Dynamic Bayesian Rule Learning for Interpretable Time Series Prediction

Alexander L. Hayes   Lucas Newman-Johnson   David Crandall

Computer Vision Lab, Indiana University, Bloomington

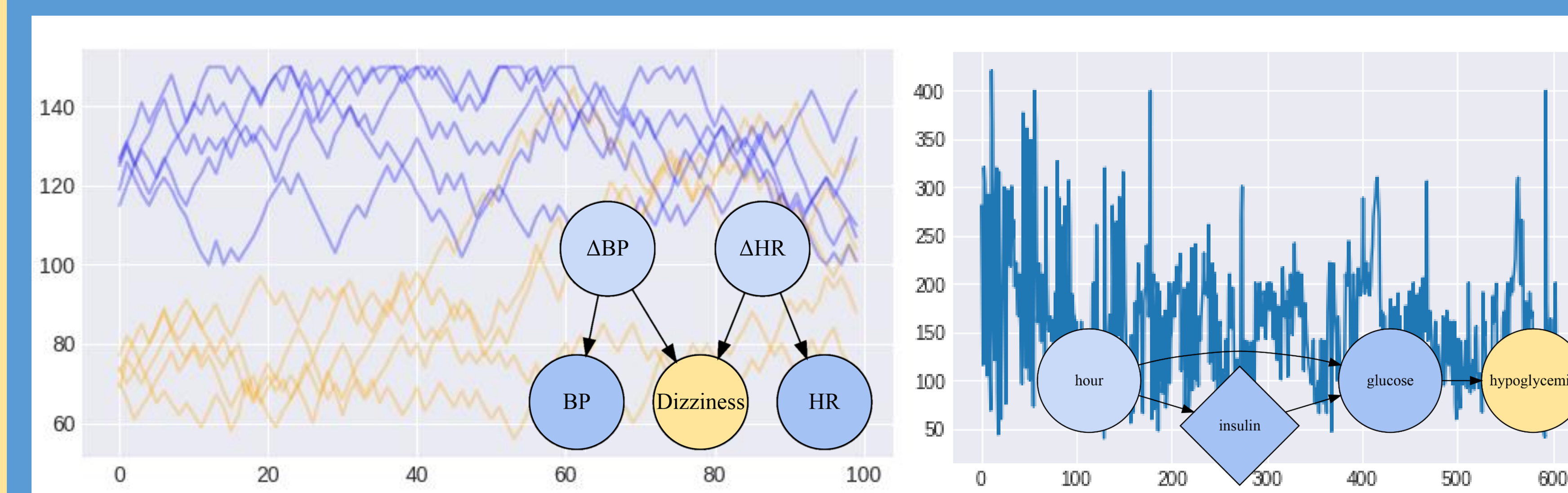
All prediction problems involve time

Many need to model time explicitly: weather, object tracking, or outcomes of clinical trials. We're interested in ***learning interpretable models***[1] for how variables interact to influence an outcome. We extend ideas for learning constrained Bayesian networks[2] to the time series domain, resulting in ***Dynamic Bayesian Networks*** from which we extract ***decision lists*** and likelihoods from.[3]

```
IF X THEN Y:
   Likelihood = 2.0
```

We motivate with two datasets

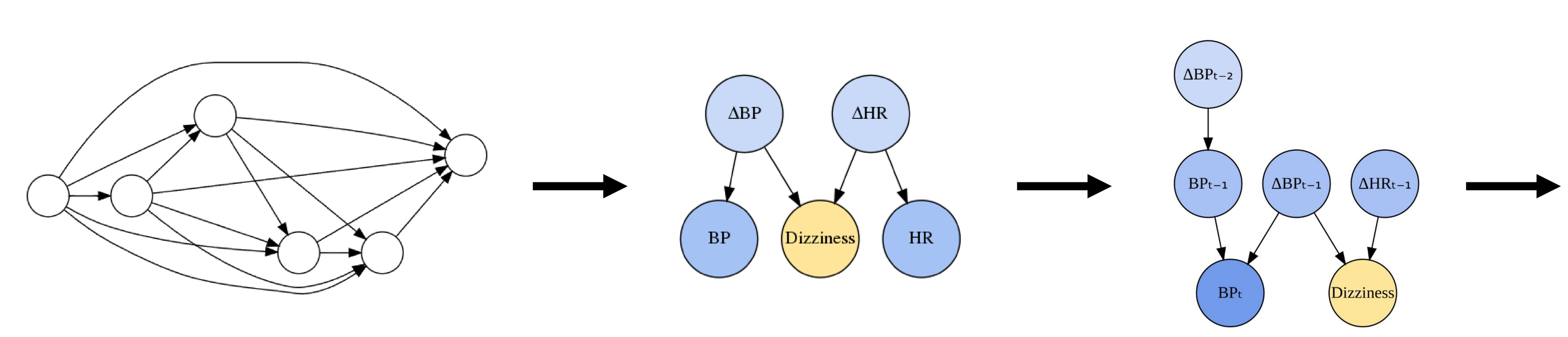
(1) A synthetic set where symptoms are caused by changes in blood pressure and heart rate, and (2) the UCI Diabetes dataset that tracked glucose, insulin doses, and hypoglycemia symptoms.



(Left) Five users from the synthetic set: heart rate in blue, blood pressure in orange.
(Right) User 55's blood glucose measurements over six months.

Extracting *Dynamic Bayesian Rules* (DBRs)

The order of observations implies a ***constraint graph*** on how variables can interact. We fit a structure on (time, variable, value) triples, unroll time steps into a Bayesian network, perform local search on a target variable, and extract rules from the final structure.



(1) *Constraint Graph*   (2) *Dynamic Bayes Net*   (3) *Bayes Net*

```
IF (ΔBP_t-1 in (3.0 - 8.0) ^ ΔHR_t-1 in (-8.0 - -3.0)) THEN (Dizziness = True)
    Likelihood = 21.71
ELSE (Dizziness = False)
    Likelihood = infinite
```

(4) *Dynamic Bayesian Rules*

Learn more about the last two steps:

We define success as *maximizing accuracy* and *minimizing the number of rules*

Decision trees learn CNF rules, so we compare with trees learned on the trajectories under cross validation. DBRs are shorter, but may be less accurate in settings with limited training data.

		Decision Tree (Baseline)	Dynamic Bayesian Rules
Dizziness (Synthetic, 1500 examples)	# Rules (Mode)	4	**3**
	Accuracy (5-fold CV)	93.6	**97.7**
Hypoglycemia (UCI Diabetes dataset. User 55. 10 examples)	# Rules (Mode)	5	**3**
	Accuracy (LOOCV)	**80.0**	60.0

Future Work

Incorporating interventions could be a window into learning causal relationships, and advice about context-specific independences may be helpful in the limited data settings.

References

[1] Rudin, C. *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead.* Nature Mach Intelligence, (2019)
[2] Schreiber JM and Noble WS. *Finding the optimal Bayesian network given a constraint graph.* PeerJ Computer Science, (2017)
[3] Gopalakrishnan, V., et al. *Bayesian rule learning for biomedical data mining.* Bioinformatics, Volume 26, Issue 5, (2010)

This work was funded by the US Department of Defense (Contract W52P1J2093009).