

価格個人化推薦システム

神畷 敏弘[†] 赤穂昭太郎[†]

[†] 産業技術総合研究所, <http://www.kamishima.net/jp/>, E-mail: mail@kamishima.net

あらまし 既存の推薦システムは、顧客の行動履歴に基づき、顧客が好む商品を見つ出す。これに加え、値引き提案というより積極的な対応ができるような拡張を提案する。

キーワード 価格個人化, 推薦システム, 協調フィルタリング, 多腕バンディット

Personalized Pricing Recommender System

Toshihiro KAMISHIMA[†] and Shotaro AKAHO[†]

[†] National Institute of Advanced Industrial Science and Technology (AIST), <http://www.kamishima.net/>

Abstract Recommender systems try to find items that customers would prefer by exploiting the customers' behavioral records. In this paper, we extend such recommender systems so that they can offer discounting as well.

Key words personalized pricing, recommender system, collaborative filtering, multi-armed bandit

1. はじめに

推薦システムとは、顧客が好むであろう商品や情報などを、顧客との対話履歴や商品の特徵に基づいて予測し、それらを顧客の目的に合わせて提示する [1]。90 年代中頃以降から多くの手法が研究レベルで提案され、今世紀に入ってから顧客へのサービスとして、多くの電子商取引サイトで導入されている。

このような成功をおさめてきた推薦システムではあるが、基本的には顧客にアイテムを提示するだけで、実店舗での店員のような高度な提案はできない。例えば、服飾店でコーディネートについて提案したり、自動車販売で車のオプションについて相談したりといったことは、現状ではできない。

ここでは、推薦以外の行動として**価格個人化** (price personalization) [2] に注目する。価格個人化とは、実店舗の店員の価格交渉のように、顧客ごとに価格を調整することである。ここでは、顧客が割引価格の場合にのみ購入すると予測された場合に限り割引価格を提示するシステムについて論じる。

本論文では、価格個人化機能を持つ推薦システムである**価格個人化推薦システム** (personalized pricing recommender system; PPRS) [3], [4] について論じる。このシステムは既存の推薦アルゴリズムにより顧客の嗜好パターンを抽出し、そのパターンと顧客の購買履歴を用いて割り引きを提示するかを決定する。[3] の結果ではベースライン以下の性能であったが、今回はベースラインを上回る性能を達成したので、これを報告する。

2. 節は価格個人化とその利点, 3. 節では PPRS の課題と実装, 4. 節では半人工データでの実験結果, そして 5. 節ではまとめを述べる。

2. 価格個人化

価格個人化 (Price personalization) とは、顧客や取引ごとに商品の価格を調整することであり [2], dynamic pricing や price customization とも呼ばれる。価格を下げると、販売数は一般に増加する。よって、一定の定価で販売するときには、価格と販売数の積を最大にする価格を定価とすることで、利益を最大化できる。ここで、定価では買わないが、割引価格なら買う顧客のみに対し割引をすることで、追加の利益を得ることが価格個人化の目的である。

この価格個人化は、販売地域や顧客の性別・年齢などによってその商品価格を変える価格差別の一種である。例えば、チェーン店が、ハンバーガーの価格を地域の物価に応じて変えたりするのが価格差別である。この価格差別で問題となるのは転売で、安い地域で購入したものを高い地域で転売されると、自身の潜在的な需要を失うことになる。そこで、ハンバーガーの例では、価格差のある地域を離しておくことで、運搬すると商品の価値が下がり転売でないようにしているといった工夫をしている。

こうした従来の価格差別とは異なり、価格個人化では電子商取引が主な対象である。これは、各顧客ごとの販売数量の管理が容易だったり、航空機チケットのように記名式だったりするため転売が困難となるためである。そして、顧客セグメントではなく個人ごとに価格を変えることや、価格に対する反応の調査が容易なためでもある。

電子商取引では、価格個人化はすでに導入されているが、問題を抱えている [5], [6]。その原因は、顧客に対して不誠実であったこと、割引を提供する状況の問題と考えている。これら

の事例では、価格個人化の実施を知らせていないといった不誠実な対応があった。一方で、その実施を知らせ、常に定価からの値引きだけをするような、誠実な価格個人化であれば、実店舗では受け入れられている。例えば、優良顧客には値引き、クーポン、おまけを提示されていることや、自動車販売店が状況に合わせた価格を提示していることを、他の顧客も受け入れられている。こうした対応を電子商取引でも行う必要があると考える。さらに、各商品を顧客が最初に参照したときのみ割引の機会を提供し、その後はその商品に対しては定価販売をするようにする。これにより、割引が提示されるまで購入を延期する *delaying purchasing* の問題 [7] を回避でき、さらに *price grabber* や *価格.com* (注1) の影響も回避できる。

