



テキストからの知識の自動獲得と意見分析への応用



研究者所属・職名：情報理工学院・教授

ふりがな おかざき なおあき

氏名：岡崎 直観

主な採択課題：

- [基盤研究\(A\)「常識的知識を活用した言語理解・推論に基づく議論マイニングの新展開」\(2019-2023\)](#)
- [若手研究\(A\)「分散表現と構成的意味計算に基づくテキストと知識ベースの頑健なグランディング」\(2015-2017\)](#)
- [若手研究\(B\)「語彙データベースと大規模コーパスに基づく意味クラス推定器の開発」\(2011-2012\)](#)

分野：知能情報学、ウェブ情報学

キーワード：自然言語処理、人工知能、機械学習、知識獲得、意見分析

課題

● **なぜこの研究をおこなったのか？（研究の背景・目的）**
 実社会の情報を発信・消費・共有する手段として、ソーシャルネットワーキングサービス（SNS）が広く普及しました。SNSは便利ではある一方で、自分と同じ立場の意見が共有・増幅されたり、フェイクニュースが拡散するなど、新たな問題を生み出しています。そこで、コンピューターを使って人々の意見を幅広く収集し、その議論構造を分析することで、様々な価値観が混在する中での意思決定・合意形成を支援したいと考えています。

● **研究するにあたっての苦労や工夫（研究の手法）**
 人間のように文章を理解するコンピューターの実現を目指す研究分野は、自然言語処理と呼ばれます。コンピューターによる自動意見分析は、自然言語処理の応用と見ることができます。

近年の自然言語処理の研究では、大量の訓練データと計算パワーに支えられた深層学習に基づくアプローチが主流となっています。ところが、「総議員の2/3以上の賛成で発議というのはおかしい」という発言から憲法96条の改正に「賛成」であることを認識するには、憲法96条に関する知識が必要になります。自然言語処理の最先端の研究でも、常識的知識を獲得・蓄積し、その知識を活用しながら推論を行う有効な枠組みは、確立されていません。

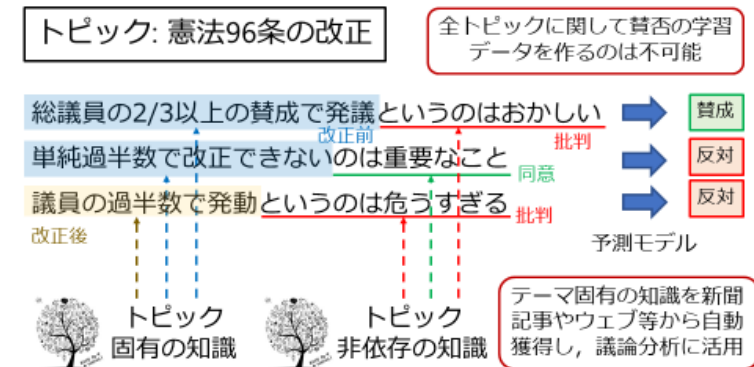


図1 意見分析で常識的な知識が必要になる例



テキストからの知識の自動獲得と意見分析への応用

研究成果

●どんな成果がでたか？どんな発見があったか？

Wikipediaなどを知識ベースとして活用し、テキストと知識ベースの対応付けや因果関係知識の自動獲得などの基盤研究、獲得した知識を活用した意見分析などの応用研究を行いました。

まず、文章中に含まれる人物名、会社名、場所名、制度名などの固有表現をWikipedia記事へ自動的に対応付ける手法を確立しました。例えば、文章中で「トランプ」という表現が出てきたとき、その表現がカードゲームのことを指すのか、米国の大統領のことを指すのか曖昧ですが、この手法では与えられた文章の文脈を深層ニューラルネットワークに埋め込むことで、テキストの情報をWikipediaに正確に結びつけることが可能になりました。

また、「手洗いがウィルスの流行を抑える」と「手洗いがウィルスの感染のリスクを減らす」といった文章から、「抑制(手洗い,ウィルス)」という因果関係知識を獲得するため、文の意味を構成的に計算するモデルを開発しました。さらに、Wikipediaの記事をコンピュータが読解し、「TPPが自由貿易を促進する」「TPPが日本の農業を衰退させる」といった因果関係の知識を自動的に獲得する手法を開発しました。

これらの常識的知識を活用しながら意見分析を行う手法として、文章を深層ニューラルネットワークで読み込みながら、必要な知識を参照するアーキテクチャを提案し、常識的知識が意見分析の精度を高めることを実証しました。なお、これらの研究成果は言語資源として、一般に公開しています。

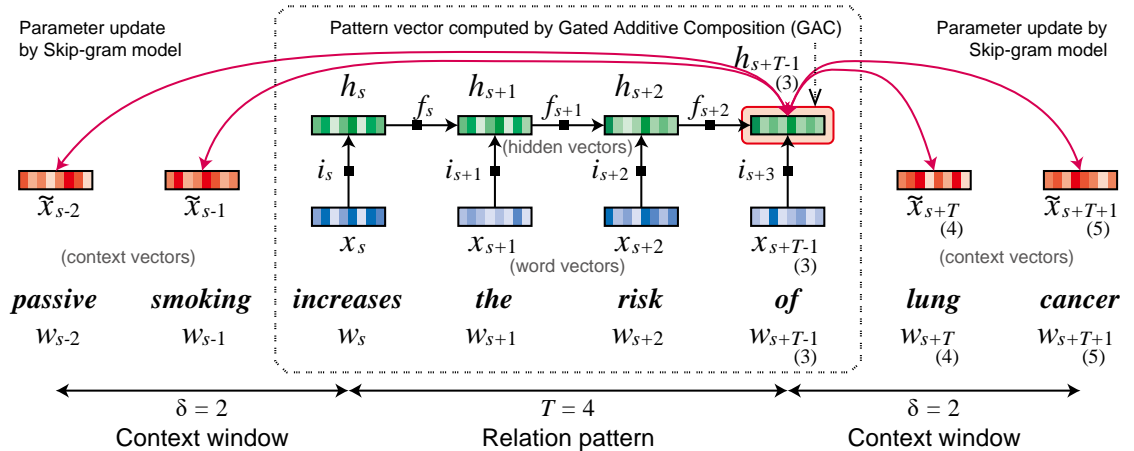


図2 深層ニューラルネットワークで“increase the risk of”という句の因果関係の意味を計算する例

今後の展望

●今後の展望・期待される効果

真に文章を理解する人工知能を実現するには、複数の文章で述べられている知識を記憶・蓄積し、不足している知識を補いながら推論・解釈する仕組みが欠かせません。本研究を深めることで、言語の理解や推論といった人工知能の長年の難題に取り組んでいきたいと考えています。