グループ適応型システムのための満足度を考慮した推薦方式の提案

山根　康男[†]　官上　大輔[††]　河合　由起子[††]　津田　宏[†††]　田中　克己[††, ††††]

[†] ㈱富士通研究所　〒211-8588 神奈川県川崎市中原区上小田中 4-1-1
[††] 独立行政法人情報通信機構　〒619-0289 京都府相楽精華町光台 3-5
[†††] 富士通株式会社　〒105-7123 東京都港区東新橋 1-5-2
[††††] 京都大学大学院情報学研究科社会情報学専攻 〒606-8501 京都市左京区吉田本町

E-mail: [†] yamane.yasuo@jp.fujitsu.com, [††]{yukiko, kanjo}@nict.go.jp, [†††] htsuda@jp.fujitsu.com, [††††] tanaka@dl.kuis.kyoto-u.ac.jp

あらまし 検索や情報推薦の分野で､個人の嗜好や興味に合った情報提供のための個人適応化の研究が行われている。しかし、グループ全員の嗜好を考慮したグループ適応の研究はそれに比べると少ないように思われる。本論文では、グループ全体の満足度を考慮した推薦方式について論じる。個人の嗜好を表す場合、推薦対象などに対する点数付けがよく用いられている。ただし、絶対的な点数付けは煩雑で個人に負担をかけやすい。我々の方式は、より容易な順位による嗜好の指定を基本とする。ただし、逆に順位を点数に変換する必要が生じる。そこで、順位を変換するためにべき法則で用いられるべき関数を用いた。プロファイルは、各メンバーごとに推薦の対象となるものをクラスの階層構造として表現し、嗜好の順位付けはクラス単位で行う。ただし、全クラスに関する順位を得ることは実用的には難しく、これを補うための順位推定についても論じる。本研究は京都の 230 箇所の名所を具体例とする京都観光案内システムの構築を目標に検討を行ったものである。

キーワード　パーソナライゼーション、プロファイル、e-commerce、情報推薦、グループ適応

1. はじめに

Web における全文検索や e-shopping では一般に誰が入力したかによらず、返される結果が同じであり、その個人の嗜好などが考慮されることは少ない。

一方、インタネットなどの普及に伴い、データ量は指数関数的に増えており、いわゆる「情報洪水」ということが言われている。その中から自分の要求に合ったものを見つけることはより難しくなってきている。

個人適応化

これらの問題を解決するために個人適応化(personalization) [Kawai04]が研究されてきた。個人のキー入力や e-shopping で購入した商品などの情報を「個人プロファイル」として記録しておき、その情報に基づいてその個人により適応した検索結果を返したり、個人の嗜好に合った商品を推薦する。

グループ適応化

個人適応化(personalization)の技術の延長として、グループのメンバーの嗜好や意思を尊重した推薦を行うためのグループ適応(group adaptation)の研究[Jameson04] が行われているが、個人適応に比べるとまだ研究は少ないように思われる。

グループ適応を行うことによって、リーダやメンバーの持つ知識の不足を補える、メンバーの嗜好や意思をより公平に反映できるなど多くの利点が期待できる。

グループ適応では、ユーザの嗜好を表すプロファイルの構成、嗜好の指定方法と、これらの情報に基づく推薦方式が重要な課題である。

グループが適応する場合、嗜好の似た複数のサブグループに分かれて適応することも考えられる。ただし、本論文で対象とするグループは､家族やクラス会など、全体で行動することを前提としたグループであり、分割については考えないことにする。

多数決

グループのメンバーの意思を反映させるためによく用いられるのが多数決である。ただし、木目細かい視点で見た場合、多数決の結果が必ずしもグループ全体の意思を反映しているとは限らず、少数派の意見が反映されにくい。

満足度の最大化

推薦においては、ユーザの嗜好に対する情報をもとに、対象物に対して点数付けを行い、点数の順に推薦するという方法がよく用いられる。グループ適応においてもこの方法が用いられている[Jameson04]。この方法によれば、多数決の少数派の意見が取り入れられにくいという問題を軽減できることが期待できる。我々もこの方式を採用した。我々は、点数を満足度と考えている。

嗜好の指定における問題点

ユーザが嗜好を指定する際、その作業が面倒と感じるようでは、グループ適応の実用化は難しい。この指定をなるべく簡便にすることが重要と考えている。

ユーザの嗜好を表す場合、点数付けがよく用いられ

る。[Jameson04]でも、ユーザの嗜好の指定において、５段階の点数付け(rating)が用いられている。我々は、この点数付けは、一般ユーザには煩雑であると考えている。また、嗜好を指定する場合、設問や指定する項目全てに対して答えることは､ユーザの負担を増やし、わかる部分だけについて指定するなどを考慮する必要がある。

