LSTM-based Transaction Bot

Vineet Kumar, PhD

Implementation: https://github.com/vineetk1/fairseq/tree/dialog

Contact: Email me through Meetup (private message, not to the group)

Agenda

- Problem
- A Solution (high-level view)
- Open Source Implementation
- Word Embeddings
- Encoder
- Decoder
- Attention
- Beam Search
- Results   (Dataset: One dialog that passed; One dialog that failed)
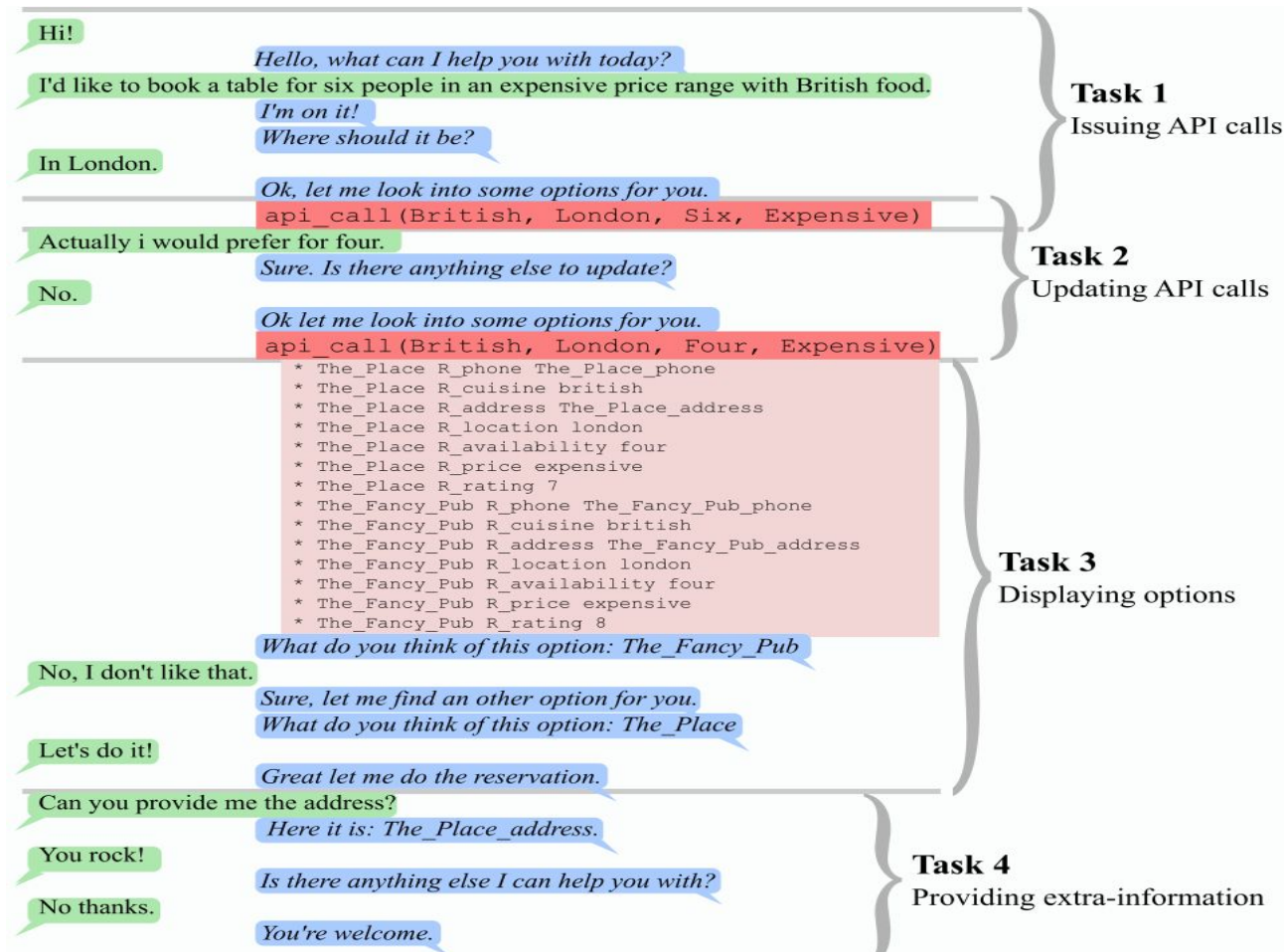- Relevant Publications

Problem

- Build Bot using AI
  - Makes reservation for table at restaurant
- Dialog State Tracking Challenge 2 (DSTC 2)
  - 1000 dialogs for training, 1000 for validation and 1000 for testing
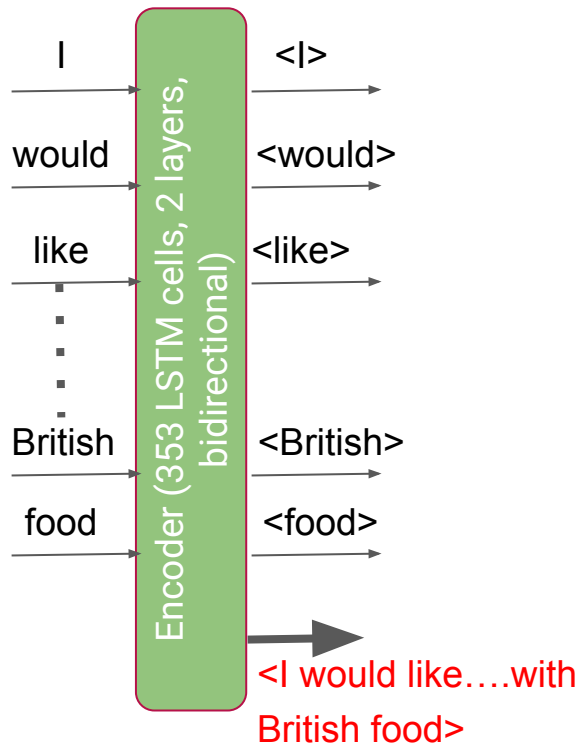 - Dialog: Client and Customer Service Agent

