

アナロジーによる飲食店メニューの 国際的な表記方式の提案とその自動生成

信本 健輔[†] 加藤 大受^{††,†††} 遠藤 雅樹^{††††} 莊司 慶行[†] 廣田 雅春^{†††††}
石川 博^{††}

[†] 首都大学東京 システムデザイン学部 〒191-0065 東京都日野市旭が丘 6-6

^{††} 首都大学東京大学院 システムデザイン研究科 〒191-0065 東京都日野市旭が丘 6-6

^{†††} ウイングアーク1st株式会社 〒150-0031 東京都渋谷区桜丘町 20-1 渋谷インフォスタワー

^{††††} 職業能力開発総合大学校 基盤ものづくり系 〒187-0035 東京都小平市小川西町 2-32-1

^{†††††} 大分工業高等専門学校 情報工学科 〒870-0152 大分県大分市大字牧 1666

E-mail: [†]nobumoto-kensuke@ed.tmu.ac.jp, ^{††}y-shoji@tmu.ac.jp, ^{†††}ishikawa-hiroshi@tmu.ac.jp,

^{†††††}kato.d@wingarc.com, ^{†††††}tendou@uitec.ac.jp, ^{††††††}m-hirota@oita-ct.ac.jp,

あらまし 本稿では、飲食店メニューの説明欄の新たな表記方式とその自動生成手法を提案する。本稿で提案する表記方式は、「お好み焼き = チヂミ - ニラ + キャベツ」のように、任意の料理を、ある国の類似した料理名と差分となる要素によって単純な数式で表したものである。また、本稿では、この表記方式を、投稿型レシピサイトのデータを用いて自動生成する手法も提案する。レシピデータに含まれる食材、調理工程などの6つの要素を用いて、類似した料理の抽出と、差分となる要素の算出を行う。本稿では、クックパッドのレシピデータを用いて、本表記方式の左辺を日本料理とした場合の結果を被験者実験を通じて、本表記方式の有用性と、自動生成の性能について評価を行った。

キーワード レシピサイト, 類似料理, 国際化

1. はじめに

海外への旅行者が飲食店を訪れた際、メニューを見ても、料理の概要を把握することができない場合がある。これは、飲食店がその母国語のみでメニューを記載している場合に、旅行者がその言語を理解できない場合に生じる。その場合、旅行者は、母国語で書かれていない料理の説明文を理解することは困難で、そもそも料理名すら読めないこともある。例えば、インターネットで調べようとしても、必ずしもインターネットを利用可能な状況であるとは限らない。このような状況では、旅行者は自身の意図する料理を注文することが困難である。

近年、日本は、外国人向けの訪日旅行の促進活動に力を入れている。その結果として、2016年の10月の段階で年間の訪日外国人数が2,000万人を超えた^(注1)。2020年に東京オリンピックの開催も予定しており、訪日外国人の増加傾向は今後も続くと考えられる。国土交通省観光庁が、訪日外国人に「訪日前に期待していたこと」、「訪日前に最も期待していたこと」をそれぞれ調査^(注2)したところ、どちらも最も多かったのが「日本食を食べること」(73.1%と25.6%)であった。このことから、訪日外国人は日本食に対して高い関心があることがわかる。

しかし、訪日外国人が日本の飲食店を訪れた際、前述した問題が起きている。国土交通省観光庁は、訪日外国人の日本の

受け入れ環境に対する不満のアンケート調査^(注3)を行った。その結果、訪日外国人が「旅行中困ったこと」として最も多くあげたのが、「無料公衆無線LAN環境」(36.7%)であった。また、「コミュニケーションに困った場所・場面」で最も多くあげたのが「言語一般」(49.3%)、2番目が「飲食」と「交通」(11.3%)であった。「不満に感じた点に関する主な回答」には、「レストランでは英語のメニューがなければ注文できない」という不満もあげている。

飲食店に、「無料公衆無線LAN環境」が整っていれば、「言語一般」の問題は機械翻訳、「飲食」の問題もインターネット検索を用いれば解決されることも多いと予想され、実際に総務省が訪日外国人にとって利用しやすいICT利用環境を実現する取り組み^(注4)を行っている。しかし、単純に翻訳した結果を提示しても、正確に翻訳されていない場合は、旅行者が料理を理解できないことも考えられる。

外国人の観光客を受け入れたい飲食店は、メニューを複数の言語で記載することは効果的な方法の一つであると考えられる。しかし、これを実現するためには多くの課題がある。従来の飲食店メニューの料理名や説明欄は、ほとんどが母国語のみの文章で記載されている。そのため、複数の言語の説明文を作成するためには、母国語で記載された説明文から翻訳する必要がある。しかし、これには、翻訳する言語の単語や文法の知識が必

