1L-03

多角的な視点に基づくニュース理解支援のための可視化手法の提案

森　一輝　　赤石　美奈
法政大学　情報科学部

1. はじめに

近年，SNS の利用用途は情報発信よりも情報収集の傾向が強くなってきている．[1]特に，Twitter はキーワードを検索すると，キーワードに関連した意見や記事等が容易に得られるが，得られる情報には様々な種類があり，ユーザがその中から必要な情報を得る事は困難である．

これらの問題に対し，ツイート発信者に対する興味により，価値のある情報を得られる岩木ら[2]の研究や，話題に関連する単語をユーザに提示し，役立つ情報を得られる鈴木ら[3]の研究が存在する．どちらも，有用な情報を選別し，効率よく情報を得ることを目的としている．しかしながら，選別された情報に対する理解については言及されていない．

そこで，本研究では，話題の理解を支援するために，ツイートの傾向から情報を分割し単語の連想関係としてネットワーク図に可視化する手法を提案する．多様な傾向のある情報も分割することで，傾向別のネットワーク図(以下分節図)に可視化され，全体では重要視されていない，もしくは気付かなかった情報を得られる場合がある．さらに，可視化された単語の連想関係から得られた新たなキーワードを再度システムに通して，繰り返し可視化することで，話題に対する知識を広げ，繰り返し分割することで，より深く理解することが可能になると考える．

2. 提案手法

提案手法の流れを図 1 に示す．本研究はユーザが気になる話題についての情報を収集し，分割して可視化することでユーザの話題理解を支援する．話題についての情報は Twitter から収集する．Twitter は多様なユーザによる意見が豊富であるため，様々な傾向の意見を収集するのに適していると考えられる．しかし，様々な傾向のあるツイート群を，1 つの図に可視化したとしても，必要な情報の判別が困難であり，有用な情報を得ることは難しい．そこで，本研究は，ツイートを傾向別に分節化することでそれぞれの傾向別の可視化を可能にする．可視化は，ツイート群の重要な単語をノード，重要な単語間の共起度をリンクとした単語の連想関係を示す．分節図より得られた情報から，ユーザがさらに詳細な情報を得たい場合，再度クラスタリングを行うことで情報は細分化され，より具体的な情報が可視化される．分節図より話題に関連する新たなキーワードを得られた場合，キーワードを加えて再度ツイートを収集することで，よりユーザが求める内容に近い情報が可視化される．

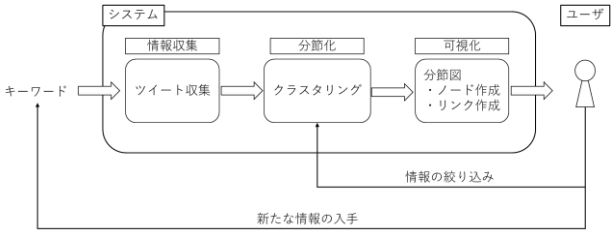

図 1:提案手法の流れ

2.1 ツイート収集

ツイートの収集は，リツイート，リプライを除いたキーワードを含むツイートを収集する．キーワードは単一の単語だけでなく，複数の単語が入力可能であり，複数の単語がキーワードである程ツイート内容を絞り込めることから，ユーザが求めるツイートが収集しやすい．収集したツイートは janome を用いて形態素解析を行い，名詞,動詞の原形,形容詞の原形,複合名詞のみを取得する．

2.2 クラスタリング

収集したツイートの内容を Doc2Vec を用いて，ベクトル化させ，K-means 法でクラスタリングを行う．キーワードが少ない場合 Doc2Vec によるベクトルの分散図は明確なクラスタに分かれず，最適なクラスタ数を決定する事が出来ない．そのため，クラスタ数はユーザが任意のクラスタ数を指定する．

Proposal of visualization method to support news understanding from multiple perspectives
Kazuki Mori and Mina Akaishi
Faculty of Computer and Information Sciences,Hosei University

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.

