

# 自然言語からの緯度経度推定に向けた説明文収集とパターン分析

坂根 和光<sup>†</sup> 山本 岳洋<sup>††</sup> 澤田 祥一<sup>†††</sup> 大塚 一路<sup>†††</sup> 山本 光穂<sup>†††</sup>

大島 裕明<sup>†,††</sup>

<sup>†</sup> 兵庫県立大学 応用情報科学研究科 〒650-0047 兵庫県神戸市中央区港島南町 7-1-28

<sup>††</sup> 兵庫県立大学 社会情報科学部 〒651-2197 神戸市西区学園西町 8-2-1

<sup>†††</sup> 三井住友海上火災保険株式会社 〒101-8011 東京都千代田区神田駿河台 3-9

E-mail: <sup>†</sup>{aa20y505, ohshima}@ai.u-hyogo.ac.jp, <sup>††</sup>t.yamamoto@sis.u-hyogo.ac.jp,

<sup>†††</sup>{shoichi.sawada, kazumichi-ohtsuka, mitsuo.yamamoto}@ms-ins.com

**あらまし** 本研究では、「計算科学センタービルの入り口の自転車置き場の横」といった、自然言語で書かれたある地点についての説明文から、その地点がどこであるかの緯度経度を推定する問題に取り組む。この問題は、町丁名までの住所情報と、数十文字程度からなる説明文が与えられ、それに対して緯度経度を返すというものである。本稿では、まず、本研究で用いるデータとして、ある地点についての説明文をクラウドソーシングを用いて収集した。具体的には、交通事故に関するヒヤリ・ハット事例に基づく274カ所の地点を収集した。さらに、得られた説明文を第三者に提示することで、その説明文から実際の緯度経度を推定可能かに関する調査についても行った。これらの収集された説明文と緯度経度において、位置を推定することに資する情報がどのように出現するかを分析を行う。また、交通事故と関連が深い地点以外の地点に関する説明文についても収集する実験を行った。

**キーワード** ジオコーディング, GIS, 自然言語処理

## 1 はじめに

Google マップや Yahoo!地図といった地図サービス、全地球測位システム (GPS) や WiFi などによる位置情報を推定可能なスマートフォンの普及により、ある地点がどこにあるかを探することは容易になってきている。その一方で、「計算科学センタービルの入り口の自転車置き場の横」のように、テキストあるいは口頭で表現される自然言語により自らの場所や目的地を表したいことは多い。たとえば、屋外で交通事故に合い救急車を呼ぶときや、スマートフォンを持っていない人に道案内するときなど、日常、非日常を問わず自然言語で場所を人に伝える状況は存在する。

本研究の大きな目的は、地点を表す説明文の緯度経度を計算機が自動的に推定するための技術を開発することである。すなわち、「計算科学センタービルの入り口の自転車置き場の横」のような地点を表す説明文を入力として受け取り、緯度 34.653999、経度 135.220103 と出力する技術を実現することである。これが実現できれば、ある人が事故の場所を電話先のオペレータに伝えた際に、オペレータ側の計算機で自動で事故の地点の候補を推測したり、人がロボットに対して場所を口頭で指示する、といったことが可能になる。たとえば、総務省消防庁が刊行した「救急救助の現況」の「事故発生場所別の搬送人員構成比」の道路に注目すると、平成 30 年では 721,778 名の救急救助が発生しており<sup>1</sup>、口頭での説明から精度良く場所を推定する技術は重要であると考えられる。

この目的を実現するため、本研究ではさまざまな地点についてその地点の場所を説明する説明文を収集する。そして、収集した説明文から緯度経度を予測するモデルとしてどのようなモデルが妥当なのかを検証するため、説明文としてどのような単語やフレーズが含まれるか、説明文から人が推定する緯度経度と実際の緯度経度との誤差といったパターンを分析する。

具体的には、まず、交通事故に関するヒヤリ・ハット事例が発生した地点についてその地点の緯度経度と、その場所を説明する説明文を収集するタスクをクラウドソーシングとして実施する。その結果、274カ所の地点に関する説明文を収集した。その後、収集した説明文を与えてそこから緯度経度を推定するタスクを再度クラウドソーシングによって実施した。これにより、収集した説明文に対して、実際の緯度経度と近い説明文とそうでない説明文を分析することができる。本稿では、収集した274カ所地点に関する説明文と、その説明文から推測された緯度経度の基礎的分析を行い、説明文から緯度経度を正確に推測するために必要な説明文の特徴や推定に際して考慮すべき特徴の分析を行った。

また、交通事故と関連が深い地点以外の地点に関する説明文についても収集する実験を行った。具体的には、兵庫県神戸市中央区港島中町7丁目を対象に、その範囲にある著者らが指定した歩道の141カ所の地点に関する説明文を収集した。

