# Neural Cross-Lingual Entity Discovery and Linking

**Avi Sil**

Joint work with: Georgiana Dinu, Gourab Kundu and Radu Florian

IBM Research AI

# Outline

- Architecture for the IBM Entity Discovery & Linking (EDL) System
  - Model & Results
    - Mention Detection
    - In doc Coref Resolution
    - Entity Linking & Clustering

# Outline

- Architecture for the IBM Entity Discovery & Linking (EDL) System
  - Model & Results
    - Mention Detection
    - In doc Coref Resolution
    - Entity Linking & Clustering

Neural & Traditional Models

# Mention Detection (By: Avi, Georgiana, Hans)

- Standard IOB sequence classifier, trained on the task

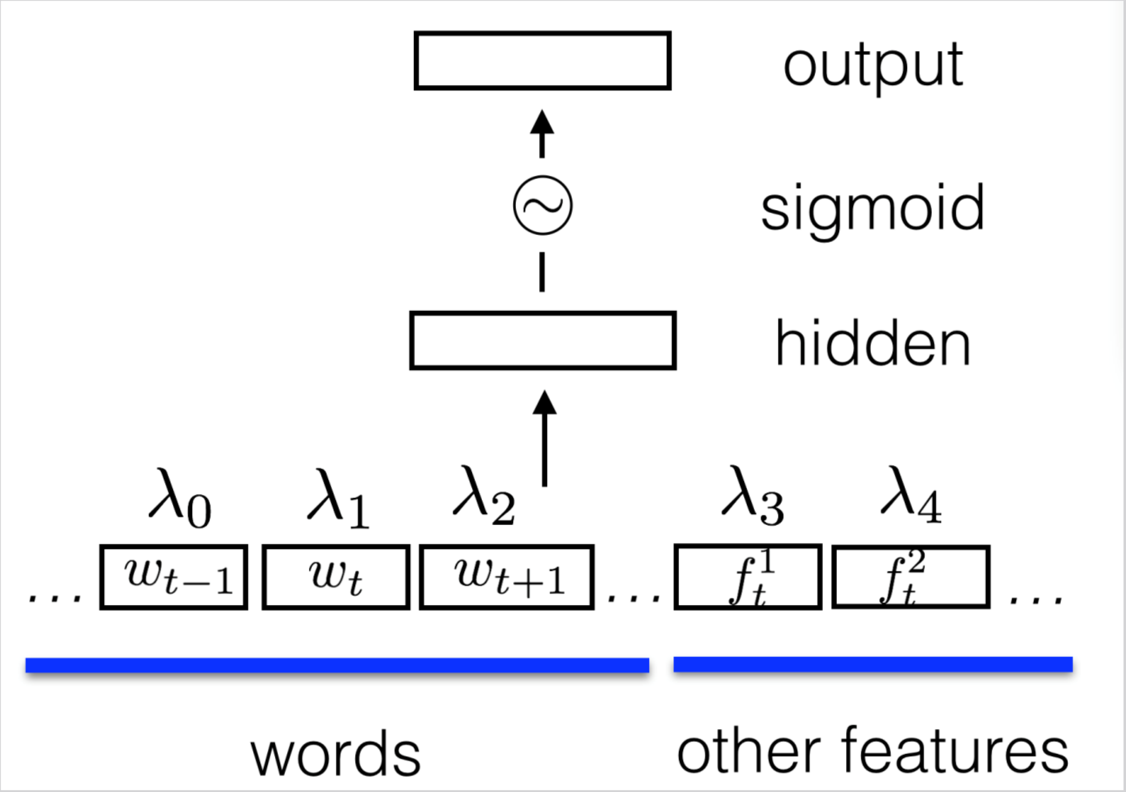
- 2 main classifiers: CRF and Neural Network-based

# Mention Detection (NN)

- Model probability:

$$P(y_t \mid X, y_{t-1})$$

- Additional features: Gazetteers, Character-level LSTMs
- Recurrence: previous 2 labels are embedded and added as input



# System Combination for Mention Detection

- Both systems (CRF, NN) have high precision
- We combine them as follows
  - Start with the "best" system
  - For each consequent system
    - Add any mentions that do not overlap with the current output

| | 2016 | | 2017 | |
|---|---|---|---|---|
| | CRF - dev | NN - dev | NN+CRF - tst | |
| English | 0.803 | 0.843 | 0.806 | |
| Spanish | 0.785 | 0.809 | 0.785 | |
| Chinese | 0.811 | 0.843 | 0.699 | 0.75* CharCNNs |

The Lample model didn't produce better results on our dev data.

# Pilot Task – minimally supervised transfer

- Train monolingual embeddings in En and foreign language
- Use a small dictionary to train a map from a foreign language into the English embedding space (Mikolov 13)
- Train a En mention detection model
- Decode new languages using the En model and mapped embeddings

# Mention Detection for Pilot Task

- Weak classifiers:
  - Silver-data (Pan et.al16) trained NN models
  - Cross-lingual transfer of models with: 1. TAC data and 2. In-house mention detection data

- Train a NN classifier to combine all the weak classifier outputs
- Use Spanish as a test case, apply to all other languages

| | Silver-trained | Best transfer | Combination | Supervised |
|---|---|---|---|---|
| Spanish | 0.335 | 0.609 | 0.704 | 0.809 |

Pan et.al_ACL16

# Outline

- Architecture for the IBM Entity Discovery & Linking (EDL) System
  - Model & Results
    - ✓ Mention Detection
    - **In doc Coref Resolution**
    - Entity Linking & Clustering

