Information Extraction

Lecture 9 – (Übung and) Event Extraction

CIS, LMU München
Winter Semester 2015-2016

Dr. Alexander Fraser, CIS

Administravia

- Seminar:
- Hausarbeit is due 3 weeks after your presentation
- However, Xmas break (24th to 6th) does not count towards your three weeks
  - Add two more weeks if your working period touches these dates

- Today we will start with a quick review of the Übung to make sure you have the key concepts
- If you are one of the few people who are not in the Seminar
  - You will still be able to follow what I am discussing
  - You can try doing the Übung (Exercise2) by simply going to the Seminar web page and downloading the relevant materials

Review of Übung

- In the Übung last week, we used the open source machine learning package Wapiti
- We worked on a binary learning task: finding <stime> tags
- We looked at:
    - Basic setup (compiling Wapiti, create sa-tagged directory) – "make prep"
    - How to run experiments (train, development, test) – "make"
    - Basic feature extraction code "extract_003.pl"
    - Wapiti pattern files "unigram_bigram_pattern.txt"

How to run experiments (train, development, test)

- Ideally you should run shell scripts like this:

  bash myscript.sh >& myscript.sh.log

- This saves the output into a log file (I always do this, and none of my scripts take parameters)
  - Even better would be to have the extractor print version numbers (and maybe use source control)

Basic feature extraction code

- We looked at extract_003.pl
- This extracts a raw representation which I sometimes refer to as the "features", but which should really be referred to differently
  - Let's call what this outputs the extract file
- The extract file is used to build the actual features used by Wapiti (and contains the gold-standard labels for training data or test data where we want Wapiti to calculate precision/recall and F)

Wapiti pattern files

- Wapiti pattern files are a level of indirection that allow us to:

  1) specify whether a column in the extract file is used

  - This is useful to "comment out" features in the extract file
  - Otherwise it is annoying – you have to remember to explicitly enable each new column as a feature

  2) create features that combine columns (so-called "compound" features)

  - Two features put together is often called a bigram

Beyond binary classification

- Wapiti supports multi-class classification
- You can just change the label in the last column in the "extract" file to any string
- Then retrain
- Very abstractly, it is doing something like one-against-all as I explained in class
  - The details are more complicated, in fact it is a multi-class maximum entropy model
  - I will skip the details (at least for now)

Sequence classification

- There is also a script that does sequence classification
- When using sequence classification, you have several rows like in the extract file
  - But without blank lines between them
  - This is a sequence
- You define a special feature which says "look at the previous label" (this feature starts with the letter "b" in the Wapiti pattern file, because it is defining a feature on the previous label and current label, which is a *label* bigram feature)
- You'll notice that the extract is much simpler, because we can refer to the word in the previous example, or the word in the next example (instead of including these as columns as we did previously)
- We will look at sequence classification in a further lab after the break

Conclusion

- Wapiti is a very interesting package for multi-class and sequential multi-class classification
- It is also quite easy to use
  - Except the annoying bug that we engineered around (where we added a single letter to very simple features like "isUpper" or "isNotUpper")
- Read the manual to see what it can do
- A further detail for avoiding overfitting the training corpus is a technique called "regularization"
  - See the Wapiti paper (cited on the website) for more about this

