DERIVING LOCAL RELATIONAL SURFACE FORMS FROM DEPENDENCY-BASED ENTITY EMBEDDINGS FOR UNSUPERVISED SPOKEN LANGUAGE UNDERSTANDING

YUN-NUNG (VIVIAN) CHEN & DILEK HAKKANI-TÜR

My Background

Yun-Nung (Vivian) Chen

PhD student advised by Prof. Alexander I Rudnicky

Language Technologies Institute, School of Computer Science, Carnegie Mellon University

Research focus: spoken dialogue system, unsupervised spoken language understanding

Outline

Introduction

- Main Idea
- Semantic Knowledge Graph
- Semantic Interpretation via Relation

Proposed Approach

- Relation Inference from Gazetteers
- Relational Surface Form Derivation
- Probabilistic Enrichment
- Boostrapping

Experiments

Conclusions

Ongoing & Future Work

Main Idea

Relation Detection for Unsupervised SLU

Spoken Language Understanding (SLU): convert automatic speech recognition (ASR) outputs into pre-defined semantic output format

"when was james cameron's avatar released"

Intent: FIND_RELEASE_DATE
Slot-Val: MOVIE_NAME="avatar", DIRECTOR_NAME="james cameron"

Relation: semantic interpretation of input utterances

- movie.release_date, movie.name, movie.directed_by, director.name

Unsupervised SLU: utilize external knowledge to help relation detection without labelled data

Semantic Knowledge Graph

Priors for SLU

What are knowledge graphs?

- Graphs with
  - strongly typed and uniquely identified entities (nodes)
  - facts/literals connected by relations (edge)

Examples:

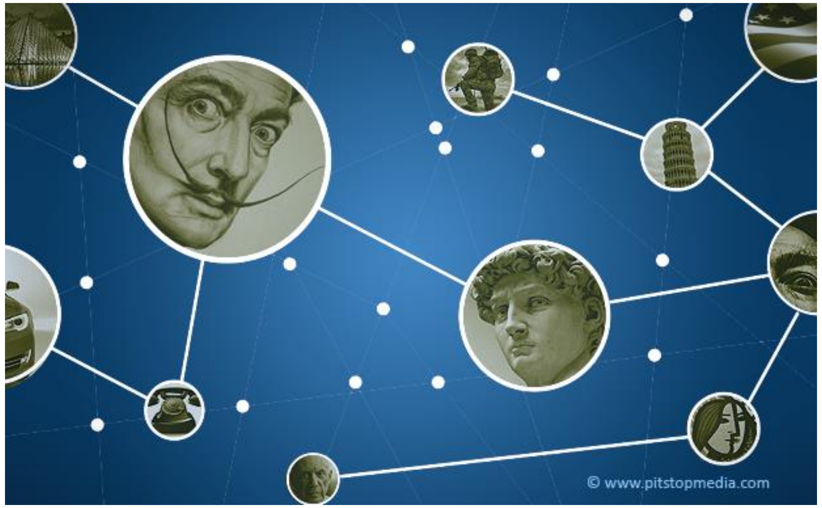
- Satori, Google KG, Facebook Open Graph, Freebase

How large?

- > 500M entities, >1.5B relations, > 5B facts

How broad?

- Wikipedia-breadth: “American Football” ←→ “Zoos”



- Slides of Larry Heck, Dilek Hakkani-Tur, and Gokhan Tur, Leveraging Knowledge Graphs for Web-Scale Unsupervised Semantic Parsing, in *Proceedings of Interspeech,* 2013.

Semantic Interpretation via Relations

Two Examples

- differentiate two examples by including the **originating node types** in the relation

User Utterance:

find movies produced by james cameron

SPARQL Query (simplified):

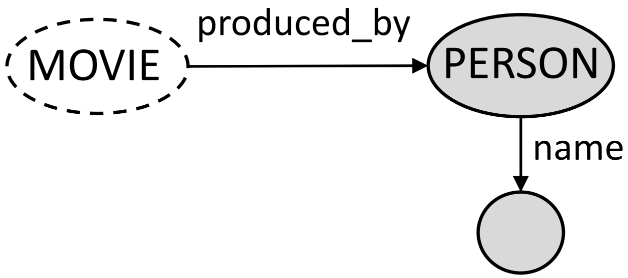
SELECT ?movie {?movie. ?movie.produced_by?producer. ?producer.name"James Cameron".}

Logical Form:

λx. ∃y. movie.produced_by(x, y) ∧ person.name(y, z) ∧ z="James Cameron"

Relation:

movie.produced_by producer.name



User Utterance:

who produced avatar

SPARQL Query (simplified):

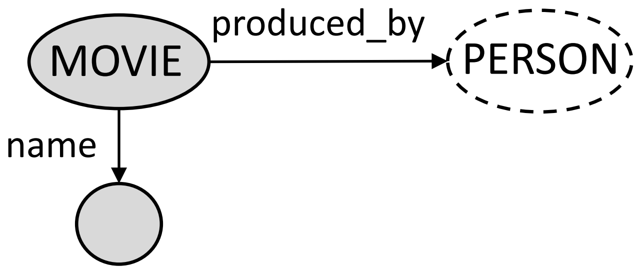
SELECT ?producer {?movie.name"Avatar". ?movie.produced_by?producer.}

Logical Form:

