Neural Dialog Models: A Survey

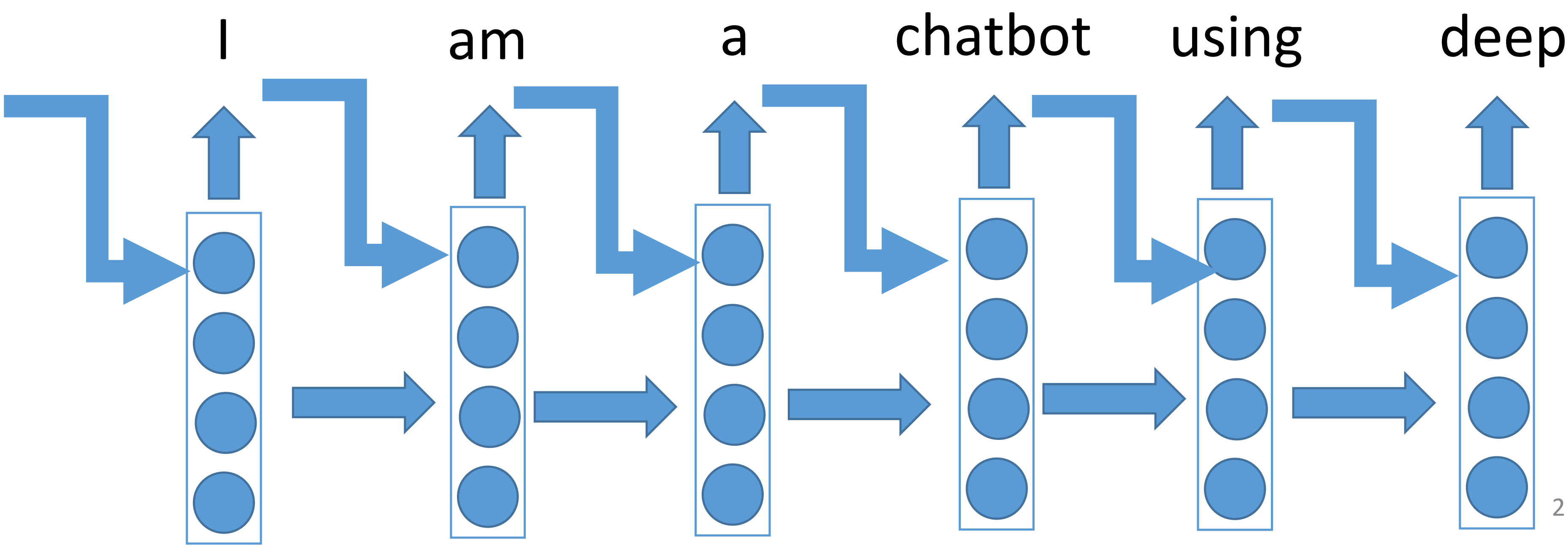
- Imitation Learning Approach -

Yuta Tsuboi (IBM Research – Tokyo) and

Hiroki Ouchi (NAIST)

Neural Language models

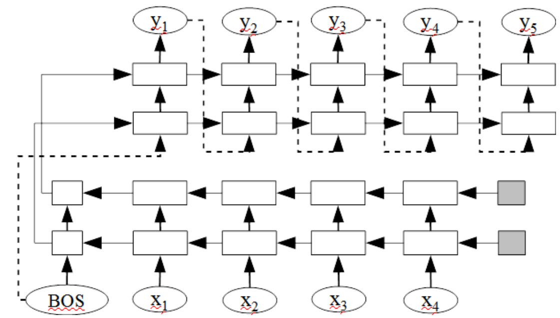
- Feed-forward LM [Bengio et al., 2003]
  - Next word predictor
- RNN-based LM [Mikolov et al., 2013]



Encoder-decoder approach for sequence learning

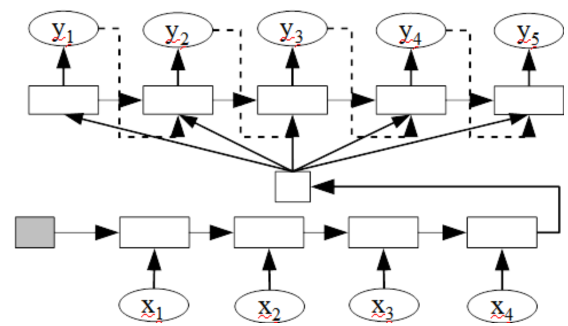
1. (Sutskever et al., 2014)
   The tail output of encoder is the initial state of decoder

(1)



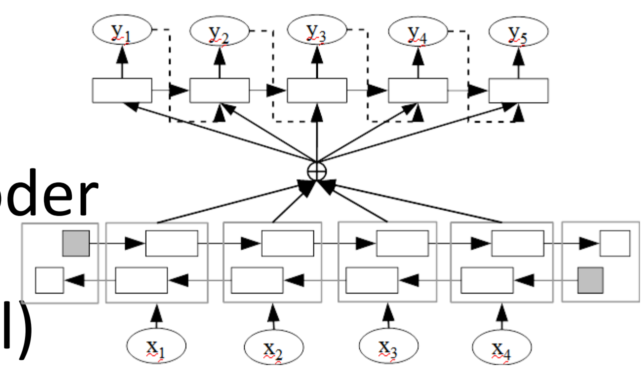
2. (Cho et al., 2014)
   The tail output of encoder is referred at all output prediction

