

購買履歴情報を二次利用するためのプライバシー保護技術の研究

長村 春紀†

谷口 洋司‡

矢島 敬士†

†東京電機大学

120-8551 東京都足立区千住旭町 5 番

12fi039@ms.dendai.ac.jp, yajima@im.dendai.ac.jp

‡第一工業大学

110-0005 東京都台東区上野 7-7-4

y.taniguchi@ueno.daiichi-koudai.ac.jp

あらまし ポイントカード管理会社は、顧客の購買履歴情報を二次利用する際、匿名化などのデータ加工を行い、顧客のプライバシー保護を心がけている。しかし、顧客自身の判断で SNS(Social Networking Service)上に公開された一部の個人情報と、二次利用目的で匿名化された購買履歴情報が結び付けられることにより、顧客のプライバシーが侵害される可能性がある。本研究では、二次利用する購買履歴情報を作成する場合に、公開してもよいと判断される情報(ノーマル情報)と公開したくない情報(センシティブ情報)に分類した上で、 k -匿名性や l -多様性を考慮した匿名化を行うことにより、SNS上の情報と結び付けられにくくする方式を提案する。

Privacy Preserving for Secondary Use of Buying History Information

Haruki Osamura†

Yoji Taniguchi‡

Hiroshi Yajima†

†Tokyo Denki University

5 Senju Asahi-cho, Adachi-ku, Tokyo 120-8551 JAPAN

12fi039@ms.dendai.ac.jp, yajima@im.dendai.ac.jp

‡Daiichi Institute of Technology

7-7-4 Ueno, Taitou-ku, Tokyo 110-0005, JAPAN

y.taniguchi@ueno.daiichi-koudai.ac.jp

Abstract When loyalty card management companies use their customer's Buying History Information secondarily, they try to preserve the customer's privacy. However, the information which a customer opened on the SNS by himself and the personal information anonymized for secondary use are able to be combined, and his privacy may be invaded. When Buying History Information is made for secondary use, we classify it into the information which can be opened (normal information) and the information which can not to be opened (sensitive information), and we anonymize them by K -anonymization and L -Diversity.

1 背景

個人情報の保護を図りつつ、パーソナルデータの利活用を促進することによる、新産業・新サービスの創出と国の安全・安心の向上の実現及びマイナンバーの利用事務拡充のために、個人情報保護法(平成15年施行)の改正案が国会に提出されている[1]。しかしながら、個人を特定しないよう加工して第三者提供しても、SNS等の普及によりインターネット上に流通する膨大なパーソナルデータと突合することで個人を特定出来る可能性が増大している[2]。

2 研究の目的

コンビニエンスストアやレンタルビデオ店などにおける購買履歴情報を管理しているポイントカード管理会社などが、プライバシーを保護しつつ購買履歴情報を有効活用(二次利用)するために、SNSなどで公開された情報と結び付けられにくくする方式を提案する。

3 購買履歴二次利用上のプライバシー問題

3.1 ポイントカード管理会社のしくみ

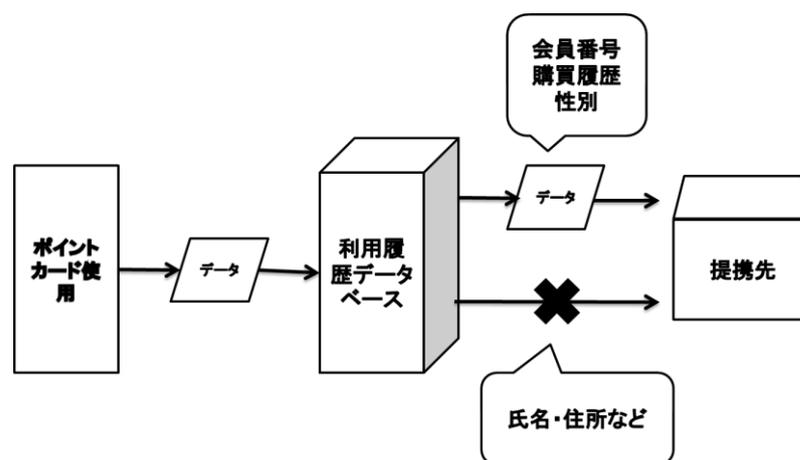


図1 ポイントカード管理会社による個人情報の取り扱い

従来より流通などの店舗はスタンプカードを発行し、割引サービスをするなど、消費者に囲い込みを図ってきた。昨今ではあらかじめ顧客の属性(性別、年齢、住所など)を登録し、POS(Point of Sales)データと結び付けられるようにしたポイントカードが普及している。消費者は購買額に応じてポイントを獲得し、それに応じたサービスを受けられるだけでなく、店舗やメーカーは購買履歴情報を分析し、顧客の固定化や売り上げ向上に活用している。ポイントカード管理会社は流通業を中心とした様々な企業と提携して購買履歴情報を収集し、ポイントカード利用履歴データベースで集中管理している。ポイントカード管理会社がデータを第三者機関(提携企業)に渡すのは、氏名や住所などの個人情報は削除し、会員ID、性別、年齢などの属性および購買履歴情報のみである[3]。

