Unsupervised Learning and Modeling of Knowledge and Intent for Spoken Dialogue Systems

Yun-Nung (Vivian) Chen

Email yvchen@cs.cmu.edu
Website http://vivianchen.idv.tw

Carnegie Mellon

OUTLINE

- Introduction
- Ontology Induction: Frame-Semantic Parsing
- Structure Learning: Knowledge Graph Propagation
- Spoken Language Understanding (SLU): Matrix Factorization
- Experiments
- Conclusions

OUTLINE

- **Introduction**
- Ontology Induction: Frame-Semantic Parsing
- Structure Learning: Knowledge Graph Propagation
- Spoken Language Understanding (SLU): Matrix Factorization
- Experiments
- Conclusions

A POPULAR ROBOT - BAYMAX

Baymax is capable of maintaining a good **spoken dialogue system** and **learning** new knowledge for better **understanding** and **interacting** with people.

Big Hero 6 -- Video content owned and licensed by Disney Entertainment, Marvel Entertainment, LLC, etc

SPOKEN DIALOGUE SYSTEM (SDS)

Spoken dialogue systems are the intelligent agents that are able to help users finish tasks more efficiently via speech interactions.

Spoken dialogue systems are being incorporated into various devices (smart-phones, smart TVs, in-car navigating system, etc).

Apple’s Siri

Microsoft’s Cortana

Microsoft’s XBOX Kinect

Amazon’s Echo

Samsung’s SMART TV

Google Now

https://www.apple.com/ios/siri/
http://www.windowsphone.com/en-us/how-to/wp8/cortana/meet-cortana
http://www.xbox.com/en-US/
http://www.amazon.com/oc/echo/
http://www.samsung.com/us/experience/smart-tv/
https://www.google.com/landing/now/

LARGE SMART DEVICE POPULATION

The number of global smartphone users will surpass 2 billion in 2016.

As of 2012, there are 1.1 billion automobiles on the earth.

The more **natural** and **convenient** input of the devices evolves towards speech

CHALLENGES FOR SDS

An SDS in a new domain requires

1) A hand-crafted domain ontology
2) Utterances labeled with semantic representations
3) An SLU component for mapping utterances into semantic representations

With increasing spoken interactions, building domain ontologies and annotating utterances cost a lot so that the data does not scale up.

The goal is to enable an SDS to automatically learn this knowledge so that open domain requests can be handled.

INTERACTION EXAMPLE

User

find an inexpensive eating place for taiwanese food

Inexpensive Taiwanese eating places include Din Tai Fung, etc. What do you want to choose? I can help you go there.

Q: How does a dialogue system process this request?

