Tailoring Privacy in Personalized Systems to User Preferences and Privacy Regulations

Alfred Kobsa
Institute for Software Research
University of California, Irvine

(Joint research with Yang Wang and Max Teltzrow)
**“Traditional” personalization on the World Wide Web**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tailored email alerts</td>
<td>64%</td>
</tr>
<tr>
<td>Customized content</td>
<td>48%</td>
</tr>
<tr>
<td>Account access</td>
<td>48%</td>
</tr>
<tr>
<td>Personal productivity tools</td>
<td>23%</td>
</tr>
<tr>
<td>Wish lists</td>
<td>23%</td>
</tr>
<tr>
<td>Product recommendations</td>
<td>23%</td>
</tr>
<tr>
<td>Saved links</td>
<td>20%</td>
</tr>
<tr>
<td>Express transactions</td>
<td>16%</td>
</tr>
<tr>
<td>Targeted marketing/advertising</td>
<td>11%</td>
</tr>
<tr>
<td>Custom pricing</td>
<td>9%</td>
</tr>
<tr>
<td>Personalized content through non-PC devices</td>
<td>7%</td>
</tr>
<tr>
<td>News clipping services</td>
<td>5%</td>
</tr>
</tbody>
</table>

Source: Forrester Research

---

**Hello, Kobsa, Alfred.** We have **DVD Recommendations** for you.

---

**Kobsa's Gold Box**

---

**My Recent Quotes**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Change</th>
<th>Last News</th>
<th>Charts</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM</td>
<td>+0.30</td>
<td>31.05</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>+4.23</td>
<td>135.30</td>
<td></td>
</tr>
<tr>
<td>MSFT</td>
<td>+0.51</td>
<td>28.81</td>
<td></td>
</tr>
<tr>
<td>$QSGDEV</td>
<td>+1.75</td>
<td>1,138.22</td>
<td></td>
</tr>
</tbody>
</table>

---

**Percent of 44 companies interviewed (multiple responses accepted)**
Recent deployments of personalization

• Personalized search
• Web courses that tailor their teaching strategy to each individual student
• Information and recommendations by portable devices that consider users’ location and habits
• Personalized news (on mobile devices)
• Product descriptions whose complexity is geared towards the presumed level of user expertise
• Tailored presentations that take into account the user’s preferences regarding product presentation and media types (text, graphics, video)
Current personalization methods (in 60 seconds)

Data sources
- Explicit user input
- User interaction logs

Methods
- Assignment to user groups
- Rule-based inferences
- Machine learning

Storage of data about users
- *Persistent* user profile
- Updated over time
Web personalization delivers benefits for both users and web vendors

**Jupiter Communications, 1998:** Personalization at 25 consumer e-commerce sites increased the number of new customers by 47% in the first year, and revenues by 52%.

**Nielsen NetRatings, 1999:**
- Registered visitors to portal sites spend over 3 times longer at their home portal than other users, and view 3 to 4 times more pages at their portal.
- E-commerce sites offering personalized services convert significantly more visitors into buyers than those that don’t.

**Choicestream 2004, 2005:**
- 80% interested in personalized content
- 60% willing to spend a least 2 minutes answering questions about themselves

**Downside:**
*Personalized sites collect significantly more personal data than regular websites, and do this often in a very inconspicuous manner.*
Many computer users are concerned about their privacy online

Number of users who reported:

- being extremely or very concerned about divulging personal information online:
  67% (Forrester 1999), 74% (AARP 2000)
- being (extremely) concerned about being tracked online:
  77% (AARP 2000)
- leaving web sites that require registration information:
  41% (Boston Consulting 1997)
- having entered fake registration information:
  40% (GVU 1998), 27% (Boston Consulting 1997), 32% (Forrester 1999)
- having refrained from shopping online due to privacy concerns, or bought less:
  32% (Forrester 1999), 32% (GVU 1998), 35% (Forrester 1999), 54% (IBM 1999), 24% (AARP 2000)
- wanting internet sites ask for permission to use personal data: 81% (Pew 2000)
- being willing to give out personal data for getting something valuable in return:
  31% (GVU 1998), 30% (Forrester 99), 51% (Personalization Consortium)
Privacy surveys do not predict people’s privacy-related actions very well

Surveys generally, and privacy surveys in particular, suffer from the “talk is cheap” problem. It costs a consumer nothing to express a desire for a law to protect privacy. After all, who would not state that he is “concerned” in some sense about privacy?

Harper and Singleton, 2001
Personalization Consortium

• In several privacy studies in E-commerce contexts, discrepancies have already been observed between users stating high privacy concerns but subsequently disclosing personal data carelessly.

• Several authors therefore challenge the genuineness of such reported privacy attitudes and emphasize the need for experiments that allow for an observation of actual online disclosure behavior.
Either Personalization or Privacy?

Personal data of computer users are indispensable for personalized interaction.

Computer users are reluctant to give out personal data.

Tradeoff between privacy and personalization?
The tension between privacy and personalization is more complex than that...

