Towards an Integer Approximation of Undirected Graphical Models

Nico Piatkowski and Katharina Morik

Artificial Intelligence Group, TU Dortmund University, 44227 Dortmund, Germany
{nico.piatkowski, katharina.morik}@cs.tu-dortmund.de
http://www-ai.cs.tu-dortmund.de

Data analytics for streaming sensor data brings challenges for the resource efficiency of algorithms in terms of execution time and the energy consumption simultaneously. Fortunately, optimizations which reduce the number of CPU cycles also reduce energy consumption. When reviewing the specifications of processing units, one finds that integer arithmetic is usually cheaper in terms of instruction latency, i.e. it needs a small number of clock cycles until the result of an arithmetic instruction is ready. This motivates the reduction of CPU cycles in which code is executed when designing a new, resource-aware learning algorithm. Beside clock cycle reduction, limited memory usage is also an important factor for small devices.

Outsourcing parts of data analysis from data centers to ubiquitous devices that actually measure data would reduce the communication costs and thus energy consumption. If, for instance, a mobile medical device or smartphone can build a probabilistic model of the usage behavior of its user, energy models can be made more accurate and power management can be more efficient. The biggest hurdle in doing this, are the heavily restricted computational capabilities of very small devices—some do not even have a floating point processor. Consequently, computationally simple machine learning approaches have to be considered. Low complexity of machine learning models is usually achieved by independence assumptions among features or labels. In contrast, the joint prediction of multiple dependent variables based on multiple observed inputs is an ubiquitous subtask in real world problems from various domains. Probabilistic graphical models are well suited for such tasks, but they suffer from the high complexity of probabilistic inference.

In the extended abstract at hand, we show how to write the joint probability mass function of undirected graphical models as rational number, if the parameters are integers. More details on the integer parametrization of undirected graphical models can be found in [1]. Inference algorithms and a new optimization scheme are proposed, that allow the learning of integer parameters without the need for any floating point computation. This opens up the opportunity of running machine learning tasks on very small, resource-constrained devices. To be more precise, based only on integers, it is possible to compute approximations to marginal probabilities, to maximum-a-posteriori (MAP) assignments and maximum likelihood estimate either via an approximate closed form solution or an integer variant of the stochastic gradient descent (SGD) algorithm. It turns

 In: T. Seidl, M. Hassani, C. Beecks (Eds.): Proceedings of the LWA 2014 Workshops: KDML, IR, FGWM, Aachen, Germany, 8-10 September 2014, published at http://ceur-ws.org

out that the integer approximations use less memory and deliver a reasonable quality while being around twice as fast as their floating point counterparts. To the best of our knowledge, there is nothing like an integer undirected model so far.

Many approximate approaches to probabilistic inference based on belief propagation were proposed in the last decade. Unfortunately, most of these methods are by no means suited for embedded or resource constrained environments. In contrast to these approaches, the model class that is proposed in the paper at hand has the same asymptotic complexity as the vanilla inference methods, but it uses cheaper operations. Inspired by work from the signal processing community, the underlying model class is restricted to the integers, which results in a reduced runtime and energy savings, while keeping a good performance. This new approach should not be confused with models that are designed for integer state spaces, in which case the state space $\mathcal{X}$ is a subset of the natural numbers or, more generally, is a metric space. Here, the state space may be an arbitrary discrete space without any additional constraints.

In their book on graphical models, Wainwright and Jordan stated that "It is important to understand that for a general undirected graph the compatibility functions ψ_C need not have any obvious or direct relation to marginal or conditional distributions defined over the graph cliques. This property should be contrasted with the directed factorization, where the factors correspond to conditional probabilities over the child-parent sets." This raises hope that we might find meaningful probabilistic models, even when we restrict the model parameters to be integers. For excluding every floating point computation, the identification of integer parameters is not enough. That is, the computations for training and prediction have to be based on integer arithmetic.

The first step towards integer models is directly related to the above statement. Strictly speaking, the parameter domain Ω is restricted to the set of integers $\mathbb{N}$ and a new potential function is defined as

$$\overline{\psi}_C(x_C) := 2^{\langle \theta_C, \phi_C(x) \rangle} = \exp\left(\ln(2)\langle \theta_C, \phi_C(x) \rangle\right).$$

Considering parameters $\theta \in \mathbb{R}^d$ of a model that has potential function $\psi_C(x_C)$, it is easy to see that replacing $\psi_C(x_C)$ with $\overline{\psi}_C(x_C)$ does not alter the marginal probabilities as long as the parameters are scaled by $1/\ln 2$. By this, it is possible to convert integer parameters that are estimated with $\overline{\psi}_C(x_C)$ to $\psi_C(x_C)$ (and vice versa), without altering the resulting probabilities. Notice that $\overline{\psi}_C(x_C)$ can be computed by logical bit shift operations which consume less clock cycles than the corresponding transcendental functions required to compute $\psi_C(x_C)$.

Acknowledgements This work has been supported by Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 "Providing Information by Resource-Constrained Data Analysis", project A1.

References

1. Piatkowski, N., Sangkyun, L., Morik, K.: The integer approximation of undirected graphical models. In: De Marsico, M., Tabbone, A., Fred, A. (eds.) 3rd International Conference on Pattern Recognition Applications and Methods (ICPRAM). SciTePress (2014)