Frontiers of Statistical Physics and Information Processing
– probabilistic description for nature and information processing –

Pairwise MRF Models selection and Traffic Inference

Cyril Furtlehner
INRIA Saclay, Tao team & Laboratoire de Rechecherches en Informatique
Paris-Sud University

We investigate different ways of generating approximate solutions to the inverse problem of pairwise Markov random field (MRF) model selection. Motivated by traffic inference we design in parallel two concurrent models an Ising and a Gaussian model with the constraint that they are suitable for "belief-propagation" (BP) based inference.

1 Problem statement

Once a joint probability measure is given, the belief propagation algorithm can be very efficient for inferring hidden variables while observing the other variables, but in real applications it is often the case that we have first to build the model. This is precisely the case for the application that motivates this work. This deals with the reconstruction and prediction of traffic congestion conditions, typically from sparse observations on the secondary network where no fixed sensors are installed. Data are obtained from vehicles embedded in the traffic, equipped with GPS and able to exchange data through cellular phone connections for example, in the form of so-called Floating Car Data (FCD) by sending their speed along with their position. The goal then is to be able to provide at any time a travel time for each unobserved segment of the network and a short term forecast for all segments.

Our approach to this objective is to build a MRF based on past observations. Each variable representing a travel time is attached to a segment possibly at various discretized time in the day. We assume that the collected data allows one for a statistical modelling of each segment and a certain number of pairs of segments. The FCD sent by probe vehicles concerning some area of interest are continuously collected over a reasonable period of time (one year or so) such as to allow a finite fraction (a few percents) of road segments to be covered in real time. Schematically the inference method works as follows:

- Historical FCD are used to compute empirical dependencies between contiguous segments of the road network.
- These dependencies are encoded into a graphical model, which vertices are (segment,timestamps) pairs attached with a traffic index variable, like e.g. the binary state CONGESTED/NOT-CONGESTED.
- Congestion probabilities of segments that are unvisited or sit in the short-term future are computed with BP, conditionally to real-time data.

On the factor-graph, the information is propagated both temporally and spatially. In this perspective, reconstruction and prediction are on the same footing, even though prediction is expected to be less precise than reconstruction.

2 MRF models

Since the distribution of travel time tt_ℓ for a given road segment ℓ is not given by a simple parametric and identical model for all segments we do not consider the inference problem in the space of travel time directly but consider instead various mapping $x_\ell(tt_\ell)$ attached to each segment which we call traffic indexes. This leads us to consider basically 2 different models:

- A Gaussian MRF (GMRF): it is the most straightforward approach. First each travel time tt_ℓ is mapped via its empirical cumulative distribution onto a normalized Gaussian variable $x_\ell = \mathcal{N}(0,1)$. The GMRF is then build based on a subset of pairwise empirical covariance $\hat{Cov}(x_\ell, x_{\ell'})$ computed in this space from the observations. In principle BP can be used in closed form in this case, the messages exchanged between variables corresponding to the mean and the variance of these variables.
  Pros: this is well suited real variable inference for BP. If BP converges the marginal are exact marginals of the GMRF model. The inverse mapping directly delivers travel time predictions along with an estimation of the error.
  Cons: there is a strong assumption that the data in the traffic index space have a single mode, which is probably not true for traffic data, where we expect instead to have different modes corresponding to different congestion patterns.

- An Ising model for traffic: this approach is based on the intuition that a natural binary latent state $s_\ell = \pm 1$ for NON-CONGESTED/CONGESTED is attached to each segment. Then, instead of encoding the full dependency between different travel times, we instead use this latent state as a proxy to encode dependencies between segments. Based on the travel time cumulative distribution we propose different ways of defining these latent state like: (i) a simple fixed threshold with $s_i \stackrel{\text{def}}{=} 2P(tt_\ell < tt_\ell^\star) - 1$ where $tt_\ell^\star$ is e.g. the median (ii) a random threshold allowing one to map the belief $b(s_\ell = 1)$ directly to a travel time tt_ℓ through the inverse cumulative distribution. The correlations between latent states variables needed to feed the MRF model are obtained either by moment matching or via en expectation maximization procedure.
  Pros: natural and appealing binary description and a very light interaction model, particularly well suited for BP when a multi-modal distribution is expected to be associated to macrostate of congestion patterns.
  Cons: variables to predict are real travel time and the mutual information between two travel times is reduced to at most one single bit of mutual information, at least when a fixed threshold (i) is used to define the latent state.

3 The Inverse Problem and a heuristic solution

In both cases we face a difficult inverse problem, where both the MRF's graph structure and parameters have to be found. In the Ising case these are the local fields and coupling while in the Gaussian case this is the so called "precision matrix" i.e. the inverse covariance matrix which has to be found. In the latter case, there are two issues which prevent the direct use of the inverse covariance matrix: (i) the empirical covariance matrix may not be necessarily full, in our context we expect many missing entries. (ii) the compatibility with BP is not insured i.e. BP might not converge if the precision matrix is too dense. Therefore in the Gaussian case we look for a good trade-off between likelihood and sparsity of the model.

