科学研究費助成事業　　研究成果報告書

平成 ２６ 年　 ４ 月 ３０ 日現在

機関番号：１２６０１
研究種目：挑戦的萌芽研究
研究期間：2012～2013
課題番号：２４６５００６４

研究課題名（和文）相関を持つデータベースに対する差分プライバシーに関する研究

研究課題名（英文）A Study on Differential Privacy for Data Base with Correlation

研究代表者
中川　裕志（NAKAGAWA, Hiroshi）

東京大学・情報基盤センター・教授

研究者番号：２０１３４８９３

交付決定額（研究期間全体）：（直接経費）　　3,000,000円、（間接経費）　　900,000円

研究成果の概要（和文）：プライバシー保護データマイニングのひとつである差分プライバシーは有望な方法であるが、データベースのレコード間に相関がある場合の分析があまり進んでいなかった。本研究では、相関がある場合に従来の差分プライバシーを適用した場合、データ入手を狙う攻撃者が相関に関する背景知識を少なく持っているほうが、流出する情報が大きいという直感に反する状況を明らかにし、この状況を改善するために背景知識も考慮したベイズ型差分プライバシーの数理モデルを確立した。この数理モデルにおいて情報漏洩の確率を与えられた閾値以下にする加算すべきラプラス雑音のパラメタを求める近似的アルゴリズムを示した。

研究成果の概要（英文）：The differential privacy is the technology of adding noise to an answer for the given query. The differential privacy for data base with correlated data records, however, has not get much attention. In this research, we find the counter intuitive phenomena that an adversary with small amount of background knowledge about correlations can get more privacy information than the adversary with big amount of these knowledge. Then, we build the Bayesian privacy model which explains this kind of phenomena and improve this situation. We also show the approximation algorithm that gives us the proper parameter value of Laplace distribution employed in differential privacy which make the provability of information leakage less than previously determined threshold.
In addition, we investigated the problems which are caused by added noise in differential privacy.

研究分野：総合領域

科研費の分科・細目：情報学･知能情報学

キーワード：プライバシー保護　データマイニング　差分プライバシー　ビッグデータ　個人情報保護　データベース　匿名化　相関

１．研究開始当初の背景

(1)ビッグデータの重要性が認識されるにつれて，その重要な部分でありかつビジネスでも役立つパーソナルデータの利活用に注目が集まってきている．たとえば，病院における感染症に関するデータ，公共事業者における種々の事故情報，企業における販売，流通に関する情報などである．これらの生データは個人情報保護，機関毎の機密性，守秘性の観点から公開できない．しかし，複数機関がこれらのデータを部分的にせよ機関を跨って参照できれば，公共安全，企業ないし業界の発展に資する例は枚挙にいとまがない．たとえば，複数の病院のインフルエンザ患者の来院数の時間的変化が小さな時間遅れで参照できれば，流行の予兆検知，ワクチン量の適正な配分などが可能になるであろう．また，同業種における競合会社であっても，機密性に触れない情報を複数の企業の間で相互参照することによって，日本企業全体として，その業種の競争力が増すことも視野に入ってくる．

(2)上記の目的を実現する技術として 2000 年代に入ってプライバシー保護データマイニング(PPDM)の研究が盛んになってきた．PPDM において解析を行う主体はデータベースを保有する組織である．一般に，各主体は保有するデータベースにおけるプライバシーは保護したいと考えるが，上述のように複数の主体間で各自のデータベースの情報を相互に参照し有効利用できれば有益となる場合が多い．

(3)PPDM は暗号型，入力摂動型，出力摂動型に大別される．暗号型は，複数の主体が自分自身の保有するデータは暗号化したうえで, 主体間で適切なプロトコルを用いて, データは他の主体に漏洩せずにデータマイニングの結果だけを得る理想的な方法だが, 暗号化および復号化を多数回繰り返すため計算量が多く実用的ではない．入力摂動型は, 公開するデータベースの内容を変更（これを摂動と呼ぶ）して，データベースの内容から個人情報が特定されないようにする手法である．代表的な k-匿名化では，個人名を匿名化した上で，同じ属性集合（例えば，郵便番号，居住地の町名，年齢など）を持つ個人が少なくとも k 人存在するようにデータベースを改変する．個人特定がしにくいという点では有望だが，データベース改変の計算量が NP であり，データマイニングにおいて強いバイアスが作用して精度が劣化する．

(4)出力摂動は，主に差分プライバシー技術で実現され，データベースへの検索質問に対する結果にその都度異なる雑音を加算する. データベースを改変する手間が不要で, バイアスもなく有望な方法である. ただし，差分プライバシーにおいてはデータベースのレコード間に相関がある場合の分析があまり進んでいなかった.

