

ユーザレビューの分散表現を用いた 主観的特徴の意味演算による観光スポット検索システム

吉田 朋史[†] 北山 大輔^{††} 中島 伸介^{†††} 角谷 和俊^{††††}

[†] 工学院大学大学院工学研究科 〒163-8677 東京都新宿区西新宿 1-24-2

^{††} 工学院大学情報学部 〒163-8677 東京都新宿区西新宿 1-24-2

^{†††} 京都産業大学コンピュータ理工学部 〒603-8555 京都市北区上賀茂本山

^{††††} 関西学院大学総合政策学部・社会情報学研究センター 〒669-1337 兵庫県三田市学園 2-1

E-mail: [†]em15020@ns.kogakuin.ac.jp, ^{††}kitayama@cc.kogakuin.ac.jp, ^{†††}nakajima@cse.kyoto-su.ac.jp,

^{††††}sumiya@kwansei.ac.jp

あらまし 観光スポットが持つ様々な特徴の中で、タグなどのメタデータでは推測困難な「夜景が美しい」「待ち合わせ場所に向いている」などの主観的特徴を考慮した観光スポット検索システムを提案する。提案システムでは、文書の分散表現獲得手法である Paragraph Vector を用いて観光スポットに対して投稿されたユーザレビューから生成した各スポットの特徴ベクトルと、ユーザが入力したキーワードをレビュー中に含むスポットの特徴ベクトルを用いて生成した入力キーワードの特徴ベクトルによる意味演算を行うことで、ユーザの入力キーワードが表す主観的特徴を考慮した観光スポット検索を行う。

キーワード 観光スポット検索, 分散表現, ユーザレビュー

1. はじめに

観光地に関する情報を収集する上で、Web 上に存在する情報の重要性は年々高まっている。Sanderson ら [1] や Gravano ら [2] の研究結果によれば、検索エンジンを用いた Web 検索の 20% 近くは特定の場所に関係する問い合わせであると言われており、Web 上で観光情報を検索するユーザが求める情報を適切に提示することは情報検索・情報推薦システムにとって重要なタスクであると考えられる。

このようなユーザからの需要に応じて、近年じゃらん^(注1) や TripAdvisor^(注2) などの観光情報サイトが普及し、これらのサイトには観光スポット（以下スポットと表す）についての情報やユーザレビューが数多く投稿されている。一般的な観光情報サイトでは、スポットの特徴を表現するメタデータとしてスポットの所在地を表す地域タグ（都道府県など）や種類を表すカテゴリタグ（「寺社仏閣」など）が付与されており、ユーザが目的に合うスポットを検索する際は、キーワードによる検索とこれらのタグによる絞り込みを組み合わせて使うことが多いと考えられる。

しかし、スポットが持つ特徴には所在地やカテゴリから推測することが困難な特徴も存在し、これらの特徴を考慮したスポット検索を行う際には地域タグやカテゴリタグに基づく検索が有効に働かない事が考えられる。例えば、大阪府大阪市に所在する観光スポット「梅田 BIGMAN」は大型液晶ディスプレイを備えたイベントスペースであり、じゃらんでは地域タグ

「大阪市」、カテゴリタグ「観光施設・名所巡り」などが付与されている。一方、「梅田 BIGMAN」は JR 大阪駅近郊の待ち合わせ場所としても著名なスポットであるが、「大阪市」や「観光施設・名所巡り」などの付与されたタグから「待ち合わせ場所」という特徴を類推することは困難であり、投稿されたレビューなどを実際に精読しない限りは「待ち合わせ場所」という特徴をユーザが知ることは困難であると考えられる。

このような「待ち合わせ場所」や「景色が綺麗」などの特徴を、本研究ではスポットの主観的特徴と定義する。主観的特徴は観光スポットが持つ機能や観光地としての役割に関連する特徴であり、ユーザがその観光スポットを訪問する目的や、訪問することで得られる感情・印象などを表す特徴であると考えられる。また、これらの主観的特徴はスポットの所在地や所属カテゴリなどの客観的かつ普遍的な情報から類推することが困難であり、ユーザレビューや SNS への投稿などユーザの主観的な印象や個人的な体験についての記述が多く含まれるテキストに含まれやすいと考えられる。

