

文書・文間及びカテゴリ間の関係を 考慮したレーティング予測

知能数理研究室 12056 外山 洋太

研究背景

商品レビューによる評判分類

- ▶ 対象問題：複数のカテゴリにおけるレーティング予測
- ▶ 文字から文書に渡る様々な言語要素間の関係，及び，カテゴリ間の関係が重要
- ▶ 従来手法はそれらを十分に考慮できていない

ホテルの雰囲気はとてもよく食事もおいしかったです。部屋についても、窓からの見晴らしがよく海がとても綺麗でした。チェックイン当日、入口のフロアの汚れが気になりましたが、翌日にはきちんと清掃されていました。機会があれば、また利用したいと思います。

総合	★★★★★	4
サービス		3
立地		5
部屋		4
設備・アメニティ		4
風呂		3
食事		-

提案手法

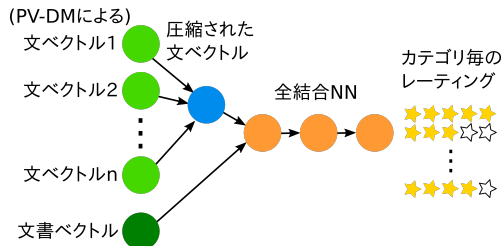
目的

以下を考慮した分類の実現

- ▶ 文章・文間の関係
- ▶ カテゴリ間の関係

方法

- ▶ パラグラフベクトル
 - ▶ 文や文書を，その意味を表す実数ベクトルに変換する手法
 - ▶ 評判分類において優れる
- ▶ ニューラルネットワーク
 - ▶ 神経回路を模した機械学習手法
 - ▶ 分類問題に適用可能
 - ▶ 文書・文間やカテゴリ間の複雑な関係を考慮



実験及び結果

実験設定

- ▶ 7 カテゴリ 6 クラスのレーティング予測の正答率を測定
- ▶ データセット：楽天トラベルにおけるレビュー約 330,000 件

結果

- ▶ 提案手法が従来手法より **高い正答率**を示す
- ▶ **文の並び**が予測のために重要
- ▶ 文書ベクトルと文ベクトルを同時に素性として用いることが有効

手法	正答率
従来手法	0.4832
提案手法	0.5030