Google Maps のクチコミを用いた美味しいものの発見

柿本航太郎† 井嶋 蒼† 横山 昌平†

† 首都大学東京システムデザイン学部システムデザイン学科 〒 191–0065 東京都日野市旭が丘 6-6
E-mail: †{kotaro.kakimoto,soxup0708}@gmail.com, ††shohei@tmu.ac.jp

あらまし 旅行や観光で食事をする際に，その地域における特産物やおすすめの料理を選ぶことが多い．しかし，SNS を通じて新たな流行が次々と生まれる現在，初めて訪れる地点での料理や食材の情報をすべて把握することは難しい．本研究では，Google Places API を用いて分析対象とする複数の都市からレストランやホテルなど様々なカテゴリのクチコミを取得し，それらの都市において注目されている食べ物を見つける手法を提案する．具体的には，得られたクチコミを地理軸・カテゴリ軸，双方において TF-IDF を適用し，単語の重要度を算出する．この仕組みにより，ユーザーの与えた場所における美味しいものを発見できる．さらに，結果をワードクラウドで効果的に可視化するシステムを実装した．

キーワード 地理情報，観光

1 はじめに

米国グーグル社が提供する地図情報サービスである Google Maps では，地図だけではなく，店の住所や電話番号・営業時間・クチコミなど，様々な情報を提供している．2017 年 10 月の時点では，日本で約 3300 万人近くのユーザが Google Maps を利用している．[1] また，スマホで地図アプリといったら「Google マップ」と答えるユーザーが 67 %もいることが，The Manifest の調査で明らかになっている [2]．ゆえに，Google Maps は，広く普及している位置情報システムであると言える．

また，人が旅行やその計画を立てる際に，何をどこで食べるかで悩むことがある．その際に，その旅行先での特産物やおすすめの料理を選ぶことが多い．しかし，その場所の情報をよく知らなかったり，あるいは自分の知識以上の事を知ろうとする場合，旅行雑誌などに頼ることがある．しかしながら，スマホが普及している現代において，新しい情報の多くは Web や SNS 上に存在する．そのため，我々は Web や SNS を用いて情報を探すことがあるが，インターネット上には莫大な量の情報が遍在しており，どのサイトにいっても，欲しい情報を得るまでに時間がかかってしまうことがある．例えば，Google Maps においては，特定の地域内の複数の飲食店を全て調べるには時間がかかり，すぐに特産物やおすすめの料理を把握するのは難しい．

以上より，我々は Google Maps 上で任意の地域を指定し，その範囲内にある多くの店のクチコミからその地域特有の美味しい料理や素材を抽出する手法を提案する．Google Maps ではローカルガイドと呼ばれるユーザーが日々大量のクチコミや写真を投稿している．図 1 に新宿のとある鰻屋のクチコミが並んだページの例を示す．図 1 にあらわれているように，クチコミには「鰻」や「肝吸い」などの食材や料理名などがよく現れている．一方で，「安い」や「オススメ」などの一般的な語も頻出するので，単に単語の出現回数だけを見ても美味しいものを発見するという本研究の目的は達成できない．

図 1 Google Maps 上でのクチコミの並んだページの一部

そこで，本研究では，Google Maps Platform 内にある Google Places API を用いる．Google Places API には様々な機能が存在する．キーワードを使って近くのスポットを検索する Text Search，指定した地点から周辺にあるスポットを検索する Nearby Search，名称や電話番号を使ってピンポイントに施設を検索することができる Find Place， PlaceID を指定してスポットの詳細情報を取得する Place Details などである．PlaceID とは Google Map 上に存在する施設一つ一つに与えられたテキスト形式の識別子のことである．Find Place と Place Details では価格によって取得できる情報が異なる．Find Place と Place Details でそれぞれの価格で取得できる情報を表 1 と表 2 に示す．

表 1 Find Place におけるそれぞれの価格で得られる情報

Type	Fields
Basic	formatted address, geometry, icon, id, name, permanently closed, photos, place id, plus code, scope, types
Contact	opening hours
Atmosphere	price level, rating

表 1 と表 2 より、価格を変えれば，その施設の住所や名前だけでなく，閉店時間やクチコミを得ることができることがわか

表 2　Place Details におけるそれぞれの価格で得られる情報

Type	Fields
Basic	address component, adr address, alt id, formatted address, geometry, icon, id, name, permanently closed, photo, place id, plus code, scope, type, url, utc offset, vicinity
Contact	formatted phone number, international phone number, opening hours, website
Atmosphere	price level, rating, reviews

る．また，Text Search では施設の Type を入力することで条件を絞って検索することができる．この Type の例としてはレストランやホテル，水族館などがある．Google Maps Platform のサイトにおける type の一覧が乗ったページの一部を図 (2) に示す．

