F-013

# 地域活性化のためのスマートフォンアプリを用いた実店舗および商品の推薦

## Brick-and-Mortar Shops/Goods Recommendation on Smart-Phone for Regional Promotion

酒井 政裕† 高 明淑† 西沢 孝浩† 阿部 真美子†
Masahiro Sakai Myungsook Ko Takahiro Nishizawa Mamiko Abe

## 1. まえがき

川崎市と東芝は，コミュニティ内の商業施設全体を活性化すると共に，消費者であるコミュニティ住民の生活を豊かで楽しくすることを目指し，川崎駅周辺の複数商業施設をクラウド上で仮想的に連携させ，ユーザの嗜好に応じた情報をスマートフォンアプリ上に配信し，購買行動などの検証を行う実証実験「川崎グランシティモール™」を行った．本稿では，当実証実験の一環として行った，利用者のアンケート回答および店舗・商品に対する評価履歴の情報に基づいた実店舗および商品の情報の推薦について，用いた推薦アルゴリズム，得られた結果と課題について検討する．

## 2. 実証実験の概要

実証実験全体の概要図を図 1 に示す．実証実験では，iPhone向けおよびAndroid™向けのアプリ(図 2)を配布し，利用者にダウンロードして利用してもらい，利用者から登録してもらった情報と利用者の行動履歴の情報をもとに，その利用者に適合すると思われる推薦情報を配信した．

利用者は，性別・生年月日・ニックネームを登録してユーザ登録を行うことでアプリの利用を開始することができ，参加している店舗・商品の一覧や検索，店舗・商品の詳細情報，各施設のフロアマップ表示，利用者による口コミ情報や写真を投稿可能な「まちツイ」機能，それらを含む新着情報を一覧できる「ニュースフィード」機能，利用者毎に推薦する店舗・商品の情報を表示する「おすすめ一覧」などの機能を利用することができる．

運営者からは利用者に任意でアンケートへの回答を依頼しており．アンケートでは川崎駅周辺での各カテゴリ(例: 衣料・アパレル店)の店舗利用頻度，興味のあるレジャー・スポーツ・文化芸術活動，好きな映画ジャンル，好きなファッションのジャンル/スタイル，外食の頻度やジャンル，職業や同居家族などの回答を得た．今回の実証実験では，対象エリア内に三つのシネマコンプレックスがあり，また対象とする商業施設には飲食店および衣料・アパレル点が多いことから，アンケートではそれらに関係する項目を多く用意した．

利用者からは，店舗・商品の詳細画面では「いいね」「行った・買った」といったフィードバックを行うことができ，店舗の詳細画面では店舗のお気に入り登録ができる（お気に入り登録された店舗の更新情報は「ニュースフィード画面」で優先的に表示される）．

また，「おすすめ一覧」画面では，自分に合わなかった推薦に対して，「ゴミ箱」のアイコンをタッチすることでそのことをフィードバックができる．

実験の規模は，最終的に川崎駅前 7 商業施設，約 500 店舗の参加にのぼり，合計約 900 個の商品/サービス/イベント情報を掲載した．また，アプリ登録者は約 2 千人にのぼった．

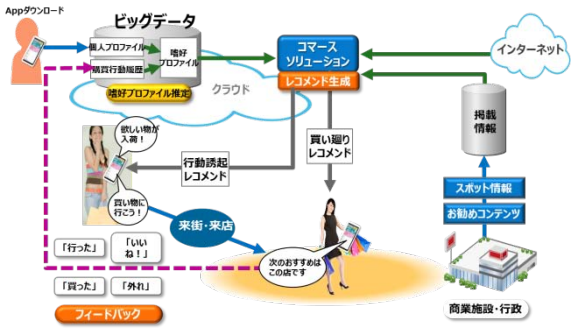


図 1: 実証実験概要


図 2: 実証実験アプリ（仮想の店舗によるイメージ図）

## 3. 推薦システム

本節では実証実験システムの推薦部分の構成と実装について述べる．

### 3.1 推薦に用いる情報

推薦で用いるユーザ情報は，デモグラフィック情報として性別・年齢の情報，アンケート回答の情報，ユーザからの各店舗・商品への「行った／買った」「いいね」「はずれ」等のフィードバック情報を用いた．

