Event Extraction from Texts

Kang Liu
Institute of Automation, Chinese Academy of Sciences
2017.12.8

Texts to Knowledge

习近平指出，中巴两国友谊源远流长。中国是最早支持巴勒斯坦人民正义事业、最早承认巴勒斯坦解放组织和巴勒斯坦国的国家之一。两国人民相互理解、相互信赖、相互支持，是真正的好朋友、好伙伴、好兄弟。近年来，中巴两国高层交往密切，政治互信更加巩固，各领域合作得到稳步发展。

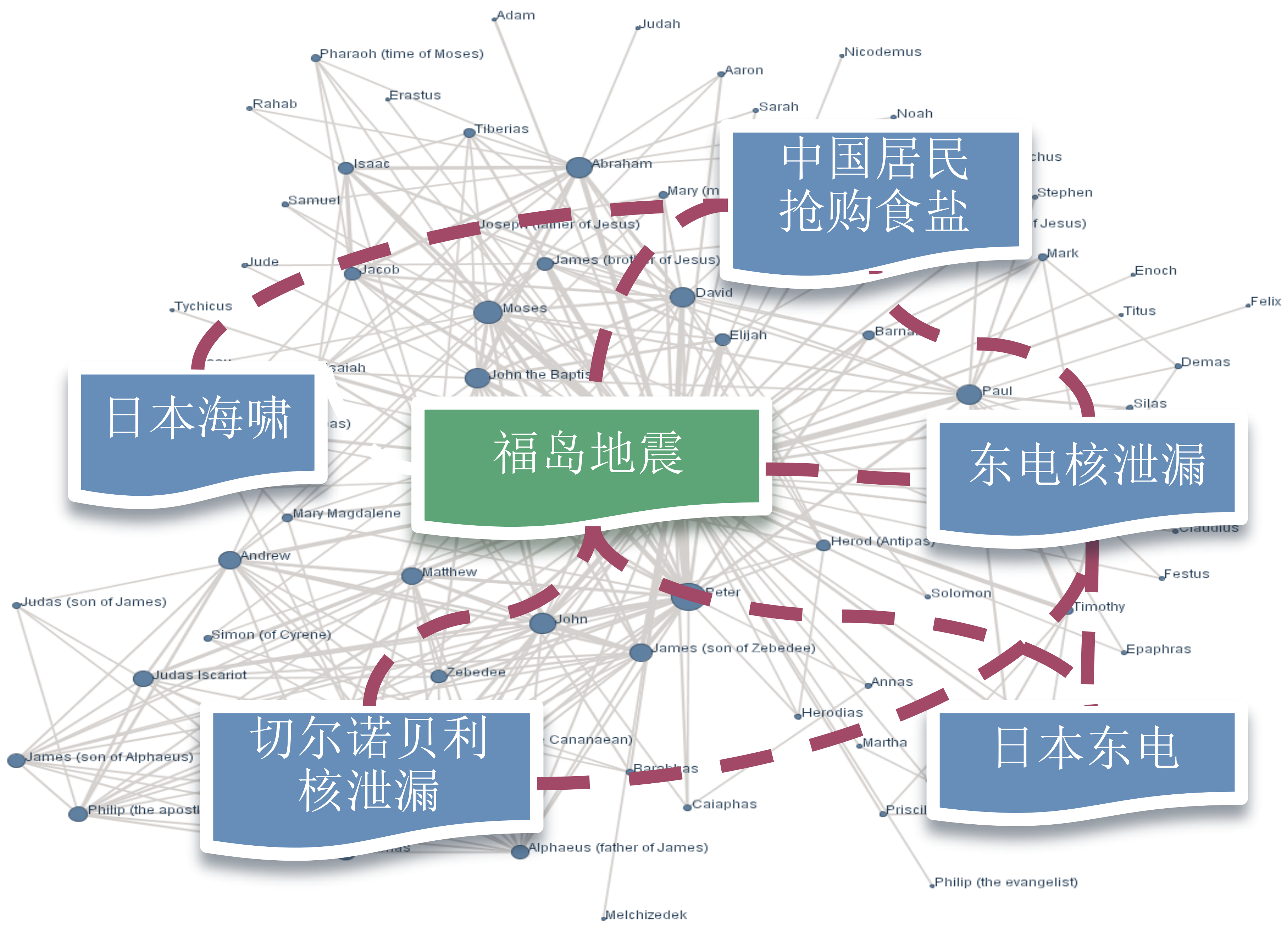
习近平强调，中方愿同巴方一道努力，在政治上继续坚定相互支持，加强协调配合，保持高层交往，不断推动两国各领域合作。中方赞赏巴方始终奉行一个中国政策，将一如既往支持巴勒斯坦人民恢复民族合法权利的正义事业。中方愿与巴方共建“一带一路”，支持有实力、有条件的企业到巴勒斯坦开展投资合作，实现互利共赢。我们愿同巴方在工业区建设、人才培训和太阳能电站项目等方面合作，帮助巴方提升自主发展能力。双方要继续加强在文化、教育、科研、党际、地方、民间、青年等各领域各层次交流合作，不断增进两国人民友谊。

习近平就推动解决巴勒斯坦问题提出四点主张：

第一，坚定推进以“两国方案”为基础的政治解决。中方坚定支持“两国方案”，支持建立以1967年边界为基础、以东耶路撒冷为首都、拥有完全主权的、独立的巴勒斯坦国，将一如既往地为解决巴勒斯坦问题发挥建设性作用。

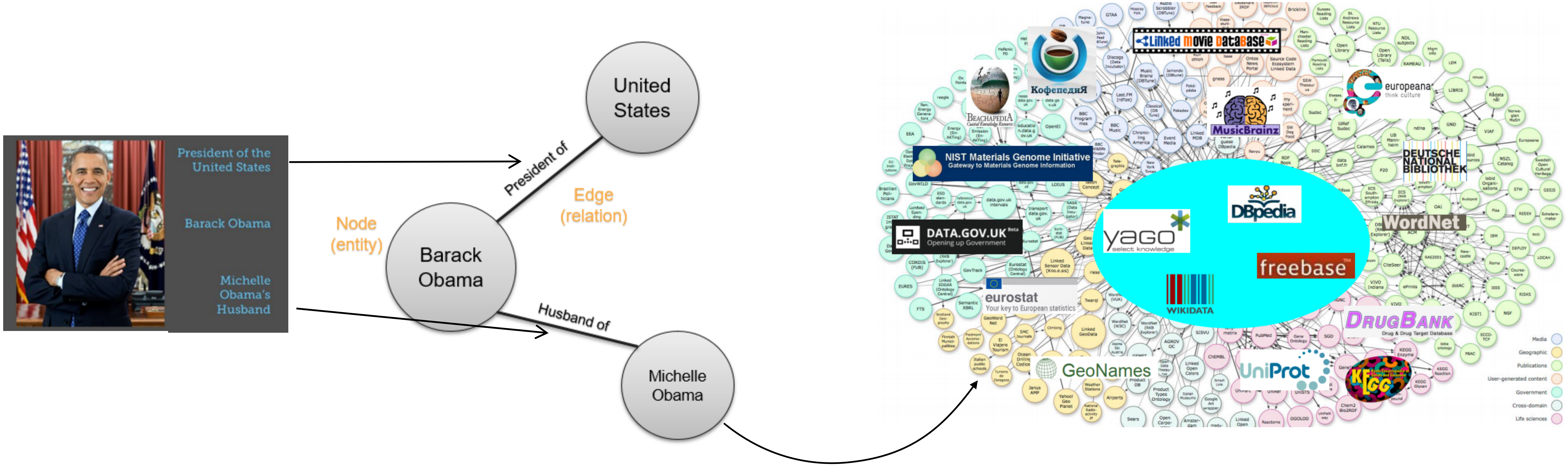
第二，坚持共同、综合、合作、可持续的安全观。中方呼吁切实落实联合国安理会第2334号决议，立即停止在被占领土上一切定居点活动，立即采取措施，防止针对平民的暴力行为。尽快复谈，加快政治解决巴勒斯坦问题，从根本上实现共同持久的安全。

