# A Shortest-path Method for Arc-factored Semantic Role Labeling

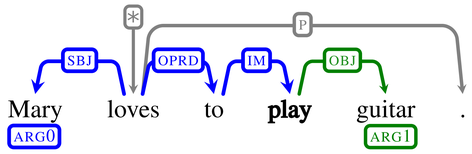

Xavier Lluís ··· Universitat Politècnica de Catalunya
Xavier Carreras ··· Xerox Research Centre Europe
Lluís Màrquez ··· Qatar Computing Research Institute


## — Abstract —


- We propose a Semantic Role Labeling (SRL) model to jointly find semantic roles and syntactic paths
  - SRL as the end goal
  - Syntactic parsing as an intermediate step to compute SRL features
  - Assumption: arc-factored SRL features
- Contributions:
  - ⋆ SRL in terms of shortest-path inference
  - ⋆ Use of the efficient Dijkstra's algorithm
  - ⋆ Ability to capture unrestricted predicate-argument paths


## — Motivation —

- Predict together syntactic paths and semantic roles
  - Assume a given predicate, e.g., *play*
  - Find the arguments *Mary* and *guitar* and also their predicate-argument paths




- Syntactic path as the source of features
- Distribution of path patterns in CoNLL-2009 data
  - English: Xue and Palmer (2004) rule
  - Czech: many ↓↓ ending patterns
  - Japanese: long tail of infrequent patterns

| English | | | Czech | | | Japanese | | |
|---|---|---|---|---|---|---|---|---|
| ∑ % | % | path | ∑ % | % | path | ∑ % | % | path |
| 63.63 | 63.6298 | ↓ | 63.90 | 63.8956 | ↓ | 37.20 | 37.1977 | ↓↓ |
| 73.97 | 10.3429 | ↑↓ | 86.26 | 22.3613 | ↓↓ | 51.52 | 14.3230 | ↓ |
| 80.63 | 6.65915 | ○ | 90.24 | 3.98078 | ↑↓ | 60.79 | 9.27270 | ↓↓↓ |
| 85.97 | 5.33352 | ↑ | 93.95 | 3.71713 | ↓↓↓ | 70.03 | 9.23857 | ↑ |
| 90.78 | 4.81104 | ↑↑↓ | 95.48 | 1.52168 | ↑↓↓ | 74.17 | 4.13359 | ↓↓↓↓ |
| 93.10 | 2.31928 | ↑↑↑↓ | 96.92 | 1.44091 | ↑ | 76.76 | 2.59117 | ↑↑ |
| 95.19 | 2.09043 | ↑↑ | 97.68 | 0.76714 | ↑↑↓ | 78.82 | 2.06111 | ↑↑↓↓ |
| 96.26 | 1.07468 | ↑↑↑↑↓ | 98.28 | 0.59684 | ↓↓↓↓ | 80.85 | 2.03381 | ↓↓↓↓↓ |
| 97.19 | 0.92482 | ↓↓ | 98.60 | 0.31759 | ↑↓↓↓ | 82.66 | 1.80631 | ↑↓↓ |
| 97.93 | 0.74041 | ↑↑↑ | 98.88 | 0.28227 | ↑↑↓↓ | 83.71 | 1.05558 | ↑↑↑ |
| 98.41 | 0.48565 | ↑↑↑↑↑↓ | 99.15 | 0.26721 | ↑↑↑↓ | 84.74 | 1.02828 | ↑↑↑↓↓ |
| 98.71 | 0.29769 | ↑↑↑↑ | 99.27 | 0.12430 | ↓↓↓↓↓ | 85.68 | 0.93500 | ↑↑↓↓↓ |
| 98.94 | 0.22733 | ↑↑↑↑↑↑↓ | 99.37 | 0.10103 | ↑↑↑↑↓ | 86.61 | 0.93273 | ↓↓↓↓↓↓ |
| 99.11 | 0.17805 | ↑↓↓ | 99.47 | 0.09747 | ↑↑ | 87.29 | 0.68249 | ↑↑↑↑↓↓ |
| 99.27 | 0.15316 | ↓↓↓ | 99.56 | 0.08515 | ↑↑↓↓↓ | 87.90 | 0.60969 | ↑↓↓↓ |
| 99.39 | 0.12065 | ↑↑↑↑↑ | 99.63 | 0.07419 | ↑↑↑↓↓ | 88.47 | 0.56646 | ↑↑↓↓↓↓ |

- ↑ upwards dependency from modifier to head
- ↓ downwards dependency from head to modifier
- ○ the argument is the predicate itself

## — Arc-factored SRL —

- Notation
  - x sentence
  - p predicate
  - a argument
  - r role
- Given fixed $x, p, r$ find best argument

$$\operatorname*{argmax}_{a} s(x, p, r, a)$$

- Decompose $s$ in the sum of $s_0$ and a $s_{syn}$

$$s(x, p, r, a) = s_0(x, p, r, a) + \max_{\pi} s_{syn}(x, p, r, a, \pi)$$

- The path $\pi$ is scored by $s_{syn}$
- First-order factorization of the path scores

$$s_{syn}(x, p, r, a, \pi) = \sum_{\langle h,m,l\rangle \in \pi} s_{syn}(x, p, r, a, \langle h, m, l\rangle)$$

