

集合型 Web オブジェクトの構成要素に対する 追加・削除要素の推薦とそのレシピデータへの応用

佃 洸撰[†] 中村 聡史^{††} 田中 克己^{††}

[†] 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

^{††} 京都大学大学院情報学研究科社会情報学専攻 〒 606-8501 京都府京都市左京区吉田本町

E-mail: [†]{tsukuda,nakamura,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 本論文では、集合型オブジェクトに対して、追加・削除可能な構成要素を発見し推薦する手法を提案する。Web 上にある料理レシピ情報などは、複数・異種の食材の集合型オブジェクトと位置づけることができる。レシピに対して食材をあと1つ追加または削除したいということはよくあるが、従来のシステムではこうした食材の発見は困難である。本研究では料理レシピのような集合型オブジェクトの構造を分析することで、追加・削除可能な食材を推薦するシステムについて述べる。例えばこれにより、ある料理レシピに対して食材を追加・削除することで、ある料理タイプに近づけることが可能となる。また、検証実験により、システムによるオブジェクトの構造分析の評価や、システムが推薦する追加・削除食材の妥当性の評価を行った。これにより、どのようなタイプのオブジェクトに対して提案手法が有効に働くのかがわかった。

キーワード 集合型 Web オブジェクト, オブジェクト構造, カテゴリ

1. はじめに

近年、Web 上で提供される情報やサービスは飛躍的に増加している。ユーザはそれらの情報について記述されたページを検索及び閲覧することにより、京都の桜の観光に関する情報や新撰組の池田屋事件に関する情報、料理のレシピなどのような多様な情報を得ることができる。

ここで、ユーザがあるページを閲覧した場合、ユーザはそのページに記述されている情報しか得ることができないため、「もう少し記述を追加した情報が欲しい」または「ここから何か1つを省きたい」という意図をユーザがもつことが多々ある。ユーザが追加または削除を望む状況としては、具体的に以下のような場合が挙げられる。

- あるユーザは春の京都を観光するため、京都の桜の観光地のコースとして清水寺、南禅寺、平安神宮、円山公園が紹介されているページに辿りついた。このユーザはすべての場所に行く時間はないので、この中から1つ削除するとしたらどこなのを知りたいと思っている。
- あるユーザが池田屋事件について調べている。ユーザが辿りついたページには、近藤勇、沖田総司、永倉新八が記述されているが、登場人物が少なく物足りない。そこでこのユーザはあと1人このメンバーに追加したようなストーリーが記述されたページに辿りつきたいと考えている。
- あるユーザがおでんを作るため、おでんのレシピページに辿りついた。しかし、このユーザはこのレシピで使われている食材に物足りなさを感じたので、あと1つ食材を追加するとしたら何を追加するべきかを知りたいと思っている。

もちろん、上記以外にも多数の状況が考えられる。しかし、現時点で上記のような検索を可能とするサービスは存在しない。また、一般的な検索エンジンを利用して上記のような検索を行

うのは非常に困難である。そこで本研究では、ユーザが現在閲覧している情報をよりユーザの求めるものへと近づけることを可能とするため、その情報に含まれている要素集合に注目し、何らかの要素を1つ追加したり1つ削除したりする手法を実現する。ここで我々が取り扱う情報を“集合型 Web オブジェクト”と呼ぶ。集合型 Web オブジェクトとは、多くの要素集合により成り立つものであり、その組み合わせによりその特性が変化するというものである。ユーザは、要素の組み合わせを変更することにより、目的とする集合型 Web オブジェクトへと辿りつくことが可能となる。

また、本研究では集合型オブジェクトの構造という考えを導入し、オブジェクトの構造の安定度を、オブジェクトを構成する要素の組み合わせを用いて計算する。そして、その安定度の変化を利用することで追加・削除すべき要素の発見を行う。要素の追加及び削除については、その根拠として、ユーザが着目している集合型 Web オブジェクトの構造の安定度と、追加・削除可能な要素を追加または削除した後の構造の安定度を提示する。この構造の安定度を参考にすることで、ユーザは自分が着目している集合型 Web オブジェクトを典型的なオブジェクトに近付けたり、非典型的なオブジェクトに近付けることが可能となる。本研究では特に集合型 Web オブジェクトとしてレシピに注目し、あるレシピに対する食材の追加・削除を実現する。

2. 集合型 Web オブジェクト

2.1 集合型 Web オブジェクトとは

ユーザが「京都の桜の観光」について調べているとき、「京都の左京区の桜の観光地」、「ライトアップされる京都の桜の観光地」、「京都の穴場な桜の観光地」などのように、京都の桜の観光に関する観光コースは無数に存在する。このとき、それぞれの観点に基づいた観光コースは、「二条城」や「平安神宮」、「清

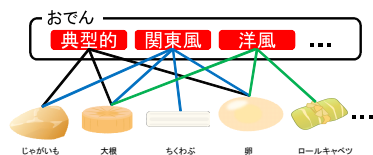


図 1 カテゴリー「おでん」における集合型 Web オブジェクト

水寺」,「大河内山荘」などのように多くの京都の桜に関する観光地を要素とし、その組み合わせにより変化するものである。そして、それぞれの観光地の組み合わせは「京都の桜の観光」という同じカテゴリ内のものである。同じカテゴリ内のものであっても、観光地の組み合わせによって、その組み合わせがもつ特性は変化する。同様に、池田屋事件というトピックに関する物語も「新撰組から見た池田屋事件」,「長州藩士から見た池田屋事件」,「坂本龍馬から見た池田屋事件」などがある。こうした各種の物語も、様々な登場人物の組み合わせによって構成されており、登場人物の組み合わせにより変化する。また、「おでん」に関するレシピについても「京風おでん」や「関東風おでん」,「洋風おでん」といったバリエーションがあり、それぞれは大根やじゃがいも、ちくわぶや卵、海老芋やロールキャベツなど各種の具材の組み合わせにより変化する(図 1)。