(2)

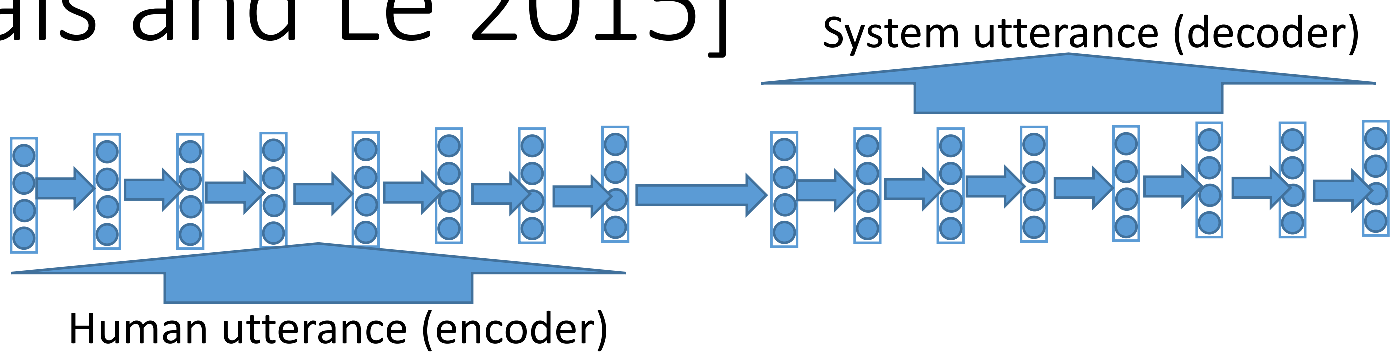


3. (Bahdanau et al., 2014)
   The weighted combination of encoder outputs is referred at all the output prediction (attention model)

(3)



A Neural Conversation Model [Vinyals and Le 2015]



- An example dialog using IT Helpdesk data [Vinyals and Le 2015]
  - **Describe your problem:** i am having issues accessing vpn
  - **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 ? …

A Neural Conversation Model [Vinyals and Le 2015]

- Although NN just mimic the utterances of human operators, it looks solving the problem.
  - ….
  - **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*
  - (The <*URL*> indeed contains information about vpn access.)
- Special tokens indicating turn taking and actor.
- Perplexity: 8 (Neural) vs. 18 (n-gram model)

