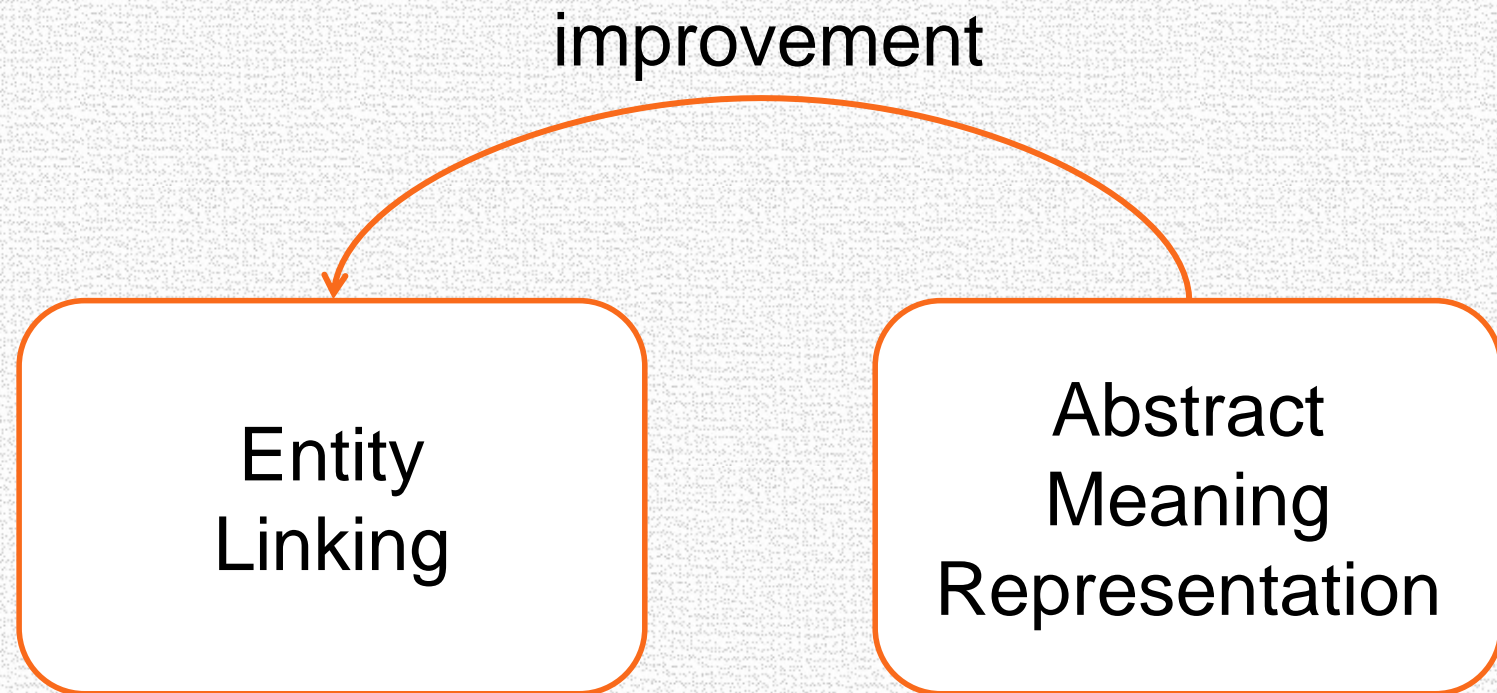


UNSUPERVISED ENTITY LINKING WITH ABSTRACT MEANING REPRESENTATION

Xiaoman Pan, Taylor Cassidy, Ulf
Hermjakob, Heng Ji, Kevin Knight
NAACL-HLT 2015

紹介者:
相澤彰子(国立情報学研究所)
aizawa@nii.ac.jp

要約



Entity Linkingとは

- テキスト中出现する固有表現を、外部知識ベースの項目に対応づける

文書中の
テキスト

Where would *McCain* be without *Sarah*?