このような主観的特徴を考慮した観光スポットの検索手法として、我々はある観光スポットと主観的特徴が類似する他の地域・カテゴリに属する観光スポットの検索手法を提案した [3]。文献 [3] で提案した手法では、文書を表現する特徴ベクトル（分散表現）をニューラルネットワークによって学習する Paragraph Vector モデル [4] を用いて、ある観光スポットに対して投稿されたユーザレビューを学習データとして各観光スポットやスポットに付与された地域タグ・カテゴリタグの特徴ベクトルを算出する。算出した観光スポットの特徴ベクトルと地域タグ・カテゴリタグの特徴ベクトルで加減算を行い、演算結果のベクトルと各スポットの特徴ベクトル間のコサイン類似

(注1): <http://www.jalan.net/kankou/>

(注2): <https://www.tripadvisor.jp/>

度を算出することで、他の地域・カテゴリにおいて入力スポットと主観的特徴が類似するスポットを検索する（このような演算方法は入力クエリの意味を考慮したベクトル演算と考えられるため、本研究ではこのような演算方法を意味演算と呼ぶ）。例えば、東京都内において「梅田 BIGMAN」と同様に待ち合わせ場所として著名な観光スポットを検索したい場合には、「梅田 BIGMAN- 大阪府 + 東京都」のようなベクトルの加減算を行い、演算結果のベクトルとコサイン類似度が高い観光スポットとして「八チ公の銅像」や「西郷隆盛像」のような東京都内において待ち合わせ場所として著名な観光スポットを抽出することを旨とした。

文献 [3] での実験結果から、Paragraph Vector モデルはスポットの主観的特徴を考慮した特徴ベクトルを算出できること、Paragraph Vector モデルで算出したベクトル同士の演算によって精度良く意味演算が行えることが確認された。一方、文献 [3] で提案した手法の問題点として以下のような点が確認された。

(1) 考慮したい主観的特徴を持つ何らかのスポットをユーザが事前に把握している必要がある

(2) 地域タグ・カテゴリタグが表す特徴以外を全て主観的特徴とみなしてしまう

(1) については、提案手法では入力スポットと他の地域・カテゴリにおいて主観的特徴が類似するスポットを検索するため、入力スポットの主観的特徴がどのような特徴であるかをユーザが把握している必要がある。例えば「待ち合わせ場所」という主観的特徴に基いてスポットの検索を行う場合には、「八チ公の銅像」や「梅田 BIGMAN」など待ち合わせ場所として著名なスポットをユーザが入力スポットとして選定する必要がある。このため、考慮したい主観的特徴は明確だがその主観的特徴を持つスポットの具体例が想起しにくい場合には、提案手法による検索が行いづらいことが想定される。

(2) については、提案手法は Paragraph Vector モデルで算出した観光スポットの特徴ベクトルに含まれる特徴の種類が「地域的特徴」「カテゴリ的特徴」「主観的特徴」の3種類であるという仮定に基づいている。しかし、実際にユーザレビュー中に含まれる記述にはこの3種類以外のスポットの主観的特徴への言及が含まれると考えられる。例えば、特定の季節に訪れやすいスポットでは「夏」「雪」「寒い」など季節に関する言及が含まれやすく、巨大な観光スポットに対しては「大きい」「広い」などスポット自体の大きさ・広さに関する言及が含まれやすいことが考えられる。Paragraph Vector モデルは文書中の各単語の出現確率に基づき文書の特徴ベクトルを学習するため、これらの主観的特徴として適切でないと考えられる特徴の類似によってもベクトル間の類似度が上昇する可能性が考えられる。

