

ソーシャルメディアの分析によるイベント開催支援

藤原 泰士[†] 難波 英嗣[†] 竹澤 寿幸[†]

[†] 広島市立大学大学院情報科学研究科 〒731-3194 広島県広島市安佐南区大塚東 3-4-1

E-mail: † {fujiwara, nanba, takezawa}@ls.info.hiroshima-cu.ac.jp

あらまし 近年、民間事業者や公共団体などがイベントを開催する際、イベント参加者の声を把握するために、ブログなどのソーシャルメディアの情報を、テキストマイニング技術を用いて分析し、イベントの運営改善に活用する事例が増加している。しかし、ソーシャルメディア上の情報は膨大であるため、例えば、イベントに対する改善案や要望など、イベントの改善に役立つ情報を人手で収集するのは容易ではない。そこで、本研究では、こうした情報を、Twitter, YouTube のコメント、旅行ブログから効率的に抽出する手法を提案する。

キーワード Twitter, YouTube, ブログ, テキストマイニング, イベント開催支援

1. はじめに

近年、イベントを開催した際のイベント参加者の声を把握する手段として、Twitter やブログなどのソーシャルメディアを用いる事例が増加している。しかし、それらのソーシャルメディアの情報は膨大に存在しており、全てを人手で判定し収集を行うには多大なコストを要する。

例えば、毎年、広島で開催される宮島水中花火大会について、ソーシャルメディア上では、以下のようなコメントが投稿されている。

- (1) 人が多くて花火が見えない
- (2) 鳥居と被っていて花火がよく見えなかった

(1)(2)は 2 つともイベント参加者の否定的な声の例である。(1)のようなイベント参加者の反応はイベント主催者側とすれば容易に想定できる内容だと考えられる。しかし、(2)は、そのイベントの象徴といっても良い宮島の鳥居が逆に不満に繋がることを示しており、このような情報はイベント主催者側にとっては想定外の内容である。この場合、鳥居と花火が重ならない場所を観覧場所としてイベント参加者に提示するようになるといったイベントの開催支援に利用できると思われる。

そこで本研究では、Twitter, YouTube, 旅行ブログから、上記のようなイベント参加者の声を抽出し、イベント主催者側に提示するシステムの構築を試みる。Twitter からは、リアルタイムで情報を抽出することが可能である。YouTube からは、イベント参加者が投稿した動画に対する大量のコメントから、そのイベントに関する反応を得ることができる。また、旅行ブログは、Twitter と比べるとイベントの開催日とブログが投稿される日にタイムラグが存在するものの、Twitter や YouTube と比べると、ある程度まとまった意見が得ら

れる可能性がある。以上述べた 3 つのソーシャルメディアから得られるデータを対象に、以下に述べる二つの手法を用いて意見分析を行う。一つ目は、Kobayashi ら[1]ブログを対象とした意見文か否かを判定し、意見文と判定された情報を本研究ではイベント開催支援に有益な情報として扱い分析を行う。二つ目は、日本アイ・ビー・エム株式会社で開発されたテキストマイニングツール TAKMI[2]を用いて文の構成要素の相関関係を分析することで、イベント開催支援に有益な情報を抽出し分析を行う。

本論文の構成は以下の通りである。2 節では関連研究、3 節では本研究で提案する有益な情報の抽出手法、4 節では抽出した情報の分析とその考察について述べ、5 節で本稿をまとめる。

2. 関連研究

本節では、本研究の関連研究として、Web から意見や要望を抽出する研究、イベントに関連する情報を抽出する研究について紹介する。

2.1 Web からの意見文の抽出

膨大な情報の中から特定の情報を抽出するという研究は多数行われている。そのひとつに、ブログや新聞記事から著者の意見文を自動的に抽出するという Kobayashi らの研究がある。ソーシャルメディア上には、本研究で必要となる有益な情報とイベント参加者のコメントでも要望とは関係のない情報が存在すると考えられる。そこで本研究では、有益な情報であるか否かという判定を行うために、その文章が意見文か否かを判断する。実験結果として、ブログからの意見文抽出は精度 0.771、再現率 0.413、新聞記事からの意見文抽出では、精度 0.619、再現率 0.513 という結果が得

