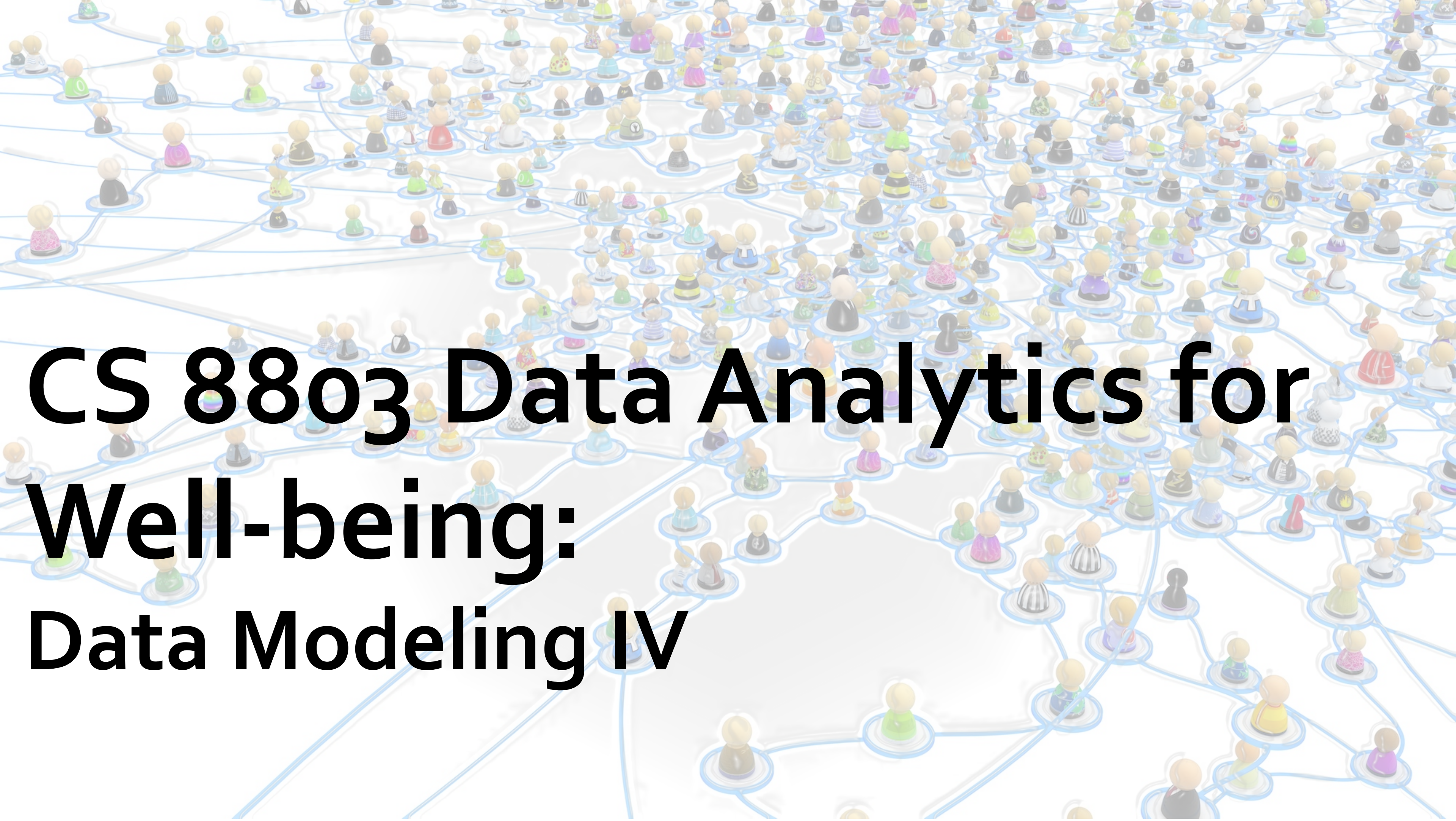


CS 8803 Data Analytics for Well-being:

Data Modeling IV

Munmun De Choudhury

munmund@gatech.edu

Week 5 | February 10, 2016

Class Presentation Signup

- 5% of overall grade.
- Pick any date/topic – list available on class website.
- Present in pairs – you can decide beforehand or see with whose schedule your date/topic matches with.
- Link: http://bit.ly/1nSO9DQ

Facebook Displays as Predictors of Binge Drinking: From the Virtual to the Visceral

Main idea

- *Goal*: to understand the role that one's own Facebook alcohol displays (e.g., a positive comment or picture about alcohol) play predicting binge drinking.
- *Contribution*: Towards this purpose authors applied constructs from the Theory of Reasoned Action to determine whether and where the construct of Facebook alcohol displays fits in predicting binge drinking as an outcome among a longitudinal sample of college students from two universities.

