How the Context Matters Language & Interaction in Dialogues

國立臺灣大學
National Taiwan University

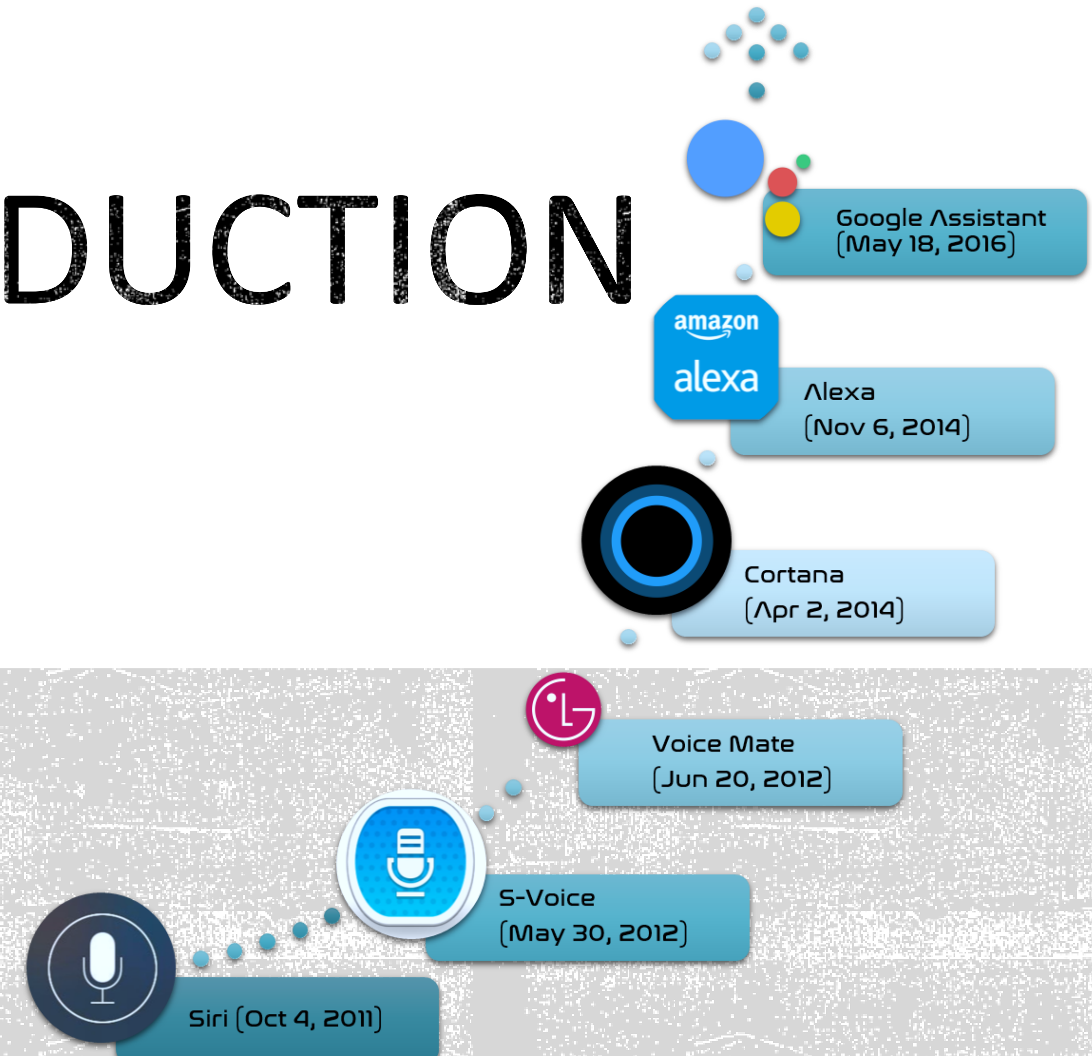
YUN-NUNG (VIVIAN) CHEN

HTTP://VIVIANCHEN.IDV.TW

Outline

- Introduction
- Word-Level Contexts in Sentences
  - Learning from Prior Knowledge –
    Knowledge-Guided **S**tructural **A**ttention **N**etworks (K-SAN) [Chen et al., ‘16]
  - Learning from Observations –
    Modularizing **U**nsupervised **S**ense **E**mbedding (MUSE) [Lee & Chen, ‘17]
- Sentence-Level Contexts in Dialogues
  - Inference –
    Leveraging Behavioral Patterns for Personalized Understanding [Chen et al., ‘15]
  - Investigation of Understanding Impact –
    Reinforcement Learning Based Neural Dialogue System [Li et al., ‘17]
    Misunderstanding Impact [Li et al., ‘17]
- Conclusion

3 INTRODUCTION



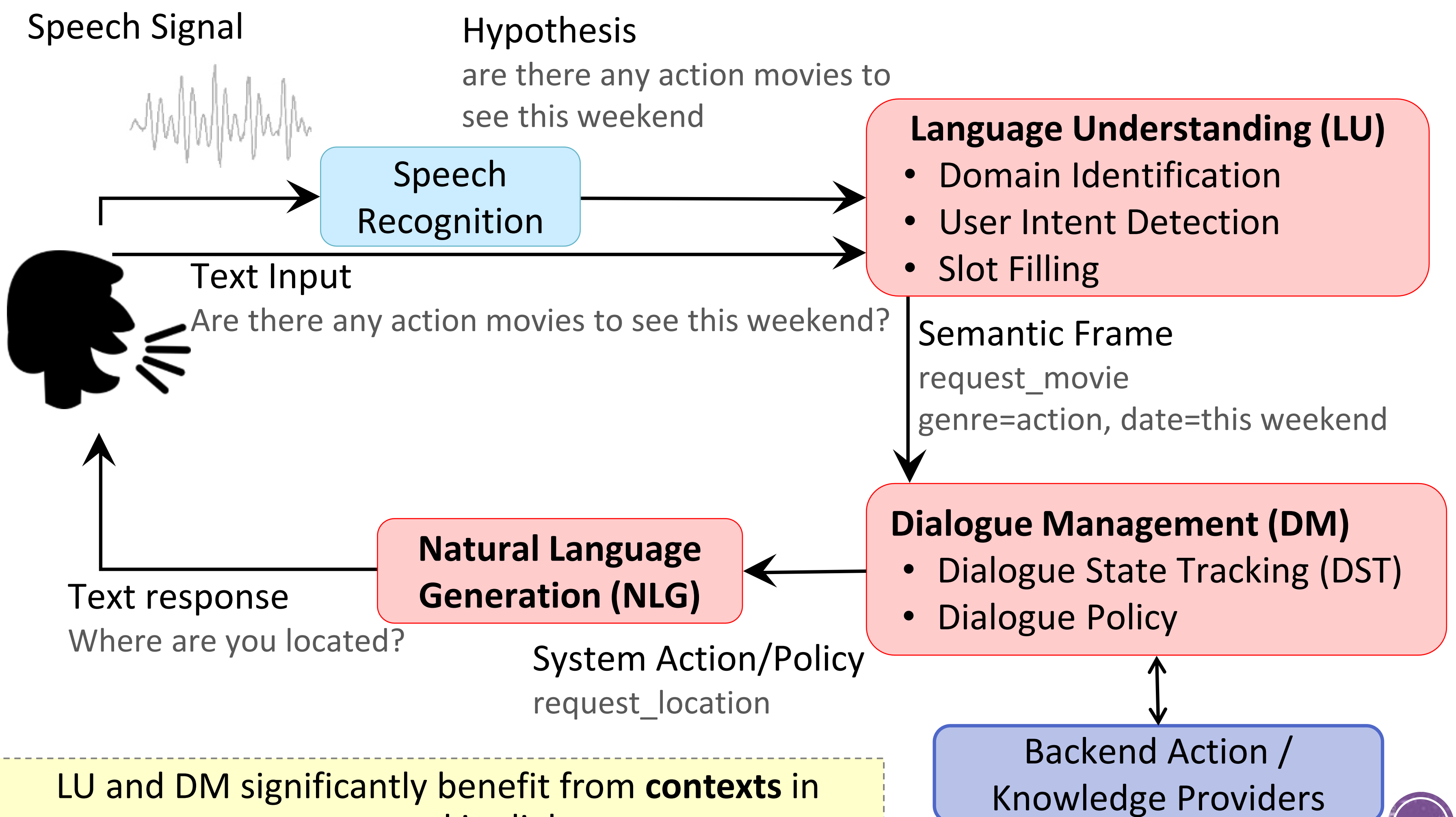
Task-Oriented Dialogue System

- Dialogue systems are intelligent agents that are able to help users finish tasks more efficiently via conversational interactions.
- Dialogue systems are being incorporated into various devices (smart-phones, smart TVs, in-car navigating system, etc).

JARVIS – Iron Man's Personal Assistant

Baymax – Personal Healthcare Companion

Dialogue System Framework



LU and DM significantly benefit from **contexts** in sentences and in dialogues

Context in Language

- Word-level context
  - Prior knowledge such as linguistic syntax

    show me the flights from seattle to san francisco

  - Collocated words

    Smartphone companies including apple, blueberry, and sony will be invited.

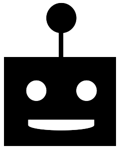
Contexts provide informative cues for better understanding

- Sentence-level context

(browsing movie reviews…)
Find me a good action movie this weekend

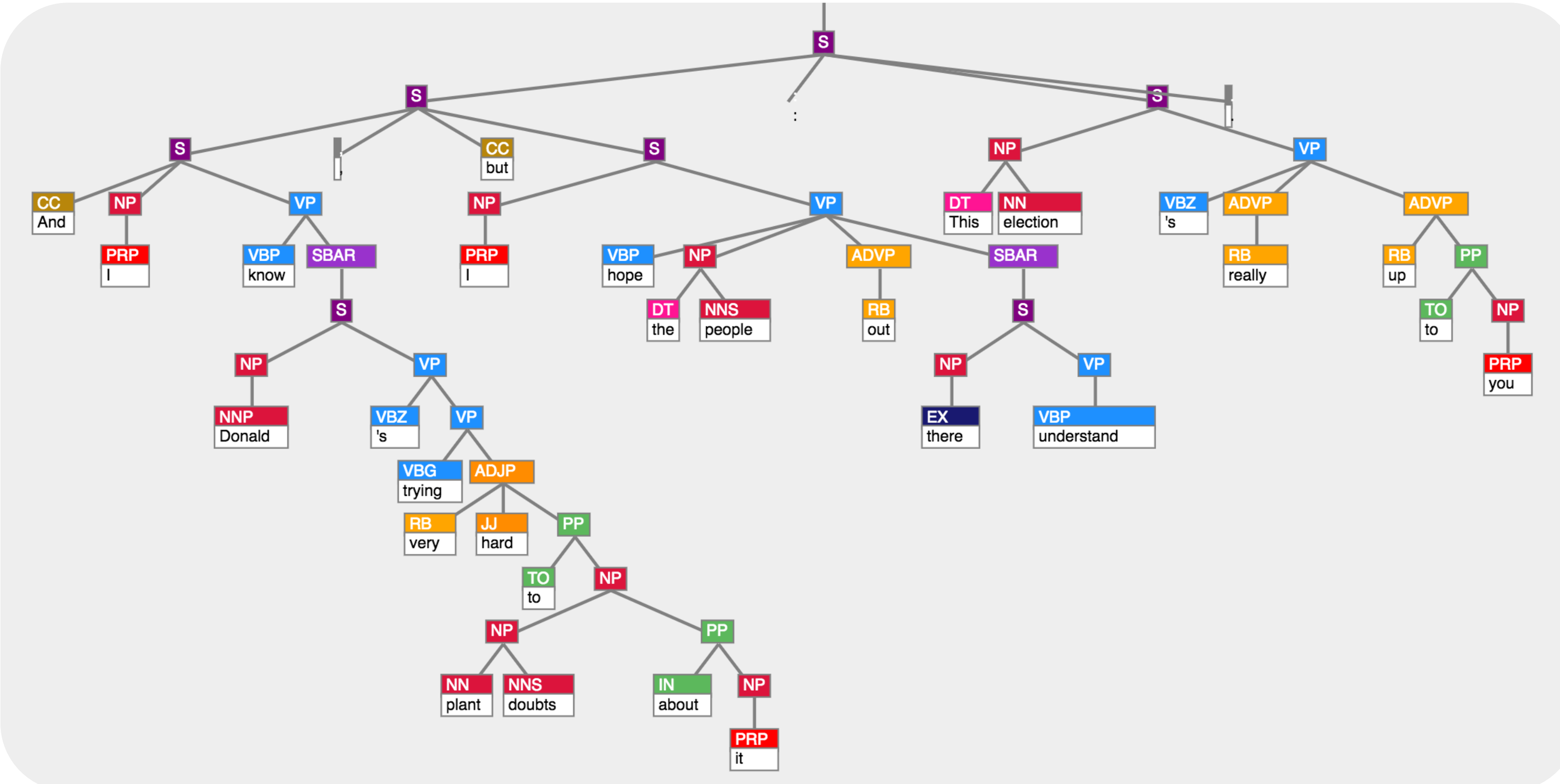
request_movie
(genre=action,
date=this weekend)

London Has Fallen is currently the number 1 action movie in America



How behavioral contexts influences the user intent

How misunderstanding influences the dialogue system performance



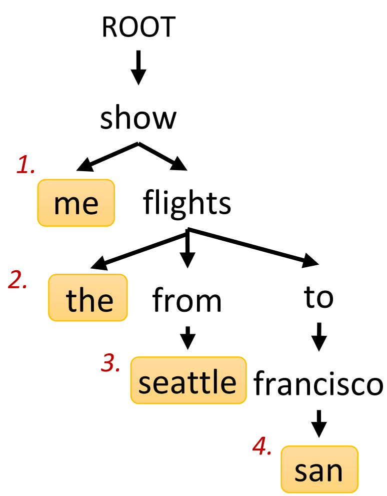
Knowledge-Guided Structural Attention Network (K-SAN)

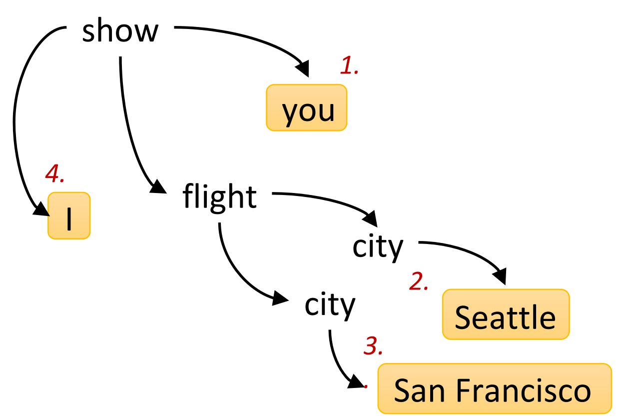
Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

Prior Structural Knowledge

Sentence *s* show me the flights from seattle to san francisco

- Syntax (Dependency Tree)
- Semantics (AMR Graph)



