

# Weakly Supervised Neural Networks for Text Mining and Public Health

**Giannis Karamanolakis, Daniel Hsu, Luis Gravano**

Department of Computer Science, Columbia University

{gkaraman, djhsu, gravano}@cs.columbia.edu

*Presented at the AT&T Data Science Symposium (December 2019)*



# Outline

## 1. Fine-Grained Text Mining & Applications

## 2. Training Neural Networks with Weak Supervision

**“Weakly supervised attention networks for fine-grained opinion mining and public health”, Karamanolakis et al. (WNUT 2019)**

**“Training neural networks for aspect extraction using descriptive keywords only”, Karamanolakis et al. (LLD 2019)**

**“Leveraging just a few keywords for fine-grained aspect detection through weakly supervised co-training”, Karamanolakis et al. (EMNLP 2019)**

## 3. Conclusions and Future Work

# Online User-Generated Reviews



## 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

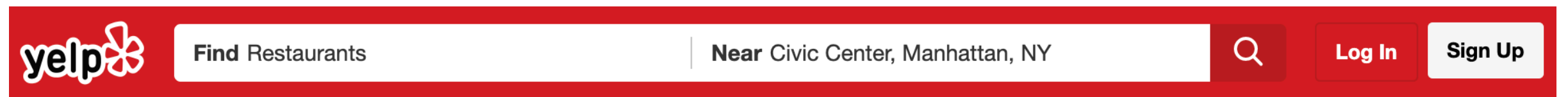
### ★★★★☆ A great TV for an amazing price

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 thereof). 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 is 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.



## Tavern on the Green

★★★★☆ 1377 reviews

\$\$\$ · [American \(New\)](#), [Cocktail Bars](#) [Edit](#)

[★ Write a Review](#) [Add Photo](#) [Share](#) [Save](#)



**Kerrie D.**  
Portland, ME  
281 friends  
560 reviews  
1108 photos  
[Elite '2019](#)

★★★★☆ 7/24/2018  
1 check-in

This is a review for Tavern on the Go, a convenient and could-be wonderful space to stop for a drink or snack while walking through Central Park.

So...right next to Tavern on the Green is a smaller building with a private garden and tables. There is a service window to order and pick up beverages and a few snacks. It is a beautiful and restful space, a great spot to unplug for a few minutes and enjoy the beauty around you. The prices are, as you might imagine, kind of nuts, BUT, your order gains you access to this sanctuary space, which makes the price point quite worth it.

Why two stars, you might be thinking?? The server working the window was cold, disinterested and gave off a VERY rude vibe! Not only to me, but to the group ordering before me which included a woman in a wheelchair! She gave the feeling that she was doing us all a BIG favor by taking our orders and the taking our money. There seemed to be a secondary server behind her in the window who was trying very hard to be customer focused. Her efforts were noticed, but she was completely over shadowed by her colleague working the front.

I would urge Tavern on the Green to check in with the servers working their "on the go" window, to make sure they are representing your brand of service in a way that you would expect!



**Stephanie G.**  
New York, NY  
80 friends  
58 reviews  
60 photos  
[Elite '2019](#)


★★★★☆ 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.


# Online User-Generated Reviews



Electronics ▾Televisions

Q

EN



### 32-Inch 720p LED TV

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

## TV Aspects


1. Price
2. Image Quality
3. Ease of Use
4. ...

★★★★☆ **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.



Find Restaurants

Near Civic Center, Manhattan, NY

Q

Log In

Sign Up

### Tavern on the Green

Claimed

★★★★☆ 1377 reviews

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

[Write a Review](#) [Add Photo](#) [Share](#) [Save](#)



## Restaurant Aspects

1. Food Quality
2. Ambience
3. Service
4. ...

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

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.

Elite '2019

[Share review](#) [Embed review](#)

Useful Funny Cool



# Opinion Mining Application: Aspect Detection



## SEGMENT-LEVEL ASPECT DETECTION



Verified Purchase

Great price for an excellent LED TV

### Sentence

### Aspect

Great Tv for the price. ....

Price

Easy to setup. ....

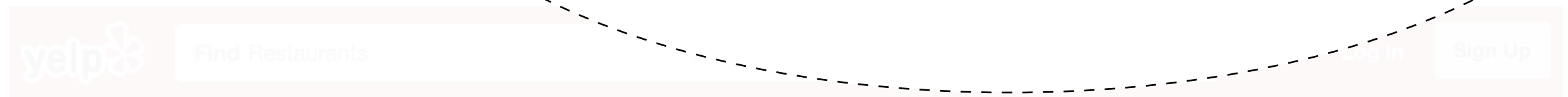
Ease of Use

The audio is ok for the tiny speakers. ....

Sound Quality

Much better than the 20" tube tv. ....

General



- **Motivation:** Segments (e.g., sentences) of a review may discuss different aspects
- **Task:** Classify individual segments to  $K$  pre-defined aspect classes



This is a review for Tavern on the Green, a convenient and could-be wonderful space to stop for a drink or snack while walking through Central Park.

So, didn't need to Tavern on the Green is a smaller building, but it's still a great place to stop for a drink or snack while walking through Central Park.

The prices are, as you might imagine, kind of high, BUT, your order gets you access to this sanctuary space, which is a great value.

VERY nice vibe. Not only to me, but to the group ordering before me which included a woman in a wheelchair. She gave the feeling that she was doing us all a BIG favor by taking our orders and the taking our money.

There seemed to be a secondary server behind her in the window who was trying very hard to be customer focused. Her efforts were noticed, but she was completely over shadowed by her colleague working the front.

I would urge Tavern on the Green to check in with the servers working their "on the go" window, to make sure they are representing your brand of service in a way that you would expect!



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 attentive. The food was good, but not amazing. I would have liked to see more variety in the menu.

My husband also ordered the salmon, and he got the squash blossom 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.

# Opinion Mining Application: Sentiment Analysis

## SEGMENT-LEVEL SENTIMENT ANALYSIS



★ ★ ★ ★ ★ 7/24/2018  
1 check-in

This is a review for Tavern on the Go, a convenient and could-be wonderful space to stop for a drink or snack while walking through Central Park.

So...right next to Tavern on the Green is a smaller building with a private garden and tables. There is a service window to order and pick up beverages and a few snacks. It is a beautiful and restful space, a great spot to unplug for a few minutes and enjoy the beauty around you. The prices are, as you might imagine, kind of nuts, BUT, your order gains you access to this sanctuary space, which makes the price point quite worth it.

Why two stars, you might be thinking?? The server working the window was cold, disinterested and gave off a VERY rude vibe! Not only to me, but to the group ordering before me which included a woman in a wheelchair! She gave the feeling that she was doing us all a BIG favor by taking our orders and the taking our money. There seemed to be a secondary server behind her in the window who was trying very hard to be customer focused. Her efforts were noticed, but she was completely over shadowed by her colleague working the front.

— positive  
— neutral  
— negative



Kerrie D.  
Portland, ME  
281 friends  
560 reviews  
1108 photos  
2019

★ ★ ★ ★ ★ 7/24/2018  
1 check-in

This is a review for Tavern on the Go, a convenient and could-be wonderful space to stop for a drink or snack while walking through Central Park.

So...right next to Tavern on the Green is a smaller building with a private garden and tables. There is a service window to order and pick up beverages and a few snacks. It is a beautiful and restful space, a great spot to unplug for a few minutes and enjoy the beauty around you. The prices are, as you might imagine, kind of nuts, BUT, your order gains you access to this sanctuary space, which makes the price point quite worth it.

Why two stars, you might be thinking?? The server working the window was cold, disinterested and gave off a VERY rude vibe! Not only to me, but to the group ordering before me which included a woman in a wheelchair! She gave the feeling that she was doing us all a BIG favor by taking our orders and the taking our money. There seemed to be a secondary server behind her in the window who was trying very hard to be customer focused. Her efforts were noticed, but she was completely over shadowed by her colleague working the front.


I would urge Tavern on the Green to check in with the servers working their "on the go" window, to make sure they are representing your brand of service in a way that you would expect!

esse J.  
rk, NJ  
7 friends  
7 reviews

★ ★ ★ ★ ★ 4/11/2017

Waited at the bar to be seated. Drink was very nice. Very strong delicious drink. People were all friendly. Our server Papa was amazing. Unfortunately I have been up half the night and suffering all day due to food poisoning. I'm assuming it was the shrimp. Its been a waterfall out of both ends and for the price I would expect better quality. Thus even making me late for school drop off and pick up today. My "medium rare" steak was too tough, more like medium well and the shrimp also was slightly over cooked. Both to the point I had to spit them out. Manager did take 50% off the steak. Great atmosphere. Just wish my bf and I weren't suffering.

# Public Health Application



Tavern on the Green

\$\$\$ · Italian, American (New York City)

★ Write a Review

## FOODBORNE ILLNESS DISCOVERY

Not Sick

Sick

Waited at the bar to be seated. Drink was very nice. Very strong delicious drink. People were all friendly. Our server Papa was amazing. Unfortunately I have been up half the night and suffering all day due to food poisoning. I'm assuming it was the shrimp. Its been a waterfall out of both ends and for the price I would expect better quality. Thus even making me late for school drop off and pick up today. My "medium rare" steak was too tough, more like medium well and the shrimp also was slightly over cooked. Both to the point I had to spit them out. Manager did take 50% off the steak. Great atmosphere. Just wish my bf and I weren't suffering.

7/24/2018

1 check-in

This is a review for Tavern on the Go, a convenient and could-be wonderful space to stop for a drink or snack while walking through Central Park.

