

コミュニケーション支援のための Tweet と スポットの関連度の予測

片山一^{1,a)} 牛尼剛聡²

概要：本論文では Twitter 上で対象ユーザとそのフォロワーの交流が活発する可能性が高い活動を推薦する手法について述べる。具体的には観光サイトに投稿された口コミを利用し、事前学習済みの BERT に口コミの内容とスポットの対応付けを追加学習させることで、スポット予測モデルを構築する。そして、スポット予測モデルに対して対象とするユーザが過去のツイートを入力することで、ユーザが興味を持つスポットを予測する。同様に対象ユーザのフォロワーがツイートに対し「いいね」する確率を学習させたモデルに観光スポットの口コミを入力することで、フォロワーが興味を持つ可能性が高いスポットを予測する。これらの2つの結果に基づいて、対象ユーザとそのフォロワーが相互の趣味嗜好と各スポットの関連度を予測することで、Twitter 上でのコミュニケーションを活発化させる可能性が高い実空間上での活動を推薦する。

キーワード：BERT, Twitter, 活動推薦

1. はじめに

近年、SNS(Social Networking Service)が広く普及し人々が日常的に使用するサービスになった。Twitter は多くのユーザを有する代表的な SNS の一つである。Twitter は情報の収集や発信、他人とのコミュニケーション等様々な目的のために利用されている。Twitter にはリアルタイムで世界中から最新の情報が投稿されるため Twitter 上に存在する膨大な情報の中からユーザの目的に適した情報を発見・習得するために検索機能、トレンド表示機能など、様々な機能が提供されている。これらは、Twitter における情報収集を主たる目的とした機能であると考えられる。一方、Twitter は、知人との日常的なコミュニケーションを目的として利用されることも多い。これまで、Twitter の情報収集に関しては、多くの機能が提供され、活発に研究が行われている。しかし、Twitter におけるコミュニケーションを活発化するための機能に関しては、十分に研究が行われているわけではない。本研究では、Twitter におけるフォロー/フォロワー間のユーザ同士の活発なコミュニケーションを促すために手法を開発することを目的とする。

コミュニケーションが盛り上がるための要因には様々なものが考えられる。代表的な要因の一つとして、対象となるユーザグループの構成員において共通の要素が存在することがある。例えば、ユーザグループのメンバー間に共通の趣味がある場合や、同じ体験を行ったことがある場合である。グループのメンバー間で共通の体験が存在するのであればその体験に関する話題によってコミュニケーションが盛り上がりと考えられる。本研究では、ユーザに、Twitter 上でのコミュニケーションを盛り上げることに繋がる経

験をもたらす実空間での活動に着目し、ユーザに対して、そのような活動を推薦することで、ユーザが、実空間上で推薦された活動を行い、それに関する投稿を行うことで、その活動に興味を持つフォロワーからのリプライなどのリアクションが行われることが期待できる。

これまでも、SNS を利用した情報推薦に関する研究は活発に行われている。特に観光地などの人々が興味を持つスポット(Point of Interest:POI)に関する情報はユーザにとって需要が高く、SNS ユーザの趣味嗜好に合わせた POI を算出する様々な手法が提案されている[1][2]。しかし、既存手法の多くはユーザが興味を持つ可能性が高い POI を推薦することに着目しており、対象とする POI において具体的などのような活動が行え、どの活動をユーザが行うべきかについて推薦を行っているわけではない。

そこで、本研究では、対象ユーザとそのフォロワーコミュニケーションが活発化することが期待される投稿につながる経験をもたらす実空間上での活動を推薦する手法を開発することを目的とする。本研究ではそれぞれの POI で行うことが可能な活動とその特徴を抽出するために、観光サイト「じゃらん net」^{*1}に紹介されている口コミを利用する。

本論文では、活動推薦を行うための前提として、過去の Twitter への投稿から得たユーザ情報とスポットとの関連度の予測し、関連度の高いスポットの中で、どの口コミに強く影響されているのかを検証する。例えば、個々のユーザに対し関連度の高い POI として複合施設が提示された際に、一方のユーザは複合施設内にある映画館に対する口コミに反応し、もう一方のユーザは施設内のスポーツショップの口コミに反応したことが分かれば、同じスポットでも