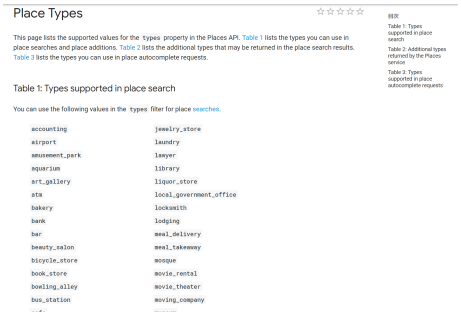

図 2　Google Maps Platform のサイトにおける type の一覧が乗ったページの一部

Google Places API を用いることで対象となる地域にあるレストランのクチコミだけでなく，複数の地域のデータ，またレストランだけでなくホテルのどの他のカテゴリにおけるクチコミを含めて分析することによって，地域ならではの美味しいものを発見する手法を提案する．さらに，得られた結果をワードクラウドで可視化することで，ユーザがその地域のを特徴づける「美味しいもの」を容易に把握できるシステムを実装した．

本論文の構成として，第 2 章で関連研究の紹介，第 3 章では手法で用いる TF-IDF のついての具体的な説明，第 4 章で提案手法についての説明，第 5 章で本手法の有効性を確かめるための実験の説明と結果の評価，第 6 章でまとめと今後の課題について述べる．

2　関 連 研 究

ネット上，または SNS 上のデータを解析して新たな情報を得るために，テキストを解析する技術についての研究が行われている．矢野ら [3] は Web 上に存在する店舗情報・評判情報から評価情報を抽出する研究を行った．川又ら [4] は，飲食店に対する口コミから料理に対する評価を行う研究を行った．どちらも口コミ内の複数の品詞から解析を行っているが，本研究では各料理名・食材名に着目しているため名詞のみを抽出して解析している．

また，GooglePlaceAPI を用いた研究も数多く行われている．新井ら [5] は，Google Places API と Yahoo!知恵袋質問検索 API から観光情報のデータセットを用意し，ユーザーの Twitter のつぶやきから観光ルートを推薦する研究を行った．

また，本研究では TF-IDF を用いて各単語に対するスコアを算出しているが，TF-IDF に対する研究も数多く行われている．影山らは [6] は，研究テーマのキーワードを Term，研究機関を Document とし，TF-IDF が張るベクトル値の内積計算を行うことで，研究機関の特徴抽出を試みた．宮崎ら [7] は，Web ページ集合に関する idf をより一般的な web ページ集合に対する idf 値に近似することと，クラスタ化された単語について全体のページの tf の近似を検討することで，話題判定に有効なキーワードを探す手法を提案した．本研究では，必要な情報を抽出するために 2 つの次元から成る TF-IDF の式を組み合わせて，単語のスコアを算出している．

3　TF-IDF

本研究では，TF-IDF を用いて単語の重要度を算出していく．TF-IDF とは，各文書内にある単語がどのくらい頻出したのかを示す TF（Term Frequency）とそれぞれの単語が全ての文書のうちでどのくらい共通して使われているかを示す IDF（Inverse Document Frequency）を掛けあわせて任意の単語の重要度を算出するものである．この重要度が高いほど，各文書を特徴づける語となる．

この章では具体的な例を用いて TF-IDF について詳しく説明していく．

3.1　TF

まず，TF について説明する．TF の式は以下のように式（1）のように表される．

$$tf = \frac{n_{t,d}}{\sum_{s \in d} n_{s,d}} \tag{1}$$

各値の説明については以下のようになる．

tf　：　文書 d 内のある単語 t の TF 値
$n_{t,d}$　：　ある単語 t の文書 d 内での出現回数
$\sum_{s \in d} n_{s,d}$　：　文書 d 内のすべての単語の出現回数の和

$n_{t,d}$ の値が大きいほど tf の値は大きくなる．つまり，文書内で単語が多く出現する程重要度が大きくなるということである．

3.2　IDF

次に，IDF について説明する．IDF の式は以下のように式（2）のように表される．

$$idf = log \frac{N}{df(t)} \tag{2}$$

各値の説明については以下のようになる．

idf　：　ある単語 t の IDF 値

N ： 全文書数
$df(t)$ ： ある単語 t が出現する文書の数

$df(t)$ の値が大きいほど，idf も値は小さくなる．つまり，多くの文書で使われている単語ほど重要度が小さくなるということである．

3.3 TFIDF

TFIDF は TF の値と IDF の値のかけ合わせによって求められるため，式（3）のように表される．

$$tfidf = tf * idf \tag{3}$$

TFIDF の値は，文書内においては頻出度は高いが他の文書においては頻出度が低い単語ほど高くなることになる．単に単語の頻出度が高いだけではその文書を特徴づける語にはなりえない．TFIDF は主に情報検索やトピック分析などの分野で用いられている．