## — SRL as Shortest Path —

- Assume a fixed predicate $p$ and role $r$
- Find paths from $p$ to $a$ with the smaller penalty $\theta$

$$\min_{\pi} \sum_{\langle h,m,l\rangle \in \pi} \theta_{\langle h,m,l\rangle}$$

- Given $\bar{x} = \langle x, p, r, a\rangle$, define a non-negative penalty as

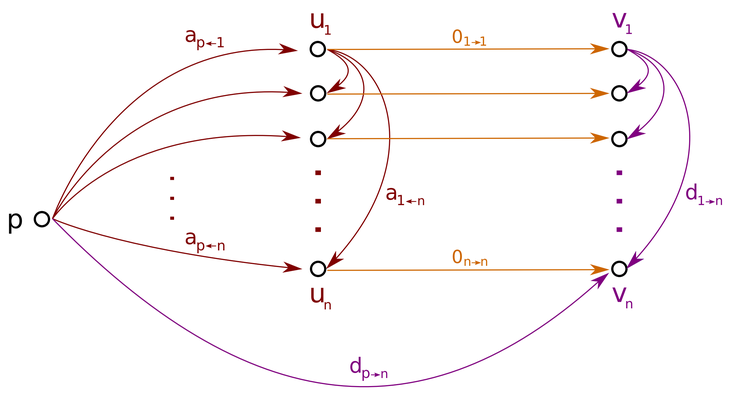
$$\theta_{\langle h,m,l\rangle} = -w \cdot f(\bar{x}, \langle h, m, l\rangle) + \theta_0$$

- where $\theta_0$ shifts all predicted scores to the positive side

$$\theta_0 = \max_{\langle h,m,l\rangle} w \cdot f(\bar{x}, \langle h, m, l\rangle)$$

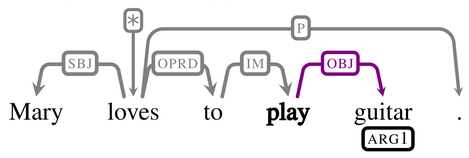
- Perceptron to learn $w$
- Dijkstra's algorithm finds the optimal path efficiently in $O(V^2)$

## — Graph Definition and Construction Algorithm —

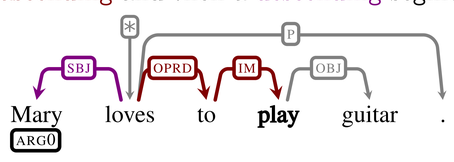



1. Add predicate $p$
2. Add ascending segments
   2.1 add nodes $u_1, \ldots, u_n$
   2.2 add ascending arcs $a_{i \leftarrow j}$
3. Add descending segments
   3.1 add nodes $v_1, \ldots, v_n$
   3.2 add descending arcs $d_{i \rightarrow j}$
4. Connect ascending and descending segments with $0_{i \rightarrow i}$

- From predicate *play* to argument *guitar* there is a single-edge descending path




- From predicate *play* to argument *Mary* there is an ascending and then a descending segment




## — Experimental Setting —

- Filter dependencies by a probabilistic parser
- Standard path features from Johansson (2009)
- Structured averaged perceptron for learning
- CoNLL-2009 Shared Task English verbal data

## — Results —

- Filter dependencies by $\gamma$, a probability factor

| Threshold $\gamma$ | 1 | 0.9 | 0.5 | 0.1 | 0.01 |
|---|---|---|---|---|---|
| Ratio | 1 | 1.014 | 1.103 | 1.500 | 2.843 |

  - $\gamma$=1 is similar to a pipeline
- Development set results

| Threshold | prec (%) | rec (%) | $F_1$ |
|---|---|---|---|
| training $\gamma = 1$ | | | |
| 1 | 77.91 | 73.97 | 75.89 |
| 0.9 | 77.23 | 74.17 | 75.67 |
| 0.5 | 73.30 | 75.03 | 74.16 |
| 0.1 | 58.22 | 68.75 | 63.05 |
| 0.01 | 32.83 | 53.69 | 40.74 |
| training $\gamma = 0.1$ | | | |
| 1 | 84.03 | 72.52 | 77.85 |
| 0.9 | 83.76 | 72.66 | 77.82 |
| 0.5 | 82.75 | 73.33 | 77.75 |
| 0.1 | 77.25 | 72.20 | 74.64 |
| 0.01 | 63.90 | 65.98 | 64.92 |
| training $\gamma = 0.01$ | | | |
| 1 | 81.62 | 69.06 | 74.82 |
| 0.9 | 81.45 | 69.19 | 74.82 |
| 0.5 | 80.80 | 69.80 | 74.90 |
| 0.1 | 77.92 | 68.94 | 73.16 |
| 0.01 | 74.12 | 65.92 | 69.78 |

  - Better to train with some variability
  - Better to test with the most restricted syntax
- Test set results

| System | prec(%) | rec(%) | $F_1$ |
|---|---|---|---|
| Non-factored | 86.96 | 75.92 | 81.06 |
| Factored $\gamma = 1$ | 79.88 | 76.12 | 77.96 |
| Factored best | 85.26 | 74.41 | 79.46 |

  - Overall, we are below the best shared task results: 83.97 $F_1$ (Zhao et al., 2009)
    - Using arc-factored features penalizes by 1.6 $F_1$ points
    - Our model uses linear classifiers and less feature engineering
    - Potentially more robust classifiers could be learned

## — Future Work —

- Apply the model to other semantic tasks
- Multilingual evaluation
- Probabilistic path modeling


**Acknowledgements:** This work was financed by the European Commission for the XLike project (FP7-288342); and by the Spanish Government for projects Tacardi (TIN2012-38523-C02-00) and Skater (TIN2012-38584-C06-01). For a large part of this work Xavier Carreras was at the Universitat Politècnica de Catalunya under a Ramón y Cajal contract (RYC-2008-02223).