

観光ガイドマッチング方法に関する考察

園田泰子^{†1} 桑田和^{†2} 長尾将宏^{†1} 川村武人^{†3}
柴田祥宏^{†3} 成瑞珈^{†3} 小谷勇人^{†3} 本橋洋介^{†1}

概要：観光ガイドとは、観光客に対して観光地におけるスポットの推薦や、各観光スポットの説明を詳細に行う。観光客と観光ガイドの相性が良いことによって、観光時の満足度向上やリピート率向上など観光地に対して良い効果があると考えられる。そこで本研究では、観光客と観光ガイドの相性が良くなるマッチング方法について検討する。観光客の属性、観光に対するコメント、観光客と観光ガイドの過去のガイド実績の一致度などを基にマッチングの適切さについて検討を実施した。本稿では、検討したアルゴリズムおよび実験結果を説明する。

キーワード：自然言語処理, 形態素解析, 観光業, 観光ガイド, マッチング

A research of method of matching guides and tourists for sightseeing

YASUKO SONODA^{†1} IZUMI KUWATA^{†2} MASAHIRO NAGAO^{†1}
TAKETO KAWAMURA^{†3} YOSHIHIRO SHIBATA^{†3} CHENG RUIJIA^{†3}
HAYATO KOTANI^{†3} YOSUKE MOTOHASHI^{†1}

Abstract: Guides recommend tourist spots and provide information about tourist spots for tourists. There is a possibility that positive effects on tourist spots such as improvement of satisfaction and repetition rate during sightseeing are obtained by affinity between guides and tourists. Therefore, this paper reviews method of matching guides and tourists for sightseeing. We examine the appropriateness of matching based on the attributes of tourists, comments on sightseeing, and the degree of coincidence between guides and tourists in the past. This paper explains about the algorithms and experimental results.

Keywords: Natural Language Processing, NLP, Morphological Analysis, Sightseeing, Tourist guide

1. はじめに

観光ガイドとは、観光客に対して観光地におけるスポットの推薦や、各観光スポットの説明を詳細に行う。観光客と観光ガイドの相性が良いことにより、観光時の満足度向上やリピート率向上など観光地に対して良い効果があると考えられる。そこで本研究では、観光客と観光ガイドの相性が良くなるマッチング方法を検討する。

図1に本研究で検証する観光ガイドシステムの全体像を示す。観光客がガイド依頼申請画面にて観光客自身の属性や行きたい観光スポットを入力すると、観光ガイドの属性と観光ガイドのこれまでのガイド実績データを踏まえて、観光客と相性が良さそうな観光ガイドを出力する。観光ガイドを出力することができれば、観光客に対して、システムが自動で観光ガイドをアサインすることが可能となる。

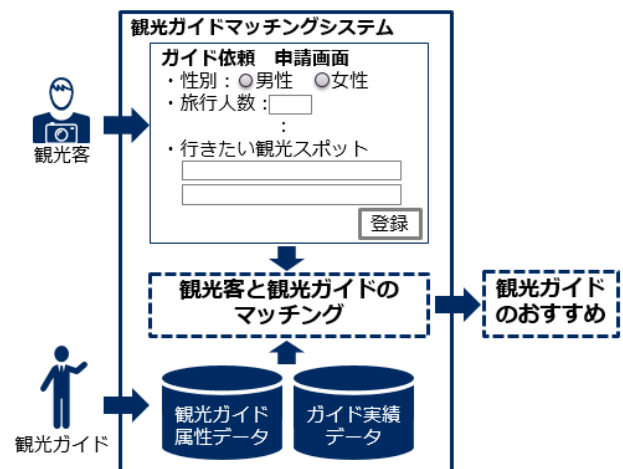


図1 観光ガイドマッチングシステム

Figure 1 Guides and Guests Matching System.

