

観光地の口コミに含まれる観光地への アクセス関連情報の抽出手法

塩田雄基 福本淳一 西原陽子

立命館大学大学院情報理工学研究科, 立命館大学情報理工学部

研究の背景(1/2)

観光地の公式サイトの情報だけでは得られないアクセス関連情報が口コミにはある

清水寺までの道中は

坂道になっており、

お土産屋などの店が並んでいる

 しかし...

清水寺の公式サイトでは

駅から徒歩で何分かかかるかしか書かれていない

画像1: 清水寺までの道中のようす

京阪電鉄 清水五条駅から

— 約徒歩25分

画像2: 清水寺公式サイトの交通案内

研究の背景(2/2)

しかし、ロコミサイトには

公式サイトに書かれている情報に加えて、駅から観光地までの道中の情報が書かれている。

アクセスは京阪電車だと「清水五条駅」ですが徒歩20分以上ですし清水寺のまわりは山の坂道でお寺内も坂道なのでゆっくり景色やお店などで過ごすのが良いでしょう。

画像3:じゃらの清水寺のロコミ

このような情報をユーザーに提示するシステムが必要



どのようにして大量のロコミからアクセス関連情報をとるのか

研究のアプローチ(1/2)

行くまでの坂道が結構きついです。でものぼっている途中の道なりでも、たくさんの喫茶店やお土産屋さんがあるので退屈はしませんでした。境内から見る景色はとてもきれいです。

必ずうかがう場所です。バス停から坂道を登り距離はありますがいろんなお店があるので飽きずに行けます。到着すると清水の舞台から飛び降りるという言葉があるように京都が一望できるような景色です。

画像4,5:じゃらの清水寺のロコミ

アクセス関連情報は「行くまで」や「バス停」などから始まっている

「境内」や「到着」など観光地に着いたことが分かる単語より前の文に書かれている



アクセスに関する文の始点と終点になっているキーワードを指定

その間の文がアクセス関連情報ではないか

研究のアプローチ(2/2)

開始キーワードと終了キーワードを用いて抽出



- ・アクセスに直接関連しない表現
- ・単純に場所を述べるだけの口コミ

例:「徒歩で清水寺」など、アクセス手段を示しているが、
具体的な評価や体験に関する情報ではない



名詞と形容詞、名詞と動詞の組を含む口コミのみを抽出

研究目的

- 口コミデータからアクセス関連情報を含む文を開始キーワードと終了キーワードを用いて抽出する
- 抽出された部分口コミからアクセス関連情報として不要な文を除去し、類似する口コミをまとめて代表文を抽出する

