Efficient Machine Teaching Frameworks for Natural Language Processing

Doctoral Thesis Defense

Giannis Karamanolakis

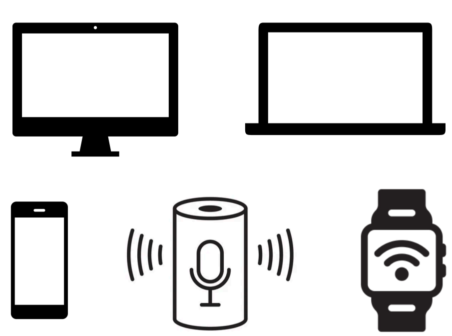
Columbia University

Efficient Machine Teaching Frameworks for Natural Language Processing

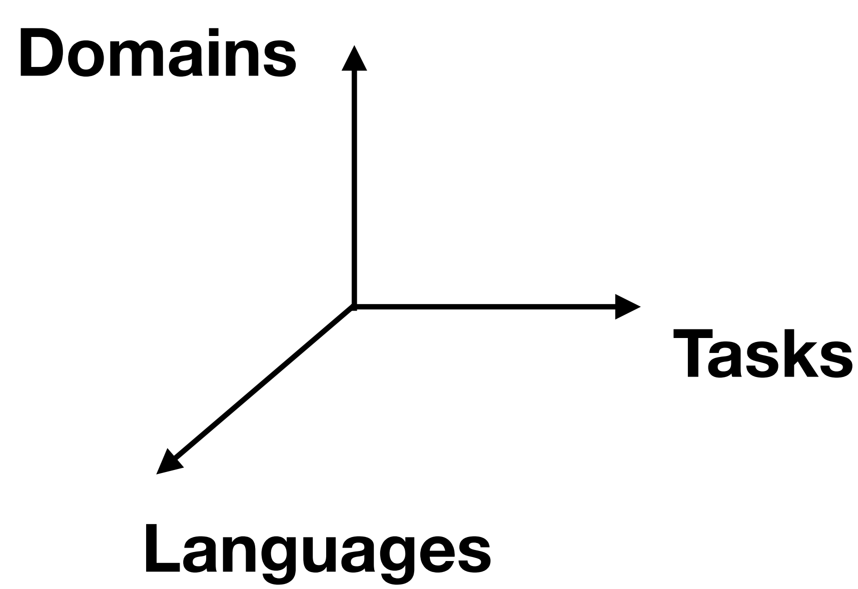
Natural Language Processing (NLP) at scale

Online text data

User interfaces

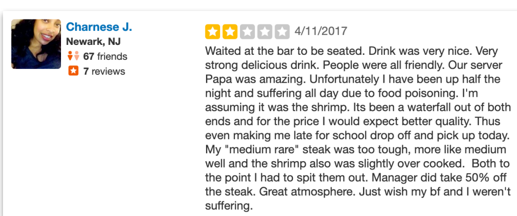


NLP across…



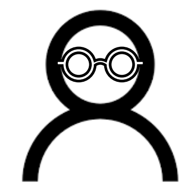
Social media analysis for public health

Restaurant Reviews


Charnese J.
Newark, NJ
67 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.

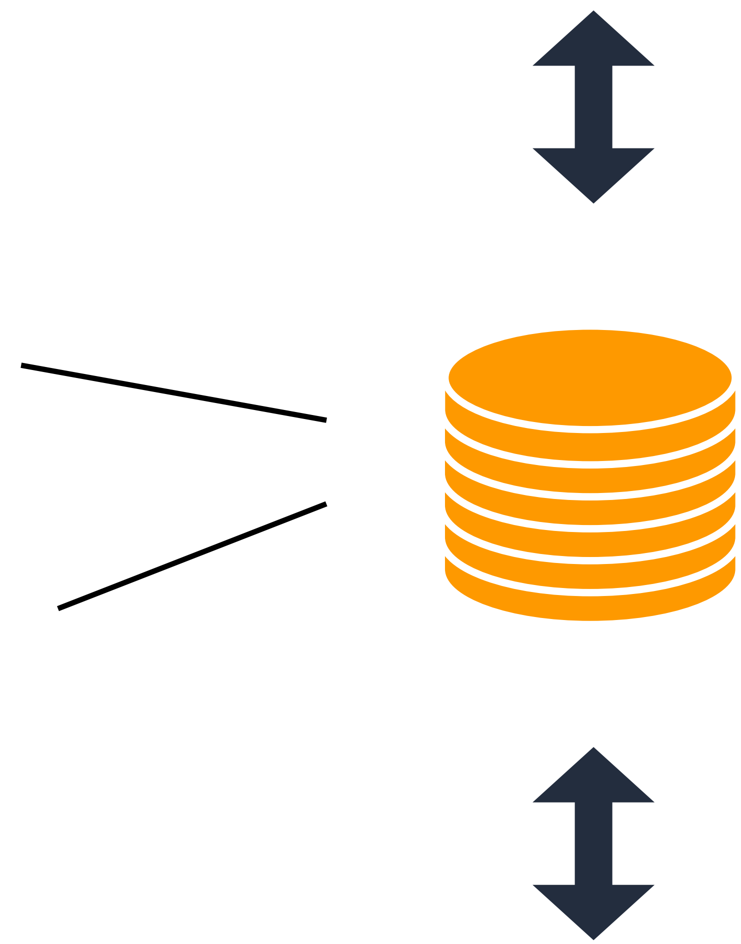
Epidemiologists



*"… Unfortunately I have been up half the night and suffering all day due to **food poisoning** …"*

Structured knowledge extraction for 10M products

flavor: “chocolate”



ingredients: “biotin”, “argan oil”, …

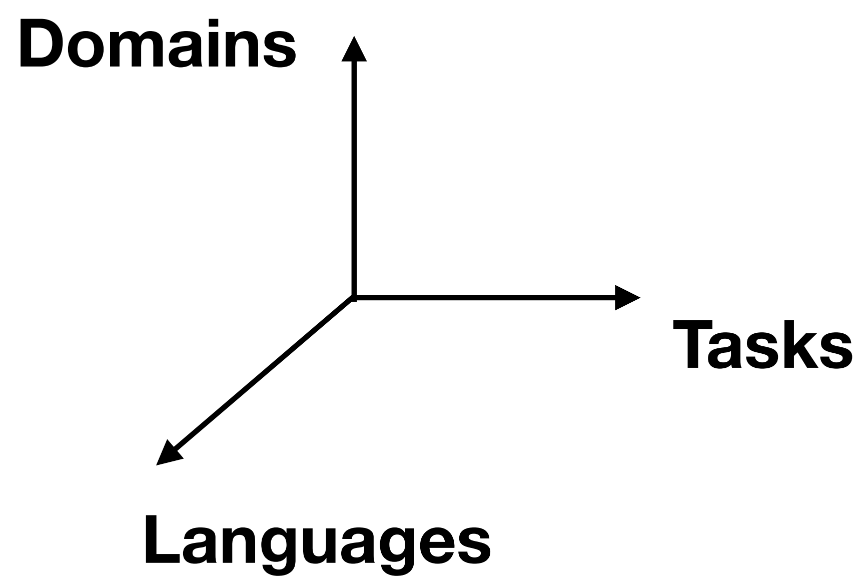
“Alexa, which shampoos contain argan oil?”

Information extraction for low-resource languages

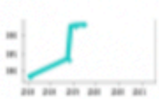
- Detecting sudden **medical emergencies** (e.g., in Uyghur)
- Extracting **flood damage** information for vulnerable countries (e.g., Bangladesh)

source: http://endangeredlanguages.com/

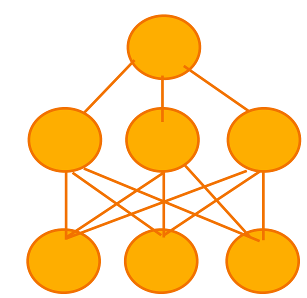
New NLP applications across…



Machine Learning (ML) is successful in NLP Benchmarks

Trend	Task	Dataset Variant	Best Model
	**Text Classification**	GLUE	distilbert-base-uncased-finetuned-sst-2-english
	**Sentiment Analysis**	SST-2 Binary classification	SMART-RoBERTa Large
	**Semantic Textual Similarity**	STS Benchmark	SMART-RoBERTa Large
	**Natural Language Inference**	MultiNLI	T5-11B
	**Semantic Textual Similarity**	MRPC	SMART-RoBERTa Large

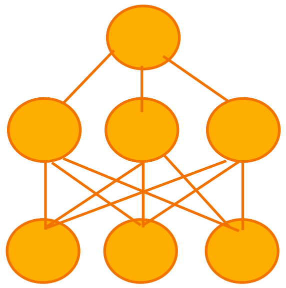
Deep neural networks



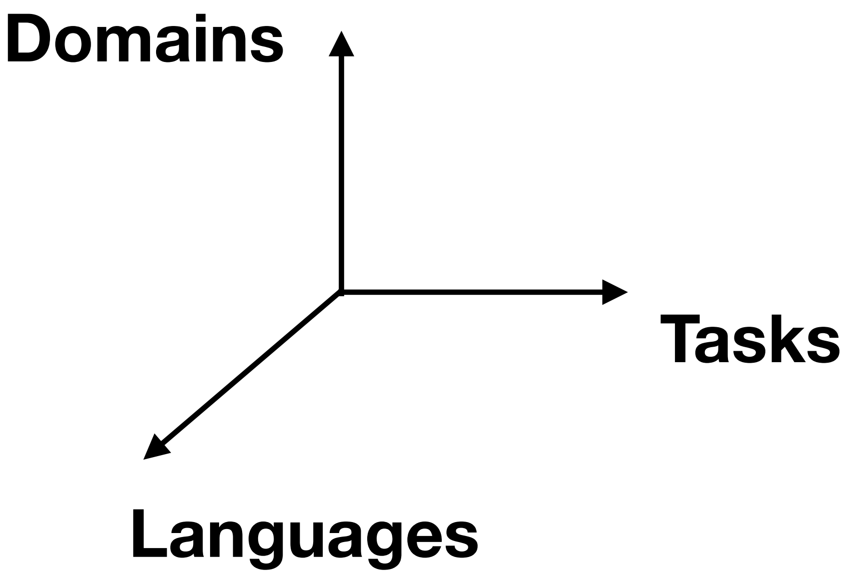
source: https://paperswithcode.com/sota/ (July 2022)

???

NLP benchmarks



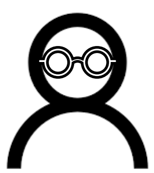
New NLP applications across…



Typical paradigm for teaching ML models: data labeling

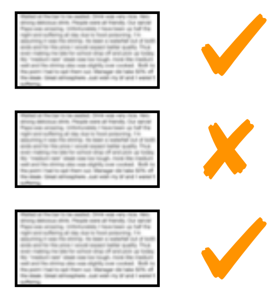
Task
(e.g., document-level, binary, food poisoning classification)

Human
(e.g., domain expert)



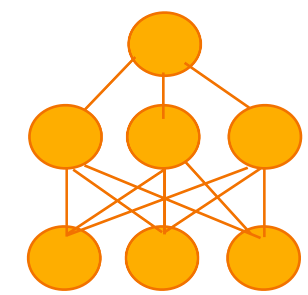
Labeled Data

Text Label



…

ML Model



Typical paradigm for teaching ML models: data labeling

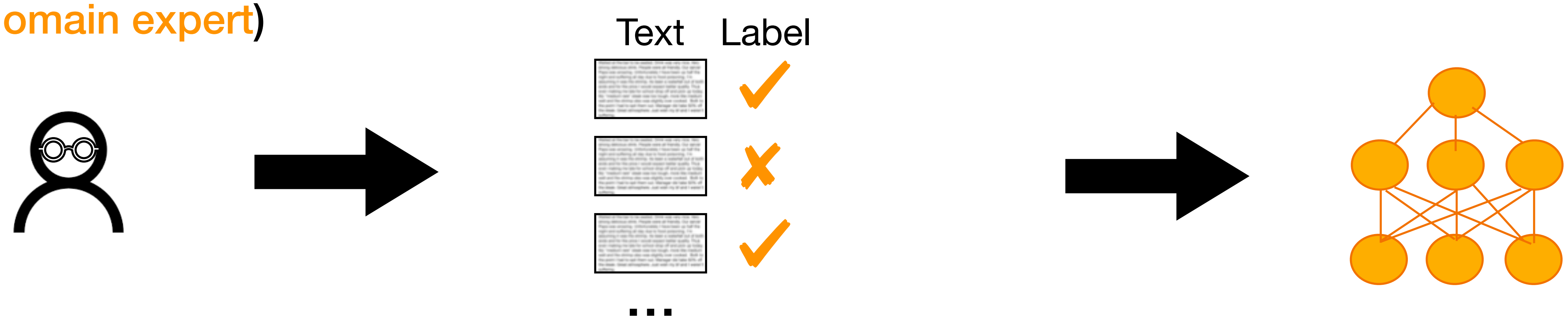
Task

(e.g., document-level, binary, food poisoning classification)

Human
(e.g., domain expert)

Labeled Data

ML Model



NLP Benchmarks

- Fixed task specifications
- Large-scale labeled data

Typical paradigm for teaching ML models: data labeling

Task

(e.g., document-level, binary, food poisoning classification)

NLP Benchmarks

- Fixed task specifications
- Large-scale labeled data

New Real-World Applications

- Dynamic task specifications
- Limited or no labeled data

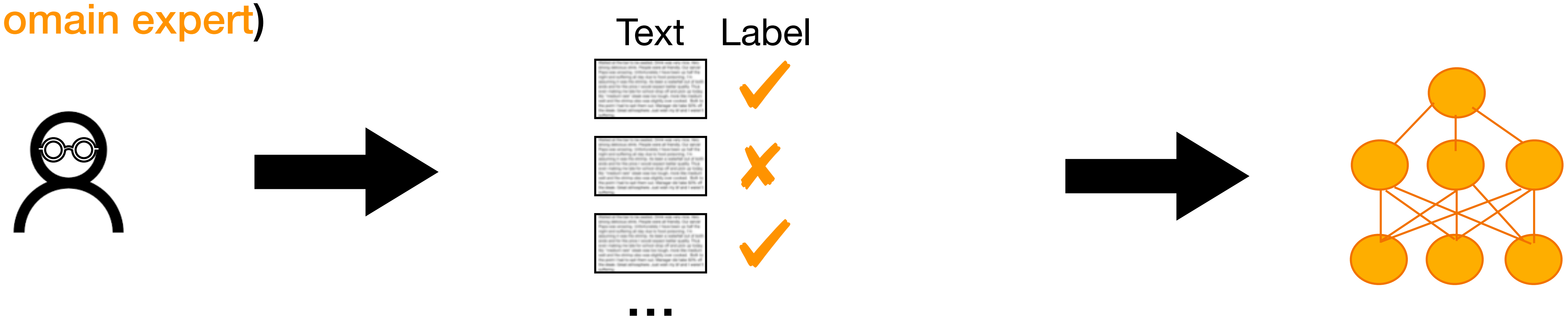
Typical paradigm for teaching ML models: data labeling

Task
(e.g., document-level, binary, food poisoning classification)

Human
(e.g., domain expert)

Labeled Data

ML Model



labeled data bottleneck

NLP Benchmarks

- Fixed task specifications
- Large-scale labeled data

New Real-World Applications

- Dynamic task specifications
- Limited or no labeled data

Typical paradigm for teaching ML models: data labeling

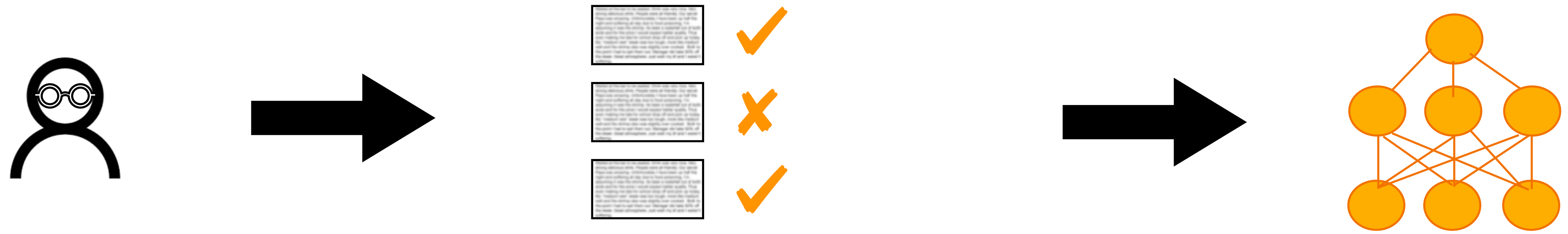
Task
(e.g., document-level, binary, food poisoning classification)

Human
(e.g., domain expert)

Labeled Data

Text Label

ML Model



...

(-) **expensive**
(-) **time-consuming**

labeled data bottleneck

NLP Benchmarks

- Fixed task specifications
- Large-scale labeled data

New Real-World Applications

- Dynamic task specifications
- Limited or no labeled data

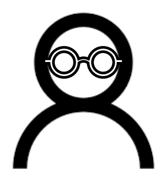
Typical paradigm for teaching ML models: data labeling

Task

(e.g., document-level, binary, food poisoning classification)

Human

(e.g., domain expert)

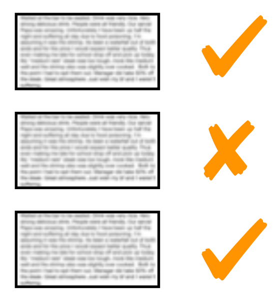


(-) expensive

(-) time-consuming

Labeled Data

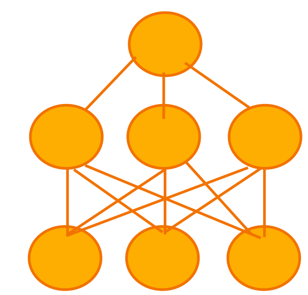
Text Label



...

(-) static

ML Model



labeled data bottleneck

NLP Benchmarks

- Fixed task specifications
- Large-scale labeled data

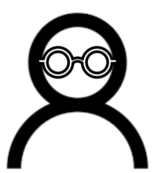
New Real-World Applications

- Dynamic task specifications
- Limited or no labeled data

Typical paradigm for teaching ML models: data labeling

Task
(e.g., ~~document-~~level, binary, food poisoning classification)
sentence-

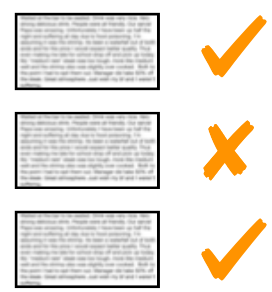
Human
(e.g., domain expert)



(-) **expensive**
(-) **time-consuming**

Labeled Data

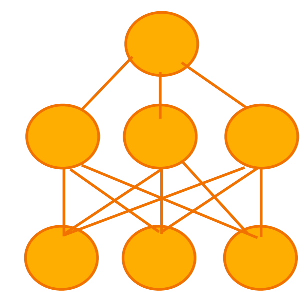
Text Label



...

(-) **static**

ML Model



labeled data bottleneck

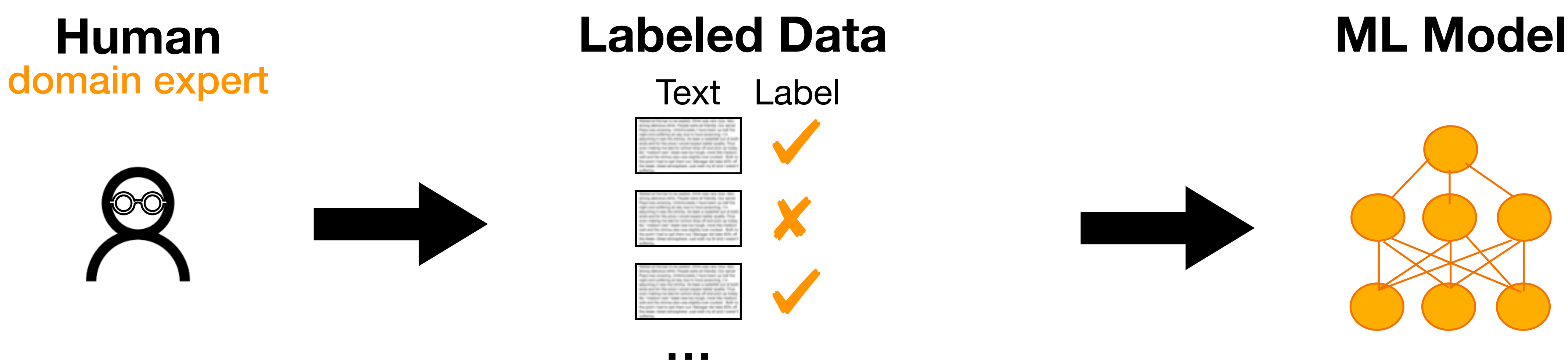
NLP Benchmarks

- Fixed task specifications
- Large-scale labeled data

New Real-World Applications

- Dynamic task specifications
- Limited or no labeled data

How to relax the labeled data bottleneck?



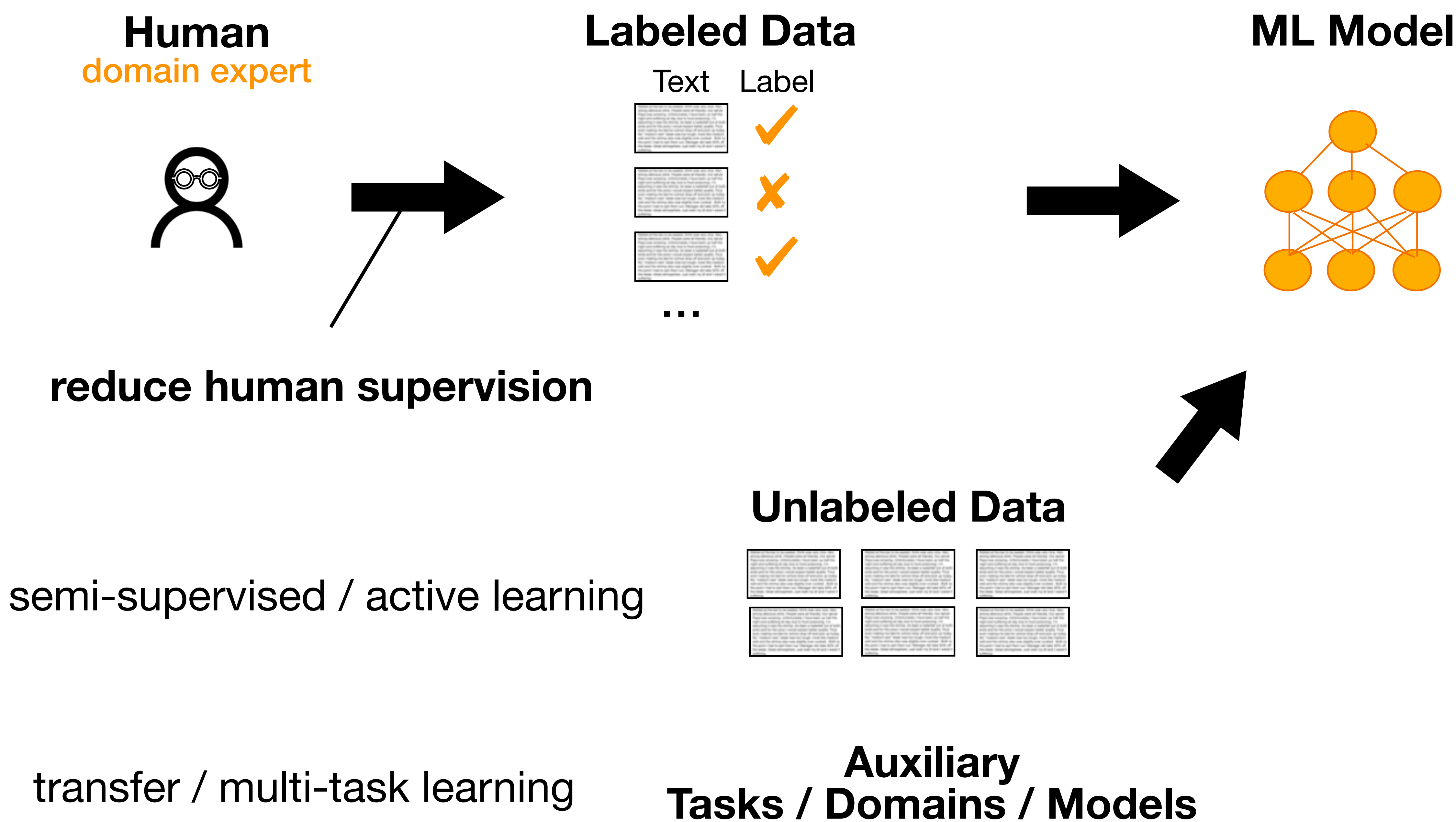
How to relax the labeled data bottleneck?

- **Common approaches:** learning with **limited** labeled data

reduce human supervision

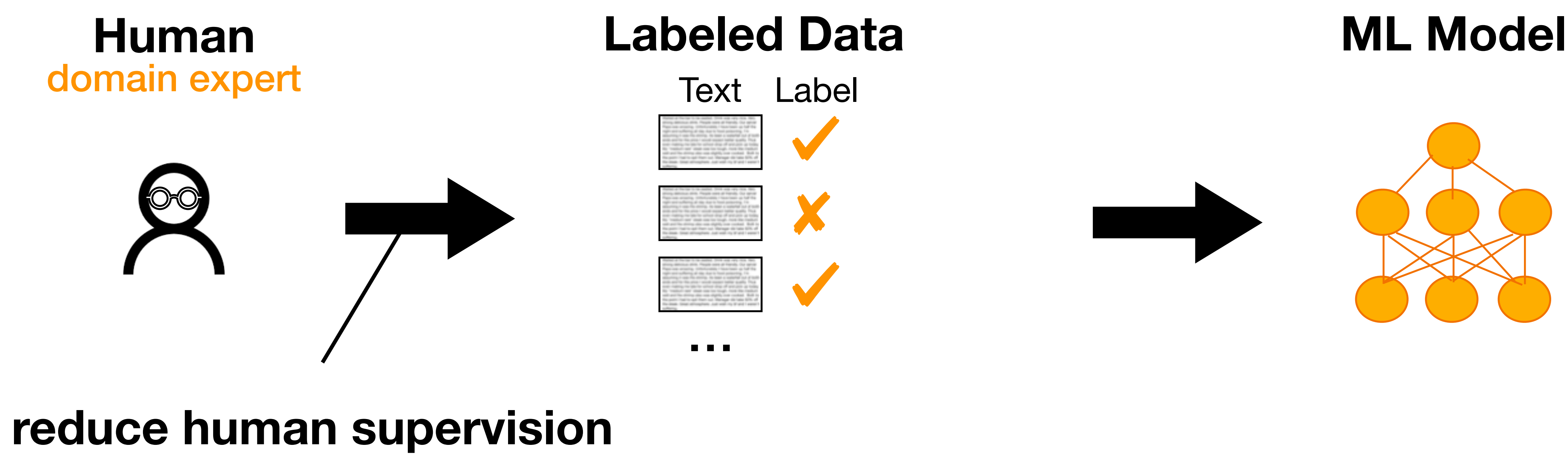
How to relax the labeled data bottleneck?

- **Common approaches:** learning with **limited** labeled data



How to relax the labeled data bottleneck?

- **Common approaches:** learning with **limited** labeled data

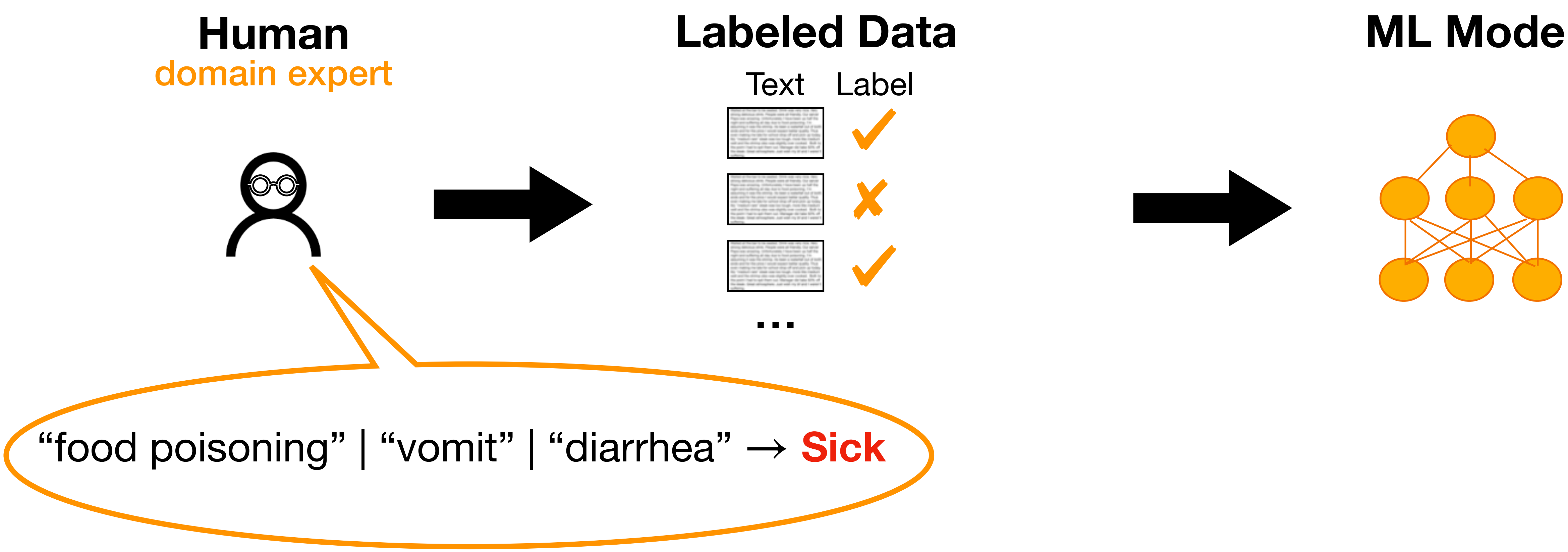


(-) Expert feedback is restricted to labeling individual instances

(-) Instance labels carry limited information

How to relax the labeled data bottleneck?

- **Common approaches:** learning with **limited** labeled data



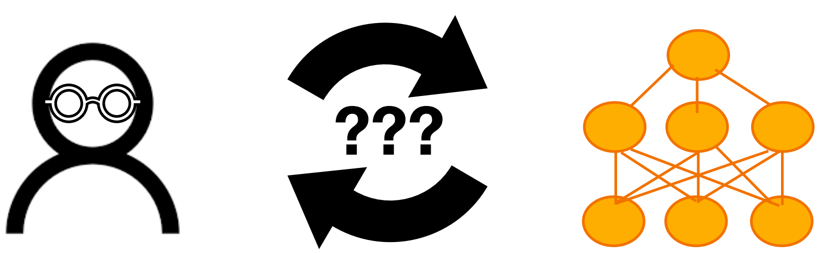
(-) Expert feedback is restricted to labeling individual instances

(-) Instance labels carry limited information

Our goal

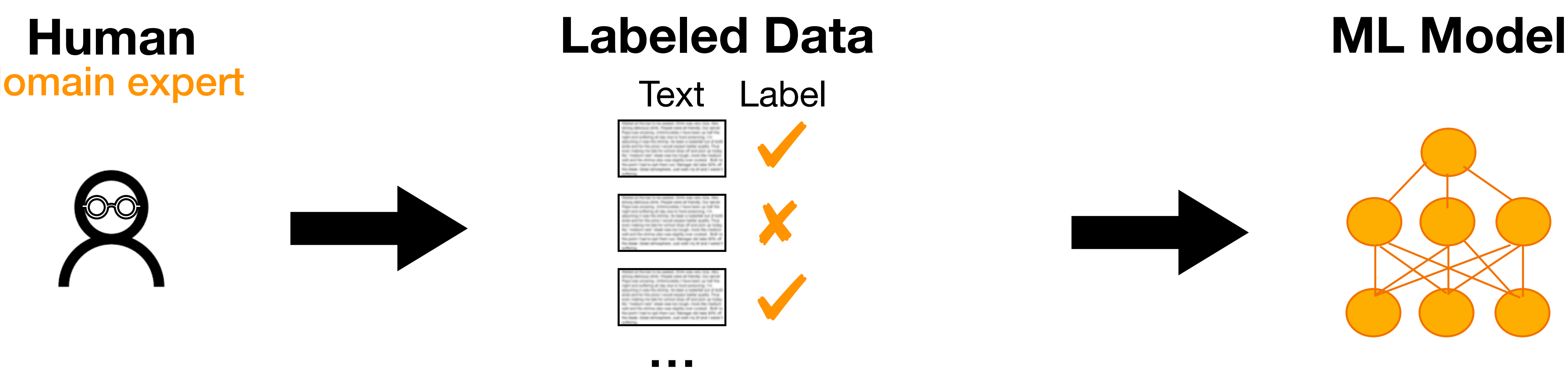
Assist humans in teaching ML models

via **more flexible types of interaction**



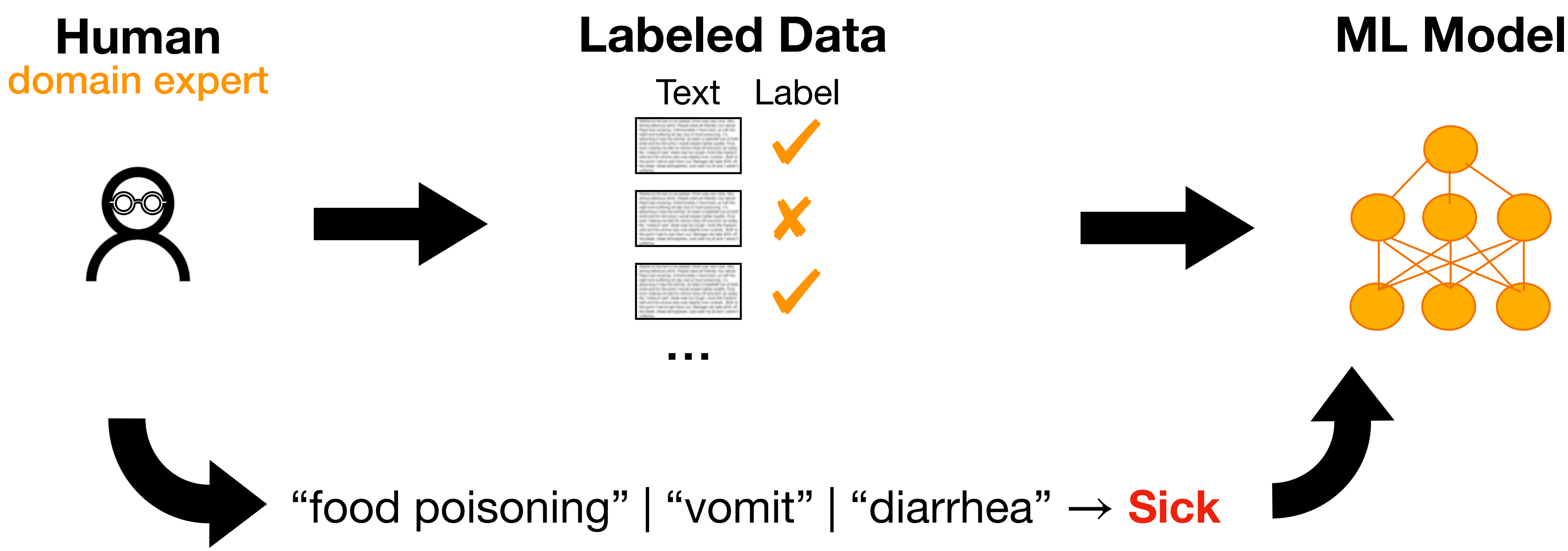
This dissertation

- **Common approaches:** learning with **limited** labeled data
- **This dissertation:** we support **more types of supervision** for teaching ML models



This dissertation

- **Common approaches:** learning with **limited** labeled data
- **This dissertation:** we support **more types of supervision** for teaching ML models

