Text Understanding with the Attention Sum Reader Network

성유연

Reading Comprehension Question

Text Understanding with the Attention Sum Reader Network

Rudolf Kadlec, Martin Schmid, Ondrej Bajgar & Jan Kleindienst
IBM Watson
V Parku 4, Prague, Czech Republic
{rudolf_kadlec,martin.schmid,obajgar,jankle}@cz.ibm.com

Abstract

Several large cloze-style context-question-answer datasets have been introduced recently: the CNN and Daily Mail news data and the Children's Book Test. Thanks to the size of these datasets, the associated text comprehension task is well suited for deep-learning techniques that currently seem to outperform all alternative approaches. We present a new, simple model that uses attention to directly pick the answer from the context as opposed to computing the answer using a blended representation of words in the document as is usual in similar models. This makes the model particularly suitable for question-answering problems where the answer is a single word from the document. Ensemble of our models sets new state of the art on all evaluated datasets.

Document: What was supposed to be a fantasy sports car ride at Walt Disney World Speedway turned deadly when a Lamborghini crashed into a guardrail. The crash took place Sunday at the Exotic Driving Experience, which bills itself as a chance to drive your dream car on a racetrack. The Lamborghini's passenger, 36-year-old Gary Terry of Davenport, Florida, died at the scene, Florida Highway Patrol said. The driver of the Lamborghini, 24-year-old Tavon Watson of Kissimmee, Florida, lost control of the vehicle, the Highway Patrol said. (...)

Question: Officials say the driver, 24-year-old Tavon Watson, lost control of a _____

Answer candidates: Tavon Watson, Walt Disney World Speedway, Highway Patrol, Lamborghini, Florida, (...)

Answer: Lamborghini

Figure 1: Each example consists of a context document, question, answer cadidates and, in the training data, the correct answer. This example was taken from the CNN dataset (Hermann et al., 2015). Anonymization of this example that makes

Kadlec, Schmid, Bajgar, Kleindienst. Text Understanding with the Attention Sum Reader Network. 2016. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics

Reading Comprehension Question

Task

Answering a cloze-style question

Document: What was supposed to be a fantasy sports car ride at Walt Disney World Speedway turned deadly when a Lamborghini crashed into a guardrail. The crash took place Sunday at the Exotic Driving Experience, which bills itself as a chance to drive your dream car on a racetrack. The Lamborghini's passenger, 36-year-old Gary Terry of Davenport, Florida, died at the scene. Florida Highway Patrol said. The driver of the Lamborghini, 24-year-old Tavon Watson of Kissimmee, Florida, lost control of the vehicle, the Highway Patrol said. (...)
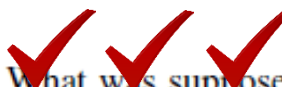
Question: Officials say the driver, 24-year-old Tavon Watson, lost control of a _____
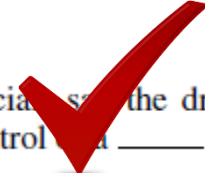
A   Tavon Watson

B   Highway Patrol

C   Walt Disney World Speedway

D   Lamborghini

Reading Comprehension Question

Task

Detect relationship between the question and the words in given document

Document: What was supposed to be a fantasy sports car ride at Walt Disney World Speedway turned deadly when a Lamborghini crashed into a guardrail. The crash took place Sunday at the Exotic Driving Experience, which bills itself as a chance to drive your dream car on a racetrack. The Lamborghini's passenger, 36-year-old Gary Terry of Davenport, Florida, died at the scene, Florida Highway Patrol said. The driver of the Lamborghini 24-year-old Tavon Watson of Kissimmee, Florida, lost control of the vehicle, the Highway Patrol said. (...)

Question: Official say the driver, 24-year-old Tavon Watson, lost control of a _____

Words to Machine

Vector embedding

Covert every word in text data into one hot vector

One hot Vector

	1	2	3	4	5	6	7	8	9
man	1	0	0	0	0	0	0	0	0
woman	0	1	0	0	0	0	0	0	0
boy	0	0	1	0	0	0	0	0	0
girl	0	0	0	1	0	0	0	0	0
prince	0	0	0	0	1	0	0	0	0
princess	0	0	0	0	0	1	0	0	0
queen	0	0	0	0	0	0	1	0	0
king	0	0	0	0	0	0	0	1	0
monarch	0	0	0	0	0	0	0	0	1

Vocabulary:
Man, woman, boy, girl, prince, princess, queen, king, monarch

Each word gets a 1x9 vector representation

Vector embedding

Covert every word in text data into one hot vector

One hot Vector

Vocabulary:
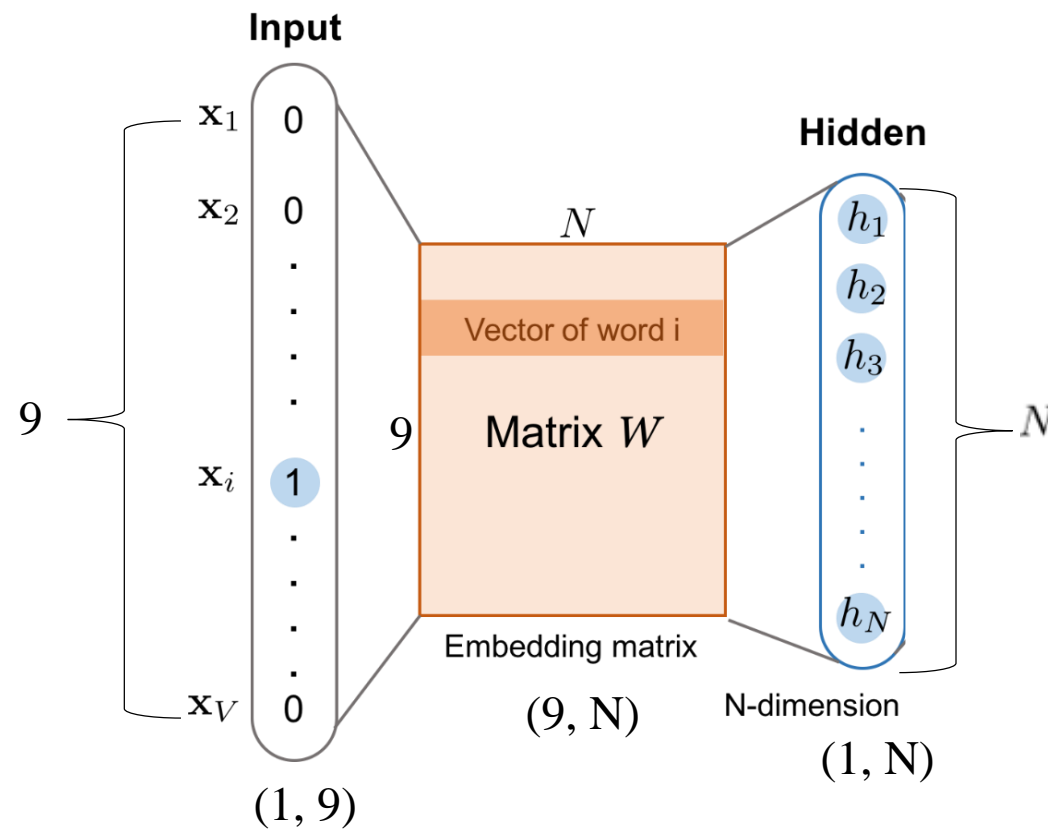Man, woman, boy, girl, prince, princess, queen, king, monarch

	1	2	3	4	5	6	7	8	9
man	1	0	0	0	0	0	0	0	0
woman	0	1	0	0	0	0	0	0	0
boy	0	0	1	0	0	0	0	0	0
girl	0	0	0	1	0	0	0	0	0
prince	0	0	0	0	1	0	0	0	0
princess	0	0	0	0	0	1	0	0	0
queen	0	0	0	0	0	0	1	0	0
king	0	0	0	0	0	0	0	1	0
monarch	0	0	0	0	0	0	0	0	1

Each word gets a 1x9 vector representation

Words to machine

Vector embedding

Embed one hot vector into the hidden size : 'princess'

Input

$\mathbf{x}_1$ 0

$\mathbf{x}_2$ 0

.