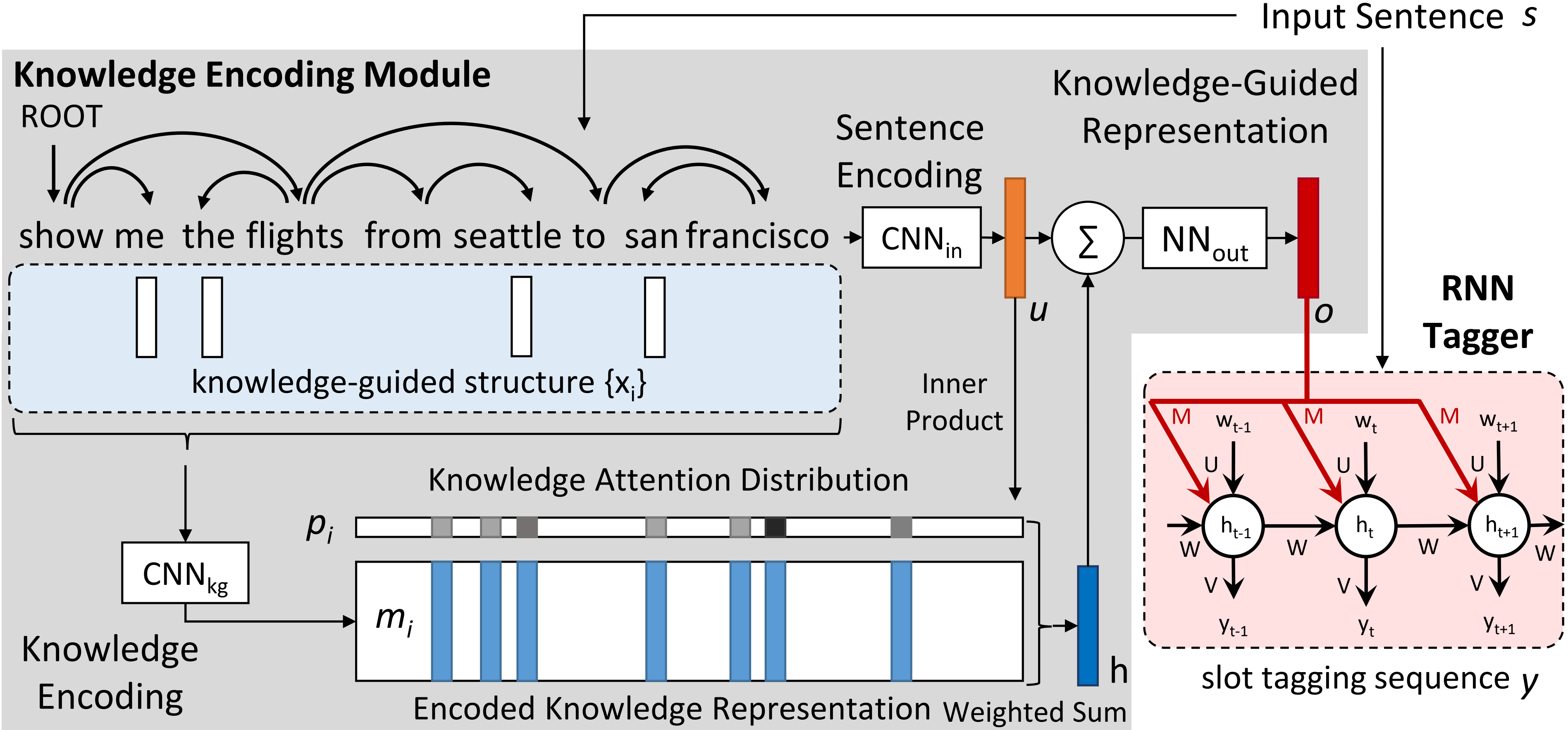


Prior knowledge about syntax or semantics may guide understanding

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

K-SAN: Knowledge-Guided Structural Attention Networks

- Prior knowledge as a teacher

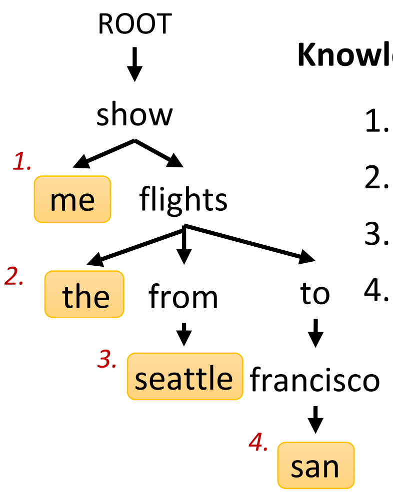


Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

Sentence Structural Knowledge

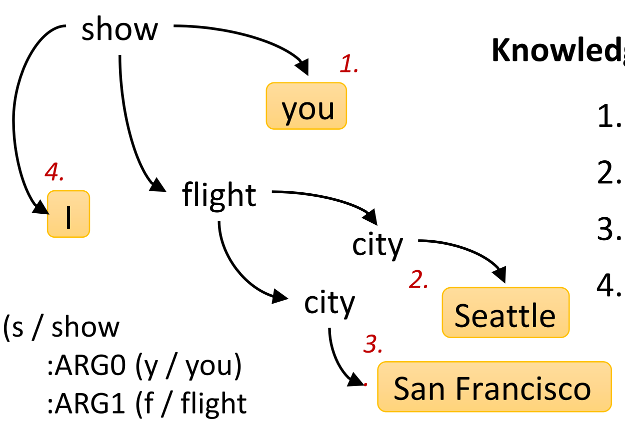
- Syntax (Dependency Tree)
- Semantics (AMR Graph)

Sentence s show me the flights from seattle to san francisco



Knowledge-Guided Substructure x_i

1. show me
2. show flights the
3. show flights from seattle
4. show flights to francisco san



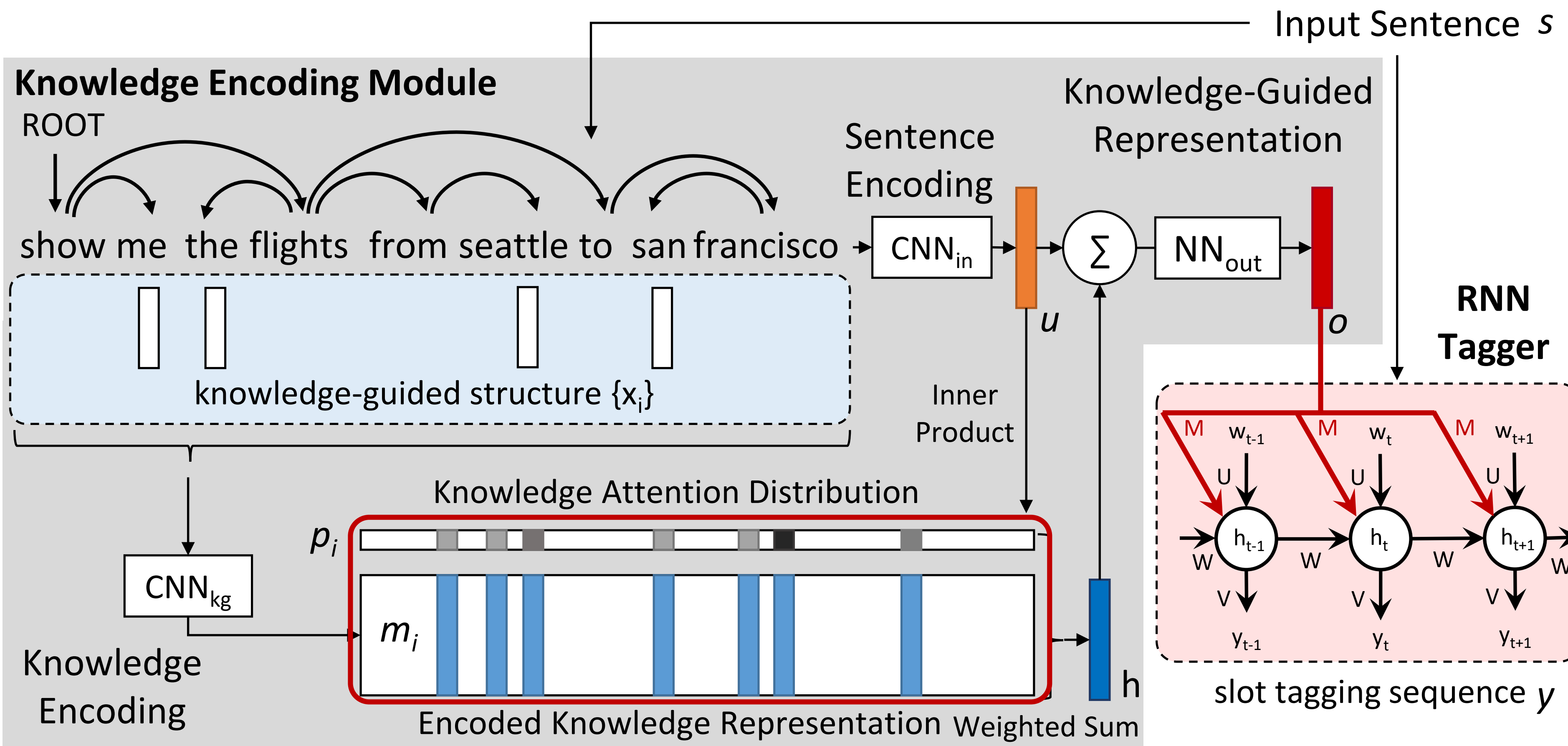
(s / show
  :ARG0 (y / you)
  :ARG1 (f / flight
    :source (c / city
      :name (d / name :op1 Seattle))
    :destination (c2 / city
      :name (s2 / name :op1 San :op2 Francisco)))
  :ARG2 (i / I)
  :mode imperative)

Knowledge-Guided Substructure x_i

1. show you
2. show flight seattle
3. show flight san francisco
4. show i

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, “Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks,” preprint arXiv: 1609.00777, 2016.

Knowledge-Guided Structures



The model will pay more attention to more important substructures that may be crucial for slot tagging.

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

K-SAN Experiments (Chen et al., 2016)

ATIS Dataset (F1 slot filling)	Small (1/40)	Medium (1/10)	Large
Tagger (GRU)	73.83	85.55	93.11
Encoder-Tagger (GRU)	72.79	88.26	94.75

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

K-SAN Experiments (Chen et al., 2016)

ATIS Dataset (F1 slot filling)	Small (1/40)	Medium (1/10)	Large
Tagger (GRU)	73.83	85.55	93.11
Encoder-Tagger (GRU)	72.79	88.26	94.75
K-SAN (Stanford dep)	**74.60**+	87.99	94.86+
K-SAN (Syntaxnet dep)	74.35+	**88.40**+	**95.00**+

Syntax provides richer knowledge and more general guidance when less training data.

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

K-SAN Experiments (Chen et al., 2016)

ATIS Dataset (F1 slot filling)	Small (1/40)	Medium (1/10)	Large
Tagger (GRU)	73.83	85.55	93.11
Encoder-Tagger (GRU)	72.79	88.26	94.75
K-SAN (Stanford dep)	**74.60**+	87.99	94.86+
K-SAN (Syntaxnet dep)	74.35+	**88.40**+	**95.00**+
K-SAN (AMR)	74.32+	88.14	94.85+
K-SAN (JAMR)	74.27+	88.27+	94.89+

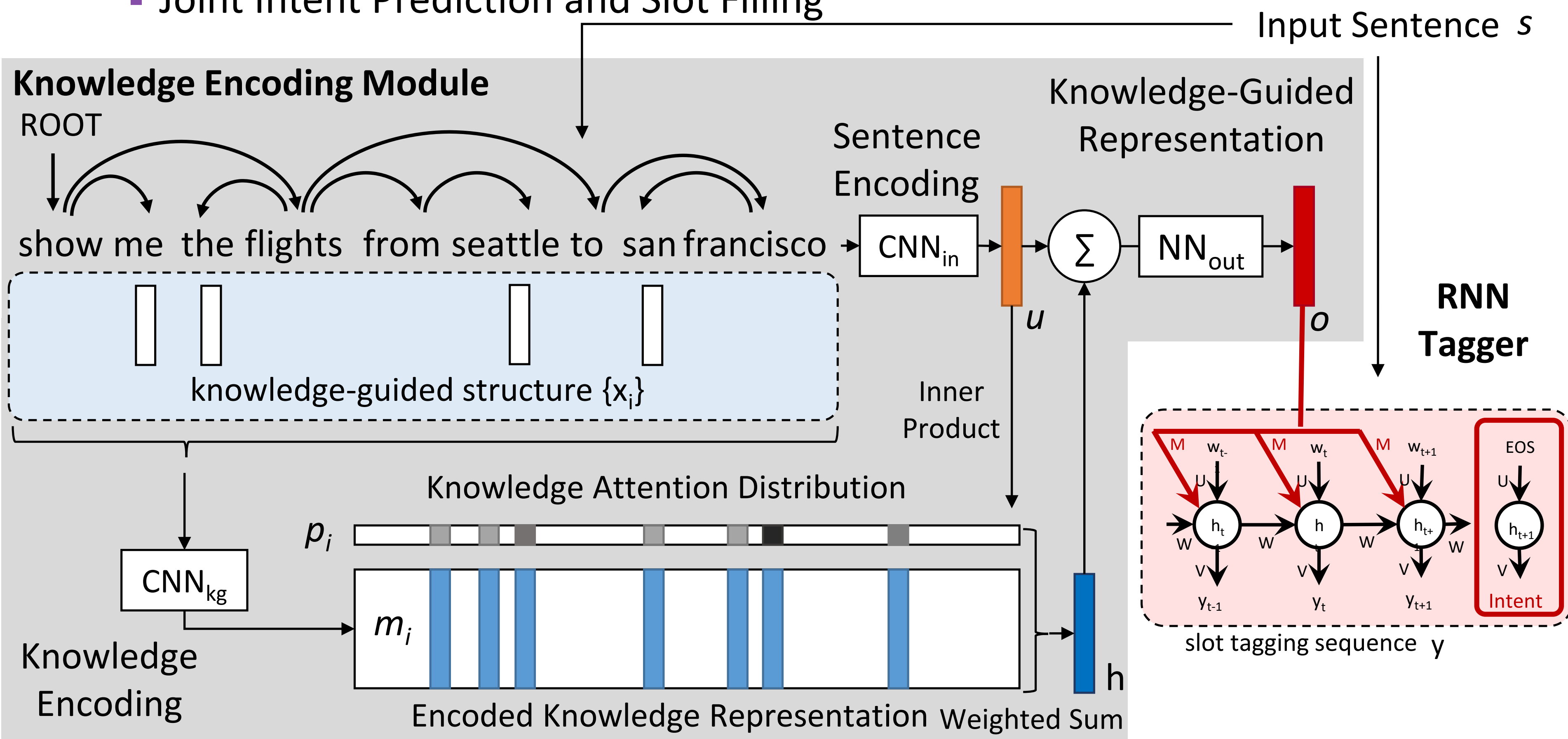
Syntax provides richer knowledge and more general guidance when less training data.

Semantics captures the most salient info so it achieves similar performance with much less substructures

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

K-SAN for Joint Semantic Frame Parsing

- Joint Intent Prediction and Slot Filling



Extend the K-SAN model for joint semantic frame parsing by outputting the user intent at last timestamp (Hakkani-Tur et al.).