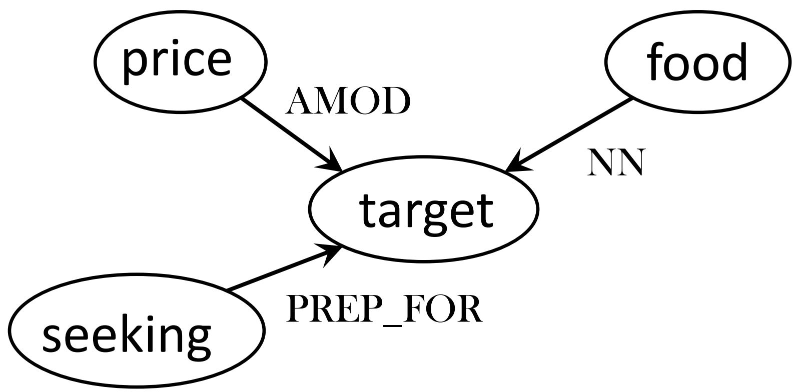
Intelligent Agent

SDS PROCESS – AVAILABLE DOMAIN ONTOLOGY

User

find an inexpensive eating place for taiwanese food

Intelligent Agent



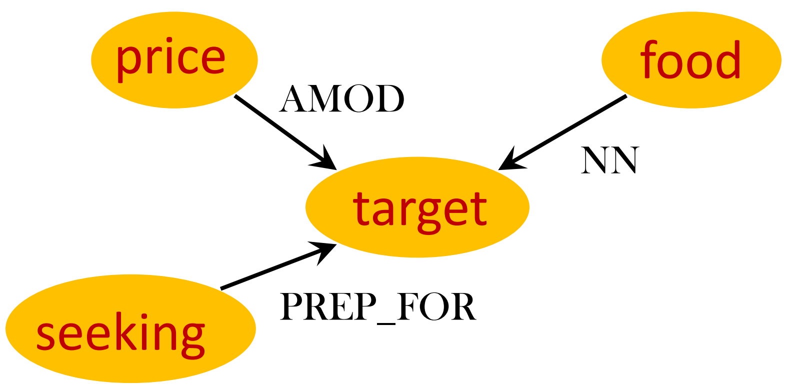
Organized Domain Knowledge

SDS PROCESS – AVAILABLE DOMAIN ONTOLOGY

User

find an inexpensive eating place for taiwanese food

Ontology Induction *(semantic slot)*



Intelligent Agent

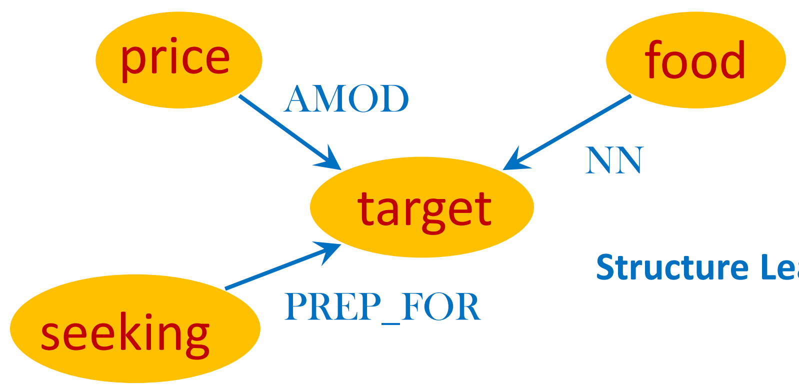
Organized Domain Knowledge

SDS PROCESS – AVAILABLE DOMAIN ONTOLOGY

User

find an inexpensive eating place for taiwanese food

Ontology Induction *(semantic slot)*



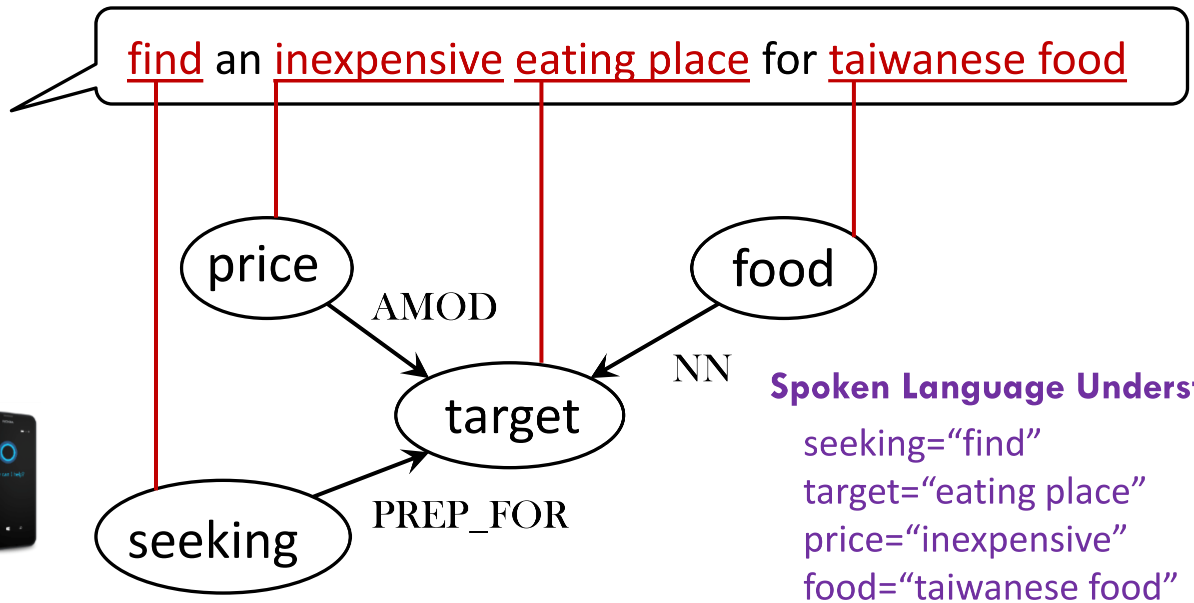
Structure Learning *(inter-slot relation)*

Intelligent Agent

Organized Domain Knowledge

SDS PROCESS – SPOKEN LANGUAGE UNDERSTANDING (SLU)

User



Spoken Language Understanding

seeking="find"
target="eating place"
price="inexpensive"
food="taiwanese food"

Intelligent Agent

Organized Domain Knowledge

SDS PROCESS – DIALOGUE MANAGEMENT (DM)

User

find an inexpensive eating place for taiwanese food

SELECT restaurant {
 restaurant.price="inexpensive"
 restaurant.food="taiwanese food"
}

Din Tai Fung
Boiling Point
:
:

Predicted behavior: navigation

Behavior Prediction

Intelligent Agent

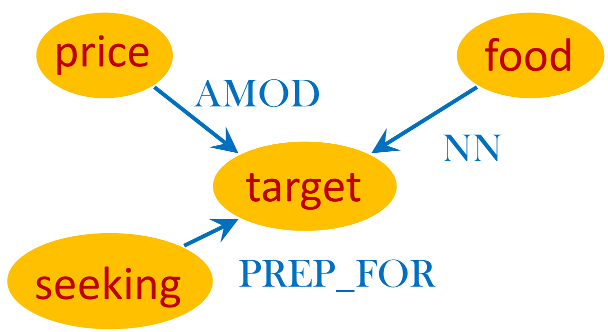
Inexpensive Taiwanese eating places include Din Tai Fung, Boiling Point, etc. What do you want to choose? I can help you go there. (navigation)

GOALS

User

find an inexpensive eating place for taiwanese food

1. Ontology Induction *(semantic slot)*



2. Structure Learning

(inter-slot relation)

3. Spoken Language Understanding

SELECT restaurant {
restaurant.price="inexpensive"
restaurant.food="taiwanese food"
}

Predicted behavior: navigation

4. Behavior Prediction

GOALS

User

find an inexpensive eating place for taiwanese food

1. Ontology Induction
2. Structure Learning

Knowledge Acquisition

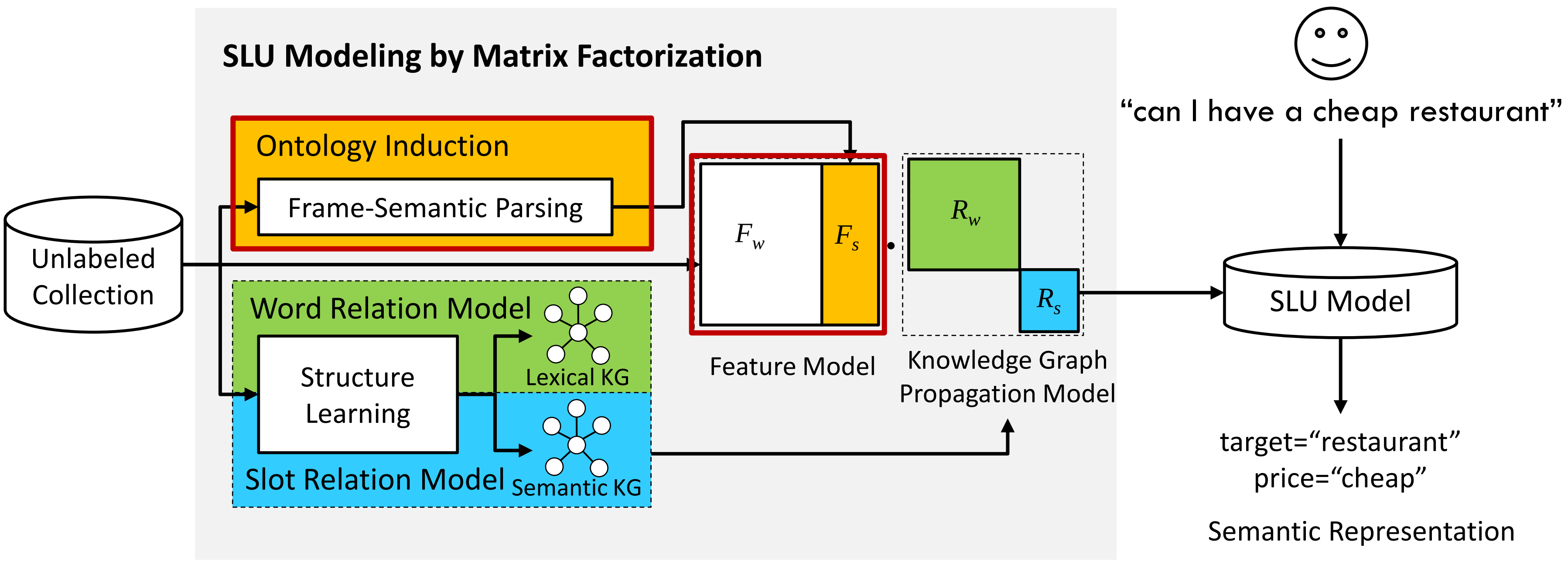
3. Semantic Decoding
4. Behavior Prediction

SLU Modeling

SPOKEN LANGUAGE UNDERSTANDING

Input: user utterances

Output: the domain-specific semantic concepts included in each utterance



OUTLINE

- Introduction
- **Ontology Induction: Frame-Semantic Parsing**
- Structure Learning: Knowledge Graph Propagation
- Spoken Language Understanding (SLU): Matrix Factorization
- Experiments
- Conclusions

PROBABILISTIC FRAME-SEMANTIC PARSING

FrameNet [Baker et al., 1998]

- a linguistically semantic resource, based on the frame-semantics theory
- words/phrases can be represented as frames
- “low fat milk” → “milk” evokes the “food” frame;
  “low fat” fills the descriptor frame element

SEMAFOR [Das et al., 2014]

- a state-of-the-art frame-semantics parser, trained on manually annotated FrameNet sentences

Baker et al., "The berkeley framenet project," in Proc. of International Conference on Computational linguistics, 1998.
Das et al., " Frame-semantic parsing," in Proc. of Computational Linguistics, 2014.

