AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

chip

VE RI TAS

Toward Large-scale and Multi-facet Analysis of First Person Alcohol Drinking

Multi-Factorial Risk Analysis and Identification

S82

Hadi Amiri, Kara Magane, Lauren Wisk, Guergana Savova, Elissa Weitzman
Computational Health Informatics Program (CHIP)
Boston Children's Hospital, Harvard Medical School
{firstname.lastname}@childrens.harvard.edu

NIAAA R21AA023901, NIGMS R01GM114355

Disclosure

I and my spouse/partner have no relevant relationships with commercial interests to disclose.

Learning Objectives

1. **Characterize important indicators related to first-person alcohol drinking in social networks**
   - Drinking context
   - Consumption level
   - Their cross correlation

2. **Develop effective computational models to identify**
   - First-person reports of alcohol consumption
   - Drinking context
   - Consumption level

First person: alcohol consumption of one or more individuals

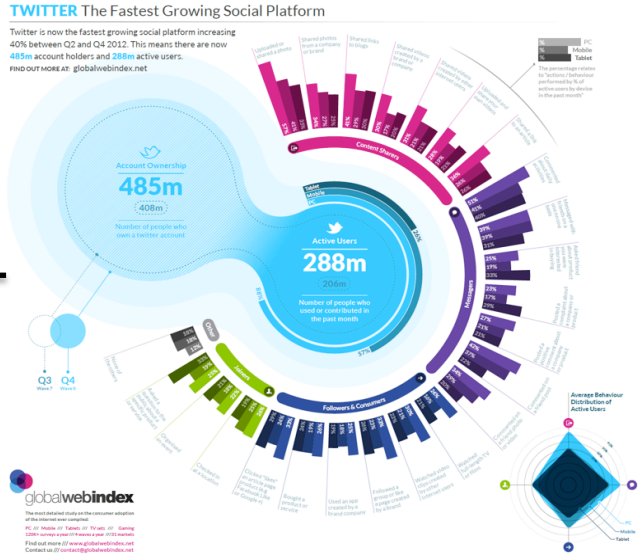
Motivation

Traditional Survey Data

	A	B	C	D	E	F	G	H	I	J	K	L
1	Resp ID	Gender	Location	Generation	Weight	Q0	Q1	Q2	Q3	Q4	Q5	Q
2	2	Male	South America	Generation X	1	0	0	0	1	0	1	
3	4	Female	South America	Baby Boomers	1.44	0	1	1	1	0	0	
4	5	Female	South America	Generation X	1	1	0	1	1	1	0	
5	6	Male	Antarctica	Baby Boomers	1.44	2	0	1	1	0	0	
6	9	Female	Europe	Baby Boomers	1.32	1	1	1	1	1	0	
7	12	Female	Europe	Baby Boomers	1.56	0	0	0	1	1	0	
8	15	Male	North America	Baby Boomers	1.56							
9	16	Male	Antarctica	Baby Boomers	1.44	1	1	1	0	0	0	
10	17	Female	Europe	Baby Boomers	1.32	2	1	1	1	0	0	
11	18	Male	North America	Traditionalists	0.595	0	0	1	0	0	0	
12	22	Male	South America	Generation X	1.32	0	0	0	1	1	0	
13	25	Female	South America	Generation X	1.32	1	1	1	1	1	1	
14	26	Female	South America	Millenials	0.765	1	0	1	0	0	0	
15	27	Male	Europe	Baby Boomers	1.56	1	1	1	1	0	0	
16	29	Male	Europe	Generation X	1							
17	30	Male	Europe	Baby Boomers	1.32	0	0	0	0	0	1	
18	31	Male	Europe	Millenials	0.68	1	0	0	0	0	0	
19	33	Male	North America	Generation X	1							
20	34	Male	North America	Generation X	1.32							
21	36	Female	North America	Millenials	1	1	0	1	1	1	0	
22	37	Female	North America	Millenials	0.765							

Complementary

Health-related Forum Posts



- **Challenge of traditional surveys**: low resolution, time-consuming, behavior change through time.
- **Deliver low-cost** & **high-resolution** views into **population behaviors** - first-person alcohol drinking.

