The AI/Data Science Professional

Security and Privacy

Acknowledgement of Country

The AI/Data Science Professional

Course Coordinator: Flora Salim

What's next...

RMIT UNIVERSITY

RMIT University acknowledges the people of the Woi wurrung and Boon wurrung language groups of the eastern Kulin Nation on whose unceded lands we conduct the business of the University.

RMIT University respectfully acknowledges their Ancestors and Elders, past and present. RMIT also acknowledges the Traditional Custodians and their Ancestors of the lands and waters across Australia where we conduct our business.

Ngarara Place

The AI/Data Science Professional – Week 5: Security and Privacy

Responsible AI/Data Science

Agenda - Privacy and Security

- Measuring Privacy
- Consent and the Privacy Act
- Identification and tracking
- Privacy breach through inference
- Data breaches/leakage
- Privacy Protection and Security Technologies
- What to do this week

Data is at the core of AI

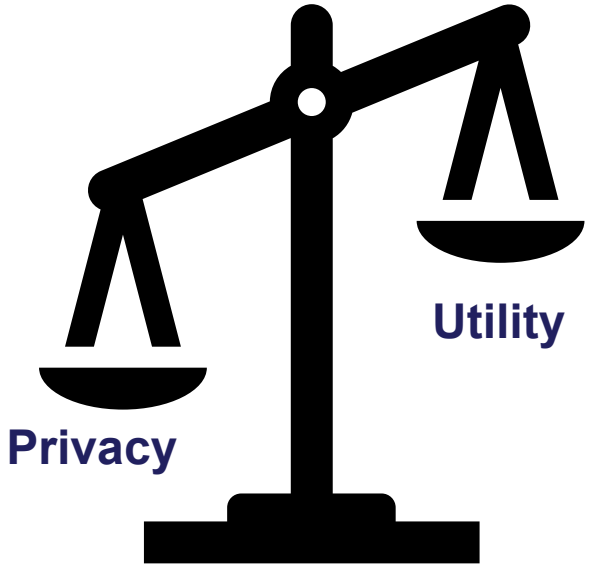
“The recent advances in key AI capabilities such as deep learning have been made possible by vast troves of data. This data has to be collected and used, which means issues related to AI are closely intertwined with those that relate to privacy and data.” - CSIRO & Data61, Innovation, Department of Industry, and Australian Government Science

Measuring Privacy

- Secrecy: it concerns information that others may gather about us
  - the probability of a data item being accessed
  - the change in knowledge of an adversary/third party upon seeing the data
- Anonymity: it addresses how much in the public gaze we are
  - the privacy leakage is measured in terms of the size of the blurring accompanying the release of data
- Solitude: it measures the degree to which others have physical access to us

(Xintao Wu, University of Arkansas)

Privacy vs. Utility Tradeoff



Data Privacy in the real world

Application	Data Collector	Third Party (adversary)	Private Information	Function (utility)
Medical	Hospital	Epidemiologist	Disease	Correlation between disease and geography
Genome analysis	Hospital	Statistician/ Researcher	Genome	Correlation between genome and disease
Advertising	Google/FB/Y!	Advertiser	Clicks/Browsing	Number of clicks on an ad by age/region/gender …
Social Recommen-dations	Facebook	Another user	Friend links / profile	Recommend other users or ads to users based on social network

Credit: Ashwin Machanavajjhala, Daniel Kifer, Bolin Ding, Xi He, 2013

Privacy breach through inference

Explicit vs. Implicit data

- Implicit data is information that is not provided intentionally but gathered from available data streams, either directly or through analysis of explicit data.
- Explicit data is information that is provided intentionally, for example through surveys and membership registration forms.

[Discussion]

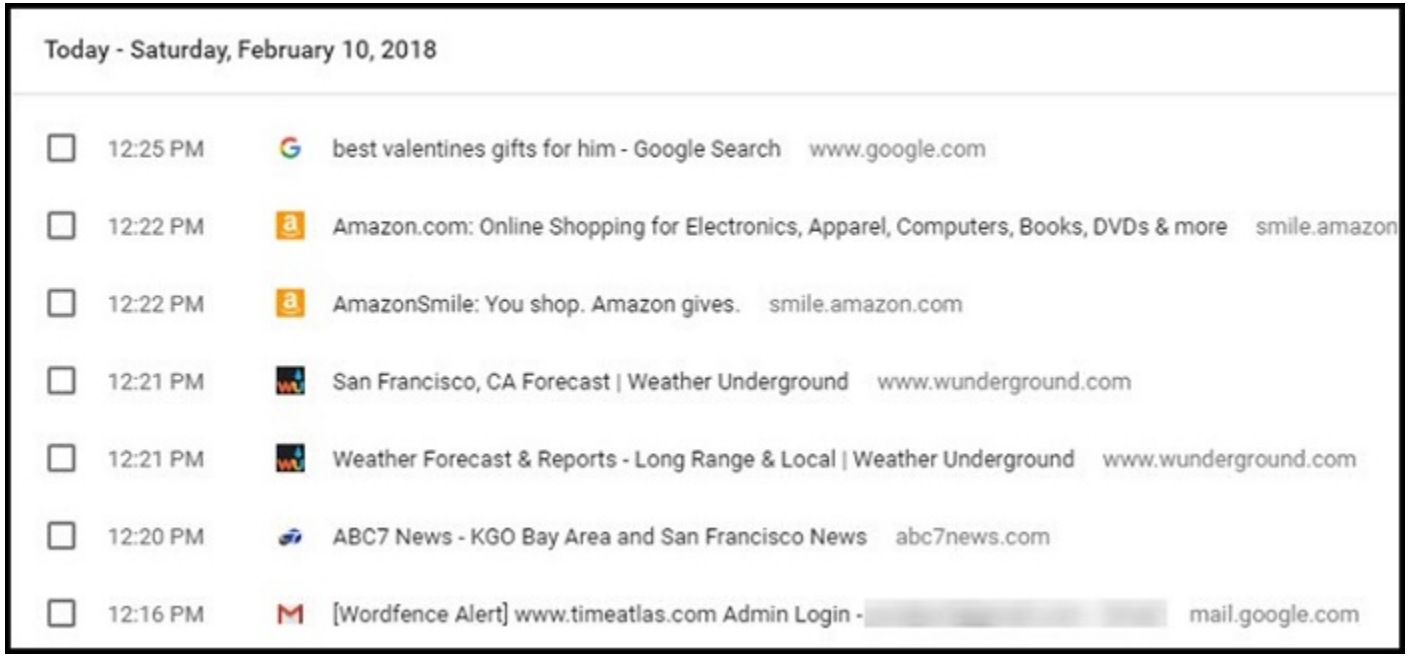
Jill posts to her Facebook page, “Jill is going to lunch early with her best friend Megan at Iron Pit BBQ!.” From this Facebook status the explicit data tells us that Jill and Megan are going to eat barbeque for lunch and where they're going to eat.

Q1: So what kind of implicit data you could probably derive from Jill’s post?

Q2: Imagine you are the CEO of Facebook fielding questions from the media. A socially conscious journalist from a magazine claims that users should own their own data and control its use. Give a reason for why Facebook would choose/not choose to do this?

