



HOKKAIDO UNIVERSITY

Title	選好度に基づく集合推薦と調理レシピへの応用
Author(s)	白井, 康之; 鶴間, 浩二; 高嶋, 宏之 他
Citation	電子情報通信学会論文誌D, J96-D(5), 1246-1254
Issue Date	2013-05-01
Doc URL	https://hdl.handle.net/2115/68339
Rights	copyright©2013 IEICE
Type	journal article
File Information	j96-d_5_1246.pdf



選好度に基づく集合推薦と調理レシピへの応用

白井 康之^{†,††a)} 鶴間 浩二^{††} 高嶋 宏之[†] 小山 聡^{††}

Set Recommendation Based on Data Preferences and Its Application to Cooking Recipes

Yasuyuki SHIRAI^{†,††a)}, Koji TSURUMA^{††}, Hiroyuki TAKASHIMA[†],
and Satoshi OYAMA^{††}

あらまし 大量の既存データをもとにしたソーシャルフィルタリングは、過去の類似データから関連アイテムを領域知識なしで推薦することができる。大量データから推測されるアイテムは一般に関連度も高く、インターネット上のショッピングサイトなどにおいて、幅広く実用化されている。一方、調理レシピや運動行動履歴などのように、各データに付随して、「簡単な調理」「パーティ向けの料理」「減量に成功した行動パターン」といったような解釈、すなわち様々な軸による評価が与えられているデータも近年では数多く蓄積されている。本論文では、このような評価結果が与えられているデータに対して、選好度をもとにした集合推薦機能と二分決定グラフを用いた効率的な計算手法を提案する。複雑な計算を必要とする集合推薦において、本論文で提案する計算手法は、出現アイテムに偏りがある現実的なデータにおいてより効率的である。また、本手法を調理レシピに対して適用した実験結果を示す。

キーワード 情報推薦, 集合推薦, 調理レシピ, 二分決定グラフ

1. ま え が き

大量の既存データをもとにしたソーシャルフィルタリングは、過去の類似データから関連アイテムを領域知識なしで推薦することができる。大量データから推測されるアイテムは一般に関連度も高く、インターネット上のショッピングサイトなどにおいて、幅広く実用化されている [1], [2]。

一方、調理レシピや運動行動履歴などのように、各データに対する評価結果や解釈が与えられているデータも近年では数多く蓄積されている。例えば、調理レシピでは、「簡単な調理」「パーティ向けの料理」「目新しい料理」といった評価軸があり得るし、運動行動履歴では、「減量に成功した行動パターン」といった評価

軸があり得る。

本論文では、このような評価結果が与えられているデータに対して、選好度をもとにした集合推薦機能の枠組みを提案し、二分決定グラフを用いた効率的な計算手法を提案する。複雑な計算を必要とする集合推薦において、本論文で提案する二分決定グラフを用いた計算手法は、出現アイテムに偏りがある現実的なデータにおいてより効率的である。

また、本手法を実問題に適用した例として、調理レシピに対する実験結果を示す。近年、インターネット上のサービスとして、調理レシピに関する情報交換が盛んに行われている。ユーザの視点に立てば、ユーザの置かれた状況、例えば食材の入手可能性や調理者の置かれた環境や技量などに応じたレシピ推薦が必要と考えられる。本論文では、食材の入手のしやすさや、レシピへの反響などの観点からレシピ改善に関して提言を行う実験システムについて報告する。

本論文の構成は以下のとおりである。まず 2. では、本論文で対象とする情報推薦の仕組みについて概説する。特に、既存の協調フィルタリングを拡張した選好度に基づく集合推薦の仕組み、また、ゼロサプレステype二分決定グラフ (ZDD : Zero-Suppressed Binary

[†] 科学技術振興機構 ERATO 湊離散構造処理系プロジェクト, 札幌市

JST-ERATO MINATO Discrete Structure Manipulation System Project, Sapporo-shi, 060-0814 Japan

^{††} 北海道大学大学院情報科学研究科, 札幌市

Graduate School of Information Science and Technology, Hokkaido University, Sapporo-shi, 060-0814 Japan

^{†††} NEC ソフトウェア北海道, 札幌市

NEC Software Hokkaido, Sapporo-shi, 060-0808 Japan

a) E-mail: shirai@erato.ist.hokudai.ac.jp

Decision Diagram) [5], [7], [8] を用いた効率的な実装手段について説明する．3. では、本手法の適用例として、クックパッド^(注1) のレシピデータを用いた情報推薦に関する実験結果について報告する．4. はまとめと今後の課題である．

なお、本論文に先行する筆者らの論文として [3] がある．本論文は [3] で示した手法を選好度に基づく集合推薦として整理し直すとともに、調理レシピに対する適用結果を新たに示したものである．

