社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2005－ＩＣＳ－140（３）
2005／8／1

プライバシー保護実現に向けた秘匿性定量化手法: LooM

今田 美幸　太田 昌克　山口 正泰

NTT 未来ねっと研究所 〒180-8585 東京都武蔵野市緑町 3-9-11
E-mail: imada@ma.onlab.ntt.co.jp, {ohta.masakatsu, yamaguch.masayasu}@lab.ntt.co.jp

あらまし ユーザの代理としてのプライバシー情報開示交渉機能実現に向けて，ユーザ集合の特性に応じてプライバシーの保護レベルを評価する秘匿性定量化手法 LooM[1]を提案する．LooM では，ユーザ集合から特定個人が絞り込まれる危険性が，開示するプライバシー情報種別（属性）に依存するという考えの下に，「プライバシー保護の問題」を「秘匿したい属性の値によってユーザ集合をクラス分類する問題」として扱う．本稿では，LooM について，ユーザのプライバシー情報データベースの規模と属性値分布の依存性の観点から評価し，有効性を示す．LooM を用いた実用システムでは，秘匿性閾値を規定し，秘匿性閾値に応じてプライバシー情報の開示の可否判断を行う．ここでは，ユーザアンケート調査に基づいた秘匿性閾値設定方法について述べる．

キーワード プライバシー保護，決定木，秘匿性，均衡モデル

LooM: An Anonymity Quantification Method for Privacy Protection

Miyuki IMADA　Masakatsu OHTA　Masayasu YAMAGUCHI

NTT Network Innovation Labs.　Midori-cho 3-9-11, Musashino-shi, Tokyo, 180-8585 Japan
E-mail: imada@ma.onlab.ntt.co.jp, {ohta.masakatsu, yamaguch.masayasu}@lab.ntt.co.jp

Abstract We propose a novel anonymity quantification method which calls LooM. The LooM is a method that evaluates the privacy protection levels every property of user collection for realizing a disclosure negotiation function of private information for user agents. Its main feature is that it can quantitatively control anonymity by a single value (disclosure threshold value) using a classification algorithm of the decision tree. In this paper, we show that the LooM is hardly affected by size of privacy information database or the attribute values distribution of users' private information by using artificial database. In order to decide the disclosure threshold value on practical systems, we show a method for setting the value based on web questionnaire data.

Keyword Privacy Protection, Decision Tree, Anonymity, Equilibrium Model

1. はじめに

インターネット上では，ユーザがプライバシー情報を提供する代わりに，サービスプロバイダ（以下プロバイダと略す）からユーザ嗜好にあったサービスを受けるというサービス利用形態が徐々に定着しつつある．例えば，氏名，年齢，住所，等の情報を入力し，会員登録を行うことで，割安な旅行プランを提示するようなサービスが提供されている．しかし，プライバシー情報を提供することにより，サービスの質や利便性が高まる反面，悪質なプロバイダによりプライバシーが侵害されるというリスクを伴うことになる．

ここでいう「プライバシーの侵害」とは，以下の通りである．まず，「ユーザ」とは，ユーザ本人，プライバシー情報，ユーザ ID（ネットワーク上での識別子）の３つの要素で表現できる．これらが独立な情報としてプロバイダに知られても，ユーザに危害が及ぶことは少ないが，相互に関連付けられた場合，人はプライバシーを侵害されたと感じると考えられる．すなわち，実世界の個人が特定され，かつその個人に関する情報が明らかになることが「プライバシーの侵害」に当たると考えられる．ただし，ネットワーク上でユーザ情報を管理する際には，ユーザ ID とプライバシー情報の２つだけが関連付けられた状態で漏洩した場合でも，個人情報保護法に違反する可能性があるため，本稿ではこのような場合も「プライバシーの侵害」に含めることとする．

筆者らは，このようなプライバシーの侵害を回避することを目的として，プライバシー情報を開示する際に，開示する情報量と開示により低下する秘匿性の度合いを定量化するプライバシー保護手法 LooM を提案

[1] LooM：Loosely managed privacy protection Method [名] [a ～] ほんやりと現われること．（研究社英和辞典より）

