

# 好奇心と安心感を両立する料理推薦

坂井 優斗<sup>†</sup> 馬 強<sup>††</sup>

<sup>†</sup> 京都大学工学部情報学科 〒 606-8501 京都府京都市左京区吉田本町

<sup>††</sup> 京都大学大学院情報学研究科 〒 606-8501 京都府京都市左京区吉田本町

E-mail: <sup>†</sup>y.sakai@db.soc.i.kyoto-u.ac.jp, <sup>††</sup>qiang@i.kyoto-u.ac.jp

**あらまし** 旅行先で今まで食べたことがない料理や名物を楽しみたいが、その料理を食べられるかについて不安を感じる場合がある。料理の推薦に関する研究はユーザの好みや料理の人気度に基づくものが多く、旅先など知らない土地の料理を安心して楽しめるかに関する考慮が十分ではない。そこで本研究では、安心感と好奇心を両立して料理を推薦するために、画像やレシピデータおよびユーザの食履歴から、料理の味と素材に対する受容性と冒険性を推定するモデルを提案する。また、推定された受容性と冒険性を用いて安心感と好奇心を両立する料理推薦の実現可能性を検証する。

**キーワード** 料理推薦, 旅行, 受容性, 冒険性, マルチメディア

## 1 はじめに

旅行において食事は楽しみの1つである。特に、旅行先で今まで食べたことがない知らない料理や名物を食すことは旅行の醍醐味であり、自らの好奇心を満たす行為である。しかし、知らない料理を食べれば必ず良質な経験が得られるわけではない。口にすると料理は自分が食べることのできる食材で構成されていて、自分の味覚に合うような味であることが望ましい。これは旅行という場面で特に強く求められる。なぜならば、旅行は多くの場合一度きりのものであり、替えが効かないからである。また、不慣れた旅先では、健康に与える影響を心配する場面がある。日常生活における食事であれば、仮に料理に満足しなかったとしても我慢することは容易く、別の機会に再挑戦することも可能である。しかし、旅先での食事は、口に合わなかったからといって新たな機会を設けることは難しく、食べる料理の選択を失敗したくないという気持ちが強く働く。従って、その料理を自分が受け入れることができると安心できることが重要になってくる。知らない料理ではあるが、旅行先で人気である、有名であるという理由で食べようと試みる冒険性と、自分がその料理を満足して食べることができるであろうという安心感はトレードオフの関係にある。

近年、料理の推薦の研究が盛んとなっている。X.Gao ら [1] は、食事の推薦をユーザがレシピに対して持つ嗜好を予想する問題であると定義して、ユーザの食の嗜好の推定を試みた。Ueda ら [2] も X.Gao ら [1] の研究と同様に、ユーザが持つ食の嗜好に基づいた料理の推薦手法を提案している。旅先での食事において、ここで紹介した従来手法のようにユーザの食の嗜好を用いて推薦を行えば、推薦によって得られる料理はそのユーザの好みの味に近く、美味しく食べることができるであろうという安心性は確保できる。しかし、普段から好んで食べている料理の味と似ていることから好奇心は伴わない。

また、Correia ら [3] は旅行に訪れた観光客への料理の推薦の

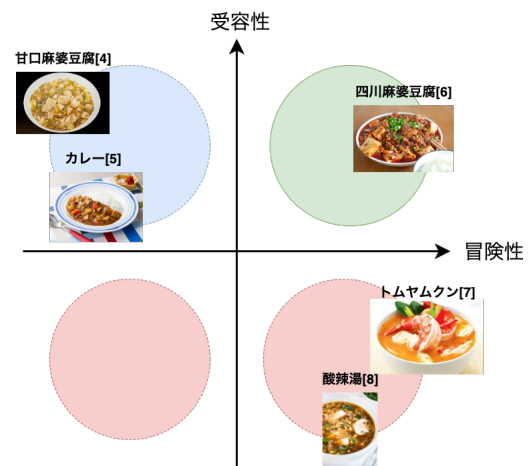


図1 推薦軸による料理例 (出典: [4] [5] [6] [7] [8])

要素の中に、食の独自性や伝統があると結論づけた。その土地固有の料理や、その土地で人気な料理をベースに推薦を行うことで、ユーザの好奇心や冒険心を満たすような料理の推薦はできる。しかし、その推薦結果の中にはユーザが食べられない食材で作られている料理や、口に合わないような癖の強い料理が含まれているかもしれない。従って、安心感は担保されない。

このように我々の知る限りでは、旅行先で知らない料理に対して好奇心と安心感の双方を備えるような推薦を行う手法が未だない。図1で推薦の要因とその結果得られる推薦料理の対応例を示す。日常的に辛い料理を好み、その中には辛口のカレーが含まれ、またエスニック料理はあまり好まないユーザの存在を仮定する。このとき、図1の赤色の円形に該当する領域である、第三象限、第四象限に含まれる料理はユーザが受け入れることが難しい料理が含まれているため、推薦しないことが望ましい。また、図1の青色の円形に該当する領域である、第二象限に含まれる料理はユーザが受け入れやすいが冒険性に乏しい料理となっていて、味の嗜好を考慮したような従来の推薦手法ではこの領域の料理が推薦されることになる。好奇心と安心感

を両立させた料理というのは、図 1 の緑色の円形に該当する第一象限に含まれる料理のことであり、図の例で言えば中国本場の四川麻婆豆腐にあたる。

本研究では安心性と好奇心を両立した料理推薦を実現するために、料理の味を表すベクトルである“味表現ベクトル”と食材を表す“食材表現ベクトル”を求める味ネットワークモデルを提案する。提案モデルは料理の味を構成する辛味、甘味、酸味の程度の識別を訓練タスクとして課すことで味表現ベクトルを学習する。