1.1 使用データ

本研究では、以下の観光データおよび、以下 (b) の各スポットの説明情報を使用する。

- (a) 観光客, 観光ガイド, ツアーの属性データ
- (b) 実際に観光したスポット情報
- (c) 各ツアーにおける消費金額
- (d) 各ツアーにおける観光ガイドの所感

^{†1} 日本電気株式会社
NEC Corporation

^{†2} 早稲田大学
WASEDA University

^{†3} NEC ソリューションイノベータ株式会社
NEC Solution Innovators, Ltd

(a), (b), (c) および (d) は、外国人観光客を対象にしたあるガイドサービスの観光実績データを使用する。観光客と観光ガイドが1日で訪れたスポットをまとめて「ツアー」と呼ぶ(図2)。なお、上記データのうち、2回以上ガイドしたことがある観光ガイドの実績データを対象とする。また、(b)の各スポットの説明情報として、各スポットのWikipediaの概要文を使用する。

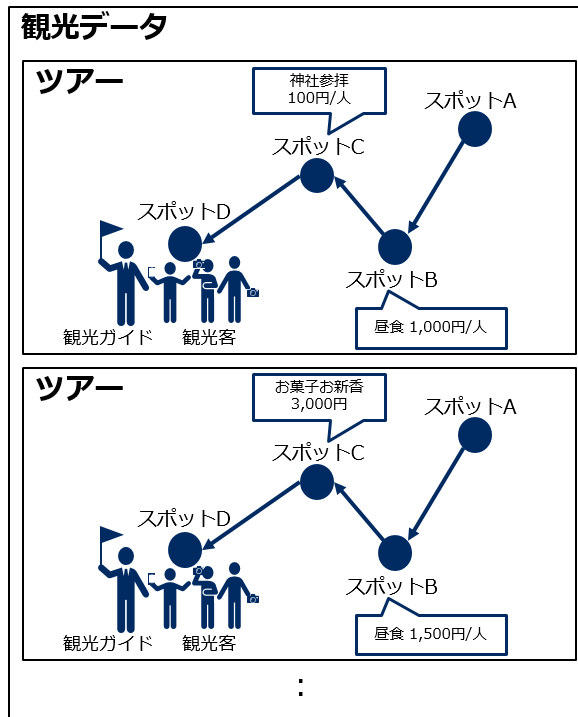


図2 観光データ
Figure 2 Guide Data.

2. 関連研究・事例

観光業におけるAIや機械学習などデータを活用した取り組みには、主に観光地や観光ルートの推薦に関するものがある。また、人のマッチングに関するAIや機械学習の事例は、教育や人材紹介等におけるものがある。

2.1 観光業における関連研究・事例

観光ナビゲーションシステムにおいて、個人の嗜好にカスタマイズされたランドマークを推薦する研究がある[1]。最近では、チャット形式で観光地やレストラン等をおすすめし、観光ルートを検索できるスマートフォンアプリが提供されている[2]。これらは観光客に対して観光地や観光地までのルートを推薦するものである。

2.2 人のマッチングにおける事例

人と人のマッチングに関する事例では、ビジネスパートナーを増やしたい人に向けたビジネスマッチングアプリが挙げられる[3]。これは、初回登録時にユーザが入力した情報から、既に登録済みのユーザとの相性などを算出するものである。

しかしながら、観光ガイドと観光客の適切なマッチングに関する研究はまだない。

3. マッチング手法

3.1 手法概要

図1における観光客と観光ガイドのマッチングの手法として、教師なし学習、教師あり学習の3種類でマッチングを検証する(表1)。

表1 観光客と観光ガイドのマッチング手法

Table 1 Method of Matching Guides and Guests.

