Answering Ad Hoc Causal Questions in Web Search

Emre Kıcıman
emrek@microsoft.com
Microsoft Research
should i buy a fixer upper or new home?
should i rent or buy a home in 2014
should i get preapproved before house hunting
should i refinance my mortgage
should i get a divorce
should i break up with my boyfriend
should i pay off my mortgage
should i file bankruptcy
should i join a gym
should i eat before or after working out
should i buy a chromebook
should i text him
should i buy a house
should i cut my hair short
should i retire at 65
should i go to college
should i go to law school
should i take a multivitamin
should i text him or wait for him to text me
should i join the military
should i leave my husband
should i get married
should i pop a blister
should i get bangs
should i wash my hair before coloring it
should i buy a house
should i pop a burn blister
should i see a doctor
should i quit drinking coffee
should i shave my head
These queries are asking ...
“what happens after _____”
Expectation Exploration Searches

Everyday, people find themselves in unfamiliar situations and consider the potential outcomes of actions.

- Individuals ask themselves (and Bing!) “What’ll happen if I do that?”
- Policy-makers ask “What happens when someone does that?”

Goal is to inform.
- Understand consequences of a considered action
- Seeking validation / support for ongoing consequences
- Explore hypotheticals
Databases answer other questions

But, no database or knowledge base answers expectation exploration questions

• Individuals rely on articles, coaches, friends’ advice, and gut instinct.

• Policy-makers rely on experiments, data analyses, and gut instinct.
Capturing outcomes from longitudinal social media data

The information we need is already being recorded

• Millions of people frequently and publicly report the actions they take and, over time, the outcomes they experience
• ... and they have been doing this for years

We can *compute* answers to people’s questions about an experience

• Use causal inference methods to compare timelines of people who reported the experience to timelines of people who did not.
• Applies to varied data: Twitter, LinkedIn, search logs, financial, ...
Building block for IR experiences

Do people buy new cars after a raise?

(a) Timeline answer

<table>
<thead>
<tr>
<th>Luxury cars</th>
<th>Small cars</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Love the dog</td>
<td>Early walk</td>
</tr>
<tr>
<td>Enjoy walks</td>
<td>Scratch</td>
</tr>
</tbody>
</table>

(b) Pros / cons list

Hey, I sprained my ankle badly

When will I play football again?

People start to mention playing football after 8 weeks

(c) Conversational agent

[KDD’15, DESIRES’18]
A new analysis task for IR

To answer “what if”, “should I”, and other expectation exploration tasks

Corpus: Rich, Individual-Level Longitudinal Data
   Today: Social media
   Tomorrow: Many data sources

Query: What happens after experience $T$?
   Need enough information to recognize “people who did $T$”

Output: Expectation Map captures changes in outcomes over time
   Building block for many user experiences
What happens after ...
• $T$ is a binary treatment status, identified by a search query
• $X$ is all social media messages that occurred before $T$, represented as bag of n-grams
• $Y$ is all social media messages that occur after $T$. E.g., separate effect per word or phrase
Rest of this talk

• Causal inference over social media data

• Experiences applying to social sciences problems

• Looking forward
- Treatment identification: ~query-matching (IR / NLP)
- Inference of treatment effects: causal inference
- Interpretability of effects: HCI / IR
1. Build individual timelines of user activities
2. Treatment identification
3. Extract covariates
4. Identify similar groups of users
5. Calculate population average outcomes

For every observed outcome \( y \in \mathcal{Y} \), compute \( P(y=1|T=1) \) per strata.
Output: Expectation Map

Example treatment “ankle sprained”

