

細粒度のユーザ属性に基づく観光地レビューの検索 Search for Sightseeing Spot Reviews Based on Fine-grained User Attributes

徳留 壮哉¹⁾ 山本 岳洋¹⁾ 湯本 高行¹⁾
Souya Tokutome Takehiro Yamamoto Takayuki Yumoto

1 はじめに

近年インターネットの普及に伴い、インターネット上で情報を収集することが増えている。その一つに、旅行に行く際にじゃらん²⁾や Tripadvisor³⁾などの観光地サイトに掲載されているレビューを閲覧し参考にすることがある。

実際に Tripadvisor は、どのくらいの人がレビューを参考にしているか、レビューの影響力を調査⁴⁾するため、アメリカ・イギリス・イタリア・シンガポール・オーストラリアの五か国で、9000 人を対象にアンケートを実施した。その結果、図 1 のように、9000 人の内、レストランの口コミでは 70%、観光スポットでは 77%、宿泊施設では 82%の人が、口コミは「きわめて重要、またはとても重要」と回答した。これらのことから、多くの人がインターネットで情報を収集する際に、レビューを参照し、どこに行くか決めていることが多いと考えられる。しかし、人気の観光スポット、レストラン、宿泊施設では、膨大な数と種類のレビューが寄せられ、自分自身の特徴に合ったレビューを探すのには時間や手間がかかってしまう。その解決策として、じゃらんや Tripadvisor などでは、「投稿日」や「行った時期」でレビューの見やすさを変更できる並び替え機能があるが、これらの機能を駆使してもレビューを閲覧する数は変わらず、自分自身に合ったレビューを探すのには時間がかかってしまう。

調査結果

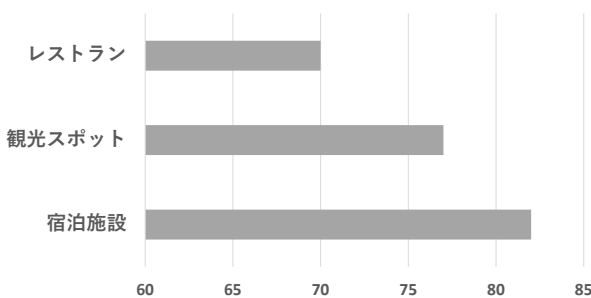


図 1 Tripadvisor : 口コミの影響力

そこで本研究では、大規模言語モデルを使用し、膨大な数と種類が寄せられたレビューの中から、ユーザの属性に合ったレビューを検索することを目的とする。

例えば、表 1 をある観光地スポットに寄せられたレビューとする。表 1 に寄せられたレビューの中から、今回は「足腰の悪い高齢者」という属性に推薦できるレビューを獲得する。「足腰の悪い高齢者」は、「砂利道が

多いと段差が多く躓いてしまう」や「駐車場から観光地までの距離が遠かった場合疲れてしまう」という属性の特徴がある。そのため本研究では、表 1 から表 2 のようなレビューを「足腰の悪い高齢者」という属性に対して推薦できるレビューとして獲得することを目的とする。

このように膨大な数と種類が存在する観光地レビューからユーザの属性に合ったレビューを獲得するため本稿では、大規模言語モデルで属性に関連するレビューを生成し、大規模言語モデルによって生成されたレビューと、観光地サイトに掲載されているレビューとの類似度を計測し、類似度を並び替えた時に上位に来たレビューをユーザに推薦できるレビューとして推薦する手法の提案を行い、他 2 種類との手法と提案手法との比較検証を行うことで、提案手法の妥当性を検証した。

本論文の構成について説明する。第 2 章では、本研究と関連する研究を紹介する。第 3 章では、本論文で提案する手法と類似度の計算方法について述べ、第 4 章では提案した手法の評価について述べる。第 5 章ではまとめと今後の展望について述べる。

表 1 観光地サイトに寄せられたレビュー

- ①空気が綺麗ですごく気持ちよかったです。
- ②周辺には飲食店が多くて嬉しかったです。
- ③境内は広く、砂利が多かったのでやや苦労しました。
- ④景色が綺麗でした！もう一度行きたいです。