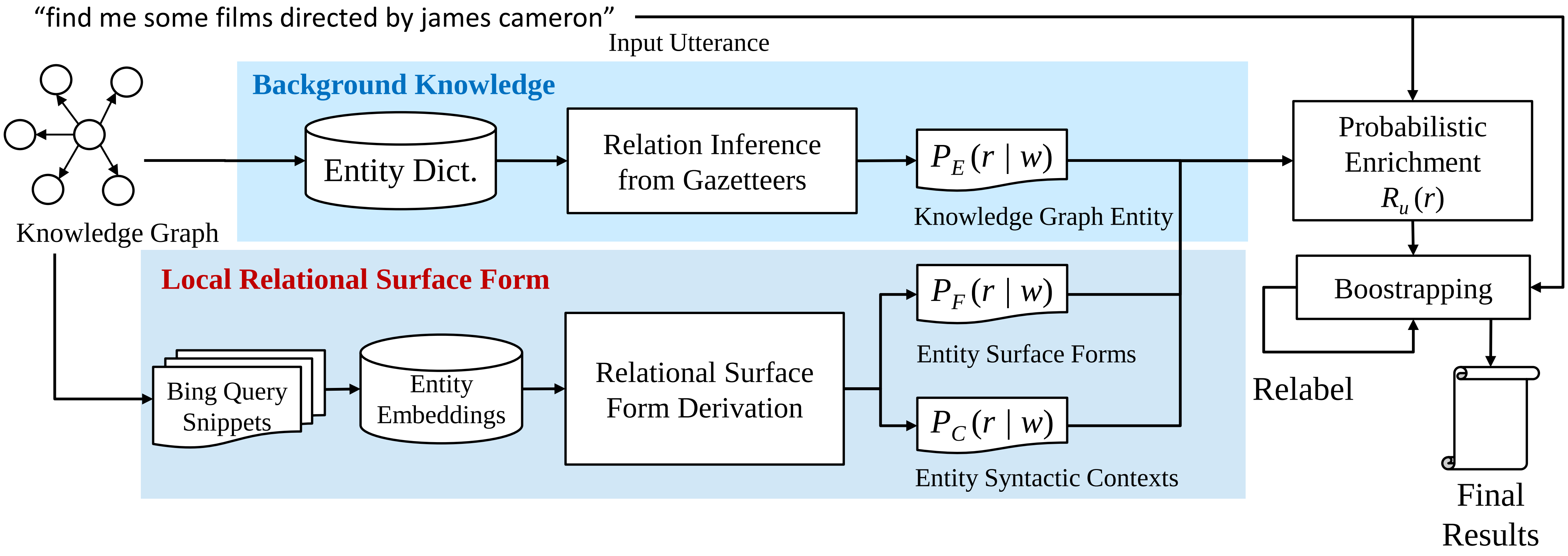
λy. ∃x. movie.produced_by(x, y) ∧ movie.name(x, z) ∧ z="Avatar"

Relation:

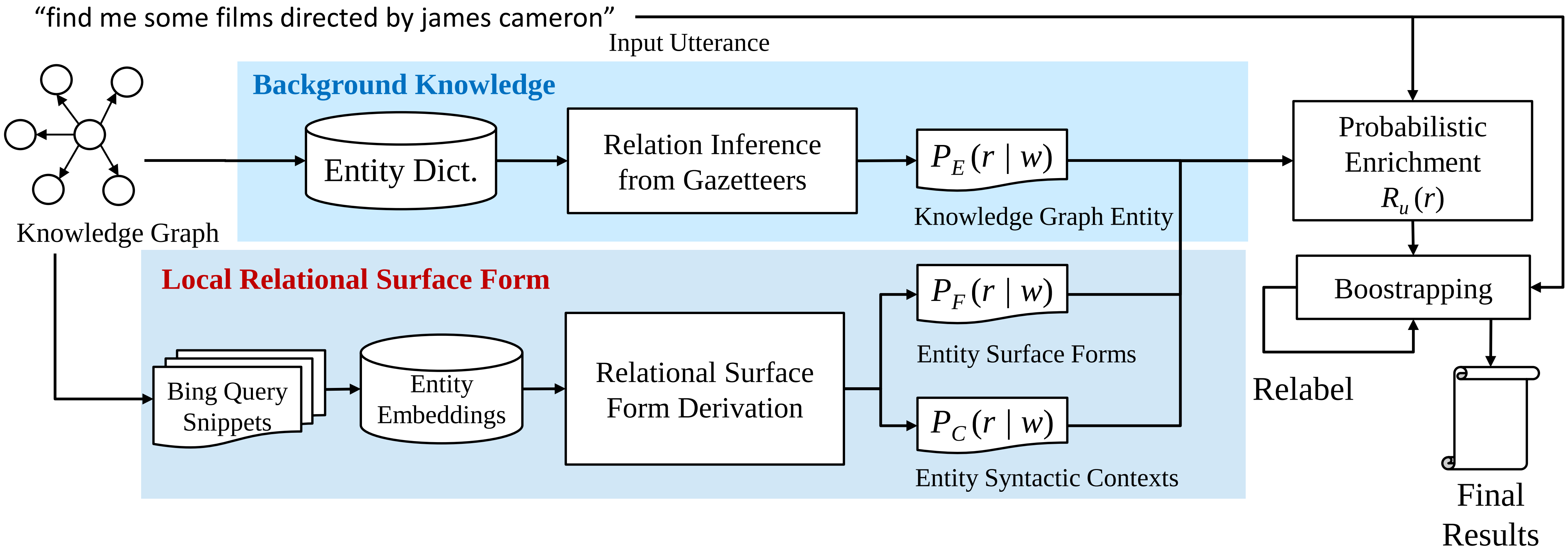
movie.name movie.produced_by



Proposed Framework



Proposed Framework



Relation Inference from Gazetteers

Gazetteers (entity lists)