そこで、本研究ではこのような問題点を改善することを目的として、Paragraph Vector モデルにより算出した各スポットの特徴ベクトルを用いて任意の主観的特徴の特徴ベクトルを動的に算出し、ある観光スポットと比較して対象の主観的特徴を強く持つ他の観光スポットを検索する手法を提案する。提案手法の具体的なアルゴリズムを以下に示す。

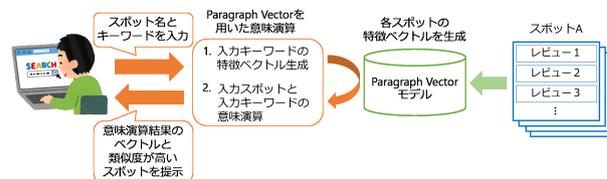


図 1 提案手法の概念図

(1) 任意の観光スポット名と考慮したい主観的特徴に関するキーワードをユーザが入力し、各スポットに対するレビュー文中での入力キーワードの単語頻度を算出

(2) 算出した単語頻度が一定の閾値以上のスポットの特徴ベクトルを抽出し、これらのベクトルの平均値などを用いて入力キーワードの特徴ベクトルを算出

(3) 入力された観光スポットの特徴ベクトルと算出したキーワードの特徴ベクトルの加算を行い、演算結果のベクトルと各スポットの特徴ベクトル間でのコサイン類似度に基づきユーザへの提示スポットを抽出

例えば、「ディズニーランド」と同種のテーマパークの中で、ジェットコースターなどのアトラクション要素が少なくよりキャラクターに特化したテーマパークを検索したい際は、レビュー文中での単語「アトラクション」の単語頻度に基づき抽出したスポットの特徴ベクトルを用いて「アトラクション」の特徴ベクトルを算出し、「ディズニーランド + アトラクション」のようなベクトル同士の加算を行うことで、ディズニーランドと比較してより「アトラクション」に特化したスポットを検索する。提案手法の概念図を図 1 に示す。

本研究では、スポットの特徴ベクトルを用いた主観的特徴の特徴ベクトル算出手法について、(1) 各ベクトルの相加平均 (2) ベクトルの各次元の最大値に対する割合に基づく重み付き平均 (3) 入力キーワードの単語頻度に基づく重み付き平均の3手法を提案し、各手法によるスポット検索の精度を実験により評価する。

本論文の構成は下記のとおりである。2 節では、関連研究との比較を通して本研究の位置付けについて述べる。3 節では、Paragraph Vector モデルを用いた観光スポットの特徴ベクトルの算出について述べる。4 節では、3 節で算出したスポットの特徴ベクトルを用いた主観的特徴に関するキーワードの特徴ベクトルの算出について述べる。5 節では提案システムによる観光スポットの検索例を、6 節ではまとめと今後の課題について述べる。

2. 関連研究

観光情報の検索・推薦に関する関連研究として、土田ら [5] は word2vec による地域を表す単語の加算による他の地域における類似スポットの発見手法を、石野ら [6] はガイドブックなどでは知ることが難しい観光スポットの詳細な情報を効率良くユーザに提示するための情報拡張システムを、伊達ら [7] はブログ記事から観光スポットの印象を表す単語を抽出し印象に基づく観光スポット推薦手法を提案している。また、あるドメインに

対応するコンテンツを別のドメインから発見するような、相対的検索やアナロジー検索に関する研究も存在する。Nakajimaら[8]は、あるドメインの空間を別のドメインに写像することで2つのドメインの相対的関係性を用いて検索する手法を提案している。また、Katoら[9]は2つのキーワードとその関係を表すキーワードに着目し、その関係キーワードをWeb検索結果から自動的に抽出する手法、および関係キーワードを用いる関係となる別のキーワードを発見する手法を提案している。本研究はこれらの先行研究で議論されていなかった観光スポットの主観的特徴に着目し、先行研究例の少なかった Paragraph Vector モデルの観光スポットへの適用を試みている点や、観光スポットの主観的特徴を表すベクトルを動的に生成し、ベクトルの加算による意味演算として観光スポットの相対的検索を試みる点でこれらの研究と異なる。