Q3: Give your views on how to protect user privacy and confidentiality with social media data. Will simply making everyone’s data private will solve the problem?

[Discussion] Let's infer some private information about this user



Ghazaleh Beigi and Huan Liu, https://www.kdnuggets.com/2019/01/privacy-preserving-personalized-services.html

Privacy attacks

There are many definitions of these attacks

- Linkage attack
- Background knowledge attack
- minimality / reconstruction attack
- composition attack
- etc

(Ashwin Machanavajjhala, Daniel Kifer, Bolin Ding, Xi He, 2013)

Consent and the Privacy Act

- Four key terms
  - The individual is adequately informed before giving consent.
  - The individual gives consent voluntarily.
  - The consent is current and specific.
  - The individual has the capacity to understand and communicate their consent.
- The ‘right to be **forgotten’**

Case study: Cambridge Analytica and public trust

Through a Facebook app, a Cambridge University researcher was able to gain access to the personal information of not only users who agreed to take the survey, but also the people in those users' Facebook social networks. In this way, the app harvested data from millions of Facebook users. Various reports indicate that these data were then used to develop targeted advertising for various political campaigns run by Cambridge Analytica.

Case study: Cambridge Analytica and public trust

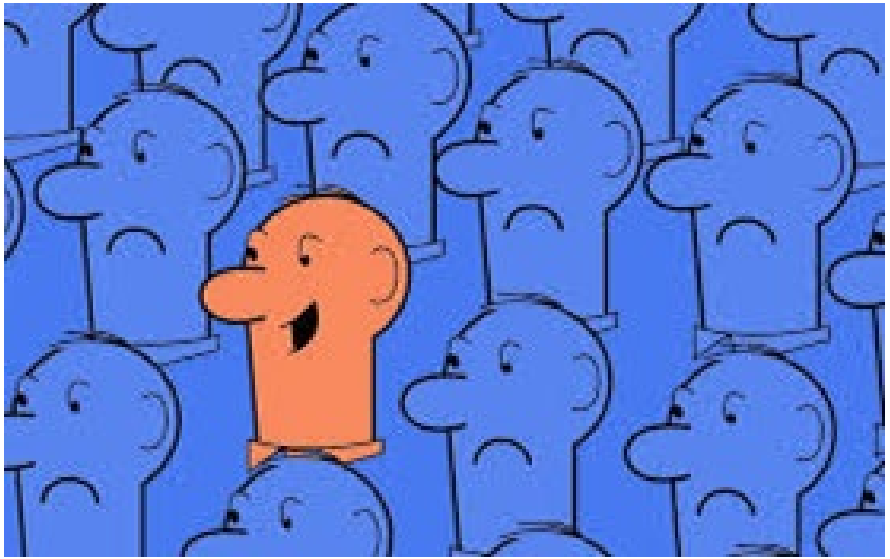
[Discussion]

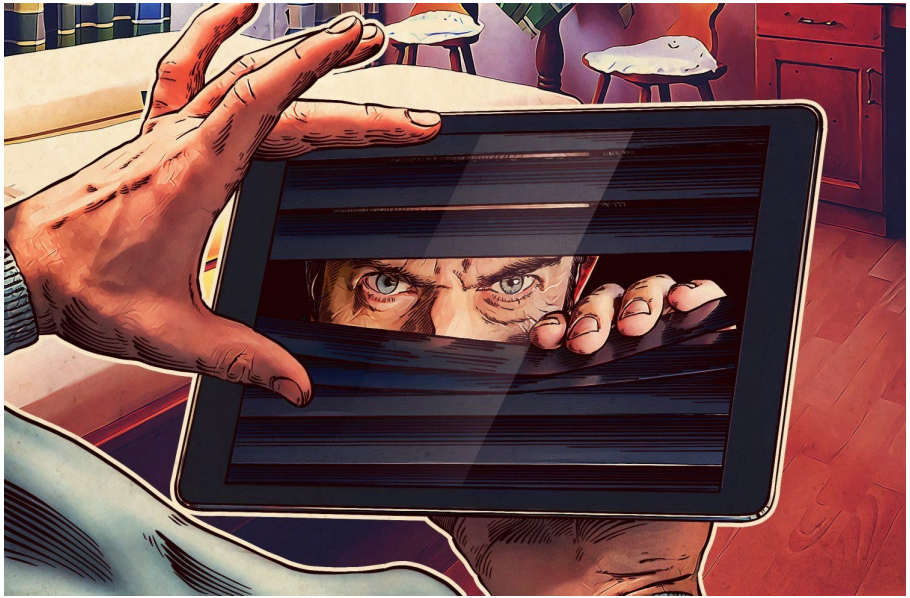
- Are there potential ethical concerns in relation to data privacy?
  - Not provided a transparent consent process
  - One user to effectively give consent for the use of others' data was particularly concerning
- Are there any concerns with regards to the intended use of this data and analysis for advertisement?
  - The use of personal data to profile and target political advertising to the users without appropriate consent

Case study: Cambridge Analytica and public trust

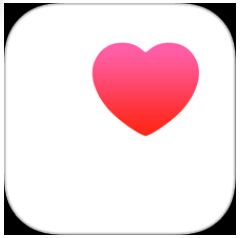
- Lesson we learned
  - Cost of inadequate data protection policies
  - Not sufficient to merely follow the letter of the law
- Code of practices
  - Australian Privacy Act
  - Consent process: current, specific, transparent
  - Regular review of data collection and usage policies
  - Balance between privacy protecting and technology innovation

