Improving Question Answering by Commonsense-Based Pre-Training

Wanjun Zhong, Duyu Tang, Nan Duan, Ming Zhou, Jian Yin et.

Wanjun Zhong Oct 20

Microsoft

Content

Part A: What's question answering?

Part B: What's common sense? Why we need it?

Part C: Introduce the dataset and commonsense knowledge base

Part D: Introduce related work

Part E: Introduce our work (Motivation, Model, Experiment)

Content

Part A: What's question answering?

Part B: What's common sense? Why we need it?

Part C: Introduce the dataset and commonsense knowledge base

Part D: Introduce related work

Part E: Introduce our work (Motivation, Model, Experiment)

What's question answering?

One type – machine reading based question answering

Input : passage, question, several candidate answers

Output: predicted answer with highest score

Example:

Passage

I was finally able to get my driving permit and it was time for my first driving lesson I was so excited to meet my instructor and drive their car for the first time I got behind the wheel made sure I checked the mirrors so I could see everything around me I put my seat belt on and told the instructor to put theirs on too I adjusted my seat so I could reach the pedals and steering wheel comfortably It was time to put the key in the ignition and start the car After the car was on I checked the mirrors to make sure I would not hit anything and backed out the parking spot He instructed me to drive the car around the block to make sure I knew the basics of driving After he felt comfortable we went out onto the road We drove for a few miles before going back to the school"

Question

Why did they take the driving lesson

Candidate Answers

(A) working towards a driver 's license
(B) just for fun

Correct Answer

(A) working towards a drivers 's license

Machine Reading Based Question Answering

- Most state–of–art method: Deep–Learning
- Tacled as a multiclass–classification problem: We want the answer with highest probability be the predicted answer.

$$p(a|d,q) \sim exp(W(a)g(d,q)), s.t. a \in V$$

(a denotes answer, d denotes document, q denotes question)