このように、ユーザが最終的に辿り着く“オブジェクト”「京都の典型的な桜の観光コース」,「新撰組から見た池田屋事件」,「京風おでん」は「京都の桜の観光」,「池田屋事件」,「おでん」などのようなある“カテゴリ”に属しており、そのオブジェクトを構成する「二条城」,「近藤勇」,「大根」といった“要素”の組み合わせにより、その特性を変えるものである。我々は、このように各要素の組み合わせにより構成されるオブジェクトを“集合型 Web オブジェクト”とよぶ。図 2 はこれをモデル化したものである。つまり、あるカテゴリを規定すると、そのカテゴリに属する集合型 Web オブジェクトは無数に存在し、各集合型 Web オブジェクトはいくつかの要素から構成される。そして、どのような要素から構成されるかによって、集合型 Web オブジェクトのもつ特性が決まる。

2.2 集合型 Web オブジェクトの構造と構造の安定度

構成要素の中には、あるカテゴリに含まれるほぼ全ての集合型 Web オブジェクトに含まれる重要な要素と、いくつかの集合型 Web オブジェクトにのみ含まれる修飾的な要素が存在する。例えば「池田屋事件」というカテゴリにおいて「新撰組から見た池田屋事件」,「長州藩士から見た池田屋事件」,「坂本龍馬から見た池田屋事件」のいずれの観点で池田屋事件を語る場合でも「近藤勇」という人物は欠かせない要素である。また、「おでん」というカテゴリに含まれる集合型 Web オブジェクトについても「京風おでん」や「関東風おでん」,「洋風おでん」など様々な種類のおでんが存在するが、どのような種類のおでんをつくる場合でも「卵」や「大根」といった食材は欠かせない要素である。

そこで本研究では、集合型 Web オブジェクトをその構成要素がある 2 層構造を成すものとして考える。ここで、2 層とは構造の土台となる層とその上の層のことである。そして、構造の土台を構成する要素を、カテゴリ内のいずれの集合型 Web オブジェクトにも欠かせない要素とし、土台の上の層を構成する要素を、カテゴリ内のいくつかの集合型 Web オブジェクトにのみ含まれる要素とする。例えばおでんの場合「卵」や「大

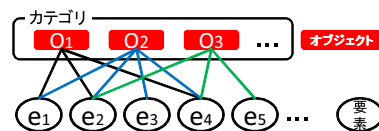


図 2 一般の集合型 Web オブジェクトモデル

根」などはどのような種類のおでんにも欠かせない食材であるので、土台を構成する要素となり、「はんぺん」や「海老芋」,「ロールキャベツ」などは、すべての種類のおでんに必要とは限らないので上の層を構成する要素となる。

ここで、土台となる要素をすべて含んだ集合型 Web オブジェクトは安定した構造をしており、土台となる要素が欠けている集合型 Web オブジェクトは不安定な構造をしていると考える。また、土台の上の層を構成する要素の中でも、多数の集合型 Web オブジェクトに含まれる要素のみから成る集合型 Web オブジェクトほど構造が安定しており、より少数の集合型 Web オブジェクトにのみ含まれる要素から成る集合型 Web オブジェクトほど構造が不安定になる。以上より、オブジェクトの構造は安定度の高い順に以下に挙げる 4 種類に分類可能になる。

(a) 土台と上層が共に安定している集合型 Web オブジェクト

土台となる要素を全て含んでおり、かつ上層にある要素も十分に含んでいる集合型 Web オブジェクト。構造が最も安定している。おでんの例では、「一般のおでん」がこれにあたる。

(b) 土台は安定しているが、上層が不安定な集合型 Web オブジェクト

土台となる要素を全て含んでおり、上層の要素に非典型的な要素が含まれている集合型 Web オブジェクト。おでんの例では、「京風おでん」や「洋風おでん」などがこれにあたる。

(c) 土台が不安定で、上層が安定している集合型 Web オブジェクト

土台となる要素で欠けている要素があるが、上層の要素は典型的な要素のみからなる集合型 Web オブジェクト。おでんの例では、卵を入れていない「我が家のおでん」などがこれにあたる。

(d) 土台と上層が共に不安定な集合型 Web オブジェクト

土台となる要素が欠けており、かつ上層の要素に非典型的な要素が含まれている集合型 Web オブジェクト。構造が最も不安定である。おでんの例では、卵や大根が入っておらず、ロールキャベツが入っている「創作洋風おでん」などがこれにあたる。

以上の (a) ~ (b) をそれぞれ図で表すと、図 3 のようになる。本研究では、この土台となる要素及び上層となる要素、構造の安定度については、あるカテゴリの集合型 Web オブジェクトを大規模に収集し、集約することで求める。

2.3 集合型 Web オブジェクトに対する要素の追加・削除

ある集合型 Web オブジェクトに対して、要素を 1 つ追加または削除すると、その前後で構造の安定度は変化する。そこで本研究では、ユーザが注目しているオブジェクトの構造とその安定度を求め、要素を追加または削除した後の構造の安定度に注目した追加・削除を行う。これにより、ユーザは自分が注目している集合型 Web オブジェクトを典型的なものに近付けたり、非典型的なものに近付けることが可能となる。

本研究では、カテゴリと集合型 Web オブジェクト、及びそ

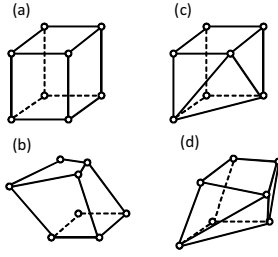


図3 集合型 Web オブジェクトの様々な構造

の構成要素が明確である料理のレシピを対象とする．将来的には，ユーザが興味を持った任意のオブジェクトに対する要素の追加及び削除を可能にする．

3. 典型度に基づく集合型 Web オブジェクトの構造の構築及び安定度の数値化

本章では，あるカテゴリに属する集合型 Web オブジェクト群と，各集合型 Web オブジェクトの構成要素が得られた場合に，集合型 Web オブジェクト構造を構築する方法及び，その安定度を数値化する方法について述べる．

3.1 追加・削除要素を推薦するまでの流れ

ユーザが着目している集合型 Web オブジェクトに対して，システムが追加及び削除可能な要素を推薦するまでの流れは以下の通りである．

- (1) ユーザが着目している集合型 Web オブジェクトが含まれるカテゴリを求める
- (2) (1) で求めたカテゴリに含まれる集合型 Web オブジェクト群を収集する
- (3) 各集合型 Web オブジェクトを構成する要素を抽出する
- (4) (2), (3) を元に，対象とするカテゴリで利用される各要素が土台として使われるものか上層で使われるものかを判断し，最も安定している構造を求める
- (5) (2), (3), (4) を元に，ユーザが着目している集合型 Web オブジェクトの構造及びその安定度を求める
- (6) ユーザが着目しているオブジェクトに対して追加及び削除可能な要素を求める

