SMU
SINGAPORE MANAGEMENT UNIVERSITY

School of
Computing and
Information Systems

Group 3 – Solution Presentation

Presenter: Jianyu, Fan, Andy

Our Group's Sharing:

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

- Problem Analysis
- Adopting Domain Knowledge
- Experiments and Our Model
- Conclusion and Limitations

Problem Analysis

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

- Claim is a conclusive statement
- Evidence includes the claims about one or few subjects that are capable or not capable of achieving reasonable supports
- Not a simple classification problem based on textual features, strong domain knowledge is needed to train a good model
- Training dataset given may not be enough to ensure model learned these complex domain knowledge
- To source for external data/pre-trained model to enhance domain knowledge in models

Class Labels

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

- SUPPORT:
  - Correlation between Claim & Evidence
  - Evidence semantically support Claim
- CONTRADICT:
  - Correlation between Claim & Evidence
  - Evidence semantically oppose Claim
- NOINFO:
  - Some or no correlation between Claim & Evidence
  - Evidence semantically irrelevant to Claim

Adopting Domain Knowledge

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

- Sci-Fact data
  - train
  - dev
  - few-shot
- Pre-trained models on biomedical domain knowledge

Inspection on Sci-Fact corpus

The Claim sentence cannot be found in the corpus, while Evidence have the format of *doc_id : [sentence, label]*

- For SUPPORT/CONTRADICT samples, evidence is one of the evidence lists
  - generate instances with each evidence
- For the NOINFO samples, we randomly generate sentence pairs

	id	claim	doc_ids	evidence
0	0	0-dimensional biomaterials lack inductive prop...	[31715818]	{}
1	2	1 in 5 million in UK have abnormal PrP positiv...	[13734012]	{'13734012': [{'sentences': [4], 'label': 'CON...

