Coreference Resolution

Slides are modified from Prof. Claire Cardie's

Plan for the Talk

- Linguistic background for coreference resolution
- supervised machine learning approach
- weakly supervised approaches

Reference resolution

- **Reference**: the process by which speakers use expressions like "John Simon" and "his" to denote a real-world entity
  - **Referring expressions**: NL expression used to perform reference
  - **Referent**: the entity that is referred to
  - **Shorthand form**: *his* refers to John Simon

?

John Simon, Chief Financial Officer of Prime Corp. since 1986, saw his pay jump 20%, to $1.3 million, as the 37-year-old also became the financial-services company's president...

Coreference

- **Coreference:** two referring expressions that are used to refer to the same entity are said to corefer
- *John Simon* is the **antecedent** of *his*.
- Reference to an entity that has been previously introduced into the discourse is called **anaphora**; and the referring expression used is said to be **anaphoric**.

John Simon, Chief Financial Officer of Prime Corp. since 1986, saw his pay jump 20%, to $1.3 million, as the 37-year-old also became the financial-services company's president...

Types of referring expressions

- Definite Noun Phrases
- Indefinite Noun Phrases
- Pronouns
- Demonstrative pronouns
- One-Anaphora

Indefinite noun phrases

- Introduce entities that are new to the hearer into the discourse context
  - I saw *a Subaru WRX* today.
  - I saw *this awesome Subaru WRX* today.

Definite noun phrases

- Refer to an entity that is identifiable to the hearer
- It has already been mentioned in the discourse
- It is contained in the hearer's set of beliefs about the world
- The uniqueness of the object is implied by the description itself
  - I saw a Subaru WRX today. *The WRX* was blue and needed a wash.
  - *The Indy 500* is the most popular car race in the US.
  - *The fastest car in the Indy 500* was a Subaru WRX.

<u>Pronouns</u>

- Another form of ***definite*** reference
- Also known as **Anaphora**
- Referent must have a high degree of activation or **salience** in the discourse model
  - John went to Bob's party, and parked next to a beautiful Subaru WRX. He went inside and talked to Bob for more than an hour. Bob told him that he recently got engaged.

    ➔ (a)?? He also said that he bought <u>*it*</u> yesterday.

    ➔ (a') He also said that he bought <u>*the WRX*</u> yesterday.

- **Cataphora**: referring expression is mentioned before its referent
  - Before <u>*he*</u> bought <u>*it*</u>, John checked over the WRX carefully.

Types of referring expressions

- Definite Noun Phrases
- Indefinite Noun Phrases
- Pronouns
- Demonstrative pronouns
- One-Anaphora

Demonstrative pronouns

- Behave somewhat differently than simple definite pronouns
- Can appear alone or as determiners
- Choice of *this* or *that* depends on some notion of spatial or temporal proximity
  - I bought a WRX yesterday. It's similar to the one I bought a year ago. *That one* was really nice, but I like *this one* even better.

One-anaphora

- Blends properties of definite and indefinite reference
  - I saw no fewer than 6 Subaru WRX's today. Now I want *one*.
- May introduce a new entity into the discourse, but it is dependent on an existing referent for the description of this new entity.

Noun Phrase Coreference Resolution

- Identify all phrases that refer to each real-world entity mentioned in the text

John Simon, Chief Financial Officer of Prime Corp. since 1986, saw his pay jump 20%, to $1.3 million, as the 37-year-old also became the financial-services company's president...

Why It's Hard

Many sources of information play a role

- head noun matches
  - IBM *executives* = the *executives*
  - Microsoft *executives*
- syntactic constraints
  - John helped himself to...
  - John helped him to...
- discourse focus, recency, syntactic parallelism, semantic class, agreement, world knowledge, …

Why It’s Hard

No single source is a completely reliable indicator

- semantic preferences
  - Mr. Callahan = president =? the carrier
- number and gender
  - assassination (of Jesuit priests) = these murders
  - the woman = she = Mary =? the chairman

Why It's Hard

Coreference strategies differ depending on the type of referring NP

- definiteness of NPs
  - … Then Mark saw the man walking down the street.
  - … Then Mark saw a man walking down the street.

- pronoun resolution alone is notoriously difficult
  - resolution strategies differ for each type of pronoun
  - some pronouns refer to nothing in the text

I went outside and it was snowing.

Types of referents: complications

- Inferable
  - A referring expression does not refer to an entity in the text, but to one that is inferentially related to it.
    - I almost bought a WRX today, but *a door* had a dent and *the engine* seemed noisy.
    - Mix the flour, butter, and water. Stir *the batter* until all lumps are gone.
- Discontinuous sets
  - Referents may have been evoked in discontinuous phrases
    - John has a Volvo, and Mary has a Mazda. *They* drive *them* all the time.
- Generics – refer to a class of entities
  - I saw no fewer than 6 WRX's today. *They* are the coolest cars.

Traditional Knowledge-Based Approaches

Lappin and Leass [1994]

- hand-crafted heuristics and filters
  - syntactic filters [Lappin and McCord 1990a]
  - morphological filter
  - pleonastic pronoun filter ("It was raining.")
  - procedure for identifying possible antecedents [Lappin and McCord 1990b]
  - salience assignment w.r.t. grammatical role, proximity, parallelism,etc.

- decision procedure

Problems with hand-written rules

- Portability
- Robustness
- Few large-scale evaluations
- Evaluations make a number of simplifying assumptions
  - perfect parse
  - omit many difficult cases, e.g. pleonastic pronouns
- **Impose coreference resolution strategies rather than learn them empirically**

Plan for the Talk

- Linguistic background for coreference resolution
- supervised machine learning approach
- weakly supervised approaches

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband,
King George VI, into a viable monarch. Logue,
a renowned speech therapist, was summoned to help
the King overcome his speech impediment...

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming **her** husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband, **King George VI**, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming **her husband**, **King George VI**, into a viable monarch. Logue, a renowned speech therapist, was summoned to help **the King** overcome **his** speech impediment...

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. **Logue**, a renowned speech therapist, was summoned to help the King overcome his speech impediment...

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. **Logue**, a **renowned speech therapist**, was summoned to help the King overcome his speech impediment...

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband, King George VI, into <u>a viable monarch</u>. Logue, a renowned speech therapist, was summoned to help the King overcome his <u>speech impediment</u>...

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband,
King George VI, into a viable monarch. Logue,
a renowned speech therapist, was summoned to help
the King overcome his speech impediment...

Singletons!

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...

A Machine Learning Approach

Typical Steps:

- Step1: Noun Phrase Identification
- Step2: Pairwise Classification
- Step3: Clustering (Why?)

A Machine Learning Approach

- Step1: Find all noun phrases
  - Using "***partial parsers***" or "***chunkers***"

[Queen Elizabeth] set about transforming [her] [husband], ...

A Machine Learning Approach

- Step2: Pair-wise Classification (using machine learning)
  - given a description of two noun phrases, *NP_i* and *NP_j*, classify the pair as *coreferent* or *not coreferent*

