Multi-Label Transfer Learning for Multi-Relational Semantic Similarity

Li “Harry” Zhang, Steven R. Wilson, Rada Mihalcea
University of Michigan

*SEM 2019
06/06/2019
Minneapolis, USA

Semantic Similarity Task

- Given two texts, rate the degree of equivalence in meaning
- Dataset: pairs of text & human annotated similarity, e.g. 0 – 5 scale
- Example
  - I will give her a ride to work.
  - I will drive her to the company.
  - Similarity: 5
- Output: A machine predicts similarity scores for all pairs
- Evaluation: Pearson/Spearman's correlation
- Existing datasets: Finkelstein et al. 2012, Agirre et al. 2012-2016, Cer et al. 2017, Hill et al. 2015, Leviant et al. 2015, etc.

Multi-Relational Semantic Similarity Task

- "Similarity" can be defined in different ways, i.e. **relations**
- Some datasets are annotated in multiple relations of similarity
  - **Human Activity**: similarity, relatedness, motivation, actor (Wilson et al. 2017)
  - **SICK**: relatedness, entailment (Marelli et al. 2014)
  - **Typed Similarity**: general, author, people, time, location, event, action, subject, description (Agirre et al. 2013)

Human Activity

- Similarity: do the two activities describe the same thing?
- Relatedness: are the two activities related to one another?
- Motivation: are the two activities done with the same motivation?
- Actor: are the two activities likely to done by the same person?

“Check email” vs. “write email” (scale of 0-4):

Similarity	Relatedness	Motivation	Actor
1.8	3.3	2.6	3.2

SICK

- Sentences Involving Compositional Knowledge
- Relatedness: are the two texts related to one another? (scale 1-5)
- Entailment: does one text entail the other? (three-way)

“Two dogs are wrestling and hugging” vs. “There is no dog wrestling and hugging

Relatedness	Entailment
3.3	Contradict

Typed Similarity

- A collection of meta-data describing books, paintings, films, museum objects and archival records (scale of 0-5)

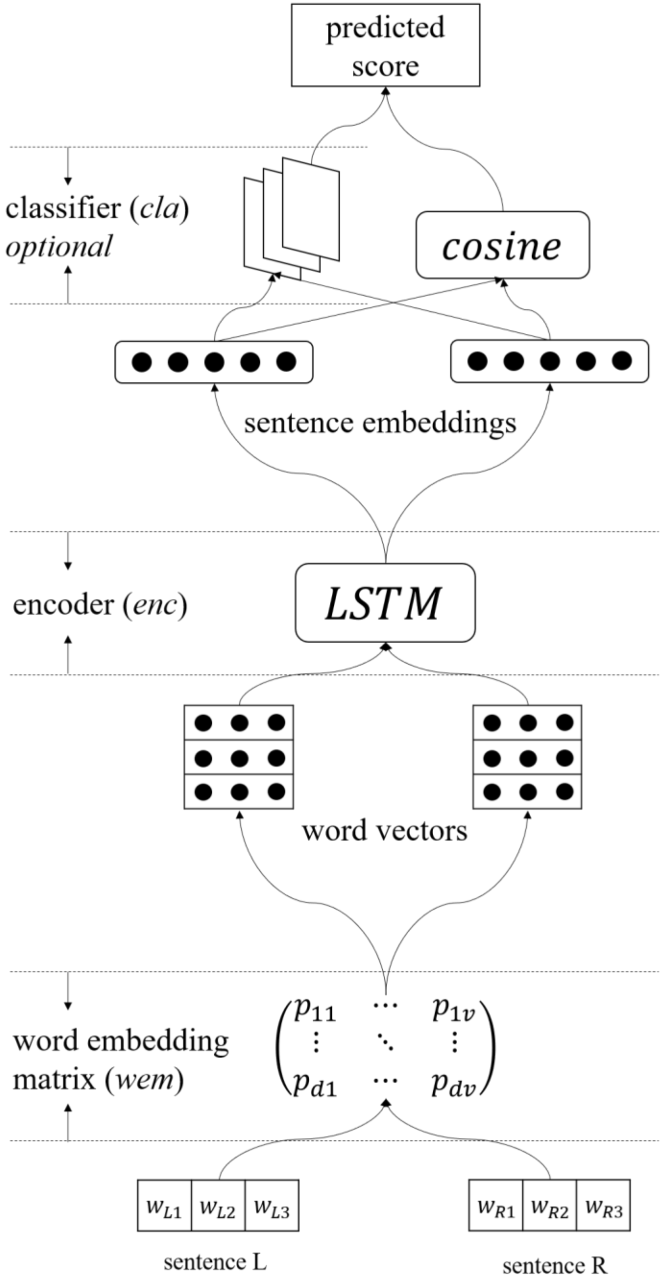
Title: London Bridge, City of London
Creator: not known
Description: A view of London Bridge which is packed with horse-drawn traffic and pedestrians. This bridge replaced the earlier medieval bridge upstream. It was built by John Rennie in 1823-31. A new bridge, built in the late 1960s now stands on this site today.

