

地域特産メニュー抽出手法の提案

宇部雅彦[†] 村田嘉利[†] 鈴木彰真[†]

岩手県立大学ソフトウェア情報学部[†]

1. はじめに

旅行や出張の際に地域特産料理を食べることは旅行者にとって楽しみの一つであるだけでなく、地域経済の振興の面からみても推進されるべきである。特に、無名な地域特産料理の知名度向上は地域活性化につながる。しかしながら、そのような料理を出す店舗を見つけることが難しい。また、地元の人に日常食として認識されていた場合、地域特産の料理として宣伝されない。現在、位置情報付きコンテンツからの地域限定語句の抽出[1]が提案されているが、地域特産料理の推薦は行われていない。そこで、知名度の低い地域特産メニューを抽出し、検索サイトとして提供するシステムを提案する。本稿では、店舗のメニューをインターネット上から情報収集し、地域特産メニューを抽出するアルゴリズムを考案した。また、抽出実験によって提案アルゴリズムの評価を行った。

2. 先行研究

一般的な商品情報抽出として、HTML 文書からの商品情報抽出方式の提案[2]がある。この論文では、オンラインショップの HTML 文書について、商品情報の記述における特徴を調査し、それを用いた抽出手法を提案している。

この商品情報抽出方式を利用し、店舗情報サイトからメニュー情報とメニューを取扱う店舗情報を抽出する。HTML から必要部分以外の情報を削り取る Web スクレイピング手法との併用により、地域単位でメニュー情報を比較することが可能になり、地域特産メニューを抽出することができる。

3. 地域特産メニュー抽出システム

3.1. メニュー情報の収集

地域特産メニューの抽出を行うためには、店舗ごとのメニュー情報を収集しなければならない。また、地域特産メニューは県レベルから市町村レベルまで存在する。全てのレベルに対応できるように、詳細な店舗住所の収集も行う必要がある。本研究では、食ベログ[3]からこれら

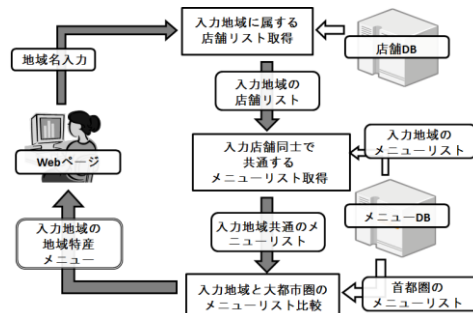


図1 地域特産メニュー抽出の概要

の情報を収集し、それぞれを対応付けた。

3.2. 地域特産メニュー抽出アルゴリズム

地域名入力による地域特産メニュー抽出の概要を図1に示す。まず、入力地域内の店舗リストを取得し、地域内全店舗のメニューリストを取得する。次に、店舗同士のメニューリストを比較し、単独店舗が取り扱う独自のメニューを除外することで、地域内で共通するメニューリストを抽出する。その上で、地域内で共通するメニューリストを首都圏のメニューリストと比較し、一般に普及しているメニューとして首都圏のメニューを除外する。比較結果は、入力地域の地域特産メニューとして出力される。

3.3. 抽出システムの設計

抽出アルゴリズムを実現するプログラムを作成した。プログラムの構成は、図2に示す通りにメニュー情報収集部分は利用者の地域名入力から、店舗情報と店舗住所情報を食ベログ[3]から抽出し、データベースに格納する。その後、店舗毎にメニュー情報を抽出し、データベースに格納する。

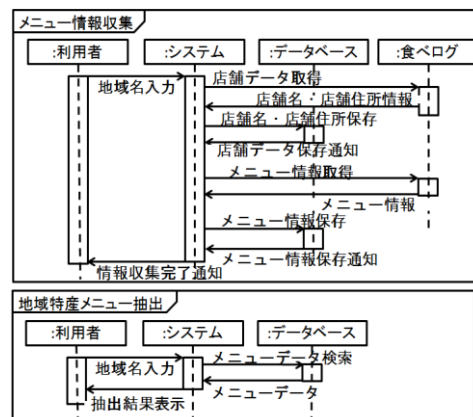


図2 プログラムのシーケンス図

Proposal of Local Foods Extraction Method
M.Ube[†], Y.Murata[†] and A.Suzuki[†]
[†]Iwate Prefectural University

