DEWS2005 6-A-01

プライバシとサービス品質のトレードオフを考慮した個人情報制御機構の提案

宮本　崇弘† 竹内　亨† 奥田　剛†† 春本　要††† 有吉　勇介†††† 下條　真司†††††

† 大阪大学大学院情報科学研究科　〒 565–0871 大阪府吹田市山田丘 1–5
†† 奈良先端科学技術大学院大学情報科学研究科　〒 630–0192 奈良県生駒市高山町 8916–5
††† 大阪大学大学院工学研究科　〒 565–0871 大阪府吹田市山田丘 2–1
†††† 尾道大学経済情報学部　〒 722–8506 広島県尾道市久山田町 1600
††††† 大阪大学サイバーメディアセンター　〒 565–0047 大阪府茨木市美穂ヶ丘 5–1
E-mail: †{miyamoto,stakeuti}@ist.osaka-u.ac.jp, ††okuda@is.naist.jp, †††harumoto@eng.osaka-u.ac.jp,
††††y-ariyoshi@onomichi-u.ac.jp, †††††shimojo@cmc.osaka-u.ac.jp

あらまし　情報推薦などの個人化サービスを受けるためには個人情報を開示する必要がある．多くの個人情報を開示することで，さらに個人化されたサービスを受けることができる可能性があるが，一方で情報漏洩や情報流用などによりプライバシが侵害される危険性も秘めている．そこで本研究では，開示する個人情報の種類やその粒度を制御する機構 GrIP（Granularity Control Mechanism based on Person Identification Probability）を提案し，シミュレーションによってその有効性を示した．

キーワード　プライバシ, プロファイル, 情報制御, 情報粒度

A Proposal for Profile Control Mechanism Considering Privacy and Quality of Personalization Services

Takahiro MIYAMOTO†, Susumu TAKEUCHI†, Takeshi OKUDA††, Kaname HARUMOTO†††, Yusuke ARIYOSHI††††, and Shinji SHIMOJO†††††

† Graduate School of Information Science and Technology, Osaka University
1–5 Yamadaoka, Suita-shi, Osaka, 565–0871 Japan
†† Graduate School of Information Science, Nara Institute of Science and Technology
8916–5 Takayama, Ikoma-shi, Nara, 630–0192 Japan
††† Graduate School of Engineering, Osaka University
2–1 Yamadaoka, Suita-shi, Osaka, 565–0871 Japan
†††† Faculty of Economics, Management and Information Science, Onomichi University
1600 Hisayamada, Onomichi-shi, Hiroshima, 722–8506 Japan
††††† Cybermedia Center, Osaka University　5–1 Mihogaoka, Ibaragi-shi, Osaka, 567–0047 Japan
E-mail: †{miyamoto,stakeuti}@ist.osaka-u.ac.jp, ††okuda@is.naist.jp, †††harumoto@eng.osaka-u.ac.jp,
††††y-ariyoshi@onomichi-u.ac.jp, †††††shimojo@cmc.osaka-u.ac.jp

Abstract　In order to enjoy the benefit of personalization services such as recommendation services, it is necessary to disclose personal information. In general, the more personal data we disclose, the more the quality of personalization services may improve. However, it also increases privacy concerns such as information leak. To cope with this problem, we propose a profile control mechanism GrIP (Granularity Control Mechanism based on Person Identification Probability) that controls the number and granularity of disclosing personal information.

Key words　privacy, profile, information control, granularity of information

1. はじめに

近年，それぞれのユーザに合わせた形で様々なサービスを提供する個人化サービスが増えてきている．実社会においては，地域にあわせて味が異なる食料品や，TV 番組の視聴者の年齢層に応じた CM などがあげられる．またインターネット上では，登録した個人情報にあわせた情報配信サービスや，購入履歴に応じた商品推薦サービスなどが提供されている．今後ユビキタス環境が整うにつれて，さらに多くのサービスが個人に合わせて提供されていくと思われる．

これらの個人化サービスを受けるためには，ユーザはサービス提供者に何らかの形で個人情報を開示する必要がある．現在は，年齢や性別のような属性情報を入力したり，カテゴリ分けされたキーワードを選択したり，商品を購入したりすることでサービス提供者に個人情報を開示している．今後ユビキタス環境が整うにつれて，年齢，性別，住所のような静的な属性情報だけでなく，現在地のような動的な属性情報，体温，心拍数のような生体情報，行動履歴や購入履歴のような履歴情報，周囲の気温，湿度のような環境情報など，様々なコンテキスト情報が容易に収集，利用できるようになっていくと考えられる．

このような様々なコンテキスト情報を個人化サービスから透過的に扱えるようにするために，我々はコンテキスト情報の異種性を隠蔽して扱うことのできるサービス記述言語 SDL-UP（Service Description Language for User Profile）を提案してきた [1]．SDL-UP を用いることで，サービス提供者はユーザに要求する個人情報を容易に指定することができる．しかし SDL-UP ではサービス提供者の要求のみ考慮しているため，個人情報の取り扱いに対するユーザの要求を反映することが困難である．さらに，多くのサービスを利用することで個人情報を開示する機会が増えるため，サービスに開示した個人情報がユーザの意図しない用途に流用される危険性が増えることが考えられる．したがって，サービス側の要求だけでなく，ユーザ側の要求であるプライバシ保護を考慮した個人情報の開示が必要である．

