

# ユーザの検索行動に基づく嗜好推定を用いた 複数人での Web 検索における意見集約支援システムの提案

本田 博之<sup>†</sup> 岩田 麻佑<sup>††</sup> 原 隆浩<sup>†</sup> 西尾章治郎<sup>†</sup>

<sup>†</sup> 大阪大学大学院情報科学研究科 〒 565-0871 吹田市山田丘 1-5

<sup>††</sup> KDDI 株式会社 〒 163-8003 東京都新宿区西新宿 2-3-2

E-mail: <sup>†</sup>{honda.hiroyuki,hara,nishio}@ist.osaka-u.ac.jp, <sup>††</sup>ma-iwata@kddi.com

**あらまし** 実環境では、複数のユーザが Web 検索を行い、グループで食事するレストランを決定する場合など、各ユーザの嗜好に基づいて意見を集約するといったことがよく起こる。この際、複数のユーザがまず個別に Web 検索を行い候補を選んだ後、グループで議論することが一般的である。しかし、複数の候補の中から全員が納得する意見集約を行うことは容易ではない。そこで本研究では、ユーザの検索行動に基づく嗜好推定及び、意見集約の手法を提案する。提案手法ではまず、閲覧時間などの検索行動履歴の偏りに着目した嗜好推定を行う。その後、ユーザの嗜好を階層的に表現する手法を用い、グループにとって公平な意見集約を行う。提案手法に基づくシステムを設計・実装し、被験者実験によってシステムの有効性を検証する。

**キーワード** 複数人 Web 検索, 検索行動, 嗜好推定, 意見集約

## 1. はじめに

近年の Web 上における情報量の爆発的増加に伴い、ユーザは、膨大な情報の中から、適切なものを Web 検索によって選び出すことを求められている。適切な情報を得る手段の一つとして、複数人による Web 検索がある。例えば、グループで食事するために、複数人で協力して情報サイトを用いてレストランを探すといった状況が考えられる。複数人による Web 検索では、複数人が挙げた多数の選択候補の中から、グループに属するメンバ全員が満足できるような適切な候補を決定する必要がある。本稿では、選択候補として挙げられる個々のコンテンツをアイテムと定義する。情報サイトなどはアイテムの集合を対象とした検索サービスとみなせる。

先行研究において、口頭の議論において複数のメンバ全員が満足できるような意見の集約を目的とした手法 [1] や、参加者の発言を要約するシステム [2] が提案されている。しかし、これらはいずれも口頭で実際の議論を行うこと前提としており、議論のための時間拘束やユーザの負荷を考慮していない。個人での意思決定に比べ、グループにおける意思決定すなわち意見集約ははるかに複雑である [3]。メンバごとの嗜好の違いにより価値判断が異なる場合、メンバ間で相互理解を得たり互いに譲歩したりするなど複雑な課題が多く、意見を集約する上でユーザには精神的負荷がかかる。また実際にユーザが向き合っている口頭の議論では、自らが無意識に重視している要素を提示できない、全員の意見が公平に反映されないといった問題が考えられる。そのため、ユーザの負荷を軽減しつつ、メンバ全員の満足を得るような意見集約システムの構築は有効であると考えられる。

メンバ全員が満足できる意見に集約するためには、メンバ個人の嗜好を考慮する必要がある。嗜好とは、ユーザがどのよう

なアイテムを好むのか、あるいはアイテムを構成するどのような要素を重視するのかという特徴を指す。たとえば、ユーザはファストフード店が好きなのかそれとも焼肉店が好きなのか、あるいはファストフード店のメニューが好きなのか安めの価格帯が好きなのかといった点で嗜好が異なると考えられる。

これまでに筆者らの研究グループでは、複数人での Web 検索を行うユーザの行動の特徴を調査し、閲覧時間・ブックマーク対象の属性のばらつきなどのユーザの嗜好推定に有用ないくつかの要素を明らかにした [4]。本稿ではさらに、意見集約に関する先行研究を参考にして複数のユーザの意見集約のアルゴリズムを検討し、嗜好推定を基に複数人の意見集約を行うシステムを提案する。意見集約とは、最終的にグループに所属するメンバ全員が納得し、満足度が最大になる解を得ることである。

提案手法ではまず、これまでに明らかにしてきた個人の嗜好推定に有用な要素を用いて、ユーザ個人のアイテムへの評価の構造をモデル化する。本稿における提案手法では、ユーザの嗜好すなわちアイテムに対する評価は、複数の評価軸から構成され、さらにそれぞれの評価軸に対するこだわりの強さ、及び各評価軸においてどのような値を重視するかという要素に分解される。各評価軸の各要素について、実験を通して推定手法を検討する。続いて、意見集約に関する研究ならびに、本研究においてユーザに対し実施したアンケート結果に基づき、意見集約のアルゴリズムを検討する。嗜好推定を基に意見集約を行うシステムを実装し、被験者実験によってその有効性を検証する。

2 章ではまず、関連研究をいくつか紹介する。3 章で、これまでに筆者らの研究で明らかになった知見を述べ、意見集約システムの検討に向けて重要な要素を整理する。続いて 4 章で意見集約システムの設計について述べ、5 章でシステムを用いた被験者実験について述べる。最後に、6 章で本稿のまとめと今後の展望を述べる。

## 2. 関連研究

### 2.1 Web 検索における個人の嗜好推定

特定ページや特定サイトの閲覧時間やクリックスルー、検索クエリなどユーザの検索行動に基づいて嗜好を推定する研究が行われている。

Morita ら [5] は、ニュース記事において閲覧時間がユーザの興味のレベルと強い相関関係にあることを示している。ニュース記事のように、ほぼ同等の規模のアイテムが複数存在するような状況では同じ傾向であることが期待できる。例えばレストラン検索サイトではほぼ同等の情報量を持つ店舗情報が複数存在するので、閲覧時間と嗜好に相関があると期待される。