3.4 具 体 例

次に具体例を用いて TFIDF について説明する．
2 つの文書 A，B があるとする．各文書の内容は以下のようとする．

文書 A :『犬　猫』
文書 B :『犬　犬　猿』

ここでの各単語における tf と idf の値は以下のようになる．

tf(犬, 文書 A)=$\frac{1}{2}$=0.5
tf(猫, 文書 A)=$\frac{1}{2}$=0.5
tf(犬, 文書 B)=$\frac{2}{3}$ ≒ 0.66
tf(猿, 文書 B)=$\frac{1}{3}$ ≒ 0.33

idf(犬)=$log\frac{2}{2}$=0
idf(猫)=$log\frac{2}{1}$ ≒ 0.3
idf(猿)=$log\frac{2}{1}$ ≒ 0.3

これより，各文書内における各単語の TFIDF の値は以下のようになる．

$tfidf$(犬, 文書 A)=tf(犬, 文書 A)*idf(犬)=0
$tfidf$(猫, 文書 A)=tf(猫, 文書 A)*idf(猫) ≒ 0.15
$tfidf$(犬, 文書 B)=tf(犬, 文書 B)*idf(犬)=0
$tfidf$(猿, 文書 B)=tf(猿, 文書 B)*idf(猿) ≒ 0.1

文書 A，文書 B ともに出現している『犬』という単語はどちらか一方だけに出現している『猫』,『猿』という単語より $tfidf$ が低いため，重要度が低いことがわかる．また『猿』という単語は『猫』という単語よりも各文書内における頻出回数が高いため，重要度は高いということがわかる．

4 提 案 手 法

本研究を行う上での全体の流れを図 3 に示す．

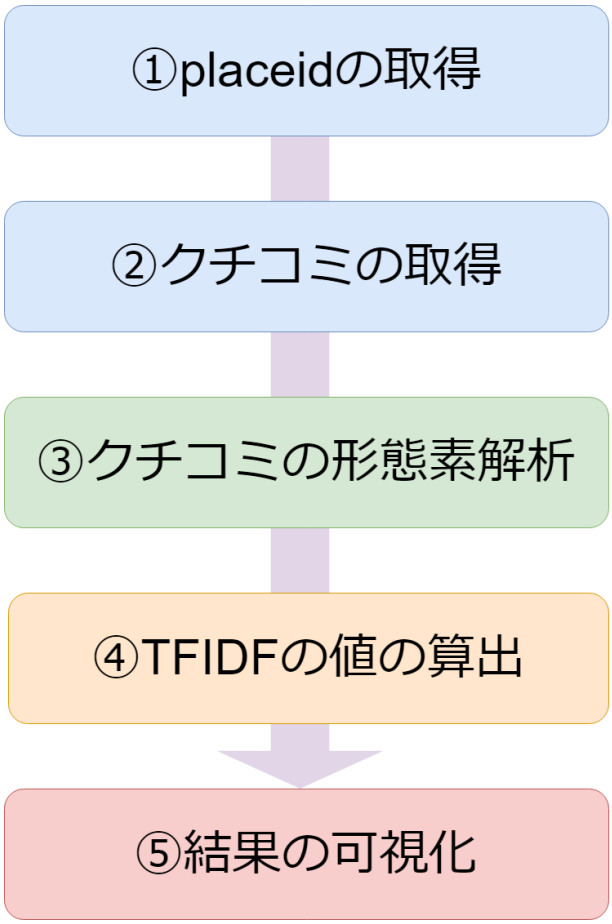

図 3　本研究の流れ

① placeid の取得と②クチコミの取得は Google Places API を用いて行う．具体的な収集方法については第 5 章にて記載する．③クチコミの形態素解析は②で収集したクチコミをデータセットとして mecab を用いて行う．④ TFIDF を算出については本章で具体的に説明する．⑤結果の可視化は wordcloud を用いて行う．

ある地域内におけるレストランのクチコミで頻出する特有の語を抽出するには，指定した地域の他に複数の地域におけるクチコミを収集し，TF-IDF を適用することで可能となる．しかし，得られた結果の中には食材や料理名だけでなく，本研究にとって不要な語，例えば地名や近隣の観光地などの単語も含まれることが予想できる．そこで，不要な語を排除するために，食材や料理名の語はあまり頻出されないが地名や観光地の単語はレストラン同様出現すると考えられる他のカテゴリにおけるクチコミに TF-IDF を適用する．これにより，不要な語の出現を抑えつつ美味しいものの単語のみを抽出することができる．

また本稿では，各地域ごとのレストランのクチコミから成る文書群を地理軸，同じ地域内のレストランとは違うジャンルにおけるクチコミから成る文書群をカテゴリ軸と呼ぶ．この地理

