Estimating the Parameters in a Mathematical Product Design Model

Takashi SHINZATO\textsuperscript{1}, Dongxiao JIANG\textsuperscript{2}, Mitsuhiro HOSHINO\textsuperscript{3}, and Ikou KAKU\textsuperscript{4}

\textsuperscript{1} Mori Arinori Center for Higher Education and Global Mobility, Hitotsubashi University, Kunitachi, Tokyo, Japan
\textsuperscript{2} Huawei Central Software Laboratory in ChinaSoft International Corporation, China
\textsuperscript{3} Department of Management Science and Engineering, Akita Prefectural University, Yurihonjo, Akita, Japan
\textsuperscript{4} Department of Environmental Management, Tokyo City University, Yokohama, Kanagawa, Japan

Abstract: Mathematical approaches for solving product design problem can be used by not only elder designers who have implicit/ambiguous knowledge of design experienced, but also younger designers even they have not so much experience, because such mathematical methods lead designers to obtain an optimal design automatically. However, there is a very serious lake to use such optimal methods in practice where the optimal solution has been obtained by a set of given coefficient parameters, which are so hard to be fixed accurately by even those most-experienced designers. In this paper, a set of simultaneous equations is built to represent the relations between product designs and correlative category labels. Parameter estimation is to fix the parameters that making all of relations to be feasible. A Boltzmann machine algorithm with belief propagation is used to perform the parameter estimation, where the number of simultaneous equations may be less than the number of parameters, to be solved possibly. Numerical experiments are provided to show the efficiency and utility of the algorithm. As a result, the value of training error is approaching to 0 slowly and stably, so it can be foresaw that parameters are estimated.

Key Words: TeX Users Group, markup

1. Introduction

Product design based on architecture theory is widely studied in past decades. Several papers have been published on product modularity and development of modular products and systems (see [1]–[5]). However, quite a few studies make a mention of cooperating integrity into the design process. The reason is that even so called global performance can be achieved through integral architectures[2]. It is hard to design a product through integral architecture. Sometimes a heavy weight system integrator[6] is necessary to coordinate the decomposed components. For designing through integral architecture, a mathematical model has been proposed in which the integrity is treated as the influence among interfaces[7]. Then a branch and bound method has been provided to solve the quadratic product design model[8]. Such mathematical approach for solving product design model is expected to improve both efficiency and utility of the algorithm. As a result, the value of training error is approaching to 0 slowly and stably, so it can be foresaw that parameters are estimated.

2. Estimation model based on product designs and their category labels

Let us set up the several variables and product design model so as to discuss briefly and systematically. We handle here product design of product that is composed by $D$ distinct devices, for instance, some assemblies in computer such as CPU, memory unit, fan unit. Intuitively, it is known that desirable function of the product is not always related with one device but also is complicatedly related with some devices. Thus we focus the complicated association between their devices as modeling of product design. As our first step of proposed motivation, the interface is represented using the pair of devices, for instance,
(a, b), where a, b ∈ {1, 2, · · · , D} are the device indices, that is we could characterize the performance of product design using the connecting state of whole interfaces. In general, although the number of interfaces is D(D − 1)/2, it may be implicitly and empirically known that the connecting state of the particular interface of optimal product design is obviously connected and/or disconnected. Then $\mathbf{S} = (S_1, S_2, · · · , S_N) \in \{±1\}^N$ is the connection of product design (it is hereafter shortened as product design) which is composed of N interfaces, that is, we do not need to handle whole interface since we exclude the interface that connecting state is already decided, here the number of already-decided interface states is $|\mathbf{h}|$. Moreover, $S_i = 1$ means interface $i$ is connected and $S_i = -1$ means it is disconnected.