し，基本機能を実装した[13]．本稿では，LooMを実システムに適用する場合を想定して，データベース規模や属性値分布に対する依存性について評価するとともに，情報の開示判断の基準となる秘匿性閾値の設定方法について述べる．

以下，2章では本稿の前提条件について述べ，3章では，プライバシー情報開示の際に，ユーザ個人やユーザ属性の秘匿性を定量的に評価する秘匿性定量化手法LooMとその適正評価について述べ，4章ではLooMを実用システムに適用する際の秘匿性閾値の決定方法について述べる．5章では関連研究について述べ，最後にまとめを述べる．

2. 前提条件

2.1. プライバシー情報の定義

プライバシー情報には，ユーザが端末から入力する静的情報と，ユーザの挙動や振る舞いをシステムが検知することで得られる動的情報がある．ここでは，前者を静的プライバシー情報(SP : Static privacy)，後者を動的プライバシー情報(DP : Dynamic privacy)と定義する．SPには，名前，住所，年齢，性別，病歴，家族構成，ユーザの嗜好，意図，およびこれらの履歴などが含まれる．DPには，位置，生体情報，温度，人検知などセンサによって取得された情報，電子マネーによる購入記録，乗車記録などのユーザの行動に付随して取得される情報，およびこれらの履歴などが含まれる．ただし，システムがユーザを一意に識別するためのユーザIDは，ユーザが端末から入力した情報ではないので，プライバシー情報に含めない．一旦取得したSPとDPから協調フィルタリングなどによる分析によって，不明なSPやDPを推測することも可能であるが，推測結果の信憑性は分析方法に依存する．よって本稿で扱うDP，SPは，ユーザの端末やセンサから直接取得した情報およびその履歴のみとする．

2.2. 前提とするシステム

本稿で前提とするシステムは，複数の端末，センサ，ネットワーク，サーバから成る．ネットワークには，インターネットの他に，ホームネットワーク，センサネットワーク，アドホックネットワークといったような様々な種類のネットワークがあり，それらはゲートウェイなどを介して相互接続されている（図1参照）．ネットワーク上には，幾つかのサーバが配置してあり，サーバは，ロケーション管理，コンテキスト管理，センス情報管理，認証，攻撃防御のための監視，攻撃者の追跡・特定などを行っている．固定端末や携帯端末は，ネットワークに接続されており，ユーザやプロバイダは，端末経由でサービスを利用提供する．ユーザが保持している携帯端末には，ユーザ自身が入力したSPが格納されている．センサやGPSにより取得したDPは，ネットワーク上の安全なサーバで一元管理する．

次にサービス例として，携帯端末を所持したユーザが，家から町のレストランに移動する場合について示す．ユーザの移動に伴い，携帯端末は当初接続されていたネットワークから切り離され，レストラン到着時に再度ネットワークに接続される．ユーザは，レストランに入り，ディスプレイ付きテーブルに座る．この様子をGPSやセンサが検出し，位置情報やセンス情報としてサーバに送信する．サーバは，本システムに接続しているコンテンツ提供サービスプロバイダに対し，これらの情報を定期的に通知する．プロバイダはレストランにいるユーザの携帯端末から嗜好情報を収集し，ユーザの好みのコンテンツを各ディスプレイ付きテーブルに送信する．このようにして，レストランにいるユーザは，食事が出てくるまでの間，好みのコンテンツを見て過ごすことができる．

上記のようなシステムでのプライバシー保護は，既にいくつか研究されている[1][2][3][4]．ユーザとプロバイダ間で行うプライバシー情報開示交渉の多くは，先ずプロバイダが開示して欲しいプライバシー情報種別をユーザに提示し，ユーザが開示の可否判断を行った後，自身のプライバシー情報をプロバイダに渡すという方式で行っている．しかし，上記システムにおけるセンサは，ユーザの行動に関するDPをユーザの許可なく一方的に採取し，利用している．またユーザは，商品購入時のユーザ登録やアンケート回答により，同じプライバシー情報をプロバイダ毎に何度も登録しなければならない．さらに，ユーザにとっては，どのプライバシー情報をどのプロバイダに登録したかという，登録履歴の把握が難しい．また，提供したプライバシー情報の管理は，基本的にプロバイダに一任されていることが多いため，収集したプライバシー情報を厳重に管理していないプロバイダから情報漏洩する危険性

