Neural Question Answering

CS 287

Review: Pooling

One common class of operations in neural network models is known as *pooling*. Informally a pooling layer consists of aggregation unit, typically unparameterized, that reduces the input to a smaller size.

Consider three pooling functions of the form $f : \mathbb{R}^n \mapsto \mathbb{R}$,

1. $f(\mathbf{x}) = \max_i x_i$
2. $f(\mathbf{x}) = \min_i x_i$
3. $f(\mathbf{x}) = \sum_i x_i / n$

Review: Neural MT

Variable-length source encoding vectors from (bi)-LSTM,

$$\mathbf{s}_1^s, \mathbf{s}_2^s, \ldots \mathbf{s}_n^s$$

Want to construct single vector $\mathbf{s}_{pool}^s$.

However we also have $\mathbf{s}^t$, last RNN state

1. Could use standard naive average pooling.
2. Could use current context to help out.

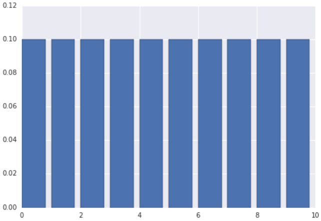
Average Pooling

Variable-length source encoding vectors from (bi)-LSTM,

$$\mathbf{s}_1^s, \mathbf{s}_2^s, \ldots \mathbf{s}_n^s$$

- Average pooling (expectation under uniform: $\alpha_j = \frac{1}{n}$)

$$f(\mathbf{s}^s) = \sum_{j=1}^{n} \frac{1}{n} \mathbf{s}_j^s = \sum_{j=1}^{n} \alpha_j \mathbf{s}_j^s$$



Hard-Attention Pooling

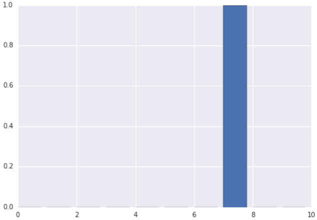
Compute score to determine selection of pooling,

$$z_j = s(\mathbf{s}_j^s, \mathbf{s}^t) = \tanh([\mathbf{s}^t, \mathbf{s}_j^s]\mathbf{W} + \mathbf{b})$$

$$j^* = \arg\max_j z_j$$

- Selection pooling (expectation under one-hot: $\alpha_j = \delta(j = j^*)$)

$$f(\mathbf{s}^s) = \mathbf{s}_{j^*}^s = \sum_{j=1}^{n} \alpha_j \mathbf{s}_j^s$$



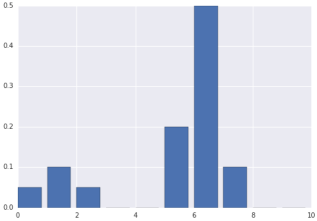
Attention-Based Pooling

Compute score to determine selection of pooling,

$$z_j = s(\mathbf{s}_j^s, \mathbf{s}^t) = \tanh([\mathbf{s}^t, \mathbf{s}_j^s]\mathbf{W} + \mathbf{b})$$

- Selection pooling (expectation under softmax: $\boldsymbol{\alpha} = \text{softmax}(\mathbf{z})$)

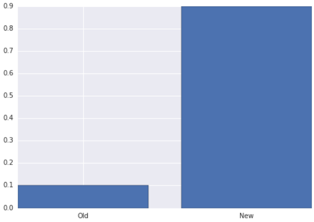
$$f(\mathbf{s}^s) = \sum_{j=1}^{n} \alpha_j \mathbf{s}_j^s$$



Recall: Dynamic Skip-Connections

$$
\begin{aligned}
NN_{sl2}(\mathbf{x}) &= (1-t)\,\text{ReLU}(\mathbf{x}\mathbf{W}^1 + \mathbf{b}^1) + t\mathbf{x} \\
t &= \sigma(\mathbf{x}\mathbf{W}^t + b^t) \\
\mathbf{W}^1 &\in \mathbb{R}^{d_{\text{hid}} \times d_{\text{hid}}} \\
\mathbf{W}^t &\in \mathbb{R}^{d_{\text{hid}} \times 1}
\end{aligned}
$$

- Here attention is between keeping state and updating.



Soft versus Hard Pooling

"Soft" attention-based Pooling

- Can backprop through to learn params.
- Allows combination of elements.

