

タグによる嗜好抽出を用いた観光イベント推薦システムの提案 Recommendation System Using Preference Extraction by Tagging

吉田 耕陽[†] 遠藤 聡志[‡] 當間 愛昇[‡] 赤嶺 有平[‡] 山田 孝治[‡]
Koyo Yoshida Satoshi Endo Naruaki Toma Yuhei Akamine Koji Yamada

1. はじめに

祭り、音楽、講演などのさまざまなイベントは、地域観光資源の1つである。イベントと観光者を結びつけるためには、ユーザーの嗜好やイベント間の関係性を反映した推薦が必要である。本研究では、イベント情報を固化的に配信、推薦するシステム構築のため、沖縄県で実際に過去に行われたイベントデータを用いてイベント配信データフォーマットをXMLにより定義し、それを活用したイベント情報の性質に有効な推薦システムの提案を行う。

提案するシステムでは、イベントのメタデータをタグとして表し、それを利用したユーザーの嗜好抽出によりイベント情報を推薦する。ユーザーアクションから、タグの重みおよびユーザー嗜好を収集し、新規イベントに対しても、そのメタデータとの比較により推薦度を抽出することが可能となる。本研究でのアプローチはコンテンツベースフィルタリングと言えるが、タグベクトルの作成にユーザーの協調行動を取り入れることで、情報の意外性や流行なども考慮している。被験者実験では、実験用サイトを閲覧することによりユーザーの嗜好を抽出し、推薦に対するアンケートを行った。

2. 提案手法

2.1 イベント情報について

イベント情報の構成は多種多様であり、イベント名称、開催場所、日時、出演者などの基本情報の他、チケット料金やスポンサー情報なども含まれる。扱う際の特徴として、複合的要素や例外的要素も含まれ、カテゴライズが難しい。また、イベントの特性からデータ投入の迅速性が求められ、井上ら[1]によるデータフローのワンストップ化を目指すためのデータフォーマットが提案されている。本研究は、イベント情報の多様性、開催日を過ぎると陳腐化する特性に着目し、ユーザーへの推薦システムの提案を試みる。イベント情報の性質から、従来の推薦手法が最適とは言い難い。代表的な推薦アルゴリズムである協調フィルタリングには、新規アイテムに対してデータが乏しく、効果的な推薦ができないというコールドスタート問題がある。情報の提示から開催日までの時間制限があるイベント情報に対しては、この問題は深刻であり、解決法としてタグの利用を提案する。

利用するデータとして株式会社エフエム那覇の運営する沖縄県内のイベントカレンダー「筥柄暦」を元に、イベントデータを再定義した。筥柄暦では、イベントデータ、開催場所データ、アーティストデータの3つが独立に定義さ

れている。イベントのみの情報では疎なデータも多いため、イベントの開催場所や出演者のタグ情報を紐付けて表現することで情報量を補完する。

2.2 タグの利用

イベントの検索や分類を目的として、データ定義の段階で3つのデータに対してタグ情報が定義してある。これは、1) イベント情報生成の際に機械的に付与、2) データ入力者が付与、3) イベント企画者、参加者が宣伝等のためにWeb上で追加、といった作成方法が上げられる。例えば、出演アーティストや、開催場所情報、地域などは自動的に付与される。この操作で、朝・昼・水曜日・週末といった時間情報や、アート、音楽、芸能などのジャンル、また、ロック・ジャズなど出演者のジャンルなど、様々な種類のタグをもったイベントタグセットを作成する。また、重みつきタグはユーザーに対しても付与され、それらを用いてユーザー・イベント間の関連度を抽出し推薦を行う。以下に、関連度抽出の概念を示す。

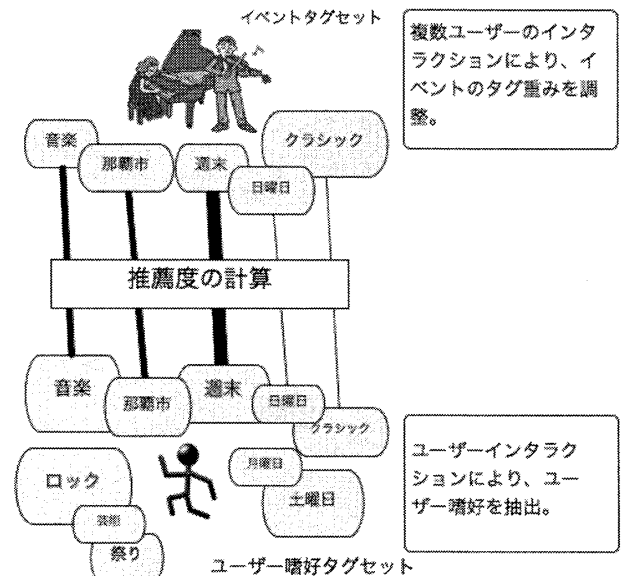


図1 関連度抽出の概念

2.3 ユーザー嗜好タグベクトルの作成

イベントに対するユーザーのインタラクションを用いて、イベントに含まれるタグをユーザーに付与し、ユーザー嗜好として推薦に利用する。本稿では、シンプルなアクションとして、イベント閲覧を利用する。イベントを閲覧したユーザーに対して、イベントに含まれるタグを付与する。イベントの閲覧を繰り返すことで、ユーザーのタグの重み