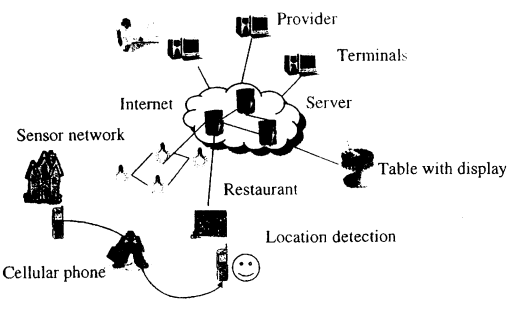

図1　前提とするシステム

がある．

一方，プロバイダは，収集したプライバシー情報を商品の品揃えやキャンペーン展開などに役立てるために活用する．しかし，収集したプライバシー情報は，センサの数や情報収集間隔，蓄積年数に応じて膨大な量になるため，プライバシー情報の維持管理が煩雑になる．また，時々刻々変わる DP の開示可否判断をユーザ個別に実行することは，事実上困難である．

このような問題を解決するためには，ユーザ個人に代わってプライバシー情報の開示交渉を行う機能が必要となる．本機能の実現のために，プライバシーの保護レベル（秘匿性）を客観的に評価する手法を確立する必要がある．

3. 秘匿性定量化手法 LooM

3.1. 秘匿性の確保とクラス分類問題

属性と属性値のペアとして管理するプライバシー情報データベース（PDB）において，プライバシー情報開示における安全性確保の問題は，ユーザのクラス分類問題ととらえることができる．ユーザ ID の特定のし易さは，どのプライバシー情報の属性をどのような順序で開示するかに依存する．例えば，女性が少ないユーザ集合を考えた場合，女性ユーザ Alice が性別属性を開示すると，Alice のユーザ ID の特定が容易になる．よって，Alice のユーザ ID を秘匿するためには，性別属性をプロバイダに対して隠蔽することが必要である．また Alice が大学生で，PDB が Alice と同じ大学にいる大学生により構成されているならば，大学生という属性を開示しても Alice のユーザ ID の秘匿性に影響はない．このように，開示する属性とユーザ ID との属性間の距離を判別する方法としては，決定木を用いた分類学習[5]と同じ考え方が適用できる．決定木を作る際は，事例集合を属性で効率的に分類するために，情報利得(information gain)を計算し，情報利得が大きくなる属性を分割属性として優先的に選択し，これをノードとして決定木を作成していく．

1 章でも述べたように，ユーザがプロバイダに対して秘匿したいのは，ユーザ ID とプライバシー情報の関連付けである．この関連付け防止に決定木学習アルゴリズムを利用するには，ユーザ ID を特定しにくいユーザ属性を優先的に開示すればよい．つまり，PDB の中でユーザ ID を特定し易い属性を秘匿属性とし，この秘匿属性に対して決定木を効率悪く作ることができれば，秘匿属性を特定することが困難になるため，ユーザ ID との関連付けも困難になる．具体的には，決定木のノードに情報利得の少ない属性を用いて効率の悪いクラス分類を行うことで，プロバイダはユーザが秘匿属性の特定に多くの労力を費やすことになるため，結果としてプライバシー保護が実現できる．LooM では，秘匿属性に対して，プロバイダから開示要求された属性の情報利得を計算し，情報利得が少ない属性であれば，開示を許可する．

どの属性を秘匿属性とするかは，ユーザの意思に依存するので，ユーザ自身が設定する．ユーザ ID を特定されたくなければ，秘匿属性は，名前のようなユーザ ID との 1 対 1 の関連がつけ易い属性とする．既婚か未婚かのようなユーザが個人的に明らかにしたくない属性の場合，秘匿属性は，ユーザ ID と 1 対 1 に対応しない属性とする．LooM は，秘匿属性が前述のいずれであっても，同じ手法で秘匿性の計算ができる．

3.2. PDB 依存の少ない秘匿性尺度

2.2 節で述べたような人の移動などによって PDB 規模や属性値分布が時間とともに変動する環境において，エントロピーをこれらに依存しない量として扱うために，LooM では，エントロピーを正規化した値を使う．正規化することは，異なる PDB 規模のエントロピー値を同一オーダーで扱うことができる．我々が採用した正規化手法は，以下の通りである．

