Context-Aware Online Collective Inference for Templated Graphical Models

Charles Dickens*, Connor Pryor*, Eriq Augustine, Alex Miller, & Lise Getoor
University of California, Santa Cruz
*Equal Contribution

Introduction

Structured prediction algorithms utilize the underlying relational properties of the data to improve predictive performance and satisfy domain constraints. In this work, we examine **online collective inference**, the problem of maintaining and performing inference over a sequence of evolving graphical models for conducting online structured prediction.

Contributions

- Define and analyze online collective inference using templated graphical models.
- Derive stability bounds on MAP states of graphical models subject to model updates.
- Propose principled approximations for updating existing templated graphical models.
- Bound the loss incurred by performing approximate model updates.
- Implement an online collective inference system with Probabilistic Soft Logic.
- Empirical evaluation of methods on three real-world datasets.

Templated Graphical Models

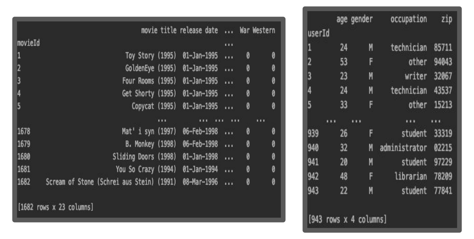
Templated graphical models (TGMs) are a general framework for defining complex probabilistic graphical models. Dependencies between variables are encoded using functions called template factors that are commonly expressed as weighted logical rule and instantiated with data.

Template Factors

```
w_d: Donates(A, P) -> Votes(A, P)
w_t: Mentions(A, "A H") -> Votes(A, "D")
w_t: Mentions(A, "T C") -> Votes(A, "R")
```

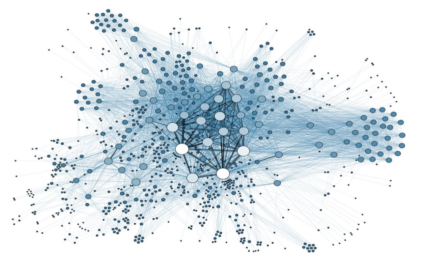
+

Data



=

Graphical Model



Probability Distribution

Instantiated TGMs define a distribution over variables Y and conditioned on variables X.

$$P(Y|X) = \frac{1}{Z(X)} \prod_{i=1}^{m} \phi_i(Y, X)$$

MAP Inference

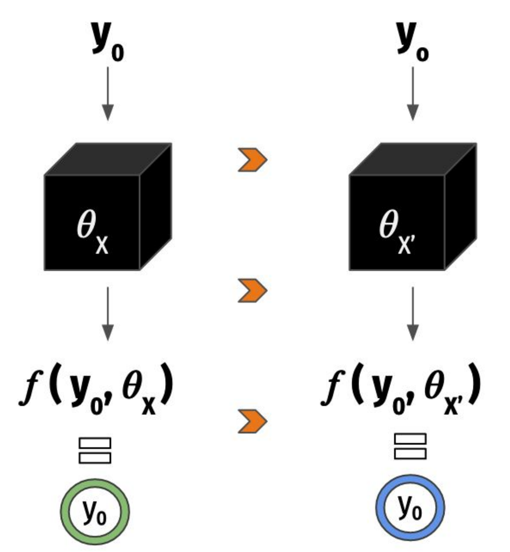
Maximum-a-posteriori (MAP) inference is performed to obtain structured predictions.

$$\mathbf{y}^* = \arg\max_{\mathbf{y}} P(Y = \mathbf{y} | X = \mathbf{x})$$

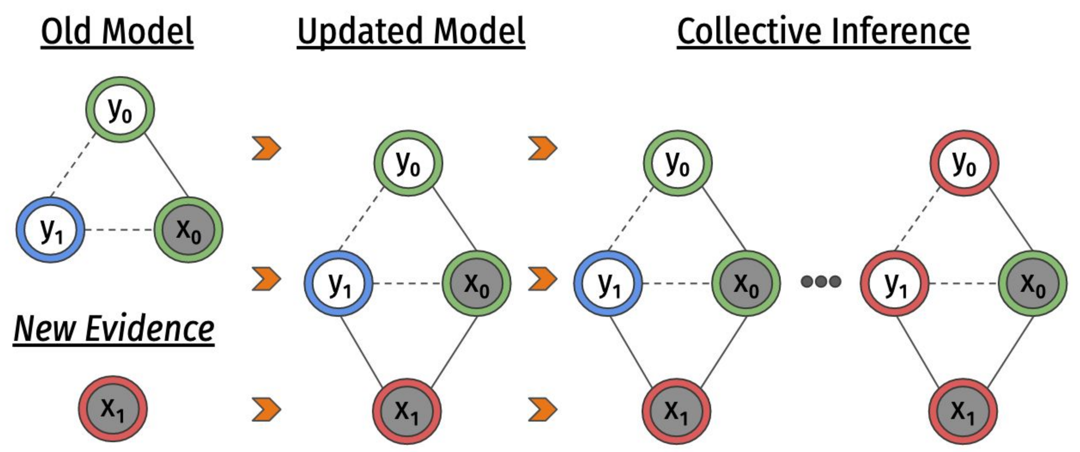
Online Collective Inference

In structured settings, predictions are not only a function of the example features and the model parameters, but they also depend on the features and predictions of other variables. New evidence has cascading effect on predictions, requiring collective inference.