.

.

9

$\mathbf{x}_i$ 1

.

.

.

$\mathbf{x}_V$ 0

(1, 9)

N

9 Vector of word i

Matrix W

Embedding matrix

(9, N)

Hidden

h_1

h_2

h_3

.

.

.

.

.

.

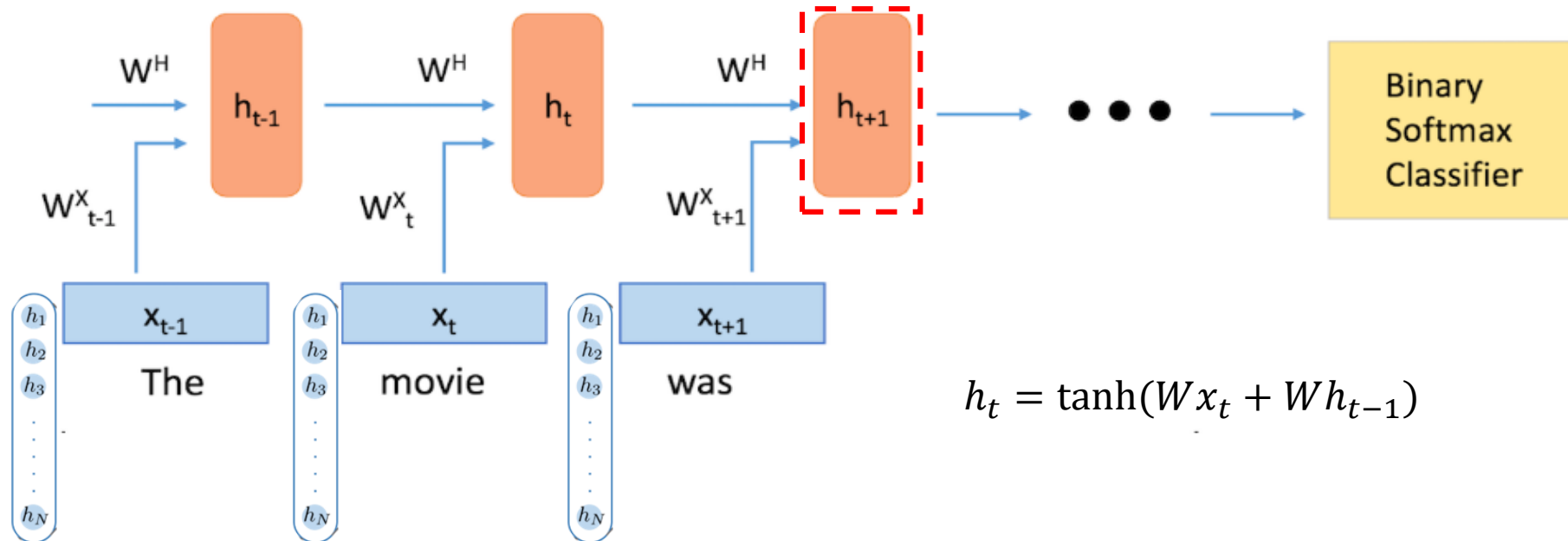h_N

N-dimension

(1, N)

N

Words to machine

Vector embedding

Embed one hot vector into the hidden size : 'princess'
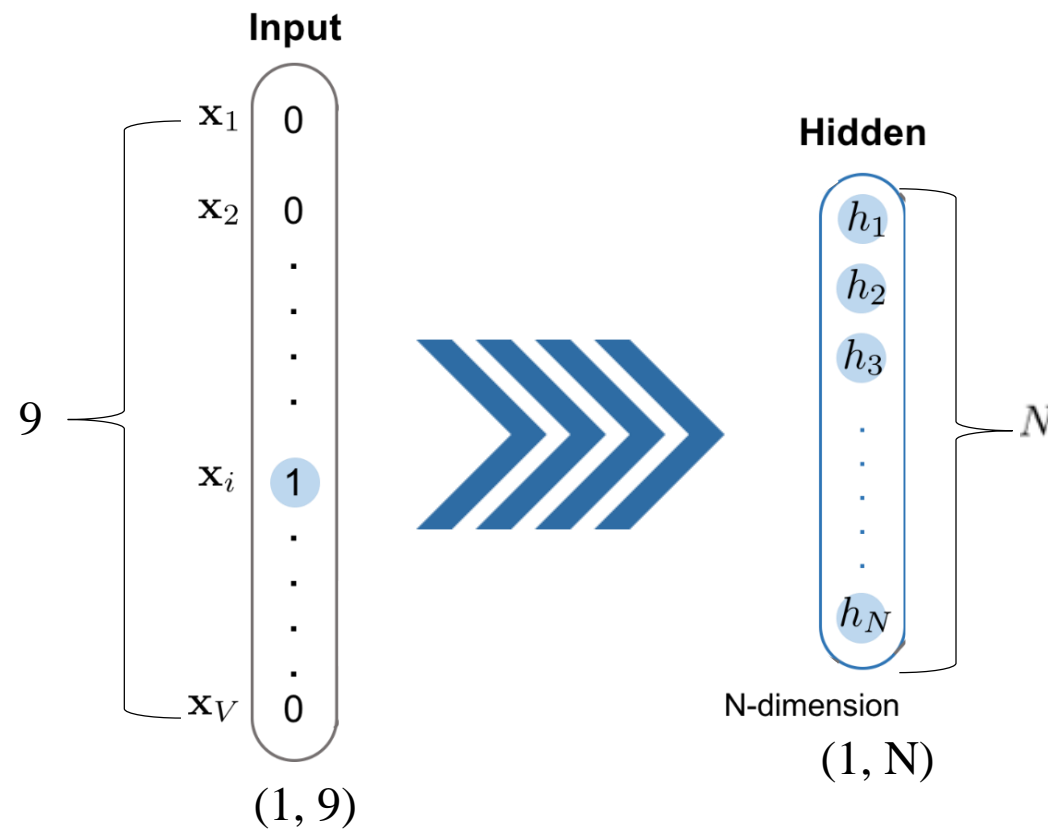
Vector embedding to RNN model

