Aspect Detection via Weakly Supervised Co-Training

Daniel Hsu
Columbia University

Yahoo! FREP Speaker Series
August 18, 2020

Joint work with Giannis Karamanolakis and Luis Gravano
Ongoing work with Alina Beygelzimer, Giannis Karamanolakis, and others

Many slides / figures courtesy of Giannis Karamanolakis

User generated reviews

Dimensions ("aspects") of a review:

1. Food quality
2. Ambience
3. Service
4. …

yelp | Find Restaurants | Near Civic Center, Manhattan, NY | Log In | Sign Up

Tavern on the Green Claimed

1377 reviews

$$$ · American (New), Cocktail Bars Edit

Write a Review | Add Photo | Share | Save

Britni R.
Houston, TX
325 friends
3 reviews
2 photos
Share review
Embed review

10/8/2019

I went to Tavern on the green before they closed during the recession and remember it very fondly. On my recent visit to New York, I went for brunch and was incredibly disappointed. It has become very touristy, the booths and chairs are rundown, glasses were not clean, the food is mediocre food and the staff are so rude.

The location is ideal, but other than that, there is no reason for going.

Useful | Funny | Cool

Rich H.
New York, NY
145 friends
236 reviews
126 photos
Elite '2019
Share review
Embed review

10/11/2019
1 check-in

This is a great venue for food, drinks, Sunday Brunch and impressing your out of town guests. It's right in Central Park so gives you a good opportunity to walk around the park either before or after. Outdoor seating on nice days is your best option. Great service, great food and great ambiance. If you go during brunch, give their Bloody Mary a try. It's one of my favorites!

Useful | Funny | Cool

Stephanie G.
New York, NY
80 friends
58 reviews
60 photos
Elite '2019
Share review
Embed review

7/30/2019
3 photos

I would actually rate this place a 3.5 stars if I could. I came here for lunch with my husband during restaurant week. The restaurant and the patio are absolutely beautiful, and it is definitely a perfect date spot. The service was friendly and efficient. The food was good, but not amazing. I ordered a watermelon appetizer and got a salmon dish as my entree. The salmon was great, and it was on top of a corn chowder/bacon sauce. Very tasty. The watermelon feta salad however, was pretty bad. It was maybe 20 pieces of watermelon cubed, with 6 pieces of feta, drizzled with balsamic vinegar. I am obsessed with watermelon feta salads, and I make them frequently at home. I'm a pretty poor cook, so the fact that my salad is 10x better than theirs says a lot.

My husband also ordered the salmon, and he got the squash flatbread for his appetizer. He was very happy with both dishes. He also ordered the wine pairing for $17, and wow. The wine was spectacular. Very delicious.

All in all, we left satisfied. The watermelon salad really left a bad taste in my mouth for the restaurant, because it's pretty shocking that they would serve something that was so bad, when other dishes were rather tasty. I definitely think it's worth a visit because it's a classic NY restaurant, but if you're just in it for the food, I think there are many other restaurants with better tasting, cheaper dishes.

User generated reviews

Dimensions ("aspects") of a review:
1. Price
2. Image quality
3. Ease of use
4. …

amazon prime | Electronics | Televisions | EN

32-Inch 720p LED TV

★☆☆☆☆ **Great tv for six months then out goes the sound**
March 18, 2016
Size: 32-Inch | Style: TV | Verified Purchase

Great tv for six months then out goes the sound. Called samsung for help and the best they could do was send me on a 240 mile round trip to a samsung approved service center that is only open in the day time so I would have to take half a day off work twice to get it fixed. So unless you live close to a service center think twice about this tv.