奥ら [6] は、ユーザごとに重要視している評価基準を考慮した情報推薦システムを提案している。このシステムでは、ユーザの置かれている状況も考慮し、ユーザの嗜好と状況両方に応じた情報推薦を実現している。ただしこのシステムは、モデル構築のため、ユーザが本来の Web 検索に関係のない評価作業を行う必要があり、ユーザの負荷が大きい。

Agichtein ら [7] は、クリックスルー、閲覧時間、検索クエリを用いた嗜好推定手法を提案している。ユーザごとに事前の機械学習を前提としているが、多様なユーザ行動を考慮することで高精度な嗜好推定が可能となっている。

本研究では、これらの研究を参考に、閲覧時間やブックマークといったユーザの検索行動に基づき、嗜好推定を行う手法を検討する。複数人による意見集約を円滑に行うため、どのアイテムを重視しているのかだけでなく、アイテムのどのような要素を重視しているのかという点にも着目した嗜好推定を行う。

### 2.2 グループの意思決定

小柴ら [8] は、グループの意思決定の際、「譲歩」の量がメンバー間で均衡すればグループ全体の満足度が向上するという「互惠性」を根拠にしたグループ意思決定支援機能を提案している。また福野らの研究 [9] によれば、相手よりも有利な解を得ることよりも、適度な譲歩によってグループの意思決定に貢献するような解のほうが、満足度が向上することが分かっている。これらの研究に基づくと、グループの意思決定においてグループ全体の譲歩量を調整することは重要な要素であることが分かる。

グループによる意思決定支援のために、Saaty [10] や山田ら [11] は AHP (Analytic Hierarchy Process) を用いる手法を提案している。AHP とは、Saaty [12] [13] [14] によって提案された、複数の要素から構成されるアイテムの集合から、ユーザが望む最適なアイテムを選び出す意思決定手法である。評価対象のアイテム、評価基準、総合評価の関係を階層的に表現し、同じ階層の要素同士で一対比較を行い、アイテムの重要度を決定する。本研究では、AHP の手法に着想を得て、評価対象、評価基準、さらに評価者 (ユーザ) の 3 階層からなるアルゴリズムを提案する。

## 3. 個人ユーザの嗜好推定

複数人 Web 検索において、メンバ全員の満足を得る意見集約を実現するためには、個人レベルでの嗜好推定とそれに基づ

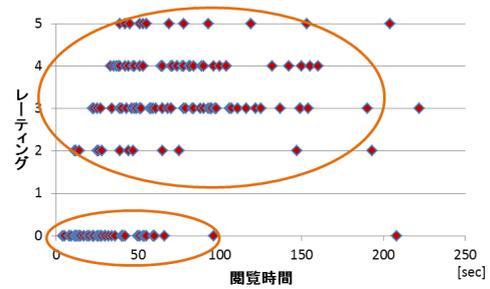


図 1 閲覧時間とレーティングの関係

いたグループとしての嗜好を決定する必要がある。筆者らの研究グループはこれまでに、ユーザ個人の嗜好推定に重要な要素を調査するために、被験者を用いた実験並びにアンケート調査を行った [4]。本稿においても、飲食店の検索を具体的な検索行動として取り上げ、ぐるなび API<sup>(注1)</sup>を用いてシステムを実装した。実験では、3 人一組のユーザのグループに、個人での検索をした後にグループでの議論を行ってもらった。また、個人での検索や嗜好に関するアンケートを実施した。

先行研究や、実験で実施したアンケートの回答を基に、ユーザの検索行動の中で嗜好推定に有用な要素を整理し、実際の実験データについて解析を行った。本章では、実験と解析によって得られた幾つかの知見を述べる。

### 3.1 複数の評価軸

アンケートで「どのような点を重視して (どのようなお店を) 検索しましたか?」という点について調査した結果、評価軸がユーザ間である程度共通していることが分かった。具体的には、「値段 (平均予算)」、「駅の近さ (徒歩時間)」、「雰囲気」、「店のジャンル」が挙げられる。この結果より、ユーザは飲食店を検索する際いくつかの評価軸を基準に店舗を選定しており、さらに、その評価軸は全てのユーザである程度共通なものであるといえる。したがって、嗜好推定を行うためにユーザ間で共通の評価軸を用い、さらに、その評価軸はシステム側で静的に設定しても十分であると考えられる。

多くの評価軸を用いることでより詳細なユーザの嗜好を推定できると考えられるが、本研究ではこれ以降、「平均予算」、「ジャンル」の 2 つを評価軸として設定し、検証を行う。他の評価軸については、今後の拡張における検討対象とする。

### 3.2 閲覧時間と嗜好の関係

続いて、先行研究 [5] [7] においても示されている、閲覧時間と嗜好の関係性について述べる。実験におけるユーザの閲覧時間履歴の長さ、ユーザがブックマークの際に付与した 5 段階評価のレーティングの関係に注目すると、閲覧時間が長いほど、ユーザの評価も高い傾向にあることが分かった (図 1)。閲覧時間とレーティングの相関係数は 0.363 と、中程度の相関を示す値であった。このことから、閲覧時間を嗜好推定の際の重みとして利用できると考えられる。提案手法における閲覧時間の使い方については、3.3 節で詳しく述べる。

(注 1) : ぐるなび Web サービス <http://api.gnavi.co.jp/api/service.htm>

### 3.3 閲覧履歴におけるアイテムの属性の“ばらつき”と嗜好の関係

次に、ユーザが各評価軸をどの程度重視して店舗の評価を行っているか（各評価軸へのこだわりの強さ）について考察を行う。ユーザが各評価軸をどの程度重視しているかという度合いを推定する方法について、ユーザが閲覧した店舗の持つ情報（APIで各店舗に付与された「平均予算」、「ジャンル」）、ユーザの検索・閲覧履歴、およびユーザのアンケート調査への回答を踏まえて考察する。具体的には、ユーザがある評価軸について一定の値へ強いこだわりをもっているほど、よく閲覧をしたり高いレーティングを与えたりする店舗の種類に偏りが生じ、結果的にユーザの閲覧した店舗の持つ情報の値のばらつきが小さくなるという仮説を立てる。そこで、ある評価軸に対するこだわりの強さと、ユーザが閲覧した店舗集合におけるその評価軸の値のばらつきの小ささには相関関係があるという仮説を立てる。例えば、あるユーザが安い店舗にこだわりがあるのなら、そのユーザは「平均予算」の低い店舗を優先して閲覧し、高いレーティングを与えると予想される。結果として、このユーザの全閲覧履歴の店舗の「平均予算」のばらつきは、他のユーザの閲覧した店舗の「平均予算」のばらつきよりも小さくなると考えられる。