h_{t+1} remembers weights of h_{t-1} and h_t

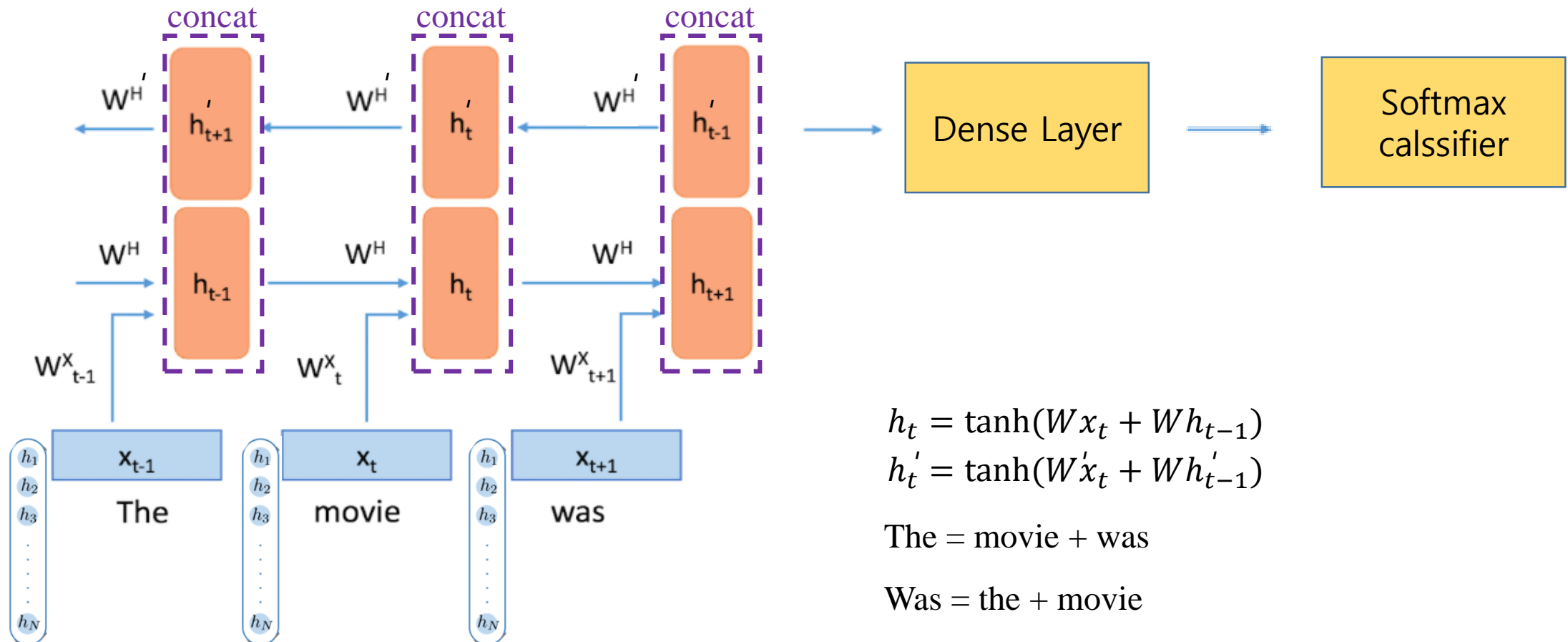$$h_t = \tanh(W x_t + W h_{t-1})$$

Vector embedding to Bi directional RNN model

h_{t+1} remembers weights of h_{t-1} and h_t

$$h_t = \tanh(Wx_t + Wh_{t-1})$$
$$h_t' = \tanh(Wx_t' + Wh_{t-1}')$$

The = movie + was
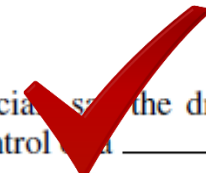
Was = the + movie

Intuitive Explanation

Task

Answering a cloze-style question

The answer to which depends on the **understanding of a context document provided with the question**

Document: What was supposed to be a fantasy sports car ride at Walt Disney World Speedway turned deadly when a Lamborghini crashed into a guardrail. The crash took place Sunday at the Exotic Driving Experience, which bills itself as a chance to drive your dream car on a racetrack. The Lamborghini's passenger, 36-year-old Gary Terry of Davenport, Florida, died at the scene, Florida Highway Patrol said. The driver of the Lamborghini, 24-year-old Tavon Watson of Kissimmee, Florida, lost control of the vehicle, the Highway Patrol said. (...)

Question: Officials said the driver, 24-year-old Tavon Watson, lost control of a _____

Intuitive Explanation

Task

Detect relationship between the question and the words in given document

1. Embed the words in document

2. Embed the whole question

3. Calculate the similarity between the two embedded vectors

Document: What was supposed to be a fantasy sports car ride at Walt Disney World Speedway turned deadly when a Lamborghini crashed into a guardrail. The crash took place Sunday at the Exotic Driving Experience, which bills itself as a chance to drive your dream car on a racetrack. The Lamborghini's passenger, 36-year-old Gary Terry of Davenport, Florida, died at the scene, Florida Highway Patrol said. The driver of the Lamborghini, 24-year-old Tavon Watson of Kissimmee, Florida, lost control of the vehicle, the Highway Patrol said. (...)

Question: Officials say the driver, 24-year-old Tavon Watson, lost control of a _____

DMQA Seminar

Task

Detect relationship between the question and the words in given document

Model guideline: set of possible answers from the document

Model guideline

Lamborghini   Walt Disney   World Speedway

Highway Patrol   Tavon Watson

Document: What was supposed to be a fantasy sports car ride at Walt Disney World Speedway turned deadly when a Lamborghini crashed into a guardrail. The crash took place Sunday at the Exotic Driving Experience, which bills itself as a chance to drive your dream car on a racetrack. The Lamborghini's passenger, 36-year-old Gary Terry of Davenport, Florida, died at the scene, Florida Highway Patrol said. The driver of the Lamborghini, 24-year-old Tavon Watson of Kissimmee, Florida, lost control of the vehicle, the Highway Patrol said. (...)

Question: Officials say the driver, 24-year-old Tavon Watson, lost control of a _____

Document

1 When she got home she shut herself up in her room and cried 2 There was nothing for her to do but resign , she thought dismally . 3 On the following **Saturday** Esther went for an afternoon walk , carrying her kodak with her . 4 It was a brilliantly fine autumn day , and woods and fields were basking in a mellow haze . 5 Esther went across lots to Mrs. Charley Cropper 's house , intending to make a call . 6 But the house was locked up and evidently deserted , so she rambled past it to the back fields . 7 Passing through a grove of maples she came out among leafy young saplings on the other side . 8 Just beyond her , with its laden boughs hanging over the line fence , was the famous plum tree . 9 Esther looked at it for a moment . 10 Then an odd smile gleamed over her face and she lifted her kodak . 11 Monday evening Esther called on Mr. Cropper again . 12 After the preliminary remarks in which he indulged , she said , with seeming irrelevance , that **Saturday** had been a fine day . 13 `` There was an excellent light for snapshots , " she went on coolly . 14 `` I went out with my kodak and was lucky enough to get a good negative . 15 I have brought you up a proof . 16 I thought you would be interested in it . '' 17 She rose and placed the proof on the table before Mr. Cropper . 18 The plum tree came out clearly . 19 Bob and Alf Cropper were up among the boughs picking the plums . 20 On the ground beneath them stood their father with a basket of fruit in his hand .

Question

Mr. Cropper looked at the XXXXX and from it to Esther .