Motivation

- College students frequently post references to alcohol use on Facebook, including references to excessive alcohol use or binge drinking behaviors
- Given the popularity and potential influence of Facebook on college students’ binge drinking, efforts are needed to understand what role new communications such as social media play in existing behavioral models.

Theory of Reasoned Action

- The theory of Reasoned Action was developed by Martin Fishbein and Icek Ajzen (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975).
- Reasoned Action predicts that behavioral intent is created or caused by two factors: our attitudes and our subjective norms.
  - Attitudes have two components: the evaluation and strength of a belief.
  - Subjective norms, also have two components: normative beliefs (what I think others would want or expect me to do) and motivation

- One theory of behavior change that is applied to alcohol use is the Theory of Planned Behavior (TPB) (Ajzen, 1985; Ajzen, 1991) – derived from TRA – it indicates that the above construct predict future behavior
  - Previous work suggests empirical support for constructs within TPB; the model accounts for between 22% and 65% of the variation in binge-drinking behaviors (Armitage & Conner, 2001; Armitage & Conner, 2002).

Hypotheses

- H1: Greater perceived approval of alcohol by friends prior to freshman year (T1) is associated with a greater intention to drink at that time (T1).
- H2: Greater attitude towards alcohol in the summer prior to freshman year (T1) is associated with a greater intention to drink at that time (T1).
- H3: Greater intention to drink alcohol in the summer prior to freshman year (T1) is associated with increased episodes of binge drinking at the conclusion of freshman year (T2).
- H4: Greater attitude towards alcohol in the summer prior to freshman year (T1) is associated with increased episodes of binge drinking at the conclusion of freshman year (T2).
- H5: A greater number of Facebook alcohol displays in the period prior to starting freshman year of college is associated with a greater intention to use alcohol freshman year.

Method

- A codebook was used to evaluate displayed alcohol references.
  - displayed alcohol content referring to attitudes, intentions or behaviors regarding alcohol were considered displayed alcohol references.
  - Example references included personal photographs in which the profile owner was drinking from a beer bottle, or text references describing drinking vodka at a party

- A total of 7 coders evaluated profiles in this study, and all had undergone a minimum 3 month training period.
- Phone interviews were conducted with all participants at the time of enrollment as a baseline evaluation, and then again at the conclusion of the student's freshman year of college.

Findings

Correlation matrix.

Variables		Baseline EtOH Displays	Baseline Intention	Baseline EtOH Attitude	Baseline EtOH Approval	Binge Episodes Year	Drinking Freshman
Baseline Displays	EtOH	1					
Baseline Intention		.176*	1				
Baseline Attitude	EToH	.176*	.638*	1			
Baseline Approval	EtOH	.193*	.529*	.474*	1		
Binge Episodes	Drinking	.194*	.481*	.438*	.337*	1	

