Data-perturbative, privacy-enhancing mechanisms for personalized recommendation systems

Javier Parra-Arnau

Joint work with Jordi Forné and David Rebollo-Monedero

javier.parra-arnau@inria.fr

INRIA Grenoble – Rhône-Alpes

Paris, France

December 16, 2014

Outline

- Motivation
- Information Leakage and Data Perturbation
  - Adversary Model
  - Data Perturbation
  - Quantitative Measures of Privacy and Anonymity for User Profiles
- Forgery and Suppression of Ratings in Recommendation Systems
  - Optimal Trade-Off between Privacy and Utility
- Experimental Analysis
- Conclusions

Outline

- Motivation
- Information Leakage and Data Perturbation
  - Adversary Model
  - Data Perturbation
  - Quantitative Measures of Privacy and Anonymity for User Profiles
- Forgery and Suppression of Ratings in Recommendation Systems
  - Optimal Trade-Off between Privacy and Utility
- Experimental Analysis
- Conclusions

Information Overload

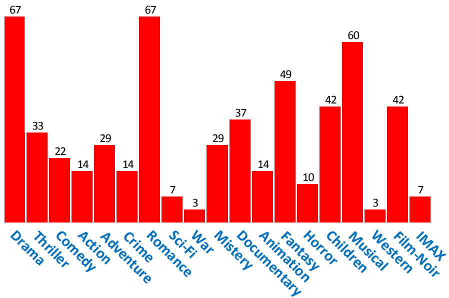
- IBM claims that “90% of the data in the world today has been created in the last two years alone” (2012) [1]

Personalized Information Systems

- A personalized information system is an information system that tailors the information-exchange functionality to meet the specific interests of their users
  - Examples of personalization include recommendation systems, tagging systems, personalized Web search and personalized news

Examples of Personalized Information Systems

User Profiles



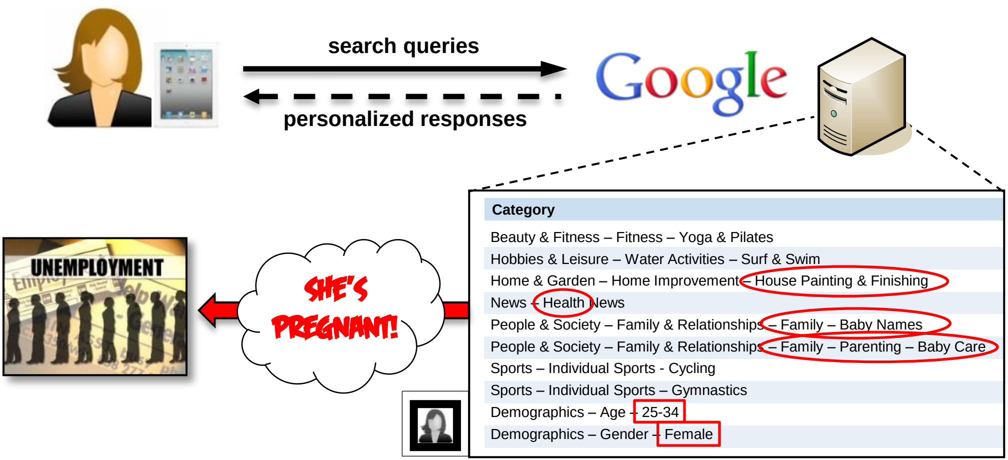
Your categories

Below you can edit the interests and inferred demographics that Google has associated with your cookie:

Category	
Beauty & Fitness – Fitness – Yoga & Pilates	Remove
Hobbies & Leisure – Water Activities – Surf & Swim	Remove
Home & Garden – Home Improvement – House Painting & Finishing	Remove
News – Health News	Remove
People & Society – Family & Relationships – Family – Baby Names	Remove
People & Society – Family & Relationships – Family – Parenting – Baby Care	Remove
Sports – Individual Sports - Cycling	Remove
Sports – Individual Sports – Gymnastics	Remove
Science – Mathematics	Remove
Technology - Smartphones	Remove

Privacy Risk

- Profiling is therefore what enables those systems to determine what information is relevant to users, but at the same time, it is the source of serious privacy concerns



Outline

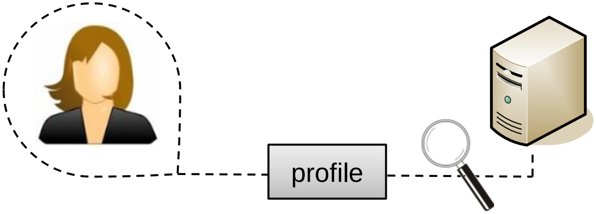
- Motivation
- Information Leakage and Data Perturbation
  - Adversary Model
  - Data Perturbation
  - Quantitative Measures of Privacy and Anonymity for User Profiles
- Forgery and Suppression of Ratings in Recommendation Systems
  - Optimal Trade-Off between Privacy and Utility
- Experimental Analysis
- Conclusions

Adversary Model

- We justify and interpret Kullback-Leibler (KL) divergence and Shannon's entropy as privacy and anonymity metrics in the application of personalized information systems
- The level of privacy provided by a PET is measured with respect to an adversary model
  - What scenario is assumed?
  - Who can be the privacy attacker?
  - How does the attacker model user interests?
  - What is the attacker after when profiling users?