Answer candidates

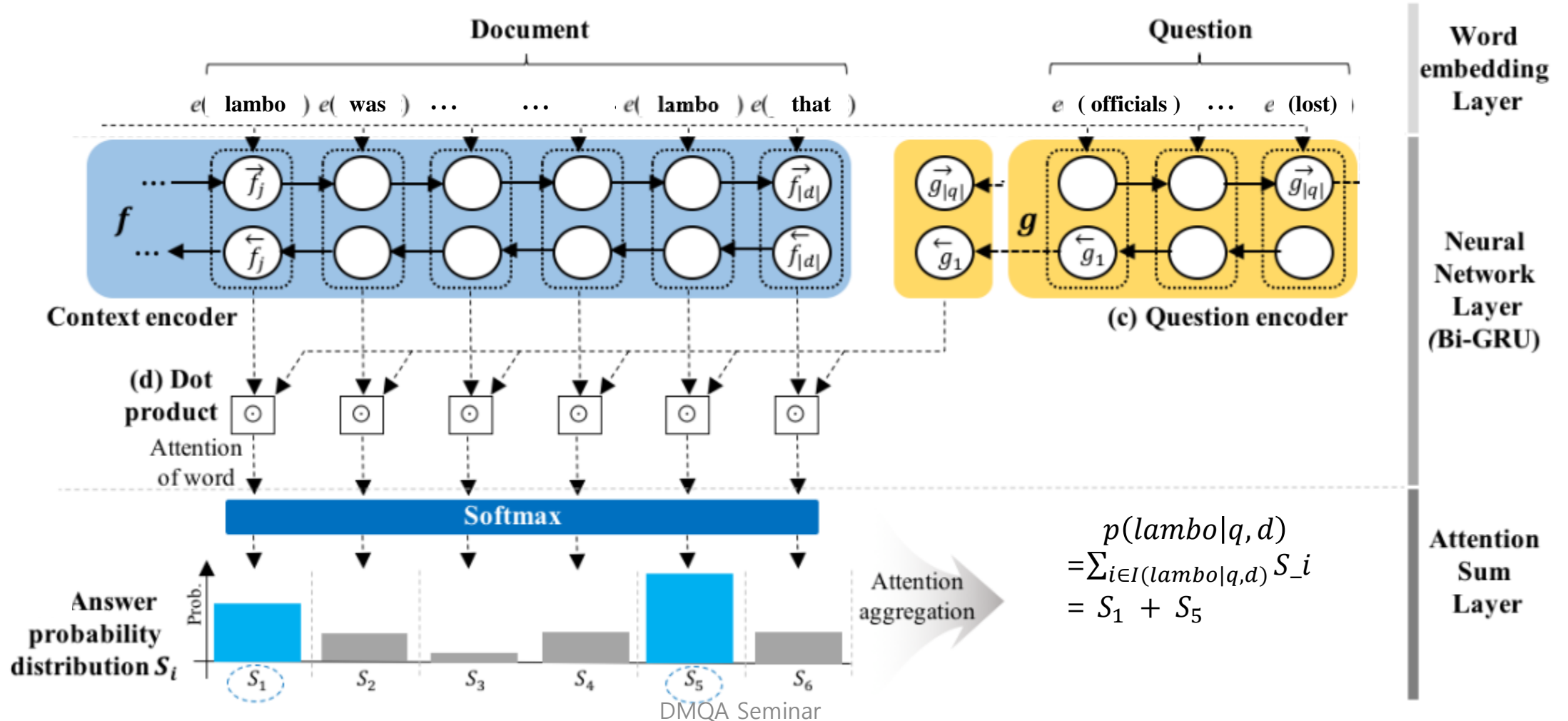**Saturday**, boughs, face, father, home, nothing, proof, remarks, smile, woods

Attention Sum Reader Architecture

Document
What was supposed to be a fantasy sports car ride ⋯

Question
Officials say the driver ⋯ lost control of a **xxxxx**
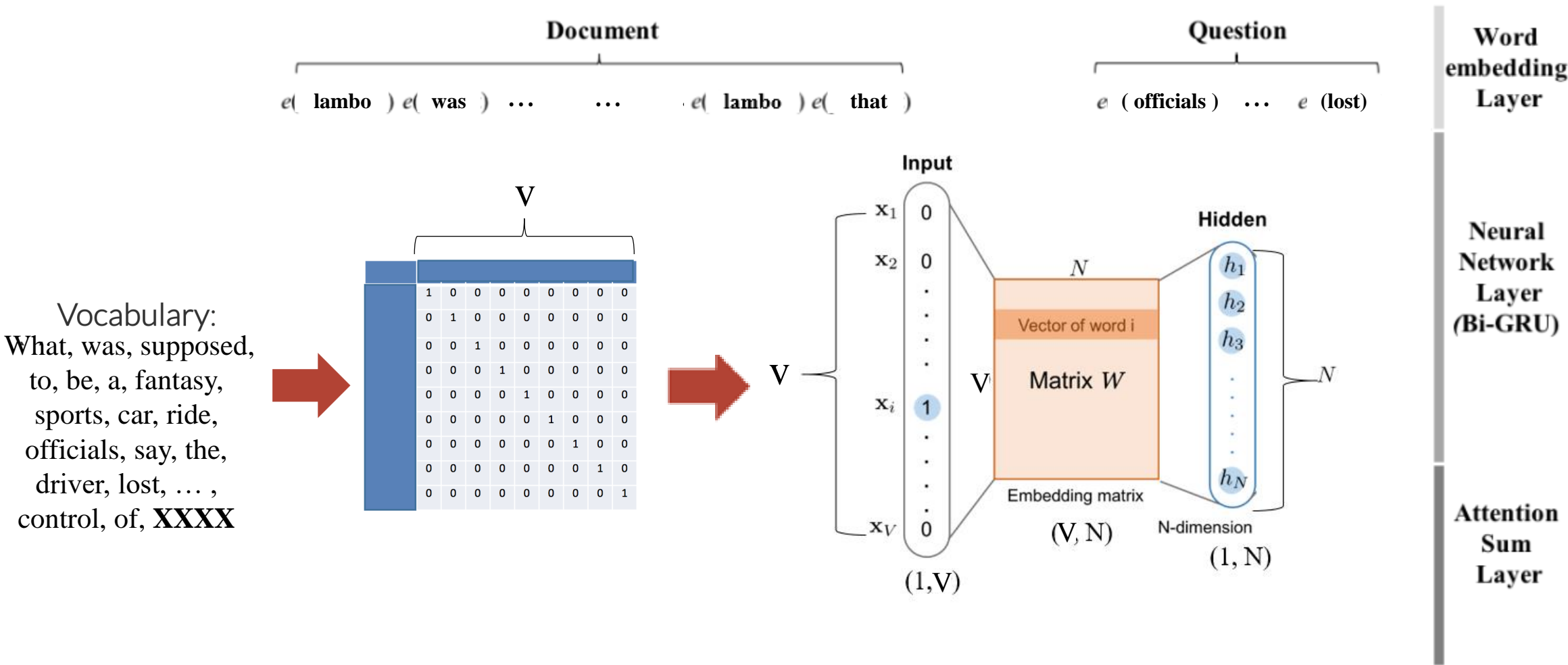
Attention Sum Reader Architecture