Online IID

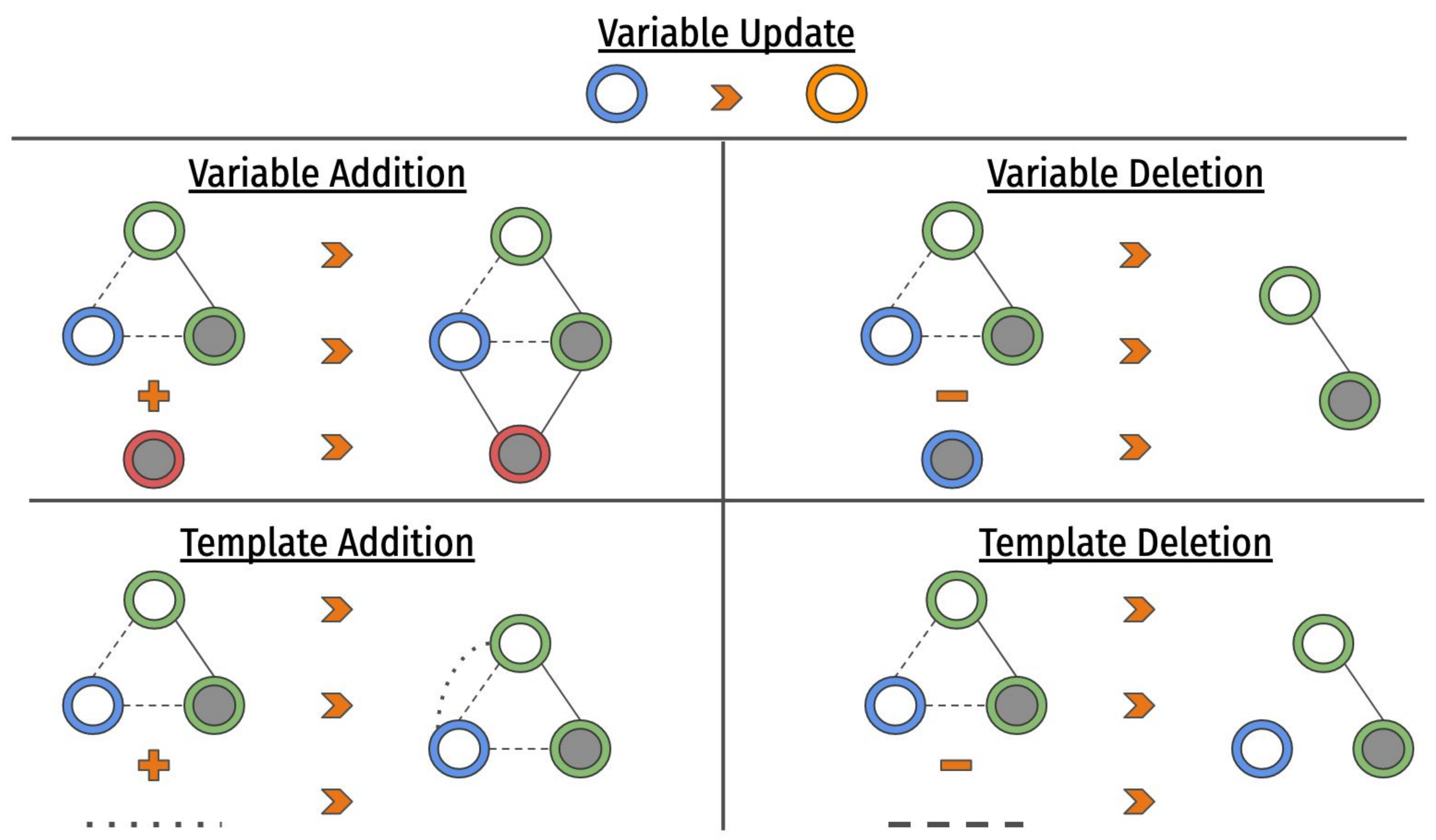


Online Collective Inference



Online TGM Updates

Five types of online TGM updates:



