Natural Language Processing and Information Retrieval

Semantic Role Labeling

Alessandro Moschitti
Department of information and communication technology
University of Trento
Email: moschitti@dit.unitn.it

Motivations for Shallow Semantic Parsing

- The extraction of semantics from text is difficult
- Too many representations:
  - α met β.
  - α and β met.
  - A meeting between α and β took place.
  - α had a meeting with β.
  - α and β had a meeting.
- Semantic arguments identify the participants in the event no matter how they were syntactically expressed.

Motivations Con't

- Two well defined resources
  - PropBank
  - FrameNet
  - High classification accuracy

Motivations (Kernel Methods)

- Semantics are connected to syntactic structures How to represent them?
- Flat feature representation
  - A deep knowledge and intuitions is required
  - Engineering problems when the phenomenon is described by many features
- Structures represented in terms of substructures
  - High complex space
  - Solution: convolution kernels (NEXT)

Predicate Argument Structures

- Given an event:
  - some words describe relation among its different entities
  - the participants are often seen as predicate's arguments.
- Example:

Paul gives a lecture in Rome

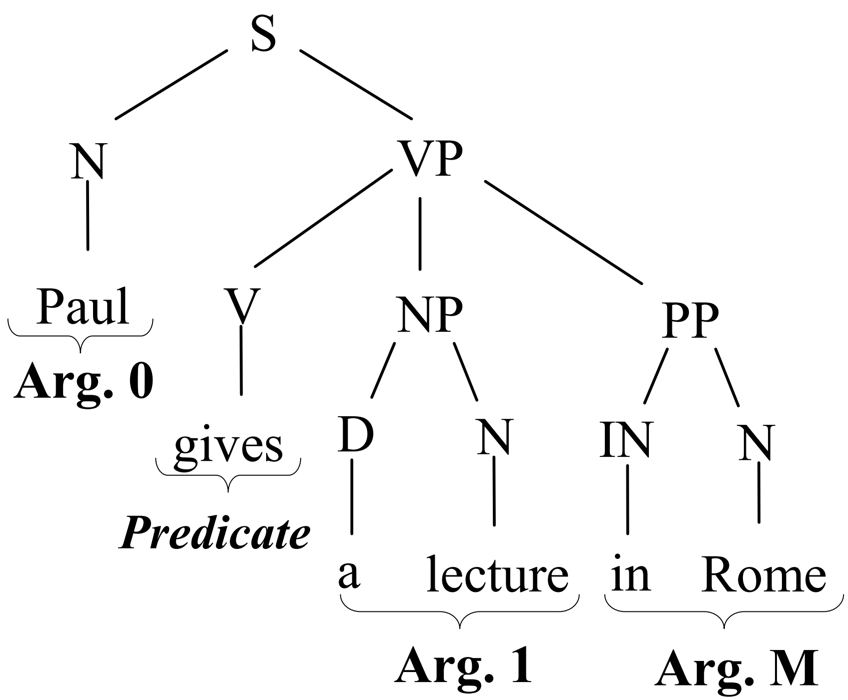
Predicate Argument Structures

- Given an event:
  - some words describe relation among its different entities
  - the participants are often seen as predicate's arguments.
- Example:

  [$_{Arg0}$ Paul] [$_{predicate}$ gives [$_{Arg1}$ a lecture] [$_{ArgM}$ in Rome]

Predicate Argument Structures (con't)

- Semantics are connected to syntax via parse trees



- Two different "standards": PropBank and FrameNet

PropBank

- 1 million-word corpus of Wall Street Journal articles
- The annotation is ***based on*** the Levin's classes.
- The arguments range from Arg0 to Arg9, ArgM.
- Lower numbered arguments more regular e.g.
  - Arg0 → subject and Arg1 → direct object.
- Higher numbered arguments are less consistent
  - assigned per-verb basis.

What does “based on Levin” mean?

- The semantic roles of verbs inside a Levin class are the same.
- The Levin clusters are formed at grammatical level according to diathesis alternation criteria.
- Diathesis alternations are variations in the way verbal-arguments are grammatically expressed

Diathesis Alternations

- Middle Alternation
  - [$_{\textit{Subject, Arg}0\textit{, Agent}}$ The butcher] cuts [$_{\textit{Direct Object, Arg}1\textit{, Patient}}$ the meat].
  - [$_{\textit{Subject, Arg}1\textit{, Patient}}$ The meat] cuts easily.
- Causative/inchoative Alternation
  - [$_{\textit{Subject, Arg}0\textit{, Agent}}$ Janet] broke [$_{\textit{Direct Object, Arg}1\text{, Patient,}}$ the cup]
  - [$_{\textit{Subject, Arg}1\textit{, Patient}}$ The cup] broke.