## 2 関連研究

テキスト情報と地理情報に関連する研究として、これまで様々な問題が取り組まれてきた。

<sup>1</sup> : <https://www.fdma.go.jp/publication/>

本研究でも対象としているジオコーディングとは、様々な地理情報が与えられたときに、緯度経度を特定することである。たとえば、テキスト情報に含まれる住所やランドマークなどを手がかりとして、緯度経度の特定を行う。相良ら[20]は、住所情報が完全でない場合などでも、部分マッチングを行うことによって、ジオコーディングを行える手法を提案した。細川ら[11]は、文書に含まれるランドマーク情報などから緯度経度を特定する手法を提案した。文書に含まれる情報との類似性を計算することで、再現率の高いジオコーディングの実現を目指している。Nesi ら[3]は、ウェブページの中に出現するテキストからジオコーディングを行う手法を提案している。その際、自然言語処理技術や、機械学習技術を用いるとともに、ウェブホストの地理情報や、IP アドレスやドメイン名の登録者などを利用することで、ジオコーディングの精度を上げることを試みた。野秋ら[26]は、歩道ネットワークデータベースを用いて、文書データから経路記述のジオコーディングの基本的な枠組みを提案した。南ら[25]は、場所を示す表現か道路名のどちらかの情報が不十分な場合でも、相互補完し、場所の特定を支援できる、路線ジオコーディングを行った。小林ら[15]は、線路や道路、区間などの道路の情報を入力とし、それを構造化して座標と一体化したデータを整備し、様々な目印の座標との関係から道路上の地点を出力する道路ジオコーダーの開発を行った。有川ら[27]は、歩道ネットワークデータベースを用いて、経路記述の解析を行い、その結果をもとにしてジオコーディングを行うシステムを開発した。

文書に含まれる地理情報を抽出し、それをもとにして様々な地理情報サービスを行う研究も数多く行われている。今井ら[10]らは、目的の POI 周辺の道のりに詳しくない人に向けて、対象の POI の大まかな位置を出すことを目指した。中村ら[23]は、ウェブページの URL と HTML 構造の分析を行い、電子掲示板の犯行予告と判定された文書から、文字列を抽出し、GeocoderAPI にかけて緯度経度を出すことで、文書と地理情報との関連を明らかにする手法を提案した。Woodruff ら[5]は、テキストから地名を含む単語を自動的に抽出して、地理参照文書の索引・検索を支援するための GIPSY システムを提案した。Nesi ら[2]は、ウェブページに関連する地理情報を取得する手法を提案した。ウェブページのテキストデータから、組織の住所などの場所に関連する情報を抽出し、それらの緯度経度を求めるということを行った。河野ら[8]は、ツイート文に含まれている地名だけではなく、都道府県ごとの独自の名物や、イベントから都道府県の推定に取り組んだ。奥村ら[6]は、全ての都道府県から各々1000人分の500個のツイート文を学習させ、固有名詞抽出を行った後、Doc2Vecを用いて特徴抽出を行い、ツイート文から都道府県単位のジオコーディングを試みた。河野ら[7]は、ツイート文に含まれる駅名のような位置情報を含む語と、投稿時間の間隔を利用し、ツイート文が投稿された位置を推定し、候補の中からスコアが最大の地点を正解と仮定し、位置情報の推定を試みた。原ら[9]は、災害について投稿しているツイートにおける緯度経度情報を取得する手法を提案した。字や町名などのキーとなる情報から災害に関連するツイ

トを取得し、自然言語処理などを用いて地理情報を取得した。Panasyuk[4]は、Twitter ユーザのユーザのプロフィール情報から、そのユーザに関連する地理情報を取得する研究を行った。

新垣ら[16]は、容易に目的地にたどり着けるように効率的なナビゲーションシステムの提案を行った。その過程の実験結果で、近くにある説明しやすい地点を現在地であるとして説明してしまうことがあるということを報告している。徐[14]は、道案内における違いについて調査した。まず、被験者の道案内のタスクを与え、被験者が行った道案内の音声データを収集した。その結果、道案内には教会のような目印となる建物を利用する人が多いことがわかった。森永ら[17]は、従来の郵便局のような局所的なランドマークと、電波塔のような広域的なランドマークに加え、電車通りのような線形的なランドマークを利用した、ルート・ナビゲーション・システムを提案した。杉谷ら[18]は、地名を含むツイートに対し、文字数などの特徴を用いることで、SVM(Support Vector Machine)によってユーザ位置の推定を行った。その際、位置情報付きツイートを利用して、機械的に正解か不正解のラベルを付与し、学習用データを大量に作成した。Moncla ら[1]は、地名集に見られる地名を明確にすると同時に、地名集のリソースに見られない他の細粒地名の空間フットプリントを推定するために、前置詞や動詞の存在などの言語的特徴に基づいたクラスタリング技術を用いた教師なしジオコーディングアルゴリズムを提案している。相良ら[19]は位置情報の情報を収集するために、電話帳を店舗データベースとして利用することで実世界に存在する店舗の住所情報を Web から収集する手法を提案している。谷[22]は Google Maps API を利用して施設の名称や、住所を入力して、その場所の緯度経度情報を出しており、それをウェブサイト上でユーザが利用できるようにしている。増山[21]は Google Maps Distance Matrix API あるいは OpenStreetMap データと QGIS の組み合わせによって、徒歩あるいは自家用車による移動距離・時間を計算する方法を示し、それぞれの問題点を指摘するとともに、その精度を評価している。三浦[12]は実データを用いて2地点間の距離を計算を行う場合、地球を完全な球体と仮定し計算する、回転楕円体と仮定した測地線距離の計算を行う、ヒュベニの公式を用いる3つの手法を紹介し、誤差比率でそれぞれを評価した。長谷川ら[24]は経路と交差点の形状に応じ、アンケート結果をもとに、目につきやすいランドマークを利用して、道案内文を生成する手法を提案している。児島ら[13]は携帯電話の GPS から大まかな位置情報を取得した後、近くの道路標識とランドマークを入力してもらうことで位置情報を推定する手法を提案している。

### 3 交通事故に関するヒヤリ・ハット事例に基づくデータセット構築

本節では、説明文からの緯度経度推定のためのデータセット収集について説明する。本研究では、説明文と緯度経度のペアを収集するためにクラウドソーシングを用いた。このクラウドソーシングの目的は、ある地点を説明した説明文およびその緯

