

研究指導 中澤 真 助教授

オンラインモールにおける 顧客の負担を軽減したリコメンデーションシステム

加藤 美希

1 はじめに

近年、Webマーケティングの発展に伴いオンラインモールの利用客が増加している。中でも楽天市場²は品揃えが豊富であり、同一商品に対して複数のショップを検索できるため、顧客が効率的に目的の商品情報を収集することができる。また、商品を購入した顧客による商品レビューによって、目的の商品を購入するか否かの判断材料も得ることができる。そのため商品の購入に限らず、買い物をする際の判断材料を得るためにオンラインモールを活用する利用客も増えている[12]。しかも増加しているのは利用客だけではない。楽天市場への出店数は2005年までの2年間で倍増しており、Yahoo!ショッピング³も出店料の安価を武器に、一気に出店数を伸ばした[11]ことから、オンラインモールは急成長を遂げている市場であることがわかる。

オンラインモールは、物理的制約を受けないため品揃えが豊富である。しかし、この種類の多さが顧客にとって不都合になる場合もある。それは、顧客が自分にとって最適な商品を効率的に探し出せないという問題である。顧客が気になっていた商品や興味のあるジャンルの商品を閲覧する場合であれば、検索バーや商品ジャンルから目的の商品に辿り着くことができるが、モール内の商品を全て把握することはできないため、本来ならば気に入るだろう商品を顧客が探し出せないのである。この問題は、結果的にショップ側の機会損失にもなり得るため、品揃えの豊富さは顧客とショップ双方のメリットであり、デメリットでもあるといえる。

そこで、個々の顧客の嗜好に合った商品を適切なタイミングで推奨するシステムが現在注目を集めており、そのようなシステムはリコメンデーションシステムとよばれている[10]。しかし、このシステムに関する研究の多くは、過去の顧客の購入履歴や商品評価などを元に推奨商品を抽出する仕組みであり、未購入客へは商品を推奨できないものとなっている[7]。また現在のオンラインモールでは、ある商品閲覧した全ての顧客に同様の商品が推奨されるため、個人の嗜好が反映されず、その上各商品への評価やWebアンケートへの回答を顧客が一つひとつ手作業でする場合もあり[3][4][5][6]、顧客に負担がかかるという問題が生じている。

よって本研究では、顧客の負担を軽減すると共に、より各々の顧客の嗜好に適した商品を推奨することを可能とするリコメンデーションシステムの構築を目的とする。条件として、顧客の嗜好を示す大量のデータと一定種

類以上の商品が必要となるため、大規模で利用者数の多いオンラインモールを研究の対象とする。

2 リコメンデーションシステムの概要

リコメンデーションシステムは顧客から明示的に、あるいは暗黙的に顧客のデータを取得し、それらを元に各顧客の嗜好に適した商品を推奨するシステムのことである。リコメンデーションシステムの手法は、主に二つある。以下にその概要を説明すると共に、それぞれの利点と欠点を挙げる。

2.1 協調フィルタリング

協調フィルタリングは膨大な情報の中からユーザの好みに合う情報を推薦するシステムとして、情報検索など様々な分野での研究が進んでいる[8][9]。リコメンデーションシステムに用いられる協調フィルタリングには、顧客間協調フィルタリングと商品間協調フィルタリングがある。前者は、顧客プロフィールとよばれる顧客の情報を、購買履歴や商品評価などのデータから生成する。プロフィール生成にはWebアンケートなどによって年齢や性別などの情報を明示的に用いる場合と、嗜好情報などの情報を暗黙的に用いる場合がある。このように生成したプロフィールに基づき、推奨対象顧客と類似のプロフィールを持つ顧客を抽出し、その顧客が購入・高評価している商品の中から、対象顧客が未購入の商品を推奨する。これは「Aさんが興味を持っている商品は、Aさんと同じ商品を購入したBさんもきっと好むだろう」というコンセプトに基づいている。後者の商品間協調フィルタリングはAmazon.com⁴で利用されている手法⁵で、ある商品と合わせて購入される頻度が高いものを推奨商品として選出している[1]。この手法では商品についてのプロフィールを生成しており、そのプロフィールから顧客が閲覧した商品と類似した商品を抽出し、推奨している。

