Yuanfudao at SemEval-2018 Task 11: Three-way Attention and Relational Knowledge for Commonsense Machine Comprehension

Liang Wang, Meng Sun, Wei Zhao, Kewei Shen, Jingming Liu

Yuanfudao Research

Brief overview of task 11

My backyard was looking a little empty, so I decided I would plant something. I went out and bought tree seeds. I found a spot in my yard that looked like it would get enough sunshine. There, I dug a hole for the seeds. Once that was done, I took my watering can and watered the seeds.

A. Why was the tree planted in that spot?

1. to get enough sunshine
2. there was no other space

B. What was used to dig the hole?

1. a shovel
2. their bare hands

Passage

Question

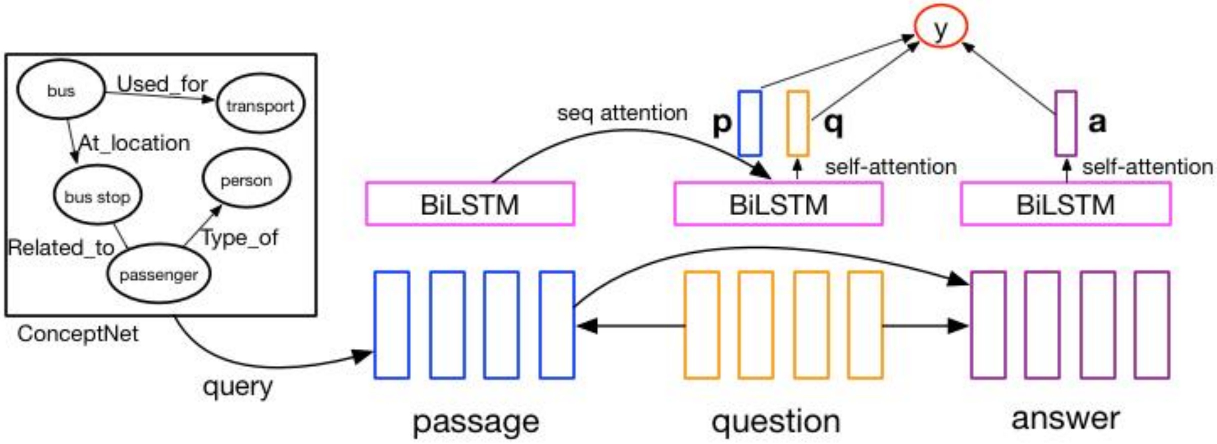
A multiple-choice reading comprehension task

Two candidate answers

Two key questions

- How to model interactions between different parts?
- How to incorporate commonsense knowledge?

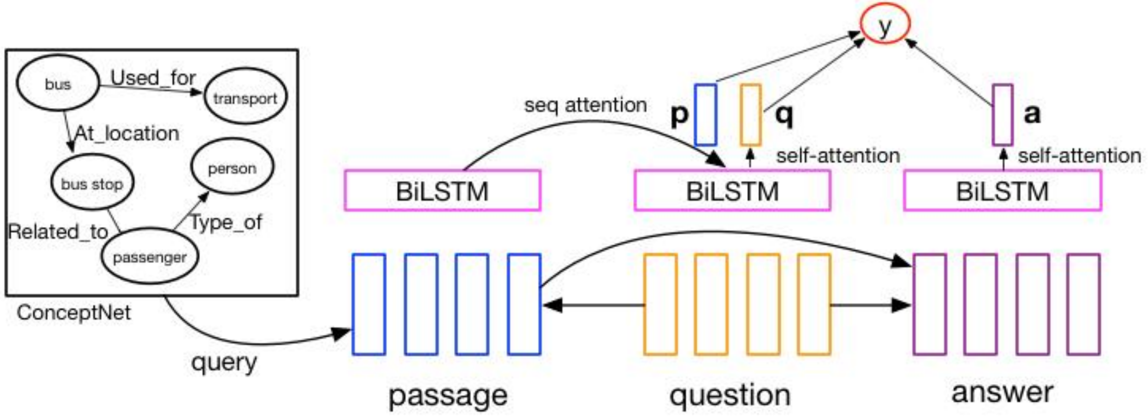
Model Architecture



Our proposed Three-way Attentive Networks (TriAN)

