曖昧な地理表現に類似した旅行記探索システムの構築

青山 蓮[1,a)]　伊藤 正彦[1,b)]

概要：旅行者の観光ルートの決定を支援するための研究は数多く行われている．従来の提案手法では，ユーザーの嗜好に合わせて観光スポットの満足度を計算し，最も満足度が高かったものを結果として提案するものや，観光スポット間の類似度に着目し，偏りが無いように観光ルートを生成するものなどが存在する．しかし，従来の方法では事前に与えられた選択肢から嗜好を決定するものが殆どであり，旅行者の希望に沿った旅行を決定するためにはまだまだ改善が必要だと言える．そこで，本研究では，ユーザーが自由に文章を入力し，その文章と実際の旅行記との類似度を BERT を用いて計算することで，最適な旅行記を提案するシステムを構築する．

1. はじめに

文章には様々な地理表現がある．ここでいう文章とは，小説やエッセイ，旅行記といったものを指す．同時に，機械が文章を読み込んで処理をする技術である自然言語処理についても多くの研究がおこなわれている．機械が地名を正しく理解することは自然言語処理をするうえで重要な要素の一つだと言える．

実際に，文章から地名を抽出する方法の一つに固有表現抽出と呼ばれるものがある．固有表現抽出とは，人名や地名などの固有名詞や，日付や時間などの数値表現を抽出する技術である．例として「太郎」が人名，「北海道」が場所にあたる．このような固有表現は文章のほとんどに出てくるが，これに当てはまらないものも存在する．例えば，「海の見える神社に行きたい」という文章があった場合，この文章自体の意味は分かるが，実際にどこを指した言葉であるのかを特定するのは容易ではない．このように，一目見ただけでは一意に場所を特定できない文章を，本研究では「曖昧な地理表現」と定義する．

実際にこの表現が現れる例として，旅行の計画などが挙げられる．どこに行きたいかは決まっていないが，何をしたいかが頭の中に思いついている場合に曖昧な地理表現は用いられやすい．また，このような表現が増える程，自身で旅行先を決定するのも難しくなる．ここで，曖昧な地理表現に対する相応しい提案方法として，実際にどういった旅行が行われたかをユーザーに可視化することだと推察した．実際の旅行記を見ることで，入力として与えた曖昧な地理表現に対して一つの答えを提示するとともに，旅行全体をイメージしてもらえることが期待できる．そのため本研究では，入力された文章と，データセットに格納された旅行記との類似性について考慮し，類似性が高い旅行記に関して，それを提案する手法を行う．

[1] 北海道情報大学
[a)] s2121054@s.do-johodai.ac.jp
[b)] imash@do-johodai.ac.jp

2. 関連研究

大内ら [1] は，文章内の地名を読み取り，それを実世界の地図に反映（ジオグラウンディング）させる課題に取り組んでいる．また，その中でデータセットを公開している [2]．「地球の歩き方旅行記データセット [3]」と呼ばれる「地球の歩き方 web」に投稿された旅行記をテキストデータに変換したものがあり，彼らはその中に現れた地名表現に対して，関連する OpenStreetMap（以下 OSM とする）等のリンク情報付与（アノテーション）などの処理を施している．これによって，実際の地理情報と関連して文章を分析することが可能になった．実際の旅行記を用いていることもあり，ここまで日本の地理情報に対応しているデータセットは他にない．そのため，本研究でも同様のデータセットを用いることにする．

倉田 [4] は，ユーザーの嗜好をもとに観光ルートを自動生成し，対話的にルートの変更を行うことができるシステム CT-Planer3 を構築している．システムの中では，旅行全体の所要時間や観光スポットに対して訪れたいか否かをユーザーが設定でき，その設定内容に応じて観光ルートが自動生成される．観光ルートの生成には遺伝的アルゴリズムを用いている．

また，武信ら [5] は観光ルートを自動生成する中で，提案された観光ルートに偏りが生まれないことを目的とした

© 2025 Information Processing Society of Japan.

