

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

- ▶ レーティング予測：商品のレーティングを予測すること
- ▶ カテゴリ：レーティングの付く各項目のこと
- ▶ 文同士の位置関係及びカテゴリ間の関係を考慮したい

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。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

文同士的位置関係の例

- ▶ 後半の文章が「総合」カテゴリのレーティングに影響しやすい場合

レビューの例

食事が美味しかったという文章

立地が悪かったという文章

総合



カテゴリ間の関係の例

- ▶ 食事◎ ⇒ サービス◎
- ▶ 設備・アメニティ◎ ⇒ サービス◎



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

- ▶ 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
- ▶ 評判分類において優れた性能

ニューラルネットワーク

- ▶ 入力間・出力間の複雑な関係を考慮

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

目的

- ▶ 文書・文間及びカテゴリ間の関係を考慮した分類

レーティング予測の流れ

(1) パラグラフベクトル

により生成

文ベクトル1

文ベクトル2

⋮

文ベクトルn

文書ベクトル

(2) 文ベクトルを

重み付け平均

→ 文の数を統一

(3) ニューラル
ネットワーク
により分類

カテゴリ毎の
レーティング



⋮



実験と今後の課題

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	0.4832
提案手法	0.5030

今後の課題

- ▶ 文間あるいは単語間等のより多様で複雑な関係を考慮
→ 文書・文ベクトルの生成と分類のモデルを統合