Input Layer

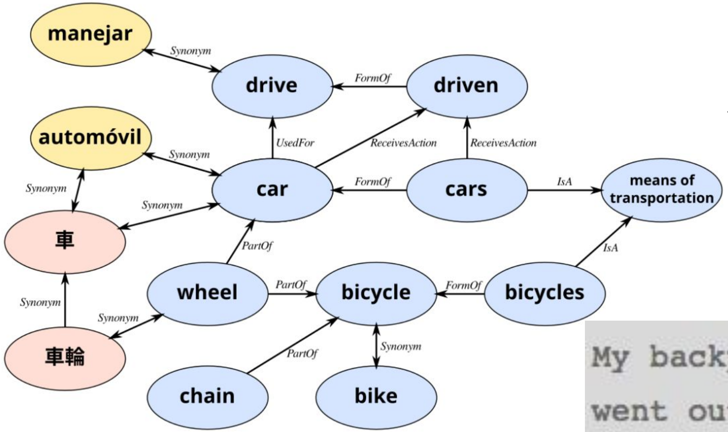
➢ Combine multiple sources of information
- GloVe embeddings
- Part-of-Speech embedding
- Named-entity embedding
- ConceptNet relation embedding
- Word-match features
- Term frequency features



$$\mathbf{w}_{P_i} = [\mathbf{E}^{glove}_{P_i}; \mathbf{E}^{pos}_{P_i}; \mathbf{E}^{ner}_{P_i}; \mathbf{E}^{rel}_{P_i}; \mathbf{f}_{P_i}]$$

Our proposed Three-way Attentive Networks (TriAN)

ConceptNet as a commonsense knowledge base



Response: Is_capable_of

My backyard was looking a little empty, so I decided I would plan
went out and bought tree seeds. I found a spot in my yard that lo
would get enough sunshine. There, I dug a hole for the seeds. Onc
I took my watering can and watered the seeds.

Query: <dig a hole, shovel>

A. Why was the tree planted in that spot?
1. to get enough sunshine
2. there was no other space

B. What was used to dig the hole?
1. a shovel
2. their bare hands

Attention Layer

- Sequence attention

$$Att_{seq}(\mathbf{u}, \{\mathbf{v}_i\}_{i=1}^{n}) = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$$

$$\alpha_i = \text{softmax}_i(f(\mathbf{W}_1 \mathbf{u})^T f(\mathbf{W}_1 \mathbf{v}_i))$$

$$\mathbf{h}^q = \text{BiLSTM}(\{\mathbf{w}_{Q_i}\}_{i=1}^{|Q|})$$

$$\mathbf{h}^p = \text{BiLSTM}(\{[\mathbf{w}_{P_i}; \mathbf{w}_{P_i}^q]\}_{i=1}^{|P|})$$

$$\mathbf{h}^a = \text{BiLSTM}(\{[\mathbf{w}_{A_i}; \mathbf{w}_{A_i}^p; \mathbf{w}_{A_i}^q]\}_{i=1}^{|A|})$$

Question-aware passage representation $\mathbf{w}_{P_i}^q = Att_{seq}(\mathbf{E}_{P_i}^{glove}, \{\mathbf{E}_{Q_i}^{glove}\}_{i=1}^{|Q|})$

Passage-aware answer representation $\mathbf{w}_{A_i}^p = Att_{seq}(\mathbf{E}_{A_i}^{glove}, \{\mathbf{E}_{P_i}^{glove}\}_{i=1}^{|P|})$

Question-aware answer representation $\mathbf{w}_{A_i}^q = Att_{seq}(\mathbf{E}_{A_i}^{glove}, \{\mathbf{E}_{Q_i}^{glove}\}_{i=1}^{|Q|})$.