- How to model question and document? – **Basic Architecture**

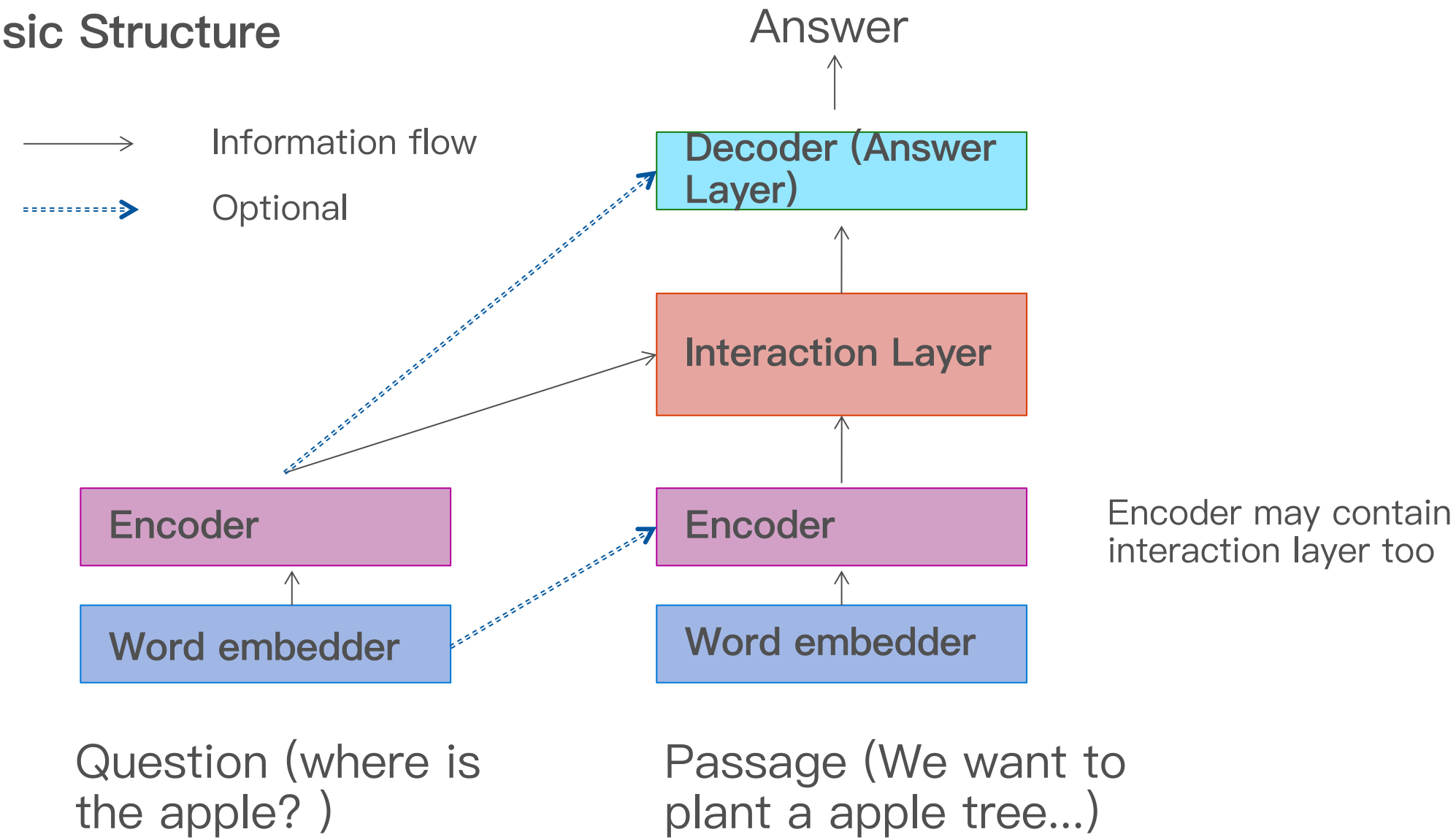
– **Encoder** – Read it carefully

– **Attention Layer** – Read the passage with question

– **Decoder (Answer Layer)** – Write down the answer

Machine Reading Based Question Answering

- **Basic Structure**



Machine Reading Based Question Answering

- **Function of each component**

– **Embedder** – Embed the word into vectors (glove, Elmo etc)

– **Encoder** – Embed the text into vector space (most simple: LSTM)

– **Attention Layer** – Attention layer to learn joint representation of question and passage, generate question–aware passage representation

– **Decoder (Answer Layer)** – predict the answer by question–aware passage representation

Content

Part A: What's question answering?

Part B: What's common sense? Why we need it?

Part C: Introduce the dataset and commonsense knowledge base

Part D: Introduce related work

Part E: Introduce our work (Motivation, Model, Experiment)

Why we need common sense?

We want the machine can answer question beyond the passage.

Just like humans do

What's common sense? : Common sense is our background knowledge about the world. For instance, we know “cat eats fish”,“apple grows on the tree” and “car has four tyres”.

Problem: Current deep learning models lack the ability to answer question and do reasoning using common sense.

Trace: Unsupervised learning and common sense reasoning becoming more and more important in recent years (EMNLP 2018 best paper: unsupervised machine translation). Moreover, several datasets needs common sense are proposed in 17 and 18. For example: ARC, SemEval2018 Task 11, SWAG, Event2Mind, Propara etc.

Why we need common sense?

Passage

I look out my window at my front yard and notice that my lawn is growing a little too tall I think it is time to mow it I check the sky for rainy clouds to make sure it will not rain soon and also check the Internet for any rain later today It will not rain today so it is safe to mow the lawn on my lawnmower My lawnmower has a motor and a seat for me to ride in so it is much easier to use I go into my garage to make sure my lawnmower has enough gas to work The lawnmower has plenty of gas and is in good shape so I turn the key and start its engine The lawnmower does n't cut the grass unless you pull the lever that controls the blade so I pull that lever and start to drive the lawnmower over the grass I make sure the grass is being cut after I drive over it and carefully steer the lawnmower over all the tall grass in my yard until the grass is cut"

Question

Where did they find the mower

Candidate Answers

(A) In the garage
(B) In the basement

Correct Answer

(A) In the garage

Knowledge

[window, AtLocation, garage], [lawnmower, AtLocation, garage],[mower, AtLocation, garage]

Where can we find common sense? – external structured commonsense knowledge base and unstructured text

Content

Part A: What's question answering?

Part B: What's common sense? Why we need it?

Part C: Introduce the dataset and commonsense knowledge base

Part D: Introduce related work

Part E: Introduce our work (Motivation, Model, Experiment)

Dataset – ARC (AI2 Reasoning Challenge)

AI2 Reasoning Challenge (ARC)

Dataset requires more powerful knowledge and reasoning, contains only natural and grade–school science questions. It is separated into Challenge Set and Easy Set. Challenge Set is hard to be solved by traditional method.

