Customized Privacy Preserving for Classification Based Applications

Zaobo He, Zhipeng Cai, Yingshu Li

Department of Computer Science
Georgia State University

Workshop on Privacy-Aware Mobile Computing (PAMCO) 2016
Classification-based Applications (CBS)

Sharing Profile with Recommender

- Learn User Profile
  - IMDb

- Uploading Age, Gender, Location, Movie genre, ...

Sharing Profile with Healthcare System

- Uploading Glucose, BMI, Pressure, Steps ...

E.g., Content Based Recommendation (Logistic Regression, SVM, Decision tree etc), Collaborative filtering, Matrix Factorization
Example: IMDB Profile Loading

Uploading Age, Gender, Location, Movie genre, Review..., get recommendation

Imdb.com reaches over 12 million U.S. monthly people
Source: https://www.quantcast.com/imdb.com
Threat

Profile privacy is not only to preserve the privacy of sensitive attributes in profiles.

Malicious applications can mine sensitive information from profiles just in classification process.

The contextual profiles attached to an user tell much about his habits, interests, activities, and relationships.
Two Types of Privacy

- **Inherent-data Privacy**
  - Protect the privacy of sensitive attributes in user’s profile
  - E.g., $X = \{\text{age, gender, location}\}$

- **Latent-data Privacy**
  - Latent sensitive information which are not expected to be exposed in classification process
  - E.g., classification results: pornographic movies, AIDS

- **Service Quality**
  - Service-related latent information which are expected to be inferred accurately in classification process
Privacy Protection Mechanisms

• **Anonymization** (removing the user’s identity)
  – It has been shown inadequate, as a single defense
  – The profiles can be de-anonymized, given an adversary with some knowledge on the users

• **Obfuscation** (reporting a fake profile)
  – Service Quality?
  – Users share their locations to receive some services back. Obfuscation degrades the service quality in favor of profile privacy
Designing a Protection Mechanism

• **Challenges**
  – Respect users’ required service quality
  – Inherent-data privacy protection
  – Latent-data privacy protection
  – Real-time protection

• **Common Pitfall**
  – Disregard either latent or inherent data privacy
  – Ignore adversary knowledge
    • Adversary can invert the obfuscation mechanism
  – Disregard optimal attack
    • Given a protection mechanism, attacker designs an attack to minimize his estimation error in his inference attack
Our Objective:
Design Optimal Protection Strategy
Considering both Inherent-data and Latent-data Privacy

A defense mechanism that
• **anticipates** the attacks that can happen against it,
• and **maximizes** the users’ service quality against the most effective attack,
• and respects the users’ **privacy** constraints against the most effective attack.
Outline

• Assumptions
• Model
  – Service Quality and Privacy Metrics
  – Inherent-data Privacy
  – Latent Inference Attack
• Problem Statement
• Solution: Optimal strategy for user
• Evaluation
Assumptions

• CBS: Sporadic Profile Exposure
  – Movie recommendation, healthcare monitoring, ...

• Adversary: Service provider
  – Or any entity who eavesdrops on the users’ CBS accesses

• Attack: Distinguishing and Inference Attack
  – What is the user’s sensitive attributes and latent information in carried by profiles when accessing LBS?

• Protection: User-centric obfuscation mechanism
  – So, we focus on a single user
Metrics

- **Service Quality Metric:**
  - Expected classification-error between real profiles and fake profiles

- **Inherent-data Privacy Metric:**
  - The **indistinguishability** of user’s sensitive attributes

- **Latent-data Privacy Metric:**
  - Adversary’s expected error in estimating the user’s latent information, given the user’s access pattern and observed fake-profile
Adversary Knowledge: User’s “Access Pattern”

$$\psi(X)$$

Probability with profile $X$ when accessing the CBS, $X = (c_1, \ldots, c_m, s_{t+1}, \ldots, s_n)$

- Non-sensitive attributes
- Sensitive attributes
Profile Obfuscation Mechanism

Probability of replacing profile $X = (c_1, \ldots, c_m, s_{t+1}, \ldots, s_n)$ with fake-profile $X = (c_1, \ldots, c_m, s'_{t+1}, \ldots, s'_n)$

Consequence: "Service Quality Loss"

$$Q_{loss}(f(.), \psi(X), d_q) = \sum_{X, X'} \psi(X)f(X'|X)d_q(X, X')$$

quality loss due to replacing $X$ with $X'$
Inherent-data Privacy

$d_\chi$-differential privacy: $f(X'|X)$ produces similar fake profiles for similar real profiles

$\epsilon d_\chi$-inherent-data privacy: an obfuscation mechanism $f(X'|X)$ satisfies $\epsilon d_\chi$-inherent-data privacy iff:

$$f(X'|X_i) \leq e^{\epsilon d_\chi(X_i,X_j)} f(X'|X_j) \quad X_i, X_j, X' \in \chi, i \neq j$$

“$\epsilon$-inherent-data privacy”
Latent Inference Attack

Posterior probability, given observed fake profile $X'$

$$Pr(X|X') = \frac{Pr(X, X')}{Pr(X')} = \frac{\int f(X'|X)\,\psi(X)}{\sum_X f(X'|X)\,\psi(X)}$$

Inference error of sensitive information due to guessing $X$ with $\hat{X}$

$$\min_{\hat{X}} \sum_X Pr(X|X')d_p(X, \hat{X})$$

User’s conditional expected privacy given $\hat{X}$

$$\sum_{X'} Pr(X')\min_{\hat{X}} \sum_X Pr(X|X')d_p(X, \hat{X})$$

User’s unconditional expected privacy (averaged over all $X'$)

$$\sum_{X'} \min_{\hat{X}} \sum_X \psi(X)f(X'|X)d_p(X, \hat{X}) \geq \delta$$

"$\delta$-latent-data privacy"
Problem Statement

- Given, the user’s profile $\psi(X)$ known to adversary

- Find obfuscation function $f(X'|X)$ that
  - Maximizes service quality, according to distortion
  - Respects inherent-data and latent-data privacy bounds $\varepsilon$ and $\delta$

