3N3-OS-28a-4

ソーシャルセンシングを用いた能動的イベント情報抽出手法の提案

Extracting Social Event Information by Active Accessing to Users in SNS

伊藤千輝 *1 Kazuki Ito　篠田 孝祐 *1 Kosuke Shinoda　小野 良太 *2 Ryota Ono　川村 秀憲 *2 Hidenori Kawamura　栗原 聡 *1 Satoshi Kurihara

*1電気通信大学 The University of Electro-Communications　*2北海道大学 Hokkaido University

Nowadays, there are many local event information introducing sites on the Internet. It has a lot of events such as festival of shrine, music live of individual citizen. Local information site, it is necessary to completely cover the area information. In the spread of the Internet, we are able to notice a lot of information in the Web, and we are able to search for information. But,local events information is not much, and it do not has been officially published on the Internet. But, information diffusion speed of local event information is slow, and event organizers do not have enough idea how to announce event information so widely by using the Internet.
In this paper, we will propose an event information extracting framework by using social sensing. In this framework, we actively access users to get detail information about events as well as extract information widely from SNS. As a result, we made it possible to obtain a wide variety of event information by this approach.

1. はじめに

インターネットの普及により，我々は多くの情報をホームページやソーシャルメディアで告知，検索できるようになり，観光分野においても個人や団体の発信した情報が活用されている．イベント情報は街に多く溢れているが，イベント開催側のノウハウの不足などから情報公開の遅れや，Web 上で公式な情報が未公開のものが多い．地域密着イベント情報サイトは，地域に隠れたイベント情報を告知し人を流入するため，日々イベント情報を抽出し続けている．

本研究では，地域密着イベント情報サイトの一つ「あなた情報マガジンびもーる *1」（以下「びもーる」）を研究対象とする．びもーるのシステムは，調和技研の関連研究機関である北海道大学調和系工学研究室での基礎研究に基づいて独自に開発した「興味解析エンジン」[1] を核に構成され，地域密着イベント情報サイトの課題解決に，少ない運営人数で大きな効果をもたらすため，IT を利活用した研究が多くされている．本研究では，「びもーる」のイベント情報を収集する既存システムの全体像と利用者の利用状況を分析し，びもーるの持つ課題を解決するための手法を検討した．

2. 地域密着イベント情報サイトの課題

2.1 「びもーる」の現状システム

「びもーる」の現在のシステム構成について説明する．システムは大きく分けて 3 つのシステムで成り立っている．まずはイベント施設などの Web ページや，地物新聞社からの情報提供でイベント情報を抽出する．次に得られたイベント情報を人の手作業でデータベースに整理し格納する．この時足りない情報や，ホームページ上で見つからなかったものはメールや電話などで直接問い合わせて情報を埋めるようにしている．作成されたイベント情報は開催期日の前までデータベース内で保管し，日時が近づくと情報推薦システムを通してサービス利用者にイベント情報を提供するようになっている．

連絡先: 電気通信大学大学院情報システム学研究科社会知能情報学専攻 栗原研究室 伊藤千輝 kazuki@ni.is.uec.ac.jp

*1　北海道大学発のベンチャー企業株式会社調和技研が運営するイベント情報サイト http://bemall.jp/

2.2 課題と解決手法の検討

「びもーる」の抱える 3 つの課題とその解決手法を述べる

1. 地域の情報を知るには土地勘が欠かせないが，少ない人数で情報源を網羅するには限度がある．そのため，地域住民によるＳＮＳへの情報発信を活用することで新たなイベント情報源の確保を行う．
2. イベントの種類は多岐にわたり，個人の趣味性の強い情報ほど Web 上では手に入れにくい．びもーるの利用者は多種多様なイベント情報を求め「びもーる」にアクセスする．そのため，ジャンルが偏らないように多種多様なイベントや趣味性の高い情報の取得を行うために，SNS から個人主催のイベント情報抽出を行う．
3. 多くの利用者の満足度を上げるには十分な量だけでなく質も重要である．そのため，掲載するイベント情報の記事をもとに，クラウドソーシング *2 に的アプローチより再度情報の編集を行うことで情報の質の向上を図る．

