

# 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

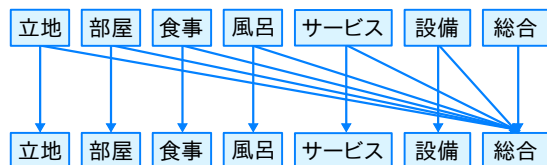
## 背景と目的

- ▶ 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



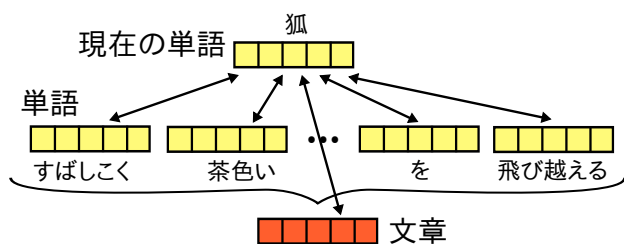
## 関連研究

- ▶ 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**させその関係を考慮



## ▶ 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
- ▶ **レーティング予測において優れた性能**



## 提案手法

- ▶ 特徴
  - ▶ 位置によって重み付け平均された文ベクトル  
→ **文同士的位置関係**を考慮
  - ▶ ニューラルネットワークによる予測  
→ **文書・文間及びカテゴリ間の関係**を考慮

{ 入力：レビューの文書  
出力：カテゴリ 毎のレーティング

### (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

位置によって  
重み付け平均

カテゴリ毎の  
レーティング

(3) ニューラル  
ネットワーク  
により予測

提案手法における予測モデル

## 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

- ▶ 提案手法が従来手法より  
**高い正答率**を示す

| 手法       | 正答率           |
|----------|---------------|
| 従来手法 [1] | 0.4832        |
| 提案手法     | <b>0.5030</b> |

## まとめ

- ▶ 多カテゴリにおけるレーティング予測について，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  
**文間，単語間，文字間等のより多様で複雑な関係**を考慮  
→ レビューの特徴の抽出と分類の**モデルを統合**

## 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.