Stability

For TGMs instantiating distributions from a log-concave and smooth exponential family, we derive the following bounds on the distance between their respective MAP states:

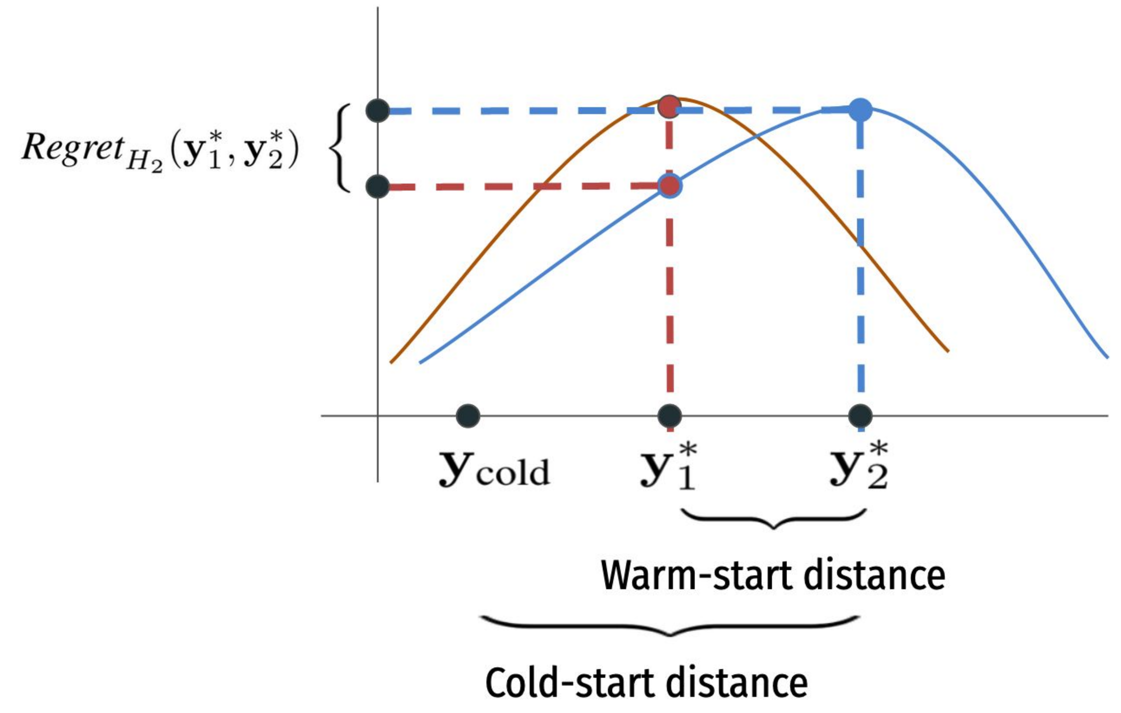
$$\|\tilde{\mathbf{y}}_1^* - \mathbf{y}_2^*\|_2 \leq 2\frac{\boxed{\beta}}{\boxed{\alpha}}\boxed{\delta} + \frac{2}{\boxed{\alpha}}\boxed{\|\nabla_{\mathbf{y}} H_{\Delta}(\tilde{\mathbf{y}}_1^*, \tilde{\mathbf{x}}_1)\|_2}$$

Strong convexity and smoothness parameters of the updated model's MAP Objective.

Terms summarizing the model updates relating two TGMs

Warm-Starts and Regret

The stability results are applied to bound the regret incurred when a model approximation is used, and to verify the advantage of using the MAP state of a related TGM as a warm-start.

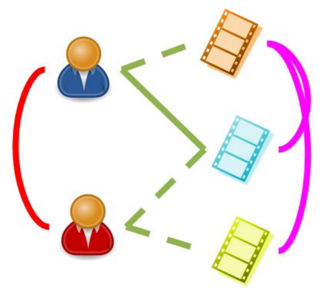


Lipschitz constant of updated model's MAP Objective

$$Regret_{H_2}(\mathbf{y}_1^*, \mathbf{y}_2^*) \leq 2\frac{\boxed{L}}{\alpha}\|\nabla_{\mathbf{y}} H_{\Delta}(\mathbf{y}_1^*, \mathbf{x})\|_2$$

Empirical Evaluation

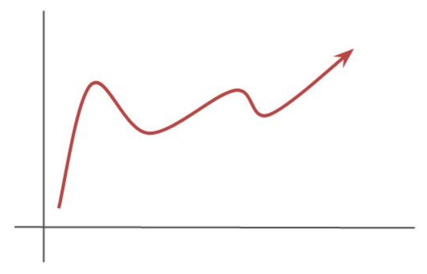
We implement an online collective inference system using **Probabilistic Soft Logic (PSL)** and test the system on 3 real-world datasets and online tasks.