本研究では 1. と 2. に注目し，地域住民をセンサとして活用する「ソーシャルセンシング」を用いた地域密着イベント情報の抽出手法を提案する．

3. 関連研究

榊ら [2] は Twitter のユーザーをセンサとして実世界の観測データを抽出するソーシャルセンサを提案した．Twitter の情報から地震発生時の震源地を予測し，物理センサー同様の機能を持つセンサであると述べた．

長野ら [3] は鉄道などの交通情報へのソーシャルセンサの利用に関して，Twitter ユーザの人口集中による偏りや，ソーシャルメディア内の情報の信頼性などの課題があげられるが，他の物理センサと工夫して組み合わせることで解決できると述べた．ソーシャルセンサは物理センサ同様の機能を持つだけではなく，物理センサと組み合わせるなど利用方法を工夫をすることで，従来では観測しきれない現象を観測できる可能性を述

*2　インターネットを利用して不特定多数の人に業務を発注したり，受注者の募集を行うこと

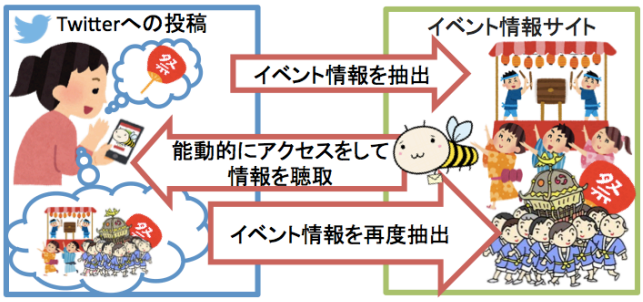

図 1: 新しいソーシャルセンセンシングの提案

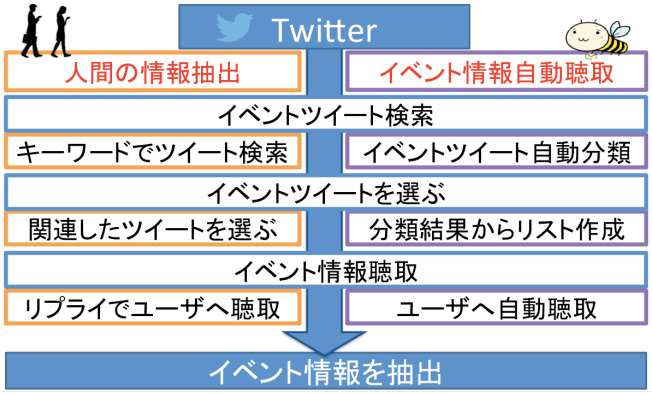

図 2: 自動抽出イメージ

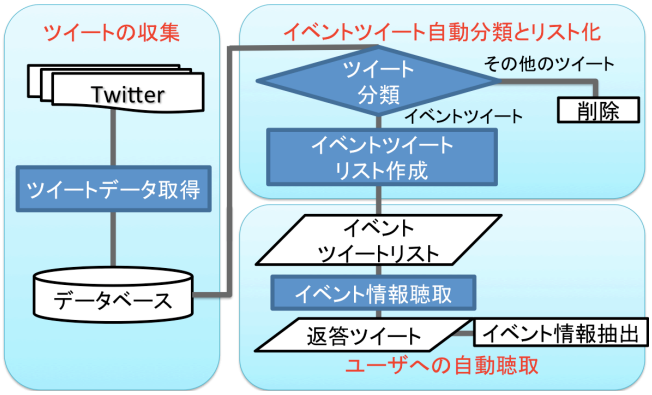

図 3: システムの全体構成

べた.

また，榊ら [4] は，ソーシャルセンシングを用いたイベント情報の抽出手法を提案した．榊らは，予告型イベントの収集とそのイベントに参加しているユーザと参加状態をテキストマッチングを用いて自動抽出させる基礎的研究を行った．

ソーシャルメディアからの観光情報の自動抽出を目的とした研究として，北海道観光振興機構と共同で普及に取り組んでいる，川村ら [5] の「キュンちゃんねる」の活用があげられる．利用者に Twitter や Facebook などの SNS の投稿本文に「＃キュン旅 (地名)」とハッシュタグをつけてもらいイベント情報を自動抽出した．実際には標準タグの利用者数の伸び悩みなどから，いかに標準タグを普及するかが課題となった．