3. Paragraph Vector モデルによる観光スポットの特徴ベクトル算出

Paragraph Vector モデルは、Mikolov らによって考案された単語の特徴ベクトル学習手法である word2vec [10] を拡張し、Le ら [4] によって提案された文書の特徴ベクトル学習手法である。Paragraph Vector モデルでは「同じ文脈で出現する単語は類似した意味を持つ」というハリスの分布仮説 [11] に基づき、「ある文書中である単語列が与えられた時、次に出現する単語を予測する」というタスクをニューラルネットワークに学習させることで、文脈や単語の語順を考慮した文書の特徴ベクトルを生成することが出来る。Paragraph Vector モデルでは、生成した文書の特徴ベクトルを文書の分散表現 (Distributed Representation) と呼ぶ。

観光スポットへの Paragraph Vector モデルの適応手法として、我々は [3] において、ある観光スポットに対して投稿された全レビュー文を1文書とみなし各文書を Paragraph Vector モデルの学習データとして各スポットの特徴ベクトルを生成する手法を提案した。図2に示すように、「八チ公の銅像」の特徴ベクトルを生成する際には「八チ公の銅像」に対して投稿された全てのユーザーレビューを連結したものを「八チ公の銅像」に対応する1文書とみなし、Paragraph Vector モデルによって学習したこの文書の分散表現を「八チ公の銅像」の特徴ベクトルとする。表1は、国内の代表的な観光情報サイトであるじゃらんに掲載されている43,759件の観光スポットに対して投稿された1,481,831レビューを学習データに用いた Paragraph Vector モデルにおいて、「八チ公の銅像」の特徴ベクトルとコサイン類似度が高い特徴ベクトルを持つ観光スポット上位5件である。

「八チ公の銅像」は待ち合わせ場所として全国的に著名な観光スポットであり、これは「八チ公の銅像」に付与された「渋谷区」などの地域タグや「観光施設・名所巡り」などのカテゴリタグだけでは予測することが難しい「八チ公の銅像」の主観的特徴であると考えられる。表1より、「八チ公の銅像」の類似スポットとして「モヤイ像」「西郷隆盛像」「梅田 BIGMAN」などの待ち合わせ場所として同様に著名な観光スポットが抽出されている。これらのスポットはレビュー中に待ち合わせにつ

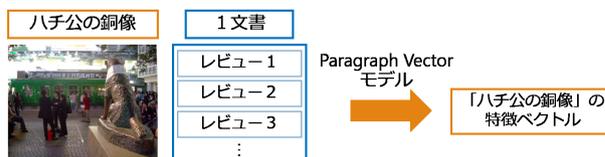


図2 Paragraph Vector モデルによる「八チ公の銅像」の特徴ベクトル生成例

表1 「八チ公の銅像」と類似度が高いスポット

スポット名	類似度	地域	カテゴリ
八チ公の銅像	-	渋谷区	観光施設・名所巡り
モヤイ像	0.7825	渋谷区	観光施設・名所巡り
八チ公ファミリーレリーフ	0.6197	渋谷区	その他
西郷隆盛像	0.5719	台東区	観光施設・名所巡り
梅田 BIGMAN	0.4992	大阪市	観光施設・名所巡り
大館駅前八チ公像	0.4675	大館市	観光施設・名所巡り

いての記述を多く含んでおり、Paragraph Vector モデルがこのようなレビュー中の記述に基づきスポットの主観的特徴を考慮した特徴ベクトルを生成できていると考えられる。

