Training Neural Networks for Aspect Extraction Using Descriptive Keywords Only

Giannis Karamanolakis
Columbia University
gkaraman@cs.columbia.edu

Daniel Hsu
Columbia University
djhsu@cs.columbia.edu

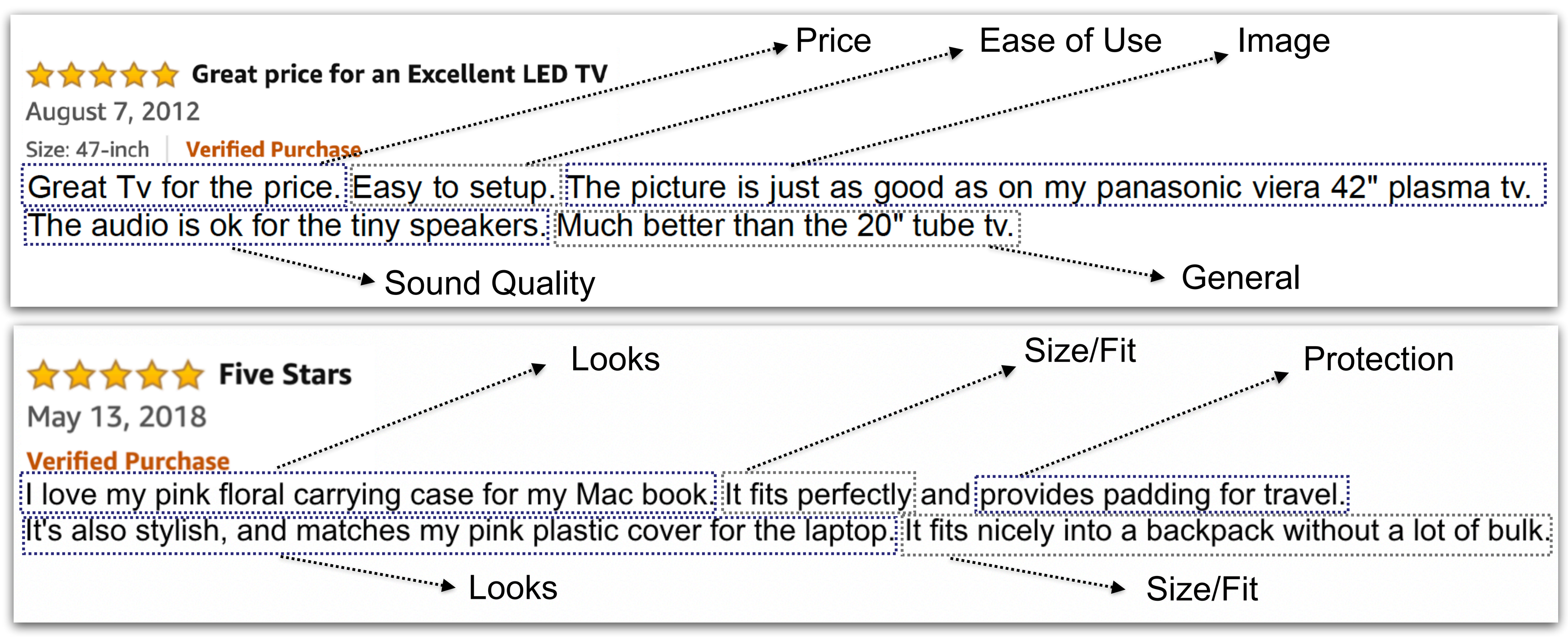
Luis Gravano
Columbia University
gravano@cs.columbia.edu

LEARNING WITH LIMITED LABELED DATA
ICLR 2019

Aspect Extraction in Online Product Reviews

Goal: Identify which product **aspects** (e.g., price, quality) are discussed in individual **segments** (e.g., sentences, clauses) of the product's reviews.

