

# クーポン共同購入サイトにおける ステークホルダ間の利得を考慮したレコメンド方式の提案

松島 裕<sup>1,a)</sup> 中村 達也<sup>1,b)</sup> 矢実 貴志<sup>1,c)</sup> 生田目 崇<sup>2,d)</sup>

**概要：**従来のオンラインショッピングサイトのレコメンドサービスは、サイト運営者の収益性が最大になるように商品・サービスレコメンド対象を決定していることが多く、いわばサイト運営者の都合を重視した方式となっている。インターネットの EC 市場においては、サイト運営者は仲介業者である場合も多く、その場合、利害関係者として、サイト運営者、情報掲載企業、消費者の 3 つのステークホルダが存在する。このような場合においても、通常のレコメンドではこれら 3 つのステークホルダの便益の獲得は考慮されていない。本稿では、近年売上を伸ばしている、クーポン共同購入サイトを念頭に、各ステークホルダの利得を指標化した上でそれらを最大にするような、サイト訪問者に対するクーポンレコメンドロジックを提案する。本稿の提案ロジックでは、フラッシュマーケティングビジネスの特徴である時限性や各クーポンの特徴の差を考慮している。また、提案した方法を実データに当てはめ、その有効性について考察する。

**キーワード：**クーポン共同購入サイト、レコメンドーション、マイニング、非構造データ

## 1. はじめに

2010 年に日本初のクーポン共同購入サイトが登場して以来、こうしたサービスは現在に至るまで国内だけで 200 サイト以上が登場するなど活発化している。クーポン共同購入サイトビジネスはクーポン供給側である情報掲載店舗と需要側である一般消費者をマッチングする仲介業者であるサイト運営業者によって行われる。こうしたサイトでは一般に大幅な割引が行われ、消費者にとっては通常価格に比べ格安に商品・サービスを購入することができる。また目当てのクーポン以外にも同時に複数のクーポンがリストアップされ、魅力的な商品に偶然出会えるといったメリットがある。一方で、クーポン掲載店舗にとっては比較的安価な宣伝費用で多くの見込み客に自店舗の商品やサービスをプロモーションすることができ、多くの新規顧客を獲得する機会となりうるというメリットがある。サイト運営者はクーポン販売金額に応じた手数料を徴収しており、より多くのクーポンが購入されるほど利益が大きくなる仕組み

である。

従来のインターネットショッピングビジネスでは、消費者のニーズを刺激し売上を向上させるためにサイト訪問者に対する商品のレコメンドーションが広く行われている。しかし、クーポン販売時の手数料収入に依存するクーポン共同購入サイトのビジネスでは、運営側の利益を優先した場合、利益率が良い、もしくは人気の高いクーポンを優先的にレコメンドする傾向もあり、結果、レコメンドされる商品に偏りが生じたり、利益を得ることができない掲載店舗が発生するという問題が発生する。CRM (Customer Relationship Management) の観点から見れば、顧客は常に同じようなクーポンを表示されることになりかねず、長期間魅力あるサイトとして見られない危険性も生じる。

また、従来のインターネットショッピング向けのレコメンドーションでは、商品ジャンルなどの属性を主要素としてレコメンドを行っているため、クーポン共同購入サイトの特徴の一つである短期間の販売期間という時限性は十分に考慮されていない。さらには、クーポン共同購入サイト利用者に共通の訴求商品があるわけではなく、あくまで個々人の興味や嗜好によって購入商品が決定される傾向にあるため、消費者が明示していない購買傾向をとらえきれない、といった問題点がある。