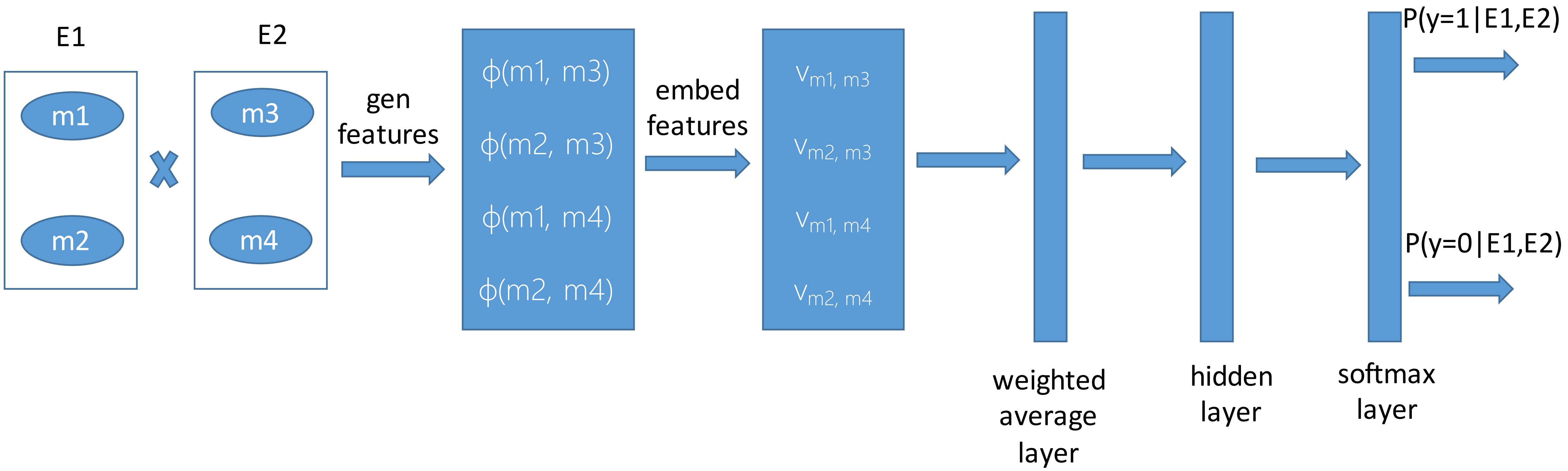
# In document Coreference Resolution (By: Gourab)

- All mentions in a document are clustered into entities using an in document coreference system
- The canonical mention of an entity is linked using EL system
- The link of canonical mention is assigned to all mentions in the entity
- We use 2 different coreference systems in this evaluation
  - MaxEnt Model
  - Neural network based Model

# Neural Network Model

- This model is used for languages without any gold standard training data
  - low resource languages like Nepali
- This model is trained over English coreference data using multilingual embeddings
- Subsequently, the model is tested over data from new language without **any retraining**

# Network Architecture




$$\sum_{i=1,2}\sum_{j=1,2}\rho(type_{m_j})\, v_{m_i,m_j}$$

# Results of NN model

- Model is trained with multilingual embeddings over
  - TAC 15 training portion of English coreference data
  - TAC 16 test portion of English coreference data
- Model is tested over
  - TAC 15 test portion of 3 languages

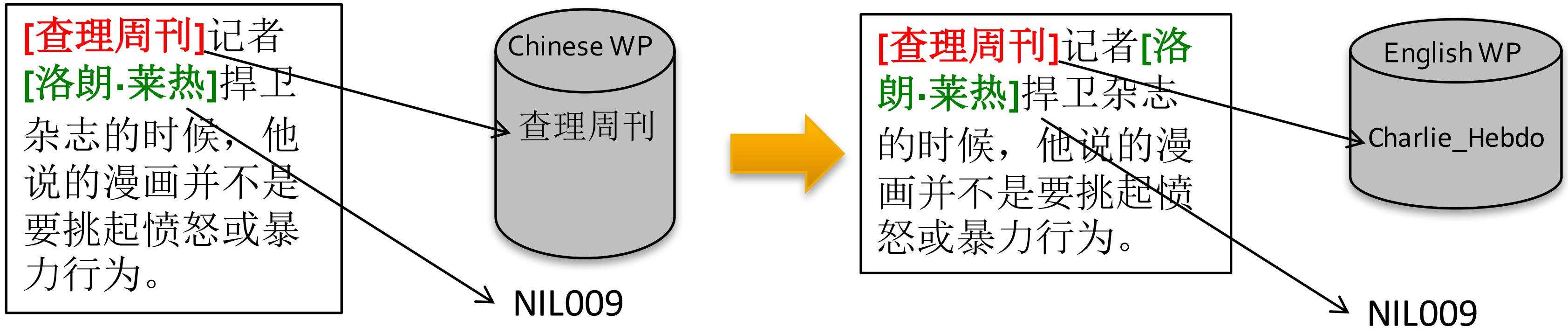
| Language | MUC | B3 | CEAF |
|---|---|---|---|
| TAC 15- test-Eng | 0.9 | 0.89 | 0.84 |
| TAC 15-test-Spa | 0.91 | 0.92 | 0.88 |
| TAC 15- test-Cmn | 0.97 | 0.96 | 0.91 |

# Outline

- Architecture for the IBM Entity Discovery & Linking (EDL) System
  - Model & Results
    - ✓ Mention Detection
    - ✓ In doc Coref Resolution
    - **Entity Linking & Clustering**

# IBM LIEL system (ACL 2016)

- Language Independent EL system: LIEL (Sil & Florian,16)
  - Collective disambiguation model based on Maximum Entropy




- SOTA performance on TAC evaluation & other benchmarks

# IBM NN EL system (AAAI 2018)

- New system
- Neural Cross-lingual Entity Linking
  - Zero-shot model
  - Avi Sil, Gourab Kundu, Radu Florian, Wael Hamza
  - AAAI 2018