Output Layer

- Self attention for summarizing vectors

$$Att_{self}(\{\mathbf{u}_i\}_{i=1}^{n}) = \sum_{i=1}^{n} \alpha_i \mathbf{u}_i$$

$$\alpha_i = \mathrm{softmax}_i(\mathbf{W}_2^T \mathbf{u}_i)$$

$$y = \sigma(\mathbf{p}^T \mathbf{W}_3 \mathbf{a} + \mathbf{q}^T \mathbf{W}_4 \mathbf{a})$$

Final prediction

Question representation: $\mathbf{q} = Att_{self}(\{\mathbf{h}_i^q\}_{i=1}^{|Q|})$

Answer representation: $\mathbf{a} = Att_{self}(\{\mathbf{h}_i^a\}_{i=1}^{|A|})$

Passage representation: $\mathbf{p} = Att_{seq}(\mathbf{q}, \{\mathbf{h}_i^p\}_{i=1}^{|P|})$.

Pretraining as a way for knowledge transfer

- RACE dataset
  - Multiple-choice reading comprehension dataset with ~100k questions.

- Pretrain on RACE and use its parameters for initialization

RACE dataset is available at http://www.cs.cmu.edu/~glai1/data/race/

Experimental Results

model	dev	test
Random	50.00%	50.00%
TriAN-RACE	64.78%	64.28%
TriAN-single	83.84%	81.94%
TriAN-ensemble	**85.27%**	**83.95%**
Human	–	98.00%

Table 2: Main results. *TriAN-RACE* only use *RACE* dataset for training The evaluation metric is accuracy.

Much better than the baseline, far from the human performance

Experimental Results

Rank	Team name	Main model	Commonsense	Other resources	Acc.
1	Yuanfudao	LSTM	ConceptNet	GloVe, Wikipedia, POS and NE tagging	0.84
2	MITRE	LSTM	–	word2vec, Twitter, stemming	0.82
3	Jiangnan	LSTM	–	GloVe, CoVe, POS and NE tagging	0.81
4	ELiRF-UPV	LSTM	ConceptNet	–	0.75
5	YNU_Deep	LSTM	–	GloVe	0.75
6	ZMU	LSTM	–	word2vec, GloVe	0.74
7	ECNU	LSTM	–	GloVe	0.73
8	YNU_AI1799	LSTM/CNN	–	word2vec, GloVe	0.72
9	YNU-HPCC	LSTM/CNN	–	word2vec	0.71
10	CSReader	LSTM	–	lemmatization, GloVe	0.63
11	IUCM	k-means	DeScript, MCScript[2]	NLTK	0.61

Experimental Results

Rank	Team name	Total	Commonsense	Text	Out of Domain
1	Yuanfudao	**0.84***	**0.82**	**0.85**	**0.79**
2	MITRE	0.82	0.79	0.83*	0.78
3	Jiangnan	0.81*	0.80	0.81*	0.75*
4	ELiRF-UVP	0.75	**0.82**	0.73	0.70
5	YNU_Deep	0.75	0.79	0.73	0.66
6	ZMU	0.74	0.80	0.72	0.66
7	ECNU	0.73	0.77	0.72	0.69
8	YNU_AI1799	0.72	0.76	0.71	0.67
9	YNU_HPCC	0.71*	0.78*	0.69*	0.64*
10	CSReader	0.63	0.64*	0.63	0.59
11	IUCM	0.61	0.54	0.64	0.58
–	Attentive Reader	0.72	0.75	0.71	0.69
–	Sliding Window	0.55	0.53	0.56	0.52
–	Human Performance	0.98			