推薦システムへ価格個人化を導入する利点について述べる前に、推薦システムの商業的継続可能性について論じる。商業的に継続可能であるためには、その運用コストを上回る利益が得られる必要がある。しかし、この条件は現状では満たされていないと考える。現在は、顧客にとって有用な情報をシステムが提供することで顧客忠誠度を向上させ、その優良顧客から販売者は追加利益を得る一方で、その対価として、販売者は運用コストを負担している。しかし、顧客忠誠度の向上の効果は間接的で不確実なため、コストに対して得られる利益が十分とはいえない場合も生じる。もしここで、価格個人化によって追加利益を得ることができれば、運用コストを補うだけの利益を得られ、システムの商業的継続可能性は高まるだろう。

推薦システムが商業的に継続可能でなければ、顧客にとって最善の推薦をしないインセンティブが生じてしまうので、継続可能性は顧客のシステムへの信頼にとって重要な要因であると考え。販売者が運営する推薦システムには、顧客の要求の満足度を最大化するのではなく、より多くの利益を販売者にもたらすインセンティブが短期的には存在する。実際に、顧客の満足よりも、他の効用を考慮する推薦システムも研究されている [8]。推薦システムが商業的に継続可能な状態でなければ、より一層の利益を得る必要があるため、このインセンティブはより強くなるだろう。しかし、価格個人化によって継続可能性が維持されるようになるのであれば、顧客の長期の忠誠度を下げるリスクを負っても、こうした不誠実な推薦をする販売者のインセンティブは減るだろう。このように顧客は価格個人化の導入によって、より信頼できる推薦を受けられる可能性が高まると考える。さらに、割引を提案される機会をも得ることもできる。

3. 価格個人化推薦システム

この節では、価格個人化の機能を導入した価格個人化推薦システムについて述べる。

3.1 価格個人化推薦システムの定式化

顧客が商品を参照する度に、適切な価格を提示し、顧客の行動に応じて得られる累積報酬を最大化することが PPRS の目標である。本論文では、最初の試みとして最も単純な PPRS を考える。ある顧客が現在見ている商品を対象に受動的に PPRS

は呼び出され、その顧客が割引を提示された場合に限り購入すると予測したならばシステムは割引を提示する。なお、割引をすれば購入される商品を能動的に選ぶことも考えられるが、これは今後の課題としたい。音楽ダウンロードや *pay-per-view* ビデオにみられるように、全ての商品の価格は同一と仮定する。さらに、価格は定価と割引価格の2段階のみを想定する。

この仮定の下で、定価か割引価格かを提示されたあと、顧客は商品を購入するかどうかを決定する。ある商品についての購買行動に基づいて、顧客を次の3種類に分類する。

- (1) **定価** (*standard*): 定価と割引価格のどちらを提示されるかに関わらず購入する顧客
- (2) **割引** (*discount*): 定価では購入せず、割引価格を提示された場合にのみ購入する顧客
- (3) **不買** (*indifferent*): 定価と割引価格のどちらを提示されても購入しない顧客

なお、これらの振る舞いは最終消費者としてのものである。もし自身で消費せず、割引価格を提示された場合に、商品を定価で転売するために、不買顧客が購入する可能性はあるとする。これらの最終消費者としての反応をまとめると次のようになる。

提示価格 \ 顧客型	定価	割引	不買
定価	購入	不買	不買
割引	購入	購入	不買

次に、顧客からの応答に応じてシステムが得る報酬 (*reward*) について考察する。システムは、対象商品に対する顧客の型を予測し、その予測に応じて価格を決定する。定価と割引価格で、商品が売れたときの利益は、それぞれ α と β ($\alpha > \beta$) であるとする。定価顧客に、割引をすると潜在的に $\alpha - \beta$ の損失を生じるので、定価顧客には定価を提示すべきである。そして、定価顧客が購入すれば、報酬 α が得られ、そうでなければ報酬は 0 となる。次に、割引顧客は割引がなければ購入しないので、この場合は割引価格を提示し、購入すれば報酬 β が得られ、そうでなければ報酬は 0 となる。最後に、不買顧客にはシステムは定価を提示する。これは、もし割引価格を提示すると、最終消費者としてではなく、転売目的で購入する可能性が高まるためである。顧客が購入しなければ報酬は 0 だが、購入すると転売の可能性が生じるため潜在的な損失、すなわち負の報酬が生じる。しかし、負の報酬の扱いは技術的に不便な場合が多いため、両方の報酬に定数を加え、購入したときには報酬 0 を得て、そうでなければ報酬 γ を得るとしておく。この潜在的に得る報酬は非常に小さなものとする、すなわち $\alpha, \beta \gg \gamma$ 。以上の報酬をまとめると次のようになる。