Title: Serpentine Bridge, Hyde Park, Westminster, Greater London
Creator: de Mare, Eric
Subject: Waterscape Animals Bridge Gardens And Parks
Description: The Serpentine Bridge in Hyde Park seen from the bank. It was built by George and John Rennie, the sons of the geat architect John Rennie, in 1825-8.

general	author	people	time	location	event	subject	description
4.2	2.6	3.0	5.0	4.8	2.8	4.0	3.2

Existing Model: Single Task

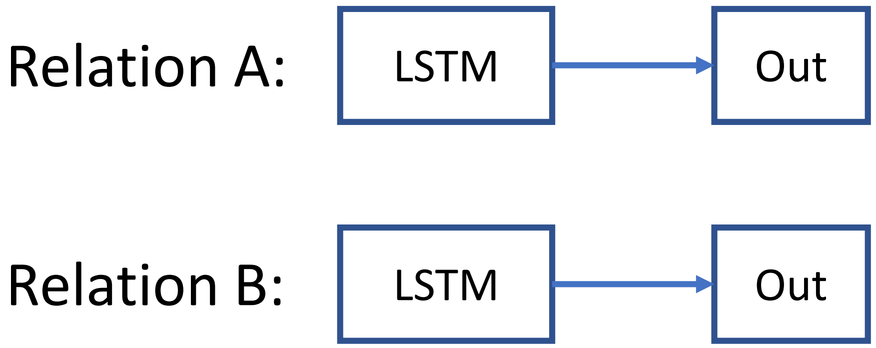
- Fine-tuning with pre-trained sentence encoder / sentence embeddings
- InferSent: Bi-LSTM with max pooling (Conneau et al. 2017)
- A logistic regression layer is used as the output layer
- All parameters are being tuned during transfer learning



