Cross-lingual Predicate Cluster Acquisition to Improve Bilingual Event Extraction by Inductive Learning

Heng Ji
hengji@cs.qc.cuny.edu
Computer Science Department
Queens College and the Graduate Center
City University of New York
June 5, 2009

‘Tricky Word’ Dilemma

This statement is SPECULATING, we’d better remove it.

What did you say? What’s the meaning of ‘SPECULATE’?☹

I mean the statement is CONJECTURING

What does ‘CONJECTURE’ mean?☹

I mean the statement is GUESSING

Ah-huh, I know what you mean☺

BLENDER
Cross-lingual Cross-document IE Lab

Outline

- Task Definition
- Approach Overview
- Cross-lingual Predicate Cluster Acquisition
  - From Bilingual Parallel Corpora
  - From Cross-lingual IE
- Using Predicate Clusters for Bi-lingual Event Extraction
  - Motivation
  - Baseline
  - Inductive Learning
- Experimental Results
- Related Work
- Conclusion and Future Work

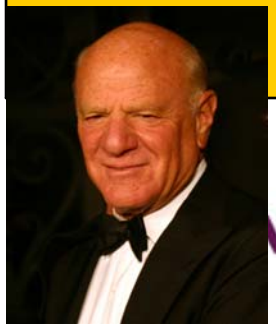
BLENDER
Cross-lingual Cross-document IE Lab

Event Extraction: 'Traditional' IE

Barry Diller Wednesday ***quit*** as chief of Viver **Vivendi Universal Entertainment**

