UNIVERSITY OF CAMBRIDGE

Neural Language Generation for Spoken Dialogue Systems

Tsung-Hsien (Shawn) Wen and Steve Young

Dialogue Systems Group

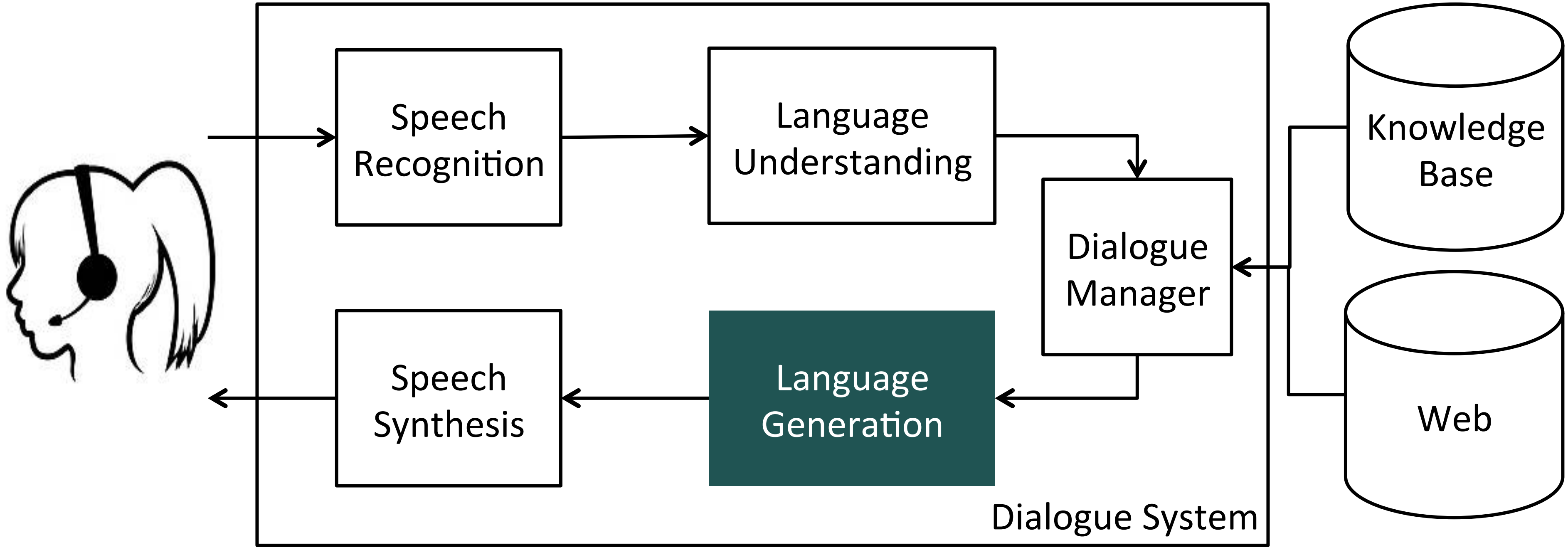
Outline

- Intro
- RNN Generator
- Semantically Conditioned LSTM
- Experiments
- Adaptation – A preliminary work
- Conclusion

Outline

- **Intro**
- RNN Generator
- Semantically Conditioned LSTM
- Experiments
- Adaptation – A preliminary work
- Conclusion

Spoken Dialogue System



NLG: Problem Definition

- Given a meaning representation, map it into natural language utterances.

Dialogue Act

Realisations

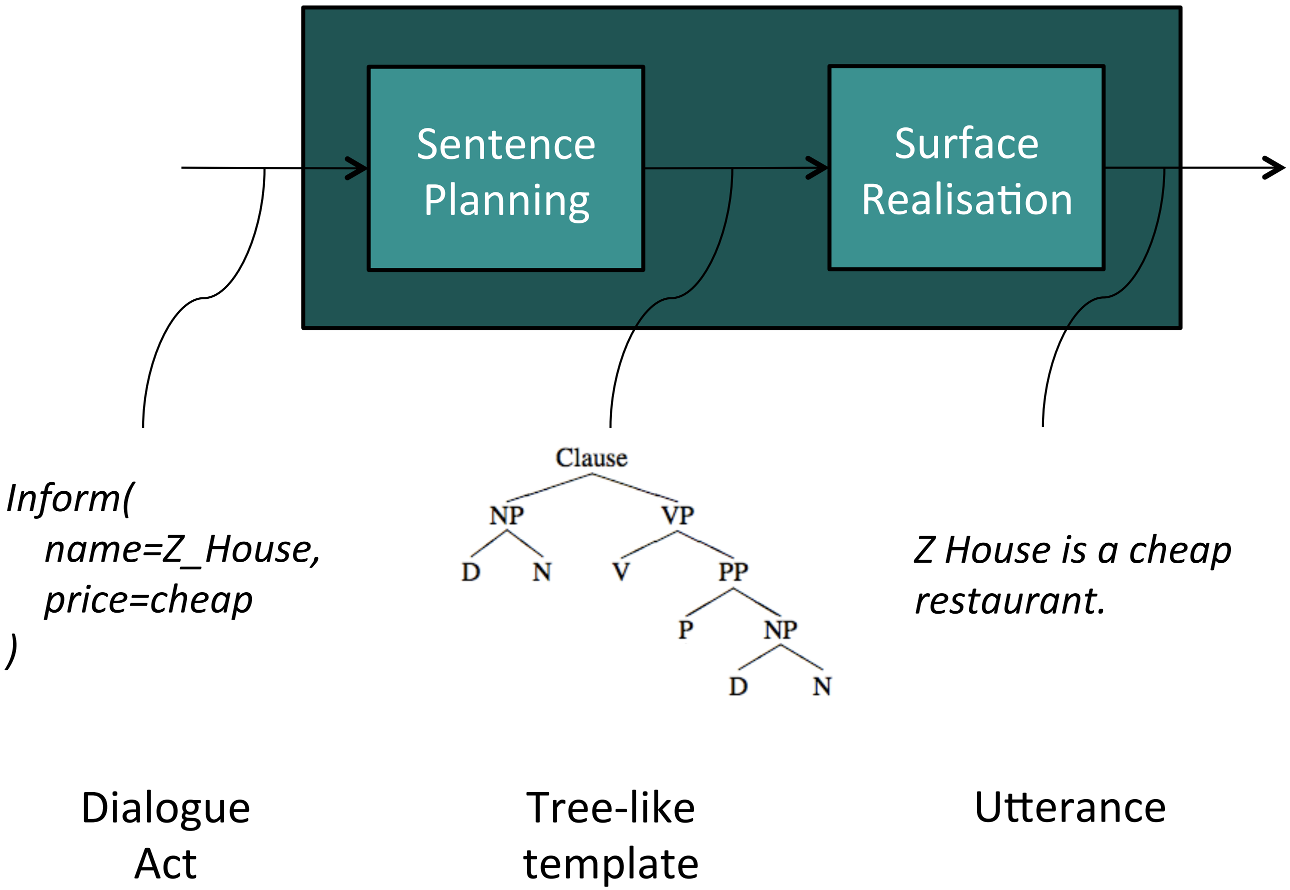
Inform(restaurant=Seven_days, food=Chinese)

Seven days is a restaurant serving Chinese.

Seven days is a Chinese restaurant.