Document
What was supposed to be a fantasy sports car ride ⋯

Question
Officials say the driver ⋯ lost control of a **xxxxx**

Word embedding Layer

Document
$e($ lambo $) e($ was $)$ ⋯ ⋯ $e($ lambo $) e($ that $)$

Question
e (officials) ⋯ e (lost)

Neural Network Layer (Bi-GRU)

Vocabulary:
What, was, supposed, to, be, a, fantasy, sports, car, ride, officials, say, the, driver, lost, … , control, of, **XXXX**

V

1	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0
0	0	0	0	1	0	0	0	0
0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	1

Input

x_1 0
x_2 0
⋮
x_i 1
⋮
x_V 0
$(1,V)$

N
Vector of word i
Matrix W
Embedding matrix
(V, N)

Hidden
h_1
h_2
h_3
⋮
h_N
N-dimension
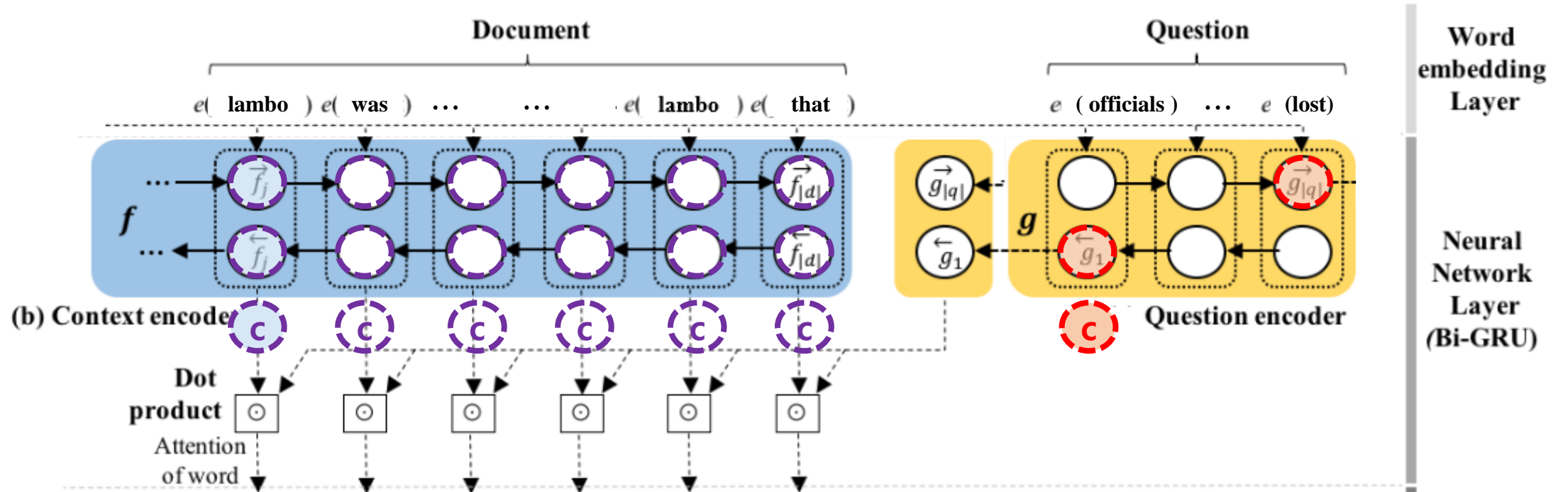$(1, N)$

Attention Sum Layer

Attention Sum Reader Architecture

Document
What was supposed to be a fantasy sports car ride ⋯

Question
Officials say the driver ⋯ lost control of a **xxxxx**
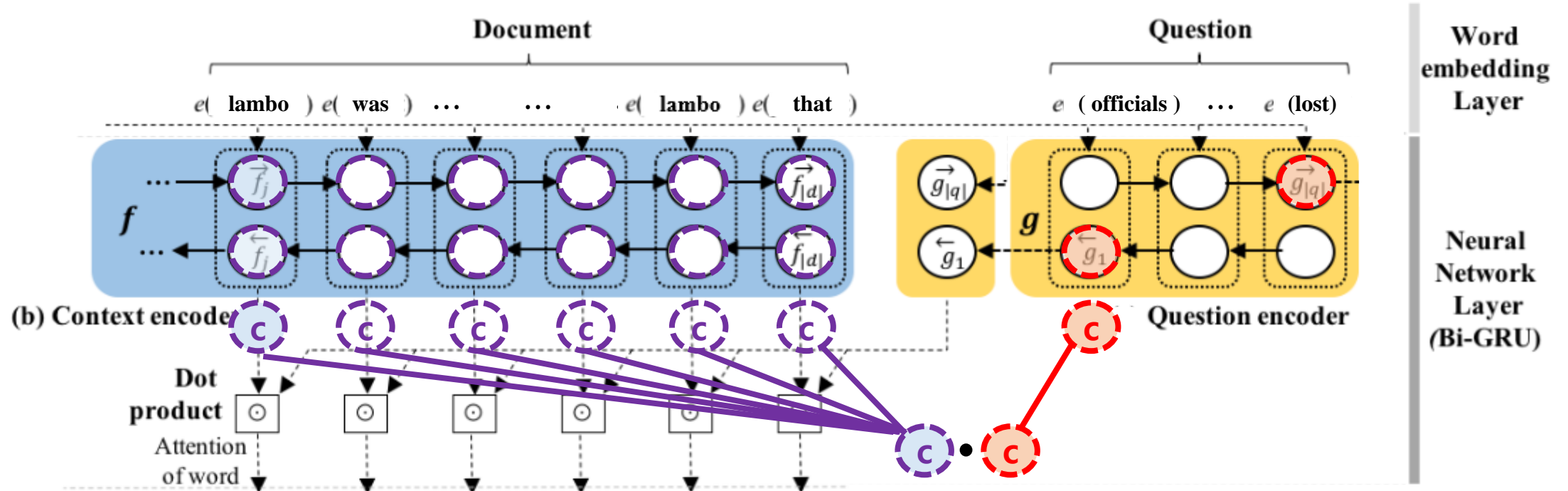
Attention Sum Reader Architecture

Document
What was supposed to be a fantasy sports car ride ⋯

Question
Officials say the driver ⋯ lost control of a **xxxxx**

Similarity calculation between the question embedding and word embeddings of documents

$$f_i(\mathbf{d}) = \overrightarrow{f_i}(\mathbf{d}) \ || \ \overleftarrow{f_i}(\mathbf{d})$$