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

Bottleneck: Low Trigger Recall

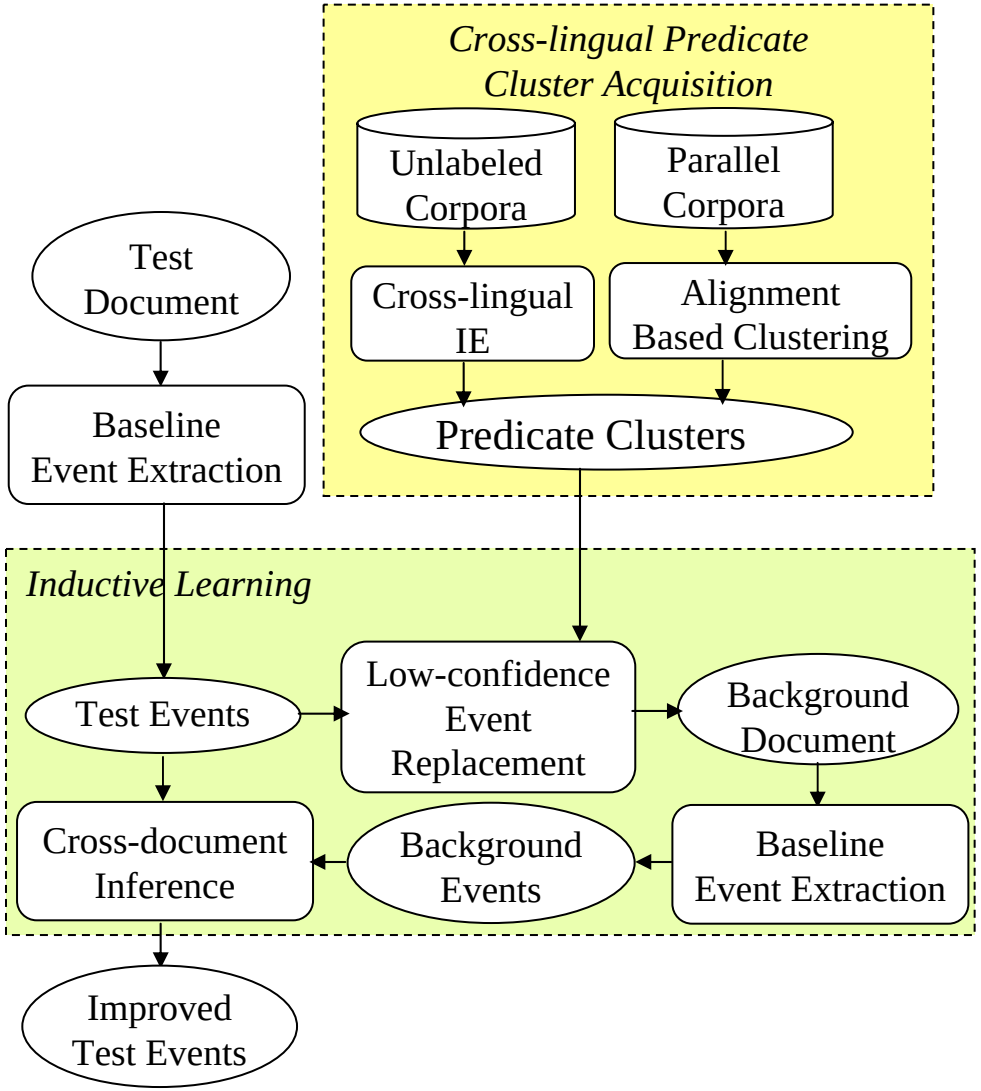


vivendi

- Target: 33 different types of Automatic Content Extraction (ACE) events

BLENDER
Cross-lingual Cross-document IE Lab

Approach Overview



BLENDER
Cross-lingual Cross-document IE Lab

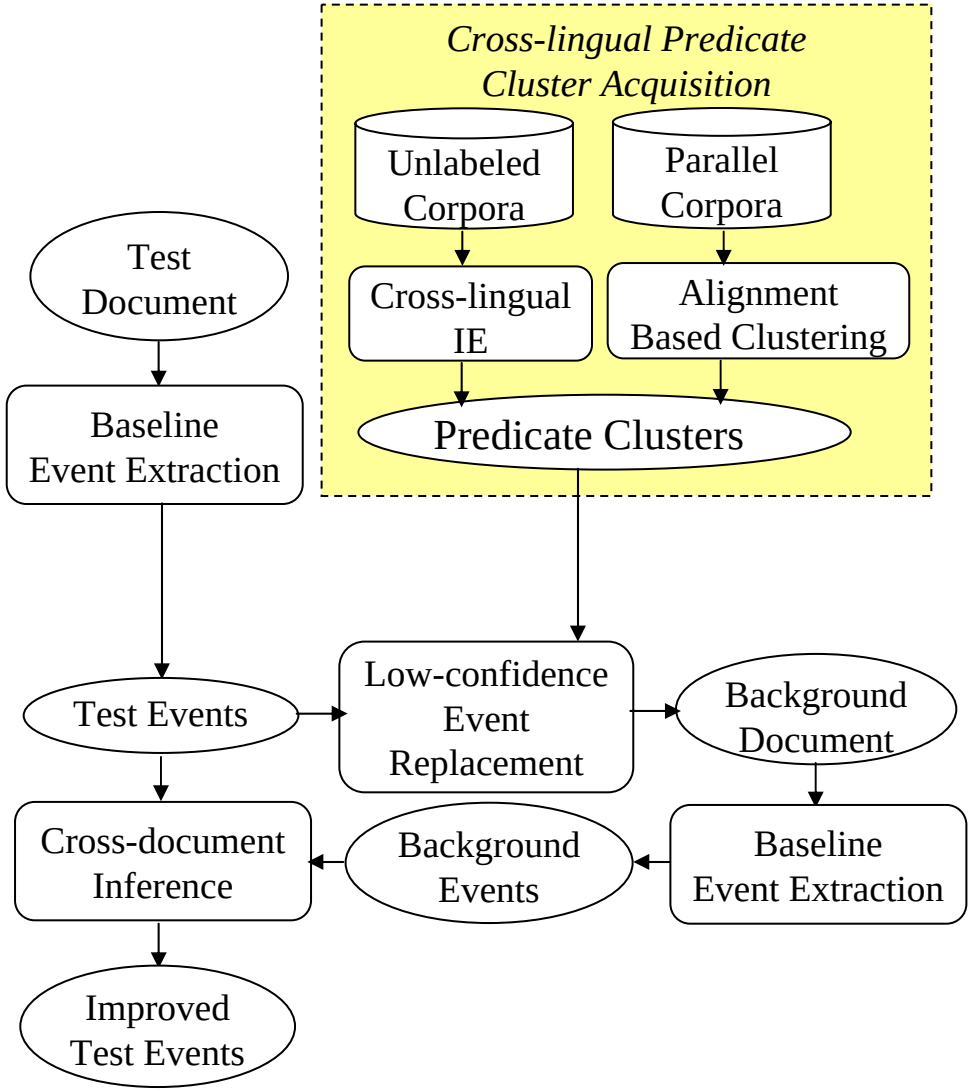
Baseline Bi-lingual Event Extraction System

