

Context Representation for Named Entity Linking

Xiao Ling

Sameer Singh

Daniel S. Weld

University of Washington



Named Entity Linking

Seattle beat Portland yesterday.

Candidates	String Similarity	Context Similarity	Coherence Score	Relational Score	Final Score
Seattle (city)	0.6	0.3	0.1	0.0	1.0
Seattle Sounders	0.3	0.7	0.2	0.1	1.3
Seattle-Tacoma (airport)	0.4	0.2	0.0	0.1	0.7

Overview

- ▶ We propose a novel context representation using **dependency features** in replace of bag-of-words.
- ▶ To combat sparsity, we perform **matrix completion** via a low-dimensional embedding of entities and features.
- ▶ We present preliminary results showing promises of the new representation.

Context Representation



Pattern Type	Example
Dependency Path	From the mention's head to a content word outside the mention. E.g. "X -[nsubj]->beat"
Premodifier	E.g. the title of a person : "President Barack Obama"
Apposition	E.g. a textual description of the entity: "Russell Wilson, the second-year quarterback"



Matrix Completion

Features \ Entities	X serve in Senate	fly into X	X serve the city of LOC	X be seat of LOC	X beat	person	soccer team	airport
Seattle (city)				1				
Seattle Sounders					?		1	
Sea-Tac (airport)		1	1					1
JFK (airport)		1	?					1
JFK (president)	1					1		
New York Red Bulls					1		1	

Features: Syntactic Patterns, FB Types

Entities: v_e

Features: v_f

Matrix size: $\sim 3m \times 700k$

Each matrix entry $E_{e,f}$ is defined as:

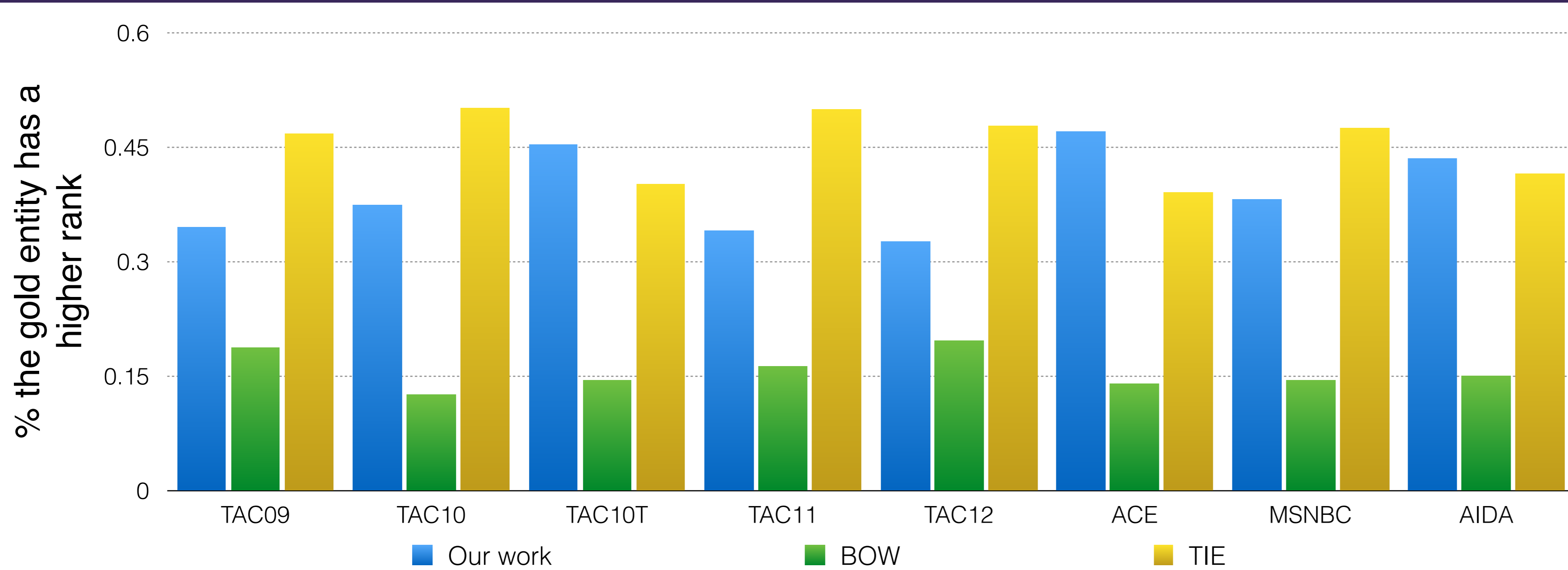
$$E_{e,f} = \frac{1}{1 + \exp(-v_e \cdot v_f)}$$

We fill in missing values via learning a low-dimensional embedding of both entities (v_e) and features (v_f) from correlations of a Freebase type and dependency feature, of two dependency features and of two similar entities.

Nearest Neighbors in the Entity Matrix

Washington (state)		Washington, D.C.		Georgia (U.S. state)		Georgia (country)	
Oregon	0.2917	Philadelphia	0.2089	Tennessee	0.2834	Azerbaijan	0.2669
Washington	0.2609	San Francisco	0.2022	Florida	0.2831	Armenia	0.2637
Idaho	0.2309	Los Angeles	0.1969	Illinois	0.2779	Ukraine	0.2569
Louisiana	0.2307	Houston	0.1957	Mississippi	0.2773	Belarus	0.2395
Georgia (U.S. state)	0.2302	Boston	0.1924	Texas	0.2760	Bulgaria	0.2376
Colorado	0.2237	Tampa, Florida	0.1915	Pennsylvania	0.2677	Kyrgyzstan	0.2359
Arizona	0.2221	New York City	0.1909	Arkansas	0.2675	Serbia	0.2339
Texas	0.2214	Baltimore	0.1882	North Carolina	0.2662	Romania	0.2324

Preliminary Experimental Results



End-to-End NEL Accuracy

Data Set	state-of-the-art	this work
TAC09	82.2	81.8
TAC10	86.8	87.3
TAC10T	-	89.2
TAC11	86.8	85.8
TAC12	76.6	65.5
AIDA	82.8	75.0
ACE	85.9	87.0
MSNBC	84.6	75.8