

言語学的モダリティに基づくデマの検知に関する研究

Fakenews Detection Based on Linguistic Modality

松田美慧・ネットワーク分科会・中央大学大学院

Disinformation is a recent information security threat, and the key point of countermeasures is emotional incitement. In this study, we identified the textual features of a disinformation that appeal to emotions, and verified the applicability of these characteristics to detection. Modality labels, which represent the psychological attitudes of the sender of the information, were defined and assigned to three datasets: the original text of the hoax, corrected information, and a review of a novel. We then compared and analyzed the frequency and order of appearance of the modality labels, and conducted machine learning using the modality labels. The results show the possibility that some modality labels are effective in identifying hoaxes [force] and others are not [will], [ability], or [none], based on their composition frequency and order of occurrence. In machine learning, the accuracy of simple class classification is insufficient, at around 46%, but if time-series information is utilized for learning and a certain level of accuracy is ensured, it is possible to detect hoaxes independently of the truth or falsehood of information and the topic. Future work includes increasing the number of data, automatically assigning modality labels, and adjusting the weights in the learning model.

1. 研究背景

インターネット上の不確かな情報の拡散は、近年のセキュリティ脅威である。その拡散要因は右図の7つに整理され、特に感情煽動はリテラシー教育やファクトチェックの選定および訂正情報の拡散などの対策を阻む大きな要因になっている。



2. 研究目的

本研究では、デマの感情煽動の引因となる特徴量を特定し、その特徴量が検知に応用可能かを検討する。そのためにモダリティを活用する。

モダリティとは、「話者の心理態度表現」と定義される言語学上の概念である

デマ例文：新型コロナ対策としてすぐ緑茶を飲んでください！！
新型コロナ対策としてすぐ緑茶を飲む + **ください！！** (モダリティ)
命題内容

3. 提案手法 (流れは以下5段階)

①データ収集

データ名称	平均文数	平均文字数
fake24	約5.17文	約144.96文字
news24	約4.38文	約248.54文字
review24	約7.71文	約291.04文字

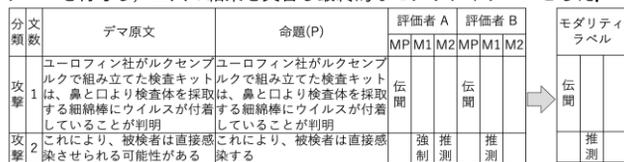
②モダリティラベルの定義

Palmerらの分類を参考に、10種類のモダリティラベル(下図赤字)を定義した。

Palmerの分類	ML	定義	例	
対事的	命題的	断言	単なる事実ではなく、揺るぎない主張を有する箇所	のだ、わけだ、ことだ、つまり
		推測	自分の意見を意見だとわかる形で強調して記述している箇所	だろう、にちがいない、はずだ、かもしれない、まい、とみられる
	証拠的	経験	文脈から自分の経験を根拠に記述している箇所	(～し) そうだ
		伝聞	当事者でなく、他者の経験を聞き伝えている箇所	らしい、ようだ、そうだ、によると
事象的	束縛的	他者に対して行動を示唆、強要する箇所	べきだ、ほうがよい、なければならない	
	誘導	他者に対して行動をお願いする箇所	てもいい、ましよう、てください、くれ	
	力動的	意志	—	つもりだ、したい
		能力	—	できる
対人的	対人	話口調に関する箇所	ね、よ、!、とても、か	
その他	なし	—	—	