Joint K-SAN Parsing Experiments

ATIS Dataset (train: 4478/dev: 500/test: 893)	Small (1/40)			Medium (1/10)			Large		
	Slot (Indep)	Slot (Joint)	Frame	Slot (Indep)	Slot (Joint)	Frame	Slot (Indep)	Slot (Joint)	Frame
Tagger	73.8	73.0		85.6	86.4		93.1	93.4	
Encoder-Tagger	72.8	71.9		88.3	87.5		94.8	93.1	
K-SAN (Syntax)	74.4^{+}	**74.6**$^{+}$		**88.4**$^{+}$	88.2^{+}		95.0^{+}	**95.4**$^{+}$	
K-SAN (Semantics)	74.3^{+}	73.4^{+}		88.3	88.1^{+}		94.9^{+}	95.1^{+}	

Communication (train: 10479/dev: 1000/test: 2300)	Small (1/40)			Medium (1/10)			Large		
	Slot (Indep)	Slot (Joint)	Frame	Slot (Indep)	Slot (Joint)	Frame	Slot (Indep)	Slot (Joint)	Frame
Tagger	45.5	50.3		69.0	69.8		80.4	79.8	
Encoder-Tagger	45.5	47.7		69.4	73.1		85.7	**86.0**	
K-SAN (Syntax)	45.0	**55.1**$^{+}$		69.5^{+}	**75.3**$^{+}$		85.0	84.5	
K-SAN (Semantics)	45.1	55.0		69.1	74.3^{+}		85.3	85.2	

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

Joint K-SAN Parsing Experiments

ATIS Dataset (train: 4478/dev: 500/test: 893)	Small (1/40)			Medium (1/10)			Large		
	Slot (Indep)	Slot (Joint)	Frame	Slot (Indep)	Slot (Joint)	Frame	Slot (Indep)	Slot (Joint)	Frame
Tagger	73.8	73.0	33.5	85.6	86.4	58.5	93.1	93.4	79.7
Encoder-Tagger	72.8	71.9	35.2	88.3	87.5	61.9	94.8	93.1	82.5
K-SAN (Syntax)	74.4^{+}	**74.6**$^{+}$	**37.6**$^{+}$	**88.4**$^{+}$	88.2^{+}	63.5^{+}	95.0^{+}	**95.4**$^{+}$	**84.3**$^{+}$
K-SAN (Semantics)	74.3^{+}	73.4^{+}	37.1^{+}	88.3	88.1^{+}	**63.6**$^{+}$	94.9^{+}	95.1^{+}	83.8^{+}

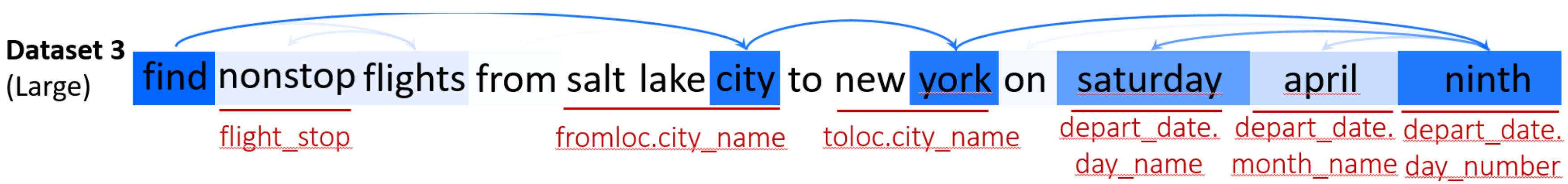
Communication (train: 10479/dev: 1000/test: 2300)	Small (1/40)			Medium (1/10)			Large		
	Slot (Indep)	Slot (Joint)	Frame	Slot (Indep)	Slot (Joint)	Frame	Slot (Indep)	Slot (Joint)	Frame
Tagger	45.5	50.3	48.9	69.0	69.8	68.2	80.4	79.8	79.5
Encoder-Tagger	45.5	47.7	52.7	69.4	73.1	71.4	85.7	**86.0**	**83.9**
K-SAN (Syntax)	45.0	**55.1**$^{+}$	**57.2**$^{+}$	69.5^{+}	**75.3**$^{+}$	73.5^{+}	85.0	84.5	84.5
K-SAN (Semantics)	45.1	55.0	54.1^{+}	69.1	74.3^{+}	**73.8**$^{+}$	85.3	85.2	83.4

When data is scare, K-SAN with joint parsing significantly improves the performance (slot & frame)

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, “Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks,” preprint arXiv: 1609.00777, 2016.

Attention Analysis

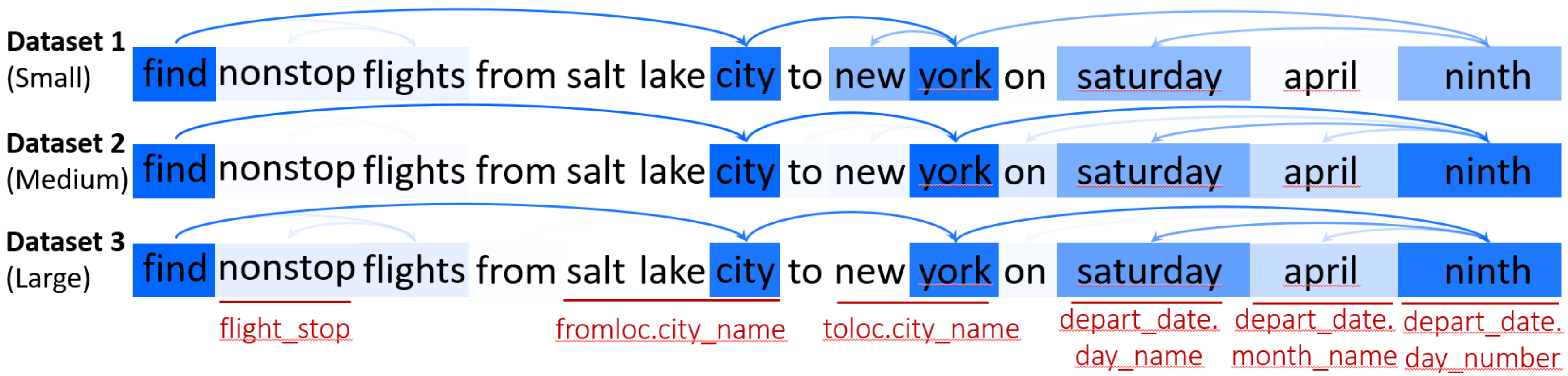
- Darker blocks and lines correspond to higher attention weights



Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

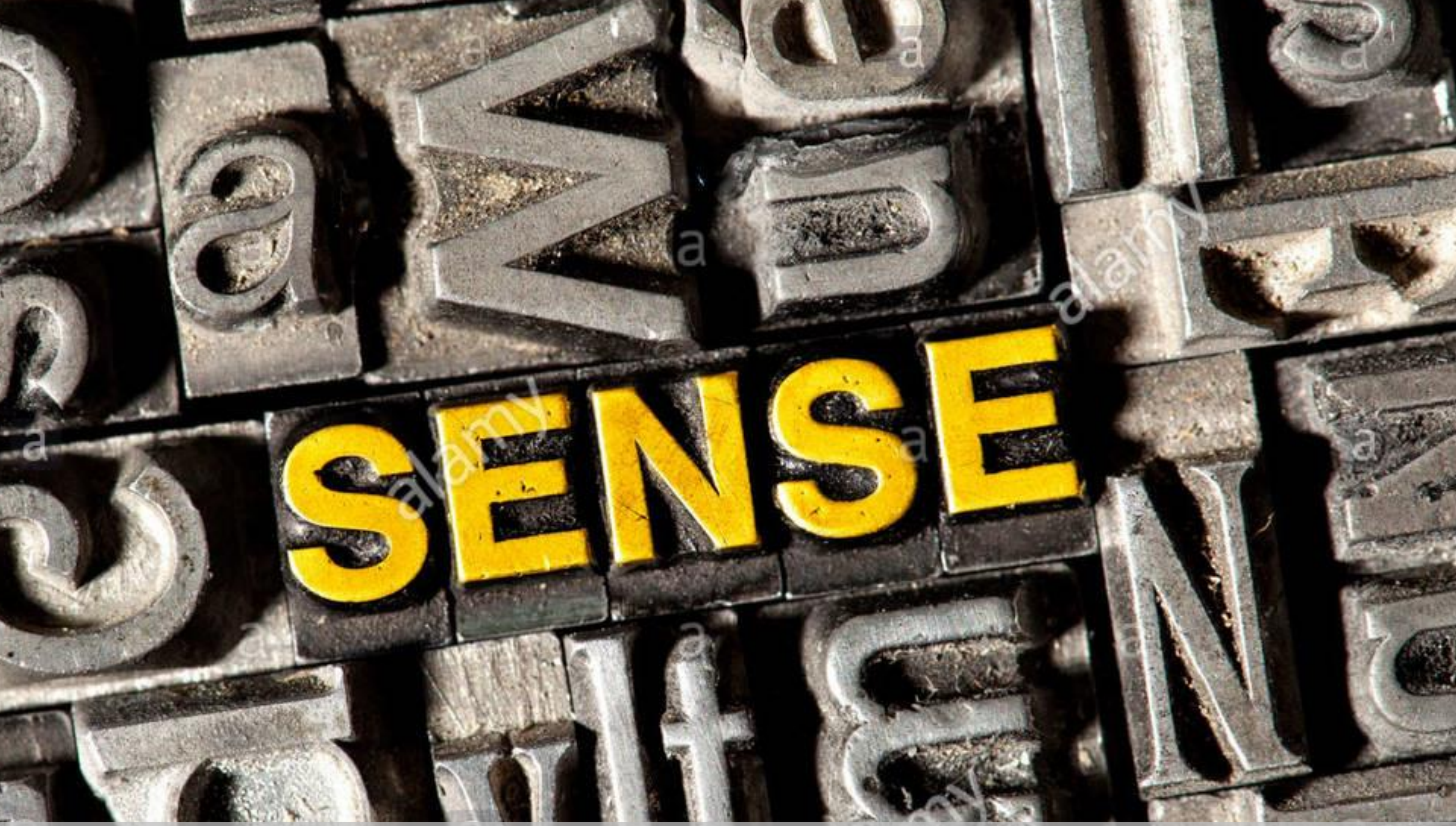
Attention Analysis

- Darker blocks and lines correspond to higher attention weights



Using less training data with K-SAN allows the model pay the similar attention to the salient substructures that are important for tagging.

Y.-N. Chen, D. Hakkani-Tur, G. Tur, A. Celikyilmaz, J. Gao, and L. Deng, "Knowledge as a Teacher: Knowledge-Guided Structural Attention Networks," preprint arXiv: 1609.00777, 2016.

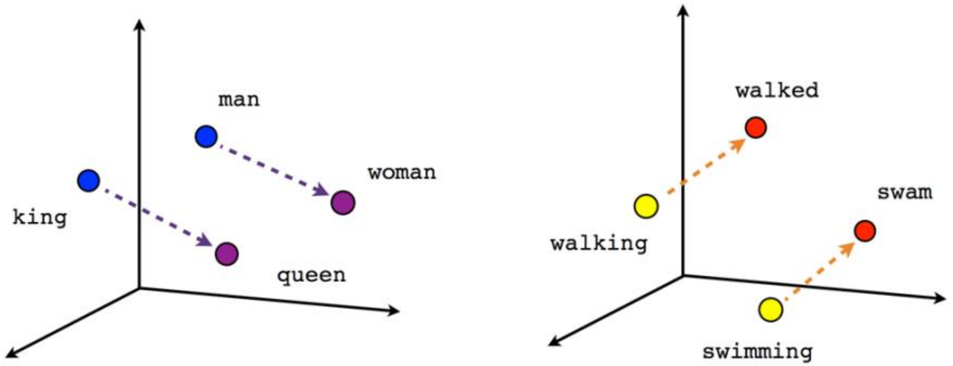


Modularizing Unsupervised Sense Embeddings (MUSE)

G.-H. Lee and Y.-N. Chen, "MUSE: Modularizing Unsupervised Sense Embeddings," preprint arXiv: 1704.04601, 2017.

Word Embedding

- Word embeddings are trained on a corpus in an unsupervised manner



Finally I chose Google instead of Apple.

Can you buy me a bag of apples, oranges, and bananas?

- Using the **same embeddings** for **different senses** for NLP tasks, e.g. NLU, POS tagging

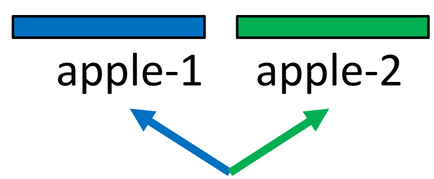
Words with different senses should correspond different embeddings

G.-H. Lee and Y.-N. Chen, "MUSE: Modularizing Unsupervised Sense Embeddings," preprint arXiv: 1704.04601, 2017.

Task – Unsupervised Sense Embeddings

- Input: unannotated text corpus
- Two key mechanisms
  - **Sense selection** given a text context
  - **Sense representation** to embed statistical characteristics of sense identity

sense embedding



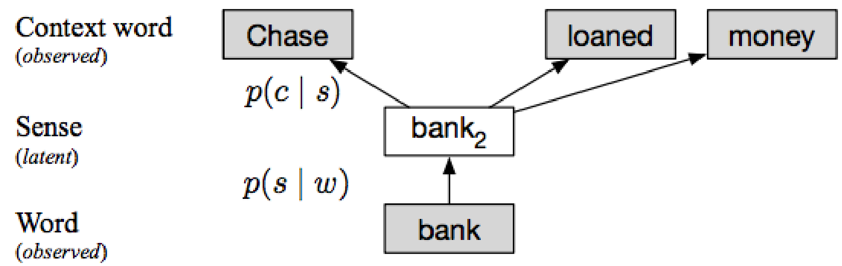
sense selection

Smartphone companies including apple blackberry, and sony will be invited.

G.-H. Lee and Y.-N. Chen, "MUSE: Modularizing Unsupervised Sense Embeddings," preprint arXiv: 1704.04601, 2017.