So...right next to Tavern on the Green is a smaller building with a private garden and tables. There is a service window to order and pick up beverages and a few snacks. It is a beautiful and restful space, a great spot to unplug a few minutes and enjoy the beauty around you. Prices are, as you might imagine, kind of nuts, BUT, order gains you access to this sanctuary space, which is the price point quite worth it.

Two stars, you might be thinking?? The server by the window was cold, disinterested and gave off a rude vibe! Not only to me, but to the group ordering with me which included a woman in a wheelchair! She was feeling that she was doing us all a BIG favor by taking our orders and the taking our money. The server seemed to be a secondary server behind her in the kitchen who was trying very hard to be customer focused. Her efforts were noticed, but she was completely overwhelmed by her colleague working the front.

I would urge Tavern on the Green to check in with the servers working their "on the go" window, to make sure they are representing your brand of service in a way that you would expect!

4/11/2017

Waited at the bar to be seated. Drink was very nice. Very strong delicious drink. People were all friendly. Our server Papa was amazing. Unfortunately I have been up half the night and suffering all day due to food poisoning. I'm assuming it was the shrimp. Its been a waterfall out of both ends and for the price I would expect better quality. Thus even making me late for school drop off and pick up today. My "medium rare" steak was too tough, more like medium well and the shrimp also was slightly over cooked. Both to the point I had to spit them out. Manager did take 50% off the steak. Great atmosphere. Just wish my bf and I weren't suffering.

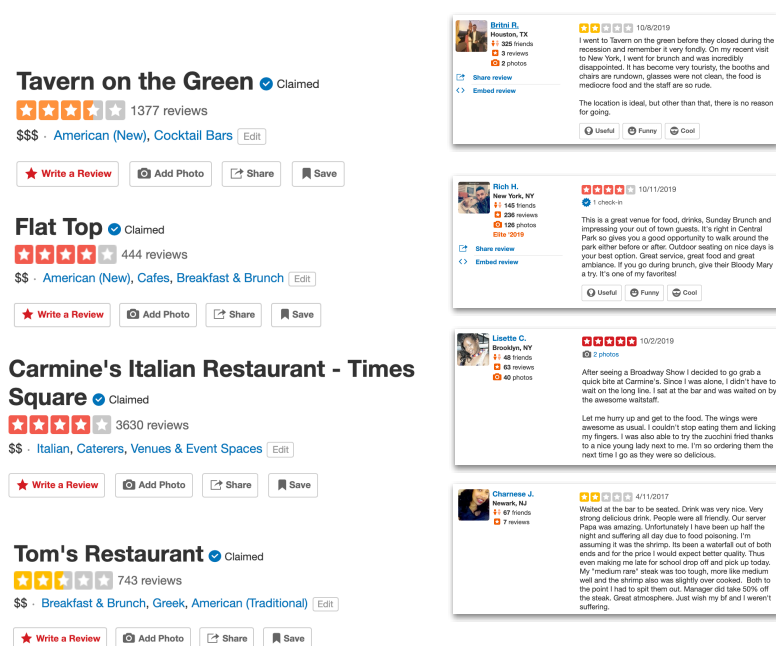


# Public Health Application

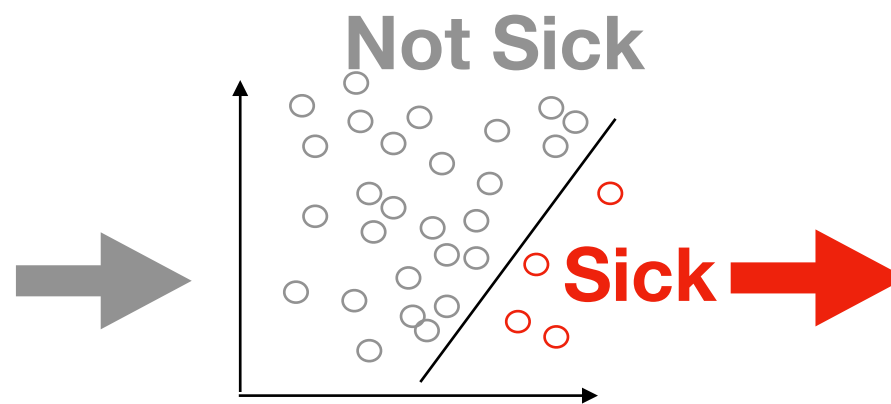
## Detecting foodborne illness from Yelp restaurant reviews



### NYC Restaurants



### ML Classifier



### Epidemiologists



I wish I could give it zero stars. I actually created a yelp account to write this review! At first I thought it was great that we got a table for 5 morning of on a Saturday. The food was okay- the poached eggs on the Benedict were a little over cooked, but nothing to complain about. The service was good, it was overall fine. That is- until I got home and me and boy friend spent the rest of the day/night and into the morning hunched over or sitting on the toilet! I have never experienced such violent food poisoning in my life! That was the only place we ate or drank anything at that day, so I know it was from this restaurant. By far the most miserable I've been- chills and crippling abdominal pain along with uncontrollable vomiting and something worse out the other end for my boyfriend! Whatever you do, do not eat here, it is not worth the risk of ending up so unwell. To clarify what I believe caused this- we both had carrot juice randomly. I know more than one person who has gotten food poisoning recently from carrot juice- especially if its raw or cold pressed.



# Public Health Application

Detecting foodborne illness from Yelp restaurant reviews



Restaurant Inspectors



Epidemiologists



NYC Restaurants

ML Classifier

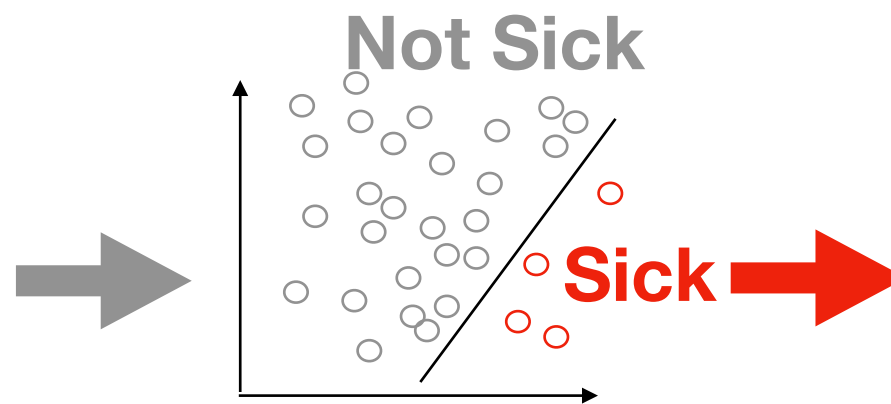
**Tavern on the Green** Claimed  
1377 reviews  
\$\$\$ · American (New), Cocktail Bars

**Flat Top** Claimed  
444 reviews  
\$\$ · American (New), Cafes, Breakfast & Brunch

**Carmine's Italian Restaurant - Times Square** Claimed  
3630 reviews  
\$\$ · Italian, Caterers, Venues & Event Spaces

**Tom's Restaurant** Claimed  
743 reviews  
\$\$ · Breakfast & Brunch, Greek, American (Traditional)

Four sample Yelp reviews are shown, detailing various dining experiences and complaints.



I wish I could give it zero stars. I actually created a yelp account to write this review! At first I thought it was great that we got a table for 5 morning of on a Saturday. The food was okay- the poached eggs on the Benedict were a little over cooked, but nothing to complain about. The service was good, it was overall fine. That is- until I got home and me and boy friend spent the rest of the day/night and into the morning hunched over or sitting on the toilet! I have never experienced such violent food poisoning in my life! That was the only place we ate or drank anything at that day, so I know it was from this restaurant. By far the most miserable I've been- chills and crippling abdominal pain along with uncontrollable vomiting and something worse out the other end for my boyfriend! Whatever you do, do not eat here, it is not worth the risk of ending up so unwell. To clarify what I believe caused this- we both had carrot juice randomly. I know more than one person who has gotten food poisoning recently from carrot juice- especially if its raw or cold pressed.

# Fine-Grained Text Classification

- **Motivation:** Segments of a review may have different labels

## ASPECT DETECTION

Labels: K aspect classes

<u>Sentence</u>	<u>Aspect</u>
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

## SENTIMENT CLASSIFICATION

Labels:  positive  
 neutral  
 negative

So...right next to Tavern on the Green is a smaller building with a private garden and tables. There is a service window to order and pick up beverages and a few snacks. It is a beautiful and restful space, a great spot to unplug for a few minutes and enjoy the beauty around you. The prices are, as you might imagine, kind of nuts, BUT, your order gains you access to this sanctuary space, which makes the price point quite worth it.

## FOODBORNE CLASSIFICATION

Labels:  Not Sick  
 Sick

Waited at the bar to be seated. Drink was very nice. Very strong delicious drink. People were all friendly. Our server Papa was amazing. Unfortunately I have been up half the night and suffering all day due to food poisoning. I'm assuming it was the shrimp. Its been a waterfall out of both ends and for the price I would expect better quality. Thus even making me late for school drop off and pick up today.