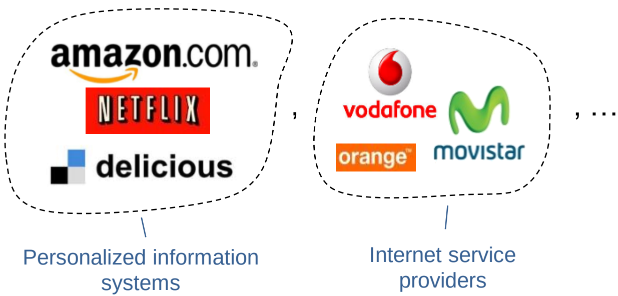
Adversary Model

- We justify and interpret Kullback-Leibler (KL) divergence and Shannon's entropy as privacy and anonymity metrics in the application of personalized information systems
- The level of privacy provided by a PET is measured with respect to an adversary model
  - What scenario is assumed?
  - Who can be the privacy attacker?
  - How does the attacker model user interests?
  - What is the attacker after when profiling users?



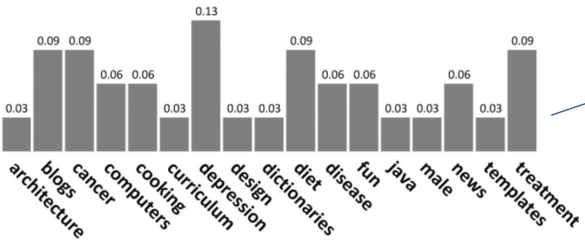
Adversary Model

- We justify and interpret Kullback-Leibler (KL) divergence and Shannon's entropy as privacy and anonymity metrics in the application of personalized information systems
- The level of privacy provided by a PET is measured with respect to an adversary model
  - What scenario is assumed?
  - Who can be the privacy attacker?
  - How does the attacker model user interests?
  - What is the attacker after when profiling users?



Adversary Model

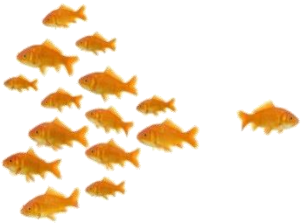
- We justify and interpret Kullback-Leibler (KL) divergence and Shannon's entropy as privacy and anonymity metrics in the application of personalized information systems
- The level of privacy provided by a PET is measured with respect to an adversary model
  - What scenario is assumed?
  - Who can be the privacy attacker?
  - How does the attacker model user interests?
  - What is the attacker after when profiling users?



user profile
=
probability mass function (PMF)

Adversary Model

- We justify and interpret Kullback-Leibler (KL) divergence and Shannon's entropy as privacy and anonymity metrics in the application of personalized information systems
- The level of privacy provided by a PET is measured with respect to an adversary model
  - What scenario is assumed?
  - Who can be the privacy attacker?
  - How does the attacker model user interests?
  - What is the attacker after when profiling users?



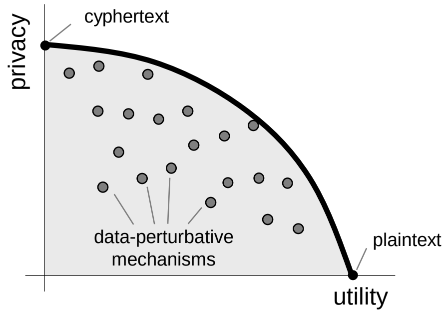
individuation

Privacy via Perturbation

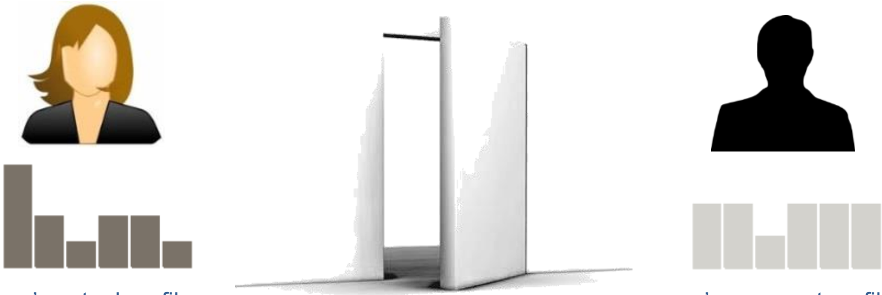
- In personalized information systems, the intended recipient of sensitive information may not be fully trusted

