

地域ユーザに着目したロコミツイートの拡張

長島 里奈^{†,a} 関 洋平^{†,b} 猪 圭[‡]

[†]筑波大学情報学群知識情報・図書館学類 [‡]筑波大学図書館情報メディア系 ^{‡‡}つくば市役所

a) s1211530@u.tsukuba.ac.jp b) yohei@slis.tsukuba.ac.jp

概要 本研究は、Twitter上で、店舗についての評判やリアルタイムな状況について投稿した、ロコミツイートを収集する手法を提案する。応用として、つくば市の地域活性化イベントに協力するかたちで、つくば市のラーメン店舗のロコミツイートを収集した。実験では、地域ユーザを利用してロコミツイートを収集する手法について、つくば市民と推定したアカウントを利用して、評価を行った。地域ユーザの投稿に着目した手法は、従来手法を参考にしたベースラインと比較して、収集件数は平均で225.0%増加し、平均精度は0.81であった。さらに、地域ユーザのツイート群から抽出したキーワードにより拡張した結果、ベースラインと比較して、収集件数は平均で253.3%増加し、平均精度は0.71であった。

キーワード Twitter, 地域活性化, 地域ユーザ, ロコミ

1 はじめに

消費者によるロコミは地域の店舗に集客をもたらし、その集客は地域活性化に結びつく。本研究ではこのような効果のあるロコミの中でも、Twitter上のロコミ（以下、ロコミツイートとする）に着目した。

実際に地域活性化にロコミツイートが活用された例として「つくば福来らーめん紀行」[1]が挙げられる。この例では、著者らが出展店舗についてのロコミツイートを収集し、つくば市と協力して公式サイトに掲載した。

ロコミツイート収集の過程で、ツイートが店舗のロコミに適合するものか否かを人手で分類する必要がある。しかし、人手による分類には膨大な手間がかかる上、作業実行者に依存した個人差が発生してしまうという問題点もある。本研究では、これらの課題を解決するために、地域ユーザに着目することにより、自動的にロコミツイートを拡張する手法を提案する。

2 関連研究

地域の情報を利用してロコミツイートの収集を行う研究には、地域名を用いて飲食店の評判を抽出した那須川ら[2]の研究、地域に特色のある語から観光情報を抽出した小原ら[3]の研究、位置情報から観光ルートを抽出した中嶋ら[4]の研究がある。以上のように、ロコミツイートの収集にあたっては、地域に特色のある語や位置情報などが手がかりとして用いられている。

3 提案手法

本研究では、先行研究で用いられた地域名、位置情報に加え、地域ユーザが発信するツイートに着目して、以下の2段階の手法に基づき、ロコミツイートを拡張する。

1. 地域名、位置情報、地域ユーザを用いて、地域店舗のツイートを収集する。

2. 1の手法で収集した地域店舗のツイートから抽出したキーワードを用いて、ロコミツイートを拡張する。

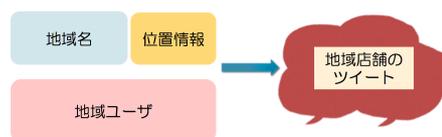


図1 地域ユーザによる地域店舗のツイートの収集

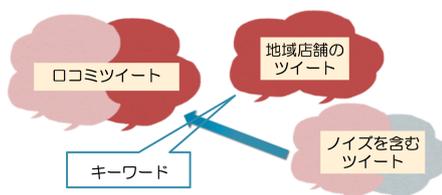


図2 キーワードによる地域店舗のツイートの拡張

4 地域ユーザに着目したツイートの収集

4.1 目的

地域ユーザの活用により地域店舗のツイート件数を高精度で増やせるかを検証する。

4.2 方法

地域ユーザ、地域名、位置情報を組み合わせた提案手法と、地域名、位置情報を組み合わせたベースラインの収集件数を比較し、増加率を確認する。

4.3 使用したデータ

実験の対象とするツイートデータは、「つくば福来らーめん紀行」[1]に出展した12店舗の店舗名を含むツイートとし、2015年8月1日～9月30日の期間収集した。また、地域ユーザのデータとしては、Twitterのプロフィール検索サービスであるツイプロ[5]に含まれるアカウントのうち、つくば市民に分類されている4,548名のデー

タから、フォロワー及びフォロワーのフォロワーについて拡張を行い、合計 30,555 件のアカウントを収集した。ただし、以下の条件のいずれかを満たすアカウントは削除した。これらの条件の閾値は、経験的に決定している。

1. フォロワーが 3 万人以上
2. フレンドが 3 万人以上
3. フォロワー数 \geq ツイプロつくば市民フォロー数 \times 50
4. フレンド数 \geq ツイプロつくば市民フォロー数 \times 50