図 1 システムの利用手順

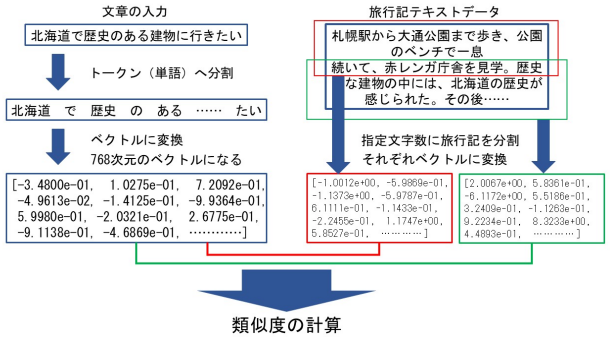

図 2 類似度計算の概要図

図 3 スクレイピングの概要図

研究を行っている．偏りがあるかどうかについては観光スポット間の類似性に着目しており，実際の観光情報サイトに掲載されている説明文を用いることで，各スポット間の説明文が似ているものを類似性が高いとみなしている．説明文が似ているかどうかの分析については，BERT と呼ばれる自然言語処理モデルの一つを用いている [6]．

本研究では，旅行記の文章とユーザーが自由に入力した文章との類似度について BERT を用いて計算していく．旅行記は，その作者が実際に観光スポットを訪れた感想や，その場所についての説明が記載されている．そのため，旅行記の文章はユーザーの嗜好を考慮するうえで一定の有用性があると考えている．

3. システム概要

本研究で作成するシステムの手順を図 1 に示す．まず，文章を入力してもらう．ここで入力する文章とは，「自然豊かな街へ行きたい」のような，具体的な地名を指していない，つまり本研究で定義した曖昧な地理表現を想定している．入力された文章に基づいて，旅行記データセット [3] との類似度計算を行う．これは，BERT に文章を与え，獲得したベクトル同士のコサイン類似度を文章間の類似度として算出している（図 2）．範囲は-1 から 1 の範囲で表され，値が 1 に近づくほど，二つの文章は似ていると言える．また，ユーザーには変数 n の値を入力してもらう．n は類似度が高かった結果の表示を上位 n 位まで表すための変数であり，例えば n = 5 にした場合は類似度が高い旅行記上位 5 個を表示する．その後に，類似度が高い旅行記を画面上に表示し，その文章内に現れた地名を [2] を用いて地図上に表示する．これは，地名にアノテーションされた OSM 情報をスクレイピングすることで座標を取得し，それを地図に反映させる手法を行う（図 3）．地図の表示には folium[*1] を使用する．

ベクトルを獲得する際には，文章の前後関係を失わないようにするため，旅行記全体を BERT への入力として与えている．しかし，旅行記の文章量は作者によってさまざまであり，約 800 文字を超える内容に関してはそのまま扱うことができないことが分かった．これは，BERT モデルが扱える文字数に限界があることが問題であると考えられる．そこで，文字数が多い旅行記に関しては，指定の文字数に区切る処理を施すことですべての旅行記に対応できるようにする．また，ある程度意味を失わないようにするため，文章を区切る際は一定文字数を被せての分割を行う（図 2（右））．分割した場合，それぞれの文章のまとまりと入力した文章との類似度を算出し，そのなかで一番値が高かった類似度を，文章全体の類似度として扱うことにする．

4. 実行事例

ここでは，曖昧な地理表現を三つの種類に分類し，それぞれについて実際にシステムを用いた結果を示す．本研究の定義である曖昧な地理表現は，以下に分類される．

- 地名を用いていない文章
  例：「海が見える神社に行きたい」「スイーツが美味し

[*1] Python で地図を可視化するためのライブラリ https://python-visualization.github.io/folium/latest/

© 2025 Information Processing Society of Japan.

いホテルに行きたい」

- 地名を用いた文章
  例 :「札幌で歴史のある建物に行きたい」「東京で景色が良い場所に行きたい」
- 地名を形容表現として用いた文章
  例 :「神戸のような街に行きたい」「函館山みたいなきれいな景色を見たい」

人によって旅行の決め方は様々であり，すでに行きたい都道府県や街が決まっている場合もあれば，過去に訪れたことのある場所が印象に残っており，似たような場所はないかと場所や地名を形容詞として用いて調べる場合もある．そのため，上記の三つに分類することで，本システムはユーザーの多種多様な希望に沿った結果が出力されるのかについて調べていく．また，システムを使用する際は例として挙げたものを用いる．n の値は 1 とし，類似度が一番高かった旅行記のみを結果として表示する．

