情報処理学会研究報告
IPSJ SIG Technical Report

Vol.2010-CSEC-51 No.8
2010/12/10

セキュアマッチングとナイーブベイズ識別器を用いたプライバシー保護リコメンド方式

山口高康[†1]　寺田雅之[†1]

近年，暗号プロトコルや統計的開示制御技術の導入により，組織の枠を超えてリコメンドを提供する方式が注目されている．本稿では，保持しているデータ量や通信・計算リソースが異なる三者を想定し，各々のプライバシーを保護しつつリコメンド機能を提供する方式を検討した．提案手法により，顧客情報を持つ Alice と販売情報をもつ Bob と来訪客 Charlie との間における処理において，プライバシーを保護できる安全性と，実績のあるナイーブベイズの精度と，顧客数と商品数の増加に対するスケールメリットが得られる．

An Efficient Privacy-Preserving Recommendation Method based on Secure Matching and Naive-Bayes Classifier

Takayasu Yamaguchi[†1] and Masayuki Terada[†1]

Preserving privacy of respondents is indispensable for cross-organization recommendations. We propose an efficient privacy-preserving recommendation method based on secure matching and naive-bayes classifier for heterogeneous environments. The proposed method is scalable and works quite efficiently in environments consist of servers, workstations and mobile terminals, each of which has different amount of data, communication speed and computation resources.

†1 (株)NTT ドコモ 先進技術研究所
NTT DOCOMO Research Laboratories

1. はじめに

1.1 背　景

ユーザがモバイル環境で有用な情報をタイムリーに手に入れるために，情報提示機能の高度化に対する要求が高まっている．モバイルでは端末の画面サイズに制限があるため，効率の良い情報提示機能が必要である．情報提示機能の高度化の実現方法として，ユーザの属性や商品の購入履歴に基づいて，ユーザの趣向に合う商品を推薦するリコメンド技術が提案されている．

しかし，ユーザ属性や商品の購入履歴といった推薦の元となるデータは，複数の企業にわたっている場合も多い．それ故，実際にはユーザのプライバシーや企業のデータを保護する為に，組織の枠を超えてリコメンドを行うことは難しかった．

ところが，近年，暗号プロトコルや統計的開示制御技術の導入により，組織の枠を超えてリコメンドを提供する方式が注目されている．

1.2 研究の位置づけと目的

本稿では，モバイル環境における情報提示機能の高度化に向けて，保持しているデータ量や通信・計算リソースが異なる三者が，各々のプライバシーを保護しつつリコメンド機能を提供する方式を提案する．

リコメンドには，協調フィルタリングとコンテンツベースの方法がある．前者の協調フィルタリングは，計算機が，あるユーザと購買履歴の似ているユーザを探し出し，その購買履歴の似ているユーザが購入した商品をリコメンドする．後者のコンテンツベースは，計算機が，あるユーザが好む商品と似ている商品を探し出し，その商品をリコメンドする．後者は前者に比べて新商品のような履歴が少ない商品でもリコメンドし易いという利点がある．

両者ともユーザのプライベートな情報を取り扱うため，ユーザのプライバシーを保護する必要がある．前者は，秘匿積集合プロトコルを利用してプライバシを保護しながら協調フィルタリングを行う方法[6)] が提案されている．後者は，ユーザのプライバシを保護するべく，ユーザの好みと商品のマッチングを，例えば洋服の寸法を S,M,L というサイズや，料理を和，洋，中というカテゴリで行う方法はあるが，ユーザのプライバシーを保護しつつ，効率的なリコメンドを行う方式は提案されていない．

そこで本稿では，コンテンツベースの方式に着目し，モバイルで遭遇する，履歴が十分にあるとは限らない様々な商品を，来店者の属性や履歴を開示することなく，的確かつ安全にリコメンドできるようにする．

1.3 ユースケースとメリット

本稿で取り組む，三者でのリコメンドのユースケースと，本ユースケースにおける，三者のそれぞれのメリットについて述べる．

今，顧客のユーザ属性を管理している Alice と，店舗を経営している Bob と，Bob の店に買い物にやってきた Charlie の三者がいて，Bob は Charlie へのリコメンドにより，店の売り上げを上げたいと考えているとする．

