

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

武信 雄平[†] 奥野 拓^{††}

[†] 公立はこだて未来大学大学院システム情報科学研究科 〒041-8655 北海道函館市亀田中野町116番地2

^{††} 公立はこだて未来大学システム情報科学部 〒041-8655 北海道函館市亀田中野町116番地2

E-mail: [†]{g2121033,okuno}@fun.ac.jp

あらまし 観光客の負担軽減を目的とした、観光ルートの作成を支援する研究が多く行われている。従来の支援手法では、ユーザの嗜好から観光スポットがユーザに与える満足度を算出し、最も満足度が高い組合せの観光ルートを推薦している。しかし、満足度を個別に算出しているため、観光ルートに含まれる観光スポットに偏りが生まれる場合がある。そこで本研究では、観光スポット間の類似性に着目した推薦を行う。まず、多目的遺伝的アルゴリズムを用いて観光ルートを生成する。その後、観光スポット同士が類似しているほど、ユーザに与える満足度を減少させることで、ユーザの嗜好に合った多様性のある観光ルートを推薦する。観光スポット間の類似度は、BERTを用いて観光スポットの説明文から獲得したベクトルを用いて算出する。

キーワード 観光、経路推薦、Web情報システム

1 はじめに

観光には、時間に限りがある場合がほとんどであり、観光先の全ての観光スポットを訪問することは難しい。そのため、事前に訪問する主要な観光スポットを決定し、効率的な観光ルートを作成しておくことが多い。効率的な観光ルートを作成するためには、観光スポットの場所や営業時間、公共交通機関の運行時間などを把握し、適切に組み合わせる必要がある。しかし、これらの情報は様々なWebサイト上に分散しているため、観光ルートを一から作ることは観光客にとって大きな負担となる。そのため、観光客の嗜好を考慮した観光ルート作りを支援する研究が多く行われている。

倉田は、ユーザの観光スポットに対する嗜好を基にした観光ルートの自動生成と、対話的なルート変更が可能なCT-Planner3というシステムを構築している[1]。このシステムでは、各観光スポットにおける滞在時間と各観光スポット間の移動時間の和が制限時間内に収まるように、ユーザに与える推定効用が高い観光ルートを生成している。観光スポットがユーザに与える推定効用は、ユーザの嗜好と観光スポットに対する評価を用いて算出している。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。住友らは、観光スポットの魅力度が最も高い観光スポットから優先的に訪問し、制限時間内に収まるよう観光ルートを自動生成する手法を提案している[2]。この手法では、観光スポットの魅力度をユーザの観光カテゴリに対する興味の度合いと、観光中に重視する感情の度合いを用いて算出している。

これらの手法では、それぞれの観光スポットのみを訪問した際に得られる満足度を基準に観光ルートを生成している。そのため、観光スポット間の類似性が考慮されておらず、観光ルート内の観光スポットに偏りが生まれる場合がある。そのような場合、限られた時間内に経験できる観光体験の種類が少なく

なってしまい、観光客が得られる満足感が低下してしまう可能性がある。そこで本研究では、ユーザの嗜好に加えて観光スポット間の類似度を考慮して観光ルートを生成することにより、ユーザにとってより満足度が高いと考えられる多様性のある観光ルートを推薦する。これを実現するために、観光スポット間の説明文間の類似度を考慮した観光ルートを自動生成・推薦するシステムを構築する。

2 関連研究

川崎は、商品推薦の手法として、単独の商品ではなく、複数の商品を組合せた全体価値を考慮した推薦手法を提案している[3]。この研究では、被験者に実際に商品を選択してもらう実験を通して、商品選択の方法と順序を分析している。この実験では、30種類のケーキを対象に、20種類の素材が使用されているか否かを基に、クラスタ分けと類似度の算出を行っている。その後、それらの結果を考慮して、ユーザの商品選択方法を分析している。その分析結果として、「直前に選択された商品を含むクラスタとの距離が近いクラスタから商品が選択されやすく、距離の離れたクラスタからは商品が選択されにくい」、「連続で同じクラスタから商品が選択されにくい」、「クラスタ内で距離の近い商品が選択されやすい」、「ユーザの商品選択は、既に選択した商品の数によって変化する」という4つの特徴があると述べられている。

庄司は、複数の商品を組合せる過程を分析することにより、高い組合せ価値を生み出す汎用的な手法を提案している[4]。組合せ価値とは、組合せによって生じる価値のことである。この研究では、楽曲やファッショントピックを対象とした商品選択の実験から、複数の商品を組合せる過程に共通する傾向として「全体として統一感や方向性があること」、「個々の要素に適度な多様性があること」、「主役と脇役の分担が適正であること」の3つを挙げている。

