NQA Seminar: ProPara

Ahmed Sohail Anwari, Anan Schütt

University of Saarland

s8ahanwa, s8anscue@stud.uni-saarland.de

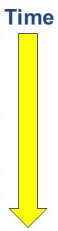
July 14, 2019

Overview

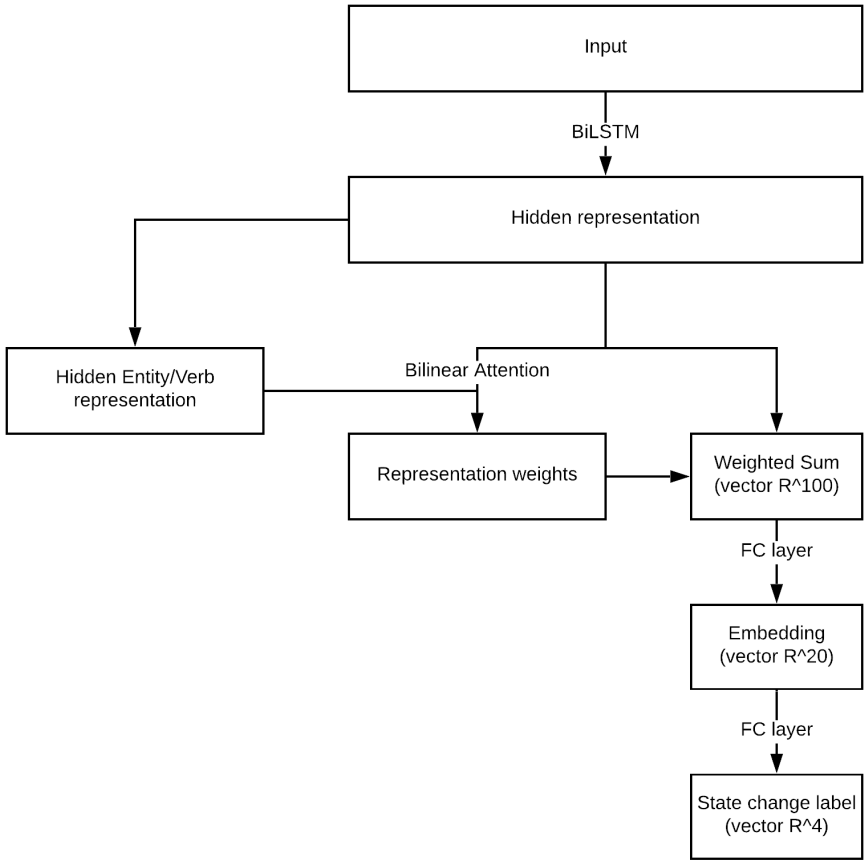
ProPara Dataset

- Procedural text, try to track objects during the paragraph.
- 2k training, 300 testing samples.

		Participants:				
Paragraph (seq. of steps):		**water**	**light**	**CO2**	**mixture**	**sugar**
	state0	**soil**	**sun**	**?**	**-**	**-**
Roots absorb water from soil						
	state1	**roots**	**sun**	**?**	**-**	**-**
The water flows to the leaf.						
	state2	**leaf**	**sun**	**?**	**-**	**-**
Light from the sun and CO2 enter the leaf.						
	state3	**leaf**	**leaf**	**leaf**	**-**	**-**
The light, water, and CO2 combine into a mixture.						
	state4	**-**	**-**	**-**	**leaf**	**-**
Mixture forms sugar.						
	state5	**-**	**-**	**-**	**-**	**leaf**