- In traditional approaches to privacy, users or designers decide whether certain sensitive information is to be made available or not. The availability of this data enables certain functionality. Its unavailability produces the highest level of privacy
  - but when intended yet untrusted recipients…

- Data perturbation is a completely different approach to more conventional privacy and security strategies
  - contemplates the possibility of exposing only portions of the data, or somewhat distorted versions of it,
  - to gain privacy at the cost of data utility



Actual and Apparent Profiles



user's actual profile q user's apparent profile t

- Users counter the adversary by distorting their private data locally
- Next, the KL divergence and Shannon's entropy are interpreted as measures of privacy and anonymity

Shannon's entropy

$$\mathrm{H}(t) = -\sum_i t_i \log t_i$$

$$\mathrm{D}(t \,\|\, p) = \sum_i t_i \log \frac{t_i}{p_i}$$

$$\mathrm{D}(t \,\|\, u) = \log n - \mathrm{H}(t)$$

KL divergence

uniform distribution

Anonymity Criteria against Individuation (I)

- The probability of a profile (distribution) may be a measure of its anonymity
- But this PMF of distributions is usually unknown…
- The maximum-entropy method is a general-purpose method for making inferences or predictions based on incomplete information
  - Its origins lie in statistical mechanics but it is present in diverse areas such as statistical physics, signal processing and spectral estimation
- Jaynes' rationale behind entropy maximization [2]
  - $X_1, \ldots, X_k$ is a sequence of i.i.d. drawings of a uniform r.v. on $\{1, \ldots, n\}$
  - Let k_i be the number of times symbol i appears in a sequence $x_1, \ldots, x_k$
  - The type t of a sequence is $t = \left(\frac{k_1}{k}, \ldots, \frac{k_n}{k}\right)$

$$\underbrace{\mathrm{H}}_{\text{Shannon's entropy}}(\underbrace{t}_{\substack{\text{empirical distribution}\\\text{(user's apparent profile)}}}) = \mathrm{H}\left(\frac{k_1}{k}, \ldots, \frac{k_n}{k}\right) \simeq \frac{1}{\underbrace{k}_{\substack{\text{\# i.i.d. drawings}\\\text{of a uniform r.v.}}}} \log \underbrace{\frac{k!}{k_1! \cdots k_n!}}_{\substack{\text{\# sequences}\\\text{with type } t}} \quad \text{for } k \gg 1$$

Anonymity Criteria against Individuation (II)

- Jaynes somehow justifies the principle of insufficient reason. But his argument is restricted to uniformly distributed drawings
- Extension of Jaynes' argument to KL divergence
  - A prior knowledge of an arbitrary PMF p of the samples $X_1, \ldots, X_k$
  - The type T of an i.i.d. drawing is an r.v. We may define its PMF $p_T(t) = \mathrm{P}\{T = t\}$
  - The expected type is $\mathrm{E}T = p$

$$-\frac{1}{k} \log p_T(t) \simeq \mathrm{D}(t \,\|\, p) \quad \text{for } k \gg 1$$

probability of a type

KL divergence

reference, average distribution

- Under this argument
  - KL divergence $\mathrm{D}(t \,\|\, p)$ may be interpreted as a measure of privacy, more precisely anonymity
  - roughly speaking, $\downarrow \mathrm{D}(t \,\|\, p) \;\Rightarrow \uparrow p_T(t) \;\Rightarrow \uparrow$ # users with this profile t
  - KL divergence regarded as a measure of anonymity, not in the sense of identifiability

Outline

- Motivation
- Information Leakage and Data Perturbation
  - Adversary Model
  - Data Perturbation
  - Quantitative Measures of Privacy and Anonymity for User Profiles
- Forgery and Suppression of Ratings in Recommendation Systems
  - Optimal Trade-Off between Privacy and Utility
- Experimental Analysis
- Conclusions

Data Perturbation in Recommendation Systems

- We focus on recommendation systems, possibly the most popular personalized information systems, and propose a mechanism that allows users to simultaneously
  - submit ratings of items that do not reflect their interests – forgery of ratings
  - skip rating certain genuine items – suppression of ratings

apply **suppression** to

some items **aligned** with your interests

apply **forgery** to

some items **not aligned** with your interests

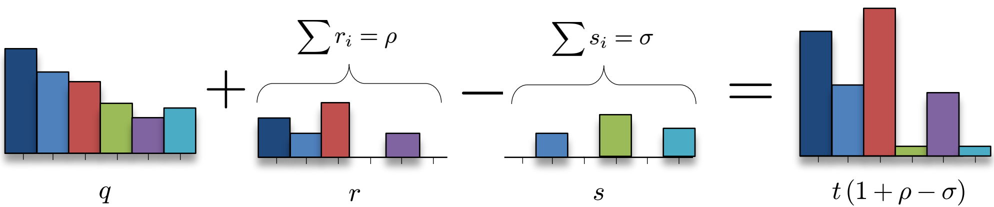
Optimal Privacy-Utility Trade-Off (I)