- Pattern matching
  - Build a pattern from each ACE training example of an event
    - British and US forces reported gains in the advance on Baghdad
      → PER report gain in advance on LOC
- MaxEnt models
  - Trigger Labeling
    - to distinguish event instances from non-events, to classify event instances by type
  - Argument Identification
    - to distinguish arguments from non-arguments
  - Argument Classification
    - to classify arguments by argument role
  - Reportable-Event Classifier
    - to determine whether there is a reportable event instance

(Ji and Grishman, 2008; Chen and Ji, 2009)

BLENDER
Cross-lingual Cross-document IE Lab

Approach Overview



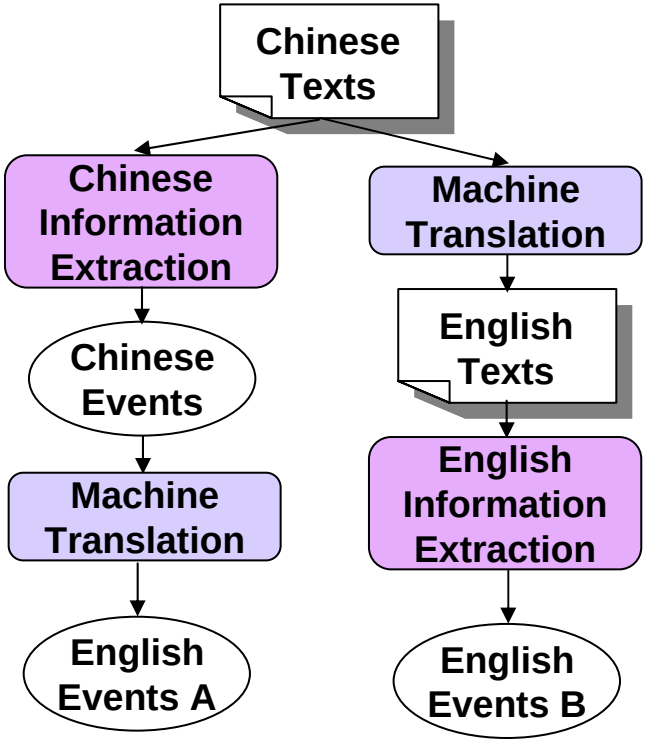
BLENDER
Cross-lingual Cross-document IE Lab

Acquisition from Bilingual Parallel Corpora

- Bootstrapping
  - Anchor set: 852 Chinese event trigger words in ACE05 training corpora
  - For each Chinese trigger, search its automatically aligned English words from a parallel corpus including 50,000 sentence pairs; In each cluster record the frequency of each unique English word
  - Conduct the same procedure in the other direction to construct Chinese predicate clusters anchored by English triggers
- Filtering
  - State-of-the-art Chinese-English word alignment error rate is about 40%
  - Filter the clusters by only keeping those predicates including the original predicate forms in ACE or English/Chinese Propbank

