

調理レパートリー拡大のための戦略的レシピ推薦法の提案

中岡 義貴[†] 佐藤 哲司^{††}[†] 筑波大学 情報学群 知識情報・図書館学類 〒305-8550 茨城県つくば市春日 1-2^{††} 筑波大学 図書館情報メディア研究科 〒305-8550 茨城県つくば市春日 1-2E-mail: [†]{nakaoka,satoh}@ce.slis.tsukuba.ac.jp

あらまし インターネット上のレシピサイトでは、膨大な量のレシピを様々な条件から検索することができる。しかし、複数の条件を組み合わせて行う検索には限界があり、検索結果に調理できないレシピが含まれていることも多かった。本論文では、調理者に合わせたレシピ選択を支援することを目的に、調理できるレシピのレパートリーを順次拡大するためのレシピ推薦法を提案する。レシピに記載される食材とレシピの関係を明らかにし、食材の使い勝手やレシピの馴染み深さの指標である定番度を算出する。提案手法では、ランキング手法である HITS アルゴリズムを重み付きに改良し、レシピの順位付けに特化した手法を実現した。公開されているレシピサイトのデータを使用して提案法を評価した結果、定番度に基づくレシピ推薦の有効性を確認できたので報告する。

キーワード レシピ推薦, HITS アルゴリズム, ネットワーク分析

Strategic recipe recommendation method for Cooking repertory expansion

Yoshiki NAKAOKA[†] and Tetsuji SATOH^{††}[†] College of Knowledge and Library Sciences, School of Informatics 1-2 Kasuga, Tsukuba City, Ibaraki, 305-8550, Japan^{††} Graduate school of Library Information and Media Studies, University of Tsukuba 1-2 Kasuga, Tsukuba City, Ibaraki, 305-8550, JapanE-mail: [†]{nakaoka,satoh}@ce.slis.tsukuba.ac.jp

Abstract In this paper, the recipe recommending method for expanding the repertory of the recipe which can be cooked for the purpose of supporting the recipe selection united with the cooking person is proposed. By the proposing method, improving with dignity the HITS which is the ranking technique realized the technique of having specialized in ranking attachment of a recipe. Since the validity of recipe recommendation was validate, it reports.

Key words

1. はじめに

近年、インターネット上で料理レシピを公開するレシピサイトの普及により、多くのレシピを簡単に探すことができる。レシピサイトはその提供元によって、食品会社が提供するサイトや、料理番組が運営するサイト、ユーザがオリジナルレシピを投稿するレシピコミュニティサイト、個人のブログなどに分類できる。食品会社である AJINOMOTO が提供する「レシピ大百科^(注1)」には、2012 年 12 月現在で 9,500 件以上のレシピが掲載されている。NHK の料理番組であるきょうの料理が運

営する「みんなのきょうの料理^(注2)」には、2012 年 12 月現在で 14,000 件以上のレシピが掲載され、放送のたびにレシピが登録されるだけでなく、ユーザによつての投稿もされている。また、ユーザがオリジナルレシピを投稿できるレシピコミュニティサイト COOKPAD^(注3)には、日々膨大な数のレシピが登録され、2012 年 12 月現在で 135 万件以上のレシピが掲載されている。これらのサイトでは、レシピ名や、食材と食材の使用量、和食や中華などのジャンル、煮るや焼くなどの調理手順など、条件を指定してレシピを検索することができる。しかし、

(注1) : レシピ大百科,AJINOMOTO CO.,INC,
<http://park.ajinomoto.co.jp/>

(注2) : みんなのきょうの料理,NHK エデュケーショナル,
<http://www.kyounoryouri.jp/>

(注3) : COOKPAD, COOKPAD Inc.,<http://cookpad.com/>

様々な条件を組み合わせてレシピを検索することは、調理者の負担となっている。例えば、過去に調理したことのあるレシピでも、忙しくて調理をする時間がないという場合や、調理したいレシピがあったとしても、調理工程の難易度が高い、食材が手に入らない、などの場合が考えられる。

この問題の解決として、調理経験の少ないレパートリーの少ない初級者は、限られた調理経験に基づいて、少しだけ難しいレシピを推薦すれば調理可能と考えられる。一方で、調理経験が豊富で調理できるレシピの多い上級者には、調理したことのない食材を使用するレシピを推薦することで、調理者の意欲を喚起することが期待できる。このように、レパートリーの拡大には、調理者に合わせた推薦を行う必要がある。

本研究では、調理者に合わせたレシピ推薦を行うために、レシピに記載される食材とレシピの関係を明らかにすることで、食材の使い勝手やレシピの馴染み深さの指標である**定番度**を算出する。定番度の高い食材は多くのレシピで使用される使い勝手の良い食材であり、定番度の低い食材はあまり使用されない珍しい食材である。定番度に基づいて順位付けを行い、推薦を行うためための評価を行う。提案手法では、Web ページのランキング手法である HITS アルゴリズムを重み付きに改良し、重みとして食材の使用量を用いることで、レシピの順位付けに特化した手法を実現した。

