Commonsense Reasoning with Implicit Knowledge in Natural Language

*Pratyay Banerjee, *Swaroop Mishra, *Kuntal Pal,

*Arindam Mitra, Chitta Baral

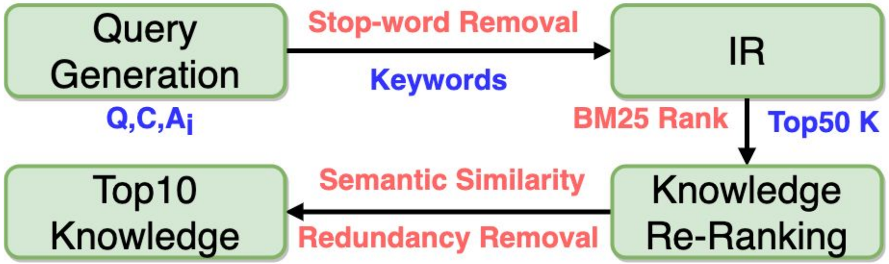
Arizona State University

Motivation

Abductive NLI	Social IQA	Physical IQA
Obs1: Jim was working on a project. ✔ Jim found he was missing an item. ✘ Jim needed a certain animal for it. **Obs2**: Luckily, he found it on a nearby shelf **Knowledge**: Peyton eventually found it before Peyton needed to determine that something is missing. Kendall never found it, as a result Kendall wants to lodge a missing complaint.	**Context**: Remy was an expert fisherman and was on the water with Kai. Remy baited Kai's hook. **Question**: What will Remy want to do next? ✔ cast the line ✘ put the boat in the water ✘ invite Kai out on the boat **Knowledge**: Alex baits Pat's hook as a result others want to cast their line.	**Goal**: When doing sit-ups: ✔ place your tongue in the roof of your mouth. It will stop you from straining your neck. ✘ place your elbow in the roof of of your mouth. It will stop you from straining your neck. **Knowledge**: How to Do Superbrain Yoga. Place your tongue on the roof of your mouth.

- Can we answer commonsense questions regarding Abductive, Social Interactions and Physical Interactions?
- Can we improve our answers with implicit knowledge in natural language form ?
- How can we find relevant knowledge ?
- How to reason with retrieved implicit relevant knowledge ?
- What kind of training strategies can we use ?
- How do they compare with respect to large / complex models ?

Knowledge Retrieval Pipeline



- **Knowledge Source**

aNLI → ***ROC Stories* + *Story Cloze*** (Directly Derived)
SIQA → ***Atomic*** (Partially Derived)
PIQA → ***Wikihow*** (Relevant)

Knowledge Infusion Modes

- The top 10 knowledge statements(K_{ij}) retrieved for each options (A_i) are infused with the transformer encoders in four different ways

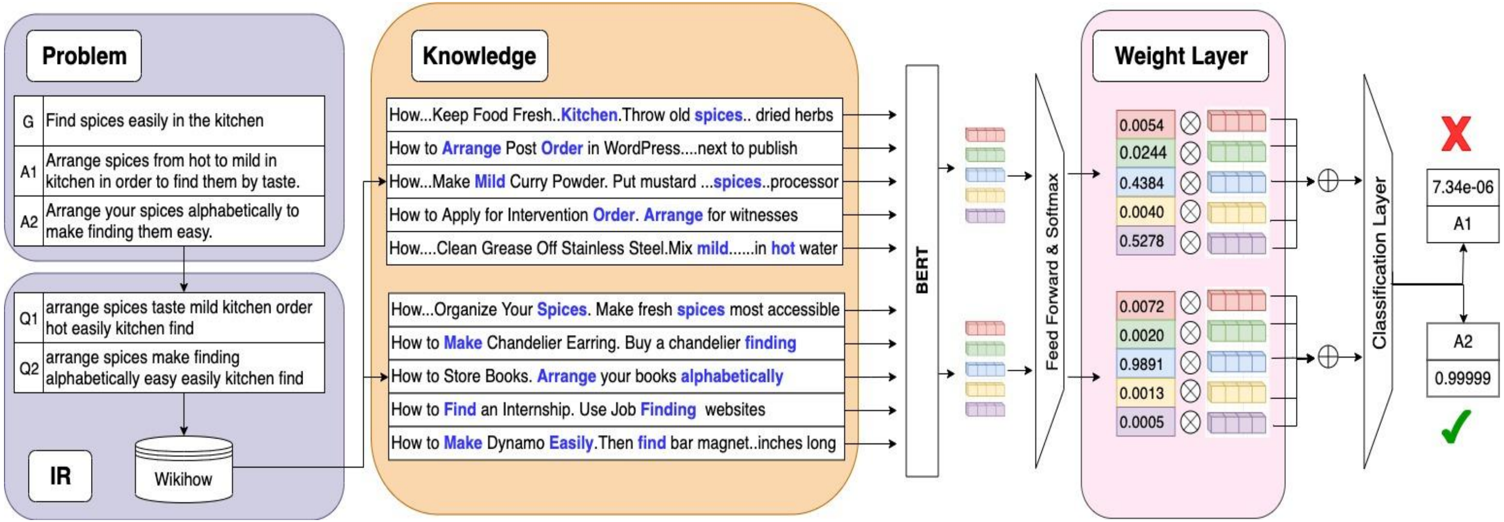
CONCAT [CLS] K_i [SEP] Q A_i [SEP]
PARALLEL-MAX [CLS] K_{ij} [SEP] Q A_i [SEP] → max (CLS for each Ai)
SIMPLE-SUM [CLS] K_{ij} [SEP] Q A_i [SEP] → sum (CLS for each Ai)
WEIGHTED-SUM [CLS] K_{ij} [SEP] Q A_i [SEP] → wtd-sum (CLS for each Ai)