- Indirect relationship between privacy and personalization
- Situation-dependent
- Many mitigating factors

People use complex “privacy calculus” to decide whether or not to disclose personal data, e.g. for personalization purposes.
Privacy-Enhanced Personalization

Can we have good personalization and good privacy at the same time?

How can personalized systems maximize their personalization benefits, while at the same time being compliant with the privacy constraints that are in effect?
Privacy constraints

A. People’s privacy preferences in a given situation (and factors that influence them)
B. Privacy norms (laws, self-regulation, principles)

Reconciliation of privacy and personalization

1. Use of privacy-enhancing technology
2. Privacy-minded user interaction design
Privacy norms

• Privacy laws
  More than 40 countries worldwide

• Industry self-regulations
  Companies, industry sectors (NAI)

• Privacy principles
  – supra-national (OECD, APEC)
  – national (Australia, Canada, New Zealand…)
  – member organizations (ACM)

Several privacy norms disallow a number of frequently used personalization methods (unless the user’s consents to them)
Privacy laws and regulations restrict the permissibility of personalization methods

- Usage logs must be deleted after each session
- Usage logs of different services may not be combined (except for accounting purposes)
- User profiles are permissible only if pseudonyms are used. (Profiles retrievable under pseudonyms shall not be combined with data relating to the bearer of the pseudonym.)
- No fully automated individual decisions are allowed that produce legal effects concerning the data subject or significantly affect him and which are based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.
- Anonymous or pseudonymous access and payment must be offered if technically possible and reasonable.
- Users must be able to withdraw their consent on processing traffic or location data at any time
Existing approaches for catering to privacy constraints

• Largest permissible dominator (e.g., Disney)
  – Infeasible if a large number of jurisdictions are involved, since the largest permissible denominator would be very small
  – Individual preferences not taken into account

• Different country/region versions (e.g., IBM)
  – Infeasible as soon as the number of countries/regions, and hence the number of different versions of the personalized system, increases
  – Individual preferences not taken into account

• Anonymous personalization (users are not identified)
  – Nearly full personalization possible
  – Harbors the risk of misuse
  – Slightly difficult to implement if physical shipments are involved
  – Practical extent of protection unclear
  – Individual user preferences not taken into account
Different methods differ in their data requirements, quality of predictions, and also their *privacy implications*.

<table>
<thead>
<tr>
<th>user modeling component</th>
<th>methods used</th>
<th>data used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>demographic data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>user-supplied data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>visited pages</td>
</tr>
<tr>
<td>UMC₁</td>
<td>clustering</td>
<td>X</td>
</tr>
<tr>
<td>UMC₂</td>
<td>rule-based reasoning</td>
<td>X</td>
</tr>
<tr>
<td>UMC₃</td>
<td>fuzzy reasoning with uncertainty</td>
<td>X</td>
</tr>
<tr>
<td>UMC₄</td>
<td>rule-based reasoning</td>
<td>X</td>
</tr>
<tr>
<td>UMC₅</td>
<td>fuzzy reasoning with uncertainty</td>
<td>X</td>
</tr>
<tr>
<td>UMC₆</td>
<td>incremental machine learning</td>
<td>X</td>
</tr>
<tr>
<td>UMC₇</td>
<td>one-time machine learning across sessions</td>
<td>X</td>
</tr>
<tr>
<td>UMC₈</td>
<td>one-time machine learning + fuzzy reasoning with uncertainty</td>
<td>X</td>
</tr>
</tbody>
</table>
Our approach

Develop a mechanism that dynamically selects those user modeling methods that *comply with the currently prevailing privacy constraints*:

- the user’s individual privacy preferences
- the privacy norms that apply to the user
### User modeling methods

<table>
<thead>
<tr>
<th>user modeling component</th>
<th>methods used</th>
<th>data used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>demographic data</td>
</tr>
<tr>
<td>UMC_1</td>
<td>clustering</td>
<td>X</td>
</tr>
<tr>
<td>UMC_2</td>
<td>rule-based reasoning</td>
<td>X</td>
</tr>
<tr>
<td>UMC_3</td>
<td>fuzzy reasoning with uncertainty</td>
<td>X</td>
</tr>
<tr>
<td>UMC_4</td>
<td>rule-based reasoning</td>
<td>X</td>
</tr>
<tr>
<td>UMC_5</td>
<td>fuzzy reasoning with uncertainty</td>
<td>X</td>
</tr>
<tr>
<td>UMC_6</td>
<td>incremental machine learning</td>
<td>X</td>
</tr>
<tr>
<td>UMC_7</td>
<td>one-time machine learning across sessions</td>
<td>X</td>
</tr>
<tr>
<td>UMC_8</td>
<td>one-time machine learning + fuzzy reasoning with uncertainty</td>
<td>X</td>
</tr>
</tbody>
</table>

Different methods differ in their data requirements, quality of predictions, and also their *privacy implications*.
Product line architecture

“The common architecture for a set of related products or systems developed by an organization.” [Bosch, 2000]

A PLA includes

– Stable core: basic functionalities
– Options: optional features/qualities
– Variants: alternative features/qualities