Texts



Knowledge

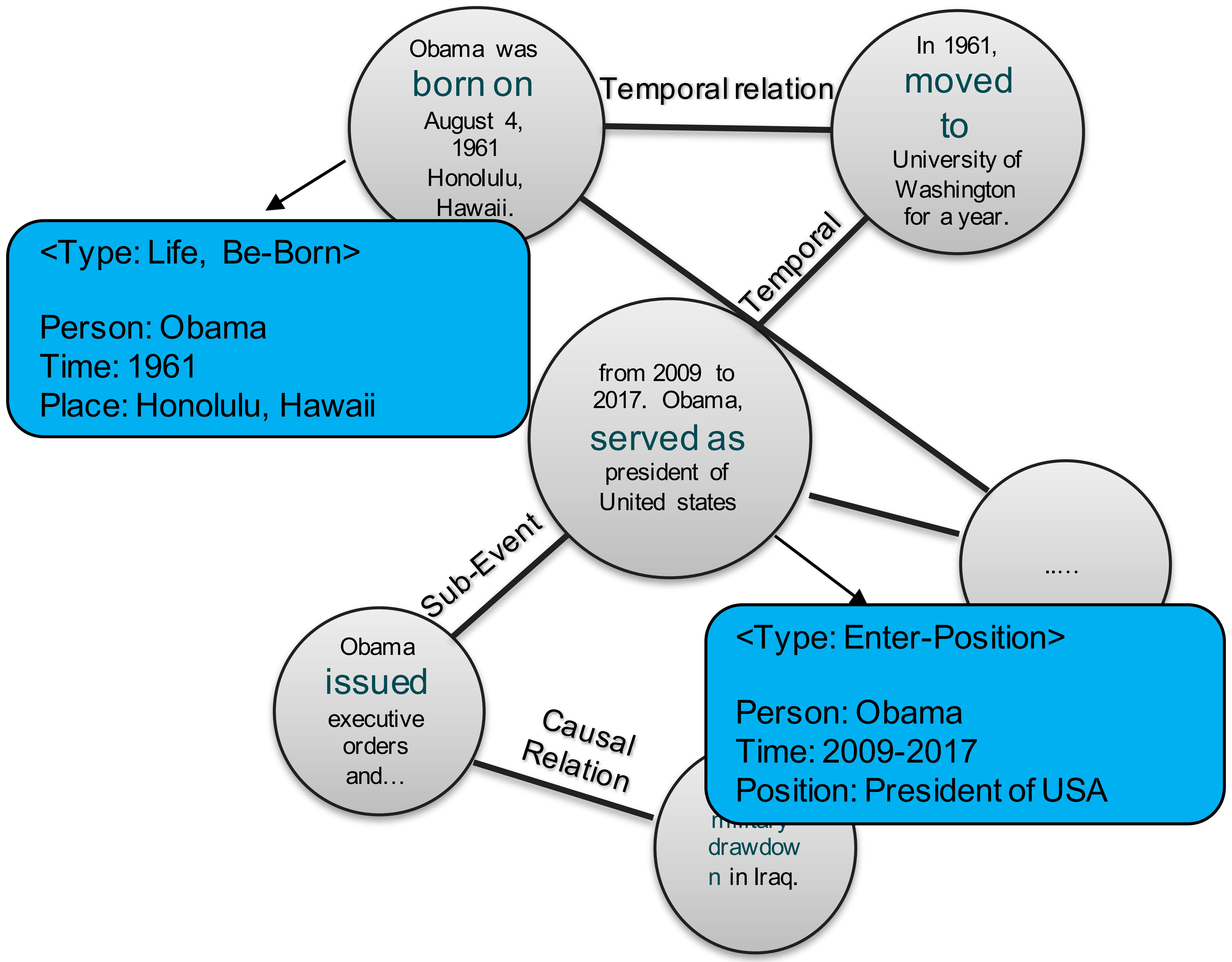
Statistic Knowledge: Entity-Centric Knowledge Graph



(Barack Obama, Spouse, Michelle Obama)

Head Entity **Relation** **Tail Entity**

Dynamic Knowledge: Event-Centric Knowledge Graph



出生事件
- 出生日期
- 出生地点
- 姓名

结婚事件
- 结婚日期
- 结婚地点
- 男方
- 女方

离职事件
- 离职日期
- 公司
- 职位

地震事件
- 震中
- 震级
- 震源
- 伤亡人数
- 财产损失

暴恐事件
- 地点
- 时间
- 伤亡人数
- 被攻击方
- 实施方

收购事件
- 收购金额
- 收购方
- 被收购方
- 时间

Event Frame (Script)

Extract Events from Texts

Barry Diller on Wednesday ***quit*** as chief of Vivendi Universal Entertainment.

Trigger	Quit (a "Personnel/End-Position" event)	
Arguments	Role = Person	Barry Diller
	Role = Organization	Vivendi Universal Entertainment
	Role = Position	Chief
	Role = Time-within	Wednesday (2003-03-04)

Extract Events from Texts

Organization

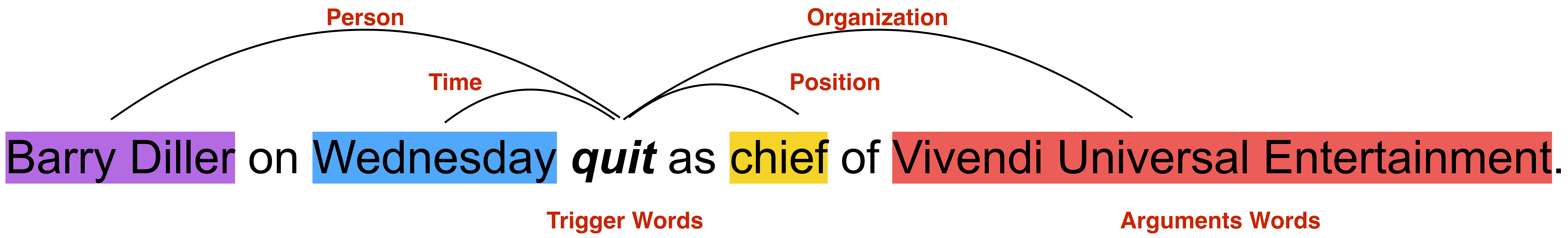
Barry Diller on Wednesday ***quit*** as chief of Vivendi Universal Entertainment.

Trigger Words

Arguments Words

Trigger	Quit (a "Personnel/End-Position" event)	
Arguments	Role = Person	Barry Diller
	Role = Organization	Vivendi Universal Entertainment
	Role = Position	Chief
	Role = Time-within	Wednesday (2003-03-04)

Extract Events from Texts



Trigger	Quit (a "Personnel/End-Position" event)	
Arguments	Role = Person	Barry Diller
	Role = Organization	Vivendi Universal Entertainment
	Role = Position	Chief
	Role = Time-within	Wednesday (2003-03-04)

Task Definition

- Definition (ACE)
  - An event is defined as a specific occurrence involving participants.
  - Event trigger, Event Type, Event argument, Argument role

Barry Diller on Wednesday ***quit*** as chief of Vivendi Universal Entertainment.

Trigger	Quit (a "Personnel/End-Position" event)	
Arguments	Role = Person	Barry Diller
	Role = Organization	Vivendi Universal Entertainment
	Role = Position	Chief
	Role = Time-within	Wednesday (2003-03-04)

1. Event Identification (Trigger Words)
2. Event Type Identification
3. Argument Identification
4. Argument Role Identification

Event Extraction vs. Relation Extraction

- Relation Extraction
  - Identify the relation between **two given entities**

/business/company/founder

Steve Jobs was the co-funder of Apple Inc.

entity1 entity2

- Event Extraction
  - Identify the relation between **an event and an entity**

Person Organization

Time Position

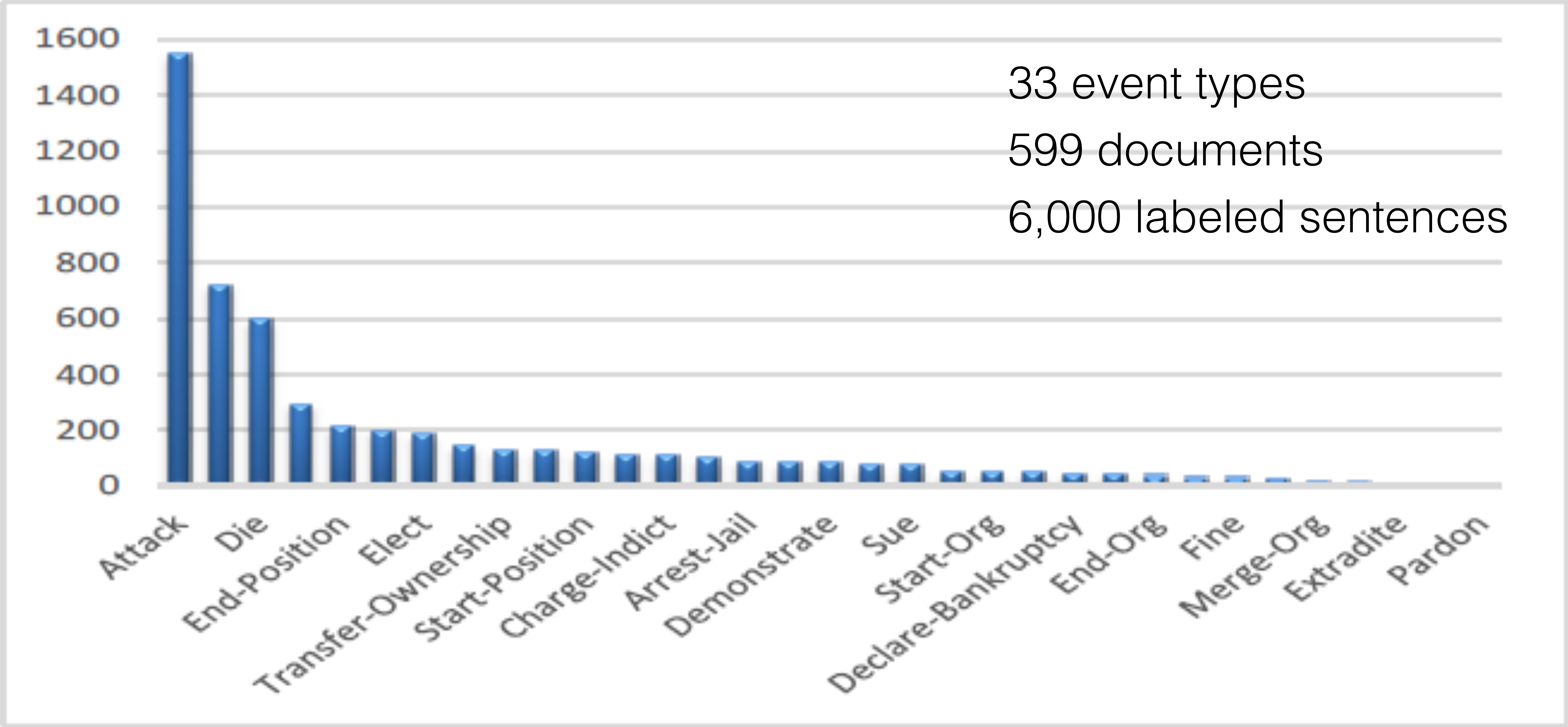
Barry Diller on Wednesday ***quit*** as chief of Vivendi Universal Entertainment.