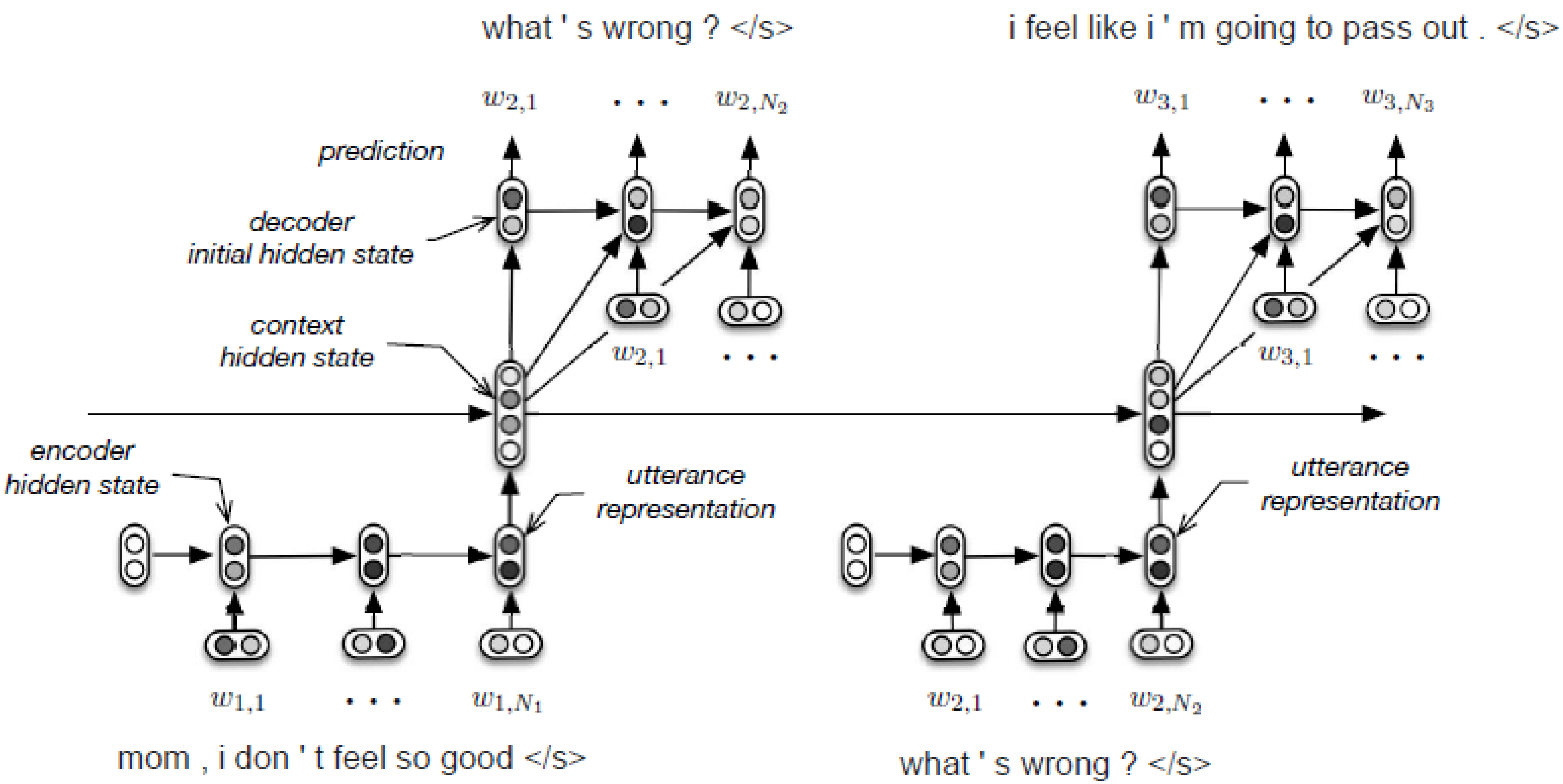
Imitation learning for dialog system

- Conventional approach: goal-driven dialog for travel assistance, technical support, etc.
  Requires 1) a measurable goal and 2) task-specific annotated conversations
- Imitation learning approach: mimic human conversations
  - Can be applicable to tasks which are not easy to numerically evaluate the result of conversations (difficult reward design)
  - Can be a unsupervised representation learning for goal-driven dialog system
  - Can be a human simulator for POMDP-based dialog system

Generation task is difficult to evaluate

- Evaluation metrics so far
  - BLEU/METEOR over reference response [Sordoni+ 2015]
  - Perplexity / word classification [Vinyals 2015] [Serban+ 2015]
  - Human evaluation [Shang+ 2015]
  - Response selection task [Lowe+ 2015]
  - Dialog act prediction task [Kalchbrenner+ 2013]

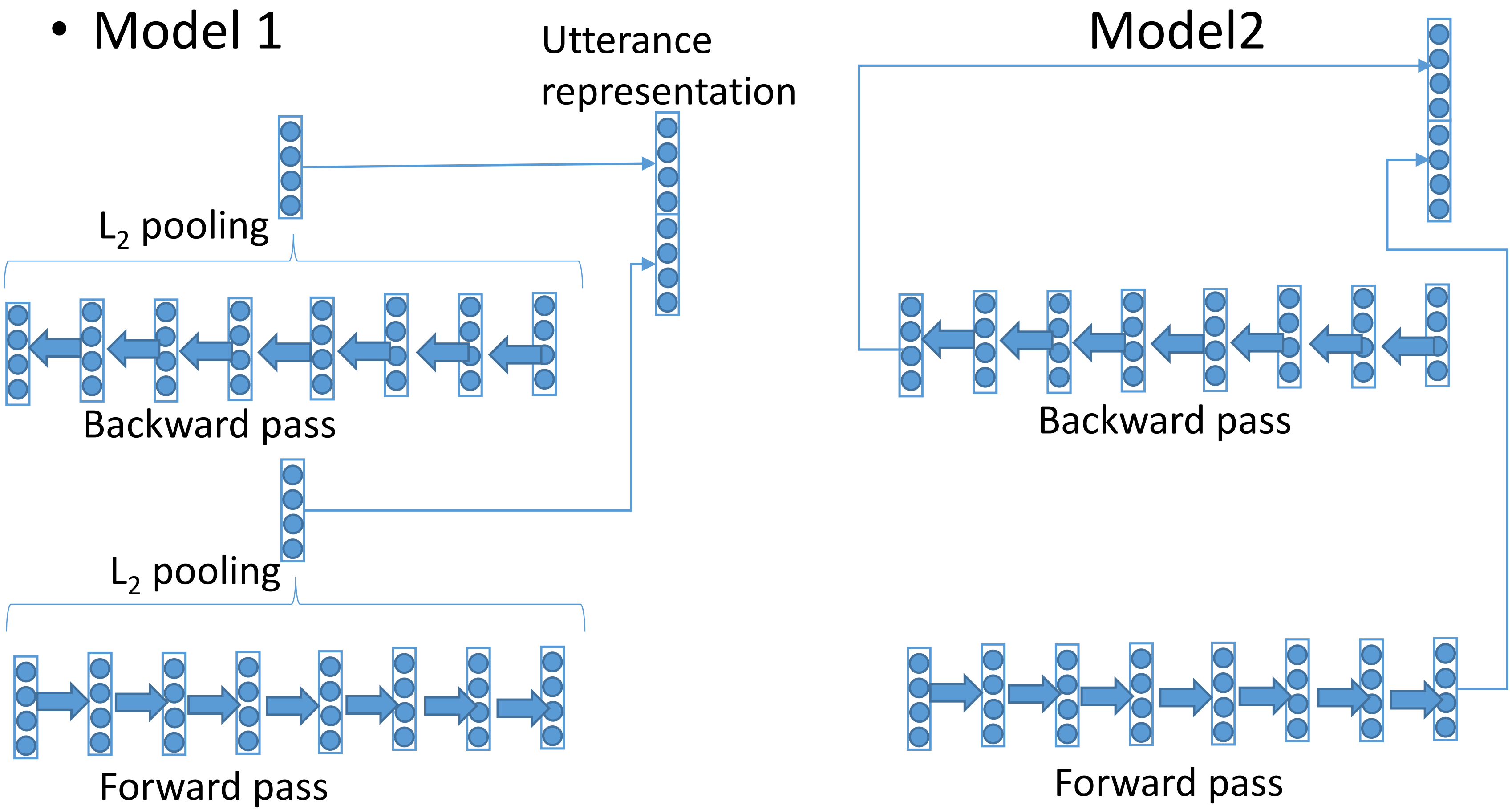
HRED: hierarchical recurrent encoder decoder [Serban+ 2015]