白石 [4] は、フラッシュマーケティングとレコメンドーションの仕組みとしてのクーポン共同購入サイトについて

<sup>1</sup> (株) NTT データ  
NTT DATA Corporation

<sup>2</sup> 中央大学  
Chuo University

a) matsushimahrs@nttdata.co.jp

b) nakamurattf@nttdata.co.jp

c) yazanet@nttdata.co.jp

d) nama@indsys.chuo-u.ac.jp

まとめ、その中で、レコメンデーションに内包する問題として、レコメンデーションに適さない商品の販売促進をどう行うか、また、対象顧客の価値観やニーズと推奨商品のミスマッチが判明した際にどのように対応するかという2つの問題を提起している。特に後者においては、当該顧客および当該商品に関してどのようにフォローするかという仲介業者たるサイト運営企業において、供給者である情報掲載店舗と需要者である消費者の両者に対する適切な対応の重要性を指摘している。

そこで本稿では、上記問題を解決するために各ユーザの潜在的な購買傾向と各ステークホルダの利得を考慮した上で、クーポンの特有の特徴の活用したリアルタイムレコメンド手法を提案する。本提案手法ではクーポンのジャンルだけでなく、価格帯や販売期間等の消費者の潜在的な購買意欲と関連性の深い情報を利用することで、個人の嗜好を反映したレコメンドを実施する。また、クーポン共同購入サイトのビジネスのステークホルダである「サイト運営者」「クーポン掲載店舗」「消費者」の各々の利得を指標化し、レコメンド商品の決定に用いることで、各ステークホルダの利得を担保したレコメンドの実施が可能になる。

本稿は、経営科学系研究部会連合大会主催の平成24年度データ解析コンペティションの課題フリー部門の研究成果である。本コンペティションでは、(株)リクルートライフスタイル<sup>\*1</sup>が運営するクーポン共同購入サイト「ポンパレ」におけるウェブアクセスデータ、購買データ、クーポンマスタ、顧客マスタが提供された。本稿では、このデータを用いて議論する。

本稿の構成は以下の通りである。第2節では本稿において用いたデータの概要と掲載クーポンの特徴について示す。第3節ではクラスタリング手法を用いたクーポンの購買要因に関する特徴に基づいたグルーピング法について示す。第4節では、各ステークホルダの利得に関する指標を定義し、上記指標と第3章のクラスタリング結果を用いたレコメンド手法を提案する。第5節では本稿の提案手法と従来手法との比較検証結果を示す。最後に、第6節で本稿のまとめと今後の展望を示す。

## 2. 使用したデータの概要

本稿において用いたデータの概要を以下に示す。

期間： 2011年7月1日～2012年6月30日

アクセスログデータ： 2,889,405件

購買ログデータ： 171,599件

クーポンマスタ： 19,723件

顧客マスタ： 22,873人

なお、クーポンはあらかじめジャンルが付与されており、クーポン種類の多い順に「宅配」「グルメ」「ホテル・

旅館」「ヘアサロン」「リラクゼーション」「その他」「エステ」「レッスン」「レジャー」「ネイル・アイ」「ギフトカード」「健康・医療」「ビューティー」である。クーポンマスタには、定価、販売価格、店舗フラグ、住所、購入条件、使用条件、宣伝テキストなどが含まれる。また、顧客マスタには顧客ID、顧客の年齢、性別が含まれる。アクセスログはサイト内のすべてのアクセスではなく、特定のクーポンのページへのアクセスデータのみが提供されている。したがって、トップページからの遷移やサイト内での検索行動などについては捕捉することはできない。

## 3. 嗜好に基づいたクーポンクラスタリング

クーポン購入は、お得感の強い商品を好むといった利用者の嗜好とクーポンの特徴が一致したときに起こると考えられる。そこで本稿では、各利用者の各アクセス時点での嗜好を推定し、その嗜好となるべく一致したクーポンをレコメンドすることを目的としたロジックを提案する。嗜好については、ジャンルや価格・利用条件といった各クーポンについて数値化もしくはカテゴリ化されている属性情報もさることながら、感覚に影響を与える定性敵情報も購買意思決定に影響を与える。定性的データの例としては、五感に訴えるような画像・動画や臭い、触感、文章などが考えられるが、本データにおいては各クーポンページに表示される利用者への訴求コメント（以降、「宣伝テキスト」と呼ぶ）が含まれている。

