

自由記述文の意味に基づく高速自動分類・検索技術



総合情報学部 総合情報学科

安達 由洋 教授 Yoshihiro Adachi

研究概要 自由記述文の特徴を低次元ベクトルで意味分散表現して、話題や内容に基づき高速に分類・検索する技術を開発しました。

研究シーズの内容

自然言語処理と意味分散表現手法を用いて、自由記述文を超高速(10,000 文を1秒以内)に分類・検索する技術を開発しました。これを用いれば、授業で随時アンケートを取り分析して受講生の疑問や理解状況を反映した授業を進めること、数千人が参加する講演会や集会で参加者の意見や疑問などにリアルタイムに答える講演をすること、Web サービスで口コミ・レビューを高速自動分析してユーザーの意見や意図を考慮に入れた高度で知的なサービスを提供すること、SNS 上で発信される自由記述文を高速自動分析して世論やトレンドなどを反映した商品・サービスを開発することなどができます。

本技術による分類・検索処理の流れ:

(分類処理)

処理1:各自由記述文を形態素解析して単語に分割する

処理2:文を構成する各単語の意味分散表現ベクトルの相加平均をとって文ベクトルを計算する

処理3:文ベクトルに基づいて自由記述文をK-means++法などによりクラスタリングする。クラスタラベル候補集合から各クラスタのセントロイドに近いラベルを付ける

(検索処理)

処理4:入力した問合せ文から処理1~2で問い合わせ文ベクトルを求め、コサイン類似度により類似度の高い自由記述文を検索する。自由記述問合せ文により話題と内容による検索ができる

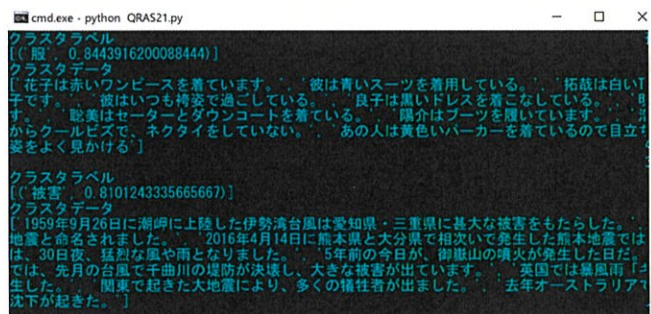


図1:分類処理結果の例



図2:検索処理結果の例

研究シーズの応用例・産業界へのアピールポイント

授業や講演会での参加者の疑問や意見、SNS 上の世論、Web 上のレビューなどを高速分析できます。

特記事項(関連する発表論文・特許名称・出願番号等)

“Development and evaluation of a real-time analysis method for free-description questionnaire responses”, IEEE ICCSE 2020.

特願 2020-085337、「文を分類する方法、装置およびプログラム」.