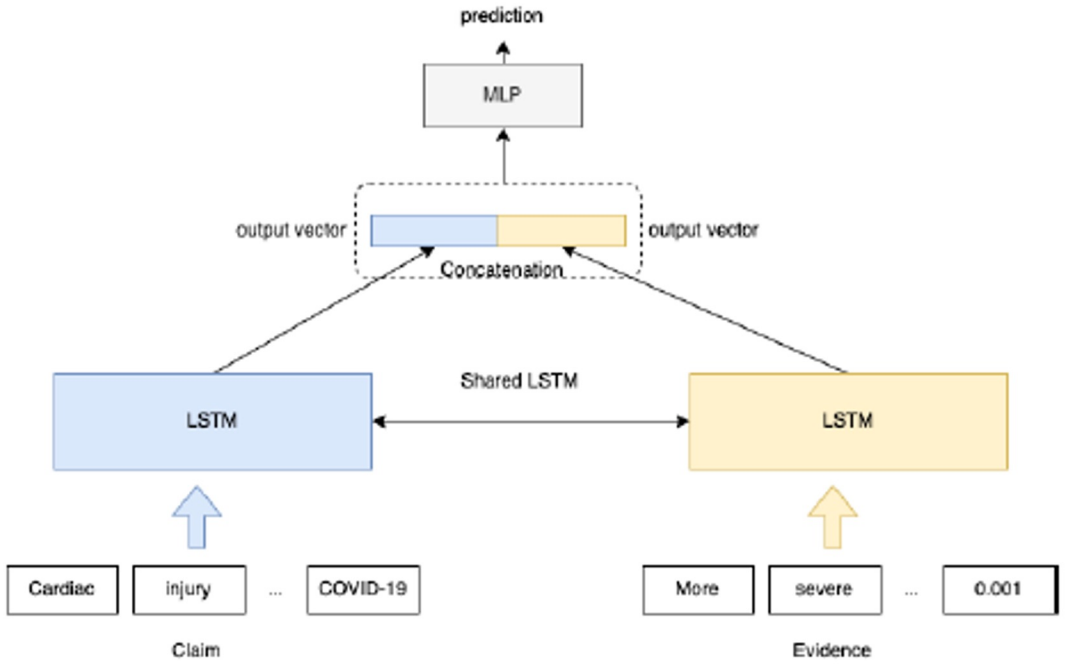
Experiments and Discussions

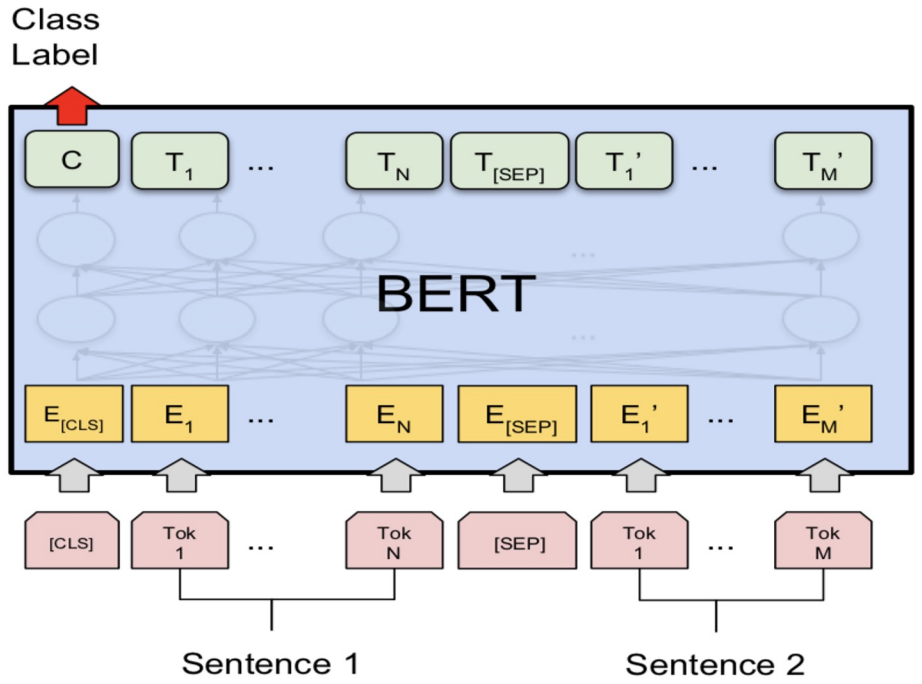
SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

- Baseline LSTM
- Baseline LSTM (with Sci-Fact data)
- Traditional LSTM, Deep Memory Network
- BERT, RoBERTa and other pre-trained Transformer models
- Best model (with Sci-Fact data)

Sentence Pair Training

- Simple understanding is to concatenate Claim sentence and Evidence sentence as the input
- In the baseline LSTM model, the concatenation is inside the model structure, in BERT, the concatenation happens after tokenization





Baseline LSTM

- Model Performance

Table 2: Best Performance of Baseline shared-LSTM under Different Settings

sci-fact proportion	sci-fact raw/processed	score
-	-	0.541846
train	raw	**0.973902**
train+dev	raw	0.940909
train+dev+fewshot	raw	0.893294
train	processed	0.893374

Models Performance Comparison

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Performance for various models

- GPU, Colab Pro / Valset ration 0.10 / epoch 10

Algorithm	**Performance**	**Running time**	**Notes**
LSTM	0.29	24 s	
MemNet	0.49	12 s	
BERT_spc	***0.725644***	3 mins	next step to experiment the different parameter settings
AEN_BERT	0.53	5 mins	

BERT_SPC Model

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Performance for various settings

- GPU, Colab Pro

Parameter settings	Performance	Running time	Notes
val ration 0.10 / epoch 10 (initial setting)	0.725644	3 mins	
val ration 0.15 / epoch 10	**0.744593**	3 mins	
val ration 0.20 / epoch 10	0.67	3 mins	
val ration 0.17 / epoch 10	0.60	3 mins	
val ration 0.10 / epoch 20	0.725644	5 mins	(early stop at epoch 14), 20 would overfitting

