

Text classification in the wild

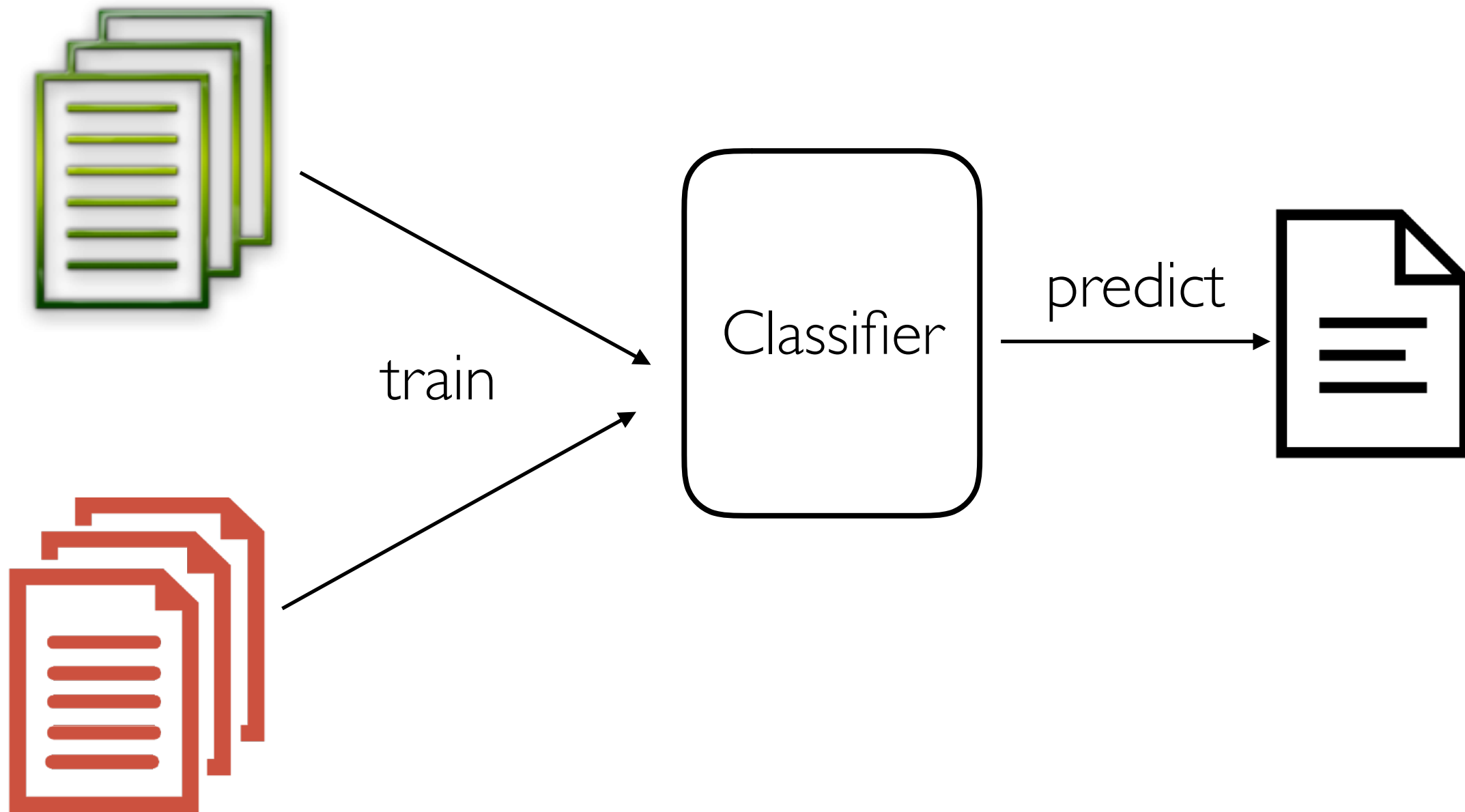
March 24, 2017

Aron Culotta

Assistant Professor
Department of Computer Science
Illinois Institute of Technology
aculotta@iit.edu
<http://cs.iit.edu/~culotta/>

