11-830 Computational Ethics for NLP

Lecture 2: Ethical Challenges in NLP

History and Philosophy

Carnegie Mellon University
Language Technologies Institute

Defining Ethics

It's the **good** things
It's the **right** things

So what are the right things?

Is there some absolute definition of right?

Aristotle’s “Nicomachean Ethics”

- Hard to study subjects such as Ethics and Politics
- Involving what is thought to be **beautiful** and **just**
- Start with what people with good up-bringing and experience in life agree on
- Try to codify rules that are ethical
  - Define ethics to be equal to law or truth (or other way round)

Ethics in Philosophy

- Continuous subject of study
  - Is studied in the abstract, but
  - Often writer gets carried away with **their own** ethics

Ethics in Law

- Also continuous subject of study
  - Laws start off to be codified ethics for society
  - But language is never precise
  - Language changes over time: (it says “man” but meant “person”)
  - Adversarial Lawyer looks for loopholes
    - Both sides try to change the interpretation of the law to their advantage

Ethics in Religion

- Successful religions usually promote their own society
  - Religious laws reflect survival of community (mostly)
  - '*Thou shalt not kill*' seems clear
    - Does it refer also to non-believers?
  - What about copying copyrighted material you already own?
    - Most religions don't comment on this
    - Not all laws can envisage future issues.

Can We Define Ethics?

- Let’s look at morality and legality
  - **Illegal+immoral**:
  - **legal+immoral**:
  - **illegal+moral**:
  - **legal+moral**:

Can We Define Ethics?

- Let’s look at morality and legality
  - **Illegal+immoral**: murder
  - **legal+immoral**: cheating on a spouse
  - **illegal+moral**: civil disobedience
  - **legal+moral**: eating ice cream

Can We Define Ethics?

- Let's look at morality and legality
  - **Illegal+immoral**: murder
    capital punishment
  - **legal+immoral**: cheating on a spouse
    cancelling Game of Thrones
  - **illegal+moral**: civil disobedience
    assassination of a dictator
  - **legal+moral**: eating an ice cream
    eating the last ice cream in the freezer

Can We Define Ethics?

Can We Define Ethics?

- Probably not

Can We Define Ethics?

- Probably not (well not within one semester)

Can We Define Ethics?

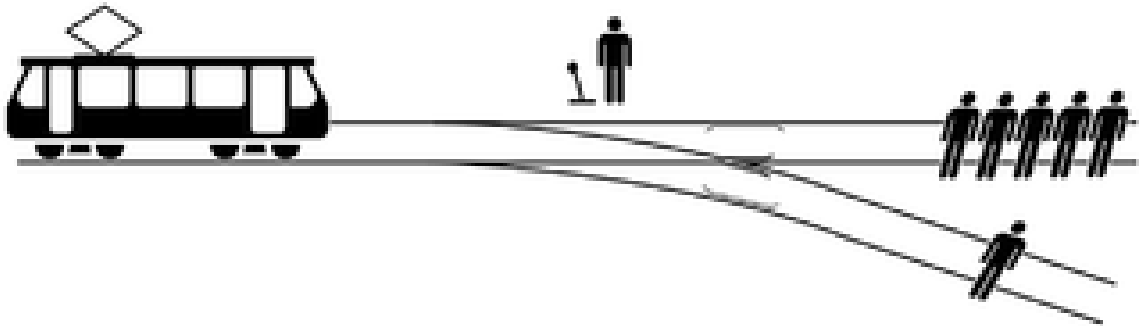
- Probably not (well not within one semester)
- So is it hopeless?

Can We Define Ethics?

- Probably not (well not within one semester)
- So is it hopeless?
- No: it is another problem with an ill-defined answer
  - It still has some definition of good and bad
  - Not everyone agrees on all examples
  - They do agree on **some** examples
  - They do have some correlation between people
- Is this different from other Language Technology Problems
  - Summarization, QA, Dialog, Speech Synthesis ...

The Trolley Problem

Should you pull the lever to divert the trolley?



[from Wikipedia]

