原著論文

アナロジーによる料理の国際的な表記方式の提案[†]

信本 健輔^{*1}・廣田 雅春^{*2}・加藤 大受^{*3}・石川 博^{*1}

本稿では、飲食店のメニュー等に含まれる料理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、投稿型レシピサイトのデータを用いてこの表記方式を生成する2つの手法も提案する。どちらの手法もレシピデータに含まれる食材、調理工程などの6つの要素に基づ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により、本表記方式の有用性と2つの手法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に訪れた際、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でかつ、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、その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、インターネットで料理の情報を調べるとしても、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、2016年の10月の段階で年間の訪日外国人人数が2,000万人を超えた¹。ここで、国土交通省観光庁の調査²によると、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」は、どちらも「日本食を食べること」（73.1%と25.6%）が最多であり、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、訪日外国人が飲食店を訪れた際、前述した問題が起こっている。訪日外国人の受け入れ環境に対する不満のアンケート調査³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る。しかし、単純に料理名やその説明文の翻訳結果を飲食店の店員に提示しても十分に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。

他言語の話者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1つと考えられる。しかし、これを実現するためには多くの課題がある。多く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、翻訳先の言語の単語や文法の知識が必要となる。また、機械翻訳を用いて翻訳する方法も存在するが、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困難である。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⁴や、楽天レシピ⁵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。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、

[†] Multilingualization of Cooking by Analogical Description
Kensuke NOBUMOTO, Masaharu HIROTA, Daiju KATO,
and Hiroshi ISHIKAWA

^{*1} 首都大学東京大学院システムデザイン研究科
Faculty of System Design, Tokyo Metropolitan University

^{*2} 岡山理科大学総合情報学部
Faculty of Informatics, Okayama University of Science

^{*3} ウイングアーク 1st 株式会社
WingArc1st Inc.

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

2 訪日外国人消費動向報告書（2016年 7-9）
<http://www.mlit.go.jp/common/001149546.pdf>

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>

4 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/>

5 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

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。

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。提案する表記方式を以下に示す。

お好み焼き = チヂミ - ニラ + キャベツ (1)

式 1 において、左辺はある国の任意の 1 つの料理を示す。右辺は、左辺に類似した別の国の料理名と、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、翻訳の容易化と読者の負担の軽減の 2 つの利点がある。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可能である。

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となるためである。

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料理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可能である。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

本章では、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

の手順で、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、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. 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収集し、 L 個の料理を含む 1 つの料理リストにする。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をレシピ特有の単語とする。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日本語の Wikipedia⁶ 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異度を示すスコア $sc_{dr}(w)$ を

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

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。 $sc_{dr}(w)$ の上位 500 件の単語をレシピ特有の単語として扱う。

表 1 に抽出したレシピ特有の単語の上位 5 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。本稿では、これらのレシピ特有の単語を食材や調理工程

6 ウィキペディア - Wikipedia <https://ja.wikipedia.org/wiki/>

表 1 レシピ特有の単語の例

レシピ特有の単語	レシピ特有度
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

などの数種類のカテゴリに人手により分類した。

3.1.3 特徴ベクトルの生成手法 1

本稿では、3.1.1 項で生成したある国の料理リストを用いて、2 種類の各料理名の特徴ベクトルの生成手法を提案する。ここでは、料理の重みを考慮した特徴ベクトルについて説明する(以下、手法 1)。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 $\mathbf{v}_a(c, l)$ を

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

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

より生成する。ここで、 $\text{df}(w_{c,m}, l)$ は料理名 l をレシピ名に含むレシピデータのうちレシピ特有の単語 $w_{c,m}$ を調理手順の文章に含むレシピデータ数、 $|N_l|$ は料理名 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 $\mathbf{v}_a(c, q)$ も、同様に生成する。

