

類似性に基づく観光イベント情報推薦システムの提案

A Proposal of Event Recommend System based on Similarity

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

1 概要

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 XML により定義し、それを活用したイベント推薦システムの提案を行う。提案する推薦システムでは、個々のイベントの要素である時間・場所・出演者などの特徴情報をタグとして抽出し、重みづけタグベクトルを作成する。さらに Simpson 係数を拡張したイベント間類似度を定義し、他のデータに対して関連度を抽出する。また抽出した関連度の利用例として、情報配信の際のフィルタリング手法を提案する。

2 イベント情報について

2.1 イベント情報とは

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、時間、出演者などの基本情報の他、チケット情報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多く、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、井上ら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 [1]。本研究ではイベント情報の多用途に着目し、ユーザーへの推薦システムの実装を試みる。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「筐柄暦」を元に、イベントデータを再定義した。筐柄暦では、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。イベント参加の意思決定には場所、出演者が大きく関ると考え、新しい定義でも 3 つに分割してそれぞれ定義した。

[†] 琉球大学 大学院 理工学研究所

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

2.2 イベント間の類似度

2.2.1 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) データ入力者が追加、3) イベント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挙げられる。例えば、出演アーティストや、開催場所名、地域などは自動的に付与される。ジャンルやタイプ等、付加情報は人の手に寄って付与される。それぞれのタグに対して、どの程度イベントの特徴を表しているかを考慮し、重要度が含まれる。今回はこのタグと重みを用いて関連度を抽出する。

2.2.2 関連の定義

共起度や類似性を計る尺度としてよく利用されるものに、Jaccard 係数、Simpson 係数、コサイン類似度などがあるが、今回のような重みのあるタグの関連度計算に直接使うことはできない。そこで、Jaccard 係数を拡張・一般化した tanimoto 係数を利用する。

しかし、tanimoto 係数は、Jaccard 係数の『A と B の間に大きな差があるとき類似度の値が低くなりやすい』[2] といった特徴がある。これでは、例えばイベント主催者がイベントの詳細タグを付与すると、そのイベントに対する類似度が低くなり孤立していく、といった問題がある。

この問題を解決するため分母に min をとっている Simpson 係数を拡張、一般化し、新しい係数を定義し、利用した。

$$sim(A, B) = \frac{\Sigma \min(X_{AK}, X_{BK})}{\min(\Sigma X_{AK}, \Sigma X_{BK})} \quad (1)$$

$$\text{但し } \min(\Sigma X_{AK}, \Sigma X_{BK}) > k$$

提案の係数は Simpson 係数に似て、『単独でのタグ重みの和が少ないイベントとの関連度は大きくなりやすい』という問題があるため、閾値を設定し、イベント間の関連を計算した。

3 実験・評価

3.1 提案手法の評価

筐柄暦の過去データを提案したフォーマットに整形し、6 つの比較用イベントを抜き出し、提案手法と tanimoto 係数による関連度を比較した。

イベントごとの説明及びタグ (重み) の一部

A	沖縄県内で開催された音楽イベント 音楽 (4), 県内 (8), 宜野湾市 (2), K-mind(3)
B	県内のライブイベントで、A と同じ市で開催されている 音楽 (4), 県内 (8), 宜野湾市 (2), HumanStage(3)
C	県外で開催された音楽イベント 音楽 (4), 県外 (8), BEGIN(2), 週末 (2)
D	小規模な祭り 祭り (4), 那覇市 (2), 行事 (4), 昼 (2)
E	自治体主催の大規模な祭り、音楽的要素も含む 祭り (4), 音楽 (2), 行事 (4), 週末 (2), 国際通り (3)
F	A に対して、ジャンル、メンバー名などが追加されている A のタグに、ロック (2)、ジャズ (2) 等を追加

	A	B	C	D	E	F
A	1.00	0.85	0.33	0.41	0.45	1.00
B	0.85	1.00	0.29	0.54	0.46	0.85
C	0.33	0.29	1.00	0.00	0.24	0.33
D	0.41	0.54	0.00	1.00	0.65	0.41
E	0.45	0.46	0.24	0.65	1.00	0.45
F	1.00	0.85	0.33	0.41	0.45	1.00

表 1: 提案係数による関連度

	A	B	C	D	E	F
A	1.00	0.80	0.10	0.44	0.41	0.54
B	0.80	1.00	0.10	0.48	0.43	0.44
C	0.10	0.10	1.00	0.00	0.08	0.06
D	0.44	0.48	0.00	1.00	0.66	0.28
E	0.41	0.43	0.08	0.66	1.00	0.26
F	0.54	0.44	0.06	0.28	0.26	1.00

表 2: tanimoto 係数による関連度

イベント A 両係数ともに、ジャンルや開催場所の類似度が近いイベントの類似度が高くなるという結果がでている。タグが付与されたイベント F と他のイベントとの関連を比較すると、tanimoto 係数では他の値との関連度が低くなっている。これでは入力者やイベント企画者が詳細情報をタグとして付与すればするほど孤立してしまうという結果である。一方、提案手法では、タグを付与していない A と同じ値がでている。この結果から、提案手法はイベント間類似度を抽出する係数として有用である。

3.2 関連度のフィルタ利用

筐柄暦は月 1000~1500 件のイベント情報を扱う網羅性の高いカレンダーである。その中から 1 部の情報を選択し、利用したいというニーズがある。例えば、沖

縄コンベンションビューローの沖縄観光情報 Web サイト「mahae plus」では、より公共性を重視した選択的
情報提供を必要としている。mahae plus では、県外観光客向けに、月に 20~30 件程度、公共性重視の県内イベントを提示している。筐柄暦から mahae plus が必要とするイベント情報を提供するためのフィルタリング手法として、提案した関係度抽出手法を利用する。

3.2.1 フィルタリング法において

mahae plus 担当者により毎月掲載されているイベントを分析すると、一定の特徴がある。それらの特徴を包含する架空のイベントを定義し、関連性の高い実在のイベントを mahae 掲載用リストとして抽出する。2009 年 10 月~12 月の開催イベントについて、関連度の高い上位 20 イベントを抽出し、mahae plus に実際に掲載されたイベントと比較した。フィルタリング性能の評価指標として、適合率 (precision) と再現率 (recall) を採用した。

	掲載されたイベント数	再現率	適合率
10 月	30	0.33	0.5
11 月	23	0.52	0.65
12 月	14	0.5	0.35

どちらも再現率か適合率が 0.5 を超える結果となった。また、mahae で『久米島町産業まつり』は表示されても『伊江村産業まつり』は表示されない場合があるが、抽出したイベントには両方とも含まれている。また、mahae に表示されても遜色ないような『第 44 回琉球古典芸能祭』、『第 33 回 東平安名崎タートルマラソン大会』などのイベントも同時に抽出することができた。この結果より、手動で行われてきたイベント収集・選別作業を支援するツールとして利用可能と考えている。

4 まとめ

イベントデータの定義、関連付け手法の提案およびその評価を行った。また、利用例としてのフィルタリング手法を提示した。今回は筆者の経験からタグの重みをつけたが、重みづけは関連度に大きく関わってくるため、タグの重み付けの最適化も必要になってくる。

参考文献

- [1] 井上明、吉村考昌、永井智子、石田達朗、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] 松尾豊、友部博教、橋田浩一、中島秀之、石塚満、「Web から人間関係ネットワークの抽出と情報支援」、人工知能学会誌 20 巻 1 号 E 2005 年