

# 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

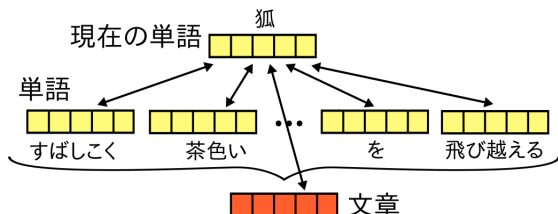
知能数理研究室 12056 外山 洋太

## 背景と目的

- ▶ 対象問題  
多カテゴリにおける商品レビューのレーティング予測
- ▶ 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  
(例1) 食事が美味しかった。  
しかし、それよりも良かったことがある。  
部屋からの眺めが素晴らしかった。  
強調
  - (例2) 部屋からの眺めが素晴らしかった。  
しかし、それよりも良かったことがある。  
食事が美味しかった。  
交換 強調
- ▶ カテゴリ間の関係  
総合 ★★★★★☆ 4  
サービス ★★★★★ 5  
立地 ★★★★★ 3  
部屋 ★★★★★ 4  
設備・アメニティ ★★★★★ 4  
風呂 ★★★★★ 3  
食事 ★★★★★ 5  
影響

## 関連研究

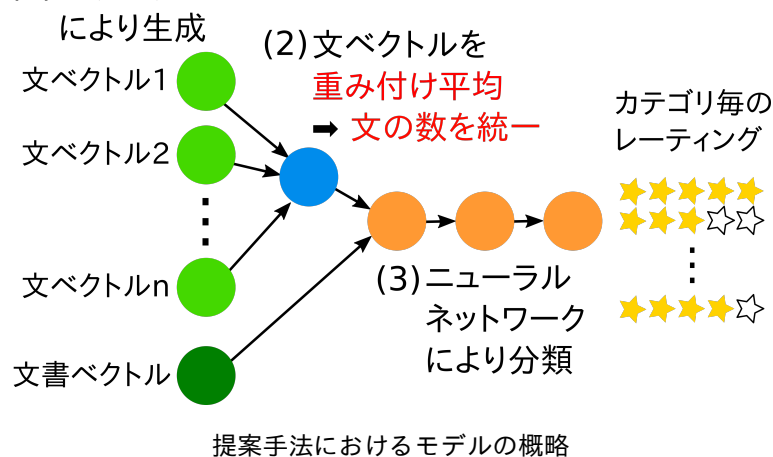
- ▶ 隠れ状態を用いたホテルレビューのレーティング予測 [1]
  - ▶ 文毎のレーティングからレビュー全体のレーティングを予測
  - ▶ カテゴリ間の繋がりを手調整によって変化させカテゴリ間の関係性を考慮
- ▶ パラグラフベクトル [2]
  - ▶ 文や文書を、その意味を表す実数ベクトルに変換
  - ▶ 評判分類において優れる
- ▶ ニューラルネットワーク
  - ▶ 神経回路を模した機械学習手法



## 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

### (1) パラグラフベクトル



## 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：楽天トラベルのレビュー約 330,000 件
- ▶ 結果

手法	正答率
従来手法 [1]	0.4832
提案手法	0.5030

  - ▶ 提案手法が従来手法より高い正答率を示す

## 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

## 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.