一方、地域特産メニュー抽出部分では、入力地域内で共通するメニューリストをデータベースから検索する。抽出地域範囲は要求に合わせて県から市区町村までの指定を自由に変更できるようにした。その後、検索結果から首都圏で存在するメニュー名を除外し、地域特産メニューを抽出する。抽出結果は文字列として出力される。

4. 盛岡市における地域特産メニューの評価

4.1. 評価の概要と結果

抽出アルゴリズムの有効性を評価するため、岩手県盛岡市の地域特産メニュー抽出を行った。抽出結果から、得られたメニュー名を特産品情報サイト[4]より調べることで、メニュー名が地域特産であるかどうか評価した。盛岡における抽出実験の結果と評価結果を表1に示す。表1の左右でそれぞれ地域特産メニューとそれ以外に分類した。

表1の右に示すメニューを除外するため、フィルタリングが必要になる。表1の「シュー♪」など、検索結果からメニューではない文字列や首都圏にない一般メニューがあった。また、表1の「唐揚げBIGプレート」など、東北地方で展開しているチェーン店のメニュー名が複数取得され、首都圏では取得できずに抽出結果から除外された場合があった。そのため、これらは地域特産メニューとしては除外されなければならない。加えて、「コーヒーはサービス」など、複数のメニューが組み合わせられた場合は接続詞で分割を行い、どちらも首都圏に存在するときは除外すべきである。

一方、「ちいたんたん」などの地域特産メニューは同一料理でありながら、異なった複数のメニュー名が抽出されたため、メニュー名の部分一致を用いて1つのメニュー名に統合すべきである。

4.2. 追加のフィルタリングによる再評価

そこで、抽出結果から該当データを除外するフィルタリングを実装した。また、複数抽出された同一メニューを統合する機能を実装した。実装したフィルタリングを表1の抽出結果に適用した結果を表2に示す。表2に示されたように、岩手県盛岡市の地域特産メニューが抽出できた。さらに、同県久慈市、北上市についても同様の手法で地域特産メニューが抽出できた。しかし、首都圏で見られる対象地域の地域特産メニューが一部除外される結果となった。首都圏に存在するメニューの中に、抽出地域の地域特産メニューが含まれており、今後その除外方

表1 地域特産メニュー抽出結果と評価結果

地域特産メニュー	地域特産メニュー以外
じゃじゃ麺小	コーヒーはサービス
じゃじゃ麺中500円	しゃかしゃかフライドポテト
チータン	シュー(▽)
ちいたんたん	シュー♪
ちーたんたん	どんこの肝和え
ちーたんの卵	ほや刺
ひつつみ	ロシアンたこ焼20
ひつつみ定食	家族でワイワイボリューム満点!
ペアレソビール	海老汁
芋の子汁	御開帳〜
盛岡冷麺(辛味別)	盛岡燗
別辛	福神漬&らっきょうorz
	唐揚げBIGプレート

表2 抽出結果のフィルタリング結果

じゃじゃ麺
ちいたんたん
ひつつみ定食
ペアレソビール
盛岡冷麺
芋の子汁

法については検討する。また、今回抽出実験を行った地域以外の地域についても抽出手法とフィルタリングの精度について検討を行い、最適化を行っていく。

5. おわりに

本稿では、知名度の低い地域特産を取り扱う店舗とメニューを抽出するため、それらの情報をウェブサイトから収集し、提案手法による岩手県盛岡市、久慈市、北上市の地域特産メニュー抽出を行った。提案した手法では、抽出結果に地域特産メニューが混在するという結果が得られた。そのため、上記の抽出結果に対してフィルタリングを実施し、地域特産メニューのみが抽出できた。

今後の展望として、一般に普及するメニューの判定方法の検討をおこなう。また、本稿での抽出地域以外の地域についてのシステムとフィルタリングの評価を行うことで、より最適な抽出をすることを目標とする。

その上で、地域特産メニューの分布を店舗の住所情報から抽出し、地域特産メニューの知名度を推定するシステムに応用していく。

参考文献

- 1) 奥, 健太.; 西崎, 剛司.; 服部, 文夫.; 地域限定性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 5(3), 2012-09-28, 97-116
- 2) 富田, 一.; 手塚, 祐.; 山本, 修.; 長岡, 満. HTML文書から商品情報抽出方式の提案. 電子情報通信学会技術研究報告 1998, 97 (502), 15-22.
- 3) グルメ・レストランガイド[食ベログ]. <http://tabelog.com>.
- 4) 特産品情報サイト★たくさんとくさん★. <http://www.takusan.net>.