以下で、3.1節で述べた2つの評価軸「平均予算」「ジャンル」について、3.2節および3.3節の議論に基づいたデータのばらつきについての検証結果を述べる。

#### 3.3.1 平均予算

「平均予算」にこだわりをもつユーザは、自らが望む一定程度の範囲の予算額にあてはまる店舗をよく閲覧・ブックマークしたり、高いレーティングを与えるものと考えられる。そこで、ユーザの閲覧した店舗における平均予算のばらつき（標準偏差）を計算し、その標準偏差値が「平均予算をどの程度重要視しますか」というアンケートへの回答とどのような関係にあるかを検証する。標準偏差を求めるための平均値は、ユーザが与えたレーティングの値、および3.2項で述べた閲覧時間を重みとして与えて算出する。レーティングの値はそのまま乗算して重みとし、閲覧時間  $T[sec]$  については以下の式で重み  $t$  を求めた。

$$t = T/10 \quad (1)$$

ユーザ  $u$  が閲覧した店舗  $R_{ui}(i = 0, 1, 2, \dots, n-1|n$  はユーザ  $u$  の総閲覧数) の「平均予算」の値を  $b_{ui}$ 、その店舗にユーザが付与したレーティングの値を  $r_{ui}(r = 1, 2, 3, 4, 5, \text{レーティングなしは} 1 \text{ とみなす})$ 、その店舗のユーザの閲覧時間を  $t_{ui}$  とした時、重みを考慮した平均  $A_{bu}$ 、分散  $V_{bu}$ 、標準偏差  $S_{bu}$  を以下のように算出する。

$$A_{bu} = \frac{\sum_{i=0}^{n-1} b_{ui} * r_{ui} * t_{ui}}{\sum_{i=0}^{n-1} r_{ui} * \frac{t_{ui}}{10}} \quad (2)$$

$$V_{bu} = \frac{\sum_{i=0}^{n-1} A_{bu} - b_{ui}}{n} \quad (3)$$

$$S_{bu} = \sqrt{V_{bu}} \quad (4)$$

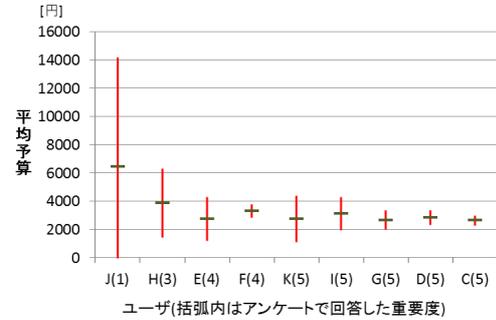


図2 平均予算に対する重要度と閲覧履歴の平均予算のばらつきの関係

図2に、各ユーザのアンケートの回答と、閲覧結果から導出した「平均予算」の標準偏差との関係を示す。横軸はユーザを表し、縦軸は「平均予算」[円]を表す。各ユーザのグラフにおいて、水平方向の短い線分は閲覧履歴における「平均予算」の平均値を表し、その線分より垂直方向上下に伸びる棒グラフが閲覧履歴における「平均予算」の標準偏差を表す。グラフのラベルについて、括弧内はそのユーザがアンケートで回答した「平均予算」の重要度を示す。また、ユーザはアンケート回答の重要度の値の大きさにソートしてある。ユーザ  $F$  を除き、アンケートで回答された重要度が高いほど、閲覧した店舗の「平均予算」の平均値が小さく、かつ標準偏差が小さい傾向にあることが分かる。アンケートで回答された重要度の、平均値との相関係数は  $-0.944$ 、標準偏差との相関係数は  $-0.915$  である。これより、「平均予算」を重視するユーザほど、ユーザが理想とする「平均予算」の値は小さい、すなわちより安い店舗を望んでいるといえる。なお、「平均予算」を重視するほどより低価格を希望するという傾向については、実験のシチュエーション設定に依存すると考えられる。例えば、「海外からの来客を饗すレストランを決定する」といった状況であれば、「平均予算」が高めの高級レストランを希望すると予想される。

また、「平均予算」を重視するユーザほど、高いレーティングを与えた店舗の「平均予算」が一定値に偏っているといえる。逆に、「平均予算」に強いこだわりがないユーザは、他の評価軸に注目して閲覧を行なっているため、高いレーティングを与えられた店舗の「平均予算」が一定値付近に集中することがなく、ばらつきが大きくなる傾向がある。

なお、ユーザ  $F$  について、アンケートで回答された重要度が4であるにもかかわらず、重要度が5と回答したユーザ  $C$  などと同程度に標準偏差、すなわちばらつきが小さい。ユーザ  $F$  は単独検索フェーズ後の自由記述のアンケートの設問「満足行くまでお店の検索ができましたか?」に対して5段階評価で4と高い値を回答しており、その理由として「(グループのメンバー)ある程度価格帯がはっきりしているメンバーだから」と答えている。また、別の設問「自分の意見はまとまりましたか?」に対しては5段階評価で5と回答しており、これには理由として「価格も内容もちよほど良いお店が多くあったから」と答えている。これらの結果より、ユーザ  $F$  は主にグループのメンバーを考慮し、閲覧したり高いレーティングを与える価格帯を固定

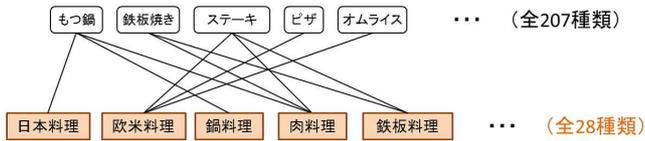


図3 属性表の概念図

していたことが分かる。このことから、「平均予算」の重要度は4ではあるが、その価格帯が検索開始時点から一貫して一定値に定められていたために、そのばらつきが小さくなったと考えられる。