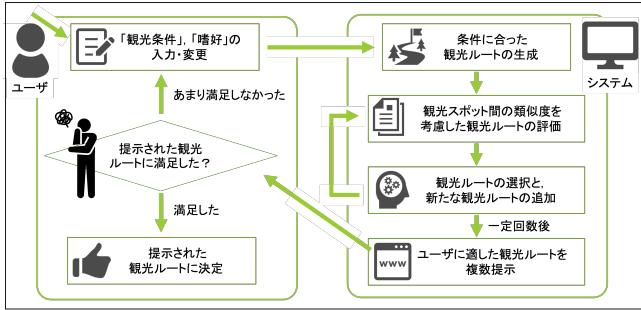


図 1 観光スポットの多様性を考慮した観光ルートシステムの使用手順

これらの研究から、複数の商品を組み合わせる際に、ある程度異なる特徴を持つ商品を組み合わせることが効果的であることが分かる。そのため、本研究では多様性を考慮した観光ルートを生成することにより、ユーザに与える満足度の高い観光ルートを生成する。

橋口らは、妊娠中あるいは産後間もない時期の悩みを対象に、コミュニケーション型質問応答サイトから類似した悩みの発見を支援する手法を提案している[5]。この手法では、BERT[6]を用いて学習した事前学習済みモデルをファインチューニングし、悩みを記述した文章間の類似度を算出することにより、類似した悩みを発見し、推薦している。BERTの事前学習済みモデルには、日本語版 Wikipedia を用いて事前学習したモデル¹を用いている。また、悩みの類似性に特化した類似度を算出するために、BERTのファインチューニングの手法として、悩みの状況を表す単語かどうかを推定する手法と、2文が類似しているかを直接推定する手法の2つを提案している。本研究に、BERTの事前学習済みモデルをファインチューニングすることにより、観光スポット間の類似度を算出する。

3 多様性を考慮した観光ルート推薦システム

本研究で構築するシステムの使用手順を図1に示す。システムではまず、観光条件として出発・終着地点、出発・終着時刻をユーザに入力してもらう。次に、ユーザの嗜好として、事前に設定した観光スポットのカテゴリをそれぞれどの程度好むかと、観光ルートに含まれる観光スポットにどの程度多様性があることを好むかをそれぞれスライダーを用いて入力してもらう。その後、入力された観光条件とユーザの嗜好から複数の観光ルートを生成する。観光ルートの生成手法は、4章にて詳しく説明する。生成した観光ルートから、ユーザの嗜好に合った観光ルートを複数提示する。提示する観光ルートの決定方法は、7章にて詳しく説明する。提示した観光ルートをユーザに確認してもらい、ユーザの満足する観光ルートではなかった場合、観光条件とユーザの嗜好を調整してもらう。その後、再度観光ルートを生成し、提示する。これをユーザが満足する観光ルートが提示されるまで繰り返す。

4 多様性のある観光ルート生成手法

本研究では、訪問する観光スポットとそれらを訪問する順序を決定することにより、観光ルートを自動生成する。これを総当たりで行う場合、観光スポット数が増えるほど、爆発的に計算量が増加する。この問題を解決するために、倉田の研究では遺伝的アルゴリズムを用いて観光ルートを生成している[1]。本研究では、筆者らが、観光スポットと移動経路に対する嗜好を考慮した観光ルートを自動生成している研究[7]と同様に、遺伝的アルゴリズムの一種である NSGA-III[8]を用いて観光ルートを生成する。観光ルートを自動生成する手順は以下の通りである。

- (1) 観光スポットをランダムに一つずつ抽出し、観光時間の制約を超えない場合は、抽出した観光スポットを観光ルートの末尾に加える。これを全ての観光スポットが抽出されるまで繰り返し、観光ルートを生成する。
- (2) (1)を事前に設定した生存個体数に達するまで繰り返し、初期の観光ルート群(生存個体群)を生成する。
- (3) 生存個体群からランダムに選択した個体を用いて交叉と突然変異を行い、新たな観光ルートを生成し、生存個体群に追加する。
- (4) 各生存個体を、各カテゴリに対する適合度(以降、カテゴリ適合度と呼ぶ)と、観光スポット間の類似度を用いて評価する。観光スポット間の類似度については5章、カテゴリ適合度の算出方法については、6章にて詳しく述べる。
- (5) reference pointと呼ばれる、評価値空間上に均等に配置した点ごとに、生存個体のカテゴリ適合度を用いて観光ルートを抽出し、次世代の観光ルート群とする。
- (6) (3)から(5)を繰り返すことにより、観光ルートを生成する。