mention
という

知識ベースの
entity
ex. Wikipedia



John McCain

Sarah Palin



- 詳しく知りたい人のためのポイント

- <http://ejmeij.github.io/entity-linking-and-retrieval-tutorial/>
- <http://nlp.cs.rpi.edu/kbp/2014/elreading.html>
- <http://www.coling-2014.org/COLING%202014%20Tutorial%20-%20Roberto%20Navigli.pdf>

- Xiao Ling, Sameer Singh, Daniel S. Weld. Design Challenges for Entity Linking. To appear in Transactions of the Association for Computational Linguistics, 2015.

ユニークに
候補が決まる
とは限らない

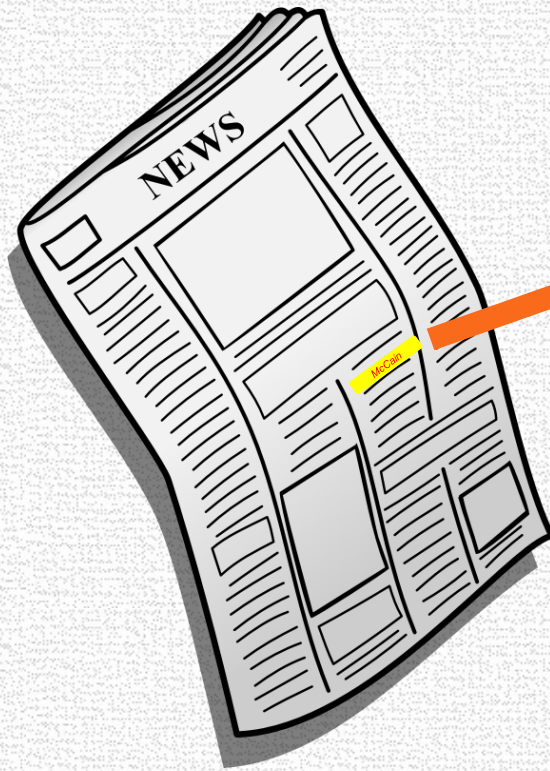
Entity Linking の一般的な処理の流れ

文書中の記述 (mention) m に対して、知識ベース中の対応しそうなエンティティ (entity) の候補を N 個あげる。

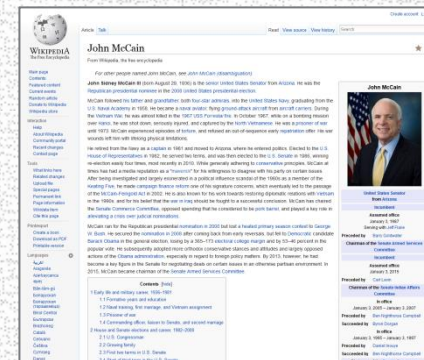


N 個の候補をある基準で並び替えて、スコアが最良のものにリンクをはる。

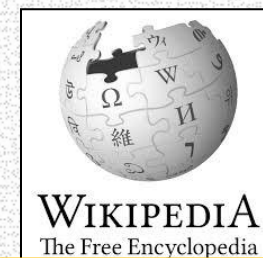
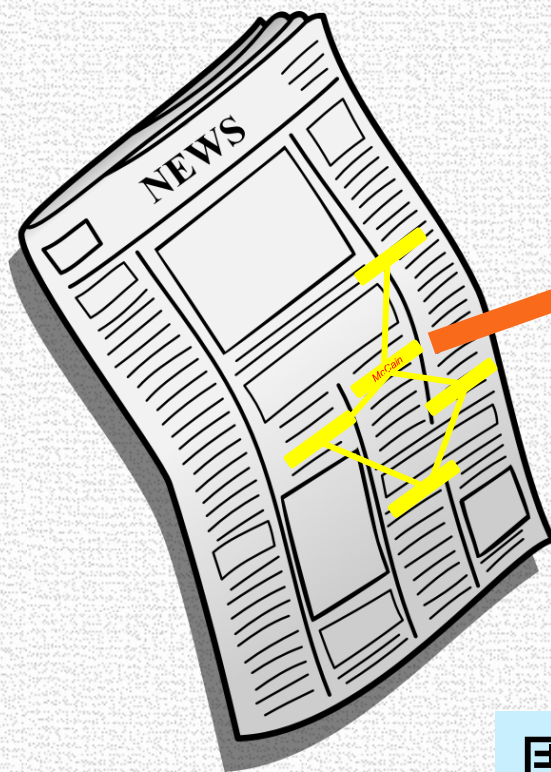
mention と entity の関連度



