Joint Compositional Learning from Text and Knowledge Bases

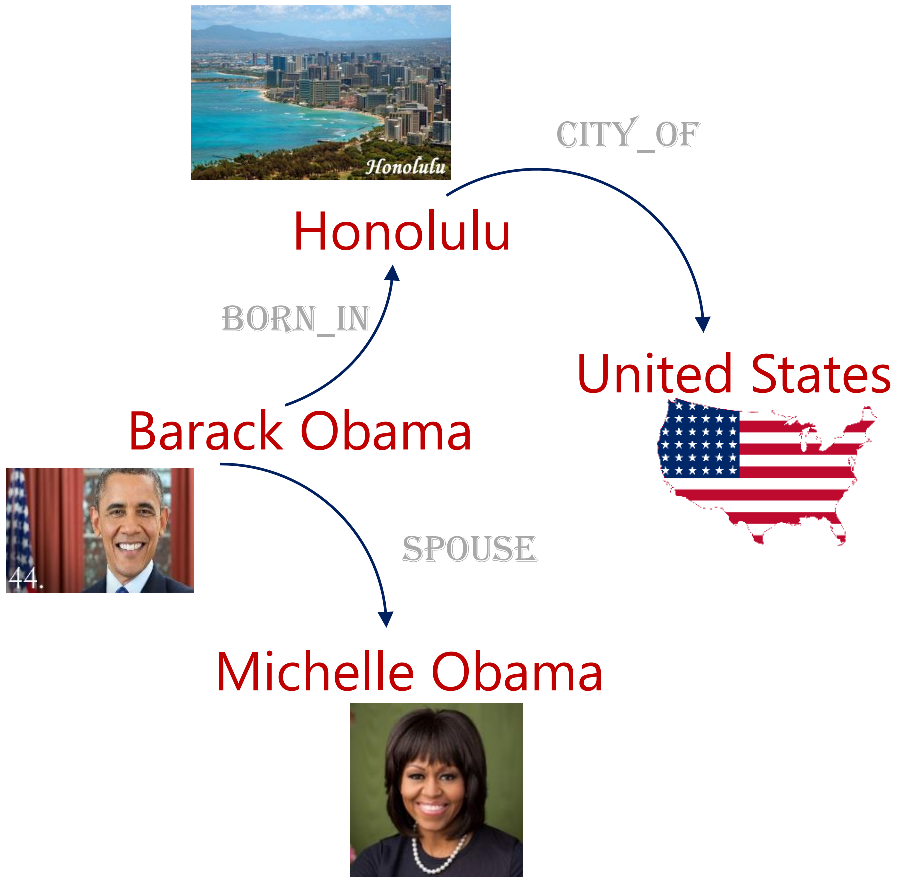
Kristina Toutanova

Joint work with

Danqi Chen, Victoria Lin, Scott Yih, Hoifung Poon, Chris Quirk, Patrick Pantel, Michael Gamon, Pallavi Choudhury

Knowledge bases

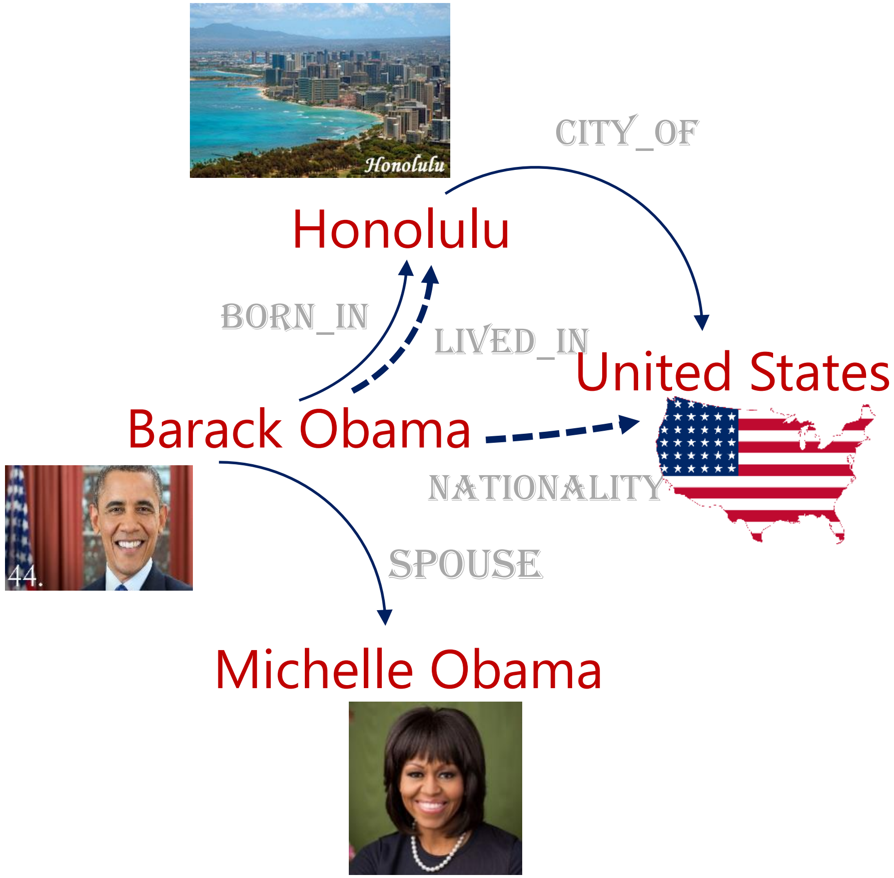
Knowledge bases with entities and relations



