CS11-747 Neural Networks for NLP

Models of Dialog and Conversation

Graham Neubig

Carnegie Mellon University
Language Technologies Institute

Site
https://phontron.com/class/nn4nlp2017/

Types of Dialog

- Who is talking?
  - Human-human
  - Human-computer
- Why are they talking?
  - Task driven
  - Chat

Models of Chat

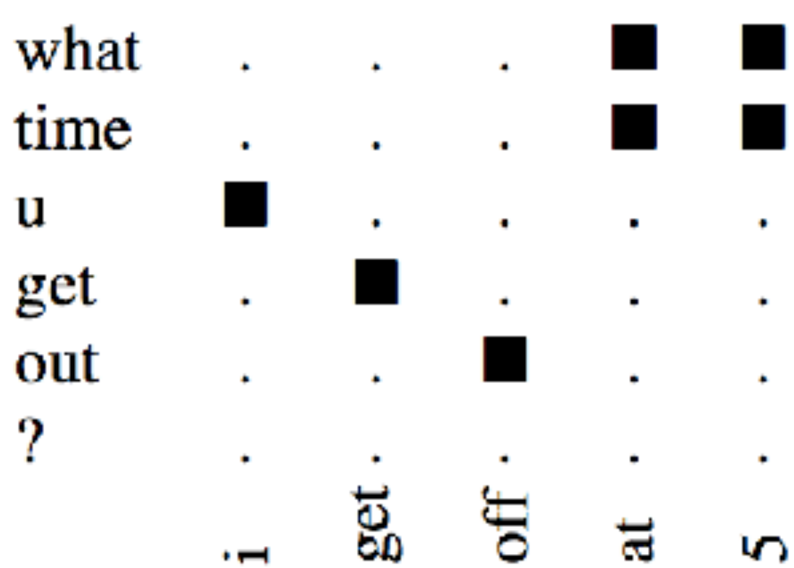
Two Paradigms

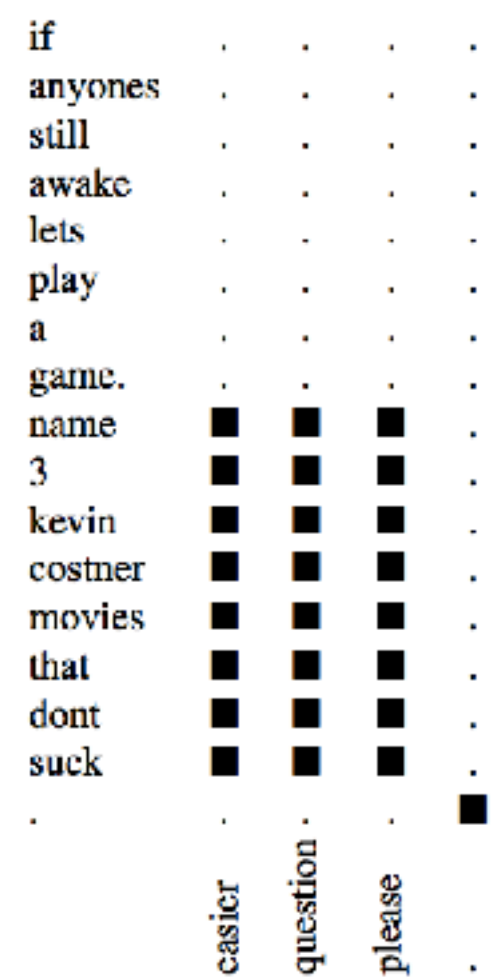
- Generation-based models
  - Take input, generate output
  - Good if you want to be creative
- Retrieval-based models
  - Take input, find most appropriate output
  - Good if you want to be safe

Generation-based Models

(Ritter et al. 2011)

- Train phrase-based machine translation system to perform translation from utterance to response
- Lots of filtering, etc., to make sure that the extracted translation rules are reliable





Neural Models for Dialog Response Generation

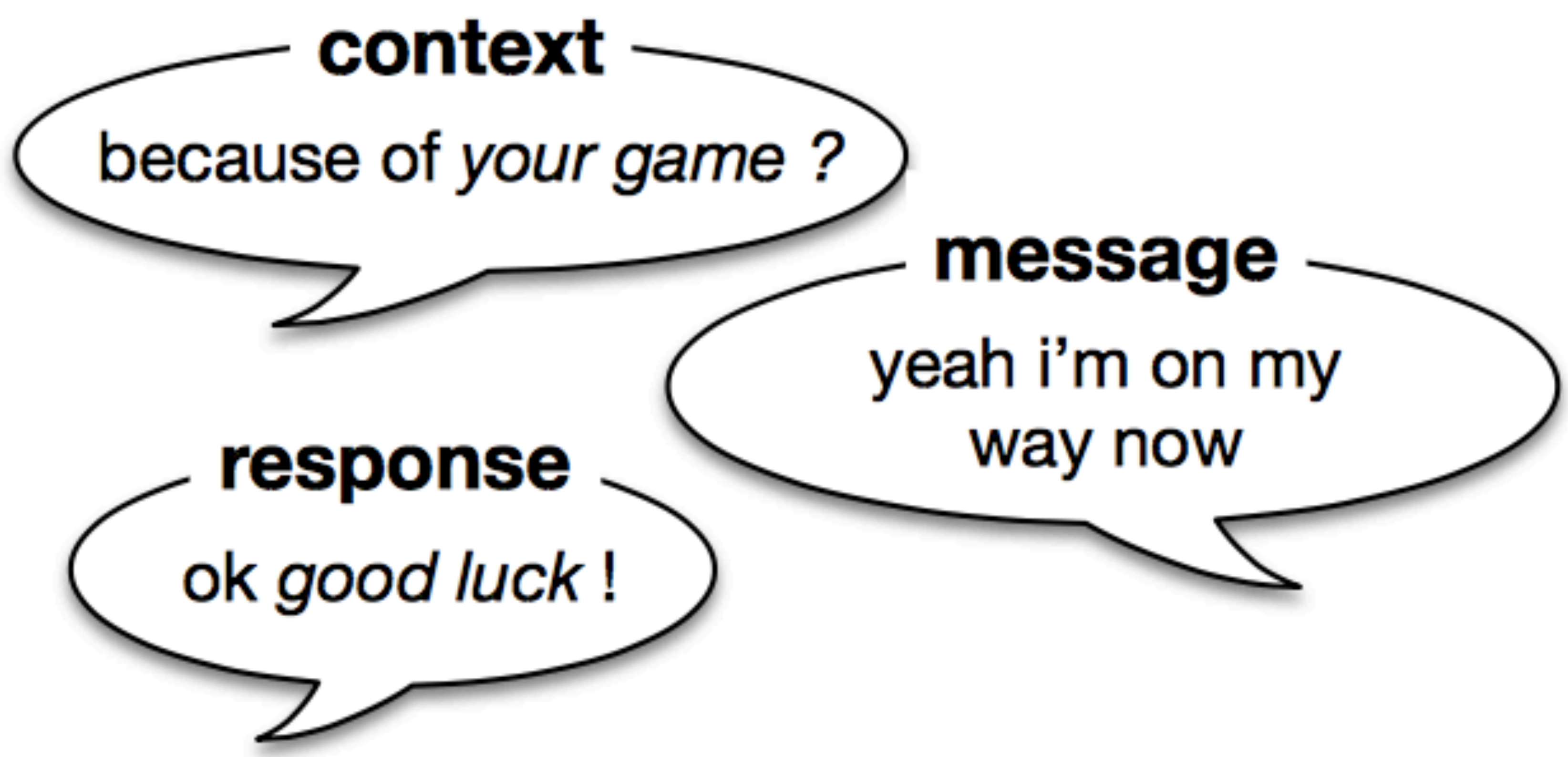
(Sordoni et al. 2015, Sheng et al. 2015, Vinyals and Le 2015)