Then performance of product designs $\mathbf{S}$ (named as cost function) can be evaluated by as follows;

$$H(\mathbf{S}, \mathbf{h}, J) = -\sum_{i=1}^{N} h_i S_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} J_{ij} S_i S_j,$$

$$= -\mathbf{h}^T \mathbf{S} - \frac{1}{2} \mathbf{S}^T J \mathbf{S}, \quad (1)$$

where $N$ dimensional vector $\mathbf{h} = (h_1, h_2, · · · , h_N)^T \in \mathbb{R}^N$ and an $N$-by-$N$ symmetric matrix $J = [J_{ij}] \in \mathbb{M}_{N \times N}$ are used, two kinds of coefficient parameters, $h_i$ and $J_{ij}$, represent single cost of interface itself and correlation cost between pair of interfaces $i, j$, respectively. Furthermore, since self-correlation is ignored, $J_{ii} = 0$ is defined and/or since constant terms in cost function is not influenced to optimality of product design, their terms are ignored. Notice that three distinct cost terms related with already-decided interface connection, that is (a) single cost term of already-decided interface, (b) correlation cost between already-decided and still-undecided interfaces, and (c) correlation cost between already-decided ones, are decomposed into constant cost term ((a) and (c) cases) and single cost term of still-undecided interface (case (b)).

If the single cost vector $\mathbf{h}$ and the correlation matrix $J$ are given, the optimal product design is characterized by product design $\mathbf{S}^*$ of the minimum cost function using proposed method[7],[8]. However, in practical product design environment, cost parameters $\mathbf{h}$ and $J$ are not easy to be decided even by those most-experienced designers. So we assume that several pairs of product design and its category label (instead of the value of cost function), for instance, good (in two categories case) or better-good-bad-worse (in four categories case) are obtained, since category label of given product design could be more easily and accurately provided than the value of cost function of product design by those most-experienced designers. Furthermore we discuss how to estimate effectively the parameters $\mathbf{h}$ and $J$ using the given pairs.

$M$ sets of pair of product design and its category label are prepared as follows;

$$d^\mu = (\mathbf{S}^\mu, K^\mu), \quad \mu = 1, · · · , M, \quad (2)$$

$$\mathbf{S}^\mu = (S^\mu_1, · · · , S^\mu_N) \in \{±1\}^N, \quad (3)$$

where $K^\mu$ is the correlated category label of product design $\mathbf{S}^\mu$ and $d^\mu$ is $\mu$th pair. In addition, we postulate that the relation between category label $K^\nu$ and cost function $H(\mathbf{S}^\nu, \mathbf{h}, J)$ is held as follows;

$$K^\nu = f \left( \frac{1}{N} H(\mathbf{S}^\nu, \mathbf{h}, J) \right), \quad (4)$$

where label function $f(s)$ is decomposed as

$$f(s) = \sum_{\mu=1}^{p} f_\nu(s),$$

$$f_\nu(s) = \frac{1}{1 + e^\gamma(s-\nu)}, \quad \gamma_\nu > 0, \quad (6)$$

where logistic function $f_\nu(s)$ which is differentiable possible is used. In limit case of $\gamma_1, · · · , \gamma_p \to \infty$, label function $f(s)$ will become Heaviside step function of step $p + 1$, that is $p + 1$ is the category number. For example, if $p = 1$, $K^\nu$ is two stages category label and $K^\nu = \lim_{\gamma \to \infty} f \left( \frac{1}{N} H(\mathbf{S}^\nu, \mathbf{h}, J) \right) \in \{0, 1\}$. $K^\nu = 1$ means product design is good and $K^\nu = 0$ means the opposite. And if $p = 3, K^\nu \in \{0, 1, 2, 3\}$, for instance, $K^\nu = 3$ means better, $K^\nu = 2$ does good, $K^\nu = 1$ does bad and $K^\nu = 0$ does worse. Because $f(s)$ is monotonically nonincreasing function with respect to $s$, that is $\frac{df(s)}{ds} = -\gamma f_\nu(s)(1 - f_\nu(s)) \leq 0$, intuitively it turns out that product design which holds smaller cost function always has higher category label.

