Learning Importance of Business Aspects from Yelp Reviews

Stephanie Mallard (smallard@stanford.edu), Bonnie Nortz (bnortz@stanford.edu)
CS 229, Stanford University

Motivation

8/12/2016

My boyfriend had been raving about Evvia since his last trip to Palo Alto a year ago and insisted we come here for a late dinner despite the jetlag.

And boy was it worth it, despite my very unwilling attitude initially. I had the impression that the Bay area tended to close earlier, and was worried we'd be the last people in the restaurant at 10:30pm but I was so very wrong.

I'm normally pretty indifferent about Greek food, but this was done so right. The ingredients were fresh and cooked simply, letting it speak for itself. I have to give the lamb chops another shout out here as well, which were probably the best I've had (I normally don't really like lamb that much either).

It was also a nice touch that the manager was walking around and checking in on things all night, even late into the evening when there were only a handful of tables left. It's always nice to see that level of dedication imo.

- What is important to this customer?
  - food, manager dedication, hours
- *Goal*: Supervised learning algorithm to detect what aspects of a business are important to customers

Results

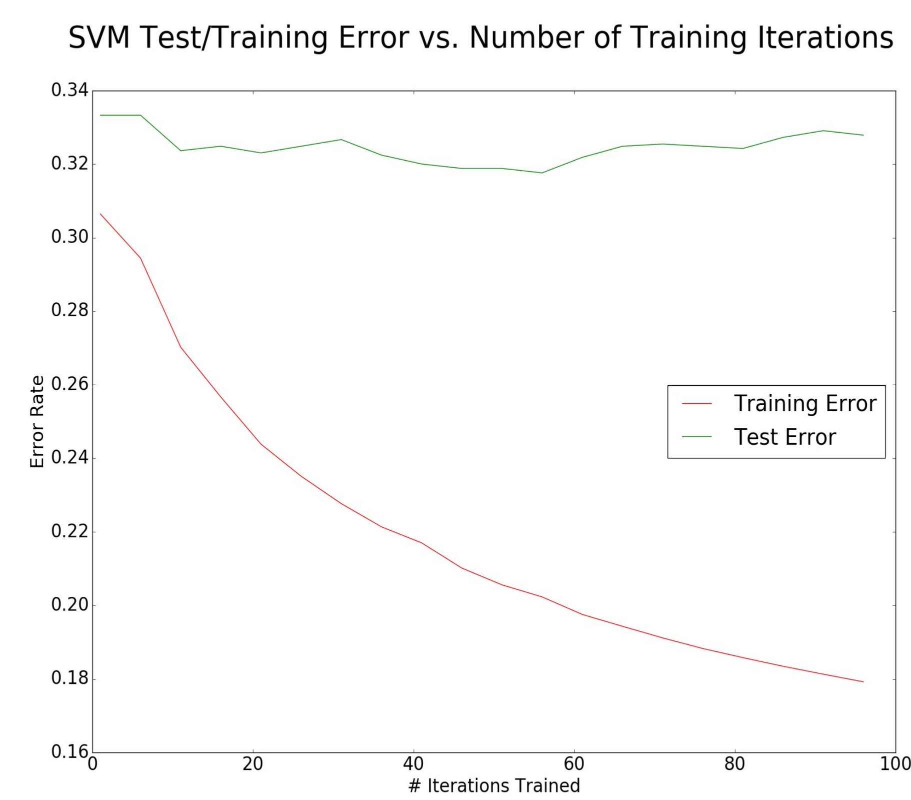
Naive Bayes

```
Extreme: 1,5 stars
Neutral: 2, 3, 4 stars
26.5% test error
Predicted 55 more extreme,
110 less extreme

Extreme: 1, 2, 5 stars
Neutral: 3, 4 stars
31.7% test error
Predicted 85 more extreme,
112 less extreme

Extreme: 1, 4, 5
Neutral: 2, 3
24.8% test error
Predicted 94 more extreme,
60 less extreme
```

