

# 旅行ブログと特許からの地域性の検出

伊藤 拓真<sup>†</sup> 難波 英嗣<sup>††</sup> 石野 亜耶<sup>†††</sup> 竹澤 寿幸<sup>††</sup>

<sup>†</sup> 広島市立大学 情報科学部 〒731-3194 広島県広島市安佐南区大塚東 3-4-1

<sup>††</sup> 広島市立大学大学院 情報科学研究科 〒731-3194 広島県広島市安佐南区大塚東 3-4-1

<sup>†††</sup> 広島経済大学 ビジネス情報学科 〒731-0138 広島県広島市安佐南区祇園 5-37-1

E-mail: {ito, nanba, ishino, takezawa}@ls.info.hiroshima-cu.ac.jp

**あらまし** 本研究では、特許の開発者の住所と発明の内容に何らかの関連性があると仮定し、旅行ブログを利用して特許からの地域性の検出を試みる。特許にはその土地の産業に根差したものが少なくない。そのような地域性を有する特許を旅行ブログエントリと比較・選別することで、特許の地域性が検出できる。地域性の高い特許は、その地域特有の産業と関連性が高いことが多いため、観光を産業という観点からとらえることで地域産業の活性化も期待できる。そこで本研究では、旅行ブログを利用して、入力に対して関連性のある地域を推定するシステムを開発する。また、特許と旅行ブログの地域性を視覚的に捉えるためにシステムを開発する。これらのシステムにより、地域性の高い特許と、その特許の関連地域を検出することができた。

**キーワード** 特許, 旅行ブログ, 観光

## 1. はじめに

特許には地域産業に根差すものが少なくない。例えば、愛媛県は都道府県別のみかんの生産量が2位であり、みかんは愛媛県の特産品として有名である。また、愛媛大学農学部はみかん関連の特許を多く出願しており、この特許は地域性を有していると考えられる。産業という観点から観光を捉え、その地域特有の産業を観光に結び付けることは、新たな観光資源の開拓につながるため、地域産業の活性化が期待できる。

そこで、本研究では、特許と旅行ブログエントリを用い、特許の地域性を検出するシステムを開発する。ある地域を訪れた旅行者が書いた旅行ブログエントリには、その地域における様々な旅行経験が綴られている。一方、地域性を有する特許には、その発明の内容と発明者の住所に何らかの関連性があると思われる。そこで、本研究では、この両者を重ね合わせることで、産業という観点からの地域の特徴を検出し、また、地域性を有する特許を検出する。

## 2. 提案システムを用いた分析例

本節では本研究で構築するシステムを用いた分析例について説明する。本研究では、地域性の検出のために2つのシステムを開発した。

表1: “okonomiyaki”, “pasta”の地域推定結果

入力	okonomiyaki	pasta
推定 地域	Japan-Hiroshima	Italy-Roma
	Japan-Hirakata	Italy-Florence
	Japan-Kyoto	US-Chicago
	Japan-Osaka	Italy-Venice
	Japan-Setagaya	Italy-Bologna
	Japan-Miyajima	Italy-Urbino
	Japan-Akasaka	Italy-Genoa

1つ目のシステムは、地域推定システムである。地域カテゴリ付きの旅行ブログデータを用いており、テキスト（語句）を入力として、入力に対する関連地域をk近傍法によって推定し、出力する。表1は、このシステムに“okonomiyaki”（お好み焼き）と“pasta”を入力したときの出力結果である。

このシステムの入力に特許を利用することで、その特許の地域性を検出することができる。例えば、入力として粒子加速器に関する特許のタイトルとアブストラクトを使用した場合、旅行ブログ側でスイス-ジュネーブにおいて図1に示すような「CERNの展示会に訪れ、世界最大の粒子加速器を見てきた」というような内容のブログが投稿されているので、入力特許の関連地域として、“Switzerland-Geneva”が推定される。