No	手法	概要
1	教師なし学習	観光データ (b) と各スポットの説明情報を使用する。スポットのクラスタリング結果を基に、観光客および観光ガイドが訪れたスポットを集計する。集計結果をベクトル値ととらえ、各ベクトルのユークリッド距離を算出する。距離が近い観光客と観光ガイドの組合せとそうでない組合せで、観光データ (c) や観光データ (d) を基に算出した満足度の差を比較する。
2	教師あり学習	観光データ (a) (b) を説明変数とする。観光データ (c) を目的変数とするモデルと、観光データ (d) を基に算出した満足度を目的変数とするモデルを作成する。
3	教師あり学習 (説明変数追加)	観光データ (a) (b) に加え、教師なし学習で算出した観光客と観光ガイドのベクトル情報を説明変数とする。目的変数は、No.2の教師あり学習と同一とする。

3.2 満足度の算出方法

教師なし学習の評価対象であり、教師あり学習の目的変数である「満足度」の算出方法について述べる。

まず、観光データ (d) を形態素解析し、単語の活用形を終止形に変換する。終止形に変換した単語群を単語感情極性対応表[4]に基づき、所感が持つ単語ごとの極性スコア(ポジティブな単語ほど1に近く、ネガティブな単語ほど-1に近い値を取る)の平均値を算出する。これを満足度スコア1とする。

次に、日本語評価極性辞書(用言編) ver.1.0 [5][6]および日本語評価極性辞書(名詞編) ver.1.0 [7][8]による文章のネガポジ判別を実装したPythonライブラリ[9]を使用し、所感のネガポジスコア(ポジティブな文章ほど1に近く、ネガティブな文章ほど-1に近い値を取る)を出力する。これを満足度スコア2とする。

最後に、満足度スコア1と満足度スコア2を加算し、正規化した値を満足度とする。満足度は0~1の値を取り、値が大きいくほど満足度が高いとみなす。

4. 教師なし学習

マッチング手法 No.1 教師なし学習について述べる。

まず、ツアー単位のデータを学習用データと評価用データに分割する。学習用データの観光ガイドを (A) 観光客を (A') と定義する。また、評価用データの観光ガイドを (B)、観光客を (B') と定義する (図 3)。

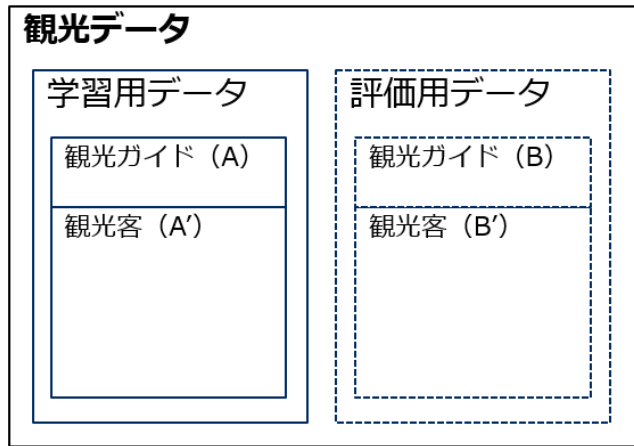


図 3 観光データの分割

Figure 3 Separation of Guide Data.

教師なし学習の概要を図 4 に示す。

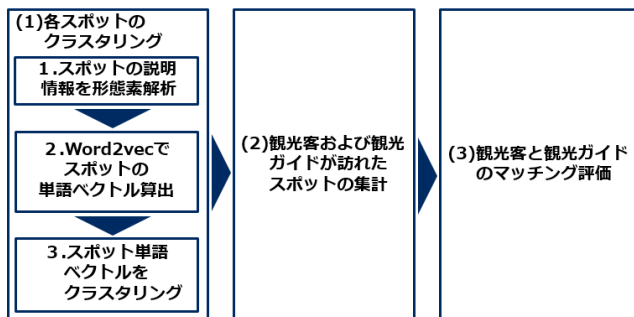


図 4 教師なし学習概要

Figure 4 Overview of Unsupervised Learning.