Restaurant Reservation System
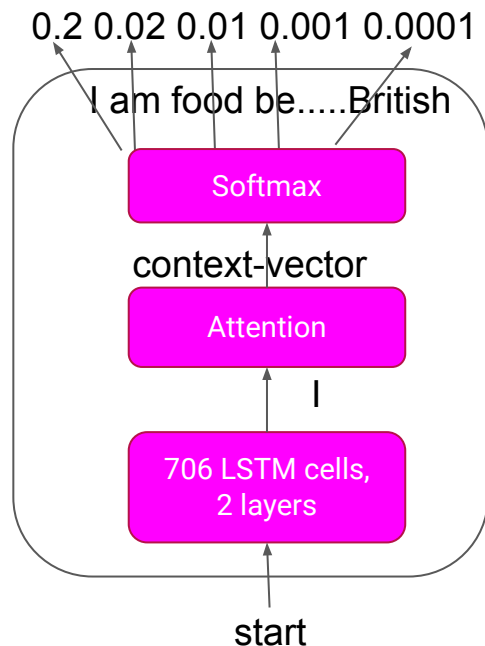
Client   **Customer Service Agent**

Hi!

Hello, what can I help you with today?

I'd like to book a table for six people in an expensive price range with British food.

I'm on it!

Where should it be?

In London.

Ok, let me look into some options for you.

```
api_call(British, London, Six, Expensive)
```

Task 1
Issuing API calls

Actually i would prefer for four.

Sure. Is there anything else to update?

No.

Ok let me look into some options for you.

```
api_call(British, London, Four, Expensive)
```

Task 2
Updating API calls

```
* The_Place R_phone The_Place_phone
* The_Place R_cuisine british
* The_Place R_address The_Place_address
* The_Place R_location london
* The_Place R_availability four
* The_Place R_price expensive
* The_Place R_rating 7
* The_Fancy_Pub R_phone The_Fancy_Pub_phone
* The_Fancy_Pub R_cuisine british
* The_Fancy_Pub R_address The_Fancy_Pub_address
* The_Fancy_Pub R_location london
* The_Fancy_Pub R_availability four
* The_Fancy_Pub R_price expensive
* The_Fancy_Pub R_rating 8
```

What do you think of this option: The_Fancy_Pub

No, I don't like that.

Sure, let me find an other option for you.

What do you think of this option: The_Place

Let's do it!

Great let me do the reservation.

Task 3
Displaying options

Can you provide me the address?

Here it is: The_Place_address.

You rock!

Is there anything else I can help you with?

No thanks.

You're welcome.

Task 4
Providing extra-information

4

A Solution (high-level view)

