

分散表現を用いたパーソナライズド観光ルート推薦

中川 智也[†] 新妻 弘崇[†] 太田 学[†]

[†] 岡山大学大学院自然科学研究科 〒700-8530 岡山県岡山市北区津島中 3-1-1

E-mail: †{nakagawa, niitsuma, ohta}@de.cs.okayama-u.ac.jp

あらまし 新井らは Twitter に投稿された観光体験情報に着目し、ツイートから観光スポットがよく訪れられている時間帯、タイムラインから共に訪れられやすい観光スポット、ツイートと手掛かり語から求めた観光スポットの観光目的の3種類のスコアを提案し、そのスコアを用いて観光ルート推薦を行った。しかし、手掛かり語の収集は Wikipedia と Weblio 類語辞典に依存するため、解説ページ等が存在しない語はスコアに反映できない。そこで本稿では、Yahoo!知恵袋の質問文の Paragraph Vector を用いて算出する、類似度スコアと NN カテゴリスコアを提案する。実験では、新井らのスコアのみによる推薦観光ルートと、新井らのスコアに提案したスコアを加えた推薦観光ルートをそれぞれ生成し、どの観光ルートを回りたいかアンケート調査を行った。アンケート調査の結果、NN カテゴリスコアの有用性を示すことはできなかったが、類似度スコアの有用性は示された。

キーワード 観光体験, doc2vec, SNS

1. はじめに

近年、スマートフォン等のモバイル通信機器の普及率が9割を超え [1]、インターネット上の情報にアクセスできる人口が大幅に増加している。さらに、SNS 等の様々なネットサービスが登場し、人々が簡単にインターネット上で意見や価値観を共有できるようになった。こうした多くの人々に利用されているインターネット上の情報の一つとして、本稿では観光体験情報に注目する。

新井ら [2] は旅行者の好みに合った観光ルートを推薦するために、Twitter^(注1) に投稿されたツイートの中から観光に関するものを収集し分析した。ツイートが投稿された時間から観光に適した時間帯を算出し、ツイート文に含まれる単語から観光スポットを訪れた人の観光目的を推定した。さらに、この分析結果を利用して、新井らは推薦観光ルートを生成した。新井らはツイートの分析の中で、旅行者の観光スポットの好みを観光ルートに反映するために、どのような目的で観光スポットを訪れられやすいか分析した。その結果、観光目的を「食事」、「景観」、「土産」、「行動」の4カテゴリと定義し、これら4つについて日本語版 Wikipedia^(注2) (以下、Wikipedia) と Weblio 類語辞典^(注3) から各カテゴリの手掛かり語を収集した。そして、各観光スポットについてのツイートの中で、いずれかの手掛かり語を含むツイート数に対する、各カテゴリの手掛かり語を含むツイート数の割合をその観光スポットの観光目的のスコアとした。この方法では Wikipedia の解説ページ等が存在しなければ手掛かり語が収集できず、まだ解説ページのない新語は観光目的を求めるのに利用できない。また、Weblio 類語辞典にも同様の問題があった。

そこで、本稿では Wikipedia や Weblio 類語辞典に依存せず、旅行者の観光スポットの好みを観光ルート推薦に反映する手法を提案する。提案手法では、一般の人が気軽に投稿でき、新語も含まれている Yahoo!知恵袋^(注4) の質問文を用いる。観光スポット名を含む Yahoo!知恵袋の質問文から Paragraph Vector の実装の1つである doc2vec でそれぞれの質問文の分散表現を得て、それらのベクトルを使って各観光スポットの特徴ベクトルを生成する。そして、その特徴ベクトルを使って観光スポット間の類似度やどのような目的で訪れられやすいかを定量化し、旅行者の観光スポットの好みを観光ルート推薦に反映する。

本稿では、まず2節で関連研究について述べる。次に、3節で新井らが提案したスコアと推薦観光ルート生成手法を説明し、4節で本稿の提案手法を説明する。そして、5節で評価実験について述べ、6節でまとめる。

2. 関連研究

2.1 観光推薦

島田らは [3] は、ユーザがお気に入りの観光地を入力すると、それと類似した観光スポットを推薦する手法を提案した。具体的には、彼らは各観光スポットについて、Yahoo!知恵袋でよく共起するキーワード、Yahoo!知恵袋の質問の時系列分布、Yahoo!知恵袋の質問のカテゴリ情報、各観光スポット周辺にある施設、各観光スポット周辺の地図画像の色ヒストグラムを求め、特徴ベクトルを生成した。そして、ユーザが入力した観光スポットの特徴ベクトルとコサイン類似度が高い観光スポットを推薦した。

藤坂ら [4] は位置情報付きツイートをを用いて観光情報の抽出を試みた。彼らは、K-means 法により分割した日本の各領域に対して、ツイート数、Twitter ユーザ数、Twitter ユーザの移