FRAME-SEMANTIC PARSING FOR UTTERANCES

Good!

Good!

?

can i have a cheap restaurant

Frame: capability
FT LU: can FE LU: i

Frame: expensiveness
FT LU: cheap

Frame: locale by use
FT/FE LU: restaurant

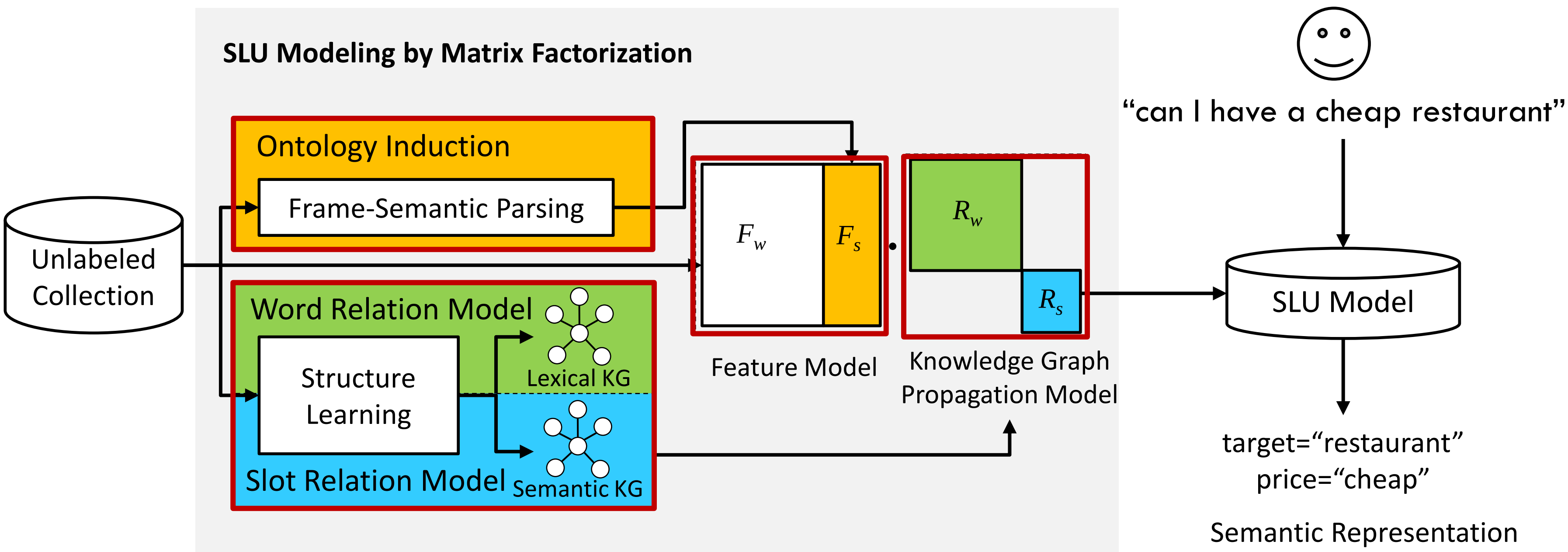
FT: Frame Target; FE: Frame Element; LU: Lexical Unit

1st Issue: adapting ***generic*** frames to ***domain-specific*** settings for SDSs

SPOKEN LANGUAGE UNDERSTANDING

Input: user utterances

Output: the domain-specific semantic concepts included in each utterance



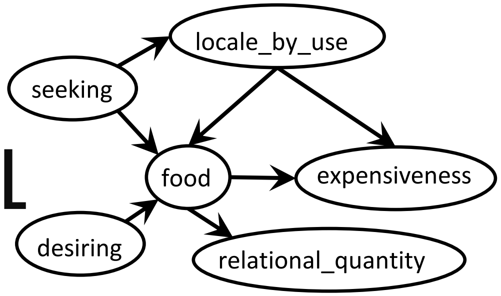
Y.-N. Chen et al., "Matrix Factorization with Knowledge Graph Propagation for Unsupervised Spoken Language Understanding," in *Proc. of ACL-IJCNLP*, 2015.

OUTLINE

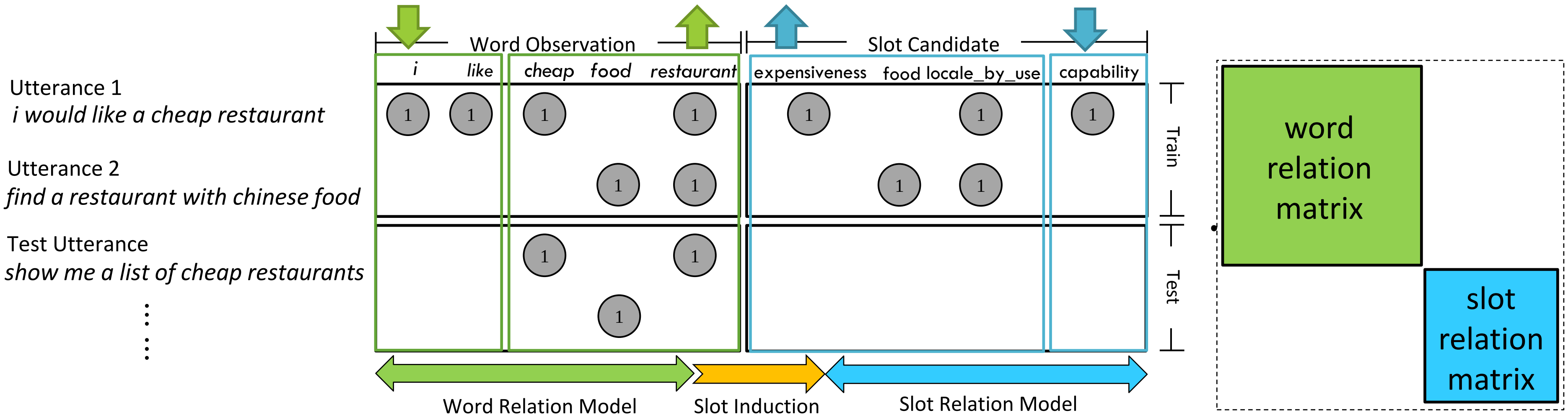
- Introduction
- Ontology Induction: Frame-Semantic Parsing
- **Structure Learning: Knowledge Graph Propagation**
  (for 1st issue)
- Spoken Language Understanding (SLU): Matrix Factorization
- Experiments
- Conclusions

1ST ISSUE: HOW TO ADAPT GENERIC SLOTS TO DOMAIN-SPECIFIC SETTING?