coref ? *coref ?*

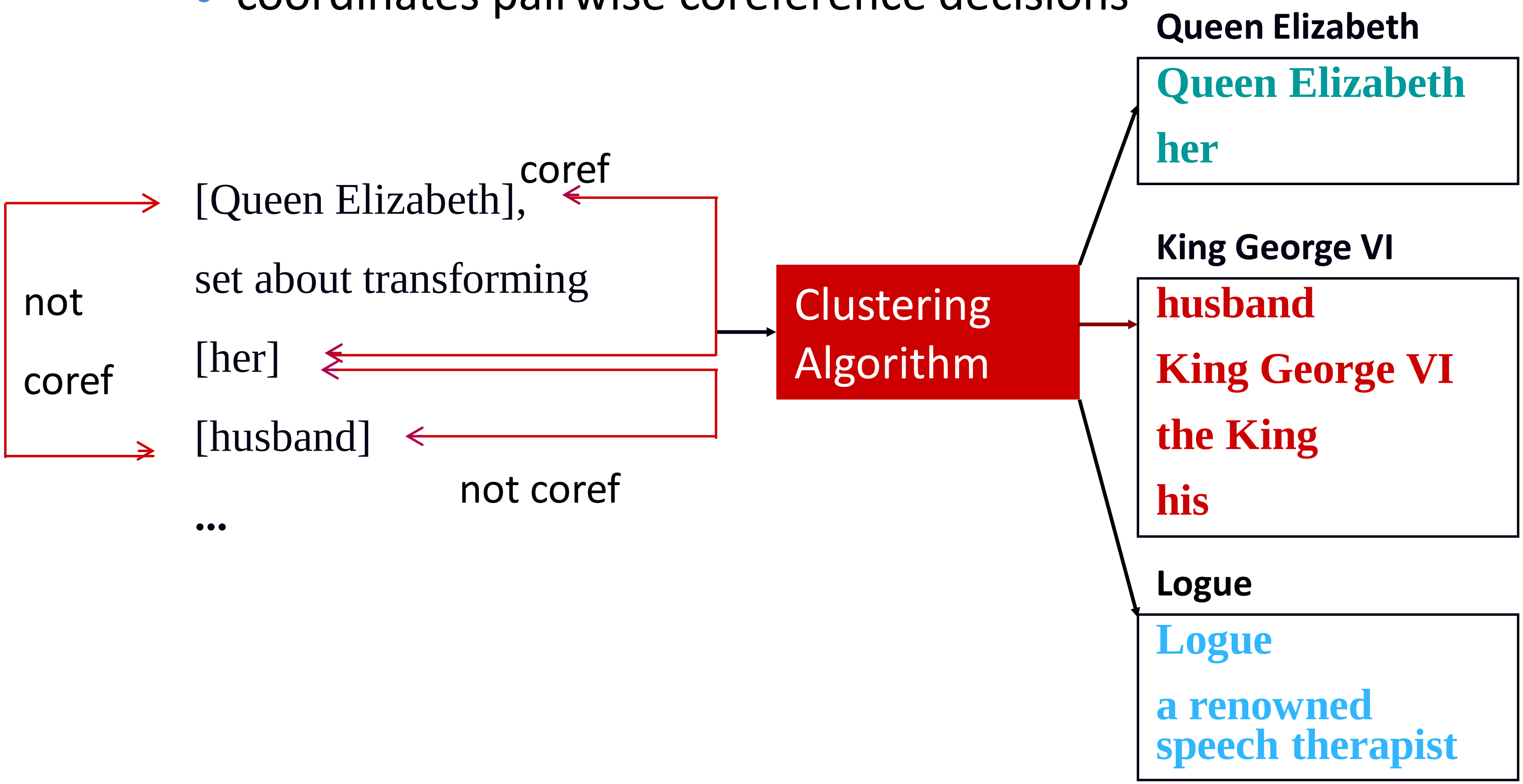
[Queen Elizabeth] set about transforming [her] [husband], ...

coref ?

Aone & Bennett [1995]; Connolly et al. [1994]; McCarthy & Lehnert [1995];
Soon et al. [2001]; Ng & Cardie [2002]; ...

A Machine Learning Approach

- Step3: Clustering
  - coordinates pairwise coreference decisions



Machine Learning Issues

- Training data creation
- Instance representation
- Learning algorithm for pair-wise decisions
- Clustering algorithm (to combine pair-wise decisions)

Supervised Inductive Learning

Examples of NP pairs (features + class)

ML Algorithm

Concept description

(novel) pair of NPs
(features)

(program)

class

Training Data Creation

- Creating training instances
  - texts annotated with coreference information

candidate antecedent anaphor

- one instance *inst*(NP_i, NP_j) for each *ordered* pair of NPs
  - NP_i precedes NP_j
  - feature vector: describes the two NPs and context
  - class value:

    coref pairs on the same coreference chain

    not coref otherwise

Instance Representation

- lexical
  - string matching for pronouns, proper names, common nouns
- grammatical
  - pronoun_1, pronoun_2, demonstrative_2, indefinite_2, …
  - number, gender, animacy
  - appositive, predicate nominative
  - binding constraints, simple contra-indexing constraints, …
  - span, maximalnp, …
- semantic
  - same WordNet class
  - alias
- positional
  - distance between the NPs in terms of # of sentences
- knowledge-based
  - naïve pronoun resolution algorithm

Why It's Hard

Many sources of information play a role

- string matching, syntactic constraints, semantic class,
- number agreement, gender agreement,
- discourse focus, recency,
- world knowledge…

- No single source is a completely reliable indicator
- Identifying each of these features automatically, accurately, and in context, is hard

Clustering Algorithm

- Best-first single-link clustering
  - Mark each NP_j as belonging to its own class: $NP_j \in c_j$
  - Proceed through the NPs in left-to-right order.
    - For each NP, NP_j, create test instances, $inst(NP_i, NP_j)$, for all of its preceding NPs, NP_i.
    - Select as the antecedent for NP_j the highest-confidence coreferent NP, NP_i, according to the coreference classifier (or none if all have below .5 confidence);
    - Merge c_j and c_j .

➔ Pros?

➔ Cons?

Clustering Algorithm

- Best-first single-link clustering

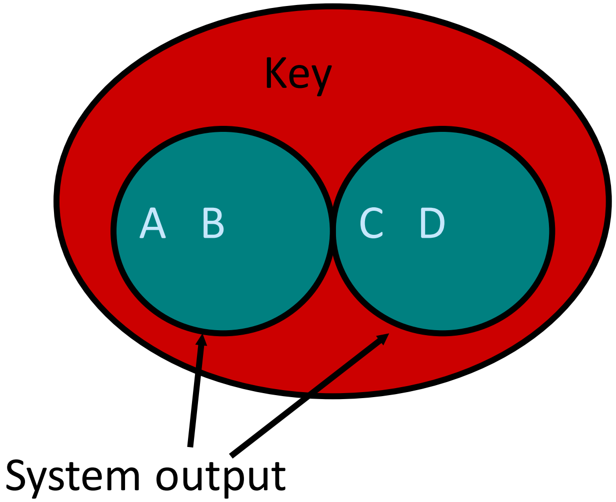
  → Pros: Simple but works surprisingly well!

  → Cons: Can’t go back and revise previous decisions

- Clustering algorithms that make collective decisions:
  - Corelational Clustering
  - Multi-cut
  - NP-hard, often hard to beat single-link clustering

Evaluation

- MUC-6 and MUC-7 coreference data sets
- documents annotated w.r.t. coreference
- 30 + 30 training texts (dry run)
- 30 + 20 test texts (formal evaluation)
- scoring program
  - recall
  - precision
  - F-measure: 2PR/(P+R)

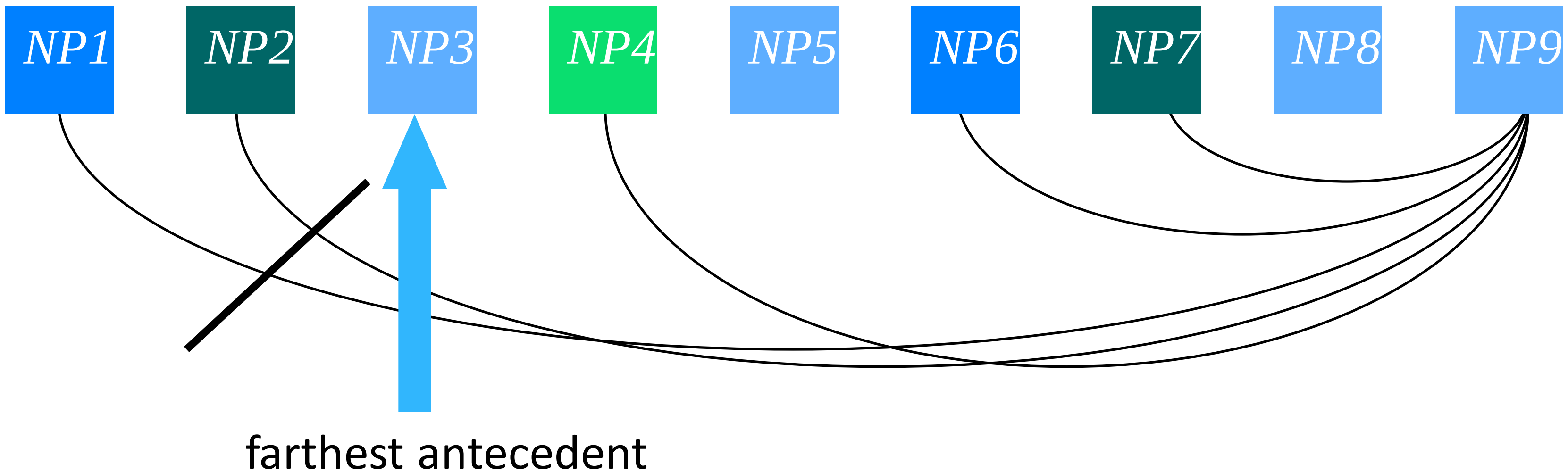


Baseline Results

	MUC-6			MUC-7		
	R	P	F	R	P	F
Baseline	40.7	73.5	**52.4**	27.2	86.3	**41.3**
Worst MUC System	36	44	40	52.5	21.4	30.4
Best MUC System	59	72	65	56.1	68.8	61.8

Problem 1

- Coreference is a rare relation
  - skewed class distributions (2% positive instances)
  - *remove some negative instances*



Problem 2

- Which pair do you think is harder for computers to learn/predict?