基本的な考え方

我々のグループ適応に対して提案する方式の基本的な考え方は以下のとおりである。

(1) 順位による嗜好の指定

ユーザの嗜好を順位付け(ordering)で指定する。順位は相対的なものであり、絶対的な点数付けに比べ、ユーザの負担が少ないと考えられる。

(2) 順位の満足度への変換

順位で指定することにすると､順位を点数（満足度）に変換する必要が生じる。我々は、順位を点数に変換するために､２章で述べるように､べき法則(power law)で用いられるべき関数を用いて、順位を満足度に変換する方式を用いた。この方式は心理学的に明確な根拠に基づくものではないが､[Omura85]に述べられている人間の嗜好はジップの法則(Zipf's law)に適合する傾向があるということにヒントを得たものである。我々は比較的我々の経験に合っていると考えている。

(3) 順位の推定によるユーザ負担の軽減

全ての項目に対して、順位をつけることはユーザの負担を増す。我々は、ユーザが指定したい項目のみ指定し、指定されていないものの順位を推定する方式を採用した。

[Kamishima05]でも順位の指定を行い､順位相関を協調フィルタリングに用いている。我々の方式の大きな特徴は、順位を満足度を表す点数に変換することにある。また、点数の指定は自由度が高く、自分に有利になるよう操作される可能性がある。一方、順位を変換した点数にはべき変換による制約がかかっており、順位の点数への変換により、そうした行為の影響を軽減することが期待できる。

我々は、京都観光案内システムを具体的な目標として、システムを構築中である。このシステムでは、京都の 230 箇所の名所から、グループ全員の満足度を最大にする複数の名所を選び、それらを結ぶ観光ルートを画面上に表示する。本論文で述べる推薦方式は、このシステムに組み込むために検討し、採用したものである。

２．満足度の計算

この章では、推薦の対象となる個々の候補に対するグループの満足度の計算の方法について述べる。

２.１．グループの満足度

以下、グループはm人の個人

$$p_1, p_2, \cdots, p_m$$

から構成されるものとする。グループへの推薦の対象となるものの集合を

$$A = \{a_1, a_2, ..., a_n\}$$

とする。たとえば、旅行の際の観光地や、テレビの場合の番組である。個々の a_i をインスタンスと呼ぶことにする。最終的目標は、これらの中からグループの満足度を最大にするような g 個を選出して推薦することである。ただし、個々の対象物の満足度を定義することは面倒である。我々の基本的な考え方は A を

$$A = C_1 \cup C_2 \cup \cdots \cup C_N$$

$$C_i \cap C_j = \phi \quad (i \neq j)$$

と、複数のクラス $C_1, C_2, \cdots, C_N$ に分類し、これらのクラスから g 個のクラスの集合

$$C_{i_1}, C_{i_2}, \cdots, C_{i_g}$$

を選ぶことである。この g 個のクラスの組を「クラスプラン」と呼ぶ。たとえば、旅行の観光地の場合、観光地を平安、鎌倉、江戸、幕末というように時代を表すクラスにより分類することが考えられる。クラスプランは、たとえば、g=2 として、平安、幕末がこれに当たる。そして、グループの満足度を最大化するような最適クラスプランを選出し、その各クラスからそれぞれインスタンスを１つ選ぶ。なお、複数のインスタンスを選ぶことも考えられ、そのために我々は飽きを考慮しているが、このことについては簡単のため省くことにする。この最後に選ばれた g 個のインスタンスを「最適インスタンスプラン」と呼ぶ。この最適インスタンスプランを求めることが最終的な目標である。

各クラス C_i $(1 \le i \le N)$ に対しては、個人 p_h ごとの満足度 $s(p_h, C_i)$ が後述のように(2.2)として定義される。また、個人 p_h のクラスプランに対する満足度は、

$$ps(p_h, C_{i_1}, C_{i_2}, \cdots, C_{i_g}) = \sum_{j=1}^{g} s(p_h, C_{i_j})$$

によって定義する。すなわち、クラスプランの各クラスの満足度の総和である｡この個人の満足度をもとに、グループ全体のクラスプランに対する満足度は

$$gs(C_{i_1}, C_{i_2}, \cdots, C_{i_g}) = \sum_{j=1}^{g} ps(p_h, C_{i_1}, C_{i_2}, \cdots, C_{i_g}) \qquad (2.1)$$

によって定義する。すなわち、各個人のクラスプランに対する満足度の総和である｡各個人を審査員と考え、個人の満足度がクラスプランに対する評価と考えれば、総和で評価するという考え方はスポーツなどの世界でもよく採られる方式であり、自然であろう。

個人の優遇

(2.1)は、各個人を公平に扱ったものである。ただし、特定の個人を優遇したい場合がある。たとえば、金婚式を祝う家族旅行や、子供を楽しませるための旅行といった場合である。これを実現するために、個人ごとの満足度に重みをつけることが考えられるが、本論文では、簡単のためこの重みについては論じない。