以上の結果より、「平均予算」という評価軸については、ユーザの閲覧した店舗の「平均予算」のばらつきを求めることで、ユーザの評価軸「平均予算」へのこだわりを数値化できる可能性があることを確認した。

### 3.3.2 ジャンル

続いて、「ジャンル」のばらつきについての検証結果を述べる。「ジャンル」は数値データではないので、「平均予算」のように、値の平均や分散、標準偏差を求めることができない。そこで、各店舗のジャンルをベクトルで表現し、その分散値を求めることにした。具体的には、まず、ぐるなびAPIで定義されている207種類の「ジャンル（以降、子ジャンルと呼ぶ）」を、手で28の親ジャンルに対応付けた。この対応表を、ジャンルの属性表と呼び、図3に対応関係の概念図を示す。各子ジャンルは複数の親ジャンルに対応し、その対応数は子ジャンルにより異なる。

属性表作成にあたり、WikipediaのCategoryの記事<sup>(注2)</sup>、及びYahoo!JAPANカテゴリ<sup>(注3)</sup>の分類を参考にした。例えば、「鉄板焼き」、「焼き鳥」、「串焼き」、「炭火焼・ろばた焼き」などの子ジャンルは、共通して「鉄板料理」という親ジャンルに対応する。そして、レストラン*r*のジャンルを親ジャンルの28次元のベクトル $\hat{\mathbf{G}}_r$ で表現する。例えば、子ジャンルを基に対応付けをした結果「日本料理」「鉄板料理」という親ジャンルが該当する店舗であれば $\hat{\mathbf{G}}_r = (1, 0, 0, \dots, 1, \dots, 0, 0, 0)$ のように表現する。ここで、APIにおける登録数の多い子ジャンルが対応する親ジャンルのスコアが高くなる傾向があり、また対応する子ジャンルの数が多い親ジャンルほどスコアが高くなる傾向があることが分かった。そこで、全店舗中での子ジャンルの出現頻度、および属性表における対応数に応じて重みを与えることとした。出現頻度に関する重みは、ぐるなびAPIに登録されている全店舗の子ジャンルを種類別に合計し、子ジャンルの出現頻度に基づいたidf値とした。対応数に関する重みは、属性表において対応する子ジャンルの種類*n*に対する $1/n$ とした。

#### ● 分散による導出

これまでの調査で、上記で定義したベクトルを用いて単純に求めた分散値では、うまくばらつきを表現できないことが分かった。ばらつきを考え方が適用できなかった原因として、ベクトルの次元が高く、親ジャンルと子ジャンルの対応数が複数

であるため、特徴量として非常に複雑であることが挙げられる。“ある特定の要素（親ジャンル）にスコアが偏ってはいるものの、他の要素のスコア構成が異なる”といった場合、分散値が大きくなることが考えられる。

#### ● ジニ係数による導出

ここで、ユーザの嗜好を店舗の「ジャンル」ベクトルと同様に表現して「ジャンル」についてのユーザのパラメタと定義し、パラメタの各要素によるヒストグラムがどのようにばらついているかを考える。ユーザのパラメタは、店舗のジャンルのベクトル要素に、5段階評価のレーティングと閲覧時間*t*を重みとして加えたものを、ユーザが閲覧した全ての店舗について加算して求めた。ユーザ*u*の、「ジャンル」に対する嗜好のパラメタベクトルを $\hat{\mathbf{G}}_u$ とする。すなわち、ユーザのジャンルについてのパラメタは、店舗のジャンルについてのベクトルと同じ28次元の要素を持つベクトルである。そして特定の要素にスコアが偏っているほど、評価軸「ジャンル」についての嗜好は強いと定義できる。

$$\hat{G}_u = \sum_{i=1}^{n-1} (\hat{G}_r * r_{ui} * t_{ui}) \quad (5)$$

$$= (g_{u1}, g_{u2}, \dots, g_{u28}) \quad (6)$$

スコアの偏りは、ジニ係数を用いて導出する。ジニ係数とは、特定の地域の全住民に対する資産の分配がどの程度偏っているかを表す指標[15]で、横軸に全人口の累積割合、縦軸に資産の累積割合を取った2次元平面上に表現される。すなわち、ヒストグラムの偏りを表す指標とみなせる。 $y = x$ の直線を均衡直線と呼び、データが均衡直線上にある時、ヒストグラムは偏りなく完全に均等であることを示す。実際のデータに基づいてプロットされた点を繋いだ線をローレンツ曲線と呼び、ローレンツ曲線と均衡直線で囲まれる面積の2倍がジニ係数である。すなわち、データが特定のパラメタに偏っているほどジニ係数は増加する。ジニ係数は[0, 1]区間の値を取る。

本研究においては、ジニ係数の考え方を応用し、横軸に累積ジャンル数割合、縦軸にスコアの累積割合を用いることで、どの程度特定のジャンルにスコアが偏っているかを求める。横軸が離散的になるため、ジニ係数に該当する面積は近似によって求める。ジャンル数*m*(本稿では28)に対して、以下の式で計算する。

$$Gini_u = \sum_{j=1}^m \left( \frac{j}{m} - g_{uj} * r_{uj} * t_{uj} \right) \quad (7)$$

図4に、各ユーザの閲覧結果から導出した「ジャンル」のジニ係数、アンケートの回答の関係を示す横軸はユーザを表し、縦軸はジニ係数を表す。横軸は、ユーザがアンケートで回答した「ジャンル」の重要度で並べ替えてある。アンケートで「ジャンル」を重視すると回答したユーザほど、ジニ係数が高い傾向にあるといえる。

