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

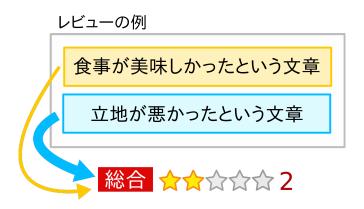
多カテゴリ における 商品レ<u>ビューのレーティング</u> 予測

- ▶ 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
- ▶ レビューとレーティング間の様々な関係が存在
 - → 文書・文間の関係及びカテゴリ間の関係に着目

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

文同士の位置関係の例

▶ 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



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.

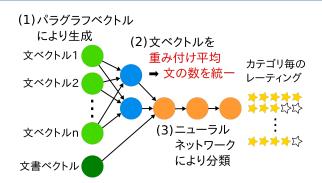
²Quoc Le et al., Distributed representations of sentences and documents. ICML 2014, 2014.

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



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

- ▶ 文間や単語間, 文字間等のより多様な関係を考慮
 - → レビューの意味表現の生成と分類のモデルを統合