- Like other translation tasks, dialog response generation can be done with encoder-decoders
- Sheng et al. (2015) present simplest model, translating from previous utterance

ID	Post	NRM-hyb
1	每年元旦都来高烧。 High fever attacks me every New Year's Day	祝早日康复，身体健康！ Get well soon and stay healthy!
2	又老了一岁，感谢我们组，我很高兴 I gain one more year. Grateful to my group, so happy	老了老了，岁月不饶人啊！ Getting old now. Time has no mercy.
3	吃素第一天，坚持住，崔朵拉。 First day of being a vegetarian. Hold on, **CuiDuoLa**	每天都是素食主义者 I have been being a vegetarian everyday
4	这种天气真应该带个帅锅出去踏青赏花 We should go out with some cute guys to enjoy a great outing in such a nice weather.	这种天气真不是盖的。 It is indeed a very nice weather.

Problem 1: Dialog More Dependent on Global Coherence

- Considering only a single previous utterance will lead to locally coherent but globally incoherent output
- Necessary to consider more context! (Sordoni et al. 2015)



- Contrast to MT, where context sometimes is (Matsuzaki et al. 2015) and sometimes isn't (Jean et al. 2015) helpful

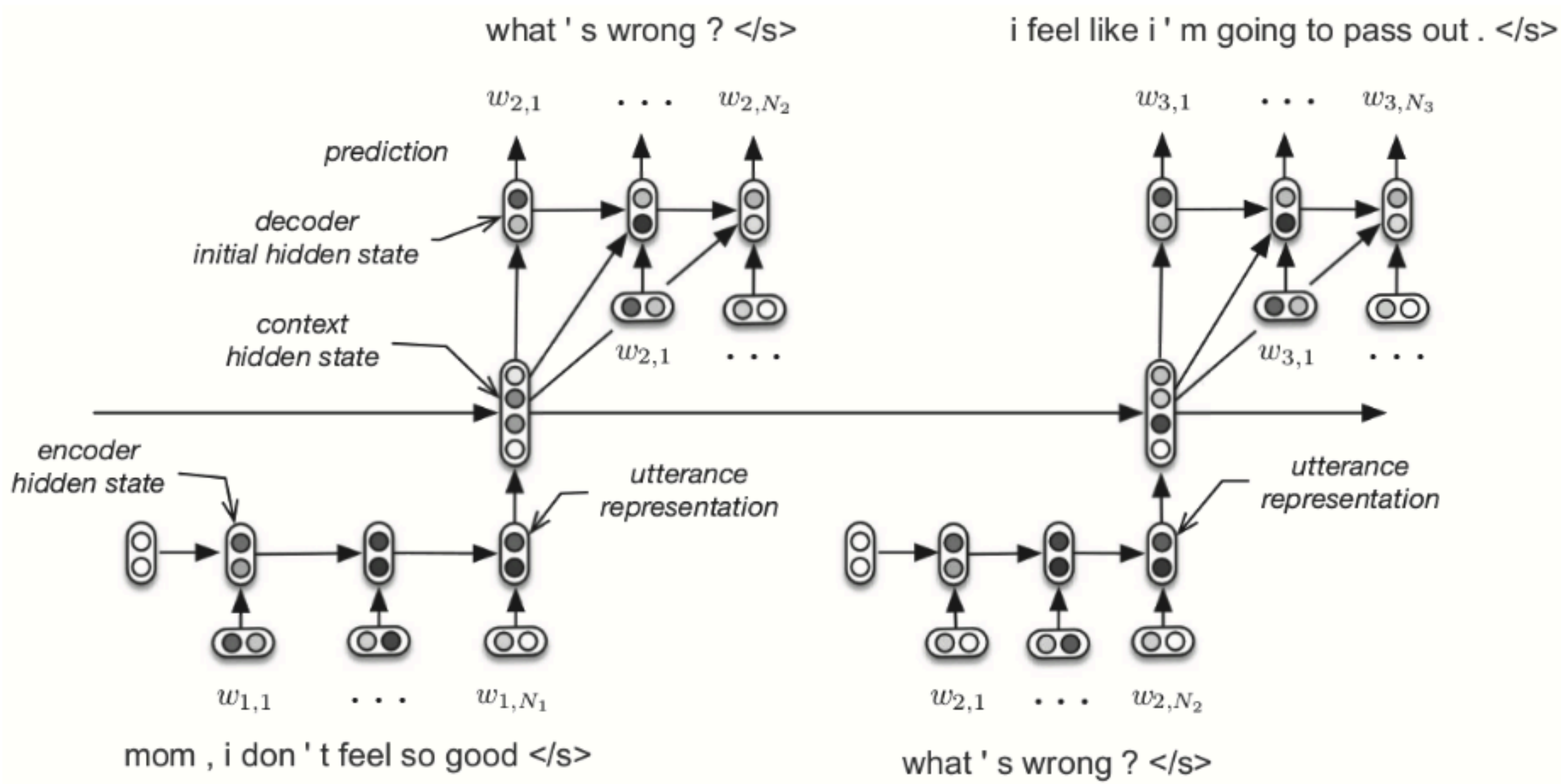
One Solution: Use Standard Architecture w/ More Context