度経度を収集し、また、得られた説明文からどの程度その緯度経度を推定可能かを分析することである。

まず、本研究で実施したクラウドソーシングの概要について説明し、その後実際のタスクの詳細を述べる。

### 3.1 交通に関するヒヤリ・ハット事例に基づく説明文収集

さまざまな地点に関する説明文を収集するため、また、「計算科学センタービル」のようなランドマークだけではなく、「計算科学センタービルの入り口と逆側の近く、医療センター方向に向かう交差点を渡る手前の地点」のような、より細かな粒度の地点を説明した説明文を収集する必要がある。このような細かい粒度の地点に関する説明文および緯度経度を多数収集するため、本研究では、交通事故や交通事故になりかけた（以下、交通ヒヤリ・ハットと呼ぶ）地点に関する説明をクラウドソーシングのワークに回答してもらうことを行った。

#### 3.1.1 タスク実施手順

具体的なタスクの実施手順は以下の通りである。まず、ワークに以下の概要文を見せた。

本タスクは、回答者ご自身が実際に見られた、もしくは、体験された、交通事故や、もしかしたら交通事故になっていたかもしれない事案（交通ヒヤリハット事案）が発生した地点についてご回答いただくものです。ここで、交通ヒヤリハット事案とは、たとえば、自転車で車道を走っていて車に幅寄せされてヒヤッとしたということや、自動車で走っていたらだいたい先の方で歩行者が赤信号を無視して横断していたということや、歩道を歩いていて他の人とぶつかりそうになったということまで、様々な事案まで含むと考えてください。ご自身が実際に見られた、もしくは、体験されたもののみを対象とし、たとえば、他人から聞いたことや、ニュースで見たことは対象外とします。

その後、以下の2つの設問により、ワークが選んだ交通ヒヤリ・ハット地点に関する緯度経度と地点の説明文を回答してもらった。緯度経度については、Google マップを使用した緯度経度確認方法<sup>2</sup>をワークに提示し、その方法で緯度経度を数値として回答してもらった。説明文に関しては以下の設問をワークに提示し、回答してもらった。

その地点を特定するような説明文は、その地点を特定するための説明を日本語の文章で書いてください。他の人（救急車や友人など）にそこに来てもらうために、電話越しに口頭で説明するという状況を想定してご回答ください。その際、その地点の町名や丁目まではすでに特定されているという前提で説明してください。そして、説明文では、緯度経度と住所を使わないようにしてください。交通事故・交通ヒヤリハット事案そのものの説明や、日時の説明、その他、個人が特定できるような情報は一切書かないようにしてください。

#### 3.1.2 タスクの実施と収集したデータ

クラウドソーシングサービスの1つであるランサーズ<sup>3</sup>を用いて3.1.1節で述べたタスクを2020年8月21日から11月21日にかけて3回にわけて実施した。第1回と第3回については、1回のタスクあたり3カ所の地点について緯度経度と説明文を回答してもらった。また、第2回については4カ所の地点について回答してもらった。

クラウドソーシングを実施し、274カ所の説明文とその緯度経度の組を収集した。表1の「説明文」と「緯度経度」の列に実際に得られた説明文と緯度経度の例を示す。表にあるように、「入り口付近」や「駐車場を渡る信号部分」のような、単なるランドマークではないより細かな粒度の説明文が得られていることが分かる。

### 3.2 人手による説明文からの緯度経度の推定

本研究の目的の1つは、地点に関する説明文中にどのようなフレーズや単語が出現していれば、その説明文の地点を表す緯度経度（以降、真の緯度経度と呼ぶ）を推定できるのかの知見を得ることである。そこで、前節で収集した説明文を第三者のワークに提示し、そのワークに説明文が指し示していると思われる緯度経度を収集するタスクを実施した。

#### 3.2.1 タスク実施手順

具体的なタスクの手順は以下の通りである。まず、ワークに以下の概要文を提示した。

本タスクは、場所について述べた説明文を見て、その説明文が指し示す地点を特定して回答していただくものです。ある程度の住所の情報は与えられています。その住所の中で、説明文に該当する地点を、Google マップを開いて探していただき、緯度経度を取得していただいて、回答していただきます。

次に、3.1節で収集した説明文を1つ提示し、用意したGoogle マップのリンクを開き、説明文が指し示していると思われる緯度経度を回答してもらった。このとき、ワークに緯度経度を推定する地点の範囲が分かるように、Google マップの初期状態として、説明文の真の緯度経度に対して、リバーズジオコーディングを行い、得られた住所に関する字（あざ）の範囲を、字が存在しない場合は、大字の範囲をハイライト表示した。リバーズジオコーディングにはYahoo!リバーズジオコーダ API<sup>4</sup>を利用した。たとえば、緯度34.689246、経度135.501673であれば、得られる町丁名は「大阪府大阪市中央区伏見町3丁目」となり、字までの情報が得られ、緯度34.924161、経度135.710621であれば、得られる町名は「京都府長岡京市神足」となり、大字までの情報を得る。また、リバーズジオコーディングで得られた範囲に真の緯度経度が含まれていない場合は、著者が住所を直接修正し、字や大字、あるいはそれに近い範囲の領域がハイライトされるハイライト表示の範囲を修正した。

3 : <https://lancers.jp>

4 : <https://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/reversegeocoder.html>

