

イベント情報収集アプリケーションの開発

田中成典[†] 北川悦司[‡] 吉田博哉^{††} 中村健二^{‡‡} 松田貴寛[†]
 関西大学総合情報学部[†] 阪南大学経営情報学部[‡]
 神戸情報大学院大学情報技術研究所^{††} 立命館大学情報理工学部^{‡‡}

1. はじめに

近年、祭や花火大会などのイベント情報がポータルサイトを介して提供されている。これらのポータルサイトは、知名度が高く規模の大きなイベントの情報を提供している。そのため、知名度が低く地域に密着したローカルなイベント情報は、ユーザが容易に取得できない問題[1]がある。そこで、Web 上からイベントの知名度と規模に関わりなく、網羅的にイベント情報を抽出する研究[2]-[4]が行われている。これらの研究では、イベント情報の記載されたブログやソーシャルメディアを収集対象として、Web ページのリンク構造やタグ情報などから判定を行う。しかし、ローカルなイベント情報を提供する Web ページは、リンク構造やタグ情報に規則性がなく、これらの手法を適用できない Web ページも多い。そこで、本研究では、Web ページ内で特徴的な単語を用いて情報を抽出する手法[5]をイベントの分野に適用する。イベントのジャンル毎に特徴的な単語を抽出することで、Web ページの構造に依存することなく、Web 上からイベント情報を抽出し、ジャンル毎に分類する手法を提案する。そして、本手法にて収集したイベント情報を提供するアプリケーションの開発を行う。イベント情報は、手軽で素早い検索が望まれるため、需要が拡大している Windows Phone にてアプリケーションの実装する。



図1 実行イメージ

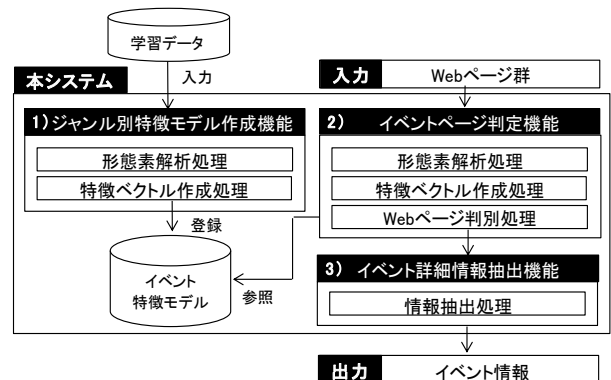


図2 本システムの概要

2. 研究の概要

本研究では、Web 上からイベント情報を抽出し、Windows Phone にて提供するアプリケーションを開発する。本アプリケーションの実行イメージを図1に示す。本アプリケーションの入力は都道府県名とイベントのジャンルとし、出力は入力した条件に該当するイベント情報とする。次に、本システムの概要を図2に示す。本アプリケーションを実現するため、本システムは、1) ジャンル別特徴モデル作成機能、2) イベントページ判定機能、3) イベント詳細情報抽出機能により構成される。

2.1 イベント特徴モデル作成機能

本機能では、イベント情報を掲載している Web ページの判定機能に利用するジャンル毎にイベントの特徴情報を作成する。まず、本機能の学習データとして、イベント情報を掲載している Web ページ群を入力する。次に、入力した Web ページ群の内容を形態素解析し、各形態素

Development of Smartphone Application for Collecting Various Events Information

[†] Shigenori Tanaka, Takahiro Matsuda
 Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-cho, Takatsuki City, Osaka 569-1095, Japan

[‡] Etsuji Kitagawa
 Faculty of Management Information, Hannan University, 5-4-33 Amamihigashi, Matsubara City, Osaka 580-8502, Japan

^{††} Hiroya Yoshida
 Institute of Computing; Graduate School of Information Technology, 2-2-7 Kano-cho, Chuo-ku, Kobe City, Hyogo 650-0001, Japan

^{‡‡} Kenji Nakamura
 College of Information Science and Engineering, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu City, Shiga, 525-8577, Japan

の TF/IDF 値を算出する。そして、ジャンル毎に TF/IDF 値の上位 1,000 件の形態素を抽出する。最後に、抽出した形態素を素性とする特徴ベクトルをジャンル毎に作成し、SVM (Support Vector Machine) を用いて、ジャンル別の特徴モデルを作成する。

2.2 イベントページ判定機能

本機能では、まず、入力された Web ページに対し、イベント特徴モデル作成機能と同様の処理を行い、特徴ベクトルを作成する。次に、イベント特徴モデルを参照し、SVM にてイベント情報が掲載されているかを判定しジャンル毎にイベント情報を分類する。