- **Dataset component:** question, multiple choice (4), large science corpus (14M scientific facts)
- **Example**:

A student riding a bicycle observes that it moves faster on a smooth road than on a rough road. This happens because the smooth road has (A) less gravity (B) more gravity (C) less friction [correct] (D) more friction

http://data.allenai.org/arc/

Dataset – SemEval 2018 Task 11

In this task, systems will be presented with narrative texts about everyday activities and are required to answer multiple-choice questions based on this text. This task assesses how the inclusion of commonsense knowledge in the form of script knowledge

- **Dataset component:** passage, question, multiple choice (2)
- **Example**:

```
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.
```

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

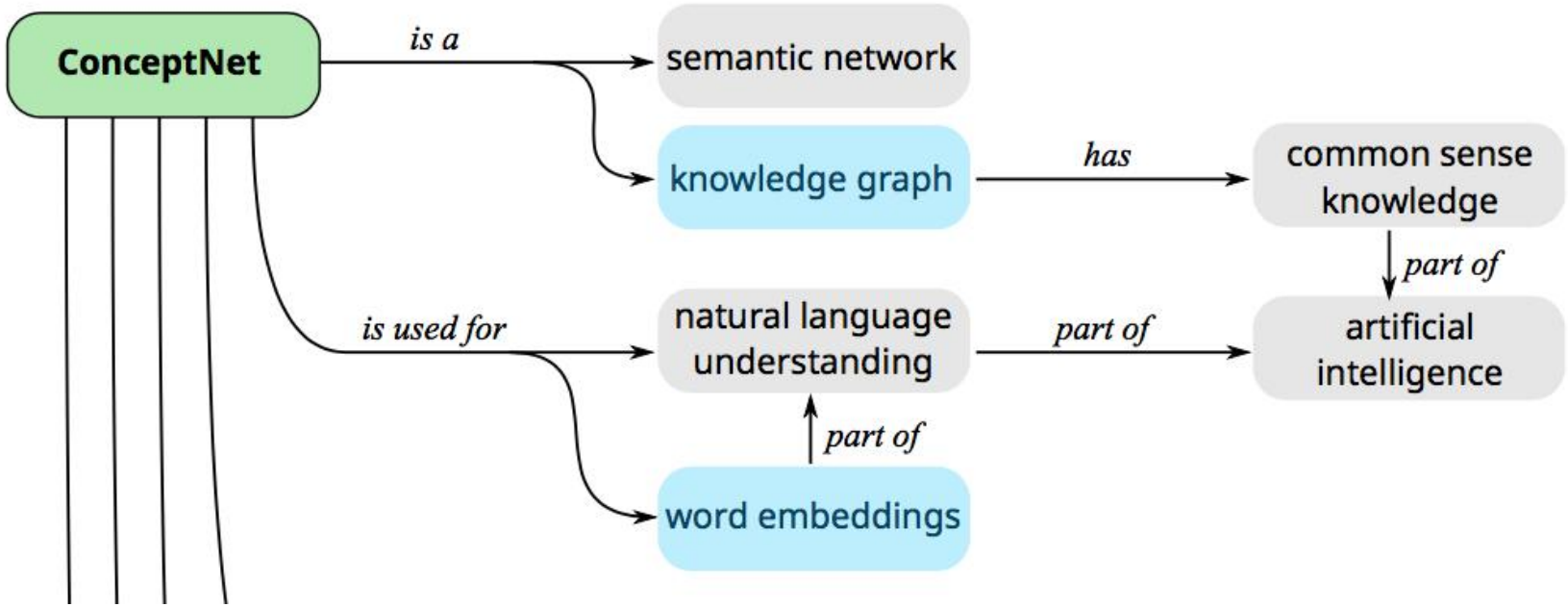
https://competitions.codalab.org/competitions/17184

Commonsense knowledge base – ConceptNet

ConceptNet is a freely–available semantic network, designed to help computers understand the meanings of words that people use

Data

Knowledge base components: 657367 triples and 39 types of relations (IsA, PartOf etc..)



http://conceptnet.io

Commonsense knowledge base – WebChild

WebChild is a large collection of commonsense knowledge, automatically extracted and disambiguated from Web contents. WebChild contains triples that connect nouns with adjectives via fine–grained relations like hasShape, hasTaste, evokesEmotion, etc