Queen Elizabeth set about transforming **her husband**,
King George VI, into a viable monarch. Logue,
the renowned speech therapist, was summoned to help
the King overcome his speech impediment...

Problem 2

- Order the following in the order of difficulties: (assuming best-first single-link clustering)
  - Pronouns
  - Proper Nouns
  - Common nouns

Queen Elizabeth set about transforming **her husband**, **King George VI**, into a viable monarch. Logue, the renowned speech therapist, was summoned to help **the King** overcome his speech impediment...

Problem 2

- Order the following in the order of difficulties

  → common nouns < pronouns < proper nouns

  (hardest) (easest)

Queen Elizabeth set about transforming **her husband**, **King George VI**, into a viable monarch. Logue, the renowned speech therapist, was summoned to help **the King** overcome his speech impediment...

Problem 2

- Coreference is a discourse-level problem with different solutions for different types of NPs
- positive example selection: selects easy positive training instances (cf. Harabagiu *et al.* (2001))

Queen Elizabeth set about transforming **her husband**, **King George VI**, into a viable monarch. Logue, the renowned speech therapist, was summoned to help **the King** overcome his speech impediment...

Problem 3

- Coreference is an ***equivalence relation***
  - loss of transitivity during pair-wise classification
  - need to tighten the connection between classification and clustering

coref ? *coref* ?

[Queen Elizabeth] set about transforming [her] [husband], ...

not coref ?

Results

	MUC-6			MUC-7		
	R	P	F	R	P	F
Baseline	40.7	73.5	52.4	27.2	86.3	41.3
NEG-SELECT	46.5	67.8	55.2	37.4	59.7	46.0
POS-SELECT	53.1	80.8	64.1	41.1	78.0	53.8
NEG-SELECT + POS-SELECT	63.4	76.3	69.3	59.5	55.1	57.2
NEG-SELECT + POS-SELECT + RULE-SELECT	63.3	76.9	**69.5**	54.2	76.3	**63.4**

- Ultimately: large increase in F-measure, due to gains in recall

Comparison with Best MUC Systems

	MUC-6			MUC-7		
	R	P	F	R	P	F
NEG-SELECT + POS-SELECT + RULE-SELECT	63.3	76.9	**69.5**	54.2	76.3	**63.4**
Best MUC System	59	72	**65**	56.1	68.8	**61.8**

Plan for the Talk

- noun phrase coreference resolution
- a (supervised) machine learning approach
- weakly supervised approaches
  - background
  - two techniques
  - evaluation

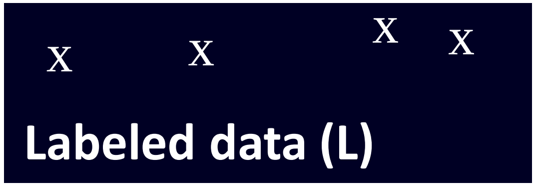
Weakly Supervised Approaches

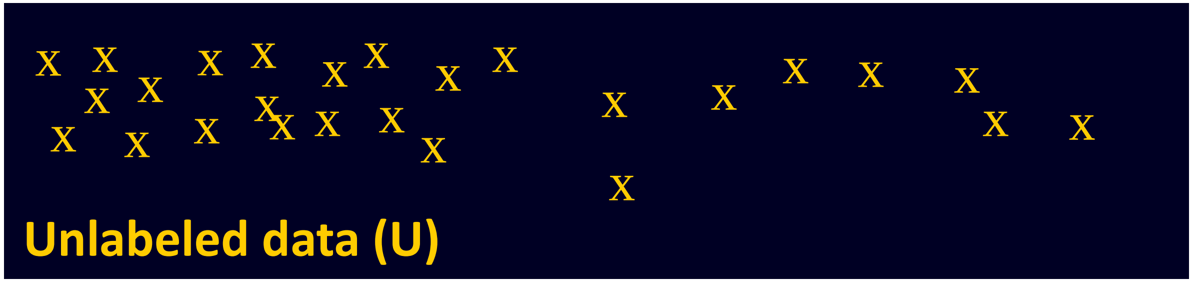
- Idea:

  bootstrap (NP coreference) classifiers using a *small amount of labeled data* (expensive) and a *large amount of unlabeled data* (cheap)

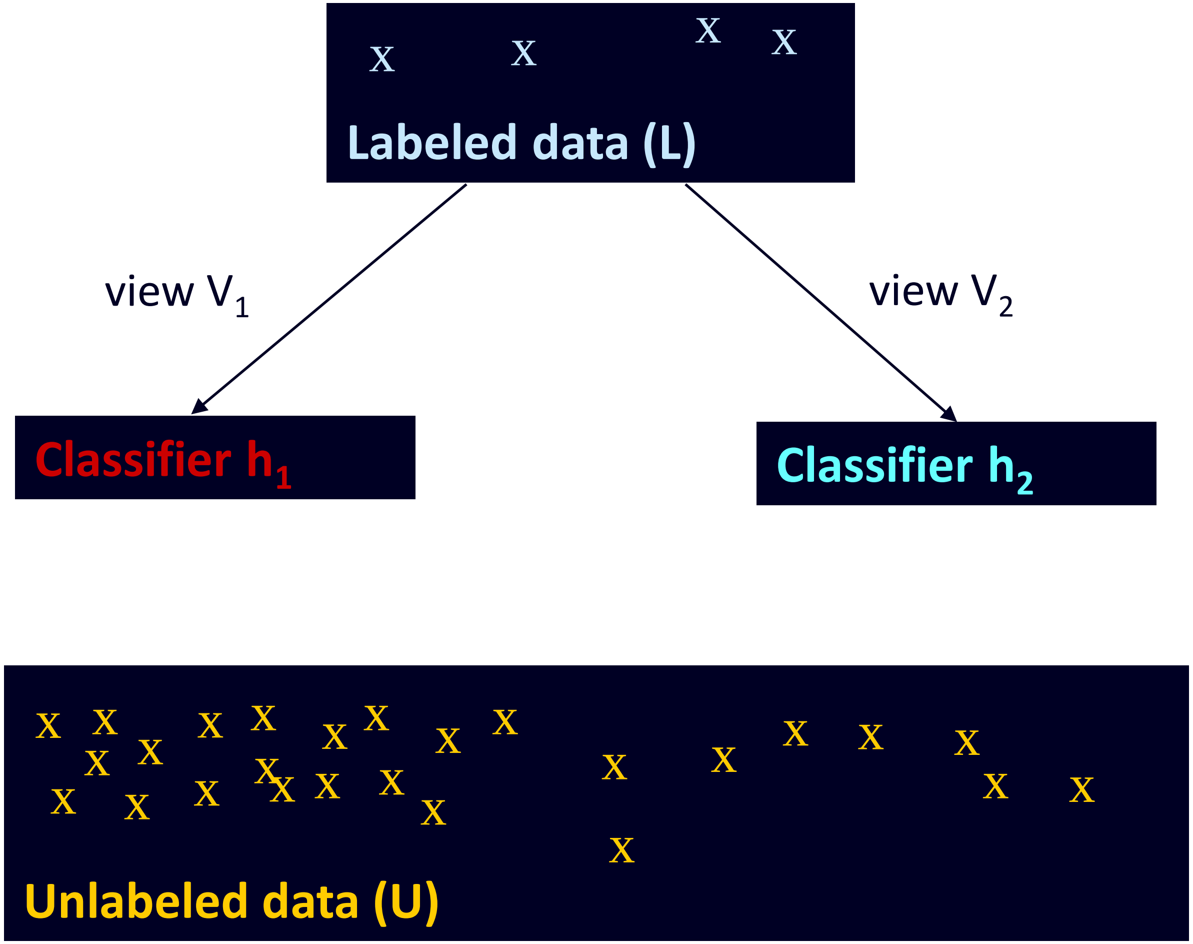
- Methods
  - Co-training
  - Self-training

Co-Training [Blum and Mitchell, 1998]