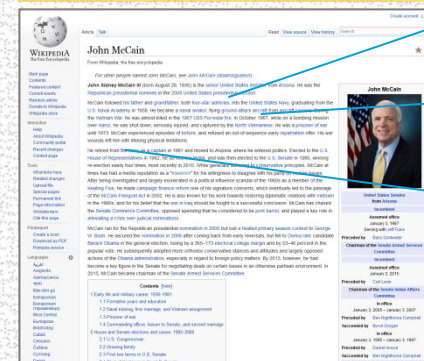
John McCain



mention と entity の関連度



John McCain



周辺情報を手がかりにする

mention と entity の関連度

- 想定されるアプローチ ([Cheng and Roth, 2013], [Pan et. al., 2015], etc.)
 - 局所戦略
 - independent disambiguation
 - context collaborator based re-ranking
 - 与えられた mention に対して最適な entity を決める
 - ‘mention’ を文内の coreferencial chain とする場合もある
 - 大域戦略
 - collective disambiguation
 - context coherence based re-ranking
 - 与えられた mention 集合に対して最適な entity 集合を決める



基本は文脈ベクトルどうしの比較

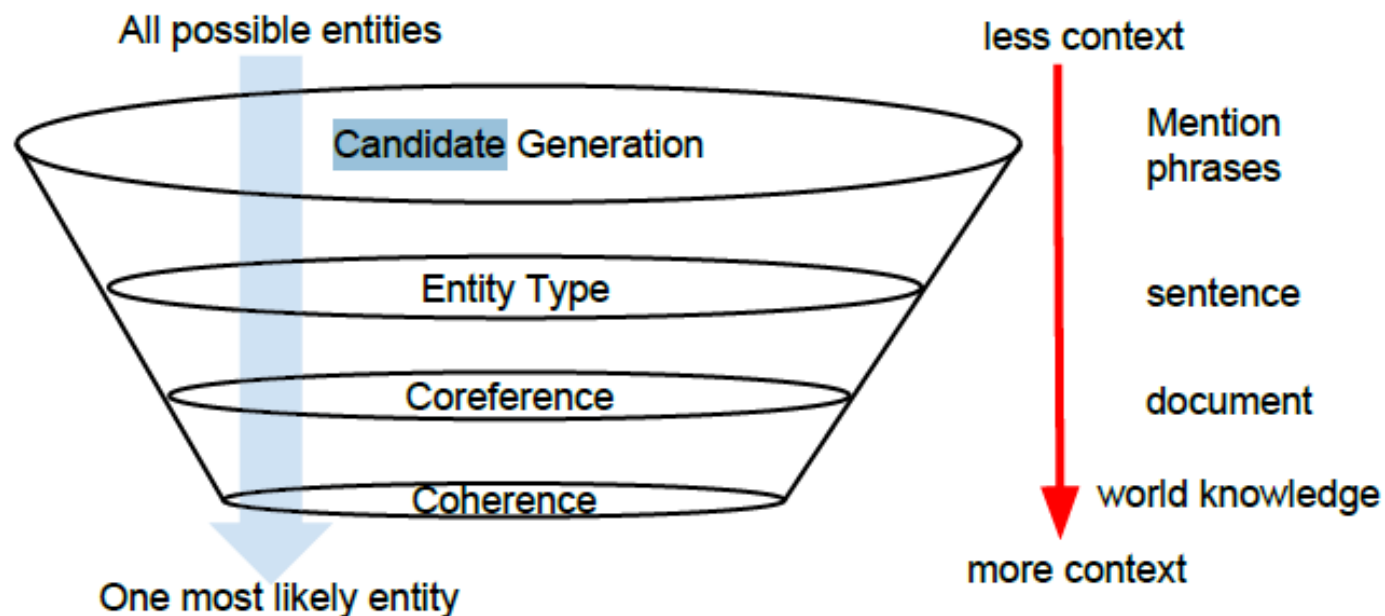
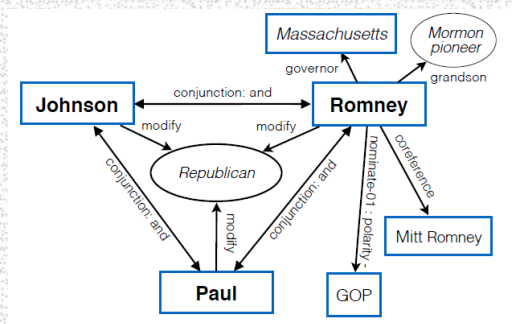