(注1) : Twitter, <http://twitter.com/>

(注2) : Wikipedia, <https://ja.wikipedia.org/wiki/>

(注3) : Weblio 類語辞典, <https://thesaurus.weblio.jp/>

(注4) : Yahoo!知恵袋, <https://chiebukuro.yahoo.co.jp/>

動量からノーマルパターンを規定し、ある時間におけるツイート数、Twitter ユーザ数、Twitter ユーザの移動量をそれぞれのノーマルパターンと比較することで地域イベントが行われている領域を検出した。

石野ら [5] は、ANPI NLP^(注5)で提供される震災情報に関わるツイートを利用して、被災時の避難経路を自動抽出する手法を提案した。石野らは機械学習を用いて、移動元、移動先、移動手段のタグを東日本大震災に関連するツイートに自動付与することによって、被災者の行動経路を抽出した。

郡ら [6] は行動計画の立案支援として、ブログからユーザの旅行時の代表的なルートとその文脈を抽出し、地図上にマッピングして提示するシステムを提案した。郡らの手法では、ブログ内に出現する各地名に対して、旅行者が実際にその場所を訪れたかどうかを文脈から判定し、訪れたと判定した場合はその地名をルート要素とする。その後、ルート要素に対して順序付けを行い、地名の系列パターンを抽出した。

Lee ら [7] は電気自動車の充電待ち時間を目的関数とし、それを最小化する観光ルートを遺伝的アルゴリズムを用いて推薦した。Lee らの手法は、旅行者の観光ルートの中の代表的な観光スポットを出発点として観光ルートの突然変異や進化を行い、より良いルートを探した。

三富ら [8] は Twitter と Flickr の投稿の位置情報を比較することで、FreeWiFi が無い観光スポットを発見する手法を提案した。ツイートを発信するときはデータ通信が必要なため、ツイートに位置情報が付与されている場所はそこに FreeWiFi がある。Flickr に投稿されている写真に位置情報を付与するときはデータ通信は特に必要なく、観光スポットで撮られる写真に位置情報が付与されやすいことから、位置情報が付与されている写真が多い場所は観光スポットとする。以上より、位置情報が付与されている写真が多い観光スポットの中からツイートの少ない観光スポットを探して、FreeWiFi が無い観光スポットを発見した。

2.2 文書の分散表現

Mikolov ら [9] [10] は、ニューラルネットワークで単語の特徴を低次元な数値ベクトルで表現する word2vec を提案した。word2vec は、文章中の単語を周辺の単語の前後関係から予測する問題をニューラルネットワークで学習し、そのニューラルネットワークの中間層を単語の特徴ベクトルとして出力する。また word2vec の生成する単語ベクトルはベクトル同士の演算ができ、その演算が単語間の意味の関係を表わせる。例えば以下のような計算式が成り立つ。

$$\text{king} - \text{man} + \text{woman} = \text{queen} \quad (1)$$

式 (1) は king から man を引き、woman を足すと queen のベクトルとなるということを示す。また、似たような意味を持つ単語をそれぞれベクトル化した場合、互いに似たようなベクトルが生成される。

Paragraph Vector は word2vec のベクトル化の対象を単語で

はなく、文章に拡張したものである [11]。似たような文章からは似たようなベクトルが生成され、文章の特徴をベクトルで表すことができる。例えば、渋谷の待ち合わせ場所として有名なハチ公像のレビュー文書をベクトル化したものは、同様に有名な渋谷のモヤイ像や上野公園の西郷隆盛像のレビュー文書をベクトル化したものとのコサイン類似が高くなるという実験結果が報告されている [12]。モヤイ像はハチ公像と同じ渋谷にある有名な待ち合わせ場所で、渋谷の地名や待ち合わせに関する単語がレビュー中に出現するため類似度が高くなる。西郷隆盛像は所在地は渋谷ではないが、こちらも待ち合わせ場所として知られており、待ち合わせに関する単語がレビュー中に出現するために類似度が高くなる。

3. 新井らのスコアと推薦観光ルート生成

本節では新井らが提案したスコア [2] と以前我々が提案した新井らが提案したスコアを用いる推薦観光ルート生成手法 [14] について説明する。

新井らが提案したスコアは、観光スポットについて呟いているツイートを収集し、それを用いて各観光スポットの観光スポットとしての良さをスコア化した。そして、推薦観光ルート生成にはそのスコアを利用して観光ルートを生成する。具体的には以下のようなものである。

3.1 観光スポットについてのツイート収集

最初に、推薦に利用するツイートの収集 [2] について説明する。各観光スポットについて、Twitter API [13] を用いて観光スポット名をクエリとしてツイートを収集する。さらに、位置情報共有サービスの URL や、「～なう」を含み、実際にその観光スポットで投稿されたツイートに関しては、そのユーザのタイムラインも併せて収集する。

