GrIP: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

GrIP: A Profile Control Mechanism for User Privacy Protection and Quality of Personalization Services

宮本 崇弘* 竹内 享†
奥田 剛‡ 春本 要§
有吉 勇介** 下條 真司††

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

情報推薦などの個人化サービスを受けるためにユーザ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．

In order to enjoy the benefit of personalization services such as recommendation services, it is necessary for users 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.

1. はじめに

近年，商品推薦サービスのように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．サービスの個人化を行う際には，ユーザは年齢や性別などの個人情報を個人化サービスに開示する必要がある．今後ユビキタス環境が整うにつれ，様々な種類の個人情報が異なるサービスに分散して保存されていると考えられる．したがって，ユーザはそのような個人情報を収集して，ユーザプロファイルを生成し，個人化サービスに開示する必要がある．このような様々な個人情報を個人化サービスが透過的に扱えるようにするために，われわれは個人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，個人化サービス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．ユーザは指定された個人情報の含むユーザプロファイルを生成すればよい．しかし，SDL-UP はユーザのプライバシに対する要求を反映することができない．

一般的に開示する個人情報が多いほどより個人に適したサービスを受けることができると考えられるが，プライバシの観点からユーザには多くの個人情報を開示したくないという要求がある．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) [2]がある．PPNP では，開示する情報の種類やその粒度を変更することで開示する個人情報を制限しプライバシを保護しているが，それによってユーザが受けることができるサービスの品質が低下する可能性がある．このように，ユーザプロファイルを生成するためには，プライバシと個人化サービスの品質のトレードオフを考慮する必要がある．

そこで本論文では，プライバシの保護とサービス品質のトレードオフの解決を目的とした，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．

2. 個人情報制御機構

1 章で述べたように，ユーザプロファイルを生成する機構にはプライバシとサービス品質のトレードオフを解決する仕組みが必要であると考えられる．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．また，ここでのプライバシとは，3 章で述べる特定確率を指すものとする．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．

- 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を事前に知ることはできない．したがって，ユーザの要求するプライバシを保護しつつ，個人情報の種類や粒度による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やユーザプロファイルの生成などを行うユーザエージェントの一機能として提供される．事前にユーザは要求する特定

* 正会員 大阪大学大学院情報科学研究科
miyamoto@ais.cmc.osaka-u.ac.jp
† 大阪大学大学院情報科学研究科博士後期課程
stakeuti@ist.osaka-u.ac.jp
‡ 奈良先端科学技術大学院大学情報科学研究科 okuda@is.naist.jp
§ 大阪大学大学院工学研究科 harumoto@eng.osaka-u.ac.jp
** 尾道大学経済情報学部 y-ariyoshi@onomichi-u.ac.jp
†† 大阪大学サイバーメディアセンター
shimojo@cmc.osaka-u.ac.jp

確率を指定する．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個人情報の種類やその粒度の組み合わせを探索し，開示する個人情報を制御する．制御された個人情報を開示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．

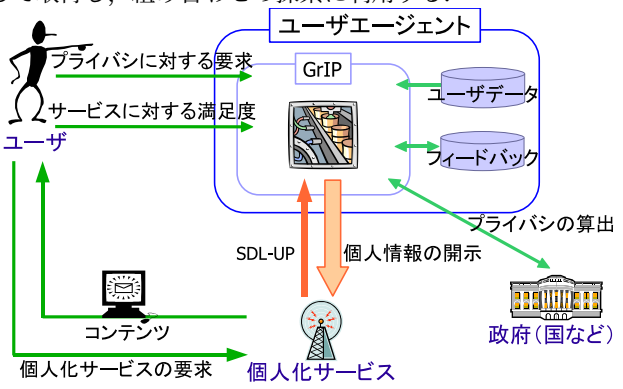

図 1 GrIP の概要
Fig.1 Overview of GrIP

3. 特定確率

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．開示された属性情報の値と同じ値を持つ人数を該当人数と呼ぶ．一般的に該当人数が少ない場合，該当人数が 1 人増えることで開示リスクは大きく減少するが，該当人数が多い場合は該当人数が 1 人増えても開示リスクはあまり変化しない．これらを考慮するために特定確率は該当人数の逆数で算出される．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測する．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Pは，以下の式を用いて推測する．