★★★★★ **Good picture and very lightweight. But it would not stay powered off.**
December 31, 2015
Size: 32-Inch | Style: TV | Verified Purchase

I had this installed on the wall bracket in the exercise room and was watching a show while finishing up the installation. It is very light and seems well built cosmetically. Problem was when I used the remote to shut it off, it did go off, no audio or picture. Then 3 seconds later it came right back on again... spooky I know. So I shut it down again, it went off and right back on again. looking around to see if my wife had the smart remote from upstairs from the Samsung big screen playing games with my psyche, not to be found. So I boxed it up, sent Amazon a return request, they sent UPS the next day and I had the replacement in 2 days no charge for anything. This TV shuts off and is a welcome addition to the man cave and the exercise room. Note, this TV has a 19 volt DC power converter similar to your laptop charger. It is not a 120VAC direct power cord to the TV from the wall. The power supply sits on the floor and a cord to the TV. 5 stars for the replacement TV

★★★★☆ **A great TV for an amazing price**
January 16, 2016
Size: 32-Inch | Style: TV | Verified Purchase

I purchased this to replace an older Philips LCD TV of the same size; after going on 8 years it finally died. The first time I ordered this it arrived with a cracked screen, which I'm suspecting was due to the packaging on Amazon's end (or lack there of). The replacement arrived in perfect condition though and it's extremely user-friendly and easy to set up. It only has 2 HDMI ports, but since this a second TV that I use in my bedroom, it works fine for me. I don't have cable in my bedroom and the two HDMI ports work perfectly for my Roku and DVD player.

The picture quality of this is very good and the sound is exceptionally good for a TV that is so thin. The actual TV itself only weighs maybe five or six pounds. My only complaint is that the legs of the TV can only be mounted to the very ends, making it difficult to fit on a smaller surface; it barely fit on the top of my dresser. The legs are also only about an inch high, making it impractical to put anything underneath of it, as it will block the picture. (To visualize, my tiny Roku 2 barely fits underneath it without blocking the screen). Other than that, it's great TV and exceptionally well priced.

User generated reviews

- Users evaluate restaurants / products along different dimensions
- Review is unstructured text; overall rating is user-specific aggregate

A great TV for an amazing price

January 16, 2016

Size: 32-Inch | Style: TV | **Verified Purchase**

I purchased this to replace an older Philips LCD TV of the same size; after going on 8 years it finally died. The first time I ordered this it arrived with a cracked screen, which I'm suspecting was due to the packaging on Amazon's end (or lack there of). The replacement arrived in perfect condition though and it's extremely user-friendly and easy to set up. It only has 2 HDMI ports, but since this a second TV that I use in my bedroom, it works fine for me. I don't have cable in my bedroom and the two HDMI ports work perfectly for my Roku and DVD player.

The picture quality of this is very good and the sound is exceptionally good for a TV that is so thin. The actual TV itself only weighs maybe five or six pounds. My only complaint is that the legs of the TV can only be mounted to the very ends, making it difficult to fit on a smaller surface; it barely fit on the top of my dresser. The legs are also only about an inch high, making it impractical to put anything underneath of it, as it will block the picture. (To visualize, my tiny Roku 2 barely fits underneath it without blocking the screen). Other than that, it's great TV and exceptionally well priced.

Problem: Fine-grained aspect detection

- What is the aspect being addressed in a given segment of a review?
- **Task**: classify review segments into pre-defined aspect classes

Sentence	Aspect
Great Tv for the price.	**Price**
Easy to setup.	**Ease of Use**
The audio is ok for the tiny speakers.	**Sound Quality**
Much better than the 20" tube tv.	**General**

Canonical machine learning approaches

- **Supervised learning**:
  - Manually label review segments
  - Then fit a multi-class classification model
  - ☹️ Expensive annotation cost
- **Unsupervised learning**:
  - Fit a topic model to review segments
  - Then manually map topics to aspects
  - ☹️ Topics may not correspond to aspects of interest

Aspects may be specific to (say) a product;
annotation/modeling efforts may only be useful for specific product.

Our approach (Karamanolakis, H., Gravano, EMNLP 2019)

- **"Weakly-supervised" learning**
  - Ask users to provide, for each aspect, indicative "seed words" that appear in many review segments
  - Use seed words to automatically label review segments
  - Fit multi-class classifier to automatically-labeled review segments
- Building on ideas from:
  - Co-training (Blum & Mitchell, 1998)
  - "Seed word"-based weak supervision (Angelidis & Lapata, 2018)

Outline

1. Weak supervision via seed words
2. Interpretation as co-training
3. Empirical evaluation on product and restaurant reviews
4. Planned work on hidden bias detection

Outline

1. Weak supervision via seed words
2. Interpretation as co-training
3. Empirical evaluation on product and restaurant reviews
4. Planned work on hidden bias detection

What is a seed word?

