

複合施設における店舗およびフロアに関する Web ページへのツイートマッピング手法の検討

藤原優^{†1}安井豪基^{†1}王元元^{†2}河合由起子^{†1}秋山豊和^{†1}^{†1} 京都産業大学^{†2} 名古屋大学

1 はじめに

実世界において場所に関連した内容の情報をリアルタイムで発信する SNS ユーザは多数おり、それらの発言は、リアルタイムでその場所の現状を知ることができ、即時性の高い有益な情報といえる。Twitter に投稿されたツイートを分析する研究は活発に行われており、イベントの検出や位置情報の取得を試みた研究も数多くある。Arakawa ら [1] の位置情報ツイートから位置依存性の高い文字列を抽出する手法、Yamamoto ら [2] の Twitter に投稿された実生活情報から有用性の高いものを抽出し局面に応じた記事をユーザに提示するシステム、Yamaguchi ら [3] の位置情報が既知であるユーザのツイートを用いてローカルイベントの検出を行い、位置情報が未知であるユーザの位置情報を推定する手法などが提案されている。

我々はこれまで、発信場所に関するツイート発見手法の提案と、それら場所に関する Web ページへのマッピングを提案し、さらにツイートユーザと、発信された場所に関連する Web ページを閲覧している Web 閲覧ユーザに対して、SNS と Web の異種メディア横断型コミュニケーションシステムを構築してきた [4]。本研究では、さらに複合施設における各フロアに関連するツイートの分類を行い、Foursquare 分析に基づく Venue 辞書を構築し、Venue 辞書を用いて関連する施設内の店舗やフロアに関する Web ページを検出し、そのページ上に該当するツイートの集約情報を提示することで、システムの拡張を目指す。

2 システム概要

図 1 にシステムの概要を示す。ツイートを発信すると、ツイートの内容と発生場所の位置情報に基づき関連する Web ページを検出し、検出したページ上にそのツイートがリアルタイムに提示される。また、取得した位置情報付きツイートに対してフロアならびに店舗名を判定することで各フロアや店舗に対応する Web ページへのマッピングが可能になる。図 1 では、東京スカイツリー付近にいるユーザが展望台でツイートを発信した場合に、そのツイートが東京スカイツリーの展望台のページと関連付けられ、ショップで発信されたツイートは、東京スカイツリーのショップのページに関連付けられ、Web ブラウザに提示されている。Web 閲覧ユーザは提示されたツイートを閲覧することで、東京スカイツリーといった場所全体だけでなくフロアや店舗に関する現状を把握できる。

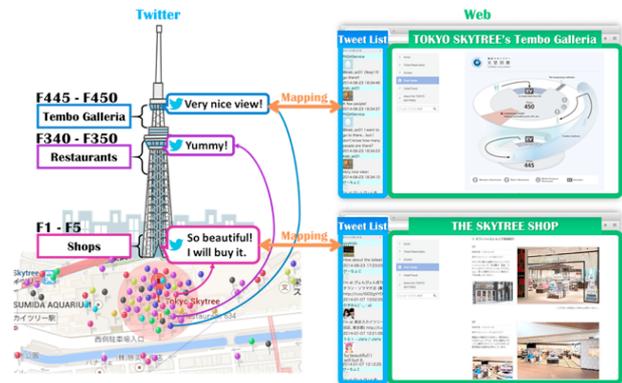


図 1: 複合施設のフロアならびに店舗に関するツイートのページへの対応

3 Web ページへのツイートマッピング

3.1 ツイートの緯度経度と内容に基づくツイート選別

本論文では、位置情報に基づく問合せを目的としており、ページとツイートを位置情報に基づき関連付ける。まず、指定地域から重複を除いた緯度経度情報を含むストリーミングツイートを The Streaming APIs を用いて取得し、取得したツイートに対して位置情報に基づいた内容判定を行い、ページと関連付ける。ツイートが発信された場所名と関連するか否かをツイートの内容から判定することで、ツイート発生場所と関係性の低いツイートの除去を行う。位置情報に基づいたツイート内容判定法は、一定範囲内の一定時間のツイートに多く出現する単語は関連性が高いと考え、場所名に対する特徴語として抽出し、特徴語を多く含むツイートを場所名に関連するツイートとして選択する。下記の式によりツイートの重要度を算出する。ツイート t に出現する各単語 i のツイートに出現する頻度を抽出し、その平均値を算出する。また、特徴的な単語が出現しても単語数が多い場合は、ツイートの重要度が低下するため、シグモイド関数を用いる。

$$\sum_{i=1}^m \left(\frac{\text{単語 } i \text{ が出現するツイート数}}{\text{ツイート総数}} \times \frac{1}{1 + e^{-x}} \right) \times \frac{1}{m}$$

m はツイート t に出現する単語総数である。

3.2 Web ページとツイートの対応付け

まず、店舗名を抽出するため、全てのツイートの緯度経度を用いて、Foursquare から場所に関する情報を取得し、それらから VenueID、店舗名、店舗の緯度経度、店舗の住所を抽出し、さらに店舗の階高情報、複合施設名を抽出し、重複除去し Venue 辞書を構築する。

An Evaluation of Tweet Mapping to Specific Web Pages based on Spatio-Temporal Analysis

^{†1} Yuu FUJIWARA ^{†1} Gouki YASUI ^{†2} Yuanyuan WANG ^{†1} Yukiko KAWAI ^{†1} Toyokazu AKIYAMA

