B5-1

Twitterからの消費者ニーズの抽出手法に関する提案

川島崇秀 †1 佐藤哲司 †1 神門典子 †2

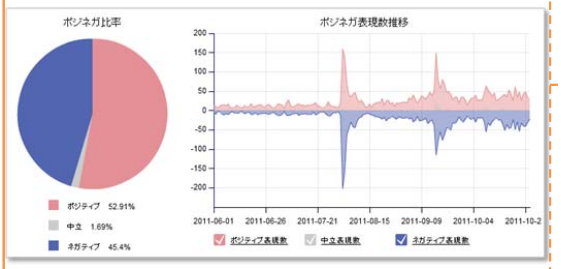
†1 筑波大学　†2 国立情報学研究所

背景・目的

1. 背景①

1.ソーシャルリスニングへの注目

・・・SNS上の口コミ分析をマーケティング活動へ活かす



(例) センチメント分析

2. 背景②

要望は消費者のニーズを直接的に表す重要な情報

➡ But 従来手法では抽出精度に課題がある…

3. 目的

より高い精度での要望抽出を目指す

- 消費者ニーズの理解を支援
- マーケティング活動
- 新規事業創造

4. 着眼点

要望に出現する特徴的な文末表現に注目

・職場周辺の悪臭が、去年の夏からひどい。なんとかして欲しい
・大阪市は南港の交通の便をお台場レベルにまで早急に引き上げるべき
・大通りを深夜に徘徊してる暴走族がうるさくて寝られない。なんとかしろ

➡ 特徴表現を使って半自動的に教師データを収集&学習させる

提案手法

1. Distant Supervisionを用いた教師データの半自動収集

Distant Supervision

- 「Distant Supervision」は、半教師あり学習の一つ。少数の知識ベースを元に半自動的に大量の教師データを収集
- 本手法を適用することによって、より低コストな機械学習の実現を目指す

1.要望表現辞書の作成

先行研究の文末表現リストを参考に、要望表現辞書を作成

2.教師データの収集

2-1. 要望表現を含むツイートの抽出

2-2. 手がかり表現が文末から3-Gram以内に出現するツイートの抽出

-「ない」はネガティブな単語との共起を追加条件に設定

<<使用した手がかり表現一覧>>

欲しい,～しろ,～たい,～てくれ,～でくれ,
いいのに,といい,いいと思う,して,べき,で下さい
～ない,～づらい,辛い,きつい

2. 機械学習のアルゴリズムを用いた分類器の構築

SVM

- 分類器の生成には、情報抽出のタスクにおける有効性が知られている SVM を使用
- 本手法を適用することによって、より高い精度での分類を目指す

3. 分類器の構築

素性として、要望を含む投稿内に出現する単語の中から、名詞・動詞・形容詞・形容動詞・助動詞の 5 つの品詞に該当する単語を用いる

評価実験

評価対象

- 要望抽出対象としてソーシャルゲームのタイトル名を使用
- 学習データにはソーシャルゲーム10タイトルに関するツイート4万件を使用

評価方法

- 1000件のツイートにクラウドソーシングサービスを利用して，人手でラベル付けを行い評価データを作成
- 構築した分類器を用いて分類を行い、適合率・再現率・F値を算出して辞書ベースの分類器と性能を比較

実験結果

- 適合率/再現率/F値が向上
- 再現率が大幅に向上

	適合率	再現率	F値
提案法	0.26	**0.69**	0.38
従来法	0.19	0.09	0.09

課題

- 詳細なルール設計(適合率の低さ)
- 新たな手がかり表現の追加(さらなる精度向上)
- 素性の追加