Figure 2: The process of finding the best entity for a mention. All possible entities are sifted through as VINCULUM proceeds at each stage with a widening range of context in consideration.

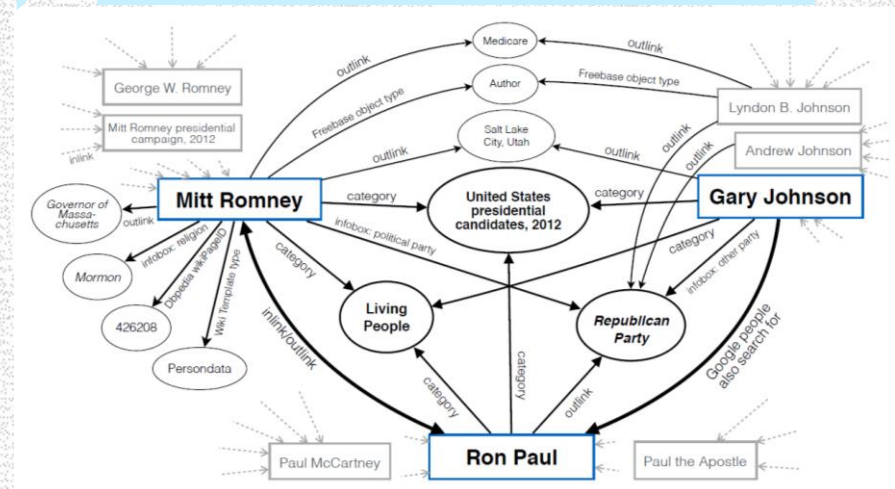
- Xiao Ling, Sameer Singh, Daniel S. Weld. Design Challenges for Entity Linking. To appear in Transactions of the Association for Computational Linguistics, 2015.

mention と entity の関連度(紹介論文)