- We seek a mathematically optimal mechanism in the sense that utility is maximized for a given privacy constraint, and vice versa
- Assume that the attacker wishes to individuate users (i.e., find uncommon users), and that p is known to users
- Denote by q the user's actual profile and define
  - rating-forgery rate $\rho \in [0, \infty]$, as the ratio of forged ratings to total genuine ratings that a user consents to submit
  - rating-suppression rate $\sigma \in [0, 1)$, as the ratio of genuine ratings agreed to eliminate
- User's apparent item distribution

a forgery strategy

$$t = \frac{q + r - s}{1 + \rho - \sigma}$$

a suppression strategy



Optimal Privacy-Utility Trade-Off (II)

- Privacy risk, or more precisely anonymity loss, is measured as the KL divergence between t and p
- Loss in utility measured as the rates of forgery and suppression
  - mathematically tractable measures of utility
- Assuming that the population of users is large enough, the privacy-forgery-suppression function is defined as

average profile of the population

$$\mathcal{R}(\rho, \sigma) = \min_{\substack{r,s \\ r_i \geqslant 0,\ s_i \geqslant 0, \\ q_i + r_i - s_i \geqslant 0, \\ \sum r_i = \rho,\ \sum s_i = \sigma}} \mathrm{D}\left(\frac{q + r - s}{1 + \rho - \sigma} \,\middle\|\, p \right)$$

privacy risk (anonymity loss)

KL divergence (inverse indicator of the likelihood of t within a population)

  - which characterizes the optimal trade-off among privacy, forgery rate and suppression rate

Theoretical Results (I)

- Explicit-form solution to the optimization problem and characterization of the optimal trade-off surface among privacy, forgery rate and suppression rate
- In the closure of the noncritical-privacy region
  - Assume w.o.l.o.g. $q_1/p_1 \leqslant \cdots \leqslant q_n/p_n$
  - Define $Q_i = \sum_{k=1}^{i} q_k \, , \bar{Q}_i = \sum_{k=i}^{n} q_k$ and $P_i \, , \bar{P}_i$ analogously, and
  - Based on resource allocation argument
  - The optimal forgery and suppression strategies yield

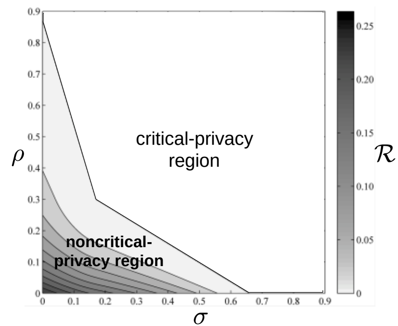
$$
\begin{aligned}
\tilde{q} &= (\; Q_i \;,\; q_{i+1} \;,\; \ldots \;,\; q_{j-1} \;,\; \bar{Q}_j \;) \\
\tilde{r} &= (\; \rho \;,\; 0 \;,\; \ldots \;,\; 0 \;,\; 0 \;) \\
\tilde{s} &= (\; 0 \;,\; 0 \;,\; \ldots \;,\; 0 \;,\; \sigma \;) \\
\tilde{p} &= (\; P_i \;,\; p_{i+1} \;,\; \ldots \;,\; p_{j-1} \;,\; \bar{P}_j \;)
\end{aligned}
$$

$$
r_k^* = \begin{cases} \frac{p_k}{P_i}(Q_i + \rho) - q_k \;, & k = 1, \ldots, i \\ 0 \;, & k = i+1, \ldots, n \end{cases}
$$

$$
s_k^* = \begin{cases} 0 \;, & k = 1, \ldots, j-1 \\ q_k - \frac{p_k}{P_j}(\bar{Q}_j - \sigma), & k = j, \ldots, n \end{cases}
$$

add false ratings where $\frac{q_k}{p_k}$ is low

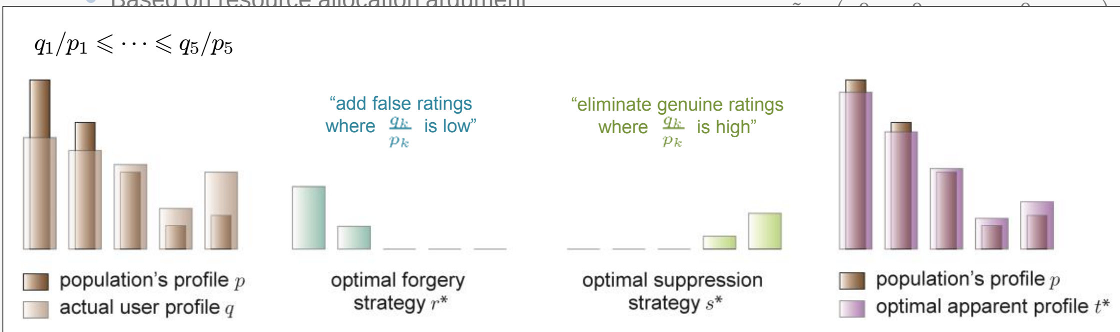
eliminate genuine ratings where $\frac{q_k}{p_k}$ is high