Trigger Words Arguments Words

Previous Event Extraction Task

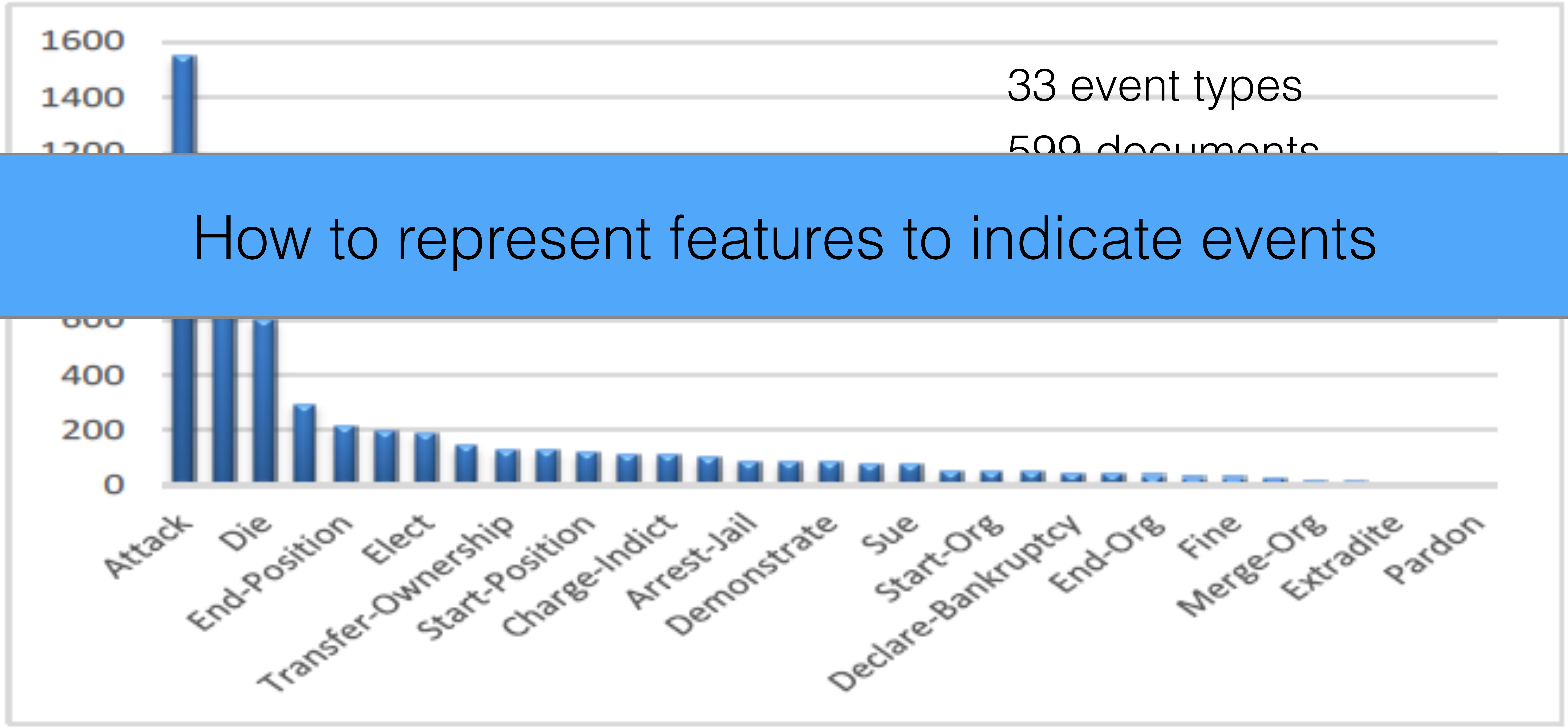
	MUC Message Understanding Conference	TDT Topic Detection and Tracking	ACE(KBP) Automatic Content Extraction
Organizer	DARPA	DAPRA	NIST
Period	1987-1997	1998-2004	ACE:2000-2008 KBP:2014-2017
Content	抽取指定的事件，包括参与这些事件的各个实体、属性和关系。例如：MUC-2是从海军军事情报中抽取事件填入预定义模板中，共10个槽	将文本切割为不同的新闻报道，监控其中新事件的报道，并且将同一话题下的分散的报道按照某种结构有效组织起来。 TDT-3:240个topic	指定的源语言数据中发现特定类型的事件，并且识别出与事件相关的信息填入预设的事件模板中。 ACE中共计8大类33个小类的事件

Challenges in Event Extraction (Open Domain)



English data in ACE 2005

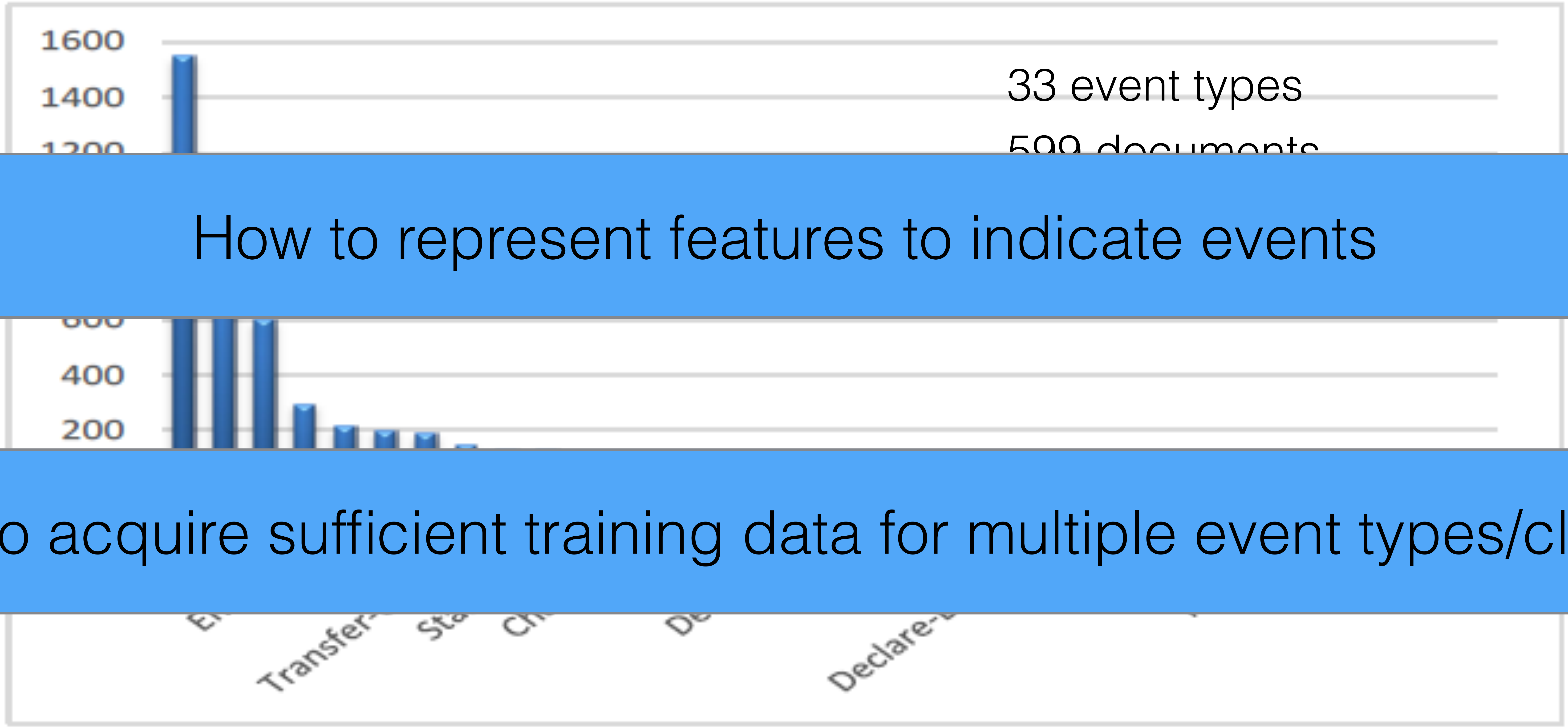
Challenges in Event Extraction (Open Domain)