このようなプライバシの保護を実現する手法として，すべての個人情報を開示するのではなく，サービスを受けるために最低限必要な情報のみを開示するという手段が考えられる．たとえば，道案内サービスのような位置に応じたサービスの場合，現在地や今後のスケジュールから得られる目的地といった情報のみをサービス提供者に開示すれば，ユーザが満足するサービスを受けられると思われる．さらに，サービスに開示する情報の粒度を変更することでプライバシ保護を実現する手段も考えられる．たとえば，天気予報サービスの場合，詳細な位置情報を開示しても得られるコンテンツは高々市区町村レベルであるため，開示する位置情報を市区町村レベルの粒度に変更してもユーザが要求するコンテンツは十分含まれていると考えられる．しかし，飲食店推薦サービスの場合，現在地情報を都道府県レベルに変更するとユーザの要求しないコンテンツも多く含まれるため，ユーザの満足度は減少すると考えられる．このように，サービスによっては登録した個人情報や収集されたコンテキスト情報がすべて必要とされないが，サービスの品質に大きく影響する個人情報の種類もあると考えられる．したがって，開示する個人情報の種類やその粒度を適切に制御することで，プライバシを保護した上で，十分な品質のサービスを受けることができると考えられる．

このように，個人情報の種類やその粒度を制御するためには，個人化サービスの要求とユーザの意図のどちらも反映させる必要がある．しかし，個人情報の種類やその粒度だけでなく，個人化サービスも多種多様であると考えられ，ユーザが利用したことのない種類の個人情報やサービスも考慮する必要がある．さらに，どの個人情報を開示することでどの程度プライバシ侵害の恐れがあるかをユーザが把握することは困難である．したがって，ユーザの意図を反映させるためには，これらの複雑な条件を網羅しつつ，ユーザによる判断が容易な制御方法が必要であると考えられる．

そこで本研究では，プライバシの保護を目的とした，開示する個人情報の種類やその粒度を制御する機構 GrIP（Granularity Control Mechanism based on Person Identification Probability）を提案する．本機構では，情報開示に伴うプライバシ侵害のリスクの客観的な評価基準を定義し，個人情報開示のリスクと個人化サービスの品質のトレードオフを考慮した個人情報の粒度調整を実現する．

2. 個人情報制御機構

本章では，提案する個人情報制御機構 GrIP の概要について述べる．さらに，既存の個人情報制御手法の問題点について述べる．

2.1 提案機構の概要

一般的に，個人化サービスはユーザが開示した情報をもとにサービスをユーザに適応させるため，個人情報を多く開示することでより個人化されたサービスが受けることができると思われる．しかし，1 章で述べたように，ユーザにはできるだけ個人情報を開示せずに個人化サービスを利用したいという要求がある．このように，ユーザが要求する個人情報の制御を実現しようとすると，プライバシ保護とサービス品質のトレードオフが生じる．ただし，ここでのサービス品質とは，個人化サービスを受けた結果得られたコンテンツに対してのユーザの満足度を指すものとする．

このトレードオフを解決するためには，まず，ユーザが要求するプライバシを客観的に評価できる基準が必要である．評価基準はユーザが指定するものであるため，1 章で述べたように，様々な情報の種類やその粒度を統一的に扱うことができ，かつ，ユーザが容易に設定できる必要がある．その評価基準をもとに個人情報の種類の選別やその粒度の変更を行い，プライバシ保護を実現する．そこで本研究では，このような評価基準として 3 章で述べる特定確率を提案する．ユーザが要求するプライバシ保護を実現するということは，特定確率がユーザの指定よりも低くなるように，開示する個人情報を制御することを指す．

本研究では，プライバシ保護とサービス品質のトレードオフを特定確率を用いて以下のように解釈する．

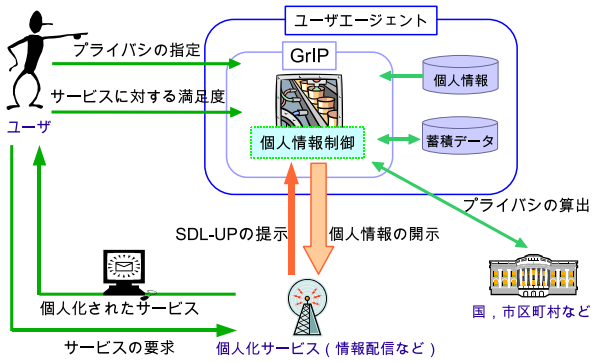

図 1　GrIP の概要

Fig. 1　Overview of GrIP

- 特定確率がユーザの指定よりも高い情報に制御された場合，ユーザはその情報の開示を受け入れない．
- 特定確率がユーザの指定よりも低い情報に制御された場合，ユーザは最低限満足する．
- 特定確率がユーザの指定よりも低い情報に制御された場合，ユーザはサービス品質が高い情報に制御するほうがより満足する．
- サービス品質が同程度の場合，特定確率が低い開示情報に制御するほうがユーザは満足する．

ユーザが指定した特定確率を満たす開示情報の組み合わせが複数存在する場合，その中からトレードオフを解決している情報に制御する必要がある．しかし，どの個人情報を開示することでどの程度ユーザが満足できるサービスを受けることができるかはわからない．したがって，ユーザの要求するプライバシを保護しつつ，属性情報ごとのサービス品質への影響度を調べ，トレードオフを解決する個人情報に制御するアルゴリズムが必要である．

以上のような要求をもとに，提案機構である GrIP の概要を図 1 に示す．GrIP は，ユーザの端末内で個人情報の管理などを行うユーザエージェントの一機能として提供する．ユーザは要求するプライバシの開示レベルを事前に指定する．GrIP では，ユーザの要求に適した範囲内でプライバシとサービス品質のトレードオフを解決する情報の種類やその粒度の組み合わせを探索し，個人情報を制御する．制御した結果得られたコンテンツに対してユーザの満足度をフィードバックとして取得し，組み合わせの探索に利用する．GrIP では個人情報の粒度の変更のみを行う．これによって，既存の個人化サービスは利用している属性情報に関して様々な粒度を受け付けるように変更するだけでよい．また，制御された情報は粒度が異なるだけであるため，制御結果をユーザが容易に理解し，監視することができる．

