Learning Semantic Entity Representations with Knowledge Graph and Deep Neural Networks and its Application to Named Entity Disambiguation

Hongzhao Huang[1] and Larry Heck[2]

Computer Science Department, Rensselaer Polytechnic Institute[1]

Microsoft Research[2]

{huangh9@rpi.edu, Larry.Heck@microsoft.com}

Specific Thanks Yelong Shen and Gustavo Abrego for the help on deep neural network related issues

Word Embeddings

- Standard word representation
  - "One-hot" representation
    - Microsoft [0, 0, 0, 0,…,0, 1, 0,…,0]
- Neural word embeddings
  - Distributed representation
    - Microsoft [0.453, -0.292, 0.732,…, -0.243]
  - Represent a word by its contextual surrounding words

"You shall know a word by the company it keeps"

(J. R. Firth 1957: 11)

- government debt problems turning into banking crises as has happened in
- saying that Europe needs unified banking regulation to replace the hodgepodge

Examples from (Socher et al, NAACL2013 turorial)

From Word Embeddings to Entity Embeddings

- How about entities?
  - Usually composed of multiple words
    - Microsoft Research, James Cameron, Atlanta Hawks

!=

+

  - Entities play crucial role in many applications
    - Entity Linking, Relation Extraction, Question & Answering…
- Our goal
  - Learn task specific accurate semantic entity representations

How can we represent entities?

- How we learn about a new entity/concept?

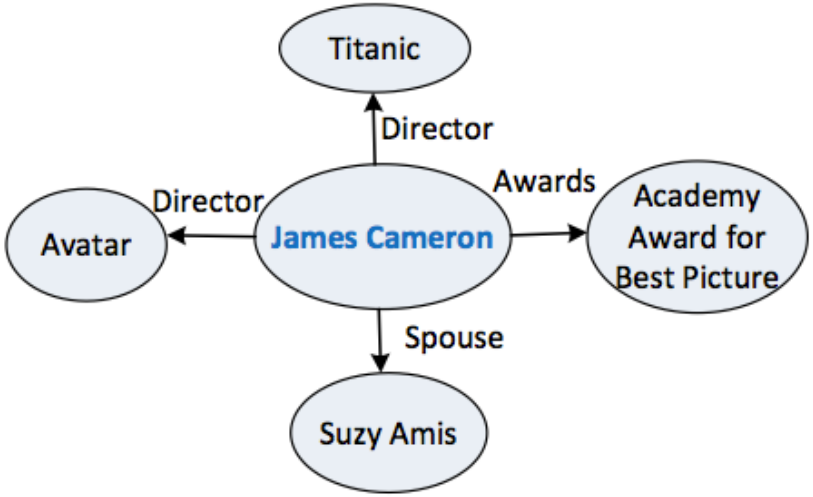
James Cameron

Film director

James Francis Cameron is a Canadian film director, film producer, deep-sea explorer, screenwriter, and editor who has directed the two biggest box office films of all time. He first found success with the science-fiction hit The Terminator. Wikipedia

- <James Cameron, film director, Titanic>
- <James Cameron, won awards, Academy Award for Best Picture>

....



Semantic Knowledge Graphs (KGs)

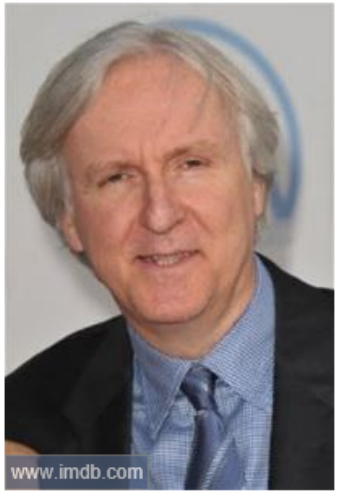
- A graph composed of:
  - Nodes: uniquely identified entities or literals
  - Edges: semantic relations
    - E.g., film director, film producer, CEO of…
- Many rich and clean KGs
  - Satori, Google KG, Freebase, Dbpedia….
- Broad applications to natural language processing and spoken language understanding
  - E.g., Unsupervised semantic parsing (Heck et al, 2012)
    - Use KG to guide automatic labeling of training instances
- This work: encode world knowledge from KG to assist deep understanding and accurate semantic representations of entities

Semantic Knowledge Graphs: An Example

SATORI http://knowledge.microsoft.com/22175d10-ee6b-77c9-e548-89b0ff670090

Open in Bing | Entity Details | Edit Entity | Edit Side Streams | Entity Cluster | Entity Log | Compare to Previous