How to represent features to indicate events

English data in ACE 2005

Challenges in Event Extraction (Open Domain)



How to represent features to indicate events

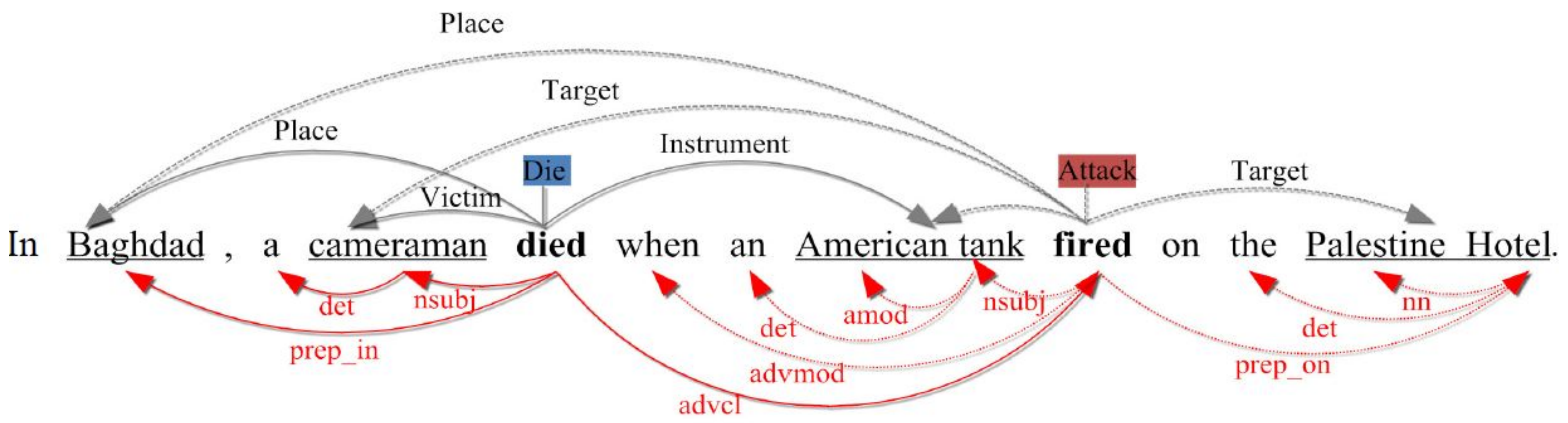
How to acquire sufficient training data for multiple event types/classes

English data in ACE 2005

Feature Representation

Traditional Methods for Feature Representation

- Human designed features



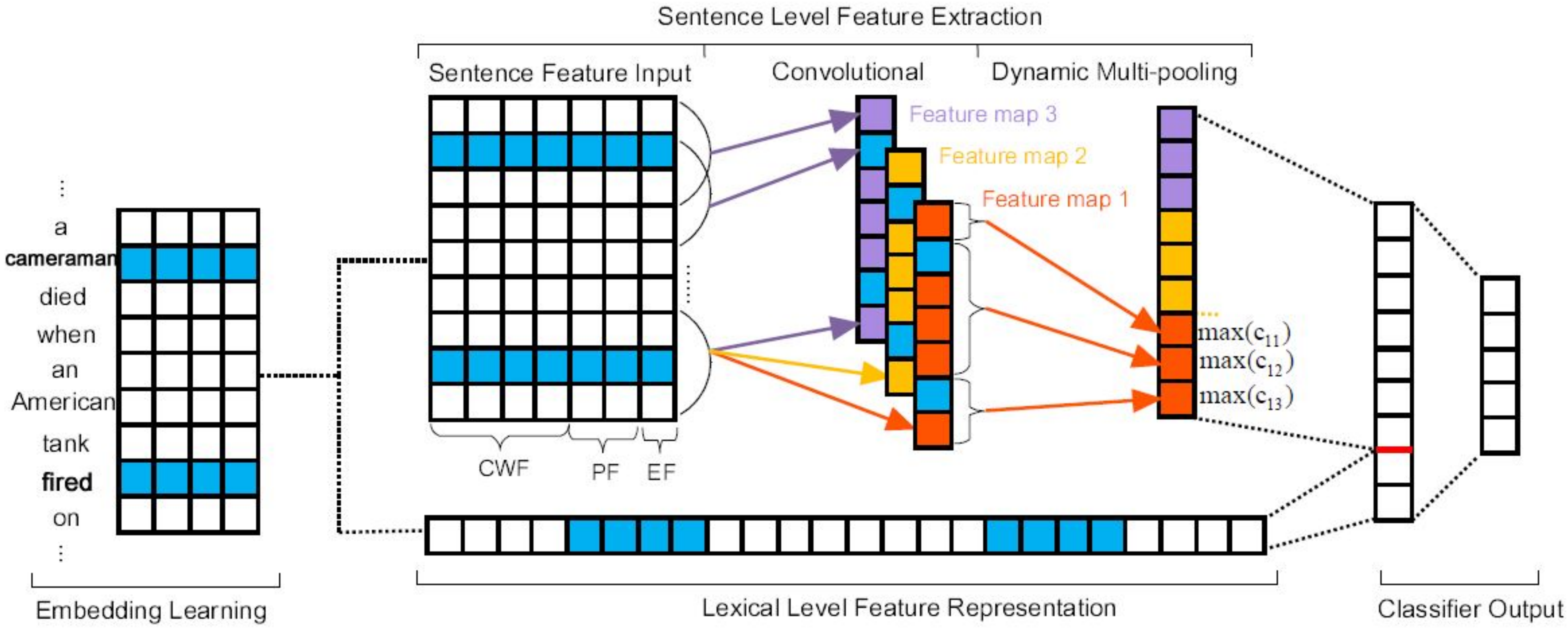
nsubj -> (cameraman plays the Victim role in die event)

????? -> (cameraman plays the Target role in Attack event)

- Too much rely on imprecise NLP tools for feature extraction
- Limitations for low-resources languages

Dynamic Multi-pooling Convolutional Neural Network

- We propose a Dynamic Multi-pooling Convolutional Neural Network(DMCNN) to automatically capture the lexical-level and sentence-level features without POS tagging/Syntactic Parsing/… (ACL-2015)



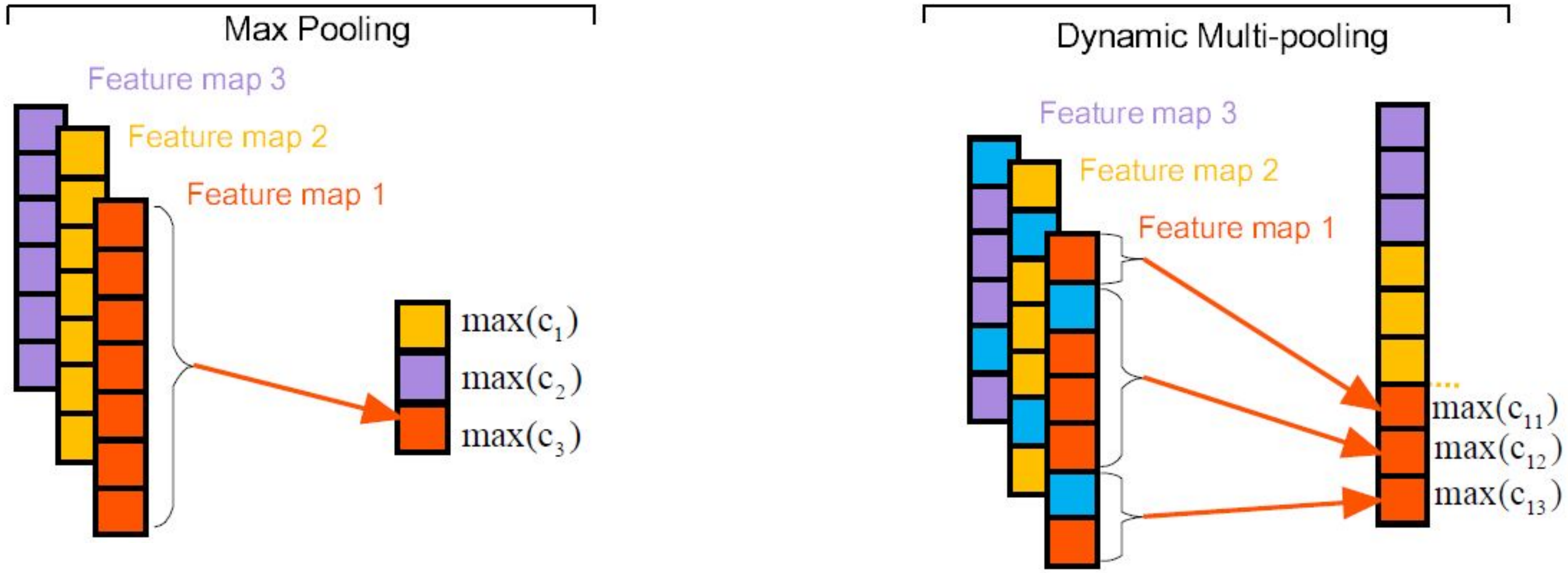
Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng and Jun Zhao, Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks, in Proceedings of ACL-IJCNLP 2015, Beijing, China, July, 26-29