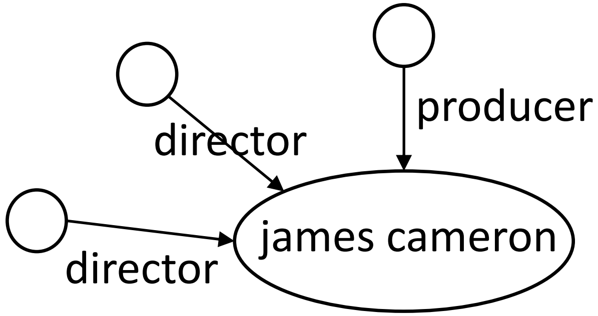
“james cameron”

#movies James Cameron directed

$$P_E(t_i \mid w) = \frac{\boxed{C(w, t_i)}}{\sum_{t_k \in T(w)} C(w, t_k)}$$

director

director
producer
:



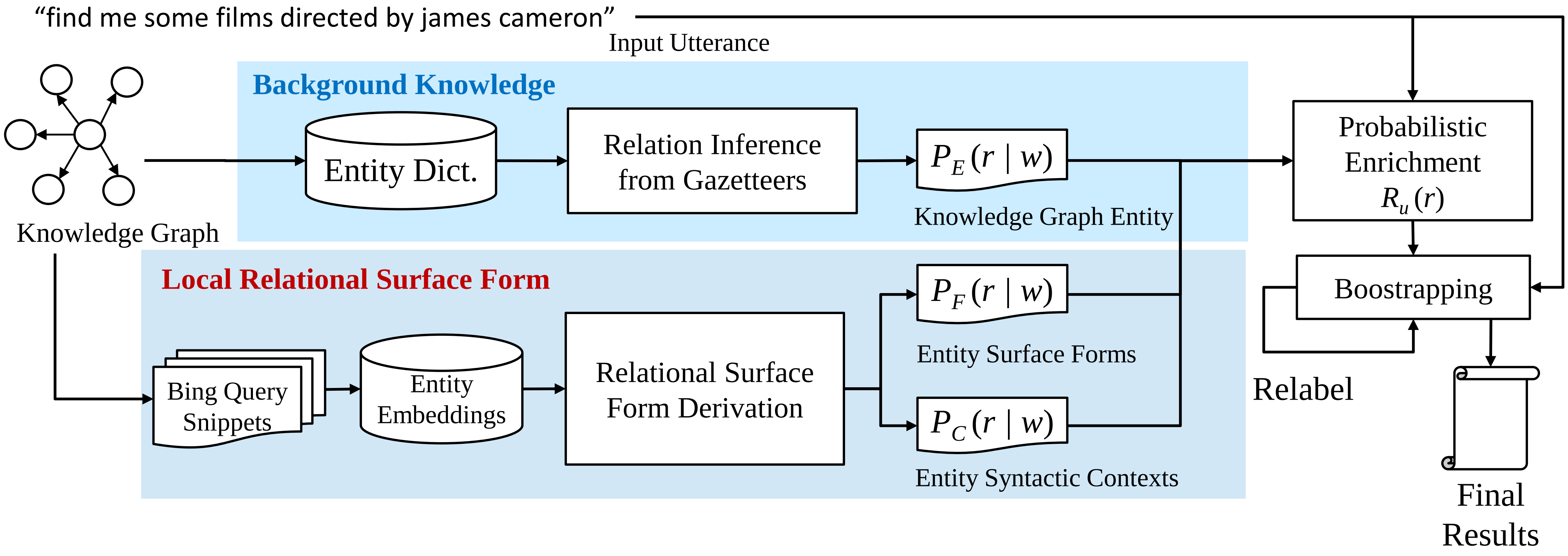
$$P_E(r_i \mid w) = P_E(t_i \mid w)$$

movie.directed_by
director.name

director

- Dilek Hakkani-Tur, Asli Celikyilmaz, Larry Heck, and Gokhan Tur, Probabilistic enrichment of knowledge graph entities for relation detection in conversational understanding, in *Proceedings of Interspeech,* 2014.

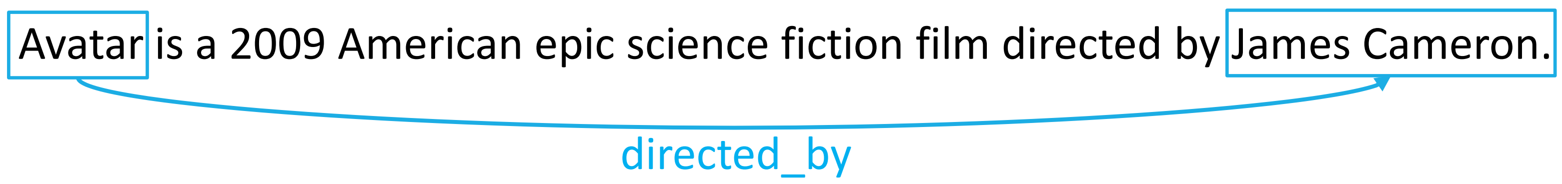
Proposed Framework



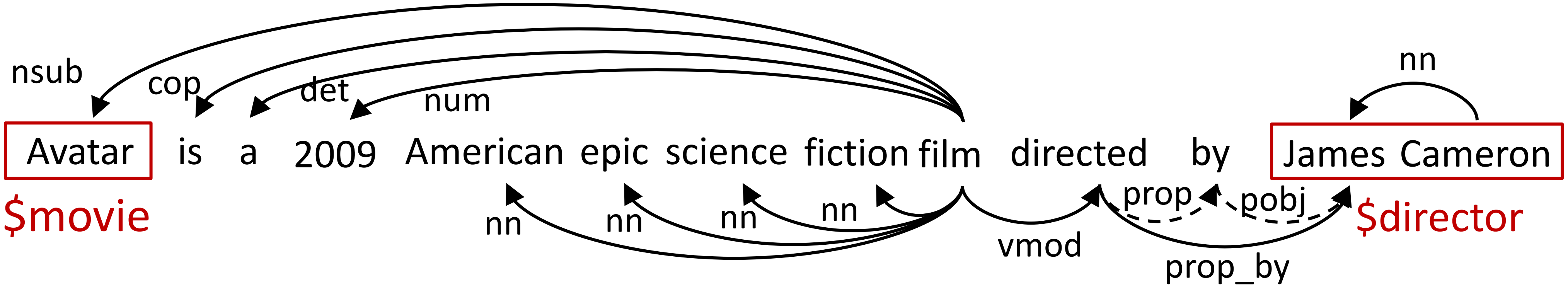
Relational Surface Form Derivation

Web Resource Mining

Bing query snippets including entity pairs connected with specific relations in KG