- Sordoni et al. (2015) consider one additional previous context utterance concatenated together
- Vinyals et al. (2015) just concatenate together all previous utterances and hope an RNN an learn

Describe your problem: i am having issues accessing vpn
Machine: *hi*
Human: *hello*
Machine: *could you please let me know what are the operating systems you are using ?*
Human: *i am using linux*
Machine: *ok it was in the ssl vpn on which you stated it was broken*
Machine: *please check if you are logged in to corp network*
Machine: *what is the error that you are running please*
Human: *i am seeing an error related to vpn*
Machine: *what is the error message that you are getting when connecting to vpn using network connect ?*
Human: *connection refused or something like that*
Machine: *may i know the version of network connect you connect ?*
Human: *i am not sure i know that*
Machine: *is the network connect prompting for any user certificate*
Machine: *?*
Human: *yes*
Machine: *are you at home or at work at the moment ?*
Machine: *office ?*
Human: *i am at home*
Machine: *try this*
Machine: *goto <URL> page and enter your name and password*
Human: *done*
Machine: *try logging in with <NAME> and check if its working fine now*
Human: *yes , now it works !*
Machine: *great*
Machine: *anything else that i can help*
Human: *no , that was very useful*
Machine: *you are other than <NAME> :)*
Machine: *have a nice night*

Hierarchical Encoder-decoder Model (Serban et al. 2016)

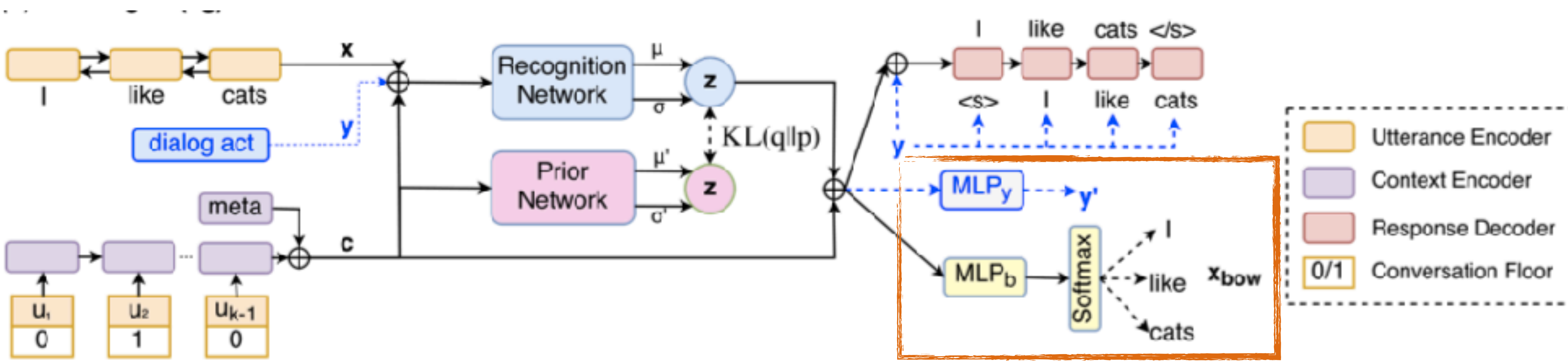
- Also have utterance-level RNN track overall dialog state



Discourse-level VAE Model

(Zhao et al. 2017)

- Encode entire previous dialog context as latent variable in VAE
- Also meta-information such as dialog acts



Also, bag-of-words loss

Problem 2: Dialog allows Much More Varied Responses

- For translation, there is lexical variation but content remains the same
- For dialog, content will also be different! (e.g. Li et al. 2016)

Input: What are you doing?			
−0.86	I don't know.	−1.09	Get out of here.
−1.03	I don't know!	−1.09	I'm going home.
−1.06	Nothing.	−1.09	Oh my god!
−1.09	Get out of the way.	−1.10	I'm talking to you.
Input: what is your name?			
−0.91	I don't know.	...	
−0.92	I don't know!	−1.55	My name is Robert.
−0.92	I don't know, sir.	−1.58	My name is John.
−0.97	Oh, my god!	−1.59	My name's John.
Input: How old are you?			
−0.79	I don't know.	...	
−1.06	I'm fine.	−1.64	Twenty-five.
−1.17	I'm all right.	−1.66	Five.
−1.17	I'm not sure.	−1.71	Eight.

Diversity Promoting Objective for Conversation (Li et al. 2016)

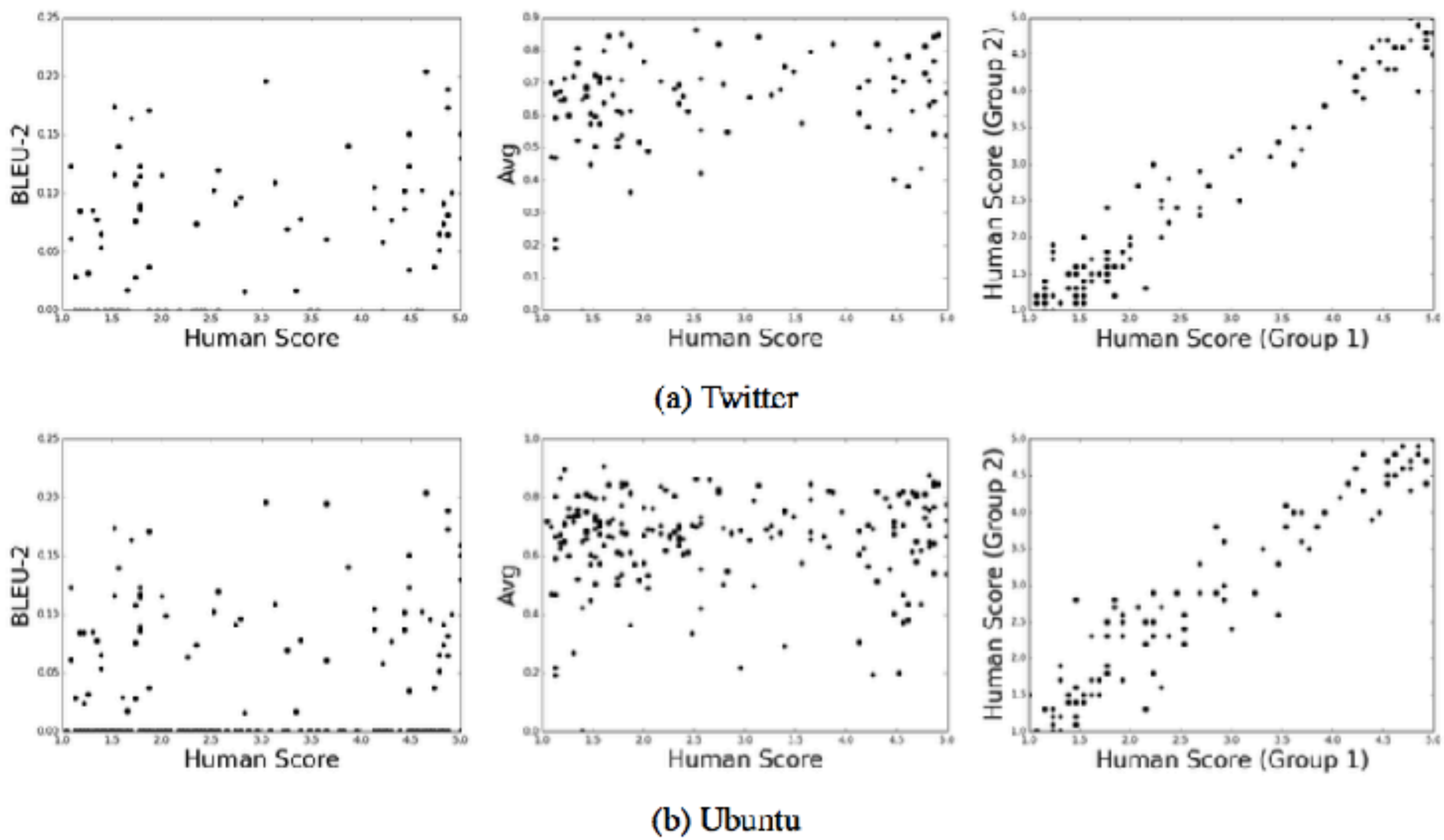
- Basic idea: we want responses that are likely given the context, unlikely otherwise
- Method: subtract weighted unconditioned log probability from conditioned probability (calculated only on first few words)

