

レストラン検索サイトの メニュー名の統計的分析による食の地域性の考察

田中 優里 山西 良典 西原 陽子

立命館大学情報理工学部

{is0315xr@ed, ryama@fc, nishihara@fc}.ritsumei.ac.jp

概要 本稿では、レストラン検索サイトのメニュー名を統計的に分析する手法を提案し、食の地域性を考察する。行き先の地域での特徴的な食材や料理は旅の楽しみの1つである。そのような食事を調べるとき、レストラン検索サイトを利用することが多いが、レストランの情報は一覧できるものの、食の地域性（つまり、地域ごとに特色のある料理や食材、あるいは、調理方法）を直感的に把握することは難しい。そこで、メニューに用いられる文字列を統計的に分析することで、食の地域性を発見可能にする。具体的には、文字列が出現する地域の頻度や任意の地域内での文字列の出現頻度、任意の地域内で文字列が出現するレストランの頻度などの統計的指標をもとに考察可能なしくみを構築した。

キーワード 食メディア、食の地域性、情報可視化、TF-IDF の応用

1 はじめに

食は、人間が生きていくために必要な生命活動の一つであり、娯楽、教育、健康、文化、コミュニケーションなどの様々な分野においても人間の生活に密着している。しかし、高齢化などの社会の変化に伴って、食を通じたコミュニケーションの場や食文化の継承の機会、健康管理などへの意識が失われつつあり、食に関する諸問題は現代社会が抱える喫緊の課題の一つと捉えられている。

Cookpad によるレシピデータの研究者向けの公開などを受け、電子情報通信学会 HCG 食メディア研究会をはじめとして、食に関して様々な観点での研究が行われている。立命館大学には、2018 年度より食マネジメント学部が開設されるなど、食に対する科学的なアプローチはこれまでにない盛り上がりを見せている。

一方で、これまでの研究の多くは、調理や料理といった食コンテンツそのものを扱ったものが多かった。レシピ検索サイトと同様に、食に関連した情報に触れる場としてはレストラン検索サイトが存在する。海外のレストラン検索サイトである Yelp は、Dataset を公開し、Photo Classification, Natural Language Processing & Sentiment Analysis, Graph Mining といったテーマでレストラン検索サイトのデータを有効活用したコンペティションを開催している¹。本稿でも、レストラン検索サイトのデータを活用した研究事例を報告する。

食は文化や歴史、気候風土と関連して地域ごとに味付けやメニューが異なり、特色のある料理や食材や調理方法が存在する。つまり、食には「地域性」が存在すると考えられる。例えば、古くから伝わる滋賀の「鮎ずし」

のように郷土料理が存在したり、地域ごとで出汁の取り方が異なることでの同じ味噌汁であっても味に地域性が生まれていたりする。外食産業や食品産業にとっては食の地域性は国内での流通においても重要な課題であり、日清食品では地域の味を利用した東日本と西日本で異なる出汁を使った商品を発売している。食の地域性が明らかになれば、よりユーザの嗜好に沿ったレストラン推薦やメニュー推薦が可能になると考えられる。

本研究の最終目的は、Web 上のレストラン検索サイトから得られる情報をもとに食の地域性をマップとして可視化するしくみを構築することである。本稿では、その端緒として、レストラン検索サイトの統計的分析によって、メニュー名に現れる地域性を明らかにする手法を提案する。

2 関連研究

単語の地域性に関連する研究として、川久保ら [1] の代表的な位置情報付き画像データから単語概念に対応する視覚（画像）の地域性を定量化する研究がある。地域別代表画像を選出し、選出された地域別代表画像の部分領域分布を分析することで、単語概念に対応する視覚の地域差の大きさを求める。地域差の分析という点では類似しているが、本研究とは、扱うデータの性質やドメインが異なっており、新たな手法を提案する必要がある。

中岡ら [2] は、地域で異なるラーメンの味について着目し、レストラン検索サイトの口コミ情報から地域の特徴を分析している。味に関して地域の特徴を捉えるため、口コミ情報から対応分析と自己組織化マップを用いた分析が行われている。福田ら [3] は、写真データと、位置と料理ジャンルの情報をもつ膨大なレストランデータから、階層的な主要撮影地域と人気グルメのジャンルを抽

Copyright is held by the author(s).

The article has been published without reviewing.