3. Estimating parameters using Kullback-Leibler distance

Since those most-experienced designers are experimentally producing the product design which has a possibility of minimizing the product cost with implication, trial product designs have the possibility of including the clue of optimal product design. Thus, we assume that the observable data are generated from Boltzmann distribution which is conformed to the cost function of product design. Namely, we suppose that $\mathbf{S}^1, · · · , \mathbf{S}^M$ are generated by the following Boltzmann distribution with inverse temperature $\beta(> 0)$ and true cost parameters $\mathbf{R}^0$ and $J^0$;

$$P(\mathbf{S}, \mathbf{R}^0, J^0) = \frac{e^{-\beta H(\mathbf{S}, \mathbf{R}^0, J^0)}}{Z(\mathbf{R}^0, J^0)}, \quad (7)$$

where the partition function is defined as the normalized constant as follows;

$$Z(\mathbf{R}^0, J^0) = \sum_{\mathbf{S}} e^{-\beta H(\mathbf{S}, \mathbf{R}^0, J^0)} \quad (8)$$

Moreover the notation $\sum_\mathbf{S}$ is the sum over whole configuration of product design. In this definition, it turns out that the probability of product design of smaller cost function is comparatively higher since $e^{-\beta}$ is a monotonically decreasing function.

However if the number of observed data $M$ is not large enough (but also not small enough), that is trial product design is not inexhaustibly produced since the firm desires to reduce whole development cost as possible, empirical distribution

$$P^e(\mathbf{S}) = \frac{1}{M} \sum_{\mu=1}^{M} \prod_{i=1}^{N} \delta(\mathbf{S}_i, \mathbf{S}^\mu_i), \quad (9)$$

$M \geq N(N + 1)/2$, since the number of unknown parameters $h_i$ and $J_{ij}$ is less than the one of trial product design or linear equations from eq. (4), we resolve strictly them without Boltzmann machine learning.
is not always consistent with true distribution \( P(\bar{S}, \bar{J}, J^0) \), where Kronecker’s delta \( \delta(a, b) \) is already used, if \( a = b \), \( \delta(a, b) = 1 \) and otherwise \( \delta(a, b) = 0 \). Although empirical distribution is not consistent with true distribution in the case of finite \( M \) (but not small enough), if we are intuitively possible to regard that this data-driven distribution is similar to true distribution in the case of sufficiently-large \( M \), we learn from observed data so that we can estimate true cost parameters.

In learning phase, since it is well-known that cost parameters can be efficiently estimated using Kullback-Leibler distance [10], in this present paper we also use this distance. Kullback-Leibler distance (or relative entropy) is one of the most popular criteria of measuring similarity (or distance) of two probabilities in the cross-disciplinary fields of machine learning and information communication,

\[
KL(\bar{h}, J) = -\sum_S P^0(\bar{S}) \log \frac{P^0(\bar{S})}{P(\bar{S}, \bar{h}, J)} \tag{10}
\]

If \( P^0(\bar{S}) = P(\bar{S}, \bar{h}, J^0) \) is satisfied or \( M \rightarrow \infty \), we could minimize \( KL(\bar{h}, J) \) when and only when \( \bar{h} = \bar{h}^0 \) and \( J = J^0 \), and \( KL(\bar{h}^0, J^0) = 0 \) is obtained. On the other hand, in the case of finite \( M \), if empirical distribution \( P(\bar{S}) \) is not always consistent with and is similar to true distribution \( P(\bar{S}, \bar{h}, J^0) \), minimizing Kullback-Leibler distance with respect to cost parameters \( \bar{h} \) and \( J \) plays an important role for estimation of product design model. Kullback-Leibler distance is expanded as follows;

\[
KL(\bar{h}, J) = -\sum_S P^0(\bar{S}) \log P(\bar{S}, \bar{h}, J) - H[P^0] - \bar{\beta} \bar{h}^T \left( \sum_S P^0(\bar{S}) S \right) - \frac{\beta}{2} \text{Tr} \left( \sum_S P^0(\bar{S}) SS^T \right) + \log Z(\bar{h}, J) - H[P^0] - \bar{\beta} \bar{m}^0 - \frac{\beta}{2} T \text{M}^0 + \log Z(\bar{h}, J) - H[P^0], \tag{11}
\]

