察 | 7 |
| 4.2 課題 | 7 |
| 第 5 章 おわりに | 8 |
| 5.1 本論文のまとめ | 8 |
| 5.2 今後の展望 | 8 |
| 謝辞 | 9 |

第 1 章

序論

1.1 背景

昨今の SNS の普及により，誰もが情報を発信・収集できるようになった．特に最近ではテキストのみならず，画像や動画と併せて情報の発信が可能である．一般論として，テキスト単体と比べて画像や動画と併せて発信されたマルチメディア情報の方が多くの注目を得やすい．逆にこれを利用して，故意に情報を捏造して発信することによって人々を誤った方向へ扇動するフェイクニュースも存在する．フェイクニュースが広まると、大規模なマイナスの影響が出る可能性があり、場合によっては重要な公共の出来事に影響を及ぼしたり、操作したりすることさえある．例えば 2016 年の米国大統領選では，2 名の候補者を支持させるためのフェイクニュースが多く拡散され，とりわけ Facebook 上では 3700 万回以上共有された [1]．

虚偽の情報ながら，扇動ではなく皮肉や風刺を込めたジョークニュースも存在する．有名な発信メディアとしては，英語では the Onion，日本語では虚構新聞が該当する．あくまで扇動ではなく笑いを提供するためのものであり，多くの場合それは批判的にはなりにくい．しかしながら，ジョークニュースはフェイクニュースと同じく限りなく真実を模した形式をとるため，同じく SNS 上で拡散されやすい傾向にある．

当研究では，扇動のために故意に情報を捏造して発信された情報をフェイクニュース，事実を発信した情報を正しいニュース，そして風刺や皮肉を込めて発信された情報をジョークニュースとして定義する．

1.2 先行研究

フェイクニュースに限らず，風評や web ページの信憑性を評価するモデルの構築の研究は数多く行われている．例えば，福島らの研究 [10] では，web ページの体裁から信頼性を評価するモデルが提案されている．また，機械学習による分類が非常に盛んに行われている．なかでも Granik らの研究 [3] や Gilda の研究 [2]，そして松尾の研究 [9] により，単語埋め込みと

ナイーブベイズ分類器や SVM, 決定木といった教師あり学習を組み合わせることによって, フェイクニュースや流言を分類するタスクで優秀な分類成果を挙げることが報告されている. ほかにも Wu らの研究 [8] によると, SNS 上で拡散された情報に対して, “誰が・どのような経緯で拡散したか” という情報から信憑性を判断するモデルも提案されている. Rubin らの研究 [6] によれば, 正しいニュース・ジョークニュースの分類にもこのアプローチが有効であることが示されている. 正しいニュース・フェイクニュース・ジョークニュースの 3 カテゴリ分類においても研究が行われている. 特に Horne と Sibel の研究 [4] によると, フェイクニュースは正しいニュースよりジョークニュースに近い性質をもち, 真実に近い形式をとるほど高い説得力をもつことが示されている.

上記の機械学習を使った研究では, いずれもテキストのみの情報を対象としている. 別の対象として, テキスト・画像を併せた情報を分類する機械学習モデルの検討も数多く行われている. 大まかな形としては, まずテキスト・画像を何らかの方法でベクトル化する. その後 2 種のベクトルを結合し, 真偽判定を行うモデルに渡す形をとっている. 例えば Jin らの研究 [5] では, テキストでは LSTM, 画像では VGG-19 を使用してベクトル化しており, 更に Attention とソーシャルコンテキスト (ハッシュタグ, URL 等) によって更に高精度な分類を行うモデルが提案されている. また Wang らの研究 [7] では, EANN というモデルが提案されている. これは画像のベクトル化においては同じく VGG-19 を使用しているが, テキストではテキスト CNN を使用している.

1.3 課題

上記の EANN モデルのような画像・テキスト双方を扱うモデルでは, 実際に真実・フェイクとのカテゴリ分類において画像単独・テキスト単独の分類に比べて優秀な成績を収めている [7]. しかしながら, あくまで “真実なのかそうでないのか” という 2 カテゴリで分類しているため, “他者を欺くための情報なのか, 皮肉・風刺を込めた情報なのか” という観点での分析がなされていない.

本研究では, 画像つきで発信された情報に対して, 正しい情報か・フェイクニュースか・ジョークニュースかを判断するモデルを構築する. このモデルを使い, 従来から画像・テキスト複合のデータセットに対して 3 カテゴリでも優秀な分類が行えることを示すことを目指す. それにより, SNS ユーザの情報収集を支援するエージェントの開発につなげることが可能となる.