1 九州大学大学院芸術工学府
Kyushu University, Fukuoka, Fukuoka 815-8540

2 九州大学大学院芸術工学研究院
Kyushu University, Fukuoka, Fukuoka 815-8540

a) katayama.hajime.498@s.kyushu-u.ac.jp

*1 <https://www.jalan.net>

それぞれのユーザの対し、推薦すべき活動が異なると考えられる。

本研究では各 Twitter ユーザに対して活動の推薦を行うことを目標とする。ここでの活動とは、「ある特定のスポットでユーザが行える体験」と定義する。一つのスポットにおいて行える活動候補は複数あると考えられ、それらをユーザへの推薦対象とする。

2. 関連研究

Twitter などの SNS の情報を分析し、推薦や検索のために POI やイベントを検出する研究は数多くみられる。

川崎ら[3]の研究では、チェックインに関する投稿だけでは大部分の POI に言及されないという問題に対し、チェックインに関する投稿だけでなく、通常の投稿の本文も解析し POI への言及の有無を判定するとともに、SNS 上で言及されていない POI に対しては回帰決定木モデルによって言及されていた際の言及数を推定し、各 POI の人気度の推定を行った。山田ら[1]は Twitter を利用し、日本語ツイートから地名に関するツイートを抽出した後イベント情報を抽出する手法を提案している。この手法では地名を含むツイートを抽出した後、Support Vector Machine[4]と Conditional Random Fields[5]を用いてイベントの名称、開催場所、開催期間の情報抽出を行った。山岸ら[2]はユーザ体験型推薦に着目し、Flickr の投稿データから Who, When, Where, What の4つの要素を分析することで趣味嗜好を推論するシステムを提案した。

観光サイトのレビューを用いた手法として吉田ら[6]の研究がある。この研究では Paragraph Vector モデル[7]を用いて観光サイトのレビューから「夜景がきれい」「待ち合わせ場所」といった各スポットが持つ特徴を算出し、これらの特徴を考慮した観光スポット検索を行う手法を提案した。

本研究は、Twitter の情報から対象ユーザの趣味嗜好を分析し、「じゃらん net」のロコミから対象ユーザの POI とそこでの活動を推薦することを目指している。共通の POI でもユーザの嗜好によって推薦される内容が変わるという点でこれらの研究とは異なる。

3. 提案手法

本節ではユーザの Twitter データを利用して、関連度の高い POI を推測し、どのロコミに強く反応を示すか検証する手法を説明する。提案手法の概要を図1に示す。

3.1 POI の推定

ユーザが興味を持つ可能性が高い POI の推定を行うために、自然言語処理モデルである BERT[8]を用いる。具体的には事前学習済みの BERT に対してロコミを入力として、そのロコミが対象としている POI 予測するというタスクでファインチューニングを行う。そして、推薦対象となる Twitter ユーザの過去の投稿や、「いいね」を押した投稿 wo

上記で学習した POI 予測モデル(BERT)に入力する。これらの投稿を利用する理由は、それらにはユーザの趣味嗜好を反映した内容が多く含まれると考えられるためである。POI 予測モデルに基づいて出力される POI の確率をその投稿と POI の関連度の強さと考える。対象ユーザの全ての投稿に対し同様の処理を行い、それぞれの POI が1位となった回数が多い順にランキングする。そして、ランキング上位の POI を対象ユーザの嗜好に合致する可能性が高い POI と推定する。

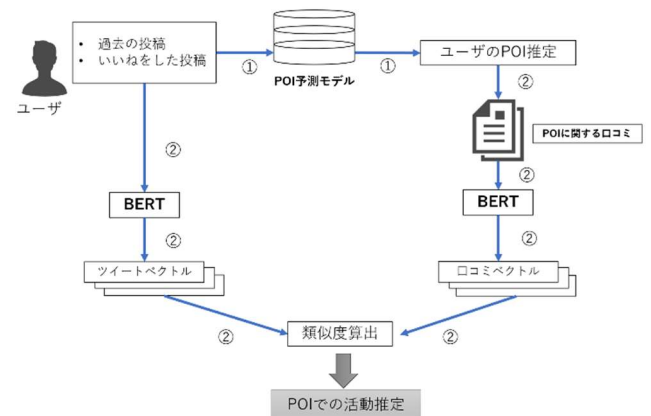


図1 提案手法の概要

3.2 POI のロコミとツイートの類似度測定

3.1 節でユーザの嗜好に合致する可能性が高いと推定した POI に関シテ、ロコミと Twitter の投稿を利用して、対象ユーザが興味を持つ可能性が高い活動を推定する。