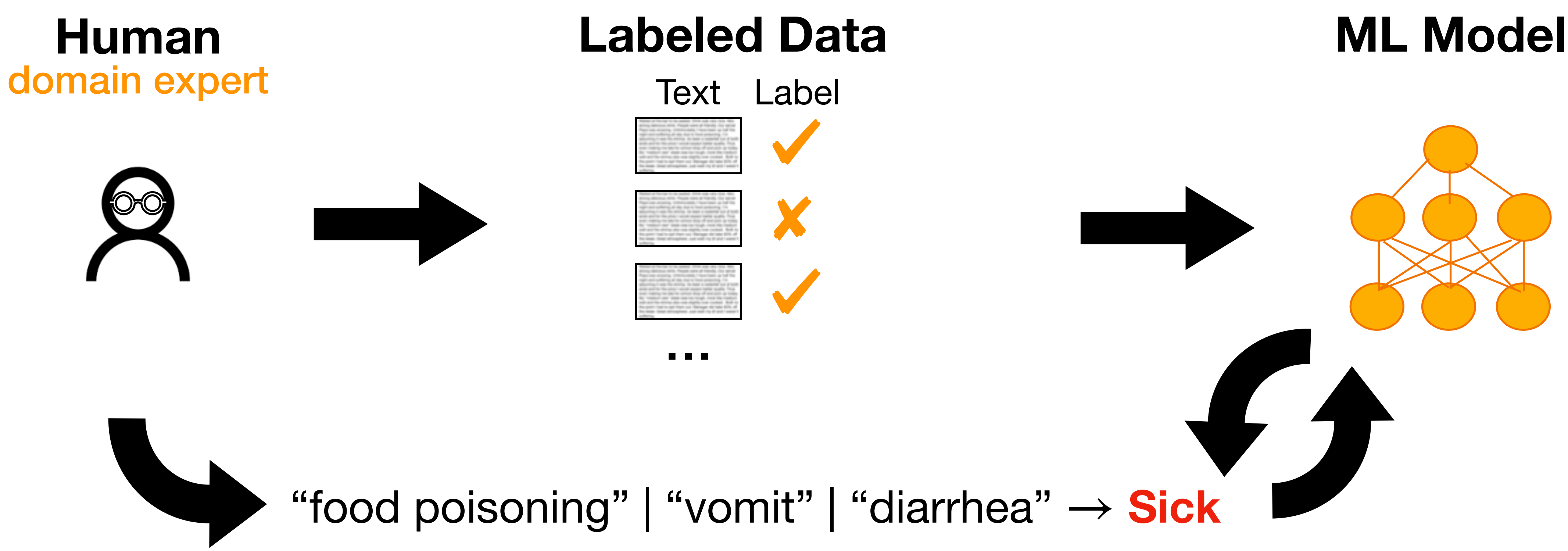


minimal domain expertise

beyond standard labels

This dissertation

- **Common approaches:** learning with **limited** labeled data
- **This dissertation:** we support **more types of supervision** for teaching ML models



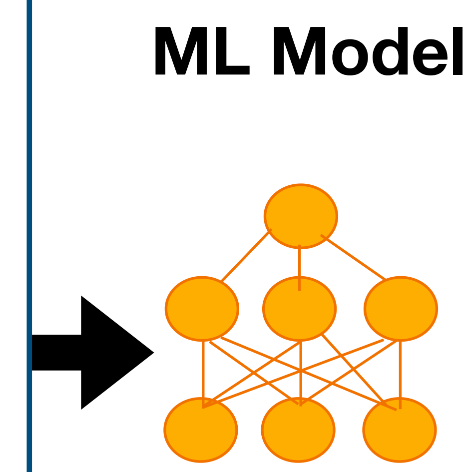
minimal domain expertise
beyond standard labels

+

generalization power
of deep neural networks

Contributions of this dissertation

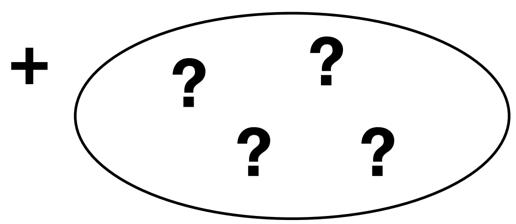
We develop **efficient** frameworks for teaching ML models using…



Contributions of this dissertation

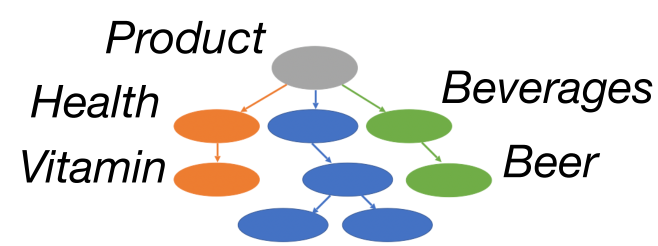
We develop **efficient** frameworks for teaching ML models using…

Coarse labels (Ch. 3)



[Karamanolakis et al. WNUT '19]

Hierarchical taxonomies (Ch. 4)



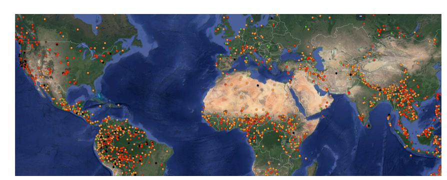
[Karamanolakis et al. ACL '20]

Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP '19]

Word translations (Ch. 6)

"injured"  "يارىلانغان"

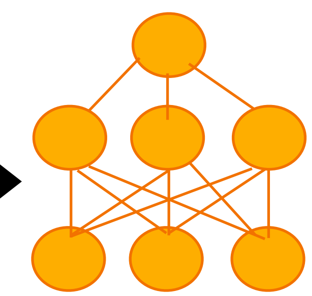
[Karamanolakis et al. Findings of EMNLP '20]

Labeling rules (Ch. 7, 8)

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

[Karamanolakis et al. NAACL '21] + work with Daniel Hsu and Luis Gravano

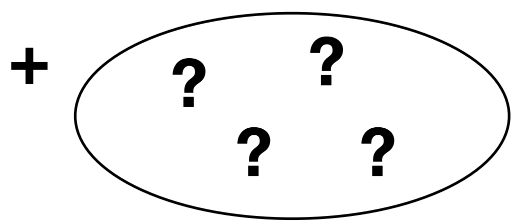
ML Model



Contributions of this dissertation

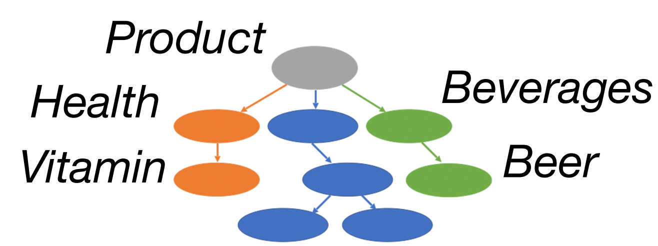
We develop **efficient** frameworks for teaching ML models using…

Coarse labels (Ch. 3)



[Karamanolakis et al. WNUT ’19]

Hierarchical taxonomies (Ch. 4)



[Karamanolakis et al. ACL ’20]

Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP ’19]

Word translations (Ch. 6)

“injured”

“يارىلانغان”

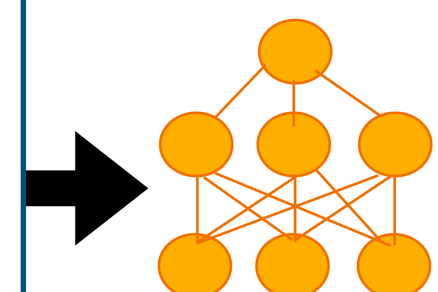
[Karamanolakis et al. Findings of EMNLP ’20]

Labeling rules (Ch. 7, 8)

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

[Karamanolakis et al. NAACL ’21] + work with Daniel Hsu and Luis Gravano

ML Model



Contributions of this dissertation

We develop **efficient** frameworks for teaching ML models using…

Coarse labels (Ch. 3)

[Karamanolakis et al. WNUT ’19]

Hierarchical taxonomies (Ch. 4)

[Karamanolakis et al. ACL ’20]

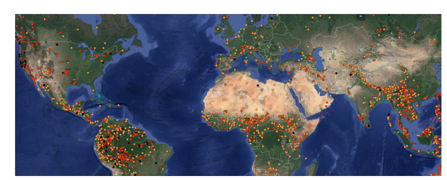
Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP ’19]

Word translations (Ch. 6)

“injured”



“يارىلانغان”

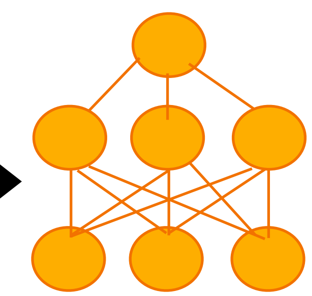
[Karamanolakis et al. Findings of EMNLP ’20]

Labeling rules (Ch. 7, 8)

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

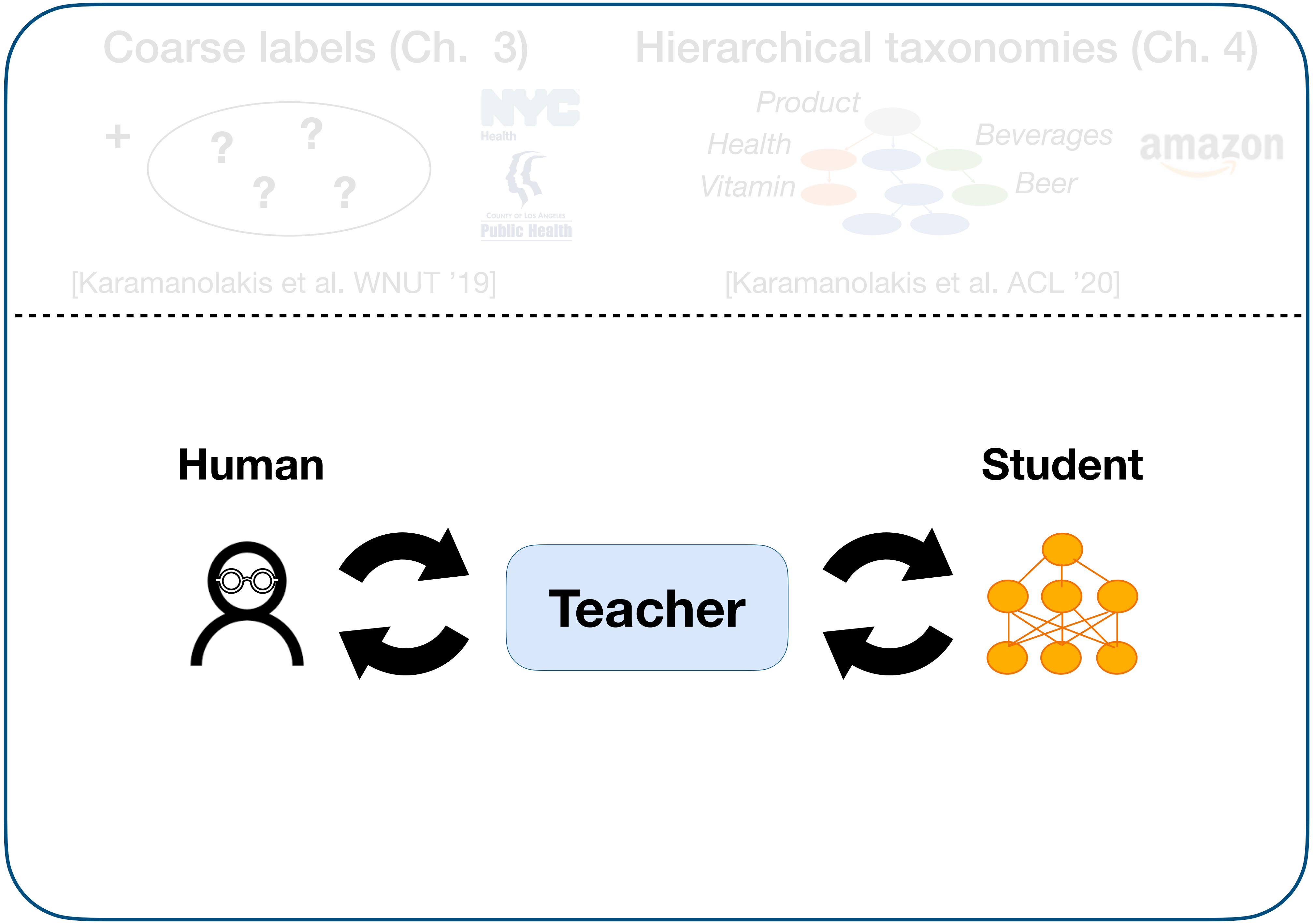
[Karamanolakis et al. NAACL ’21] + work with Daniel Hsu and Luis Gravano

ML Model



Contributions of this dissertation

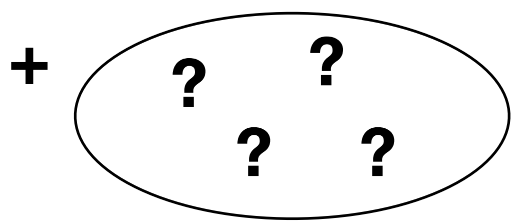
We develop **efficient** frameworks for teaching ML models using…



Contributions of this dissertation

We develop **efficient** frameworks for teaching ML models using…

Coarse labels (Ch. 3)



[Karamanolakis et al. WNUT ’19]

Hierarchical taxonomies (Ch. 4)

[Karamanolakis et al. ACL ’20]

Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP ’19]

Word translations (Ch. 6)

“injured”

“يارىلانغان”

[Karamanolakis et al. Findings of EMNLP ’20]

Labeling rules (Ch. 7, 8)

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

[Karamanolakis et al. NAACL ’21] + work with Daniel Hsu and Luis Gravano

Fine-grained classification with coarse labels

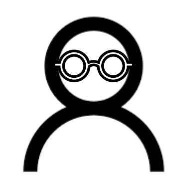
Fine-grained classification with coarse labels

- Epidemiologists read restaurant reviews for evidence of **food poisoning**

Restaurant Review

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.

Epidemiologist



Fine-grained classification with coarse labels

- Epidemiologists read restaurant reviews for evidence of **food poisoning**

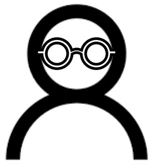
It is **time-consuming** to read long reviews

Restaurant Review

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.

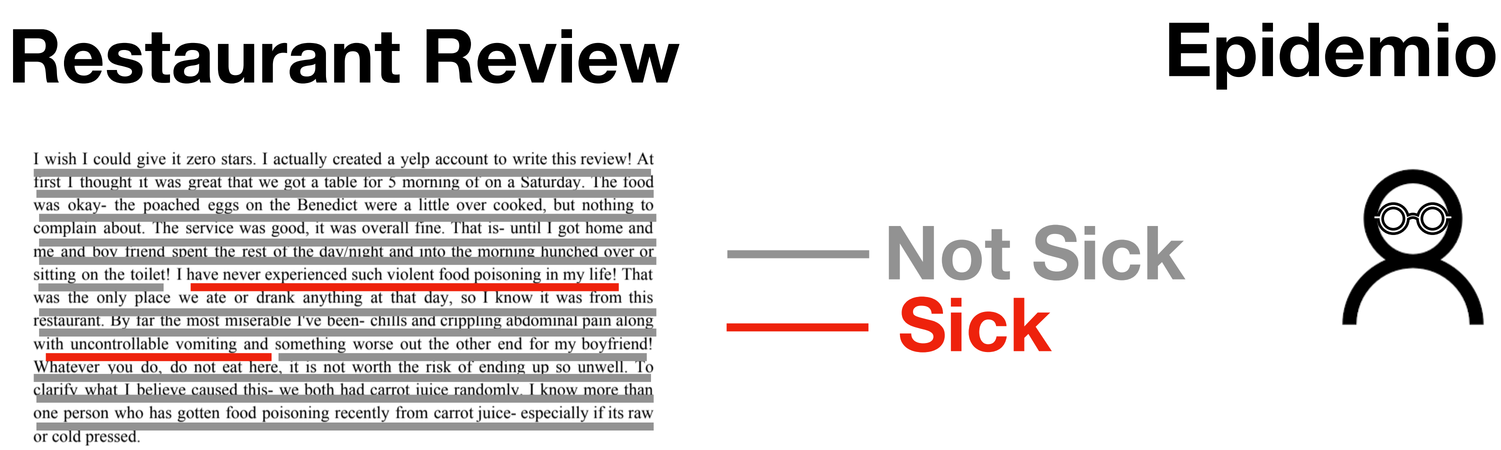
Sick?

Epidemiologist



Fine-grained classification with coarse labels

- Epidemiologists read restaurant reviews for evidence of **food poisoning**
- Task: train a **sentence** classifier



Fine-grained classification with coarse labels

- Epidemiologists read restaurant reviews for evidence of **food poisoning**
- Task: train a **sentence** classifier

Sentence labels are **expensive** to obtain for training

Restaurant Review

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.

Epidemiologist

Fine-grained classification with coarse labels

- Epidemiologists read restaurant reviews for evidence of **food poisoning**
- Task: train a **sentence** classifier

Sentence labels are **expensive** to obtain for training

We use **coarse review labels** that are **easier** to obtain

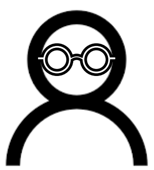
Restaurant Review (Sick)

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.

Not Sick

Sick

Epidemiologist



How to train sentence classifiers using coarse review labels?

review label (observed)		p		
sentence labels (unobserved)	p_1	p_2	$\cdots$	p_M

How to train sentence classifiers using coarse review labels?

- Assume all sentences in a review have the same label? $p_1 = p_2 = \ldots = p$

review label
(observed)

p

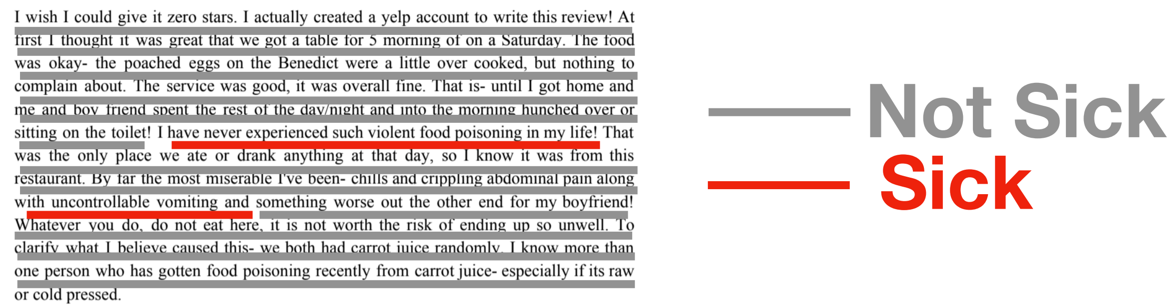
sentence labels
(unobserved)

$p_1 \qquad p_2 \qquad \cdots \qquad p_M$

How to train sentence classifiers using coarse review labels?

- Assume all sentences in a review have the same label? $p_1 = p_2 = \ldots = p$

  (-) not effective

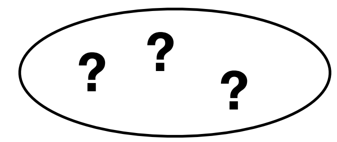


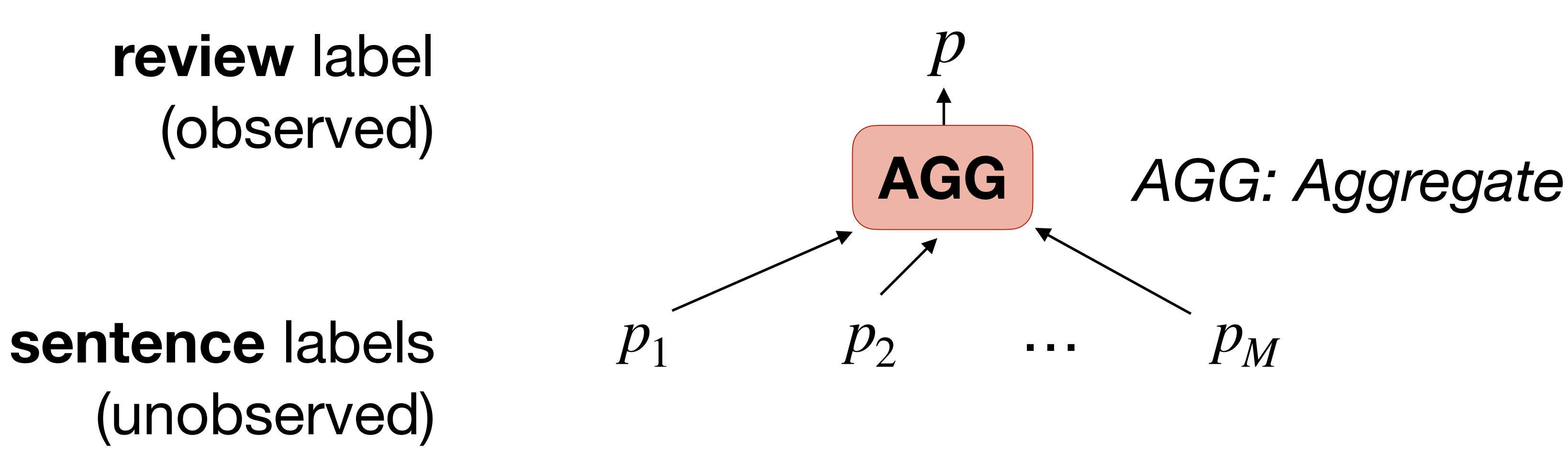
review label (observed) p

sentence labels (unobserved) $p_1 \quad p_2 \quad \cdots \quad p_M$

How to train sentence classifiers using coarse review labels?

- Assume all sentences in a review have the same label?
- We employ Multiple Instance Learning (MIL) +

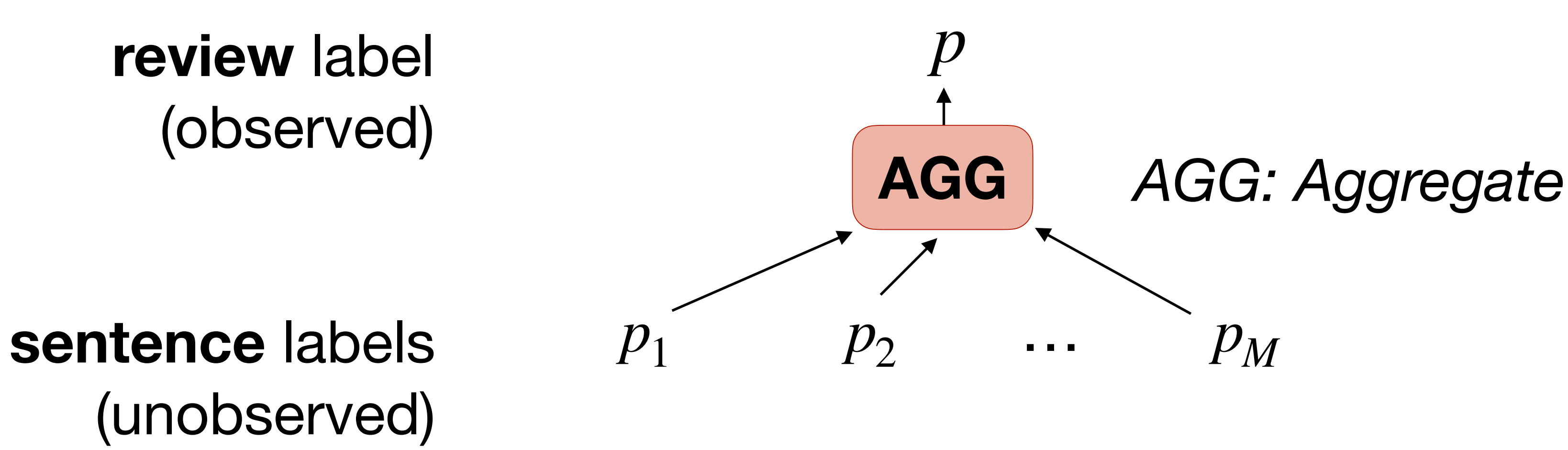




How to train sentence classifiers using coarse review labels?

- Assume all sentences in a review have the same label?
- We employ Multiple Instance Learning (MIL)
  - State-of-the-art methods: hierarchical MIL networks

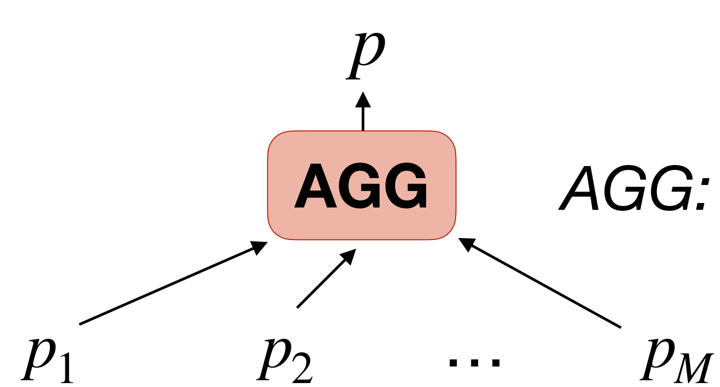
[Pappas and Popescu-belis, 2014;2017]
[Angelidis and Lapata, 2018]



How to train sentence classifiers using coarse review labels?

- Assume all sentences in a review have the same label?
- We employ Multiple Instance Learning (MIL)
  - State-of-the-art methods: hierarchical MIL networks [Pappas and Popescu-belis, 2014;2017] [Angelidis and Lapata, 2018]
  - Our work: MIL network with new AGG function

review label (observed)



AGG: Aggregate

sentence labels (unobserved)

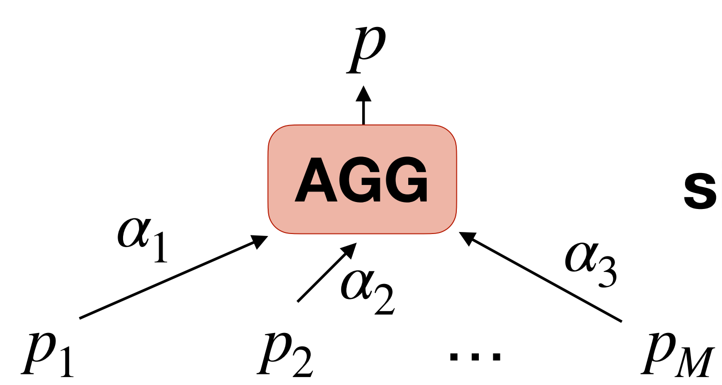
"Weakly supervised attention networks for fine-grained opinion mining and public health", Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (WNUT@EMNLP 2019)

How to train sentence classifiers using coarse review labels?

- Assume all sentences in a review have the same label?
- We employ Multiple Instance Learning (MIL)
  - State-of-the-art methods: hierarchical MIL networks
  - Our work: MIL network with new AGG function

[Pappas and Popescu-belis, 2014;2017]
[Angelidis and Lapata, 2018]

review label
(observed)



sigmoid attention

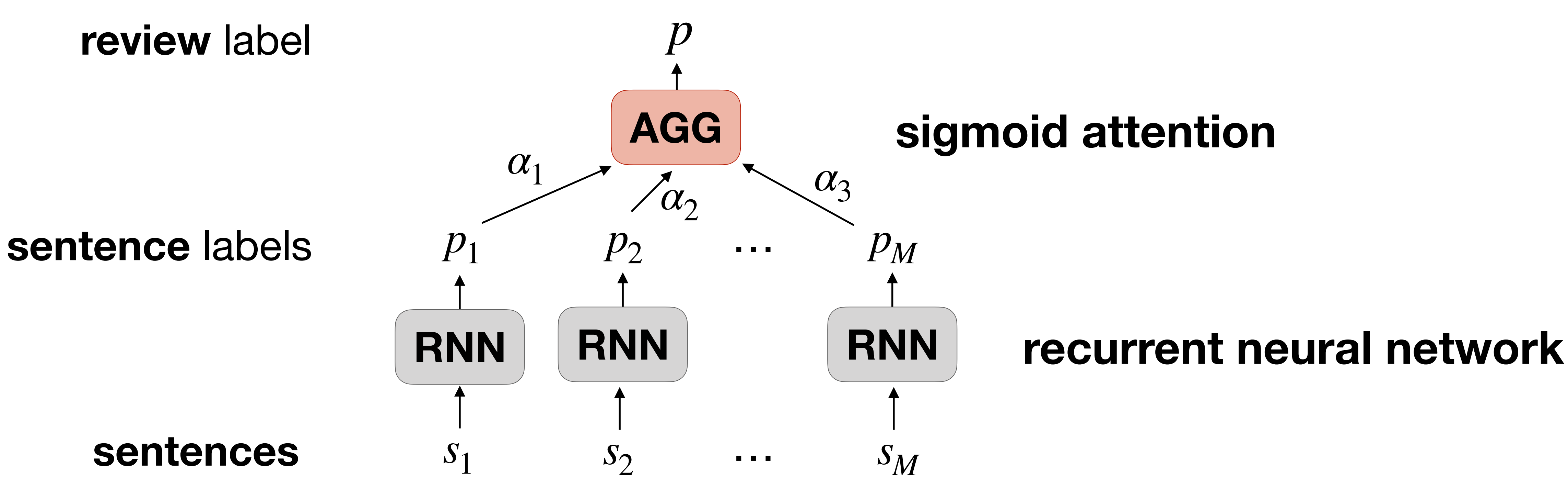
sentence labels
(unobserved)

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

How to train sentence classifiers using coarse review labels?

- Assume all sentences in a review have the same label?
- We employ Multiple Instance Learning (MIL)
  - State-of-the-art methods: hierarchical MIL networks [Pappas and Popescu-belis, 2014;2017] [Angelidis and Lapata, 2018]
  - Our work: MIL network with new AGG function

Hierarchical Sigmoid Attention Network (HSAN)



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

Evaluation of HSAN for opinion mining and public health

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

Evaluation of HSAN for opinion mining and public health

1. We improve previous methods in standard **opinion mining benchmarks**

Sentiment Classification (Yelp)

Aggregation Function **AGG**	**F1**
average (Kotzias et al., 2015)	46.8
softmax (Angelidis and Lapata, 2018)	59.9
sigmoid (our approach)	**63.3 (+5.6%)**

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

Evaluation of HSAN for opinion mining and public health

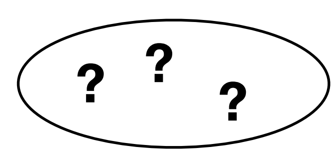
1. We improve previous methods in standard **opinion mining benchmarks**

Sentiment Classification (Yelp)

Aggregation Function **AGG**	**F1**
average (Kotzias et al., 2015)	46.8
softmax (Angelidis and Lapata, 2018)	59.9
sigmoid (our approach)	**63.3** **(+5.6%)**

Sigmoid attention is effective in **MIL** when the **witness rate** is high

+



% of positive instances in positive bags

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

Evaluation of HSAN for opinion mining and public health

1. We improve previous methods in standard **opinion mining benchmarks**
2. We apply HSAN to detect **food poisoning** in Yelp restaurant reviews

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

Evaluation of HSAN for opinion mining and public health

1. We improve previous methods in standard **opinion mining benchmarks**
2. We apply HSAN to detect **food poisoning** in Yelp restaurant reviews
    - Achieve **48.6% higher recall** than previous classifier [Effland et al. 2018]

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

Evaluation of HSAN for opinion mining and public health

1. We improve previous methods in standard **opinion mining benchmarks**
2. We apply HSAN to detect **food poisoning** in Yelp restaurant reviews
   - Achieve **48.6% higher recall** than previous classifier [Effland et al. 2018]
   - Deploy HSAN for inspections in health departments in NYC / LA County

🤢 : Sick ✓ : Not Sick

Pred	Text
✓	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.

