

オープンキャンパス案内支援システムに基づく 情報推薦アルゴリズムの検討

小澤 和也[†] 田高 周[†] 南谷 和毅[†] 小舘 俊[‡] 岡村 容伸[†]

安澤 隼人[‡] 佐藤 壮真[‡] 池野 直人[‡] 大林 武[†] 木下 賢吾[†]

東北大学大学院情報科学研究科[†] 東北大学工学部情報知能システム総合学科[‡]

1. はじめに

オープンキャンパスは、学校への入学を検討している生徒が数多く参加するイベントであり、参加者が関心を持つ分野に応じた「個別化案内」を実施することが望ましい。そこで我々は、オープンキャンパス展示の案内を支援することを目的として、協調フィルタリングによる展示の推薦を行う web アプリケーション SmartCampus を開発した。本稿では、SmartCampus の概略とオープンキャンパスで実施した際のデータでの利用者の動向、情報推薦アルゴリズムの課題について述べる。

2. SmartCampus

2.1 SmartCampus の概要

SmartCampus は、オープンキャンパス用に開発された展示案内支援システムである(図 1)。

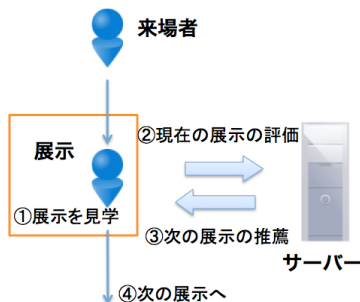


図 1. SmartCampus 概略

利用者は、各展示に設置された QR コードをスマートフォンなどの携帯端末で読み込み SmartCampus にアクセスすることで、展示内容の評価を行い、その評価内容に基づいた展示の推薦情報を得ることができる。また、人気の展示のランキングをリアルタイムで閲覧することも可能である。このシステムを導入することによって、利用者と開催者側双方に次のようなメリットが期待できる。

A study of information recommendation algorithm based on a new open campus guidance supporting system
[†] Kazuya Ozawa, Shu Tadaka, Kazuki Minamiya, Yasunobu Okamura, Takeshi Obayashi, Kengo Kinoshita, Graduate School of Information Sciences, Tohoku University.
[‡] Shun Kodate, Hayato Anzawa, Soma Sato, Naoto Ikeno, Department of Information and Intelligent Systems, Tohoku University.

利用者(来場者、主に高校生)側のメリット

- ・自分の興味にあった展示を発見できる
- ・人気の展示を知ることができる

運営(学校)側のメリット

- ・来場者評価を展示者へフィードバックできる
- ・適切な展示案内が生徒の入学意欲向上に繋がる

2.2 協調フィルタリングによる展示情報推薦

内容ベースフィルタリングにより展示を推薦する際には、2つの課題を検討する必要がある。1つは、各展示を評価・分類する指標が存在しない点、もう1つは、推薦時に参加者の嗜好を判断する難しさである。

これに対して我々は、アイテムベース協調フィルタリングを採用した。このアルゴリズムは、展示への利用者の評価情報を使用し、高く評価された展示との類似性を考慮して推薦する。評価情報から展示間の類似度と利用者の嗜好を判断するため、前述の2つの課題が解決される。

具体的には、利用者と展示の間に、評価値の対応関係を考える。この関係から利用者 u_i に対して、まだ訪問していない展示 e_j を推薦する度合いを以下の $score_{i,j}$ で定義する(式 1 と図 2)。

$$score_{i,j} = \sum_k s_{i,k} r_{j,k} \quad (式 1)$$

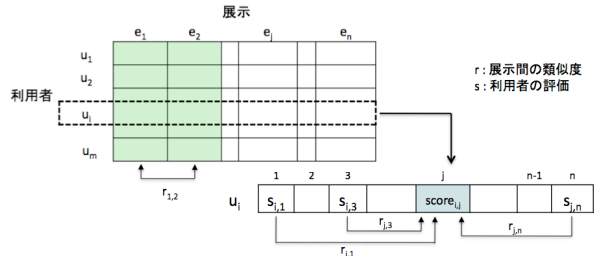


図 2. 評価情報からのお勧め度計算

$s_{i,k}$ は利用者 u_i の展示 e_k に対して投票を行った際の評価値であり、今回は 0, 1, 2 の 3 段階とした。 $r_{j,k}$ は展示 e_j と e_k の 2 つの展示を訪問した利用者らによる評価値のピアソンの相関係数である。この $score_{i,j}$ は、利用者がこれまで高く評価した展示と類似した展示ほど高くなる。推薦を行う際には、利用者が評価していない全ての展示に対して計算を行い、高い $score_{i,j}$ を持つ展示から推薦をおこなう。