2 : <https://ohshimalab.github.io/coordinate/>

表 1 交通ヒヤリ・ハット事例に基づくクラウドソーシングで収集した説明文と緯度経度の例。

説明文	説明文が示す緯度経度		説明文から人手で推測された緯度経度	
	緯度	経度	緯度	経度
東邦銀行本店営業部の到岸寺の中間の細い道。並木通りの入り口付近	37.75403	140.467499	37.754095	140.466836
環七の方から大森駅に向かって歩くととても細い道と別の道が交差する地点	35.5822311	139.7250052	35.580913	139.726834
御堂筋沿いの備後町3の交差点でデイルーヤマザキと駐車場を渡る信号部分	34.685295	135.500975	34.685273	135.500937

### 3.2.2 タスク実施と収集したデータ

今述べたタスクを2021年1月24日にランサーズを用いて実施した。具体的には、3.1節で3回にわたり収集した274件の説明文に対して、説明文からの緯度経度の推定のタスクを実施した。後の分析で、説明文から推測される緯度経度のワークごとのばらつきを分析するために、1地点につき3名のワークが緯度経度の回答を行った。また、1つのタスクにつき1名のワークは4件の説明文に対する緯度経度を回答してもらった。また、1名のワークがタスクに参加できるのは4回までとした。

表1の「説明文から人手で推測された緯度経度」列に得られた緯度経度の例を示す。本研究では、このようにして得られた、ある地点の説明文とその緯度経度、そして説明文から推測された緯度経度のデータを分析することで、精度良く緯度経度を推測するために必要な言語的特徴や良い説明文を作るための特徴などを明らかにする。

## 4 説明文と説明文から推測される緯度経度のパターン分析

本研究の大きな目的は、説明文と町丁名までの住所情報が与えられた際に、その説明文が示す地点の緯度経度を推定するモデルを作成することである。そのために、本稿では3節で収集した説明文およびその説明文から人が推測した緯度経度のデータを分析することで、人手による緯度経度の推定に有用な情報として、どのような言語的特徴が説明文に表れるのかの基礎的分析を行った。この分析を進めることで、多くの人にとって正確な場所が推測可能な説明文の特徴や、説明文から緯度経度を推定するモデルを構築する際に考慮すべき特徴などを明らかにしていくことができると考えている。

### 4.1 誤差の計算方法

分析にあたり説明文が示す真の緯度経度と、説明文からワークが推定した緯度経度との誤差を求めた。誤差は、地球の楕円体モデルの表面上での最短距離である測地線距離を2地点間で算出した。具体的には、Pythonのジオコーディングライブラリgeopy<sup>5</sup>が提供する楕円体上の2点間の距離計算法Karney手法を用いた。測地系は日本の法律でも採用され、多くの国が基準としているGRS80を用いた。また、算出する数値の単位はメートルとした。

本研究では、1つの説明文に対して3名のワークが緯度経度を推定している。3名のワークそれぞれについて真の緯度経度との距離を測定し、その平均を算出することで、説明文と説明

文から推測される緯度経度の誤差を計算した。この値が小さいほど、人間にとって説明文から真の緯度経度を推定することが容易であると考えられる。なお、図3はその平均と標準偏差をグラフで表したものである。

### 4.2 真の緯度経度と説明文と誤差の関係

4.1節で述べた方法を用いて収集した説明文に対して真の緯度経度との誤差を求めた。表2に真の緯度経度との誤差が小さかった説明文上位5件を、表3に誤差の大きかった説明文上位5件を示す。

まず、表2に示した誤差の小さい説明文について分析する。この5つの説明文を見てみると、いずれの説明文もランドマークとなる、「深田内科」や「阪急電車の踏切」のような表現が出現しており、そのランドマークと説明している地点の位置関係が、「目の前」や「すぐの」のような、ランドマークからあまり移動しない表現が出現していることが分かる。また、移動する場合でも、「北」のような方角情報と、「橋」や「踏切」など、どこを通るのかを明確に説明している。このように、誤差の小さい説明文は、始点が明確で、真の地点への進行方向が明記しているなどの特徴があると考えられる。

次に、表3に示した真の緯度経度との誤差が大きい説明文について分析する。この5つの説明文を見てみると、ランドマークとなる表現があまり使用されていないことが分かる。また、使用されているランドマークも、「鉄工所」や「中古車販売店」のような地図で見ると一目ではどこか分からないランドマークが出現している。このような地図で見ると分からない表現はランドマーク以外にも出現している。たとえば、「夜はかなり暗く」や「傾斜15度以上の急傾斜」は、普段その道を通っている人は印象に残っていて、他者に説明するときも特徴として使用することはあり得るが、説明している地点の周辺に詳しくない人で、なおかつ、地図を見ても分かりにくい特徴は場所の推定が困難であることが分かる。さらに、「運転が危ない」や「ミッション車」のような説明している人の感想が出現しており、説明している当事者以外には真の緯度経度の推定が困難な分かりにくい余分な情報が含まれた説明文になっている。

また、同じ説明文からおのおのが推定した地点の個人差を調べるため、5番目に誤差が大きかった説明文「国道41号と国道179号が交わる交差点の南側、41号沿いにある、中古車販売店手前で細い左側の道へと枝分かれする地点」に対して3名のワークがどのような地点を緯度経度として回答したのかを見てみた。3名中2名のワークは、図1にある真の緯度経度と近い緯度経度を推定し回答したが、1名のワークが図2を見れば分かるように、図2にある2つのピンのうち、真の緯度経度

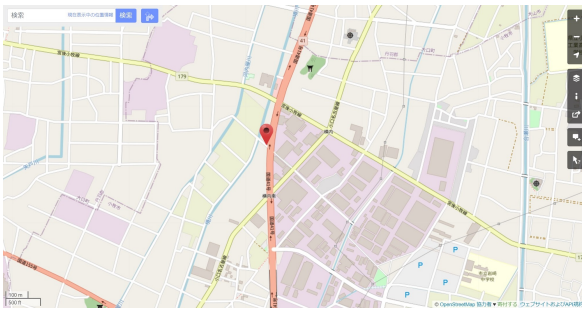
5: <https://github.com/geopy/geopy>

表 2 真の緯度経度と説明文から 3 名のワーカが推定した緯度経度とのずれが小さい説明文.