KNOWLEDGE GRAPH PROPAGATION MODEL



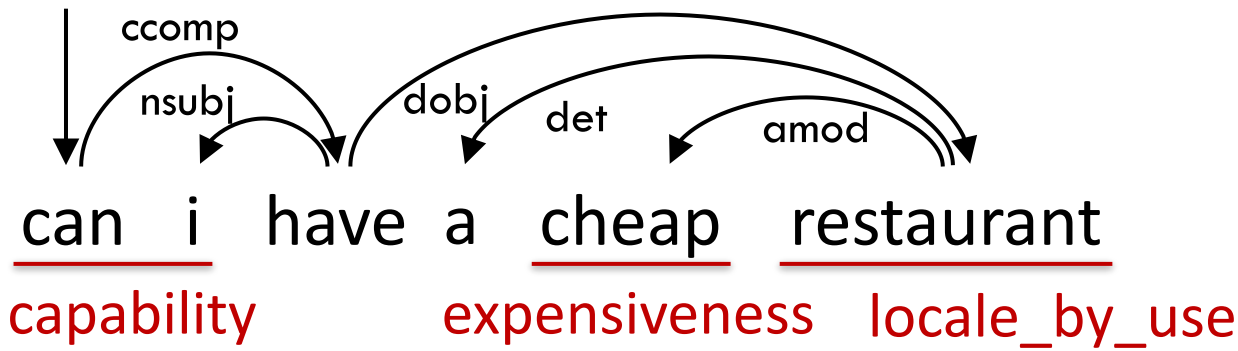
Assumption: The domain-specific words/slots have more dependency to each other.

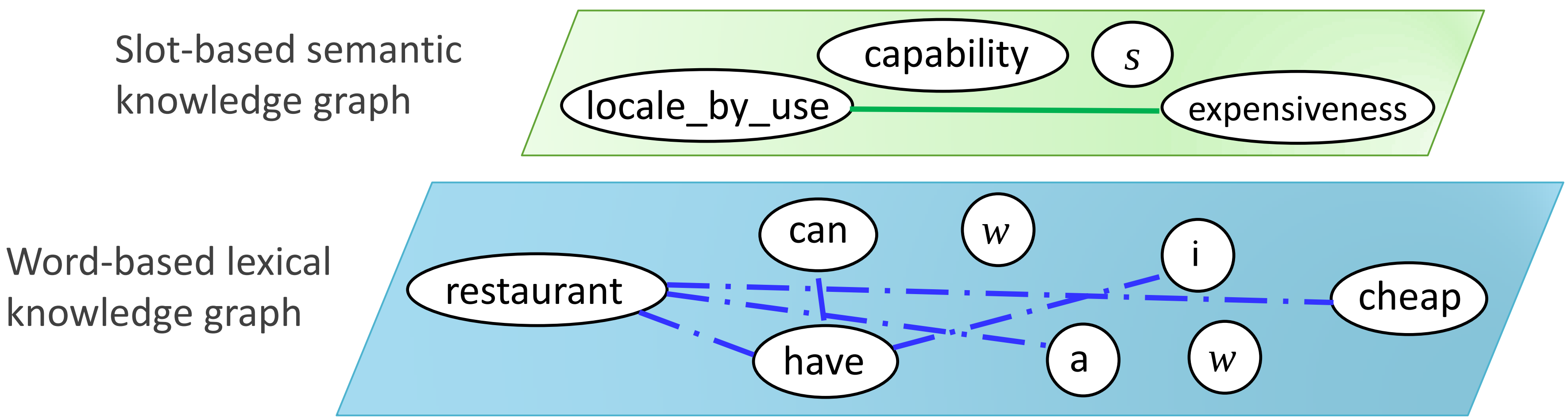


Relation matrices allow each node to propagate scores to its neighbors in the knowledge graph, so that domain-specific words/slots have higher scores after matrix multiplication.

KNOWLEDGE GRAPH CONSTRUCTION

Syntactic dependency parsing on utterances

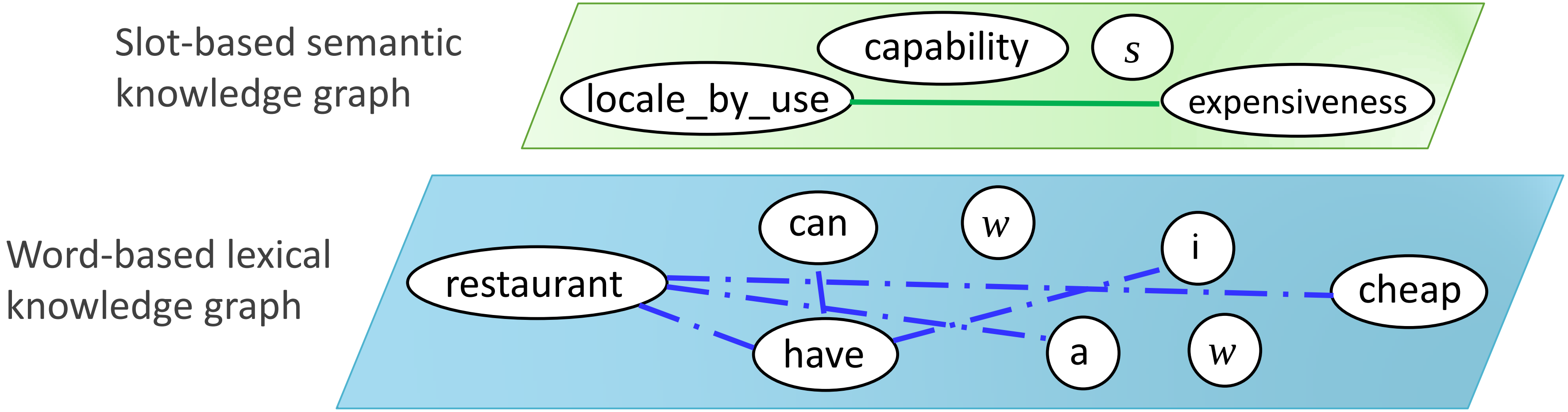




KNOWLEDGE GRAPH CONSTRUCTION

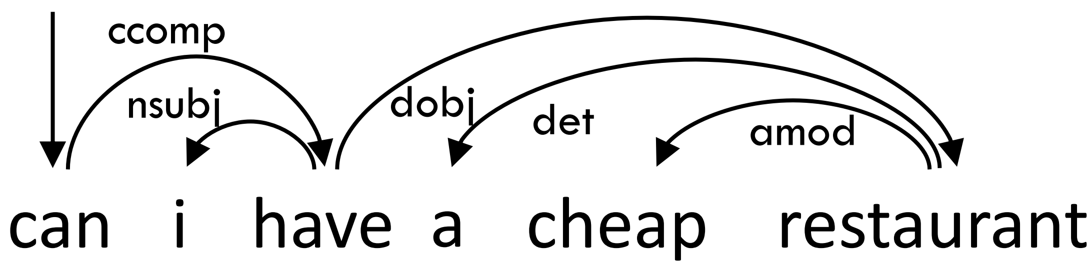
The edge between a node pair is weighted as relation importance to propagate the scores via a relation matrix

How to decide the weights to represent relation importance?