Motivation

Alcohol use - a significant source of global morbidity and mortality

Rehm et al. Addiction 2017

Alcohol use behaviors in social media

- First-person reports of alcohol use
- Temporal patterns
- Behavioral nuances

Liu et al. CSCW 2017

West et al. Preventive Medicine 2012

Pang et al. IEEE 2015

Dearth of evidence

- Drinking context
- Consumption level

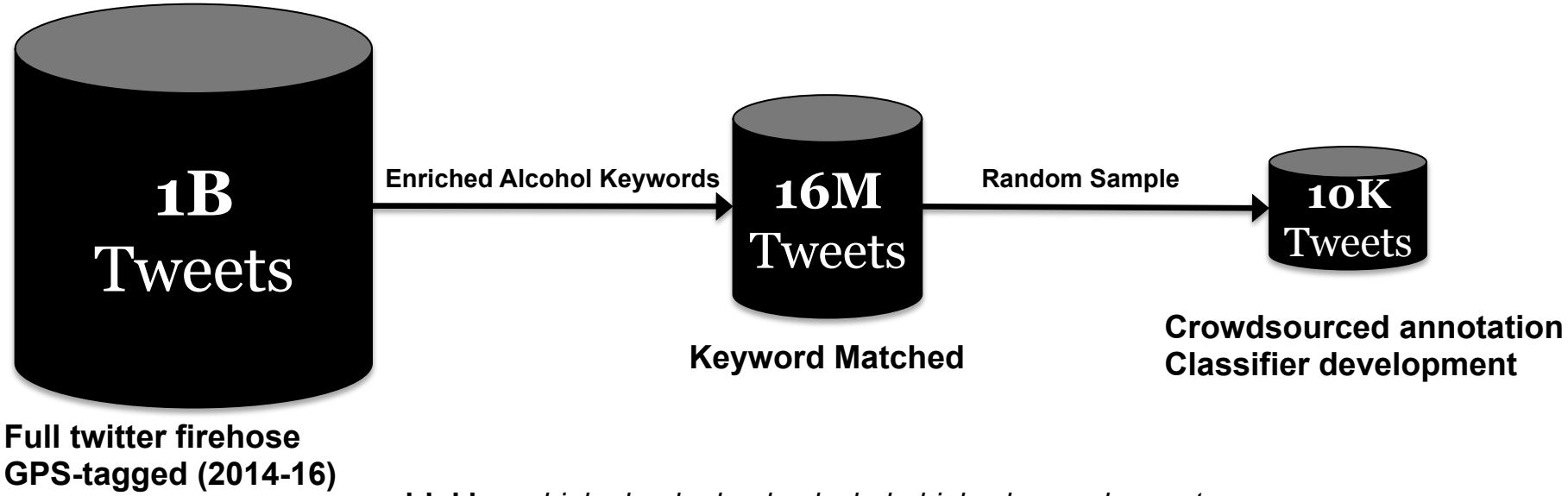
Outline

Dataset Development
- Crowdsourcing process
- Data analysis

Modeling
- Task and challenges
- Solution

Experiments
- Results and Analysis

Dataset Development – Crowdsourcing



