

ハイブリッド位置情報サービスのための POI 情報の統合プラットフォーム

一円 真治† 梶 克彦‡ 河口 信夫‡
名古屋大学工学部† 名古屋大学大学院工学研究科‡

1 はじめに

現在、スマートフォンなど GPS 機能を持ったデバイスの普及により位置情報の取得が容易になっている。これに伴い飲食店やホテルなどの POI(Point Of Interest)に関する情報を提供する位置情報サービスが多く登場している。例えば Yahoo! Local Search, Hotpepper の周辺の飲食店検索などがある。またこれらのサービスは Web API(以下 API) を公開しており、各サービスが保持しているコンテンツを利用できる。しかし、POI に関する Web サービスには網羅性や情報の質に問題がある [1]。我々の調査の結果、以下の課題が解決されていない。

1. API 利用者が欲しい情報を個別の API からは網羅的に取得できない
2. 各サービスが持つ属性情報が異なる
3. API により POI の属性情報の記述フォーマットが異なる

これらの課題を解決するために、複数の位置情報サービスから POI 情報を取得し統合的に利用可能とするプラットフォームを提案し、同一地点を表す複数の POI 情報を統合する手法、その評価と今後の課題を述べる。

2 POI 情報の分析

ぐるなび、食べログ、Hotpepper の各 API に対して、名古屋駅を中心とし半径 1km (食べログは API の仕様上 1.5km) の範囲内に含まれる POI 情報を取得した。該当 POI 数はぐるなび: 448 件, Hotpepper: 525 件, 食べログ: 1022 件であった、これより各 API から取得できる POI 情報が異なることがわかる。

POI の属性情報を複数の API から取得し、表 1 にその違いを示す。ぐるなびから取得できた名前情報には店舗のキャッチコピーが含まれている。住所情報では県名や建物名の有無や丁番号の表記の違いがある。Hotpepper からは電話番号情報を取得できない。これらより POI に関して取得できる属性情報とそのフォーマットは各 API によって異なることがわかる。

表 1: POI の属性情報 [赤から名駅店]

属性情報	API		
	ぐるなび	Hotpepper	Yahoo!
名前	赤から鍋と せせり焼き 赤から名駅店	赤から 名駅店	赤から名駅店
住所	愛知県 名古屋市 中村区 名駅 3-14-16 東洋ビルディング B1	名古屋市 中村区 名駅 3-14-16 東洋ビル B101	愛知県 名古屋市 中村区名駅 3 丁目 14-16
緯度	35.172474	35.172474	35.1723811
経度	136.883723	136.883723	136.8835944
電話番号	052-588-xxxx	なし	052-588-xxxx
その他	営業時間 アクセス PR 文 etc	予算 収容人数 クーポン etc	最寄り駅 ジャンル etc

3 POI 情報統合プラットフォーム

本プラットフォームの概要を図 1 に示す。クライアントとは一般ユーザ、開発者を指す。POI 情報統合システムは複数の位置情報サービスとクライアントとのハブの役割を果たし、クライアントからのリクエストを受け、複数の位置情報サービスとやり取りを行う。そして得られたレスポンス情報を統合しクライアント側に、より網羅性の高い情報を返す。複数の位置情報サービスから同じ POI に関する情報が得られた場合、それらの属性情報を統合し多様な POI 情報を返す。

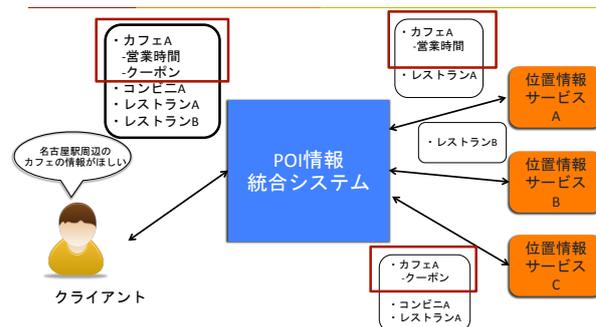


図 1: POI 統合プラットフォーム

3.1 Web アプリとしてのプロトタイプの実装

現在、Web アプリとしてプロトタイプの開発中である (図 2)。地図上に POI がマーカーとして表示され、クリックすると POI 情報閲覧ウィンドウが表示される。POI の名前、住所などの基本情報が表示されており、API アイコンから該当 API から取得できた属性情報を閲覧