Figures are cited from original paper

Bidirectional HRED [Serban+ 2015]

- Bidirectional RNN for utterance representation
  - Model 1



Experimental results [Serban+ 2015]

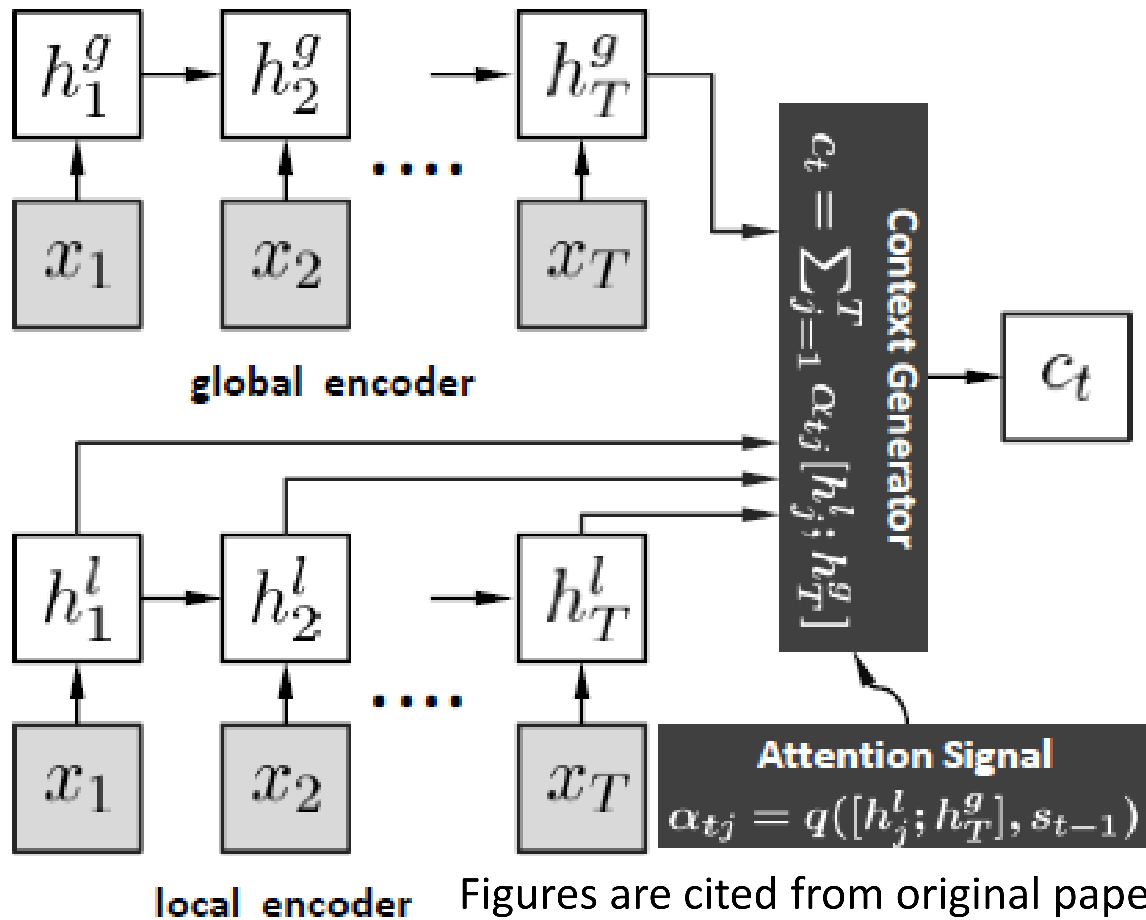
- MovieTriples dataset: 3 consecutive utterances between two interlocutors from movies
  - 196,308 triples for training, 53 tokens/triples
- Pre-training
  - word embedding (word2vec)
  - 5M Q-A pairs (movie subtitles)
- Results: Note that the majority of the predictions are generic, such as *I don't know* or *I'm sorry*.