ここまでグループ全体の満足度について述べてきた。残った問題として、個人の満足度をどう計算するかがある。以下この問題について述べる。

2.2. 個人の満足度

１章で述べたように、我々の方式では、順位を満足度に変換する必要がある。変換の方式は以下のとおりである。

個人の満足度をクラスをもとに以下のように定義する。今、N 個のクラスを、個人 p_h が好む順に

$$C_1, C_2, ..., C_N$$

と並べたとする[1]。このとき、C_k が選ばれることによる p_h の満足度が

$$s(p_h, C_k) = 1/k^w \quad (0 < w) \qquad (2.2)$$

に従うものと仮定する。図１に w=0.5, 1, 2 の場合のグラフを示す。なお、w を 0 に近づけると(2.2)のグラフは y=1 という直線に近づく。このことは、その属性の影響が少なくなることを意味する。

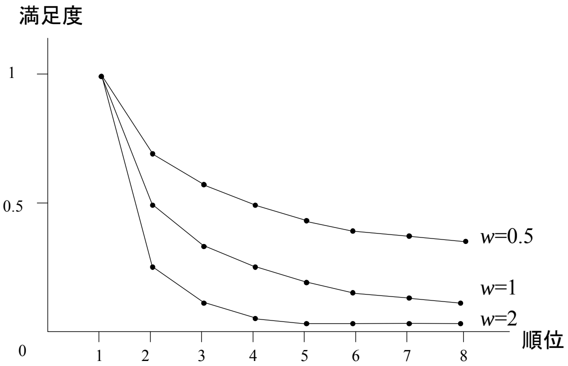

図1 べき法則のグラフ

変換式の背景

前述のように、(2.2)は[Omura85] に書かれている人間の嗜好もこのジップの法則に適合する傾向があるということにヒントを得たものである。ここでジップの法則やそれを一般化したべき法則(power law) について簡単に説明しておく。

ジップの法則は、小説「ユリシーズ」の中に現れる単語を頻度順に並べたとき、単語の順位 k と頻度の割合を%で表したものを f とすると、

$$f = 10/k$$

という関係が成り立つという発見からきている。すなわち、１番目の単語は小説の中に 10%, 2 番目は 5%, 3 番目は 3.3%、・・・現れることを意味している。順位が上なほど差がはっきりしているが、順位が下になるほどどんぐりの背比べになる。ジップの法則は、(2.2)で $w = 1$ とした場合に当たる。

確かに、好きな球団や歌手、音楽などを挙げる場合、１番目、２番目、３番目くらいまではははっきりしているが、それ以降は次第に区別が曖昧になり、どんぐりの背比べになるということが多いように思われる。

ジップの法則を一般化したべき法則は河川や都市の人口、所得などにもおおよそ当てはまると言われている。(2.2)はちょうどべき法則を表す式になっている。

嗜好の自己相似性

最近ネットワーク理論 [Barabasi02]が注目を集め、ネットワークにおいてべき法則が成り立つことが多いことがわかってきている。たとえば、Web のリンク数とそのリンク数をもつサイトの頻度の間にべき法則が成り立つ。さらに自己相似性ということが言われ、このべき法則は大きなネットワークで成り立つとともに、その部分的なネットワークでも成り立つということが言われている。

我々はこの理論にヒントを得て、個人の嗜好に対してもこの自己相似性が成り立つものと仮定する。この自己相似性の仮定は後で述べるクラス階層において下位クラスの順位を満足度に変換する際に意味を持つ。(2.2)と同様、この仮定も心理学的根拠に基づくものではないが、我々は経験によく合っていると考えている。

たとえば音楽を例にすると、前述のように音楽のジャンル（クラシック、ジャズなど）に対する順位を１，２、３位までは付けられるが、それ以降は次第に曖昧になるであろう。音楽の１つのジャンル、たとえばクラシックに絞って好きな作曲家を上げることを考えると、やはり同じ傾向があるように思われる。さらに、モーツァルトの曲に絞っても同様である。嗜好の自己相似性はこうしたことを意味する仮定である。

多数決と満足度による例

ここでは、簡単な例をもとに、多数決と満足度の比較を試みることにする。A, B, C 三人の旅行で A, B は平安時代に興味があり、第１、第２、第３希望は京都御所、平安神宮、八坂神社とし、幕末には全く興味がないものとする。C は幕末に興味があり、第一希望は寺田屋で、平安には全く興味がないものとする。多数決であれば、平安時代の３つの観光地が選ばれることになる。(2.2)において $w = 1$ とする。この場合、上記の仮定に基づくと３人の満足度の合計は、

