PHYSICAL REVIEW E **74**, 042103 (2006)

Benefits and pitfalls of belief-propagation-mediated superparamagnetic clustering

Thomas Ott and Ruedi Stoop
Institute of Neuroinformatics, UNI/ETH Zurich, Switzerland
(Received 16 June 2006; published 30 October 2006)

Belief propagation is a candidate for the accelerated evaluation of statistical averages if compared to Monte Carlo approaches. We show that superparamagnetic clustering can be accelerated substantially by means of belief propagation. However, critical slowing down of the method is observed. Origin and strategies of avoidance of the critical slowing down are explained from models of periodic Ising grids. We show that belief-propagation-mediated superparamagnetic clustering provides identical clusters with Monte Carlo-mediated clustering only at the coarse-grained level. Concurring results between the Monte Carlo and the belief-propagation methods can be obtained by means of the sequential superparamagnetic-clustering approach.

DOI: 10.1103/PhysRevE.74.042103 PACS number(s): 02.50.Tt, 05.70.Fh, 89.70.+c

Superparamagnetic clustering algorithms (SC) [1], and in particular, the sequential superparamagnetic clustering variant (SSC) [2,3], are inherently time consuming and therefore natural candidates for acceleration by means of the belief-propagation (BeP) method [4], which, however, is known to often run into problems of convergence. How to optimize the convergence properties of BeP applied to superparamagnetic clustering is the focus of this contribution.

Clustering is the fundamental cognitive process of binding data items to classes. Algorithms designed for achieving this computationally can be divided into two classes: the *parametric* and the *nonparametric* approaches. Whereas in the first class one starts from *a priori* assumptions about the global structure of the data (such as the number of clusters, or the intracluster distribution properties), nonparametric algorithms rely entirely on the similarity of data items, i.e., on local information. Recently, a nonparametric algorithm has been developed, which is based on the picture of an inhomogeneous ferromagnetic spin system. In this so-called superparamagnetic clustering approach (SC) [1], the clustering process is identified with the pattern-formation process that a s-state Potts spin system undergoes when the temperature T is used as the control parameter. Increasing T from zero leads the system through different phase transitions. For small $0 \leq T < T_{ferro}$, the system is in the ferromagnetic phase and all spins form one single cluster. For $T_{ferro} < T < T_{para}$, the system is in the superparamagnetic phase, where groups of aligned spins form local clusters. Couplings among spins of local clusters are significantly stronger than to spins outside. Upon the increase of T, the local clusters tend to break up. This leads to a natural clustering hierarchy into ever more and smaller clusters, in which only the most strongly coupled spins remain aligned. The process ends at $T = T_{para}$, from whereon the spins are no longer correlated and form singletons (paramagnetic phase). In order to map a clustering problem onto this setting, imagine that the data items are described by property vectors, so that a pairwise (dis)similarity function d_{ij} can be defined (e.g., by the Euclidean distance). Each data item is now represented by a spin variable s_i, which will be coupled to its q nearest neighbors s_j evaluated by means of the similarity function d_{ij}.

In the superparamagnetic clustering algorithm, the coupling strength J is a decreasing Gaussian function of the similarities

$$J_{ij} = J_{ji} = \frac{1}{\hat{K}} \exp\left(\frac{-d_{ij}^2}{2a^2}\right), \tag{1}$$

where $\hat{K}$ is the average number of coupled neighbors per site and a is the similarity average of the coupled spins. The probability of a spin configuration s is given by the Boltzmann distribution

$$p(s) = \frac{1}{Z} e^{-H(s)/T}, \tag{2}$$

with the Hamiltonian $H(s) = -\sum_{i,j} J_{ij} \delta_{s_i s_j}$. For a given T, two coupled spins i, j are called aligned if the pairwise spin correlation

$$p_{ij} = \sum_s p(s) \delta_{s_i s_j} \tag{3}$$

exceeds the threshold Θ, i.e., $p_{ij} > \Theta$. Sets of pairwise aligned spins define the clusters. The choice of Θ is uncritical [1,2], a good value is $\Theta = 0.7$. The core quantities $p_{ij} = p(s_i = s_j)$ are traditionally evaluated by means of naturally time-consuming Monte Carlo (MC) methods (e.g., the Swendsen-Wang algorithm [6]).