James Cameron



James Francis Cameron (born August 16, 1954) is a Canadian film director, film producer, deep-sea explorer, screenwriter, and editor who has directed the two biggest box office films of all time. He first found success with the science-fiction hit The Terminator (1984). He then became a popular Hollywood director and was hired to write and direct Aliens (1986); three years later he followed up with The Abyss (1989).
en.wikipedia.org




Born	August 16, 1954	Kapuskasing
Height	6' 2" (1.88 M)	
Spouse	Suzy Amis (2000)	Linda Hamilton (1997 - 1999)
	Kathryn Bigelow (1989 - 1991)	Gale Anne Hurd (1985 - 1989)
	Sharon Williams (1978 - 1984)	

Awards	Category	Year
	Mainichi Film Concours Readers' Choice Award for Best Film	1999
	Academy Award for Best Film Editing	1997
	Academy Award for Best Director	1997
	Academy Award for Best Picture	1997

Films

Solartaxi: Around the World with the Sun (2010)

Explorers: From the Titanic to the Moon (2006)

Ray Harryhausen: Special Effects Titan (2011)

Entity Facts Injected Facts

(11371) Legend: Not Published - Low - Med - High Editorial Mode

XML (Download XML) - JSON (Download JSON) - CDB Json (use mainline XML)

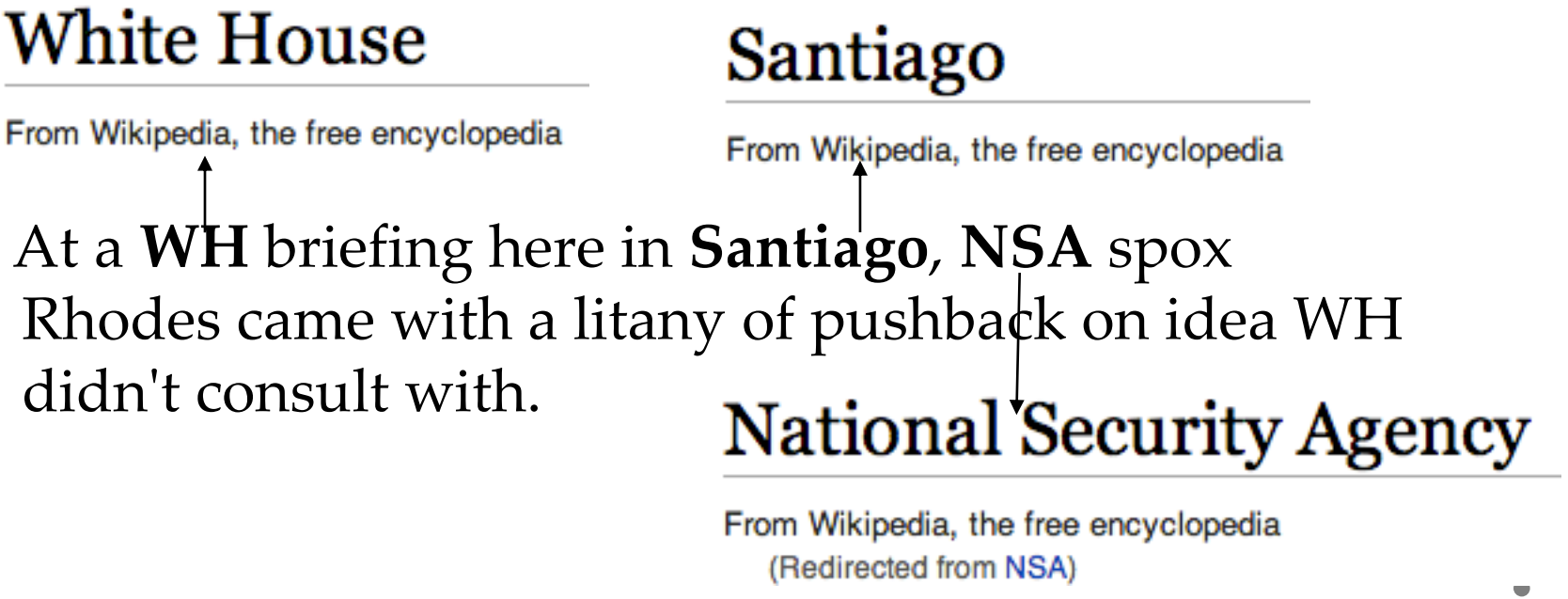
award.nominee.award_nominations

category: Writers Guild of America Award for Best Original Screenplay
ceremony: Writers Guild of America Awards 1997
nominated_work: Titanic
year: 1998Z

category: London Film Critics Circle for Director of the Year
nominated_work: Titanic

