

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

多カテゴリにおける商品レビューのレーティング予測

- ▶ 研究意義：企業における文書からの商品の評判分析
- ▶ レビュー内の文・単語・レーティング等の間で様々な関係が存在
→ 文書・文間の関係及びカテゴリ間の関係に着目

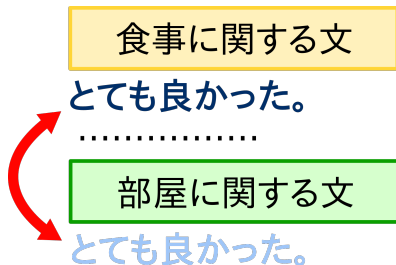
	カテゴリ	レーティング
ホテルの雰囲気はとてもよく食事もおいしかったです。部屋についても、窓からの見晴らしがよく海がとても綺麗でした。フロアの汚れが気になりましたが、翌日にはきちんと清掃されていました。機会があれば、また利用したいと思います。	総合	★★★★☆ 4
	サービス	3
	立地	5
	部屋	4
	設備・アメニティ	4
	風呂	3
	食事	5

文間・カテゴリ間の関係

文間の関係

「とても良かった」の文が

- ▶ 食事に関する文の下に存在 ⇒ 食事◎
- ▶ 部屋に関する文の下に存在 ⇒ 部屋◎



カテゴリ間の関係

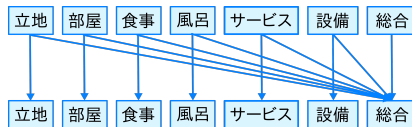
- ▶ 他のカテゴリ ◎
⇒ 「総合」カテゴリ ◎



関連研究

隠れ状態を用いたホテルレビューのレーティング予測¹ (従来手法)

- ▶ 文毎のレーティングからレビュー全体のレーティングを予測
- ▶ カテゴリ間の繋がりを**手調整で変化**させて考慮



パラグラフベクトル²

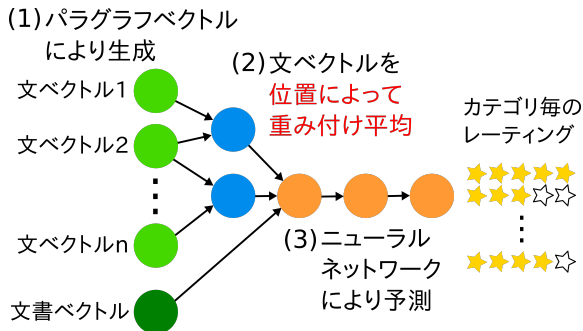
- ▶ 文や文書を実数ベクトルに変換する手法
- ▶ **レーティング予測において優れた性能**

¹藤谷宣典ら, 隠れ状態を用いたホテルレビューのレーティング予測. 言語処理学会第 21 回年次大会, 2015.

²Quoc Le et al., Distributed representations of sentences and documents. ICML 2014, 2014.

提案手法

- ▶ 文書・文間及びカテゴリ間の関係を自動で考慮したレーティング予測
- ▶ パラグラフベクトルと **入出力間の複雑な関係を考慮**できるニューラルネットワークを利用



実験

実験設定

- ▶ 7カテゴリにおける 0~5 点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**を示した

手法	正答率 [%]
従来手法	48.32
提案手法	50.30

まとめと今後の課題

まとめ

- ▶ 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手法を提案
- ▶ 従来手法より高い正答率

今後の課題

- ▶ 文間や単語間、文字間等のより多様な関係を考慮
→ レビューの特徴の抽出と分類のモデルを統合