

# EC サイトにおける 店舗推薦システムに向けた類似店舗抽出手法

大河原一輝<sup>†1</sup> 牟田将史<sup>†3</sup> 平野廣美<sup>†3</sup> 益子宗<sup>†3</sup> 星野准一<sup>†2</sup>

ショッピングモール型の E コマースは数万規模の多数の店舗により構成される。店舗が販売する商品には店舗特徴（品揃え、価格帯、テイストなど）が反映されており、幅広い商品ジャンルをカバーしている。ユーザは店舗単位で商品を開覧することで商品ジャンルの関連性にとらわれない幅広い商品を開覧することができる。しかし、従来の E コマースでは店舗数が膨大であるため、ユーザが興味を持つような店舗を探すことが難しいという問題がある。ユーザが興味を持つような店舗の発見性を向上させることにより、ユーザが好みと感じる商品の発見性が向上すると考えられる。そこで本稿では、ユーザが過去にアクセスした店舗に類似した店舗を推薦できるような仕組みとして、類似店舗抽出手法を提案する。評価実験では、店舗をランダムに選んだ場合と提案手法が選んだ場合の好みの店舗の発見率を比較し、提案手法の方が高い発見率となることを確認した。また、提案手法のパラメータ変化に対する発見率の関係を調査し、提案手法を実際の店舗推薦システムに応用するための知見を得た。

## The Similar Store Extraction Method for The Store Recommendation System on E-Commerce

KAZUKI OOKAWARA<sup>†1</sup> MASAHUMI MUTA<sup>†3</sup>  
HIROMI HIRANO<sup>†3</sup> SOH MASUKO<sup>†3</sup> JUNICHI HOSHINO<sup>†2</sup>

Shopping mall type electronic commerce sites (EC-malls) has many stores created by store owners. The products on each store reflect the characteristics of the stores (selection, pricing, styles, etc.). And each store covers a wide range of product categories. Customers can view the items in the store unit, and can search for a wide range of items. However, it is difficult to identify preferred stores, because EC-Mall has many stores. We thought customers can search for favorite items if they can find preferred stores. In this paper, we proposed the similar store extraction method for the store recommendation system. In order to validate the effectiveness of the method, we conducted comparative experiments with store recommender system used the method and the random recommender system. The results show that the system used the proposed method finds preferred stores better, and the system has favorable conditions.

### 1. はじめに

ショッピングモール型の E コマースサイト（以降、EC-Mall と略す）は数万規模の多数の店舗により構成される。店舗が販売する商品には、品揃え、価格帯、商品テイストなどの店舗の特徴が反映されている。また、店舗は幅広い商品を扱っている。ユーザは店舗単位で商品を開覧することで商品ジャンル[a]の関連性にとらわれない幅広い商品を探ることができる。店舗回遊は幅広い商品への気づきを増やすことができると考えられる。

幅広い商品への気づきを増やすことは、ユーザの商品探索行動の 1 つである「探索型検索」の支援委繋がると言われている[1]。探索型検索 (Exploratory Search) とは、欲しい物が曖昧な（または決まっていない）状態で商品を探ることである。例えば、好みの服を探したい、便利な家具を見つけないなどの曖昧な要求に対して商品を探る行動は探索型検索である。文献[2]によると、探索型検索で

は明確な検索対象が存在しないため、まず欲しいと思える商品を探して目的を明確化させていく必要があるとされている。また、曖昧な目的は検索クエリの作成を困難にするため、キーワード検索などの参照型検索 (Lookup Search) では探索型検索の支援は難しいとされている。探索型検索を支援する為には、ユーザ自身で幅広い商品を探さず、好みの商品を見て回る仕組みが必要であると考えられる。

EC-Mall における店舗は幅広い商品を扱っている。ユーザは好みと感じる店舗を回遊することにより、興味のある商品を幅広く見て回ることが可能であると考えられる。しかし、EC-Mall では店舗数が膨大であるため、ユーザが好みと感じる店舗を探すことが難しいという問題がある。

本稿では、好みの店舗を見つけるための仕組みとして、ユーザの好みの店舗に類似した店舗を推薦することができる類似店舗抽出手法を提案する。本手法では、店舗ごとに取得した商品説明文から形態素を Bag of Words (単語と頻度を合わせた情報) として取り出し、TF-IDF 法によって店舗の特徴語を抽出する。そして、店舗間の特徴語を Set of Words (単語のみに限定した情報) として類似度を計算することにより、類似店舗の情報を抽出する。本手法は店舗類似度を商品ジャンル別に計算してから統合することにより、店舗が扱う商品ジャンルの違いを考慮した類似店舗を抽出

<sup>†1</sup> 筑波大学情報学群情報メディア創成学類  
College of Media Arts, Science and Technology, School of Informatics,  
University of Tsukuba

<sup>†2</sup> 筑波大学大学院システム情報系  
Faculty of Engineering Information and Systems, University of Tsukuba

<sup>†3</sup> 楽天株式会社楽天技術研究所  
Rakuten Institute of Technology, Rakuten, Inc.

a) 商品毎に付与される女性服、雑貨、食品等といった商品情報を指す。

することができる。店舗が扱う商品ジャンルの違いを考慮することで男性向け店舗と女性向け店舗を分けることが可能であり、より精度の高い類似店舗が抽出されることが期待できる。また、単語辞書を利用せずに店舗の特徴語を作成し、類似店舗を抽出することも可能である。

