

# Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training

**Giannis Karamanolakis**  
Columbia University  
gkaraman@cs.columbia.edu

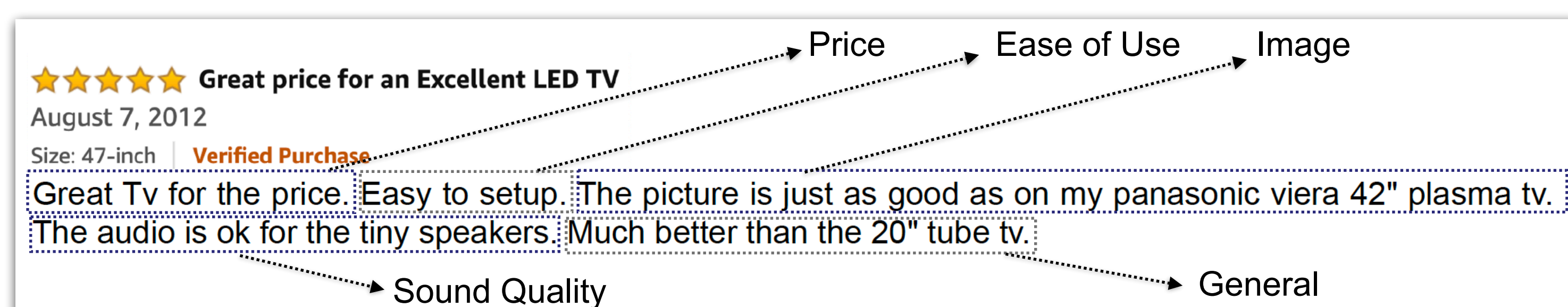
**Daniel Hsu**  
Columbia University  
djhsu@cs.columbia.edu

**Luis Gravano**  
Columbia University  
gravano@cs.columbia.edu

## Aspect Detection in Online Reviews

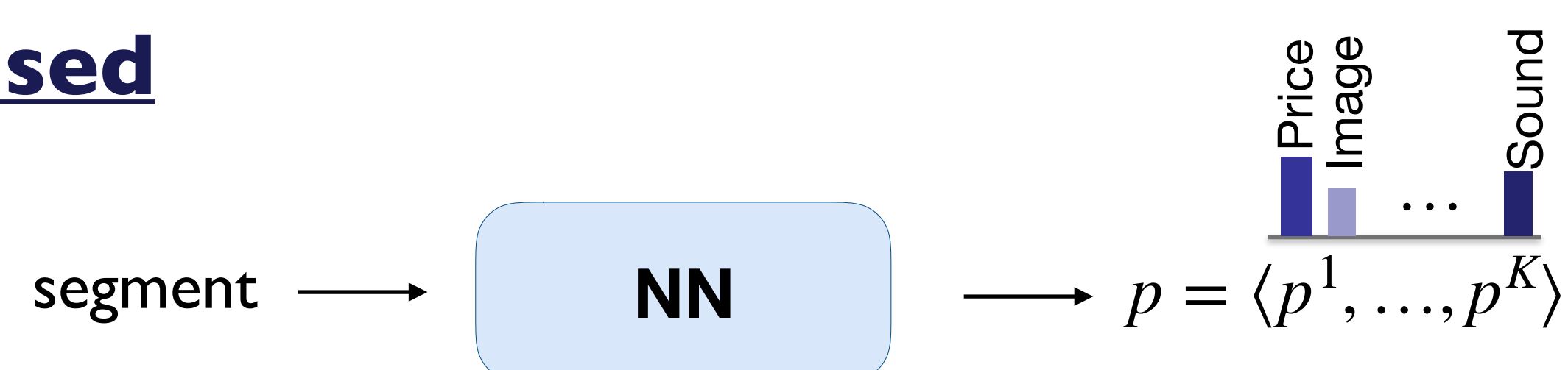
**Goal:** Identify which product **aspects** (e.g., price, quality) are discussed in individual **segments** (e.g., sentences, clauses) of a review.