4. 提案手法

従来のソーシャルセンシングを用いた手法はツイートからの情報抽出が主であったが，わずか 140 字の Twitter のツイートデータでは，情報が十分でないものや情報として不完全なものが多く情報抽出は困難である．また，抽出できたイベント情報があったとしても，情報量が少ないイベント情報は信頼性にかける．そこで本研究では，イベントツイートからイベント情報を抽出後に再度ユーザへの情報聴取を行い，ソーシャルセンシングを用いて Twitter 内のユーザへ能動的にアクセスしてイベント情報を引き出す手法を提案する．(図 1)

イベント情報を Twitter から取得するために我々人間は，イベントツイートをイベントのキーワードから検索し，ツイートの中からイベントに関するツイートを選び，最後に気になるツイートへリプライを用いて情報聴取を行う．自動聴取を行うためには，自動分類により得たイベントツイートから聴取するためのリストを作成し，ユーザへ自動聴取を行う．人間が Twitter 内から情報取得する際と同じ流れでイベント情報を聴取を用いて抽出する（図 2）.

5. システムの構成

本提案手法をもとにイベント情報を自動で聴取するシステムを作成した（図 3）.

1. **ツイート収集**

Twitter の居住地登録やツイートの位置情報の正確な登録者はごくわずかであり，プロフィールやツイートからユーザの居住地を選定するのは極めて難しい．「びもーる」の各地の Twitter アカウントのフォロワーは，各地域の住民が多く，イベント情報に興味を持ったユーザが多いと考えられる．そこで本研究では，「びもーる」の Twitter アカウントの中でも一番フォロワーの多い「びもーる札幌版」の Twitter アカウントのフォロワーからツイートデータを取得する．ツイートは常時収集され，データベース内に格納される．

2. **イベントツイート自動分類とリスト化**

取得したツイートデータから，イベント情報が含まれている可能性の高い「イベントツイート」と「その他のツイート」を分類し，イベントツイートのみを抽出する．イベントツイートの分類には，榊ら [4] の提案した「イベント三要素」を用いる．本提案手法では，パターンが有限である「開催場所」と「開催日時」と，パターンが無限である「イベント名」を用いてキーワードマッチングによってイートを分類する．「開催場所」に関してはイベント情報を抽出する Twitter アカウントが「びもーる札幌版」のため，札幌である．「開催日時」は（表 1）からテキストにマッチした日付表現から分類する．「イベント名」はイベント名の「手がかり語」である「開催」「イベント」「ライブ」を元に分類する．

次に抽出されたイベントツイートから，イベントを聴取するツイートリストを作成する．リスト作成時にスパムに間違われないように，同じツイートやユーザに対して同時期に何度もアクセスして聴取しないように設定した（図 4）.

表 1: 日時表現パターン

正規表現	取得できる日付例
mm 月 dd 日	12 月 1 日
mm/dd	12/12
mm.dd	12.10
dd 日	30 日
dd 日	25 日

Step1:ツイートデータベース内のツイートの重複除去
Step2:ツイートデータベースを日付順に並べる
Step3:前日 1 日分のツイートを取得しテーブルに書き出す
Step4:テーブルから同じユーザがツイートしたものを削除
Step5:テーブルから過去 7 日間でリストにツイートが入ったユーザを削除
Step6:テーブルから「キーワード」が含まれないものを削除
Step7:テーブルから日時表現パターンにマッチしないものを削除

図 4: イベントツイートリスト作成手順

3. **ユーザへの自動聴取**

作成したリストを用いて再度ユーザに聴取を行いイベント情報を取得する．ダイレクトメッセージを用いた聴取や，イベント入力専用ページのリンクをツイートに貼り付けて聴取する手法も試したが，API 制限 [*3] の問題や利用者の反応が悪く，リプライによる聴取を行った．作成されたリストへリプライを用いて聴取を行い,「1, イベント名 2, 日時 3, 場所 4, その他 (主催者，HP 等)」を埋めてもらうことで情報抽出処理の効率化を図った（図 5）．

聴取する時間を設定する際には，ユーザの聴取に反応しやすい時間帯に設定する．びもーるのユーザは，12 時台と 18 時以降がツイート数が多かったため，この時間を情報聴取時間とする．