$$g(\mathbf{q}) = \overrightarrow{g_{|\mathbf{q}|}}(\mathbf{q}) \ || \ \overleftarrow{g_1}(\mathbf{q})$$
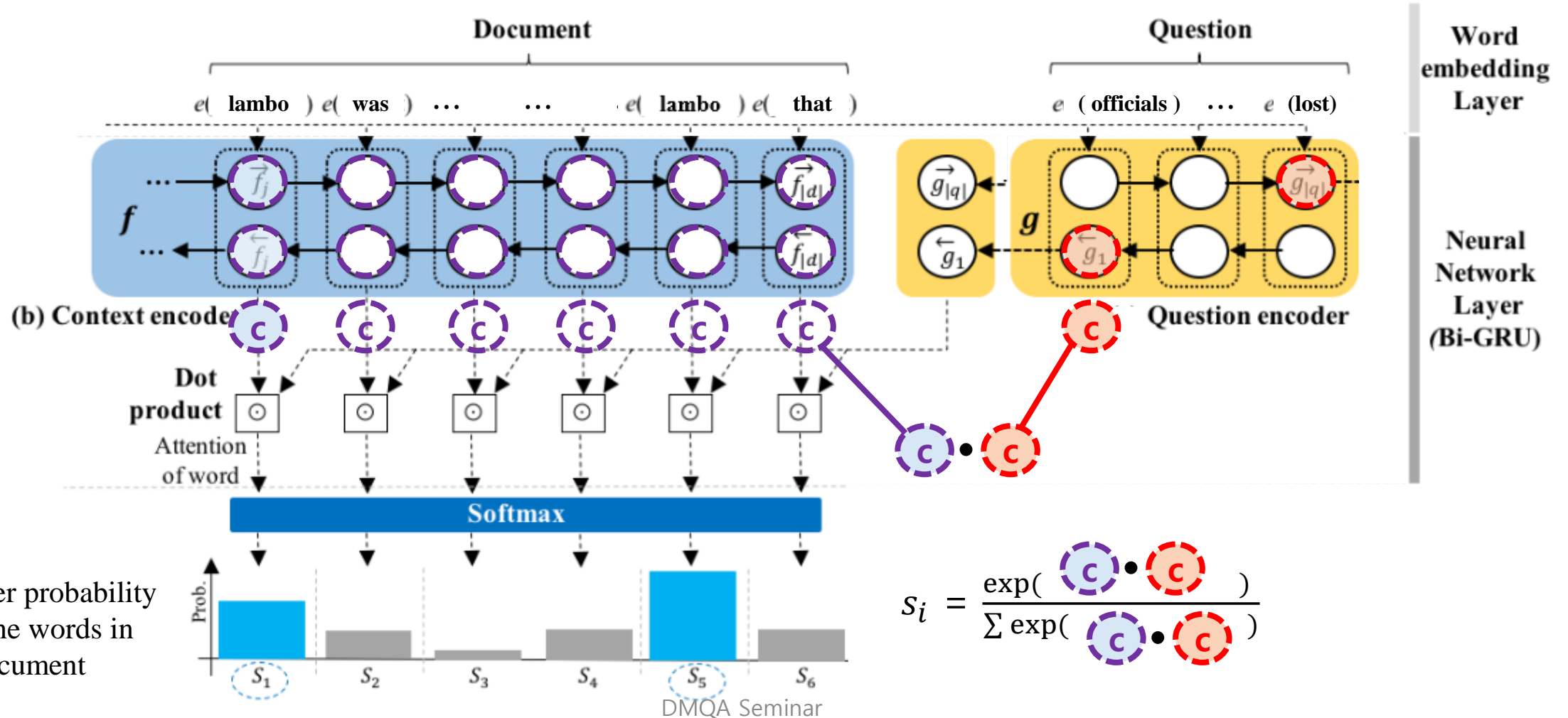
Attention Sum Reader Architecture

Document
What was supposed to be a fantasy sports car ride ⋯

Question
Officials say the driver ⋯ lost control of a **xxxxx**

Answer probability over the words in the document

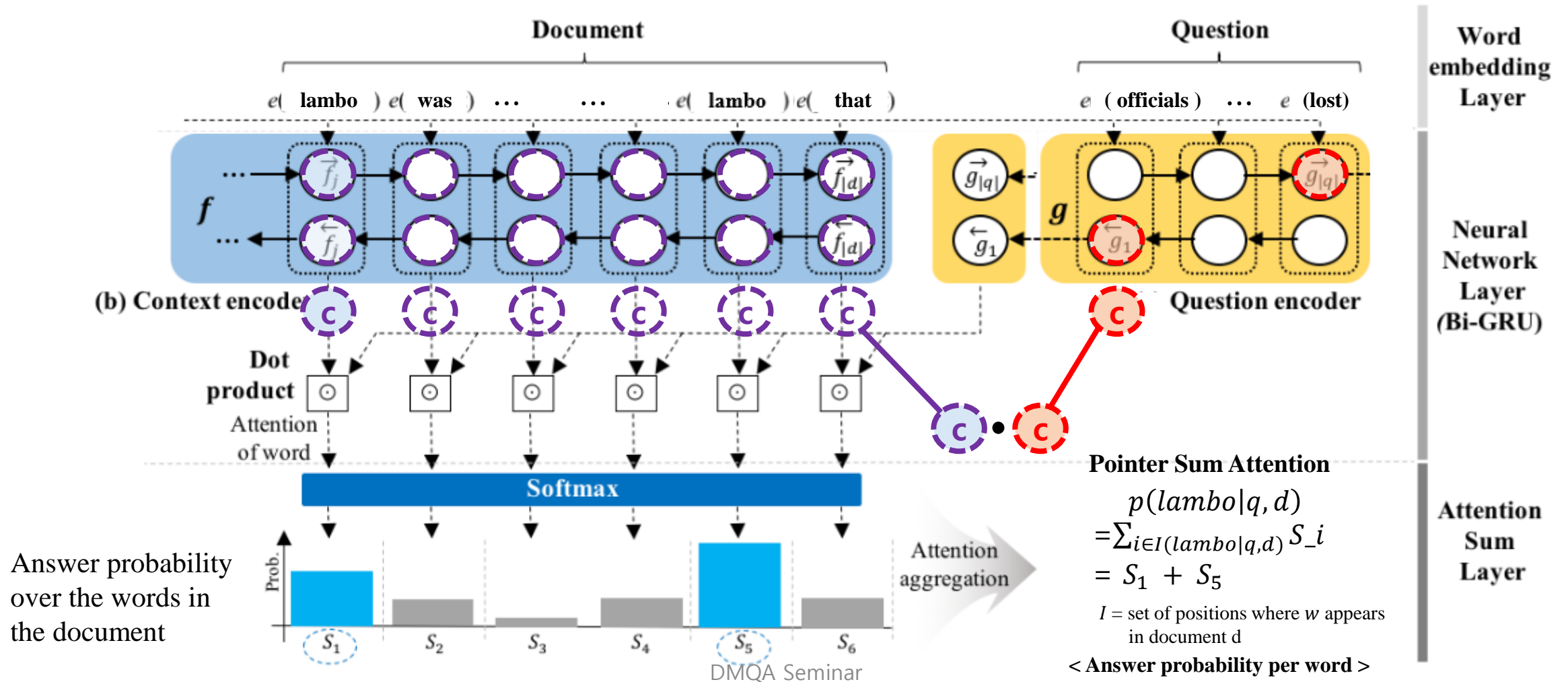$$s_i = \frac{\exp(\; c \bullet c \;)}{\sum \exp(\; c \bullet c \;)}$$

Attention Sum Reader Architecture

Document
What was supposed to be a fantasy sports car ride ⋯

Question
Officials say the driver ⋯ lost control of a **xxxxx**

Document

$e($ lambo $)$ $e($ was $)$ ⋯ ⋯ $e($ lambo $)$ $e($ that $)$

Question

e (officials) ⋯ e (lost)

f

$\overrightarrow{f_j}$ $\overrightarrow{f_{|d|}}$

$\overleftarrow{f_j}$ $\overleftarrow{f_{|d|}}$

(b) Context encoder

C C C C C C

$\overrightarrow{g_{|q|}}$ $\overrightarrow{g_{|q|}}$

g

$\overleftarrow{g_1}$ $\overleftarrow{g_1}$

C

Question encoder

Word embedding Layer

Neural Network Layer (Bi-GRU)

Dot product ⊙ ⊙ ⊙ ⊙ ⊙ ⊙

Attention of word

C • C

Softmax

Answer probability over the words in the document

Prob.