1　各クラスが１つのインスタンスからなる場合は、インスタンスレベルで考えたことになる。

$(1+1/2+1/3)\times 2 = 3.67$

となる。満足度を最大化するのは、第三希望の八坂神社の代わりにＣの第一希望を入れる場合で、

$(1+1/2)\times 2 + 1\times 1 = 4.0$

である。 $w=2$ の場合は、前者は 2.72、後者は 3.5 と計算され、差が広がる。

人間はこうした判断を自然に行っているように思われる。上記のような状況の場合、１つはＣに譲るという場合が多いであろう。

属性とクラスの複合

クラスに対する個人の満足度を(2.2)によって定義した。しかし、前述のように N の値が大きくなると、後になるほど嗜好の順位を定めることが難しくなる。そこで以下に述べるような属性、クラス積というものを用いて、これをより容易にすることを考えた。

対象となるものには、一般に複数の属性がある。観光地であれば、時代、種類、重要度と言ったものである。ここでは、属性は

$T_1, T_2, \cdots, T_M$

の M 個からなるものとする。上の例で言えば、T_1 が時代、T_2 が種類、T_3 が重要度といった具合である。

この属性ごとの視点からクラスに分類することを考える。たとえば、時代であれば古代、中世、近代あるいはより詳細に、平安、鎌倉、・・・、幕末といったクラスに分類する。

属性 $T_i \ (1 \le i \le M)$ に対応するクラスの数を t_i、対応するクラスを

$C_{i1}, C_{i2}, \cdots, C_{it_i}$

と表すことにする。我々の狙いは、属性ごとのクラス数 t_i を比較的少なく抑えることにより、嗜好の順位を付けやすくすることにある。また、我々の仮定から、各属性ごとのクラスに対しても、(2.2)は成り立つ。ただし、重み w は一般に属性ごとに異なる。全体としては属性、およびそれぞれの属性に対応するクラスは

属性	クラス
T_1	$C_{11}, C_{12}, \cdots, C_{1t_1}$
T_2	$C_{21}, C_{22}, \cdots, C_{2t_2}$
	・・・
T_m	$C_{m1}, C_{m2}, \cdots, C_{mt_m}$

と表される。そして、各属性ごとに１つのクラス $C_{1i_1}, C_{2i_2}, \cdots, C_{Mi_M}$ を選び、それらの共通部分として得られる

$C_{1i_1} \cap C_{2i_2} \cap \cdots \cap C_{Mi_M}$

をクラス積と呼び、$< C_{1i_1}, C_{2i_2}, \cdots, C_{mi_m} >$ と表すことにする。このクラス積も一つのクラスであり、属性ごとのある特徴を共通に持つ対象の集合と考えられる。たとえば、＜古代、寺、世界遺産＞であれば、古代の世界遺産である寺の集合を意味する。クラス積の数は全体で、 $N = t_1 t_2 \cdots t_M$ 個となる。

クラス積に対する満足度

属性ごとのクラスには、個人ごとに嗜好の順位が付けられているものとする（順位の付け方については３章を参照されたい)。すなわち各クラスの満足度を計算することはできる。ここで問題になるのは、クラス積の満足度をどう計算するかということである。

基本的な考え方は、クラス積の満足度は各クラスの満足度の積になるということである。こう考える根拠について以下説明する。

		T_2: C_{21} 1	C_{22} 1/2	C_{23} 1/3
T_1	C_{11} 1	1	1/2	1/3
	C_{12} 1/2	1/2	1/4	1/6
	C_{13} 1/3	1/3	1/6	1/9

図**2** クラス積と満足度

簡単な例として、２つの属性 T_1、T_2 を考える。それぞれ、時代、種類と考えることにする。また、それぞれ C_{11}, C_{12}, C_{13}（それぞれ、中世、近代、古代）および C_{21}, C_{22}, C_{23}（それぞれ、寺、史跡、神社）と分類され、ある個人 p_h の嗜好もこの順番であるものとする。(2.2)における重み w は両属性とも１であるものとする。このとき、各クラス積の満足度を積によって計算したものを図２である。行が T_1 のクラス C_{11}, C_{12}, C_{13} に対応し、それぞれの満足度が横に示されている。また、列が T_2 のクラス C_{21}, C_{22}, C_{23} に対応し、同様に満足度が下に書かれている。表のマスが行と列のそれぞれのクラスに対応したクラス積を表し、中の値は積として計算された満足度を表している。本来であれば、これら９個のクラス積には、1, 1/2, 1/3, ・・・, 1/7, 1/8, 1/9 と満足度がつくはずである。計算された値はそうはなっていないが、我々は本来の値を近似あるいは推定するものと考えている。なお、１位同士、あるいは最下位同士のクラス積の満足度は 1, 1/9 と合っている。