そして、ユーザがこれまでに食べてきた料理に対して味表現ベクトルと食材表現ベクトルを求め、ターゲットとなる料理の味表現ベクトル、食材表現ベクトルとの間の位置関係から、ユーザがターゲットの料理をどれくらい受け入れられるかを表す受容性と、どれくらい好奇心が満たされるかを表す冒険性という指標を推定して、安心感と好奇心を両立する料理推薦の実現を検討する。

本研究の貢献は次の通りである。

- 味ネットワークモデルを利用することで、料理の画像データとレシピデータから辛味、甘味、酸味の程度を識別することができる。(3 節)
- 画像データとレシピデータを用いて、味と食材という料理を構成する 2 種類の要素を表す表現ベクトルを出力するニューラルネットワークモデルを提案する。(3 節)
- ユーザの食履歴データを用いて、対象とした料理の受容性と冒険性の分析を行う。(4 節)

## 2 関連研究

X.Gao ら [1] は食事の推薦をユーザがレシピに対して持つ嗜好を予想する問題であると定義している。そして、食の嗜好を推定するために Hierarchical Attention based Food Recommendation (HAFR) というニューラルネットワークを基にした推薦システムを提案している。HAFR は、ユーザが特定のレシピの料理を食べたかどうかを表す履歴データ、料理の材料データ、料理の画像データの 3 つの入力から、あるレシピをユーザが受け入れる確率を推定する。

Ueda ら [2] も食の嗜好をベースにしたレシピの推薦手法を提案している。ユーザが過去に閲覧したことのあるレシピや調理したことのあるレシピからそのユーザが好きな食材及び嫌いな食材、つまり食の好みを推定し、さらに好みの食材が多く含まれるほどレシピのスコアが高くなるような推薦を行なっている。

Kauppinen ら [9] は記憶に残っている素晴らしい食事や食事経験を探索し、過去の経験の性質を理解することを目的に、フォーカスグループインタビューを行い、参加者からポジティブな食事経験に関する話やそれがポジティブであった理由などのデータを集めた。フォーカスグループインタビューの参加者は 25 名の男女で、年齢層は 21～58 歳であった。インタビューによって集められたデータの分析結果から、食に関連する経験は大きく 4 つのテーマに分かれていて、その中には“観光”が含まれると主張した。また、観光客としての食事や食事経験を

説明する典型的な特徴に、新しい種類の食事へ挑戦する意欲を挙げた。

Jiménez ら [10] はガストロノミー、文化、観光の間の関係性の分析の研究を行なった。スペインのコルドバ歴史地区で外国人観光客を対象にフィールドワークを行った結果、観光客は料理を観光地の文化的アイデンティティの一部として重要視していることが判明した。

Correia ら [3] はファジィ集合質の比較分析で、地元料理の属性が客の満足度に与える影響を探索した。香港を訪れた外国人観光客を対象としたデータを分析した結果、推薦につながる要因には料理の質、独自性、伝統、サービスの質の 4 つがあるという結果が得られた。特に、料理の独自性と伝統は 4 つの属性の中でも重要な特徴であると主張している。

Javier Marín ら [11] は、料理がテキストで表されるレシピデータと、画像データという異なる形式の 2 種類のデータで構成されるときに、レシピデータと画像データそれぞれを同一の共有空間へ同じ形状のベクトルとして埋め込むことによって、料理のレシピから料理画像の検索や、料理画像からレシピの検索手法を提案している。また、Javier Marín らはクロスモーダルな検索手法を実現するために、ニューラルネットワークモデル Joint Neural Embedding Model(本論文ではこれ以降 JNEM と表記する。)を提案している。JNEM は料理を構成する食材及び料理を形づくる調理手順を示すテキストデータからレシピ表現ベクトルを、画像データから料理画像表現ベクトルを出力する。同一の料理から出力されるレシピ表現ベクトルと画像表現ベクトルは 2 つのコサイン類似度が小さくなるように学習されているため、テキストと画像という異なる形式のデータから得られる表現ベクトルが同一の空間で扱えるようになっている。JNEM の一部は本研究の提案モデルで利用され、その具体的な構造の説明は 3.1 節で行う。

## 3 提案手法

本節では、味表現ベクトルや食材表現ベクトルを学習するモデルである味ネットワークを説明して、それを利用して受容性と冒険性を推定する手法を述べる。味ネットワークは、料理の食材データ、調理手順データ、画像データを入力として、料理を構成する食材の表現ベクトルと、料理の味の表現ベクトルを出力する。味ネットワークの構造図は図 2 のようになっている。以下、3.1 節で味ネットワークモデルに組み込まれている、Javier Marín ら [11] の提案モデル JNEM の説明を行い、3.2 節で、モデルを構成するコンポーネント内の処理をまとめる。そして、3.3 節では、食材表現ベクトルと味表現ベクトルの推定手法を説明する。3.4 節では、提案モデルの訓練タスクを説明し、3.5 節では提案モデルの学習及びテストに用いるデータセットについて説明し、最後に 3.6 節で提案モデルの学習方法と評価結果をまとめる。