### 3.1 属性概要

提供されたデータにおいて、クーポンは「グルメ」「物販」「ホテル・旅館」のカテゴリが大半を占めていた。クーポンの販売期間は72時間（3日）の場合が最も多く、ほとんどが1週間以内である。定価に対する割引率は大半が50%以上であるが、特にエステやビューティなどのサービス系ジャンルについては高割引率のクーポンが多く存在した。また、その他の属性においてもジャンルごとに各属性値に違いがあることから、同一ジャンルでも嗜好によってターゲットが異なっていると想定される。

### 3.2 宣伝テキストの加工

宣伝テキスト情報は他の属性情報とは異なり非構造データである。したがって、このデータを用いて直接クーポンをクラスタリングに用いることは困難である。そこで本稿では、形態素解析を用いて頻出単語を抽出し、嗜好性に関する重要な単語を抽出する。そして、類似する表現や意味を持つ単語をグループ化する。それらの単語グループごとに「嗜好性に関する単語」とし、それらの単語が各クーポンに含まれるかどうかを新たなクーポン属性として付与する。手順は下記の通りである。

(1) クーポンの紹介用テキストに対してクーポンのジャン

<sup>\*1</sup> <http://www.recruit-lifestyle.co.jp/>

ルごとにオープンソースの形態素解析システムである MeCab[2] 用いて単語ごとに分割する。そして、およそ 5%のクーポンに含まれることを閾値として、TF (単語出現頻度) の上位 30 位の単語を抽出する

(2) 抽出した単語のうち、本稿にて定義した 7つの嗜好性に関する判断基準「質」「価格」「量」「行動」「利便性」「評判」「付加価値」と一致する単語を同定し、各判断基準に一致する単語があるかどうかを 2 値の属性情報としてクーポンに付与する

### 3.3 嗜好性によるクラスタリングと解釈

前述の嗜好性に関する単語の有無に加え、そのほかのクーポン属性情報を付加し、非階層型クラスタ分析である k-means 法 [3] によるクラスタリングを実施した。クラスタの大きさ並びに解釈の観点から 6 つのクラスタを作成した。各属性におけるクラスタの構成割合を図 1、各クラスタにおけるクーポンジャンルの構成割合を図 2 に示す。これらから、各属性毎に関連性の強いクラスタが存在すること、各クラスタ毎に構成するクーポンの属性値の偏りがあることが分かる。本稿では、これらの集計結果と特徴的な単語の割合から、各クラスタを以下のように解釈した。

