Predictive Modeling of Spatio-Temporal Sensor Network Data using Markov Random Fields

Manuel Claeys Bouuaert[1], Tijs Neutens[1], Nico Van de Weghe[1], and Bernard De Baets[2]

[1] Dept. of Geography, Ghent University,
Krijgslaan 281-S8, 9000 Ghent, Belgium
[2] Dept. of Mathematical Modelling, Statistics and Bioinformatics, Ghent University,
Coupure links 653, 9000 Ghent, Belgium
manuel.claeysbouuaert@ugent.be

Abstract. Recently, a spatio-temporal application of Markov Random Fields for modeling of spatio-temporal phenomena from collected sensor network data has been proposed by Piatkowski et al. The model employs stated conditional dependence relations between sensor readings as domain knowledge and allows for a wide variety of interesting inference queries, making it a useful predictive modeling approach. Although some of the model's properties have been investigated, important questions remain, such as the motivation for and influence of the stated conditional dependence assumptions, the feasibility of parameter estimation and such aspects as accuracy and overfitting. We propose to take a deeper look at this application and its performance on a variety of synthetic datasets. A better understanding of this model will enable to gain insights into when and how it can be applied to real world data.

Keywords: Sensor Networks, Spatio-Temporal Data, Markov Random Fields, Undirected Graphical Models, Prediction

1 Introduction

Sensor networks are used to collect data on a wide range of physical and social phenomena, including traffic jams, air pollution, sea temperatures, water levels in river networks and crowding at mass-events. Models of these phenomena can make use of such data to help us investigate their dynamics and make predictive statements on their evolution. Recently, a spatio-temporal application of Markov Random Fields (MRF) has been proposed for this purpose by Piatkowski et al. [1, 2]. We start by giving a brief overview of this model, and will indicate aspects of our current and future research in the next section.

This research is partly conducted at the CartoGIS research group of Ghent University, where over the last years research was done on measuring and analysing the dynamics of pedestrian crowds at mass events through proximity-based Bluetooth tracking [8]. Since the interest in this modeling approach grew from the research question of how to model such types of phenomena, we will investigate the possibilities for applying spatio-temporal MRF models on this data in the light of the required data and network properties.

This research was started just recently, and we hope to collect the first results in the coming weeks.

Brief Overview of the Proposed Model

Markov Random Fields are a type of Probabilistic Graphical Models. They describe a multi-variate probability distribution and use a graph of which the nodes represent random variables and the edges specify conditional dependence relations between these random variables. A good introduction to such models can be found in [4] and [5].

The proposed model introduced in [1, 2] is a specific application of Markov Random Fields for spatio-temporal periodic data from sensor networks, where every sensor reading at every timestamp is represented by a random variable. The model makes two critical choices. Firstly, it uses singleton and pairwise indicator functions (one for every node and edge and their possible states) in the factorisation of the multivariate probability function. In this parametric model, every such function is accompanied by one parameter. A second choice considers the way in which the spatial and temporal dependence and independence relations are built up among the different nodes (i.e. among the different variables and thus among the different sensor readings). In short, the model proposes to start from a set of spatial relations assumed to hold at every timestamp (called the *protograph*), and introduces basic temporal relations whilst making a kind of first-order temporal Markov assumption. Knowledge on the spatial and temporal dynamics of the observed phenomenon can then be cleverly exploited to define a graph that has as few edges as possible (creating as few parameters as possible), but that is still able to capture all important dependence relations of the phenomenon's behaviour.

The parameters of the model can then be learned from data by Maximum Likelihood Estimation (MLE). Since Markov Random Fields are a member of the family of exponential models, they have elegant properties which lead to a convex form of the likelihood and an interpretation of the MLE condition known as the *moment matching equation*. The MLE solution can be found using standard convex analysis optimisation. In every optimisation step some terms must be computed using exact or approximate inference algorithms, depending on the size and structure of the graph. For the high dimensional and loopy graphs that will be common in this setting, variants of Loopy Belief Propagation (LBP) are the standard approach [4]. Interestingly, during this computation only specific averages of the historic data are used, the full historical data must not be stored.

