A Hybrid Neural Model for Type Classification of Entity Mentions

Li Dong[†*] **Furu Wei**[‡] **Hong Sun**[$] **Ming Zhou**[‡] **Ke Xu**[†]
[†]State Key Lab of Software Development Environment, Beihang University, Beijing, China
[‡]Microsoft Research, Beijing, China
[$]Microsoft Corporation, Beijing, China
donglixp@gmail.com {fuwei,hosu,mingzhou}@microsoft.com kexu@nlsde.buaa.edu.cn

Motivation

- Types group entities to categories
- Entity types are important for various NLP tasks
  - Question answering
  - Relation extraction
  - Semantic role labeling
  - …
- Our task
  - predict an entity mention's type

Type Classification of Entity Mentions

- Input
  - $\underbrace{[c_{-S} \dots c_{-1}]}_{\text{left context}}$ $\underbrace{[w_1 \dots w_n]}_{\text{mention}}$ $\underbrace{[c_1 \dots c_S]}_{\text{right context}}$
- Output
  - Type

- [an initiative sponsored by][Bill & Melinda Gates Foundation][to fight HIV infection]

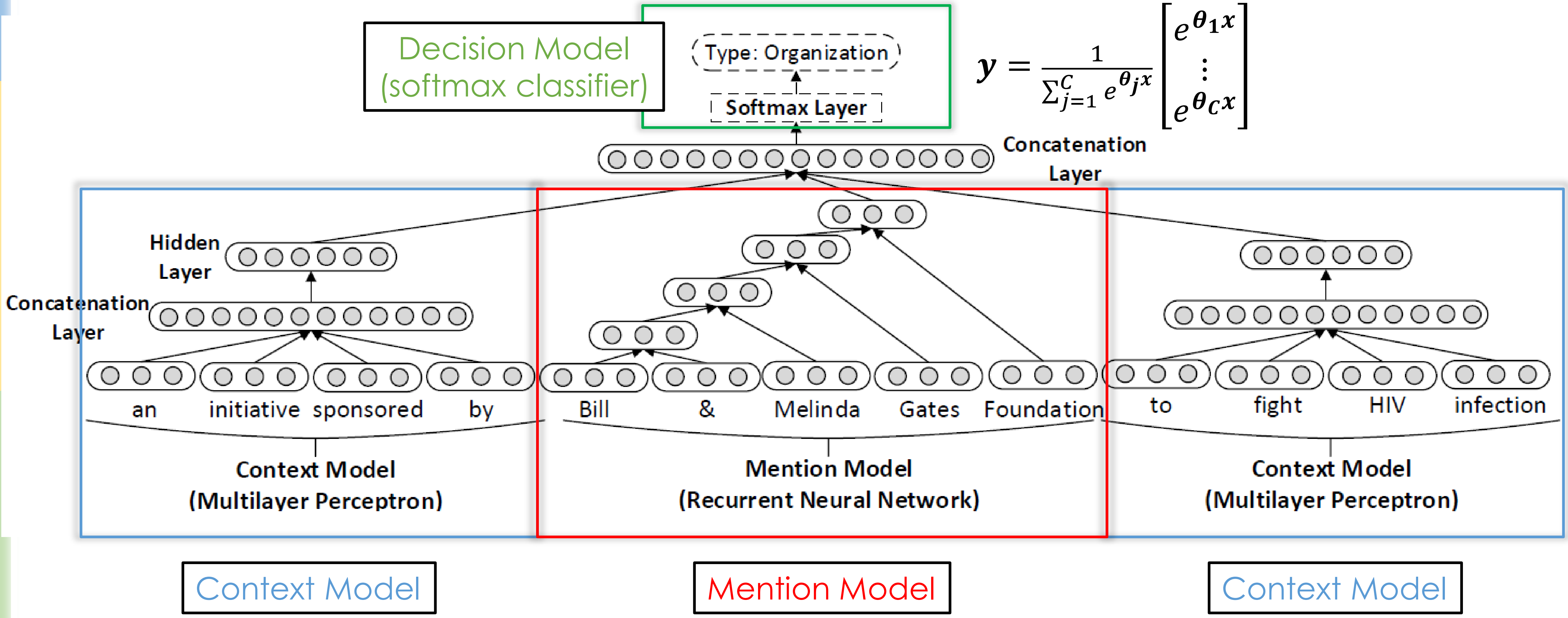
Organization

- Mention
  - Bill & Melinda Gates Foundation (*Organization*)
  - Bill, Melinda, Gates -> {Person Name}
  - {Person Name} + Foundation -> Organization
- Context
  - [The greater part of][**Gates**][' population is in Marion County .] (*Location*)
  - [**Gates**][was a baseball player .] (*Person*)