応答 \ 顧客型	定価	割引	不買
購入	α	β	0
不買	0	0	γ

3.2 PPRS の三つの設計上の課題

PPRS の三つの設計上の課題、観測の曖昧性、クラス不均衡、および活用-探索トレードオフについて順に述べる。

3.2.1 観測の曖昧性

観測の曖昧性とは、顧客の型の識別不能性にとまなう問題で

(注1) : <http://www.pricegrabber.com> と <http://kakaku.com>

ある。真の顧客の型は観測できないので、提示した価格に対する顧客の応答から推定する必要がある。ここで、システムが提示できるのは定価か割引価格かのいずれかである。定価を提示すると、定価顧客は購入するが、割引顧客と不買顧客は購入しないので、このときの応答からは割引顧客と不買顧客は区別できない。同様に、割引価格を提示したときには、定価顧客と割引顧客は区別できなくなる。このように、顧客の応答から完全には顧客の型を識別できない識別不能性の問題がある。本論文では、この問題に対し、多段階分類アプローチで対応する。

3.2.2 クラス不均衡

分類問題で、クラスの頻度に大きな偏りがあるとき予測精度が低下することはクラス不均衡問題として知られている [9]。これは、一方のクラスの事例が、他方と比べて非常に少ないとき、少数派クラスの事例が多数派クラスに分類されやすくなるという問題である。定価顧客や割引顧客と比べて、不買顧客は圧倒的に多く、クラス不均衡の状態にあるため、定価顧客や割引顧客が高い頻度で不買顧客に誤分類されてしまう。本論文では、この問題に対し、予備選別とクラス重み付けで対応する。

3.2.3 活用-探索トレードオフ

活用-探索トレードオフの問題は訓練データの収集時に生じる。割引顧客と推定した顧客が実は定価顧客であるとき、割引価格を提示すると購入する。これは予期された応答なので、システムは同様の状況のとき今後も誤識別を続けてしまい、潜在的に $\alpha - \beta$ の損失を出し続けることになる。よって、現在の予測が本当に正しいかを調べるため、予測した顧客型に対して最適ではない価格も提示して、訓練データを収集する必要がある。逆に、こうした非最適行動をあまりに頻繁に行うと、得られる総報酬は減ることとなる。この問題は活用-探索のトレードオフと呼ばれ、最適行動とデータの収集との釣り合いをうまく調節しなくてはならない。

この活用-探索トレードオフを最適化する問題は、多腕バンディット問題と呼ばれる [10],[11]。アームとも呼ばれる行動の候補集合から、一つの行動を選び実行し、その行動に対する報酬を得ることを繰り返し行うとする。このとき、もし各行動に対して得られる報酬が完全に分かっていたら、候補行動の中で最も報酬が高い行動を選択し続けられればよい。しかし、報酬は未知で予測する必要がある。そのため、報酬を受け取る度に、システムは報酬の予測モデルを更新し、その更新したモデルで次回に選択する行動を、活用-探索トレードオフを考慮して決定する。報酬が既知のときに獲得可能な総報酬の最大値から、ある時刻までに、この手続きで得られる総報酬を引いた値であるリグレットを最小化するのが多腕バンディット問題の目標である。

3.3 PPRS の実装

本節では PPRS の実装について述べる。PPRS は、アイテムの嗜好データと顧客の購入履歴の 2 種類のデータを利用する。嗜好データから推薦モデルを獲得し、このモデルと購入履歴を訓練データとして、顧客型を予測する分類器を学習する。PPRS の導入以前では、販売者はすでに、顧客に定価販売しつつ、既存の推薦システムを運用しているとする。このため、PPRS の開始時から、初期嗜好データを利用でき、また常に定