積と考える理由のもう一つは、自己相似性の仮定で

ある。自己相似性の仮定からは積で計算することが自然に思われる。以上をまとめると以下の通りである。

個人 p_h のクラス積 $<C_{1i_1}, C_{2i_2}, \cdots, C_{Mi_M}>$ に対する満足度は、

$$s(p_h, <C_{1i_1}, C_{2i_2}, \cdots, C_{Mi_M}>) = \prod_{j=1}^{M} s(p_h, C_{ji_j}) = \prod_{j=1}^{M} 1/(k_j^{\,w_j})$$

で表される。なお、ここで個人 p_h の各属性 T_j ごとの C_{ji_j} に対する嗜好の順位は k_j、また属性の式(2.2)における重みを w_j であるとしている。

なお、個人のクラス積に対する満足度の簡単な例を表１に示す。なお、簡単のために $w_j = 1$ とした。

表１ クラス積の満足度の例

順位	クラス積に対する満足度
全て１位	1
1つだけ2位で残りは１位	1/2
1つだけ3で残りは1位	1/3
全て２位	$(1/2)^M$

３．プロファイルと順位付け

この章では、グループ適応化の基盤となるプロファイルについて述べる。

グループプロファイル

グループプロファイルは、個人プロファイルの情報をもとに、合成されるものである。本論文で用いるものは、その一番単純な形態としての個人プロファイルの集合体である。

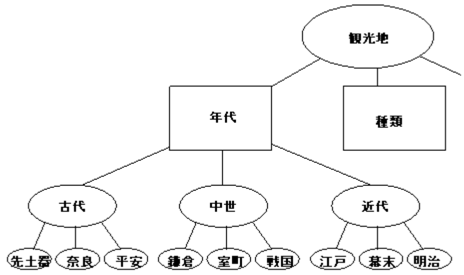

図3 静的なクラス構成による個人プロファイル

個人プロファイル

本論文における個人プロファイルは、図３に示すように、ユーザオントロジー[Kanjo04]の考え方に沿うもので、クラスの階層構造として形成される。ただし、ユーザオントロジーは追加されるインスタンスによりクラス構成を動的に変える。これに対し、本論文における個人プロファイルのクラス構成は各個人に共通で不変、すなわち静的なものである。

楕円がクラス、長方形が属性を表す。クラス構成は属性ごとに形成される。なお、今まで満足度を計算する説明ではクラス階層は属性ごとに１段のものとして説明してきた。しかし、図３に示すようにクラス階層は一般に属性ごとに多段である。満足度を計算する場合、全てのレベルのクラスを使うのではなく、あるレベルのクラス（たとえば最下層のクラス）を用いる。

今まで時代を大きく古代、中世、近代に分類してきたが、古代はさらに、先土器、奈良、平安とより詳細に分類することが可能である。必要であれば、さらに分類することも可能である。

クラスの順位付け

個人ごとの嗜好は、前述のように、各クラスに対する個人の嗜好の順位によって表されるものと考える。順位付けは、アンケートなどによる手入力によるものと、クラスに含まれるインスタンス数によって順位付けるなどの自動的なものが考えられる。自動的なものは本論文のテーマ外であるので、以下では手入力による順位付けに焦点を当てて述べることにする。順位付けは大きく、以下で述べる３つの方式が考えられる。各方式では、次のことを共通に考慮した。

(1) 任意のレベルの指定が可能

階層のどのレベルに順位付けしても構わない。よく知らない人は大雑把に指定したいであろう。図３で言えば、古代、中世、近代の３つの指定である。しかし、歴史に詳しい人はさらに細かく指定したいであろう。詳しい人でも他の時代には疎い可能性もある。これらがどのレベルを指定してもよいとする理由である。

(2) 部分的な順位付けが可能

手入力で順位を付けることを考えた場合、面倒であり、推薦において影響する重要な部分についてだけ最小限の手間で付けることを望むであろう。順位付けを部分的でよいとした理由はこのためである。順位を付けなかったものに対しては、４章で述べる順位推定により補完する。

３.１． 相対順位方式

この方式による順位付けの例を図４に示す。クラス階層において兄弟関係にあるものについて１，２、３、・・・と順に付けていく。この順位は親クラスの下の兄弟間での順位であり、相対的なものである。したがって、この順位を「相対順位」と呼ぶことにする。

この方式の利点は、順位が付けやすいということである。人間は順位を付ける場合、順位がトップに近い場合はつけやすいが、順位が下がってくるとどんぐりの背比べで順位がつけにくい。クラスを限り、その中で順位をつけることにより容易になる。

３.２．絶対順位方式

この方式による順位付けの例を図５に示す（図４、図５および図６とも同一の個人の嗜好を表すものである）。長方形で表された属性以下のクラス階層全体

を対象として１、２、３、・・・と順位を付けていく。同じ順位を付けることは許されるが、それは同程度に嗜好・興味を持っていることを意味する。この順位は、属性以下の全てのクラスを対象とした絶対的な順位であり、「絶対順位」と呼ぶことにする。

