

情報推薦におけるコンテキスト依存型ランキング方式の提案

奥 健太[†] 中島 伸介[†] 宮崎 純[†] 植村 俊亮^{††} 加藤 博一[†]

[†] 奈良先端科学技術大学院大学情報科学研究科

^{††} 奈良産業大学情報学部情報学科

あらまし 本稿では、コンテキスト依存型情報推薦システムに対し、有効なランキング方式を提案する。コンテキスト依存型情報推薦システムは、ユーザの嗜好およびコンテキストに適合するコンテンツをユーザに推薦するものである。ただし、推薦候補として判定されたコンテンツが膨大に存在する場合、その中から、欲しいものをユーザ自信が見つけ出す必要が生じるため、ユーザにとって負担となる。ゆえに、推薦候補に対し、ユーザの好みの程度に応じてランキング処理を施すことが重要である。ユーザの好みの程度は、コンテキストによっても変わってくるため、推薦候補に対するランキングそのものも、やはりコンテキスト依存となる。提案方式では、ユーザがコンテンツの各特徴を、どの程度重視しているかを、コンテキストごとに学習する。この学習成果に基づいて、推薦候補の各コンテンツのランキングスコアを算出する。そして、このランキングスコアを基準に、推薦候補に対しランキング処理を行う。さらに、単純なランキング方式との比較による評価実験を行い、提案方式の有効性を確認した。

キーワード 情報推薦, コンテキスト依存型情報推薦, Context-Awareness, ランキング

A Proposal of Context-Aware Ranking Method for Information Recommendation

Kenta OKU[†], Shinsuke NAKAJIMA[†], Jun MIYAZAKI[†], Shunsuke UEMURA^{††}, and Hirokazu KATO[†]

[†] Graduate School of Information Science, Nara Institute of Science and Technology

^{††} Department of Informatics, Faculty of Informatics, Nara Sangyo University

Abstract In this paper, we propose a context-aware result ranking method for use in our SVM based context-aware restaurant recommendation system. In addition to being important when selecting recommendations, user context is also very important when ranking the selected results. Thus, we propose a result ranking system which uses user feedback to determine the user's current context. From this, the system accommodates this context by adjusting the weights of the various restaurant parameters and then finally re-ranking the results accordingly. Our experiments have shown that our method is effective in providing relevant results to end-users.

Key words Information Recommendation, Context-Aware Recommendation, Context-Awareness, Ranking

1. はじめに

我々は、これまでに、「人の嗜好は、そのときおかれているコンテキスト（時間帯や場所、同伴者など）によって変化する」という考えに基づき、コンテキスト依存型情報推薦システムを提案してきた [1] [2]。ユーザへの推薦候補の判定には、2クラス分類手法である Support Vector Machine (SVM) [3] を用いている。この SVM により、ユーザにとって「満足」であると判定されたコンテンツをユーザに推薦する。

しかし、推薦候補として判定されたコンテンツが膨大にあった場合、その中から、欲しいものをユーザ自信が見つけ出す必

要が生じるため、ユーザにとって負担となる。ゆえに、推薦候補に対し、ユーザの好みの程度に応じてランキング処理を施すことが重要である。

ランキング処理に対する基本的な考えを、飲食店情報推薦を例に説明する。たとえば、あるユーザは「安さ」を重視して飲食店を選ぶとする。すると、推薦システムは推薦候補の中から「平均予算」が安い飲食店を見つけ、それらを上位に提示する。しかしながら、ユーザが何を重視して飲食店を選ぶか、といったことは、やはりそのときのコンテキストによって変わってくる。一人でいるときには、「安さ」を重視して飲食店を選ぶが、恋人と一緒にいるときには、「安さ」には拘らず、特に「雰囲気

気」を重視する、といった例が挙げられる。さらには、一人では、「平均予算」が「1,000円未満」の安い店に満足するが、恋人と一緒にいるときには、平均予算が「3,000~5,000円」と比較的高級な店に満足する、といった例も挙げられる。