Knowledge base components: Millions of knowledge contain 5 types of relation.

relation	range	domain	assertions
hasTaste	sweet-a^1	strawberry-n^1	(chocolate-n^1,creamy-a^2)
	hot-a^9	chili-n^1	(pizza-n^1,delectable-a^1)
	sour-a^2	salsa-n^1	(salsa-n^1,spicy-a^2)
	salty-a^3	sushi-n^1	(burger-n^1,tasty-a^1)
	lemony-a^1	java-n^2	(biscuit-n^2,sweet-a^1)
hasShape	triangular-a^1	leaf-n^1	(palace-n^1,domed-a^1)
	meandering-a^1	circle-n^1	(table-n^2,flat-a^1)
	crescent-a^1	ring-n^8	(jeans-n^2,tapered-a^1)
	obtuse-a^2	egg-n^1	(tv-n^2,flat-a^1)
	tapered-a^1	face-n^1	(lens-n^1,spherical-a^2)

http://resources.mpi–inf.mpg.de/d5/webchild/

Content

Part A: What's question answering?

Part B: What's common sense? Why we need it?

Part C: Introduce the dataset and commonsense knowledge base

Part D: Introduce related work

Part E: Introduce our work (Motivation, Model, Experiment)

Related work considering commonsense knowledge

Modeling common sense is a relatively new topic in recent years. Existing method roughly fall into two groups:

1. **Enhance the basic computational units (word or noun phrase) (Majority)**
2. **Support external signals at the top layer before the model makes the final decision (ours)**

Our work also relates to top-performing machine reading model, which applied deep-learning.

Related work considering commonsense knowledge

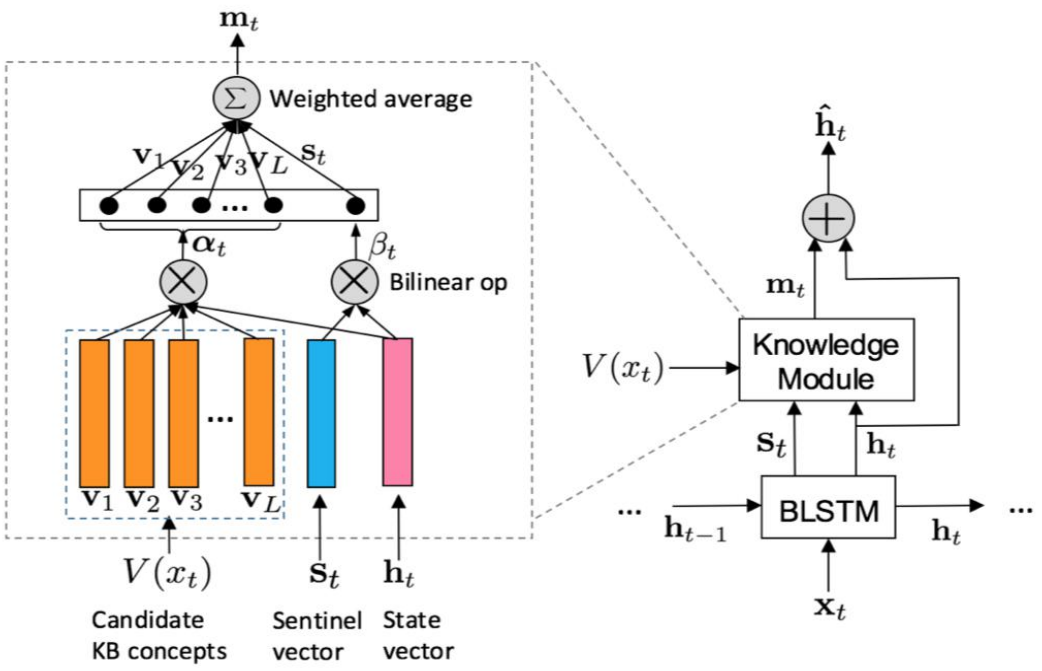


Figure 1: Architecture of the KBLSTM model. As each time step t, the knowledge module retrieves a set of candidate KB concepts $V(x_t)$ that are related to the current input x_t, and then computes a knowledge state vector $\mathbf{m}_t$ that integrates the embeddings of the candidate KB concepts $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_L$ and the current context vector $\mathbf{s}_t$. See Section § 3.2 for details.

Weakness of this kind of method:

1. Depend highly on the distribution of the data.
2. Can't perform complex reasoning
3. Hard to explain

Content

Part A: What's question answering?

