

推薦関係グラフのパターン抽出に基づく商品間の関係判定手法

北山 大輔^{†,a} 角谷 和俊^{‡,b}

[†] 工学院大学情報学部 [‡] 兵庫県立大学環境人間学部

a) kitayama@cc.kogakuin.ac.jp b) sumiya@shse.u-hyogo.ac.jp

概要 Amazon.com や楽天市場などのオンラインショッピングサイトでは閲覧商品に対して推薦商品が提示される。我々はこの推薦関係をエッジとし、商品集合は推薦による有向グラフの構造を持つと考えた。本稿では、この推薦関係グラフにおける商品の推薦、被推薦のパターンおよび商品カテゴリを分析することで商品間の関係を判定する手法を提案する。例えば、ある「プリンタ」に対してのみ「DVD-R」が推薦されており、その他のプリンタには推薦されていない場合、このプリンタに対して DVD-R は「レーベル印刷」の機能を表現する関係を持つ。本稿では、商品間の一般的もしくは特有な機能を表現する関係、上位/下位関係、同等関係を定義し、それらの判定手法について議論する。また、プロトタイプシステムを構築し、実験を行った。

キーワード 情報推薦システム, 商品特徴抽出, 推薦関係グラフ, オンラインショッピング

1 はじめに

近年、楽天市場¹や Amazon²などのオンラインショッピングサイトが普及している。これらのサイトでは、ある商品を閲覧している際に協調フィルタリング [1, 2] による商品の推薦が行われている。多くのユーザは、ある目的に基づいて関連する商品を買うと考えられる。例えば、「プリンタ A」という商品は用紙の印刷や写真印刷、レーベル印刷などで使用されるので、推薦商品には「印刷用紙」「デジタルカメラ」「DVD-R」などの商品が推薦されている。また、類似する機能を持つ「プリンタ B」も推薦される。我々は、このような商品推薦システムの推薦商品にはユーザから見た商品間の関係が反映されていると考えた。すなわち、「印刷用紙」「デジタルカメラ」「DVD-R」とプリンタ Aの間にはプリンタ A に対する機能という関係があり、「プリンタ B」とプリンタ Aの間にはライバル・姉妹品（同等関係）や上位機種・人気商品（上位・下位関係）という関係があると考えられる。本研究では商品間の推薦と被推薦の関係をグラフとして表現し、商品カテゴリと組み合わせることにより、ある商品と他の商品間の関係を判定する手法を提案する。具体的には、他の類似商品には推薦されていない商品や、多数の類似商品によく推薦される商品といった推薦関係グラフ上のパターンが商品間の関係の特有度合いを表すと考えた。例えばある「プリンタ」にのみ「デジタルカメラ」が推薦され、ある「レインコート」に対してのみ「双眼鏡」が推薦されていた場合、この「プリンタ」は特に「デジタルカメラ」の写真印刷に使用され、この「レインコート」は、特に「双眼鏡」で動かずに使用する時に使われるものであると推測できる。また、「プリン

タ」およびその類似商品から「プリンタ用インク」が推薦されている場合、一般的な機能として印刷があることが推測できる。さらに、カテゴリの同一性が商品全体に対する関係、もしくは商品の機能に対する関係を表すと考えた。提案手法の応用として、類似商品集合に対し商品間の関係を明示し、なおかつその機能の差異を提示する商品比較アプリケーションが考えられる。例えば、プリンタを検索している場合に検索結果としてサイズ違いの類似商品や追加機能がある類似商品が提示される。しかし、これらの商品集合の明確な違いを画像や商品名からのみ把握するのは難しく、ユーザはその商品ページを辿り、商品の詳細情報を閲覧する必要がある。この時、例えば、一方のプリンタは業務用で用紙印刷などによく用いられており、他方のプリンタは写真印刷などによく用いられているということが分かれば、写真印刷用のプリンタを求めているユーザは前者のプリンタより後者のプリンタの方が自身の使用目的に適していると判断ができる。用紙印刷という機能を表すために「印刷用紙」を提示し、写真印刷という機能を表すために「デジタルカメラ」を提示することにより商品選択支援を行う。本稿の構成を以下に示す。まず 2 章で想定する商品比較アプリケーションを説明する。次に 3 章において、推薦関係グラフと商品カテゴリを定義する。4 章では推薦関係グラフのパターンと商品間の関係の判定手法について議論する。5 章で実験結果とその考察を述べ、6 章でまとめと今後の課題について述べる。