I have never experienced such violent food poisoning in my life

… crippling abdominal pain along with uncontrollable vomiting …

https://publichealth.cs.columbia.edu/

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

Evaluation of HSAN for opinion mining and public health

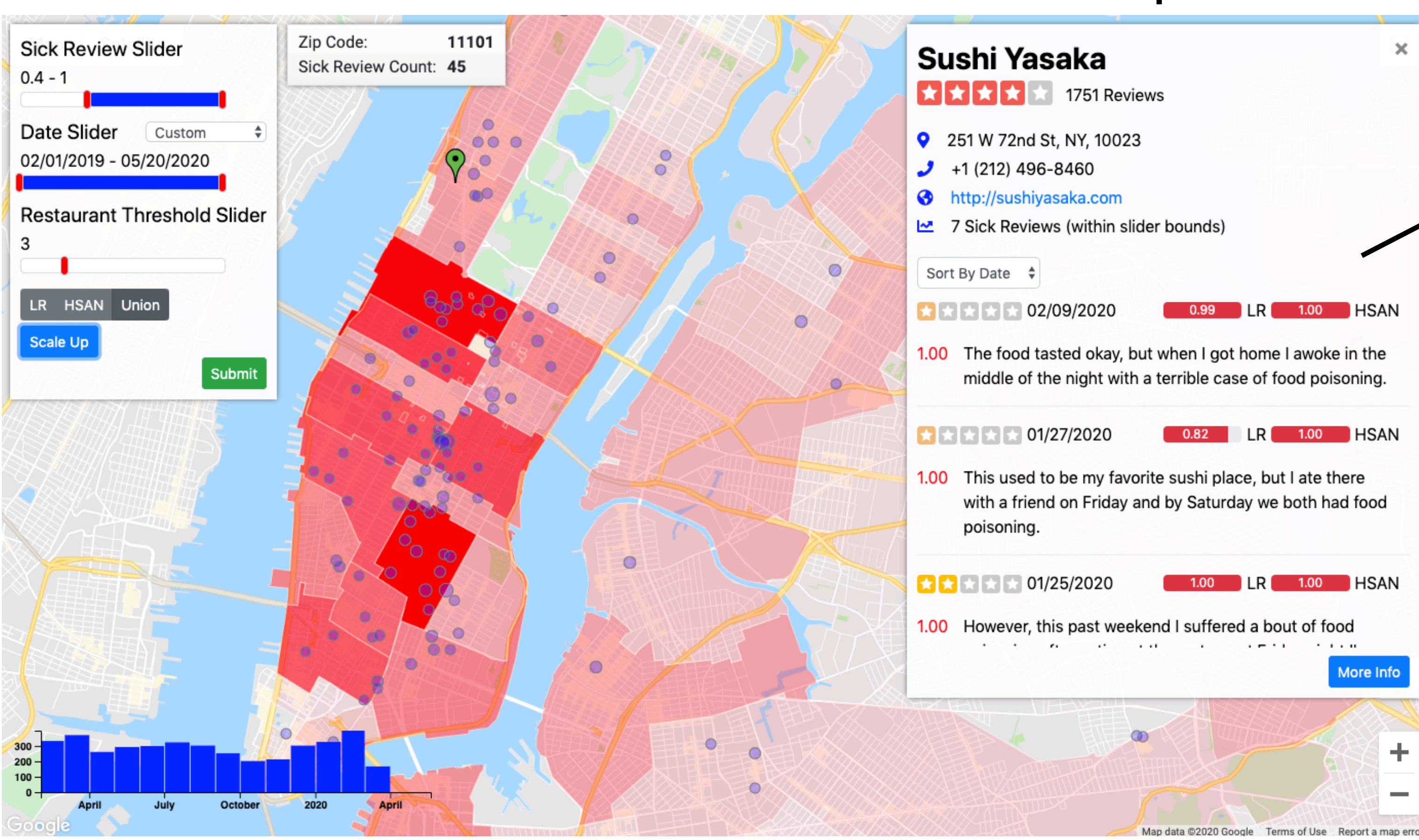
1. We improve previous methods in standard **opinion mining benchmarks**
2. We apply HSAN to detect **food poisoning** in Yelp restaurant reviews

- Achieve **48.6% higher recall** than previous classifier [Effland et al. 2018]
- Deploy HSAN for inspections in health departments in NYC / LA County

NYC Health

County of Los Angeles Public Health

Interactive Foodborne Illness Map



Visualize only important sentences!

work with Sam Raab

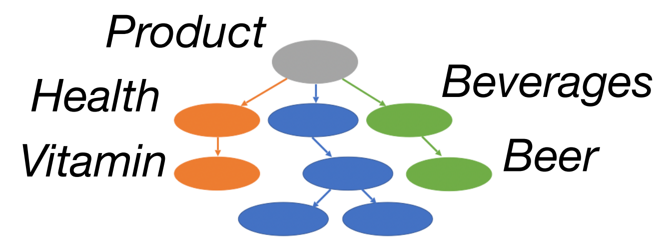
https://publichealth.cs.columbia.edu/

Efficient machine teaching frameworks for NLP

Coarse labels (Ch. 3)

[Karamanolakis et al. WNUT ’19]

Hierarchical taxonomies (Ch. 4)



[Karamanolakis et al. ACL ’20]

(3) Seed words

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

[Karamanolakis et al. EMNLP ’19]

(4) Word translations

“injured”

“ياربلانغان”

[Karamanolakis et al. Findings of EMNLP ’20]

(5) Labeling rules

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x) else ABSTAIN
```

[Karamanolakis et al. NAACL ’21]

Extraction of product attributes from descriptions

Ben & Jerry's Strawberry Cheesecake Ice Cream 16 oz

Brand	Flavor	Size

Extraction of product attributes from descriptions

Ben & Jerry's Strawberry Cheesecake Ice Cream 16 oz

Brand	Flavor	Size
Ben & Jerry’s	strawberry cheesecake	16 oz

Extraction of product attributes from descriptions

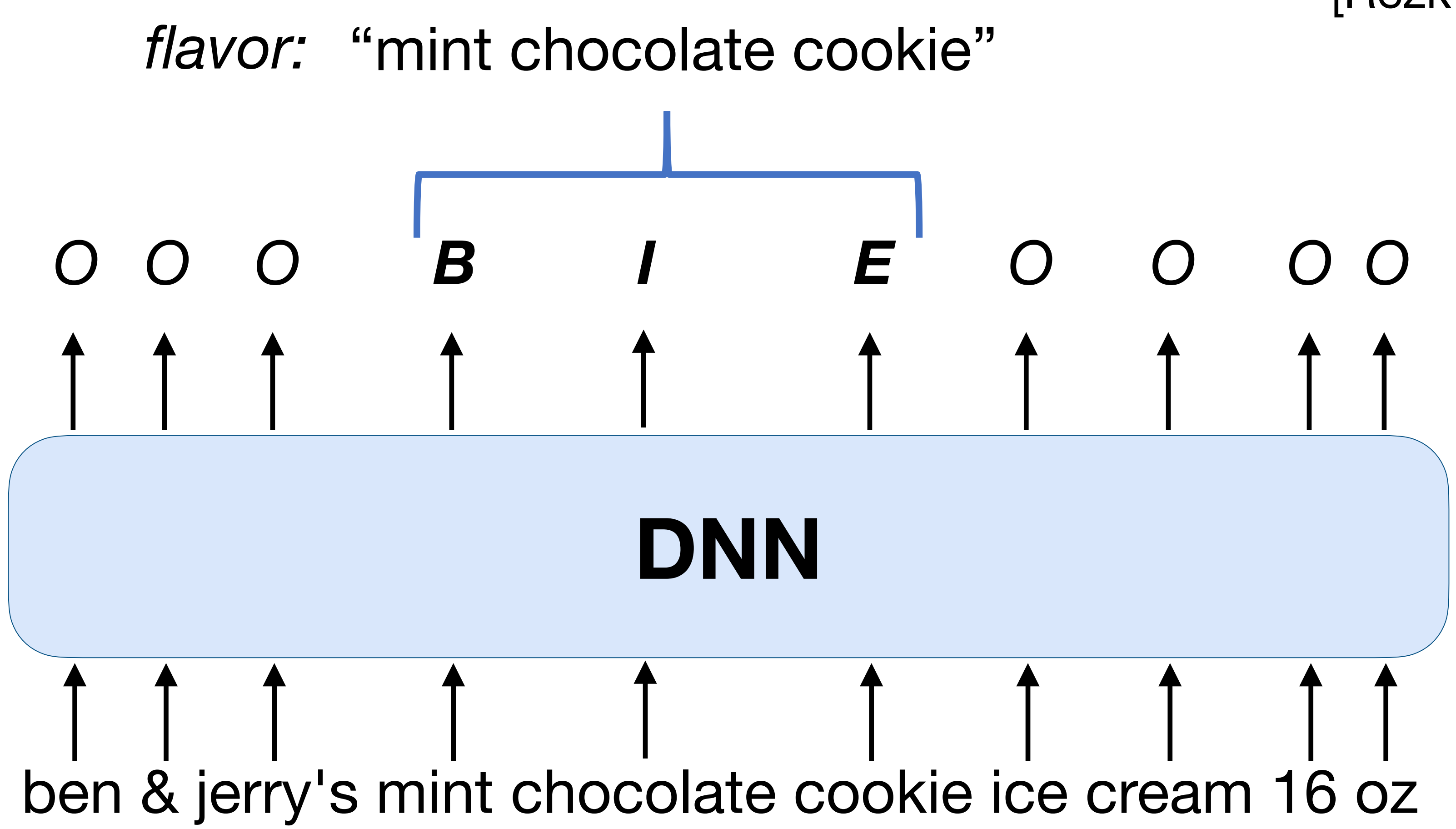
- **Previous work:** deep neural networks for **sequence tagging**

[Zheng et al., KDD'18]
[Xu et al., ACL'19]
[Rezk et al., ICDE'19]

Extraction of product attributes from descriptions

- **Previous work:** deep neural networks for **sequence tagging**

[Zheng et al., KDD'18]
[Xu et al., ACL'19]
[Rezk et al., ICDE'19]



Extraction of product attributes from descriptions

- **Previous work:** deep neural networks for **sequence tagging**

[Zheng et al., KDD'18]
[Xu et al., ACL'19]
[Rezk et al., ICDE'19]

- **Limitations of previous work:**

  (-) designed for products under a **single** product category

OpenTag [Zheng et al., KDD'18]

"dog food" category → training data *for "dog food" products* $\xrightarrow{train()}$ **DNN** *for "dog food" products*

Extraction of product attributes from descriptions

- **Previous work:** deep neural networks for **sequence tagging**

[Zheng et al., KDD'18]
[Xu et al., ACL'19]
[Rezk et al., ICDE'19]

- **Limitations of previous work:**
  - **(-)** designed for products under a **single** product category
  - **(-)** hard to scale to large e-commerce companies

amazon

- 10M+ products
- 10K+ categories

We scale up extraction to 4K product categories

“TXtract: taxonomy-aware knowledge extraction from thousands of product categories”,
Giannis Karamanolakis, Jun Ma, and Xin Luna Dong. (ACL 2020)

We scale up extraction to 4K product categories

“TXtract: taxonomy-aware knowledge extraction from thousands of product categories”,
Giannis Karamanolakis, Jun Ma, and Xin Luna Dong. (ACL 2020) amazon

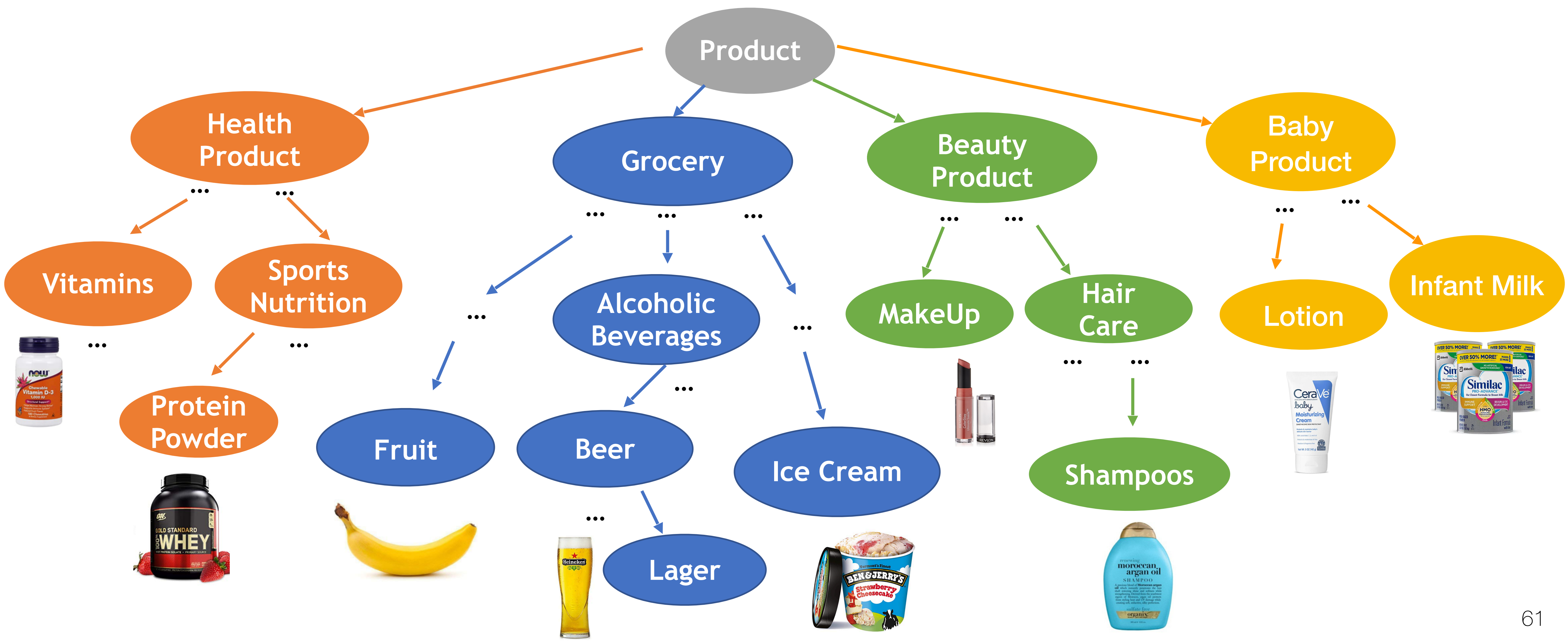
Train 4K models?

(-) expensive

Assume a single “flat” space?

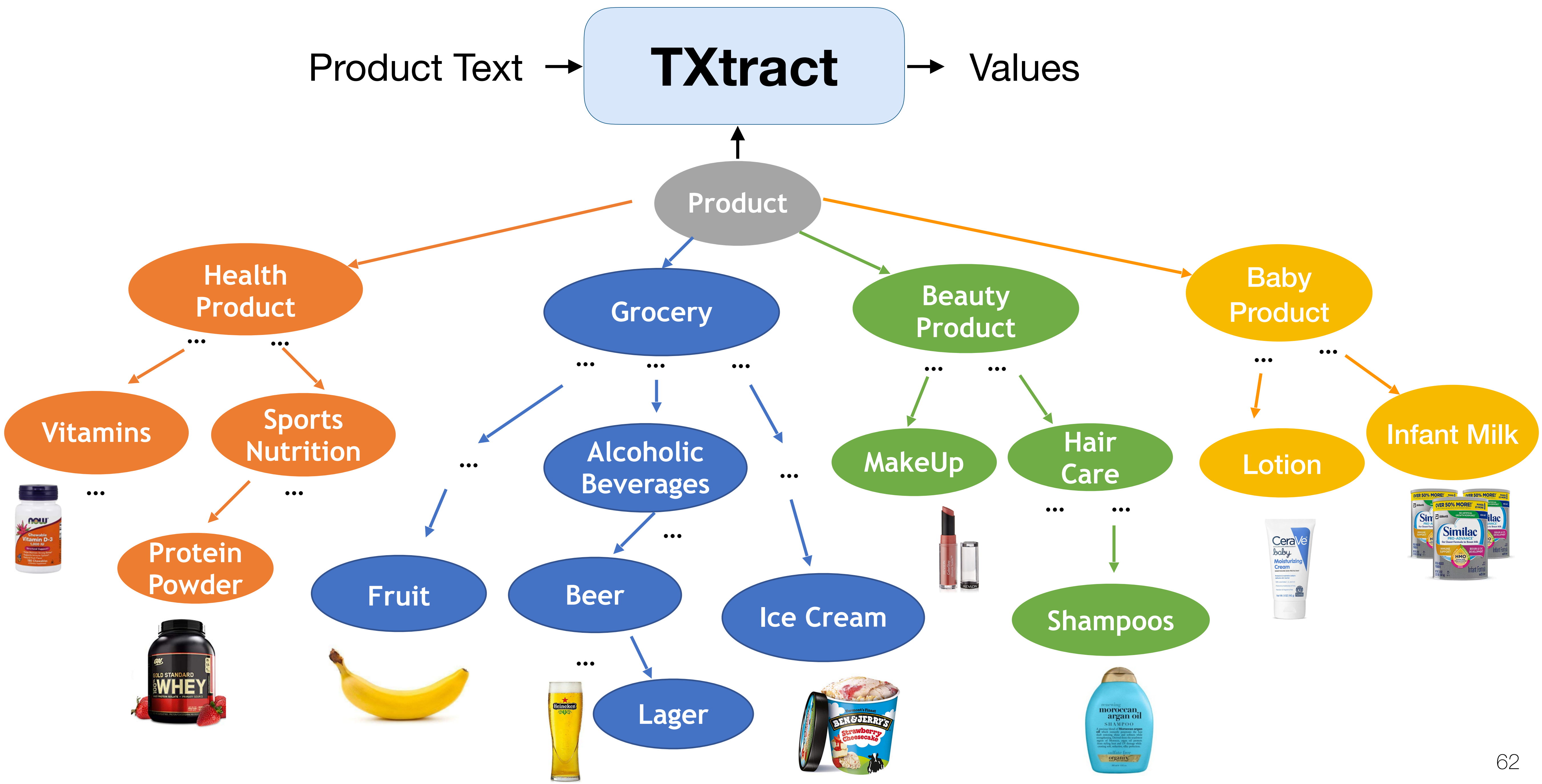
(-) not effective

Taxonomy



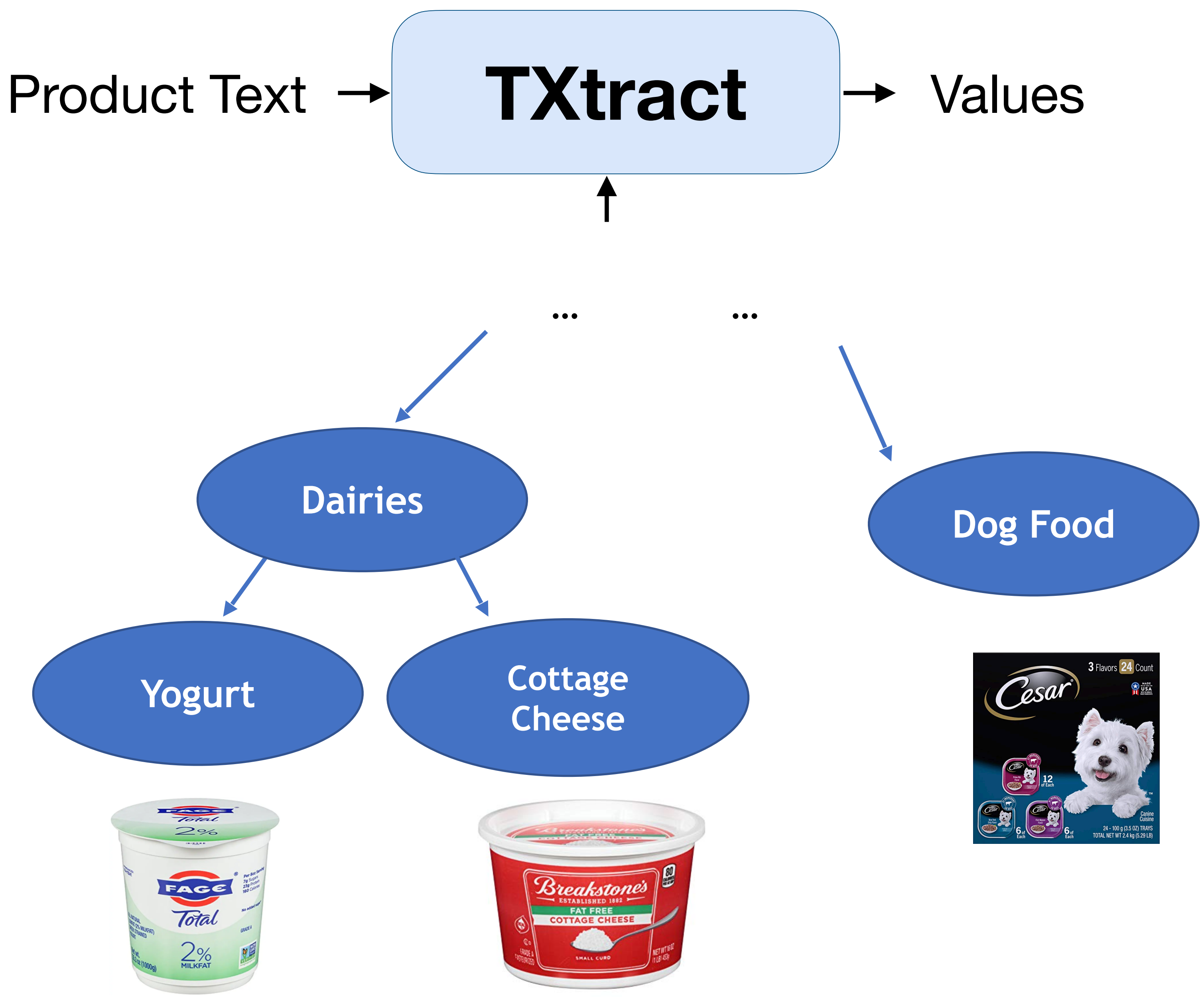
We scale up extraction to 4K product categories

"TXtract: taxonomy-aware knowledge extraction from thousands of product categories", Giannis Karamanolakis, Jun Ma, and Xin Luna Dong. (ACL 2020) amazon



We scale up extraction to 4K product categories

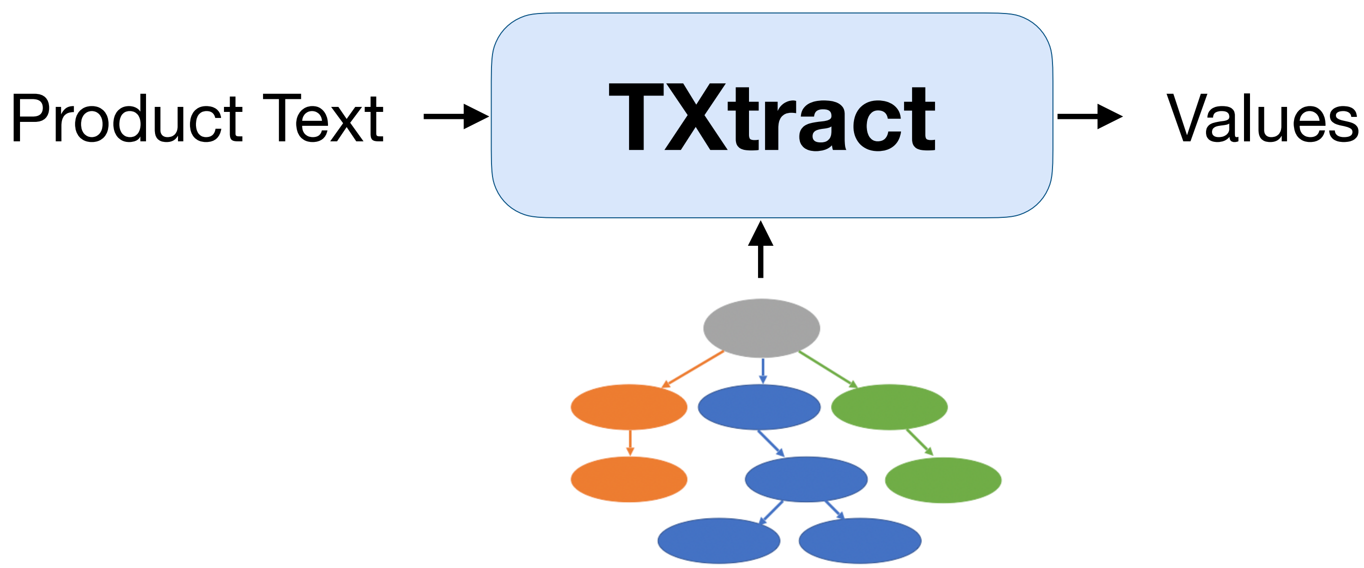
“TXtract: taxonomy-aware knowledge extraction from thousands of product categories”,
Giannis Karamanolakis, Jun Ma, and Xin Luna Dong. (ACL 2020) amazon



We scale up extraction to 4K product categories

(+) effective

extract **category-specific** values

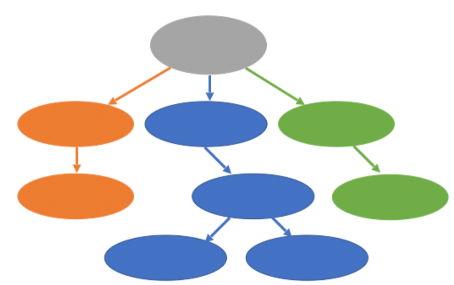


We scale up extraction to 4K product categories

(+) effective

extract **category-specific** values

Product Text → **TXtract** → Values



Digital Camera

flavor?

Not applicable

Vitamin

flavor: “fruit”

Fruit

flavor: ~~“fruit”~~

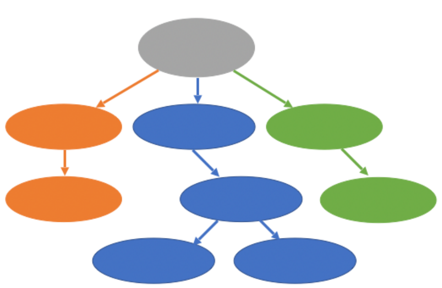
Not valid

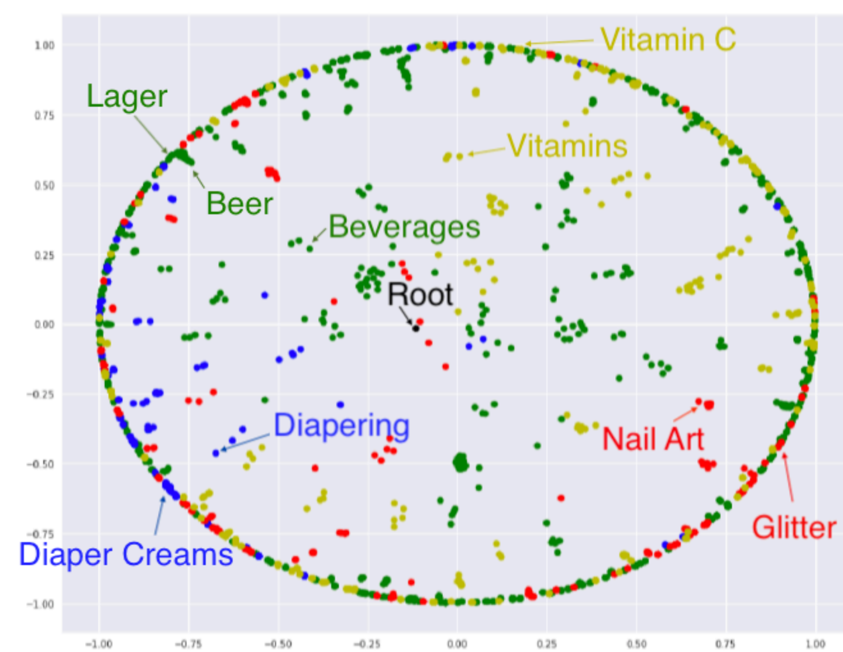
We scale up extraction to 4K product categories

(+) effective

extract **category-specific** values

Product Text → **TXtract** → Values





Category Embeddings
in Poincaré Space

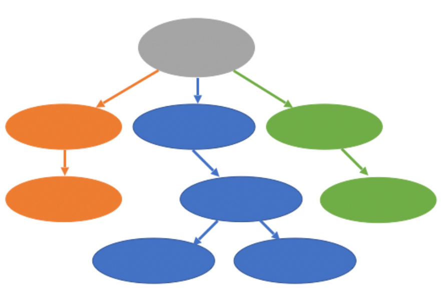
We scale up extraction to 4K product categories

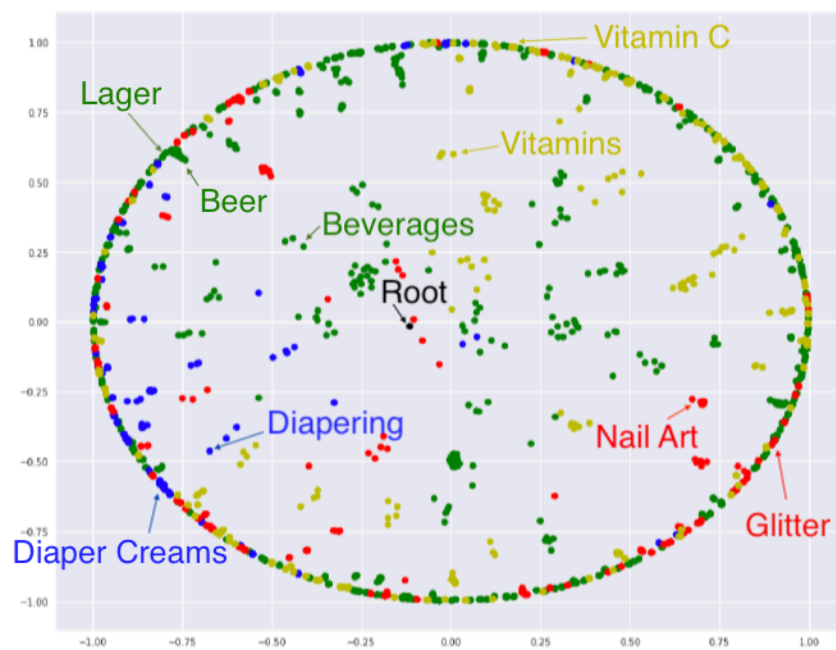
(+) effective

extract **category-specific** values

Product Text → **Conditional self-attention** → **Category-specific** Values

Product Category ↑





Category Embeddings
in Poincaré Space

We scale up extraction to 4K product categories

(+) effective

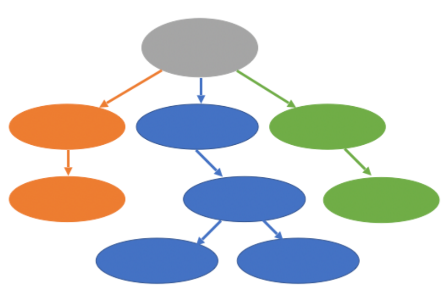
extract **category-specific** values

(+) efficient

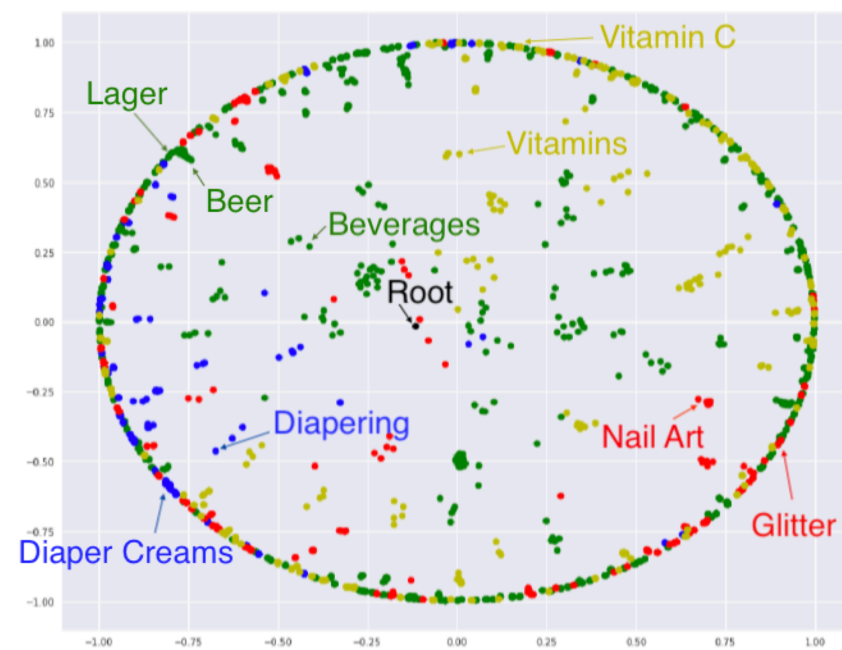
address **all** categories in a **single** model

Product Text → **Conditional self-attention** → **Category-specific** Values

Product Category ↑



→



Category Embeddings
in Poincaré Space

We scale up extraction to 4K product categories

- TXtract improves extraction performance across 4K categories

[Zheng et al., KDD'18]
Assumes "**flat**" space →
We **consider** categories →

	Coverage (%)	**F1 (%)**
OpenTag	73.0	46.6
TXtract	**81.6** (**+11.7%**)	**49.7** (**+10.4%**)

"TXtract: taxonomy-aware knowledge extraction from thousands of product categories", Giannis Karamanolakis, Jun Ma, and Xin Luna Dong. (ACL 2020)

We scale up extraction to 4K product categories

- TXtract improves extraction performance across 4K categories
- TXtract is part of the “AutoKnow” knowledge graph

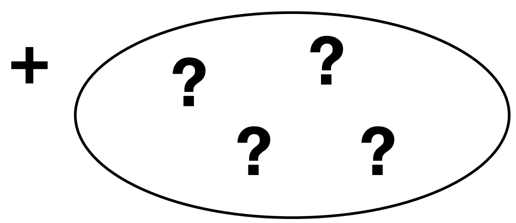
→ Amazon search
→ Amazon product detail pages

“AutoKnow: Self-Driving Knowledge Collection for Products of Thousands of Types”, Dong et al. (KDD 2020)
https://www.amazon.science/blog/building-product-graphs-automatically

“TXtract: taxonomy-aware knowledge extraction from thousands of product categories”,
Giannis Karamanolakis, Jun Ma, and Xin Luna Dong. (ACL 2020)

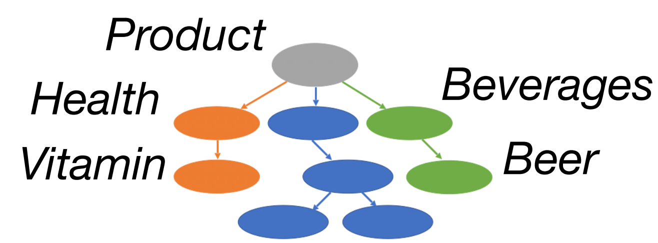
Efficient machine teaching frameworks for NLP

Coarse labels (Ch. 3)



[Karamanolakis et al. WNUT ’19]

Hierarchical taxonomies (Ch. 4)



[Karamanolakis et al. ACL ’20]

Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP ’19]

Word translations (Ch. 6)

“injured”

“ياربلانغان”

[Karamanolakis et al. Findings of EMNLP ’20]

Labeling rules (Ch. 7, 8)

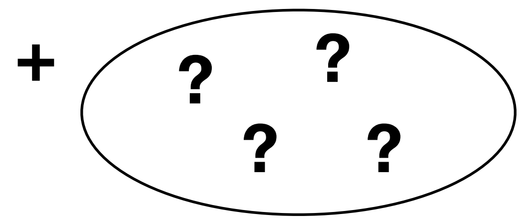
```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

[Karamanolakis et al. NAACL ’21] + work with Daniel Hsu and Luis Gravano

Efficient machine teaching frameworks for NLP

Can we teach ML models
without any ground-truth (coarse or fine) labels?

Coarse labels (Ch. 3)



[Karamanolakis et al. WNUT ’19]

Hierarchical taxonomies (Ch. 4)

[Karamanolakis et al. ACL ’20]

Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP ’19]

Word translations (Ch. 6)

“injured”

“يارىلانغان”

[Karamanolakis et al. Findings of EMNLP ’20]

Labeling rules (Ch. 7, 8)

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

[Karamanolakis et al. NAACL ’21] + work with Daniel Hsu and Luis Gravano

Efficient machine teaching frameworks for NLP

Coarse labels (Ch. 3)

\+ ? ? ? ?

[Karamanolakis et al. WNUT ’19]

Hierarchical taxonomies (Ch. 4)

[Karamanolakis et al. ACL ’20]

Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP ’19]

Word translations (Ch. 6)

“injured” “ياربلانغان”

[Karamanolakis et al. Findings of EMNLP ’20]

Labeling rules (Ch. 7, 8)

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

[Karamanolakis et al. NAACL ’21] + work with Daniel Hsu and Luis Gravano

Fine-grained aspect detection in online reviews

- Online reviews discuss various **aspects** of the entities

amazon prime

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
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.

TV Aspects

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

yelp

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 we
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.

Restaurant Aspects

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

Fine-grained aspect detection in online reviews

- Online reviews discuss various **aspects** of the entities
- **Task:** Classify individual sentences to pre-defined aspects

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**

Fine-grained aspect detection in online reviews

- Online reviews discuss various **aspects** of the entities
- **Task:** Classify individual sentences to pre-defined aspects

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**

- **Goal:** train neural networks **without** any ground-truth labels (coarse or fine)

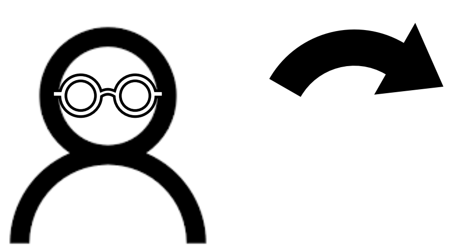
Fine-grained aspect detection in online reviews

- Online reviews discuss various **aspects** of the entities
- **Task:** Classify individual sentences to pre-defined aspects

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**

- **Goal:** train neural networks **without** any ground-truth labels (coarse or fine)

- We assume a small set of indicative **seed words** per aspect



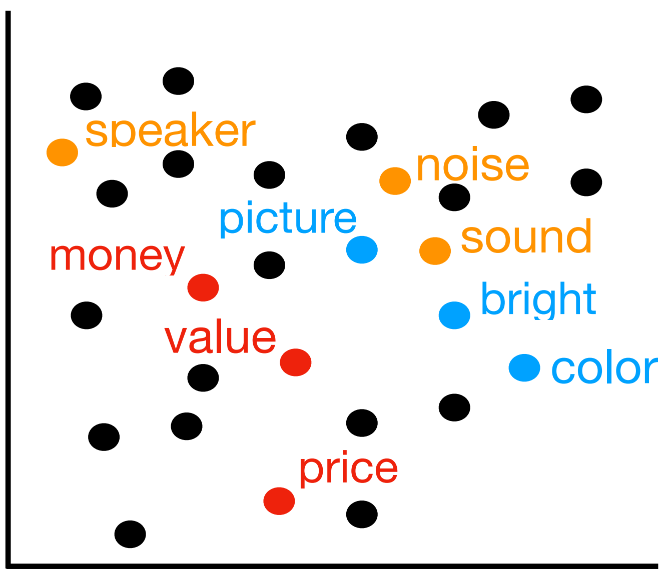
Aspect	Seed Words
Price	price, value, money, worth, paid
Image	picture, color, quality, black, bright
Sound	sound, speaker, noise, loud, volume

How to leverage seed words for neural networks?

How to leverage seed words for neural networks?

- **Previous approaches:** use seed words for initialization

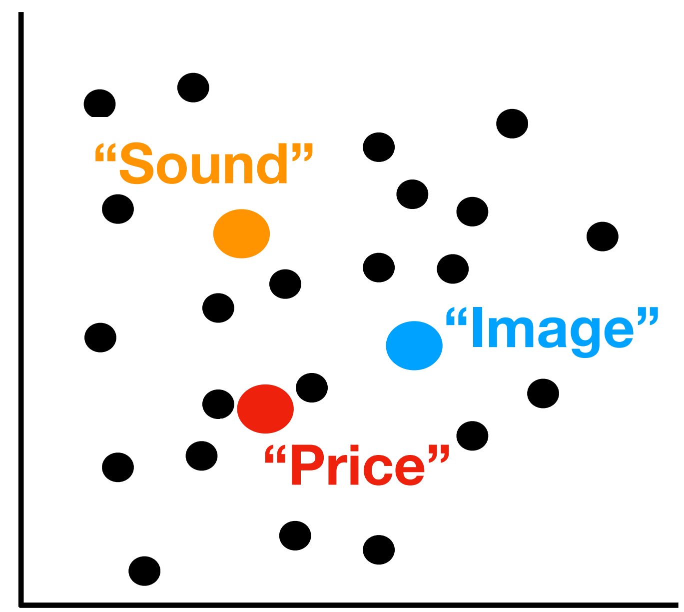
[Lund et al., 2017]
[Angelidis and Lapata, 2018]



Pre-Trained Word Embeddings

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

unweighted / weighted average

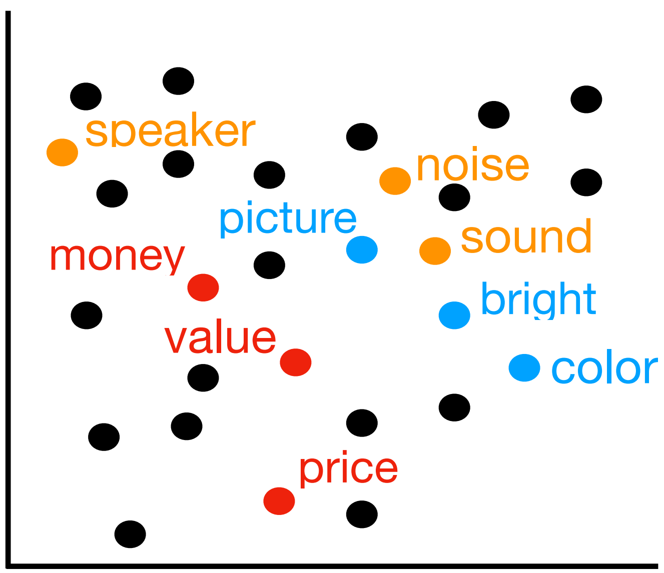


Initial Aspect Embeddings

How to leverage seed words for neural networks?