Alice は顧客から預かっているユーザ属性を漏洩したくない．

Bob は，Bob の店の売り上げ情報と，Bob の店で商品を購入したユーザ属性の統計値とを漏洩したくない．Bob は，顧客情報を持つ Alice とのコミュニケーションを通じて，Bob の店で商品を購入したユーザ属性の統計値が得られる．ただし，ユーザ属性の統計値は，ある人数以上の顧客を集計対象とした場合にのみに限られる．

Charlie は，Charlie 自身のユーザ属性と，Charlie が閲覧する商品の履歴とを漏洩したくない．Charlie は，Bob の店に初めて来店したとしても，面倒な会員申し込み手続きなどをせずにスムーズにリコメンドを受けたい．ただし，Charlie は Alice の顧客であるとは限らない．

上記のユースケースを実現するには，三者が以下のようなメリットを享受できるようにしなければならない．

(1) Alice は顧客から預かっているユーザ属性の情報漏えいを防止できる
(2) Bob は，Bob の店で販売した商品の情報を Alice に渡さずに，Bob の店の商品を購入する可能性の高いユーザ属性の傾向を知ることができる
(3) Charlie は，Charlie のプライバシー情報を Alice や Bob に渡さずに Charlie が欲しいもの順に商品を閲覧できる

1.4 要求条件と課題

1.3 小節で述べたユースケースにおいて，Alice, Bob, Charlie がメリットを享受するための要求条件を以下に纏める．

(1) 安全性
ユーザのプライバシーと店舗のノウハウを保護できる
(2) 精度
精度良くリコメンドできる
(3) スケール
顧客数と商品数が多くなっても通信量と計算量の増加が少なく，軽快に動作できる

1.3 小節で述べたユースケースを従来方式の秘匿内積計算を用いたナイーブベイズ識別器で実現しても，顧客数や商品数が多い場合にスケールしなくなるという問題がある．そこで本稿では，特に Alice と Bob の間での処理速度向上を課題として取り上げて，その解決を試みる．

2 章では準備として，本稿で用いる記号と用語を定義して，適用範囲と前提条件を述べる．3 章では従来方式として，秘匿内積計算を用いたナイーブベイズ識別器や，セキュアマッチングを用いたクロス集計を述べる．4 章では提案方式として，Alice と Bob の間での処理速度を，秘匿内積計算を用いたナイーブベイズ識別器よりも向上させる方式を述べる．5 章では評価として，要求条件に照らした各方式の評価を述べる．6 章ではまとめを述べる．

2. 準　　備

2.1 記号と用語の定義

- $\boldsymbol{t}_A \in \mathcal{Z}^{N_A}$:会員の端末識別子
- $X \in \mathcal{R}^{V \times N_A}$:会員のプロファイル $(x_{v,n_A} > 0)$
- $\boldsymbol{t}_B \in \mathcal{Z}^{N_B}$:購入者の端末識別子
- $Y \in \{0,1\}^{L \times N_B}$:購入された商品 $\left(\sum_{l=1}^{L} y_{l,n_B} = 1\right)$
- $\acute{\boldsymbol{x}} \in \mathcal{R}^{V}$:Charlie のプロファイル $(\acute{x}_v > 0)$
- $\boldsymbol{\pi}^{(l)} \in \mathcal{R}^{V}$:l 番目の商品についての会員のプロファイルの集計値
- $\hat{\boldsymbol{\theta}}^{(l)} \in \mathcal{R}^{V}$:l 番目の商品についてのリコメンドのパラメータ $\left(\sum_{v=1}^{V} \theta_v^{(l)} = 1\right)$
- $P(\acute{y}^{(l)}|\acute{\boldsymbol{x}}, \hat{\boldsymbol{\theta}}^{(l)})$:l 番目の商品についての Charlie へのリコメンド値