"Hard" attention-based

- Forces a single-correct answer.
- Can train attention separately (if you have annotations)
- Can also train with reinforcement learning (see use in Xu et al (2015), REINFORCE (Williams, 1992))

A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

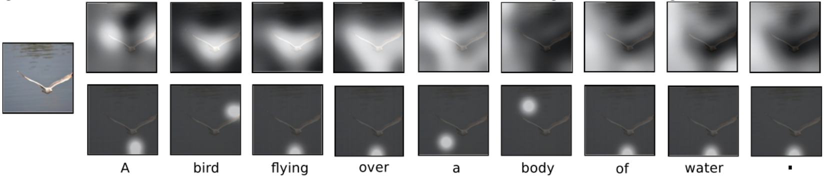
A stop sign is on a road with a mountain in the background.

A little girl sitting on a bed with a teddy bear.

A group of people sitting on a boat in the water.

A giraffe standing in a forest with trees in the background.

Figure 2. Attention over time. As the model generates each word, its attention changes to reflect the relevant parts of the image. "soft" (top row) vs "hard" (bottom row) attention. (Note that both models generated the same captions in this example.)



Dataset	Model	BLEU				METEOR
		B-1	B-2	B-3	B-4	
Flickr8k	Google NIC(Vinyals et al., 2014)$^{\dagger\Sigma}$	63	41	27	—	—
	Log Bilinear (Kiros et al., 2014a)$^{\circ}$	65.6	42.4	27.7	17.7	17.31
	Soft-Attention	**67**	44.8	29.9	19.5	18.93
	Hard-Attention	**67**	**45.7**	**31.4**	**21.3**	**20.30**
Flickr30k	Google NIC$^{\dagger\circ\Sigma}$	66.3	42.3	27.7	18.3	—
	Log Bilinear	60.0	38	25.4	17.1	16.88
	Soft-Attention	66.7	43.4	28.8	19.1	**18.49**
	Hard-Attention	**66.9**	**43.9**	**29.6**	**19.9**	18.46
COCO	CMU/MS Research (Chen & Zitnick, 2014)^{a}	—	—	—	—	20.41
	MS Research (Fang et al., 2014)$^{\dagger a}$	—	—	—	—	20.71
	BRNN (Karpathy & Li, 2014)$^{\circ}$	64.2	45.1	30.4	20.3	—
	Google NIC$^{\dagger\circ\Sigma}$	66.6	46.1	32.9	24.6	—
	Log Bilinear$^{\circ}$	70.8	48.9	34.4	24.3	20.03
	Soft-Attention	70.7	49.2	34.4	24.3	**23.90**
	Hard-Attention	**71.8**	**50.4**	**35.7**	**25.0**	23.04

Quiz: Attention

You have coded up an attention-based neural translation model and a non-attention-based encoder-decoder model. At runtime you produce translations with greedy search.

- How does the time complexity and space complexity compare between the two models?
- How might impact this approach if you switched to a much larger problem (say translating full documents)?

Last Lecture: Semantics

Branch of linguistic focused on meaning

- Compositional Semantics
  - Meaning of utterances
  - Concerned with the relations of meaning
  - Often expressed with logical relations

Neural methods for semantic-tasks with no explicit logic.

Today's Lecture

Neural Question Answering

- MemNN and on simple tasks
- Neural approaches to WebQuestions
- Other QA Domains

Contents

MemNN

Factoid QA

Read and Comprehend

Memory Networks (Weston et al, 2014)

- General architecture for models with "memory"
- Memory is encoded in an explicit store
- Network repeatedly processes input produces output.

The Architecture

- I Read a symbolic input value
- G Produce a new generalized memory
- O Create an output representation from memories
- R Respond with a symbolic output.

NMT as MemNN

Assume we start with source-memory representations $\mathbf{s}_j^s$

- I Read last target word
- G Update target RNN representation
- O Pool over source “memories” (either soft- or hard)
- R Respond with next target word.

bAbI Tasks (Weston et al, preprint)

```
1 Mary moved to the bathroom.
2 John went to the hallway.
3 Where is Mary?        bathroom        1
4 Daniel went back to the hallway.
5 Sandra moved to the garden.
6 Where is Daniel?      hallway 4
7 John moved to the office.
8 Sandra journeyed to the bathroom.
9 Where is Daniel?      hallway 4
10 Mary moved to the hallway.
11 Daniel travelled to the office.
12 Where is Daniel?     office  11
```

bAbI

- Synthetic tasks to test MemNN type architectures
- Running reading comprehension from a synthetic domain.
- 20 different tasks ranging in complexity

TASK

1 -3 Supporting Facts

4-5 - Arg. Relations

6 - Yes/No Questions

7 - Counting

8 - Lists/Sets

9 - Simple Negation

10 - Indefinite Knowledge

11 - Basic Coreference

12 - Conjunction

13 - Compound Coref.