- **Previous approaches:** use seed words for initialization

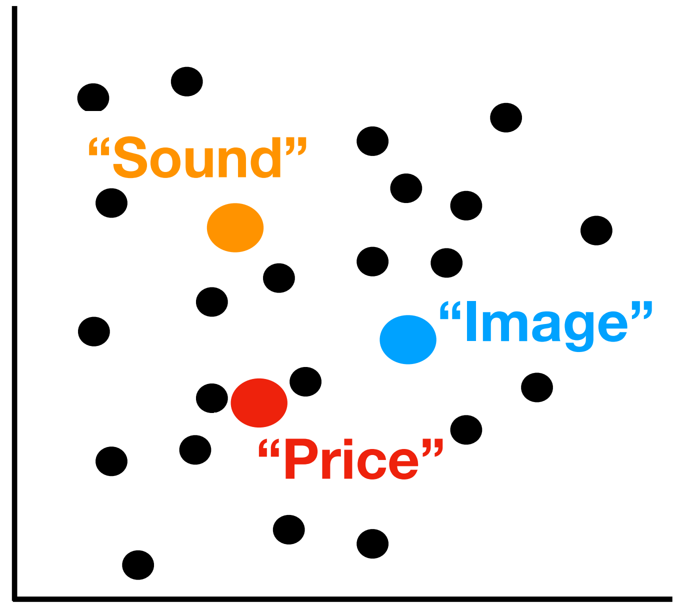
[Lund et al., 2017]
[Angelidis and Lapata, 2018]



Pre-Trained Word Embeddings

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

unweighted / weighted average



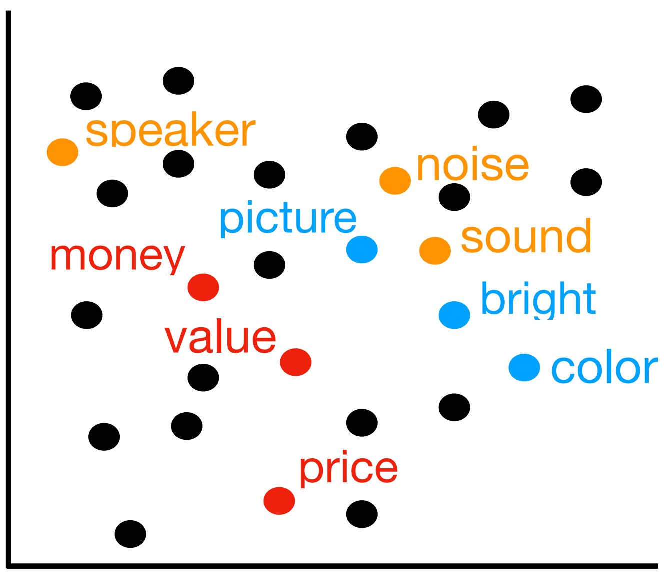
Initial Aspect Embeddings

Train Unsupervised Model

How to leverage seed words for neural networks?

- **Previous approaches:** use seed words for initialization

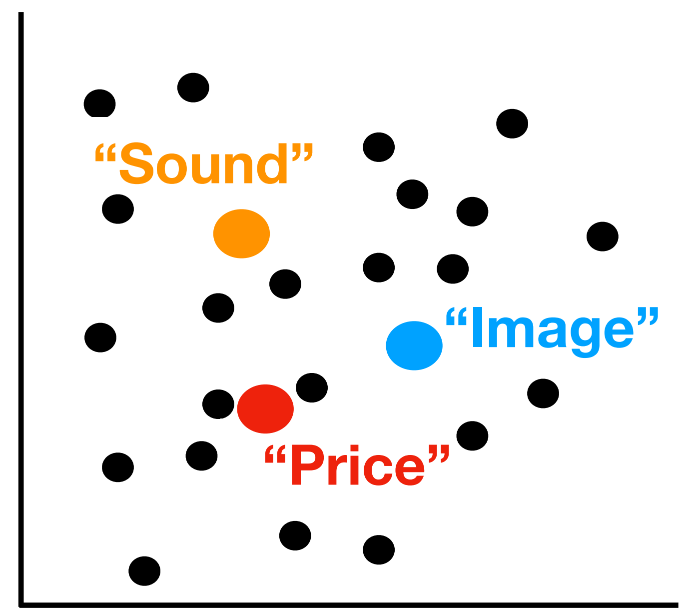
[Lund et al., 2017]
[Angelidis and Lapata, 2018]



Pre-Trained Word Embeddings

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

unweighted / weighted average



Initial Aspect Embeddings

Train Unsupervised Model

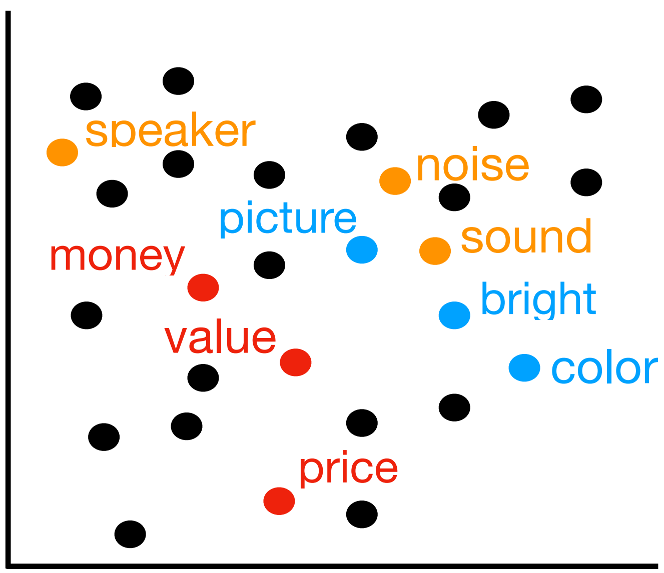
- **Limitations of previous approaches:**

  (-) Individual seed words are **not** used during training

How to leverage seed words for neural networks?

- **Previous approaches:** use seed words for initialization

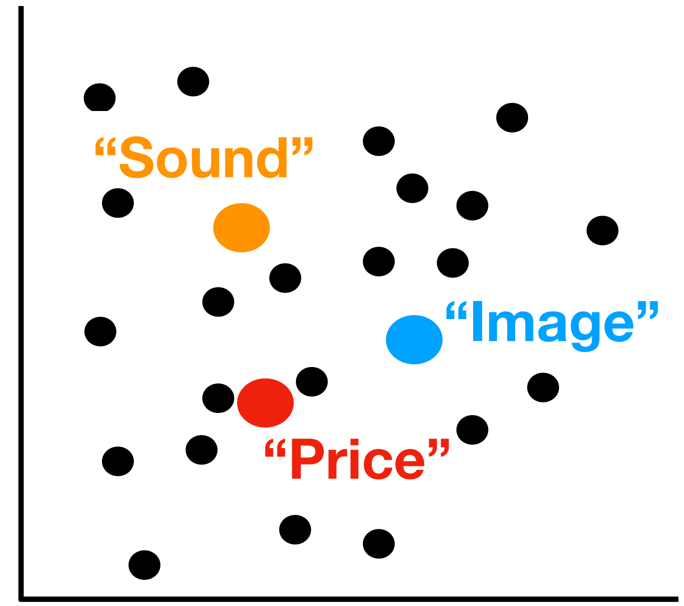
[Lund et al., 2017]
[Angelidis and Lapata, 2018]



Pre-Trained Word Embeddings

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

unweighted / weighted average



Initial Aspect Embeddings

Train Unsupervised Model

- **Limitations of previous approaches:**

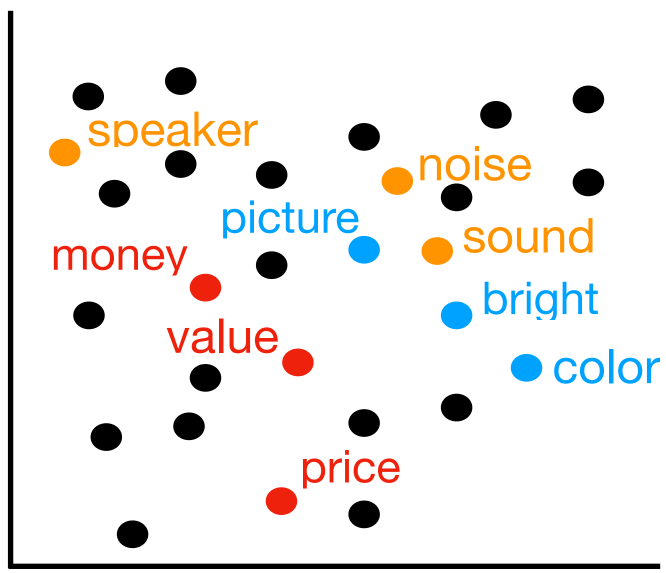
  (-) Individual seed words are **not** used during training

  (-) Aspect embedding quality is **sensitive** to word embedding quality

How to leverage seed words for neural networks?

- **Previous approaches:** use seed words for initialization

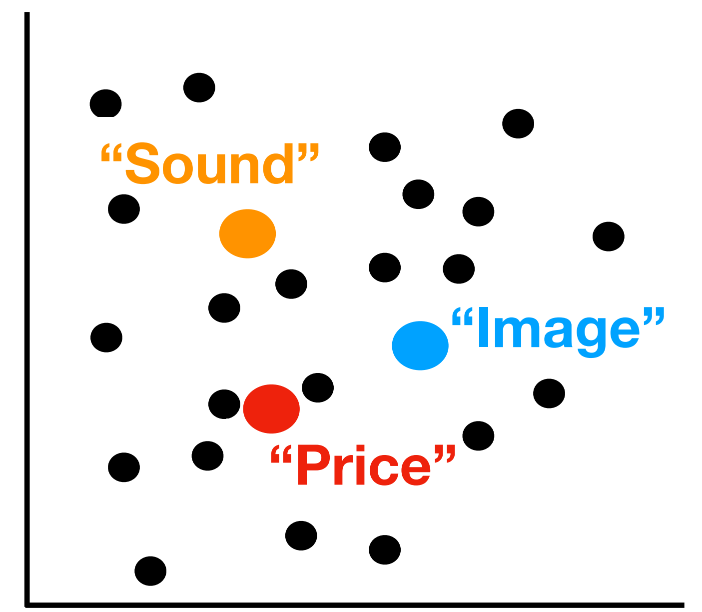
[Lund et al., 2017]
[Angelidis and Lapata, 2018]



Pre-Trained Word Embeddings

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

unweighted / weighted average



Initial Aspect Embeddings

Train Unsupervised Model

- **Limitations of previous approaches:**
  - **(-)** Individual seed words are **not** used during training
  - **(-)** Aspect embedding quality is **sensitive** to word embedding quality
  - **(-)** Aspect embedding fine-tuning risks from **diverging** from pre-defined aspects

How to leverage seed words for neural networks?

- **Previous approaches:** use seed words for initialization [Lund et al., 2017] [Angelidis and Lapata, 2018]
- **Our approach:** weakly-supervised co-training [Karamanolakis, Hsu, Gravano, EMNLP ’19]

We explore the **predictive** power
of each **individual** seed word
during training

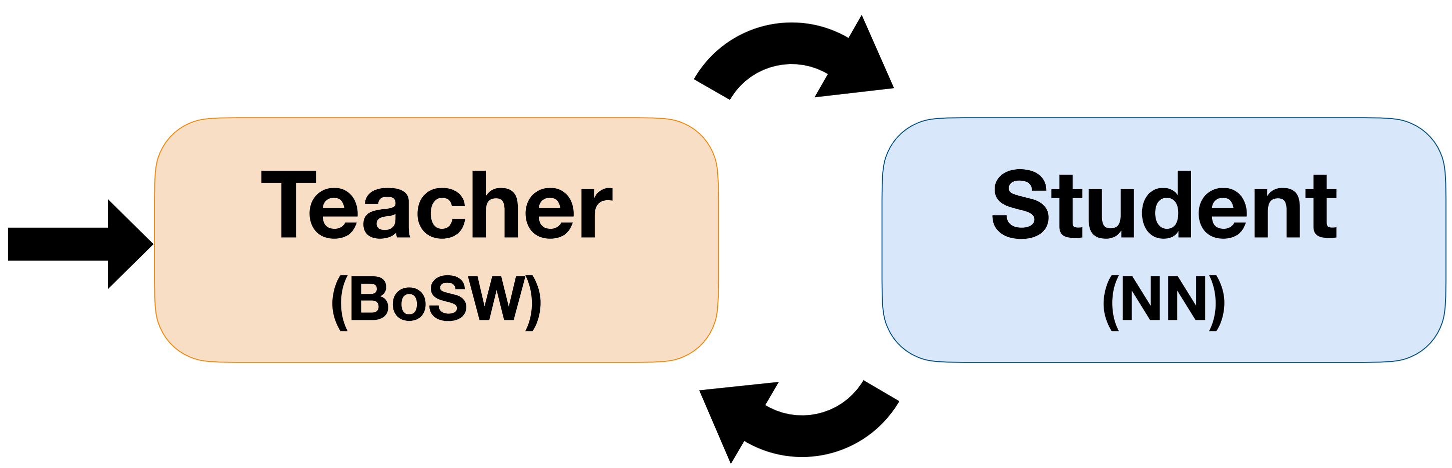
“Leveraging just a few keywords for fine-grained aspect detection through weakly-supervised co-training”
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (EMNLP 2019)

How to leverage seed words for neural networks?

- **Previous approaches:** use seed words for initialization [Lund et al., 2017] [Angelidis and Lapata, 2018]
- **Our approach:** weakly-supervised co-training [Karamanolakis, Hsu, Gravano, EMNLP ’19]

+ Leverage **seed words** in a bag-of-seed-words classifier (**Teacher**)
+ Use **Teacher**’s predictions to train a neural network (**Student**)
+ Adaptively cope with **noisy** seed words through **iterative co-training**

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

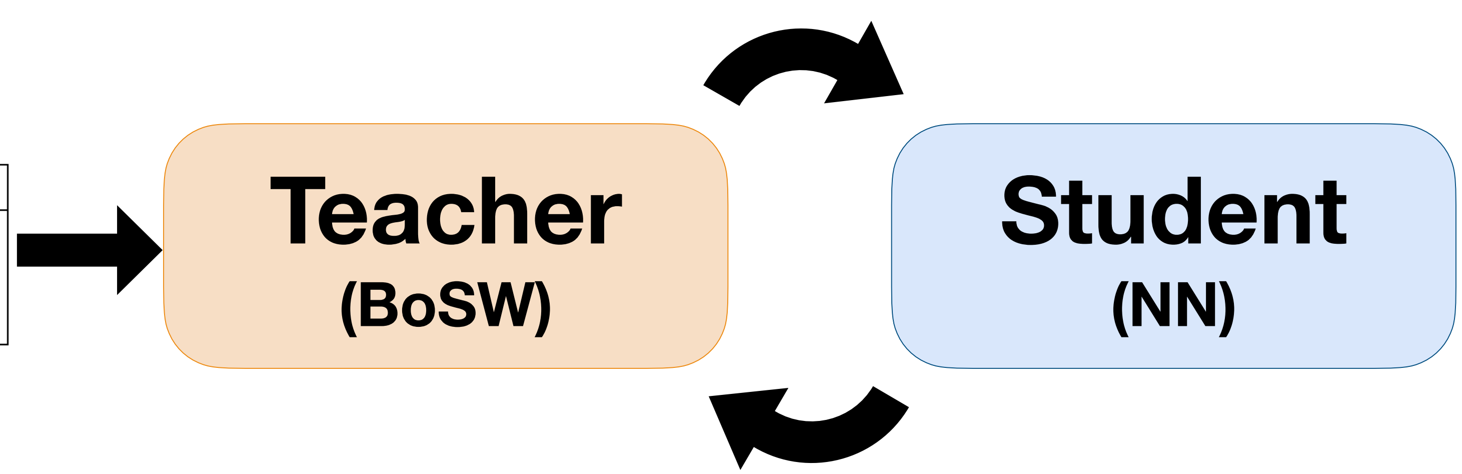


“Leveraging just a few keywords for fine-grained aspect detection through weakly-supervised co-training”
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (EMNLP 2019)

Weakly supervised co-training framework

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

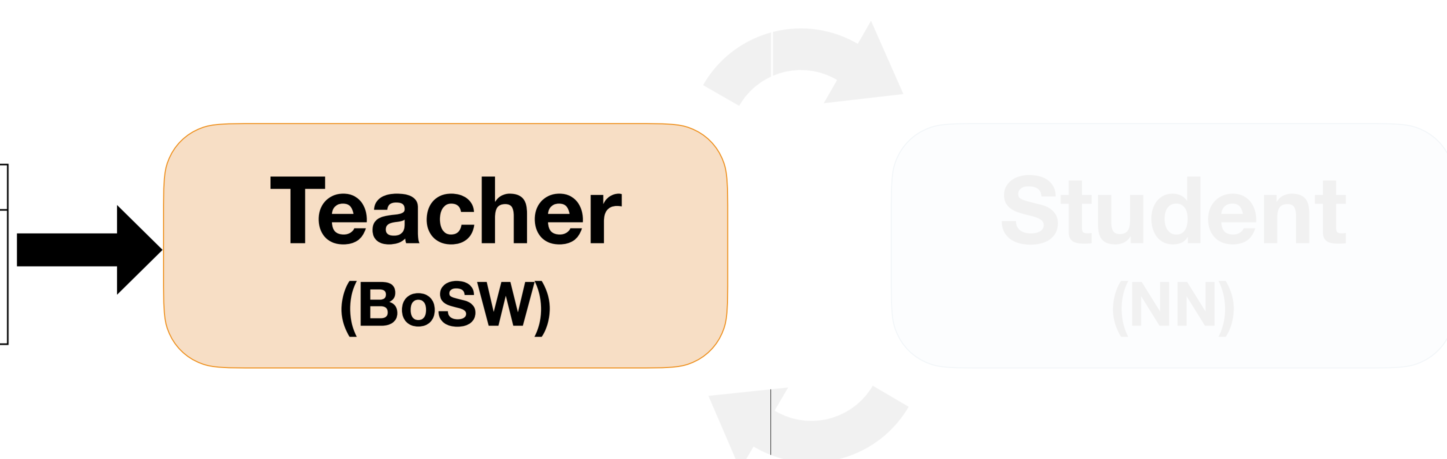
Aspect	Seed Words
Price	price, value, money, ...
Image	picture, color, bright, ...
Sound	sound, speaker, noise, ...



Weakly supervised co-training framework

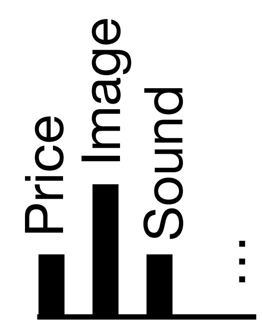
1. **Teacher**
2. Student
3. Teacher -> Student
4. Student -> Teacher

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



Teacher: leverages seed words to predict aspects

- Teacher: **Bag-of-Seed-Words (BoSW)** classifier



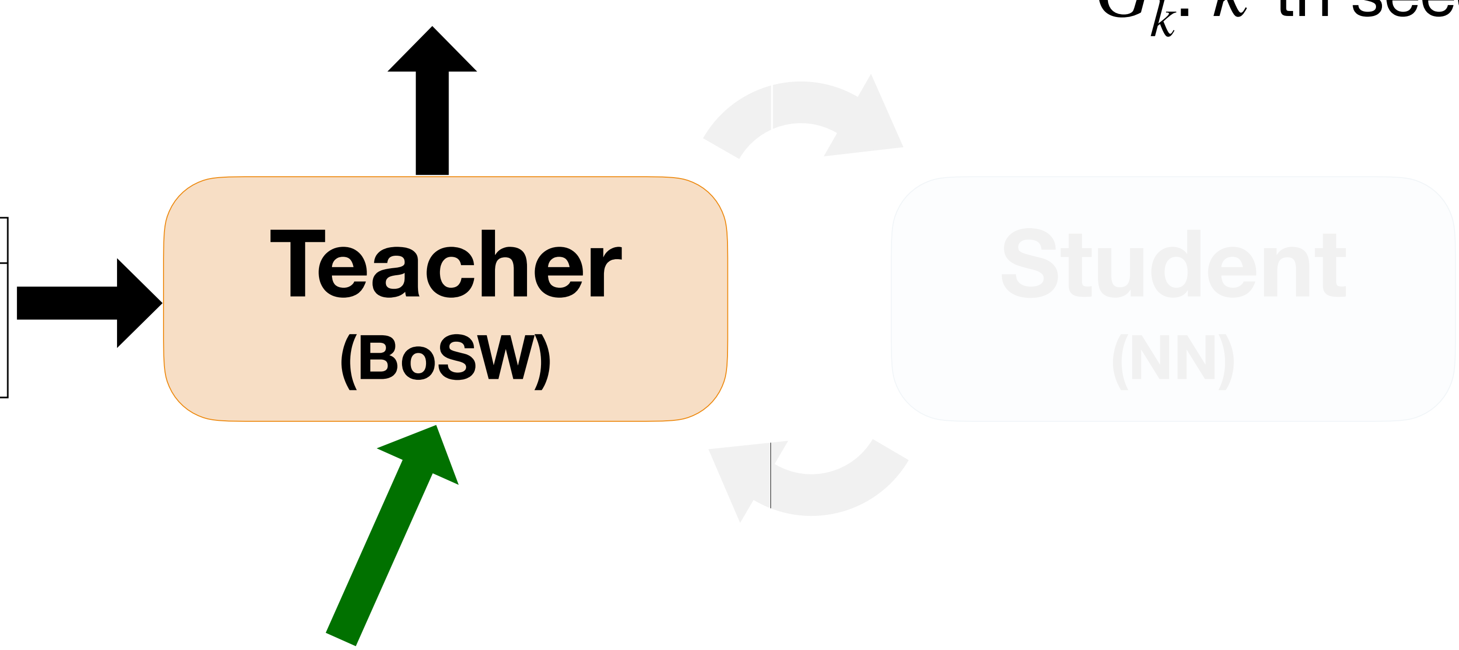
aspect probabilities $\langle q^1, \ldots, q^K \rangle$

$$q^k = \frac{\exp(\sum_j \mathbb{1}\{j \in G_k\} \cdot c^j)}{\sum_{k'} \exp(\sum_j \mathbb{1}\{j \in G_{k'}\} \cdot c^j)}$$

c^j: count of seed word j

G_k: k-th seed set

Aspect	Seed Words
Price	price, value, money, ...
Image	$\boxed{\text{picture,}}$ color, bright, ...
Sound	sound, speaker, noise, ...



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

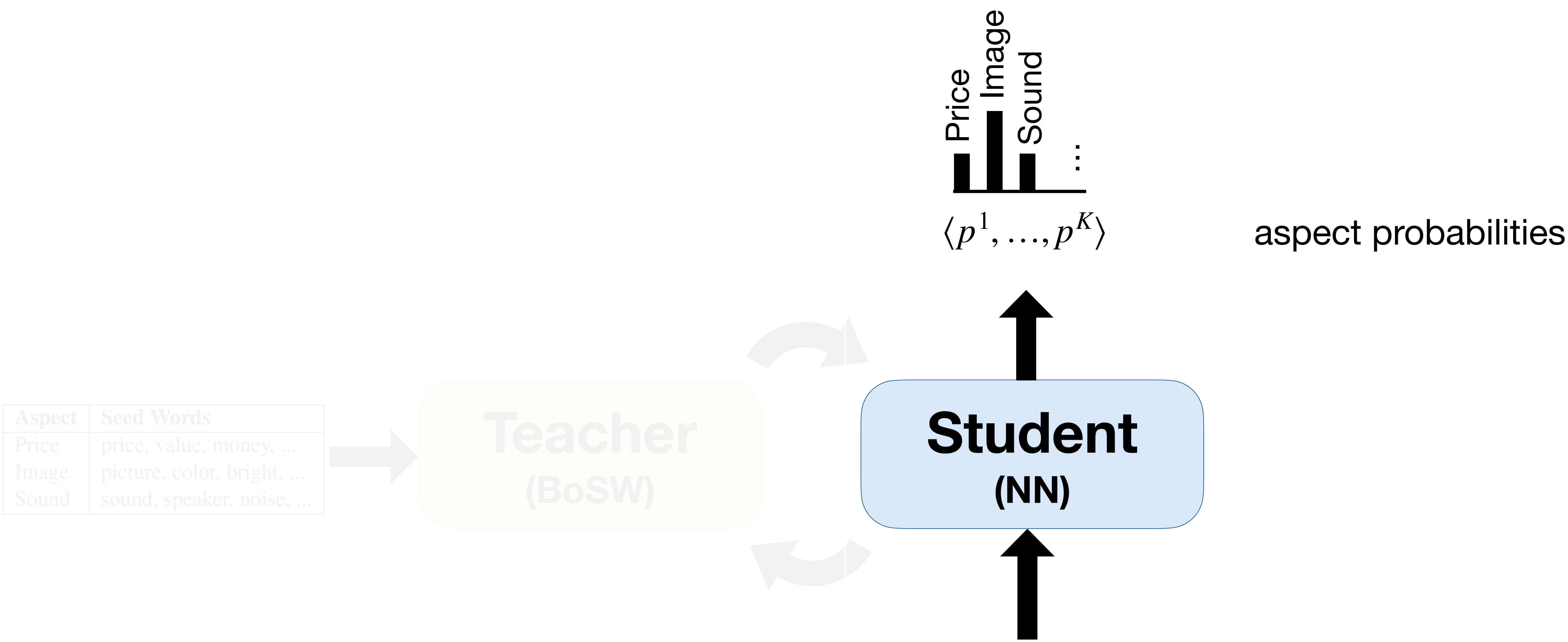
Weakly supervised co-training

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

Student: leverages seed words and non-seed words

- **Student:** an embedding-based neural network

Uses **seed words** and **non-seed words** (context) to predict aspects

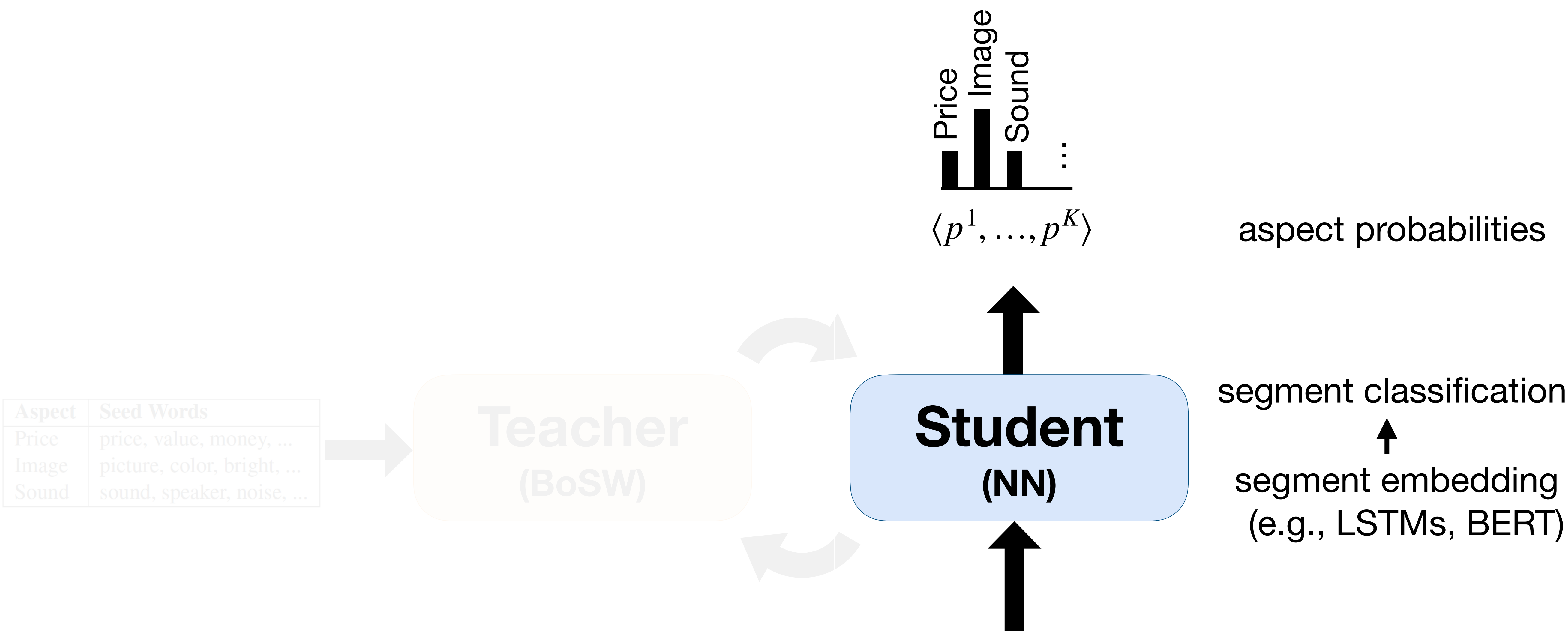


segment: *"The **picture** looks very pixelated when playing blu-ray movies"*

Student: leverages seed words and non-seed words

- **Student:** an embedding-based neural network

Uses **seed words** and **non-seed words** (context) to predict aspects

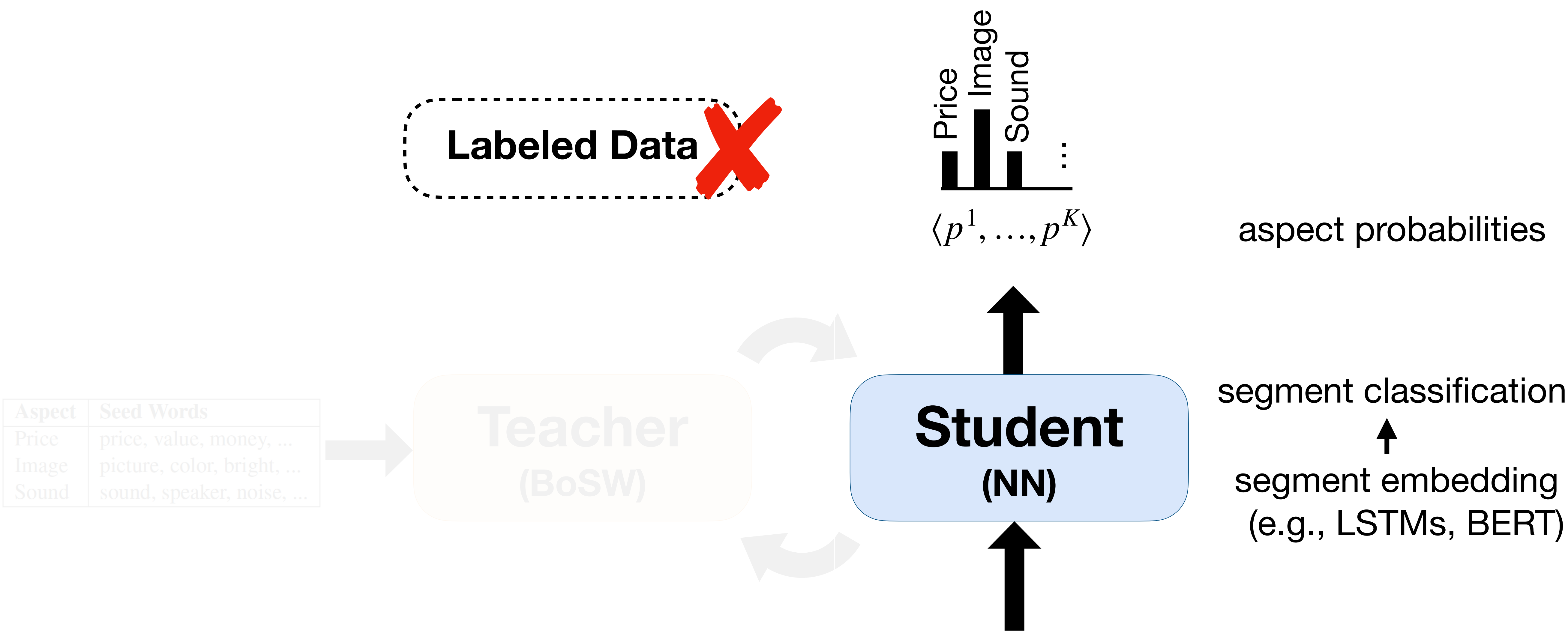


segment: “*The **picture** looks very pixelated when playing blu-ray movies*”

Student: leverages seed words and non-seed words

- **Student:** an embedding-based neural network

Uses **seed words** and **non-seed words** (context) to predict aspects

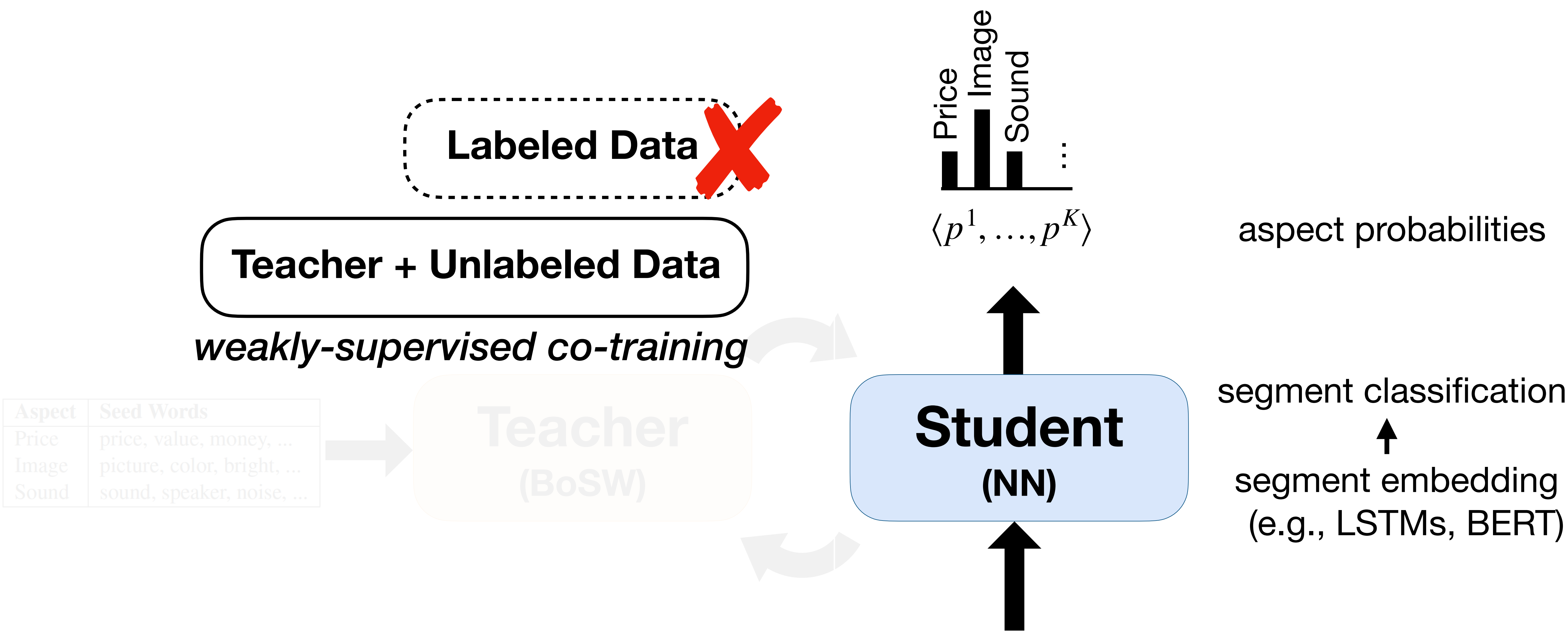


segment: *"The **picture** looks very pixelated when playing blu-ray movies"*

Student: leverages seed words and non-seed words

- **Student:** an embedding-based neural network

Uses **seed words** and **non-seed words** (context) to predict aspects



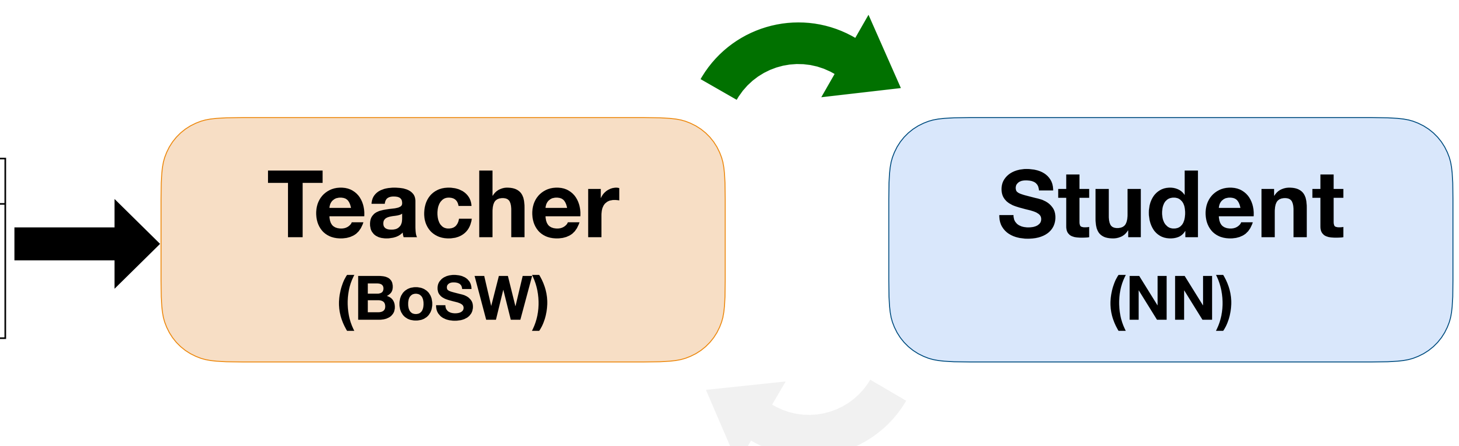
segment: *"The **picture** looks very pixelated when playing blu-ray movies"*

