Modeling Syntactic-Semantic Dependency Correlations in Semantic Role Labeling Using Mixture Models

Junjie Chen [1] Xiangheng He [2] Yusuke Miyao [1]

[1]The University of Tokyo [2]Imperial College London

Syntax-Aware Semantic Role Labeling

- Semantic Role Labeling predicts semantic dependencies
- Semantic dependencies correlate with syntactic dependencies

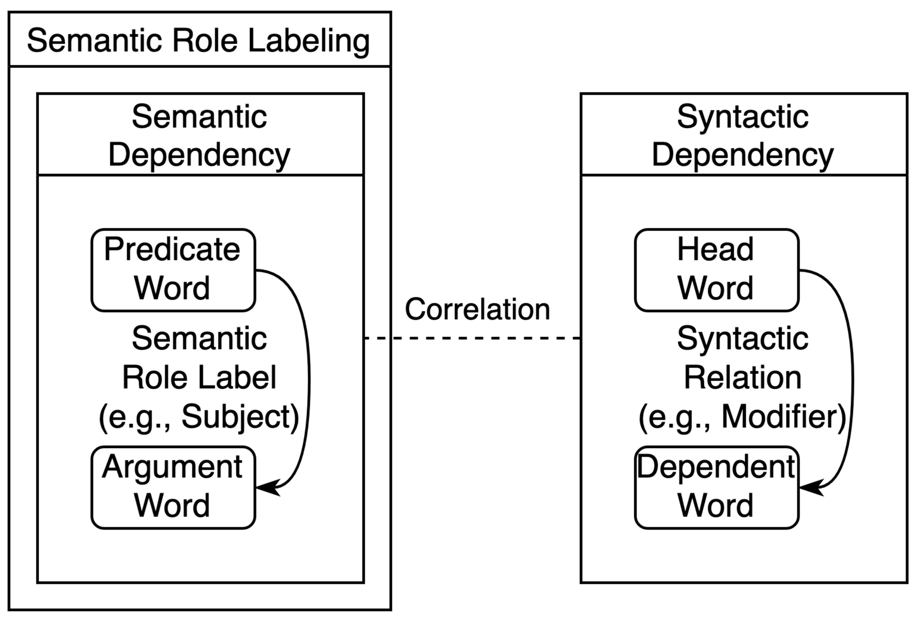

Figure 1. Semantic role labeling and two dependencies

Research question: How to *define* and *utilize* the Syntactic-Semantic Dependency Correlation for semantic dependency predictions

Concepts

- Semantic labels denote the relation of semantic dependencies
- Hop patterns denote the transition of shortest syntactic dependency paths (SSDP)

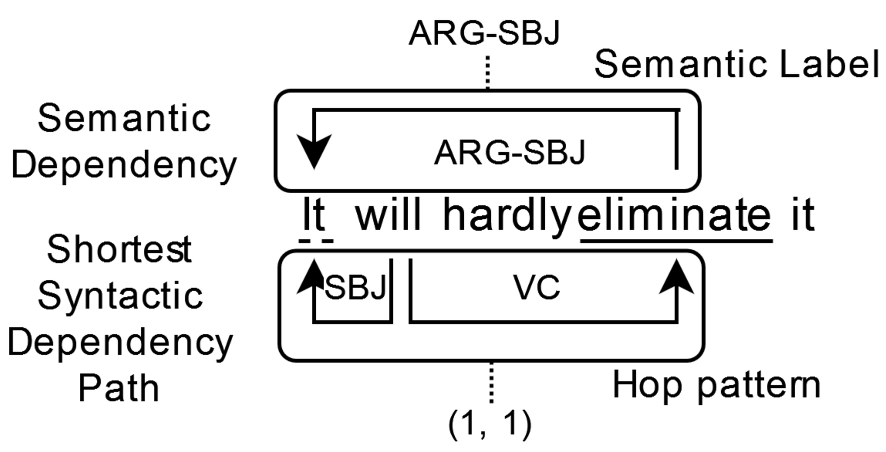

Figure 2. Relevant concepts. The solid underline highlights the predicate word and the dashed underline highlights the argument word

Proposal Overview

We propose a mixture model-based semantic parser (MM) that

- Clusters hop patterns with similar semantic label distributions using mixture weights
- Decouples semantic label distributions for different hop patterns using component distributions