推薦対象の店舗・商品については， 0-1 の属性を「タグ」として百数十個程度用意し，運用者らが手作業で属性を設定した．タグには「イタリアンレストラン」など店舗ジャンルなどに関するもの，「クラシック音楽」「ノスタルジー映画」「ガーリー系」など嗜好ジャンルに関わるもの，ターゲット年齢層や性別を表すもの，価格帯を表すもの，「割引キャンペーン」などの特殊な属性等が存在する．以降ではすべてのタグの集合を TAGS で表す．

### 3.2 推薦アルゴリズム

推薦アルゴリズムには主に，内容ベースフィルタリング方式と，協調フィルタリングがあるが，本実証実験では内容ベースフィルタリングとして，後述するタグマッチ方式，

†（株）東芝, Toshiba Corporation

タグスコア方式，TD-IDF を用いた方式を，協調フィルタリングに関してはユーザ間協調フィルタリング方式を実装し，それらを組み合わせたハイブリッド方式を運用に用いた．

タグマッチ方式

推薦候補であるコンテンツに付与したタグと，ユーザが初期に登録したアンケート情報のマッチングをとるためのルールを複数定義し，いずれかによってマッチングが取れたコンテンツを推薦する，もっとも単純な推薦方式である．

今回は「男性ファッション」「女性ファッション」「ファッション雑貨」「日用品，趣味系」「サービス」「映画」「レストラン」の 7 つのカテゴリ毎に異なったルールを定義した．例えば，「レストラン」であれば，コンテンツの対象年齢層[1]・性別と利用者の年齢・性別がマッチしたうえで，「カフェ」などのタグと「川崎駅前のカフェ利用頻度」「平日昼間に外食でカフェに行くか」といったアンケート回答のいずれかでマッチした場合に，マッチすると判定した．また，これに加えて，以下のような工夫も行っている．

(1) 扶養家族の年齢層の考慮：カテゴリによっては年齢層のマッチングの際に，年齢だけでなく扶養家族の年齢層との OR にする（子供服や赤ちゃん用品を扱う店舗などを考慮）．
(2) 閾値の設定：「川崎駅前のカフェ利用頻度」のような段階を持つアンケート項目のマッチングの際には，項目毎に回答の分布が異なること[2]を想定し，事前に別途実施したアンケート回答の情報を用いて閾値を設定した．具体的には，年齢性別趣味などで抽出した対象層での第 1 四分位点（ただし第 1 四分位点となる選択肢が「利用しない」など最も低い選択肢である場合には一つ上の選択肢）以上である際にマッチするとした．
(3) 推薦リストの多様性：推薦リストに記載されるコンテンツが特定のジャンルに偏ってしまい，利用者がたまたまそのジャンルに興味を持っていないときに推薦リストが価値のないもの（例えばレストランばかりを推薦する推薦リストは昼食直後には価値が低い可能性がある）になることを避けるため，ジャンルごとに上限（例：スイーツ店は 4 件まで，ホームセンターは 1 件まで）を設けてランダム選択を行う．

タグスコア方式

タグマッチ方式は，マッチの度合いを考慮せず，結果がマッチするかしないかのいずれでしかないため，嗜好の強さの順に推薦するといったことができない．そこで，ユーザ毎の各タグへの嗜好の強さを表すスコア(以下，タグスコア)を計算しておき，各ユーザへの推薦生成時には各コンテンツについて付与されたタグのスコアの総和を計算し，最も大きいものから順に推薦する方式を実装した．

タグスコアの初期値は，ユーザの嗜好アンケートをもとに，それと対応するタグスコアに加算して計算する．また，アンケート回答の情報に加えて，利用者がコンテンツに対して行う「行った／買った」「いいね」「はずれ」等のフィードバックの情報についても反映させることとした．具体的には，フィードバック対象のコンテンツに付与されていたタグの(そのユーザの)点数に対して，ポジティブなフィードバックであれば加算，ネガティブなフィードバックであれば減算する．

また，マッチしたものから毎回ランダムに選択を行うタグマッチ方式と比べ，固定的なスコアとなるタグスコア方式では利用者から見た新鮮味が失われてしまうため，推薦リストの上位から順にコンテンツを埋めていく際に，一旦選択したコンテンツと同一のカテゴリのコンテンツのスコアを一定の係数をかけることで減算する機構も実装した．