where \( H[P^0] = -\sum_S P^0(\bar{S}) \log P^0(\bar{S}) \) is Shannon information entropy of empirical distribution, and \( \bar{m}^0 = (m^0_1, m^0_2, \ldots, m^0_M)^T \in \mathbb{R}^M \) and \( M_0 = \{M^0_{ij}\} \in \mathbb{M}_{N \times N} \) are the first and second moments of empirical distribution, respectively. Furthermore, using observed data \( \bar{S}^1, \ldots, \bar{S}^M \),

\[
m^0_i = \sum_S S_i P^0(\bar{S}) = \frac{1}{M} \sum_{\mu=1}^M S^\mu_i \tag{12}
\]

\[
M^0_{ij} = \sum_S S_i S_j P^0(\bar{S}) = \frac{1}{M} \sum_{\mu=1}^M S^\mu_i S^\mu_j \tag{13}
\]

are readily calculated. In these definitions, it turns out that Shannon information entropy, the first and second moments of empirical distribution are possible to be regarded as constant with respect to cost parameters \( \bar{h} \) and \( J \).

**4. Boltzmann machine learning and proposed method**

**4.1 Boltzmann machine learning**

Boltzmann machine is a kind of stochastic recurrent neural network developed by Ackley, Hinton and Sejnowski[9] and is defined on undirected complete graph with rich structure.

![Fig. 1 Estimation of true distribution using Boltzmann machine learning.](image)

Moreover, shown in figure 1, the learning policy of Boltzmann machine is to estimate model parameter (of trial distribution) which is able to minimize Kullback-Leibler distance between trial distribution and experimental distribution evaluated using sample data generated from generative model (or generative distribution) of which statistical property is not known. Thus, from the extremum conditions of Kullback-Leibler distance,

\[
\frac{\partial KL}{\partial h_i} = -\beta m^0_i + \frac{\partial G}{\partial h_i}, \tag{14}
\]

\[
\frac{\partial KL}{\partial j_{ij}} = -\frac{\beta}{2} M^0_{ij} + \frac{\partial G}{\partial j_{ij}}, \tag{15}
\]

our desired model parameters are derived where \( KL = KL(\bar{h}, J) \) and \( G = \log Z(\bar{h}, J) \) are already used. However in general, when the number of sample data \( M \) is not limited to infinity, that is empirical distribution is not consistent with true distribution, it is not always a unique solution of model parameters. In actual situation, using a certain criterion we need to determine the appropriate solution in solution space of optimal model parameters.

Our proposed analytical idea is simple, that is, if given sample data generated from true distribution and these related information, for instance, category labels of observed data (see eq. (4), we are possible to derive the appropriate solution based on these data-driven information. The fundamental approach and our proposed approach of Boltzmann machine learning are explained as follows;

**(1) Boltzmann machine learning without constraint conditions**

A general approach of Boltzmann machine learning is to minimize Kullback-Leibler distance with respect to model parameters as follows;

\[
(\bar{h}^*, J^*) = \arg \min_{\bar{h}, J} KL(\bar{h}, J). \tag{16}
\]

Moreover Kullback-Leibler distance is regarded as analytical function on model parameters, one is able to apply steepest descent method as follows;

\[
h_i^{t+1} = h_i^t - \eta h \frac{\partial KL}{\partial h_i}, \tag{17}
\]

\[
J_{ij}^{t+1} = J_{ij}^t - \eta j \frac{\partial KL}{\partial J_{ij}}, \tag{18}
\]
where small non-negative learning steps \( \eta_0 \) and \( \eta_1 \) are used and \( h_i^t \) and \( J_{ij}^t \) are single cost parameter and correlation cost parameter at iteration of \( t \) step, respectively.

(2) Boltzmann machine learning with constraint conditions

On the other hand, our proposed approach is Boltzmann machine learning with several constraint conditions as follows;

\[
(h^{**}, J^{**}) = \arg \min_{h_i, J_{ij}} KL(h, J)
\]

subject to \( M \) constraints in eq. (4), (19)

where we need to estimate the optimal model parameters with several constraint conditions.