(注1)：国籍/月別 訪日外客数 (2003年 - 2016年) http://www.jnto.go.jp/jpn/statistics/since2003_tourists.pdf

(注2)：訪日外国人消費動向報告書 (H28.7-9) <http://www.mlit.go.jp/common/001149546.pdf>

(注3)：外国人旅行者に対するアンケート調査結果について <http://www.mlit.go.jp/common/000190659.pdf>

(注4)：訪日外国人のICT利用環境整備に向けたアクションプラン http://www.soumu.go.jp/main_content/000296265.pdf

要となる。また、機械翻訳器を用いて翻訳する方法も存在するが、上でも述べたように、正確に翻訳されない可能性がある。さらに、予備知識の無い言語への翻訳であれば、その翻訳結果が正確な翻訳かどうかの判断は困難である。したがって、複数の言語で記載されたメニューを作成することは、外国人旅行者を獲得する飲食店にとって重要であるが、その作成は困難である。本稿では、多言語に容易に拡張可能かつ、類似した料理から容易に類推可能な、アナロジーによる国際的な表記方式を提案する。

提案する表記方式を生成するためには、説明したい料理と、アナロジーに用いられる料理の両方に関する様々な知識が必要であり、これを人手で行うのは難しい。そこで、本稿では、ユーザ投稿型レシピサイトのデータを用いて、この表記を自動生成する手法も提案する。近年、クックパッド^(注5)や、楽天レシピ^(注6)に代表されるユーザ投稿型レシピサイトが普及している。中でも日本最大のユーザ投稿型レシピサイトであるクックパッドは、現在 250 万品を超える投稿レシピデータを保持している。これは日本の料理のレシピのみでなく、世界中の料理のレシピが含まれているため、このレシピデータを用いる。

以下、2. 節では提案するアナロジーによる国際的な表記方式について述べる。3. 節では、提案する表記方式を、ユーザ投稿型レシピサイトのレシピデータから生成する手法について述べる。4. 節では、3. 節で述べる手法を用いて行った実験について述べる。5. 節は、考察を述べる。6. 節では、関連研究について述べる。7. 節では、本稿のまとめを述べる。

2. アナロジーによる国際的な表記方式の提案

本稿では、飲食店のメニューを複数の言語へ拡張を容易にするため、アナロジーを用いた表記方式を提案する。提案する表記方式は、

$$\text{お好み焼き} = \text{チヂミー} - \text{ニラ} + \text{キャベツ} \quad (1)$$

のように、左辺のある国の任意の料理を、別の国の類似した料理名と、それらの料理の差分となる要素によって数式で表したものである。この表記方式は以下の 2 つの利点がある。

- 翻訳の容易化
- 読者の負担の軽減

これらの利点により、飲食店のメニューを複数の言語へ容易に拡張できる。類似した料理名に翻訳先の国の料理を用いて式を生成し、翻訳したものを列挙することで、図 1 のような国際的なメニューを作ることができる。

2.1 翻訳の容易化

本稿で提案する表記方式は、従来の文章による表記方式よりも、簡潔な表記方法を採用するため、翻訳を容易に行うことが可能であると考えられる。これは、「ニラ」や、「キャベツ」などの単語のみで表記するため、文の解釈が必要なく、ドメインを料理に絞った翻訳が可能であるためである。

(注5)：レシピ検索 No.1 /料理レシピ載せるなら クックパッド <https://cookpad.com/>

(注6)：楽天レシピ -料理レシピ検索サイト <http://recipe.rakuten.co.jp/>



図 1 提案する表記方式を用いた飲食店メニューのイメージ

2.2 読者の負担の軽減

提案する表記方式は、式 1 のように記述量が少ないため、読者が読む必要のあるメニューの説明の分量が少ない。従来の文章による料理の説明文の場合、料理を詳細に説明するためには文書量を増加させる必要があるが、読者の立場では、読む分量が増加するため負担である。一方、読む分量を少なくするために、単純に説明文を短くすると、料理を説明するための情報量が減ってしまい、読者は十分にその料理を理解できない可能性がある。しかし、提案する表記方式は、読者が知っていると想定される類似した料理名を未知の料理名の説明に用いるため、記述量を最小限にでき、読者の読む量を減らすことができる。さらに、説明したい料理の情報の多くを類似した料理名に代替させることで、文書量の削減に伴う情報量の低下を抑えることが可能である。

