Neural Open Information Extraction

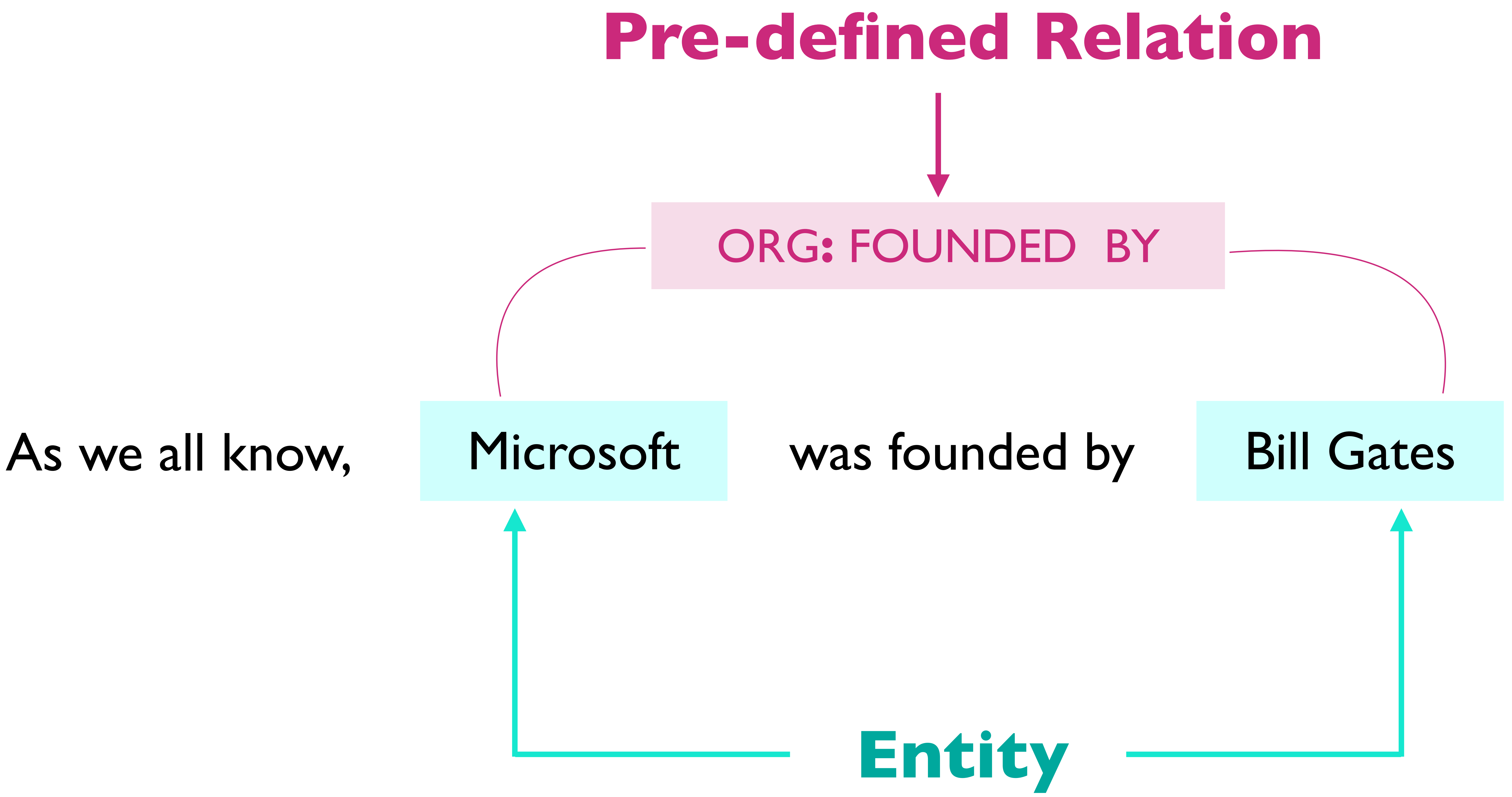
Speaker: 杨晰

xyang41@stu.ecnu.edu.cn

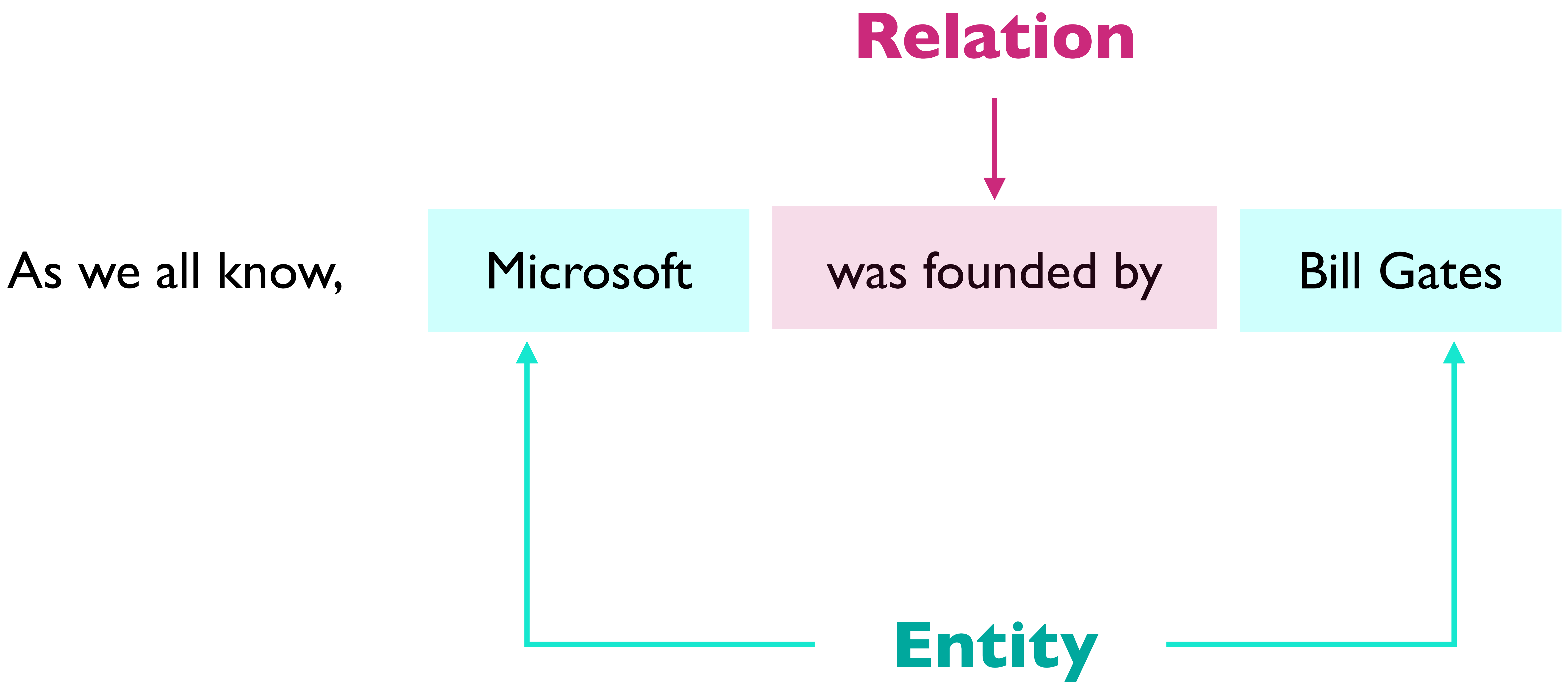
Outline

- Background: Open Information Extraction (OIE)
- Works
  - *Neural Open Information Extraction*
  - *Systematic Comparison of Neural Architectures and Training Approaches for Open Information Extraction*
  - *OpenIE6: Iterative Grid Labeling and Coordination Analysis for Open Information Extraction*
  - *Maximal Clique Based Non-Autoregressive Open Information Extraction*
- Conclusions

BG: Relation Extraction



BG: Open Information Extraction (OIE)



BG: OIE

- Guidelines
  - Assertedness: extracted propositions should be asserted by the original sentence
  - Minimal propositions: each slot as minimal as possible, as long as the original information is preserved
  - Completeness and open lexicon: extract all asserted propositions from a sentence

Stanovsky, Gabriel, and Ido Dagan. "Creating a large benchmark for open information extraction." *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2016.

BG: OIE

- Guidelines

Bell , instead of Jimmy , distributes sweet candies and small toys .

~~Assertedness~~: (Jimmy, distributes, sweet candies)
~~Minimal propositions~~: (Bell, distributes, sweet candies and small toys)
(Bell, distributes, sweet candies)
(Bell, distributes, small toys)

BG: OIE

- Neural vs Hand-crafted Patterns OIE: error propagation
- Neural OIE Methods
  - Generation-based: sequence-to-sequence task
    - *Neural Open Information Extraction* ✅
    - *IMoJIE*
  - Labeling-based
    - *Systematic Comparison of Neural Architectures and Training Approaches for Open Information Extraction* ✅
    - *OpenIE6* ✅
    - *Maximal Clique Based Non-Autoregressive Open Information Extraction* ✅

BG: OIE

- *Multi2OIE*: first predicts all the relation arguments using BERT, then predicts arguments associated with each relation using multi-head attention blocks

- Span-based: sequence-to-sequence task

- *SpanOIE*: uses a predicate module to first choose potential candidate relation spans, and for each relation span, classifies all possible spans of the sentence as subject or object.