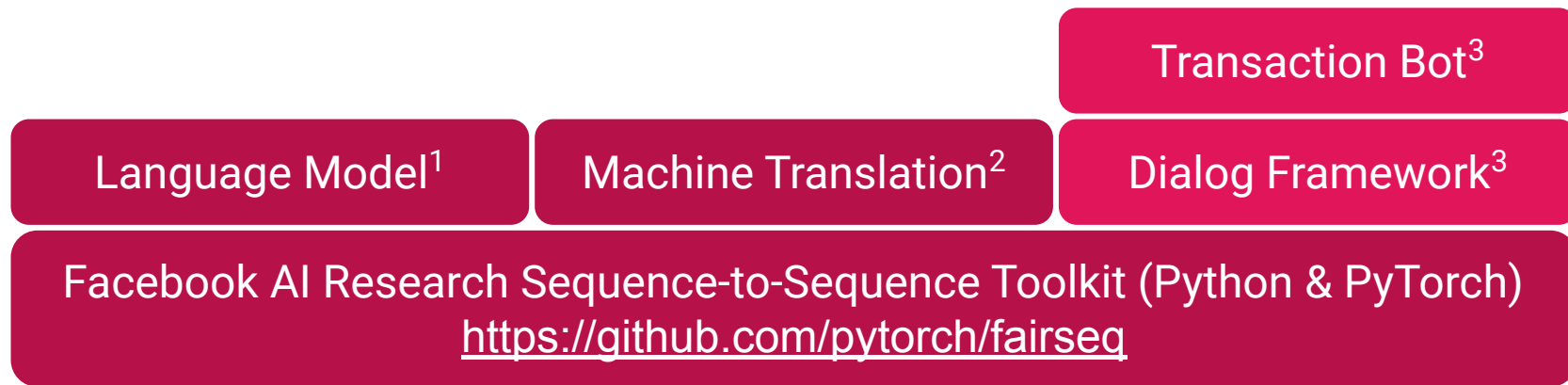**Encoder**

Decoder with Attention

1. Forward propagation
2. Calculate Loss (e.g. using "label smoothed cross entropy" function)
3. Backward propagation for gradients
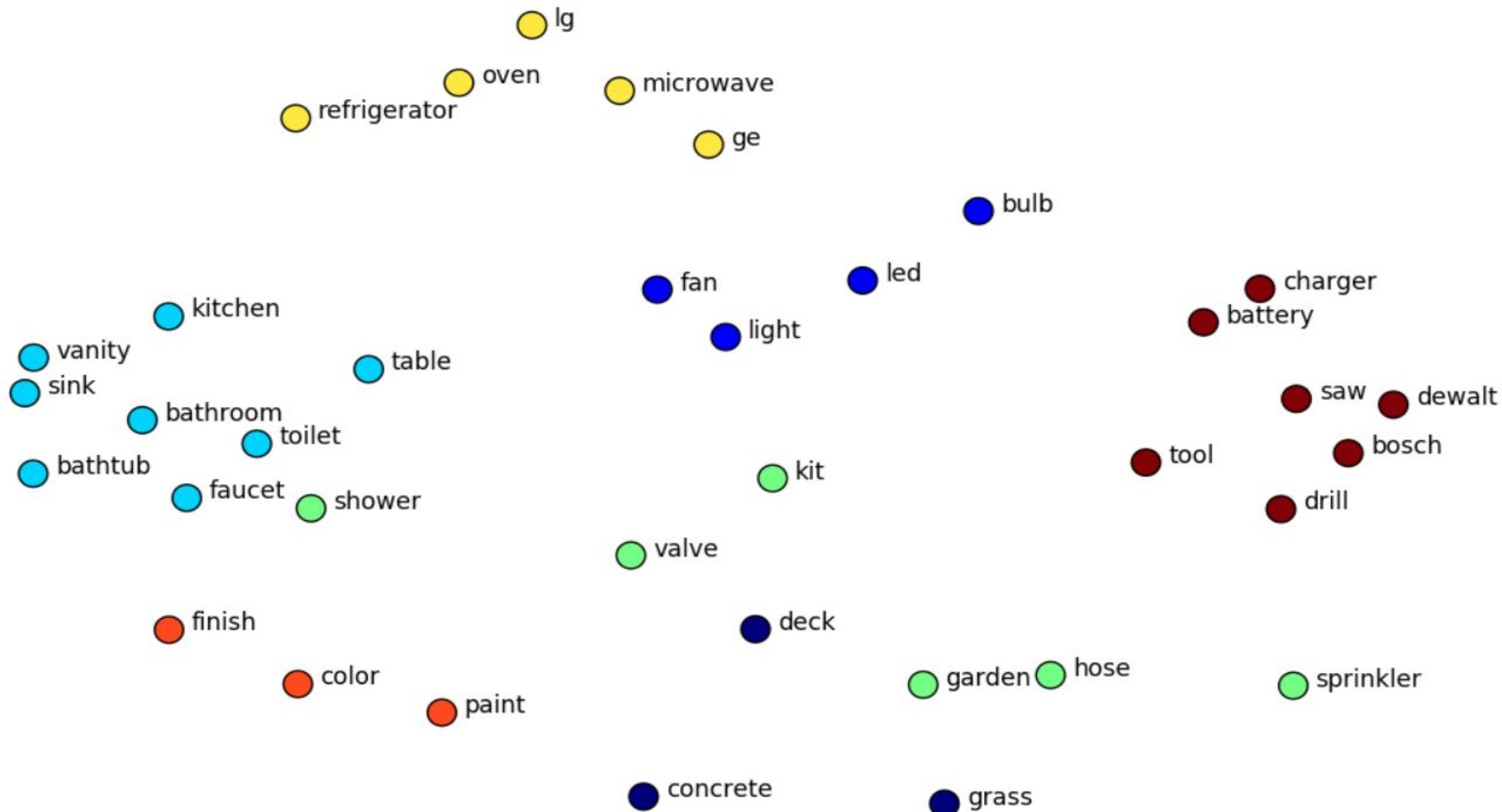4. Update parameters using Optimizer (e.g. nag)
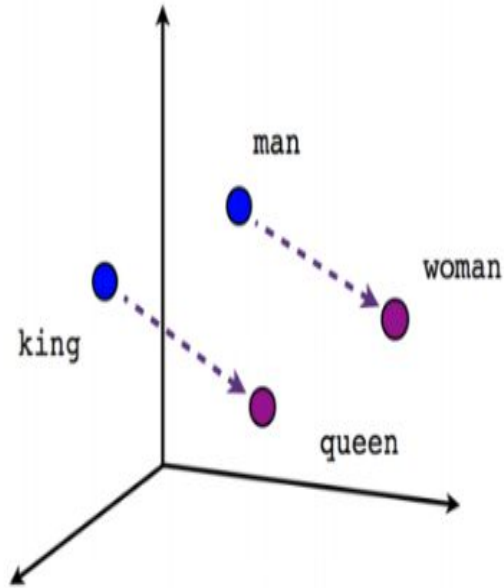
Open Source Implementation

Transaction Bot[3]

Language Model[1]　　　Machine Translation[2]　　　Dialog Framework[3]

Facebook AI Research Sequence-to-Sequence Toolkit (Python & PyTorch)
https://github.com/pytorch/fairseq

Implementation of Transaction Bot: https://github.com/vineetk1/fairseq/tree/dialog

1. Predict the next word in a sequence of words
2. Given a source sequence (e.g. English), predict a target sequence (e.g. French)
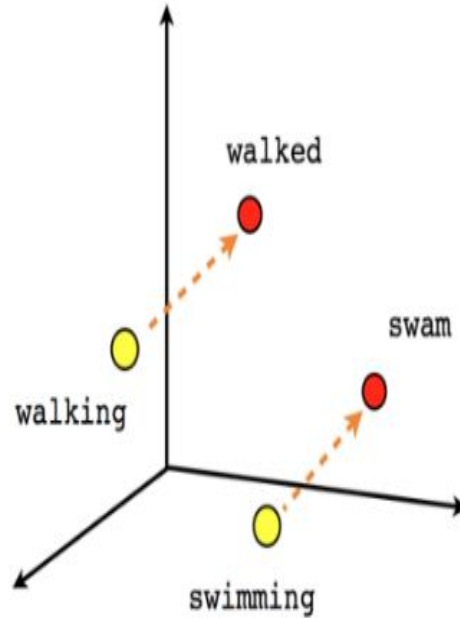3. Given a source sequence plus all previous sequences, predict a target sequence
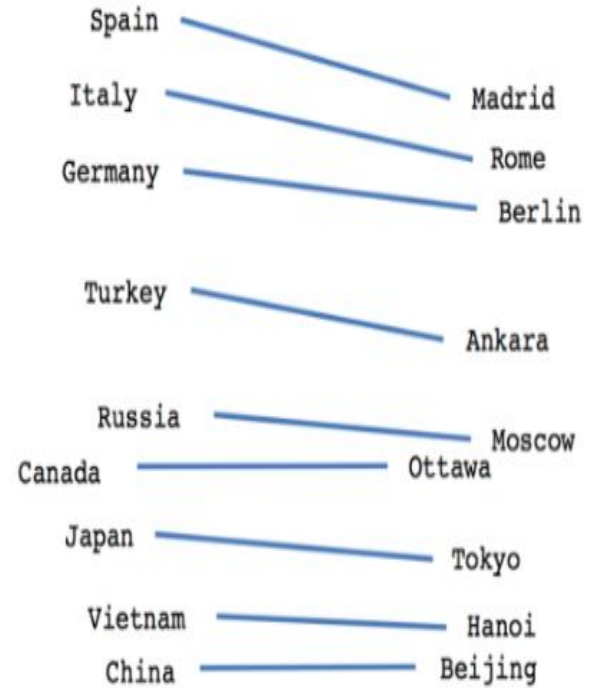
Word Embeddings using Glove; word -> vector

lg
oven
microwave
refrigerator
ge
bulb
led
charger
fan
battery
kitchen
light
vanity
table
sink
saw
dewalt
bathroom
bosch
toilet
tool
bathtub
kit
drill
faucet
shower
valve
finish
deck
color
garden
hose
sprinkler
paint
concrete
grass