3. アナロジーによる国際的な表記方式の自動生成手法の提案

本節では、本表記方式の自動生成手法を提案する。本手法は、ユーザ投稿型レシピサイトのレシピデータを用いて、

- (1) 料理の特徴ベクトルの生成
- (2) 類似した料理の抽出
- (3) 差分の出力

の手順から、式の生成を行う。

3.1 料理の特徴ベクトルの生成

ここでは、レシピ特有の単語を用いた料理の特徴量を求める方法を述べる。料理は、食材や、調理工程、味や、調理器具などの様々な要素から成り立っていると考えられる。そこで、本研究では、レシピデータからレシピ特有の単語を抽出し、それらを食材や調理工程などのカテゴリに分類する。その後、カテゴリごとに、単語の出現頻度を素性とする特徴ベクトルを生成する。以下に特徴ベクトルの生成の手順を示す。

- (1) 各国の料理リストの作成
- (2) レシピ特有の単語の抽出
- (3) レシピ特有の単語のカテゴリ分け
- (4) 特徴ベクトル生成

3.1.1 各国の料理リストの作成

本稿で用いる各国の料理名のリストは、あらかじめ作成する。実際の手順として、Web サイトや書籍から、各国の料理名を取

レシピ特有の単語	レシピ特有意度
中火	8142.1
しんなり	4883.1
溶く	4735.5
流し入れる	3383.6
ゆでる	2584.3
炒める	2387.1
乱切り	2355.3
弱火	2318.0
泡立てる	2280.7
まぜる	1892.9

集し、1つのリストにする。出力結果に含まれる別の国の料理名は、このリストに含まれている必要がある。説明したい料理(式1の左辺)に類似した別の国の料理(式1の右辺の最も左の項)をこのリスト内から探索する。これ以降では、説明したい料理をクエリと記す。

3.1.2 レシピ特有の単語の抽出

本研究では、レシピ特有の単語を、レシピデータの調理手順の文章に出現する単語の中で、一般的な文章ではあまり出現しないが、調理手順の文章には頻出する単語とする。このレシピ特有の単語を抽出するために、レシピデータに加え、一般的な文章として Wikipedia^(注7) のデータを用いる。レシピデータの調理手順と Wikipedia の文章に対して形態素解析を適用し、レシピデータの調理手順と Wikipedia のどちらにも出現する単語 w のそれぞれのレシピ特有意度を示すスコア $sc_{df}(w)$ を

$$sc_{df}(w) = \frac{df_{recipe}(w)}{df_{wp}(w)} \quad (2)$$

より求める。このとき、 $df_{recipe}(w)$ は、 w が出現するレシピデータの調理手順の数、 $df_{wp}(w)$ は、 w が出現する Wikipedia のページ数とする。 $sc_{df}(w)$ の上位 500 件の単語をレシピ特有の単語として扱う。

3.1.3 レシピ特有の単語のカテゴリ分け

表1に抽出したレシピ特有の単語の上位10件を示す。表1に含まれる、レシピ特有の単語は、名詞や動詞などが混在している。これらのレシピ特有の単語を、食材や調理工程などの数種類のカテゴリに人手により分類する。

3.1.4 特徴ベクトルの生成

3.1.1節で生成した、ある国の料理リストの各料理名の特徴ベクトルを生成する。このベクトルはカテゴリ分けしたレシピ特有の単語の出現頻度を素性とする。 C 種類のカテゴリに含まれるカテゴリ c ($1 \leq c \leq C$)、カテゴリ c のレシピ特有の単語 $w_{c,m}$ ($1 \leq m \leq M$)、ある国の料理名リストの料理名 l ($1 \leq l \leq L$) を用い、カテゴリ c での料理名 l の特徴ベクトル $\mathbf{v}(c, l)$ を

$$\mathbf{v}(c, l) = (a_1, a_2, a_3 \dots, a_m, \dots, a_M) \quad (3)$$

$$a_m = \frac{df(w_{c,m}, l)}{|N_l|} \quad (4)$$

より生成する。ここで、 $df(w_{c,m}, l)$ は l をレシピ名に含むレシピデータのうち $w_{c,m}$ を調理手順の文章に含むレシピデータ数、 $|N_l|$ は l をレシピ名に含むレシピデータの全データとする。また、上式の l をクエリの料理名 q と置き換え、クエリの特徴ベクトル $\mathbf{v}(c, q)$ も、同様に生成する。

3.2 類似した料理の抽出

3.1節で生成した特徴ベクトルを用い、料理名リストの各料理名とクエリとの類似度を求める。3.1節で述べた、料理を特徴づける要素の重要さは異なると考えられる。例えば、食材と調理器具を比較すると、食材の方が重要であると考えられる。そこで、本研究では、どのカテゴリが類似した料理の抽出に重要かを示す重要度を予備実験により求める。その重要度を各ベクトルの重みとして類似度の計算を行い、類似した料理を最終的に抽出する。カテゴリ c での料理名 l とクエリ q の類似度 $sim_c(l, q)$ を各特徴ベクトル $\mathbf{v}(c, l)$ と $\mathbf{v}(c, q)$ のコサイン類似度とする。この $sim_c(l, q)$ と、カテゴリ c の重要度 α_c を用い、各料理名 l とクエリ q の類似度 $sim_{all}(l, q)$ を

$$sim_{all}(l, q) = \prod_{i=1}^C sim_i(l, q)^{\alpha_i} \quad (5)$$

より求める。ここで用いる重要度 α_c は、予備実験により求めたものである。予備実験は、4.2.3節で述べる。得られた $sim_{all}(l, q)$ が最も高い l をクエリ q に類似した料理 s として抽出する。

