

Randomized Deep Structured Prediction for Discourse-level Processing

Manuel Widmoser*¹, Maria Leonor Pacheco*², Jean Honorio², Dan Goldwasser²

¹University of Salzburg, Austria

²Purdue University, USA

*equal contribution

Problems

- Many discourse-level tasks require accounting for longer texts
[instead of just sentences]
- Learning highly expressive models is challenging
[deep structured prediction]
- Constrained inference comes at high computational cost
[the more dependencies, the higher the cost]

Contributions

- Combine expressive representations and inference
- Substitute exact inference by randomized inference
- Explore two challenging structured prediction problems:
 1. Argument Mining
 2. Stance Prediction

Argument Mining

essay

Living and studying overseas

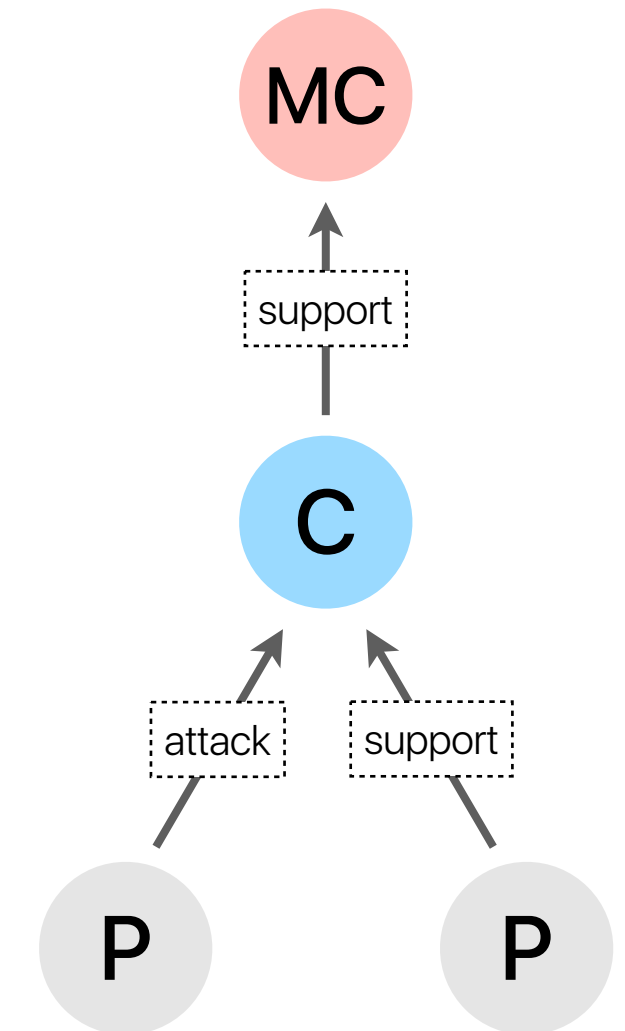
It is every student's desire to study at a good university and experience a new environment. While some students study and live overseas to achieve this, some prefer to study home because of the difficulties of living and studying overseas. In my opinion, one who studies overseas will gain many skills throughout this experience for several reasons.

First, living and studying overseas is an irreplaceable experience when it comes to learn standing on your own feet. One who is living overseas will of course struggle with loneliness, living away from family and friends but those difficulties will turn into valuable experiences in the following steps of life. Moreover, the one will learn living without depending on anyone else. Also, employers are mostly looking for people who have international and language skills. Becoming successful in this study will give the student an edge in job market. Therefore, one who has studied and lived overseas will become more eligible for the job than his/her peers.