図2は「じゃらん net」に掲載されている「キャナルシティ博多」に関するロコミを抜粋したものである。ロコミには「ドローン体験」や「トマトの酸味と野菜・豚肉の旨味がギュッと詰まったスープに、少し太めの綿が絶妙にマッチして美味!」といった、対象の観光スポットで行った活動や実際に現地に足を運ばないと知りえない情報が書かれている。また、その観光スポットで何を行ったかはユーザによって異なるため、ロコミに記載されている内容はそのスポットで行うことができる活動を幅広く表している。

そこで本手法では、対象となる POI に関するロコミの中から対象ユーザの興味と類似性があるものを抽出する。具体的には図3のようにツイートとロコミの投稿から類似度を計算し、その値が大きいほどユーザにとって関連度の高いロコミを興味のある活動と予測する。

ツイートとロコミの類似度計算を行う前処理として、ロコミのセンテンスに分割する。分割されたセンテンスと対象ユーザが Twitter で投稿・いいねをしたテキストを、それぞれ BERT を用いてベクトル化する。センテンスごとに区切りベクトル化したロコミテキストの集合を $R = \{r_1, r_2, r_3, \dots, r_n\}$ 、ベクトル化した Twitter のテキストの集合

を $T=\{t_1, t_2, t_3, \dots, t_m\}$ とするとき、 $r_i \in R$ と $t_j \in T$ の類似度を
コサイン類似度を利用して式(1)のように定義する。

$$\text{sim}(r_i, t_j) = \frac{\sum_k r_{i,k} t_{j,k}}{\sqrt{\sum_k r_{i,k}^2} \sqrt{\sum_k t_{j,k}^2}} \quad (1)$$

対象ユーザが投稿・いいねしたテキストベクトル t_j に対し、対象とする POI に関する全ての口コミベクトル r_i との類似度を求め、類似度が高い口コミを順に並べてユーザに推薦すべき活動として提示する。

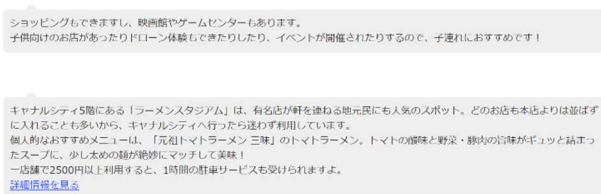


図 2 口コミの例

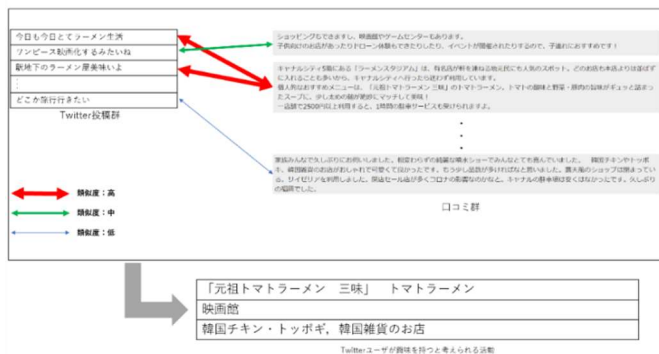


図 3 口コミとツイートから類似性ある活動を推定する例

3.3 予備実験

本実験では「じゃらん net」等の口コミサイトに投稿された口コミを利用して分類モデルを学習させ、学習したモデルにユーザの Twitter の投稿を入力する。このとき出力として予測された POI が、入力テキストの内容と関連性の高い POI であることが重要である。そこで予備実験として、以下の実験を行った。

- 福岡の観光スポット 13 件の分類モデル構成
- Twitter 投稿と見立てたテキストを入力し、出力結果の整合性の調査

表 1 にデータセットとして用いた観光スポットを示す。このデータセットには 1300 件の口コミが含まれ、学習時には訓練データ 60%、検証データ 20%、テストデータ 20% に分割し学習とそのモデルの評価を行った。利用する自然言語処理モデルとして、東北大学の乾研究室が公開している日本語事前学習済み BERT モデル*2 を用いた。

学習させたモデルにテストデータを入力した結果、81% の精度で分類を行うことができた。このことから、口コミ内容をもとに観光スポットの予測はある程度可能であると考えられる。

次に学習させたモデルに表 2 のような実験サンプルとして作成した Twitter 投稿テキストを入力する。入力結果とその分類スコアを表 3 に示す。結果から野球に関する内容のテキストは野球場である「PayPay ドーム」に分類され、「いい散歩コース知ってる人教えて」という内容のテキストは広い園内で景色などを楽しみながら散歩ができる「海の中道海浜公園」に分類されている。このように対象のスポットに関して直接レビューした文章では無いにも関わらず、テキストの内容から整合性のあるスポット分類が行えていると考えられる。