ここで、ユーザのアンケートの回答は5段階のうち[2,5]の間に分布しているが、[0,1]の区間幅を取りうるジニ係数が、お

(注2) : Wikipedia <http://ja.wikipedia.org/wiki/Category:飲食店>

(注3) : Yahoo!JAPAN カテゴリ <http://dir.yahoo.co.jp/>

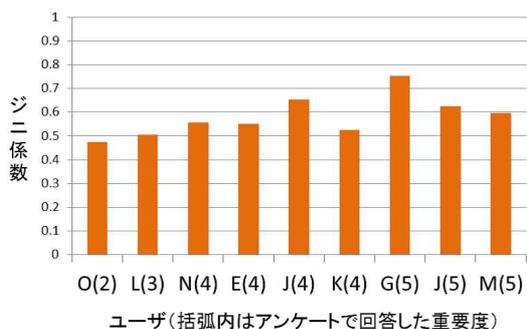


図4 ジェンルに対する重要度と閲覧履歴のジャンルに関するジニ係数の関係

よそ 0.45~ 0.75 の間のごく限られた範囲に分布していることが分かる。ジニ係数は、完全平等の時に 0 を、完全不平等の時に 1 をとる。ユーザが自らの嗜好に従って店舗を閲覧している以上、完全平等、すなわち全ての「ジャンル」に該当する店舗を均等に閲覧するとは考えにくく、逆に、唯一の「ジャンル」に該当する店舗のみを閲覧することも、実環境での Web 検索を考慮すると、自然とはいえない。よって、ジニ係数の値がある程度の範囲内であることは妥当といえる。

以上の検証から、「ジャンル」に対する嗜好のパラメタベクトル  $\hat{G}_u$  の要素を用いたジニ係数の計算によって、ユーザの「ジャンル」に対するこだわりを推定できる可能性を示した。

### 3.4 各評価軸でユーザが真に望む値の推定

評価軸の重要度を推定するだけでなく、各評価軸において、ユーザが求める具体的な値（以降、「真値」と定義）についても推定する必要がある。例えば、共に「平均予算」を重視し、安めの店舗を望む 2 ユーザがいる場合、一方のユーザは 2000 円程度を望んでいるのに対して、もう一方は 1000 円程度を望んでいるという状況が考えられる。ここでは、「平均予算」および「ジャンル」の真値について検討する。

#### 3.4.1 平均予算

「平均予算」については、ばらついた値の重心はその集合の代表値の一つであることから、3.3 項で求めた、重み付き平均を真値として算出することを考える。図 5 に、各ユーザの「平均予算をどの程度重要視しましたか」というアンケートへの回答、ユーザが理想とする「平均予算」の区間値、推定した真値の関係を示す。棒グラフが、ユーザがアンケートで回答した、希望する平均予算の区間値を表し、グラフ中で一定値を示す細い線分が、重み付き平均値を表す。なお、ユーザ F については、アンケートにおいて平均予算の希望区間値を回答しなかったため、回答をグラフに表示していない。真値がユーザの理想とする「平均予算」の区間の中央付近の値であることが分かる。このことから、「平均予算」に関しては、重み付き平均を真値とし、一定幅の前後値も含めた区間値をユーザの望む値の推定値とみなすことができると考えられる。また、アンケートへの回答の重要度が大きいユーザほど、理想とする値の区間幅が小さい。すなわち、ユーザの望む値を区間値として推定する際、その区間幅は重要度を用いて決定できると考えられる。一方、ユーザ

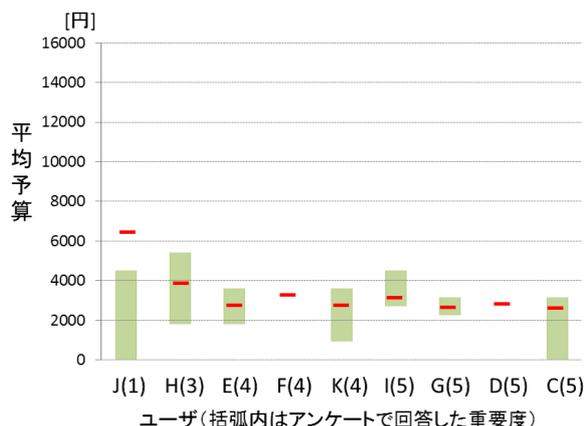


図5 平均予算の重要度と希望区間値および推定真値の関係

J について、重み付き平均値がユーザの希望値の範囲に含まれていないことがわかる。これは、ユーザ J が「平均予算」をほとんど重視しておらず（アンケートでの重要度の回答が 1）、低予算な店舗から高予算な店舗までを閲覧した結果、閲覧履歴が大きくばらつき、重み付き平均値が精度を欠いたためであると考えられる。このようなユーザの場合、評価軸「平均予算」の重要度が低いため、真値の推定がやや精度を欠いたものでも嗜好推定に大きな影響はないと考えられる。

#### 3.4.2 ジェンル

「ジャンル」の真値は、3.3.2 項でユーザのパラメタとして定義したベクトルとした。どのジャンルにどれほど嗜好が強いかをベクトル要素のスコアに表れており、ユーザが真に望む値としてふさわしいと考えられる。

この定義の妥当性および精度について、ベクトルの要素ひとつひとつに対して「〇〇をどの程度重要視しましたか」というアンケートを実施して検証するのは適当でないと考えられる。システムで定義した 28 のジャンルを意識し、かつそれらにスコアを割り振る、という思考プロセスがユーザの意識にあるとは考えにくく、アンケートが正解データとして妥当とは考えにくいためである。そのため、「ジャンル」の真値の定義の妥当性・精度については、ユーザのパラメタベクトルを真値とする定義を用いてシステムを実装し、実際の店舗を提示しそれに対する満足度を調査する、という実用に近い方法での検証を行う。

## 4. グループの意見集約システムの検討

ここまでで、ユーザ個人の嗜好推定に有用な要素、ならびにその具体的な推定手法について述べた。本章では、個人の嗜好推定によって得られた情報を複数人で集約する方法について、集団意思決定の先行研究に基づき検討する。

本研究においては飲食店の検索を対象にしており、意見集約の結果として、対象のグループにとって最も満足度の高い飲食店を出力するシステムの設計を目指す。グループにとっての満足度の高い解を出力できるシステムというのは、各ユーザの嗜好を的確に判断し、かつ各ユーザの主張を実環境と同等に反映するシステムである。提案システムの処理は大きく分けて、ユーザの嗜好推定と、各ユーザの主張の反映という 2 つの