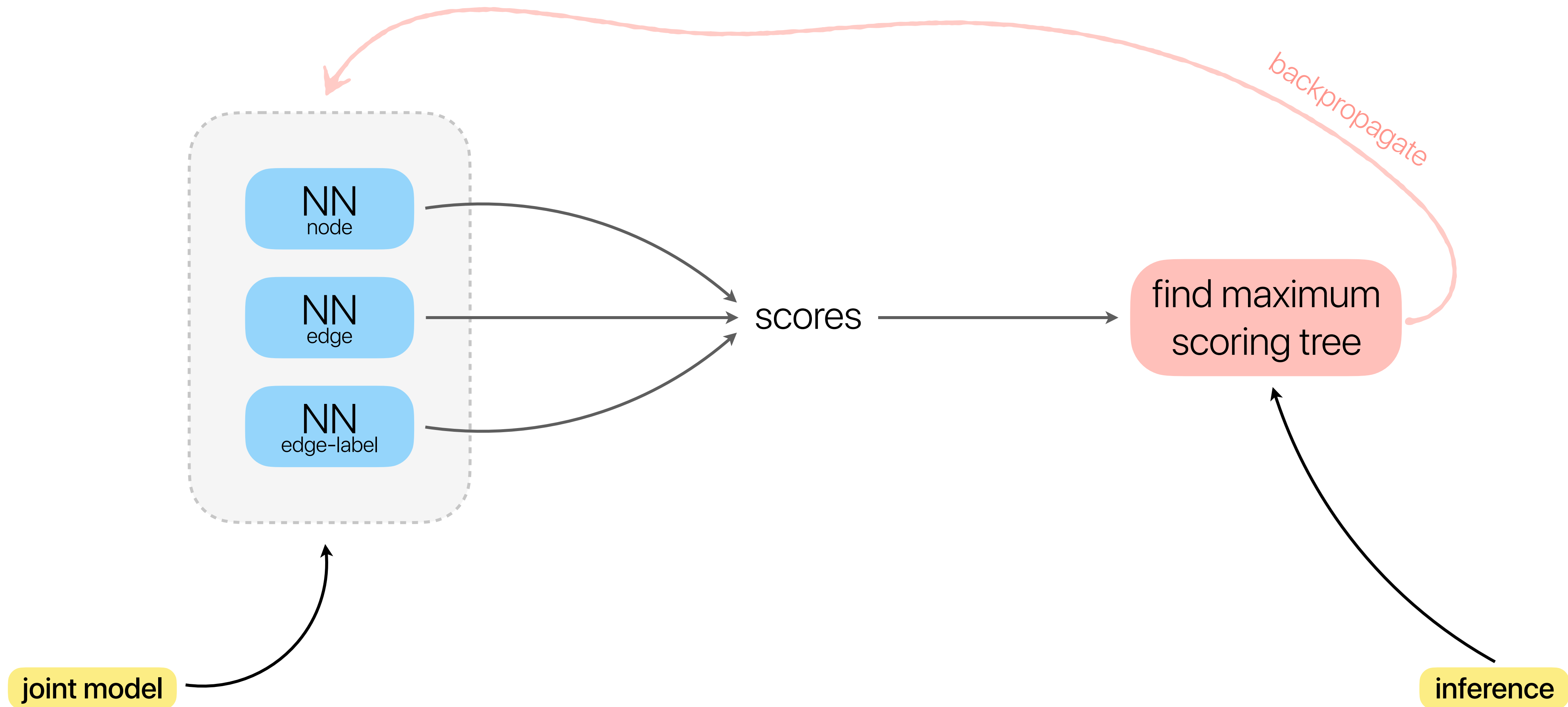
...

- $N = \{(N_1, MC), (N_2, C), (N_3, P), (N_4, P), \dots\}$
- $E = \{(E_1, (N_2, N_1)), (E_2, (N_3, N_2)), (E_3, (N_4, N_2)), \dots\}$
- $L = \{(E_1, \text{support}), (E_2, \text{attack}), (E_3, \text{support}), \dots\}$

predict those



Learning



Objective Function

The diagram illustrates the objective function $L(\mathbf{x}, \mathbf{y}, \hat{\mathbf{y}}; \theta)$ with several annotations:

- structured hinge loss**: A yellow box pointing to the $\max(\mathbf{0}, \dots)$ part of the function.
- hamming loss**: A green box pointing to the $\Delta(\mathbf{y}, \hat{\mathbf{y}})$ term.
- gold structure**: A red box pointing to the $\sum_{\psi_i \in \Psi} \rho_i(\mathbf{x}_i, \mathbf{y}_i; \theta^i)$ term.
- set of parameters**: A red box pointing to the θ parameter.
- score for potential ψ_i from a forward pass**: A blue box pointing to the $\rho_i(\mathbf{x}_i, \hat{\mathbf{y}}_i; \theta^i)$ term.

$$L(\mathbf{x}, \mathbf{y}, \hat{\mathbf{y}}; \theta) := \max\left(\mathbf{0}, \max_{\hat{\mathbf{y}} \in \mathbf{Y}} \left(\Delta(\mathbf{y}, \hat{\mathbf{y}}) + \sum_{\psi_i \in \Psi} \rho_i(\mathbf{x}_i, \hat{\mathbf{y}}_i; \theta^i) \right) - \sum_{\psi_i \in \Psi} \rho_i(\mathbf{x}_i, \mathbf{y}_i; \theta^i) \right)$$

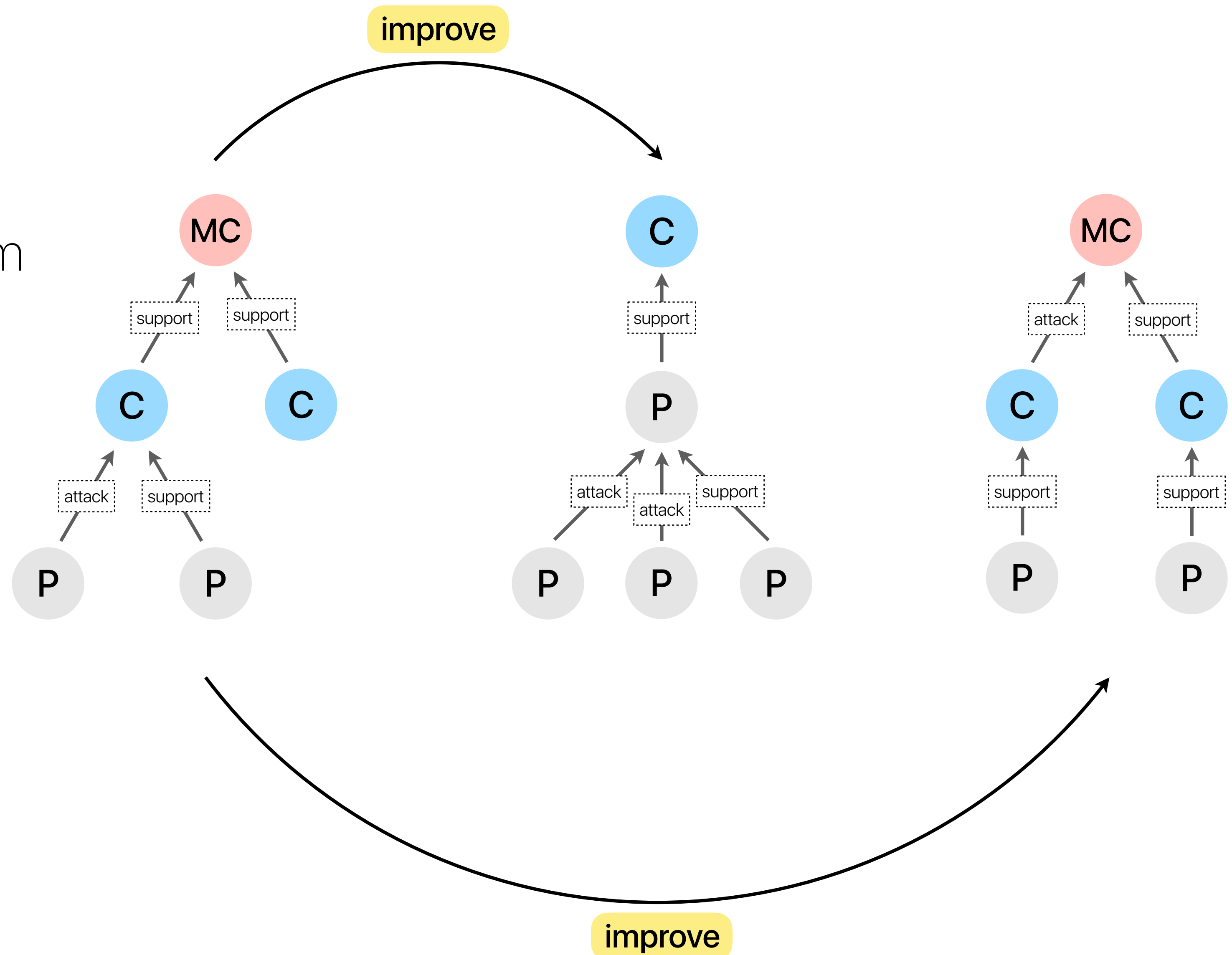
Inference

- **ILP** define inference problem as integer linear program
[exact inference]
- **AD³** translate ILP into an AD³ instance
[approximate inference]
- **Rand** randomized inference

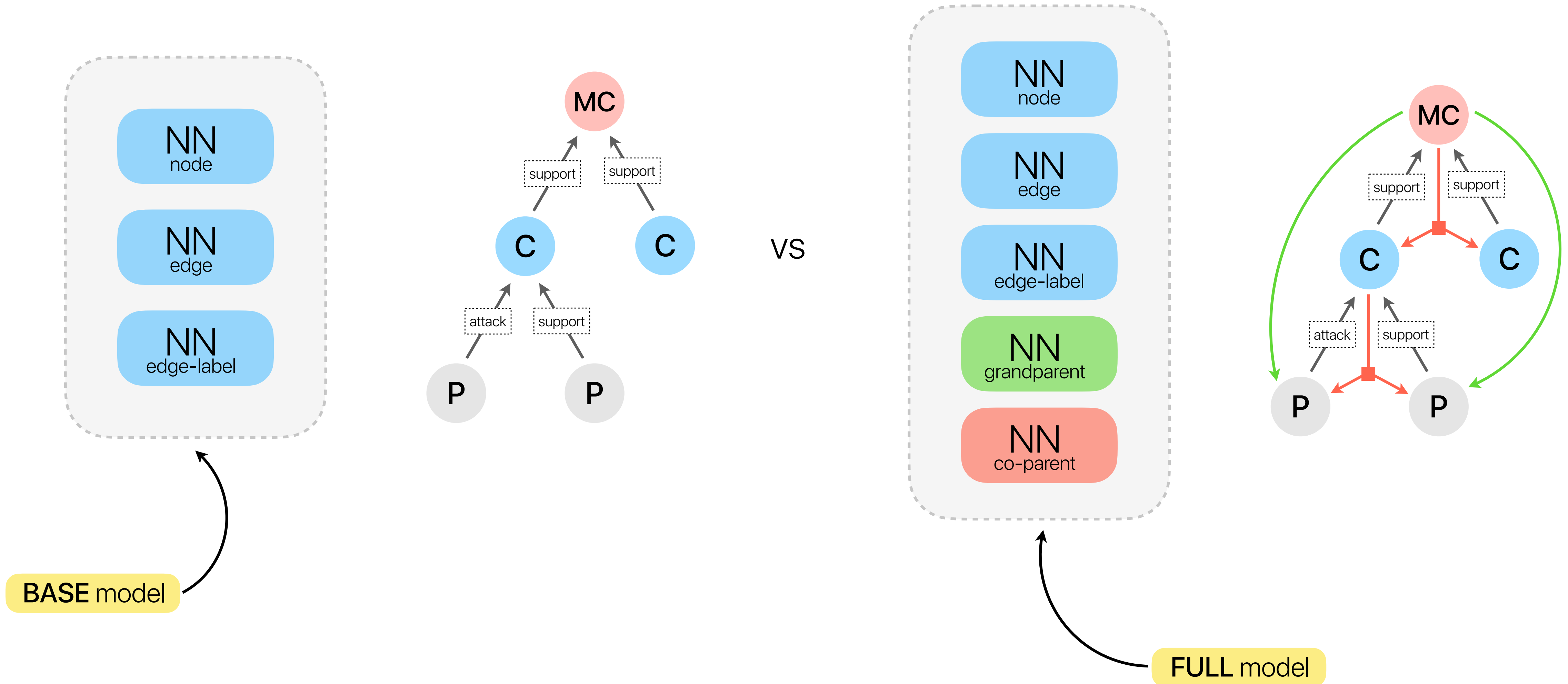
Can we get a competitive model?