7

Some features of vectors

Male-Female
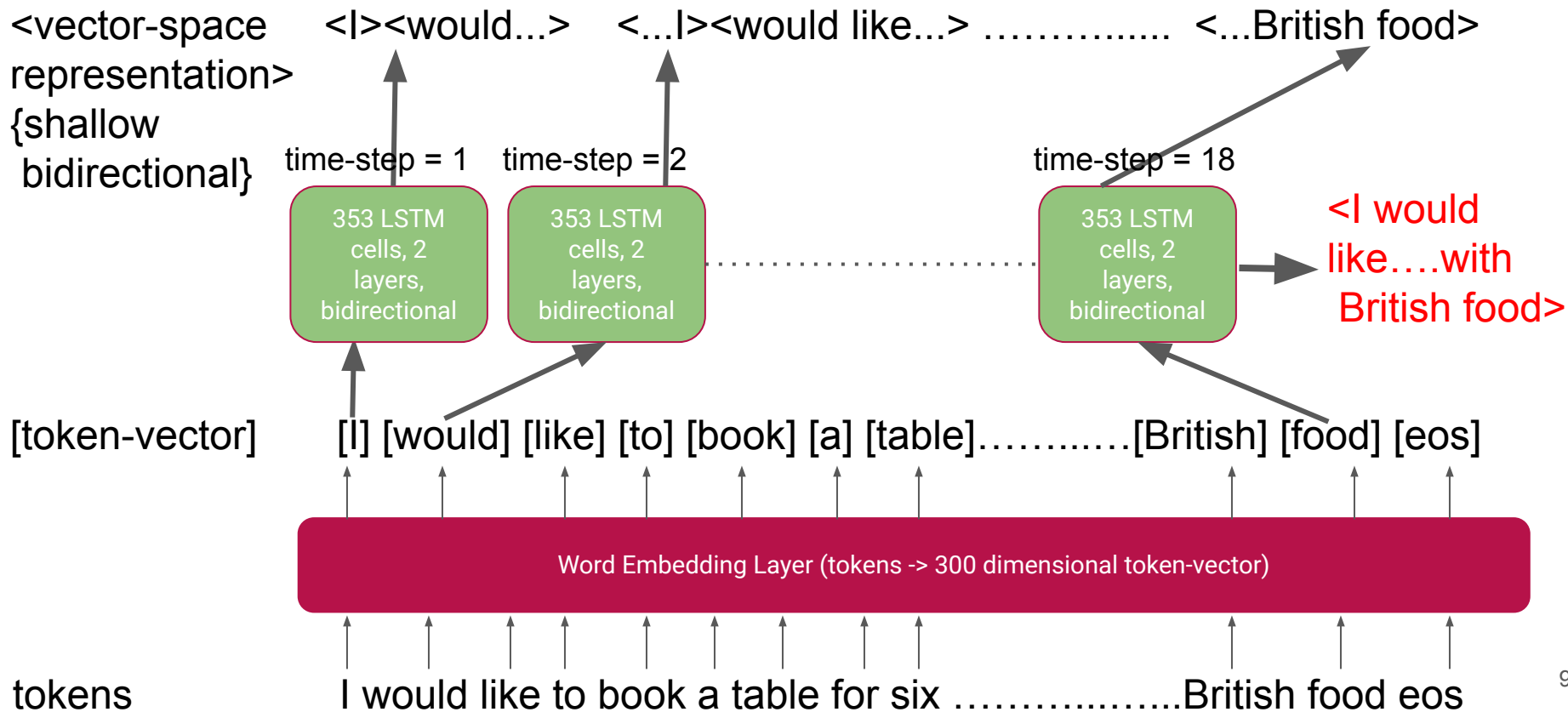
Verb tense
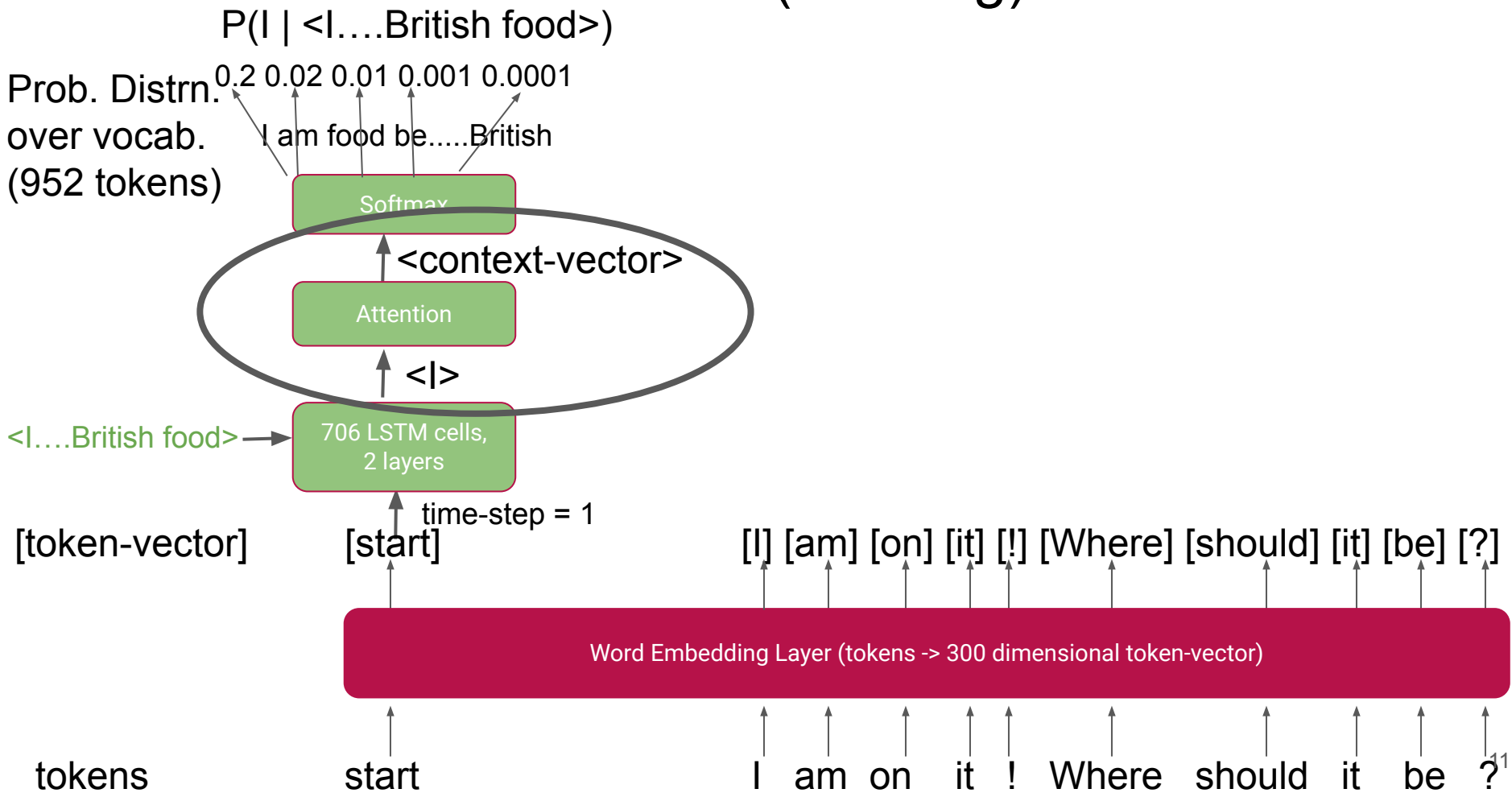
Country-Capital