2.2　関 連 研 究

ユーザが要求するプライバシ保護を記述する言語として P3P（Platform for Privacy Preferences）[2] があげられる．P3P では，個人情報を Web サーバに収集する時の基準を規定している．ユーザは事前にポリシを設定しておき，サイトから個人情報が要求されるとそのサイトのプライバシポリシを調べてユーザのポリシに沿っているかを判断できるようにする．この判断を自動的に行うための規約として APPEL（A P3P Preference Exchange Language）[3] がある．P3P や APPEL では，ユーザとあらゆるエンティティ間で直接的または間接的にプライバシポリシのネゴシエーションができる仕組みと，受けるサービスに直接関係ない個人情報の提供を拒否できる統一された仕組みの提供を目指している．しかし，プライバシレベルの定義と通信相手や利用端末，時間，場所等によってプライバシレベルを変える仕組みは提供されていない．また文献 [4] では，サイトのプライバシポリシにそってデータが収集されているかの確認方法がない，サイトを見るためにユーザからの個人情報の提供を強要する可能性があるなどの問題点も指摘されている．

個人情報の情報量によってプライバシ侵害のリスクを数値化し，それをもとにユーザが要求するプライバシ保護を指定させる研究として，LooM（Loosely managed privacy protection Method）[5], k-anonymity [6] などがあげられる．LooM では，個人情報の情報エントロピーを計算することで数値化を行っている．エントロピーを用いることで情報の種類に依存しない数値化を行うことができる．しかし，LooM では開示する情報に対するリスク評価とその評価にもとづいた開示する情報の選別しか行っていない．また，情報の開示によって得られるサービスの品質が考慮されていない．k-anonymity では同じ情報を持つ人数を一定以上に制御することでプライバシ保護を行っている．k-anonymity は特定のサーバに含まれる複数のユーザの情報を開示する際の制御を目的としているため，制御を要求するユーザの全ての情報を考慮して制御を行う．しかし，本研究ではそれぞれのユーザが個別に多数のサービスを利用する環境を想定しているが，k-anonymity を利用するためには全てのユーザの属性情報を知る必要があるため不可能であり，現実的ではない．

プライバシを考慮して個人情報を制御する研究として PPNP（Privacy Profile Negotiation Protocol）[7] がある．この研究ではユーザとサービス提供者の間でネゴシエーションを行い，個人情報を制御している．サービス提供者はどの情報，粒度を要求するかを指定し，ユーザ側でユーザが設定したポリシにしたがって開示する情報の粒度を選択する．しかし，開示する情報の粒度をユーザ側で設定することはできるが，設定をサービスごとに行わなくてはならず，新しいサービスを受ける場合には新たにルールを設定する必要がある．さらに，ユーザが意図した粒度に制御するためには，それぞれの属性情報ごとにユーザがルールを設定する必要があるため，ユーザが設定する項目は莫大なものになってしまう．サービス側もユーザに要求する情報の種類や粒度をリストアップするか情報を加工する制御ルールを記述する必要がある．また，ルールによって加工された情報がユーザの要求を満たしているかを確認することは困難である．したがって，処理結果によってどの程度プライバシが保護されているかをさらに評価する必要がある．

また，開示する情報の粒度に応じてサービスの内容を変更する研究がある [8]．この研究ではサービス享受と個人情報提供の間のトレードオフに注目して情報制御を行っている．しかし，いくつかのシナリオについてのトレードオフを例示しているのみで具体的な制御について述べられていない．

3. 特定確率

本章では，プライバシの評価基準である特定確率を定義し，その特徴について述べる．特定確率を用いることで，図 1 におけるユーザによるプライバシの指定を実現する．

3.1 定　　義

個人情報を開示した場合，その開示した情報によってどの程度プライバシが侵害される可能性があるかを本研究では開示リスクと呼ぶ．開示リスクには，開示した情報をもとに個人が特定されるリスク，私生活が暴露されるリスクなど，様々なものが考えられる．本研究ではこれらの開示リスクのうち，開示した情報をもとに個人が特定されるリスクについて考える．このリスクを特定確率と定義する．特定確率とは，ある属性情報を持つ集団において属性情報を開示した当人に行き当たる確率のことである．このように数値化をすることで，客観的な評価基準をもとにユーザは要求するプライバシ保護を容易に判断することができる．特定確率を式 1 に示す．

$$\text{特定確率} = \frac{1}{\text{該当人数}} \tag{1}$$

該当人数とは，開示された個人情報と同じ属性値を持つ人数を指す．一般的に該当人数が少ない場合，1 人該当人数が増えることで開示リスクは大きく減少する．一方，該当人数が多い場合は 1 人該当人数が増えても開示リスクは大きく変化しない．特定確率では該当人数の逆数をとることで，これらを考慮したリスクを表現している．

3.2 算 出 方 法

個人情報に含まれる属性情報としては，年齢情報や，性別情報，職業情報，位置情報などが考えられる．特定確率を算出するためには，これらの属性情報に関してユーザと同じ値を持つ人数を調査する必要がある．年齢情報や性別情報，職業情報などは現在でも国勢調査や市区町村の統計情報として得ることができる．位置情報は，位置をもとにした P2P ネットワーク [9] や位置情報のプライバシ制御を行う研究 [10] で提案されているフレームワークなどを用いることで，プライバシを考慮したうえで特定の地域に存在する人数を調査することが可能である．これらをもとに属性情報ごとの該当人数を求める．