3.2 観光スポットのスコア

次に、[2] における観光スポットの観光スポットとしての良さを示すスコアの算出方法を説明する。新井らは各観光スポットについて、時間帯スコア・共起スコア・カテゴリスコアの3つのスコアを定義した。時間帯スコアは、観光スポット名を含む全ツイート数に対する1日を8等分した3時間ごとの各時間帯におけるツイート数の割合で、その観光スポットを訪れるのに良い時間帯を表すスコアである。共起スコアは、とある観光スポットについてツイートしたユーザのタイムラインに出てくる他の観光スポットの共起確率で、共に訪れられやすい観光スポットを表すスコアである。カテゴリスコアは、手掛かり語を元にツイートを「食事」、「景観」、「土産」、「行動」の4カテゴリに分類したときの各カテゴリの割合である。この手掛かり語とは Wikipedia と Weblio 類語辞典から収集したものである。例えば、「食事」カテゴリの手掛かり語には Wikipedia の「料理」カテゴリ下の語、「景観」カテゴリの手掛かり語には Weblio 類語辞典で「綺麗」と類似する語等が含まれる。カテゴリスコアは、旅行者がどういう目的で観光スポットを訪れているかを表すスコアである。これら3つのスコアを観光ルート推薦に使用する。また、これら3つのスコアを合計したものを本稿では TCC スコアと以下では呼ぶこととする。

(注5) : ANPI NLP, http://trans-aid.jp/ANPI_NLP/index.php/

3.3 スコアを用いたルート生成

求めたスコアを用いた推薦観光ルートの生成手法 [14] について図 1 を使って説明する。これは新井らのルート生成手法を基に、我々が改良したものである。

観光ルートの被推薦者は訪れたい観光スポット 1 つと観光の出発地、出発時刻、到着地、到着時刻を入力する。入力した観光スポットを選択スポットと呼び、選択スポットのカテゴリスコアが最大のカテゴリを被推薦者が興味を持つ興味カテゴリと呼ぶ。

まず、STEP1 で出発地→選択スポット→到着地の初期ルートを作成する。

次に、STEP2 で出発地→選択スポット、選択スポット→到着地の各部分経路の間に訪れることができる観光スポットの中から、3.2 節で説明したスコアの和が最大となる観光スポットを探す。図 1 では S_A と S_B が各部分経路でスコアの和が最大の観光スポットである。

次に、STEP3 で S_A と S_B でスコアの和が大きい方をルートに加え、出発地→ S_A →選択スポット→到着地というルートになる。

次に STEP4 で再び、各部分経路の間に訪れることができる観光スポットの中から、スコアの和が最大となる観光スポット S_C, S_D, S_E を探し、STEP5 でさらにその中でスコアの和が最大となる S_E をルートに加える。

加えられる観光スポットがなくなるまで、STEP4 と STEP5 を交互に繰り返し、ルートを生成する。そして、STEP6 でこのルートを保存しておく。

このとき、局所最適解に陥る可能性があるため、STEP7 でルートから観光スポットを無作為に削除し、STEP8 で再び、STEP4 と同じように各部分経路でスコアの和が最大となる観光スポットを探し、その中でスコアの和が最大となる観光スポットをルートに加える。そして、加えられる観光スポットがなくなるまで繰り返してルートを生成する。STEP9 でこのルートを保存し、互いに異なる 100 ルートが保存されるまで、推薦観光スポットの削除とルートの生成を繰り返す。そして、保存したルートのうち、観光スポットのスコアの合計が最も高いルートを被推薦者に推薦する。このランダム探索アルゴリズムは遺伝的アルゴリズムの 1 つと言える。

4. 提案手法

本節では、手掛かり語に依存せずに旅行者の観光スポットの好みを推薦観光ルートに反映するスコアを提案する。そのため、本稿では Yahoo!知恵袋に投稿された質問文を doc2vec [11] でベクトル化し、各観光スポットの特徴ベクトルを算出する。そして、観光スポットの特徴ベクトルの類似度を算出した観光スポットの類似度スコアと、観光スポットの特徴ベクトルをニューラルネットワークによって「食事」、「景観」、「土産」、「行動」の 4 つに分類した出力の NN カテゴリスコアを提案する。