Thus the following Lagrange multiplier function is defined as

\[
L = -\beta h^T \Phi \Phi^T h - \frac{\beta}{2} T_{ve} J^T M^T \Phi \Phi^T G^T M J + G + \frac{1}{M} \sum_{i=1}^{M} \epsilon_{ij}(K_{0}-f(\phi^t)),
\]

where \( \phi^t = \frac{1}{N} H(\tilde{S}^h, \tilde{S}^v, \tilde{H}) \) is abbreviated and the label of true model which has single cost \( h^0 \) and correlation cost \( J^0 \),

\[
K_{0} = f \left( \frac{1}{N} H(\tilde{S}^h, \tilde{S}^v, \tilde{H}), J^0 \right)
\]

is used.

Three points should be noted here. First, Kullback-Leibler distance includes constant term \( H[\tilde{P}_M] \) and non-constant terms (the others) with respect to model parameters, since the extremum of Kullback-Leibler distance is not related with the constant term, we ignore the constant term in optimization. Second, equality constraints of eq. (4) are represented as added auxiliary function and auxiliary variable \( \epsilon_{ij} \) is non-negative number. Thus, the extremum of Lagrange multiplier function with respect to single cost, correlation cost and auxiliary variable are assessed as follows;

\[
\frac{\partial L}{\partial h_i} = -\beta \partial m_i^0 \Phi^T \Phi h + \frac{1}{N M} \sum_{j=1}^{M} \epsilon_{ij}(\phi^t) S_i^0,
\]

\[
\frac{\partial L}{\partial J_{ij}} = -\beta \frac{J_{ij}}{2} M_i^0 + \frac{1}{2 N M} \sum_{j=1}^{M} \epsilon_{ij}(\phi^t) S_i^0 S_j^0,
\]

\[
\frac{\partial L}{\partial \epsilon_{ij}} = K_{0} - f(\phi^t),
\]

where \( \partial m_i^0 = -S_i^0/N \) and \( \partial M_i^0 = -S_i^0 S_j^0/2N \) are utilized. In addition, in order to resolve the optimization problem, we employ steepest descend method as follows;

\[
h_i^{t+1} = h_i^t - \eta_0 \frac{\partial L}{\partial h_i},
\]

\[
J_{ij}^{t+1} = J_{ij}^t - \eta_1 \frac{\partial L}{\partial J_{ij}},
\]

\[
\epsilon_{ij}^{t+1} = \epsilon_{ij}^t - \eta_2 \frac{\partial L}{\partial \epsilon_{ij}},
\]

where \( \eta_0, \eta_1 \) and \( \eta_2 \) are non-negative learning steps and \( h_i^t \), \( J_{ij}^t \) and \( \epsilon_{ij}^t \) are also single cost estimation, correlation cost estimation and auxiliary variable at iteration of \( t \) step, respectively.

Lastly, it turns out that the partial derivatives of \( G \) in eq. (22) and eq. (23) are so hard to be assessed straightforwardly since from the definition of \( G \), namely, \( G = \log Z(\tilde{H}, \tilde{J}) \), the calculation quantity needed by evaluation of the partition function \( Z(\tilde{H}, \tilde{J}) \) is \( 2^N \) from eq. (8). Thus, we asymptotically analyze the partial derivatives of logarithm function of partition function using Thouless-Anderson-Palmer equation developed in statistical mechanics.

4.2 Thouless-Anderson-Palmer equation

Here we introduce Thouless-Anderson-Palmer equation invented in statistical mechanics, which can reveal the macroscopic properties of many-body systems observed in practical physical experiments, in order to solve the above-mentioned NP-hard problem[12]. Firstly, the expectation of coupled state of interface \( i \) is represented as \( m_i \), that is, \( m_i = \sum_S P(\tilde{S}^i, \tilde{H}, J)S_i \). Based on Thouless-Anderson-Palmer equation [8],[11], if fixed \( \tilde{H} \) and \( J \), \( S_i \) is affected by the sum of three terms, the single cost \( \beta h_i \), the correlation cost of the whole neighborhood \( \beta \sum_{j=1}^{N} J_{ij} m_j \) and Onsager reaction term \( \beta^2 \sigma_j^2 (1-q) m_i \) [11],[12], hence \( m_i \) is estimated as follows;