2. 関連研究

本研究は、レシピと食材によるリンク構造に基づいて、レシピに順位付けを行い、調理者に合わせたレシピを推薦する。調理者に合わせてレシピ選択の支援を行う関連研究として、調理の難易度を考慮した研究や [1] [2]、調理者の嗜好を考慮した研究 [3] などがある。大山ら [1] は、調理者の調理に関する知識や技術を考慮した、レシピ推薦システムを開発している。調理に関する知識と技術を「調理スキル」、レシピを構成する各作業を「調理作業」と定義し、調理スキルに対する調理作業の難易度を決定している。調理スキルと調理作業の関係を明らかにすることで、調理作業の難易度に基づくレシピ選択を実現している。また、岩本ら [2] は、レシピの難易度を決定するために、調理工程に相当する調理動作を抽出し、調理動作の難易度を決定している。調理動作数と調理動作の難易度を考慮し、検索結果候補を順位付けすることで、難易度によるレシピの検索を実現している。

一方、上田ら [3] は、調理者の嗜好を考慮したレシピ推薦を行うために、食材の使用履歴を用いて食材の得点付けを行なっている。得点付けの手法として、*TF-IDF* を応用した *FF-IRF* を提案している。調理経過日時を考慮した食材の出現頻度である *FF* (食材利用頻度: *Foodstuff Frequency*) と、食材の特異性を示す *IRF* (食材の特異度: *Inverted Recipe Frequency*) から得点を算出する。正規化した食材の得点を加算することでレシピの得点を算出し、得点の高いレシピから推薦することで、調理者の嗜好に合わせたレシピの推薦を実現している。

八木ら [4] は、POS データから得られる相関ルールから 2 部グラフを作成し、HITS アルゴリズムによって各ルールの重要

度を算出している。強い相関を持つルール同士をリンクでつなぎ、相関の強さをリンクの重みとして表現することで、重み付きの隣接行列を作成している。そして、隣接行列に HITS アルゴリズムを適用することで、各ルールのスコアを算出し、相関ルールの原因を hub 値、結果を authority 値として、重要度の高い相関ルールを抽出している。

山口ら [5] は、サーチエンジンのクエリログから作成した 2 部グラフに、HITS アルゴリズムを適用することで、同位語と話題語の発見を実現している。HITS アルゴリズムで算出される authority 値の上位の語を同意語とし、hub 値の上位の語を話題語としている。サーチエンジンから同時に検索される語の関係から、それぞれの語をノードとする 2 部グラフを作成している。作成した 2 部グラフの特徴は、HITS アルゴリズムを適用するに相応しいリンク構造であると述べている。

レシピ選択に関する研究の多くは、調理者とレシピとの関係を用いてレシピ選択の支援を行なっている。本研究は、レシピと食材によるリンク構造を用いて、調理者のレシピ選択を支援する。調理者に合わせたレシピ推薦のために、レシピと食材の関係からネットワーク分析を行い、レシピの順位付けを行う点で大きく異なる。

3. 提案手法

本章では、レシピと食材の順位付けを行うための、定番度算出手法について説明する。料理のレシピは通常複数の要素で構成され、料理の名前や調理時間、食材と食材の使用量、調理道具、調理の手順、などが挙げられる。本研究ではこれらの要素の中から、レシピで使用される食材に着目する。レシピと食材の関係を明らかにすることで、それぞれの定番度を算出し、レシピの定番度に基づいて順位付けを行う。食材は、食材固有の使い勝手の良さを持っていると考え、より多くのレシピで使用される食材は、使い勝手が良く、幅広い調理に適した食材であると仮定する。同様に、幅広い調理に適した食材で構成されているレシピは、多くの人に指示される馴染み深いレシピであると仮定する。この 2 つの仮定から、食材の使い勝手やレシピの馴染み深さである定番度を、レシピと食材の関係をを用いて算出する。

以下、3.1 節で、レシピと食材の関係について説明する。次に 3.2 節で、リンク構造に基づく定番度の算出手法について述べ、3.3 節で、重み付きの定番度算出手法について説明する。

3.1 レシピと食材の関係

本論文では、仮定を検証するために、いくつかのレシピとレシピで使用されている食材について様々なレシピサイト ここでは COOKPAD^(注4)、Yahoo! レシピ^(注5)、レシピ大百科^(注6)、みんなのきょうの料理^(注7)、ホームクッキング^(注8) の 5 種類のサイトについて調査を行った。使用されているレシピ数を調査し