Encoder (Training)

Dialog State
<vector-space
representation>
{shallow
  bidirectional}

<I><would...>   <...I><would like...> ….…........ <...British food>

time-step = 1   time-step = 2   time-step = 18

| 353 LSTM cells, 2 layers, bidirectional | 353 LSTM cells, 2 layers, bidirectional | .... | 353 LSTM cells, 2 layers, bidirectional |

<I would like….with British food>

[token-vector]   [I] [would] [like] [to] [book] [a] [table]….….....[British] [food] [eos]

Word Embedding Layer (tokens -> 300 dimensional token-vector)

tokens   I would like to book a table for six ….……........British food eos

9

Decoder (Training)

P(I | <I....British food>) P(am | I, <I....British food>) P(be | should,...,am,I,<I....British food>)

0.2 0.02 0.01 0.001 0.0001   0.02 0.2 0.01 0.001 0.0001   0.002 0.001 0.01 0.3 0.0001

Prob. Distrn. over vocab. (952 tokens)

I am food be.....British   I am food be.....British   I am food be.....British

| Softmax | Softmax | Softmax |

<vector-space> <I>   <am>   <be>

<I....British food> →

| 706 LSTM cells, 2 layers | 706 LSTM cells, 2 layers | .... | 706 LSTM cells, 2 layers |

time-step = 1   time-step = 2   time-step = 9

[token-vector] [start]   [I] [am] [on] [it] [!] [Where] [should] [it] [be] [?]

| Word Embedding Layer (tokens -> 300 dimensional token-vector) |

tokens   start   I am on it ! Where should it be ?

Attention (Training)

P(I | <I….British food>)

Prob. Distrn. over vocab. (952 tokens)

0.2 0.02 0.01 0.001 0.0001

I am food be.....British

Softmax

<context-vector>

Attention

<I>

<I….British food> → 706 LSTM cells, 2 layers

time-step = 1

[token-vector]   [start]

[I] [am] [on] [it] [!] [Where] [should] [it] [be] [?]

Word Embedding Layer (tokens -> 300 dimensional token-vector)

tokens   start   I am on it ! Where should it be ?

Luong's Attention (Training)

Context
Vector   tanh(W(<attention vector><I>))

Attn.   ……… 0.5(<I would>) + 0.2(<I would like..>) +...+ 0.04(<..British food>)
Vector

Normalized   0.5   0.2   …….…....   0.04
Attn. Score

Attn. Score   500   212   …….…....   30

Dot Product   ….. (<I would>.<I>) (<I would like..>.<I>) …….….... (<..British food>.<I>)

Dialog State …. <I would> <I would like..> …… <..with British food..> <..British food>

Vinyal's Attention (Training)

Context Vector	tanh(W(<attention vector><l>))		
Attn. Vector	0.5(<l would>.<l>) + 0.2(<l would like..>.<l>) +...+ 0.04(<..British food>.<l>)		
Normalized Attn. Score	0.5	0.2 …............	0.04

Vinyal's Attn. Score v^T(tanh(W(<l would><l>))) where vector v is trainable parameter

~~Luong's Attn. Score~~	~~500~~	~~212~~ ~~…...........~~	~~30~~
~~Dot Product~~	~~(<l would>.<l>)~~	~~(<l would like..>.<l>)~~ ~~…...........~~	~~(<..British food>.<l>)~~

Dialog State …..<l would> <l would like..> …… <..with British food..> <..British food>[13]

Attention (Training)

P(I | <I….British food>)

Prob. Distrn.
over vocab.
(952 tokens)

0.2 0.02 0.01 0.001 0.0001

I am food be.....British

Softmax

<context-vector>

Attention

<I>

<I….British food> → 706 LSTM cells, 2 layers

time-step = 1

[token-vector]   [start]

(<context-vector>[I]) [am] [on] [it] [!] [Where] ……..