Prior Approaches

- Efficient sense selection [Neelakantan et al., 2014; Li and Jurafsky, 2015]
  - Use word embeddings as input to update the sense posterior given words
  - Introduce ambiguity



- Purely sense-level embedding [Qiu et al., 2016]
  - Inefficient sense selection → exponential time complexity

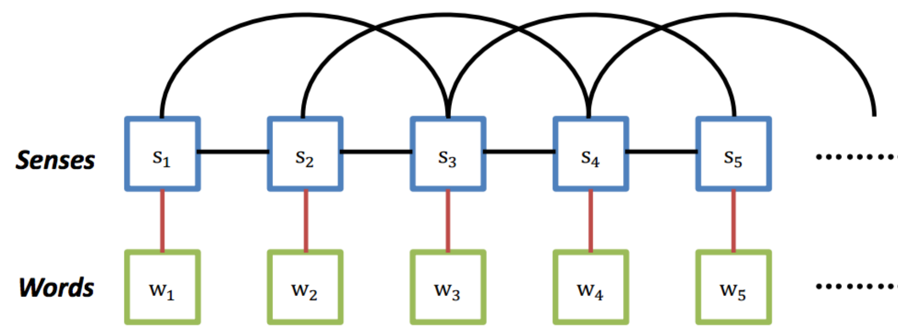
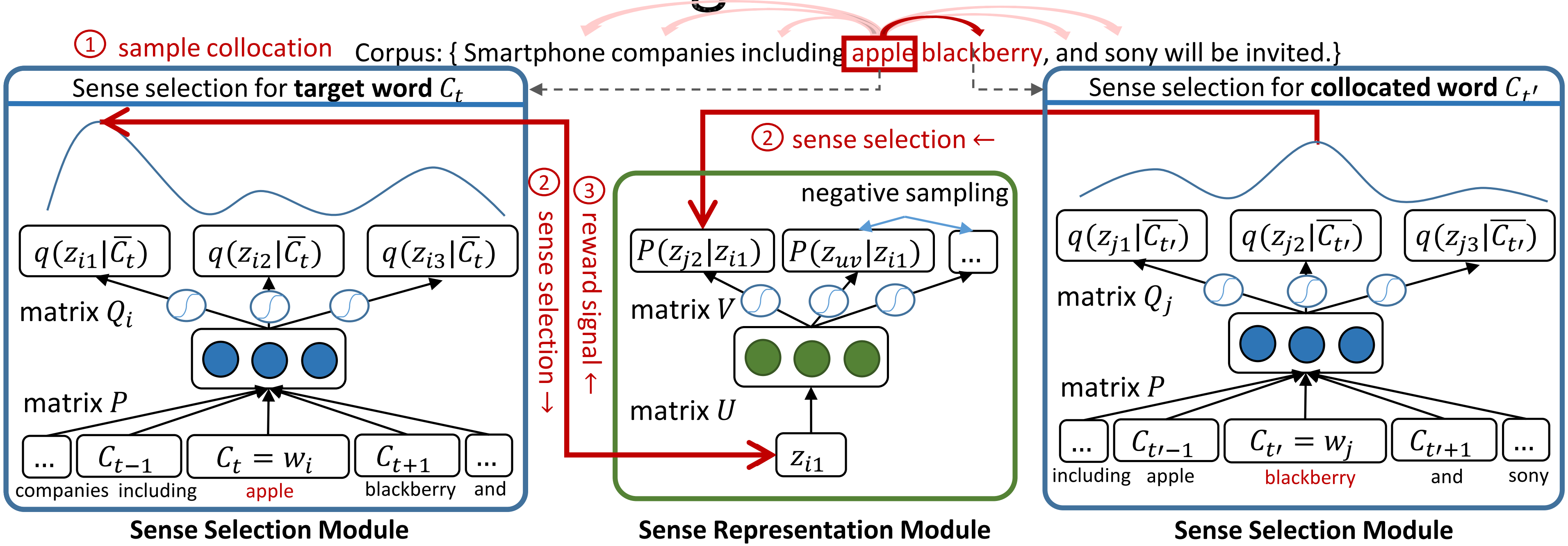


Figure 1: Context-Dependent Sense Embedding Model with window size $k = 1$

The prior approaches have disadvantages about either *ambiguity* or *inefficiency*

G.-H. Lee and Y.-N. Chen, "MUSE: Modularizing Unsupervised Sense Embeddings," preprint arXiv: 1704.04601, 2017.

MUSE: Modularizing Unsupervised Sense Embeddings



- Sense selection
  - Policy-based

$$\pi(z_{ik} \mid \bar{C}_t) = \frac{\exp(Q_{ik}^T \sum_{j \in \bar{C}_t} P_j)}{\sum_{k' \in Z_i} \exp(Q_{ik'}^T \sum_{j \in \bar{C}_t} P_j)}$$

  - Value-based

$$q(z_{ik} \mid \bar{C}_t) = \sigma(Q_{ik}^T \sum_{j \in \bar{C}_t} P_j)$$

- Sense representation learning

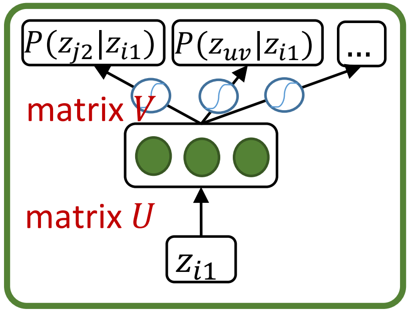
$$\log \mathcal{L}(z_{jl} \mid z_{ik}) = \log \frac{\exp(U_{z_{ik}}^T V_{z_{jl}})}{\sum_{z_{uv}} \exp(U_{z_{ik}}^T V_{z_{uv}})}$$

  - Skip-gram approximation

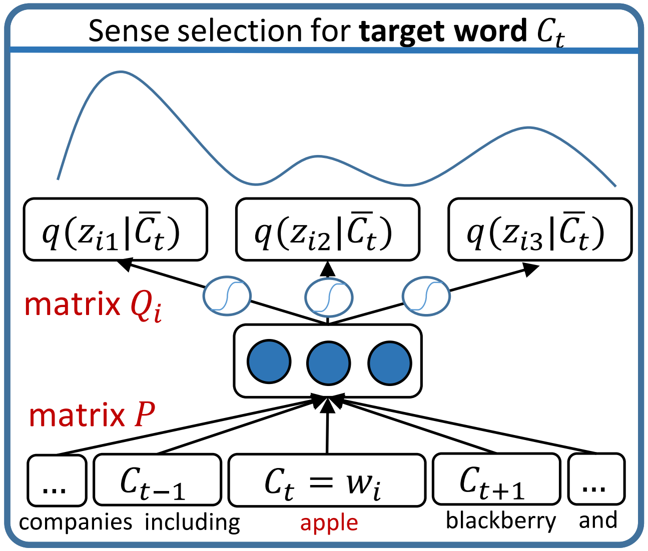
$$\log \bar{\mathcal{L}}(z_{jl} \mid z_{ik}) = \log \sigma(U_{z_{ik}}^T V_{z_{jl}}) + \sum^{M} \mathbb{E}_{z_{uv} \sim p_{neg}(z)} [\log \sigma(-U_{z_{ik}}^T V_{z_{uv}})]$$

Collocated likelihood serves as a reward signal to optimize the sense selection module.

Joint Learning



Sense Representation Module



Sense Selection Module

- Learning algorithm

Algorithm 1: Learning Algorithm

for $w_i = C_t \in C$ **do**
- sample $w_j = C_{t'} (0 < |t' - t| \leq m)$;
- $z_{ik} = \text{select}(C_t, w_i)$;
- $z_{jl} = \text{select}(C_{t'}, w_j)$;
- optimize U, V by (4) for the sense representation module;
- optimize P, Q by (5) or (7) for the sense selection module;

- Sense selection strategy
  - **Stochastic policy**: selects the sense based on the probability distribution
  - **Greedy**: selects the sense with the largest Q-value (no exploration)
  - **ε-Greedy**: selects a random sense with ε probability, and adopts the greedy strategy
  - **Boltzmann**: samples the sense based on the Boltzmann distribution modeled by Q-value

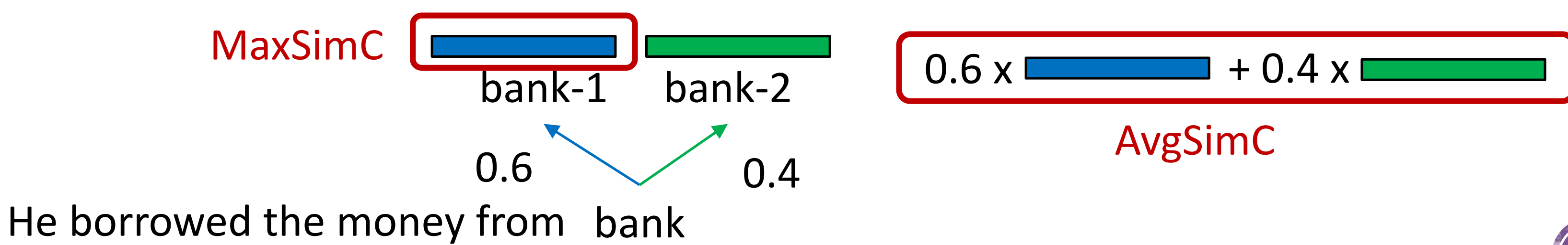
Contextual Word Similarity Experiments

- Dataset: SCWS for multi-sense embedding evaluation

He borrowed the money from **banks**. | I live near to a **river**. | correlation=?

Baseline

Approach	MaxSimC	AvgSimC
Huang et al., 2012	26.1	65.7
Neelakantan et al., 2014	60.1	**69.3**
Tian et al., 2014	63.6	65.4
Li & Jurafsky, 2015	**66.6**	66.8
Bartunov et al., 2016	53.8	61.2
Qiu et al., 2016	64.9	66.1



Contextual Word Similarity Experiments

- Dataset: SCWS for multi-sense embedding evaluation

He borrowed the money from **banks**.	I live near to a **river**.	correlation=?

Approach	MaxSimC	AvgSimC
Huang et al., 2012	26.1	65.7
Neelakantan et al., 2014	60.1	**69.3**
Tian et al., 2014	63.6	65.4
Li & Jurafsky, 2015	**66.6**	66.8
Bartunov et al., 2016	53.8	61.2
Qiu et al., 2016	64.9	66.1
MUSE-Policy	66.1	67.4

Contextual Word Similarity Experiments

- Dataset: SCWS for multi-sense embedding evaluation

He borrowed the money from **banks**.	I live near to a **river**.	correlation=?

Approach	MaxSimC	AvgSimC
Huang et al., 2012	26.1	65.7
Neelakantan et al., 2014	60.1	**69.3**
Tian et al., 2014	63.6	65.4
Li & Jurafsky, 2015	**66.6**	66.8
Bartunov et al., 2016	53.8	61.2
Qiu et al., 2016	64.9	66.1
MUSE-Policy	66.1	67.4
MUSE-Greedy	66.3	68.3

Contextual Word Similarity Experiments

- Dataset: SCWS for multi-sense embedding evaluation

He borrowed the money from **banks**. | I live near to a **river**. | correlation=?

Approach	MaxSimC	AvgSimC
Huang et al., 2012	26.1	65.7
Neelakantan et al., 2014	60.1	**69.3**
Tian et al., 2014	63.6	65.4
Li & Jurafsky, 2015	66.6	66.8
Bartunov et al., 2016	53.8	61.2
Qiu et al., 2016	64.9	66.1
MUSE-Policy	66.1	67.4
MUSE-Greedy	66.3	68.3
MUSE-ε-Greedy	**67.4**+	68.6

Contextual Word Similarity Experiments

- Dataset: SCWS for multi-sense embedding evaluation

He borrowed the money from **banks**. | I live near to a **river**. | correlation=?

Approach	MaxSimC	AvgSimC
Huang et al., 2012	26.1	65.7
Neelakantan et al., 2014	60.1	**69.3**
Tian et al., 2014	63.6	65.4
Li & Jurafsky, 2015	66.6	66.8
Bartunov et al., 2016	53.8	61.2
Qiu et al., 2016	64.9	66.1
MUSE-Policy	66.1	67.4
MUSE-Greedy	66.3	68.3
MUSE-ε-Greedy	67.4+	68.6
MUSE-Boltzmann	**67.9**+	68.7

MUSE with exploration achieves the best sense embeddings in MaxSimC.

Synonym Selection Experiments

	Approach	ESL-50	RD-300	TOEFL-80
Conventional Word Embedding	Global Context	47.73	45.07	60.87
	SkipGram	52.08	55.66	66.67
Word Sense Disambiguation	IMS+SkipGram	41.67	53.77	66.67
Unsupervised Sense Embedding	EM	27.08	33.96	40.00
	MSSG (Neelakantan et al., 2014)	<u>57.14</u>	<u>58.93</u>	78.26
	CRP (Li & Jurafsky, 2015)	50.00	55.36	<u>82.61</u>
	MUSE-Policy	52.38	51.79	79.71
	MUSE-Greedy	57.14	58.93	79.71
	MUSE-ε-Greedy	61.90+	62.50+	84.06+
	MUSE-Boltzmann	**64.29**+	**66.07**+	**88.41**+
Supervised Sense Embedding	Retro-GlobalContext	63.64	66.20	71.01
	Retro-SkipGram	56.25	65.09	73.33

MUSE with exploration achieves the state-of-the-art results for synonym selection.

Qualitative Analysis

- KNN senses sorted by collocation likelihood

Context	KNN Senses
… braves finish the season in **tie** with the los angeles dodgers …	scoreless otl shootout 6-6 hingis 3-3 7-7 0-0
… his later years proudly wore **tie** with the chinese characters for …	pants trousers shirt juventus blazer socks anfield
… of the mulberry or the **blackberry** and minos sent him to …	cranberries maple vaccinium apricot apple
… of the large number of **blackberry** users in the us federal …	smartphones sap microsoft ipv6 smartphone
… ladies wore extravagant **head** ornaments combs pearl necklaces face …	venter thorax neck spear millimeters fusiform
… appoint john pope republican as **head** of the new army of …	multi-party appoints unicameral beria appointed

MUSE learns sense embeddings in an ***unsupervised*** way and achieves the first ***purely sense-level*** representation learning system with ***linear-time sense selection***

Y.-N. Chen, S. Ming, A. I Rudnicky, and A. Gershman, "Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding," in *Proc. of ICMI*, pages 83-86, 2015. ACM.

Leveraging Behavior Patterns of Mobile Apps for Personalized Spoken Language Understanding

Introduction

- Task: user intent prediction
- Challenge: language ambiguity

send to vivian

Communication

Email? v.s. Message?

① **User preference**
  - ✓ Some people prefer "Message" to "Email"
  - ✓ Some people prefer "Outlook" to "Gmail"

② **App-level contexts**
  - ✓ "Message" is more likely to follow "Camera"
  - ✓ "Email" is more likely to follow "Excel"

Considering behavioral patterns in history to model SLU for intent prediction.

Y.-N. Chen, S. Ming, A. I Rudnicky, and A. Gershman, "Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding," in *Proc. of ICMI*, pages 83-86, 2015. ACM.