BLENDER
Cross-lingual Cross-document IE Lab

Acquisition from Cross-lingual IE

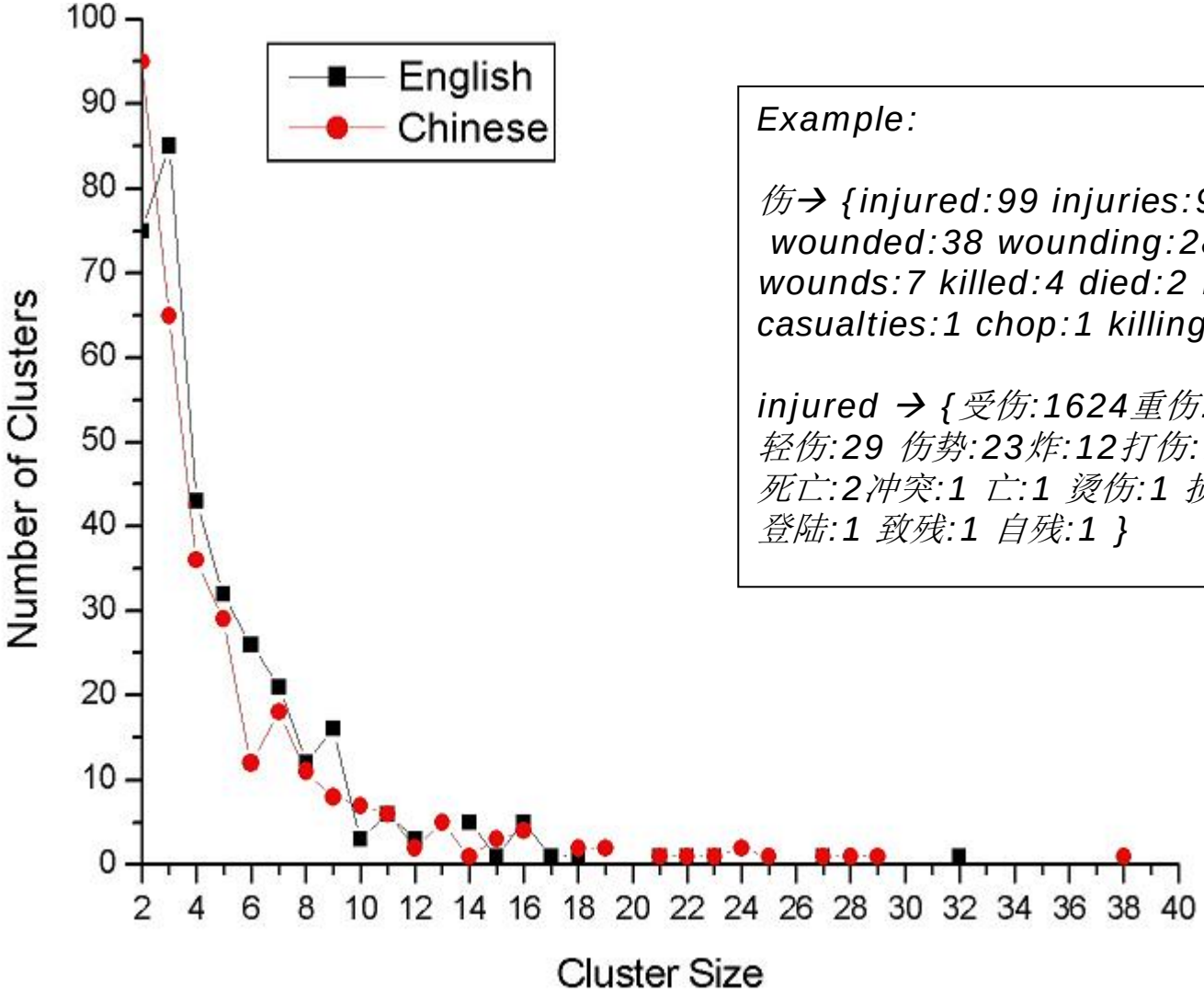


(Zens and Ney, 2004)