4.1 質問文収集

Yahoo!知恵袋は一般の人々が生活の中で疑問に思ったことをその質問のカテゴリを指定して投稿するサービスである。カ

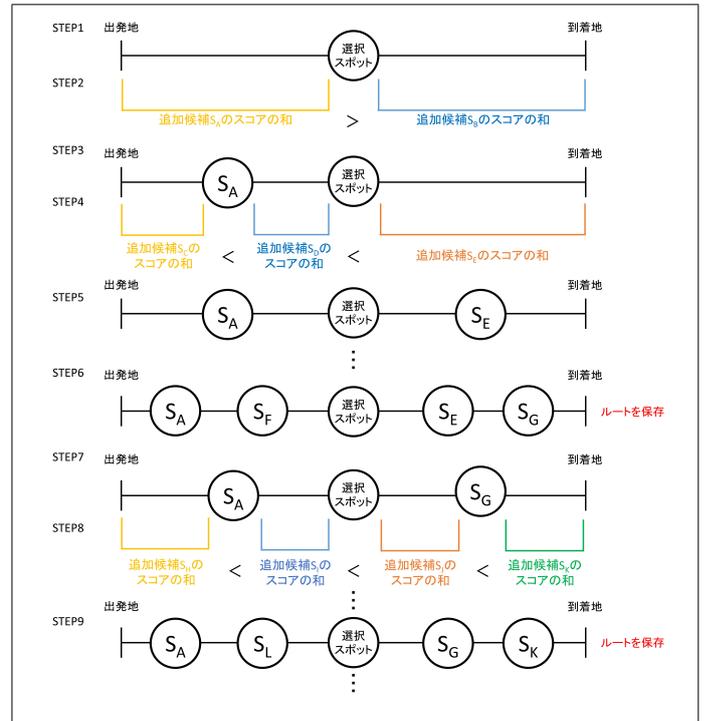


図 1 ルート生成アルゴリズム

テゴリの中には旅行に関するカテゴリもあり、観光スポットの特色や観光目的が質問文に記載されている。そこで、本稿では「地域、旅行、おでかけ」カテゴリ下の「国内」カテゴリで投稿された質問文を利用する。Yahoo!知恵袋の質問検索 API [15] を用いて、各観光スポット名をクエリとして「国内」カテゴリ下の質問文を収集する。このとき、観光スポットの中には観光スポット名が曖昧なものが存在する。例えば、岡山県内のレストラン「小樽」で検索すると、北海道の「小樽」に関する質問もヒットする。そこで、[16] で提案した、doc2vec でベクトル化した質問を SVM を用いて分類する手法でこの曖昧性を解消する。

4.2 観光スポットの特徴ベクトル

次に、4.1 節の手法で収集した質問文から各観光スポットの特徴ベクトルを求める。具体的な方法は以下のようである。

4.2.1 質問文のベクトル化

まず、収集した全観光スポットについての質問を Sen [17] を用いて、形態素解析する。次に、形態素解析した質問文を Paragraph Vector の実装の 1 つである doc2vec を用いてベクトル化する。似た文章からは似たベクトルが生成される性質があり、例えば城や寺社仏閣に関する文章には歴史に関する単語が共通して現れるため、似たベクトルが生成される。逆に、似たベクトルの質問文に含まれる単語は似ているといえる。

4.2.2 観光スポットの特徴ベクトル

質問文のベクトルから観光スポットの特徴を表す特徴ベクトルを生成する。観光スポットについての質問は、その観光スポットで何が楽しめるか尋ねるもの以外にも、地理的な情報やその観光スポットの歴史等様々な内容を含む。つまり、質問ベクトルには、その観光スポットの特徴が含まれている。よって、本稿では、各観光スポットについての質問ベクトルの重心ベク

トルを求め、それをその観光スポットの特徴ベクトルとする。例えば、歴史ある観光スポットは、歴史に関する話題を含む質問が多い。よって、特徴ベクトルにはこのような特徴が反映されていると考えられる。

4.3 提案スコア

本節では、本稿で提案する類似度スコアと NN カテゴリスコアについて説明する。

4.3.1 観光スポットの類似度スコア

類似度スコアは観光スポットの特徴がどの程度似ているかを示すスコアである。特徴ベクトルには各観光スポットの観光目的や地理的な情報等が含まれているため、特徴ベクトルが似ている観光スポットは、それらが似ているといえる。そこで、観光スポットの特徴ベクトルの類似度を求め、それを観光スポットの類似度とする。観光客のお気に入りの観光スポットとの類似度が高い観光スポットも、その観光客は気に入る可能性が高い。

本稿では観光スポットの特徴ベクトルの類似度にはコサイン類似度を用いる。特徴ベクトルの類似度を観光スポットの全ての組み合わせのペアで求める。そのため、スポット数を n とすると nC_2 の組み合わせの類似度を算出することになる。

4.3.2 NN カテゴリスコア

NN カテゴリスコアは手掛かり語を用いずに、ニューラルネットワークにより、各観光スポットの観光目的を 3.3 節で説明した「食事」、「景観」、「土産」、「行動」の 4 カテゴリに分類したときの出力である。本稿で利用するニューラルネットワークのモデルの概略図を図 2 に示す。 \mathbf{x} は観光スポットの特徴ベクトル、 \mathbf{y} はニューラルネットワークの出力ベクトルである。入力層のユニット数を観光スポットの特徴ベクトルの次元数の 100、出力層のユニット数をカテゴリの数の 4、中間層のユニット数を 4、活性化関数をソフトマックスとした。

まず、各観光スポットについて観光客の主な観光目的を「食事」、「景観」、「土産」、「行動」の 4 カテゴリに人手で分類する。次に、各観光スポットの特徴ベクトルと人手での分類結果でニューラルネットワークを学習する。そして、このニューラルネットワークに特徴ベクトルを入力し、出力を得る。このニューラルネットワークの各出力ユニットの出力が NN カテゴリスコアの各カテゴリのスコアに相当する。例えば、「食事」、「景観」、「土産」、「行動」の各カテゴリに相当するユニットの出力がそれぞれ 0.4, 0.3, 0.2, 0.1 のとき、NN カテゴリスコアの「食事」、「景観」、「土産」、「行動」の各カテゴリのスコアが 0.4, 0.3, 0.2, 0.1 となる。