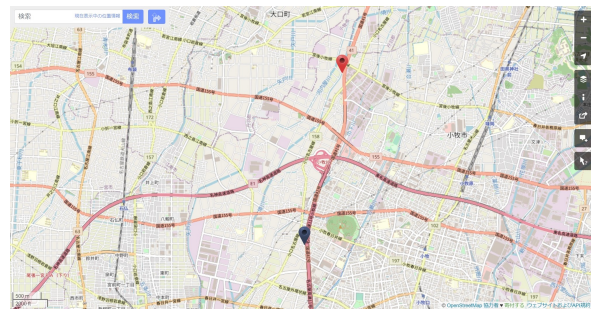
説明文	真の緯度経度からのずれの平均 (m)
深田内科、すずらん薬局、岡山交通の目の前の交差点	0.987
仲の池緑地の南にある阪急電車の踏切を北側に渡ってすぐ、線路沿いの道と交差するところ。	1.599
土屋金物店前の交差点地点	1.692
中部中学校に隣接する母子寮と王子製紙の北側にある地藏川をまたぐ橋を越えてすぐの十字路。眼鏡屋の目の前の交差点。	1.752
久方中学グラウンド側の交差点、戸笠公園寄り。	1.857

表 3 真の緯度経度と説明文から 3 名のワーカが推定した緯度経度とのずれが大きい説明文.

説明文	真の緯度経度からのずれの平均 (m)	重心からの距離の平均 (m)
有馬街道の神戸三田線と、六甲トンネルから下ってきた車道との合流地点。	1660.545	1383.733
急なカーブが続く場所でこの近辺にはラブホテルがあったり鉄工所があり夜はかなり暗くて運転が危ない場所です。	1498.843	1946.437
松江市国道 485 線からだんだん道路に乗る分岐点	1341.716	784.845
宜野湾市と浦添市の市境の地点。ミッション車はもちろん、オートマチック車でかなり下がる傾斜 15 度以上の急傾斜。	1068.324	1255.954
国道 4 1 号と国道 1 7 9 号が交わる交差点の南側、4 1 号沿いにある、中古車販売店手前で細い左側の道へと枝分かれする地点	1008.519	1328.210

図 1 差異が 5 番目に大きい説明文での真の緯度経度の地点  
地図画像は OpenStreetMap より作成

©OpenStreetMap contributors, Base map and data from OpenStreetMap and OpenStreetMap Foundation/CC BY-SA 2.0 <sup>6</sup>.

図 2 図 1 を縮小し真の位置と最も離れた推定地点を示した図  
地図画像は OpenStreetMap より作成

©OpenStreetMap contributors, Base map and data from OpenStreetMap and OpenStreetMap Foundation/CC BY-SA 2.0 <sup>7</sup>.

の地点を示している北の方にあるピンよりもかなりの距離が離れた南の地点を回答しており、誤差が大きくなっていった。このワーカが推定した地点には、説明文にもある「中古車販売店」が近くにあり、なおかつ、中古車以外にも自動車の販売店や自転車の販売店が存在しており、「細い左側の道へと枝分かれする地点」を推定し回答していた。すなわち、このワーカは説明文から緯度経度を推定しようとしているものの、固有名詞ではないランドマークを使用した説明文は、推定するとき基準となる点が明確ではないといった理由により、この説明文が示す地点とは離れた地点を回答してしまったのではないかと考えられる。

#### 4.3 3 名のワーカの回答のばらつき

説明文から推定される地点の傾向を分析するため、3 名のワーカの回答のばらつきと、真の緯度経度と 4.1 節で計算される誤差との関係性を分析した。

3 名のワーカの回答のばらつきの計算は以下の方法で行った。まず、3 名のワーカが回答した 3 つの緯度経度の重心を求める。

ここでは、簡単のため緯度、経度それぞれについて平均を求めることで重心を求めた。具体的には、3 名のワーカの回答の緯度経度を  $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$  とすると、以下の式で重心  $(x_a, y_a)$  を求める。

$$(x_a, y_a) = \left( \frac{x_1 + x_2 + x_3}{3}, \frac{y_1 + y_2 + y_3}{3} \right)$$

ここで、 $(x_i, y_i)$  は 1 名のワーカが回答した (緯度, 経度) の組を表す。その後、この重心  $(x_a, y_a)$  とそれぞれのワーカの回答との測地線距離を 4.1 節で述べた方法で測定し、平均を求める。重心と 3 名のワーカが回答した地点の測地線距離の平均が小さければ、その説明文はどのワーカも近い地点を回答したような説明文であり、大きければワーカによって推定する地点がさまざまであるような説明文であると考えられる。

表 3 では、上記で説明した重心からの距離の平均も掲載している。この表から分かるように、重心からの距離の平均値は、4 つの説明文で 1 キロメートルを超えていることが分かる。このことは、真の緯度経度からのずれが大きい説明文から地点を推定した場合、3 名全員が同じような地点を推定するということではなく、各人が異なる地点と考えたということが予想される。ただし、真の緯度経度からのずれよりも、重心からの距離の方