られた。本研究では、Kobayashi らの意見文抽出器をソーシャルメディアの分析に用いる。

2.2 イベント情報の抽出

本研究と同じくイベントに関連する研究として、Nanba らの研究がある[3]。Nanba らはイベント情報の情報源として新聞記事と Web からのイベント情報の抽出を行っている。また、Web ページからイベント名や日時、開催場所などの情報を抽出することで、これから開催されるイベントを収集している。イベント情報を抽出する手法として、人手で収集した手掛かり語を素性を用いて、機械学習によるイベント記事の検出を提案している。吉田[4]らは、必要性の高いイベントの情報リストを作成し、イベント名の前後関係を用いてブログ記事と Web ページからイベント情報抽出の手法を提案している。岡本ら[5]は、常に新しいイベント情報を追加し、有益なイベント情報を提供するために、地域に着目したイベント情報の抽出を行っている。上記の研究は、イベント名や開催日時に焦点を当てているが、本研究ではそのイベント情報からイベント開催支援に役立つ有益な情報を抽出するといった点で異なる。

2.3 観光情報の自動抽出

本研究に関連する研究として、石野らの研究がある[6]。石野らは、言語解析技術を用いて、観光ガイドブックのページ分類を行っている。石野らは、この手法を組み込む際、各ブログエントリを日付け順に並べ、時系列データとして、系列ラベルリング問題として解き、機械学習に CRF を用いる手法を提案している。これにより、精度 0.867、再現率 0.381 を得ている。

また、抽出された旅行ブログから観光情報として地域名と土産物の対を効率的に抽出する。提案手法の有効性を示すため、石野らは地名と土産物の対を 482 対抽出し、旅行ブログ、一般ブログ、Web から抽出された地域名と土産物の中で新しい地域名と土産物を比較する実験を行った。旅行ブログからの観光情報の抽出に関しては、抽出された上位 100 種類の土産物において精度 0.740 を得ることができており、旅行ブログは観光情報の情報源として有効であると述べている。

3. イベント参加者から意見の収集と分析

3.1 システムの概要

本節では、本研究で提案するテキスト分析手法について説明する。本節の構成は以下の通りである。3.2 節で意見文抽出器について説明し、3.3 節でテキストマイニングツール(TAKMI)について説明する。異なる分析ツールを利用することで、より多用な観点からの

意見の分析を目指す。入力として、YouTube のコメント、Twitter、ブログの 3 種類を利用する。3 つの異なる利点を持ったソーシャルメディアを用いることで幅広い意見の収集を目指す。以上 2 つの手法を用いて類似したイベントの比較を行い、各イベント固有の問題を洗い出すことで特定のイベントの知見が得られると思われる。図 1 はシステムの概要を示しており、3.2 節は左側、3.3 節は右側の処理を示している。

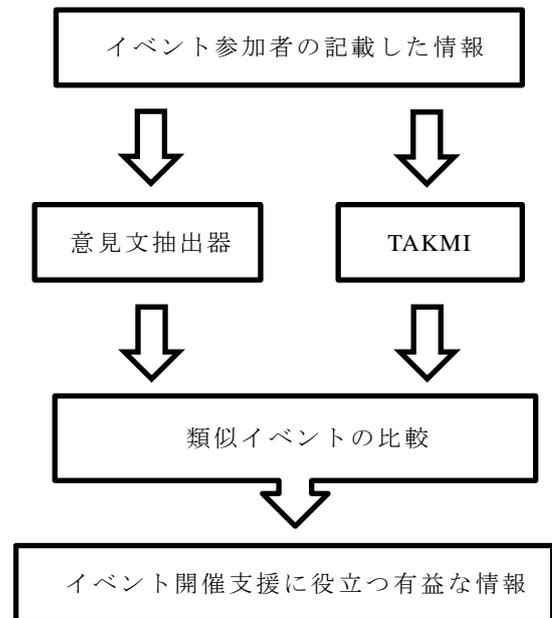


図 1: システム概要

3.2 意見文抽出器

本研究では、Kobayashi らの手法を用いることで、膨大なテキスト情報から有益となる意見文の抽出を行う。Kobayashi らは、ブログと新聞記事をそれぞれ対象にした意見文抽出器を構築しているが、本研究では、ブログ用の意見文抽出器を用いる。ブログには、新聞記事では記述されない特有な表現がある。例えば、新聞記事において、「宮島の紅葉がとても綺麗です。」と書かれた文がある場合、ブログでは「宮島の紅葉めっちゃキレイ☆彡(≧▽≦)o」というような、くだけた表現で記述されることが多い。特に、「☆彡(≧▽≦)o」といった顔文字表現は、ブログ固有のものである。これに対処するために、Kobayashi らは、Wikipedia から収集した観光名所の名前を用いて評価対象リストを作成している。文章の感情を表現するような顔文字や「!!!」や「……………」といった記号を連続して用いる強調表現、ブログ固有の評価対象リストなどを素性に用いて、機械学習を行うことで意見文の抽出を行う。抽出された意見文を本研究では、イベント開催に役立つ有益な情報としてイベント主催者に提示する。