14 - Time Reasoning

15-16 - Basic Deduction / Basic Induction

17 - Positional Reasoning

18 - Size Reasoning

19 - Path Finding

Simple Model (Non-Questions)

- Mary moved to the bathroom.

I Read in words

G Construct and append CBoW sentence representation

$$\mathbf{s}_j = G(\mathbf{x}^0) = \sum_{i=1}^{n} \mathbf{W}\mathbf{x}_i^0$$

O,R Nothing

Simple Model (Questions)

- Where is Mary?

I Read in words

G Construct CBoW sentence representation

$$\mathbf{x} = G(\mathbf{x}^0) = \sum_{i=1}^{n} \mathbf{W}\mathbf{x}_i^0$$

O Find best sentence match and apply hard attention

$$j^* = \arg\max_j s(\mathbf{x}, \mathbf{s}_j)$$

R Respond by classification over possible outputs

$$\hat{\mathbf{y}} = \text{softmax}(NN_{MLP}([\mathbf{x}; \mathbf{s}_{j^*}]))$$

bAbI

1. Mary moved to the bathroom.
2. John went to the hallway.
3. Where is Mary?

- Use CBoW to match and to help predict answer (bathroom)

Multiple Hops

Task 18: Size Reasoning

1. The football fits in the suitcase.
2. The suitcase fits in the cupboard.
3. The box is smaller than the football.
4. Will the box fit in the suitcase? A:yes
5. Will the cupboard fit in the box? A:no

Multi-Hop Model (Non-Questions)

- 3 Will the box fit in the suitcase?

I Read in words

G Construct CBoW sentence representation

$$\mathbf{x} = G(\mathbf{x}^0) = \sum_{i=1}^{n} \mathbf{W}\mathbf{x}_i^0$$

O Find best sentence match and apply hard attention

$$j^* = \arg\max_j s(\mathbf{x}, \mathbf{s}_j)$$

$$k^* = \arg\max_k s(\mathbf{x}, \mathbf{s}_{j^*}, \mathbf{s}_k)$$

R Respond by classification over possible outputs

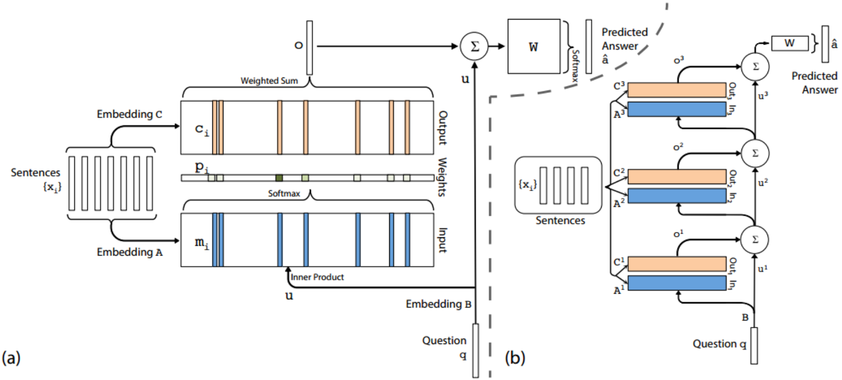
$$\hat{\mathbf{y}} = \mathrm{softmax}(NN_{MLP}([\mathbf{x}; \mathbf{s}_{j^*}; \mathbf{s}_{k^*}]))$$

How do we learn the hard-attention?

- Strong supervision
  - Hard-attention is trained in O step
  - No explicit backprop through decisions

- Weak supervision (End-to-End)
  - Soft-attention with no intermediary answers given
  - Use softmax to combine possible memories.

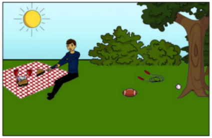
End-to-End MemNN (Sukhbaatar, 2015)



Other Tasks: Visual QA

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?

Contents

MemNN

Factoid QA

Read and Comprehend

Freebase | Find... | Browse | Query | Help | Sign In or Sign Up | English

Important! Freebase is read-only and will be shut-down. More.