- What do we care about?
  - adequacy, fluency, readability, variation
    (Stent et al 2005)

Traditional pipeline approach



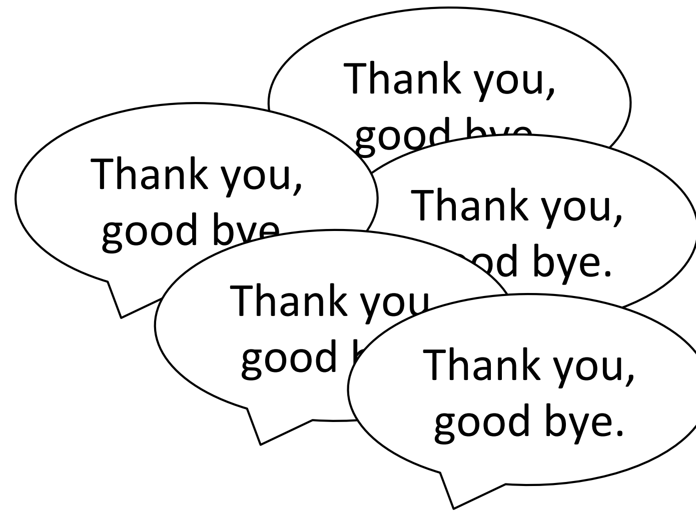
Problems

- Scalability
  - Grammars are handcrafted.
  - Require expert knowledge.

$$
\begin{array}{lllllllll}
A & \rightarrow & mm, & \Pr(0.11) & | & mh, & \Pr(0.67) & | & hh, \quad \Pr(0.22) \\
B & \rightarrow & mm, & \Pr(0.68) & | & hm, & \Pr(0.23) & | & hh, \quad \Pr(0.09) \\
C & \rightarrow & mm, & \Pr(0.58) & | & hm, & \Pr(0.42) & & \\
 & & & & & & & & \\
T & \rightarrow & hQA, & \Pr(0.12) & | & hQB, & \Pr(0.18) & | & APm, \quad \Pr(0.16) \\
 & & & & | & BPh, & \Pr(0.39) & | & hOm, \quad \Pr(0.15) \\
U & \rightarrow & ARC, & \Pr(0.13) & | & BRB, & \Pr(0.44) & | & BRC, \quad \Pr(0.36) \\
 & & & & | & CRC, & \Pr(0.07) & & \\
V & \rightarrow & ARA, & \Pr(0.16) & | & ARB, & \Pr(0.66) & | & CRB, \quad \Pr(0.08) \\
 & & & & | & hQA, & \Pr(0.10) & & \\
W & \rightarrow & BRA, & \Pr(0.10) & | & CRA, & \Pr(0.08) & | & CRB, \quad \Pr(0.07) \\
 & & & & | & X]\{X, & \Pr(0.75) & & \\
 & & & & & & & & \\
R & \rightarrow & lWm, & \Pr(0.14) & | & mWm, & \Pr(0.22) & | & mWh, \quad \Pr(0.23) \\
 & & & & | & hWm, & \Pr(0.17) & | & hWh, \quad \Pr(0.24)\, ' \\
Q & \rightarrow & AVh, & \Pr(0.28) & | & BVm, & \Pr(0.55) & | & BVh, \quad \Pr(0.06) \\
 & & & & | & CVh, & \Pr(0.10) & & \\
P & \rightarrow & lUB, & \Pr(0.14) & | & mUC, & \Pr(0.22) & | & hUA, \quad \Pr(0.20) \\
 & & & & | & hUC, & \Pr(0.44) & & \\
O & \rightarrow & ATA, & \Pr(0.86) & | & CTC, & \Pr(0.14) & & \\
 & & & & & & & & \\
X & \rightarrow & xX, & \Pr(0.35) & | & \epsilon, & \Pr(0.65) & & \\
S & \rightarrow & [XTX\}, & \Pr(1.00) & & & & &
\end{array}
$$

Problems

- Boring
  - Frequent repetition of outputs.
  - Non-colloquial, awkward utterances.

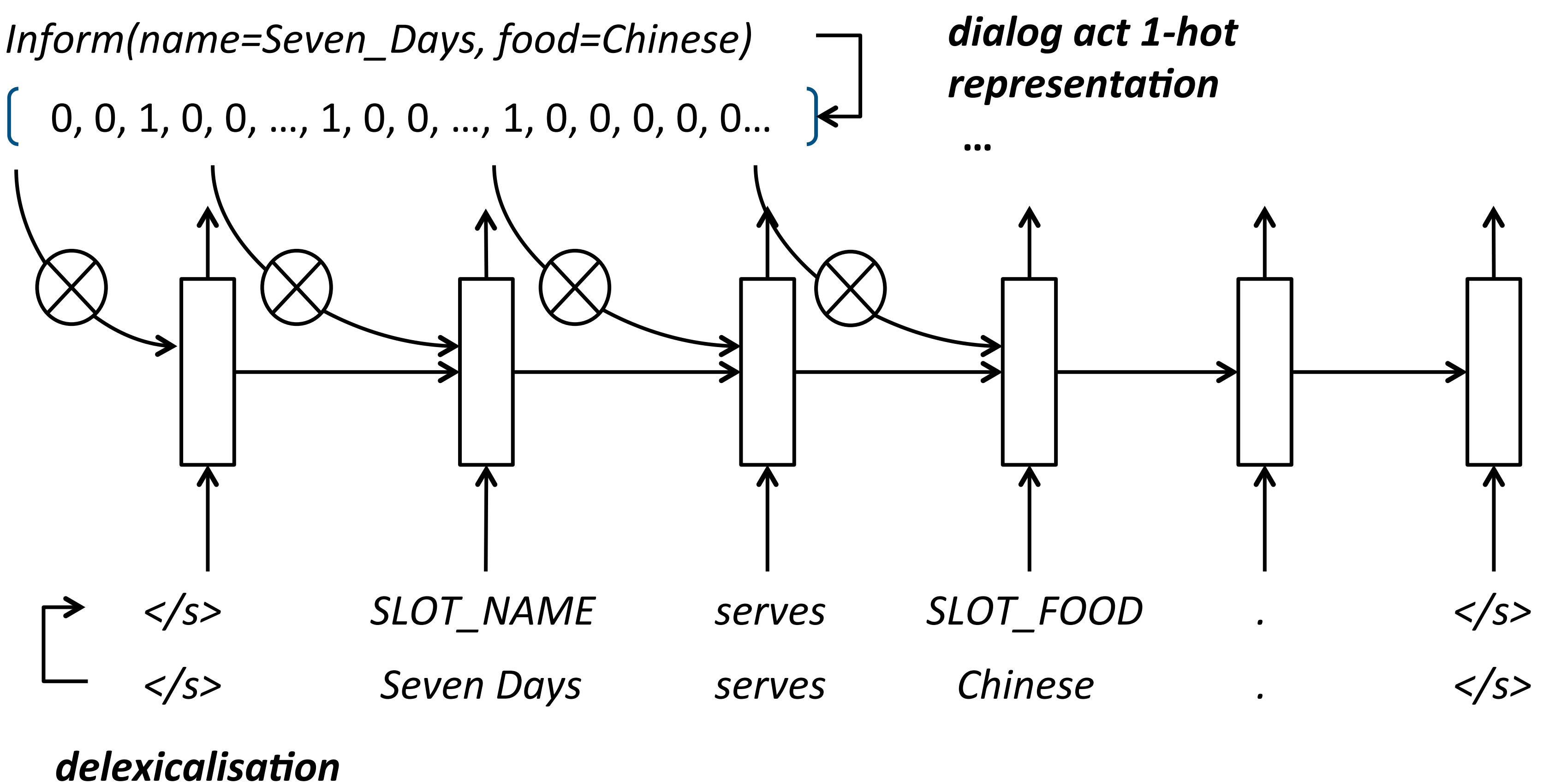


Seven Days is a nice restaurant in the expensive price range, in the north part of the town, if you don't care about what food they serve.

Outline

- Intro
- **RNN Generator**
- Semantically Conditioned LSTM
- Experiments
- Adaptation – A preliminary work
- Conclusion

Recurrent Generation Model



RNNLM (Mikolov et al 2010)

Recurrent Generation Model

- Gates are controlled by exact matching of features and generated tokens.
- Apply a decay factor δ<1 on feature values.

- Binary slots/special values need to be additionally handled.

Over-generation & Reranking

- Generate a bunch of candidate utterances.
- Rerank them!