表 2 足腰の悪い高齢者という属性に該当するレビュー

- ③境内は広く砂利が多かったのでやや苦労しました。

2 関連研究

ユーザの特徴に合ったレビューを推薦する研究は、盛んに行われている。

2.1 スコアを用いたレビュー推薦

安倍ら [1] はホテルに対するレビューの分析、評価項目に詳細化した上で評価項目別にスコアを付与し、キーワード、特徴、程度などホテルの評価に関する評価表現辞書を項目別に作成する提案、およびそれを用いてレビューのスコアリングを行った。その後、評価表現辞書に基づきスコアリングされた結果を用いることで、既存のホテル推薦手法の妥当性の確認実験を実施した。

林ら [2] はユーザがホテル宿泊後レビューを投稿し、投稿された不満レビューを評価表現辞書を用いてスコアリングしレビュースコアを作成した。その後、同様の評

1) 兵庫県立大学 University of Hyogo

2) <https://www.jalan.net/>

3) <https://www.tripadvisor.jp/>

4) <https://www.tripadvisor.jp/powerofreviews.pdf>

価表現辞書を用いてホテルスコアを作成し、レビュースコアとホテルスコアから不満点を解消する不満解消スコアを作成し、不満解消スコアと一致度の高いホテルを検索しレビュー投稿者に推薦するシステムを提案した。

住友ら [3] は観光者の嗜好性に合わせた観光スポットの魅力度を数値化し、得られた観光スポットの魅力度を用いて観光者の移動時間を考慮した観光経路の構成アルゴリズムを提案した。

野津ら [4] は Twitter に投稿された TV 番組についての tweets の感情を 10 種類の感情軸に分類し、作成した 10 次元の感情ベクトルを作成した。その後 TV 番組間の感情ベクトルのコサイン類似度を求めることで、TV 番組の推薦を行う手法を提案した。

田熊ら [5] はユーザが順位付けしたホテルの評価項目を基にホテルを推薦する手法を提案している。投稿されたレビューから評価項目ごとの名詞の数をカウントし、数が多い順に評価項目の順位付けを行い、ユーザが類似度が高いレビューを推薦している。

2.2 ベクトルを用いたレビュー推薦

村瀬ら [6] はマンガのストーリーに応じた漫画を、利用者の指定した作品を基にベクトル空間モデルを用いて類似度を判定し、推薦する手法を提案している。

櫛見ら [7] は投稿されたレビューを用いて映画の特徴ベクトルを生成し、ユーザが選択した映画の特徴ベクトルを他の映画の特徴ベクトルの類似度を算出することで、同様の好みの要素を持つ映画を推薦できる手法を提案した。

大山ら [8] は、ゲームレビューサイトから抽出したレビュー文をもとにゲーム辞書を作成した。その後 Word2vec を用いた類似度計算によって入力データとゲーム間の類似度が高いゲームの抽出・提示を行うシステムの提案をした。

2.3 ユーザの特徴を用いたレビュー推薦

林部 [9] はユーザの旅行の状況を聞き出し、その状況に合わせた宿を提案するシステムの構築を目指すため、「肯定的事実」と「推薦対象」という言語知識を宿のレビューテキストから抽出する手法を提案し、肯定的事実を推薦対象にアピールポイントとして用いることが妥当であるかの評価をクラウドソーシングを用いて実施した。

村中ら [10] はユーザの映画視聴履歴と映画評価からユーザの嗜好を抽出し、ユーザ特有の評価基準を考慮した映画推薦手法の提案を行った。

上述の通り、ユーザによって投稿されたレビューや画像、特徴ベクトル、スコアリングなどを分析し、情報推薦を行う研究が数多く行われている。しかし、本研究にて実施している大規模言語モデルを用いてユーザ属性の特徴を踏まえたレビューを生成し、観光地サイトに寄せられたレビューと類似度を比較し、ユーザ属性に合った観光地レビューを推薦する研究は行われていない。

3 属性に該当するレビュー検索

本研究では、観光地に訪れたことがある人がその観光地についてどのような印象を抱き、どのような体験をしたかを記録したレビューを扱いたかったため、株式会社リクルートが提供する観光地サイトじゃらん²⁾でスクレイピングを行い、愛知県名古屋市の観光地名所「徳川美術館」に寄せられたレビュー 539 件を用いた。