本研究では上記の (1)～(6) のうち，特に (4)～(6) に注目し，手法を提案する．

3.2 集合型 Web オブジェクトの構造の構築

集合型 Web オブジェクトの構造を分析するためには，まずあるカテゴリにおいて土台を構成する要素と上層を構成する要素を分離する必要がある．そのため，本研究では要素間の共起度に基づいて，要素の典型度を分析し，土台と上層を構成する要素を分類する．また，ユーザが着目している集合型 Web オブジェクトの安定度を求めるために，そのオブジェクトが属するカテゴリ内において，最も典型的な集合型 Web オブジェクトを構成する要素集合を求める．なお，以下では，あるカテゴリに属する集合型 Web オブジェクトの集合を $O = \{o_1, o_2, \dots, o_n\}$ ，集合型 Web オブジェクト集合 O 内の集合型 Web オブジェクトの構成要素として 1 度でも出現する全要素集合を $E = \{e_1, e_2, \dots, e_m\}$ とする．ここで， n はカテゴリ内の集合型 Web オブジェクトの数であり， m はカテゴリ内の集合型 Web オブジェクトの構成要素の数である．

2 つの異なる要素 e_i と e_j の共起度を求める手法として，本

稿では次の式を用いる．

$$co(e_i, e_j) = \frac{|R(e_i) \cap R(e_j)|}{\min(|R(e_i)|, |R(e_j)|)} \quad (1)$$

ただし， $R(e)$ はカテゴリ内で要素 e を含む集合型 Web オブジェクトの集合を表す．

3.2.1 土台を構成する要素の抽出

土台を構成する要素集合 B を発見する方法は以下の通りである．

- (1) 全ての要素 $e \in E$ の中で， $|R(e)|$ の値が閾値 α を越えるものを土台候補集合 S に分類する．
- (2) $B \leftarrow \phi$ とし，以下のアルゴリズムを繰り返す
 - (a) S の中から， $|R(e)|$ が最大，つまり $e_i = \operatorname{argmax}_{e \in S} |R(e)|$ となる要素 e_i を求める．
 - (b) (a) で求めた要素を S から B へ移動させる．
 - (c) S に含まれる各要素 $e \in S$ について， B に含まれる要素群との共起度が最大となる要素 e_{max} と，共起度 c を求める．
 - (d) c が閾値 β_1 以上のとき， $S \leftarrow S - \{e_j\}$ ， $B \leftarrow B \cup \{e_j\}$ とし，手続き (c) に戻る．
 - (e) c が閾値 β_1 未満であれば終了し，手続き (3) に移動する．
- (3) 最終的に得られた要素集合 B を土台を構成する要素集合とする．

3.2.2 典型的集合型 Web オブジェクトを構成する要素集合の抽出

あるカテゴリにおいて，土台となる要素集合は，そのカテゴリ内のどのような集合型 Web オブジェクトにも含まれるべき要素集合である．よって，カテゴリ内で最も典型的な集合型 Web オブジェクトも必ずそのカテゴリで土台となる要素を全て含んでいる．そこで，3.2.1 節の手法により抽出された，カテゴリ内の集合型 Web オブジェクトの土台となる要素集合 B を用いて，典型的集合型 Web オブジェクトを構成する要素集合 T を以下の方法により求める．

- (1) 全ての要素 $e \in E$ の中で， $|R(e)|$ の値が閾値 α を越えるものを典型的集合型 Web オブジェクトを構成する要素の候補集合 S に分類する．ここで，集合 B に含まれる要素は集合 T に含まれることは決まっているので，集合 S の要素から集合 B に含まれる要素は除く．
- (2) $T \leftarrow B$ とし，以下のアルゴリズムを繰り返す．
 - (a) S に含まれる各要素 $e \in S$ について， T に含まれる要素群との共起度が最大となる要素 e_{max} と，その時の共起度 c を求める．
 - (b) c が閾値 β_2 以上のとき， $S \leftarrow S - \{e_j\}$ ， $T \leftarrow T \cup \{e_j\}$ とし，手続き (a) に戻る．
 - (c) c が閾値 β_2 未満であれば終了し，手続き (3) に移動する．
- (3) 最終的に得られた要素集合 T を典型的集合型 Web オブジェクトを構成する要素集合とする．

3.3 集合型 Web オブジェクトの構造の安定度の数値化

本研究では，ある集合型 Web オブジェクトの構造の安定度を求めるために，安定度を求めたい集合型 Web オブジェクトに含まれている要素集合と，そのカテゴリ内の典型的集合型 Web オブジェクトを構成する要素集合との差異を元に，安定度を求める．このとき，最も典型的な要素集合から成る集合型

Web オブジェクトの安定度を 1 とし、安定度が 1 に近いほど安定した構造をしており、値が小さくなるほど構造が不安定であるとする。ある集合型 Web オブジェクト o に含まれる要素の集合を $E_o = \{e_1, e_2, \dots, e_l\}$ とすると、要素集合 E_o の安定度 $Stability(E_o)$ は次式により求められる。ただし、 l は安定度を求める集合型 Web オブジェクトに含まれる要素の数である。

$$Stability(E_o) = affinity(E_o) - \delta \cdot difference(E_o) \quad (2)$$

$$-\delta \leq Stability(E_o) \leq 1$$

ここで、 $affinity(E_o)$ は要素集合 E_o の全ての要素間の相性を数値化したものであり、 $difference(E_o)$ は要素集合 E_o と T の差異を数値化したものである。これらの値の求め方を以下に述べる。

$affinity(E_o)$ は、要素集合 E_o の全ての要素間の共起度の平均であり、次式により求める。

$$affinity(E_o) = \frac{1}{|E_o|C_2} \sum_{e_i, e_j \in E_o} co_1(e_i, e_j) \quad (3)$$

$$0 \leq affinity(E_o) \leq 1$$

なお、 $co_1(e_i, e_j)$ は以下のように定義する。

$$co_1(e_i, e_j) = \begin{cases} 1 & co(e_i, e_j) > \theta \\ 0 & otherwise \end{cases} \quad (4)$$