As an alternative, BePs have recently obtained widespread attention (e.g., Ref. [5]), by offering approximate marginals that can be obtained with a significant speedup over MC methods. BeP is based on the exchange of messages $m_{i\to j}(s_j)$ and $m_{j\to i}(s_i)$ between pairs of coupled sites i, j, containing a recommendation about in what state the other site's spin should be. Given the set of messages at time t, $\{m_{i\to j}^t(s_j)\}$, the outgoing messages at time $t+1$ are determined by the weighted products of incoming messages as

$$m_{i\to j}^{t+1}(s_j) = \sum_{s_i} e^{2J_{ij}\delta_{s_i s_j}/T} \prod_{k \in N(i)\backslash j} m_{k\to i}^t(s_i), \tag{4}$$

where $N(i)\backslash j$ denotes all neighboring sites of i without j. Once BeP has converged to $\{m_{i\to j}^\infty(s_j)\}$, the pair correlation (3) can be approximated by means of the pairwise beliefs $b_{ij}(s_i, s_j)$, where for the case of Ising spins as the relevant example considered below, $p_{ij} = b_{ij}(1,1) + b_{ij}(-1,-1)$ and

$$b_{ij}(s_i,s_j) = ce^{2J_{ij}\delta_{s_i s_j}/T} \prod_{k\in N(i)\setminus j} m^{\infty}_{k\to i}(s_i) \prod_{l\in N(j)\setminus i} m^{\infty}_{l\to j}(s_j) \quad (5)$$

(c is a normalization constant).

Generally, the proof of convergence of BeP is a delicate matter and has only been established for very special problem classes. On treelike coupling structures, BeP provides the solution Eq. (3) within a finite number of iterations [4]. If the underlying graphs are loopy, it is important to know for which initial conditions BeP converges, and how closely the solution matches the analytical solution. The aim of this paper is to show that for binary (Ising) spin systems, convergence of BeP-mediated SC can be achieved by a clever choice of the initial conditions at all temperatures T. Based on numerical simulations we suggest that this situation extends to the general Potts spin case.

From the point of view of statistical physics, the explicit evaluation of the pair correlations p_{ij} would provide the exact clustering solution (if we ignore nonrigorous criteria such as the threshold Θ). Since the evaluation of this solution is unfeasible for large data sets, the MC solution was taken as the point of reference, on the basis of the Swendsen-Wang algorithm. This approach provides stable solutions after $\sim$200 steps for grids of up to a few thousand data points [2].

Our primary observation with respect to BeP-mediated SC is that suitable termination thresholds $\epsilon = \sum_{i,j,s}|m^{t+1}_{i\to j}(s) - m^{t}_{i\to j}(s)|/(4N\hat{K})$ restrict the process to usually $N_{it} \leqslant 20$ iterations. At some temperatures T, the speed of convergence is, however, hampered by the occurrence of a critical slowing-down phenomenon [see Fig. 1(b)]. If we terminate BeP here after the same number of iterations, the clustering has already achieved its asymptotic form and makes further iterations useless. As is indicated in Fig. 2, for SC, the BeP and MC approaches coincide on the main clustering structures. Rather irrespective of the system size, the clusters that persist over larger-temperature intervals contain the same elements for both approaches (although the corresponding temperature intervals need not be coincident). Clusters of any size whose existence is limited to only a short part of the T axis are generally unresolved by BeP. Since the *sequential* superparamagnetic clustering algorithm (SSC) [2,3] exclusively relies on stable clusters, it is not affected by this shortcoming. This motivates us to base the investigation of the behavior BeP-mediated superparamagnetic clustering on SSC. BeP-mediated SSC leads to a substantial speedup by a factor of 10–20 if compared to the Swendsen-Wang approach, providing robust and reliable clustering results.