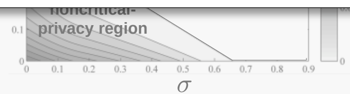
  - Optimal trade-off $\mathcal{R}(\rho, \sigma) = \mathrm{D}\left(\frac{\tilde{q} + \tilde{r} - \tilde{s}}{1 + \rho - \sigma} \middle\| \tilde{p} \right)$

Theoretical Results (I)

- **Explicit-form solution** to the optimization problem and characterization of the optimal trade-off surface among privacy, forgery rate and suppression rate
- In the closure of the noncritical-privacy region
  - Assume w.o.l.o.g. $q_1/p_1 \leqslant \cdots \leqslant q_n/p_n$
  - Define $Q_i = \sum_{k=1}^{i} q_k \, , \bar{Q}_i = \sum_{k=i}^{n} q_k$ and $P_i \, , \bar{P}_i$ analogously, and $\tilde{q} = (\, Q_i \, , q_{i+1} \, , \dots , q_{j-1} \, , \bar{Q}_j \,)$
    $\tilde{r} = (\, \rho \, , \, 0 \, , \dots , \, 0 \, , \, 0 \,)$
  - Based on resource allocation argument



  - Optimal trade-off $\mathcal{R}(\rho, \sigma) = \mathrm{D}\left(\frac{\tilde{q} + \tilde{r} - \tilde{s}}{1 + \rho - \sigma} \middle\| \tilde{p}\right)$



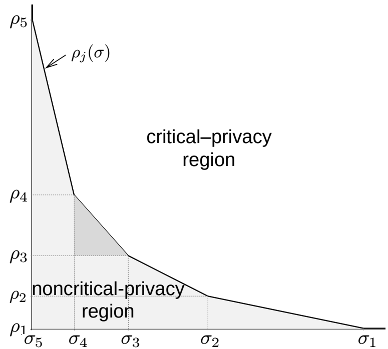
Theoretical Results (II)

- The critical-privacy region is convex. Its boundary is a convex, piecewise linear function of σ, determined by some forgery and suppression thresholds
  - For $j = 2, \ldots, n$ define the forgery thresholds

$$\rho_i = \begin{cases} P_i \frac{q_i}{p_i} - Q_i & , \quad i = 1, \ldots, j-1 \\ \frac{P_{j-1}}{P_j}(\bar{Q}_j - \sigma) - Q_{j-1} & , \quad i = j \\ \infty & , \quad i = j+1 \end{cases}$$

  - For $j = 1, \ldots, n$ define the suppression thresholds

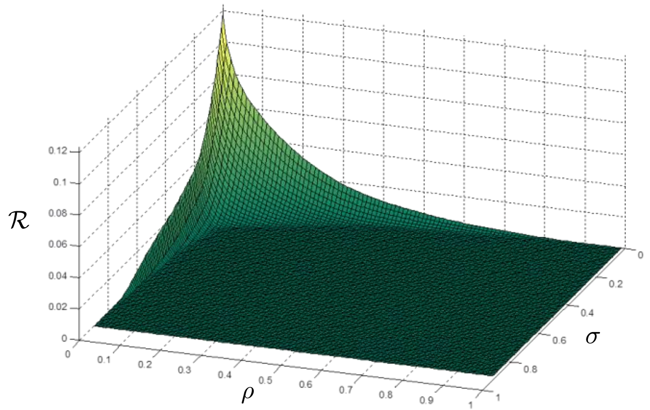
$$\sigma_j = \bar{Q}_j - \bar{P}_j \frac{q_j}{p_j}$$



- First-order Taylor approximation at the origin in the nontrivial case when $q \neq p$
  - $$\underbrace{\frac{\mathrm{D}(q \,\|\, p) - \mathcal{R}(\rho, \sigma)}{\mathrm{D}(q \,\|\, p)}}_{\text{relative reduction in privacy risk}} \simeq \rho \underbrace{\left(1 - \frac{\log \frac{q_1}{p_1}}{\mathrm{D}(q \,\|\, p)}\right)}_{\delta_\rho > 1} + \sigma \underbrace{\left(\frac{\log \frac{q_n}{p_n}}{\mathrm{D}(q \,\|\, p)} - 1\right)}_{\delta_\sigma > 0}$$