(注4) : COOKPAD, COOKPAD Inc., <http://cookpad.com/>

(注5) : Yahoo! レシピ, Yahoo! JAPAN, <http://recipe.gourmet.yahoo.co.jp/>

(注6) : レシピ大百科, AJINOMOTO CO., INC, <http://park.ajinomoto.co.jp/>

(注7) : みんなのきょうの料理, NHK エデュケーショナル, <http://www.kyounoryouri.jp/>

(注8) : ホームクッキング, キッコーマン, <http://www.kikkoman.co.jp/homecook/>

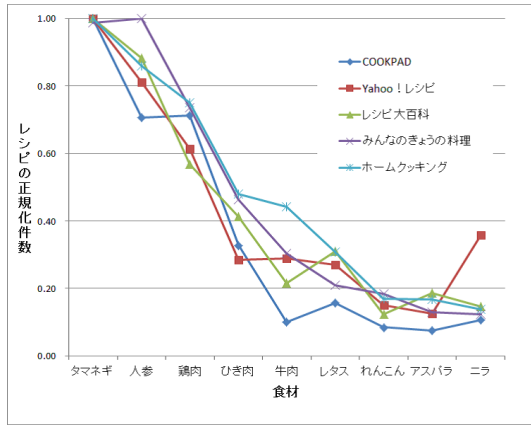


図 1 レシピサイトごとの食材による検索結果

た結果を図 1 に示す。

図の横軸は、タマネギからニラまで 9 種類の食材を取り、それぞれの食材が使用されているレシピ数^(注9)を示したものである。図の縦軸は、各サイトで検索されたレシピ数の最大値で正規化した値である。図 1 の結果から明らかなように、いずれのサイトにおいても、いくつかの例外を除いて、右肩下がりグラフとなっていることがわかる。ホームクッキングにおいては、件数は単調減少となっている。調査した 9 種類の食材では、タマネギ、人参の順で多くのレシピで使用されており、これらの食材は定番度が高いと考えられる。さらに、より詳細な食材の相関を見るために、料理番組が運営するみんなのきょうの料理、投稿型のコミュニティサイトである COOKPAD、企業が提供するレシピ大百科のように、形態や掲載レシピ数の異なるレシピサイトにおいて、食材数を増やし調査を行った。ここでは、みんなのきょうの料理に掲載されている 100 種類の食材から、野菜や、魚などの曖昧なものを除いた 93 種類の食材を使用しているレシピ数を調査した。各レシピサイトの順位相関を検証するために、スピアマンの順位相関係数を算出した。各サイト間の順位相関係数を表 1 に示す。いずれの相関係数も 0.7 以上と高い相関を示し、レシピ大百科は最も高い値となっている。図はレシピ大百科を右肩下がりとなるように食材を配置しているが、他のサイトも同様に右肩下がりとなっている。例外として COOKPAD では卵とさけ(鮭)の件数が多く、みんなのきょうの料理ではふきの件数が多いことがわかる。卵とさけについては、非常に安価で購入可能であり、簡単な調理でレシピが完成するため、調理者が投稿を行う COOKPAD の件数が多くなったと考えられる。みんなのきょうの料理のふきは、NHK の料理番組がレシピサイトの運営を行っており、利用者層の偏りによるものなどの理由が考えられる。

以上の結果では、食材の順位は極一部の例外を除いて、レシピサイトに依存していないことを示している。この性質は、より多くの食材、レシピを対象にネットワーク分析を行うことで、食材そのものが持つ使い勝手の良さを評価できるとことを指し

(注9)：いずれのサイトにおいても検索語「タマネギ」と「たまねぎ」では、検索されるレシピ数が一致することから、各サイトでは表記の揺れを吸収していると考えられる。

表 1 レシピサイトごとの順位相関係数

	レシピ大百科	きょうの料理	COOKPAD	(平均)
レシピ大百科	-	0.817	0.821	0.819
きょうの料理	(0.817)	-	0.729	0.773
COOKPAD	(0.821)	(0.729)	-	0.775

ている。同様に、レシピそのものについても、固有の馴染み深さである、定番度を求めることができると考えられる。

3.2 リンク構造に基づく定番度算出手法

レシピと食材は、レシピから使用する食材へリンクを貼れば、それぞれをノードとする 2 部グラフであるといえる。この 2 部グラフから 2 値の隣接行列を作成し、HITS アルゴリズムを適用することで、レシピと食材の定番度を算出し、レシピ定番度に基づいて順位付けを行う。本節では、HITS アルゴリズムの概略と、HITS アルゴリズムを用いた定番度算出手法について述べる。

3.2.1 HITS アルゴリズム

HITS アルゴリズムは Web ページのリンク構造を用いて、ページの重要度といえるスコアを算出する手法であり、Kleinberg [6] によって提案された手法である。Web ページ間のリンクの様子を図 2 に示す。HITS アルゴリズムでは、被リンクの指標である authority 値の高い Web ページを authority、発リンクの指標である hub 値の高い Web ページを hub として定義する。ここで、良い authority は、多くの被リンクを持つ authority 値の高いページであり、良い hub は、多くの発リンクを持つ hub 値の高いページである。Web ページ p から Web ページ q へリンクされているとき、ページ p の authority 値と hub 値のスコアは以下のように定義されている。

$$\begin{aligned} auth(p) &= \sum_{q, p \rightarrow q} hub(q) \\ hub(p) &= \sum_{q, p \rightarrow q} auth(q) \end{aligned} \quad (1)$$

実際の計算は、リンク構造を隣接行列 L として表現し、 L の転置行列である L^T を用いて再帰的に求めることができる。図 2 の Web ページ間のリンクを、リンク構造として表現したものを図 3 に示す。このリンク構造は、以下のような隣接行列で

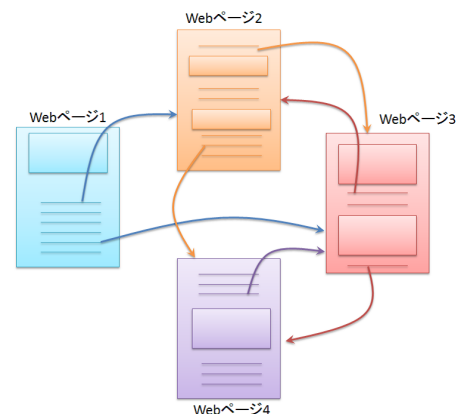


図 2 4 つの Web ページ間のリンク

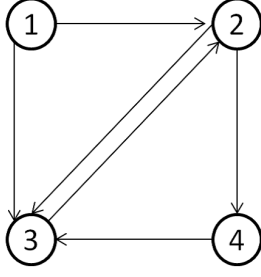


図 3 4つの Web ページによるリンク構造

表現することができ、式 3 を繰り返し計算することで、各ページの authority 値と hub 値を求めることができる。

$$L = \begin{matrix} & P_1 & P_2 & P_3 & P_4 \\ \begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{matrix} & \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix} \quad (2)$$

$$\begin{aligned} auth^k &= L^T L x^{k-1} \\ hub^k &= L x^k \end{aligned} \quad (3)$$