2 商品選択支援システムの概念

我々は推薦関係グラフのパターンから抽出した商品間の関係を用いて、類似商品間の機能の差異を提示するシステムを開発する。具体的には類似商品間の同等、上位・下位関係、個々の商品の機能を表現する異種の商品

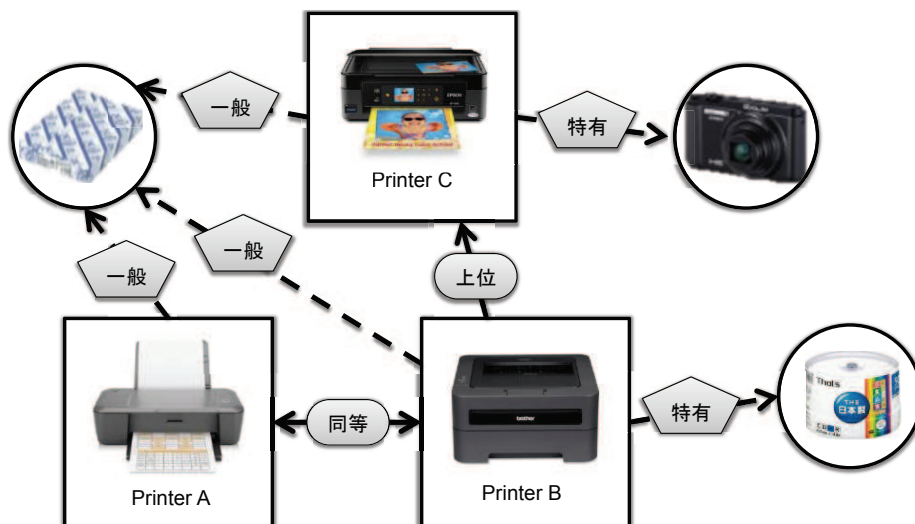


図1 商品間の関係を用いた商品選択支援システム

を提示するシステムを想定している。このような支援を行うことで、商品説明文や商品レビューには出現してこない、購入者やショッピングサイト側が想定する機能などを抽出することができ、ユーザは購入のための新たな参考情報を取得できる可能性がある。図1に想定するアプリケーションのイメージを示す。ここではプリンタを検索することを想定する。このとき、システムが判定した各類似商品間の関係およびその機能を提示する。図1の例ではプリンタAは業務用の用紙印刷などの機能を持つプリンタであり、プリンタBはデジタルカメラの写真印刷などの機能を持つプリンタということがわかる。この時写真をよく印刷するユーザの場合であれば後者のプリンタを容易に選択できる。この様に、特徴的な機能を表す商品を提示してやることで商品の特徴を把握し易くし、選択支援を行う。このシステムを実現するために、本稿では商品間の同等、上位・下位関係、一般・特殊な機能を表す関係判定する手法について議論する。

3 アプローチ

3.1 推薦関係グラフと商品カテゴリ

ショッピングサイトの商品推薦にはユーザの購入履歴が用いられており、同時に購入される商品や、見比べて閲覧される商品が推薦されていると考えられる。そのため、それらの推薦商品には、推薦元の商品と間に何らかの機能を表現するという関係が存在する場合がある。例えば、「プリンタ」に対して「DVD-R」が推薦されている場合、このプリンタにはDVDラベルの印刷機能があることがわかる。このような推薦商品からわかる機能には、予め付与された商品説明文や、ユーザの使用感を記述する商品レビューには表れにくい商品の機能が出現している場合がある。例えば、「ファービー」という人形の