### 3.1 JNEM

JNEM は図 3 に示すように、主に 3 つのエンコーダー部と

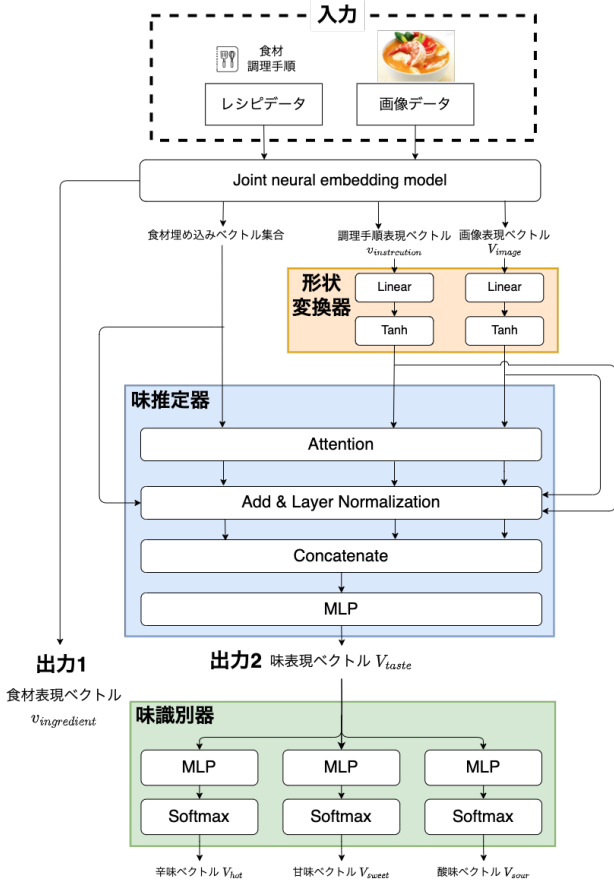


図 2 味ネットワークの構造図

全結合層から構成される。入力料理に含まれる食材集合と調理手順から成るレシピデータ及び、料理の画像データである。各エンコーダー部の仕組みは次のとおりである。

### 3.1.1 食材エンコーダー

はじめに、レシピの食材集合に含まれる各食材は word2vec により、300 次元の意味ベクトルへと変換される。これらを、 $\{v_{ingr}^1, v_{ingr}^2, \dots, v_{ingr}^m\}$  とする。ただし、 $m$  はレシピに含まれる食材の数とする。その後、食材集合が非順序関係であることを考慮し、双方向 LSTM に各食材ベクトル  $v_{ingr}^1, v_{ingr}^2, \dots, v_{ingr}^m$  を入力し、出力として食材表現ベクトルである単一の 600 次元ベクトル  $v_{ingredient}$  を得る。

### 3.1.2 調理手順エンコーダー

一般に、レシピに含まれる調理手順は複数の文、手順から構成されている。そこで、まずは各手順を示す文章のエンコードを Skip-thought [12] を応用した手法で行う。Skip-thought のモデルには GRU が含まれるが、その部分を LSTM に変えて、さらに調理手順の始まりと終わりを表す指示を追加するという変更を行なったモデルで、各調理手順の文を 1024 次元のベクトルへと変換する。最後に、先ほど得られたベクトルを通常の LSTM に入力し、レシピの調理手順表現ベクトルである単一の 1024 次元のベクトルを得る。ここで、調理手順表現ベクトルを  $v_{instruction}$  としておく。

### 3.1.3 画像エンコーダー

料理画像は、データセット ImageNet [13] による事前学習済

みモデル ResNet50 [14] からソフトマックス関数による分類層を取り除いたモデルで 2048 次元のベクトル  $v_{image}$  にエンコードされる。

### 3.1.4 共有空間への埋め込み

各エンコーダによって得られた表現ベクトルは以下のようにして同一の共有空間へと埋め込まれる。

- レシピの食材と調理手順を埋め込みベクトル  $v_{ingredient}$  と  $v_{instruction}$  でそれぞれ表現する。そして、それらを結合して、1 層の線形層と活性化関数として  $\tanh$  関数に通すことで、1024 次元のレシピデータ埋め込みベクトル  $V_{recipe}$  を得る。

- 画像データはレシピデータの埋め込みベクトルの次元に合わせるために、3.1.3 節で得られた  $v_{image}$  を 1 層の線形層と活性化関数として  $\tanh$  関数に通して、1024 次元の画像データ埋め込みベクトル  $V_{image}$  を得る。

そして、レシピデータと画像データのペアを用いて、それら 2 つのデータが同一の料理のものであるときに、 $V_{recipe}, V_{image}$  のコサイン類似度が小さくなるようにモデルの訓練を行う。

## 3.2 コンポーネントの説明

味ネットワークは、形状変換器、味推定器、味識別器の 3 つのコンポーネントから成る。本節では、各コンポーネント内の処理を説明する。

### 3.2.1 形状変換器

形状変換器では、 $v_{instruction}, V_{image}$  を入力にとり、それらを 1 層の線形層と活性化関数  $\tanh$  関数に通すことで、300 次元ベクトルへと変換した  $v'_{instruction}, V'_{image}$  を出力する。

$$v'_{instruction} = \tanh(W_{instruction} v_{instruction} + b_{instruction})$$

$$V'_{image} = \tanh(W_{image} V_{image} + b_{image})$$

### 3.2.2 味推定器

味推定器は  $\{v_{ingr}^1, v_{ingr}^2, \dots, v_{ingr}^{20}\}, v'_{instruction}, V'_{image}$  を入力にとり、味表現ベクトル  $V_{taste}$  を出力する。

まず、入力ベクトルを Attention 層に通して、入力ベクトルを重み付けした結果を出力から得る。このとき、Attention 層の出力は大きさが 22 行 300 列の行列であり、各行には  $\{v_{ingr}^1, v_{ingr}^2, \dots, v_{ingr}^{20}\}, v'_{instruction}, V'_{image}$  を重み付けしたベクトルが入っている。

$$query = [v_{ingr}^1, v_{ingr}^2, \dots, v_{ingr}^{20}, v'_{instruction}, V'_{image}]$$

$$key = [v_{ingr}^1, v_{ingr}^2, \dots, v_{ingr}^{20}, v'_{instruction}, V'_{image}]$$

$$value = [v_{ingr}^1, v_{ingr}^2, \dots, v_{ingr}^{20}, v'_{instruction}, V'_{image}]$$

$$V^{attention} = \text{Attention}(query, key, value)$$

次に、Attention 層の入力と出力の和をとり、Layer Normalization を用いて各行ベクトルの特徴量の正規化を行い、 $V_{LN}$  を得る。