本研究では、「ユーザが何を重視してコンテンツを選択しているかは、そのときのユーザのコンテキストに依存する」という考えに基づき、コンテキスト依存型ランキング方式を提案する。本方式では、ユーザがコンテンツの各特徴をどの程度重視しているかを、コンテキストごとに学習する。この学習成果に基づいて、推薦候補の各コンテンツのランキングスコアを算出する。そして、このランキングスコアを基準に、推薦候補に対しランキング処理を行う。さらに、単純なランキング方式との比較による評価実験を行い、提案方式の有効性を確認する。

2. 関連事項

2.1 コンテキスト依存型情報推薦システム

我々は、これまでに、コンテキスト依存型ユーザ嗜好モデリング手法(図1)を提案してきた[1][2]。この手法は、2クラス分類手法であるSVM[3]を用い、コンテンツに対するユーザの好み(「満足」・「不満足」)を学習するものである。この手法では、「同じコンテンツであっても、評価時のコンテキストが異なれば、異なるデータ」として扱っている。具体的には、多次元の特徴パラメータで表現されたコンテンツデータに対し、評価時のコンテキストの特徴パラメータを統合したものを、学習データとして扱っている。本研究では、飲食店の特徴パラメータは、「ジャンル」や「雰囲気」、「平均予算」などの28次元、コンテキストの特徴パラメータは、「時刻」や「天気」、「予算」、「同伴者」などの24次元で表現されている。この手法により、コンテキストごとに異なるユーザの好みを適切に学習させることを可能にした。

我々は、また、この手法に基づいたコンテキスト依存型情報推薦システムを実装した。このシステムにより、ユーザが満足すると推測されたコンテンツデータを、推薦候補として扱う。本研究で提案するランキング方式は、この得られた推薦候補に対して適用されるものである。

ユーザのコンテキストに依存したアプリケーションに関して、Context-Awareアプリケーションに関連した研究[4]が多く行われている。Shopping Assistant[5]では、店内の買物客の位置情報に基づいて、商品情報を提示する。Cyberguide[6]では、現在時刻や旅行者の位置情報に基づいて、観光ガイドなどを提示する。しかし、天気や同伴者、予算、ユーザの目的、気分など、ユーザの嗜好に直接影響を及ぼすようなコンテキストを考慮したアプリケーションはあまり見当たらない。

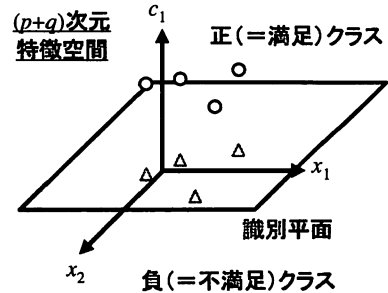
2.2 単純なランキング方式

推薦候補に対する単純なランキング方式として、下記のものが挙げられる。

- a) 多クラス分類に基づくランキング方式
- b) SVMの識別平面からのユークリッド距離に基づくランキング方式

SVMは、本来、2クラスを対象とした分類手法であるが、 n

コンテキストごとに評価が異なるデータを分離可能



x_1, x_2, \dots, x_p : コンテンツデータの p 次元特徴パラメータ軸
(図は2次元の場合:e.g. ジャンル, 雰囲気, 平均予算)

c_1, c_2, \dots, c_q : コンテキストの q 次元特徴パラメータ軸
(図は1次元の場合:e.g. 時刻, 天気, 予算, 同伴者)

図1 コンテキスト依存型ユーザ嗜好モデリング

段階に分けて学習させることにより、 $n+1$ クラスを対象とした分類が可能となる。a)の方法により、たとえば5段階の満足度評価値(不満足1-2-3-4-5満足)に対応して、5クラスの分類を行う。このとき、最も評価値の高いクラスから上位にランク付けすることで、ランキング処理を行う。

