

時空間情報を含む「つぶやき」からの推測による お出かけスポット情報の生成

深谷昭宏[†] 浦正広[‡] 遠藤守[†] 山田雅之[†] 宮崎慎也[†] 安田孝美[‡]

中京大学情報理工学部[†] 名古屋大学大学院情報科学研究科[‡]

1. はじめに

出かけ先を選定する際、公式サイトからは発信者の伝えたいポイントに関する情報を収集できる。しかし、利用者からの評価やコメントなどがなく、判断材料として十分ではない。一方、口コミサイトではそのような情報が集まる場となっているが、利用者の自発的な情報提供により成立するため、内容に偏りが生じたり、情報収集自体が円滑に行えない点が問題として挙げられる。これを解決する手段として、ブログ記事から自動的に広く情報を収集し、旅行記のみを抽出する手法が提案されている[1]。しかしながら、あらかじめデータベースに登録された観光スポットのみを対象とするため、抽出は特定のスポットの旅行記に限定される。また、ブログ記事は長文になりやすいため、抽出情報の体系化も容易ではない。

そこで本研究では、“つぶやきメディア”を対象に、位置情報を含んだ‘つぶやき’からその対象となるスポットを特徴づける(i)名詞を抽出するとともに、(ii)場所や(iii)営業時間を推測し、それらをまとめた‘お出かけスポット情報’を自動生成するシステムを提案する。“つぶやきメディア”では、その時に思ったことを気軽に発言でき、人々の日常生活の様々な内容が Web に集約されているため、幅広い情報の収集が期待できる。また、“つぶやき”は一般に文字数が短く、位置情報がメタデータとして付加されたものもあることから、抽出情報の体系化に適している。

2. 提案システム

情報フィルタリングにより“つぶやき”から“お出かけスポット”に関係のある情報を抽出し、位置情報と時間情報を用いて“お出かけスポット情報”を生成する。

2.1. 情報フィルタリング

システムに人間の感性を取り入れることが重要であるとされており[2]、テキストマイニングにおいても重要な要素である。本システムでは、すべての“つぶやき”から、お出かけに関する名詞を抽出するため、フィルタリングの条件としてキーワード及び品詞の組み合わせの 2 つの条件を用いたフィルタリング処理を行う。キーワードはつぎのように定める。位置情報付加機能を有する Timelog, Himatch, コエポッド, locaty, staaaaar の 5 つのサービス上の“つぶやき”から、お出かけに関する“つぶやき”を人力で集め、それらに形態素解析を施し[3]、頻出した上位 10 件の形容詞と動詞をキーワードとする(表 1)。また、お出かけ情報に多く含まれる品詞の組み合わせを利用する(表 2)。上記 2 つのフィルタの組み合わせにより、“つぶやき”から“お出かけスポット情報”となり得る可能性の高い名詞の抽出を行う。フィルタリングの手順は、まず、“つぶやき”に形容詞または動詞のキーワードが含まれているか判断し、含まれている場合、品詞の組み合わせを参照する。2 つの条件にマッチした場合に名詞を抽出し、キーワードをその名詞に対する評価とする。

表 1 頻出単語の例

形容詞				動詞			
単語	数	単語	数	単語	数	単語	数
うまい	53	安い	16	し	128	飲み	30
いい	44	高い	15	ある	57	行っ	29
美味しい	23	旨かつ	15	食べ	48	来	28
旨い	22	すごい	13	い	46	なっ	28
良い	18	楽しかつ	13	てる	39	やっ	23

表 2 品詞の組み合わせと例

品詞の組み合わせ	例
(副詞)+形容詞+名詞	(とても)+美味しい…
名詞+係助詞+(副詞)+形容詞	…が+(とても)+うまい
名詞+係助詞+(形容詞)+動詞	…に+(はやく)+来た

A Generation Method of Sightseeing Spots Information from Short Messages including Geolocation

[†]Akihiro Fukaya, Mamoru Endo, Masashi Yamada, Shinya Miyazaki : Chukyo Univ.

[‡]Masahiro Ura, Takami Yasuda : Graduate School of Nagoya Univ.