- Seed word for an aspect: a weakly positive indicator of the aspect
  - "We can think of [seed words] as query terms that someone would use to search for segments discussing [the aspect]." (Angelidis & Lapata, 2018)
  - Domain-specific
  - Indicative, but not necessarily highly accurate

Aspect	**Seed Words**
Price	price, value, money, worth, paid
Image	picture, color, quality, black, bright
Sound	sound, speaker, noise, loud, volume

- Our method starts with a small set of seed words for each aspect.

How to get seed words?

1. Manually provided by domain expert

Aspect	Seed Words
Price	price, value, money, worth, paid
Image	picture, color, quality, black, bright
Sound	sound, speaker, noise, loud, volume

2. Automatically from small, labeled corpus (Angelidis & Lapata, 2018)

Great price for an excellent LED TV
Verified Purchase

Sentence	Aspect
Great Tv for the price.	**Price**
Easy to setup.	**Ease of Use**
The audio is ok for the tiny speakers.	**Sound Quality**
Much better than the 20" tube tv.	**General**

Aspect	Seed Words
Price	price, value, money, worth, paid
Image	picture, color, quality, black, bright
Sound	sound, speaker, noise, loud, volume

Why seed words?

- Potentially more valuable than aspect annotations for individual review segments
  - A seed word provides information about potentially many review segments
  - The aspect label for a review segment is only useful for that review segment

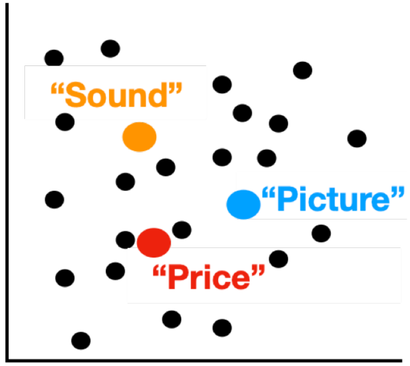
"**paid**"
~
[price]

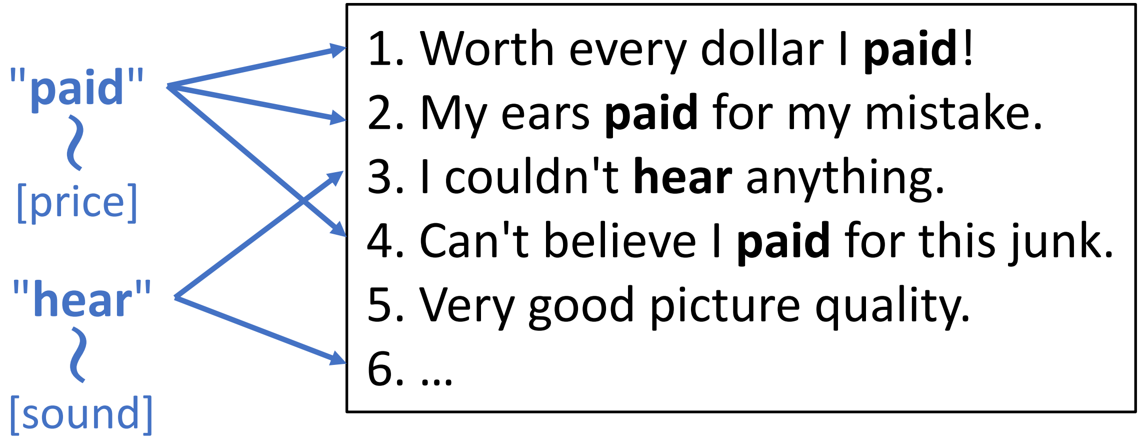
1. Worth every dollar I **paid**! [price]
2. My ears **paid** for my mistake.
3. I couldn't hear anything.
4. Can't believe I **paid** for this junk.
5. Very good picture quality.
6. …

- (Aspect labels still necessary for validation.)

How to use seed words?

- **Recent approaches**:
  - (Lund, Cook, Seppi, Boyd-Graber, 2017; Angelidis & Lapata, 2018)
  - Use seed words to **initialize** topic models or embedding models
- **Our approach**:
  - Fit multi-class model to a corpus weakly-labeled by seed words
  - (How? Why?)





