6W-02

データ提供時のプライバシ設定の適正度を定量化する指標の提案

萱場　啓太 [†1]　生出 拓馬 [†1,†2]　阿部 亨 [†1,†3]　菅沼 拓夫 [†1,†3]

[†1] 東北大学大学院情報科学研究科　[†2] 日本学術振興会特別研究員 DC
[†3] 東北大学サイバーサイエンスセンター

1　はじめに

IoT 技術等の発展により，利用者に関するデータを利活用するサービス（位置情報を利用したナビゲーションサービスなど）が急増している．こういったサービスの多くには，データ提供に同意するか否かを利用者自身が選択できるプライバシ設定機能が実装されているため，利用者がプライバシ設定を行う機会は今後さらに増えると考えられる．このとき利用者は，サービスから提供を要求されたデータ項目ごとのプライバシ設定を，データ提供に関する独自のルール（データ提供ポリシ）に基づき行うと考えられるが（図 1），プライバシ設定の機会が増えることで，利用者がデータ提供ポリシに反する誤ったプライバシ設定を行ってしまう可能性が高まると予想される．

本研究の目的は，利用者が自身のデータ提供ポリシに基づく適正なプライバシ設定を行えるよう支援する手法の実現である．本稿では，この手法を実現するため，利用者のプライバシ設定が適正である程度（適正度）を定量化する指標を提案する．これにより，プライバシ設定に関する利用者の意思を反映した適正プライバシ設定支援手法の実現が期待できる．

2　関連研究

様々な利用者が行ったプライバシ設定のデータセットをもとに協調フィルタリングやサポートベクトルマシンを利用することで，利用者のデータ提供ポリシに基づくプライバシ設定を推薦する手法が提案されている [1, 2]．しかしこれらの手法は，(i) 推薦したプライバシ設定（推薦プライバシ設定）の適用の可否を決定するための判断基準を示していないため，利用者がこの可否決定に自身の意思を反映することが困難であると考えられる．また，(ii) 推薦プライバシ設定適用の同意を利用者から度々取得しており，この同意取得手続きが利用者にとって負担になると考えられる．

これらの課題に対し，本稿では，利用者がこれまでに行ったプライバシ設定に基づき利用者のプライバシ設定の適正度を定量化する指標を提案する．

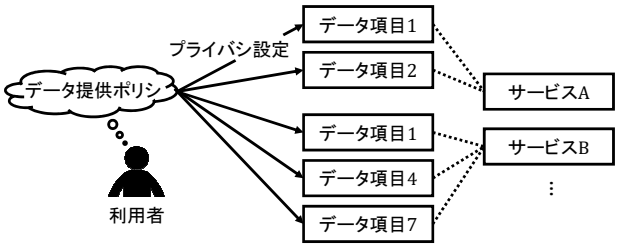

図 1　データ提供ポリシに基づくプライバシ設定

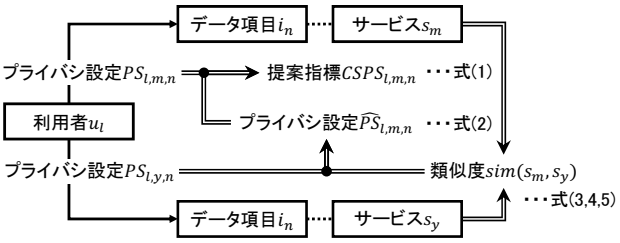

図 2　提案指標の算出手法の概要

3　提案

提案指標の算出手法の概要を図 2 に示す．利用者は，利用するサービスから要求されたデータ項目に対するプライバシ設定（データ提供の可否）を自由に変更できるものとする．以下，利用者の集合を $U = \{u_l \mid l = 1, 2, ..., L\}$，サービスの集合を $S = \{s_m \mid m = 1, 2, ..., M\}$，サービスが利用者へ要求するデータ項目の集合を $I = \{i_n \mid n = 1, 2, ..., N\}$ と各々表す．

利用者 u_l がサービス s_m から要求されたデータ項目 i_n に対して行うプライバシ設定を $PS_{l,m,n} \in \{0 = \text{提供拒否}, 1 = \text{提供許可}\}$，$u_l$ にとって適正であると推測されるプライバシ設定を $\widehat{PS}_{l,m,n}$ とし，プライバシ設定の適正度を定量化する提案指標 $CSPS_{l,m,n}$ を式 (1) で求める．

$$CSPS_{l,m,n} = 1 - |\widehat{PS}_{l,m,n} - PS_{l,m,n}| \quad (1)$$

$CSPS_{l,m,n}$ は，$\widehat{PS}_{l,m,n}$ と $PS_{l,m,n}$ の差が小さいほど高い値となる．これは，利用者の行うプライバシ設定が，適正であると推測されるプライバシ設定と近しいほど，利用者のプライバシ設定が適正であることを意味する．

本稿では，$\widehat{PS}_{l,m,n}$ の算出に，ユークリッド距離によるアイテム間類似度を基に利用者が各アイテム

A Proposal of a Measurement to Quantify Conformance of Sensible Privacy Settings
Keita KAYABA[†1], Takuma OIDE[†1,†2], Toru ABE[†1,†3], and Takuo SUGANUMA[†1,†3]
[†1]Graduate School of Information Sciences, Tohoku University
[†2]Research Fellow of Japan Society for the Promotion of Science
[†3]Cyberscience Center, Tohoku University

に対して行う評価を推測する手法 [3] を用いる．具体的には，$\widehat{PS}_{l,m,n}$ を式 (2) で求める．

$$\widehat{PS}_{l,m,n} = \frac{\sum_{s_y \in S^{(l,n)}} sim(s_m, s_y) \cdot PS_{l,y,n}}{\sum_{s_y \in S^{(l,n)}} sim(s_m, s_y)} \quad (2)$$