秘匿したいユーザと同じ属性値をもつユーザを正，そうでない属性値をもつユーザを負の 2 つのクラスに別ける．ユーザ集合において，該当する属性値が正である確率を P_+，負である確率を P_- とした場合，秘匿したいユーザの正負に関するエントロピー H_0 は，式（1）のようになる．

$$H_0 = -P_+ \log_2 P_+ - P_- \log_2 P_- \quad (1)$$

ある属性 F の属性値の集合を $\{v_1, v_2, v_3, v_4, \cdots, v_n\}$ とした場合，属性 F が v_i である確率を P_i，F が v_i である集合において正である確率を P_{i+}，負である確率を P_{i-} とした場合の属性 F のエントロピー H_F は，式（2）のようになる．

$$H_F = \sum_{i=1}^{n} P_i (-P_{i+} \log_2 P_{i+} - P_{i-} \log_2 P_{i-}) \quad (2)$$

F を知ることによって得られる情報利得は，$H_0 - H_F$ となるが，これを H_0 に対して正規化すると，正規化した情報利得 G_N は，式（3）のようになる（H_0, H_F, D_F の関係は，図 2 参照）．

$$G_N = 1 - \frac{H_F}{H_0} = 1 - D_F \quad (3)$$

式（3）より，D_F が大きい値だと G_N が小さい値になるめ，効率の悪い分類が実現できる．例えば属性の開示をスケジューリングする際，D_F の値が大きい属性から順次開示すれば，秘匿性確保が持続できるので，より多くの属性を開示できる．LooM では，秘匿性を定量的に評価する量として D_F を用いる．

秘匿属性

User ID	Gender	Occupation	…	Location of8:00a.m.	name
12	F	Student	…	Park	Alice
345	M	Office worker	…	Cafe	Bob
6	M	Office worker	…	Station	John
7890	M	Student	…	Park	Bob

H_F H_0

図 2　D_F の計算方法

3.3. 適正評価

本節では，3.2 節で提案した D_F が秘匿性評価尺度として適正であるか否かについて評価する．ある時刻 T に PDB へ登録したユーザのプライバシー情報は，模擬的に以下の 10 種類とした．SP の属性と属性値は，「氏名：Alice, Bob, ...」，「性別：男女」，「血液型：A, B, O, AB」，「職業：会社員，学生」，学生の場合「学校種別：小学校，中学校，高等学校」，DP の属性と属性値は，「空腹：yes, no」，「経由した駅名：新宿，池袋，上野，東京の各駅」，「9 時から 12 時の各時間帯にいた場所の履歴：学校，会社，公園，ショッピング街」とした．秘匿属性とその属性値は，（氏名，Alice）とし，該当する人は，PDB 中に 1 人しかいないとした．評価は，DB の規模依存性と，属性値分布依存性の観点から行った．

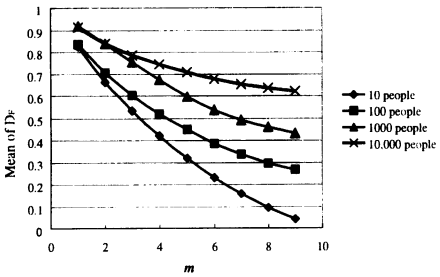

図 3　規模依存性

図 3 に，規模依存性に関する D_F の評価結果を示す．

横軸は開示属性数，縦軸は開示属性数ごとに計算した D_F の平均値（属性の組合せパターンすべてについて計算）である．属性値分布は一様分布とし，PDB 規模は 10 人，100 人，1,000 人，10,000 人とした．

ユーザ数が多い場合，多くの属性を開示しても D_F の平均値の減少率は低い．ユーザ数が少ない場合，少しの属性開示で，D_F の平均値の減少率は高い．よって，PDB 規模が大きい方がより多くの属性を開示できる．