顧客間協調フィルタリングの利点は、顧客間の類似性を比較するため、顧客にとって意外性のある商品を推奨できるという点である。一方、商品間協調フィルタリングは大量の顧客データを必要とせず、仕組みが単純であるゆえ処理速度が速い。両者共通の欠点は、プロフィールの生成手法の多くが、個々の商品に対しての評価を要求しているため顧客に負担がかかることと、購入履歴のない新商品の推奨は難しいという点である。商品进行评估することは、商品への好き嫌いを知ることができるため精度の高いプロフィールの生成が期待できるが、その反面顧客が時間と労力を費さなければならぬ。また、商品間協調フィルタリングにおいては、推奨商品が同種類の商品であるなどクロスセル的な販売方法になりがちであり、各々の顧客に対して嗜好に合致した商品が推奨されないといった欠点も挙げられる。

¹ 電子商店街。複数のオンラインショップが軒を連ねるサイトのこと。

² <http://www.rakuten.co.jp/>

³ <http://shopping.yahoo.co.jp/>

⁴ <http://www.amazon.com/>

⁵ 「Customers who bought this item also bought」という商品リストの提示が、Amazon.comのリコメンデーションである。

2.2 コンテンツベースフィルタリング

コンテンツベースフィルタリングは商品の属性データと顧客プロフィールを照合して商品を推奨する手法である。まず、コンテンツベースとなる商品をあらかじめ属性ごとにカテゴリに分類する。書籍を例にすると、カテゴリとは著者名や諸分野の集合のことである。次に、顧客が閲覧したWebページ内の特徴的なキーワードを顧客の嗜好情報として蓄積し、プロフィールを生成する。最後に、プロフィールの上位にあるキーワードとカテゴリ分けした商品の属性とを照合し、一致したカテゴリに属している商品を推奨する。

この手法の利点は、商品間協調フィルタリングの利点と同じく大量の顧客データを必要としない点である。欠点もまた商品間協調フィルタリングと同様で、推奨商品が同種類の商品になる傾向がある。さらに、映画や音楽などは比較的属性の設定が容易であるが、ファッションなど感性に差異があるものは、カテゴリに分類することが難しいといった欠点もある。

3 顧客の負担を軽減する顧客プロフィールの生成法

前章で指摘した問題を解決するため、ここでは顧客に負担を軽減すると共に、各々の顧客の嗜好に適した商品を推奨するプロフィール生成の方法を示す。

3.1 顧客プロフィールの概要と収集データ

本研究では、嗜好情報収集のためのデータとして商品の閲覧履歴を用いる。これには四つの理由がある。第一に、商品の閲覧は全ての顧客において通常の行為であり、顧客に負担がかからない。第二に、短期間で大量のデータ収集が容易であり、閲覧回数を重ねるほど顧客の嗜好を推察しやすい。第三に、購入履歴だけでは、オンラインモールでまだ商品を購入したことのない顧客へは商品を推奨できないが、閲覧履歴では未購入客へも商品を推奨できる。第四に、商品広告の見出しや画像を見て興味を持ち、商品について詳しく知りたいという動機が閲覧という行動に結びつき、より多様な嗜好情報を収集できる。