Existing Model: Single Task

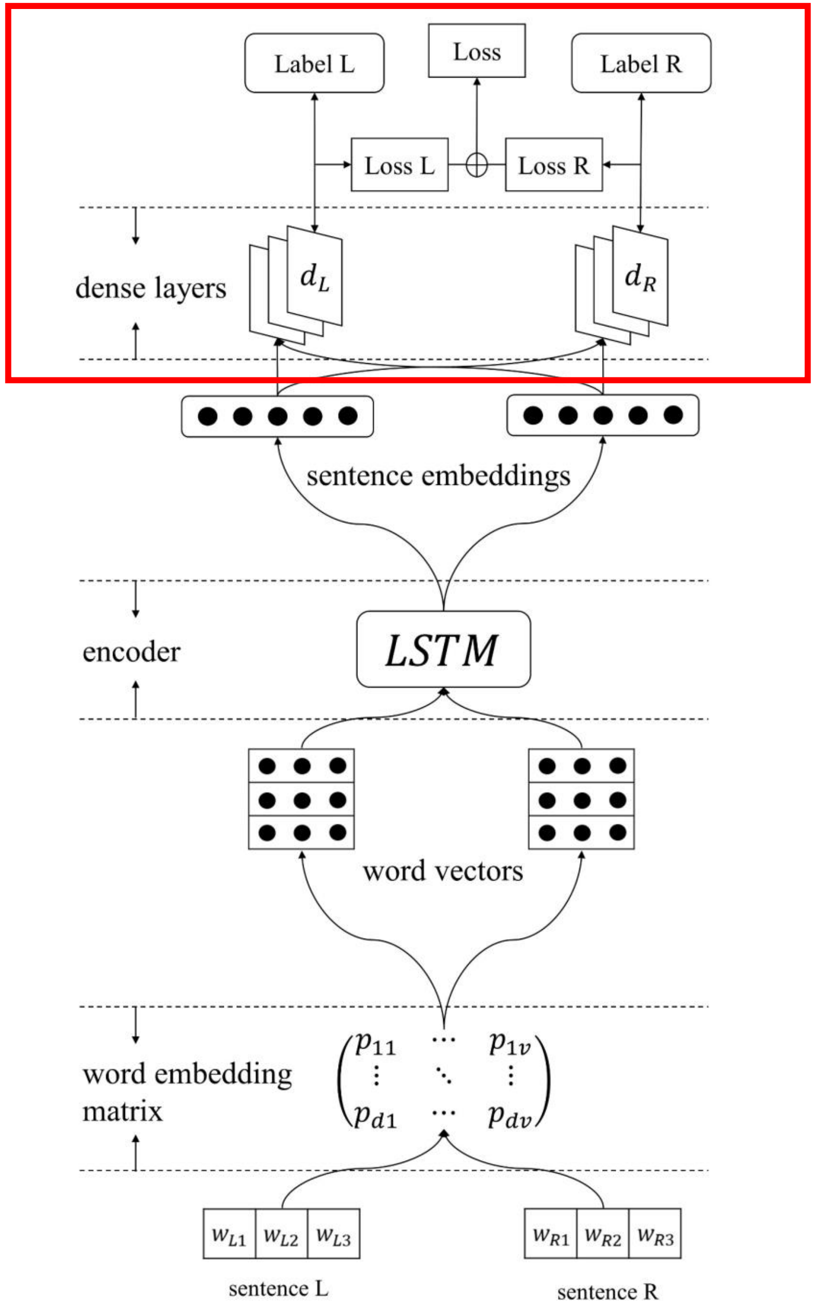
- Treats each relation as a single separate task
- No parameter or information is shared among relations of similarity
- The **Single-Task** baseline

- Question: can we learn across different relations, by sharing parameters?



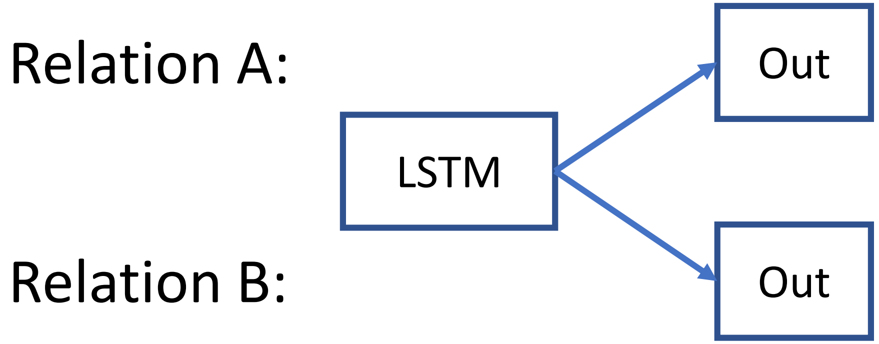
Proposed Multi-Label Model

- Same sentence encoder model
- All relations share the lower-level parameters in the LSTM
- **Each relation has its own output layers**
- Each output layer makes a prediction at the same time