¹<https://www.yelp.com/dataset/challenge>

出して可視化する手法を提案している。階層的な主要撮影地域とは、位置情報を用いて、写真データをクラスタリングすることにより、階層的に主要撮影地域を抽出したものである。福田らはレストランのジャンルについて分析を行っているが、本研究ではさらに詳細な料理や食材、調理方法の抽出を目指す。奥ら [4] の研究では、対象とする位置に対し、現地での出現頻度が相対的に高い語句（地域限定語句）の抽出が行われている。情報源として、レストラン検索サイトやスポット情報サイトなどから収集される位置情報付きコンテンツ（位置情報、PR 文やレビュー）を使用し、語句の地域限定性を表すスコアを算出している。特徴語句抽出によく用いられる IDF および奥らが定義する地域関連重みの組合せにより対象地域に関連しつつ特異な語句の抽出に成功している。本研究では、奥らの地域限定語句の抽出方法を参考に、地域で特色のある食の地域性をレストラン検索サイトの情報から抽出する。

3 提案手法

提案手法では、以下の手順に従ってレストラン検索サイトのメニュー情報を分析した。

- Hotpepper グルメの API を用いて、指定した地域の検索上位 100 件のレストランのメニューを取得する。
- 取得したメニューに対して、N-gram を抽出する。N の値は、1 からメニュー名の最大文字列長までとした。ここで、辞書を用いた形態素解析ではなく N-gram を用いる理由は、辞書には登録されていない未知のメニューが存在する可能性を考えたためである。
- 抽出された N-gram について、地域性を算出するための統計的指標として TF(Term Frequency), RF(Restaurant Frequency), LF(Local Frequency) の値をそれぞれ算出する。

文書から特徴的な単語を抽出する方法として、単語の出現頻度を用いる TF-IDF 法がある。TF-IDF 法では、「各文書における単語の頻度」を表す TF (Term Frequency) と、「ある単語が複数の文書館での希少度」を表す IDF (Inverse Document Frequency) の積を単語の重要度として扱っている。TF-IDF は、一般的な機械学習で必要とされる重要な単語についてのアノテーション付きの学習データを必要とせず、分析対象となる文書集合のみを用いて算出可能な単語重要度として言語処理分野で長らく用いられてきている。

提案手法では、この TF-IDF 法を食の地域性を分析するための再解釈を行う。地域性が高い食は、その食を示

表 1 分析対象とした全 63 地域の一覧。下線は、考察対象とした 16 地域を示す。

都道府県名	地域名	都道府県名	地域名
東京	御茶ノ水	東京	神保町
島根	松江駅	山口	山口駅
東京	表参道	東京	田園調布
東京	新宿西口	東京	北千住
東京	大岡山	千葉	津田沼
神奈川	湘南台	愛知	名古屋駅
鳥取	鳥取駅	千種	千種
愛知	鶴舞	北海道	札幌駅
北海道	新札幌	北海道	旭川市中心
宮城	青葉区	茨城	つくば駅北・筑波大学
香川	高松駅・北浜	徳島	徳島駅
富山	富山駅	福岡	博多駅 (博多口)
福岡	福岡市東区	福岡	小倉・平和通駅・魚町
長崎	浦上	大分	府内町・大手町・金池
宮崎	宮崎駅	鹿児島	鹿児島中央
新潟	新潟駅前	石川	金沢駅
京都	北野天満宮	北海道	函館駅
北海道	五稜郭	福島	会津若松駅
岩手	盛岡駅	東京	早稲田
東京	西早稲田	東京	目白
東京	後楽園	東京	船堀
東京	東小金井	東京	八王子みなみ野
東京	八王子市その他	東京	永田町
東京	赤羽橋	東京	半蔵門
神奈川	川崎市多摩区その他	神奈川	横浜市港北区その他
埼玉	高坂	千葉	野田市その他
滋賀	草津市	京都	出町柳
京都	宝ヶ池	京都	京田辺
大阪	樟葉	愛知	豊田市その他
宮城	太白区その他	広島	五日市
山口	湯田温泉	愛媛	松山駅
福岡	戸畑		

す文字列がその地域の多くのレストランのメニュー名に出現すると考えられる。また、一方で、地域性が高い食は、その食を示す文字列がその他の地域ではあまり用いられない傾向にあると考えられる。上記の考え方に則ると、TF に代わる指標として単語が出現したレストラン数 (RF), DF に代わる指標として単語が出現した地域数 (LF) をそれぞれ定義できる。