2.2 適用範囲と前提条件

1.3 小節で述べた登場人物と，2.1 小節で定義した各種情報とを図 1 に示す．

本稿では，簡単のため，Bob の店にやってきた客は，1 度の買い物で L 種類の商品の中から 1 種類の商品だけを購入することとする．

1.4 小節で述べた通り，精度良いリコメンドが要求されるため，Charlie へのリコメンド情報を算出する際には古くから実績のあるナイーブベイズ[2] を用いることとする．

ベイズのアプローチでは分布の仮定を置く必要がある．人の属性を表すプロファイルは多項分布に従うこととする．すなわち，プロファイルを V 個の項目からなるベクトルで表し

Vol.2010-CSEC-51 No.8
2010/12/10

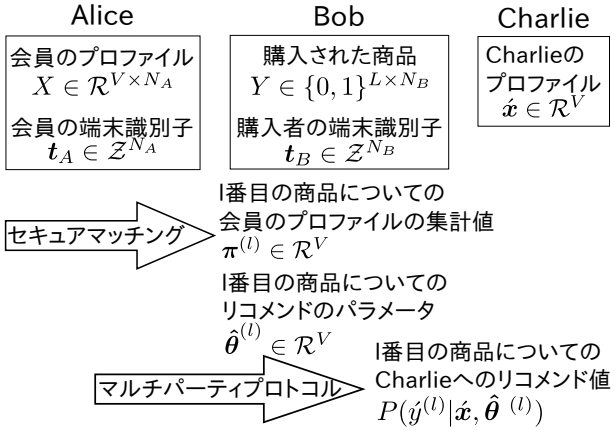

図 1 登場人物と各種情報の関係

て，各項目の値を 0 以上とする．また，リコメンドのパラメータの分布は，一般に用いられている共役事前分布を用いる[8]．多項分布の共役事前分布はディレクレ分布であるので，リコメンドのパラメータを V 個の項目からなるベクトルで表して，各項目の値の和を 1 とする．

リコメンドのパラメータを求める際には，事後確率を最大にする観点から，l 番目の商品についてのリコメンドのパラメータ $\hat{\theta}_v^{(l)}$ を MAP 推定で求めることとする．本稿で用いるベイズアプローチ識別器を一般的な表現で記述すると付録 A.1 のようになる．

ナイーブベイズは説明変数の独立性を仮定する手法である．そのため，Alice が有する会員のプロファイルの V 個の項目は独立であるとみなされるとする．すなわち，l 番目の商品を購入した会員の v 番目のプロファイルの値を独立に集計できれば良い．1.4 小節で述べた通り，Alice にプロファイルを預ける会員数 N_A や Bob の店での購入者数 N_B が多くなってもスケールすることが要求されるため，Alice が Bob に会員のプロファイルの集計値を開示するのにセキュアマッチング[4] を用いることとする．

多項分布は指数分布族であるので，l 番目の商品についての Charlie へのリコメンド値を算出する際に，Charlie のプロファイルとリコメンドのパラメータの対数との内積計算が必要となる．1.4 小節で述べた通り，ユーザのプライバシーと，店舗のノウハウとを保護できることが要求されるため，Bob が Charlie にリコメンド情報を提供するのに秘匿内積計算[7] を用いることとする．

3. 関 連 手 法

3.1 秘匿内積計算を用いたナイーブベイズ識別器

● Alice-Bob 間

Vaidya らは，分割したデータベースを互いに秘匿したままでナイーブベイズ識別器を実現する方式を提案している[7]．属性情報を保持する Alice とクラス情報を保持する Bob とが協力して，ナイーブベイズの計算を行う．ただし，Alice と Bob とがそれぞれ持っているデータは同期していることとする．

ナイーブベイズの独立性の仮定から，$x_v^{(l)}$ のそれぞれの条件下における $y^{(l)}$ の条件付確率を求め，未知パラメータを学習する．

$$P(y^{(l)}|x_v^{(l)}) = \frac{P(y^{(l)}, x_v^{(l)})}{P(x_v^{(l)})} = \frac{y^{(l)} \cdot x_v^{(l)}}{N} \frac{N}{|x_v^{(l)}|} \tag{1}$$