3.1 前処理

今回使用するレビューは句点で区切り、区切られた文章同士で類似度を計測するため、生成されたレビュー、実レビューそれぞれを表 3 のように句点で区切る。本研究では、じゃらん寄せられた観光地レビューを実レビューとする。句点で区切った後に類似度比較の段階に移行する。

表 3 分割レビューの詳細

分割前レビュー

徳川美術館は館内が広く、快適に過ごすことができ、非常に充実した時間を過ごせました。雰囲気も良く、再度訪れたいと思える場所でした。

分割後レビュー

- ①徳川美術館は館内が広く、快適に過ごすことができ、非常に充実した時間を過ごせました
- ②雰囲気も良く、再度訪れたいと思える場所でした。

3.2 大規模言語モデルによる仮想レビュー生成

ユーザ属性に合った観光地レビューを推薦するため本研究では、大規模言語モデルによって生成されたレビューと、じゃらん寄せられた観光地レビューとの類似度を計測し、類似度が高い順に並び替えた際に上位にくるレビューを、ユーザに推薦できるレビューとして考えた。大規模言語モデルとしては、OpenAI⁶⁾が提供する gpt-3.5-turbo1 を用い、大規模言語モデルに表 4 のようなプロンプトを与え、属性の特徴を踏まえたレビュー生成を行う。

表 4 レビュー生成に用いるプロンプト

あなたは観光地レビューから人間の特徴に合ったレビューを抽出しようとしています。

上記の情報を踏まえて、「車を持っていない大学生」という属性に対して、推薦できるレビューを 10 個考えてください。またレビューを生成する際に、「車を持っていない大学生」という単語を含まず、レビューを生成してください。

またプロンプトにおいて、「属性名に関連するレビューを生成してください」という単純なプロンプトでは、属性名を含めたレビューが生成されることが多かった。そのため、生成レビューと実レビューとの類似度計測において、生成レビューに属性名が含まれると属性名と類似している表現が含まれた実レビューの類似度が高く計測され、正解データが獲得できる割合が低かった。

そこで本研究では、プロンプトにおいて、「属性名を含まずレビューを生成してください」という文をプロンプトに挿入した。その結果、属性名が含まれてレビュー

6) <https://platform.openai.com/docs/api-reference/chat>

表 5 足腰の悪い高齢者 関連度上位 10 件 ※関連度は第五位で四捨五入

順位	実レビュー	関連度
1	徳川家に伝わる大名道具や工芸品などの展示があり、貴重な品々を見ることができました。歴史好きな方には特におすすめです。レストランやカフェもあり、休憩もできます。	0.9147
2	隣に徳川園があるので、美術館で作品鑑賞をした後、庭園を散策。一日ゆっくり過ごすことができました。街の喧騒を忘れ、静かに癒される素敵な場所です。また訪れたいです。	0.9145
3	日本の風情が楽しめて、ゆったりした気持ちになれました。刀や兜、茶の道具などたくさんの展示物がありました。アクセスも、街中にあるため、訪れやすかったです。	0.9106
4	尾張徳川家の品々が展示されております、交通の便もよく近くにカフェともありのんびりとした休日過ごすことができます。	0.9105
5	歴史的な資料が所狭しと並んでいます。想像していたよりも、見ごたえ十分です。時間をかけてゆっくりと見るのがオススメです。	0.9077
6	徳川美術館は徳川家の展示物が多数あって、歴史好きにはたまらない美術館になっています。ゆっくり見ることができました。	0.9055
7	徳川家の品々が展示されている徳川美術館。歴史が好きな方にはとてもおすすめです。ゆっくり見ることができました。	0.9055
8	静かでゆったりとし佇まいです。徳川の関係物だけでなく能楽堂などもありました。館内は広くゆっくりと展示品を見学できます。	0.9040
9	名古屋の観光スポットです。徳川の素晴らしいものが、展示されています。美術館の周囲は庭園になっており、散策するのもいいです。	0.9030
10	場所がいいところにあり、景観が楽しめました。企画展も興味深いものやっていることが多いです。徳川にまつわる物が多く、刀やよろいなど珍しいものがたくさん展示してありました。	0.9012