おもちゃにおいてはその商品説明文には遊び方やその機能などが記述されており、商品レビューには実際に遊んだ子どもの反応などが記述されている。一方で、商品推薦システムでは上記の商品説明文や商品レビューの記述とは関係の無い、「ももいろクローバーZ」というアイドルのCDやDVDが提示される。この時、推薦理由はこの商品のCMでこのアイドルが登場していたからだと考えられ、アイドルを応援する目的で購入された商品であると考えられる。このように我々はショッピングサイトの商品推薦には、商品説明文や商品レビューに記述されていない機能が出現している場合があると考えた。我々は商品をノード、推薦関係をエッジとして、有向グラフで表現し、これを推薦関係グラフと定義する。図2では実際の商品を用いた推薦関係のグラフの例を示す。エッジの先が推薦商品を表し、エッジの元の商品に推薦されているとする。商品間に双方向のエッジが存在する場合、その商品は双方向の推薦関係を持つ。例えば、図2の中央の「Printer A」には「DVD-R」や種類の違うプリンタが推薦されている。また「Printer C」にも「Printer A」が推薦されているので、両者には双方向の推薦関係が存在する。本手法では推薦関係グラフに加えて商品カテゴリを用いる。商品カテゴリとは予めショッピングサイトにおいて割り振られている商品分類のことであり、すべての商品はいずれかの商品カテゴリに分類されている。例えば、「Printer A」は「パソコン・周辺機器 > プリンタ > インクジェットプリンタ」という商品カテゴリに分類される。また、「DVD-R」は「パソコン・周辺機器 > PCアクセサリ・サプライ > 記録メディア > DVD-R」に分類される。本稿では商品カテゴリの最下層に対する所属関係のみを用いる。



図2 推薦関係グラフの例

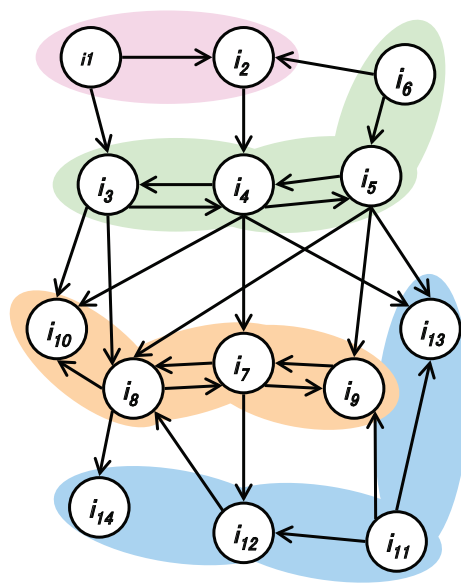


図3 推薦関係グラフと商品カテゴリ

3.2 関連研究

河内ら [3] はネットオークションにおける出品者と落札者の関係を、ユーザをノード、購入関係をエッジとし、有向グラフと考え、HITS アルゴリズムを適用することでネットワークの解析を行った。本研究では商品をノードとし、推薦関係を有向グラフと考えるが、全体の傾向ではなく、一部の推薦パターンに着目してグラフを解析する点が異なる。また、HITS アルゴリズムを推薦関係グラフに適用した場合、商品群の中で人気（推薦されやすい）の商品を見つけることは可能であると考えられ、効率的に上位・下位関係を発見できると考えられる。我々は、それに加えて機能を表現する商品の判定、および同等品の判定を行う。佃ら [4] は主題語に対する意外な関連語を発見する手法を提案している。Wikipedia のグラフの解析において、主題語とその同位語、また主題語から関連語への辿り着きづらさと関連語の認知度を算出することで主題語と関連語の意外度を求めている。主題語とその同意語との差分を取る考え方が本研究と類似しているが、商品カテゴリの同一性からその種類の分類を行う点で異なる。オンラインショッピングサイトにおける商品の特徴を容易に把握するための研究が平山ら [5] によって行われている。ユーザが商品に対して付与する商品レビューを解析し、その評価観点や評価（極性）を可視化し、ユーザに対して商品の評判情報の理解支援を行うが、本研究では商品間の関係を用いて商品の比較を可能とするものである。レビュー分析を相補的に用いることでより効果的になると考えられる。土方らは [6] 情報推薦を行う際の理由として、ユーザに対して新しい発見が起るような商品推薦を行う手法の提案を行って

る。本研究における商品選択支援システムにおいても、ユーザの知らない発見性の高い関係でフィルタリングするなど相補的に用いることで有用に働く可能性がある。庄司らは [7] オンラインショッピングにおけるユーザの意思決定支援を行うためのインタフェースの改良を行うための実験を行っている。欲しい商品が決定していないユーザを対象とし、店員とのコミュニケーションモデルによっていくつかの仮説を立てている。本研究は、このうちの「商品コンセプトに関わる場面情報」に相当する情報の提示を目的としているといえる。