5 観光スポット間の類似度算出方法

本研究では、人の感覚に近い類似度を算出する必要がある。人の感覚に近づけるためには、人が観光スポットに対して感じる特徴や観光した際に感じる魅力に基づいて類似度を算出することが有効である。そこで、観光情報サイトに記述されている観光スポットの説明文間の類似度を観光スポット間の類似度として用いる。一般に観光スポットの説明文には、各観光スポットについて簡潔に記述されている説明文(概要文)と、より詳細に観光スポットの特徴や魅力などが記述されている説明文(詳細文)が存在する。例として、函館市の公式観光情報サイトである「はこぶら」[9]に掲載されている観光スポットの説明文の一例を図2に示す。本研究では、図2のような概要文と詳細文を用いて類似度を算出する。観光スポットの説明文には、自治体が作成している信頼性が高い観光情報サイトに記載されている説明文を用いる。本研究では、「はこぶら」に掲載されている観光スポットのうち、「観光スポット(見る)」に分類されているものを対象とし、2021年6月に収集した205件の観光スポットの説明文を用いる。観光スポット間の類似度の算出には、

1 : https://nlp.ist.i.kyoto-u.ac.jp/index.php?ku_bert_japanese



図 2 「はこぶら」に掲載されている観光スポット情報の例

TF-IDF, Doc2Vec, BERT の 3 つの手法を使用し、算出した類似度を比較することにより適切な類似度算出方法を選択する。5.1 節から 5.3 節では、各手法を用いた類似度の算出方法について述べる。5.4 節では、各手法を用いて算出した類似度の比較結果について述べる。

5.1 TF-IDF を用いた類似度の算出

観光スポットの説明文の中で特徴的な単語が類似していれば、観光スポット同士にも類似性があると考えられる。そのため、各観光スポットの説明文の中で特徴的な単語に着目可能な TF-IDF を用いて類似度を算出する。算出に使用する説明文には、観光スポットのより詳細な部分まで考慮した類似度を算出できる詳細文を用いる。まず、説明文を形態素解析し、名詞、動詞、形容詞のみを抽出する。抽出した各単語に対して TF-IDF 値を算出する。BoW の各成分を、算出した TF-IDF 値に置き換えたベクトルを観光スポットのベクトルとする。観光スポットのベクトル間のコサイン類似度を観光スポット間の類似度とする。

5.2 Doc2Vec を用いた類似度の算出

Doc2Vec は文章からベクトル表現を獲得するための手法であり、獲得したベクトル表現同士の類似度を算出することにより、観光スポット間の類似度を算出する。算出時に使用する説明文には、5.1 節と同様に詳細文を用いる。まず、説明文を形態素解析した単語列を用いて Doc2Vec のモデルを生成する。モデル生成時のベクトルの次元数、エポック数、学習時に使用する単語の最小出現回数は、それぞれ複数のパラメータを設定して学習を行い、最も評価の高いモデルを使用する。今回のモデルを生成した際に使用した各パラメータの値を表 1 に示す。本研究では、算出した類似度が高い観光スポットの組み合わせが、「はこぶら」上で同一のカテゴリであるかを基準として算出した F 値が最も高かった、次元数が 150、エポック数が 70、最小出現回数が 2 のモデルを使用する。

表 1 Doc2Vec のモデル生成時に使用した各パラメータ

次元数	エポック数	最小出現回数
50, 100, 150,	10, 20, 30, 40, 50,	
200, 250, 300,	60, 70, 80, 90, 100	
350, 400	110, 120, 130, 140, 150	1, 2

表 2 「はこぶら」の下位カテゴリに該当する観光スポットの総数

カテゴリ名	観光スポット数	
	追加前	追加後
景色	47	47
坂	19	38
歴史的建造物	45	45
史跡・碑	41	41
教会・修道院	7	46
神社・寺	22	42
博物館	24	46

表 3 BERT のモデルをファインチューニングする際の各パラメータ

	パラメータ
最大トークン数	512
バッチサイズ	32
学習率	1×10^{-5}

5.3 BERT を用いた類似度の算出

BERT は、大規模コーパスを用いて作成した事前学習済みモデルを、対象とするタスクに合わせてファインチューニングすることにより、様々な言語処理タスクにおいて高い精度を達成している手法 [6] である。本研究では、BERT のモデルを用いて各説明文に含まれる各単語の分散表現を平均したベクトルを観光スポットのベクトルとして獲得する。単語の分散表現は、BERT モデルの隠れ層の最終層から獲得する。獲得した観光スポットのベクトル間のコサイン類似度を観光スポット間の類似度として算出する。