FrameNet (Fillmore, 1982)

- Lexical database
- Extensive semantic analysis of verbs, nouns and adjectives.
- Case-frame representations:
  - words evoke particular situations and participants (semantic roles)
- E.g.: Theft frame →

  7 diamonds were reportedly stolen

  from Bulgari in Rome

FrameNet (Fillmore, 1982)

- Lexical database
- Extensive semantic analysis of verbs, nouns and adjectives.
- Case-frame representations:
  - words evoke particular situations and participants (semantic roles)
- E.g.: Theft frame →

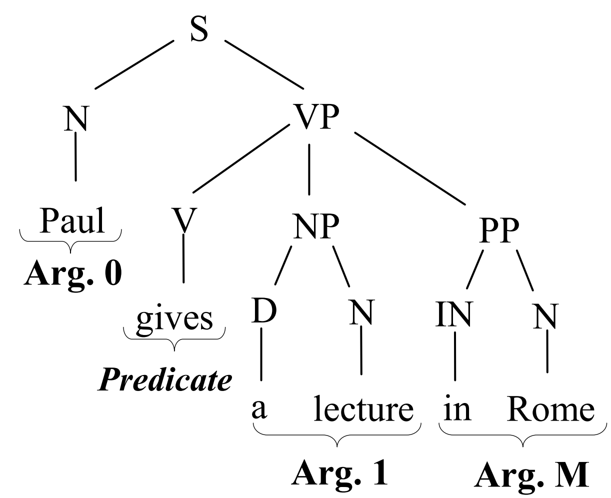
[$_{\text{Goods}}$ 7 diamonds] were reportedly [$_{\text{predicate}}$ stolen] [$_{\text{Victim}}$ from Bulgari] [$_{\text{Source}}$ in Rome].

Can we assign semantic arguments automatically?

- Yes….many machine learning approaches
  - Gildea and Jurasfky, 2002
  - Gildea and Palmer, 2002
  - Surdeanu et al., 2003
  - Fleischman et al 2003
  - Chen and Ranbow, 2003
  - Pradhan et al, 2004
  - Moschitti, 2004
  - Interesting developments in CoNLL 2004/2005
  - …

Automatic Predicate Argument Extraction

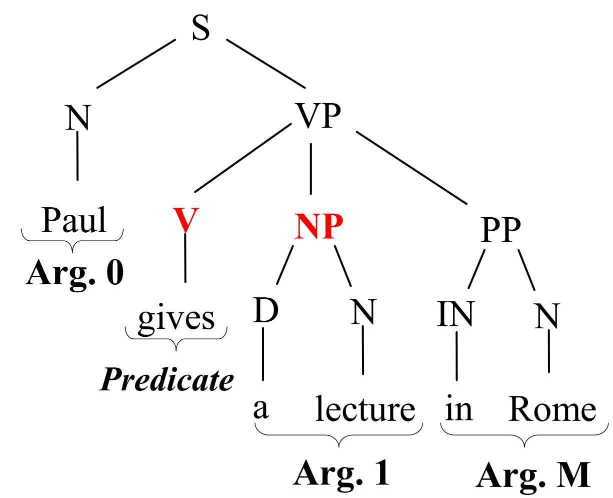
- Boundary Detection
  - One binary classifier
- Argument Type Classification
  - Multi-classification problem
  - *n* binary classifiers (ONE-vs-ALL)
  - Select the argument with maximum score



Predicate-Argument Feature Representation

Given a sentence, a predicate *p*:

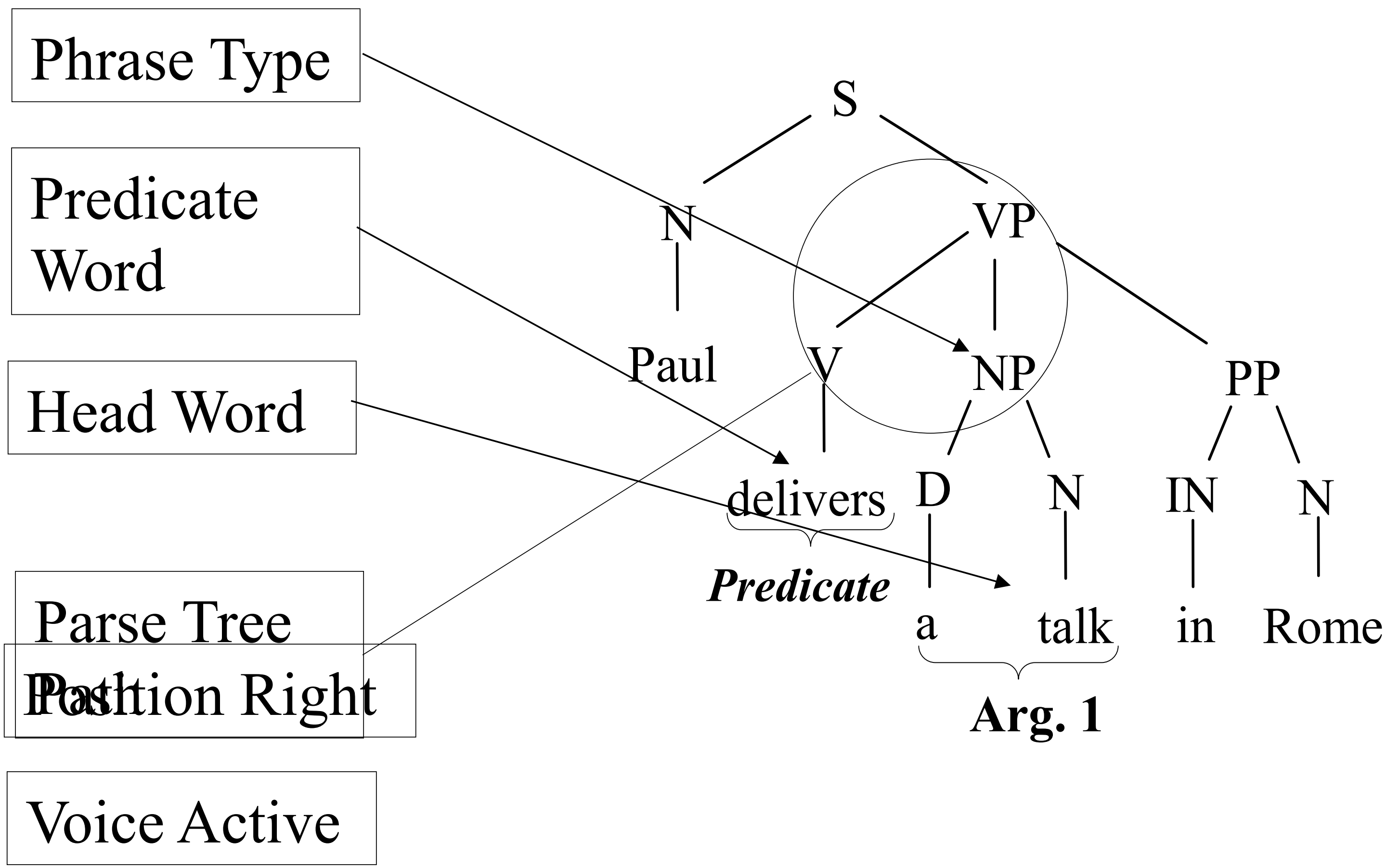
1. Derive the sentence parse tree
2. For each node pair <N_p,N_x>
   a. Extract a feature representation set *F*
   b. If N_x exactly covers the Arg-i, *F* is one of its positive examples
   c. *F* is a negative example otherwise



Typical standard flat features
(Gildea & Jurasfky, 2002)

- Phrase Type of the argument
- Parse Tree Path, between the predicate and the argument
- Head word
- Predicate Word
- Position
- Voice

An example



Phrase Type
Predicate Word
Head Word
Parse Tree Path
Position Right
Voice Active
S
N
VP
Paul
V
NP
PP
delivers
D
N
IN
N
Predicate
a
talk
in
Rome
Arg. 1

Flat features (Linear Kernel)

- To each example is associated a vector of 6 feature types

$$\vec{x} = (0,..,1,..,0,..,0,..,1,..,0,..,0,..,1,..,0,..,0,..,1,..,0,..,1,\ 1)$$

$$\text{PT} \qquad \text{PTP} \qquad \text{HW} \qquad \text{PW} \qquad \text{P} \quad \text{V}$$

- The dot product counts the number of features in common

$$\vec{x} \cdot \vec{z}$$

Feature Conjunction (polynomial Kernel)