軸・カテゴリ軸における各文書の概観を図 4 に示す．

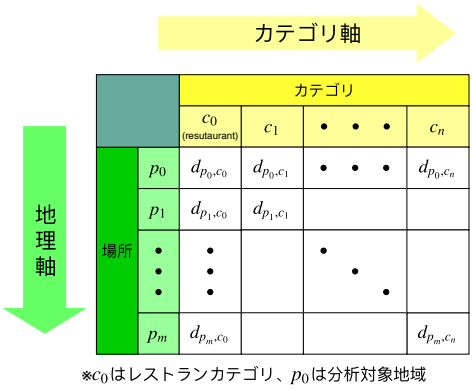

図 4　地理軸・カテゴリ軸における各文書の概観図

図 4 のように，地理軸においては分析対象地域以外の m 地点，カテゴリ軸においてはレストラン以外の n 種類のカテゴリのクチコミを収集する．ここで，分析対象地域は p_0，レストランは c_0 であり，クチコミ文書群を d_{p_0,c_0} として表している．そして，d_{p_0,c_0} に含まれる美味しいものを表す語を発見するためのスコア計算式を以下に示す．

$$S = S_{geo} * S_{cat} \tag{4}$$

提案手法におけるスコア S は，地理軸におけるスコア S_{geo} とカテゴリ軸におけるスコア S_{cat} の掛け合わせによって求められる．

次に，S_{geo} と S_{cat} で用いる TF-IDF による単語の重み付けについて説明する．

まず，地理軸上での単語の重み付けを行う．様々な場所におけるレストランにおけるクチコミ文書に TF-IDF を適用し，各地域における特産物やおすすめの料理の単語の重要度を算出していく．地理軸上レストランカテゴリにおける全文書 N_{c_0} 内の単語 t の TF-IDF の計算式は以下のように表される．

$$tf_{p_0,c_0} = \frac{n_{t,d_{p_0,c_0}}}{\sum_{s \in d_{p_0,c_0}} n_{s,d_{p_0,c_0}}} \tag{5}$$

$$idf_{p_0,c_0} = log\frac{N_{c_0}}{df(t)} \tag{6}$$

$n_{t,d_{p_0,c_0}}$ はある単語 t の文書 d_{p_0,c_0} 内での出現回数を表し，$df(t)$ はある単語 t が出現する地理軸上の文書の数を表す．地理軸における TF-IDF の式を利用した S_{geo} は式 (5) と式 (6) を用いて以下のように表される．

$$S_{geo} = tf_g * idf_g \tag{7}$$

式 (7) は一般的に使われる TF-IDF 値を求める計算式と同様である．

しかし，地理軸上の文書での TF-IDF を計算するだけでは，その地域の特産物やおすすめの料理といったものでなく，その地域の地名や料理以外の名詞など本研究を目的を達成する上では不要な語が上位に来る可能性がある．そこで，レストランのカテゴリにおいて顕著に現れる単語の重要度を上げるために，カテゴリ軸での TF-IDF の計算式を利用する．具体的に説明すると，クチコミ収集でレストランでは無い他のカテゴリにおける文書から出てくる語には，その地域の特色を表すが本研究上で不要な語が含まれる．その一方で，特産物やおすすめの料理の単語が出てくる可能性は低い．そのため，カテゴリ軸上での TF-IDF を利用して，必要な語のみ重要度が上がると良い．しかしながら，我々のもつカテゴリはレストランとホテルのみであり，TF-IDF をそのまま適応することができない．これはカテゴリ軸における idf が 2 つの値しか取らないため，単語の重要度を算出するには idf の値に影響されやすくなってしまうからである．そこで現時点ではレストランとホテルにおける tf のみを用いてカテゴリ軸のスコア S_{cat} を算出する．カテゴリ軸上における単語 t の tf の計算式は地理軸上における tf 値と等しいため，式 (5) と同じである．カテゴリ軸における tf の式を利用した S_{cat} は式 (5) を用いて以下のように表される．

$$S_{cat} = \frac{2 * tf_{p_0,c_0}}{\sum tf_{p_0}} - 1 \tag{8}$$

ここでの $\sum tf_{p_0}$ はカテゴリ軸におけるある単語 t の tf の値の総和を意味する．本研究では，カテゴリ軸をレストランとホテルの 2 つの文書から考えるため，式 (8) は式 (9) のようになる．

$$S_{cat} = \frac{2 * tf_{p_0,c_0}}{tf_{p_0,c_0} + tf_{p_0,c_1}} - 1 \tag{9}$$

ここで，c_0 はレストラン，c_1 はホテルを表している．また，S_{cat} は-1 から 1 の値をとる．つまり，同地域のレストランとホテルの文書のうちホテルでより見られる単語はマイナスの値，レストランとホテルでほぼ同じ割合で見られる単語には 0 に近い値が乗じられる．これにより，レストランのクチコミで頻出される料理名などのスコアは変えず，ホテルとレストランの両方で見られる地域名などのスコアは下げることができる．