2.2 情報生成

本システムにおける“お出かけスポット情報”の生成手順を以下に示す。まず，“つぶやきメディア”から情報フィルタリングにより，お出かけに関する名詞を含んだ“つぶやき”を抽出し，その集合を T とする。つぎに，T の各“つぶやき”について，一定距離内にある“つぶやき”を T から取り出し，それらの集合を U とする (図 1)。U の各“つぶやき”についてお出かけに関する名詞を抽出し，“お出かけスポット情報”L に追加する。このとき，“行く”や“来る”といった移動が推測されるキーワードの場合には，抽出される情報は“お出かけスポット”の名称と推測する。また，U の各“つぶやき”の位置情報により“お出かけスポット”の場所を推測する。投稿時間について一定時間内に連続した U の各“つぶやき”を取り出し，それらの集合により営業時間を推測し，リスト L に追加する。また U の各“つぶやき”が連続していない場合，投稿時間の周辺は営業時間であると推測する (図 2)。

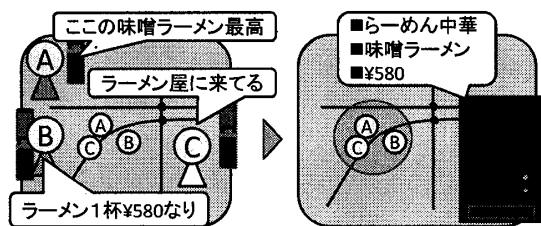


図 1 抽出された名詞のリスト化

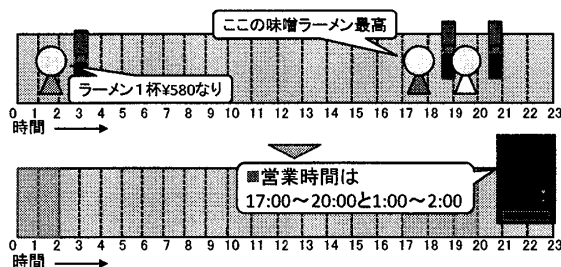


図 2 営業時間の推測

3. お出かけスポット情報の生成例

実際の“つぶやき”を例に“お出かけスポット情報”の生成までの処理の過程を示す (表 3)。

表 3 つぶやきの例

時間情報	つぶやき
2009.12.28(月)22:53	行きつけの南陽に来ています。 ここの台湾ラーメンは美味しい
2010.01.06(水)21:32	にんにくチャーハンが超旨い

“おすすめスポット情報”は，1つの“つぶやき”から一定範囲内の“つぶやき”を集約し，情報の抽出を行うことで生成される。上段の“つぶやき”を評価対象とした場合，まず，句点もしくは改行により文単位に分割し，それぞれについてフィルタリングを行う。結果，第 1 文はキーワード「来て」を含み「名詞+係助詞+(形容詞)+動詞」に該当したため，“南陽”が抽出され，第 2 文はキーワード「美味しい」を含み「名詞+係助詞+(副詞)+形容詞」に該当したため，“台湾ラーメン”が抽出された。つぎに，上段の“つぶやき”の位置を基準に周辺の“つぶやき”を絞り込んだところ，下段の“つぶやき”が該当し，この“つぶやき”からは「にんにくチャーハン」が抽出された。これら 2 つの抽出情報の組み合わせにより“おすすめスポット情報”を生成することができた (表 4)。

表 4 生成された“おすすめスポット情報”

名称	南陽
所在地	豊田市山之手 6-66
特徴	台湾ラーメン:美味しい にんにくチャーハン:超旨い
日時	(月)(水), 21:32~22:53

4. まとめ

本研究では，時空間情報を付加した“つぶやき”から目的となる“おすすめスポット情報”を生成するシステムを提案した。提案したシステムにより“つぶやき”から情報を抽出し“おすすめスポット情報”生成を行い，その有効性を確認した。今後の課題としてフィルタリングの精度向上や生成情報の向上のため，“つぶやき”の収集を行う。

謝辞

本研究の一部は財団法人栢森情報科学振興財団研究助成，財団法人日比科学技術振興財団研究助成，財団法人 JKA 補助事業による。

参考文献

- [1] 竹葉大輔: 観光評判情報の抽出に関する研究, 高知工科大学 2008 年度学位論文, 2009.02.
- [2] 遠藤守ほか: 人工知能技術に関する調査等研究委員会報告書財団法人, 人工知能研究振興財団, 2009.
- [3] Taku Kudo, et al.: Applying Conditional Random Fields to Japanese Morphological Analysis, Proc. of EMNLP-2004, pp.230-237, 2004.