2. 集合推薦

図 1 に、一般的な協調フィルタリングと、近年提案されている集合推薦の仕組み、また本論文で対象とする選好度に基づく集合推薦の仕組みの概要を示す．伝統的な協調フィルタリングでは、アイテムはユーザーの嗜好や過去の履歴との関連性に基づき推薦される [1], [2]. ここでは、各アイテムは互いに独立であるとされ、アイテム間の関連性は考慮されず、共起確率が高いものが推薦されるのが一般的である．例えば図 1 の (1) に示した「協調フィルタリング」においては、推薦されるアイテム f と h の関連性は考慮されることはなく、推薦対象であるアイテム集合 $\{a, b, c, d, e\}$ に対して、データベース中の頻出パターンから推薦すべきアイテムが決定される．アイテム間の相互関連性が薄く、それぞれ独立なアイテムとして考えられる場合には、こうした協調フィルタリングの仕組みが有益である．

これに対して、アイテム間の関連性や依存関係を考慮した情報推薦方法も検討されている (図 1 の (2)). 例えば、Xie ら [10] は、旅行プランの策定において、コスト的な制約を考慮した情報推薦手法を提案している．図 1 の (2) の例では、観光地 f と g 、あるいは観光地 h と j は同時に訪問した事例があることから集合としての推薦対象になるが、一方、 f と h はともに頻出する有名な観光地であっても、同時に訪問することがコスト的に困難で、かつ過去にこれらを同時に訪問した例がなければ、 f と h をセットで推薦することはない．また、Parameswaran ら [11] は、大学の履修科目選択において、科目間の関連性 (前提とする知識の依存関係)、例えば、同時履修すべき科目や前提となる履修科目といった制約を考慮して、科目集合の推薦を行う仕組みを提案している．このような情報推薦の仕組みは、集合推薦 (Set Recommendation) と呼ばれている．

一方、本論文で扱う選好度に基づく集合推薦は、こ

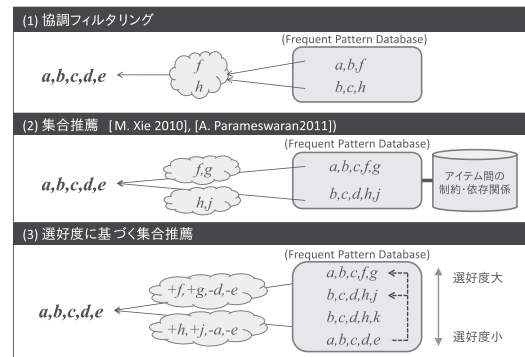


図 1 協調フィルタリングと集合推薦

Fig. 1 Collaborative Filtering and Set Recommendation.

うした集合推薦の考え方、すなわち意味のあるアイテム集合を推薦するという考え方に基づくが、データに何らかの選好度に基づいた評価が与えられている場合を扱う (図 1 の (3)). 選好度に基づく情報推薦では、集合推薦の枠組みは以下のように拡張される [3].

まず各アイテム集合には、優良さやコストなど、何らかの選好度が付与されているものとする．ここで、あるデータが与えられたとき、どのようにアイテムを変更すれば選好度を高めることができるかを提示することが本情報推薦の目的である．アイテムの変更とは具体的には、アイテムの追加あるいは削除を表す．ショッピングサイトにおける情報推薦では、“今後追加すべきアイテム”のみを推薦することになるが、日常生活行動パターンなどに関する情報推薦ではアイテムの追加のみならず、削除もまた重要となる．例えば、減量を目的とした日常行動記録から、減量に成功したグループと成功していないグループ間の差異に着目することにより、あるユーザーに対して、どのような生活改善をすれば減量に成功したグループに近づくかを推薦する．こうした例においては、ユーザーの行動パターンはできるだけ変更が少ないことが望ましい．営業職で夜の付き合いが多いユーザーに対して、生活パターンや食生活をがらりと変えるようにアドバイスすることは現実的ではない．このため、アイテムの追加あるいは削除の数を制限し、その制約の中でより選好度を高めるためにはどうしたらよいかを提示する．

以下、本章ではまず我々が用いた集合推薦手法について述べ、次章でレシピ情報への適用について紹介

(注1) : <http://cookpad.com>

する。

2.1 選好度に基づく集合推薦手法

インターネットショッピングにおける顧客数やレシピサイトにおけるレシピ提供数に見られるように、一般にインスタンス数は膨大な数にのぼる。本論文で示す情報推薦を実現するためには、膨大なインスタンス集合を効率良く管理し、与えられた制約のもとで比較・探索を行う機能が必要である。本節では、まず我々が提案する集合推薦の例を示し、次節以降で、ゼロサプレス型二分決定グラフ (ZDD : Zero-Suppressed Binary Decision Diagram) [5], [7], [8] を用いた推薦アルゴリズムを示すこととする。

(例) 今、アイテム集合を $\Sigma = \{a, b, c, d, e\}$ とし、データベースに、以下のようにパターン集合と選好度が定義されているとする。

パターン	選好度
$\{a, b, c\}$	1
$\{b, c\}$	2
$\{c\}$	1
$\{d, e\}$	2
$\{e\}$	3

今、 $D = \{a, c\}$ なるアイテム集合を考える。追加アイテムと削除アイテムの制約をそれぞれ $N_{add}^D = N_{delete}^D = 1$ (アイテムの追加と削除はそれぞれ 1 以下であること) と係数制約 $M = 1$ (選好度 1 以上をもつアイテム集合への推薦) を満たす推薦は、以下の 3 通りがあり得る。