Randomized Inference

- Draw tree structure at random
- Label and score the tree
- Improve greedily locally
- Repeat

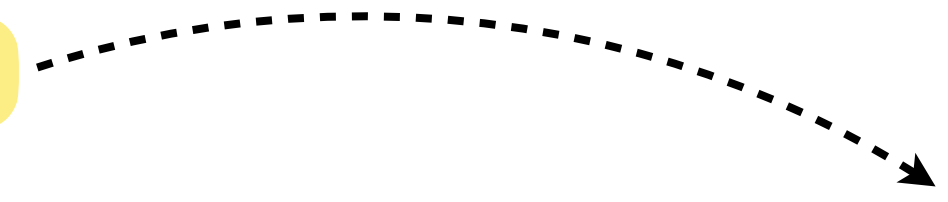


Base vs Full



Stance Prediction

thread



Does private gun ownership deter crime?

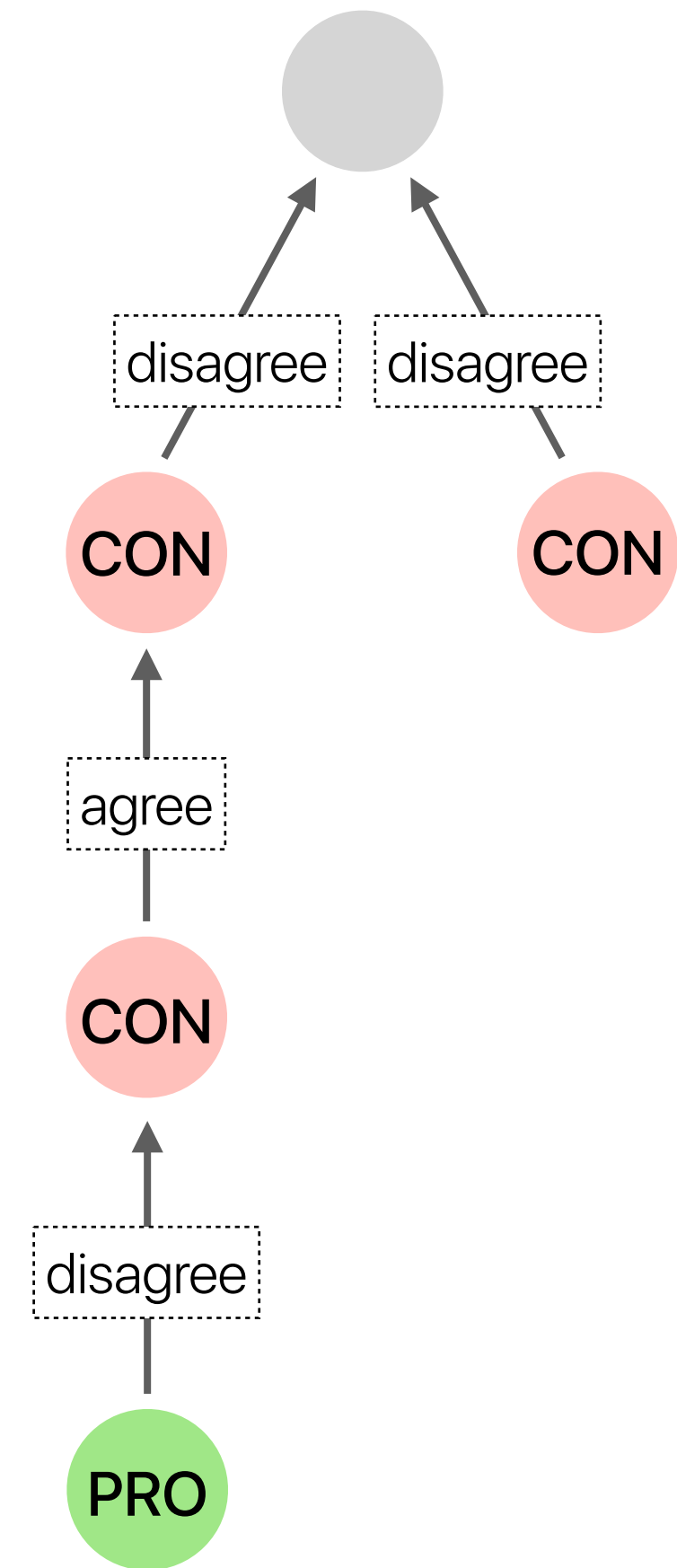
I believe yes, because a robber might not rob a house if he/she knows that there is a gun inside that house.