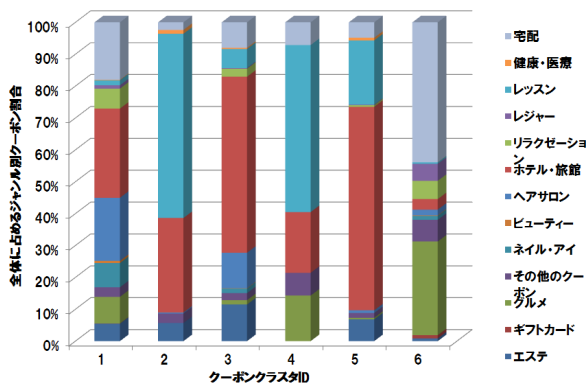


図 1 クラスタ別クーポンジャンル構成割合

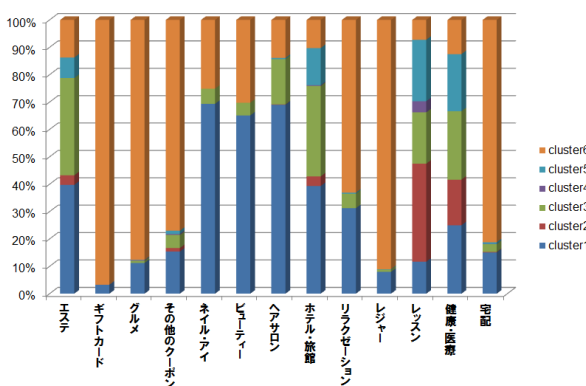


図 2 ジャンル別クラスタ構成割合

#### クラスタ 1

概要： 美的生活型

特徴： エステやサロン等のビューティ系が多い  
利用可能曜日等の制約無し

#### クラスタ 2

概要： 学び場型

特徴： レッスン系が多い  
掲載エリア数が多く、日付限定が多い

#### クラスタ 3

概要： エンジョイ型

特徴： ホテルやエステ系が多い  
物販やギフトカードが少ない

#### クラスタ 4

概要： スペシャルティ型

特徴： レッスン, ホテル, グルメ, 物販が多い  
価格が高く、付加価値の大きい商品が多い

#### クラスタ 5

概要： お買い得型

特徴： レッスン, ホテル, エステ系が多い  
割引率が高い商品が多い

#### クラスタ 6

概要： 日用品型

特徴： 物品, ギフトカード, グルメ系が多い  
販売金額が小さく、販売上限数が多い

## 4. ステークホルダの利得を考慮したレコメンドロジック

本節では、前節のクーポンのクラスタリングの結果を用いて、消費者の潜在的な購買傾向に基づき各ステークホルダの利得を考慮したレコメンドロジックを提案する。本提案手法では、「利用者は同一セッションにて閲覧したクーポンと同様のクラスタに属するクーポンに対して“閲覧時点での購買意欲”を持っている」「過去の同一セッション内でのクーポンクラスタの遷移状況から“潜在的に購買意欲”のあるクーポンを判別できる」と仮定し、サイト内の同一セッションの閲覧履歴を基にレコメンドすべきクーポンの選定を行う。

### 4.1 レコメンドロジックの概要

本稿では、3 節で示したクーポンのクラスタリング結果を用いて顧客ごとに潜在的な購買意欲に則したクーポンのレコメンドを実施する。本稿は、以下で説明するような現時点での購買意欲に合ったクーポンをレコメンドする“追従型ロジック”と潜在的な購買意欲に合ったクーポンをレコメンドする“プロアクティブ型ロジック”を複合した方法を提案する。

#### 4.1.1 追従型レコメンドロジック概要

本稿における追従型ロジックは、「利用者は同一セッションの直前に閲覧したクーポンと同じクラスタに属するクーポンに対して閲覧時点での購買意欲がある」という仮定の

もとでクーポンを Recommend するロジックである。本稿では、同一セッションにおいて直前に参照したクーポンクラスタと同様のクラスタを Recommend するものとする。

#### 4.1.2 プロアクティブ型 Recommend ロジック

本稿におけるプロアクティブ型手法とは、同一セッション内の過去のクーポン閲覧情報を利用者の持つ嗜好パターンとしてとらえ、閲覧クーポンから潜在的な購買意欲と合ったクーポンを Recommend する手法である。プロアクティブ型手法の各ステップのロジックは下記の通りである(図3)。

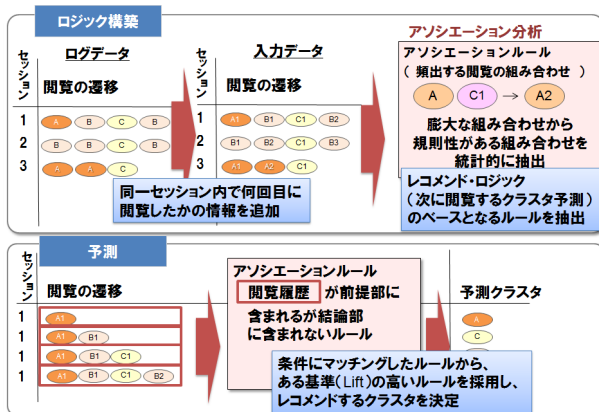


図3 提案するプロアクティブ型 Recommend ロジック

##### 4.1.2.1 プロアクティブ型ロジック学習ステップ

###### 入力

全利用者のサイトの遷移ログ\*2

###### 出力

頻出クーポンクラスタパターン

###### ロジック

- (1) 同一セッションにおいて閲覧したクーポンをクーポンクラスタに変換し、セッション開始から現在までのクーポンクラスタの集合を1つのアイテム集合とする。ただし、同じセッション内で同一クーポンを連続して参照している場合は重複を削除する
- (2) アソシエーション分析の一種である Apriori 法 [1] を用いてリフト値が最も高いパターン（頻出パターン）を抽出