	All	
Model	**Perplexity**	**Perplex.@U_3**
RNN	27.09 ± 0.13	26.67 ± 0.19
HRED	27.14 ± 0.12	26.60 ± 0.19
HRED-Bi.	$\mathbf{26.81 \pm 0.11}$	$\mathbf{26.31 \pm 0.19}$

Table is cited from original paper

Attention-based decoder for one round of conversation [Shang+ 2015]

- Dynamically select and linearly combine different parts of the input utterance



Figures are cited from original paper

Response Selection Task (1)

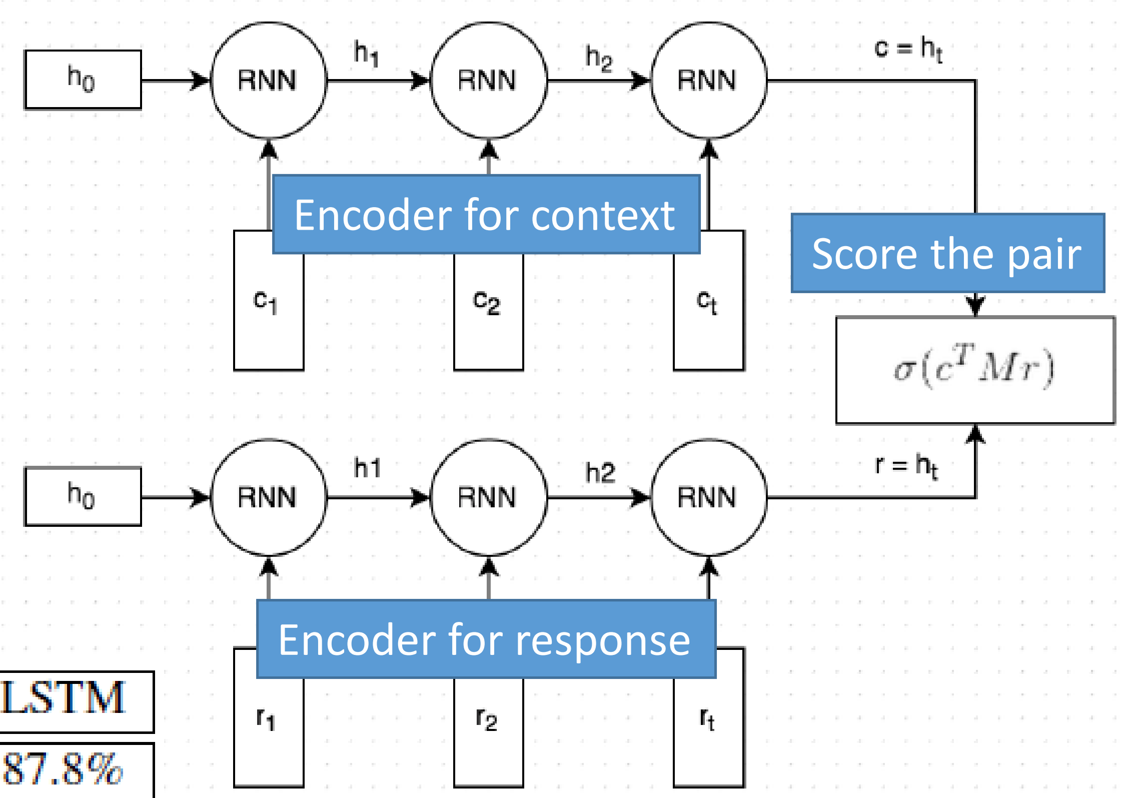
- The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems [Lowe+ 2015]
- Best response selection corresponding to a given dialog context
  - BLEU can provide a very low score for reasonable responses
  - Avoid the direct evaluation of generation task

Context	Response	Flag
well, can I move the drives? __EOS__ ah not like that	I guess I could just get an enclosure and copy via USB	1
well, can I move the drives? __EOS__ ah not like that	you can use "ps ax" and "kill (PID #)"	0

Figures are cited from original paper

Response Selection Task (2) [Lowe+ 2015]

- Encoders for both context and response (no decoder)



Method	TF-IDF	RNN	LSTM
1 in 2 R@1	65.9%	76.8%	87.8%
1 in 10 R@1	41.0%	40.3%	60.4%
1 in 10 R@2	54.5%	54.7%	74.5%
1 in 10 R@5	70.8%	81.9%	92.6%