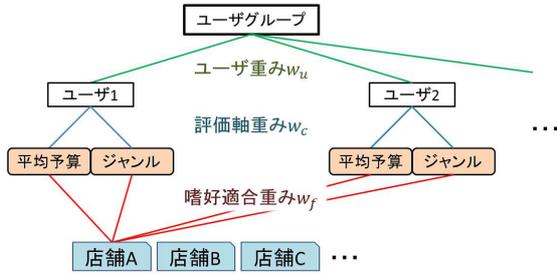


図6 複数人での意見集約を目的としたユーザの嗜好モデル

フェーズからなる。

#### 4.1 嗜好モデルの構成

複数の要素から構成されるアイテムの集合から、ユーザが望む最適なアイテムを選び出す意思決定手法として、Saaty [12] [13] [14] によって提案された、階層分析法 (AHP; Analytic Hierarchy Process) がある。評価対象のアイテム、評価基準、総合評価の関係を階層的に表現し、同じ階層の要素同士で対比較を行い、アイテムの重要度を決定する。本研究では、ある対象への評価を階層的に表すというアイデアに基づき、以下のような、複数人で意見を集約することを目的とした嗜好モデルを提案する (図6)。ユーザの嗜好を階層的に表現し、より単純な要素に因数分解することで、「ユーザは〇〇を好む」という曖昧な事象を数値に置き換えて扱うことができる。また、各要素毎に導出手法の検討が出来るため、システムとしてのチューニングも容易になると考えられる。

提案する嗜好モデルの要素構造は、ユーザ、評価軸、店舗の3階層で構成されており、 $W_U$  (ユーザ重み)、 $W_C$  (評価軸重み)、 $W_F$  (嗜好適合度重み) の3つの重みを定義する。3つの重みは全て区間  $[0,1]$  で正規化し、階層間で重みに差がないようにする。

#### 4.2 各階層における重みの定義

前節で定義したユーザの嗜好モデルにおける、各重みの定義について述べる。3章で述べたユーザ個人の嗜好推定の結果を用いて各重みを定義する。重みを導出する際の入力データは、ユーザの閲覧履歴における「閲覧時間」、「飲食店の平均予算」、「飲食店のジャンル」、「ブックマークの有無」、「5段階レーティングの値」である。

##### 4.2.1 ユーザ重み

グループのメンバ間において、どのユーザの意見がどの程度重要であるかを表すパラメータである。ユーザ間の意見の重要度の差は、実環境の人間関係やユーザの性格など、提案モデルに含まれる要素の中で最も人文学的性格が強く、複雑であると考えられる。そのため、まずはユーザの明示的な嗜好に沿う単純な方法でユーザ重みを決定し、ユーザ重みそのものの有効性を検証することが望ましい。本稿においては、実験において自らの意見をどの程度主張するつもりか/したかについて問い、それに対する回答をシステムへの入力とした。具体的には、議論の前においては「自分の意見は議論でどの程度主張するつもりですか?[5段階]」と問い、これに対する回答を、議論後においては「自分の意見はどの程度主張できましたか?[5段階]」と問

い、これに対する回答を、それぞれ各ユーザのユーザ重み  $W_U$  とした。

##### 4.2.2 評価軸重み

ユーザがどのような評価軸をどの程度重視しているかを示すパラメータである。提案モデルでは、3.3節において議論した閲覧履歴の「ばらつき」に基づき定義されたユーザの「こだわり」を表すパラメータを基に評価軸重みを定義する。本稿においては、「平均予算」と「ジャンル」の2つの評価軸を設定し、スコアを与える。ここで、ユーザが自らの嗜好に基づいて飲食店を「平均予算」、「ジャンル」という2軸で評価する際、どちらの評価軸をどの程度重視したかの度合いは、2つの評価軸間で相対的に決定されると考えられる。「ユーザの嗜好」を全体とした時、「平均予算」、「ジャンル」はそれぞれ嗜好のどの程度の割合を占めるか、という考え方をすることができる。そこで、「平均予算」「ジャンル」の2軸のスコアの合計が1.0となるよう、スコアを正規化する。

具体的な値の導出方法は以下のとおりである。

- 平均予算

$$W_{Cb} = \log_{10}\left(\frac{1}{S_u}\right) \quad (S_u \text{は標準偏差 式(4)}) \quad (8)$$

- ジャンル

$$W_{Cg} = Gini \quad \text{式(5)} \quad (9)$$

##### 4.2.3 嗜好適合度重み

ユーザの嗜好に基づいて飲食店を評価した時、各店舗がどの程度ユーザの嗜好に適合しているかを表すパラメータである。「各ユーザが、実際に店舗を評価した時に与えるであろうスコア」の推定値であるといえる。5節で求めたユーザが真に望む値 (= 真値) と API によって付与された飲食店のパラメータを比較し、二者間で「平均予算」、「ジャンル」の各要素の差が小さいほど、当該飲食店はユーザの嗜好に近いと判断できるため、嗜好適合度重みに高いスコアを与える。

具体的な値の導出方法は以下のとおりである。

- 平均予算

ユーザのパラメータ (= 真値) を  $b_u$  [円]、店舗のパラメータを  $b_r$  [円] とし、次式で算出する。

$$W_{Fb} = 1 - \sqrt{(b_u - b_r)^2} \quad (10)$$

- ジャンル

ジャンルについては、ユーザと店舗それぞれのパラメータベクトル  $\hat{G}_u$ 、 $\hat{G}_r$  のコサイン類似度で表し、次式で算出する。

$$W_{Fg} = \cos(\hat{G}_u, \hat{G}_r) \quad (11)$$

#### 4.3 集約した意見の提示

システムの出力は、ユーザのグループにとって最も満足度が高いと考えられる飲食店の提示である。まず、4.1節で定義された各パラメータを基に、各飲食店に対する総合評価のスコアを算出する。飲食店に対するスコアの算出方法は、 $u$  をユーザ数とすると以下のとおりである。

$$Score = \sum_{v=0}^{u-1} (W_{C_{bv}} * W_{F_{bv}} + W_{C_{gv}} * W_{F_{gv}}) \quad (12)$$

