

QAサイトにおける質問に適した回答の判定

西原陽子^{1,2} 松村真宏² 谷内田正彦²

1日本学術振興会 2大阪大学



1. 研究目的

質問者にベストアンサー(BA)と選ばれる回答の判定手法を提案

□従来手法が適用しにくい理由

- ・質問者が気に入る回答がBAになる
- ・正解回答がBAになるとは限らない

□提案手法の特徴

- ・質問と回答の相性を含まれる文末表現で評価
- ・BAと似ている回答を除いて学習



2. ベストアンサーの判定法

入力：BAとノーマルアンサー(NA)の集合

出力：BAの判定器



(1) QとBAのペアを文末表現でクラスタリング

- ・茶釜で、文末表現の3-gramを抽出
- ・QとBAのペアを、階層的にクラスタリング
- ・クラスタ間距離は、Ward法で算出

(2) クラスタごとにBAとNAを学習

- ・SVMlightを使用
- ・素性は名詞、動詞、形容詞の1,2,3-gramと文末表現の1,2,3-gram
- ・BAを正例、NAを負例
- ・同じQで、BAとのコサインが高いNAは除く



3. 評価実験

BAの判定精度を確認

□比較手法

BL：単語と文末表現でSVM

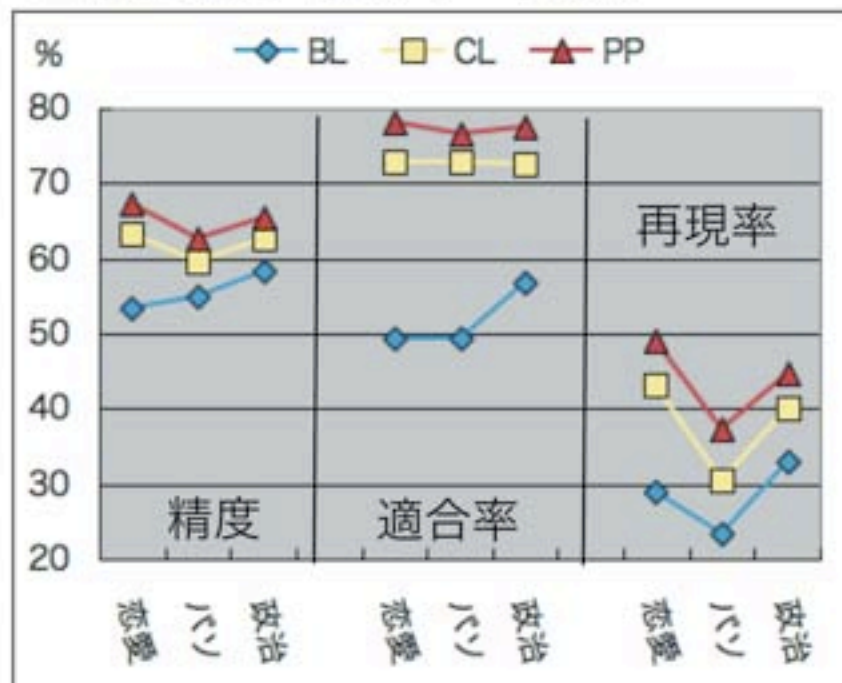
CL：クラスタリングして、単語と文末表現でSVM

PP：提案手法

□使用データ

- ・Yahoo!知恵袋での質問カテゴリ「恋愛相談、人間関係の悩み」「パソコン、周辺機器」「政治、社会問題」に含まれたQA
- ・BAとNAを5,000件ずつ使用

□判定精度・適合率・再現率



クラスタリング、BAと似ているNAの除去で、判定精度が向上

4. 結論と今後の課題

- ・BAとなる回答の判定手法を提案
- ・QとBAをクラスタリングし、BAと似ているNAを除くと、判定精度が向上した
- ・今後の課題はBAに似たNAの処理