データは同期している前提なので，この条件付確率の計算式では N が打ち消しあう．また，Bob は単独で $|x_v^{(l)}|$ を計算できる．よって，$y^{(l)} \cdot x_v^{(l)}$ の内積計算さえできれば，未知パラメータを学習できる．

この内積計算は秘匿内積計算プロトコルで実現する．秘匿内積計算プロトコルの一般的な記述を付録 A.2 に示す．秘匿内積計算は原理的にべき乗計算が必要となるため，計算効率を高める事は困難である．

● Bob-Charlie 間

データの識別は，二者間秘匿回路計算プロトコル[9] を実行して，最大尤度のクラスを求める．しかし，この計算コストも高い．

3.2 セキュアマッチングを用いたクロス集計

● Alice-Bob 間

セキュアマッチングプロトコルは秘匿識別子集計の実現方式である．データを秘匿したまま積集合を求めるプロトコルには多項式評価[5] があるが，多項式の大きさ n に対して $O(n^2)$ の計算量がかかった．それ故，秘匿識別子集計では大きなクロス集計表の作成は非効率であると考えられてきた．

しかし，Agrawal らは，可換な一方向性関数を用いて，ハッシュ値の生成に $O(n)$，照合に $O(\log n)$ のコストで計算できる照合タグ方式を提案した[1]．また，千田らは，ランダムオラクルモデルの上で照合タグ方式の安全性を証明し[3]，さらに計算

量と通信量を削減して，クロス集計表を効率的に作成する方式を提案している[4]．

● Bob-Charlie 間

クロス集計表から，安全にデータを識別する方法は提案されていない．

4. 提 案 方 式

提案方式は，Alice-Bob 間，Bob-Charlie 間での処理に Bob 内での処理を加えて，下記の 3 つのフェーズで構成する．

● Alice-Bob 間

Bob は，Alice とのセキュアマッチングにより，l 番目の商品についての会員のプロファイルの集計値を得る．一般的なプロトコルでは Alice と Bob の両方が集計値を得るが，ここでは Bob のみが集計値を得るようにする (付録 A.3)．すなわち，Bob の店の営業上のノウハウ等が，Alice に流出することはない．

$$\pi_v^{(l)} = |x_{n,v} \cap y^{(l)}|_n = \sum_{n=1}^{N^{(l)}} x_{n,v}^{(l)} \tag{2}$$

● Bob 内

Bob は，l 番目の商品についての会員のプロファイルの集計値を用いて，店舗のノウハウである l 番目の商品についてのリコメンドのパラメータを学習する (付録 A.6)．

$$\hat{\theta}_v^{(l)} = \frac{\pi_v^{(l)} + (\xi_v^{(l)} - 1)}{\left(\sum_{v=1}^{V} \pi_v^{(l)}\right) + V(\xi_v^{(l)} - 1)} \tag{3}$$

● Bob-Charlie 間

Charlie は，Bob との秘匿内積計算により，Charlie のプロファイルと，学習済みの l 番目の商品についてのリコメンドのパラメータとを用いて，l 番目の商品についての Charlie へのリコメンド値を算出する．(付録 A.2)．ここで，Charlie 自身のユーザ属性と，Charlie の閲覧する商品の履歴とは，Bob に知られることはないため，Charlie のプライバシーは保護される．

$$P(\acute{y}^{(l)}|\acute{\boldsymbol{x}}, \hat{\boldsymbol{\theta}}^{(l)}) = \left(\sum_{v=1}^{V} \acute{x}_v \log \hat{\theta}_v^{(l)}\right) + \left((\xi_v^{(l)} - 1) \sum_{v=1}^{V} \log \hat{\theta}_v^{(l)}\right) \tag{4}$$

表 1 各手法の比較評価結果

評価項目	SIP+NB	SMP+CA	提案方式
安全性	○	△	△
精度	○	△	○
スケール	×	○	○

5. 評　　価

1.4 小節で述べた要求条件に照らして，本稿で述べた各手法の比較を通じて評価する．

安全性の面では，秘匿内積計算を用いたナイーブベイズ識別器 (SIP+NB) が，お互いの情報を一切漏らさないので安全である．セキュアマッチングを用いたクロス集計 (SMP+CA) と提案方式は，クロス集計値が Alice の情報のうちいくらかの情報を Bob に与える．ただし，Alice の DB 内の顧客のプライバシーが漏れることはない．

