

カスタム価格設定推薦システム — 簡単な実装と予備実験 —

神畷 敏弘[†] 赤穂昭太郎[†] 佐久間 淳^{††,†††}

[†] 産業技術総合研究所 〒 305-8568 茨城県つくば市梅園 1-1-1 産総研つくば中央第 2
^{††} 筑波大学 大学院 システム情報工学研究科 〒 305-8571 茨城県つくば市天王台 1-1-1
^{†††} 科学技術振興機構 〒 102-8666 東京都千代田区四番町 5-3

あらまし 推薦システムは、顧客が好むであろうアイテムを提案する。本研究では、このシステムに、価格の値引きをする新たな機能を加える。この新システムでは、この値引きをするかどうか判断を各顧客ごとに行うが、こうした顧客によって販売価格を変える価格設定手法は価格カスタム化と呼ばれている。このカスタマイズ価格設定推薦システムが、顧客と販売側の両方にもたらすであろう利益について論じる。また、標準的な推薦システムと多腕バンディット手法を組み合わせることで、こうしたシステムを実現する方法を提案する。簡単なシステムを実装し、半人工データ上で予備的な実験を行った結果を報告する。

キーワード 推薦システム, 協調フィルタリング, 価格カスタム化, 多腕バンディット

Customized Pricing Recommender System — Simple Implementation and Preliminary Experiments —

Toshihiro KAMISHIMA[†], Shotaro AKAHO[†], and Jun SAKUMA^{††,†††}

[†] National Institute of Advanced Industrial Science and Technology (AIST)
AIST Tsukuba Central 2, Umezono 1-1-1, Tsukuba, Ibaraki, 305-8568 Japan

^{††} Graduate School of SIE, University of Tsukuba 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577 Japan

^{†††} Japan Science and Technology Agency 5-3, Yonban-cho, Chiyoda-ku, Tokyo, 102-8666 Japan

Abstract Recommender systems suggests items that would be preferred to customers. Here, we propose to add new function, price discounting, to these systems. This new system determines whether it offers discounting for each customer, and thus this price-setting scheme is called price customization. We discuss the benefits that this customized pricing recommender system will bring for both customers and dealers. We propose to realize such systems by combining standard recommendation algorithm and multi-armed bandit approaches. We implemented a simple system and performed preliminary experiments on semi-simulated data.

Key words recommender system, collaborative filtering, price customization, multi-armed bandit

1. はじめに

推薦システム (recommender system) とは、利用者が好むと予測されるアイテムや情報などを、利用者の反応やアイテムの特徴に基づいて見つけ出し、それらを利用者の目的に合わせた形式で提示するシステムである [1]~[3]. GroupLens [4] をはじめ 90 年代中頃以降から多くの手法やシステムが研究レベルで提案され、また、今世紀に入ってから多くの電子商取引サイトで、顧客へのサービスとして幅広く導入されるようになった。

最初に、電子商取引サイトが推薦システムを導入する動機と、顧客がシステムを利用する理由について論じる。マーケティング

ツールは、顧客をグループにわけ、それらの特徴を分析し、それぞれに販売戦略を定めて、販売側の効用を最大化するのが目的である。それに対し、推薦システムは、各顧客のプロファイルを考慮して顧客に対して有用な情報を提供するのが目的であるべきとの考えがある [3]. そして、システムを利用する代償として、販売履歴やデモグラフィックな情報などの個人情報、顧客は販売側に提示する。その一方、販売側は、適切なアイテムを提示し、顧客の要求を満たすことで顧客忠誠度の向上を期待できる。Bergemann と Ozmen は、アイテムに対する顧客の知識の不確実性を減らすことによる付加価値の影響を、ゲーム理論に基づいて論じ、いくつかの市場占有率の均衡点が存在

すると論じた [5].

こうした交換が成立するには、第一に、顧客と販売側の間で交換が対等に行われ、第二に、少なくとも支払うコスト以上の利益を顧客と販売の両方が得られなければならないだろう。これらの点について、まず、顧客と販売側の交換が対等であるかについて考えよう。販売側の基本的な目標は利益の最大化であるが、これは利用者の要求と対立する可能性がある。例えば、顧客が必要とする以上に高価なものを顧客に提示することを、販売側が推薦システムにさせるといった状況である。実際に、顧客の現在の要求に最適なものの代わりに、将来的により多くの利益を販売側にもたらすアイテムを提示して、長期間での利益の最大化をめざすような、推薦システムも研究されている [6]. 推薦システムを運営するコストよりも、多くの追加利益を販売側は少なくとも得なくてはならないため、顧客の要求を最大化する代わりに、自身により多くの利益をもたらす推薦をするインセンティブが販売側には基本的にある。そして、もし自身への利益を最大化はしないこうした推薦をされたなら、顧客は個人情報を提供する動機を失うだろう。そうなれば、推薦システムは顧客の志向パターンを分析できなくなって有用な推薦ができなくなり、結果として、顧客の意志決定を補助できなくなり、推薦システムは役に立たないものになるだろう。