Data Collection

- Subjects' app invocation is logged on a daily basis
- Subjects annotate their app activities with
  - **Task Structure**: link applications that serve a common goal
  - **Task Description**: briefly describe the goal or intention of the task

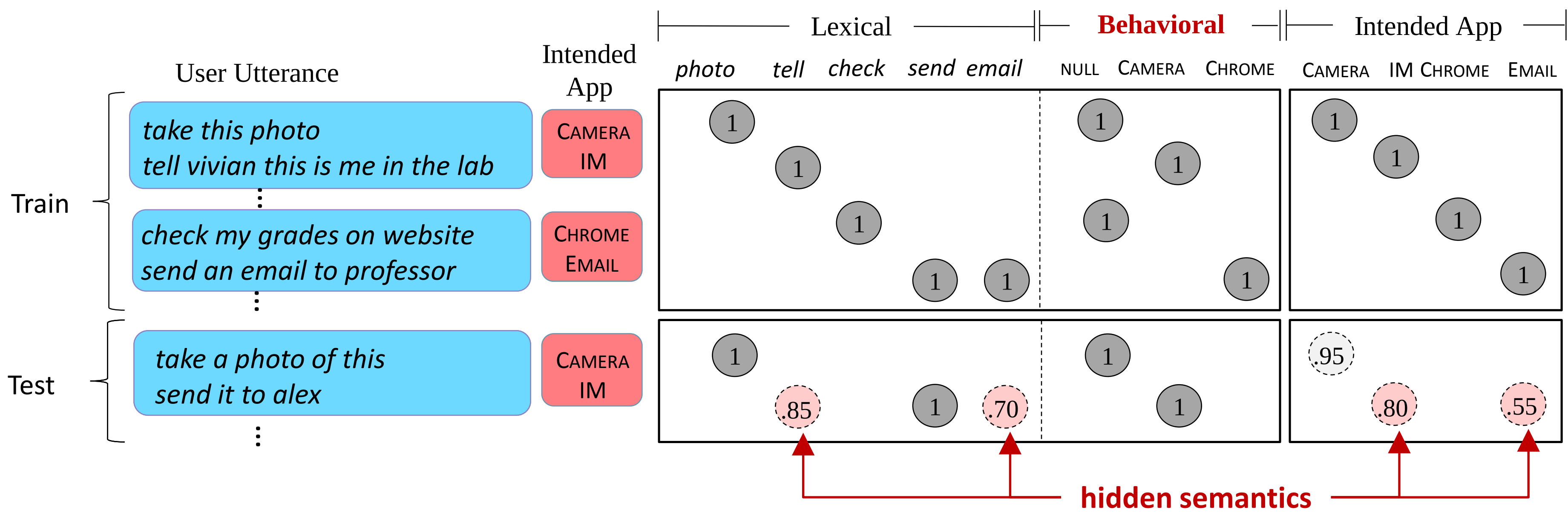
Meta	TASK59; 20150203; 1; Tuesday; 10:48
App	com.android.settings → com.lge.music
Desc	play music via bluetooth speaker

- Subjects use a wizard system to perform the annotated task by speech

Dialogue			
	$\mathbf{W_1}$:	Ready.	
	$\mathbf{U_1}$:	Connect my phone to bluetooth speaker.	SETTINGS
	$\mathbf{W_2}$:	Connected to bluetooth speaker.	
	$\mathbf{U_2}$:	And play music.	MUSIC
	$\mathbf{W_3}$:	What music would you like to play?	
	$\mathbf{U_3}$:	Shuffle playlist.	MUSIC
	$\mathbf{W_4}$:	I will play the music for you.	

Y.-N. Chen, S. Ming, A. I Rudnicky, and A. Gershman, "Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding," in *Proc. of ICMI*, pages 83-86, 2015. ACM.

SLU for Intent Prediction



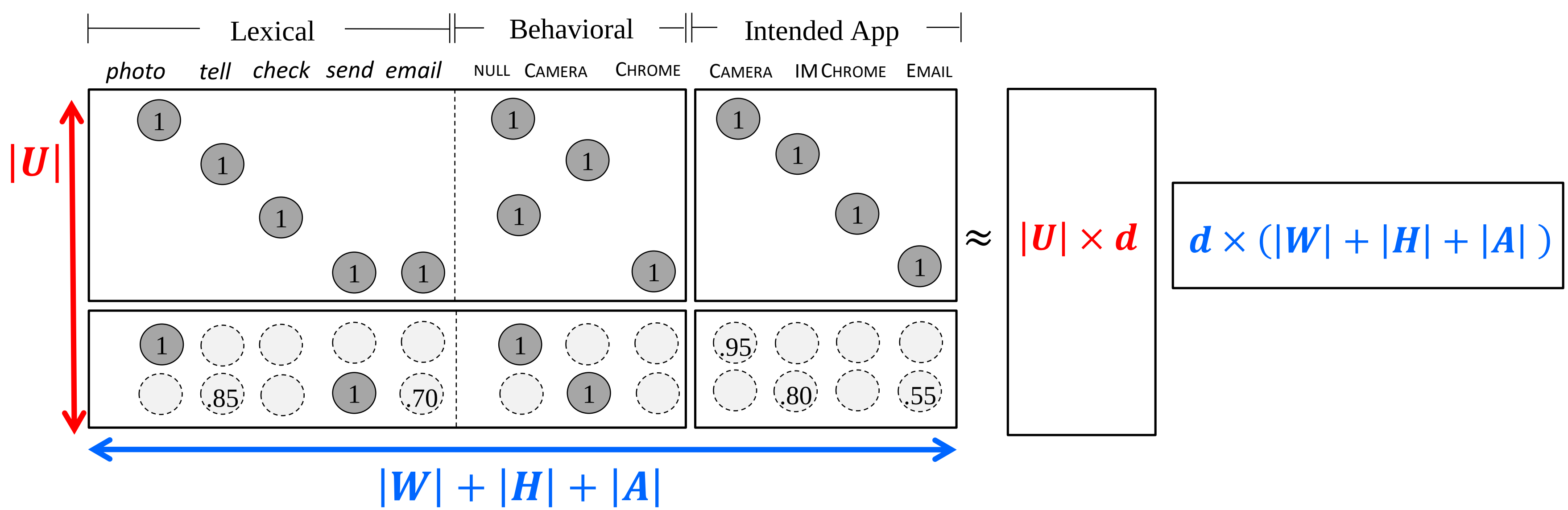
Issue: unobserved hidden semantics may benefit understanding

Solution: use matrix factorization to complete a partially-missing matrix based on a low-rank latent semantics assumption.

Y.-N. Chen, S. Ming, A. I Rudnicky, and A. Gershman, "Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding," in *Proc. of ICMI*, pages 83-86, 2015. ACM.

Matrix Factorization (MF)

- The decomposed matrices represent low-rank latent semantics for utterances and words/histories/apps respectively
- The product of two matrices fills the probability of hidden semantics



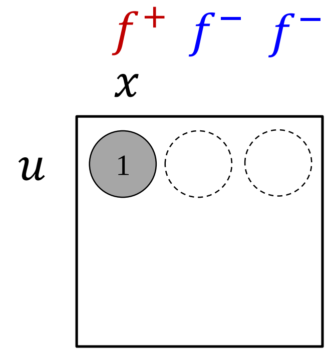
Y.-N. Chen, S. Ming, A. I Rudnicky, and A. Gershman, "Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding," in *Proc. of ICMI*, pages 83-86, 2015. ACM.

Parameter Estimation

- Model implicit feedback by completing the matrix
  - not treat unobserved facts as negative samples (true or false)
  - give observed facts higher scores than unobserved facts

$$\begin{aligned} f^+ &= \langle u, x^+ \rangle \\ f^- &= \langle u, x^- \rangle \end{aligned} \Rightarrow p(f^+) > p(f^-)$$

$$p(M_{u,x} = 1 \mid \theta_{u,x}) = \sigma(\theta_{u,x}) = \frac{1}{1 + \exp(-\theta_{u,x})}$$



- Objective:

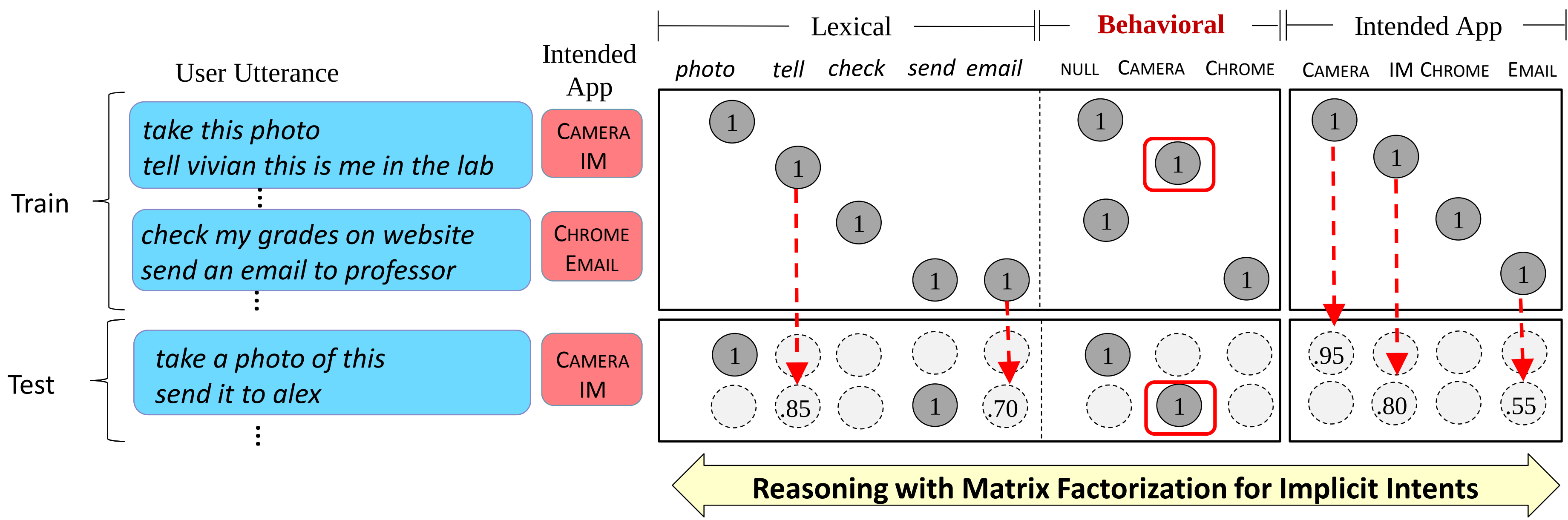
$$\sum_{f^+ \in \mathcal{O}} \sum_{f^- \notin \mathcal{O}} \ln \sigma(\theta_{f^+} - \theta_{f^-})$$

  - the model can be achieved by SGD updates with fact pairs

The objective is to learn a set of well-ranked apps per utterance.

Y.-N. Chen, S. Ming, A. I Rudnicky, and A. Gershman, "Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding," in *Proc. of ICMI*, pages 83-86, 2015. ACM.

SLU Modeling by MF



Y.-N. Chen, S. Ming, A. I Rudnicky, and A. Gershman, "Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding," in *Proc. of ICMI*, pages 83-86, 2015. ACM.

Experiments

- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC)
  mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (**MLE**)

Approach			Lexical	Behavioral	All
(a)	MLE	User-Indep		13.5 / 19.6	
(b)		User-Dep		20.2 / 27.9	

The user-dependent model is better than the user-independent model.

Y.-N. Chen, S. Ming, A. I Rudnicky, and A. Gershman, "Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding," in *Proc. of ICMI*, pages 83-86, 2015. ACM.

Experiments

- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC)
  mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (**MLE**)
  Multinomial Logistic Regression (**MLR**)

Approach			Lexical	Behavioral	All
(a)	MLE	User-Indep		13.5 / 19.6	
(b)		User-Dep		20.2 / 27.9	
(c)	MLR	User-Indep	42.8 / 46.4	14.9 / 18.7	
(d)		User-Dep	48.2 / 52.1	19.3 / 25.2	

Lexical features are useful to predict intended apps for both user-independent and user-dependent models.

Y.-N. Chen, S. Ming, A. I Rudnicky, and A. Gershman, "Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding," in *Proc. of ICMI*, pages 83-86, 2015. ACM.

Experiments

- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC)
  mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (**MLE**)
  Multinomial Logistic Regression (**MLR**)

Approach			Lexical	Behavioral	All
(a)	MLE	User-Indep		13.5 / 19.6	
(b)		User-Dep		20.2 / 27.9	
(c)	MLR	User-Indep	42.8 / 46.4	14.9 / 18.7	**46.2^{+} / 50.1^{+}**
(d)		User-Dep	48.2 / 52.1	19.3 / 25.2	**50.1^{+} / 53.9^{+}**

Combining lexical and behavioral features improves MLR performance, which models explicit information from observations.

Y.-N. Chen, S. Ming, A. I Rudnicky, and A. Gershman, "Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding," in *Proc. of ICMI*, pages 83-86, 2015. ACM.

Experiments

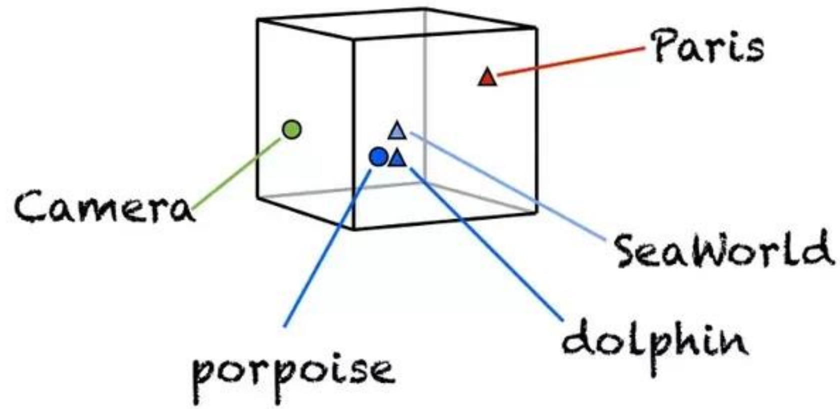
- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC)
  mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (**MLE**)
  Multinomial Logistic Regression (**MLR**)

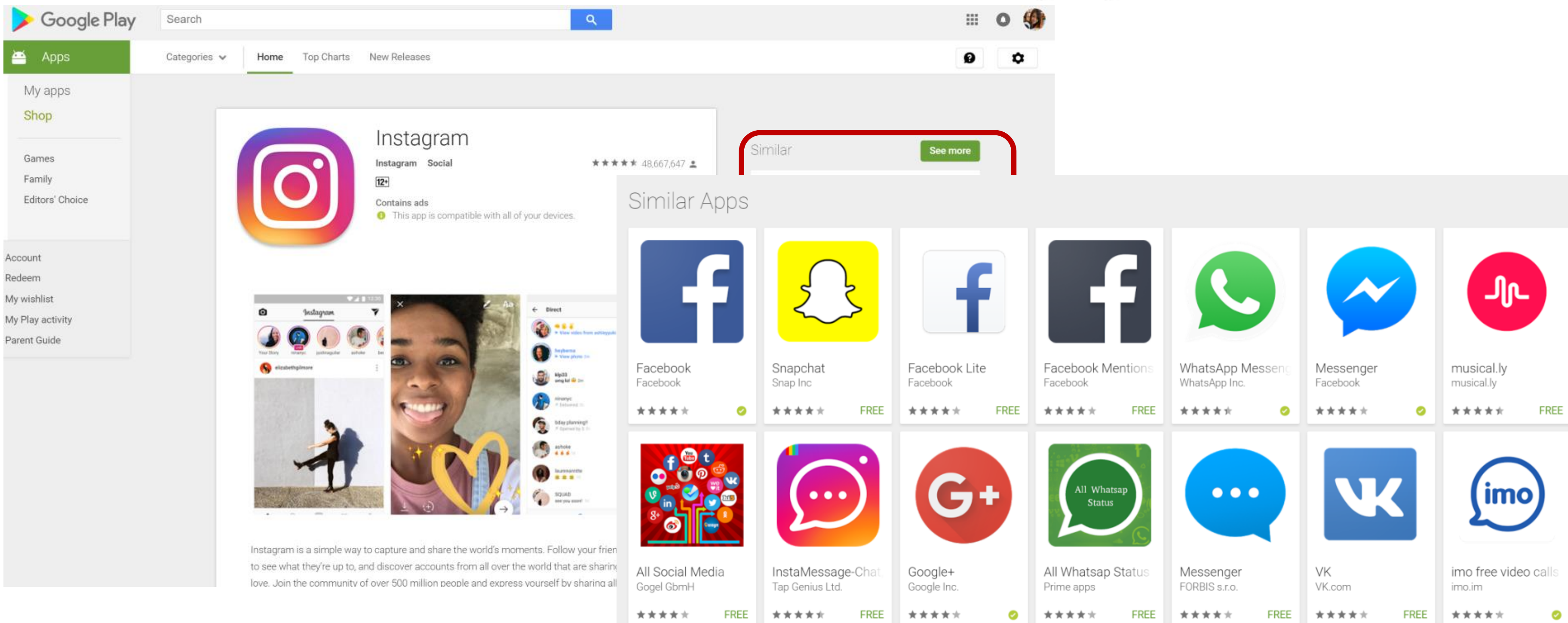
Approach			Lexical	Behavioral	All
(a)	MLE	User-Indep	13.5 / 19.6		
(b)		User-Dep	20.2 / 27.9		
(c)	MLR	User-Indep	42.8 / 46.4	14.9 / 18.7	46.2^{+} / 50.1^{+}
(d)		User-Dep	48.2 / 52.1	19.3 / 25.2	50.1^{+} / 53.9^{+}
(e)	(c) + Personalized MF		**47.6 / 51.1**	**16.4 / 20.3**	**50.3^{+*} / 54.2^{+*}**
(f)	(d) + Personalized MF		**48.3 / 52.7**	**20.6 / 26.7**	**51.9^{+*} / 55.7^{+*}**

Personalized MF significantly improves MLR results by considering hidden semantics.