評価実験では、店舗をランダムに選んだ場合と提案手法が選んだ場合の好みの店舗の発見率を比較し、提案手法の方が高い発見率となることを確認した。また、提案手法のパラメータ変化に対する発見率の変化を調査し、本手法を実際の店舗推薦システムに応用するための知見を得た。

## 2. 関連研究

### 2.1 商品推薦手法に関する研究

幅広い商品の気づきを支援するために多くの商品推薦技術が研究されている。商品推薦を行うための情報フィルタリング技術は、協調フィルタリングと内容ベースフィルタリングがある[3]。協調フィルタリングは、ユーザの行動履歴の関連性を用いて商品間の関連性を抽出する手法である[4]。協調フィルタリングでは、ユーザの行動履歴を用いて商品の関連性を抽出するため、商品ジャンルによる関連性を越えた幅広い商品の繋がりを抽出することができる[5]。また、コンテンツを解析せずに関連性を抽出できる手軽さから多くの手法が提案されている([8][9][10]など)。例えば、Hijikataらはユーザの知らないような商品の発見性を向上させる手法を提案している[7]。Zieglerらは推薦リストのトピックを多様化させ、ユーザの満足度を向上させている[11]。Amazon.comやGoogle Newsが提供するシステムにも協調フィルタリングが用いられている[12][13]。協調フィルタリングは特徴量を抽出することが困難なアイテムに対しても、高い精度で商品を推薦できると言われている[14]。しかし、高い推薦精度が災いし、似たような商品ばかり推薦されてしまうといった問題が多くの商用システムで確認されている[14]。評価付けがされていないアイテムが推薦候補に入らない cold start 問題や、新規利用者の推薦の質が低くとどまる first rater 問題などの問題も存在する[6]。

一方、内容ベースのフィルタリングは、商品情報の解析を行うことで商品間の関連性を抽出する手法である[15]。内容ベースフィルタリングでは、ユーザの行動履歴を利用せずに商品間の関連性を抽出するため、評価付けされていない新商品も推薦することができる[6]。また、少数派の嗜好パターンをもつ利用者にも適切な情報を推薦することができる[16]。内容ベースフィルタリングを用いた例として、MooneyらはAmazon.comから取得した書籍情報の特徴を利用した推薦システムを提案している[15]。Caoらはコンテンツ間の類似する概念にファジィ理論を適用し、偏りの少ないコンテンツの揭示手法を提案している[17]。内容ベースフィルタリングでは、ユーザプロフィールとコンテンツモデル(商品情報から抽出した商品の関連性)を直接的

に比較するため、新規性や意外性の高い商品を推薦することが難しいと言われている[14]。

上述のように、協調フィルタリングと内容ベースフィルタリングは両者とも利点と欠点が異なる。本稿で提案する手法も店舗の関連性を抽出するために情報フィルタリングを扱う必要がある。上記の知見をもとに、提案手法が扱う情報フィルタリングについて3.2節で検討する。

### 2.2 商品探索支援に関する研究

ユーザの情報探索行動を支援するために、キーワード検索に用いるクエリを推薦する研究が行われている。例えば、Parikhらは入力したクエリのSemanticな関係性を3種の類似指標により抽出し、関連したクエリを推薦している[18]。Schraefelらはユーザが閲覧しているコンテンツから属性を抽出し、それをパネルとして掲示するシステムを提案した[19]。Cutrellらは検索結果に含まれるメタデータを掲示し、検索条件を容易に変更できるシステムを提案している[20]。キーワード検索結果の商品配列を変更することにより、商品探索を支援する研究も存在する。Bhattacharyaらは検索クエリの情報を特定属性と不特定属性という独自の指標で分類し、不特定属性の類似度が高いものから商品表示を行う手法を提案している[21]。これらの研究では、キーワード検索のクエリ生成を支援することができる。しかし、ユーザの情報要求が曖昧な場合の商品探索ではクエリとユーザの需要が一致せず、ユーザを満足させられるような商品を探索することが難しいと考えられる。

大河原らは店舗単位の商品探索の可能性を論じ、店舗の回遊性を向上させるインタフェースを提案している[22]。この研究では、店舗を素早く回遊させることにより多くの商品への気づきの支援を可能にしているが、数万規模の店舗の回遊が考慮されていない。また、掲示する店舗を限定するための店舗推薦手法の必要性について述べている。本研究では、幅広い商品探索を支援するための店舗推薦システムの実現に向け、店舗推薦手法を提案している。

## 3. 類似店舗抽出における要件分析

### 3.1 提案手法の対象について

ユーザが好みと感じる店舗の類似店舗を推薦することにより、ユーザは多くの好みの店舗を発見することができる。好みの商品を幅広く発見できると考えられる。提案手法では、抽出される類似店舗がユーザの好みによる繋がりを考慮できるような手法を目指す。欲しいと感じる商品を好みの商品とし、リピートしたいと感じる店舗を好みの店舗と定義する。本稿では、その具体例としてEC-Mallにおける実商品のうち最も需要の多いアパレル店舗[b]を対象とする。アパレル店舗はユーザによって商品の嗜好が大きく異なり、その判断基準も曖昧で個人差があるため、本手法が支援する探索型検索が行われやすいと考えられる。

