

Structured Prediction in Time Series Data

Jia Li

Department of Computer Science
 University of Illinois at Chicago
 Chicago, IL, 60607

Introduction

Time series data is common in a wide range of disciplines including finance, biology, sociology, and computer science. Analyzing and modeling time series data is fundamental for studying various problems in those fields. For instance, studying time series physiological data can be used to discriminate patients' abnormal recovery trajectories and normal ones (Hripcsak, Albers, and Perotte 2015). GPS data are useful for studying collective decision making of group-living animals (Strandburg-Peshkin et al. 2015).

There are different methods for studying time series data such as clustering, regression, and anomaly detection. In this proposal, we are interested in structured prediction problems in time series data. Structured prediction focuses on prediction task where the outputs are structured and interdependent, contrary to the non-structured prediction which assumes that the outputs are independent of other predicted outputs. Structured prediction is an important problem as there are structures inherently existing in time series data. One difficulty for structured prediction is that the number of possible outputs can be exponential which makes modeling all the potential outputs intractable.

Related Work

For classification and prediction task, one common approach is to treat each data point in a time series independently. This is typically known as non-structured prediction. Some classic approaches include support vector machines (SVM), logistic regression, and decision tree. These approaches do not consider the structure in the data. Hidden Markov Model (HMM) is a generative model proposed to solve the structured predictions tasks. It takes the transition probability between different timesteps into account. Maximum Entropy Markov Model (MEMM) is a discriminative model combining HMM and maximum entropy model. In MEMM, it predicts the most likely label at the current timestep only considers the history which leads to a problem known as label bias. Conditional Random Fields (Sutton and McCallum 2010), as a discriminative model, is proposed to solve the label bias problem. This model uses a global normalizer which provides the best possible sequence label. Structured SVM

(Tsochantaridis et al. 2004) is a maximum margin approach for structured prediction.

These approaches are all based on the idea of empirical risk minimization (ERM). However, optimizing many loss measures such as hamming loss under ERM is NP-hard. Thus, most existing approaches use approximate convex loss function to replace the original one. For example, CRF uses log-loss and Structured SVM (SSVM) uses hinge loss. Unfortunately, such replacements lead to mismatch between the predictions with the applications. Recent adversarial methods (Asif et al. 2015) reverse the thinking of empirical risk minimization by instead optimizing exact performance measures on approximate training data.

Problem Definition

In this paper, we focus on the following problem:

Given: training data instances with input $\mathbf{X}_{1:t}$ which is a time series data from timestep 1 to timestep t , and corresponding structured outputs \mathbf{y} .

Learn: a model θ which provides a prediction $\hat{\mathbf{y}}$ on unseen data that minimizes some desired loss function $loss(\hat{\mathbf{y}}, \mathbf{y})$, where \mathbf{y} are the correct outputs.

Current Progress

We studied two problems in time series data: sequence tagging and dynamic network prediction.

Sequence tagging is the problem of assigning a class label to each timestep in a time series. Activity recognition, which is one example of sequence tagging, translates raw sensor data (GPS, accelerometer data) into higher level activity labels such as walking and running. In sequence tagging, one loss measure that people care about is hamming loss which has very direct interpretation as the number of timesteps that the predictor is mismatched with the correct labels. In (Li et al. 2016), we proposed an adversarial sequence tagging approach. Instead of choosing a predictor's parametric form and using ERM on training data to select its parameters, we obtain the predictor that performs the best for the worst-case choice of conditional label distributions that match statistics measured from available training data. Under such formulation, the original non-convex and NP hard optimization problem can be solved efficiently. For a sequence with length L and $|Y|$ possible labels for each position, the number of possible labels for the sequence is $|Y|^L$. As a result,

even numerating the whole adversarial game is not possible. We used double oracle and single oracle (McMahan, Gordon, and Blum 2003) approaches for solving this problem. These two approaches both use constraint generation methods to incrementally produce a sub-game of the exponentially large full game. In our experiments, we used two activity recognition datasets. One uses sensors for human activities and the other one uses GPS and accelerometer sensor for baboon activity recognition. Also, we showed the proposed adversarial sequence tagging approach can be applied to natural language process tasks where there are many sequence tagging problems such as Name Entity recognition. We demonstrate superior performance compared to CRF and SSVM. Additionally, we proved that our proposed approach has a nice property which is Fisher Consistency. Specifically, if the predictor is provided with the actual data distribution and an arbitrarily rich feature representation, the resulting predictor will minimize the expected loss. On the contrary, SSVM does not provide such property guarantee.

