

Data Science for Social Good

Experience Sharing

Dah Ming Chiu

Data Science Research Centre, CIHE

Feb 15, 2023

What is data science

- Analyzing data help us
 - Understand what is going on
 - Make better decisions
- Many examples, everywhere:
 - Daily life example: how to pick an avocado?
 - Medical example: is the vaccine safe?
 - Government's use of data and policy setting: GDP, CPI, UR, LE..., budget...
 - Business intelligence: banks, insurance, e-commerce, online social media
- Methods
 - Probability, Statistics
 - Computer science, data mining, clustering, machine learning

Types of jobs for data science

- Prediction
 - What factors affect an outcome?
 - Correlation analysis – help judge which factors are more important
- Hypothesis testing
 - For something that affects lots of people, e.g. drug, minimal wage
 - Carefully do some small scale experiment, estimate consequence with confidence
- Optimizing a complex system using data
 - Maximize view count, click-through-rate
 - Often machine learning applied to learn more complicated relationships
- Understanding the causes
 - How to explain some outcome? E.g. Why HK has highest Life Expectancy?
 - It is outcome from a very complex system
 - Need to get causal relationships right

The popularization of data science

- Due to continual computerization, IT development
 - Vast amount of data automatically collected (different from census, surveys, measurements)
 - Creating opportunities to study various problems
 - Note:
 - The automatically available data may not help answer your questions
 - May have privacy, and fairness issues
- Great commercial success (e-commerce, social media...)
 - Resulting in market demand
- Data science: merging methods from different disciplines
 - What is data science, different people may explain differently

Data Science Research Centre at CIHE

- Formed 2022: <https://cihe.edu.hk/dsrc>
- Mission
 - Promote interdisciplinary research, initially focusing on social science
 - Collaborate with social service sector, do research with social impact
- Progress so far
 - Formed a small team of data scientist/engineers
 - Doing several projects, collaborating with NGO, social enterprise
 - Collaborate with social science faculty, on policy related research
 - With short course on data science
 - Running a online talk series

Project 1: data warehousing

- Collaborating with CFSC
- NGOs may provide several services
 - Each service has its own database
 - Data may not be in standardized format
 - Data needs cleaning
- Data warehouse
 - Standardize data, put data into a common place (logically)
 - Dashboard for visualization and analysis – “know your customer”
 - Helpful for operation evaluation, problem/root cause analysis, planning


Project 2: recommendation system (ongoing)

- Collaborating with Big Silver
 - Big Silver provides a mobile app and learning material for carers
- Recommendation system for individuals
 - Suitable learning material, topics/posts/threads, friends
- Recommendation system is a well-studied problem
 - Various machine-learning algorithms can be applied for given KPIs
 - In the social context, objective is more vague
 - Adopt a subjective evaluation, user feedback based approach

The screenshot displays the web interface for the 'Recommender System For Carers Chat'. On the left is a navigation sidebar with the following items: Home, About, Dashboard, Data, Users, Items, Search User By Serial, Search Item By Serial, Check Items By User History Views, Upload File, Recommend (highlighted with a star), Using Method A (Item CF), Item CB (New), and Using Method C (User CF). The main content area features a header with 'version: 0.2.3' and the title 'Recommender System For Carers Chat'. Below the header is a large circular logo with the Chinese characters '喺傾' (Choi King). The main text reads: '歡迎來到喺傾的推薦系統' (Welcome to the Chai King recommendation system) and '照顧者可以來休息，與朋友聊天，好好感受自己，準備照顧的旅程。' (Carers can come to rest, chat with friends, feel good about themselves, and prepare for the journey of care). Under the heading 'App功能' (App Features), it states: '情緒支持 - 照顧路上百般感受，不是很多人明白，這裡都是願意了解的朋友，可以抒發感受、發表意見。實用可靠 - 每月一專題課程，按自己步伐掌握所需知識，睇到又聽到，知道未來的照顧路線圖，及早準備。' (Emotional support - various feelings on the care journey, not many understand, here are friends willing to understand, can express feelings, share opinions. Practical and reliable - monthly special course, follow your own pace to master needed knowledge, see and hear, know the future care route map, prepare early). The '推薦功能' (Recommendation Features) section lists: '基於內容' (Content-based), '基於協同過濾' (Collaborative filtering), '基於平台評分' (Platform rating), '基於熱度冷卻' (Heat/cold), and '實時監控' (Real-time monitoring). The '合作方' (Partners) section lists: '太銀' (Tai Yin), '明愛專上學院' (Ming Ai Specialized College), and '香港中文大學' (The Chinese University of Hong Kong).

Project 3: SDU rental analysis (ongoing)

- Collaborating with HKCSS, they collected data about Subdivided Units (SDU)
- Some of the questions:
 - What are the factors determining the rent, for different areas?
 - Can suitable rent be predictable?
 - How to provide useful information for stakeholders?
- Progress
 - Learned rent policies in Netherland project
 - Met others who studied this problem in the past
 - Tried correlation analysis for different factors and locations; and data visualization



The screenshot shows a website interface for 'Subdivided Flats in Hong Kong'. It features two photographs of narrow hallways with doors. The first photo has the Chinese characters '劏房' (SDU) overlaid in red. To the right is a table listing tenants:

租客甲 Tenant A	租客乙 Tenant B	租客丙 Tenant C	租客丁 Tenant D	
租客戊 Tenant E	租客己 Tenant F	租客庚 Tenant G	租客辛 Tenant H	租客壬 Tenant I

Below the table is a search bar with '简体中文' and '执行' buttons, and a 'Login or register' link. The main content area is titled 'CIHE DSRC SDU Agent' and includes an 'Operation Menu' dropdown, a breadcrumb '首页 / Sdfs / Sdfs', and a section for 'Sdfs' with a search bar and a map of Hong Kong showing several red location markers.

Project 4: data extraction system (ongoing)

- Extract data automatically from scanned documents
 - E.g. Surveys, forms, receipts, bank statements
 - Automatic extraction saves lots of man power
- Available tools from open source and cloud
 - OCR, machine learning for document understanding
 - Our approach: human-in-the-loop auto-extraction

Social policy research

- Life expectancy (LE) and Healthy life expectancy (HLE)
 - Wrote a paper about why HK does well in LE
 - Need to understand the situation with HLE
 - Are there data from social, medical and insurance services that can help monitor HLE (need due to increasing aging population)
- Can we understand marriage rate decline (in China) using data science?
 - Gross model
 - Learn parameters from available data

Conclusions

- Data science for social good
 - Data science widely and successfully used commercially
 - Opportunities in applying data science in social services?
 - For single NGO, business intelligence (KYC) type of applications
 - Across NGOs, data may help policies, smart city
- Challenges
 - Data privacy
 - Fairness