また Twitter の投稿には、その投稿内容自体にユーザの趣味嗜好を反映していないものも多くみられる。今回の例では「眠い」というテキストからはユーザの特性を読み取る事は難しく、出力された分類スコアも低い値となっている。本研究ではこのような投稿をノイズと考え、本実験の際は分類スコアのスケールリングを行い、閾値を超えた投稿のみを POI の推定に利用する。

表 1 福岡の観光スポット

マリンワールド海の中道	鷲尾愛宕神社
リパークルーズ	キャナルシティ博多
福岡市動植物園	マリノアシティ福岡
屋台	PayPay ドーム
博多シティ	海の中道海浜公園
大宰府天満宮	中州
九州国立博物館	

表 2 POI 推測のための予備実験の結果

テキスト	観光スポット	分類スコア
今日も今日とでラーメン生活	屋台	2.8856
山口の水餃館に行ってきました	マリンワールド海の中道	4.1039
ホークス優勝おめでとう	PayPay ドーム	4.0981
今年も甲子園の奇跡が来たね	PayPay ドーム	4.1216
カーブ三連放は笑えん	PayPay ドーム	3.6925
いい散歩コース知ってる人教えて	海の中道海浜公園	2.7704
無印良品が一番信頼できる	マリノアシティ福岡	1.4635
今期のアニメは面白い作品が多い	九州国立博物館	1.3802
京都に旅行行きたい	キャナルシティ博多	1.3986
神社巡りしたい	鷲尾愛宕神社	2.3971
眠い	キャナルシティ博多	0.5791
九州大学受かった	九州国立博物館	1.8187

*2 <https://github.com/el-tohoku/bert-japanese>

4. 評価実験と実験結果

4.1 実験方法

本提案手法の有用性を評価するために被験者実験を行う。被験者が投稿・いいねをしたツイート集合を入力として与えられたとして、Twitter テキストの入力からユーザーの特性と整合性のあると考えられる POI を推定できているか、主観評価により評価する。

POI に関するロコミと Twitter 投稿との類似度測、適切に活動に関するロコミを抽出できているか、抽出した活動とユーザーの特性の間に類似性がみられるかを主観評価により評価する。

4.1.1 POI 予測モデルの構成

評価実験ではユーザーの POI 推測モデルの構成のデータセットとして京都府の観光スポット 41 件に関するロコミ 4064 件を用いた。表 3 にデータセットとして用いた観光スポット(POI)を記載する。

予備実験と同様にデータセットを訓練データ 60%、検証データ 20%、テストデータ 20%に分割し、BERT に分類タスクで学習させた結果、精度 81%で分類を行うことができた。この結果からスポット数が増加した場合でも高い精度で分類を行うことができることが分かった。このモデルを用いて Twitter ユーザーの POI を推定する。

4.1.2 Twitter ユーザーの POI 推測

Twitter ユーザーの POI を推定するためのツイート取得のために Twitter API^{*3}を用いた。対象ユーザーとなる被験者の過去の投稿と、「いいね」をしたツイートを取得し、前処理としてリプライと英文ツイートの除去、ハッシュタグや URL、絵文字除去のテキスト処理を行った。前処理を行った 4861 件のツイートを 4.1 節で述べた POI 分類モデルの入力として用いる。

最も分類スコアが高いスポットとそのラベルを出力する。前章で述べたように分類スコアには幅があるため以下の処理でユーザーの POI を推定する。

- 分類スコアを式(2)によって正規化を行う
- 最大値を $\max(x)$ 、最小値を $\min(x)$

$$x'_i = \frac{x_i}{\max(x) - \min(x)} \quad (2)$$

- 閾値を 0.7 とし、正規化された分類スコアが閾値以上のとき POI 推定に利用する

閾値を超えたツイートは 236 件あり、分類回数が上位 5 件のスポットを表 4 に示す。

*3 <https://developer.twitter.com/ja>

4.1.3 POI のロコミとツイートの類似度測定実験

6.2 節によって求められた POI の中から、上位 3 件である「映画村」「京都国際漫画ミュージアム」「錦市場」のロコミとツイートの類似度を測定する。ロコミをセンテンスごとに分割を行うライブラリとして GiNZA^{*4}を用いた。