推薦システムを通じた価値交換の成立には、もう一つ、販売側と顧客の両方にとって、推薦システムで得られる利益が、それを利用するコストより大きいことも必要だろう。この点について、販売側と顧客のそれぞれの立場から考えよう。推薦システムの導入で販売側が得る追加利益は、顧客忠誠度の向上によって得られると期待されているが、それはシステムの運用コストを必ずしも超えるとは限らない。一方、個人情報の提供の代わりに、不必要な可能性もあるアイテムの推薦という、完全には信用できない情報を顧客が得ているにすぎない。このように、販売側と顧客の両方にとって、利益がコストを上回っているかは自明とはいえないと、我々は考えている。

そこで、推薦システムにおいて、利用者と販売側との間で価値の新たな交換を行う枠組みを示す。この新たな価値交換について鍵になるのが価格カスタム化である。これは、同じ商品ではあるが、取引や個人ごとにその販売価格を変えるというものである。この価格カスタム化を導入した推薦システムによる、価値交換の新たな枠組みと、その実現について論じる。また、標準的な推薦システムと多腕バンディット手法を組み合わせることで、こうしたシステムを実現する方法を提案する。簡単なシステムを実装し、半人工データ上で予備的な実験を行った結果を報告する。

ここで推薦システムと価格カスタム化を組み合わせる利点を二つ述べておく。第一に、自身が好むものを顧客は購入するので、顧客の嗜好パターンが価格カスタム化に役立つというのは妥当だろう。こうした嗜好パターンは、推薦の過程で利用者の行動履歴から獲得されているので、それを価格カスタム化に導入するのは自然だろう。第二に、アイテムを購入するかどうかを予測するための訓練データが、疎になる問題に対し、推薦システムとの統合は貢献する。長い利用者の行動履歴の中で、利

用者がアイテムを購入する事象は非常に低頻度でしか生じない。こうした事象を予測する問題が困難であることは、クラス不均衡問題 [7] として知られている。推薦システムの予測に基づいて利用者が好まず、購入されないことがほぼ確実なアイテムを除外することにより、このクラス不均衡問題の解消に、推薦システムとの統合は役立つ。

本稿の構成は以下のとおりである。2. 節では、価格カスタム化と、販売側と顧客の価値交換の枠組みについて論じる。カスタム価格設定推薦システムを 3. 節で提案し、実験結果を 4. 節で示す。5. と 6. 節では、関連研究と議論・まとめを述べる。

2. 価格カスタム化

前節で述べたように価格カスタム化 (price customization) [8], [9] とは、顧客ごとに販売価格を変えることであり、動的価格設定 (dynamic pricing) や価格個人化 (price personalization) とも呼ばれている。取引が個人との間で行われるようになった電子商取引においてその実現が容易になり、実際に導入もされている^(注1)。

この価格カスタム化について述べる。顧客が商品を購入する場合に、その商品の価値を測るために心理的に比較する価格のことを参照価格という。この参照価格を中心に、ある下側のしきい値を超えるまでは安くなったと感じることはなく、逆に、ある上側のしきい値を超えるまでは高くなったと感じることはない。Dahana と照井は、これらの価格しきい値を推定するためのモデルを提案している [8]。そして、これらのしきい値が推定できれば、個人ごとに下側のしきい値よりわずかに下の価格を設定すれば、価格を大きく下げることなく販売量を増やし、総売上げを向上させることができる。逆に、個人ごとに上側のしきい値よりわずかに下の価格を設定すれば、販売量を大きく下げることなく販売価格を上げることができるので、やはり総売上げを向上させることができる。

この価格カスタム化は、従来からある価格差別の一種ともいえる。従来の価格差別では、販売地域や顧客の性別・年齢などによってその販売価格を変えていた。例えば、チェーン店のハンバーガーの価格を地域の所得に応じて変えたり、レディース・デイやシニア割引などの追加サービスを提供したりする。この価格差別で問題となるのは転売である。例えば、価格の安い地域で購入し、それを高い地域で販売することで、他の業者が収益を得ると、自身の潜在的な需要を失うことになり、売上げは低下してしまう。そこで、ハンバーガーチェーンの例では、価格差のある地域を離しておくことで、運搬にかかる時間のために商品の価値が低下し転売を困難にするといった工夫をしている。

価格カスタム化では、従来の価格差別とは異なり、主に電子商取引を対象としている。そのため、各顧客ごとに販売数量を管理することができ、多数の商品を転売することが難しい。また、航空機のチケットのように記名式になっていて転売できな

