

## タグを用いた嗜好抽出による観光イベント推薦システム

吉田 耕陽<sup>†1</sup> 遠藤 聡志<sup>†2</sup> 當間 愛晃<sup>†2</sup>  
赤嶺 有平<sup>†2</sup> 山田 考治<sup>†2</sup>

イベント情報は多種多様であるが、ユーザーの嗜好に応じてこれらの推薦を行うことで、情報収集を補助し、参加意欲を高めることが期待される。そこで本研究ではイベント情報の特徴をタグとして抽出し重み付けを行い、それを用いた関連度計算によってイベント情報を推薦する手法の提案を行う。イベント情報を配信する Web サイトを構築し被験者実験を行い、既存のアルゴリズムと比較することで、提案手法の有用性を示した。

### Event Reccomend System based on Preference Extraction Using Tags

KOYO YOSHIDA,<sup>†1,?</sup> SATOSHI ENDO,<sup>†2</sup>  
NARUAKI TOMA,<sup>†2</sup> YUHEI AKAMINE<sup>†2</sup>  
and KOJI YAMADA<sup>†2</sup>

Event information include various kinds of features. For recommendation of events, it is necessary to extract user's profile correctly. In this paper, we propose method to extract user's preference using tags and recommend system that based on similarities between user and events. Based on user testing, our system provides more suitable recommendations, than collaboration filtering recommendation methods.

<sup>†1</sup> 琉球大学 大学院 理工学研究科

Graduate School of Science and Engineering University of Ryukyus

<sup>†2</sup> 琉球大学工学部情報工学科

Department of Information Engineering University of Ryukyus

### 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、イベント情報を効果的に配信、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用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。

近年、情報推薦の研究は活発化し、Amazon.com に代表される商品推薦や、SNS におけるユーザー推薦、また Web ページ推薦<sup>1)</sup> など様々なコンテンツについての研究がなされている。イベント情報を効果的に推薦するには情報コンテンツの集約・詳細化が求められると筆者は考える。イベント主催者や開場・自治体・観光会社などの Web サイトがそれぞれ個別で配信するデータを一元的に管理し、必要な情報を API を通して配信することで、データ入力者の負担を軽減し個々の情報の詳細化を行うことができる。集約されたイベント情報の一部を他の観光情報サイトに配信して利用する研究として、筆者は月 1000~1500 件の沖縄県のイベントを網羅するイベントカレンダー『篋柄暦』<sup>7)</sup> のデータを元に、ワンストップサービスとしてホテルや自治体が各サイトが配信するイベント情報を推薦する手法を提案した<sup>2)</sup>。本研究の目標は、一定地域のイベント情報を網羅する詳細データを収集し、再掲載目的でデータを利用する媒体やイベント参加者にそれぞれが必要とする情報を配信するシステムを構築することである。

本稿で提案するシステムでは、Web サイト上でのイベント参加者への情報推薦に着目し、イベントの場所・ジャンル・時間情報・出演者などをタグとして表し、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。ユーザーのインタラクションからタグの重み調整およびユーザー嗜好を収集し、新規イベントに対してもタグとの比較により推薦度を抽出することが可能となる。

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、ユーザー嗜好からイベント関連度を計算した。提案手法と一般的なアルゴリズムである協調フィルタリングの 2 つでそれぞれお推薦イベントリストを作成し被験者にアンケート調査を行い、提案手法の性質、有用性を示した。

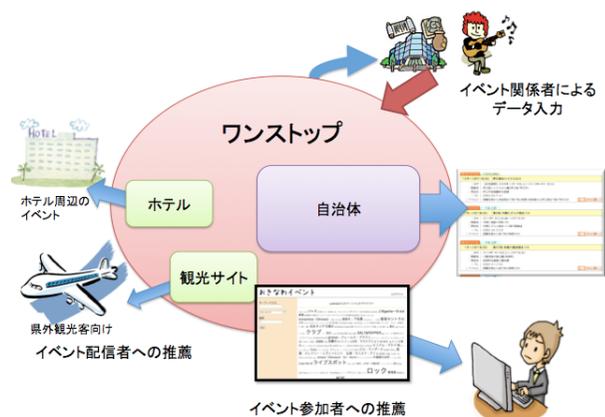


図1 イベント配信の概念図

## 2. 研究背景

### 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、カテゴリ化が難しい。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、イベント開催日を過ぎると情報としての価値が極端に下がり、陳腐化するという性質を持つ。イベント情報を一元化しデータ再投入・再加工を減らす目的として、井上ら<sup>3)</sup>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。また、ブログなどのテキスト情報からイベント情報を抽出する試みも活発で、森近ら [3] は、機械学習と集合知を組み合わせた情報抽出手法を提案し、イベント情報構造化システムの実装を行っている。