Weak supervision via seed words

- Each seed word is associated with exactly one aspect
- Treat a review segment as a "bag of seed words"
  - NB: Some segments contain no seed words ☹. We label these "no aspect".
- Assign "soft label" $q = (q_1, \ldots, q_K)$ to review segment, where

$$q_k \propto \exp(\#\text{ words in seg. that are seed words for aspect } k)$$

$(q_1, \ldots, q_K)$

Fitting a multi-class model

- So far:
  1. Obtain seed words for each aspect
  2. Automatically assign "soft labels" q to all review segments x
- Now fit multi-class model (e.g., logistic model) to these weakly-labeled review segments (e.g., by minimizing cross entropy objective)

$$J(p) = \sum_{(x,q)\in S} \sum_{k=1}^{K} q_k \log p_k(x)$$

Highly reminiscent of co-training (Blum & Mitchell, 1998)!

Overall method

1. Obtain seed words for each aspect
2. Assign "soft labels" to all review segments
3. Fit multi-class model to these weakly-labeled review segments

+ Only Step 1 requires human supervision
+ In Step 3, model learns to predict aspects from non-seed words (and other possible context features as well)
+ We also propose an iterative (E-M type) scheme that refines the "soft labels" and then refines the multi-class model.

Outline

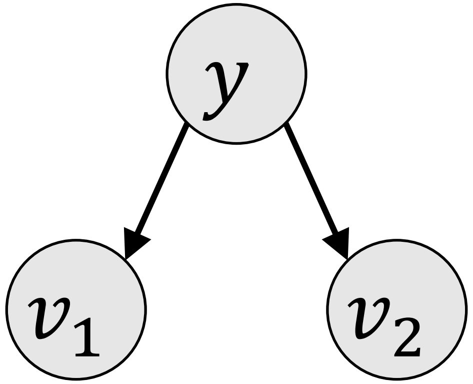
1. Weak supervision via seed words
2. Interpretation as co-training
3. Empirical evaluation on product and restaurant reviews
4. Planned work on hidden bias detection

Co-training

- Each data point has two somewhat redundant "views"
  - E.g., web pages:
    View 1 = words appearing on page
    View 2 = anchor text attached to links that point to the page
- **How to leverage redundancy?**

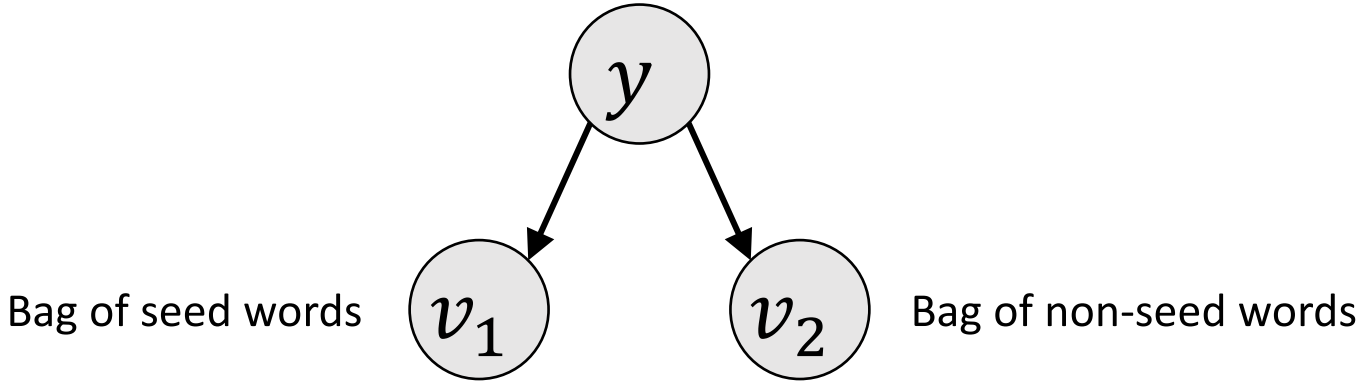
(Blum & Mitchell, 1998)

- Assume views v_1 and v_2 are cond. independent given label y.
- Weak classifier based on v_1 gives a useful (noisy) label for a classifier based on v_2