# Fine-Grained Text Classification

- **Motivation:** Segments of a review may have different labels
- **Goal:** provide **segment**-level predictions

# Fine-Grained Text Classification

- **Motivation:** Segments of a review may have different labels
- **Goal:** provide **segment**-level predictions
- **Issue:** ground truth segment labels are **expensive** to obtain



# Fine-Grained Text Classification

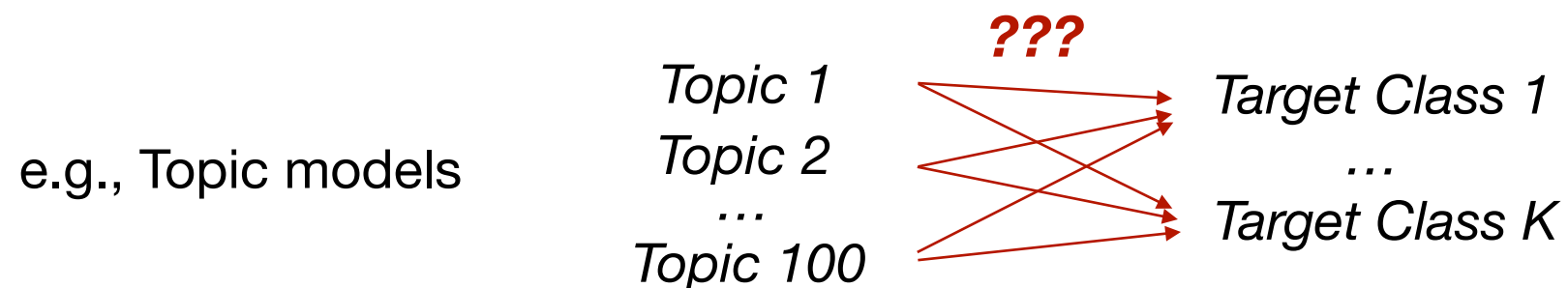
- **Motivation:** Segments of a review may have different labels
- **Goal:** provide **segment**-level predictions
- **Issue:** ground truth segment labels are **expensive** to obtain
  - **Supervised learning?** (-) require (many) segment labels



- >1B products
- >10K categories
- New products added every day

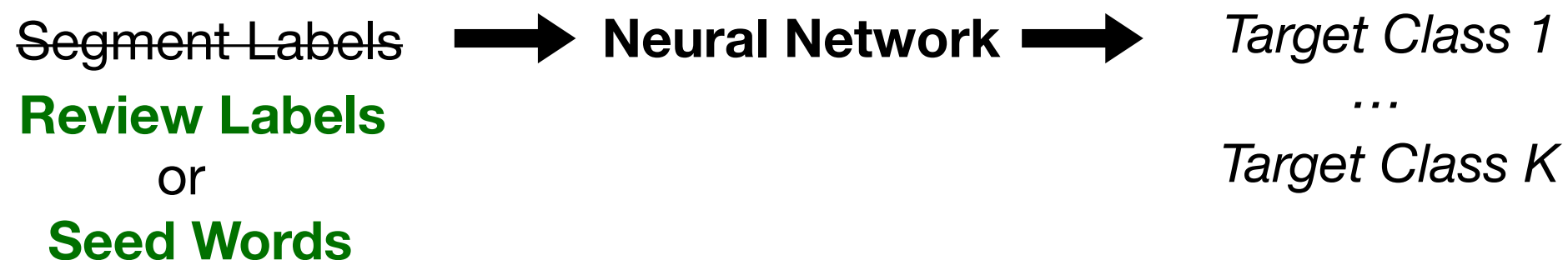
# Fine-Grained Text Classification

- **Motivation:** Segments of a review may have different labels
- **Goal:** provide **segment**-level predictions
- **Issue:** ground truth segment labels are **expensive** to obtain
  - **Supervised learning?** ~~(-) require (many) segment labels~~
  - **Unsupervised learning?** (-) may not capture target classes



# Fine-Grained Text Classification

- **Motivation:** Segments of a review may have different labels
- **Goal:** provide **segment**-level predictions
- **Issue:** ground truth segment labels are **expensive** to obtain
  - **Supervised learning?** ~~————— (-) require (many) segment labels~~
  - **Unsupervised learning?** ~~————— (-) may not capture target classes~~
  - **Weakly supervised learning** (+) leverage cheaper supervision sources



# Outline

## 1. Fine-Grained Text Mining & Applications

## 2. Training Neural Networks with Weak Supervision

**“Weakly supervised attention networks for fine-grained opinion mining and public health”, Karamanolakis et al. (WNUT 2019)**

**“Training neural networks for aspect extraction using descriptive keywords only”, Karamanolakis et al. (LLD 2019)**

**“Leveraging just a few keywords for fine-grained aspect detection through weakly supervised co-training”, Karamanolakis et al. (EMNLP 2019)**

## 3. Conclusions and Future Work



# Weakly Supervised Text Classification

## Segment Classification Without Ground Truth Segment Labels

- We leverage cheaper supervision sources:

1. Review labels: (+) already available

### SENTIMENT CLASSIFICATION

User rating



### FOODBORNE CLASSIFICATION

Epidemiologist diagnosis



# Weakly Supervised Text Classification

## Segment Classification Without Ground Truth Segment Labels

- We leverage cheaper supervision sources:

1. Review labels: (+) already available

### SENTIMENT CLASSIFICATION

User rating



### FOODBORNE CLASSIFICATION

Epidemiologist diagnosis



or

2. Seed words: (+) easier to obtain manually or automatically

### ASPECT DETECTION

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

# Outline

## 1. Fine-Grained Text Mining & Applications

## 2. Training Neural Networks with Weak Supervision

- Using Review Labels

**“Weakly supervised attention networks for fine-grained opinion mining and public health”**,  
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (WNUT 2019)

- Using Seed Words

**“Training neural networks for aspect extraction using descriptive keywords only”**,  
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano (LLD 2019)

**“Leveraging just a few keywords for fine-grained aspect detection through weakly supervised co-training”**, Giannis Karamanolakis, Daniel Hsu, and Luis Gravano (EMNLP 2019)

## 3. Conclusions and Future Work

# Outline

## 1. Fine-Grained Text Mining & Applications

## 2. Training Neural Networks with Weak Supervision

### ▸ Using Review Labels

1. **“Weakly supervised attention networks for fine-grained opinion mining and public health”**, Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (WNUT 2019)

### ▸ Using Seed Words

1. **“Training neural networks for aspect extraction using descriptive keywords only”**, Giannis Karamanolakis, Daniel Hsu, and Luis Gravano (LLD 2019)
2. **“Leveraging just a few keywords for fine-grained aspect detection through weakly supervised co-training”**, Giannis Karamanolakis, Daniel Hsu, and Luis Gravano (EMNLP 2019)

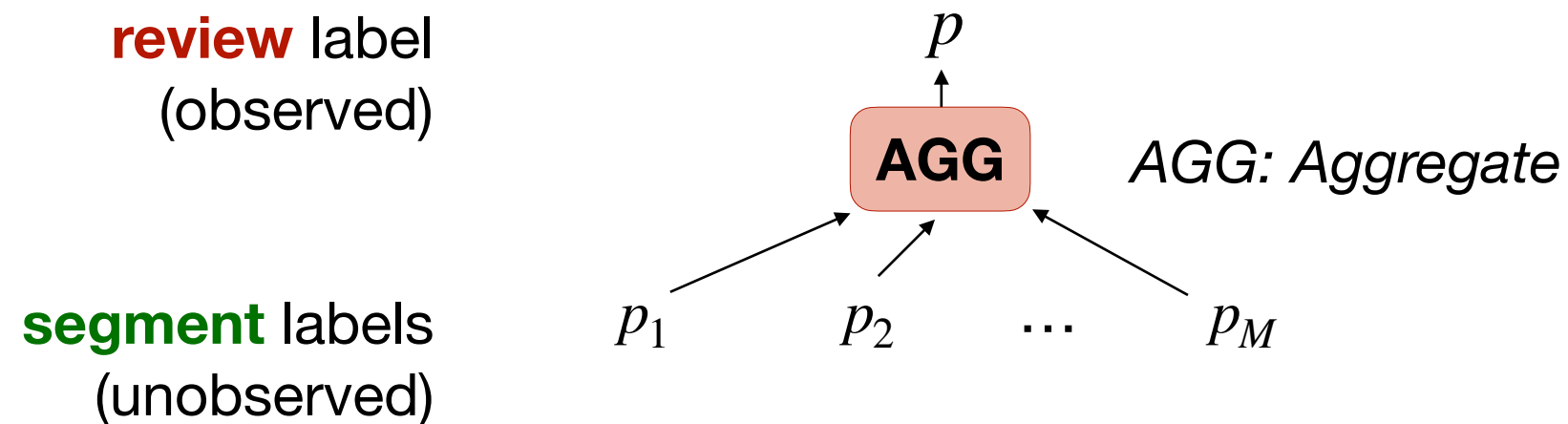
## 3. Conclusions and Future Work



# How to Train **Segment** Classifiers Using **Review** Labels?

- **Previous work:** Multiple Instance Learning (MIL)