*$p < .01$

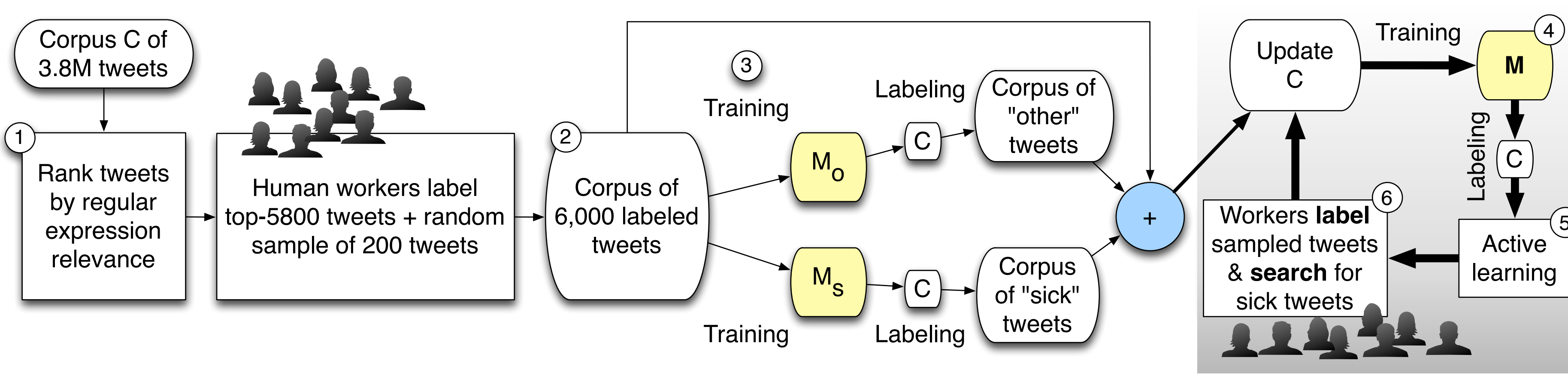
- Positive attitude towards alcohol predicted binge drinking through a direct path as well as through intention to drink.
- The overall fit of the model increases with the inclusion of Facebook alcohol displays as a direct predictor of alcohol behavior compared to a model without this path.
- Broadly, Facebook posts predict future binge drinking behavior directly.

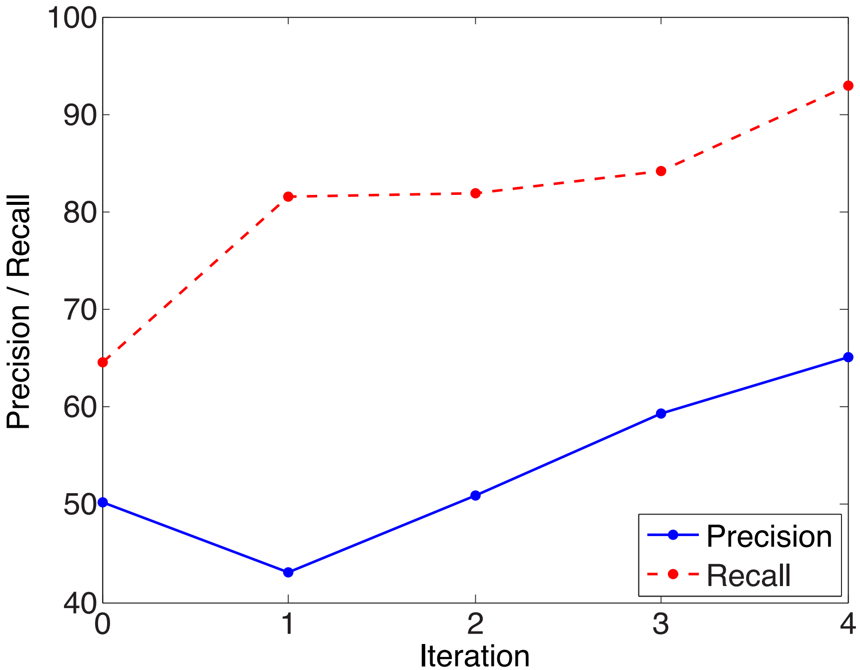
nEmesis: Which Restaurants Should You Avoid Today?