表 1: 各ソーシャルメディアのイベント名と件数

イベント名	Twitter(件)	YouTube(件)	ブログ(件)
なばなの里 ウィンターイルミネーション	50	500	100
神戸ルミナリエ	713	500	102
時之栖イルミネーション	188	500	98
東京ミッドタウン イルミネーション	685	500	100
SENDAI 光のページェント	354	500	96
光のルネサンス	910	500	97
丸の内イルミネーション	141	500	97
ひろしまドリミネーション	58	500	95
キャナルシティ博多 イルミネーション	19	500	95
さがみ湖イルミリオン	225	500	94
東京ミチテラス	1,048	500	93
榛名湖イルミネーションフェスタ	57	500	99
いとまんピースフルイルミネーション	48	263	94
おおたイルミネーション	85	500	96
イルミネーションロードくれ	7	172	93
合計	4,588	6,935	1,449

3.3 テキストマイニングツール(TAKMI)

本研究では、日本アイ・ビー・エムが開発した TAKMI を用いてテキストの解析を行う。TAKMI は、品詞の関連性や構成要素の相関関係などを解析し、可視化するツールである。例として、「詳しい地図が欲しい」という文に対して「欲しい」という動詞と「詳しい...地図」という対の頻度と相関が出力される。本研究では、文の構成要素の相関関係から情報を分析し有益な情報の抽出を試みる。

4. 実験

本節では、3 節で説明した手法を用いて、分析を行う。4.1 節で実験データについて説明し、4.2 節で分析により得られた知見について説明する。

4.1 実験データ

本節では、分析に使用したデータの詳細を記述する。本研究では、様々なイベントの中から、全国各地の「イルミネーション」に関する情報を収集した。「なばなの里 ウィンターイルミネーション」や「神戸ルミナリエ」など、15 種類のイベントを対象に、Twitter から 4,588 件、YouTube から 6,935 件、ブログから 1,446 件の情報を人手で収集した。詳細を表 1 に示す。

4.2 分析結果

4.2.1 意見文抽出器を用いた分析

表 1 で記載した各ソーシャルメディアの情報を、3.1 節で説明した意見文抽出器を用いて分析を行う。

• Twitter

意見文抽出器を用いて、4,588 件の Twitter データから抽出した結果、137 件の意見文が検出された。以下に、抽出した意見文の例を示す。

- (3) そんなに人がいなくてよかったけど道路がカチカチに凍ってて怖がって奥には行けずでした。
- (4) 神戸ルミナリエは来年の 1 月くらいまで、やって欲しかったですね。
- (5) 去年の綺麗なトンネルなくなっちゃってたね…

(3)は、イベント会場の不備を表す意見文から、危険な場所の整備・保守を行うことが可能であり、イベント時の事故を予防できると考えられる。また、(4)の意見文から、イベント開催期間が短いという指摘を推測することができ、(5)の意見文から、次のイベント開催の際の改善点になると考えられる。

• YouTube

意見文抽出器を用いて、6,935 件の Twitter データから意見文を抽出した結果、66 件の意見文が検出された。以下に、抽出した意見文の例を示す。

(6) 昨夜のふれびゅーとは大違いだけど以前程の美しさを感じなかった。特に建物前の昔はソリとかあったところが酷い。

この意見文から、以前のイベントと比較して改善点を抽出することが可能であると考えられる。また、抽出された意見文の数が少ないということもあるが、全体的にイベントに関する情報よりも、YouTube に投稿された動画そのものに対する意見(撮影機材や撮影環境など)が多い傾向にあった。

・ブログ

意見文抽出器を用いて、1,449 件のブログデータから意見文を抽出した結果、210 件の意見文が抽出された。以下に、抽出した意見文の例を示す。

(7) なばなの里のウインター・イルミネーションでは三脚の使用が禁止されています。そうして10分間くらい違う場所で撮った後、ふと見たら、その4人組が正々堂々と三脚を使って撮影しているではありませんか！こいつら、忠告されているのを知りながら、改めようとしません。

(7)は、イベント会場における迷惑行為に関する意見文である。(7)のような情報は、イベント参加者にしか感じられない事象が多いと考えられる。このような情報をイベント主催者に提示することにより、特定の迷惑行為に対する対策を促せると考えられる。