- b を追加する。
- a を削除し、 b を追加する。
- a を削除する。

これらの変更を D に対して施すことにより、変更されたアイテム集合は、選好度 1 以上という制約を満たす。ここで、もし係数制約を $M = 2$ とするのであれば、“ a を削除し、 b を追加”のみが候補となる。

このような推薦アイテム集合の検出は、いわゆる条件付きの類似検索に相当するものであるが、アイテム集合が膨大になるケースや、追加アイテム、削除アイテムの制約が複雑になるケースでは、効率的な計算方法が求められる。次節では、ZDD に基づく処理系 VSOP (Valued-Sum Of Products calculator) を用いた実装方法について概観する。

2.2 ZDD と VSOP

二分決定グラフ (Binary decision diagrams) [4], [5]

は、大規模ブール関数を効率的に表現し取り扱うことができる仕組みとしてよく知られたデータ構造であり、変数を表す節点と各変数の真理値を表す枝 (0 枝と 1 枝) から構成されている。BDD の簡約化ルールは、“冗長な節点の削除” (1 枝と 0 枝が同じ節点を指しているとき、その節点を冗長なものとして削除する) と“等価な節点の共有” (等価なサブグラフを共有する) からなっている。一方、ゼロサプレス型 BDD (ZDD) [5], [7], [8] は、疎な組合せデータ集合に特化した、いわば BDD の変形であるが、その意味論は BDD とは若干異なっている。ZDD の簡約化ルール、並びに、ブール関数 $F = (a \wedge b) \vee (a \wedge c) \vee c = ab + ac + c$ (a, b, c を論理変数とする) に対する ZDD による簡約化の例を図 2 に示す。図中の “Binary Decision Tree” は、上記 F を二分木で表したもので、ZDD の簡約化ルールを適用することにより、“ZDD Representation” で示すような簡約化が可能であることを表している。ZDD での簡約化ルールは、以下の二つから構成されている。

- 等価な節点の共有 (BDD と同様) (図 2[1])。
- 1-枝が直接 0 終端節点を指している節点を削除 (図 2[2])。

一方、BDD での簡約化ルールにあるような“冗長な節点の削除”は行わない。すなわち、BDD では存在しない節点は、1 でも 0 でも成り立っていたのに対して、ZDD では存在しない節点は 0 のみ成り立つことを表している。このような BDD の改良は、特に、疎なデータ構造に対しては、BDD と比較して極めてコンパクトに表現できることが分かっている。例えば、インターネット上の購買履歴では、商品数は膨大にあるものの、実際に購入されたアイテムはごく一部である。購入されないアイテムに対して、BDD のように節点を残しておくことは明らかに非効率であり、ZDD のように購入されたアイテムのみの関連性を示した方

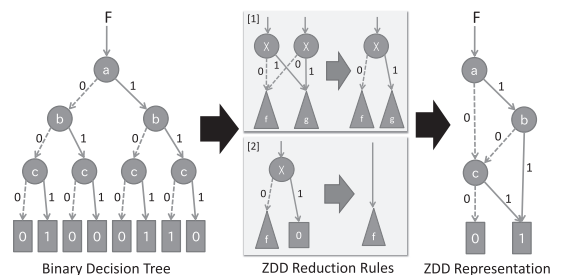


図 2 二分木表現と ZDD の簡約ルール
Fig. 2 Binary tree and ZDD reduction rules.

が効率的である．レコード数やアイテムの分布にも依存するが，アイテムの出現に偏りがある場合には，数百倍，数千倍もの圧縮効果があることも多い．

VSOP (Valued-Sum-Of-Products Calculator) [9] は，ZDD 構造に基づき，組合せ集合を多項式として表現し，ZDD 構造に対する演算，あるいは ZDD 構造間の演算を行う処理系である．多項式の各項は，アイテムの組合せを表し，係数は，重みを表している（係数には負数も許す）．

例えば，多項式 $abc - ac + 2bc + c + 2de + 3e - abd$ は，七つのアイテム集合 $abc, ac, bc, c, de, e, abd$ からなり，各アイテム集合はそれぞれ $1, -1, 2, 1, 2, 3, -1$ といった重みをもっている．

VSOP においては，各項の係数を表現するために (-2) 進表現 [8] を用いている．すなわち，VSOP の最上位節点は， $1(= (-2)^0), -2(= (-2)^1), 4(= (-2)^2), -8(= (-2)^3), 16(= (-2)^4), \dots$ の各ビットにそれぞれ対応している．例えば， -12 は， $(-2)^5 + (-2)^4 + (-2)^2$ と一意に表現することが可能である．

図 3 は，式 $abc - ac + 2bc + c + 2de + 3e - abd$ に対する VSOP 表現を表している．ここで例えば， ac なるアイテム集合を考えると，これは最上位節点である $+1$ 並びに -2 のみを真とするので， ac の係数は， $+1 - 2 = -1$ と計算することができる．

以上のように，VSOP では，最上位節点が (-2) 進数でラベル付けされた ZDD 構造により，整数を係数とする多項式を効率的に表現することが可能である．また，VSOP 代数に基づく和，差，積，商，比較演算等の多項式間の演算処理が実装されている [7], [9]．

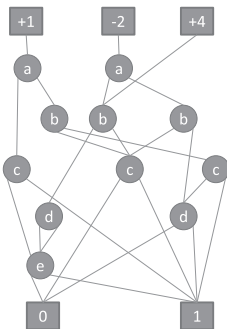


図 3 VSOP による $abc - ac + 2bc + c + 2de + 3e - abd$ の多項式表現

Fig. 3 Polynomial representations in VSOP for $abc - ac + 2bc + c + 2de + 3e - abd$.