Knowledge network for coherent mentions



Knowledge network for constructed from KB

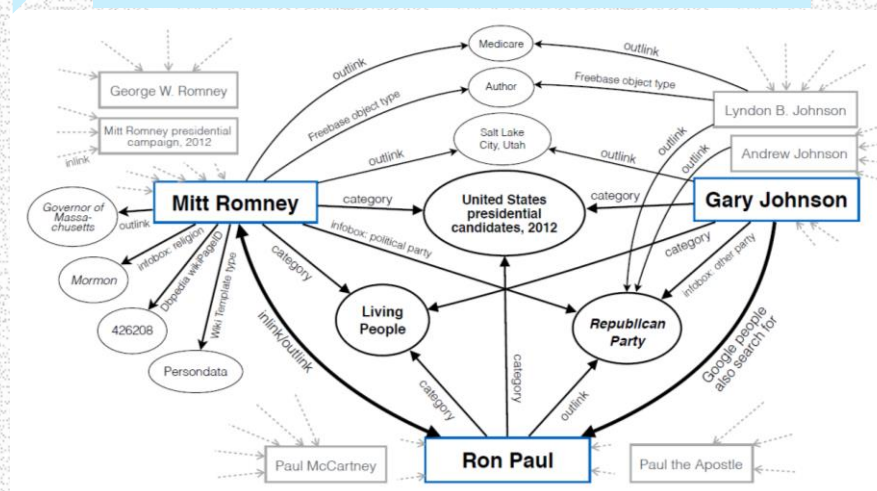
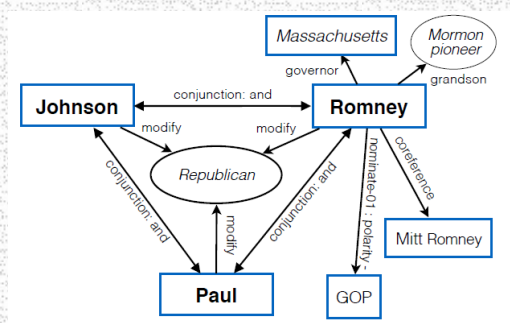


mention と entity の関連度(紹介論文)

Knowledge network
for coherent mentions



Knowledge network
for constructed from KB



- Entity linking のための knowledge network construction
 - Seed entity: person, organization, geo-political entitiesの3種類
 - Knowledge graph は seed entity をルートにしたスター状のグラフで、AMRグラフ上で到達可能な(任意のタイプの)エンティティが葉になっている。

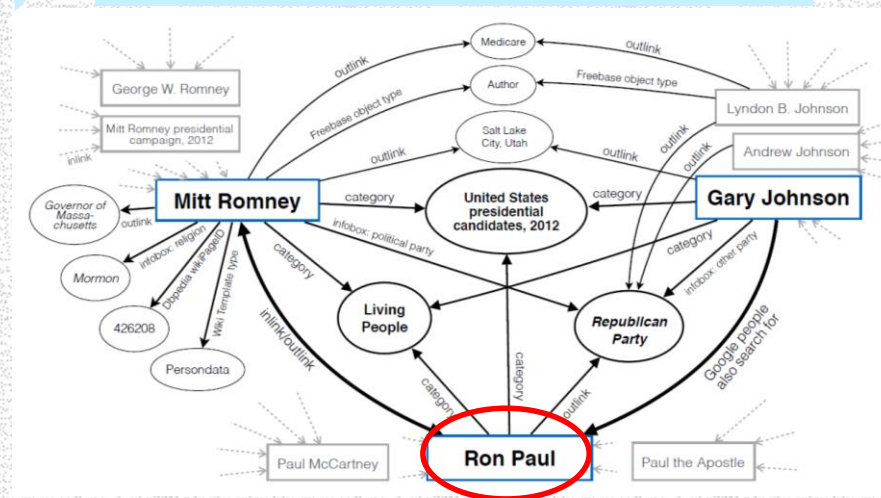
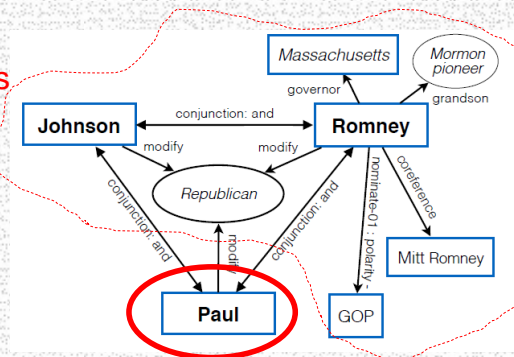
mention と entity の関連度(紹介論文)