図 4 に，属性値分布依存性に関する D_F の評価結果を示す．ここでは，PDB 規模を固定とし，Poisson 分布における λ（1 回の試行での整数値 k の発生する平均）に依存した属性値分布に偏りをつけた場合について評価した．属性値分布に偏りをつける際に Poisson 分布を選んだ理由は，λ 値を変更するだけで分布の形状を変えることが可能であるからである．PDB 規模は，1,000 とした．λ＝1, 5, 10 とし，λ＝1 の場合は，（氏名，Alice）の人と同じ属性値の人が多い場合を示し，λ＝10 の場合は，（氏名，Alice）の人と違う属性の人が多いとした．横軸と縦軸は，規模依存性の評価の場合と同じである．

λ＝1 の場合は，D_F の平均値の減少率は低い．λ＝10 の場合，D_F の平均値の減少率は高い．これらのことから，秘匿したい属性値と同じ属性値の人が多い場合，多くの属性を開示できるが，少ない場合は，ほとんどの属性を開示できない．

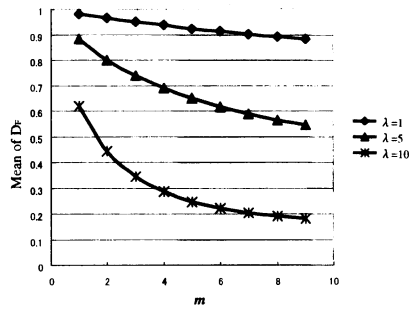

図 4　属性分布依存性

上記の結果より，D_F は，秘匿性の評価尺度として適切であると考える．

4. 秘匿性閾値

実用システムでは，D_F のある値を秘匿性閾値 D_{FT} として用い，D_{FT} により開示の可否判断を行う．D_{FT} は，適用するサービスやプロバイダが PDB に対してどの程度背景知識を持っているかなどによって値が異なる．ここでは，プロバイダが PDB に対して，背景知識を持たない場合について，サービス毎の閾値の決定方法について述べる．

秘匿性閾値を決定するためには，（1）サービス利用/提供に必要な開示属性数の上限値/下限値の明確化と，（2）開示した際の秘匿性レベルの把握が必要である．

4.1. 開示属性数の限界値

先ず，開示属性数の限界値（上限または下限）について述べる．ユーザとしては，できればプライバシー情報を開示しないで，サービスを受けたい．プロバイダは，サービスを提供する代わりに，できるだけ多くのプライバシー情報を開示して欲しい．つまり，ユー

ザとプロバイダのプライバシー情報開示に対する要求は相反する．そこで，両者のプライバシー情報開示に関する要求の均衡を図るための均衡モデルを提案する．均衡モデルは，割引などのサービス特典を与えた場合，ユーザとプロバイダはそれぞれどの程度の数のプライバシー情報を開示してもよい，または開示して欲しいかの均衡を求めるモデルである（図5参照）．均衡モデルでは，ミクロ経済学の市場均衡[6]を参考に，あるサービス特典に対して，ユーザがプライバシー情報を最大どの程度開示してもよいかを供給（supply）とし，プロバイダが少なくともどの程度開示して欲しいかを需要（demand）とみなし，需要と供給の均衡点から開示属性数を求める．均衡点の開示属性数は，サービス毎に異なると考えられる．そこで，代表的な業種である外食，旅行，ダイエットのそれぞれに対し，金銭・ポイント，情報提供（例：広告配送），景品，便利系（例：飲酒した後のタクシーを手配する居酒屋＋タクシーサービス），キャンペーン系（例：誕生日やクリスマスキャンペーン）の5つの特典別に，ユーザとプロバイダのそれぞれの観点から，プライバシー情報開示要求の均衡点をwebアンケートにより求めることとした．特典は，既存の各産業における特典を元に列挙した．