4 推薦関係グラフのパターン抽出

4.1 商品の特有度算出

ある商品 s と商品 t の関係を判定するために、推薦関係グラフのパターンを用いる。本稿では、商品カテゴリが同一の商品を類似商品とする。我々は、類似商品群には推薦されておらず、ある商品 s からのみ推薦される商品 t はその商品特有の機能を表すと考えた。この考えに基づき、商品 t の特有度を以下の式で算出する。

$$Score_t = 1 - \frac{|SimItem(s) \cap RecItem(t)|}{|SimItem(s)|} \quad (1)$$

ここで、 $SimItem$ 関数は商品 s の類似商品集合を返す関数であり、 $RecItem$ 関数は商品 t を推薦対象に含む商品集合を返す関数である。すなわち、類似商品数に対し、商品 t を推薦対象とする商品の割合で商品 t の特有度 $Score_t$ を決定する。特有度は 1.0 から 0.0 までの値をとる。少数の類似商品のみが商品 t を推薦対象とするならば、特有度は高くなり、商品 t は商品 s 特有の機能を表していると判定することが可能である。一方、多

表1 特有度と商品カテゴリを用いた関係判定

		特有度	
		$\geq \alpha$	$\beta \geq$
商品 カテゴリ	同	同等関係 姉妹品・ライバル商品等	上位・下位関係 上位機種・人気商品等
	異	特有機能関係 プリンタに対する DVD-R 等	一般機能関係 プリンタに対するコピー用紙等

表2 データセット：使用した商品カテゴリ

商品カテゴリ	商品数
食品 & 飲料 > めん類 > ラーメン	10
食品 & 飲料 > 米 & 雑穀 > ご飯パック	10
おもちゃ > ぬいぐるみ	10
パソコン & 周辺機器 > プリンタ	10
家電 & カメラ > カメラ > 双眼鏡・望遠鏡・光学機器 > 双眼鏡	10
家電 & カメラ > カメラ > デジタルカメラ	10

数の類似商品から商品 t が推薦されているならば、特有度は低くなり、商品 t は商品 s の類似商品群に対する一般的な機能であると考えられる。例えば、「プリンタ A」を商品 s 、「DVD-R」を商品 t とした場合、「DVD-R」は「プリンタ A」に対してのみ推薦されており、「プリンタ A」の類似商品である「プリンタ B」や「プリンタ C」、「プリンタ D」には推薦されていない。この時、「プリンタ A」と「DVD-R」はこのパターンを満たす。図2を用いてこのパターンを説明する。図3のエッジは推薦関係を示し、エッジの先が推薦されている商品、エッジの元が推薦される商品である。同じ商品カテゴリの商品は同色で囲まれている。 i_4 を商品 s の「プリンタ A」とする。 i_4 の類似商品となるのは同じ商品カテゴリである i_3, i_5, i_6 である。 i_4 に推薦されている商品 i_7, i_{10}, i_{13} の中で類似商品に対して推薦されていないのは i_7 のみである。この時 i_4 と i_7 はこのパターンを満たし、特有度として1.0が算出される。閾値 α 以上の特有度をもつ商品を、その商品特有の機能を表す商品と判定し、閾値 β 以下の特有度の商品を、一般的な機能を表す商品と判定する。ただし、 α と β は $\alpha > \beta$ の関係である。

4.2 特有度と商品カテゴリを用いた関係判定

推薦関係グラフの形を用いることで算出した特有度と、商品 s と商品 t 間の商品カテゴリ上の関係を用いることで、商品 t の関係の種類を判定する。例えば、商品 s をあるメーカーのプリンタとする。商品 t を印刷用紙とする。商品 t は類似商品からよく推薦され、かつ商品 s と商品 t は異なる商品カテゴリであると考えられる。このような場合、商品 t は商品 s の商品カテゴリにおいて一般的な機能を表す商品であると考えられる。一方、商品 t が人気機種のプリンタであったとする。このような