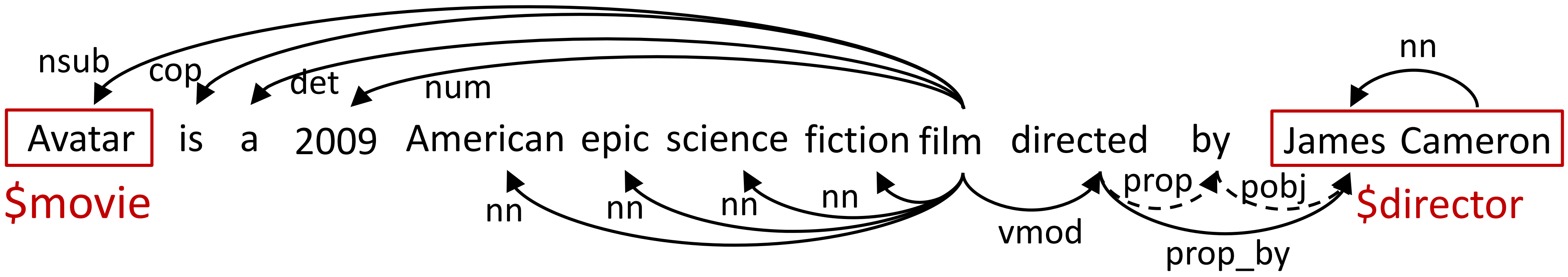
Dependency Parsing



Relational Surface Form Derivation (cont.)

Dependency-Based Entity Embeddings

1) Word & Context Extraction



Word	Contexts
$movie	film/nsub^{-1}
is	film/cop^{-1}
a	film/det^{-1}
2009	film/num^{-1}
american, epic, science, fiction	film/nn^{-1}

Word	Contexts
film	film/nsub, is/cop, a/det, 2009/num, american/nn, epic/nn, science/nn, fiction/nn, directed/vmod
directed	$director/prep_by
$director	directed/prep_by^{-1}

Relational Surface Form Derivation (cont.)

Dependency-Based Entity Embeddings

2) Training Process
   - Each word w is associated with a vector v_w and each context c is represented as a vector v_c
   - Learn vector representations for both words and contexts such that the dot product $v_w \cdot v_c$ associated with good word-context pairs belonging to the training data D is maximized
   - Objective function:

$$\arg\max_{v_w, v_c} \sum_{(w,c)\in D} \log \frac{1}{1+\exp(-v_c \cdot v_w)}$$

Word	Contexts
$movie	film/nsub^{-1}
is	film/cop^{-1}
a	film/det^{-1}
2009	film/num^{-1}
american, epic, science, fiction	film/nn^{-1}

Word	Contexts
film	film/nsub, is/cop, a/det, 2009/num, american/nn, epic/nn, science/nn, fiction/nn, directed/vmod
directed	$director/prep_by
$director	directed/prep_by^{-1}

Relational Surface Form Derivation (cont.)

Surface Form Derivation

Entity Surface Forms

- learn the surface forms corresponding to entities

$char, $director, etc.

$$S_i^F(w_j) = \boxed{\frac{\mathrm{sim}(w_j, e_i)}{\sum_{e_k \in E} \mathrm{sim}(w_j, e_k)}}$$

$$P^F(r_i \mid w_j)$$

based on word vector v_w

$char: “character”, “role”, “who”
$director: “director”, “filmmaker”
$genre: “action”, “fiction”

→ with similar contexts

Entity Syntactic Contexts

- learn the important contexts of entities

$$S_i^C(w_j) = \boxed{\frac{\mathrm{sim}(\hat{w}_j, e_i)}{\sum_{e_k \in E} \mathrm{sim}(\hat{w}_j, e_k)}}$$