本研究では、事前学習済みモデルとして東北大大学が公開している、日本語 Wikipedia を用いて学習したモデル²を用いる。また、ファインチューニング用のタスクには、入力した説明文がどのカテゴリに分類された観光スポットのものであるかを推定するタスクを用いる。学習用の説明文には、観光スポットの概要文を使用する。しかし、表 2 の追加前の観光スポット数に示す通り、ファインチューニングに使用するカテゴリごとに観光スポット数に偏りがあるため、観光スポット数が多いものに重点を置いた学習が行われる可能性がある。そこで、観光スポット数が 40 未満のカテゴリに対して、「Yahoo! トラベル」[10] から収集した説明文を追加した。追加後の各カテゴリごとの観光スポット数は、表 2 に示す通りである。ファインチューニング時のパラメータは表 3 に示す通りであり、最適化関数には Adam を用いた。

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

表 4 各手法を用いて算出した mAP

類似度算出手法	mAP
TF-IDF	0.449
Doc2Vec	0.558
事前学習済み BERT モデル	0.558
ファインチューニング済み BERT モデル	0.788

5.4 類似度算出結果の比較

最も適切な類似度が算出できる手法を選択するために、5.1 節から 5.3 節の手法で算出した類似度を用いてモデルを評価した。同一のカテゴリに分類された観光スポットは類似していると仮定し、モデルの評価基準には、類似度が高い観光スポット同士が同一のカテゴリに属しているかを用いる。具体的な評価方法として、mean Average Precision (mAP) を用いることにより、各観光スポットごとに同一のカテゴリに分類された観光スポットとの類似度が高いほど、評価を高くする。

5.1 節から 5.3 節までの 3 つの類似度算出手法に、5.3 節の手法のファインチューニング済み BERT モデルを事前学習済み BERT モデルを用いて類似度を算出した場合を加えた結果を表 4 に示す。評価の結果、mAP の値は TF-IDF が 0.449 と最も低く、ファインチューニング済み BERT モデルが 0.788 と最も高かった。TF-IDF については、固有名詞の影響が大きく、歴史上の人物や近隣の観光スポットの名称が大きくなり過ぎてしまうこと、類似した意味を持つ単語が全く別の単語として扱われてしまうこと、見ることに関連した観光スポットのみを抽出しているため、それに関連した単語の影響が小さくなり過ぎてしまうことなどが原因で、適切な類似度が算出できなかったと考えられる。ファインチューニング後のモデルについては、カテゴリ分けのタスクを用いたファインチューニングによって、観光との関連が強い単語の影響が大きくなり、そうではない単語の影響は小さくなることにより、適切な観光スポット間の類似度を算出できたと考えられる。

6 観光ルートがユーザに与える満足度の算出方法

本研究では、ユーザの嗜好に合った多様性のある観光ルートを生成する。そのため、ユーザの嗜好に合った観光ルートであるかを評価するために、観光ルートがユーザに与える満足度を算出する必要がある。観光ルートの満足度を算出するために、まず各観光スポットに対してカテゴリごとの分類スコアを算出する。分類スコアは、観光スポットが各カテゴリとの関連性の高さを表す値である。カテゴリには、「はこぶら」の「スポット情報」の上位カテゴリである「観光スポット(見る)」に含まれる「景色」や「歴史的建造物」などの 7 つの下位カテゴリを用いる。算出した分類スコアに対して、観光ルートに含まれる観光スポットとの類似度に応じて補正をかける。補正後の分類スコアの総和を観光ルートのカテゴリ適合度とする。カテゴリ適合度は、各カテゴリにどの程度適した観光ルートであるかを表す値である。算出した観光ルートのカテゴリ適合度とユーザの嗜好を用いてユーザに与える満足度を算出する。6.1 節では、

観光スポットに対する分類スコアの算出方法について述べる。

6.2 節では、6.1 節で算出した分類スコアと観光スポット間の類似度を用いて、観光ルートがユーザに与える満足度を算出する。

6.1 各カテゴリに対する各観光スポットの分類スコア

5.3 節で作成したファインチューニング済み BERT モデルを使用して、各カテゴリに対する各観光スポットの分類スコアを算出する。あるカテゴリに対する分類スコアは、そのカテゴリとの関連性の高さを表す値である。分類スコアは、詳細文をファインチューニング済み BERT モデルに入力することにより獲得する。その後、カテゴリごとに分類スコアを、最小値を 0、最大値をカテゴリごとの最も大きな分類スコアとして正規化する。

