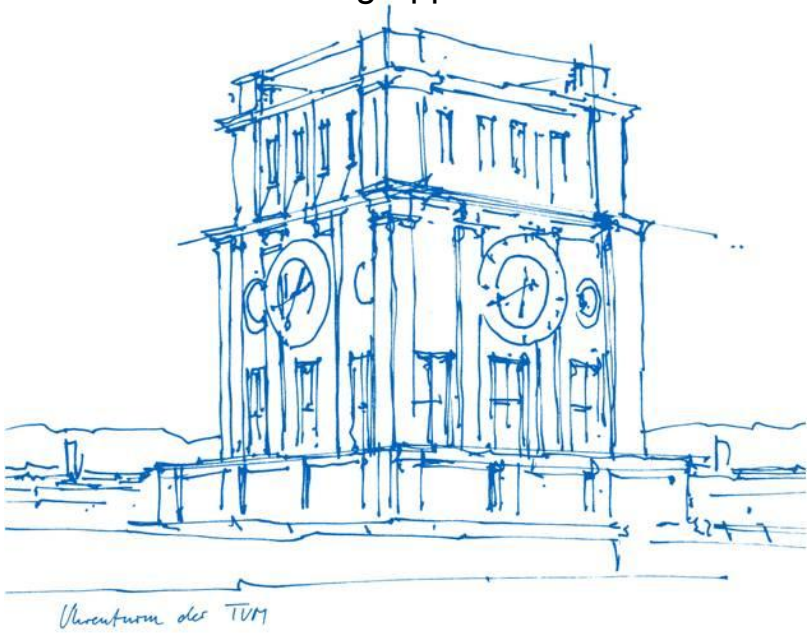
Opinion Mining Lab Group 1.3

Topic: Weakly Supervised Aspect Extraction Using a Student-Teacher Co-Training Approach

Group members: Jingpei Wu, Ke Xin Chen, Kevin George

31.05.2021



For the past two weeks

- Fixed seedwords methods as suggested in last lab session.
- Manually combined and selected seedwords from different methods.
- Ran and tested official GitHub repo from [2] and unofficial code [3] for paper [1]
  - [2] was written in Python2, manually modified the code to Python3
  - Applied dataset from repo [2] to repo [3], to make sure the model performs the same as described in the paper [1].
  - Tried to understand and fix repo [3], as it is unofficial and some methods did not performed the same as described in the paper [1]

[1] Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training, Karamanolakis et al.
[2] Summarizing Opinions: Aspect Extraction Meets Sentiment Prediction and They Are Both Weakly Supervised, Angelidis et al.
[3] https://github.com/aqweteddy/LeverageJustAFewKeywords

Current Selected Seedwords

Price: price cost expensive money pay demand charge expensive extra profit
Taste: taste better flavor delicious tasty sweeter flavour feel like try
Animal: animal cow chicken feed outdoors cage hen support beef goat
Dairy: milk cow raw drink dairy almond baby soy cheese cream
Health life: vegan eat meat animal healthy diet chicken fruit vegetable ingredient
Chemical: antibiotic chemical pesticide fertilizer herbicide toxic synthetic science residue natural
Business: market consumer industry supply production business customer grocery economic supply
Label: label certified ingredient company choice brand business product factory origin
Information: research information point article study target problem question reason comment
Env: environment organic water fruit sustainable soil destructive equally farm life
Freshness & quality: fresh quality nutrition vitamin source nature science way nutrient organic
Farming: plant seed crop farm population grow grass fruit vegetable land
General: gmo monsanto non online start year day thousand thank lot

Test Results: Teacher

- bag of seed words model (30 seed words each aspect)
- no trainable parameters
- metric: micro-averaged F1

Method	Product Review Domain						AVG
	Bags	Keyboards	Boots	Headsets	TVs	Vacuums	
LDA-Anchors (Lund et al., 2017)	33.5	34.7	31.7	38.4	29.8	30.1	33.0
ABAE (He et al., 2017)	38.1	38.6	35.2	37.6	39.5	38.1	37.9
MATE (Angelidis and Lapata, 2018)	46.2	43.5	45.6	52.2	48.8	42.3	46.4
MATE-unweighted	41.6	41.3	41.2	48.5	45.7	40.6	43.2
MATE-MT (best performing)	48.6	45.3	46.4	54.5	51.8	47.7	49.1
Teacher	55.1	52.0	44.5	50.1	56.8	54.5	52.2
experiments using Teacher	52.68	54.12	50.55	52.15	58.12	56.14	

[1] Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training, Karamanolakis et al.

[2] https://github.com/aqweteddy/LeverageJustAFewKeywords

Change to Reference Code

- change calc_z

$$z_j^k = \frac{\sum_{i=1}^{N} \mathbb{1}\{c_i^j > 0\}\mathbb{1}\{t_i = k\}}{\sum_{k'} \sum_{i=1}^{N} \mathbb{1}\{c_i^j > 0\}\mathbb{1}\{t_i = k'\}}.$$

```
r = torch.stack([(bow[torch.where(idx == k)] > 0).float().sum(0)
                 for k in range(num_asp)])  # [asp_cnt, bow_size]
# bsum = r.sum(-1).view(-1, 1)    # [asp_cnt, 1]
bsum = r.sum(0).view(1, -1)     # [1, bow_size]
bsum = bsum.masked_fill(bsum == 0., 1e-10)
z = r / bsum
```

- general aspect assignment

```
mask = bow.sum(1) == 0  # bow: [B, bow_size] -> mask: [B]
# result[mask, self.general_asp] = 1 # pretend that general words appear once
result = torch.softmax(result, -1)
result[mask, :] = 0
result[mask, self.general_asp] = 1
```

[1] https://github.com/aqweteddy/LeverageJustAFewKeywords

Test results: Training

- Dataset: OPOSUM

Method	Product Review Domain						AVG
	Bags	Keyboards	Boots	Headsets	TVs	Vacuums	
Student-BoW	57.3	56.2	48.8	59.8	59.6	55.8	56.3
Student-W2V	59.3	57.0	48.3	**66.8**	**64.0**	57.0	58.7
Student-W2V-RSW	51.3	57.2	46.6	63.0	62.1	57.1	56.2
Student-ATT	60.1	55.6	49.9	66.6	63.4	58.2	58.9
Student-BERT	**61.4**	**57.5**	**52.0**	66.5	63.0	**60.4**	**60.2**

original edition (3 epochs)	50.5 (*)	49.8
change calc_z (3 epochs)	56.5	
general aspect assignment (3 eps)	56.8	

(*) README report 59, different batch size

[1] Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training, Karamanolakis et al.
[2] https://github.com/aqweteddy/LeverageJustAFewKeywords

Current Challenges

- Our test results are different than the results provided by the paper.

- Further idea verification: the unofficial repo applied teacher-student method to each batch, which is different than described in the paper. As the teacher method should run over all dataset first, then apply student method.

- Readme in unofficial repo is not clear, need more testing.

- Created a dataset contain each sentence with its aspect based on annotated dataset, for further verify on model performance.

Plans for the next two weeks

- Fix current code until get similar results as provided in the paper.
- Apply our own dataset.
- Add options for word2vec.