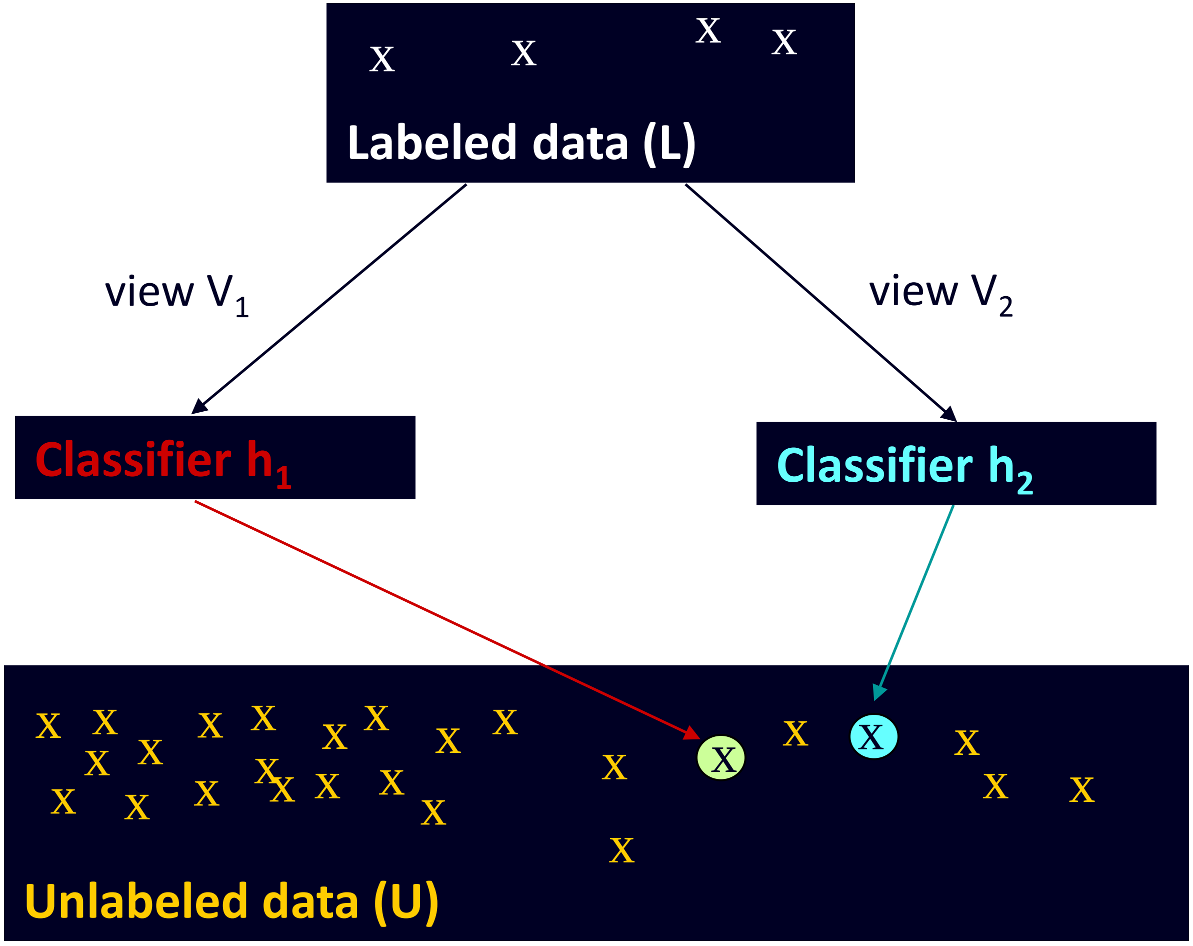




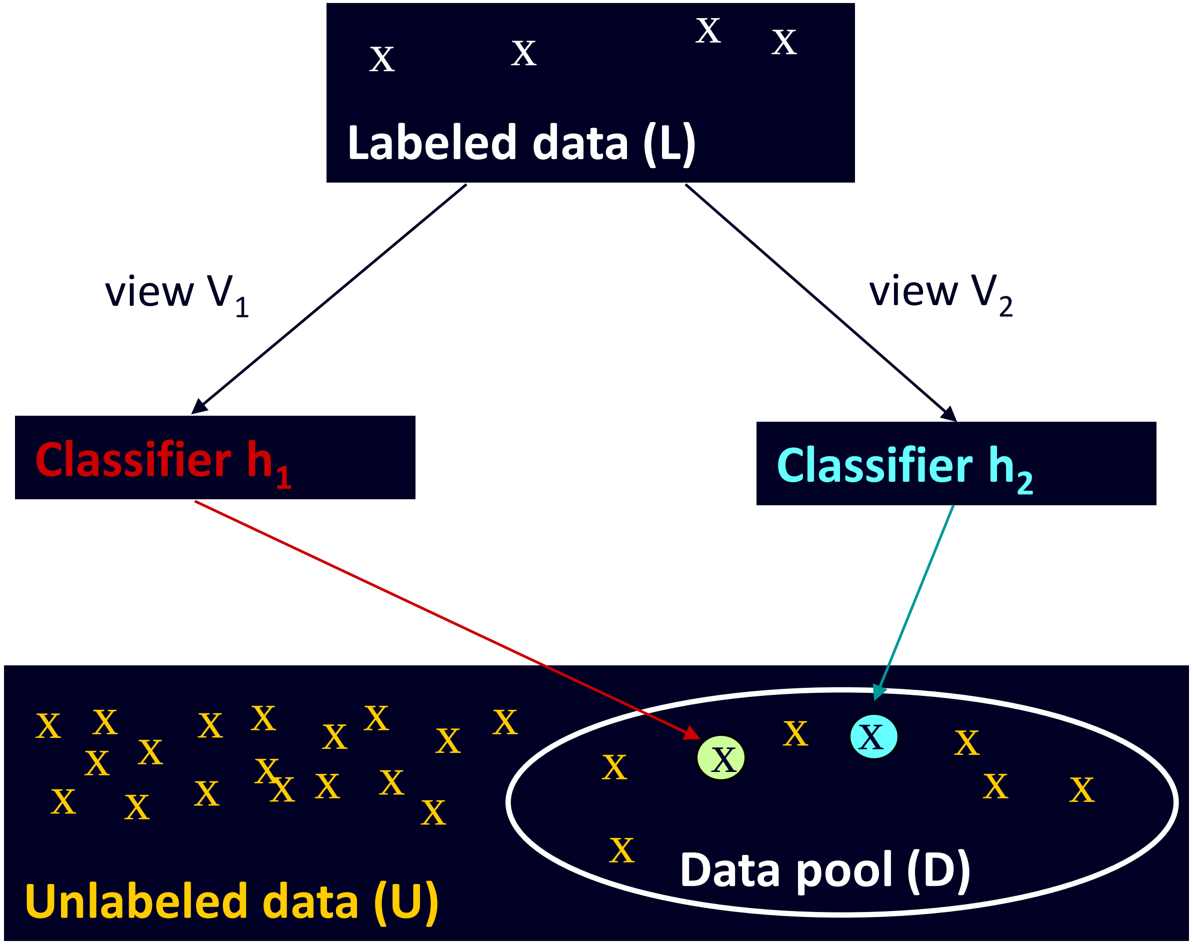
Co-Training [Blum and Mitchell, 1998]



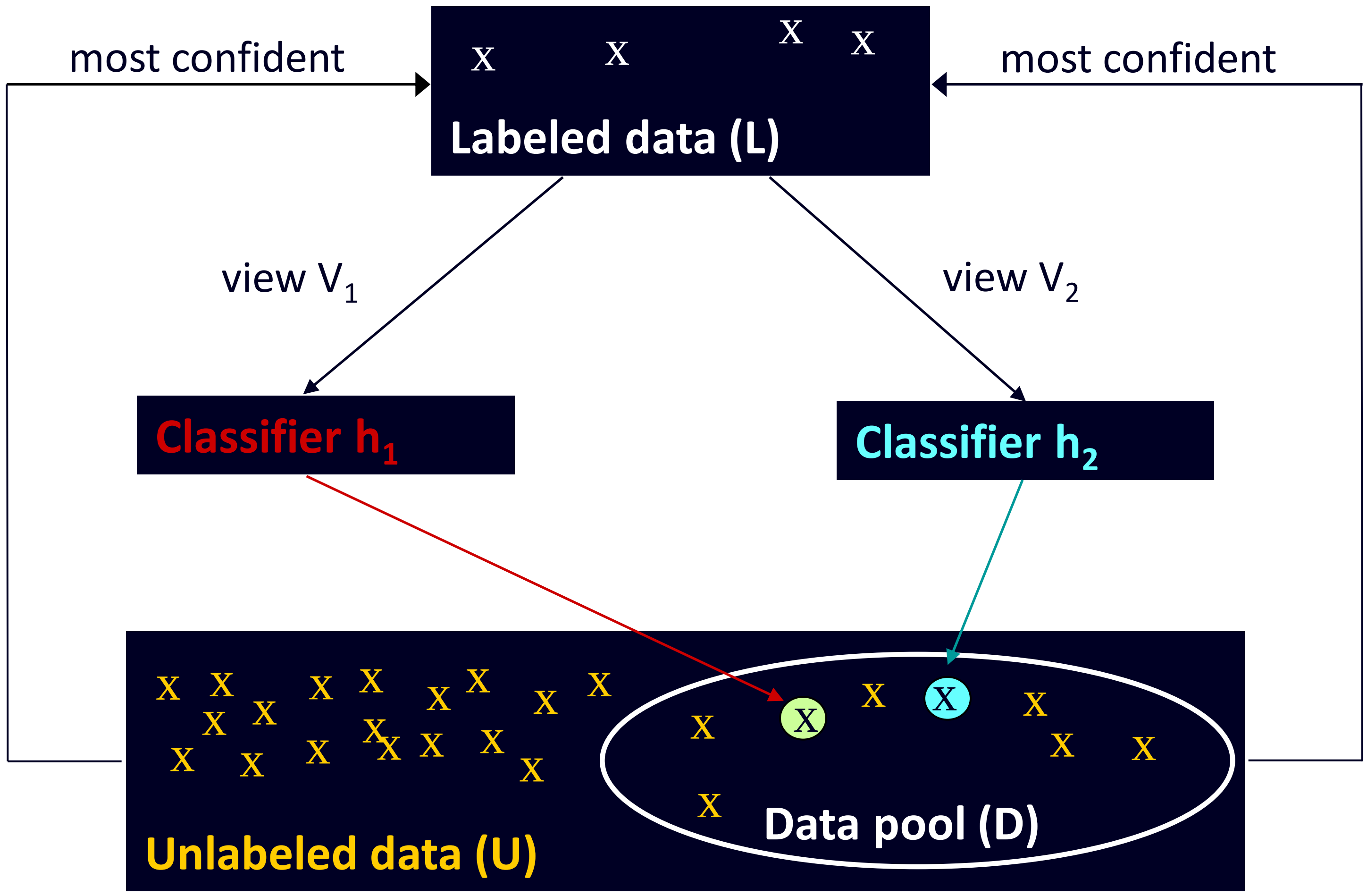
Co-Training [Blum and Mitchell, 1998]