２．研究の目的

(1)本応募研究では，この相関がある場合に差分プライバシーを拡張するための問題点を明らかにし，攻撃側が事前知識を持つ場合にも情報漏洩の確率を与えられた閾値以下にする数理モデルとしてベイズ型差分プライバシーの数理モデルを確立することを目的とした.

(2)相関があるデータベースや雑音加算などの方法が誘発する問題についても検討し, パーソナルデータを利活用する際の問題点を明らかにする.

３．研究の方法

(1)データベースにおける相関の度合いが従来の差分プライバシーに与える影響を，まず分析する．差分プライバシーでは，１要素の値だけ異なる２個のデータベース D_1，D_2 に対する質問　　f の答え $f(D_1)$と $f(D_2)$の間に次式が成立し，質問結果から D_1, D_2 のどちらを使って得た答えかが区別できる確率を制限する.

$$\forall r \quad e^{-\varepsilon} \le \frac{\Pr(M(f(D_1))=r)}{\Pr(M(f(D_2))=r)} \le e^{\varepsilon} \qquad (1)$$
$$M(f(D)) = f(D) + Lap(1/\varepsilon)$$

ただし, Lap は下式のラプラス分布である.

$$Lap(1/\varepsilon) = \frac{\varepsilon}{2}\exp(-\varepsilon|x|) \quad x\text{が確率変数}$$

式(1)によれば，D_1，D_2 の差分を保護の度合いがラプラス分布のパラメタ ε によって制御できることが特徴である．つまり，ε が大きいと大きな雑音が要求されることになる．しかし，これはデータベースにおける複数のレコードの値の間に相関がない場合の結果である．よって，相関があるとどのような現象が発生するかを明らかにする数理モデルが必要となるため，本研究ではそのようなモデル化を試みる．すなわち攻撃者がデータベースにおける相関を事前知識として持っている場合と持っていない場合の各々で，式(1)の ε で表される差分プライバシーの強度との関係を，ベイズ統計を用いて数理モデル化する．これによって，相関を持つデータベースにおける差分プライバシーの設計指針を明らかにする.

(2)相関のあるデータベースに差分プライバシーを適用したときに誘発する副作用について分析する．これは，研究を進めるうちに明らかになり，かつ差分プライバシーの実用上考慮すべき現象である．このように将来的課題についてもその萌芽を明らか

にすることを試みる．

４．研究成果

(1)相関のあるデータベースにおける差分プライバシー

まず，相関のあるデータベースを以下のように定義する．すなわち，データベース中の任意の２レコードの値d_1とd_2に相関がない場合は

$$\Pr(d_1,d_2)= \Pr(d_1)\Pr(d_2)$$

であるのに対して，相関がある場合は

$$\Pr(d_1,d_2)> \Pr(d_1)\Pr(d_2)$$

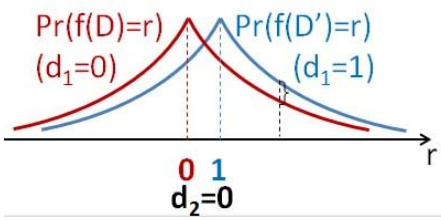
である．

ここで，攻撃者はd_1の真の値を知ることが目的であるとする．このとき，強い攻撃者はd_2の値を知っており，弱い攻撃者は知らないとする．この両者が差分プライバシーで得る情報を比較してみる．

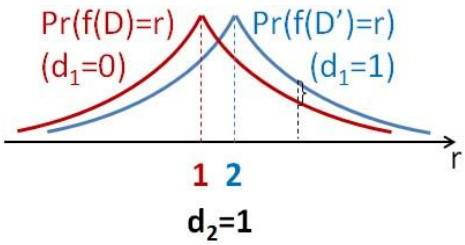
設定条件：d_1とd_2は0，1のいずれかの値をとるとする．また，両者の間には$d_1=d_2$という正の相関があるとする．また，質問は，d_1+d_2すなわち$f(D)= d_1+d_2$ とする．下の図1，図２でDは$d_1=0$のデータベース，D’は$d_1=1$のデータベースであるとする．