6.2 観光スポット間の類似度を考慮した観光ルートの満足度算出方法

ユーザが好む多様性の度合い応じて、観光スポットの多様性を考慮した観光ルートのカテゴリ適合度を算出する。そのため、6.1 節で観光スポットごとに算出した分類スコアに対して、ユーザが好む多様性の度合いと観光ルートに含まれる他の観光スポットとの類似性に応じて補正をかける。その際、1 つの観光スポットで多様な観光体験が可能な場合、他の多くの観光スポットとの類似度が高くなってしまう可能性がある。そのような場合、ユーザにとって満足度が高い観光スポットであっても、補正が大きくなることにより観光ルートから取り除かれてしまう可能性がある。これを防ぐために、2 つの観光スポット間の類似度を算出した際に、ユーザに与える満足度が低い観光スポットにのみ補正をかける。

まず、観光スポット y を個別に訪問した際にユーザに与える満足度(以降、基礎満足度と呼ぶ)を算出する。基礎満足度 S_y の算出には、倉田の手法[1]を参考に、観光スポット y のカテゴリ j に対する分類スコアを B_{yj} 、ユーザの嗜好を U_j として、式(1)を用いる。

$$S_y = \sum_{j=1}^M B_{yj} U_j \quad (1)$$

算出した基礎満足度 S_y を基に、観光スポット y に対して、同一の観光ルートに含まれる、 S_y よりも基礎満足度の高い観光スポット y'_i を抽出する。観光スポット y_i に対して抽出した観光スポット群を Y' とし、観光スポット y と y'_i の類似度を $L_{y'_i}$ としたとき、式(2)を用いて類似度による減少補正值 L_y を算出する。

$$L_y = \begin{cases} \prod_{i=1}^{Y'} \min\left(1, \frac{(1-y'_i)}{(1-\alpha)}\right) & \alpha < 1 \text{ のとき} \\ 1 & \alpha = 1 \text{ のとき} \end{cases} \quad (2)$$

α はユーザが観光ルートに含まれる観光スポットにどの程度多様性があることを好むかを表す数値であり、0 から 1 の範囲で、ユーザが多様性を好むほど値を小さくする。減少補正值 L_y は、基礎満足度が最も高い観光スポットの場合、その観光スポットよりも基礎満足度が高い観光スポットは存在しないため、必ず 1 となる。基礎満足度が最も低い観光スポットの場合は、他の

全ての観光スポットとの類似度を用いて算出する。観光ルート y に対して算出した減少補正値 L_y と分類スコア B_{yj} を用いて、カテゴリ j に対するカテゴリ適合度 C_j を式(3)を用いて算出する。

$$C_j = \sum B_{yj} L_y \quad (3)$$

算出した C_j を式(1)の B_{yj} と置き換えて算出した S_y を、観光ルートがユーザに与える満足度とする。

7 観光ルートの推薦方法

本研究では、4章の手法で生成した観光ルート群から、ユーザに合った観光ルートを複数提示する。提示する観光スポットの選択手順は以下の通りである。

(1) 生成された複数の観光ルート群から、過去に推薦した観光ルートと観光スポットの組み合わせが同じ観光ルートを削除する。

(2) 残った観光ルート群を6章の手法を用いて評価し、最も満足度の高い観光ルートを1つ目の推薦観光ルートとして抽出する。

(3) (2)で選択されなかった観光ルート群から、抽出した観光ルートに対して、異なる観光スポットが最低でも1つ以上含まれる観光ルートの中から、最も満足度が高い観光ルートを選択する。

(4) 選択した観光ルートの満足度が、1つ目の推薦観光ルートに対して70%以上であれば、次点の推薦観光ルートとして抽出する。

(5) 選択した観光ルートの満足度が、70%未満であるか、推薦観光ルートが一定の数になるまで(3),(4)を繰り返す。

8 観光ルートの生成実験

7章まで述べた手法を用いて、多様性のある観光ルートの生成が可能であるかを検証するために実験を行った。実験では、ユーザが好む観光スポットの多様性の度合いが異なる条件下で観光ルートを生成し、生成された中で6章の手法を用いて算出した満足度が最も高い観光ルートを比較した。

8.1 生成する観光ルートの設定

生成する観光ルートは、出発・終着地点を「函館駅」、観光時間を180分とした。使用する観光スポットは、「はこぶら」の「観光スポット(見る)」に分類されている観光スポット205件を用いた。各観光スポットにおける滞在時間は一律で20分とし、観光スポット間の移動は徒歩のみとした。想定ユーザの各カテゴリに対する嗜好は表5に示す通りである。想定ユーザの嗜好は、景色を強く、歴史的建造物を少し、それ以外のカテゴリを僅かに好む嗜好とした。

