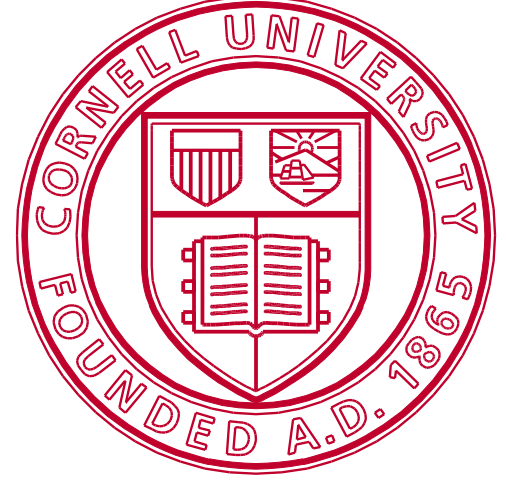# Leveraging Semantic Web Search and Browse Sessions for Multi-Turn Spoken Dialog Systems

Cornell University

Lu Wang
luwang@cs.cornell.edu

Larry Heck
larry.heck@ieee.org

Dilek Hakkani-Tür
dilek@ieee.org

Microsoft Research

Introduction

Goal: leverage behavioral <u>web search and browse sessions</u> to bootstrap <u>multi-turn spoken dialog systems (SDS)</u>

— S_1 : **Web Search and Browse Session**
T_1 (Q): [the great gatsby]$_{film}$ 2013
T_2 (C): [www.imdb.com/title/tt1343092/]$_{film}$
T_3 (Q): movies by [leonardo dicaprio]$_{actor}$
T_4 (C): [www.imdb.com/name/nm0000138/]$_{actor}$
T_5 (Q): [titanic]$_{film}$
T_6 (C): [www.imdb.com/title/tt0120338/]$_{film}$
T_7 (C): [http://www.imdb.com/character/ch0002338/]$_{character}$
T_8 (Q): does [leonardo dicaprio]$_{actor}$ have an award

— S_2 : **Spoken Dialog System Session**
U_1: go to bing dot com
U_2: search for [the great gatsby]$_{film}$
U_3: click that ([www.imdb.com/title/tt1343092/]$_{film}$)
U_4: show me movies by [leonardo dicaprio]$_{actor}$
U_5: scroll down
U_6: show me information on [titanic]$_{film}$
U_7: go back
U_8: show me information on [inception]$_{film}$
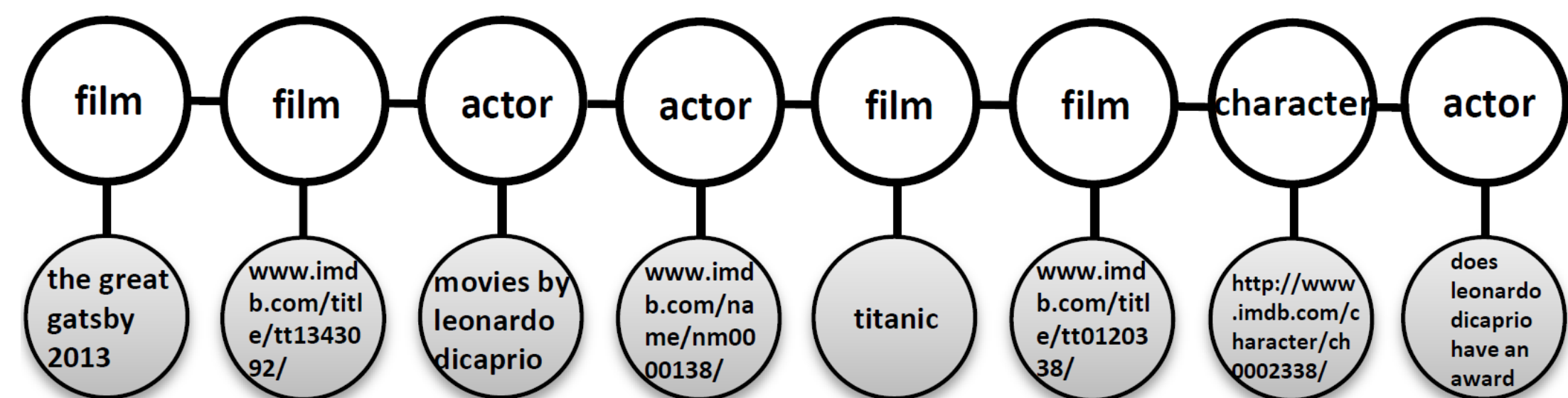
The Query Interpretation Model

Entity Mention Identification

- **Linear-chain Conditional Random Fields (CRF)**: identify all the entity mentions in each query.