- For any Chinese trigger *ch-trigger*, if it has the same translation *en-trigger* in A and B, add *en-trigger* into the cluster anchored by *ch-trigger bv*

BLENDER
Cross-lingual Cross-document IE Lab

Derived Cross-lingual Clusters



Example:

伤→ {injured:99 injuries:96 injury:76 wounded:38 wounding:28 injuring:14 wounds:7 killed:4 died:2 mutilated:1 casualties:1 chop:1 killing:1 shot:1}

injured → {受伤:1624重伤:102伤:99 轻伤:29 伤势:23炸:12打伤:10爆炸:6 伤害:3 死亡:2冲突:1 亡:1 烫伤:1 损失:1 出席:1 登陆:1 致残:1 自残:1 }

BLENDER
Cross-lingual Cross-document IE Lab

Motivation: Improve Rare Trigger Labeling

- Test Sentence:

  Identified as "Conflict-Attack" Event with Confidence=0:

  He told AFP that Israeli intelligence had been dealing with at least 40 tip-offs of impending attacks when the Haifa bus was blown up.

- Trigger Cluster

  *炸毁 → { **blown up**:4 bombing:3 blew:2 destroying:1 destroyed:1 }*

- Replaced Sentences

  Identified as "Conflict-Attack" Event with Confidence=0.799:

  *He told AFP that Israeli intelligence had been dealing with at least 40 tip-offs of impending attacks when the Haifa bus was **destroyed**.*

  ...

BLENDER
Cross-lingual Cross-document IE Lab

Improve Rare Trigger Labeling (Cont')

- Test Sentence:

  Identified as "Justice-Release_Parole" Event with Confidence=0:

  这名嫌犯因为侵害案件假释出狱却又犯下了重罪. 。 (This suspect was released because of the violation case but committed a felony again.)

- Trigger Cluster

  releasing → {假释:4 释放:1 }

- Replaced Sentences

  Identified as "Justice-Release_Parole" Event with Confidence=0.964:

  这名嫌犯因为侵害案件释放出狱却又犯下了重罪

BLENDER
Cross-lingual Cross-document IE Lab

Improve Frequent Trigger Labeling

- Test Sentence:

  Identified as "Personnel-End_Position" Event with Confidence=0:

  *Barry Diller on Wednesday **step** from chief of Vivendi Universal Entertainment, the entertainment unit of French giant Vivendi Universal.*