[ACL18]Neural Open Information Extraction

Lei Cui, Furu Wei, and Ming Zhou

Microsoft Research Asia

Task Definition

- Open IE is cast as a sequence-to-sequence generation problem

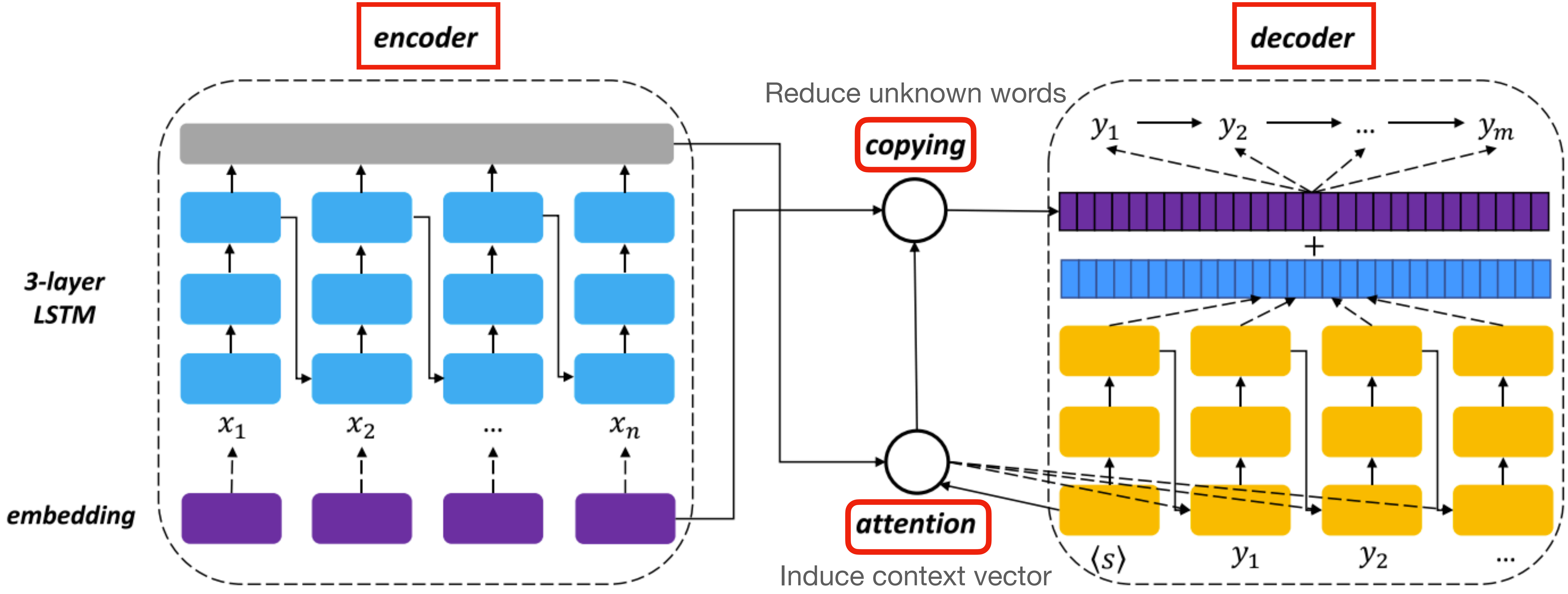
$$
\begin{aligned}
P(Y|X) &= P(Y|x_1, x_2, ..., x_m) \\
&= \prod_{i=1}^{n} p(y_i|y_1, y_2, ..., y_{i-1}; x_1, x_2, ...x_m)
\end{aligned}
$$

Input: *Deep learning is a subfield of machine learning*

↓

<arg1> *Deep learning* ***</arg1>*** ***<rel>*** *is a subfield of* ***</rel>*** ***<arg2>*** *machine learning* ***</arg2>***

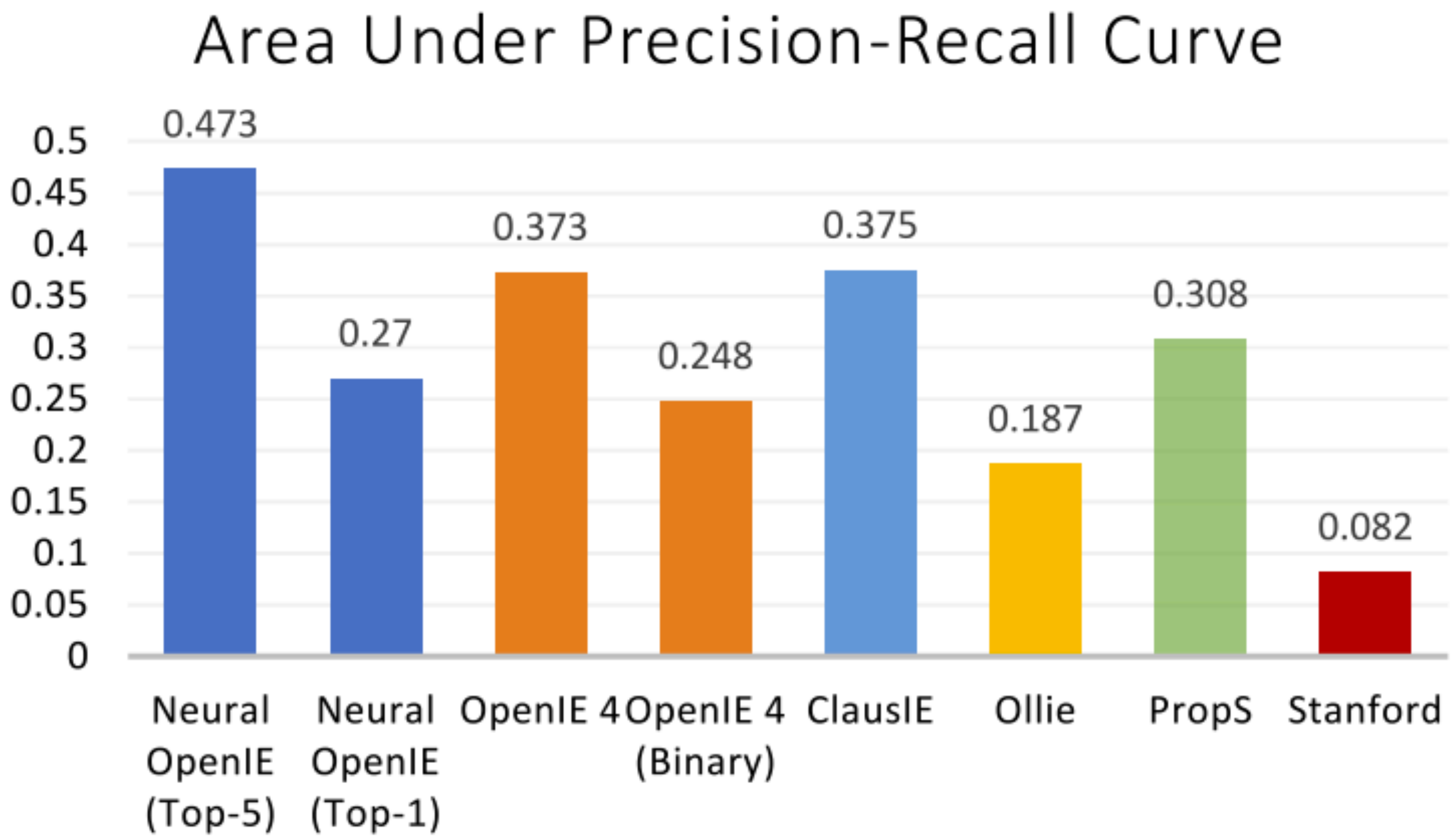
Architecture: Neural Open IE



Experiments

- Training Data: Wikipedia dump 20180101 (OPENIE4 system) [36,247,584 <sentence, tuple>]
- Test Data: OIE16 [3200 sentences, 10359 extractions]

Experiments



System	**Device**	**Time**
Stanford	CPU	234s
Ollie	CPU	160s
ClausIE	CPU	960s
PropS	CPU	432s
OpenIE4	CPU	181s
Neural Open IE	GPU	172s

[EMNLP20]Systematic Comparison of Neural Architectures and Training Approaches for Open Information Extraction

Patrick Hohenecker,[1,2] Frank Mtumbuka,[1] Vid Kocijan,[1] Thomas Lukasiewicz[1,2]

[1] *University of Oxford, Oxford, UK*
[2] *Serein AI, London, UK*

Task Definition

- Open IE is cast as a sequence tagging problem, each word is labelled with a BIO tag.

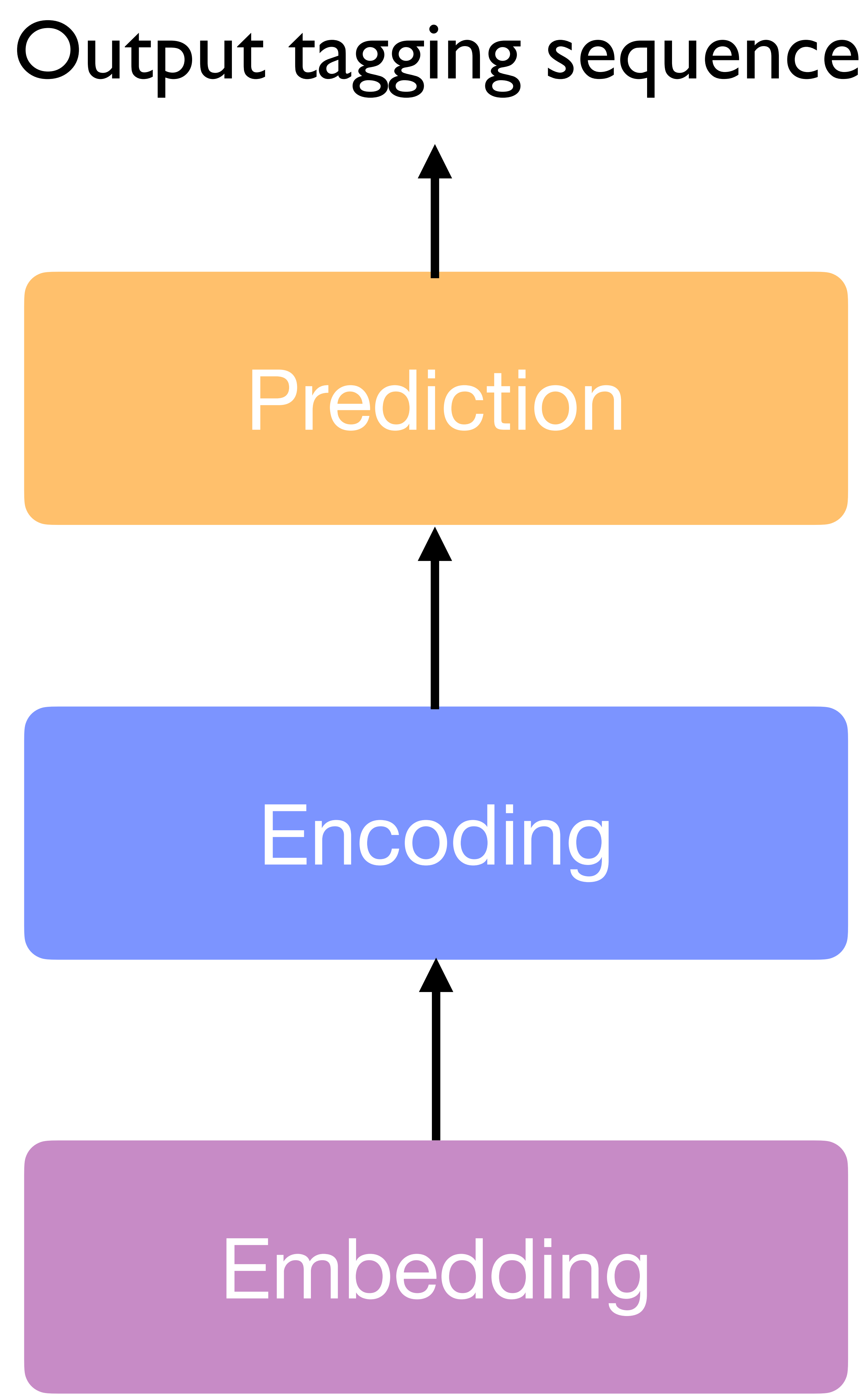
Ludwig	van	Beethoven	was	a	world	-	famous	composer	of	classical	music	.
A0-B	*A0-I*	*A0-I*	*P-B*	*A1-B*	*A1-I*	*A1-I*	*A1-I*	*A1-I*	*O*	*O*	*O*	*O*