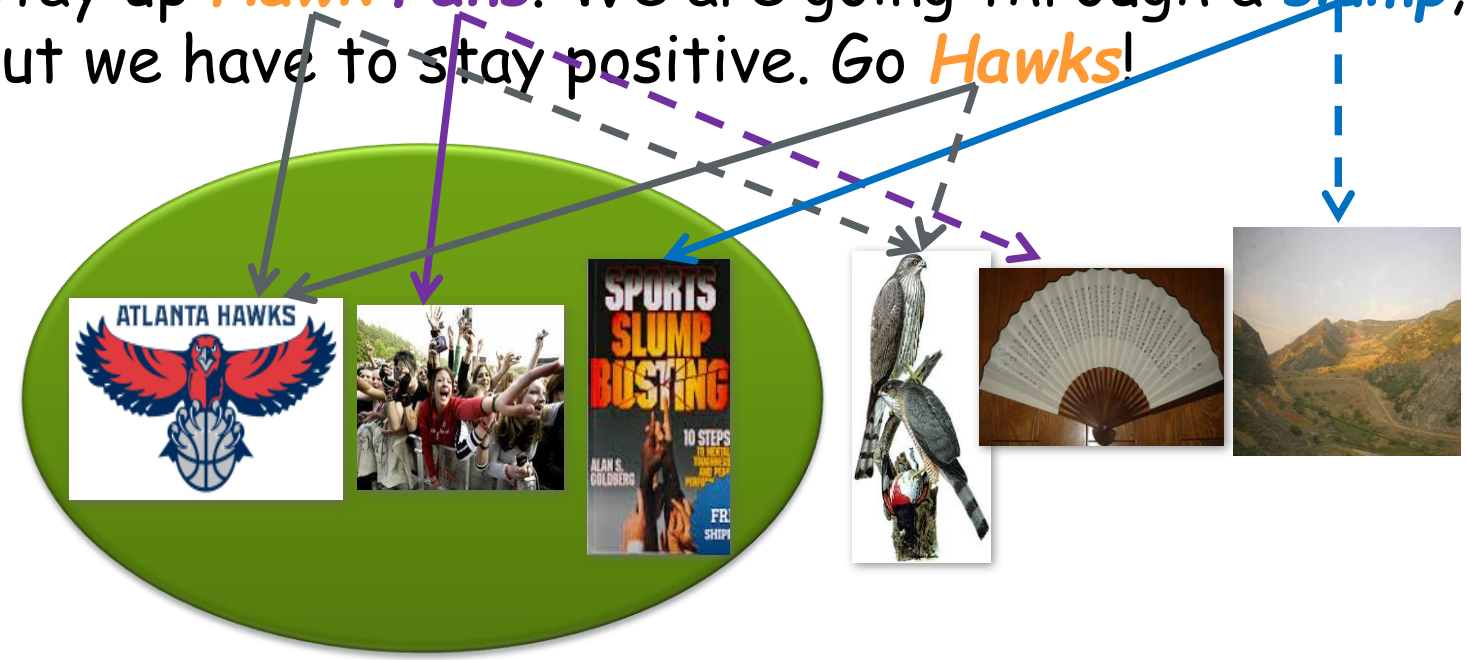
Named Entity Disambiguation (NED): Task Definition

- Disambiguate linkable mentions from a specific context to their referent entities in a Knowledge Base
  - A mention: a phrase referring to something in the world
    - Named entity (person, organization), object, event…
  - An entity: a page in a Knowledge Base



Entity Semantic Relatedness is Crucial for NED

- Stay up *Hawk Fans*. We are going through a *slump*, but we have to stay positive. Go *Hawks*!



