Physica A 436 (2015) 321–326

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Simple method for inference in inverse Ising problem using full data

Hirohito Kiwata

Division of Natural Science, Osaka Kyoiku University, Kashiwara, Osaka 582-8582, Japan

HIGHLIGHTS

- A new method for inference in the inverse Ising problem.
- The accuracy of the inference by our method is similar to that by pseudolikelihood method.
- The required computational task of the inference is less than that of pseudolikelihood method.

ARTICLE INFO

Article history:
Received 8 November 2014
Received in revised form 20 March 2015
Available online 8 May 2015

Keywords:
Inverse Ising problem
Frequency distribution
Overdetermined system
Pseudolikelihood maximization

ABSTRACT

We consider inference in the inverse Ising problem using full data, which means incorporating sets of spin configurations. We approximate the Boltzmann distribution of the system to generate a frequency distribution derived from the given data. Then, the ratio between two Boltzmann distributions with different spin configurations eliminates the partition function and we obtain linear equations which can be solved to yield statistical parameters. Our method is applicable to cases where the absolute values of the coupling parameters and external fields are large. Compared to pseudolikelihood maximization, the accuracy of the inference obtained from our method is similar, although our approach is less labor intensive.

1. Introduction

The connection between computer science and statistical physics is increasing on a daily basis [1,2]. Among the many subjects which attract both computer scientists' and physicists' attention, the study of neural networks has a long history [1]. The neural network has been recognized as a purely computational model rather than a model imitating the brain. The neural network as a computational model is roughly divided into two classes: a feedforward neural network and a feedback neural network. In the feedforward neural network, signals are passed through the layers of the neural network in a single direction while, in the feedback neural network, signals are passed through the units of the neural network in interactive directions. Furthermore, the feedback neural network models can be roughly classified as being based on the Hopfield network or the Boltzmann machine. In the Hopfield network, the state of a unit of the neural network is renewed obeying a deterministic rule, but this according to a stochastic rule in the Boltzmann machine [3,4]. Because of the likeness between the Boltzmann machine and the spin glass model in statistical physics, the Boltzmann machine is of particular interest to physicists.

For the Boltzmann machine to achieve the expected performance, we must determine its statistical parameters in advance. The determination of the statistical parameters of the Boltzmann machine from observed data corresponds to machine learning from the viewpoint of computer science or an inverse Ising problem from a physics perspective. The inverse Ising problem has attracted much attention from physicists. In a direct problem, provided that the Boltzmann distribution

E-mail address: kiwata@cc.osaka-kyoiku.ac.jp.

http://dx.doi.org/10.1016/j.physa.2015.04.026
0378-4371/

is given, various statistics such as on-site magnetization and correlation between Ising spins are evaluated using a probability distribution. In an inverse problem, the statistical parameters of the Boltzmann distribution are evaluated using given statistics, i.e., sets of data, and, in statistical science, the inference of statistical parameters relies on a maximum likelihood estimation. Also, for the inverse Ising problem, the maximum likelihood estimation of statistical parameters is effective. However, a daunting computational workload is involved in the estimation of the statistical parameters. The likelihood function is represented by the Gibbs free energy of the Boltzmann machine, and, therefore, the inference of statistical parameters leads to evaluation of the Gibbs free energy. When the number of constituents of the system is large, evaluation of the Gibbs free energy requires considerable computation. In order to estimate statistical parameters, approximate methods, which have been developed in physics, have been applied to the inverse Ising problem [5–15]. However, these methods restrict the efficacy of the inference of statistical parameters from given data owing to coverage of approximations.

As a reliable inference method for the inverse Ising problem, pseudolikelihood maximization (PLM) has attracted much attention recently [16–20]. In PLM, the likelihood function is replaced with an approximate likelihood, so that PLM avoids the large computational cost involved in evaluation of the likelihood function. In addition, it has a wider range of applicability than other approximate methods. In the present paper, we propose an alternative inference method for the inverse Ising problem. Our method has the property that the accuracy of the inference is similar to that of PLM, although the computational cost of determining the inference is less.