Related Work

- Named Entity Recognition
  - Limited types
    - *Location, Person, Organization, Misc*
  - (e.g.) Question Answering
    - Questions are classified into more answer types
- Named Entity Linking (1. Link to an entity in knowledge base 2. Query its entity type)
  - Performance drops for uncommon entities
  - (e.g.) Question Answering
    - Extracted answer candidate may not appear in knowledge base
  - NEL is a harder problem than type classification
- Design rich features
  - N-gram, morphological features, gazetteers, WordNet, ReVerb patterns, POS tags, dependency parsing results, etc.

Architecture

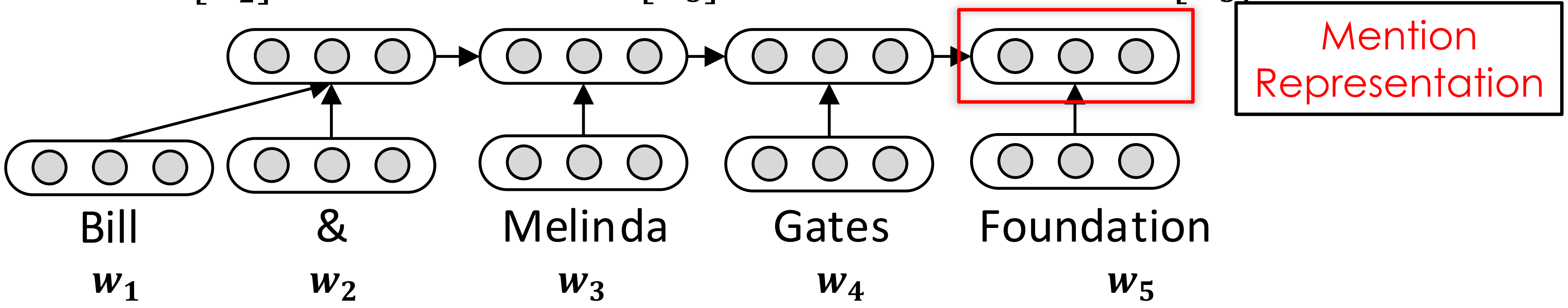


$$\boldsymbol{y} = \frac{1}{\sum_{j=1}^{C} e^{\theta_j x}} \begin{bmatrix} e^{\theta_1 x} \\ \vdots \\ e^{\theta_C x} \end{bmatrix}$$

RNN-based Mention Model

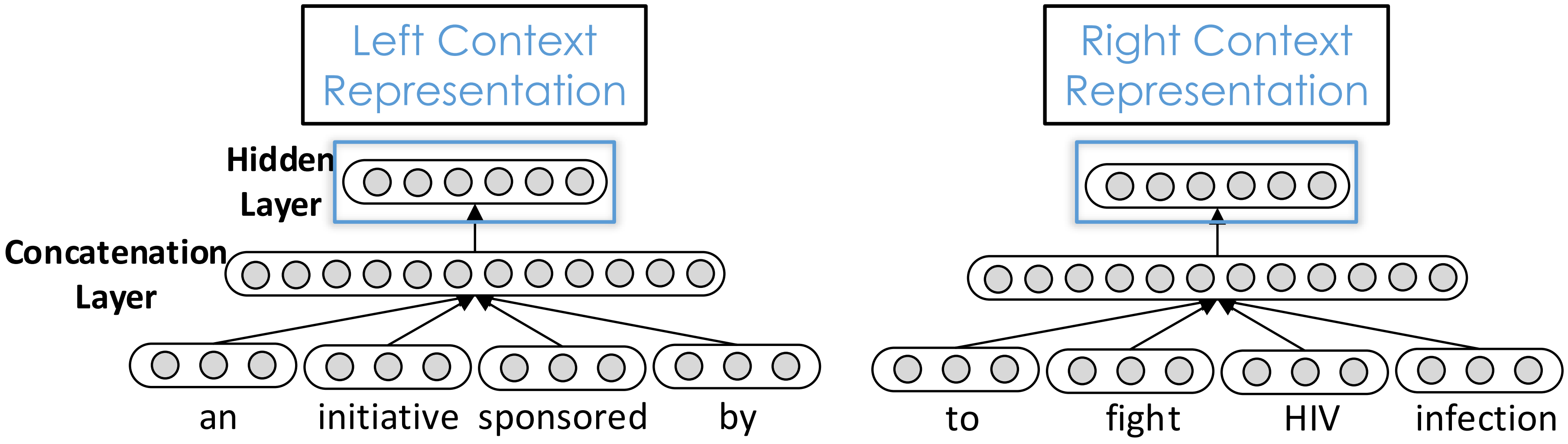
- Learn composition patterns for entity mention
  - {Name} + Foundation / University -> (Organization)
  - {Body Region} + {Disease} -> (Disease)
- **R**ecurrent **N**eural **N**etworks (Elman Networks)
  - Use a global composition matrix to compute representation recurrently
  - A natural way to learn composition patterns