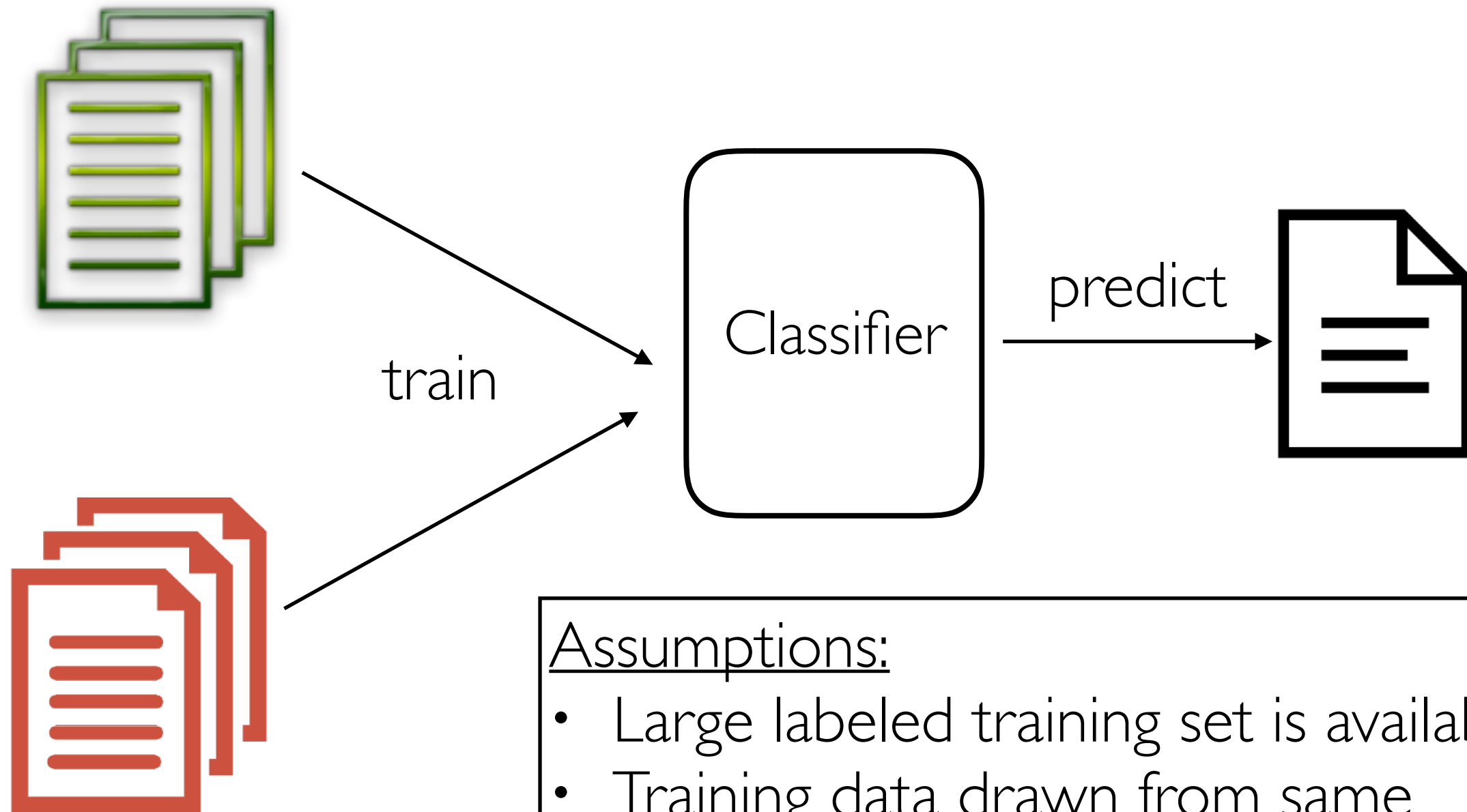
Labeled training data

Unlabeled testing data



Labeled training data

Unlabeled testing data

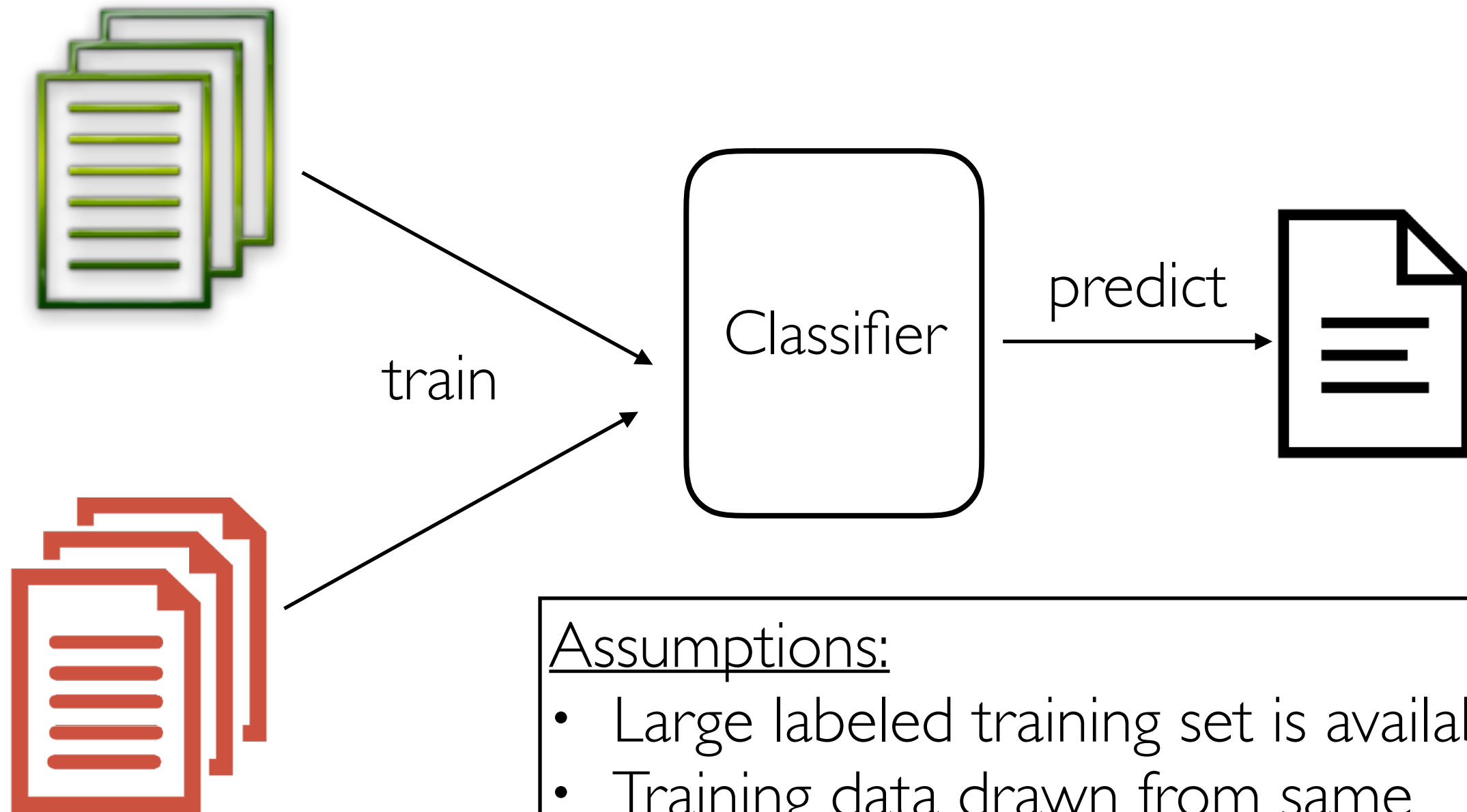


Assumptions:

- Large labeled training set is available
- Training data drawn from same distribution of testing data
- Testing data is non-stationary

Labeled training data

Unlabeled testing data



Assumptions:

- Large labeled training set is available
- Training data drawn from same distribution of testing data
- Testing data is non-stationary

Rarely (if ever) true in practice

Outline

- Domain adaptation
 - Identifying hazardous products from Amazon reviews
- Confounding variables
 - Predicting health attributes of social media users
- Learning from bags of instances
 - Predicting demographics of social media users

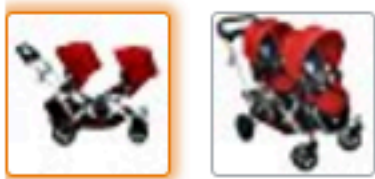
Identifying Amazon reviews that precede product recalls

[Bhat & Culotta 2017]

Contours Options Tandem Stroller, Ruby



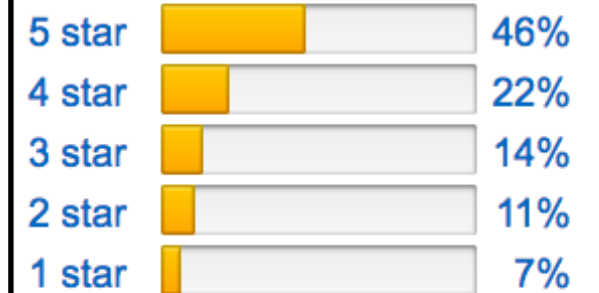
106 customer reviews



Customer Reviews

4.5 stars 106

3.5 out of 5 stars



[See all verified purchase reviews](#)



United States CONSUMER PRODUCT SAFETY COMMISSION

cpsc.gov

Recall date: July 24, 2012

Recall number: 12-233



Recall Summary

Name of product:

Contours Options LT Tandem Strollers

Hazard:

The front wheel assembly can break, posing a fall hazard to the child in the stroller. In addition, for strollers manufactured in January and February 2012, the nuts that hold the stroller's basket support screws in place can detach. Detached nuts can pose a choking hazard to young children.



United States CONSUMER PRODUCT SAFETY COMMISSION

cpsc.gov

Recall date: July 24, 2012

Recall number: 12-233



Recall Summary

Name of product:

Contours Options LT Tandem Strollers

Hazard:

The front wheel assembly can break, posing a fall hazard to the child in the stroller. In addition, for strollers manufactured in January and February 2012, the nuts that hold the stroller's basket support screws in place can detach. Detached nuts can pose a choking hazard to young children.

Customer Review

4 of 6 people found the following review helpful

★☆☆☆☆ **Flimsy, and the company is unresponsive - waste of \$,**
December 10, 2010

By [Madfoot](#)

This review is from: Contours Options Tandem Stroller, Ruby (Older Version)
(Discontinued by Manufacturer) (Baby Product)

I was so excited when I found this stroller. I loved that I could turn the seats forward or back, so the kids could face each other or me. It was a bit of a bear to get in and out of the back of my car, but I didn't mind because it was really handy. But after less than 4 months of use, it fell apart: the front end collapsed because the two pins holding it in place popped out. This happened just as I was leaving the house with my parents, and it was very embarrassing; they clearly thought I was trying to take their granddaughters out in a death-trap, and I am inclined to agree.

Customer Review

4 of 6 people found the following review helpful

★☆☆☆☆ **Flimsy, and the company is unresponsive - waste of \$,**

December 10, 2010

By [Madfoot](#)

This review is from: Contours Options Tandem Stroller, Ruby (Older Version)
(Discontinued by Manufacturer) (Baby Product)

I was so excited when I found this stroller. I loved that I could turn the seats forward or back, so the kids could face each other or me. It was a bit of a bear to get in and out of the back of my car, but I didn't mind because it was really handy. But after less than 4 months of use, it fell apart: the front end collapsed because the two pins holding it in place popped out. This happened just as I was leaving the house with my parents, and it was very embarrassing; they clearly thought I was trying to take their granddaughters out in a death-trap, and I am inclined to agree.

Customer Review

4 of 6 people found the following review helpful

★☆☆☆☆ **Flimsy, and the company is unresponsive - waste of \$,**

December 10, 2010

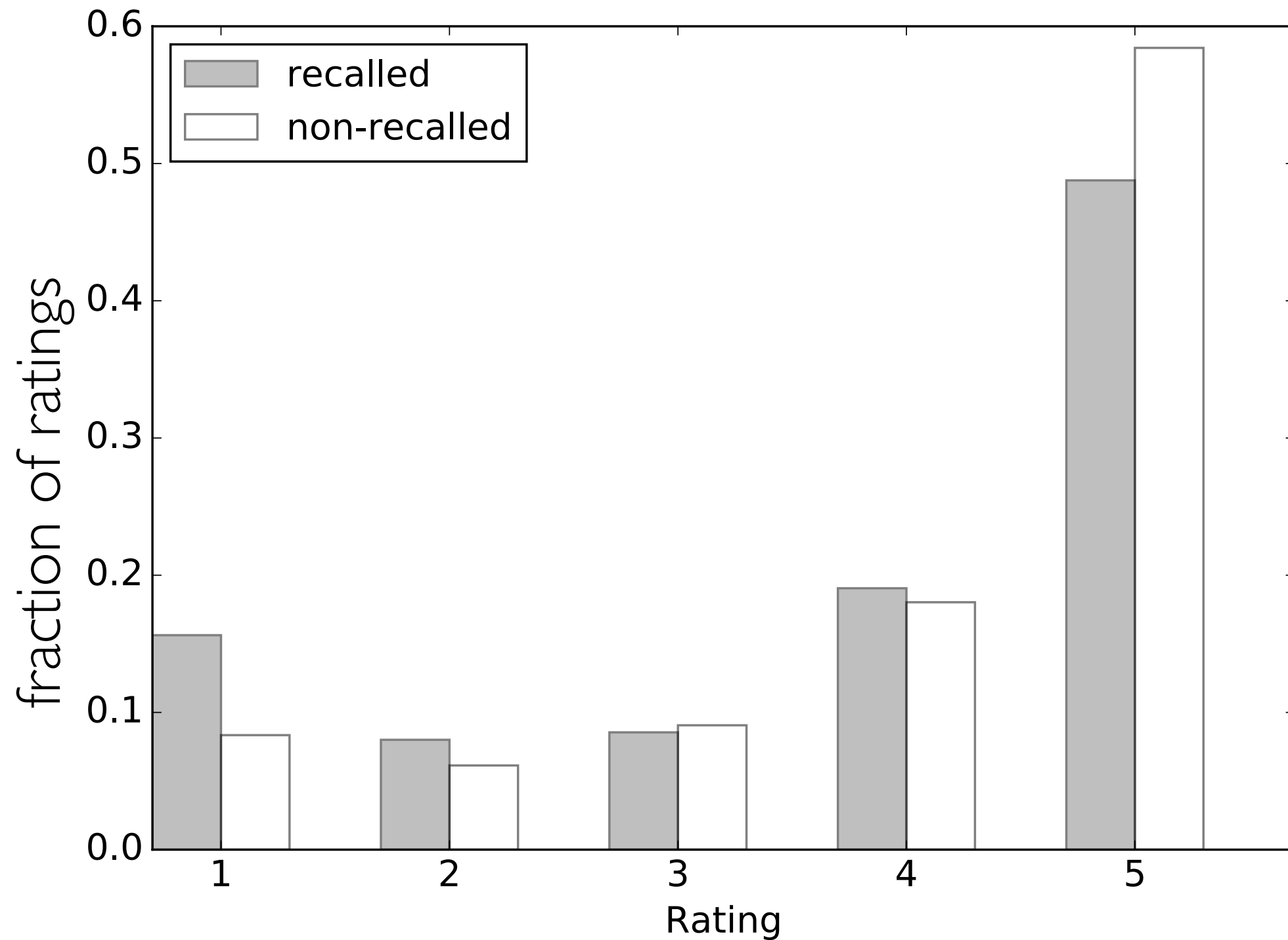
By [Madfoot](#)

This review is from: Contours Options Tandem Stroller, Ruby (Older Version) (Discontinued by Manufacturer) (Baby Product)

I was so excited when I found this stroller. I loved that I could turn the seats forward or back, so the kids could face each other or me. It was a bit of a bear to get in and out of the back of my car, but I didn't mind because it was really handy. But after less than 4 months of use, it fell apart: the front end collapsed because the two pins holding it in place popped out. This happened just as I was leaving the house with my parents, and it was very embarrassing; they clearly thought I was trying to take their granddaughters out in a death-trap, and I am inclined to agree.