Part B: What's common sense? Why we need it?

Part C: Introduce the dataset and commonsense knowledge base

Part D: Introduce related work

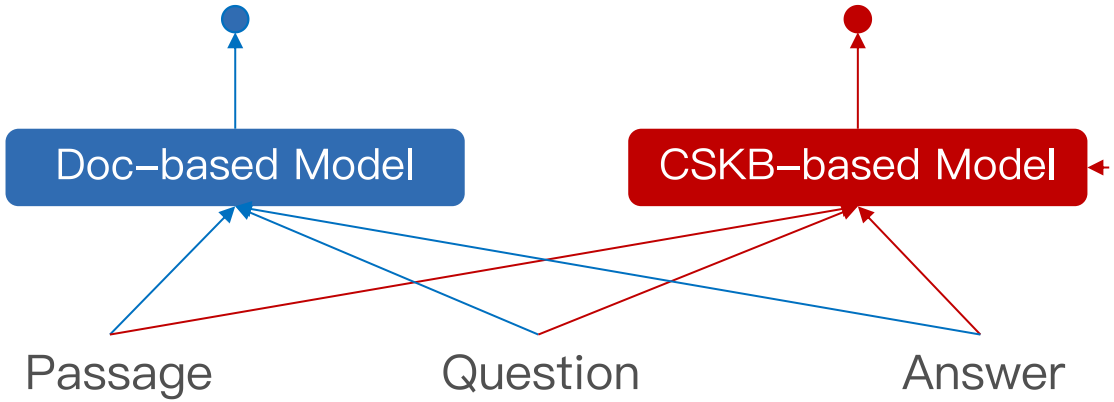
Part E: Introduce our work (Motivation, Model, Experiment)

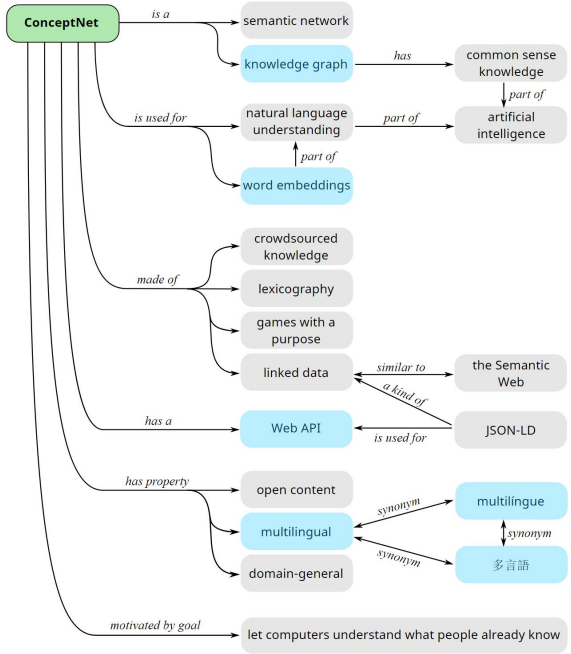
Motivation

- To perform commonsense reasoning, we leverage external commonsense knowledge, such as ConceptNet , to improve the commonsense reasoning capability of a question answering (QA) system.

To achieve this goal, we pre-train a generic model from external commonsense knowledge about the world

$$f(p,q,c) = \alpha \times f_{doc}(p,q,c) + \beta \times f_{kb}(p,q,c)$$





Motivation

Our model performs commonsense reasoning by modeling the connection between different concepts. These connection can be divided into (a) direct and (b) indirect ones