3. 実施結果

3.1 実施内容

2014年7月30・31日の東北大学工学部情報知能システム総合学科のオープンキャンパスでSmartCampusを運用した。入場者4,307名中、851名がSmartCampusを利用し計2,140票の投票情報が得られた(図3)。

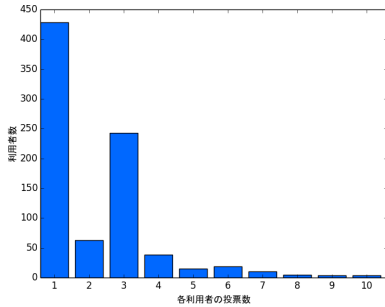


図3. 利用者の投票数分布

投票のインセンティブとして3票以上投票することでプレゼント当選への応募権利が獲得できるキャンペーンを実施した。その影響と思われるピークが投票数3に見られる一方、多くの参加者が1度の投票でやめていることが見て取れる。

3.2 投票データの解析結果

利用者と展示の関係を見るため、相関係数を使用して距離を定義し、クラスタ解析とヒートマップの描画を行った(図4)。

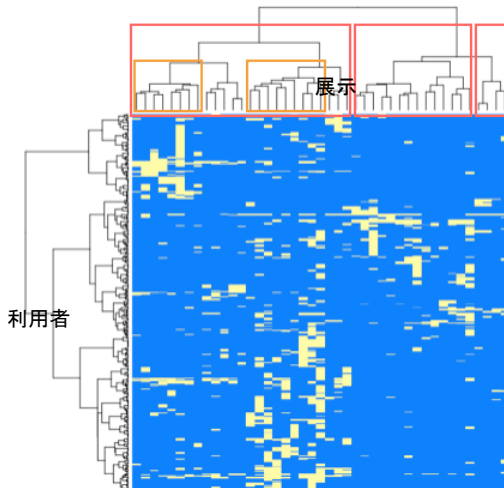


図4. 投票情報によるヒートマップ
(黄：高評価、青：低評価・無投票、赤枠：同建物、橙枠：同分野)

利用者と展示の関係として、評価が存在する場合にはその値(0~2)を使用し、評価されていない場合には-1を使用した。また、利用者と展示のクラスタ間の分離度を一定以上保つため、3票以上投票した利用者360名と10票以上投票があった43展示のみを利用した。この図からは、投票に偏りが大きい点と、展示がある建物ごと

のクラスタ(赤枠)や同じ階/部屋のサブクラスタが見られた。しかし、一部には位置によらず似た展示分野のクラスタ(橙枠)も観察することができた。また、利用者側のクラスタでも同様に訪問した展示の位置的な偏りが存在することが観察された。

3.3 推薦システムの効果

各利用者が展示間の移動の際に、システムによる推薦に従った割合を求めると、0.103と非常に低かった。この結果から、システムの推薦情報は、利用者が次に訪問する展示を決める際にほとんど影響を与えなかったと考えられる。

原因としては相関係数の評価の問題が考えられる。相関係数は、どちらか一方の評価データが一定であると定義できないため、本システムでは、相関係数が定義できない場合には0としている。全展示ペアで評価値の相関係数を計算すると、95%の展示間で0となっており、その結果として、一部の展示ばかりが推薦される状態となっていた。

評価値が一定である問題の他にも、展示ペアの相関係数の計算時に両展示を評価している利用者数自体が少ないことも問題である。相関係数の適切な初期値を検討することが必要だと考えられる。

4. 考察・まとめ

オープンキャンパス案内支援システムを開発し、実際に東北大学工学部情報知能システム総合学科のオープンキャンパスで運用をおこなった。そのデータを解析した結果、参加者が訪問した展示には位置的な偏りが見られた。また同じ建物内など距離が近い範囲では、似た分野の展示を連続して訪問する傾向も見られ、利用者が一定の興味のもとに行動し、無作為に近辺の展示を巡回している訳ではないことも分かる。

これらの結果から、より良い推薦システムにするための課題としては、展示内容だけでなく、物理的に近いか否かの情報も考慮して推薦アルゴリズムを改善する必要があることが分かる。また、展示間の相関係数は利用者の評価に依存するため、より多くの利用者を獲得する必要がある。今回の実施では、1度の投票でやめている利用者が約半数であり、継続的に利用してもらうためにアプリケーション内の利用促進の工夫も必要であることが示唆された。

参考文献

[1] smartcampus-framework, <https://bitbucket.org/smartcampus-framework/smartcampus/overview> (2014年12月1日)
[2] Toby Segaran, 當山仁健, 鴨澤眞夫, “集合知プログラミング”, オライリージャパン, 2008