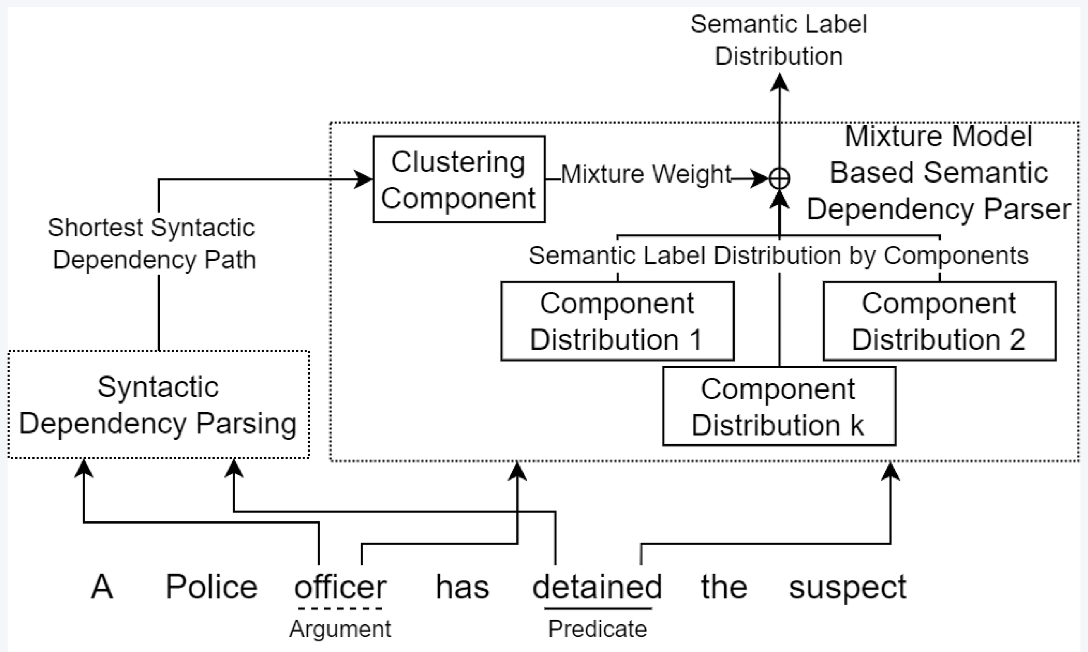

Figure 3. Illustration of the proposed model.

References

[1] Shexia He, Zuchao Li, Hai Zhao, and Hongxiao Bai. Syntax for semantic role labeling, to be, or not to be. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2061–2071, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[2] Michael Roth and Mirella Lapata. Neural semantic role labeling with dependency path embeddings. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1192–1202, Berlin, Germany, August 2016. Association for Computational Linguistics.

[3] Emma Strubell, Patrick Verga, Daniel Andor, David Weiss, and Andrew McCallum. Linguistically-informed self-attention for semantic role labeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5027–5038, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

Correlation: Semantic label distributions differ by hop patterns

- Semantic label distributions vary significantly across hop patterns
- Semantic label distributions are similar for long SSDPs