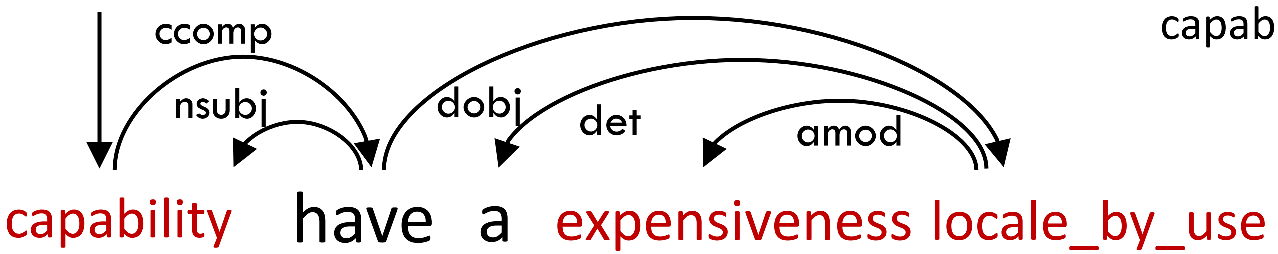
WEIGHT MEASUREMENT BY EMBEDDINGS

Dependency-based word embeddings



can = $[0.8 \quad \ldots \quad 0.24]$
have = $[0.3 \quad \ldots \quad 0.21]$
:
:

Dependency-based slot embeddings



expensiveness = $[0.12 \quad \ldots \quad 0.7]$
capability = $[0.3 \quad \ldots \quad 0.6]$
:
:

Levy and Goldberg, " Dependency-Based Word Embeddings," in *Proc. of ACL*, 2014.

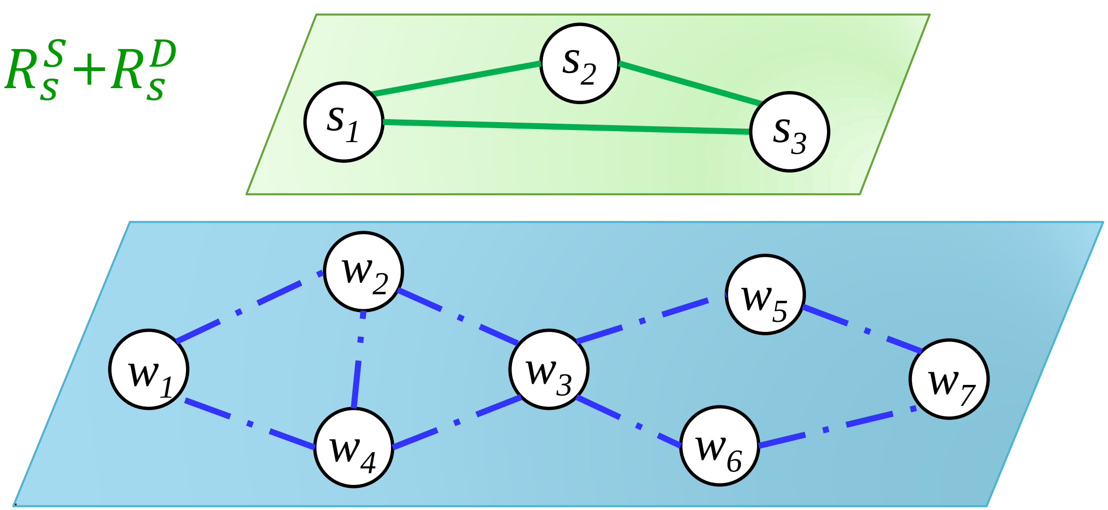
WEIGHT MEASUREMENT BY EMBEDDINGS

Compute edge weights to represent relation importance

- Slot-to-slot semantic relation R_s^S: similarity between slot embeddings
- Slot-to-slot dependency relation R_s^D: dependency score between slot embeddings
- Word-to-word semantic relation R_w^S: similarity between word embeddings
- Word-to-word dependency relation R_w^D: dependency score between word embeddings

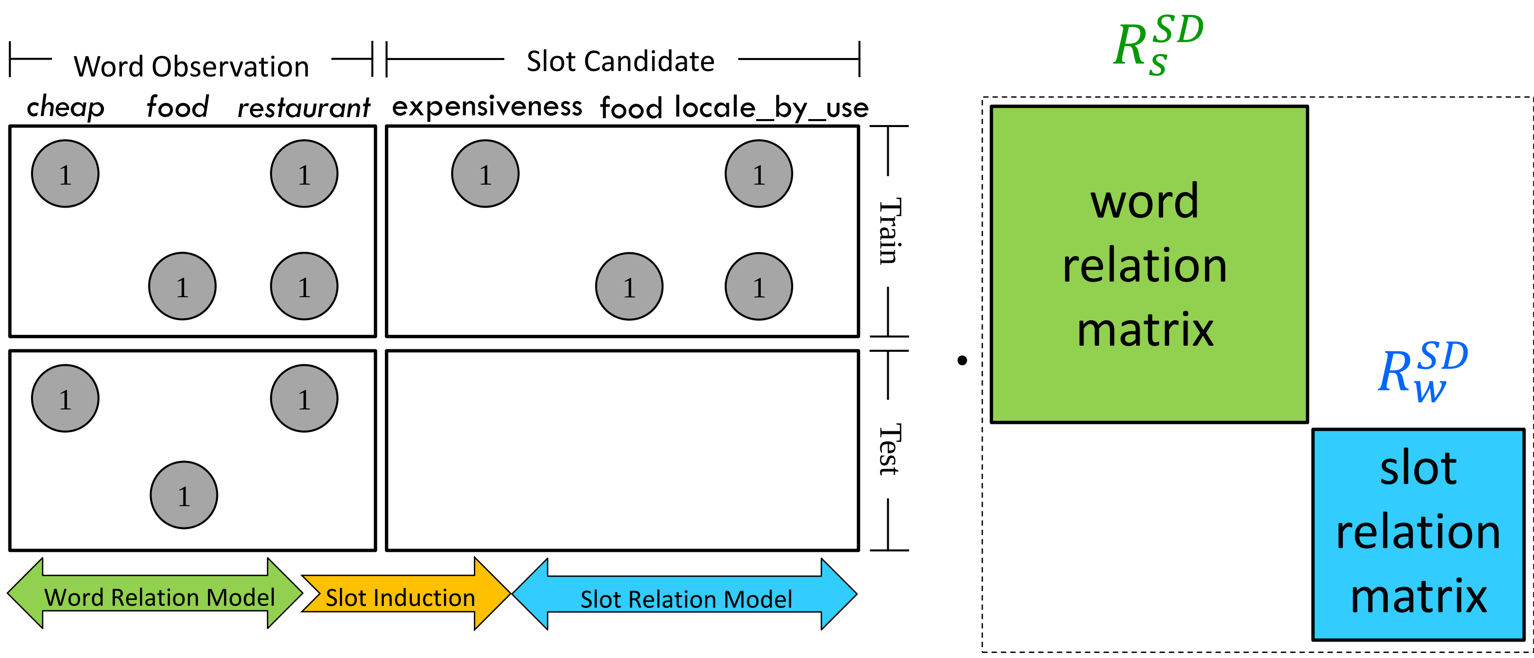
$$R_s^{SD} = R_s^S + R_s^D$$

$$R_w^{SD} = R_w^S + R_w^D$$

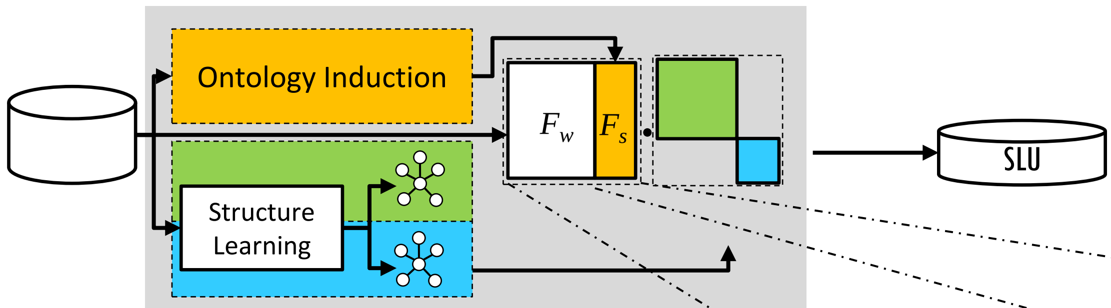


Y.-N. Chen et al., “Jointly Modeling Inter-Slot Relations by Random Walk on Knowledge Graphs for Unsupervised Spoken Language Understanding," in *Proc. of NAACL*, 2015.