Going deeper with attention

➢ Treat the output of BiLSTMs as new input representations

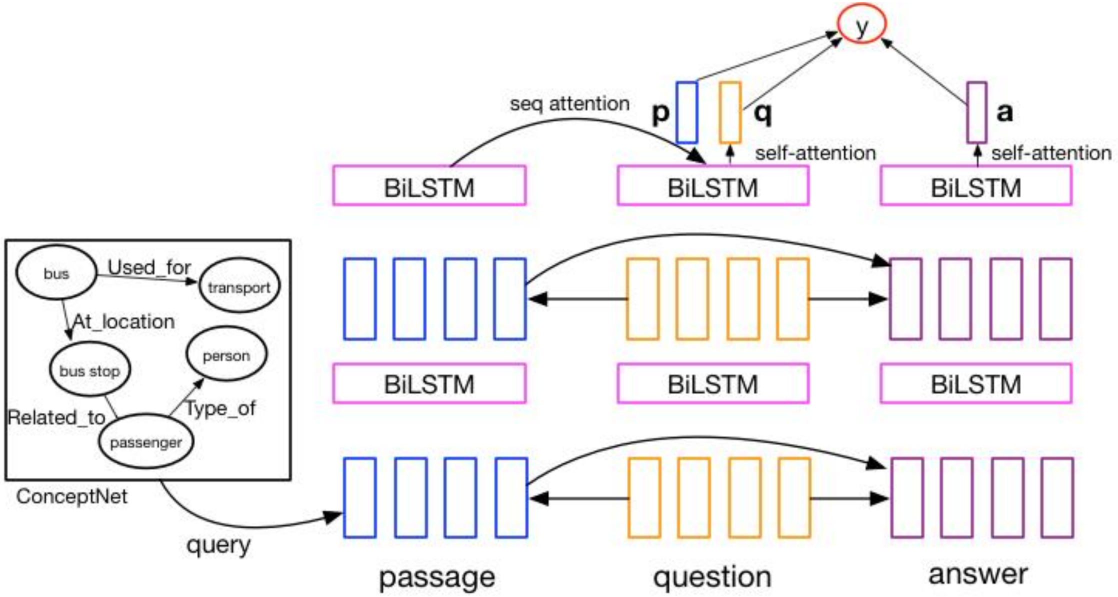


Illustration of two-layer TriAN architecture

Experimental Results

model	dev	test
1-layer TriAN-single	**83.84%**	**81.94%**
2-layer TriAN-single	82.71%	80.55%

Table 3: Accuracy comparison of shallow and deep TriAN models.

Ablation Study

model	dev	test
TriAN-single	**83.84%**	**81.94%**
w/o pretraining	82.71%	80.51%
w/o ConceptNet	82.78%	81.08%
w/o POS	82.84%	81.27%
w/o features	82.92%	81.35%
w/o NER	83.60%	81.87%

Table 4: Ablation study for input representation.

Pretraining matters

Commonsense knowledge matters

Ablation Study

model	dev	test
TriAN-single	**83.84%**	81.94%
w/o passage-question attention	83.51%	**82.20%**
w/o passage-answer attention	83.07%	81.39%
w/o question-answer attention	83.23%	81.84%
w/o attention	81.93%	80.65%

Attention matters

Table 5: Ablation study for attention. The last one "*w/o attention*" removes all word-level attentions.