$$\log p(T|S) - \lambda \log U(T)$$

message	SEQ2SEQ	MMI $P(T\|S) - \lambda P(T)$
I did not get the report from an mi-6 agent.	I don't think that's a good idea.	You did the right thing, did you?
You haven't been given an assignment in this case.	I don't know what you are talking about.	I've been looking all over for you.
I'm losing my grip.	I don't know what you are talking about.	I'm the only one in the world.
I am ready to help.	Come on, come on	I have something we need to talk about.
You programmed me to gather intelligence. That's all I've ever done.	You know that, don't you?	You do have fun, don't you?
I mean, we'd have to talk to him.	I mean, I don't know.	I mean, he's a good guy.
This has nothing to do with him.	You don't know anything about him, do you?	I know him as much.
How come you never say it?	I don't know	Because I don't want to hurt you
He is the one behind this scheme to crash x-net.	I don't know what you are talking about	Yes, he is, isn't he?

Diversity is a Problem for Evaluation!

- Translation uses BLEU score; while imperfect, not horrible
- In dialog, BLEU shows very little correlation (Liu et al. 2016)



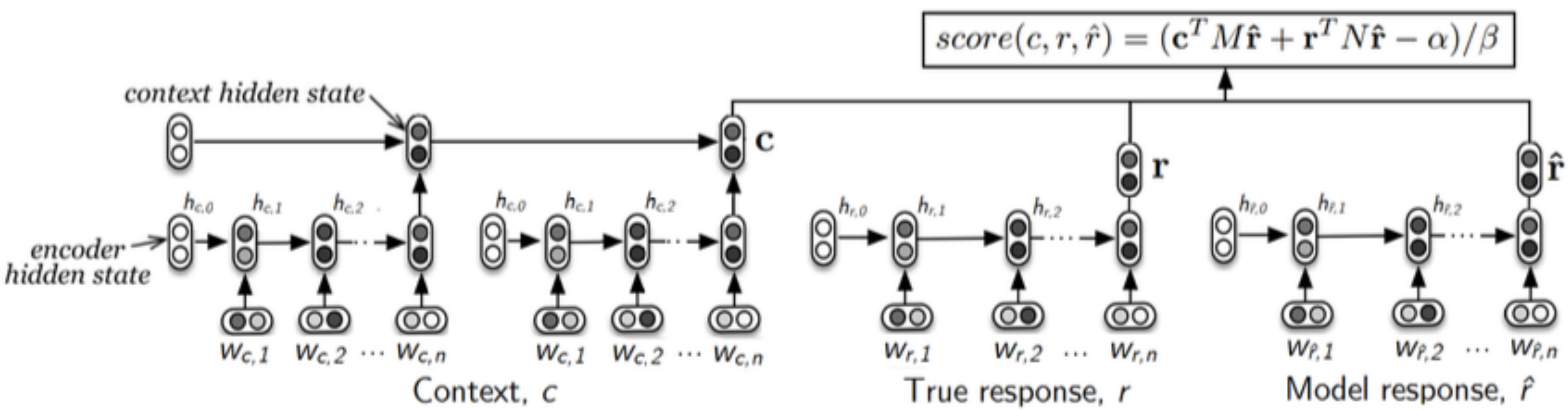
Using Multiple References with Human Evaluation Scores (Galley et al. 2015)

- Retrieve good-looking responses, perform human evaluation, up-weight good ones, down-weight bad ones

Context c	**Message** m	**Response** r	**Score**
i was about to text you and my two cousins got excited cause they thought you were "rihanna"	aww, i can imagine their disappointment	they were very disappointed!!!	0.6
yes. my ex-boyfriend, killed my cat. like i say, it was the start of a bad time...	*i can imagine!*	*yes. luckily, the whole thing feels very much of the past now.*	0.8
its good.. for some reason i can't name stand out tracks but i've been playing it since it dropped	*i can imagine, banks doesn't disappoint*	*na this is anything but a disappointment..*	0.6
at my lil cousins dancing to "dance for you". these kids are a mess.	*lmaoo i can imagine.*	*they were belly rolling, filarious.*	0.4
what's sick about it?? do you know how long it is?? no so how is it sick?	*i can imagine it*	*your imagination is wrong, very wrong at that.*	-0.1
it's saad oh yeah the snow is very beautiful	*yeah i can imagine*	*the weather in russia is very cool.*	-0.7

Learning to Evaluate

- Use context, true response, and actual response to learn a regressor that predicts goodness (Lowe et al. 2017)
- Important: similar to model, but has access to reference!



- Adversarial evaluation: try to determine whether response is true or fake (Li et al. 2017)
- One caveat from MT: learnable metrics tend to overfit

Problem 3: Dialog Agents should have Personality

- If we train on all of our data, our agent will be a mish-mash of personalities (e.g. Li et al. 2016)

message	Where do you live now?
response	I live in Los Angeles.
message	In which city do you live now?
response	I live in Madrid.
message	In which country do you live now?
response	England, you?