Freebase

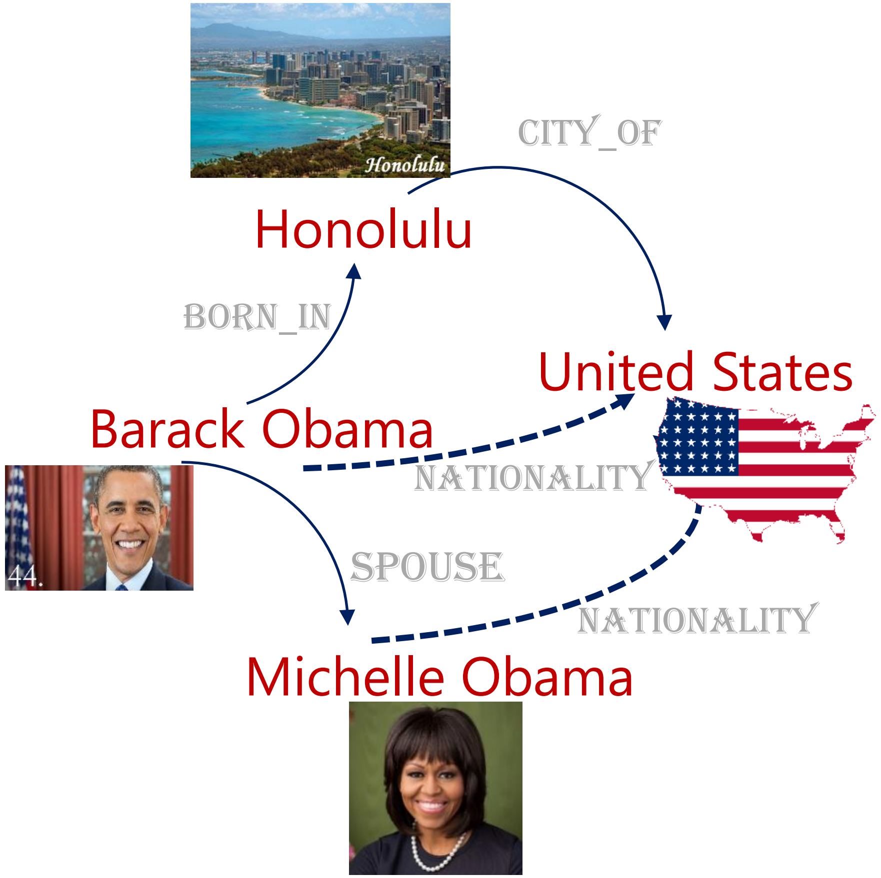
悟 SATORI

Knowledge bases are incomplete

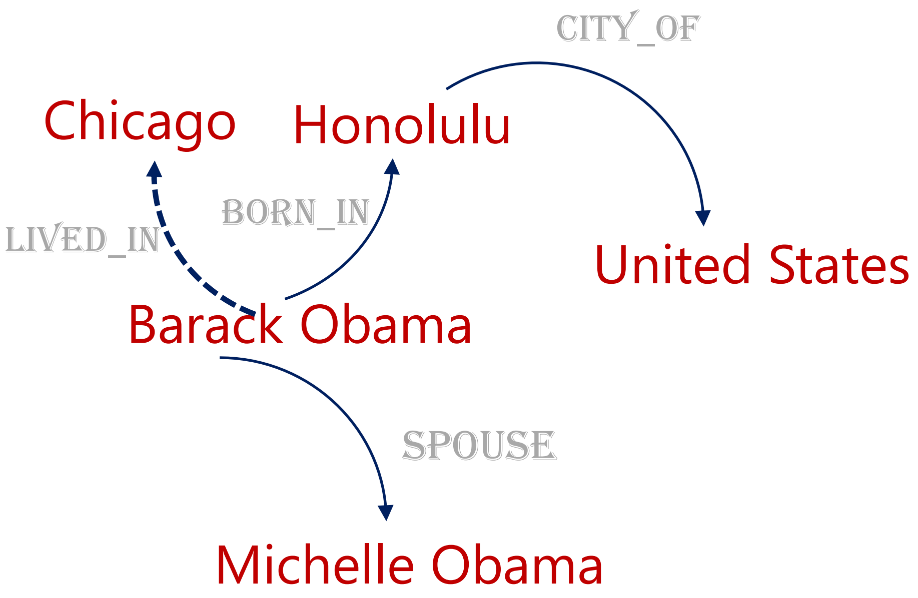


Task: Automatically predict missing relation links
to extend the coverage of the KB.

Information Source: Existing knowledge from KB



Information source: Facts stated in text



Barack Obama worked in Chicago.

A photo of Barack Obama's Chicago house.

Goal: use both sources of information. [Lao et al. 2012], [Gardner et al. 2013,2014] [Riedel et al. 2013] [Neelakantan et al 2015]

Outline

Background

- Embedding-based models for KB completion
  - Using text: Universal Schema

Problems/advances

- Sparsity of textual relations
  - Compositional representations of text
- Are node/relation embeddings sufficient
  - Using observed features
- Inference from multi-step relation paths from KB and text
  - Efficient compositional representation and learning

Conclusion

Embedding-based models for KB completion

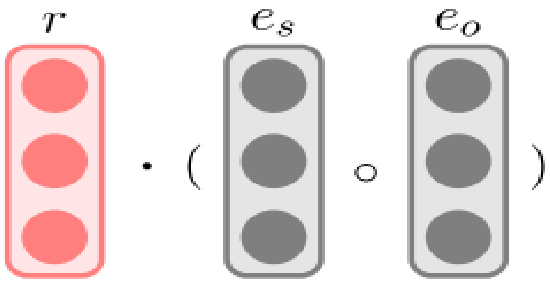
Michelle Obama

Encoding relevant properties of the entities, predictive of their relationships.