2.3 ZDD 構造に基づく集合推薦

本節では，ZDD 構造に基づく集合推薦の概要について説明する．はじめに，前述の選好度をもったデータ集合は，以下のような多項式で表現することができる．

$$C = abc + 2bc + c + 2de + 3e$$

ここである与えられたアイテム集合 D に対して， N_{add}^D と N_{delete}^D 並びに係数 M の制約のもとで， D を修正（アイテムの追加と削除）して得られるアイテム集合の候補を，上記の C より抽出することが目的となる．

図 4 は， $N_{add}^D = N_{delete}^D = 1$ かつ $D = ac$ において， C を ZDD 構造上で探索するプロセスを表している．探索は，各最上位節点 $+1, -2, +4$ からそれぞれ開始する縦型探索として行われ，各枝上で，追加アイテム数，削除アイテム数を管理する．制約を満たさないことが分かった場合には，それ以降の枝の探索をまとめて枝刈りすることができる．例えば，図 4 左側（図中の * 箇所）の c から伸びている 0-枝は， $D = ac$ に対して a, c を共に削除する枝であるため， $(add, delete) = (0, 2)$ であり， N_{delete}^D の条件を満たさない．したがって，この枝から下の探索をまとめて省略することができる．

以上の手続きに従って，各最上位節点に対して，true (1) となる候補を求めると，以下ようになる．

最上位節点	アイテム集合
+1	: abc, ac, c
-2	: bc, ac
+4	: bc

上記の結果より，各アイテム集合に対して以下の係

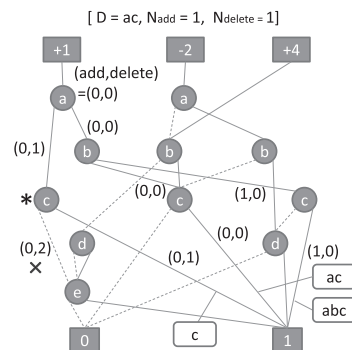


図 4 ZDD 構造上の探索
Fig. 4 Search on ZDD structures.

表 1 実行結果 (実行効率評価)
Table 1 Experimental results. (Performance evaluation)

(a) 乱数パターン (Time : ms)

	本提案手法 (ZDD based)			フラット検索		
	1 M	5 M	10 M	1 M	5 M	10 M
$N_{add} = 1, N_{delete} = 1, M = 1$	27	34	36	101	439	855
$N_{add} = 2, N_{delete} = 2, M = 1$	201	235	247	168	539	998
$N_{add} = 3, N_{delete} = 3, M = 1$	255	443	604	179	580	1084
$N_{add} = 3, N_{delete} = 3, M = 2$	290	457	597	176	587	1092
$N_{add} = 3, N_{delete} = 3, M = 3$	249	458	602	178	586	1089

(b) 固定パターン (Time : ms)

	本提案手法 (ZDD based)			フラット検索		
	1 M	5 M	10 M	1 M	5 M	10 M
$N_{add} = 1, N_{delete} = 1, M = 1$	24	30	34	103	433	841
$N_{add} = 2, N_{delete} = 2, M = 1$	114	133	151	110	468	903
$N_{add} = 3, N_{delete} = 3, M = 1$	217	232	258	178	562	1042
$N_{add} = 3, N_{delete} = 3, M = 2$	205	238	261	183	567	1040
$N_{add} = 3, N_{delete} = 3, M = 3$	220	251	252	185	570	1065

数を得る.

アイテム集合	重み
bc	: +2
abc	: +1
c	: +1
ac	: -1

したがって、係数制約を $M = 2$ とすると、これを満足する解は、 bc のみとなり、 $M = 1$ とすると、 bc , abc , c が推薦候補となる.

2.4 実行効率

上記で説明した ZDD をベースとした探索手法に關して、人工的に構成された問題に対して実行効率評価を行う. ここで対象とする問題は以下のとおりである. まず、アイテムは合計で 170 種類から構成され、データレコードは、それぞれ七つのアイテムを含むとする. データレコードは簡単のため、クラス P(positive) とクラス N(negative) に分割されているものとする. ここで、クラス P とクラス N との差分を取ることに、各アイテム集合の選好度が求まるとする. 実験では、推薦対象であるデータに対して、選好度を上げるように、アイテムの追加・削除を求めることが目的となる.