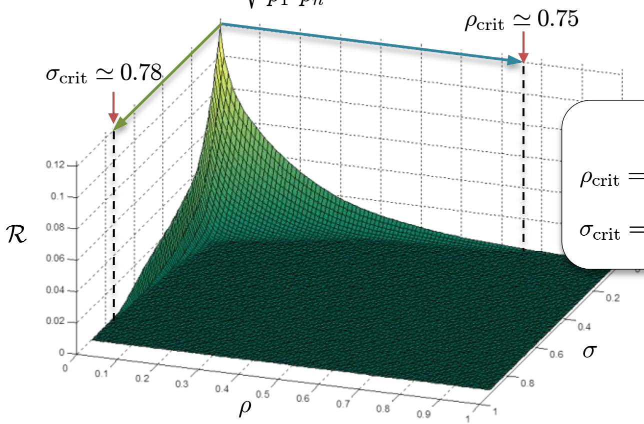
Theoretical Results (III)

- Forgery and suppression as pure strategies, i.e., operate alone
  - Which is the pure strategy causing the minimum distortion to attain the critical-privacy region?
    - Choose forgery if, and only if, $\dfrac{q_1/p_1 + q_n/p_n}{2} < 1$
  - Which is the pure strategy providing better privacy protection at low rates?
    - Choose forgery if, and only if, $\sqrt{\dfrac{q_1}{p_1}\dfrac{q_n}{p_n}} < 2^{\mathrm{D}(q\,\|\,p)}$



Theoretical Results (III)

- Forgery and suppression as pure strategies, i.e., operate alone
  - Which is the pure strategy causing the minimum distortion to attain the critical-privacy region?
    - Choose forgery if, and only if, $\frac{q_1/p_1 + q_n/p_n}{2} < 1$
  - Which is the pure strategy providing better privacy protection at low rates?
    - Choose forgery if, and only if, $\sqrt{\frac{q_1}{p_1}\frac{q_n}{p_n}} < 2^{\mathrm{D}(q\,\|\,p)}$



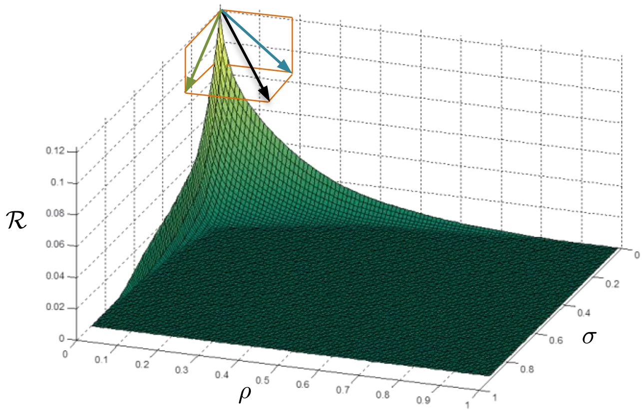
critical rates

$$\rho_{\text{crit}} = \min\{\rho : \mathcal{R}(\rho, 0) = 0\}$$

$$\sigma_{\text{crit}} = \min\{\sigma : \mathcal{R}(0, \sigma) = 0\}$$

Theoretical Results (III)

- Forgery and suppression as pure strategies, i.e., operate alone
  - Which is the pure strategy causing the minimum distortion to attain the critical-privacy region?
    - Choose forgery if, and only if, $\frac{q_1/p_1 + q_n/p_n}{2} < 1$
  - Which is the pure strategy providing better privacy protection at low rates?
    - Choose forgery if, and only if, $\sqrt{\frac{q_1}{p_1}\frac{q_n}{p_n}} < 2^{\mathrm{D}(q\,\|\,p)}$



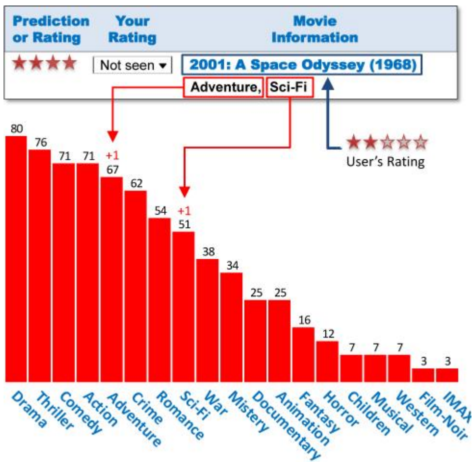
Outline

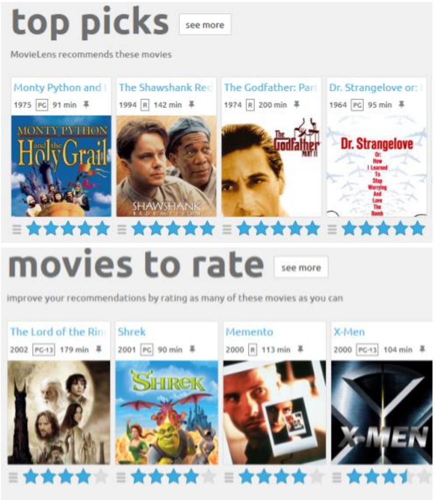
- Motivation
- Information Leakage and Data Perturbation
  - Adversary Model
  - Data Perturbation
  - Quantitative Measures of Privacy and Anonymity for User Profiles
- Forgery and Suppression of Ratings in Recommendation Systems
  - Optimal Trade-Off between Privacy and Utility
- Experimental Analysis
- Conclusions