Event Extraction

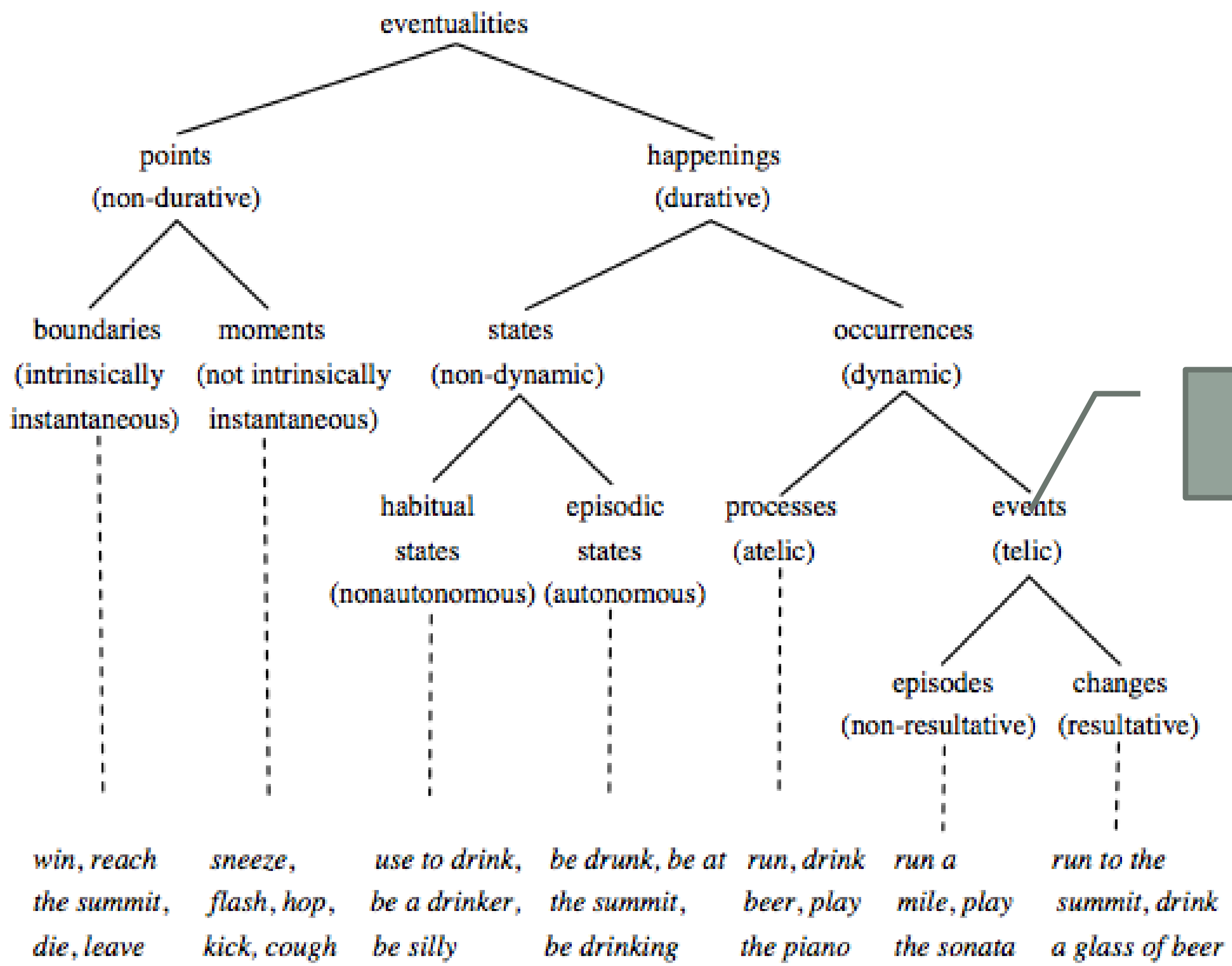
- We'll now discuss event extraction, as defined in state-of-the-art statistical systems

Outline

- Event Definition
- Event Knowledge Networks Construction
- Basic Event Extraction Approach
- Advanced Event Extraction Approaches
  - Information Redundancy for Inference
  - Co-training

General Event Definition

- An Event is a specific occurrence involving participants.
- An Event is something that happens.
- An Event can frequently be described as a change of state.



Most of current NLP work focused on this

Chart from (Dölling, 2011)

Event Extraction

- Task Definition
- Basic Event Extraction Approach
- Advanced Event Extraction Approaches
  - Information Redundancy for Inference
  - Co-training

Event Mention Extraction: Task

- An event is specific occurrence that implies a change of states
- **event trigger**: the main word which most clearly expresses an event occurrence
- **event arguments**: the mentions that are involved in an event (participants)
- **event mention**: a phrase or sentence within which an event is described, including trigger and arguments
- Automatic Content Extraction defined 8 types of events, with 33 subtypes

Argument, role=victim *trigger*

ACE event type/subtype	*Event Mention Example*
Life/Die	**Kurt Schork died** in **Sierra Leone yesterday**
Transaction/Transfer	**GM sold the company** in **Nov 1998** to **LLC**
Movement/Transport	**Homeless people** have been **moved** to **schools**
Business/Start-Org	**Schweitzer founded a hospital** in **1913**
Conflict/Attack	the **attack** on **Gaza** killed **13**
Contact/Meet	**Arafat's cabinet met** for **4 hours**
Personnel/Start-Position	**She** later **recruited the nursing student**
Justice/Arrest	**Faison** was wrongly **arrested** on suspicion of murder

Slide from Heng Ji

Supervised Event Mention Extraction: Methods

- Staged classifiers
  - Trigger Classifier
    - to distinguish event instances from non-events, to classify event instances by type
  - Argument Classifier
    - to distinguish arguments from non-arguments
  - Role Classifier
    - to classify arguments by argument role
  - Reportable-Event Classifier
    - to determine whether there is a reportable event instance
- Can choose any supervised learning methods such as MaxEnt and SVMs

(Ji and Grishman, 2008)

Slide from Heng Ji

Typical Event Mention Extraction Features

- Trigger Labeling
  - Lexical
    - Tokens and POS tags of candidate trigger and context words
  - Dictionaries
    - Trigger list, synonym gazetteers
  - Syntactic
    - the depth of the trigger in the parse tree
    - the path from the node of the trigger to the root in the parse tree
    - the phrase structure expanded by the parent node of the trigger
    - the phrase type of the trigger
  - Entity
    - the entity type of the syntactically nearest entity to the trigger in the parse tree
    - the entity type of the physically nearest entity to the trigger in the sentence

- Argument Labeling
  - Event type and trigger
    - Trigger tokens
    - Event type and subtype
  - Entity
    - Entity type and subtype
    - Head word of the entity mention
  - Context
    - Context words of the argument candidate
  - Syntactic
    - the phrase structure expanding the parent of the trigger
    - the relative position of the entity regarding to the trigger (before or after)
    - the minimal path from the entity to the trigger
    - the shortest length from the entity to the trigger in the parse tree

(Chen and Ji, 2009)

Slide from Heng Ji

Why Trigger Labeling is so Hard?