# Problem Formulation (English EL)

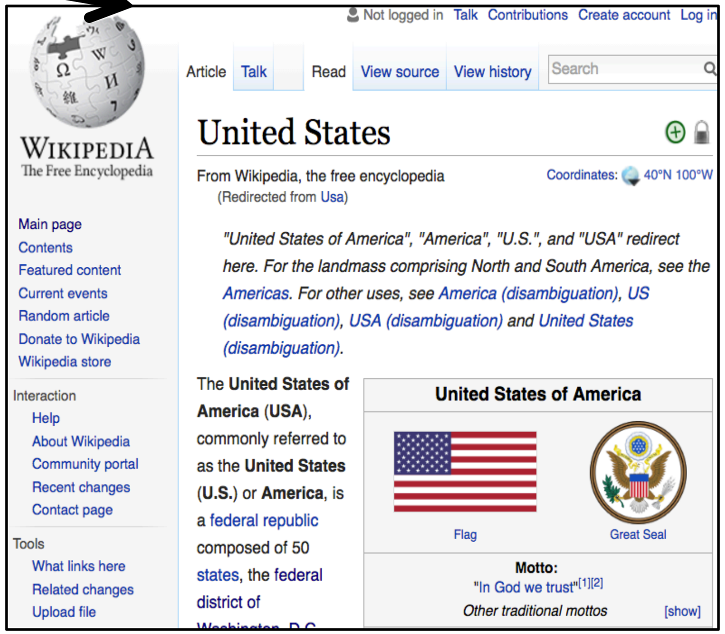
- **Given**: Query mention $\boldsymbol{m}$ and a document $\boldsymbol{D} \in \boldsymbol{en}$ and Wikipedia $\boldsymbol{KB_{en}}$
- **Step 1 (Fast Search)**: Extract the most likely list of links $\boldsymbol{l_{j_1}}$,.., $\boldsymbol{l_{j_m}}$ for $\boldsymbol{m}$ in $\boldsymbol{D}$
- **Step 2 (Ranking)**: Estimate: $$\hat{l}^m = \arg\max_j P(C|m, D, l_j^{(m)})$$
- where "$C$" is the consistency measure **for matching contexts between**:
  - the pair $(\boldsymbol{m}, \boldsymbol{D})$ and a Wikipedia link $\boldsymbol{l_j}$

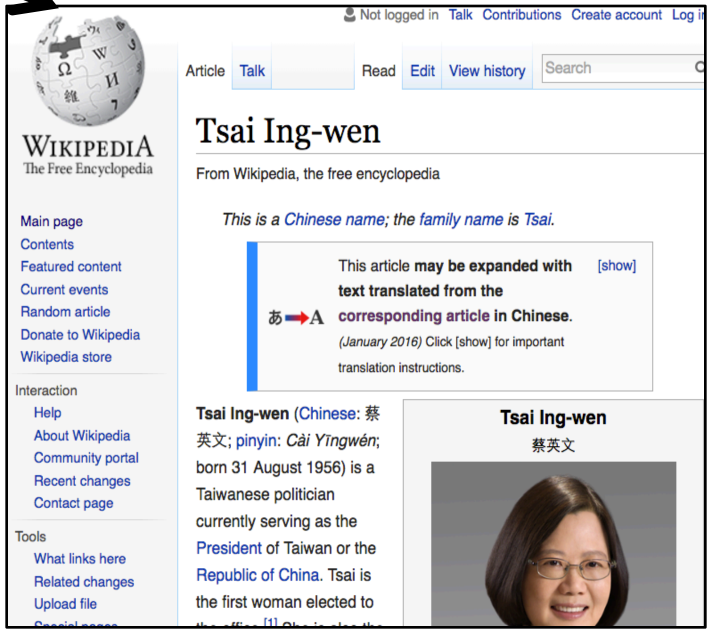
# Problem Formulation (cross-lingual EL)

- **Given**: Query mention $m$ and a document $D \in$ tr and Wikipedia $KB_{en}$
- **Step 1 (Fast Search)**: Extract the most likely list of links $l_{j_1}, .., l_{j_m}$ for $m$ in $D$
- **Step 2 (Ranking)**: Estimate: $$\hat{l}^m = \arg\max_j P(C|m, D, l_j^{(m)})$$
- where "$C$" is the consistency measure **for matching contexts between**:
  - the pair $(m, D)$ and a Wikipedia link $l_j$

# Cross-Lingual Entity Linking

Tayvan, ABD ve İngiltere'de hukuk okuması, Tsai'ye bir LL.B. kazandırdı ...