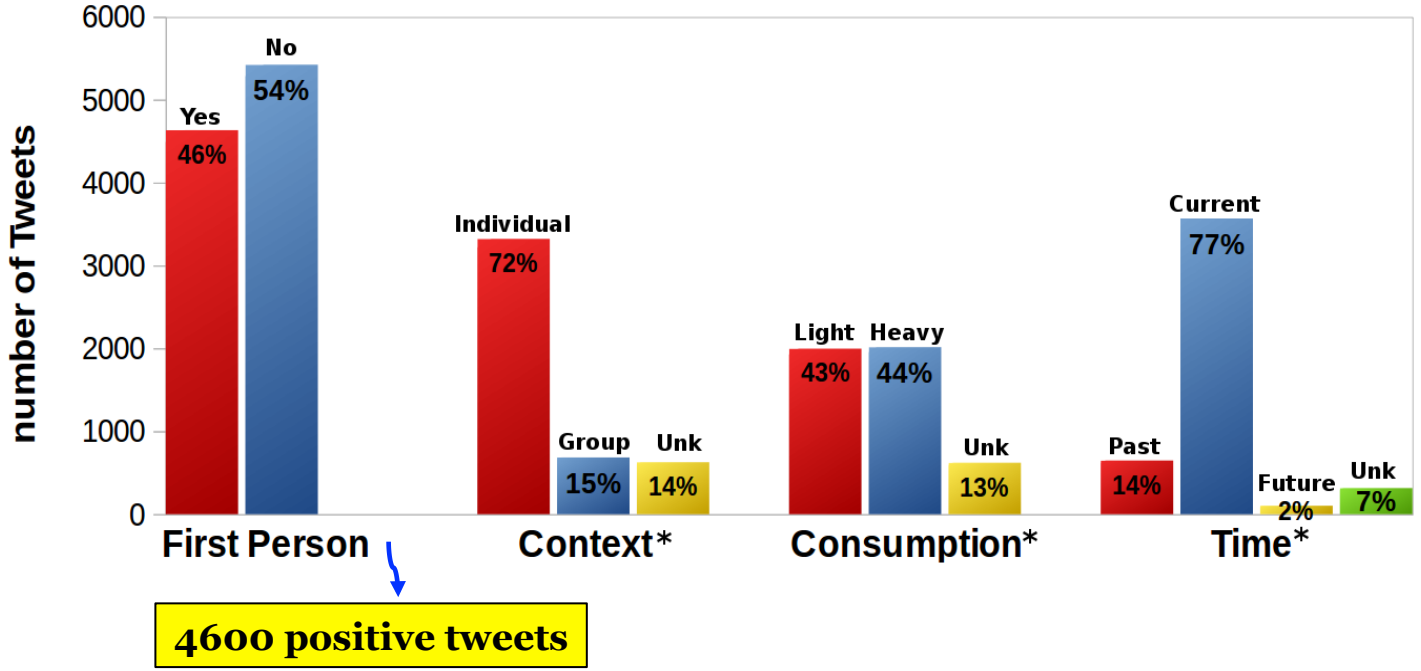
drinking: *drink, drank, drunk, alcohol, drinks, booze, beer, etc.*

First-person alcohol positive: *It's our fourth beer and we're **drinking** some more!*
First-person alcohol negative: *I think **drinking** water can cure my depression!*

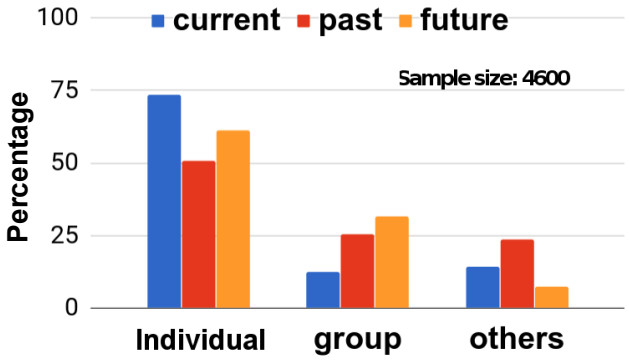
Dataset Quality Assurance – 10K

- **10K tweets** crowdsourced for annotation
  - **500 tweets** annotated and injected into 10K dataset
  - Annotators **maintain 80% accuracy** on these tweets
  - Each tweet labeled by **3 annotators**
  - **Fleiss' kappa: 76.1%**, substantial agreement