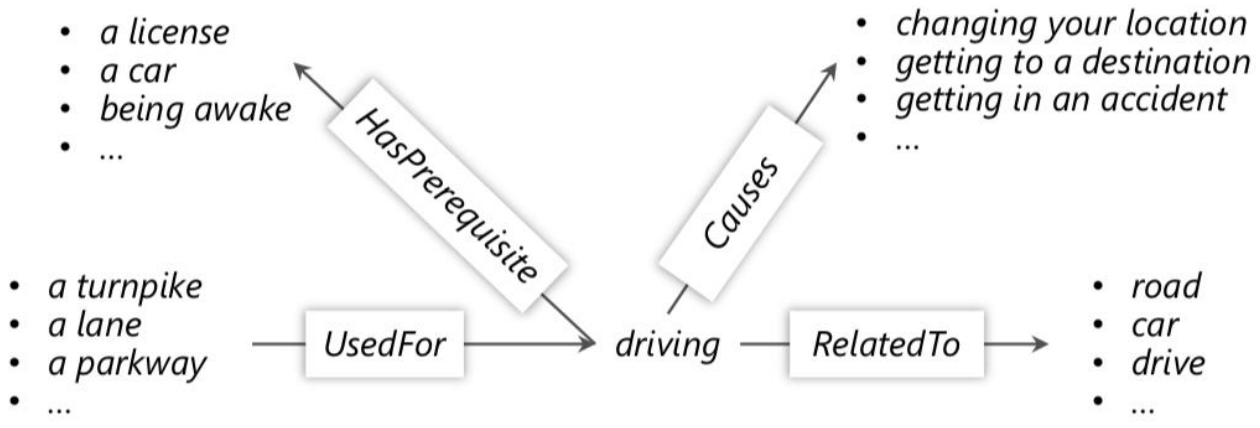


Figure 2: A sampled subgraph from ConceptNet with "*driving*" as the central word.

Direct connection:

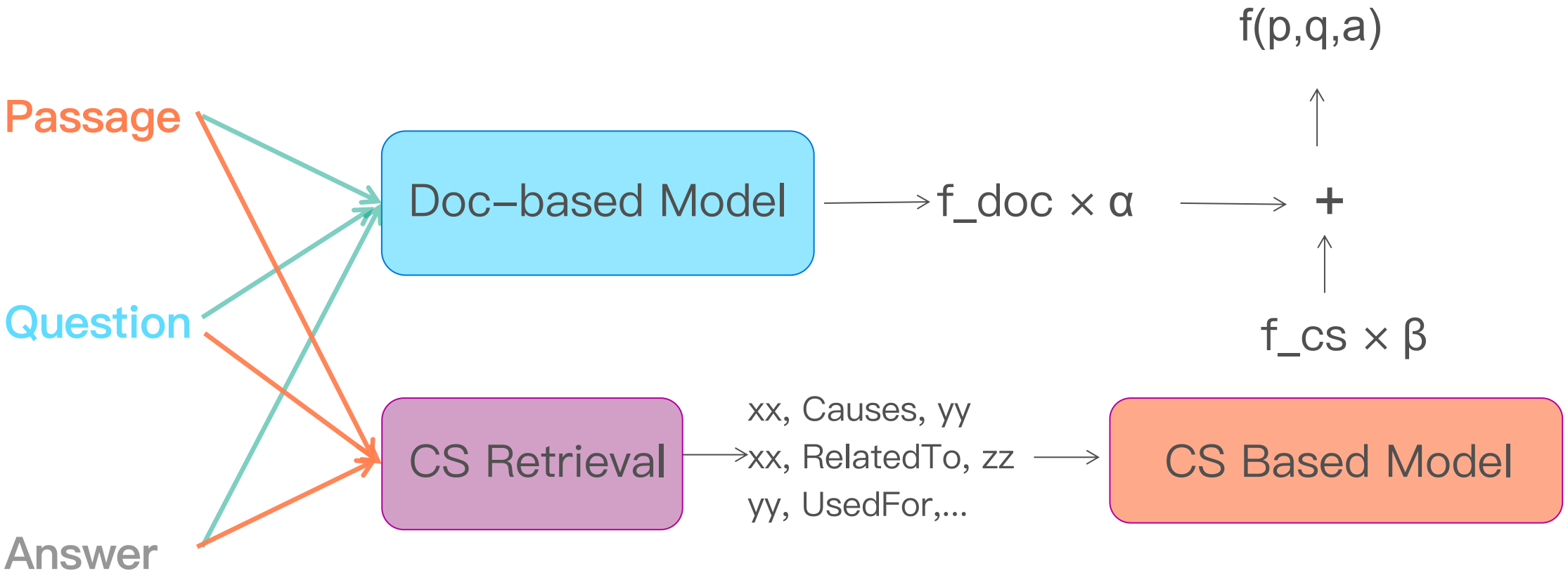
{"driving", "a license"}

{"driving", "road"}

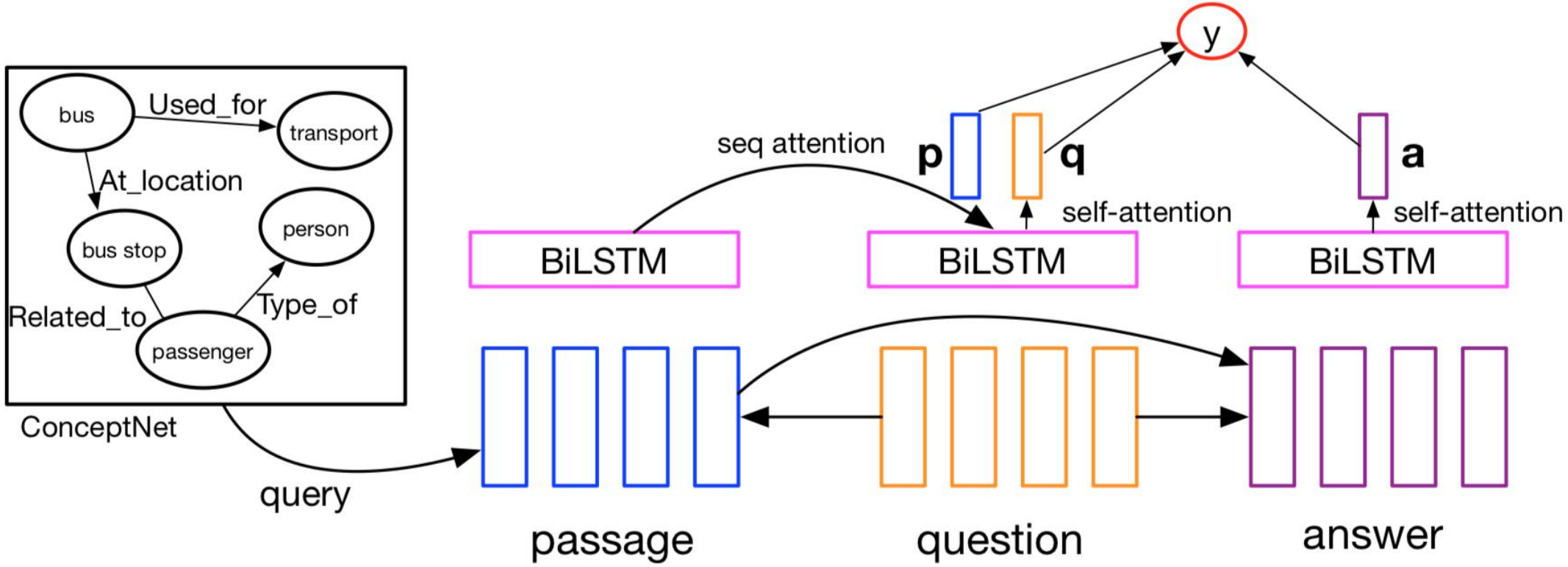
Indirect connection:

{"a car", "getting to a destination"} conncted by pivot "driving"

Approach Overview

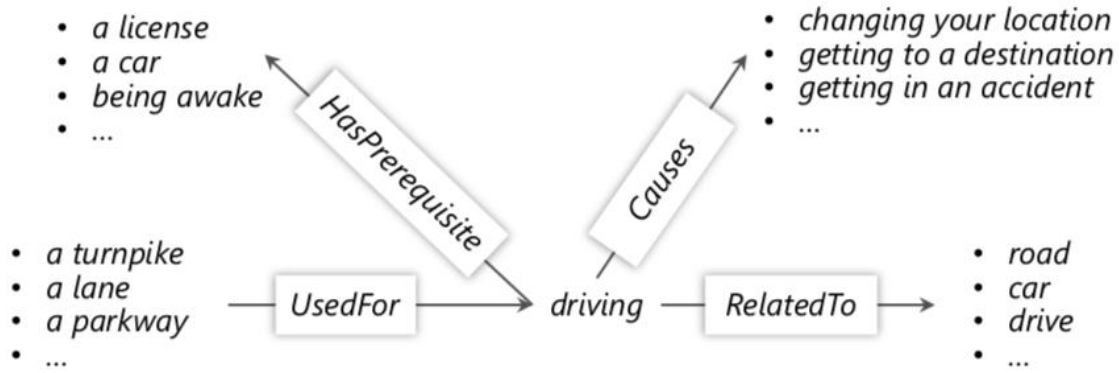


Doc–based Model



Commonsense based Model

Concept "driving" & neighbouring graph structure



Step 1: Sample selector

positive instance: pairs of direct–connected concepts

negative instance: pairs of randomly selected disconnected concepts

Step 2: Encoder of internal words and external graph structure

$$h^w(c) = BiLSTM(Emb(c))$$

$$h^n(c) = \sum_{c' \in NBR(c)} (W^{r(c,c')} h^w(c') + b^{r(c,c')})$$

$$Enc(c) = [h^w(c); h^n(c)]$$

Commonsense based Model

Step 3: Score Computing of two concepts

$$f_{cs}(c_1, c_2) = Enc(c_1) \odot Enc(c_2)$$

Step 4: Objective function (Max–Margin Loss)