The above observations are astonishing, since BeP is known to have convergence problems in loopy data systems. In order to explore the origins of this well-behavedness, we first note that BeP convergence is always guaranteed in the paramagnetic phase [10]; in the ferromagnetic and superparamagnetic phases, it, however, may fail. Our claim is that if the initialization is by means of site-independent messages $m_{i\to j}(1) = 1-y$ and $m_{i\to j}(-1) = y$ with $0 < y < 0.5$ for all connections $i \to j$, BeP also converges in the latter regimes, showing a robust and efficient behavior. This surprising result warrants a detailed further examination.

Since clusters are areas of a relatively homogeneous density, which translate into regions of homogeneous couplings, a simple generic model to study the behavior of BeP on SC and SSC are regular Ising grids. The extension of our studies to systems with nonuniform coupling requires further analysis and is beyond the scope of this contribution. In our analysis, we focus on site-independent message initialization and impose periodic boundary conditions. With this, we eliminate boundary effects, so that under the temporal evolution of map (4) the messages remain to be site independent. For this model, it is convenient to parametrize the messages as

$$\tanh \nu_{ij} = m_{i\to j}(s_j = 1) - m_{i\to j}(s_j = -1), \quad (6)$$

so that the update rule (4) for ν defines the map

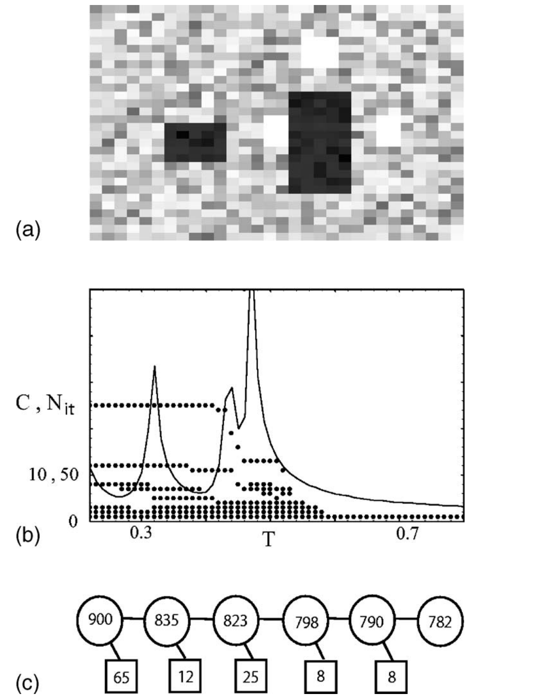

FIG. 1. (a) Clustering analysis of a visual toy scene (five rectangular embedded objects), by means of SSC; (b) cluster stability and breakup and BeP convergence speed (cluster size C and required iterations N_{it}, respectively); (c) SSC-BeP output, identifying five natural clusters of sizes as indicated in the boxes.

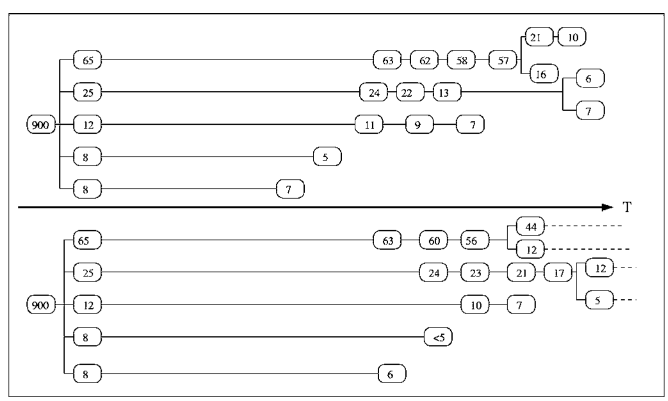

FIG. 2. Dendrogram comparison between SC using BeP and MC (upper and lower panels), respectively. Agreement is limited to the coarse-cluster structure (clusters of sizes $n = 65, 25, 12, 8, 8$).