- DT this “this is the largest pro-troops demonstration that has ever been in San Francisco”
- RP forward “We've had an absolutely terrific story, pushing forward north toward Baghdad”
- WP what “what happened in”
- RB back “his men back to their compound”
- IN over “his tenure at the United Nations is over”
- IN out “the state department is ordering all non-essential diplomats”
- CD nine eleven “nine eleven”
- RB formerly “McCarthy was formerly a top civil servant at”

Slide from Heng Ji

Why Trigger Labeling is so Hard?

- A suicide bomber detonated explosives at the entrance to a crowded
- medical teams carting away dozens of wounded victims
- dozens of Israeli tanks advanced into thenorthern Gaza Strip
- Many nouns such as “death”, “deaths”, “blast”, “injuries” are missing

Slide from Heng Ji

Why Argument Labeling is so Hard?

- Two 13-year-old children were among those killed in the Haifa bus bombing, Israeli public radio said, adding that most of the victims were youngsters
- Israeli forces staged a bloody raid into a refugee camp in central Gaza targeting a founding member of Hamas
- Israel's night-time raid in Gaza involving around 40 tanks and armoured vehicles
- Eight people, including a pregnant woman and a 13-year-old child were killed in Monday's Gaza raid
- At least 19 people were killed and 114 people were wounded in Tuesday's southern Philippines airport

Slide modified from Heng Ji

Why Argument Labeling is so Hard?

- Two 13-year-old children were among those killed in the Haifa bus bombing, Israeli public radio said, adding that most of the victims were youngsters
- Fifteen people were killed and more than 30 wounded Wednesday as a suicide bomber blew himself up on a student bus in the northern town of Haifa
- Two 13-year-old children were among those killed in the Haifa bus bombing

Slide from Heng Ji

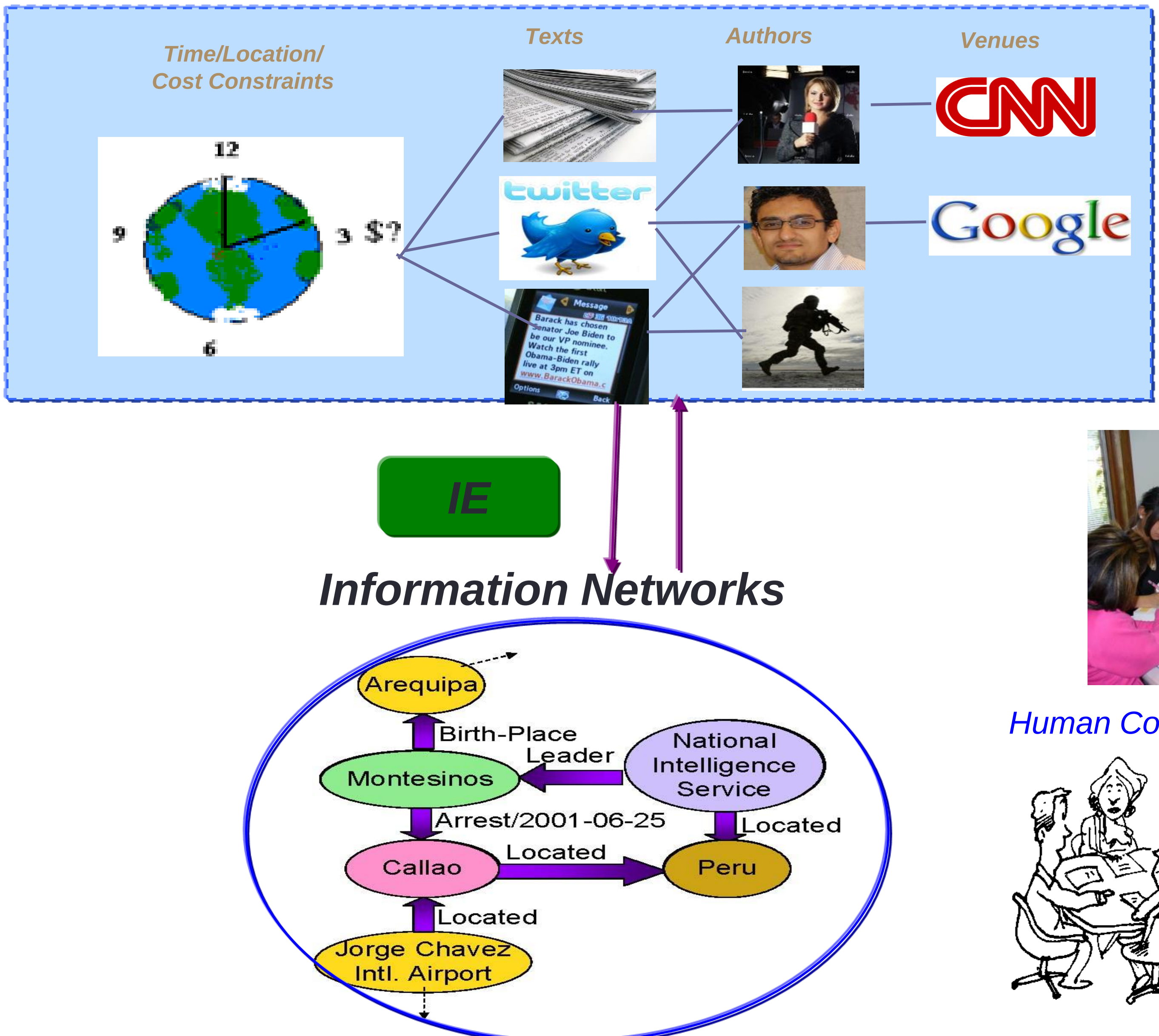
State-of-the-art and Remaining Challenges