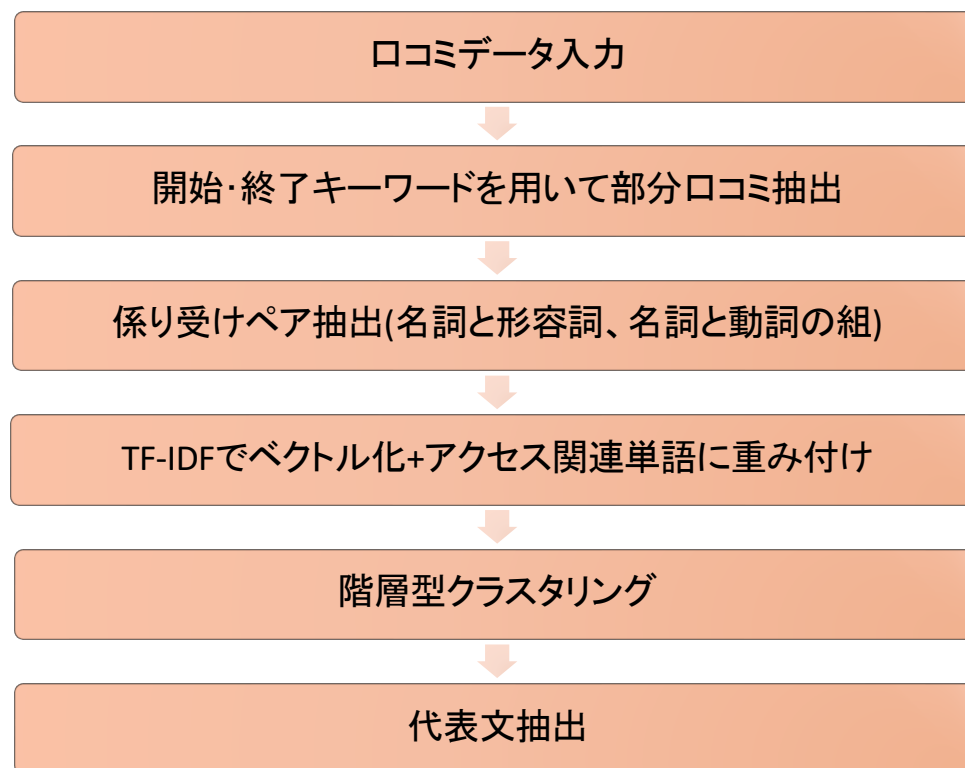


- 公式サイトでは得にくいアクセス関連情報(坂道の勾配、混雑、道中の店舗など)をまとめて提示する仕組みを構築する

関連研究

研究分野	主な研究内容	限界	本研究の位置づけ
口コミ分析 (Ye 2009, Banerjee 2016, Xiang 2017など)	ホテル予約行動への影響、施設評価、消費者意思決定	アクセス関連情報の抽出は対象外	アクセス関連情報を対象に分析
アクセス情報研究 (Zhang 2019, Yamada 2018, Isoda 2020など)	経路選択、観光ルート解析、混雑考慮	地理情報や移動パターン中心で、口コミの体験情報は未活用	口コミから坂道・土産屋・混雑状況などの体験要素を抽出

提案手法の概要



口コミデータの取得

旅行サイト「じゃらんnet」の”クチコミ”欄にある口コミが書かれた部分のみを取得
(下の画像のグレーの背景の部分)



[何度訪れても良い](#)

★★★★☆ 4.0 カップル・夫婦

建仁寺からバスで向かった。
こちら閉館間際だったので、バタバタと拝観。
何度訪れても清水の舞台からの景色は良い。夕暮れ時も相まってとても良い雰囲気だった。
御朱印も直書でいただいた。お寺さんらしく達筆でダイナミックで素晴らしいのだが、正直何て書いてあるのか分からなかった。

行った時期：2024年9月

投稿日：2024年12月4日



しゅーとさん 男性／30代

このクチコミは参考になりましたか？ はい 0

画像6:じゃらんの清水寺の口コミ

アクセスに関する口コミの抽出

手動で抽出したアクセスに関する口コミを参考に、アクセスに関する文の始点となっているキーワードと終点となっているキーワードを決定

始点は「駅」や「バス停」、移動手段などがキーワード

終点は観光地名や観光地の中の場所、「着きました」などの報告がキーワード

開始キーワード	清水道，京都駅，清水五条駅，祇園四条駅，バス停，駅から， 最寄りのバス停，徒歩で，歩いて，タクシーで，車で，バスで， 電車で，自転車で，清水寺まで，行くまで，清水寺への，参道，道中
終了キーワード	清水寺，到着，着きました，着きます，到着した，到着する，舞台，境内 行けます，行きました，辿り着きました，着いた，着く，おすすめ

画像7:開始キーワードと終了キーワードのリストの例(清水寺)

係り受けペア抽出(名詞-形容詞/名詞-動詞)

観光客の観光地のアクセスに関する評価や体験に関連する表現を抽出したい

抽出した際にとれてしまった必要の無い単語のみの口コミを除去したい

例: 徒歩で清水寺



名詞と形容詞(形容動詞含む), 及び名詞と動詞という組み合わせを含む

口コミのみを抽出

例: 坂道→大変、土産屋→多い

階層型クラスタリング

類似する口コミをまとめて、ユーザーに提示する口コミ数を削減する



- TF-IDFを用いてベクトル化
- アクセス関連単語(バス停、徒歩など)に重み付け
- Ward法による階層型クラスタリング

予備実験：部分口コミ抽出の精度評価

- 実験目的：開始キーワードと終了キーワードを用いて抽出した部分口コミがどの程度アクセス関連情報を持つのか調べる
- 実験手順：
 - 五つの観光地についてアクセス関連情報の抽出を行う
 - 抽出された部分口コミを手でアクセスに関して有益か判別
 - 適合率を調べる

$$\text{適合率} = \frac{\text{アクセス関連情報として有益だと判断できるもの}}{\text{抽出された部分口コミ数}}$$

- 実験データ：清水寺4,840件, 伏見稲荷4,837件, 平安神宮1,066件, 金閣寺2,095件, 銀閣寺1,185件

予備実験：実験結果

- 清水寺において抽出された部分口コミは、約79%の適合率
- 伏見稲荷や銀閣寺などでは適合率が50%台～40%台と低め
- アクセスに関する口コミが多い観光地は適合率も高くなっている
- 適合率の平均は61.6%で概ねアクセスに関する部分口コミが抽出可能