上記の提案する情報分類システムを検証するために, 事前に用意されたデータセットを用いて 10 分割交差検定によって分析を行う. また上記システムの分類性能を評価するために, 画像・テキスト単独で分類を行った結果と比較することで, 提案システムが目標に適していることを示す. その結果テキスト単独でのマクロ F 値が約 0.22, 画像単独でのマクロ F 値が約 0.40 であったのに比べ, 提案モデルのマクロ F 値は約 0.93 という数値を出し, 提案モデルの

有効性が示された。

第 2 章

提案手法

2.1 モデル概観

2.2 特徴生成器

2.3 ニュース分類器

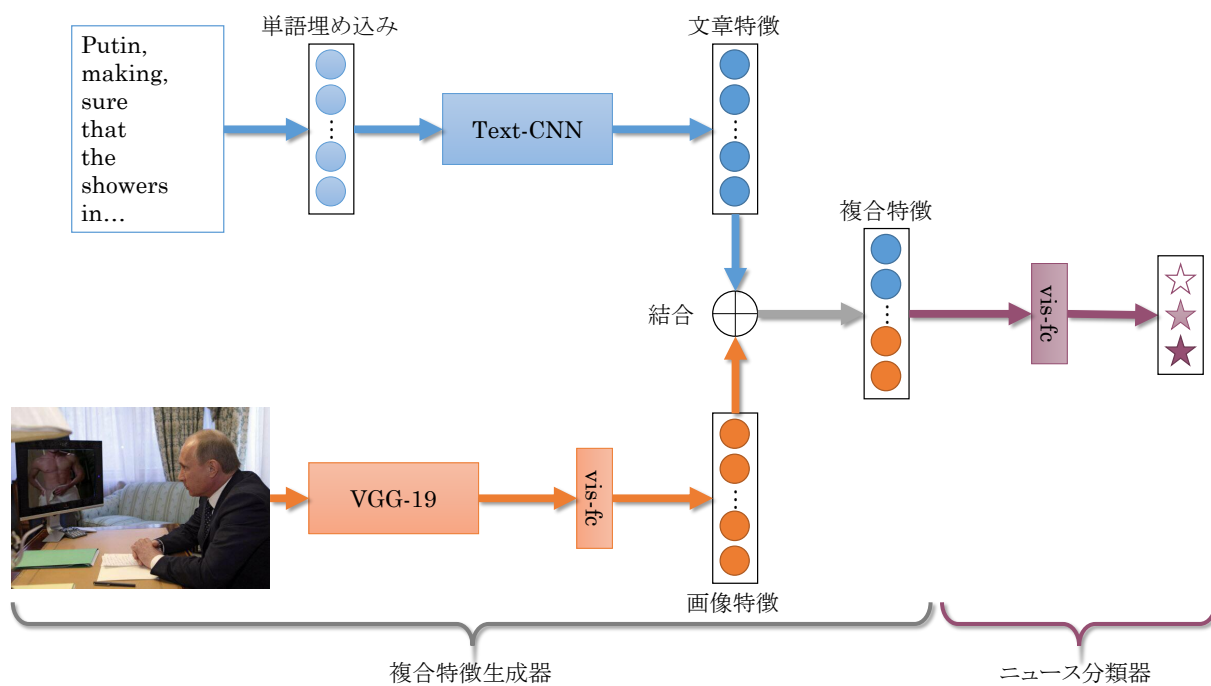


図 2.1 提案モデル図. 青色がテキスト特徴生成器, 橙色が画像特徴生成器, 紫色がニュース分類器である.

第 3 章

評価実験

3.1 データセット

3.2 比較対象手法

3.3 実験条件

3.4 実験結果

第 4 章

評価

4.1 考察

4.2 課題

第 5 章

おわりに

5.1 本論文のまとめ

5.2 今後の展望

謝辞

本研究を行うにあたり，ご多忙の中，終始適切かつ丁寧なご指導をして下さった大須賀昭彦教授，田原康之准教授，清雄一准教授に深く感謝致します．貴重な勉学の機会を与えてくださったことに深く御礼申し上げます．

参考文献

- [1] Hunt Allcott and Matthew Gentzkow. “Social media and fake news in the 2016 election”. In: *Journal of Economic Perspectives* 31.2 (2017), pp. 211–36.
- [2] S. Gilda. “Evaluating machine learning algorithms for fake news detection”. In: *2017 IEEE 15th Student Conference on Research and Development (SCORED)*. Dec. 2017, pp. 110–115. DOI: 10.1109/SCORED.2017.8305411.
- [3] M. Granik and V. Mesyura. “Fake news detection using naive Bayes classifier”. In: *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*. May 2017, pp. 900–903. DOI: 10.1109/UKRCON.2017.8100379.
- [4] Benjamin D Horne and Sibel Adali. “This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news”. In: *arXiv preprint arXiv:1703.09398* (2017).
- [5] Zhiwei Jin et al. “Multimodal fusion with recurrent neural networks for rumor detection on microblogs”. In: *Proceedings of the 2017 ACM on Multimedia Conference*. ACM. 2017, pp. 795–816.
- [6] Victoria Rubin et al. “Fake news or truth? using satirical cues to detect potentially misleading news”. In: *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*. 2016, pp. 7–17.
- [7] Yaqing Wang et al. “EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection”. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM. 2018, pp. 849–857.
- [8] Liang Wu and Huan Liu. “Tracing fake-news footprints: Characterizing social media messages by how they propagate”. In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM. 2018, pp. 637–645.
- [9] 松尾省吾. “機械学習を用いた流言の検出に関する研究”. MA thesis. 電気通信大学院, 2018.
- [10] 福島隆寛 and 内海彰. “Web ページの信頼性の自動推定”. In: *知能と情報* 19.3 (2007), pp. 239–249.