また、SVMでは、2クラスを分類するような識別平面を構成する。b)は、正クラスに属するデータのうち、この識別平面からのユークリッド距離が遠いものほど、上位にランク付けすることで、ランキング処理を行う。これは、識別平面付近は、正負の判定が曖昧な領域であるのに対し、遠くに位置するほど、その判定が明瞭なものになる、という考えに基づく。

3. コンテキスト依存型ランキング方式

3.1 基本概念

図2は、例として、ある1次元の特徴パラメータに着目したときの、あるコンテキストにおける学習データの正事例および負事例の分布を示したものである。横軸は、その特徴パラメータの特徴量 x を示し、 $\{x: 0 \leq x \leq 1\}$ の連続値で表現される。正事例はユーザにとって満足なデータを意味し、負事例はユーザにとって不満足なデータを意味する。

(a)の場合では、正事例が $\{0 \leq x < 0.3\}$ の範囲で、負事例が $\{0.3 \leq x < 0.7\}$ の範囲で、それぞれの分布が偏っている。このような偏ったデータ分布から、この特徴パラメータはユーザの満足・不満足の評価に影響を及ぼしていることが分かる。

一方、(b)の場合では、正事例と負事例とが混在している。したがって、このような混在したデータ分布から、この特徴パラメータはユーザの満足・不満足の評価には何ら影響を与えていないことが分かる。以上の例をふまえ、各特徴パラメータに着目したときのデータ分布について、以下の仮説を立てる。

仮説I 正事例(負事例)の分布が偏っている。⇒その特徴パラメータはユーザの満足(不満足)評価に影響を及ぼしている。

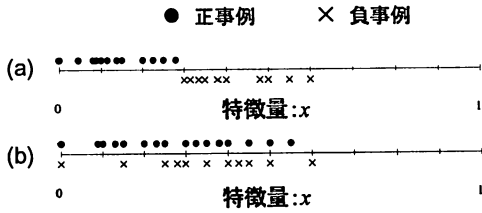


図2 ある1次元の特徴パラメータに着目したときの、学習データの正事例および負事例の分布例

仮説 II 正事例と負事例とが混在している。⇒ その特徴パラメータはユーザの満足・不満足の評価には影響を及ぼしていない。

以上のように、各特徴パラメータに着目したときの正事例と負事例の分布の偏りを調べることで、その特徴パラメータをユーザがどの程度重要視しているか（すなわち、その特徴パラメータに対するユーザの評価基準の重み）を知ることができる。この各特徴パラメータに対する評価基準の重みをコンテキストごとに学習し、それに基づいて、各コンテンツデータのランキングスコアを算出することにより、コンテキスト依存型のランキング処理が可能になる。

なお、図2は、飲食店の特徴パラメータである「平均予算(0~10,000円)」に着目したときの、あるユーザの実データに基づくものである。(a)、(b)は、それぞれ別のユーザのあるコンテキストにおける学習データの分布を示したものである。実際、(a)のユーザは、このコンテキストでは、「平均予算」が3,000円以内の飲食店を好む傾向にあった。一方で、(b)のユーザは、このコンテキストでは、「平均予算」は特に重要視していなかった。これらのことから、上記の仮説は、事実に基づいているといえる。

3.2 コンテンツ特徴パラメータに対するユーザの評価基準の重みの学習

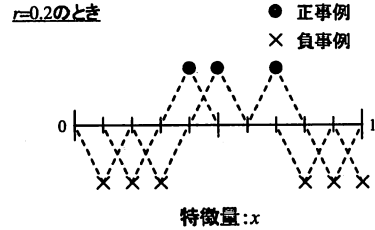
本節では、コンテンツの各特徴パラメータに着目したとき、あるコンテキストにおける学習データ集合から、その特徴パラメータに対するユーザの評価基準の重みを学習する方法について述べる。3.1節で立てた仮説に基づいた以下の方式を提案する。

方式 I データ分布が仮説 I に当てはまる。⇒ その特徴パラメータに対する正(負)の評価基準の重みを大きくする ($w \rightarrow +1$ ($w \rightarrow -1$))。