精度の面では，SIP+NB と提案方式が，実績のあるナイーブベイズの手法を実装できるので精度が高い．SMP+CA はクロス集計表からリコメンドに用いる識別結果を直接得られないので精度が高くない．

スケールの面では，SMP+CA と提案方式が，大きなクロス集計表を効率的に作成できるのでスケールする．SIP+NB は秘匿内積計算のコストが高いのでスケーラビリティに劣る．

6. ま と め

本稿では，保持しているデータ量や通信・計算リソースが異なる三者を想定し，各々のプライバシーを保護しつつリコメンド機能を提供する方式を検討した．新たに Bob 内でナイーブベイズの学習パラメーターを算出する処理を加えることによって，Alice-Bob 間での処理を，スケーラビリティの低い秘匿内積計算ではなく，スケーラビリティの高いセキュアマッチングを用いることができるようになる．この工夫を取り入れた提案手法によれば，顧客情報を持つ Alice と販売情報をもつ Bob と来訪客 Charlie との間における処理において，プライバシーを保護できる安全性

と，実績のあるナイーブベイズの精度と，顧客数と商品数の増加に対するスケールメリットが得られる．提案方式の定量評価は今後の課題である．

付　　録

A.1 ベイズアプローチ識別器

$\acute{\boldsymbol{d}}$ の事後の予測分布

$$p(\acute{\boldsymbol{d}}|\mathbb{D}) = \int_{\Theta} m(\hat{\boldsymbol{d}}|\Theta)p(\Theta|\mathbb{D})d\Theta \tag{5}$$

学習データ $\mathbb{D}$ を用いて，未知パラメータマトリクス Θ を事後確率最大の観点で学習する．ここで，学習データ $\mathbb{D}$ はユーザ属性マトリクス $X \in \mathcal{R}^{V\times N}$ と購買動向マトリクス $Y \in \{0,1\}^{L\times N}$ の組である．すなわち，$\mathbb{D} = \{X, Y\}$ である．

学習データ $\mathbb{D}$ を得た後の未知パラメータ Θ の事後確率はベイズの定理より，以下の比例関係が得られる．

$$p(\Theta|\mathbb{D}) = \frac{p(\mathbb{D}|\Theta)p(\Theta)}{p(\Theta)} \propto p(\mathbb{D}|\Theta)p(\Theta) \tag{6}$$

上式の比例関係の対数をとっても大小関係は変わらない．そこで，上式の比例関係の対数の式で未知パラメータマトリクス $\Theta \in \mathcal{R}^{V\times L}$ を動かし，事後確率を最大とする未知パラメータ $\hat{\Theta}$ を求める．

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} P(\Theta|\mathbb{D}) = \underset{\Theta}{\operatorname{argmax}} \left[\mathcal{L}(\mathbb{D};\Theta) + \log P(\Theta)\right] \tag{7}$$

D の事後の予測分布に $\hat{\Theta}$ を代入すれば，未知の属性 $\hat{\boldsymbol{x}}$ を L 個のクラスに識別するベイズ識別器は以下のようになる．

$$\acute{y} = \underset{l}{\operatorname{argmax}} \left[\mathcal{L}(\hat{\boldsymbol{x}};\hat{\boldsymbol{\theta}}^{(l)}) + \log P(\hat{\boldsymbol{\theta}}^{(l)}|\mathbb{D})\right] \tag{8}$$

A.2 秘匿内積計算

入力 :Alice が持つ $\boldsymbol{x} = (x_1, ..., x_n)$ と Bob が持つ $\boldsymbol{y} = (y_1, ..., y_n)$
出力 :$r_A + r_B = \boldsymbol{x} \cdot \boldsymbol{y}$ を満たす，Alice が持つ r_A と Bob が持つ r_B