Figures are cited from original paper

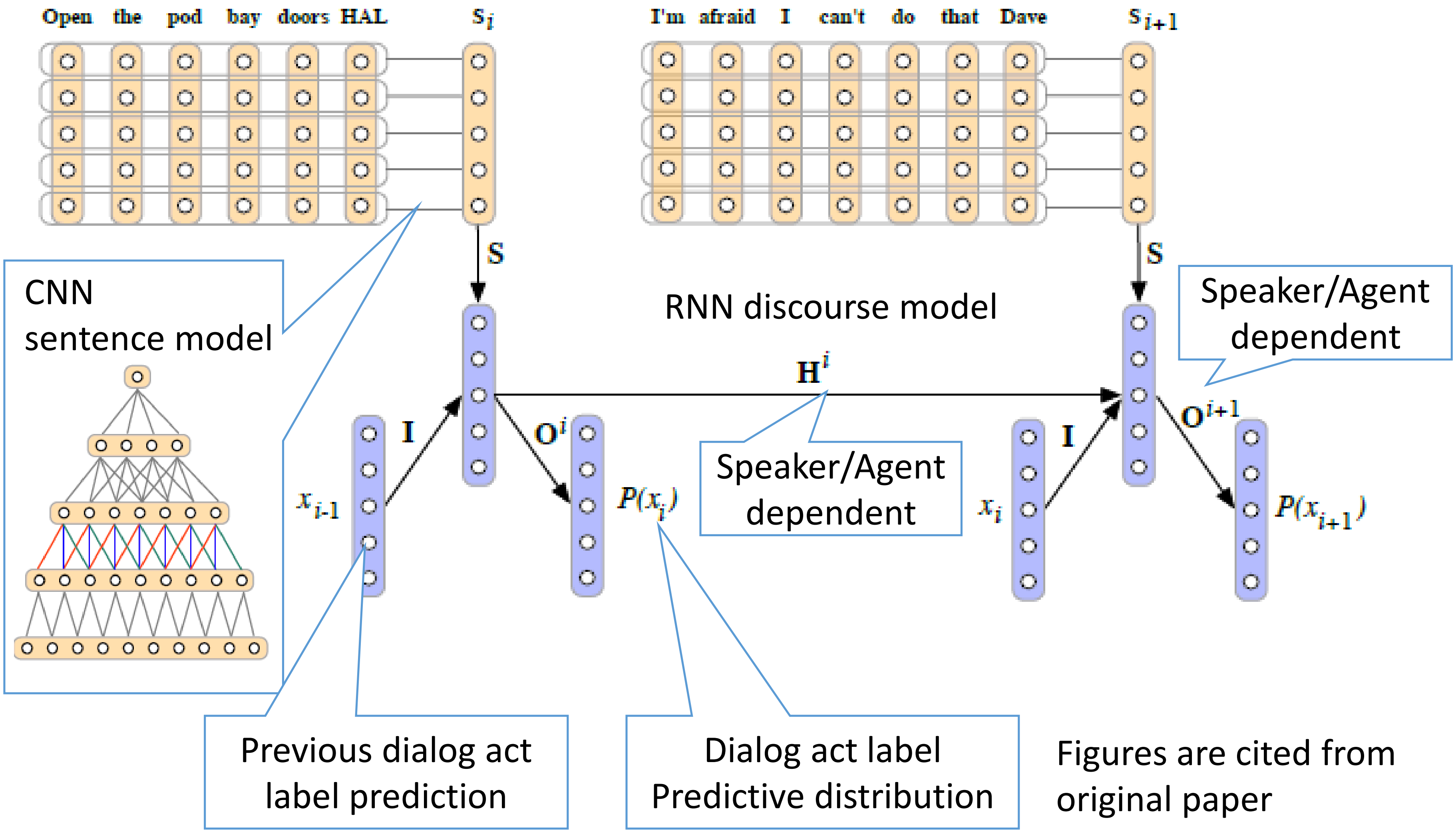
Dialog Act Prediction Task (1)

- Recurrent Convolutional Neural Networks for Discourse Compositionality [Kalchbrenner+ 2013]

Dialogue Act Label	Example
Statement	*And, uh, it's a legal firm office.*
Backchannel/Acknowledge	*Yeah, anything could happen.*
Opinion	*I think that would be great.*
Abandoned/Uninterpretable	*So, -*
Agreement/Accept	*Yes, exactly.*
Appreciation	*Wow.*
Yes – No – Question	*Is that what you do?*
Non – Verbal	*[Laughter], [Throat-clearing]*

Figures are cited from original paper

Dialog Act Prediction Task (2) [Kalchbrenner+ 2013]



Figures are cited from original paper

Dialog Act Prediction Task (3) [Kalchbrenner+ 2013]

- A dialogue and its nearest neighbors over discourse vector
  - semantically different
  - pragmatically similar

Center Dialogue	A: *Do you repair your own car?* B: *I try to, whenever I can.*
First NN	A: *Do you do it every day?* B: *I try to every day.*
Second NN	A: *Well, do you have any children?* B: *I've got one.*
Third NN	A: *Do you manage the money?* B: *Well, I, we talk about it.*
Fourth NN	A: *Um, do you watch it every Sunday?* B: *[Breathing] Uh, when I can.*

	Accuracy (%)
RCNN	**73.9**
LM-HMM trigram	71.0
LM-HMM bigram	70.6
LM-HMM unigram	68.2
Majority baseline	31.5
Random baseline	2.4

Figures are cited from original paper

RNN-based Natural Language Generation for Conventional Dialog System (1) [Wen+ 2015]

- Task: Mapping an abstract dialogue act into an appropriate surface text
- Data: dialogues collected in a user trial of a statistical dialogue manager