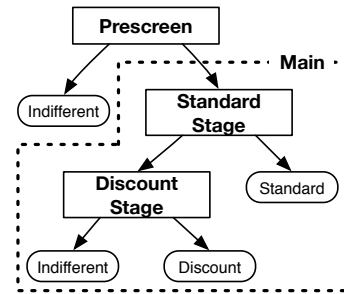


図 1 顧客の識別の過程

価を提示した場合に購入履歴データを保持していると仮定する。PPRS の稼働後は、追加の嗜好データと、定価と割引価格を適切に提示したときの購買履歴が得られる。

嗜好データから推薦モデルを学習し、このモデルを既存の推薦と、割引実施の決定に利用する。顧客がある商品を最初に参照したとき、PPRS はその商品-顧客の対に対して割引販売するかどうかを決定する。この判断は、商品-顧客の対に関する特徴、より具体的には、デモグラフィックなどの顧客自身の特徴と、嗜好データから得た推薦モデルの、対象商品-顧客対のモデルパラメータを用いる。これは、推薦モデルのパラメータの含まれる顧客の嗜好情報と、顧客自身の情報に、顧客の価格感度 (price sensitivity) が依存するのは現実的な仮定だと考えるためである。この判断のための分類器を、顧客の応答に応じて更新する。PPRS は、顧客が商品を参照する度に、定価と割引価格のいずれかを、活用-探索トレードオフを考慮しながら提示することを繰り返す。

3.2 節の問題に対処しつつ、3.1 節の目標を満たすように PPRS を設計する。観測の曖昧性とクラス不均衡の緩和のために採用する多段階分類、活用-探索トレードオフ対策の多腕バンディット、およびクラス不均衡問題に対処するためのクラス重み付けについて順次述べる。

3.3.1 多段階分類

本論文の PPRS では、図 1 のように、前処理段階 (prescreen) と主段階 (main) で構成された過程で顧客型を識別する。前処理段階は、クラス不均衡を緩和するため、自明な不買顧客を除外し、主段階では二つの分類器を用いて、観測の曖昧性を回避しつつ顧客型を識別する。

前処理段階では、明らかな不買顧客を、嗜好データから学習した推薦モデルを利用して除外する。嫌いな商品は買わないという仮定に基づき、ある商品-顧客の対について、商品への予測評価値が非常に低ければ、その顧客を不買顧客とみなす。この予測評価値が低い顧客の除外により、半分以上の不買顧客を除外でき、クラス不均衡を大きく緩和できる。これは、価格個人化を推薦システムに導入する重要な利点の一つである。

観測の曖昧性に対処するため、主段階を、さらに定価分類器 (standard classifier) と割引分類器 (discount classifier) の二つに分けた。前段の定価分類器は、定価顧客とそれ以外の顧客に顧客を分類し、後段の割引分類器は、定価以外の顧客を割引と不買に分類する。

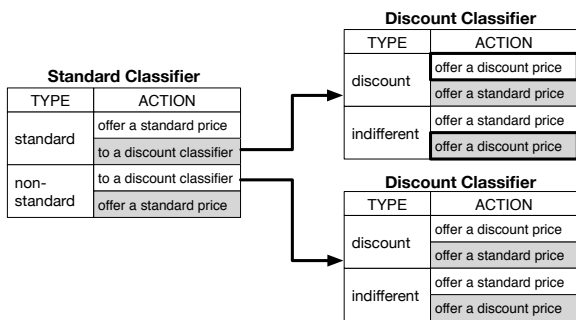


図2 図1中の主段階での行動選択の詳細

注：図1中の主段階の各分類器を小さな表で示した。列“TYPE”には、分類器の予測した顧客型を、列“ACTION”はシステムが選択する行動をそれぞれ示した。白地と灰地の背景は、予測した顧客型に対して、それぞれ活用と探索に相当する選択を表す。

定価分類器と割引分類器の訓練データの獲得について述べる。定価分類器の訓練データは、定価を提示した商品への顧客の応答履歴から生成する。定価で顧客が購入した場合を正例、それ以外を負例とする。さらに、割引価格を提示したときでも、顧客が買わなかった場合は、定価でも買わないのは当然と考えられる。そこで、この場合も定価分類器の負例として利用する。