### 2.2 情報推薦について

推薦アルゴリズムの代表的なものとして 1) 協調フィルタリング、2) コンテンツベースフィルタリングの 2 種類がある。前者は、ユーザーのコンテンツ評価履歴を利用しあるユーザと

嗜好の類似したユーザーから推薦コンテンツを抽出する手法で、コンテンツの内容を考慮しないことも特徴である。しかし、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り、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る。後者は、コンテンツの属性に基づきコンテンツ間の類似度を求め推薦に利用する手法で、適切な特徴抽出が難しい、似たようなコンテンツばかりが推薦されやすいといった欠点がある。推薦システムの精度を表す指標の代表的なものとして、適合判定されたアイテムのうち利用者に有用である情報の割合である適合率 (precision)、すべての適合データのうち実際にシステムが抽出できた割合である再現率 (recall) がある。また、推薦特有の評価指標として、推薦されたアイテムの目新しさの指標である新規性 (Novelty) やユーザー本人だけでは予測・発見することができないアイテムを提示するセレンディピティ: 意外性 (Serendipity) も重要とされている。先述した 2 つのアルゴリズムでは、協調フィルタリングはユーザー自身が知らないジャンルでも他のユーザーの知識を通じて知ることができ、よりセレンディピティがある推薦ができるとされている。<sup>5)</sup> 他の評価指標としては、システムがどれだけのアイテムを予測可能であることを示す Coverage や、リスト内のコンテンツがどれだけ似ているか・似ていないかの指標である多様性 (Diversity) が挙げられる。

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実装を試みる。後述する提案手法は、コンテンツの属性としてタグを利用するためコンテンツベースのフィルタリングといえるが、タグの重みの調整に複数ユーザーの行動を取り入れることで、推薦の意外性や流行なども考慮している。

## 3. イベント情報推薦

### 3.1 利用するデータ

本研究では、実際に解析、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』のデータを利用した。筥柄暦は月に 1000~1500 のイベントを扱う網羅性の高いデータだが、時間と場所のみで詳細内容が無いデータなど、疎なデータも多く含まれる。

### 3.2 XML によるデータ構造の定義

現在、Web 上ではイベント情報の配信元として、自治体やイベント開場、イベント主催



- 2010年8月～12月のイベント数 5320
- 付与されたタグの種類 2942
- 1イベントに対するタグ数の平均 8.018個

タグ	付与されたイベント数	idf 値
音楽	2025	0.693
土曜日	997	1.609
島唄ポップス	364	2.639
美術館	268	2.944
ロック	263	2.996
宜野湾市	232	3.091
ライブ	222	3.135
アコースティック	128	3.713
フリーマーケット	35	5.024
桜坂セントラル	19	5.635
BEGIN	18	5.687
MONGOL800	2	7.886

表 1 付与されたタグの例

ユーザーがイベント閲覧等のインタラクションを行うと、係数を利用してユーザーの嗜好タグベクトルが調整される。一般性の高い単語ほど重みは低くなり、逆にアーティスト名や開催場所名は重みが高くなる。この性質により、あるアーティストが好きなユーザーや、音楽イベントに興味があり満遍なくチェックをするユーザー等の嗜好モデルのバランス調整が可能となる。

### 3.4 ユーザータグベクトルの作成

本システムではユーザーの嗜好モデルを重み付きタグのベクトルとして表現する。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される。ユーザーがタグにイベントを通してタグにアクセスした回数と idf を用いてタグごとの重みを調整し、インタラクションを繰り返すことで嗜好モデルが構築される。

ユーザーに対するあるタグ  $t$  の重みは、次の式で表される。

$$W_U(t) = \frac{I^t * idf}{\sum_{t_i \in U} I^{t_i} * idf} \quad (2)$$

ここで  $I_t$  は、タグを含むイベントに対するユーザーの閲覧・評価等のシステムに応じたインタラクションであり、タグの重みの平方の和が1となるよう正規化を行う。