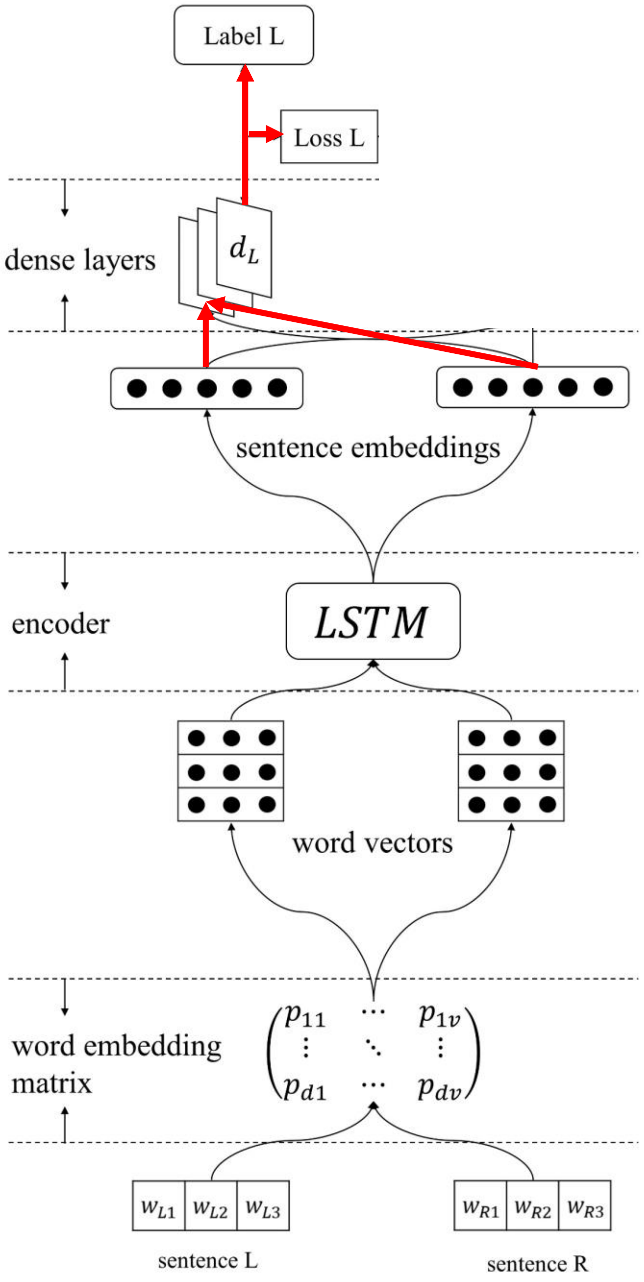
Proposed Multi-Label Model

- Assuming 2 relations (A and B)
- One output layer per relation
- The rest of the parameters are shared between the 2 relations
- The 2 losses are summed as the final loss
- All parameters in the model are updated
- The **Multi-Label** model



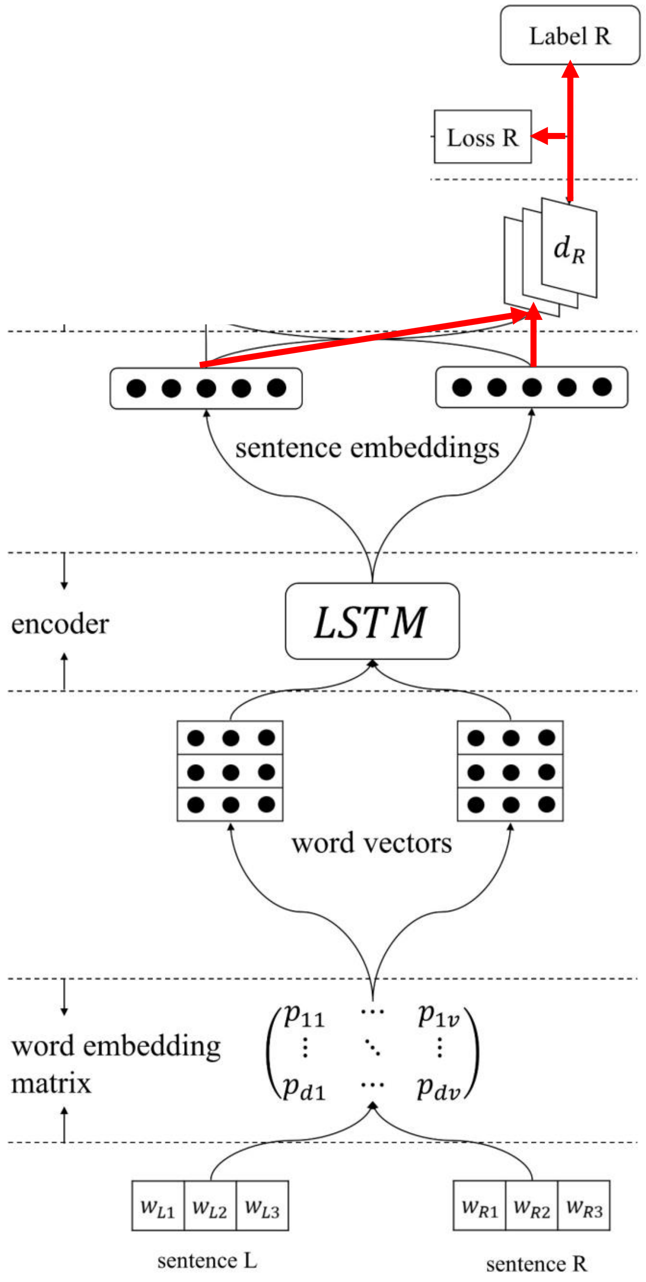
Alternative Multi-Task Model

- Same sentence encoder model
- Alternate between batches of different relations
- Update the related parameters each time