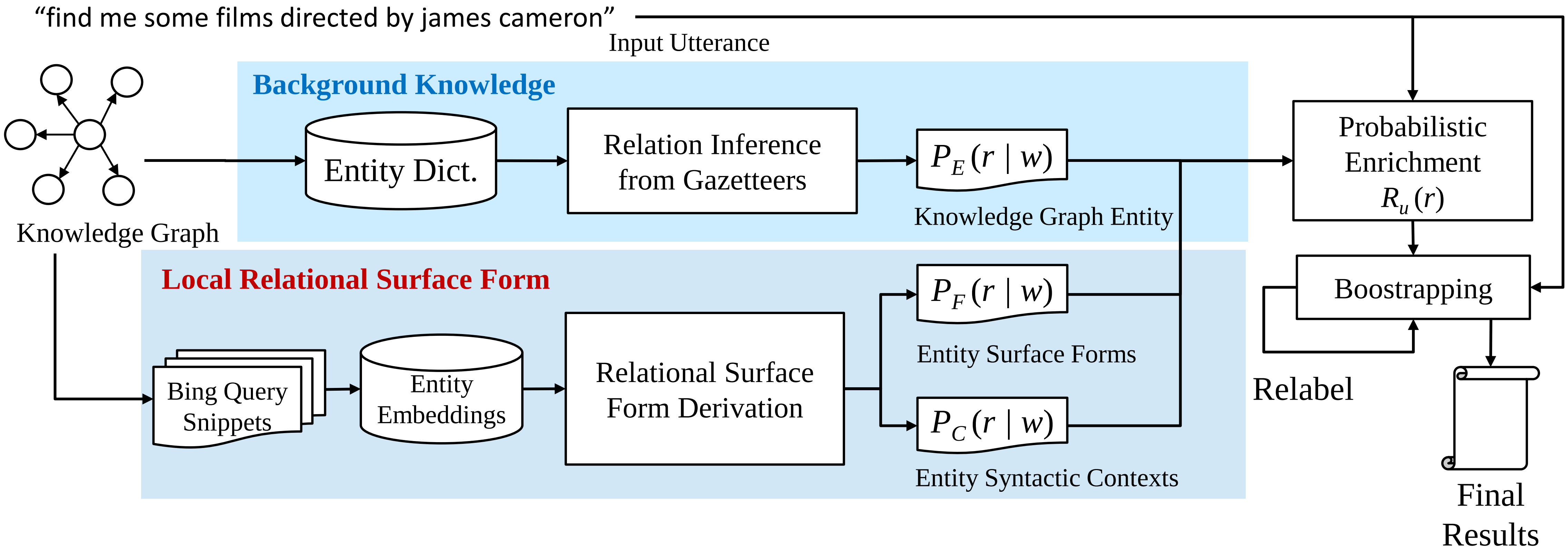
$$P^C(r_i \mid w_j)$$

based on context vector v_c

$char: “played”
$director: “directed”

→ frequently occurring together

Proposed Framework



Probabilistic Enrichment

Integrate relations from

- Prior knowledge $P_E(r \mid w)$
- Entity surface forms $P_F(r \mid w)$
- Entity syntactic contexts $P_C(r \mid w)$

r	actor	produced_by	location
$P_E(r \mid w)$	0.7	0.3	0
$P_F(r \mid w)$	0.4	0	0.6
$P_C(r \mid w)$	0	0	0
Unweighted $R_w(r)$	1	1	1
Weighted $R_w(r)$	0.7	0.3	0.6
Highest Weighted $R_w(r)$	0.7	0	0.6

Integrated Relations for Words by

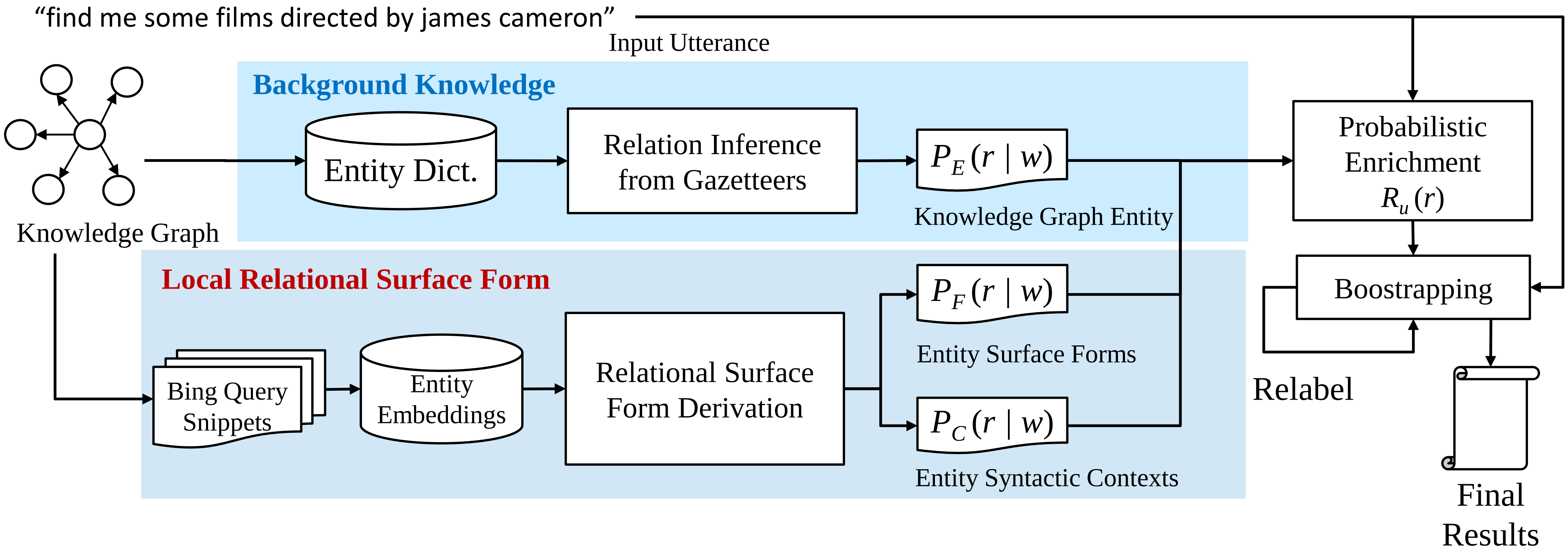
- **Unweighted**: combine all relations with binary values
- **Weighted**: combine all relations and keep the highest weights of relations
- **Highest Weighted**: combine the most possible relation of each word

Integrated Relations for Utterances by

$$R_u(r_i) = \max_{w \in u} R_w(r_i)$$

- Dilek Hakkani-Tur, Asli Celikyilmaz, Larry Heck, and Gokhan Tur, Probabilistic enrichment of knowledge graph entities for relation detection in conversational understanding, in *Proceedings of Interspeech,* 2014.