開示する個人情報の該当人数は，それぞれの属性情報の該当人数を用いて推定する．属性情報を一つしか持たない場合は，その属性情報の該当人数が開示する個人情報の該当人数になる．複数の属性情報 $(a_1, a_2, \cdots, a_n)$ を持つ個人情報の該当人数 $\mathbf{P}$ は，式 2 を用いて推定する．

$$\mathbf{P} = P(a_1) \times \prod_{i=2}^{n} \frac{P(a_i)}{Q(a_i)} \qquad (n \geqq 2) \tag{2}$$

$Q(a_i)$ とは，属性情報 a_i の該当人数を求めた統計情報の母集団の人数を指す．たとえば，国勢調査を用いて性別情報の該当人数を求めた場合，日本の人口や吹田市の人口などを指す．$P(a_i)$ とは，$Q(a_i)$ のうち，ユーザと同じ値を持つ人数を指す．たとえば，日本の人口のうち「男性」という値を持つ人数を指す．

ここで利用する属性情報ごとの該当人数を厳密に求めることは困難である．しかし，特定確率は指標であり，また該当人数は時々刻々と変化すると考えられるため，厳密な値を扱う必要はない．したがって，それぞれの属性情報の該当人数は独立であると考え，統計データなどを用いて概算すればよい．

3.3 特　　徴

特定確率には以下のような特徴があると考えられる．

- ユーザが要求するプライバシの容易な指定

個人情報は年齢情報や位置情報などがひとつだけ含まれている場合や，年齢情報と位置情報のどちらも含まれている場合，位置履歴情報が含まれている場合など，様々なものが考えられる．ユーザがルールを用いて要求する情報開示リスクを設定する場合には，これらの組み合わせをすべて網羅しなくてはならない．しかし，特定確率を用いることでこれらの違いを意識することなく，要求する特定確率をひとつ設定するだけで統一的に個人情報を制御することができる．

- リスク評価基準

特定確率を用いることで，ユーザの要求する情報開示リスクの設定だけでなく，既存の個人情報制御機構の処理結果に対するリスク評価を行うことができる．

- 個人情報の分散管理

特定確率を算出するためには国や市区町村，位置情報管理フレームワーク [10] などに個人情報を開示する必要がある．しかし，それぞれの属性情報ごとに開示することができるため，複数の情報を同時に開示する場合に比べて特定確率が低くなり，安全であると思われる．また，制御されていない個人情報をそれぞれのユーザの端末で管理，制御することができるため，ユーザごとに個人情報を分散して管理することができる．

4. 粒度調整アルゴリズム

本章では，プライバシ保護とサービス品質のトレードオフを解決するためのアルゴリズムについて述べる．

属性情報はそれぞれ異なる粒度を持ち，その粒度の組み合わせも複数存在する．2.1 節で述べたように，ユーザの指定した特定確率を満たす組み合わせが複数存在する場合，その組み合わせの中からプライバシ保護とサービス品質のトレードオフを解決した組み合わせを選択する必要がある．しかし，個人化サービスによっては，そのサービス品質は開示情報の粒度だけでなく，その組み合わせにも応じて変化することが考えられる．したがって，未知の個人化サービスを利用するごとに属性情報の粒度がサービス品質にどの程度影響するかを調べる必要がある．そこで本研究では，それぞれの属性情報のサービス品質への影響度を調べ，トレードオフを解決する粒度の組み合わせを探索するランダム探索アルゴリズムとトップダウン探索アルゴリズムの二種類の粒度調整アルゴリズムを提案する．

4.1 ランダム探索アルゴリズム

ランダム探索アルゴリズムでは，ランダムに粒度の組み合わせを選択していき，トレードオフを解決している組み合わせを探索していく．特定確率とサービス品質のトレードオフは式 3 で定義する適格度を用いて解決する．

$$\text{適格度} = \alpha \times \text{サービス品質} + (1 - \text{特定確率}) \tag{3}$$

α はサービス品質に対する重みである．α の大きさによって同程度のサービス品質の範囲が決まる．サービス品質が高く，特定確率が低いほど適格度が大きくなるため，適格度が大きい組み合わせを選択することでトレードオフを解決している組み合わせを選択することができる．

以下にアルゴリズムの手順を示す．

（1） 要求された特定確率を満たす粒度の組み合わせを選別

（2） 各組み合わせの適格度を算出

（3） 適格度をもとに確率的に組み合わせを決定

（4） 特定確率とユーザからのフィードバックを蓄積

本研究では，確率的な組み合わせの決定方法に，適格度の大きい組み合わせほど選択される確率が高くなるソフトマックス行動選択規則 [11] を利用する．ソフトマックス行動選択規則を式 4 に示す．

$$P(a) = \frac{\exp(Q(a)/\tau)}{\sum_{i \in V} \exp(Q(i)/\tau)} \tag{4}$$

ここで $Q(a)$ は組み合わせ a における適格度，τ は正定数，V は組み合わせの集合である．

ユーザからサービスに対する満足度のフィードバックを蓄積する方法について，本研究では粒度の組み合わせごとの蓄積とそれぞれの属性ごとの蓄積の二種類を提案する．

粒度の組み合わせごとの蓄積方法では，その組み合わせの実際のサービス品質をもとに最適な組み合わせを決定する．この蓄積方法では，蓄積データから組み合わせの過去の正確なユーザ満足度が取得できる．しかし，一回のフィードバックからひとつの組み合わせに対してしか蓄積することができないため，探索速度は遅く，確率的な組み合わせの決定方法によっては局所解に陥ることが多くなると考えられる．また，位置情報のように時間とともに属性情報が変化する環境においては，同じ粒度の組み合わせであってもユーザ満足度が変化することが考えられる．そのため，蓄積データから取得したユーザ満足度が現在のユーザ満足度と一致しない可能性がある．