$$P = Q(a_1) \times \sum_{i=2}^{n} \frac{Q(a_i)}{R(a_i)} \qquad (n \geq 2)$$

$R(a_i)$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し，$Q(a_i)$は$R(a_i)$のうち，ユーザと同じ値を持つ人数を指す．たとえば，国勢調査を用いて「男性」という属性情報に対する該当人数を求める場合，R(性別)は日本の人口を，Q(性別)は日本の男性の人数を指す．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．

特定確率を用いることで，開示する属性情報の種類や数が異なる場合でも，同じ基準で開示リスクを比較することができる．したがって，ユーザは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．

4. 粒度調整アルゴリズム

属性情報はそれぞれ異なる粒度を持つため，ある二つの属性情報を開示する場合でも，プライバシを保護したうえで開示できる粒度の組み合わせは複数存在する．2 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，サービス品質は開示情報の粒度やその組み合わせ，利用する個人化サービスによって変化すると考えられる．そこで本研究では，それぞれの粒度の組み合わせでサービスを利用した場合のサービス品質を調べ，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，確率的に粒度の組み合わせを探索していく．2 節で述べたように，トレードオフを解決した組み合わせはサービス品質が高く，特定確率が低くなると考えられる．そこで，特定確率とサービス品質のトレードオフを以下の式で定義する適格度を用いて表現する．

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

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．確率的に組み合わせを探索していくために，適格度が大きいほど選ばれる確率が高くなるソフトマックス行動選択規則[3]を利用する．組み合わせ cが探索される確率は以下の式から求められる．

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

$Q(c)$は組み合わせ c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．

以下にアルゴリズムを示す．

(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フィードバックされるサービス品質を蓄積

このアルゴリズムでは，ユーザからのフィードバックされるサービス品質は属性情報の粒度ごとに蓄積する．属性情報の粒度ごとに蓄積することで，サービス品質に大きく影響する属性情報の粒度を見つけることができる．また，一度のフィードバックに対して複数の組み合わせに対して蓄積が行えるため，少ない探索回数でトレードオフを解決した組み合わせを発見できると考えられる．

4.2 トップダウン探索アルゴリズム

一般的に，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．そこでトップダウン探索アルゴリズムでは，そのような特定確率とサービス品質の関係を利用して，開示する個人情報の粒度が細かい組み合わせから探索を行う．

以下に，アルゴリズムを示す．

(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選別
(2) 選別した中から開示する組み合わせを選択
(3) 選別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サービス品質が高いグループに属するある組み合わせを選択し，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，ほかのサービス品質の高い組み合わせの粒度を変更する．全てのサービス品質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．サービス品質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フィードバックされるサービス品質を蓄積

特定確率が高い組み合わせの中にサービス品質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探索する．全ての最も粒度が細かい組み合わせのサービス品質を取得した後，サービス品質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．(4)を実行することで，未知の組み合わせのサービス品質を取得できるため，トレードオフを解決している組み合わせを探索することができる．しかし，サービス品質が低い組み合わせが選択される可能性がある．また，(5)を実行することで，サービス品質の高い組み合わせが選択される．しかし，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．今回はサービス品質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．

5. シミュレーション評価

GrIP の有効性を示すためプロトタイプを実装し，シミュレーションにより評価を行った．

5.1 環境

シミュレーションでは個人情報として，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」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．位置情報はサービスを受けるごとに東京都内と大阪市内の計 7294 地点からランダムに選択した．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 2 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：“PPCCCSSSN”という 9 桁の ID を設定した．“PP”は都道府県の ID を，“CCC”は市区町村の ID を，“SSS”は町名の ID を，“N”は丁目をそれぞれ指す．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．