6 : <https://www.openstreetmap.org/copyright>

7 : <https://www.openstreetmap.org/copyright>



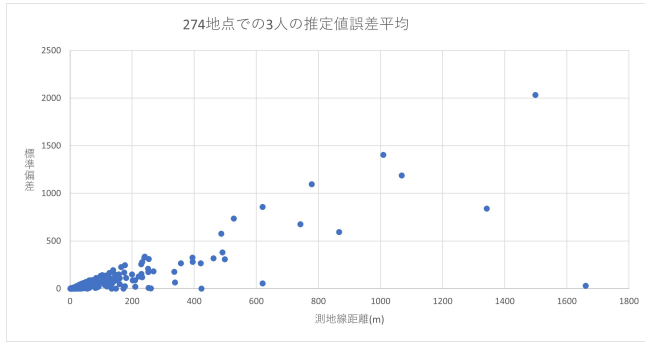


図 3 真の位置と推定位置のずれを視覚化.

が大きい場合には、3名のワーカのうち2名は真の緯度経度に比較的近い地点を回答したが、1名のみ大きく異なる地点を回答したという場合であると考えられる。具体的には、「急なカーブが続く場所…」という説明文や、「国道41号と国道179号が…」という説明文がそれに該当する。逆に、重心からの距離の方が小さい説明文では、3名のワーカのおおのが別々の地点を回答したと考えられる。たとえば、「松江市国道485線からだんだん…」という説明文3人のワーカが別々の地点を回答していた。

人手による緯度経度の推定に有用な情報として、どのような言語的特徴が説明文に表れるのかの分析を行った結果は、固有名詞で書かれたランドマークとなる建物が説明文に使用されており、なおかつ、ランドマークの周辺の地点が真の緯度経度だと、推定が容易となる。その一方で、固有名詞で書かれたランドマークとなる建物が出現せず、地図を見るだけでは確認ができない情報が出現している説明文は、人手による緯度経度の推定は困難となる。

## 5 特定の範囲の地点に基づくデータセット構築

本稿では3節で述べた方法でさまざまな地点の説明文と緯度経度を収集し、4節では説明文とそこから推測される緯度経度の関係についての基礎的な分析を行った。一方、今回実施した、地点および説明文の収集方法は得られる地点に偏りがあると考えられる。具体的には、交通ヒヤリ・ハット事例を対象にデータを収集しているため、表1に示すように、交差点や信号など、交通事故と関連が深い地点に関する説明文が多く得られる。また、3.1節で行ったタスクでは、我々から地点を与えるのではなく、ワーカに交通ヒヤリ・ハット事例としての地点をあげてもらった。これは、地点として説明しやすい地点に関する説明文だけを収集している可能性が考えられる。

そこで、今後本研究を進めるにあたり、前節の方法で収集する説明文に加え、異なる性質の地点に関する説明文についても収集する。具体的には、特定の範囲にある地点を網羅的に選り各地点の説明文を収集する、という方法によるデータセット構築を行う。3節の交通ヒヤリ・ハット事例を用いたデータセット収集と異なる点は、我々から地点をワーカに与える点、また、網羅的に地点を選りため、交通事故が起こりやすい場所という偏りを低減できるという点である。

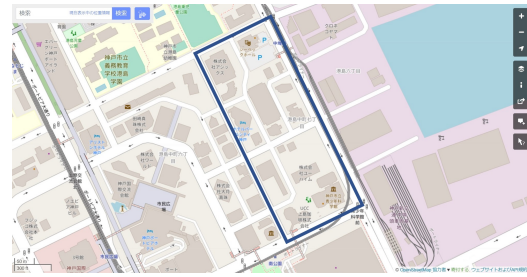


図 4 説明文収集対象となる港島中町7丁目の範囲

地図画像は OpenStreetMap より作成

©OpenStreetMap contributors, Base map and data from OpenStreetMap and OpenStreetMap Foundation/CC BY-SA 2.0 <sup>8</sup>.



図 5 図4の一部の範囲を拡大して説明文収集地点の一部を示した図

地図画像は OpenStreetMap より作成

©OpenStreetMap contributors, Base map and data from OpenStreetMap and OpenStreetMap Foundation/CC BY-SA 2.0 <sup>9</sup>.



図 6 図5の1カ所目での360度写真の北向き。

### 5.1 対象とした範囲

著者が所属する兵庫県立大学情報科学キャンパスの近くである、兵庫県神戸市中央区港島中町7丁目を対象に、その範囲にある歩道141カ所の地点に関する説明文を収集する。図4の枠は説明文収集の対象とする港島中町7丁目の範囲を示したものである。また、図4の一部の範囲を拡大した図5に示し、説明文の収集対象となる141カ所の地点のうち9カ所目までを示す。そして、図5の1の地点に立ち、北を向いたときの景色を図6に示す。

### 5.2 各地点周辺の写真撮影

この141カ所を対象に説明文のタスクを実施するために、そ

8 : <https://www.openstreetmap.org/copyright>

9 : <https://www.openstreetmap.org/copyright>

表 4 神戸市中央区港島中町 7 丁目を対象としてクラウドソーシングで収集した説明文の例.