4.4 実験結果

実験の結果を表 1 に示す。

表 1 提案手法のベースラインに対する収集件数の増加率

店舗名	提案手法		ベースライン		収集件数の増加率 (%)
	収集件数	精度	収集件数	精度	
小五郎	31	0.61	4	0.80	675
鶏々	151	1.00	33	1.00	358
たちばな	12	0.67	3	1.00	300
松辰	90	1.00	23	1.00	291
油虎	483	1.00	124	1.00	289
活龍	291	1.00	84	1.00	246
大勝軒	49	0.43	21	0.88	133
喜元門	247	1.00	107	1.00	131
おび屋	12	1.00	6	1.00	100
がむしゃ	86	1.00	44	1.00	95
五衛門	13	1.00	7	1.00	86
大樹	7	0.03	7	0.04	0
平均値	122.7*	0.81	38.6	0.89	225.0

* ベースラインに対して、t 検定で有意差があることを確認 (両側検定で有意水準 5%)

4.5 考察

平均増加率は 225.0%、平均精度は 0.81 であった。地域ユーザを活用することにより、高い精度を保ったまま口コミツイートの収集件数を増加させることができた。

5 キーワードによるツイートの拡張

5.1 目的

本節では、3 節で述べた提案手法の 1 段階目で収集したツイート集合 (4 節の実験の結果) から、キーワードを抽出することにより、口コミツイートを拡張し、高い精度で収集件数を拡張できるかを検証する。

5.2 方法

1. まず、地域店舗名を含むツイート集合のうち、3 節で述べた提案手法の 1 段階目で収集したツイート集合に含まれないツイートを収集する。このうち、提案手法の 1 段階目で収集したツイート集合から抽出したキーワードを含むツイートを、拡張する。
2. 収集件数と精度を評価し、ベースラインおよび、提案手法の 1 段階目の収集件数に対する増加率を確認する。

表 2 キーワードにより拡張したツイートの収集件数増加率

店舗名	収集件数	精度	収集件数の増加率 (%)	
			対ベースライン	対 4 節の手法
喜元門	308	1.00	188	25
活龍	362	1.00	331	20
がむしゃ	102	1.00	132	19
大勝軒	57	0.29	171	16
五衛門	15	0.54	114	15
鶏々	164	0.98	397	9
油虎	520	1.00	319	8
松辰	95	1.00	313	6
小五郎	31	0.01	675	0
大樹	7	0.00	0	0
たちばな	12	0.67	300	0
おび屋	12	1.00	100	0
平均値	140.4*	0.71	253.3	9.8

* ベースラインに対して、t 検定で有意差があることを確認 (両側検定で有意水準 5%)
たちばなとおび屋は、キーワードが抽出できなかったため、拡張を行わない。

5.3 実験結果

実験の結果を表 2 に示す。

5.4 考察

平均増加率はベースラインと比較して、253.3%、4 節の手法と比較して、9.8%であった。平均精度は 0.71 である。キーワードの活用により高い精度を保つことはできたが、口コミツイートの増加量はごくわずかであった。

6 おわりに

本研究では、地域ユーザのツイート集合を活用して、地域の店舗についての口コミツイートを集約する手法を提案した。実験では、地域ユーザ、地域名、位置情報を手がかりとして口コミツイートを収集し、収集したツイート群に現れるキーワードを活用することで、口コミツイートを高い精度で拡張できることを確認した。

今後は、キーワード抽出方法を改善し、収集件数のさらなる拡張について検討する。また、一般市民が利用できるインタフェースの実装を行う予定である。

謝辞

本研究の一部は、筑波大学研究基盤支援プログラム (B タイプ)、科学研究費補助金基盤研究 B (課題番号 25280110)、萌芽研究 (課題番号 25540159) の助成を受けて遂行された。

参考文献

- [1] つくば福来らーめん紀行。
<http://tsukuba-honmono.jp/noodle/> (参照 2015-09-10)。
- [2] 那須川哲哉, 吉田一星, 西山莉紗ほか: 大量のつぶやきから日本酒の美味しい店を発掘する, 言語処理学会第 21 回年次大会発表論文集, pp.820-823, 2015。
- [3] 小原基季, 森田和宏, 泓田正雄, 青江順一: Twitter 本文を用いた観光情報抽出及び分析システムの構築, 第 29 回人工知能学会全国大会, 4M1-4, 2015。
- [4] 中嶋勇人, 新妻弘崇, 太田学: 位置情報付きツイートを利用した観光ルート推薦, 情報処理学会第 158 回データベースシステム研究発表会, No.28, 2013。
- [5] ツイプロ. <http://twpro.jp> (参照 2015-09-10)。