$\Rightarrow$ ⟨*Ludwig van Beethoven, was, a world-famous composer*⟩

Beethoven	,	who	died	in	1827	,	composed	the	Ode	to	Joy	.
A0-B	*O*	*O*	*P-B*	*P-I*	*A1-B*	*O*	*O*	*O*	*O*	*O*	*O*	
A0-B	*O*	*O*	*O*	*O*	*O*	*O*	*P-B*	*O*	*A1-B*	*A1-I*	*A1-I*	*O*

$\Rightarrow$ ⟨*Beethoven, died in, 1827*⟩
⟨*Beethoven, composed, Ode to Joy*⟩

OIE Architecture



Training Loss

- Loss function: the negative log-likelihood (NLL)
  - Outside-tag appear more frequently than other tags.

(a)	*O*	*O*	*A0-B*	*A0-I*	*A0-I*	*O*	*O*	*O*	*O*	*P-B*	*P-I*	*O*	*A1-B*	*A1-I*	*O*
(b)	*O*	*O*	*A0-B*	*A0-I*	*A0-I*	*O*	*O*	*O*	*O*	*P-B*	*P-I*	*O*	*A1-B*	*A1-I*	*O*
(c)	*O*	*O*	*A0-B*	*A0-I*	*A0-I*	*O*	*O*	*O*	*O*	*P-B*	*P-I*	*O*	*A1-B*	*A1-I*	*O*

Figure 2: (a) Excluding all *O* tags when computing the loss. (b) Considering only the tags in the transitions between different tags (c) Considering only the tags in the transitions between different tags that are not *O*.

Experiments

- Training Data: Wikipedia dump 20180101 (OPENIE4 system) [36,247,584 <sentence, tuple>] + OIE16 training data
- Test Data: OIE16 test data

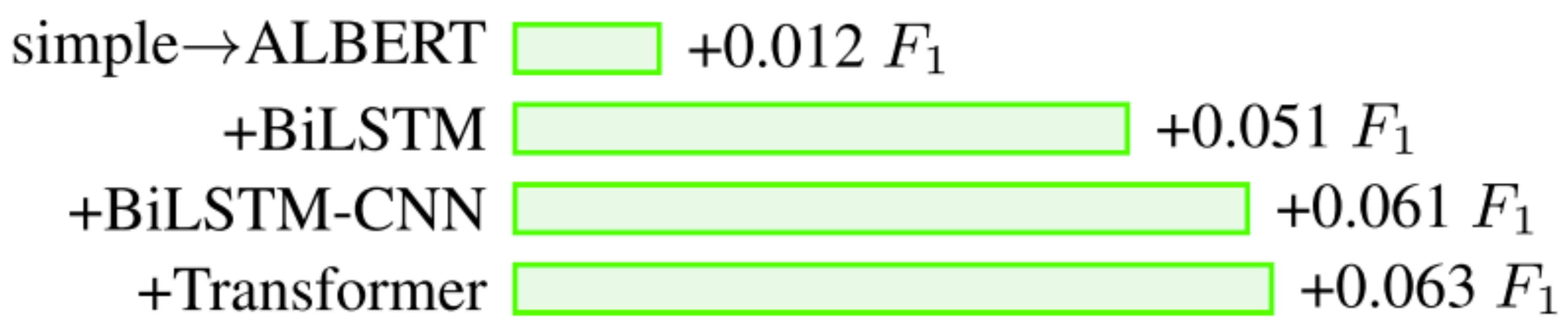
Experiments

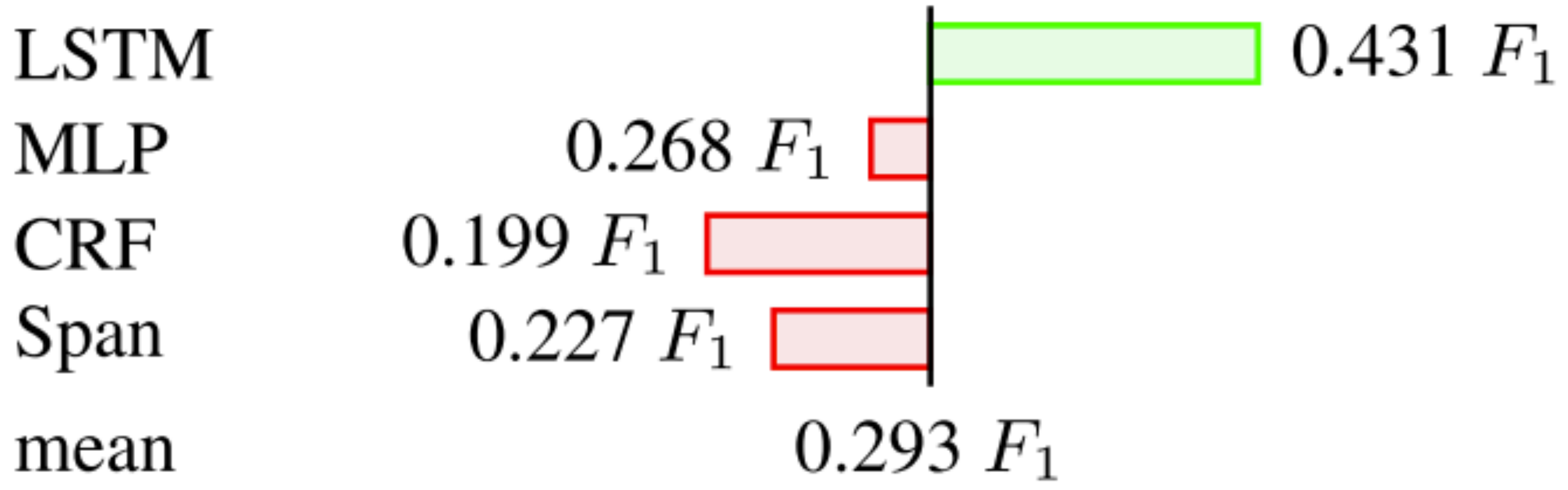
encoding block	prediction block	LSTM (a)	LSTM (b)	LSTM (c)	LSTM (d)	MLP (a)	MLP (b)	MLP (c)	MLP (d)	CRF (a)	Span (a)
simple embedding block											
none	F_1	0.325	0.402	0.320	0.411	0.125	0.224	0.225	0.325	0.199	0.223
	AUC-PR	0.184	0.172	0.181	0.178	0.113	0.112	0.110	0.114	**0.136**	0.177
BiLSTM	F_1	0.211	0.535	0.346	0.598	0.123	0.276	0.213	0.460	0.181	0.221
	AUC-PR	0.140	0.524	0.277	0.602	0.055	0.219	0.213	0.435	0.119	**0.195**
BiLSTM-CNN	F_1	0.225	0.486	0.588	0.589	0.113	0.250	0.241	0.452	0.201	0.230
	AUC-PR	0.112	0.474	0.574	**0.610**	0.065	0.198	0.196	0.420	0.124	0.189
Transformer	F_1	0.332	0.539	0.351	**0.601**	0.126	0.281	0.260	**0.471**	**0.205**	**0.242**
	AUC-PR	0.132	0.534	0.321	0.597	0.070	0.183	0.206	**0.439**	0.117	0.178
ALBERT embedding block											
none	F_1	0.329	0.406	0.323	0.415	0.126	0.226	0.225	0.326	0.203	0.256
	AUC-PR	0.195	0.175	0.193	0.187	0.107	0.106	0.105	0.107	0.145	0.187
BiLSTM	F_1	0.333	0.541	0.349	0.623	0.161	0.281	0.263	0.463	0.185	0.253
	AUC-PR	0.250	0.504	0.286	0.610	0.120	0.220	0.201	0.435	0.107	**0.205**
BiLSTM-CNN	F_1	0.186	0.463	0.596	0.610	0.123	0.276	0.258	0.468	0.203	0.253
	AUC-PR	0.081	0.397	0.582	0.614	0.054	0.219	0.211	0.431	0.131	0.193
Transformer	F_1	0.351	0.555	0.362	**0.628**	0.117	0.292	0.274	**0.476**	**0.217**	**0.273**
	AUC-PR	0.242	0.515	0.278	**0.644**	0.044	0.204	0.218	**0.436**	**0.149**	0.198

References: Angiras (2018) Stanovsky et al. (2018) Jia and Xiang (2019) Zhan and Zhao (2019)