- We would like our agents to be consistent!

Personality Infused Dialog

(Mairesse et al. 2007)

- Train a generation system with controllable "knobs" based on personality traits

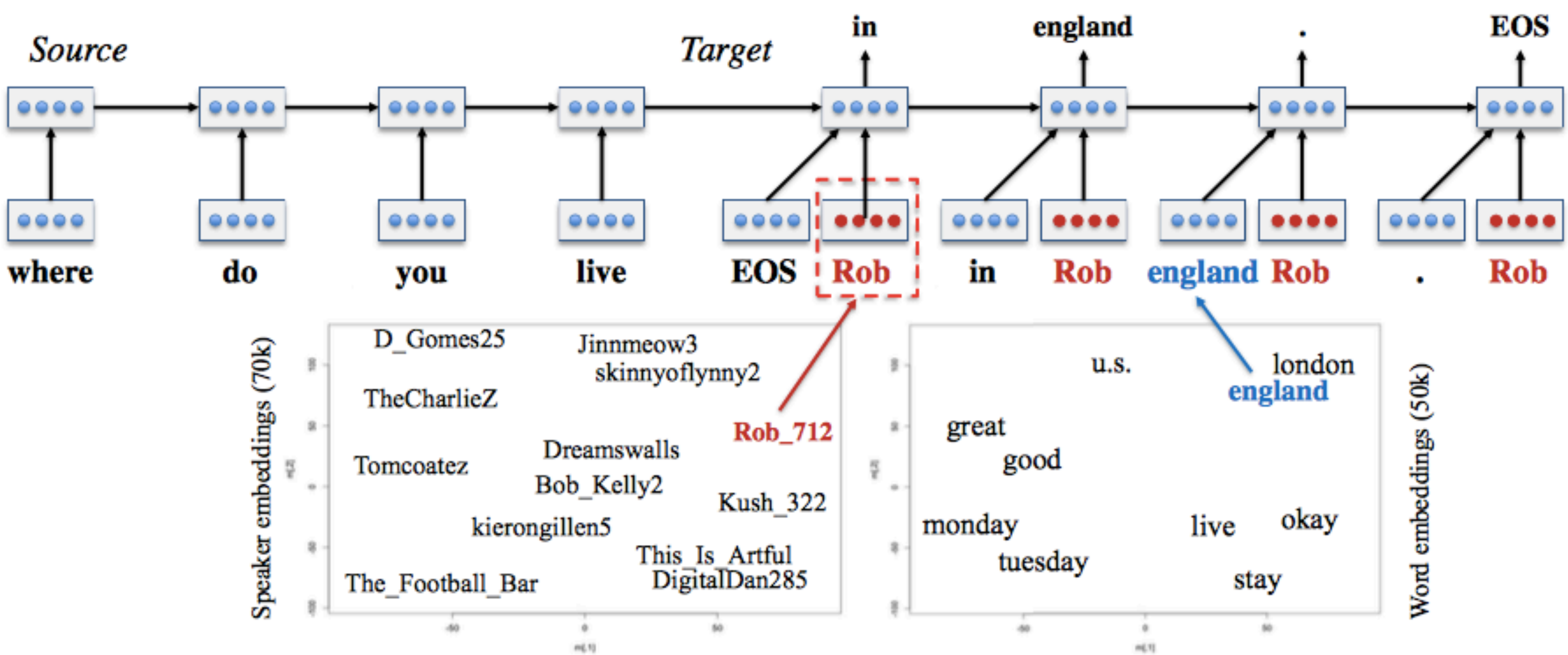
- e.g. Extraversion:

- Non-neural, but well done and perhaps applicable

Alt	Realization	Extra
5	Err... it seems to me that Le Marais isn't as bad as the others.	1.83
4	Right, I mean, Le Marais is the only restaurant that is any good.	2.83
8	Ok, I mean, Le Marais is a quite french, kosher and steak house place, you know and the atmosphere isn't nasty, it has nice atmosphere. It has friendly service. It seems to me that the service is nice. It isn't as bad as the others, is it?	5.17
9	Well, it seems to me that I am sure you would like Le Marais. It has good food, the food is sort of rather tasty, the ambience is nice, the atmosphere isn't sort of nasty, it features rather friendly servers and its price is around 44 dollars.	5.83
3	I am sure you would like Le Marais, you know. The atmosphere is acceptable, the servers are nice and it's a french, kosher and steak house place. Actually, the food is good, even if its price is 44 dollars.	6.00
10	It seems to me that Le Marais isn't as bad as the others. It's a french, kosher and steak house place. It has friendly servers, you know but it's somewhat expensive, you know!	6.17
2	Basically, actually, I am sure you would like Le Marais. It features friendly service and acceptable atmosphere and it's a french, kosher and steak house place. Even if its price is 44 dollars, it just has really good food, nice food.	6.17

Persona-based Neural Dialog Model (Li et al. 2017)

- Model each speaker in embedding space



- Also model who the speaker is speaking to in speaker-addressee model

Retrieval-based Models

Dialog Response Retrieval

- Idea: many things can be answered with template
- Simply find most relevant response out of existing ones in corpus