Entity Type Determination

- **CRFs trained with missing labels:** replace the turns with no entity matched with missing labels.

Relation Extraction

- **Relation Templates**
  - Entity-Relation-Entity (E-R-E)
    - e.g. "*Leonardo Dicaprio in Titanic*"
  - Type-Relation-Entity (T-R-E)
    - e.g. "*movies by James Cameron*"
- **Multiclass logistic regression** classifiers (one vs. all strategy)

Web Search Query Interpretation

Data

- **Web Search Session Log**
  - 12 weeks of IE search/browse sessions from March 18 to June 9, 2013
  - Domain: Film.
  - 297,352 sessions in total.
- **Knowledge Base**: extension from Freebase.
- **Training via Distant Supervision**

	# Entity Type	# Entity	# Relation Type	# Relation
Training	44	215,053	33	20,197
Test	44	104,657	27	5,255
Test Only	0	23,557	0	1,104

Pilot Study: *Entity-based Search Session Modeling*
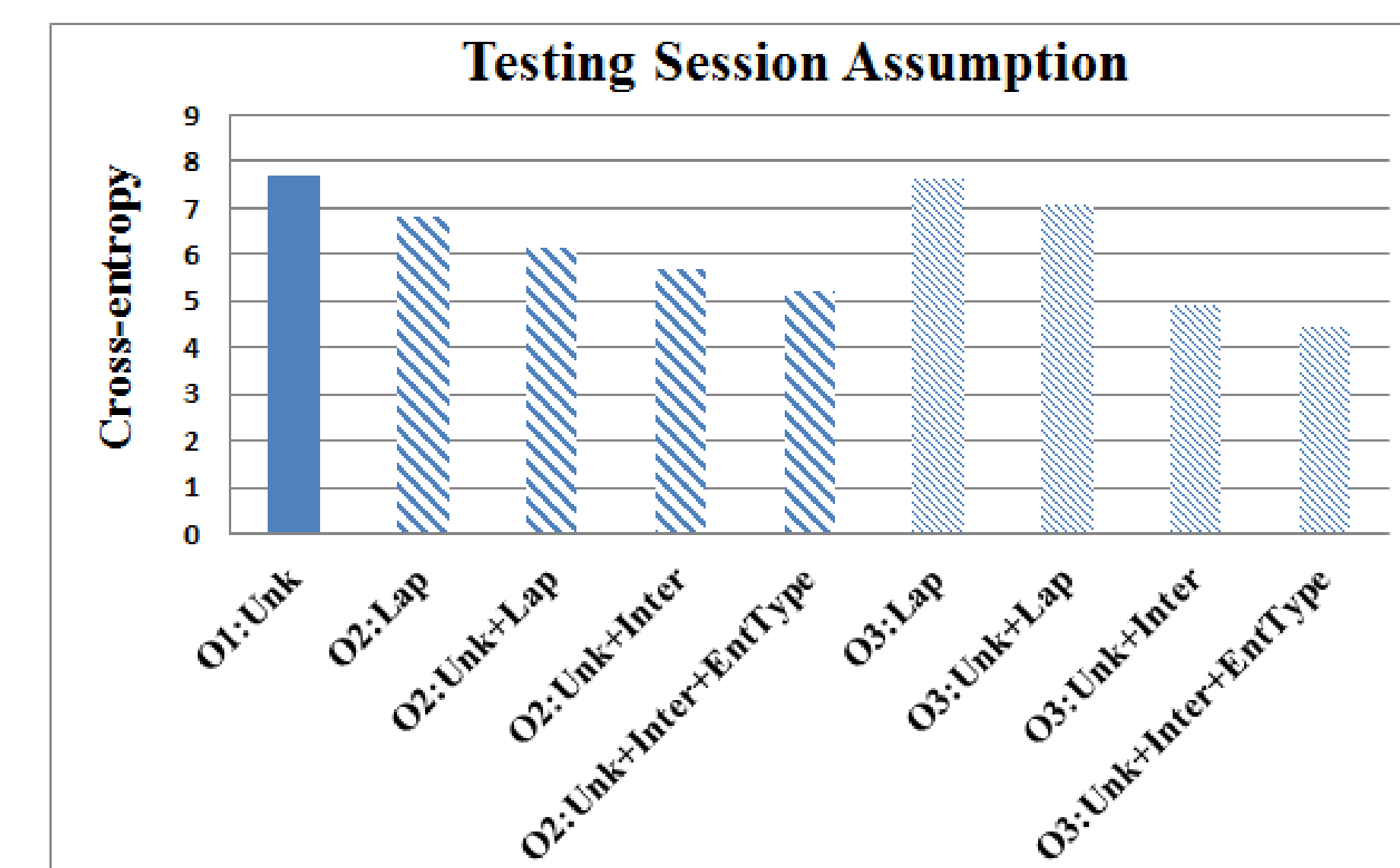
- Session = sequence of entities.
- Session likelihoods: Markov models with different history.
- 2 smoothing techniques: additive smoothing & linear interpolation.