- State-of-the-art Performance (F-score)
  - English: Trigger 70%, Argument 45%
  - Chinese: Trigger 68%, Argument 52%
  - Single human annotator: Trigger 72%, Argument 62%
- Remaining Challenges
  - Trigger Identification
    - Generic verbs
    - Support verbs such as "take" and "get" which can only represent an event mention together with other verbs or nouns
    - Nouns and adjectives based triggers
  - Trigger Classification
    - "named" represents a "Personnel_Nominate" or "Personnel_Start-Position"?
    - "hacked to death" represents a "Life_Die" or "Conflict_Attack"?
  - Argument Identification
    - Capture long contexts
  - Argument Classification
    - Capture long contexts
    - Temporal roles

(Ji, 2009; Li et al., 2011)

Slide from Heng Ji

IE in Rich Contexts



Human Collaborative Learning

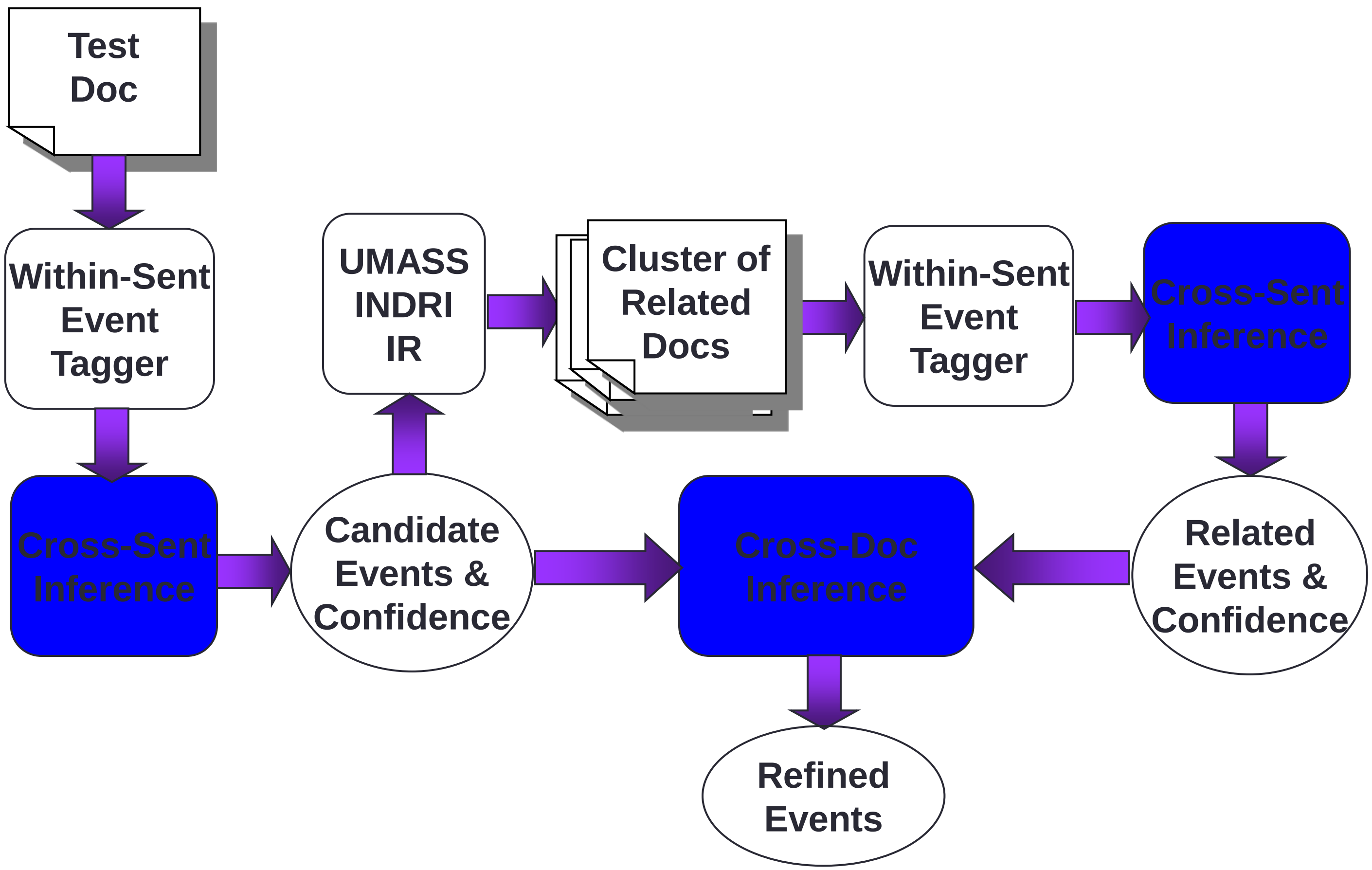
Slide from Heng Ji

Capture Information Redundancy

- When the data grows beyond some certain size, IE task is naturally embedded in rich contexts; the extracted facts become inter-dependent
- Leverage Information Redundancy from:
  - Large Scale Data (Chen and Ji, 2011)
  - Background Knowledge (Chan and Roth, 2010; Rahman and Ng, 2011)
  - Inter-connected facts (Li and Ji, 2011; Li et al., 2011; e.g. Roth and Yih, 2004; Gupta and Ji, 2009; Liao and Grishman, 2010; Hong et al., 2011)
  - Diverse Documents (Downey et al., 2005; Yangarber, 2006; Patwardhan and Riloff, 2009; Mann, 2007; Ji and Grishman, 2008)
  - Diverse Systems (Tamang and Ji, 2011)
  - Diverse Languages (Snover et al., 2011)
  - Diverse Data Modalities (text, image, speech, video…)
- But how? Such knowledge might be overwhelming…

Slide from Heng Ji

Cross-Sent/Cross-Doc Event Inference Architecture



Slide from Heng Ji

Baseline Within-Sentence Event Extraction

1. 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
2. MaxEnt models
   1. Trigger Classifier
      - to distinguish event instances from non-events, to classify event instances by type
   2. Argument Classifier
      - to distinguish arguments from non-arguments
   3. Role Classifier
      - to classify arguments by argument role
   4. Reportable-Event Classifier
      - to determine whether there is a reportable event instance

Slide from Heng Ji

Global Confidence Estimation