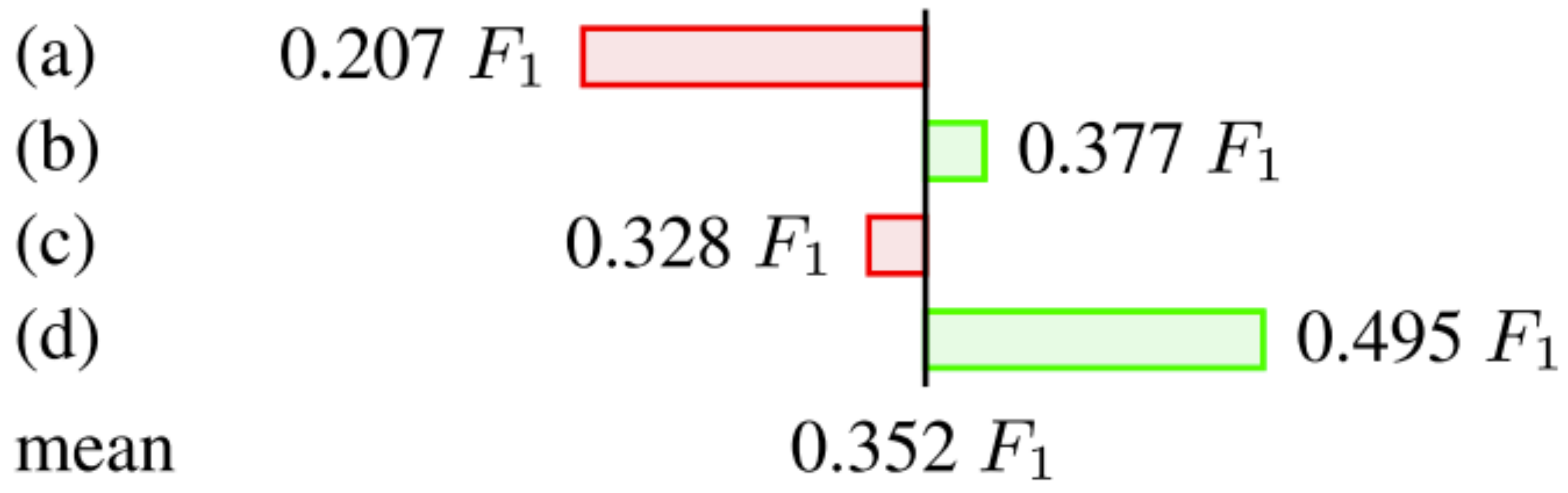
Table 1: The results of our experimental evaluation, where the different columns correspond with the different training schemes: (a) standard NLL (i.e., considering all labels), (b) disregarding O-tags, (c) optimizing transitions only, and (d) considering start and end of a triple's elements only. The best results are underlined for each of the encoding blocks, and printed boldface for each prediction block.

Ablation Study





- Embedding blocks, Encoder blocks
- Prediction blocks



O	*O*	*A0-B*	*A0-I*	*A0-I*	*O*	*O*	*O*	*O*	*P-B*	*P-I*	*O*	*A1-B*	*A1-I*	*O*
O	*O*	*A0-B*	*A0-I*	*A0-I*	*O*	*O*	*O*	*O*	*P-B*	*P-I*	*O*	*A1-B*	*A1-I*	*O*
O	*O*	*A0-B*	*A0-I*	*A0-I*	*O*	*O*	*O*	*O*	*P-B*	*P-I*	*O*	*A1-B*	*A1-I*	*O*

- Training schemes

[EMNLP20]OpenIE6: Iterative Grid Labeling and Coordination Analysis for Open Information Extraction

Keshav Kolluru,[1] Vaibhav Adlakha,[1] Samarth Aggarwal,[1] Mausam,[1] and Soumen Chakrabarti [2]

[1] Indian Institute of Technology Delhi
[2] Indian Institute of Technology Bombay

Contributions

- is based on our novel IGL architecture,
- is trained with constraints to improve recall,
- handles conjunctive sentences with our new state-of-art coordination analyzer, which is 12.3 pts better in F1, and
- is $10\times$ faster compared to current state of the art and improves F1 score by as much as 4.0 pts.

Motivations

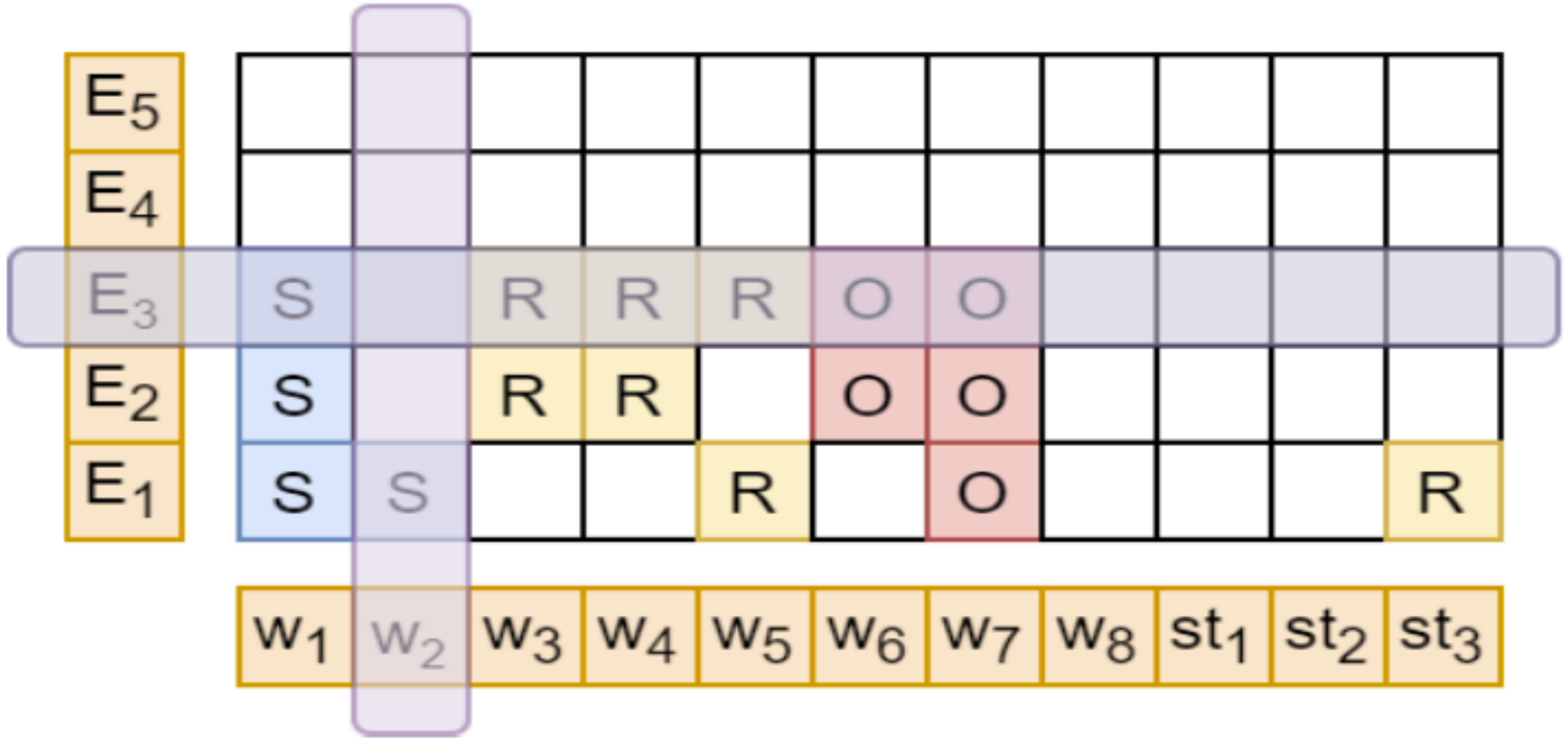
- Generation systems (IMoJIE): Re-encode ➡️ Capture dependencies among extractions / Slow
- Labeling systems (RnnOIE, SenseOIE): Labeling ➡️ Ignore dependencies among extractions / Fast / Less accurate

↓

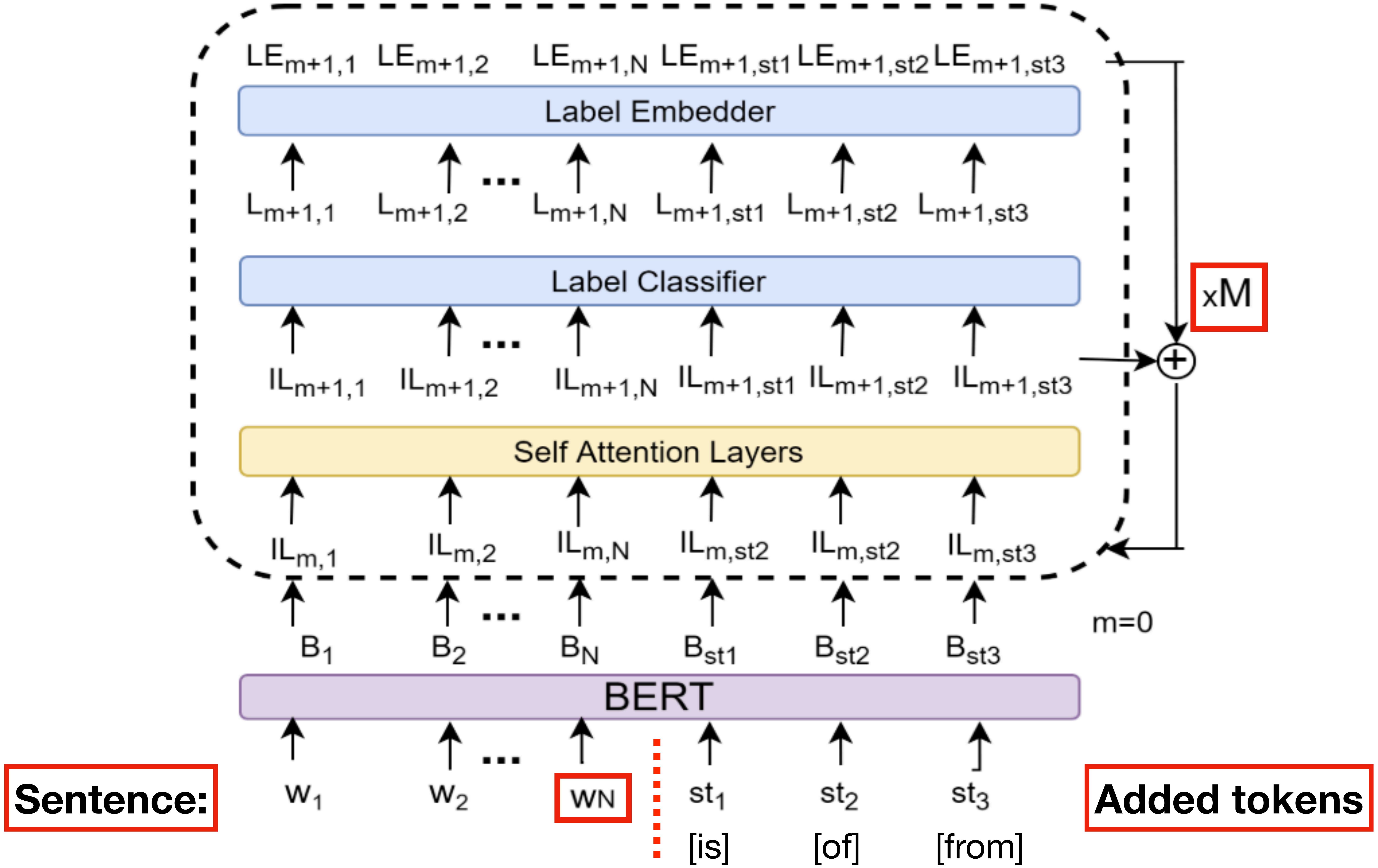
Trade-off: **Iterative Grid Labeling (IGL)**