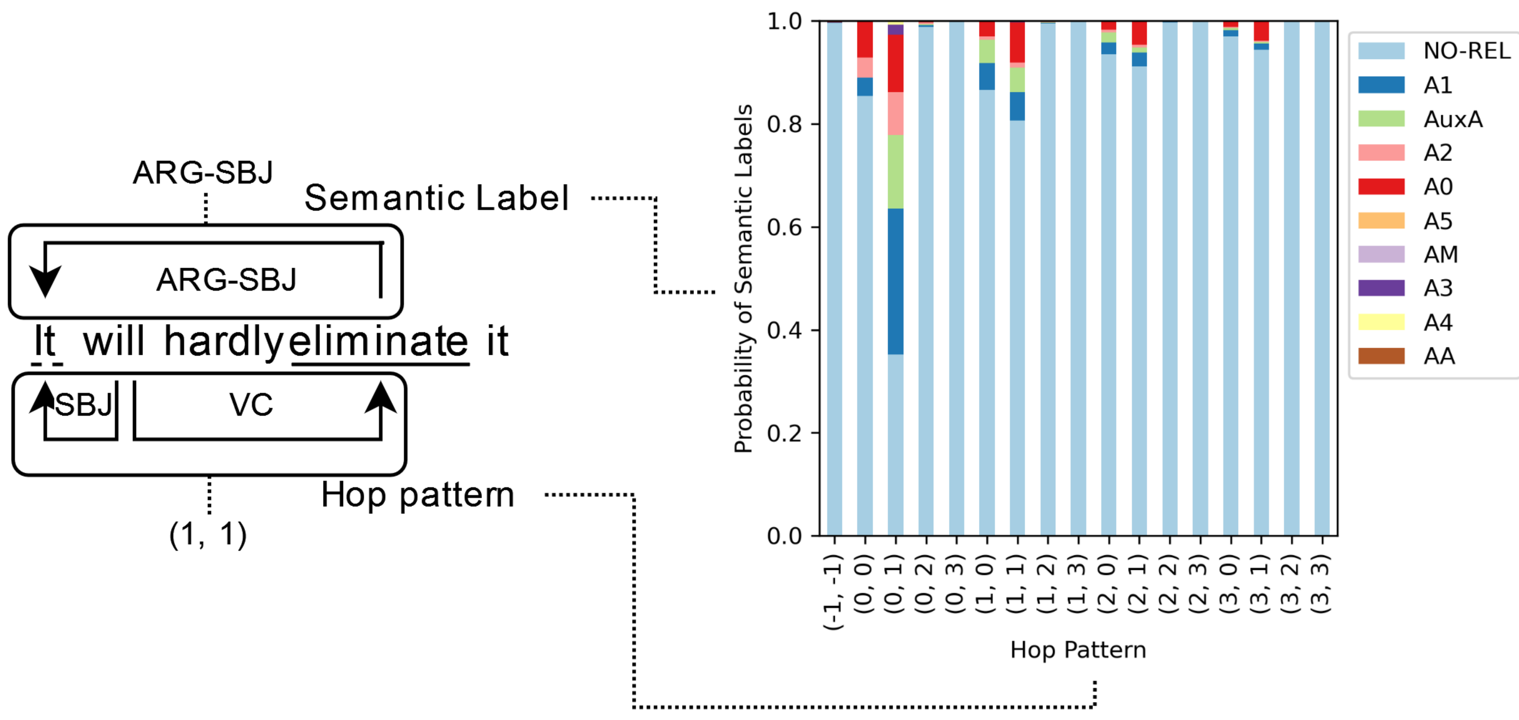

Figure 4. Semantic label distributions by hop patterns. (-1, -1) indicates hop patterns beyond (3, 3).

Mutual Information Gain of Hop Patterns

Mutual information gain measures the impact of hop patterns on semantic label distributions

Analysis:

- Hop pattern (0,1) has the highest information gain of 0.149bits
- Long hop patterns have near-zero information gains
- Short hop patterns have diverse non-zero information gains, ranging from 0.011 bits to 0.149bits

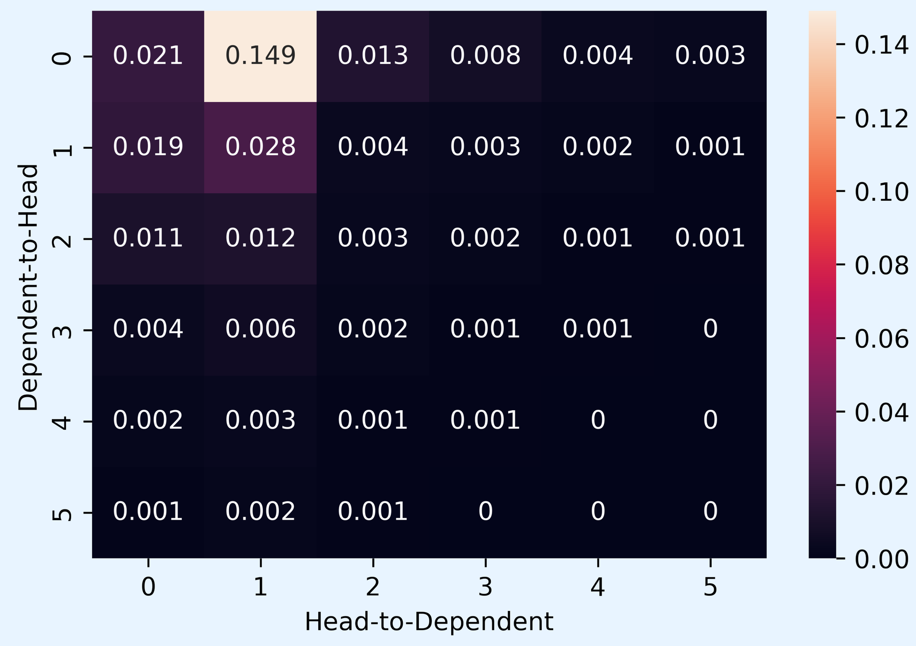

Figure 5. Mutual information gain of hop patterns up to (5, 5)