b) 衣類・服飾品全般を販売する店舗として用いている。

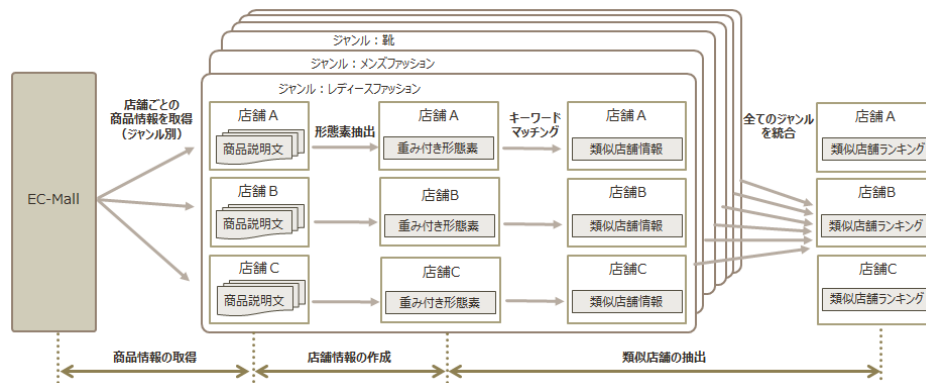


図1 類似店舗抽出手法の概要図  
Figure1 The Similar Store Extraction Method.

### 3.2 扱う手法の検討

情報推薦を行うための情報フィルタリング技術は協調フィルタリングと内容ベースフィルタリングに分けられる(2.1節を参照)。協調フィルタリングでは、ユーザの行動履歴にない新しい情報の関連性を抽出することが難しく、行動履歴に多く含まれる情報ほど優先的に抽出されてしまうという問題点が指摘されている[6][14]。協調フィルタリングによる手法では、新規参入店舗が考慮されにくく、老舗店舗が優先されやすくなってしまふことが懸念される。

一方、内容ベースフィルタリングでは、幅広い商品の繋がりが抽出されにくいことが指摘されている[14]。商品情報を用いた場合、取得した情報が商品の特徴を全て表現できているとは限らない。その不足した情報による関連性では幅広い商品の繋がりが抽出されにくいと考えられる。提案手法では店舗単位の膨大な商品情報を用いて関連性を抽出するため、この問題は改善される可能性が高いと考えられる。提案する類似店舗抽出手法には内容ベースフィルタリングを適用する。扱う情報は、TF-IDF法(文書内の特徴語に重みを付ける手法)[23]などによって店舗単位の情報にまとめることが容易である商品説明文を利用する。

本稿では、以上の要件を満たした類似店舗抽出手法を提案する。ユーザの好みが反映された類似店舗の抽出が可能であるか実験により調査し、本手法の有効性を検討する。

## 4. 類似店舗抽出手法

### 4.1 手法概要

提案手法では、店舗ごとに取得した商品情報から店舗情報を作成し、店舗情報の類似度を計算することで類似店舗を抽出する。図1に本手法の概要図を示す。

本手法は、商品情報の取得、店舗情報の作成、類似店舗の抽出の3つのフェーズから構成される。商品情報の取得を行うフェーズでは、EC-Mallから店舗ごとの商品説明文を商品ジャンル別に取得する。店舗が扱う商品のジャンルの偏りを考慮した類似店舗を抽出するため、店舗ごとの商品数も商品ジャンル別に取得する。店舗情報を作成する

フェーズでは、店舗ごとの商品説明文に対して形態素解析を行い、名詞と形容詞をBag of Words(単語と頻度を組み合わせた情報)として抽出する。抽出した品詞にTF-IDF法を適用し、店舗の特徴語を表した形態素集合を作成する。店舗ごとの形態素集合から任意の特徴語を取り出して店舗情報を作成する。類似店舗を抽出するフェーズでは、店舗の特徴語をSet of Words(単語のみの情報)と考えてキーワードの照合を行い、店舗間の類似度を計算する。最後に、店舗が扱う商品のジャンルの偏りを利用して類似度の調整を行い、類似店舗を抽出する。商品情報の取得についての詳細を4.2節、店舗情報の作成についての詳細を4.3節、類似店舗の抽出についての詳細を4.4節で述べる。

### 4.2 商品情報の取得

EC-Mallでは店舗によって扱う商品のジャンルが大きく異なる。商品説明文に含まれるキーワードによる関連性では、似た特徴語を持つ店舗を類似店舗として抽出することはできるが、商品ジャンルの違いまで考慮することは難しいと考えられる。提案手法では、男性向けの商品を販売する店舗と女性向けの店舗が類似店舗として抽出されることは好ましくない。そこで商品ジャンル別に店舗の類似度を計算することにより、ジャンルが大きく異なる店舗が類似店舗として抽出されないようにする。提案手法では、店舗が扱う商品の商品説明文を商品ジャンル別に取得し、各店舗における商品ジャンルごとの商品数も取得する。

### 4.3 店舗情報の作成

#### 4.3.1 形態素の抽出

提案手法では、店舗の特徴語の関連性を見出すことにより類似店舗を抽出する。店舗の特徴語は店舗が扱う商品の説明文から抽出する。抽出する特徴語は店舗が扱う商品の特徴を表した語句が望ましいと考えられる。商品特徴を表す語句としては、商品名や対象ユーザ、商品の素材などの情報を表すことができる名詞(固有名詞)や、商品の質感や雰囲気などの情報を表現することができる形容詞などが考えられる。そのため本手法では、商品説明文に対して形

態素解析を行い、商品の特徴を表す語句を多く含んでいると考えられる名詞および形容詞を抽出する。形容詞に関しては様々な活用が存在するため、原型に変形したものを抽出する。抽出した語句の重要度を計算できるようにするために語句の出現回数も記録し、単語と頻度を組み合わせた情報である Bag of Words として表現する。