特定確率は平成 12 年の国勢調査のデータをもとに求めた．位置情報の該当人数は昼間人口のデータを利用した．個人化サービスとして 4 種類のコンテンツ推薦サービスを用意した．また，サービス品質は個人情報が全て開示された際に推薦するコンテンツと，開示する個人情報の制御を行った際に推薦するコンテンツとの F 値と定義する．

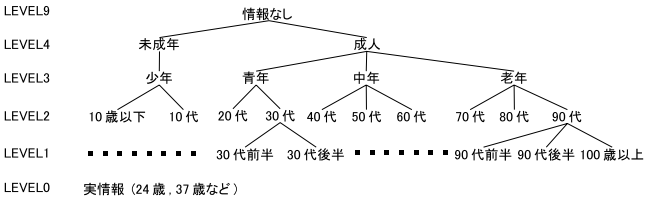

図 2 年齢情報の粒度
Fig.2 Granularity of Age

5.2 結果と考察

5.2.1 特定確率とサービス品質の相関関係

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サービス品質の分布を図 3 に示す．それぞれの図は異なる推薦サービスにおける分布を示す．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サービス品質も上昇する傾向にあることがわかる．しかし，同程度のサービス品質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，サービス品質が大きく変化する属性情報が存在するためである．(b)の推薦サービスでは，大きくふたつに分布がわかれている．このサービスの場合，職業情報を開示することでサービス品質が高くなる傾向にあるが，職業情報を開示しない場合はサービス品質が低くなる傾向にあるためである．

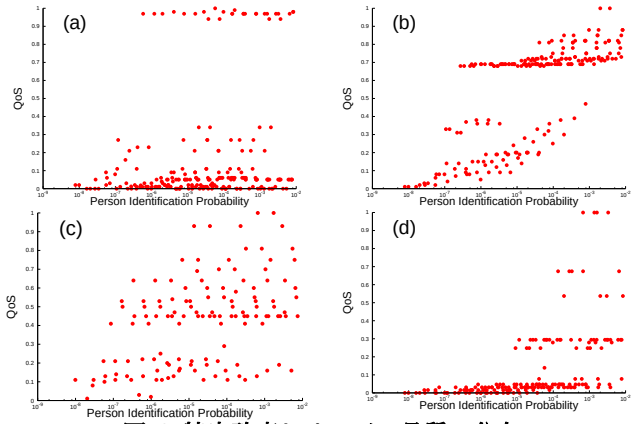

図 3 特定確率とサービス品質の分布
Fig.3 Distribution of PIP and QoS

(c)の推薦サービスを用いた場合にサービス品質が 0.9 以上の組み合わせを表 1 に示す．表より，この場合サービス品質を高くするためには年齢情報と性別情報を詳細に開示する必要があることがわかる．一方，位置情報は詳細な情報を必要しないと考えられる．このように，細かい粒度で開示したほうがサービス品質が高くなる属性情報と，粒度を変更してもサービス品質に大きく影響を与えない属性情報がある．したがって，このようなサービス品質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．

表 1 サービス品質 0.9 以上の粒度の組み合わせ
Table 1. Combinations of Granularities (QoS≧0.9)

年齢情報	性別情報	職業情報	位置情報	QoS
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

5.2.2 シミュレーション結果

試行回数とサービス品質のグラフを図 4 に, 試行回数と特定確率のグラフを図 5 に示す．100 回実行し，その平均をグラフにしている．図における highest PIP とは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．最も特定確率が高い組み合わせのサービス品質と特定確率は，各試行の平均で求めている．

図より，最も特定確率が高い組み合わせのサービス品質が低いことがわかる．これは，多くの個人情報を開示することでサービス品質が高くなるとは限らないことを示している．ランダム探索アルゴリズムでは，探索が進むにつれサービス品質が高くなっていることがわかる．このアルゴリズムでは，どの属性がサービス品質に高く影響するかを反映した蓄積を行っているため，影響度の低い属性の粒度が変更され，影響度の高い属性の粒度は細かく開示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サービス品質が低く，特定確率が高くなっている．しかし，探索が進むにつれランダム探索アルゴリズムより高いサービス品質で低い特定確率の組み合わせを選択している．