4.4 観光ルート推薦への導入

4.3 節で説明した類似度スコアと NN カテゴリスコアを 3.3 節の推薦観光ルート生成で利用するスコアに加える。具体的には、推薦観光ルート生成時に選択スポット以外の観光スポットを追加するとき、訪問時間の時間帯スコア、選択スポットとの共起スコア、興味カテゴリのカテゴリスコアの 3 スコアに、本稿で提案する類似度スコアと NN カテゴリスコアを加えたスコアの和が最大の観光スポットをルートに追加する。

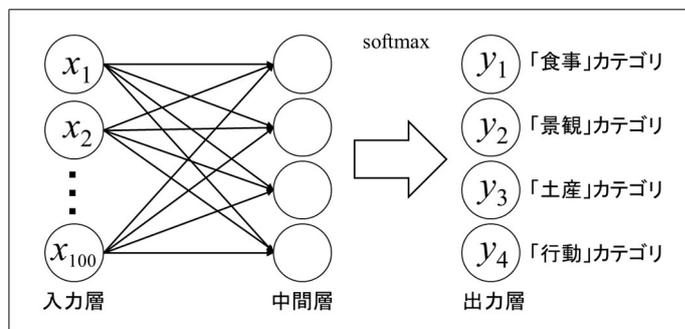


図 2 ニューラルネットワークのモデルの概略図

表 1 岡山県内の代表的な観光スポットの例とその質問数

観光スポット	質問数
岡山後楽園	54
岡山城	280
美観地区	622
大原美術館	118
湯原温泉	176

5. 評価実験

本稿では、岡山県内の 161 の観光スポットを用いて、実験を行う。まず、161 の観光スポットについて特徴ベクトルから類似度スコア、NN カテゴリスコアを求める。そして、3.3 節のルート生成に用いる TCC スコアに類似度スコアと NN カテゴリスコアを追加して推薦観光ルートを生成する。追加した 2 つのスコアの有用性を評価するために、

- (1) TCC スコアのみを用いた推薦ルート
- (2) TCC スコアに類似度スコアを加えた場合の推薦ルート
- (3) TCC スコアに NN カテゴリスコアを加えた場合の推薦ルート
- (4) TCC スコアに類似度スコアと NN カテゴリスコアを加えた場合の推薦ルート

の 4 種類のルートを生成し、どのルートを観光してみたいか、被験者にアンケート調査を行う。

5.1 Yahoo!知恵袋の質問文収集

岡山県内の 161 の観光スポットについて、観光スポット名を含む Yahoo!知恵袋の質問を収集した。岡山県内のいくつかの有名な観光スポットについて収集した質問数を表 1 に示す。

5.2 類似度スコア

岡山県内の代表的な観光スポットである岡山市の岡山後楽園と類似した観光スポットとその類似度を表 2 に示す。岡山後楽園は日本三名園のうちの一つで、江戸時代初期に造営された庭園であり、広い芝生地や季節で移り変わる植物等、景観を楽しむことができる観光スポットである。

これとの類似度が最も高い岡山城は岡山後楽園に隣接する安土桃山時代に建てられた城である。こちらも歴史のある観光スポットで、城の造りや天守閣から岡山後楽園や岡山の街並みが望め、景観を楽しむ観光スポットと言える。岡山市の観光情報サイト [18] では岡山後楽園と岡山城の両方を訪れるコースをモデルコースとして提案しており、岡山後楽園を訪れる観光客に

表 2 岡山後楽園に類似した観光スポット上位 5 件

観光スポット	類似度
岡山城	0.92
大原美術館	0.89
美観地区	0.85
井倉洞	0.77
吉備津神社	0.75

表 3 ニューラルネットワークによる観光スポットのカテゴリ分類

		正解カテゴリ分類			
		食事	景観	土産	行動
出力最大 カテゴリ	食事	1	6	0	1
	景観	6	34	4	15
	土産	0	1	1	1
	行動	0	2	0	10
計		7	43	5	27

岡山城を薦めるのは適切と言える。

一方、大原美術館と美観地区は岡山市に隣接する倉敷市に位置しており、歴史的価値のある美術品や江戸時代の街並みを楽しむことができる。これらも観光情報サイト [18] のモデルコースで岡山後楽園と共に訪れるよう提案されている。

5.3 NN カテゴリスコア

人手による各観光スポットの観光目的の正解カテゴリへの分類結果と、その観光スポットの特徴ベクトルを入力したときのニューラルネットワークによる出力のうち最大の値をとるカテゴリ (出力最大カテゴリ) を表 3 に示す。学習に用いたデータ数が多かった正解カテゴリが「景観」カテゴリである観光スポットの中で、正解カテゴリと出力最大カテゴリが一致したのは 43 件中 34 件だった。一方、学習に用いたデータ数が少なかった正解カテゴリが「食事」カテゴリである観光スポットの中で、正解カテゴリと出力最大カテゴリが一致したのは 7 件中 1 件、正解カテゴリが「土産」カテゴリである観光スポットの中で、正解カテゴリと出力最大カテゴリが一致したのは 5 件中 1 件だった。学習に用いたデータ数が少ないカテゴリの観光スポットに関しては、人手による観光目的の分類とニューラルネットワークによる出力では大きな違いがみられた。