強い攻撃者の場合

d_2の値を知っているので，その値に応じて雑音$Lap(1/\varepsilon)$を加算すると図１a,bのどちらかになる．



a ． d_2=0の場合



b. d_2=１の場合

図１　d_2の値が知られている場合に攻撃者に分かるd_1値の分布の差異

この図のようにd_1=0，１の場合の分布の差異は同様に小さい．

弱い攻撃者の場合

この場合，質問fに対する答えは0か2である．よって，回答の分布は図2のようになる．

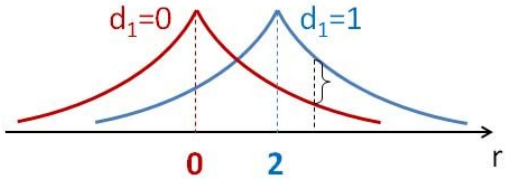


図２　d_2の値を知らない弱い攻撃者に分かるd_1値の分布の差異

図2の場合は，質問の結果が0,2と離れているため，同じεに対する$Lap(1/\varepsilon)$の加算では，d_1値の分布の差異が強い攻撃者の場合より大きくなり，攻撃者に漏れる情報は大きくなる．

以上の例から分かるように，差分プライバシーはデータベースのレコード間の相関がある場合，相関の情報を知らない弱い攻撃者に対して，相関の情報を知っている強い攻撃者に大きな情報を与えてしまう．これは，通常の差分プライバシーの弱点であり，この研究を通して得られた重要な知見である．

この結果は弱い攻撃者に多くの情報が流出するという直感に反するものなので，そのことを導出するプライバシーの流出する差分プライバシーの評価式(1)の改善を試みた．

データベースDからデータレコードd_iを除いたものをD_{-i}とする．攻撃者が相関を知っているデータレコードの集合をK_iとする．当然，K_iはD_{-i}の部分集合である．また，fを質問とするときrは質問者に返される答えであり，差分プライバシーの場合は$f(D)+ Lap(1/\varepsilon)$である．ここで，次式で与えられるε-CBPと命名した評価基準を導入する．

$$\left|\log\frac{\Pr(d_i \mid r,K_i)}{\Pr(d_i \mid K_i)}-\log\frac{\Pr(d_i' \mid r,K_i)}{\Pr(d_i' \mid K_i)}\right|\le\varepsilon$$

この式は次のように書き換えられる．

$$CBP=\frac{\sum_{U_i}\Pr(r \mid d_i,D_{-i})\Pr(U_i \mid d_i,K_i)}{\sum_{U_i}\Pr(r \mid d_i',D_{-i})\Pr(U_i \mid d_i',K_i)}$$ ただし，

$$e^{-\varepsilon}\le CBP\le e^{\varepsilon}$$

$$U_i=D_i-K_i \qquad (2)$$

ここで一例として，加算する雑音$Lap(1/\varepsilon)$をε=0.1とし，d_1とd_2の相関が次の相関行列で与えられる場合を想定してみる．

$\Pr(d_1,d_2)$	d_1=0	d_1=1
d_2=0	0.49	0.01
d_2=1	0.01	0.49

ここで，d_2=1を知っている強い攻撃者の場合，以下の表のような条件付き確率を知っていることになる．

	d_1=0	d_1=1
Pr(d_1|r=2,d_2)	0.018	0.982
Pr(d_1|d_2)	0.02	0.98

この結果から強い攻撃者を想定して計算してみると，CBP≈$e^{-0.1}$　となる．
一方，d_2=1を知らない弱い攻撃者の場合に知られる条件付き確率は下の表のようになる．

	d_1=0	d_1=1
Pr(d_1|r=2)	0.45	0.55
Pr(d_1)	0.50	0.50

この結果から弱い攻撃者を想定して計算してみると，CBP≈$e^{-0.2}$　　となる．
つまり，弱い攻撃者のほうがCBPの下側の限界を小さくとる必要があり，雑音$Lap(1/\varepsilon)$のεは大きな値をとる必要がある．ということは同じ安全性すなわち情報の漏洩確率を確保するためには，加算される雑音の絶対値は多くしなければならない．換言すれば，CBPという評価尺度はデータレコードに相関がある場合は，より厳しいものとなり，直感に合致する．