Recalled products also have good reviews.



Can we build a text classifier to identify Amazon reviews indicating a potential health or safety hazard?

Desired labeled data

“...My son almost suffocated to death in this...”

“...My baby rolled forward and hit his head...”

“...My daughter got a bump on her head...”

positive

review mentioning a hazard

“...awesome stroller!!...”

“I’m sending this product back today!!! ... This seat was so heavy I could hardly get it out of the box.”

“It’s cheaply made. I washed it on the gentle cycle and it began to fall apart :(“

negative

no mention of hazard
(though can be other
problems with product)

Problem

Positive examples are rare.



► Report Details

► Product Details

▼ Incident Details

Incident Description

Britax B-ready stroller seat hinge came apart completely as my child was climbing into the stroller. The footrest dropped and the child fell from the stroller, scratching her leg and arm as she fell. There is no way to repair the hinge. The footrest no longer works properly.

Incident Date 5/21/2015

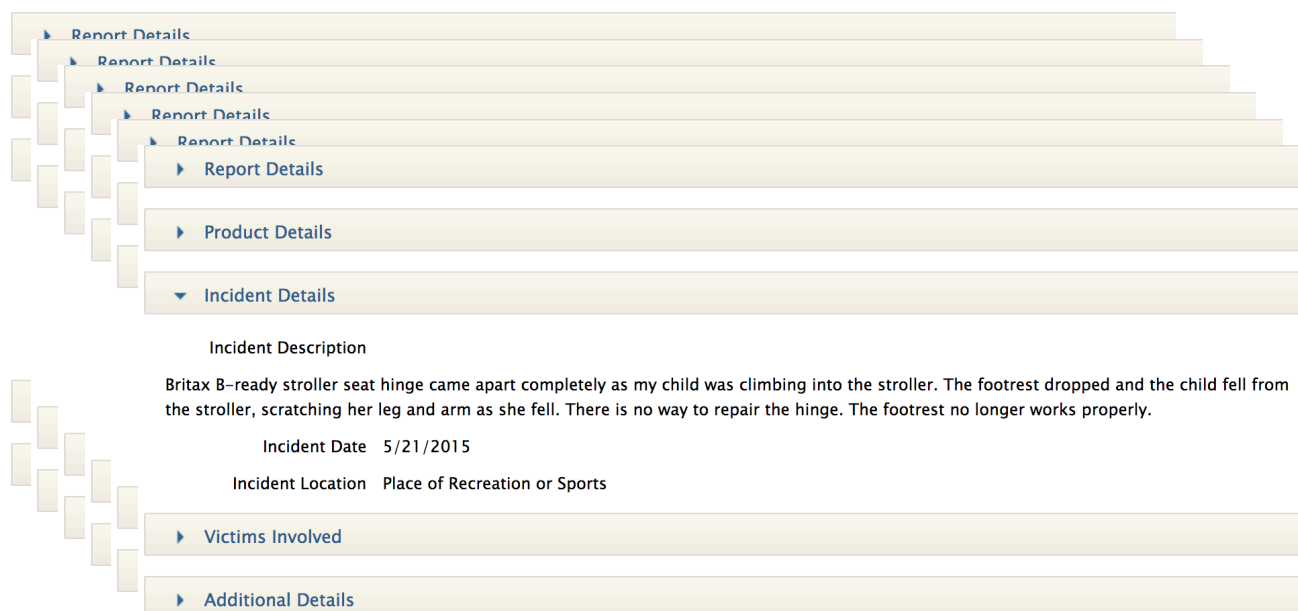
Incident Location Place of Recreation or Sports

► Victims Involved

► Additional Details

Positive/Unlabeled Learning

[Li & Liu 2005]



Positive examples from
one domain



4 of 4 people found the following review helpful

★★★★★ **Fantastic :) use with Phil&Teds Navigator,**
February 15, 2013

By [Erica Helf](#)

Verified Purchase ([What's this?](#))

This review is from: Phil&teds TS3 Car Seat Adapter for Graco Classic Connect To Sport, Dot, Classic and Navigator Strollers - Main Seat (Baby Product)

After much searching I found wording that all accessories made for phil&teds explorer will fit the newer model P&T's

Unlabeled examples from
another domain
assumed to be negative

Baseline Classifier

- Logistic regression
- Positive examples from SaferProducts.gov
 - “Babies & Kids” category
- Negative examples: random sample from Amazon reviews
 - “Baby” category
 - Optionally restrict to those with star rating $> T$

Problem with the baseline classifier

Top terms for positive class:

mold

pampers

fell

crib

rock

dangerous

night light

hazard

gate

rash

Problem with the baseline classifier

Top terms for positive class:

mold

pampers

fell

crib

rock

dangerous

night light

hazard

gate

rash

Selection bias

- Certain products are more common in complaints to CPSC than on Amazon
 - cribs, night lights, gates
- Traditional algorithms assume training data are drawn from same distribution as testing data
- Here, $p_{\text{train}}(y|x) \neq p_{\text{test}}(y|x)$

Solution: Feature Weighting

Idea:

- increase the feature values for terms that correlate with positive instances in the Amazon data

“This crib is very dangerous”

“crib”	“is”	“very dangerous”	“zebra”	“very”	...
1	1	1	0	1	
1.2	1.01	17.4	0	1.5	

Solution: Feature Weighting

Of course, we don't know which reviews are positive.
So, we estimate with the baseline classifier:

$$p(y = 1 \mid x_i = 1) \approx \frac{C(\hat{y} = 1 \wedge x_i = 1)}{C(x_i = 1)}$$

- \hat{y} class label predicted by baseline
- x_i feature i
- $C(\cdot)$ document count

E.g., $p(y = 1 \mid \text{pampers} = 1) = .02$

Compared with .91 in original training data

Solution: Feature Weighting

With additional transformations, we scale probabilities for each term to be non-negative and have mean=1.

We use the transformed values and train on the same data as the baseline model.

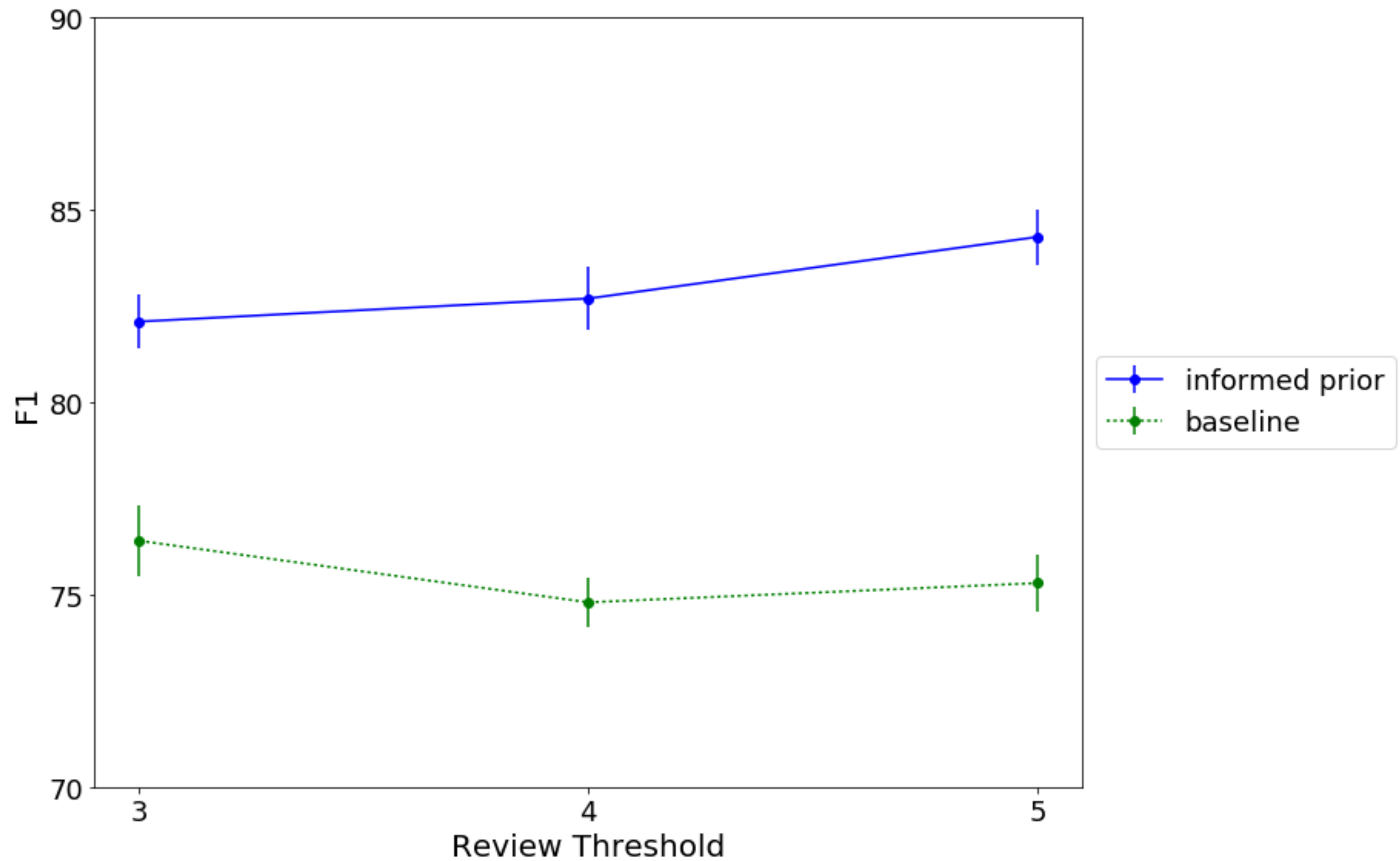
We also transform the values of the testing data.