#### 4.3.2 形態素の重みづけ

商品説明文から抽出した形態素には、商品の特徴を表していないような不要な形態素が多く含まれる。また、商品の特徴を表しているが大半の店舗でも利用されているような形態素も多く含まれていると考えられる。これらの形態素は店舗の特徴語として好ましくない情報であると考えられる。そこで、抽出した形態素に重み付けを行い、店舗の特徴語を取り出すための情報を作成する。

形態素の重み付けには TF-IDF 法[23]を採用する。TF-IDF 法では、特定の文書における語句の出現頻度を表す  $tf$  値 (Term Frequency) と全ての文書に対する語句の網羅性を表す  $idf$  値 (Inverse Document Frequency) を利用して語句の重み (TF-IDF 値) を計算する。本手法では、店舗ごとの商品説明文を 1 つの文書とすることにより、各店舗における形態素の重みを式(1)のように求めることができる。TF-IDF 法を用いることにより、各店舗の形態素の重みは、その店舗の商品説明文に多く使われている形態素ほど大きくなり、多くの店舗で使われている形態素ほど小さくなる。つまり、多くの店舗で用いられていないような各店舗の形態素を店舗の特徴語として表現することができる。

$$w(m, s) = tf(m, s) \cdot idf(m)$$

$$tf(m, s) = \frac{n_{m,s}}{\sum_{d \in s} n_{d,s}} \quad (1)$$

$$idf(m) = \log \frac{N}{df(m)} + 1$$

$w(m, s)$  : 店舗  $s$  における形態素  $m$  の重み (TF-IDF 値),

$tf(m, s)$  : 店舗  $s$  における形態素  $m$  の  $tf$  値,

$n_{m,s}$  : 店舗  $s$  における形態素  $m$  の出現頻度,

$idf(m)$  : 形態素  $m$  の  $idf$  値,

$N$  : 全店舗数,  $df(m)$  : 形態素  $m$  が出現する店舗の数

#### 4.3.3 店舗特徴語の抽出

各店舗における重み付き形態素情報から任意の数の形態素を取り出し、店舗の特徴語として扱う。しかし、どの程度の重みを持つ形態素が店舗の特徴語として有効であるかは不明である。また、店舗によって形態素の重みの分布が異なると考えられるため、単純に閾値以上の形態素を店舗の特徴語として良いかは疑問である。そこで本稿では、条件①「重み順で上位  $X$  位までの形態素」または条件②「重みが閾値  $Y$  以上である形態素」を特徴語として扱うことにし、好ましい  $X$  値,  $Y$  値については 6 章の実験結果を通し

て検討していくこととする。ただし、TF-IDF 法による重みが 0 となった形態素に対しては、店舗の特徴を表した言葉でない判断できるため、特徴語としないことにする。

### 4.4 類似店舗の抽出

#### 4.4.1 商品ジャンル別の店舗類似度の計算

抽出した店舗の特徴語には TF-IDF 法による重みが付与されている。4.3.3 項では有効であると考えられる形態素を店舗の特徴語として抽出しているが、店舗の特徴を表していないような形態素が全て除外されたわけではない。特徴語の重みを考慮して類似店舗を抽出してしまうと、好ましくない情報の影響が大きくなってしまふ可能性がある。そこで、抽出した店舗の特徴語を単語だけの情報 (Set of Words) と見なし、式(2)のように商品ジャンル別の店舗間の類似度を計算する。店舗の特徴語を Set of Words とすることにより、好ましい特徴語と好ましくない特徴語の影響を同等に扱うことができる。好ましい特徴語が多く抽出できていれば類似店舗抽出の精度が向上すると考えられる。

$$sim(a, b, j) = \frac{n(\text{keyword}(a, j) \cap \text{keyword}(b, j))}{n(\text{keyword}(a, j) \cup \text{keyword}(b, j))} \quad (2)$$

$sim(a, b, j)$  : 商品ジャンル  $j$  における店舗  $a, b$  の類似度,

$\text{keyword}(a, j)$  : 商品ジャンル  $j$  における店舗  $a$  の特徴語,

$n(X)$  : 集合  $X$  の要素数

#### 4.4.2 店舗類似度の計算

商品ジャンル別の店舗類似度を式(3)のように統合することにより各店舗の類似店舗を抽出する。店舗によって扱う商品のジャンルの偏りが異なるため、式(3)では店舗ごとに異なる店舗類似度が抽出される。例えば、商品ジャンルの偏りが異なる店舗 A と店舗 B の類似度を計算すると、店舗 A における店舗 B の類似度、店舗 B における店舗 A の類似度はそれぞれ異なった値となる。本手法では、式(3)によって得られる類似度が高い店舗を類似店舗として扱う。

$$sim^l(a, b) = \sum_{j \in J} sim(a, b, j) \cdot per(a, j)$$

$$per(a, j) = \frac{n_{j,a}}{\sum_{k \in J} n_{k,a}} \quad (3)$$

$sim^l(a, b)$  : 店舗  $a$  における店舗  $b$  の類似度,

$sim(a, b, j)$  : 商品ジャンル  $j$  における店舗  $a, b$  の類似度,

$per(a, j)$  : 店舗  $a$  における商品ジャンル  $j$  の商品数の割合,

$n_{j,a}$  : 店舗  $a$  における商品ジャンル  $j$  の商品数,

$J$  : 対象とする商品ジャンルの集合

$N$  : 全店舗数,  $df(m)$  : 形態素  $m$  が出現する店舗の数

### 5. 適用事例

本章では提案手法の適用事例について述べる。今回、提案手法に用いる EC-Mall は、現在の日本において最も利用者の多い EC-Mall である楽天市場[24]とした。本手法に必

表1 取得した商品ジャンル一覧

Table1 The acquired product genre

商品ジャンル名	楽天ジャンル ID [c]
キッズ・ベビー・マタニティ	100533
バッグ・小物・ブランド雑貨	216131
レディースファッション	100371
インナー・下着・ナイトウェア	100433
メンズファッション	551177
靴	558885

表2 店舗の特徴語抽出に用いるパラメータ

Table2 Parameters for the feature word extraction.