Weakly supervised co-training

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

apply Teacher on **unlabeled** data to get weak labels

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

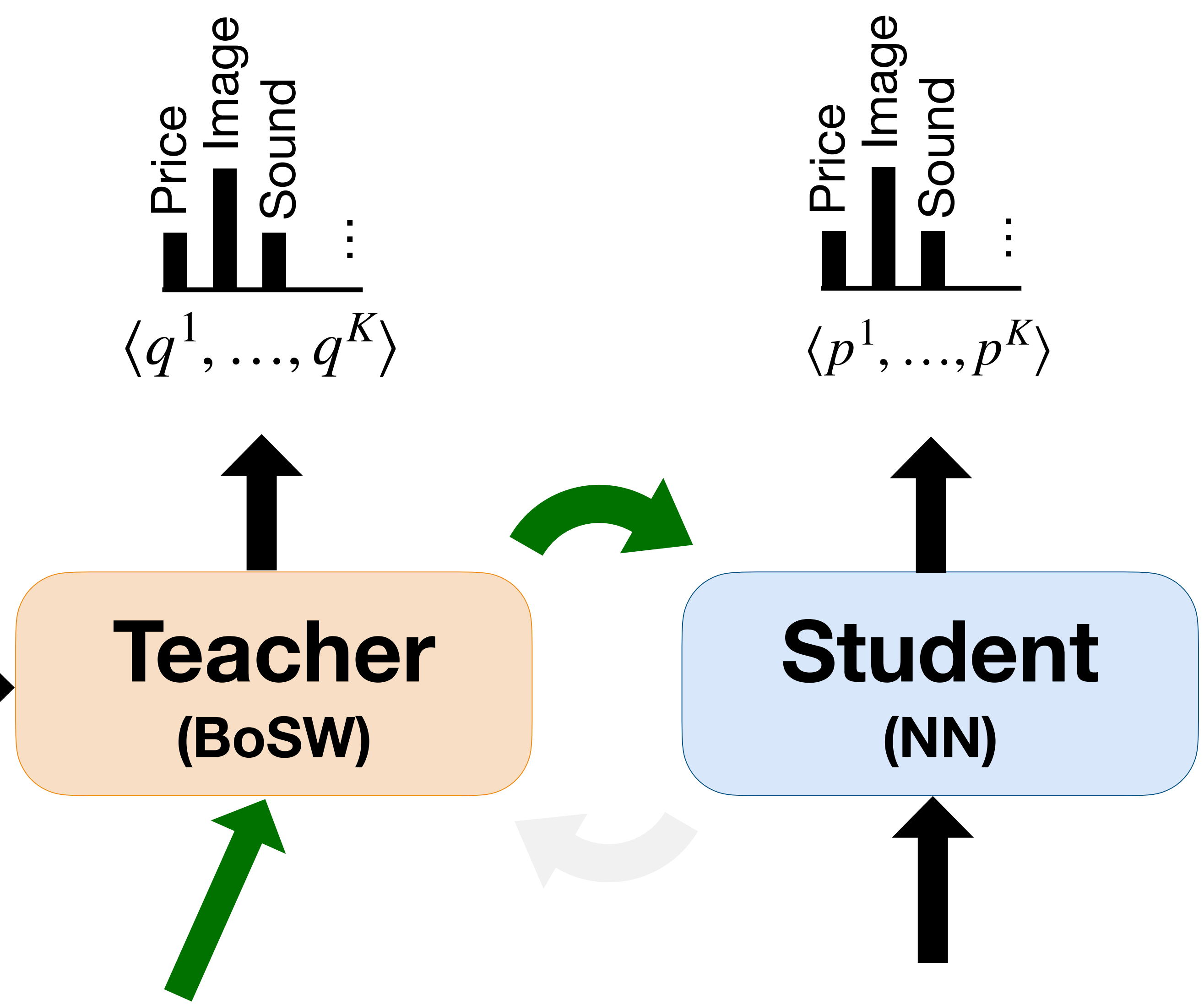


Training Student using Teacher's predictions

- **Student** is trained through the "distillation" objective [Ba & Caruana, 2014; Hinton et al 2015]

$$H(q_i, p_i) = -\sum_k q_i^k \log p_i^k$$

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

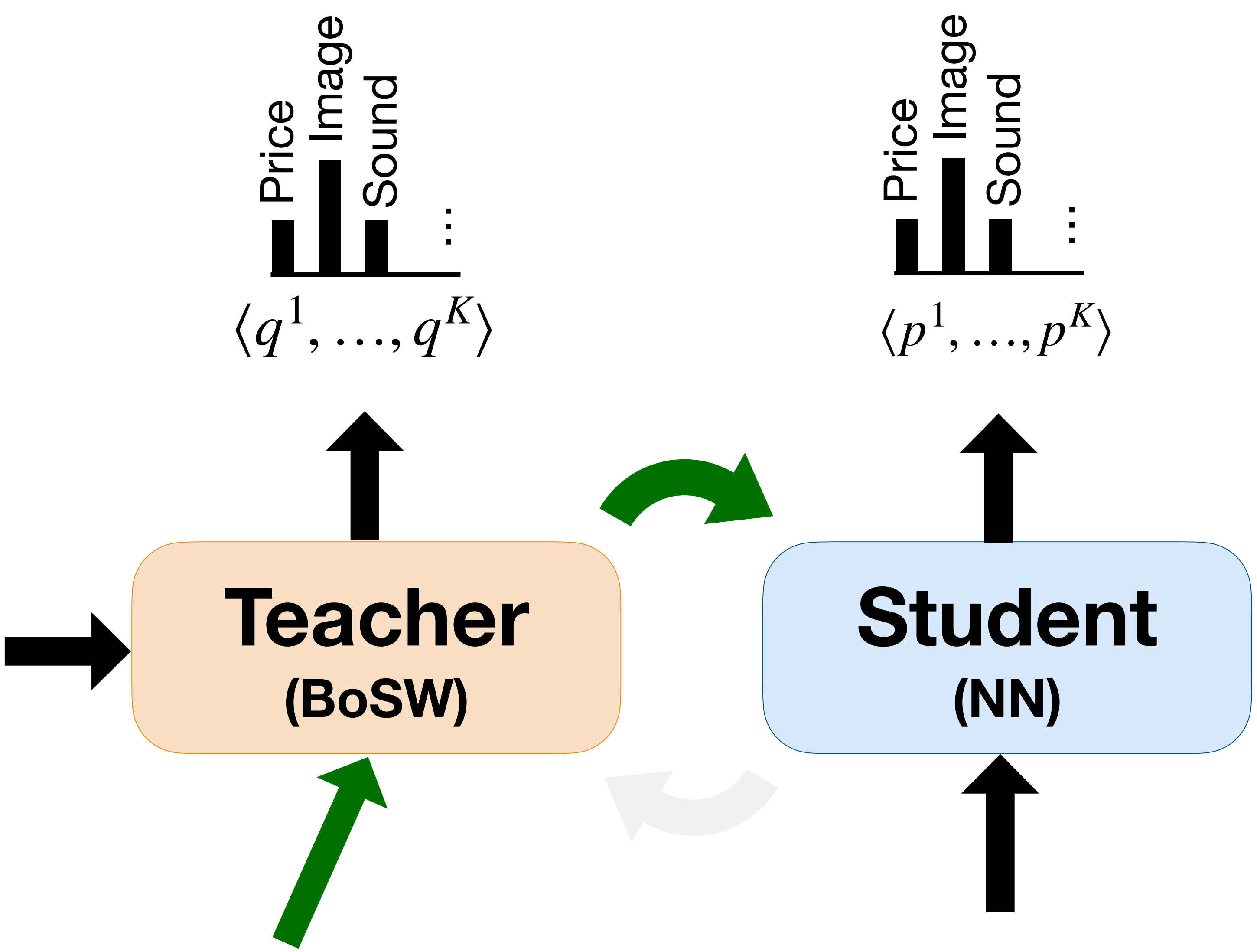


segment: *"The **picture** looks very pixelated when playing blu-ray movies"*

Training Student using Teacher's predictions

- **Student** is trained through the "distillation" objective [Ba & Caruana, 2014; Hinton et al 2015]
- **Student** also considers the **context** of **seed words**

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

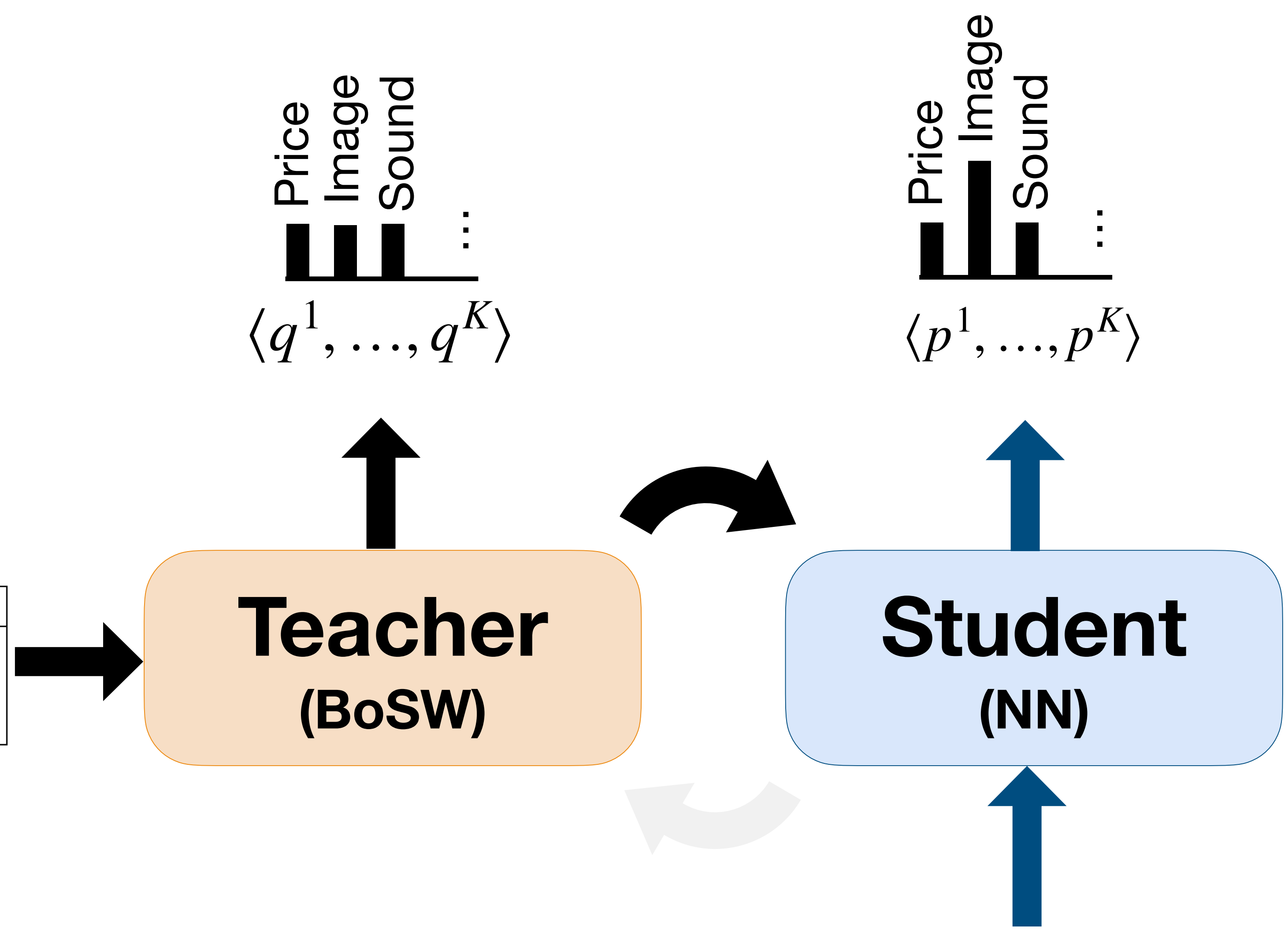


segment: "*The **picture** looks very pixelated when playing blu-ray movies*"

Training Student using Teacher's predictions

- **Student** is trained through the "distillation" objective [Ba & Caruana, 2014; Hinton et al 2015]
- **Student** also considers the **context** of **seed words**
- **Student** predicts aspects even if no **seed words** appear

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

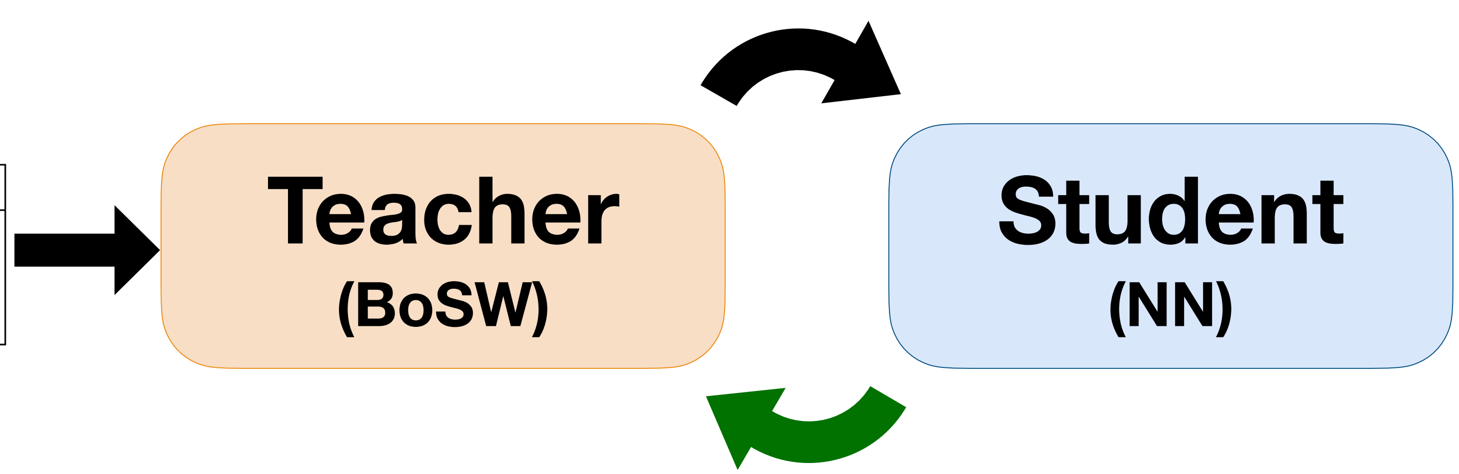


segment: *"The <UNK> looks very **pixelated** when playing **blu-ray movies**"*

Weakly supervised co-training

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

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



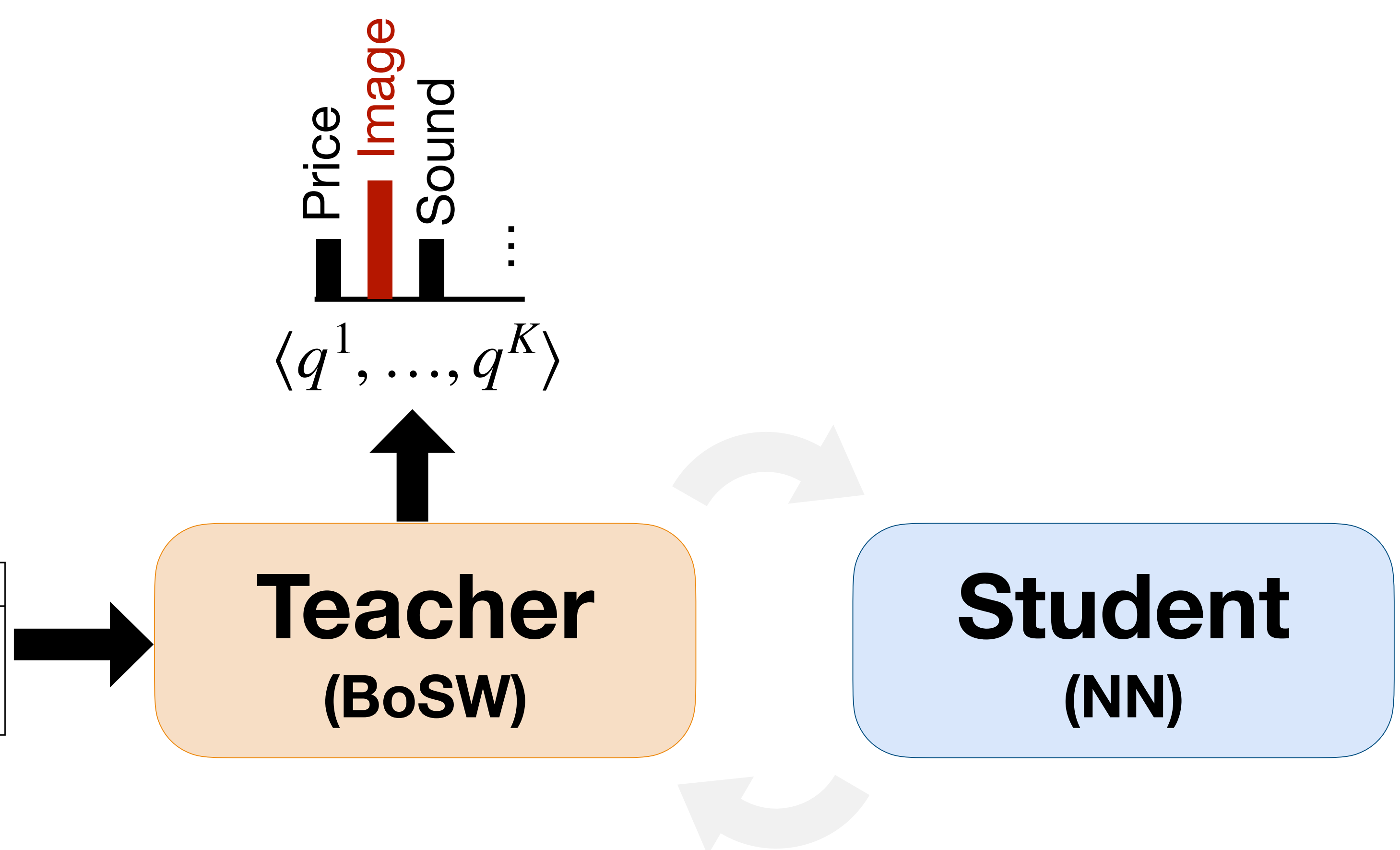
Weakly supervised co-training

- Seed words may be **noisy**: only weakly indicative of aspects

Qualities	Price	Image	Sound
bright	**0%**	**100%**	**0%**

“bright” should also be indicative of **“Sound”** aspect

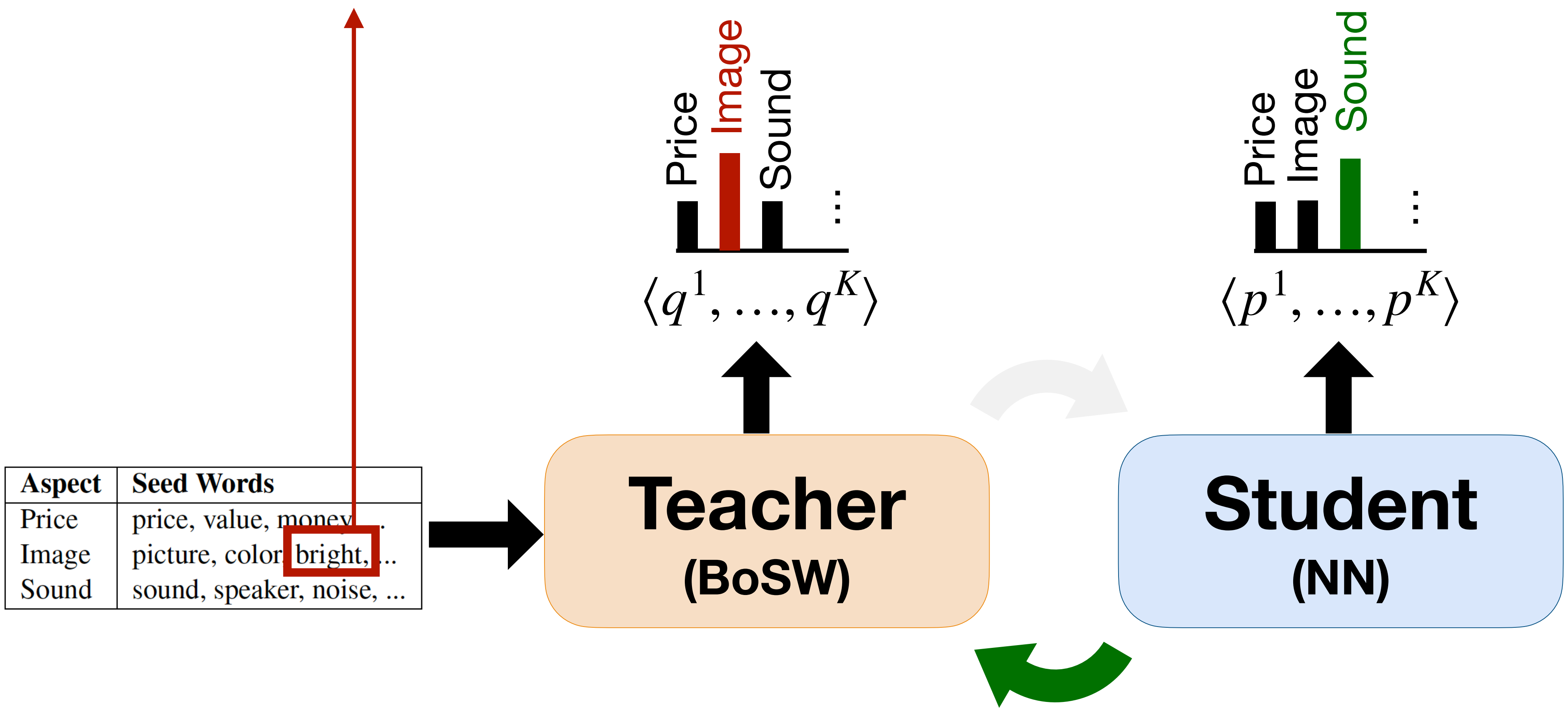
Aspect	Seed Words
Price	price, value, money, ...
Image	picture, color, bright, ...
Sound	sound, speaker, noise, ...



Weakly supervised co-training

- Seed words may be **noisy**: only weakly indicative of aspects
- **Teacher** estimates seed word **qualities** … using **Student**'s predictions

Qualities	Price	Image	Sound
bright	0%	100%	0%

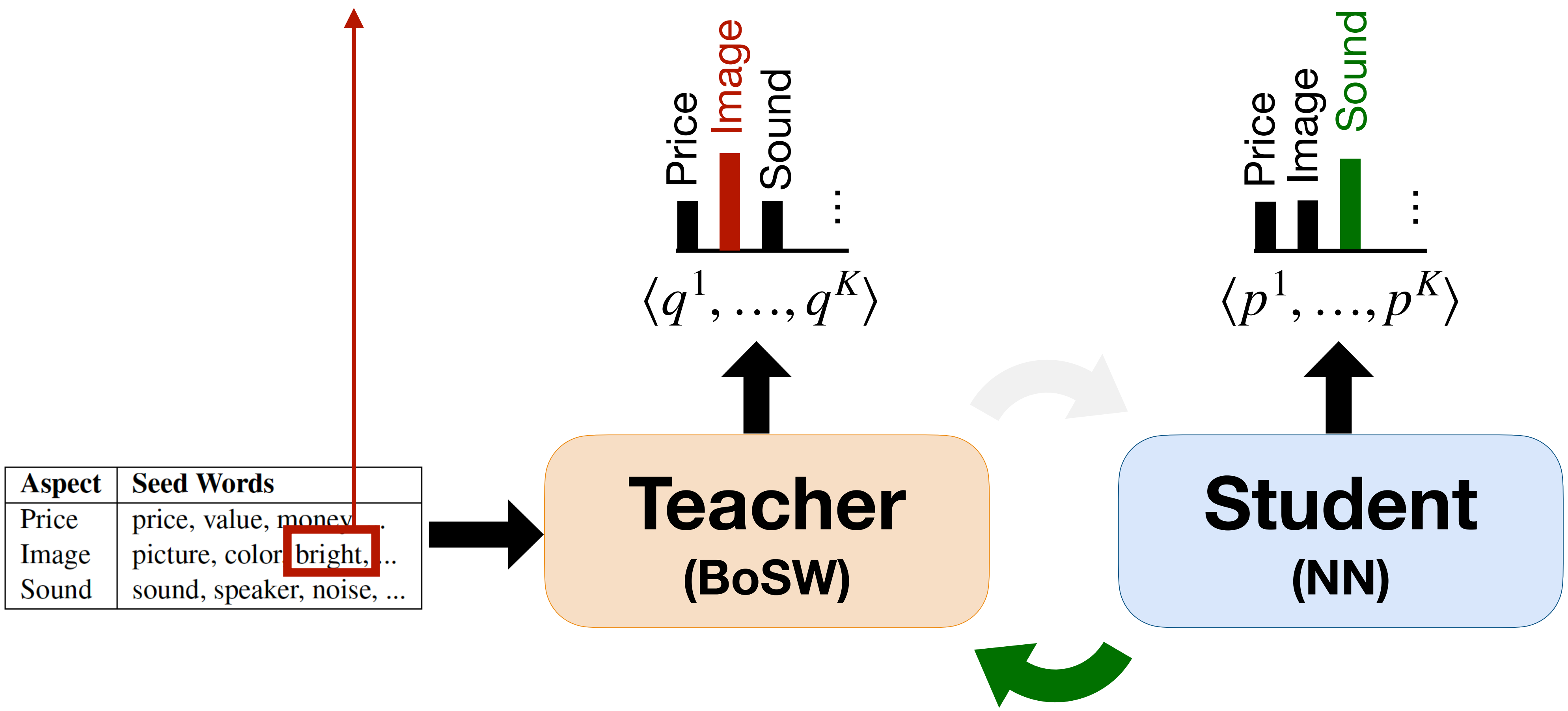


Weakly supervised co-training

- Seed words may be **noisy**: only weakly indicative of aspects
- **Teacher** estimates seed word **qualities** … using **Student**'s predictions

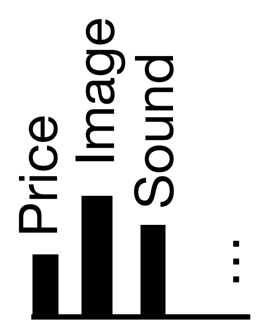
Qualities	**Price**	**Image**	**Sound**
~~**bright**~~	~~**0%**~~	~~**100%**~~	~~**0%**~~
bright	**0%**	**70%**	**30%**

Model Bootstrapped EM



Weakly supervised co-training

- **Improved Teacher:** uses **seed words** and **quality estimates** to predict aspects



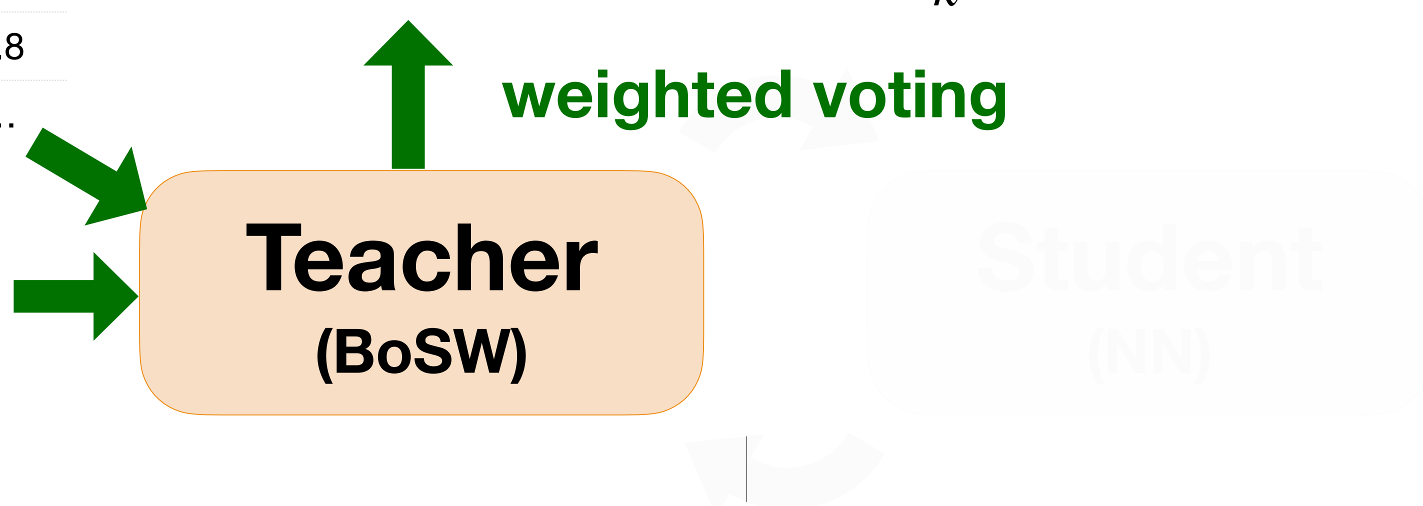
$$\langle q^1, \ldots, q^K \rangle$$

$$q^k = \frac{\exp \sum_j \mathbb{1}\{j \in G_k\} \cdot \hat{z}_j^k \cdot c^j}{\sum_{k'} \exp \sum_j \mathbb{1}\{j \in G_{k'}\} \cdot \hat{z}_j^{k'} \cdot c^j}$$

z_k^j: weight of seed word j for class k

Weight $\hat{z}$	Price	Image	Sound
bright	0.0	0.7	0.3
noise	0.0	0.2	0.8
...	...	...	...

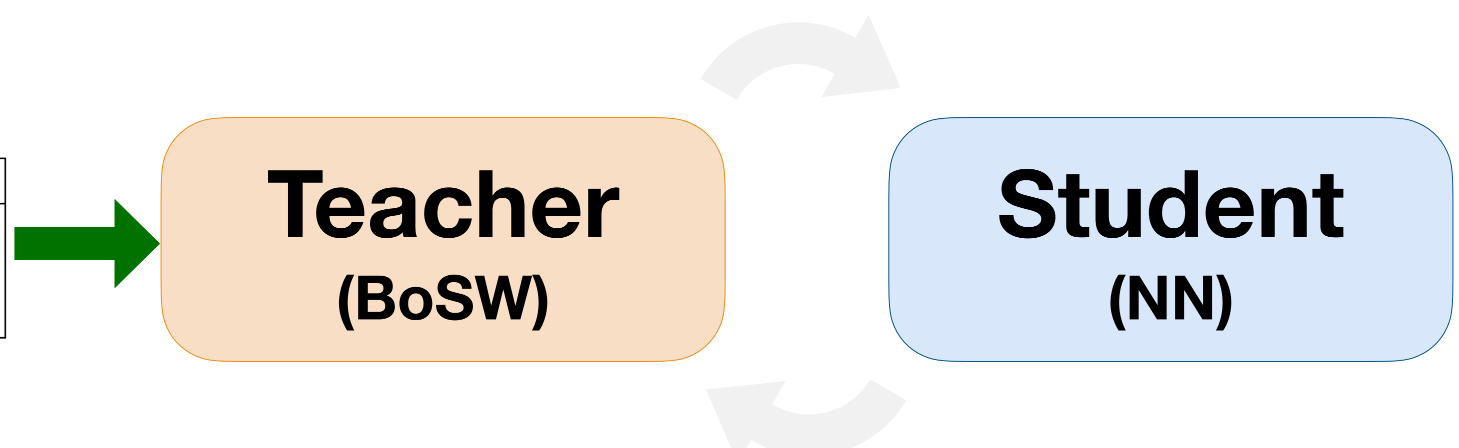
Aspect	Seed Words
Price	price, value, money, ...
Image	picture, color, bright, ...
Sound	sound, speaker, noise, ...



Iterative co-training pipeline

- **T0:** Apply **Teacher** on unlabeled data (assuming “perfect” seed words)
- **S0:** Train **Student** using **Teacher**’s predictions
- **T1:** Update **Teacher**’s weights (seed word qualities) using **Student**’s predictions
- **S1:** … Iterate until convergence

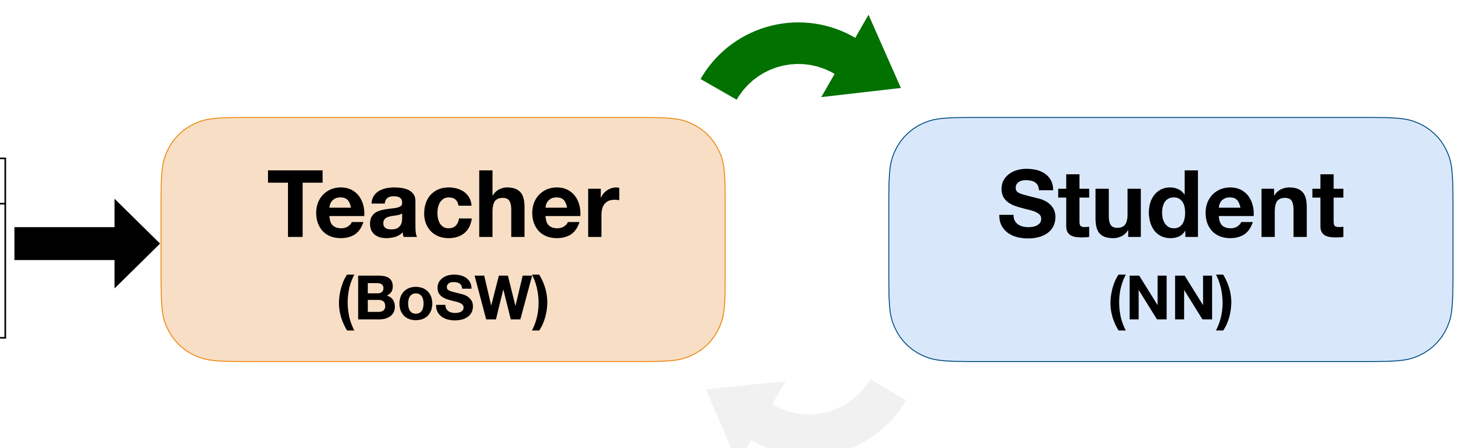
Aspect	Seed Words
Price	price, value, money, ...
Image	picture, color, bright, ...
Sound	sound, speaker, noise, ...