Dynamic Multi-Pooling Layer

- Traditional CNNs use max pooling
- We perform Dynamic Multi-Pooling

Barry Diller on Wednesday ***quit*** as **chief** of Vivendi Universal Entertainment.

$$p_{ji} = \max(c_{ji})$$



Experimental Results

- ACE Dataset
  - 40 newswire articles from ACE 2005 as test set
  - 30 other documents from different genres as development set
  - The rest (529) documents for training

Methods	Trigger Identification(%)			Trigger Identification + Classification(%)			Argument Identification(%)			Argument Role(%)		
	P	R	F	P	R	F	P	R	F	P	R	F
Li's baseline	76.2	60.5	67.4	74.5	59.1	65.9	74.1	37.4	49.7	65.4	33.1	43.9
Liao's cross-event	N/A			68.7	68.9	68.8	50.9	49.7	50.3	45.1	44.1	44.6
Hong's cross-entity	N/A			72.9	64.3	68.3	53.4	52.9	53.1	51.6	45.5	48.3
Li's structure	76.9	65.0	70.4	73.7	62.3	67.5	69.8	47.9	56.8	64.7	44.4	52.7
DMCNN model	80.4	67.7	**73.5**	75.6	63.6	**69.1**	68.8	51.9	**59.1**	62.2	46.9	**53.5**

Compared with the state-of-the-arts

Stage	**Method**	**1/1**	**1/N**	**all**
		F_1	F_1	F_1
Trigger	Embedding+T	68.1	25.5	59.8
	CNN	72.5	43.1	66.3
	DMCNN	**74.3**	**50.9**	**69.1**
Argument	Embedding+T	37.4	15.5	32.6
	CNN	51.6	36.6	48.9
	DMCNN	**54.6**	**48.7**	**53.5**

Effect of dynamic max-pooling

Our method achieves the best performance without the need of the existing NLP tools

Event arguments are important to the Event Detection

Mohanmad fired Anwar, his former protege, in 1998.

Attack or *End-Position*?

- If we consider the argument phrase “former protege” (Role=Position), we will have more confidence to predict it as an End-Position event.

Shulin Liu, Yubo Chen, Shizhu He, Kang Liu and Jun Zhao, Leveraging FrameNet to Improve Automatic Event Detection, in Proceedings of ACL 2016, Berlin, Germany, August, 7-12

More Attentions on Argument Words

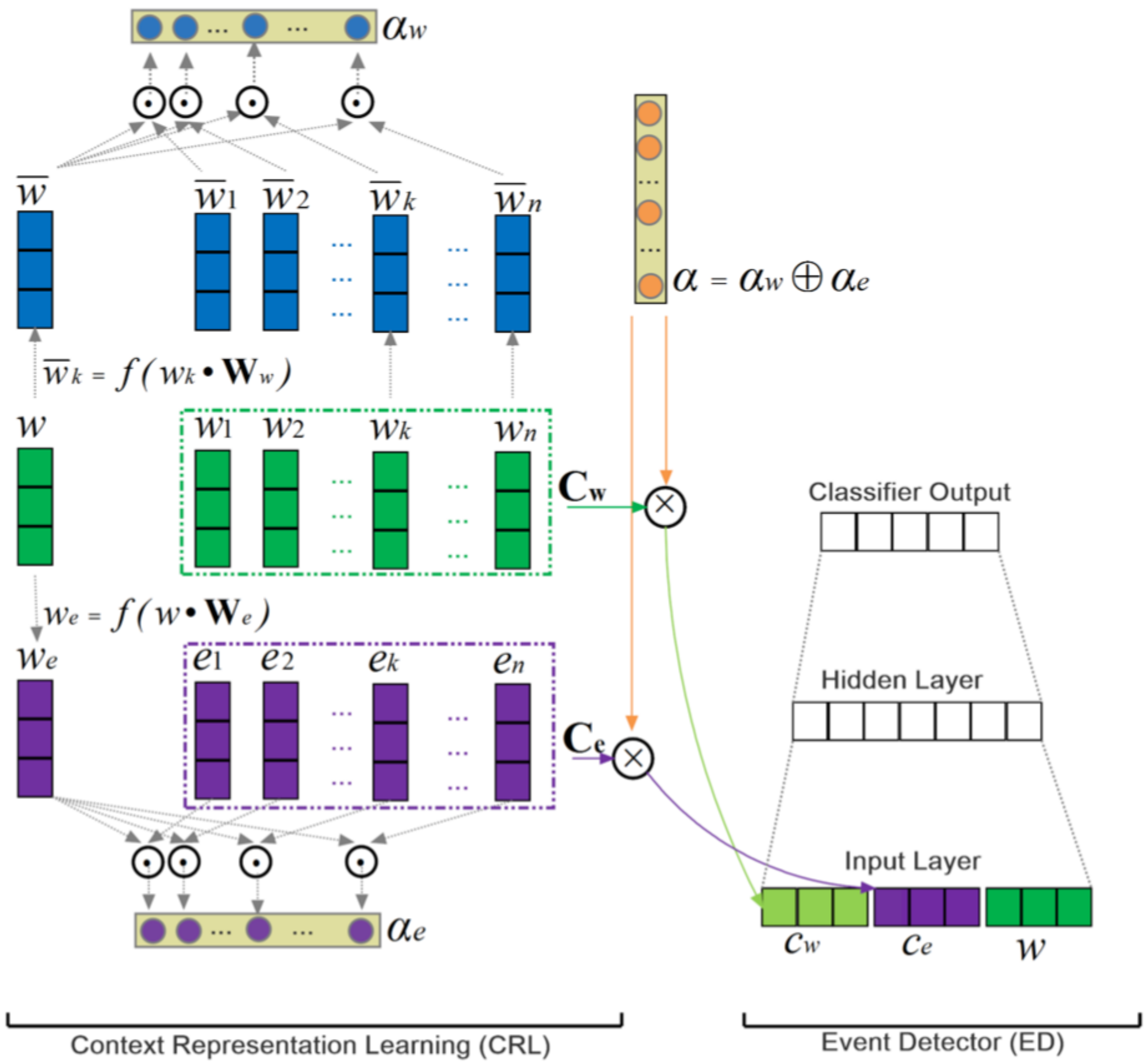
- The rp of trigger candidate w
    - the embedding of w
- The rp of the contextual words

$$c_w = \mathbf{C_w} \alpha^{\mathrm{T}}$$

- The rp of the contextual entities

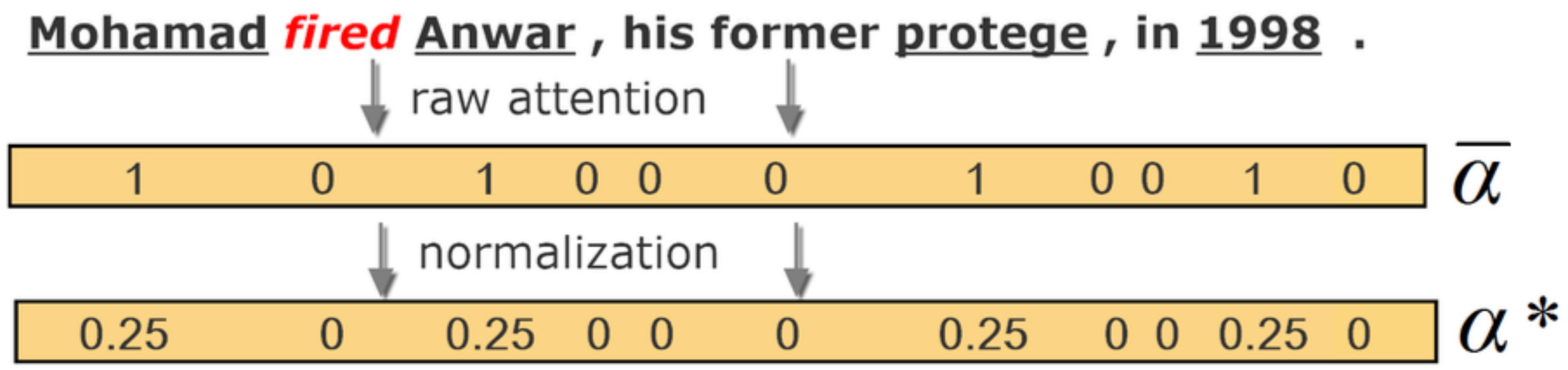
$$c_e = \mathbf{C_e} \alpha^{\mathrm{T}}$$

Cw and Ce are contextual words and entities, respectively.
α is the attention vector, which is computed as the manner illustrated in the right figure.