が生成される確率は低下した。また本研究で実行したプロンプトでは、属性の特徴を踏まえたレビューを大規模言語モデルに 10 件生成してもらうよう記述した。

3.3 レビュー類似度比較

生成レビューと実レビューの類似性を計測するため、コサイン類似度を用いる。

本研究では、句点で区切られた生成レビューと、句点で区切られた実レビューとの類似度比較を OpenAI 社が提供する埋め込みモデル text-embedding-ada-002 を使用し、句点で区切られたレビューをベクトル化する。その後、ベクトル化されたレビューをコサイン類似度で計測する。

例えば表 3 の文章が生成レビューだとすると、分割後レビューにおいては、2 つの文章が獲得できる。そして獲得した後、2 つの文章をベクトル化する。同様に、実レビューにおいても、分割後に獲得できた文章が n 個だった場合、 n 個の文章をベクトル化する。こうして獲得できたベクトルを、分割した 1 個目の生成レビュー文章と分割した 1 個目の実レビュー文章との類似度をコサイン類似度で計測する。同様に、分割した 1 個目の生成レビューと分割した 2 個目の実レビューとの類似度をコサイン類似度で計測し、この作業を分割した 2 個目の生成レビューと分割した n 個目の実レビューとの類似度をコサイン類似度で計測するまで行う。すべての通りでコサイン類似度が計測されたのちに、一番類似度が高く計測された類似度をその実レビューの類似度とする。その後すべての実レビューでこの類似度比較を行い、全ての実レビューで類似度を計測する。今回の研究では実レビューが 539 件あったため、この流れを 539 回繰り返し、合計 539 個のコサイン類似度を抽出した。そして、計測された類似度を基に、類似度が高く計測された順に並び替え、上位に来た実レビューがユーザーに推薦できるレビューと考えた。

以下に分割した生成レビューと分割した実レビューとの類似度計算、ならびに一番類似度が高く計測された実

レビューの文章を取得する式を示す。式においては、実レビューを r 、生成レビューを g とし、分割後の実レビュー r_i が実レビューに含まれていることを、 $r_i \subseteq r$ と表記する。

$$sim(r, g) = \max_{r_i \subseteq r, g_j \subseteq g} sim(r_i, g_j)$$

次に、ユーザ属性に対するレビューの関連度を表現する式を以下に示す。表 5, 6 の関連度においても同様の意味で扱う。式においては、 a はユーザ属性、 $gen(a)$ はユーザ属性 a に関して生成されたレビューの集合として表記する。

$$rel(a, r) = \max_{g \in gen(a)} sim(r, g)$$

4 実験

4.1 実験方法

本稿では生成レビューと実レビューを句点で区切った後、類似度を計測し、類似度が高く計測された実レビューをユーザ属性に該当するレビューとして検索を行った。この提案手法の性能を比較するため、他 2 種類の手法でも属性に推薦できるレビューが獲得できるかの検証を行った。

一つ目の比較手法は、生成レビューと実レビューとの類似度を計測する手法である。前述の通り本稿では、生成レビューと実レビューを句点で区切り、類似度を計測した。そこで比較手法として、生成レビューと実レビューを句点で区切らず、レビュー文のまま類似度計測を行い、ユーザーに推薦できるレビュー獲得を試みた。

二つ目の比較手法は属性名と実レビューとの類似度を計測する手法である。属性名が「足腰の悪い高齢者」だった場合、この属性名と類似しているレビューを獲得する手法である。