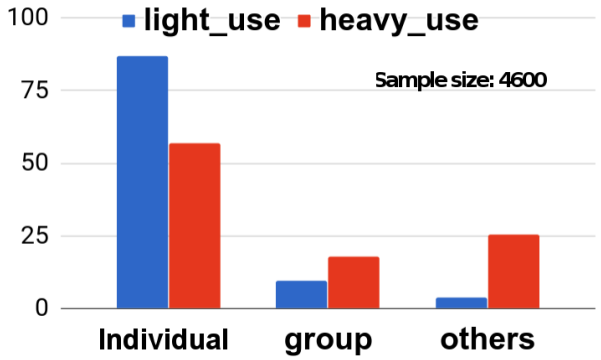
Dataset – 10K


number of Tweets
6000
5000
4000
3000
2000
1000
0
Yes
46%
No
54%
Individual
72%
Group
15%
Unk
14%
Light
43%
Heavy
44%
Unk
13%
Past
14%
Current
77%
Future
2%
Unk
7%
First Person
Context*
Consumption*
Time*
4600 positive tweets

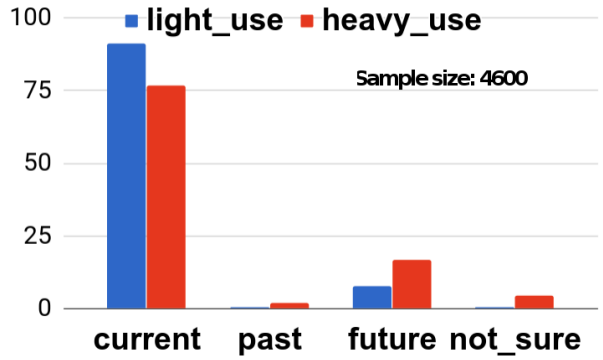
Dataset Analysis – 10K



- In solitary contexts, time of drinking is often current;
- In group context, time of drinking is often past or future.

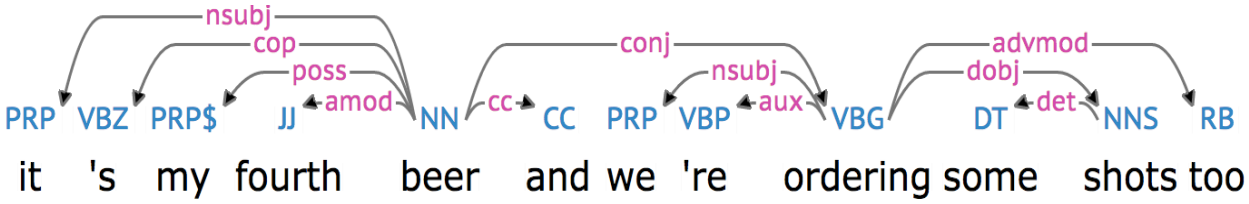


- In solitary contexts, consumption tends to be light;
- In group context, consumption tends to be heavy.



- Current drinking often indicates light consumption;
- Future and past drinking often indicate heavy consumption.