つまり、 E_o 内に相性の悪い要素数が多くなるほど、 $affinity(E_o)$ の値は小さくなる。

また、 $difference(E_o)$ は次式により求められる。

$$difference(E_o) = (1 - \mu) \cdot \sum_{e_i \in E_o - T} \frac{1 - R'(e_i)}{|E_o - T|} + \mu \cdot \frac{\sum_{e_i \in T - E_o} R'(e_i)}{\sum_{e_i \in T} R'(e_i)} \quad (5)$$

$$0 \leq difference(E_o) \leq 1$$

$R'(e)$ は、カテゴリ内で $|R(e)|$ の値が最大になる要素 e_{max} を求め、次のように定義する。

$$R'(e) = \frac{|R(e)|}{|R(e_{max})|} \quad (6)$$

式 (5) の右辺第一項は、集合 E_o に含まれていて集合 T に含まれない要素がある場合に 0 から 1 の値をとり、集合 $E_o - T$ の中にカテゴリ内で出現頻度の低い要素が多くなるほど大きな値をとる。つまり、集合 E_o に、そのカテゴリ内で珍しい要素が多く含まれているほど、集合 T とのずれが大きくなるとみなす。また、式 (5) の右辺第二項は、集合 T に含まれていて集合 E_o に含まれない要素がある場合に 0 から 1 の値をとり、集合 $T - E_o$ の中にカテゴリ内で出現頻度の高い要素が多くなるほど大きな値をとる。つまり、集合 E_o に、そのカテゴリ内で重要な要素が多く欠けているほど、集合 T とのずれが大きくなるとみなす。

3.4 追加・削除要素の抽出

本節では、ユーザが着目している集合型 Web オブジェクトに対して、追加及び削除可能な要素を抽出する方法について述べる。

3.4.1 追加要素の抽出

追加する要素として推薦すべき要素の性質は、追加すること

で現在の集合型 Web オブジェクトが典型的・非典型的な集合型 Web オブジェクトに近づく要素である。つまり、追加後の集合型 Web オブジェクトの安定度が最大となる要素と最小となる要素を推薦する。ただし、追加後の集合型 Web オブジェクトの安定度が低くなりすぎる要素は、追加する要素の候補からはずす。これらの条件を満たすように、追加要素を抽出する方法は以下の通りである。

- (1) ユーザが着目している集合型 Web オブジェクトが含まれるカテゴリにおいて土台となる要素を 3.2 節の方法で求める。土台となる要素の集合を B とする。
- (2) ユーザが着目している集合型 Web オブジェクトが含まれるカテゴリ内の集合型 Web オブジェクトを構成する全要素集合 $E = \{e_1, e_2, \dots, e_m\}$ の中で、 $|R(e_i)|$ の値が閾値 γ を越えていて、かつユーザが着目している集合型 Web オブジェクト E_o に含まれていない要素集合を追加候補集合 S に分類する。つまり、

$$S = \{e | e \in E - E_o \wedge |R(e)| > \gamma\} \quad (7)$$

- (3) 集合 S の各要素 e について、集合 E_o の要素で $|R(e_i)|$ の値が閾値 γ を越えている全ての要素との共起度を式 (1) を用いて求め、共起度が 0 となる要素が集合 E_o に含まれる場合は要素 e を集合 S から除く。
- (4) 集合 S に含まれる各要素を E_o に追加した場合の安定度を 3.3 節の方法により求め、追加後の安定度が高くなる順にランク付けする。

3.4.2 削除要素の抽出

ユーザが着目している集合型 Web オブジェクトから削除する要素として推薦すべき要素の性質は、土台以外の要素であり、かつ削除する要素間で安定度の差が大きくなる要素である。削除する要素を抽出する方法は以下の通りである。

- (1) ユーザが着目している集合型 Web オブジェクトが含まれるカテゴリにおいて土台となる要素を 3.2 節の方法で求める。土台となる要素集合を B とする。
- (2) ユーザが着目している集合型 Web オブジェクトを構成する要素の内、土台となる要素以外の要素集合を削除の候補とし、削除候補集合 D に分類する。つまり、

$$D = \{e | e \in E_o - B\} \quad (8)$$

- (3) D に含まれる要素について、各要素を E_o から削除した場合の安定度を 3.3 節の方法により求め、安定度が高くなる順にランク付けする。

4. レシピへの適用

集合型 Web オブジェクトをあるレシピとしたとき、カテゴリはその料理の一般的な名前であり、要素は利用される食材となる。また、そのカテゴリに含まれる集合型 Web オブジェクト群はその料理の一般名で分類されている全てのレシピとなる。そこで、本研究では我々の手法の有効性を検証するため、3 章で提案した手法をレシピに適用し、あるレシピに対して追加・削除可能な食材を推薦するシステムを実現する。

4.1 レシピの抽出

レシピ構造を構築する前処理として、COOKPAD [1] に登録されている約 6 万種類のレシピについて、それぞれ以下の項目を抽出し、データベースに格納する。

- レシピの ID

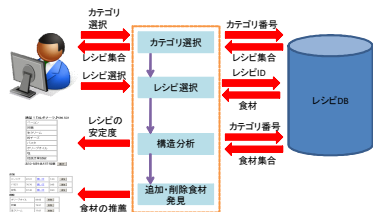


図 4 システムのイメージ図

- レシピのカテゴリ番号
- レシピのカテゴリ名
- レシピ名
- 使われている食材

なお、ここでは食材の表記ゆれの問題をあらかじめ解決し、データベースに格納している。

4.2 システムの流れ

ユーザがシステムを用いてレシピへの食材の追加及び削除を行う流れは以下ようになる。システムのイメージを図 4 に示す。

- (1) ブラウジングにより、ユーザはあるレシピを選択する。
- (2) システムは (1) のレシピが含まれるカテゴリ内の全てのレシピを取得し、次にその全てのレシピで使われている食材を取得する。
- (3) システムはユーザが選択したレシピの構造を分析し、ユーザに現在ユーザが選択しているレシピの安定度を提示する。
- (4) システムは追加・削除することで構造が安定する食材、不安定になる食材を求め、追加・削除可能な食材の候補を、追加・削除後の構造の安定度の数値と共にユーザに提示する。
- (5) ユーザが現在のレシピに満足していないとき、ユーザは提示された食材の中から追加または削除したい食材をひとつ選択する。このとき、追加または削除した要素に基づき、システムは新たなレシピを求め、ユーザに提示し、(3) へ戻る。ユーザが満足しているときはここで終了する。