(1) 各スポットのクラスタリング

全データの各スポットの説明情報を形態素解析し、名詞、動詞、形容詞、形容動詞である単語を対象に 200 次元の単語ベクトルを作成する。次に、単語ベクトルを k-means 法で 65 個のクラスタに分割する。そして、分割結果を基に、各スポットの説明情報に含まれる単語がどのクラスタに属しているかを集計する。

(2) 観光客および観光ガイドの訪れたスポットの集計

評価用データの観光客 (B') と学習用データの観光ガイド (A) が実際に観光したスポットを使い、観光客および観光ガイドがどのスポットクラスタを何回観光しているかを集計する。これを観光客ベクトルおよび観光ガイドベクトルと呼ぶ。

(3) 観光客と観光ガイドのマッチング評価

最後に、各観光客と観光ガイドのマッチング度合いを、観光ガイドベクトルと観光客のユークリッド距離によって算出する。観光客ベクトルと観光ガイドベクトルのユークリッド距離に近いほどマッチング度合いが高いとみなす。評価用データの観光ガイド (B) と観光客 (B') において、マッチング度合いの高い組合せとそうでない組合せを比較して評価する。

5. 教師あり学習

マッチング手法 No. 2 および No. 3 の教師あり学習について述べる。

説明変数の異なる 2 パターンについて、目的変数として「消費金額」「満足度」の 2 項目を設定した計 4 種類のモデルを作成した。学習アルゴリズムには、勾配ブースティングを用いた。表 2 に、教師あり学習モデルの作成に関する詳細な情報を述べる。

表 2 教師あり学習モデルの詳細

Table 2 Detail of Supervised Learning.

	マッチング手法 No.2	マッチング手法 No.3
アルゴリズム (ライブラリ)	LightGBM	LightGBM
学習データ数	① 消費金額：6515 ② 満足度：6153	① 消費金額：6515 ② 満足度：6153
検証データ数	① 消費金額：6164 ② 満足度：5799	① 消費金額：6164 ② 満足度：5799
実際に観光したスポットを含むツアー情報の主要な説明変数 (○がついている変数を使用)		
スポットのエリア・休日フラグなど 4 種類	○	○
観光ガイドベクトル-観光客ベクトル間の平均絶対誤差 (数値)		○
観光客の主要な説明変数 (○がついている変数を使用)		
使用言語・旅行人数・子供有無など 11 種類	○	○
観光客ベクトル (数値)		○
観光ガイドの主要な説明変数 (○がついている変数を使用)		
年代・エリア・使用可能言語など 6 種類	○	○
観光ガイドベクトル (数値)		○
目的変数 (○がついている変数を使用) ※1 目的変数別でモデルを作成		
消費金額 (数値)	○ (※1)	○ (※1)
満足度 (数値)	○ (※1)	○ (※1)

観光客と観光ガイドの組合せにおいて、予測した消費金額や満足度が高いほどマッチング度合いが高いとみなす。

5.1 マッチング手法 No.2

観光客の属性データは、表 2 の観光客の主要な説明変数を含む計 11 種類のデータを使用した。また、観光ガイドの属性データは、表 2 の観光ガイドの主要な説明変数を含む計 6 種類のデータを使用した。ツアーに関するデータとして、表 2 の実際に観光したスポットを含むツアー情報の主要な説明変数 4 種類を使用した。

上記の観光客の属性データ、観光ガイドの属性データを使用して勾配ブースティングモデルを作成した。

5.2 マッチング手法 No.3

(1) で使用した観光データ (a) (b) に加え、4. 教師なし学習で作成した観光ガイドベクトル、観光客ベクトルを使用して勾配ブースティングモデルを作成した。追加した説明変数は、スポットクラスごとに観光ガイドベクトルおよび観光客ベクトルの平均値と、両者の平均絶対誤差を算出結果の 3 種類である。

6. 評価結果

6.1 教師なし学習

評価用データに属する観光ガイド (B) が、実際に評価用データの観光客 (B') をガイドしていた群を「一致グループ」とし、そうでない群を「不一致グループ」とし、各グループの消費金額の中央値と平均値を示す (表 3)。

表 3 教師なし学習結果 (消費金額の比較)

Table 3 Result of Unsupervised Learning
(Comparison of Consumption).