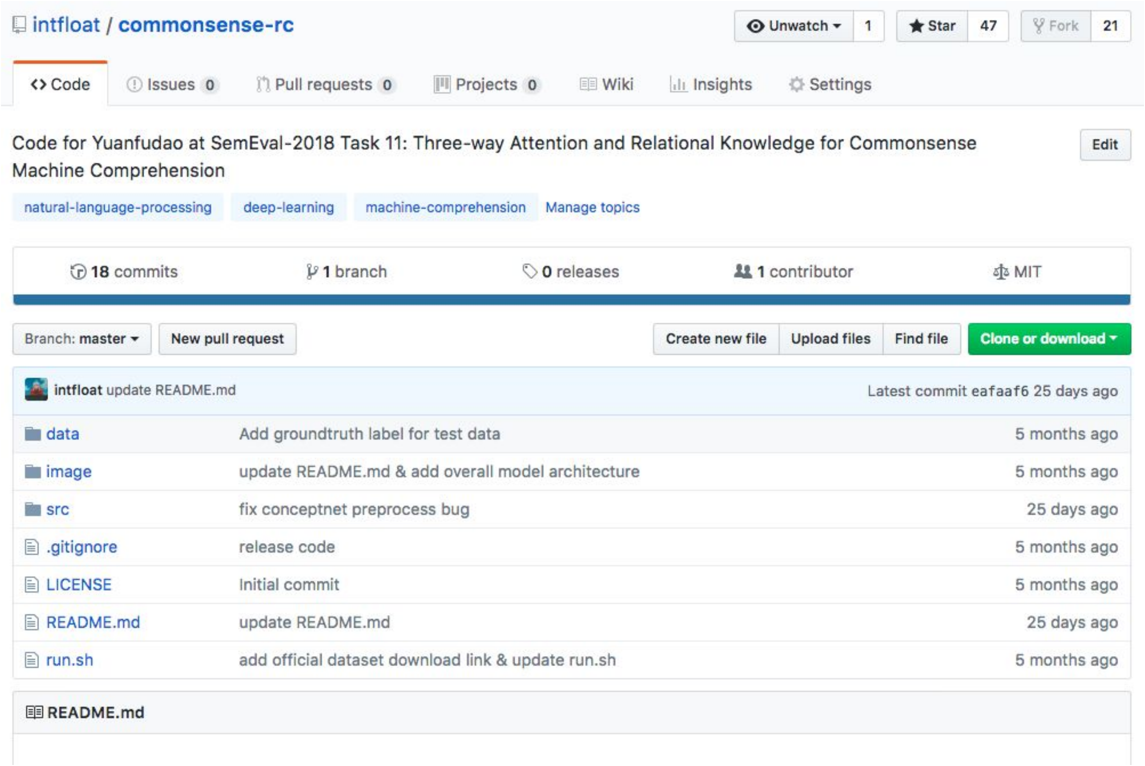
Further Discussions

- What is the appropriate way to model commonsense knowledge?
  - Commonsense knowledge as additional input
  - Event calculus
- Transfer learning among related but different datasets
  - Pretraining
  - Data augmentation + weighted sampling
  - Domain adaptation

Conclusion

- An intuitive attention-based method for multiple-choice reading comprehension
- ConceptNet as commonsense knowledge base
- Pretraining as a simple and effective way for knowledge transfer

Code Available



https://github.com/intfloat/commonsense-rc

Recent Advances on Machine Reading Comprehension

- ➢ NAACL 2018
- ➢ IJCAI 2018
- ➢ ACL 2018

NAACL 2018

- *Looking Beyond the Surface: A Challenge Set for Reading Comprehension over Multiple Sentences*
- *CliCR: a Dataset of Clinical Case Reports for Machine Reading Comprehension*
- *Tracking State Changes in Procedural Text: a Challenge Dataset and Models for Process Paragraph Comprehension*
- *The Argument Reasoning Comprehension Task: Identification and Reconstruction of Implicit Warrants*
- *Challenging Reading Comprehension on Daily Conversation: Passage Completion on Multiparty Dialog*
- *Contextualized Word Representations for Reading Comprehension*
- *Robust Machine Comprehension Models via Adversarial Training*
- *Simple and Effective Semi-Supervised Question Answering*
- *Deep Contextualized Word Representations*

NAACL 2018

- ➢ *Looking Beyond the Surface: A Challenge Set for Reading Comprehension over Multiple Sentences*
- ➢ *CliCR: a Dataset of Clinical Case Reports for Machine Reading Comprehension*
- ➢ *Tracking State Changes in Procedural Text: a Challenge Dataset and Models for Process Paragraph Comprehension*
- ➢ *The Argument Reasoning Comprehension Task: Identification and Reconstruction of Implicit Warrants*
- ➢ *Challenging Reading Comprehension on Daily Conversation: Passage Completion on Multiparty Dialog*
- ~~➢ *Contextualized Word Representations for Reading Comprehension*~~
- ➢ *Robust Machine Comprehension Models via Adversarial Training (make model more robust to adversarial examples)*
- ~~➢ *Simple and Effective Semi-Supervised Question Answering*~~
- ~~➢ *Deep Contextualized Word Representations*~~

Harder Reading Comprehension Datasets

- *Looking Beyond the Surface: A Challenge Set for Reading Comprehension over Multiple Sentences (multiple choice, 6k questions, 7 domains)*
- *CliCR: a Dataset of Clinical Case Reports for Machine Reading Comprehension (cloze-style, 100k questions, clinical case reports)*
- *Tracking State Changes in Procedural Text: a Challenge Dataset and Models for Process Paragraph Comprehension (procedural text, 81k datapoints)*
- *The Argument Reasoning Comprehension Task: Identification and Reconstruction of Implicit Warrants (2k arguments, used for SemEval-2018 Task 12)*
- *Challenging Reading Comprehension on Daily Conversation: Passage Completion on Multiparty Dialog (cloze-style, 13k questions)*