方式 II データ分布が仮説 II に当てはまる。⇒ その特徴パラメータに対する正負の評価基準の重みを“0”にする ($w = 0$)。

ここで、 w は、特徴パラメータに対するユーザの評価基準の重みを表し、この値が“+1” (“-1”) に近いほど、その特徴パラメータがユーザの満足(不満足)評価に、より影響を及ぼすことを意味する。

さらに、この特徴パラメータに対する評価基準の重みは、その特徴パラメータの特徴量に依存するものである。たとえば、図2(a)に示す x 軸が、飲食店の特徴パラメータである「平



$m^{++}=2$	$m_{max}^{++}=3$	$p^{++}=0.667$	$p^{+-}=0.389$
$m^{-}=6$	$m_{max}^{-}=15$	$p^{-}=0.400$	$p^{--}=0.122$
$m^{+}=5$	$m_{max}^{+}=18$	$p^{*}=0.278$	

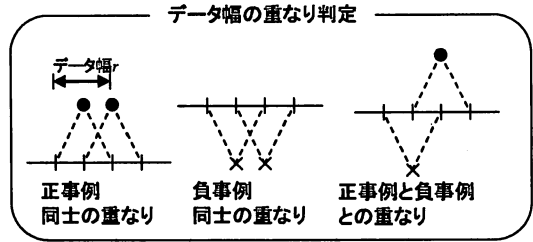


図3 正事例および負事例の分布の偏りの調査

均予算(0~10,000円)」を表すとした場合を考える。図2(a)から、このユーザは、「平均予算」が3,000円未満の飲食店は好むが、3,000円以上の飲食店は好まないといえる。このように、特徴パラメータに対する評価基準の重みが、どのようにその特徴量に依存しているかを調べる必要がある。

本手法は、大きく以下の2つの手順で構成される。

- (1) 特徴パラメータに対する正負の評価基準の重みの算出
- (2) 特徴パラメータの特徴量に依存した重みの算出

3.2.1 特徴パラメータに対する正負の評価基準の重みの算出

本項では、特徴パラメータに対する正負の評価基準の重みの算出方法を、図3を用いて説明する。図3に示すように、3個の正事例および6個の負事例が分布している。

step 1 いま、各データ同士の重なりを調べるために、仮想のデータ幅 r ($r: 0 \leq r \leq 1$) を持たせる。そのときのデータ幅の重なるの程度に基づいてデータ分布の偏りを調べる。(図3は $r = 0.2$ のときの例を示している)。

step 2 図3に示す重なり判定を参考に、

$$\begin{aligned}
 m^{++} &: \text{正事例同士の重なり数} \\
 m^{--} &: \text{負事例同士の重なり数} \\
 m^{+-} &: \text{正事例と負事例との重なり数}
 \end{aligned}$$

を、それぞれ求める。ここで、 $r = 1$ のときの各重なり数を、最大重なり数とし、それぞれ、 m_{max}^{++} 、 m_{max}^{--} 、 m_{max}^{+-} とする。