Task Definition

- Open IE is cast as a 2D grid labeling problem of size (M, N).
  - M = 5: pre-defined maximum number of extractions
  - N: sentence length
  - Each extraction corresponds to one row in the grid



Architecture: Iterative Grid Labeling (IGL-OIE)



Constrained Training: CIGL-OIE

- IGL: High precision, low recall ➡️ Inject prior knowledge as soft constrains
  - POS Coverage (POSC): words with nouns, verbs, adjectives, or adverbs POS tags should be covered. spaCy for POS
  - Head Verb Coverage (HVC): Head verb should be present in the relation span of some (not too many) extractions
  - Head Verb Exclusivity (HVE): The relation span of one extraction can contain at most one head verb
  - Extraction Count (EC): Num(extractions with head verbs in the relation span) > Num(head verbs)

Coordination Boundary Detection: IGL-CA

- Coordination Analysis is cast as a hierarchical labeling / 2D grid labeling problem of size (M, N).
  - M = 3: pre-defined maximum depth of hierarchy
  - N: sentence length
  - Labels: CC (coordinated conjunction), CONJ (belonging to a conjunct span), or N (None)
  - IGL-CA: generate simple (non-conjunctive) sentences

Architecture: OpenIE6

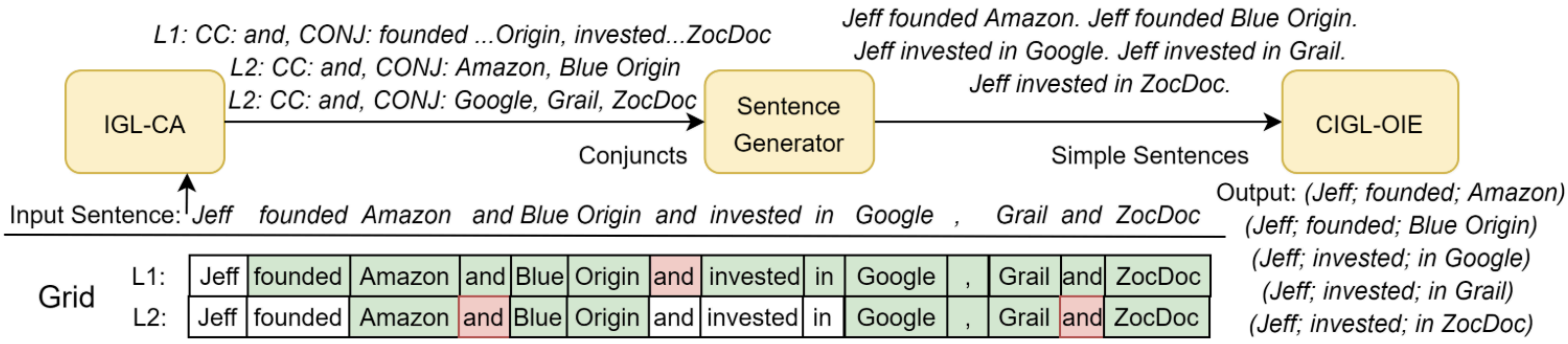


Figure 4: The final OpenIE system. IGL-CA identifies conjunct boundaries by labeling a 2-D grid. This generates simple sentences and CIGL-OIE emits the final extractions.

Experiments: Datasets

- Training: OpenIE4 training dataset(following IMoJIE), 92,774 Wiki sentences, 190,661 extractions
- Evaluation: CaRB reference set(re-annotated OIE16 corpus via crowd-sourcing)
- Metrics: OIE16, Wire57, CaRB, CaRB(1-1)

Experiments: Evaluation

- Dataset: CaRB (re-annotated OIE16 corpus via crowd-sourcing)

System	CaRB		CaRB(1-1)		OIE16-C		Wire57-C	Speed
	F1	AUC	F1	AUC	F1	AUC	F1	Sentences/sec.
MinIE	41.9	-	38.4	-	52.3	-	28.5	8.9
ClausIE	45.0	22.0	40.2	17.7	61.0	38.0	33.2	4.0
OpenIE4	51.6	29.5	40.5	20.1	54.3	37.1	34.4	20.1
OpenIE5	48.0	25.0	42.7	20.6	59.9	39.9	35.4	3.1
SenseOIE	28.2	-	23.9	-	31.1	-	10.7	-
SpanOIE	48.5	-	37.9	-	54.0	-	31.9	19.4
RnnOIE	49.0	26.0	39.5	18.3	56.0	32.0	26.4	**149.2**
(Cui et al., 2018)	51.6	32.8	38.7	19.8	53.5	37.0	33.3	11.5
IMoJIE	53.5	33.3	41.4	22.2	56.8	39.6	36.0	2.6
IGL-OIE	52.4	33.7	41.1	22.9	55.0	36.0	34.9	142.0
CIGL-OIE	**54.0**	**35.7**	42.8	24.6	59.2	40.0	36.8	142.0
CIGL-OIE + IGL-CA (OpenIE6)	52.7	33.7	**46.4**	**26.8**	**65.6**	**48.4**	**40.0**	31.7

Experiments: Ablation Study

Coordination Analyzer	IMoJIE	CIGL-OIE
None	36.0	36.8
CalmIE	37.7	38.0
(Teranishi et al., 2019)	36.1	36.5
IGL-CA	**39.5**	**40.0**

Table 6: Wire57 F1 scores of IMoJIE and CIGL-OIE with addition of different coordination analyzers. IGL-CA improves both of the OpenIE extractors.

System	Wire57-C	CaRB		Constraint Violations					Num. of Extrs
	F1	F1	AUC	POSC	HVC	HVE	EC	HVC+HVE+EC	
IMoJIE	36.0	53.5	33.3	687	521	105	330	957	1354
IGL-OIE	34.9	52.4	33.7	1494	375	**128**	284	787	1401
IGL-OIE (POSC)	36.7	49.6	33.4	**396**	303	200	**243**	746	**1577**
IGL-OIE (HVC,HVE,EC)	35.8	53.2	32.7	1170	295	144	246	**655**	1509
CIGL-OIE	**36.8**	**54.0**	**35.7**	766	**274**	157	237	668	1531
Gold	100	100	100	371	324	272	224	820	2714

Table 4: Performance and number of constraint violations for training with different sets of constraints. CIGL-OIE represents training IGL architecture based OpenIE extractor with all the constraints - POSC, HVC, HVE and EC

[EMNLP21]Maximal Clique Based Non-Autoregressive Open Information Extraction

Bowen Yu,[1,2] Yucheng Wang,[1,2] Tingwen Liu,[1,2] Hongsong Zhu,[1,2] Limin Sun,[1,2] Bin Wang[3]

[1]*Institute of Information Engineering, Chinese Academy of Sciences*
[2]*School of Cyber Security, University of Chinese Academy of Sciences*
[3]*Xiaomi AI Lab, Xiaomi Inc., Beijing, China*

Motivations

- Complicated facts: Overlapping; Discontinuous; Nested
- Auto-regressive systems: enforce an unnecessary order on the facts; error accumulation

John is the premier and first minister of British Columbia

Subject	Predicate	Object
John	premier of	British Columbia
John	first minister of	British Columbia