A bag-of-words model for review segments

- Assume words in review segment about aspect k are drawn iid from distribution P_k over a vocabulary
  - Some words in vocab are seed words; rest are non-seed words.
  - View 1 = "bag of seed words"
  - View 2 = "bag of non-seed words"
- **Under what conditions does our "weak supervision via seed words" act as a weak classifier?**



Seed word utility and robustness

- **Proposition**: A review segment of length L about aspect k^* is correctly (hard) labeled with probability $> 1/2$ if

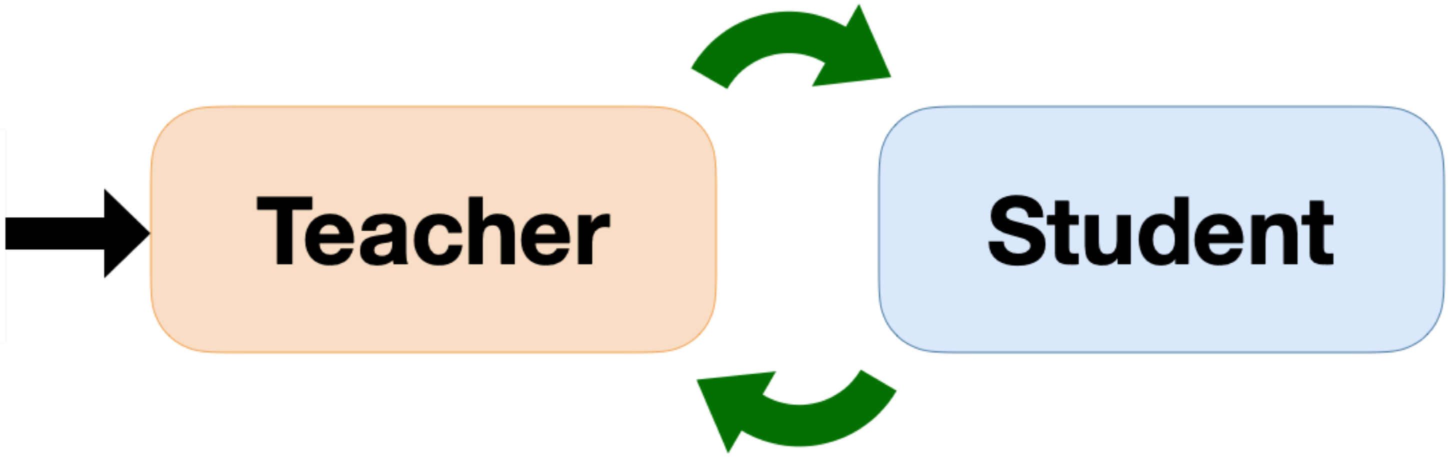
$$P_{k^*}(\mathrm{SW}_{k^*}) > \max_{k \neq k^*} P_{k^*}(\mathrm{SW}_k) + O\left(\sqrt{\frac{P_{k^*}(\mathrm{SW}_k)\log K}{L}} + \frac{\log K}{L}\right)$$

+ Probability condition only scales logarithmically with K
+ Only depends on mass assigned by P_{k^*} to all seed words of an aspect; not on any individual seed word probability (c.f. implicit "anchor word" assumption in Lund *et al*, 2017)

Other interpretations

- Distillation / model compression (Bucilua, Caruana, Niculescu-Mizil, 2006; Ba and Caruana, 2014; Hinton, Vinyals, Dean, 2015; …)
  - **Teacher**: "seed word"-based weak supervision
  - **Student**: multi-class classification model
- E-M algorithm (Dempster, Laird, Rubin, 1977; Seeger, 2000; …)

Aspect	Seed Words
Price	price, value, money, ...
Image	picture, color, bright, ...
Sound	sound, speaker, noise, ...



