

レーティング予測によるフォントを基盤としたレビュー解析

豊田工業大学 知能数理研究室 外山 洋太, 三輪 誠, 佐々木 裕

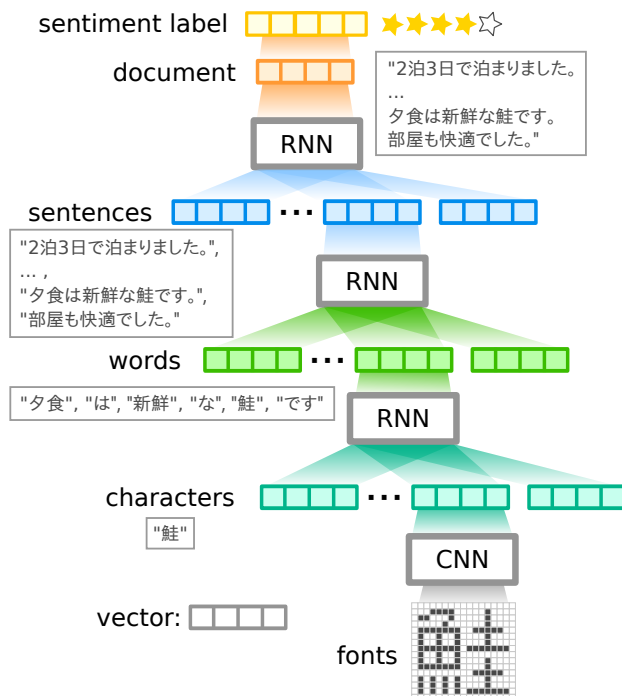
背景と目的

- ▶ 対象タスク：表意，表語文字を含む言語におけるレーティング予測
- ▶ 応用例：企業における文書からの商品の評判分析
- ▶ 目的：文字の表層情報を利用したレーティング予測の実現



提案手法

- ▶ 入力：フォント 画像で表現されたレビュー
- ▶ 出力：予測したレーティングラベル
- ▶ 特徴
 - ▶ フォント 画像から文字の意味情報を抽出
 - ▶ HAN[1] の手法を基に文字からの文書の階層構造を利用
- ▶ 予測手順
 - (1) foo
 - (2) bar



関連研究

- ▶ Hierarchical Attention Network (HAN) [1]
 - ▶ Attention 構造付きの Recurrent Neural Network を用いた文書分類モデル
 - ▶ 文字または単語から文，文書までの階層的構造を利用
 - 文字の表層情報の利用ができていない
- ▶ Radical-Enhanced Chinese Character Embedding [2]
 - ▶ 漢字-部首辞書を利用した漢字埋め込みの生成手法
 - ▶ 対象タスクと漢字の部首当てについて同時に学習
 - 漢字-部首辞書が余分に必要

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
 - ▶ 提案手法の従来手法より 高い正答率

手法	正答率	RMSE
HAN [1]	0.483	0.81
提案手法	0.503	0.73

まとめ

- ▶ フォント 画像を用いたレーティング予測及びレビュー解析の手法を提案
- ▶ 提案手法による従来手法 [1] より高い正答率
- ▶ アテンションの可視化により
- ▶ 今後の予定
 - ▶ 可視化の方法を洗練

参考文献

- [1] Zichao Yang et al., Hierarchical Attention Networks for Document Classification. NAACL 2016, 2016.
- [2] Yaming Sun et al., Radical-Enhanced Chinese Character Embedding. ICONIP 2014, 2014.