以上三つの手法で比較を行い、提案手法の妥当性を評価する。

表 6 車を持っていない大学生 関連度上位 10 件 ※関連度は第五位で四捨五入

順位	実レビュー	関連度
1	場所もわかりやすく駐車場、美術館内も広いところがよかったです。美術館はあまり行かないのですが、すごく楽しめました	0.8958
2	徳川的美術品を数多く展示していると聞き、名古屋へ行くなら、ぜひ行きたいと思っていた美術館。美術館までの道のりも、とても趣があり、紅葉や苔の生えた道を歩くだけでも、静かで癒されました。中に入ると、歴史的な美術品が多くありました。千利休や紫式部の直筆や、徳川家康や織田信長が使っていたという価値のある美術品がぞくぞく。教科書で見たことのある人物のゆかりの品々をみて、はるか昔の時代の生活や文化にふれる、素敵な時間を過ごしました。	0.8957
3	行った時は、ご即位記念ということで、宮廷で生まれた文化についてわかりやすく説明されていました。名古屋駅から市営バスで行くと便利です。	0.8946
4	ボランティアガイドさんによる案内がある時間に合わせて伺いました。ガイドの方は、すごく良く勉強をされていて、とても面白かったです。売店の奥に無料ロッカーがあるので、荷物を預けて身軽に見学すると良いです。	0.8941
5	貴重な展示品や徳川家関連の品など時間を忘れるくらいゆっくりと見る事が出来ました。都市部にあるので交通の便はいいのですが、観覧料が少し高いかなという印象。	0.8940
6	館内は予想以上に広く、展示品も充実しています。駅バスターミナルから出る観光巡回バス「メーグル」が他も立ち寄るのでアクセスもしやすいです。	0.8938
7	場所がいいところにあり、景観が楽しめました。企画展も興味深いものやっていることが多いです。徳川にまつわる物が多く、刀やよろいなど珍しいものがたくさん展示してありました。	0.8937
8	日本の風情が楽しめて、ゆったりした気持ちになりました。刀や兜、茶の道具などたくさん展示物がありました。アクセスも、街中にあるため、訪れやすかったです。	0.8932
9	地下鉄の駅からは少し離れているので、車かバス、めーぐるでの来館をお勧めします。大変貴重な物が展示されており、見ごたえがあります。そこまで広くないです。	0.8906
10	徳川のお殿様のお宝満載。「戦国武将」や「武家の文化」が大好きな方には、たまらない施設です。地味ですが、隣接する蓬左文庫や、お殿様のお屋敷のお庭・徳川園もぜひ見ておきたい場所です。最寄りの地下鉄等の駅からは、やや離れていますが、観光ルートバス・メーグルバス（金ピカで派手な外観にややびっくりしますが）のルートに入っていますので、名古屋駅発で名古屋城などと合わせて観光される計画なら、このメーグルバスの 1 Day チケット（大人 500 円）を購入して移動に使うのが便利だと思います。また、このチケットで入館料が割引になりますし、その他いくつかの施設の入場割引や、レストラン等の優待サービスもあってお得です。	0.8905

4.2 実験結果

本研究では足腰の悪い高齢者「車を持っていない大学生」二種類の属性で、属性の特徴に該当するレビューを獲得を実施した。表 5 に分割した生成レビューと分割した実レビューとの類似度を計測し、類似度が高く計測された順番に並び替えた際に出現した「足腰の悪い高齢者」という属性に対し推薦できる 10 件のレビューを示す。一方で表 6 には、分割した生成レビューと分割した実レビューとの類似度を計測し、関連度が高く計測された順番に並び替えた際に出現した「車をもっていない大学生」という属性に対し推薦できる 10 件のレビューを示す。

表 5 の結果からは、足腰の悪い高齢者に推薦できるレビューは 2 件と非常に少ない数での獲得ができた。表 6 の結果からは、「車を持っていない大学生」という属性において推薦できるレビューは 6 件獲得することができた。次に生成レビュー・実レビューを句点で分割し類似度を計測する手法を提案手法、生成レビュー・実レビ

表 8 車をもっていない大学 正解データ獲得件数

レビュー件数	提案手法	比較手法①	比較手法②
10 件	6	3	2
20 件	9	7	6
30 件	14	8	10
40 件	15	10	11
50 件	22	13	12

表 9 足腰の悪い高齢者 適合率

レビュー件数	提案手法	比較手法①	比較手法②
10 件	0.200	0.500	0.100
20 件	0.200	0.400	0.050
30 件	0.267	0.300	0.067
40 件	0.275	0.275	0.050
50 件	0.300	0.260	0.080

表 7 足腰の悪い高齢者 正解データ獲得件数

レビュー件数	提案手法	比較手法①	比較手法②
10 件	2	5	1
20 件	4	8	1
30 件	8	9	2
40 件	11	11	2
50 件	15	13	4

表 10 車をもっていない大学 適合率

レビュー件数	提案手法	比較手法①	比較手法②
10 件	0.600	0.300	0.200
20 件	0.450	0.350	0.300
30 件	0.467	0.267	0.333
40 件	0.375	0.250	0.275
50 件	0.440	0.260	0.240