4.3 モジュール

4.3.1 カテゴリ選択モジュール

カテゴリ選択モジュールでは、ユーザがカテゴリを選択すると、そのカテゴリに属する全てのレシピのタイトルをデータベースから取得し、表示する。つまり、レシピの 1 つ 1 つが、食材を構成要素とする集合型 Web オブジェクトであるとみなし、カテゴリ内の全ての集合型 Web オブジェクトをユーザに提示する。

4.3.2 レシピ選択モジュール

レシピ選択モジュールでは、ユーザがレシピを選択すると、そのレシピで使われている食材をデータベースから取得する。これは、ユーザが着目する集合型 Web オブジェクトとその構成要素を取得することに相当する。レシピ選択モジュールは、取得した食材集合を構造分析モジュールに渡す。

4.3.3 構造分析モジュール

構造分析モジュールは、レシピ選択モジュールから受け取った情報とカテゴリ内の各レシピで使われている食材の組み合わせを元に、3.3 節で述べた手法を用いて、ユーザが選択したレシピの構造の安定度を求めてユーザに提示するモジュールである。構造分析モジュールは、ユーザが選択したレシピで使われている食材や、土台となる食材の集合、典型的な食材の集合な



図 5 「カルボナーラ」のカテゴリ内のあるレシピに対する実行結果

どを追加・削除食材発見モジュールに渡す。

4.3.4 追加・削除食材発見モジュール

追加・削除食材発見モジュールでは、構造分析モジュールから受け取った情報をもとに、3.4 節で述べた手法を用いて、ユーザが選択したレシピに対して追加及び削除可能な食材を発見し、ユーザに提示する。ユーザはシステムが推薦した食材に対して追加または削除を実行する。

4.4 実行例

システムの実行例を図 5 に示す。まず、ユーザはいずれかのカテゴリを選択し、さらにカテゴリ内のレシピ群から 1 つのレシピを選択する。次に、システムはユーザが選択したレシピの構造の安定度を提示し、ユーザが選択したレシピに対して追加・削除可能な食材を推薦する。

5. 検証及び考察

本章では、あるカテゴリに属する集合型 Web オブジェクトには、欠かすことのできない土台となる要素及び典型的な集合型 Web オブジェクトを構成する要素集合が存在すること、そして典型的な集合型 Web オブジェクトほど構造の安定度が高くなるということ、さらにシステムが推薦する食材によってレシピを典型的・非典型的なレシピに近付けることができるかということを実験により検証した。これらを評価するにあたり、COOKPAD から「カルボナーラ」、「ナポリタン」、「豚汁」、「ミネストローネ」、「トマトのサラダ」、「ツナを使ったサラダ」の 6 つを対象カテゴリとして選択した。それぞれのカテゴリに含まれているレシピ数は 73, 59, 141, 76, 80, 85 である。

5.1 土台となる食材及び典型的なレシピに使われる食材の抽出

あるカテゴリにおいて、提案手法でどの程度正確に土台となる食材を抽出できるかを調べるために、実験を行った。まず、正解セットを作るために、評価の対象とするカテゴリ内のレシピで使われている食材の中で、そのカテゴリ内の 20%以上のレシピで使われている食材を抽出する。ただし、カテゴリ内で 20%以上使われている食材が 10 個未満であれば、使われているレシピの数が上位 10 件の食材を自動的に抽出する。次に、抽出した食材をランダムに並べて被験者に見せ、そのカテゴリ内のどのようなタイプの料理を作るときにも必要であると思う食材を選択してもらう。10 名の被験者に上記のタスクに取り組んでもらい、土台となる食材として 7 名以上の被験者が選択した食材を正解とした。被験者によって、土台となる食材として選択された食材は表 1 の通りである。次に、3.2 節で述べた土台を抽出する手法を用いて、閾値を変化させたときに、被験者らによって作成された正解土台食材を最も高い精度で抽出できる閾値を求めた。精度の評価のため、閾値の値を 0 から 1 ま

表 1 評価実験により土台となる食材として選択された食材

カテゴリ	食材
カルボナーラ	パスタ, 卵, 粉チーズ
ナポリタン	パスタ, ケチャップ
豚汁	味噌, 豚肉
ミネストローネ	トマト, コンソメ
トマトのサラダ	トマト
ツナを使ったサラダ	ツナ

表 2 システムが土台となる食材として抽出した食材

カテゴリ	食材
カルボナーラ	パスタ, 卵, 粉チーズ, ベーコン, 粗挽き黒胡椒
ナポリタン	パスタ, ケチャップ, 玉ねぎ
豚汁	味噌, 豚肉, ニンジン
ミネストローネ	トマト, 玉ねぎ, ニンジン
トマトのサラダ	トマト
ツナを使ったサラダ	ツナ

表 3 被験者が典型的なレシピに必要な食材として選択した食材

カテゴリ	食材
カルボナーラ	パスタ, 卵, 粉チーズ, ベーコン, 粗挽き黒胡椒, 生クリーム
ナポリタン	パスタ, ケチャップ, ビーマン, ウインナー, 玉ねぎ, 粗挽き黒胡椒
豚汁	味噌, 豚肉, ニンジン, こんにゃく, ごぼう
ミネストローネ	トマト, コンソメ, ベーコン, ニンジン, 玉ねぎ
トマトのサラダ	トマト
ツナを使ったサラダ	ツナ, マヨネーズ

表 4 システムが典型的なレシピに必要な食材として抽出した食材

カテゴリ	食材
カルボナーラ	パスタ, 卵, 粉チーズ, ベーコン, 粗挽き黒胡椒, 生クリーム
ナポリタン	パスタ, ケチャップ, ビーマン, ウインナー, 玉ねぎ
豚汁	味噌, 豚肉, ニンジン, 大根, だし
ミネストローネ	トマト, コンソメ, ベーコン, ニンジン, 玉ねぎ
トマトのサラダ	トマト, 玉ねぎ
ツナを使ったサラダ	ツナ, マヨネーズ