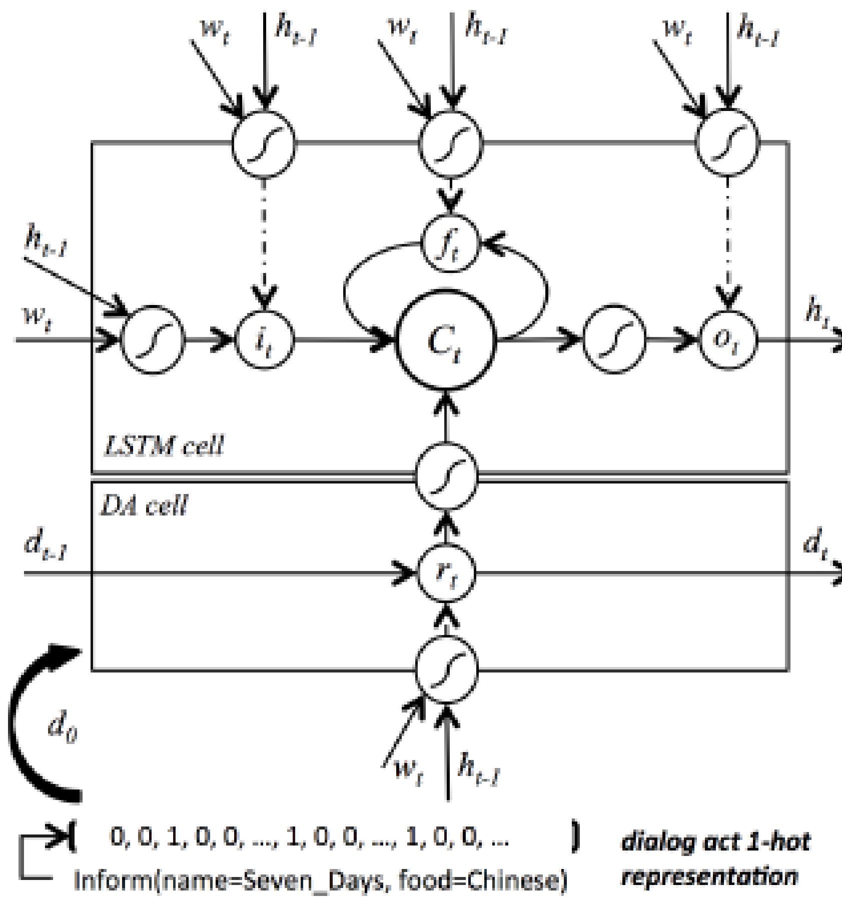
Example Dialogue Acts and Realizations from SF Restaurant Domain
inform(name="red door cafe", goodformeal="breakfast", area="cathedral hill", kidsallowed="no")
red door cafe is a good restaurant for breakfast in the area of cathedral hill and does not allow children .
red door cafe is a good restaurant for breakfast in the cathedral hill area and does not allow children .
red door cafe is a good restaurant for breakfast in the cathedral hill area and does not allow kids .
red door cafe is good for breakfast and is in the area of cathedral hill and does not allow children .
red door cafe does not allow kids and is in the cathedral hill area and is good for breakfast .

Figures are cited from original paper

RNN-based Natural Language Generation for Conventional Dialog System (2) [Wen+ 2015]

- Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems
  - Additional gated memory for dialog act into LSTM
- Real user assessment

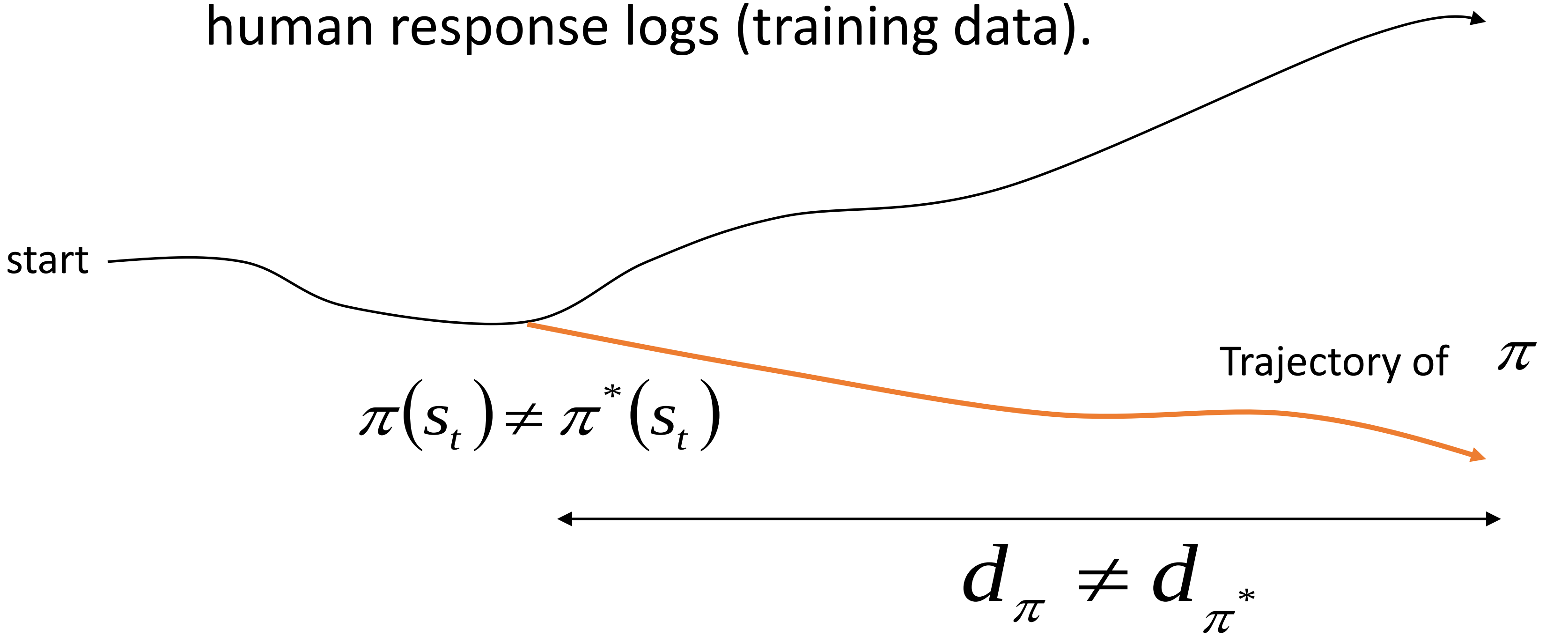
Method	Informativeness	Naturalness
+deep	2.58	**2.51**
sc-lstm	**2.59**	2.50
rnn w/	2.53	2.42*
classlm	2.46**	2.45