LIVED_IN

Encoding relevant properties of the relations that help define the set of entity pairs for which the relation holds.

LIVED_IN Michelle Chicago



f(Michelle Obama, LIVED_IN, Chicago)

RESCAL [Nickel et al 2011], TransE [Bordes et al 2011, 2013], Bilinear-diagonal [Yang et al. 2015]

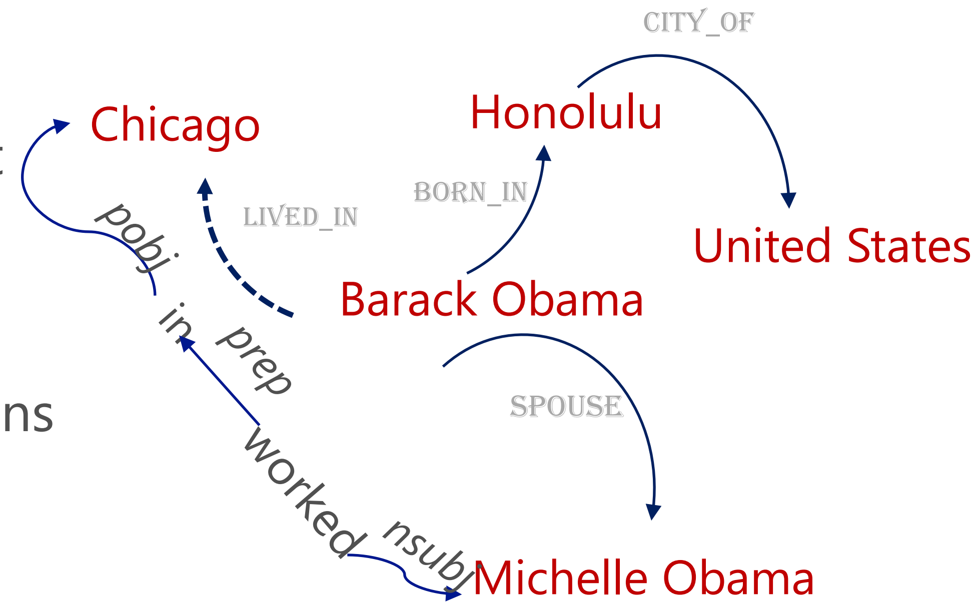
Modeling text in embedding-based models for KB Completion
Universal Schema [Riedel, Yao, Marlin, McCallum 2013]

- Detect mentions of entity pairs in large document collections

  *"Here is a look at the life of **Bill Gates, co-founder of Microsoft**"*

- Represent textual mentions by the linguistic relationship expressed. Define a new "textual relation" for each distinct linguistic relationship.

$$\text{SUBJECT} \xrightarrow{\text{dep}} \text{co-founder} \xrightarrow{\text{prep}} \text{of} \xrightarrow{\text{pobj}} \text{OBJECT}$$



- Learn continuous representations of KB and textual relations

FOUNDER

$$\text{SUBJECT} \xrightarrow{\text{dep}} \text{co-founder} \xrightarrow{\text{prep}} \text{of} \xrightarrow{\text{pobj}} \text{OBJECT}$$