4.1 地名を用いていない文章

用いる文章は「海が見える神社に行きたい」とし，類似度が一番高かった旅行記を図 4 に示す．この旅行記との類似度は 0.538 であった．実際に文章を読んでいくと，この旅行記の筆者は「塩釜神社」に訪れている．そこで「高台にある塩釜神社からは，遥か海が望めます」と述べている．つまり，今回の用いた文章である「海の見える神社に行きたい」についてはしっかりと希望に沿ったものが現れていることが分かる．

4.2 地名を用いた文章

用いる文章は「北海道で歴史のある建物に行きたい」とし，類似度が一番高かった旅行記を図 5 に示す．この旅行記との類似度は 0.565 であった．こちらの文章を読んでいくと，作者は北海道庁旧本庁舎に訪れており，そこでの感想を「記念室はとても重厚な雰囲気です．カーテンなどの調度品も歴史を感じます.」と述べている．つまり，地名を用いた文章の場合でも，希望に沿った結果が現れていることが分かる．

4.3 地名を形容表現として用いた文章

用いる文章は「神戸のような街に行きたい」とし，類似度が一番高かった旅行記を図 6 に示す．この旅行記との類似度は 0.56 であった．文章を読んでいくと，作者は神戸に訪れており，今回の文章である神戸のような街という希望は叶っていないことが分かる．これは,「みたいな」や「のような」といった形容表現に対応する旅行記が無く，入力した文章に存在する名詞に対して，それに合致する旅行記があれば類似度が高くなってしまったのが原因だと思われる．今回であれば，旅行記データセットの中で「神戸のような街」と感想を述べている旅行記は無く,「神戸」「街」とそれぞれの単語が含まれている旅行記を類似度が高いものとして扱われてしまったのが原因ではないかと考える．

4.4 地図の表示

次に， 4.1 説の実行事例において類似度が高かった旅行記の文章中に現れた場所を地図上に示す（図 7，図 8）．図 7 および図 8 では，図 4 の文章の中に現れる「塩釜神社」や「仁王島」「瑞巌寺」といった地名が，地図上にプロットされていることがわかる．塩釜神社と瑞巌寺の距離は約 10km 程であり，この文章だけでなくとも，実際に旅行記の中に現れた地名と地名との距離感は，文章を読んだだけでは簡単に分からない．そのため，旅行記の表示と地図の表示の組み合わせは有用性が高いと言える．

5. おわりに

本研究では, ユーザーが自由に文章を入力し, 旅行記をテキストデータとしたデータセットとの類似度について計算することで, 類似した旅行先を表示するシステムを提案した. 類似度の計算には BERT を使用している．4 章にて実際に本研究のシステムを使用してみた結果, 入力した文章に関してしっかりと希望に沿った旅行記が提案されており, この手法には一定の有用性があると考えられる．しかし， 4.3 節のように，特定の文章ではうまく類似度がはかれなかったことも分かった。今回は「神戸のような街に行きたい」という形容表現を用いたが，今後は他の地名を用いた形容表現についても実行事例を調査し，もし同様の結果が出た場合はユーザーの入力に一定の制約を設けるのも検討している．

今後の展望として, 地図の表示結果に関してシステムを改良する必要がある．今回は, 場所に OSM の node が付与されているもののみ地図上に表示を行ったが, 実際には way や relation など, 特定の座標一点のみを表しているわけではない情報が多数付与されている．しかし, 実際に, OSM の way リンクにアクセスすると, Wikidata が載っているため, そこにアクセスすることで緯度経度の値を取得できると考えている．今回はシステムの作成のみにとどまっているが, 本当に有用性があるのかについてもしっかりと実験を行う必要がある．

謝辞 本研究では，国立情報学研究所の IDR データセット提供サービスにより株式会社地球の歩き方から提供を受けた「地球の歩き方旅行記データセット」を利用しました．貴重なデータを提供いただき, この場を借りて感謝の意を表します..

参考文献

[1] 大内啓樹. 「地理空間情報と自然言語処理」プロジェクト. https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-22H03648/.

© 2025 Information Processing Society of Japan.