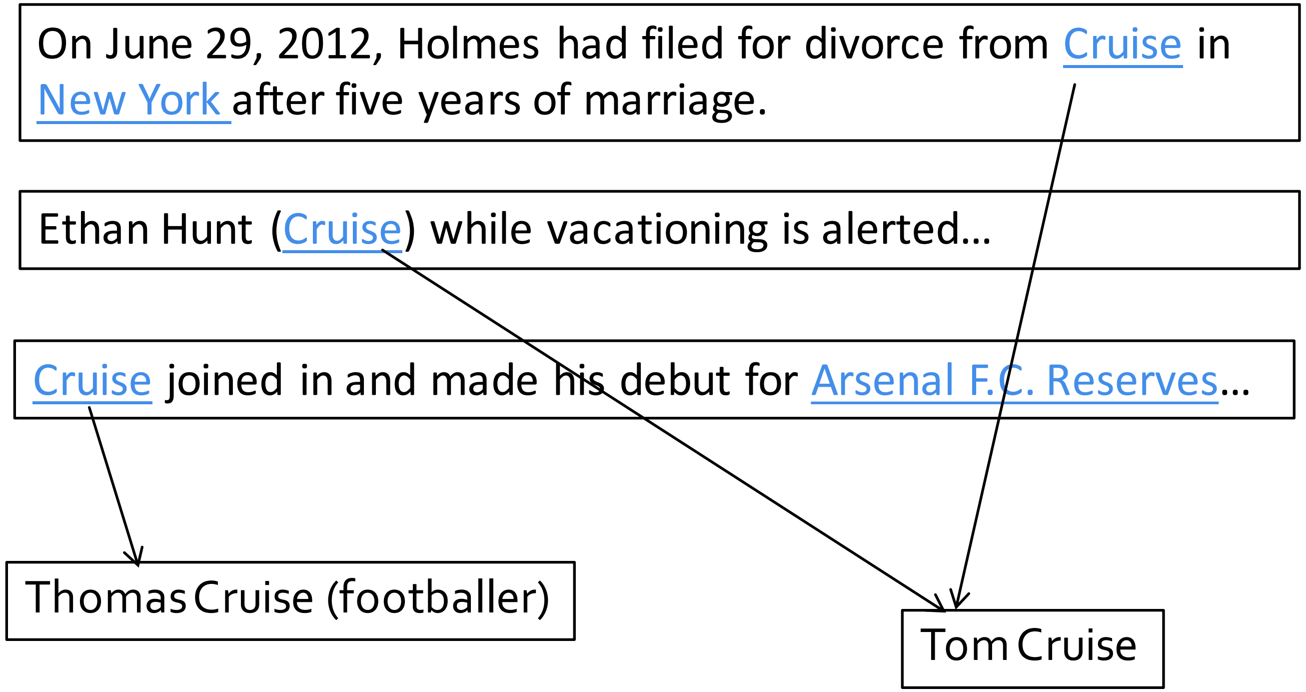

Example by Tsai & Roth'16

- **Challenges:**
  - Link to the English Wikipedia
  - Comparing non-English words to English Wikipedia titles

# EL Talk Outline

- Problem Formulation
  - Fast Search
- Word Embeddings
- Modeling Contexts
- Cross-Lingual Entity Linking
  - Model
  - Feature Abstraction layer
- Experiments

# Fast Search (English)

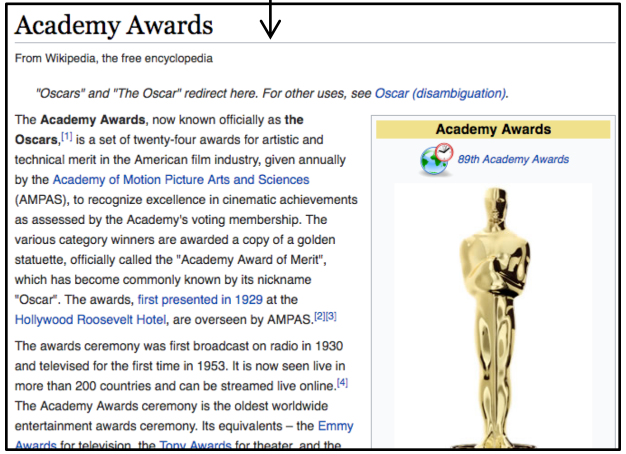



Cruise:

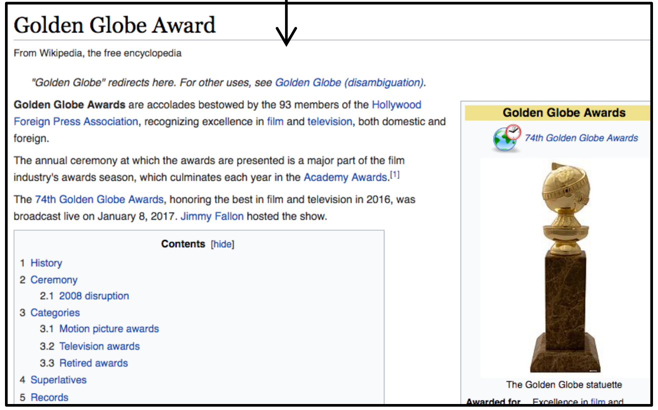
- en/Tom_Cruise (probability: 0.66)
- en/Thomas_Cruise_(footballer) (probability: 0.33)

# Fast Search (Cross-Lingual)

..a los Premios Óscar y en cuatro a los Premios Globo de Oro, su significativa presencia..


## Interlanguage Links


Premios Oscar: en/Academy_Awards (probability: 1.0)

Premios Globo de Oro: en/Golden_Globe_Awards (probability: 1.0)

# Outline

- Problem Formulation
  - Fast Search
- Word Embeddings
- Modeling Contexts
- Cross-Lingual Entity Linking
  - Model
  - Feature Abstraction layer
- Experiments

# Word Embeddings

- Mono-lingual (English)
  - CBOW Word2Vec

- Multi-Lingual
  - Canonical Correlation Analysis (CCA) (Faruqui & Dyer, 14; Tsai & Roth, 16):
    - Alignment using Wikipedia title mapping obtained from inter-language links
  - Multi-CCA (Ammar et.al, 16)
    - Project pre-trained monolingual embeddings in each language (except English) to the vector space of pre-trained English word embeddings
  - Weighted Least Squares (LS) (Mikolov et.al, 13)

# Outline

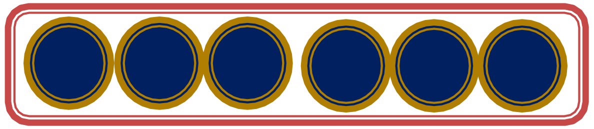
- Problem Formulation
  - Fast Search
- Word Embeddings
- Modeling Contexts
- Cross-Lingual Entity Linking
  - Model
  - Feature Abstraction layer
- Experiments