KB relations   Textual relations

Outline

Background

- Embedding-based models for KB completion
  - Using text: Universal Schema

Problems/advances

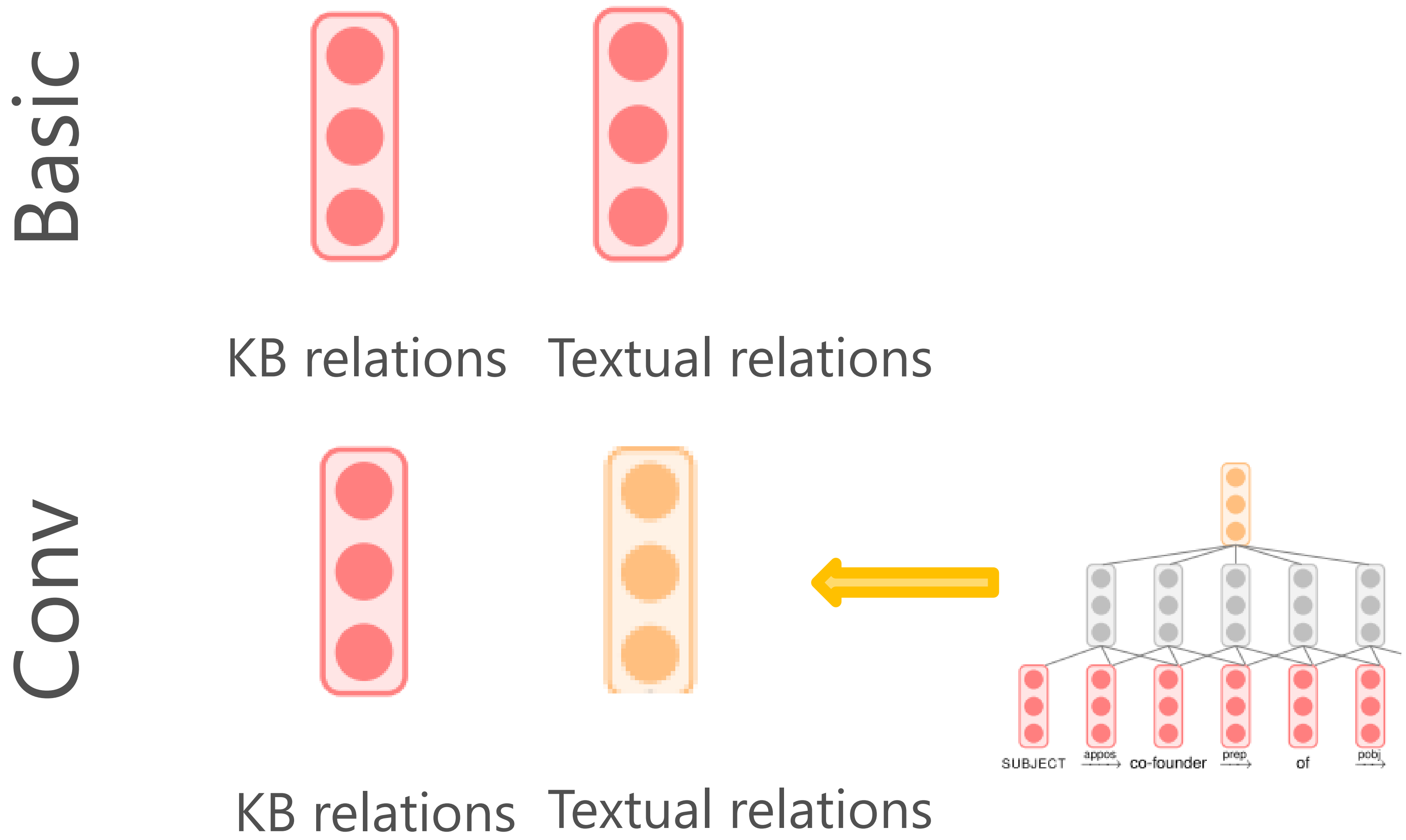
- Sparsity of textual relations
  - Compositional representations of text
- Are node/relation embeddings sufficient
  - Using observed features
- Inference from multi-step relation paths from KB and text
  - Efficient compositional representation and learning

Conclusion

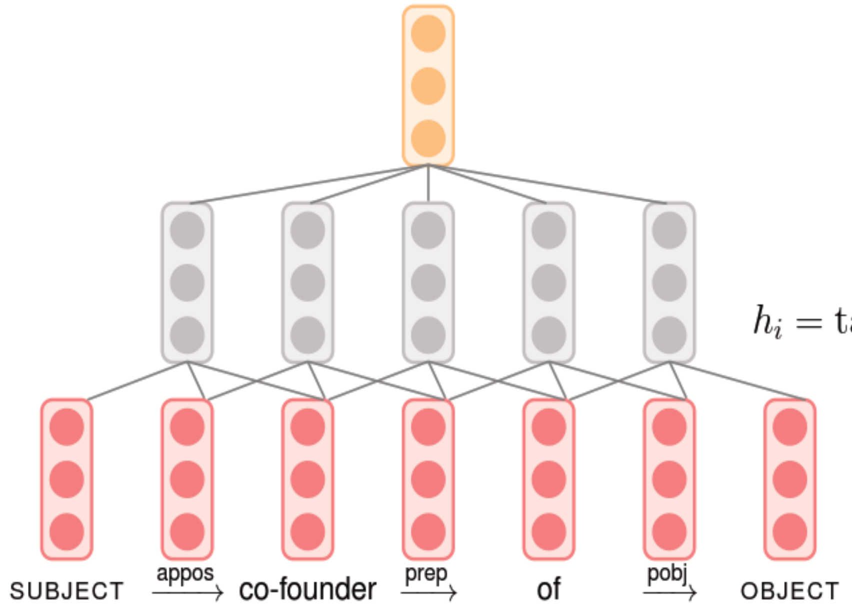
Sparsity of textual relations

Textual Pattern	Count
SUBJECT $\xrightarrow{appos}$ founder $\xrightarrow{prep}$ of $\xrightarrow{pobj}$ OBJECT	12
SUBJECT $\xleftarrow{nsubj}$ co-founded $\xrightarrow{dobj}$ OBJECT	3
SUBJECT $\xrightarrow{appos}$ co-founder $\xrightarrow{prep}$ of $\xrightarrow{pobj}$ OBJECT	3
SUBJECT $\xrightarrow{conj}$ co-founder $\xrightarrow{prep}$ of $\xrightarrow{pobj}$ OBJECT	3
SUBJECT $\xleftarrow{pobj}$ with $\xleftarrow{prep}$ co-founded $\xrightarrow{dobj}$ OBJECT	2
SUBJECT $\xleftarrow{nsubj}$ signed $\xrightarrow{xcomp}$ establishing $\xrightarrow{dobj}$ OBJECT	2
SUBJECT $\xleftarrow{pobj}$ with $\xleftarrow{prep}$ founders $\xrightarrow{prep}$ of $\xrightarrow{pobj}$ OBJECT	2
SUBJECT $\xrightarrow{appos}$ founders $\xrightarrow{prep}$ of $\xrightarrow{pobj}$ OBJECT	2
SUBJECT $\xleftarrow{nsubj}$ one $\xrightarrow{prep}$ of $\xrightarrow{pobj}$ founders $\xrightarrow{prep}$ of $\xrightarrow{pobj}$ OBJECT	2
SUBJECT $\xleftarrow{nsubj}$ founded $\xrightarrow{dobj}$ production $\xrightarrow{conj}$ OBJECT	2
SUBJECT $\xleftarrow{appos}$ partner $\xleftarrow{pobj}$ with $\xleftarrow{prep}$ founded $\xrightarrow{dobj}$ production $\xrightarrow{conj}$ OBJECT	2