<table>
<thead>
<tr>
<th>Elapsed Time</th>
<th>Outcome Y</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>crutches</td>
<td>50x</td>
</tr>
<tr>
<td>Day 1</td>
<td>play soccer</td>
<td>0.10x</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 42</td>
<td>crutches</td>
<td>1x</td>
</tr>
</tbody>
</table>
What words do people mention more once they say they have gout?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Count</th>
<th>Absolute Increase</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flare_up</td>
<td>35</td>
<td>4.1%</td>
<td>12.33</td>
</tr>
<tr>
<td>Uric_acid</td>
<td>27</td>
<td>2.9%</td>
<td>10.36</td>
</tr>
<tr>
<td>Uric</td>
<td>28</td>
<td>2.9%</td>
<td>10.11</td>
</tr>
<tr>
<td>Flare</td>
<td>81</td>
<td>4.9%</td>
<td>9.92</td>
</tr>
<tr>
<td>Big_toe</td>
<td>38</td>
<td>2.9%</td>
<td>9.86</td>
</tr>
<tr>
<td>Joint</td>
<td>301</td>
<td>7.2%</td>
<td>7.22</td>
</tr>
<tr>
<td>Aged</td>
<td>32</td>
<td>1.7%</td>
<td>6.51</td>
</tr>
<tr>
<td>Correlation</td>
<td>45</td>
<td>2.8%</td>
<td>6.11</td>
</tr>
<tr>
<td>Bollock</td>
<td>53</td>
<td>2.5%</td>
<td>5.96</td>
</tr>
<tr>
<td>Shite</td>
<td>108</td>
<td>3.4%</td>
<td>5.93</td>
</tr>
</tbody>
</table>
What words do people mention more once they say they have high triglyceride levels?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Count</th>
<th>Absolute Increase</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your_risk</td>
<td>46</td>
<td>24.8%</td>
<td>18.12</td>
</tr>
<tr>
<td>Statin</td>
<td>48</td>
<td>23.1%</td>
<td>17.69</td>
</tr>
<tr>
<td>Lower</td>
<td>120</td>
<td>35.9%</td>
<td>17.18</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>54</td>
<td>23.0%</td>
<td>16.72</td>
</tr>
<tr>
<td>Healthy_diet</td>
<td>55</td>
<td>19.3%</td>
<td>16.54</td>
</tr>
<tr>
<td>Fatty_acid</td>
<td>29</td>
<td>18.3%</td>
<td>16.37</td>
</tr>
<tr>
<td>Help_prevent</td>
<td>73</td>
<td>26.9%</td>
<td>16.01</td>
</tr>
<tr>
<td>Risk_factor</td>
<td>33</td>
<td>18.3%</td>
<td>15.55</td>
</tr>
<tr>
<td>Fish_oil</td>
<td>48</td>
<td>24.4%</td>
<td>15.42</td>
</tr>
<tr>
<td>inflammation</td>
<td>78</td>
<td>25.1%</td>
<td>15.30</td>
</tr>
</tbody>
</table>
Rest of this talk

• Causal inference over social media data

• *Experiences applying to social sciences problems*

• Looking forward
Applications in computational social science

• Transitions from mental health discussions to suicidal ideation
  • Explores heterogeneous effects [CHI16], and human assessments of balance [ICWSM17]
  • w/De Choudhury (GATech) et al.

• Daily Behaviors and Sleep and Exercise Quality
  • Combines query logs, location traces and MS Band data [JDSA 18]
  • w/Farajtabar, Nathan, and White

• Conjunction of factors triggering waves of seasonal influenza
  • Explores geographic flow, juxtaposition of multiple large-scale datasets, county-level matching [eLife 2018]
  • w/Chattopadhyay (UChicago), Elliott (UChicago), Shaman (Columbia), Rzhetsky (U Chicago)
College

• Success in college predicts individual career success, career happiness and economic achievement.
• Drives macro-economic growth too.
• But 1 in 3 college students leaves without earning a degree
• Many factors jeopardize college success: family responsibilities, financial pressures, individual behaviors
Alcohol in College

• College students drink more than non-college peers

• Persistent public health issue: hangovers, lowered academic performance, DUI arrests, risky sexual behavior, sexual and other assaults.

• Excessive alcohol consumption is negatively associated with college success

• Studies of college behavior through self-report surveys and interviews; in-person or deliberate recruiting.
What happens to students who mention alcohol more?

**Corpus:** 5 years of tweets by students starting college in Fall 2010
- 63k users, Aug 2010 – May 2015
- ~650M total tweets
- ~12 tweets/day per active person

**Treatment:** Mention alcohol keywords often during first semester

**Outcome:** Focus on topics known to be relevant to college success
Identify alcohol keywords in Fall 2010 (9/15-12/15/2010)
Compare top 1/3 of alcohol mentioners (by % active days w/alcohol keywords)
... to bottom 1/3 of alcohol mentioners (no alcohol mentions)
### Outcome Topics