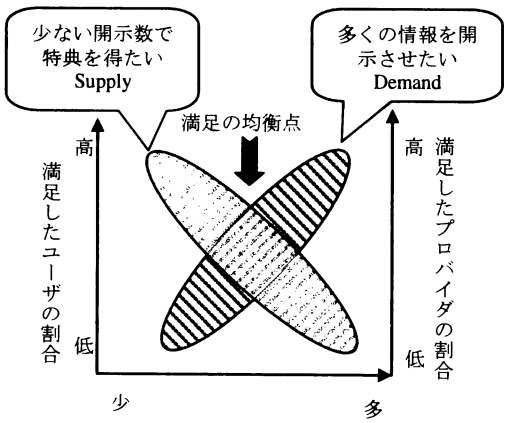

図5　均衡モデル

webアンケートは，2005.1～3月に実施し，ユーザ約16,000人，3業種において企画業務に携わっているプロバイダ社員約200～600人からの回答を得た．

均衡モデルの適用例として，業種：外食，特典：金銭･ポイントのアンケート結果を図6に示す．横軸は，金銭・ポイント特典に対してユーザが開示してもよいと考える最大の属性数，およびプロバイダが開示して欲しい最小の属性数である．縦軸は，ユーザとプロバイダの当該特典に対して満足した人の割合である．この結果，ユーザとプロバイダの均衡点における開示属性数は，5～6個となった．

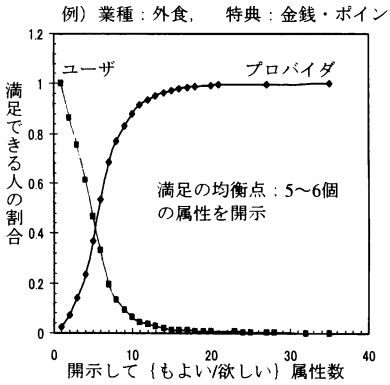

図6　均衡モデルに外食産業：金銭・ポイント特典を適用した場合

表1に3業種5特典の均衡点における開示属性数を示す．この結果，外食の場合平均の開示属性数が少なく，旅行やダイエットの場合多いことが分かった．外食の開示属性数が少ない理由は，現状，何のプライバシー情報をプロバイダに提供しなくても，路上やチラシで割引券を入手できるというサービス利用／提供形態が定着しているためと考えられる．旅行やダイエットは，個人の嗜好や生活スタイルに合わせた個人指向型のサービス利用提供形態なので，多くの属性を提供/要求すると考えられる．

表1　3業種5特典の属性開示個数

特典＼業種	外食	旅行	ダイエット
金銭・ポイント	5～6	17～18	20～21
商品	4	5～6	12～13
便利系	6～7	14～15	13～14
キャンペーン	5～6	18～19	14
情報提供	5～6	14	15～16

4.2. 秘匿性レベル

4.1節で述べたようなユーザとプロバイダの両方が満足する開示属性数について論ずる際，ユーザはどの程度の秘匿性を前提としていたのであろうか．均衡点モデルにおいて，あるn個の属性を開示した際に，ユーザの要求する秘匿性を確保できないのであれば，秘匿性閾値として十分ではない．そこで本節では，プライバシー情報の開示に当たって，ユーザが求める秘匿性の程度を明らかにする．そこで，自分が何人に1人にまでに絞りこまれると不安を感じるかについて4.1節で述べたwebアンケート調査を実施した．

回答者数は，約11.000人である．調査の結果，セキュリティ対策を何も講じていない場合，1.000人に1人，個人情報に対して暗号などの処理を施した場合，

100 人に 1 人，氏名や電話番号など個人特定につながる可能性のある個人情報を一切ださない場合，100 人に 1 人に，それぞれ絞り込まれると不安に感じることが分かった．よって，1,000 人に 1 人に絞り込まれると不安に感じる場合が D_F の上限に相当し，100 人に 1 人が D_F の下限に相当すると考えられる．

4.3. 秘匿性閾値の計算例

以下に，1 人のユーザ Alice を秘匿する場合の秘匿性閾値 D_{FT} の計算例を示す．

先ず，4.2 節のユーザが要求する秘匿性から D_F 値を求める．プロバイダに PDB へ背景知識がないと仮定した場合，式 (1)，(2)，(3) より，PDB 規模が 10,000 人の場合は，D_F＝0.77(1,000 人に 1 人に絞られた場合)～0.35（100 人に 1 人に絞られた場合），100,000 人の場合は，D_F＝0.63（1,000 人に 1 人に絞られた場合）～0.29（100 人に 1 人に絞られた場合）となる．