次に加算する雑音の分布を決めるεを計算するために式(2)のCBPの最大値と最小値の計算が必要だが，総和Σを領域U_iにおいて計算する必要がある．はU_iはデータベース全体からK_iを差し引いた残りの部分であり，非常に大きいため，総和計算の量は莫大である．具体的には次式の計算量となる．
d_iのとりうる値の集合をD*，rがとりうる値の集合をR*，データベース内のデータレコード数をnとする．すると，解くべき式の数は

$$n \bullet |\mathrm{R}^*| \bullet |\mathrm{D}^*|^n \bullet 2^{n-1}$$

となり，計算の複雑さは禁止的である．したがって，問題を近似して計算の複雑さを減らすことを考える．

まず，d_iに相関のあるデータレコード集合U^dと相関のないデータレコードU^cに分離する．次にデータベース内の相関のあるデータレコード全体をKとするとき，d_iに関係する部分をK^d，関係しない部分をK^cとする．こうすると，近似計算は，以下のように分解して行える．

$$\varepsilon = \varepsilon_1 + \varepsilon_2$$

$$e^{-\varepsilon_1} \le \frac{A_i}{A_i'} \le e^{-\varepsilon_1} \quad e^{-\varepsilon_2} \le \frac{B_i}{B_i'} \le e^{-\varepsilon_2}$$

$$\Rightarrow e^{-\varepsilon} \le \frac{\sum_i A_i \cdot B_i}{\sum_i A_i' \cdot B_i'} \le e^{-\varepsilon} \quad (3)$$

ただし，式(3)の各要素は以下のように定義される．

$$A_i = \sum_{U^d} \left(\begin{array}{l} \Pr(r \mid d_i, D_{-i}) \\ \cdot \Pr(U^d \mid d_i, U^C, K^d, K^c) \end{array} \right)$$

$$A_i' = \sum_{U^d} \left(\begin{array}{l} \Pr(r \mid d_i', D_{-i}) \\ \cdot \Pr(U^d \mid d_i', U^C, K^d, K^c) \end{array} \right)$$

$$B_i = \sum_{U^c} \Pr(U^c \mid d_i, K^d, K^c)$$

$$B_i' = \sum_{U^c} \Pr(U^c \mid d_i', K^d, K^c)$$

式(3)の近似計算によって解くべき式の数は

$$n \bullet |\mathrm{R}^*|^3 \bullet |\mathrm{D}^*|^M \bullet 2^{M-1}$$

となる．ただし，Mはd_iに関係するデータレコード数であり，$n<<M$　であることが期待されるので，解くべき式の数は大きく減っている．ただし，d_iに関係するデータレコードを求める作業，K^d，K^cを求める作業が必要であり，これは一般的には解けない．解決できる特殊な場合として，1) d_iが隣接するd_{i-1}，d_{i+1}とだけしか相関しない場合および，　2) 相関関係をリンクとしたデータレコードが木構造をしている場合，について効率良くd_iに関係するデータレコード，K^d，K^cを求める方法を探し出したが，これらは非常に限定的な結果でしかないといえよう．

(2)差分プライバシーは誘発する風評被害

2010年以降，プライバシー保護データマイニングの研究分野では，その社会的影響に関する研究がヨーロッパでは盛んになり，奨学金獲得者の人選は公平に行われたどうかを評価するモデルの研究が進んできている．このアイデアを一歩進めてみたとき，差分プライバシーあるいはより実用に近いk-匿名化においてデータベースへの質問結果，ないしデータベースそのものに加えられる雑音が引き起こす悪影響を考慮すべきだという着想にいたった．悪影響としてはインターネットにおける風評被害や濡れ衣が考えられる．これについては学会発表④を行った．以下では差分プライバシーによる雑音加算が誘発する風評被害について詳述する．
質問者は差分プライバシーで得られた結果をどのように使おうと自由である．この状況で以下の 2 つのケースを考えてみよう．
評価が連続する場合

例えば，ある企業で出荷された製品 A で発見された不良品数が質問されたとしよう．月刊 10 万個が出荷され，不良品が 100 個だったところを，雑音加算によって 102 個としようが 99 個としようが，これが毎月類似に値であれば社会的にさほど大きな影響はでない．つまり，差分プライバシーによってその企業の評価，あるいは製品 A の評価はほとんど変動しない．もちろん，長期的な観測で不良品が増加傾向であれば，評価が下がり，減少傾向であれば評価は上がるが，これは想