On its own, not so much. However when combined with laws that reimburse the private individuals and leave criminals with no options whatsoever, when it's codified that the legal gun owner was well within their rights to shoot and/or kill the person attacking them and that they face no charges, then that helps deter crime even more.

I think this is pretty much common sense. I think some study is in order, though. Everybody...put a sign in your window that says "This is a gun free home" and we will see if the incidents of robberies goes up among group members.

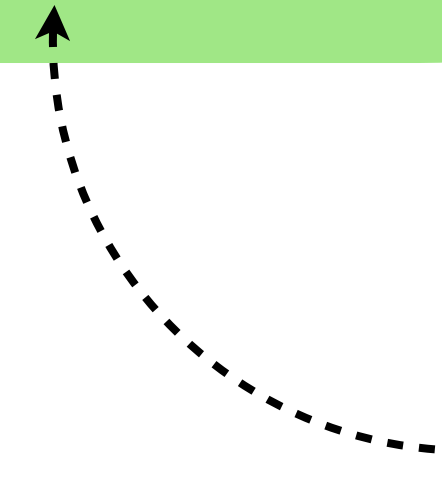
yeah but they also might not care! i mean they might have a gun themselves!

- Label nodes and edges
- Tree structure given
- Much larger trees



Base vs AC

- **BASE model** learns node and edge labels
- **AC model** enforces the same stance for all posts by the same author

 [author constraints]

Experiments: Argument Mining

		nodes	edges	edge-labels
	ILP	83.0	57.6	68.0
BASE model	AD ³	83.2	58.2	68.4
	Rand	82.8	58.4	68.4
	ILP	83.1	61.2	69.2
FULL model	AD ³	83.7	62.0	68.5
	Rand	83.8	62.6	68.4