Iterative co-training pipeline

- **T0:** Apply **Teacher** on unlabeled data (assuming “perfect” seed words)
- **S0:** Train **Student** using **Teacher**’s predictions
- **T1:** Update **Teacher**’s weights (seed word qualities) using **Student**’s predictions
- **S1:** … Iterate until convergence

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

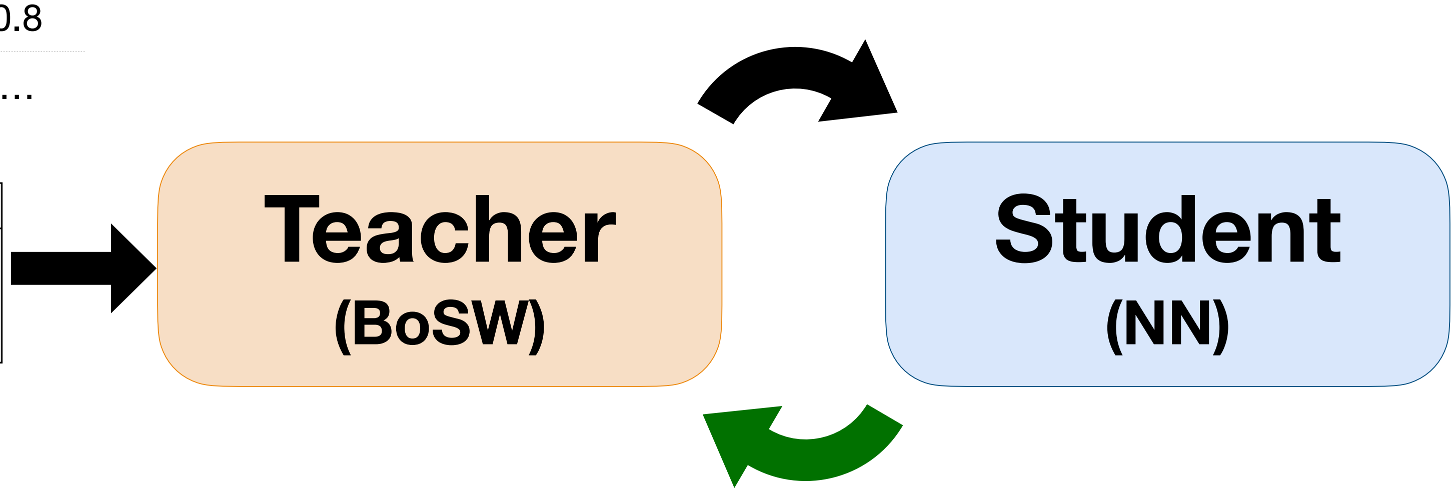


Iterative co-training pipeline

- **T0:** Apply **Teacher** on unlabeled data (assuming “perfect” seed words)
- **S0:** Train **Student** using **Teacher**’s predictions
- **T1:** Update **Teacher**’s weights (seed word qualities) using **Student**’s predictions
- **S1:** … Iterate until convergence

Qualities	Price	Image	Sound
bright	0.0	0.7	0.3
noise	0.0	0.2	0.8
…	…	…	…

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

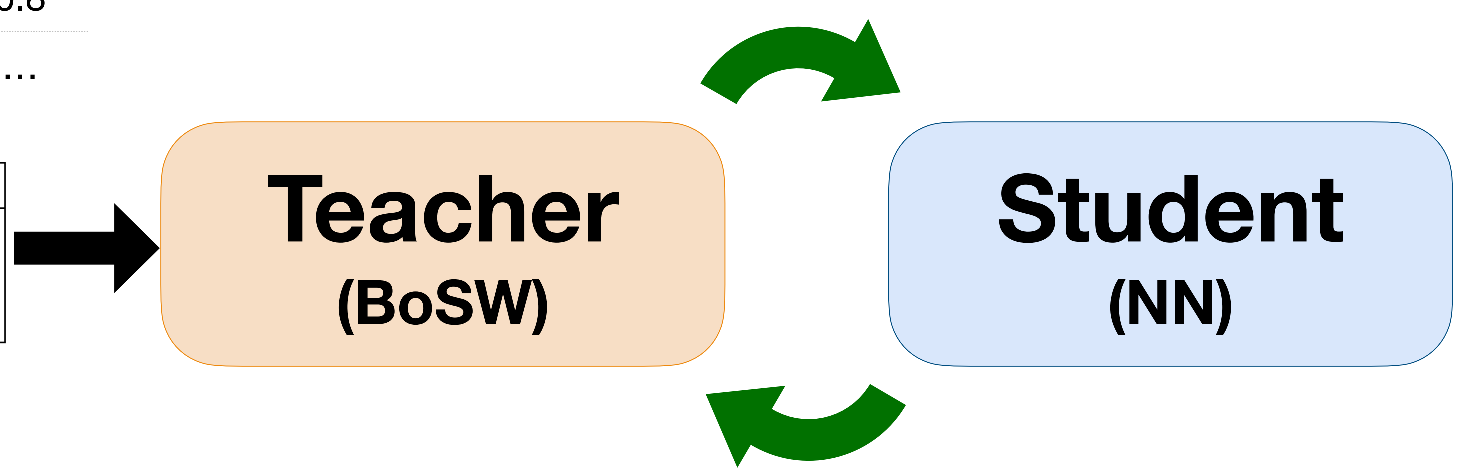


Iterative co-training pipeline

- **T0:** Apply **Teacher** on unlabeled data (assuming “perfect” seed words)
- **S0:** Train **Student** using **Teacher**’s predictions
- **T1:** Update **Teacher**’s weights (seed word qualities) using **Student**’s predictions
- **S1:** … Iterate until convergence

Qualities	Price	Image	Sound
bright	0.0	0.7	0.3
noise	0.0	0.2	0.8
…	…	…	…

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



Experiments: segment-level aspect detection

- **Datasets:**
  1. OPOSUM-Bags&Cases
  2. OPOSUM-Keyboards
  3. OPOSUM-Boots
  4. OPOSUM-Bluetooth Headsets
  5. OPOSUM-TVs
  6. OPOSUM-Vacuums
  7. SemEval-Restaurants-English
  8. SemEval-Restaurants-Spanish
  9. SemEval-Restaurants-French
  10. SemEval-Restaurants-Russian
  11. SemEval-Restaurants-Dutch
  12. SemEval-Restaurants-Turkish

OPOSUM
Amazon product reviews
9 aspects / domain: Quality, Looks, Price, …

SemEval-2016
Restaurant reviews
12 aspects / language: Ambience, Service, Food, …

Experimental procedure

- **Methodology:** experiments on 12 datasets separately
  - Training: no labeled data, 1M **unlabeled** segments, 30 **seed words** per aspect
  - Evaluation: micro-averaged F1 score on labeled test segments (5 runs)

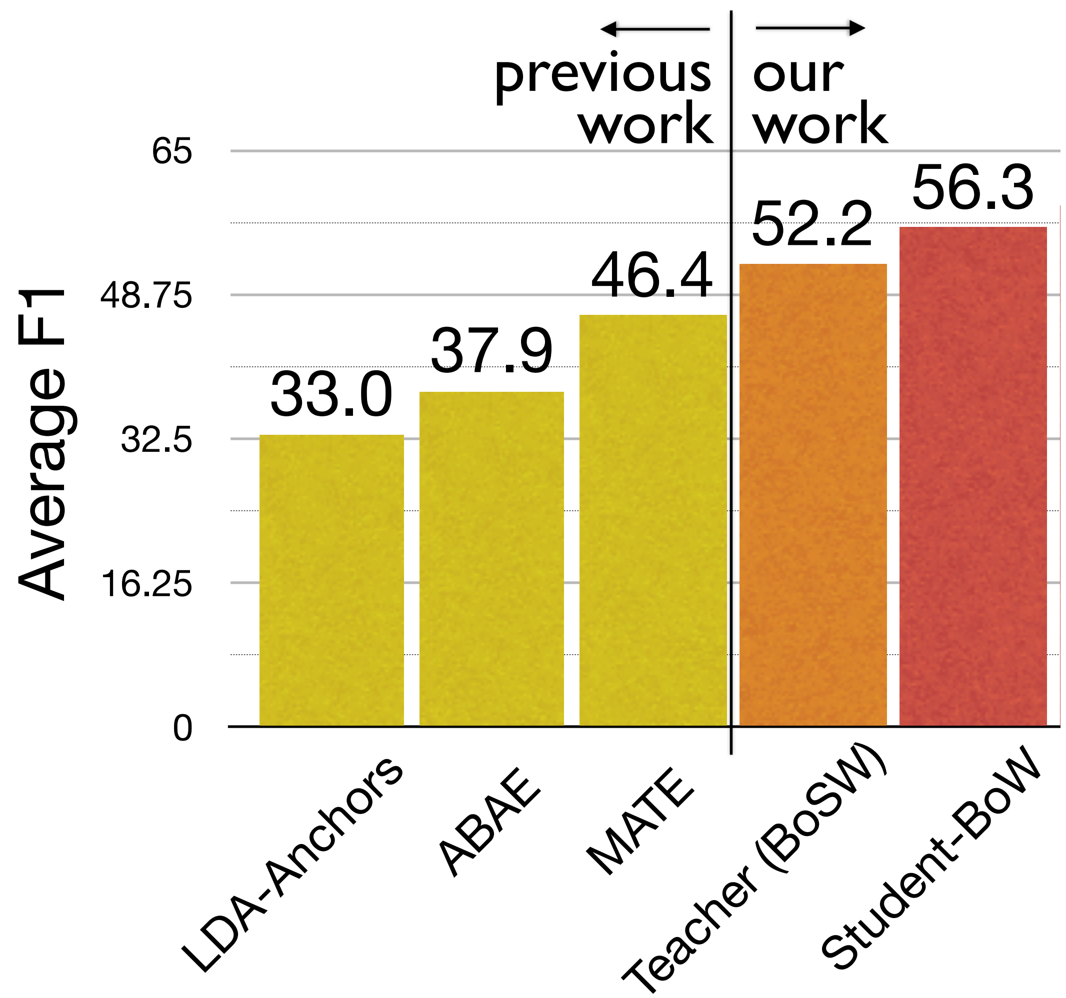
Experimental procedure

- **Methodology:** experiments on 12 datasets separately
  - Training: no labeled data, 1M **unlabeled** segments, 30 **seed words** per aspect
  - Evaluation: micro-averaged F1 score on labeled test segments (5 runs)

- **Model Comparison:** how to use seed words?
  - **LDA-Anchors**: Initialize “anchors” in topic models (Lund et al., 2017)
  - **ABAE:** Map learned topics to aspects post-hoc (He et al., 2017)
  - **MATE:** Initialize neural autoencoder with seed embeddings (Angelidis & Lapata, 2018)
  - Our **Teacher-Student** framework: use seed words in Teacher to train Student

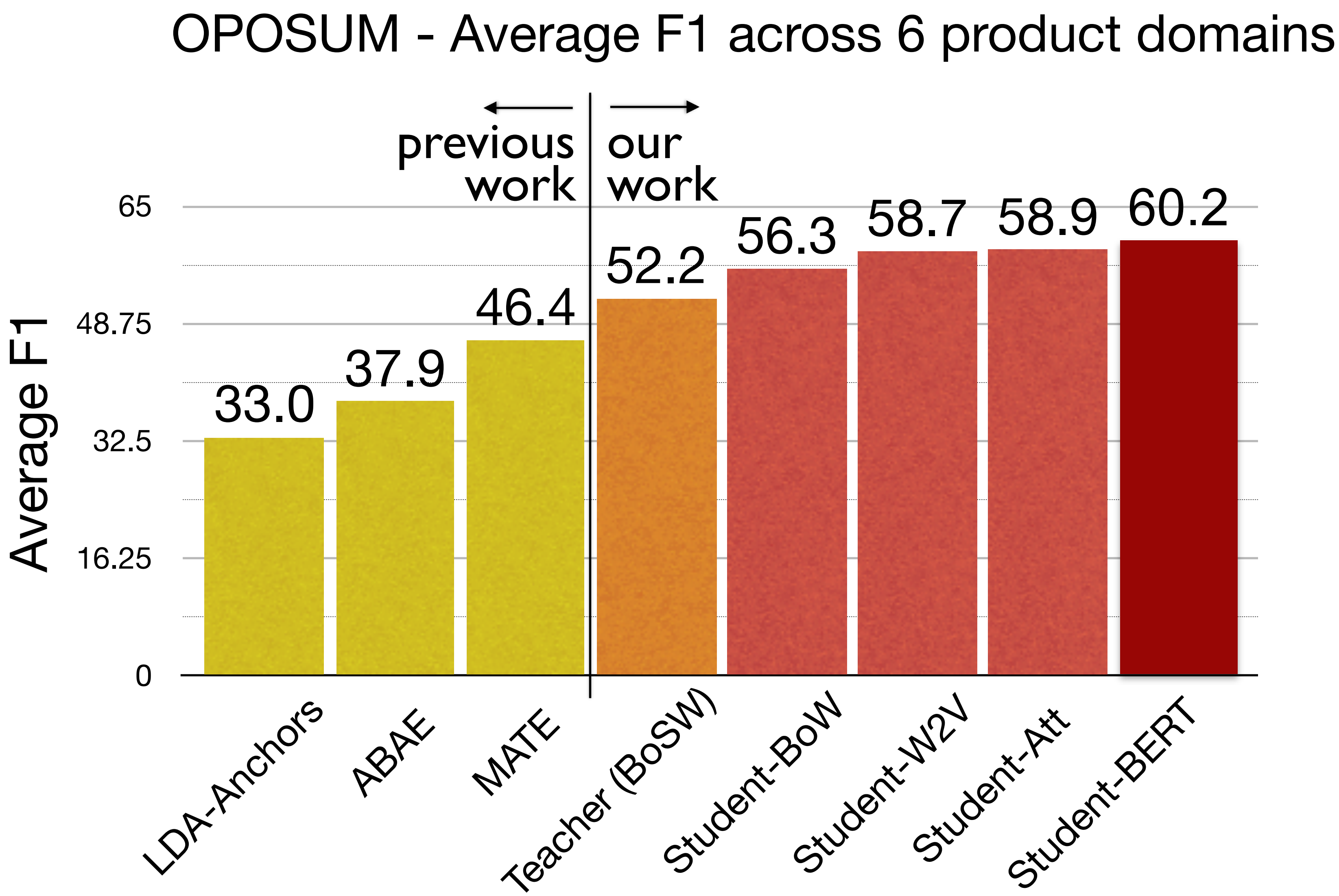
We leverage the predictive power of seed words

OPOSUM - Average F1 across 6 product domains



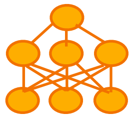
- **Teacher** and **Student-BoW (bag-of-words)** outperform previous approaches

Better embedding techniques boost the student's performance



- Our framework can be used with **any** type of Student classifier
- **Student** outperforms **MATE** by **29.7%** in F1 across 6 product domains
- **Student** outperforms **MATE** across **all** domains and languages by 14.1 in F1 on average

The student generalizes beyond seed words

	F1 (12-class)
Teacher-BoSW (bag of seed words)	52.2
Student-BERT 	**60.2 (+15%)**

No labeled data!

The student generalizes beyond seed words

	F1 (12-class)
Teacher-BoSW (bag of seed words)	52.2
Student-BERT	**60.2 (+15%)** ⟵ *No labeled data!*

Example (1): *"The only drawback is that this place is really* ***expensive****."*

expensive → **Teacher-BoSW:** Price ✓

→ **Student-BERT:** Price ✓

Example (2): *"They* ***pay*** *such detail to everything from miso soup to complex rolls"*

pay → **Teacher-BoSW:** Price ✗

→ **Student-BERT:** Food Quality ✓

Example (3): *"It's great if you spent the day there and didn't want to drive to eat"*

(no seed words) → **Teacher-BoSW:** - ✗

→ **Student-BERT:** Location ✓

Efficient machine teaching frameworks for NLP

Coarse labels (Ch. 3)

[Karamanolakis et al. WNUT '19]

Hierarchical taxonomies (Ch. 4)

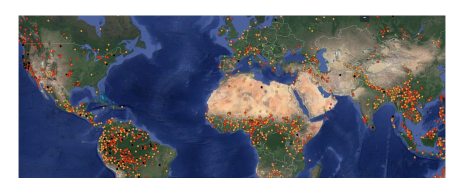
[Karamanolakis et al. ACL '20]

Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP '19]

Word translations (Ch. 6)

"injured"  "يارىلانغان"

[Karamanolakis et al. Findings of EMNLP '20]

Labeling rules (Ch. 7, 8)

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

[Karamanolakis et al. NAACL '21] + work with Daniel Hsu and Luis Gravano

Document classification beyond english

- It is **expensive** to obtain labeled documents for each **target** language
- **Cross-lingual classification:** use labeled documents from a **source** language

source: http://endangeredlanguages.com/

Cross-lingual text classification with minimal resources

- Challenge: how to bridge the source and target languages?

??

Source Language

“We need medical supplies”

Target Language

“بىزگە داۋالاش بۇيۇملىرى لازىم”

Cross-lingual text classification with minimal resources

- Challenge: how to bridge the source and target languages?

??

Source Language

“We need medical supplies”

Target Language

“بىزگە داۋالاش بۇيۇملىرى لازىم”

Cross-lingual resources required!

- Parallel corpora
- Machine translation systems
- Bilingual dictionaries
- Pre-trained multilingual language models
- …

Cross-lingual text classification with minimal resources

- Challenge: how to bridge the source and target languages?

??

Source Language

“We need medical supplies”

Target Language

“بىزگە داۋالاش بۇيۇملىرى لازىم”

Cross-lingual resources required!

- Parallel corpo
- Machine transl   stems
- Bilingual diction
- Pre-trained m   l language models
- …

“Cross-lingual text classification with minimal resources by transferring a sparse teacher”
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (Findings of EMNLP 2020)

Cross-lingual text classification with minimal resources

- Challenge: how to bridge the source and target languages?
- Our approach: translate just the **most indicative** seed words

??

Source Language

“We need medical supplies”

Target Language

“بىزگە داۋالاش بۇيۇملىرى لازىم”

Cross-lingual resources required!

- Parallel corpo
- Machine transl stems
- Bilingual diction
- Pre-trained m l language models
- …

“Cross-lingual text classification with minimal resources by transferring a sparse teacher”
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (Findings of EMNLP 2020)

Cross-lingual text classification with minimal resources

- Challenge: how to bridge the source and target languages?
- Our approach: translate just the **most indicative** seed words

Translation budget, e.g., 3 words

Source Language | **Target Language**

“Cross-lingual text classification with minimal resources by transferring a sparse teacher”
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (Findings of EMNLP 2020)

Cross-lingual text classification with minimal resources

- Challenge: how to bridge the source and target languages?
- Our approach: translate just the **most indicative** seed words

Translation budget, e.g., 3 words

Source Language		**Target Language**
1. injured	→	1. يارىلانغان
2. attacks	→	2. ھۇجۇملار
3. disease	→	3. كېسەل
4. night ✗		
5. … ✗		

“Cross-lingual text classification with minimal resources by transferring a sparse teacher”
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (Findings of EMNLP 2020)

Cross-lingual text classification with minimal resources

- Challenge: how to bridge the source and target languages?
- Our approach: translate just the **most indicative** seed words

Translation budget, e.g., 3 words

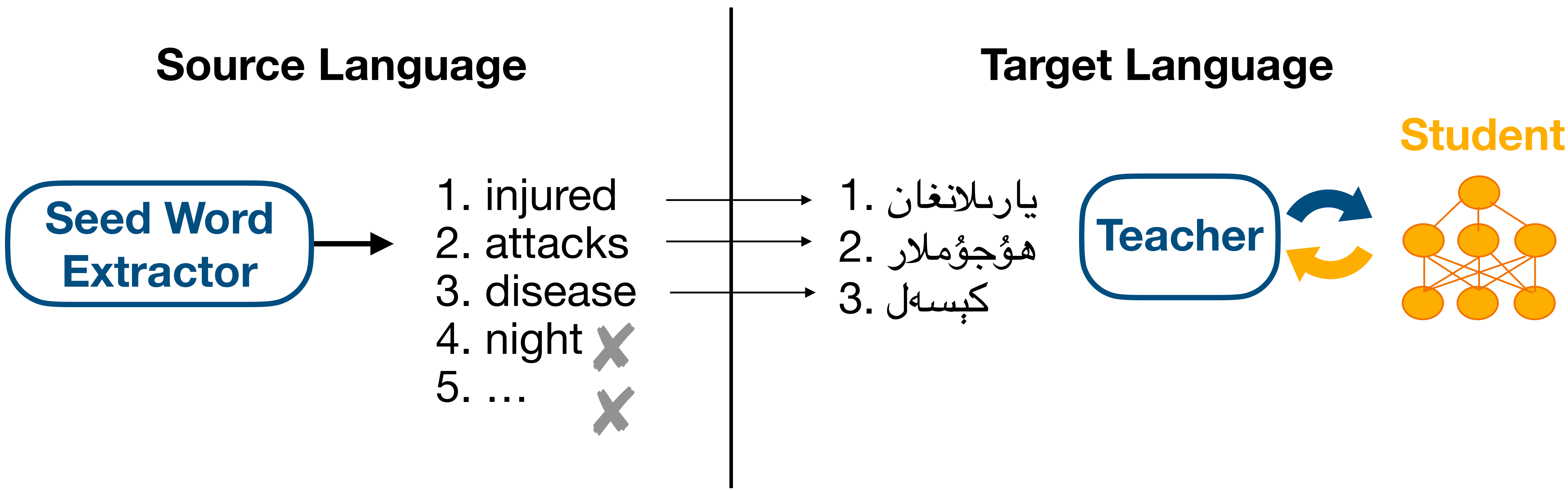
Source Language			**Target Language**
		???	
???	1. injured 2. attacks 3. disease 4. night ✗ 5. … ✗	→ 1. يارىلانغان → 2. ھۇجۇملار → 3. كېسەل	**???**

"Cross-lingual text classification with minimal resources by transferring a sparse teacher"
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (Findings of EMNLP 2020)

Cross-lingual text classification with minimal resources

- Challenge: how to bridge the source and target languages?
- Our approach: translate just the **most indicative** seed words

Cross-lingual Teacher Student (CLTS)

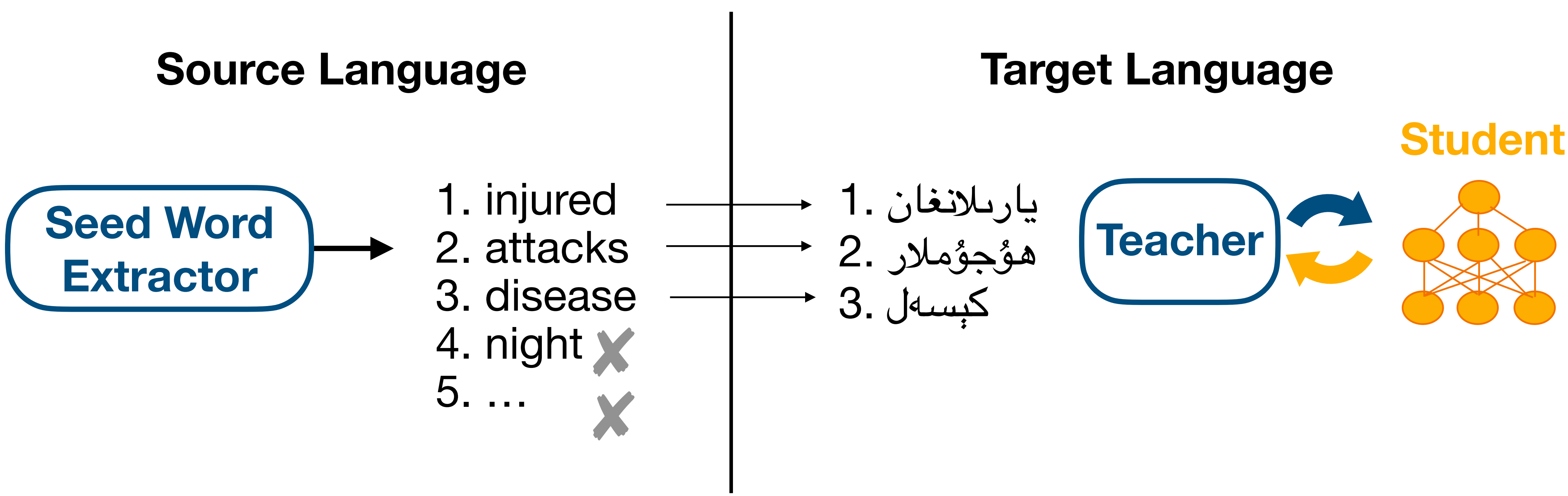


“Cross-lingual text classification with minimal resources by transferring a sparse teacher”
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (Findings of EMNLP 2020)

Cross-lingual text classification with minimal resources

- Challenge: how to bridge the source and target languages?
- Our approach: translate just the **most indicative** seed words

Cross-lingual Teacher Student (CLTS)



(+) Does **not** require parallel corpora / machine translation / multilingual representations

"Cross-lingual text classification with minimal resources by transferring a sparse teacher"
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (Findings of EMNLP 2020)

Evaluation of CLTS on 18 languages and 4 tasks

“Cross-lingual text classification with minimal resources by transferring a sparse teacher”
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (Findings of EMNLP 2020)

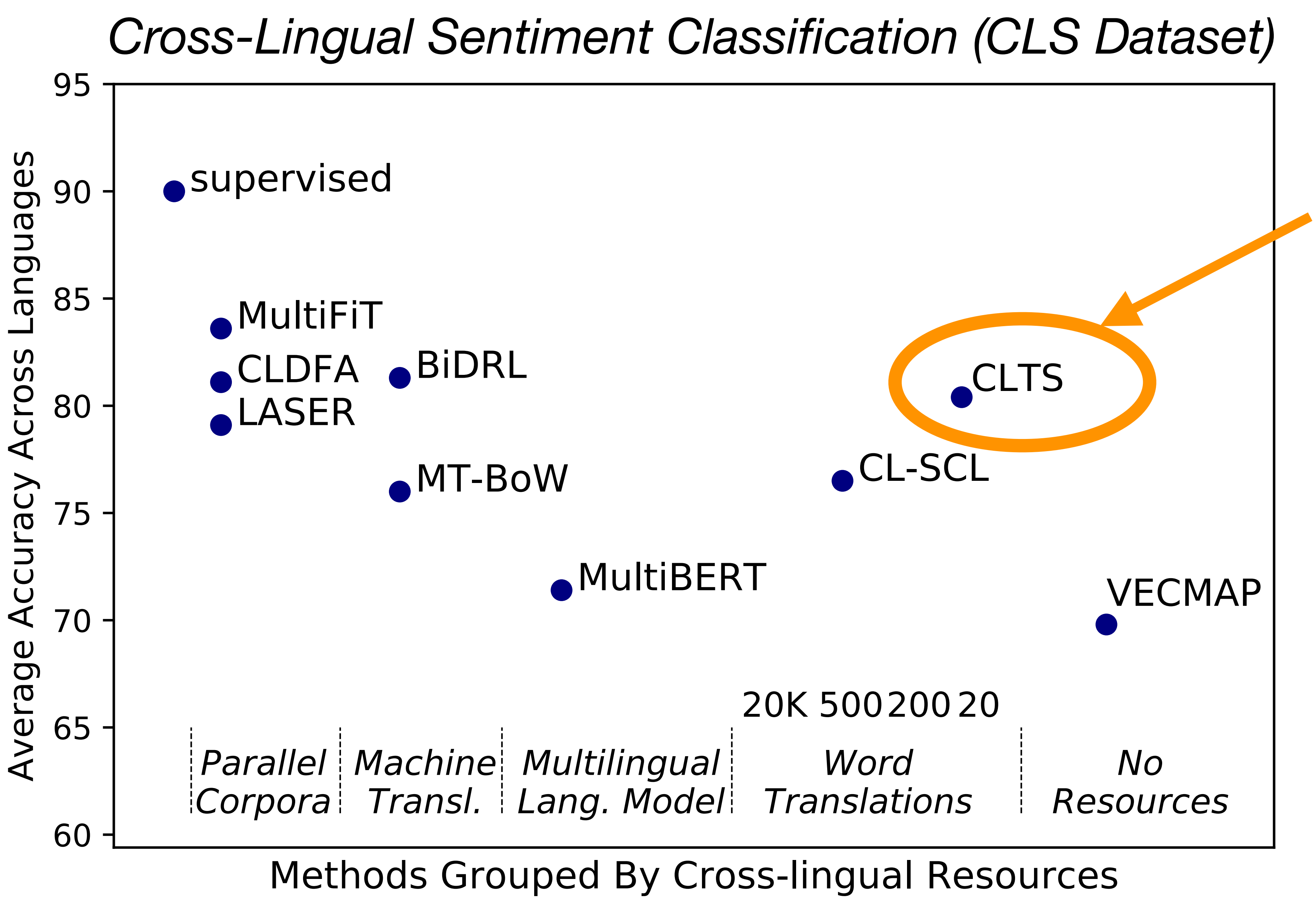
Evaluation of CLTS on 18 languages and 4 tasks

- Student outperforms Teacher by 56% on average across 18 languages

"Cross-lingual text classification with minimal resources by transferring a sparse teacher"
Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. (Findings of EMNLP 2020)

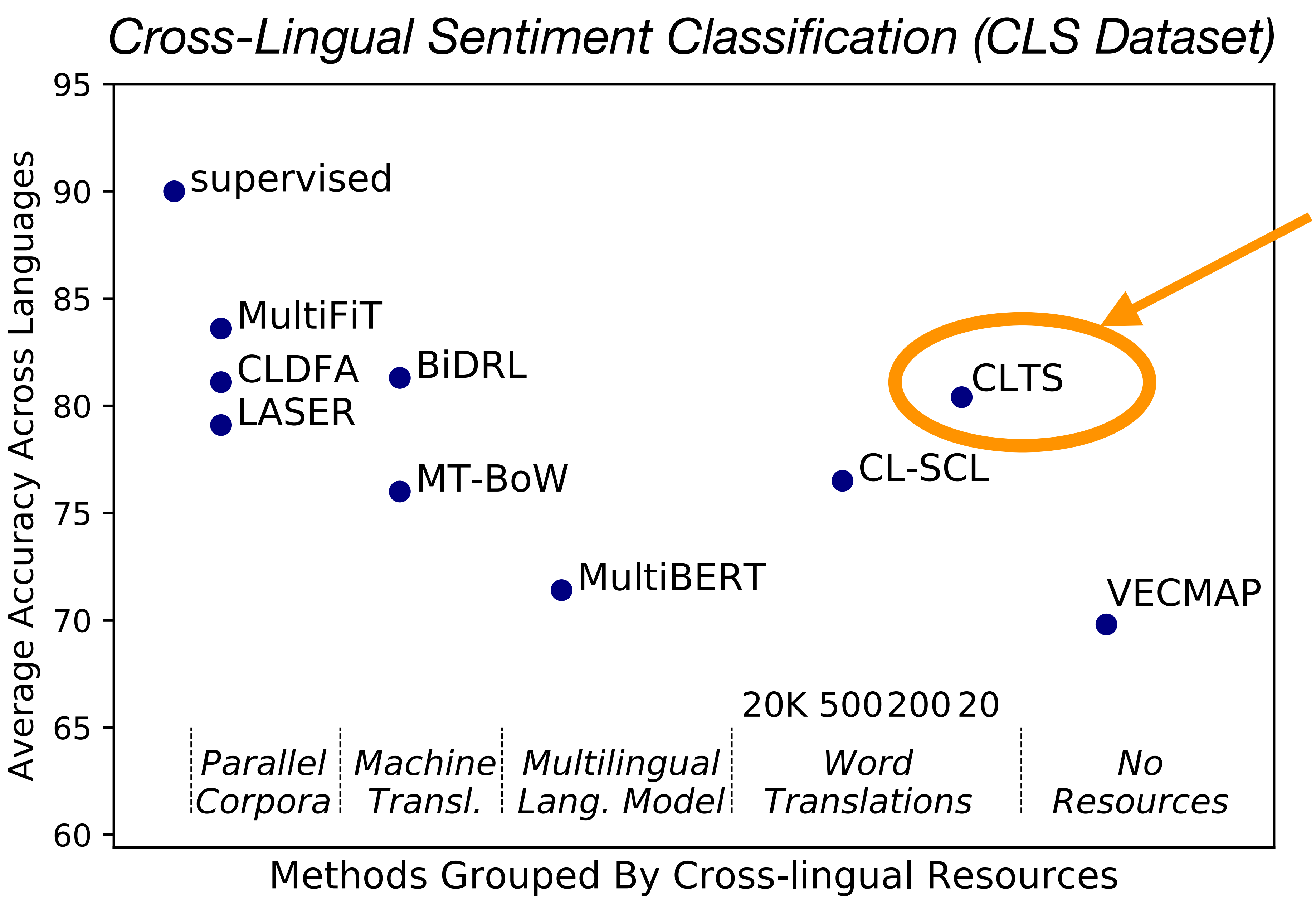
Evaluation of CLTS on 18 languages and 4 tasks

- Student outperforms Teacher by 56% on average across 18 languages
- CLTS is effective with as few as 20 word translations



Evaluation of CLTS on 18 languages and 4 tasks

- Student outperforms Teacher by 56% on average across 18 languages
- CLTS is effective with as few as 20 word translations
- CLTS sometimes outperforms **even more expensive** approaches

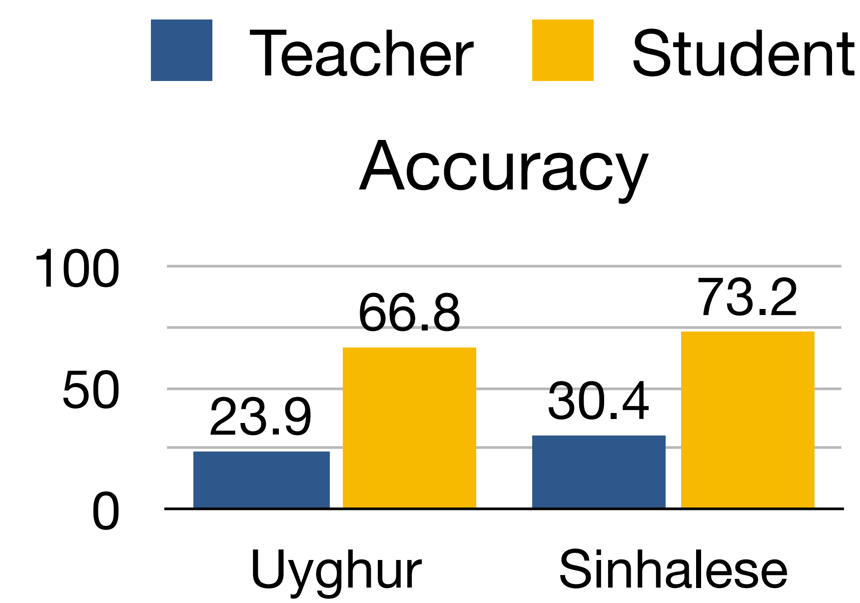


Evaluation of CLTS on 18 languages and 4 tasks

- Student outperforms Teacher by 56% on average across 18 languages
- CLTS is effective with as few as 20 word translations
- CLTS sometimes outperforms **even more expensive** approaches
- CLTS can be applied in **low-resource** languages

Medical emergency situation detection

English	->	Uyghur	Sinhalese
1. injured	->	يارىلانغان	තුවාල ලැබුවා
2. attacks	->	ھۇجۇملار	ප්‍රහාර
3. medical	->	medical	වෛද්‍ය
4. crisis	->	كرىزىس	අර්බුදය
5. disease	->	كېسەل	රෝගය
6. malaria	->	بەزگەك كېسىلى	මැලේරියාව
7. health	->	ساغلاملىق	සෞඛ්‍යය
8. injuring	->	يارىلىنىش	තුවාල වීම
9. yemen	->	يەمەن	යේමනය
10. hospitals	->	دوختۇرخانىلار	රෝහල්



Food poisoning detection across languages

cross-lingual transfer

English Labeled Data → Multi-lingual classifiers

Chinese

Basha – Sherbrooke Unclaimed

4/4/2018

千！万！别！去！我男朋友昨天晚上点了个shawarma plate，从凌晨三点开始上吐下泻到现在。我认识他五年，连感冒都没见他得过。珍爱生命远离这家餐馆吧。

Spanish

La Mojarra Loca Grill Unclaimed

7/23/2017

Este lugar la verdad no se los recomiendo y más si se trata para los niños. Fui con mi familia al lunch y mi niño pidió chicken nuggets y de verdad se los digo esos pedazos de pollo estaban asquerosos parece que los tenían de hace mucho tiempo y el de inmediato empezó a vomitar es increíble que un niño de 4 años te diga que la comida no sirve eso para el chef. ...