3.2 従来技術(k-匿名化)の適用

上記、第三者機関に渡されるデータは氏名・住所などを削除するという基本的な匿名化処理がなされる。さらに、SNS などの公開情報との突合によるプライバシーの侵害リスクを減らすために、k-匿名化の適用が図られている[4]。

例えば、以下の2つの表では、表1はk-匿名化を行う前のデータ、表2はk-匿名化した後のデータである。

本事例では、プライバシーとして隠しておきたいセンシティブな情報は「タバコ(の購入履歴)」とする。この時、匿名化を行う前にある人物 A さんが SNS で「X 社パン①と H 社飲料①おいしかった」などとつぶやいてしまうと、表1には一人しか存在せず、瞬時に個人を特定されてしまい(30 歳女性と分かる)喫煙者だとわかってしまう。従って、表2のように年齢を曖昧化、購入履歴を一般化へ加工し、同じ属性が k 人以上となるようにする。このように同じ属性が k 人以上いる状態を k-匿名性といい、そのようにデータを加工することを「k-匿名化」とよんでいる。

表2の場合では k=2 すなわち 2-匿名化である。2-匿名化することによって、表2では A さんがタバコを購入したかどうかを特定することはできない。つまりタバコ購入確率は 1/2 である。

表1 k-匿名化前のデータ

年齢	性別	購買履歴		
30	女性	X 社パン①	H 社飲料①	A 社タバコ①
34	男性	X 社パン②	I 社飲料②	S 社雑誌②
38	女性	Y 社パン①	J 社飲料①	Z 社ガム①
45	男性	Z 社パン①		B 社タバコ②
47	女性		H 社飲料③	C 社タバコ①
49	女性		J 社飲料④	

表2 k-匿名化後のデータ(2-匿名化)

年齢	性別	購買履歴		
30 代	女性	パン	飲料	タバコ
30 代	男性	パン	飲料	雑誌
30 代	女性	パン	飲料	ガム
40 代	男性	パン		タバコ
40 代	女性		飲料	タバコ
40 代	女性		飲料	

3.3 従来技術の課題

k-匿名化を行うと、プライバシー侵害リスクを回避できる一方で、情報損失が発生し、マーケティング価値が下がることとなる。本事例では、タバコメーカーにとって貴重となる、購買されたタバコの銘柄情報が欠落してしまう。ポイントカード管理会社が第三者機関(例えばタバコメーカー)に渡すデー

タはマーケティング価値をできる限り下げることなくプライバシー侵害リスクを回避できる必要がある。

4 提案方式

4.1 提案方式の考え方

本研究では、プライバシー侵害に繋がるセンシティブな情報は SNS などで公開されにくいと考えた。日本政府主催のパーソナルデータに関する検討会において、センシティブな情報としては、政治的見解、信教、労働組合への加盟、人種及び民族、門地及び本籍地、保険医療及び性生活、犯罪歴、社会的身分、収入記録、その他の社会的差別の原因となるおそれのある事項などが挙げられている [2]。このようなセンシティブな情報は本人自ら SNS などで公開することは稀であると考えられる。

提案方式では、ポイントカード管理会社が扱う購買履歴情報の中でセンシティブな情報と、むしろ SNS などで公開されやすい情報を分類し、公開されやすい情報に関しては突合され個人特定につながりやすいと考え、k-匿名化を行う。

4.2 提案方式

提案方式のフローを以下に示す(図2)。

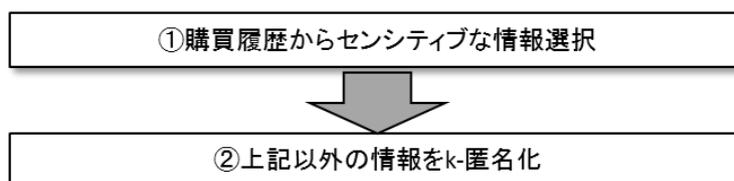


図2 提案方式のフロー

①購買履歴からセンシティブな情報選択

表3にポイントカード管理会社の購買履歴データ項目の例を示す。これらから氏名・住所・電話番号などは削除され、会員番号、性別、年齢、購買商品などが購買履歴情報として第三者機関に渡される。