この方式では、上位クラスとその下位クラスの両方に同時に順位を指定することは許されない。これは、上位クラスと下位クラス間で不整合が起きることを防ぐためである。指定されなかったクラスについては、４章で述べる順位推定を行う。

この方式の利点は、嗜好の順位が明確な場合、嗜好をより正確に表すことができることである。

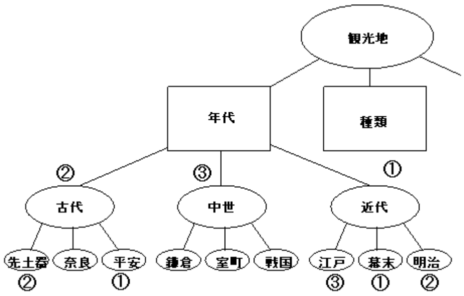

図4 相対順位方式

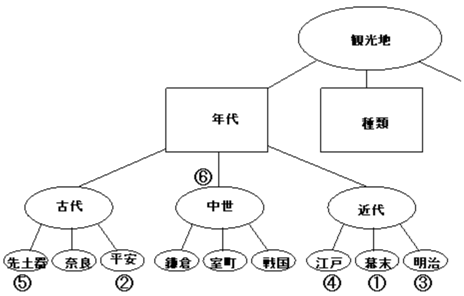

図5 絶対順位方式

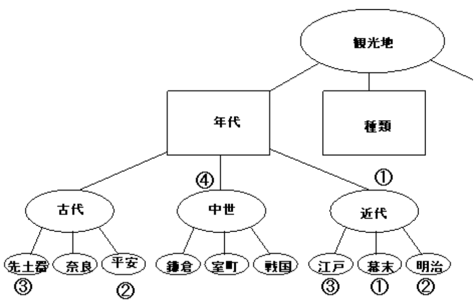

図6 混合順位方式

３.３. 混合順位方式

この方式は上述の２方式を混合させた方式で、両者の利点を取り入れることができる。この方式による順位付けの例を示したのが図６である。この方式では絶対順位方式によって、クラス階層全般に番号を付けていく。そして、上位クラスと下位クラスの両方に順位を付ける場合、下位クラスの順位を相対順位にする。この方式により、絶対順位方式ではできなかったクラスの下位クラスの順位をさらに詳細に指定することが可能となる。

４. 順位推定

この章では、順位が指定されていない部分の順位推定の方式について述べる。この順位推定により、ユーザの指定の自由度を増すことができる。

まず、推定規則を列挙することにする。なお、説明を簡単にするために、(2.2)の重みを省いて議論する。重みを考慮した方式はここでの議論から容易に導くことができる。

４.１ 上位・下位クラスの順位関係

順位推定について述べる前に、その際に重要となる上位クラス C とその n 個ある下位クラス $C_1, C_2, \ldots, C_n$ の関係についての基本的な考え方を述べることにする。上位クラスの絶対順位を R とする。下位クラスには、$C_1, C_2, \ldots$の順で相対順位がついているものとする。

まず、上位クラスの順位 R や他の情報から C_1 の絶対順位を推定する。この推定方法はいろいろ考えられる。最も単純な方法の一つは、順位 R を C_1 の順位とすることである。この方式を「単純遺伝方式」と呼ぶ。

推定された C_1 の絶対順位を a (C_1)と表すことにする。このとき、C_2 , C_3, …の絶対的順位は $2a$ (C_1), $3a$ (C_1), …あるとするのが基本的な考え方である。こう考えると、C_1 , C_2 , C_3, …の満足度は、$\alpha = a$ (C_1)とすれば、(2.2)から、それぞれ $1/\alpha$, $1/(2\alpha)$, $1/(3\alpha)$, … となる。

この結果は、３.２で述べた自己相似性と整合性の取れたものである。また、２番目、３番目の満足度は、それぞれ一番目の満足度 $1/\alpha$ の 1/2, 1/3 になっており、(2.2)とちょうど整合性が取れている。

４.２ 順位の推定

ここでは、指定されていない順位の推定の方法について述べる。

４.２.１ デフォルト順位

同じレベルのクラスの一部が指定されてそれ以外が指定されていない場合について説明する。

(1) 相対順位方式の場合

兄弟関係にあるサブクラスの一部の順位が指定されている場合、指定された最大の順位を m とすると、指定されていないサブクラスの相対順位は全て m+1 であるものとする。

例４.１ 図７の(a)に相対順位方式におけるデフォルト順位の例を示す。この図は上位クラスの順位が 2, 下位クラスに左から 1, 2, 3 位が付けられているが、数字が書かれていないクラスは、順位が指定されていないことを意味する。最大の順位は m = 3 であるので、図