5.4 観光ルート推薦

本節では、以下に示す出発地や選択スポット等の観光ルートの被推薦者による入力 2 通りで、3.3 節で説明したルート生成で利用するスコアを (1) から (4) のように変えて推薦観光ルートを生成する。

(I) 10:00 に岡山駅を出発、19:00 に岡山駅に到着、選択スポットが岡山後楽園

(II) 9:00 に岡山大学を出発、21:00 に岡山駅に到着、選択スポットが湯原温泉

入力 I の条件でルート生成で利用するスコアを変えて生成された推薦観光ルートと Google Maps API [19] を用いて出力したそのルートマップを表 4 から表 7、図 3 から図 6 に、II の条件でルート生成で利用するスコアを変えて生成された推薦観光ルートと Google Maps API を用いて出力したそのルートマッ

表 4 (I-1) TCC スコアのみを用いた推薦ルート

マーカー	スポット名	到着時刻	出発時刻
A	岡山駅		10:00
B	岡山県立美術館	10:04	11:24
C	岡山後楽園	11:28	12:32
D	吉備津彦神社	12:56	13:37
E	安養院	14:02	15:32
F	渋川海岸	16:11	17:11
G	石山公園	18:04	18:53
H	岡山駅	19:00	

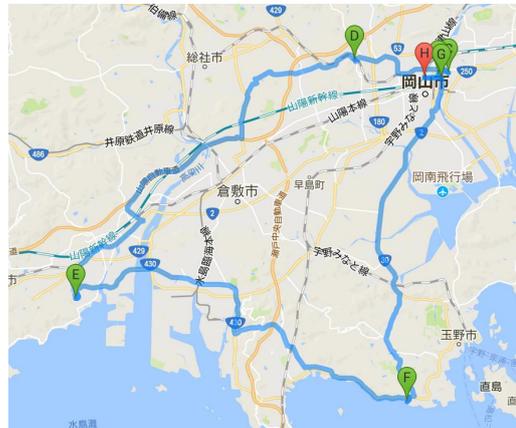


図 3 (I-1) TCC スコアのみを用いた推薦ルートマップ

表 5 (I-2) TCC スコアに類似度スコアを加えた場合の推薦ルート

マーカー	スポット名	到着時刻	出発時刻
A	岡山駅		10:00
B	夢二郷土美術館	10:06	12:25
C	岡山後楽園	12:26	13:30
D	岡山城	13:37	15:12
E	吉備津神社	15:43	17:16
F	石山公園	17:45	18:53
G	岡山駅	19:00	



図 4 (I-2) TCC スコアに類似度スコアを加えた場合の推薦ルートマップ

プを表 8 から表 11、図 7 から図 10 に示す。

(I) 10:00 に岡山駅を出発、19:00 に岡山駅に到着、選択スポットが岡山後楽園

(I-1) TCC スコアのみを用いた推薦ルート

(I-2) TCC スコアに類似度スコアを加えた場合の推薦ルート

(I-3) TCC スコアに NN カテゴリスコアを加えた場合の推薦ルート

表 6 (I-3) TCC スコアに NN カテゴリスコアを加えた場合の推薦ルート

マーカー	スポット名	到着時刻	出発時刻
A	岡山駅		10:00
B	招き猫美術館	10:21	11:07
C	岡山後楽園	11:28	12:32
D	吉備津彦神社	12:56	13:37
E	常盤公園	14:05	15:05
F	御崎神社	15:15	16:15
G	備中国分寺	16:21	18:25
H	岡山駅	19:00	

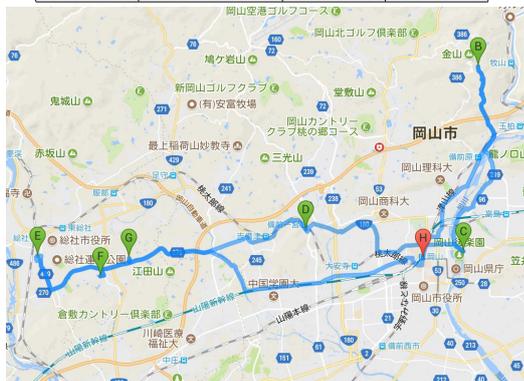


図 5 (I-3) TCC スコアに NN カテゴリスコアを加えた場合の推薦ルートマップ

表 7 (I-4) TCC スコアに類似度スコアと NN カテゴリスコアを加えた場合の推薦ルート

マーカー	スポット名	到着時刻	出発時刻
A	岡山駅		10:00
B	岡山城	10:08	11:43
C	岡山後楽園	11:50	12:54
D	吉備津彦神社	13:18	13:59
E	御崎神社	14:19	15:19
F	備中国分寺	15:25	17:01
G	龍泉寺	17:19	18:29
H	岡山駅	19:00	