購買商品の中でプライバシーの侵害に関わるセンシティブな情報を選択する。例えば、保険医療及び性生活に関連する商品や、社会的差別の原因となるおそれのある商品などである。具体的には、避妊具、アダルト関連雑誌、タバコ(女性)、特別な趣味関連商品などが考えられる。

②上記以外の情報を k-匿名化

センシティブな情報以外の購買商品や年齢、性別に関して k-匿名化を行う。

各購買商品の k-匿名化に関しては、商品の分類コード[5]に基づくレベルを定義し、必要に応じて一般化(レベルを変更)することにより、k-匿名性を確認する。表4に購買商品のレベル化の例を示す。

k-匿名化後の購買履歴情報について、センシティブな情報に関する 1-多様性についても確認する。つまり、センシティブな情報以外の購買商品や年齢、性別に関して k-匿名化を行っても、この k 人がいずれもセンシティブな購買履歴を含んでいる場合、プライバシーが侵害される可能性がある。その

ような状況を排除するために、k 人の中にセンシティブな購買履歴のある人とセンシティブな購買履歴のない人の両方を含んでいることを確認する。

表3 ポイントカード管理会社の購買履歴データ項目例

識別子					準識別子			購買履歴		
氏名	会員番号	住所詳細	電話番号	メールアドレス	年齢	性別	住所抽象	店	日時	購買商品

表4 購買商品のレベル化

レベル	具体例(商品名「加藤園うまいお茶」)
レベル1	加藤園うまいお茶
レベル2	加藤園緑茶
レベル3	緑茶飲料
レベル4	茶系飲料
レベル5	飲料

5 検証実験内容

5.1 検証データ

提案方式の有効性を検証するために、実際の購買履歴情報を用いた。検証データの内容を以下に示す。

- ・元データの種類： コンビニエンス・ストアのレシート
- ・データ収集方法： 研究室の同僚およびその家族から収集(1週間分)
- ・収集件数： 計170件(男性:120件, 女性:50件)
 - ⇒ 性別のみわかるように収集
 - ⇒ 男女は延べ件数(同一人物かどうか区別せず)

表5 検証データの一部(商品名を仮名に変換)

性別	購買履歴					
女	SP ブルーベリー	SP ブールパン	タコピラフおにぎり	SP もりだくさん五目いなり	リッチな銀のミルクキャンディ	
女	昭和おいしい牛乳	フェイスティッシュ				
男	スティックスイカパンチョコ	SP 手巻きおにぎりタコマヨ	SP いなりずし梅干し	SP カップみそ汁2種の野菜		
女	北アルプスデザート水	SP ソーセージおむすび	ダークワイルド メンソール BOX	アントニーボスカフェオレ		
男	紫野菜のグラタン	世界の台所色んなライチ	出来立てコロッケロール	レーズンマスタード		
女	SP ミネラルサブライ水	カンメントライトメンソール	まいとーの鮮やかなプリン	SP ラグジュアリソフトティッシュ		
男	エイトスター					
女	ストールケーキ	林水バピーソフトクリーム風味	カントンミルクティー			
男	SP 冷奴セット	SP 焚火焼肉カルビ弁当	赤缶クラシック しらす入サケ			
男	昆布弁当	赤飯おこわおにぎり	しゃかりこチーズL	週刊中年ラン		
男	昭和いちごオレ	大石井コーヒー	SP カフェアイスコーヒーR			
女	ツーンボロ ホワイトメンソール	ホットグラスト8Box	粗おろし大根素麺	SP 辛子明太子	SP 半熟煮卵	満月ブルー25度

5.2 検証方法

検証データを分析した結果、想定していたセンシティブ情報(避妊具, アダルト関連雑誌, タバコ(女性), 特別な趣味関連商品など)の中で「女性によるタバコ購入」が散見されたので、それをセンシティブ情報として選択した。

検証データに含まれる女性データ50件のうちタバコ購買履歴ありは14件であった。タバコを購入する女性の中で、自動販売機でタバコを購入する女性は12.6%である[6]との報告もあり、タバコをコンビニエンス・ストアやスーパーマーケットなどで他の商品に紛れて購入している女性が多いことが想定できる。頻繁にタバコを購入していることがばれることにセンシティブな女性も多いと仮定した。