$\boldsymbol{v_2} = tanh(W \begin{bmatrix} \boldsymbol{w_1} \\ \boldsymbol{w_2} \end{bmatrix} + \boldsymbol{b^m})$ $\boldsymbol{v_3} = tanh(W \begin{bmatrix} \boldsymbol{v_2} \\ \boldsymbol{w_3} \end{bmatrix} + \boldsymbol{b^m})$ ... $\boldsymbol{v_5} = tanh(W \begin{bmatrix} \boldsymbol{v_4} \\ \boldsymbol{w_5} \end{bmatrix} + \boldsymbol{b^m})$



MLP-based Context Model

- Use context to disambiguate
  - [The greater part of][**Gates**][' population is in Marion County .] (*Location*)
  - [**Gates**][was a baseball player .] (*Person*)
- **M**ulti**L**ayer **P**erceptrons
  - Location-aware, jointly trained



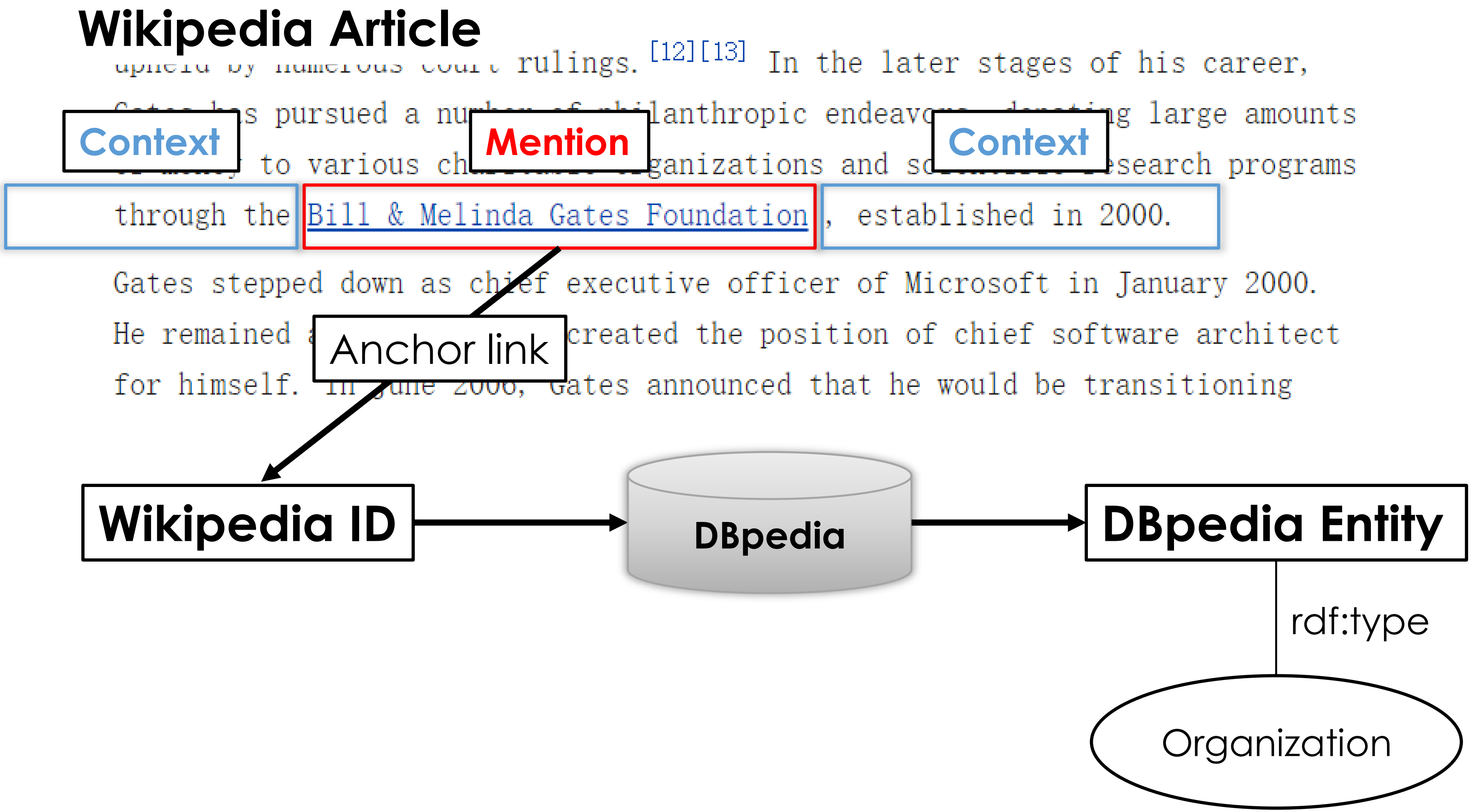
Model Training

- Objective function

$$\underset{\theta}{\text{minimize}} \underbrace{-\sum_i \sum_j \boldsymbol{t}_j^i \log \boldsymbol{y}_j^i}_{\text{cross entropy loss}} + \underbrace{\frac{\lambda_\theta}{2} \|\theta\|_2^2}_{\text{regularization}}$$

- Back-propagation algorithm
  - Back-propagate errors of softmax classifier to other layers
- Optimization
  - Mini-batched AdaGrad

Automatically Generating Training Data



Automatically Generating Training Data

- DBpedia ontology
  - 22 top-level types

Organisation, MeanOfTransportation, Holiday, Work, Food, Award, AnatomicalStructure, Device, Colour, Language, TopicalConcept, EthnicGroup, Currency, Disease, Drug, Person, Place, Activity, CelestialBody, Event, Species, BioChemSubstance

- Wiki-22
  - #Train: 2 million
  - #Dev: 0.1 million
  - #Test: 0.28 million

Evaluation on Wiki-22