Knowledge network
for coherent mentions



Knowledge network
for constructed from KB

Paulの
collaborators



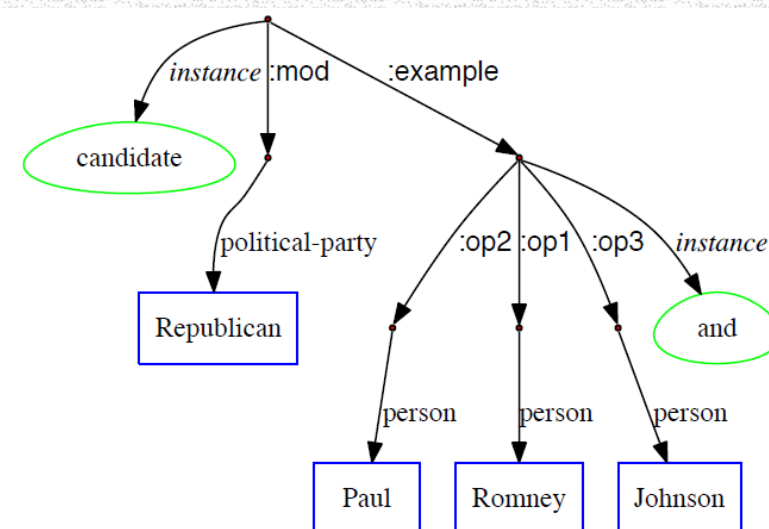
mention と entity どうしの文脈ベクトルの構成方法

⇒ 関連性が低い語が文脈ベクトルにはいってくるとノイズになる

⇒ AMRで構成した knowledge network 上の 'collaborators'を使う

Abstract Meaning Representation (AMR)

- dependency parsing, semantic role labeling と比較して
 - 「ノード」は「語」ではなく「エンティティ」
 - concept と instance を区別
 - concept は単語、PropBankのframeset、数量
 - ex. "**b** / **boy**" for an instance **b** of the concept **boy**
 - 「エッジ」の種類が豊富
 - frame arguments
 - general semantic relations
 - relations for quantities
 - relations for date-entities
 - relations for lists



(d) *Republican candidates like **Romney**, **Paul**, and **Johnson**...*

AMR から knowledge networkの生成(メモ)

- **Entity nodes**

- 言語解析からわかるものについて name expansion (Philippine and Pacific Plates → Philippine Plate)
- AMR の entity についているタイプを利用。8 main types + 100以上の subtypes。人手で AMR typeとDBpedia typeを対応づける。

- **Semantic roles**

- OntoNotes Semantic Layer(若干変更) core roles + non-core semantic relations

- **Background time and location**

- テキスト中に明記されていない場合には、文書の発行時間・場所を利用

- **Coreference**

- Coreferential chain of mentions はグループ化して1つのmentionとしてまとめる。ただしAMRは文内のcoreferenceしかアノテーションしていないので、文をまたがる場合には、部分文字列と略記だけを考慮する。

- **Knowledge network for coherent mentions**

- and, or などの conjunction relations を抽出してエンティティどうしをリンクでつなぐ

KBからの knowledge network の生成(メモ)

