Using Simpson’s Paradox to Discover Interesting Patterns in Behavioral Data

Kristina Lerman
USC Information Sciences Institute

Thanks to Nazanin Alipourfard, Peter G. Fennell,
IC2S2 2018

http://www.isi.edu/~lerman
Can you trust the trend?

- Do Stack Exchange users write better answers later in a session?  
- Do additional exposures by friends suppress a Twitter user’s use of a hashtag?  
- Do Reddit users become more active over time?

1. Alipourfard, Fennell, & Lerman (2018) “Can you trust the trend: Discovering simpson’s paradoxes in social data” in WSDM.
**SIMPSON’S PARADOX**

A TREND APPEARS IN DIFFERENT SUB-GROUPS OF DATA BUT DISAPPEARS OR REVERSES WHEN THESE SUB-GROUPS ARE COMBINED.*

* Simpson (1951). "The Interpretation of Interaction in Contingency Tables". *JRSS*
## SIMPSON’S PARADOX, AN ILLUSTRATION

Which treatment should doctor recommend for kidney stones*?

<table>
<thead>
<tr>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>78% (273/350)</td>
<td>83% (289/350)</td>
</tr>
</tbody>
</table>

* Wikipedia
SIMPSON’S PARADOX, AN ILLUSTRATION

Which treatment should doctor recommend for kidney stones*? After accounting for the confounder—stone size—the best choice reverses.

<table>
<thead>
<tr>
<th></th>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Stones</td>
<td>93% (81/87)</td>
<td>87% (234/270)</td>
</tr>
<tr>
<td>Large Stones</td>
<td>73% (192/263)</td>
<td>69% (55/80)</td>
</tr>
<tr>
<td>All</td>
<td>78% (273/350)</td>
<td>83% (289/350)</td>
</tr>
</tbody>
</table>

* Wikipedia
WHY SIMPSON’S REVERSAL OCCURS

- Small stones are more easily treated, especially by treatment A

- But are overwhelmingly assigned to treatment B

![Pie charts showing treatment success and failure for small and large stones for treatments A and B.](attachment:image.png)
Central ideas and papers

• Simpson’s paradox implies that functional differences exist within the population.
  
  E.g., small kidney stones are easier to treat

• “Computational Social Scientist Beware: Simpson's Paradox in Behavioral Data”, in *J. Computational Social Science* 2018

• We developed an algorithm to automatically identify Simpson’s paradoxes data.
  
  • “Using Simpson’s paradox to discover interesting behavioral patterns in data” in *ICWSM* 2018
  
  • “Can you Trust the Trend? Discovering Simpson’s Paradoxes in Social Data” in *WSDM* 2018
Method to discover Simpson’s paradoxes in data

For each covariate $X_p$

- Step 1: Estimate trend of outcome $Y$ with respect to a covariate $X_p$
Method to discover Simpson’s paradoxes in data

For each covariate $X_p$

• Step 1: Estimate trend of outcome $Y$ with respect to a covariate $X_p$
  For each remaining covariate $X_c$

• Step 2: Disaggregate data by conditioning on $X_c$
Method to discover Simpson’s paradoxes in data

For each covariate $X_p$

- Step 1: Estimate trend of outcome $Y$ with respect to a covariate $X_p$

For each remaining covariate $X_c$

- Step 2: Disaggregate data by conditioning on $X_c$
- Step 3: Compare trends of the outcome within disaggregated bins to the trend in the aggregated data
Empirical validations

Problem solving
• 2 years
Outcome Y:
• Is the problem solved correctly on the first attempt?
11 features:
• Day, month, first five, number of problems solved, ...

Question answering
• 6 years
Outcome Y:
• Will the answer be accepted as best answer by the asker?
19 features:
• words, code lines, session length, reputation, tenure, number of answers written, ...

Language learning
• 2 weeks
Outcome Y:
• Are all words correctly recalled in a lesson
22 features:
• Day, month, session length, first five, distinct words, hour24, ...
Stack Overflow – Simpson’s pairs

<table>
<thead>
<tr>
<th>Pseudo-$R^2$</th>
<th>Covariate of Performance</th>
<th>Conditioning variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.03</td>
<td>Answer Position w/in Session</td>
<td>Experience (answers written)</td>
</tr>
<tr>
<td>0.03</td>
<td>Session Length</td>
<td>Experience (answers written)</td>
</tr>
<tr>
<td>0.02</td>
<td>Experience</td>
<td>Reputation</td>
</tr>
<tr>
<td>0.02</td>
<td>Answer Position w/in Session</td>
<td>Reputation</td>
</tr>
<tr>
<td>0.02</td>
<td>Session Length</td>
<td>Reputation</td>
</tr>
<tr>
<td>&lt;0.01</td>
<td>Answer Position w/in Session</td>
<td>Session Length</td>
</tr>
<tr>
<td>&lt;0.01</td>
<td>Time since Previous Answer</td>
<td>Answer Position w/in Session</td>
</tr>
</tbody>
</table>

* Ferrara, et al. (2017) “Dynamics of content quality in collaborative knowledge production” in ICWSM.
Stack Overflow Simpson’s Paradox
# Khan Academy – Simpson’s pairs

<table>
<thead>
<tr>
<th>Pseudo-R²</th>
<th>Covariate of Performance</th>
<th>Conditioning variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04</td>
<td>Hour24</td>
<td>Problems Correct</td>
</tr>
<tr>
<td>0.03</td>
<td>Month</td>
<td>Five First Attempts</td>
</tr>
<tr>
<td>0.01</td>
<td>Month</td>
<td>Session Index</td>
</tr>
<tr>
<td>0.01</td>
<td>Month</td>
<td>Total Solve Time</td>
</tr>
<tr>
<td>0.02</td>
<td>Session Number</td>
<td>Number of Problems</td>
</tr>
<tr>
<td>0.01</td>
<td>Session Number</td>
<td>Tenure</td>
</tr>
<tr>
<td>0.01</td>
<td>Session Number</td>
<td>Total Solve Time</td>
</tr>
</tbody>
</table>
Khan Academy Simpson’s Paradox
Can you trust the trend?

Do Stack Exchange users write better answers later in a session?¹

Do additional exposures by friends suppress a Twitter user’s response?²

Do Reddit users become more active over time?⁴

Algorithm to find natural experiments in data

Performance over time in Khan Academy:
- a) performance declines in aggregate; but
- b) increases for ‘slow’ users. Subgroups automatically identified by our algorithm.

Slow users become more persistent:
they are more likely to continue working on problems they got wrong on their first attempt.

Hypothesis: User interface change in Khan Academy in April 2013 made slow users more “gritty”: more likely to continue working on a problem they got wrong. As a result, they performance on other problems increases.
To summarize

https://github.com/ninoch/Trend-Simpsons-Paradox/

• Simpson’s paradox occurs when an association/trend observed in the subgroups disappears or reverses when the subgroups are combined into one.

• Algorithm to automatically identify subgroups with different trends
  • A tool for data-driven discovery
  • And to formulate new hypotheses about data.

• Algorithm available!
  https://github.com/ninoch/Trend-Simpsons-Paradox/
  • Works for binary outcomes, linear models
THANK YOU!

Sponsors

NSF: CIF-1217605
ARO: W911NF-15-1-0142, W911NF-16-1-0306

Questions?
lerman@isi.edu