その後聴取の返答を取得し，イベント情報を抽出し，実働する「びもーる」のイベント情報データベースへ登録する．

6. 実験と考察

本提案手法のイベントツイートの分類精度を確かめるために，提案手法である「イベント三要素」によるツイート分類と機械学習によるツイートの分類，二つの精度を比較する実験を行った．また，二つの分類手法でリストを作成して情報聴取を行い，分類手法によって利用者の反応の違いを確かめた．最後に，本研究手法である能動的なアクセスを用いたイベント情報聴取手法の評価を行った．

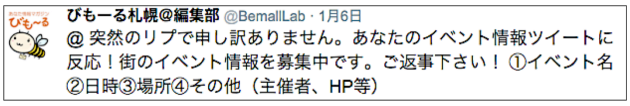

図 5: 実際のリプライイメージ

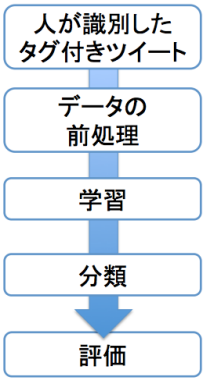

図 6: 機械学習によるツイート分類手順

6.1 イベントツイート分類手法

図 6 の手順で機械学習によるツイート分類を行った． 本提案手法のツイート分類と機械学習 3 つ挙げ比較した．機械学習には．分類時の重要項目が人間の目で見てわかり易い決定木，文書分類に多く用いられるナイーブベイズ分類器，2 値分類において高い精度を持つ SVM(Support Vector Machine) を用いた．機械学習の教師データとしてランダムに抽出した 2014 件のツイートデータを人間の手作業でイベントツイート（384 件）とその他のツイート（1630 件）に分類した．
学習データの前処理として，テキストの並び順は無視した単語の集合「Bag-of- Words」を用いたテキスト処理を行った．素性として，出現頻度上位 200 個の動詞と名詞を用いた（図 7）．

Step1:データベースからイベントツイートを取得し書くツイートを形態素解析にかける．
大通り｜で｜1｜2｜日｜から｜祭り｜が｜開催

Step2:形態素解析の結果から，各ツイートのキーワード（名詞と動詞）の集合リスト $W = \{w_n\}$ を作成する．
(大通り,1,2, 日, 祭り, 開催)

Step3:各文章におけるキーワード w_n の出現回数 w_di をカウントし文書ベクトル $w_d = \{c_d1, c_d2, ..., c_{d\#(W)}\}$ を得る．

図 7: 単語文書行列作成アルゴリズム

決定木による分類結果から,「開催，ライブ，イベント」が分類時の重要項目の上位に位置した．これは，本提案手法で手掛かり後として用いたキーワードと同じ単語であり，分類精度も同等の精度となった．機械学習によるイベントツイート分類

*3 Twitter の運営サーバーに負荷をかけないようにするための利用規制

表 2: 分類結果比較

	提案手法	決定木	ナイーブベイズ	SVM poly kernel
Precision	0.845	0.829	0.81	0.842
Recall	0.854	0.846	0.814	0.855
F-Measure	0.828	0.824	0.812	0.837

の中では，SVM がやや優れており，本提案手法と比べると機械学習の分類の方が F 値が高かった．全体の F 値を比べるとあまり大きな差は見られなかった．本実験の教師データで出現していないキーワードもツイート分類の要素に含まれるため，教師データを増やすことで，より高精度でツイートを分類できると考える．(表 2)

6.2 聴取リスト作成手法

ツイートの分類手法によってイベント聴取時の返答の仕方に大きな変化は見られなかった．

本研究における本来リストに入れなければならないユーザーの発信したツイートとは「イベント情報が含まれるもの」ではなく，「イベント情報をリプライによって提供してくれるであろうツイート」であり，学習データを「イベントツイートデータ」ではなく「イベント情報を返してくれたユーザの特徴データ」に変える必要がある．そのためには，聴取により返答してくれたユーザの日々のツイートの特徴や，イベント予告ツイートの特徴を学習させることで可能になると考えられる．