Topic

Abraham Lincoln en

Created by book_bot on 5/7/2009

mid: /m/0gzh **notable type:** /government/us_president **notable for:** /government/us_president **on the web:** wikipedia.org

Abraham Lincoln was the 16th president of the United States, serving from March 1861 until his assassination in April 1865. Lincoln led the United States through its Civil War—its bloodiest war and its greatest moral, constitutional and political crisis. In doing so, he preserved the Union, abolished slavery, strengthened the federal government, and modernized the economy. Reared in a poor family on the western frontier, Lincoln was a self-educated lawyer in Illinois, a Whig Party leader, state legislator during the 1830s. Lincoln was elected to Congress in 1846, where he promoted rapid modernization of the economy through banks, tariffs, and railroads. He had originally agreed not to run for a second term and his opposition to the Mexican–American War was unpopular among the voters. He returned to Springfield and concentrated on his successful law practice throughout central Illinois. He returned to politics in 1854, and was a leader in building up the new Republican Party, which was had a statewide majority. After a series of highly publicized debates in 1858, during which Lincoln spoke out against the expansion of slavery, he lost the U.S. [-]

Properties | I18n | Keys | Links

View and edit specific domains, types, or properties...

Filter options: ☐ Show all domains and properties

Common /common

Freebase Commons

Topic /common/topic

- **Also known as** /common/topic/alias

Also known as

Honest Abe
Abe Lincoln
The Buffoon
Caesar
Father Abraham
The Flatboat Man
The Grand Wrestler
The Great Emancipator
The Illinois Baboon
The Jester

58 values total »

Description /common/topic/description

Abraham Lincoln was the 16th president of the United States, serving from March 1861 until his assassination in April 1865. Lincoln led the United States through its Civil War—its bloodiest war and its greatest moral, constitutional and political crisis. In doing so, he preserved the Union, abolished slavery, strengthened the federal government, and modernized the economy. Reared in a poor family on the western frontier, Lincoln was a self-educated lawyer in Illinois, a Whig Party leader, state legislator during the 1830s. Lincoln was elected to Congress in 1846, where he promoted rapid modernization of the economy through banks, tariffs, and railroads. He

Types:

Common
- Topic

Film
- Film story contributor
- Film subject

Government
- US President
- Politician
- Political Appointer
- U.S. Congressperson

Medicine
- Public figure with medical condition

Military
- Military Person
- Military Commander

Books
- Literature Subject

WebQuestions (Berant, 2013)

Questions:

- what high school did president bill clinton attend?
- what form of government does russia have today?
- what movies does taylor lautner play in?

Answers:

- Hot Springs High School
  `http://www.freebase.com/view/en/bill_clinton`
- Constitutional republic
  `http://www.freebase.com/view/en/russia`
- Eclipse, Valentine's Day, The Twilight Saga: Breaking Dawn - Part 1, New Moon
  `http://www.freebase.com/view/en/taylor_lautner`

Freebase Triples

(subject, relationship, object)

`What American cartoonist is the creator of Andy Lippincott?`

`(andy lippincott, character created by, garry trudeau)`

`Which forest is Fires Creek in?`

`(fires creek, containedby, nantahala national forest)`

`What is an active ingredient in childrens earache relief ?`

`(childrens earache relief, active ingredients, capsicum)`

`What does Jimmy Neutron do?`

`(jimmy neutron, fictional character occupation, inventor)`

Method:

- Embed Freebase entities and relations.
- Learn mapping between question words and freebase
- Include large amounts of semi-supervision to learn embeddings

Existing approaches for QA from KBs use learnable components to either transform the question into a structured KB query (Berant et al., 2013) or learn to embed questions and facts in a low dimensional vector space and retrieve the answer by computing similarities in this embedding space (Bordes et al., 2014a).

Start with memory containing all freebase triples

- I Input the questions
- G Compute continuous bag of n-grams
- O Score all triples based on embeddings

$$j^* = \arg\max_j s(\mathbf{x}, \mathbf{s}_j) = \arg\max_j \cos(\mathbf{x}, \mathbf{s}_j)$$

- R Respond with the matched triple result.

Hence, in this paper, we introduce a new dataset of much larger scale for the task of simple QA called SimpleQuestions. 2 This dataset consists of a total of 108,442 questions written in natural language by human English-speaking annotators each paired with a corresponding fact from FB2M that provides the answer and explains it.

The term simple QA refers to the simplicity of the reasoning process needed to answer questions, since it involves a single fact. However, this does not mean that the QA problem is easy per se, since retrieving this single supporting fact can be very challenging as it involves to search over millions of alternatives given a query expressed in natural language.