##### 4.1.2.2 プロアクティブ型ロジック予測ステップ

###### 入力

対象顧客の現在の遷移ログ

###### 出力

被 Recommend クーポン集合

###### ロジック

- (1) 入力ログデータを学習ステップ時と同様のトランザクションデータに変換

\*2 ただし前述のとおりすべてのアクセスログデータではないため、閲覧までのプロセスなどは把握できず、クーポン間の遷移データとなる。

- (2) 学習した頻出パターンの前提部に (1) が包含されるパターンを抽出
- (3) 抽出したパターンのうちリフト値が最大となるパターンの結論部のクーポンクラスタ（被 Recommend クラスタ）を出力

本提案手法では、追従型ロジック、プロアクティブ型ロジックの双方で決定したクーポンクラスタを被 Recommend クラスタとして最終的に出力する。上記ロジックを用いることで、クーポンの購買特徴を形成する「ジャンル」「限定性」といった利用者の嗜好に合ったクーポンをクラスタ単位で加味しつつ、かつ過去の閲覧パターンから各利用者のニーズに合っていると見えるクーポンクラスタも同時に Recommend することが可能になる。

##### 4.1.3 各ステークホルダの利得の指標化

4.1.2 節では、各利用者の潜在的な購買意欲にマッチしたクーポンクラスタを Recommend するロジックを提案した。本節では、被 Recommend クラスタに属するクーポンの中から実際に Recommend するクーポンを決定するための指標として、ステークホルダの利得を指標化する。

本稿では、各ステークホルダの利得を下記のように解釈し、評価関数として以下を用いる。

###### サイト運営業者

利得： 利益向上

特徴： 総売上期待値

サイト運営業者の利得としては、全体の利益向上である。利益は売上に比例するため、売上期待値を最大にすることで、最大利益を得られる。そこで、下記のように、クーポン価格に対して、アクセス数で購買枚数を割った成約率を乗じたものを利得の指標とした。

$$(\text{運営業者利得}) = (\text{価格}) \times \frac{(\text{購買成立枚数})}{(\text{アクセス数})}$$

###### クーポン掲載店舗

利得： 広告宣伝, 損失補てん

特徴： 残り時間当たりのクーポン在庫数

クーポン掲載店舗にとっては、新たな顧客の獲得のための宣伝、機会損失を補てんする機会として、クーポン発行を行うことが考えられる。したがって、想定したクーポン枚数を売り切ることが重要となる。そこで、指標値として、下記のように時点時点における購買上限数および残りの販売時間に対する販売可能枚数という効率指標を与える。

$$(\text{掲載店舗利得}) = \frac{(\text{販売上限数}) - (\text{購買数})}{(\text{購買上限数}) \times (\text{残り販売時間})}$$

###### 訪問者者

利得： 各訪問者のニーズに合ったクーポンを入手

利得： 地理上購入者に近いクーポン

訪問者に関する属性データは十分に得られなかったため、アクセスしたクーポンと地理的属性に近いクーポンを提示

することを考えた。これは、店舗の利用においては立地が大きな制約条件になることを考慮したものである。ただし、商品販売のように、立地の影響がないカテゴリーの場合は、該当する商品の中からランダムに Recommend することにする。なお、次節で示す本稿の実験では、各クーポンの住所データから市町村レベルでの緯度経度に変換している。