TF-IDF方式

今回の実証実験では「レストラン」「ファッション」「映画」を主要な推薦対象としていたことから，それらのコンテンツに関しては，その中のより細かいジャンルも属性として付与し，また，ユーザへのアンケート回答に関しても対応するような設問を用意した．例えば，レストランに関しては「イタリアン」「フレンチ」「フードコード」といった属性（「タグ」）を，女性向けファッションに関しては「カジュアル系」「ガーリー系」「裏原系」といった属性をそれぞれ用意・付与している．それに対して，それ以外のコンテンツ，例えば，「時計店」「スポーツジム」「食料品」「惣菜」などについてはより細かなジャンルの属性は用意していない．

タグスコア方式では，各コンテンツのスコアを計算する際に，そのコンテンツに付与したタグに対するユーザの嗜好度の和を計算している[3]という仕組み上，以下のような問題が発生していた．

- 細かい属性まで用意されているジャンルのコンテンツは，他のジャンルのコンテンツよりも，スコアが高くなりがちな傾向があり，その他のコンテンツが推薦されにくくなる．
- 「女性向け」「男性向け」「女性向け」「ターゲット年齢層F2/M2[4]」「中価格帯」など，大量のコンテンツに対して付与されているタグについては，フィードバックに基づく嗜好度の加算が頻繁に行われ、他のタグに比べてスコアが高くなってしまう．

この問題を避けるため，最終的には，文章内の単語の重みづけ手法である TF-IDF (Term Frequency – Inverse Document Frequency)[2]を用いてタグの重みづけを行うことで解決した．TF-IDF の基本的な考え方は，単語 t の文章 d 中における出現頻度を tf(t,d)，その単語を含む文章の出現頻度を df(t) として，文章中の単語の重みを

[1]広告業界におけるマーケティングで用いられる C1, C2, F1/M1, F2/M2,…などの区分を利用
[2] たとえば，食料品店は週に 1 回でも少ない方だが，花屋の場合，3 ヶ月に一回でも多いと言える．
[3] タグスコア推薦による各コンテンツのスコアの計算は，各タグに対する利用者の嗜好度合のベクトルと，コンテンツに各タグが付与されているかどうかの 0-1 のベクトルの内積の計算となっていることに注意されたい．
[4] F2 層，M2 層はそれぞれ 35 歳～49 歳の女性と男性の区分．

$$tfidf(t,d) = tf(t,d) \times \log(1 / df(t))$$

とするというものであり，tf(t,d) 部分は文章中における出現頻度が高い単語ほど重要度が高くなる作用を持ち，log(1 / df(t)) は多くの文章に出現するような単語ほど重要度を下げる作用を持つ．

ここでは，各コンテンツcを文章，コンテンツに付与されたタグを文章が含む単語として考え， TF-IDFを用いてベクトル $\langle tfidf(tag,c)\rangle_{tag\in TAGS}$ を計算し，これを単位ベクトルに正規化したものを，コンテンツcの特徴ベクトル $v_c$ として用いることとした．

また，各ユーザのフィードバック情報のタグスコアへの反映の際にも，フィードバック対象cの特徴ベクトル $v_c$に応じた重みでスコアを加減算することとした．得られた結果を正規化した結果をユーザの特徴ベクトル $v_u$ とする．

ユーザuのコンテンツcへのスコアは，二つのベクトル$v_c$と$v_u$のコサイン類似度

$$\cos\theta = \frac{v_u \cdot v_c}{|v_u||v_c|} = v_u \cdot v_c$$

として定義する[1]．

ユーザ間協調フィルタリング方式

ユーザ間協調フィルタリングは，あるユーザに対して，似た嗜好を持つユーザが高く評価しているコンテンツを推薦する方式である．今回は協調フィルタリングの手法のうち k-近傍法の方式を採用し，ユーザごとに，事前アンケートの回答結果から類似度を計算し，類似度の高い k 人のユーザのコンテンツ評価値（後述）を参照してコンテンツごとにスコアを計算する（図 3）．

ユーザの特徴ベクトルとしては，事前アンケート項目のうち，趣味，映画への興味，ファッションへの興味，飲食店への興味を用い[2]，ユーザ間の類似度は，特徴ベクトルの間のユークリッド距離の逆数，コサイン類似度などを検討したが，最終的には値の分布と以下の理由からユーザ毎の平均値を差し引いたうえでのコサイン類似度を用いた．

- 特徴ベクトルの各成分が，「現在楽しんでいる」(=3)，「今後やってみたい」(=2)，「今はあまり興味はない」(=1)という 3 段階の値であり，常に正の値のため，コサイン類似度では角度の差が出にくい．
- 利用者によって，積極的な回答をする率に大きな差があり，各人の回答の基準が一致していないことが予想され，それを補正するため．