地域 a における文字列 t の RF: $rf_a(t)$ は式 (1) で算出される。

$$rf_a(t) = \frac{n_a(t)}{\sum_{i \in T_a} n_a(i)}, \quad (1)$$

ここで、 $n_a(i)$ は地域 a において文字列が出現したレストランの頻度を示し、 T_a は地域 a に出現する文字列の集合を示す。

文字列 t の LF: $lf(t)$ の逆数である $ilf(t)$ は式 (2) で算出される。

$$ilf(t) = \log \frac{N}{lf(t)} + 1, \quad (2)$$

ここで、 $lf(t)$ は文字列 t が出現した地域数を示す。

式 (1) と式 (2) で算出されたそれぞれの値の積を式 (3) によって求めることで、文字列 t の地域 a における地域性を示す指標 $rf \cdot ilf$ を算出する。

$$rf \cdot ilf_a(t) = rf_a(t) \times ilf(t). \quad (3)$$

4 分析と考察

分析対象としたデータは、Hotpepper グルメの Web サイトに存在するレストランのメニュー情報とした。分析対

象の地域には、過去に情報処理学会の全学会大会、あるいは、人工知能学会の全学会大会が開催された63地域についてHotpepper グルメのAPI (<https://webservice.recruit.co.jp/hotpepper/>) で取得可能な学会開催地から一番近い地域を用いた。表1に、本研究で扱った全63地域(APIで取得した地域名)の一覧を示す。

本稿では、ここから分析対象とする地域が日本の各地区(北海道、東北、関東、中部、近畿、中国、四国、九州・沖縄)に分散するように都道府県を選択した。選択した都道府県に存在する都市から1都市を無作為に分析対象として決定し、合計16地域を詳細な考察の対象とした。考察対象とした16地域は、表1中で下線で示す。全63地域の上位10件のN-gramと16地域から上位5件ずつのN-gramから得られた文字列について $tf \cdot ilf$ の値を考察した。

4.1 文字列としての考察

表2に、文字列「カル」「ルビ」「カルビ」についての各地域それぞれでの $rf \cdot ilf$ を示す。まず、「ルビ」について見てみると、どの地域であっても「カルビ」と概ね同値の $rf \cdot ilf$ を示しており、「ルビ」という文字列がほぼ全ての場合において「カルビ」に含まれた文字列であると示唆される。一方で、「カル」については「カルビ」について高い $rf \cdot ilf$ を示した松江駅、名古屋駅、会津福島駅以外の金沢駅で高い値を示した。金沢駅の「カル」が含まれるメニューには「カルビ」だけでなく、「マスカルポーネ」、「エスカルゴ」などのほか、特に「カルパッチョ」という文字列が用いられたメニューが多く存在したためと考えられる。また、金沢駅では、表3に示した「いか」「海老」という文字列についての $rf \cdot ilf$ についても、他の15地域よりも高い値を示しており、海鮮類を多く用いたメニューが多く存在すると推察される。

表4に、文字列「ステ」「テー」「ーキ」「ステーキ」についての各地域それぞれでの $rf \cdot ilf$ を示す。どの地域でも、どの文字列に対しても、おおよそ同程度の値を示しており、メニューの一部に「ステーキ」という単語が含まれていることが推察される。「ステーキ」の $rf \cdot ilf$ について、最高値を示したのは松江駅であった。松江駅のメニューについて見てみると、他の地域と比べて「ステーキ」が多いというわけではなかった。本実験では、1地域につき上位100件のレストラン情報を取得したが、松江駅はレストラン情報が27件しか取得できず、他の地域よりも少ない情報量の中で分析が行われており、メニューに偏りが出ている可能性があることがわかった。

4.2 単語としての考察

表5に、文字列の中でも単語として解釈可能な「鳴門」「カレー」について $rf \cdot ilf$ を示す。まず、「鳴門」については草津市で低い値が見られるものの、徳島駅で突

出して高い値を示しており、他の全ての地域で値が0となっている。徳島県には、地名として「鳴門」があり、メニューの一部に「鳴門」を用いた「鳴門金時」「鳴門わかめ」「鳴門蓮根」「鳴門鯛」などの地場産品が多いことが明らかになった。

「カレー」について見てみると、会津若松駅、松江駅、鳥取駅という順で高い値を示していた。鳥取市は、1世帯あたりのカレー年間支出金額および購入数量(2014年~2016年平均)²で日本1位になっており、カレーを食べる機会が多いため高い値が示されたと推察される。会津若松駅について、「カレー」が含まれるメニューを見てみると「カレー焼きそば」や「カレーピザ」「カレーそば」などの様々なメニューに「カレー」が使用されていた。さらに、会津若松駅には「会津カレー」というメニューが存在しており、ご当地のカレーメニューが存在することがわかった。一方で、松江駅における「カレー」を含むメニューについて見てみると、カツカレーやビーフカレーといった定番メニューが多く、鳥取駅や会津若松駅ほどの「カレー」との深い関連は見られず、単純に「カレー」が他の地域よりも一般的なメニューとして食べられていると考えられる。