POI の上位 3 件の各ロコミを分割したテキストに対するそれぞれのツイートの類似度の式(1)によって求め、値が高い上位 3 件のロコミを表 5 に示す。

4.2 結果と考察

4.2.1 ユーザーの POI 推定

ここでは 4.1.2 の方法で求めた観光スポットについて POI として妥当性について考察する。今回用いた Twitter ユーザーの特性として、映画やアニメ・漫画に関する内容の投稿を行っている、もしくは「いいね」を押していることが確認できた。それらの投稿から上位に「映画村」「京都漫画国際ミュージアム」が上位となった結果は妥当性があると考えられる。しかし、これらのスポットに関する話題はどのユーザーに対しても比較的话题に上がりやすい内容と考えられる。対して、今回の実験で上位にランキングされた「錦市場」「瑞光窯」、「京都水族館」といったスポットは意外性の高い結果となった。説明可能性モデルなどを用いて今回のシステムに妥当性が認められた場合、提案手法によって意外性のある POI 推定が行うことが可能であると考えられるため、今後さらに検討を行う必要がある。

4.2.2 POI のロコミと類似度測定結果に関する考察

4.1.3 によって求められた、POI と推定された観光スポットのロコミと Twitter の投稿との類似度測定の結果について考察する。今回の実験結果では類似度が高いと計算された二つのペアのテキスト間に、類似性が強くみられるとは言えなかった。しかし、抽出されたロコミはその観光スポットで行える活動に加え、時間帯による客の多さやお店の閉店時間など現地に行くうえで知っておきたい情報を多く含む結果となった。これらの結果から、今回の手法のように 2 つのテキストを直接利用するのではなく、別の手法によって観光スポットでの活動を Twitter ユーザーの趣味嗜好に合わせて抽出できる可能性があると考えている。

5. まとめ

本研究ではコミュニケーション支援のための Twitter ユーザーの POI 及び POI での活動推薦を行うための前提として、「じゃらん net」に掲載されているロコミから Twitter ユーザーの興味嗜好に沿った観光スポットを推測し、ロコミのどの内容に類似性を持つかを検証した。

POI の推測は妥当性が認められたが、意外性のある POI の推測、POI に関するロコミと Twitter テキストとの類似性の調査にはさらに検討をする必要がある結果となった。今後は大量のロコミデータセットを利用したロコミモデルを構築し、より高精度な対応付けを行うことを計画している。

表 3 京都府の観光スポット

南座	下鴨神社
八坂神社	鴨川公園
京都府京都文化博物館	金閣寺
遊覧船	よーじや
京都国立博物館	キザクラカッパカントリー
保津川下り	錦市場
桂離宮	京都タワー
伏見稲荷	嵐山
琴引浜	三十三間堂
青蓮院	平和台公園
俵屋吉富(京都菓子資料館)	日吉ダム
サントリー京都工場	瑞光窯
東寺	京都るり溪温泉
アサヒビール大山崎美術館	平等院
京都水族館	清水寺
映画村	銀閣
伊根の舟屋	京都駅ビル
京都植物園	レンタル着物と可菜
京都国際漫画ミュージアム	かに一番
日本文化体験	京都鉄道博物館
天橋立	

表 4 分類された回数が多いスポット上位 5 件

順位	スポット
1	映画村
2	京都国際漫画ミュージアム
3	錦市場
4	瑞光窯
5	京都水族館

参考文献

- [1] 山田 渉, 菊池 悠, 落合 桂一, 鳥居 大祐, 稲村 浩, 太田 賢. マイクロブログを用いたイベント情報抽出技術. 情報処理学会論文誌, Vol57, No1, 123-132(Jan.2016)
- [2] 山岸 立, 馬 強. 個人適応型観光のためのユーザ体験推薦. 日本データベース学会和文論文誌. Vol.20-J, Article No.15, 2022 年 3 月
- [3] 川崎 仁嗣, 榎園 健, 深澤 祐介, 豊田 正史. SNS 上での言及を考慮した施設人気度の推定手法. 情報処理学会研究報告. Vol2019-DPS-179 No9, Vol2019-MBL-91 No9, Vol2019-ITS-77 No9. 2019/5/23
- [4] Corinna,C. Vladimir,V. Support-Vector Networks, Machine Learning, Vol.20. pp.273-297(1995)
- [5] John,L, Andrew,M, Feramdo.C. Conditional randomfields:Probablistic models for segmenting and labeling sequence data. Proc. 18th Internal Conference on Machine Learning, pp.282-289(2001)
- [6] 吉田 朋史, 北山大 輔, 中島 伸介, 角谷 和俊. ユーザレビューの分散表現を用いた主観的特徴の意味演算による観光スポット検索システム. DEIM Forum 2017 P6-5
- [7] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. In Proceedings of the 31th International Conference on Machine Learning, ICML 2014, pp. 1188–1196, 2014.
- [8] Devlin, J., Chang, M., Lee, K. and Toutanova,K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, CoRR,Vol. abs/1810.04805 (online), available from (http://arxiv.org/abs/1810.04805) (2018)