定の範囲内の社会的評価であるといえよう．

評価が不連続な場合

ある食品業者の製品 B には多量に摂取すると危険な物質Cが含まれているが，安全基準D以下なら危険ではないとされているとしよう．危険な物質Cの含有量D以上の出荷ロット数を質問したとしよう．これをさらに 2 つの場合に分けて考える．ただし，雑音加算によって負の値になった場合は回答を 0 とする．

1) データベースに記載されていた物質Cの含有量D以上の出荷ロット数＝0 であったにもかかわらず，雑音加算によって 1 以上にという回答が出力されたとしよう．これは，この食品業者の社会的評価を著しく損なうことになる．いわゆる風評被害である．

同様のことは，レストランにおける食中毒の件数，パソコンやスマホのハードウェアやソフトウエアの不具合報告数などでも生じうることで，0 であったにもかかわらず，1 以上になると利用者が不必要に用心してしまい，レストランや病院にマイナス評価を与えかねない．

2)データベースに記載されていた物質Cの含有量D以上の出荷ロット数＝１であったにもかかわらず，雑音加算によって 0 という回答が出力されたとしよう．これは，この食品業者の実態を隠蔽する．（以後瑕疵隠蔽と呼ぶ．）

社会的影響という観点で問題になるのは上記 2)の場合である．通常の差分プライバシーでは

min (｜回答の誤差｜＋データベース内部の 1 個のデータの特定できる確率) (4)

という最適化の問題を考え，加算される雑音の分布である上記のラプラス分布のパラメタ λ を調整する．データマイニングにおいて最小化する目標となる精度は，式(4)の回答の誤差はその絶対値が最適化の対象になるため誤差自体の正負は考慮されない. しかし, 上記 2)の問題では，誤差の正負によって社会的影響の正負が非対称に異なってくる．以下で，この問題を扱う．

風評被害と瑕疵隠蔽のモデル化

上記 2)の状況を縦軸の正の方向に風評被害額，負の方向に瑕疵隠蔽の利得額:y をとり，横軸に物質Cの含有量D以上の出荷ロット数:x をとって図示したのが図 3 である．

図 3 で，d=0 の場合が，雑音加算のない場合の不良ロット数に対する風評被害額（瑕疵隠蔽利得額）を表す折れ線である．

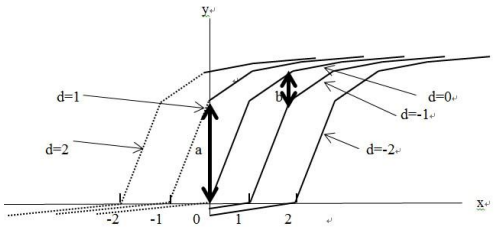


図 3　加算する雑音 d を変化させたときの物質Cの含有量D以上の出荷ロット数:x vs 風評被害額（瑕疵隠蔽利得額）:y

一方，x＜0 の領域は不良ロットが出荷されないので，上記の議論が当てはまらない．ただし，不良ロットがない状態で多数の出荷があればあるほど食品業者側に信用という利得が生ずる．この信用は不良品出荷に比べて，ロット当たりに換算すれば非常に小さな値と考えるのが適当であろうから，図では x<0 の部分の傾きは緩やかになる．

図 3 による分析により，データベースへの質問への回答に雑音が加算された場合の風評被害（瑕疵隠蔽利得）の構造が明らかになった．

風評被害と瑕疵隠蔽における差分プライバシーの設計

ここで, プライバシー保護という観点からは上記のデータベースへの質問への回答から，個別の食品業者が特定されないことが目標となる．食品衛生の観点からはこのような保護はよくないが，レストランの評判分析，商店などサービス業のサービス度合いに適用し，レストラン業界全体あるいはサービス業界全体の動向のデータマイニングを行うにあたっては，個別店舗の特定は好ましくない．このことを考慮した，風評被害（瑕疵隠蔽利得）まで考慮した雑音加算の最適化は以下の式となる．ただし，瑕疵隠蔽利得は消費者側の損失であるので, 両者を同列に扱うためには絶対値をとる必要があり，次式(5)となる．ただし，y(x)は真のデータベース質問への回答の値である．identify(d(λ))は λ をパラメタとするラプラス分布による雑音 d(λ)をパラメタとしたとき店舗などが特定できる確率である．今後は，このようなモデルの汎用化，精密化，あるいは評価実験を行うとともにこの議論が適用できる社会現象を網羅していく作業を行う予定である．

$$\lambda = \arg\min_{\lambda} \sum_{x,d(\lambda)} \begin{pmatrix} \left| y(x + Lap(\lambda)) \right| \\ + identity(Lap(\lambda)) \end{pmatrix} \quad (5)$$