Template responses

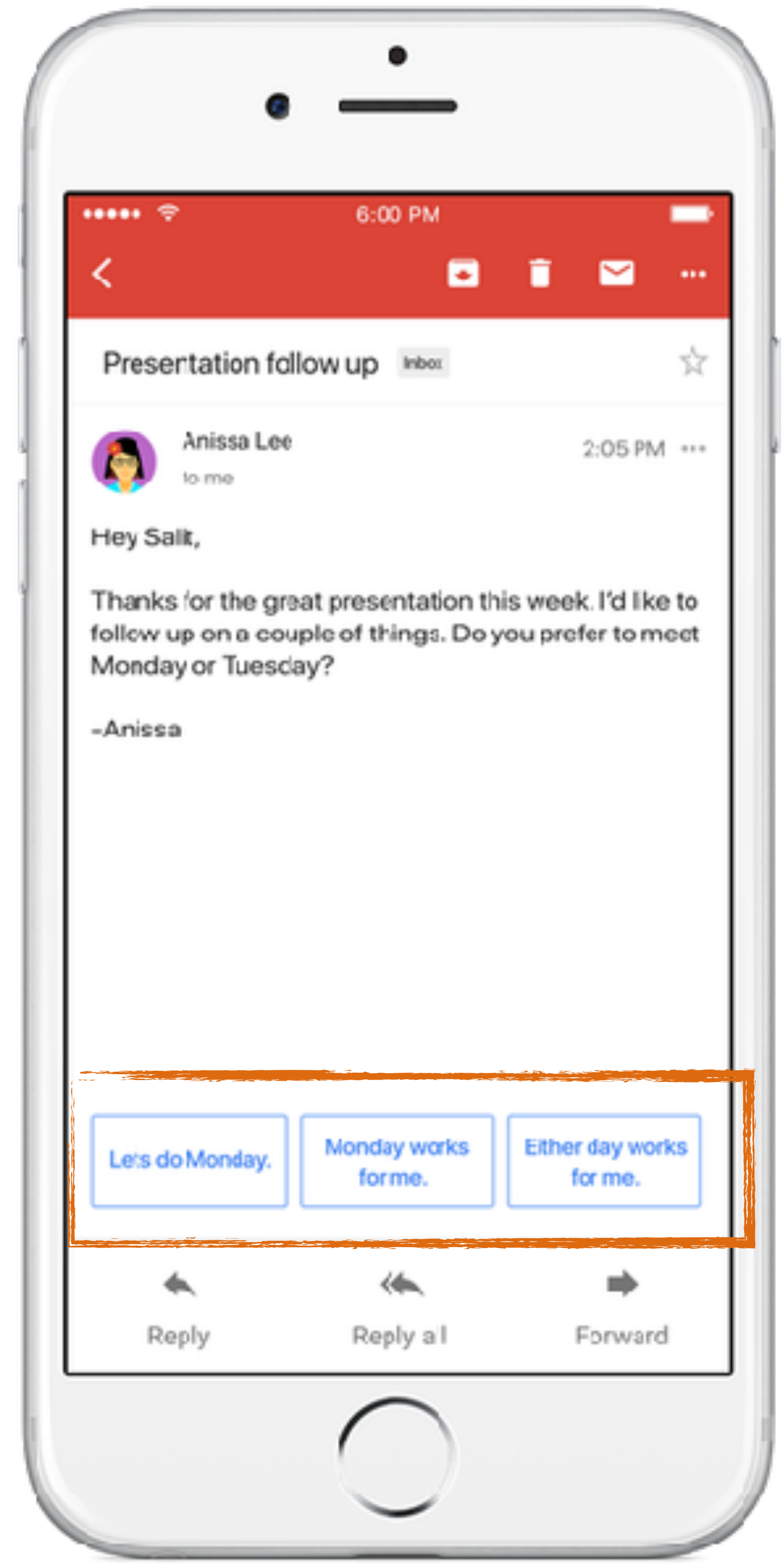
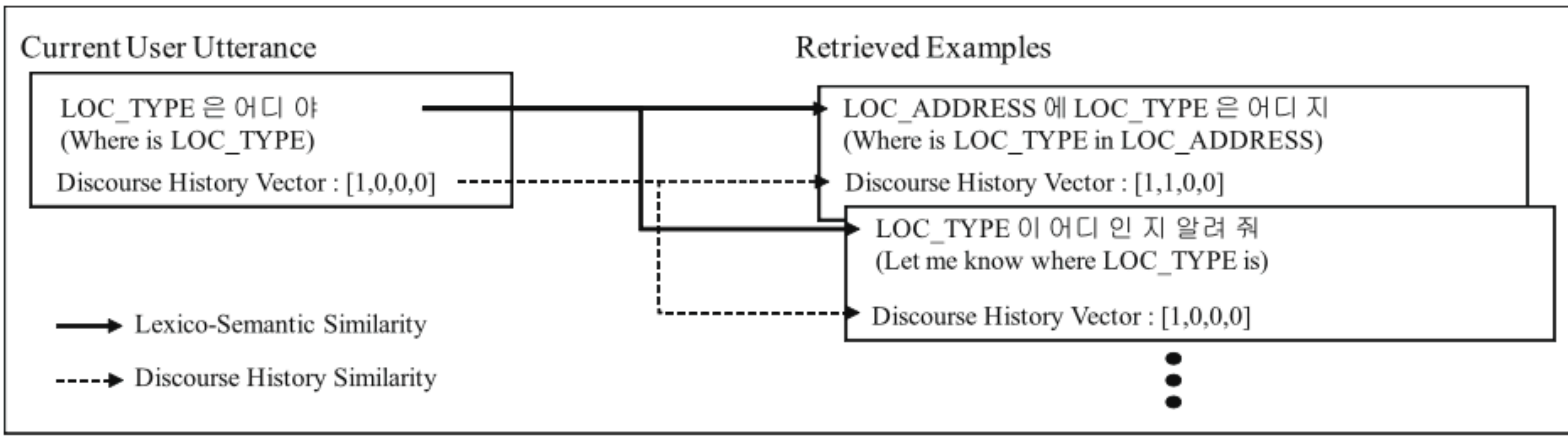


Image Credit: Google

Retrieval-based Chat
(Lee et al. 2009)

- Basic idea: given an utterance, find the most similar in the database and return it

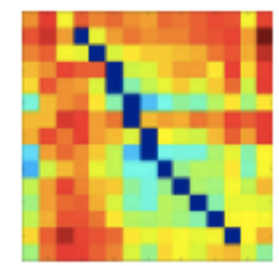
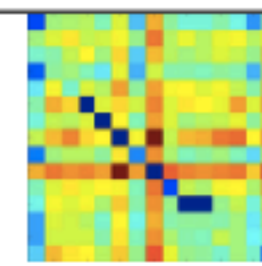
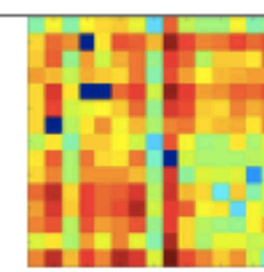


- Similarity based on exact word match, plus extracted features regarding discourse

Neural Response Retrieval

(Nio et al. 2014)

- Idea: use neural models to soften the connection between input and output and do more flexible matching

sim	Sentences	Matrix
0.94	S_1) Captain, we can not keep going fast on these icy roads. S_2) We can not keep going fast on these icy roads!	
0.60	S_1) Hold your fire! He's got a girl. S_2) Looks like he's got a hostage.	
0.38	S_1) Yes, I can see that too and I don't think it's so terrible. S_2) That's why I do all the thinking.	