本稿においては、スコアを与える対象の母集団は「ユーザグループのいずれかのメンバが閲覧もしくはブックマークした店舗」の集合とした。これは、実環境においても、ユーザ同士が議論の対象とするのはいずれかのユーザが閲覧した、あるいは候補として挙げたアイテムであることが一般的であるからである。

以上の方法で各飲食店に与えられたスコアを基準に、飲食店を降順に並べ替え、上位のものをユーザのグループに提示する。

## 5. 評価実験

4. 章で提案したシステムの有効性を評価するため、ユーザを用いた被験者実験を行った。

### 5.1 実験設定

#### 5.1.1 条件設定

実験は被験者 3 名を 1 グループとし、計 9 名からメンバを入れ替えて 4 グループを作成して実験を行った。被験者の内訳は大阪府在住の 20 代男性 9 名である。

タスクは「グループの 3 名で食事を行うための店探し」と指定し、ユーザが既知の店舗を候補に入れることを極力避け、なおかつ可能な限り多くの候補からの選択を行ってもらうため、店舗の立地を東京都に指定した。具体的には、まず店舗情報を個人で検索し、気に入った店舗があれば、レーティングの付与やブックマークをしてもらい（単独検索フェーズ）、その後、ブックマークを共有して口頭で議論を行ってもらいグループとしての解を決定してもらった。

#### 5.1.2 実験システム

実験システムのインターフェースは、筆者らの研究グループがこれまで行ってきた研究と同様のものを用いる [4]。ユーザは本システムを用いて、飲食店を検索する。

#### 5.1.3 実験手順

##### (1) 単独検索 (30 分)

各ユーザが、検索システムを用いて自由に閲覧・ブックマークを行う。

##### (2) 議論 (40 分)

議論用のブックマーク共有システム [4] を用い、ユーザ同士で実際に向き合って議論してもらった。ユーザのグループにとって満足度の高い上位 10 件の候補を決めてもらった。

##### (3) システムによる候補の提示・システムの評価 (後述)

### 5.2 比較手法

提案システムにおける意見集約手法（提案手法）を以下の手法と比較する。

#### 5.2.1 口頭議論

実験の議論フェーズにおいて、ユーザ間の口頭の議論で決定されたランキング。ユーザ同士が実際に議論を行って決定したランキングなので、議論を経ることを前提とした上での正解データの一つと言える。

#### 5.2.2 ランダム手法

グループのいずれかのユーザが閲覧もしくはブックマークした店舗の集合から、ランダムに店舗を抽出し、ランキングを生成する。

#### 5.2.3 明示的手法

ユーザに、各店舗に対する明示的な評価をしてもらい、それを基にランキングを生成する。ユーザに評価の手間と負担を与え、グループのメンバ全員にとって満足度の高い候補が選ばれるため、本研究における理論上の正解データの一つと言える。

具体的には、グループのいずれかのユーザが閲覧もしくはブックマークした店舗の集合から、店舗を選び出して 1 つずつユーザに提示し、ユーザにはその店舗に対する満足度を 5 段階で評価してもらう。この 5 段階評価のユーザ間での単純合計が高い上位 10 件でランキングを生成する。この際、店舗の提示順はランダムとし、ユーザ間でも異なる順で提示した。

この作業は、ユーザ同士の議論の前と後でそれぞれ行い、ランキングもそれぞれ別に生成する。これら 2 つのランキングを、比較手法の「明示的手法（議論前）」、「明示的手法（議論後）」とする。明示的手法（議論前）は、各ユーザの議論前の純粋な嗜好を反映し、明示的手法（議論後）は議論の結果に影響を受けた、納得や妥協を含む満足度を反映している。

#### 5.2.4 提案手法 1(ユーザ重みなし)

提案手法において、ユーザ重みを 1:1:1 としてランキングを生成する。4.2.1 項でも述べたとおり、ユーザ重みは提案モデルに含まれる要素の中で最も人文学的性格が強く、複雑であると考えられるため、ユーザ重みの影響を排除した手法も評価対象に加え、ユーザ重みの有効性を検証する。

#### 5.2.5 提案手法 2(ユーザ重み議論前)

提案手法において、ユーザ重みを議論の前の段階で 4.2.1 項で述べた手法を用いて定義して入力し、ランキングを生成する。議論前のユーザ重みを入力とするため、特に各ユーザの「本音」での意見や主張の強さが反映されることが予想される。

#### 5.2.6 提案手法 3(ユーザ重み議論後)

提案手法において、ユーザ重みを議論を終えた後の段階で 4.2.1 項で述べた手法で定義して入力し、ランキングを生成する。議論で実際に意見が反映された度合いを入力とするため、議論がユーザ間の意見の強さの相対関係に与える影響を検証する。

### 5.3 実験結果と考察

提案手法と各比較手法によるランキングに対する、各ユーザの満足度を図 7 に示す。グラフは、ユーザ A, C, D の 3 名のグループに対する実験の結果で、各手法によって提示されたランキングに対する 3 名の満足度の総合評価を示す。また、各手法の一番右側のグラフは 3 名のユーザの満足度の平均値を示す。全てのグループにおいて、提案手法、明示的手法（議論前）、明示的手法（議論後）がいずれもランダム手法より高い満足度を示していることが分かる。また提案手法は、明示的手法（議論後）には劣るものの、明示的手法（議論前）と同程度程度の満足度が得られている事が分かる。このことから、提案手法は理想的な解は得られていないものの、嗜好推定ならびに意見集約の