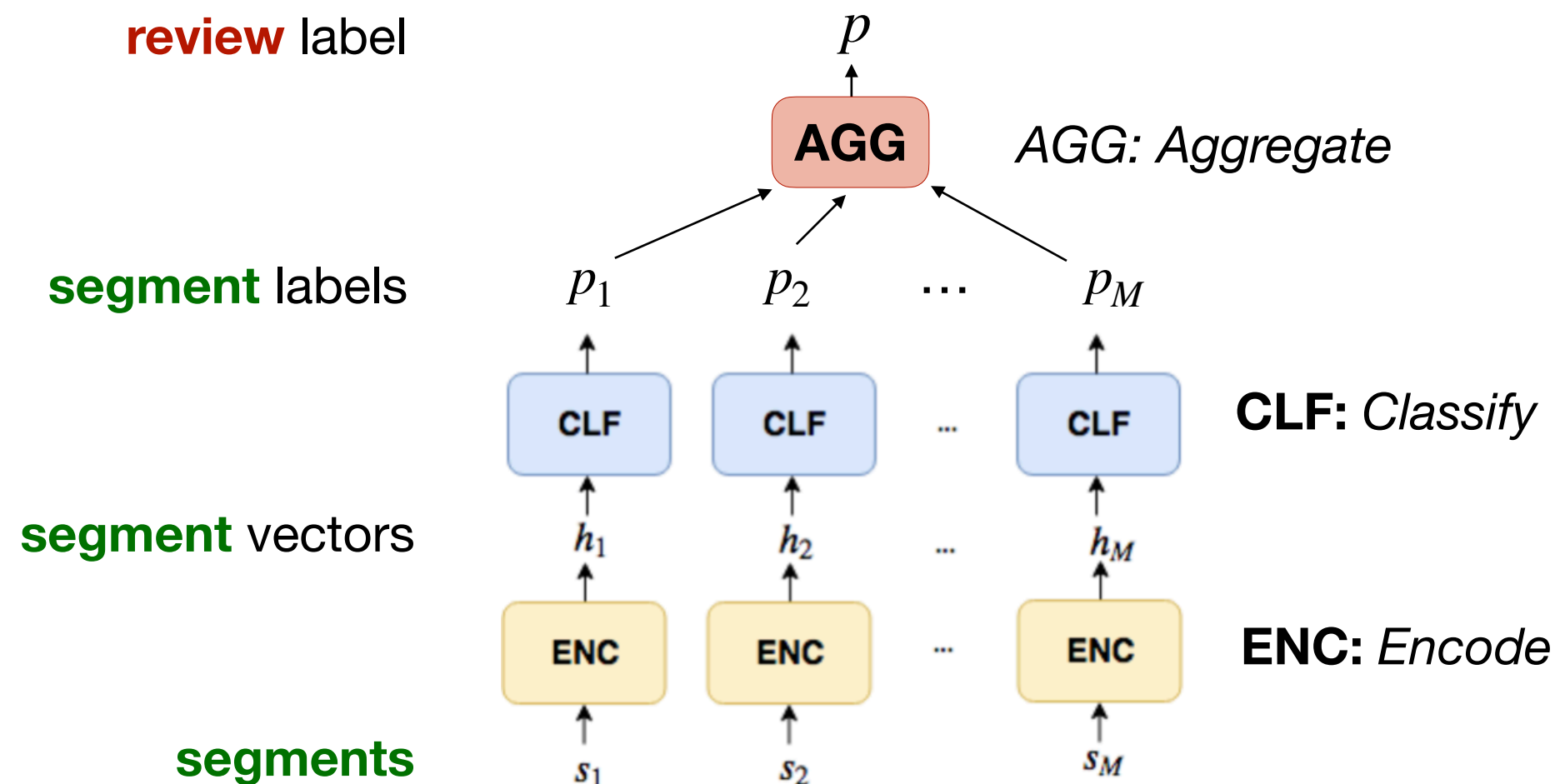
- Bags: reviews
- Instances: segments



# How to Train **Segment** Classifiers Using **Review** Labels?

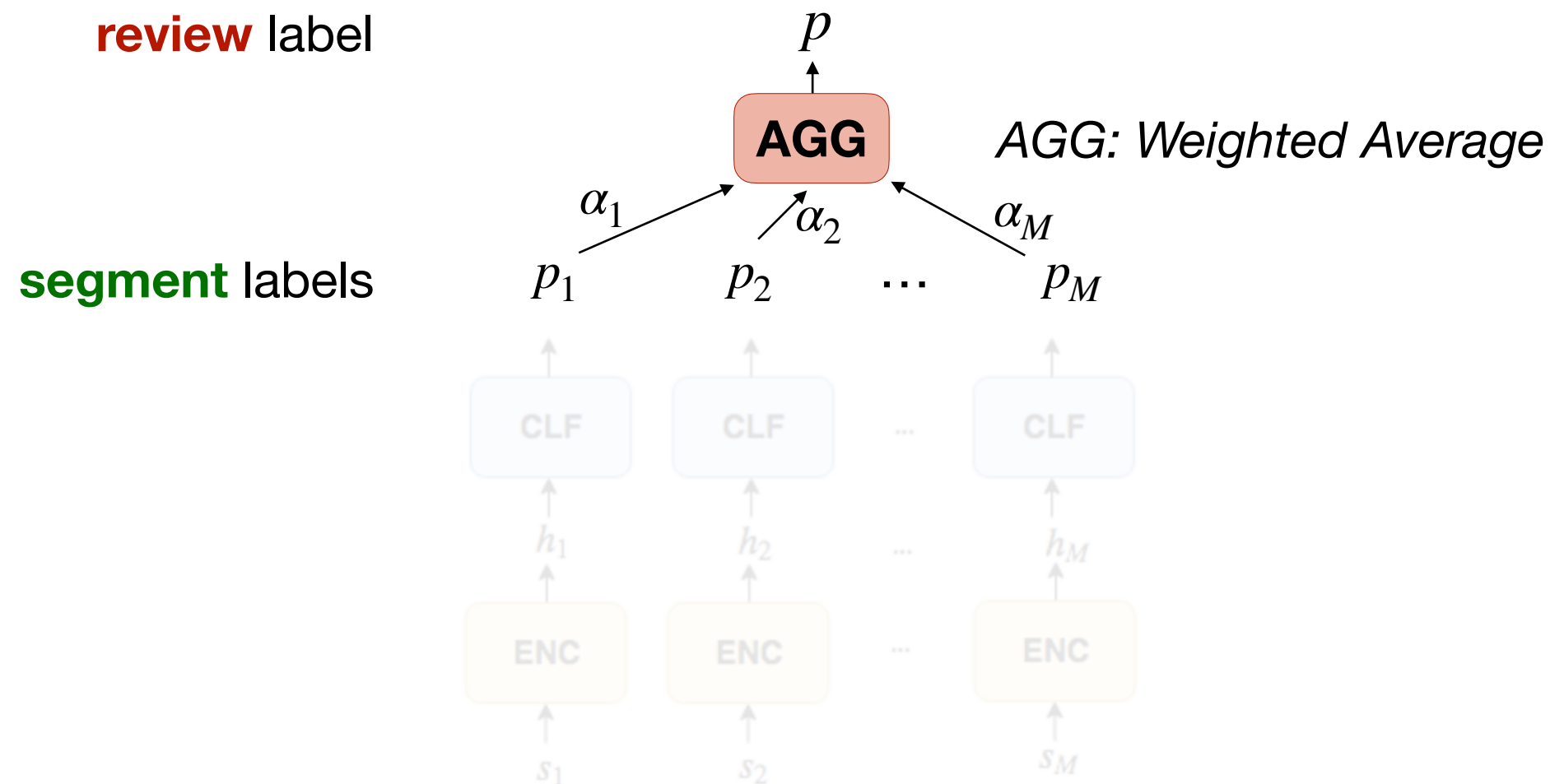
- **Previous work:** Multiple Instance Learning (MIL)

- State-of-the-art approaches: hierarchical MIL networks [Pappas and Popescu-belis, 2014;2017]  
[Angelidis and Lapata, 2018]



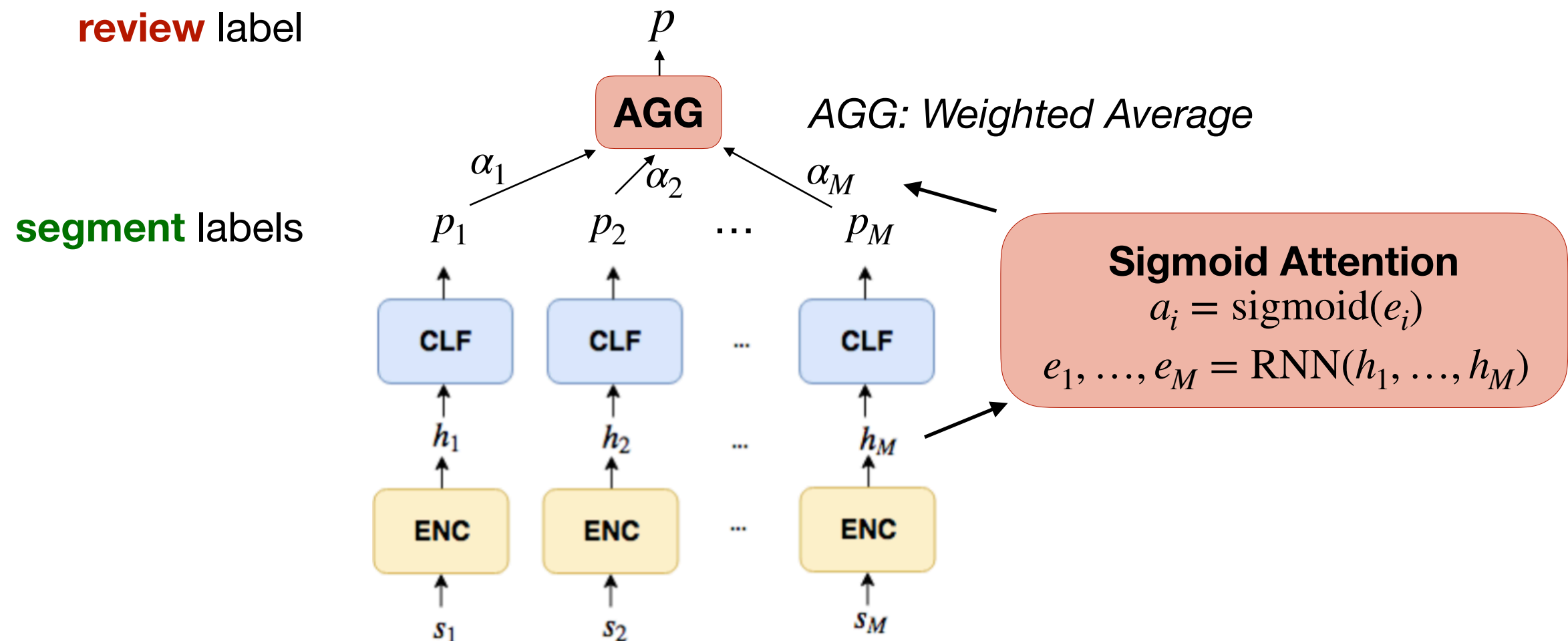
# We Improve MIL Networks by Changing the AGG Function