step 3 この最大重なり数に対するそれぞれの重なり数の割合 p^{++} 、 p^{--} 、 p^{+-} を、次式により求める。

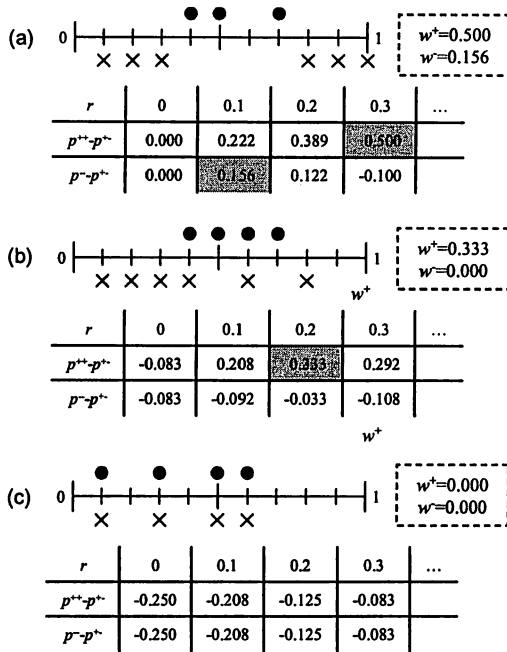


図4 特徴パラメータに対する正負の評価基準の重みの算出例

$$p^{++} = m^{++} / m_{\max}^{++}$$

$$p^{--} = m^{--} / m_{\max}^{--}$$

$$p^{+-} = m^{+-} / m_{\max}^{+-}$$

データ幅 r が大きいほど、これらの割合 p^{++} , p^{--} , p^{+-} も大きくなる。データ幅 r が小さいうちに、 p^{++} (p^{--}) が大きいということは、それだけ、正事例 (負事例) の分布が偏っているということを示す。一方、 p^{+-} が大きいということは、それだけ、正事例と負事例とが混在しているということを示す。

step 4 $\{r : 0 \leq r \leq 1\}$ の範囲において、 p^{++} (p^{--}) と p^{+-} との差 $p^{++} - p^{+-}$ ($p^{--} - p^{+-}$) を最大とするデータ幅 r^{++} (r^{--}) を見つける。ここで、 $p^{++} - p^{+-}$ ($p^{--} - p^{+-}$) は、特徴パラメータに対する評価基準の正 (負) の重みを表しているといえ、

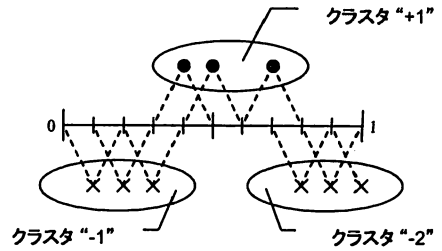
$r = r^{++}$ のとき

$$w^+ = \begin{cases} p^{++} - p^{+-} & (p^{++} - p^{+-} > 0 \text{ のとき}) \\ 0 & (p^{++} - p^{+-} \leq 0 \text{ のとき}) \end{cases}$$

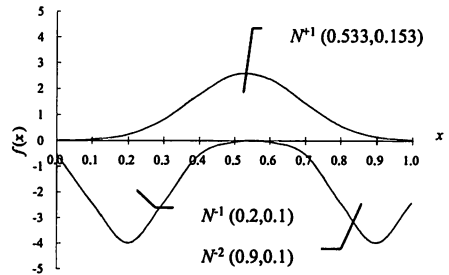
$r = r^{--}$ のとき

$$w^- = \begin{cases} p^{--} - p^{+-} & (p^{--} - p^{+-} > 0 \text{ のとき}) \\ 0 & (p^{--} - p^{+-} \leq 0 \text{ のとき}) \end{cases}$$

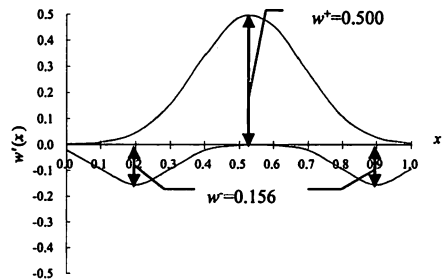
とおける。ここで、 w^+ は正の評価基準の重みを意味し、この値が大きければ、この特徴パラメータはユーザの満足評価に大きく影響を及ぼしているといえる。一方、 w^- は負の評価基準の重みを意味し、この値が大きければ、このパラメータはユーザの不満足評価に大きく影響を及ぼしているといえる。



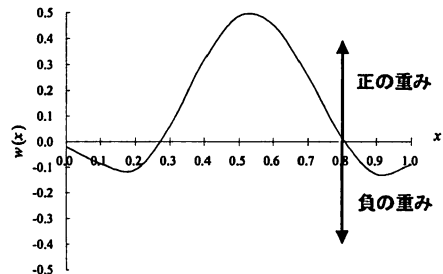
step 1: クラスタを作成する。



step 2: 各クラスタに対する正規分布を算出する。



step 3: 正規分布を正規化する。



step 4: すべての分布を合成する。

図5 特徴パラメータ値に依存した重みの算出

図4は、例として (a) および (b), (c) の3つのデータ分布に対し、正負の重み w^+ , w^- を算出した例である。この例からも、 w^+ および w^- について、上記のことが確認できる。

