

Web上のテキストデータからの重要語・評判の抽出

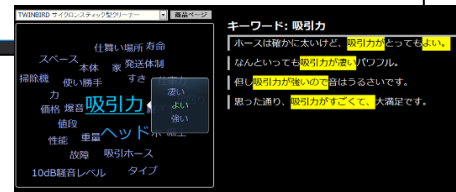
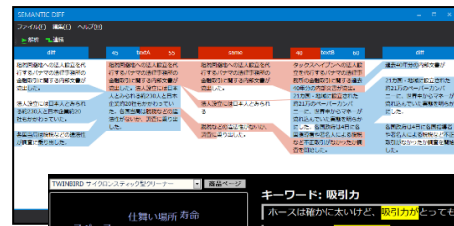
工学研究科 電子情報工学専攻 湯本 高行

キーワード 自然言語処理、機械学習、情報抽出、評判情報分析、Web

研究概要

文書データに対して、統計的な分析や機械学習を行い、重要な語句の抽出や印象の推定を行うアルゴリズムの開発を行っている。具体的には以下のような課題に取り組んでいる。

- 2つのニュース記事から意味的な差異および共通点を抽出し、その重要度を算出する手法の開発
 - ツイートが特定のトピックについて肯定的か否定的かを分類する手法の開発
 - 商品レビューからユーザがどの特徴に対してどのような意見を持っているかを抽出し、どのような特徴が特に重要視されているかを推定する手法の開発
- また、これらのアルゴリズムを応用した情報システムの開発も行っている。



アピールポイント

現在は、商品レビュー、新聞記事データ、ソーシャルメディアデータなどWeb上のデータを対象にしているが、他の分野であっても、ある程度の規模の文書データであれば応用は可能である。

応用分野

機械学習、評判情報分析、テキストマイニング、自然言語処理を伴う各種情報分析