(<https://www.travelblog.org/Europe/Switzerland/South-West/Geneva/blog-850730.html>)

図1: 粒子加速器に関するブログの例

このように、特許を旅行ブログによって分類し、両者の位置情報の距離を考慮に入れた分析を行うことで、地域性の高い特許を検出することが可能になる。

2 つ目のマッピングシステムについて紹介する。このシステムは、入力された検索クエリと適合度の高い特許と旅行ブログエントリそれぞれを検索し、頻度に応じた色で地図上にプロットする。地域性のある地域が赤く表示されている。図 2 に特許のマッピングシステムに対して、検索クエリとして“vehicle”を入力したときの結果を示す。

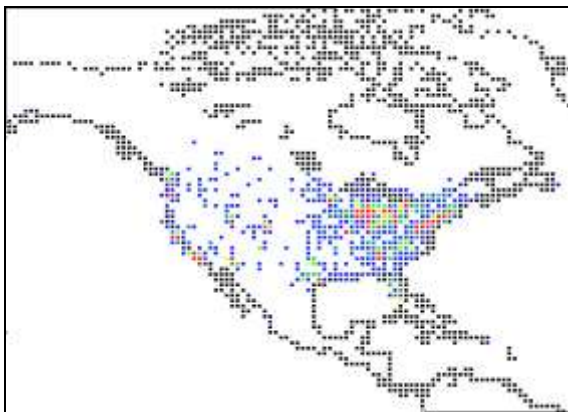


図 2：特許マッピングシステムに“vehicle”と入力したときの出力例

図 2 の出力結果の場合は、自動車産業の栄えているミシガン州周辺で地域性が検出できていることが分かる。また、旅行ブログにおいても同様に、“vehicle”を入力として使用した場合のマッピングシステムの出力結果を確認してみると、ミシガン州周辺で車に関するブログエントリが多く投稿されていることが分かった。それらのブログエントリの具体的な内容を閲覧してみると、「自動車歴史博物館を訪問した」というような、自動車に関する観光の情報が存在することが分かり、特許と旅行ブログの重なり合いが見つかった。

### 3. 関連研究

本研究では、特許から地域性を検出することを目的としているが、これまでも地域特有の情報の抽出や、地域性の判定に関する研究が行われている。

本研究の関連研究として、3.1 節では、Web を情報源とした分析を行っている研究、3.2 節では、旅行ブログエントリを情報源とした分析を行っている研究をそれぞれ紹介する。

#### 3.1. Web を情報源とした分析

手塚ら[1]は、任意のキーワード型検索クエリに対するウェブページの検索結果や、検索クエリ自体が持つ空間的属性である地域性を推定する手法を提案している。例えば、ユーザーが「うどん」が有名な場所を調べたいという場合、「うどん」というキーワードを入力すると、「うどん」に対するウェブページの検索結果を

実時間で解析し、「うどん」が有名な地域や、関連度の高い地域を地図上に表示させる。任意の検索クエリを入力として、地域性を検出している点で、本研究と非常に類似している。しかし、こちらの研究では、ウェブページを対象とし、マッピングに利用する位置情報として文書中に現れる一般の地物名を使用している。本研究では、特許を対象とし、特許内に表記されている開発者の住所を利用しており、旅行ブログエントリとの重ね合わせで地域性を検出している点で異なる。また、森本ら[2]は、任意のキーワードをクエリとしてロボットを用いてウェブページのクロールを行い、ユーザーから指定された種類の施設に関して情報を収集することで、既存のウェブ上の施設検索サイトには存在しない情報を、ユーザーに提示するシステムを構築している。地域性の推定にはページ中に含まれる住所表記を利用している。ウェブページを情報源として解析を行っているが、住所情報から地域性の推定を行っている点で本研究と似ているが、特許と旅行ブログエントリの位置情報の比較によって、産業という観点からの地域性の検出を試みている点で異なる。

