

Graphical Recurrent Inference Networks

Roger Milroy
Department of Computer Science
Royal Holloway, University of London

2019
November

1 Introduction

The technique introduced by the paper Combining Generative and Discriminative Models for Hybrid Inference [1] is aimed at maximising the relative strengths of two separate approaches to the same problem. The first approach is the traditional mathematical approach of formulating a model that describes the phenomenon and then using that model to predict on new information. We consider the model to be correct when it accurately describes new data. The other approach is that exemplified by Deep Learning. That is creating some architecture of a Neural Network (or other algorithm if we look at Machine Learning in general) and training an explicit mapping from data to result. There are advantages and disadvantages to both, and the aim of the Graphical Recurrent Inference Network is to maximise the relative strengths of each and thus achieve results better than either even in the regimes in which they are best suited.

2 Graphical Recurrent Inference Networks

One key challenge in combining approaches is that often they are in different and in the worst case mutually incompatible representations. This happens when the abstraction that one approach uses cannot work with the other technique and vice versa. Fortunately many mathematical models of real world dynamical systems can be formulated or reformulated into a graphical interpretation. At the same time Graphical Neural Networks have been developed that allow inference on arbitrary graphs. So what the authors do is they create two graphs. The mathematical model as one graph and another copy of the graph which is what the GNN will perform inference on. In addition to this, they use a Gated Recurrent Unit (GRU) in order to take information from the previous time step into the next. The output of the GNN is then combined into the prediction of the mathematical model, meaning that the GNN is left the task of modeling and predicting the residual error. This can be caused by Gaussian noise or from

aspects of the dynamic system that the mathematical model has ignored for the sake of simplicity.

This structure is exemplified well on Kalman filters which is the key mathematical model used in my project. In fact the authors of the paper used a Kalman Filter as one of the examples where it outperformed both plain Kalman Filters and explicit mappings. Note while the explicit mapping was able to reach the same level of error, it took much more data to do so.

3 Analysis of the Technique

This technique is particularly interesting as it is the first technique (that I am aware of) that manages to fuse existing knowledge in a domain area with the advantages that Deep Learning has shown in regimes with large amounts of data. In the past the expert domain knowledge was usually deployed in feature engineering. Trying to manipulate the data to fit a clean linear classifier.

That is not to say that the technique has no flaws or challenges. It is certainly a computationally expensive technique. This is due mainly to the use of RNN components (the Gated Recurrent Unit or GRU). And the subsequent need to aggregate several cycles of updates in order to make predictions. In the context of my project, it is doubly expensive due to the use of the Extended Kalman Filter for sensor fusion. This is computationally expensive on its own due to the need for continuous Taylor Expansions. Adding this technique on top adds to that and may be overly expensive for onboard computation.

References

- [1] GARCIA SATORRAS, V., WELLING, M., AND AKATA, Z. Combining generative and discriminative models for hybrid inference. In *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 13802–13812.