Idea: Represent the compositional structure of textual relations (using convolution)



Compositional parametrization of textual relations



$$r = \max\{h_i\}$$

$$h_i = \tanh(W^{-1}v_{i-1} + W^0 v_i + W^1 v_{i+1} + b)$$

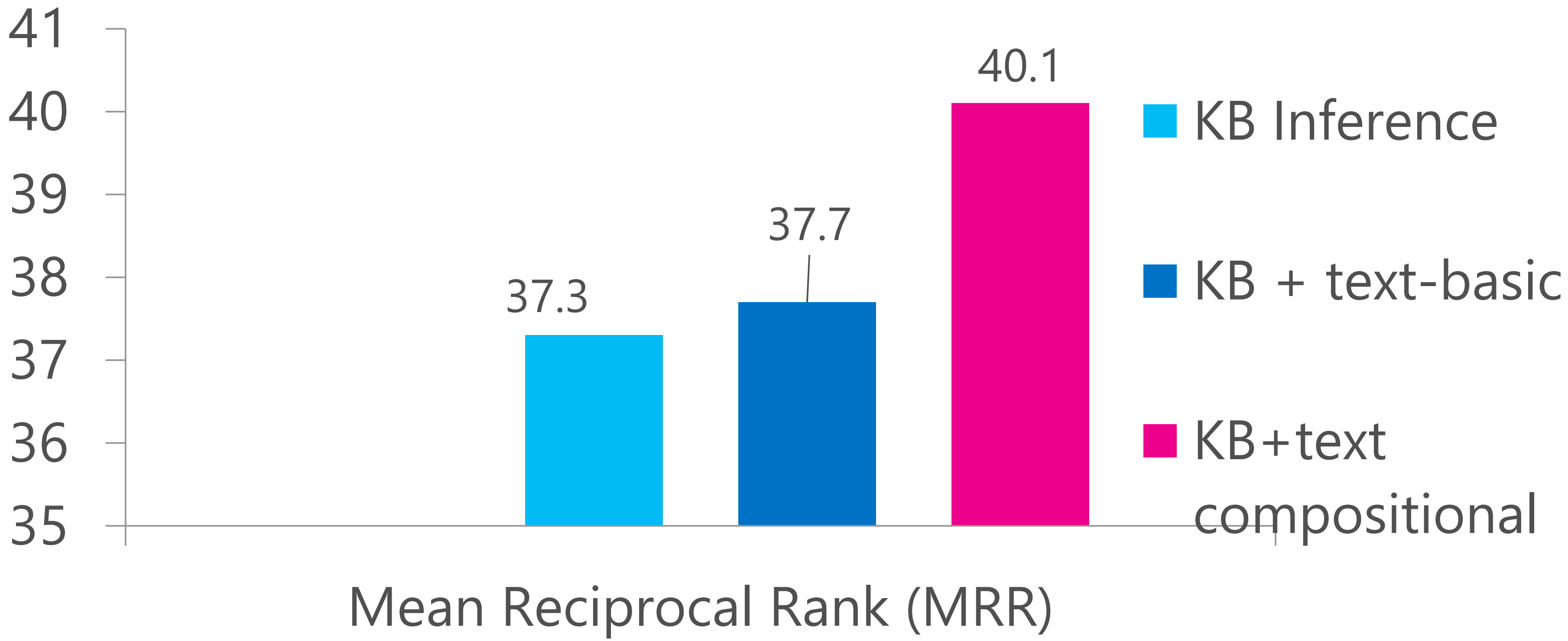
$$v_i = Ve_i$$

[Toutanova, Chen, Pantel, Poon, Choudhury, Gamon, EMNLP 2015]

Results: Using compositional representations of text

Evaluation on held out queries : Where did Michelle Obama live?

Mean reciprocal rank of first correct answer (times 100)



FB15K-237 new dataset from Freebase and ClueWeb.
http://research.microsoft.com/en-us/downloads/3a9bf02d-b791-4e95-b88d-389feef3e421/

Outline

Background

- Embedding-based models for KB completion
  - Using text: Universal Schema

Problems/advances

- Sparsity of textual relations
  - Compositional representations of text
- Are node/relation embeddings sufficient
  - Using observed features
- Inference from multi-step relation paths from KB and text
  - Efficient compositional representation and learning

Conclusion

Can learned embeddings of entities and relations encode sufficient relevant structure from KB+text?