2. Theory

We consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, which consists of a set of vertices, $\mathcal{V}$, labeled by $\{1, 2, \ldots, N\}$. Each vertex is connected to another vertex by an element of a set of edges, $\mathcal{E}$. An edge joining vertices i and j is represented by $e = (i, j)$, where $e \in \mathcal{E}$. Given the graph $\mathcal{G}$, we consider a model where an Ising spin, $s_i \in \{\pm 1\}$, is situated on the ith vertex. The energy of the system is given by

$$E = -\sum_{(i,j)\in\mathcal{E}} J_{ij} s_i s_j - \sum_{i\in\mathcal{V}} h_i s_i, \tag{1}$$

where J_{ij} corresponds to a coupling parameter between s_i and s_j, and h_i is an external field applied to s_i. In the case that no edge exists between the ith and jth vertices, the value of J_{ij} becomes zero. For the sake of simplicity, we define a set of Ising spins as $\boldsymbol{s} = (s_1, s_2, \ldots, s_N)$, and name it a spin configuration. We assume that the probability of the system being found in $\boldsymbol{s}$ is given by the following Boltzmann distribution:

$$P(\boldsymbol{s}|\{J_{ij}\}, \{h_i\}) = \frac{\exp[-\beta E(\boldsymbol{s})]}{Z(\beta, \{J_{ij}\}, \{h_i\})}, \quad Z(\beta, \{J_{ij}\}, \{h_i\}) = \prod_{i\in\mathcal{V}} \sum_{s_i=\pm 1} \exp[-\beta E(\boldsymbol{s})], \tag{2}$$

where β is inverse temperature and, in order to explicitly show the dependence of the energy on $\boldsymbol{s}$, we represent the energy of the system by $E(\boldsymbol{s})$. The expectation values such as the on-site magnetization and correlation between Ising spins are evaluated using the Boltzmann distribution of Eq. (2). The expectation values depend on the coupling parameters, $\{J_{ij}\}$, and external fields, $\{h_i\}$.

In the inverse problem, the coupling parameters and the external fields are estimated using the given data. We consider the case in which M sets of the Ising spin configurations, $\{\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)}, \ldots, \boldsymbol{s}^{(M)}\}$, are given as the data, which are generated by an independent identical Boltzmann distribution, $P(\boldsymbol{s}|\{J_{ij}\}, \{h_i\})$. We infer the coupling parameters and the external fields from these data. For the inference, the maximum likelihood estimation is useful. Given $\{\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)}, \ldots, \boldsymbol{s}^{(M)}\}$, we define a histogram corresponding to a frequency distribution function, such that

$$Q(\boldsymbol{s}) = \frac{1}{M} \sum_{\mu=1}^{M} \delta(\boldsymbol{s}, \boldsymbol{s}^{(\mu)}), \tag{3}$$

where $\delta(\boldsymbol{s}, \boldsymbol{s}')$ is the Kronecker delta, which is equal to 1 when $\boldsymbol{s} = \boldsymbol{s}'$ and 0 when $\boldsymbol{s} \neq \boldsymbol{s}'$. With the frequency distribution function, we define the log-likelihood as

$$\begin{aligned} \mathcal{L}(\beta, \{J_{ij}\}, \{h_i\}) &= \prod_{i\in\mathcal{V}} \sum_{s_i} Q(\boldsymbol{s}) \ln P(\boldsymbol{s}|\{J_{ij}\}, \{h_i\}), \\ &= \frac{1}{M} \sum_{\mu=1}^{M} \beta \left(\sum_{(i,j)\in\mathcal{E}} J_{ij} s_i^{(\mu)} s_j^{(\mu)} + \sum_{i\in\mathcal{V}} h_i s_i^{(\mu)} \right) - \log Z(\beta, \{J_{ij}\}, \{h_i\}). \end{aligned} \tag{4}$$

To find $\{\hat{J}_{ij}\}$ and $\{\hat{h}_i\}$, which maximize the log-likelihood, Eq. (4), we differentiate this equation with respect to J_{ij} or h_i and, respectively, find

$$\frac{1}{M} \sum_{\mu=1}^{M} s_i^{(\mu)} s_j^{(\mu)} = \frac{1}{Z(\beta, \{J_{ij}\}, \{h_i\})} \frac{1}{\beta} \frac{\partial Z(\beta, \{J_{ij}\}, \{h_i\})}{\partial J_{ij}}, \tag{5}$$