KNOWLEDGE GRAPH PROPAGATION MODEL



FEATURE MODEL



Utterance 1

i would like a cheap restaurant

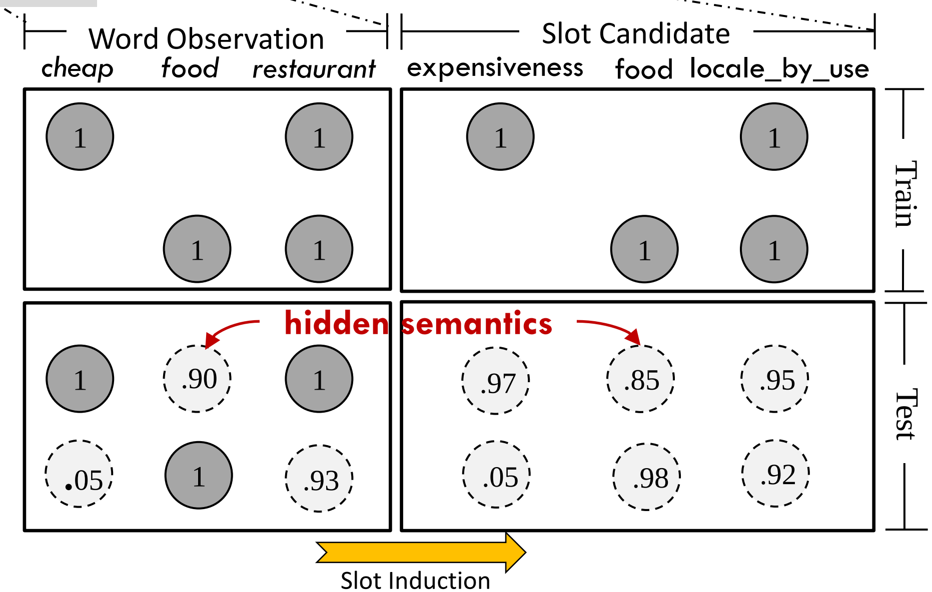
Utterance 2

find a restaurant with chinese food

Test Utterance

show me a list of cheap restaurants

⋮



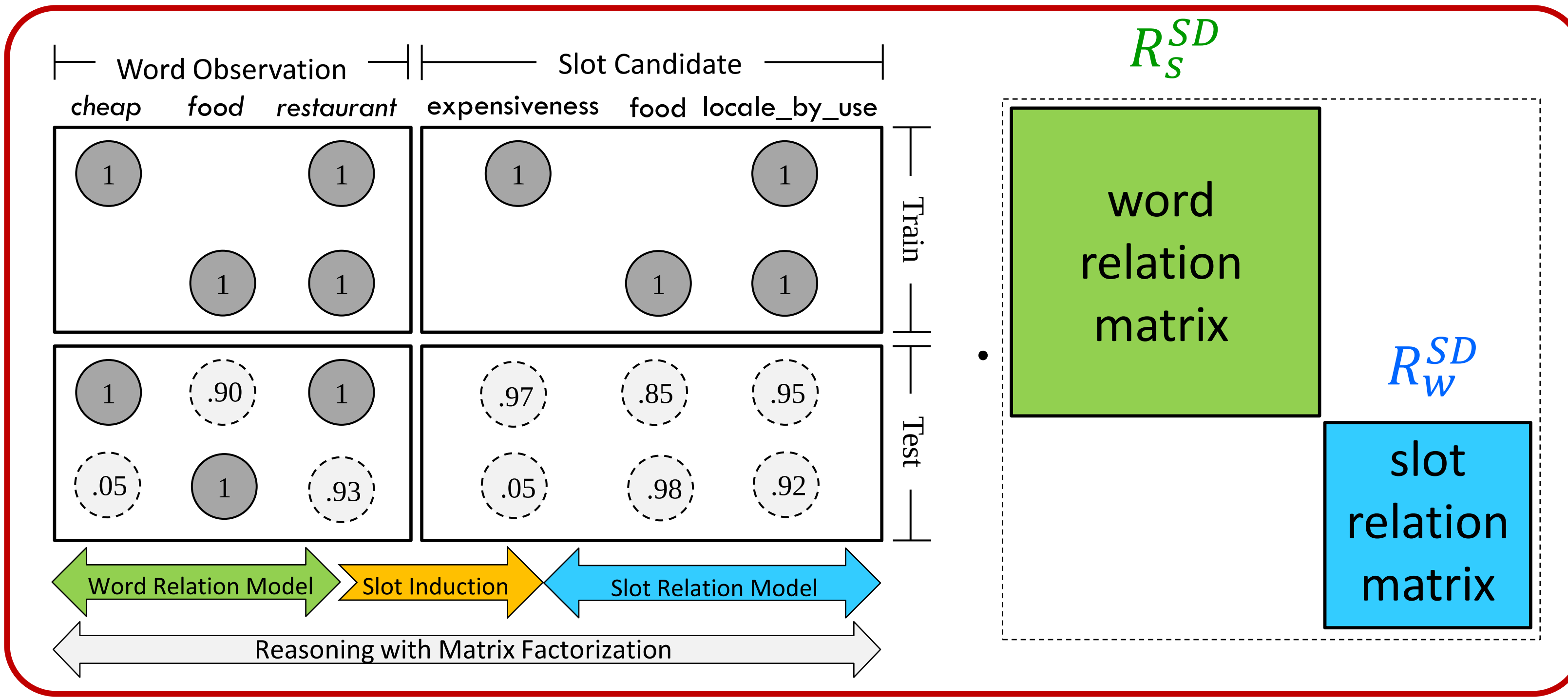
2nd Issue: unobserved hidden semantics may benefit understanding

OUTLINE

- Introduction
- Ontology Induction: Frame-Semantic Parsing
- Structure Learning: Knowledge Graph Propagation
- **Spoken Language Understanding (SLU): Matrix Factorization**
  (for 2nd issue)
- Experiments
- Conclusions

2ND ISSUE: HOW TO LEARN IMPLICIT SEMANTICS?