	中央値		平均値	
	一人当たり	グループ全体	一人当たり	グループ全体
一致グループ (770 件)	2200	5058	4354	11388
不一致グループ (5394 件)	2180	4800	4585	11828

平均値で比較すると、不一致グループの方が消費金額は高い結果となった。これは、消費金額のはずれ値 (高額商品の購入等) が不一致グループに集中しているためと考えられる。中央値で比較すると、わずかながら一致グループの方が高い結果となった。

次に、各グループの満足度の中央値と平均値を示す (表 4)。

表 4 教師なし学習結果 (満足度の比較)

Table 4 Result of Unsupervised Learning
(Comparison of Satisfaction).

	中央値	平均値
一致グループ (727 件)	0.4798	0.4797
不一致グループ (5072 件)	0.4773	0.4767

満足度においては、中央値、平均値ともにわずかながら一致グループの方が高い結果となった。実際に所感に記載されている文章を観察すると、ポジティブな文章が書かれていることが多かったため、所感を基に作成した満足度には大きな差が見られなかったと考えられる。

6.2 教師あり学習

観光データ (a) (b) を使用したモデル、観光データ (a) (b) およびスポットクラスタを使用したモデルの RMSE と MAE を表 5 に示す。

表 5 教師あり学習結果

Table 5 Result of supervised Learning.

目的変数	マッチング手法 No.2		マッチング手法 No.3	
	RMSE	MAE	RMSE	MAE
消費金額	5845	1535	5747	1511
満足度	0.0890	0.0674	0.0883	0.0666

説明変数を追加したマッチング手法 No.3 の方が、消費金額・満足度のいずれも RMSE, MAE とともに誤差が小さい結果となった。消費金額の RMSE と MAE の差が大きくなってるのは、高額商品の購入等のはずれ値の影響が表れているためと考えられる。

次に、マッチング手法 No.2, No.3 において、目的変数を消費金額としたモデルの説明変数の重要度の上位 10 件を表 6 に示す。

表 6 重要度の高い説明変数

Table 6 Explanatory Variables with High Importance.

順位	マッチング手法 No.2	マッチング手法 No.3
1	ガイドプランニング経験値	スポットの区分が買物である
2	スポットの区分が買物である	スポットクラスタ 3 における観光客ベクトルの平均値
3	スポットの区分が食事である	スポットクラスタ 4 における観光ガイドベクトルと観光客ベクトルの平均絶対誤差
4	スポットの区分が体験である	スポットクラスタ 16 における観光客ベクトルの平均値
5	エリアが新橋・浜松町・田町である	スポットクラスタ 4 における観光客ベクトルの平均値

6	エリアが浅草橋・両国・押上である	スポットクラスタ 25 における観光ガイドベクトルと観光ベクトルの平均絶対誤差
7	観光ガイドが 80 年代以上である	スポットクラスタ 52 における観光ガイドベクトルの平均値
8	エリアが飯田橋・神楽坂・四ツ谷である	スポットクラスタ 41 における観光客ベクトルの平均値
9	ゲストの使用言語が英語である	スポットクラスタ 18 における観光ガイドベクトルと観光ベクトルの平均絶対誤差
10	エリアが鎌倉・大船である	スポットクラスタ 31 における観光客ベクトルの平均値

マッチング手法 No.2 のモデルでは、スポットの区分やスポットのエリアの重要度が高い結果となった。観光ガイドが観光プランを立てた経験が多く、買物・食事・体験を目的として観光スポットを訪れている場合、消費金額が高い傾向にあると言える。

一方で、スポットクラスタ情報を説明変数に追加したマッチング手法 No.3 のモデルでは、スポットクラスタを基に追加した説明変数の重要度が高いという結果になった。まず、スポットクラスタ 3 における観光客ベクトルの平均値について述べる。重要度の高いスポットクラスタ 3 のベクトル値が高いスポットは美術館や博物館であった。美術館や博物館に行けば入館料を払うため、消費金額を予測するモデルにおいて重要度が高くなったと考えられる。実際に、スポットクラスタ 3 のベクトル値上位 10 スポットを訪れた観光客のうち 51.6%が数百円～数千円を支払っている。しかし、金額の使用用途は入館料だけではなく、隣接するカフェなどでの食事代やお土産代が含まれているデータもあった。