4.2.2 TAKMI を用いた分析

本節では、TAKMI を用いて、表 1 に記載した各ソーシャルメディアの情報の分析を行い、その結果に対する考察について述べる。解析には、相関関係とその情報の頻度の数値を用いる。図 2 に TAKMI の動作例を示す。行と列には解析する項目を選択することができる。例では、行に形容詞、列に句の構成要素を選択しており、その二つの対の頻度と相関関係の数値を表示している。

行:形容詞	列:句の構成要素	頻度	相関
美しい	美しい...光	15	15.1
詳しい	詳しい...地図	22	14.6
少ない	人...少ない	18	13.3
明るい	まだ...明るい	12	10.4
すごい	すごい...人	16	7.8
大きな	大きな...タワー	15	7.4
幻想的に	幻想的に...光	18	6.4
多い	人...多い	53	6.3

図 1 : TAKMI の動作例

・Twitter

表 2 に、Twitter データの相関関係の高い上位 5 件を示す。最も相関関係が高い「人...多い」という情報は、イベントの状況を把握するうえで必要である。しかし、比較的どのようなイベントでも想定できる内容であり、既出の情報であると言える。

相関関係は上位と比べると低くなるが、「マッピング...すごい」といった対を抽出している。このような対が記述されているツイートを調べたところ、大阪城 3D マッピングのイルミネーションに関して、「光のルネサンス」という特定のイベントについての良さについて記述されている。このような情報は各イルミネーションごとに抽出された差分を見ることで抽出が可能となっている。このように類似イベントを比較することでそのイベントにしか得られない有益な情報をイベント主催者に提示することにより、他のイベントには無い長所を更に生かすイベント作りが出来ると考えられる。

表 2 : Twitter データにおける相関関係

品詞(全品詞)	句の構成要素	頻度	相関
多い	人...多い	27	41.8
照らす	聖夜...照らす	39	40.5
ブログ	ブログ...更新する	37	38.4
多い	人-が...多い	16	35.5
ござる	イベント-が...開催する	15	33.2
～中略～			
3D	マッピング...すごい	5	13.4

・YouTube

表 3 に、YouTube のコメントにおける相関関係を示す。全体的に、「動画」や「YouTube」といった情報が多く見られた。内容としては、イベントに関する情報よりも動画を撮影したカメラやツールについての記載が多く、イベント開催支援に有用な情報を見つけ出すことが困難となっている。

表 3 : YouTube データにおける上位 5 件の相関関係

品詞(形容詞)	句の構成要素	頻度	相関
最大級だ	関東...最大級だ	9	327.7
新	大学...学	8	274.7
苦手だ	ブログ...苦手だ	7	198.1
上品だ	大手町...結ぶ	7	186.8
雑だ	里...デートだ	19	179.1

・ブログ

表 4 に、ブログデータにおける相関関係を示す。表 4 において、まず、「詳しい...地図」について見ていく。内容としては、東京ミッドタウンのイルミネーションに関して詳しい地図を見ながらイルミネーションを見て回れたといった実施したサービスの評価について知ることができるというものであった。次に「人...少ない」を見ていくと、人が少なくなる状況や時間、環境をイベントごとに知ることができ、その状況や環境に合わせたそのイベント独自のサービスを行うことでイベントを活発化させることが可能である。また混雑を避けたい人にその情報を提供することでイベントに対する評価が向上するのではないかと考えられる。

表 4：ブログデータにおける上位 5 件の相関関係

品詞(形容詞)	句の構成要素	頻度	相関
青い	青い...光	15	15.1
詳しい	詳しい...地図	22	14.6
少ない	人...少ない	18	13.3
明るい	まだ...明るい	12	10.4
すごい	すごい...人	16	7.8

4.2.3 意見文抽出器と TAKMI で抽出された情報の比較

本節では、4.2.1 項と 4.2.2 項で紹介した有益な情報と 3 つのソーシャルメディアを用いたことで得られた有益な情報の比較を行う。まず、4.2.1 項では意見文抽出器を用いて抽出された意見文は 3 つのソーシャルメディアともイベントの開催支援に役立つものが抽出できた。また、4.2.2 項の TAKMI を用いて抽出された情報は 4.2.1 項で抽出した情報に比べて有益な情報と判断されるものの割合が少ない結果となった。しかし、意見文抽出器では抽出されなかった「大阪城の 3D マッピングが良かった」といった情報を、TAKMI を用いたことにより、有益な情報として得られた。