③アノテーション

2人の作業者がモダリティラベルの定義に従って、下図の流れで一文に1つずつラベルを付与し、二人の結果を突合し最終的なモダリティラベルとした。



④ラベルの比較分析

モダリティラベルの構成比率と出現順序を比較した。出現頻度の比較では、文章を5区間に分けて出現頻度を比べた文章単位と文単位がある。

⑤機械学習での検知実験

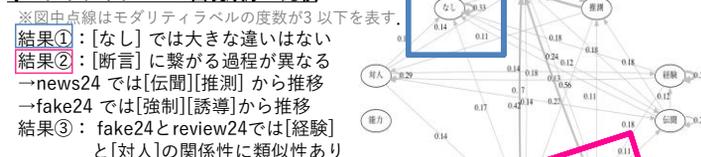
- 学習手法：教師あり機械学習3種とニューラルネットワーク1種
- 学習データ1：DATA10(全モダリティラベルの文章ごと)の構成比率)
- 学習データ2：DATA07([意志, 能力, なし]を除く7つのラベルの構成比率)
- 教師データ：データセットの属性(レビュー：0, 訂正情報：1, デマ：2)
- テストデータ：全データ72文章のうちランダムに抽出した18文章

4. 結果・考察

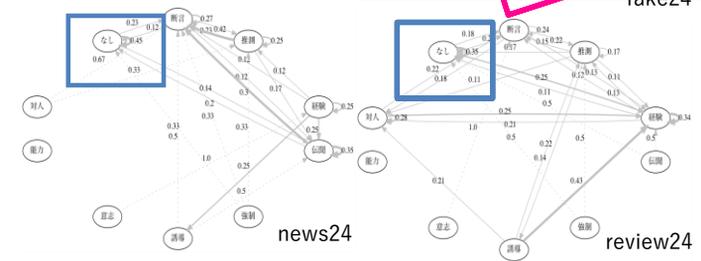
◆モダリティラベルの構成頻度の比較

- 共通点①：認識的M[断言][推測]の比率は比較的高い
- 共通点②：力動的M[意志][能力]の比率は非常に低い
- 共通点③：[なし]の比率は約2割程度

◆モダリティラベルの出現順序の比較



結果①：[なし]では大きな違いはない
結果②：[断言]に繋がる過程が異なる
→news24では[伝聞][推測]から推移
→fake24では[強制][誘導]から推移
結果③：fake24とreview24では[経験]と[対人]の関係性に類似性あり



◆機械学習による検知実験

学習回数は1000回で、評価指標(正解率, 適合率, 再現率, F値)の平均を下図に示す。図中名称の添字は、其々0はレビュー、1は訂正情報、2はデマ原文を示す。

DATA 10	正解率	適合率0	適合率1	適合率2	再現率0	再現率1	再現率2	F値0	F値1	F値2
K近傍	0.44	0.51	0.41	0.31	0.8	0.36	0.18	0.62	0.38	0.23
決定木	0.46	0.63	0.34	0.42	0.63	0.4	0.35	0.63	0.37	0.38
ランダムフォレスト	0.48	0.65	0.39	0.39	0.69	0.41	0.35	0.67	0.4	0.37
LSTM	0.56	0.85	0.42	0.76	0.44	0.83	0.41	0.57	0.56	0.53

DATA07	正解率	適合率0	適合率1	適合率2	再現率0	再現率1	再現率2	F値0	F値1	F値2
K近傍	0.46	0.52	0.42	0.34	0.82	0.39	0.16	0.64	0.4	0.22
決定木	0.46	0.63	0.34	0.43	0.65	0.41	0.34	0.64	0.37	0.38
ランダムフォレスト	0.52	0.71	0.42	0.45	0.7	0.48	0.39	0.7	0.44	0.42
LSTM	0.71	0.93	0.55	0.88	0.66	0.9	0.58	0.77	0.68	0.7

5. まとめ

本研究では、デマの感情煽動の引因となる特徴量をモダリティを用いて特定し、その特徴量が検知に応用可能かを教師あり学習を用いて検討した。その結果、モダリティラベルのうち、デマの識別に有効な[強制], 有効でない[意志][能力][なし]がある可能性を示し、検知においては時系列情報の学習により、性能向上が図れることを示した。今後の課題としてはデータ数の拡充、モダリティラベルの自動付与、学習モデルにおける重みづけの調整などがある。