- The most important feature used for NED
  - Non-collective approaches (Ferragina & Scaiella, 2010; Milne and Witten, 2008; Guo et.al., 2013)
  - Collective Approaches (Cucerzan, 2007; Milne and Witten, 2008b; Kulkarni et al., 2009; Pennacchiotti and Pantel, 2009; Ferragina and Scaiella, 2010; Cucerzan, 2011; Guo et al.,2011; Han and Sun, 2011; Han et al., 2011; Ratinovet al., 2011; Chen and Ji, 2011; Kozareva et al., 2011; Shen et al., 2013; Liu et al., 2013, Huang et al., 2014)

The State-of-the-art Approaches for Entity Semantic Relatedness

- (Milne and Witten, 2008): V
  unsupervised method

Limitation I: Ingore the world knowledge from the rich Knowledge Graphs

$$SR(c_i, c_j) = 1 - \frac{\log \max(|C_i|, |C_j|) - \log |C_i \cap C_j|}{\log(|C|) - \log \min(|C_i|, |C_j|)}$$

  - C_i: the set of incoming links t
- Supervised Method (Ceccarel
  - Formulate as a learning-to-ra
  - Explore a set of link-based features

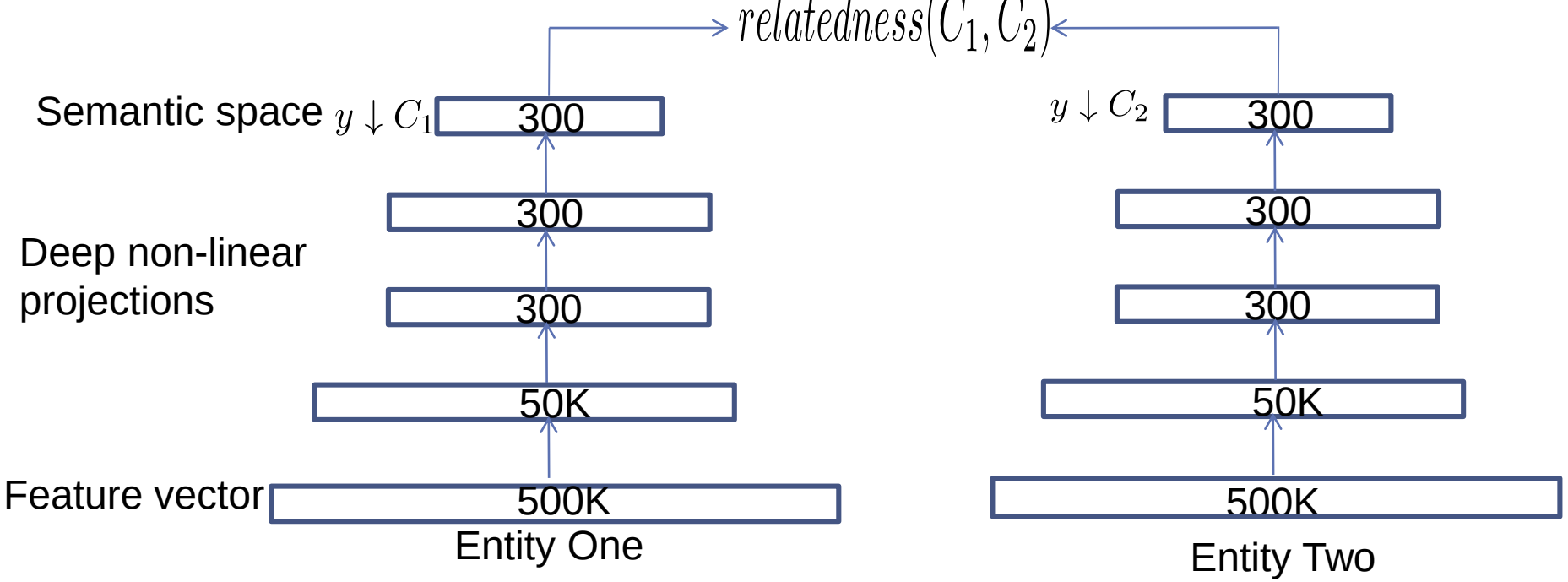
Limitation II: what if we donot have anchor links?

Boot used "imperialism" to describe United States policy, not only in th
This embrace of empire is made by others neoconservatives, including
a and Mark Steyn. It is also made by some liberal hawks, such as polit