Extension

- App functionality modeling
  - Learning app embeddings





App embeddings encoding functionality help user-independent understanding

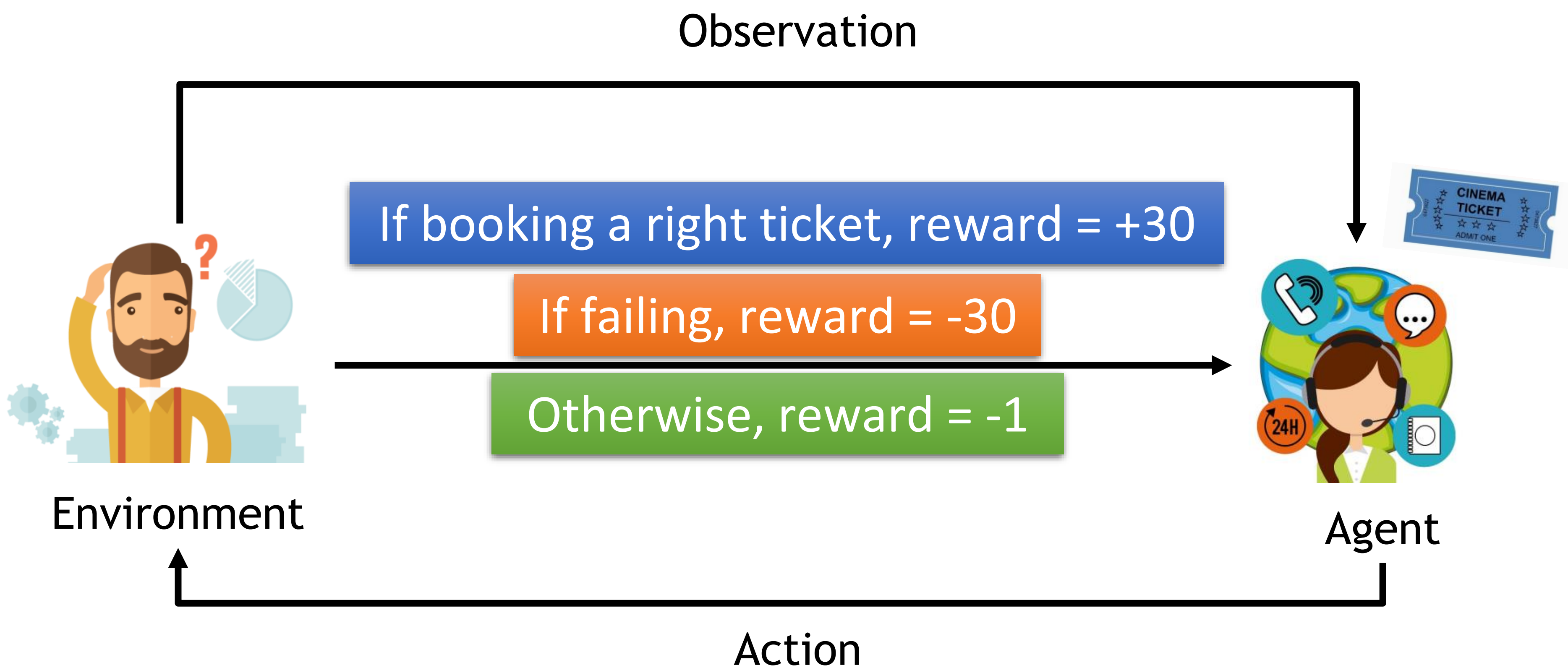
Y.-N. Chen, S. Ming, A. I Rudnicky, and A. Gershman, "Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding," in *Proc. of ICMI*, pages 83-86, 2015. ACM.

Investigation of Language Understanding Impact for Reinforcement Learning Based Dialogue Systems

X. Li, Y.-N. Chen, L. Li, and J. Gao, "End-to-End Task-Completion Neural Dialogue Systems," preprint arXiv: 1703.01008, 2017.
X. Li, Y.-N. Chen, L. Li, J. Gao, and A. Celikyilmaz, "Investigation of Language Understanding Impact for Reinforcement Learning Based Dialogue Systems," preprint arXiv: 1703.07055, 2017.

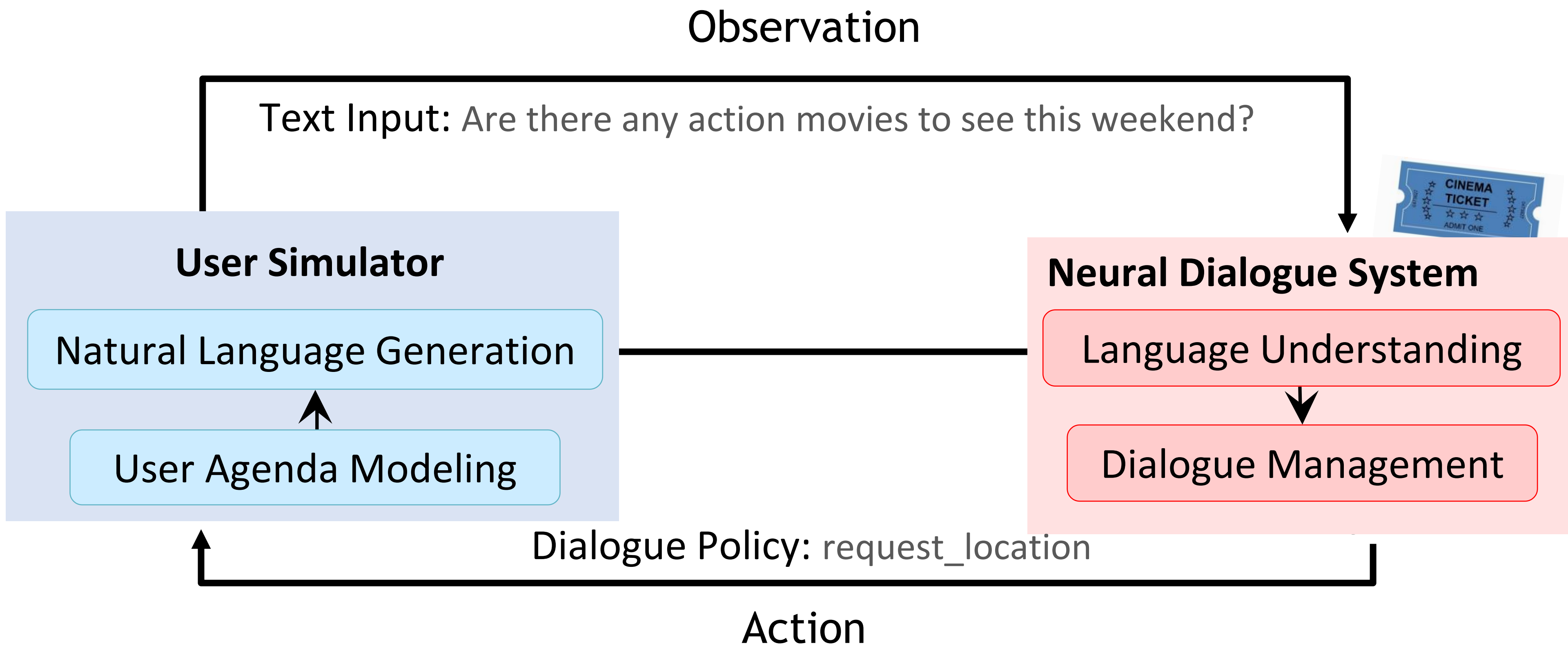
E2E Neural Dialogue System

- **Dialogue management** is framed as a **reinforcement learning** task
- Agent learns to select actions to **maximize the expected reward**



E2E Neural Dialogue System

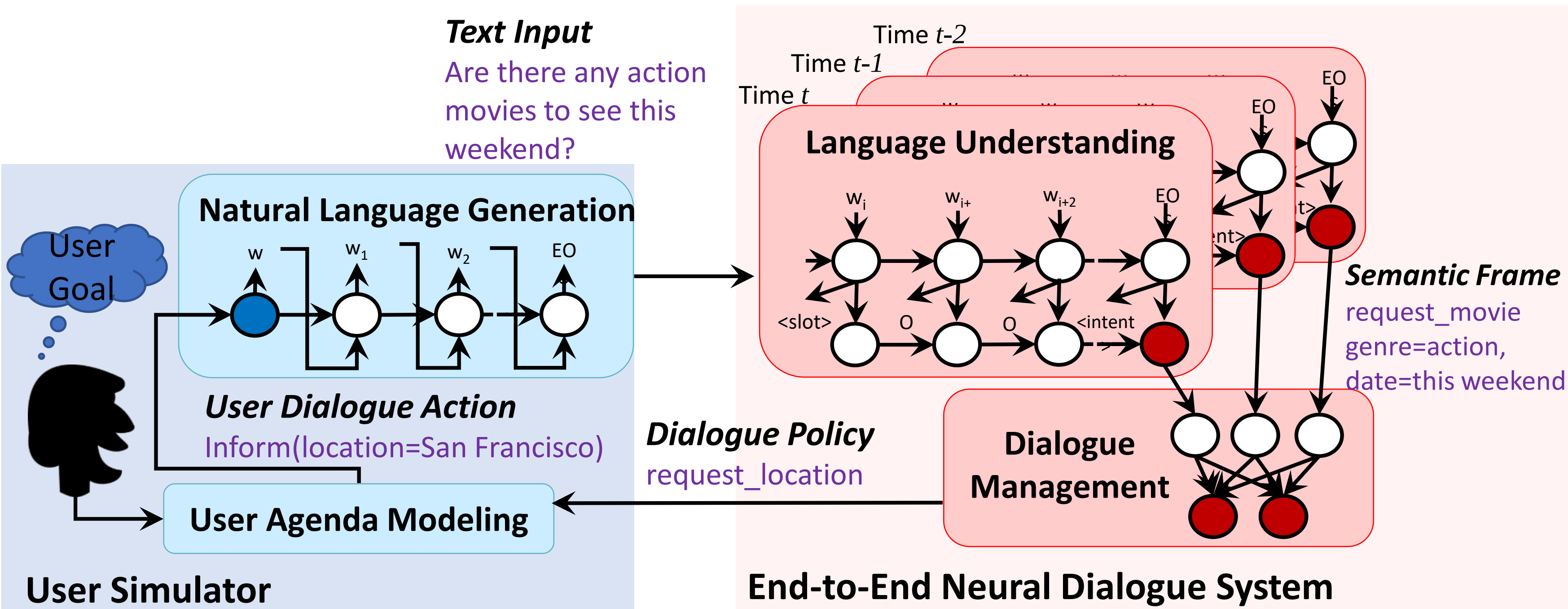
- **Dialogue management** is framed as a **reinforcement learning** task
- Agent learns to select actions to **maximize the expected reward**



X. Li, Y.-N. Chen, L. Li, and J. Gao, "End-to-End Task-Completion Neural Dialogue Systems," preprint arXiv: 1703.01008, 2017.

E2E Neural Dialogue System

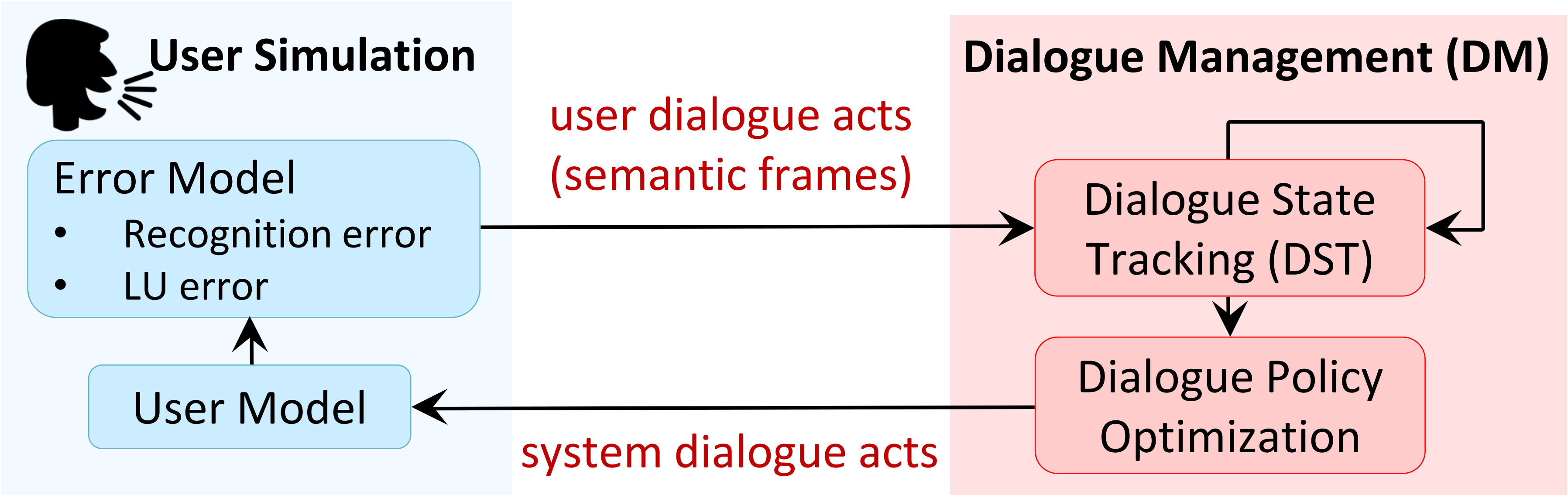
- NLU and NLG are trained in a supervised manner
- DM is trained in a reinforcement learning framework (NLU and NLG can be fine tuned)



X. Li, Y.-N. Chen, L. Li, and J. Gao, "End-to-End Task-Completion Neural Dialogue Systems," preprint arXiv: 1703.01008, 2017.