- **This work:** We change AGG in MIL networks to capture label hierarchy:
  - **Multiple** segments may contribute to review label with **different** weights  $\alpha_1, \dots, \alpha_M$



# We Improve MIL Networks by Changing the AGG Function

- **This work:** We change AGG in MIL networks to capture label hierarchy:
  - **Multiple** segments may contribute to review label with **different** weights  $\alpha_1, \dots, \alpha_M$
  - We compute segment weights using **sigmoid** (instead of softmax) attention



# Sigmoid Attention Networks Outperform Previous Networks

## Sentiment Classification - Yelp

Model	F1
Non-Hierarchical Baseline	61.5
MIL : AGG=uniform average (Kotzias et al., 2015)	46.8 <b>(-14.8)</b>
MIL : AGG=softmax attention (Angelidis and Lapata, 2018)	59.9 <b>(-1.6)</b>
MIL : AGG=sigmoid attention (our model)	<b>63.3 (+1.8)</b>

# Sigmoid Attention Networks Outperform Previous Networks

## Sentiment Classification - Yelp

Model	F1
Non-Hierarchical Baseline	61.5
MIL : AGG=uniform average (Kotzias et al., 2015)	46.8 <b>(-14.8)</b>
MIL : AGG=softmax attention (Angelidis and Lapata, 2018)	59.9 <b>(-1.6)</b>
MIL : AGG=sigmoid attention (our model)	<b>63.3 (+1.8)</b>

- Sigmoid attention is crucial for MIL networks to outperform our simple baseline

*More experiments (including IMDB review classification) in our paper.*

# We Demonstrate Advantages of Our Model for Public Health

- Apply MIL networks for **foodborne illness detection** in Yelp restaurant reviews
  - We achieve **48.6% higher recall** than best model in Effland et al. 2018



# We Demonstrate Advantages of Our Model for Public Health

- Apply MIL networks for **foodborne illness detection** in Yelp restaurant reviews
  - We achieve **48.6% higher recall** than best model in Effland et al. 2018
  - We highlight importance sentences

**I have never experienced such violent food poisoning in my life**

**By far the most miserable I've been- chills and crippling abdominal pain along with uncontrollable vomiting ...**

Pred	Att	Text
✓	0.00	I wish I could give it zero stars. 🤢 : Sick ✓ : Not Sick
✓	0.00	I actually created a yelp account to write this review!
✓	0.00	At first I thought it was great that we got a table for 5 morning of on a Saturday.
✓	0.00	The food was okay- the poached eggs on the Benedict were a little over cooked, but nothing to complain about.
✓	0.00	The service was good, it was overall fine.
✓	0.00	That is- until I got home and me and boy friend spent the rest of the day/night and into the morning hunched over or sitting on the toilet!
🤢	0.18	<b>I have never experienced such violent food poisoning in my life!</b>
✓	0.00	That was the only place we ate or drank anything at that day, so I know it was from this restaurant.
🤢	0.82	<b>By far the most miserable I've been- chills and crippling abdominal pain along with uncontrollable vomiting and something worse out the other end for my boyfriend!</b>
🤢	0.00	Whatever you do, do not eat here, it is not worth the risk of ending up so unwell.
✓	0.00	To clarify what I believe caused this- we both had carrot juice randomly.
🤢	0.00	I know more than one person who has gotten food poisoning recently from carrot juice- especially if its raw or cold pressed.

# We Demonstrate Advantages of Our Model for Public Health

- Apply MIL networks for **foodborne illness detection** in Yelp restaurant reviews
  - ▶ We achieve **48.6% higher recall** than best model in Effland et al. 2018
  - ▶ We highlight importance sentences
  - ▶ Fine-grained predictions could help foodborne inspections in health departments

**I have never experienced such violent food poisoning in my life**

**By far the most miserable I've been- chills and crippling abdominal pain along with uncontrollable vomiting ...**

Pred	Att	Text
✓	0.00	I wish I could give it zero stars. 🤢 : Sick ✓ : Not Sick
✓	0.00	I actually created a yelp account to write this review!
✓	0.00	At first I thought it was great that we got a table for 5 morning of on a Saturday.
✓	0.00	The food was okay- the poached eggs on the Benedict were a little over cooked, but nothing to complain about.
✓	0.00	The service was good, it was overall fine.
✓	0.00	That is- until I got home and me and boy friend spent the rest of the day/night and into the morning hunched over or sitting on the toilet!
🤢	0.18	<b>I have never experienced such violent food poisoning in my life!</b>
✓	0.00	That was the only place we ate or drank anything at that day, so I know it was from this restaurant.
🤢	0.82	<b>By far the most miserable I've been- chills and crippling abdominal pain along with uncontrollable vomiting and something worse out the other end for my boyfriend!</b>
🤢	0.00	Whatever you do, do not eat here, it is not worth the risk of ending up so unwell.
✓	0.00	To clarify what I believe caused this- we both had carrot juice randomly.
🤢	0.00	I know more than one person who has gotten food poisoning recently from carrot juice- especially if its raw or cold pressed.

# Outline

## 1. Fine-Grained Text Mining & Applications

## 2. Training Neural Networks with Weak Supervision

- Using Review Labels

1. **“Weakly supervised attention networks for fine-grained opinion mining and public health”**,  
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (WNUT 2019)

- **Using Seed Words**

1. **“Training neural networks for aspect extraction using descriptive keywords only”**,  
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano (LLD 2019)
2. **“Leveraging just a few keywords for fine-grained aspect detection through weakly supervised co-training”**, Giannis Karamanolakis, Daniel Hsu, and Luis Gravano (EMNLP 2019)

## 3. Conclusions and Future Work

# Segment-Level Aspect Detection with Seed Words

- **Goal:** Classify individual segments to  $K$  pre-defined aspect classes



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**

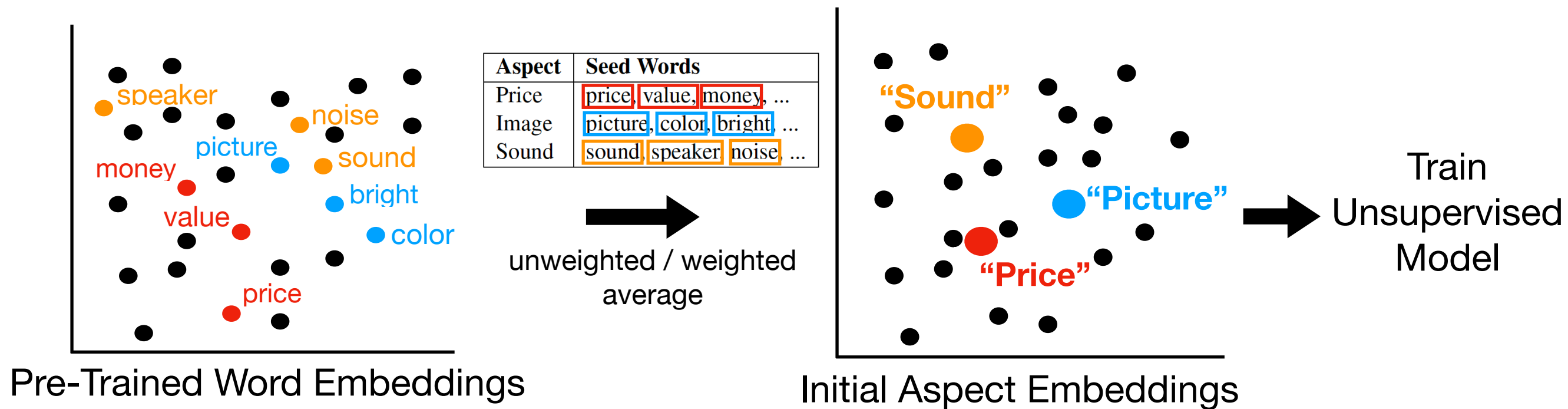
- **Issue:** Ground truth aspect labels may not be available (even at review-level)
- **This work:** We only assume a small set of indicative **seed words** per aspect

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

# How to Leverage Seed Words in Neural Networks?

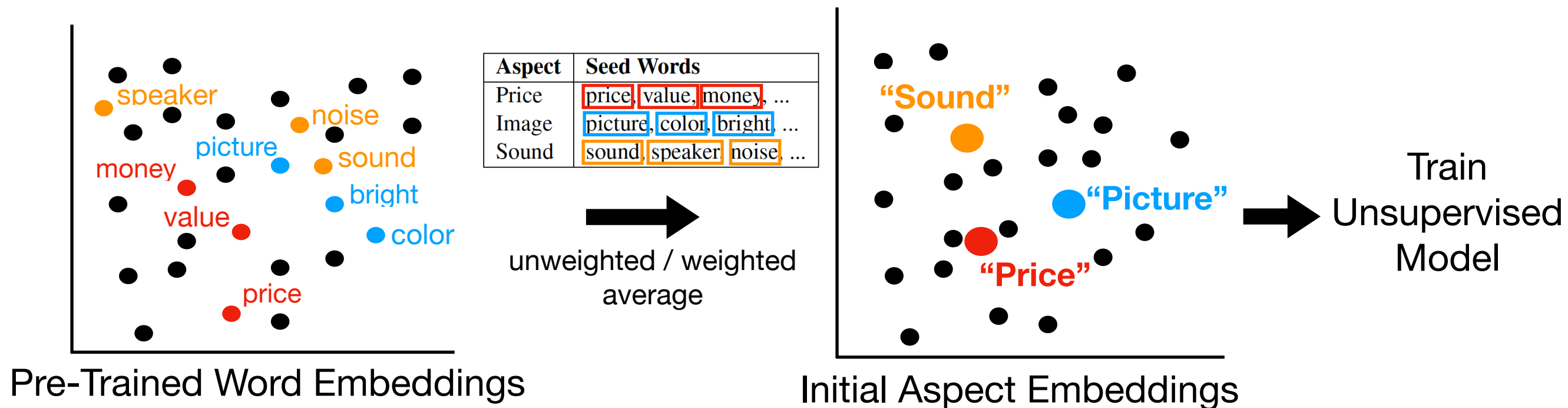
# How to Leverage Seed Words in Neural Networks?