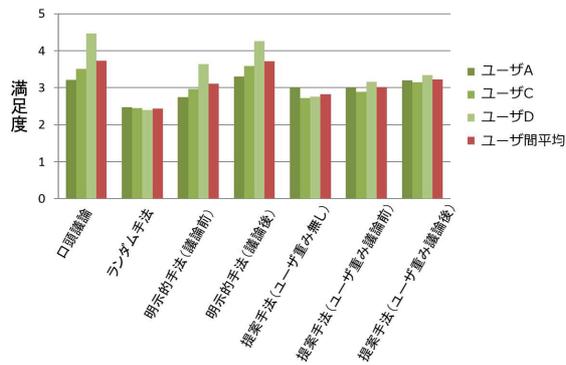


図7 各手法の満足度評価の結果

手法として一定の性能を示しているといえる。

いずれのグループにおいても、ユーザ重みを考慮した手法は、ユーザ重みを考慮しない手法よりも高い満足度を示していることが分かる。この理由として、ユーザ同士の意見集約の際、ユーザ間の性格の違いや人間関係を考慮した方が、ユーザにとって最適の解が導けると考えられることが挙げられる。このことから、単に平等に意見を反映させるだけでなく、ユーザ同士の意見の強さを考慮に入れることは、ユーザグループにとって適切な候補を選び出す上で有効であるといえる。

また、明示的手法（議論前）と明示的手法（議論後）の満足度を比較すると、明示的手法（議論後）の満足度の方が高いことが分かる。議論を経ることで満足度が上がっていることから、議論による納得や妥協を経て、グループ内の各ユーザの嗜好が変化し、同じような嗜好に偏向した可能性があるといえる。したがって、ユーザの嗜好は短時間で変化する可能性があり、またそれは他者からの「説得」「提案」によって生じると考えられる。このことから、今後システムの性能向上を検討する上で、ユーザの嗜好の変化を考慮に入れなければならないと考えられる。また一方で、「システムからの提案」によって、ユーザの嗜好が変化し、より高い満足度を得られる可能性があることを示唆しており、本システムがユーザ間の議論の代替となりうるということを示している。今後、議論前後で明示的手法におけるユーザの満足度が向上した要因についてさらに調査し、提案システムを拡張する必要がある。

## 6. おわりに

本稿では複数人による Web 検索における意見集約の支援システムの構築を目的として、ユーザの嗜好と検索行動との関係性を被験者実験を通して検証し、また集団意思決定の手法を基に、意見集約のための嗜好モデルを構築した。さらに、構築した嗜好モデルに基づいて、意見集約システムを提案した。

提案手法に基づいて構築したシステムを用いて被験者実験を行い、システムの有効性を検証した。システム構築のためのユーザの検索行動の解析ならびに、提案システム及び他の手法を用いた実験から以下のような結果・考察が得られた。

- ユーザの嗜好パラメタとアイテムの属性を比較することで、ユーザの嗜好に近いアイテムであるかどうか判断できる

と考えられ、ユーザの嗜好パラメタに近い属性をもつアイテムほど、ユーザにとって理想のアイテムであるといえる。

- 提案手法を用いたシステムは、ユーザから一定の満足度を得ることができた。また提案手法は、ユーザに手間をかけさせる明示的な手法と同程度の満足度を得られており、提案手法がユーザの負担を軽減する有効な手法であることが示唆された。

- ユーザの嗜好は、他者との議論における提案や説得によって短時間で変化する可能性がある。

- ユーザ間で意見の重要度に差があり、それを適切に定義することで、複数ユーザのグループにとって満足度の高いアイテムを推定する際の精度を向上させると予想される。

今後は、被験者実験を通してシステムの拡張・改良を行い、システムの性能を向上させる予定である。特に、人文的側面からも調査を行い、ユーザ間での意見の重要度を決定する適切な手法を検討する。

## 文 献

- 高野伸栄, 鈴木聡士, 代替案修正ベクトル法による合意形成システムに関する研究, 土木学会論文集, Vol.716, pp.11-20, 2002.
- 古田一雄, 前原基芳, 高嶋亮祐, 中田圭一, 知的支援機能を備えた電子会議システム, 社会技術研究論文集, Vol.1, No.0, pp.299-306, 2003.
- 加藤直孝, 合意形成プロセスにおける参加者の視点情報の共有に基づくグループ意思決定支援システムの研究, 博士論文, 1998.
- 本田博之, 岩田麻佑, 原隆浩, 西尾章治郎, ユーザの検索行動に基づく嗜好推定を用いた複数人 Web 検索における意見集約支援システムの検討, データ工学と情報マネジメントに関するフォーラム, B4-4, 2013
- M.Morita, Y.Shinoda, Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval, Proceedings of SIGIR '94, pp. 272-281, 1994.
- 奥健太, 中島伸介, 宮崎純, 植村俊亮, 加藤博一, 情報推薦におけるユーザの価値判断基準モデルに基づくコンテキスト依存型ランキング方式, 情報処理学会論文誌データベース, Vol.2, No.1, pp.57-80, 2009.
- E.Agichtein, E.Brill, S.Dumais, R.Ragno, Learning User Interaction Models for Predicting Web Search Result Preferences, Proceedings of SIGIR '06, pp.3-10, 2006.
- 小柴等, 加藤直孝, 國藤進, 互恵性を用いたグループ意思決定支援機能の提案, 情報処理学会論文誌, Vol.50, No.1, 268-277, 2009.
- 福野光輝, 大淵憲一, 最終提案交渉における受け手の拒否動機分析: 同一性保護の観点から, 社会心理学研究, Vol.16, No.3, pp.184-192, 2001.
- Saaty T.L., Group Decision Making And The AHP, The Analytic Hierarchy Process, Springer-Verlag, pp.59-67, 1989.
- 山田善靖, 杉山学, 八巻直一, 合意形成モデルを用いたグループ AHP, Journal of the Operations Research Society of Japan, Vol.40, No.2, pp.236-244, 1997.
- Saaty T.L., A Scaling Method For Priorities In Hierarchical Structures, Journal of Mathematical Psychology, Vol.15, No.3, pp.237-281, 1977.
- Saaty T.L., The Analytic Hierarchy Process, McGraw-Hill, 1980.
- Saaty T.L., Decision Making With The Analytic Hierarchy Process, International Journal of Services Sciences, Vol.1, No.1, 2008.
- 中村良平, 都市・地域における経済集積の測度, 岡山大学経済学会雑誌, Vol.39, No.4, pp.99-121, 2008