Co-Training [Blum and Mitchell, 1998]



Co-Training [Blum and Mitchell, 1998]



Potential Problems with Co-Training

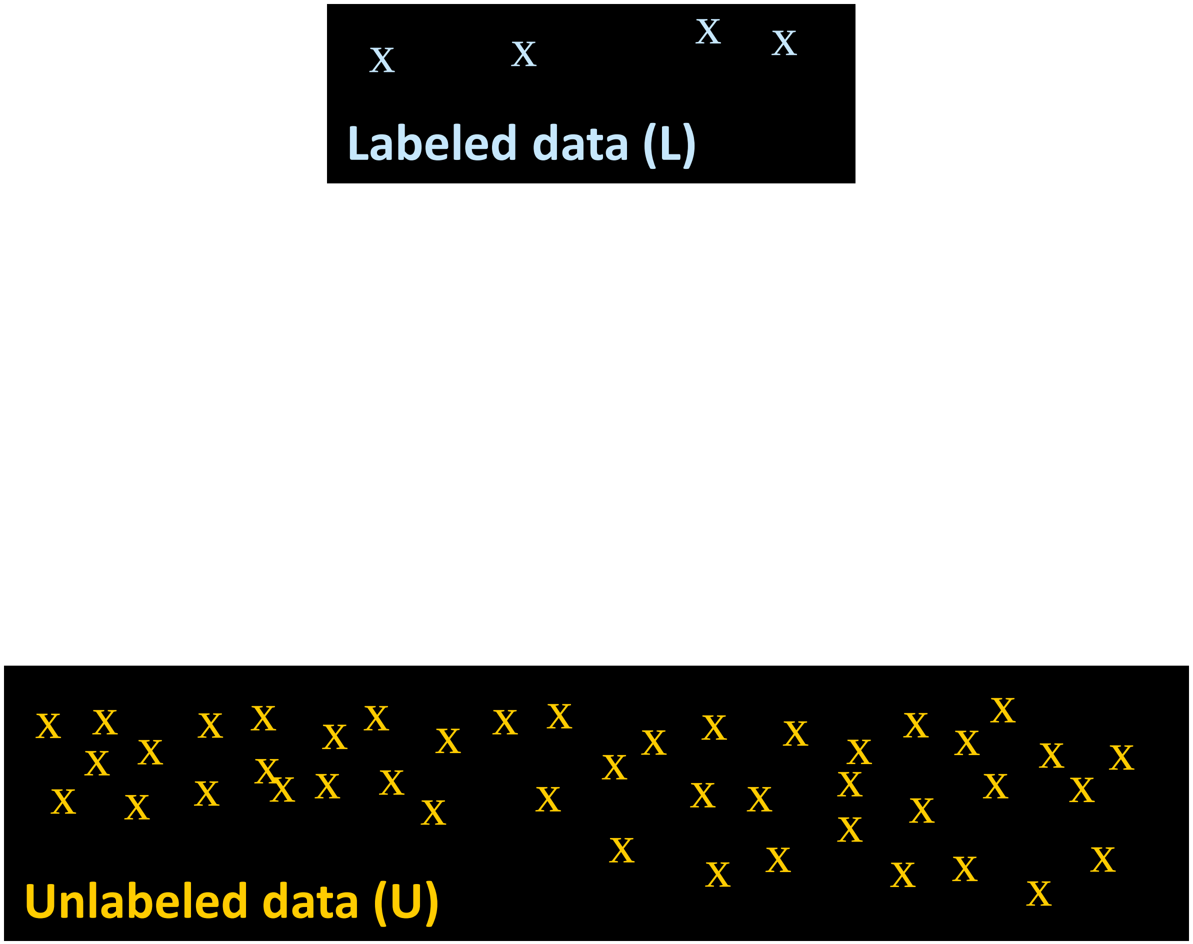
- Strong assumptions on the "views" (Blum and Mitchell, 1998)
  - each view must be sufficient for learning the target concept
  - the views must be conditionally independent given the class
  - empirically shown to be sensitive to these assumptions (Muslea *et al*., 2002)

- A number of parameters need to be tuned
  - views, data pool size, growth size, number of iterations, initial size of labeled data
  - algorithm is sensitive to its input parameters (Nigam and Ghani, 2000; Pierce and Cardie, 2001; Pierce 2003)

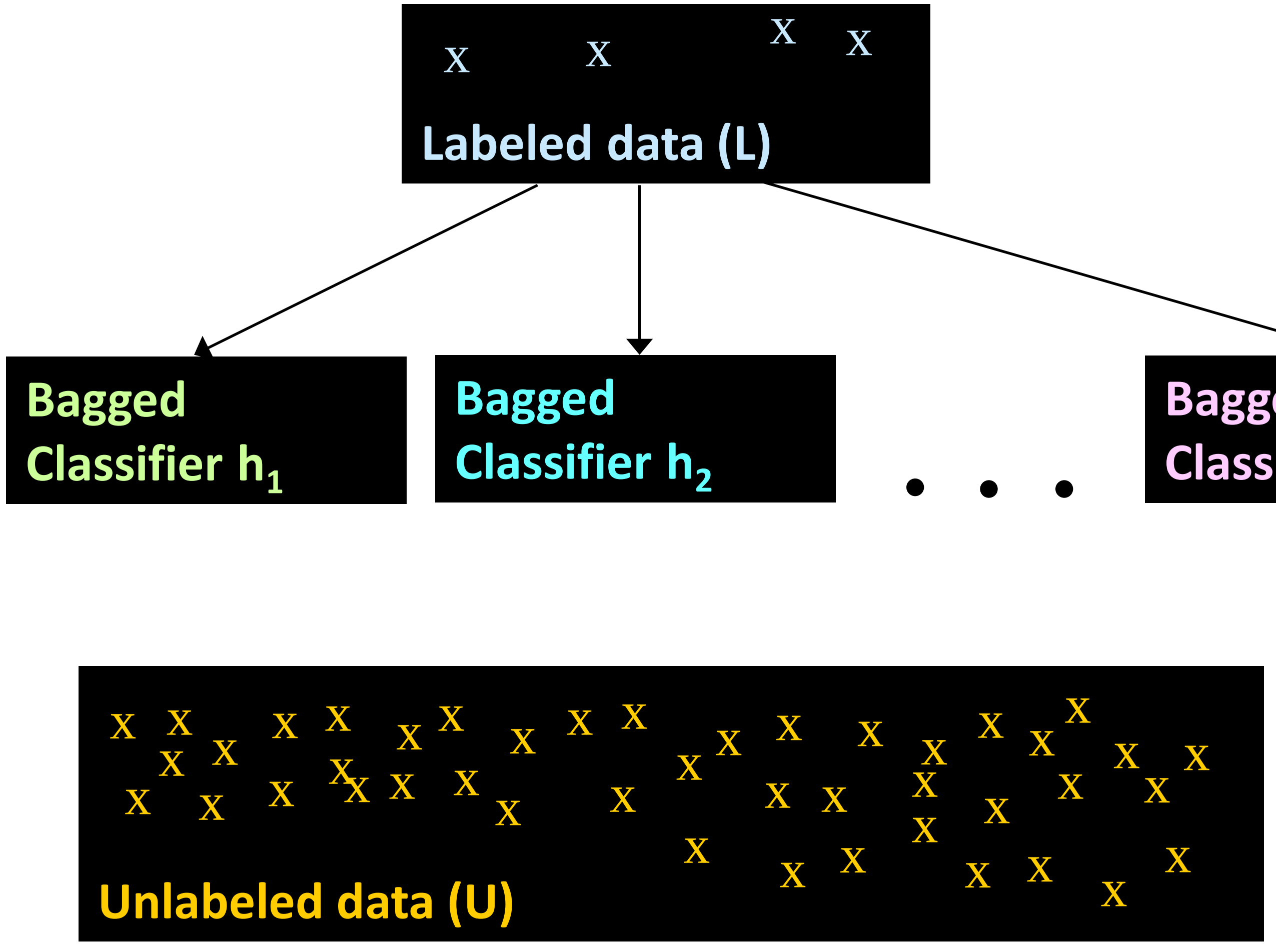
Potential Problems with Co-Training

- Multi-view algorithm
  - Is there any natural feature split for NP coreference?
    - view factorization is a non-trivial problem for coreference
      - Mueller *et al.*'s (2002) greedy method