S_1 S_2 S_3 S_4 S_5 S_6

Attention aggregation

Pointer Sum Attention

$$p(lambo|q,d)$$
$$=\sum_{i \in I(lambo|q,d)} S\_i$$
$$= S_1 + S_5$$

I = set of positions where w appears in document d

< Answer probability per word >

Attention Sum Layer

DMQA Seminar

Attention Sum Reader Training Details

1. Answer Candidate and pointer sum attention (word probability among document)

Lambo, Walt Disney, World Speedway, Highway Patrol, Tavon Watson, nothing, proof, remarks, smile, woods

answer   model output

[1, 0, 0, 0, 0, … , 0, 0 ,0 ,0]   [0.6, 0.05, 0.1, 0.0, … , 0.1, 0.0, 0.1 ,0.0 ,0.0]

# of answer candidates (10)   # of answer candidates (10)

2. Minimize negative log likelihood

$$- log P_\theta (a | \mathbf{q}, \mathbf{d}),$$

where a is the correct answer probability for the question q and document d,

P_θ is the word probability of the word from answer candidate

Attention Sum Reader Evaluation

1. Single model

2. Ensemble model : use simple averaging of the answer probabilities predicted by ensemble members

Average Ensemble (average weight)

Choose top 20% of all trained models from validation

Greedy Ensemble (average weight through greedy learning

Add the 70% of best performing model that had not been previously tried

Attention Sum Reader (Ensemble)

Attention Sum Reader   Attention Sum Reader   Attention Sum Reader   Attention Sum Reader
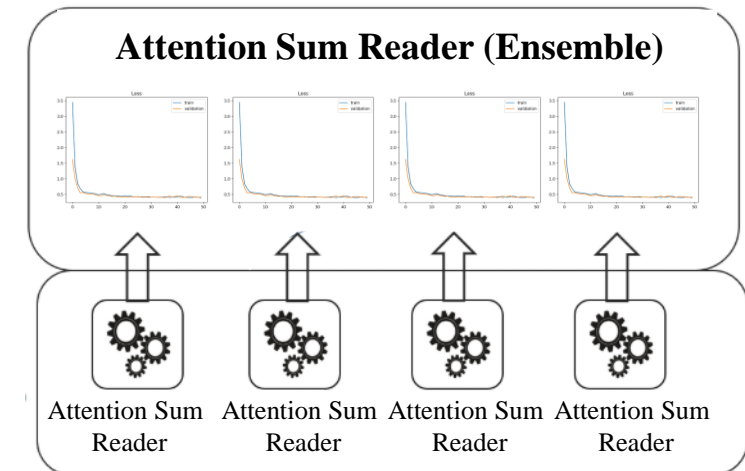
Attention Sum Reader Datasets

CBT Data set

Story understanding dataset based on 118 children books from project Gutenberg

1 `` phebe beckoned to him ; i saw her , '' cried rose , staring hard at the door .
2 `` is it more presents coming ? ''
3 asked jamie , just as his brother re-appeared , looking more excited than ever .
4 `` yes ; a present for mother , and here it is ! ''
5 roared archie , flinging wide the door to let in a tall man , who cried out , '' where 's my little woman ?
6 the first kiss for her , then the rest may come on as fast as they like . ''
7 before the words were out of his mouth , mrs. jessie was half-hidden under his rough great-coat , and four boys were prancing about him clamouring for their turn .
8 of course , there was a joyful tumult for a time , during which rose slipped into the window recess and watched what went on , as if it were a chapter in a christmas story .
9 it was good to see bluff uncle jem look proudly at his tall son , and fondly hug the little ones .
10 it was better still to see him shake his brothers ' hands as if he would never leave off , and kiss all the sisters in a way that made even solemn aunt myra brighten up for a minute .

11 but it was best of all to see him finally established in grandfather 's chair , with his '' little woman '' beside him , his three youngest boys in his lap , and _____ hovering over him like a large-sized cherub .

faith | brothers | rose | **archie** | rest | mouth | way | mother | sisters | george

CNN data set

Question	Passage
it 's clear @entity0 is leaning toward **@placeholder** , says an expert who monitors @entity0	... @entity116 , who follows @entity0 's operations and propaganda closely , recently told @entity3 , *it 's clear @entity0 is leaning toward* **@entity60** in terms of doctrine , ideology and an emphasis on holding territory after operations . ... .
@placeholder says he understands why @entity0 wo n't play at his tournament	... @entity0 called me personally to let me know that he would n't be playing here at @entity23 , '' **@entity3** said on his @entity21 event 's website . ... .
a tv movie based on @entity2 's book **@placeholder** casts a @entity76 actor as @entity5	... to @entity12 @entity2 professed that his **@entity11** is not a religious book . ... .
he 's doing a his - and - her duet all by himself , @entity6 said of **@placeholder**	... we got some groundbreaking performances , here too , tonight , @entity6 said . we got **@entity17** , who will be doing some musical performances . he 's doing a his - and - her duet all by himself . ... .
rapper **@placeholder** '' disgusted , '' cancels upcoming show for @entity280	... with hip - hop star **@entity246** saying on @entity247 that he was canceling an upcoming show for the @entity249 . ... (but @entity249 = @entity280 = SAEs)
pilot error and snow were reasons stated for **@placeholder** plane crash	... a small aircraft carrying **@entity5** , @entity6 and @entity7 the @entity12 @entity3 crashed a few miles from @entity9 , near @entity10 , @entity11 . ... .

Daily mail data set

Passage

(@entity4) if you feel a ripple in the force today , it may be the news that the official @entity6 is getting its first gay character . according to the sci-fi website @entity9 , the upcoming novel " @entity11 " will feature a capable but flawed @entity13 official named @entity14 who " also happens to be a lesbian . " the character is the first gay figure in the official @entity6 -- the movies , television shows , comics and books approved by @entity6 franchise owner @entity22 -- according to @entity24 , editor of " @entity6 " books at @entity28 imprint @entity26 .