Word Embedding Layer (tokens -> 300 dimensional token-vector)

tokens   start   I   am   on   it   !   Where   should   it   be   ?

14

Inference

I am on it ! Where should it be ?

Encoder   **Decoder with Attention**

15
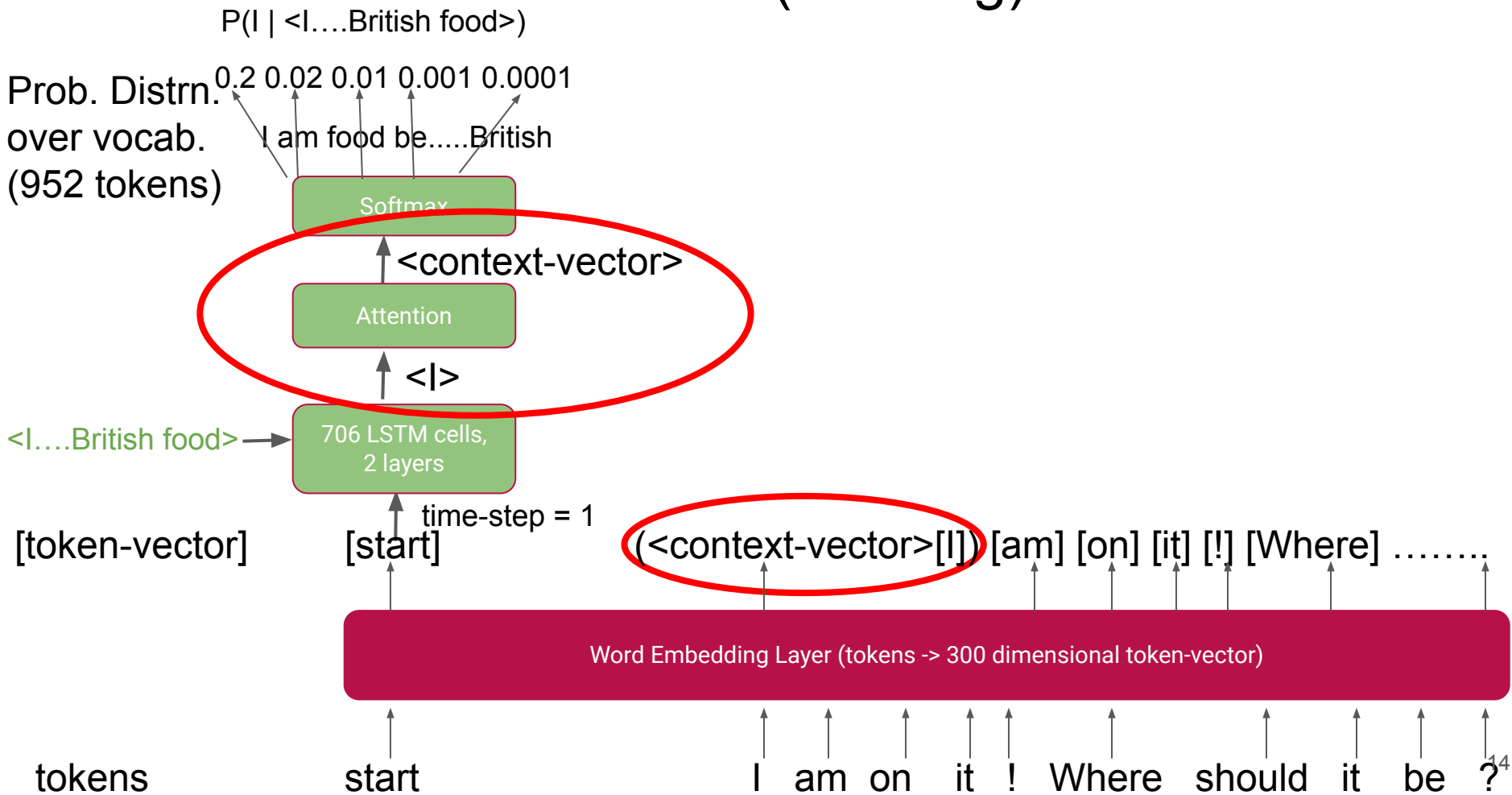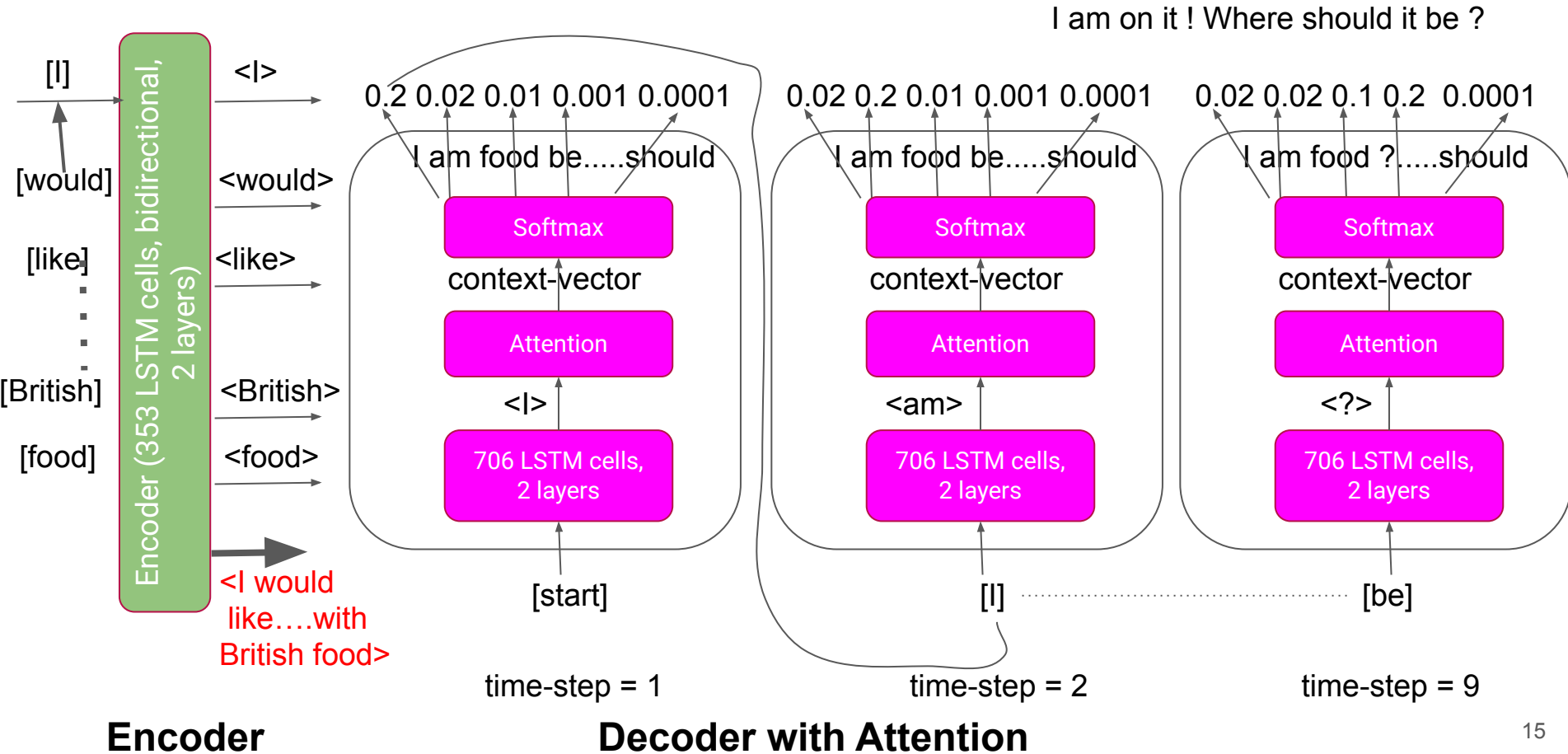
Beam Search (Inference); beam-width = 3