(注1) : CNN.com - Web sites change prices based on customers' habits: <http://edition.cnn.com/2005/LAW/06/24/ramasastry.website.prices/>

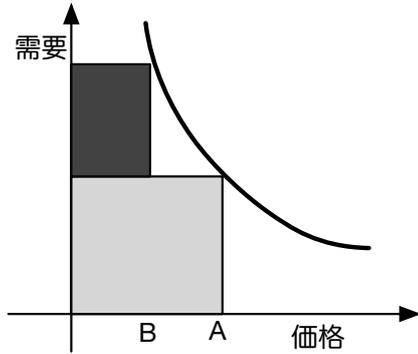


図1 価格カスタム化によって得られる追加利益

いものを対象にしている。同時に、電子商取引では、ランダムに価格を変動させて、その価格での販売に顧客が応じるかどうかをサンプリングしてデータを取得し、価格しきい値の推定に利用することもできるようになる。

2.1 価値交換の新たな枠組み

この価格カスタム化によって、販売側と顧客の間で生じる価値交換について議論する。まず、販売側の立場から考えよう。既存の推薦システムを導入するとき、システムの運用コストに対して、顧客忠誠度の向上による追加利益が得られた。価格カスタム化の導入で得られる追加利益を図1で示す。このグラフは、横軸が販売価格で、縦軸がそのときの需要、すなわち販売数量である。価格を全顧客に対して固定している場合には、価格×需要の四角形の面積が最大になるAの価格に設定すべきである。ここで価格を個人化して、価格Aでは購入しないが、価格Bなら購入する顧客にのみ、価格Bで販売する。すると従来の灰色の四角形の利益に加え、黒い部分の四角形の利益が追加で得られることになる。

次に顧客の立場から考えよう。既存の推薦システムを利用するには、個人情報を提供することコストの代償として、意志決定のための推薦を得ていた。価格カスタム化を導入すると、販売側が価格Aでは購入しないが、価格Bでなら購入するとシステムが判断した場合、利用者は(A - B)だけ商品を安く購入することができる。すなわち、自身の個人情報と交換に、確率的に割引購入をできる権利を得ているといえる。ここで、価格Aで購入した顧客にとっては不公平ではないかという指摘も考えられる。この指摘に対しては、もし顧客が別の取引で割引を提示されることが確率的に保証されているなら、我々はこの種の交換は顧客間で公平だと考えている。

この新たな交換について、顧客と販売側の交換が対等であること、両方にとってコスト以上の利益が得られるか考えてみよう。既存の推薦システムでは、提示するアイテムを変えることだけが販売側の利益向上の手段であったため、顧客の要求を必ずしも最大化しない強いインセンティブがあった。それに対し、割引の提案という新たな手段を提供することで、顧客忠誠度を下げる危険を冒してまで、顧客の意図に沿わない推薦をするインセンティブは減るだろう。さらに、この交換によって、販売側の顧客忠誠度の向上や、顧客側の意志決定のための情報

表1 各型の顧客の応答に対する報酬

応答	顧客		
	定価	割引	不買
買	α	β	0
不買	0	0	γ

といった間接的な利益ではなく、売上げの増加や、割引販売といった直接的な利益になっている。よって、互いに直接的で明確な追加利益が得られているため、従来の推薦システムより、互いのコストに見合った交換をしていると考えている。

3. 価格カスタム化推薦システム

この節では、この価格カスタム化を導入したカスタム価格設定推薦システム(CPRS; Customized Pricing Recommender System)について述べる。

CPRSには能動型と受動型が考えられる。能動型CPRSでは、顧客に提示するアイテム自体も、システムが選択するものである。顧客が非常に好むと予測されるが、価格面で選択しないと判断されるアイテムなどを積極的に提示するなど、多様な提案が考えられる。受動型のCPRSでは、閲覧など顧客がアクセス中のアイテムに対して、そのアイテムを顧客が好むと予測される場合に起動され、必要に応じて割引を提案する。能動型CPRSには多様な発展が考えられるが、その一方で実現にあたって解決すべき課題も多い。そこで、まず第一段階として、受動型CPRSについて検討する。

価格カスタム化を導入した推薦システムは、我々の知る限りは、まだ提案されていないので、最も簡潔な場合を考える。まず、全てのアイテムの価格は同一とする。音楽のダウンロード販売などの状況には当てはまるだろう。さらに問題を簡単にするため、価格は2段階で、定価と割引価格で得られる利益を、それぞれ α と β とする。なお、 $\beta < \alpha$ であり、割引価格は参照価格の下側のしきい値より低く設定されているものとする。

顧客を次の三種類に分類する：

- (1) **定価** (standard)：定価でも購入する。
- (2) **割引** (discount)：定価では購入しないが、割引があれば購入する。
- (3) **不買** (indifferent)：最終消費者として購入の意志がない。