- Trigger Cluster

  *下台 → { resign:6 **step**:5 **quit**:3}*

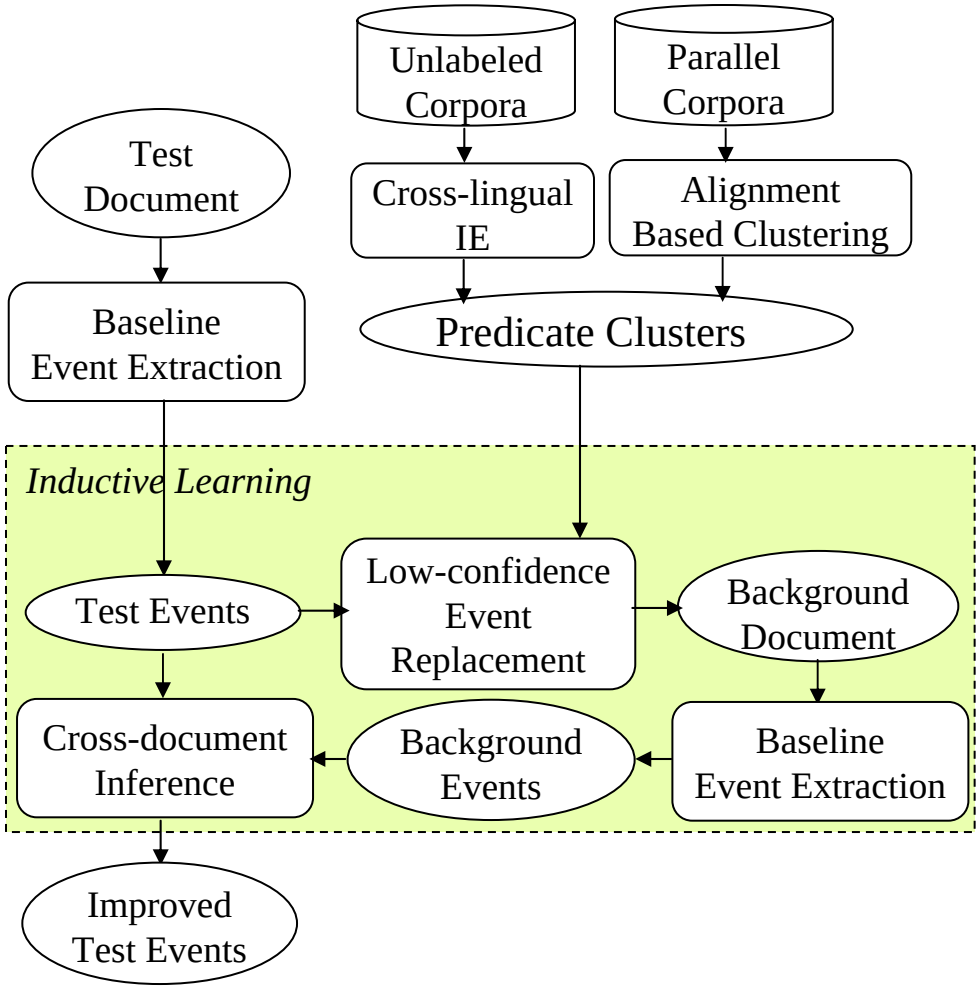
- Replaced Sentences

  Identified as "Personnel-End_Position" Event with Confidence=0.564:

  *Barry Diller on Wednesday **quit** from chief of Vivendi Universal Entertainment, the entertainment unit of French giant Vivendi Universal.*

BLENDER
Cross-lingual Cross-document IE Lab

Approach Overview



BLENDER
Cross-lingual Cross-document IE Lab

Inductive Learning: Background Document Generation

- For each event mention in a test document, the baseline event tagger produces confidence value:

  LConf(trigger, etype): The probability of a string *trigger* indicating an event mention with type *etype* in a context sentence *S*

- If *LConf(trigger, etype)* is lower than a threshold, and it belongs to a predicate cluster *C*, for each *predicatei* $\in$ *C*, replace *trigger* with *predicatei* in *S* to generate new sentence *S'*, and add *S'* into *background document*

BLENDER
Cross-lingual Cross-document IE Lab

Inductive Learning: Cross-document Inference

- Compose a test document and its background documents as a topically-related set
- Hypotheses
  - One Trigger Sense Per Set
  - One Argument Role Per Set
- Document-wide and Set-wide Cluster Confidence
  - Frequency weighted by local confidence
  - Count frequency of trigger in each cluster with a particular event type
  - For each argument and its coreferred names, count frequency of event type
  - For each argument and its coreferred names, count frequency of event type and role

BLENDER
Cross-lingual Cross-document IE Lab

Cross-document Inference (Cont')

- Inference Actions
  - Aggregate similar events across documents and conduct statistical global inference
  - Remove triggers and argument annotations with local or cross-document cluster confidence lower than thresholds
  - Propagate highly consistent and frequent triggers and arguments with high global cluster confidence to override other, lower confidence, extraction results