time-step = 1

time-step = 2

time-step = 3 …………………….

Start

working 2×10^{-6}
How 2×10^{-1}
Where 9×10^{-2}
She 32×10^{-9}
I 2×10^{-3}

How 2×10^{-1}

should 3×10^{-9}
many 2×10^{-6}
was 9×10^{-12}
am 4×10^{-9}
like 2×10^{-13}

Where 9×10^{-2}

should 3×10^{-1}
many 2×10^{-9}
was 9×10^{-12}
am 4×10^{-9}
like 2×10^{-13}

I 2×10^{-3}

should 3×10^{-6}
many 2×10^{-9}
was 9×10^{-12}
am 4×10^{-1}
like 2×10^{-2}

Where should $9 \times 10^{-2} \times 3 \times 10^{-1} = 27 \times 10^{-3}$

it 3×10^{-1}
many 2×10^{-9}
was 9×10^{-12}
on 4×10^{-9}
like 2×10^{-13}

I like $2 \times 10^{-3} \times 2 \times 10^{-2} = 4 \times 10^{-5}$

it 3×10^{-1}
many 2×10^{-9}
was 9×10^{-12}
on 4×10^{-9}
like 2×10^{-13}

I am $2 \times 10^{-3} \times 4 \times 10^{-1} = 8 \times 10^{-4}$

it 3×10^{-1}
many 2×10^{-9}
was 9×10^{-12}
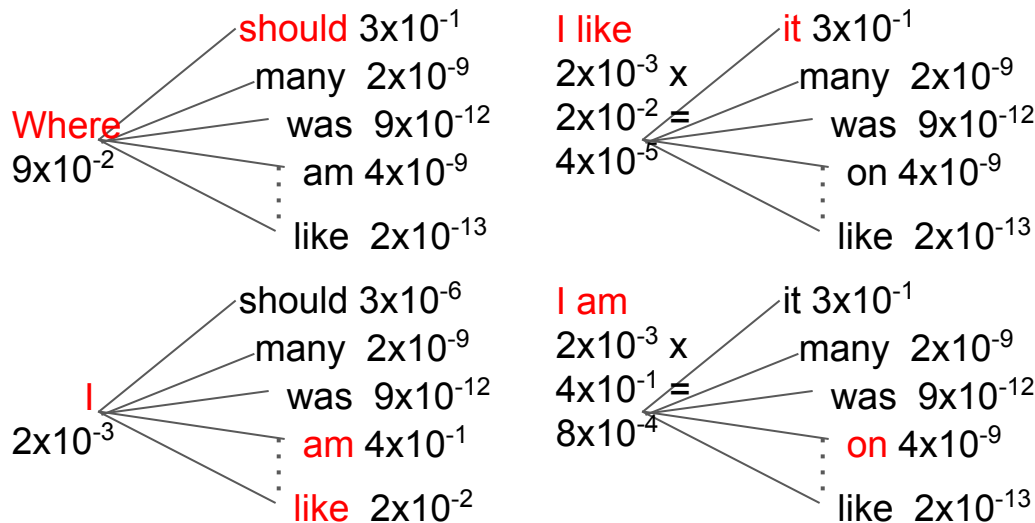**on** 4×10^{-9}
like 2×10^{-13}

---

x = I would like .. British food

time-step = 1
$P(y^1 \mid x)$

time-step = 2
$(\arg \max y) \; P(y^1, y^2 \mid x) = P(y^1 \mid x).P(y^2 \mid x, y^1)$

time-step = 3
$(\arg \max y) \; P(y^1, y^2, y^3 \mid x) = P(y^1 \mid x).P(y^2 \mid x, y^1).P(y^3 \mid x, y^1, y^2)$

Vocabulary of 952 tokens; if each sequence is 10 words long then 952^{10} combination of sequences:
How many people are in your party?
Where should it be?

I am on it! Where should it be?
I like it! Where should it be?