- Adversary observes $X'$, and finds optimal $\hat{X}$ to minimize the user’s privacy who uses $f(X'|X)$
Optimal Strategy for the User

Minimize:

\[
\sum_{X, X' \in \chi} \psi(X) f(X'|X) \sum_{k_i \in G} d_{q_i}^{k_i}(X, X')
\]

Subject to:

\[
f(X'|X_i) \leq e^{\epsilon d_{\chi}(X_i, X_j)} f(X'|X_j)
\quad X_i, X_j, X' \in \chi, i \neq j
\]

\[
\delta \text{-latent-data privacy}
\]

\[
\sum_{X'} \min_{\hat{X}} \sum_{X} \psi(X) f(X'|X) \sum_{p_i \in G} d_{p_i}^{p_i}(X, \hat{X}) \geq \delta
\]

\[
\sum_{X' \in \chi} f(X'|X) = 1, \quad X \in \chi
\]

\[
f(X'|X) \geq 0 \quad X, X' \in \chi
\]

Proper probability distribution

\(\epsilon\)-inherent-data privacy
Algorithm 1 $(\epsilon, \delta)$-privacy preserving with optimal service quality

**Input:** user’s access pattern $\psi(.)$, attribute vector $X = (c_1, \ldots, c_t; s_1, \ldots, s_{t+1})$, inherent-privacy level $\epsilon$ and latent-privacy level $\delta$.

**Output:** $O_f : f(X'|X) \in O_f$.

1. $O^l_f = \text{LaPriCheck}(\psi(.), X, \delta)$;
2. **for** each $X'$ with probability $f(X'|X) \in O^l_f$ **do**
   3. $O^i_f = \text{InPriCheck}(\psi(.), X, \epsilon, O^i_f)$;
   4. **end for**
5. $f(X'|X) = \arg \min_{f(X'|X)\in O^l_f} Q_{loss}(f(.), \psi(X), d_q)$;
6. $O_f = O_f \cup f(X'|X)$
7. Choose a pseudo-attribute vector $X'$ by sampling from $\hat{X}$ probability $f(X'|X) \in O_f$
8. **return** $O_f$
Evaluation: Obfuscation Function

- **Optimal**
  - Solve the linear optimization problem (using Matlab)

- **Basic (k-anonymity)**
  - Hide profile $X$ among the k-1 nearest profiles (with positive $\psi$ probability)
Evaluation

• Optimal attack against optimal obfuscation
  – Given tow privacy level constraints

• Bayesian attack against any obfuscation

\[ h(\hat{X} | X') = \frac{Pr(X, X')}{Pr(X')} = \frac{f(X'|X)\psi(X)}{\sum_X f(X'|X)\psi(X)} \]

• A consolidated inherent-data privacy metric

\[ \text{AdvErrIn}(f, \psi, d_p) = \min \sum_{\hat{X}, X', X} \psi(X)f(X'|X)h(\hat{X}|X')d(\hat{X}, X) \]
Inherent-data Privacy: Optimal vs. K-anonymity

Inherent-data privacy threshold is set to be same for Optimal and K-anonymity.
Latent-data Privacy: Optimal vs. K-anonymity

- A consolidated latent-data privacy metric

<table>
<thead>
<tr>
<th>Quality loss: $ql$</th>
<th>$\delta$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2516</td>
<td>0.71</td>
<td>85</td>
</tr>
<tr>
<td>3.2840</td>
<td>0.72</td>
<td>87</td>
</tr>
<tr>
<td>3.2925</td>
<td>0.73</td>
<td>88</td>
</tr>
<tr>
<td>3.2925</td>
<td>0.74</td>
<td>88</td>
</tr>
<tr>
<td>3.3020</td>
<td>0.75</td>
<td>88</td>
</tr>
<tr>
<td>3.3067</td>
<td>0.76</td>
<td>88</td>
</tr>
<tr>
<td>3.3091</td>
<td>0.77</td>
<td>89</td>
</tr>
<tr>
<td>3.4181</td>
<td>0.78</td>
<td>97</td>
</tr>
<tr>
<td>3.4491</td>
<td>0.79</td>
<td>97</td>
</tr>
<tr>
<td>3.5240</td>
<td>0.80</td>
<td>102</td>
</tr>
<tr>
<td>3.5748</td>
<td>0.81</td>
<td>104</td>
</tr>
<tr>
<td>3.6177</td>
<td>0.82</td>
<td>106</td>
</tr>
<tr>
<td>3.7198</td>
<td>0.83</td>
<td>110</td>
</tr>
<tr>
<td>3.7284</td>
<td>0.84</td>
<td>111</td>
</tr>
<tr>
<td>3.8722</td>
<td>0.85</td>
<td>112</td>
</tr>
</tbody>
</table>

Latent-data privacy threshold is set to be same for Optimal and k-anonymity.
Latent-data Privacy: Optimal vs. K-anonymity

Latent-data privacy threshold is set to be same for Optimal and k-anonymity.
Conclusion

• (Attribute) Privacy is an undisputable issue, with more people uploading their profile more regularly

• Disregarding adversary’s background knowledge, strategies, capabilities, limits the privacy protection

• Our mechanism achieve the best possible service quality with both inherent-data privacy and latent-data privacy preserved