3.1.4 特徴ベクトルの生成手法 2

次に、もう 1 つの特徴ベクトルの生成手法として、調理手順の文書による重みを考慮した特徴ベクトルを

$$b_m = \frac{\text{df}(w_{c,m}, l)}{\text{df}(w_{c,m}, \text{all})} \quad (5)$$

より生成する(以下、手法 2)。 $\text{df}(w_{c,m}, \text{all})$ は全レシピデータのうちレシピ特有の単語 $w_{c,m}$ を調理手順の文章に含むレシピデータ数である。手法 2 も手法 1 と同様にクエリの特徴ベクトルを生成する。

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て料理名リストの各料理名とクエリとの類似度を求める。

はじめに、3.1 節で述べた、料理を特徴づける食材や調理工程などの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、食材の方が料理に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 $\text{sim}_c(l, q)$ 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 $\text{sim}_c(l, q)$ 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 $\text{sim}_{\text{all}}(l, q)$ を

$$\text{sim}_{\text{all}}(l, q) = \prod_{i=1}^C \text{sim}_i(l, q)^{\alpha_i} \quad (6)$$

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。予備実験の結果は、4.2.2 項で述べる。得られた $\text{sim}_{\text{all}}(l, q)$ 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。本稿では、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する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 $\mathbf{v}_{\text{diff}}(s, q)$ を

$$\mathbf{v}_{\text{diff}}(s, q) = \mathbf{v}(s) - \mathbf{v}(q) \quad (7)$$

より生成する。得られた $\mathbf{v}_{\text{diff}}(s, q)$ の要素の値の絶対値が大きい順に P 件の要素 p に対応する食材 w_p ($0 \leq p \leq P$) を s と q の料理の差分となる要素として抽出する。これらの要素を用いて式 1 のような出力を生成する。式 1 の右辺の表現を r とする。 r は $r \leftarrow s$ で初期化し、順に

$$\begin{cases} r \leftarrow r + w_p & (a_p < 0) \\ r \leftarrow r & (a_p = 0) \\ r \leftarrow r - w_p & (a_p > 0) \end{cases} \quad (8)$$

にしたがって、 w_p を連結していくことで生成する。この a_p は要素 p の値を指す。

4. 実験条件

本章では、レシピデータを用いて 2 章で提案した表記方式を 3 章の 2 つの手法より作成し、その結果に対する評価実験の手順について述べる。本稿では、被験者実験により、2 つの手法の性能を比較した。

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」⁷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 項で説明した料理名リストは、韓国とイタリアの 2 カ国のリストを Web サイトを参考に作成した。このとき、クックパッドのレシピデー

7 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>

表2 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	しょう油	466.1	麺棒	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	オイスターソース	422.3	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	ボン酢	319.4	菜箸	490.1	水気	1045.4	手早い	1694.8

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は、韓国が59、イタリアが45となった。

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⁸を用いた。

4.2 実装

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て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項の文章の形態素解析には、MeCab⁹を用いた。3.1.2項の分類は、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 ($C = 6$) のカテゴリとした。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。

各カテゴリの5単語を表2に示す。表2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 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。

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、インターネットによる検索で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

表3 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

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、6つのカテゴリの重要度の総和が1になるように正規化したものを各カテゴリの重要度とする。表3に各カテゴリのレシピ特有の単語数と重要度を示す。

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

本実験では、手法1と手法2の2つの手法により生成したそれぞれの本表記方式を6名の被験者がそれぞれ評価する。被験者は6名全員日本在住年数が20年以上23年以下であり、被験者6名の平均年齢は21.5歳である。2つの手法により生成した本表記方式の評価を比較し、考察する。クエリは、「刺身」や「焼肉」などの代表的な日本料理を25件用いた。

4.3.1 類似した料理の抽出の評価手順

2つの提案手法について、各クエリに類似した料理を2カ国の料理リストから抽出する。被験者は、2つの提案手法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）で評価する。

4.3.2 差分の出力の性能評価手順

2つの提案手法に、従来手法である料理の説明文を機械翻訳により翻訳する方法を加えた3つの手法による出力を比較し、考察する。

提案手法の出力は、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、2つの提案手法より、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）

⁸ 日本の Wikipedia 情報ダウンロードページ
<https://dumps.wikimedia.org/jawiki/latest/>

⁹ MeCab: Yet Another Part-of-Speech and Morphological Analyzer <http://taku910.github.io/mecab/>