このうち、定価顧客には定価を、割引顧客には割引価格を、不買顧客には一部は転売目的の場合もあるため購入されにくいように定価を提示すればよい。こうした行動をとったときに、定価顧客と割引顧客に対してはその予測が正解であったときに、それぞれ α と β の報酬を得る。不買顧客の場合には、潜在需要と新たな顧客との接触の機会を失うので、売れたときに損失となる。だが、負の報酬は扱いにくいので、売れたときの報酬を0、売れなかった場合の報酬を γ としておく。これらの報酬をまとめると、表1ようになる。

3.1 2段階分類と多腕バンディット

顧客が三つの型のうちどれであるか分かっているならば、定価・不買顧客には定価を、割引顧客には割引価格を提示しておけば、

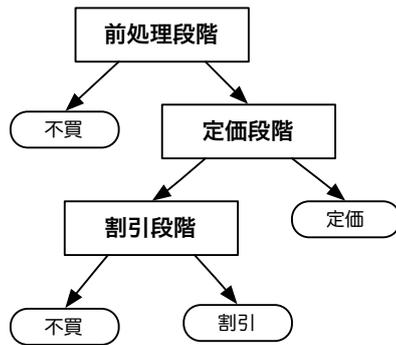


図2 顧客分類の過程

得られる報酬は最大化される。しかし、顧客がどの型であるかは予測しなければならない。

この予測のために、各型の顧客の行動を整理する。定価顧客は定価と割引価格のどちらを提示されても購入するが、不買顧客はどちらでも購入しない。割引顧客は定価を提示されても購入しないが、割引価格を提示されたら購入する。

顧客の行動規則に基づき、顧客を図2に示す手順で分類する。ある顧客がある商品を開覧したとき、その顧客のその商品に対する顧客タイプを分類することが目的であるので、対象アイテムが変わると、同じ顧客が不買タイプであったり、定価タイプであったりすることに注意されたい。

前処理段階では、明らかな不買顧客を除外する。この段階は、クラス間の事例数の差が非常に大きい場合、低頻度のクラスの事例への予測精度が低下するクラス不均衡問題 (class imbalance problem) [7] に対処するために導入した。顧客の分類では、不買顧客が定価・割引顧客よりずっと多い場合、この問題への対処は重要である。そこで、対象商品への嗜好スコアが非常に低い場合に、好きではないアイテムを購入することはないとの仮定の下、それらの顧客を不買顧客として除去する。この嗜好スコアは、標準的な推薦システムにより推定できる。

残る定価段階と割引段階の2段階が主たる段階である。第1段階の定価段階の分類器は、定価顧客とそれ以外を区別する。この分類器は、定価を顧客に提示したときに、そのアイテムを購入した場合を正例、それ以外の場合を負例とした訓練事例集合から獲得する。

割引段階の分類器は、定価顧客ではない定価段階で分類された場合に起動され、割引顧客と不買顧客を分類する。この分類器は、定価分類器に定価顧客でないかと判断され、割引価格を提示したときに、そのアイテムを購入した場合を正例、それ以外の場合を負例とした訓練事例集合から獲得する。

さらに、もう一つの問題がある。割引価格を提示していれば、定価顧客も購入するため、定価顧客を割引顧客に誤分類していれば潜在的に $\alpha - \beta$ の損失を出していることになる。そこで、割引顧客と分類されていても、時々定価顧客である可能性を考慮して定価を提示して、顧客から応答を得てデータを収集する必要がある。その一方で、最適と予測される以外の行動を増やしすぎると、得られる報酬は減ってしまう。同様に、定価顧

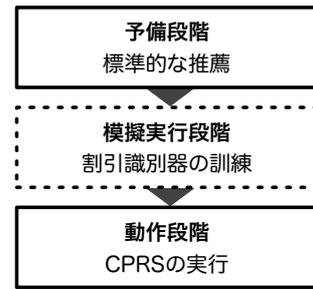


図3 CPRSの実行段階

客と誤分類した割引顧客に定価を提示しても購入されないの、定価顧客と分類されていても、誤分類の可能性を考慮して割引価格を提示する必要がある。ところが、これを行いつぎると、やはり潜在的な損失が大きくなってしまふ。このように、予測に基づく行動とデータ収集のバランスをとる必要があるが、これは 利用-探索 (exploit-explore) のトレードオフとして知られる。

このトレードオフを調整を最適化する問題を多腕バンディット (multi-armed bandit) 問題 [10], [11] という。この多腕バンディット問題では、本来は未知の情報を利用できたとして、各回に最も最適な行動をとった場合に得られる報酬の総量とくらべて、どれだけ実際にとった行動による報酬の総量が少なかったかというリグレットと呼ばれる量の最小化を試みる。softmax 行動選択, UCB1, Gittins 尺度などの基準が知られている。

