

## 大規模なテキストデータ分析に基づく情報の抽出・可視化技術

社会情報科学部 湯本 高行



## キーワード

自然言語処理、機械学習、データマイニング、テキストマイニング、情報抽出

## 研究概要

さまざまなテキストデータに対して、データマイニング、統計的手法、機械学習などを用いて、

- (1) 文章の重要箇所やユーザが必要とする情報などの自動抽出技術
- (2) ユーザが閲覧している情報の位置付けの理解のためのスコア化(社会的な認知度, 信頼度など)や関連情報の推薦手法
- (3) 抽出した情報やスコアをユーザにわかりやすく提示する手法

などの研究を行っています。具体的なテーマとしては、以下のようなものがあります。

- ・語の共起関係に基づく情報の典型性/意外性の推定
- ・商品レビューの効率的な閲覧支援のための重要観点の抽出・可視化



## アピールポイント

Webページ、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、さまざまなテキストデータを対象とした分析技術を研究しています。また、特定の業界・分野に特化した分析技術の開発も行っています。

例: 経済分野・震災関連に特化したニュース記事分析、製造業における代替工場推薦、QAサイトでの病気の相談からの関連症状の発見、など

## 応用分野

- ・テキストデータからの情報抽出・知識発見、テキストデータを対象とした情報検索・情報推薦
- ・特定の業界・分野に特化した分析とその応用(例: 経済分野・震災関連のニュース記事、医療健康関連のWebページの分析など)