(1) Bob は加法準同型性を満たす Bob の公開鍵で $\boldsymbol{y}$ を暗号化して，Alice へ $E(y_1), ..., E(y_n)$ を送る
(2) Alice は，乱数 r_A を選び，Bob へ $c = E(y_1)^{x_1}, ..., E(y_n)^{x_n}$ を送る
(3) Bob は $r_B = D(c) = x_1y_1 + ... + x_ny_n$ を得る

A.3 セキュアマッチングによる秘匿集計

入力 :Alice が持つ $X = (x_1, ..., x_{N_A})$ と Bob が持つ $Y = (y_1, ..., y_{N_B})$
出力 :$|X \cap Y|$

(1) Bob は乱数 $r_B \in Z_q$ を選び，位数 q の巡回群 G と，G を値域とするハッシュ関数 H で，$\boldsymbol{y}$ を暗号化して，Alice へ $H(y_1)^{r_B}, ..., H(y_{N_B})^{r_B}$ を送る
(2) Alice は，乱数 $r_A \in Z_q$ を選び，Bob へ $H(x_1)^{r_A}, ..., H(x_{N_A})^{r_A}$ と $H(y_1)^{r_Br_A}, ..., H(y_{N_B})^{r_Ar_B}$ とをシャッフルして送る
(3) Bob は $H(x_v)^{r_Ar_B} = H(y_v)^{r_Br_A}$ を満たす個数 $|X \cap Y|$ を得る

A.4 多項分布を仮定した尤度関数

$$Multi\{N, \boldsymbol{x}^{(l)}; \boldsymbol{\theta}^{(l)}\} = \frac{N^{(l)}!}{\prod_v x_v^{(l)}!}\{\theta_v^{(l)}\}^{x_v^{(l)}} \tag{9}$$

$$P(\boldsymbol{x}^{(l)}|\boldsymbol{\theta}^{(l)}) \propto \prod_{v=1}^{V}(\theta_v^{(l)})^{x_v^{(l)}} \tag{10}$$

$$\log P(\boldsymbol{x}^{(l)}|\boldsymbol{\theta}^{(l)}) \propto \sum_{v=1}^{V} x_v^{(l)} \log \theta_v^{(l)} \tag{11}$$

$$\mathcal{L}(X^{(l)};\boldsymbol{\theta}^{(l)}) \triangleq \log P(X^{(l)}|\boldsymbol{\theta}^{(l)}) \propto \sum_{n=1}^{N^{(l)}}\sum_{v=1}^{V} x_{n,v}^{(l)} \log \theta_v^{(l)} \tag{12}$$

A.5 ディレクレ分布を仮定した事前確率の対数

$$Dirichlet\{\boldsymbol{\theta}^{(l)}; \boldsymbol{\xi}^{(l)}\} = \frac{\Gamma(\sum_v \xi_v^{(l)})}{\prod_v \Gamma(\xi_v^{(l)})}\prod_v \{\theta_v^{(l)}\}^{\xi_v^{(l)}-1} \tag{13}$$

$$P(\boldsymbol{\theta}^{(l)}) \propto \prod_{v=1}^{V}\{\theta_v^{(l)}\}^{\xi_v^{(l)}-1} \tag{14}$$

$$\log P(\boldsymbol{\theta}^{(l)}) \propto (\xi_v^{(l)} - 1)\sum_{v=1}^{V} \log \theta_v^{(l)} \tag{15}$$

A.6 最大の事後確率を得る未知パラメータの推定

$$\hat{\boldsymbol{\theta}}^{(l)} = \underset{\boldsymbol{\theta}^{(l)}}{\operatorname{argmax}}\{\mathcal{L}(X^{(l)};\boldsymbol{\theta}^{(l)}) + \log P(\boldsymbol{\theta}^{(l)})\} \tag{16}$$

最大化したい上式の中括弧の中に，尤度関数と事前確率の対数を代入する．

$$\mathcal{L}(X^{(l)};\boldsymbol{\theta}^{(l)}) + \log P(\boldsymbol{\theta}^{(l)}) = \left(\sum_{n=1}^{N^{(l)}}\sum_{v=1}^{V} x_{n,v}^{(l)} \log \theta_v^{(l)}\right) + \left((\xi_v^{(l)} - 1)\sum_{v=1}^{V} \log \theta_v^{(l)}\right) \tag{17}$$