Different BERT Models Performance

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Performance for various settings

- GPU, Colab Pro / Valset ration 0.15 / epoch 10

Parameter settings	**Performance**	**Running time**
bert-base-uncased / bert-dim 768 (original)	0.744593	3 mins
bert-large-uncased / bert-dim 1024	***0.748357***	11 mins
bert-large-uncased / bert-dim 1024 (epoch 20)	0.734332	21 mins

RoBERTa Model

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Performance for various settings

- GPU, Colab Pro / Valset ration 0.10 / epoch 5

Parameter settings	**Performance**	**Running time**
roberta-base / epoch 5 / lr 2e-5 (original) / batch size 16	0.831402	2 mins
roberta-base / epoch 20 / lr 2e-5 / batch size 16	**0.835563**	5 mins
roberta-base / epoch 20 / lr 1e-5 / batch size 8	0.809204	6 mins

RoBERTa Model (with Sci-fact)

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Performance for various settings

- GPU, Colab Pro / Valset ration 0.10 / epoch 20

Parameter settings	**Performance**	**Running time**	**Notes**
roberta-base / epoch 20 / lr 2e-5 / batch size 16 (original)	**0.835563**	5 mins	various
roberta-base / **epoch 20** / lr 2e-5 / batch size 16 + sci-fact	0.786538	12 mins	training set of sci-fact -> train & dev
roberta-base / **epoch 5** / lr 2e-5 / batch size 16 + sci-fact	0.786538	3 mins	training set of sci-fact -> train & dev
roberta-base / epoch 5 / lr 2e-5 / batch size 16 + sci-fact	0.790319	3 mins	training set of sci-fact -> only train

RoBERTa Model (with Sci-fact)

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

- Analysis
  - Hyperparameter tuning has **no effect on** the score, two score are the same
  - Assumption: For the pre-trained roberta, adding training data won't affect prediction result of the model; or the form of added data is causing the model to confuse.
- Inspect the added sci-fact data
- Next Step
  - Try to fix the evidence column problems
  - Try other roberta pre-trained models

```
df_claims_train_scifact
```

	id	claim	doc_ids	evidence
0	0	0-dimensional biomaterials lack inductive prop...	[31715818]	{}
1	2	1 in 5 million in UK have abnormal PrP positiv...	[13734012]	{'13734012': [{'sentences': [4], 'label': 'CON...
2	4	1-1% of colorectal cancer patients are diagnos...	[22942787]	{}
3	6	10% of sudden infant death syndrome (SIDS) dea...	[2613775]	{}
4	9	32% of liver transplantation programs required...	[44265107]	{'44265107': [{'sentences': [15], 'label': 'SU...
...	...	...	...	...
804	1403	siRNA knockdown of A20 accelerates tumor progr...	[33370, 38355793]	{'33370': [{'sentences': [6], 'label': 'CONTRA...
805	1404	siRNA knockdown of A20 slows tumor progression...	[33370, 38355793]	{'33370': [{'sentences': [6], 'label': 'SUPPOR...
806	1405	taa1 tar1 tar2 triple mutants in Arabidopsis d...	[10504681]	{}
807	1406	β-sheet opening occurs during pleurotolysin po...	[2617858]	{'2617858': [{'sentences': [4], 'label': 'SUPP...
808	1407	β1/Ketel is able to bind microtubules.	[8087082, 29863668]	{}

809 rows × 4 columns

RoBERTa Model (with Sci-fact)

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Original evidence column in sci-fact corpus:

id	claim	doc_ids	evidence
12	40mg/day dosage of folic acid and 2mg/day dosa...	33409100	{'33409100': [{'sentences': [8], 'label': 'SUP...
28	A T helper 2 cell (Th2) environment impedes di...	12670680	{'12670680': [{'sentences': [1], 'label': 'CON...
28	A T helper 2 cell (Th2) environment impedes di...	12670680	{'12670680': [{'sentences': [1], 'label': 'CON...
30	A breast cancer patient's capacity to metaboli...	24341590	{'24341590': [{'sentences': [10], 'label': 'SU...
30	A breast cancer patient's capacity to metaboli...	24341590	{'24341590': [{'sentences': [10], 'label': 'SU...