BLENDER
Cross-lingual Cross-document IE Lab

Experiments: Data and Scoring Metric

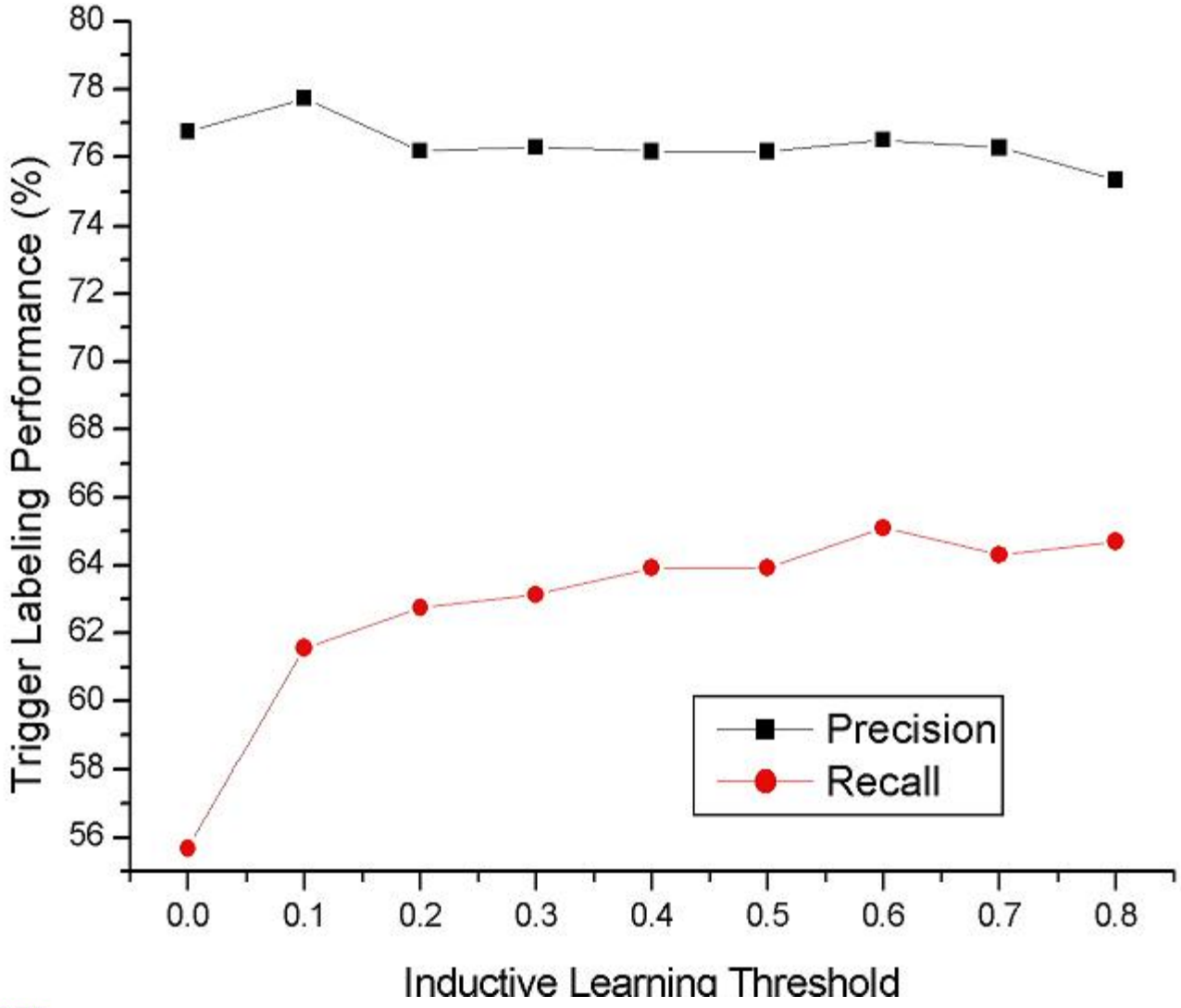
- Data: ACE 2005 training corpora (number of documents)

Language	Training Set	Development Set	Test Set
English	525	33	66
Chinese	500	10	40

- Scoring
  - *A trigger is correctly labeled* if its event type and position in the document match a reference trigger.
  - *An argument is correctly identified* if its event type and position in the document match any of the reference argument mentions.
  - *An argument is correctly identified and classified* if its event type, position in the document, and role match any of the reference argument mentions.