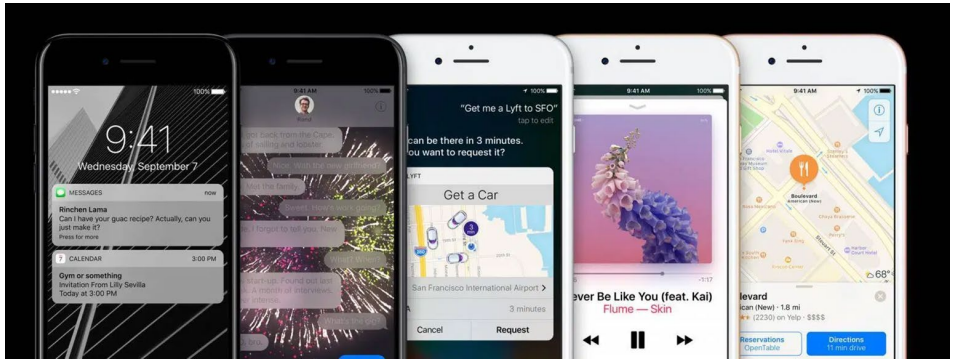
Identification and Tracking

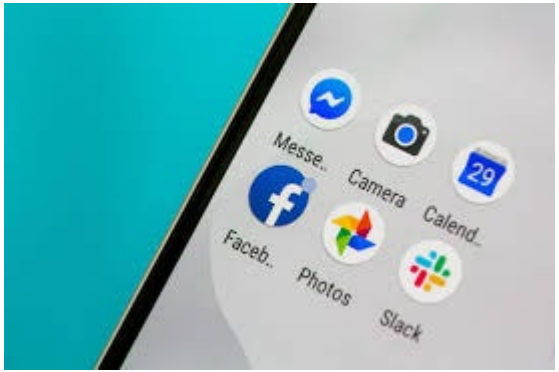


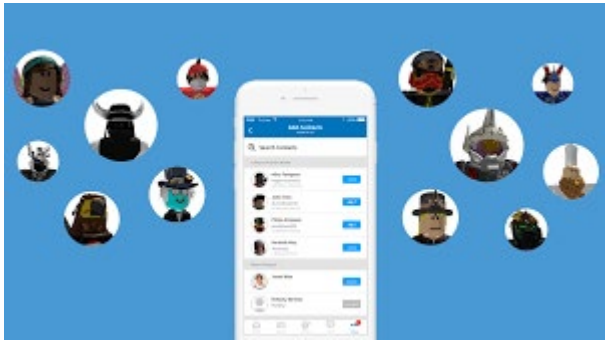


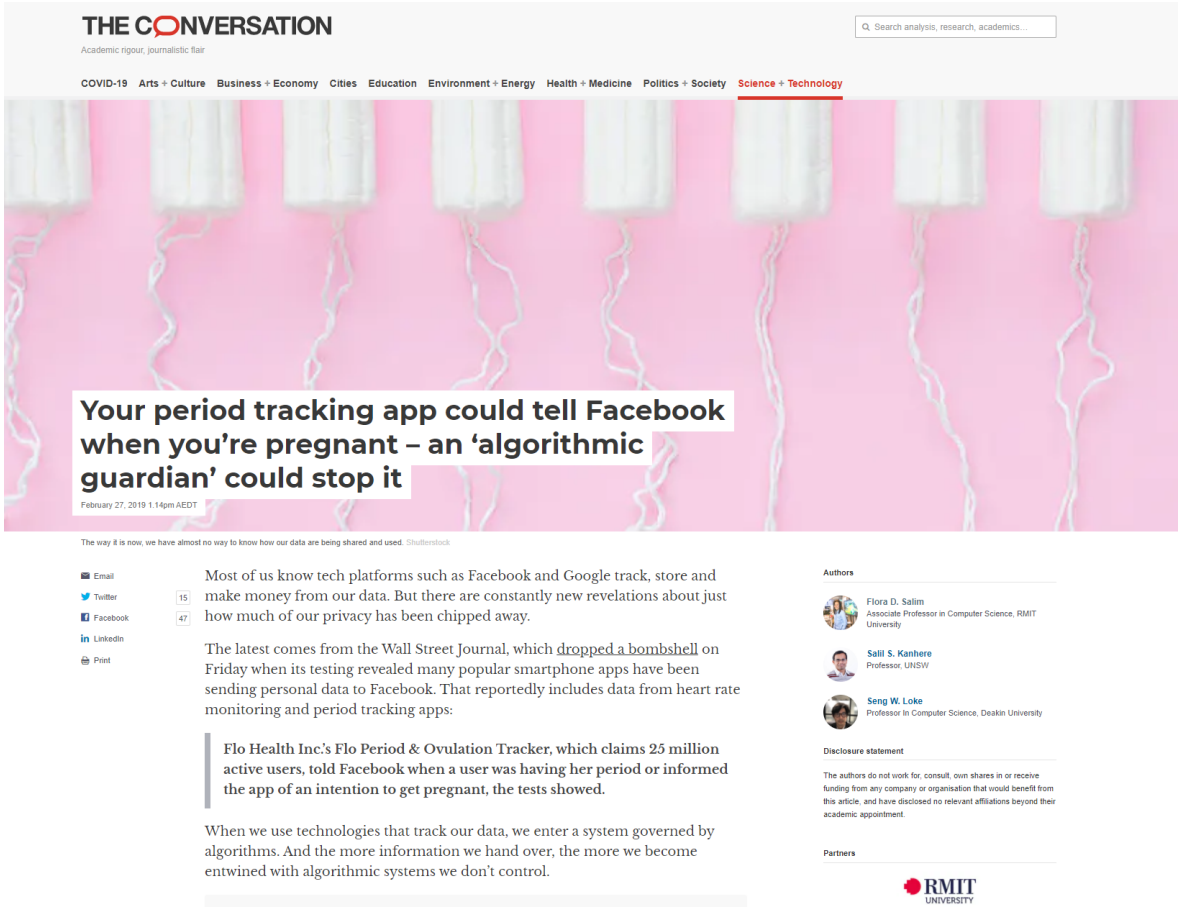
Apps that track you











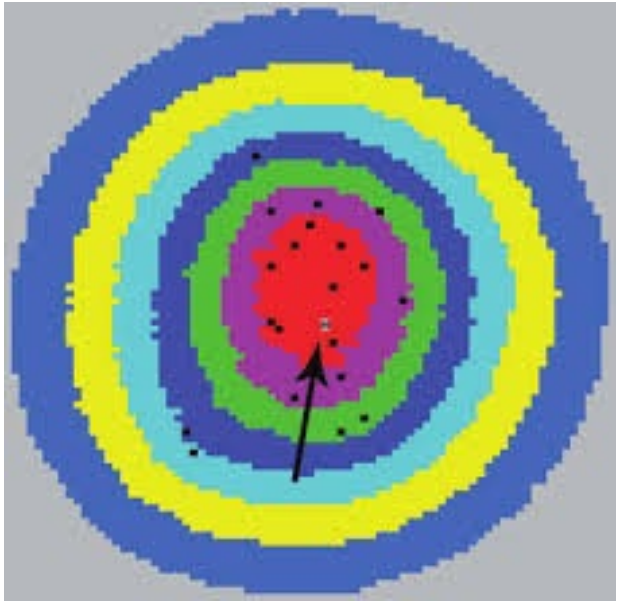
https://theconversation.com/your-period-tracking-app-could-tell-facebook-when-youre-pregnant-an-algorithmic-guardian-could-stop-it-111815

CS[1]: Locating people via geo-profiling

A recently published research uses geo-profiles generated by publicly available information to (possibly) identify the artist Banksy, who has chosen to remain anonymous. The study was framed as an investigation of the use of geo-profiling to solve a “mystery of modern art.” The authors suggest that these methods could be used by law enforcement to locate terrorist bases based on terrorist graffiti.

CS[1]: Locating people via geo-profiling

- Geo-profiling
  - a criminal investigative methodology that analyzes the locations of a connected series of crimes to determine the most probable area of offender residence.
- Assumptions
  - Not in buffer zone
  - Not too far from home



CS[1]: Locating people via geo-profiling

- Lessons to learn
  - How non-personal data will be shared and with whom.
  - How non-personal data could be used in conjunction with other data about the same individual.
  - The use of privacy-by-design: ensure privacy is proactively managed and addressed through organisational culture, practices, processes and systems.

CS[2]: Targeted Marketing through Face Recognition

At present, advertisements are made more personal by using facial recognition technology. In 2014, several companies are already bringing these ideas to (digital) life. Companies like UK grocer Tesco has introduced a digital screen outside its Express store in Lincoln, allowing it to promote its own ranges and engage with visiting customers. These screens will use inbuilt cameras equipped with facial recognition algorithms to ascertain the age and gender of individual shoppers.