Contents

MemNN

Factoid QA

Read and Comprehend

	CNN			Daily Mail		
	train	valid	test	train	valid	test
# months	95	1	1	56	1	1
# documents	90,266	1,220	1,093	196,961	12,148	10,397
# queries	380,298	3,924	3,198	879,450	64,835	53,182
Max # entities	527	187	396	371	232	245
Avg # entities	26.4	26.5	24.5	26.5	25.5	26.0
Avg # tokens	762	763	716	813	774	780
Vocab size		118,497			208,045	

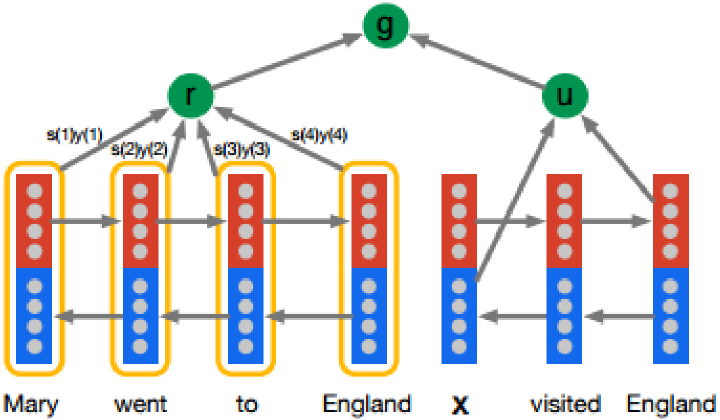
	Original Version	Anonymised Version
Context		
	The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the "Top Gear" host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon "to an unprovoked physical and verbal attack." …	the *ent381* producer allegedly struck by *ent212* will not press charges against the " *ent153* " host , his lawyer said friday . *ent212* , who hosted one of the most - watched television shows in the world , was dropped by the *ent381* wednesday after an internal investigation by the *ent180* broadcaster found he had subjected producer *ent193* " to an unprovoked physical and verbal attack . " …
Query		
	Producer **X** will not press charges against Jeremy Clarkson, his lawyer says.	producer **X** will not press charges against *ent212* , his lawyer says .
Answer		
	Oisin Tymon	*ent193*

The Attentive Reader can be viewed as a generalisation of the application of Memory Networks to question answering [3]. That model employs an attention mechanism at the sentence level where each sentence is represented by a bag of embeddings. The Attentive Reader employs a finer grained token level attention mechanism where the tokens are embedded given their entire future and past context in the input document.

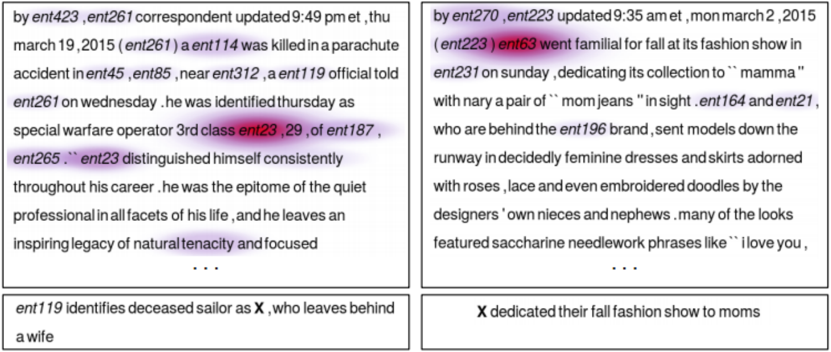
The Architecture

Read all bi-directional document positions

- I Read source with blank
- G Run RNN over the source
- O Find source entity with best match.
- R Score match at that position.



g
r
u
s(1)y(1)
s(2)y(2)
s(3)y(3)
s(4)y(4)
Mary
went
to
England
x
visited
England


by ent423 , ent261 correspondent updated 9:49 pm et , thu march 19 , 2015 (ent261) a ent114 was killed in a parachute accident in ent45 , ent85 , near ent312 , a ent119 official told ent261 on wednesday . he was identified thursday as special warfare operator 3rd class ent23 , 29 , of ent187 , ent265 . `` ent23 distinguished himself consistently throughout his career . he was the epitome of the quiet professional in all facets of his life , and he leaves an inspiring legacy of natural tenacity and focused
. . .
ent119 identifies deceased sailor as X , who leaves behind a wife
by ent270 , ent223 updated 9:35 am et , mon march 2 , 2015 (ent223) ent63 went familial for fall at its fashion show in ent231 on sunday , dedicating its collection to `` mamma " with nary a pair of `` mom jeans " in sight . ent164 and ent21 , who are behind the ent196 brand , sent models down the runway in decidedly feminine dresses and skirts adorned with roses , lace and even embroidered doodles by the designers ' own nieces and nephews . many of the looks featured saccharine needlework phrases like `` i love you ,
. . .
X dedicated their fall fashion show to moms