図 5 の位置	説明文
1	UCC コーヒーアカデミー神戸校の道路を挟んで北側にある駐車場の南南東側の歩道の地点
2	UCC コーヒーアカデミー神戸校の北西側にあるボールが建ち並んでいる歩道の地点
3	UCC コーヒーアカデミー神戸校の道路を挟んで北側にある駐車場の南南東側の歩道から少し北東に進んで、歩道が左に曲がり始めた地点

の地点が分かるようその地点周辺の写真が必要である。そこで、2020 年 11 月 22 日の午前 9 時から午後 3 時にかけて、以下に示す手順で 141 カ所周辺の写真を撮影した。なお、撮影にあたり、道路使用許可申請書を兵庫県神戸市水上警察署に事前に提出し認可を得た。

実際の撮影は、各地点の歩道上に一脚を立て、そこに 360 度カメラである RICOH THETA S を設置する。そして、撮影者が写らないよう撮影者自身は遠方に退避し、スマートフォンから RICOH THETA S の操作を行い撮影を行った。

### 5.3 得られた写真を利用した説明文収集

141 地点の 360 度写真を用いて、前節と同様の説明文収集タスクを実施した。具体的には、得られた写真 Google Street View にアップロードし、写真データを 1 つずつ確認し、人物の顔や、ナンバープレートに、ぼかしを入れるなど適切な処理を起こったうえで、各地点の様子を Google Street View を通じてワークに提示し、同時に、回答してもらう各地点の位置が分かる地図を提示し、1 回のタスクあたり 3 カ所の地点について説明文を記入して回答してもらう、というタスクを実施した。

#### 5.3.1 タスク実施手順

具体的なタスクの実施手順は以下の通りである。まず、ワークに以下の概要文を見せた。

本タスクでは、こちらで指定した地点について、その地点を示す地図と、その地点で撮影された 360 度写真を見ていただき、その地点を特定するような説明文を作成していただきます。指定される地点は 3 カ所あり、それぞれにおいて説明文の作成を行っていただきます。

その後、以下の注意事項を見せ、兵庫県神戸市中央区港島中町 7 丁目の我々が指定した地点に対しての、説明文を回答してもらった。

- 他の人（救急車や友人など）にピンポイントにその点に来てもらうために、電話越しに口頭で説明するという状況を想定してご回答下さい。
- 町名や丁目まではすでに特定されているという前提で説明して下さい。
- 作成していただく説明文では、緯度経度や、住所を使わないようにして下さい。
- 町名と丁目が分かっている、その説明文を見れば、誰にでも、誤差なく、ピンポイントにその地点を特定できるような説明をしてください。
- 指定された 3 カ所は全て異なる地点ですので、必ず、異なる説明文を作成して下さい。
- 作成していただく説明文では、他の指定された地点や、

その説明文を参照することはせず、それぞれ独立に回答して下さい。例えば、「1 カ所目から北に 20m ほど先の地点」のような回答はしないようにして下さい。

- 作成していただく説明文では、自動車や人のような短時間で位置が変化すると想定されるものを参照しないで下さい。例えば、「赤い車のすぐ横」のような回答はしないようにして下さい。

- 地図は、地図上にアイコンを合わせて、マウスを上下左右にドラッグすることで、地図を動かすことができます。また、マウスのホイールボタンや、画面左下の＋、－ボタンで地図の拡大、縮小ができます。

- 360 度写真では、画面上でマウスを上下左右にドラッグするか、画面右下にあるコンパスの左右にある矢印をクリックすることで、周囲を見回すことができます。また、マウスのホイールボタンや、画面右下の＋、－ボタンで画像の拡大、縮小ができます。

#### 5.3.2 タスクの実施と収集したデータ

今述べたタスクをランサーズを用いて 2021 年 2 月 5 日に実施した。具体的には、図 4 の 141 カ所に対して、1 回のタスクあたり図 4 の中から我々が指定した 3 カ所の地点について説明文を回答してもらった。その際ワークには、図 4 の中から指定した 3 カ所の地点が分かる地図と、図 6 のような 360 度写真を提示した。また、1 名がタスクに参加できるのは 3 回までとした。タスク 1 件につき実際に得られた説明文の例は、表 4 の「説明文」の列に示す。表にあるように、ランドマークは「UCC コーヒーアカデミー神戸校」で共通しているが、「南南東側」や「北西側」のような、細かな位置情報のある説明文が得られていることが分かる。

## 6 まとめと今後の課題

本研究では、自然言語で書かれたある地点についての説明文から、その地点がどこであるかの緯度経度を推定する問題の取り組みにあたって必要となるデータの収集と、収集されたデータの分析を行った。

現在、合計 415 カ所の地点に対する説明文を収集した。今後もし引き続き説明文を収集するとともに、得られた説明文をもとに人が推測する緯度経度についても収集する。現時点では、新たなデータセットの収集に向けて、三宮周辺で網羅的に点を決めて、各地点の説明文を収集する活動を実施する予定である。その後、収集された説明文と緯度経度において、位置を推定することに資する情報がどのように出現するかの分析をさらに進める。