図 6 (I-4) TCC スコアに NN カテゴリスコアを加えた場合の推薦ルートマップ

(I-4) TCC スコアに類似度スコアと NN カテゴリスコアを加えた場合の推薦ルート

(II) 9:00 に岡山大学を出発, 21:00 に岡山駅に到着, 選択スポットが湯原温泉

表 8 (II-1) TCC スコアのみを用いた推薦ルート

マーカー	スポット名	到着時刻	出発時刻
A	岡山駅		9:00
B	渋川動物公園	9:55	10:55
C	みかどや	11:22	12:46
D	渋川公園	13:09	14:16
E	渋川海岸	14:19	15:19
F	湯原温泉	17:01	18:42
G	石山公園	20:10	20:53
H	岡山駅	21:00	



図 7 (II-1) TCC スコアのみを用いた推薦ルートマップ

(II-1) TCC スコアのみを用いた推薦ルート

(II-2) TCC スコアに類似度スコアを加えた場合の推薦ルート

(II-3) TCC スコアに NN カテゴリスコアを加えた場合の推薦ルート

(II-4) TCC スコアに類似度スコアと NN カテゴリスコアを加えた場合の推薦ルート

5.4.1 被験者によるアンケート

生成した推薦ルートについて, どのルートを観光してみたいか被験者にアンケート調査を行った. 被験者は 2018 年 1 月の時点で, 岡山大学工学部情報系学科の 4 年生 4 名と同大学院自然科学研究科の博士前期課程 1 年生 2 名, 2 年生 2 名の計 8 名である.

入力 I で生成した観光ルート (I-1) から (I-4) を提示し, 実際に回りたいルート順にランク付けしてもらった. 同様に, 入力 II で生成した観光ルートについてもランク付けしてもらった.

それぞれの入力で用いるスコアを変えて生成した各推薦観光ルートの平均順位を表 12 に示す. 平均順位が最も高いのは入力 I は TCC スコアに類似度スコアを加えた推薦観光ルート, 入力 II は TCC スコアに類似度スコアと NN カテゴリスコアを加えた推薦観光ルートだった. 入力 II で 2 番目に平均順位が高いのは TCC スコアに類似度スコアを加えた推薦観光ルートで, TCC スコアに類似度スコアを加えた推薦観光ルートは入力 I, II のどちらにおいても高い評価を得た.

一方, TCC スコアに NN カテゴリスコアを加えた推薦観光

表 9 (II-2) TCC スコアに類似度スコアを加えた場合の推薦ルート

マーカー	スポット名	到着時刻	出発時刻
A	岡山駅		9:00
B	岡山後楽園	9:12	10:16
C	吉備津彦神社	10:40	11:21
D	白糸の滝	12:20	13:16
E	満奇洞	14:14	15:14
F	湯原温泉	16:05	17:46
G	湯郷温泉	18:39	19:43
H	岡山駅	21:00	



図 8 (II-2) TCC スコアに類似度スコアを加えた場合の推薦ルートマップ

表 12 各ルートの平均順位

	(1)	(2)	(3)	(4)
入力 I	2.50	1.38	3.25	2.88
入力 II	2.88	1.88	3.63	1.63

ルートはどちらの入力においても、平均順位が最下位で、有用性を示すことができなかった。

6. まとめ

本稿では、旅行者の観光スポットの好みを推薦観光ルートに反映するために、観光スポットのスコアとして、観光スポットの類似度スコアと NN カテゴリスコアを提案した。さらに、これらのスコア算出のために、各観光スポットに関する Yahoo!知恵袋の質問ベクトルから各観光スポットの特徴を表す特徴ベクトルを算出した。

類似度スコアは、この特徴ベクトルによる類似度であり、Yahoo!知恵袋の質問文において、他の観光スポットとどの程度似ているかを示すスコアである。本研究では、特徴ベクトルが算出できた観光スポットの全ての組み合わせで特徴ベクトルのコサイン類似度を算出し、類似度スコアとした。

NN カテゴリスコアは観光スポットの特徴ベクトルとその観光スポットの観光目的「食事」、「景観」、「土産」、「行動」の人手による分類結果を学習データとして、ニューラルネットワークを学習し、各観光スポットの特徴ベクトルを入力したときの

表 10 (II-3) TCC スコアに NN カテゴリスコアを加えた場合の推薦ルート

マーカー	スポット名	到着時刻	出発時刻
A	岡山駅		9:00
B	招き猫美術館	9:17	10:06
C	法輪寺	10:30	11:20
D	夢二郷土美術館	11:28	13:47
E	浄土寺	14:01	14:35
F	吉備津彦神社	15:07	15:48
G	湯原温泉	17:04	19:34
H	岡山駅	21:00	



図 9 (II-3) TCC スコアに NN カテゴリスコアを加えた場合の推薦ルートマップ

各ユニットの出力を「食事」、「景観」、「土産」、「行動」カテゴリのスコアとしたものである。

これらのスコアの有用性を確かめるために、新井らが提案した TCC スコアのみで推薦された推薦観光ルート、TCC スコアにこれらのスコアを加えた推薦観光ルートを生成し、どのルートを観光してみたいか被験者にアンケート調査を行った。