場合、商品 t は類似商品から比較的良好に推薦され、商品 s と商品 t は同じ商品カテゴリであると考えられる。この場合は、商品 t は商品 s の商品カテゴリにおいて、より上位の位置づけにある商品であると考えられる。このように、我々は特有度の大小と、商品カテゴリの同一性から、商品 s と商品 t の関係の種類を判定する。表1に関係の種類をまとめる。

5 商品間関係判定の例

5.1 データセット

次のデータセットを用い特有度の算出を行い、商品カテゴリの同一性と組合せて関係の判定を行った。商品およびその推薦商品を Amazon Product Advertising API を用いて収集し、推薦関係グラフの構築を行った。商品カテゴリ内の商品をシードとし、そこから API から取得可能な推薦商品を2ホップ分収集することで、シードの商品に対する推薦関係の網羅度を高めた。データの規模は商品データ数が9,848件分であり推薦商品を含む総データ数は95,050件である。検証に用いる商品データは指定した各商品カテゴリから10件ずつ無作為に選んだシードとして用いた商品である(表2)。この計60件の商品データそれぞれの推薦商品について特有度を求めた。

5.2 算出結果と考察

算出結果の一部を表3, 4, および5に示す。表3では、実験に用いた「プリンタ」に対し「デジタルカメラ」や「写真用印刷用紙」が特有度の上位に判定されている。これらはこのプリンタの「写真印刷」という機能を反映していると考えられる。また、「複合機」が上位に判定されているが、これはインクの型番を共有する機種であり、

Proceedings of ARG WI2

表3 実験結果: Canon インクジェットプリンタ PIXUS IX6530 A3 ノビ対応 5色 W 黒インク コンパクトビジネス

推薦商品	特有度	カテゴリ
RICOH デジタルカメラ G700, 12.1メガピクセル 工学ズーム 5倍, 広角 28 mm, 防水 5 m 耐衝撃 2 m, 防塵	1.000	異
Canon 写真用紙 グロス ゴールド A3, 20枚 GL-101A320	1.000	異
Canon 写真用紙 グロス ゴールド A3 ノビ, 20枚 GL-101A3N20	1.000	異
Canon 高品位専用紙 HR-101 A3 ノビ HR-101A3N0BI	1.000	異
Canon 普通紙 ホワイト A3	1.000	異
Canon インクタンク BCI-326 (BK/C/M/Y) + BCI-325 マルチパック BCI-326+325/5MP	0.959	異
Canon インクジェット 複合機 PIXUS MX893 自動両面对応, ADF 搭載 FAX 有線・無線 LAN	0.951	同
ELECOM 詰め替えインク CANON BCI-325&326 互換 5色セット THC-MG5230RSET	0.894	同

表4 実験結果: CASIO デジタルカメラ EXLIM EX-ZS6 ブラック EX-ZS6BK

推薦商品	特有度	カテゴリ
サンワサプライ DG-BGM2BK コンパクトマルチケース (黒)	0.994	異
CASIO NP-80/NP-82 互換充電器	0.975	異
CASIO デジタルカメラ 充電器 (USB-AC アダプタ) AD-C53U	0.963	異
CASIO EXILIM EX-Z270 EX-Z1 NP-80/NP-82 互換バッテリー	0.963	異
CASIO EXILIM デジタルカメラケース EX-ZS6 ZS10 ZS12 ZS20 N1 N10	0.963	異
CASIO デジタルカメラ EXLIM EX-ZS6 シルバー EX-ZS6SR	0.956	同
CASIO デジタルカメラ EXLIM EX-ZS6 ピンク EX-ZS6PK	0.938	同
HAKUBA デジタルポーチ CS S-size Black SDP-CSS-BK	0.925	異
ELECOM 2.7 インチ液晶保護フィルム グロス DGP-007G	0.906	異
Transcend SDHC カード 16 GB Class10 TS16GSDHC10E	0.300	異

同世代の姉妹品であると考えられ、これらの同等関係として妥当であると考えられる。インクは機種毎に型番が異なるのが一般的であるため、インク類も上位となっているが、インクに関しては一般的な機能を表すと考えられる。そのため、特有度の上位に来ることは誤りである。今後の課題として、同商品カテゴリの商品が類似商品群から推薦されているかを判定する尺度を組み込むことが必要であると考えられる。