- Extreme cases: symmetric and inverse relationships

$$\text{Barack Obama} \xrightarrow{\text{BORN\_IN}} \text{Honolulu}$$

$$\text{Barack Obama} \xrightarrow{\text{SPOUSE}} \text{Michelle Obama}$$

Training KB

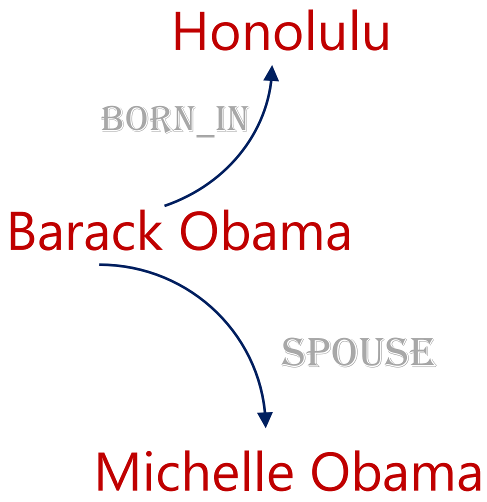
$$\text{Michelle Obama} \xrightarrow{\text{SPOUSE}} ?$$

$$\text{Honolulu} \xrightarrow{\text{PEOPLE\_BORN\_IN}} ?$$

Test queries

Simple observed features model (impoverished PRA)

- Extreme cases: symmetric and inverse relationships



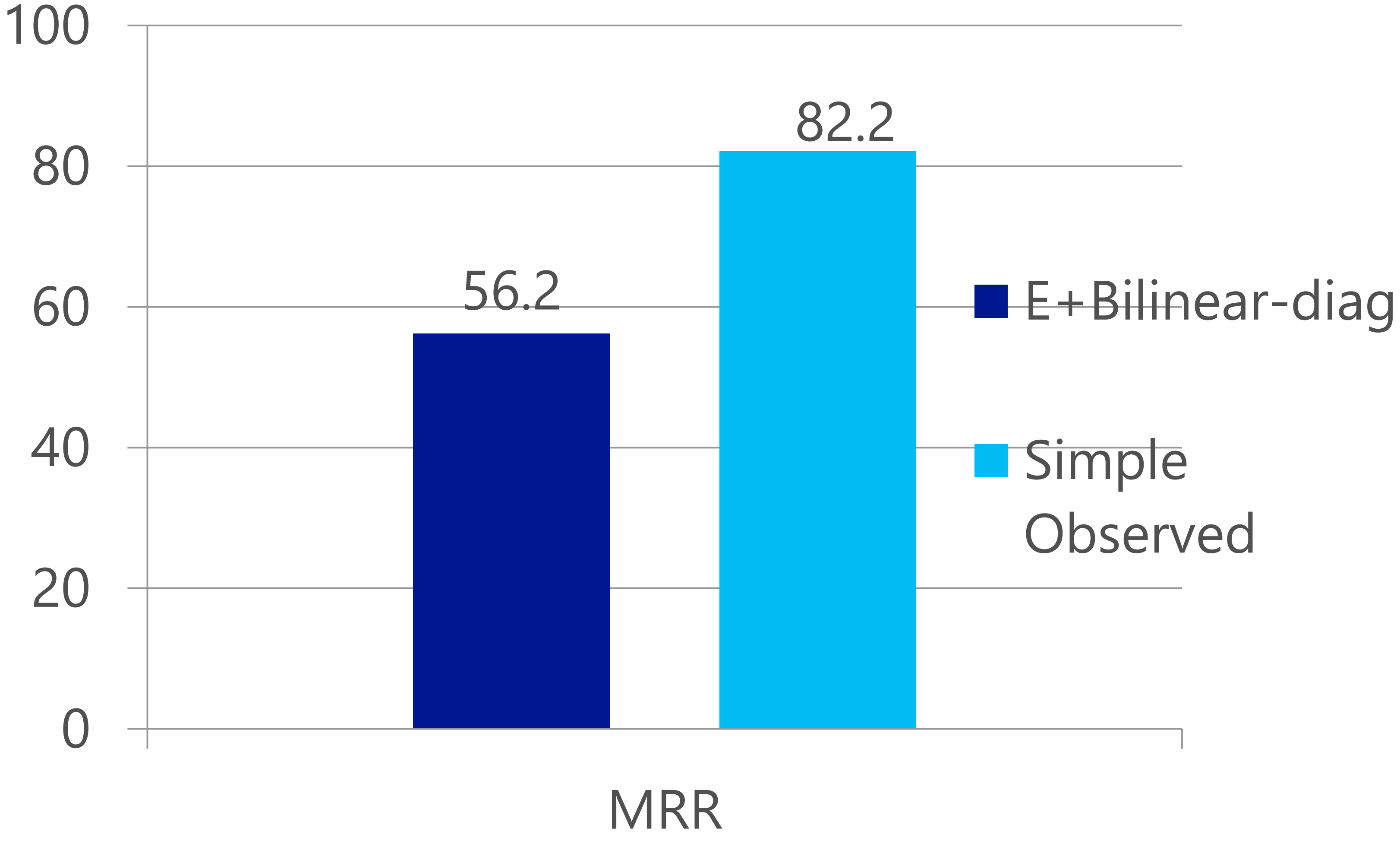
f(x=Michelle Obama, SPOUSE, y=Barack Obama) = $w_1 \phi_1 = 99.8$

ϕ_1 $\quad$ y $\xrightarrow{\text{SPOUSE}}$ x $\in K$ $= 1$

K Training KB

Toutanova and Chen, CVSM Workshop, 2015

Embedding versus simple observed features on FB15K [Bordes et al 2013]



Result largely due to redundancy of knowledge base, but gains from observed features (e.g. direct text links) also seen in other datasets in our work and prior work [e.g. Knowledge Vault]