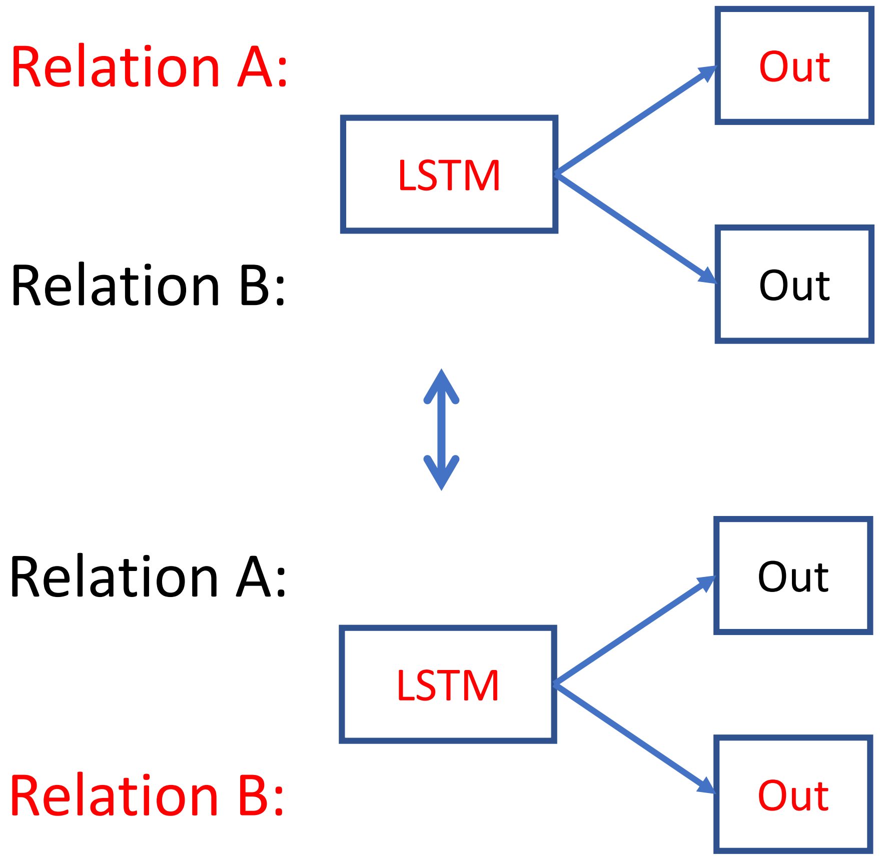
Alternative Multi-Task Model

- Same sentence encoder model
- Alternate between batches of different relations
- Update the related parameters each time