*4 <https://megagonlabs.github.io/ginza/>

表 5 ロコミとツイートの類似度

	ツイート	ロコミ	類似度
映画村	忙しすぎて鬱になりそう..自分の誕生日、もし可能なら1日、地上で寝たい。でもそうしたら色々な人たちに迷惑をかけてしまう。涙。弱音を吐いてごめん。頑張るしかないか！	5歳の娘には難しすぎてすぐ断念したので、あと一つ好きなやっけていいよ！といって順番待ちしているとその男性スタッフが、娘の首元のハーネスを無理やり持ち上げ退室する方は引っぱりあげ強制退室、娘は急に引っぱりあげられびっくりしていました	0.946
	新婚から思い出の詰まった鹿児島で最後に大きな買い物をして地元に戻ってきました。愛車とはこの事かというくらい愛着が湧いてます♪転勤族と結婚して大変な分、それ以上にその地で色んな場所に行けて沢山経験出来て慣れてしまえば楽しいです。次の赴任先が決まるまでは広島でも楽しく過ごします	コラボイベント期間中だけだとは思いますが、村内に隠れリラクマがたくさんいるとスタッフの方に声をかけていただけだったので、閉村時間ギリギリまで頑張って探してしまいました。	0.946
	アメリカの通信会社であるVerizonによる「通信ラグのある世界」を表現したCM。久々に新しいタイプの映像を見せられた感があって、一度見ただけで鮮烈な印象が残りました。とにかく映像のクラフト力が素晴らしい。今後ひとつの指標として残っていく仕事だと思います	また、単なる時代劇のロケセットの町並み、撮影所というのではなく、忍者屋敷やゲーム体験、今流行りの時代劇衣装へのコスプレ体験など、体験型アミューズメント施設となっている感じが凄くしました。	0.945
京都国際漫画ミュージアム	ついなんです。もち。それなりに苦労はしたけど、売り物みたいな出来栄えにはなりません。それでも自分なりに愛していました。それを友人のプロのカメラマンに振る舞いました。こういう写真に撮らせてしまったのは愛情でしょうか？気を遣わせてしまったのでしょうか？	調べていなかったこちらに落ち度はありますが、祝日もやってくれたら嬉しかったなと思いました。	0.947
	『いつか～one fine day』を観劇。いや、素晴らしかった。ミュージカルに興味がある方は、まだやってみないなので、張り付けておきます。	漫画好きな方は一日たっぷり漫画が読めるので本当にいいと思いますよ。	0.938
	いわた書店さんの一万円選書応募完了第三者の目線から本を選んでいただけるので、自分だと選ばない本に出会えると思うと凄く楽しみ！ご縁がありますように！	数年前にに利用させていただきましたが、自由に読める漫画本がたくさんあり、どこで読んでも良い自由さと途中で退館しても再入場もでき、ランチは外で済ませてから再入場もありなことも嬉しいです。	0.936
錦市場	vol.2終わりました。うーん、悔しさも残る場面もあるが皆さんからの後押しで乗り切れたのは自分でもびっくり。つくづく幸せ者です。離れの建築もだいぶええ感じになってきたので母屋のほうも覗きに行かないとね。とにかく本当にありがとう！岡野昭仁	あと、錦平野というところで朝食を食べたかったのですがそこはあさでもすごい行列で今回は諦めました。	0.942
	今日2回目のワクチンを打ってきました昼過ぎに打ったけど、もう副反応出ている…！これ以上酷くなりませんように	夕方6時すぎに着いたらもう店がやってなくて次の日にもいきませんが無駄な時間を費やしました。	0.937
	今日は島牧村の道の駅「よってけ！島牧」に中継で行ってきました。お店の中にはファイターズの顔ハメパネルがありました。何やら使い方を間違えたみたいです	ミニドーナツが沢山入って300円のやつでしたが、終わりがけに行ったのもあり、おまけしてもらえました！	0.936