Our Approach

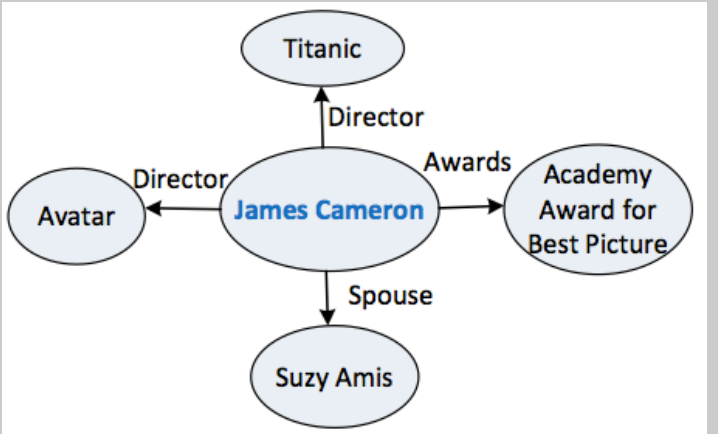
- Learn entity representations with supervised DNN and KG
  - Non-linear DNN proven to have more expressive power than the linear models
  - Directly to optimize parameters for semantic relatedness
- The DNN-based Semantic Similarity Model (DSSM) (Huang et al, 2013)

$$relatedness(C_1, C_2) = Cosine(y \downarrow C_1, y \downarrow C_2)$$



Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck, Learning Deep Structured Semantic Models for Web Search using Clickthrough Data Proc. CIKM2013

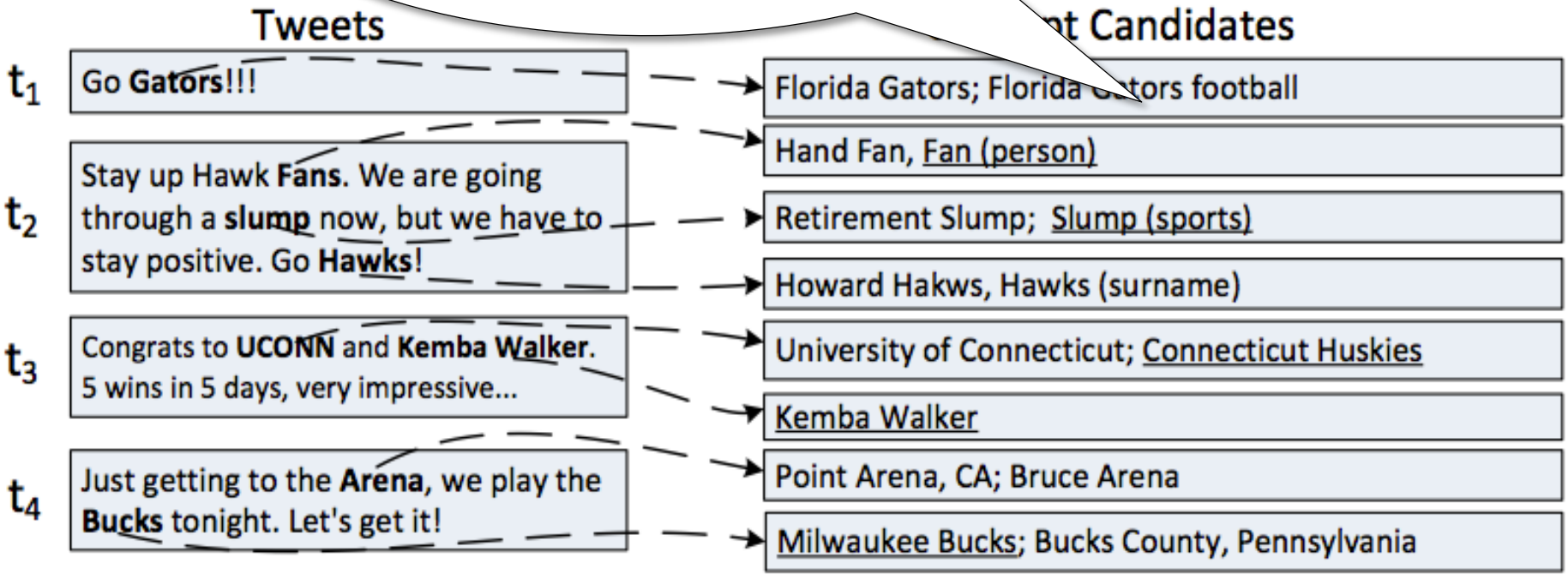
Encoding Knowledge from Knowledge Graph

Knowledge	Representation	Example
Description	Letter tri-gram vector	dog = <#do, dog, og#> <0,…,1,1,…,0,1,…,0>
Entity Type	1-of-V vector	<0,….,0,…,1,…,0,…>
Subgraph	1-of-V vector for relation Letter tri-gram for entities	