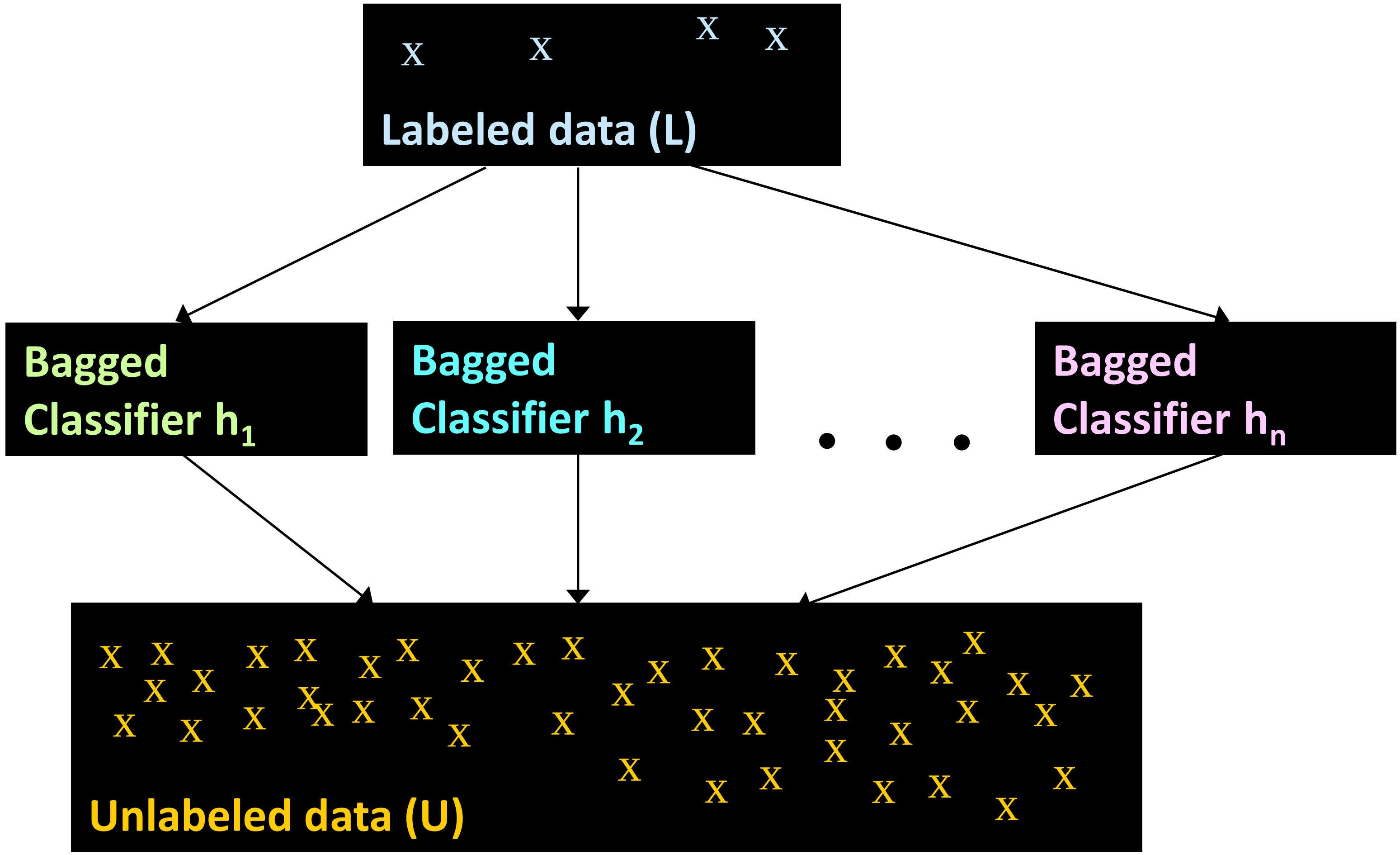
Self-Training with Bagging [Banko and Brill, 2001]



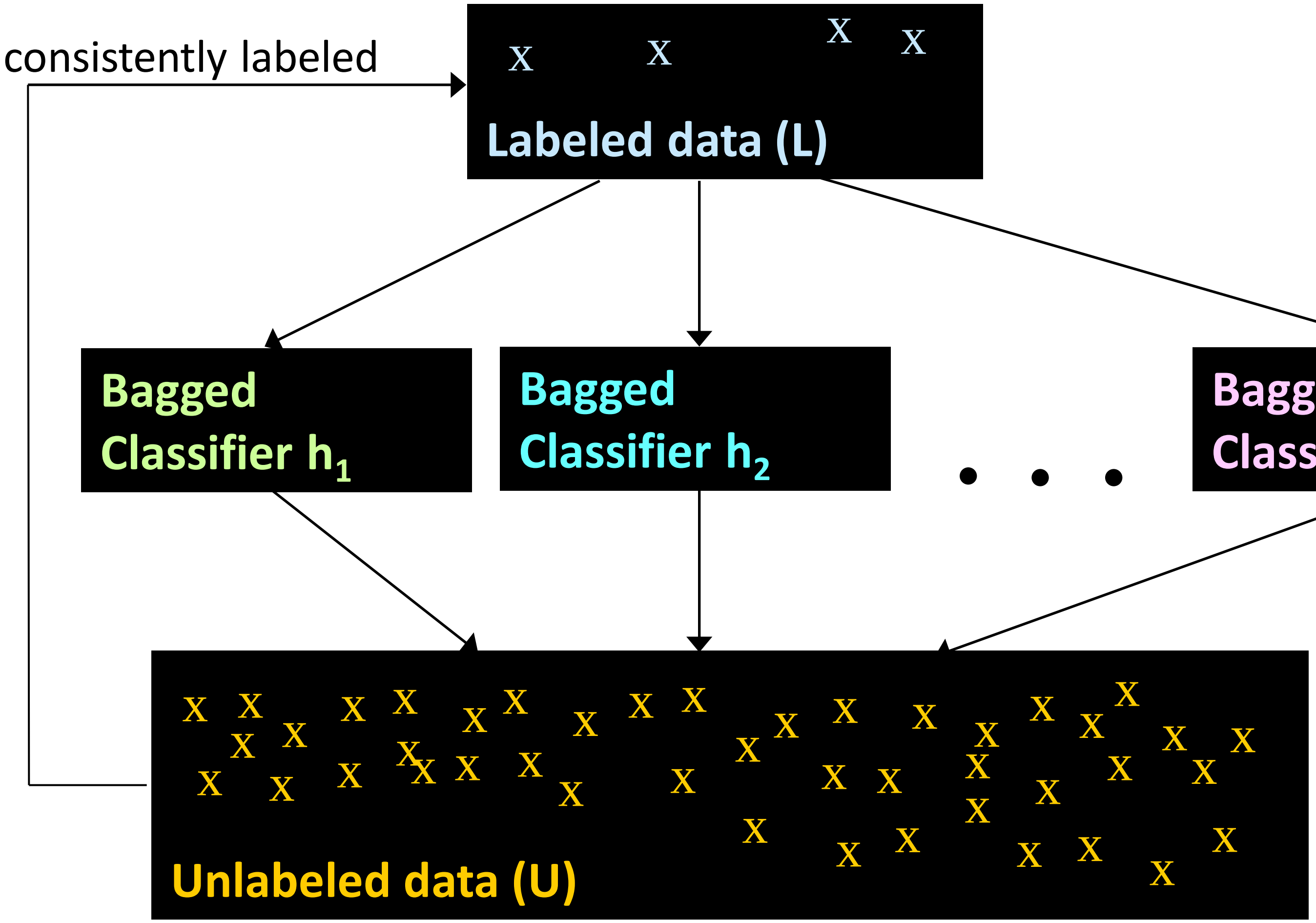
Self-Training with Bagging [Banko and Brill, 2001]



