Data Minimization Could Increase Unfairness!

Defining Data Minimization in Personalized Systems, Applying it through Active Learning, and Studying its Trade-off with Fairness

Nasim Sonboli, Department of Computer Science, Tufts University
Mentor: Bert Huang

Data Minimization, a legal obligation.

More data, more utility?
Utility-theoretic methods that balance the costs of sharing personal data with online services in return for the benefits of personalization

Data Minimization incurs only minimal quality loss!
Utility (average click entropy reduction) according to a greedy ordering of attributes in order of minimum incremental cost

Cost Comparison according to greedy ordering

Active Learning General Framework

Active Learning as a means of operationalizing Data Minimization

Hypothesis: Data Minimization could increase Consumer Unfairness due to data imbalance!

Initial Result: Consumer Unfairness in Active Learning methods

Goal: Designing Active Learning strategies to operationalize the Data Minimization principle in personalized systems while preserving accuracy and preventing/decreasing Unfairness.

The European Union’s General Data Protection Regulation (GDPR) mandates

- **Fairness** — Processing must be fair to the data subject.
- **Purpose limitation** — Process data for the legitimate purposes specified explicitly to the data subject when you collected it.
- **Data minimization** — Collect and process only as much data as necessary for the purposes specified.

Fairness — Processing must be fair to the data subject.
Purpose limitation — Process data for the legitimate purposes specified explicitly to the data subject when you collected it.
Data minimization — Collect and process only as much data as necessary for the purposes specified.

US Data Privacy laws such as Massachusetts Information Privacy Act (MIPA), Colorado Privacy Act (ColoPA), California Privacy Rights Act (CPRA), Virginia Consumer Data Protection Act (VCDPA), Utah Consumer Privacy Act (UCPA)

References