- The initial vectors are the same
- They are mapped in

$$\Phi(< x_1, x_2 >) \rightarrow (x_1^2, x_2^2, \sqrt{2}x_1x_2, x_1, x_2, 1)$$

- This corresponds to …

$$\Phi(\vec{x}) \cdot \Phi(\vec{z}) =$$

$$x_1^2 z_1^2 + x_2^2 z_2^2 + 2x_1x_2z_1z_2 + x_1z_1 + x_2z_2 + 1 =$$

$$= (x_1z_1 + x_2z_2 + 1)^2 = (\vec{x} \cdot \vec{z} + 1)^2 = K_{Poly}(\vec{x}, \vec{z})$$

- More expressive, e.g. Voice+Position feature

  (used explicitly in [Xue and Palmer, 2004])

Polynomial vs. Linear

- Polynomial is more expressive.
- Example, only two features C_{Arg0} ($\cong$ the logical subject)
  - Voice and Position
- Without loss of generality we can assume:
  - `Voice` = 1 $\Leftrightarrow$ active and 0 $\Leftrightarrow$ passive
  - `Position` =1 $\Leftrightarrow$ the argument is after the predicate and 0 otherwise.
- C_{Arg0} = `Position` **XOR** `Voice`
  - non-linear separable
  - separable with the polynomial kernel