8.2 推薦結果

ユーザが多様性を好む度合いに応じて生成された観光ルートごとの観光スポットを訪問する順序を表6に示す。多様性を好まない場合($\alpha=0$)は、景色としての分類スコアが高い、とも

表5 想定するユーザの各評価項目に対する嗜好の度合い

評価項目	嗜好の度合い
景色	5
坂	1
歴史的建造物	2
史跡・碑	1
教会	1
神社仏閣	1
博物館	1

表6 ユーザが好む多様性の度合いに応じて生成された観光ルートごとの観光スポットを訪問する順序

ユーザが好む多様性の度合い	観光スポットを訪問する順序
多様性を好まない($\alpha=1$)	ともえ大橋, 旧桟橋(東浜桟橋), 遺愛幼稚園, 元町配水場, 函館海産商同業協同組合事務所
半分程度 多様性を好む ($\alpha=0.5$)	明治天皇御上陸記念碑, 函館市北方民族資料館, 遺愛幼稚園, 函館海産商同業協同組合事務所, ともえ大橋
多様性を好む ($\alpha=0$)	函館山, 太刀川家住宅店舗

え大橋、旧桟橋(東浜桟橋)、元町配水場のような観光スポットが最も多く、歴史的建造物としての分類スコアが高い、遺愛幼稚園、函館海産商同業協同組合事務所が含まれた観光ルートが推薦された。多様性を半分程度好む場合($\alpha=0.5$)、景色としての分類スコアが高い観光スポットが減り、史跡・碑と博物館の観光スポットが新たに追加された。多様性を好む場合($\alpha=1$)、訪問可能な範囲で景色、歴史的建造物としての分類スコアがそれぞれ最も高い観光スポットのみの観光ルートが推薦された。

8.3 考察

多様性を考慮することにより、ユーザが好む景色や歴史的建造物の分類スコアが高い観光スポットが多く含まれつつ、他のカテゴリに対する分類スコアが高い観光スポットも含まれた多様性のある観光ルートが生成できた。多様性を半分程度好む場合、ユーザが最も好む景色の分類スコアが高い観光スポットの数が、歴史的建造物の分類スコアが高い観光スポットよりも少なくなっている。これは、訪問可能な景色の分類スコアが高い観光スポットが少なく、それらの観光スポット間の類似性が高いことが要因となっていると考えられる。多様性を好む場合の観光ルートは、ユーザが好むカテゴリに対する分類スコアが高い観光スポットが優先的に選ばれる傾向があると考えられる。また、多様性を好まない場合や半分程度多様性を好む場合の観光ルートに比べて観光スポット数が少ない原因是、函館山にあると考えられる。函館山は、ユーザが好む景色に対する分類スコアが高いため、優先的に選択されやすい。また、移動手段を徒歩のみとしており、函館駅から離れた位置にある函館山が他の観光スポットよりも優先的に選択されやすい。その結果、他の観光スポットを訪問するための時間が短くなり、観光スポット数が少なくなっていると考えられる。

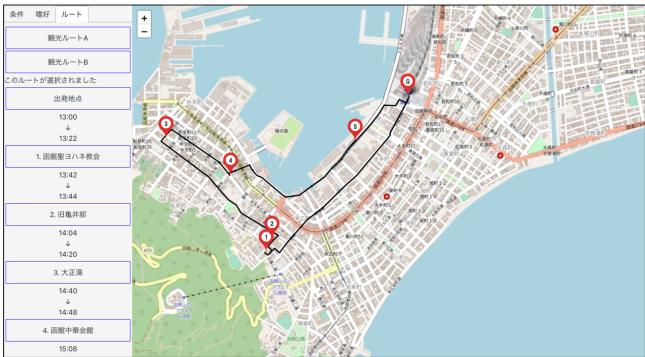


図3 観光ルートの詳細を提示する画面

9 提案する観光ルート推薦システムを用いた評価実験

情報系の学生 11 名を対象に、本研究で提案する観光ルート推薦システムの評価実験を行った。実験では、観光ルートを生成する際に多様性を考慮することが有用であるかと、推薦された観光ルートに含まれる観光スポットが適切であるかを検証した。

9.1 評価実験の方法

まず実験の設定として、下記の4つの条件を被験者に説明する。

- 函館に訪問するのは初めてとする。
 - 出発・終着地点は「函館駅」とする。
 - 観光の時間は 13 時から 16 時までとする。
 - 移動手段は徒歩のみとする。