**Key task in:** Sentiment Analysis, Opinion Mining, Review Summarization.



## Neural Networks For Aspect Detection

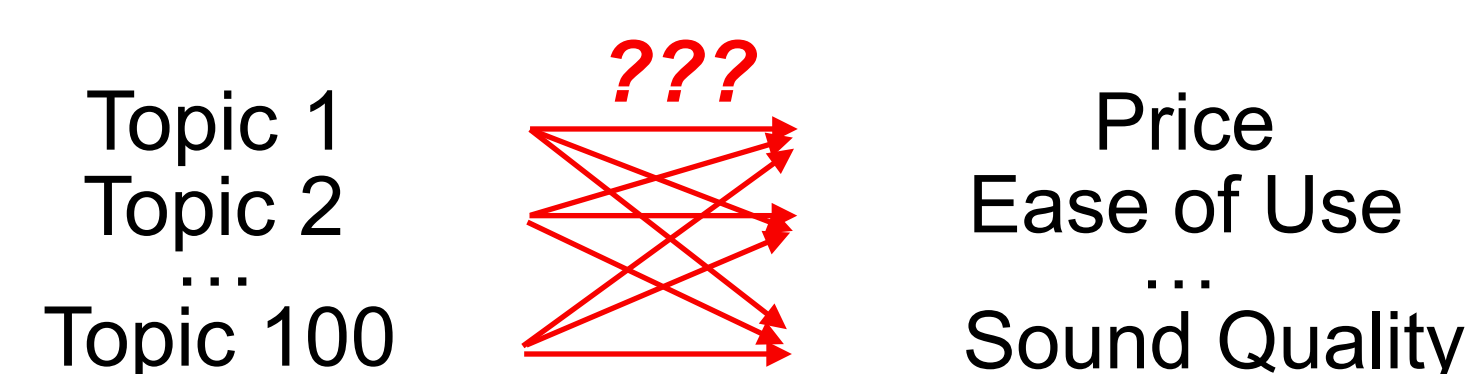
### I. Supervised



**Issue: Ground-truth aspect labels are not readily available.**

- Manual segment annotation is expensive and not scalable.

### 2. Unsupervised (Neural Topic Models)



**Issue: May not capture the K aspects of interest.**

### 3. Weakly Supervised: Learning with Seed Words

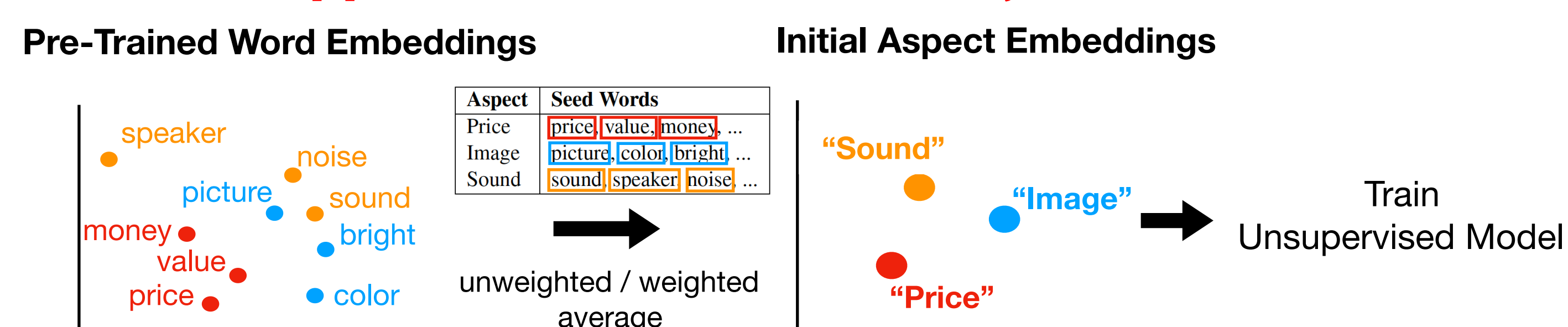
**Idea:** Leverage descriptive **seed words** as a “weak” source of supervision.

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

Easier to collect **seed words** than **ground-truth aspect labels**.

- manually (humans) or automatically (small number of labeled segments).