A Californian startup called Emotient meanwhile focuses on the area of facial expression analysis. Incorporated into next-generation TVs by way of a webcam, this technology could potentially be used to monitor viewer engagement levels with whatever entertainment is placed in front of them.

CS[2]: Targeted Marketing through Face Recognition

Image from https://www.essentialretail.com/news/53c530217bd0c-tesco-using-digital-screen-technology-to-promote-special-offers/

CS[3]: Social media and the loss of confidentiality

- The curly fry conundrum: Why social media “likes” say more than you might think

https://www.ted.com/talks/jennifer_golbeck_your_social_media_likes_expose_more_than_you_think/transcript?language=en

CS[4]: Retailers' Predictions

How Target uses habits in its stores

https://www.youtube.com/watch?v=RC5HNTj3Dag

CS[5]: Mood and Perception Manipulation

A controversial research study (Kramer, Guillory, and Hancock 2014) used Facebook's platform to demonstrate that users' moods can be manipulated by filtering their feeds (comments, videos, pictures and web links posted by their Facebook friends). The social media company altered the news feeds (the main page users land on for a stream of updates from friends) of nearly 700,000 users. Feeds were changed to reflect more "positive" or "negative" content, to determine if seeing more sad messages makes a person sadder. The study shows that reducing exposure to feeds with positive content led to the user posting fewer positive posts, and the same pattern occurred for negative content.

CS[5]: Mood and Perception Manipulation

- Not provided informed consent
- ‘filtering practices’
- How about if it goes beyond targeted marketing?
  - fake news phenomenon
- Code of practices
  - Require new controls to ensure user trust
  - Informed of advertising tactics, e.g., similar to the EU Cookie Law
- Technologies:
  - Fake news/videos detection

Data breaches/leakage

A data breach happens when personal information is accessed, disclosed without authorization, or is lost, for example:

- a USB or mobile phone that holds an individual's personal information is stolen
- a database containing personal information is hacked
- someone's personal information is sent to the wrong person

Cost of a Data Breach

USD 3.92 million

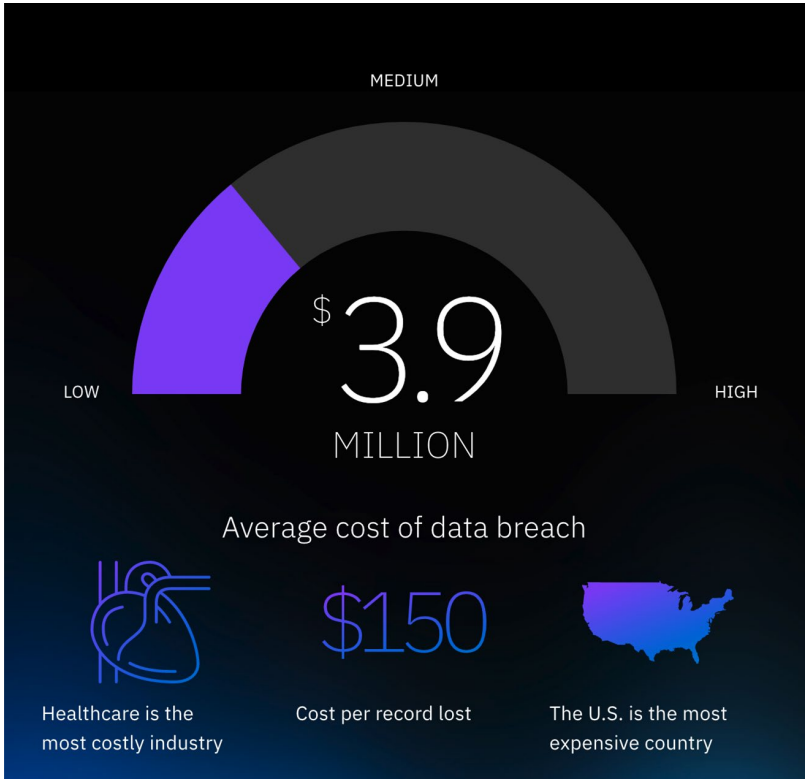
Average total cost of a data breach

United States

Most expensive country: USD 8.19 million

Healthcare

Most expensive industry: USD 6.45 million



Source from https://www.ibm.com/security/data-breach

How long is the lifecycle of a data breach?

Source from https://www.ibm.com/security/data-breach

The longtail costs of a data breach

Source from https://www.ibm.com/security/data-breach

CS[6]: Equifax data breach

In 2017 Equifax, a US-based credit reporting agency, experienced a data breach affecting at least 145.5 million individuals, with various degrees of sensitive personal information compromised. In addition, due to the huge number of people affected, it took several weeks to identify the individuals and notify the public that the breach had occurred. The cost of the breach was estimated to be in the realm of US$275 million.

How the Massive Equifax Data Breach Happened - SciShow [0:00 - 2:50]
https://www.youtube.com/watch?v=_6QbsIgpw8U

Open data

- What is open data
- Benefits of open data
  - Performance
  - Economy
  - Social welfare