# Modeling Context from the query Document

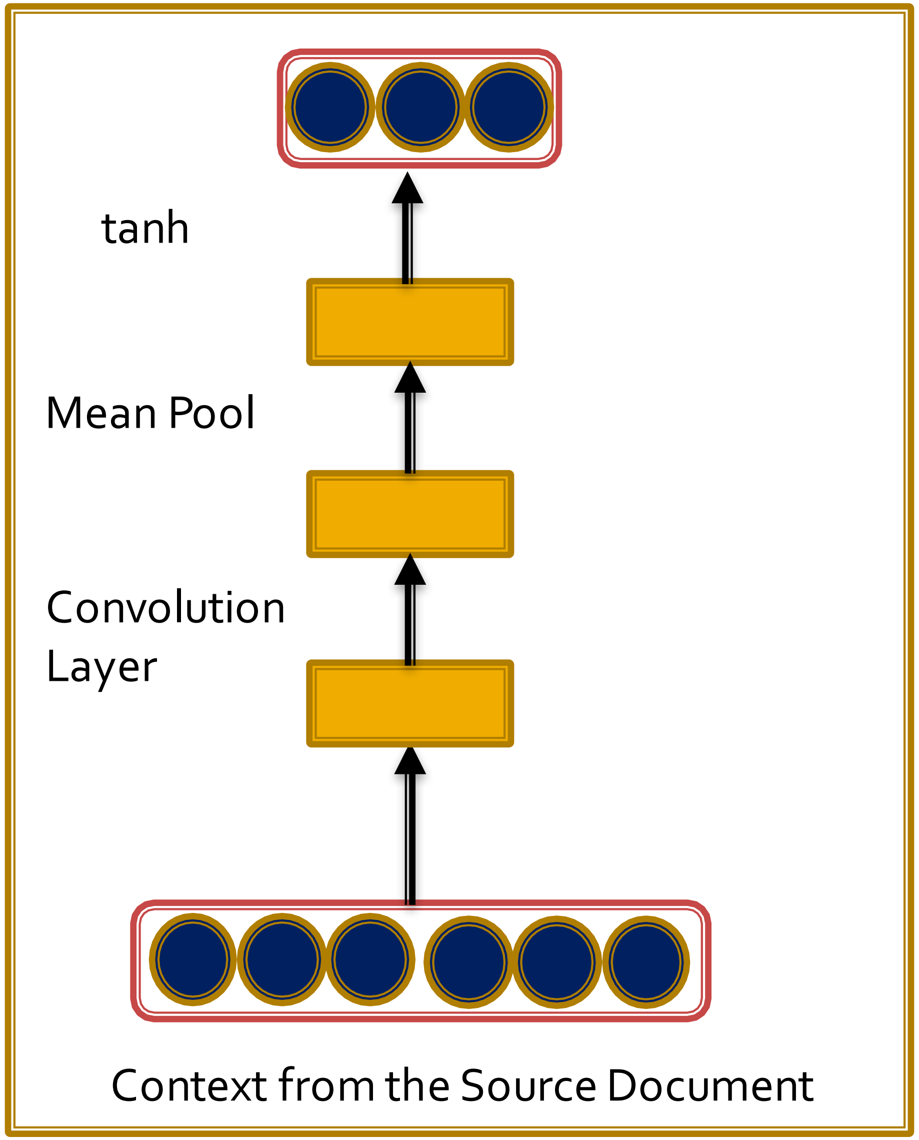
- Get all sentences from the entity coref chain

" [Broad] catapulted [England] to a 74-run win over [Australia]...
[Broad] sent captain [Michael Clarke]'s off stump cart-wheeling before [Steve Smith]..
[Broad] and [Bresnan] found their stride in the evening session.."

- Concatenate them together
  - Get a variable length representation



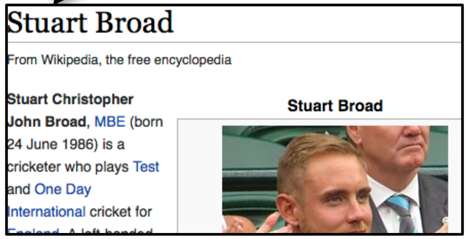
# Modeling Context from the query Document



# Modeling Context from target Wikipedia page

- Get all possible links of the mention from the KB

" [Broad] catapulted [England] to a 74-run win over [Australia]…

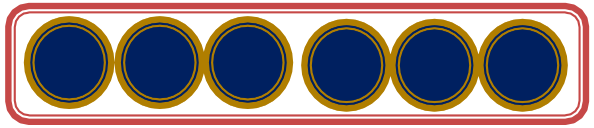



# Modeling Context from target Wikipedia page

- Extract the first paragraph of the current link/page

> **Stuart Christopher John Broad**, MBE (born 24 June 1986) is a cricketer who plays Test and One Day International cricket for England. A left-handed batsman and right-arm seam bowler, Broad's professional career started at Leicestershire, the team attached to his school, Oakham School; in 2008 he transferred to Nottinghamshire, the county of his birth and the team for which his father played. In August 2006 he was voted the Cricket Writers' Club Young Cricketer of the Year.