4. 主観的特徴に関する入力キーワードの特徴ベクトル算出

この節では、3節で算出した各スポットの特徴ベクトルを用いて、ユーザが入力した主観的特徴に関するキーワードの特徴ベクトルの算出について述べる。本研究では、投稿されたレビュー中での入力キーワードの単語頻度を考慮して観光スポットを抽出し、抽出したスポットの特徴ベクトルを用いてキーワードの特徴ベクトルを算出する。本研究では、式1の条件式を満たすあるスポット s の特徴ベクトル v_s を、入力キーワードの特徴ベクトル算出に用いるスポットの特徴ベクトルとして抽出する。

$$V_{kw} = \{v_s | tf(s, kw) \geq \theta_{tf}\} \quad (1)$$

ここで、 $tf(s, kw)$ はスポット s に投稿された全レビュー中での入力キーワード kw の単語頻度の合計値、 θ_{tf} は単語頻度に関する閾値、 V_{kw} は式1の条件を満たす特徴ベクトルの集合を表す。例えば、キーワード「夜景」の特徴ベクトルの算出には、投稿された全レビュー中での「夜景」の単語頻度の合計値が閾値 θ_{tf} 以上のスポットを用いる。

我々は文献 [3] において観光スポットに付与された地域タグやカテゴリタグの算出手法を比較検討し、対象のタグが付与されたスポットの特徴ベクトルを相加平均する手法が最も精度良く意味演算が行えることを実験的に確かめた。一方、相加平均による地域・カテゴリタグは意味演算に与える影響度が小さすぎるという欠点も実験結果から同様に明らかとなった。この結果に基づき、本研究では入力キーワードの特徴ベクトル v_{kw} の算出手法として、入力キーワードがレビュー中に含まれるスポットの特徴ベクトルの相加平均を基本とし、新たに提案する各次元の最大値や単語頻度などによる重み付き平均を含めた3つの手法をそれぞれ4.1節、4.2節、4.3節で提案する。算出した入力キーワードの特徴ベクトルと対象の観光スポットの特徴

ベクトルの加算を行うことで、対象の観光スポットと比較して入力キーワードが表す主観的特徴を強く持つ他の観光スポットを検索する。

4.1 相加平均による入力キーワードの特徴ベクトル

式 2 は、集合 V_{kw} に含まれる全てのベクトルの相加平均を用いた入力キーワードの特徴ベクトル算出式である。ここで、 $|V_{kw}|$ は集合 V_{kw} に含まれるスポット数を表す。

$$v_{kw} = \frac{1}{|V_{kw}|} \sum_{v_s \in V_{kw}} v_s \quad (2)$$

4.2 各次元の最大値に基づく重み付き平均による入力キーワードの特徴ベクトル

特徴ベクトル v_{s_i} の次元数を N 、 n 次元目の値を $d_{v_{s_i}n}$ 、集合 V_{kw} に属する全スポットの特徴ベクトルの n 次元目の値の集合を D_n 、 D_n の最大値を $\max D_n$ とした時、集合 V_{kw} に属する特徴ベクトル s_i の n 次元目の重み $w(s_i, n)$ と、 $w(s_i, n)$ を用いた重み付き平均で計算される v_{kw} の n 次元目の値 d_n は式 3、4 で表される。

$$w(s_i, n) = \left| \frac{d_{v_{s_i}n}}{\max D_n} \right| \quad (3)$$

$$d_n = \frac{\sum_{i=1}^{|V_{kw}|} w(s_i, n) d_{v_{s_i}n}}{\sum_{i=1}^{|V_{kw}|} w(s_i, n)} \quad (4)$$

式 4 で算出した v_n を用いて、 v_{kw} は式 5 で表される。

$$v_{kw} = (d_1 \quad d_2 \quad \dots \quad d_n \quad \dots \quad d_N)^T \quad (5)$$

4.3 単語頻度に基づく重み付き平均による入力キーワードの特徴ベクトル

式 6 は、スポット s_i に投稿された全レビュー（総単語数 $|s_i|$ ）中での入力キーワード kw の単語頻度に基づいたスポット s_i の重み w_{s_i} の算出式である。

$$w_{s_i} = \frac{tf(s_i, kw)}{\log_{10} |s_i|} \quad (6)$$

算出した重みを用いた集合 V_{kw} 内の特徴ベクトルの重み付き平均として、入力キーワードの特徴ベクトルを式 7 で算出する。

$$v_{kw} = \frac{\sum_{i=1}^{|V_{kw}|} w_{s_i} v_{s_i}}{\sum_{i=1}^{|V_{kw}|} w_{s_i}} \quad (7)$$