で 0.1 ずつ変化させた時の F 値を求め、その値が最大となる閾値を調べた。この結果、閾値の値が 0.6 のときに F 値は最大となった。以上より、以降の評価実験では閾値 β_1 の値として 0.6 を採用する。閾値 β_1 の値を 0.6 にしたときに、システムによって土台となる食材として抽出された食材は表 2 の通りである。

典型的なレシピに使われる食材の評価に関しても同様に、カテゴリの最も典型的なレシピを作るときに必要なであると思う食材を選択してもらい、正解セットを作った。結果を表 3 に示す。また、3.2 節で述べた提案手法の閾値 β_2 の値が 0.3 のときに F 値は最大となった。以上より、以降の評価実験では閾値 β_2 の値を 0.3 とする。このとき、システムによって典型的なレシピに使われる食材として抽出された食材を表 4 に示す。

5.2 レシピ構造の安定度の評価

構造の安定度が高いレシピほど、典型的なレシピであるという仮説を検証するために、次のような実験を行った。まず、あるカテゴリ内のすべてのレシピの安定度を、5.1 節で得られた結果を元にして求め、その値によってレシピを 4 つのグループに分ける。次に、各グループの中からレシピをランダムに 5 つずつ選択する。そして、5 つのレシピからランダムに 1 つずつを選択し、被験者にその 4 つのレシピを見せ、より典型的なレシピだと思う順に順位づけしてもらい、システムが求めた各レシピの安定度の順位との相関を調べる。4 人の被験者には 6 つのカテゴリについて、各カテゴリの 4 つのレシピの組を 2 つずつ評価してもらった。ベースラインとしては、あるレシピで使われている食材の集合を $E = \{e_1, e_2, \dots, e_k\}$ としたとき、各レシピのスコア $f_{tf}(E)$ を次式により求め、その値が高いほど典

型的なレシピであるとして順位付けを行う。ここで、 k はそのレシピで使われている食材の数である。

$$f_{tf}(E) = \frac{1}{|E|} \sum_{e_i \in E} |R(e_i)| \quad (9)$$

被験者が求めた順位と、提案手法を用いて求めた順位及びベースラインを用いて求めた順位との相関関係を見るために、スピアマンの順位相関係数を用いた。結果を表 5 に示す。

表 5 カテゴリ毎の被験者による典型度の順位との相関

手法	カテゴリ	カルボナーラ	ナポリタン	豚汁	ミネストローネ	トマトのサラダ	ツナを使ったサラダ
提案手法		1.00	0.316	0.948	1.00	0.632	1.00
ベースライン		1.00	-0.200	0.800	0.800	0.800	0.800

この結果を見ると、提案手法は「トマトのサラダ」以外のカテゴリではベースラインの手法以上の相関をもつことがわかる。以下、カテゴリ毎に実験結果の考察を行う。

「カルボナーラ」のカテゴリでは、5.1 節の評価実験において被験者が選択した典型的な食材集合と、システムが抽出した典型的な食材集合が完全に一致していたため、被験者の順位との相関が 1 になった。ベースラインの相関係数も 1 になったのは、カルボナーラのレシピは多くのレシピ間で使われている食材が共通しているため、典型的な食材以外の食材は、使用される頻度が低い食材である場合が多いためである。つまり、使われる回数の少ない珍しい食材が使われているレシピがあれば、そのレシピのスコアも低くなり、被験者も非典型的なレシピであるとみなすためだと考えられる。「ナポリタン」のカテゴリでは、提案手法とベースライン共に相関は低かったが、今回の実験では、相関が低くなった原因の特定はできなかった。

「豚汁」「ミネストローネ」「トマトのサラダ」「ツナを使ったサラダ」の 4 つのカテゴリは、1 つのレシピに多くの食材が使われ、使う食材の自由度が高いという共通した特徴があるが、「トマトのサラダ」では提案手法がベースラインを下回っていた。ここで、これら 4 つのカテゴリの違いを考えると、土台となる食材及び典型的なレシピで使われる食材が「豚汁」と「ミネストローネ」に比べて「トマトのサラダ」と「ツナを使ったサラダ」は少ない点が挙げられる。土台となる食材及び典型的なレシピで使われる食材が多い場合、それらの食材を多く使ったレシピほど、提案手法により求められるレシピの安定度が高くなり、また被験者もそれらの食材を多く含むほど典型的であると考えられる。また、レシピに使う食材の自由度は高いが、典型的な食材が多く使われているレシピであれば、典型的な食材以外の食材が使われていても、極端に使用される頻度が低い食材でなければ安定度は高い値となり、被験者も典型度は高いと考える。ベースラインの手法では、使われている食材の数が少なくなるほどスコアが高くなる傾向が見られたが、多くの食材を使うレシピが多いカテゴリでは、使われている食材の数が少なすぎると被験者は非典型的であるとしたため、提案手法を下回ったと考えられる。一方、土台となる食材及び典型的なレシピで使われる食材が少ない場合、それら以外の食材が使われると、そのレシピの安定度は大きく低下する傾向が見られた。特に「トマトのサラダ」や「ツナを使ったサラダ」のカテゴリでは、トマトとツナが使われるレシピ数が他の食材より極端に多いため、その傾向が強かったと考えられる。また、典型的な食材の数が少ないので、その中で使われていない食材があるレシピの安定度も低くなる傾向が見られた。さらに、評価実験でも

土台となる食材及び典型的な食材として選択された食材が少なかったように、被験者によって典型的なレシピで使うと考える食材が大きく異なることも相関が低くなった原因であると考えられる。

5.3 システムが推薦する追加・削除食材の評価

システムが推薦する追加・削除食材の妥当性を検証するために、あるレシピに対する食材の追加・削除がユーザにどの程度受け入れられるかを実験により調べた。ここでは、まず、各カテゴリのレシピをランダムに 20 個ずつ抽出する。被験者には、それぞれのカテゴリについてレシピを 3 つずつ見せ、以下の 4 つの項目について、提案手法とベースライン手法により推薦される食材のうち、各項目により適した食材を選択してもらう。

- (1) 追加することで典型的なレシピに近づく食材
- (2) 追加することで非典型的なレシピに近づく食材
- (3) 削除することで典型的なレシピに近づく食材
- (4) 削除することで非典型的なレシピに近づく食材