- Model uses Socher et al. (2011) recursive auto-encoder + dynamic pooling

Smart Reply for Email Retrieval (Kannan et al. 2016)

- Implemented in GMail smart reply
- Similar response model with LSTM seq2seq scoring, but many improvements
    - **Beam search** over response space for scalability
    - **Canonicalization** of syntactic variants and **clustering** of similar responses
    - **Human curation** of responses
    - Enforcement of diversity through **omission of redundant responses** and **enforcing positive/negative**

Task-driven Dialog

Chat vs. Task Completion

- Chat is basically to keep the user entertained
- What if we want to do an actual task?
  - Book a flight
  - Access information from a database

Traditional Task-completion Dialog Framework

- In semantic frame based dialog:
- **Natural language understanding** to fill the slots in the frame based on the user utterance
- **Dialog state tracking** to keep track of the overall dialog state over multiple turns
- **Dialog control** to decide the next action based on state
- **Natural language generation** to generate utterances based on current state

NLU (for Slot Filling) w/ Neural Nets (Mesnil et al. 2015)

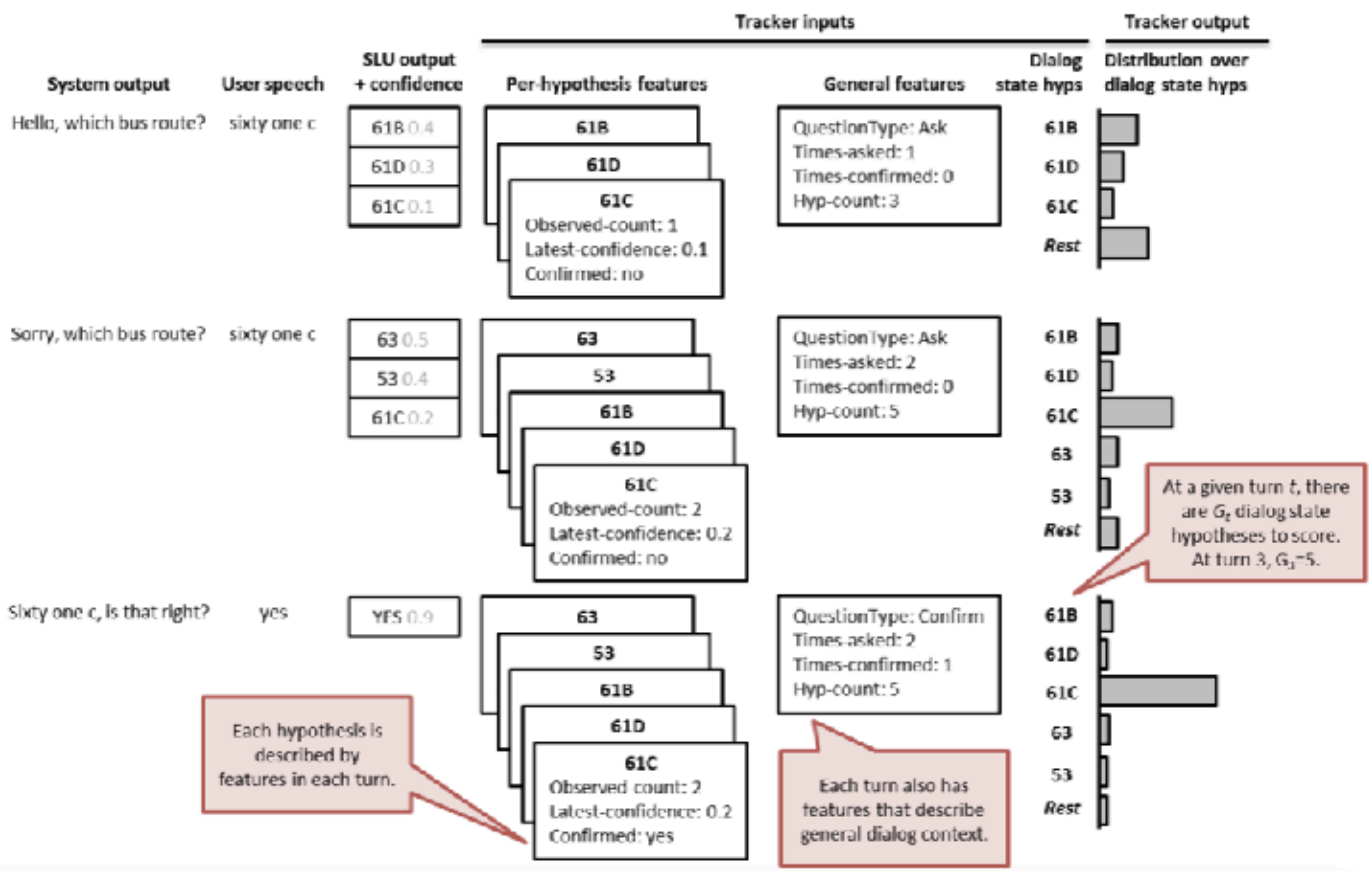
- Slot filing expressed as BIO scheme

Sentence	*show*	*flights*	*from*	*Boston*	*To*	*New*	*York*	*today*
Slots/Concepts	O	O	O	B-dept	O	B-arr	I-arr	B-date
Named Entity	O	O	O	B-city	O	B-city	I-city	O
Intent	*Find_Flight*							
Domain	*Airline Travel*							

