

Appendix

Hyperparameters in Algorithm 1

Here we provide more details on the hyper-parameters used in Algorithm 1.

1. To reduce the number of parameters while also maintaining fairness, we did not tune over the number of iterations separately (i.e., the loop from Lines 2-9). We found in practice that a very small number of iterations would yield the best performance, so we finally used only 1 iteration for both Scheme I and Scheme II.
2. INFERENCE() subroutine on Line 5: We did exact inference via ILP with hard constraints, so there are no hyper-parameters. We used hard constraints since Line 7 is from partial annotations that are already assumed to be noiseless.
3. LEARN() subroutine: We fixed the parameters to be the default parameters used in CogCompNLP without tuning any of them (which were tuned on completely annotated datasets and are slightly unfair to the partial scheme II).

Example Usage of I_k Analysis in Semantic Role Classification (SRC)

One thing we did not expand in the paper was that the slope of the $I_k - I_{k-1}$ curve shown in Fig. 4 depends on the specific configurations. Take the structure required by SRC (i.e., a bipartite graph) as an example. The number of argument of a predicate will be the parameter d in Example 2. The reason we chose $d = 4$ when we plotted Fig. 4 was that the average number of arguments in SRC is roughly this number, but if we choose a larger number, then the slope of $I_k - I_{k-1}$ is steeper, as shown in Fig. 8. This is consistent with one’s intuition, since when d is very small, e.g., $d = 1$, there is hardly any structural constraint, and when d is large, the structure of the arguments for the same predicate is complicated. To verify this argument, we further tested ESPA on those predicates with more than 5 arguments, and as shown in Fig. 8, the improvement was indeed larger. The analysis here shows that the mutual information analysis introduced in this paper is a very useful metric in practice.

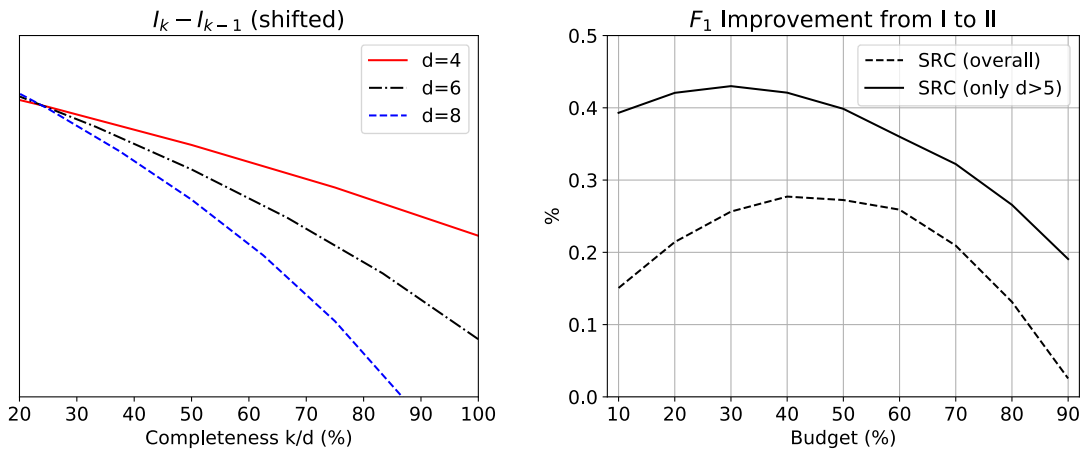


Figure 8: Left: The shape of $I_k - I_{k-1}$ for the SRC task given different k 's. Right: The improvement brought by ESPA for the entire SRC test set and for the subset of only predicates with more than 5 arguments.