Figures are cited from original paper

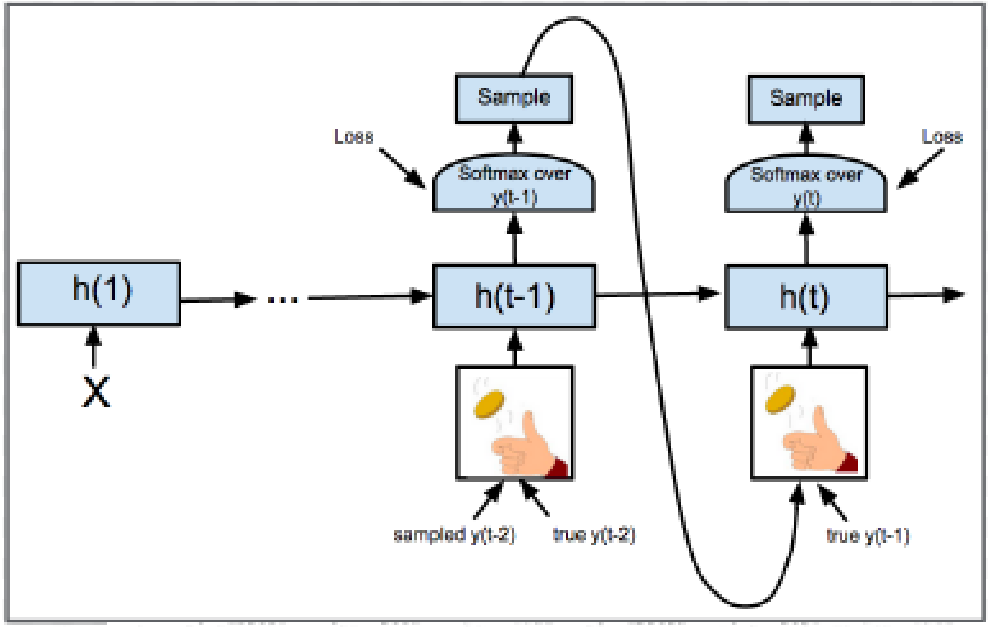
Discussion (1)

- Currently they do not care about error propagation
  - Once a dialog system returns wrong response, it have to handle states not observed in human response logs (training data).



Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks [Samy Bengio+, 2015]

- Randomly select the true previous token or an estimated token coming from the model itself.
- However, it is not trivial to get true human utterance after erroneous system utterance.



Figures are cited from original paper

References

- Neural dialog models
  - Oriol Vinyals and Quoc Le.
    A neural conversational model. In Deep Learning Workshop (ICML 2015).
  - Iulian V. Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, Joelle Pineau.
    Hierarchical Neural Network Generative Models for Movie Dialogues. In arxiv 1507.04808.
  - Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Meg Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan.
    A neural network approach to context-sensitive generation of conversational responses. In NAACL-HLT 2015.
  - Lifeng Shang, Zhengdong Lu, and Hang Li.
    Neural Responding Machine for Short Text Conversation. In ACL-IJCNLP 2015.
- Response Selection
  - Ryan Lowe, Nissan Pow, Iulian Serban, Joelle Pineau
    The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems. In SIGDIAL 2015.
- Dialog Act Prediction
  - N Kalchbrenner, and P Blunsom.
    Recurrent convolutional neural networks for discourse compositionality. In arxiv 1306.3584.
- Natural Language Generation for Conventional Dialog System
  - Tsung-Hsien Wen, Milica Gasic, Nikola Mrksic, Pei-Hao Su, David Vandyke, Steve Young.
    Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems. In EMNLP 2015.
- Imitation Learning Issue
  - Samy Bengio, Oriol Vinyals, Navdeep Jaitly, Noam Shazeer.
    Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks. In NIPS 2015

Other related work

- Matthew Henderson, Blaise Thomson, and Steve Young.
  Word-based dialog state tracking with recurrent neural networks. In SIGDIAL 2014
- P-H, Su, D. Vandyke, M. Gasic, D. Kim, N Mrksic, T-H Wen and S. Young
  Learning from Real Users: Rating Dialogue Success with Neural Networks for Reinforcement Learning in Spoken Dialogue Systems. In Interspeech 2015