- RNN-CRF based model for tags

Dialog State Tracking

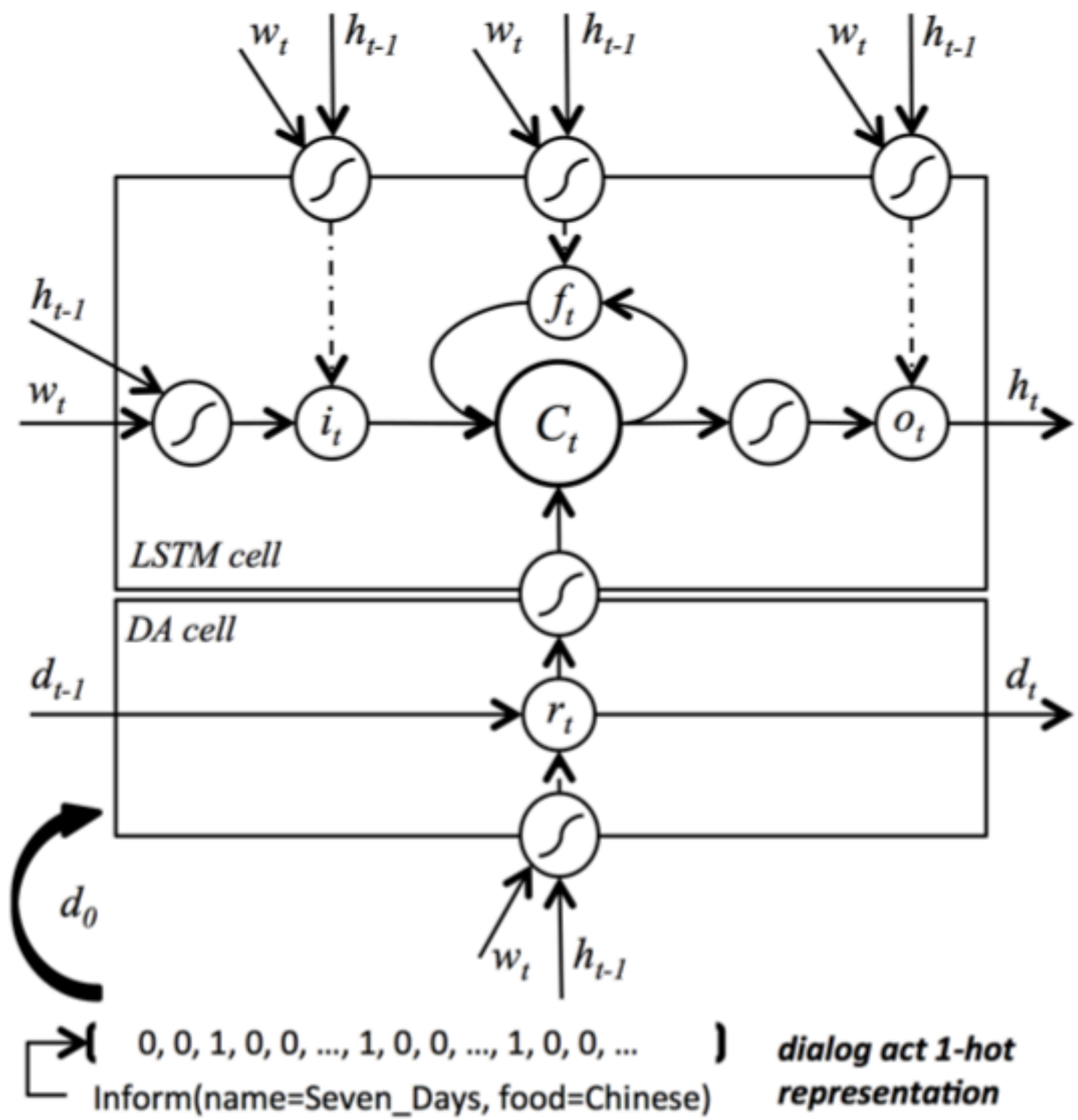
- Track the belief about our current frame-filling state (Williams et al. 2013)



- Henderson et al. (2014) present RNN model that encodes multiple ASR hypotheses and generalizes by abstracting details

Language Generation from Dialog State w/ Neural Nets (Wen et al. 2015)

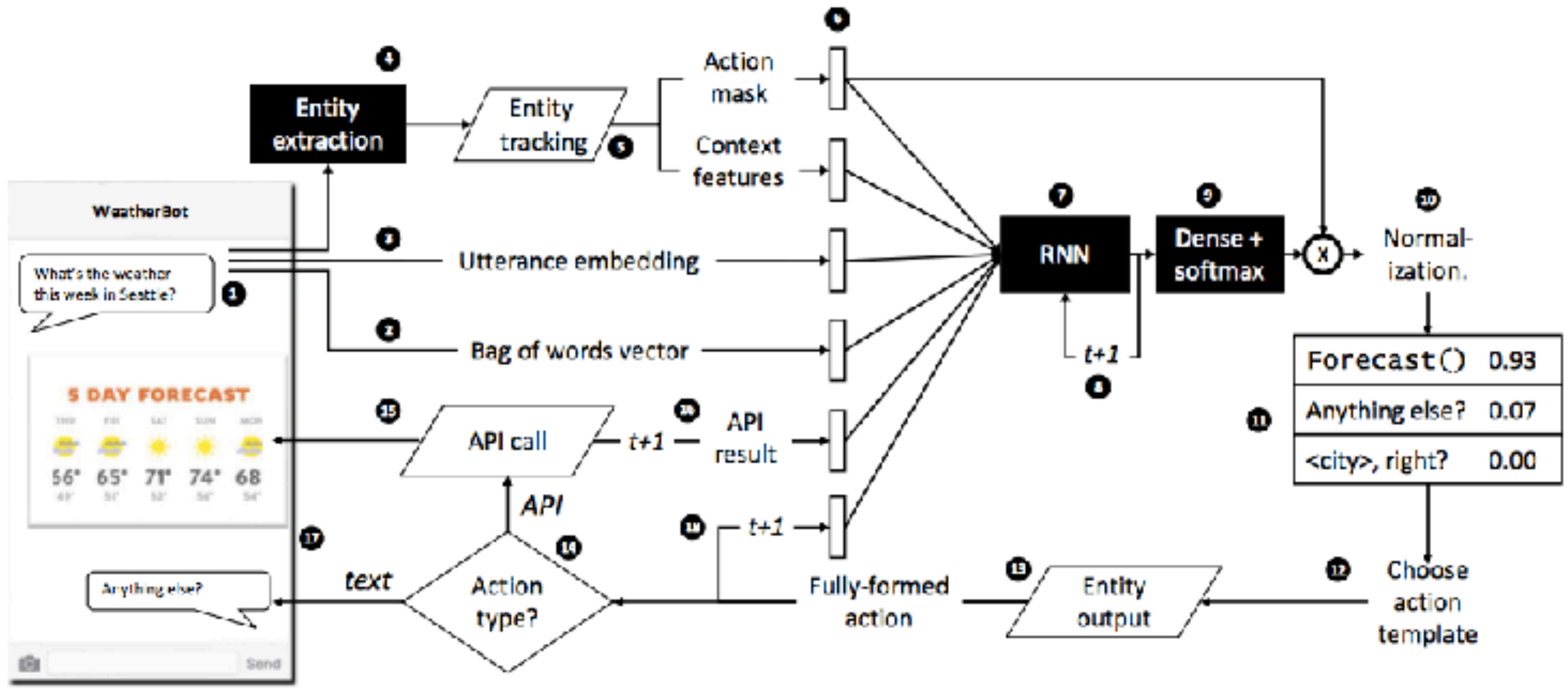
- Condition LSTM units based on the dialog input, output English



End-to-end Dialog Control

(Williams et al. 2017)

- Train an LSTM that takes in text and entities and directly chooses an action to take (reply or API call)



- Trained using combination of supervised and reinforcement learning

Questions?