5. 提案システムによる観光スポット検索

提案する観光スポット検索システムのユーザインタフェースを図 3 に示す。図の左部では提案システムの入力スポットとして利用可能なスポットの一覧を検索することが可能であり、各スポットのカテゴリ・地域・レビュー数などが補足情報として表示される。図の中央上部は入力スポットと入力キーワードをシステムへの入力部である。「検索」ボタンをクリックすることで入力スポットと入力キーワードの特徴ベクトルを算出しベクトル間の加算を行い、加算結果のベクトルとの類似度に基づき検索されたスポットが中央下部の地図上に表示される。検索に用いる入力キーワードの特徴ベクトルは、4.1 節で示した相加

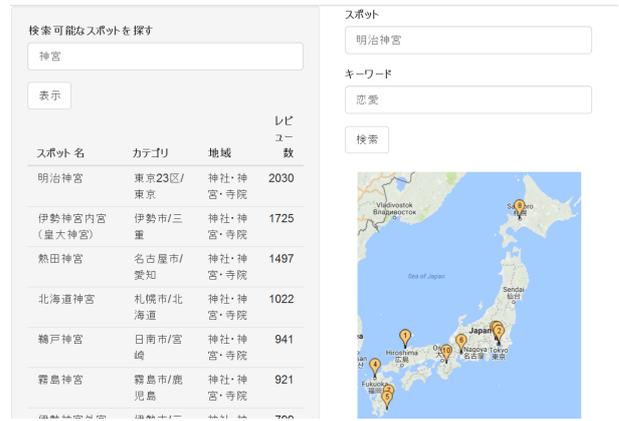


図 3 提案する観光スポット検索システムのユーザインタフェース

表 2 提案システムによる“明治神宮 + 恋愛”の検索結果

スポット名	元順位	相加平均	最大値	単語頻度
出雲大社	15	△5	△8	△7
東京大神宮	5	△3	△4	△4
川越氷川神社	55	△26	△30	△38
榊原神社（富山県）	16	△5	△10	△11
榊原神社	50	△33	△41	△41
熱田神宮	2	▽1	△0	▽1
宮崎神宮	3	△0	△0	△1
北海道神宮	4	▽1	▽1	▽3
武蔵一宮氷川神社	14	△2	△4	▽3
榊原神宮	8	▽1	▽5	▽13

平均により算出された入力キーワードの特徴ベクトル、4.2 節で示した各次元の最大値に基づく重み付き平均により算出された入力キーワードの特徴ベクトル、4.3 節で示した単語頻度に基づく重み付き平均により算出された入力キーワードの特徴ベクトルのいずれかを用い、3 手法のどれを用いるかは選択することが出来る。

提案システムを用いたスポット検索結果例として、表 2 に東京都内の著名な神社である明治神宮を入力スポットとした“明治神宮 + 恋愛”の検索結果を示す。表中の各観光スポットのコサイン類似度に基づく順位は Paragraph Vector モデルで特徴ベクトルを作成した全 43,759 スポット中での順位を表す。また、表中の記号はそれぞれの手法で算出した入力キーワード「恋愛」の特徴ベクトルを用いた加算による順位付け結果が「元順位」に比べて何位分上昇/下降したかを表している。例えば、「△5」は「元順位」に比べて順位が 5 位上昇したことを、「▽5」は「元順位」に比べて順位が 5 位下降したことを表している。

表 2 において、全ての手法による順位付けで順位が上昇した「出雲大社」「川越氷川神社」などの 5 つのスポットはいずれも縁結びや恋愛成就などで著名な神社であり、入力キーワード「恋愛」との関連が深いスポットであると考えられる。一方、「熱田神宮」「宮崎神宮」など前述したスポット以外の 5 つのスポットは縁結びや恋愛成就などの特徴を一般に持たないと考えられているスポットである。この結果から、“明治神宮 + 恋愛”の演算を行うことで縁結びや恋愛成就などの「恋愛」と関連する主観的特徴を考慮した観光スポット検索が行えていると考え