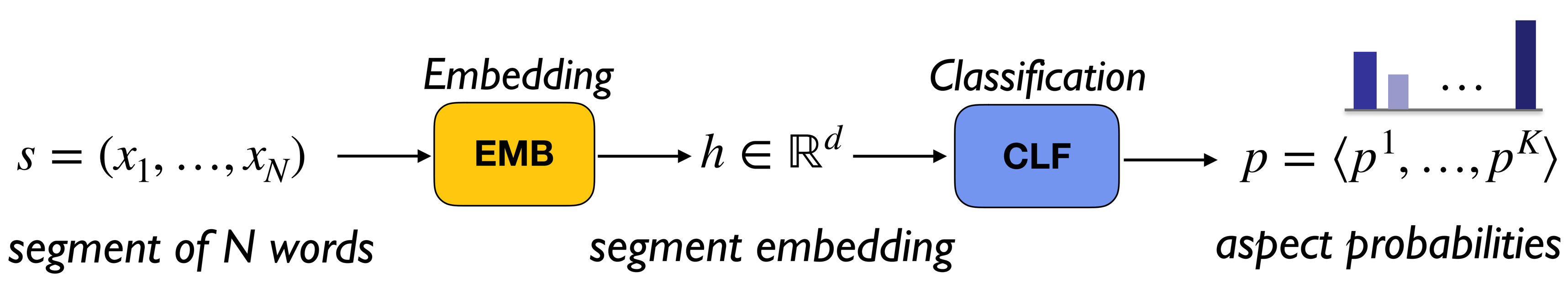
Key task in: Sentiment Analysis, Opinion Mining, Review Summarization.



Neural Networks For Aspect Extraction

1. Supervised

Deep neural networks learn good representations of text for classification.

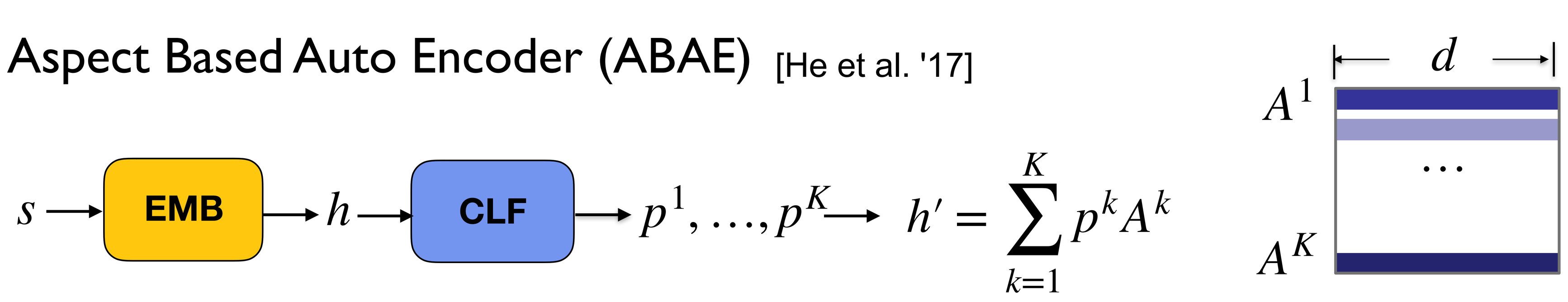


Issue: Ground-truth aspect labels are not inherently available.

- Manual segment annotation is expensive and not scalable.

2. Unsupervised (Neural Topic Models)

Aspect Based Auto Encoder (ABAE) [He et al. '17]



- Aspect embeddings $A^k \in R^d$ initialized using **k-means**.

Issue: ABAE fails to capture the particular aspects of interest.

- Probabilities $p^1, \ldots, p^K$ are **not** mapped to the K aspects of interest.

3. Weakly Supervised

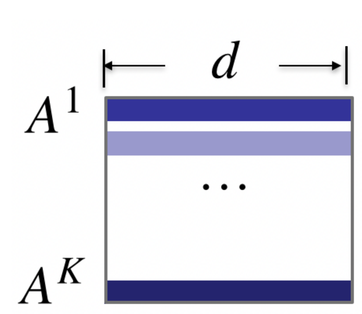
Idea: Use seed words as a "weak" source of supervision.