(a)の矢印の下に示すように、イタリックで書かれた順位としてデフォルト順位を定める。

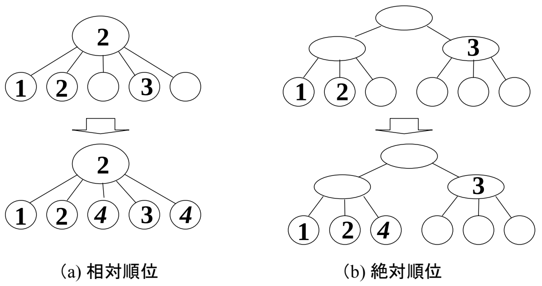

図7 デフォルト順位

(2)絶対順位方式の場合

絶対順位の中で指定された最大の順位を m として、他は全て m+1 であるものとする。相対順位の場合と同様に、図７の(b)に絶対順位のデフォルト順位を定める例を示す。他の部分は、以下に述べる遺伝、逆遺伝の規則から推定される。なお、何も指定されていない場合は、最大順位は 0 とする。

(3) 混合順位方式の場合

上記(１), (２) を適宜適用して、順位を推定する。

４.２.２ 遺伝

ここでは、下位クラスの絶対順位がわかっていない場合の推定について述べる。

上位クラス C の絶対順位が R とする。その n 個ある下位クラス $C_1, C_2, \ldots, C_n$ の相対順位が指定されていない場合と指定されている場合の２つに分けてその推定について説明する。なお、相対順位の場合、一番上の兄弟関係にあるクラスの順位は、相対順位であるが、これを絶対的な順位と見なすことにする。

(1) 相対順位が指定されていない場合

指定はされていないが、実際には順位はあるはずである。その一番目の絶対順位は R や他の情報から推定される。その値をαとする。すると、この下位クラスの順位は、４.１で述べたことから、α、2α, ..., nαと順位付けされるはずである。ただし、どのサブクラスがその順位かはわからない。そこで、下位クラスの平均的な絶対順位 r や満足度 s を考えることにする。平均的な満足度 s は

$$s = (1/n)\sum_{i=1}^{n} 1/(i\alpha)$$

で与えられる。

$$\zeta(n) = (1/n)\sum_{i=1}^{n} 1/(i)$$

とおくことにする。この値は、１位から n 位まで絶対順位が付けられたクラスの平均満足度と考えられるものである。したがって、s は

$$s = \zeta(n)/\alpha \qquad (4.1)$$

と表される。絶対順位と満足度の関係は (2.2)で与えられるので、$r = 1/s$ である。すなわち、(4.1)から、

$$r = \alpha/\zeta(n)$$

で与えられる。すなわち、n 個のサブクラスはすべて絶対的順位 r および満足度 s を持つものと推定する。単純遺伝方式の場合、$\alpha = R$ であるから、

$$s = \zeta(n)/R$$
$$r = R/\zeta(n)$$

と計算される。

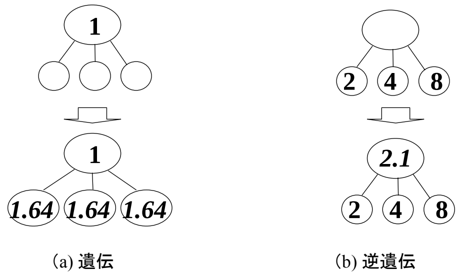

図8 遺伝と逆遺伝

例４.２ 図８(a)に例を示す。単純遺伝方式の場合、$\zeta(3) = (1/3)(1+1/2+1/3) = 11/18$ より、$r = 1.64$ と計算され、図８(a)の矢印の下のように絶対順位が推定される。

(2)相対順位が指定されている場合

（１）の場合と同様、その一番目の絶対順位を R から推定し、その値をαとする。また、指定された下位クラス $C_1, C_2, \ldots, C_n$ の相対順位を $r_1, r_2, \ldots, r_n$ とする。このとき、それぞれのサブクラスの絶対順位は順に、αr_1, αr_2, ..., αr_n と推定される。

４.２.３. 逆遺伝

ここでは、上位クラスの絶対順位がわからない場合の推定について述べる。

上位クラス C の絶対順位がわからず、その n 個ある下位クラス $C_1, C_2, \ldots, C_n$ の絶対順位がそれぞれ $r_1, r_2, \ldots, r_n$ と指定されているものとする。このとき、上位クラス C の絶対順位 R を次のように推定する。

下位クラスの平均満足度 s は、

$$s = (1/n)\sum_{i=1}^{n} 1/r_i \qquad (4.2)$$

で計算される。(4.1)より、

$$\alpha = \zeta(n)/s \qquad (4.3)$$

と計算される。この α の値は、サブクラスの中で一番順位が高いものの順位を、４.１で述べた基本的な考え方から逆に割り出したものである。一般には、r_1 の値とは異なるが、サブクラスの n 個の順位から割り出したものであり、こちらの方がよりサブクラス全体の性