次に 3 つのソーシャルメディアから得られた情報の比較を行う。3 つのソーシャルメディア共通で得られた内容としては「人が多い」といった情報が多かったが、それぞれの利点が繁栄された書き方が多々見られた。Twitter では人が多い時間と場所、YouTube ではその様子が動画で見ることができ、ブログではどの時間に何故多いのかといった内容が記載されている場合があった。

次に個別のメディアで得られた有益な情報としては、Twitter から得られた情報は路面の状況といったリアルタイムで知れると有益な情報を抽出出来た。YouTube では、(6)のような以前のものととの比較に関し

て記載されたコメントを抽出しており、実際に動画で視覚的に確かめられるといったメリットを生かせる情報を抽出している。ブログでは、Twitter では文字数の制限から詳細に説明できないような会場の迷惑行為といった情報を抽出出来ている。

5. おわりに

本研究では、様々なソーシャルメディアの情報から、イベント開催時に役立つ有益な情報をイベント主催者に提示する手法を提案した。本研究の手法では、テキスト情報から意見文抽出器を用いて意見文を抽出する手法と TAKMI を用いたテキストの統計解析の 2 つの手法から分析を試みた。前者の手法では、実際にイベント主催者に提示すればイベント開催支援に役立つと考えられる情報が抽出できた。後者の手法では、有益な情報も含まれているが相関関係の上位を見ただけでは判断しづらい結果となった。

今後の課題として、意見文抽出器にかけた出力結果を TAKMI により解析を行うことで、よりイベント主催者が有益な情報だと判断しやすくなるのではないかと考えられる。現状では、抽出結果が 100 件前後と非常に件数が少ないので、データ数を増加させることによって抽出結果を増やす必要があると考えられる。

次に、TAKMI で解析を行った結果から品詞に有益な情報との関連性があるのではないかと考えられた。例えば、4.2.2 節のブログのデータ解析の際に説明した「人...少ない」の項目の品詞を見ると「少ない」となっている。「少ない」という単語は否定的な単語であると考えられ、このように否定的または肯定的な単語からその項目の重要性を計ることが出来れば提示手法として有効なのではないかと思われる。鍛冶ら[7]が構築した評価表現辞書を用いることで肯定的か否定的かの判定が可能であり、品詞に着目して情報を制限したいと考えている。

最後に、特定のイベントに特化した意見文を抽出することによって、そのイベント特有の情報が得られると考えられる。例えば、今回は 15 種類のイルミネーションのデータをもとに実験を行っている。神戸ルミナリエについて抽出された情報とその他 14 種類のイベントから抽出された情報を比べて、類似するものを除去すれば、神戸ルミナリエ特有の情報が得られイベント開催支援に役立つのではないかと考えられる。

参考文献

- [1] Kobayashi, D., Nanba, H., and Takezawa, T. "Extraction of Opinion Sentences using Machine Learning: Hiroshima City University at NTCIR-7 MOAT". In Proceedings of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question

Answering and Cross-lingual Information Access, pp.255-259, 2008.

- [2] Zhu W.-D., Iwai A., Leyba T., Magdalen J., MacNeil K., Nasukawa T., (Nita) P. N., and Sugano K.. IBM Content Analytics Version 2.2: “Discovering Actionable Insight from Your Content.” IBM Redbooks, 2011.
- [3] Nanba, H., Saito, R., Ishino, A., and Takezawa, T. “Automatic Extraction of Event Information from Newspaper Articles and Web Pages”. ICADL 2013, LNCS 8279, pp.171-175, 2013.
- [4] 吉田将人, 福原知宏, 増田英考, “ブログ記事とWeb ページを用いたイベント情報抽出手法の提案”, 情報処理学会研究報告, デジタルドキュメント, Vol.2009, No.35, pp.37-44, 2009.
- [5] 岡本昌之, 菊池匡晃, “ブログからの地域イベント情報抽出”, 情報処理, Vol.51, No.1, pp.14-17, 2010.
- [6] 石野亜耶, 難波英嗣, 竹澤寿幸, “旅行ブログエントリーからの観光情報の自動抽出”, 日本知能情報ファジィ学会誌, Vol.22, No.6, pp.667-679, 2010.
- [7] 鍛冶伸裕, 喜連川優, “自動構築した評価文コーパスからの評価表現辞書の構築”, 日本データベース学会 Letters, Vol.6, No.1, pp.41-44, 2007.