

Twitter を用いたリアルタイム情報収集による観光地情報推薦システム

免田 哲矢† Kryssanov V.V. ‡ 林 勇吾‡ 小川 均‡

立命館大学大学院 理工学研究科 情報理工学専攻†

立命館大学 情報理工学部‡

1 はじめに

近年、Twitter^[4]に代表されるマイクロブログサービスが web 上で世界中の人々から注目を集めている。Twitter の特徴は、「ツイート」と呼ばれる 140 文字以内に制限された文章を、メッセージとしてインターネット上に発信することができる点である。更新が容易で短い文章を発信するため、blog や SNS (Social Network Service) よりも手軽に投稿できる^[1]ことから、ユーザは身の回りで起こった出来事やユーザ自身の感情などを頻繁に投稿する傾向にある。

本研究では、Twitter に投稿された観光地に関するツイートを利用し観光地の情報の推薦を行うシステムの開発を行うことを目的とする。

2 システム構成

システムの構成を図 1 に示す。本システムでは、Twitter 上のデータを用いて、他のユーザの観光地情報をリアルタイムで収集し、他のユーザから発信された観光地情報に関連するデータを形態素解析により抽出する。そして、抽出されたデータに対して観光地に対して好印象と判断された点、悪印象と判断された点を判断し、

他のユーザから発信された観光地情報を地図上に表示するシステムの提案を行う。観光地に対しての好印象と判断された点、悪印象と判断された点を明確にするために、現在は評価極性を使用することを検討している。評価極性とは、抽出された名詞に対しての評価を数値化したものである。また、投稿時間をフィルタリングすることで出現頻度や共起頻度を計算し旬な話題や関連する話題などを発見することができる。最終的には、ユーザに観光地を推薦するだけでなく、ユーザが選択した観光地の周辺情報も Twitter から抽出し推薦できるようにする。そして、選択された観光地に対して最適なルートを推薦するシステムの構築を行う。このシステムを用いることで従来のブログや SNS とは異なり、より即時的な意見の収集を行えることが見込める。

3 マイクロブログからの情報取得

Twitter から取得する観光地の対象を京都府に存在する地名や建造物名に限定して開発を行う。

3.1 処理の流れ

ここでは京都に実在する“金閣寺”という観光地についてのツイートを取得するものとする^[2,3]。次に Twitter4j^[5]を使用し Twitter 上から文章データを取得する。取得した文書データに対して形態素解析を実施しデータを品詞ごとに分解する。ここでの形態素解析には Sen という形態素解析機を使用した。その中であらかじめ京都の地名・建造物名を入れた固有名詞辞書とマッチングを行いキーワードである “金閣寺”という固有名詞を含むつぶやきを抽出しデータベースに格納する。以上の手法で対象となるツイートを取得することが可能である。

3.2 つぶやき取得結果

3.1 のような処理を行うことで表 1 のような結果を得ることができた。ツイートに“金閣寺”という名詞が入っているツイートを抽出しそれぞれの投稿時間、ユーザ名、ツイートをデータベースに格納した。このように京都の地名や建造物などの観光地情報を Twitter 上から抽出することができた。

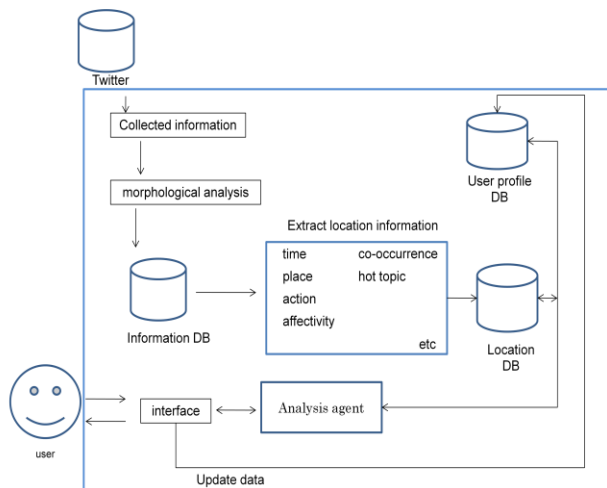


図 1：システム構成図

A recommender system for sightseeing using real time information extracted from Twitter

†Graduate School of Science and Engineering, Ritsumeikan
‡College of Information Science and Engineering,

Ritsumeikan University

Tetsuya Menda

Kryssanov V.V.