6.3 ソーシャルセンシングを用いた地域密着イベント情報抽出

表 3 より，10 月後半から 11 月にかけては，ハロウィンなどの秋のイベントが多かったため, イベント聴取へのユーザの反応 *4 が良かったと考えられる．本提案手法を用いて，最初からデータ化されたイベント情報だけではなく，こちらから能動的アクセスを行うことで従来取得できなかったイベント情報を月に平均 10 件程抽出することが出来た．従来手法では，大きな団体や市町村のイベントや，広告費をかけ新聞や Web 上にイベントを告知したイベントのみが多く抽出され，種類も万人受けするものが多かった．本提案手法では，個人主催の講座, ライブ, 地元ショップの催しなどジャンルは様々であり，多種多様なイベント情報が抽出することができた．本手法の実装を機に，イベント情報サイトの存在を知ったイベント主催者から直接イベント情報の投稿も多くみられるようになった（図 8）.

表 3: 本手法を用いた抽出結果 (10/8-1/17)

月	「びもーる」掲載記事	聴取返答	平均インプレッション	平均エンゲージメント
10	2	10	25.49	0.79
11	14	40	31.76	1.59
12	7	45	22.85	1.06
1	5	33	17.02	0.95

*4 「びもーる」へ掲載：取得できたイベント数
リプライへの返事：聴取への返答
インプレッション：ツイートがユーザのタイムラインに表示された回数
エンゲージメント：タイムラインのツイートを見て，起こしたアクションの回数（返信，お気に入りなど）

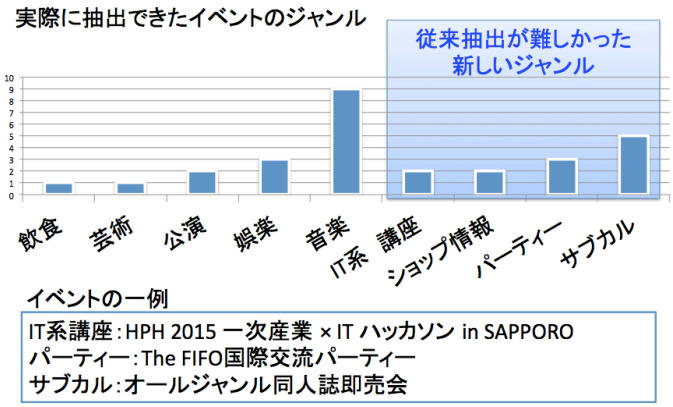

図 8: 本手法を用いて抽出されたイベントのジャンルと一例 (10/8-1/17)

7. おわりに

本研究では，イベント情報サイトを地域住民とイベント主催者などに利活用してもらうことで，多くの人が地域のイベントに参加してもらうために，地域密着イベント情報サイトの現状の課題からソーシャルセンシングを用いたイベント情報抽出を提案した．Twitter のデータから単にイベント情報を抽出するだけではなく，より詳細なイベント情報を抽出するため，ユーザに対して能動的にアクセスしてイベント情報を聞き取る方法を提案した．結果として, 従来手法では取得できなかった地域で開催される個人主催のライブなどを地域住民の情報提供によって多種多様なイベント情報を抽出することが可能となった．

今後の課題として，イベントを聴取する際のツイートの分類を精密に機械学習を用いて行うことと，どういったユーザがイベント情報を返すのかを分析し，機械学習の教師データとすることで，より多くのイベント情報を抽出することがあげられる．

参考文献

[1] Ryota Ono, Kei Hirata, Hidenori Kawamura and Keiji Suzuki ：Scoring Algorithm for Event Notice Recommender System,The 2nd international conference on Serviceology(ICServ2014), Yokohama,2014

[2] Takeshi Sakaki,Makoto Okazaki,Yutaka Matsuo：Earthquake shakes Twitter users: real-time event detection by social sensors,Proceedings of the 19th international conference on World wide web,2010.

[3] 長野伸一, 上野 晃嗣, 長健太：ソーシャルセンサからの鉄道運行情報検出システムの開発, 電子情報通信学会論文誌,2013.

[4] 榊剛史, 松尾豊：ソーシャルメディアの予告型イベント及び参加条件の抽出手法,JSAI'2013.

[5] 川村秀憲：北海道の観光情報における標準タグの普及の取り組みとキュンチャンネルの開発, 情報処理学会デジタルプラクティス, 2012.