

クックパッドデータセットの情報推薦システムへの活用に向けて

太田 葵, 吉本 篤史, 柴田 佑樹, 高間 康史 (東京都立大学)

概要

クックパッドデータの情報推薦への応用を検討

研究1: 知識グラフ埋め込みの利用: 副菜に合う主菜の推薦

研究2: 協調フィルタリングのための人工データセット生成

知識グラフ埋め込みの利用: 副菜に合う主菜の推薦

目的

潜在的 (= 献立データに含まれない) 主菜-副菜の関係を推薦

仮説: 献立データを正しく学習 → 潜在的関係を予測可能

手順

1. TransE (Translations in the Embedding space) による献立埋め込み表現の構築

献立データからトリプル群を作成: 3種類の relation を定義

(副菜, fit_main, 主菜), (レシピ, category_is, カテゴリ), (カテゴリ1, parent_is, カテゴリ2)

2. 既存の献立を正確に学習できるようにモデルを調整

学習に使用したデータセットをテストデータとして MRR と recall@10 を計算

3. 高性能モデルの予測上位を確認

予測上位10件の妥当性を人手で判断 → 適合率を計算

結果

「タコとアボカドのサラダ 柚子胡椒風味。」の 主菜予測結果(下線は既存関係)
コンビーフのさっぱりパテ
ふわふわ卵としらすのホットミルクかけご飯
ダイエットに★ ご飯→キャベツで麻婆丼
【我が家の定番】鳥ミンチdeミートソース
ポークソテー 本格☆あめ色たまねぎソース
白身魚のオープン焼き・ソレント風♪
油揚げの挟み焼き 洋風ver.

今後の予定

潜在的関係の予測精度の評価: アンケート実施

データセット

クックパッドデータセットに含まれる

研究1: 献立データ, レシピカテゴリデータ, カテゴリの階層構造データ

研究2: つくれぽデータ(調理履歴), 食材データ

協調フィルタリングのための人工データセット生成

目的

・人工的な評価値行列の生成 → 協調フィルタリングによるレシピ推薦

・ユーザの合理的行動を仮定: 好きな食材を利用し, 嫌いな食材は利用しないレシピを選択

手順

1. 実データからアイテムプロファイルの構築: 食材を次元とするベクトル

2. 実データからユーザプロファイルの構築: 調理履歴内レシピの平均ベクトル

3. アイテムベクトルとユーザベクトルのコサイン類似度を利用して人工データを生成

3-1. 各ユーザ: 実データ(調理履歴)と同数のレシピを選択(コサイン類似度の高い順)

3-2. ランダム性の追加: レシピの半数を無作為抽出で選択

4. 協調フィルタリングの適用 → 推薦精度を算出

結果

■ 実データと人工データの推薦精度の比較

	RMSE	MAE	Precision@5	Recall@5
実データ	0.277	0.159	0.202	0.015
人工データ①	0.113	0.040	0.995	0.054
人工データ②	0.212	0.153	0.874	0.049

人工データ①: コサイン類似度の高いアイテムから順に生成(3-1)

人工データ②: ランダム性の追加(3-2)

人工データ①: 実データよりかなり高い精度 → 合理的行動の予測は容易

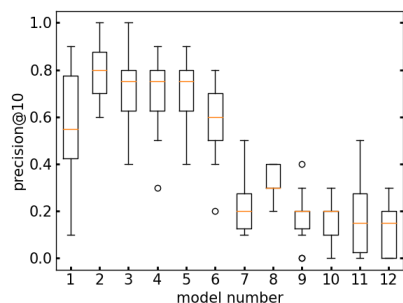
人工データ②: RMSE, MAEは実データに近づく

今後の予定

・実データと人工データセットの類似性の調査: 評価値分布, 人気レシピの比較など

・実データをテストデータとした評価の実施

■ 既存関係の予測精度と潜在的関係の予測精度の関係



高 ← MRR, recall@10 → 低

仮説が概ね成立することを確認