Examples of open data in action

Image source from https://www.europeandataportal.eu/en/training/what-open-data

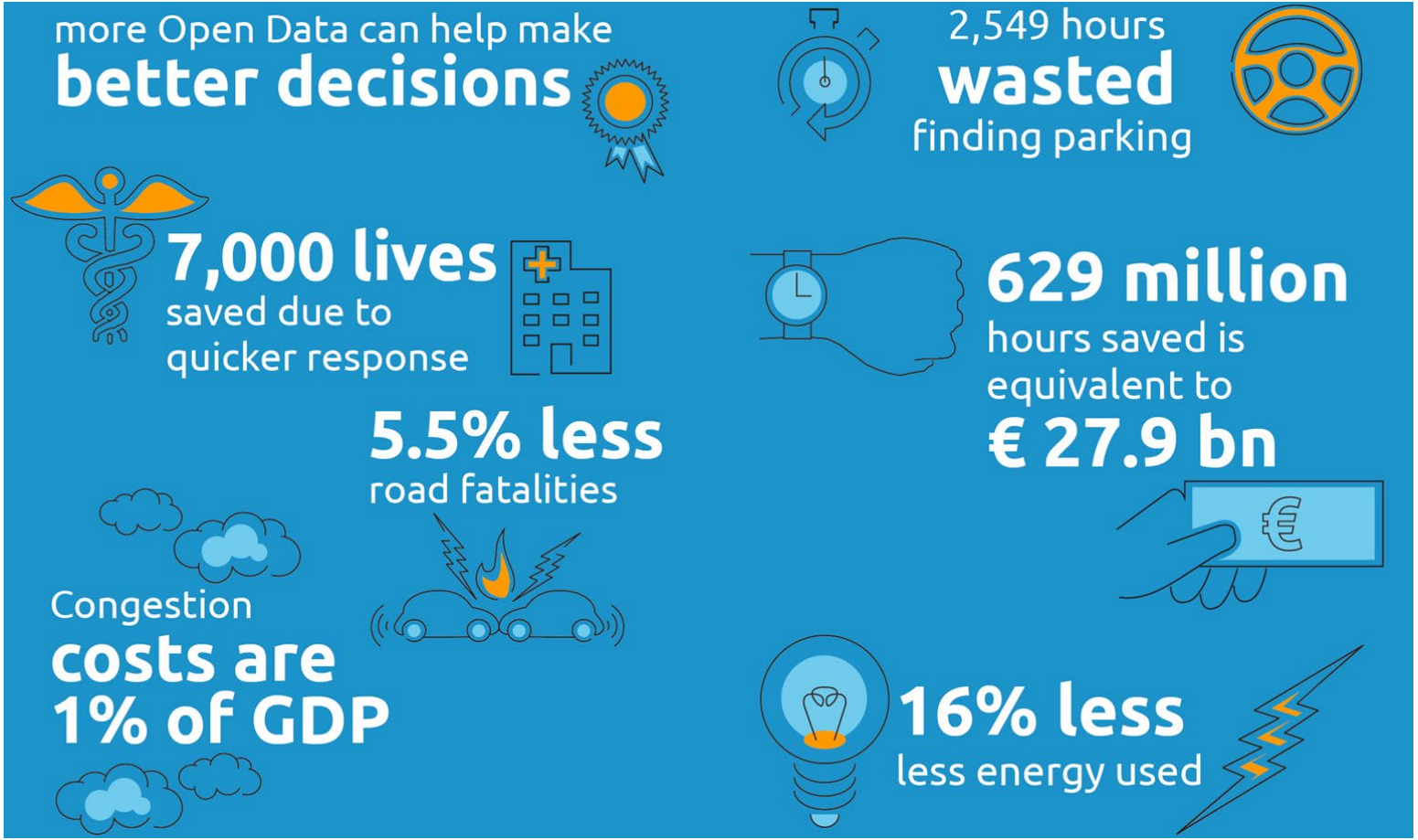


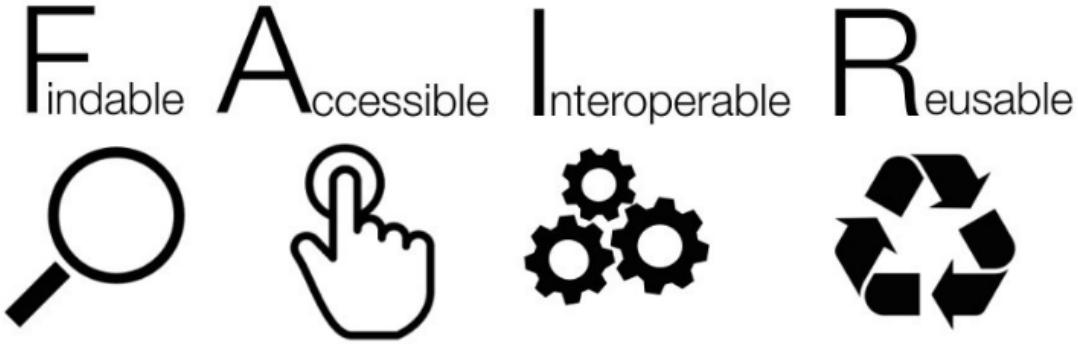
Image source from https://www.europeandataportal.eu/en/training/what-open-data

How to Use Open Data

- Your purpose
  - Define your purpose
  - Identify Data labels
- Open licence
  - Check Openness
  - Check Attribution requirements
  - Check Share-Alike requirements
- File format
- Data Quality

Source from https://www.europeandataportal.eu/en/training/what-open-data

FAIR Principles of Open Data



https://www.nature.com/articles/sdata201618
https://www.go-fair.org/fair-principles/

Re-identification

- Location datasets
- Medical data
- Social media data

CS[7]: De-identified Medicare and PBS open data

On 1 August 2016, the Department of Health published on data.gov.au a collection of Medicare Benefits Schedule and Pharmaceutical Benefits Schedule related data. The data consisted of claims information for a 10% sample of people who had made a claim for payment of Medicare Benefits since 1984, or for payment of Pharmaceutical Benefits since 2003.

A range of steps were taken by the Department of Health to de-identify the dataset before its public release. However, one month after the dataset was published, researchers in the University of Melbourne identified a weakness in the technique used to encrypt Medicare service provider numbers in the dataset, allowing the encryption to be reversed…….

[Discussion]

Do medical data analytics has different data curation requirements to data analytics for astrophysics?

ADDRESSING OPEN DATA CHALLENGES

- Re-identification
  - Technical, legal, and administrative safeguards
  - Expert evaluation
  - Tools to de-identify unstructured or dynamic data types
  - Policies and procedures for evaluating re-identification risk across databases
- Data Quality: accurate, complete, and current
- Equity and Fairness.
- Public Trust

Reference from https://fpf.org/2018/01/31/if-you-cant-take-the-heat-map-benefits-risks-of-releasing-location-datasets/

CS[8] Strava Global Heatmap

A portion of the Strava Labs heat map from Kandahar Airfield in Afghanistan, made by tracking activities. (Screenshot)

https://www.washingtonpost.com/world/a-map-showing-the-users-of-fitness-devices-lets-the-world-see-where-us-soldiers-are-and-what-they-are-doing/2018/01/28/86915662-0441-11e8-aa61-f3391373867e_story.html

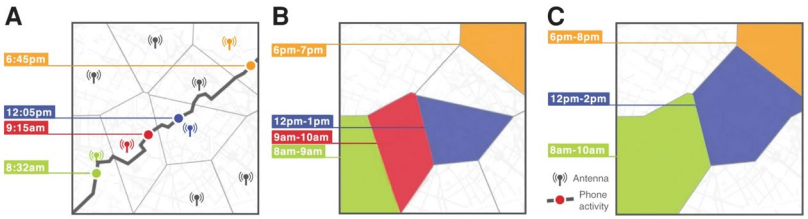
Unique in the Crowd
The privacy bounds of human mobility

“Yet, if individual's patterns are unique enough, outside information can be used to link the data back to an individual.” (Montjoye et al. 2013, Nature)

https://www.nature.com/articles/srep01376

Lesson: Coarsening, scaling, and aggregation alone are not sufficient for ensuring anonymity

From: Unique in the Crowd: The privacy bounds of human mobility



(A) Trace of an anonymized mobile phone user during a day.The dots represent the times and locations where the user made or received a call. Every time the user has such an interaction, the closest antenna that routes the call is recorded. (B) The same user's trace as recorded in a mobility database. The Voronoi lattice, represented by the grey lines, are an approximation of the antennas reception areas, the most precise location information available to us. The user's interaction times are here recorded with a precision of one hour. (C) The same individual's trace when we lower the resolution of our dataset through spatial and temporal aggregation. Antennas are aggregated in clusters of size two and their associated regions are merged. The user's interaction are recorded with a precision of two hours. Such spatial and temporal aggregation render the 8:32 am and 9:15 am interactions indistinguishable.