次に、調理方法を意味する文字列について着目してみる。表6に、「揚げ」「焼き」「炒め」についての各地域それぞれでの $rf \cdot ilf$ を示す。「揚げ」は、松江駅で高い値を示している。松江駅での「揚げ」を含むメニューには、「揚げ鶏」や「竜田揚げ」「揚げごぼう」「揚げバナナ」などの様々な揚げ物が見られた。また、「焼き」は、会津若松駅で高い値を示していた。「焼きそば」や「お好み焼き」「玉子焼き」「ちゃんちゃん焼き」など、一見地域に深い関連のないようなメニューが多かったが、さらに詳しく見てみると「会津地鶏 レバー串焼き」「会津地鶏の玉子焼き」「【会津B級グルメ】カレー焼きそば」を発見した。会津の地鶏を使用したメニューや、会津若松駅で高い重要度を示していたカレーを使ったメニューが存在することが明らかになった。「炒め」は、新潟駅前で高い値を示していた。「炒め」を含むメニューについて見てみると、「海老と野菜のあっさり塩炒め」や「しじみガーリック炒め」「ベビーホタテのガーリックバター炒め」などの海鮮類を使用した炒めものが多いことがわかった。また、炒めもの特徴として、キムチや味噌、ガーリック、マヨネーズなどの様々な味付けがされていたが、その中でもガーリックとバターを使用した炒めものが多いという二次的な発見も得られた。このように、調理方法にも地域性が存在することが明らかとなった。

²<http://www.stat.go.jp/data/kakei/5.html> (2018年10月24日確認)

表2 「カル」「ルビ」「カルビ」についての各地域それぞれでの $rf \cdot ilf$.

	島根県 松江駅	東京都 大岡山	愛知県 名古屋駅	鳥取県 鳥取駅	北海道 旭川中心部	茨城県 つくば駅北 筑波大学	徳島県 徳島駅	福岡県 博多駅 (博多口)	鹿児島県 鹿児島中央	新潟県 新潟駅前	石川県 金沢駅	福島県 会津若松駅	岩手県 盛岡駅	滋賀県 草津市	京都府 京田辺	愛媛県 松山駅
カル	0.0174	0.0166	0.0168	0.0198	0.0187	0.0136	0.0181	0.0166	0.0144	0.0170	0.0220	0.0170	0.0135	0.0185	0.0191	0.0114
ルビ	0.0244	0.0072	0.0232	0.0164	0.0112	0.0100	0.0092	0.0152	0.0132	0.0140	0.0164	0.0216	0.0116	0.0196	0.0140	0.0000
カルビ	0.0254	0.0074	0.0237	0.0168	0.0115	0.0065	0.0094	0.0151	0.0127	0.0147	0.0151	0.0221	0.0119	0.0205	0.0143	0.0000

表3 「いか」「海老」についての各地域それぞれでの $rf \cdot ilf$.

	島根県 松江駅	東京都 大岡山	愛知県 名古屋駅	鳥取県 鳥取駅	北海道 旭川中心部	茨城県 つくば駅北 筑波大学	徳島県 徳島駅	福岡県 博多駅 (博多口)	鹿児島県 鹿児島中央	新潟県 新潟駅前	石川県 金沢駅	福島県 会津若松駅	岩手県 盛岡駅	滋賀県 草津市	京都府 京田辺	愛媛県 松山駅
いか	0.0209	0.0205	0.0099	0.0254	0.0167	0.0072	0.0156	0.0159	0.0102	0.0182	0.0277	0.0231	0.0193	0.0209	0.0129	0.0057
海老	0.0145	0.0169	0.0141	0.0167	0.0114	0.0088	0.0167	0.0181	0.0154	0.0177	0.0202	0.0158	0.0177	0.0167	0.0065	0.0204

表4 「ステ」「テー」「ーキ」「ステーキ」についての各地域それぞれでの $rf \cdot ilf$.

	島根県 松江駅	東京都 大岡山	愛知県 名古屋駅	鳥取県 鳥取駅	北海道 旭川中心部	茨城県 つくば駅北 筑波大学	徳島県 徳島駅	福岡県 博多駅 (博多口)	鹿児島県 鹿児島中央	新潟県 新潟駅前	石川県 金沢駅	福島県 会津若松駅	岩手県 盛岡駅	滋賀県 草津市	京都府 京田辺	愛媛県 松山駅
ステ	0.0155	0.0101	0.0151	0.0126	0.0188	0.0117	0.0220	0.0176	0.0189	0.0172	0.0212	0.0172	0.0176	0.0163	0.0193	0.0260
テー	0.0169	0.0138	0.0161	0.0154	0.0154	0.0099	0.0196	0.0159	0.0193	0.0161	0.0204	0.0161	0.0138	0.0165	0.0196	0.0235
ーキ	0.0156	0.0149	0.0159	0.0154	0.0169	0.0081	0.0181	0.0152	0.0176	0.0154	0.0183	0.0190	0.0132	0.0176	0.0169	0.0256
ステーキ	0.0150	0.0122	0.0166	0.0141	0.0171	0.0069	0.0217	0.0159	0.0196	0.0166	0.0192	0.0173	0.0148	0.0169	0.0233	0.0280

表5 「鳴門」「カレー」についての各地域それぞれでの $rf \cdot ilf$.