表4では、実験に用いた「デジタルカメラ」に対し「SDカード」が特有度の下位に判定されている。「SDカード」は写真撮影時のデータ記録に用いるものであり、一般的な機能という関係であると考えられる。デジタルカメラとSDカードは別カテゴリの商品であり、提案手法により一般的な機能という関係に正しく判定されている。

表5では、実験に用いた「インスタントラメン」に対し「モバイルバッテリー」が特有度の上位に判定されているが、「インスタントラメン」の特有な機能を表しているとは考えにくい。このような無関係な商品が判定される理由としては、特有度が「他の類似商品に無い推

薦商品ほど特徴的である」という考え方に基づくため、無関係だがまとめ買いされやすい商品が推薦関係グラフに出現した場合に特有度が高くなってしまいうからである。このような、無関係な推薦商品が全体にどの程度出現し、提案手法の精度にどのような影響を与えるのかを今度調査する必要がある。

これらのことより、我々は商品の特有度と商品カテゴリの同一性から商品間の関係を適切に判定できる可能性を確認した。今後は大規模なデータを用い実験することで、特有度の算出方法の改良および関係の判定精度の改善を行う予定である。

6 まとめ

本稿では商品間の関係を判定するために、ショッピングサイトでの推薦、被推薦の関係を推薦関係グラフとして表現し、類似商品などを含む推薦関係のパターンを挙げ、判定を行う手法の提案を行った。商品間の関係を明示することで、類似商品集合が与えられた場合にユーザの選択支援を行うシステムの提案を行った。実験を行い、

表 5 実験結果: 藤原製麺 札幌円山動物園白クマ塩ラーメン 112.4 g × 10 袋

推薦商品	特有度	カテゴリ
マルチデバイス対応 Cheero Power Plus 2 10400mAh (ブラック) 大容量モバイルバッテリー	1.000	異
藤原製麺 札幌円山動物園白クマ塩ラーメン 112.4 g × 10 袋 [特別価格] [数量限定]	0.970	同

特有度を用いた関係判定が正しく動作する可能性を確認した。しかし、一般的な目的を表す商品を特有であると誤判定する場合に対し、商品の粒度を適切に扱う手法が必要であることがわかった。また、無関係の商品が推薦グラフに含まれる場合が存在することがわかった。そのため、その量および提案手法への影響を調査する必要があると考えられる。今後の課題として、提案した以外のパターンの検証、提案アプリケーションの実装による有用性の評価などが挙げられる。

謝辞

本研究の一部は、平成 25 年度科研費若手研究 (B) 「ユーザ閲覧行動と周辺コンテンツの質的評価に基づく協調型画像撮影アシスト方式」(課題番号:24700098) によるものです。ここに記して謝意を表すものとします。

参考文献

- [1] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, GroupLens: open architecture for collaborative filtering of Netnews, In Proc. Conference on Computer Supported Cooperative Work, pp. 175–186, 1994.
- [2] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, Item-based collaborative filtering recommendation algorithms, In Proc. 10th International Conference on World Wide Web, pp. 285–295, 2001.
- [3] 河内 佑美, 森本 浩之, 吉井 伸一郎, ネットワーク構造解析による電子商取引システムユーザの特性分化, 情報処理学会研究報告, ICS [知能と複雑系], 2006(2), pp. 9-16, 2006.
- [4] K. Tsukuda, H. Ohshima, M. Yamamoto, H. Iwasaki, and K. Tanaka, Discovering unexpected information on the basis of popularity/unpopularity analysis of coordinate objects and their relationships, In Proc. 28th ACM Symposium on Applied Computing (SAC 2013), pp. 878–885, 2013.
- [5] 平山 拓央, 湯本 高行, 新居学, 佐藤邦弘, 語の共起と極性に基づく商品レビュー閲覧支援システム, 情報処理学会研究報告, データベース・システム研究会報告, 2012-DBS-155(3), 1-9. 2012.
- [6] Y. Hijikata, T. Shimizu, and S. Nishida, Discovery-oriented collaborative filtering for improving user satisfaction, In Proc. International Conference on Intelligent User Interfaces 2009 (IUI2009), pp. 67–76, 2009.
- [7] H. Shoji, S. Conart, An interaction method that facilitates concept articulation in shopping online, AI & Society, Vol. 19, No. 1, pp. 65–83, Springer-Verlag, 2005.