式 2 の隣接行列 L に HITS アルゴリズムを適用し、算出した authority 値と hub 値のスコアとスコアに基づいて Web ページを順位付けした結果を以下に示す。

$$\begin{aligned} auth^T &= \begin{pmatrix} 0 & 0.283 & 0.532 & 0.185 \end{pmatrix} \\ hub^T &= \begin{pmatrix} 0.347 & 0.305 & 0.121 & 0.227 \end{pmatrix} \end{aligned}$$

$$authority \text{ の順位} = P_3 > P_2 > P_4 > P_1$$

$$hub \text{ の順位} = P_1 > P_2 > P_4 > P_3$$

この結果から、図 3 のリンク構造では、ページ 3 が最も良い authority であり、ページ 1 が最も良い hub であることがわかる。

3.2.2 定番度算出手法

これまでに説明した HITS アルゴリズムを用いて計算を行う対象は、Web ページのような同じ要素の集合で構成されるリンク構造である。本研究では、レシピと食材のような異なる要素で構成されるリンク構造である、2 部グラフに対して HITS アルゴリズムを適用する。レシピで使用される食材とレシピをノードとし、食材とその食材を使用しているレシピとの間にリンクを張ることで、レシピと食材は 2 部グラフとなる。このとき、authorities を食材、hubs をレシピとしてそれぞれのスコアを算出する。スコアの特徴として、スコアの低いレシピはスコアの低い食材を多く使用し、スコアの低い食材はスコアの低い多くのレシピで使用される。この 2 部グラフの例を図 4 に示す。2 部グラフにおいて食材はレシピとの間にもみリンクを持

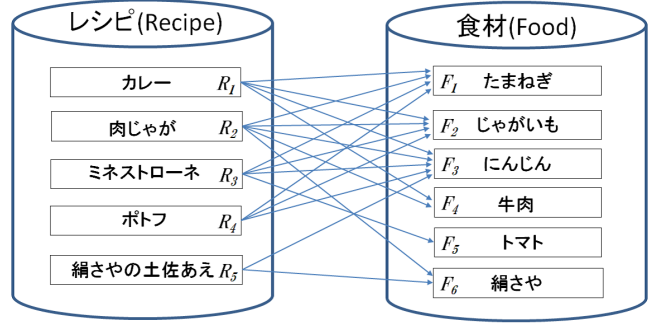


図 4 レシピと食材の 2 部グラフ

ち、レシピは食材との間にもみリンクを持つ。図 4 から作成される隣接行列 L を以下に示す。

$$L = \begin{matrix} & F_1 & F_2 & F_3 & F_4 & F_5 & F_6 \\ \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{matrix} & \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \end{matrix} \quad (4)$$

次に、式 4 の隣接行列 L に対して HITS アルゴリズムを適用し、authority 値と hub 値を算出する。それぞれのスコアを以下に示す。

$$\begin{aligned} auth^T &= \begin{pmatrix} 0.234 & 0.234 & 0.259 & 0.127 & 0.055 & 0.092 \end{pmatrix} \\ hub^T &= \begin{pmatrix} 0.233 & 0.259 & 0.214 & 0.199 & 0.096 \end{pmatrix} \end{aligned}$$

算出したスコアをレシピと食材の定番度とし、定番度に基づいて順位付けを行う。順位付けの結果を以下に示す。

$$\begin{aligned} authority \text{ の順位} &= F_3 > F_1 > F_2 > F_4 > F_6 > F_5 \\ hub \text{ の順位} &= R_2 > R_1 > R_3 > R_4 > R_5 \end{aligned}$$

図 3 の 2 部グラフによる計算では、authority 値が最も良いのは F_3 であるにんじんとなり、hub 値が最も良いのは R_2 である肉じゃがという結果が得られた。この結果から、スコアの一番高いにんじんは使い勝手の良い食材で、スコアの一番高い肉じゃがは使い勝手の良い食材を多く使用している。一方、トマトはミネストローネのみで使用される食材であり、そのスコアは小さくなっている。絹さやの土佐あえはスコアの低い珍しい食材で構成されたレシピであると考えられる。しかし、3 章の始めに述べたように、レシピには食材とともに食材の使用量が記載されている。例として、スコアの近いミネストローネとポトフのレシピに記載されている使用量を表 2 に示す。

ミネストローネでは、たまねぎやにんじんなどの定番度の高

表 2 食材の使用量 (一人前)