$$V^{add} = [v_{ingr}^1, v_{ingr}^2, \dots, v_{ingr}^{20}, v'_{instruction}, V'_{image}] + V^{attention}$$

$$V^{LN} = \text{LayerNormalization}(V^{add})$$

そして、 $V^{LN}$  の各行ベクトルを連結して単一ベクトル

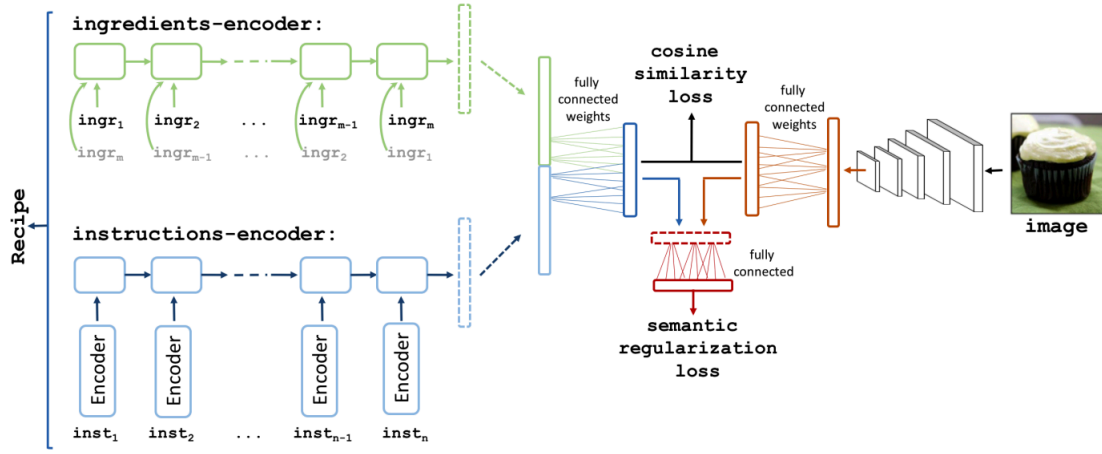


図3 JNEMの構造図 (出典: [11])

$V'^{attention}$ へと変換する。

$$V'^{attention} = \text{concatenate}(V^{LN})$$

最後に、 $V'^{attention}$ を隠れ層の活性化関数が $ReLU$ 関数である3層の多層パーセプトロンに通す。隠れ層には、非活性化ノードの割合を0.5に設定しているDropoutを適用している。MLPの出力から6次元の味表現ベクトル $V_{taste}$ を得て、味推定器の出力とする。

### 3.2.3 味識別器

味識別器は味表現ベクトル $V_{taste}$ を入力にとり、辛味、甘味、酸味の程度を示す3次元ベクトル $V_{hot}, V_{sweet}, V_{sour}$ を出力する。

はじめに、 $V_{taste}$ を隠れ層の活性化関数が $ReLU$ 関数である3層のMLPに入力し、出力層では $softmax$ 関数を適用し $V_{hot}$ を得る。ここで、 $V_{hot}$ は第一要素から順に、辛くない、やや辛い、とても辛いを表す辛味ベクトルである。 $V_{hot}$ と同様に、 $V_{taste}$ をMLPに入力して、 $V_{sweet}, V_{sour}$ を得る。 $V_{sweet}$ は第一要素から順に、甘くない、やや甘い、とても甘いを表す甘味ベクトル、 $V_{sour}$ は第一要素から順に、酸っぱくない、やや酸っぱい、とても酸っぱいを表す酸味ベクトルである。 $V_{hot}, V_{sweet}, V_{sour}$ は、3.4節で説明するモデルの学習に用いる。

### 3.3 食材表現ベクトル・味表現ベクトルの推定方法

本節では、味ネットワークが出力する食材表現ベクトルと味表現ベクトルの推定方法について説明する。

**食材表現ベクトル** 食材表現ベクトルはJNEMから出力される $v_{ingrdient}$ とする。

**味表現ベクトル** はじめに、味表現ベクトルを求めたい料理のレシピデータ、画像データをJNEMに入力して、食材の集合をword2vecで処理した結果である $\{v_{ingr}^1, v_{ingr}^2, \dots, v_{ingr}^m, v_{ingr}^{m+1}, \dots, v_{ingr}^{20}\}$ 、調理手順のエンコード結果である $v_{instruction}$ 、画像のエンコード結果である $V_{image}$ を得る。ただし、 $m$ は食材の数で、 $\{v_{ingr}^{m+1}, \dots, v_{ingr}^{20}\}$ は食材集合の系列長を揃えるために用意された、ダミー食材(食材ID: 0)に対するword2vecによる埋め込みベクトル

である。次に、 $v_{instruction}, V_{image}$ を形状変換器に入力して、300次元に変換した $v'_{instruction}, V'_{image}$ を得る。最後に、 $\{v_{ingr}^1, v_{ingr}^2, \dots, v_{ingr}^m, v_{ingr}^{m+1}, \dots, v_{ingr}^{20}\}, v'_{instruction}, V'_{image}$ を味推定器に入力して、出力から味表現ベクトル $V_{taste}$ を得る。

### 3.4 味ネットワークの訓練タスク

一般に、人間が舌で感じる味は甘味・塩味・酸味・苦味・うま味の5つとされているが、食の嗜好には辛味の要素も欠かせないと判断し、味を甘味・塩味・酸味・苦味・うま味・辛味の6要素で構成されるものと捉えた。さらに、それら6要素のうち、辛味、甘味、酸味は特に人間の味の受け入れやすさに強く影響すると考え、味ネットワークの訓練タスクには、入力した料理に対して辛味、甘味、酸味がどの程度かを識別する訓練タスクを課することで味表現ベクトル $V_{taste}$ を出力できるように学習を行う。

辛味、甘味、酸味は全て次の1, 2, 3のいずれかに分類されると設定した。

1. 全く(辛く、甘く、酸っぱく)ない
2. 少し(辛い、甘い、酸っぱい)
3. とても(辛い、甘い、酸っぱい)