Modeling



Classification Tasks

1. First-person
2. Drinking context
3. Consumption level

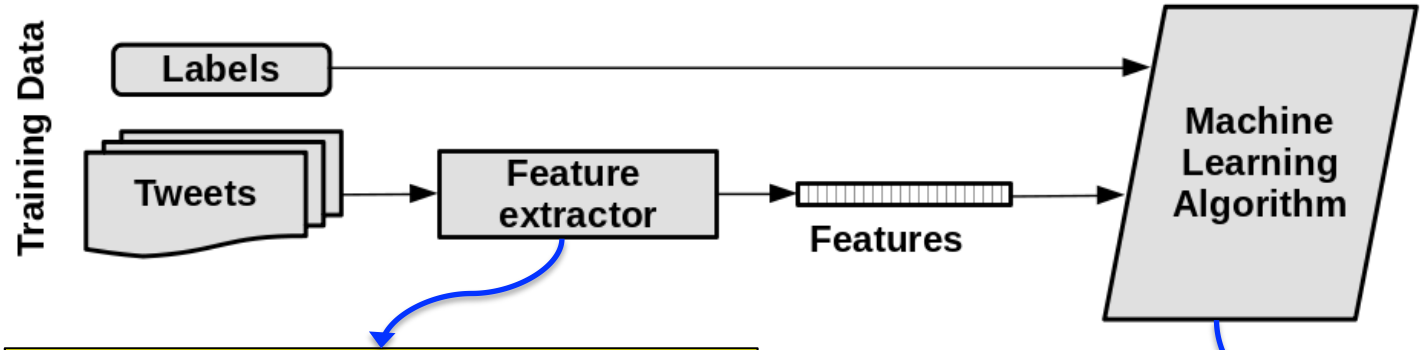
Challenge

- Complexity of human language

Solution

- Linguistically-aware machine learning

Modeling



Basic Features Liu et al. (2017)
- Character & word ngrams
- Topic model features

Enhanced Features
- Basic Features
- Syntactic and language bias features
- Sentiment & word embeddings features
- Quadratic version of these features

Classification Algorithm
- SVMs, linear kernel
- Grid search to optimize parameters

Results – 10K

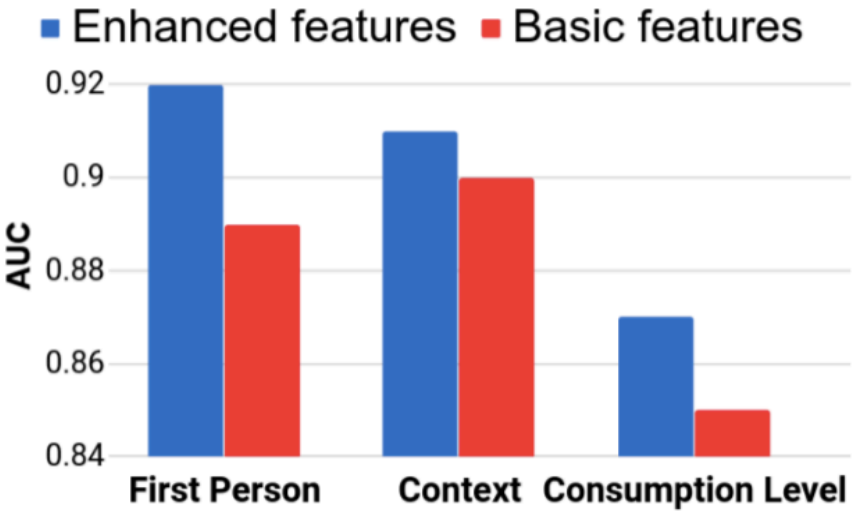


Figure 3: AUC classification performance.

Ablation analysis showed:

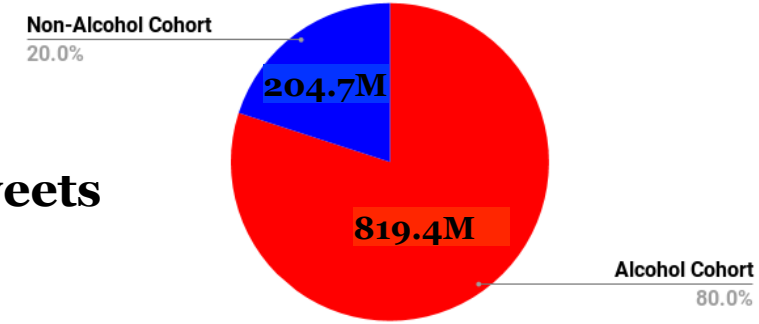
1. **Syntactic and**
2. **Word embedding**

features are key source of improvement.