Question

characters in " @placeholder " movies have gradually become more diverse

Answer

@entity6

Accuracy Measure

	CNN		Daily Mail	
	valid	test	valid	test
Attentive Reader [†]	61.6	63.0	70.5	69.0
Impatient Reader [†]	61.8	63.8	69.0	68.0
MemNNs (single model) [‡]	63.4	66.8	NA	NA
MemNNs (ensemble) [‡]	66.2	69.4	NA	NA
Dynamic Entity Repres. (max-pool) [♯]	71.2	70.7	NA	NA
Dynamic Entity Repres. (max-pool + byway)[♯]	70.8	72.0	NA	NA
Dynamic Entity Repres. + w2v [♯]	71.3	72.9	NA	NA
Chen et al. (2016) (single model)	72.4	72.4	76.9	75.8
AS Reader (single model)	68.6	69.5	75.0	73.9
AS Reader (avg for top 20%)	68.4	69.9	74.5	73.5
AS Reader (avg ensemble)	73.9	**75.4**	78.1	77.1
AS Reader (greedy ensemble)	74.5	74.8	78.7	**77.7**

	Named entity		Common noun	
	valid	test	valid	test
Humans (query) [(*)]	NA	52.0	NA	64.4
Humans (context+query) [(*)]	NA	*81.6*	NA	*81.6*
LSTMs (context+query) [‡]	51.2	41.8	62.6	56.0
MemNNs (window memory + self-sup.) [‡]	70.4	66.6	64.2	63.0
AS Reader (single model)	73.8	68.6	68.8	63.4
AS Reader (avg for top 20%)	73.3	68.4	67.7	63.2
AS Reader (avg ensemble)	74.5	70.6	71.1	**68.9**
AS Reader (greedy ensemble)	76.2	**71.0**	72.4	67.5

Attention Sum Reader Evaluation

Good case

what was supposed to be a fantasy sports car ride at @entity3 turned deadly when a @entity4 crashed into a guardrail . the crash took place sunday at the @entity8 , which bills itself as a chance to drive your dream car on a racetrack . the @entity4 's passenger , 36 - year - old @entity14 of @entity15 , @entity16 , died at the scene , @entity13 said . the driver of the @entity4 , 24 - year - old @entity18 of @entity19 , @entity16 , lost control of the vehicle , the @entity13 said .

...

0 ▭▬▬▬ 1

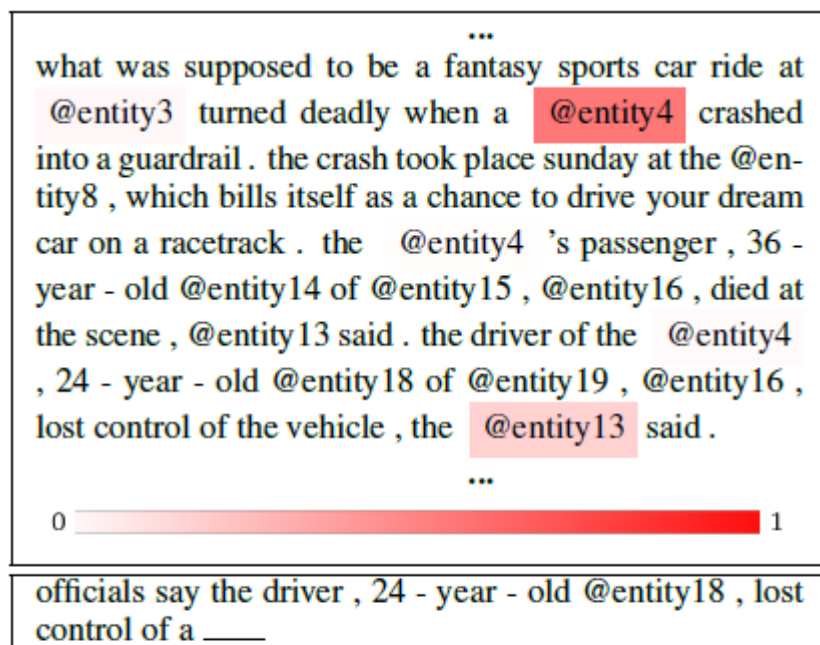officials say the driver , 24 - year - old @entity18 , lost control of a ____

Figure 3: Attention in an example with anonymized entities where our system selected the correct answer. Note that the attention is focused only on named entities.

Fine case

@entity11 film critic @entity29 writes in his review that "anyone nostalgic for childhood dreams of trans- formation will find something to enjoy in an uplifting movie that invests warm sentiment in universal themes of loss and resilience , experience and maturity . " more : the best and worst adaptations of "@entity" @entity43, @entity44 and @entity46 star in director @entity48's crime film about a hit man trying to save his estranged son from a revenge plot. @entity11 chief film critic @entity52 writes in his review that the film

...

0 ▭▬▬▬ 1

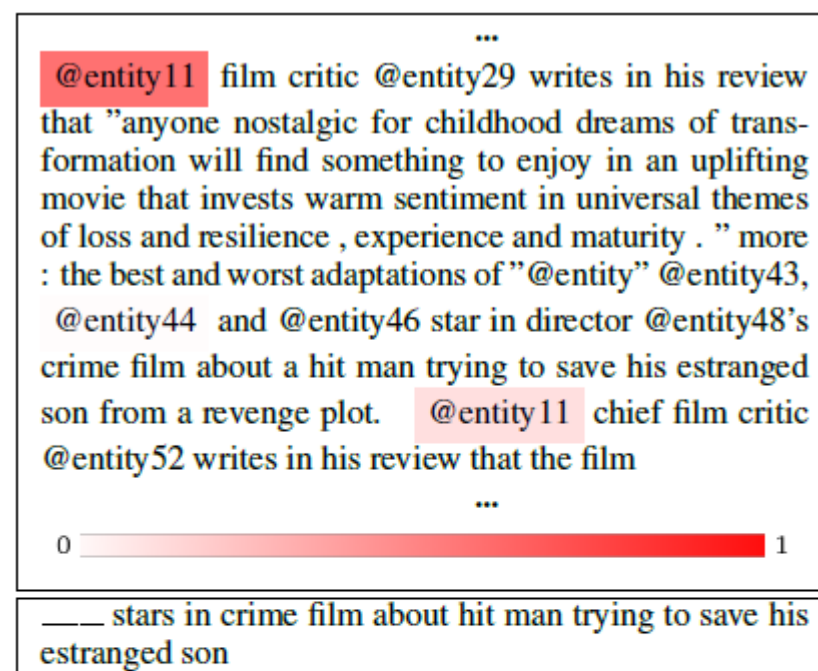____ stars in crime film about hit man trying to save his estranged son

Figure 4: Attention over an example where our system failed to select the correct answer (en- tity43). The system was probably mislead by the co-occurring word 'film'. Namely, entity11 occurs 7 times in the whole document and 6 times it is to- gether with the word 'film'. On the other hand, the correct answer occurs only 3 times in total and only once together with 'film'.

DMQA Seminar

Conclusion

1. Natural Language text comprehension task

   Answer to cloze-style question from a document

2. Simple but explicit neural model

3. An analysis by Chen et al., 2016 suggests that CNN and Daily Mail dataset are difficult to understand for humans and the model may achieve the maximal accuracy

Thank you

Appendix

	CNN			Daily Mail			CBT CN			CBT NE		
	train	valid	test	train	valid	test	train	valid	test	train	valid	test
# queries	380,298	3,924	3,198	879,450	64,835	53,182	120,769	2,000	2,500	108,719	2,000	2,500
Max # options	527	187	396	371	232	245	10	10	10	10	10	10
Avg # options	26.4	26.5	24.5	26.5	25.5	26.0	10	10	10	10	10	10
Avg # tokens	762	763	716	813	774	780	470	448	461	433	412	424
Vocab. size	118,497			208,045			53,185			53,063		

CBT CN : CBT Common Nouns
CBT NE : CBT Named Entities

Appendix

Vocab dimension 52,263

Embedding dimension 200

Hidden dimension 128

RNN type GRU

Gradient clipping 10

Embedding matrix initialization: uniformly random