Once the parameters are learned, interesting queries can be performed on the model, as demonstrated in [1]. These include the computation of joint, marginal and conditional probabilities, the computation of the most probable multivariate state and sampling of multivariate states.

Although some of the properties of this model have been investigated [2], important questions remain. Here, we will outline a few of them and indicate aspects which we are currently investigating in further detail, or planning to investigate in the near future.

2 Current and Future Research

Influence of Graph

In [1, 2], graph design is divided in a spatial and temporal stage. First, a protograph is proposed that "represents the physical deployment of the sensors" of the studied phenomenon, constructed by means of a k-NN graph or based of the network on which the phenomenon takes place (a street network). Then a temporal Markov assumption is made by connecting nodes representing the same physical sensor across consecutive time-steps. In [2] additional temporal edges are created between nodes of consecutive timestamps when the protograph contains a spatial edge between their corresponding physical sensors.

This represents only a small amount of the possible spatio-temporal relationships. We set out to investigate the potential of other reasonings when building up the graph, not necessarily following the protograph routine.

A first example is to use knowledge on the speed of information: when something happens at a sensor at a specific time, which future sensor readings can be influenced in a direct way by this event, taking into account the speed at which the information travels (e.g. based on average wind speeds and particle diffusion speeds for air pollution or visibility and the speed of traffic waves for traffic jams). When direct influence between two physical sensors is possible but the necessary time is more than the time-steps between two readings, temporal edges will span multiple time-steps and hence no first-order temporal Markov assumption is made.

Benchmarks on Synthetic Datasets

The conditions under which a set of parameters exist that honour the moment matching condition (i.e. the log-likelihood function attains a maximum for finite parameter values) have been investigated in literature [6], but are difficult to verify for realistically scaled MRFs. We are currently implementing MRFs in available open-source software packages [9, 10], creating synthetic input data and testing various approximate inference algorithms to investigate if parameter estimation is feasible at realistic scales and under which conditions a solution is found.

We start by creating synthetic input data in two steps. First, a graph is built by connecting a set of nodes (sensor readings) following a certain reasoning. Then, a spatio-temporal field is chosen, where examples include travelling waves, diffusion and drift processes, threshold processes and combinations of these phenomena. Data samples are created by reading the values of that field at all sensor locations and timestamps. Every sample (collection of sensor readings) represents a period of the phenomenon. By adding some form of randomness to the field over the different samples (e.g. variable intensity or spatial or temporal initialisation), all periods differ but originate from the same distribution.

We can then train a model using these samples. Once parameters are found that honour the moment matching condition, we evaluate the model accuracy in multiple ways. One is to generate new samples from the trained model using a sampling algorithm, and to check if these samples exhibit the same dynamics as the training samples generated by the initial spatio-temporal field. Another way is to compute conditional probabilities and compared the output to the expected behaviour of the original field. Moreover we can compute a measure of the prediction accuracy of the model using a training and test set (see next section). This way we can investigate which types of phenomena can be learned (i.e. trained and predicted successfully) and how this depends on dataset specifications such as training sample size, spatial and temporal resolution of the sensor network, number of discrete states.

When training a model on data by MLE parameter estimation, an important practical choice we have to make is that of the algorithm to be used to calculate the value of the log likelihood and gradient at every step of the convex optimisation (and at the end to check the moment matching condition). Only for the smallest dimensions exact algorithms can be used, in other cases one will have to turn to approximate algorithms. A good approximate algorithm must be fast, have high probability of convergence, converge to the correct value and keep the convexity of the log likelihood [7]. Identifying such algorithms is one of the subjects of our research.

Another crucial question regards the necessary amount of input samples. Certain states of sensor readings and (more crucially) certain state combinations between two adjacent sensor readings can be rare, necessitating many input samples to correctly estimate the empirical expectations of the corresponding sufficient statistics, which is a fundamental part of the moment matching condition and MLE optimization. The importance of this issue for practical application and insights on determining a sufficient sample size are subject of our research.