提案手法により各項目で推薦される食材は、まずあるレシピに対して追加及び削除可能な食材を 3.4 節の手法により抽出し、追加候補の食材の中で追加後の安定度が最も高くなる食材を上記の (1) の食材として推薦し、追加後の安定度が最も低くなる食材を (2) の食材として推薦する。また、削除候補の食材の中で削除後の安定度が最も高くなる食材を (3) の食材として推薦し、削除後の安定度が最も低くなる食材を (4) の食材として推薦する。各項目にあてはまる食材が複数個ある場合、(1) と (4) については使われているレシピの数が多い食材を推薦し、(2) と (3) については使われているレシピの数が少ない食材を推薦する。使われているレシピの数も同じ食材が複数個あれば、ランダムに 1 つの食材を抽出して推薦する。

ベースラインとなる手法は、まず、評価の対象となるレシピが含まれるカテゴリ内の食材の中で、使われるレシピの数が閾値以上の食材を抽出し、その食材集合から、評価の対象となるレシピで使われている食材を取り除く。そして、残った食材の集合の中で、使われているレシピの数が最も多い食材を上記の (1) の食材として推薦し、使われているレシピの数が最も少ない食材を (2) の食材として推薦する。また、評価の対象となるレシピで使われている食材から、そのレシピが属するカテゴリにおいて土台となる食材を除いた食材集合の中から、使われているレシピの数が最も少ない食材を (3) の食材として推薦し、使われているレシピの数が最も多い食材を (4) の食材として推薦する。追加・削除共に候補が複数個あるときはランダムに 1 つの食材を抽出して推薦する。

提案手法とベースラインで同じ食材が推薦された場合は、2 つの食材が共に該当項目に適していると思えば両方を選択し、共に不適切であると思えば両方を未選択にするようにした。3 人の被験者に対して評価実験を行った結果は表 6 の通りである。

実験全体を通して言えることとして、追加することで典型的なレシピに近づく食材と、削除することで非典型的なレシピに近づく食材は、提案手法とベースラインの手法で同じ食材が推薦されるレシピが多かった点が挙げられる。これは、追加の場合は追加の候補となる食材集合の中で、使われているレシピの数が極端に多い食材がある場合は、その食材を加えることで安定度が最も高くなることが多く、削除の場合は、削除の候補となる食材集合の中で、使われているレシピの数が最も多い食材を削除すると、安定度が最も低くなるが多いためである。このため、典型的なレシピに近づく食材として推薦される食材

表 6 ユーザが選択したカテゴリ毎の食材の割合

		追加 典型	追加 非典型	削除 典型	削除 非典型
カルボナーラ	提案手法	100.0%	55.6%	66.7%	0.0%
	ベースライン	100.0%	33.3%	44.4%	0.0%
	未選択	0.0%	11.1%	11.1%	100.0%
ナポリタン	提案手法	100.0%	44.4%	55.6%	44.4%
	ベースライン	88.9%	44.4%	55.6%	55.6%
	未選択	0.0%	11.1%	22.2%	33.3%
豚汁	提案手法	88.9%	55.6%	77.8%	44.4%
	ベースライン	77.8%	44.4%	55.6%	55.6%
	未選択	0.0%	0.0%	11.1%	44.4%
ミネストローネ	提案手法	88.9%	77.8%	77.8%	22.2%
	ベースライン	88.9%	22.2%	22.2%	0.0%
	未選択	0.0%	0.0%	0.0%	77.8%
トマトのサラダ	提案手法	66.7%	44.4%	44.4%	22.2%
	ベースライン	55.6%	55.6%	55.6%	33.3%
	未選択	0.0%	0.0%	22.2%	66.7%
ツナを使ったサラダ	提案手法	55.6%	33.3%	55.6%	22.2%
	ベースライン	77.8%	66.7%	66.7%	22.2%
	未選択	0.0%	0.0%	33.3%	66.7%

は、少なくとも片方の食材は選択されたが、削除することで非典型的なレシピに近づく食材として推薦される食材は、その食材を削除すると料理として成立しなくなる食材が多いので、未選択となることが多く見られた。また、一般に食材を削除することで非典型的なレシピに近付けたいという欲求をユーザが持つことが少ないことも、未選択となる場合が多くなった理由として考えられる。「トマトのサラダ」と「ツナを使ったサラダ」のカテゴリでは、追加することで典型的なレシピに近づく食材として 2 つの手法で異なる食材が推薦されることが多かったが、使用する食材の自由度が高く、追加することで典型的なレシピに近づくと考えられる食材が被験者間で大きく異なるため、提案手法とベースラインの手法の間に殆ど差は見られなかった。

追加すると非典型的なレシピに近づく食材と、削除することで典型的なレシピに近づく食材で 2 つの手法の間で大きな差が出たカテゴリとしては、「ミネストローネ」が挙げられる。5.2 節の考察で述べたように、類似した性質をもつ「豚汁」とは結果に大きな差が出た。その理由としては、一般に豚汁は味噌とだし以外は肉類や野菜などの具材を入れることで作られるため、推薦される食材が多いものの、豚汁の場合はおでんとは異なり、「京風」や「洋風」といった分類も殆どないので、具材の組み合わせに関しては、相性の良し悪しは大きな問題とならないこと、また、豚汁はサラダと同様にどちらの食材を選択するかは被験者に左右されるというのも一因である。これに対してミネストローネは、肉類や野菜類などの具材も多く使われているが、砂糖や醤油、ケチャップなどの調味料も多く使われている。このため、推薦される食材としては、豚汁のときよりも調味料が多く見られた。調味料であれば、組み合わせによっては料理のバランスが大きく崩れることもあるので、提案手法で推薦される食材の方が選ばれることが多かったのではないかと考えられる。

6. 一般の集合型 Web オブジェクトへの適用

本章では、5 章の評価実験で得られた知見を元に、一般の集合型 Web オブジェクトに対して提案手法を適用する場合、どのようなタイプの集合型 Web オブジェクトであれば本手法が有効であるか、あるいは有効でないかについて述べる。

本手法が有効に働く集合型 Web オブジェクトとして、次のような例が挙げられる。あるユーザが池田屋事件について記述

されたページを読んでいて、そのページに出てくる人物にあと1人追加したようなストーリーが記述されたページに辿りつきたいと考えている。「新撰組の物語」というカテゴリには、集合型 Web オブジェクトとして様々な側面から語られる新撰組の物語があり、物語によって構成要素である登場人物は異なる。しかし、どのような側面から池田屋事件を語る場合でも、近藤勇や沖田総司、土方歳三は欠かせない人物であり、また同時に語られることが多い人物や少ない人物の組み合わせがあると考えられる。よって、ユーザの意図に応じて、ユーザが今見ているページの登場人物に1人追加することで、今読んでいる話よりも有名な話を読んだり、裏話を読んだりすることが可能になる。