MATRIX FACTORIZATION (MF)

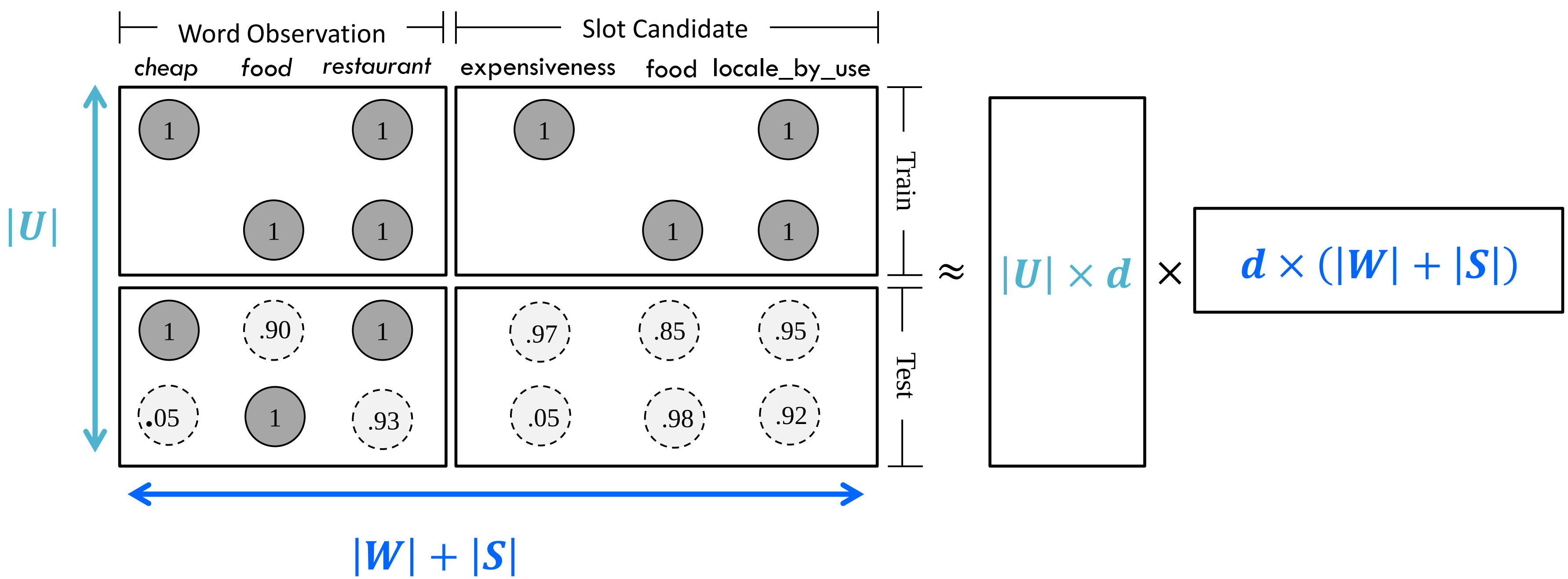


MF method completes a partially-missing matrix based on a low-rank latent semantics assumption.

MATRIX FACTORIZATION (MF)

The decomposed matrices represent low-rank latent semantics for utterances and words/slots respectively

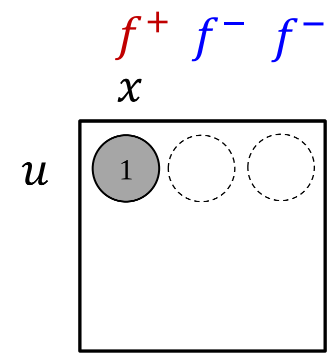
The product of two matrices fills the probability of hidden semantics



BAYESIAN PERSONALIZED RANKING FOR MF

Model implicit feedback

- not treat unobserved facts as negative samples (true or false)
- give observed facts higher scores than unobserved facts



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

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

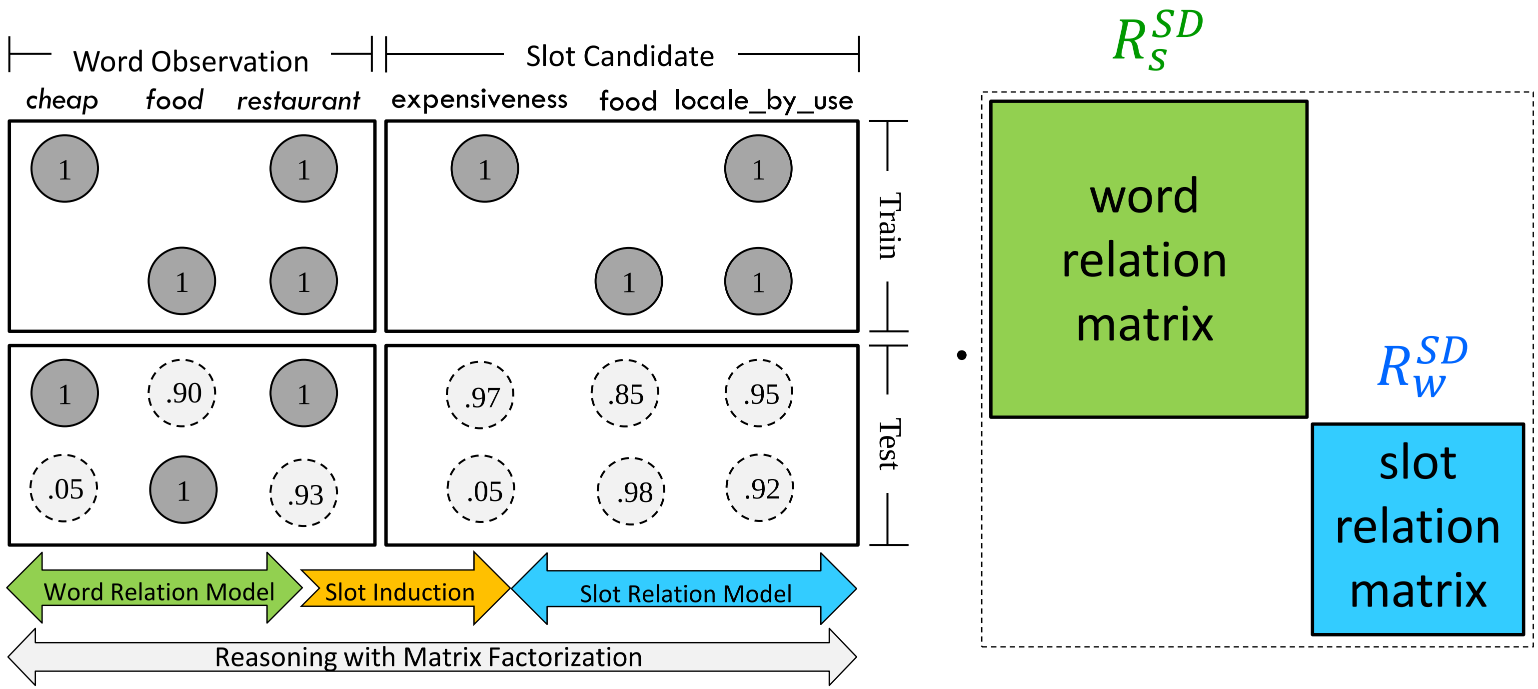
Objective:

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

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

2ND ISSUE: HOW TO LEARN IMPLICIT SEMANTICS?