Self-Training with Bagging [Banko and Brill, 2001]



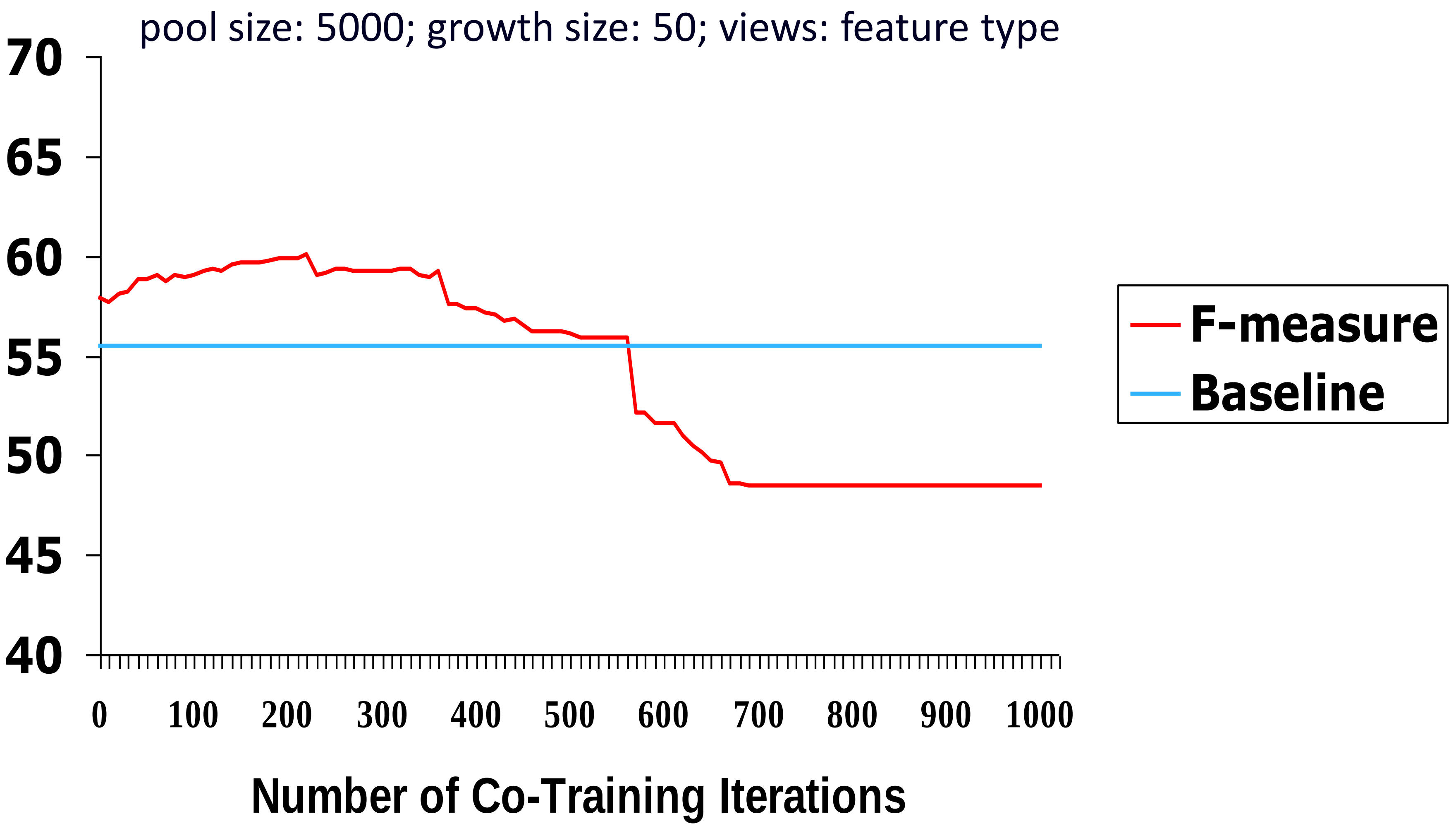
Self-Training with Bagging [Banko and Brill, 2001]



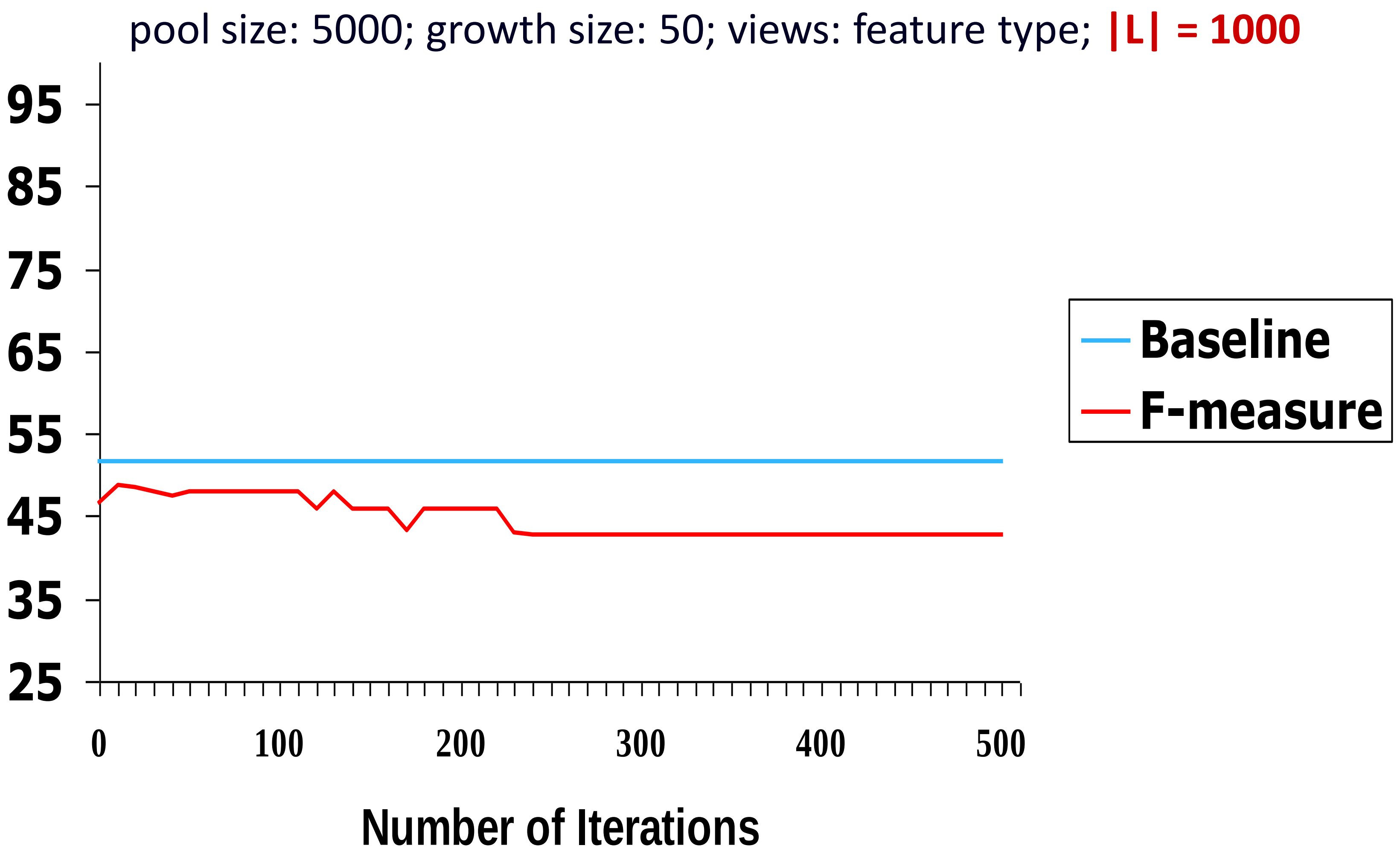
Evaluation

- MUC-6 and MUC-7 coreference data sets
- labeled data (L): one dryrun text
    - 3500-3700 instances
- unlabeled data (U): remaining 29 dryrun texts
- vs. fully supervised ML
    - ~500,000 instances (30 dryrun texts)