- **Previous approaches:** only use seed words for initialization [Lund et al., 2017]  
[Angelidis and Lapata, 2018b]



# How to Leverage Seed Words in Neural Networks?

- **Previous approaches:** only use seed words for initialization [Lund et al., 2017]  
[Angelidis and Lapata, 2018b]



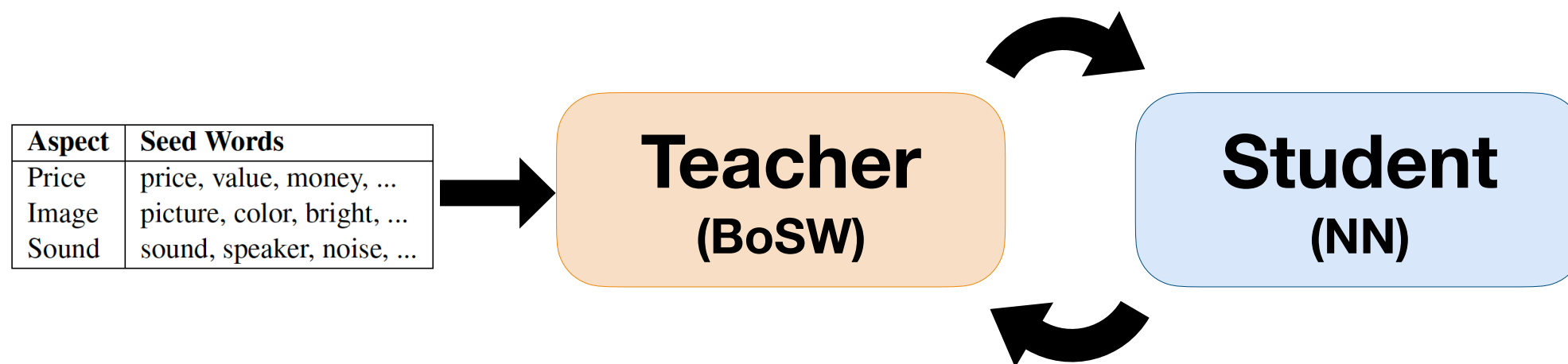
- **Our approach:** Weakly Supervised Co-Training

- + We exploit the **predictive** power ...
- + of each individual seed word ...
- + during training



# Weakly Supervised Co-Training

1. Teacher
2. Student
3. Teacher -> Student
4. Student -> Teacher



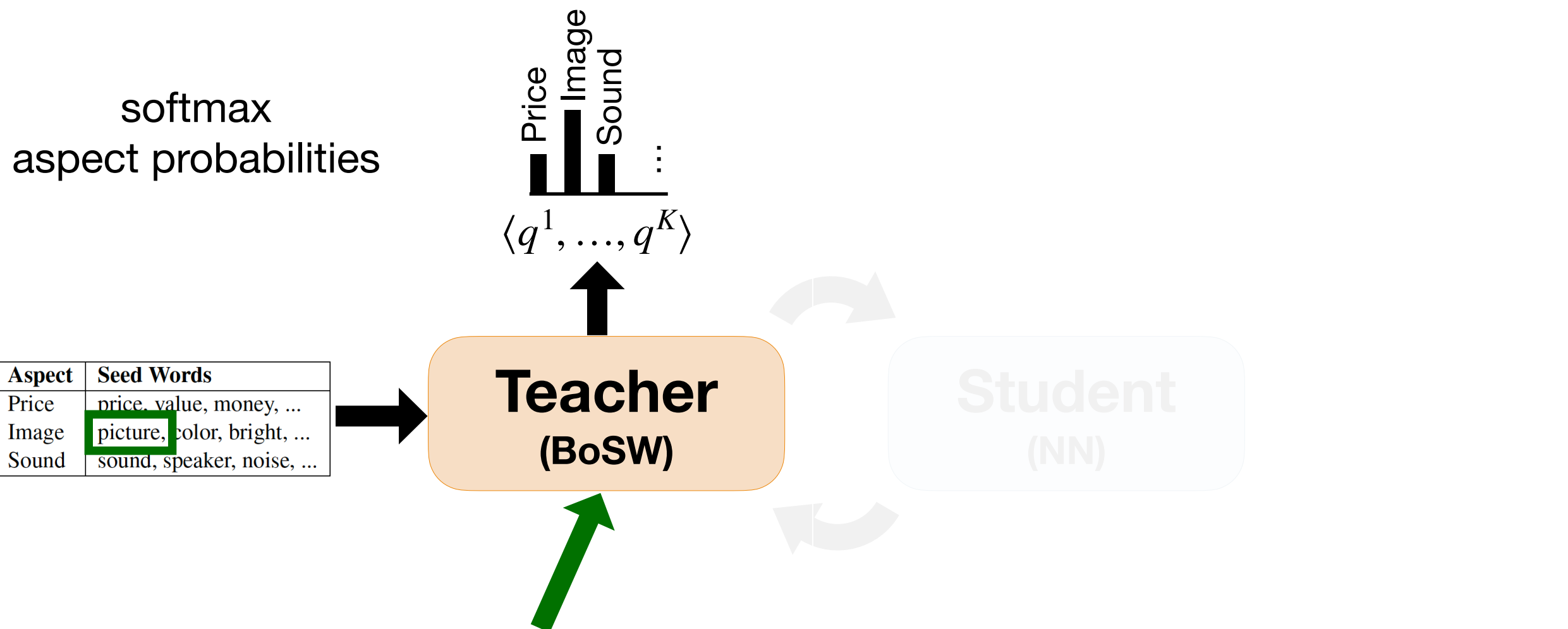
# Weakly Supervised Co-Training

1. **Teacher:** a bag-of-seed-words (BoSW) classifier

2. Student

3. Teacher  $\rightarrow$  Student

4. Student  $\rightarrow$  Teacher



segment: “The **picture** looks very pixelated when playing blu-ray movies”