## 5. 検証実験

本節では、前節までに提案した Recommend ロジックに対して、その精度評価、及び各ステークホルダの利得獲得に関する効果検証結果を示す。

### 5.1 予測精度検証

本稿の提案手法による Recommend 結果の精度を評価するために、過去のサイトの閲覧履歴を用いて Recommend クラスターの正答率の比較を行った。

#### 5.1.1 検証概要

予測精度の検証のために用いたデータの概要、ロジックのパラメータの設定は下記の通りである。

##### 学習用データ

- 期間：2011年7月1日～2012年3月31日(9カ月)
- データサイズ：389,169 レコード

##### アプリアリ法パラメータ

- 最小サポート：1%
- 最小信頼度：0%
- ルール長上限：8

また、検証方法の概要は下記の通りである。

- (1) 学習用データの各セクションに対して、最後に閲覧したクーポンを予測の正答データ、その他のクーポンを4章にて示したデータ加工法と同様の処理にてトランザクションデータ化する。
- (2) (1)において生成したアイテム集合を入力として、提案手法とランダム Recommend を用いてそれぞれ被 Recommend クーポンを出力
- (3) (2)の正答データと各手法の出力結果を比較し、一致した割合を比較

#### 5.1.2 検証結果

検証の結果、ランダム Recommend ロジックが正答率 40.7%であったのに対し、本稿で提案したプロアクティブ型 Recommend 手法の正答率は 68.5%であり、より高精度な Recommend が可能であることが示された。また、直前に閲覧したクラスターと同じクラスターのクーポンを Recommend する追従型 Recommend とプロアクティブ型 Recommend の結果を比較した際、追従型は 62.2%、プロアクティブ型は 46.6%であった。一般的にクーポン共同購入サイトでは同じようなクーポンを連続して閲覧する傾向にあることから、全体の評価では追従型手法のほうが精度が高い。しかし一方で、他のクラスターにスイッチする場合を比較すれば、

プロアクティブ型にて Recommend したクーポンを多く閲覧している。連続した閲覧を当てるということも重要であるが、CRM の観点に立てば、訪問者が気づいていないニーズを掘り起こし驚きや気づきを誘発することも重要であり、本稿のロジックはこうした新規のニーズ開拓に対して一定の効果があると考えられる。したがって、両者を複合的に用いることにより、ランダム Recommend ロジックよりも十分に高い有効性を担保していると考えられる。

### 5.2 ステークホルダの利得獲得に関する検証

本節では、本稿にてフラッシュマーケティングサイト Recommend の課題点として挙げている「各ステークホルダの利得の有無」に対する本提案手法の効果を検証した。

#### 5.2.1 検証概要

予測精度の検証のために用いたデータの概要、ロジックのパラメータの設定は下記の通りである。

##### 学習用データ

- 期間：2011年7月1日～2012年3月31日(9カ月)
- データサイズ：389,169 レコード

##### 検証用データ

- 期間：2012年4月29日～2012年5月5日(1週間)
- セッション数：47,039 セッション

##### アプリアリ法パラメータ

- 最小サポート：1%
- 最小信頼度：0%
- ルール長上限：8

##### その他検証用パラメータ

- クーポン参照時購入確率：7.5%

また、本稿にて実施した各ステークホルダの利得獲得比較の検証方法は下記の通りである。

- (1) サイト閲覧時のクーポン購入確率を算出
- (2) 各セッションにおいて上記購入確率からクーポンの購入確率を算出
- (3) 下記2つの値を算出して比較
  - (a) 総売上額  
— 検証期間内における購買クーポンの売上額
  - (b) 販売クーポンの未達成率  
— Recommend されたクーポンの販売未達成率(販売上限枚数に占める現状の販売割合)の比較

利得獲得検証の結果について、セッションの進行に合わせた Recommend 精度の時系列推移を図4, 5にまとめる。売上期待値に関しては、ランダム Recommend に比べて本稿の提案手法のほうが高く、セッション数が進むにつれてより差異が顕著になっている。また、各セッションにおける販売未達成率の比較に関しては、ランダム Recommend に比べて本稿の提案手法のほうがより現時点で販売未達成率の高い、クーポン掲載元店舗にとって在庫リスクの高いクーポンを Recommend していることが分かる。