3.2.2 特徴パラメータの特徴量に依存した重みの算出

図5は、図3の例に対して、特徴パラメータの特徴量に依存した重みを算出する過程を示したものである。以下、この図5を用いて説明する。

step 1 $r = r^{++}$ ($r = r^{--}$) のとき、重なり合っている正事例 (負事例) から構成されるクラスタを作る。図5の場合では、図に示すような3つのクラスタ (クラスタ "+1" およびクラス

タ“-1”, クラスタ“-2”が作られる。

step 2 データ分布が正規分布に従うと仮定し、各クラスタに対する正規分布を求める。図5の場合では、図に示すような3つの正規分布 ($N^{+1}(0.533, 0.153)$ および $N^{-1}(0.2, 0.1)$, $N^{-2}(0.9, 0.1)$) が得られる。ここでは、便宜上、正事例によるクラスタの正規分布を正方向に、負事例によるクラスタの正規分布を負方向に示している。

step 3 正事例(負事例)によるクラスタの正規分布の最大値を w^{+} (w^{-}) に置き換えることにより、正規分布を正規化する。

step 4 すべての分布を合成する。すると、図5のstep 4のような分布が得られる。

この合成された分布を、特徴パラメータに対する重み分布 $w(x)$ と定義する。 $w(x)$ は以下の性質を持つ。

$w(x) > 0$ ($w(x) < 0$) \Rightarrow 特徴量 x はユーザの満足(不満足) 評価に影響を及ぼす。

$w(x) = 0 \Rightarrow$ 特徴量 x はユーザの満足・不満足の評価には影響を及ぼさない。

たとえば、図5に示す x 軸が、飲食店の特徴パラメータである「平均予算(0~10,000円)」を表すとした場合を考える。図5の例は、あるコンテキストにおいて、「このユーザは「平均予算」が3,000円以上8,000円未満である飲食店に満足し、「平均予算」が3,000円未満もしくは8,000円以上である飲食店には不満足となる」ことを表している。

3.3 各特徴パラメータに対する重み分布に基づくランキングスコアの算出

特徴パラメータに対する重み分布に基づいて、推薦候補の各コンテンツのランキングスコアを算出する。

いま、コンテンツの特徴パラメータ i に着目したときの、その特徴パラメータに対する重み分布が、 $w_i(x) (0 \leq x \leq 1)$ で与えられていたとする。また、ランキングスコアを求めたいコンテンツを k とし、そのコンテンツの特徴パラメータ i が、 x_{ki} で与えられていたとする。すると、重み分布を考慮したときの、コンテンツ k のランキングスコア s_k は次式で表される。

$$s_k = \sum_i w_i(x_{ki}) \quad (1)$$

このランキングスコアを推薦候補の各コンテンツに対して求める。最後に、各コンテンツのランキングスコアを基準に、推薦候補に対し、ランキング処理を施す。

4. 評価実験

本章では、評価実験により提案手法の有効性を確認する。提案手法を含め、比較対象に用いたランキング方式を以下に示す。

I コンテキスト依存型ランキング方式

II 多クラス分類に基づくランキング方式

III SVMの識別平面からのユークリッド距離に基づくランキング方式

表1 評価実験に用いたデータセット

データセット	データ数	1コンテキストあたりのデータ数
学習データセット	100	20
推薦データセット	150	30

4.1 実験方法

評価実験に用いたデータセットは、表1に示すとおりである。飲食店の特徴パラメータは28次元、コンテキストの特徴パラメータは24次元で、それぞれ表現される。各データに対し、被験者が5段階の満足度評価値(不満足1-2-3-4-5満足)を与えている。ここで、評価値が3以上のデータを正事例、3未満のデータを負事例として扱う。なお、本実験では5つのコンテキストパターンを想定して行った。被験者は、想定された各コンテキストパターンにおいて満足度評価を行った。1コンテキストパターンあたりの評価データ数は、表1に示すとおりである。なお、被験者は、研究室の学生および教官で、22歳から35歳の9人である。