旅の第一目的地は塩釜神社です。
神社創建の年代は詳らかではありませんが、平安時代初期頃の書物にはその名前があるそうで、歴史の古い由緒正しき神社なんですねー。
神社の雰囲気って、いいですよね。本宮と別宮がコの字型に並んでいます。それぞれにお祭りしている神様がいらっしゃいます。
塩業や漁業の守護、家内安全、延命長寿、交通安全、厄除け、とりわけ安産守護の神様として信仰されているそうです。
同上。
高台にある塩釜神社からは、遥か海が望めます。庭園も美しいです。
・・・・・・
鐘島。仁王島。海猫には餌付けができます。餌はもちろんかっぱえびせん。海から見た五大堂です。素敵な風景ですね！
・・・・・・
ホテルをチェックアウトしてから塩釜水産物仲卸市場に行ってきました。
ホームページによると、ここは決して、“観光化されたお土産市場”ではなく、厳しい目利きの“プロ”が買付に来ている市場だそうです。
・・・・・・

図 4　4.1 節の文章と類似度が高かった旅行記 一部省略

旭川空港着陸、10数分前の北海道の大地です。北海道初上陸からなのか、果てしなく続く雪景色に感動してしまいます。
航空機のｺｯｸﾋﾟｯﾄ内で1時間40分遅れの到着ですが、雄大な大地がそんなちっぽけなことを忘れさせてくれそうです。
予定していたﾊﾞｽは乗り遅れてしまいましたが、たまたま同じように旭山動物園に向かう方々とﾀｸｼｰ料金をｼｪｱして旭山動物園に到着です。
・・・・・・
北海道庁旧本庁舎は現在使われている新庁舎ができるまで約80年間に渡って道政を担ってきた赤ﾚﾝｶﾞつくりの建物です。
記念室はとても重厚な雰囲気です。ｶｰﾃﾝなどの調度品も歴史を感じます。
そして、北海道庁旧市庁舎で一番目を引くのは、１階の玄関を入ってすぐのところにあるこの階段かもしれません。とても立派な階段です。
通称「赤れんが庁舎」と言われるこの建物は1888(明治21)年にｱﾒﾘｶ風ﾈｵ・ﾊﾞﾛｯｸ様式の建物です。
館内は無料公開されていて樺太や北方領土の展示もされています。
札幌市時計台は正式名称を「旧札幌農学校演舞場」と言います。現在は高層ﾋﾞﾙ群に囲まれていて鐘の音も時計台周辺でしか聞こえないのだそうです。
・・・・・・

図 5　4.2 節の文章と類似度が高かった旅行記 一部省略

元教会を改装して作られたカフェ兼パン屋さんです。
神戸好きなら誰もが知っている、かなり有名なＳＰＯＴです。そしてここのパンやドイツ系でかなりおいしいんです！！特にバケット系が最高でした。
元教会のカフェの中です。まさに教会そのもの。
窓から差し込む優しい光が、雰囲気を盛り上げます。二階部にはパイプオルガンが置かれていたところもわかります。
ここにステンドガラスを入れれば、一体どんな雰囲気なんでしょう？なんて想像しちゃいました。
・・・・・・
背後には六甲山から連なる高い山もあり、神戸は山も海もあり、本当に美しい町です。
小高い山の上から、神戸の町を見下ろしています。本当にきれいですね。
港町であり、とても上品な町です。山と海と両方有する神戸ならではの景色です。夜景も美しかったです。
神戸の夜景は「日本三大夜景」の一つとして挙げられています。
・・・・・・

図 6　4.3 節の文章と類似度が高かった旅行記 一部省略

図 7　図 4 の旅行記内に現れた場所を出力した地図

図 8　図 7 の A を拡大した地図

[2] 東山翔平, 他. 日本語旅行記ジオパージングデータセット atd-mcl. 言語処理学会 第 30 回年次大会, 2024.
[3] 株式会社地球の歩き方. 地球の歩き方旅行記データセット, nov 2022.
[4] 倉田陽平. Ct-planer 3：Web 上での対話的な旅行プラン作成支援. 観光科学研究, No. 5, pp. 159–165, 03 2012.
[5] 武信雄平, 奥野拓. 観光スポットの多様性を考慮した観光ルート推薦システムの構築. In *DEIM2023*.
[6] Jacob Devlin, et al. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019*, 2019.

© 2025 Information Processing Society of Japan.