次に、スポットクラスタにおける観光ガイドベクトルと観光客ベクトルの平均絶対誤差について述べる。この説明変数は、観光ガイドが過去にガイドしたスポットと観光客の訪れたスポットが似ているかどうかという指標である。スポットクラスタ 4 における観光ガイドベクトルと観光客ベクトルの平均絶対誤差について考察する。スポットクラスタ 4 のベクトル値が高いスポットは、レンタル着物店や古民家、恩賜庭園であった。スポットクラスタ 4 のベクトル値上位 10 スポットを訪れた観光客の金額の使用用途は、入園料や着物体験、抹茶セットが挙げられる。観光ガイドの詳細な説明が必要となりうるスポットにおいては、観光ガイドが過去にガイドしたスポットと、観光客の訪れたスポットの近さが消費金額に影響すると考えられる。

6.3 マッチング手法別結果比較

各マッチング手法の結果を基に、観光客の消費金額、満足度を比較する。

マッチング手法別、観光ガイド・観光客ペア別の消費金額の中央値を比較した結果は図 5 のとおりである。

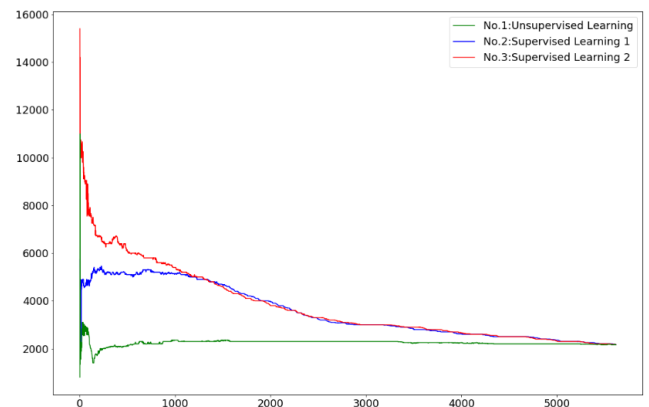


図 5 マッチング手法別 観光ガイド・観光客ペア別消費金額の中央値比較

Figure 5 Comparison of Median Consumption by Matching Method / Tourist and Guide Pair.

横軸は、マッチング度合いの高い上位 n 件までを縦軸の値の算出対象としたかを示す。マッチング度合いの定義は、教師あり学習の場合、観光ガイドと観光客の組合せにおいて予測結果の消費金額が高い順に並べ、予測消費金額が高いほどマッチング度合いが高いと判断する。教師なし学習の場合、観光ガイドベクトルと観光客ベクトルのユークリッド距離の近い組合せ順に並べた結果である。縦軸は実際の消費金額の中央値を表している。つまり、算出対象が少ない状態では縦軸が高い数値で、算出対象が増えると徐々に実際の消費金額の中央値が下がっていく状態が良い形状と言える。本研究においては、スポットクラスタを説明変数に加えた教師あり学習（図 5 の赤線）が最も良いモデルと言える。

次に、マッチング手法別、観光ガイド・観光客ペア別の満足度の平均値を比較した結果を図 6 に示す。

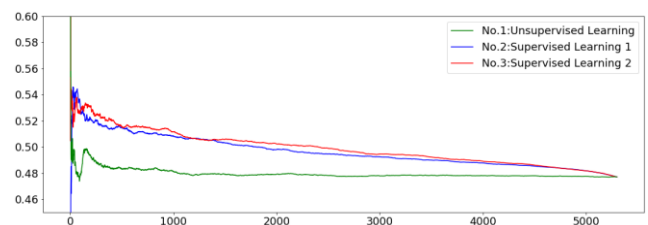


図 6 マッチング手法別 観光ガイド・観光客ペア別満足度の平均値比較

Figure 6 Comparison of Satisfaction Averages by Matching Method / Tourist and Guide Pair.

