

語句の関係性に着目した SNS 投稿の真偽分類の検討

牛込龍太郎^{†1} 松田健^{†2} 園田道夫^{†3} 趙晋輝^{†4}
 中央大学^{†1 †3 †4} 情報通信研究機構^{†1} 長崎県立大学^{†2}

1. はじめに

SNS(Social Network Service)はユーザー間のコミュニケーションツールのみならず、不特定多数のユーザーに対する情報発信を迅速に行うためのツールとしても利用されるため、マーケティングや災害時の情報伝達など、様々な分野で必要不可欠なツールとなっている。しかし常に膨大な数の投稿がなされる SNS において、投稿内容の真偽をリアルタイムに検証することは困難であるため、デマや誤りを含む情報が拡散されることに関して何らかの対策が必要である。本研究では、デマを含む単語の関係性についての分析結果から、SNS の 1 つである Twitter への投稿「ツイート」を用いてどの程度 false positive が発生するか検証を行った。

2. 関連研究

Twitter 上での誤った情報の流布についての研究は数多くなされている。それらの中でも小嶋ら[2]は、リツイート頻度の高いものに着目し、リツイートの反復回数やツイート文が含む語句の感情極性からツイートのデマ推定を行っている。本研究では、リツイート以外のツイートも情報拡散の一翼を担い得ると考え、それらを含めたデマ推定を行うために通常ツイート文に着目した語句間の関係性について検証を行う。

3. ベイジアンネットワーク

ベイジアンネットワーク (BN) とは確率的グラフィカルモデルの 1 つであり、グラフの各ノードが確率変数、リンクが確率変数間の確率的関係を表現する。グラフィカルモ

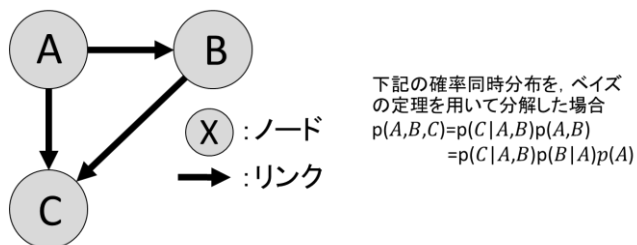


図 1 ベイジアンネットワークの例
 Figure 1 An Example of Bayesian Networks

デルへ可視化することによって、その構造から確率変数間の条件付き独立性などの知見を得ることが可能である。BN の特徴として、各リンクが向きを持っているためノード間の因果関係を表現するのに適していることが挙げられる。図 1 にベイジアンネットワークの例を示す。

4. 提案手法

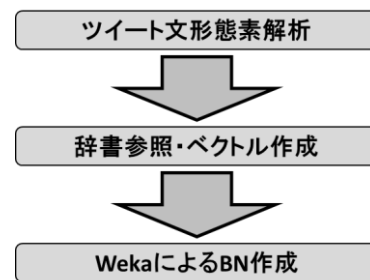


図 2 提案手法の流れ

Figure 2 Procedure of the proposed method

4.1 ツイートデータ

本研究で使用するツイートデータは[1]で使用したデマに関連するツイート 314 件と、通常ツイートデータ 8901 件である。

4.2 辞書データ

本研究の提案手法において各ツイート文が参照する辞書データは、先行研究[1]において作成したものを使用する。本データは日本語評価極性辞書[3][4]から、Twitter 上においてユーザー同士で罵倒するのに用いられ得る表現や、特定の話題に関連の深い語句を選択し、またツイートデータを形態素解析した際に出現頻度が高く、極性辞書に含まれていない表現を加えたものである。話題別に自然災害関連の話題に適用させる辞書 “disaster” 及び事件・事故関連の話題に適用させる辞書 “accident” を作成しており、各辞書には 262 語が含まれている。ツイートデータの話題の観点から、本研究では “disaster” を使用する。

4.3 ベイジアンネットワークの作成

各ツイート文に対して形態素解析を行い、辞書に含まれる各語句に対しツイート文にその語句が含まれているかどうか、またそのツイート文がデマに関連するものかどうかについて計 263 次元のベクトルを各ツイート文に対して作成する。作成したベクトルを機械学習ソフトウェア Weka に読み込ませ、BN を構築する。そして、分割数を 5 として交差検証を行い、構築した BN の精度を False Positive に着目して評価する。