$$f:\nu_{ij}^{t+1}=\tanh^{-1}\left[\tanh(J/T)\tanh\left(\sum_{k\in N(i)\setminus j}\nu_{ki}^{t}\right)\right],\qquad(7)$$

where J is the coupling strength, constant across the grid. In the case of an $n\times n$ grid with q nearest-neighbor coupling, the mapping (7) is $\mathbb{R}^{n^2q}\to\mathbb{R}^{n^2q}$. Convergence of ν is necessary for the messages m to converge and, as a consequence, also for the beliefs (6). In the reparametrized form, the site-independent messages obtain the form $\nu^0=\nu_{ij}^0=u\in]0,\infty[$. Due to the periodic boundary conditions, the n^2q-dimensional map (7) boils down to the one-dimensional map

$$g:\nu^{t+1}=\tanh^{-1}\{\tanh(J/T)\tanh[(q-1)\nu^t]\}.\qquad(8)$$

For general finite grids with open boundary conditions, map (8) would not be appropriate. However, for grids larger than about 8×8, the messages remain under Eq. (7) almost site independent, so that boundary effects are negligible and the dynamics is well approximated by Eq. (8). Map (8) has the unique fixed point $\nu_0=0$ at high T (BeP-paramagnetic phase), whereas in the BeP-ferromagnetic phase (low T), two additional fixed points $\nu_\pm$ emerge. For any initialization ν^0, g converges to one of the three fixed points. For a two-dimensional grid with $q=4$, the additional fixed points evaluate to $\nu_\pm=\pm\mathrm{arcosh}\{[-3+\tanh(J/T)]/[-4+4\tanh(J/T)]\}^{1/2}$. The critical condition $g'(\nu_0)=1$ shows that the fixed point ν_0 loses stability if the temperature approaches the value $T_{crit}=2.88539\,J$ from above [7].

Map g can be obtained from restricting map f to the diagonal $d:=\{\boldsymbol{\nu}_d=(\nu_d,\ldots,\nu_d)\in\mathbb{R}^{n^2q}\}$ (where $\nu_d\in\mathbb{R}$). It can be shown that the largest eigenvalue λ_1 of the Jacobian $df(\boldsymbol{\nu}_d)=(\partial\nu_{ij}/\partial\nu_{kl})|_{\boldsymbol{\nu}=\boldsymbol{\nu}_d}$ on d always coincides with $g'(\nu_d)$. Calculation confirms that due to the general Frobenius-Perron theorem in any point $\boldsymbol{\nu}_d$ the eigenvector v_1 associated with λ_1 is unique, and that v_1 aligns with d. In the BeP-paramagnetic phase, f is a contraction [10], so that Eq. (7) always converges to the fixed point $\boldsymbol{\nu}_0$. In the BeP-ferromagnetic phase, $\boldsymbol{\nu}_0$ is unstable under f, with maximal expansion into the direction of d. For $|\nu_d|>\nu_{crit}^T$, d is stable against perpendicular perturbations $\Delta\boldsymbol{\nu}$ (see Fig. 3). In this range, df is bounded by g',

$$|df(\boldsymbol{\nu}_d)\Delta\boldsymbol{\nu}|<|\lambda_1\Delta\boldsymbol{\nu}|=|g'(\nu_d)\Delta\boldsymbol{\nu}|<\Delta\boldsymbol{\nu},\qquad(9)$$

since $|g'(\nu_d)|<1$ for $|\nu_d|>\nu_{crit}^T$. Since irregularities can be viewed as sustained perturbations, this suggests that BeP reliably converges for site-independent initialization $\boldsymbol{\nu}_d^0$ also in the case of large, but finite, Ising grids with nonperiodic boundary conditions. This renders the observed convergence reliability of SC (SSC) plausible. We conjecture that using site-independent initial conditions as described above will always guarantee the convergence of BeP in the context of superparamagnetic clustering.