Prediction Accuracy

A measure for prediction accuracy on ordinal data called the 'instant-wise prediction accuracy' is proposed in [2] and determined for models on two real world datasets using a training- and test set routine. This measure allows us to compare the model with a standard naive k-NN predictor. The results indicated that for those two specific datasets the prediction accuracies are comparable. We think that additional test are desirable on synthetic and real datasets to compare the model's prediction accuracy to that of standard predictors (baseline predictors such as random walks or historical mean, Vector Auto Regression, k-NN, ...) and indicate in which cases it has distinct advantages.

Overfitting

For realistically sized datasets, the amount of parameters used by the proposed model grows extremely fast and can easily attain orders of magnitude of e.g. 10^7. This makes these models prone to overfitting: many degrees of freedom may twist the model such that it is very good at predicting training data, but fails when predicting test data. Along with investigating prediction accuracy, we will also check wether overfitting actually occurs. If so, some techniques will be tested to prevent overfitting (regularisation, early stopping, cross-validation and Bayesian priors, ...).

There exist two papers where elastic net regularisation (combined ℓ_1 and ℓ_2 regularisation) is used for this model, but the validation of this choice is not discussed in detail [1, 3].
Note that tackling overfitting simply by providing more input data is not an option, since only historic averages are used during model training. Of course, providing more data will still yield more accurate values for the empirical expectations of the sufficient statistic, used during training.

Should such techniques offer no salvation, the option to share parameters across nodes and edges can be investigated, since this significantly reduces the total number of parameters. Note that this allows for example to make the model locally or globally time-independent.

3 Conclusions

Recently, a spatio-temporal application of Markov Random Fields has been proposed for the task of modeling spatio-temporal phenomena from collected sensor network data [1, 2]. In this model, the assumed conditional dependence relations between the different sensor readings are used as domain knowledge. We outline the ideas behind this approach, discuss some of its properties and indicate the aspects of the model which we are currently investigating, or will investigate in the near future. To this extent, we are studying their performance on synthetic data, and datasets are introduced based on some general spatio-temporal phenomena. Also, we propose to investigate other spatio-temporal reasonings behind the build up of the model's graph. This can allow for models that are better suited to capture the dynamics present in the studied phenomena. Also, we are investigating some practical aspects of the application of the model to real world data, such as identifying the optimal approximate inference algorithms and assessing overall feasibility of parameter estimation. Finally, strategies for aspects such as prediction accuracy and overfitting are discussed.

This research was started just recently, and we hope to collect the first results in the coming weeks.

We believe that a systematic investigation of the properties of this model will allow us to validate its use on real world datasets.

4 Acknowledgments

Manuel Claeys Bouuaert is a Fellow of the Fund for Scientific Research-Flanders (FWO–Vlaanderen).

References

1. Piatkowski, N., Lee, S., Morik, K.: Spatio-temporal models for sustainability. Proc. SustKDD Work. within ACM KDD (2012)
2. Piatkowski, N., Lee, S., Morik, K.: Spatio-temporal random fields: compressible representation and distributed estimation. Mach. Learn. 93, 115–139 (2013)
3. Piatkowski, N.: iST-MRF: interactive spatio-temporal probabilistic models for sensor networks. In International workshop at ECML PKDD 2012 on instant interactive data mining, (2012)
4. Bishop, C.: Pattern Recognition and Machine Learning. Springer (2006)
5. Koller, D., Friedman, N.: Probabilistic graphical models: principles and techniques. MIT press (2009)
6. Wainwright, M.J., Jordan, M.I.: Graphical Models, Exponential Families, and Variational Inference. Found. Trends Mach. Learn. 1 (2007)
7. Wainwright, M.J., Tommi S.J., Alan S.W.: Tree-reweighted belief propagation algorithms and approximate ML estimation by pseudo-moment matching. Workshop on Artificial Intelligence and Statistics. 21, Np: Society for Artificial Intelligence and Statistics (2003)
8. Versichele, M., Neutens, T., Delafontaine, M., Van de Weghe, N.: The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. Appl. Geogr. 32, 208–220 (2012)
9. Schmidt, M.: UGM: Matlab code for undirected graphical models (2014) `http://www.di.ens.fr/~mschmidt/Software/UGM.html`
10. Piatkowski N.: Spatio-Temporal Random Fields (STRF), Artificial Intelligence Group of the TU Dortmund University, Germany (2014) `http://sfb876.tu-dortmund.de/strf/index.html`