"Discovering foodborne illness complaints in multiple languages using English annotations only"
Ziyi Liu, **Giannis Karamanolakis**, Daniel Hsu, and Luis Gravano. (LOUHI@EMNLP 2020)

Efficient machine teaching frameworks for NLP

Coarse labels (Ch. 3)

[Karamanolakis et al. WNUT ’19]

Hierarchical taxonomies (Ch. 4)

[Karamanolakis et al. ACL ’20]

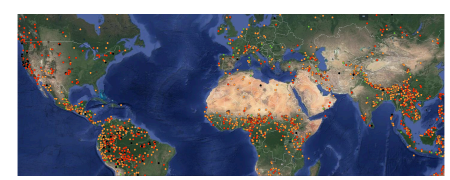
Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP ’19]

Word translations (Ch. 6)

“injured”



“يارىلانغان”

[Karamanolakis et al. Findings of EMNLP ’20]

Labeling rules (Ch. 7, 8)

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

[Karamanolakis et al. NAACL ’21] + work with Daniel Hsu and Luis Gravano

Efficient machine teaching frameworks for NLP

It is not always easy to express domain knowledge as seed words

Coarse labels (Ch. 3)

[Karamanolakis et al. WNUT '19]

Hierarchical taxonomies (Ch. 4)

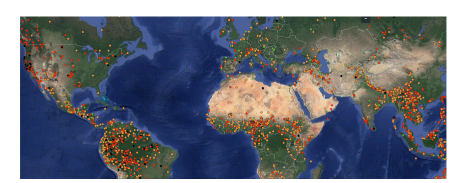
[Karamanolakis et al. ACL '20]

Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP '19]

Word translations (Ch. 6)

"injured"  "يارىلانغان"

[Karamanolakis et al. Findings of EMNLP '20]

Labeling rules (Ch. 7, 8)

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

[Karamanolakis et al. NAACL '21] + work with Daniel Hsu and Luis Gravano

Efficient machine teaching frameworks for NLP

It is not always easy to express domain knowledge as seed words

Coarse labels (Ch. 3)

[Karamanolakis et al. WNUT '19]

Hierarchical taxonomies (Ch. 4)

[Karamanolakis et al. ACL '20]

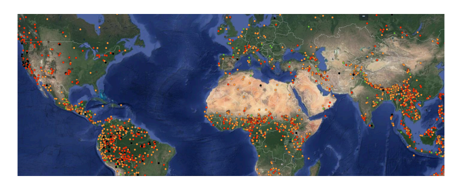
Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP '19]

Word translations (Ch. 6)

"injured"



"يارىلانغان"

[Karamanolakis et al. Findings of EMNLP '20]

Labeling rules (Ch. 7, 8)

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

[Karamanolakis et al. NAACL '21] + work with Daniel Hsu and Luis Gravano

Efficient machine teaching frameworks for NLP

It is not always easy to express domain knowledge as seed words

Coarse labels (Ch. 3)

[Karamanolakis et al. WNUT ’19]

Hierarchical taxonomies (Ch. 4)

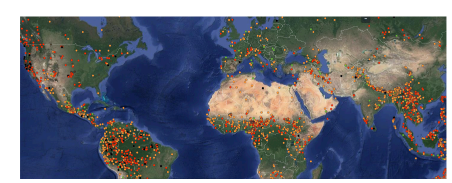
[Karamanolakis et al. ACL ’20]

Seed words (Ch. 5)

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

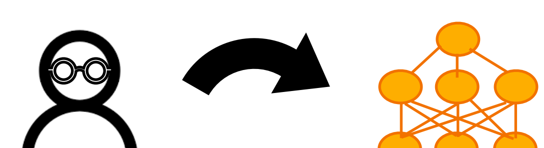
[Karamanolakis et al. EMNLP ’19]

Word translations (Ch. 6)

“injured”  “يارىلانغان”

[Karamanolakis et al. Findings of EMNLP ’20]

Labeling rules (Ch. 7, 8)



[Karamanolakis et al. NAACL ’21] + work with Daniel Hsu and Luis Gravano

Efficient machine teaching frameworks for NLP

It is not always easy to express domain knowledge as seed words

Coarse labels (Ch. 3)

[Karamanolakis et al. WNUT ’19]

Hierarchical taxonomies (Ch. 4)

[Karamanolakis et al. ACL ’20]

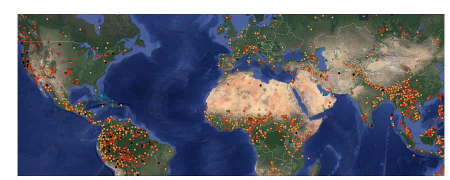
Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP ’19]

Word translations (Ch. 6)

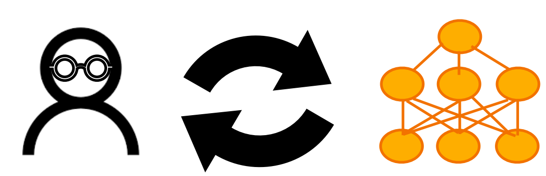
“injured”



“يارىلانغان”

[Karamanolakis et al. Findings of EMNLP ’20]

Labeling rules (Ch. 7, 8)



[Karamanolakis et al. NAACL ’21] + work with Daniel Hsu and Luis Gravano

From seed words to general labeling rules

- Rules: heuristic labeling functions written by **domain experts**
- Rules can be used to automatically create labels for **unlabeled** data x

Example: regular expression patterns

Spam classification

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x) else ABSTAIN
```

Question type classification

```
def numeric_question(x):
    return NUMERIC if x.startswith("when") else ABSTAIN
```

From seed words to general labeling rules

- Rules: heuristic labeling functions written by **domain experts**
- Rules can be used to automatically create labels for **unlabeled** data x

Example: regular expression patterns

Spam classification

```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x) else ABSTAIN
```

Question type classification

```
def numeric_question(x):
    return NUMERIC if x.startswith("when") else ABSTAIN
```

Rules based on external lexicons / models / knowledge bases

Sentiment classification