また，得られた結果を基に WordCloud を用いて結果を可視化する．WordCloud は通常，形態素解析された単語の文書をテキストデータとして，結果を出力する．しかし，今回得られる結果は，式 (4) から算出される各単語のスコアを表す 0〜1 の間の数字となる．そこで今回は式 (4) から得られた結果を整数となるまで乗算し，その数字の数だけ対応する単語を表示した文書をテキストデータとして用いる．これを WordCloud に適応することで各単語の頻出度の比率から，それぞれの単語の文字の大きさを決定し可視化させた結果を得ることができる．

5　評 価 実 験

5.1　データセットの準備

本実験では，新宿（東京）・梅田（大阪）・札幌（北海道）・高知市（高知）・高松（香川）・博多（福岡）・那覇（沖縄）の 7 地点の繁華街を対象としたデータセットを用いる．クチコミのデータを収集する際，指定した地点を中心におよそ 1000 m× 1000 mの範囲のデータを収集する．ただし，Google Places API に

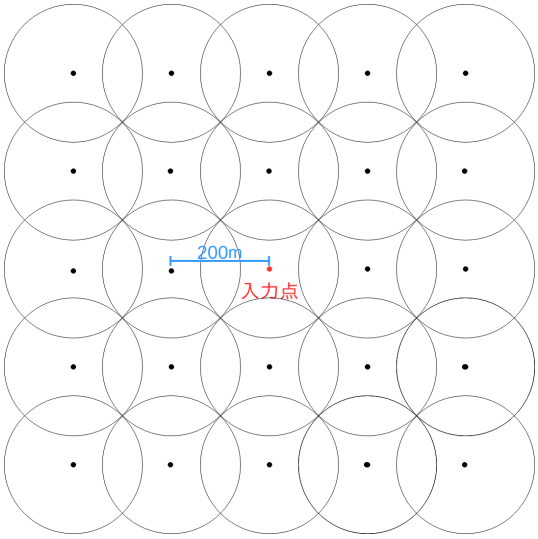

図 5　店のクチコミ収集時の範囲指定図

おけるクチコミを収集するクエリは地点と半径を入力にとり，それ対して 20 の店，かつ 1 つの店に対して 5 件のクチコミしか収集できない．そこで，クチコミを収集する際により多くの情報を得るために対象地域を細かく分割し，データを収集する（図 5）．

1 つの緯度経度を指定した後，そこから図 1 のように指定した地点の周りに 24 個の点をとる．各点から同じ半径内の領域で同様にクチコミを収集していく．特定の範囲内の店の選出は，おすすめ順に 20 個までを取るため，重複したデータは削除する．本実験では，各点の間隔を縦・横共に 200 mとするため，クエリの半径はおよそ 142 mとする．

今回集めた各地域のレストランの数及びクチコミの数を表 3 に示す．

表 3　各地域のレストランの数及びクチコミの数

場所	レストランの数	レストランのクチコミ数
新宿	94	449
梅田	118	550
札幌	121	570
高知	120	442
高松	107	476
博多	106	491
那覇	96	456

また，実験では本研究で提案した手法の有効性の検証を行うため，高松と札幌において地理軸のみを用いた結果と地理軸・カテゴリ軸の 2 つを用いた結果を比較する．高松と札幌におけるホテルの数及びホテルのクチコミ数を表 4 に示す．また，クチコミを収集する際に指定した Google Map における高松と札幌のそれぞれのデータ収集範囲における中心点を図 6 と図 7 に示す．