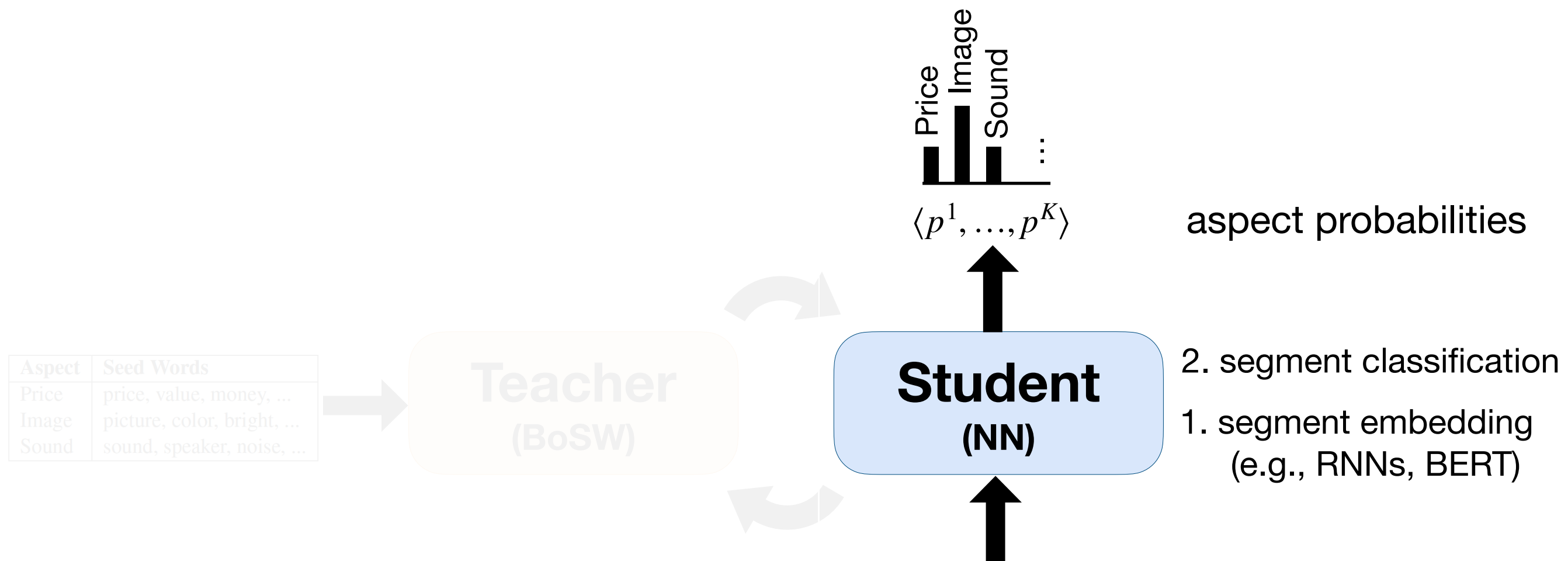
# Weakly Supervised Co-Training

1. Teacher

**2. Student:** an embedding-based neural network (NN)

3. Teacher  $\rightarrow$  Student

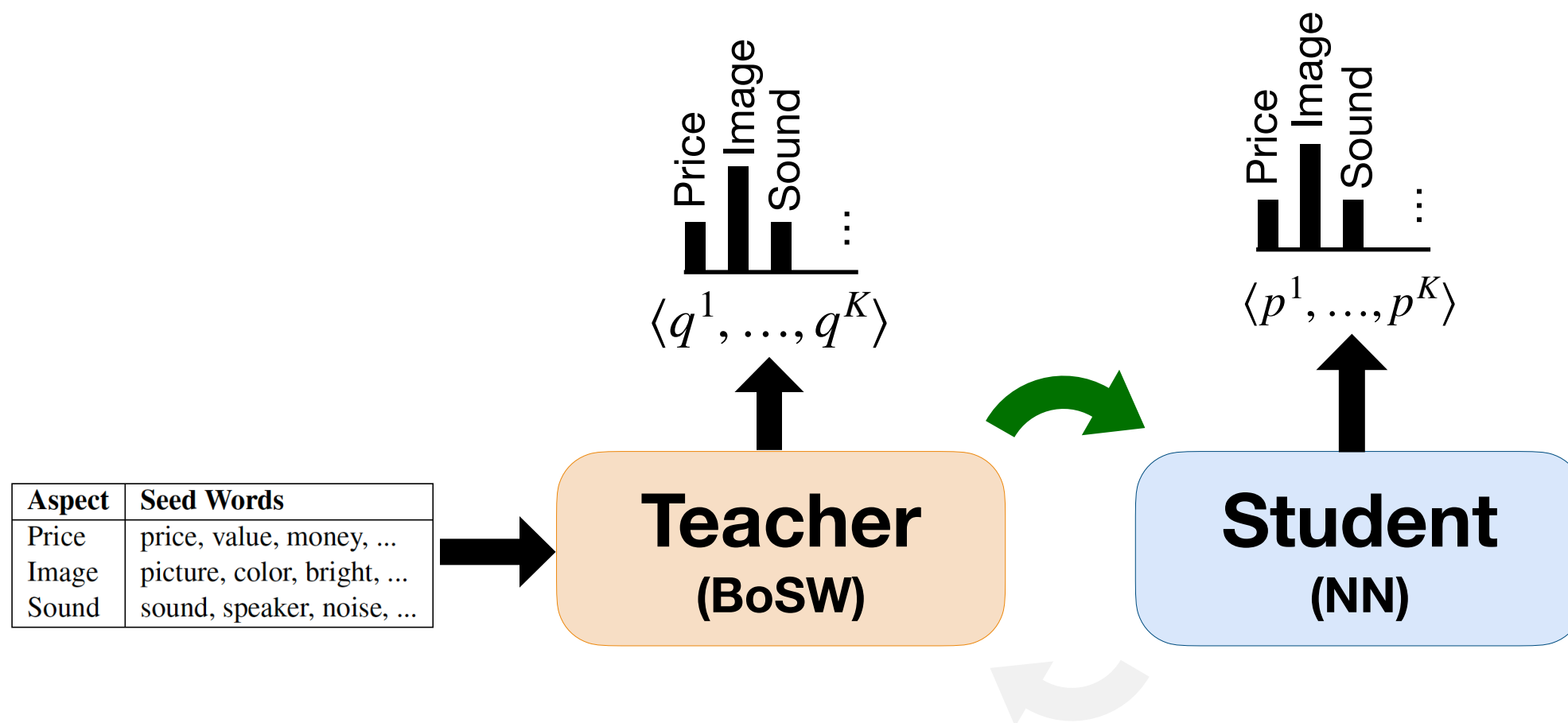
4. Student  $\rightarrow$  Teacher



segment: “*The picture looks very pixelated when playing blu-ray movies*”

# Weakly Supervised Co-Training

1. Teacher
2. Student
- 3. Teacher -> Student:** predict Teacher's aspect probabilities
4. Student -> Teacher

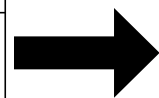


# Weakly Supervised Co-Training

1. Teacher
2. Student
- 3. Teacher -> Student**
4. Student -> Teacher

uses seed words only

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



**Teacher**  
(BoSW)

**Student**  
(NN)



segment: “The **picture** looks very pixelated when playing blu-ray movies”

# Weakly Supervised Co-Training

1. Teacher
2. Student
- 3. Teacher -> Student**
4. Student -> Teacher

uses seed words only

uses seed words  
and context!

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

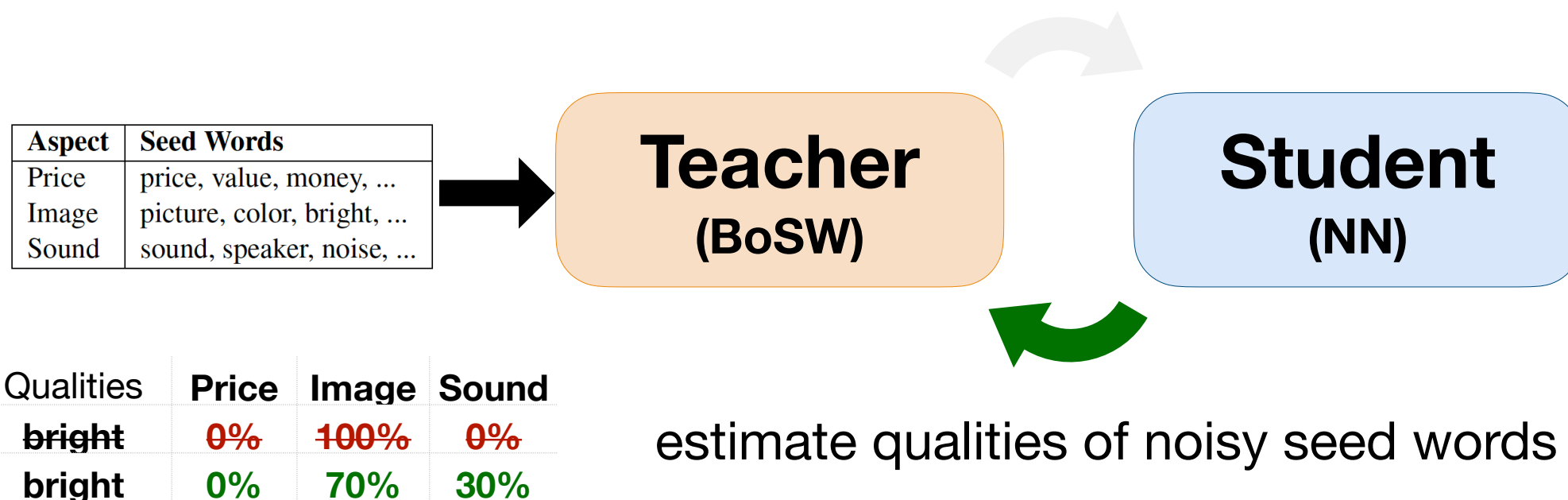
**Teacher**  
(BoSW)

**Student**  
(NN)

segment: “*The **picture** looks very pixelated when playing blu-ray movies*”