- Within-Sentence IE system produces local confidence
- IR engine returns a cluster of related docs for each test doc
- Document-wide and Cluster-wide Confidence
  - Frequency weighted by local confidence
  - *XDoc-Trigger-Freq(trigger, etype):* The weighted frequency of string *trigger* appearing as the trigger of an event of type *etype* across all related documents
  - *XDoc-Arg-Freq(arg, etype):* The weighted frequency of *arg* appearing as an argument of an event of type *etype* across all related documents
  - *XDoc-Role-Freq(arg, etype, role):* The weighted frequency of *arg* appearing as an argument of an event of type *etype* with role *role* across all related documents
  - *Margin* between the most frequent value and the second most frequent value, applied to resolve classification ambiguities
  - ……

Slide from Heng Ji

Cross-Sent/Cross-Doc Event Inference Procedure

- Remove triggers and argument annotations with local or cross-doc confidence lower than thresholds
  - *Local-Remove*: Remove annotations with low local confidence
  - *XDoc-Remove*: Remove annotations with low cross-doc confidence

- Adjust trigger and argument identification and classification to achieve document-wide and cluster-wide consistency
  - *XSent-Iden/XDoc-Iden*: If the highest frequency is larger than a threshold, propagate the most frequent type to all unlabeled candidates with the same strings
  - *XSent-Class/XDoc-Class*: If the margin value is higher than a threshold, propagate the most frequent type and role to replace low-confidence annotations

Slide from Heng Ji

Experiments: Data and Setting

- Within-Sentence baseline IE trained from 500 English ACE05 texts (from March – May of 2003)
- Use 10 ACE05 newswire texts as development set to optimize the global confidence thresholds and apply them for blind test
- Blind test on 40 ACE05 texts, for each test text, retrieved 25 related texts from TDT5 corpus (278,108 texts, from April-Sept. of 2003)

Experiments: Trigger Labeling

System/Human \ Performance	**Precision**	**Recall**	**F-Measure**
Within-Sent IE (Baseline)	67.6	53.5	59.7
After Cross-Sent Inference	64.3	59.4	61.8
After Cross-Doc Inference	60.2	76.4	67.3
Human Annotator 1	59.2	59.4	59.3
Human Annotator 2	69.2	75.0	72.0
Inter-Adjudicator Agreement	83.2	74.8	78.8