アンケートの結果、TCC スコアに類似度スコアを加えた推薦観光ルートは TCC スコアのみで推薦された観光ルートよりも評価が高く、類似度スコアの有用性が示された。しかし、TCC スコアに NN カテゴリスコアを加えた推薦観光ルートは TCC スコアのみで推薦された観光ルートよりも評価が低かった。学習に用いた特徴ベクトルにカテゴリごとの偏りがあったことがその原因と考えられる。

今後の課題としては、Yahoo!知恵袋の質問文以外の文章も NN カテゴリスコアのニューラルネットワークの学習に用いることで、学習用データの偏りをなくし、NN カテゴリスコアを改善することが挙げられる。

文献

- [1] 総務省, “平成 29 年度情報通信白書”, <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/nc262110.html>
- [2] 新井晃平, 新妻弘崇, 太田学, “Twitter を利用した観光ルート推薦の一手法”, 第 7 回データ工学と情報マネジメントに関するフォーラム, G7-6, pp. 1-8, 2015.

表 11 (II-4) TCC スコアに類似度スコアと NN カテゴリスコアを加えた場合の推薦ルート

マーカー	スポット名	到着時刻	出発時刻
A	岡山大学		9:00
B	招き猫美術館	9:17	9:59
C	満奇洞	11:13	12:13
D	神庭の滝	13:05	14:06
E	湯原温泉	14:36	16:17
F	龍泉寺	17:28	18:28
G	岡山後楽園	19:02	20:06
H	石山公園	20:11	20:53
I	岡山駅	21:00	



図 10 (II-4) TCC スコアに NN カテゴリスコアを加えた場合の推薦ルートマップ

International Conference on Machine Learning, pp. 1188-1196, 2014.

- [12] 工学院大学インタラクティブメディア研究室, “Python と gensim で doc2vec を使う”, <https://kitayamalab.wordpress.com/2016/11/14/python-と-gensim-で-doc2vec-を使う/>
- [13] Twitter Developer, <https://developer.twitter.com/>
- [14] 中川智也, 新妻弘崇, 太田学, “マイクロブログを利用した観光ルート推薦における移動効率の改善”, 第 8 回データ工学と情報マネジメントに関するフォーラム, 2016.
- [15] Yahoo!デベロッパーネットワーク, <http://developer.yahoo.co.jp/webapi/chiebukuro/v1/questionsearch.html>
- [16] 中川智也, 新妻弘崇, 太田学, “Yahoo!知恵袋を利用した施設名の曖昧性解消手法の提案”, 第 9 回データ工学と情報マネジメントに関するフォーラム, C4-4, 2017.
- [17] 形態素解析システム Sen, <https://www.mlab.im.dendai.ac.jp/~yamada/ir/MorphologicalAnalyzer/Sen.html>
- [18] 岡山県観光連盟, “おかやま観光ネット”, <https://okayama-kanko.net/sightseeing/index.php>
- [19] Google Maps API, <https://developers.google.com/maps/?hl=ja>

- [3] 嶋田和孝, 上原尚, 遠藤勉, “集合知に基づく観光地推薦システムの構築”, 観光と情報: 観光情報学会誌 10(1), pp.113-124, 2014.
- [4] 藤坂達也, 李龍, 角谷和俊, “地域イベント発見のためのジオタグ付マイクロブログを用いたノーマルパターン検出手法”, 平成 22 年度情報処理学会関西支部大会, Vol. 2010, E-04, 2010.
- [5] 石野亜耶, 小田原周平, 難波英嗣, 竹澤寿幸, “Twitter からの被災時の行動経路の自動抽出および可視化”, 言語処理学会 第 18 回年次大会, A4-6, pp. 907-910, 2012.
- [6] 郡宏志, 服部峻, 手塚太郎, 田島敬史, 田中克己, “ブログからのビジターの代表的な行動経路とそのコンテキストの抽出”, 電子情報通信学会技術研究報告, Vol. 106, No. 149, pp. 29-34, 2006.
- [7] Lee, J., Park, G., “A tour recommendation service for electric vehicles based on a hybrid orienteering model”, Proceedings of the 28th Annual ACM Symposium on Applied Computing, SAC '13, pp. 1652-1654, 2013.
- [8] 三富恵佑, 遠藤雅樹, 江原遥, 廣田雅春, 横山昌平, 石川博, “外国人にアクセシブルな FreeWiFi がない観光スポットの発見”, 第 8 回データ工学と情報マネジメントに関するフォーラム, H4-5, pp. 1-6, 2016.
- [9] Mikolov, T., Sutskever, I., Chen, K., Corrado G. and Dean, J., “Distributed representations of words and phrases and their compositionality”, Advances in Neural Information Processing Systems 26, pp. 3111-3119, 2013.
- [10] Mikolov, T., Chen, K., Corrado G. and Dean, J. “Efficient estimation of word representations in vector space”, arXiv preprint arXiv:1301.3781, pp. 1-12, 2013.
- [11] Quoc V, Le and Tomas Mikolov, “Distributed Representations of Sentences and Documents”, Proceedings of the 31st