<Entity, Edge Label, Node>

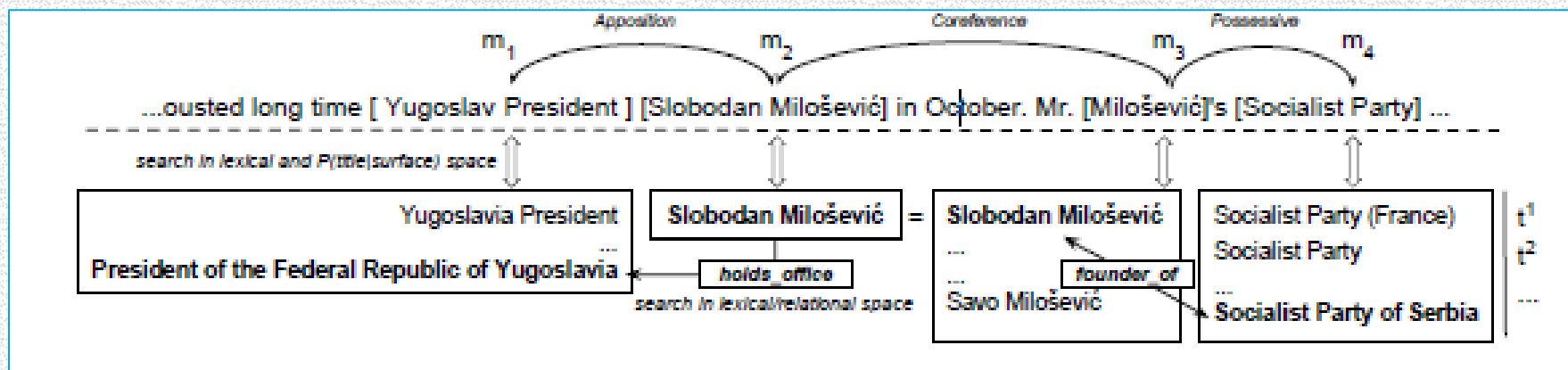
- Wikipedia infobox (+Dbpedia, Freebase)

KBどうしの比較

- **Entity salience**
 - Commonness: wikipedia中のリンクの数(上位N件の数)
- **Context collaborator re-ranking**
 - Mention と entity の collaborators の集合類似度(Jaccard係数)
- **Context coherence re-ranking**
 - coherentなmentionの集合 R_m が与えられたとき、候補となるentityの組み合わせの集合 C_m について、 $R_m, c \in C_m$ から構成される2つのknowledge networkの類似度(=collaborator共通度に基づくJaccard係数)を計算して、類似度がもっとも高い c を選ぶ
- いまの段階で、coherence はconjunction relations に限定されているため、coherentなmentionの集合が得られるのは限られた場合だけ

提案手法は 'unsupervised'

- Supervisedなapproachをとっている従来研究: [Cheng and Roth: Rational Inference for Wikification, EMNLP-2013, Figure 1]



mention の周辺にある semantic relations をとる

⇒そのままでは無関係なものが沢山はいってしまうので重みをつける

上記論文では、ILPを使って collaborative ranking (listwise pairing) を行うが、その目的関数を学習していることに相当

実験による評価

- 人手による EL アノテーションを含む AMR R3 corpus を利用して評価(全部で 1,613 entities)、accuracy

実験結果(表2)