	島根県 松江駅	東京都 大岡山	愛知県 名古屋駅	鳥取県 鳥取駅	北海道 旭川中心部	茨城県 つくば駅北 筑波大学	徳島県 徳島駅	福岡県 博多駅 (博多口)	鹿児島県 鹿児島中央	新潟県 新潟駅前	石川県 金沢駅	福島県 会津若松駅	岩手県 盛岡駅	滋賀県 草津市	京都府 京田辺	愛媛県 松山駅
鳴門	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.7989	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0326	0.0000	0.0000
カレー	0.0105	0.0226	0.0051	0.0239	0.0051	0.0184	0.0108	0.0045	0.0115	0.0115	0.0115	0.0242	0.0067	0.0118	0.0000	0.0242

表6 「揚げ」「焼き」「炒め」についての各地域それぞれでの $rf \cdot ilf$.

	島根県 松江駅	東京都 大岡山	愛知県 名古屋駅	鳥取県 鳥取駅	北海道 旭川中心部	茨城県 つくば駅北 筑波大学	徳島県 徳島駅	福岡県 博多駅 (博多口)	鹿児島県 鹿児島中央	新潟県 新潟駅前	石川県 金沢駅	福島県 会津若松駅	岩手県 盛岡駅	滋賀県 草津市	京都府 京田辺	愛媛県 松山駅
揚げ	0.0203	0.0111	0.0153	0.0175	0.0153	0.0125	0.0168	0.0176	0.0198	0.0193	0.0166	0.0225	0.0177	0.0151	0.0126	0.0246
焼き	0.0175	0.0170	0.0137	0.0150	0.0154	0.0136	0.0152	0.0166	0.0181	0.0164	0.0172	0.0209	0.0167	0.0151	0.0144	0.0195
炒め	0.0118	0.0129	0.0093	0.0143	0.0136	0.0186	0.0154	0.0154	0.0204	0.0233	0.0129	0.0190	0.0179	0.0143	0.0000	0.0054

5 おわりに

本稿では、食の地域性をマップとして可視化することを目標とし、レストラン検索サイトから得られるメニュー情報を統計的に分析することで食の地域性を抽出する方法を提案した。全63地域のレストラン情報を用いて分析指標を算出し、選出した16地域について食の地域性についての考察を行った。その結果、提案した分析指標を参照することで、各地域の特性となりうる食材や調理方法などが発見できる可能性が示唆された。

今後の展望として、外国人観光客に対して食の地域性を用いたおもてなしを可能にするべく、海外のレストラン検索サイトに対してもメニュー分析を行っていく。また、本稿で扱った63地域をさらに広範囲へ広げ様々な地域での分析や、N-gramではなく形態素解析によって分割された単語レベルでの分析なども行う。

謝辞

本研究では、ホットペッパー Web サービスから提供頂いたレストラン情報を用いた。本研究の遂行にあたり、立命館大学食マネジメント学部教授井澤裕司氏、立命館

大学大学院河合俊樹氏の協力を得た。本研究は、一部、すかいらーくフードサイエンス研究所の支援のもと行われた。記して、謝意を表す。

参考文献

- [1] 川久保秀敏, 柳井啓司: 地域別代表画像を用いた単語概念の地域性の分析, 研究報告コンピュータビジョンとイメージメディア (CVIM), vol.2011-CVIM-176, no.6, pp.1-8, 2011.
- [2] 中岡伊織, 中本晋太郎: 自己組織化マップおよび対応分析による口コミ情報を用いた地域性特徴分析, 日本知能情報ファジィ学会第29回ファジィシステムシンポジウム講演論文集, pp.891-894, 2013.
- [3] 福田哲也, 小野景子, 熊野雅仁, 木村昌弘: ソーシャルメディアデータに基づく主要観光地域の人気グルメ抽出, 第76回全国大会講演論文集, vol.2014, no.1, pp.577-578, 2014.
- [4] 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 (TOD), vol.5, no.3, pp.97-116, 2012.