前段階の分類器で定価顧客以外と識別された顧客に対し、割引価格を提示したときの応答事例を、割引分類器の訓練データとして用いる。これは、定価顧客以外の顧客を分類することが、割引分類器の目的であるからである。この分類器の正例は、割引価格を提示されたときに購入した場合であり、それ以外の場合が負例となる。なお、定価分類器が誤分類することも考えられ、また真の顧客型は観測できないので、この誤分類を検出することはできない。すると、割引分類器の訓練事例としては不適当となってしまうが、定価分類器の訓練事例は初期的に与えられる訓練事例なども利用できるため、定価分類器の精度は高く、このことは大きな問題にはならないと考えている。

3.3.2 多腕バンディットとクラス重み付け

次に、活用-探索トレードオフを扱うために多腕バンディットアルゴリズムを導入する。通常バンディットの設定では、行動の選択肢であるアームの数と、報酬を返す主体であるバンディットの数と同じだが、ここでは3種類の顧客型に対して、行動の選択肢は定価と割引価格の2種類と一致していない。そこで、図1の定価分類器と割引分類器のそれぞれについて、個別に標準的なバンディットアルゴリズムを適用する。この方法では、個別の分類器ごとに累積報酬を最大化しており、二つの分類器を合わせた全体として最適な結果を得られるわけではないが、実験的にはうまく動作した。

図1の過程に、活用と探索の選択も含めて示したものが、図2である。前処理段階を通過した商品-顧客の対は、左の定価分類器に入力される。この対を定価顧客と識別したとき、識別結果を活用するなら定価を提示し、探索をするなら対を割引分類器に送る。逆に、対が定価顧客以外と識別したときは、活用と探索ではちょうど反対の行動を選択することになる。入力が割

引分類器に渡されると、顧客は割引型と不買型に分類される。活用と探索のいずれを選択したかと、この予測顧客型とに応じて、最終的な行動を決定する。

アーム数とバンディット数が一致しない問題に加えて、コンテキストを考慮する必要もある。すなわち、基本型のバンディットでは、報酬は応答履歴のみに依存して予測するが、ここでは、商品や顧客の特徴も考慮する必要がある。このような問題は、特に bandit with covariates や contextual bandit などとも呼ばれる [12]。この問題に対するアルゴリズムも開発されているが、クラス不均衡問題には対処できない。そこで、単純なバンディットである ϵ -greedy [11] と通常分類器とを組み合わせる。このアプローチでは、まず商品や顧客の特徴に基づき通常分類器で顧客型を予測し、その型で報酬を最大にする行動を決める。その行動を確率 $1 - \epsilon$ ($0 < \epsilon \leq 0.5$) で選択し、それ以外の行動を確率 ϵ で選択する。なお、 ϵ -greedy では活用-探索のトレードオフは自動的に調整できず、パラメータ ϵ を手で調整する必要がある。 ϵ -greedy は単純ではあるが、調整がよければ、十分な性能を示すと報告されている [12], [13]。

最後に、クラス不均衡問題を扱うため、クラスの重み付けを導入する。定価分類器と割引分類器に、それぞれしきい値 S_{Th} と D_{Th} を導入する。そして、予測確率が大きい顧客型に分類するのではなく、予測確率がしきい値確率より大きかったときに、定価分類器では定価型へ、割引分類器では割引型へ分類する。 S_{Th} を小さくすることは定価型を重視することに相当し、一方で D_{Th} を小さくすることは割引型を重視することに相当する。このしきい値の導入により、多数派クラスに誤分類されやすいというクラス不均衡問題に対処できる。

4. 実験

簡単なシステムを実装し、半人工のデータに適用した。

4.1 実験条件

4.1.1 手続き

実験手続きは、価格個人化導入前の推薦システムを運用している状態である準備フェーズと、PPRS が稼働している主フェーズとがある。準備フェーズでは、販売者は既存の推薦システムを運用しており、全ての顧客に定価を提示している。これにより、顧客の商品に対する嗜好データと、定価を提示したときの購入履歴が獲得できる。嗜好データは推薦システムのモデルの学習に用い、購入履歴は初期定価分類器の学習に用いる。主フェーズでは、PPRS が、割引をするかどうかを逐次的に判断し、顧客の応答に応じて定価分類器と割引分類器を更新する。主フェーズ開始時では、準備フェーズでも訓練されているため定価分類器はすでに高精度だが、割引分類器は未学習で低精度である。なお、この実験では簡単のため主フェーズでは推薦モデルの更新は停止し、準備フェーズ終了時のものを使った。