相良ら[3]はタウンページ情報を辞書として用いることで、Web から特定地域の特定業種に関する空間コンテンツを効率的に収集し、HTML 構造を解析して順位付けを行う手法を開発し、実用性の高い空間サーチエンジンを構築している。つまり、タウンページを用いて、Web ページを解析して出力を得ている。本研究では、旅行ブログエントリを用いて、関連地域を推定するシステムに特許を入力し、出力を分析することで、特許からの地域性の検出を試みている。

奥ら[4]は語句の地域限定性に着目し、旅行先などの対象地域固有の語句を自動的に抽出するための尺度として地域限定スコアを提案している。地域限定性スコアは、IDF と地域との関連性の重みを組み合わせることで定義される。例えば、三重県松坂市の松坂駅周辺で夕食をとりたいとき、数ある飲食店の中から「松坂牛」のような、松坂市固有の特産品を提供してくれる店を提示してくれる。つまり、指定された現地に存在するコンテンツの語句の中から対象地域固有の限定語句をグルメ情報サイトや、スポット情報サイトから自動的に抽出し、地図上に視覚化している。本研究では、任意の検索クエリから地域性を検出することを試みているので、入力と出力が異なる。

#### 3.2. 旅行ブログエントリを情報源とした分析

新田ら[5]の研究では旅行ブログエントリから「訪れた場所」、「食べたもの」、「購入したもの」の情報が記載してあるものについて手掛かり語を用いて収集し、機械学習を利用することで旅行者の行動情報を自動抽出し、抽出した旅行ブログエントリを地図上に表示す

るシステムを構築することで、旅行者の行動に基づいた地域性を判定している。Fujiiraら[6]は旅行ブログエントリについて、投稿したブロガーの属性と、「買う」、「食べる」、「体験する」、「泊まる」、「見る」といった「観光タイプ」をブログエントリ本体の属性として自動判定し、判定された属性に基づいた分析を行うことにより旅行者の特徴や各都市の地域性を判定している。Haoら[7]は、確率トピックモデルに基づく地域トピックモデル（Location-Topic model）を用いて、地域を特徴づけるキーワードを旅行ブログから抽出する手法を提案している。また、ブログ以外のソースから関連画像を取得し、情報推薦の拡張を行っている。これらの研究は、主に旅行ブログを情報源とした分析を行うことで地域性の推定や、地域特有の情報の抽出を行っている。本研究では、旅行ブログと特許の重ね合わせによって特許からの地域性の検出を試みている点で異なる。

岡本ら[8]は「東京」や「秋葉原」などの地名をクエリとして、地名を含むブログエントリを収集し、それらを情報源として利用している。同じ話題の代表語が別地域でも多く抽出されるかどうかで地域性を判定し、地域特有のイベント情報を抽出する手法を提案している。こちらの研究が地域名を入力として、旅行ブログを情報源に地域特有の単語を抽出して地域性を検出しているのに対して、本研究では、任意のキーワードを入力として、特許と旅行ブログの2つの情報源から地域性を検出している。

#### 4. 旅行ブログと特許からの地域性の検出

本研究では、2種類の方法で特許から地域性を検出することを試みる。以下、4.1節および4.2節で、それぞれの手法について述べる。

##### 4.1. 旅行ブログと特許の地域性の視覚化

特許の地域性を視覚的に捉えるため、入力された検索クエリと適合度の高い特許と旅行ブログそれぞれについて、地図上にマッピングするシステムを構築した。図3を用いてマッピングシステムの概要を説明する。このシステムは、入力された検索クエリと適合度の高い上位n件の特許と旅行ブログを検索し、その頻度の高い地域から低い地域までを、赤から青のグラデーションでプロットすることで、入力された検索クエリに関する地域性を視覚的に確認することができる。

図3中の検索システムのコーパスについて説明する。特許検索システムにおいては、特許のタイトルとアブストラクトを抽出したものを使用している。旅行ブログ検索システムでは、ブログエントリのタイトルと本文を使用している。マッピングに用いる位置情報は、特許の開発者（複数人いる場合は最初のひとり）の住所と、ブログ投稿時にユーザーが指定する地域カ