We call this the “**informed prior**” method.

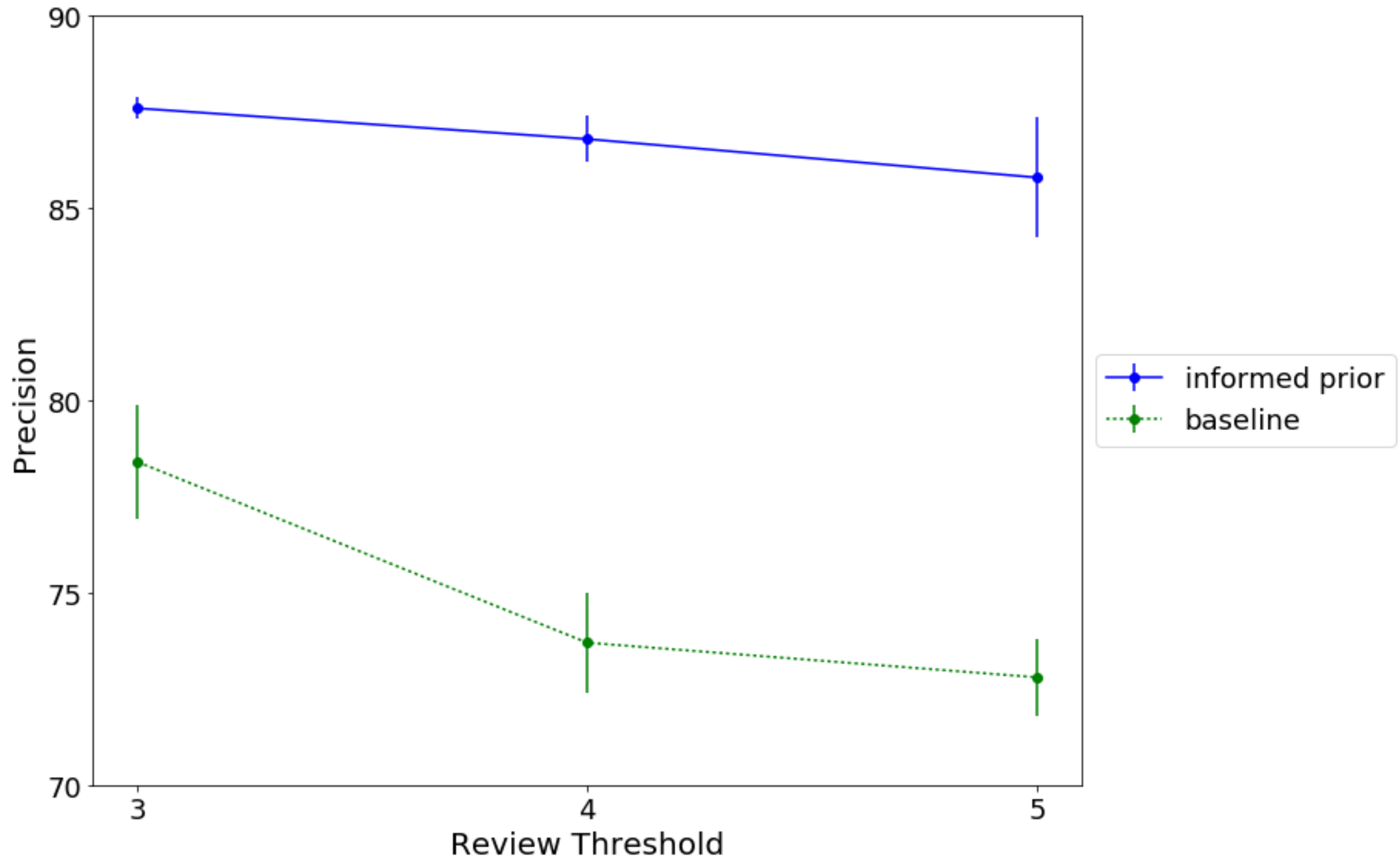
“crib”	“is”	“very dangerous”	“zebra”	“very”	...
1.2	1.01	17.4	0	1.5	

Model	Top terms
Informed prior	very dangerous, cpsc, mold, smacked, swallow it, emergency room, recalled, recall, was playing, hazard, is unsafe, snapped, leaned forward, the consumer, got stuck, was hanging, burnt, injured, exploded, was chewing
Baseline	mold, <i>pampers</i> , fell, <i>crib</i> , rock, dangerous, <i>night light</i> , hazard, broke, happened, <i>gate</i> , rash, <i>light</i> , recall, <i>model</i> , stuck, unsafe, caused, noticed, choking