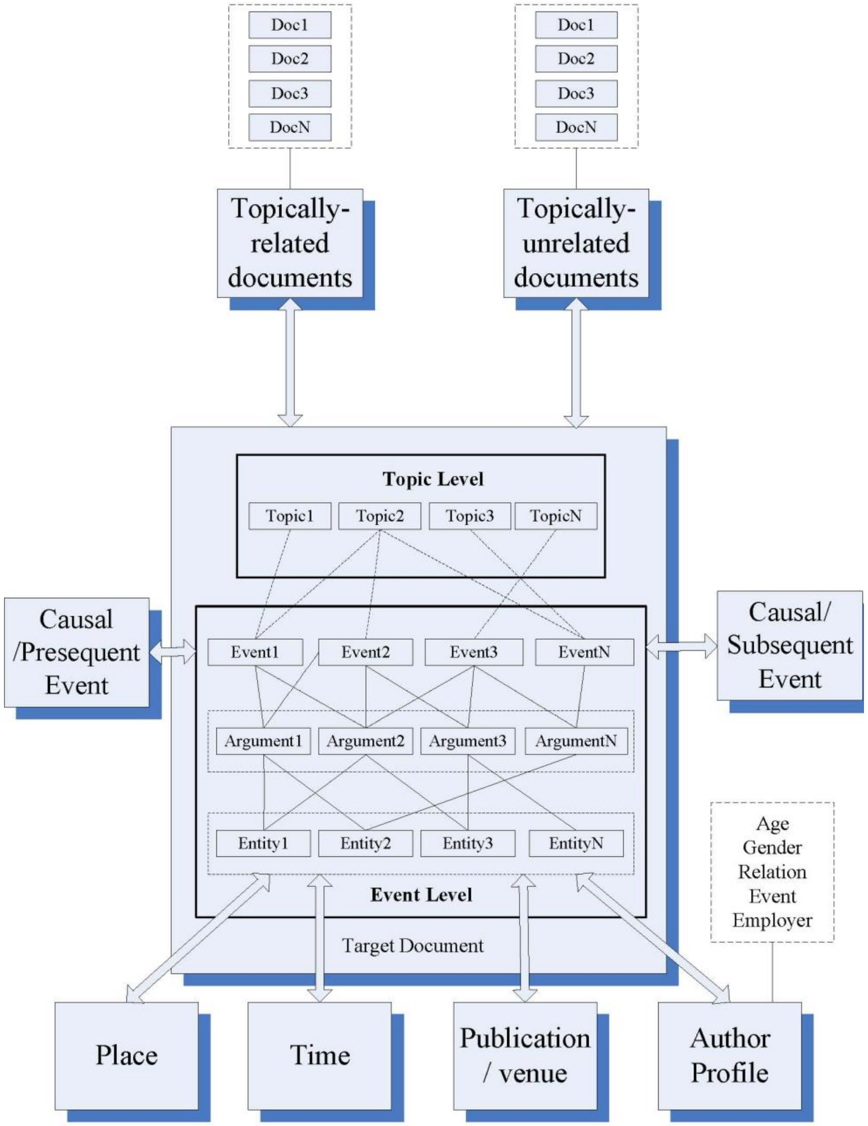
Slide from Heng Ji

Experiments: Argument Labeling

Performance / System/Human	Argument Identification			Argument Classification Accuracy	Argument Identification +Classification		
	P	**R**	**F**		**P**	**R**	**F**
Within-Sent IE	47.8	38.3	42.5	86.0	41.2	32.9	36.3
After Cross-Sent Inference	54.6	38.5	45.1	90.2	49.2	34.7	40.7
After Cross-Doc Inference	55.7	39.5	46.2	92.1	51.3	36.4	42.6
Human Annotator 1	60.0	69.4	64.4	85.8	51.6	59.5	55.3
Human Annotator 2	62.7	85.4	72.3	86.3	54.1	73.7	62.4
Inter-Adjudicator Agreement	72.2	71.4	71.8	91.8	66.3	65.6	65.9

Slide from Heng Ji

Global Knowledge based Inference for Event Extraction



- Cross-document inference (Ji and Grishman, 2008)
- Cross-event inference (Liao and Grishman, 2010)
- Cross-entity inference (Hong et al., 2011)
- All-together (Li et al., 2011)

Slide from Heng Ji

Bootstrapping Event Extraction

- Both systems rely on expensive human labeled data, thus suffers from data scarcity

  (much more expensive than other NLP tasks due to the extra tagging tasks of entities and temporal expressions)

Questions:

- Can the monolingual system benefit from bootstrapping techniques with a relative small set of training data?
- Can a monolingual system (in our case, the Chinese event extraction system) benefit from the other resource-rich monolingual system (English system)?