推薦対象ユーザ v のコンテンツ i に対するスコアを以下の式で計算する．

$$P(v,i) = \sum_{u\in kNN(v)} sim(u,v) \times F(u,i)$$

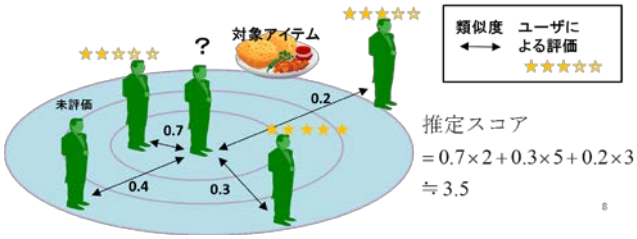


図 3: ユーザ間協調フィルタリング

ここで sim(u,v)はユーザ u とユーザ v の類似度，F(u,i)はユーザ u のコンテンツ i への評価値(数値化したフィードバックの総和)，kNN(v)はコンテンツ i への評価を行っているユーザのうち v への類似度の高い順の最大 k 人の集合である．

### 3.3 推薦アルゴリズムの運用

実際の運用においては，当初はもっとも単純な方式であるタグマッチ方式で運用を開始し，その後，行動履歴情報がある程度蓄積されたのちに，それらを活用できるタグスコア方式・ユーザ間協調フィルタリング方式のハイブリッド方式（組み合わせ方に関して，今回は結果を交互に並べる手法を採用）へと変更した．その後，タグスコア方式の TF-IDF 方式への変更，なお多様性を増すためのスコア減算方式の導入などを行った．

## 4. 評価

### 4.1 推薦内容の精度

精度を評価するため，利用者uへの推薦リストを$L_u$，$L_u$の$_j$番目(j∈{1,…})の要素であるコンテンツを$L_{u,j}\in L_u$，利用者uがフィードバックを行っているコンテンツの集合をFB(u)，そのときのコンテンツiへの評価値をF(u,i)として，以下の 2 つの評価指標を用いて評価実験を行った．それぞれの指標は各利用者への推薦リスト毎に計算し，その平均を用いる．

Precision (精度):

$$P = \frac{|\{i \in L_u | F(u,i) > 0\}|}{|L_u|}$$

Normalized Discounted Cumulative Gain (nDCG)[3]:

$$nDCG = \frac{1}{Z_u}\sum_j \frac{F(u,L_{u,j})}{\log_2(j)}$$

ここで$Z_u$は完全にスコア順になっているリストのときの nDCG が 1 とするような定数．

2014 年 1 月 16 日から 2014 年 4 月 24 日までの期間で，実際に推薦リストを作成した日付にしたがって，推薦方式毎に最大 50 コンテンツからなる推薦リストを生成し[3]，これら指標を計測した．評価は，推薦リストを作成した日から 1 ヶ月の間のフィードバックを用いて計算した．評価対象ユーザは，評価期間に 1 回以上ポジティブなフィードバックをしているユーザとした．

[1] なお，両ベクトルとも単位ベクトルに正規化済みのため，$|v_u|=|v_c|=1$ である

[2] なお、TF-IDF 方式で述べた特徴ベクトルと異なったものを利用しているのは開発が並行していたためである．

[3] 実際にその期間に利用者に対して推薦されていたリストとは異なることに注意

結果を図 4 上に示す．横軸は 2014 年 1 月 16 日からの経過日数を表している．方式名のうち uu はユーザ間協調フィルタリング方式を表している．

結果を見ると，評価期間の後半で全ての推薦方式について評価値が減少していく様子が見られる．これは，実験の初期段階で活発にコンテンツに対してフィードバックしていたのが，後半に行くにしがたい，見たことのあるコンテンツの割合が増えるなどして，フィードバック数が減っているためと考えられる．

手法間の比較では，類似する手法であるタグスコア方式と TF-IDF 方式は評価値も似た振る舞いを示している．タグマッチ方式の評価値が低いのは，ルールに合致するコンテンツ全体からランダムにコンテンツを選択するため，比較的多様なコンテンツをユーザの推薦リストに含むことが出来る一方で，明確に優先順位を決める他方式より精度が低下していると考えられる．