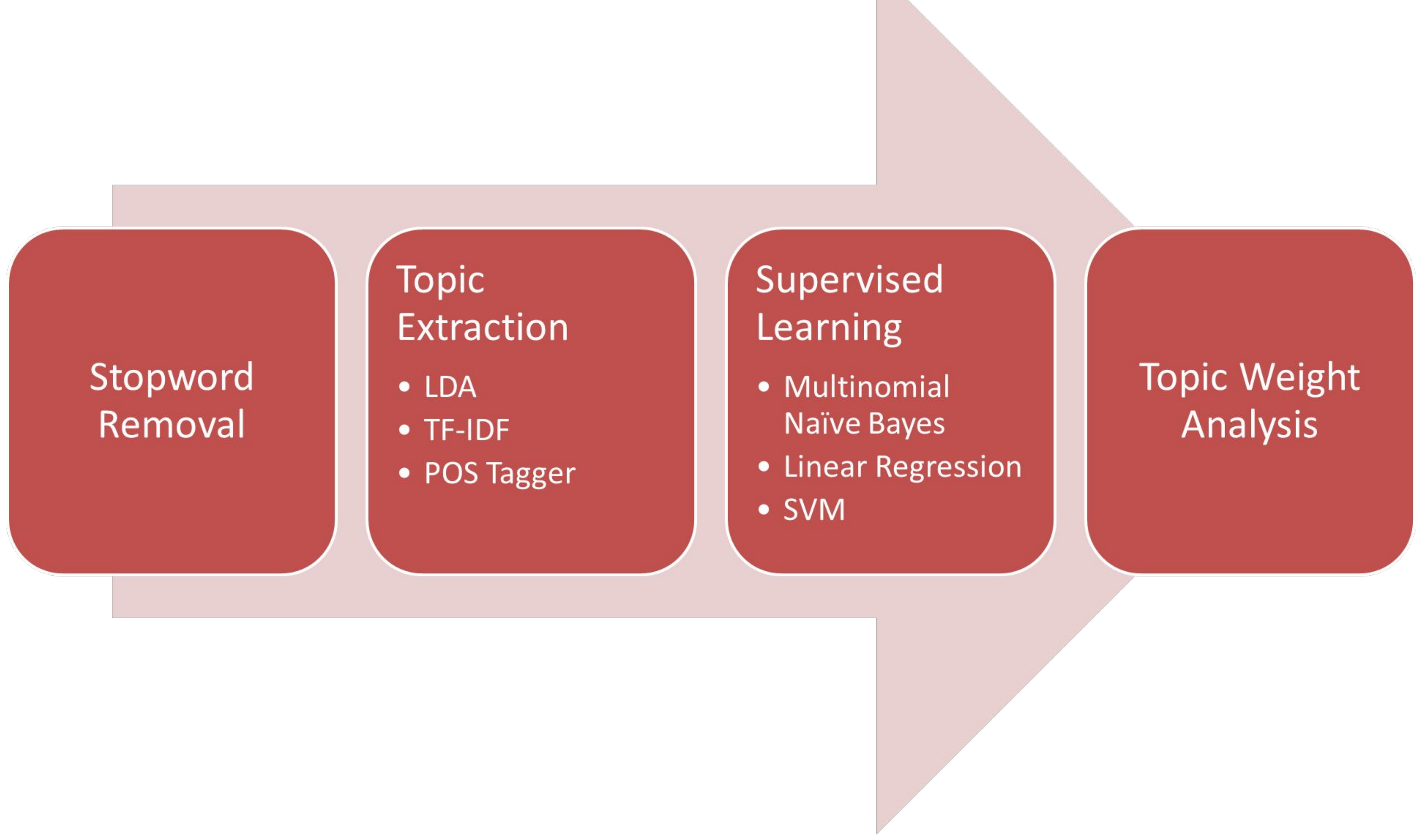
- Used plain preprocessed text as well as POS-tagged reviews as input
- Challenges: which star ratings do we label 'extreme' and which are 'neutral'?
- Possible solution: add noise into dataset (flip extreme labels to neutral or other way around) to make extreme and neutral reviews more similar
- We tried 5%, 10%, and 20% noise, result was more even prediction errors, but worse overall error

SVM

- Used plain preprocessed text as well as POS-tagged reviews as input
- Challenges: how many iterations of SGD should we run so we don't overfit?
- Best result was with 50 iterations and step size 0.0005: 21% training error, 32% test error



Learning Pipeline



Key idea: topics in very good or very bad reviews are more important to customers than topics in more neutral reviews.

Analysis

- We had similar error results for just the preprocessed text and for the POS input, which tells us that this is a difficult prediction task
- Even though our error was relatively high, our main goal is to determine which topics are important, so the weights for topics still tell us something
- Example topics from the SVM give us a glimpse of what customers of these restaurants found important and what they did not
- Areas for future work: improve topic modeling so the analysis of topics requires less human scrutiny, and run a linear regression on individual businesses to determine where they succeed and fail

Example weights from SVM

Extreme positive weight implies 'important', extreme negative weight implies 'neutral'

owner	0.5625	parking	-0.5215
welcome	0.5115	options	-0.4825
clean	0.445	pricey	-0.469
delicious	0.418	expensive	-0.3155
employees	0.373	refills	-0.307
delectable	0.3245	beverage	-0.286
yum	0.32	decor	-0.2725
fresh	0.305	buffet	-0.238
poisoning	0.2985	retro	-0.233
manager	0.2795	overpriced	-0.2235