教師データとなる料理の辛味、甘味、酸味は3次元のOne-hotベクトルで表現し、味識別器で出力される $V_{hot}, V_{sweet}, V_{sour}$ とのクロスエントロピー誤差をLossとして訓練を行う。

### 3.5 データセット

2節で紹介したJavier Marínら[11]は、クロスモーダルな検索タスクの実現だけではなく、Recipe1M, Recipe1M+という料理に関する大規模コーパスデータセットの構築を行なった。Recipe1Mは、24の料理ウェブサイトから料理のレシピデータと画像データを集めることで構成されており、レシピの数はおよそ100万個、画像の数はおよそ80万枚である。さらに、検索エンジンを用いてRecipe1Mに含まれる料理の画像検索を行い、検索結果を料理の画像データとして補強したデータセットがRecipe1M+である。Recipe1M及びRecipe1M+には次のようなデータが含まれている。

(料理ID, 料理名, 食材, 調理手順, 料理画像, コースラベル)



表 1 辛味代表料理レシピデータの例

recipeID	00089de6f7
title	Spicy Korean Temple Noodles
ingredient	2 tablespoons gochujang (Korean chile paste), 2 tablespoons sugar, preferably organic, 9 ounces thin udon, :
instruction	1. In a large bowl, mix together the gochujang, ... 2. Bring to a boil 3 quarts of water ... 3. Add the noodles, and when the water returns ... :



図 4 辛味代表料理画像データの例 (出典: [15])

各レシピは料理 ID で一意に識別され、料理名や食材、調理手順、前菜やデザートといったコースラベルのデータを含む。また、1 つのレシピに対して複数の料理画像が紐づけられている。

本研究では、モデルの学習用教師データ及びテストデータに Recipe1M の一部の料理を使用する。Recipe1M には辛味、甘味、酸味の程度を示すデータは存在しないので、筆者が味のラベルづけを行い、教師データ及びテストデータとしている。

辛味、甘味、酸味に対して効率的に学習を行うために、ラベル 2, 3 に該当する料理、つまり各味が強く現れている料理を正例の教師データとして選定した。その具体的な方法を辛味の場合を例に次に示す。

(1) Recipe1M に含まれる料理から辛味が顕著に表れている料理を代表料理として 3 つ抽出する。代表料理には、例えば表 1, 図 4 の料理が存在する。

(2) Recipe1M 内の各料理データに対して、 $V_{im}^{tar}, V_{rec}^{tar}$  を計算し、 $V_{im}^{tar}, V_{rec}^{tar}$  とする。(1) で抽出した代表料理 3 つも同様に計算し、それぞれ  $V_{im}^i, V_{rec}^i (i = 1, 2, 3)$  とする。

(3) 代表料理との間の画像に関する類似度  $sim_{image}$ 、レシピに関する類似度  $sim_{recipe}$  をベクトル間のコサイン類似度を計算することによって求める。

$$sim_{image} = \frac{\cos(V_{im}^{tar}, V_{im}^1) + \cos(V_{im}^{tar}, V_{im}^2) + \cos(V_{im}^{tar}, V_{im}^3)}{3}$$

$$sim_{recipe} = \frac{\cos(V_{rec}^{tar}, V_{rec}^1) + \cos(V_{rec}^{tar}, V_{rec}^2) + \cos(V_{rec}^{tar}, V_{rec}^3)}{3}$$

表 2 各味ラベルの数と割合

味	全くない (1)	少し (2)	とても (3)
辛味	51 件 (85%)	7 件 (11.7%)	2 件 (3.3%)
甘味	25 件 (41.7%)	31 件 (51.7%)	4 件 (6.7%)
酸味	40 件 (66.6%)	18 件 (30.0%)	2 件 (3.3%)

表 3 交差検証での辛味、甘味、酸味の正解率

fold	hot_accuracy	sweet_accuracy	sour_accuracy
1	83.3%	85.0%	71.6%
2	81.7%	86.7%	66.7%
3	85.0%	75.0%	78.3%
4	90.0%	80.0%	75.0%
5	81.7%	73.3%	73.3%
6	90.0%	86.7%	73.3%
7	73.3%	88.3%	68.3%
8	88.3%	71.7%	71.6%
9	78.3%	83.3%	75.0%
10	83.3%	78.3%	73.3%
Average	83.5%	80.8%	72.7%

さらに、料理としての類似度  $sim_{food}$  を  $sim_{image}, sim_{recipe}$  の加重平均をとることで求める。

$$sim_{food} = \frac{1}{10} sim_{image} + \frac{9}{10} sim_{recipe}$$

(4) (3) で得た  $sim_{food}$  が高いものを上から順に 100 件採用する。

また、正例のサンプリングだけでなく、負例のサンプリングも行うために、 $sim_{food}$  が低い料理も順に 100 件採用した。以上の手順で、計 600 件の料理で構成されるデータセットを構築して利用する。

### 3.6 味ネットワークの学習と評価

3.5 節で扱ったデータセット 600 件をランダムに 9:1 に分割し、9 割に当たる 540 件の料理データを学習用データ、残りの 1 割をモデルの性能を測定するテスト用データとした。損失関数はクロスエントロピー誤差、最適化手法は Adam、学習率は 0.001 で固定して学習を行なった。

学習したモデルの性能をテスト用データ 60 件で測定したときの正解率は、辛味ラベルに対しては 85%、甘味ラベルに対しては 85%、酸味ラベルに対しては 75% であった。ただし、テストデータの各ラベルの数と割合は表 2 に示されるとおりである。