質を表しているものと考えられる。このαの値から、（２）で用いた方法を逆に用いて、Rを推定する。

例４.３　図８(b)に逆遺伝の例を示す。(4.2)より、$s = 7/24$である。したがって、(4.3)より、$\zeta(3) == 11/18$であるので、$\alpha = \zeta(n)/s == 2.10$と計算される。単純遺伝方式の場合$R = \alpha$であるから、上位クラス C の絶対順位は 2.1 と推定される。

５．具体例による説明

本章では、提案方式による動作について観光地を例に具体的に説明する。観光地の属性としては、時代と種類を用いる。時代のクラスは、古代、中世、近代、また種類のクラスは神社、寺、史跡とする。また、グループを構成するメンバーは、A, B, C, D, E, F の６名とし、各個人の嗜好の順位は表２のとおりとする。６名は大きく３つのサブグループ S1, S2, S3 に分けられる。S1 の嗜好は S2 とはちょうど逆になっている。S3 は中間的だが、S2 に近い嗜好を持っている。表は２つの属性のそれぞれに対する嗜好の順位を示している。

表２　個人の嗜好の順位

サブグループ	個人	嗜好の順位		
		１位	２位	３位
S1	A, B, C	古代 神社	中世 寺	近代 史跡
S2	D, E	近代 史跡	中世 寺	古代 神社
S3	F	中世 寺	近代 史跡	古代 神社

表3 クラス積ごとの満足度

(a) ケース１（年代のw=1）

クラス積		グループ全体の満足度	ランキング
時代	種類		
古代	神社	3.33	1
	寺	2.17	4
	史跡	1.83	8
中世	神社	2.17	4
	寺	2.25	3
	史跡	2.00	6
近代	神社	1.83	8
	寺	2.00	6
	史跡	2.58	2

(b) ケース２（年代のw=2）

クラス積		グループ全体の満足度	ランキング
時代	種類		
古代	神社	3.11	1
	寺	1.72	3
	史跡	1.27	6
中世	神社	1.25	7
	寺	1.63	4
	史跡	1.25	7
近代	神社	1.08	9
	寺	1.42	5
	史跡	2.17	2

なお、(2.2)における重みを次の２つについて計算する。

ケース１：　時代、種類ともに w=1

ケース２：　時代 w=2, 種類 w=1

このときのクラス積ごとのグループ全体の満足度は、表３に示すとおりである。

まず、時代の重みが違うケース１と２で、クラス積のランキングが変動していることがわかる。時代の重みが重いケース２では、時代の影響が強く出ている。ケース１では、１，２，３位がそれぞれサブグループ S1, S2, S3 の１位に対応しており、多数派の影響がケース２に比べ少ない。ただ、どちらのケースにしても、多数派が有利であることは確かであるが、少数派の嗜好も反映されていることがわかる。この結果は、グループの満足度最大化という観点から概ね妥当であると考えている。

６．まとめ

本論文では、嗜好の順位を基本としたグループ適応のための推薦方式について論じた。提案方式の特徴は以下のとおりである。

・嗜好の指定が容易

順位による指定は、相対的であり、人間には指定しやすい。順位推定により、さらに補完される。

・順位の満足度への変換

満足度最大化を目的とするため、嗜好の順位を満足度に変換する必要がある。このため、べき関数を用いて順位を満足度に変換した。

今後の課題として、実際のシステムによる有効性の検証が必要であると考えている。

文献

[Barabasi02] Albert-Laszlo Barabasi, “新ネットワーク思考—世界のしくみを読み解く”, NHK 出版, (2002).

[Kanjo04]　官上 大輔, 河合 由起子, 田中 克己, “A3: オントロジーの共有によるユーザ適応のためのフレームワークの提案”, DEWS2004 I-3-2, (2004).

[Kamishima05]　神嶌敏弘, “なんとなく協調フィルタリング--- 複数の順序応答に基く推薦”、人工知能学会全国大会（第１９回）1F4-05, (2005).

[Kawai04]　河合 由起子, 官上 大輔, 田中 克己,“興味と好みに基づく複数 Web ページの情報融合・提示システムの検討”, DEWS2004 2-C-1, (2004).

[Jameson04] Anthony Jameson: More than the sum of its members: challenges for group recommender systems, Proc. of the working conference on Advanced Visual Interfaces, AVI2004, pp.48-54, (2004).

[Masthoff04] Selecting News to Suit a Group of Criteria: An Exploration, http://www.di.unito.it/~liliana/TV04/FINAL/masthoff.pdf.

[Omura85]　大村 平, “多変量解析のはなし”, 日科技連出版社、pp.33-34, (1985).