条件内容	条件	略称
条件①：重み順で上位 X 位までの形態素を特徴語として抽出する。	X = 100	N100
	X = 300	N300
	X = 500	N500
	X = 700	N700
	X = 900	N900
条件②：重みの閾値が Y より大きい形態素を特徴語として抽出する。	Y = 0.0	W00
	Y = 0.2	W02
	Y = 0.4	W04
	Y = 0.6	W06
	Y = 0.8	W08
	Y = 1.0	W10

要な商品情報は、楽天株式会社が提供している楽天 API[25]を用いることで取得した。また、本稿ではアパレル店舗を対象としているため、取得する商品ジャンルは6種類(表1)に限定した。対象とする店舗については、表1の商品ジャンルにおいて販売ランキング1位~1020位までの商品を販売している店舗を重複しないように取得し、得られた1033店舗を用いた。提案手法における形態素抽出には、形態素解析エンジンである MeCab[27]を利用した。

### 5.1 類似店舗抽出結果

提案手法の適用事例を図2に示す。図2では、女性向け店舗を2件、男性向け店舗を1件の適用事例を示している。ベース店舗の類似店舗を上位2件まで挙げ、ベース店舗との特徴語のマッチ率を明記した。図中の商品画像は、筆者らが楽天市場の店舗ページ(TOPページ)を閲覧することにより、店舗の個性を表していると感じた商品を抽出した。図2の(a)(b)では、レディースファッションを扱う異なる店舗が類似店舗として抽出されている。(c)のメンズファッションを扱う店舗の例では、メンズファッションを扱う店舗が類似店舗として抽出されていることが分かる。店舗間の特徴語マッチ率については、店舗によってバラつきがあることが読み取れる。



(a) レディースファッションを主に扱う店舗の類似店舗①



(b) レディースファッションを主に扱う店舗の類似店舗②



(c) メンズファッションを主に扱う店舗の類似店舗③

図2 提案手法の適用事例

Figure2 Application case of the proposed method.

### 5.2 特徴語マッチ率

提案手法では、店舗の特徴語を抽出する際に任意に変更できるパラメータが存在する(4.3.3項を参照)。本稿では、パラメータの変化によって類似度の分布がどのように変化するか調査するため、11パターンのパラメータを用意した(表2)。以後、各パラメータ条件は表2における略称を用いる。これらの条件別に提案手法の実験を行うことにより、提案手法における好ましいパラメータについて考察する。

### 6. 評価実験

抽出される類似店舗がユーザの好みによる繋がりを表現できているか検討するため、2種類の実験を行った。提案手法の実装要件については5章(適用事例)で述べたものと同様にした。今回の実験は、ECサイトを2,3ヶ月に1回以上利用する20代の男女14名(男性8名、女性6名)に対して行った。複数の実験による疲労などの影響を考慮し、各実験おきにユーザの疲労具合を口答で適宜確認した。また、実験中に疲労を感じた場合は適宜報告するように指示をした。実験中に男性2名が疲労を感じたことを報告したため、その2名を除外した12名(男性6名、女性6名)の実験結果を本章では示す。

#### 6.1 店舗の嗜好判断に関する実験

本手法では商品情報を用いて類似店舗を抽出しており、商品の好みは店舗の好みに影響することを前提としている。

c) 楽天APIを通して商品を取得する際に用いる商品ジャンル番号。

そこで、まず商品の好みと店舗の好みの関連性の有無について確認する実験を行う。具体的には、好みと感ずる店舗に好みと感ずる商品の数が関係しているか調査し、好みの商品と好みの店舗の関係性の有無について検討する。

### 6.1.1 実験内容

本実験では、ランダムに選定した店舗の商品を表示する評価用システムを用いた(図3)。評価用システムでは、画面スクロールなどの無駄な処理を減らすため、表示する商品数を15商品として1画面に収まるように調整した[27]。被験者には、欲しいと感ずる商品を選択してもらい、その店舗が好みかどうか評価してもらった。店舗名や商品価格などの影響をなくすため、店舗名の表記はせずに商品画像のみ表示した。リピートしたいと思う店舗を好みの店舗とするように被験者に指示した。店舗の好みに関する評価値には、「嫌い(1点)」、「やや嫌い(2点)」、「やや好き(3点)」、「好き(4点)」の4件法を採用し、曖昧な回答が選ばれないように中央値を外した[29]。各評価値におけるデータを取得するため、被験者には全ての評価値を最低3つ選ぶまで実験をしてもらった。