さらに  $k = 10$  とした  $k$  分割交差検証を行なった結果、表 3 の正解率が得られた。

交差検証での計 10 個の分割の平均正解率は、辛味ラベルが 83.5%、甘味ラベルが 80.8%、酸味ラベルが 72.7% であり、テストデータ及び交差検証の結果から、各味に対しておよそ 7 割から 8 割程度の正解率を性能として持つ味識別器が実現できたと考えられる。

## 4 ユーザ実験とケーススタディ

### 4.1 ユーザ実験

3.6 節では、味ネットワークの性能をテストデータ及び交差検証を用いることで評価した。本節ではさらにユーザ実験を行うことで辛味、甘味、酸味の味分類がどれだけ正しくできているかどうかを評価する。

はじめに、ユーザ実験で用いるデータセットについて説明する。本実験では Kaggle で公開されているデータセット “foodRecSys-V1”<sup>1</sup> を使用する。このデータセットはインターネット上のレシピサイトで大きな規模を誇る “Allrecipes.com”<sup>2</sup> から、2000 年～2018 年の間に投稿されたレシピデータを収集し、構築された。foodRecSys-V1 には、およそ 116 万人のユーザ、5 万個のレシピ、そして約 380 万件のユーザとレシピとのインタラクションが次の形で格納されている。

- レシピファイル: レシピ ID, 料理名, 平均評価, 画像 URL, 評価数, 食材, 調理手順, 栄養成分, レビュー
- インタラクションファイル: ユーザ ID, レシピ ID, 評価, 最終更新日時

レシピファイル及びインタラクションファイルから、10 種類以上の料理の食事履歴が存在していて、さらにその中に料理の偏りがある、もしくは味や食材といった観点で特徴ある料理が含まれているユーザを対象に、5 人のユーザと彼らの食事履歴データを抽出し、Recipe1M の同料理または同名の料理に置き換えることでユーザ実験の対象データを用意した。

ユーザ実験の被験者は、学部生 1 名、修士 2 名を対象とした。(以降、被験者 A, 被験者 B, 被験者 C とする。) 筆者が用意した Google フォームで計 59 個の料理を提示し、全ての料理に対して辛味、甘味、酸味それぞれに“(辛く, 甘く, 酸っぱく) ない”, “少し (辛い, 甘い, 酸っぱい)”, “とても (辛い, 甘い, 酸っぱい)” の 3 段階の回答をしてもらった。

被験者 A, B, C の回答及び被験者 A, B, C の回答のうち最も票が多かったラベルに対する提案モデルの識別正解率は図 5 のような結果が得られた。ただし、多数派を決めるときに 3 つのラベルが全て 1 票ずつで同票だった場合は計算から除いた。

多数決によって得られたラベルに対する正解率は、3.6 節での交差検証によって得られた平均正解率に比べて、辛味が 7.6%, 甘味が 13.6%, 酸味が 17.5% 低い結果となった。これは、味には主観的に定まる側面があり、定量的な測定が難しいことが理由であると考えられる。教師データの辛味、甘味、酸味ラベルは筆者がつけたものであり、その教師データで学習したモデルと被験者の味覚の間には少なからず差が存在すると考えている。

### 4.2 ケーススタディ

本節では受容性と冒険性を両立した料理の提案をケーススタディとして行う。最初に検討した方法は、実際に被験者を募って、過去に食べたことのある料理データを収集し、提案手法に

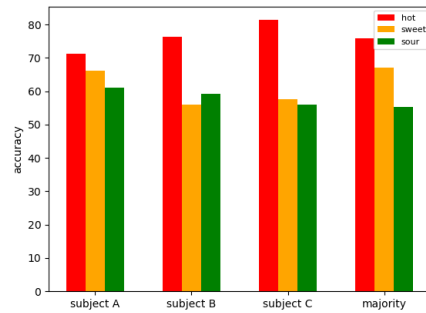


図 5 辛味、甘味、酸味ごとの正解率

よって推薦できると判断された料理を提示して、その料理が受容性と冒険性を備えているか回答してもらうという方法である。しかし、Recipe1M に含まれる料理には海外の料理が多いことを踏まえて、4.1 節のユーザ実験でも使用した 5 人のユーザの食事の履歴データを用いて、ケーススタディを行なった。その手順は次のとおりである。

- 食履歴データに含まれる料理全てに対して味表現ベクトル  $V_{taste}$  を求め、それらが 1 つのクラスタを形成していると見なし、全料理の味表現ベクトルの重心  $V_{center}$  を計算する。そして、各料理の  $V_{taste}$  と  $V_{center}$  との間のマハラノビス距離を測り、それらの中の最大値を  $d_{max}$  とする。さらに、ターゲットとなる料理の  $V_{taste}$  と  $V_{center}$  との間のマハラノビス距離  $d'$  を計算し、 $d' \leq d_{max}$  であれば、対象料理は受容できると判定する。

- 食履歴データに含まれる料理全てに対して食材表現ベクトル  $v_{ingredient}$  を求め、主成分分析を行うことで 2 次元のベクトルへと次元を圧縮して散布図上にプロットする。そして、同様にターゲットとなる料理の食材表現ベクトル  $v_{ingredient}$  も PCA で次元圧縮を行い、同一平面上にプロットしたときに、履歴データによって形成される点の位置の偏りから外れていたそれはユーザにとって冒険性のある料理であると判定する。

計 5 つのケースのうち、受容性と冒険性を両立した料理の提案が上手くいったケースと、失敗したケースの 2 つの分析を以下にまとめる。

**ケース 1** ケース 1 で扱うユーザ 1 の履歴データには、ケーキやクッキー、パイといったデザートが多く含まれている。履歴データに含まれる 16 個の料理とターゲットの料理の食材表現ベクトルの散布図を図 6 に、料理名とマハラノビス距離を表 4 にまとめた。