Outline

Background

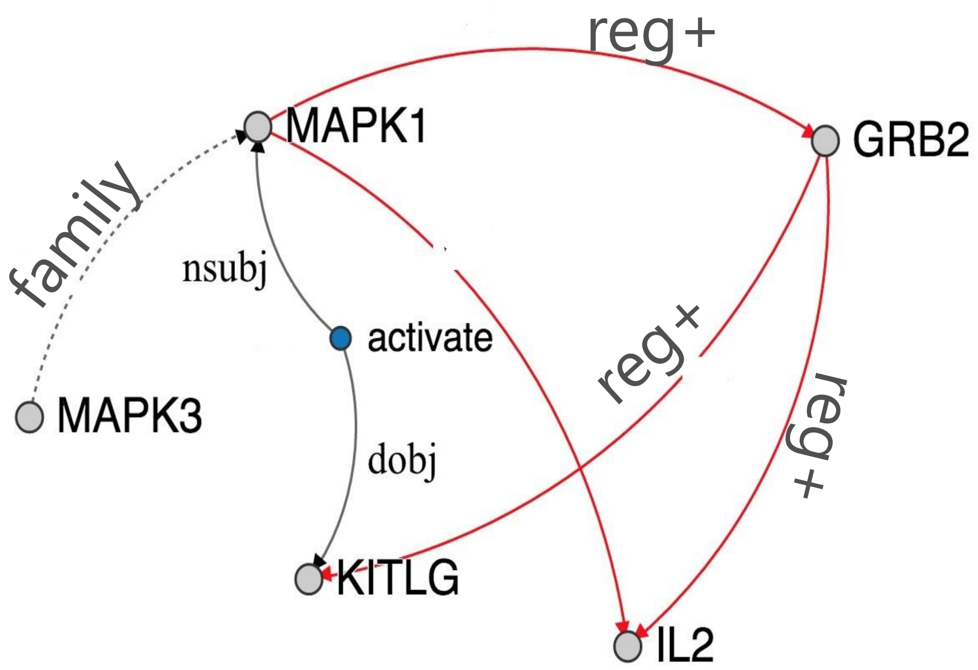
- Embedding-based models for KB completion
  - Using text: Universal Schema

Problems/advances

- Sparsity of textual relations
  - Compositional representations of text
- Are node/relation embeddings sufficient
  - Using observed features
- Inference from multi-step relation paths from KB and text
  - Efficient compositional representation and learning

Conclusion

Relation paths between genes in genomics KB+text



Paths from GRB2 to MAPK3

P1: GRB2 reg+ IL2 _reg+ MAPK1 _family MAPK3

P2: GRB2 reg+ KITLG _dobj-[activate]-nsubj MAPK1 _family MAPK3

P3: GRB2 _reg+ MAPK1 _family MAPK3

Prior work on learning with relation+text paths: Path Ranking Algorithm [Lao & Cohen 2010, Lao et al 2012]

- For an ordered entity pair GRB2, MAPK3, compute path-constrained random walk probabilities to reach MAPK3 from GRB2 given each allowed path type (sequence of relation types)

P1: GRB2 <reg+, _reg+, _family> MAPK3	0.30
P2: GRB2 <reg+, _dobj-[activate]-nsubj, _family> MAPK3	0.40
P3: GRB2 <_reg+,_family> MAPK3	0.05

- Learn weights for path features to predict relationships between entities.

Prior work on learning with relation+text paths

Sparsity of path types: exponential number of path types; text makes problem especially challenging

- Computational efficiency problem: memory to store path features, time to compute path features and score node pairs
  - Select a limited number of allowable path types [Lao et al 2011,Lin et al 2015, Guu et al 2015];
  - Faster ways to compute random walk probabilities [Lao 2012] or forgo use of random walk probabilities [Gardner and Mitchell 2015]
- Statistical estimation problem: too many parameters to learn without parameter sharing
  - Integrate vector space similarity in random walks: [Gardner et al 2014]
  - Learn compositional representations of path features using RNN neural networks [Neelakantan et al 2015] or simpler addition of edge embeddings [Lin et al 2015]

Can we do even more with compositional representation of path types?

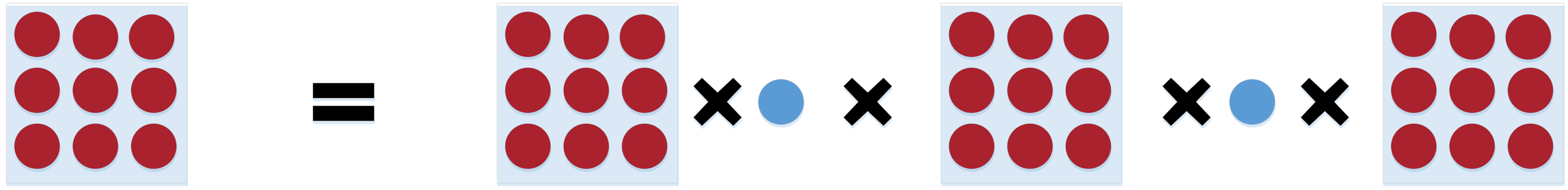
- Finer-grained information from relation paths:
  - What nodes does a path pass through?

P1: GRB2 <reg+, IL2, _reg+, MAPK1, _family> MAPK3 0.30

  - Paths passing through particular genes might be more important
  - But in traditional approach blows up memory required to store paths and parameters to learn.