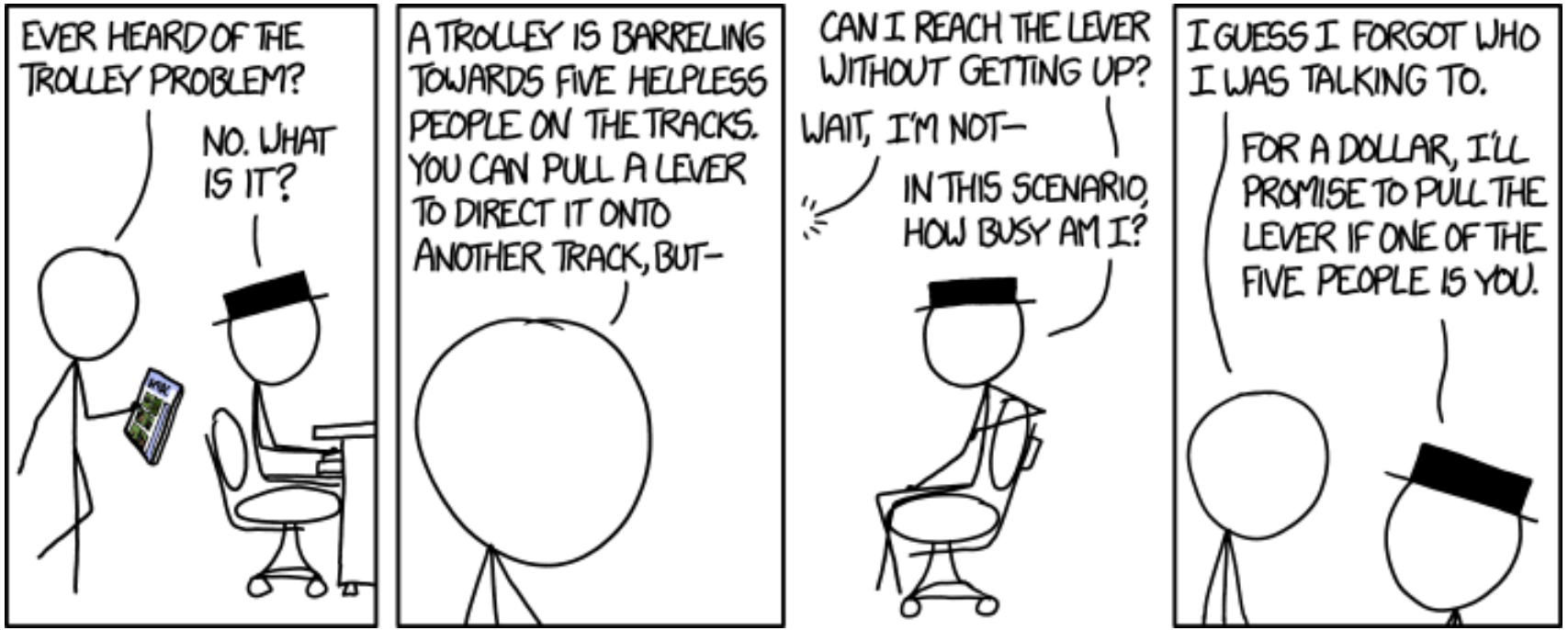
Trolley Problem

- One issue:
  - Actively participating if you pull the lever
  - Actively participating if you could pull the lever
- Does it make a difference with the number of people
  - Or the age of the people (or how well you know them)
- Is it different if you push “Homer Simpson” in front of the train
  - Much more explicitly killing “Homer Simpson”

Trolley Problem

- Not every one agrees on the same solution

Trolley Problem



xkcd

Prisoner’s Dilemma

- Two criminals are caught and sent to prison
- If one confesses, he goes free and the other gets 3 years
- If they both confess, they both get 2 years
- If they both stay silent, then they both only get 1 year each

Prisoner’s Dilemma

- Best action is to *both* stay silent
- So rational choice is to stay silent
- But if that’s the rational choice you should confess

Problem requires trust, which might not exist

Iterative Prisoner's Dilemma

- Same rules but you collect points
- You play the game multiple times
  - If you "trade" you give a point to the other
  - If you "defect" you loose nothing
  - If the other "trades" you get a point
  - If the other "defects" you don't get a point.
- What is the best strategy
- You can build trust, you build history

Iterative Merchant's Dilemma

- Is it different if the players are merchants vs prisoners
- Defect every turn is safest
  - Never any loss
  - But not wealth creation
- Trade every turn is risky
  - Other might work you only trade, so they defect
- Trade and do what the other did the last time
  - Might work if the other has the same strategy
- Application of game theory
  - Note that the results are very different
  - Optimized by turn vs over the game

Defining Ethics

- No absolute answer
- (and probably never can be one)
- Be aware of what you think is ethical might not be for others'
- But don't give up
- At least ensure ethical choices are deliberate

Watch This Talk

‘‘Hey, new question,’’ Barbie said. ‘‘Do you have any sisters?’’
‘‘Yeah,’’ Tiara said. ‘‘I only have one.’’
‘‘What’s something nice that your sister does for you?’’ Barbie asked.
‘‘She does nothing nice to me,’’ Tiara said tensely.
Barbie forged ahead. ‘‘Well, what is the last nice thing your sister did?’’
‘‘She helped me with my project — and then she *destroyed* it.’’
‘‘Oh, yeah, tell me more!’’ Barbie said, oblivious to Tiara’s unhappiness.
‘‘That’s it, Barbie,’’ Tiara said.
‘‘Have you told your sister lately how cool she is?’’
‘‘No. She is *not* cool,’’ Tiara said, gritting her teeth.
‘‘You never know, she might appreciate hearing it,’’ Barbie said.

Intelligent Systems: Design & Ethical Challenges

Barbara Grosz NYT article: Barbie Wants to Get to Know Your Child

Your Homework Before The Next Lecture

NIPS Keynote: Kate Crawford, The Trouble with Bias

https://goo.gl/qqeMKQ