$d_{max}$  は、料理 6: My Crab Cakes のマハラノビス距離 3.31 に等しい。また、今回ターゲット料理として設定した California Roll は  $d' = 3.03 (\leq d_{max})$  であり、さらに図 6 を見ると、これまで食べてきた料理の食材表現ベクトルが平面左下に偏っているのに対して、ターゲット料理はそれらから大きく離れた右端に位置している。従って、California Roll はユーザが受け入れることのできる受容性を保ちながらも、冒険性を織り込んだ料理である可能性が高く、安心感と好奇心を両立した料理の候補

1: <https://www.kaggle.com/datasets/elisaxxygao/foodrecsysv1>

2: <https://www.allrecipes.com/>

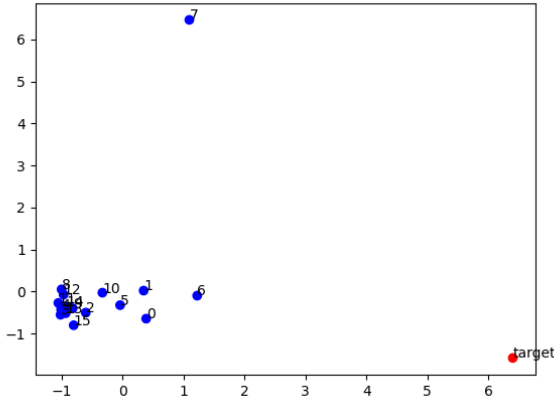


図 6 ユーザ 1 食材表現ベクトルの散布図

表 4 ユーザ 1 の食履歴

index	title	mahalanobis distance
0	Red Velvet cupcake	2.58
1	Turkish Delight	2.06
2	Dark Chocolate cake	2.53
3	Best Big, Fat, Chewy Chocolate Chip Cookie	2.73
4	Molasses Cookies	1.02
5	Tiramisu cheesecake	1.77
6	My Crab Cakes	3.31
7	Mie Goreng	2.86
8	Sweet Potato Casserole	2.18
9	Favorite Old Fashioned Gingerbread	1.93
10	Chocolate Truffle Pie	2.47
11	Bread Pudding	1.12
12	Sweet Potato Pie	1.76
13	Green Tomato Cake	3.23
14	Oatmeal Raisin Cookies	2.40
15	Easy blueberry pie	1.90
target	California Roll	3.03

表 5 ユーザ 2 の食履歴

index	title	mahalanobis distance
0	Pork Barbeque	1.95
1	Spicy Salsa	2.83
2	Italian Pork Tenderloin	2.87
3	Pumpkin, Sweet Potato, Leek and Coconut Milk Soup	1.53
4	Flank Steak a la Willyboy	2.88
5	Emily's Famous Sloppy Joes	1.90
6	Canadian Cod Au Gratin	2.24
7	The Ultimate Chili	1.47
8	Wheat Bran and Walnut Bagels	2.52
9	Easy blueberry pie	1.82
target	Taiwan Ramen	2.98

として挙げることができる。

**ケース 2** ケース 2 で扱うユーザ 2 の履歴データには肉料理やサルサ、チリコンカーンといった辛味が強い料理が含まれている。履歴データに含まれる 10 個の料理とターゲットの料理の食材表現ベクトルの散布図を図 7 に、料理名とマハラノビス距離を表 5 にまとめた。

$d_{max}$  は、料理 4: Flank Steak a la Willyboy のマハラノ

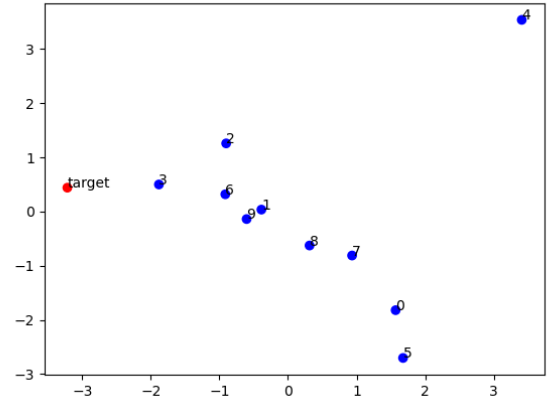


図 7 ユーザ 2 食材表現ベクトルの散布図

ビス距離 2.88 に等しい。今回ターゲット料理として設定した Taiwan Ramen は  $d' = 2.98 (> d_{max})$  であるので、受容性に乏しいと判断できる。また、図 7 を見ると、ターゲット料理の食材表現ベクトルは、普段食している料理の表現ベクトルの偏りから外れている。従って、Taiwan Ramen は冒険性は存在するが、安心感と好奇心を両立した料理としては推薦できないと考えられる。

**結論及び考察** 5 つのケースから、食材表現ベクトルを主成分分析によって次元削減したベクトルは平面上で一定の偏りを形成し、味表現ベクトルからは味の違いを過去の分布との距離を測ることで推定できることがわかった。ただし、10 個を超える料理の中には、例えば、ケース 1 の料理 8: Sweet Potato Casserole と料理 12 の Sweet Potato Pie のように似たような味を持つものにも関わらず食履歴分布の重心からの距離が異なるような場合も存在する。味表現ベクトルが 6 次元のベクトルであることを踏まえると、1 ユーザあたりの食事の履歴データの数にさらに多いデータを使用することが望ましいと考えられ、今後はそのようにデータを収集する予定である。

## 5 おわりに

本研究では、料理に対して味表現ベクトルと食材表現ベクトルを求めて、それらを用いることでユーザの安心感と好奇心をどちらも満たすような料理を発見するという課題に取り組んでいる。味表現ベクトルと食材表現ベクトルの推定は、既存研究のモデルを一部組み込んだ、独自のニューラルネットワークモデルである味ネットワークを実装することで実現した。また、味ネットワークの学習には料理の辛味、甘味、酸味を分類するという訓練タスクを設定し、料理の辛味、甘味、酸味に対する識別器を実装した。この識別器の性能は筆者が用意したテストデータとユーザ実験で評価を行い、テストデータに対しては平均正解率が 81.7%、ユーザ実験で集められた回答に対しては平均正解率が 68.5% という結果であった。