Agreeing

Results:

- Models assign a unique component to the hop pattern (0, 1), the hop pattern with the highest information gain
- Models assign long SSDPs with near-zero information gains to a single component
- Models assign SSDPs with diverse information gains to different components

(a) FastText

α \ β	0	1	2	3
0	3	1	0	0
1	2	2	0	0
2	2	2	0	0
3	2	0	0	0
4	0	0	0	0
5	0	0	0	0

(b) ELMo

α \ β	0	1	2	3
0	2	0	4	4
1	2	1	4	4
2	3	3	4	4
3	4	4	4	4
4	4	4	4	4
5	4	4	4	4

(c) BERT

α \ β	0	1	2	3
0	4	0	2	2
1	4	3	2	2
2	4	3	2	2
3	4	4	2	2
4	2	2	2	2
5	2	2	2	2

Table 1. Component assignments extracted from models

Conclusion: Semantic label distributions of different hop patterns have unique properties

CoNLL-2009 Experiments

MM outperforms baseline methods in most cases

- MM outperforms baselines over FastText, ELMo, and Bert embeddings (Figure. 6)
- MM outperforms baselines over many languages (Figure. 7)
  - MM outperforms Multitask and Transformer in German, Spanish, and Catalan
  - MM outperforms Transformer in Chinese
  - MM fails to learn in Czech
- MM improves in predicting short-distance semantic dependencies (Figure. 8)
  - MM is the only syntax-aware method improving over Transformer on FastText and ELMo embeddings
  - MM retains the advantage in long-distance semantic dependencies of syntax-aware methods

Method	Syntax-aware	Method	Syntax-aware
Transformer	No	Multitask	Yes
LISA[3]	Yes	PathLSTM[2]	Yes
Pruning[1]	Yes	MM	Yes

Table 2. List of methods

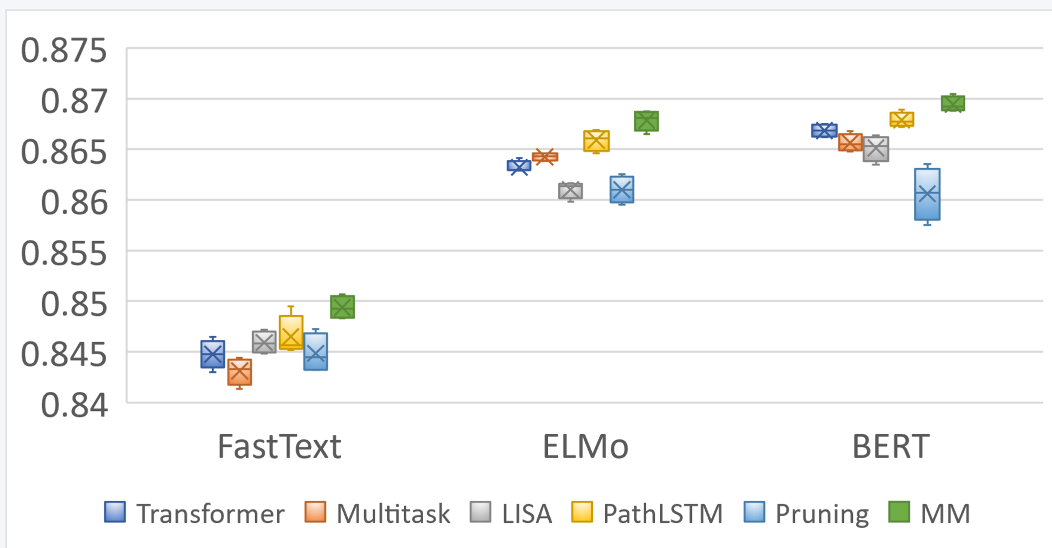

Figure 6. English LAS comparison on input embeddings.