一方，属性ごとの蓄積方法では，その属性がサービス品質に与える影響を蓄積する．この蓄積方法では，一回のフィードバックから複数の組み合わせに対して蓄積が行えるため，探索速度を速めることができる．属性ごとの蓄積方法では，その属性の粒度の特定確率をキー値としてユーザ満足度を蓄積する．図 2 に特定確率をキー値とした蓄積方法を示す．(1) ～ (3) の各点はそれぞれの特定確率における蓄積されているユーザ満足度の値を示す．蓄積されている特定確率の最小値（1）よりも小さい特定確率（a）のユーザ満足度を取得する場合は，0 を返す．蓄積されている特定確率の最大値（3）よりも大きい特定確率（c）のユーザ満足度を取得する場合は，(3) の値を返す．特定確率（b）のユーザ満足度を取得する場合は，蓄積されている値のうち（b）よりも小さい特定確率である（2）の値を返す．

4.2 トップダウン探索アルゴリズム

トップダウン探索アルゴリズムでは，特定確率とサービス品質の関係を利用して，ユーザ満足度が高い組み合わせからトレードオフを解決している組み合わせを探索していく．一般的に，個人化サービスではユーザが開示した情報をもとにサービスを変更するため，個人情報を多く開示することでより個人化されたサービスが受けることができると思われる．そこで，特定確率が上昇するとサービス品質も上昇するという知識を利用したアルゴリズムを以下に示す．

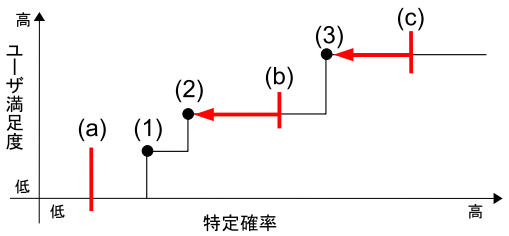

図 2 特定確率をキー値としたユーザ満足度の蓄積方法

Fig. 2 Method to store user's satisfacation for services by person identification probability

（1） 要求された特定確率を満たし，粒度が最も細かい組み合わせを抽出

（2） 抽出した中から開示する組み合わせを選択

（3） 抽出した組み合わせを全て選択，探索した場合，以下の（4）, (5）のどちらかを実行

（4） 探索済みの組み合わせの中でユーザ満足度が高いグループに属する組み合わせについて，属性情報の粒度を粗いものに変更する．粒度を変更した組み合わせも探索済みであった場合はほかの属性情報の粒度を粗くするか，ほかのユーザ満足度の高い組み合わせの粒度を変更する．全てのユーザ満足度の高い組み合わせに対して探索が終了した場合, (5）を実行する．

（5） 現時点でトレードオフを解決している組み合わせを選択する．ユーザ満足度が高いグループに属する組み合わせの中で最も特定確率が低い組み合わせをトレードオフを解決している組み合わせとする．

（6） 特定確率とユーザからのフィードバックを蓄積

特定確率が高い組み合わせの中にユーザ満足度が最も高い組み合わせが存在すると考えられるため，まず, (1）(2）にて粒度が最も細かい組み合わせを優先的に探索する．全ての最も粒度が細かい組み合わせのユーザ満足度を取得した後，ユーザ満足度が同程度で特定確率が低い組み合わせを探すために（3）を実行する．

（3）において（4）を選択する確率をあげることで，多くの組み合わせを選択，探索できるため，トレードオフを解決している組み合わせを早く見つけることができる．しかし，トレードオフを解決している可能性がある組み合わせの中には，ユーザ満足度が低い組み合わせも含まれるため，可能性がある組み合わせを全て探索するまではユーザ満足度が低くなってしまう問題がある．

また, (5）を選択する確率をあげることで，ユーザ満足度が高い組み合わせを優先的に選択することができる．しかし，トレードオフを解決している組み合わせではないため，特定確率は高くなり，またトレードオフを解決している組み合わせの発見が遅くなる問題がある．今回はユーザ満足度が高くなることを優先させるために, (5）が選択される確率を高くした．

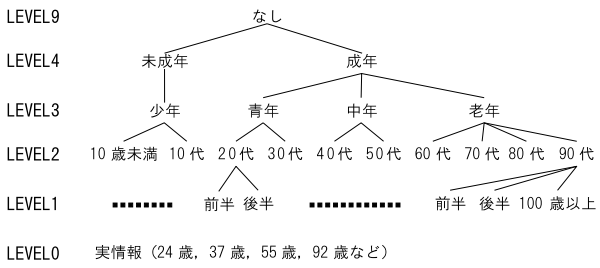

図 3　年齢情報の粒度
Fig. 3　Granularity of age

図 4　位置情報の ID
Fig. 4　Location ID

5.　シミュレーション評価

GrIP の有効性をシミュレーションにより評価した．

5.1　環　　境

ここでは，シミュレーションで用いたユーザ情報，評価指標，特定確率を推定するための統計情報，推薦方式について述べる．

5.1.1　ユ　ー　ザ

シミュレーションには，個人情報のうち，年齢情報，性別情報，職業情報，位置情報の 4 つの属性情報を利用した．「23 歳，男性，学生，大阪府大阪市北区梅田 1 丁目」という個人情報を持ち，0.01 以下の特定確率を要求する仮想的なユーザを用意した．また，動的な状況の変化への対応を実証するために，位置情報をサービスを受けるごとに変更させた環境を用意した．位置情報は東京都内と大阪市内の計 7294 地点からランダムに選択を行った．