られる。

6. ま と め

本稿では、観光スポットが持つ様々な特徴の中でタグなどのメタデータでは推測困難な「夜景が美しい」「待ち合わせ場所に向いている」などの主観的特徴を考慮した観光スポット検索を行うため、Paragraph Vector モデルにより算出した各スポットの特徴ベクトルを用いて任意の主観的特徴の特徴ベクトルを動的に算出し、ある観光スポットと比較して対象の主観的特徴を強く持つ他の観光スポットを検索する手法を提案した。今後の課題として、提案した観光スポット検索手法によるスポット検索手法の精度を定量的に評価する実験を行う予定である。また、使用するユーザ個人個人の嗜好を考慮したよりパーソナライズされたオブジェクト検索を行えるよう改善することが考えられる。提案手法ではベクトル空間モデルに基づく特徴ベクトルを用いたコサイン類似度による検索を行ったが、この特徴ベクトルはユーザレビューから抽出したオブジェクトが持つ特徴のみを考慮して算出されているためユーザ個人個人の嗜好は反映されていない。しかし、本研究で定義する主観的特徴は本来ユーザが対象オブジェクトを実際に利用することでユーザが得られる印象・感情・体験などを表すため、オブジェクトを検索するユーザ個人個人の嗜好と組み合わせて用いることでより効果的な検索が行えると考えられる。

謝 辞

本研究の一部は、平成 28 年度科研費基盤研究 (B)(課題番号: 26280042) によるものです。ここに記して謝意を表すものとします。

文 献

- [1] Mark Sanderson and Janet Kohler. Analyzing geographic queries. In *Proceedings of the SIGIR 2004 Workshop on Geographic Information Retrieval*, 2004.
- [2] Luis Gravano, Vasileios Hatzivassiloglou, and Richard Lichtenstein. Categorizing web queries according to geographical locality. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, CIKM '03, pp. 325–333, 2003.
- [3] 吉田朋史, 北山大輔. ユーザレビューの分散表現を用いた役割的に類似する観光スポット検索手法. 観光情報学会第 14 回研究発表会講演論文集, pp. 78–81, 2016.
- [4] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014*, pp. 1188–1196, 2014.
- [5] 土田崇仁, 遠藤雅樹, 加藤大受, 江原遥, 廣田雅春, 横山昌平, 石川博. Word2vec を用いた地域やランドマークの意味演算. DEIM Forum 2016 最終論文集, pp. H5–1, 2016.
- [6] 石野亜耶, 藤井一輝, 藤原泰士, 前田剛, 難波英嗣, 竹澤寿幸. 旅行ブログエントリと質問応答コンテンツを利用した旅行ガイドブックの情報拡張. 人工知能学会論文誌, Vol. 29, No. 3, pp. 328–342, 2014.
- [7] 伊達賢志, 北須賀輝明, 糸川剛, 有次正義. 旅先での観光地選り支援のためのブログを用いた観光地の印象抽出手法. マルチメディア, 分散協調とモバイルシンポジウム 2011 論文集, pp. 1566–1579, 2011.
- [8] Shinsuke Nakajima and Katsumi Tanaka. Relative queries and the relative cluster-mapping method. In YoonJoon Lee, Jianzhong Li, Kyu-Young Whang, and Doheon Lee, editors, *Database Systems for Advanced Applications: 9th International Conference, DASFAA 2004, Jeju Island, Korea, March 17-19, 2003. Proceedings.*, pp. 843–856. Springer Berlin Heidelberg, 2004.
- [9] Makoto P. Kato, Hiroaki Ohshima, Satoshi Oyama, and Katsumi Tanaka. Query by analogical example: Relational search using web search engine indices. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pp. 27–36, 2009.
- [10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pp. 3111–3119. 2013.
- [11] Zellig S Harris. Distributional structure. *Word*, Vol. 10, No. 2-3, pp. 146–162, 1954.