	ミネストローネでの使用量	ポトフでの使用量
たまねぎ	25g	100g
じゃがいも	40g	100g
にんじん	10g	50g
トマト	150g	-

い食材を使用しているが、ポトフと比較すると約 4 分の 1 の使用量となっているうえに、スコアの低い食材であるトマトを多く使用している。HITS アルゴリズムをそのまま適用すると、レシピで使用している食材を同等に扱った計算となるため、式 5 のようにポトフよりミネストローネの順位が高くなってしまふ。ミネストローネとポトフの例のように、レシピで使用されている食材はレシピに対して、それぞれが持つ影響の違いが存在すると考えられる。本研究では、この問題を解決するために、レシピの定番度の算出する際に使用する食材の使用量を考慮する。これは、多くのレシピで使用する食材の定番度を高くするだけでなく、多くのレシピにより影響を与える食材の定番度を高くするためである。使用量を重みとすることで、レシピで使用する食材の影響を考慮する。

3.3 重み付き定番度算出手法

本節では、より使用するレシピに対して影響を与える食材の定番度を高くするための、使用量を考慮したレシピの定番度の算出手法について説明する。

3.3.1 使用量に基づく定番度

本研究では、レシピと食材に張られているリンクに重みを付けることで、食材の使用量を考慮した計算を実現する。リンクの重みには、レシピにおける食材の総使用量の平均を用いる。ミネストローネとポトフで使用する食材の使用量を図 5 に示し、リンクに重みを付けた 2 部グラフの例を図 6 に示す。

このような重みの付いた 2 部グラフを行列で表現する際に、行列の各要素の値にリンクの重みを用いる。図 6 の 2 部グラフは、以下のように重みを要素とする行列で表現することができる。

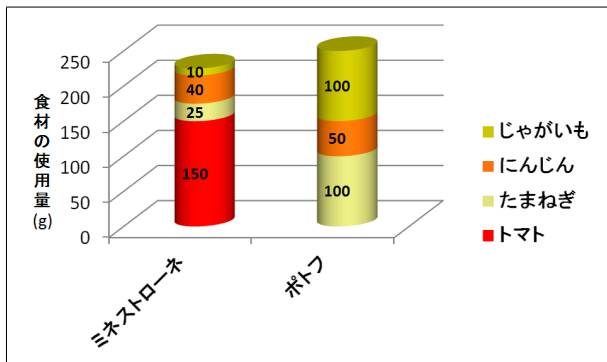


図 5 ミネストローネとポトフの食材使用量

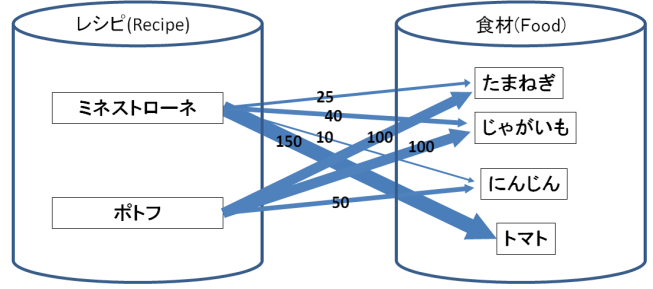


図 6 使用量を重みに持つ 2 部グラフ

$$L = \begin{matrix} & F_1 & F_2 & F_3 & F_4 & F_5 & F_6 \\ \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{matrix} & \begin{pmatrix} 70 & 80 & 50 & 80 & 0 & 0 \\ 50 & 110 & 40 & 40 & 0 & 0 \\ 25 & 40 & 10 & 0 & 150 & 0 \\ 100 & 100 & 50 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 20 \end{pmatrix} \end{matrix} \quad (5)$$

実際の計算では、全レシピの総重量の平均で正規化した値を用いる。式 5 の隣接行列 L を用いて、算出したそれぞれのスコアを以下に示し、スコアに基づいた順位を式 6 に示す。総重量の平均は 205g である。

$$auth^T = \begin{matrix} & F_1 & F_2 & F_3 & F_4 & F_5 & F_6 \\ \begin{pmatrix} 0.264 & 0.349 & 0.164 & 0.131 & 0.091 & 0.001 \end{pmatrix} \end{matrix}$$

$$hub^T = \begin{matrix} & R_1 & R_2 & R_3 & R_4 & R_5 \\ \begin{pmatrix} 0.276 & 0.269 & 0.152 & 0.295 & 0.007 \end{pmatrix} \end{matrix}$$

$$authority \text{ の順位} = F_2 > F_1 > F_3 > F_4 > F_5 > F_6$$

$$hub \text{ の順位} = R_4 > R_1 > R_2 > R_3 > R_5$$

使用量を重みとして用いることで、レシピで使用する食材の重要度を区別し、繰り返しの計算の際に使用量に応じてスコアを与えることができる。

4. 評価実験と考察

3 章で提案した定番度算出手法を実装し、評価実験を行うことで提案手法の有効性を確認する。以下、4.1 節では、実験用データと食材の抽出について説明する。4.2 節は、定番度算出手法の評価のために本実験における適切な推薦順について述べ、HITS アルゴリズムを適用して実験した定番度の結果および考察を示す。4.3 節は、重み付きの定番度算出手法を用いて実験した結果および考察を示す。そして 4.4 節では、定番度に基づくレシピの順位を用いたレシピ推薦の評価および考察を示す。

4.1 レシピの収集

4.1.1 実験用データ

本実験は、3 章で述べた算出した順位相関の高いレシピサイトである、レシピ大百科のレシピを対象とする。レシピ大百科に記載されているレシピの HTML ファイルを取得し、HTML ソースの解析を行うことで、レシピ名とレシピで使用する食材を抽出した。食材にはレシピサイト特有の表記の揺れがあつ