^{†1} Kyoto Sangyo University

^{†2} Nagoya University

表 1: 各フロアのツイート集合から抽出された出現頻度の高い上位 5 単語

時間帯	フロア	コスメ, フード スイーツ (1 階)	レディス, メンズ ファッション (2~8 階)	ブックス, ライフ スタイルグッズ (9 階)	レストラン (10 階)	その他
(3) 12~14 時		朝	あかり	紀伊国屋	腹ごしらえ	好きだ
		蘭	週	フロッピー	山芋	パスタ
		贈り物	姉さん	寄り道	スンドゥヴ	月
		七夕	兄さん	ジム	韓国	菜
		天の川	新しい	イヤホン	たこ焼き	開花
(5) 18~20 時		丸	ジム	電球	回転	百百
		珈琲	ダーク	客	寿司	段
		但馬	マシュマロ	普段	カブリチョーザ	おれ
		ごちそう	ゆるい	放送	あたり	十
		パラダイス	スープ	うるさい	たこ焼き	小鳥

この Venue 辞書を用いて, Web 閲覧ユーザの閲覧している Web ページの場所名および店舗名を抽出する.

Web ページの場所名とフロア, 店舗名の抽出は, まず, ユーザの閲覧している Web ページの URL を取得し, その Web ページのスニペットを取得し, スニペットから出現頻度の高い単語を特徴語として抽出する. 形態素解析よりその特徴語の中から地名を判別し, 該当する単語をそのページの場所名とする. なお, 複数地名が抽出された場合は全てを場所名とする. そして, Venue 辞書を用いてフロアと店舗名を付与する.

また, 3.1 節で選別した発信場所に関連するツイートに対して, Venue 辞書を用いてシード語を選出し, k-近傍法による分類を行い, 各ツイートのフロアや店舗名を付与する. 以上により, Web ページならびにツイートに対する場所名 (例: スカイツリー), フロア (例: 1 階), 店舗名 (例: スカイショップ) 情報が付与され, Web ページとツイートが対応づけられる.

4 実験および考察

複合施設内のフロアおよび店舗に関する Web ページに集約情報を提示するため, 大阪駅直結の大型ショッピングモール「LUCUA」内のツイートに対して, 本手法を用いて抽出したツイートの特徴を検証する. 期間は 2015 年 10 月の一ヶ月分のデータから (1) 6~8 時, (2) 9~11 時, (3) 12~14 時, (4) 15~17 時, (5) 18~20 時, (6) 21~23 時の 6 つの時間帯で検証した.

(1) に関しては開店前のため, ツイートが存在しないフロアがあった. そこで (1) を除く全ての時間帯の単語の相関関係を比較したところ, (3) と (5) が最も高かったため, 本稿では (3) と (5) の時間帯によるツイート集合から出現頻度の高い特徴語を抽出した結果を表 1 に示す. (3) と (5) の時間帯における各フロアで抽出された特徴語を比較した結果, 各フロア間の上位 30 単語に対するスピアマンの順位相関係数は 0.3081~0.3776 であった. 表 1 より, 各フロアで抽出された特徴語はそのフロアに関連する単語であることが確認できる. また, フロアによっては, 時間帯が変化するにつれ話題が変化していることが確認できた. 例えば, 1 階はコスメとフードといった様々な種類の店舗が存在するが, (5) の時間帯には食に関する単語が多く出現している一方で, (3) の時間帯では食に関する単語はない. 本

屋と家電のフロア (9 階) でも同様の傾向であった. 逆に, 飲食のみに関連するフロア (10 階) ではほぼ全ての時間帯で食事に関連する単語が確認できる. 以上より, 時間帯に応じて分類されたツイートに対して適切なフロア情報を付与可能なことが確認できた.

5 まとめ

本研究では, 大型ショッピングモールや駅構内のような複合施設における各店舗などの小規模施設に関するツイートの発見を行い, ツイートに複合施設全体だけでなくフロアや店舗に関する分類を行い, そのツイートに関連する施設内のフロアや店舗の Web ページ上に該当するツイートならびにツイートの集約情報を提示するシステムの構築を目指した. Foursquare 情報を分析し, Venue 辞書を作成し, 各フロアと店舗に関する情報を取得した. さらに各フロアに関連するツイートの時間帯による分類を行い, 各フロアにおける特定の時間帯による話題性の高い単語の抽出が確認できた. 今後, 対応付けた Web ページに対するツイートの適合性について評価する予定である.

謝辞

本研究の一部は, 総務省 SCOPE (ICT イノベーション創出型研究開発), JSPS 科研費基盤研究 (B) (26280042) および基盤研究 (C) (15K00162) の助成を受けて実施された.

参考文献

- [1] Arakawa, Y., Tagashira, S. and Fukuda, A.: Relationship Analysis between User's Contexts and Real Input Words through Twitter, Proc. of UbiCoNet 2010, pp. 1751-1755, 2010.
- [2] Yamamoto, S. and Satoh, T.: Two Phase Extraction Method for Extracting Real Life Tweets using LDA, Proc. of AP-Web 2013, pp. 340-347, 2013.
- [3] Yamaguchi, Y., Amagasa, T., Kitagawa, H. and Ikawa, Y.: Online User Location Inference Exploiting Spatiotemporal Correlations in Social Streams, Proc. of CIKM 2014, pp. 1139-1148, 2014.
- [4] Wang, Y., Yasui, G., Hosokawa, Y., Kawai, Y., Akiyama, T. and Sumiya, K.: TWinChat: A Twitter and Web User Interactive Chat System, Proc. of CIKM 2014, pp. 2045-2047, 2014.