被験者への説明が終了後、提案システムを実際に使用してもらう。まず、出発・終着地点の選択と観光時間の入力、観光スポットの嗜好と多様性を好む度合いの入力をそれぞれ行ってもらう。入力後、4章の観光ルート生成手法を用いて観光ルートを生成する。その際、ユーザが入力した観光スポットの嗜好と多様性を好む度合いを用いて2つの観光ルートを生成する。1つは、ユーザが入力した観光スポットの嗜好と多様性を好む度合いをそのまま使用して生成した観光ルート（以降、多様性を考慮した観光ルートと呼ぶ）である。もう1つは、ユーザが入力した多様性を好む度合いを0に変更した観光ルート（以降、多様性を考慮しない観光ルートと呼ぶ）である。2つの観光ルートを生成した後、それぞれの観光ルートに含まれる観光スポットとその訪問順を確認してもらう（図3）。被験者の好みに合った観光ルートが生成されていた場合は、どちらの観光ルートが好みであるかを選択してもらう。被験者の好みに合った観光ルートが生成されなかった場合は、観光スポットの嗜好と多様性を好む度合いを被験者の好みに合った観光ルートが生成されるまで調整してもらった後に、どちらの観光ルートが好みであるかを選択してもらう。観光ルートを選択後、2つの観光ルートと提案システムに関するアンケートに回答してもらう。

9.2 実験結果

評価実験の結果、多様性を考慮した観光ルートが選択された

割合は 45 % であった。多様性を考慮した観光ルートを選択した理由として、「景色もよく、歴史的な建造物もあるルートだから」や「神社、教会など嗜好に合う観光地が含まれていたため」などが記述されていた。また、多様性を考慮した観光ルートを選択しなかった理由として、「3, 4 時間の移動時間が長くて効率がわるそう」や「函館山を登るルートは現実的ではないと思ったから」などが記述されていた。

多様性を考慮した観光ルートと多様性を考慮しない観光ルートそれぞれから、被験者好みに最も合った観光スポットとして選択された観光スポットが、ユーザが好むカテゴリの観光スポットである割合は 73 % であった。ユーザが好むカテゴリは、観光スポットの嗜好を入力した際にスライダーが 70 % 以上の値に設定されたカテゴリとした。被験者好みではない観光スポットとして選択された観光スポットが、ユーザが好むカテゴリの観光スポットである割合は 33 % であった。ユーザが好むカテゴリの観光スポットが選択された理由として、「これまでのスポットで十分感がある」、「名前だけ表示されているのでどのような観光地か分からぬいため」、「観光地かどうか分からぬい、嗜好に合っていない」という回答があった。

システムを使用した感想として、良かった点では「時間などを考慮して自動的にルート生成される」、「観光スポットの詳細情報が見られる」などが記述されていた。悪かった点としては、「ルート生成にかかる時間が長すぎる」、「経路に無駄な部分がある」、「既に訪問した観光スポットの一部が決まっている場合に対応できない」などが記述されていた。

9.3 考察

多様性を考慮した観光ルートと考慮しない観光ルートでは、多様性を考慮しない観光ルートが僅かに多く選ばれている。しかし、多様性を考慮した観光ルートを選択した理由として、ユーザの嗜好に合った観光ルートが推薦されたという旨的回答がされている場合もあるため、ある程度の有用性があると考えられる。多様性を考慮した観光ルートがあまり選ばれなかった原因として、観光ルートの移動経路に問題があったことが考えられる。8章の評価実験から、多様性を好む度合いが高くなるほど、ユーザが好むカテゴリに対する分類スコアが高い観光スポットを含む観光ルートが生成されることが分かっている。今回の実験では被験者の多くが景色を好む傾向にあり、多様性を好む度合いを高くしたユーザに対して8.3節でも述べた「函館山」を含む観光スポット数の少ない観光ルートが推薦されたことがあった(図4)。このような移動経路に問題がある場合に多様性を考慮しない観光ルートが選ばれてしまっており、多様性を考慮したかに関係しない部分が影響してしまっている可能性がある。

観光ルートに含まれる観光スポットの中で、好みに一番合った観光スポットがユーザが好むカテゴリの観光スポットである割合が7割を超えており、ある程度ユーザの好みに合った観光スポットが推薦できたと考えられる。しかし、被験者好みではない観光スポットとして選択された観光スポットがユーザが好むカテゴリの観光スポットからも選ばれていた。この原因と



図 4 選択されなかった多様性を考慮した観光ルートの例

して、観光スポットの評価が適切にできていない可能性が考えられる。本研究では、5.3 節で説明した手順でファインチューニングしたモデルを使用して、観光スポットの各カテゴリに対する分類スコアを算出している。このモデルでは、シングルラベルに分類するタスクを用いており、各カテゴリの分類スコアが、他のカテゴリの影響を受けてしまう。そのため、マルチラベルに分類するタスクを用いてファインチューニングすることによって、結果が改善される可能性がある。