Motivations

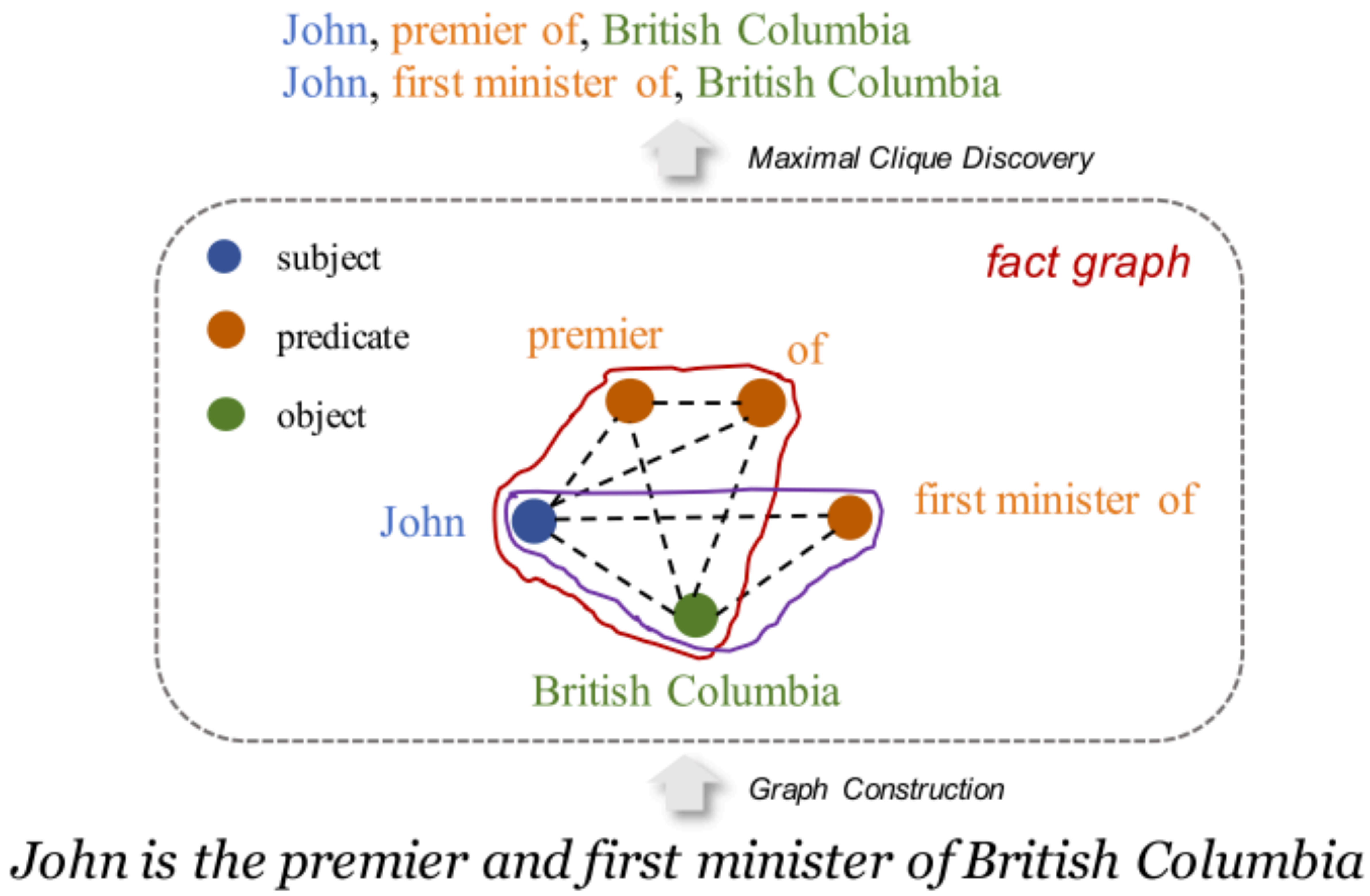
- Complicated facts: Overlapping; Discontinuous; Nested
- Auto-regressive systems: enforce an unnecessary order on the facts; error accumulation

non-autoregressive OpenIE framework

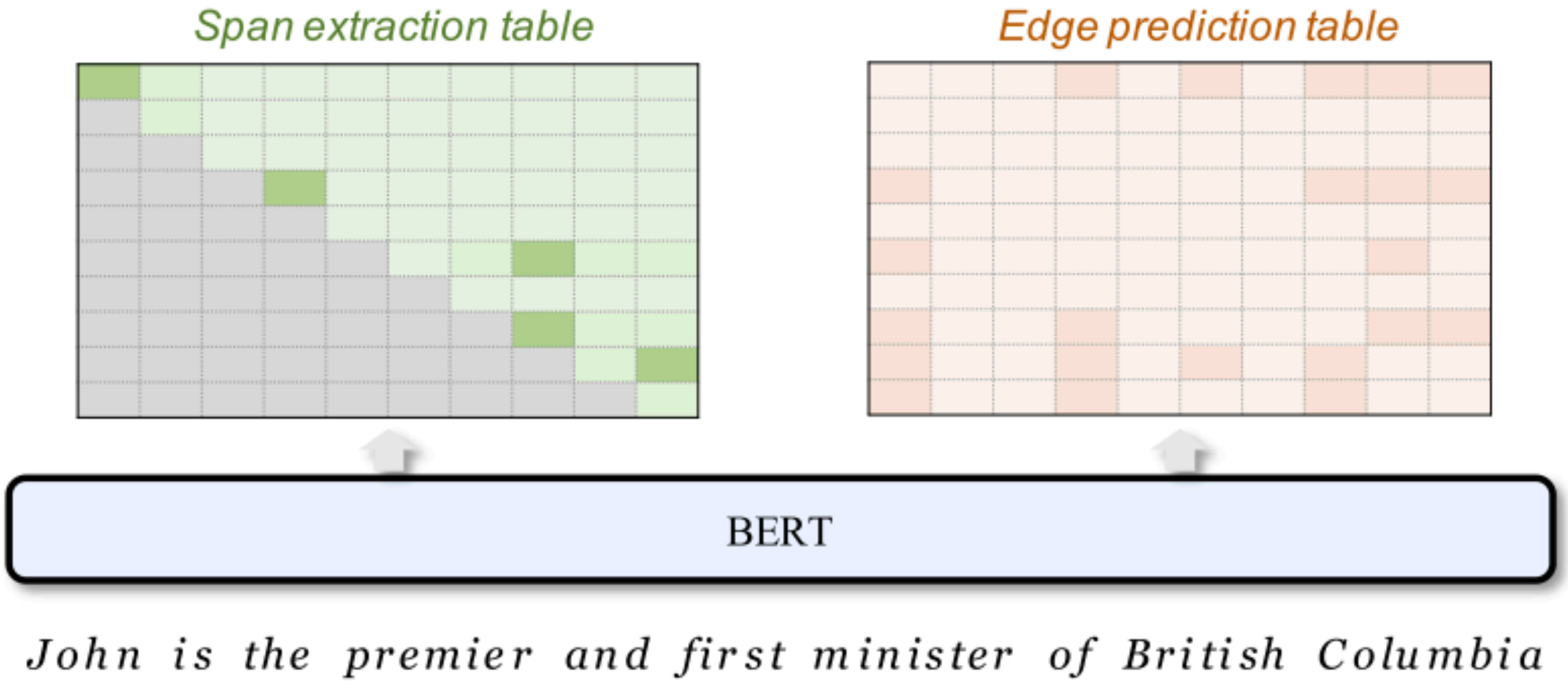
Maximal clique discovery based open Information Extractor (MacroIE)

Task Definition

- Open IE is cast as a maximal clique discovery task.
  - Node: a continuous span
  - Edge: two nodes belonging to the same fact



MacroIE Architecture



Architecture: Span Extraction

Subject	Predicate	Object
John	premier of	British Columbia
John	first minister of	British Columbia

	John	is	the	premier	and	first	minister	of	British	Columbia
John	B2E									
is										
the										
premier				B2E						
and										
first								B2E		
minister										
of								B2E		
British										B2E
Columbia										

B2E: Beginning-End tag

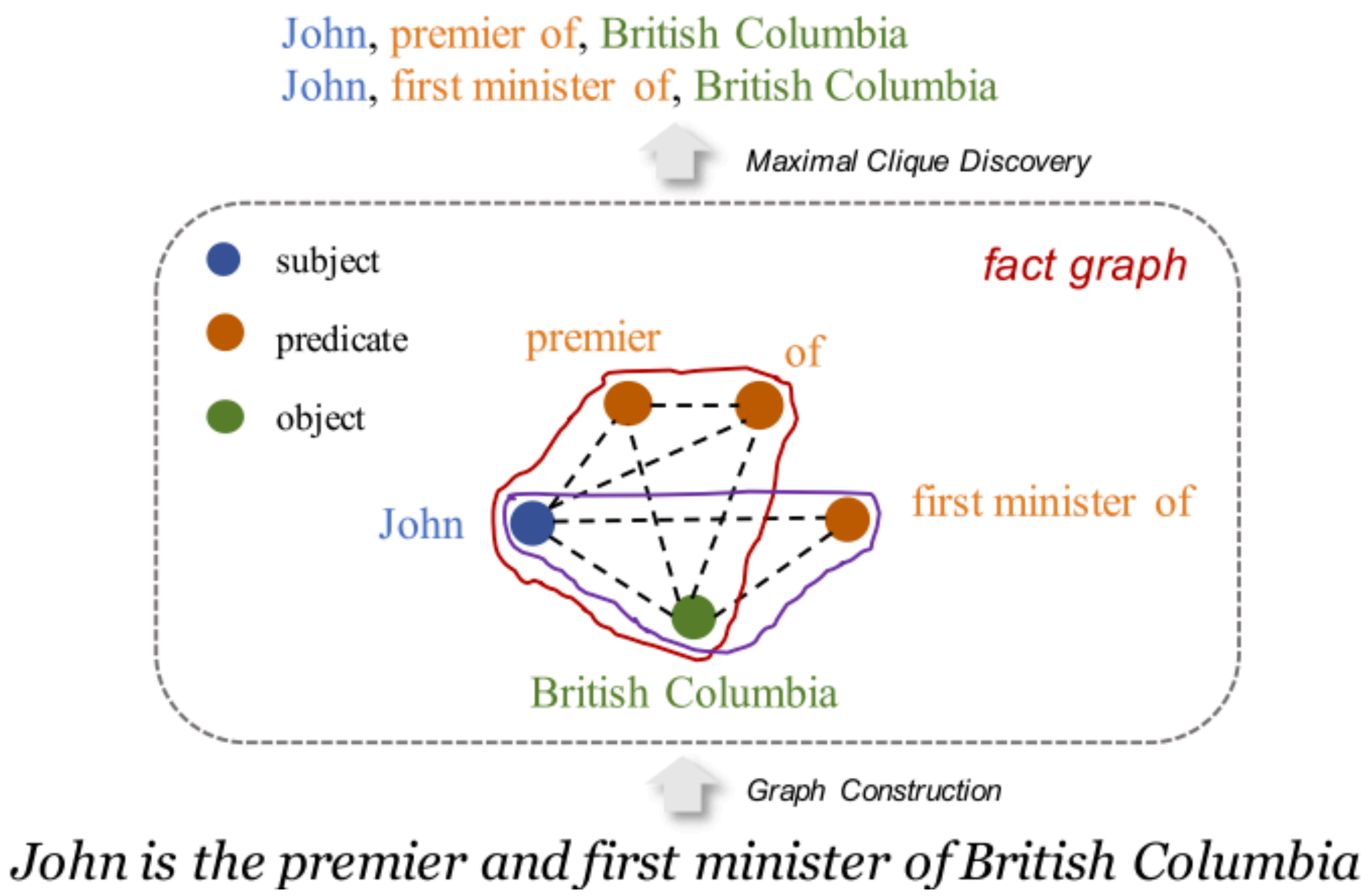
Architecture: Edge Prediction

Subject	Predicate	Object
John	premier of	British Columbia
John	first minister of	British Columbia

	John	is	the	premier	and	first	minister	of	British	Columbia
John				B-S2P E-S2P		B-S2P		B-S2P E-S2P	B-S2O	E-S2O
is										
the										
premier	B-P2S E-P2S							B-P2P E-P2P	B-P2O	E-P2O
and										
first	B-P2S								B-P2O	
minister										
of	B-P2S E-P2S			B-P2P E-P2P					B-P2O	E-P2O
British	B-O2S			B-O2P		B-O2P		B-O2P		
Columbia	E-O2S			E-O2P				E-O2P		

Tag: Position-Role
B: Beginning, E: End,
S: Subject, P: Predicate,
O: Object

MacroIE Workflow



Algorithm 1 Overall workflow

Input: Sentence $S = \{w_1, w_2, \cdots, w_n\}$
Output: The fact set expressed in S, denoted as F.