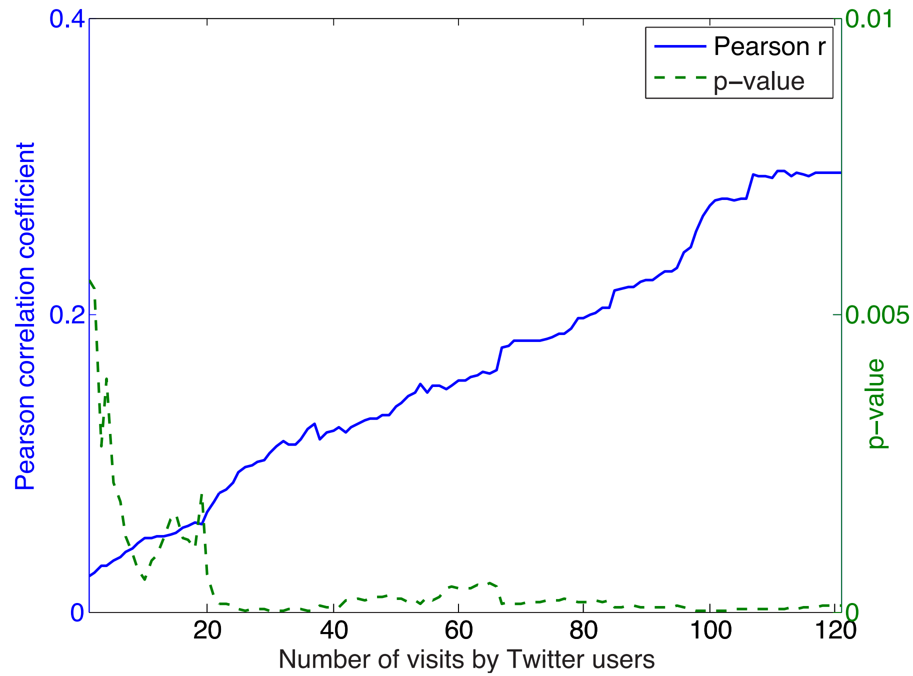
Summary

- The paper presents an end-to-end system, nEmesis, that automatically identifies restaurants posing public health risks.
- Data – Twitter. A language model is built to identify people complaining about food bourne illnesses.
- People's visits to restaurants are modeled by matching GPS data embedded in the messages with restaurant addresses (NYC).
  - A "health score" is assigned to each venue.
- Analysis reveals that the inferred health score correlates (r = 0.30) with the official inspection data from the Department of Health and Mental Hygiene (DOHMH).
- Adding attributes of online (Twitter) data with the DOHMH violation scores shows that over 23% of variance can be explained by the factors mined from Twitter

Restaurants in DOHMH inspection database	24,904
Restaurants with at least one Twitter visit	17,012
Restaurants with at least one *sick* Twitter visit	120
Number of tweets	3,843,486
Number of detected *sick* tweets	1,509
***Sick* tweets associated with a restaurant**	479
Number of unique users	94,937
Users who visited at least one restaurant	23,459







Feature	Regression Coefficient
Constant term c	$+16.1585$ ***
Number of visits	-0.0015 ***
Number of distinct visitors	-0.0014 ***
Number of *sick* visitors (f^T)	$+3.1591$ ***
Proportion of *sick* visitors (f)	$+19.3370$ ***
Number of sick days of visitors	0 ***

Aside from binge drinking, what other behaviors about college students could you study using social media?

D’Angelo et al. argue that alcohol displays in many ways could be a manifestation of their chosen identity online. Bringing in Goffman’s self-presentation related observations, to what extent these displays could simply be enacted (rather than real) and how could you account for it? (Also Auguste’s comment)

College students often live in a close knit community. Would peer effects influence alcohol use? How would you measure it?

D’Angelo et al. used a qualitatively coded list of topics to identify Facebook alcohol displays. What type of quantitative methods would be suitable for the purpose?

D'Angelo et al. used Facebook post information such as mentions of alcohol use and pictures of alcohol as their independent variables. It is not surprising that this is predicting future binge drinking (Anurag).

What other non-explicit cues may be predictive of this behavior?

How can you establish causation beyond such correlational effects?

Parisa noted why incoming Freshmen students were suitable for the study. In some ways alcohol displays need to be mediated with college adjustment. What variations would you see in a legal age population?

Sadilek et al. used Twitter data to model food bourne illnesses and obtain health score of restaurants. What types of other data will be suitable?

Visitation estimation on Sadilek et al. may be problematic (DOHMH + 4sq data). 25 meters maybe reasonable in Atlanta but in NYC with high density of restaurant in many areas, it may be a problem. People also may not report food bourne illness problems immediately. How can you counter these concerns?

In many ways, the Sadilek et al. paper is the classic paper where we need to have the correlation/causation discussion. Given the study is correlational, where can the inferences of the model go wrong?

Some of you indicated the utility of Yelp data, however is it likely that food bourne illnesses will be reported there? What would be the limitations of Yelp in this type of a study?

If you were to do either or both of the studies with anonymous geotagged social media like Yik Yak, what would be the benefits? What would be the limitations?