### 6.1.2 実験結果

6.1.1項の実験結果を図4に示す。図4において、グラフの横軸は掲示した店舗に対する好みの評価値を表し、縦軸は各評価値における被験者が好みと感ずる商品の平均数を表している。この2つの要因に相関があるか分析するため、ノンパラメトリック法によるスピアマンの順位相関係数を求めた。その結果、相関係数が0.895 ( $p < 0.01$ )であったため、この2つの要因間に正の相関があることが示された。従って、店舗に好み(欲しい)と感ずる商品が多く掲示されていると、その店舗を好み(再び訪れたい)と感ずるようになることが示された。また、図4における好み(4点)と感ずるために必要な商品数に大きなバラつきがあることから、この2つの要因の関係には大きな個人差があることが読み取れる。

ランダムに店舗を表示した場合における好みの店舗の発見率を図5に示す。図5のグラフにおいて、縦軸は店舗の発見率を表しており、エラーバーの長さは各データの標準偏差を表している。店舗の評価値のうち「嫌い(1点)」、「やや嫌い(2点)」をネガティブな評価とし、「やや好き(3点)」、「好き(4点)」をポジティブな評価とした。図5における店舗の発見率は、各評価の割合を被験者ごとに平均した値を表している。

ランダムで店舗を表示した場合に、ポジティブな評価とネガティブな評価に有意差があるか分析するため、1対の標本によるt検定を実施した。その結果、両側検定におけるp値が0.025 ( $p < 0.05$ )であったため、有意差があることが示された。従って、図5よりランダムで店舗を表示した場合は、ネガティブな評価となる店舗が多いことが分かっ



図3 評価実験用システム  
Figure3 The experiment system

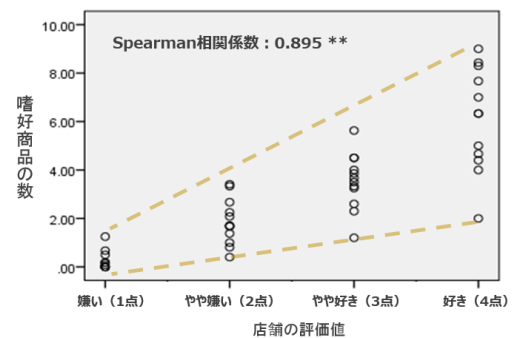


図4 嗜好商品と嗜好店舗の関係 (\*\*  $p < 0.01$ )  
Figure4 Relationship of preference items and stores.

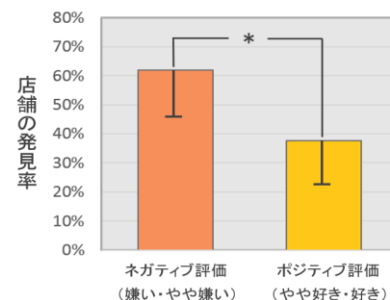


図5 ランダム店舗表示における店舗発見率 (\*  $p < 0.05$ )  
Figure5 Store discovery rate in random display.

た。今回の実験において、ポジティブな評価となる店舗の発見率は37.8%であった。

## 6.2 類似店舗抽出の精度に関する実験

抽出した類似店舗がユーザの好みを反映できているか確認する実験を行う。提案手法では推薦精度に影響するパラメータ(4.3.3項を参照)が存在するため、本実験では表2のパラメータ条件を設定して実験を行う。

### 6.2.1 実験内容

本実験では、任意の店舗の類似店舗を掲示する評価用システムを用いた。システムの外見は6.1節で用いた評価用システム(図3)と同じにした。本実験を行う前に、被験者には楽天市場から好み(リピートしたい)と感ずる店舗を選定してもらい、著者らが評価用システムにその店舗を設

定した。実験は 6.1 節と同様、好みの商品を選択してもらい、店舗の好みについて評価してもらった。また、表 2 の各パラメータ条件で抽出される類似店舗を比較するため、各パラメータにおける類似店舗を上位 5 件ずつ表示し、好みの店舗の発見率について比較した。

### 6.2.2 実験結果

各パラメータ条件における好みの店舗の発見率を図 6 に示す。縦軸は好みの店舗の発見率を表しており、エラーバーの長さは標準偏差を表している。また、本手法における好みの店舗の発見率とランダム店舗表示における発見率を比較するため、図 5 における好みの店舗の発見率も RAND として図 6 に含めた。各パラメータにおける店舗の発見率に有意差があるか分析するため、被験者内計画における 1 要因の反復測定による分散分析を行った。今回の分散分析には SPSS Statistics (Ver.2.2) を用いた。

分散分析の結果、 $F(3.893, 42.825) = 11.272, p < 0.01$  であり、有意な主効果があることが分かった。有意な主効果があるため Bonferroni の手法による多重比較を行い、条件①とランダム店舗表示の間に有意水準 1% で有意差が認められた。また、N700 は W08, W10 とも有意差が認められた。

実験結果 (図 6) から、ランダムに店舗を表示した場合より、「TF-IDF 法の重み上位 X 位までの形態素を店舗の特徴語とする」という条件①の提案手法が類似店舗を表示した場合の方が好みの店舗を発見しやすいことが示された。また、条件①における N700 のみが発見率の低い W08 および W10 と有意差があることから、今回の実験環境下では N700 の発見率が特に高いことが示された。条件②はランダムに店舗を表示した場合と有意差が見られなかったため、類似店舗抽出がうまく行っていないことが分かる。