Slide from Heng Ji

Cross-lingual Co-Training

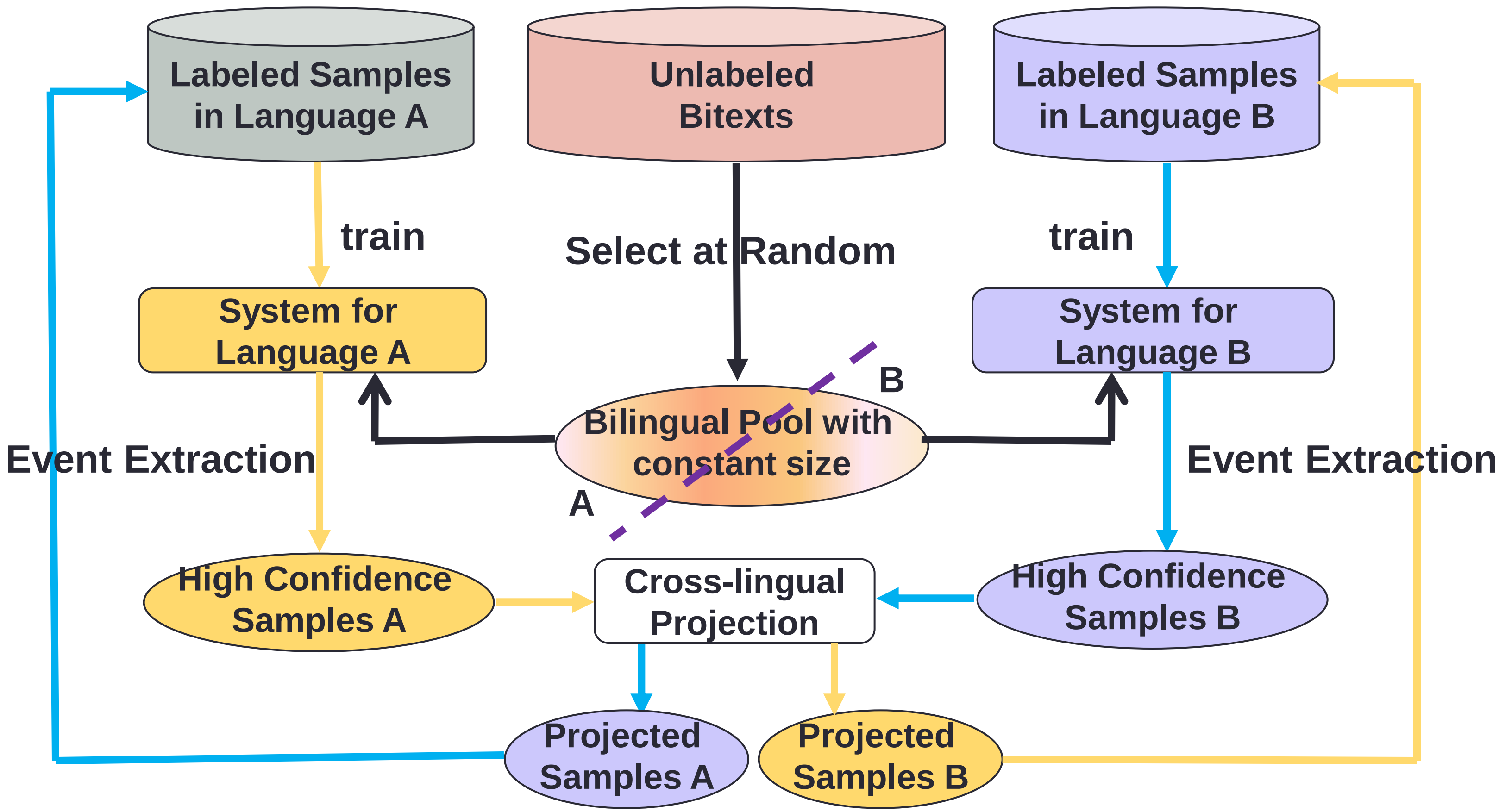
Intuition:

- The same event has different "views" described in different languages, because the lexical unit, the grammar and sentence construction differ from one language to the other.
- Satisfy the sufficiency assumption

Slide from Heng Ji

Cross-lingual Co-Training for Event Extraction

(Chen and Ji, 2009)



- Bootstrapping: n=1: trust yourself and teach yourself
- Co-training: n=2 (Blum and Mitchell,1998)
  - the two views are individually **sufficient** for classification
  - the two views are conditionally **independent** given the class

Slide from Heng Ji

Cross-lingual Projection

- A key operation in the cross-lingual co-training algorithm
- In our case, project the triggers and the arguments from one language into the other language according to the alignment information provided by bitexts.