Fill the SE table T_s and EP table T_e with Equation 3

Decode T_s to obtain the span set P
Initialize the fact graph G with P
for span $v \in P$ **do**
  for another span $g \in P$ **do**
    if Fact Graph Construction E-⋆ $\in T_e(v.\text{end}, v.\text{end})$ **then**
      Connect v and g in G
    end if
  end for
end for

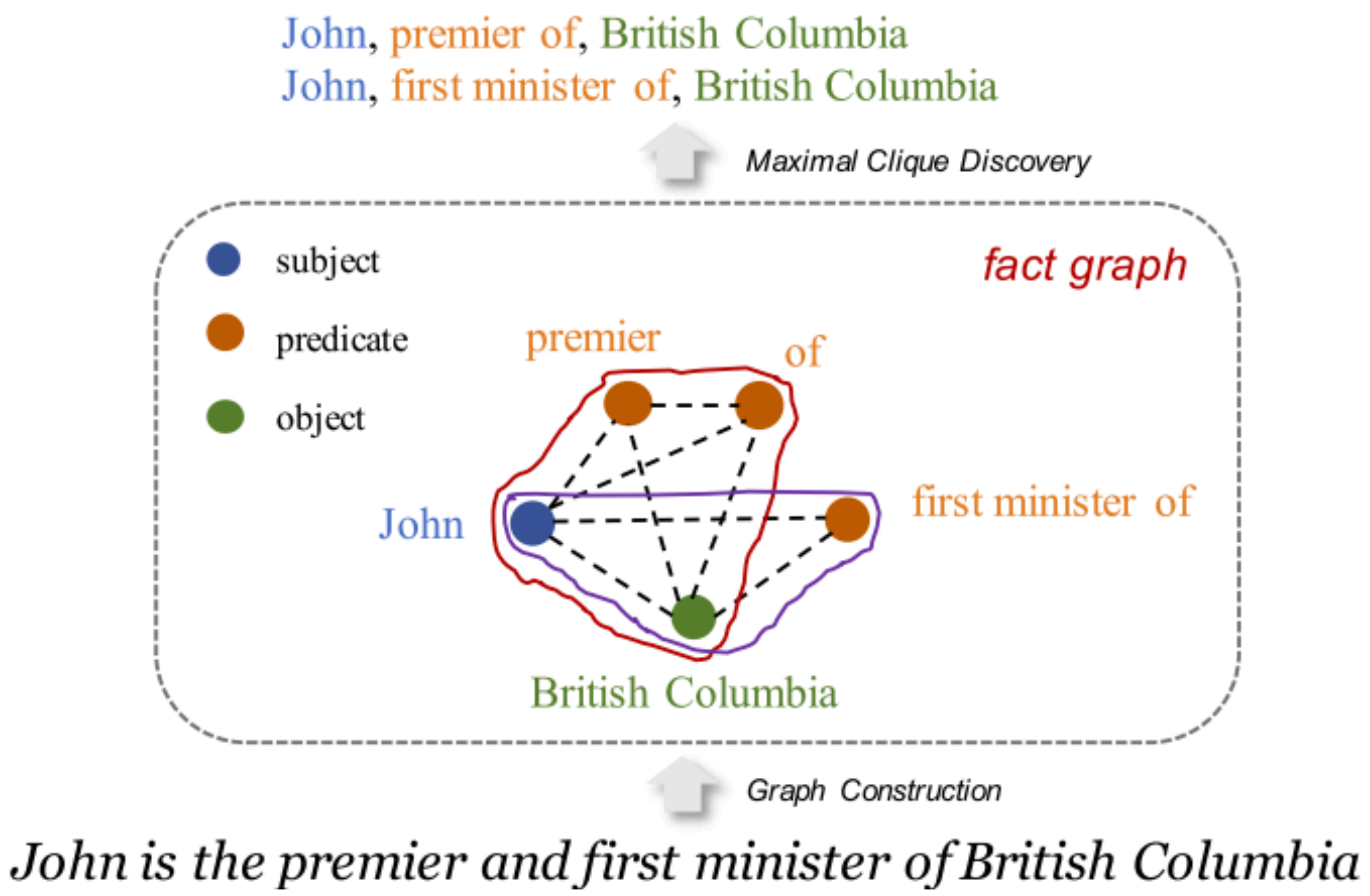
Find the maximal cliques C in G with Algorithm ??

for clique $c \in C$ **do**
  for span $v \in c$ **do**
    Initialize the role list of v with $\varnothing$, denoted as R_v
    for another span $g \in c$ **do**
      Add the outgoing role part of each tag in $T_e(v.\text{begin}, g.\text{begin})$ and $T_e(v.\text{end}, g.\text{end})$ to R_v
    end for

Clique-to-Fact Transformation

  end for
  Merge the spans of the same role type with their order in S as the fact element.
  Assemble elements to constitute a fact and add it to F
end for
return F

Workflow: Fact Graph Construction



Algorithm 1 Overall workflow

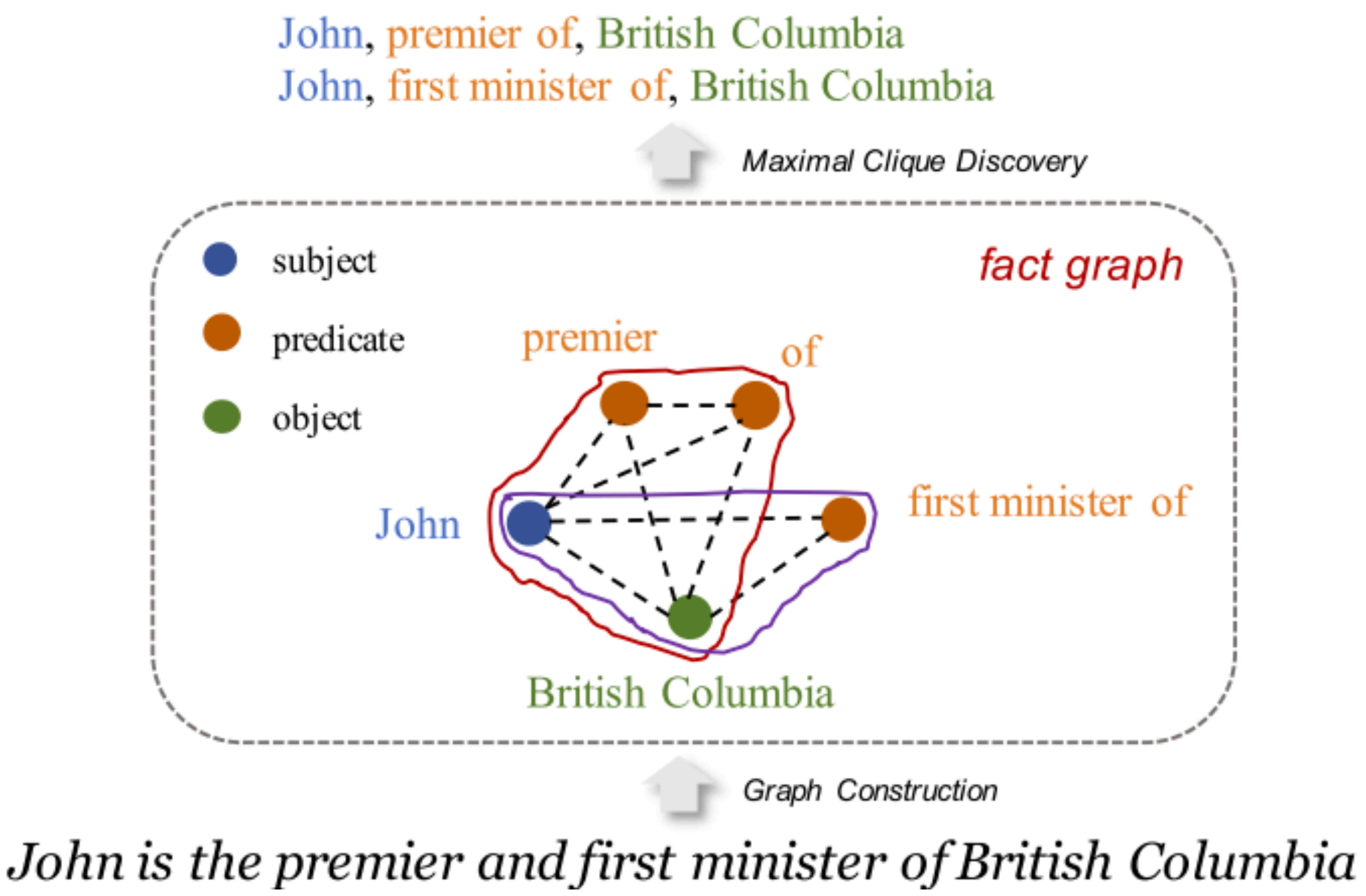
Input: Sentence $\mathrm{S} = \{w_1, w_2, \cdots, w_n\}$
Output: The fact set expressed in S, denoted as F.

