

アンケート結果に対するFactorization Machines に基づく賃貸サイトの満足度推定

久保田涼斗, 大矢航輔, 伏見卓恭(東京工科大学コンピュータサイエンス学部)

背景・目的

- ◆ どの賃貸サイトを使えばいいかわからない際に納得のいくサイトをすぐ使用したい
- ◆ 物件探しについて顧客と提供者をつなげる必要がある
- ◆ 満足度を用いて最適なサイト提示につなげるためにどの質問項目が影響しているか関係性を見出す

使用データセット： オリコンデータセット

案件名：賃貸情報サイト
調査期間：
2018年6月15日から2018年6月19日
対象者：
過去1年以内に賃貸情報サイトで住宅を探して入居した人

実施方法

Field-aware Factorization Machines (FFM)

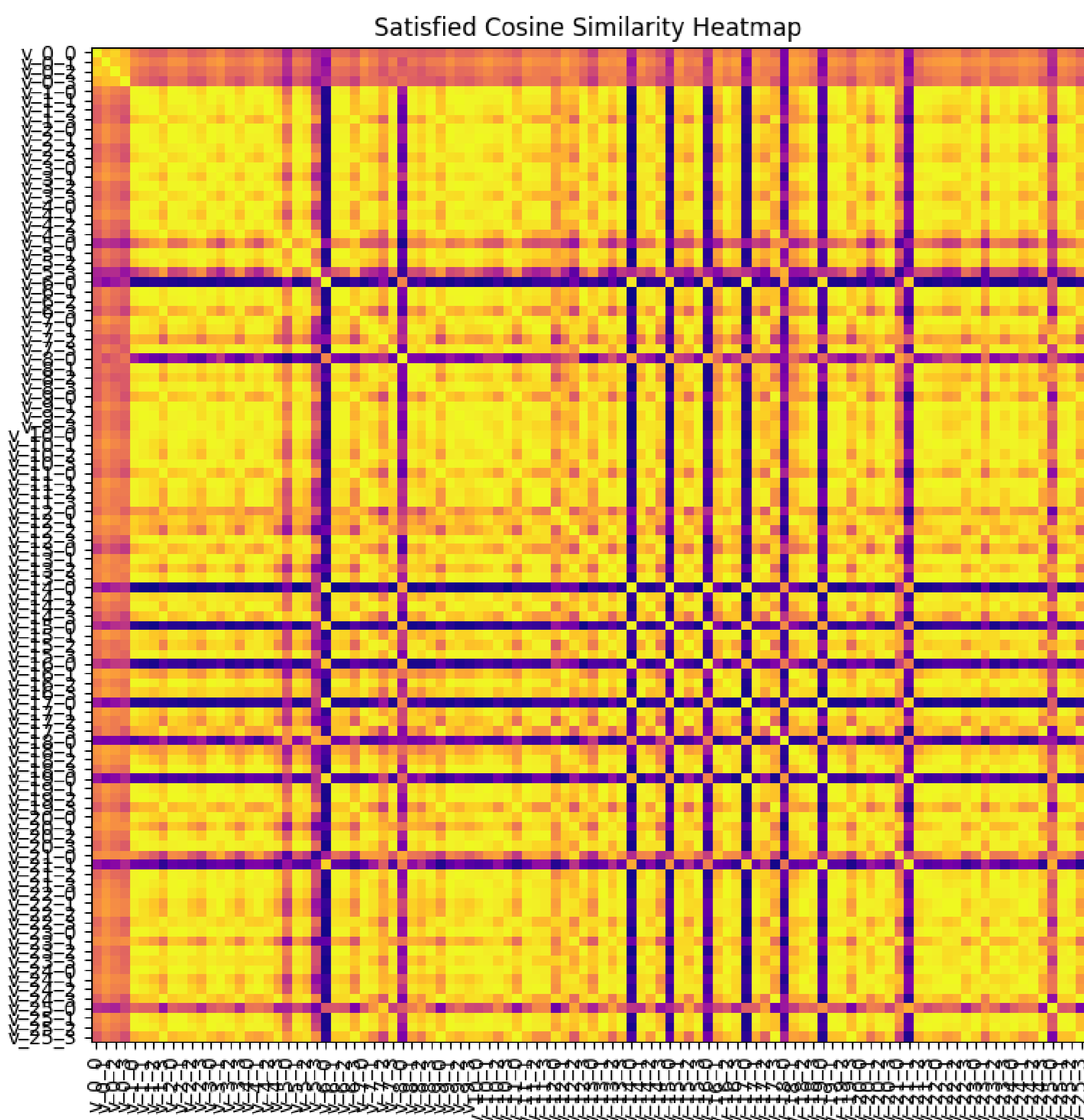
- ◆ アンケート全体の点数の内訳が低い点数では少なく、最も多いのが中間から少し高めめの6~8点に集中していたことや満足度の平均点数が6点後半から7点前半であることを考慮し全体で均等な人数になるよう4つに分類
6点以下を0、7点を1、8点を2、9点以上を3とする
- ◆ データを訓練用(80%)・評価用(20%)で分割 【データ数：1325件】
- ◆ **説明変数**：項目ごとにより分けられた満足度(1~10)
(例)：サイト情報の見やすさ、サイト内の物件の正確性、物件の情報量
- ◆ **フィールド**：回答した点数に従って質問項目を4つのフィールドに分類【計26個】
- ◆ **目的変数**：サイトの総合満足度(1~10)
- ◆ 潜在ベクトルの次元数を32次元に設定
- ◆ モデルを用いて目的変数の予測を行い、得られた潜在ベクトル v の内積(コサイン類似度)を出力