5.2　実験結果と考察

まずは，高松における地理軸のみの TF-IDF を計算し，上位 30 件を並べものの結果を図 8 に示す．また，その結果を

表 4　高松と札幌におけるホテルの数及びクチコミの数

場所	ホテルの数	ホテルのクチコミ数
札幌	32	160
高松	25	121

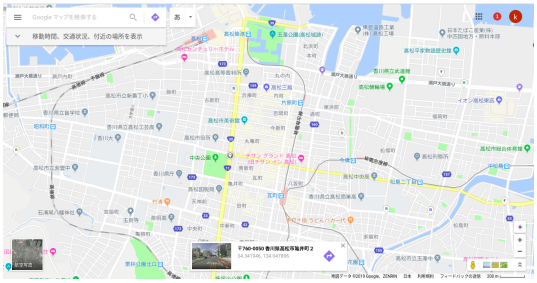

図 6　Google Map における高松のデータ収集範囲の中心点

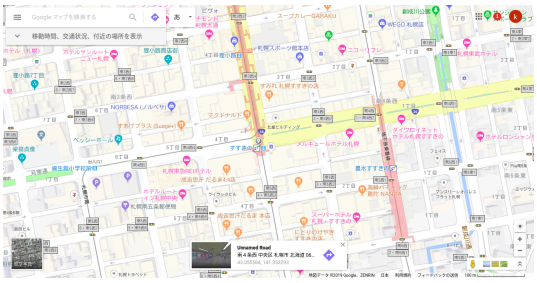

図 7　Google Map における札幌のデータ収集範囲の中心点

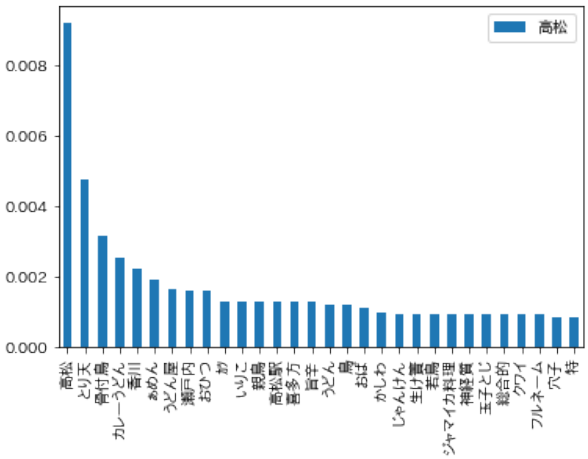

図 8　地理軸のみを用いた高松の単語重要度

wordcloud を用いて可視化したものを図 9 に示す．

上位に「とり天」や「骨付鳥」といった特有の料理名などが出ているが最重要単語に「高松」と出てしまい，的確においしいものを発見できた，とは言えない．これは，各地域におけるレストランのクチコミの文書を解析する際に，高松における結果では指定した場所が高松市であったり高松駅付近だったことによるものだと考えられる．しかし，他の地域におけるクチコミで「高松」という言葉がでてくることは基本的には考えられない．それゆえ，「高松」という単語の頻出度を計算した際に，他の地域に比べて tf_{p_0,c_0} 及び idf_{p_0,c_0} の値が共に際立って高くなり，最重要単語としてランキングの上位にあがってしまった．