Table 2: Topics linked to college success

<table>
<thead>
<tr>
<th>Concept</th>
<th>Seed words</th>
<th>Final topic words</th>
<th>Example Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer group interaction</td>
<td>friend, boyfriend,</td>
<td>boyfriend, buddy, roommate, bandmate,</td>
<td>Yup buddy I found my bandmate!!</td>
</tr>
<tr>
<td></td>
<td>girlfriend</td>
<td>fiancé, +20 more</td>
<td></td>
</tr>
<tr>
<td>Family responsibilities</td>
<td>mother, father,</td>
<td>mother, father, brother, step-</td>
<td>Thankful for my little bro and mom</td>
</tr>
<tr>
<td></td>
<td>brother, sister</td>
<td>dad, grandmother, +21 more</td>
<td>I have a sister #fact</td>
</tr>
<tr>
<td>Study habits</td>
<td>study, library,</td>
<td>study, library, math, tutor, text-</td>
<td>anyone that wants to study for</td>
</tr>
<tr>
<td></td>
<td>homework</td>
<td>books, worksheets, +49 more</td>
<td>history we’re in the library</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>but anyways ima off to study</td>
</tr>
<tr>
<td>Financial pressures</td>
<td>debt, student loans,</td>
<td>wages, afford, utilities, tuition,</td>
<td>finally my wages wooo</td>
</tr>
<tr>
<td></td>
<td>loans</td>
<td>evicted, fees, +28 more</td>
<td>@anon its all about money. Im in</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>debt. dont want more loans</td>
</tr>
<tr>
<td>Legal/criminal challenges</td>
<td>police, cops, jail,</td>
<td>cops, police, restraining, probation,</td>
<td>meeting my parole officer</td>
</tr>
<tr>
<td></td>
<td>parole</td>
<td>rehab, +15 more</td>
<td>cops pulling out breathalyzer f***k</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>we drunk</td>
</tr>
</tbody>
</table>

In addition, 195+ validated empath topics
People in the Alcohol group were much less likely to mention studying over the next several years, and somewhat less likely over the entire time period.
Criminal and Financial effects

People in the Alcohol group were more likely to mention legal and criminal challenges through most of our study; and slightly more likely to mention financial pressures over time.
Methodological Limitations (1)

• Population biases require measurement
• Analysis misses language meaning beyond tokens
  • Requires qualitative reading
• Studying changes in conversational behavior only
  • E.g., reporting biases
  • E.g., inverted causal relationships
Methodological Limitations (2)

• Assumes ignorability of unobserved confounds
• Assumes SUTVA (no network effects)
• Heterogeneous effects
• Survivor bias.
• When treatment is complex (e.g., when not a point treatment) can be confounded by other experiences
Rest of this talk

- Causal inference over longitudinal data
- Experiences applying to social sciences problems
- Looking forward
Does this generalize?

• Calculate consequences of 39 experiences from business, health and society questions
  • E.g., people taking Prozac, getting a divorce, buying life insurance, investing money
  • Experiences selected from Bing search queries

• MTurk workers evaluate surface validity based on supporting social media messages, and online reference material

More broadly, key evaluation criteria must go beyond Correctness to include Interpretability and Usefulness

[CSCW’17]
Opportunities: Core methods

• Improved automation of longitudinal analyses
  • Feature representation for longitudinal
  • Mitigation of known social data biases
  • Under scaling and performance constraints

• Personalization and heterogeneity: Identifying effect modifiers

• Integration of domain knowledge as available
Opportunities: Treatment identification

• Improving query interface

• Expanding classes of treatments beyond point-treatments
  • E.g., Multi-dose treatments
Opportunities: End-to-end task integration

• Interaction and iteration

• Interpretability of results and supporting evidence

• Evaluation methods and criteria for end-to-end task fulfillment
Summary

Everyone asks “what happens after …”
  • For decision support, for expectation setting, for exploration, ...

Causal inference over longitudinal, individual-level data can give answers
  • A building block for new IR answers and experiences

Developed/validated methods in computational social science tasks
  • Mental health; sleep and exercise quality; alcohol and college trajectories; ...

Now: Scale, interaction, interpretability to make broadly accessible
  • Many exciting research opportunities
Selected papers mentioned in the talk (all at http://kiciman.org/)

- Kıcıman, Richardson, **Towards Decision Support and Goal Achievement: Identifying Action-Outcome Relationships from Social Media**. KDD’15

- De Choudhury, Kıcıman, Dredze, Coppersmith, Kumar, **Shifts to Suicidal Ideation from Mental Health Content in Social Media**. CHI’16 *honorable mention*

- Olteanu, Varol, Kıcıman, **Distilling the Outcomes of Personal Experiences: A Propensity-scored Analysis of Social Media**. CSCW’17

- Kıcıman, Counts, Gasser, **Using Longitudinal Social Media Analysis to Understand the Effects of Early College Alcohol Use**. ICWSM’18

- Kıcıman, Thelin, **Answering What If, Should I and other Expectation Exploration Queries Using Causal Inference over Longitudinal Data**. DESIRES’18