Movielens Recommendation System

- Empirical assessment of our data-perturbative approach
  - Apply the forgery and the suppression of ratings to the popular movie recommendation system Movielens
  - Data set with 4 099 users, and profiles modeled across 19 movie genres

Example of user profile

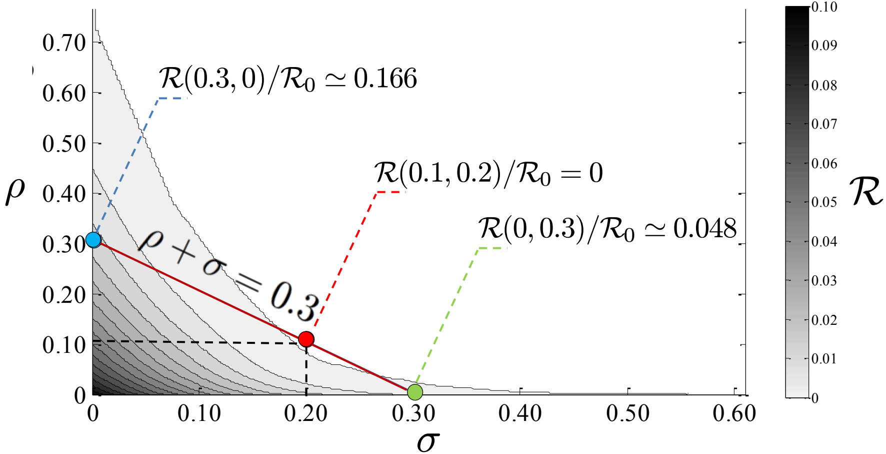




Detailed Experimental Results (I)

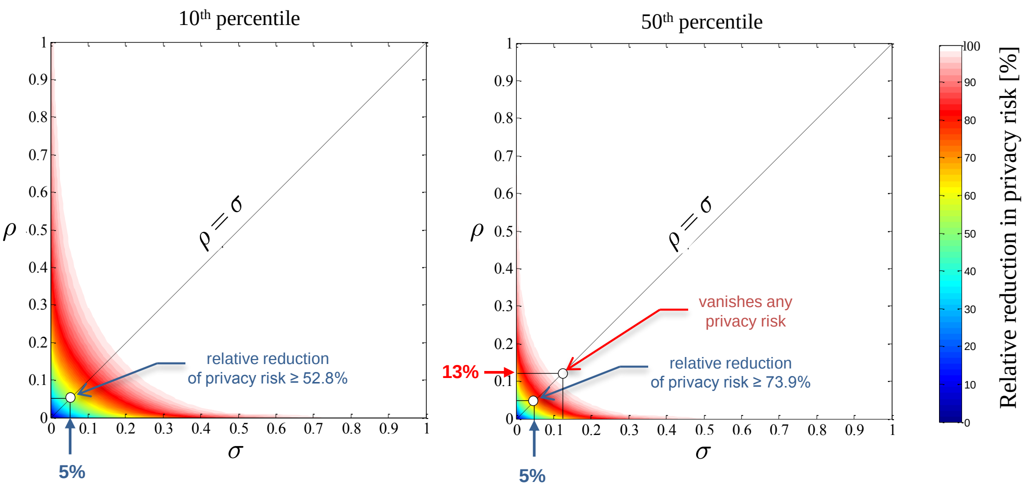
- Optimal trade-off between privacy and utility for a particular user
  - The mixed strategy may provide stronger privacy protection for the same total rate than the pure strategies, i.e.,

$$\underbrace{\mathcal{R}(\rho,\sigma)}_{\text{mixed}} \leqslant \underbrace{\mathcal{R}(\rho+\sigma,0)}_{\text{pure forgery}}, \underbrace{\mathcal{R}(0,\rho+\sigma)}_{\text{pure suppression}}$$



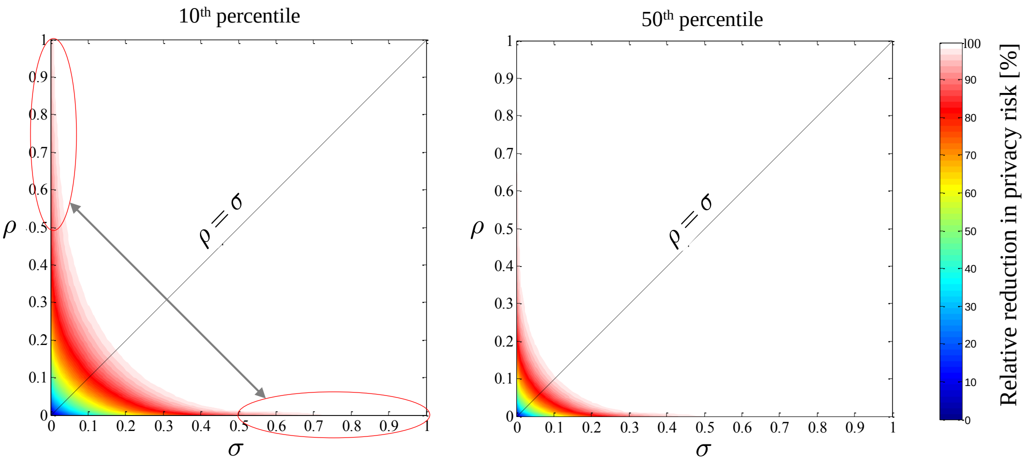
Detailed Experimental Results (II)