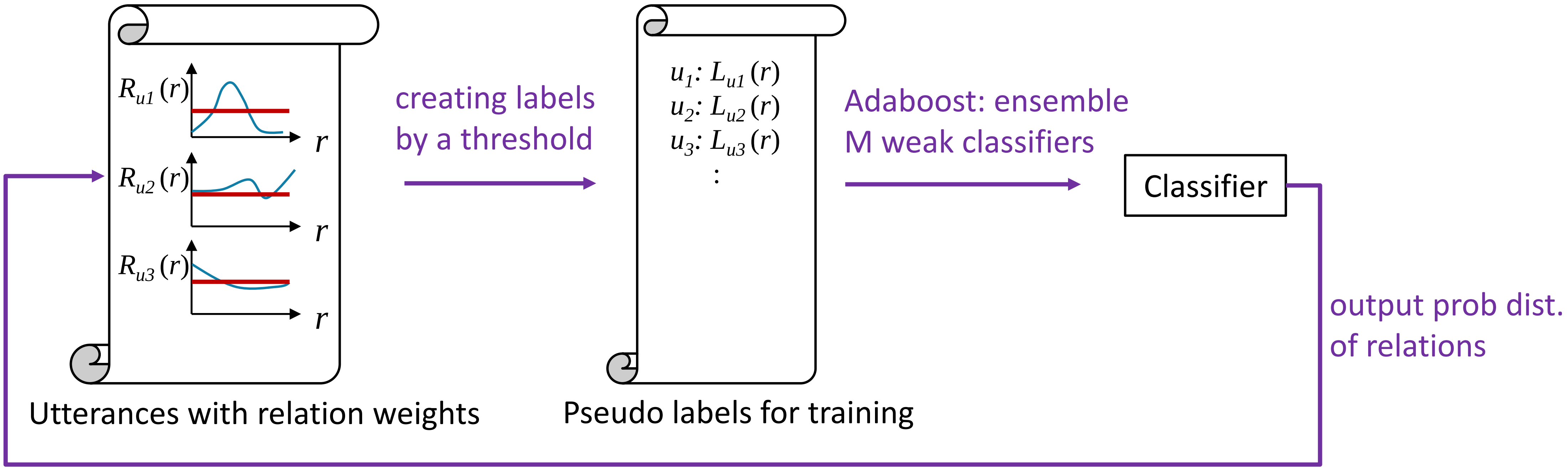
Proposed Framework



Boostrapping

Unsupervised Self-Training

Training a multi-label multi-class classifier estimating relations given an utterance



Experiments

Dataset

Knowledge Base: Freebase

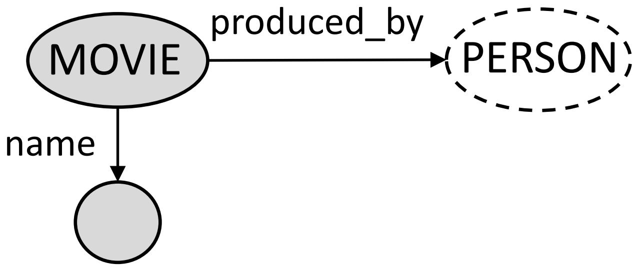
- 670K entities
- 78 entity types (movie names, actors, etc)

Relation Detection Data

- Crowd-sourced utterances
- Manually annotated with SPARQL queries → relations

Query Statistics	Dev	Test
% entity only	8.9%	10.7%
% rel only w/ specified movie names	27.1%	27.5%
% rel only w/ specified other names	39.8%	39.6%
% more complicated relations	15.4%	14.7%
% not covered	8.8%	7.6%
#utterances	3338	1084

User Utterance:
who produced avatar
Relation:
movie.name movie.produced_by



Experiments

All performance

Evaluation Metric: micro F-measure (%)

Baseline

Approach	Unweighted		Weighted		Highest Weighted	
	Ori	Boostrap	Ori	Boostrap	Ori	Boostrap
Gazetteer	35.21	36.91	37.93	40.10	36.08	38.89
Gazetteer + Weakly Supervised	25.07	37.39	39.04	39.07	39.40	39.98
Gazetteer + Entity Surface Form (Reg)	34.23	34.91	36.57	38.13	34.69	37.16

Experiments

All performance

Evaluation Metric: micro F-measure (%)

Approach	Unweighted		Weighted		Highest Weighted	
	Ori	Boostrap	Ori	Boostrap	Ori	Boostrap
Gazetteer	35.21	36.91	37.93	40.10	36.08	38.89
Gazetteer + Weakly Supervised	25.07	37.39	39.04	39.07	**39.40**	39.98
Gazetteer + Entity Surface Form (Reg)	34.23	34.91	36.57	38.13	34.69	37.16
Gazetteer + Entity Surface Form (Dep)	**37.44**	**38.37**	**41.01**	**41.10**	39.19	**42.74**

Baseline

Words derived by dependency embeddings can successfully capture the surface forms of entity tags, while words derived by regular embeddings cannot.

Experiments

All performance

Evaluation Metric: micro F-measure (%)

Approach	Unweighted		Weighted		Highest Weighted	
	Ori	Boostrap	Ori	Boostrap	Ori	Boostrap
Gazetteer	35.21	36.91	37.93	40.10	36.08	38.89
Gazetteer + Weakly Supervised	25.07	37.39	39.04	39.07	**39.40**	39.98
Gazetteer + Entity Surface Form (Reg)	34.23	34.91	36.57	38.13	34.69	37.16
Gazetteer + Entity Surface Form (Dep)	**37.44**	**38.37**	**41.01**	**41.10**	39.19	**42.74**
Gazetteer + Entity Context	35.31	37.23	38.04	38.88	37.25	38.04

Baseline

Words derived from entity contexts slightly improve performance.