Seven days is a good restaurant in the south.	0.9
There is no restaurant in the south.	0.2
Seven days is in the south part of town.	0.7

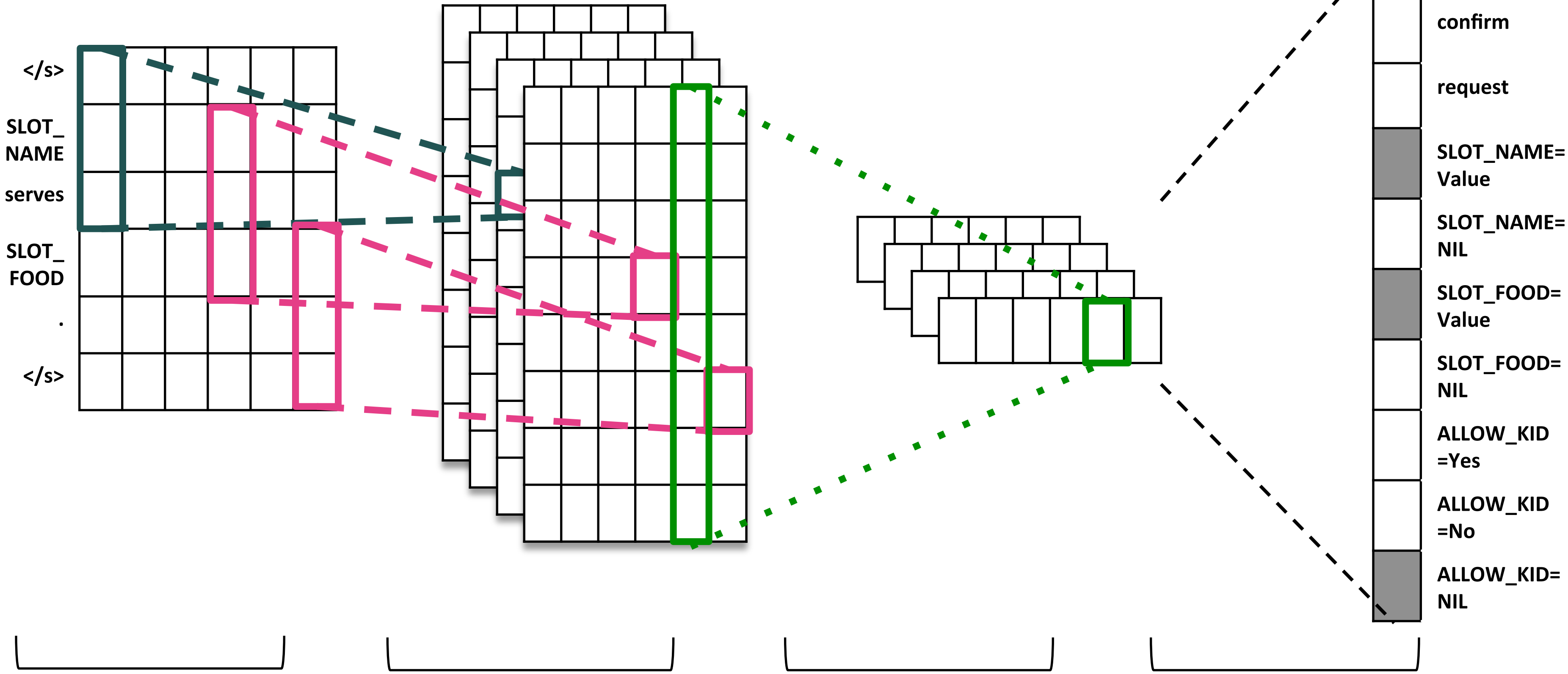
- Simple & variation included.

(Oh & Rudnicky 2000)

CNN Semantic Reranker

Target dialogue act: *inform(name=Seven_days, food=Chinese)*
Generated candidate: *</s> SLOT_NAME serves SLOT_FOOD . </s>*

(Kalchbrenner et al., 2014)



Sentence representation over delexicalised corpus

1-D convolutional layer with multiple feature maps

Average pooling over time

Fully connected layer for classifying dialogue act

Backward Reranker

- Train a RNN with utterances reversed.
  - In order to consider backward context
  - Ex. *"Seven Days is an exceptional restaurant."*
- Reranking Score:
  - $LL_{FowardRNN} + LL_{BackwardRNN} - Loss_{CNN}$

Generated Examples

inform_only(name='red victorian bed breakfast',accepts_credit_cards='yes',near='haight',has_internet='yes')

red victorian bed breakfast is the only hotel near haight and accepts credit cards and has internet .

red victorian bed breakfast is the only hotel near haight and has internet and accepts credit cards .

red victorian bed breakfast is the only hotel near haight that accept credit cards and offers internet .

the red victorian bed breakfast has internet and near haight , it does accept credit cards .

the red victorian bed breakfast is the only hotel near haight that accepts credit cards , and offers internet .

Generated Examples

inform_no_match(area=tenderloin)
there are no restaurants in the tenderloin area .
there are 0 restaurants in the tenderloin area .
unfortunately there are 0 restaurants in the tenderloin area .
i could not find any restaurants in tenderloin .
there are no restaurants in tenderloin .

Generated Examples

?request(food)

what type of food are you looking for ?

what type of food would you like ?

what kind of food are you looking for ?

what type of food are you interested in ?

what kind of food would you like ?

Outline

- Intro
- RNN Generator
- **Semantically Conditioned LSTM**
- Experiments
- Adaptation – A preliminary work
- Conclusion

SC-LSTM

- Original LSTM cell

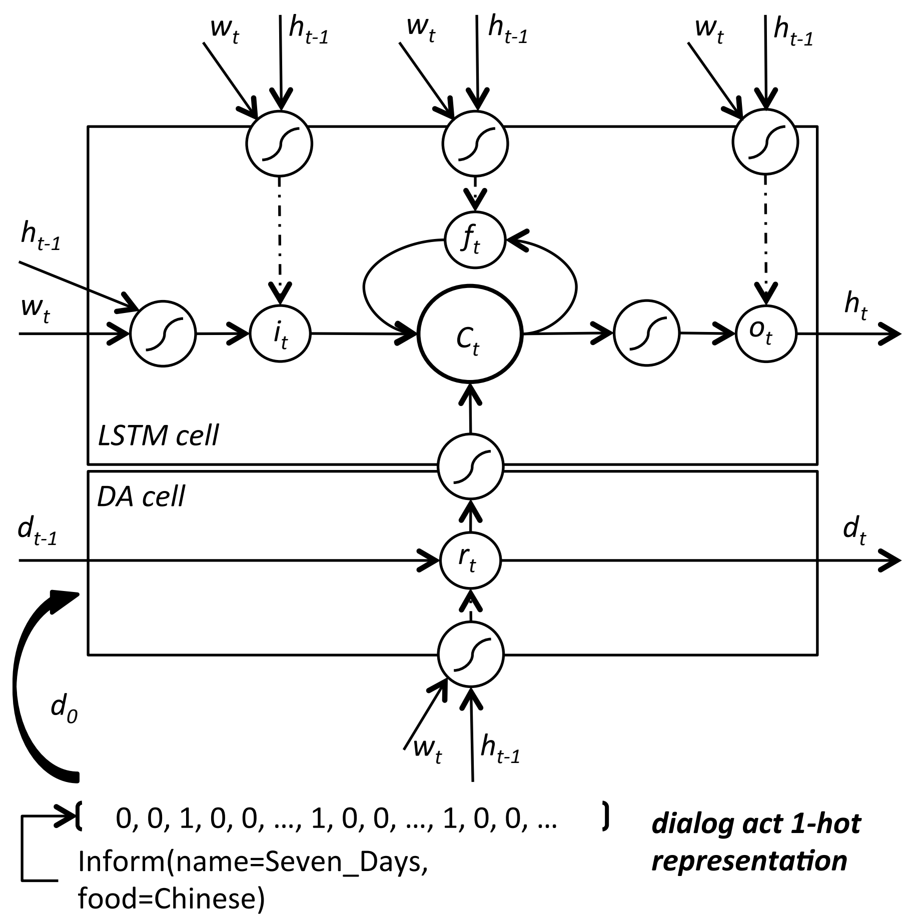
$$\mathbf{i}_t = \sigma(\mathbf{W}_{wi}\mathbf{w}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1})$$
$$\mathbf{f}_t = \sigma(\mathbf{W}_{wf}\mathbf{w}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1})$$
$$\mathbf{o}_t = \sigma(\mathbf{W}_{wo}\mathbf{w}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1})$$
$$\hat{\mathbf{c}}_t = \tanh(\mathbf{W}_{wc}\mathbf{w}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1})$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \hat{\mathbf{c}}_t$$
$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t)$$