Frame-Level Interaction

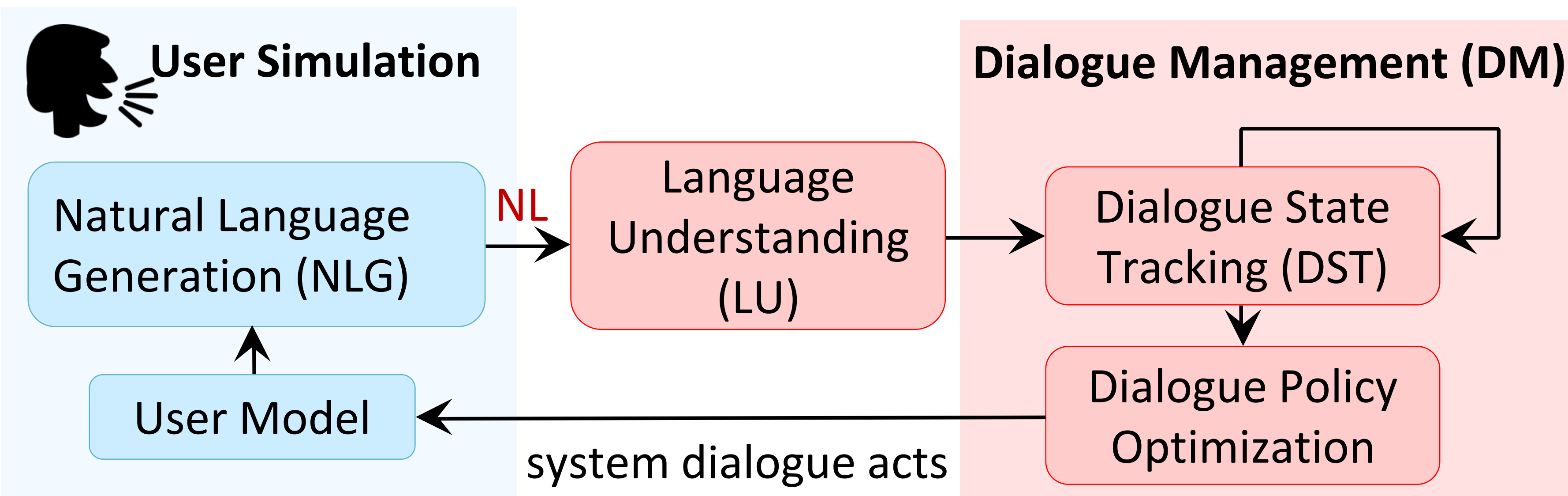
- DM receives frame-level information
  - No error model: perfect recognizer and LU
  - Error model: simulate the possible errors



X. Li, Y.-N. Chen, L. Li, and J. Gao, "End-to-End Task-Completion Neural Dialogue Systems," preprint arXiv: 1703.01008, 2017.

Natural Language Level Interaction

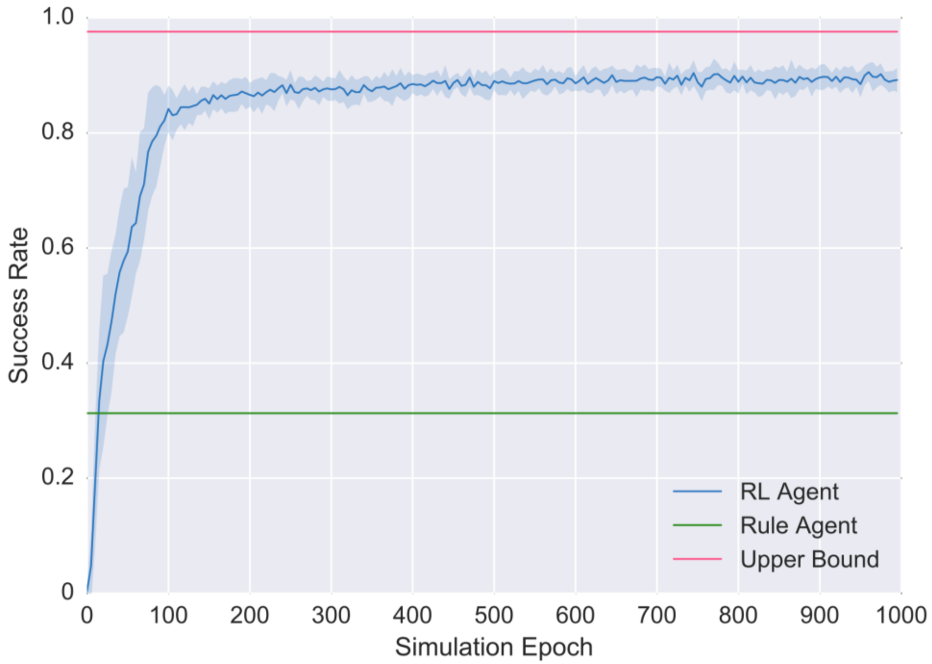
- User simulator sends natural language
  - No recognition error
  - Errors from NLG or LU



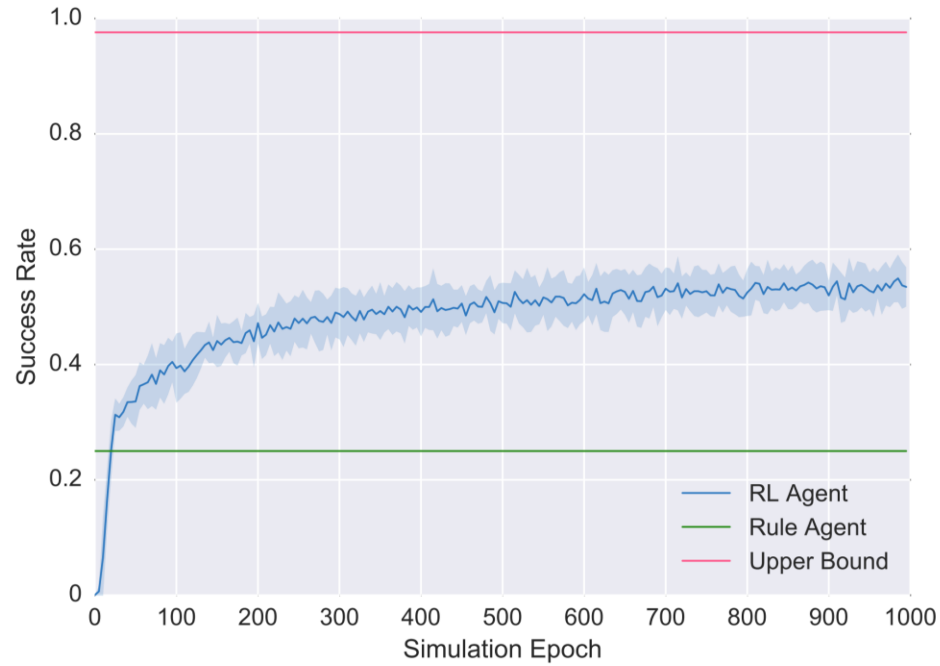
X. Li, Y.-N. Chen, L. Li, and J. Gao, "End-to-End Task-Completion Neural Dialogue Systems," preprint arXiv: 1703.01008, 2017.

End-to-End Reinforcement Learning

- Frame-level semantics



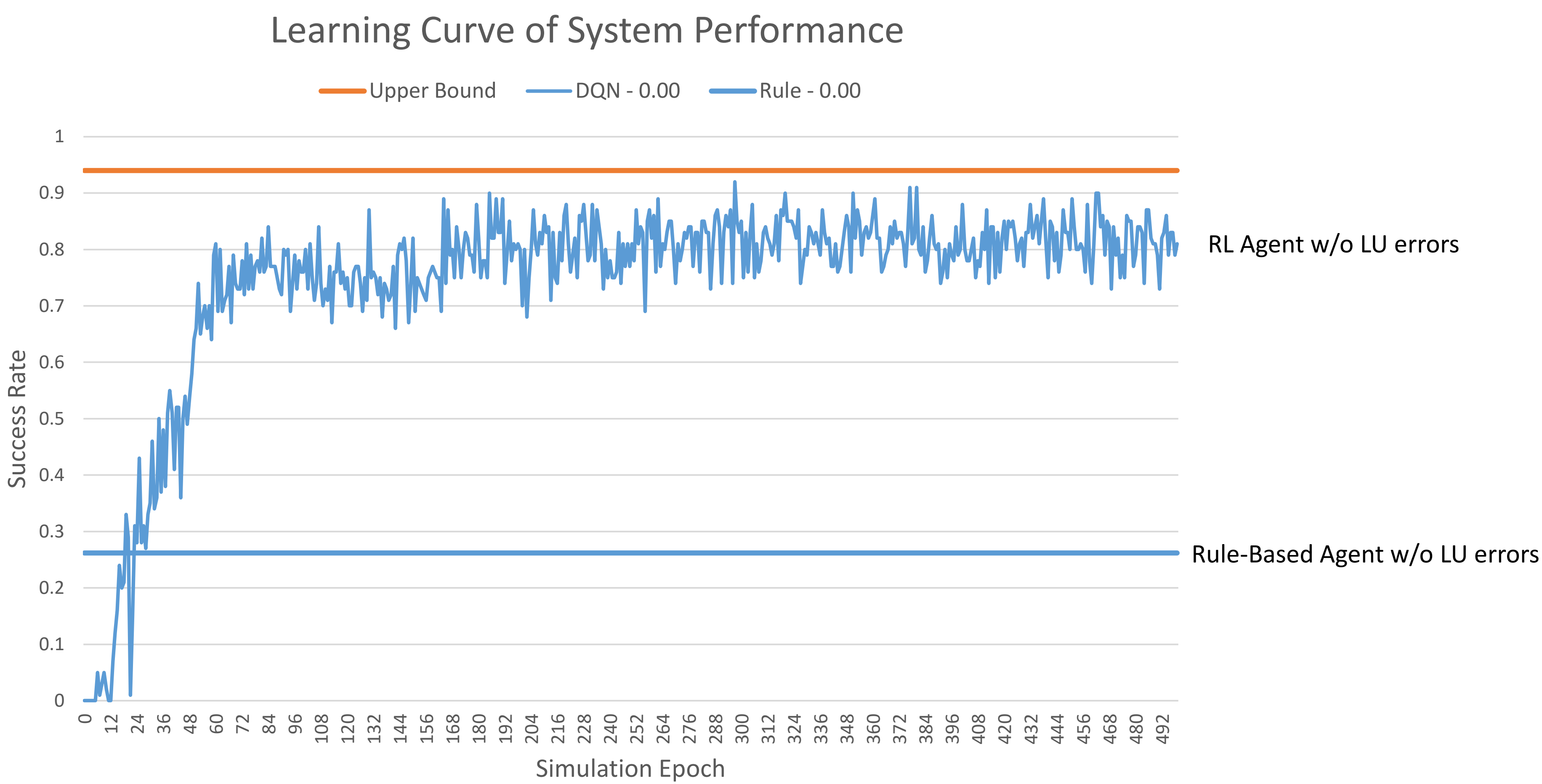
- Natural language



The RL agent is able to learn how to interact with users to complete tasks more **efficiently** and **effectively**, and outperforms the rule-based agent.

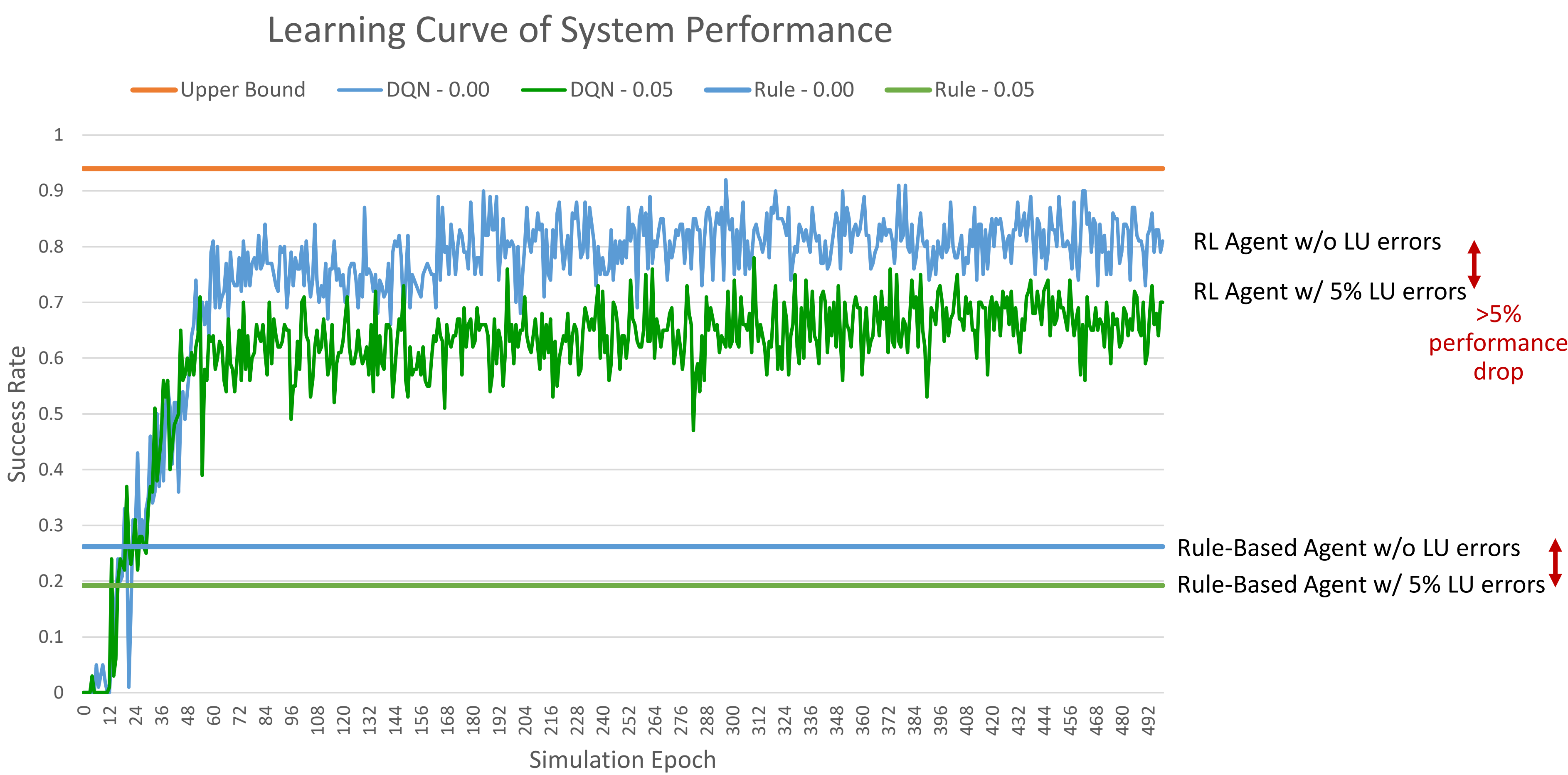
X. Li, Y.-N. Chen, L. Li, and J. Gao, "End-to-End Task-Completion Neural Dialogue Systems," preprint arXiv: 1703.01008, 2017.