4.1.2 データ集合

提案した PPRS をテストするため、MovieLens の 1M データ集合^(注2)から半人工データを生成した。嗜好データと顧客の

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

表 1 PPRS の評価スコアの一覧

(a) 真の報酬			(b) 観測報酬			(c) 利益		
	定	割 不	定	割 不	定	割 不	定	割 不
定	α	0 0	定	α β 0	定	α β α		
割	0	β 0	割	0 β γ	割	0 β 0		
不	0	0 γ	不	0 0 γ	不	0 0 0		

注：行は真の顧客型，列は予測した顧客型に相当する。『定』『割』『不』は、それぞれ定価顧客、割引顧客、および不買顧客を示す。

デモグラフィック特徴は MovieLens のものを用い、顧客の購買履歴は人工的に生成した。人工購買履歴は、最低でも次の三つの条件を満たすべきと考えた。(a) 対象商品に対する嗜好は、定価顧客、割引顧客、不買顧客の順に弱くなる。(b) 購入するかどうかは、デモグラフィックなどの顧客自身の特徴と、顧客の対象商品への嗜好の度合いに依存している。(c) 大多数は不買顧客で、残りのうち割引顧客は定価顧客より少ない。これらの条件を満たすように、まず、対象商品への嗜好が最高の 5 であり、年齢が 45 歳以上の商品-顧客対を選んだ。嗜好度が 5 であることをシステムに隠すため、評価 4 と 5 はどちらも評価 4 として扱って、推薦モデルと分類器の学習に用いた。選んだ対のうち、男性を定価顧客、女性を割引顧客とした。選ばなかったものは不買顧客とした。この購入履歴は非常に単純ではあるが、それでも 3.2 節で指摘した問題のため、システムが追加報酬を得られるかどうかは自明な問題ではないことに留意されたい。

約 100 万個の商品-顧客対のうち、半分は準備フェーズ用として推薦モデルと初期定価分類器の学習に用いた。残りの半分は、主フェーズで PPRS のテストに利用した。定価顧客と割引顧客の占める割合は、それぞれ 3.5% と 1.4% であった。

4.1.3 評価スコア

表 1 に真の報酬、観測報酬、および利益の三つの評価スコアを示す。不買顧客と予測したが、本当は割引顧客であった場合の例を述べる。この場合、表 1(a)~(c) の『不』の行と『割』の列から、真の報酬、観測報酬、そして利益は、それぞれ 0, β , 0 となる。この評価を、全ての顧客に繰り返してえたスコアの合計で性能を評価した。第 1 の真の報酬は、予測した顧客型が真のそれと一致したときのみ増加する。第 2 の観測報酬は、観測の曖昧性により真の顧客型は常には観測できないので、顧客の応答に応じて計算するスコアである。このスコアは、予測した顧客型から期待される応答が、顧客から得られたときに増加する。本来は真の報酬を最大化したいのだが、実際に観測できるのは観測報酬のみであるため、これらの二つの報酬がよく相関していることが望ましい。第 3 の利益は、転売リスクを無視した商品から得る利益の合計で、価格個人化の導入で総利益が増加することを確認するためのものである。

報酬パラメータは、条件 $\alpha > \beta \gg \gamma$ を満たすように、 $\alpha=1.0$, $\beta=0.5$, $\gamma=0.01$ とした。このとき、真の顧客型が完全に分かっているとした場合の、達成可能な報酬スコアの最大値は 25468 となる。全員に定価を提示した場合に、定価・割引・不買顧客からそれぞれ α , 0, γ の報酬を得るとしたときの総報酬は 22028 となり、この値が真の報酬と観測報酬のベースラインとなる。

表 2 3 種類のスコアの平均値

	真の報酬	観測報酬	利益
pLSA	23855.1	23897.4	20331.1
MD	23930.7	23970.2	20346.7

注：これらのスコアは観測報酬を最大にする次の条件で求めた。pLSA モデルでは、 $S_{Th} = 0.01$, $D_{Th} = 0.97$, および $\epsilon = 10^{-2.3}$ とした。MD モデルでは、 $S_{Th} = 0.003$, $D_{Th} = 0.97$, および $\epsilon = 10^{-2.4}$ とした。なお、二つの報酬と利益のベースラインが 22028 と 17268 であることを記しておく。

またこの場合の総収入は 17268 で、これが収入スコアのベースラインとなる。

4.1.4 その他の条件

嗜好モデルを学習するための推薦アルゴリズムとして pLSA と行列分解 (MD) (詳細は [14]) を用い、定価分類器と割引分類器にはロジスティック回帰を用いた。準備フェーズで学習した初期定価分類器を、主フェーズでは、5 万回の応答を顧客から得る度に更新した。割引分類器は、顧客が定価顧客ではないと予測され、かつ割引価格を提示したと場合のみのデータを利用するが、この条件が 50 回生じる度に更新した。