それぞれの属性情報の粒度を以下のように定義した．

- 年齢情報

図 3 のような階層構造を設定し，実情報から情報なしまでの 6 段階に定義した．

- 性別情報

情報あり，情報なしの 2 段階に定義した．

- 職業情報

情報あり，情報なしの 2 段階に定義した．

- 位置情報

位置情報は図 4 のような 9 桁の ID を設定した．市区町村の百の位は，区群なら 1，市群なら 2，町村群なら 3 を設定した．粒度を粗くするごとに最下位の値を削除させ，実情報から情報なしまでの 10 段階に定義した．

ユーザ満足度は個人情報が全て開示された際に推薦されるコンテンツと，開示する個人情報の制御を行った際に推薦されるコンテンツとの F 値と定義する．F 値は式 5〜7 のように再現率と適合率の調和平均で表される．

$$\text{再現率} = \frac{\text{推薦された適合するコンテンツ}}{\text{全コンテンツのうち適合コンテンツ}} \quad (5)$$

$$\text{適合率} = \frac{\text{推薦された適合するコンテンツ}}{\text{全コンテンツのうち推薦されたコンテンツ}} \quad (6)$$

$$F\text{ 値} = \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}} \quad (7)$$

F 値を用いることで，情報制御を行った場合ユーザが望んでいるコンテンツが推薦結果に含まれていないことと，ユーザが望んでいないコンテンツが推薦結果に含まれていることに関して同時に評価することができる．

5.1.2　特 定 確 率

特定確率を求めるためには各属性情報ごとに該当人数を調べる必要がある．今回は平成 12 年の国勢調査のデータをもとに特定確率を求めた．位置情報は昼間人口のデータを利用した．

5.1.3　推薦サービス

10,000 個のコンテンツからユーザに仮想的に推薦を行う個人化サービスを作成した．特定の個人化サービスに依存しない制御を実証するために，個人化サービスの重要な要素であるコンテンツと推薦アルゴリズムをそれぞれ複数用意した．

コンテンツがターゲットとする属性情報の分布による違いを調べるために，複数のコンテンツの分布を用意した．ユーザ満足度を高くするためにはより細かい粒度の情報を開示しなくてはならないようにするために，属性情報ごとにひとつの値しか持たせていない．たとえば，「24 歳，男，職業不問，271270031」といった情報は持つが，「20 代，男，学生，27127****」のような情報は持たない．それぞれの属性情報の値を，シミュレーションで利用するユーザの値を平均値とした正規分布や昼間人口の分布などにもとづいて用意した．

推薦アルゴリズムはスコアリング方法の異なる以下の 4 種類を用意した．

（a）　完全一致属性推薦

コンテンツがターゲットとする属性情報とユーザの属性情報が一致した場合に加点する．

（b）　近傍属性推薦

ユーザの属性情報と一致した場合だけでなく，属性情報に近い場合にも加点する．近いほど高いスコアを加点する．

（c）　平均近傍属性推薦

近傍属性推薦と同様に加点し，平均スコアをもとに正規化．

（d）　協調フィルタリングによる推薦

協調フィルタリングの実験結果を公開している MovieLens [12] のデータをもとに，協調フィルタリングを用いた推薦を仮想的に実現した．ここでは，開示した個人情報に一致するほかのユーザが評価した上位のコンテンツを推薦する．また，複数の属性情報が一致するユーザがいた場合，一致した属性情報の数だけそのユーザの評価を重複させる．本研究では，MovieLens の協調フィルタリングを利用するため，コンテンツは MovieLens のもののみ用意した．

5.2　結果と考察

5.2.1　特定確率とユーザ満足度の相関関係

要求された特定確率を満たす粒度の組み合わせについて，各組み合わせの特定確率とユーザ満足度の分布を図 5 に示す．図において，(a)，(b)，(c)，(d) とはそれぞれ完全一致属性推薦，

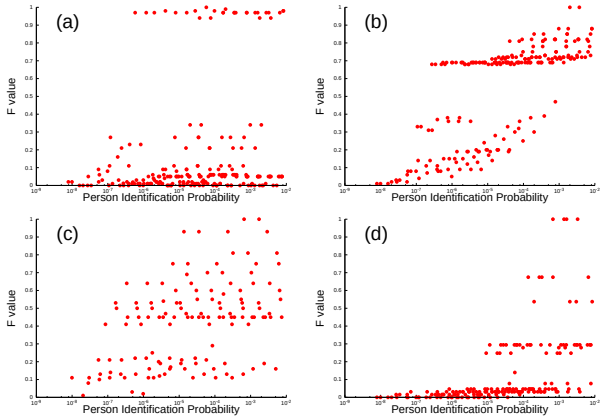

図 5　特定確率とユーザ満足度の分布

Fig. 5　Distribution of person identification probability and F value

表 1　ユーザ満足度 0.9 以上の粒度の組み合わせ

Table 1　Combinations of granularities (user's satisfacation ≧ 0.9)

年齢情報	性別情報	職業情報	位置情報	ユーザ満足度
LEVEL0	あり	あり	2712*****	1.0
LEVEL0	あり	あり	271******	1.0
LEVEL0	あり	あり	27*******	1.0
LEVEL0	あり	なし	2712700**	0.93
LEVEL0	あり	なし	271270***	0.93
LEVEL0	あり	なし	27127****	0.93
LEVEL0	あり	なし	2712*****	0.93
LEVEL0	あり	なし	271******	0.93
LEVEL0	あり	なし	27*******	0.93

