Coding Exercise: Research Assistant

April 2024

Shaoda Wang

Goals:

1. Compile a sample of 400 papers from MIT's Computer Science and Artificial Intelligence Lab.
    - URL: https://dspace.mit.edu/handle/1721.1/5458/browse?type=dateissued
    - Timeframe: focus on year 2000 to today.
2. Categorize each author's nationality or ethnicity.
3. Categorize research topics: sub-categories within AI (e.g. facial recognition, speech recognition).
4. Identify patterns of nationality/ethnicity and research topics within AI across time.
5. Randomly pick authors from the data compiled above and further identify 30 authors who are currently faculty members in top-ranked universities in the US. Construct these faculty members' career trajectories outside of academia (e.g., serve as chief scientist at Facebook).
    - You may use sources such as google scholar, faculty members' CVs, acknowledgement in the paper publication, LinkedIn, online news, personal website, institutional webpage, etc.
6. Identify patterns of published research papers of these faculty members across time, before and after the faculty members engage in activities outside of the academia.

Notes:

- This exercise is intentionally designed to be open-ended. Use your own judgment on how best to proceed. For example,
    - The data compilation task should include sufficient information of the paper for future analyses. You should decide which variables to include.
    - You should decide how to best categorize authors' nationality/ethnicity.
    - You should decide how to best categorize research topics.
- Please document the steps you take to compile the data as well as categorize the authors and paper topic. Make sure to document the criteria you use to categorize and the rationale for any decisions you make. If you can automate part of these steps, please document accordingly.
- You should not spend more than two days on this task.
- Please upload a zip file of your final outputs on Canvas. The file should include:
    - A writeup of the steps you took and the results
    - Data file
    - Code files