## 7. 考察

### 7.1 店舗の嗜好判断基準について

6.1 節の実験結果では、好みの店舗の判断基準に好みの商品数が影響することが示された。嗜好店舗の評価値に対する好みの商品数を表す図 4 より、好みの商品数が多くなるほど、店舗を好みと感じやすくなることが分かった。また、好みの店舗と感じるために必要な商品数は人によってバラつきがあることも分かった。提案手法では商品情報による繋がりをを用いて類似店舗を抽出しており、ユーザの商品嗜好と店舗嗜好に関連性があることを前提としていた。これらの知見はこの前提条件を肯定するものであり、ユーザの好みの店舗を推薦することが好みの商品を推薦することに繋がると考えられる。ただし、店舗が扱う商品数は膨大であるため、店舗の商品の掲示方法については別途考える必要があると考えられる。

図 5 の結果より、ランダムに店舗を掲示した場合は好みでない店舗が好みの店舗より多く発見されてしまうことが示された。今回の実験で対象とする店舗は、楽天市場にお

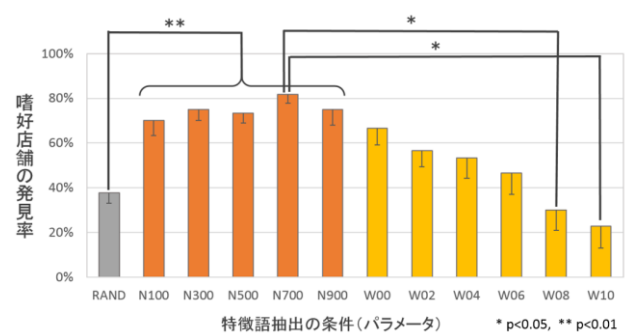


図 6 特徴語抽出の条件と嗜好店舗の発見率

Figure6 Discovery rate of preference stores and conditions of the feature word extraction.

ける販売ランキングが高い店舗としている。この結果は、多くの「人気店舗」がユーザの好みを反映していないことを示しており、好みの店舗の発見率を向上させるための支援が必要であることを示唆していると考えられる。

### 7.2 類似店舗抽出の精度について

6.2 節では、提案する類似店舗抽出手法がユーザの嗜好店舗をどの程度抽出できるか調査する実験を行った。また、6.1 節ではランダムに店舗を見ていった時の嗜好店舗の発見率についても調査した。それらの結果、ユーザはランダムに店舗を回遊する場合より提案手法が推薦した店舗を回遊する方が好みの店舗を発見しやすいことが示された。今回の実験では店舗の特徴語 (類似店舗を抽出するための情報) の選定を 2 通りの基準で決定しており (4.3.3 項を参照)、好みの店舗の発見率が優位に向上するのは「TF-IDF 法の重み上位 X 位までの形態素を店舗の特徴語とする」という条件①のみであることも示された。

図 6 から、 $X=700$  の嗜好店舗の発見率が  $X=100, 300, 500, 900$  の時より高い傾向にあることが分かる。この結果は、TF-IDF 法による重みが大きい形態素が必ずしも店舗の特徴語として有効であるわけではないことを示していると考えられる。また、重みが低すぎる形態素も店舗の特徴語としては適していないと考えられる。

「TF-IDF 法の重みが閾値 Y 以上である形態素を店舗の特徴語とする」という条件②に対しては優位差が見られなかった。店舗ごとに抽出した形態素の重みにはバラつきがあると考えられ、閾値を基準とした特徴語は類似店舗抽出の精度を不安定にしていると考えられる。本手法の TF-IDF 法によって重みづけられた形態素は重みの絶対値に意味はなく、重みによって得られた順番に意味があるといったことが考えられる。また、重み順における上位と下位の形態素より、中位の形態素の方が店舗の特徴語に向いていると考えられる。

### 7.3 店舗推薦システムに向けた拡張性について

提案手法は、幅広い商品探索を支援するための店舗推薦システムに適用できることを想定している。7.1 節における

考察では、好みの店舗を推薦するシステムがユーザの幅広い商品探索を支援できることを述べた。7.2節では、ランダムに店舗を回遊していくより、提案手法が推薦する店舗を回遊した方が嗜好店舗を発見しやすいことについて述べた。ランダムな店舗回遊は従来 EC-Mall における店舗検索システムの問題点であることが指摘されている[22]ため、この知見は提案手法が従来の店舗検索より嗜好店舗の発見性が高いことを示唆していると考えられる。7.2節では、本手法における店舗特徴語抽出の好ましいパラメータについても考察し、本手法を店舗推薦システムに適用する際の重要な知見が得られた。今回提案した手法では、一般的な構文解析によく利用される単語辞書を利用していないという利点もある。商品説明文には個性豊富な固有名詞（例えば、商品名など）が多く存在し、それは時期刻々と増大していくと考えられる。単語辞書を使わない手法は時代の変化に柔軟に対応できる可能性があると考えられる。

## 8. まとめと今後の課題

本稿では、EC-Mall における探索型検索の重要性に着目し、店舗単位の幅広い商品をユーザに推薦することができる類似店舗抽出手法を提案した。本手法では、店舗ごとに取得した商品説明文から特徴語リストを作成し、店舗間の特徴語の関連性を計算することによって類似店舗を抽出した。本手法は店舗が扱う商品ジャンルの違いを考慮した類似店舗を抽出することができ、単語辞書を利用せずに店舗の特徴語を抽出することが可能である。

評価実験では、店舗をランダムに選んだ場合と提案手法が選んだ場合の好みの店舗の発見率を比較し、提案手法の方が高い発見率となることを確認した。また、提案手法のパラメータ変化に対する発見率の変化を調査し、提案手法を実際の店舗推薦システムに応用するための知見を得た。