3.3 差分の出力

クエリの要素と、類似した料理の要素の差分を抽出し、式で出力をする。本研究では、抽出する要素を食材に絞る。3.1節で生成した食材カテゴリの特徴ベクトルは、食材の出現頻度を素性としている。したがって、クエリと類似した料理の食材カテゴリの特徴ベクトルの差を求めることで、どちらか片方の料理でのみ頻出な食材を抽出できると考えられる。クエリ q と、類似した料理として得られた料理名 s との、差分となる要素を抽出し、式で出力する。差分の抽出に、食材カテゴリの特徴ベクトル $\mathbf{v}(s)$ と $\mathbf{v}(q)$ を用いる。 s と q の差分となる要素を抽出するためのベクトル $\mathbf{v}_{diff}(s, q)$ を

$$\mathbf{v}_{diff}(s, q) = \mathbf{v}(s) - \mathbf{v}(q) \quad (6)$$

より生成する。得られた $\mathbf{v}_{diff}(s, q)$ の値の絶対値が大きい順に P 件の要素 w_p ($0 \leq p \leq P$) を、 s と q の料理の差分のとなる要素として抽出する。これらの要素を用いて式1のような出力、

$$q = s + w_1 + w_2 + \dots + w_p + \dots + w_P \quad (7)$$

$$w_p = \begin{cases} w_p & (a_p < 0) \\ 0 & (a_p = 0) \\ -w_p & (a_p > 0) \end{cases}$$

を生成する。この a_p は要素 p の値を指す。

(注7) : ウィキペディア - Wikipedia <https://ja.wikipedia.org/wiki/>

表 2 各国の料理リストの料理名数

国名	料理名数
韓国	59
イタリア	45
アメリカ	44

4. 評価実験

本実験では、レシピデータを用いて、3. 節の手法より、2. 節で提案した表記方式を生成した。その結果から、被験者実験を通して、本表記方式の有用性と、本手法の性能について評価した。また、本手法の比較対象として、既存の単語の定量化手法である Word2Vec^(注8)を用いた。

4.1 データセット

本実験は、クックパッド株式会社と国立情報学研究所が研究者に提供している「クックパッドデータ」^(注9)を用いた。これは、クックパッドで2014年9月30日までに公開された約171万件のレシピデータが含まれている。また、3.1.1節で説明した料理名リストは、韓国、イタリア、およびアメリカの3カ国を、それぞれWebサイト^(注10)、書籍^(注11)、およびWikipediaから収集した。このとき、クックパッドのレシピデータに含まれていない料理名は取り除いた。収集した各国の料理リストの料理名数を表2に示す。また、レシピ特有の単語の抽出には、日本語版Wikipediaの2016年9月20日時点での全ページの記事本文^(注12)を用いた。

4.2 実装

以下で、本手法と、Word2Vecを用いた手法の実装について説明する。また、各カテゴリの重要度を求めるために行った予備実験についても説明する。

4.2.1 本手法の実装

4.1節のデータセットに本手法を用い、本表記方式を生成した。本実験は、日本の飲食店のメニューに本表記方式を適用した場合を想定し、式1の左辺を日本料理とした出力を生成した。3.1.2節の文章の形態素解析には、MeCab^(注13)を用いた。3.1.3節のカテゴリは、食材、調理工程、調味料、調理器具、食材を修飾する語、および調理工程を修飾する語の6種類($C=6$)のカテゴリに分類した。その際に、「cm」のような単位や、「ID」のような、どのカテゴリにも属さない単語は除外した。本実験では、上記のようにカテゴリ分けしたが、ここにあげているカテゴリ以外にも、シズルワード[1]や、レシピ用語

タグ[2]などもカテゴリとして含めた方が有効な可能性もある。

各カテゴリの5単語を表3に示す。表3のとおり、食材を修飾する語とは、「しんなり」や「ひたひた」のような、調理中の食材の状態を表すような単語を指し、また、調理工程を修飾する語とは、「中火」や「弱火」のような調理中の動作の程度を表す単語や、「乱切り」や「小口切り」のような、動作の中でも特有の名が付いている単語を指す。