4. 実験

最初の段階として、簡単なシステムを実装し、半人工データで実験した結果を報告する。

4.1 タスクの定式化とその実装

このシステムには図3に示すような三つの実行段階がある。予備段階では、標準的な推薦システムがまだ運用されている段階で、割引価格を提示する価格カスタム化は行われぬ。この段階で、推薦システムを訓練するための嗜好データと、定価を提示したときに顧客が購入するかどうかを決める分類器を訓練する購入履歴を収集する。この嗜好データを用いて、2種類のモデルベースの推薦システム (付録1. の pLSA と行列分解) を訓練した。このモデルは、図2の前処理段階で明らかな不買顧客を除外するために利用するのに加え、これらのモデルパラメータを定価・割引分類器のための特徴量としても利用する。なお、以降の模擬実行や動作段階では推薦モデルの更新は行わなかった。

購買履歴からは、定価顧客を特定するための初期的な分類器を訓練する。単純であり、またオンラインでの訓練が容易なため、ここでは単純ベイズ分類器を使った。システムを利用している顧客と、閲覧中のアイテムの対に対する特徴ベクトルが必要になる。これには、顧客のデモグラフィックな特徴に加え、これらのアイテムや顧客に関連した推薦モデルのパラメータ群と、推定された嗜好スコアを利用した。例えば、付録1. の行列分解モデルでは、目的アイテムのバイアス項などを用いた。顧

客やアイテムの嗜好情報はこれらのパラメータに抽出されており、また嗜好パターンはアイテムの購入決定と密接な関連があると考えられるので、これらのパラメータなどを対象の顧客-アイテムペアの分類に採用するのは妥当だろう。

次の、図3の模擬実行段階は実行されない場合もある。CPRSの動作段階の初めでは、図2の割引分類器は全く訓練されていない。なぜなら、一度も割引価格を顧客に提示したことがないので、訓練データが全くないからである。そこで、ある一定期間、システムは連続的に割引価格を、前処理で除外されなかった全ての場合で提示し、割引分類器を訓練するための事例を収集する。なお、割引分類器も、単純ベイズで、定価分類器と同じ特徴量を利用した。

最後の動作段階は、CPRSが本格的に実行される段階である。対象の顧客とアイテムを図2の手順で分類した結果に加え、探索-活用のトレードオフも考慮して、定価と割引価格を適宜選択して顧客に提示する。定価・割引分類器は、顧客のフィードバック情報に基づいて更新される。すなわち、定価を提示したときには定価分類器を更新し、割引価格を提示したときには割引分類器を更新する。

4.2 実験結果

上記の簡潔なシステムをテストするために、Movielensの100万データ集合^(注2)から半人工のデータを生成した。顧客がアイテムを買うかどうかの判断は、アイテムへの嗜好と、顧客のデモグラフィックな特徴に密接に関係しているだろう。そこで、目標アイテムに最高の5の評価を付けていて、その年齢が45歳以上の顧客はアイテムを購入するとした。これらの顧客のうち男性を定価顧客、女性を割引顧客とした。100万データのうち半分を予備段階で使い、pLSAと行列分解の2種類の推薦モデルと、初期の定価分類器を訓練した。これらのモデルを訓練するときは、顧客の購入手動に使った5の評価を隠すため、元の評価が4と5のデータを両方とも評価4に変えた。

残りの半分のデータは模擬実行段階と動作段階で利用した。上記の基準で定価と割引顧客を選択したところ、それぞれこの後半データの3.5%と1.4%が該当した。効用 α 、 β 、および γ は、それぞれ1.0、0.5、および0.01に設定した。標準と割引分類器の訓練にはクラス不均衡問題が伴うため、単純ベイズ分類器の正・負の判定のしきい値を通常の0.5ではなく、前処理段階で訓練した初期定価分類器のF尺度を最大化するしきい値に設定した。仮に全ての顧客のタイプが既知であったとき、後半データから得られる報酬の合計は25468で、この値が理論上の報酬の上限であり、リグレットはこの値から実際に得られた報酬を引いたものとなる。システムが常に定価と割引価格を提示した場合に得られる報酬の合計は、それぞれ22028と12074になる。常に定価を提示したときの合計報酬22028を、価格カスタム化によって上回るのがベースラインの目標となる。

探索-活用のトレードオフを扱う最も単純な方法をここでは採用した。一つ目は模擬実行段階を使うアプローチである。最初の5000評価値を模擬実行段階として、割引分類器を訓練し、

その後の動作段階では図2の手順で予測された顧客タイプに対して最良の行動をとった。このアプローチでは、行列分解とpLSAのそれぞれのモデルに対し、それぞれの合計報酬は21982と20857となった。これらはどちらもベースラインの結果より悪く、模擬実行時間を長くするとさらに合計報酬は下がった。