テゴリをそれぞれ抽出し、緯度経度情報に変換することで得ている。これらの情報を利用し、検索クエリに対する地域ごとの特許の出願頻度とブログの投稿頻度を算出し、それに対応する色で地図上にプロットしている。

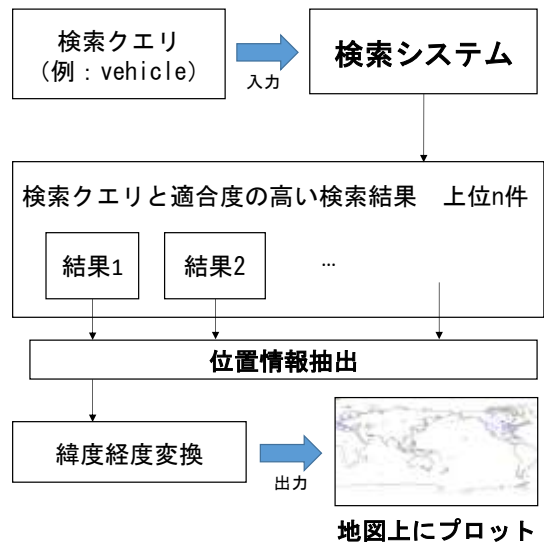


図3: マッピングシステムの概要

このシステムによって世界中の地域において、特許とブログの両者を、2節の図2で示したような同様の方法でマッピングすることができるため、同じ検索クエリを使用して得られた2つのシステムの出力結果を比較すれば、各地域の産業における特色と観光を検出・発見することができる。

##### 4.2. 地域性の高い特許の検出

入力されたテキスト（語句）に対して、関連性のある地域を推定する地域推定システムを開発し、そのシステムを利用した、地域性の高い特許の検出手法を提案する。旅行ブログサイトのひとつである TravelBlog ([www.travelblog.org](http://www.travelblog.org))は、投稿時にユーザーが各エントリの地域を選択するが、この情報を用い、k近傍法によって、入力文書と関連性の高い地域を推定するシステムを構築する。システムの概要について図4を用いて説明する。このシステムは、まず、入力にテキスト（語句）を使用し、旅行ブログ検索システムにより適合度の高いブログエントリを検索する。次に、検索結果の適合度をエントリに付与されている位置情報ごとに集計し、合計値の高い順に位置情報を出力することで、入力されたテキストや語句の関連地域を推定することができる。



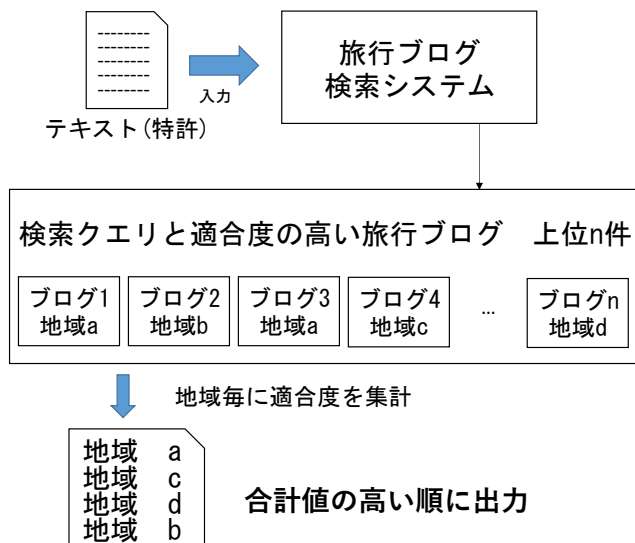


図 4：地域推定システムの概要

2 節で示した，入力として“okonomiyaki”（お好み焼き）と“pasta”を入力したときの出力結果例（表 1）を見てみると，入力に対して正しい関連地域が推定できていることが分かる．また，地域性の大きさについて，“okonomiyaki”の推定結果のように，日本国内でも，“広島”，“大阪”，“東京”のように，小さな地域性が複数存在している場合と，“pasta”のように，イタリア全体が大きな 1 つの地域性を有している場合があることが確認できた．