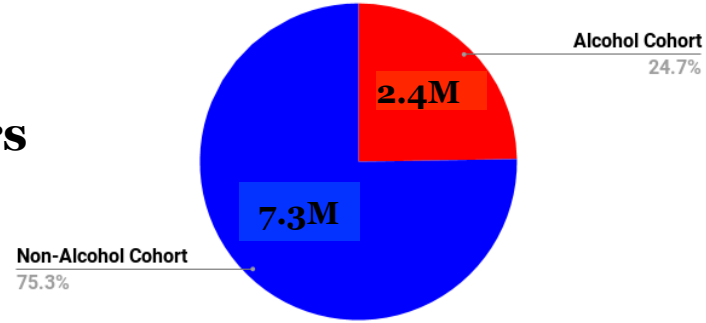
Analysis – 1B

Alcohol Cohort: users who sent at least one alcohol-relevant tweet in 1B set.

#Tweets



#Users

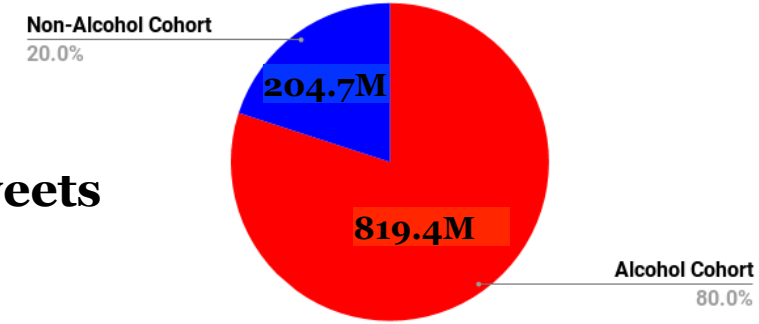


Users in **Alcohol cohort** are considerably more active (send more tweets) than those in **Non-Alcohol cohort**.

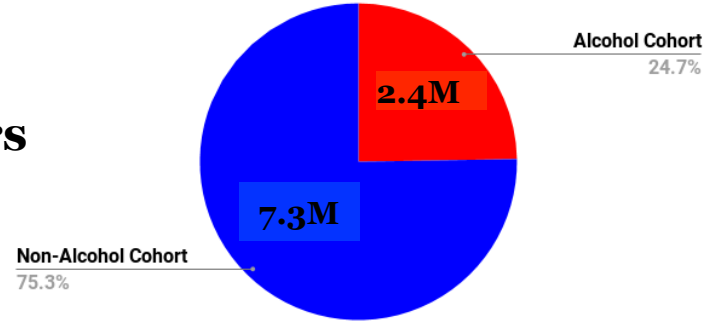
Analysis – 1B

Alcohol Cohort: users who sent at least one alcohol-relevant tweet in 1B set.

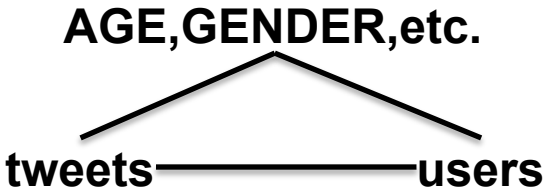
#Tweets



#Users



Users in **Alcohol cohort** are considerably more active (send more tweets) than those in **Non-Alcohol cohort**.