表4 本手法の評価値の平均と評価値が3以上のクエリの数

	評価値の平均	評価値が3以上のクエリの数
手法1	2.7	22
手法2	2.6	17

・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）で評価する。

従来手法の出力は、本実験の被験者が日本人であるため、英語で記載された日本料理の説明文を機械翻訳により日本語に翻訳したものを評価した。英語の料理の説明文は料理の説明として適切な部分を人手で抽出し、機械翻訳には、Google 翻訳¹⁰を用いた。被験者は翻訳された料理の説明文それぞれが

- ・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）
- ・文章として意味はとれるか（翻訳の性能）

の3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）で評価する。

5. 実験結果

5.1 類似した料理の抽出の結果

表3に、レシピ特有の単語をカテゴリ分けし、それぞれのカテゴリごとの重要度を求めた結果を示す。表3より、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。

25件のクエリに対し、類似した料理として抽出したイタリア料理と韓国料理それぞれ計50件の評価を総合した結果として、各手法の評価値の平均と、評価値が3以上だったクエリ数を表4に示す。表4より、各手法の評価値の平均は同程度あることがわかる。また、評価値が3以上のクエリ数は、わずかに手法1の方が大きいことがわかる。したがって、類似した料理の抽出性能は、手法1と手法2は同程度であることがわかる。

表5に、クエリに類似した料理と被験者による5段階評価の平均値の例を示す。表5より、同じクエリでも国によって類似した料理が適切に抽出できている場合とできていない場合があることがわかる。例えば、「焼肉」をクエリとしたときは、韓国料理では類似した料理を抽出できているが、イタリア料理ではできていない。これは、韓国料理のリストには「焼肉」に類似した「プルコギ」が存在しているが、イタリア料理のリストには存在しないためである。逆のことが、「天ぷら」をクエリとしたときなどにも起きている。これを改善するためには、料理リストの料理名数を増やす必要があると考えられる。本実験では、単純に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は既知の料理が類似した料理として用いられ

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

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た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。

また、表5のクエリ「ちらし寿司」について、手法2の韓国料理は「キムパブ」という韓国の海苔巻きが類似料理として抽出されている。一方、手法1は「冷麺」が類似料理で、評価値が低い。手法1の「冷麺」と「ちらし寿司」の食材のコサイン類似度は0.70である。これは、手法1の「キムパブ」と「ちらし寿司」の食材のコサイン類似度が0.18であるため、「冷麺」と「ちらし寿司」の食材の類似度が高いことがわかる。手法1で「冷麺」を抽出しているのは、「ちらし寿司」と「冷麺」が「タマゴ」、「キュウリ」、「酢」、「みりん」など、食材や調味料が類似した料理であるためだと考えられる。レシピの内容としては類似した料理であるが、主食材こそ異なり、人からは類似していない料理と判断されやすいため評価値が低い結果となったと考えられる。

5.2 差分の出力の結果

5.2.1 手法1と手法2の比較

図2に、4.3.2項より得られた、2つの提案手法において、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被験者による評価値の平均値を示す。差分となる要素の抽出の評価を行うために、図2には、4.3.1項で3以上と評価されたクエリと類似した料理の差分となる要素の出力の評価の平均値のみを反映している。図2より、手法1と手法2はどちらも、 p が増加してもイメージのしやすさはあまり変化しないが、クエリの説明の正確さは増加し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、説明として不足している点を差分となる要素がうまく補っているためと考えられる。また、クエリの説明の正確さとイメージのしやすさ、どちらも全ての要素数でわずかに手法2の方が評価の平均値が高いが、要素数の増加に伴う評価の平均値の変化は手法1と手法2ではあまり違いがないことがわかる。

また、4.3.2項で述べた本手法より得られた出力の例を表6に示す。表6より、2つの手法において、「お好み焼き」と「チヂミ」の差分となる要素の出力として「ゴマアブラ」や「キャベツ」、「アオノリ」など、適切なものが出力されていることがわかる。しかし、手法1の出力と手法2の出力を比較すると、