たため、同一の食材であると考えられた場合に統一を行った。食材の抽出から表記の統一については4.1.2節で詳述する。抽出したデータを用いて隣接行列を作成し、レシピと食材の定番度を算出する。

4.1.2 食材の抽出

レシピ大百科に記載されている食材には、「玉ねぎ小」や「玉ねぎ・大」、「玉ねぎ薄切り」、「玉ねぎのみじん切り」など、同じ食材と考えられる場合でも様々な表記で記述されている。このように、食材の使用量や調理方法などを記載している場合は、使用している食材名以外を削除した。これは、食材自体に使い勝手の良さがあるという考えに基づいている。さらに、食品会社が提供しているサイトであることから、AJINOMOTOの商品である食材を使用しているレシピも多い。今回はこのような食材は実験の対象外とする。レシピデータからの食材抽出には、正規表現を用いることで食材の統一を行った。重み付きの手法で使用する使用量は、食材と同様に抽出を行った。使用量の(g)への統一については、4.3節で詳述する。

4.2 定番度算出手法の評価

調理者に合わせたレシピの推薦を行い、レパートリーを拡大することで、調理者のレシピ選択を支援することができる。本研究の課題としては、レシピと食材の隣接行列を用いた計算による順位付けが挙げられる。実際にレシピと食材の定番度を算出し、順位付けを行った結果からレシピの推薦法を考察することで、順位付けを行うための手法の評価を行う。

4.2.1 実験方法

実際に提案手法を用いて定番度を算出し、順位付けを行うことで推薦手法として適切であるか評価を行う。HTMLファイルから抽出したデータとして、レシピ数は9,519件、食材数は1,887件である。このデータを用いて計算を行うための隣接行列を作成する。レシピをhub、食材をauthorityと考え、9,519行×1,887列の隣接行列を作成する。3章で提案した手法を用いて、定番度の算出を行った。

4.2.2 実験結果

実験によって算出したレシピと食材の定番度をそれぞれ図7と図8に示す。縦軸は算出した定番度を対数目盛で表し、横

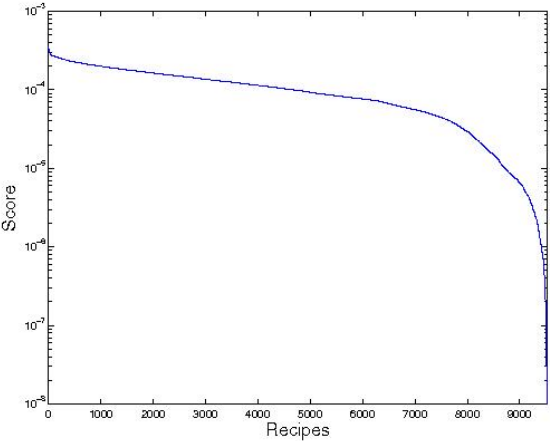


図7 レシピの定番度

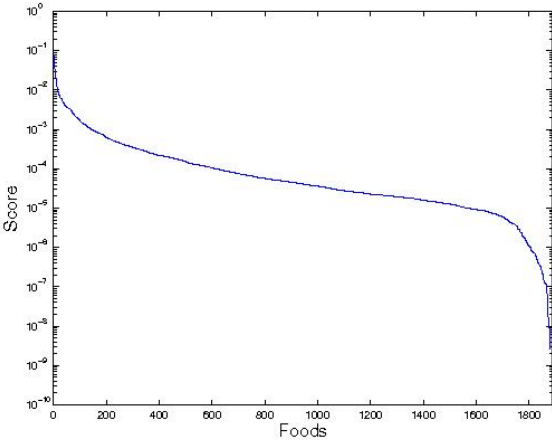


図8 食材の定番度

軸にはレシピと食材それぞれを定番度の高い順に並べた。レシピと食材の定番度に基づき、順位付けを行った結果を表3と表4に示す。詳細な結果として、定番度の上位であるレシピにおいて使用されている食材を、食材の定番度の高いものであるか確かめるために、スコアの高い上位3レシピで使用される食材を表5に示す。

表3 レシピの順位

順位	レシピ
1	鶏のちゃんこ鍋
2	まるまるミートのてるてるボール
3	スプーンで食べる和風煮込みハンバーグ
4	焼めしのカレースープかけ
5	彩色牛肉絲（色々野菜と牛肉の香り炒め）
6	ミートボールの甘酢あんかけ
7	かつおのピリ辛中華あんかけ
8	中華風肉味噌かた焼きそば
9	おからのクロック
10	酢豚

表4 食材の順位

順位	食材	使用件数
1	水	4,428
2	しょうゆ	3,050
3	酒	2,107
4	こしょう	2,669
5	砂糖	1,710
6	にんじん	1,608
7	玉ねぎ	1,651
8	みりん	976
9	ねぎ	1,088
10	片栗粉	932