Attention Supervision

- **Strategy 1: only pay attention to argument words**



- **Strategy 2: pay attention to both argument words and their surroundings**
  - Step 1: obtaining the raw attention vector in the same manner as S1
  - Step 2: creating a new vector α' with all points initialized with zero
  - Step 3: for each $\overline{\alpha}_i = 1$, we update the new vector:
  - $$\alpha'_{i+k} = \alpha'_{i+k} + g(|k|, \mu, \sigma), k \in [-w, w]$$
  - Step 4: calculating the final attention vector α* by normalizing α'

Regularization in Learning Model

- Loss function of attentions

$$D(\theta) = \sum_{i=1}^{T} \sum_{j=1}^{n} ({\alpha^*}_j^i - \alpha_j^i)^2$$

- Joint loss function

$$J'(\theta) = J(\theta) + \lambda D(\theta)$$

Compared with State-of-the-arts

- ACE2005

Methods	P	R	F_1
Li's joint model (2013)	73.7	62.3	67.5
Liu's PSL (2016)	75.3	64.4	69.4
Liu's FraneNet-Based (2016)	77.6	65.2	70.7
Ngyuen's joint model (2016)	66.0	**73.0**	69.3
Ngyuen's Skip-CNN (2016)	N/A		71.3
ANN	73.2	57.9	64.6
ANN-S1†	**81.4**	62.4	70.8
ANN-S2†	78.0	66.3	**71.7**

Training Data Generation

Generating Labeled Data from Structured KB

- Distant (Weak) Supervision in Relation Extraction

Freebase

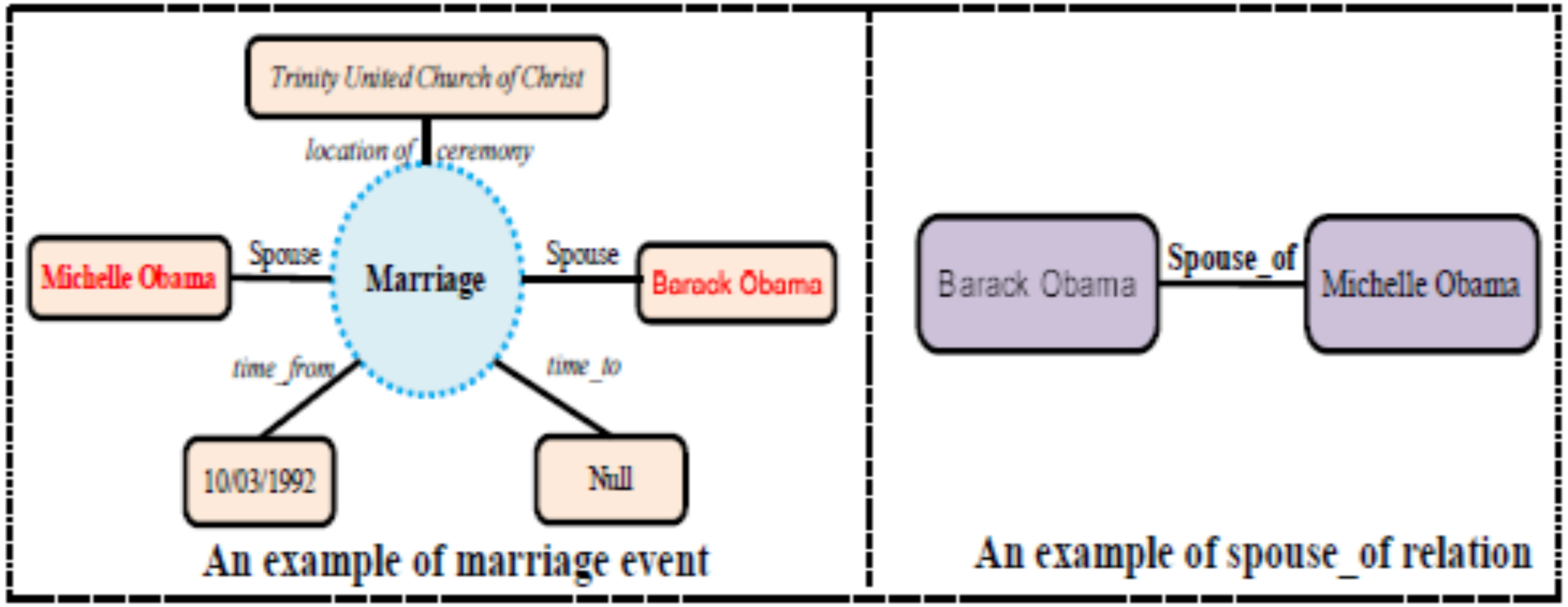
Knowledge base		
Relation	Entity 1	Entity 2
Founder	Steve Jobs	Apple
…	…	…

The New York Times

Sentence
Steve Jobs was the co-founder and CEO of **Apple** and formerly Pixar.
Steve Jobs passed away the day before **Apple** unveiled iPhone 4S.
…

The Strategy doesn't work for Event Extraction

- Triggers are not given out in existing knowledge bases



RE: (entity1, relation, entity2)
We can use **Michelle Obama** and **Barack Obama** to label back

EE:(event instance, event type; role1, argument1 ;...; rolen, argumennt)
We can not use **m:02nqglv** and **Barack Obama** to label back

In ACE, an event instance is represented as a trigger word

Training Data Generation

- Step1: Event Trigger Words Extraction

  Assumption: The sentences mention all arguments denote such events

- Step2: Argument Extraction/Role Identification
  - using Tigger words and Entities

Barry Diller on Wednesday ***quit*** as chief of Vivendi Universal Entertainment.

Solution: Using Key Arguments

Yubo Chen, Shulin Liu, Xiang Zhang, Kang Liu and Jun Zhao, Automatically Labeled Data Generation for Large Scale Event Extraction, To appear in Proceedings of ACL 2017, Vancouver, Canada, July 30-August 4.

Arguments Selection for Events

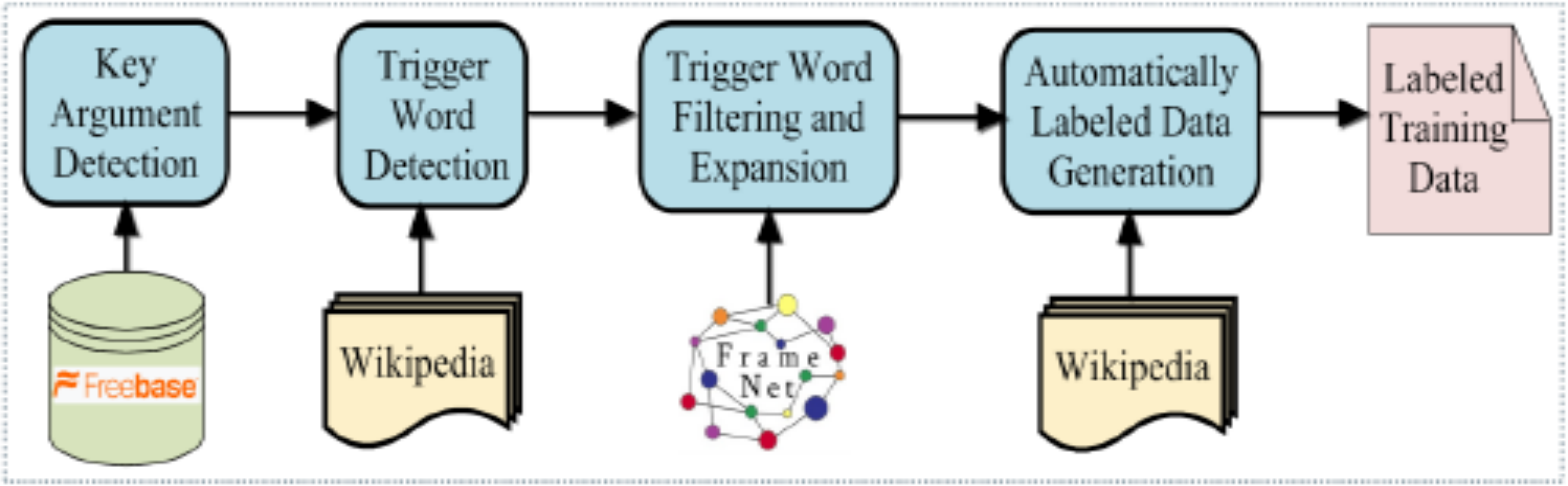
- Arguments for a specific event instance are usually mentioned in multiple sentences.

Event Type	EI#	A#	S#
education.education	530,538	8	0
film.film_crew_gig	252,948	3	8
people.marriage	152,276	5	0
...	...	...	...
military.military_service	27,933	6	0
olympics.olympic_medal_honor	20,790	5	4
sum of the selected 21 events	3,870,492	100	798

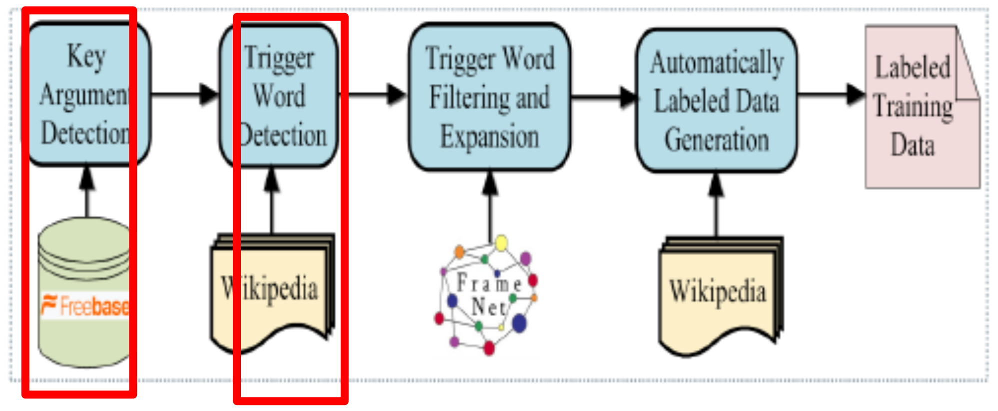
Statistics of events in Freebase.