ここで，$S^{(l,n)} \subseteq S$ は，利用者 u_l がデータ項目 i_n に対しプライバシ設定を既に行ったサービスの集合を表し，$sim(s_m, s_y)$ は，$s_m \notin S^{(l,n)}$ と $s_y \in S^{(l,n)}$ の類似度を表す．$\widehat{PS}_{l,m,n}$ は，利用者 u_l がデータ項目 i_n に対し既に行ったプライバシ設定にサービス間の類似度で重み付けした加重平均である．これにより，s_m との類似度が高いサービスに対して利用者が行ったプライバシ設定が $\widehat{PS}_{l,m,n}$ に強く反映される．

類似度 $sim(s_m, s_y)$ は式 (3) で求める．

$$sim(s_m, s_y) = \frac{\alpha}{1 + d(s_m, s_y)} \quad (3)$$

ここで，$d(s_m, s_y)$ は，利用者のプライバシ設定から求めた，サービス s_m, s_y 間のユークリッド距離を表す．サービス s_m, s_y に対しプライバシ設定を行った利用者の集合を各々 $U^{(m)}, U^{(y)}$ で，s_m, s_y が利用者に要求するデータ項目の集合を各々 $I^{(m)}, I^{(y)}$ で表し，$U^{(m \cap y)} = U^{(m)} \cap U^{(y)}, I^{(m \cap y)} = I^{(m)} \cap I^{(y)}$ とすると，$d(s_m, s_y)$ は式 (4) で求められる．

$$d(s_m, s_y) = \sqrt{\sum_{u_x \in U^{(m \cap y)}} \sum_{i_z \in I^{(m \cap y)}} (PS_{x,m,z} - PS_{x,y,z})^2} \quad (4)$$

また，式 (3) 中の α は，$I^{(m)}$ と $I^{(y)}$ の Jaccard 係数であり，式 (5) で表される．

$$\alpha = \frac{| I^{(m)} \cap I^{(y)} |}{| I^{(m)} \cup I^{(y)} |} \quad (5)$$

これにより，$| I^{(m)} \cup I^{(y)} |$ に対する $| I^{(m)} \cap I^{(y)} |$ の割合をサービス間の類似度の上限値とする．

4 考察

4.1 提案指標の有用性

(1) 推薦設定適用同意の判断基準の提供: 利用者は，推薦プライバシ設定の適用の可否を決めるための判断基準の一つとして，提案指標を参考にできる．これにより，推薦プライバシ設定の適用の可否決定に利用者の意思を反映できると考えられ，(i) の解決が期待される．

具体的には，プライバシ設定の推薦時に提案指標を提示することで，「推薦プライバシ設定を適用しますか？（適用前の適正度：0.1，適用後の適正度：0.9）」というように，推薦プライバシ設定適用の同意のための一つの判断基準を適正度という形で利用者に提供できる．

(2) 推薦設定適用同意の自動化: 利用者は，提案指標に対する閾値を自ら設定することで，推薦プライバシ設定の適用に関する同意の自動化を自らの意思を反映して行うことができる．これにより，推薦プライバシ設定適用の同意に対する利用者の負担を軽減できると考えられ，(ii) の解決が期待される．

現状のプライバシ設定推薦手法は，推薦プライバシ設定適用の同意を利用者から取得する手続きの自動化を基本的には行っていない．この一因として，利用者の意思を適切に反映した同意の自動化が困難であることが考えられる．例えば，[1, 2] に関して，「同意取得手続きの省略に同意すると，推薦プライバシ設定の適用に自動的に同意する」という同意自動化手法が考えられるが，この手法で利用者が選択できるのは同意取得手続きの省略に同意するか否かだけであり，利用者の選択肢の自由度が低いため，利用者の意思を適切に反映した同意の自動化が行えるとは言い難い．

これに対し，推薦プライバシ設定適用の同意の取得に提案指標を用いて，「利用者があらかじめ設定した適正度 0.1 を下回るプライバシ設定は，推薦プライバシ設定の適用に自動的に同意する」という同意自動化手法を考える．この手法で利用者が選択できるのは提案指標に対する閾値であり，利用者の選択肢の自由度が高いため，利用者の意思を適切に反映した同意の自動化が行えると考える．

4.2 提案指標の課題

(1) プライバシ設定推測の精度向上: 提案指標の精度は，適正なプライバシ設定の推測手法の精度に依存する．本稿で用いた単純な推測手法の精度は，疎なプライバシ設定データセットにおいて高くないため，そのようなデータセットにおいても高い精度を維持できるよう推測手法を改善する必要がある．

(2) 直感的理解の支援: 提案指標を数値のまま利用者に提示する方法では，その数値の意味を利用者が直感的に理解することは難しい．したがって，利用者が提案指標の意味を直感的に理解できるようなGUI を実装するなどの工夫が求められる．

5 おわりに

本稿では，利用者のプライバシ設定が適正である程度を定量化する指標を提案した．また，提案指標の有用性や課題について考察した．今後は，考察で述べた課題の解決，および提案指標を用いた適正プライバシ設定支援機構の構築を行う．

参考文献

[1] Liu, R., et al.: PriWe: Recommendation for Privacy Settings of Mobile Apps Based on Crowdsourced Users' Expectations, *Proc. of IEEE MS 2015*, pp. 150–157 (2015).

[2] Liu, B., et al.: Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions, *Proc. of SOUPS 2016*, pp. 27–41 (2016).

[3] Segaran, T.: Programming Collective Intelligence: Building Smart Web 2.0 Applications, Oreilly & Associates Inc (2007). 當山仁健，鴨澤眞夫（訳）：集合知プログラミング，オライリー・ジャパン（2008）．