今後の課題としては、価格やレビューといった情報も考慮できる手法の提案、探索型検索を促すような店舗・商品探索インタフェースの提案などが考えられる。今後も、ユーザ、店舗、EC-Mall の多角的な視点で店舗推薦システムについて検討する必要があると考えられる。

## 参考文献

- 1) White, R. W. and Roth, R. A.: *Exploratory Search: Beyond the Query-Response Paradigm*, Morgan and Claypool (2009).
- 2) Marchionini, G.: Exploratory search: from finding to understanding, *Communication of the ACM*, Vol.49, No.4, pp.41-46 (2006).
- 3) Riecken, D.: Personalized Views of Personalization, *Communication of the ACM*, Vol.43, No.8, pp.26-158 (2000).
- 4) Su, X. and Khoshgoftar, T. M.: A Survey of Collaborative Filtering Techniques, *Advances in Artificial Intelligence*, vol.2009, No.4, pp.1-19 (2009).
- 5) 神嶋敏弘：推薦システムのアルゴリズム（1），人工知能学会誌，Vol.22，No.6，pp.826-837（2007）。
- 6) 神嶋敏弘：推薦システムのアルゴリズム（2），人工知能学会誌，Vol.23，No.1，pp.89-103（2008）。

- 7) Hijikata, Y., Shimizu, T. and Nishida, S.: Discovery-oriented collaborative filtering for improving user satisfaction, *Proc. the 14th international conference on intelligent user interfaces*, ACM, pp.67-76 (2009).
- 8) Cho, J., Kwon, K. and Park, Y.: Collaborative filtering using dual information sources, *Intelligent Systems, IEEE*, Vol.22, No.3, pp.30-38 (2007).
- 9) Hwang, S. Y. and Chen, L. S.: Using trust for collaborative filtering in eCommerce, *Proc. the 11th International Conference on Electronic Commerce*, ACM, pp.240-248 (2009).
- 10) Murakami, T., Mori, K. and Orihara, R.: Metrics for evaluating the serendipity of recommendation lists, *New Frontiers in Artificial Intelligence*, pp.40-46 (2008).
- 11) Ziegler, C. N., et al.: Improving recommendation lists through topic diversification, *Proc. the 14th Int'l conference on World Wide Web*, ACM, pp.22-32 (2005).
- 12) Linden, G., Smith, B. and York, J.: Amazon.com recommendations: Item-to-item collaborative filtering, *IEEE Internet Computing*, IEE-E, Vol.7, No.1, pp.76-80 (2003).
- 13) Das, A. S., Datar, M. and Garg, A.: Google news personalization: Scalable online collaborative filtering, *Proc. the 16th Int'l Conf. on World Wide Web*, ACM, pp.271-280 (2007).
- 14) 土方嘉徳：嗜好抽出と情報推薦技術，情報処理学会論文誌，Vol.48，No.9，pp.972-978（2007）。
- 15) Mooney, R. J. and Roy, L.: Content-based book recommending using learning for text categorization. *Proc. the fifth ACM conference on Digital libraries*, ACM, pp.195-204 (2000).
- 16) Burke, R.: Hybrid Recommender Systems: Survey and Experiments, *Modeling and User-Adapted Interactions*, Vol.12, No.4, pp.331-370 (2002).
- 17) Cao, Y. and Li, Y.: An intelligent fuzzy-based recommendation system for consumer electronic products, *Expert Systems with Applications*, Vol.33, No.1, pp.230-240 (2007).
- 18) Parikh, N. and Sundaresan, N.: Inferring Semantic Query Relations from Collective User Behavior, *Proc. the 17th ACM conference on information and knowledge management*, ACM, pp.349-358 (2008).
- 19) Schraefel, M. C., et al.: mSpace: Improving information access to multimedia domains with multimodal exploratory search, *Communication of the ACM*, Vol.49, No.4, pp.47-49 (2006).
- 20) Cutrell, E., et al.: Fast, flexible filtering with phlat, *Proc. the SIGCHI conference on Human Factors in computing systems*, ACM, pp.261-270 (2006).
- 21) Bhattacharya, S., Gollapudi, S. and Munagala, K.: Consideration set generation in commerce search, *Proc. the 20th international conference on World Wide Web*, ACM, pp.317-126 (2011).
- 22) 大河原一輝，平野廣美，益子宗ほか，ショッピングモール型 EC サイトのための店舗情報視覚化システム，情報処理学会論文誌，Vol.56，No.3（2015）（in press）。
- 23) Salton, G.: *Automatic Text Processing: The Transformation, Analysis, and Retrieval of*, Addison-Wesley (1989).
- 24) 楽天市場：楽天市場，楽天株式会社（オンライン），入手先〈<http://www.rakuten.co.jp>〉（参照 2015-01-10）。
- 25) 楽天ウェブサービス：楽天商品検索 API，楽天株式会社（オンライン），入手先〈<http://webservice.rakuten.co.jp/api/itemsearch>〉（参照 2015-01-10）。
- 26) Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, *Proc. the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004)*, pp.230-237 (2004).
- 27) Cosley, D., et al.: Is seeing believing?: how recommender interfaces affect users' options, *Proc. the SIGCHI Conf. on Human Factors in Computing Systems*, ACM, pp.585-592 (2003).
- 28) 新目真紀，市川照久，片岡信弘ほか：量的評価のガイドライン，情報システムと社会環境研究会，情報システム有効性評価手法研究分科会，第 1.1 版（2012）。