- Run CNNs on them



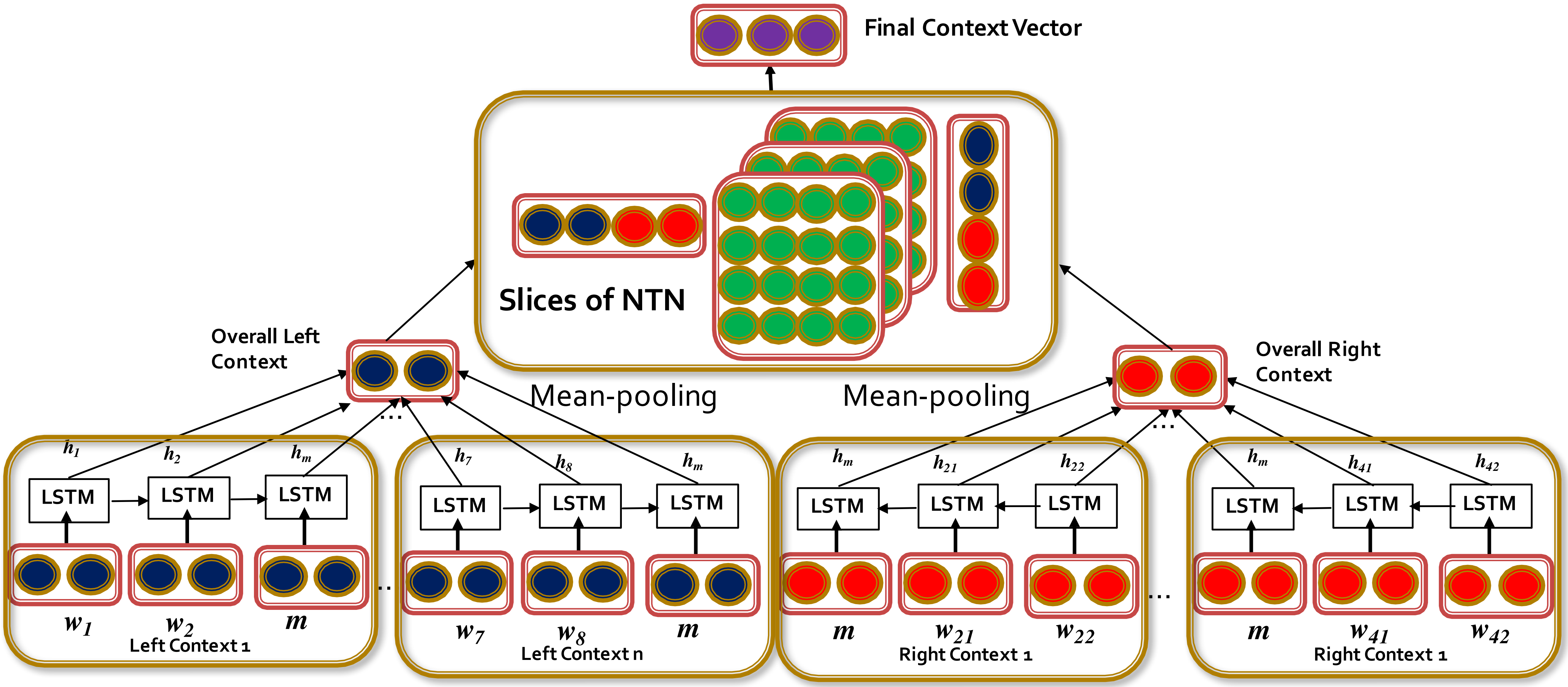
# Wikipedia Link Embeddings

- **Objective**: Model the **whole** Wikipedia page for an entity

- We compute the embeddings $\boldsymbol{e_p}$ of the page $\boldsymbol{p}$:

$$e_p = \frac{\sum_{w \in p} e_w idf_w}{\sum_{w \in p} idf_w}$$

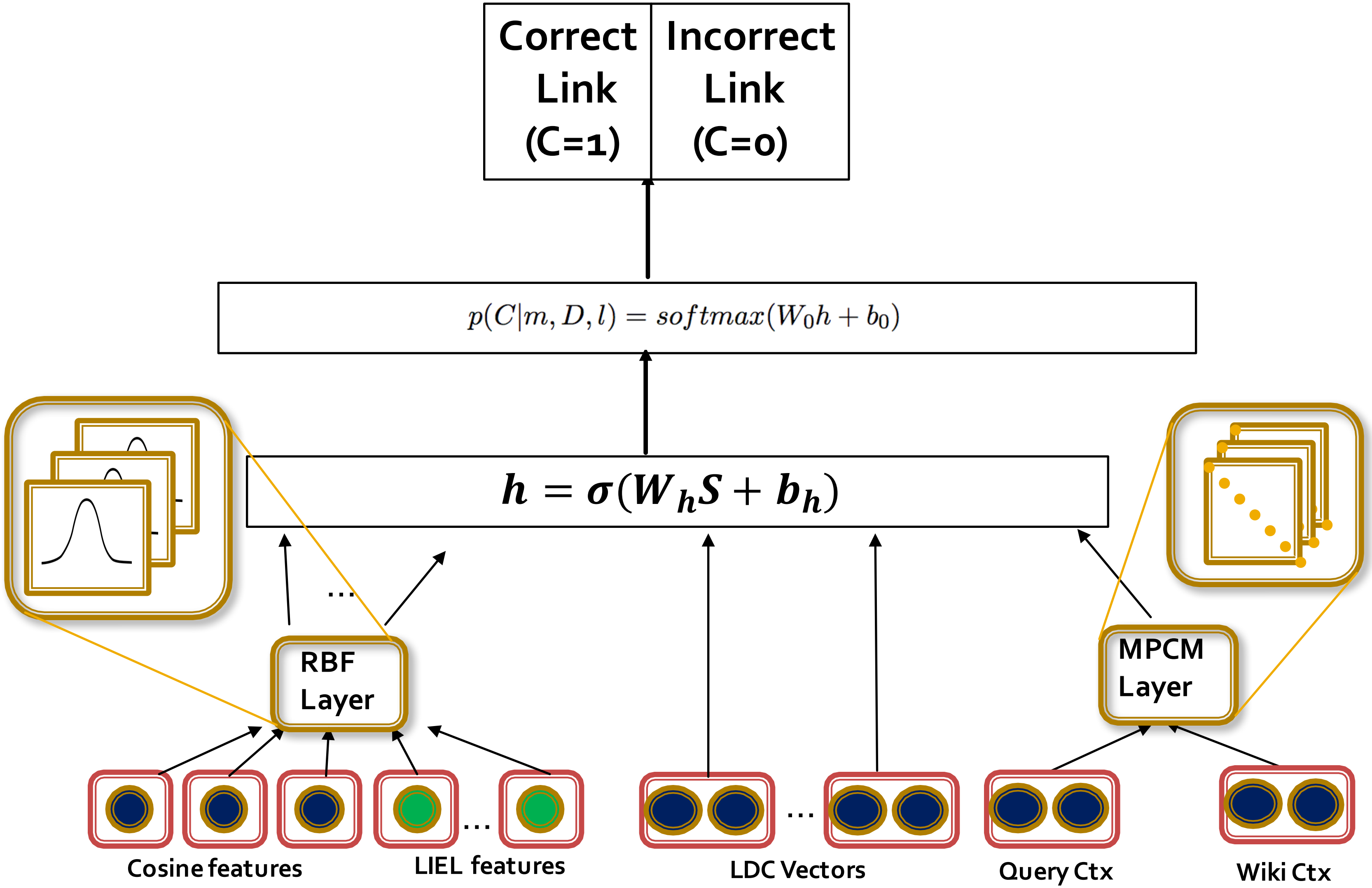
# Fine Grained Context Modeling



# Outline

- Problem Formulation
  - Fast Search
- Word Embeddings
- Modeling Contexts
- Cross-Lingual Entity Linking
  - Model
  - Feature Abstraction layer
- Experiments

# Neural Model Architecture

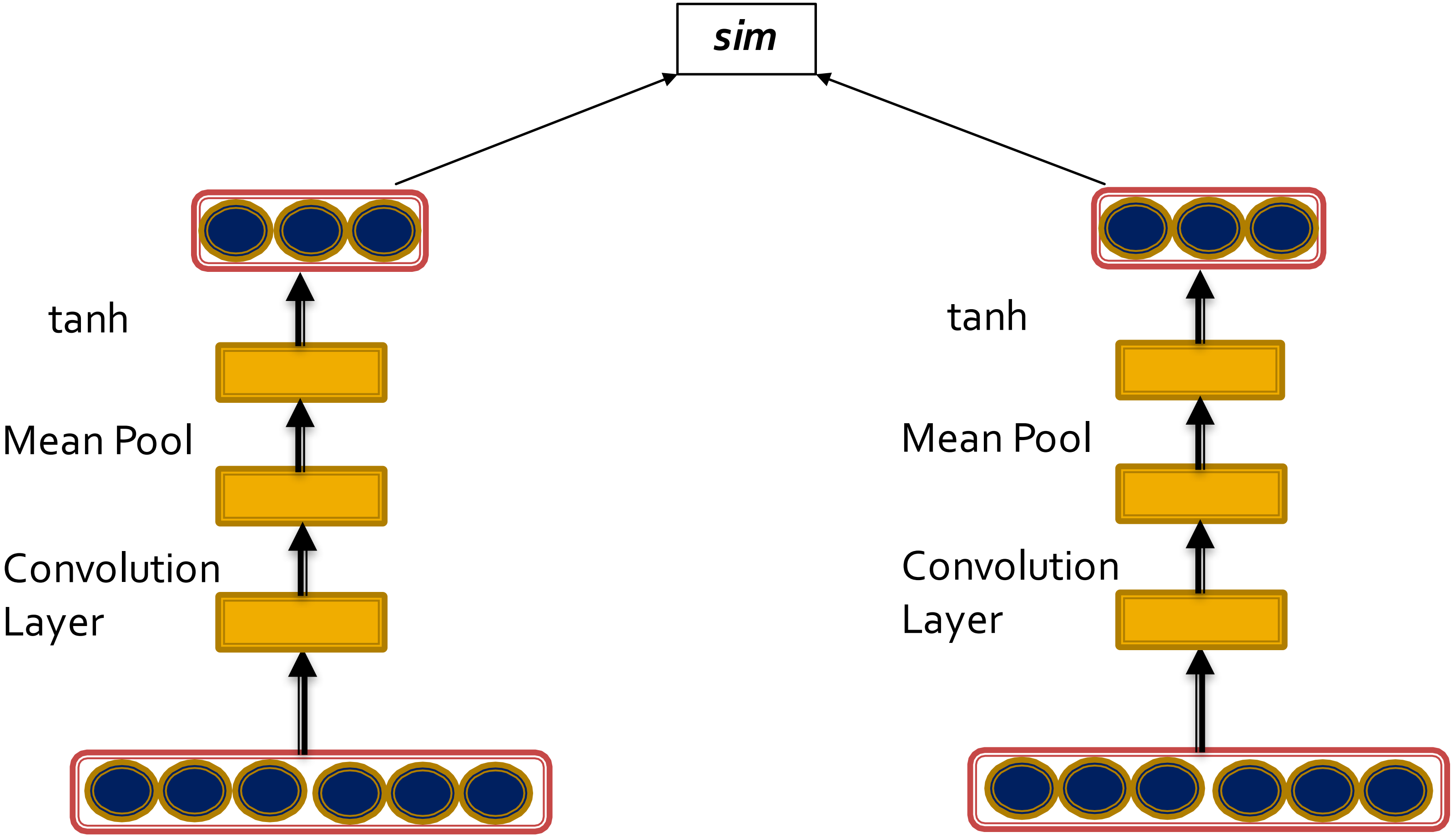


# Feature Abstraction Layer

- **Similarity Features by comparing Context Representations**
  - "Sentence context - Wiki Link" Similarity
  - **"Sentence context - Wiki First Paragraph" Similarity**
  - "Fine-grained context - Wiki Link" Similarity
  - Within-language Features (LIEL, Sil & Florian, ACL16)

- **Semantic Similarities and Dis-similarities**
  - **Lexical Decomposition and Composition (LDC)** (Wang et.al.,16a)
  - Multi-perspective Context Matching (MPCM) (Wang et.al.,16b)