4.2.2 Word2Vecの適用

本手法は、左辺のクエリとなる料理1単語を入力とし、右辺で複数単語の加減算による説明を出力するものであるが、同様な出力を行う既存手法は存在しない。しかし、類似する既存手法としてWord2Vecがあげられる。Word2Vecは、「東京 - 日本 + フランス = パリ」のように単語の概念について加減算することを目指した手法である。本手法も、料理名を単語のベクトルとして表現しており、これと類似している。本手法は1単語の入力に対し、複数単語で出力するが、Word2Vecではこのようなことができないため単純な比較はできない。しかし、本手法の一部である、クエリに類似した他国の料理を抽出する点においては、Word2Vecでも可能であるため、この点においてのみWord2Vecを本手法との比較対象に用いる。

本実験では、Word2Vecの単語の持つ意味を加減算できるという特性を生かし、「お好み焼き - 日本 + 韓国 = チヂミ」のようにして用いる。これは、ある国の料理名から国の要素を取り除き、出力として得たい類似した料理の国名を加えることで、クエリに類似した他国の料理の抽出が可能だと考えた。これにより得られた結果と、本手法の類似した料理の抽出で得られた結果とを比較した。コーパスには、4.1節で述べた日本語版Wikipediaのデータを用いた。入力する単語の長さを示す文脈窓の大きさと、単語の最小登場回数は、どちらも5に設定した。また、ベクトルの次元数は、100とした。

4.2.3 各カテゴリの重要度決定のための予備実験

予備実験により、各カテゴリに重み付けするための重要度を決定した。3.1節の手順で生成した、各カテゴリの特徴ベクトル単体で、類似した料理を抽出した際の結果を評価し、各カテゴリの重要度を決定した。この評価は、被験者実験により生成した正解データを用いて行った。この予備実験には、クエリとして10件の日本料理と、イタリア料理リストを用いた。また、被験者は3名の日本人とした。

以下で、各カテゴリの重要度の決定手順を述べる。まず被験者は、イタリア料理リストの各料理が、クエリと類似しているかを「はい」、「いいえ」の2択で回答する。その際に、被験者にとって、未知の料理名があった場合を考慮し、インターネット検索により、料理を調べても良いとした。3名の被験者のうち2名が似ていると感じた料理の組み合わせを、正解データとする。次に、各カテゴリの特徴ベクトルより、各クエリとイタリア料理リストの各料理のコサイン類似度を求め、各カテゴリ毎に上位5件のランキングを生成する。各カテゴリ毎に生成したランキングに含まれている正解データ数を正規化したものを、各カテゴリの重要度とする。表4に各カテゴリのレシピ特有の単語数と重要度を示す。

(注8) : Word2Vec : Tool for computing continuous distributed representations of words. <https://code.google.com/archive/p/word2vec/>

(注9) : 情報学研究データリポジトリ クックパッドデータセット - 国立情報学研究所 <http://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html>

(注10) : 韓国料理のレシピ - 新大久保 コリアンタウン情報なら WOW 新大久保 <http://www.wowsokb.jp/recipe/>

(注11) : 世界の料理・メニュー辞典—最新海外旅行者のための — 学研辞典編集部

(注12) : 日本の Wikipedia 情報ダウンロードページ <https://dumps.wikimedia.org/jawiki/latest/>

(注13) : MeCab: Yet Another Part-of-Speech and Morphological Analyzer <http://taku910.github.io/mecab/>

表3 6種にカテゴリ分けしたレシピ特有の単語 上位5語

食材		調理工程		調味料		調理器具		食材を修飾する語		調理工程を修飾する語	
word	score	word	score	word	score	word	score	word	score	word	score
薄力粉	1760.1	溶く	4735.5	こしょう	842.1	フライパン	1068.7	しんなり	4883.2	中火	8142.1
片栗粉	1582.6	流し入れる	3383.6	コンソメ	704.9	お鍋	636.4	ひたひた	1290.7	乱切り	2355.3
サラダ油	1287.9	ゆでる	2584.3	コショウ	581.3	麵棒	605.5	水溶き	1242.4	弱火	2318.0
ホットケーキミックス	1260.2	炒める	2387.1	しょう油	466.1	トースター	569.0	焦げ目	1137.2	小口切り	1779.9
ごま油	1152.2	泡立てる	2280.7	オイスターソース	422.3	菜箸	490.1	水気	1045.4	手早い	1694.8

表4 レシピ特有単語 500語をカテゴリ分けした時の各カテゴリのレシピ特有の単語数と重要度

カテゴリ	単語数	重要度
食材	122	0.11
調理工程	194	0.28
調味料	26	0.17
調理器具	27	0.13
食材を修飾する語	58	0.17
調理工程を修飾する語	55	0.13

