Retreat Proposal

Dynamic Structure Learning

16.01.2015

1 Motivation

Many real world Machine Learning datasets are represented as a collection of time series:

- activation scans (EEG or fMRI) from different parts of the brain;
- the aggregated activity of different people/groups/topics in social networks;
- the market share of companies within different market segments, etc.

In the times of Big Data the number of these time series and their relationships can be very complex, and sometimes acquiring extra data incurs high cost. Therefore, automated techniques are required to process the data available and find all the hidden dependencies. To improve upon the learned model, it's also important to automate this process from an active learning perspective, such that the learner is not restricted to making passive observations, but can choose data wisely to minimize the data acquisition cost.
One of the ways to understand the underlying data relationships is to learn a graphical model structure (Markov or Bayesian network) "explaining" the behaviour of a joint collection of time series [3, 7].
For the examples above the hidden dependencies and structure are the following:

- activations in some parts of the brain cause activations in others, these relations are still unknown and observing them (or deviations from the normal behaviour) can lead to the solution of many biological and medical questions [1];
- some people/groups/topics influence others, and finding such influencers allows to either predict or control social media [2];
- market shares are competitively dependent both within and between different segments, and probably addressing only few segments can change the global picture.

Using graphical models formalism one can explain all the dependencies among given time-series data, and make them interpretable (*this* causes *that*) and quantitatively valuable (change *this* by X and *that* will change by Y). In this project, we will first focus on modeling such structure learning problem from a *passive learning* perspective. Given that we can obtain an initial estimation on the underlying structure, we will then consider the problem of active learning for structured learning, specifically in the domain of time-series data analysis.
Keywords: Bayesian Networks, Gaussian Structural Learning, Dynamic Causal Modelling, Graphical Lasso, TIGER, Regularised Gaussian Processes, Stability Selection, etc.

2 Approaches

There are several approaches to modelling either a one time series or several time series with interdependencies:

- Vector autoregression VAR(p) [9], and its special case VAR(1): $X_t = \Phi X_{t-1} + Z_t$.

- Graphical Lasso [4] for estimating inverse covariance directly, assuming the time series elements to be i.i.d., $X_t \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Allows to infer only coordinate structure without the time structure, but on the other hand allows stability selection procedure [6].

- Dynamic Causal Modelling [1] allows to utilize the ODE approach, is widely used when a certain *control time series* is available: $\mathcal{D}(X_t, U_t, \boldsymbol{\theta}) = 0$, where $\mathcal{D}$ denotes a (linear) ODE of the response X_t and control U_t, parametrized by $\boldsymbol{\theta}$. Used mostly in brain signal analysis.

- Continuous-Time Diffusion Networks [2] (used mostly for social networks analysis).

- Regularized regression from X_{t-1} to X_t, such as Gaussian Processes. Regularization is incorporated for graph sparsity.

- Learning Bayesian Networks through Markov blankets [7] (a more combinatorial set of approaches – stands a little separated).

- Active learning of causal structures from interventions [5], and active learning of interaction networks through value injection queries [8].

3 Goals

Immediate goals:

- See the global picture: how all methods are related, what are their assumptions, can one be replaced with another?

- Modify Graphical Lasso approach to model conditional distribution $X_t|X_{t-1}$. Compare it with other approaches, such as regularised Gaussian Processes.

- See how the stability selection [6] is modified when applied to conditional distributions $X_t|X_{t-1}$.

- Look for relations between Continuous-Time Diffusion Networks and other approaches.

- Elaborate on the ideas of active learning for the structured data.

- See, whether it is possible to apply the techniques used in one field to techniques from another (such as stability selection).

Long term goals:

- Develop a framework for comparing and validating structural learning in a dynamic setting.

- Mix methods for structure learning: take the best of different worlds.

- Try approaches on different datasets to address the problems of the people in the institute: social graphs, active learning settings, brain EEG and brain fMRI, available market data, etc.

- Write many papers :)

References

[1] Kay H Brodersen, Thomas M Schofield, Alexander P Leff, Cheng Soon Ong, Ekaterina I Lomakina, Joachim M Buhmann, and Klaas E Stephan. Generative embedding for model-based classification of fmri data. *PLoS computational biology*, 7(6):e1002079, 2011.

[2] Nan Du, Le Song, Manuel Gomez-Rodriguez, and Hongyuan Zha. Scalable influence estimation in continuous-time diffusion networks. In *Advances in Neural Information Processing Systems*, pages 3147–3155, 2013.

[3] Byron Ellis and Wing Hung Wong. Learning causal bayesian network structures from experimental data. *Journal of the American Statistical Association*, 103(482):778–789, 2008.

[4] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[5] Alain Hauser and Peter Bühlmann. Two optimal strategies for active learning of causal models from interventional data. *International Journal of Approximate Reasoning*, 55(4):926 – 939, 2014. Special issue on the sixth European Workshop on Probabilistic Graphical Models.

[6] Nicolai Meinshausen and Peter Bühlmann. Stability selection (with discussion). *Journal of the Royal Statistical Society: Series B*, 72:417–473, 2010.

[7] Radhakrishnan Nagarajan, Marco Scutari, and Sophie Lèbre. *Bayesian Networks in R*. Springer, 2013.

[8] Lev Reyzin. *Active Learning of Interaction Networks*. PhD thesis, Yale University, New Haven, Connecticut, 2009.

[9] Eric Zivot and Jiahui Wang. Vector autoregressive models for multivariate time series. *Modeling Financial Time Series with S-PLUS®*, pages 385–429, 2006.