Outline

1. Weak supervision via seed words
2. Interpretation as co-training
3. Empirical evaluation on product and restaurant reviews
4. Planned work on hidden bias detection

12 data sets

- OPOSUM-Bags&Cases
- OPOSUM-Keyboards
- OPOSUM-Boots
- OPOSUM-Bluetooth Headsets
- OPOSUM-TVs
- OPOSUM-Vacuums
- SemEval-Restaurants-English
- SemEval-Restaurants-Spanish
- SemEval-Restaurants-French
- SemEval-Restaurants-Russian
- SemEval-Restaurants-Dutch
- SemEval-Restaurants-Turkish

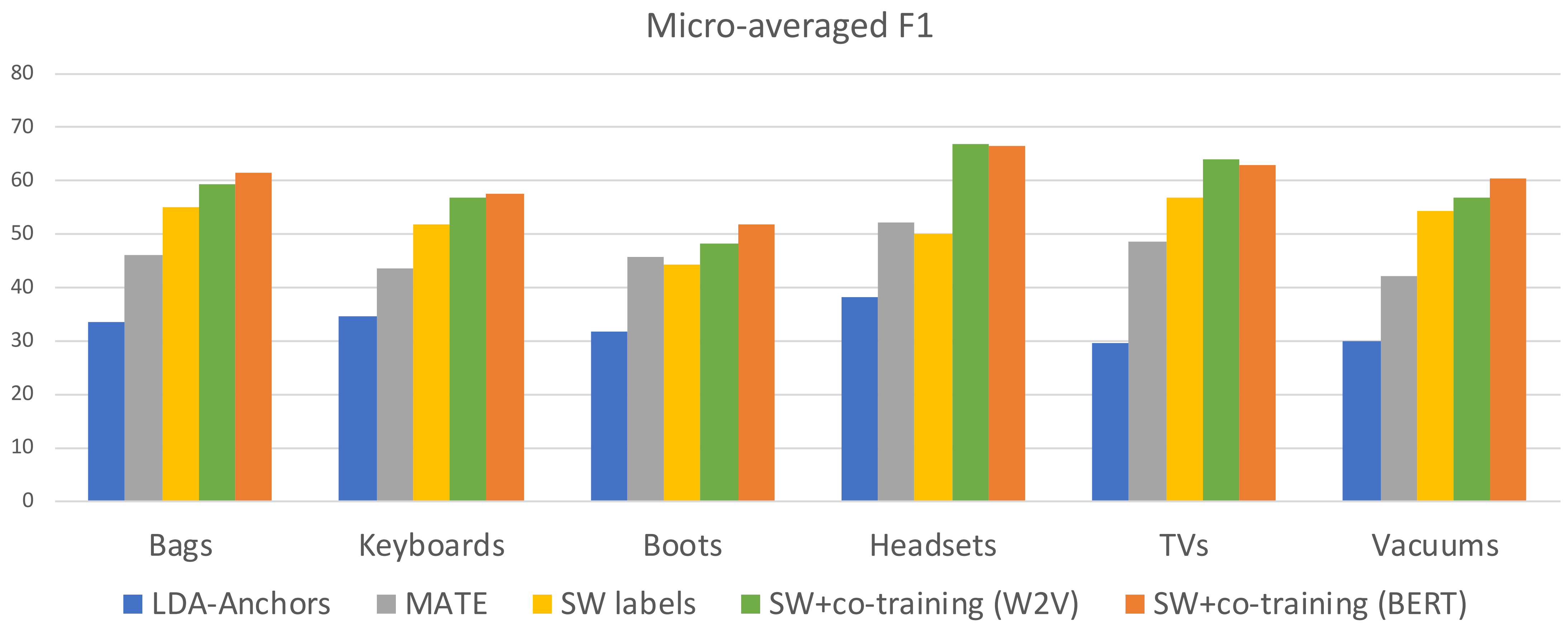
OPOSUM (product reviews)
9 aspects per domain: *quality, looks, price, …*

SemEval-2016 (restaurant reviews)
12 aspects per language: *ambience, service, food, …*