近傍属性推薦，平均近傍属性推薦，協調フィルタリングによる推薦を指す．

図より，どの推薦サービスを利用しても特定確率が上昇するとユーザ満足度も上昇する傾向にあることがわかる．しかし，同程度のユーザ満足度でも特定確率が大きく異なる組み合わせが存在する．これは，ユーザ満足度が大きく変化する属性情報が存在するためである．(b) の推薦アルゴリズムでは，大きくふたつに分布がわかれている．この場合，職業情報を開示することでユーザ満足度が高くなる傾向にあるが，職業情報を開示しない場合はユーザ満足度が低くなる傾向にあるためである．

(c) の推薦アルゴリズムを用いた場合にユーザ満足度が 0.9 以上の組み合わせを表 1 に示す．表より，この場合ユーザ満足度を高くするためには年齢情報と性別情報を実情報の開示が必要であることがわかる．一方，職業情報や位置情報は，その上で特定確率を満たすことのできる粒度に変更すればユーザ満足度が高くなると考えられる．このように，細かい粒度で開示したほうがユーザ満足度が高くなる属性情報と，粒度を変更してもユーザ満足度に大きく影響を与えない属性情報がある．したがって，このようなユーザ満足度に大きく影響する属性情報とその粒度を発見することで，トレードオフを解決した粒度の組み合わせを選択することが可能となる．

5.2.2　静的環境におけるシミュレーション結果

属性情報の値を固定して平均近傍属性推薦を利用した場合

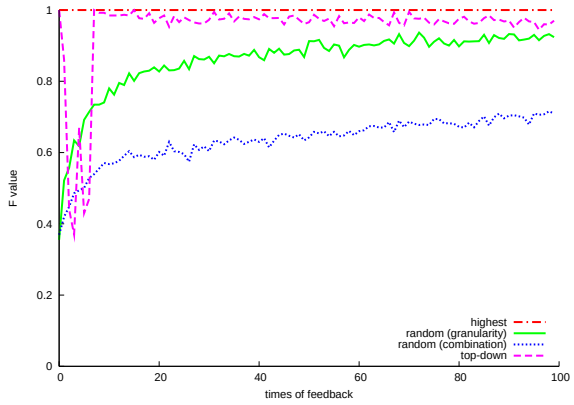

図 6　静的環境における試行回数とユーザ満足度

Fig. 6　Times of feedback and user's satisfacation at static environment

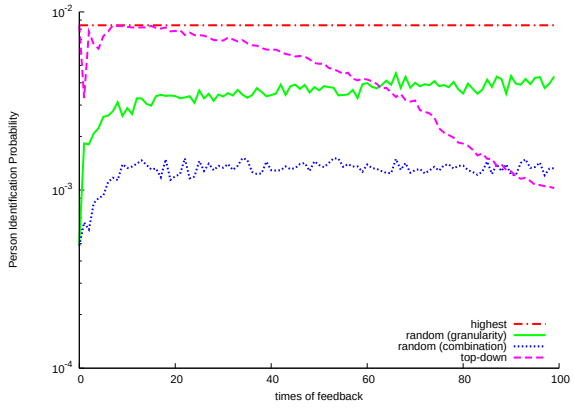

図 7　静的環境における試行回数と特定確率

Fig. 7　Times of feedback and person identification probability at static environment

の，試行回数とユーザ満足度のグラフを図 6 に，試行回数と特定確率のグラフを図 7 に示す．100 回実行し，その平均をグラフにしている．図において random（granularity）, random（combination）, top-down, highest とは 4 章で述べたランダム探索アルゴリズムの属性ごとの蓄積，ランダム探索アルゴリズムの組み合わせごとの蓄積，トップダウン探索アルゴリズム，ユーザが要求した特定確率を満たす組み合わせのうち最も特定確率が高い組み合わせを指す．

トップダウン探索アルゴリズムでは，探索が進むにつれてユーザ満足度が高いまま，特定確率が低くなっていくことがわかる．探索初期は粒度が最も細かい組み合わせを選択する必要があるため，特定確率が高くなっている．探索が進むにつれ，ユーザ満足度に大きく影響しない属性情報を発見し，ユーザ満足度が高いまま特定確率を低く抑えている．一方，ランダム探索アルゴリズムでは，探索が進むにつれユーザ満足度が高い組み合わせを選択している．特に，組み合わせごとの蓄積に比べて属性ごとの蓄積のほうがユーザ満足度が高くなっている．これは，属性ごとの蓄積では一度に複数の組み合わせについて蓄積できるためであると考えられる．どちらのアルゴリズムを用いた場合でも，探索を進めることで最も特定確率が高い組み合わせよりもユーザ満足度が高く，かつ，特定確率が低いトレードオフを解決している組み合わせを選択できることがわかる．

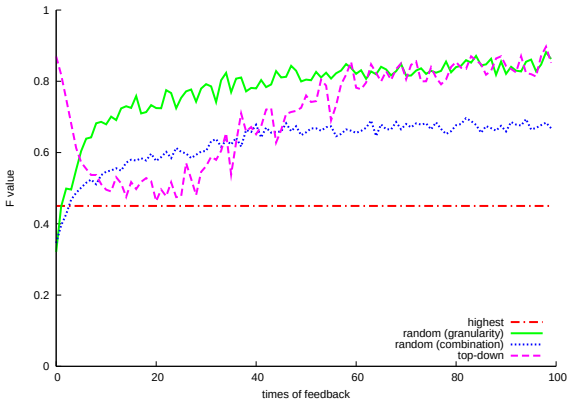

図 8　動的環境における試行回数とユーザ満足度

Fig. 8　Times of feedback and user's satisfacation at dynamic environment

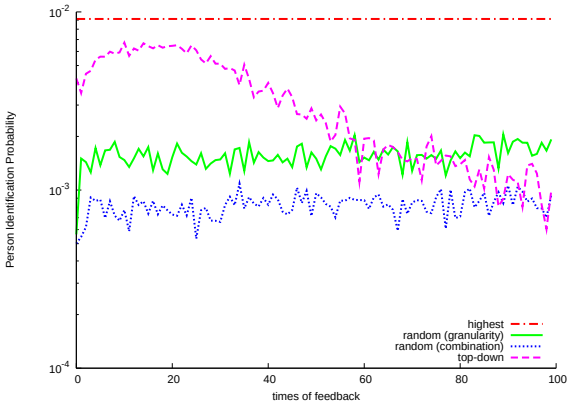

図 9　動的環境における試行回数と特定確率

Fig. 9　Times of feedback and person identification probability at dynamic environment