Privacy Protection and Security

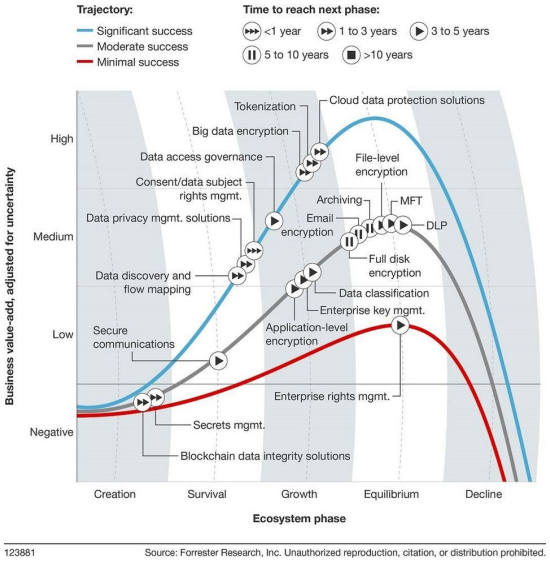
“Throughout their lifecycle, AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.”

Privacy Protection and Security

- Identification of potential security vulnerabilities, and
- Assurance of resilience to adversarial attacks.
- Security measures should account for unintended applications of AI systems, and potential abuse risks, with appropriate mitigation measures.

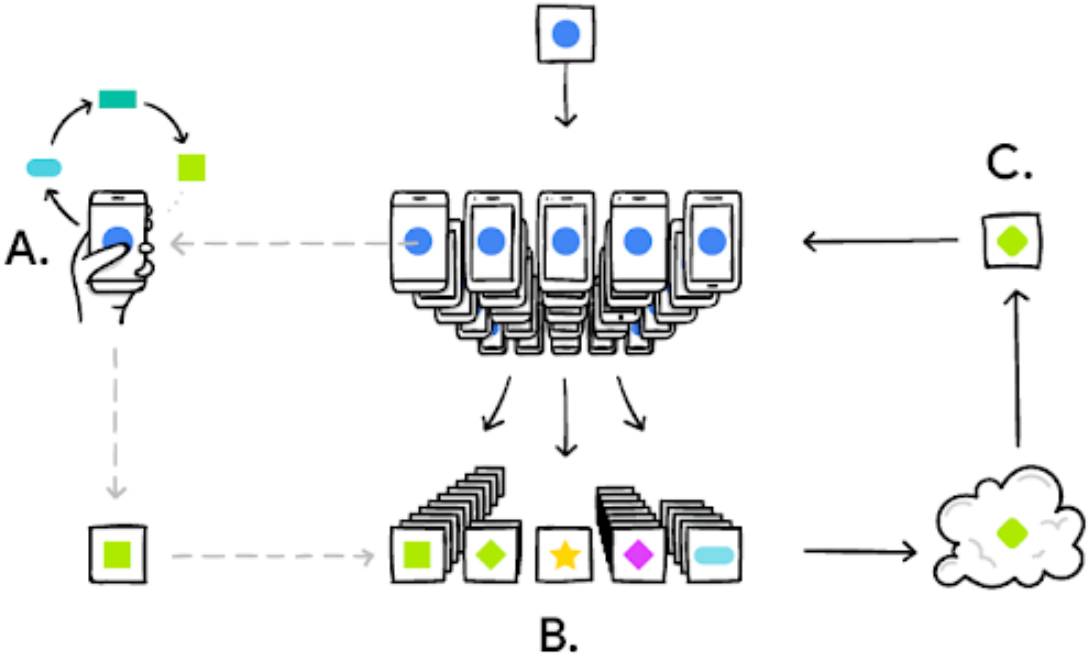
Data Security And Privacy Technologies

- Cloud data protection (CDP)
- Tokenization
- Two factor authentication
- Big data encryption
- Data access governance
- Consent/data subject rights management
- Data privacy management solutions
- Data discovery and flow mapping
- Data classification
- Enterprise key management (EKM)
- Application-level encryption



Source from https://www.forbes.com/sites/gilpress/2017/10/17/top-10-hot-data-security-and-privacy-technologies/#7a32595f6b3f

The Federated Learning Approach



Source from https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

Federated Self-Supervised Learning of Multi-Sensor Representations for Embedded Intelligence

IEEE INTERNET OF THINGS JOURNAL 1

Federated Self-Supervised Learning of Multi-Sensor Representations for Embedded Intelligence

Aaqib Saeed, Flora D. Salim *Member, IEEE*, Tanir Ozcelebi *Member, IEEE*, and Johan Lukkien *Senior Member, IEEE*

***Abstract*—Smartphones, wearables, and Internet of Things (IoT) devices produce a wealth of data that cannot be accumulated in a centralized repository for learning supervised models due to privacy, bandwidth limitations, and the prohibitive cost of annotations. Federated learning provides a compelling framework for learning models from decentralized data, but conventionally, it assumes the availability of labeled samples, whereas on-device data are generally either unlabeled or cannot be annotated readily through user interaction. To address these issues, we propose a self-supervised approach termed *scalogram-signal correspondence learning* based on wavelet transform to learn useful representations from unlabeled sensor inputs, such as electroencephalography, blood volume pulse, accelerometer, and WiFi channel state information. Our auxiliary task requires a deep temporal neural network to determine if a given pair of a signal and its complementary viewpoint (i.e., a scalogram generated with a wavelet transform) align with each other or not through optimizing a contrastive objective. We extensively assess the quality of learned features with our multi-view strategy on diverse public datasets, achieving strong performance in all domains. We demonstrate the effectiveness of representations learned from an unlabeled input collection on downstream tasks with training a linear classifier over pretrained network, usefulness in low-data regime, transfer learning, and cross-validation. Our methodology achieves competitive performance with fully-supervised networks, and it outperforms pre-training with autoencoders in both central and federated contexts. Notably, it improves the generalization in a semi-supervised setting as it reduces the volume of labeled data required through leveraging self-supervised learning.**

***Index Terms*—self-supervised learning, deep learning, federated learning, embedded intelligence, low-data regime, sensor analytics, learning representations.**

I. INTRODUCTION

LEARNING representations with deep neural networks have made tremendous improvements in the last few years on challenging real-world tasks [1]–[4], thanks to the emergence of massive datasets. In particular, the wealth of sensory data from the Internet of Things (IoT) devices are only recently being leveraged for tackling important problems in understanding context, user monitoring, health, and other predictive analytics tasks, e.g., for emotional well-being [5], [6], sleep tracking [7], and physical activity detection [8]. The success is mainly attributed to the supervised methods that utilize labeled datasets for training models in a central environment. In contrast, learning models from unlabeled decentralized data still presents a major challenge. Obtaining large, well-curated sensory data from edge devices is especially difficult owing to issues like user privacy, the prohibitive cost of labeling, bandwidth limitations, network connectivity, and the diversity of device types [9]. These factors make it significantly challenging to harness abundant data on remote devices for learning semantic features with standard supervised approaches.