表 5 定番度の大きい上位 3 レシピの使用食材

食材数	鶏のちゃんこ鍋	まるまるミート...	スプーンで食べ...
1	鶏もも肉	豚ひき肉	合いびき肉
2	鶏ひき肉	玉ねぎ	にんじん
3	ねぎ	ピーマン	玉ねぎ
4	しょうが	赤ピーマン	ピーマン
5	卵	ジャンボピーマン (黄)	しいたけ
6	酒	ねぎ	ひじき
7	こしょう	しょうが	卵
8	片栗粉	水	食パン
9	しょうゆ	しょうゆ	こしょう
10	ゆでたけのこ	こしょう	しょうゆ
11	菜の花	片栗粉	みりん
12	生わかめ	水	酒
13	干しいたけ	砂糖	砂糖
14	ねぎ	トマトケチャップ	水
15	油揚げ	酒	
16	キャベツ	しょうゆ	
17	玉ねぎ	酢	
18	水	いり白ごま	
19	みりん		
20	しょうゆ		
21	砂糖		

4.2.3 考 察

食材の定番度の順位から、定番度のスコア上位の食材の多くは調味料となっている。使用件数を見てわかるように、これらは多くのレシピで使用されている。調味料は、料理の味を決定するレシピの重要な要素である。しかし、使用量を考慮する際に他の食材との比較は困難であり、非常に多く使用されるので順位に与える影響の大きさも考慮し、重み付きの手法では除いた。

4.3 重み付き定番度算出手法の評価

4.3.1 使用量に基づく実験方法

食材の使用量を用いた計算は、食材と同様にレシピサイトから抽出を行うため、レシピサイトに記載されている使用量のみ扱う。使用量についても、食材と同様に表記の揺れが多く存在する。1 枚や、1 個、1 株など、グラム (g) 表示されていない食材については対応表を作成することにより、多くの食材の使用量を取得した。隣接行列の重みとして用いる際には、全レシピの平均重量によって正規化した値を用いる。

4.3.2 実験結果

重み付きの手法で算出した結果を図 9 と図 10 に示し、順位付けを行った結果を表 6 と表 7 に示す。縦軸は算出した定番度を対数目盛で表し、横軸はレシピと食材それぞれを定番度の高い順に並べているレシピと食材の定番度に基づき、順位付けを行った結果を表 6 と表 7 に示す。

4.3.3 考 察

表 7 の結果では、使用量が多く、多くのレシピで使用されている食材のスコアが高くなっている。同様に、スコアの低い食材の使用量が多く、スコアの低い食材をあまり使用していない

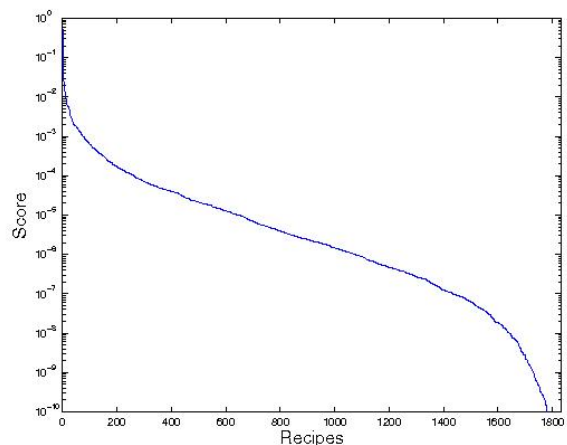


図 9 レシピの定番度 (重み付き, 調味料なし)

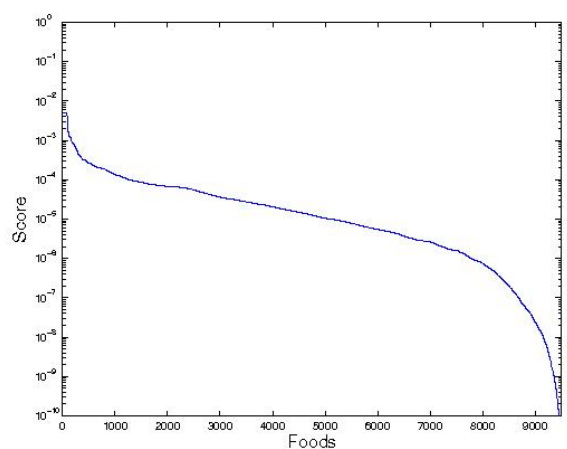


図 10 食材の定番度 (重み付き, 調味料なし)

表 6 使用量に基づくレシピの順位

順位	レシピ
1	ソーセージと野菜のスープ
2	さけのちゃんちゃん焼き
3	野菜の蒸し煮
4	ボークポトフ
5	豆腐と野菜の豆乳グラタン
6	たっぷり野菜のポトフ
7	ポルシチ
8	ザクザクキャベツと鶏のコンソメスープ
9	しめじと鶏肉の韓国炒め
10	なんでも野菜のスープ

レシピのスコアが高くなっている。この結果から、使用量の多い食材のスコアの考慮を実現し、使い勝手の良さを重視した定番度を算出できているといえる。

4.4 定番度に基づくレシピ推薦法

本手法で算出した定番度に基づくレシピの順位に対して、調理レパートリーの拡大に伴い出現する未知な食材の累積数と、使用済みの食材のみで構成されるレシピ数を用いて評価を行う。未知な食材とは、これまで調理したレシピに使用されていない

順位	食材	平均使用量 (g)
1	キャベツ	260.45
2	にんじん	113.55
3	玉ねぎ	141.63
4	大根	319.08
5	じゃがいも	175.63
6	かぼちゃ	451.71
7	もやし	179.52
8	鶏もも肉	250.39
9	小麦粉	60.308
10	ご飯	304.77