一を句点で分割せず類似度を計測する手法を比較手法①、属性名と実レビューとの類似度を計測する手法を比較手法②とし、上位 10 件、20 件、30 件、40 件、50 件でどのくらい正解データを獲得することができたかを表 7, 8 で示す。同様に上位 10 件、20 件、30 件、40 件、50 件での適合率を表 9, 10 で示す。

「足腰の悪い高齢者」という属性では、上位 10 件や、上位 20 件とランキング数が少なければ、比較手法①で正解データを獲得できる割合が高いことが判明した。しかし、ランキング数が多くなればなるほど、提案手法で獲得できるレビュー数は多いことが判明した。一方で「車をもっていない大学生」では、上位 10 件や、上位 20 件とランキング数が少なければ、「足腰の悪い高齢者」とは異なり、比較手法①より、提案手法で正解データを獲得する割合が高いことが判明した。また、ランキング数が多くなったとしても、提案手法で正解データを獲得する割合が高いことが判明した。そのため、外部の特徴が含まれた属性「車をもっていない大学生」においては、ランキング数の数問わず、提案手法を使用したほうが良いと考えられる。

また、比較手法②の属性名と実レビューの類似度を計測する手法であるが、提案手法と比較した時、正解データの獲得の数の差が大きく開いた。このことから、単純に属性名と類似するレビューを探す手法より、属性の特徴を踏まえたレビューと実レビューを比較したほうが、ユーザに推薦できるレビューを獲得する精度が高いことが判明した。

4.3 考察

表 7, 8 から、正解率としては「足腰の悪い高齢者」という属性より、「車をもっていない大学生」という属性のが高いことがわかる。この獲得できた数の少なさにの要因として、生成レビューに含まれている単語に強い影響を受けたことが挙げられる。今回大規模言語モデルには、属性の特徴を踏まえたレビューを 10 件生成させた。生成し、10 件のレビューの内に「ゆっくりと休みながら観光を楽しむことができました」という表現が含まれたレビューがあった。今回の観光地「徳川美術館」においては、「時間をかけてゆっくりみるのがオススメです」や「ゆっくりみることができました」といったレビューが数多く寄せられていた。そのため、「ゆっくり」という表現で類似度が高く計測され、上位 10 件では正解データを数多く獲得することができなかつたと考えられる。

これらのことから、「ゆっくり」といった副詞などが生成レビューに含まれてしまうと、実レビューに「ゆっくり」などの副詞が含まれている場合が多いとき、ユーザに関連性のないレビューを獲得する可能性が高くなり、正解データを獲得できる精度が低下することが判明した。

一方「車をもっていない大学生」という属性の特徴を踏まえ、大規模言語モデルによって生成されたレビューには、地下鉄やバスといった公共交通機関に関連した表現が含まれたレビューが生成された。また、実レビューにおいても、「公共交通機関を利用して来館した」というレビューが多くみられた。そのため、「足腰の悪い高齢者」という属性よりも「車をもっていない大学生」という属性の、上位 10 件に含まれる正解データの割合

が多くなったと考えられる。

これらのことから、ユーザに推薦できるレビュー獲得に当たり、精度を向上させる課題としては、大規模言語モデルを用いたレビューの生成だと考えられる。

また本研究では、生成レビューにおいて、「属性名を含まずレビューを生成してください」と単純なプロンプトを大規模言語モデルに与えた。その結果、表 11 のように「徳川美術館」という場所の情報が含まれたレビューが生成されないことがあった。例えば、徳川美術館には坂がないにもかかわらず、表 11 の①のような、観光地内には坂が含まれているというレビューが生成されることがあった。そのため、①のような生成レビューと類似している実レビューが見当たらず、本当に欲しい形の実レビューを獲得することができる可能性が低下したと考えられる。

表 11 足腰の悪い高齢者に関する生成レビュー

望まなかった形の生成レビュー例

- ①観光地内には坂がなく、歩きやすい環境でした。足の痛みを気にせず、景色を楽しむことができました。
- ②設備が整っており、エレベーターや手すりがあるので、足腰の弱い方でも安心して観光できます。
- ③観光地名所へのアクセスが良く、歩きやすいコースがあります。