- DA cell

$$\mathbf{r}_t = \sigma(\mathbf{W}_{wr}\mathbf{w}_t + \mathbf{W}_{hr}\mathbf{h}_{t-1})$$
$$\mathbf{d}_t = \mathbf{r}_t \odot \mathbf{d}_{t-1}$$

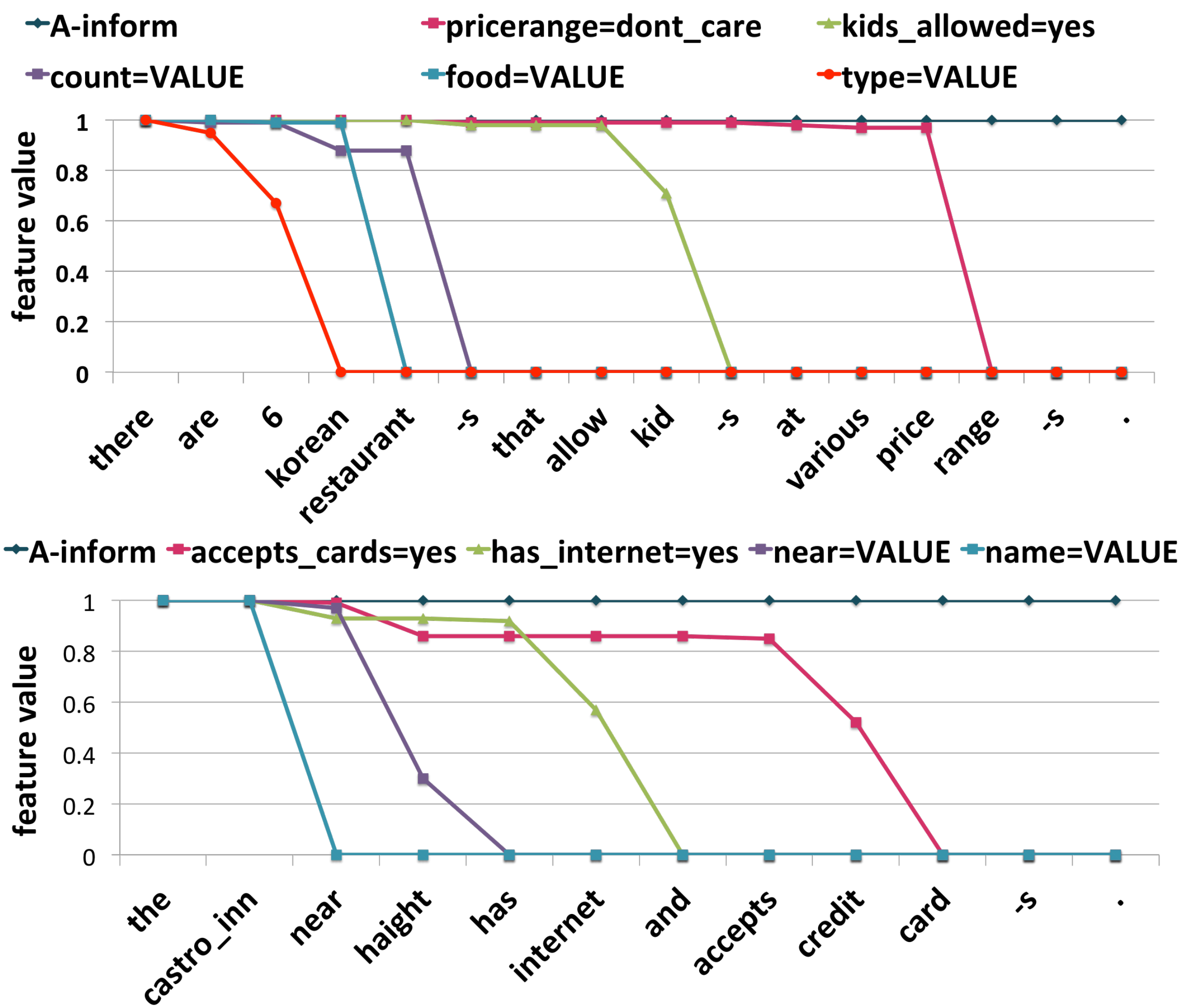
- Modify Ct

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \hat{\mathbf{c}}_t + \tanh(\mathbf{W}_{dc}\mathbf{d}_t)$$



(Hochreiter and Schmidhuber, 1997)