- Seed words: descriptive keywords.
- Assume ~30 seed words per aspect are available.
- Easier to collect than thousands of labels.

Aspect	*Seed Words*
Price	price, value, money, …
Image	picture, color, quality, …

Multi-seed Aspect Extractor (MATE) [Angelidis & Lapata '18]

- Extends ABAE.
- Aspect embeddings A^k initialized using **seed words**.
- A^k: average of seed word embeddings.

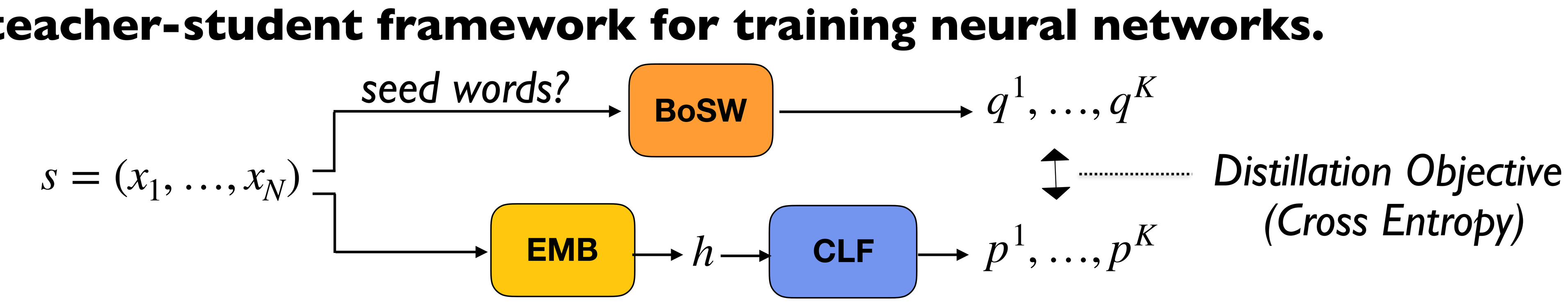


Question: Does MATE effectively leverage the seed words?

- Seed words are used just for the **initialization** of aspect embeddings.

Our Weakly Supervised Approach

A teacher-student framework for training neural networks.



- **Teacher**: A bag-of-seed-words (BoSW) logistic regression classifier
  - Binary weights: *"if a seed word appears, increase the corresponding aspect score."*

The teacher only considers the (few) seed words.

- **Student:** An embedding-based neural network
  - Mimics the teacher's predictions (distillation loss).
  - Encouraged to generalize via regularization (L2 + dropout).

The student also considers the (many) non-seed words.

Example 1: …For fifty bucks, it was a great deal… *(0 seed words)*
Example 2: …The picture looks very pixelated when playing bluray movies… *(2 seed words)*

- Experiment: We completely drop seed words (DSW) from the student's input.

Example 2: …The UNK UNK very pixelated when playing bluray movies…

We effectively leverage a few seed words as supervision for training neural networks.