Why is LU so important?



X. Li, Y.-N. Chen, L. Li, J. Gao, and A. Celikyilmaz, "Investigation of Language Understanding Impact for Reinforcement Learning Based Dialogue Systems," preprint arXiv: 1703.07055, 2017.

Why is LU so important?



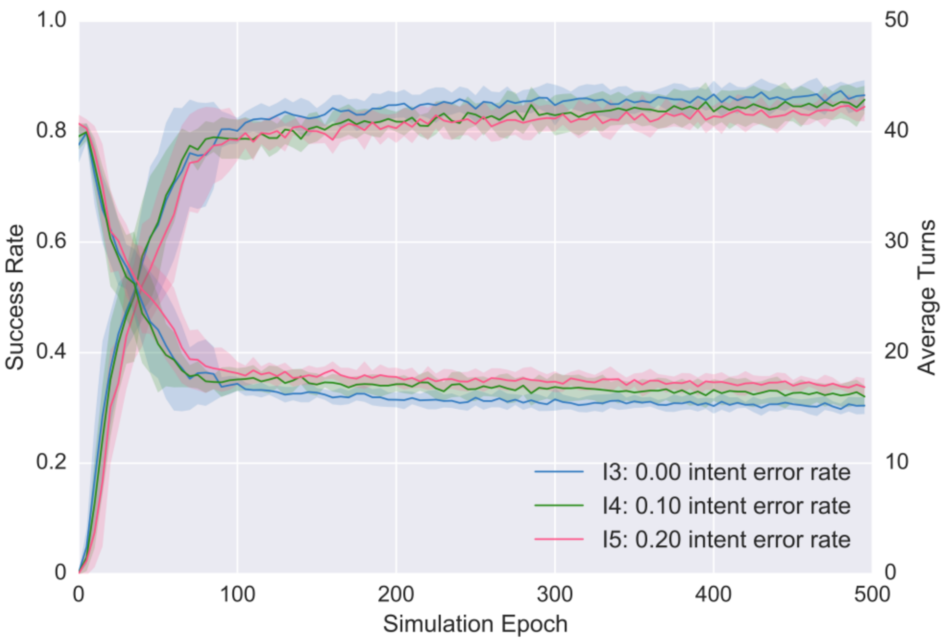
The system performance is sensitive to **LU errors (sentence-level contexts)**, for both rule-based and RL agents.

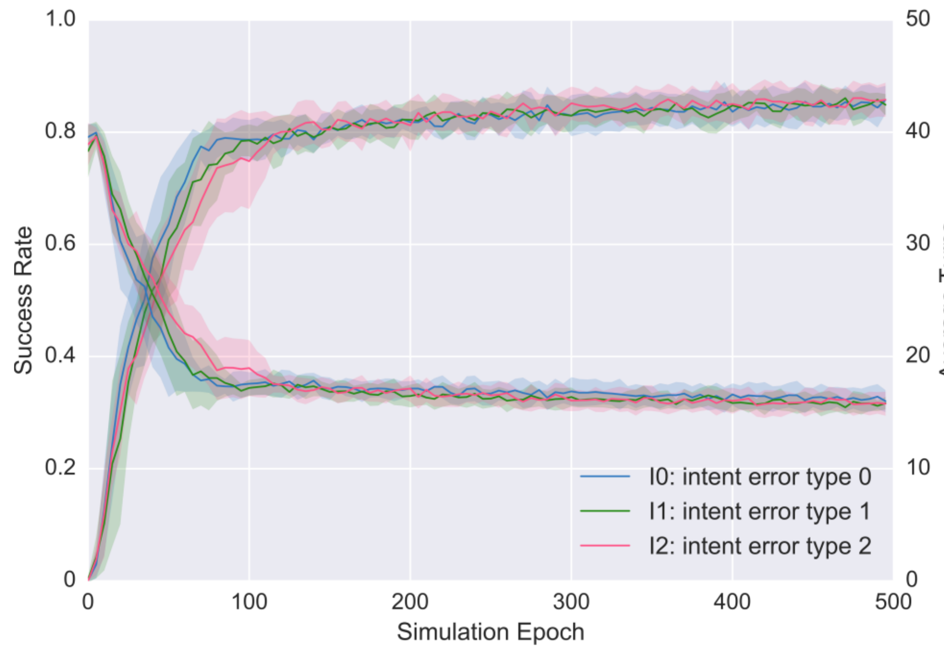
Intent Error Analysis

request_year
~~request_moviename~~(actor=Robert Downey Jr)

Group 1: greeting(), thanks(), etc
Group 2: inform(xx)
Group 3: request(xx)

- Intent error rate
  - I3: 0.00
  - I4: 0.10
  - I5: 0.20
- Intent error type
  - I0: random
  - I1: within group
  - I2: between group





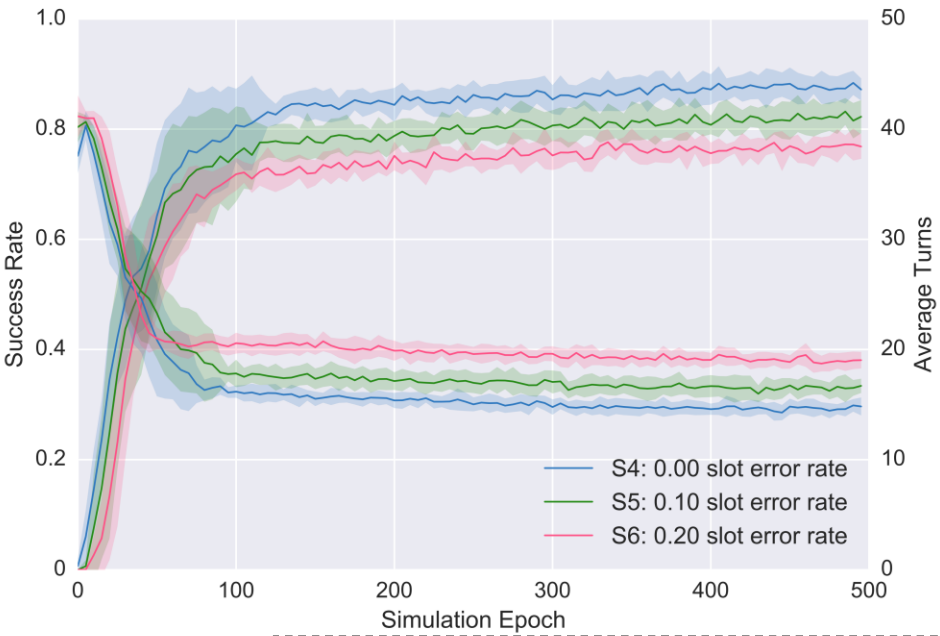
Intent errors slightly influence the RL system performance

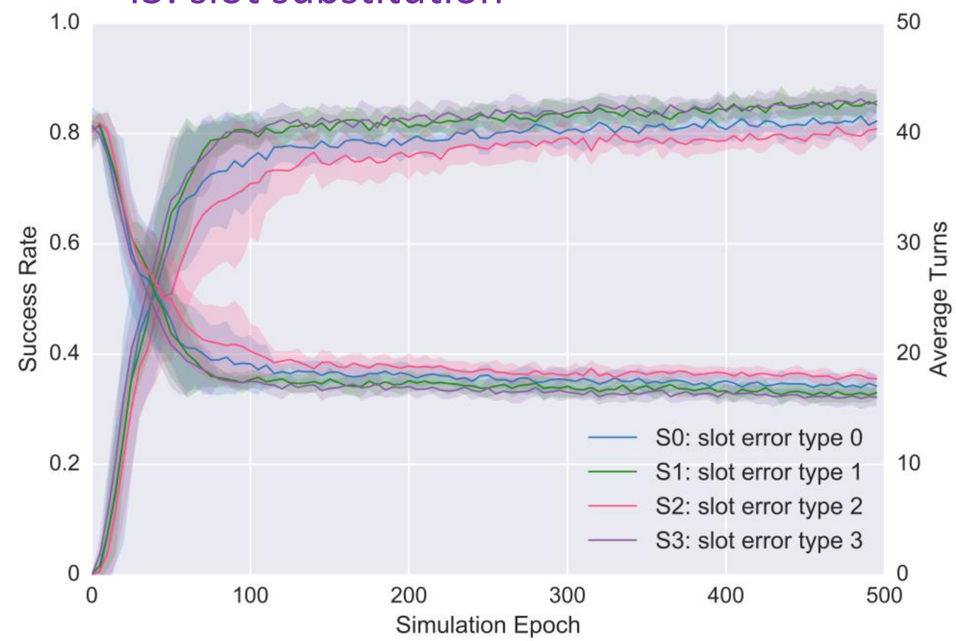
Between-group intent errors degrade the system performance more

request_moviename
(~~actor=Robert Downey Jr~~)
director Robert Downey Sr

Slot Error Analysis

- Slot error rate
  - S4: 0.00
  - S5: 0.10
  - S6: 0.20
- Slot error type
  - I0: random
  - I1: slot deletion
  - I2: value substitution
  - I3: slot substitution



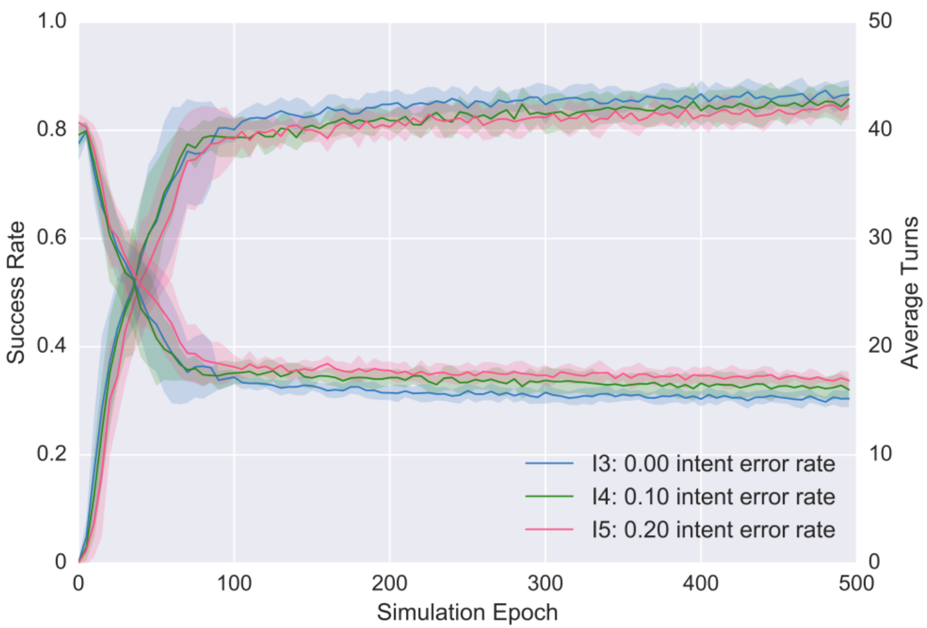


Slot errors significantly degrade the RL system performance

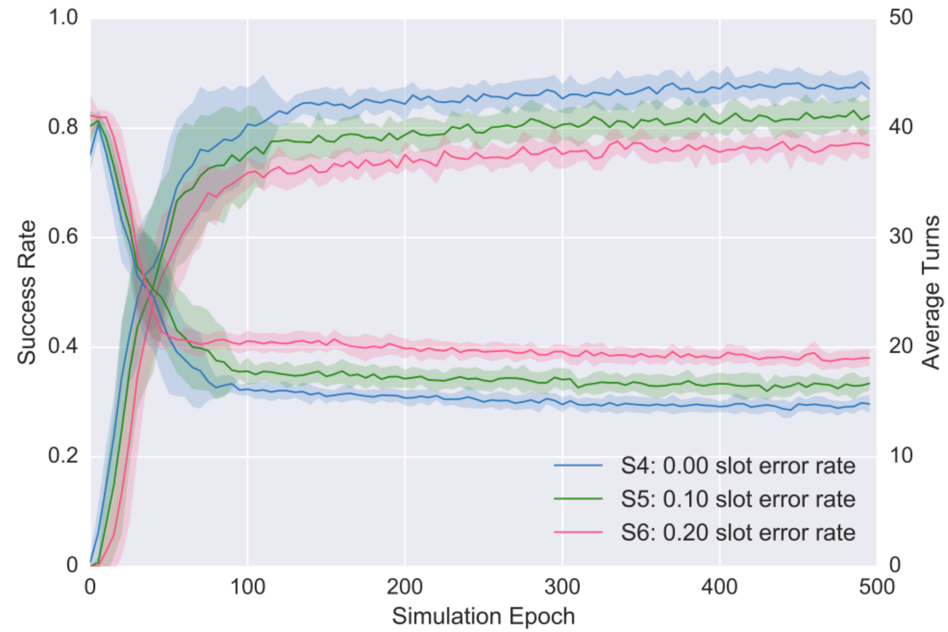
Value substitution has the largest impact on the system performance

Error Comparison

- Intent error rate



- Slot error rate



The RL agent has better robustness to intent errors in terms of dialogue-level performance

Slot filling is more important than intent detection in language understanding

Conclusion

- Word-level contexts in sentences help understand word meanings
  - Learning from Prior Knowledge –

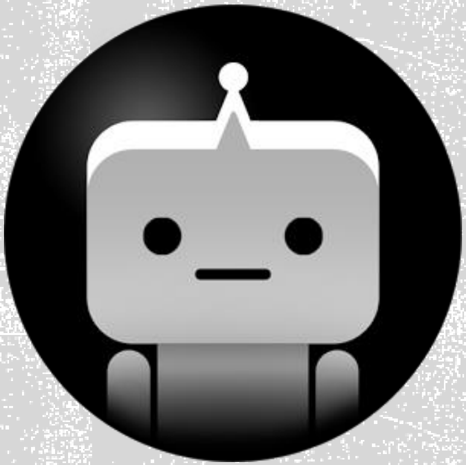
    K-SAN achieves better LU via known knowledge [Chen et al., ‘16]
  - Learning from Observations –

    MUSE learns sense embeddings with efficient sense selection [Lee & Chen, ‘17]
- Sentence-level contexts have different impacts on dialogue performance
  - Inference –

    App contexts improve personalized understanding via inference [Chen et al., ‘15]
  - Investigation of Understanding Impact –

    Slot errors degrade system performance more than intent errors [Li et al., ‘17]
- Contexts from different levels provide cues for better understanding in supervised and unsupervised ways

Thanks for Attention!



Q & A