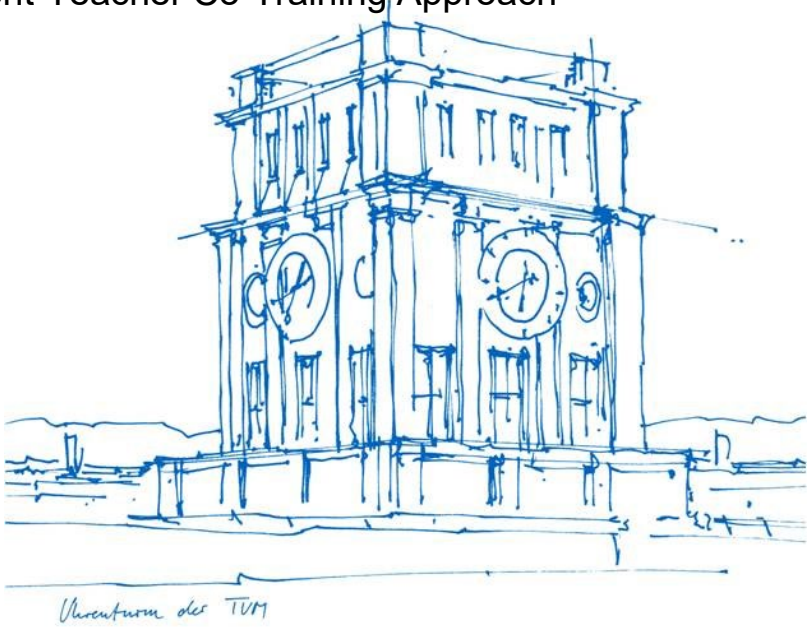
Opinion Mining Lab Group 1.3

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

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

19.04.2021



Agenda

- Problem Definition by Kexin Chen

- Specific Tasks
  - Pre-trained Word Embeddings and Seed Words Selection by Jingpei Wu
  - Iterative Seed Word Distillation (ISWD) - Co Training by Kevin George

Problem Definition

What is our task?

- Extract various aspects of the organic food products from social media comments by using a weak supervision method and specifically in a student-teacher co-training approach.

What is Weak Supervision?

- It is a method to minimize the necessary annotation effort.

What is a student-teacher co-training approach?

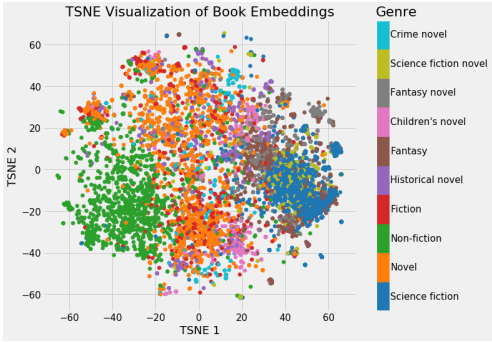
- The teacher is a bag-of-words classifier based solely on provided seed words.
- The student is an embedding-based neural network trained on data “softly” labeled by the teacher.
- Co-training is a classic multi-view learning method for semi-supervised learning.

Pre-trained Word Embeddings and Seed Words Selection

- several pre-trained word embeddings, such as word2vec, GloVe, BERT, etc.
- word embeddings fine-tuning on our own dataset to capture more task-specific meanings
- dataset: provided English organic dataset

- K clusters represented by centroids in the embedding space with K-means
- words that close to centroid as seed words

Aspect	Seed Words
Price (EN)	price, value, money, worth, paid
Image (EN)	picture, color, quality, black, bright
Food (EN)	food, delicious, pizza, cheese, sushi
Drinks (FR)	vin, bière, verre, bouteille, cocktail
Ambience (SP)	ambiente, mesas, terraza, acogedor, ruido



Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training, Karamanolakis et al.
https://devopedia.org/word-embedding

Iterative Seed Word Distillation (ISWD) - Co Training

1. Apply the teacher on unlabeled training segments to get predictions (without considering seed word qualities).
2. Train the student using the teacher's predictions (optimize the cross entropy between the teacher's (soft) predictions and the student's predictions)
3. Apply the student in the training data to get predictions.
4. Update the seed word quality parameters using the student's predictions.
5. Iterate Steps 1-4 until convergence.

- Seeds are 'noisy annotations'.

Reference

- Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training
  - https://arxiv.org/abs/1909.00415
- Weak Supervision: A New Programming Paradigm for Machine Learning
  - http://ai.stanford.edu/blog/weak-supervision/