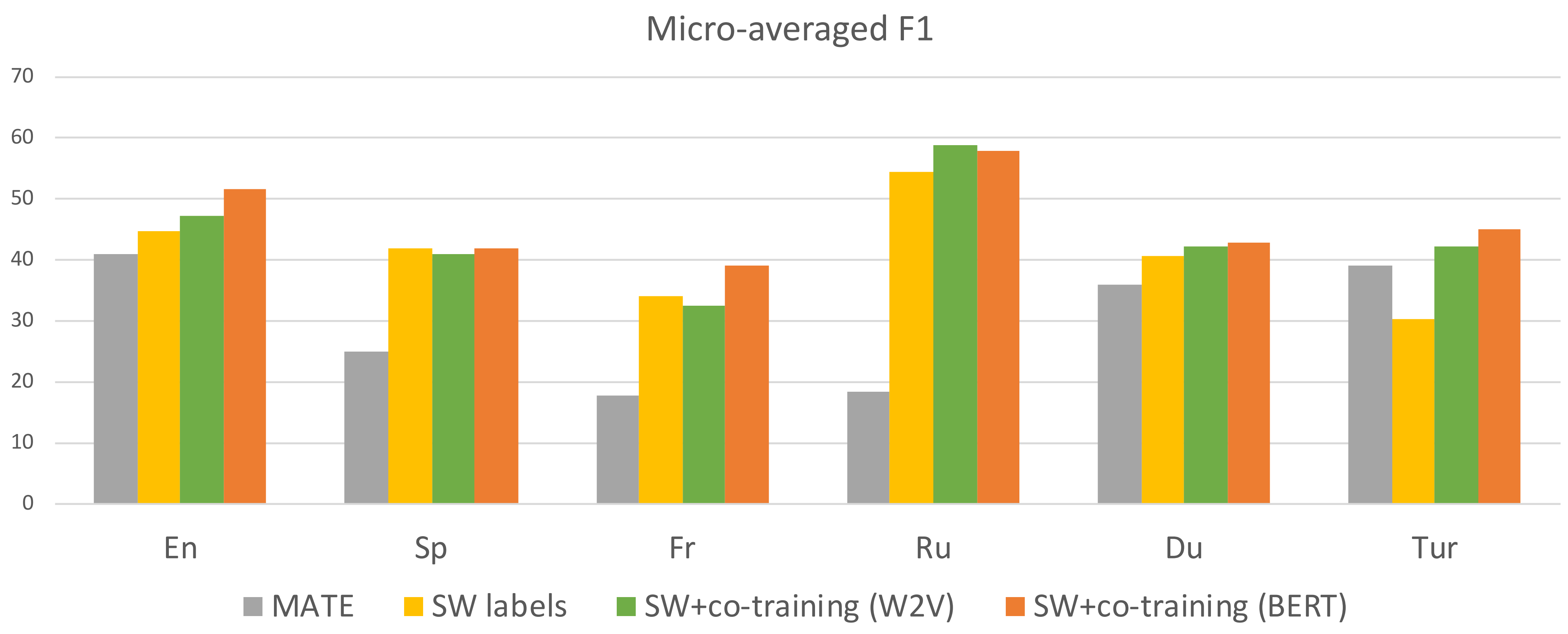
Setup

- **Training**:
  - 1M unlabeled review segments
  - 30 seed words per aspect obtained using method of Angelidis & Lapata (2018)
- **Evaluation**:
  - 750 labeled review segments
  - Performance metric: micro-averaged F1 [averaged over 5 runs]
- **Baselines**:
  - LDA-Anchors (Lund *et al*, 2017)
  - MATE: Multi-Seed Aspect Extractor (Angelidis & Lapata, 2018)
- **Multi-class classification models**:
  - Word2Vec embeddings from (Angelidis & Lapata, 2018; Ruder, Ghaffari, Breslin, 2016)
  - BERT embeddings (Devlin, Chang, Lee, Toutanova, 2019)
  - Linear model on top of embeddings; train all layers