$$f(x) = w_0 + \sum_{i=1}^d w_i x_i + \sum_{i,j} \langle v_{i,F(j)}, v_{j,F(i)} \rangle x_i x_j$$

- ◆ 重みベクトルを用いてコサイン類似度の計算
※ $a = v_{i,F(j)}, b = v_{j,F(i)}$ とする

$$\cos(a, b) = \frac{\langle a, b \rangle}{\|a\| \|b\|} = \frac{\sum_{k=1}^n a_k b_k}{\sqrt{\sum_{k=1}^n a_k^2} \sqrt{\sum_{k=1}^n b_k^2}}$$

結果



コサイン類似度の特徴量ヒートマップ

コサイン類似度の値が最も大きい組み合わせ5つ(左) 最も小さい組み合わせ5つ(右)

特徴量の組み合わせ	潜在ベクトルのコサイン類似度	特徴量の組み合わせ	潜在ベクトルのコサイン類似度
2_1 and 9_1	0.99968	8_3 and 14_0	-0.9990
2_3 and 3_3	0.99965	12_3 and 14_0	-0.9983
10_1 and 25_1	0.99951	6_2 and 14_0	-0.9978
11_1 and 25_1	0.99945	2_2 and 14_0	-0.9971
19_1 and 25_1	0.99943	1_2 and 14_0	-0.9966

予測精度評価

平均二乗誤差 (MAE) : 0.6342

二乗平均平方根誤差 (RMSE) : 0.8741

考察

- ◆ 1番特徴量の内積の重みが大きい2_1(サイトの分かりやすさ)と9_1(物件情報の豊富さ)は物件の情報量に対して物件内容が分かりやすいかどうか比例していると考えられる
- ◆ 25_1(サイト運営会社の信頼性)では10_1(物件情報の分かりやすさ)、19_1(サイトの信頼性)のようにサイトに対するイメージが絡んでおり、実際に利用したサイトを運営している会社のイメージに繋がっていると考えられる
- ◆ 14_0(物件掲載元の問い合わせのしやすさ)は8_3(物件情報の新しさ・更新頻度)や12_3(周辺環境情報の豊富さ)等物件情報で高い満足度になっていることから問い合わせる内容が少なさやサイトの質が良いと考えられる

今後の課題

- ◆ 満足度から得られる最適な物件サイトの提示
- ◆ 複数のモデルを組み合わせた予測精度の向上

謝辞

本研究では、国立情報学研究所のIDRデータセット提供サービスにより、株式会社oricon MEからご提供頂いた「オリコンデータセット」を利用しました。心より感謝いたします。