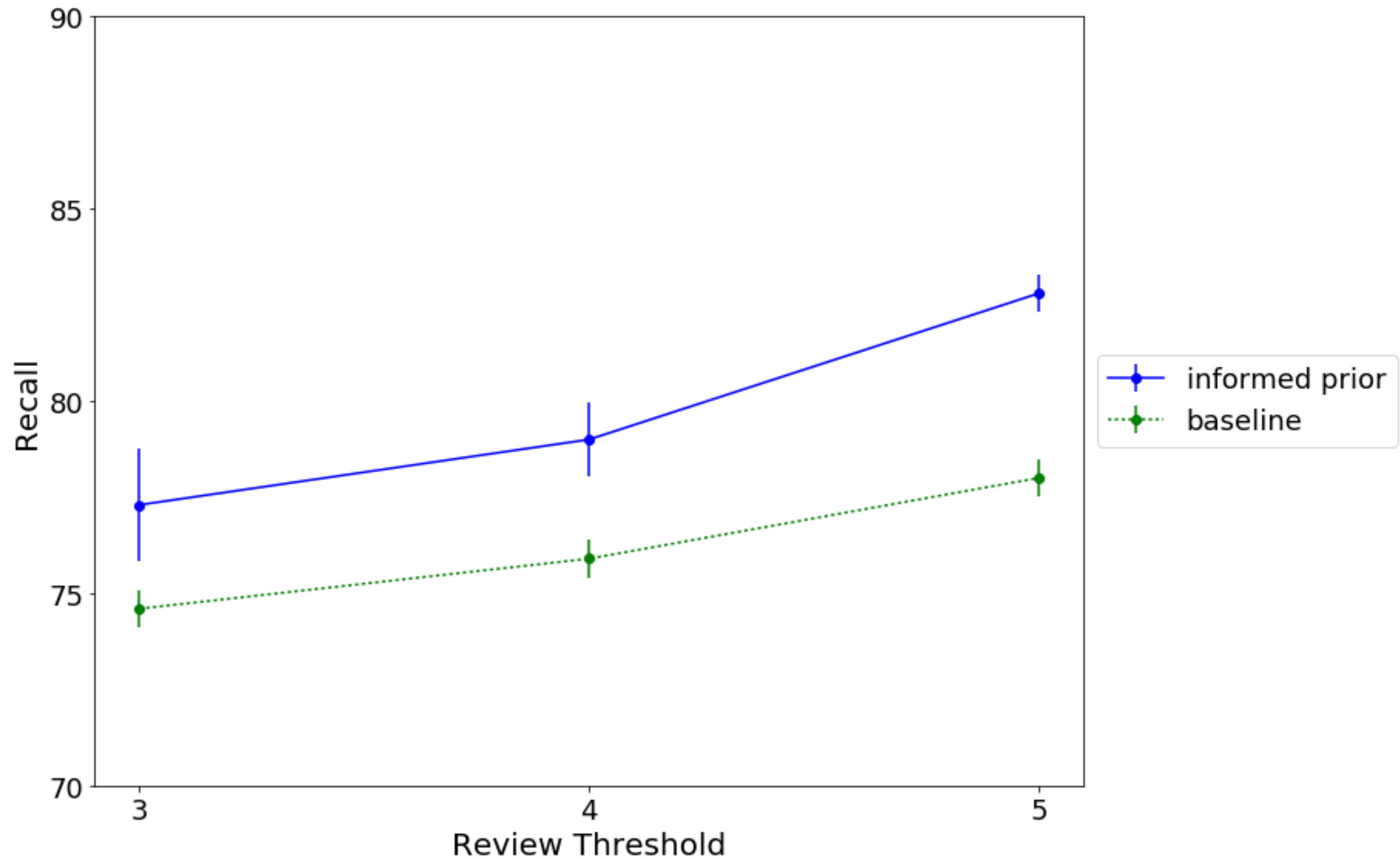
Review classification accuracy



Review classification accuracy

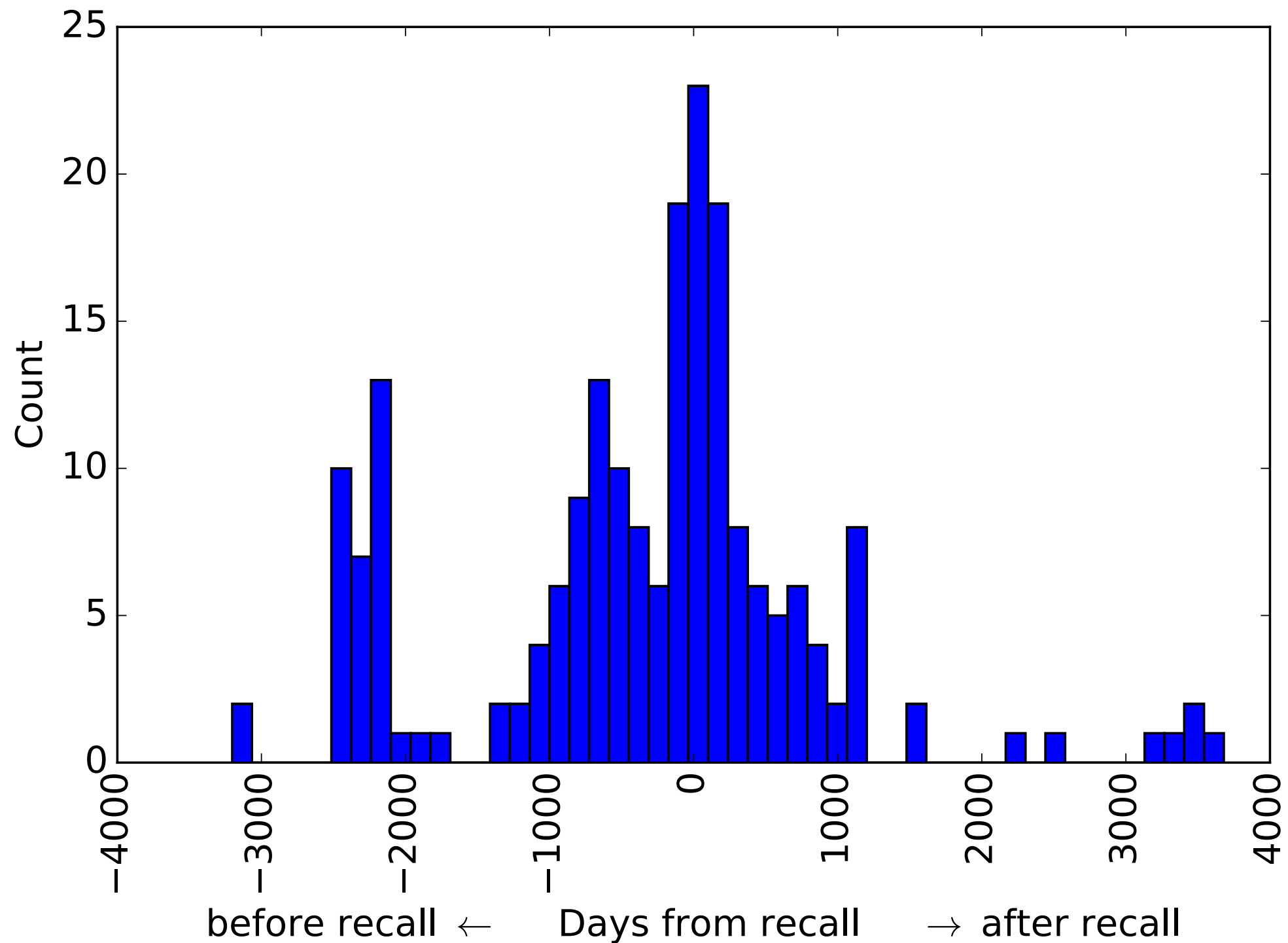


Review classification accuracy

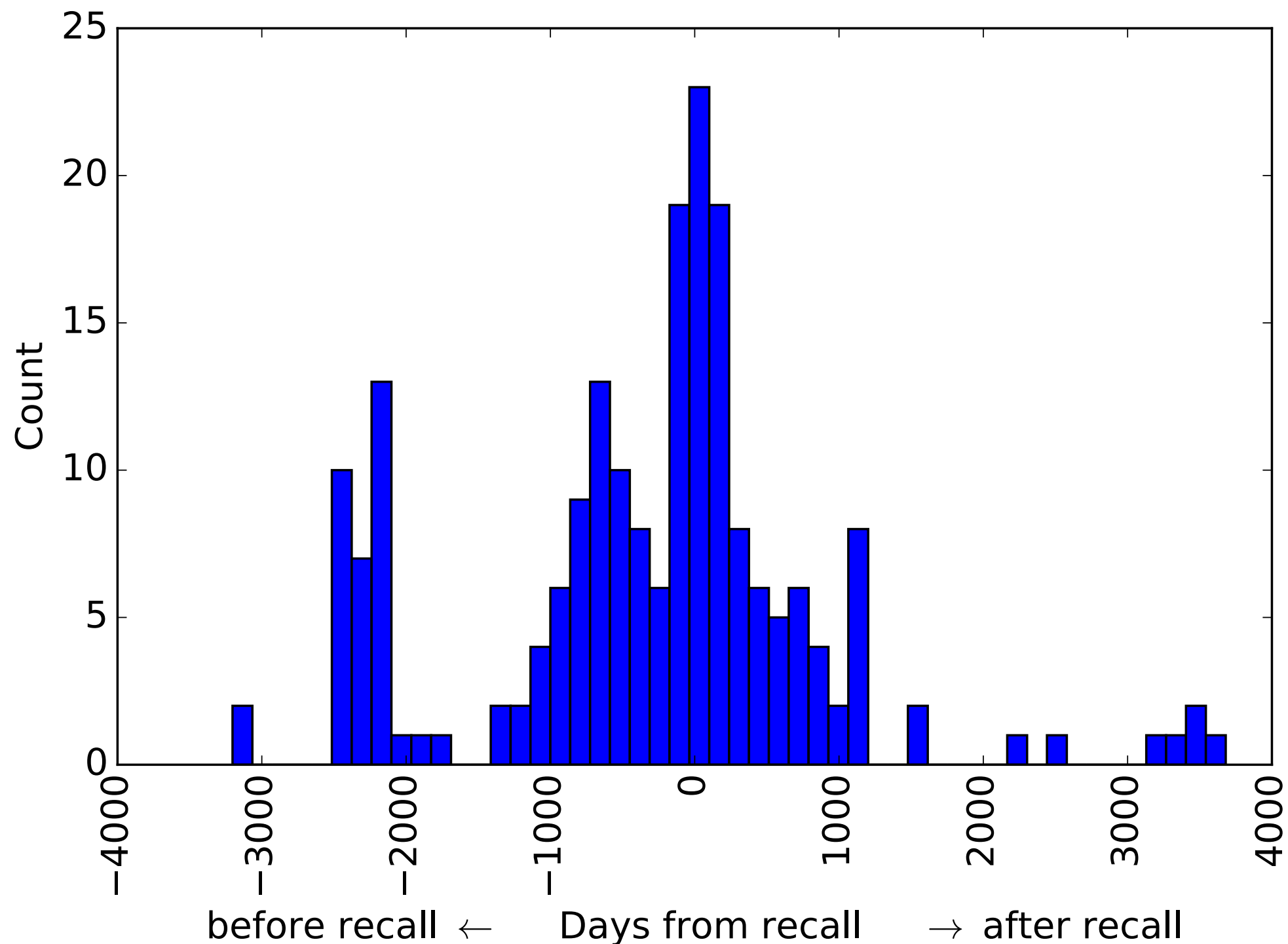


Are bad reviews leading indicators of recalls?

Are bad reviews leading indicators of recalls?