実験では二つのシナリオを用意した. 一つは、アイテムをランダムに発生させて作成したデータセットを対象とした場合であり (以下「乱数パターン」)、もう一つは、クラス P とクラス N にそれぞれ固有の長さ 3 のパターンを 20 種類ずつ用意し、クラス P のデータ、クラス N のデータはそれぞれ固有パターンのいずれかを必ず含むようにし、残りの四つのアイテムはランダ

ムに選択したものである (以下「固定パターン」とする). また、実験では、クラス P とクラス N のデータサイズをそれぞれ 100 万、500 万、1000 万のケースについて実行を計測した. 本システムは、Java で実装されており、実験は、SUSE Linux Enterprise Server 10 (quad-core AMD Opteron 3 GHz CPU, 512 GB RAM) 上で行った.

表 1 に結果を記す. 実行時間は、10 回の実行の平均を表す. また、表中で“フラット検索”は、比較のために用意した検索プログラムであり、ソートされたアイテム集合に対して、逐次的に検索を行うものである.

乱数パターンに対しては、本提案手法 (ZDD based) もフラット検索も固定パターンのケースに比べ、大きな差異は認められない. 実際、ZDD 構造は、このような偏りのないデータセットに対しては圧縮効果が低い. ZDD のサイズ (節点数) を見ても、1 M, 5 M, 10 M のデータセットそれぞれに対して、3369111, 14639216, 28694856 となっており、ほぼデータサイズに比例する形となっている. これに対して、固定パターンの場合には、1 M, 5 M, 10 M のデータセットそれぞれに対する ZDD のサイズ (節点数) は、2464685, 10562805, 18646702 となっており、データ数が増えても共有できる部分が多いために、乱数パターンのケースに比べ、圧縮率が高くなっており、共有部分が大きくなっていることを意味している. したがって、制約に従って検索した場合に、制約を満たさない部分をまとめて回避できるため、効率的な探索を行うことが可能である. 表 1 のケースでは、1 M のデータに対しては、ほとん

ど変わらないかむしろ本提案手法の方が遅くなっているが、データサイズが増加（5M, 10M）した場合には、フラット検索では、乱数パターンと同様、計算時間もほぼ比例して増加するのに対して、本提案手法の方ではほぼコンスタントな計算時間が達成できている。

以上示したように、本論文で説明した ZDD をベースとした探索手法は、特に、大規模かつ疎なデータベースを対象とした場合に、一般的なデータベースを用いた検索に比べて大きな効果があるといえる。購買履歴データや、本論文で対象とするレシピの食材データでは、ランダムなデータセットというよりは、その頻度が冪乗則に従うような偏った分布となっていることが予想されるため、本論文で示したような ZDD を用いた方法に実行効率上の効果があると考えられる。

3. レシピデータを用いた実験

本章では、上記で示した集合推薦の枠組みをレシピ提供サイトで提供されているレシピ情報に適用する。レシピ提供サイトでの情報推薦には、料理名やジャンル、食材、カロリーなどの情報を条件とした検索が可能になっているものが多いが、類似レシピ検索、特に、既に手元にあるレシピ情報に対して、どのような改善を施すことで目的とするレシピを得ることができるかを提示する推薦機能はない。現実的に調理する側にとってみれば、既に身近にある材料や自身の技量などから、ある程度作るべき料理の形が決まってしまうため、こうした既存レシピに対する改善の推薦は有用である。

また、推薦基準としては、できるだけ多くの人に反響のあるレシピにしたいというニーズもある一方で、身近な食材で作れるレシピ、あるいは作り方が容易なレシピという観点もある。本章では、クックパッドに掲載されているレシピ情報に対して、食材の入手可能性や反響の大きさ（コメント数）といった基準を求め、レシピ情報に対する様々な視点からの情報推薦を行った結果を示す。

3.1 推薦クライテリア

レシピに対する推薦のクライテリアとして、本実験では、以下の二つの視点を導入した。

- コメントの多さ（反響の大きさ）
- 食材の入手しやすさ

ここで、「コメントの多さ」はレシピ紹介記事中のコメントの個数を表す。また、食材については、表記揺れを手作業で修正した上で、約 5000 個の食材を頻度

に従って選別し、「仕事マーケットプレイス Lancers」^(注2)を利用して、アンケート結果をもとに「食材の入手しやすさ」の定量化を行った。アンケートでは、各材料について、入手可能、入手困難、不明のうちから一つを選択してもらい、10 人の回答を総合して、入手しやすいかどうかを各材料について判定した。また、メニューを構成する各食材の入手しやすさからメニュー全体の食材の入手しやすさを定量化した。

これらの視点は、全てのレシピ、あるいは全てのユーザにとって必要となる基準ではない。例えば、何らかのイベントで用いるレシピであれば、珍しい食材を用いたり、手順が複雑なことは問題にならないだろう。一方、一人暮らしのユーザからみれば、材料が容易に入手可能であることや手順が容易であることは、一般的には重要な要素となる。このように、ユーザの嗜好や価値観、あるいは調理する場面・状況に応じて、様々な視点から改善できるレシピを探すのが想定される利用方法である。