本研究では、上記検証データに関して以下の検証を行った。

- ・SNS で公開される情報
 - 公開者の実名
 - タバコ以外の1商品(具体的商品名あり)に関する感想など
- ・検証データの女性データ50件に対して上記提案方式を適用
 - センシティブな情報: 女性によるタバコ購入履歴
 - k-匿名化: k=2 とする
 - ⇒ k-匿名性を保つように必要に応じてタバコ以外の購買商品について一般化(レベル変更)を行う

6 検証実験結果

検証データに対して提案方式を適用した結果の一部を表6に示す。

センシティブな情報(タバコ)を含む購買履歴情報 14 件に含まれているタバコ以外の商品名をリストアップし、プライバシー保護のための k-匿名化により、それらが如何に一般化されたかを示しており、括弧内の数字は購入件数である。例えば、「北アルプスヨーグル」はタバコ非購入者にも購入されており、一般化されておらず(レベル1)、購買履歴情報としてマーケティング価値は高いままである。一方、「アントニーポスカフェオレ」はレベル3まで一般化され「コーヒー飲料」となっており、マーケティング価値は下がってしまっている。

表6 提案方式の適用結果の一部(商品名を仮名に変換)

レベル1	レベル2	レベル3
SP ミネラルサムライウォーター	SP ジュース系非炭酸飲料(4)	
SP いちごのむヨーグルト	SP ヨーグルト飲料(3)	
昭和P加速健康ゼリー	昭和栄養ドリンク	栄養ドリンク(5)
北アルプスヨーグル(2)		
SP 炭火ソーセージおむすび	SP おむすび(7)	
アントニーポスカフェオレ	アントニーコーヒー飲料	コーヒー飲料(5)
マイトーの爽やかプリン	マイトー プリン	プリン(2)
SP 焼きカレーパン	SP 菓子パン(7)	
SP ふわりメープルのパン	SP 菓子パン(7)	
タニーコーヒーブラック	タニーコーヒー飲料	コーヒー飲料(5)
アントニーグリーンたから(2)		
アントニーポスとけるカフェオレ	アントニーコーヒー飲料	コーヒー飲料(5)
SP 手巻明太マヨ	SP 手巻寿司(14)	

タバコを購入した 14 件のデータに含まれる購買商品の一般化の状況を以下に示す。

- ・購買商品総数： 36
- ・一般化
 - レベル1(一般化しなかった)： 8
 - レベル2： 18
 - レベル3： 10
 - レベル4： 0
 - レベル5： 0

7 考察

上記検証実験結果に示したように、実際のコンビニエンス・ストアの購買履歴情報に提案方式を適用することにより、購買履歴情報のマーケティング価値をできる限り下げないようにできることが確認できた。

具体的には、センシティブな情報自身は SNS など公開されにくいので、一般化せずに第三者に渡すことが可能である。また、センシティブな情報以外の購買履歴情報の中でも売れ筋と思われる商品は一般化しなくても k-匿名性と t-多様性が保たれているため、そのまま第三者に渡すことが可能である。

また、本研究の今後の課題を以下に示す。

- ・SNS で公開される商品情報を 2 件以上と仮定した場合の提案方式の有効性確認
- ・検証データを大規模化(1,000 件以上)した場合の検証実験
- ・コンビニエンス・ストアだけでなく、ドラッグストアやレンタルビデオ店など様々な所での検証
- ・他のセンシティブな情報(タバコだけでなく、避妊具やアダルトビデオなど)での検証
- ・提案方式のフローの自動化(プログラム作成など)
- ・購買履歴だけでなく、他分野(医療介護分野など)への展開

参考文献

- [1] 個人情報保護に関する法律及び行政手続における特定の個人を識別するための番号の利用等に関する法律の一部を改正する法律案(概要)
- [2] 日本政府会議パーソナルデータに関する検討会資料
<https://www.kantei.go.jp/jp/singi/it2/pd/index.html>
- [3] カルチュア・コンビニエンス・クラブ T 会員規約改定
<http://www.ccc.co.jp/customer/optout.html>
- [4] 廣田啓一, 保木野昌稔, 藤木由里, 松崎和賢, 吉田圭吾, 赤井健一郎, 高橋克巳, 白井康之, 山口利恵: 情報大航海プロジェクトにおける個人情報匿名化基盤の構築と検証, 情報処理学会研究報告, Vol.2010-CSEC-50, No.48, pp1-12 (2010).
- [5] 日本標準商品分類(総務省)
http://www.soumu.go.jp/main_content/000294482.pdf
- [6] taspo(タスポ)所有状況と購入チャネル変化の実態: 株式会社インテージ
<http://www.intage.co.jp/library/20080929/>