4.3 被験者実験による評価

本手法では、以下について評価するために被験者実験を行った。

- 類似した料理の抽出の性能
- 差分の出力の性能

被験者は5名で、クエリは、代表的な日本料理である「寿司」、「刺身」、「焼き肉」、「ラーメン」、「うどん」、「天ぷら」、「とんかつ」、「カレーライス」、「肉じゃが」、および「お好み焼き」の10件を用いた。

4.3.1 類似した料理の抽出の性能評価手順

本手法と4.2.2節で述べたWord2Vecの手法より、各クエリに類似した料理を、3カ国の料理リストから抽出する。被験者は、本手法とWord2Vecにより抽出された料理がクエリと類似しているかを5段階（1 = 大変悪い、2 = 悪い、3 = 普通、4 = 良い、5 = 大変良い）で評価し、その結果を考察する。

4.3.2 差分の出力の性能評価手順

3.3節で述べた差分の要素数 p を $p = \{0, 1, 2, 5, 8\}$ と設定し、本手法より、各クエリに対して、本表記方式を生成する。被験者は、生成された表記方式それぞれが

- クエリをイメージしやすいか（イメージのしやすさ）
- クエリの説明として正確か（正確さ）

の2つの観点を5段階（1 = 大変悪い、2 = 悪い、3 = 普通、4 = 良い、5 = 大変良い）で評価し、考察する。

5. 考察

5.1 料理の特徴ベクトルの考察

表4より、レシピ特有の単語をカテゴリ分けすると、食材と調理工程に分類された単語が多く、食材を修飾する語や調理工程を修飾する語が少ないことがわかる。これは、ユーザが調理手順を記す際、食材や、調理工程は必ず記すが、調理手順をより詳細に説明するために用いられる、食材を修飾する語や、調

表5 提案手法とWord2Vecを用いた手法による類似した料理抽出性能の評価の平均値

手法	平均値
提案手法	3.31
Word2Vec	1.43

理工程を修飾する語は、ユーザによって記載の有無が異なるためだと考えられる。

5.2 類似した料理の抽出の考察

表5に本手法とWord2Vecを用いた、類似した料理の抽出性能の評価の平均値を示す。表5より、本手法はWord2Vecを用いた手法より平均値が高い。Word2Vecの平均値が低いのは、Word2Vecが、「類似した文脈で利用される単語は、類似した意味を持つ」という仮説に基づいており、類似した料理の抽出において重要とされる食材や、調理工程を考慮していないためだと考えられる。また、表5より、本手法の5段階評価の平均値が3を超えている点から、適切な類似した料理が抽出できていることがわかる。

表6に、クエリに類似した料理と被験者による5段階評価の平均値を示す。表6より、同じクエリでも国によって適切な類似した料理が抽出できている場合とできていない場合があることもわかる。例えば、刺身は、イタリア料理では類似した料理を抽出できているが、アメリカ料理ではできていない。これは、刺身に類似した料理がイタリア料理リストには存在しているが、アメリカ料理には存在していないためだと考えられる。これを改善するためには、料理リストの料理名数を増やさせる必要があると考えられる。

また、本実験では、国別で料理名リストを作成したが、メニューを読む旅行者にとって既知の料理が類似した料理として用いられれば良いため、より広範囲での料理名リストや、世界中でよく知られている料理リストを生成し、本手法に適用すれば、より高性能な自動生成が行えると考えられる。

5.3 差分の出力の考察

図2に、4.3.2節より得られた、要素数 p を $p = \{0, 1, 2, 5, 8\}$ としたときの、イメージしやすさと正確さの評価の平均値を示す。図2より、要素数 p が増加するほどイメージのしやすさが下がっていることがわかる。これは、要素数 p が増えることにより、読者が処理する情報量が増え、クエリのイメージが困難になるためだと考えられる。

表 6 本手法により抽出した類似した料理とその評価の平均値

クエリ	イタリア		韓国		アメリカ	
	類似した料理	評価	類似した料理	評価	類似した料理	評価
寿司	カルパッチョ	3.8	キムパブ	4.4	カリフォルニアロール	4.0
刺身	カルパッチョ	4.8	ユッケ	3.0	コブサラダ	1.4
焼き肉	スパゲッティ	1.2	プルコギ	4.2	ポークチョップ	3.8
ラーメン	スパゲッティ	3.4	ちゃんぽん	4.6	チキンスープ	3.2
うどん	スパゲッティ	3.2	ちゃんぽん	3.8	スパゲッティ	3.2
天ぷら	フリット	4.2	キンパ	1.4	フライドチキン	3.2
とんかつ	フリット	3.6	サムギョプサル	2.8	ポークチョップ	3.4
カレーライス	ミネストローネ	2.8	カムジャタン	3.0	チャウダー	3.0
肉じゃが	カポナータ	2.8	カムジャタン	4.0	チキンスープ	2.4
お好み焼き	フリッタータ	3.2	チヂミ	4.6	トルティーヤ	3.0