Dataset

- Real human-machine dialogs vs. Synthetic
    - Bots do very well on Synthetic datasets
    - Real datasets are a challenge
- Dialog State Tracking Challenge 2 (DSTC 2)
  https://arxiv.org/pdf/1605.07683.pdf
- Noisy because of speech -> text
- Accuracy metric is unnecessarily stringent
    - F1 metric is better but not used

Result

** Translated 1117 dialogs with 11237 turns and 782520 tokens using beam=3 in
   195.9s (5.70 dialogs/s, 57.36 turns/s, 3994.54 tokens/s)
** % number of dialogs that passed = (18/1117 x 100) = 1.61%
   ** (# of turns in dialog: # of occurrences) = (5: 7), (6: 8), (7: 3)
** % number of turns that passed = (4249/11237 x 100) = 37.81%
** % number of dialogs that failed = (1099/1117 x 100) = 98.39%
      ** (# of consecutive turns that passed, counting from beginning of dialog
         / # of turns in dialog: # of occurrences) = (2/4: 1), (1/5: 10), (3/5:
         1), (1/6: 63), (2/6: 2), (3/6: 3), (1/7: 142), (2/7: 8), (3/7: 3),
         (4/7: 1), (5/7: 1), (1/8: 160), (2/8: 16), (3/8: 2), (1/9: 149), (2/9:
         14), (1/10: 118), (2/10: 11), (3/10: 1), (1/11: 89), (2/11: 13),
         (3/11: 3), (5/11: 1), (1/12: 77), (2/12: 11), (1/13: 44), (2/13: 5),
         (3/13: 1), (1/14: 37), (2/14: 4), (3/14: 1), (1/15: 23), (2/15: 5),
         (3/15: 2), (1/16: 13), (2/16: 3), (3/16: 1), (1/17: 9), (2/17: 1),
         (1/18: 9), (2/18: 2), (1/19: 11), (1/20: 6), (2/20: 1), (3/20: 1),
         (1/21: 6), (1/22: 2), (2/22: 1), (1/23: 3), (1/24: 2), (1/25: 1),
         (1/26: 1), (1/27: 2), (1/28: 1), (1/29: 1)
** Generate test with beam=3: BLEU4 = 23.11, 25.5/22.9/22.1/22.1 (BP=1.000,
   ratio=2.070, syslen=283321, reflen=136863)

Dialog Passed

Tr1-S: <SILENCE>

Tr1-T: Hello , welcome to the Cambridge restaurant system . You can ask for restaurants by area , price range or food type . How may I help you ?

Tr1-H0-P: Hello , welcome to the Cambridge restaurant system . You can ask for restaurants by area , price range or food type . How may I help you ?

Tr2-S: cheap restaurant on the east part of town   {Price, Location}

Tr2-T: api_call R_cuisine east cheap

Tr2-H0-P: api_call R_cuisine east cheap

Tr3-S: <SILENCE>

Tr3-T: the_missing_sock is a nice place in the east of town and the prices are cheap

Tr3-H0-P: the_missing_sock is a nice place in the east of town and the prices are cheap

Tr4-S: what is the address

Tr4-T: Sure , the_missing_sock is on the_missing_sock_address

Tr4-H0-P: Sure , the_missing_sock is on the_missing_sock_address

Tr5-S: phone number

Tr5-T: The phone number of the_missing_sock is the_missing_sock_phone

Tr5-H0-P: The phone number of the_missing_sock is the_missing_sock_phone

Tr6-S: what type of food   {cuisine}

Tr6-T: the_missing_sock serves international food

Tr6-H0-P: the_missing_sock serves international food

Tr7-S: thank you good bye

Tr7-T: you are welcome

Tr7-H0-P: you are welcome

Entity Types (required):
Cuisine, Location, Price
Entity Types (other):
Post-code, Phone, Address

19

Dialog Failed

Tr1-S: <SILENCE>

Tr1-T: Hello , welcome to the Cambridge restaurant system . You can ask for restaurants by area , price range or food type . How may I help you ?

Tr1-H0-P: Hello , welcome to the Cambridge restaurant system . You can ask for restaurants by area , price range or food type . How may I help you ?

Tr2-S: expensive australian restaurant   {Price, Cuisine}

Tr2-T: You are looking for a australian restaurant right?

Tr2-H0-F: What part of town do you have in mind?   {Location?}

Tr3-S: yes

Tr3-T: I'm sorry but there is no restaurant serving australian food

Tr3-H0-F: What kind of food would you like?

Tr4-S: expensive french restaurant   {Price, Cuisine}

Tr4-T: You are looking for a australian restaurant right?

Tr4-H0-F: api_call french R_location expensive

Tr5-S: french

Tr5-T: There are restaurants serving french in the expensive price range . What area would you like?

Tr5-H0-F: api_call french R_location expensive

Tr6-S: any

Tr6-T: api_call french R_location expensive

Tr6-H0-F: Would you like something in the cheap , moderate , or expensive price range?

Entity Types (required):
Cuisine, Location, Price
Entity Types (other):
Post-code, Phone, Address

Publications

- Bordes, A., Boureau, Y. L., & Weston, J. (2016). Learning end-to-end goal-oriented dialog. *arXiv preprint arXiv:1605.07683.* https://arxiv.org/pdf/1605.07683.pdf
- Eric, M., & Manning, C. D. (2017). A copy-augmented sequence-to-sequence architecture gives good performance on task-oriented dialogue. arXiv preprint arXiv:1701.04024. https://arxiv.org/pdf/1701.04024.pdf
- Vinyals, O., Kaiser, Ł., Koo, T., Petrov, S., Sutskever, I., & Hinton, G. (2015). Grammar as a foreign language. In Advances in neural information processing systems (pp. 2773-2781). https://papers.nips.cc/paper/5635-grammar-as-a-foreign-language.pdf
- Luong, M. T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025.* https://arxiv.org/pdf/1508.04025.pdf
- Bahdanau, D., & Cho, K. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv: 1409.0473. https://arxiv.org/pdf/1409.0473.pdf