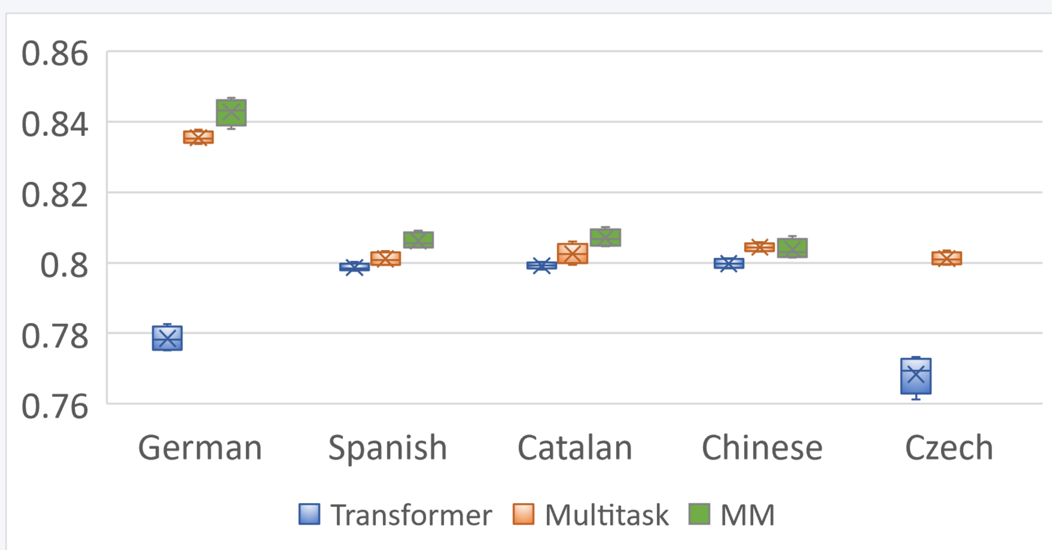

Figure 7. LAS comparison on five languages using FastText.

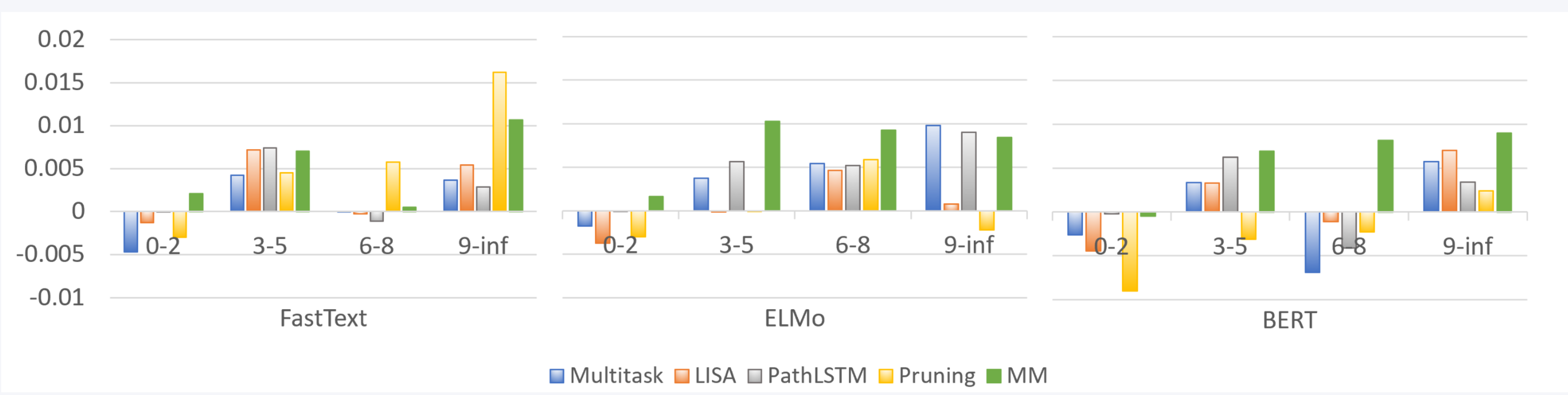

Figure 8. Relative LAS comparison by the linear distance of semantic dependencies (English, FastText).