Weakly Supervised Attention Networks for Fine-Grained Opinion Mining and Public Health

COLUMBIA UNIVERSITY

Giannis Karamanolakis
Columbia University
gkaraman@cs.columbia.edu

Daniel Hsu
Columbia University
djhsu@cs.columbia.edu

Luis Gravano
Columbia University
gravano@cs.columbia.edu

W-NUT 2019
EMNLP-IJCNLP 2019

Fine-Grained Analysis of Online User-Generated Reviews

User reviews: Consist of multiple segments (e.g., sentences, clauses).

Motivation: Different **segments** of a review may have different labels.

Carmine's Italian Restaurant
$$ · Italian, Venues & Event Spaces
200 W 44th St
New York, NY 10036

Example: Opinion Mining

→ **Overall Opinion:** negative

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.

→ **Drinks:** positive
→ **Service:** positive
→ **Food:** negative
→ **Price:** negative

→ **Ambience:** positive

Goal: Train **segment-**level classifiers.

Challenge: **No ground truth segment labels.**

- Segment labels are **not** typically available and are **expensive** to obtain.

Weakly Supervised Learning

Goal: Train **segment**-level classifiers using only **review** labels.

+ Review labels are already available.
- Review labels may not be directly relevant for segment classification.

Multiple Instance Learning (MIL)

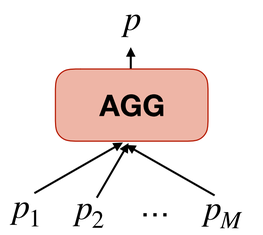
Bag: review $r = s_1, \ldots s_M$ with observed label p.
Instances: segments s_i with unobserved labels p_i .

MIL Assumption ("at least one"): $p = 1$ iff $\exists i : p_i = 1$

MIL Assumption (relaxed): $p = \mathrm{AGG}(p_1, \ldots, p_M)$

How to choose the **AGG** function?
… depends on the task.

Witness rate: proportion of positive instances in positive bags.



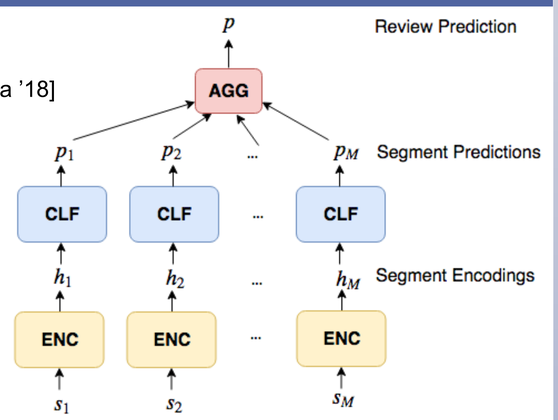
review label

segment labels

Multiple Instance Learning Networks

Hierarchical MIL Networks:
[Kotzias et al. '15, Pappas et al., '17, Angelidis and Lapata '18]

1. **ENC:** encode segments.
2. **CLF:** classify segments.
3. **AGG:** aggregate segment labels.
   - average [Kotzias et al. '15]
   - weighted average [Pappas et al., '17, Angelidis and Lapata '18]



Training: use review labels.
Evaluation: outperform simpler MIL-based classifiers (e.g., LogReg)
Question: **Where do performance gains stem from?**

1. Hierarchical structure of MIL networks (AGG)?
2. Representationsal power of deep learning components?

New baseline: Non-hierarchical deep networks may outperform hierarchical MIL Networks!

When? When **AGG** functions are not suitable for the task at hand.

We train Hierarchical Sigmoid Attention Networks for segment classification using review labels only!