google
search

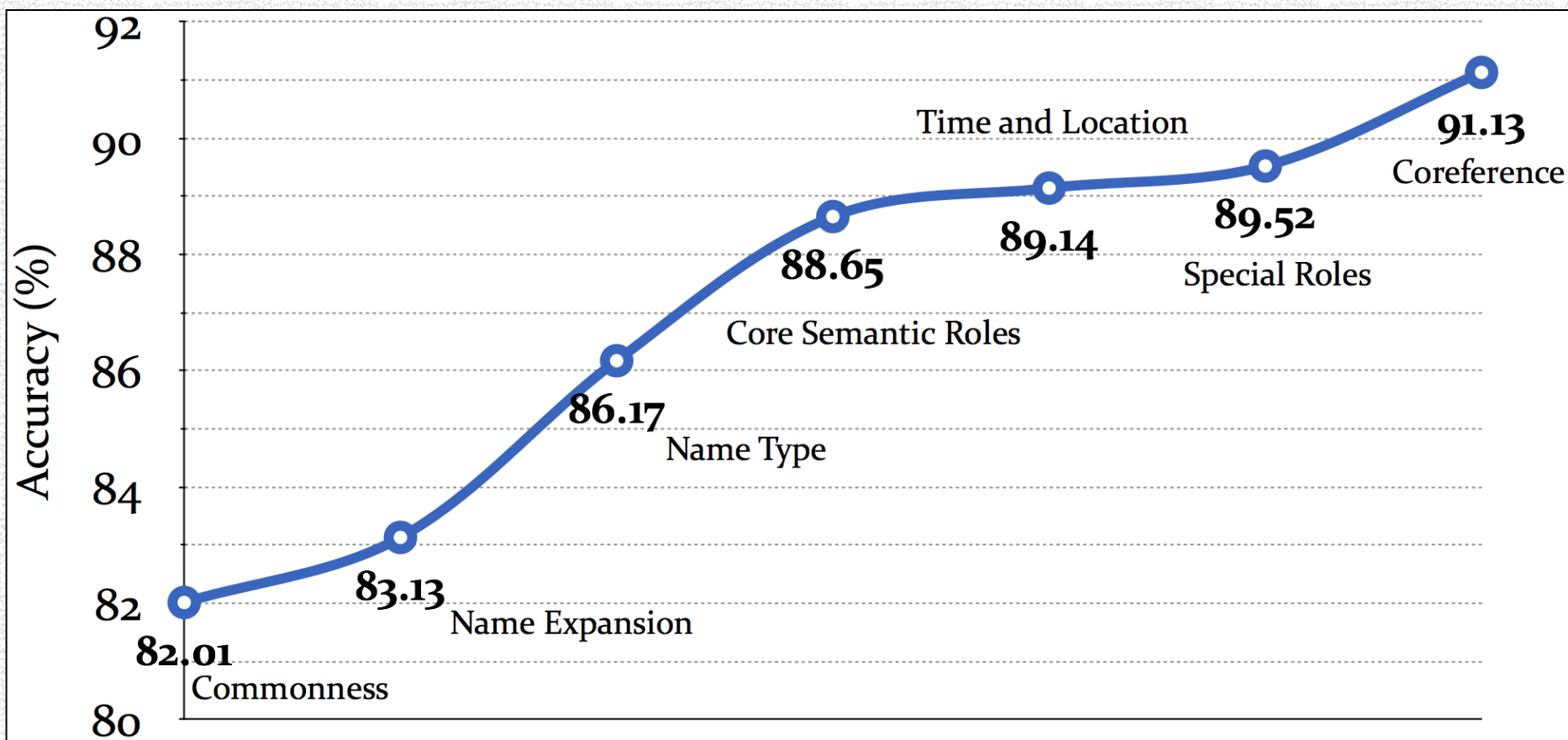
Approach		Definition	News	DF	Total
Popularity	Commonness	based on the popularity measure as described in section 5.2.	89.76	68.99	82.20
	Google Search	use the top Wikipedia page returned by Google search using the	88.10	77.17	84.12
Supervised	State-of-the-art	Human annotated core semantic roles	93.07	87.41	91.01
Unsupervised Context Collaborator Approach	Sen. Level Cooccurrence		80.77	70.31	83.31
	Doc. Level Cooccurrence	document-level co-occurrence based collaborator selection (collaborators limited to human AMR-labeled named entities)	87.51	69.37	80.90
	Human AMR	using human annotated AMR nodes and edges.	93.56	80.88	91.13
	System AMR	using AMR nodes and edges automatically generated by an AMR parser (Flanigan et al., 2014).	90.15	85.69	88.52
	Human SRL	using human annotated core semantic roles defined in PropBank (Palmer et al., 2005) and NomBank (Meyers et al., 2004): ARG0, ARG1, ARG2, ARG4 and ARG5.	93.27	71.21	85.24
Unsupervised Combined Approach	Human AMR	coherence approach used where possible (215 mentions), col-	94.34	88.25	92.12

Human annotated core semantic roles よりも System AMRの方がよい！

state-of-the-art よりよい

coherenceを考慮するとさらに 正解率が上がる

AMR アノテーションの効果 (図4)



エラー分析と結論

- Knowledge gap between the source text and KB
 - 一般的な知識を使って推論する
 - 他の外部知識源を参照する
- AMRには、まだまだいろいろな可能性がある
 - 性能はよかったけれど、志半ばのことはいくつか (relation情報を使っていない、coherenceも予備検討どまり、etc.)