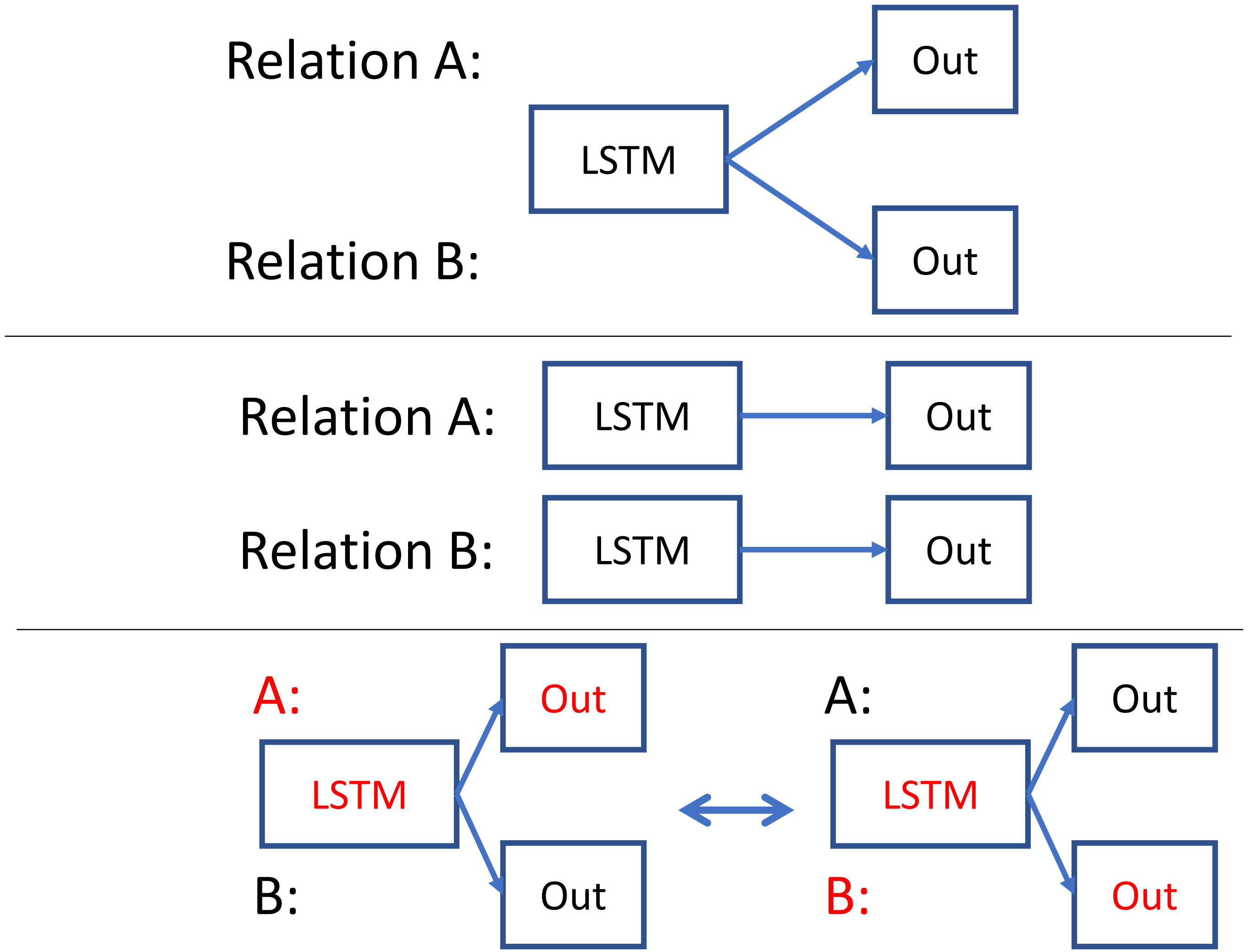
Alternative Multi-Task Model

- Same sentence encoder model
- Assuming 2 relations (A and B)
- Still 2 output layers
- Take a batch of pairs, predict relation A
- Update parameters
- Take a batch of pairs, predict relation B
- Update parameters
- The **Multi-Task** model



Comparison Between the Models

- Multi-Label Learning (MLL)
- Single-Task Learning (Single)
- Multi-Task Learning



Results

- ↑ means MLL outperforms by a statistically significant margin
- ↓ means MLL underperforms by a statistically significant margin
- Multi-Label Learning (MLL) setting has the best performance mostly

	SIM	**REL**	**MA**	**PAC**
MLL	.720	.721	.682	.557
Single	.719	.717↑	.682	.555
MTL	.683↑	.686↑	.651↑	.515↑

Human Activity dataset (Spearman’s correlation)

	Relatedness	**Entailment**
MLL	.882	86.7
Single	.874↑	86.4↑
MTL	.871↑	86.2↑

SICK dataset (Pearson’s correlation)

	general	**author**	**people**	**time**	**location**	**event**	**subject**	**description**
MLL	.744	.721	.640	.713	.751	.611	.697	.737
Single	.750↓	.690↑	.619↑	.712	.744↑	.606↑	.694↑	.718↑
MTL	.718↑	.689↑	.611↑	.697↑	.723↑	.579↑	.669↑	.714↑

Typed-Similarity dataset (Pearson’s correlation)

Discussion and Conclusion

- Multi-Label Learning is a simple but effective way to approach multi-relational semantic similarity tasks
- Learning from one similarity relation helps with learning another
- The idea can be applied to any kind of fine-tuning setting (e.g. graph encoder, language model) used in any multi-label datasets
- Further questions and discussions can be directed to Li Zhang (zharry@umich.edu)