- micro-F1 / macro-F1 score
- Baseline methods
  - Support Vector Machine (SVM)
  - Multinomial Naive Bayes (MNB)
  - Sum word vectors (ADD)
    - Use a softmax classifier
- *-mention
  - Only use mention
- *-context
  - Only use context
- *-joint
  - Use both mention and context

Method	Micro-F1	Macro-F1
SVM-mention	90.2	89.7
MNB-mention	87.0	87.6
ADD-mention	90.1	90.7
HNM-mention	**93.4**	**93.6**
SVM-context	76.3	73.3
MNB-context	72.8	70.0
ADD-context	75.4	73.1
HNM-context	**81.1**	**78.3**
SVM-joint	93.5	93.4
MNB-joint	85.9	82.8
ADD-joint	94.1	93.9
HNM-joint (our)	**96.8**	**96.5**

Comparison with Previous Systems

- HYENA [Yosef et al., 2012]
  - Support Vector Machine
  - unigrams, bigrams, and trigrams of mentions, surrounding sentences, mention paragraphs, part-of-speech tags of context words, gazetteer dictionary
- FIGER [Ling and Weld, 2012]
  - Perceptron
  - unigrams, word shapes, part-of-speech tags, length, Brown clusters, head words, dependency structures, ReVerb patterns

Dataset	Method	Micro-F1	Macro-F1
Wiki-5	HYENA	**95.2**	91.9
	HNM-joint	95.0	**93.6**
News	FIGER	72.6	80.1
	HNM-joint	**75.1**	**80.6**

Evaluation on Unseen Mentions

- Evaluate on unseen mentions (length > 2)
  - Mentions which do not appear in the train set

Method	Micro-F1	Macro-F1
SVM-mention	75.8	68.8
MNB-mention	75.5	69.0
ADD-mention	76.1	69.3
HNM-mention	**82.5**	**75.6**

- Help us deal with uncommon or unseen mentions
  - RNN-based mention model utilizes the compositional nature of mentions