Only 0.02% of instances can find all argument mentions in one sentence

Our Method for Generating Labeled Data in Event Extraction



Our Method for Generating Labeled Data in Event Extraction

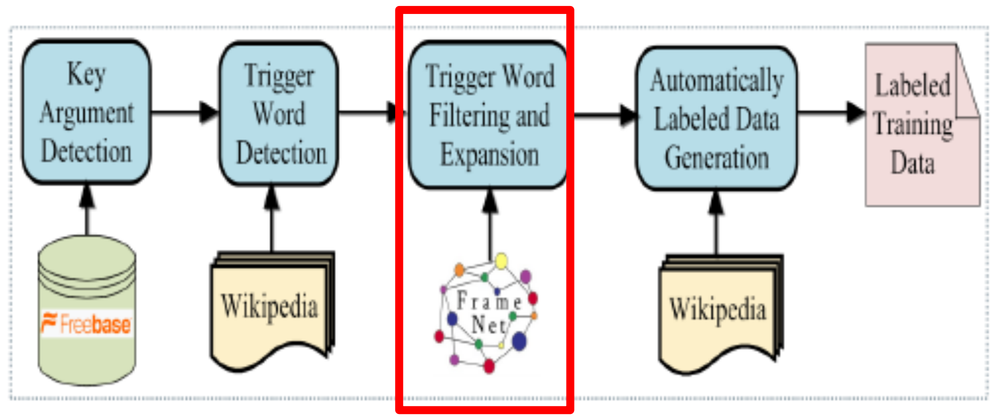


- Key Argument Detection
  - Role Saliency： $RS_{ij} = \frac{Count(A_i, ET_j)}{Count(ET_j)}$
  - Event Relevance： $ER_i = \log \frac{Sum(ET)}{1 + Count(ETCi)}$
  - Key Rate: $KR_{ij} = RS_{ij} * ER_i$

- Trigger Word Detection
  - Trigger Candidate Frequency： $TCF_{ij} = \frac{Count(V_i, ETS_j)}{Count(ETS_j)}$
  - Trigger Event Type Frequency： $TETF_i = \log \frac{Sum(ET)}{1 + Count(ETI_i)}$
  - Trigger Rate： $TR_{ij} = TCF_{ij} * TETF_i$

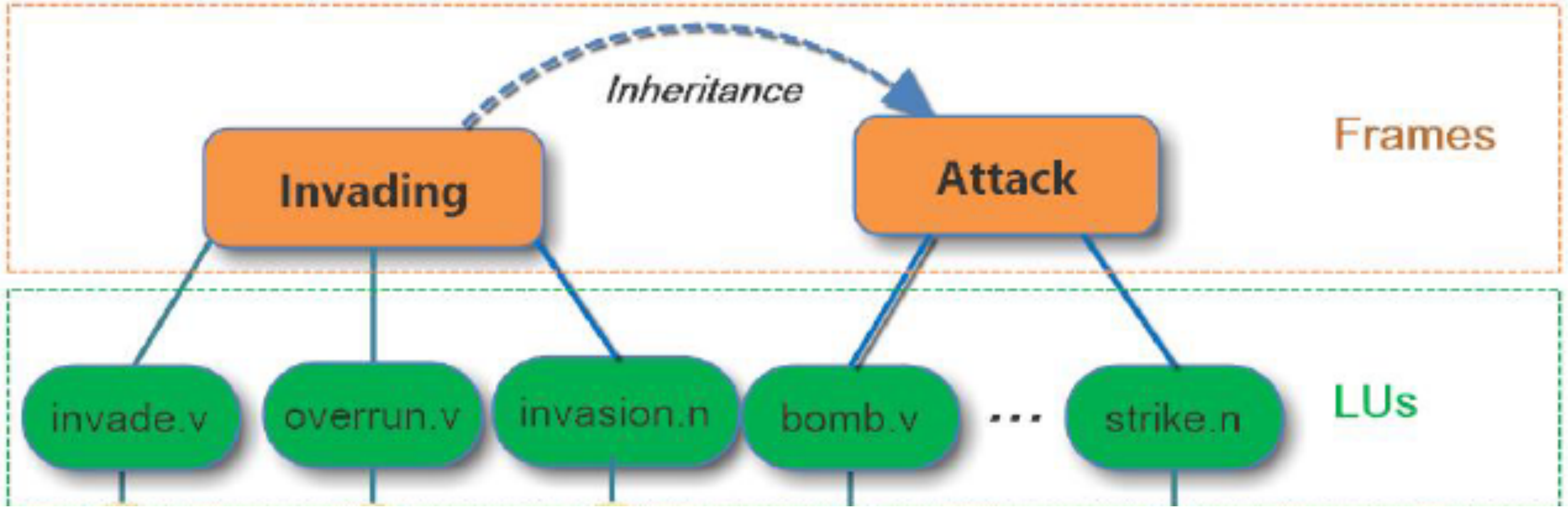
we choose verbs with high TR values as the trigger words for each event type

Our Method for Generating Labeled Data in Event Extraction

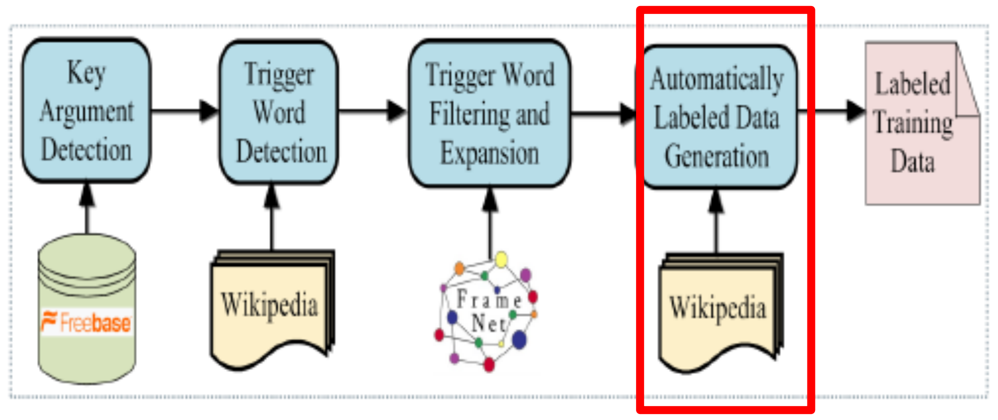


- Trigger Word Filtering and Expansion
  - We propose to use linguistic resource FrameNet to filter noisy verbal triggers and expand nominal triggers

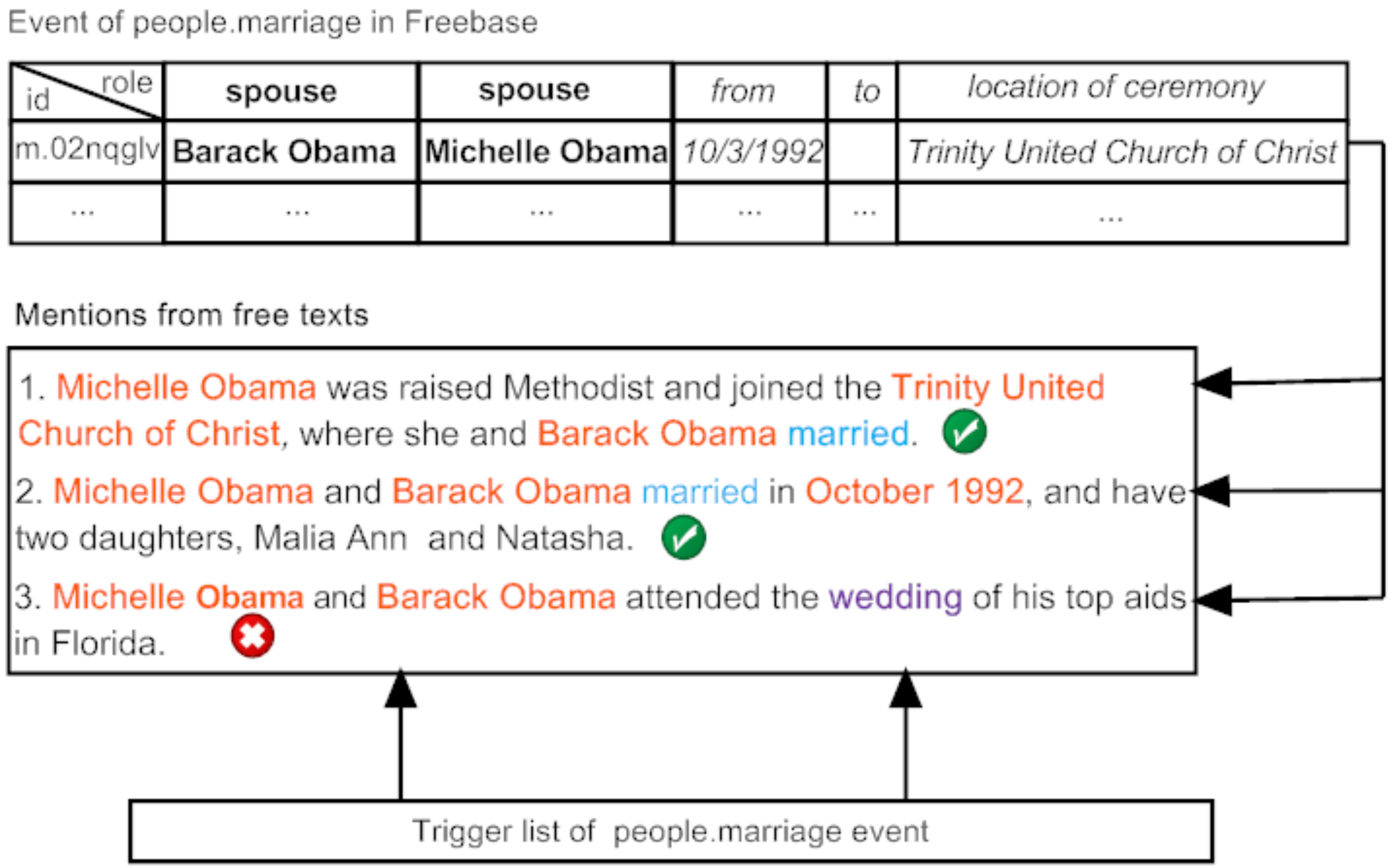
$$frame(i) = \arg\max_j(similarity(e_i, e_{j,k}))$$