3.2 実験結果

表 2 に、情報推薦結果の一部を示す。「推薦の観点」は、コメント数の多さを基準にしたものと食材入手の容易性を基準としている。コメント数の多さを基準にした場合には、現状コメントがついていない（反響がない）レシピに対して、どのように変更をすれば反響のある（あるいは目立つ・他の人にとって参考になる）レシピになるかという観点から推薦を行ったものであり、一方、食材入手の容易性を基準にした場合には、より容易に入手可能な食材をもとにした類似のレシピを推薦するものである。また、「オリジナルのスコア」「推薦レシピのスコア」はそれぞれ、コメント数、食材入手の容易性の観点から、11 段階（0～10）にスコアをつけたものである。「削除」、「追加」はそれぞれ削除・追加するアイテム（食材）を表す。なお、ここでは削除アイテム、追加アイテムの上限はそれぞれ 1 とし、係数制約は 2 以上として実行した。

結果からは以下のことが見て取れる。(1) の例では、レシピに対する反響（コメント）が全くない「かぶとにんじんのポタージュスープ」に対して、にんじんをかぶの葉に入れ換えることで、類似したコメントのあるレシピに入れ換えが可能であることを表している。ただし、推薦レシピのスコアは 2 となっているので、コメント数が大幅に増えるというわけではない。これに

(注2) : <http://www.lancers.jp>

表 2 情報推薦の例 (add ≤ 1, delete ≤ 1)
Table 2 Experimental results. (Item recommendation)

推薦の観点			オリジナル (推薦の対象)		推薦されるレシピ		
改善点	オリジナルのスコア	推薦レシピのスコア	タイトル	食材	タイトル	削除	追加
コメント数の多いレシピ	0	2	かぶとにんじんのポタージュスープ	かぶ, たまねぎ, にんじん, コンソメスープの素, バター, 塩, 胡椒, 牛乳	カブのポタージュスープ	にんじん	かぶの葉
	0	6	カスタードクッキー	バター, ベーキングパウダ, 三温糖, 卵, 薄力粉	チョコレートパウンドケーキ	三温糖	板チョコ
	0	3		バター, ベーキングパウダ, 三温糖, 卵, 薄力粉	コーヒー風味のビスコッティ	バター	インスタントコーヒー
	0	5	レーズンパン	スキムミルク, ドライイースト, バター, レーズン, 卵, 塩, 強力粉	ごはんで作るレーズン食パン	卵	ご飯
	0	3		スキムミルク, ドライイースト, バター, レーズン, 卵, 塩, 強力粉	ミルクはちみつパン	レーズン	はちみつ
	0	3		スキムミルク, ドライイースト, バター, レーズン, 卵, 塩, 強力粉	ふんわり甘いたまごパン	レーズン	グラニュー糖
	0	5		スキムミルク, ドライイースト, バター, レーズン, 卵, 塩, 強力粉	食物繊維とカルシウムが豊富なイチジクのパン	レーズン	ドライイチジク
食材入手が容易なレシピ	0	2	レーズンパン	スキムミルク, ドライイースト, バター, レーズン, 卵, 塩, 強力粉	イタリア伝統の味 パネトーネ	スキムミルク	砂糖
	0	3	かぶとにんじんのポタージュスープ	かぶ, たまねぎ, にんじん, コンソメスープの素, バター, 塩, 胡椒, 牛乳	冷蔵庫にあるもので作る野菜チャウダー	かぶ	白菜
	0	7	優しい味のかぼちゃスープ	かぼちゃ, たまねぎ, にんじん, バター, 塩, 胡椒, 牛乳, 生クリーム	かぼちゃとにんじんのポタージュ	生クリーム	コンソメキューブ
	0	2	口どけ良好な濃厚ショコラ	グラニュー糖, ココアパウダー, ベーキングパウダ, 卵, 無塩バター, 薄力粉	ガトーショコラ風チョコレートケーキ	ココアパウダー	チョコレート

対して, (2) や特に (4)(7) などの例では, 一般的な材料ではないものを入れることで, 大きな反響のある珍しいレシピを構成することが可能であることを表している。

また, 食材の入手可能性に着目した推薦結果 (8)~(11) では, スキムミルクやココアパウダーといった食材に対する代替として, 砂糖やチョコレートが提示されている。(10) では, 生クリームをコンソメキューブと入れ換えることが示されており, かつ入れ換え後のスコアも非常に高くなっている。一般に生クリームも身近な食材ではあるものの, 本分析のもととなったアンケート結果では, 生クリームが常に家庭にあるとした回答は 10 件中 0 件である一方で, コンソメキューブは 10 件中 9 件が常に家庭にあると回答している。

以上のように, 全般的に見れば, まず材料の入手については, 入手が困難な食材に対する代替手段を提供するといった傾向が認められ, 当初想定していたとおりの結果が得られているように思われる。一方, コメント数の多さを基準としたパターンでは, 変わった食材を提示することで大きな反響を呼ぶというケースが多く見られる。「目立たせる方策」という観点では, レシピ提供者にとっては有益な情報ではあるが, 目新しい料理というユーザ側の視点からみれば, 必ずしも変わった食材を組み合わせる必要はなく, 別な基準が必要になるとも考えられる。

いずれにしても, ユーザが求める何らかの基準でラ

ンキングを求めておけば, 現在のレシピ構成に対して, 与えられた追加数, 削除数, 並びに係数制約のもとで, これらを満足するレシピ構成を推薦することが可能である。

