

1. はじめに
2. 自然言語処理
3. 自然言語処理の推定
4. 関連研究
6. まとめ

修士研究について

水上 和秀 (Kazuhide Mizukani)
u355020@st.pu-toyama.ac.jp

富山県立大学 工学部 電子情報工学専攻

October 20, 2023

1.1 本研究の背景

2/7

1. はじめに
2. 自然言語処理
3. 自然言語処理の推定
4. 関連研究
6. まとめ

背景

- sns の普及により不確実な情報やデマといった「フェイクニュース」が多く存在する
- 人が発信する情報はその情報に対する発信者側の主観的な態度(モダリティ)が含まれており、情報の不確実性が増すことがある

目的

- (予定) ある情報を言語分析し、それが正しい情報か評価する

1. はじめに
2. 自然言語処理
3. 自然言語処理の推定
4. 関連研究
- 6.まとめ

日本語の自然言語処理

- テキストデータを分析する技術として自然言語処理(NLP)がある
- 近年自然言語処理の技術が向上し、英語や中国語の高度な分析が可能となったがそのほかの言語の分析精度はあまり高くない
- 特に日本語はひらがな、カタカナ、漢字などが混在しており、その難しさから日本語の自然言語処理の研究は遅れている。

3. 自然言語処理の理論

4/7

形態素分析

- 形態素分析は自然言語処理の手法の一つでテキストを単語ごとに分割する手法
- テキスト内の各単語に対して品詞タグ(名詞、形容詞など)をつけることができ、文法解析に役立つ

TF-IDF

- テキストデータ内の単語の重要性を評価するための手法
- 特定の文章内での単語の出現頻度とその単語が一般的に使われているかを考慮して単語の重みを計算する

コサイン類似度

- 2つのベクトル間の類似性を測るための手法
- 言語処理では、文章を数値化し、その数値の類似度を調べることによって二つの文章が類似しているか調べることができる

→コサイン類似度は文脈を考慮せず、単語の出現回数による文章の比較のため、単語の順序や文脈による意味の違いをとらえることができない

3. 自然言語処理の理論 2

1. はじめに
2. 自然言語処理
3. 自然言語処理の推定
4. 関連研究
6. まとめ

N-gram 法

- n-gram 法はテキストを n この文字のまとまりに分割して分析する手法
- 分解した文字列片を見出し語としてインデックスを作成する
- n-gram 法を使うことにより文脈の曖昧性を考慮した文字検索をすることができる

モダリティ解析

- 文章のうち、「文の述べ方」にかかる主観的な意味内容の部分
- 「推量」、「命令」、「疑問」などがある
- 文章のモダリティを分析することによりその情報に含まれる意味合いを分析することができる

4 関連研究

1. はじめに
2. 自然言語処理
3. 自然言語処理の推定
4. 関連研究
6. まとめ

言語概念ベクトルを用いた文書間類似度に基づく複数文書自動要約

■ a

文脈理解のための文モダリティの分類と自動判定

- 今までの小論文の自動評価システムは接続表現の主張のみを手掛かりにして文脈を把握するなど、筆者の主張について論理の展開を評価の観点に加えることに対して十分に配慮されていなかった
- 小論文の文脈理解に向けて、文章中で展開される「筆者の意見が述べられている文」を自動的に判定する手法について検討
 - 決定木学習器によって文のモダリティのカテゴリ分けを行った
 - 文章中のある分が意見文であるか叙述文であるか判定する精度が高かった

まとめ

7/7

1. はじめに
2. 自然言語処理
3. 自然言語処理の推定
4. 関連研究
6. まとめ

まとめ

- 文脈の分析方法を調べた
- まだ方向性は固まっていない
- 自然言語処理分野で進めていく

方向性(予定)

- 情報に含まれる主観的な態度(モダリティ)や文脈を考慮し、その情報の真偽を評価する
- もっと理論について調べる