以上の理由から、顧客の閲覧履歴を用いて嗜好情報を抽出する。次節ではその具体的手順について説明する。

3.2 閲覧履歴を利用したプロフィール生成法

次に顧客プロフィールの生成の具体的手順を示す。

- (1) システム管理者が、システム構築時に商品に関連づけるための特徴的要素をシステムに登録する。商品が書籍だと仮定すると、ここでの例は“ビジネス書”や“経営学”、“経理”といった要素のことである。本研究では、この要素数を n とし、顧客の嗜好を表すプロフィールを n 次元の嗜好ベクトルとして表す。
- (2) ショップ運営者は商品を追加する際に、(1)でシステム上に登録された特徴的要素の中から、商品に関連しそうな要素を任意で1個あるいは複数個選択する。その記録はシステム上に蓄積される。
- (3) m 人の顧客を C_i とし、 n 個の商品の特徴的要素を E_j とする。 i は顧客の、 j は特徴的要素のインデックスである。 C_i の E_j に対する嗜好算出値を w_{ij} とする。嗜好算出値とは、顧客 C_i の特徴的要素 E_j に対する嗜好度合いのことである。特徴的要素に関連した商品を何回閲覧したかによって定める。また、顧客 C_i の嗜好の特徴をベクトル C_i

で表す。これを嗜好ベクトルとよび、次のように定義する。

$$C_i = (w_{1i}, w_{2i}, \dots, w_{ni})$$

ベクトルの各成分 w_{ij} を特徴的要素の閲覧回数でなく、商品の閲覧回数とする方法も考えられるが、それではほとんどの顧客が一つの商品を0~3回ほどしか商品を読まないと考えられるため、顧客の嗜好を判断することが困難である。そこでベクトルの要素を商品の特徴的要素とすれば、商品の閲覧回数の多少を考慮しなくても良い。そうすることにより、異なる商品でも特徴的要素が共通していればより顧客の特徴的嗜好を推察することができる。例えば、特徴的要素が“ビジネス書”、“経営学”の商品と、“小説”、“経営学”、“歴史”の商品、“経営学”、“感動”という商品を閲覧したとすれば、この顧客の“経営学”における嗜好算出値は3となり、“経営学”という要素への嗜好が強いと推察できる。表1は、顧客プロフィールの例である。

表1:顧客プロフィールの例1

	ビジネス書	小説	経営学	経理	感動	歴史	...	E_n
C_1	340	116	24	5	9	36		0
C_2	338	123	3	46	27	9		12
C_3	120	312	21	6	9	38		2
C_4	5	112	2	0	2	24		2
C_5	180	16	68	46	3	5		0
⋮								
C_m	134	45	85	13	5	31		14

顧客同士の類似性の有無を判断する手法として、本研究では内積を利用する。しかし、 w_{ij} の値をそのまま用いることは問題がある。なぜなら“ビジネス書”や“小説”などといった要素は高頻出語であり、 w_{ij} が過大になりやすいため多くの顧客が同じように大きな値をとりやすい。そのような要素を重視すれば顧客の嗜好に偏りがなくなってしまう。一方、“経営学”や“経理”、“感動”、“歴史”といった比較的細かい要素は個性的であり、顧客の嗜好が明確に表れる。よって過大になりがちな要素よりも、細分的な特徴的要素を重視するようなベクトルに補正しなければならない。例えば表1において、“ビジネス書”という要素においては顧客 C_1 と C_2 が最も近い値であったが、“経営学”や“経理”といった要素においては C_1 と C_3 が近い。 C_1 と C_2 の“ビジネス書”に対する嗜好算出値が大きいので、“ビジネス書”という要素に対する嗜好度合いが強いと判断できる。しかし、細分的で独特な特徴的要素を個人の嗜好として適切に反映させれば、この場合は C_1 と C_3 に類似性があると判断するべきである。このような特徴的要素 E_j 間における w_{ij} の差のばらつきを補正するため、 E_j それぞれの列において w_{ij} を正規化する。