提案方式によるランキング処理手順を以下に示す。

(1) 学習データセットをコンテキストパターンごとに分割する(ここでは、5分割)。

(2) 分割した各学習データセットから各特徴パラメータに対する重み分布を算出する(3.2節参照)。

(3) コンテキストパターンごとの推薦データセット中の各データのランキングスコアを算出する(3.3節参照)。

(4) 算出したランキングスコアを基準に、コンテキストパターンごとの推薦データセットに対し、ランキング処理を行う。すなわち、各コンテキストパターンにおいて、第1~30位までのランキングが得られることになる。

本実験では、再現率・適合率曲線および R 適合率を用いて評価を行う。いま、第 R 位までのランキング結果に着目したとき、本実験における再現率および適合率は、それぞれ次式で表される。

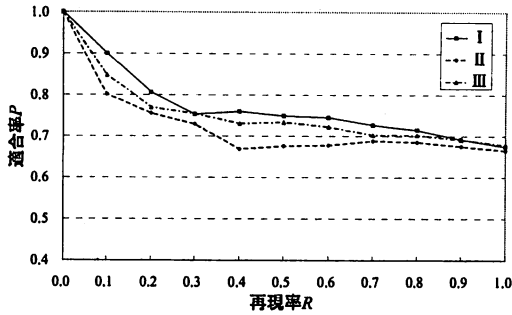
$$\text{再現率} = \frac{\text{ランキング結果に含まれる満足データの数}}{\text{推薦データセット中の満足データの数}} \quad (2)$$

$$\text{適合率} = \frac{\text{ランキング結果に含まれる満足データの数}}{\text{ランキング結果に含まれるデータの数}} \quad (3)$$

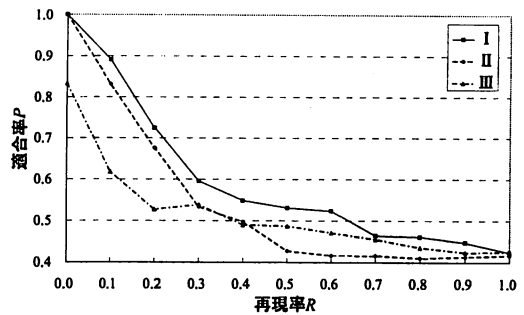
R 適合率は、上位から R 番目までのランキング結果を出力した時点での適合率のことであるから、この適合率がすなわち R 適合率ということになる。再現率・適合率曲線とは、上位から i 番目までのランキング結果を用いて算出した再現率と適合率をそれぞれ R_i , P_i で表したとき、両者の関係を示したグラフのことである。このグラフが上に位置するほど、性能が良いと評価されることになる。

4.2 実験結果および考察

4.1節で示した、各ランキング方式の、再現率・適合率曲線および R 適合率を、それぞれ、図6および図7に示す。ここでは、適合率を求める際に、推薦データセットのうち、評価値3以上のデータを正事例とした場合、評価値4以上のデータを正事例とした場合、それぞれにおける結果を示している。なお、

