DSC 180A Data Science Project I Section A02

Wikipedia Edit Wars

Keng-Chi Chang

<kechang@ucsd.edu>

2020-01-15

More on Proposals and Projects

- Set goals for yourself
  - ‣ Put a project on your resume
  - ‣ Submit to a conference
  - ‣ Apply for graduate school
- Less "fancy" compared with other sections
- But many companies are facing these questions (ex. discourage what kind of content, fact-checking), perhaps easier to stand out
- Start to think about it TODAY
- Find ideas through readings, news, etc
- Discuss with me at anytime

Assignment 1 Due Next Friday (Jan 24)

- Instructions on Github repo
- Write a survey of the data (report)
- Develop code for ingesting the data for later use (code)
- Describe and justify the data ingestion process (report)
- Ask me questions if you encounter any, especially on Friday lab sessions (I will be there at least 9-10am, just call me if you didn't sign up)

Discussions

Go to www.menti.com and enter code 61 45 29

The Wisdom of Crowds (Surowiecki)

- If you put together a big enough and diverse enough group of people and ask them to "make decisions affecting matters of general interest", that group's decisions will, over time, be "intellectually superior to the isolated individual" no matter how smart or well-informed he is.

- The best collective decisions are the product of disagreement and contest, not consensus or compromise.

- An intelligent group does not ask its members to modify their positions. Instead, it figures out how to use mechanisms produce collective judgments that represent not what any one person in the group thinks but rather, in some sense, what they all think.

Potential Ideas for Project

- Algorithms to detect areas of consensus and areas of disbutes for each controversial article
- Algorithms to classify edits into helpful/unhelpful for collective judgments or consensus building
- Does this updating process lead to some consensus? What would happen if remove unhelpful edits
  - ‣ Can be transferable to issues like fact-checking

Coordination and Quality (Kittur & Kraut)

- Explicit coordination: Communication on the article talk page
- Implicit coordination: Editing in an article is concentrated among a small subset of editors
- Quality: Community ratings by other Wikipedians, Amazon MTurk to check validity
- Find: Concentrating editing in fewer editors may enable a small core of committed contributors to focus on complex, highly interdependent tasks while allowing infrequent contributors to add value on simple, stand-alone tasks
  - Representative democracy vs. Direct democracy
  - Traditional media vs. Social media

Bias in Crowds vs. Experts (Greenstein & Zhu)

- Compare pairs of articles between Encyclopædia Britannica and Wikipedia
- Measure the bias of the articles (train a model of word usage by politicians for which ideology is known, fit the model to articles)
- Find: Crowd-based knowledge production does not result in articles with more biased than articles produced by experts when the crowd-based articles are substantially revised
  - ‣ Idea: Automatically detect and correct for algorithmic biases
  - ‣ Sendhil Mullainathan on New York Times: "Biased Algorithms Are Easier to Fix Than Biased People"