一方、土台となる要素が少なすぎるカテゴリでは、本手法は有効ではない。このため、例えばあるユーザが邪馬台国について記述されたページを読んでいて、そのページに出てくる人物にあと一人追加して、より変わったストーリーを読みたいという欲求を持ったとする。しかし、「邪馬台国の物語」というカテゴリ内のどのような種類の物語でも、それらを語る上で出てくる登場人物として、卑弥呼は絶対に欠かせない人物であり、かつ卑弥呼以外の人物で欠かせない人物はあまり存在しない。よって、ごく一部の物語にしか登場しない人物は複数人存在するが、そのいずれの人物が非典型的であると感じるかは、人によって異なるため、本手法でユーザの欲求に沿う最適な人物を推薦することは困難である。

また、集合型 Web オブジェクトの構成要素間で相性の良い悪しが存在しない場合も本手法は有効ではない。例えば、あるユーザが防災用品を揃えるために防災用品のリストが掲載されているページを見ているときに、全ての防災用品を一度に揃えるには予算が足りないため、このリストの中から1つ削除するとしたらどれが良いのかを知りたいという欲求を持ったとする。しかし、防災用品の中で、ある2つを一緒に揃えると極端に相性が悪くなるというものは殆どないと考えられる。このため、本手法によって削除の候補として推薦される防災用品は、多くの人にとって根拠のないものと感じられてしまう。

7. 関連研究

情報推薦に関する研究は多数行われている。小原ら[2]はユーザが記述した Blog 記事を収集し解析することで、ユーザの嗜好に適したニュース記事を推薦する手法を提案している。また、丹羽ら[3]は Folksonomy を利用した Web ページ推薦システムの構築手法を提案し、大村ら[4]は逐次変動するユーザの嗜好に即時対応できる Web ページ推薦手法を提案している。これらはいずれも、ニュース記事、Web ページを要素と考えると、要素の追加を目的としていると言える。しかし、いずれの研究もユーザの嗜好に最も適した要素を推薦することを目的としている点が本研究とは異なる。

志土地ら[5]は調理レシピテキストから代替食材を発見する手法を提案している。この手法では各レシピをその料理名に応じていずれかの料理群に分類し、さらに料理群ごとに素材と調理動作の関連付けを行い、素材間の類似性を考えることで新たな代替素材を得ている。志土地らはユーザの意図に関わらず、各レシピにおいて代替可能な素材の発見を行っており、また素材の追加や削除に関しては考慮していない。それに対して我々は追加または削除可能な食材を複数個提示し、レシピの構造を考慮することによりユーザの意図を反映した追加・削除を可能にしている。

また、大平ら[6]はユーザの意見、特にユーザの好みを反映した献立を作成する手法を提案している。この研究では、ユーザが料理名や調理法の希望を入力し、得られた献立をユーザの材料の保有状況を満たすように適用する。さらに、事例ベース推論を利用することで、過去のユーザの意見を反映することも可能にしている。しかし、彼らの手法では「今のレシピに物足りなさを感じるのでは何かあと1つ食材を追加したい」「今のレシピを風変りなレシピにしたい」といった、ユーザの抽象的な意図を反映することはできない。

8. まとめ

本論文では要素の組み合わせからなる集合型 Web オブジェクトの典型度について、その構造としての安定度を利用することで、要素の追加及び削除可能な要素を推薦する手法の提案を行った。そして、今回は集合型 Web オブジェクトとしてレシピに着目し、レシピに対する食材の追加及び削除を実現した。

実験では、カテゴリ内の集合型 Web オブジェクトを構成する要素として土台となる要素が一定数存在し、要素間に相性が存在するカテゴリにおいて提案手法が有効に働くことがわかった。

今回はカテゴリとカテゴリに属する集合型 Web オブジェクト、及び各集合型 Web オブジェクトの構成要素が Web 上から容易に収集可能なレシピを対象とした。しかし、今後は一般の集合型 Web オブジェクトを対象として提案手法を適用していきたいと考えている。そのためには、まずある Web ページの中でユーザが着目している要素を抽出し、そのページで扱っている話題がどのカテゴリに属するのかが発見できなければならない。そして、カテゴリが発見できた後は、そのカテゴリに属する集合型 Web オブジェクトとそれぞれの構成要素を抽出する必要がある。さらに、追加及び削除可能な要素の発見が困難であるカテゴリタイプに属する集合型 Web オブジェクトに対しても、追加及び削除可能な要素を発見できるように手法の拡張を考えていく予定である。

謝 辞

本研究の一部は、グローバル COE 拠点形成プログラム“知識循環社会のための情報学教育研究拠点”，計画研究“情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究”（研究代表者：田中克己，A01-00-02，課題番号 18049041），によるものです。ここに記して謝意を表すものとします。

文 献

- [1] COOKPAD.
<http://cookpad.com>
- [2] 小原 恭介, 山田 剛一, 絹川 博之, 中川 裕志: “Blogger の嗜好を利用した協調フィルタリングによる Web 情報推薦システム”, 2005 年, 人工知能学会
- [3] 丹羽智史, 土肥拓生, 本位田真一: “Folksonomy マイニングに基づく Web ページ推薦システム”, 報処理学会論文誌 Vol.47 No.5, 報処理学会, 2006.
- [4] 大村慶祐, 大向一輝, 武田英明: “ソーシャルブックマークのおけるインベータに注目した情報推薦手法の提案”, 人工知能学会全国大会 (第 22 回) 論文集, No. 2(D3-03), 2008 年 6 月
- [5] 志土地由香, 高橋友和, 井手一郎, 村瀬洋: “調理レシピテキストからの代替素材の発見”, 人工知能学会第 22 回全国大会論文集, 1B1-2, pp.1-4
- [6] 大平峰子, 大園忠親, 新谷虎松: “ユーザ意見を反映した事例による献立作成エージェントについて”, 第 57 回情報処理学会全国大会論文集 vol.2 pp.337-338 Oct. 1998