```
def sentiment_lexicon_score(x, sentiwordnet):
    if sentiwordnet(x) > 0.8:
        return POSITIVE
    elif sentiwordnet(x) < 0.2:
        return NEGATIVE
    else:
        ABSTAIN
```

Challenges in learning with labeling rules

(1) Noise

$x_i \rightarrow \text{rule1} \rightarrow \text{SPAM}$ ✗

True label: HAM

(2) Coverage

$x_i \rightarrow \text{rule1} \rightarrow \text{ABSTAIN}$

(3) Conflicts

x_i

rule1 → SPAM

rule2 → HAM

rule3 → SPAM

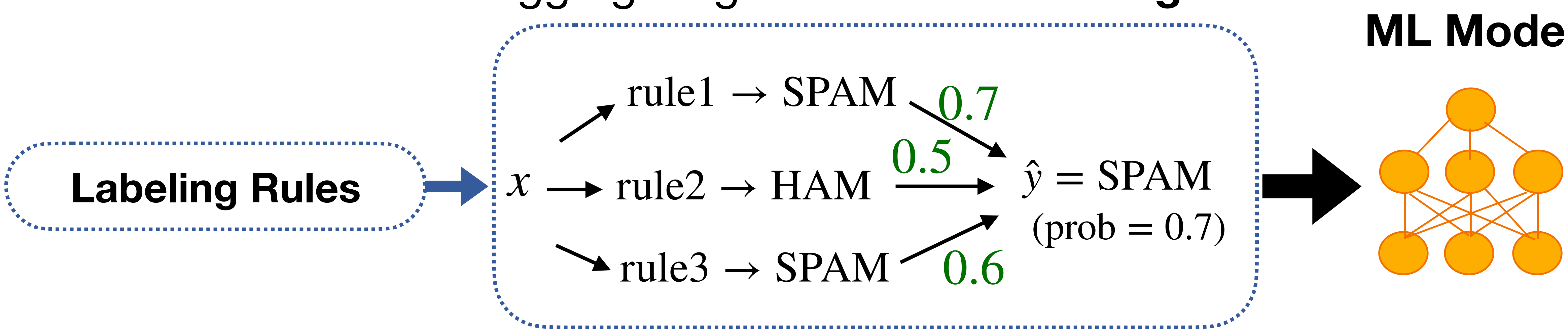
rule4 → HAM

???

How to train ML models with labeling rules?

[Ratner et al., 2017; Bach et al., 2019; Awasthi et al., 2020]

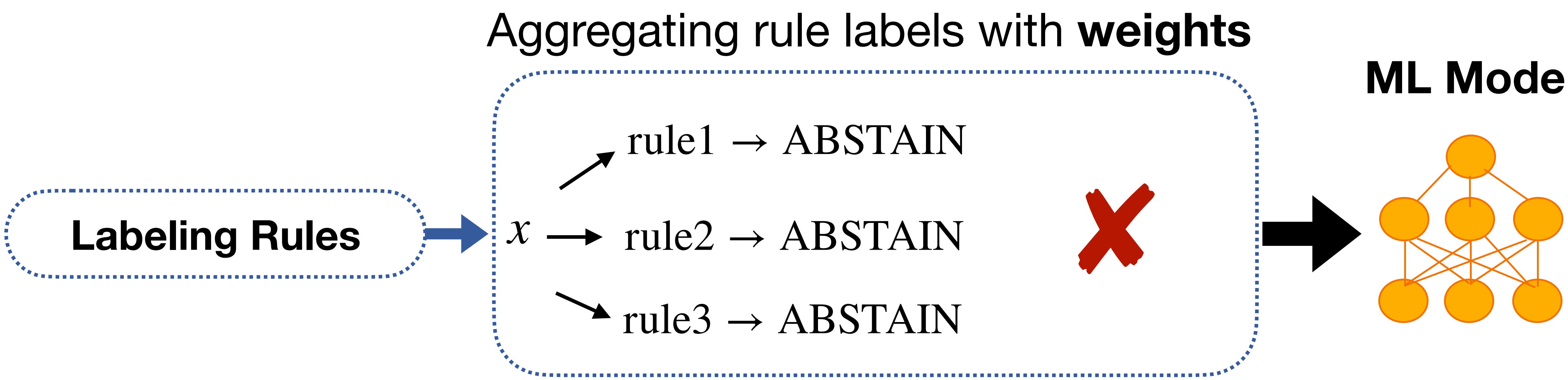
Aggregating rule labels with **weights**



How to train ML models with labeling rules?

- Previous work **ignores unlabeled instances** that are not covered by rules

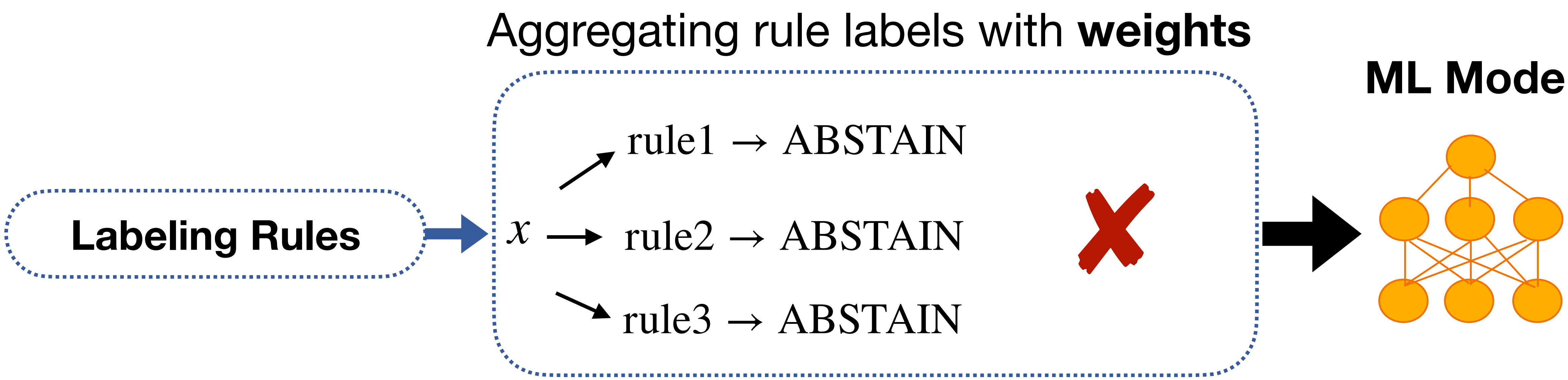
[Ratner et al., 2017; Bach et al., 2019; Awasthi et al., 2020]



How to train ML models with labeling rules?

- Previous work **ignores unlabeled instances** that are not covered by rules

[Ratner et al., 2017; Bach et al., 2019; Awasthi et al., 2020]



Expert-provided rules are often sparse

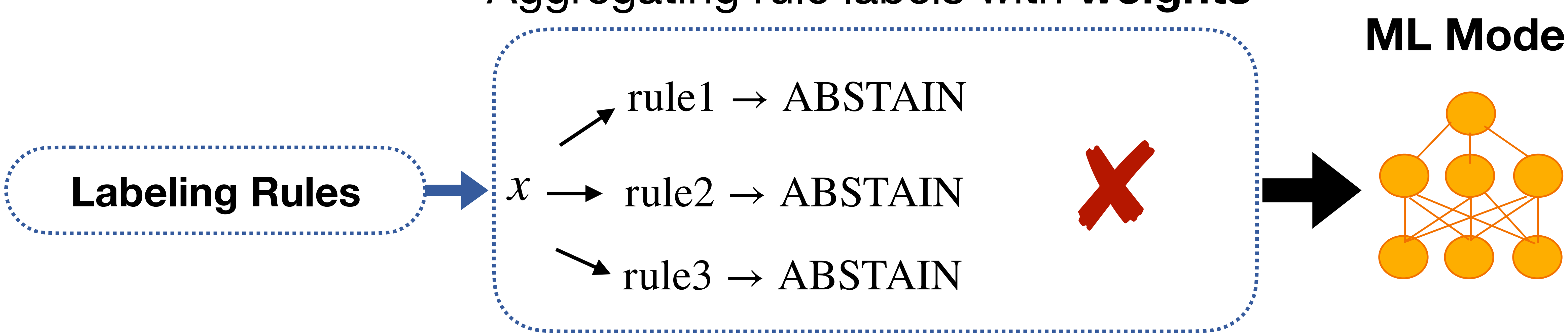
6 datasets (45 rules per dataset): 40% of instances are **ignored**

How to train ML models with labeling rules?

- Previous work **ignores unlabeled instances** that are not covered by rules

[Ratner et al., 2017; Bach et al., 2019; Awasthi et al., 2020]

Aggregating rule labels with **weights**



Expert-provided rules are often sparse

6 datasets (45 rules per dataset): 40% of instances are **ignored**

Don't throw them away!

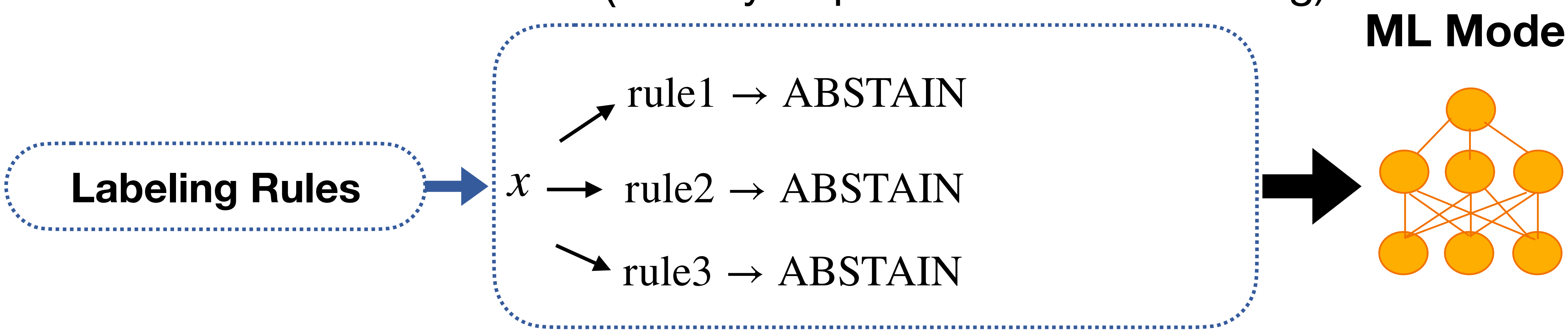
"Self-training with weak supervision",
Giannis Karamanolakis, Subhabrata Mukherjee, Guoqing Zheng, Ahmed Hassan Awadallah (NAACL '21)

Microsoft

How to train ML models with labeling rules?

- Previous work **ignores unlabeled instances** that are not covered by rules
- We **leverage all unlabeled instances** for weak supervision via **self-training**

ASTRA (we**A**kly-supervised **S**elf-**TRA**ining)

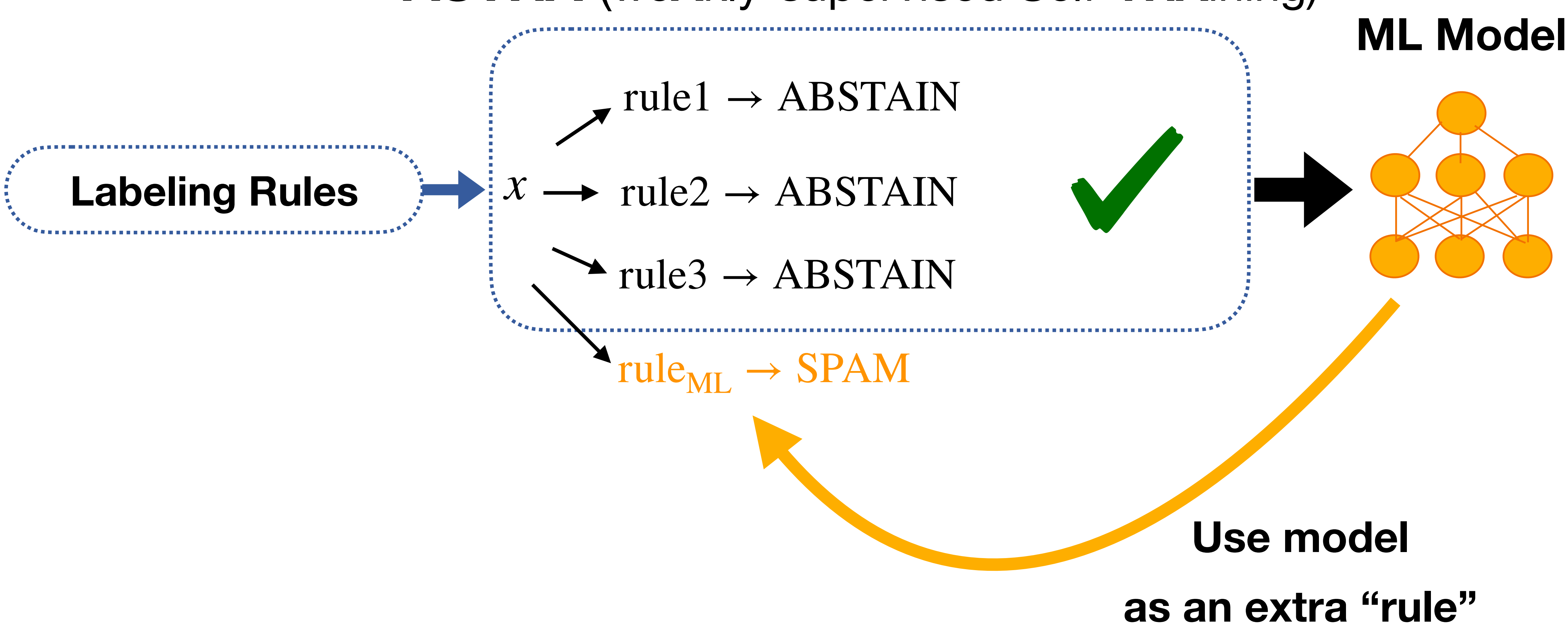


"Self-training with weak supervision",
Giannis Karamanolakis, Subhabrata Mukherjee, Guoqing Zheng, Ahmed Hassan Awadallah (NAACL '21)

How to train ML models with labeling rules?

- Previous work **ignores unlabeled instances** that are not covered by rules
- We **leverage all unlabeled instances** for weak supervision via **self-training**

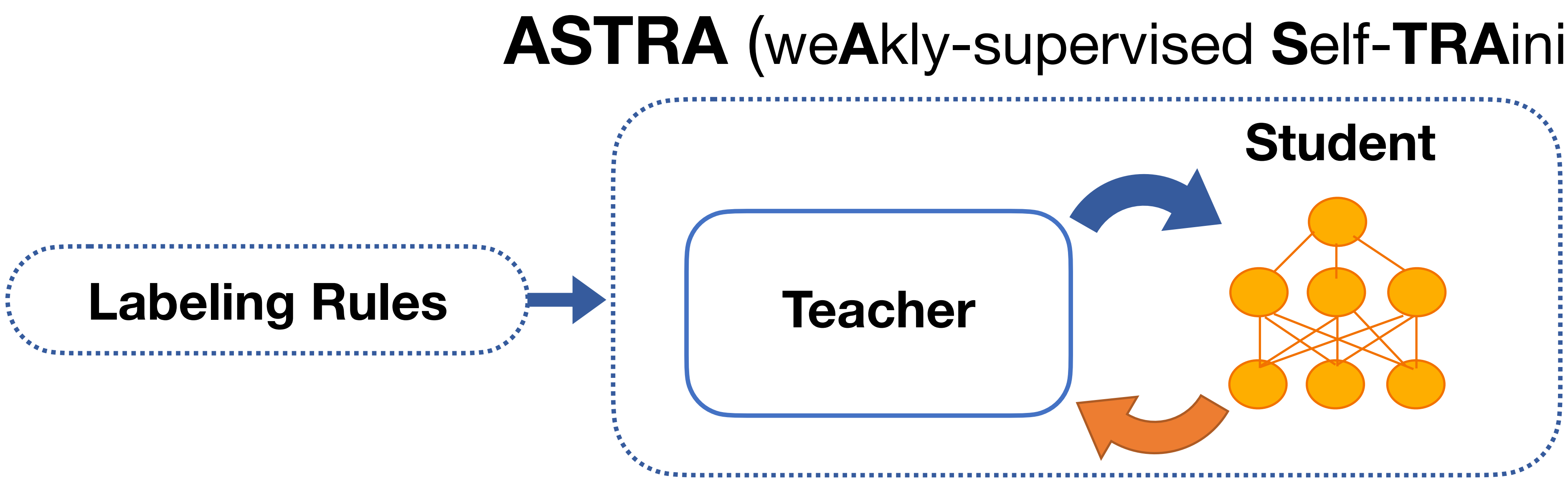
ASTRA (we**A**kly-supervised **S**elf-**TRA**ining)



"Self-training with weak supervision",
Giannis Karamanolakis, Subhabrata Mukherjee, Guoqing Zheng, Ahmed Hassan Awadallah (NAACL '21)

How to train ML models with labeling rules?

- Previous work **ignores unlabeled instances** that are not covered by rules
- We **leverage all unlabeled instances** for weak supervision via **self-training**



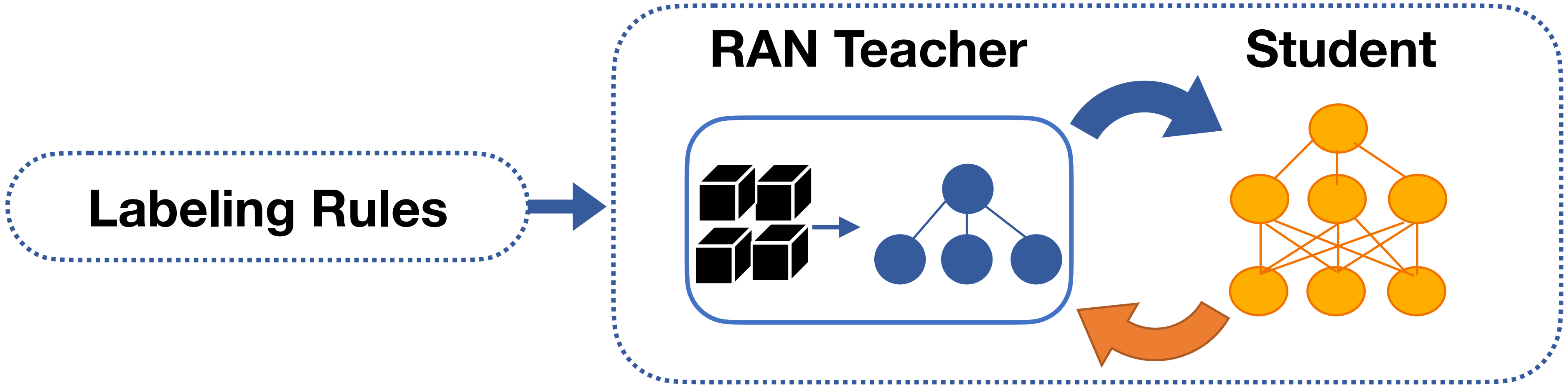
Microsoft

"Self-training with weak supervision",
Giannis Karamanolakis, Subhabrata Mukherjee, Guoqing Zheng, Ahmed Hassan Awadallah (NAACL '21)

How to train ML models with labeling rules?

- Previous work **ignores unlabeled instances** that are not covered by rules
- We **leverage all unlabeled instances** for weak supervision via **self-training**
- We develop a **rule attention network (RAN)** and an SSL training loss

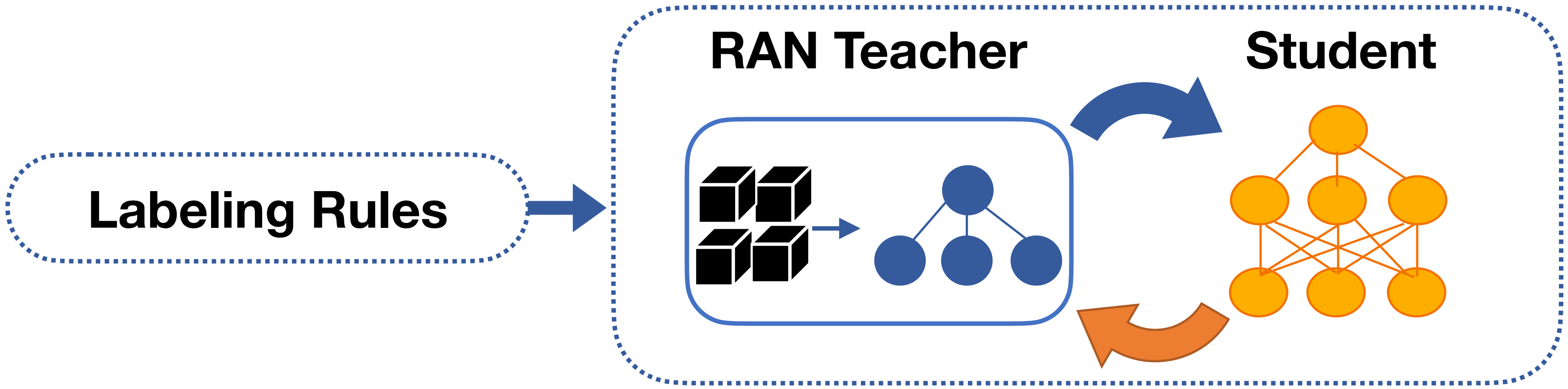
ASTRA (we**A**kly-supervised **S**elf-**TRA**ining)



"Self-training with weak supervision",
Giannis Karamanolakis, Subhabrata Mukherjee, Guoqing Zheng, Ahmed Hassan Awadallah (NAACL '21)

How to train ML models with labeling rules?

- Previous work **ignores unlabeled instances** that are not covered by rules
- We **leverage all unlabeled instances** for weak supervision via **self-training**
- We develop a **rule attention network (RAN)** and an SSL training loss
- On **6 NLP benchmarks,** ASTRA improves the state-of-the-art by 3% in accuracy

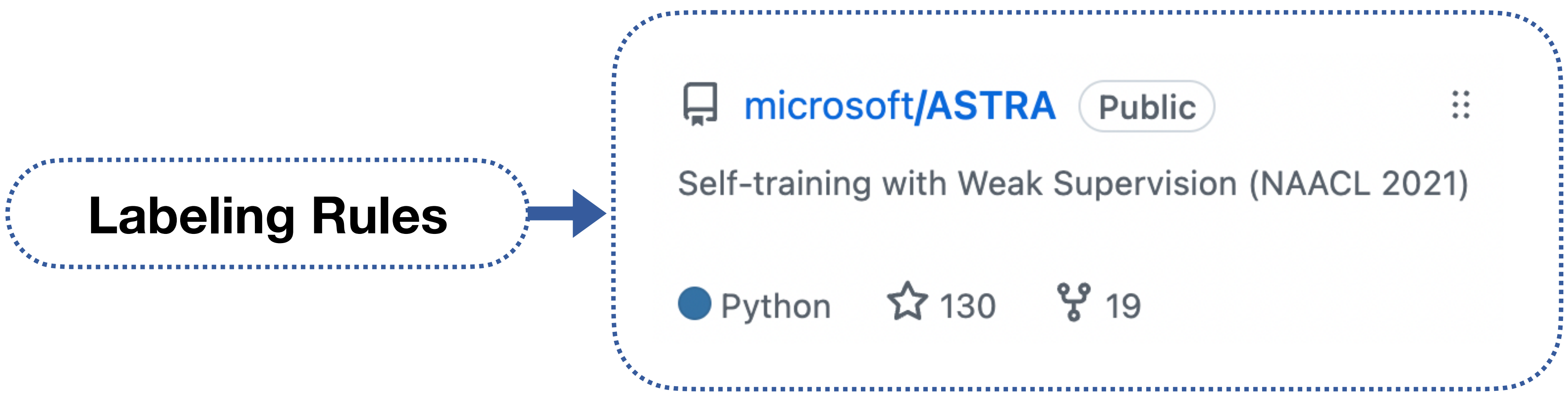


Microsoft

"Self-training with weak supervision",
Giannis Karamanolakis, Subhabrata Mukherjee, Guoqing Zheng, Ahmed Hassan Awadallah (NAACL '21)

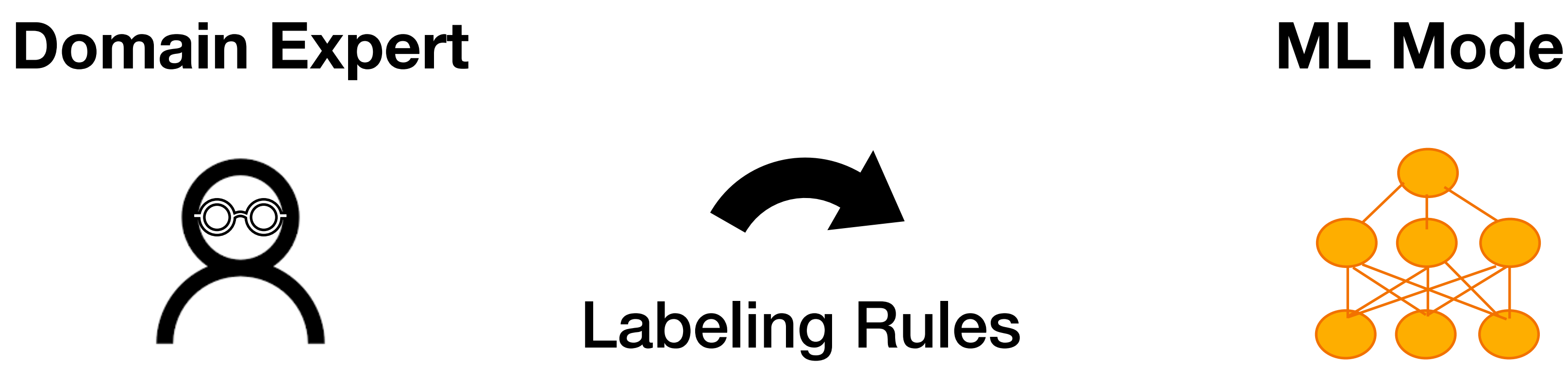
How to train ML models with labeling rules?

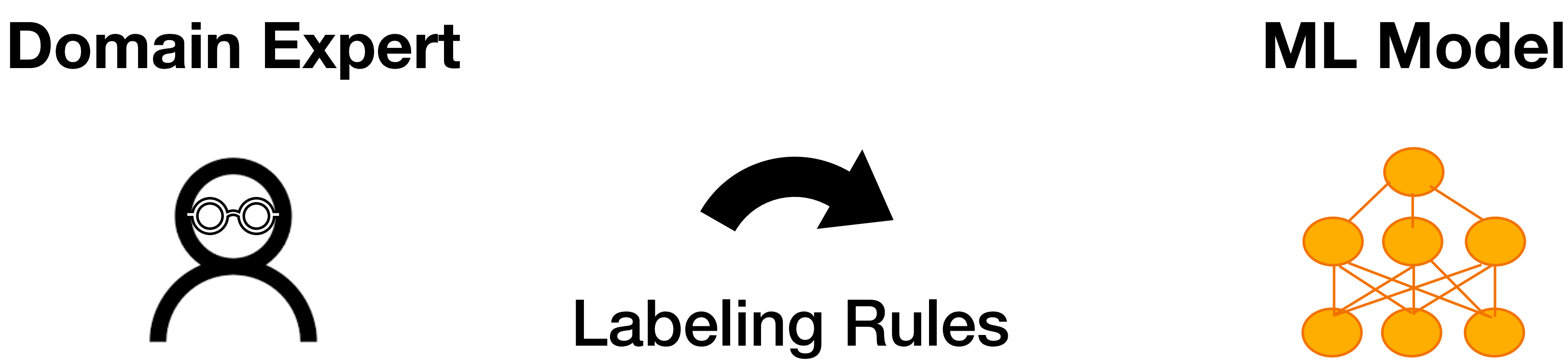
- Previous work **ignores unlabeled instances** that are not covered by rules
- We **leverage all unlabeled instances** for weak supervision via **self-training**
- We develop a **rule attention network (RAN)** and an SSL training loss
- On **6 NLP benchmarks,** ASTRA improves the state-of-the-art by 3% in accuracy



https://github.com/microsoft/ASTRA

“Self-training with weak supervision”,
Giannis Karamanolakis, Subhabrata Mukherjee, Guoqing Zheng, Ahmed Hassan Awadallah (NAACL ’21)

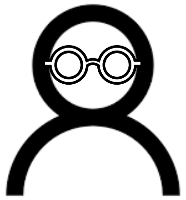

Domain Expert
Labeling Rules
ML Model



Assumption: Expert has already created a **sufficiently large** set of rules

Addressing new NLP tasks from the earliest stages

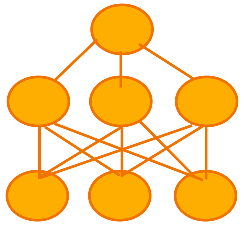
Domain Expert



Labeling Rules

???

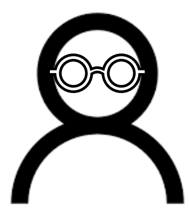
ML Model



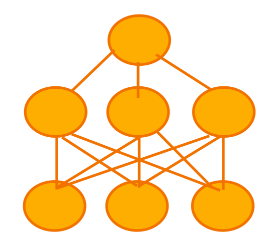
Addressing new NLP tasks from the earliest stages

- How to efficiently exploit the domain expert's limited time to teach a ML model?

Domain Expert



ML Model



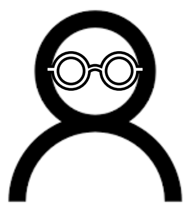
Labeling Rules

???

Addressing new NLP tasks from the earliest stages

- How to efficiently exploit the domain expert's limited time to teach a ML model?

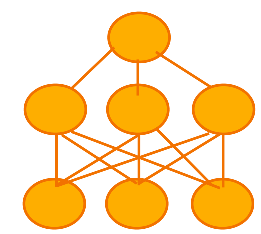
Domain Expert



Labeling Rules

???

ML Model



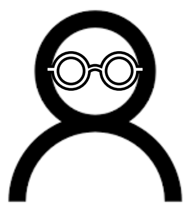
high coverage: "*check out*" → SPAM

vs. high precision: "*check out*" AND "http" AND "free" → SPAM

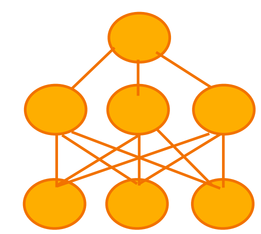
Addressing new NLP tasks from the earliest stages

- How to efficiently exploit the domain expert's limited time to teach a ML model?

Domain Expert



ML Model



Labeling Rules

???

Creating a sufficiently large set of rules requires

Time **Creativity** **Foresight**

Addressing new NLP tasks from the earliest stages

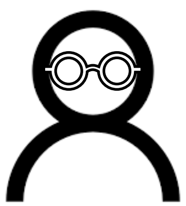
- How to efficiently exploit the domain expert's limited time to teach a ML model?

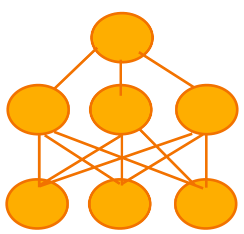
Labeled Data

Domain Expert

???

ML Model





Labeling Rules

???

Creating a sufficiently large set of rules requires

Time **Creativity** **Foresight**

Addressing new NLP tasks from the earliest stages

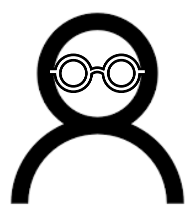
- How to efficiently exploit the domain expert's limited time to teach a ML model?

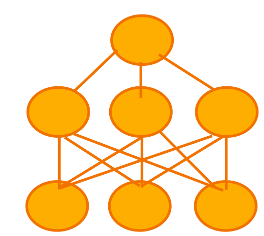
Labeled Data

Domain Expert

???

ML Model





Labeling Rules

???

Creating a sufficiently large set of rules requires

Time **Creativity** **Foresight**

- Goal: assist experts via guidelines and tools for **efficient** machine teaching

work with Daniel Hsu and Luis Gravano

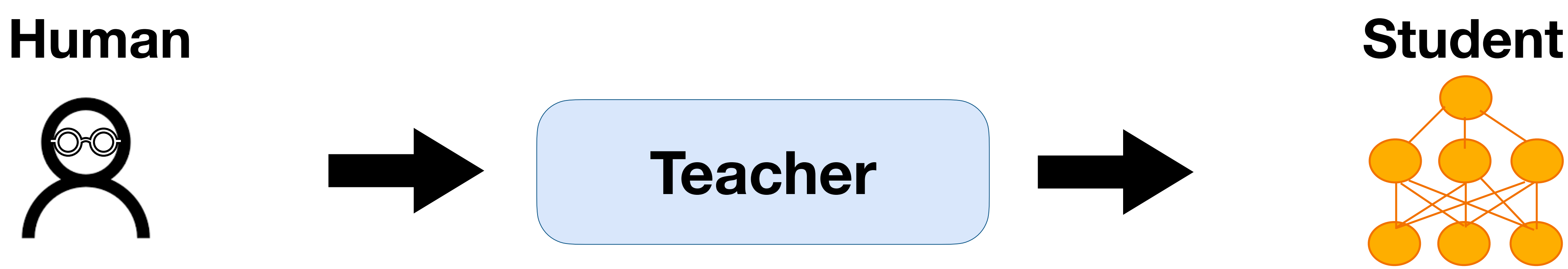
Our analysis and interactive framework for efficient machine teaching

Our analysis and interactive framework for efficient machine teaching

- We reveal patterns that could inform guidelines for rule creation

We evaluate 1000+ Human-Teacher-Student configurations

in 6 datasets with human-provided rules



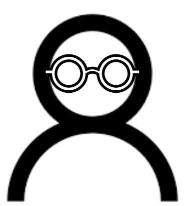
Our analysis and interactive framework for efficient machine teaching

- We reveal patterns that could inform guidelines for rule creation

We evaluate 1000+ Human-Teacher-Student configurations

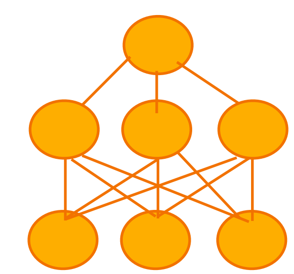
in 6 datasets with human-provided rules

Human



Teacher

Student



10 Resource sizes

Low → … →High

6 Teacher types

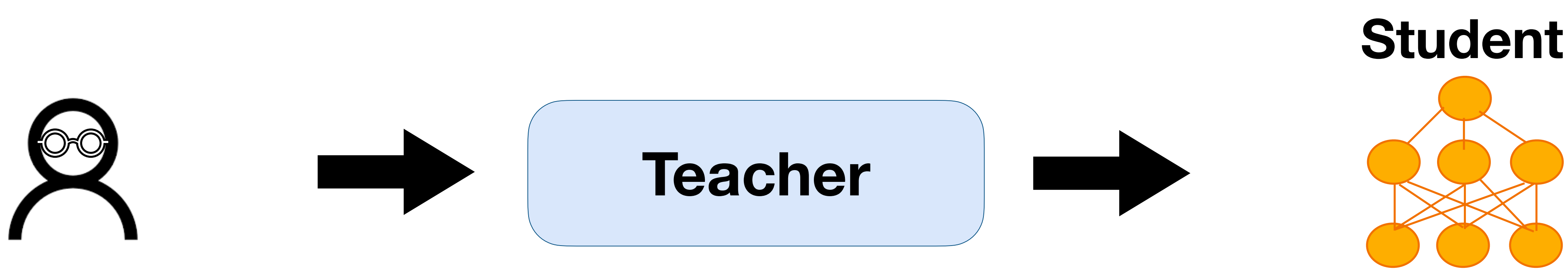
Majority voting
[Dawid, Skene, 1979]
[Ratner et al., 2017]
[Ratner et al., 2019]
[Fu et al., 2020]
[Karamanolakis et al., 2021]

3 Student types

Bag-of-words logistic regression
Multi-layer perceptron
BERT-based classifier

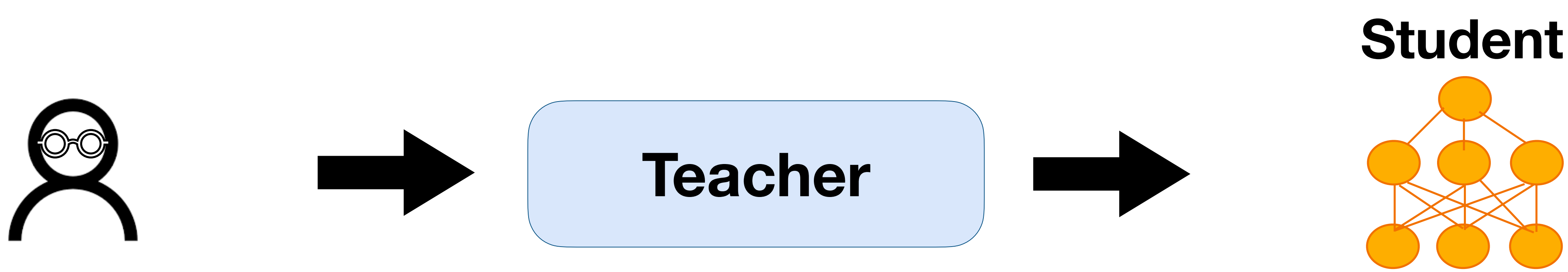
Our analysis and interactive framework for efficient machine teaching

- We reveal patterns that could inform guidelines for rule creation
  - Increasing Teacher's F1 does **not** necessarily increase Student's F1



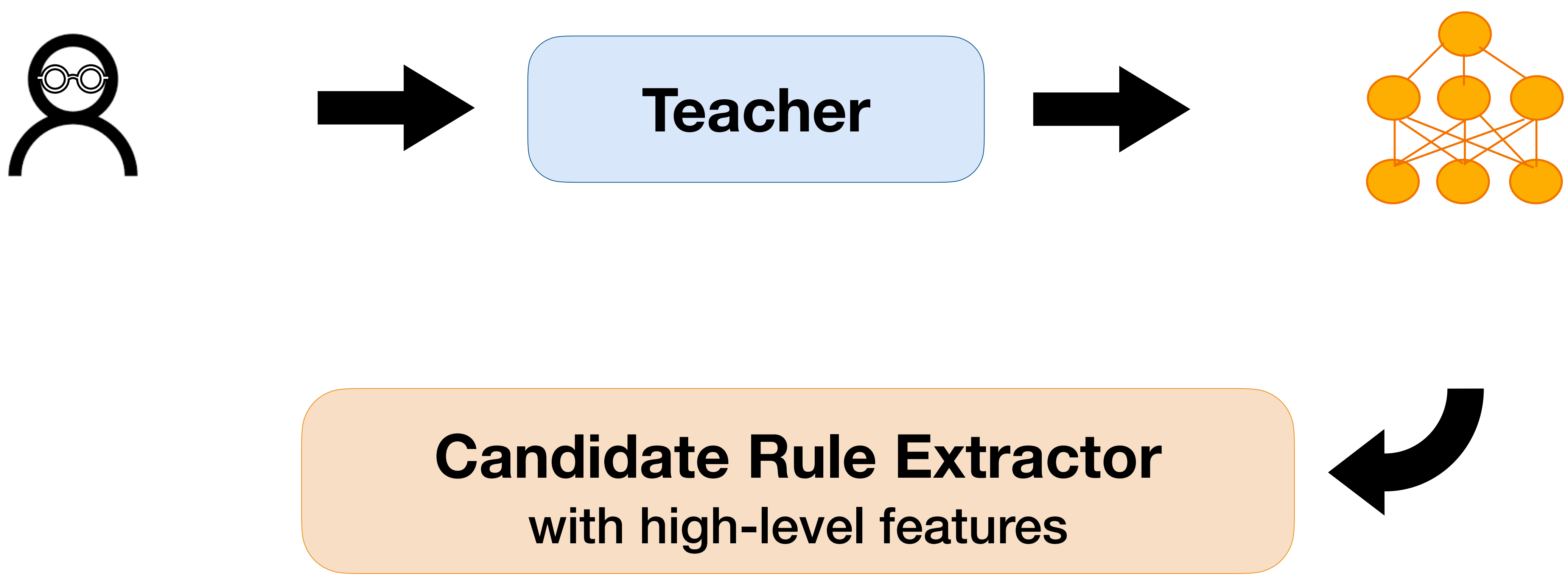
Our analysis and interactive framework for efficient machine teaching

- We reveal patterns that could inform guidelines for rule creation
    - Increasing Teacher's F1 does **not** necessarily increase Student's F1
    - Teacher's **precision** is more important than **coverage**



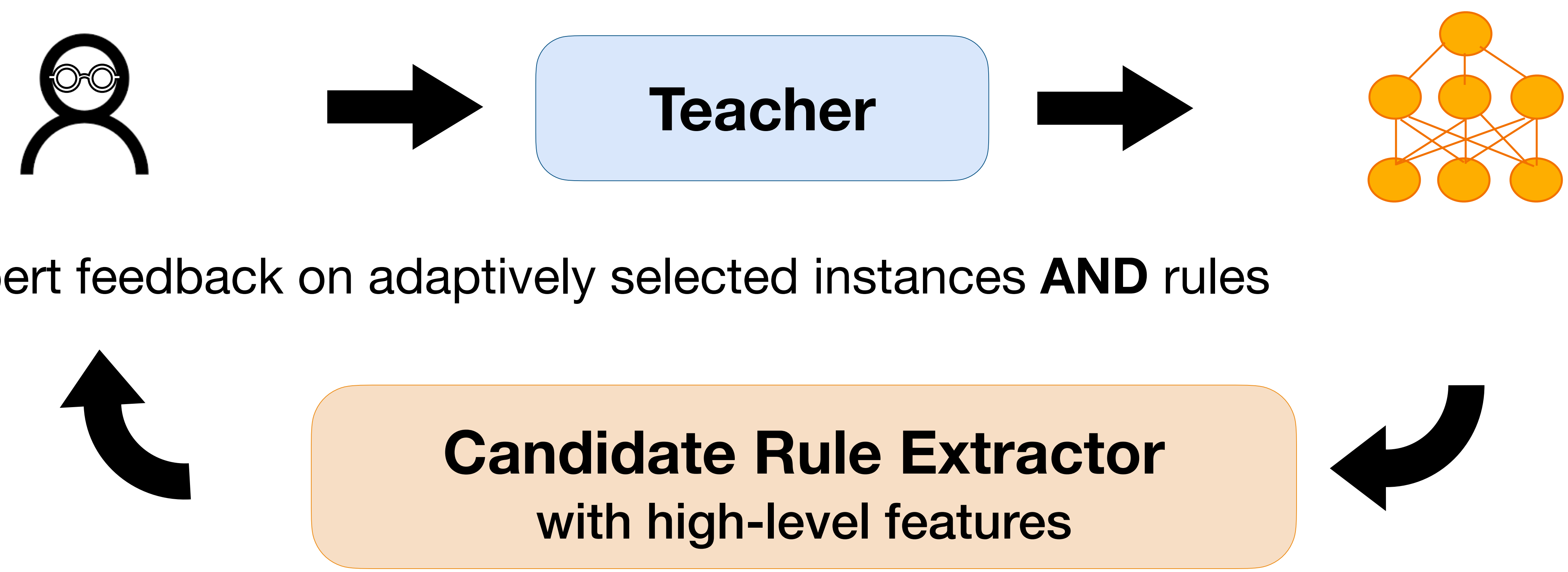
Our analysis and interactive framework for efficient machine teaching

- We reveal patterns that could inform guidelines for rule creation
  - Increasing Teacher's F1 does **not** necessarily increase Student's F1
  - Teacher's **precision** is more important than **coverage**
- We present an interactive machine teaching framework:



Our analysis and interactive framework for efficient machine teaching

- We reveal patterns that could inform guidelines for rule creation
  - Increasing Teacher's F1 does **not** necessarily increase Student's F1
  - Teacher's **precision** is more important than **coverage**
- We present an interactive machine teaching framework:



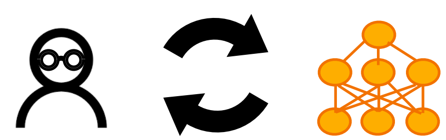
Our analysis and interactive framework for efficient machine teaching

- We reveal patterns that could inform guidelines for rule creation
  - Increasing Teacher's F1 does **not** necessarily increase Student's F1
  - Teacher's **precision** is more important than **coverage**
- We present an interactive machine teaching framework:

Rule Extractor	**F1 (6 datasets)**
n-gram rules	64.6
n-gram + linguistic + prompt rules	**77.6**

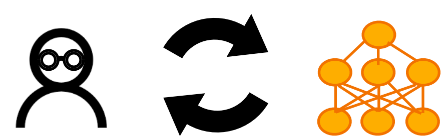
Our analysis and interactive framework for efficient machine teaching

- We reveal patterns that could inform guidelines for rule creation
  - Increasing Teacher's F1 does **not** necessarily increase Student's F1
  - Teacher's **precision** is more important than **coverage**
- We present an interactive machine teaching framework:

Interactive Method 	**F1 (6 datasets)**
Feedback on instances (active learning)	75.7
Feedback on instances + rules (ours)	**79.1**

Our analysis and interactive framework for efficient machine teaching

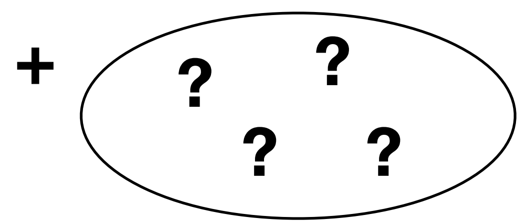
- We reveal patterns that could inform guidelines for rule creation
  - Increasing Teacher's F1 does **not** necessarily increase Student's F1
  - Teacher's **precision** is more important than **coverage**
- We present an interactive machine teaching framework:

Interactive Method 	**F1 (6 datasets)**
Feedback on instances (active learning)	75.7
Feedback on instances + rules (ours)	**79.1**

Feedback on a **single** rule leads to **several** labels

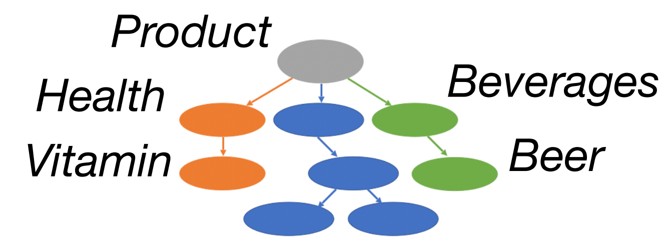
Efficient machine teaching frameworks for NLP

Coarse labels (Ch. 3)



[Karamanolakis et al. WNUT ’19]

Hierarchical taxonomies (Ch. 4)



[Karamanolakis et al. ACL ’20]

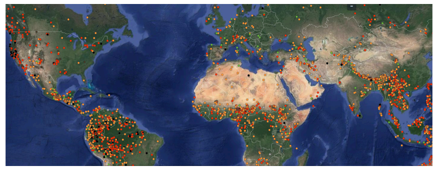
Seed words (Ch. 5)

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

[Karamanolakis et al. EMNLP ’19]

Word translations (Ch. 6)

“injured”



“يارلانغان”

[Karamanolakis et al. Findings of EMNLP ’20]

Labeling rules (Ch. 7, 8)

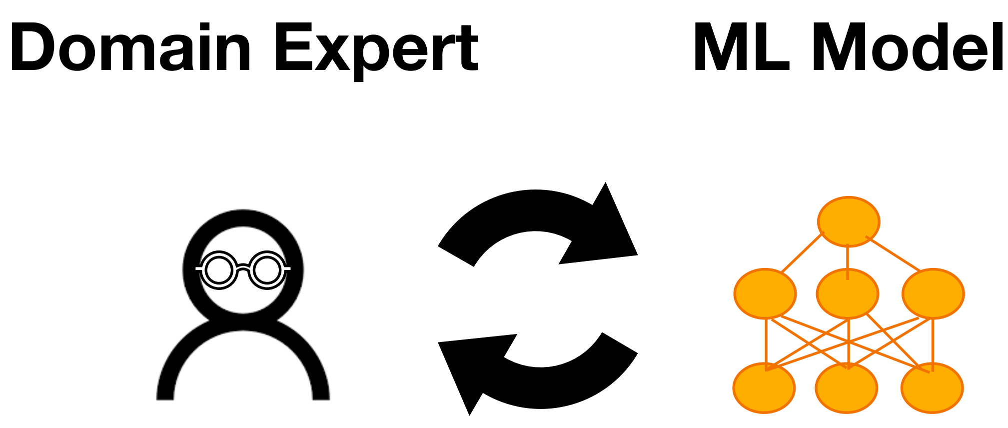
```
def regex_check_out(x):
    return SPAM if re.search("check.*out", x)
```

[Karamanolakis et al. NAACL ’21] + work with Daniel Hsu and Luis Gravano

Contributions of this dissertation

Contributions of this dissertation

- Develop **efficient** frameworks for teaching ML models using high-level supervision

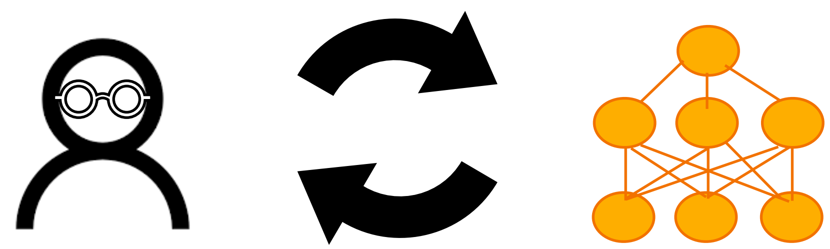


• Coarse labels	[Karamanolakis et al., WNUT '19]
• Hierarchical taxonomies	[Karamanolakis et al., ACL '20]
• Seed words	[Karamanolakis et al., EMNLP '19]
• Word translations	[Karamanolakis et al., Findings of EMNLP '20]
• Labeling rules	[Karamanolakis et al., NAACL '21]

Contributions of this dissertation

- Develop **efficient** frameworks for teaching ML models using high-level supervision

Domain Expert **ML Model**



- Coarse labels [Karamanolakis et al., WNUT ’19]
- Hierarchical taxonomies [Karamanolakis et al., ACL ’20]
- Seed words [Karamanolakis et al., EMNLP ’19]
- Word translations [Karamanolakis et al., Findings of EMNLP ’20]
- Labeling rules [Karamanolakis et al., NAACL ’21]

minimal domain expertise
beyond standard labels

+

generalization power
of deep neural networks

Contributions of this dissertation

- Develop **efficient** frameworks for teaching ML models using high-level supervision
- Integrate ML models into operational systems **without** expensive labeled data

[Karamanolakis et al., WNUT ’19]
[Liu et al., LOUHI ’20]
[Cao et al., SocialNLP ’21]

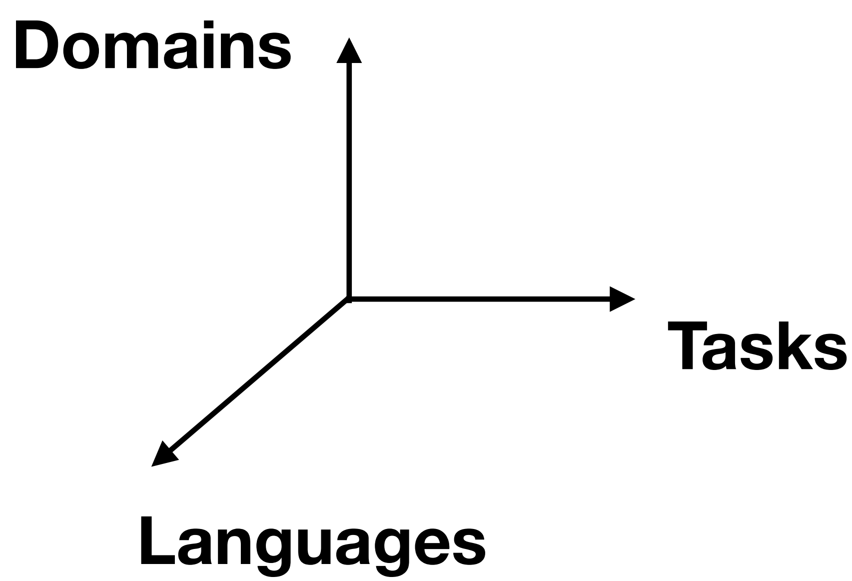
https://publichealth.cs.columbia.edu/

[Karamanolakis et al., ACL ’20]
[Dong et al., KDD ’20]

https://www.amazon.science/blog/building-product-graphs-automatically

Contributions of this dissertation

- Develop **efficient** frameworks for teaching ML models using high-level supervision
- Integrate ML models into operational systems **without** expensive labeled data
- Demonstrate the benefits of high-level supervision for scaling NLP



Contributions of this dissertation

- Develop **efficient** frameworks for teaching ML models using high-level supervision
- Integrate ML models into operational systems **without** expensive labeled data
- Demonstrate the benefits of high-level supervision for scaling NLP
- Facilitate future research on machine teaching

Public code

https://github.com/cu-publichealth/FoodborneML

https://github.com/microsoft/ASTRA

https://github.com/IRI-Bangladesh-Flood-Insurance-Research

New benchmarks

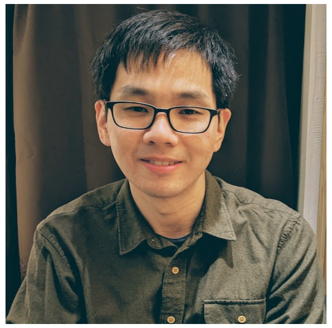
http://aka.ms/walnut_benchmark [Zheng et al., NAACL '22]

https://github.com/allenai/natural-instructions [Wang et al., '22]

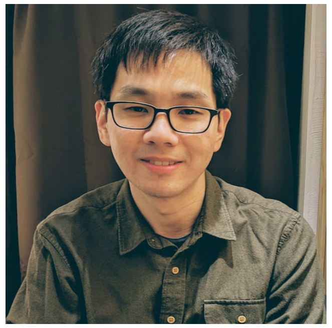
"WALNUT: a benchmark on weakly supervised learning for natural language understanding",
Guoqing Zheng, Giannis Karamanolakis, Kai Shu, Ahmed Hassan Awadallah (NAACL '22)

"Benchmarking generalization via in-context instructions on 1,600+ language tasks" Wang et al., '22

Thank you all!

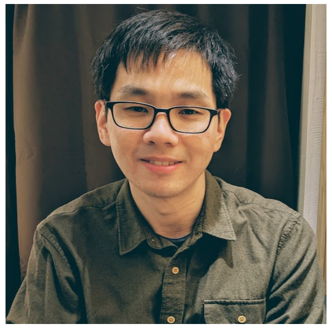


Thank you all!

Thank you all!



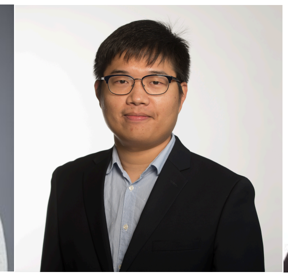
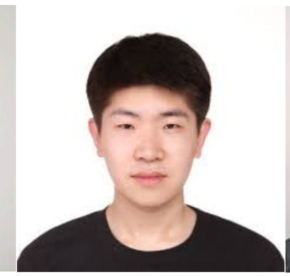



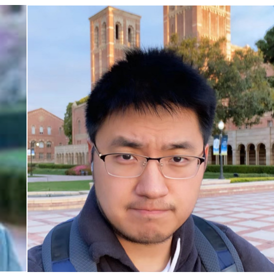
Thank you all!

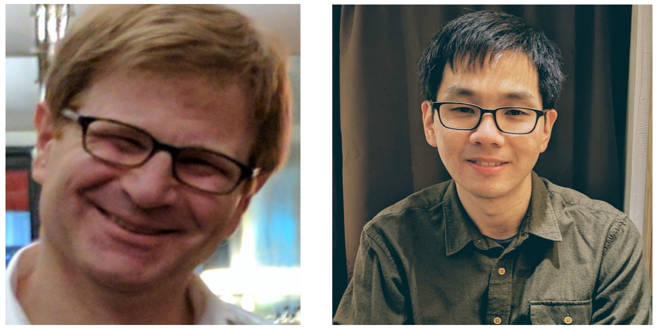




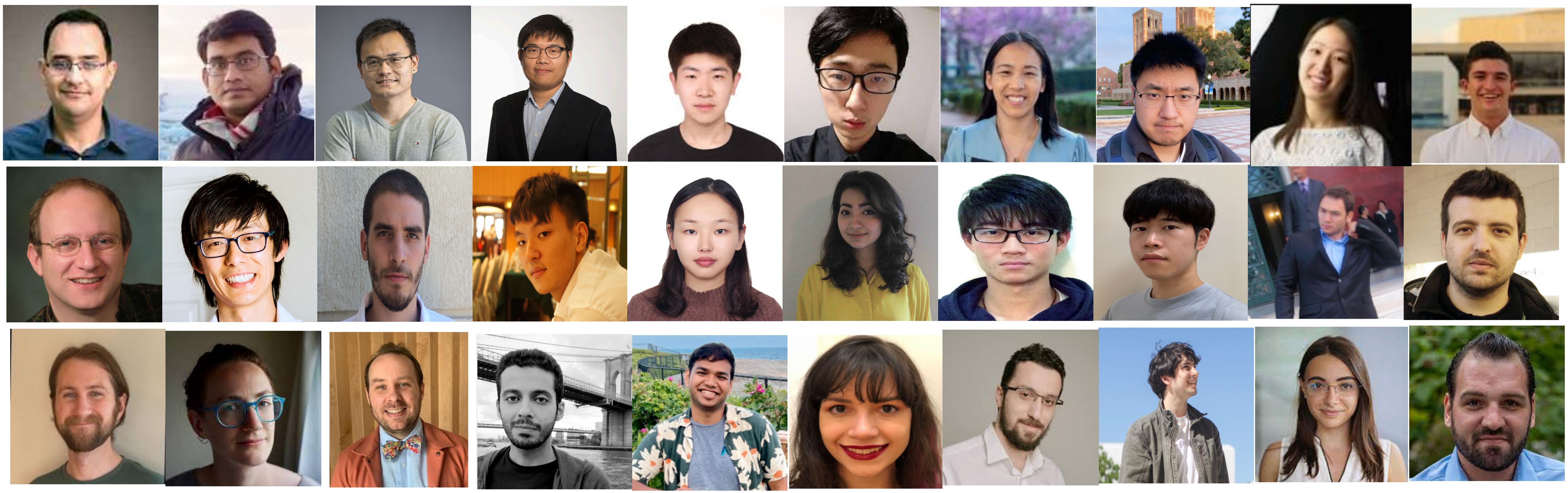


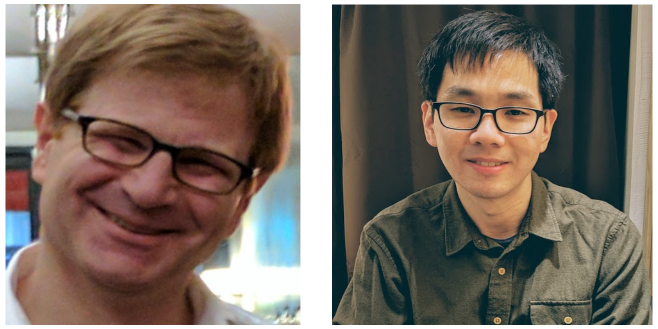



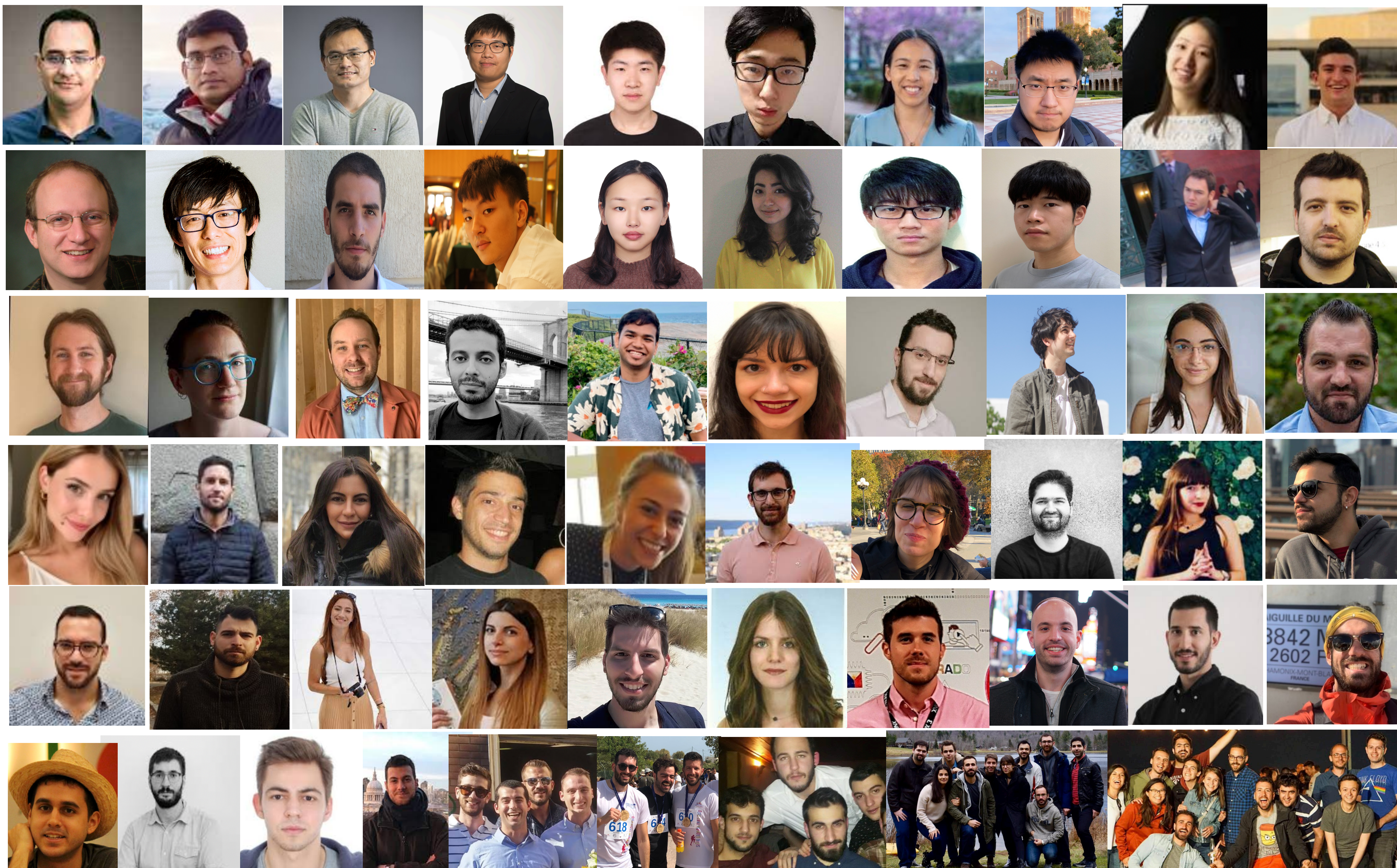


Thank you all!





Thank you all!



Publications

web: https://gkaramanolakis.github.io/

email: gkaraman@cs.columbia.edu

1. **"Leveraging just a few keywords for fine-grained aspect detection through weakly-supervised co-training"**
   Giannis Karamanolakis, Daniel Hsu, Luis Gravano. (EMNLP '19) Health Public Health
2. **"Weakly supervised attention networks for fine-grained opinion mining and public health"**
   Giannis Karamanolakis, Daniel Hsu, Luis Gravano. (WNUT@EMNLP '19) Health Public Health
3. **"TXtract: taxonomy-aware knowledge extraction from thousands of product categories"**
   Giannis Karamanolakis, Jun Ma, Xin Luna Dong. (ACL '20)
4. **"AutoKnow: Self-Driving Knowledge Collection for Products of Thousands of Types"**
   Xin Luna Dong, Xiang He, Andrey Kan, Xian Li, Yan Liang, Jun Ma, Yifan Ethan Xu, Chenwei Zhang, Tong Zhao, Gabriel Blanco Saldana, Saurabh Deshpande, Alexandre Michetti Manduca, Jay Ren, Surender Pal Singh, Fan Xiao, Haw-Shiuan Chang, Giannis Karamanolakis, Yuning Mao, Yaqing Wang, Christos Faloutsos, Andrew McCallum, Jiawei Han (KDD '20)
5. **"Detecting Foodborne Illness Complaints in Multiple Languages Using English Annotations Only"**
   Ziyi Liu, Giannis Karamanolakis, Daniel Hsu, Luis Gravano. (LOUHI@EMNLP '20) Health Public Health
6. **"Cross-lingual text classification with minimal resources by transferring a sparse teacher"**
   Giannis Karamanolakis, Daniel Hsu, Luis Gravano. (Findings of EMNLP '20) Health Public Health COLUMBIA CLIMATE SCHOOL INTERNATIONAL RESEARCH INSTITUTE FOR CLIMATE AND SOCIETY
7. **"Quantifying the effects of COVID-19 on restaurant reviews",** Health Public Health
   Ivy Cao, Zizhou Liu, Giannis Karamanolakis, Daniel Hsu, Luis Gravano (SocialNLP '21)
8. **"Self-training with weak supervision",** Microsoft
   Giannis Karamanolakis, Subhabrata Mukherjee, Guoqing Zheng, Ahmed Hassan Awadallah, (NAACL '21)
9. **"WALNUT: a benchmark on weakly supervised learning for natural language understanding",** Microsoft
   Guoqing Zheng, Giannis Karamanolakis, Kai Shu, Ahmed Hassan Awadallah (NAACL '22)
10. **"Flood event extraction from news media to support satellite-based flood insurance",**
    Jakwanul Safin, Tejit Pabari, Giannis Karamanolakis, Max Mauerman, and Beth Tellman (submitted) COLUMBIA CLIMATE SCHOOL INTERNATIONAL RESEARCH INSTITUTE FOR CLIMATE AND SOCIETY
11. **"Benchmarking generalization via in-context instructions on 1,600+ language tasks"** (submitted)