**Issue: Previous approaches use seed words just for initialization.**



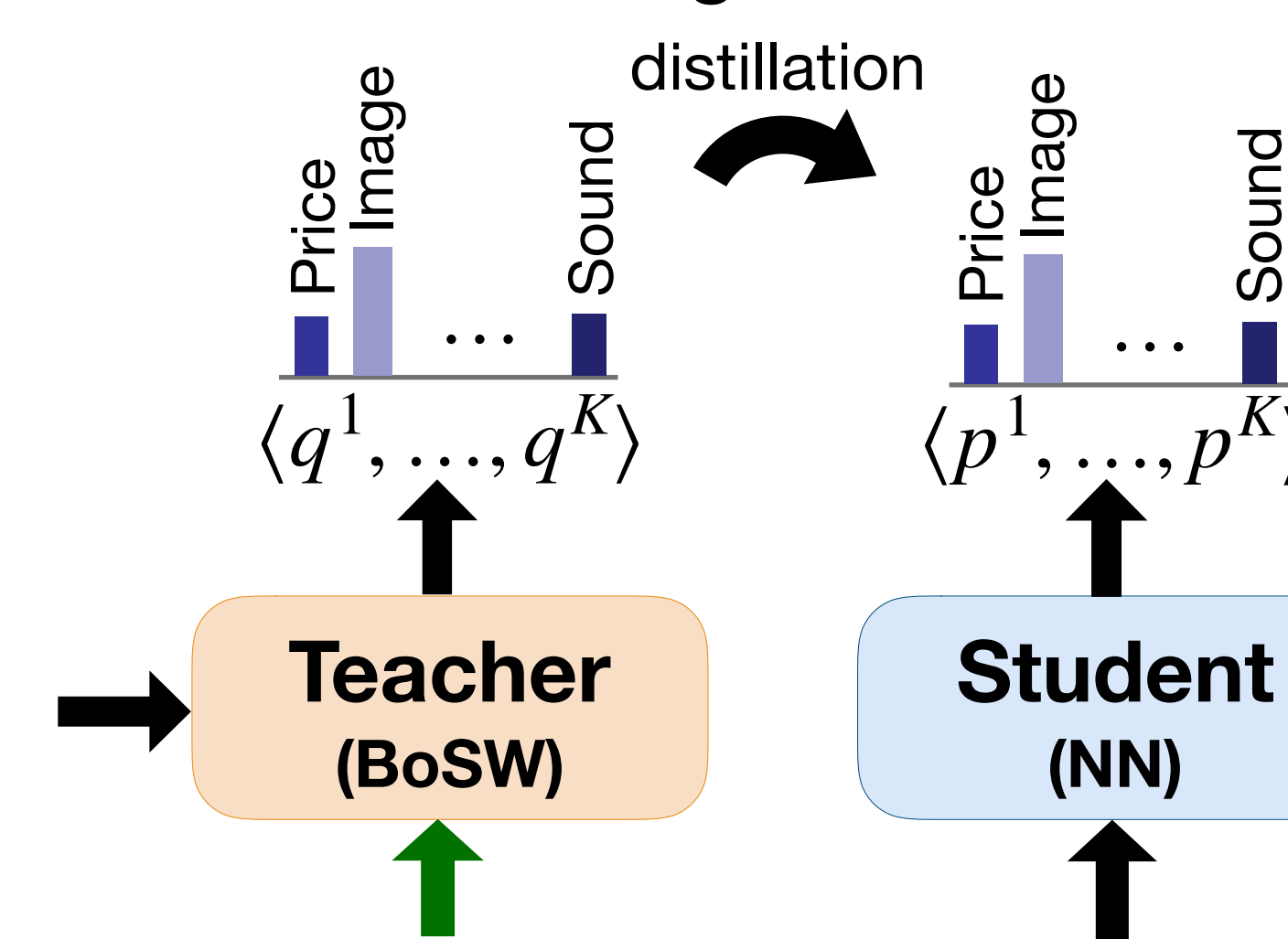
- Individual seed words are **not** used during training.

## Weakly Supervised Co-Training

A **teacher-student** framework for training neural networks.

- Teacher**
- Bag-of-Seed-Words
  - “Counts” seed words

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



- Student**
- Embedding-based NN
  - Trained through distillation [Ba & Caruana '14; Hinton et al. '15]

1. Embedding (e.g., BERT)
2. Classification

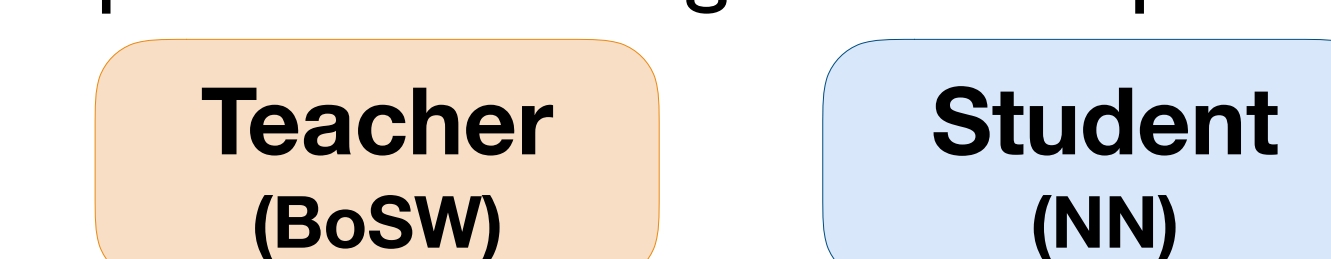
input segment: “The **picture** looks very pixelated when playing blu-ray movies”

- Teacher** only considers the (few) seed words.
- Student** also considers the **context** of the seed words.
- Student** predicts aspects even if no seed words appear.

Dropping seed words: “The **<UNK>** looks very **pixelated** when playing **blu-ray** movies”

- Improved Teacher:** estimates seed word qualities ... using Student’s predictions!

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



Model Bootstrapped EM [Khetan et al. '18]

**We leverage (noisy) seed words as weak NN supervision**

## Experiments

### Datasets

- Amazon reviews from 6 product domains (9 aspects)
- Restaurant reviews from 6 languages (12 aspects)

### Training

- 1M unlabeled segments
- 30 seed words per aspect (same across models)

### Baselines

- LDA-Anchors [Lund et al. '17]
- Aspect-Based Auto-Encoder (ABAE) [He et al. '17]
- Multi-seed Aspect Extractor (MATE) [Angelidis & Lapata '18]

### Results

- Teacher** and **Student-BoW** outperform previous approaches.
- Student-\*** outperforms **MATE** across all domains and languages (best model: **Student-BERT**).
- Most improvement achieved after **one round** of co-training.