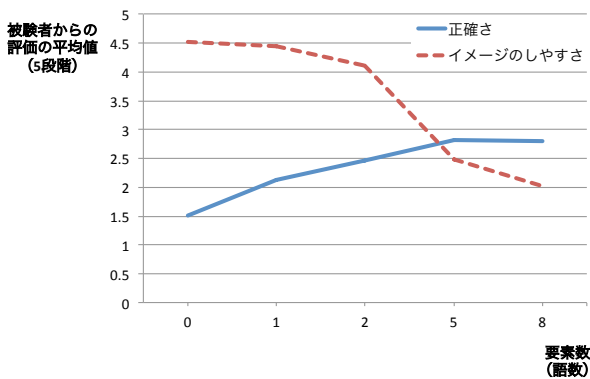


図 2 要素数 p を $p = \{0, 1, 2, 5, 8\}$ としたときのイメージのしやすさと正確さの評価の平均値

また、図 2 より、要素数 p が、増加するにあたり、クエリの説明の正確さも上がっていることがわかる。これは、類似した料理のみでは説明として不十分であるが、差分となる要素が、説明として不足している点を、うまく補う働きをしたためだと考えられる。また、要素数 p が 5 から 8 に変化する際に、正確さが減少しているのは、料理の差分を説明するためには過剰な差分も出力され、説明がわかりにくくなったためだと思われる。

また、4.3.2 節で、本手法より得られた出力の例を、表 7 に示す。表 7 において、お好み焼きと刺身をクエリとした際の出力では、差分の要素として、「キャベツ」や、「オリーブ」などの適切な食材を抽出できており、正確さの平均値も高い。しかし、カレーライスをクエリとした際の出力では、平均値が低い。ミネストローネとの差分となる食材として、「カレー粉」が抽出されるべきだと考えられるが、抽出できなかったことが原因である。これは、「カレー粉」の表現が、「カレールウ」や、「カレールー」など投稿したユーザによって異なるためだと考えられる。また、ラーメンをクエリとした際の出力では、スパゲッティとの差分となる食材として、「パスタ」が抽出されている。ラーメンとスパゲッティは、料理名こそ異なるが、類似した食材である「麺」と「パスタ」が用いられている料理であり、これらの差分となる食材として「パスタ」が抽出されるのは適切ではない。これらの問題の改善策として、料理の特徴ベクトルを生成

する際に、類似した食材をまとめた上で特徴ベクトルを生成することなどがあげられる。

6. 関連研究

6.1 レシピの特徴の抽出の研究

レシピの研究において、レシピデータのどの要素に着目し、レシピを特徴づけるかを考えることは重要である。Chung [3] は、レシピを特徴づける食材は材料リストの最初に記載されると考えた。橋ら [4] は、材料と調理器具の要素に着目し、レシピ名の修飾表現を付与した根拠となるネーミングコンセプトを抽出した。金内ら [5] は、調理手順のデータは、レシピ作成者の基づくものであるため、レシピの特徴を示す表現は、レビューのデータから抽出する方が有力だと考えた。レビュー情報の特徴抽出を用いて、目的別にレシピのカテゴリ分類を行った。

本研究は、特徴づける対象をレシピではなく、料理名としている。料理名を特徴づけることで、料理名間の類似度や差分を算出することが可能だと考えられる。また、本研究では、レシピ特有の単語を、一般的な文章ではあまり出現しないが、調理手順の文章には頻出する単語とし、これを料理を特徴づける要素としている。

6.2 類似レシピ抽出の研究

レシピを特徴づけた上で、レシピ間の類似度を算出する研究も数多く存在する。菊米ら [6] は、材料、手順に着目し、レシピ間の類似度を算出した。Wang ら [7] は、食材や調理手順を用いてレシピのグラフ化を行い、それをもとにレシピ間類似度を求めた。花井ら [8] は、類似レシピを判断する上で重要な要素は食材と調味料であるとし、これらと、食材の希少度を用いた類似レシクラスタリングの手法を提案した。福本ら [9] は、食材の重要度と食品群を考慮した、食材の分量を食材の重要度で重み付けしてコサイン類似度を求め、食品群ごとの平均分量を用いて同様にコサイン類似度を合算した値をレシピ間類似度とした。野沢ら [10] は、レシピデータの中には、一部の食材を入れ替えた類似レシピが多く存在することに着目し、特定の食材に対して他の食材の類似度を算出することにより、代替食材の発見する手法を提案した。