食材数	ソーセージと...	さけのちゃん...	野菜の蒸し煮
1	ソーセージ	生ざけ	にんじん
2	キャベツ	キャベツ	セロリ
3	玉ねぎ	玉ねぎ	玉ねぎ
4	大根	にんじん	トマト
5	トマト	ねぎ	キャベツ
6	ローリエ	もやし	かぼちゃ
7	パセリ	大根	ピーマン
8	粒マスタード	白菜	しめじ
9		にら	ベーコン
10			にんにく

食材、つまり調理したことのない食材である。定番度に基づくレシピの順に調理を行う場合に、累積食材数と使用済み食材のみで構成されるレシピ数の推移を図 11 に示す。図の縦軸は、累積食材数であり、レシピで使用する食材の累積をカウントしていく。使用した食材は使用済み食材とし、食材の累積数にはカウントしない。横軸は、使用済み食材のみで調理することができるレシピ数である。図 11 から明らかなように、初めは傾きが緩やかで途中から少しずつ傾きが急になっていることがわかる。傾きが緩やかな約 5,000 件までのレシピなどは、未知な食材の使用を増やさずにレパートリーを拡大できることから、初級者に推薦することができる。一方、傾きが急な 8,000 件か

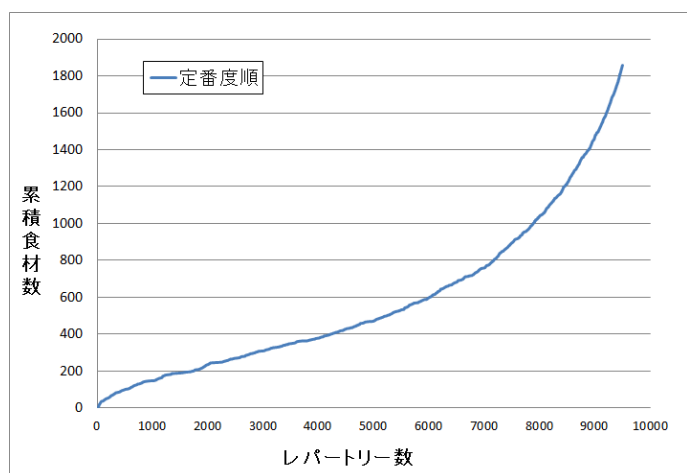


図 11 レパートリー拡大の推移

ら 9,000 件のレシピなどは、少ないレシピで使用済みレシピを増やすことができるため、上級者に推薦することができる。この様に、定番度に基づくレシピの順位を用いたレシピ推薦では、調理者に合わせた推薦をすることが可能であると考えられる。

5. ま と め

本論文では、調理レパートリー拡大のために、調理者に合わせたレシピ推薦法の提案を行った。レパートリーの拡大には、調理者の調理可能なレシピに合わせて、レシピの推薦を行うことが必要である。調理経験の少ない初級者は、限られた調理経験に基づいて、少しでも難しいレシピを推薦すれば調理可能と考えられる。レシピ推薦法の基盤として、レシピで使用される食材とレシピの関係から、レシピのスコアを算出し順位付けを行った。本研究では、レシピに記載される食材とレシピの関係を明らかにすることで、食材の使い勝手やレシピの馴染み深さの指標である定番度を算出した。

提案手法では、レシピサイトからレシピ、食材、食材の使用量を取得し、レシピと食材の 2 部グラフから隣接行列を作成した。隣接行列に HITS アルゴリズムを適用することで、2 部グラフからレシピと食材のスコアを算出した。提案手法の改良として、レシピでより影響の強い食材のスコアを高めるため、食材の使用量を重みとする隣接行列を作成し、使用量に基づく定番度を算出した。レシピサイトに掲載されているレシピデータを用いて定番度を算出し、定番度に基づく順位付けを行った。順位の評価として、未知な食材の累積数とそれに伴うレパートリー数の結果から、調理者に合わせたレシピ推薦の見通しを得た。

今後の課題は、定番度を用いたレシピ推薦の実現や、定番度の精度向上として大規模なデータを用いた定番度の算出を行っていきたい。

謝 辞

本研究の一部は、筑波大学図書館情報メディア系プロジェクト研究による助成を受けたものである。

文 献

- [1] 大山裕也, 塚原みな, 中内靖. ユーザの調理スキルを考慮した調理支援システムに関する研究. ロボティクス・メカトロニクス講演会講演概要集, Vol. 23, No. 1-4, pp. 36-46, 2009.
- [2] 岩本純也, 宮森恒. 調理の難易度を考慮したレシピ検索システムの提案. 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 (DEIM Forum 2012), E1-4, 2012.
- [3] 上田真由美, 石原和幸, 平野靖. 食材利用履歴に基づき個人の嗜好を反映するレシピ推薦手法. 日本データベース学会 letters, Vol. 1, pp. 29-32, 2008.
- [4] 八木一光, 岡野慎吾, 森本康彦. 相関ルールの 2 部グラフを用いた重要アイテムの発掘. 電子情報通信学会 第 19 回データ工学ワークショップ (DEWS2006), 6A-i7, pp. 1-8, 2008.
- [5] 山口雅史, 大島裕明, 小山聡, 田中克己. サーチエンジンのクエリログを利用した同位語・話題語の発見と可視化 (履歴応用). 電子情報通信学会技術研究報告. データ工学 (DE), Vol. 106, No. 148, pp. 121-126, 2006.
- [6] J. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *J. of the ACM*, Vol. 46, pp. 604-632, 1999.