10 おわりに

本稿では、ユーザにとってより満足度が高いと考えられる多様性のある観光ルートを推薦することを目的として、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を提案した。本システムでは、ユーザの観光スポットに対する嗜好と観光スポット間の類似度を考慮して観光ルートを推薦する。観光ルートの生成には、遺伝的アルゴリズムの一種である NSGA-III を使用する。観光スポット間の類似度算出には、最も評価の高かったファインチューニング済み BERT モデルを用いる。観光ルートがユーザに与える満足度は、ユーザの嗜好と、観光スポット間の類似度に応じて補正をかけた観光スポットの分類スコアを用いて算出する。推薦する観光ルートは、観光ルートが与える満足度と観光スポットの重複を考慮して選択する。被験者を用いた実験の結果、多様性を考慮した観光ルートよりも多様性を考慮しない観光ルートが僅かに多く選ばれたが、アンケートの結果から多様性を考慮することがある程度有用であることが示唆される。また、観光ルートに含まれる観光スポットについては、ユーザの嗜好に適した観光スポットを選択できている場合が多いが、シングルタスクを用いた BERT のモデルを用いたことにより、ユーザの嗜好にあまり適していない観光スポットが推薦される場合があった。

今後の展望として、観光スポットの評価方法や観光ルート生成アルゴリズムを改善する必要がある。観光スポットの評価方法については、マルチラベルに対応した分類タスクを用いてファインチューニングしたモデルを構築し、観光スポットの分類スコアを算出することが考えられる。これにより、各カテゴリに対して、他のカテゴリからの影響を受けない評価が可能になると考えられる。観光ルート生成アルゴリズムについては、

遺伝的アルゴリズムのパラメータの調整や並列分散処理の実装を行うことにより、より高速な観光ルート生成を実現する。パラメータの調整については、様々な条件で観光ルートを生成した結果を比較することにより、生存個体数、世代数、交叉率、突然変異率に、適切な値を設定することが考えられる。これらの改善の他に、バスなどの公共交通機関を利用した観光ルートを生成可能にすることや、各観光スポットにおける滞在時間に適切な値を設定すること、訪問する観光スポットをユーザが指定可能にすることなどが考えられる。滞在時間については、全ての観光スポットで同一の値を用いてる。そのため、観光に時間がかかる観光スポットが含まれていた場合は、想定した観光時間を超過してしまう場合がある。観光にあまり時間がかかる観光スポットが含まれていた場合は、追加で観光スポットを訪問できる可能性がある。これらの問題を解決するために、観光スポットのジャンルに応じて滞在時間を決定したり、ユーザが自由に滞在時間を設定できるようにすることが考えられる。観光スポットをユーザが指定可能することについては、倉田[1]の研究と同様に生成された観光ルートの各観光スポットに対して、「行きたい」、「行きたくない」といった要望を追加できるようにしてることや、ユーザの嗜好を基に主要な観光スポットを推薦し、その中からユーザに訪問したい観光スポットを複数決定してもらった後に観光ルートを生成することが考えられる。

文 献

- [1] 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、Vol.5, pp.159-165 (2012).
- [2] 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).
- [3] 川崎 雄太：組合せ価値を考慮した商品推薦システムの構築、大学院研究年報 理工学研究科編、Vol.44 (2014).
- [4] 庄司裕子：好適な組合せを実現するための汎用モデルに関する研究、日本感性工学会論文誌、Vol.14, No.4, pp.511-516 (2015).
- [5] 橋口友哉、山本岳洋、藤田澄男、大島裕明：CQA コンテンツからの状況が類似する悩みの検索、人工知能学会論文誌、Vol.36, No.1, p.WI2-B-1-13 (2021).
- [6] Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, NAACL-HLT 2019, pp.4171-4186 (2019).
- [7] 武信雄平、奥野拓：観光スポットと移動経路に対する嗜好を考慮した観光ルート推薦システムの構築、第 83 回全国大会講演論文集 2021, No.1, p.335-336 (2021).
- [8] Deb, K., Jain, H. : An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints, IEEE Transactions on Evolutionary Computation, Vol.18, pp.577-601(2014).
- [9] 函館市観光部：函館市公式観光情報サイトはこぶら、入手先 <<https://www.hakobura.jp/>> (参照 2022-05-20).
- [10] Yahoo Japan: 宿泊予約・ホテル予約なら Yahoo! トラベル、入手先 <<https://travel.yahoo.co.jp/>> (参照 2022-11-14).