2.3. 可視化

分節図に表示するノードは重要度の上位 50 単語と検索キーワードとする．単語の重要度には tf-idf を用いる．ツイートは 140 字以内で構成されるため，同じ単語が繰り返されることは少ない．そのため，1 ツイートの tf の重みは少ないと考えられるため，tf は 10 ツイートをまとめたツイート群に，idf はまとめたツイート群全体に用いる．ノードの大きさはキーワードを最大値とし，それ以外のノードの大きさは，tf-idf の値に応じて 5 段階で表される．また，ノードの色は名詞を赤色，動詞を青色，形容詞を緑色，キーワードを黄色として可視化する．

分節図に表示するノード間のリンクの太さと長さはノードとなる単語の共起度の強さを表す．ノードとなる単語同士の共起度には jaccard 係数を用いる．話題の理解に必要な重要な単語の多くは主語となる名詞であると考えられるため，リンクは名詞と修飾関係にある品詞の名詞-名詞,名詞-動詞,名詞-形容詞の間に作成し，共起度に比例して太く，短く可視化する．

実際に可視化した例として”コロナ,変異種”をキーワードとした分節図の 1 例を図 2 に示す．

3. 評価実験

キーワードを”コロナ,変異種”としたツイート 10,000 件を対象に 3 クラスタに分割した分節図 3 つとクラスタ分けしない可視化図(以下全体図とする)を用いて実験を行った．

3.1 実験目的

分節図が，全体図では重要視されない，もしくは可視化されない重要な単語をどれ程提示できるかを調査する．

3.2 実験手法

手法として，被験者 8 名に分節図と全体図を見せ，重要であると考える単語を 1 分以内に選出してもらう．その結果から，全体図で重要な単語とされていないが，分節図では重要な単語とされる単語(以下見落とし語とする)と，全体図で可視化されていないが，分節図では重要な単語とされる単語(以下部分重要語とする)の数を表にまとめ，評価した．

3.3 実験結果

実験結果を表 1 に示す．表から，どの被験者も部分重要語を多く選出していることがわかる．また，見落とし語も少なくはあるが 2 名を除き，選出している．見落とし語は，どれも全体図よりもノードの大きさや，リンクの太さ，数が大きい単語であったことから，見落とし語は分節図で視覚的に重要度が増したため選出されたと考えられる．この結果から，分節図は全体図では注目されない重要な単語に着目することが可能なため，多くの重要な単語を提示可能であり，任意の話題の把握や意見分析の視点を得るための支援として有効であると考えられる．

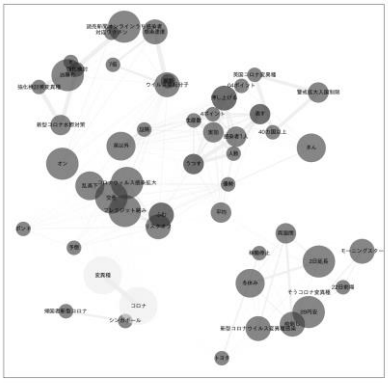

図 2:キーワード”コロナ 変異種”の分節図

表 1:見落とし語と部分重要語の数

	被験者1	被験者2	被験者3	被験者4	被験者5	被験者6	被験者7	被験者8
見落とし語	0	2	2	3	1	1	0	1
部分重要語	4	9	10	15	7	9	11	4

4. おわりに

本研究では，任意の話題の理解支援として，キーワードとなる単語を含むツイートを収集し，分割して可視化する手法を提案した．実験結果から，分節図は全体図では注目されない重要な単語を多く提示することが可能であるとわかった．今後の展望として，テキストの傾向による分節化だけではなく，ツイート発信者のユーザ特徴による分節化を行い，誰が，どの様な意見を持つかを可視化することで，さらなる話題の理解支援を行いたいと考えている．

参考文献

[1] 総務省:平成 30 年版情報通信白書,第 4 章 第 1 部第 2 節,2018

[2] 岩木祐輔，アダムヤフト，田中克己:マイクロブログにおける有用な記事の発見支援，DEIM Forum 2009

[3] 鈴木諒雅，杉本徹:話題語に関連する情報のマイクロブログからの抽出，情報処理学会第 78 回全国大会

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.