まず、 E_j における w_{ij} の和 S_j を

$$S_j = \sum_{i=1}^m w_{ij}$$

とし、 E_j の列に対し w_{ij} を正規化した w'_{ij} を求める。これは E_j の和を1としたときの w_{ij} の値であり、次式のように定義する。

$$w'_{ij} = \frac{w_{ij}}{S_j}$$

この w'_{ij} をよって嗜好ベクトルの各成分も更新することになる。次に、内積を求めるためにベクトル C_i の大きさを求める。

$$\|C_i\| = \sqrt{\sum_{j=1}^n w'_{ij}{}^2}$$

しかし、このベクトルに対し、内積を類似度として算出することには問題がある。それは、顧客 C_i それぞれが持つベクトルの大きさに差異があるため、内積の値に影響が現れるということである。例えば、全く閲覧されていない商品を閲覧する顧客ほどその値は大きくなり、逆に皆が閲覧している商品ばかりを閲覧する顧客ほど、その値は小さくなってしまふ。そこで、ベクトルの方向を変えずにベクトルの大きさを1に正規化する。正規化された C_i の商品要素 E_j に対する嗜好算出値 w''_{ij} は、

$$w''_{ij} = \frac{w'_{ij}}{\|C_i\|} = \frac{w'_{ij}}{\sqrt{\sum_{j=1}^n w'_{ij}{}^2}}$$

となる。最終的に顧客の嗜好ベクトルは、

$$C_i = (w''_{1i}, w''_{2i}, \dots, w''_{ni})$$

と定義され、最終的に顧客プロファイルは表2のようになった。

表2: 顧客プロファイルの例2

	ビジネス書	小説	経営学	経理	感動	歴史	...	E_n
C_1	0.64	0.34	0.25	0.09	0.35	0.53		0.00
C_2	0.37	0.21	0.02	0.48	0.59	0.08		0.48
C_3	0.19	0.77	0.19	0.09	0.29	0.48		0.12
C_4	0.02	0.64	0.04	0.00	0.15	0.70		0.28
C_5	0.29	0.04	0.61	0.72	0.10	0.06		0.00
⋮								
C_m	0.17	0.09	0.60	0.16	0.13	0.31		0.67

3.3 コサイン類似度を用いた類似顧客の抽出

前項で生成した顧客プロファイルから、コサイン類似度を用いて、顧客間の類似度を算出する。仮に、顧客 C_1 と C_2 の類似度を求めるとすると、

$$\cos \theta(C_1, C_2) = \frac{C_1 \cdot C_2}{1 \times 1} = \sum_{j=1}^n w''_{1j} w''_{2j}$$

となる。

この手法では、コサイン類似度が0未満になることがないため、類似度の値が1に近ければ近いほど類似性が強く、0であれば両者の間に類似性がないものとする。これを、全ての顧客に対して算出する。

この類似度に基づき、嗜好が類似している顧客を抽出するため以下の二つの条件を設定する。ここで、 α と β は運用時にシステム管理者が調整するパラメータである。

- 条件1: 類似顧客数の上限を α 人とする
- 条件2: コサイン類似度の下限を β とする

条件1だけでは、類似度の低い顧客までもが類似していると判断されてしまう可能性がある。そこで条件2を追加して、類似度の高い顧客だけを集めることができるようにした。逆に、条件2の閾値を超える顧客が何百、何千人もいれば、推奨商品を抽出する際のタスクが大きくなり、処理に負担がかかってしまう。よって条件1により類似顧客数を制限しなければならない。このような理由から2つの条件を設定した。

4 顧客の嗜好を考慮した商品推奨法

前章において、顧客プロファイルからコサイン類似度によって C_i と嗜好が類似している顧客を抽出できることを示した。ここでは、その結果を用いて C_i への商品の推奨法について述べる。

4.1 閲覧回数履歴と購入履歴を利用した商品の推奨

ここでは、顧客 C_i に商品を推奨する場合を考える。 C_i と嗜好が類似している顧客 C_j 、ある商品ジャンル内の h 個の商品を G_l とし、表3のように C_i と G_l で交差した要素を f_{il} とする。 l は商品のインデックスである。ここでジャンルとは、同族の商品をひとまとめにした商品群のことである。例えば、ジャンルを「書籍」とすると、そこに含まれる商品 G_l には、本のタイトルが並ぶ。