目的関数を J とおき，未知パラメータを求める．ラグランジュの未定係数法により，$\sum_{v=1}^{V} \theta_v^{(l)} = 1$ の条件下で J の最大化を行う．

$$J = \left(\sum_{n=1}^{N^{(l)}}\sum_{v=1}^{V} x_{n,v}^{(l)} \log \theta_v^{(l)}\right) + \left((\xi_v^{(l)} - 1)\sum_{v=1}^{V} \log \theta_v^{(l)}\right) + \lambda\left(\sum_{v=1}^{V} \theta_v^{(l)} - 1\right) \to Max \tag{18}$$

$$\frac{\partial J}{\partial \theta_v^{(l)}} = \left(\frac{1}{\hat{\theta}_v^{(l)}}\sum_{n=1}^{N^{(l)}} x_{n,v}^{(l)}\right) + (\xi_v^{(l)} - 1)\frac{1}{\hat{\theta}_v^{(l)}} + \lambda = 0 \tag{19}$$

$$\hat{\theta}_v^{(l)} = \frac{\left(\sum_{n=1}^{N^{(l)}} x_{n,v}^{(l)}\right) + (\xi_v^{(l)} - 1)}{-\lambda} \tag{20}$$

条件より，$\sum_{v=1}^{V} \hat{\theta}_v^{(l)} = 1$ であるから，未知パラメータは以下のようになる．

$$\sum_{v=1}^{V} \hat{\theta}_v^{(l)} = \sum_{v=1}^{V} \frac{\left(\sum_{n=1}^{N^{(l)}} x_{n,v}^{(l)}\right) + (\xi_v^{(l)} - 1)}{-\lambda} = 1 \tag{21}$$

$$\lambda = -\left(\sum_{v=1}^{V}\sum_{n=1}^{N^{(l)}} x_{n,v}^{(l)}\right) - V(\xi_v^{(l)} - 1) \tag{22}$$

$$\hat{\theta}_v^{(l)} = \frac{\left(\sum_{n=1}^{N^{(l)}} x_{n,v}^{(l)}\right) + (\xi_v^{(l)} - 1)}{\left(\sum_{v=1}^{V}\sum_{n=1}^{N^{(l)}} x_{n,v}^{(l)}\right) + V(\xi_v^{(l)} - 1)} \tag{23}$$

謝辞 本稿の検討にあたりアドバイスを頂きました，東海大学の菊池浩明教授と日本電信電話 (株) 情報流通プラットフォーム研究所の千田浩司氏に感謝いたします．

参 考 文 献

1) R. Agrawal, A. V. Evfimievski, and R. Srikant. Information sharing across private databases. *In ACM SIGMOD 2003*, pp. 86-97, 2003.
2) Charu C. Aggarwal, and Philip S. Yu. Privacy-Preserving Data Mining: Models and Algorithms. *Springer*, 2008.
3) 千田, 五十嵐, 高橋. 照合タグを用いた秘匿共通集合計算プロトコルとその応用. コンピュータセキュリティシンポジウム *2009*, IPSJ, 2009.
4) 千田, 寺田, 山口, 五十嵐, 濱田. セキュアマッチングを用いた組織間クロス分析. コンピュータセキュリティシンポジウム *2010*, IPSJ, 2010.
5) Michael J. Freedman, Kobbi Nissim, and Benny Pinkas. Efficient private matching and set intersection. *In Eurocrypt 2004*, IACR, 2004.
6) 木澤, 磯崎, 菊池, 秘匿積集合プロトコルを利用したプライバシ協調フィルタリングの提案. 暗号と情報セキュリティシンポジウム *2009*, SCIS, 2009.
7) Jaideep Vaidya, and Chris Clifton. Privacy preserving naive bayes classifier for vertically partitioned data. *In Society for Industrial and Applied Mathematics*, 2008.
8) 渡部. ベイズ統計学入門. 福村出版, 1999.
9) A. C. Yao. How to generate and exchange secrets (extended abstract). *In IEEE FOCS '86*, pp. 162-167, 1986.