Visualization

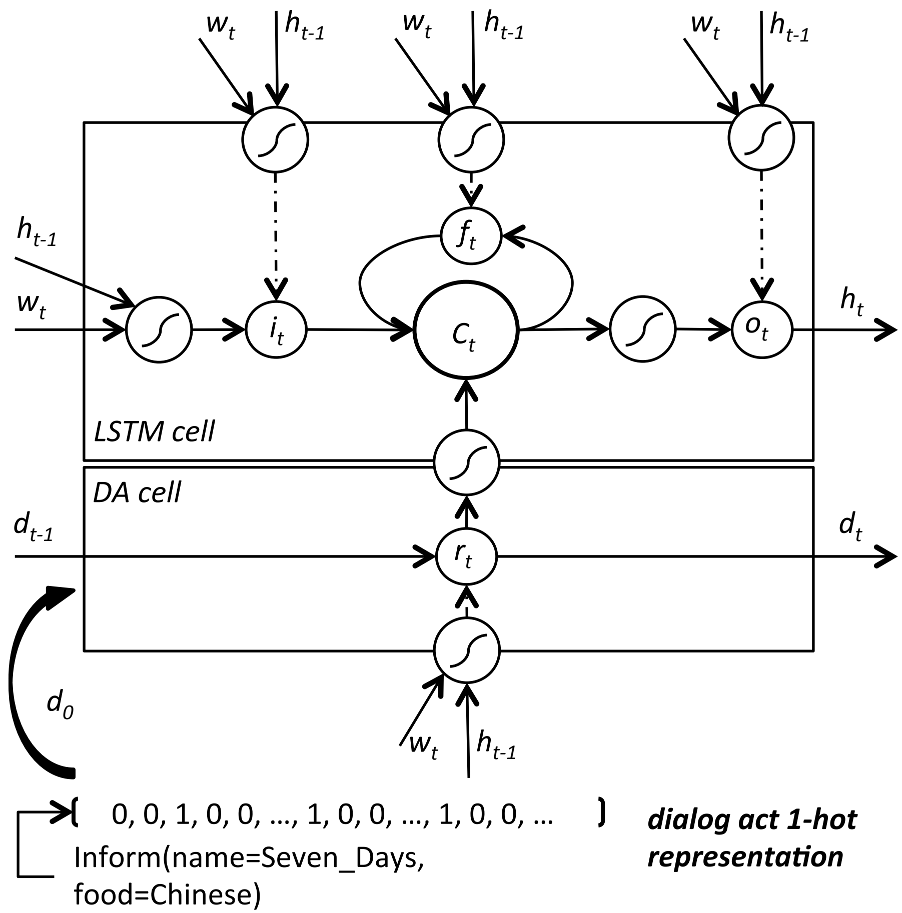


SC-LSTM

- Cost function

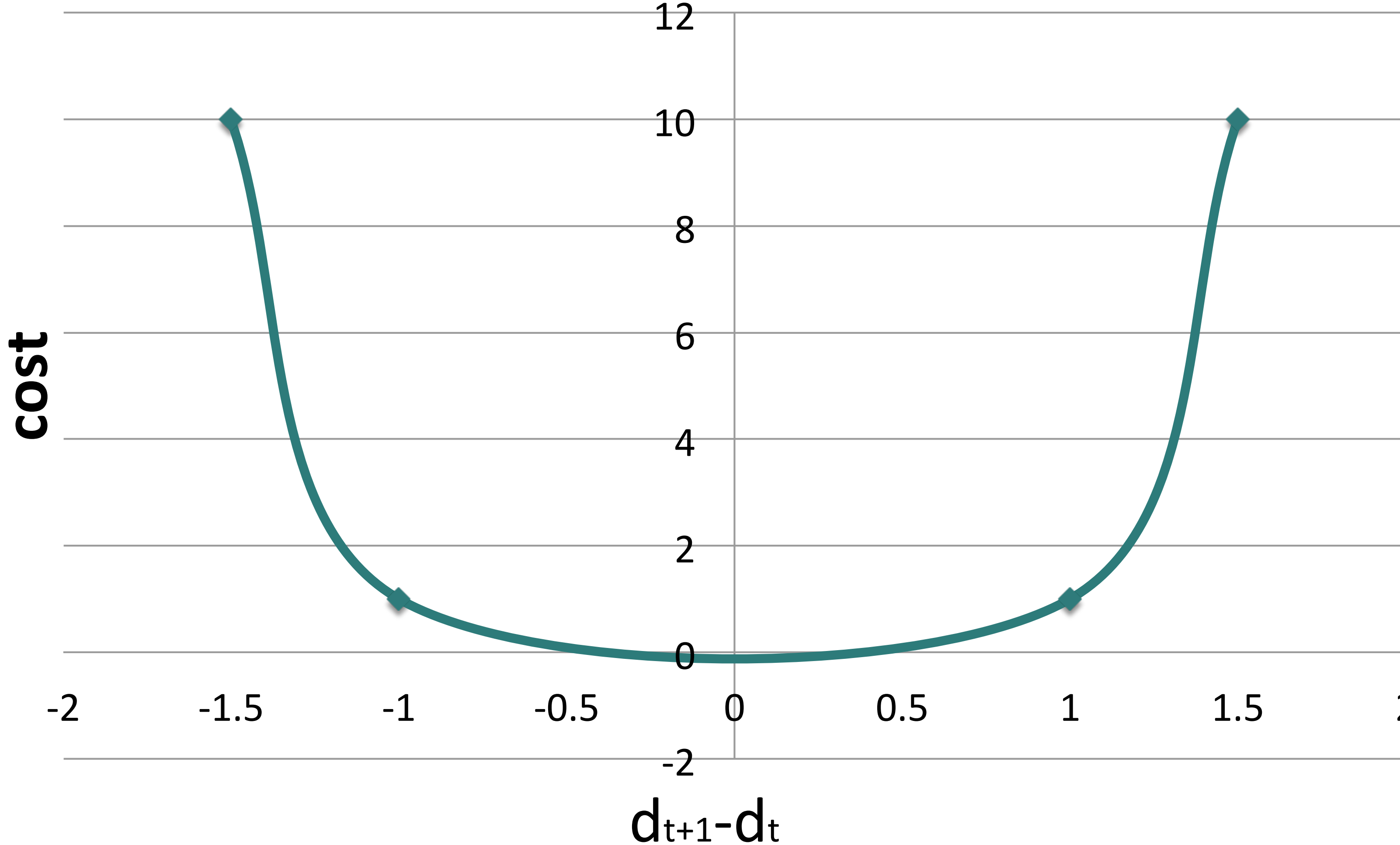
$$F(\theta) = \sum_t \mathbf{p}_t^{\intercal} log(\mathbf{y}_t)$$
$$+ \|\mathbf{d}_T\|$$
$$+ \sum_{t=0}^{T-1} \eta\xi^{\|\mathbf{d}_{t+1} - \mathbf{d}_t\|}$$

- 1st term : Log-likelihood
- 2nd term: make sure rendering all the information needed
- 3rd term: close only one gate each time step.



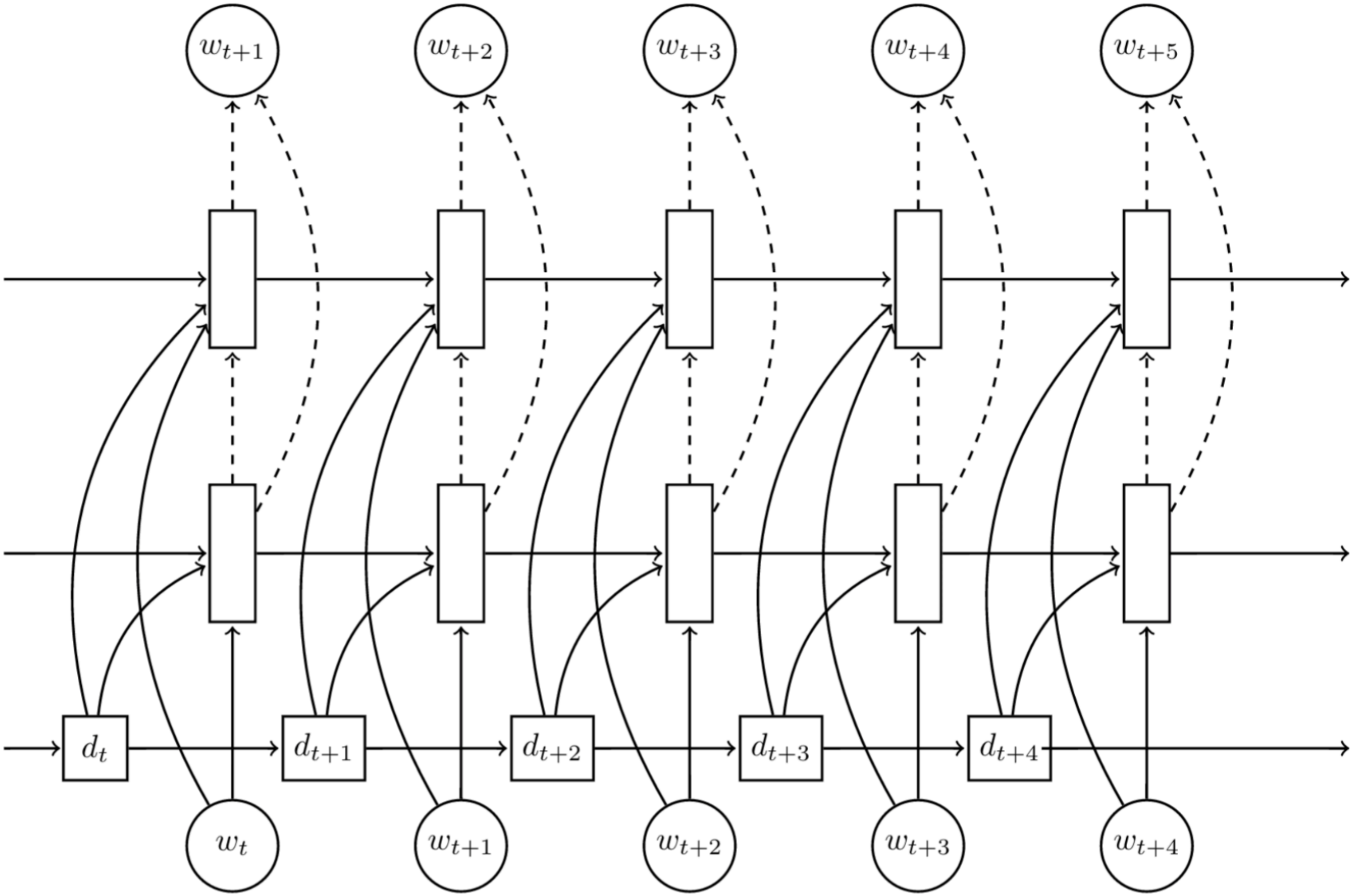
(Hochreiter and Schmidhuber, 1997)

