

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

背景と目的

- ▶ 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
- ▶ 目的
以下を考慮したレーティング予測の実現
 - ▶ 文章・文間の関係
 - ▶ カテゴリ間の複雑な関係

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	★★★☆☆	-

複数のカテゴリを持つ商品レビューの例

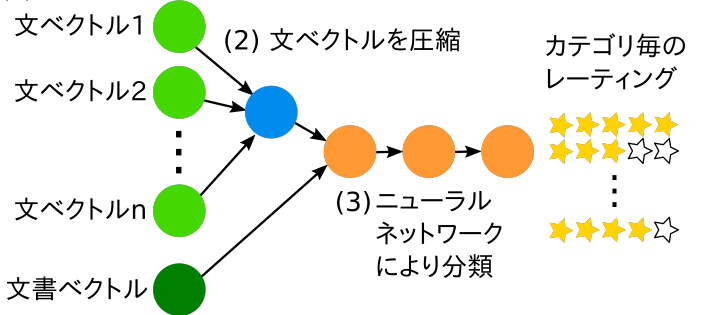
関連研究

- ▶ 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]
 - ▶ 複数のカテゴリにおけるレーティング予測
 - ▶ 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によって変化させカテゴリ間の関係性を考慮
- ▶ パラグラフベクトル [2]
 - ▶ 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

提案手法

- ▶ 特徴
 - ▶ 文毎の意味表現 → **文同士の位置関係**を考慮
 - ▶ ニューラルネットワークによる分類器 → **文書・文間及びカテゴリ間の複雑な関係**を考慮
- ▶ 入力：レビューと正解レーティングの組の集合
- ▶ 出力：各レビューについて予測されたカテゴリ毎のクラス
- ▶ レーティング予測の流れ
 - (1) パラグラフベクトルによってレビュー内の文書全体及び各文の意味表現を生成
 - (2) 文ベクトルをレビュー毎に重み付け平均 → 全てのレビューで文の数を統一
 - (3) ニューラルネットワークで多ラベル多クラス分類

(1) パラグラフベクトルで生成



提案手法におけるモデルの概略

実験

- ▶ 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

まとめ

- ▶ 多カテゴリにおける評判分類問題について、レビュー全体の文書ベクトルに加え重み付け平均された文ベクトルを用いた手法を提案
- ▶ 提案手法が従来手法 [1] より高い正答率を示した。
- ▶ 今後の課題
言語要素間のより多様で複雑な関係を考慮
→ 各レビューの意味表現を生成するモデルと分類を行うモデルを 1 つに統合
→ 学習手法の柔軟性及び正答率の向上を目指す。

参考文献

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