IJCAI 2018

- *Reinforced Mnemonic Reader for Machine Reading Comprehension*
- *Towards Reading Comprehension for Long Documents*

IJCAI 2018

- ~~*Reinforced Mnemonic Reader for Machine Reading Comprehension*~~
- *Towards Reading Comprehension for Long Documents (two-stage technique for long document RC)*

ACL 2018

- *Knowledgeable Reader: Enhancing Cloze-Style Reading Comprehension with External Commonsense Knowledge*
- *Simple and Effective Multi-Paragraph Reading Comprehension*
- *DuoRC: Towards Complex Language Understanding with Paraphrased Reading Comprehension*
- *Stochastic Answer Networks for Machine Reading Comprehension*
- *Multi-Granularity Hierarchical Attention Fusion Networks for Reading Comprehension and Question Answering*
- *Joint Training of Candidate Extraction and Answer Selection for Reading Comprehension*
- *Multi-Passage Machine Reading Comprehension with Cross-Passage Answer Verification*
- *Efficient and Robust Question Answering from Minimal Context over Documents*
- *Harvesting Paragraph-level Question-Answer Pairs from Wikipedia*
- *Multi-Relational Question Answering from Narratives: Machine Reading and Reasoning in Simulated Worlds*
- *Know What You Don’t Know: Unanswerable Questions for SQuAD*

ACL 2018

- *Knowledgeable Reader: Enhancing Cloze-Style Reading Comprehension with External Commonsense Knowledge*
- *Simple and Effective Multi-Paragraph Reading Comprehension*
- *DuoRC: Towards Complex Language Understanding with Paraphrased Reading Comprehension*
- ~~*Stochastic Answer Networks for Machine Reading Comprehension*~~
- *Multi-Granularity Hierarchical Attention Fusion Networks for Reading Comprehension and Question Answering*
- *Joint Training of Candidate Extraction and Answer Selection for Reading Comprehension*
- *Multi-Passage Machine Reading Comprehension with Cross-Passage Answer Verification*
- *Efficient and Robust Question Answering from Minimal Context over Documents*
- *Harvesting Paragraph-level Question-Answer Pairs from Wikipedia*
- *Multi-Relational Question Answering from Narratives: Machine Reading and Reasoning in Simulated Worlds*
- ~~*Know What You Don’t Know: Unanswerable Questions for SQuAD*~~

Harder Reading Comprehension Datasets

- *DuoRC: Towards Complex Language Understanding with Paraphrased Reading Comprehension (movie plots, 186k questions)*
- *Harvesting Paragraph-level Question-Answer Pairs from Wikipedia (1 million auto-generated questions)*
- *Multi-Relational Question Answering from Narratives: Machine Reading and Reasoning in Simulated Worlds (> 1 million questions in simulated world)*

ACL 2018

- *Knowledgeable Reader: Enhancing Cloze-Style Reading Comprehension with External Commonsense Knowledge*
- *Simple and Effective Multi-Paragraph Reading Comprehension (multi-document RC)*
- *Multi-Granularity Hierarchical Attention Fusion Networks for Reading Comprehension and Question Answering (single-document RC)*
- *Joint Training of Candidate Extraction and Answer Selection for Reading Comprehension (multi-document RC)*
- *Multi-Passage Machine Reading Comprehension with Cross-Passage Answer Verification (multi-document RC)*
- *Efficient and Robust Question Answering from Minimal Context over Documents*
- *Multi-Relational Question Answering from Narratives: Machine Reading and Reasoning in Simulated Worlds*

ACL 2018

- *Knowledgeable Reader: Enhancing Cloze-Style Reading Comprehension with External Commonsense Knowledge (combine with external knowledge)*
- *Simple and Effective Multi-Paragraph Reading Comprehension*
- *Multi-Granularity Hierarchical Attention Fusion Networks for Reading Comprehension and Question Answering*
- *Joint Training of Candidate Extraction and Answer Selection for Reading Comprehension*
- *Multi-Passage Machine Reading Comprehension with Cross-Passage Answer Verification*
- *Efficient and Robust Question Answering from Minimal Context over Documents (faster RC by selecting relevant sentences first)*