ランキング評価指標である nDCG に関しては，タグスコア方式と TF-IDF 方式が有利と考えていたが，実際はユーザ間協調フィルタリングが全体を通して評価が高かった．その一方で，実験の後半部分で，学習に用いたユーザの行動履歴データが増えるにつれ，タグスコア方式や TF-IDF 方式の評価値が上がる傾向も見られた．

なお，今回の評価は，実証実験で実際に運用していた推薦方式がどれであるかに影響を受けている可能性が存在する．実際に稼働していた方式によって生成されたリストに含まれていたコンテンツは利用者からのフィードバックを得やすいため，評価上も有利になりやすい可能性がある．特に，タグマッチ方式はランダム性が大きく実際に生成されていたリストと，評価実験で生成されたリストの違いが大きいと考えられる． これを根本的に解決するためには，オフラインの評価ではなく，A/B テストなどを通じたオンラインでの評価が必要と考えられる．

### 4.2 推薦内容の多様性

今回，コンテンツベースフィルタリング手法では推薦内容の多様性を向上するための対策を導入したが，一方で推薦内容の精度と多様性の間にはトレードオフの関係があると考えられる．また，ユーザ間協調フィルタリングはその仕組みから，意外性のあるコンテンツを推薦する可能性があり，それにより推薦内容が多様なものになる可能性がある．

そこで，各推薦方式によるリストの多様性の尺度として，「リスト中のコンテンツをランダムに選択し，コンテンツに付与されているタグをランダムに選択する」際の選択されるタグの情報量[4]を用い， nDCG と比較した結果を図 5 に示す．

この結果からは，TF-IDF 方式は nDCG は最も高かった一方で、多様性は他方式に比べて低い結果となっている．また，ユーザ間協調フィルタリングは精度と多様性の双方でバランスの良い結果となっている．

## 5. まとめ

本稿では，川崎グランシティモールの実証実験における推薦の概要，実験を通じて得られた知見や工夫について報告した．

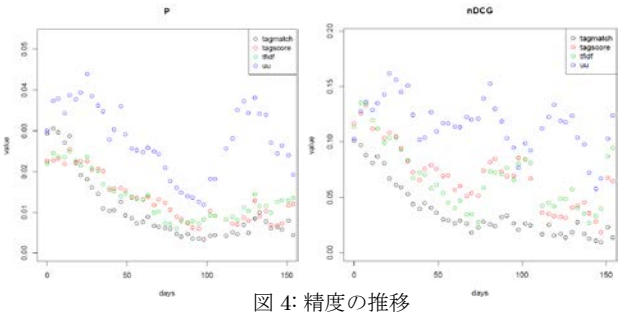


図 4: 精度の推移

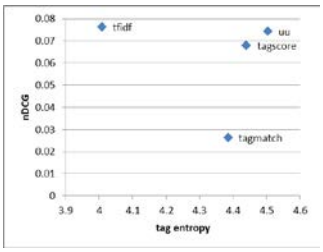


図 5: 各推薦方式の精度と多様性

予備的な評価では，今回の実証実験ではほぼ一貫してユーザ間協調フィルタリングの性能が高く，同時に多様性のある推薦内容となっていた，一方で行動履歴データが増えるにつれ，行動履歴情報を用いたコンテンツベースフィルタリング手法の性能の向上がみられたといったことが分かった．

また，本稿ではコンテンツへの明示的なフィードバックのみを用いた推薦を行ったが，実際には「コンテンツの詳細画面を開いたが，フィードバックを行わなかった」といった操作履歴の情報も暗黙的なフィードバック情報として推薦に活用できる可能性があり，別途分析を行っている．また，特にこれらの情報を統合して活用するにあたっては，今回用いた古典的なコンテンツベースフィルタリングおよび協調フィルタリングではなく，より機械学習的な手法を用いることが有用であると考えている．

### 参考文献

[1] 川崎市，川崎駅周辺地区スマートコミュニティ事業　川崎グランシティモール実証実験参加者募集について http://www.city.kawasaki.jp/200/page/0000053904.html

[2] K. Spärck Jones, “A statistical interpretation of term specificity and its application in retrieval, ” Journal of Documentation, vol. 28, no. 1, pp. 11-21, 1972.

[3] K. Järvelin and J. Kekäläinen. “IR evaluation methods for retrieving highly relevant documentsm,” In Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval", SIGIR'00, pp. 41–48. ACM, 2000.

[4] J. Konstan and M. Ekstrand, “Introduction to Recommender Systems,” Coursera, [Online] https://www.coursera.org/course/recsys.