Training Strategies

- **REVISION**
  - Pre-Trained (MLM+NSP [BERT] and MLM [RoBERTa]) on K_D
  - Fine-Tuned on D
- **OPENBOOK**
  - Fine-Tuned on D+S, S is a subset of K_D
- **REVISION+OPENBOOK**
  - Pre-Trained (MLM+NSP [BERT] and MLM [RoBERTa]) on K_D
  - Fine-Tuned on D+S, S is a subset of K_D

K_D - Respective knowledge sources for given 3 datasets D

Flow-Diagram of our Weighted-Sum Model



A sample from PIQA dataset

Results : Training Strategies with Knowledge infusion modes

		BERT				RoBERTa			
Dataset	**Strategy**	**Concat**	**Max**	**Sim-Sum**	**Wtd-Sum**	**Concat**	**Max**	**Sim-Sum**	**Wtd-Sum**
aNLI	OPENBOOK	73.9± 0.8	73.7± 0.1	73.5± 0.7	73.3± 1.0	83.9± 0.5	80.8± 0.9	81.7± 0.6	84.4± 0.4
	REVISION	72.7± 0.3	N/A	N/A	N/A	82.4	N/A	N/A	N/A
	REVISION & OPENBOOK	74.4± 0.2	74.3± 0.1	74.0± 0.9	75.1±0.4	84.2± 0.7	81.4± 0.8	82.6± 0.6	**86.7**± 0.6
PIQA	OPENBOOK	67.8± 0.4	72.4± 0.6	72.6± 1.2	72.5± 0.1	74.8± 0.5	75.2± 0.9	75.6± 0.7	77.1± 0.2
	REVISION	74.5± 0.3	N/A	N/A	N/A	75.2± 0.8	N/A	N/A	N/A
	REVISION & OPENBOOK	67.7± 0.1	73.8± 0.8	76.8± 0.5	76.8± 0.3	75.4± 0.7	76.2± 0.8	76.8± 0.4	**80.2**± 0.6
SIQA	OPENBOOK	70.1± 0.8	67.8± 0.1	70.0± 0.7	70.2± 0.4	76.5± 0.7	77.2± 0.6	77.4± 0.2	78.3± 0.5
	REVISION	69.5± 0.9	N/A	N/A	N/A	76.8± 0.3	N/A	N/A	N/A
	REVISION & OPENBOOK	68.8± 0.4	66.6± 0.4	68.9± 0.1	69.3± 0.6	78.2± 0.3	77.4± 0.9	76.7± 0.5	**79.5**± 0.9

- Revision with Openbook strategy works best across 3 datasets for each of 4 knowledge infusion modes
- Weighted sum model shows the best performance across all datasets