商品の推奨には商品の閲覧回数履歴と購入履歴の両方を利用する。 C_i が G_l を閲覧した回数が f_{il} に格納される。ただし、 f_{il} は定数 k を超えないものとし、たとえ商品を k 回以上閲覧した場合でも、 $f_{il} = k$ とする。表3では、 k を4と定めた。また商品を購入した場合には、 $f_{il} = k+1$ とし、何度購入しても値は変化しないものとする。これは購入回数によって k の値が増加し続けてしまうことを防ぎ、商品を購入したことに最も高いウエイトを置くためである。

表3: 顧客の閲覧回数履歴と購入履歴

	G_1	G_2	G_3	...	G_l	...	G_h
C_i^*	1	0	1				0
C_1	0	4	0				1
C_2	0	3	1				3
C_3	1	5	1				0
⋮							
C_i					f_{il}		5
⋮							
C_m	1	2	0	5	3		2
合計	98	338	46	15	171		256

こうして定めた f_{ij} から f_{ij} の列の和を算出し、その和が最も高い商品の上位10位を C_i^* に推奨する商品として抽出する。これは顧客間協調フィルタリングのコンセプトと同様に、 C_i^* と類似した嗜好を持つ顧客が気になっている商品、また購入している商品は、 C_i^* も気に入る商品であるという考えに基づいている。

ただし、このシステムはパーソナライズされることで成り立つシステムである。そのため顧客はあらかじめ自分のアカウントを持ち、システムにログインすることが必要となる。ログインせず商品閲覧・購入すれば、プロフィールが生成されることも、商品が推奨されることもない。

4.2 コンテンツベースによる新商品の推奨

協調フィルタリングの欠点は、まだ誰も閲覧・購入していない商品は推奨できないことであった。しかし、今回提案したリコメンデーションシステムでは、新商品の推奨も可能となる。新商品を推奨するには、新商品のプロフィールと既に述べた顧客プロフィールの二つを利用する。

新商品のプロフィールとは、新商品に関連する特徴的要素をベクトルの成分として、新商品の特徴ベクトルを表したものである。ここで、 h 個の商品 G_j に含まれるある新商品を G_{new} とすると、 n 次元の新商品の特徴ベクトル G_{new} は次のように定義される。

$$G_{new} = (g_{1new}, g_{2new}, \dots, g_{jnew})$$

ショップ運営者が新商品の特徴的要素をシステムに登録することで、このプロフィールは生成される。そのとき、新商品の特徴的要素に該当するならば1、該当しなければ0を g_{jnew} に格納する。そして、顧客ごとに $g_{jnew}=1$ となる E_j 列の w_{ij} のみの平均を求める。つまり、商品プロフィールと顧客プロフィールに共通するインデックスである E_j の嗜好算出値 w_{ij} の合計を顧客 C_i ごとに算出し、それを g_{jnew} が1である E_j の個数で割ったものに等しい。最後に、その平均が という下限を満たしている顧客 C_i に新商品を推奨する。この も運用時点において調整されるパラメータである。

なお、新商品に関連するべき要素が提示された要素候補にない場合は、ショップ運営者がシステム管理者に要請して、新たな要素を追加してもらうことで対応する。よってショップ運営者が好き勝手に要素を付加することはできない。これは、複数の類似した要素や一つの商品にしか含まれないような特異な要素が乱立することを防ぐためである。また、特徴的要素数をなるべく少なくすることで、処理の軽量化を図るためでもある。

5 むすび

本研究では顧客の負担を軽減しながらも、各顧客の嗜好に合った商品を推奨するリコメンデーションシステムを構築した。負担を軽減するために、顧客のプロフィール生成に閲覧履歴を利用し、各顧客の嗜好をより特徴的なものにするためには、商品の特徴的要素によって顧客の嗜好ベクトルを表した。このことは、新商品の推奨をも可能にした。また嗜好算出値の正規化とコサイン正規化により、細分的な特徴的要素を重視し、閲覧回数数の多少が影響しない顧客プロフィールを生成できた。さらには、類似顧客の中で最も人気の高い商品を推奨することで、各々の顧客の嗜好に適した商品