Experimental Results

- Baseline 1 (Gazetteer Only): string matching with the gazetteer.
- Baseline 2 (No Session): log-linear model using the same features.
- CRF (mention + type): jointly learn the entity mention and entity type.
- Session model (CRF): removed turns without entity match.
- Session model (CRF w/ missing label): replace turns without entity match with missing labels.

	Recall on Entity Extraction
Baseline 1 (Gazetteer Only)	84.00
Baseline 2 (No Session)	86.34
CRF (mention + type)	87.33
Session model (CRF)	88.06
Session model (CRF w/ missing label)	**90.38**

	Recall on E-R-E	Recall on T-R-E
Baseline 1 (Gazetteer Only)	79.21	74.72
Our Model	**80.81**	**78.22**

Entity Extraction in Spoken Dialog System

Data

- **Conversational Browsing Data (CB data)**
  - Collected from system where users interact by voice and gesture to perform tasks (conversational web search and browsing)
  - 6691 utterances in 136 sessions.
- **Netflix Data**
  - Collected from natural language film search application based on Netflix.
  - 13851 utterances in1895 sessions.

Experimental Results

- Baselines and session models are the same as previously defined.
- Heck et al: use training data constructed from Wikipedia.

	CB Data					Netflix Data				
	Film	Director	Actor	Character	Genre	Film	Director	Actor	Character	Genre
(# utterances)	195	16	55	3	58	2,013	196	713	85	968
Baseline 1 (Gazetteer Only)	56.61	89.66	93.20	50.00	18.75	58.55	58.55	92.82	68.22	18.01
Baseline 2 (No Session)	71.71	81.25	90.57	50.00	18.75	67.98	75.29	89.23	*70.23*	18.01
Heck et al.	-	-	-	-	-	71.72	84.62	58.61	-	29.55
Our session model (CRF)	71.71	*89.66*	95.24	50.00	18.75	67.98	91.83	*95.12*	*70.23*	29.29
Our session model (CRF with missing label)	**75.80**	*89.66*	**96.23**	**80.00**	**21.54**	**69.39**	**92.01**	*95.12*	*70.23*	**33.70**

Discussion

- Our session-based models trained on web search logs outperform two nontrivial baselines and state-of-the-art methods in most scenarios.
- Our models effectively leverage behavioral web search/browse patterns to improve the entity extraction performance in spoken dialog systems.
- Genre detection is relatively poor. Due to different context people use for genres, e.g., "Disney" is a *genre* in "Disney movie", but is a *place name* in "Disney world".
- Future Work: add co-reference resolution

Contributions

- 1st successful method to directly leverage web search & browse sessions to bootstrap statistical multi-turn models for SDS
- Significantly improved multi-turn SDS with a scalable method