Intuition behind the 3rd term



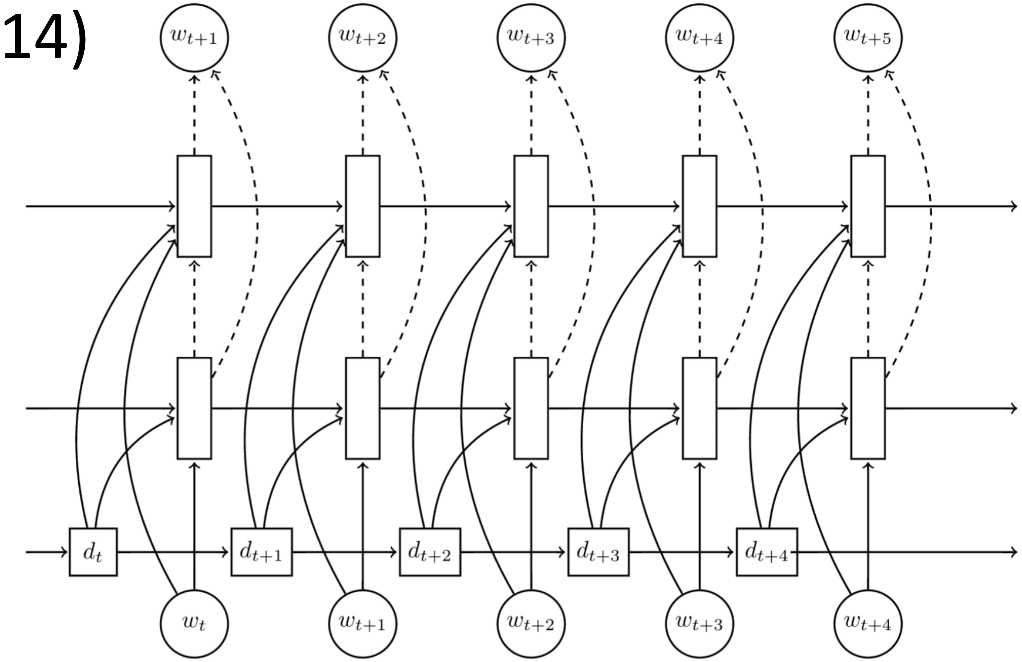
$\eta = 0.01, \xi = 100$

Deep Architecture



Deep Architecture

- Techniques applied
  - Skip connection
    (Graves et al 2013)
  - RNN dropout
    (Srivastava et al 2014)



Outline

- Intro
- RNN Generator
- Semantically Conditioned LSTM
- **Experiments**
- Adaptation – A preliminary work
- Conclusion

Setup

- Data collection:
  - SFX restaurant/hotel domains

Ontologies

	SF Restaurant	SF Hotel
act type	inform, inform_only, reject, confirm, select, request, reqmore, goodbye	
shared	name, type, *pricerange, price, phone, address, postcode, *area, *near	
specific	*food *goodformeal ***kids-allowed**	***hasinternet** ***acceptscards** ***dogs-allowed**

bold=binary slots, *=slots can take "don't care" value

Setup

- Data collection:
  - SFX restaurant/hotel domains
  - Workers recruited from Amazon MT.
  - Asked to generate system responses given a DA.
  - Result in ~5.1K utterances, 228/164 distinct acts.
- Training: BPTT, L2 reg, SGD w/ early stopping.

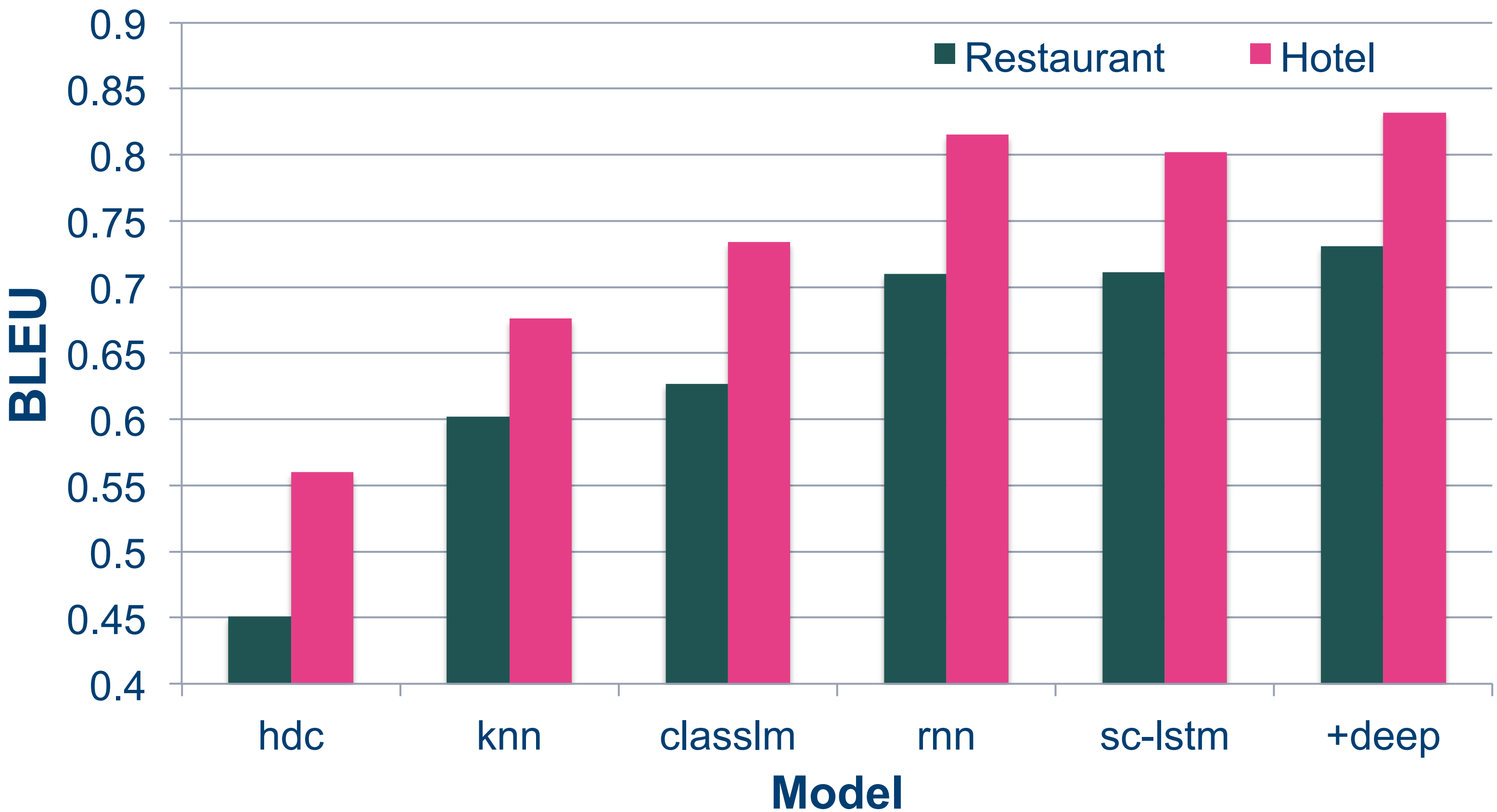
  train/valid/test: 3/1/1, data up-sampling