Yugo Hayashi

Hitoshi Ogawa

表 1. ツイート取得結果

投稿時間	ユーザ名	ツイート
Wed Dec 08 16:22:27 JST 2010	kyauco	金閣寺なう。楽しかった!
Wed Dec 08 16:03:15 JST 2010	CPA_N_Okada	さあ金閣寺はどのくらい金なのか...ワクワ
Wed Dec 08 15:32:33 JST 2010	pugitan	今から金閣寺行くよー!
Wed Dec 08 15:30:10 JST 2010	Xx261mip	金閣寺に来た〜♪とても美しい景色で感動なう(´▽`)

3.3 調査

Twitter から取得したデータがどのくらいの精度で評価を行えるのかをアンケート形式で実験した。アンケートでは実際に”金閣寺”という固有名詞が含まれているツイートを 400 件収集しその中から無作為に 40 件のツイートを選出した。次にそのツイートの”金閣寺”という言葉に対して「どれくらいポジティブか、ネガティブか」を評定してもらった。なお、ここでの「ポジティブ・ネガティブ」は”金閣寺”に対していい印象を与えるか、または悪い印象を与えるかと定義した。質問は、「非常にポジティブ」「ポジティブ」「どちらともいえない」「ネガティブ」「非常にネガティブ」の 5 件法による回答で 20 名の調査協力者（以後、評価者）に評定。をしてもらった。

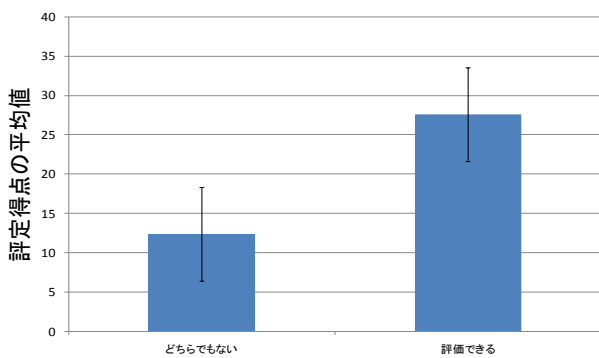


図 2: 評価者の平均

分析では、まず評価者が評定した 40 の項目を「どちらでもない」と「(非常に)ネガティブ」、「どちらともいえない」、「(非常に)ポジティブ」を合計した「評価できる」に分類し、それぞれの項目数を得点化した。図 2 は、評定者の評定得点の平均値（縦軸）を、各カテゴリー（横軸）に示したものである。次に、1 要因の分散分析を実施したところ、有意な差が認められた ($F(1, 39) = 64.479, p < .01$)。これより、評定者は、ツイートの中から、「評定できる」と判断した割合が有意に多いことを示している。次に「評価できる」と評定したツイートの内「非常にポジティブ」、「ポジティブ」、「非

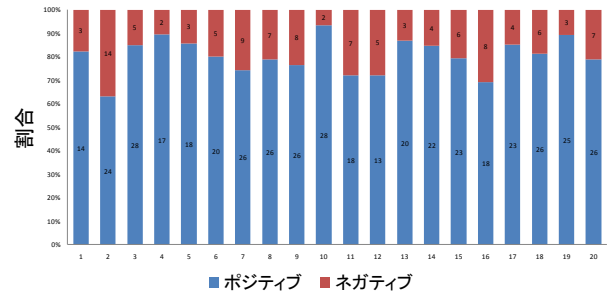


図 3: 評価できるツイートの割合

常にネガティブ」、「ネガティブ」がどれぐらいの頻度で出現しているのかを分析した。図 3 は、評定者の評定得点の平均値（縦軸）を、各カテゴリー（横軸）に示したものである。その結果、「非常にポジティブ」、「ポジティブ」な評価のほうが「非常にネガティブ」、「ネガティブ」な評価よりも多く「推薦可能なコンテンツ」が豊富に存在することを意味する。また被験者全員に対してほぼ同様の結果を得ることができた。そのため Twitter から取得したデータに対して評価手法を決定することは可能であるといえる。今後は”金閣寺”以外のキーワードでも分析を行い、評価手法を決定し、コンピュータで評価を行うと、どの程度の精度でキーワードに対しての評価を行うことができるかを考察していきたい。

4 今後の展開

現在は、Twitter4j と形態素解析機の Sen を使用し、Twitter 上から京都の地名や建造物名などが含まれたツイートを取得するシステムを開発した。今後は、検索したいキーワードが含まれるツイートだけでなく、ツイートと一緒に投稿された GPS 情報や画像情報も抽出できるように改良する。そして、抽出したツイートに対して評価表現辞書を使用し、評価極性を算出することで対象の観光地に対して評価を行っていく。

参考文献

- [1] Martin Ebner, Hermann Maurer: Can Microblogs and Weblogs change traditional scientific writing? in: Proceedings of E-Learning 2008, Las Vegas, p.768-776, 2008
- [2] 岩木 祐輔, アダム ストヤノフ, 田中 克己: マイクロブログにおける有用な記事の発見支援, DEIM Forum 2009 A6-6,(2009)
- [3] 田中 淳史, 田島 敬史: twitter のツイートに関する分類手法の提案, DEIM Forum 2010 A5-4,(2010)

参考 URL

- [4] Twitter: <http://twitter.com/>
- [5] Twitter4jAPI: <http://twitter4j.org/ja/index.html>