10 Google 翻訳 <https://translate.google.com>

表5 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の例

	韓国				イタリア			
	手法1		手法2		手法1		手法2	
クエリ	類似した料理	評価	類似した料理	評価	類似した料理	評価	類似した料理	評価
焼肉	プルコギ	3.5	プルコギ	3.5	スパゲッティ	1.7	ペペロンチーノ	1.3
天ぷら	ビビンバ	1.2	冷麺	1.0	フリット	4.3	フリット	4.3
とんかつ	サムギョプサル	3.0	サムギョプサル	3.0	フリット	4.2	フリット	4.2
ちらし寿司	冷麺	1.3	キムパブ	4.0	カルパッチョ	2.5	カルパッチョ	2.5

表6 本手法より得られた出力の例

手法	国名	出力結果
手法1	韓国	お好み焼き = チヂミ + ゴマアブラ + キャベツ + ニラ + アオノリ + カツオブシ
手法2	韓国	お好み焼き = チヂミ + アオノリ + アゲダマ + ベニシヨウガ + カツオブシ + キャベツ
手法1	イタリア	カレー = カボナータ + オリーブ + ニンニク + ズッキーニ + パプリカ + タマネギ
手法1	イタリア	ラーメン = スパゲッティ + パスタ + オリーブ + ニンニク + タマネギ + ネギ

手法1の出力の方が2つの料理の主要食材を抽出できている。これは、手法2で生成する特徴ベクトルの要素の値は、全レシピデータの調理手順の文章に含まれるレシピデータ数が高い単語ほど低くなるためだと考えられる。したがって、差分となる要素の抽出手法としては、手法1の方が適切であると考えられる。

しかし、手法1でも適切な差分となる要素が出力されない場合もある。「カレー」と「カボナータ」の差分となる食材として、「カレールウ」が2つの料理の大きな差分の1つであると考えられるが、抽出されなかった。これは、「カレールウ」という食材が全レシピデータの中でも出現頻度が低く、3.1.2項で述べたレシピ特有の単語の抽出の際に、抽出できなかったためだと考えられる。したがって、本手法では、「カレー」における「カレールウ」のような、ある料理のみに用いられる食材を抽出することが困難であると考えられる。また、「ラーメン」と「スパゲッティ」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、食材名こそ異なるが、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。

本実験は、日本料理に対する精通度合いが低い被験者での評価が行えなかったが、出力として得られた「カツオブシ」や「アゲダマ」などは、そのような被験者には理解しにくい表現である可能性がある。今後は、日本料理に対する精通度合いが低い被験者での評価を行う必要があると考えられる。

5.2.2 従来手法との比較

4.3.2項より得られた、従来手法においての被験者による評価の平均値は、正確さが3.1、イメージのしやすさが3.0、翻訳の性能が3.2となった。この結果と図2を比較することによって、正確さに関しては、2つの提案手法の要素数が5以上のとき、評価の平均値は3.3以上となり、従来手法の出力の評価の

平均値を上回ることがわかる。また、イメージのしやすさに関しては、2つの提案手法と、従来手法で評価の平均値にそれほど大きな違いはないが、要素数によっては、2つの提案手法の方が高い場合もあることがわかる。さらに、従来手法は、多少の違和感はあるものの、ある程度意味を取ることができる翻訳が可能ながわかる。

6. 関連研究

6.1 レシピの特徴の抽出の研究

レシピを用いた多くの研究において、レシピデータのどの要素を用いてレシピを特徴づけるかを考えることは重要である。Chung[1]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。橘ら[2]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[3]は、調理手順のデータはレシピ作成者の主観に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

特徴づけたレシピを用いてレシピ間の類似度を算出する研究や、その結果を用いる研究も数多く存在する。Wangら[4]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[5]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ークラスタリングの手法を提案した。福本ら[6]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 [7] 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2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、アナロジーを用いた飲食店のメニューなどの料理の説明の新たな表記方式を提案した。また、ユーザ投稿型レシピサイトのデータを用いてこの表記方式を自動生成する2つの手法の提案を行った。2つの提案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、クエリに最も類似する料理の抽出と差分の抽出について、手法1の方が正確と評価された。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れの改善を試みる。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6K00157, 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、感謝の意を表する。