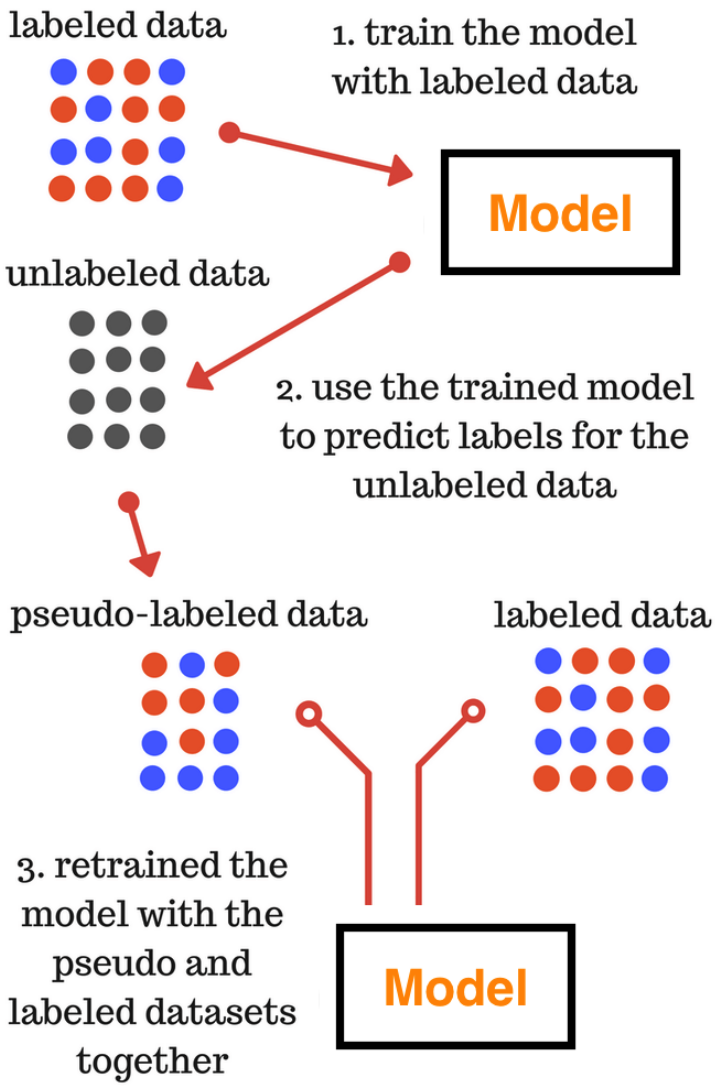


ProLocal

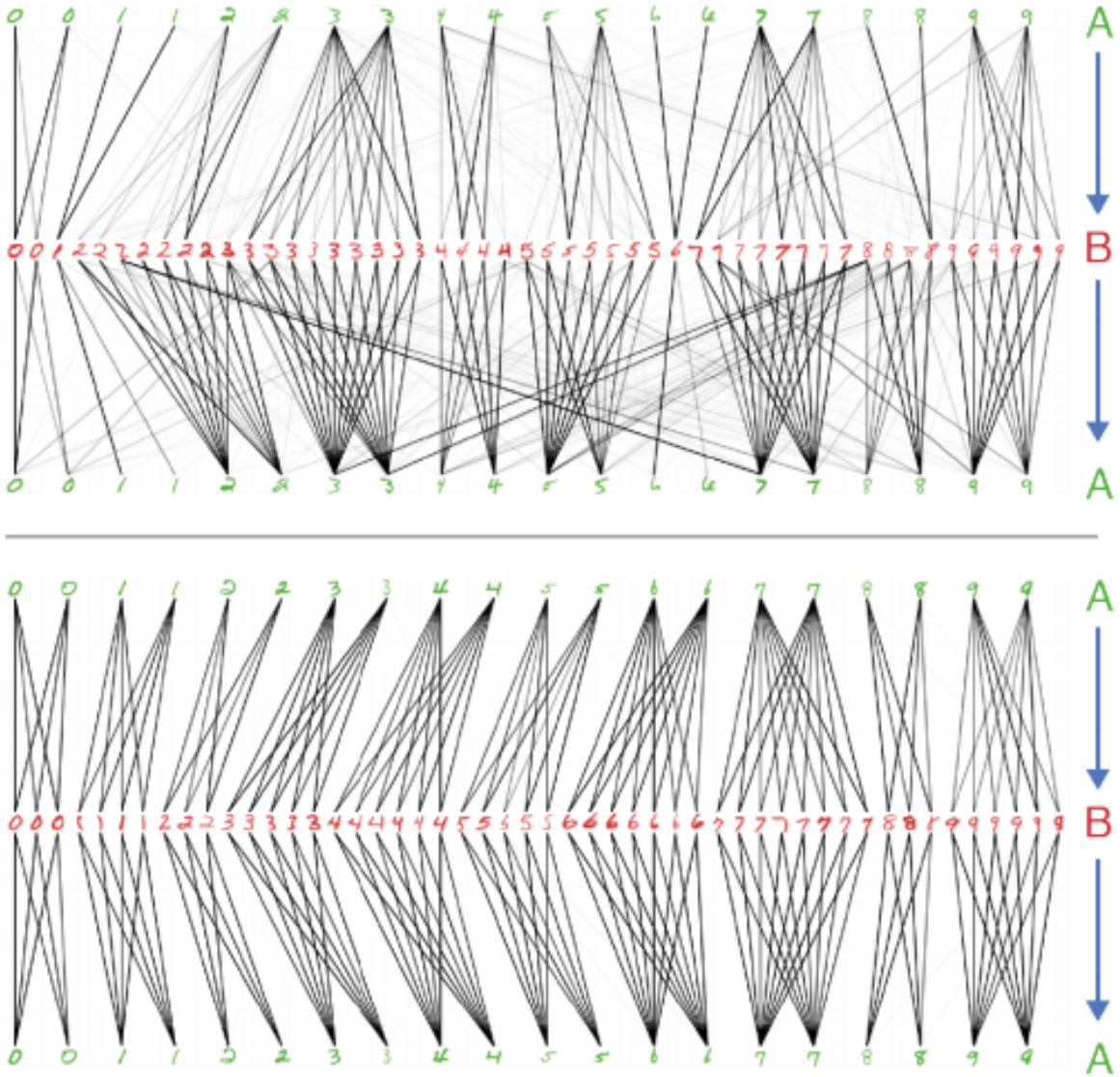

Input
BiLSTM
Hidden representation
Hidden Entity/Verb representation
Bilinear Attention
Representation weights
Weighted Sum (vector R^100)
FC layer
Embedding (vector R^20)
FC layer
State change label (vector R^4)

ProLocal + BERT

Pseudo-label



Learning by Association



Learning by Association

- Given labeled samples A, unlabeled samples B
- Calculate embeddings using NN, E_A, E_B
- Similarity between samples given by $E_{A_i} \cdot E_{B_j} =: M_{ij}$
- Define walker randomly walking between labeled and unlabeled samples by this similarity
- Probability of walking from A_i to $B_j = P_{ij}^{AB} = \frac{exp(M_{ij})}{\sum_k exp(M_{ik})}$, and similarly P^{BA}
- Probability of round-trip walking is $P^{ABA} := P^{AB} P^{BA}$
- Probability of "correct" walks is $\frac{1}{|A|} \sum_{i \sim j} P_{ij}^{ABA}$

Results - Learning by Association

F1 scores on test set

	without unlabeled	with unlabeled
256 labeled	0.353 (0.035)	0.400 (0.044)
512 labeled	0.423 (0.042)	0.436 (0.024)
all labeled	0.522 (0.018)	0.535 (0.015)

Outlook

- Apply SSL on ProLocal + BERT model
- Use other similar datasets as unlabeled samples