Available at : https://www.repository.cam.ac.uk/handle/1810/251304

Corpus-based Evaluation

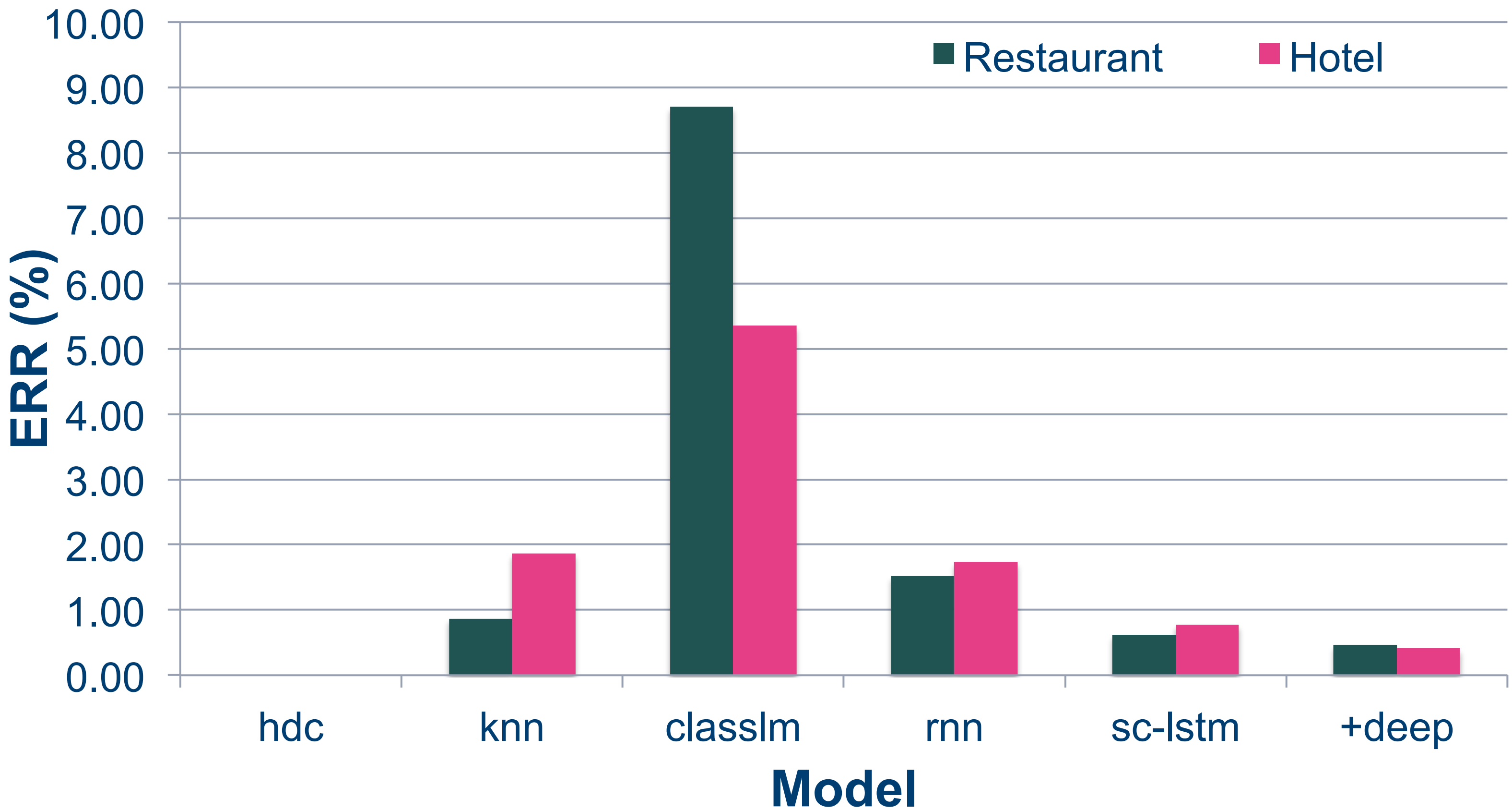
- Test set: ~1K utterances each domain
- Metrics: BLEU-4 (against multiple references), ERR(slot error rates)
- Averaged over 5 random initialised networks.
- Over-gen 20, evaluate on top-5
- Models compared:
  - handcrafted generator (hdc)
  - kNN example-based generator (kNN)
  - class-based LM generator (classlm, O&R 2000)
  - rnn-based generator (rnn, Wen et al 2015)

Corpus-based Evaluation



Selection scheme : 5/20

Corpus-based Evaluation



Selection scheme : 5/20

Human Evaluation

- Setup
  - Judges (~60) recruited from Amazon MT.
  - Asked to evaluate two system responses pairwise.
  - Comparing *classlm*, *rnn*, *sc-lstm*, and *+deep*

- Metrics:
  - Informativeness, Naturalness (rating out of 3)
  - Preference

Human Evaluation

Method	Informativeness	Naturalness
+deep	2.58	**2.51**
sc-lstm	**2.59**	2.50
rnn	2.53	2.42^{*}
classlm	2.46^{**}	2.45

$^{*}p < 0.05$ $^{**}p < 0.005$

Human Evaluation

Pref.%	classlm	rnn	sc-lstm	+deep
classlm	-	46.0	40.9^{**}	37.7^{**}
rnn	54.0	-	43.0	35.7^{*}
sc-lstm	59.1^{*}	57	-	47.6
+deep	62.3^{**}	64.3^{**}	52.4	-

$^{*}p < 0.05$ $^{**}p < 0.005$

Outline

- Intro
- RNN Generator
- Semantically Conditioned LSTM
- Experiments
- **Adaptation – A preliminary work**
- Conclusion

Attentive Encoder-Decoder

- Embedding

$$\mathbf{z}_i = \mathbf{s}_i + \mathbf{v}_i$$

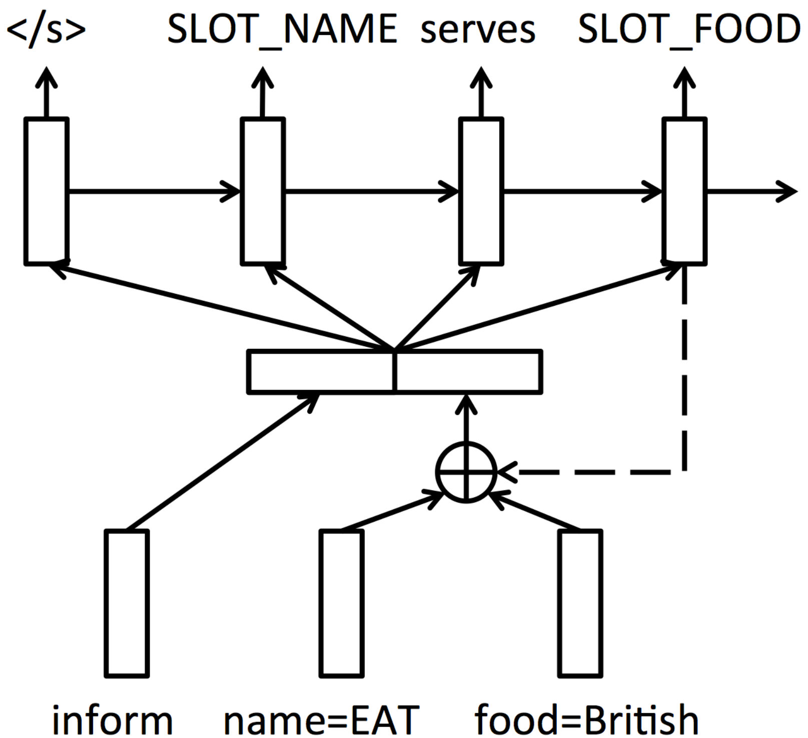
- Attention

$$\beta_{t,i} = \mathbf{q}^\intercal \tanh(\mathbf{W}_{hm}\mathbf{h}_{t-1} + \mathbf{W}_{mm}\mathbf{z}_i)$$

$$\omega_{t,i} = e^{\beta_{t,i}} / \sum_i e^{\beta_{t,i}}$$

$$\mathbf{d}_t = \mathbf{a} \oplus \sum_i \omega_{t,i}\mathbf{z}_i$$

- Generation
  - Typical LSTM



(Mei et al 2015)