To highlight the challenges associated with learning a generalizable model for a particular use case, consider this illustrative example. Let us assume that we aim to develop a robust sleep stage classification model that can be used for a larger population of users. The standard methodology is to learn a supervised model and requires example-label pairs for providing supervision so that a model can differentiate among instances of multiple classes through learning underlying patterns in the input. The procedure begins with the data collection to monitor hundreds of users for electroencephalography (EEG) or other signals as they progress through various stages of sleep and accumulate the multi-sensor data in a centralized (data center) repository for further analysis. The next step is then to get the aggregated inputs annotated by the sleep expert (i.e., generally a professional trained in analyzing physiological signals) for the sleep classes, such as wake, N1, N2, N3, and rapid eye movement. Then, the learning and evaluation phase involves several iterations of improving the model performance. Lastly, the model is deployed in the wild for user monitoring. The process of model development, from data collection to annotation, could be extremely costly owing to the difficulty in setting up an experimental (data collection) protocol. Furthermore, the domain expertise required for the labeling could be severely limited. This problem is exacerbated by the need of supervised deep neural network models for a massive amount of labeled data to learn discriminative features. It becomes painstakingly difficult to inspect and annotate hundreds of thousands of hours of multi-sensor data. Therefore, in practice, limited-sized sensor data are collected and labeled for learning the model, which could further affect its generalization. The important point to note here is that the explained strategy is only applicable when the users

Aaqib Saeed, Tanir Ozcelebi and Johan Lukkien are with the Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands. E-mail: {a.saeed, t.ozcelebi, j.j.lukkien}@tue.nl. Correspondence should be addressed to Aaqib Saeed.

Flora D. Salim is with the School of Science, RMIT University, Melbourne Australia. She co-directs the RMIT Centre for Information Discovery and Data Analytics (CIDDA). E-mail: flora.salim@rmit.edu.au.

This work is funded by SCOTT (www.scott-project.eu) project. It has received funding from the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement No 737422. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and Austria, Spain, Finland, Ireland, Sweden, Germany, Poland, Portugal, Netherlands, Belgium, Norway.

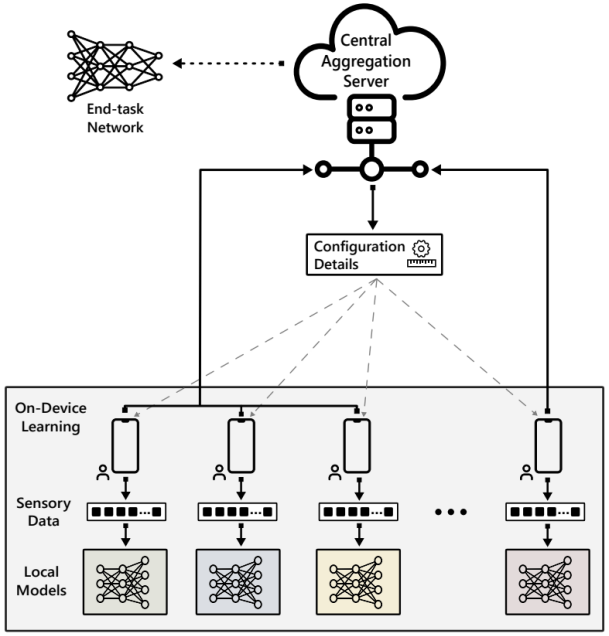


Fig. 3: Overview of federated learning framework. A central server dispatches a randomly initialized model and other training configuration details to the selected clients' devices, as depicted by dashed gray lines. The clients train local models on their private data and send the models back to the server illustrated with solid black lines. The models are aggregated to produce a unified model that is used for the end-task.

https://arxiv.org/pdf/2007.13018.pdf

Breakout Case Study

What's next...

Shopping Intent Recognition and Next Location Prediction from Cyber-Physical Activities via Wi-Fi Logs

Manpreet Kaur[1,2], **Flora Salim**[1], Yongli Ren[1], Jeffrey Chan[1], Martin Tomko[3], Mark Sanderson[1]

[1] Computer Science and Information Technology, RMIT University, Melbourne, Australia

[2] Tableau Software, US

[3] Department of Infrastructure Engineering, Melbourne University

Correspondence: flora.salim@rmit.edu.au; @flosalim; http://florasalim.com

RMIT UNIVERSITY

Online Retail vs. Brick and Mortar

© Westfield Sydney

TRIIBE (Tracking Indoor Information Behaviour)

Passive WiFi-based sensing and ISP logging

- 67 Wi-Fi access points (APs) across 90,000 square meters
- 18- million rows of Wi-Fi access logs over a 1-year period
- over 120,000 anonymized users at an inner city shopping mall

Ren, Y., Tomko, M., Salim, F.D., Ong, K., Sanderson, M. (2015). “Analyzing Web Behavior in Indoor Retail Spaces. Journal of the Association for Information Science and Technology (JASIST). Vol. 68, issue 1, Jul 2015.

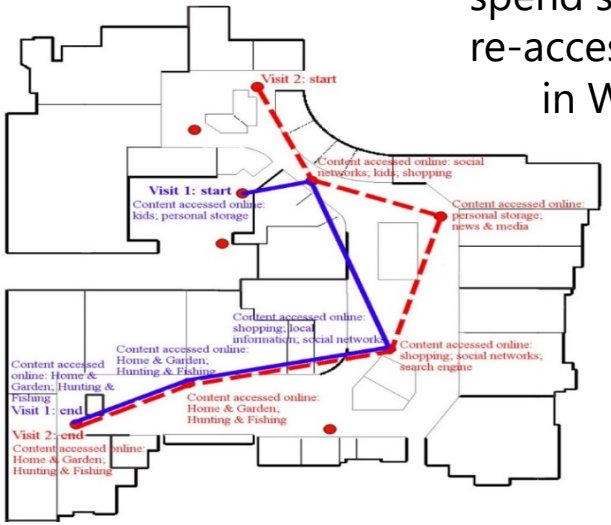
Discuss

- What are some of the open problems faced by the data owner (mall operator)?
- Discuss the wide range of the potential utility of this data
- Are there any private data / attributes that could be highly useful to solve some of the problems?
- Provide some examples of potential privacy and security issues
- How could these issues be addressed?

Initial Analysis

TABLE 1. Aggregate statistics of the AL, BL, and QL.

Wi-Fi AP Log (AL)	
No. of users	120,548
No. of AP association	907,084
No. of user visits	261,369
Web Browsing Log (BL)	
No. of users browsing	70,196 (58.3% of AL users)
No. of issued URLs	18,088,018
No. of user visits	139,004
Query Log (QL)	
No. of users searching	11,169 (9.3% of AL users)
No. of queries	119,196
No. of query sessions	20,637



In repeated visits, people tend to
- revisit similar mall locations, and
- spend similar amount of time re-access similar Web content in Web categories, e.g.
  - Social Networking
  - Home & Garden
  - Hunting & Fishing

Ren, Y., Tomko, M., Salim, F.D., Ong, K., Sanderson, M. (2015). “Analyzing Web Behavior in Indoor Retail Spaces. Journal of the Association for Information Science and Technology (JASIST). Vol. 68, issue 1, Jul 2015.