本研究は、レシピ間や、食材間の類似度ではなく、料理名間

表7 本手法より得られた出力の例

クエリ	国名	出力結果	正確さの評価の平均値
お好み焼き	韓国	お好み焼き = チヂミ + キャベツ - ごま油 - ニラ + 青のり + 豚肉	3.8
刺身	イタリア	刺身 = カルパッチョ - オリーブ	3.6
カレーライス	イタリア	カレーライス = ミネストローネ - ベーコン - キャベツ - オリーブ + ご飯 + 玉ねぎ	2.8
ラーメン	イタリア	ラーメン = スパゲッティ - パスタ - オリーブ	2.2

の類似度を算出している。上記のような、類似レシピの抽出の研究の中では、食材が重要であると示されているものが多いが、手順や、調味料を考慮しているものも存在する。これを考慮して、本研究では、食材や、調理工程をはじめとした様々な料理の要素の重要度を求めている。また、レシピ特有の単語に着目することは、有効であると考え、重要度を考慮した上で、これを用いて料理間の類似度を求めている。さらに、本研究は類似した二つの料理の差分となる要素を抽出する手法も提案している。ある料理を、類似した料理に加えて、それらの料理の差分となる要素を用いることで、より詳細な言い換えが可能になると考えられる。

7. まとめ

本稿では、飲食店メニューの説明欄の新たな表記方式とその自動生成手法を提案した。提案した表記方式は、「お好み焼き = チヂミ - ニラ + キャベツ」のように、任意の料理を、ある国の類似した料理名と差分となる要素によって単純な数式で表したものである。また、この表記方式を、ユーザ投稿型レシピサイトのデータを用いて自動生成する手法の提案を行った。提案した手法により生成した本表記方式を、被験者実験を通じて評価した。その結果、類似した料理は、比較的適切な料理が抽出されたことが確認できた。

式の差分となる食材は、食材名の表記ゆれが影響し、適切な食材が抽出できない場合があった。今後は、類似した食材をまとめた上で、料理の特徴ベクトルを生成することでこの改善を試みる。また、最適なレシピ特有の単語のカテゴリ分けについて調査する。

謝 辞

本研究は、首都大学東京傾斜的研究(全学分)学長裁量枠戦略的研究プロジェクト戦略的研究支援枠「ソーシャルビッグデータの分析・応用のための学術基盤の研究」及び JSPS 科研費 16K00157, 16K16158 による。また、本研究では、クックパッド株式会社と国立情報学研究所が提供する「クックパッドデータ」を利用した。ここに記し、感謝の意を表す。

文 献

- [1] Daisuke Kato, Mai Miyabe, Eiji Aramaki, and Akiyo Nadamoto. Sizzle word analysis from multiple internet media. In *Proceedings of the 18th International Conference on Network-Based Information Systems*, pp. 444–451. IEEE, 2015.
- [2] 笹田鉄郎, 森信介, 山肩洋子, 前田浩邦, 河原達也. レシピ用語の定義とその自動認識のためのタグ付与コーパスの構築. 自然言語処理, Vol. 22, No. 2, pp. 107–131, 2015.

- [3] Young-joo Chung. Finding food entity relationships using user-generated data in recipe service. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pp. 2611–2614. ACM, 2012.
- [4] 橘明穂, 若宮翔子, 難波英嗣, 角谷和俊. 料理名の修飾表現の関係性に基づくレシピのネーミングコンセプト抽出. 電子情報通信学会技術研究報告. DE, データ工学, Vol. 113, No. 214, pp. 19–24, 2013.
- [5] 金内萌, 難波英嗣, 角谷和俊. レビュー情報を用いた料理レシピカテゴリの構築およびレシピの自動分類. 2016 年度情報処理学会関西支部支部大会講演論文集, Vol. 2016, , 2016.
- [6] 苅米志帆乃, 藤井敦. 料理どうしの類似と組合せに基づく関連レシピ検索システム. 言語処理学会第 14 回年次大会発表論文集, pp. 959–962, 2008.
- [7] Liping Wang, Qing Li, Na Li, Guozhu Dong, and Yu Yang. Substructure similarity measurement in chinese recipes. In *Proceedings of the 17th international conference on World Wide Web*, pp. 979–988. ACM, 2008.
- [8] 花井俊介, 難波英嗣, 灘本明代. 主食材と主調味料を考慮した類似レシビクラスタリング. 第 8 回データ工学と情報マネジメントに関するフォーラム, pp. E2–7, 2016.
- [9] 福本亜紀, 井上悦子, 中川優. 食材の重要度と食品群を考慮したレシビ間類似度の算出手法. 第 4 回データ工学と情報マネジメントに関するフォーラム, pp. D9–2, 2012.
- [10] 野沢健人, 中岡義貴, 山本修平, 佐藤哲司. word2vec を用いた代替食材の発見手法の提案 (データ工学と食メディア). 電子情報通信学会技術研究報告. DE, データ工学, Vol. 114, No. 204, pp. 41–46, 2014.