ここで，料理名・食材名の単語のスコアを保持しつつ，「高

図 9　地理軸のみを用いた高松の結果を WordCloud で可視化したもの

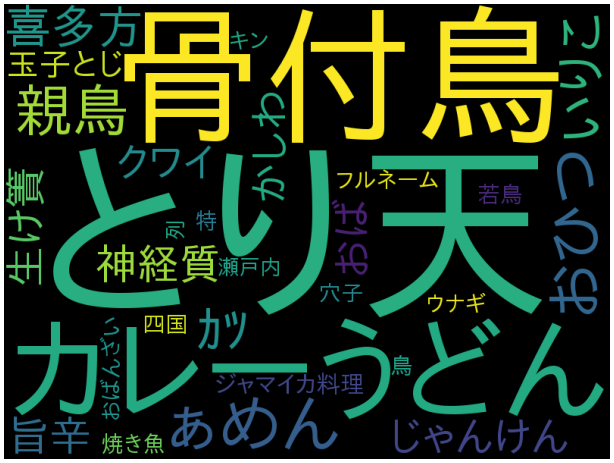

図 11　WordCloud を用いた高松の結果

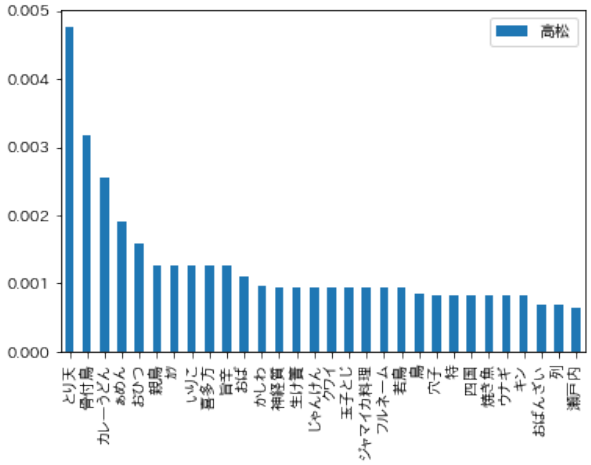

図 10　地理軸とカテゴリ軸を用いた高松の単語重要度

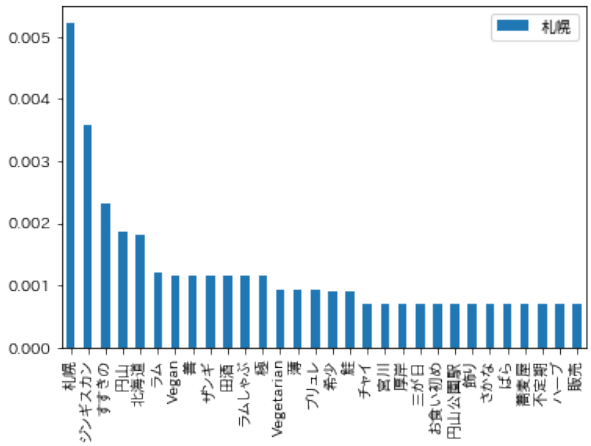

図 12　地理軸のみを用いた札幌の単語重要度

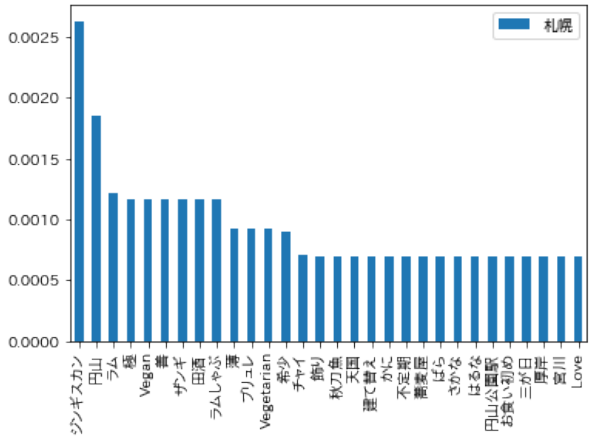

図 13　地理軸とカテゴリ軸を用いた札幌の単語重要度

松」を含んだ地名などのスコアを低くするためには，同じ地域内でレストラン以外の料理名・食材名の出ないカテゴリのクチコミの文書を利用すればよい．これが本研究の核となるアイデアである．今回はカテゴリ軸における他のジャンルとしてホテルのクチコミを利用した．これはホテルのクチコミの文書内において，「高松」という言葉はレストラン同様位置的な理由のもと頻繁に出現するが，ホテルは宿泊施設であるがゆえに料理や食材の単語は出現しにくいと考えられたからである．本実験での提案手法を用いた結果を図 10 に表示する．

図 10 では，図 8 で見られた「高松」といった地名の単語は見られなくなり，上位には料理の単語が並んでいるのがわかる．これは，先に述べたように，ホテルのクチコミ内で料理名・食材名が出現しない一方で「高松」という単語が頻繁に出現したことによるものである．しかし，上位には「ぁめん」という本研究の趣旨に沿わない料理名・食材名とは考えられない結果も出ている．これは，指定した範囲内にある「ごぞうらぁめん」という店の「旨辛らぁめん」や「味噌らぁめん」といった特殊なメニュー名によるものだと考えられる．これは，形態素解析において固有名詞を辞書に登録することで解決できる．

次に，得られた結果を WordCloud を用いて可視化した結果を図 11 に示す．図 11 より，指定した高松付近の地域において「とり天」や「骨付鳥」，「カレーうどん」が美味しいものとして挙げられることがひと目でわかる．

同様に，札幌での結果を比較したものを図 12 と図 13 に，また，手法を使う前と後での結果を WordCloud を用いて可視化した結果を図??と図 15 に示す．　図 12 では，「札幌」や「すすきの」といった地名が上位にきてしまっている．これは，高松同様指定した場所が札幌であったり，近くにすすきの駅があっ