RoBERTa Model (with Sci-fact)

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Augmented instances only select the first one evidence as the evidence column

- 0.88 score

```
train_scifact
```

	id	text	evidence	label
0	0	0-dimensional biomaterials lack inductive prop...	Alterations of the architecture of cerebral wh...	0
1	2	1 in 5 million in UK have abnormal PrP positiv...	RESULTS Of the 32,441 appendix samples 16 were...	2
2	4	1-1% of colorectal cancer patients are diagnos...	OBJECTIVES To carry out a further survey of ar...	0
3	6	10% of sudden infant death syndrome (SIDS) dea...	OBJECTIVES To carry out a further survey of ar...	0
4	9	32% of liver transplantation programs required...	Policies requiring discontinuation of methadon...	1
...	...	...	...	...
804	1403	siRNA knockdown of A20 accelerates tumor progr...	Inhibiting A20 expression by siRNAs in vitro r...	2
805	1404	siRNA knockdown of A20 slows tumor progression...	Inhibiting A20 expression by siRNAs in vitro r...	1
806	1405	taa1 tar1 tar2 triple mutants in Arabidopsis d...	OBJECTIVE A20 is a TNF-inducible primary respo...	0
807	1406	β-sheet opening occurs during pleurotolysin po...	The major conformational changes in PlyB are a...	1
808	1407	β1/Ketel is able to bind microtubules.	Membrane attack complex/perforin-like (MACPF) ...	0

809 rows × 4 columns

RoBERTa Model (with Sci-fact)

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Augmented instances select the all evidences from ‘evidence’ column

- 0.98 score (currently the best one)

	id	claim	doc_ids	evidence	Evidence	label
0	12	40mg/day dosage of folic acid and 2mg/day dosa...	33409100	{'33409100': [{'sentences': [8], 'label': 'SUP...	CONCLUSION Treatment with high doses of folic ...	0
1	28	A T helper 2 cell (Th2) environment impedes di...	12670680	{'12670680': [{'sentences': [1], 'label': 'CON...	Individuals with SLE also have elevated serum ...	0
2	28	A T helper 2 cell (Th2) environment impedes di...	12670680	{'12670680': [{'sentences': [1], 'label': 'CON...	Thus, in Lyn(-/-) mice, basophils and IgE auto...	0
3	30	A breast cancer patient's capacity to metaboli...	24341590	{'24341590': [{'sentences': [10], 'label': 'SU...	Compared with extensive metabolizers, there wa...	0
4	30	A breast cancer patient's capacity to metaboli...	24341590	{'24341590': [{'sentences': [10], 'label': 'SU...	Compared with extensive metabolizers, those wi...	0
...	...	...	...	...	...	...
1256	1403	siRNA knockdown of A20 accelerates tumor progr...	[33370, 38355793]	{'33370': [{'sentences': [6], 'label': 'CONTRA...	The tumorigenic potential of GSCs was decrease...	0
1257	1404	siRNA knockdown of A20 slows tumor progression...	[33370, 38355793]	{'33370': [{'sentences': [6], 'label': 'SUPPOR...	The tumorigenic potential of GSCs was decrease...	0
1258	1405	taa1 tar1 tar2 triple mutants in Arabidopsis d...	[10504681]	{}	Alterations of the architecture of cerebral wh...	0
1259	1406	β-sheet opening occurs during pleurotolysin po...	[2617858]	{'2617858': [{'sentences': [4], 'label': 'SUPP...	The major conformational changes in PlyB are a...	0
1260	1407	β1/Ketel is able to bind microtubules.	[8087082, 29863668]	{}	Alterations of the architecture of cerebral wh...	0

1261 rows × 6 columns

RoBERTa Model (with Sci-fact)