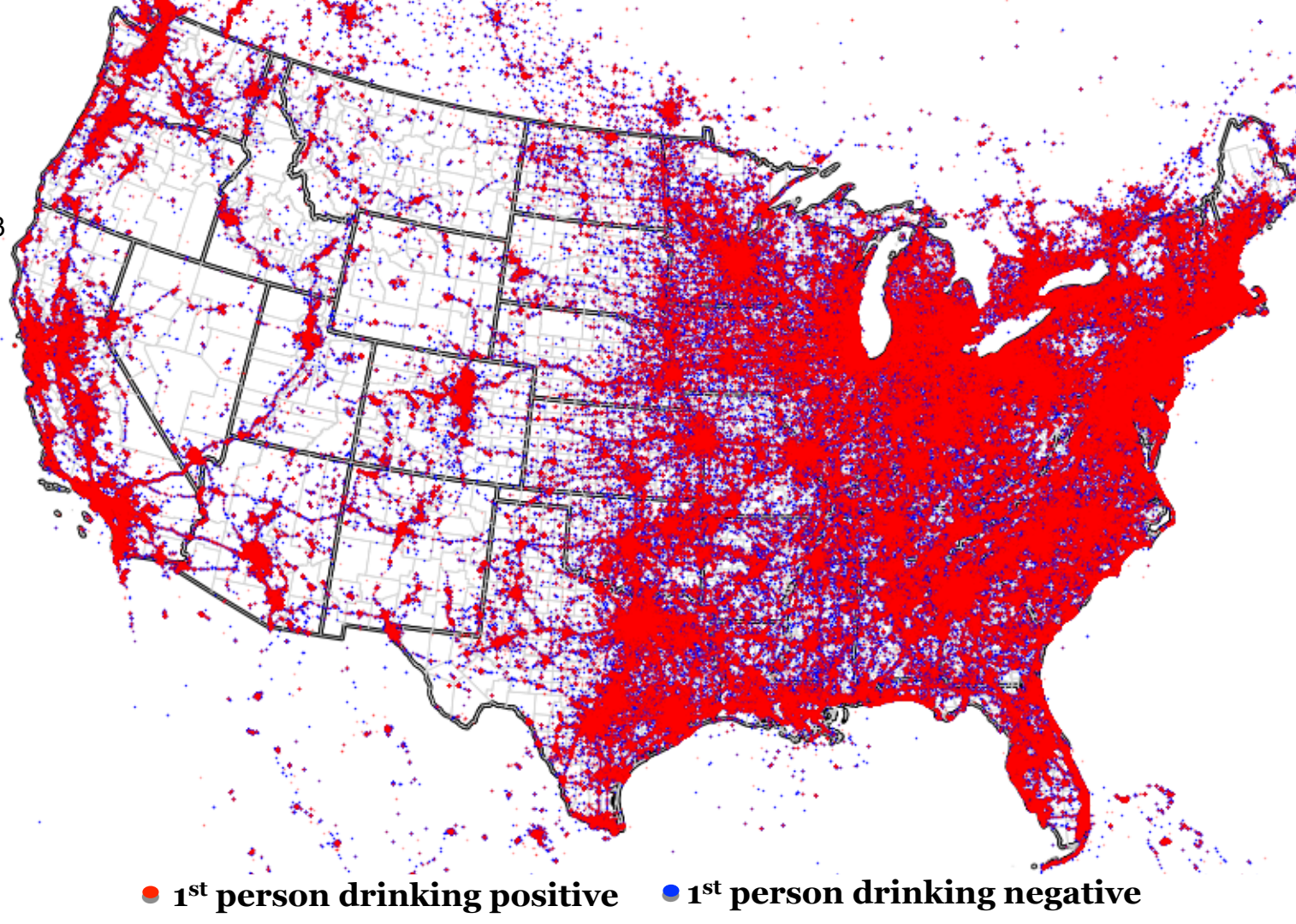


Analysis

POS PREVALENCE

California	13.3
Texas	9.3
New York	7.0
Florida	5.4
Pennsylvania	5.1
Ohio	4.6
Illinois	4.3
Michigan	3.5
New Jersey	3.4
North Carolina	2.8
Massachusetts	2.7
Georgia	2.4
Virginia	2.4
Washington	2.1
Arizona	1.9
Maryland	1.9
Indiana	1.9
Minnesota	1.7
Wisconsin	1.6
Oregon	1.5

…



Conclusion and Future Work

Conclusion

- Effective classifiers to detect first-person report of alcohol use, context, and consumption level
- Complement traditional monitoring of alcohol use
- Public health interventions and their evaluation

Future work

- Association among alcohol use and reports of problems such as *fighting*, *accident*, *vomiting*, *DUI*

Acknowledgement: This work was supported by funding from (NIAAA R21AA023901, PI: Weitzman) and (NIGMS R01GM114355, PIs: Elhadad,Savova).

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

Thank you!