MATRIX FACTORIZATION (MF)



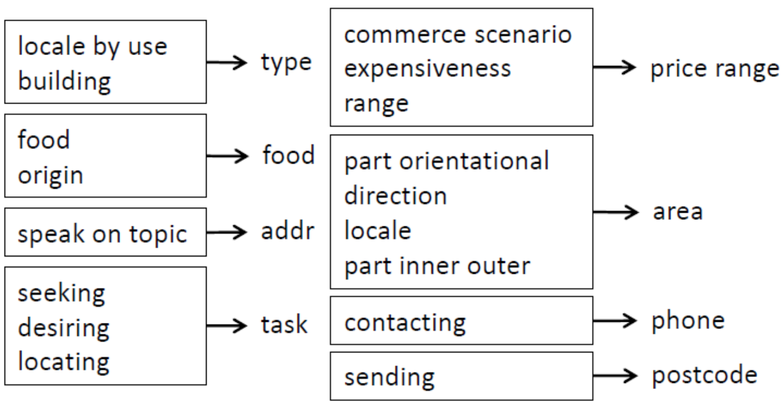
MF method completes a partially-missing matrix based on a low-rank latent semantics assumption.

OUTLINE

EXPERIMENTAL SETUP

Dataset

- Cambridge University SLU corpus [Henderson, 2012]
  - Restaurant recommendation in an in-car setting in Cambridge
    - WER = 37%
    - vocabulary size = 1868
    - 2,166 dialogues
    - 15,453 utterances
    - dialogue slot: addr, area, food, name, phone, postcode, price range, task, type



The mapping table between induced and reference slots

Henderson et al., "Discriminative spoken language understanding using word confusion networks," in *Proc. of SLT*, 2012.

EXPERIMENT 1: QUALITY OF SEMANTICS ESTIMATION

Metric: Mean Average Precision (MAP) of all estimated slot probabilities for each utterance

<table>
<tr><th colspan="2" rowspan="2">Approach</th><th colspan="2">ASR</th><th colspan="2">Manual</th></tr>
<tr><th>w/o</th><th>w/ Explicit</th><th>w/o</th><th>w/ Explicit</th></tr>
<tr><td rowspan="2">Explicit</td><td>Support Vector Machine</td><td colspan="2">32.5</td><td colspan="2">36.6</td></tr>
<tr><td>Multinomial Logistic Regression</td><td colspan="2">34.0</td><td colspan="2">38.8</td></tr>
</table>

EXPERIMENT 1: QUALITY OF SEMANTICS ESTIMATION

Metric: Mean Average Precision (MAP) of all estimated slot probabilities for each utterance

Modeling Implicit Semantics

Approach			ASR		Manual	
			w/o	w/ Explicit	w/o	w/ Explicit
Explicit	Support Vector Machine		32.5		36.6	
	Multinomial Logistic Regression		34.0		38.8	
Implicit	Baseline	Random				
		Majority				
	MF	Feature Model				
		Feature Model + Knowledge Graph Propagation				

EXPERIMENT 1: QUALITY OF SEMANTICS ESTIMATION

Metric: Mean Average Precision (MAP) of all estimated slot probabilities for each utterance

Modeling Implicit Semantics

Approach			ASR		Manual	
			w/o	w/ Explicit	w/o	w/ Explicit
Explicit	Support Vector Machine		32.5		36.6	
	Multinomial Logistic Regression		34.0		38.8	
Implicit	Baseline	Random	3.4		2.6	
		Majority	15.4		16.4	
	MF	Feature Model	24.2		22.6	
		Feature Model + Knowledge Graph Propagation	**40.5*** **(+19.1%)**		**52.1*** **(+34.3%)**	

EXPERIMENT 1: QUALITY OF SEMANTICS ESTIMATION

Metric: Mean Average Precision (MAP) of all estimated slot probabilities for each utterance

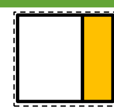
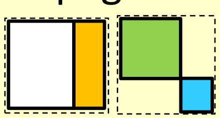
Approach			ASR		Manual	
			w/o	w/ Explicit	w/o	w/ Explicit
Explicit	Support Vector Machine		32.5		36.6	
	Multinomial Logistic Regression		34.0		38.8	
Implicit	Baseline	Random	3.4	22.5	2.6	25.1
		Majority	15.4	32.9	16.4	38.4
	MF	Feature Model	24.2	37.6*	22.6	45.3*
		Feature Model + Knowledge Graph Propagation	**40.5*** (+19.1%)	**43.5*** (+27.9%)	**52.1*** (+34.3%)	**53.4*** (+37.6%)

Modeling Implicit Semantics

The MF approach effectively models hidden semantics to improve SLU.

Adding a knowledge graph propagation model further improves performance.

EXPERIMENT 2: EFFECTIVENESS OF RELATIONS

Approach		ASR	Manual
Feature Model 		37.6	45.3
Feature + Knowledge Graph Propagation 	Semantic $\begin{bmatrix} R_w^S & 0 \\ 0 & R_s^S \end{bmatrix}$	41.4*	51.6*
	Dependency $\begin{bmatrix} R_w^D & 0 \\ 0 & R_s^D \end{bmatrix}$	41.6*	49.0*
	Word $\begin{bmatrix} R_w^{SD} & 0 \\ 0 & 0 \end{bmatrix}$	39.2*	45.2
	Slot $\begin{bmatrix} 0 & 0 \\ 0 & R_s^{SD} \end{bmatrix}$	42.1*	49.9*
	Both $\begin{bmatrix} R_w^{SD} & 0 \\ 0 & R_s^{SD} \end{bmatrix}$		

All types of relations are useful to infer hidden semantics.

EXPERIMENT 2: EFFECTIVENESS OF RELATIONS

Approach			ASR	Manual
Feature Model			37.6	45.3
Feature + Knowledge Graph Propagation	Semantic	$\begin{bmatrix} R_w^S & 0 \\ 0 & R_s^S \end{bmatrix}$	41.4*	51.6*
	Dependency	$\begin{bmatrix} R_w^D & 0 \\ 0 & R_s^D \end{bmatrix}$	41.6*	49.0*
	Word	$\begin{bmatrix} R_w^{SD} & 0 \\ 0 & 0 \end{bmatrix}$	39.2*	45.2
	Slot	$\begin{bmatrix} 0 & 0 \\ 0 & R_s^{SD} \end{bmatrix}$	42.1*	49.9*
	Both	$\begin{bmatrix} R_w^{SD} & 0 \\ 0 & R_s^{SD} \end{bmatrix}$	**43.5*** **(+15.7%)**	**53.4*** **(+17.9%)**

All types of relations are useful to infer hidden semantics.

Combining different relations further improves the performance.

OUTLINE

- Introduction
- Ontology Induction: Frame-Semantic Parsing
- Structure Learning: Knowledge Graph Propagation
- Spoken Language Understanding (SLU): Matrix Factorization
- Experiments
- **Conclusions**

CONCLUSIONS

Ontology induction and **knowledge graph construction** enable systems to automatically acquire open domain knowledge.

MF for SLU provides a principle model that is able to

- unify the automatically acquired knowledge
- adapt to a domain-specific setting
- and then allows systems to consider implicit semantics for better understanding.

The work shows the feasibility and the potential of improving *generalization*, *maintenance*, *efficiency*, and *scalability* of SDSs.

The proposed unsupervised SLU achieves 43% of MAP on ASR-transcribed conversations.

Q & A

Thanks for your attentions!!