プライバシー保護データマイニング

神嶌 敏弘 (産業技術総合研究所)
http://www.kamishima.net/

情報処理学会第73回全国大会，2011/03/03

個人データをめぐる現状

データの利活用には大きな利点がある

- インフラ・医療・流通などの効率化や，新たなエコシステム

データ管理を完全に保証することは不可能

- プライバシーポリシーで定めた基準を満たすかどうか，不明瞭な場合がどうしても存在する
- データの処理は複雑で監査会社にも完全には把握できないことも

個人情報とそれ以外の情報は本質的には区別できない

- あるデータから何を知ることができるかを予測できない

推薦システムと掲示板のID対応付け

[Frankowski+ SIGIR06]

推薦システムを使っている人のIDと映画掲示板を使っている人のIDを
その行動パターンから対応付けする

推薦システムでの
映画評価

映画掲示板での
感想の書き込み

同じ一群の映画について言及・評価している人は
同一である可能性が高い

- 一意に特定された利用者は全体の31%
- 5人のうちの一人として特定される利用者は44%
- 一意に識別されないようにするには，約88%の映画に対する評価を消す必要 ➡ システムとして役に立たない

VoIPでの通話相手の特定

[Verscheure+ ICDM06]

VoIP (Voice over Internet Protocol)

音声をインターネット上でやりとりし，電話などの機能を実現する

通話の内容や，通話相手は秘匿したい

- 会話内容を暗号化
- Onion Routing やプロキシなどでIPアドレスを秘匿

↓

Voice Activity Detection

sniff したパケットのヘッダ情報と信号処理で通話・非通話状態を特定

背景知識：一方が発話中はもう一方は発話しない

対話ペアの可能性を数値化 + クラスタリング

↓

通話対象のペアを特定できる

デモグラフィック属性のログからの抽出

[Jones ECMLPKDD09 Invited Talk]

nfl, poker, espn, ufc, railroad, prostate, football, golf, male, wrestling, compusa, saddam, a variety of adult terms

男性　　女性

fanfiction, bridal, makeup, women's, knitting, hair, ecards, glitter, yoga, diet, divorce

性別：83%

myspace, pregnancy, wikipedia, lyrics, quotes, apartments, torrent, baby, wedding, mall, soundtrack

若者　中高年

aarp, telephone, lottery, amazon.com, retirement, funeral, senior, mapquest, medicare, newspapers, repair

年齢：65%

誤差7歳以内

住所：35%

3桁郵便番号

ログからの個人の特定

[Jones ECMLPKDD09 Invited Talk]

個人の検索語の上位10人に自身が現れる割合は90%

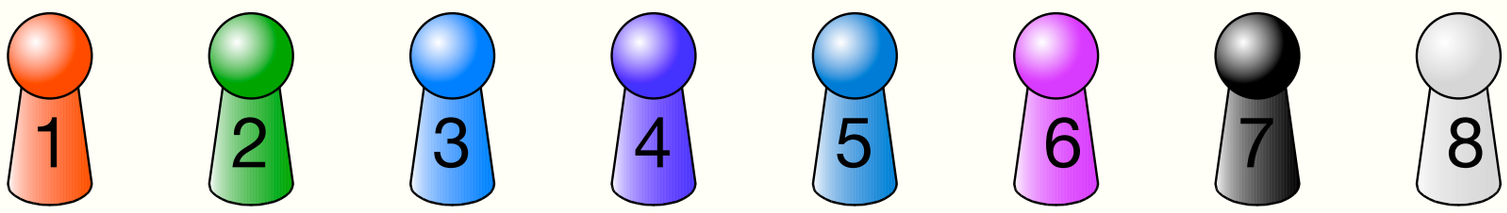
鈴木さんが検索した名前をTF-IDFでソート

鈴木  田中，山田，鈴木，佐藤 … 藤原， 高橋， 斉藤 …

上位10人を選び出す

ある名前を検索した人のリストの1位が本人である割合は85%

山本さんを検索した人を修正TF-IDFでソート



山本さん

ではどうしたらいいのか？

集めてよいデータとそうでないデータに分けて規制

↓

実効性はない

区別できない + 保証できない

↓

本来は，個人情報が漏れることより，それで起きる不都合を抑止したい

データの収集から利用までトータルで不都合が生じにくい工夫

- 収集：プライバシポリシー，データのアクセス管理，情報の自己コントロール権，サービスのアカウンタビリティ，プライバシー保護データマイニング
- 利用：オプトイン（Do not call list)

敵対的学習

[Lowd+ KDD05]

Spamメールフィルタ：メールの特徴からSpamかどうかを識別

↓

Spammer としては，どうにかすり抜けたい

- 送ったSpamメールがフィルタを通過したかどうかは分かる

  画像リンクなどの仕掛けと識別器の組み合わせ

- Spamメールとしての有効性をスコア付けできる

  Spammerにとってはリンクをクリックしたくなる語を入れたい

敵対的学習：Spamフィルタを通過するメールで，最も有効性のスコアが高いものを計算できるかどうかを調べる

↓

防御側の安全性が評価できる

差別配慮型マイニング

[Pedreschi+ KDD08]

データマイニングの結果が人生に大きく影響する場合
ローンの審査・人事採用・保険の可否
自分の個人情報によって自分に不利な判断が！

相関ルール：左辺の条件が成り立つとき右辺の条件が高い割合で成立
確信度：左辺の条件が成立しているとき右辺が成立している割合

住所 = 東京 ➡ ローン却下（確信度：0.25）

顔 = ブサイク ＆ 住所 = 東京 ➡ ローン却下（確信度：0.75）

差別的な条件を加えるとローン却下の確信度が上がった！

α-protection：差別的な条件を加えると確信度がα倍以上になる

こうした差別的な判断規則を見つけて排除する

これから作っていく枠組み

[Boyd WWW10 Keynote Talk]

共有地の悲劇：共有したデータを濫用してしまう不都合

ネット上の活動について社会学の見地から研究しているDana Boydの講演から

- 個人情報は，人間を一意に特定する PII (Personally Identifiable Information) を共有したとしても，人々を不快にさせる PEI (Personally Embarrassing Information) を放棄したわけではない
- civil inattention (by Erving Goffmann)：道ばたを歩いている人は，公共の場にはいるが，それをジロジロ見つめたりはしない
- プライバシとは**コンテキストの問題**である (by Helen Nissenbaum)

明らかな濫用を防ぐ技術・抑止するインセンティブ

利用のコンテキストから乖離しないように