A statistical physics approach to the inverse Ising problem (IIP) is given by the linear response theory combined with various hypothesis. If the coupling are assumed to be small, the perturbation expansion can be used to deliver at lowest order the mean field solution, the TAP solution at the next order ... A different type of mean-field approximation is the Bethe approximation which reduces to the TAP approximation at lowest order and which consists in to assume that the graph is locally a tree. This leads to two, possibly different mean-field solutions to the IIP: the direct one, by using the relation valid on a tree between the joint probability and the single and pairwise marginal distributions; the indirect one also called susceptibility propagation relying on the relation between the inverse susceptibility matrix and the set of susceptibility coefficients attached to the links of the tree.

These approach are mainly valid if we expect a high temperature model in its paramagnetic phase. Instead our data displays low temperature behavior and cannot be considered uni-modal. To get round this problem we have proposed first a simple heuristic deformation of the direct Bethe model with 2 parameters, K the mean connectivity of the graph and α a global rescaling of the couplings. The graph is obtained by first determining a maximum spanning tree with weights given by mutual information between variables and completing up to the targeted mean connectivity by selecting links according to these weights. The parameter $\alpha \in [0,1]$ allows one to interpolate between a model with independent variables with single marginal matching the observation for $\alpha = 0$, and the direct Bethe solution for $\alpha = 1$. Upon joint calibrating of α and K, the model is able to recover well separated modes of a multimodal distribution. In that case, the model displays many BP fixed points, in one to one correspondence with the hidden modes contained in the observation data. We can interpret this heuristic solution through some asymptotic mapping onto a Hopfield model.

4 Various Perturbations of the Bethe reference point

Minimal solution based on Maximum spanning tree with mutual information suggests it is a good starting point from which we should perturb around the Bethe approximation to find improved solutions compatible with BP. We have identified three different promising and possibly complementary way of proceeding.

Line search along the natural gradient direction: a first direction is provided by the observation that the natural gradient can be made explicit a the Bethe point. It involves up to 4-points susceptibility coefficients which we can be actually computed explicitly in term of the 2-points susceptibility coefficients at this reference point. Depending on the way the deformation of the model is then parametrized, tractable optimization strategies using this natural gradient can be defined.

Iterative proportional scaling: A second direction that we have explored consists in to proceed link-wise from this reference point. The link yielding the maximum gain in likelihood is obtained by solving a simple variational problem which solution is referred to as "iterative proportional scaling" (IPS) in the statistics literature, for solving maximum likelihood estimation problem.

In the GMRF case we found that this can be implemented efficiently due to local transformations of the precision matrix after adding one link. By comparison we implemented methods based on sparse norm penalization (L_0 or L_1) and find out that it is competitive with L_0 based approach, with a $O(N^3)$ complexity in the sparse regime. Incidentally we also found that the L_1 based method is not working well for this problem. An additional advantage of our IPS based solution is the possibility to combine it with spectral constraints like walk-summability with BP or/and graph structure constrain to enforce compatibility. For the same computational cost we can get a complete set of good trade-off between likelihood and compatibility with BP.

Concerning the Ising model IPS is too computationally expensive, and even if we separate the structure from the coupling selection, there is no satisfactory solution in the low temperature regime. It can be used only for marginal modification of a Bethe model.

Low temperature expansion and loop corrections: For the Ising model a standard way to deal with the high couplings at low temperature is provided by the low temperature expansion, which in absence of local fields leads to a dual model of binary loop variables with weak interactions. If local fields are non-zero, but a BP fixed point is given, a similar generalized loop expansion w.r.t. to this reference point has been formulated (Chertkov, Chernyak 2006). When looking at the explicit formula for the Bethe susceptibility we can see that it potentially incorporates loop corrections which are wrong already at the first loop contribution. This explains why susceptibility propagation is not working well in this domain. Our proposal in this context to provide approximate solutions to the IIP in a tractable way, is to use a minimal cluster expansion in the loop variables to exploit the weak coupling hypothesis. In this way we obtain a self consistent system of equations with a fixed point solution which can be found by iteration in some cases.

Acknowledgements: The work summarized here is done in collaboration with Jean-Marc Lasgouttes, Yufei Han and Victorin Martin, in the context of the ANR SYSCOMM Project TRAVESTI No 08-SYSC-017.

References

[1] http://travesti.gforge.inria.fr/

[2] C. Furtlehner, Yufei Han, J.-M. Lasgouttes, V. Martin. Pairwise MRF Calibration by Perturbation of the Bethe Reference Point. INRIA RR-8059, 2012.

[3] C. Furtlehner, J.M. Lasgouttes, and A. Auger. Learning multiple belief propagation fixed points for real time inference. *Physica A: Statistical Mechanics and its Applications*, 389(1):149–163, 2010.

[4] C. Furtlehner, J.-M. Lasgouttes, and A. de La Fortelle. A belief propagation approach to traffic prediction using probe vehicles. In *Proc. IEEE 10th Int. Conf. Intel. Trans. Sys.*, pages 1022–1027, 2007.