このシステムに，特許のタイトルとアブストラクトを入力すれば，その特許の関連地域を推定することができる．入力した特許（の開発者の住所）と推定地域の距離が近ければ，その特許は地域性が高いことが期待できる．

## 5. 実験

本節では，地域推定システムを用いた提案手法の有効性を示すために行った実験と，実験結果からの考察について述べる．5.1 節で実験方法について，5.2 節で実験結果と考察について述べる．

### 5.1. 実験条件

本研究では特許のタイトルとアブストラクトを 4.2 節で述べた地域推定システムの入力として使用し，地域性の高い特許を検出する手法を提案した．本節では，この手法の有効性について検証した実験について述べる．実験には，国際特許データベース（WO）に公開されている特許と，TravelBlog に登録されている旅行ブログエントリを使用した．それぞれのデータについて以下より説明する．

- 国際特許データベース（WO）

PCT 条約に基づいて，同じ形式で国際出願された特許の集合である．世界の主要な国の特許庁で受理された出願を，先行技術文献を調査する目的で公開してい

る．1998 年から 2013 年までの間に公開された，168,045 件を実験に利用する．

- TravelBlog

旅行ブログエントリを主体とした海外の Web サイトであり，世界各国に関する旅行ブログエントリが写真とともに掲載されている．また，ユーザーが旅行ブログエントリを投稿する際，どこに関する旅行ブログエントリであるかを「大陸」，「国」，「都市」の階層であらかじめ決めて投稿する仕様となっている．478,034 件を実験に利用する．

4.2 節でも述べたように，地域推定システムに国際特許のタイトルとアブストラクトを入力し，TravelBlog を利用して，関連地域を推定したとき，特許と推定地域の距離が近いとき，特許の地域性が高いことが期待される．よって，特許と推定地域の距離が一定値以内の組み合わせを抽出し，分析を行う．

### 5.2. 実験結果と考察

国際特許 168,045 件の関連地域を地域推定システムによって推定した結果，特許と推定地域の距離が一定値以内だった組み合わせの件数を表 2 に示す．

表 2：特許と推定地域の距離が近い組み合わせの件数

距離	件数
約 100km 以内	522
約 200km 以内	728
約 300km 以内	1053
約 400km 以内	1609
約 500km 以内	2314

表 2 の結果の内，距離が 100km 以内だった組み合わせについて分析を行った．分析によって実際に地域性が高いと判断できた特許の割合は 0.180 だった．低い結果になったものの，今後さらに分析を進めるためのデータを多く収集でき，いくつかの傾向を確認することができた．以下でそれらの傾向について説明する．

- 地域性の高い特許

2 節で述べた CERN の例のように，ある分野において中核を担っている企業が存在する地域と，その分野における特許の組み合わせが地域性の高いと判断できるものが多かった．また，ひとつの地域で同じ国際特許分類の特許が複数出願されている場合，その特許は地域性が高いことが多いことも分かった．シリコンバレーなどの工業地帯におけるコンピューター関連の特許がそれに該当する．

- 地域性の低い特許

特許と推定地域の距離は近いが，地域性は低いと判断された特許には，各国の主要都市や大都市に推定されているものが多かった．これは，国際出願をする企

業や団体の存在する地域や、観光先として選ばれる地域が、比較的大都市に多いことが原因だと考えられる。また、医療分野に関する研究は、様々な国と地域で進められており、国際特許分類によって医療分野に分類されている特許については、地域性を検出することが難しいと思われる。分析を行った 522 件の特許の内、地域性が低いと判断されたもの 428 件、国際特許分類によって医療分野に分類された特許は 71 件だった。地域性の低い特許の約 16.6%は医療分野のものであることがわかった。