### 3.5 イベントタグベクトルの作成

イベントには、3.3節の操作で、データ作成の段階でタグが付与してあり、ユーザータグ同様タグごとに重み調整を行う。ユーザータグベクトルがひとりのユーザーのインタラクションで決定することに対して、イベントタグは全ユーザーのインタラクションによる重み付けを行い、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。

$$W_E(t) = idf * \sum_U I^{t \in U} \quad (3)$$

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。

### 3.6 関連の抽出

ユーザータグとイベントタグのベクトルに対して、共通に持つタグの重みの積の総和を推薦度のスコアとする。ただし閲覧済みのイベントに関しては、スコアを0とする処理を行う。

$$score_{U,E} = \sum_{t \in U \&\&t \in E} W^U(t) * W^E(t) \quad (4)$$

ユーザーと推薦対象期間のイベントリストに対して上記の計算によって関連度の抽出を行い、関連度が上位のイベントをユーザーに提示するという形で推薦を行う。イベントそのものでなくタグの重み変更されるため、1度もアクセスされなくてもアクセスの多いイベントと共通のタグを持てば推薦確率は上昇する。例えば、アーティスト A が好きで多くチェックを行うユーザーがいると、『A が好きである』という嗜好を抽出し、A が出演するイベントならデータ追加と同時に推薦が可能となることが予想される。また、複数ユーザーのインタラクションで重み変更することで集合知の要素も取り入れ、データ収集期間を限定することで今旬のイベント・注目されるイベントの抽出も可能となる。

## 4. 評価実験

### 4.1 実験方法

提案手法を実装した被験者実験用サイトを構築し実験を行い、一般的なアルゴリズムである協調フィルタリングと比較・考察を行った。被験者には自由にイベント閲覧を行った後、収集した嗜好情報やアクセスログを元に、提案手法と協調フィルタそれぞれで抽出した推薦イベントリストを提示し、アンケートによる調査を行う。アンケート項目として、『有用だった』という項目のほか、発見性、意外度の調査のため『知っていた』『意外である』という項目も用意し、推薦システムの性質としての差異を比較した。なお、イベント情報の数の多さから、単純に『システム利用以前から知っていたイベント』を抽出するとその割合は非常に低くなると予想したため、関連する出演者や場所に関する情報を持っていることで『知っていた』と回答するよう実験設定を行った。意外性に関しては、自分のサイト閲覧行動に対する推薦として、妥当であるか意外であるか、という観点で評価するよう指示した。協調フィルタリングを用いた推薦リストを作成するためのライブラリとして、オープンソースの機会学習ライブラリである Apache Mahout<sup>9)</sup> を利用し、イベントに対するユーザーのアクセスログを用いて推薦リストを抽出した。

### 4.2 イベント推薦 Web サイト

実験・評価のため、上記アルゴリズムを実装したイベント配信 Web サイトを Ruby on Rails により構築した。3.2 節で解説したイベント XML データを篋柄層 API より取得し、イベント情報表示を行う。アクセス履歴やユーザー・イベントタグベクトルは、ローカルに保存し、実際のイベントデータとは分離した形で実装を進めた。

ユーザーはサイトにログイン後、日付情報やタグ情報からイベントを検索できる他、ユーザーのアクセス情報からユーザー嗜好を抽出し、スコアの高い順に推薦イベントリストの提示を行う。ユーザーのイベントに対するインタラクションの種類として、「お気に入り追加」や「評価」などが考えられるが、本システムではシンプルな形であるイベントに対する閲覧を行うことを評価の指標とした。また、抽出したユーザー嗜好を反映したタグクラウドをトップページに提示し再検索に利用させることで、ユーザーの負担を増やさずに暗黙的に嗜好モデルの補正を行うよう実装を行った。トップページに表示されるタグに実際に興味があればそれを用いて検索を行い、アクセスの無いタグは相対的に重みが減少する。

作成した Web サイトのスクリーンショットを図 3 に示す。



図 3 作成したイベント閲覧サイト

作成した Web システムにてユーザーに自由にアカウントを作成して利用をしてもらったところ、一定期間に 29 人のユーザーの利用があった。その中から 5 人に対してアンケート調査を行い、2010 年 9 月のイベントについて 2 つのアルゴリズムによって各 20 ずつ推薦イベントを抽出・提示した。回答者には、興味度を 1~4 までの点数付け、知っていた・知らなかった、意外であるか、という項目について評価してもらい、興味を持った度合いが 3 以上のイベントを正解として二値化し、推薦リスト内の適合率を導出した。推薦リスト内の知らなかったイベント、意外なイベントの割合についてもそれぞれ導出し、3 つの観点から比較を行った。