A Platform Integrating POI Information For Hybrid Location-based Service
Shinji Ichien † Katsuhiko Kaji ‡ Nobuo Kawaguchi ‡
School of Engineering, Nagoya University †
Graduatel School of Engineering, Nagoya University ‡

できるようになっている。



図 2: Web アプリでの POI 情報閲覧画面

3.2 POI の統合手法

中心地点と半径を設定し取得範囲を決め、複数の Web API から POI 情報を取得する。POI が持つ住所、電話番号、名前の属性情報を用いて POI 情報同士の同定をする。まず POI の属性情報の正規化を行う。住所情報の正規化には住所正規化 API[2] を用いる。例えば、住所情報 [名古屋市中村区名駅 3 丁目 14-16] をリクエストすると、[愛知県名古屋市中村区名駅 3-14-16] という様に足りない情報は補完され、正規化した住所情報が得られる。電話番号情報は関連性のない文字列が含まれていればできる限り削除し、ハイフンを含めた形を正規形とする (例:[052-123-4567])。

次に得られた複数の POI 情報を住所情報でグルーピングをする。各住所グループ内で POI 情報のすべての組み合わせにおいて同定する。グループ分けの理由は同定できる POI 情報同士は同じ住所情報を持つと考えられ、グループ内で同定を行うことで組み合わせ数が減り、計算効率が向上するからである。

3.3 同定判定

入力である 2 つの POI 情報が持つ正規化した電話番号情報で比較を行い、同一であれば名前情報を用いて名前文字列比較をする。一方の POI 情報に電話番号情報がなければ、店舗検索 API を活用した API 比較を行う。また両方になれば名前文字列比較を行う。それぞれの場合で POI 情報同士を同定する。

API 比較：店舗検索 API とは電話番号情報を用いて API にリクエストを送り、同じそれを持つ POI 情報を取得できる API とする。入力する 2 つの POI を A・B とし、A の電話番号情報を用いて B の店舗検索 API から POI:C(以下 C) が得られたとする。取得元 API が同じなので、C が B と同一の POI 情報ならば、保持する POI の属性情報やそのフォーマットは同一であると考えられる。POI に固有であると考えられる名前情報で B と C に対して文字列比較を行い、同一文字列であれば A と B を同定する。この判定は A が電話番号情報を保持、かつ B の取得元 API が店舗検索 API に対応している必

要がある。

名前文字列比較：2 つの POI 情報の店舗名文字列から編集距離 [3] と最長共通部分文字列長を求める。店舗名文字列とは POI の名前情報文字列から店舗所在地文字列を除いた文字列である。名前文字列をスペースで区切り、複数の文字列に分割する。店舗所在地文字列は其中で「店」を最後尾に含む文字列をさす。例えば、名前文字列「赤から名駅西口店」の場合、店舗名文字列として「赤から」が得られる。これらと真の判定結果を特徴量として与え、機械学習より得られたアルゴリズムにより同定をする。

4 評価

機械学習ソフト「Weka」に正解:50 件、不正解:450 件のデータセットを与え、分類器には「J48」を用いた (編集距離 : D, 最長共通部分列長 : L)。これより、L が 2 以上かつ D が 8 以下の場合と D が 9 以上ならば L が 5 以上である場合に同定する手法を得た。ぐるなび:448 件, Hotpepper:525 件, 食べログ:1022 件の全 POI の統合可能性判定結果と統合可能 POI 数の真値を表にまとめた。統合可能 POI は複数の API から POI 情報を取得できる POI のことである。適合率は 1.00。再現率は約 0.91 であった。

表 2: 統合判定結果と真の結果

		真の結果	
		可	否
予測結果	可	504	0
	否	51	791

5 今後の課題

POI 数の網羅性と情報の正確性向上のため、対応 API の追加とユーザからの情報投稿を可能とするクラウドソーシング活用システムの開発を考えている。一般ユーザの投稿情報とは POI 情報統合に誤判定があった場合の修正データ、POI 登録、POI の属性情報の追加を指す。

参考文献

- [1] 石田 和成. ジオメディアの分布調査および携帯端末を用いたジオローカルコンテンツ構築システムの開発. 情処研報. DD, [デジタル・ドキュメント] 2011-DD-80(11), 1-8, 2011
- [2] 住所正規化 API. http://tou.ch/developer/api_all?uri=geo%2Faddress_normalize.
- [3] 花田 博幸, 工藤 峰一. 編集距離による文字列の探索高速化に関する研究. 信学技報. PRMU, パターン認識・メディア理解 108(94), 41-45, 2008