を推奨できるシステムとなった。既存のシステムの欠点を改善したこの提案手法は、顧客間協調フィルタリングと、コンテンツベースフィルタリング両方の特性を併せ持つハイブリッド型のリコメンデーションシステムであるともいえる。

オンラインという特性を活かし、今後BtoC間の電子商取引数は益々増加の一途を辿ると考えられる。しかし、店員とコミュニケーションができる実店舗とは違い、オンラインモール(ショップ)は顧客と直接的に対応することができない。そのため価格以外での競争優位を獲得し、より顧客の心を掴むようなWebマーケティング戦略が必要となってくるだろう。その中で、「こんな商品が欲しかった」と顧客が一目で気に入る商品が推奨されれば、顧客にとっても適切であり、オンラインモールにとっても望ましいことである。また広がるWeb2.0の追い風を受け、幅広い層へのアプローチが可能である現在、Webビジネスにおけるロングテール効果は一層高まりをみせている。

提案したシステムでは、商品の特徴的要素を利用する点で、ショップ運営者にとっては若干の負担増となった。今後は、システムからショップ側への特徴的要素の提示手段や、さらなる処理の軽量化を図るための研究に取り組みたい。

参考文献

- [1] Greg Linden, Brent Smith, and Jeremy York, "Amazon.com Recommendations *Item-to-Item Collaborative Filtering*", IEEE INTERNET COMPUTING, pp. 76-90, 2003.
- [2] Eugene Agichtein, Eric Brill, Susan Dumais, and Robert Ragno, "Learning User Interaction Models for Predicting Web Search Result Preferences", SIGIR 06, August 6-11, pp. 3-16, 2006.
- [3] 石野正彦, 八巻直一, 市川照久, 水野忠則, "顧客の嗜好に合った商品推奨方法の提案", 情報処理学会研究報告2005-IS-92(6), pp. 37-42, 2005.
- [4] 石野正彦, 八巻直一, 市川照久, "顧客の嗜好に合せた商品のリコメンデーションシステムの提案", 情報処理学会第67回全国大会, 6TT2-1, pp. 619-622, 2005.
- [5] 石野正彦, 八巻直一, 市川照久, "商品属性を用いたリコメンデーションシステムの提案", 情報処理学会研究報告2004-IS-90(6), pp. 41-46, 2004.
- [6] 神高敏弘, "なんとなく協調フィルタリング - 順序応答に基づく推薦", 人工知能学会研究会資料, SIG-KBS-A304-37, pp. 1-8, 2004.
- [7] 高山毅, 小熊浩史, 池田哲夫, 三浦良介, "商品購入の前後関係を考慮したレコメンド方式と評価", 電子情報通信学会信学技報, DE2004-5, pp. 25-30, 2004.
- [8] 大杉直樹, 門田暁人, 松本健一, 森崎修司, "ソフトウェア機能の推薦システムのための協調フィルタリング", ソフトウェアシンポジウム2002議事録, pp. 15-24, 2002.
- [9] 杉山一成, 波多野賢治, 吉川正俊, 植村俊亮, "ユーザの負担なく構築したプロフィールに基づく適応的Web情報検索", 電子情報通信学会論文誌, D- No. 11, pp. 975-990, 2004.
- [10] 寺野隆雄, 他, eビジネスの理論と応用, 東京電気大学出版局, 2003.
- [11] インターネット白書2006, 株式会社インプレスR&D, 2006.
- [12] <http://www.johotsusintokei.soumu.go.jp/whitepaper/ja/h18/pdf/i1060000.pdf>, 情報通信白書平成18年度版, 総務省情報通信統計データベース, 2006.
- [13] <http://e-words.jp/>, IT用語辞典e-Words.
- [14] <http://www.atmarkit.co.jp/im/terminology/>, @IT情報マネジメント用語辞典, アイタイムメディア株式会社.