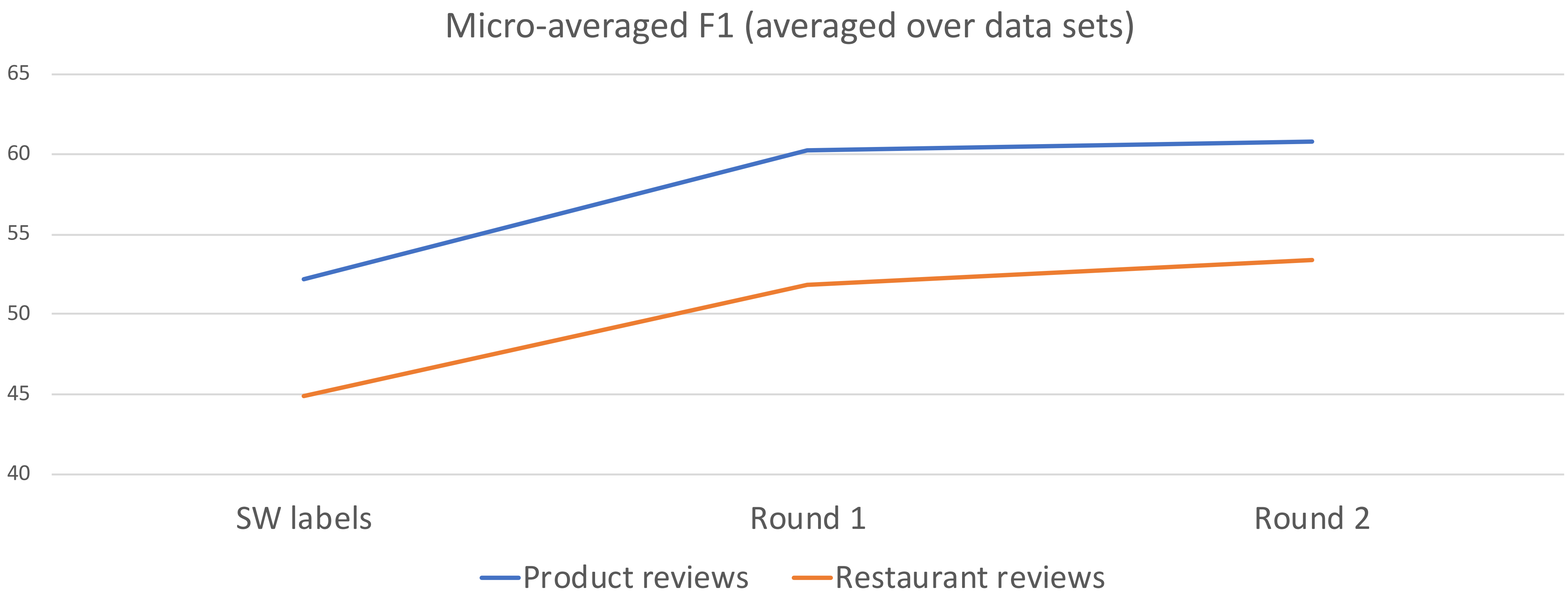
Results on product reviews



Results on restaurant reviews



Iterative co-training (BERT)



Summary

- Seed words highly useful as weak supervision
  - More effective use of seed words than as initialization for topic / embedding models
- Co-training framework allows one to leverage state-of-the-art models

Outline

1. Weak supervision via seed words
2. Interpretation as co-training
3. Empirical evaluation on product and restaurant reviews
4. Planned work on hidden bias detection

Media bias

AllSides™

- News media often comes with hard-to-detect bias
- Examples from AllSides.com:
  - Spin
  - Unsubstantiated claims
  - Opinion statements presented as fact
  - Sensationalism/emotionalism
  - …

Example

The Washington Post

As he jetted to Paris last Friday, President Trump received a congratulatory phone call aboard Air Force One. British Prime Minister Theresa May was calling to celebrate the Republican Party's wins in the midterm elections — never mind that Democrats seized control of the House — but her appeal to the American president's vanity was met with an ornery outburst.

Trump berated May for Britain not doing enough, in his assessment, to contain Iran. He questioned her over Brexit and complained about the trade deals he sees as unfair with European countries. May has endured Trump's churlish temper before, but still her aides were shaken by his especially foul mood, according to U.S. and European officials briefed on the conversation.

Potential for detection via seed words

- Many forms of bias can be detected through language
  - AllSides.com: **To stir emotions, reports often include colored, dramatic, or sensational words as a substitute for the word "said."**
  - E.g., mocked, raged, bragged, fumed, lashed out, incensed, scoffed, frustration, erupted, rant, boasted, gloated
- **Goal**: Learn to detect such forms of bias, leveraging other context information beyond known keywords

(Ongoing work)

Acknowledgements

- NSF IIS-15-63785, Yahoo FREP Award, Sloan Research Fellowship
- Becca Funke, Kapil Thadani for discussions about media bias

Thank you!