Unsupervised Collective Disambiguation with Graph Regularization

- Perform collective disambiguation for a set of topically-related tweets sim
  - H ss problems
  - I eets (e.g., via social ne

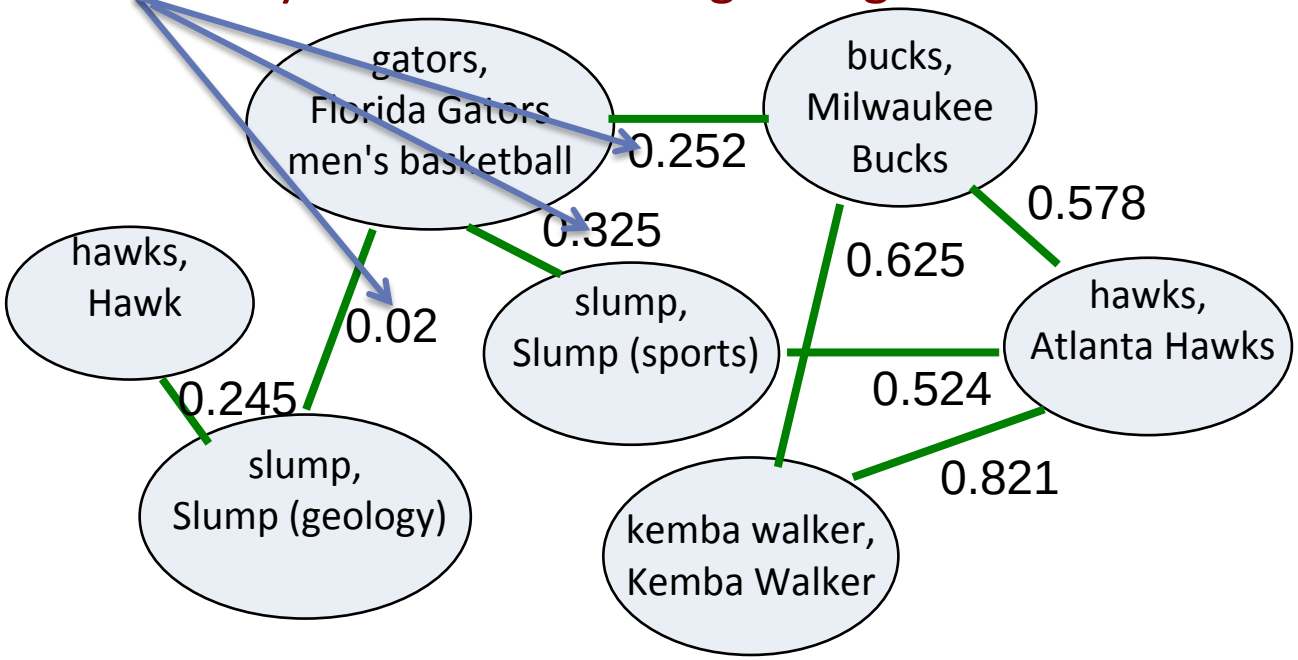
Accuracy = 0.25, tweets are short and noisy, can not provide rich context information



Underlining concepts are referent concepts

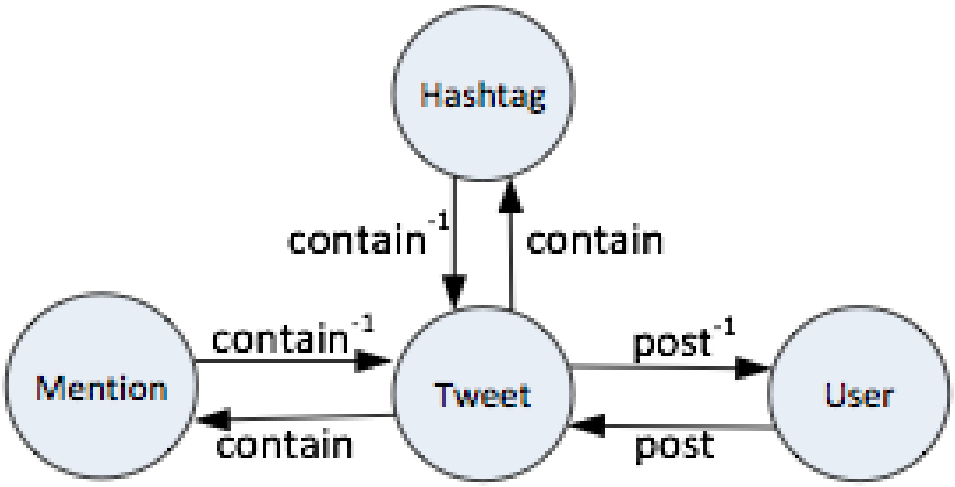
Graph Construction Over Multiple Tweets

- Each node is a pair of mention and entity candidates
  - Entity candidates are retrieved based on anchor links in Wikipedia
- An edge is created for two nodes if
  - Two mentions are relevant
    - Detect with meta path
  - And two entities are semantically related
    - Cosine similarity over semantic entity embeddings
    - Similarity is used as the edge weight