Our Method for Generating Labeled Data in Event Extraction

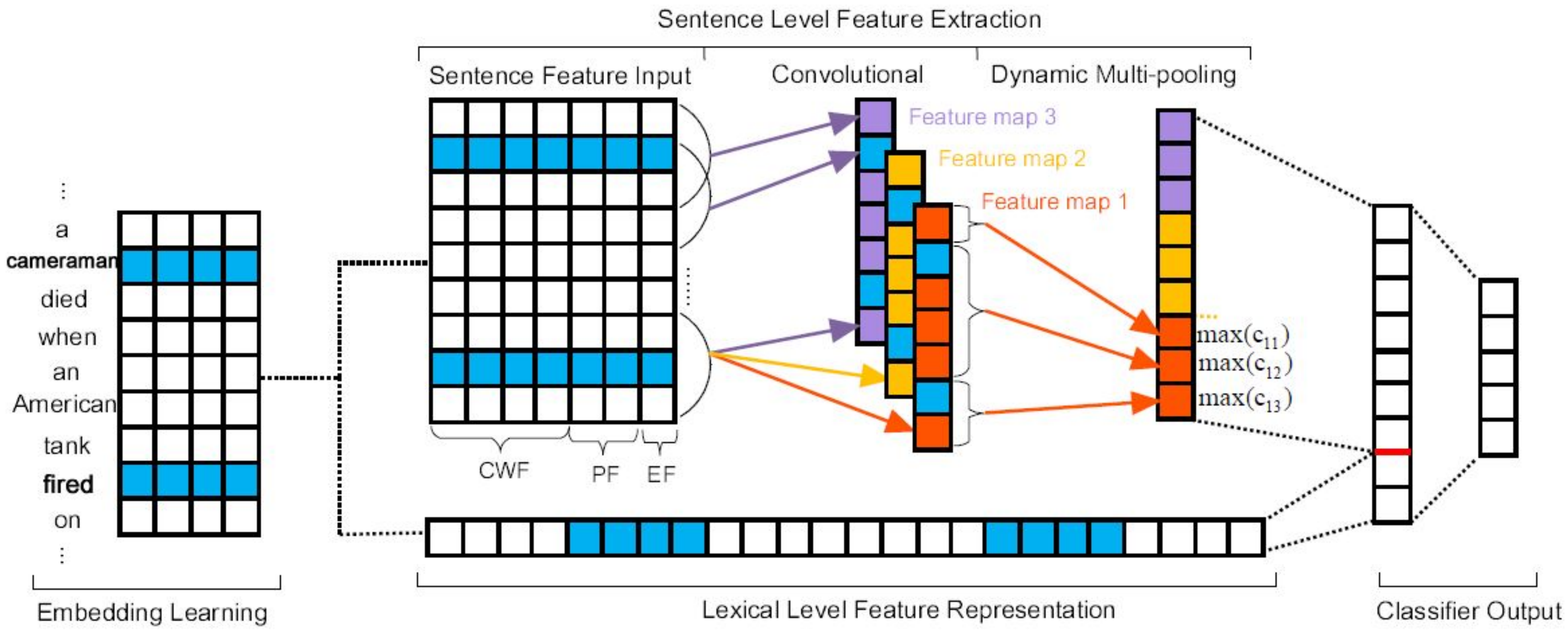


- Automatically labeled data generation
  - We propose a Soft Distant Supervision and use it to automatically generate training data



Neural Network of Event Extraction

- DMCNN in ACL-2015



Employing Multi-Instance Learning to alleviate Noisy

- Training Method: Multi-Instance Learning
  - In stage of argument classification, we take sentences containing the same argument candidate and triggers with a same event type as a bag and all instances in a bag are considered independently.

Algorithm 1 Multi-instance learning

1: Initialize θ. Partition the bags into mini-batches of size b_s.
2: Randomly choose a mini-batch, and feed the bags into the network one by one.
3: Find the j-th instance m_i^j $(1 \leq i \leq b_s)$ in each bag according to Eq. (10).
4: Update θ based on the gradients of m_i^j $(1 \leq i \leq b_s)$ via Adadelta.
5: Repeat steps 2-4 until either convergence or the maximum number of epochs is reached.

Experiments

- Generated Labeled Data

Event Type	Freebase Size	Sentences (KA)	Sentences (KA+T)	Examples of argument roles sorted by KR	Examples of triggers
people.marriage	152,276	56,837	26,349	spouse, spouse, from, to, location	marriage, marry, wed, wedding, couple,..., wife
music.group_membership	239,813	90,617	20,742	group, member, start, role, end	musician, singer, sing, sang, sung, concert,..., play
education.education	530,538	26,966	11,849	student, institution, degree,..., minor	educate, education, graduate, learn, study,..., student
organization.leadership	43,610	5,429	3,416	organization, person, title,..., to	CEO, charge, administer, govern, rule, boss,..., chair
olympics.olympic_medal_honor	20,790	4,056	2,605	medalist, olympics, event,..., country	win, winner, tie, victor, gold, silver,..., bronze
...	...	...	...	...	...
sum of 21 selected events	3,870,492	421,602	72,611	argument1, argument2 ,..., argumentN	trigger1, trigger2, trigger3, ... , triggerN

- Manual Evaluations of Labeled Data

##001 He is the uncle of [**Amal Clooney**], [wife] of the actor [**George Clooney**].
Trigger: wife **Event Type:** Marriage **MannalAnotate[Y/N]:**
Argument: Amal Clooney **Role:**Spouse **MannalAnotate[Y/N]:**
Argument: George Clooney **Role:**Spouse **MannalAnotate[Y/N]:**

##002 She was [married] to the cinematographer [**Theo Nischwitz**] and was sometimes credited as [**Gertrud Hinz-Nischwitz**].
Trigger: married **Event Type:** Marriage **MannalAnotate[Y/N]:**
Argument: Theo Nischwitz **Role:**Spouse **MannalAnotate[Y/N]:**
Argument: Gertrud Hinz-Nischwitz **Role:**Spouse **MannalAnotate[Y/N]:**

Stage	Average Precision
Trigger Labeling	88.9
Argument Labeling	85.4

Results

- Main Results

Methods	Trigger Identification(%)			Trigger Identification + Classification(%)			Argument Identification(%)			Argument Role(%)		
	P	R	F	P	R	F	P	R	F	P	R	F
Li's structure trained with ACE	76.9	65.0	70.4	73.7	62.3	67.5	69.8	47.9	56.8	64.7	44.4	52.7
Chen's DMCNN trained with ACE	80.4	67.7	73.5	75.6	63.6	69.1	68.8	51.9	59.1	62.2	46.9	53.5
Nguyen's JRNN trained with ACE	68.5	75.7	71.9	66.0	73.0	69.3	61.4	64.2	62.8	54.2	56.7	55.4
DMCNN trained with ED Only	77.6	67.7	72.3	72.9	63.7	68.0	64.9	51.7	57.6	58.7	46.7	52.0
DMCNN trained with ACE+ED	79.7	69.6	**74.3**	75.7	66.0	**70.5**	71.4	56.9	**63.3**	62.8	50.1	**55.7**

Summary

- Open Domain Event Extraction
  - Exploiting neural network for feature representation
  - Employing existing KB (Freebase) to automatically generating training data
  - This strategy could be expand to many domains and event types

Thanks for your attention

Questions?