また、今回の大規模言語モデルによる生成レビューでは、表 11 の生成レビュー例の②のように、比較的どの観光地にも当てはまりやすいレビューが生成される傾向があった。表 11 の生成レビュー例の③においては、逆の表現として考えられる「観光地名所へのアクセスが悪く」といった実レビューとの類似度が高く計測された。そのため、ネガティブな表現とポジティブな表現は表裏一体として考えられ、一つの表現で複数の正解データを獲得することができると考えられる。一方で、比較的どの観光地にも当てはまりやすいレビューが大規模言語モデルによって生成される傾向があったため、「一通り見学するには時間がかかる」などといった、時間がかかるという「足腰の悪い高齢者」にとっては苦痛を強いられてしまうという美術館ならではのレビューを獲得することができなかつた。

5 まとめと今後の展望

本研究ではユーザ属性に合ったレビューを検索するため、大規模言語モデルによって属性に関連するレビューを生成し、観光地に寄せられたレビューと生成されたレビューとの類似性を計測することによって、ユーザに推薦できるレビュー検索手法を提案した。そして、他の手法と比較において、外部の特徴を踏まえた属性、「車をもっていない大学生」に関しては推薦できるレビューを獲得できる精度が高かった。しかし、身体的特徴を踏まえた属性、「足腰の悪い高齢者」に関しては、ランキング上位で獲得することができず、比較手法の方が精度が高くなった。これは、大規模言語モデルによって生成されるレビューにおける副詞の有無によるものと考えられる。このことから、副詞を含まないレビューを生成することで、正解データをより多く獲得できることが考えら

れる。

今後は、さらに属性の特徴を踏まえたレビューを獲得するため、外部の独自の情報を組み合わせることで、大規模言語モデルの回答精度を向上させると言われている RAG などの技術を扱い、今まで獲得することができなかった観光地レビュー獲得を目指す予定である。また提案手法においては、句点毎のレビューでの類似度比較により、膨大な時間がかかってしまうという課題があった。そのため、システムとして実用できる形にしていく手法を考えていく。

謝辞

本研究は JSPS 科学研究費助成事業 JP24K03228, JP21H03775, JP22H03905, による助成を受けたものです。ここに記して謝意を表します。

参考文献

- [1] 安倍克, 中島伸介, “レビュー自動スコアリング方式に基づくホテル推薦システム”, DEIM Forum 2020 P1-28(day1 p93).
- [2] 林貴宏, 尾内理紀夫, “Web 上のレビューを利用した映画推薦システム”, 人工知能学会論文誌, vol.30, No.1, pp.102-111, (2015) .
- [3] 住友千将, 岳五一, “観光者の嗜好性を考慮した観光経路構成アルゴリズムの実装と実証実験”, パーソナルコンピュータ利用技術学会論文誌, Vol. 15, No. 1, pp. 13-20, 2021.
- [4] 野津貴之, 若原徹, “放送中に投稿された Tweet を用いた TV 番組の推薦”, 情報処理学会第 80 回全国大会, 7K-04, 2018.
- [5] 田熊浩二, 福原楓, 亀井清華, 藤田聡, “口コミを用いた嗜好抽出に基づくホテル推薦手法”, 情報処理学会第 78 回全国大会, 2016.
- [6] 村瀬尊好, 柊和佑, 安藤友晴, “マンガの概要に基づく作品推薦システム”, FIT2012 第 11 回情報科学技術フォーラム N-013,2012.
- [7] 櫛見圭司, 北山大輔, “ユーザレビューを用いた全体的・部分的観点の類似に基づく映画推薦”, DEIM Forum 2017, B5-4, 2017.
- [8] 大山光暉, 竹山佳成, 平田圭二, “レビュー分を考慮したゲーム推薦システムの実現に向けた単語類似度調整の取り組み”, エンタテインメントコンピューティングシンポジウム (EC2017), 2017.
- [9] 林部祐太, “宿レビューからの肯定的事実と推薦対象の抽出”, 言語処理学会 第 25 回年次大会, 2019.
- [10] 村中清史, 横井健, “映画データのメタデータに対するユーザ特有の重要度を用いた映画推薦手法の提案”, 第 76 回全国大会講演論文集, Vol. 2014, No. 1, pp. 619-620, 2014.