Examples: Compositionality of Mentions

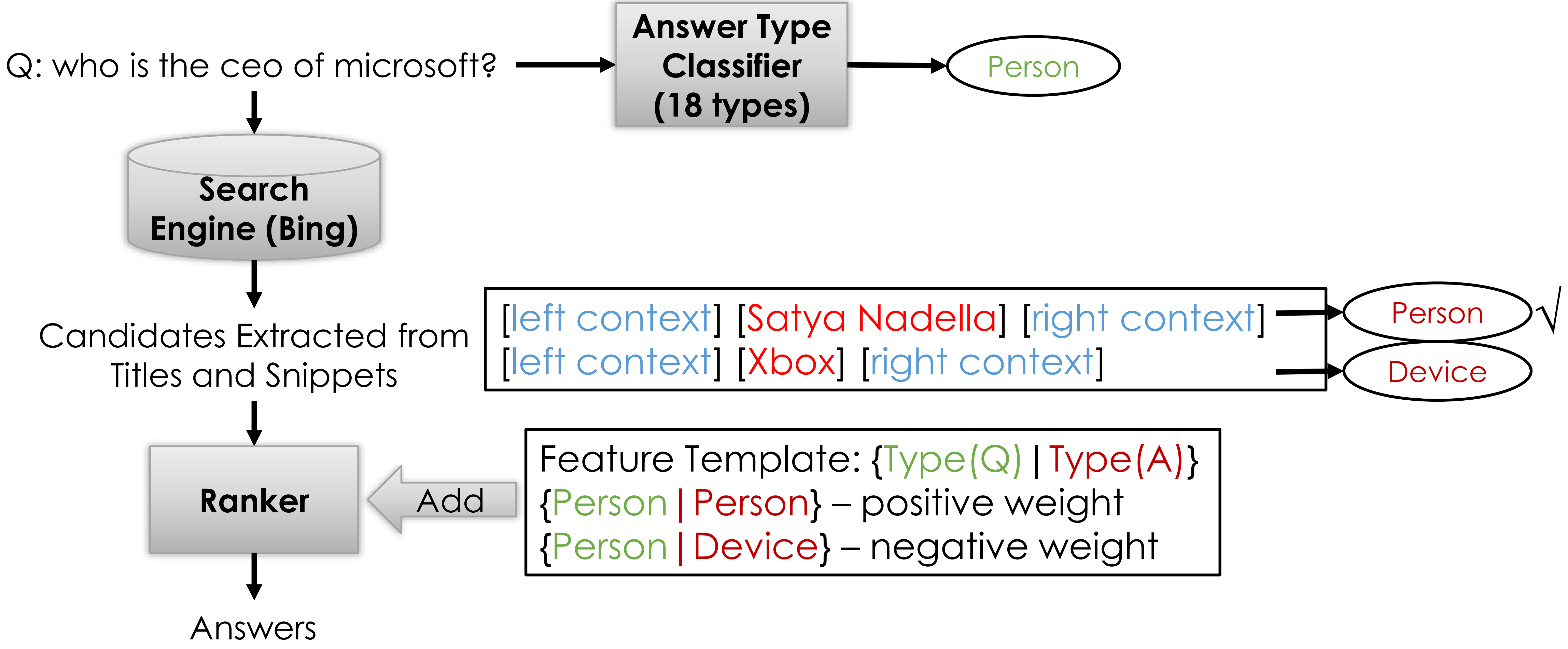
- Query similar mention examples
  - cosine similarity of mentions' vector representations

English civil war	Spanish civil war / Greek civil war / Nigerian civil war / Angolan civil war
Columbia University School of Law	Northwestern University School of Law / West Virginia University College of Law / University of Iowa College of Law / Golden Gate University School of Law
Subdural Hematoma	Intracranial Haemorrhage / Cardiac Arrhythmia / Duodenal Ulcer / Arterial Thrombosis
Joseph Jefferson Award	Margaret A. Edwards Award / Marian Engel Award / Doug Wright Award / Timothy Findley Award
Red-bellied Lemur	Oriental White-eye / Red-legged Honeycreeper / Black-crowned White-eye / Snowy Egrets

- Mentions that are of similar patterns are closer

Evaluation in Question Answering (QA)

- Web-based QA system [Cucerzan and Agichtein, 2005; Lin, 2007]
  - Add Q&A type interaction feature template



Evaluation in Question Answering (QA)

- WebQuestions dataset [Berant et al., 2013]
  - Manually annotated question-answer pairs

Method	Acc@1	Acc@3	Acc@5
w/oTYPE	29.2	50.8	61.2
w/TYPE	**33.5**	**55.6**	**64.4**

Table 6: Evaluation results on the QA task. Type information obtained by our approach improves the accuracy. w/oTYPE: Without using type features in the answer ranking model. w/TYPE: Using type features in the answer ranking model.

- Our type classifier improves the accuracy of QA systems

Conclusion and Future Work

- Conclusion
  - Recurrent Neural Networks are good at learning soft patterns
    - Compositional nature of entity mentions
    - Generalize for Unseen or uncommon mentions
  - Automatically generate training data instead of annotating manually
  - Type information is important for many NLP tasks
- Future work
  - Fine-grained type classification
    - Person -> doctor, actor, etc.
  - Utilize hierarchical taxonomy
  - Multi-label
  - Utilize global information (e.g., document topic)
  - …