Fill the SE table T_s and EP table T_e with Equation 3
Decode T_s to obtain the span set P **Node**
Initialize the fact graph G with P
for span $v \in \mathrm{P}$ **do**
  for another span $g \in \mathrm{P}$ **do**
    if $\texttt{B-}\star \in \mathrm{T}_e(v.\text{begin}, g.\text{begin})$ & $\texttt{E-}\star \in \mathrm{T}_e(v.\text{end}, v.\text{end})$ **then** **Edge**
      Connect v and g in G
    end if
  end for
end for
Find the maximal cliques C in G with Algorithm ??
for clique $c \in \mathrm{C}$ **do**
  for span $v \in c$ **do**
    Initialize the role list of v with $\varnothing$, denoted as R_v
    for another span $g \in c$ **do**
      Add the outgoing role part of each tag in $\mathrm{T}_e(v.\text{begin}, g.\text{begin})$ and $\mathrm{T}_e(v.\text{end}, g.\text{end})$ to R_v
    end for

Clique-to-Fact Transformation

  end for
  Merge the spans of the same role type with their order in S as the fact element.
  Assemble elements to constitute a fact and add it to F
end for
return F

Workflow: Clique-to-Fact Transformation



Algorithm 1 Overall workflow

Input: Sentence $\mathrm{S} = \{w_1, w_2, \cdots, w_n\}$
Output: The fact set expressed in S, denoted as F.

Fill the SE table T_s and EP table T_e with Equation 3
Decode T_s to obtain the span set P
Initialize the fact graph G with P
for span $v \in$ P **do**
    for another span $g \in$ P **do**
        if **Fact Graph Construction** E−⋆ $\in$ $\mathrm{T}_e(v.\text{end}, v.\text{end})$ **then**
            Connect v and g in G
        end if
    end for
end for
Find the maximal cliques C in G with Algorithm ??
for clique $c \in$ C **do**
    for span $v \in c$ **do**
        Initialize the role list of v with $\varnothing$, denoted as R_v
        for another span $g \in c$ **do**
            Add the outgoing role part of each tag in $\mathrm{T}_e(v.\text{begin}, g.\text{begin})$ and $\mathrm{T}_e(v.\text{end}, g.\text{end})$ to R_v
        end for
        Select the most frequent role type in R_v as the role of v in the clique c Node Role
    end for
    Merge the spans of the same role type with their order in S as the fact element.
    Assemble elements to constitute a fact and add it to F
end for
return F

Experiments

- Train data: OpenIE4, dev/test data: CaRB

	OpenIE4	
	# of sentences	# of facts
Train	91,277	180,517
Dev	638	2,548
Test	634	2,715

System	CaRB(1-1)			CaRB (word)			Gestalt (word + order)		
	F1	AUC	Opt.F1	F1	AUC	Opt.F1	F1	AUC	Opt.F1
SenseOIE (Roy et al., 2019)	23.9	-	23.9	28.2	-	28.2	-	-	-
SpanOIE (Zhan and Zhao, 2020)	37.9	-	37.9	48.5	-	48.5	-	-	-
RnnOIE (Stanovsky et al., 2018)	39.3	18.3	39.5	49.0	26.1	49.1	7.4	4.9	7.7
NOIE (Cui et al., 2018)	38.3	19.8	38.7	51.1	32.8	51.6	9.0	4.3	9.2
IMoJIE (Kolluru et al., 2020b)	41.2	22.2	41.4	53.3	33.3	53.5	9.6	5.6	9.4
IGL-OIE (Kolluru et al., 2020a)	41.0	22.9	41.1	52.2	33.7	52.4	10.1	5.4	9.7
MacroIE (ours)	**43.5**	**25.0**	**43.8**	**54.8**	**36.3**	**55.1**	**12.9**	**6.0**	**13.1**

Table 3: Main results on OpenIE4. The improvement over baselines is significant (p-value < 0.05).

Experiments

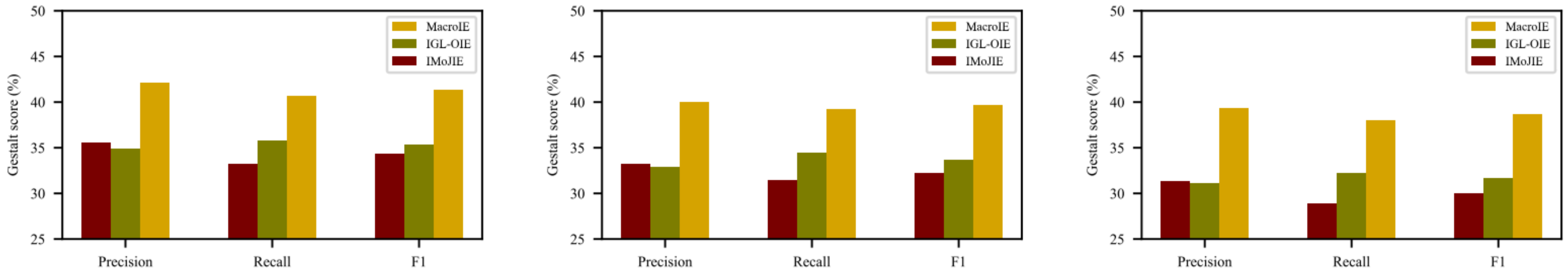
- SAOKE: largest publicly available human-annotated OIE dataset (Chinese)

SAOKE	
# of sentences	# of facts
37,544	133,400
4,693	16,563
4,693	16,407

System	CaRB(1-1)			CaRB			Gestalt		
	F1	AUC	Opt.F1	F1	AUC	Opt.F1	F1	AUC	Opt.F1
IMoJIE (Kolluru et al., 2020b)	36.6	22.6	37.0	38.7	25.4	39.5	36.4	22.5	37.3
IGL-OIE (Kolluru et al., 2020a)	37.6	22.8	38.4	39.3	25.5	40.6	37.1	23.6	38.4
MacroIE (ours)	**41.2**	**24.5**	**41.5**	**42.7**	**27.8**	**43.7**	**42.8**	**27.2**	**43.7**

Table 4: Main results on SAOKE. The improvement over baselines is significant (p-value < 0.05).

Additional Experiments



(a) Performance on Overlapping Facts (b) Performance on Discontinuous Facts (c) Performance on Nested Facts

Figure 6: Gestalt scores on (a) extracting overlapping facts, (b) identifying discontinuous facts, and (c) detecting nested facts. All the analyses are conducted on the test set of SAOKE.

Additional Experiments

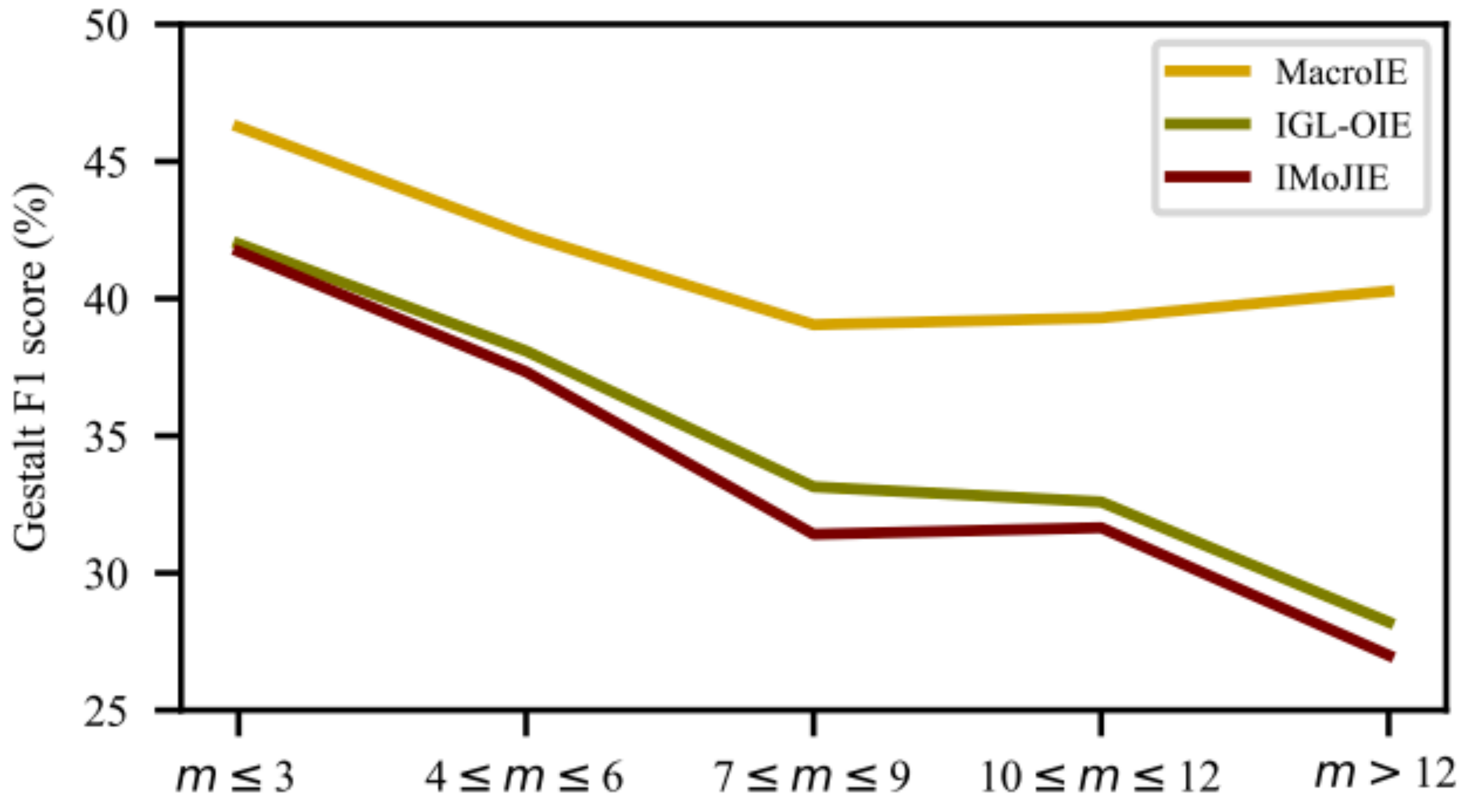


Figure 7: Gestalt F1 score of conducting extraction in sentences that contain different numbers of facts (m).

Thanks

Q & A