二つ目は、模擬実行段階を行わず、予測された顧客タイプに対して最良でない行動を1%の確率でとる、いわゆる ϵ ランダムアプローチである。このアプローチでは、行列分解とpLSAのそれぞれのモデルに対し、それぞれの合計報酬は21925と20843となった。合計報酬はやはりベースラインより下回る結果となった。なお、非最適行動をとる確率を上げるほど、合計報酬は低下した。

今回はどちらの実験結果も優れないものだったが、その理由と今後の方針について考察する。合計報酬がベースラインより低下する理由は、割引分類器を訓練するために割引価格を提示することによる報酬の低下を、割引分類器の予測精度の向上で補えないためである。この予測精度が向上しない理由は、割引価格でのみ購入される事象が生じる確率が小さく、クラス不均衡問題が原因となり高精度の予測ができていないため、まずこの問題に対処する必要がある。さらに、割引顧客である可能性の低い顧客に対して割引価格を提示し続けると探索のためのコストが増大してしまう。割引分類器の予測精度が低い段階から、割引顧客である可能性が高い顧客を重点的に探索する手法も必要だろう。

合計報酬の減少を回避するには、多くの報酬をもたらす定価顧客が重要である。よって、定価顧客を逃さないようにするため、定価分類器には高い再現率が要求される。一方、割引分類器には高い精度が要求される。極論すれば、全ての定価顧客が定価分類器で検出されていれば、一人でも確実に割引顧客を検出することで、報酬の増加が見込める。このような分類器の性質を考慮にいったバンディットの戦略の開発すべきだろう。

今後は、これらの問題に対処したアルゴリズムの開発を進めたい。

5. 関連研究

多腕バンディットアルゴリズムは幾つかの推薦システムで利用されている。推薦タスクともみなせるWebコンテンツ最適化はバンディットの手法が適用されている[12]~[14]。Webコンテンツ最適化は、バナー広告や記事などをWebページに表示するとき、閲覧者にできるだけクリックされるようなものを選び出す問題である。クリックされやすいかどうかを調べる探索と、現在分かっている確率分布の下で推定クリック率が最高のコンテンツを表示する活用とのトレードオフがあるため、多腕バンディット問題として扱われている。閲覧者からのフィードバックに遅延があるとか、バナー広告の最低表示回数制限などの制約があるため、通常多腕バンディットアルゴリズムにいろいろな工夫が加えられている。Wengら[15]は、ネットワーク上に分散して独立に稼働している推薦システムサイト群を使って推薦する方法を提案した。サイト群の中から、現在の

(注2) : GroupLens research lab: <http://www.groupLens.org/>

利用者にとってよい推薦をするサイトを見つけ出す問題を多腕バンディット問題として定式化している。多腕バンディット問題の一般化である強化学習を, Shani らは推薦システムに適用した [6]。顧客から得る報酬を, 各トランザクションごとではなく, 長期間にわたる報酬の合計を最大化するよう設計されている。どの商品を顧客に提示するかを政策と, システムの行動に対する顧客の反応に応じて報酬や状態遷移が決まるようなマルコフ決定過程を考え, 強化学習の問題として解いた。

Baydar は, 食料品店について, 顧客の過去の行動に応じて価格カスタム化をした場合の損益などをエージェントシミュレーションによって求める環境を提案した [16]。

6. 議論とまとめ

本研究では, 価格カスタム化の機能を備えた価格カスタム化推薦システム (CPRS) を提案した。この CPRS によって生じる顧客と販売側の新たな価値の交換について議論し, 簡単なシステムの実装と予備実験を行った。その結果, CPRS における多腕バンディット問題では, クラス不均衡の問題を重視すべきことや, 顧客の分類の段階に応じて, 再現率と精度のどちらを重視すべきかといった課題を示した。

今回の CPRS は, 商品価格や割引率が固定されている制限があったが, 報酬の設計や, 多腕バンディットアルゴリズムの改良によりこうした制限は取り除いてゆく必要がある。例えば, 回帰問題を扱うバンディットアルゴリズムが適用できれば, 価格自体の自動設定もできるだろう。顧客間の公平性を保証するために, 各顧客ごとに割引を提案する最低回数などの制限の導入や, 割引の有無だけでなくアイテム自体の選択も行う能動的 CPRS の開発も将来的な課題である。CPRS では, 顧客から得られる利益を報酬と考えて効用を設計したが, 効用関数の設計によっては, この枠組みの適用範囲はさらに広がるだろう。たとえば, 在庫整理をしたいときには, 在庫量も効用関数に組み込むことにより, より大幅な割引で在庫整理を行うことなども考えられる。顧客に対して単に価格面で割引をするだけでなく, セット販売, ポイントサービス, 輸送サービスなど他の要因を考慮した効用の設計も可能だろう。こうしたことが実現できれば, 推薦システムは単にアイテムを利用者に提示するだけでなく, 実際の店員のように, より洗練された提案を顧客に対して行うことができるようになるだろう。このようにより多様な提案のできる **おもてなしシステム (attendant system)** と呼ぶべきものに CPRS を進化させてゆきたいと考えている。