## 謝 辞

本研究の一部は JSPS 科学研究費助成事業 JP18H03243 による助成を受けたものです。ここに記して謝意を表します。

## 文 献

- [1] Ludovic Moncla, Walter Renteria-Agualimpia, Javier Nogueras-Iso, and Mauro Gaio. Geocoding for texts with fine-grain toponyms : an experiment on a geoparsed hiking descriptions corpus. In *Proceedings of the 22nd ACM sigspatial international conference on advances in geographic information systems*, pp. 183–192, 2014.
- [2] Paolo Nesi, Gianni Pantaleo, and Marco Tenti. Ge(o)lo(cator): Geographic information extraction from unstructured text data and web documents. In *2014 9th International Workshop on Semantic and Social Media Adaptation and Personalization*, pp. 60–65, 2014.
- [3] Paolo Nesi, Gianni Pantaleo, and Marco Tenti. Geographical localization of web domains and organization addresses recognition by employing natural language processing, pattern matching and clustering. *Engineering Applications of Artificial Intelligence*, Vol. 51, pp. 202–211, 2016.
- [4] Aleksey Panasyuk, Edmund Szu-Li Yu, and Kishan G. Mehrotra. Improving geocoding for city-level locations. In *Proceedings of the 2019 IEEE 13th International Conference on Semantic Computing*, pp. 416–421, 2019.
- [5] Allison Gyle Woodruff and Christian Plaunt. Gipsy: Georeferenced information processing system. *Journal of the American Society for Information Science*, Vol. 45, pp. 645–655, 1994.
- [6] 奥村貴俊, 彌富仁. Tweet 解析によるユーザー埋め込み表現を用いた都道府県レベルでの位置推定. 言語処理学会第 24 回年次大会発表論文集, pp. 73–76, 2018.
- [7] 河野愛樹, 中村健二, 小柳滋. マイクロブログから抽出した地物情報と投稿間隔を考慮した位置情報推定. 情報処理学会第 73 回全国大会講演論文集, pp. 785–786, 2011.
- [8] 河野慎, 植田一博. Recurrent neural network によるマイクロブログの投稿位置推定. 人工知能学会論文誌, Vol. 32, No. 1, pp. 1–8, 2017.
- [9] 原久美子, 木野泰伸, 鳥海不二夫. 字・町名をキーとした災害時 twitter 情報の抽出と地図への展開—「どこ」で「何」が起きているのかを知る—. 人工知能学会全国大会論文集, Vol. 27, pp. 1–4, 2013.
- [10] 今井良太, 数原良彦, 戸田浩之, 鷺崎誠司. Web 文書を利用した POI に関連する地名の抽出. 情報処理学会 情報アクセスシンポジウム 2013, pp. 9–13, 2013.
- [11] 細川宜秀, 高橋直久. ドキュメント・データを対象としたジオコーディング手法. 情報処理学会 データベースシステム研究会研究報告, pp. 87–93, 2003.
- [12] 三浦英俊. 緯度経度を用いた 3 つの距離計算方法. オペレーションズ・リサーチ, Vol. 60, No. 12, pp. 701–705, 2015.
- [13] 伴幸児島, 和也山根, 政生柳澤, 辰夫大附, 望戸川. 道路標識とランドマークを用いた歩行者位置特定システムと実地調査による評価. 情報処理学会論文誌, Vol. 51, No. 3, pp. 899–913, 2010.
- [14] 徐潤純. 情報伝達における日韓比較 : 道案内の場面を中心に. 同志社大学日本語・日本文化研究, No. 17, pp. 61–74, 2020.
- [15] 小林亘, 相良毅. 2016 年熊本地震における道路被災の位置特定と道路ジオコーダの開発. GIS-理論と応用, Vol. 25, No. 2, pp. 91–100, 2017.
- [16] 新垣紀子, 野島久雄. 人はいつ道を尋ねるのか : ナビゲーションにおける外的資源としての他者. 認知科学, Vol. 5, No. 3, pp. 49–58, 1998.
- [17] 森永寛紀, 若宮翔子, 谷山友規, 赤木康宏, 小野智司, 河合由起子, 川崎洋. 点と線と面のランドマークによる道に迷いにくいナビゲーション・システムとその評価. 情報処理学会論文誌, Vol. 57, No. 4, pp. 1227–1238, 2016.
- [18] 杉谷卓哉, 白川真澄, 原隆浩, 西尾章治郎. 教師あり機械学習を用いたツイート投稿時のユーザ位置推定手法. 情報処理学会 データベースシステム研究会 研究報告, pp. 1–8, 2013.
- [19] 相良毅, 有川正俊. ジオパースによる web からの空間コンテンツ獲得. 電子情報通信学会 15 回データ工学ワークショップ, 2004.
- [20] 相良毅, 有川正俊, 坂内正夫. ジオリファレンス情報を用いた空間情報抽出システム. 情報処理学会論文誌データベース (TOD), Vol. 41, No. SIG06(TOD7), pp. 69–80, 2000.
- [21] 増山篤. フリーのデータ, ツール, サービスによる地点間距離・移動時間の計算実験. GIS-理論と応用, Vol. 25, No. 2, pp. 71–78, 2017.
- [22] 谷謙二. ジオコーディングと地図化の Web サイトの構築とその活用—Google Maps API を利用して—. 埼玉大学教育学部地理学研究報告, Vol. 30, pp. 1–12, 2010.
- [23] 中村健二, 田中成典, 大谷和史, 山本雄平. セキュアライフの創出を目指した安全知の獲得に関する研究—犯行予告マップの構築—. 日本知能情報ファジィ学会第 26 回ファジィシステムシンポジウム講演論文集, pp. 708–713, 2010.
- [24] 長谷川保, 松田三恵子, 久保田浩明. 歩行者向け対話型道案内サービスのための案内テキスト生成. 映像情報メディア学会技術報告, pp. 43–48, 2001.
- [25] 南佳孝, 関本義秀, 中條覚, 柴崎亮介. 路線情報を加味した道路関連情報の位置特定に関する研究. 土木学会論文集, Vol. 67, No. 1, pp. 7–17, 2011.
- [26] 野秋浩三, 相良毅, 有川正俊. 歩道ネットワークと地名辞書を基本とした日常的な場所表現を対象としたジオコーディング手法. 電子情報通信学会第 15 回データ工学ワークショップ論文集, 2004.
- [27] 有川正俊, 野秋浩三. 歩道ネットワークデータベースを用いた道案内文のジオコーディング手法. 第 43 巻, pp. 26–29, 2005.