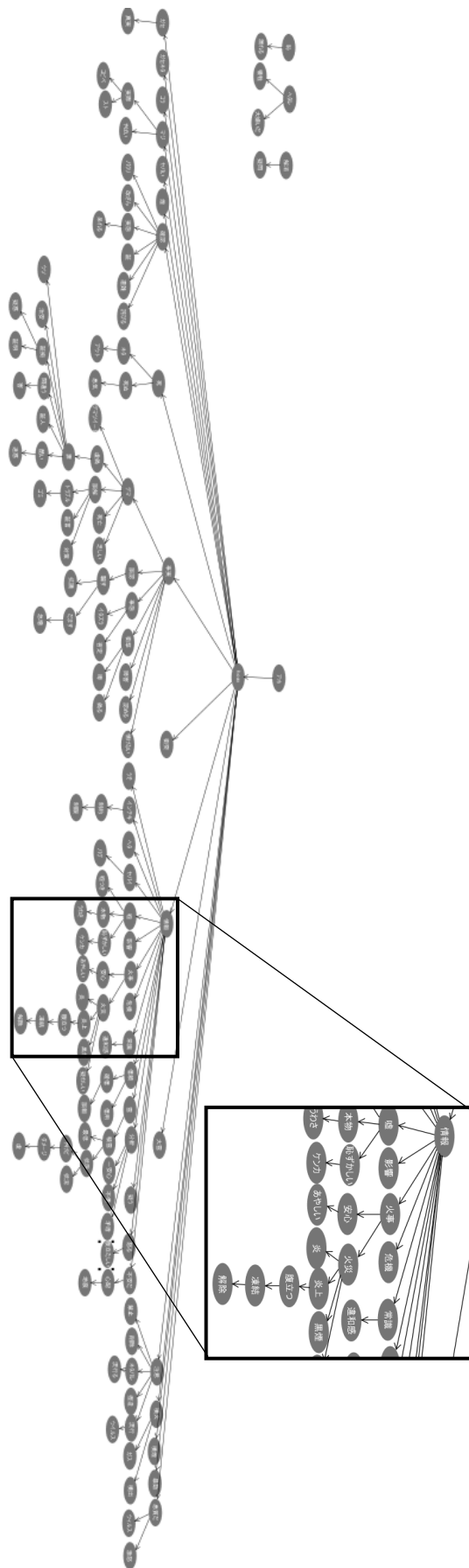


図3 ベイジアンネットワークの構築結果
Figure 3 Obtained Bayesian Network

5. 結果・考察

構築した BN を図 3 に示す. 使用したツイートデータを増加させたことにより, [1]で作成した BN よりも特にノード数, ノードの深さが増大しているほか, グラフが複数に分割された. 図 3 の黒線枠で囲んだ部分に着目すると, 「情報」, 「嘘」, 「本物」と「うわさ」では情報の伝達に関連しており, 「火事」, 「火災」, 「炎上」と「炎」では火事に関連しており, 「炎上」, 「腹立つ」, 「凍結」と「解凍」では SNS でよく用いられる, というようにそれぞれ自然言語的な意味の繋がりが BN においても構築されていることが見て取れる. 「炎上」という単語は, 物理現象としての意味と SNS 上で議論が紛糾していることの 2 種の意味で用いられるため, 2 つの繋がりに属していると考えられる.

モデルの精度として False Positive の割合, すなわち通常のツイート文をデマツイートと分類した割合は 0.016 であったが, デマツイートの件数と通常ツイートの件数の差が影響している可能性が考えられる.

6. まとめ

本研究では, SNS の 1 つである Twitter への投稿, ツイートを対象とし, 有向確率的グラフィカルモデルであるベイジアンネットワークを構築することによって, 取得したツイート文が含む語句間に存在する, 自然言語的な意味の関連性を視覚的にも得ることができた.

今後の課題としては, デマ検知への応用を目標として, 使用するデータセットを変化させた上で BN を構築し, 類似する構造の発見を行うことがあげられる.

参考文献

- [1] “ベイジアンネットワークによる SNS 上での流言の分類”, 牛込龍太郎, 松田健, 園田道夫, 趙晋輝, FIT2017.
- [2] “震災時におけるツイッターのトレンドワードと拡散情報を利用したデマ推定の一考察”, 小嶋和徳, 伊藤慶明, 石亀昌明, 鳥海不二夫, 情報処理学会第 75 回全国大会, 2012.
- [3] “日本語評価極性辞書(名詞編)”, 東山昌彦, 乾健太郎, 松本裕, “述語の選択選好性に着目した獲得”, 言語処理学会, 2008.
- [4] “日本語評価極性辞書(用言編)”, 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, “意見抽出のための評価表現の収集”, 自然言語処理, Vol.12, No.3, pp.203-222, 2005.