Experiments

All performance

Evaluation Metric: micro F-measure (%)

	Approach	Unweighted		Weighted		Highest Weighted	
		Ori	Boostrap	Ori	Boostrap	Ori	Boostrap
Baseline	Gazetteer	35.21	36.91	37.93	40.10	36.08	38.89
Baseline	Gazetteer + Weakly Supervised	25.07	37.39	39.04	39.07	39.40	39.98
Baseline	Gazetteer + Entity Surface Form (Reg)	34.23	34.91	36.57	38.13	34.69	37.16
Proposed	Gazetteer + Entity Surface Form (Dep)	37.44	38.37	**41.01**	41.10	39.19	42.74
Proposed	Gazetteer + Entity Context	35.31	37.23	38.04	38.88	37.25	38.04
Proposed	Gazetteer + Entity Surface Form + Context	**37.66**	**38.64**	40.29	**41.98**	**40.07**	**43.34**

Combining all approaches performs best, while the major improvement is from derived entity surface forms.

Experiments

All performance

Evaluation Metric: micro F-measure (%)

	Approach	Unweighted		Weighted		Highest Weighted	
		Ori	Boostrap	Ori	Boostrap	Ori	Boostrap
Baseline	Gazetteer	35.21	36.91	37.93	40.10	36.08	38.89
	Gazetteer + Weakly Supervised	25.07	37.39	39.04	39.07	39.40	39.98
	Gazetteer + Entity Surface Form (Reg)	34.23	34.91	36.57	38.13	34.69	37.16
Proposed	Gazetteer + Entity Surface Form (Dep)	37.44	38.37	**41.01**	41.10	39.19	42.74
	Gazetteer + Entity Context	35.31	37.23	38.04	38.88	37.25	38.04
	Gazetteer + Entity Surface Form + Context	**37.66**	**38.64**	40.29	**41.98**	**40.07**	**43.34**

With the same information, learning surface forms from dependency-based embedding performs better, because there's mismatch between written and spoken language.

Experiments

All performance

Evaluation Metric: micro F-measure (%)

	Approach	Unweighted		Weighted		Highest Weighted	
		Ori	Boostrap	Ori	Boostrap	Ori	Boostrap
Baseline	Gazetteer	35.21	36.91	37.93	40.10	36.08	38.89
	Gazetteer + Weakly Supervised	25.07	37.39	39.04	39.07	39.40	39.98
	Gazetteer + Entity Surface Form (Reg)	34.23	34.91	36.57	38.13	34.69	37.16
Proposed	Gazetteer + Entity Surface Form (Dep)	37.44	38.37	**41.01**	41.10	39.19	42.74
	Gazetteer + Entity Context	35.31	37.23	38.04	38.88	37.25	38.04
	Gazetteer + Entity Surface Form + Context	**37.66**	**38.64**	40.29	**41.98**	**40.07**	**43.34**

Weighted methods perform better when less features, and highest weighted methods perform better when more features.

Experiments

All performance

Evaluation Metric: micro F-measure (%)

	Approach	Unweighted		Weighted		Highest Weighted	
		Ori	Boostrap	Ori	Boostrap	Ori	Boostrap
Baseline	Gazetteer	35.21	36.91	37.93	40.10	36.08	38.89
	Gazetteer + Weakly Supervised	25.07	37.39	39.04	39.07	39.40	39.98
	Gazetteer + Entity Surface Form (Reg)	34.23	34.91	36.57	38.13	34.69	37.16
Proposed	Gazetteer + Entity Surface Form (Dep)	37.44	38.37	**41.01**	41.10	39.19	42.74
	Gazetteer + Entity Context	35.31	37.23	38.04	38.88	37.25	38.04
	Gazetteer + Entity Surface Form + Context	**37.66**	**38.64**	40.29	**41.98**	**40.07**	**43.34**
	+ Names of Entity Types					**43.03**	**46.94**

Additionally adding names of entity types helps improve performance.

Experiments (cont.)

