NCRM

COMPUTATIONAL SOCIAL SCIENCE

PROGRAMMING SKILLS FOR SOCIAL SCIENTISTS

Dr. James Allen-Robertson - University of Essex

jallenh@essex.ac.uk

@Minyall github.com/Minyall
PERSONAL TRAJECTORY

Sociology BA, MA and PhD

- Cyberculture
- Piracy & Copyright
- Utilising digital documents for qualitative research

Materiality Media Theorist

- Design in media systems
- User agency
- Memory
- Algorithmic Power Relations


PERSONAL TRAJECTORY

Sociology BA, MA and PhD
- Cyberculture
- Piracy & Copyright
- Utilising digital documents for qualitative research

Materiality Media Theorist
- Design in media systems
- User agency
- Memory
- Algorithmic Power Relations

Computational Social Science
- Building own tools for data analysis.
- Automated collection of data at scale.
- Focus particularly on text data retaining my qualitative roots.
- Much more collaborative work providing skills and tools to colleagues.
COMPUTATIONAL SOCIAL SCIENCE EXAMPLES
LARGE SCALE DOCUMENT CLUSTERING

- Uber Driver Discourse
  - 50K threads from Uber driver and Taxi driver forums.
  - Document Similarity + Dimensionality Reduction + Clustering
- What are the concerns of workers under algorithmic management?
- Where do Uber Drivers and Taxi Drivers converge and diverge in their concerns?
Computational Social Science Examples

Topic Modelling for Nuanced Discourse

- Topic Modelling the Investigatory Powers Act (HRBDT)
  - 219 submissions to the consultation categorised by organisation type.
  - NMF Topic Modelling to discover trends in discourse across groups.
  - Focusing in on HR discourse within tech firms. How is human rights being intertwined with tech?
  - In progress
Topic Modelling the Investigatory Powers Act (HRBDT)

- 219 submissions to the consultation categorised by organisation type.
- NMF Topic Modelling to discover trends in discourse across groups.
- Focusing in on HR discourse within tech firms. How is human rights being intertwined with tech?
Finding Trump Tweets with Neural Networks.

- Trained a Neural Network to recognise if a Tweet was written by Trump or his Staff.
- 76% accurate on unseen tweets.
- 2 day side project.
- Can AI support us in verifying the source of information?

COMPUTATIONAL SOCIAL SCIENCE EXAMPLES
CLASSIFYING DOCUMENTS WITH A NEURAL NETWORK

Finding Trump Tweets with Neural Networks.
- Trained a Neural Network to recognise if a Tweet was written by Trump or his Staff.
- 76% accurate on unseen tweets.
- 2 day side project.
- Can AI support us in verifying the source of information?

Collected 3,219 Tweets using #NoTap.

Protest movement against a new pipeline project in Italy.

Utilised a Twitter listener to gather Tweets live, and then extract the images from them for qualitative analysis.

Aim was to study the role of the visual in online protest.

Di Ronco, Anna, James Allen-Robertson, and Nigel South. 'Representing environmental harm and resistance on Twitter: The case of the TAP pipeline in Italy', Crime, Media, Culture, 0: 1741659018760106.
Network relations, bot detection and discourse divergence in the Italian Vaccination Debate.

Automated coding of image content using Neural Network based feature identification > Visual culture in the Catalanian Referendum (4m Tweets).

Mapping the Dark Web - Web Crawler + Trained Classifier > Ethical Issues in good abundance!

Warning: Enthusiasm for computational methods and collaboration may lead to project overload
PROGRAMMING FOR SOCIAL SCIENTISTS

INCENTIVES

- Existing software may not scale or have the feature flexibility for your project.
- Automated data extraction opens up a range of new sources and questions / possibilities.
- Currently very fast development in computer science - new techniques to be repurposed for research.
- Allows you to build custom tools specific to your research question.
- All components are free!!
PROGRAMMING FOR SOCIAL SCIENTISTS

LEARNING

- Basics from a 2 hour workshop at a conference...

- Online resources
  - Blogs
  - Tutorials
  - Code Repositories

- Online courses
  - Sage Campus (Academia Focus)
  - DataCamp - (Interactive)
  - DataQuest - (Interactive)
  - ...

NCRM - COMPUTATIONAL SOCIAL SCIENCE
PROGRAMMING FOR SOCIAL SCIENTISTS

CHALLENGES

- Translating Techniques to Academia
  - Often tutorials have Industry focus.
  - Sometimes you can utilise tools ‘unconventionally’.

- Statistics!!
  - Ultimately techniques are reliant on statistical principles.
  - Not statistically minded? > Can make statistics easier(!)

- Dirty Data
PROGRAMMING FOR SOCIAL SCIENTISTS

CHALLENGES

» Often each stage of a project requires learning something new.

» Steep initial learning curve.

» Rabbit hole

» Techniques have to change with scale.

» 100 Docs - Text Documents

» 1,000 Docs - DataFrames

» 1M Docs - Databases!

PERSONAL TIMELINE:

2 Weeks > Web Scraper Built
1.5 Year > First Paper
2 Years > Rapid Prototyping and research design.
PROGRAMMING FOR SOCIAL SCIENTISTS

CHALLENGES

- Often each stage of a project requires learning something new.
  - Steep initial learning curve.
  - Rabbit hole
- Techniques have to change with scale.
  - 100 Docs - Text Documents
  - 1,000 Docs - DataFrames
  - 1M Docs - Databases!

PERSONAL TIMELINE:

- 2 Weeks > Web Scraper Built
- 1.5 Year > First Paper
- 2 Years > Rapid Prototyping and research design.
PROGRAMMING FOR SOCIAL SCIENTISTS

BENEFITS

- Collaboration
  - Get involved in a vast range of substantive topics.
  - Talk to colleagues, find out what their approaches are, could they benefit from these techniques? - Not always!
- New frontiers == Favourable Editors (Sometimes)
  - Generally a strong positive response from reviewers and editors for methodological innovation.
- Currently in vogue - strong skills demand outside academia, research councils very favourable.

Data Science

✔

Data Science + Social Science

✔️✔️✔️
PROGRAMMING FOR SOCIAL SCIENTISTS
GETTING STARTED IN PYTHON

- Fundamental
  - Pandas - Data storage and structure
  - SciKit Learn - Classification / Regression / Clustering / Dimensionality Reduction / Model Selection / Preprocessing.
  - Seaborn - Data Visualisation
  - Requests - Web Scraping
  - Natural Language Toolkit - Text mining and linguistics.

www.anaconda.com/download

Integrates code, output, notes and text into a single document. Much more intuitive than a normal IDE.
PROGRAMMING FOR SOCIAL SCIENTISTS
GETTING STARTED IN PYTHON

- Advanced
  - Bokeh - Build Interactive Data Visualisations.
  - Scrapy - Powerful high speed web scraping.
  - Keras - Build and train Neural Networks for supervised machine learning classification.
  - Gensim - Text analysis at very large scale.

Integrates code, output, notes and text into a single document. Much more intuitive than a normal IDE.

www.anaconda.com/download
HELP!

- Stackoverflow.com
- Google Groups
- Github
- Google Search Magic

...or email me
THANKS!

NCRM

Dr. James Allen-Robertson - University of Essex

...or email me

JALENH@ESSEX.AC.UK

THANKS!