3.3 考 察

本章で述べたような調理レシピに対する情報推薦では, 実用化を考えた場合には, 代替可能な食材や食材の好き嫌いといったより現実的な基準も取り組む必要があるが, 既存の手持ちのレシピに類似していて, かつ効果が高いレシピを抽出するという点でユーザ側の現実的な要望に応え得る機能を有していると考えている。

また, 調理レシピ以外に, 本論文で示した手法が有効であると考えられる分野として, 研究論文データにおけるキーワード分析 (他の研究グループと研究対象分野やテーマを比較して今後の候補となり得る研究キーワードを提示する) や, 健康支援サービスにおける行動履歴への応用 (運動履歴, 食事履歴などをもとに, ダイエットを目的として日常生活の改善を提言する), インターネット検索のキーワードの推薦 (過去の検索キーワードとアクセスの履歴から, 検索キーワードの修正を提言) などがある。いずれも, 既存の状況に対して, 何らかの目的での改善を必要とするものであり, かつ大規模なデータ構成が想定されることから, 本論文で示した集合推薦の目的に合致したテーマであるといえる。実際のデータの入手・整備自体が現状で

は大きな課題となっているものも多いが、今後ますますこういったデータの蓄積が進んでいくものと考えられる。

4. む す び

本論文では、選好度が付与されたデータセットに対して、アイテムの追加、削除制約並びに選好度に対する制約のもとでの集合推薦の枠組みを示し、ZDDに基づく実装方法により、現実世界のデータで想定される偏りのあるデータに対して、効率的に候補集合を求めることができることを示した。また、実際のレシピデータに対する適用結果を示し、ユーザのニーズにそった形で選好度に応じた情報推薦が可能となることを示した。ただし、推薦の細部においては、今後考慮すべき点もある。例えば、調理レシピの推薦に関しては、実用上有益なアプリケーションとして、近年様々な研究が行われている。苅米ら[12],[14]は、材料や手順、あるいは食材の組合せに着目し、調理レシピの構造的解析結果との融合を試みている。また、中岡ら[15]は、定番度や珍しさ等の概念を導入した調理支援システムを提案しているほか、高橋ら[13]は、レシピの具体的な内容をもとに、レシピ名における修飾表現に関する分析を行い、レシピ検索の精度を向上させる手法を提案している。レシピ検索という観点からすれば、食材のもつ意味的情報、例えば、手に入りやすさに加え、価格、味、雰囲気、栄養バランス、更には、個人の嗜好とのマッチングといった点も実用化にあたっては必須であり、本論文で示した手法を今後レシピ推薦の実アプリケーションに応用していく場合には、こうした拡張も必須となるだろう。

情報推薦手法に関する関連研究としては、以下が挙げられる。まず、Emerging Patterns [6]は、クラス間の差異に着目し、特徴的なパターンを抽出した上で分類を行う手法である。評価づけられたデータセットの差分に着目する点で本論文の手法とモチベーションは類似しているが、我々の手法では、必ずしも emerging パターンのみを対象としているわけではなく、また、本論文で述べた手法は、実際に情報推薦を目的としている点で大きく異なっている。

近年、提案された集合推薦 (Set Recommendation) [10],[11]は、集合としての類似性に基づき、アイテムの集合を推薦するもので、アイテム間の関連性あるいは文脈を考慮しているという点で、本論文の対象分野と重なる部分も多い。ただし、一般的な集合推

薦では、過去の事例との類似性に基づきアイテム集合を推薦候補とするのに対して、本論文は、選好度を高めるための推薦を行うものである。目的は大きく異なっているものの、アイテム間の関連性の考慮あるいは背景知識の利用については、今後検討していく予定である。

また、本論文で扱う探索手法は、制約の与え方が異なるものの、考え方としては、編集距離 (Levenshtein Distance) をもとにした動的プログラミングと極めて近い。ただし、本論文で示した手法は、選好度をもとにした改善のための情報推薦を行うものであり、利用者が示した条件に従って、より効果の高い案を抽出することが可能である。また、ZDD をベースにしたアプローチでは、大規模なデータセットを ZDD としてよりコンパクトにメモリ上に展開でき、その上で、様々な制約に基づく解を探索できるという点において、単なる制約を満足する解を抽出するだけでなく、今後の応用範囲が広い。例えば、あるアイテム集合をサブセットとして含む解の抽出や、逆に、あるアイテム集合を含まない解の抽出など、既存 ZDD で定義・実装されている演算処理を応用した解の抽出が行える可能性がある。

以上、本論文では、選好度に基づく新しい集合推薦の枠組みと、調理レシピ推薦への応用について述べた。今後の技術的な課題としては、あるアイテム集合に類似したものを与えられたデータ集合から抽出するだけでなく、異なるデータ集合間での曖昧マッチングの効率的な実装方法を検討している。例えば、過去にユーザが調理したレシピ群と、データベースに含まれるレシピ群とのマッチングにより、過去のユーザのレシピ群の中から、様々な観点から改良が可能な候補解をまとめて抽出することができる。併せて、より複雑な制約を考慮した探索手法を検討していく予定である。

謝辞 本実験では、クックパッド^(注1)に掲載されているレシピデータを使用しました。ここに記して謝意を表します。

文 献

- [1] P. Melville and V. Sindhwani, "Recommender systems," in Encyclopedia of Machine Learning, pp.829–838, Springer, 2010.
- [2] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE

(注1) : <http://cookpad.com/>

Trans. Knowl. Data Eng., vol.17, no.6, pp.734-749, 2005.