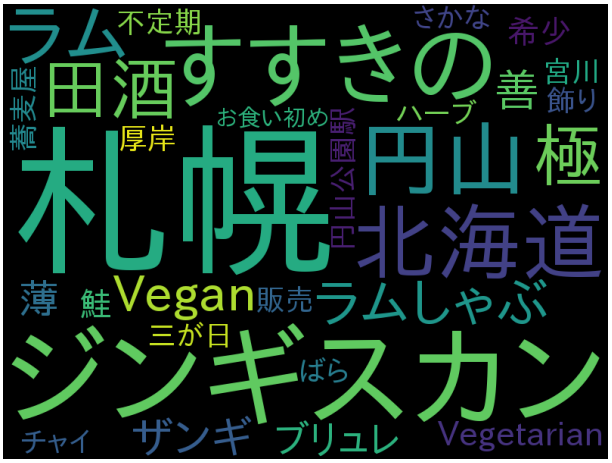

図 14　地理軸のみを用いた札幌の結果を WordCloud で可視化したもの

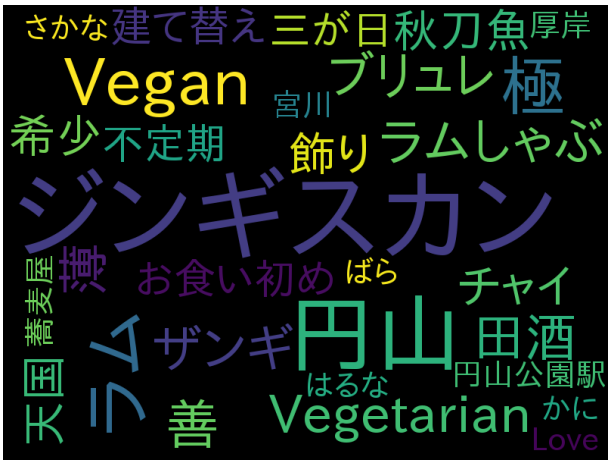

図 15　WordCloud を用いた札幌の結果

たことによるものであると考えられる．しかし，本手法を用いた図 13 では「札幌」や「すすきの」といった地名の単語は上位に見られなくなり，代わりに「ジンギスカン」や「ラム」など料理や食材の単語が上位にみられるようになった．「ジンギスカン」についで出てくる「円山」という単語は，「ステーキ円山」という店がクチコミ収集する際の対象の店として選ばれたためであり，クチコミ内に「円山」という単語が頻繁に出現することによって高いスコアが出たと考えられる．この問題については，クチコミの対象となる店名をストップワードに登録することで回避できる．また，図 15 より，札幌の対象の地域における美味しいものとして「ジンギスカン」や「ラム」などがあることがひと目でわかる．

以上の結果より，地理軸・カテゴリ軸の観点から TF-IDF を利用することで，その地域特有のおいしいものを発見することができる有効性を示すことができた．

6　ま　と　め

本論文では，Google Maps API を用いて，クチコミからその地域特有の美味しいものを発見するための手法を提案した．また，日本の 7 都市を対象としたデータセットを用いて，本手法の有効性を確認した．更に，得られた結果から WordCloud を用いて可視化を行った．

今後の課題として，以下の 7 点が挙げられる．

・カテゴリ軸のカテゴリ数の増加
・カテゴリ軸における *idf* の値の有効性の検討
・カテゴリ軸のレストランのジャンルを細分化
・スコアを算出する際のストップワードの検討
・形態素解析における『の』という単語についての検討
・word2vec の検討
・評価方法の検討

カテゴリ軸のカテゴリの数を増やすことにより，レストランのカテゴリにおける顕著に表れる単語の重要度をより上位にするための精度を上げることが可能になる．また，カテゴリ数を増やすことによって各単語の *idf* の値が結果に影響されやすくなるため，*idf* を利用した式の再検討を行う．

カテゴリ軸のレストランのジャンルを和食や中華，フレンチなど更に細分化することによって，各地域における様々なジャンルの特徴的な食事を検索することができることが可能になる．これはレストランに限らず，ホテルにおけるランクにも適用できる．

ストップワードの検討については，第 5 章で触れたように独特なメニュー名などが挙げられる．また，『カツオのたたき』という単語は形態素解析時に『カツオ』『の』『たたき』のように分解されるが，『カツオのたたき』はひとつの料理名としてそのまま抽出したい．そのため形態測定行う際に『の』という言葉を上手く扱いつつ単語が分解できるよう検討する．

また，word2vec を使い各単語をクラスタリングし，ユーザーが入力した地名の周りの単語を排除するといった TF-IDF を使用せず別の手法を用いた研究の検討，および本研究との精度の比較もする．

評価方法については今回，wordcloud で可視化した結果から判断していたが，観光雑誌や様々な地域の情報が記載されたサイトを利用した評価の検討も行う．

文　　献

[1] 2017 年日本におけるスマートフォンアプリ利用者数 top10. `https://www.netratings.co.jp/news_release/2017/12/Newsrelease20171219.html`. Accessed: 2018-12-27.
[2] Most popular navigation apps. `https://corriente.top/navigationapp-share/`. Accessed: 2018-12-27.
[3] 輝昭 相沢矢野 宏美. 嗜好を考慮した評判情報検索手法. 電子情報通信学会技術研究報告, 第 104 巻, pp. 13–18, 2004.
[4] 岡 誠 彦森 博川又 亜弓. 飲食店口コミサイトを対象にした評価文抽出. 電子情報通信学会技術研究報告, 第 115 巻, pp. 47–51, 2015.
[5] 太田 学新井 晃平. Twitter を利用した観光ルート推薦の一手法. In *DEIM Forum 2015*, 第 8 巻, 2015.
[6] 辻 洋景山 明宣. TF/IDF アルゴリズムを用いた研究機関の特徴抽出法. 電気学会論文誌, pp. 713–719, 2005.
[7] 川端 豪宮崎 将隆. Wikipedia ページへの tfidf 法の適用. 研究報告音声言語情報処理, pp. 1–6, 2009.