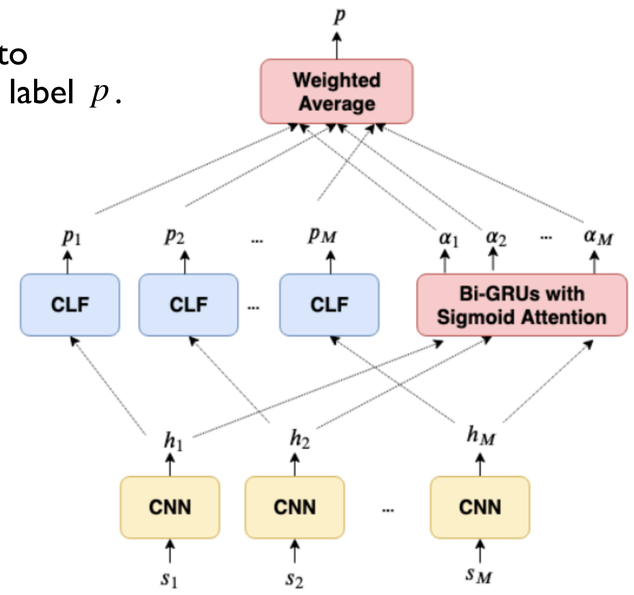
Our Weakly Supervised Approach

We use **sigmoid attention** as **AGG** function in **MIL networks.**

Sigmoid attention: allows multiple segments to contribute with different weights α_i to the review label p.

MIL attention $p = \mathrm{AGG}(p_1, \ldots, p_M)$ ≠ standard attention $h = \mathrm{AGG}(h_1, \ldots, h_M)$

Sigmoid is more suitable than **softmax** when **multiple** s_i are relevant to p.



MIL-Sigmoid

$$\alpha_i = p(z_i = 1 \mid e_1, \ldots, e_M) = \frac{1}{1 + \exp(-e_i)}$$

MIL-Softmax

$$\alpha_i = p(z = i \mid e_1, \ldots, e_M) = \frac{\exp(e_i)}{\sum_{i=1}^{M} \exp(e_i)}$$

Experiments

Models:

- Non-Hierarchical Rev-*:

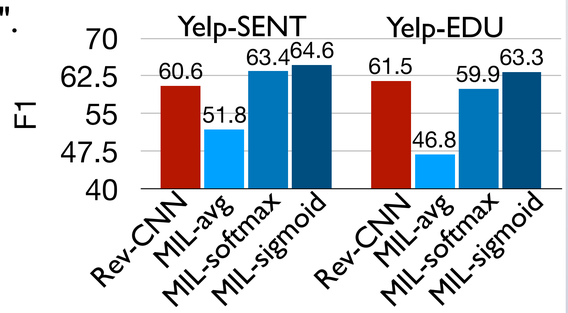
Train: r → ENC → h → CLF → p
Test: s_i → h_i → p_i

- Hierarchical MIL-* networks with various AGG functions: average, softmax, sigmoid.

1. Segment-Level Sentiment Classification

3-class classification: "positive", "neutral", "negative".
Dataset: OPOSUM - 4 datasets (Yelp/IMDB reviews).
Results:

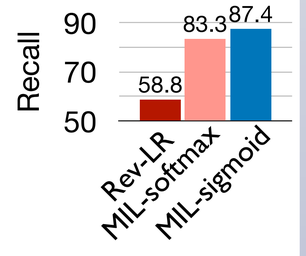
- Choice of AGG is crucial.
- Rev-CNN outperforms MIL-avg/softmax (4/4).
- MIL-sigmoid outperforms MIL-softmax.



2. Foodborne Illness Discovery

Goal: Detect foodborne illness from Yelp restaurant reviews.
Dataset: Yelp reviews labeled ("Sick"/"Not Sick") by epidemiologists.
Results:

- **MIL-sigmoid:** 48.6% higher recall than **Rev-LR** (Effland et al. 2018)
- **Highlighting important segments:**



✓ 0.00 The service was good, it was overall fine.
✓ 0.00 That is- until I got home and me and boy friend spent the rest of the day/night and into the morning hunched over or sitting on the toilet!
😊 0.18 **I have never experienced such violent food poisoning in my life!** → **"I have never experienced such violent food poisoning in my life"**
✓ 0.00 That was the only place we ate or drank anything at that day, so I know it was from this restaurant.

We increase the chances of identifying previously unknown foodborne outbreaks!

Summary of Contributions

1. We show that **non-hierarchical** baselines may outperform previous **MIL** networks.
2. We identify that **sigmoid** attention is more appropriate than **softmax** attention for MIL.
3. We demonstrate that our model could have positive impact for **public health**.

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