- [3] Y. Shirai, K. Tsuruma, Y. Sakurai, S. Oyama, and S. Minato, "Incremental set recommendation based on class differences," Proc. 16th Pacific-Asia Conference on Knowledge Discovery and Data Mining, LNAI 7301, pp.183-194, 2012.
- [4] R.E. Bryant, "Graph-based algorithms for Boolean function manipulation," IEEE Trans. Comput., vol.35, no.8, pp.677-691, Aug. 1986.
- [5] D.E. Knuth, "Zero-suppressed BDDs: A combinatorial alternative," The Art of Computer Programming: Bitwise Tricks & Techniques, vol.4, fascicle 1, pp.117-126, Addison-Wesley, 2009.
- [6] G. Dong, X. Zhang, L. Wong, and J. Li, "CAEP: Classification by aggregating emerging patterns," Proc. Second International Conference on Discovery Science, LNCS 1721, 1999.
- [7] S. Minato, "Zero-suppressed BDDs for set manipulation in combinatorial problems," Proc. 30th ACM/IEEE Design Automation Conference (DAC'93), 1993.
- [8] S. Minato, "Implicit manipulation of polynomials using zero-suppressed BDDs," Proc. IEEE European Design and Test Conference, 1995.
- [9] S. Minato, "VSOP (Valued-Sum-of-Products) calculator for knowledge processing based on zero-suppressed BDDs," Federation over the Web, LNAI 3847, 2006.
- [10] M. Xie, L.V.S. Lakshmanan, and P.T. Wood, "Breaking out of the box of recommendations: From items to packages," Proc. 4th ACM Conf. on Recommender systems, 2010.
- [11] A. Parameswaran, P. Venetis, and H. Garcia-Molina, "Recommendation systems with complex constraints: A course recommendation perspective," ACM Trans. Information Systems, vol.29-4, pp.20:1-20:33, 2011.
- [12] 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," 言語処理学会年次大会発表論文集, 2008.
- [13] 高橋良平, 小山 聡, 大島裕明, 田中克己, "投稿型レシピサイトにおけるレシピ名の修飾表現の適合性判定手法," 信学論 (A), vol.J94-A, no.7, pp.467-475, July 2011.
- [14] 苅米志帆乃, 藤井 敦, 料理レシピテキストを対象とした構造解析の高精度化, 信学技報, DE2012-8, 2012.
- [15] 中岡義貴, 杉本和香奈, 佐藤哲司, 既存レシピを活用した並行調理支援に関する提案, 信学技報, DE2012-1, 2012.
- [16] 白井康之, 鶴間浩二, 小山 聡, 高嶋宏之, "効率的な集合演算処理に基づくレシピ推薦機能," 信学技報, DE2012-12, 2012.

(平成 24 年 7 月 3 日受付, 9 月 27 日再受付)



白井 康之

昭 62 東工大・理・数学卒. 平元東工大大学院総合理工学研究科了. 同年, (株) 三菱総合研究所入社. 平 4~7 (財) 新世代コンピュータ技術開発機構出向. (株) 三菱総合研究所主席研究員を経て, 平 22 年 5 月より, (独) 科学技術振興機構 ERATO 湊離散構造処理系プロジェクト技術参事兼研究員. 平 23 年度より北海道大学情報科学研究科客員教授を兼務. 論理推論, データマイニング等の研究に従事. 博士 (工学). 情報処理学会, 人工知能学会各会員.



鶴間 浩二

平 15 室蘭工大・工・情報工学卒. 平 17 同大学院情報工学専攻了. 同年, 北海道日本電気ソフトウェア (株) 入社. 平 22~24 (独) 科学技術振興機構出向. 平 24 年 5 月より日本電気 (株) 出向. 組合せ計算アルゴリズムや情報推薦等の応用研究開発に従事.



高嶋 宏之

平 8 室蘭工大・工・応用化学卒. 平 10 北海道大学大学院理学研究科了. 同年, 北海道日本電気ソフトウェア (株) 入社. 平 24 から JST ERATO 湊離散構造処理系プロジェクト技術員, 現在に至る. 組合せ計算アルゴリズムに関する研究開発に従事.



小山 聡 (正員)

平 6 京大・工・数理卒. 平 8 同大学院工学研究科了. 平 14 同大学院情報学研究科博士後期課程了. 博士 (情報学). 日本電信電話 (株), 日本学術振興会特別研究員, スタンフォード大学 Visiting Assistant Professor, 京都大学大学院情報学研究科助教を経て, 平 21 より現職. 機械学習, データマイニング, 情報検索に興味をもつ. 平 17 年度人工知能学会論文賞. 平 21 日本データベース学会上林奨励賞. 情報処理学会, 人工知能学会, 日本データベース学会, IEEE, ACM, AAAI 各会員.