[†] 琉球大学 大学院 理工学研究科 Graduate School of Science and Engineering University of Ryukyus

[‡] 琉球大学 工学部 情報工学科 Department of Information Engineering University of Ryukyus

が追加される。また、文章中の特徴的な単語を抽出するためのアルゴリズムである tf-idf を応用し、イベント集合に対するタグの逆出現頻度を乗算することにより、「音楽」、「週末」のような多くのイベントに含まれるタグ重みを減らすよう調整する。

ユーザーに対するあるタグの重みは、次の式で表される。

$$user_weight(t) = count(t) \cdot idf(t) \quad (1)$$

$$idf(t) = \log\left(\frac{\text{全イベント数}}{\text{(タグ}t\text{を含むイベント数)}}\right) \quad (2)$$

(1)式によって求められた重みつきタグの集合をユーザーの嗜好として推薦に利用する。

2.4 イベントタグベクトルの作成

イベントタグに対しても、同様に重みを付与する。全ユーザーのアクションに対して重みをつけ、上記と同様なidfによる調整を行う。以下がイベントタグの重みの式である。

$$event_weight(t) = total_count(t) \cdot idf(t) \quad (3)$$

この操作により、参加ユーザーの集合知を利用し、多くのユーザーからアクセスの多いタグを含むイベントの推薦確率が高まることが期待される。

2.5 推薦イベントの抽出

ユーザータグとイベントタグのベクトルに対して値が0~1の間をとるよう正規化を行い、(4)式により内積をとることでスコアを計算する。また、閲覧済みのイベントに関しては、内容を知っていると仮定してスコアを0とする処理を行う。

$$score(U, E) = \sum U^{tag_w} \cdot E^{tag_w} \quad (4)$$

3. 実験

3.1 実験方法

筧柄暦のイベント情報配信 API を利用し、上記システムを実装したイベント閲覧サイトを構築した。データには予めテキスト処理により、時間的情報や、開催場所・出演者情報、曜日などを抽出し、タグを付与してある。ユーザーは、タグや開催期間での絞り込みを通してイベントの検索、閲覧が可能である。被験者に対して、それぞれの興味のあるイベントを自由に閲覧してもらい、嗜好タグベクトルを作成する。抽出したユーザータグベクトルに基づき、2010年6月の第2週、第3週、第4週における推薦度の高い上位10イベントずつ情報を提示し、推薦に対するユーザーの評価アンケートとして、「有益な情報だった」「どちらともいえない」「有益な情報ではない」の3段階に分けて回答を得た。嗜好抽出の妥当性を考察するため、イベントに定期的に参加し、実際にライブイベント企画など経験のある被験者を選択しアンケートを行った。

表1：実験データ

閲覧可能イベント数	4739
推薦対象期間イベント数	1432
ユーザーに付与されたタグ数	186

3.2 実験結果および考察

テストユーザーに付与されたタグを表2に示す。被験者の居住区周辺の地名や、音楽関係、週末のタグが上位に現れることが確認できた。また、平日やスポーツ、芸能関係のタグは値が低くなった。被験者の過去に参加したイベントなどに感覚的に近いタグ集合であり、改善の余地はあるものの、提案手法は推薦のための嗜好抽出が可能であると考えられる。

推薦に関する実験としてユーザーベクトルと期間中のイベントベクトル間類似度に基づき推薦を行った結果、抽出されたイベントのほとんどが音楽イベントであり、被験者の住む地域周辺のイベントが多いことが確認できた。また、表3で示すように全体の40%のイベントに対して有益だったとの回答を得ることができた。有益でない理由として、「音楽」という広いジャンルのイベントではあるが、その中の出演者やジャンルが好みではないという意見を得た。多くのアクセスを行いデータ収集することによるユーザー嗜好ベクトルの洗練や、出演者・開催場所タグを増やすことによりこの問題は改善すると考えられる。

表2：ユーザーに付与されたタグの一部

重みの大きいタグの例	ロック, K-mind, フォーク, 金曜日, ライブスポット, 夜, 土曜日, 宜野湾市, HY, 音楽, 北谷町
重みの小さいタグの例	沖縄市, ホテル, 水曜日, 浦添市, サッカー, 三線, フリマ, ルネッサンスリゾートオキナワ

表3：アンケート結果

	有益である	どちらともいえない	有益ではない
評価数	12	7	11

4. まとめ

本研究では、重み付きタグを利用したユーザー嗜好の抽出、イベント情報推薦システムの提案・実験およびその評価を行った。

今後の課題として、単純なイベントアクセス以外でのユーザーインタラクションの利用による推薦制度向上や、データ収集期間を限定することによる注目イベントの抽出が挙げられる。

参考文献

- [1] 井上 明, 吉村 孝昌, 永井 智子, 石田 達郎, 佐野 嘉紀, 長澤 知津子, 小林 聡, 金田 重郎, “イベント情報のワンストップ化-NewsMLを用いた広報情報発信システム-”, 情報処理学会・情報システムと社会環境研究会(SIG-IS), 情報処理学会研究報告, 2005-IS-91, 2005年3月
- [2] 株式会社エフエム那覇 “筧柄暦”, 論文誌名, <http://event.uruma.jp/>