Using Social Media To Understand Causal Relationships

...

Alankrit Johari

Causal Knowledge

Very important and hard to find.

Why important?

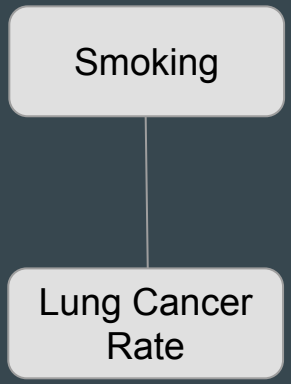
- To analyze claims like “Playing Video games result in violent behavior in children”

Why hard?

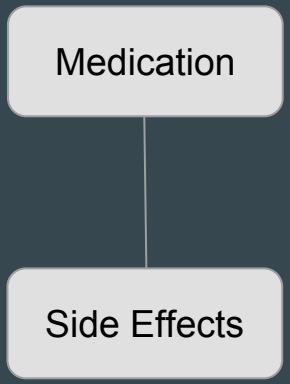
- Need domain expertise

Why do we need Causal Knowledge?

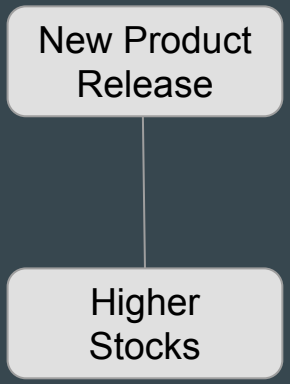
PREDICTION



EXPLANATION



INTERVENTION



What I want to do?

Main focus: medical claims

Examples:

- Meditation is good for concentration.
- Watching too much television harms your eyes.
- Playing violent video games causes violent behavior in children.

Basically, claims that are myths and sounds true.

Right Variables

- What are we finding causes between?
- It can be in format: (Cause - Problem), (Problem - Cause)
- Not necessarily related to problems: (Solution - Effect)
- Example: “Mediation is good for Concentration”
    - Meditation - Solution
    - Concentration - Effect

Text Mining

There is so much information floating around on the web.

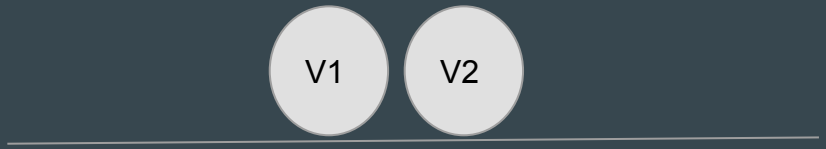
For these medical claims, people write:

- Blogs,
- Answers (eg. Yahoo Answers, Quora),
- etc.

Use text mining techniques to analyze this text and understand correlations and verify causes.

Co-occurrence and Network Analysis Method

- I will plot the strength of co-occurrence on a graph.
- The closer the two variables the stronger is the correlation.
- The further the two variables the weaker is the correlation.
- The correlation will be based on count of co-occurrence over the web.
- Some weightage to the higher reliable sources?



Data

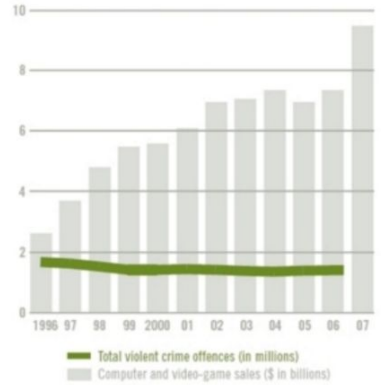
- Available public API's
  - Yahoo Answers!
  - Blogger
- Information extraction easy from these public API's
- Data extraction from Quora requires use of scraping tools
  - Selenium (for scraping)
  - BeautifulSoap (for scraping)
  - Mechanize (for navigating)

Verification

- Will compare the resulting graphs with various experimental studies.
- Correlation to Causal mapping
- Check results for highly unlikely pairs, eg. (good sleep - bad result on exams)

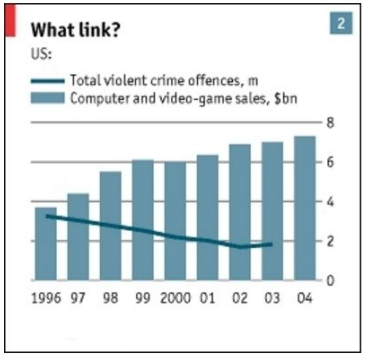
III. Youth Violence Trends

13. Violent Crime Offenses and Video Game Sales, 1996-2007



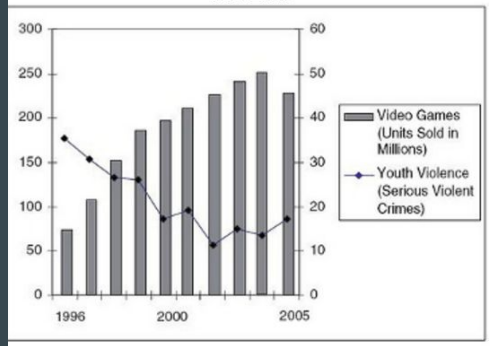
Source: Entertainment Software Association, "Essential Facts About Games and Violence," www.theesa.com, 2008

14. Violent Crime Offenses and Video Game Sales, 1996-2004



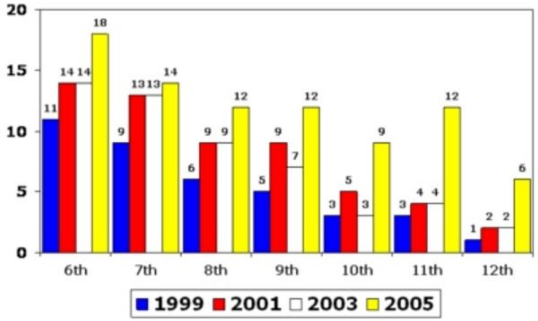
Source: "Chasing the Dream," *Economist*, Aug. 4, 2005

15. Sales of Video Games Compared to Youth Violence, 1996-2005



Source: Adam Thierer, MA, "Video Games and 'Moral Panic,'" www.pff.org, Jan. 23, 2009

16. Reports of Bullying, 1999-2005 (Rate Per 1,000 Students)



Source: Virginia Youth Violence Project, "Violence in Schools," www.youthviolence.edschool.virginia.edu (accessed Feb. 12. 2010)

Challenges

- Representation of data?
- What about negative correlations.
- How closes to causes? Just correlations?
- Relevant data?

References

- Netzer, O., Feldman, R., Goldenberg, J. and Moshe Fresko. 2012. Mine Your Own Business: Market-Structure Surveillance Through Text Mining. Marketing Science 31(3):521-543.

- Kleinberg, S. and George Hripcsak. 2011. A review of causal inference for biomedical informatics. Journal of Biomedical Informatics.

- http://videogames.procon.org/view.resource.php?resourceID=003627