Slide from Heng Ji

Experiments (Chen and Ji, 2009)

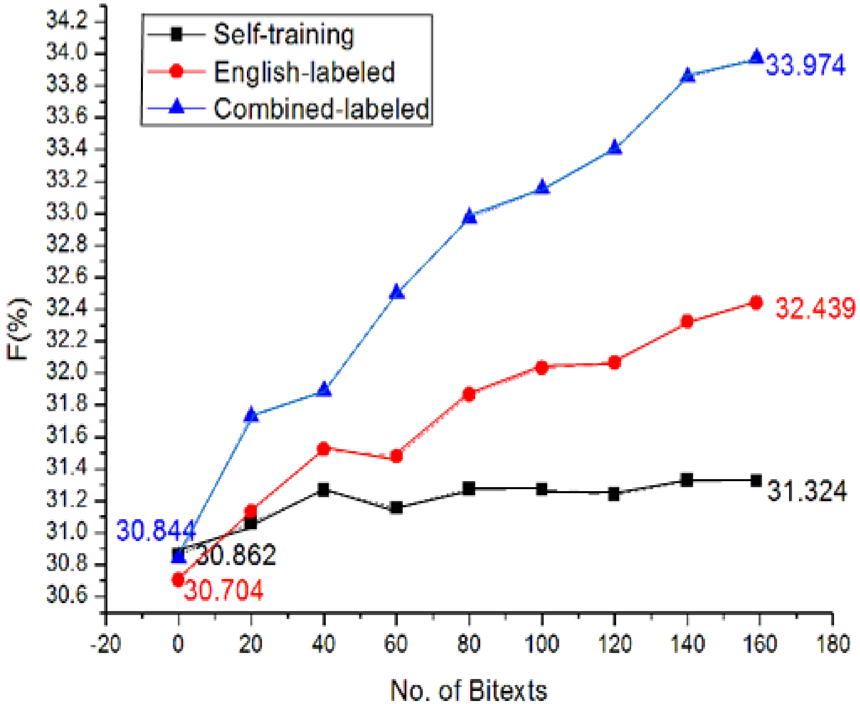
Data

- ACE 2005 corpus
  - 560 English documents
  - 633 Chinese documents

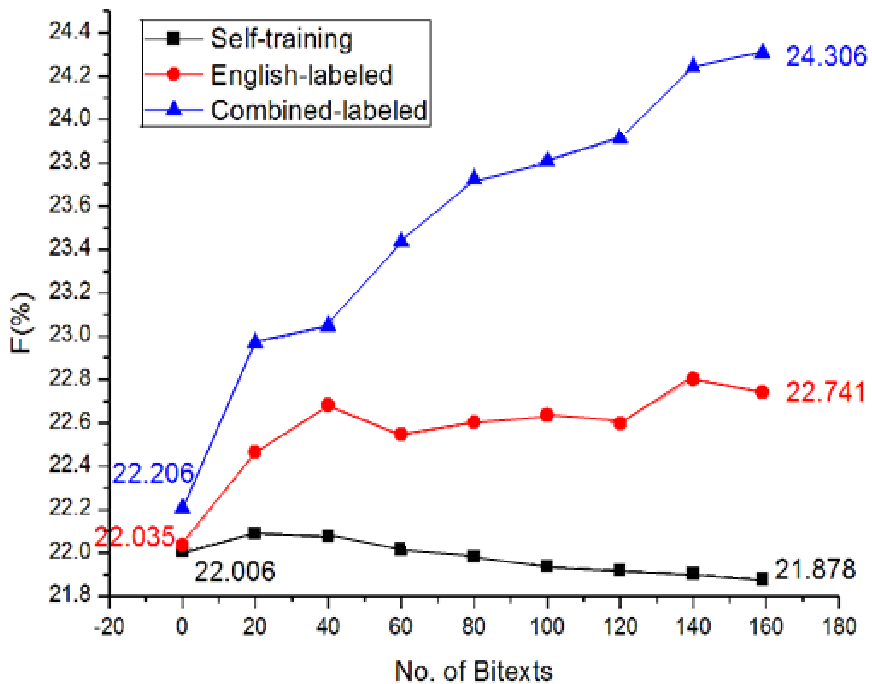
- LDC Chinese Treebank English Parallel corpus
  - 159 bitexts with manual alignment

Slide from Heng Ji

Experiment results



Self-training, and Co-training
(English- labeled & Combined-labeled)
for Trigger Labeling



Self-training, and Co-training
(English- labeled & Combined-labeled)
for Argument Labeling

Slide from Heng Ji

Analysis

- Self-training: a little gain of 0.4% above the baseline for trigger labeling and a loss of 0.1% below the baseline for argument labeling. The deterioration tendency of the self-training curve indicates that entity extraction errors do have counteractive impacts on argument labeling.
- Trust-English method: a gain of 1.7% for trigger labeling and 0.7% for argument labeling.
- Combination method: a gain of 3.1% for trigger labeling and 2.1% for argument labeling.

  The third method outperforms the second method.

Slides

- The slides for event extraction are from Heng Ji, who is a IE researcher at RPI

Summary

- Event extraction is an interesting topic which has recently started to undergo significant changes
  - In these slides we talked about cross-document reference
  - One can go further and include the web and/or ontologies (next lecture)
- It is a very difficult problem but clearly necessary if we want to reason about changes of state, rather than facts that hold over long periods of time

- Thank you for your attention!