Figure 1(b) suggests, however, that at BeP bifurcation points, cluster breakups are accompanied by a critical slowing-down phenomenon. This phenomenon can be understood from the one-dimensional map g [Eq. (8)] as well. Close to a stable fixed point ν_∞, the evolution of a perturbation ϵ_0 has the form $\epsilon_n=|g^n(\nu^0)-\nu_\infty|$, which vanishes as $\epsilon_n\approx[g'(\nu_\infty)]^n\epsilon_0$. The number of iterations n needed to fall below an arbitrary small threshold ϵ_c is therefore $n>|\ln(\epsilon_c/\epsilon_0)/\ln[g'(\nu_\infty)]|$ from which follows that

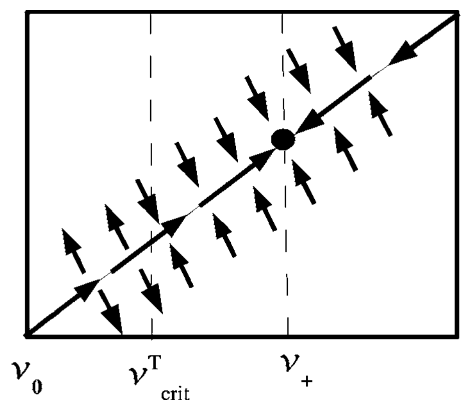

FIG. 3. Stability and convergence of map (7) for an Ising grid with periodic boundary conditions. The large and the small arrows indicate the directions of stability associated with the largest and the second-largest eigenvalue of the map, respectively.

$$n(T)\sim\frac{1}{|\ln[g'(\nu_\infty)]|},\qquad(10)$$

where $g'(\nu_\infty)\to 1$ for $T\to T_{crit}$. Expansion of $r(T):=\ln\{g[\nu_\infty(T)]\}$ around T_{crit} yields 1 as the leading order, because $r'(T_{crit})\neq 0$. For $T\to T_{crit}$, we may thus write

$$n(T)\sim\frac{1}{|T-T_{crit}|^\gamma}\quad\text{with }\gamma=1,\qquad(11)$$

and identify $\gamma=1$ with the BeP critical exponent.

In Fig. 4, for the two-dimensional Ising grid with $J=1$ and $q=4$, the critical slowing down at the BeP-ferromagnetic-paramagnetic transition is described by using the theoretical

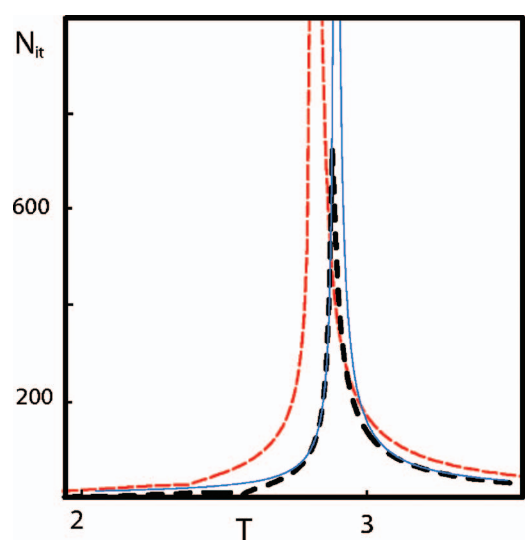

FIG. 4. (Color) Critical slowing down for $2d$ grids. Full line: theoretical result from Eq. (10); bold dashes: $n\times n$ grid ($n\geqslant 3$), periodic boundary conditions; light dashes: 20×20 finite grid, free boundary conditions.

approach Eq. (10), and by using simulations of Ising grids [BeP equations (5)] with periodic and free boundary conditions. The following observations can be made: (i) The shapes of all graphs are similar (although the specific scalings vary due to differing termination thresholds ϵ). (ii) The left-hand-side derivative at the critical peak exceeds that of the right-hand side, due to the slower decay of $g'(\nu_\infty)$ for $T > T_{crit}$. (iii) For $\epsilon \to 0$, simulation results for periodic boundary Ising grids coincide with the predictions by Eq. (10). In particular, the critical temperature matches exactly to $T_{crit} = 2.88539\ J$. The convergence behavior of finite grids is essentially independent of boundary conditions and system size. For free boundary conditions, the critical temperature is, however, slightly lowered (for simple systems, this can be verified analytically). This analysis justifies us to interpret the peaks of bad performance in Fig. 1 as BeP critical slowing down.