Dynamic runtime selection (van der Hoek 2002):
A particular architecture *instance* is selected from the product-line architecture
Our approach

(Selection Component)
Example: ogle.com cum privacy

1. Notification of LDAP bindings
2. Each user’s detected privacy constraints
3. Generate bindings
4. Bindings for each UMCs
5. Each user’s PCS vector
6. Create or assign one instance for each user
7. personalized privacy-respecting service for each user
8. Dynamic Privacy Context Detection

- UMC1: applicable to Alice
- UMC2: applicable to Cheng
- UMC3: applicable to Bob

Evaluate each UMC’s Boolean guards against its binding and then generate corresponding Privacy Constraints Satisfaction (PCS) vectors.
Run-time System Instances

<table>
<thead>
<tr>
<th>Assigned to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
</tr>
<tr>
<td>Cheng</td>
</tr>
<tr>
<td>Bob</td>
</tr>
</tbody>
</table>

(3) personalized privacy-respecting services for each user

(7) Create or assign one instance for each user

(4) Bindings for each UMCs

User Modeling Component Pool

<table>
<thead>
<tr>
<th>Alice</th>
<th>Cheng</th>
<th>Bob</th>
</tr>
</thead>
<tbody>
<tr>
<td>UMC_1</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UMC_2</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UMC_3</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UMC_4</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UMC_5</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UMC_6</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UMC_7</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UMC_8</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Evaluate each UMC's Boolean guards against its binding and then generate corresponding Privacy Constraints Satisfaction (PCS) vectors

(2) Each user's detected privacy constraints

(1) Notification of LDAP binds

(3) Generate bindings

Dynamic Privacy Context Detection

Privacy constraints applied to
Alice

Privacy constraints applied to
Cheng

Privacy constraints applied to
Bob
The privacy constraints

Privacy constraints applied to Alice

German Tele-Service Data Protection Law

Section 4(2)-4(4): profiling
Combining user profiles retrievable under pseudonyms with data relating to the bearer of the pseudonym, is prohibited.

Personal data to be erased immediately after each session except for very limited purposes.

Privacy constraints applied to Cheng

Cheng’s own privacy preferences:
“Dislike being tracked”

Privacy constraints applied to Bob

Network Advertising Initiative (NAI) Self-Regulatory Principles

Section II: NAI’s Statement of Purposes
Merging non-personally identifiable use data with personally identifiable demographic data, is prohibited unless user give prior affirmative consent.
Run-time System Instances

Assign to

Alice
Cheng
Bob

Selection Component

(3) Generate bindings

User Modeling Component Pool

(4) Bindings for each UMCs

(6) Each user’s PCS vector

Dynamic Privacy Context Detection

Privacy constraints applied to
Alice
Cheng
Bob

(5) Evaluate each UMC’s Boolean guards against its binding and then generate corresponding Privacy Constraints Satisfaction (PCS) vectors
There is no magic bullet for reconciling personalization with privacy

Effort is comparable to

… making systems secure

… making systems fast

… making systems reliable
Privacy-Enhanced Personalization: need for a process approach

1. Gain the user’s trust
   - Respect the user’s privacy attitude (and let the user know)
     • Respect privacy laws / industry privacy agreements
   - Provide benefits (including optimal personalization within the given privacy constraints)
   - Increase the user’s understanding (don’t do magic)
   - Use trust-enhancing methods
   - Give users control
   - Use privacy-enhancing technology (and let the user know)

2. Then be patient, and most users will incrementally come forward with personal data / permissions if the usage purpose for the data and the ensuing benefits are clear and valuable enough to them.
• Study the impacts of privacy laws, industry regulations and individual privacy preferences on the admissibility of personalization methods

• Provide optimal personalization while respecting privacy constraints

• Apply state-of-the-art industry practice for managing the combinatorial complexity of privacy constraints

<table>
<thead>
<tr>
<th>Country</th>
<th>Registration duties</th>
<th>Record-keeping duties</th>
<th>Reporting duties</th>
<th>Disclosure duties at website</th>
<th>Duty to respect user requests for</th>
<th>Duty to respect user requests for (“removal” or “right to be forgotten”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>✴ yes</td>
<td>?</td>
<td>?</td>
<td>✴ yes</td>
<td>✴ Inspection ✴ Correction ✴ exception</td>
<td>Argentina ✴ ?</td>
</tr>
<tr>
<td>Australia</td>
<td>?</td>
<td>✴ yes</td>
<td>✴ yes</td>
<td>✴ yes</td>
<td>✴ Inspection ✴ Inspection</td>
<td>Australia ✴ ?</td>
</tr>
<tr>
<td>Austria</td>
<td>✴ yes ✴ content</td>
<td>?</td>
<td>?</td>
<td>✴ yes</td>
<td>✴ Inspection ✴ Rectification ✴ Erasure</td>
<td>Austria ✴ yes</td>
</tr>
<tr>
<td>Canada</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>✴ yes</td>
<td>✴ Inspection</td>
<td>Canada ✴ ?</td>
</tr>
</tbody>
</table>