Experiments

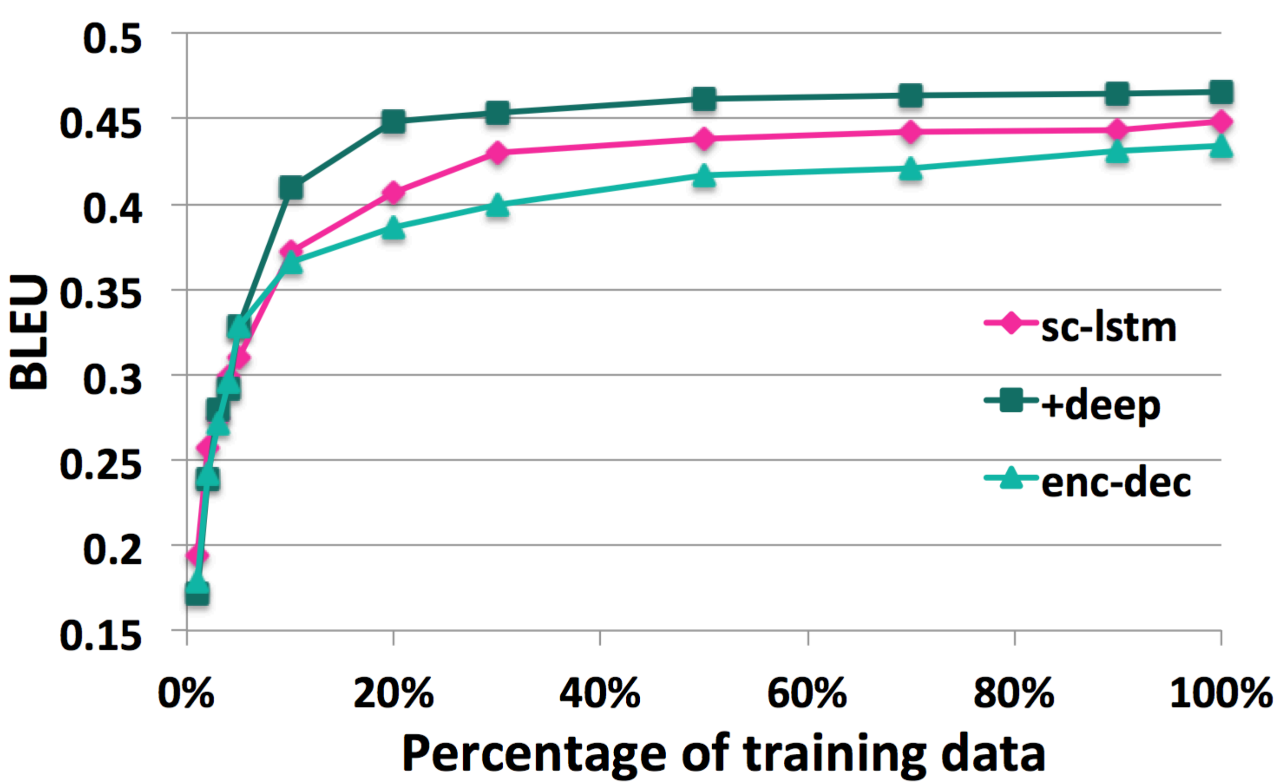
- On new laptop ontology

act type	inform, inform_only_match, inform_no_match, inform_count, inform_all, inform_no_info, recommend, compare, confirm, select, suggest, request, request_more, goodbye
slots	family*, battery_rating*, drive_range*, **is_for_business***, price_range*, weight_range*, warranty, battery, design, dimension, utility, weight, platform, memory, price, drive, processor

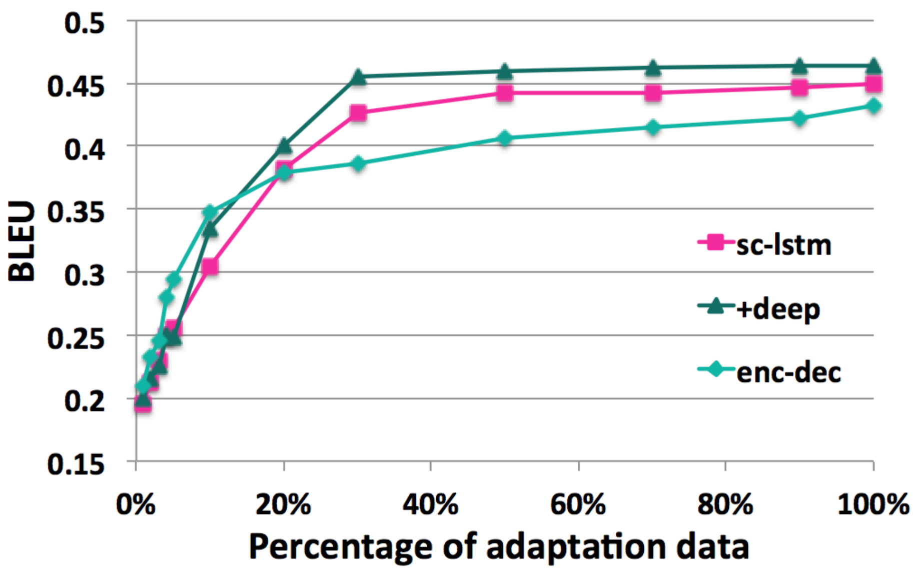
bold=binary slots, *=slots can take don't care value

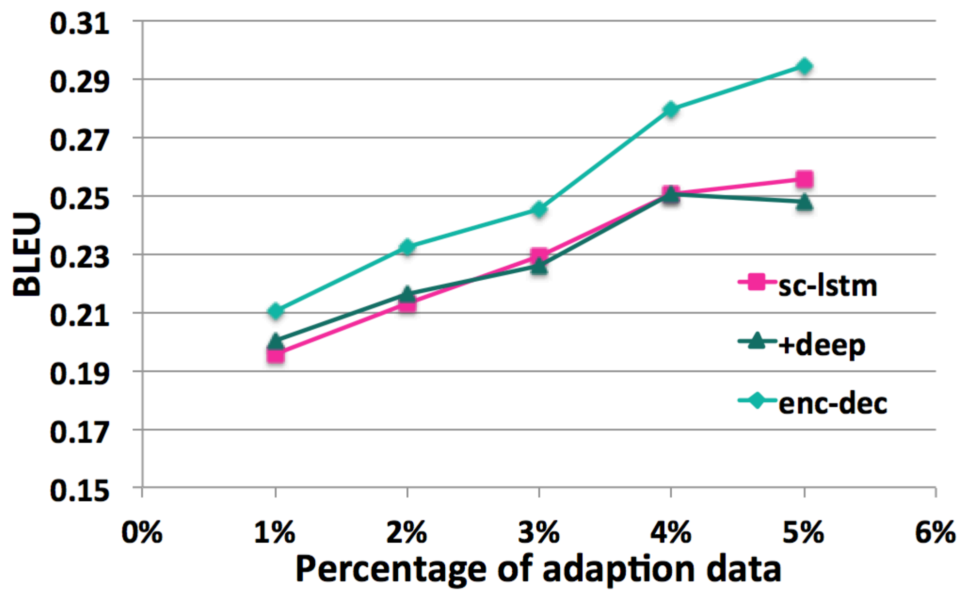
- Comparing performance and adaptation capability with SC-LSTM.

From scratch



Adapt from Rest+Hotel to Laptop





Outline

- Intro
- RNN Generator
- Semantically Conditioned LSTM
- Experiments
- Adaptation – A preliminary work
- **Conclusion**

Conclusion

- NLG can be learned N2N from data.
- Learn LM & slot gating control signal jointly
- Corpus-based/Human evaluation.
- More colloquial, more scalable.
- Potential for open domain SDS.

Papers

- Tsung-Hsien Wen, Milica Gasic , Dongho Kim, Nikola Mrksic, Pei-Hao Su, David Vandyke, and Steve Young. Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking. In *Proceedings of SIGdial 2015*.
- Tsung-Hsien Wen, Milica Gasic , Nikola Mrksic, Pei-Hao Su, David Vandyke, and Steve Young. Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems. In *Proceedings of EMNLP 2015*.
- Tsung-Hsien Wen, Milica Gasic, Nikola Mrksic, Lina M.R. Barahona, Pei-Hao Su, David Vandyke, and Steve Young. Toward Multi-domain Language Generation using Recurrent Neural Networks. To be appear in NIPS Workshop on Machine Learning for SLU and Interaction 2015.

Selected References

- Amanda Stent, Matthew Marge, and Mohit Singhai. 2005. Evaluating evaluation methods for generation in the presence of variation. In Proceedings of CICLing 2005.
- Alice H. Oh and Alexander I. Rudnicky. 2000. Stochastic language generation for spoken dialogue systems. In Proceedings of the 2000 ANLP/NAACL Workshop on Conversational Systems.
- Tomas Mikolov, Martin Karafit, Lukas Burget, Jan Cernocky, and Sanjeev Khudanpur. 2010. Recurrent neural network based language model. *In Proceedings on InterSpeech*.
- Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. Proceedings of the 52nd Annual Meeting of ACL.
- Sepp Hochreiter and Jurgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*.
- Hongyuan Mei, Mohit Bansal, Matthew R. Walter. 2015. What to talk about and how? Selective Generation using LSTMs with Coarse-to-Fine Alignment. arXiv.

UNIVERSITY OF CAMBRIDGE

Thank you! Questions?

Tsung-Hsien Wen is supported by a studentship funded by Toshiba Research Europe Ltd, Cambridge Research Laboratory