- UKP Dataset: Stab and Gurevych, 2017: https://doi.org/10.1162/COLI_a_00295

- reported scores: macro F1 for nodes and edge-labels, positive F1 for edges

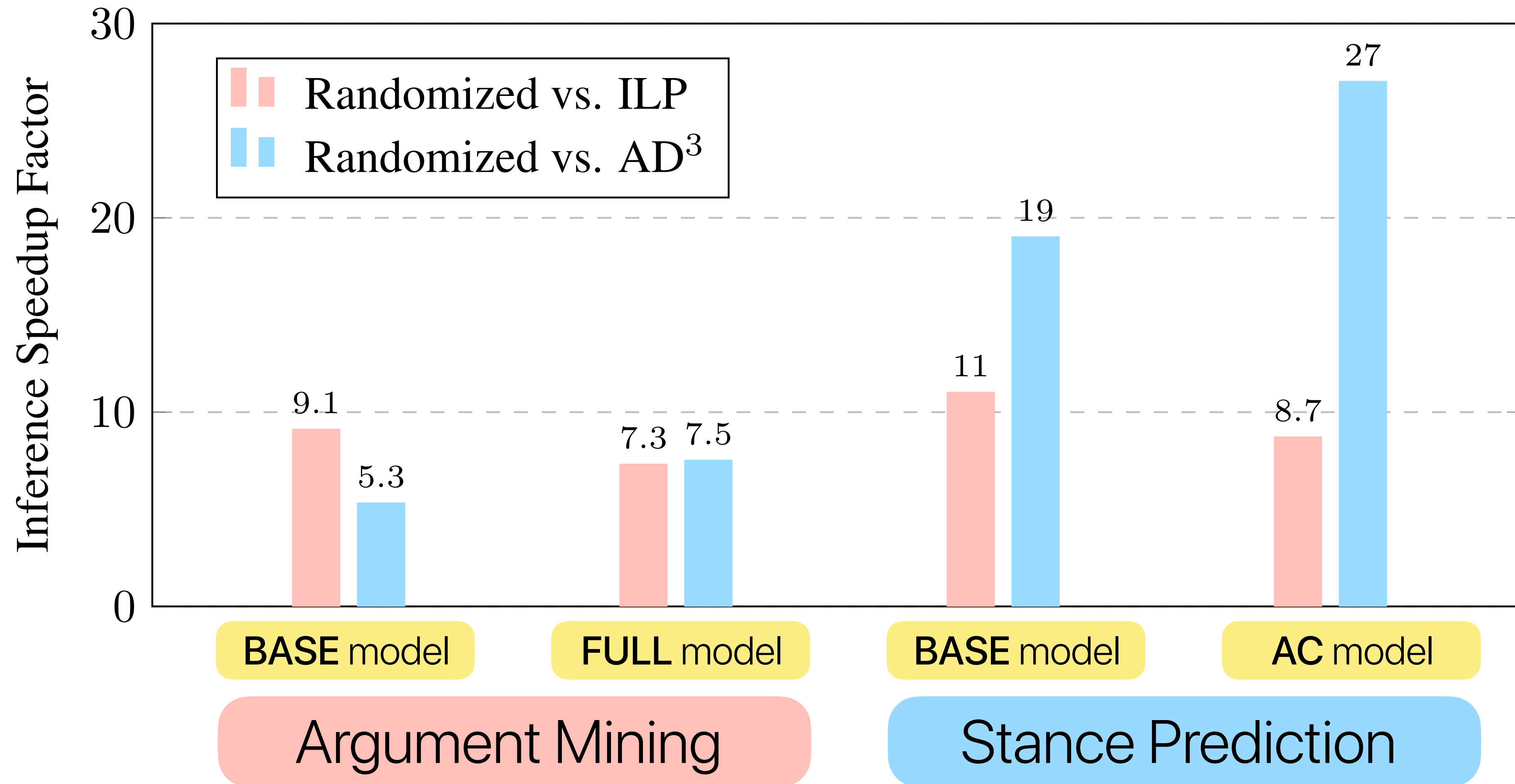
Experiments: Stance Prediction

		nodes	edge-labels
	ILP	70.7	68.3
BASE model	AD ³	70.1	68.2
	Rand	68.9	69.0
	ILP	82.3	81.5
AC model	AD ³	81.8	79.5
	Rand	81.7	80.9

- 4Forums Dataset from the Internet Argument Corpus: Walker et al., 2012: <https://www.aclweb.org/anthology/N12-1072/>

- reported scores: averaged accuracies over 4 issues (abortion, evolution, gay marriage, gun control)

Inference Analysis



Conclusion

- We studied the effectiveness of randomized inference for deep structured prediction
- Positive results for two challenging discourse-level tasks
- Highly competitive results at a lower computational cost
[an efficient alternative to exact inference]