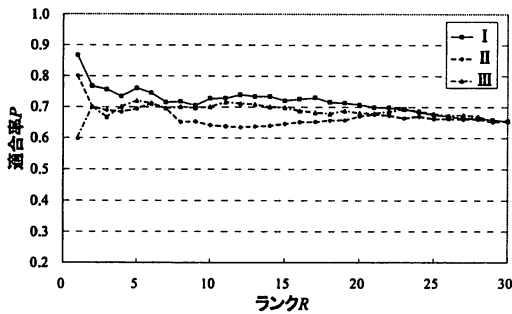


(a) 評価値3以上を正事例とした場合

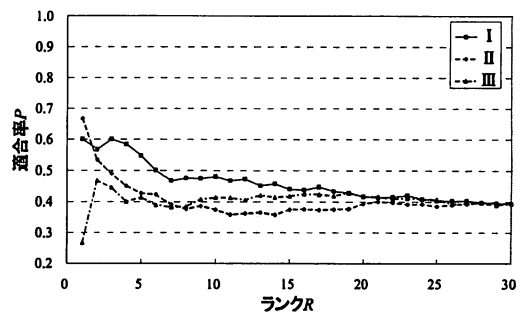


(b) 評価値4以上を正事例とした場合

図6 再現率・適合率曲線



(a) 評価値3以上を正事例とした場合



(b) 評価値4以上を正事例とした場合

図7 R 適合率

これらの結果は5つのコンテキストパターンに対して得られた結果の平均を示している。

いずれの結果からも分かるように、提案方式が比較的有効であることを示した。特に、 R 適合率に関して、図7 (a) より、たとえば、ランキング結果の上位10件の飲食店データが推薦されたとき、このユーザは7件の飲食店に対し満足する。さらに、図7 (b) より、このユーザは5件の飲食店に対し、より満足する。この結果は、実用的なものであるといえ、本実験結果より提案方式が有効であることを示した。

5. まとめ

本研究では、「ユーザが何を重視してコンテンツを選択しているかは、そのときのユーザのコンテキストに依存する」という考えに基づき、コンテキスト依存型ランキング方式を提案した。本方式では、ユーザがコンテンツの各特徴をどの程度重視しているかを、コンテキストごとに学習する。この学習成果に基づいて、推薦候補の各コンテンツのランキングスコアを算出する。そして、このランキングスコアを基準に、推薦候補に対しランキング処理を行う。さらに、単純なランキング方式との比較による評価実験を行い、提案方式の有効性を確認した。

ただし、今回は、コンテキストパターンを手動で分割し、分割後の学習データに基づいて、ランキングスコアを算出した。今後は、このコンテキストを適切にクラスタリングする方式を検討する必要がある。また、今回提案した方式により算出され

た重み分布が、適切に人の思考過程を反映したものであるか、被験者実験を通じて検証する。

謝辞 本研究の一部は、文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しいIT基盤技術の研究」、公募研究「偏在する大規模構造化文書からの高度情報抽出」(研究代表者:宮崎 純, A01-34, 課題番号 19024058) による。ここに記して謝意を表します。

文 献

- [1] Kenta Oku, Shinsuke Nakajima, Jun Miyazaki and Shunsuke Uemura: "Investigation for Designing of Context-Aware Recommendation System Using SVM", The International MultiConference of Engineers and Computer Scientists 2007 (IMECS 2007), March 2007.
- [2] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮: "状況依存型ユーザ嗜好モデリングに基づく Context-Aware 情報推薦システム", 情報処理学会論文誌 データベース, (TOD34), 情報処理学会, 2007年06月.
- [3] Cortes, C. and Vapnik, V.: "Support-Vector Networks, Machine Learning", Vol.20, No.3, pp.273.297 (1995).
- [4] Chen, G. and Kotz, D.: "A Survey of Context-Aware Mobile Computing Research", Paper TR2000-381, Department of Computer Science, Dartmouth College (2000).
- [5] Asthana, A., Cravatts, M. and Krzyzanowski, P.: "An Indoor Wireless System for Personalized Shopping Assistance", Proceedings of IEEE Workshop on Mobile Computing Systems and Applications, Santa Cruz, California, IEEE Computer Society Press, pp.69.74 (1994).
- [6] Abowd, G.D., Atkeson, C.G., Hong, J., Long, S., Kooper, R. and Pinkerton, M.: "Cyberguide: A Mobile Context-Aware Tour Guide, Wireless Networks", Vol.3, No.5, pp.421.433 (1997).