- Assume all 4099 users apply a common forgery rate and a common suppression rate
  - For relatively small values of ρ and σ (lower than 15%), a vast majority of users lowered privacy risk significantly
  - Slight asymmetry between the rates of forgery and suppression for pure strategies



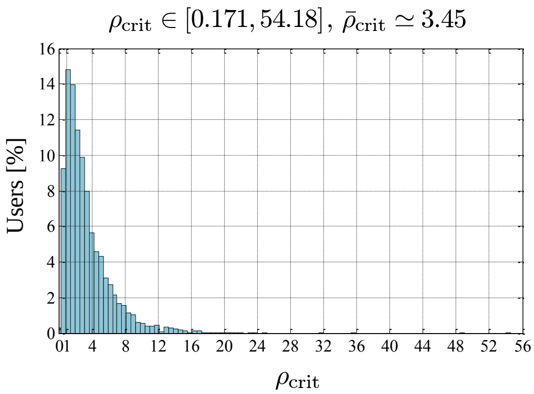
Detailed Experimental Results (II)

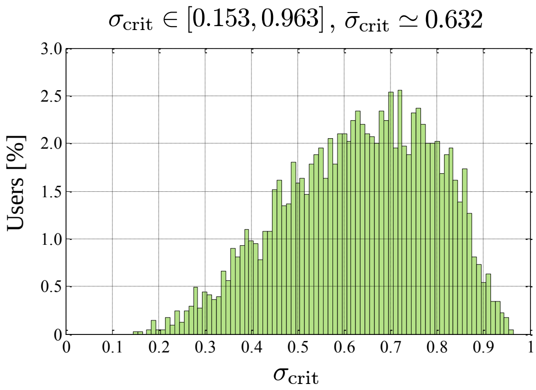
- Assume all 4099 users apply a common forgery rate and a common suppression rate
  - For relatively small values of ρ and σ (lower than 15%), a vast majority of users lowered privacy risk significantly
  - Slight asymmetry between the rates of forgery and suppression for pure strategies



Detailed Experimental Results (III)

- Pure strategies – in 95.3% of cases, suppression reached the critical-privacy region with a lower distortion than forgery did
  - Critical forgery rate $\rho_{\text{crit}} = \min\{\rho : \mathcal{R}(\rho, 0) = 0\}$
  - Critical suppression rate $\sigma_{\text{crit}} = \min\{\sigma : \mathcal{R}(0, \sigma) = 0\}$





Outline

- Motivation
- Information Leakage and Data Perturbation
  - Adversary Model
  - Data Perturbation
  - Quantitative Measures of Privacy and Anonymity for User Profiles
- Forgery and Suppression of Ratings in Recommendation Systems
  - Optimal Trade-Off between Privacy and Utility
- Experimental Analysis
- Conclusions

Conclusions

- Data-perturbative mechanism for the privacy enhancement in personalized recommendation systems
- Our mechanism has several features that make it particularly interesting to recommendation systems, but poses a trade-off between privacy and utility
- The proposed mechanism has been engineered to attain the optimal privacy-utility trade-off
  - Propose KL divergence as user-profile privacy criterion, and interpret it quantities from fundamental concepts of information theory and statistics
  - Privacy-utility trade-off modeled as optimization problems
  - Closed-form solution, by using convex-optimization techniques
- Theoretical analysis of said trade-off
- Experimental analysis carried out in Movielens

References

- [1] P. C. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch, and G. Palis, Understanding big data. McGraw-Hill, 2012. [Online]. Available: http://www-01.ibm.com/software/data/bigdata
- [2] E. T. Jaynes, "On the rationale of maximum-entropy methods," Proc. IEEE, vol. 70, no. 9, pp. 939-952, Sep. 1982

Thank you for your attention

Javier Parra-Arnau

http://sites.google.com/site/javierparraarnau

Data-perturbative, privacy-enhancing mechanisms for personalized recommendation systems

Javier Parra-Arnau

Joint work with Jordi Forné and David Rebollo-Monedero

javier.parra-arnau@inria.fr

INRIA Grenoble – Rhône-Alpes

Paris, France

December 16, 2014