BLENDER
Cross-lingual Cross-document IE Lab

Confidence Metric Thresholding on Dev set



BLENDER
Cross-lingual Cross-document IE Lab

Trigger Labeling Performance

Language/System \ Performance		Precision	Recall	F-Measure
English	Baseline	67.8	53.5	59.8
	After Using Cross-lingual Predicate Clusters	69.2	59.4	63.9
Chinese	Baseline	58.1	47.2	52.1
	After Using Cross-lingual Predicate Clusters	60.2	52.6	56.1

BLENDER
Cross-lingual Cross-document IE Lab

Argument Labeling Performance

Language/System \ Performance		Argument Identification			Argument Classification Accuracy	Argument Identification +Classification		
		P	**R**	**F**		**P**	**R**	**F**
English	Baseline	49.3	31.4	38.3	88.2	43.5	27.7	33.9
	After Using Cross-lingual Predicate Clusters	51.7	32.7	40.1	89.6	46.3	29.3	35.9
Chinese	Baseline	46.2	33.7	39.0	95.0	43.9	32.0	37.0
	After Using Cross-lingual Predicate Clusters	46.8	36.7	41.1	95.6	44.7	35.1	39.3

BLENDER
Cross-lingual Cross-document IE Lab

Discussion: Alternatives

- Self-training and Bootstrapping
  - Add high-confidence events in the test set back as additional training data and re-train the event tagger → produced 1.7% worse F-measure score for English d
  - Using the test set itself is not enough, need to explore new predicates to serve as background evidence
- Bootstrapping using relevant unlabeled data
  - Obtained limited improvement – about 1.6% F-measure gain for English.
  - Both self-training and bootstrapping methods require good data selection scheme. But not for any test set we can easily find relevant unlabeled data.
  - Our approach is less expensive – we can automatically generate background data while introducing new evidence
- Encoding the cluster membership as an additional feature in the supervised-learning procedure of the baseline event tagger
  - Similar to (Miller et al., 2004) and (Lin et al., 2009) on name tagging
  - Requires access the algorithms of the baseline system

BLENDER
Cross-lingual Cross-document IE Lab

Related Work

- Paraphrase and Word Cluster Discovery
  - From mono-lingual parallel corpora (e.g. Barzilay and McKeown, 2001; Lin and Pantel, 2001; Ibrahim et al., 2003; Pang et al., 2003; Shinyama and Sekine, 2003)
  - From cross-lingual parallel corpora (e.g. Bannard and Callison-Burch, 2005; Callison-Burch, 2008)
- Use Paraphrase and Word Cluster to Improve IE
  - Improve English Name tagging (Miller et al., 2004; Lin et al., 2009)
- Use Cluster for Local Replacement
  - Use local context replacement for pronoun resolution (Ge et al., 1998)

BLENDER
Cross-lingual Cross-document IE Lab

Conclusion and Future Work

- Described two approaches to extract cross-lingual predicate clusters
- Designed a new inductive learning framework to effectively incorporate these clusters for event extraction
- Without using any additional data or changing the baseline algorithms, we demonstrated that this method can significantly enhance the performance of a state-of-the-art bilingual event tagger.

- Future Work
  - In the future we will attempt incorporating POS tagging results and frequency information for filtering
  - Acquire clusters from very large corpora (e.g. google ngram data) (JHU summer worshop09)
  - Extract cross-lingual relation and name clusters to improve other IE tasks
  - Automatically discover new event types (non-ACE event types) or more fine-grained subtypes/attributes for existing ACE event types from the derived predicate clusters
  - Use manually generated clusters such as VerbNet (Kipper et al., 2006)

BLENDER
Cross-lingual Cross-document IE Lab

BLENDER

Cross-lingual Cross-document IE Lab

Thank you