これらの結果より、本稿の提案手法は売上増加、在庫リスクの低減の観点において、サイト運営側、クーポン掲載店舗側の双方が利得を得られる Recommend が実装できていると考えられる。

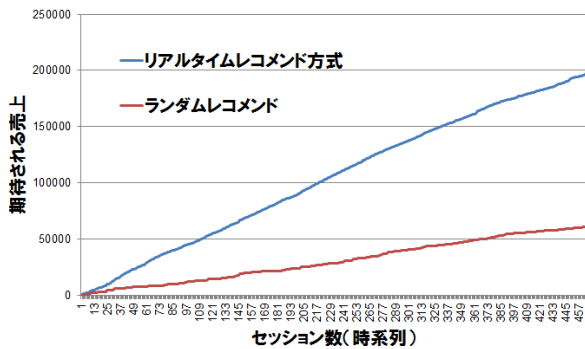


図 4 総売上額期待値の時系列推移

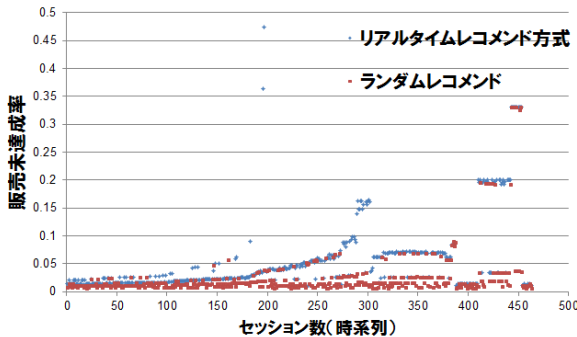


図 5 被 Recommend クーポン販売未達成率の時系列推移

## 6. まとめと今後の課題

本稿では、クーポン共同購入サイトの特徴である時限性やジャンルや割引率などの利用者の嗜好性に関わる特徴の特徴を考慮しつつ、各ステークホルダそれぞれが利得を獲得するための Recommend ロジックを提案した。精度比較と各ステークホルダの利得獲得に関する検証結果により、本稿の提案手法はランダム Recommend 手法に比べて高精度であり、かつ各ステークホルダの獲得利得が増加していることが分かった。

今後の課題としては、現実のフラッシュマーケティングサイトの特徴や充足条件を満たした Recommend ロジックに改良することである。例えば、通常のネットショッピングでは商品の検索時により上位の位置に掲載されている商品のほうがより購買される確率が高くなるという特徴がある一方、表示画面に掲載するクーポンの枚数自体に限られるため、本稿の提案手法のように現在のセッションにおける追随型の Recommend とプロアクティブ型の Recommend を同時に実施することができない可能性がある。その際、サイト運営側と掲載元店舗の利得バランスをどのようにして調整するか、またどのような順序関係にて Recommend するかなどを深堀する必要がある。また、本稿ではデータの都合

上考慮出来なかったが、サイト内での全アクセスログを用いた訪問者の購買プロセスのモデル化も今後の課題である。特に、検索などの絞り込みは訪問者の嗜好やニーズが反映されたものであり、こうした購買前のデータをフル活用した Recommend ロジックの開発が望まれる。

## 参考文献

- [1] R. Agrawal, T. Imielinski and A. Swami, Mining Association Rules Between Sets of Items in Large Database, *Proc. SIGMOD*, pp.207–216 (1993).
- [2] T. Kudo, K. Yamamoto and Y. Matsumoto, Applying Conditional Random Fields to Japanese Morphological Analysis, *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004)*, pp.230–237 (2004).
- [3] J.B. MacQueen, Some Methods for Classification and Analysis of Multivariate Observations. 1. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, pp.281–297 (1967), Retrieved 2009-04-07.
- [4] 白石弘幸, Recommend ーションとフラッシュマーケティング—ICT 支援販売促進の連携—, *金沢大学経済論集*, Vol.31, No.2, pp.133–152 (2011).