2.3 イベント詳細情報抽出機能

本機能では、イベントページ判定機能にてイベント情報が掲載されていると判定された Web ページから、イベント名、開催日時と開催場所に関する情報を正規表現にて抽出する。そして、それらの情報を Windows Phone アプリケーションへの提供に適した形式に変換する。

3. システムの実証実験と考察

本システムの有用性を実証するため、イベント情報を掲載している Web ページを判定する実験と、イベント情報をカテゴリ毎に分類する実験を行った。本実験では、正解データとしてイベント情報を掲載している Web ページをカテゴリ毎に 100 件、不正解データとしてイベント情報を掲載していない Web ページ 100 件を実験データとして用いた。なお、正解データは、各ジャンル名、開催日時と開催場所を組み合わせたクエリを用いて検索し得られた検索結果の上位 100 件の Web ページを用いた。

3.1 イベント情報判定の実証実験

本実験では、ローカルなイベント情報を掲載した Web ページの自動判定が可能かを評価するため、正解データから無作為に抽出した Web ページ 100 件と不正解データの Web ページ 100 件の合計 200 件を実験データとして評価した。判定精度は、適合率、再現率と F 値により評価した。実験結果を表 1 に示す。実験結果から、イベント情報の有無を適切に判定することがわかった。これらのことから、本提案手法がイベント情報を含む Web ページを判定する際に有用であることが示された。

3.2 イベント情報分類の実証実験

本実験では、既存のポータルサイトにに基づき、6 つのジャンルにイベント情報を分類可能であるかを評価した。本実験では、各カテゴリの正解データ 100 件と不正解データ 100 件を実験デー

表 1 実験結果

	イベント情報を含む Web ページ	イベント情報を含まない Web ページ
適合率	0.83	0.79
再現率	0.77	0.84
F 値	0.80	0.81

表 2 実験結果

	1	2	3	4	5	6
適合率	0.51	0.37	0.47	0.70	0.30	0.69
再現率	0.34	0.74	0.32	0.33	0.17	0.79
F 値	0.41	0.49	0.38	0.45	0.22	0.74

1…節句・年中行事, 2…市場・フリーマーケット,
3…娯楽施設, 4…花火・イルミネーション,
5…ネイチャー・アウトドア, 6…スポーツ

タとして用いた。分類精度は、適合率、再現率と F 値により評価した。実験結果を表 2 に示す。実験結果から、スポーツは適切に分類できたが、その他の 5 つのジャンルでは精度が大きく低下していることが分かった。これは、本提案手法で、分類のために用いたジャンルが、イベント分類に対して不適切であったことが原因と考えられるため、イベントのジャンルを適切にすることで改善できると考えられる。

4. おわりに

本研究では、Web 上からイベント情報を抽出し Windows Phone にて提供するアプリケーションを開発した。本システムでは、Web ページに含まれる特徴的な単語を用いて、イベント情報を掲載している Web ページを自動的に判定する。実証実験の結果、本提案手法がイベント情報を掲載する Web ページを識別する際に有用であることを実証する。しかし、イベントの分類においては、十分な精度を得ることができなかったため、今後は、失敗したイベント情報を分析し、適切に分類できるカテゴリの考案を行うことで、イベント情報の分類精度の向上を目指す。

参考文献

- [1] 安田宣仁, 戸田浩之: 検索位置のごく周辺を対象とした地理情報検索, 人工知能学会論文誌, 人工知能学会, Vol.23, No.5, pp.364-373, 2008.
- [2] Becker, H., Naaman, M., and Gravano, L.: Learning Similarity Metrics for Event Identification in Social Media, *Proceedings of the 3rd International ACM Conference on Web Search and Data Mining*, ACM, pp.291-300, 2010.
- [3] Rattenbury, T., Good, N., and Naaman, M.: Toward Automatic Extraction of Event and Place Semantics from Flickr Tags, *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, pp.103-110, 2007.
- [4] 石野亜耶, 難波英嗣, 竹澤寿幸: 旅行プログメントリからの観光情報の自動抽出, 知能と情報, 日本知能情報ファジィ学会, Vol.22, No.6, pp.667-679, 2010.
- [5] 寺口敏生, 田中成典, 西江将男: Web リソースと地理情報を活用した新店舗情報の自動抽出に関する研究, 知能と情報, 日本知能情報ファジィ学会, Vol.23, No.4, pp.411-427, 2011.