	全口コミ	抽出された口コミ	有益な口コミ	適合率
清水寺	4840件	228件	180件	約79%
伏見稲荷	4837件	119件	66件	約55%
平安神宮	1066件	42件	27件	約64%
金閣寺	2095件	29件	18件	約62%

評価実験

- 実験目的: 距離閾値を設定し, 適合率からこのシステムの有効性を評価する
- 実験手順:
 - 階層型クラスタリング(ワード法)を用いてクラスタリング
 - 距離閾値を2.0から2.4まで0.1ずつ変化させる
 - その結果を閾値ごとに人手でアクセス関連情報として有益か判別
 - 各観光地のクラスタ数と適合率を求める
- 実験データ: 清水寺4,840件, 伏見稲荷4,837件, 平安神宮1,066件, 金閣寺2,095件, 銀閣寺1,185件

実験結果：閾値の決定

閾値が 2.2 のとき,

- 閾値を増やしても適合率の上昇が少ない
- アクセス関連情報が少ない金閣寺において適切なクラスタ数になる



閾値を 2.2 に決定

距離閾値	2.0	2.1	2.2	2.3	2.4
クラスタ数	117	96	81	65	53
清水寺	0.51	0.53	0.69	0.70	0.73
クラスタ数	34	27	17	12	11
伏見稲荷	0.56	0.52	0.76	0.75	0.73
クラスタ数	11	8	4	2	2
平安神宮	0.55	0.625	0.75	0.50	0.50
クラスタ数	5	3	3	2	2
金閣寺	0.40	0.66	0.66	1.00	1.00
クラスタ数	13	10	6	4	3
銀閣寺	0.46	0.50	0.50	0.75	0.66

実験結果：適合率評価

●経路が多様な観光地(例：清水寺)

→ 得られた口コミが多く、効果的



●単純な経路(例：金閣寺)

→ 情報が少なく、精度低下



距離閾値	2.0	2.1	2.2	2.3	2.4
クラスタ数 清水寺	117 0.51	96 0.53	81 0.69	65 0.70	53 0.73
クラスタ数 伏見稲荷	34 0.56	27 0.52	17 0.76	12 0.75	11 0.73
クラスタ数 平安神宮	11 0.55	8 0.625	4 0.75	2 0.50	2 0.50
クラスタ数 金閣寺	5 0.40	3 0.66	3 0.66	2 1.00	2 1.00
クラスタ数 銀閣寺	13 0.46	10 0.50	6 0.50	4 0.75	3 0.66

実験結果：正しく出力された口コミ例

- 本研究で抽出したい口コミが正しく抽出された
- 実体験を元にした公式サイトには書かれていないような有用な口コミを抽出された



- 実際の観光地までの道中の様子が具体的に把握できる

観光地	例
清水寺	道中にたくさんお土産屋さんがあります
伏見稲荷	道中に景色の良い場所や神聖なスポットが随所にありますので、余裕のある方は山頂コースで巡って参拝されると良いかと思います
平安神宮	バス停も近く、地下鉄からは徒歩10分、アクセスは良し
金閣寺	道中もなかなかの見応え
銀閣寺	道中で様々なお店があるので楽しいです

実験結果：誤って抽出された口コミ例

- 京都らしさや観光地の雰囲気感想、カフェの紹介等、アクセス手段に関係ない口コミが抽出された
- アクセスに関係するが徒歩の情報ではなく、バスの情報が抽出された



- 本研究として抽出したいアクセス関連情報ではない

観光地	例1
清水寺	京都の中でもっとも京都を感じさせる場所、それが清水寺
伏見稲荷	参道の朱色の鳥居は圧巻です
平安神宮	テラス席もあるので観光に疲れたらコーヒーを飲みながら平安神宮
金閣寺	バスで、地下鉄・バス共通二日券もおすすめ
銀閣寺	バスで10分から15分ぐらいで到着します私は銀閣寺

まとめ

- 本研究ではロコミからアクセス関連情報を抽出する手法を提案した
- ロコミから開始キーワードと終了キーワードを用いて、部分ロコミを抽出し、クラスタリングを行い代表文を抽出した
- 抽出された代表文を有益か判断し、適合率を求めた
- 適合率は平均約70%となった
- この提案手法でアクセス関連情報を抽出することが可能