In applications of our clustering approach, there is no inherent interest in phase transitions. In the way our clustering algorithm is implemented (fixed maximal number of iterations), their influence is marginalized, as can be seen in the above-developed framework as follows. Although at T_{crit} the fixed point of g has marginal stability, the simple form of g implies that in the first iterations, the fixed point is quickly approached to within an acceptable distance (where the beliefs stop to change dramatically), before power-law slowed-down convergence sets in. Therefore, in order to assign two spins via p_{ij} to the same cluster, a small number of iterations is sufficient. For the example discussed in Fig. 1, we achieve absolutely identical results if we threshold the number of iterations to below 20, for all T. Based on these observations and on the conjecture that the algorithm converges reliably for a site-independent initialization, we conclude that the BeP algorithm provides satisfactory results generically for as few as 10–20 iterations.

Although Monte Carlo and BeP are both of $O(N)$ in the data points, BeP thus leads to a significant gain in speed for SC. One of the few weaknesses SC has is that the clustering resolution is generally compromised if the clusters to be detected are characterized by largely differing average coupling strengths [3]. This weakness is eliminated by our extension of SC to SSC, the sequential superparamagntic clustering approach [2]. For SSC, we first run SC for the data set from $T=0$ to the end of the superparamagnetic phase T_{para}. The stability of each cluster is measured by the quantity $s := \Delta T_{cl} / T_{para}$, where T_{cl} is the length of the temperature interval across which the cluster persists. The cluster with maximal s is selected. This cluster and its residual set are reclustered separately, with reinitialized coupling strengths based on a recalculated parameter a [Eq. (1)]. The procedure is repeated until only clusters of a size below a predetermined minimal size remain, which then are interpreted as noise, or if their stabilities s are found to be marginal. The leaves of the final (binary) tree then form the set of natural clusters. Further details on the algorithm can be found in Ref. [3]; the advantages of SSC in comparison to other approaches are discussed in detail in Ref. [2]. For the toy system shown in Fig. 1, the procedure yields a linear tree. This is due to the homogeneity of the rectangular objects that do not provide substructures. If the proposed site-independent initialization of beliefs is applied, BeP-mediated SC will converge for general clustering data. For coarse clustering, BeP works well with SC, because its approximative nature is absorbed within the reduced clustering resolution provided by SC. In order to arrive at the clustering quality provided by the MC method, BeP, however, requires the strength of SSC.

---

[1] M. Blatt, S. Wiseman, and E. Domany, Phys. Rev. Lett. **76**, 3251 (1996).

[2] T. Ott, A. Kern, A. Schuffenhauer, M. Popov, P. Acklin, E. Jacoby, and R. Stoop, J. Chem. Inf. Comput. Sci. **44**, 1356 (2004).

[3] T. Ott, A. Kern, W.-H. Steeb, and R. Stoop, J. Stat. Mech.: Theory Exp. (2005) P11014.

[4] J. S. Yedidia, W. T. Freeman, and Y. Weiss, in *Exploring Artificial Intelligence in the New Millennium*, edited by G. Lakemeyer and B. Nebel (Morgan Kaufmann, San Francisco, CA, 2003), pp. 239–269.

[5] N. Shental, A. Zomet, T. Hertz, and Y. Weiss, in *Advances in Neural Information Processing Systems 16 NIPS 2003*, edited by S. Thrun, L. K. Saul, and B. Schölkopf (MIT Press, London, 2004).

[6] S. Wang and R. H. Swendsen, Physica A **167**, 565 (1990).

[7] It is known that if BeP converges, the obtained beliefs correspond to minima of the Bethe free-energy functional [4,8]. This is the reason why the critical temperature T_{crit} for BeP on periodic Ising grids and the critical temperature of the Bethe-Peierls approximation [9] coincide [T. Ott (unpublished)].

[8] T. Heskes, Neural Comput. **16**, 2379 (2004).

[9] H. Bethe, Proc. R. Soc. London, Ser. A **150**, 552 (1935).

[10] J. Mooij and B. Kappen, e-print cs.IT/0504030.