Movielens 1M: Recommender System

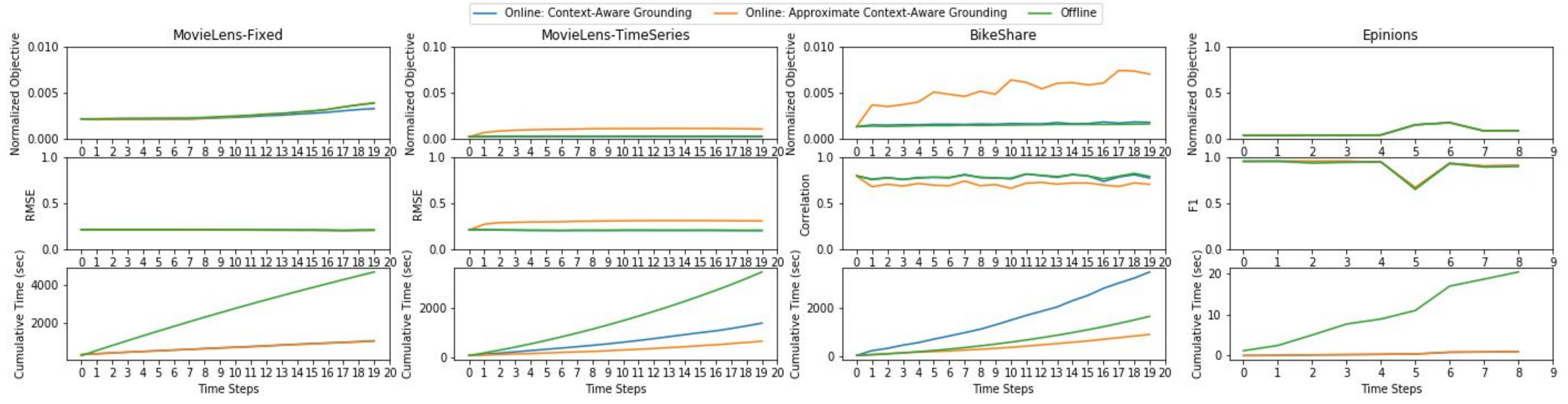
```
w_d: Donates(A, P) -> Votes(A, P)
w_t: Mentions(A, "A H") -> Votes(A, "D")
w_t: Mentions(A, "T C") -> Votes(A, "R")
```

Epinions: Rule Selection

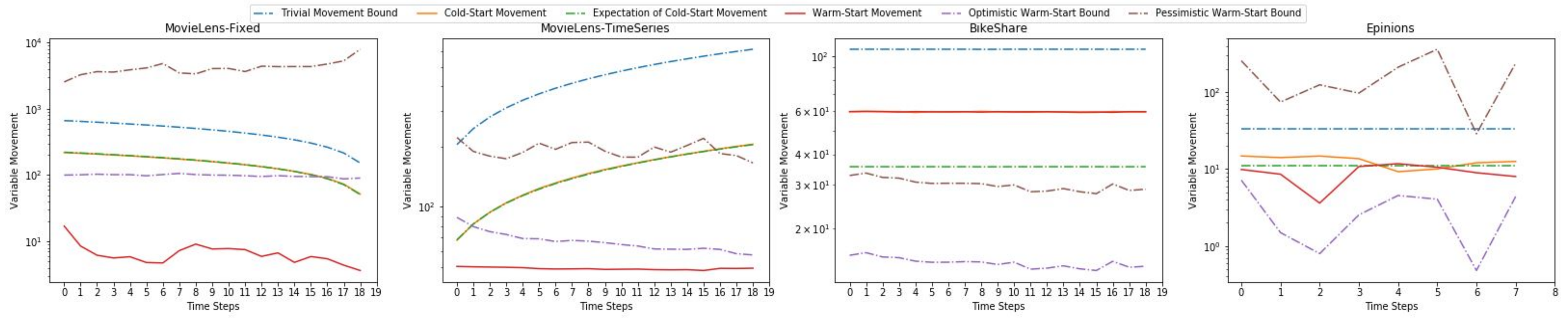


Bikeshare: Demand Forecasting

Runtime and Regret



Stability and Warm-Starts



Conclusions and Future Work

Conclusions

- Approximate methods consistently yield 2-5 times speedups over offline models and their regret can be bounded by the complexity of the model updates and approximation techniques.
- Exact methods result in 2-5 times speedups in common online settings but current context aware algorithms can be slower for some sequences of model updates.
- MAP states of related TGMs are good warm starts.

Future Work

- Explore more approximate grounding techniques
- Integrate approximate inference algorithms to achieve real-time predictions
- Online learning of template factor parameters
- Reduce memory requirements by summarizing or forgetting portions of model