- Our approach: compositional representations of paths including path types and nodes [Toutanova, Lin, Yih, Poon, Quirk, ACL 2016]

Compositional representations of paths including nodes

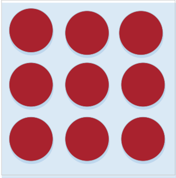


$$f(s,r,t) = x_s^T W_r x_t + \mathrm{vec}\Big(\sum_{\pi} P(t|s,\pi)\Phi(\pi)\Big)^{\mathrm{T}} \mathrm{vec}(\mathrm{W_r})$$

- Turns out we don't need to store explicit paths and compute path-constrained random walk probabilities to evaluate and learn the scoring function!

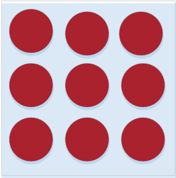
Exact algorithm to compute weighted sum of path representations

$$F_l(s,t) = \sum_{|\pi|=l} P(t|s,\pi)\Phi(\pi))$$



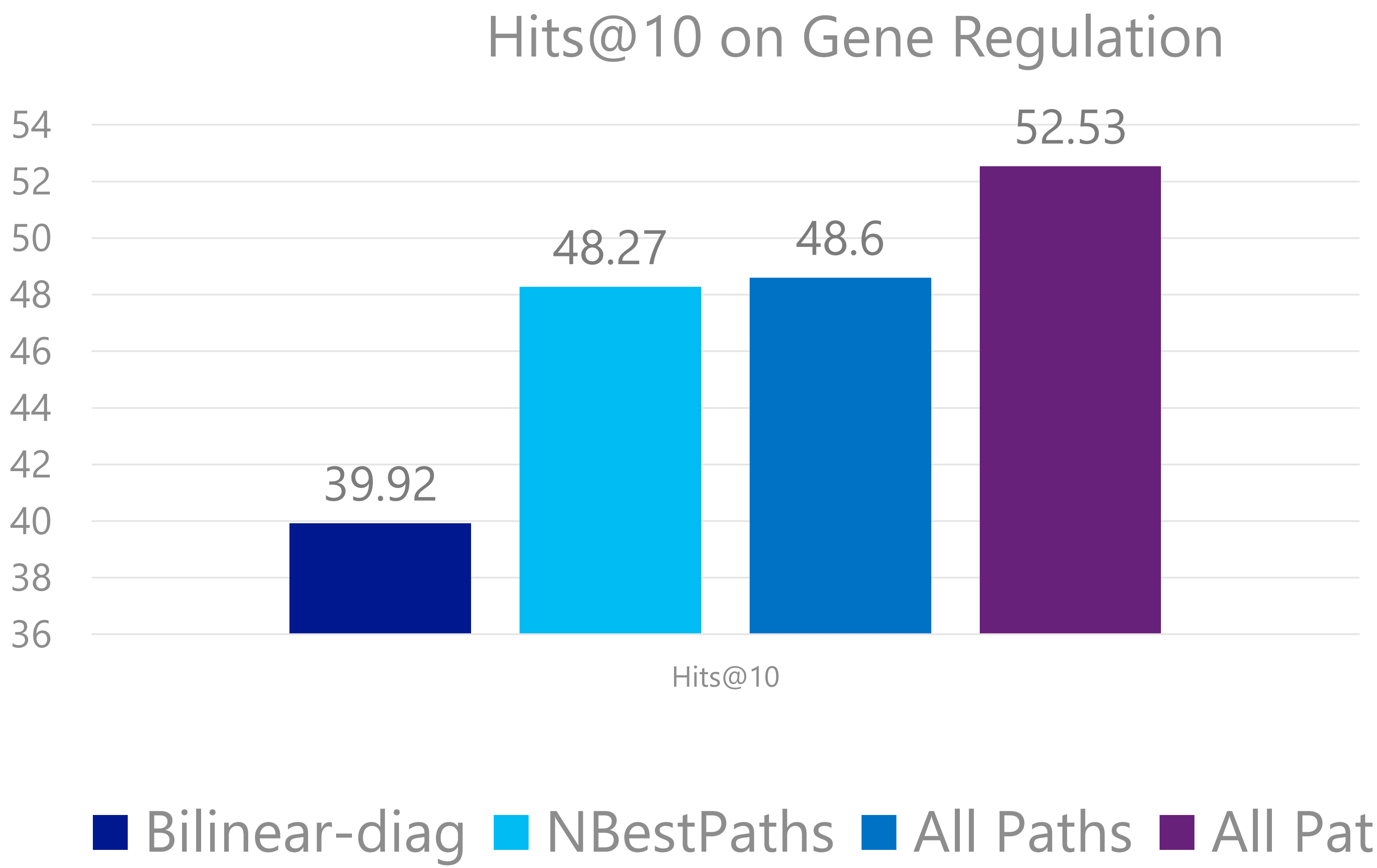
- Incrementally compute the sum of path representations by using sums of paths up to length *l* to compute sums of paths up to length *l+1*

$$F_{l+1}(s,t) = \sum_k F_1(s,k)F_l(k,t)$$



- Memory for connected entity pairs equal to dimensionality of relation embeddings for each path length: can be much lower than number of active features per pair in PRA
- The time to compute scores is $O(LEN_e)$ (not exponential with path length)
- Adding node weights does not increase asymptotic complexity.
- Can be much more efficient than standard approach for large L and a large set of possible relations R.

Results: using compositional representations of relation paths from KB and text relations



NCI-PID database + textual mentions from Pubmed

Conclusion

- Compositional representations of text help improve universal schema models
- Explicit observed features from KB+text graph can bring substantial benefits
- Compositional representations of KB+text relation paths enable richer context (path nodes) and exact efficient computation