レシピサイトにおけるユーザと料理の相互関係を保存しているデータセットから 5 人のユーザと彼らの食事履歴を抽出し、

それらに提案モデルを適用することで、受容性や冒険性が担保された料理の発見が可能であることをケーススタディで示した。

本研究の課題点とこれからの展望についてまとめる。

- 訓練タスクに用いる教師データが量、質ともに改善余地が大きい。提案モデルの学習には辛味、甘味、酸味それぞれ 540 件の料理データを用いたが、540 という数が学習に十分ではない。従って、これからの研究課題には教師データの数を増やすというタスクが存在する。この課題の解決方法には、クラウドソーシングを用いて不特定多数の人に味のラベルづけを依頼することが考えられる。また、ChatGPT など既存のツールを用いて味のラベルづけをサポートする方法も検討している。ラベルの正確さに対する信頼性に問題はあがるが、1 つの有効な支援手段であると考えられる。

さらに教師データを質的な面から見ても不十分な点はあると考えている。4.1 節で触れたように、味の感じ方は万人に共通せず、少なからず主観的に定まる部分が存在する。この性質を鑑みるに、教師データの味ラベルを本研究のように特定の 1 人が決めるのは真に正しいラベルとは言えないだろう。この問題の解決に対しても不特定多数の人間の力を借りることのできるクラウドソーシングは有効である。

その他、教師データに関する細かい問題として、ラベルづけをできている料理の種類が圧倒的に少ないことや、“(辛く、甘く、酸っぱく) ない” のラベルがつけられている料理の数が“少し(辛い、甘い、酸っぱい)”, “とても(辛い、甘い、酸っぱい)” のラベルを持つ料理の数に比べて多く存在し、全体的にラベルの偏りがあることが課題点である。

- 味表現ベクトル及び食材表現ベクトルを求めることのできる料理が Recipe1M に含まれる料理に限られているという点。現段階では調理手順のエンコード結果が記録されている Recipe1M の料理しか扱えない。本研究が生かされるユースケースとして、旅行先の飲食店でメニューに載っている料理の受容度と冒険度を推定するというシーンが想定されるので、これを実現するために任意の料理レシピに対して味ネットワークを適用できるように拡張する予定である。そして将来的な展望として、任意のレシピに対して味表現ベクトルと食材表現ベクトルを出力できるようにすることを考えている。

さらに発展的な課題及び展望として、モデルの入力を画像データもしくはレシピデータのどちらかで対応できるような変更を挙げる。現在は提案モデルの入力に料理の画像、食材、調理手順の 3 つをとっているが、上で説明したようなユースケースを想定すると、飲食店のメニューにその 3 つのデータが全て揃っている可能性はかなり低いので提案モデルを適用できる場面が少ないことが考えられる。メニューに記載されている料理の情報の中で代表的なものは料理画像または料理タイトルであるため、これらのどちらかを入力することで表現ベクトルを出力できるようなモデルの改善が望まれる。

## 謝 辞

本研究の一部は総務省 SCOPE (201607008) による。

- [1] Xiaoyan Gao, Fuli Feng, Xiangnan He, Heyan Huang, Xinyu Guan, Chong Feng, Zhaoyan Ming, and Tat-Seng Chua. Hierarchical attention network for visually-aware food recommendation. *IEEE Transactions on Multimedia*, Vol. 22, No. 6, pp. 1647–1659, 2020.
- [2] Mayumi Ueda, Syungo Asanuma, Yusuke Miyawaki, and Shinsuke Nakajima. Recipe recommendation method by considering the user's preference and ingredient quantity of target recipe. *Lecture Notes in Engineering and Computer Science*, Vol. 2209, pp. 519–523, 03 2014.
- [3] Antónia Correia, Seongseop Kim, and Metin Kozak. Gastronomy experiential traits and their effects on intentions for recommendation: A fuzzy set approach. *International Journal of Tourism Research*, Vol. 22, , 12 2019.
- [4] マーボー豆腐 〜甘口と辛口は鍋を分ける方が良い〜 - ■炎のクリエイター日記. <https://blog.goo.ne.jp/sugichan-goo/e/8309dec20051f23b692c9753544a6cf5?fm=rss>. (Accessed on 01/30/2023).
- [5] 電子レンジで作る カレー | レシピ | エスビー食品株式会社. <https://www.sbfoods.co.jp/recipe/detail/08678.html>. (Accessed on 01/30/2023).
- [6] Mapo tofu recipe — authentic chinese recipes — lost plate china. <https://lostplate.com/recipes/mapo-tofu-recipe/>. (Accessed on 01/30/2023).
- [7] Tom Yam soup thailand recipe - food.com. <https://www.food.com/recipe/tomyam-soup-thailand-479094>. (Accessed on 01/30/2023).
- [8] Hot and sour soup (ready in 20 minutes!) - little spoon farm. <https://littlespoonfarm.com/hot-and-sour-soup-recipe/>. (Accessed on 01/30/2023).
- [9] Hannele Kauppinen-Räsänen, Johanna Gummerus, and Katariina Lehtola. Remembered eating experiences described by the self, place, food, context and time. *British Food Journal*, Vol. 115, , 05 2013.
- [10] Francisco Javier Jiménez-Beltrán, Tomás López-Guzmán, and Francisco González Santa Cruz. Analysis of the relationship between tourism and food culture. *Sustainability*, Vol. 8, No. 5, 2016.
- [11] Javier Marin, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba. Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images, 2018.
- [12] Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. Skip-thought vectors, 2015.
- [13] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge, 2014.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [15] Spicy korean temple noodles recipe - nyt cooking. <https://cooking.nytimes.com/recipes/1017290-spicy-korean-temple-noodles>. (Accessed on 01/30/2023).