### 4.3 抽出したユーザー嗜好

ユーザーのアクセスによりどのように嗜好が抽出されたのか、2 人のテストユーザーについての例を表 2 に示す。

ジャンル、時間、市町村や地区、アーティスト名など、様々な角度からの嗜好を抽出されている。ユーザー A は音楽に関するアーティスト名やライブハウス、沖縄県中部周辺の地区、週末、夜といったタグが上位である。ユーザー B は、芸術や書店などでのイベントに興味を示している。嗜好がユーザーを表しているかという判断は主観によるところが大きい

ユーザー A	ロック, ライブスポット, 桜坂, SALT&PEPPER, 宜野湾市 北谷町, 桜坂セントラル, ジャズ, サカナクション, 夜, 県内
ユーザー B	講演, 映画, おもろまち, 博物館, 書店, 月曜日, 那覇市, トーク 火曜日, 沖映通り, 月曜日, ショップ, 徳尾聡 (陶芸家), 昼

表 2 ユーザーに付与された重みの高いタグの例

が、上位に出現する地名のタグはユーザーそれぞれの居住区に近いことなど、テストユーザーから感覚的に会っているという意見を得た。嗜好ベクトルを元に推薦リストを抽出したところ、ユーザー A に対しては音楽イベント、B に対してはアートや行事が中心の推薦リストが作成された。

## 5. アンケート結果と考察

	適合率	発見性	意外な推薦の割合%
提案手法	44%	86%	23%
協調フィルタリング	31%	83%	29%

表 3 アンケート結果

実験結果を表 3 に示す、提案手法が協調フィルタリングと同程度の意外性・発見性を維持するとともに、適合率は比較的高い結果となった。協調フィルタリングはアクセスの無いコンテンツが推薦されない弱点も見られたが、推薦手法では多くの種類のイベントが推薦され、Coverage が高いことも確認できた。一方、推薦されたリストを協調フィルタリングにおける嗜好が少数派のユーザーは適切な推薦が受けられないという問題が同様に提案手法にも見られた。被験者の多数は音楽関係を好む集団だったが、その他少数のユーザーに対しても音楽系のイベントが推薦され、そのイベントについての評価は不満足を示す値だった。ユーザーの協調行動を用いることで意外性のあるイベントが推薦しやすくなるが、同時に多数派の嗜好に影響され、ノイズとして現れやすくなることにも配慮が必要である。また、アンケート結果を分析すると、知っているイベントの推薦に対する満足度が上がりやすいという性質も確認できた。3.3 で『知っている』の基準を下げた影響があることも考えられるが、イベント推薦ではイベントの一部（出演者、場所など）を知っている状況で残りの情報を補完するという推薦方法も有効ではないかと仮定すると、時間・場所・ジャンルなどの

多様なタグ付け正確に行うことで満足度の高い推薦が可能になると考えられる。

## 6. まとめ

本稿では、観光イベントの配信・推薦のための XML フォーマットを定義した。また、ユーザーの嗜好を重みつきタグベクトルによって表現しユーザー・イベント間の類似度に基づく推薦システムの提案を行い、被験者実験によりその有用性を示した。今後の課題として、推薦精度向上のためのタグ自動生成方法の検討や、多様なユーザーインタラクションを用いてのユーザー嗜好抽出精度の改善があげられる。

**謝辞** 本研究は、株式会社エフエム那覇の収集したイベントデータを利用し実地されたものである。

## 参考文献

- 1) 丹羽 智史, 土肥 拓生, 本位田 真一, “Folksonomy マイニングに基づく Web ページ推薦システム”, 情報処理学会論文誌 47(5), 1382-1392, 2006-05-15
- 2) 吉田 耕陽, 遠藤 聡志, 當間 愛晃, 赤嶺 有平, 山田 孝治, “類似性に基づく観光イベント情報推薦システムの提案-”, 情報処理学会全国大会, 2010-3S-3, 2010 年 3 月
- 3) 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
- 4) 森近 憲行, 濱崎 雅弘, 亀田 堯宙, 大向 一輝, 武田 英明, “機械学習とユーザ知識を用いたイベント情報の構造化”, 第 2 4 回人工知能学会全国大会 (JSAI2010) 論文集, 2010
- 5) 神嶌 敏弘, “推薦システムのアルゴリズム”, 人工知能学会誌 23(1), 89-103, 2008-01-01
- 6) J Golbeck, B Parisa, J Hendler: Folksonomies-Cooperative Classification and Communication Through Shared Metadata, (2004).
- 7) 株式会社エフエム那覇 “窺柄曆”, <http://event.uruma.jp/>
- 8) 土方嘉徳, “嗜好抽出と情報推薦技術”, 情報処理, Vol.48, No.9 (2007).
- 9) Apache Lucene Mahout, <http://lucene.apache.org/mahout/>