# Measure the Cosine Similarity

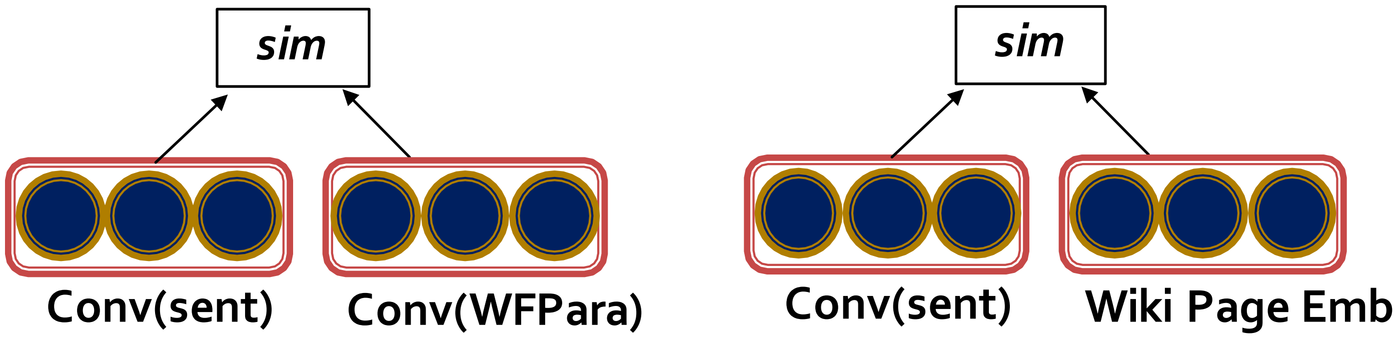



Context from the query Document

Context from the target Wiki Page

# Similarities over multiple granularities of context

- Cosine Similarity based features:




- These values are mapped to a 100-D vector using an **RBF** node
  - Smooth binning process
  - More parameters than a single cosine value

# Outline

- Problem Formulation
  - Fast Search
- Word Embeddings
- Modeling Contexts
- Cross-Lingual Entity Linking
  - Model
  - Feature Abstraction layer
- Experiments

# Experiments

- Datasets:
  - English:
    - CoNLL 2003
    - TAC 2010
  - Cross-lingual (Spanish & Chinese):
    - TAC 2015

# English Experiments

| Systems | In-KB acc. % |
| --- | --- |
| Hoffart *et al.* (2011) | 82.5 |
| He *et al.* (2013) | 85.6 |
| Francis-Landau *et al.* (2016) | 85.5 |
| Sil & Florian (2016) | 86.2 |
| Lazic *et al.* (2015) | 86.4 |
| Chisholm & Hachey (2015) | 88.7 |
| Ganea *et al.* (2015) | 87.6 |
| Pershina *et al.* (2015) | 91.8 |
| Globerson *et al.* (2016) | 92.7 |
| Yamada *et al.* (2016) | **93.1** |
| This work | 92.1 |
| This work+CtxLSTMs | 93.0 |
| This work+CtxLSTMs+LDC | 93.4 |
| This work+CtxLSTMs+LDC+MPCM | **94.0** |

(a) CoNLL2003

# English Experiments

Google

IBM

| Systems | In-KB acc. % |
| --- | --- |
| TAC Rank 1 | 79.2 |
| TAC Rank 2 | 71.6 |
| Sil & Florian (2016) | 78.6 |
| He *et al.* (2013) | 81.0 |
| Chisholm & Hachey (2015) | 80.7 |
| Yamada *et al.* (2016) | 85.2 |
| Globerson *et al.* (2016) | 87.2 |
| This work | 85.0 |
| This work+CtxLSTMs | 86.3 |
| This work+CtxLSTMs+LDC | 86.9 |
| This work+CtxLSTMs+LDC+MPCM | **87.4** |

(b) TAC2010

# Cross-lingual Experiments

| Systems | Linking Acc % |
|---|---|
| Sil & Florian (2016) / TAC Rank 1 | 80.4 |
| Tsai & Roth (2016) | 80.9 |
| This Work | **81.9** |

Table 4: **Performance comparison on the TAC 2015 Spanish dataset.**

| Systems | Linking Acc % |
|---|---|
| TAC Rank 1 | 83.1 |
| Tsai & Roth (2016) | 83.6 |
| This Work | **84.1** |

Table 5: **Performance comparison on the TAC 2015 Chinese dataset.**

# TAC 2017 Results

| | NERC | NERLC | CEAFmC |
|---|---|---|---|
| Eng | 0.806 | 0.668 | 0.713 |
| Spa | 0.785 | 0.603 | 0.664 |
| Cmn | 0.699 | 0.520 | 0.593 |

Trained once on English!

1. Second in Mention Detection (English)
2. Top score in End-end metric (English)
3. Third in Spanish Mention and EL

# 2017 Pilot Task Results

| Lang | NERC | NERLC |
|---|---|---|
| Kikuyu | 0.803 | 0.797 |
| Swahili | 0.664 | 0.51 |
| Nepali | 0.319 | 0.312 |
| All 10 langs | 0.488 | 0.401 |

1. Models:
    1. Mention: System combo
    2. Coref & EL: Purely NN
2. Second position overall end-to-end metric
3. Transfer of knowledge from English helps

# Conclusion

- Model performs zero-shot learning for x-lingual EL
  - Can be applied to any language if we have multi-lingual embeddings
  - Makes effective use of deep NNs
    - mixing CNNs and LSTMs to produce contextual representation
    - Capture similarities + dis-similarities for the task (AAAI 2018 paper)
- Obtained the top score in the English EL task
  - Competitive performance in the other languages e.g. Spanish

# Thanks!

- Questions?