$$l(c_1, c_2, c') = max(0, f_{cs}(c_1, c') - f_{cs}(c_1, c_2) + mgn)$$

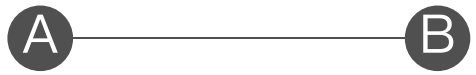
After training on ConceptNet, then we can obtain the score of two concept. And how we calculate the score between two sets of concepts?

$$f_{cs}(a_i) = \frac{1}{|E_1|} \sum_{x \in E_1} \max_{y \in E_2} (f_{cs}(x, y))$$

c_2 denote the positive instance, c' negative. E1, E2 are different sets of concepts

Evaluation on Academic Dataset

- Relation Missing



- Pivot Missing

- Results on ARC dataset (AI2)

Method	Accuracy
DGEM	26.41
KG^2 (state-of-the-art)	31.70
Fdoc	31.25
Fdoc + Fkb_rm	32.28
Fdoc + Fkb_pm	32.96
Fdoc + Fkb_rm + Fkb_pm	33.39

- Results on SemEval dataset

Method	Accuracy
Fdoc (one of top ranked systems in SemEval 18)	80.33
Fdoc + Fkb_rm	81.58
Fdoc + Fkb_pm	81.44
Fdoc + Fkb_rm + Fkb_pm	81.80

Passage

I was finally able to get my driving permit and it was time for my first driving lesson I was so excited to meet my instructor and drive their car for the first time I got behind the wheel made sure I checked the mirrors so I could see everything around me I put my seat belt on and told the instructor to put theirs on too I adjusted my seat so I could reach the pedals and steering wheel comfortably It was time to put the key in the ignition and start the car After the car was on I checked the mirrors to make sure I would not hit anything and backed out the parking spot He instructed me to drive the car around the block to make sure I knew the basics of driving After he felt comfortable we went out onto the road We drove for a few miles before going back to the school"

Question

Why did they take the driving lesson

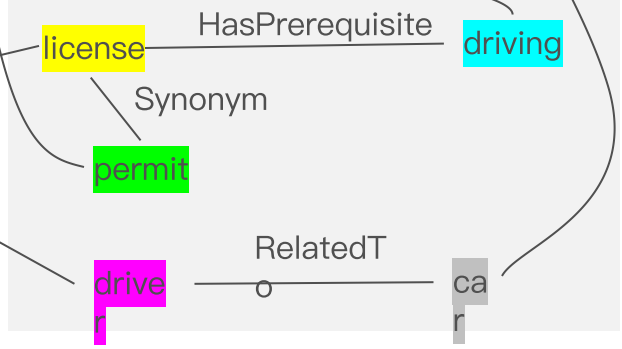
Candidate Answers

(A) working towards a driver 's license
(B) just for fun

Correct Answer

(A) working towards a drivers 's license

Retrieved Knowledge



Advantage of Our Model

- Our model is task independent and is trained locally, which means our model can be easily corporated into other task needs commonsense reasoning.
- Our model can perform commonsense reasoning of the connection of two concepts, rather than purely encode them into the model.
- Our model obtain state-of-art performance on ARC leaderboard.

Leaderboard

Rank	Model	Test Score
1	ET-RR 08/29/2018 8:00:00 AM	36.36%
2	BiLSTM Max-out 08/22/2018 8:00:00 AM	33.87%
3	TriAN + f(dir)(cs) + f(ind)(cs) 09/06/2018 8:00:00 AM	33.39%
4	Anonymous (paper under conference review) 09/20/2018 8:00:00 AM	33.20%
5	KG^2 06/01/2018 8:00:00 AM	31.70%
6	DGEM 04/02/2018 8:00:00 AM	27.11%
7	TableILP 04/02/2018 8:00:00 AM	26.97%

Thanks!

Wanjun Zhong Oct.20

Microsoft