次に，4.1 節の開示属性数の限界値から，D_F 値を求める．例えば，属性値分布が一様分布の場合，図 3 より PDB 規模 10,000 人の属性の開示個数 5～6 個（外食産業，金銭・ポイント特典）の時の D_F 値平均は，約 0.55 である．これは，上記 10,000 人の場合の D_F＝0.77～0.35 の中に含まれるので，この場合の秘匿性閾値 D_{FT} は，0.55 程度でよいと考える．D_F 値が 0.77～0.35 の間にない場合，ユーザとプロバイダの開示要求の均衡が図れないため，現状の均衡モデルを用いた秘匿性閾値設定は行うべきではない．

5. 関連研究

データマイニングの分野において，プライバシー保護を目的とした研究が行われている[7][8][9][10]．privacy preserving data mining では，元となる PDB にノイズを入れたり[11]，PDB の分割再構成時にマイニングを行うことで，情報開示時の個人のプライバシーを保護している[12]．

多くの研究は，ある固定的なひとつの PDB に対して，preserving することが目的であり，我々のような PDB 規模や属性値分布が変わる環境を前提としてはいない．さらに，LooM のような秘匿性の定量化を目指した研究はほとんど行われていない．

6. まとめ

筆者らが提案する秘匿性定量化手法 LooM について，プライバシー情報データベース（PDB）の規模や属性値分布への依存性が少なく，1 つの閾値設定で，秘匿性を評価できることを実証した．LooM の適正について，PDB 規模依存性，属性値分布依存性の観点から評価を行い，良好な結果を得た．実用システムにおける秘匿性評価尺度 D_F の秘匿性閾値 D_{FT} を規定するために，web アンケートを実施した．3 業種 5 特典に対して，均衡モデルからユーザとプロバイダの両者が満足する属性開示個数と，ユーザが要求する秘匿性について明らかにし，これらから秘匿性閾値決定方法を示した．

今後は，LooM の改良，均衡モデルの改良，アンケート結果の詳細な分析，実用システムに向けた検討を行っていく．

文　献

[1] M. Langheinrich, "Privacy by Design Principles of Privacy-Aware Ubiquitous Systems," Proc. Ubicomp 2001, pp. 273-291, Springer-Verlag LNCS 2201, 2001.

[2] A. Beresford, and F. Stajano, "Location Privacy in Pervasive Computing," IEEE Pervasive Computing, Vol. 2, No. 1, pp. 46-55, 2003.

[3] M. Langheinrich, "Privacy Invasions in Ubiquitous Computing," Privacy In Ubicomp'2002, Sept. 2002.

[4] S. Spiekermann, J. Grossklags, and B. Berendt, "E-privacy in 2nd generation E-Commerce," Proceedings of ACM Conference on Electronic Commerce, October 2001.

[5] T. M. Mitchell, "Machine Learning," WCB McGraw-Hill, 1997.

[6] Supply and Demand; http://www.netmba.com/econ/micro/supply-demand/

[7] L. Sweeney, "k-anonymity: a model for protecting privacy, "International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, pp. 557-570, Oct. 2002.

[8] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy-Preserving Mining of Association Rules," Information Systems, 29(4), June 2004.

[9] S. Oliveira and O. R. Zaiane, "Achieving Privacy Preservation When Sharing Data For Clustering," Workshop on Secure Data Management in a Connected World (SDM'04) in conjunction with VLDB'2004,Springer Verlag LNCS 3178, pp 67-82, Toronto, Canada, Aug. 30, 2004.

[10]]R. Agrawal, and R. Srikant, "Privacy-Preserving Data Mining," ACM SIGMOD Int'l Conf. on Management of Data, Dallas, May 2000.

[11]Y. Lindell, and B. Pinkas, "Privacy Preserving Data Mining," Proceedings of the 20th Annual International Cryptology Conference on Advances in Cryptology, pp. 36-54, August 20-24, 2000.

[12]C. Clifton, and D. Marks, "Security and Privacy Implication of Data Mining," Proceedings of ACM SIGMOD workshop on Data Mining and Knowledge Discovery, 1996.

[13]今田美幸，高杉耕一，太田昌克，小柳恵一，"ユビキタスネットワーキング環境におけるプライバシ保護手法　LooM，" 信学論（B），pp.563-573, 2005.3