5.2.3　動的環境におけるシミュレーション結果

位置情報がサービスを受けるごとに変化する動的環境において，平均近傍属性推薦を利用した場合の探索回数とユーザ満足度のグラフを図 8 に，探索回数と特定確率のグラフを図 9 に示す．100 回実行しその平均をグラフにしている．

ランダム探索アルゴリズムは，静的環境とほぼ同様のグラフになっている．これは，位置情報が変化してもユーザ満足度が高くなる組み合わせを発見できているためである．特に属性ごとの蓄積では，どの属性がユーザ満足度に高く影響するかを反映した蓄積ができているため，影響度の低い属性の粒度が変更される傾向にあるためである．一方，トップダウン探索アルゴリズムは，位置情報が動的に変化するため最も粒度が細かい組み合わせを探索する時間が長くなり，探索初期の段階ではユーザ満足度が低く，特定確率が高くなっている．しかし，探索が進むにつれユーザ満足度が高くなり，特定確率は低くなっている．これは，どの属性情報がユーザ満足度に高く影響するかを考慮して探索できているため，異なる属性情報の値に変更されてもトレードオフを解決している組み合わせを選択できている．

これらの結果より，探索初期においてはランダム探索アルゴリズムが有効であり，試行回数が増えた場合，トップダウン探索アルゴリズムが有効であることがわかった．したがって，サービスを受ける回数が比較的多い PUSH 型サービスにおいては，ランダム探索アルゴリズムを，サービスを受ける回数が比較的少ない PULL 型サービスにおいては，トップダウン探索アルゴリズムを用いると有効であると考えられる．

6.　ま　と　め

本研究では，プライバシ保護とサービス品質のトレードオフを考慮した個人情報制御機構 GrIP を提案した．GrIP の有効性を示すために複数の推薦アルゴリズムを用いてシミュレーションを行った．その結果，GrIP を用いることでユーザが要求したプライバシを保護した上で高いユーザ満足度のサービスを受けることができることを示した．

今後の課題として，実環境での提案機構の有効性の検証があげられる．今回はユーザ満足度を計算して求めたため特定確率が上昇するとユーザ満足度も上昇したが，実環境では異なる傾向があらわれることもありうる．さらに，利用できる属性情報は年齢情報や位置情報などに限定しているが，嗜好情報や履歴情報も考慮した実験を行う必要がある．

謝　　辞

本研究の一部は，平成 15 年度総務省「ユビキタスネットワーク認証・エージェント技術の研究開発」の研究助成によるものである．

文　　献

[1] 宮本崇弘, 山田和広, 竹内亨, 奥田剛, 春本要, 下條真司：“情報源の異種性を隠蔽し動的に粒度調整可能なユーザプロファイル生成機構”, マルチメディア，分散，協調とモバイル (DICOMO2004) シンポジウム, pp. 429–432 (2004).

[2] “Platform for privacy preferences project”. available at `http://www.w3.org/P3P/`.

[3] “A P3P preference exchange language”. available at `http://www.w3.org/TR/P3P-preferences/`.

[4] E. P. I. Center: “Pretty poor privacy: An assessment of P3P and internet privacy” (2000). available at `http://www.epic.org/reports/prettypoorprivacy.html`.

[5] 今田美幸, 高杉耕一, 太田昌克：“情報エントロピーを用いたプライバシ保護手法: LooM”, 電子情報通信学会技術研究報告 ISEC 2004-54, pp. 87–94 (2004).

[6] L. Sweeney: “k-anonymity: a model for protecting privacy”, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, **10**, 5, pp. 557–570 (2002).

[7] S. Tamaru, J. Nakazawa, K. Takashio and H. Tokuda: “PPNP: A privacy profile negotiation protocol for services in public spaces”, Proceedings of Fifth International Conference on Ubiquitous Computing (UbiComp2003) (2003).

[8] 藤原香織, 正寺朋子, 中村暢芳, 山邉哲生, 中島達夫：“ユーザプレファランスによる適応制御機構を備えたコンテクストアウェアシステムの提案”, 情報処理学会研究報告 2003-UBI-2, pp. 177–182 (2003).

[9] 金子雄, 福村真哉, 春本要, 下條真司, 西尾章治郎：“モバイル環境における端末の位置情報に基づく P2P ネットワークの提案と評価”, 電子情報通信学会データ工学ワークショップ (DEWS 2004) 論文集 (2004).

[10] 岩井将行, 高橋元, 門田昌哉, 中島達夫, 徳田英幸：“Tachyon:プライバシを考慮する電子タグ位置情報管理機構”, 情報処理学会全国大会論文集, pp. 49–53 (2004).

[11] R. S. Sutton and A. G. Barto: “Reinforcement Learning: An Introduction”, The MIT Press (1998).

[12] “Movie Lens”. available at `http://www.grouplens.org`.