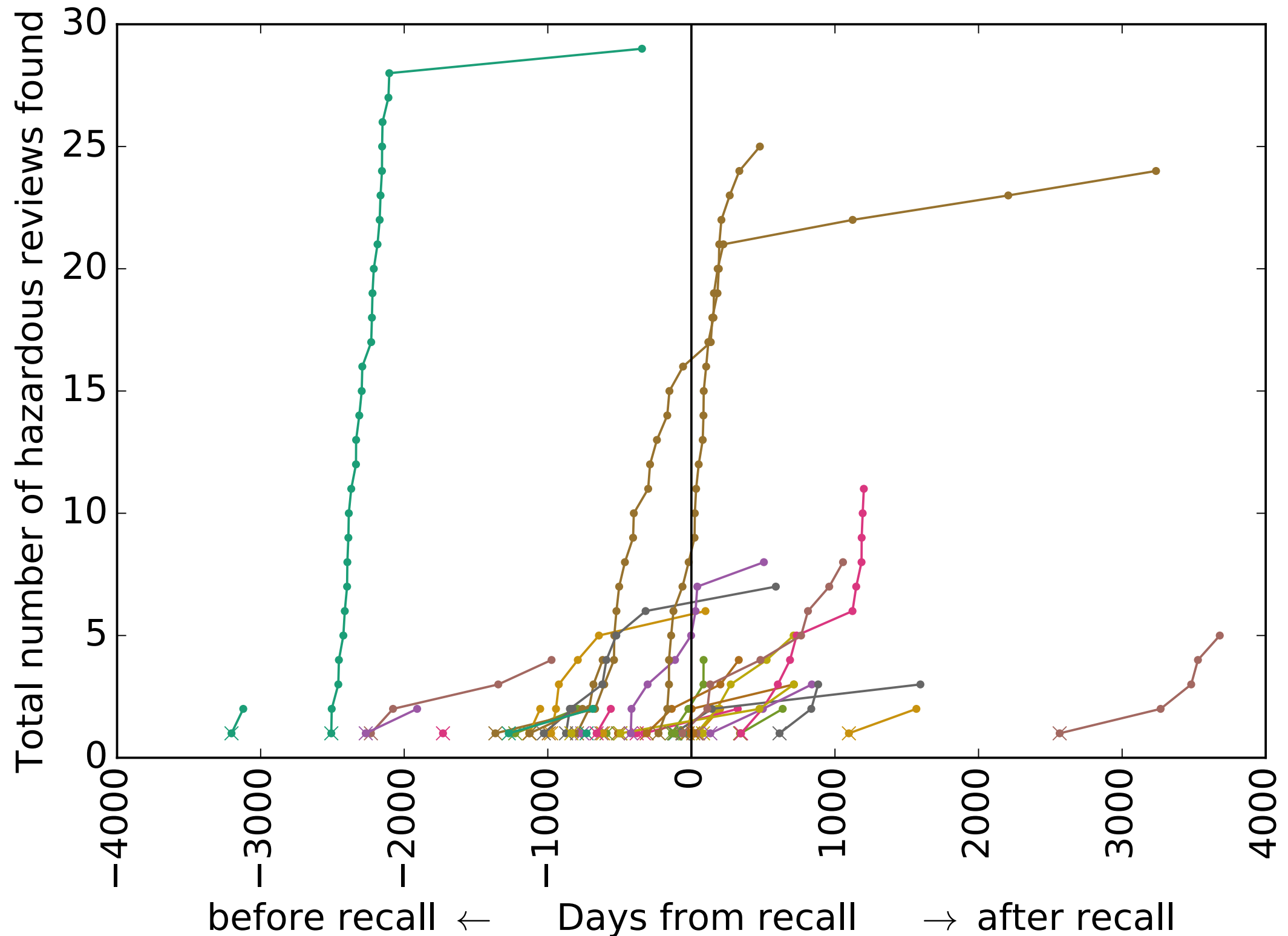


Are bad reviews leading indicators of recalls?



Hazardous review found prior to recall for 45% of products

Are bad reviews leading indicators of recalls?



Detected reviews often relate to recall

Review, 4/30/2013

“I’m not sure if this attachment has a defect or if it is only supposed to have one button on the adapter, but it makes the carseat very wobbly and unstable...Is mine defective? Everyone else seems to have great reviews, but mine is so unstable it seems dangerous.”

Recall, 6/4/2014

“The plastic adaptors used to connect an infant car seat to a stroller can crack, become unstable and break during use, posing a fall hazard to infants.”

Detected reviews often relate to recall

Review, 1/14/2007

“It’s a very poor design and needs a LOT of work. And my daughter ends up in a crumbled up ball on one side of the swing more times than not.”

Recall, 5/30/2007

“Infants can shift to one side of the swing and become caught between the frame and seat, posing an entrapment hazard.”

Conclusions: Review Classification

- Feature reweighting can adapt a model from one domain to another
- Allows us to train a classifier using “found” data
- Amazon reviews appear to provide leading indicators of subsequent product recalls

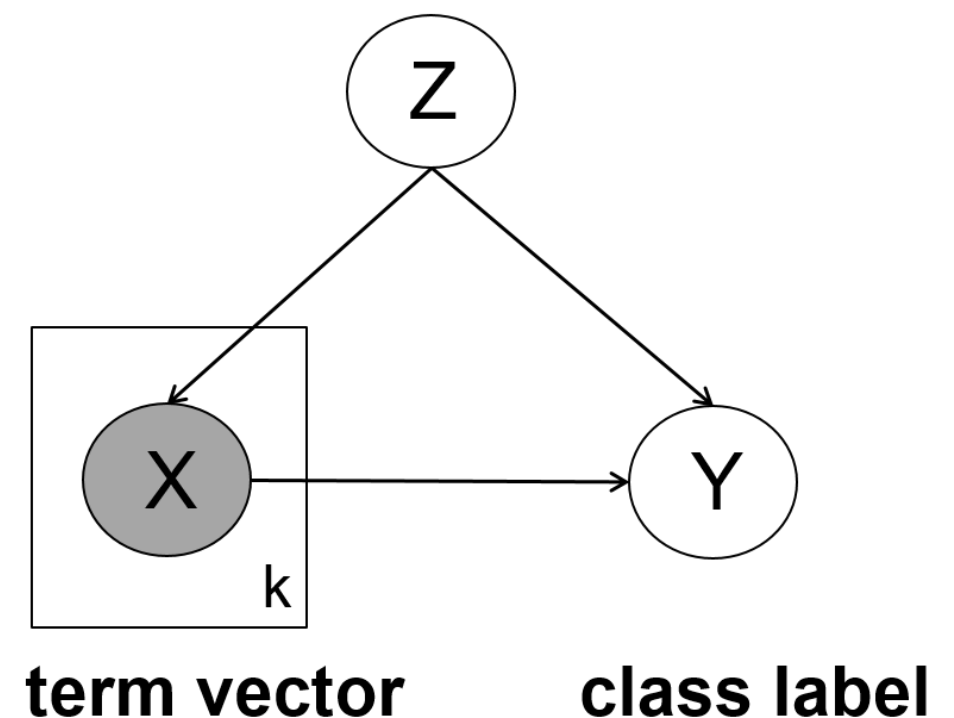
Code/data: <https://github.com/tapilab/icwsm-2017-recalls>

Controlling for confounds in text classification

[Landeiro & Culotta '16, '17]

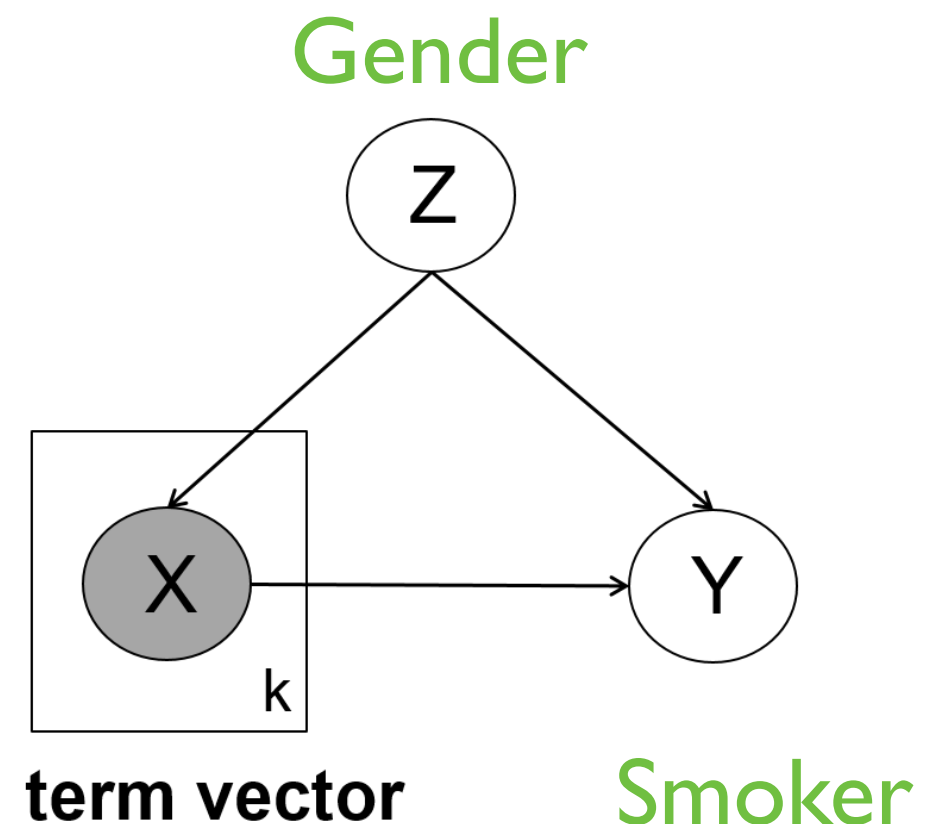
Confounds in classification

- In the review classification domain, we had $p_{\text{train}}(y|x) \neq p_{\text{test}}(y|x)$
- In other domains, there may be a confounding variable **z** such that $p_{\text{train}}(y|z) \neq p_{\text{test}}(y|z)$



Confounds in classification

- In the review classification domain, we had $p_{\text{train}}(y|x) \neq p_{\text{test}}(y|x)$
- In other domains, there may be a confounding variable z such that $p_{\text{train}}(y|z) \neq p_{\text{test}}(y|z)$