# Weakly Supervised Co-Training

1. Teacher
2. Student
3. Teacher -> Student
- 4. Student -> Teacher**



# Weakly Supervised Co-Training

1. Teacher

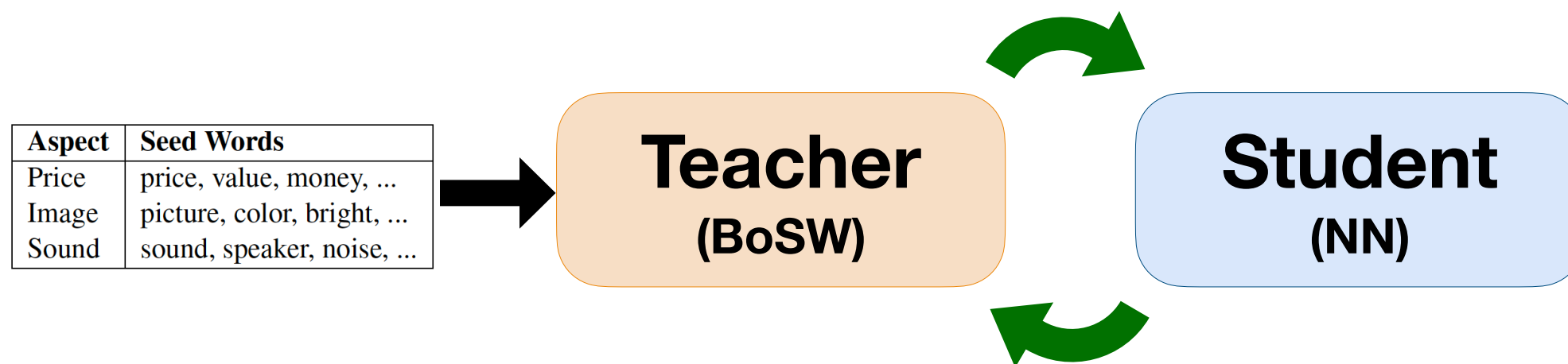
2. Student

**3. Teacher -> Student**

**4. Student -> Teacher**

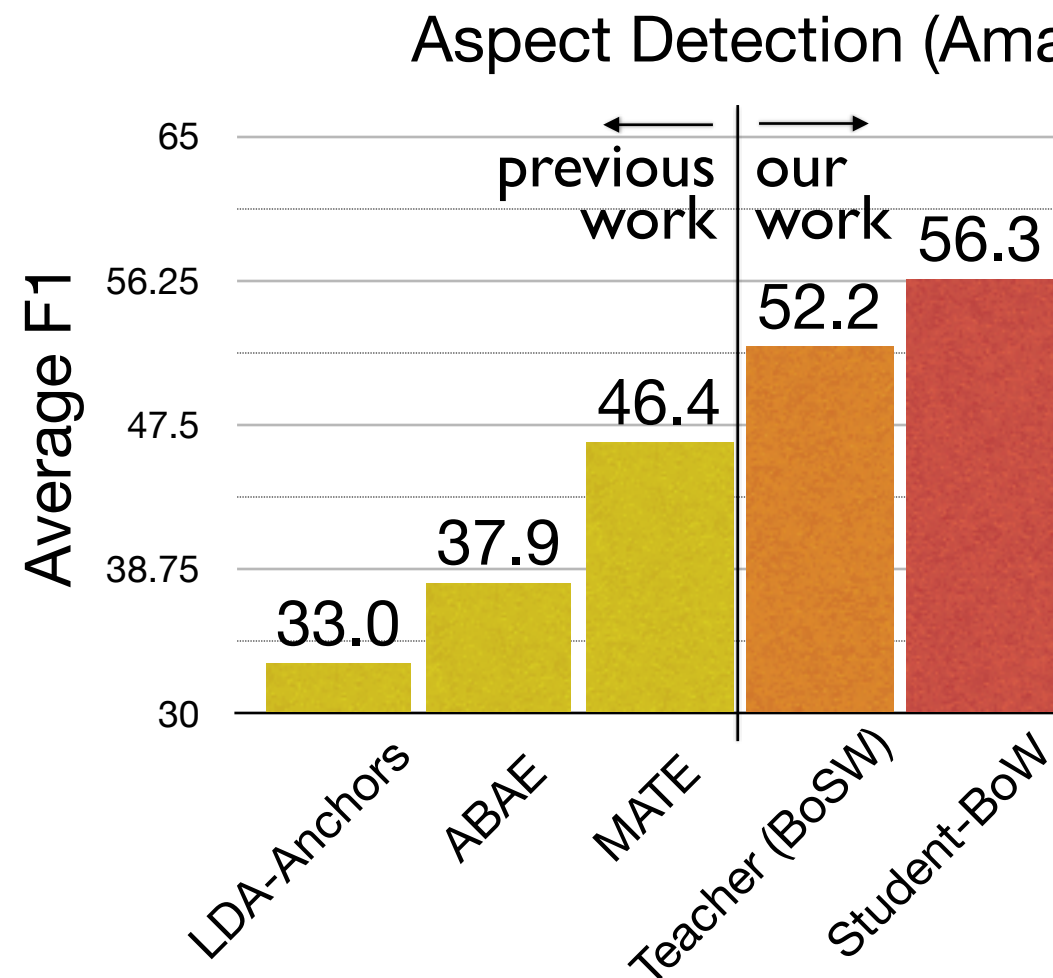
**Iterative Co-Training**

iterate until convergence (teacher-student agreement)



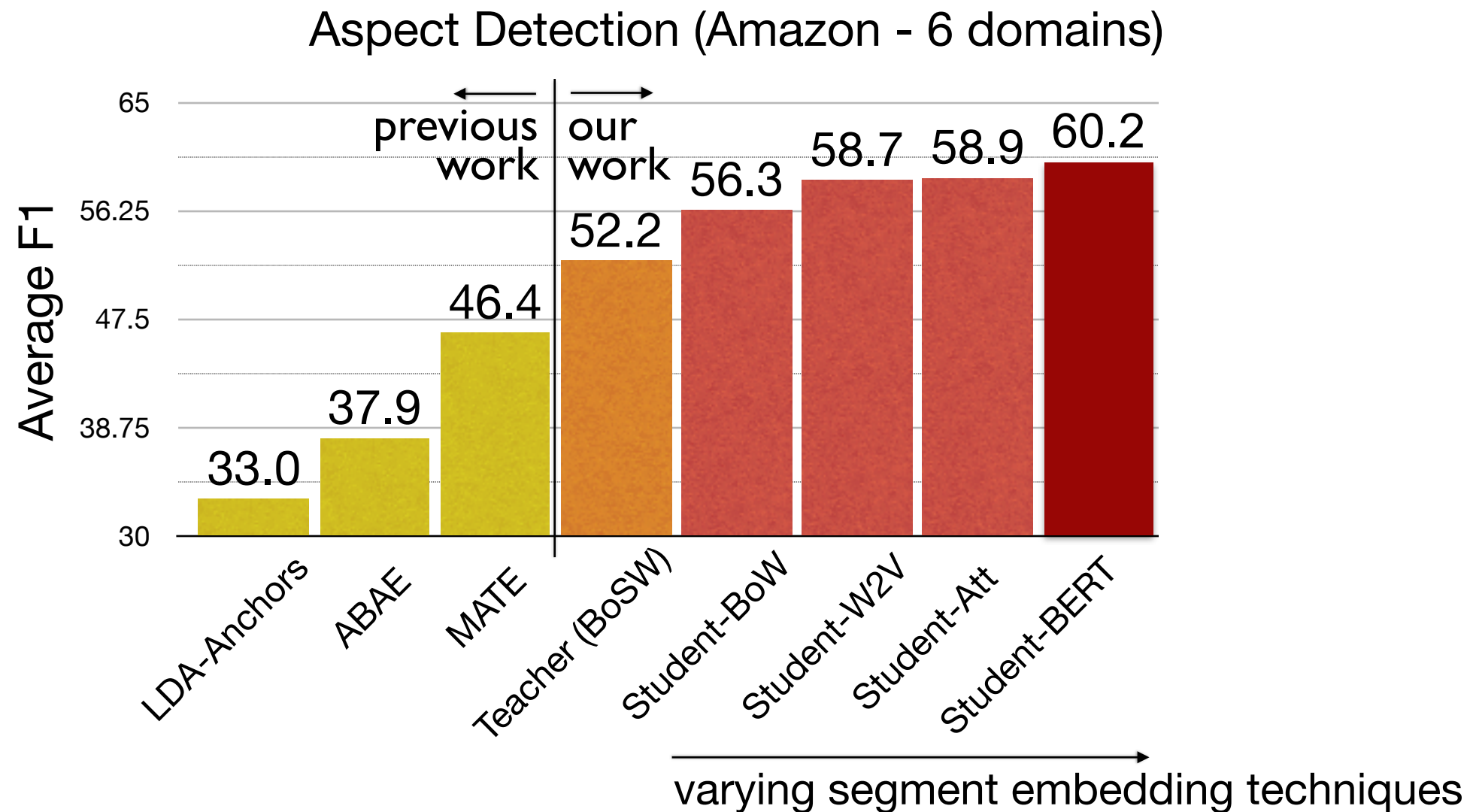


# We Leverage the Predictive Power of Seed Words



- **Teacher** and **Student-BoW** outperform previous approaches (same seed words)

# Recent Embedding Techniques Boost Student's Performance



- **Student-\*** outperform **MATE** across 12 datasets (6 domains and 6 languages)

# Outline

## 1. Fine-Grained Text Mining & Applications

## 2. Training Neural Networks with Weak Supervision

1. **“Weakly supervised attention networks for fine-grained opinion mining and public health”**,  
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (WNUT 2019)
2. **“Training neural networks for aspect extraction using descriptive keywords only”**,  
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano (LLD 2019)
3. **“Leveraging just a few keywords for fine-grained aspect detection through weakly supervised co-training”**, Giannis Karamanolakis, Daniel Hsu, and Luis Gravano (EMNLP 2019)

## 3. Conclusions and Future Work

# Training NNs for Segment Classification Without Ground Truth Segment Labels

- Leverage **review labels** in MIL networks
  - Change aggregation function (AGG) to sigmoid attention [1]
- Leverage **seed words** in neural networks
  - Use seed words in teacher-student co-training instead for initialization [2,3]
- Demonstrate advantages on 3 text mining applications
  - Segment-level aspect detection (Amazon & Yelp) [2,3]
  - Segment-level sentiment classification (Yelp & IMDB) [1]
  - Foodborne illness discovery (Yelp) [1]

[1] Karamanolakis, Hsu, and Gravano, WNUT 2019

[2] Karamanolakis, Hsu, and Gravano, LLD 2019

[3] Karamanolakis, Hsu, and Gravano, EMNLP 2019

# Current & Future Work

- **Sigmoid Attention MIL Networks:**
  - Deploy to more health departments
  - Evaluate “explainability” of fine-grained predictions
- **Teacher-Student Co-Training**
  - Discover new seed words with a human in the loop
  - Leverage logical rules in Teacher
  - Apply to more NLP tasks where classes are indicated by seed words/rules

# Thank you!

**Giannis Karamanolakis**

gkaraman@cs.columbia.edu

<https://gkaramanolakis.github.io>