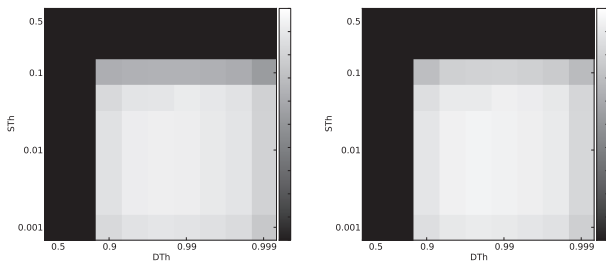
多腕バンディットアルゴリズムの ϵ -greedy の探索確率 ϵ は $[10^{-1}, 10^{-3.5}]$ の範囲で、クラスしきい値 S_{Th} と D_{Th} は、それぞれ $[0.001, 0.5]$ と $[0.5, 0.999]$ の範囲で変化させた。なお、 S_{Th} が約 0.25 のとき、初期定価分類器の F 尺度は最大であった。 S_{Th} , D_{Th} , そして ϵ の各組み合わせごとに 10 回の試行を繰り返し、その平均スコアを報告する。図 1 の前処理段階では、商品-顧客対の予測スコアが 3 未満のものを除外した。

4.2 実験結果

観測報酬を最大にするようにパラメータを定めたときの、3 種類の評価スコアを表 2 に示す。いずれの指標においても、ベースラインを上回る性能を得ることができた。これにより、3.2 節で指摘した問題に、提案した PPRS 手法がうまく対処できることが示された。

S_{Th} を小さく、かつ D_{Th} を大きくしたときにこれらのスコアが最大になる理由について考察する。定価顧客が一度の取引でもたらす報酬は大きいため、報酬の損失を回避するには、定価顧客が非常に重要となる。よって、定価顧客を逃さないように、定価分類器では再現率を重視すべきであり、しきい値 S_{Th} を小さくすべきである。一方で、割引顧客を不買と誤分類してもベースラインからの損失はないので、割引分類器では適合率が重要となり、しきい値 D_{Th} を大きくすることが望ましい。非常に極端な状況として、全ての定価顧客が定価分類器で検出されているのであれば、たった一人でも割引顧客を割引分類器で見つけることができれば、総報酬をベースラインより増やすことができる。

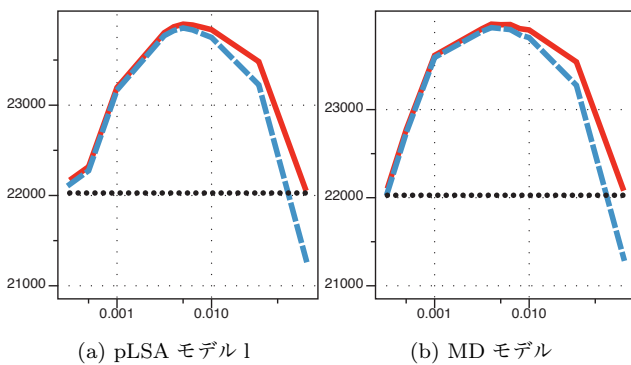
しきい値 S_{Th} と D_{Th} の変化に対する、真の報酬と観測報酬の平均値の変化を、図 3 に示す。上記のように、定価分類器と割引分類器は、それぞれ再現率と適合率を重視するようにすると、どちらの報酬も大きくなっている。逆に、一方のしきい値でもその逆の条件にすると、報酬は急激に減少する。さらに、



(a) 真の報酬

(b) 観測報酬

図3 SThとDThの変化に対する、真の報酬と観測報酬の変化
注：探索確率 ϵ は $10^{-2.3}$ に設定し、pLSAモデルを採用した。横軸と縦軸は、それぞれしきい値 DThとSThに相当する。各セルの輝度は、[22000, 24000]の範囲の報酬を示す。なお、22000未満の報酬は切りあげてある。



(a) pLSAモデル1

(b) MDモデル

図4 探索率 ϵ の変化に伴う、真の報酬と観測報酬の変化

注：横軸と縦軸は、それぞれ探索確率 ϵ と観測報酬を示す。pLSAモデルでは、しきい値は STh=0.01とDTh=0.97に、MDモデルでは、STh=0.003とDTh=0.97に設定した。破線は真の報酬、実線は観測報酬、点線は報酬をそれぞれ示す。