- Performance for various settings
  - GPU, Colab Pro / Valset ration 0.10 / epoch 20
- Sci-Fact (train) dataset, evidence fixed

Parameter settings	**Performance**	**Running time**	**Notes**
roberta-base / epoch 20 / lr 2e-5 / batch size 16 (original)	0.835563	5 mins	
roberta-base / epoch 5 / lr 2e-5 / batch size 16 + sci-fact	0.790319	3 mins	training set of sci-fact -> only train
roberta-base / epoch 5 / lr 2e-5 / batch size 16 + sci-fact (evidence fixed -simplest way)	0.881113	3 mins	training set of sci-fact -> only train, also **simply fixed the evidence problem**
roberta-base / epoch 5 / lr 2e-5 / batch size 16 + sci-fact (evidence **fully** fixed)	**0.980364**	5 mins	training set of sci-fact -> only train, also **fully fixed the evidence problem**

Further Experiment with Baseline LSTM

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

With Sci-fact corpus (train), loading the evidence column fully fixed version

- GPU: Tesla T4 15GB

Parameter settings	**Performance**	**Running time**	**Notes**
baseline LSTM + sci-fact (train)	**0.973902**	30 s	training set of sci-fact, the evidence column is either a dic or a '{}'
baseline LSTM + sci-fact (train)	0.893374	30 s	training set of sci-fact, the evidence column is meaningful and single sentence

LSTM (with Sci-Fact) Implication

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

- The previous LSTM (with Sci-Fact train) achieved 0.97 is probably caused by the leak of label information in the evidence column
- After resolving the leak of label information issue, LSTM (with Sci-Fact train) still outperforms the simple RoBERTa-SPC model, which means that adding the **better** training instances could greatly improve the performance even with the much simpler model.

Other RoBERTa Models

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Apply the roberta-large pre-trained model (previously was only roberta-base)

Parameter settings	**Performance**	**Running time**	**Notes**
roberta-base / epoch 5 / lr 2e-5 / batch size 16 + sci-fact (evidence **fully** fixed)	0.980364	5 mins	the validation is only separated from provided training data
roberta-large / epoch 5 / lr 2e-5 / batch size 8 + sci-fact (evidence **fully** fixed)	**0.987036**	12 mins	the validation is only separated from provided training data
roberta-large / epoch 10 / lr 2e-5 / batch size 8 + sci-fact (evidence **fully** fixed)	0.98684	40 mins	the validation is only separated from provided training data, **overfitting**

Other Pre-trained Models

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Apply the various pre-trained model (biomedical related) [part 1]

Parameter settings	Performance	Running time	Notes
roberta-large	**0.987036**	12 mins	
allenai/biomed_roberta_base	0.980364	12 mins	2.68 millions scientific papers from Semantic Scholar corpus via continued pretraining
PlanTL-GOB-ES/roberta-base-biomedical-clinical-es	0.947608	5 mins	in Spanish, 278k clinical documents and notes, with data cleaned before training
raynardj/pmc-med-bio-mlm-roberta-large	**0.987036**	17 mins	Pre-trained on full-text scholarly articles (5.2 millions) in biomedical and life sciences journals [1]

Limitations

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Computing and Information Systems

Data Augmentation Perspective

- Extra training instances and training epoch does improved the prediction result significantly, however too much may lead to overfitting (lower result)
- While using Sci-Fact sentence pairs with NOINFO label are generated randomly from Sci-Fact corpus, which may lead to possible bias
- The proportion of the NOINFO / SUPPORT / CONTRADICT instances may also shade influences on the training process, further experiments can be made to address this
- Our model performances seem to heavily depend on the extra training instances. Model may be overfitting and may not perform as well in phase 2.

Limitations

Architecture Perspective

- Selecting the correct model is more important than the hyperparameter tuning. It may remain a question about how to best select the pre-trained models
- New approaches of make better use of pre-trained models are emerging, like prompt learning. There can be more explorations on those new techniques.

Thanks!