横軸は予測結果の満足度が高い順に観光ガイドと観光客の組合せを並べ、上位何件までを縦軸の値の算出対象としたか、縦軸は実際の満足度の平均値を表している。図 5

と同様、算出対象が少ない状態では縦軸が高い数値で、算出対象が増えると徐々に満足度の平均値が下がっていく状態が良い形状と言える。教師あり学習モデル（図6の青線および赤線）は算出対象が増えるにつれて、満足度の平均値が下がっていることがわかる。また、教師なし学習モデル（図6の緑線）は算出対象が少ない状態では3手法のうち最も良い満足度の平均値を取るが、上位20位以降は満足度の平均値に大きな変化が見られなかった。

7. 考察とまとめ

本分析の結果から、満足度や消費金額が高くなるマッチングを実現できる可能性を確認した。教師あり学習モデルにおいて、スポットクラスごとの観光ゲストベクトルと観光ガイドベクトルの平均絶対誤差を用いた方が、より消費金額や満足度が高くなることを確認した。これにより、観光客の基本属性の他に、どこに行きたいのかといったスポットの情報を予め聞いておくことで満足度の高いマッチングが行える可能性があることが分かった。

また、教師なし学習モデルを作成したところわずかながら傾向は見て取れるが、教師あり学習のように満足度を高めるマッチングを十分作成することが難しいという結果になった。

本分析では目的変数の一つとして観光ガイドの所感から算出した満足度を使用しているが、観光客自身が入力したものではない。観光客のガイド実施後のアンケート内で回答してもらうことにより、分析の質が向上することが期待できる。

今後、本研究で分かったことを踏まえて、さらなるマッチング手法の改善を行い、運用を通して実用的なマッチングアプリケーションを開発していきたい。

参考文献

- [1] 高木修一, 益田真輝, 仲谷善雄 "個人の嗜好にカスタマイズされたランドマークを用いた観光ナビの提案", 第74回全国大会講演論文集, 2012(1), pp.305 - 306, 2012-03-06.
- [2] JTB JAPAN Trip Navigator
<https://www.jtb.co.jp/inbound/appli/index.asp>
- [3] ビジネスマッチングアプリ yenta <https://yenta.talentbase.io/yenta/>
- [4] 高村大也, 乾孝司, 奥村学 "スピンモデルによる単語の感情極性抽出", 情報処理学会論文誌ジャーナル, Vol.47 No.02 pp. 627--637, 2006.
- [5] 東北大学 乾・岡崎研究室 (2008). 日本語評価極性辞書 (用言編) ver.1.0 (2008年12月版) / Inui-Okazaki Laboratory, Tohoku University(2008). Japanese Sentiment Dictionary (Volume of Verbs and Adjectives) ver. 1.0
<https://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary>
- [6] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol.12, No.3, pp.203-222, 2005. / Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi. Collecting Evaluative Expressions for Opinion Extraction, Journal of Natural Language Processing 12(3),

- 203-222, 2005.
- [7] 東北大学 乾・岡崎研究室 (2008). 日本語評価極性辞書 (名詞編) ver.1.0 (2008年12月版) / Inui-Okazaki Laboratory, Tohoku University(2008). Japanese Sentiment Dictionary (Volume of Nouns) ver. 1.0
<https://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary>
- [8] 東山昌彦, 乾健太郎, 松本裕治, 述語の選択選好性に着目した名詞評価極性の獲得, 言語処理学会第14回年次大会論文集, pp.584-587, 2008. / Masahiko Higashiyama, Kentaro Inui, Yuji Matsumoto. Learning Sentiment of Nouns from Selectional Preferences of Verbs and Adjectives, Proceedings of the 14th Annual Meeting of the Association for Natural Language Processing, pp.584-587, 2008.
- [9] oseti · PyPI <https://pypi.org/project/oseti/>