真の報酬と観測報酬は同じしきい値の設定で共に最大化されており、これらの報酬の変化は全般的に相関がある。人工データではなく、実際に運用する場合には、本来は最大化したい真の報酬は分からず、観測報酬だけが計測できる。そのため、このように二つの報酬の変化の相関が高く、観測報酬を最大化すれば、真の報酬も最大になっていることが実験的に確かめられたことは非常に望ましい結果である。

探索確率 ϵ のさまざまな値に対する平均報酬の変化を図4に示す。同じ ϵ の値で、真の報酬と観測報酬は最大となっており、上記の実験結果と同様に望ましい結果となっている。探索確率パラメータ ϵ が、PPRSの性能にとって非常に大きいことも分かる。この ϵ を自動的に調整できるような、UCB1 [10] のようなより高度なバンディットアルゴリズムの導入も今後は検討したい。

以上のことから、パラメータを調整することでPPRSがベースライン以上の報酬を獲得できることが示された。すなわち、観測の曖昧性、クラス不均衡、および活用-探索トレードオフの問題を回避しつつ、全員に定価を提示する場合と比較して、提案手法によって追加利益を獲得できることが示された。

5. 議論とまとめ

本研究では、価格個人化機能を持つ価格個人化推薦システムを提案した。価格個人化の導入で得られる追加利益で、推薦システムの商業的な継続可能性を向上させ、このことにより顧客のシステムへの信頼性が向上することについて論じた。そして、PPRSを実装し、その性質を半人工データを用いて検証した。なお、本研究の詳細は [4] を参考にされたい。

ここでは価格を報酬に置き換えて全体の設計を行ったが、価格以外の効用を報酬として考えれば、この枠組みの適用範囲はさらに広がるだろう。例えば、在庫処分をしたいときには、在庫の維持費を考慮できるような効用関数を設計すればよいだろう。費用面以外に、セット販売や、ポイントサービス、運送面での優遇なども報酬の設計で対処できると考える。今後は、実店舗の店員のように、単に推薦をするだけでなくより洗練された提案や行動ができる、いわば **おもてなしシステム** (attendant system) と呼ぶべきものに、推薦システムは進化してゆくべきだと考えている。

文 献

- [1] 神島：“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] 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).
- [3] 神島, 赤穂, 佐久間：“カスタム価格設定推薦システム”，人工知能学会全国大会 (第24回) 論文集, 3C3-4 (2010).
- [4] T. Kamishima and S. Akaho: “Personalized pricing recommender system — multi-stage epsilon-greedy approach”, Proc. of The 2nd Int'l Workshop on Information Heterogeneity and Fusion in Recommender Systems (2011).
- [5] D. Streitfeld: “On the web, price tags blur: What you pay could depend on who you are”, The Washington Post (2000). <http://www.washingtonpost.com/ac2/wp-dyn/A15159-2000Sep25>.
- [6] “Web sites change prices based on customers' habits”, CNN.com (2005). <http://edition.cnn.com/2005/LAW/06/24/ramasastry.website.prices/>.
- [7] N. Stokey: “Intertemporal price discrimination”, The Quarterly J. of Economics, **93**, 3, pp. 355-371 (1979).
- [8] G. Shani, D. Heckerman and R. I. Brafman: “An mdp-based recommender system”, Journal of Machine Learning Research, **6**, pp. 1265-1295 (2005).
- [9] N. Japkowicz: “Learning from imbalanced data sets: A comparison of various strategies”, AAAI Workshop: Learning from Imbalanced Data Sets, pp. 10-15 (2000).
- [10] P. Auer, N. Cesa-Bianchi and P. Fischer: “Finite-time analysis of the multiarmed bandit problem”, Machine Learning, **47**, pp. 235-256 (2002).
- [11] R. S. Sutton and A. G. Barto: “Reinforcement Learning: An Introduction”, MIT Press (1998).
- [12] L. Li, W. Chu, J. Langford and R. E. Schapire: “A contextual-bandit approach to personalized news article recommendation”, Proc. of The 19th Int'l Conf. on World Wide Web, pp. 661-670 (2010).
- [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] T. Kamishima and S. Akaho: “Nantonac collaborative filtering: A model-based approach”, Proc. of The 4th ACM conference on Recommender Systems, pp. 273-276 (2010).