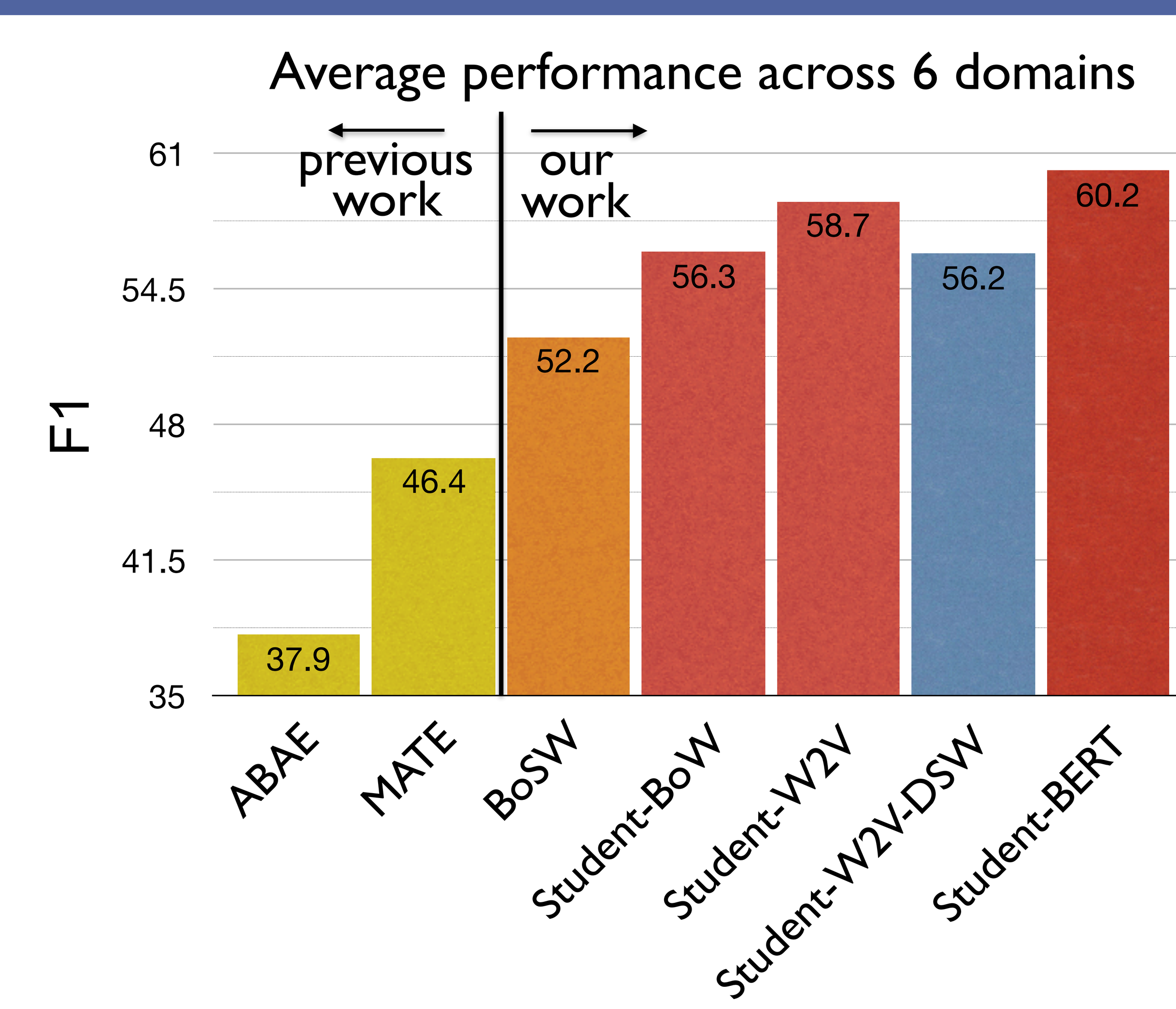
Experiments

Datasets

- Amazon product reviews
- 6 product domains
- 9 aspects per domain
- 30 seed words per aspect

Segment Embedding

- **BoW:** bag-of-words representation
- **W2V:** avg of word2vec embeddings
- **BERT:** Google's BERT (avg pooling)



Results

- **BoSW** outperforms **MATE** using only seed words.
- **Student-BoW**/**Student-W2V** outperform **BoSW** with proper regularization.
- **Student-W2V** associates non-seed words to aspects: even when dropping the seed words from the student's input (**Student-W2V-DSW**) it outperforms **BoSW**.
- **Student-BERT** outperforms **MATE** by 30%.

Conclusions & Future Work

- **Seed words** are effectively leveraged into **BoW classifiers**.
- **BoW classifiers** are used in a distillation framework to train **neural networks**.
- **Student** generalizes better than **teacher**: associates **non-seed words** with aspects.
- On-going work: handling **noisy** seed words & learning **better** seed words.

Acknowledgements: This material is based upon work supported by the National Science Foundation under Grant No. IIS-15-63785.