謝辞 本研究は科研費 21500154 の助成を受けた。

文 献

- [1] 神鳥: “推薦システムのアルゴリズム (1)~(3)”, 人工知能学会誌, Vol. 22, No. 6 ~ Vol. 23, No. 2 (2007-2008).
- [2] P. Resnick and H. R. Varian: “Recommender systems”, *Communications of The ACM*, **40**, 3, pp. 56-58 (1997).
- [3] J. Ben Schafer, J. A. Konstan and J. Riedl: “E-commerce recommendation applications”, *Data Mining and Knowledge Discovery*, **5**, pp. 115-153 (2001).
- [4] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl: “GroupLens: An open architecture for collaborative filtering of Netnews”, *Proc. of The Conf. on Computer*

Supported Cooperative Work, pp. 175-186 (1994).

- [5] D. Bergemann and D. Ozmen: “Optimal pricing with recommender system”, *ACM Conference on Electronic Commerce*, pp. 43-51 (2006).
- [6] G. Shani, D. Heckerman and R. I. Brafman: “An mdp-based recommender system”, *Journal of Machine Learning Research*, **6**, pp. 1265-1295 (2005).
- [7] N. Japkowicz: “Learning from imbalanced data sets: A comparison of various strategies”, *AAAI Workshop: Learning from Imbalanced Data Sets*, pp. 10-15 (2000).
- [8] N. Terui and W. D. Dahana: “Price customization using price thresholds estimated from scanner panel data”, *Journal of Interactive Marketing*, **20**, pp. 58-70 (2006).
- [9] 照井: “価格閾値の推定と価格カスタマイゼーションの可能性”, *日本統計学会誌*, **37**, 2, pp. 261-277 (2008).
- [10] R. S. Sutton and A. G. Barto: “強化学習”, 森北出版 (2000). (三上 貞芳, 皆川 雅章 訳)
- [11] P. Auer, N. Cesa-Bianchi and P. Fischer: “Finite-time analysis of the multiarmed bandit problem”, *Machine Learning*, **47**, pp. 235-256 (2002).
- [12] N. Abe and A. Nakamura: “Learning to optimally schedule internet banner advertisements”, *Proc. of The 16th Int'l Conf. on Machine Learning*, pp. 12-21 (1999).
- [13] D. Agarwal, B.-C. Chen and P. Elango: “Explore/exploit schemes for web content optimization”, *Proc. of The 9th IEEE Int'l Conf. on Data Mining*, pp. 1-10 (2009).
- [14] L. Li, W. Chu, J. Langford and R. E. Shapire: “A contextual-bandit approach to personalized news article recommendation”, *Proc. of The 19th Int'l Conf. on World Wide Web*, pp. 661-670 (2010).
- [15] L.-T. Weng, Y. Xu, Y. Li and R. Nayak: “Towards information enrichment through recommendation sharing”, *Data mining and Multi-agent Integration* (Ed. by L. Cao), Springer, chapter 7, pp. 103-125 (2009).
- [16] C. Baydar: “Agent-based modeling and simulation of store performance for personalized pricing”, *proc. of the Winter Simulation Conf.* (2003).
- [17] T. Hofmann and J. Puzicha: “Latent class models for collaborative filtering”, *Proc. of the 16th Int'l Joint Conf. on Artificial Intelligence*, pp. 688-693 (1999).
- [18] A. Das, M. Datar, A. Garg and S. Rajaram: “Google news personalization: Scalable online collaborative filtering”, *Proc. of The 16th Int'l Conf. on World Wide Web*, pp. 271-280 (2007).
- [19] R. M. Bell and Y. Koren: “Scalable collaborative filtering with jointly derived neighborhood interpolation weights”, *Proc. of The 7th IEEE Int'l Conf. on Data Mining*, pp. 43-52 (2007).
- [20] Y. Koren: “Collaborative filtering with temporal dynamics”, *Proc. of The 15th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 447-455 (2009).