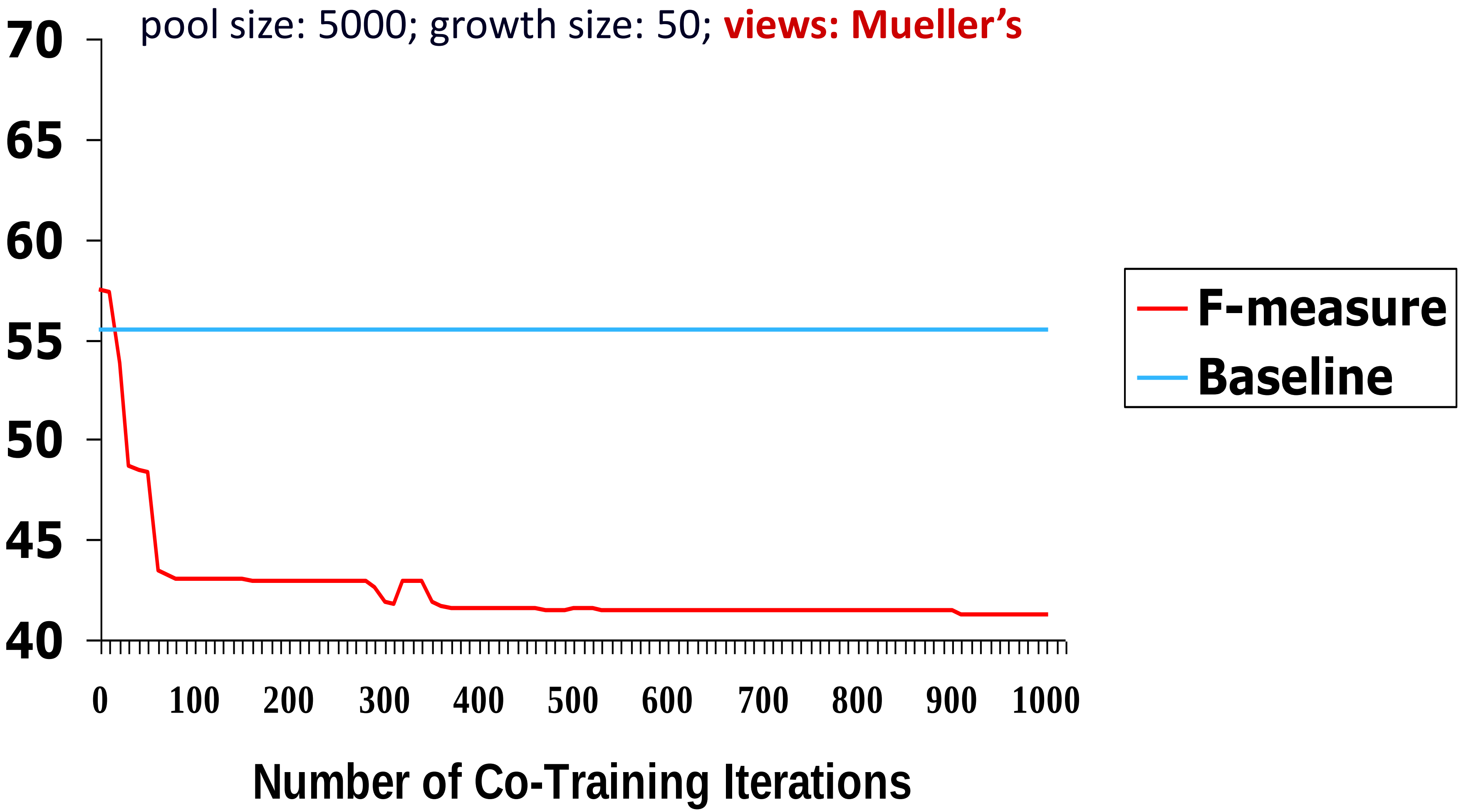
Learning Curve for Co-Training (MUC-6)



Learning Curve for Co-Training (MUC-6)



Learning Curve for Co-Training (MUC-6)



Self-Training Parameters

- Number of bags
  - tested all odd number of bags between 1 and 25

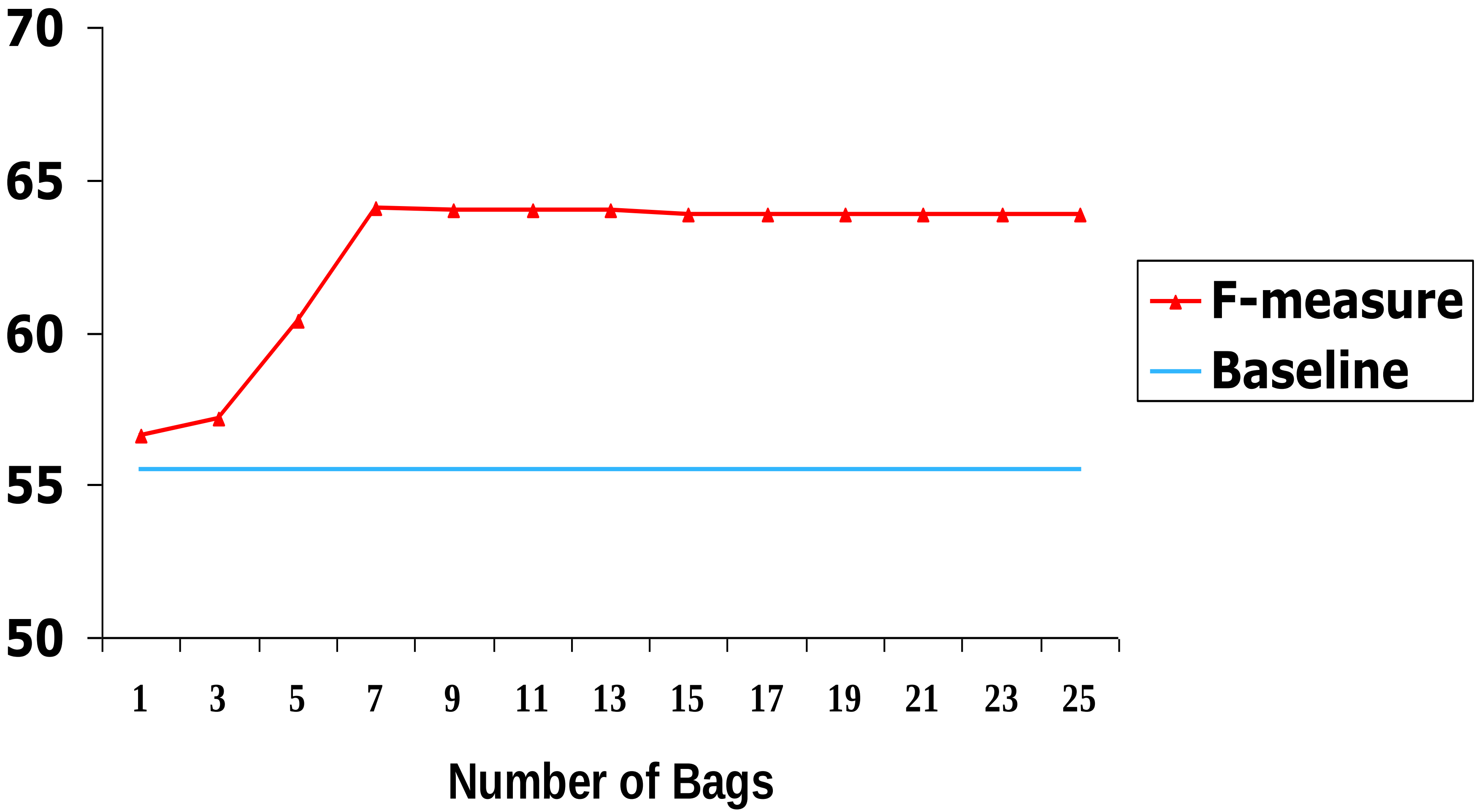
- 25 bags are sufficient for most learning tasks (Breiman, 1996)

Results (Self-Training with Bagging)

	MUC-6			MUC-7		
	R	P	F	R	P	F
Baseline	58.3	52.9	**55.5**	52.8	37.4	**43.8**
Co-Training	47.5	81.9	**60.1**	40.6	77.6	**53.3**
Self-Training with Bagging	54.1	78.6	**64.1**	54.6	62.6	**58.3**

- Self-training performs better than co-training

Self-Training: Effect of the Number of Bags (MUC-6)



Results

	MUC-6			MUC-7		
	R	P	F	R	P	F
Baseline	58.3	52.9	**55.5**	52.8	37.4	**43.8**
Co-Training	47.5	81.9	**60.1**	40.6	77.6	**53.3**
Self-Training with Bagging	54.1	78.6	**64.1**	54.6	62.6	**58.3**
Supervised ML* (~500,000 insts)	63.3	76.9	69.5	54.2	76.3	63.4

Summary

- Supervised ML approach to NP coreference resolution
  - Good performance relative to other approaches
  - Still lots of room for improvement
- Weakly supervised approaches are promising
  - Not as good performance as fully supervised, but use much less manually annotated training data
- For problems where no natural view factorization exists…
  - Single-view weakly supervised algorithms
    - Self-training with bagging