参考文献

- [1] Y. Chung: "Finding food entity relationships using user-generated data in recipe service," *Proc. of the 21st ACM Int. Conf. on Information and knowledge management*, pp. 2611-2614, 2012.
- [2] A. Tachibana, S. Wakamiya, H. Nanba, and K. Sumiya: "Extraction of concepts based on relations of modifiers in recipe titles," *IEICE Technical Report. Data engineering*, Vol.113, No.214, pp. 19-24, 2013.
- [3] M. Kanauchi, H. Nanba, and K. Sumiya: "Construction of a cooking recipe category and automatic classification of recipes," *Proc. of the 2016 IEICE Kansai Section Convention Record*, Vol.2016, 2016.
- [4] L. Wang, Q. Li, N. Li, G. Dong, and Y. Yang: "Substructure similarity measurement in chinese recipes," *Proc. of the 17th Int. Conf. on World Wide Web*, pp. 979-988, 2008.
- [5] 花井俊介, 難波英嗣, 難本明代: "主食材と主調味料を考慮した類

似レシビクラスターリング," 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.

- [6] 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ピ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.
- [7] K. Nozawa, Y. Nakaoka, S. Yamamoto, and S. Tetsuji: "Finding method of replaceable ingredients using large amounts of cooking recipes," *Proc. of the IEICE Technical Report. Data engineering*, Vol.114, No.204, pp. 41-46, 2014.

(2018年4月14日 受付)

(2018年8月18日 採録)

[問い合わせ先]

〒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

信本 健輔

TEL: 042-585-8600-4605

E-mail: nobumoto-kensuke@ed.tmu.ac.jp

著者紹介

信本 健輔 [非会員]

2017年首都大学東京システムデザイン学部卒業。現在、同大学大学院システムデザイン研究科在学中。

廣田 雅春 [非会員]

2014年静岡大学創造科学技術大学院情報科学専攻修了。大分工業高等専門学校情報工学科助教を経て、2017年より岡山理科大学総合情報学部情報科学科講師。博士(情報学)。情報処理学会、日本データベース学会、ACM、自然言語処理学会、観光情報学会 各会員。

加藤 大受 [非会員]

ウイングアーク1st株式会社 技術本部品質管理部所属。首都大学東京システムデザイン研究科 博士後期課程在学中。品質管理学会、観光情報学会、日本サービス学会 各会員。

石川 博 [非会員]

首都大学東京システムデザイン研究科教授。1979年東大理学部情報科学科卒。富士通研究所・都立大学・静岡大を経て2013年より現職。東大博士(理学)。著書に『ソーシャル・ビッグデータサイエンス入門』(コロナ社)等、ACM TODS, IEEE TKDE, VLDB, IEEE ICDE 等論文多数。情報処理学会坂井記念特別賞、科学技術庁長官賞受賞。情報処理学会フェロー、電子情報通信学会フェロー、ACM, IEEE 各会員。

Multilingualization of Cooking by Analogical Description

by

Kensuke NOBUMOTO, Masaharu HIROTA, Daiju KATO, and Hiroshi ISHIKAWA**Abstract:**

As described in this paper, we propose a new analogical description for cooking and a method to generate it automatically. The analogical description consists of a similar cooking and differences between those cooking, such as 'Okonomiyaki = Korean pancake - leek + cabbage'. To extract a similar cooking and differences for the description, we analyze recipe data of cooking sites. In this paper, we proposed two types of feature vector to calculate the similarity. We specifically examine six elements as differences of cooking and its similar cooking (e.g., foods and cooking-related verbs). We demonstrate that our proposed description is expressible in subject experiment.

Keywords: recipe site, similarity dish, internationalization

Contact Address: **Kensuke NOBUMOTO**

Faculty of System Design, Tokyo Metropolitan University
6-6 Asahigaoka, Hino-shi, Tokyo 191-0065, Japan
TEL: +81-42-585-8600-4605
E-mail: nobumoto-kensuke@ed.tmu.ac.jp