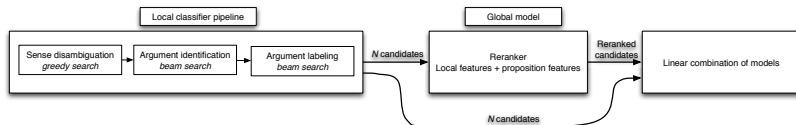# Pipeline and Reranker-based
Multilingual Semantic Role Labeling

Anders Björkelund, Love Hafdell, Pierre Nugues

June 4, 2009

Overview

- ▶ Pipeline of linear classifiers
- ▶ Beam search used to generate N candidates
- ▶ Reranker evaluates every candidate
- ▶ Pipeline and reranker scores are combined

Pipeline

- ▶ Predicate Disambiguation
  - ▶ One classifier for each lemma
  - ▶ Default sense labels for unknown lemmas
- ▶ Argument Identification
  - ▶ Binary classifier
  - ▶ No pruning
- ▶ Argument Classification
  - ▶ Multi-class classifier
  - ▶ Composite labels considered unique (Czech and Japanese)
- ▶ Specialized feature sets
  - ▶ Greedy forward selection
  - ▶ For each classifier in each language

Reranker

- ▶ Beam search used in argument identification and classification to generate pool of candidates
- ▶ Binary classifier that reranks complete propositions
- ▶ Features
    - ▶ All local AI features
    - ▶ All local AC features
    - ▶ Argument Label Sequence
- ▶ The reranker outputs a probability, $P_{Reranker}$

Generation of Candidates (AI)

- AI module generates the top k unlabeled propositions

	They	had	brandy	in	the	library	.
P(Arg)	0.979	0.00087	0.950	0.861	0.00006	0.0076	0.00009
P(¬Arg)	0.021	0.999	0.050	0.139	0.999	0.992	0.999

- $P_{AI} :=$ the product of the probabilities of all choices

Generation of Candidates (AC)

▶ AC module generates the top *l* labellings of each proposition

They	had	brandy	in	the	library	.
A0 0.999	–	A1 0.993	AM-TMP 0.471	–	–	–
A1 0.000487	–	C-A1 0.00362	AM-LOC 0.420	–	–	–
AM-DIS 0.000126	–	AM-ADV 0.000796	AM-MNR 0.0484	–	–	–
AM-ADV 0.000101	–	A0 0.000722	C-A1 0.00423	–	–	–

▶ $P_{AC} :=$ the product of the probabilities of all labels

Pipeline and Reranker combination

▶ The pipeline probability of a labeled proposition is defined as

$$P_{Local} := P_{AI} \times (P_{AC})^{1/a},$$

where a is the number of arguments

▶ P_{Local} probabilities are normalized to sum to 1, denoted P'_{Local}

▶ Final candidate is selected to maximize

$$P_{Final} := P'_{Local} \times (P_{Reranker})^{\alpha}$$

▶ $\alpha = 1$ gave best results on development set

Selecting Final Candidate

- Top ten candidates when using beam widths $k = l = 4$

Proposition	P'_{Local}	$P_{Reranker}$	P_{Final}
[They]$_{A0}$ had [brandy]$_{A1}$ [in]$_{AM-LOC}$ the library.	0.295	0.359	0.106
[They]$_{A0}$ had [brandy]$_{A1}$ [in]$_{AM-TMP}$ the library.	0.306	0.246	0.0753
[They]$_{A0}$ had [brandy]$_{A1}$ in the library.	0.0636	0.451	0.0287
[They]$_{A0}$ had [brandy]$_{A1}$ [in]$_{AM-MNR}$ the library.	0.143	0.0890	0.0128
[They]$_{A0}$ had [brandy]$_{A1}$ [in]$_{C-A1}$ the library.	0.137	0.0622	0.00854
[They]$_{A0}$ had brandy [in]$_{AM-TMP}$ the library.	0.0139	0.0206	$2.86 \cdot 10^{-4}$
[They]$_{A0}$ had brandy [in]$_{AM-LOC}$ the library.	0.0131	0.0121	$1.58 \cdot 10^{-4}$
They had [brandy]$_{A1}$ [in]$_{AM-TMP}$ the library.	0.00452	0.0226	$1.02 \cdot 10^{-4}$
They had [brandy]$_{A1}$ [in]$_{AM-LOC}$ the library.	0.00427	0.0133	$5.68 \cdot 10^{-5}$
[They]$_{A0}$ had brandy [in]$_{AM-MNR}$ the library.	0.00445	0.00364	$1.62 \cdot 10^{-5}$

Top ten propositions sorted by final score

Results

- Results and improvement by reranker (Labeled F_1 scores)

	Greedy	Reranker	Gain
Catalan	79.54	80.01	0.47
Chinese	77.84	78.60	0.76
Czech	84.99	85.41	0.42
English	84.44	85.63	1.19
German	79.01	79.71	0.70
Japanese	75.61	76.30	0.69
Spanish	79.28	76.52	-2.76
Spanish*	79.28	79.91	0.63
Average	80.10	80.31	0.21
Average*	80.10	80.80	0.70

* denotes post-evaluation figures after bug fix

Further Work

- ▶ Reranker features
  - ▶ Other feature templates
  - ▶ Feature selection
- ▶ Review combination of pipeline and reranker probabilities
- ▶ Dynamic beam width
- ▶ Argument pruning