Relevant Mention Detection: Meta Path

- A meta-path is a path defined over a network and composed of a sequence of relations between different object types (Sun et al., 2011)
  - Each meta path represent a semantic relation
- Meta paths between mention and mention
  - M-T-M
  - M-T-U-T-M-M
  - M-T-H-T-M
  - M-T-U-T-M-T-H-T-M
  - M-T-H-T-M-T-U-T-M



Schema of a Heterogeneous Information Network in Twitter

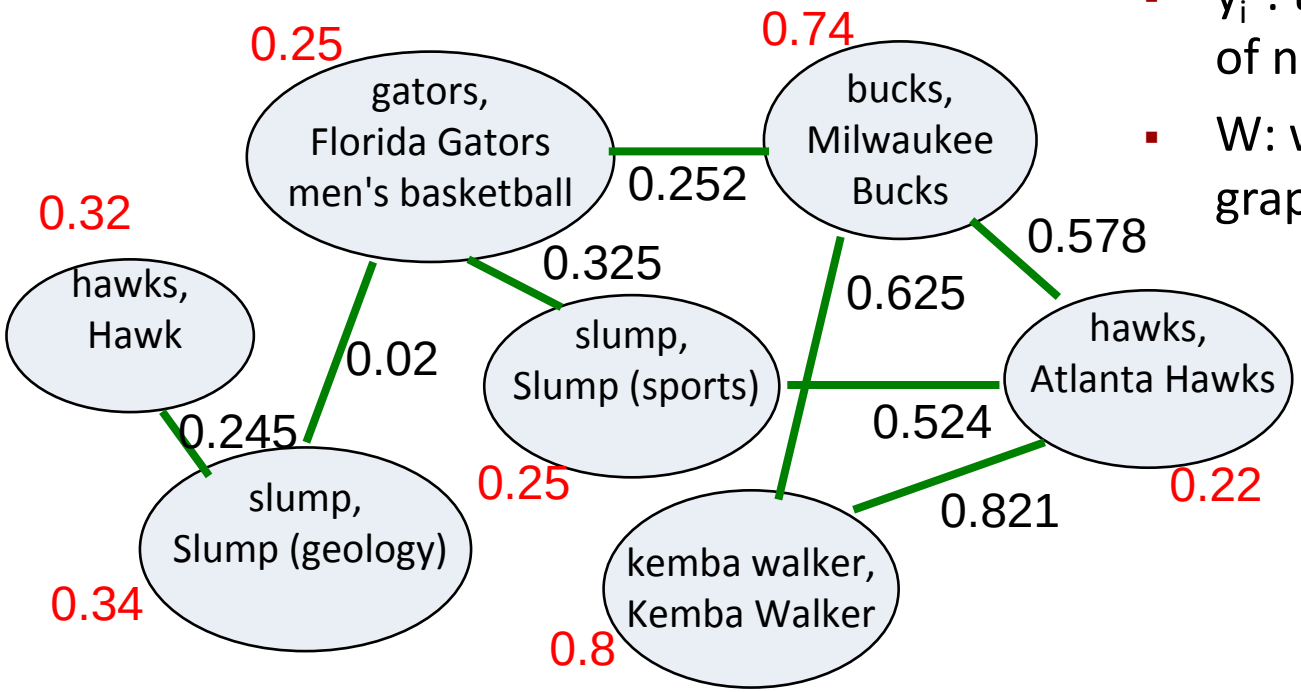
M: mention, T: tweet, U: user, H: hashtag

- Two mentions are considered as relevant if there exist at least one meta path between them

Unsupervised Graph Regularization

- The model (Adapted from Zhu et.al, 2003)
  - $$\mathcal{Q}(\mathcal{Y}) = \mu \sum_{i=l+1}^{n} (y_i - y_i^0)^2 + \frac{1}{2} \sum_{i,j} W_{ij} (y_i - y_j)^2.$$
- Initial ranking score
  - prior popularity and context similarity