Cyber Physical Social (CPS) Contexts → Behaviours

Cyber

online activities

Physical

movement activities

Social

group activities

Predicting Visitor Demographics from Their CPS Behaviours

–age: 18-24, 25-39, 40-54, 55+
–education level: Secondary/high school, Honours degree?
–income: 0-$18,200, $18,201-$37,000, $37,001-$80,000, $80,000+
–parental status: having kids?
–shopper category: Inner or Rest of Sydney resident, CBD Worker, Domestic tourist, International tourist

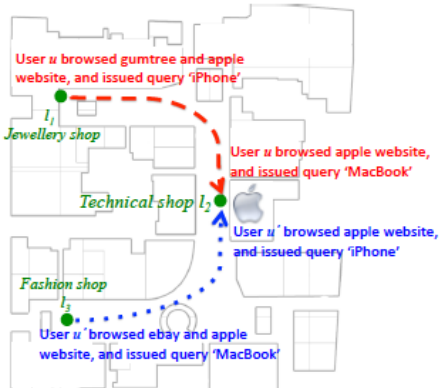
Physical: frequency, weekdays, duration, interests in shop categories

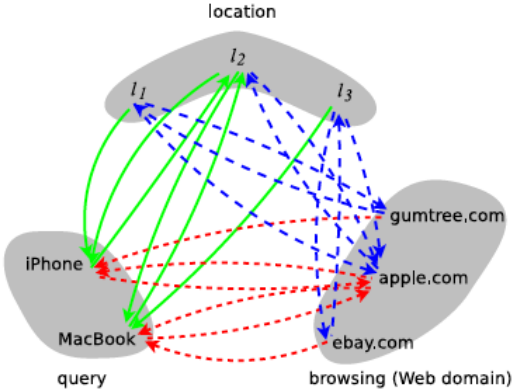
Cyber: WiFi frequency, search frequency, what to browse/search

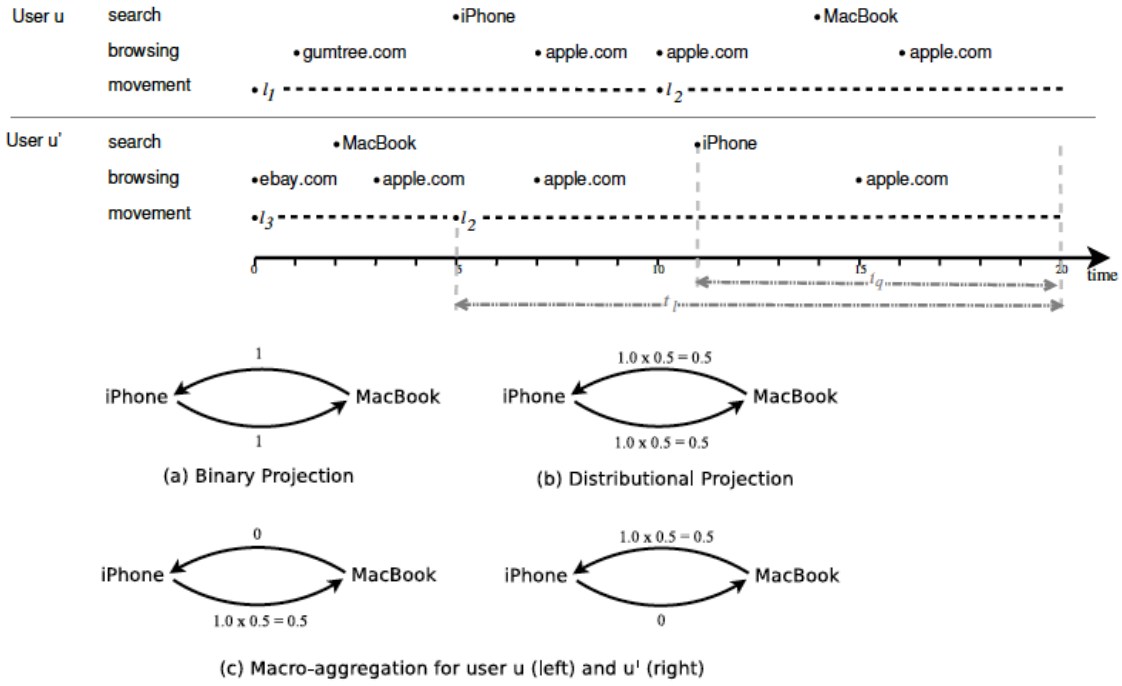
Social: single, with kids, with another adult, in a group

Yongli Ren, Martin Tomko, Flora D Salim, Jeffrey Chan, Mark Sanderson. "*Understanding the Predictability of User Demographics from Cyber-Physical-Social Behaviours in Indoor Retail Spaces*". EPJ Data Science 7(1), 2018.

Location-Query-Browse Graph for Contextual Recommendation







Y. Ren, M. Tomko, F. Salim, J. Chan, C. L.A. Clarke and M. Sanderson. "*A Location-Query-Browse Graph for Contextual Recommendation*". IEEE Transactions on Knowledge and Data Engineering (TKDE). 30(2): 204-218 (2018)

TRIIBE (Tracking Indoor Information Behaviour)

Australian Research Council (ARC) Linkage 2014-2018

Selected Publications

- Kaur MA, Salim FD, Ren Y, Chan J, Tomko M, Sanderson M. 'Joint Modelling of Cyber Activities and Physical Context to Improve Prediction of Visitor Behaviors'. ACM Transactions on Sensor Networks (TOSN), 2020
- Kaur, M. Salim, F. Ren, Y. Chan, J. Tomko, M. and Sanderson, M. 2018, 'Shopping intent recognition and location prediction from cyber-physical activities via Wi-Fi logs', 5th ACM Conference on Systems for Built Environments (BuildSys 18), pp. 130-139
- Ren, Y., Tomko, M., Salim, F., Chan, J., Clarke, C. L. A., and Sanderson, M., "A Location-Query-Browse Graph for Contextual Recommendation". *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2018.
- Ren, Y., Tomko, M., Salim, F.D., Chan, J., Sanderson, M.. "Understanding the Predictability of User Demographics from Cyber-Physical-Social Behaviours in Indoor Retail Spaces". *EPJ Data Science* 7(1), 2018.
- B. Priyogi, M. Sanderson, F. Salim, J. Chan, M. Tomko, Y. Ren. Identifying In-App User Actions from Mobile Web Logs. *Proceedings of the 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2018.
- Ren, Y., Salim, F. D., Tomko, M., Bai. Y. B., Chan, J., Qin, K. K., Sanderson, S. (2017), "D-Log: A WiFi Log-based differential scheme for enhanced indoor localization with single RSSI source and infrequent sampling rate", *Pervasive and Mobile Computing*, vol. 37, June, 2017, Elsevier, pp. 94–114.
- Ren, Y., Tomko, M., Salim, F.D., Ong, K., Sanderson, M. (2015). "Analyzing Web Behavior in Indoor Retail Spaces. *Journal of the Association for Information Science and Technology (JASIST)*. Vol. 68, issue 1, Jul 2015.
- Ren, Y., Tomko, M., Ong, K., Sanderson, M. : How People Use the Web in Large Indoor Spaces. In proceedings of the *23rd ACM Conference on Information and Knowledge Management (CIKM)*, 2014

Westfield

Australian Government

Australian Research Council

What to do this week?

- Week 5 Deadline – reminder
  - Micro-credential “Presenting Using Story”
  - Task 2 Milestone 1 – Group Project Initial Submission
- Read materials