付 録

1. モデルベース推薦システム

協調フィルタリング (collaborative filtering; CF) は, ある特定の利用者 (活動利用者; active user) の好みを, 他の利用者 (標本利用者; sample user) から集めた嗜好データに基づいて予測する。まず, 採点法で集めた評価スコアを用いた標準的な CF タスクを定式化する。利用者 x とアイテム y を, それぞれ $x \in \{1, \dots, n\}$ と $y \in \{1, \dots, m\}$ で示す。利用者 x がアイテム y に対して付けたスコアを s_{xy} と記す。このスコアは, 5段階評価と評価尺度中の一つの値をとり, 好みの度合いを表す。訓練集合はタプル集合 $\mathcal{D} = \{(x_k, y_k, s_{x_k y_k})\}$, $k = 1, \dots, N$ で

ある。任意の利用者 x と任意のアイテム y の対について、嗜好スコアを予測する関数 $\hat{s}_{xy} = f(x, y)$ を、この訓練集合 \mathcal{D} から学習することが目標である。利用者 x が評価したアイテムの集合を $\mathcal{Y}_x = \{y | (x', y, s) \in \mathcal{D} \text{ s.t. } x' = x\}$ としておく。

本稿では pLSA と 行列分解 の二つのモデルを用いた。一つ目は、語と文書の簡約表現を得る方法として提案された pLSA で、CF タスクにも適用され [17]、Google News のような商用システムでも採用されている [18]。実数値のスコアを扱えるように、[17] の利用者-アイテム-スコアの 3 方向モデルを若干変更した。形式的に、 z は離散潜在確率変数で、 $\{1, \dots, K\}$ の値をとり、ベルヌーイ分布の K 値拡張であるカテゴリ分布に従う。利用者 x とアイテム y の離散確率変数を、それぞれ x と y する。どちらも、 z が与えられたときの条件付きでカテゴリ分布に従う。加えて、スコア用の実数確率変数 s は、 z が与えられたとき条件付きで正規分布に従う。まとめると、訓練集合 \mathcal{D} に対する対数尤度関数は次式：

$$\mathcal{L}(\mathcal{D}; \Theta) = \sum_{(x, y, s) \in \mathcal{D}} \log \sum_z \Pr[z] \Pr[x|z] \Pr[y|z] \mathcal{N}(s; \mu_z, \sigma_z^2)$$

この尤度を最大化するパラメータは EM アルゴリズムで容易に計算できる。モデルパラメータが学習できれば、利用者 x によるアイテム y への嗜好スコアは、次の条件付き期待値で計算できる。

$$\hat{s}_{xy} = E[s|x, y] = \frac{\sum_z \mu_z \Pr[z] \Pr[x|z] \Pr[y|z] \mathcal{N}(s; 0, \sigma_z^2)}{\sum_z \Pr[z] \Pr[x|z] \Pr[y|z]}$$

本論文の実験では、 $K = 10$ とし、離散変数の事前分布のパラメータを 0.1 とした。

ここではさらに、採点法で取捨されたスコアのさまざまなバイアスを取り除くための [19] の技法を導入した。まず、 \mathcal{D} 中の全スコア上の平均である大域効果 b を取り除くため、各スコアからこれを引いて修正スコア $s'_{xy} = s_{xy} - b$ を得る。各修正スコアから、アイテム y に対する全修正スコアの平均 d_y を引いて、 $s''_{xy} = s'_{xy} - d_y$ を得る。さらに、利用者 x に対する、全てのこの修正スコア s''_{xy} の平均 c_x を、各修正スコアから引いて $s'''_{xy} = s''_{xy} - c_x$ を得る。訓練集合 \mathcal{D} 中のスコアをこれらの修正スコア s'''_{xy} で置換し、上記の pLSA モデルを訓練した。修正スコアの予測値 \hat{s}'''_{xy} は、修正バイアスを加えて元のスコアの推定値を得る $\hat{s}_{xy} = \hat{s}'''_{xy} + b + c_x + d_y$ 。

二つ目のモデルベース法は [20] の式 (4) の行列分解を用いた方法である。このモデルでは嗜好スコアは次式で計算する。

$$\hat{s}_{xy} = b + c_x + d_y + u_y^\top \left[v_x + \frac{1}{\sqrt{|\mathcal{Y}_x|}} \sum_{y' \in \mathcal{Y}_x} w_{y'} \right] \quad (\text{A.1})$$

ただし、 b , c_x , および d_y は大域、利用者ごと、アイテムごとのバイアスを取り除くためのパラメータ。 u_y , v_x , および $w_{y'}$ は K 次元のパラメータベクトル。本稿の実験では $K = 5$ とした。 u_y と v_x の内積は、アイテムと利用者の交差的な影響を表し、 $w_{y'}$ は各利用者がどのアイテムを評価したのかという情報を考慮する目的で導入している。これらのパラメータは次の損失関数を最小化することで最適化する。

$$\text{loss}(\mathcal{D}; \Theta) = \sum_{(x, y, s) \in \mathcal{D}} (s_{xy} - \hat{s}_{xy})^2 + \lambda R$$

ただし、 R は大域バイアス以外のパラメータの L_2 正則化項を足したもので、 λ は正則化パラメータである。パラメータが求められれば、任意の利用者と任意のアイテムの対に対する嗜好スコアは (A.1) 式で推定できる。