- y_i: the final ranking score of node i
- y_i^0: the initial ranking score of node i
- W: weight matrix of the graph



Data and Scoring Metric

- Data
  - A public data set includes 502 messages from 28 users (Meiji et al., 2012)
  - A Wikipedia dump on May 3, 2013
- Scoring Metric
  - Accuracy on top ranked entity candidates

Models for Comparison

- TagMe: an unsupervised model based on prior popularity and semantic relatedness of a single message (Ferragina and Scaiella, 2010)
- Meij: the state-of-the-art supervised approach based on the random forest model (Meij et al., 2012)
- GraphRegu: our proposed unsupervised graph regularization model

Overall Performance

- Our methods are unsupervised

Method	Accuracy
TagMe (unsupervised)	61.9%
Meiji (5 fold cross-validation)	68.4%
GraphRegu + (Milne and Witten, 2008)	64.3%

Overall Performance (con't)

- Encode Knowledge from contextual descriptions

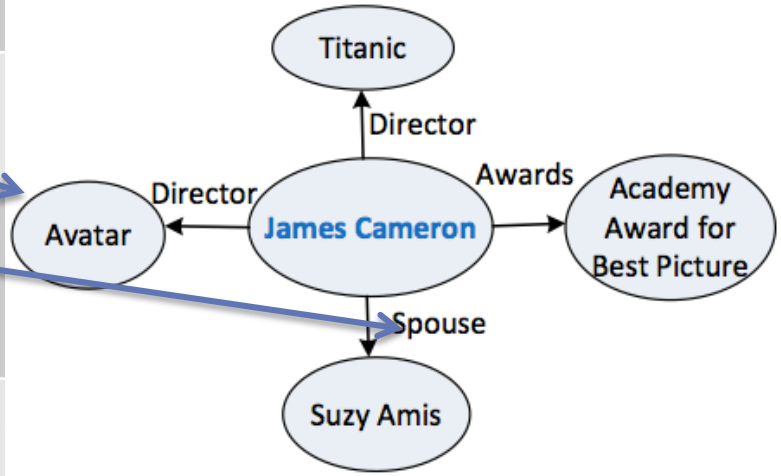
Method	Accuracy
TagMe (unsupervised)	61.9%
Meiji (5 fold cross-validation)	68.4%
GraphRegu + (Milne and Witten, 2008)	64.3%
GraphRegu + DSSM + Description	**71.8%**

- 26% error rate reduction over TagMe
- 21% error rate reduction over the standard method to compute semantic relatedness (Milne and Witten, 2008)

Overall Performance

- Encode Knowledge from structured KG

Method	Accuracy
TagMe (unsupervised)	61.9%
Meiji (5 fold cross-validation)	68.4%
GraphRegu + (Milne and Witten, 2008)	64.3%
GraphRegu + DSSM + Subgraph (Entity)	68.2%
GraphRegu + DSSM + Subgraph (Relation + Entity)	70.0%
GraphRegu + DSSM + Subgraph (Relation + Entity) + Entity Type	**70.9%**



- 23.6% error rate reduction over TagMe
- 18.5% error rate reduction over the standard method to compute semantic relatedness (Milne and Witten, 2008)

Overall Performance

- Encode all Knowledge from KG

Method	Accuracy
TagMe (unsupervised)	61.9%
Meiji (5 fold cross-validation)	68.4%
GraphRegu + (Milne and Witten, 2008)	64.3%
GraphRegu + DSSM + Description	**71.8%**
GraphRegu + DSSM + Subgraph (Entity)	68.2%
GraphRegu + DSSM + Subgraph (Relation + Entity)	70.0%
GraphRegu + DSSM + Subgraph (Relation + Entity) + Entity Type	70.9%
GraphRegu + DSSM + Description + Subgraph (Relation + Entity) + Entity Type	**71.9%**

Conclusions and Future work

- We propose to learn deep semantic entity embeddings with supervised DNN and Knowledge Graph
  - Significantly outperform the standard approach for named entity disambiguation
- Future Work
  - Encode semantic meta-paths from Kowledge Graph into DNN
    - To capture the semantic meaning of knowledge
  - Learn entity embedding with Knowledge Graph for other tasks
    - E.g., Question & Answering

Thank You !!!

Any Questions/Comments?

We will release the embedding for the whole Wikipedia Concepts Soon!!!