Dynamic networks are useful for many interesting tasks such as modeling social interactions, information diffusion, and product recommendation. Dynamic network prediction predicts the network topology given the history of the network. Existing approaches (Lahiri and Berger-Wolf 2007; Dunlavy, Kolda, and Acar 2011) do not explicitly define an error measure which is being optimized. Additionally, although different measures such as AUC, precision, recall, F1-score are used for evaluation, it is not clear whether the proposed methods are actually directly optimizing these measures. Additionally, most of the current approaches predicts each edge independently despite networks are very complex objects. We generalize the adversarial prediction to dynamic network prediction setting where the edges can be predicted jointly and demonstrate the benefits of our approach real dynamic network datasets.

Future Plans

We propose two promising problems to solve.

Active Learning in Time Series Data

Supervised learning usually requires a significant amount of labeled data to perform well. However, obtaining the labels can be expensive and tedious. Active learning actively chooses some data points to be labeled to achieve similar performance with supervised learning methods while reducing the number of labeled data points needed. Active learning for non-structured prediction has been extensively studied (Settles 2012). For structured prediction in time series data, there are more challenges. First of all, it is essential to distinguish whether sequences require entire (all timesteps in the sequence) or partial (subset of timesteps) labeling. Partial labeling can significantly reduce the labeling effort. Existing work, which is able to handle partial labeling, labels each position in the sequence independently. While this makes sense for some tasks in NLP such as translation where people can still understand one word. However, in time series data, it is much hard to label without context. Thus, there is a trade off between reducing the labeling efforts while

keeping enough context for the experts. Second, switching between sequences and different timesteps in the same time series incurs additional cost. Labeling consecutive timesteps can further reduce labeling effort. Third, the uncertainty from one part of the sequence should not impact the other parts of the sequence.

Jointly Modeling Networks and Attributes

Many networks not only have nodes and edges, but also attributes associated with nodes. For example, individual activity is one attribute for the nodes in animal networks. Activity and network are not independent events. For instance, when animals are feeding or in coordinated movement, they have different social network structures. At the same time, their individual activities are also different. The network structure will impact the changes of attributes, and vice versa. Jointly modeling attributes and networks has the potential to improve the accuracy of both tasks, thus it is an important and open problem to explore.

Acknowledgments

The research in this paper was supported in part by the NSF grants III-1514126 (Ziebart, Berger-Wolf), CNS-1248080 (Berger-Wolf), and RI-1526379 (Ziebart).

References

- Asif, K.; Xing, W.; Behpour, S.; and Ziebart, B. D. 2015. Adversarial cost-sensitive classification. In *International Conference on UAI*.
- Dunlavy, D. M.; Kolda, T. G.; and Acar, E. 2011. Temporal link prediction using matrix and tensor factorizations. *TKDD* 5(2):10.
- Hripcsak, G.; Albers, D. J.; and Perotte, A. 2015. Parameterizing time in electronic health record studies. *Journal of the American Medical Informatics Association* ocu051.
- Lahiri, M., and Berger-Wolf, T. Y. 2007. Structure prediction in temporal networks using frequent subgraphs. In *CIDM 2007.*, 35–42. IEEE.
- Li, J.; Asif, K.; Wang, H.; Ziebart, B. D.; and Berger-Wolf, T. 2016. Adversarial sequence tagging. In *IJCAI*.
- McMahan, H. B.; Gordon, G. J.; and Blum, A. 2003. Planning in the presence of cost functions controlled by an adversary. In *ICML*, 536–543.
- Settles, B. 2012. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 6(1):1–114.
- Strandburg-Peshkin, A.; Farine, D. R.; Couzin, I. D.; and Crofoot, M. C. 2015. Shared decision-making drives collective movement in wild baboons. *Science* 348(6241):1358–1361.
- Sutton, C., and McCallum, A. 2010. An introduction to conditional random fields. *arXiv preprint arXiv:1011.4088*.
- Tsochantaridis, I.; Hofmann, T.; Joachims, T.; and Altun, Y. 2004. Support vector machine learning for interdependent and structured output spaces. In *Proceedings of the twenty-first international conference on Machine learning*, 104. ACM.