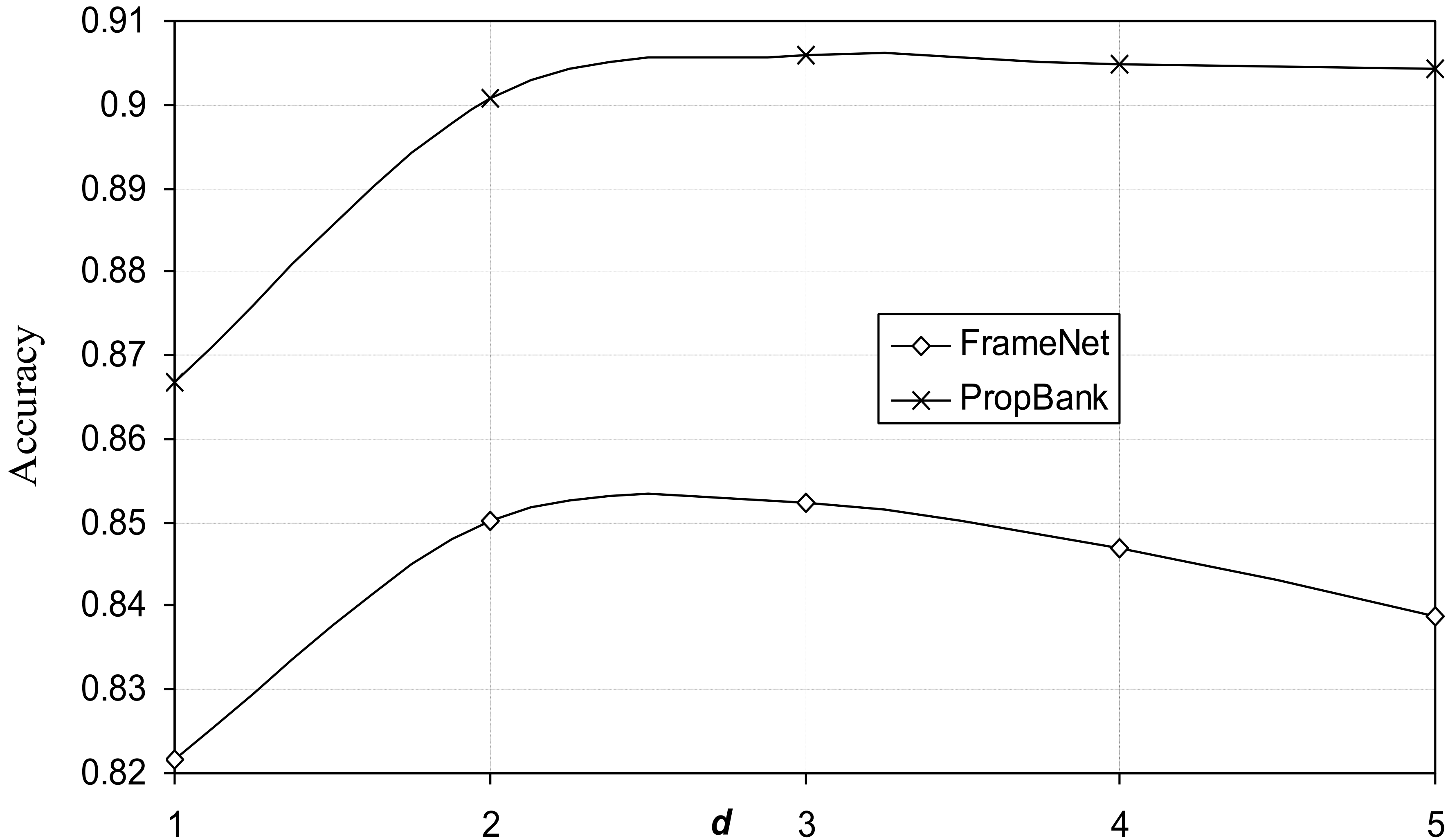
Gold Standard Tree Experiments

- PropBank and PennTree bank
  - about 53,700 sentences
  - Sections from 2 to 21 train., 23 test., 1 and 22 dev.
  - Arguments from Arg0 to Arg9, ArgA and ArgM for a total of 122,774 and 7,359
- FrameNet and Collins' automatic trees
  - 24,558 sentences from the 40 frames of Senseval 3
  - 18 roles (same names are mapped together)
  - Only verbs
  - 70% for training and 30% for testing

Boundary Classifier

- Gold trees
  - about 92 % of F1 for PropBank
- Automatic trees
  - about 80.7 % of F1 for FrameNet

Argument Classification with standard features



PropBank Results

Args	P3	PAT	PAT+P	PAT×P	SCF+P	SCF×P
Arg0	90.8	88.3	90.6	90.5	94.6	94.7
Arg1	91.1	87.4	89.9	91.2	92.9	94.1
Arg2	80.0	68.5	77.5	74.7	77.4	82.0
Arg3	57.9	56.5	55.6	49.7	56.2	56.4
Arg4	70.5	68.7	71.2	62.7	69.6	71.1
ArgM	95.4	94.1	96.2	96.2	96.1	96.3
Global Accuracy	**90.5**	**88.7**	**90.2**	**90.4**	**92.4**	**93.2**

PropBank Competition Results (CoNLL 2005)

- Automatic trees
- Boundary detection 81.3% (1/3 of training data only)
- Classification 88.6% (all training data)
- Overall:
  - 75.89 (**no heuristics applied**)
  - with heuristics [Tjong Kim Sang et al., 2005] 76.9

Other system results

	Development			Test WSJ		
	P(%)	R(%)	F_1	P(%)	R(%)	F_1
punyakanok	80.05	74.83	77.35	82.28	76.78	79.44
haghighi	77.66	75.72	76.68	79.54	77.39	78.45
marquez	78.39	75.53	76.93	79.55	76.45	77.97
pradhan	80.90	75.38	78.04	81.97	73.27	77.37
surdeanu	79.14	71.57	75.17	80.32	72.95	76.46
tsai	81.13	72.42	76.53	82.77	70.90	76.38
che	79.65	71.34	75.27	80.48	72.79	76.44
moschitti	74.95	73.10	74.01	76.55	75.24	75.89
tjongkimsang	76.79	70.01	73.24	79.03	72.03	75.37
yi	75.70	69.99	72.73	77.51	72.97	75.17
ozgencil	73.57	71.87	72.71	74.66	74.21	74.44
johansson	73.40	70.85	72.10	75.46	73.18	74.30
cohn	73.51	68.98	71.17	75.81	70.58	73.10
park	72.68	69.16	70.87	74.69	70.78	72.68
mitsumori	71.68	64.93	68.14	74.15	68.25	71.08
venkatapathy	71.88	64.76	68.14	73.76	65.52	69.40
ponzetto	71.82	61.60	66.32	75.05	64.81	69.56
lin	70.11	61.96	65.78	71.49	64.67	67.91
sutton	64.43	63.11	63.76	68.57	64.99	66.73
baseline	50.00	28.98	36.70	51.13	29.16	37.14

FrameNet Competition results Senseval 3 (2004)

- 454 roles from 386 frames
- Frame = “oracle feature”
- Winner – our system [Bejan et al 2004]
  - Classification – A = 92.5%
  - Boundary – F1 = 80.7%
  - Both tasks – F1 = 76.3 %

Competition Results

(UTDMoraresc u)	0.899	0.772	0.830674
(UAmsterdam)	0.869	0.752	0.806278
(UTDMoldovan)	0.807	0.78	0.79327
(InfoSciInst)	0.802	0.654	0.720478
(USaarland)	0.736	0.594	0.65742
(USaarland)	0.654	0.471	0.547616
(UUtah)	0.355	0.453	0.398057
(CLResearch)	0.583	0.111	0.186493