これらの傾向を参考に、地域推定を行う際に、使用する旅行ブログエントリに付与されている地域カテゴリの偏りの平滑化や、国際特許分類を用いた、入力に使用する特許の選別を行うことで、地域性の高い特許の抽出精度を上げることができると思われる。

## 6. おわりに

本研究では、特許と旅行ブログエントリそれぞれについて検索システムを構築し、両者の地域性を視覚的に捉えるために、地図上へのマッピングを行った。さらに、テキストから関連地域を推定する分類器を開発し、そのシステムに特許のタイトルとアブストラクトを入力することで、地域性の高い特許を検出する手法を提案し、特許の地域性検出に関する多くのデータを収集し、いくつかの傾向を確認することで、地域性の高い特許が検出できることを確認した。

## 7. 今後の課題

より多くの地域性の高い特許を検出するためには、ブログエントリの解析や、検索システムの改良が必要になると考えられる。本研究では、1 件の特許に対しての最上位の関連地域のみを考慮し、地域性の検出を試みたが、いくつかの候補の中から関連地域を推定することでより多くの、地域性の高い特許を検出できると考える。また、現在の旅行ブログ検索システムは、検索クエリとして入力した単語との適合度のみを検索に用いているが、それぞれのブログエントリについて、何をしているかという観光タイプや、エントリ内に掲載されている写真を考慮に入れることで、より詳しい地域性を検出できると考えられる。特許検索システムについても同様に、今回の実験では人手によって分析を行ったが、国際特許分類や特許の概要以外の部分を考慮する、検索結果が大都市に偏っている問題を平滑化の処理などによって解決することで、よりローカルな地域性の検出が可能になることが期待できる。

## 謝辞

本研究は総務省による戦略的情報通信研究開発推進制度 (SCOPE) の支援を受けて行われた。

## 参 考 文 献

- [1] 手塚太郎, 近藤浩之, 田中克己. (2008). 混合ガウス分布を用いたウェブコンテンツの地域性推定とオブジェクトレベルローカルサーチ. 情報処理学会論文誌データベース, Vol.1, No.1, pp.13-25.
- [2] 森本泰貴, 藤本典幸, 長屋務, 出原博, 萩原兼一. (2007). Webを対象としたロボット型住所関連情報検索システムの開発. 電子情報通信学会論文誌, D Vol.J90-D, No.2, pp.245-256.
- [3] 相良毅, 有川正俊. (2004). ジオパースによるWebからの空間コンテンツの獲得. 第15回データ工学ワークショップ(DEWS2004), 1-11-01.
- [4] 奥健太, 西崎剛司, 服部文夫. (2012). 地域性限定スコアに基づく位置情報付きコンテンツからの地域限定語句の抽出. 情報処理学会論文誌データベース, Vol.5, No.3, pp.87-116.
- [5] 新田崇人, 難波英嗣, 石野亜耶, 竹澤寿幸. (2016). 外国人旅行者の行動分析および地域性の判定. 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(DEIM).
- [6] Fujii Kazuki, Nanba Hidetsugu, Takezawa Toshiyuki, Ishino Aya. (2016). Travellers' Behaviour Analysis Based on Automatically Identified Attributes from Travel Blog Entries. in Proceedings of Workshop of Artificial Intelligence for Tourism, (PRICAI).
- [7] Qiang Hao, Rui Cai, Changhu Wang, Rong Xiao, Jiang-Ming Yang, Yanwei Pang, Lei Zhang. (2010). Equip Tourists with Knowledge Mind from Travelogues. in Proceedings of the 19<sup>th</sup> International Conference on World Wide Web, (WWW), pp.401-410.
- [8] 岡本昌之, 菊池匡晃. (2010). ブログからの地域イベント情報抽出. 情報処理, Vol.51, No.1, pp.14-17.