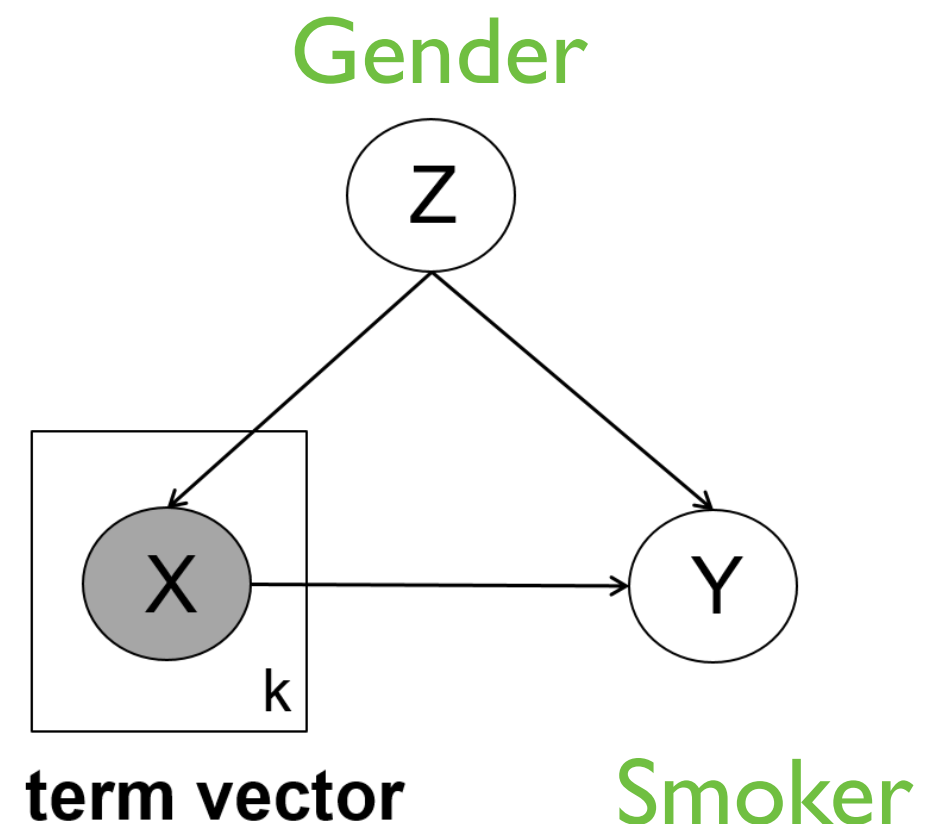


Confounds in classification

- In the review classification domain, we had $p_{\text{train}}(y|x) \neq p_{\text{test}}(y|x)$
- In other domains, there may be a confounding variable z such that $p_{\text{train}}(y|z) \neq p_{\text{test}}(y|z)$

“Observational studies from social media”

[Yom-Tov '16; De Choudhury & Kiciman '17; Altoff et. al '17; ...]



Confounds in classification

- Why does $p_{\text{train}}(y|z) \neq p_{\text{test}}(y|z)$ happen?
 - Limited training sample
 - Data drift over time

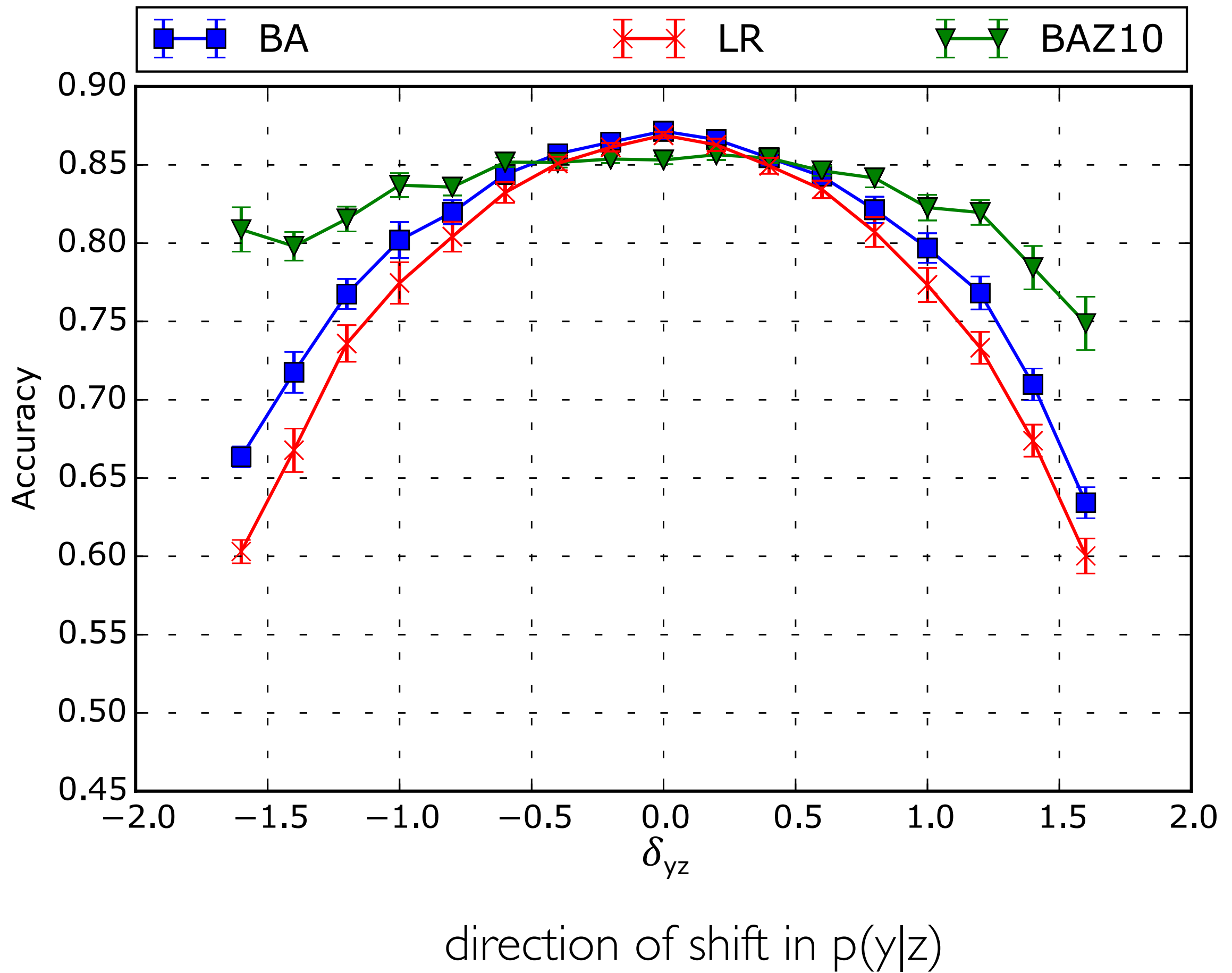
Adjusting for confounds in classification

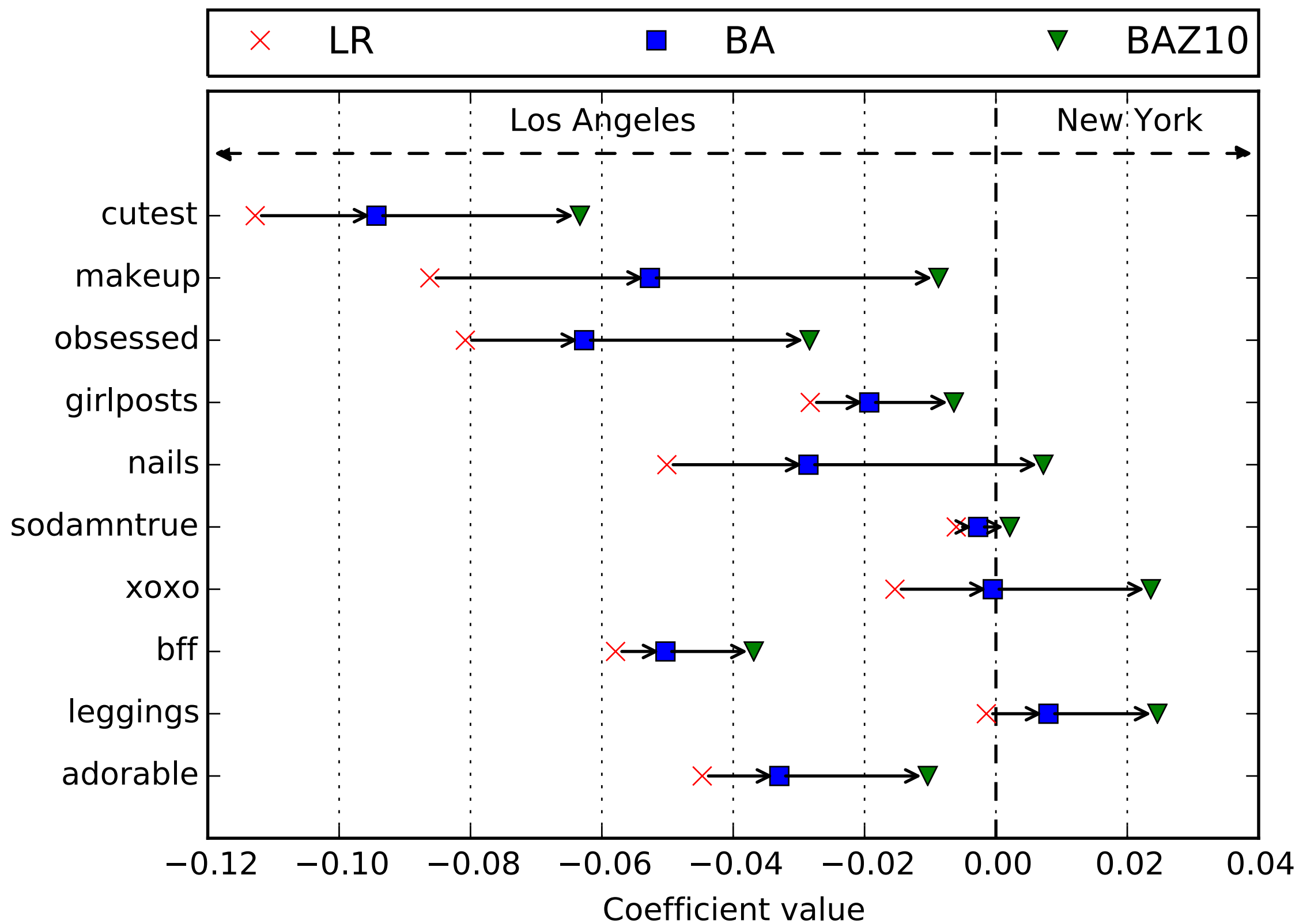
- Covariate adjustment (“backdoor adjustment”)
- Logistic regression estimates $p(y \mid \mathbf{x})$
 - Omitting \mathbf{z} leads to **omitted variable bias**
- Instead, estimate $p(y \mid \mathbf{x}, z)$
- At test time, when \mathbf{z} is unknown, compute:

$$\sum_{z \in Z} p(y \mid \mathbf{x}, z) p(z)$$

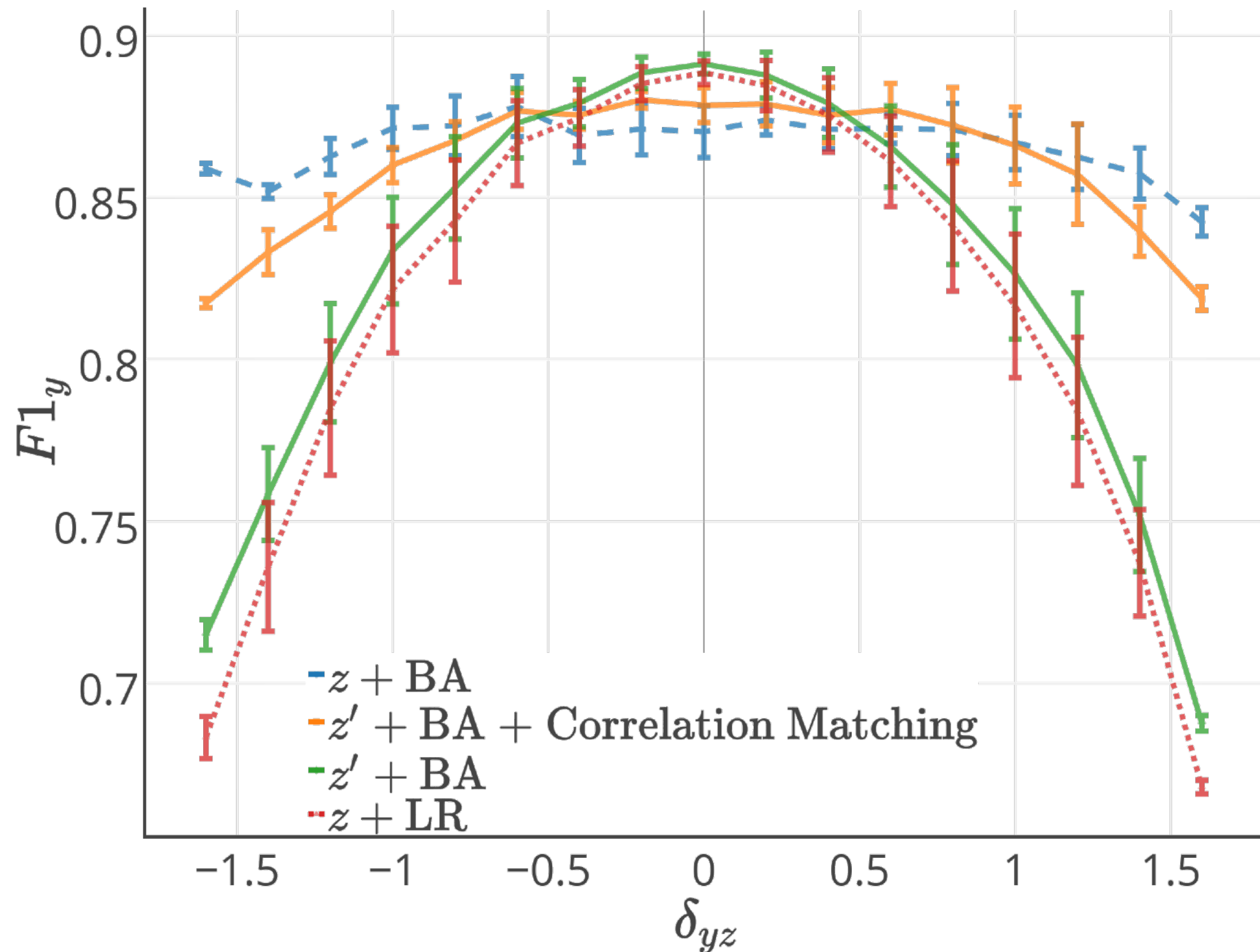
Experiments with confounders

- Task: predict the location of a Twitter user based on their tweets
 - Gender as a confounder: observed at training time, but not testing time
- To measure robustness to shift in $p(y|z)$, sample different train/test splits to vary $[p_{\text{train}}(y|z) - p_{\text{test}}(y|z)]$





Recently extended to cases where \mathbf{z} is observed with measurement error at training time. [Landeiro & Culotta '17]



Conclusions: Confounds in classification

- Omitted variable bias can significantly degrade accuracy of a text classifier.
- Covariate adjustment makes classifier robust to changes in confounding relationships
- Allows us to train a classifier using small, possibly biased datasets

Code/data: <https://github.com/tapilab/aaai-2016-robust>

Predicting Twitter user demographics with distant supervision

[Culotta et. al '16; Ardehaly & Culotta '16]

Learning from label proportions

- Often, it is difficult/unethical to get labeled training data at the instance level
 - E.g., income level of a Twitter user?
- However, it is often possible to associate population statistics with **bags** of instances.
 - E.g.:
 - 27% of households in Cook County have income >\$100k
 - Create a bag of all Twitter users from Cook County

Can we fit a model to population statistics, then use it to classify individual instances?

Learning from label proportions

- Simple model: ridge regression used for classification
- For bag i , we have
 - The mean feature vector z_i
 - The given label proportions \tilde{p}_i
- Set parameters by minimizing penalized squared error

$$E(\theta) = \sum_i (\tilde{p}_i - z_i^T \theta)^2 + \lambda \|\theta\|^2$$

- Classify a new instance as positive if $x^T \theta > .5$

LLP Experiments

- Predict 7 attributes of Twitter users
 - gender, age, income, political party, education, children, race/ethnicity
- Use population statistics from QuantCast.com
 - E.g., 19% of web users who visited lifehacker.com have a graduate degree
 - Bag consists of a sample of users that follow @lifehacker on Twitter
- In [Ardehaly & Culotta '16], we also experiment with county statistics and name statistics.

LLP Experiments

F1 results on validation data

	Friends		Text		Friends + Text	
	distant	full	distant	full	distant	full
Gender	.75	.66	.86	.84	.87	.84
Ethnicity	.60	.68	.86	.86	.81	.86
Politics	.80	.83	.56	.73	.74	.73
Average	.72	.72	.76	.81	.81	.81

distant = LLP model

full = fully supervised logistic regression

friends = use features over who a user follows

text = use features over tweets from the user

Top follow features

Education	No College College Grad School	YouTube, PlayStation, RockstarGames, Xbox, IGN StephenAtHome, WIRED, ConanOBrien, mashable nytimes, WSJ, NewYorker, TheEconomist, washingtonpost
Children	No Kids Has Kids	NewYorker, StephenAtHome, nytimes, maddow, pitchfork parenting, parentsmagazine, HuffPostParents, TheEllenShow, thepioneerwoman
Income	\$0-50k \$50-100k \$100-150k \$150k+	YouTube, PlayStation, IGN, RockstarGames, Drake AdamSchefter, cnnbrk, SportsCenter, espn, ErinAndrews WSJ, espn, AdamSchefter, SportsCenter, ErinAndrews WSJ, TheEconomist, Forbes, nytimes, business

Top text features

Age	18-24	d, haha, album, x, xd, _:, actually, stream, wanna, im
	25-34	super, dc, baby, definitely, nba, pregnancy, wedding, even, entire, nyc
	35-44	star, fans, kids, tv, bike, mind, store, awesome, screen, son
	45-54	wow, vote, american, comes, ca, santa, county, boys, nice, high
	55-64	vote, golf, red, american, country, north, county, holiday, smile, 99,999
	65+	vote, golf, @foxnews, holiday, may, american, he, family, north, national
Politics	Democrat Republican	women, u, ain't, nyc, equality, la, voice, seattle, dc, @nytimes @foxnews, christmas, #tcot, football, county, morning, family, christians, country, obama's

Conclusions: Learning from label proportions

- To avoid annotating data, we can fit models to population statistics matched with bags of instances.
- Often meet or exceed accuracy of models trained on instance-level annotations.

Code/data: <https://github.com/tapilab/jair-2016-demographics>

Overall Conclusions

- It's difficult to get the data you want for text classification
- Instead, take the imperfect data you have and adapt the models to work with it
 - Domain adaptation: transfer model from one type of data to another
 - Robustness to confounders
 - Learning from label proportions

Thanks!

aculotta@iit.edu

<http://cs.iit.edu/~culotta>

<https://github.com/tapilab>