これらの結果より，提案するアルゴリズムを用いることで，最も特定確率が高い組み合わせよりもサービス品質が高く，かつ，特定確率が低い組み合わせを探索できることがわかる．

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サービス品質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．今回はサービス品質を計算して求めたため特定確率が上昇するとサービス品質も上昇したが，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，嗜好情報や履歴情報も考慮した実験を行う必要がある．

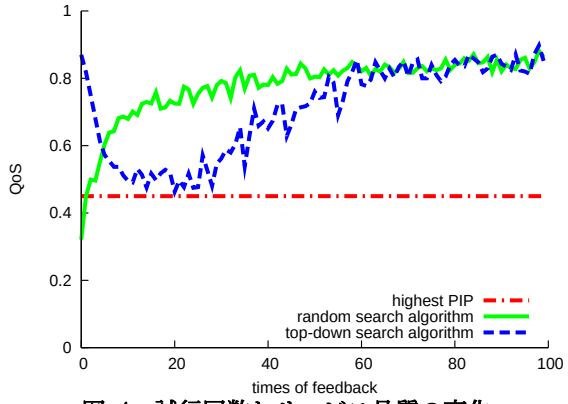

図 4 試行回数とサービス品質の変化
Fig.4 Times of Feedback and QoS

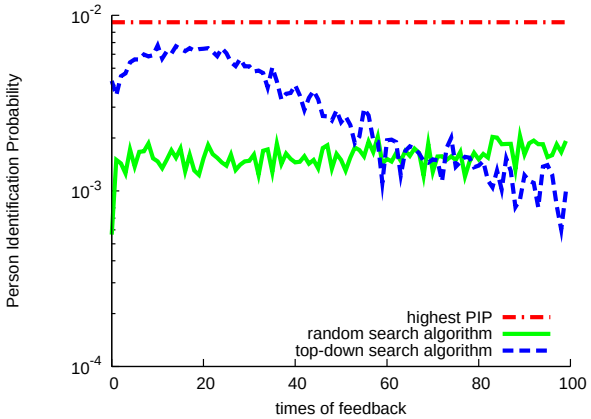

図 5 試行回数と特定確率の変化
Fig.5 Times of Feedback and PIP

[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，山田和弘，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] 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 (UbiComp2003), (2003).

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

宮本 崇弘 Takahiro MIYAMOTO
2005 大阪大学大学院情報科学研究科博士前期課程修了． 日本データベース学会会員．

竹内 亨 Susumu TAKEUCHI
大阪大学大学院情報科学研究科博士後期課程在学中．2003 大阪大学大学院基礎工学研究科博士前期課程修了．コミュニケーション支援システムの研究に従事．IEEE，情報処理学会各学生会員．

奥田 剛 Takeshi OKUDA
奈良先端科学技術大学院大学情報科学研究科助手．1998 大阪大学大学院基礎工学研究科博士前期課程修了．マルチメディア通信システムなどの研究に従事．IEEE 会員．

春本 要 Kaname HARUMOTO
大阪大学大学院工学研究科助教授． 1994 大阪大学大学院基礎工学研究科博士前期課程修了．工学博士．データベースシステム，マルチメディア情報システムなどの研究に従事．IEEE，情報処理学会，電子情報通信学会各会員．

有吉 勇介 Yusuke ARIYOSHI
尾道大学経済情報学部助教授．1995 大阪大学大学院基礎工学研究科博士後期課程単位取得満期退学．1995 日本電気株式会社入社．2003 から現職．工学博士．情報推薦，情報流通の研究に従事．情報処理学会，電子情報通信学会各会員．

下條 真司 Shimojo SHINJI
大阪大学サイバーメディアセンター教授．1986 大阪大学大学院基礎工学研究科博士後期課程修了．工学博士．LAN のアクセス方式の性能評価，分散処理システムの性能評価，分散型オペレーションシステムの研究に従事．IEEE，ACM，情報処理学会各会員．