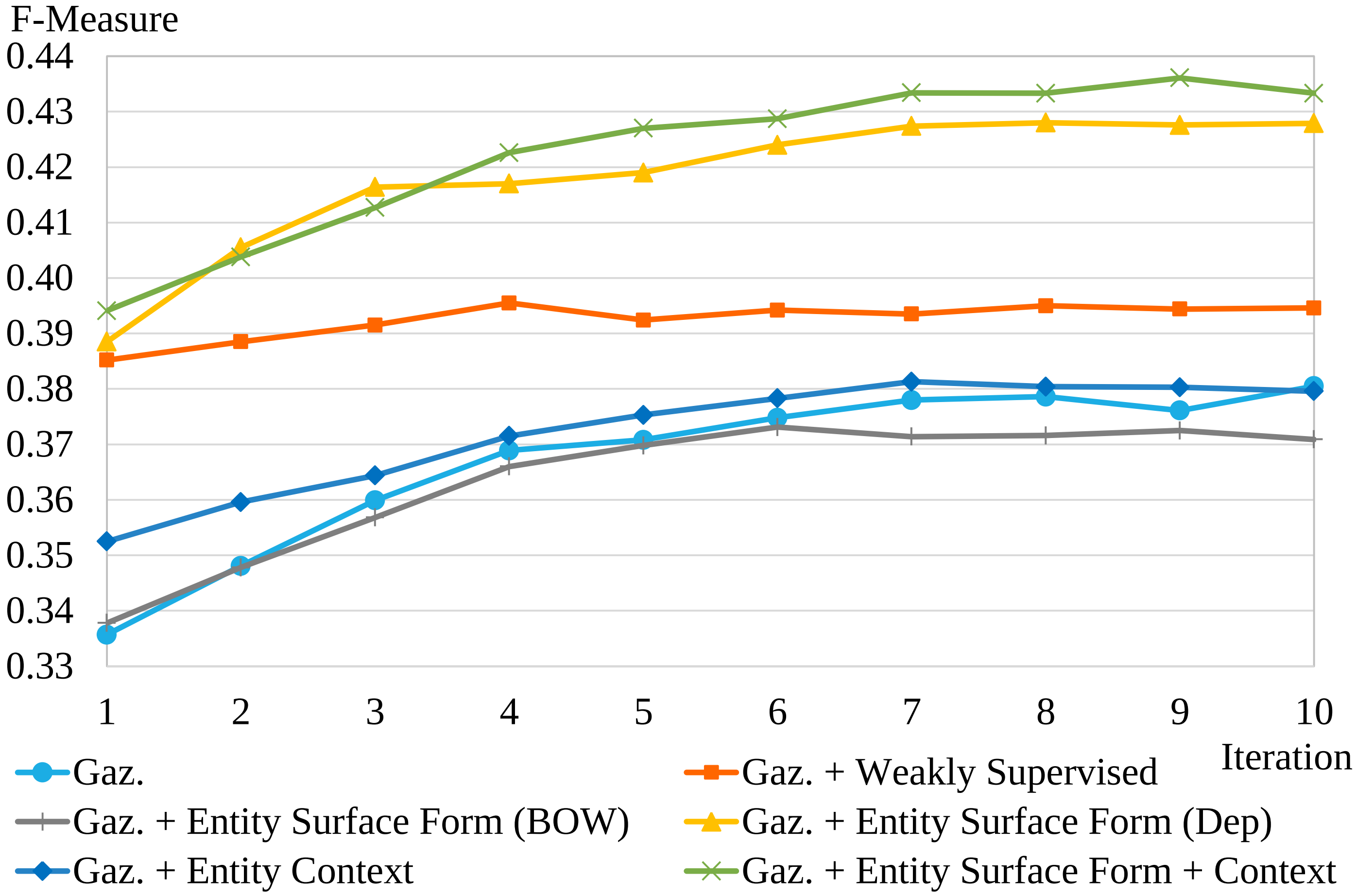
Entity Surface Forms Derived from Dependency Embeddings

The functional similarity carried by dependency-based entity embeddings effectively benefits relation detection task.

Entity Tag	Derived Word
$character	character, role, who, girl, she, he, officier
$director	director, dir, filmmaker
$genre	comedy, drama, fantasy, cartoon, horror, sci
$language	language, spanish, english, german
$producer	producer, filmmaker, screenwriter

Experiments (cont.)

Effectiveness of Boosting



- The best result is the combination of all approaches, because probabilities came from different resources can complement each other.
- Only adding entity surface forms performs similarly, showing that the major improvement comes from relational entity surface forms.
- Boosting significantly improves most performance

Conclusions

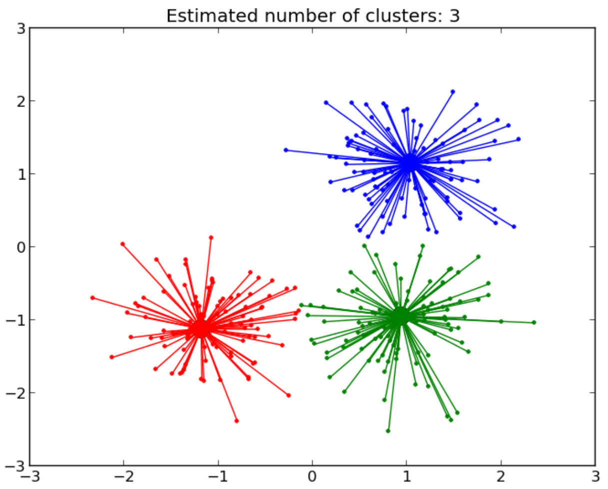
We propose an unsupervised approach to capture the relational surface forms including entity surface forms and entity contexts based on dependency-based entity embeddings.

The detected relations viewed as local observations can be integrated with background knowledge by probabilistic enrichment methods.

Experiments show that involving derived relational surface forms as local cues together with prior knowledge can significantly improve the relation detection task and help open domain SLU.

Ongoing & Future Work
Active Learning

Idea: manually label small data to boost performance



Approach

1. Extract exemplar utterances by clustering
   - Feature set: ngram, relation prob, both
   - Clustering: affinity propagation, k-means, etc.
2. Label exemplar utterances
3. Train the classifier on labelled data

Unsupervised results

- Embeddings: 0.4334
- Embeddings + Names: 0.4694

#training data (total = 3338)	5	10	15	20	25	30	35	40	45	50
Baseline: random selection	0.2892	0.3581	0.3867	0.3921	0.4306	0.4421	0.4522	0.4741	0.4810	**0.4821**
Unigram: Euclidean distance	0.1937	0.3167	0.3202	0.3252	0.3557	0.4005	0.4283	0.4447	0.4566	**0.4689**
Relation (embeddings)	0.3219	0.3545	0.4126	0.4218	0.4671	0.4907	0.4550	**0.4808**	0.4629	0.4800
Relation (names)	0.2780	0.2480	0.3686	0.3966	0.2860	0.4341	0.4490	0.4903	0.5005	**0.5150**
Relation (embeddings + names)	0.3457	0.3269	0.4552	0.4012	0.4489	0.4916	0.5191	0.5247	**0.5570**	0.5417

Thanks to…

Dilek Hakkani-Tür & Gokhan Tur for their advising

Dilek Hakkani-Tür for her help and care

Qi Li & Xiang Li for their discussion

All friends met here for enjoying summer together

MSR for this internship & XBOX ONE

Q & A ☺

THANKS FOR YOUR ATTENTIONS!!