

# 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

# 研究背景

## 多カテゴリにおける商品レビューのレーティング予測

- ▶ カテゴリ：レーティングの付く各項目のこと
- ▶ 文同士的位置関係及びカテゴリ間の関係が重要

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		5

# 研究背景

## 文同士の位置関係

(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

# 研究背景

## カテゴリ間の関係

- ▶ 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

- ▶ 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

- ▶ 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
- ▶ 入力の要素間の複雑な関係を考慮

# 提案手法

## 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

カテゴリ毎の  
レーティング



⋮



(3) ニューラル  
ネットワーク  
により分類

# 実験及び結果

## 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	<b>0.5030</b>