Results : Weighted Sum model Performance Comparison

Models/ Accuracy	aNLI		PIQA		SIQA	
	Val	Test	Val	Test	Val	Test
BERT	67.36	<u>66.75</u>	68.08	<u>69.23</u>	64.88	<u>64.50</u>
GPT-2 XL	N/A	N/A	70.20	<u>69.50</u>	47.50	<u>45.30</u>
RoBERTa	85.05	<u>83.91</u>	76.28	<u>76.80</u>	77.85	<u>76.74</u>
RoBERTa 5 Ensemble	N/A	<u>83.22</u>	N/A	79.66	N/A	78.68
$L2R^2$ [2020]	N/A	**86.81**	N/A	N/A	N/A	N/A
KagNet [2019]	N/A	N/A	N/A	N/A	65.05	<u>64.59</u>
GBR [2020]	N/A	N/A	N/A	N/A	75.64	<u>76.25</u>
UnifiedQA T5 11B [2020]	N/A	<u>80.04</u>	N/A	**89.50**	N/A	**79.75**
Ours: BERT + WS	74.60	74.96	76.82	72.28	70.21	67.22
Ours: RoBERTa + WS	85.90	84.18	80.20	78.24	79.53	78.00

- WS with BERT and RoBERTa is better than baseline BERT and RoBERTa
- Better than huge parameterized model (T5 in aNLI)
- WS shows better performance than complex Graph based models like KagNet or GBR

Analysis

Types of Error	aNLI	SIQA	PIQA
Annotation	41%	38%	10%
Model Prediction	48%	27%	29%
Distracting Knowledge	11%	35%	61%

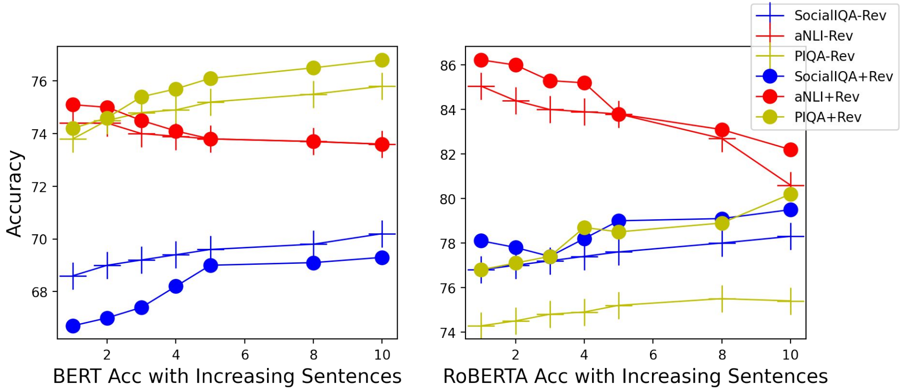
Error Percentage

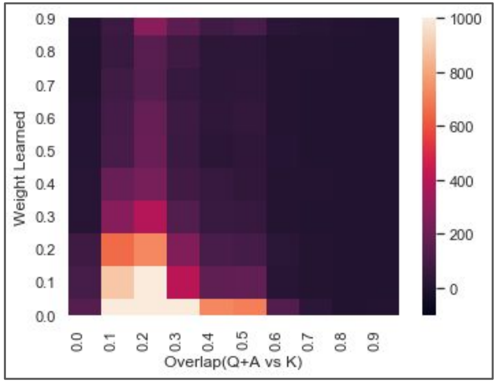
Knowledge	aNLI	SIQA	PIQA
Explicitly Present	14%	11%	10%
Implicitly Present	55%	59%	51%
Fully Irrelevant	31%	30%	39%

Knowledge Classification - Correct Predictions

- For PIQA we find most of the errors are due to distracting knowledge
- Over 50% of the correct predictions for each of the 3 datasets have implicit knowledge

Analysis of the Weighted Sum model





- We show the impact of increasing knowledge sentences on the accuracy with Revision training strategy for both BERT and RoBERTa
    - It increases for SIQA and PIQA
- We also show the weights learned by the weight layer of the WS model
    - We find that the model learns lower weights where there is very less lexical overlap between knowledge retrieved and Question and Answer options

Conclusion

In this paper through 3 commonsense reasoning dataset,

- We perform a thorough analysis of transformers’ (BERT and RoBERTa) ability to reason.
- We present four modes of knowledge infusion in transformer encoders which improves 2-9% of accuracy across the datasets
- We carry out an extensive investigation to study the impact of different knowledge sources and pre-training on such knowledge sources.

Thank You !!!

Contacts : Pratyay Banerjee, Swaroop Mishra, Kuntal Pal, Arindam Mitra, Chitta Baral

Paper : https://openreview.net/pdf?id=a4-fFL7aCi0

Email : {pbanerj6, srmishr1, kkpal, chitta}@asu.edu, arindam.mitra@microsoft.com