(3)その他の活動

プライバシー保護データマイニングの位置づけや紹介という内容で著書の執筆（図書

①），および人工知能学会大会における下記のセッションのオーガナイズを行った．刊行を行った．

(1)中川 裕志，吉田 稔，佐久間 淳（オーガナイザ）： 第 26 回人工知能学会全国大会 オーガナイズドセッション OS-20「プライバシー保護データマイニング」の主催，山口県教育会館, 2012 年 6 月

(2)中川 裕志,佐久間 淳,神嶌 敏弘,荒井 ひろみ(オーガナイザ)：第 27 回人工知能学会全国大会,オーガナイズドセッション「OS-06 情報の保護と中立性に配慮したデータ分析」．富山,2013 年 6 月

２回とも参加者は５０名程度であった．また，プライバシー保護データマイニング関連の発表(学会発表①②)を行った．

①はプライバシー保護データマイニングのうち，ネットワークにおける利用者同士の通信履歴のようなネットワーク型のデータから，プライベートなデータの流出を防いだうえで，データベースから EM アルゴリズムでデータマイニング手法である．②は複数のデータ保持者たちが，自分のデータを秘匿しつつ，準同型公開鍵暗号を用いて全員のデータを用いてデータマイニングを行う手法である．提案手法の特徴は，自分以外のデータ保持者たちが全員結託しても，自分のデータが暴露されないアルゴリズムを提案したことである．

また，本研究から得られた知見を応用した萌芽的課題として，4.⑵の風評被害をさらに発展させた濡れ衣現象に着目した課題（成果[その他]①）を見いだしたので，今後の課題として研究を続ける予定である．

５．主な発表論文等

〔雑誌論文〕（計２件）

① Yang Bin, Hiroshi Nakagawa. Privacy-Preserving EM Algorithm for Clustering on Social Network. PAKDD 2012, Part I, LNAI ,Springer-Verlag. 7301, pp. 542-553. 2012

② 小宮山 純平，佐藤 一誠，中川 裕志: ロックアップ期間による制約を考慮した確率的バンディット問題，情報処理学会論文誌数理モデル化と応用（TOM）,6(3), pp. 11-22 2013 年 12 月

〔学会発表〕（計 3 件）

① 楊 斌，佐藤 一誠，中川 裕志：Secure Clustering in Private Networks. 第 26 回人工知能学会全国大会 オーガナイズドセッション「プライバシー保護データマイニング」OS-20-2. 山口県教育会館. 山口県，山口市，2012 年 6 月 13 日

② 楊 斌，中川 裕志：Collusion-Resistant Privacy-Preserving Data Mining. 第 27 回人工知能学会全国大会, 3L3-0S-06b-1, 富山国際会議場，富山県，富山市，2013 年 6 月 6 日

③ 中川 裕志，角野 為耶: 滞在場所の k-匿名化と濡れ衣．情報処理学会.第 62 電子化知的財産・社会基盤研究発表会(EIP 研究会) Vol.2013-EIP-62, No.12． 東京工芸大学，東京都，中野区，2013 年 11 月 21 日

〔図書〕（計１件）

① 中川 裕志： 情報法，（宇賀克也，長谷部恭男 編：第 8 章 データベースサービスとコンテンツ），pp.133-159，有斐閣，2012 年 9 月

〔産業財産権〕

○出願状況（計０件）

○取得状況（計０件）

〔その他〕

① SlideShare のアップロードコンテンツ K匿名化と濡れ衣． 2014 年 2 月 2014.4.5 時点での閲覧数 :1584. http://www.slideshare.net/hirsoshnakagawa3/k-31921914

６．研究組織

(1)研究代表者
中川 裕志（Nakagawa， Hiroshi）
東京大学・情報基盤センター・教授
研究者番号：２０１３４８９３

(2)研究分担者
佐藤 一誠（Sato， Issei）
東京大学・情報基盤センター・助教
研究者番号：９０６１０１５５