\[
m_i = \tan \left( \beta h_i + \beta \sum_{j=1}^{N} J_{ij} m_j - \beta^2 \sigma_j^2 (1-q) m_i \right),
\]

\[
q = \frac{1}{N} \sum_{j=1}^{N} m_j^2,
\]

where since this calculation amount of Thouless-Anderson-Palmer equation is almost \( O(N^2) \), it turns out that it is feasible to carry out comparatively. Furthermore as the first step of our research, we assume that correlation cost is independently and identically distributed with normal distribution with mean \( \mu_j/N \) and variance \( \sigma_j^2 /N \), that is \( J_{ij} \sim N(\mu_j/N, \sigma_j^2/N) \). Or since fixed \( J_{ij} \), sample mean and sample variance of correlation cost are

\[
\frac{\mu_j}{N} = \frac{2}{N(N-1)} \sum_{j=1}^{N} \sum_{j=1}^{N} J_{ij},
\]

\[
\frac{\sigma_j^2}{N} = \frac{2}{N(N-1)} \sum_{j=1}^{N} \sum_{j=1}^{N} (J_{ij} - \mu_j/N)^2.
\]

Note that \( J_{ij} \sim O(1/N) \) since \( \sum_{j=1}^{N} J_{ij} m_j \sim O(1) \) and \( h_i \sim O(1) \) are needed in valance [8].

4.3 Proposed faster algorithm

Although in the previous subsection, if fixed \( h_i \) and \( J_{ij} \), mean of interface \( i \) with Boltzmann distribution of trial model \( P(\tilde{S}^i, \tilde{H}, J) \) is assessed using Thouless-Anderson-Palmer equation, moreover cost parameters \( h_i \) and \( J_{ij} \) and auxiliary parameter \( \epsilon_{ij} \) are calculated using the steepest descend method in subsection 4.1, since this approach is converged comparatively slowly, we need to propose the faster algorithm to solve efficiently.

From \( \frac{\partial L}{\partial h_i} = \frac{\partial L}{\partial J_{ij}} = 0 \) and \( \frac{\partial L}{\partial \epsilon_{ij}} = \frac{\partial L}{\partial \epsilon_{ij}} = \beta \sum_i S_i \sum_j P(\tilde{S}^i, \tilde{H}, J) S_i = \beta m_i \) and \( \frac{\partial L}{\partial \epsilon_{ij}} = \beta \sum_i S_i \sum_j P(\tilde{S}^i, \tilde{H}, J) S_i \),

\[
m_i = m_i^0 - \frac{1}{N M \beta} \sum_{j=1}^{M} \epsilon_{ij}(\phi^t) S_i^0,
\]

\[
\sum_i S_i P(\tilde{S}^i, \tilde{H}, J) S_i = M_i - \frac{1}{N M \beta} \sum_{j=1}^{M} \epsilon_{ij}(\phi^t) S_i^0 S_i^0.
\]
are obtained. That is, \( m_j \) is derived from eq. (32) not eq. (28),
\[
h_i = \frac{1}{\beta} \tanh^{-1} m_i - \sum_{j=1}^{N} J_{ij} m_j + \beta \sigma_j^2 (1 - q) m_i
\]
(34)
is calculated where \( \tanh^{-1} u = \frac{1}{2} \log \frac{1 + u}{1 - u} \) is used. On the other hand, the correlation cost is also assessed. That is, the covariance of interface \( i \) and interface \( j \),
\[
\chi_{ij} = \sum_{S} P(S, \tilde{h}, J) S_i S_j - m_i m_j
\]
is rewritten as
\[
\chi_{ij} = \sum_{S} \left( \frac{\partial^2 G}{\partial \tilde{h} \partial \tilde{h}} - \frac{\partial^2 G}{\partial \tilde{h} \partial \phi} \right)_{S_i S_j}
\]
(36)
Then,
\[
J = -\frac{1}{\beta} \chi^{-1} + \frac{1}{\beta} \left( \Lambda + \beta^2 \sigma_j^2 (1 - q) I - \frac{2 \beta^2 \sigma_j^2}{N} \bar{m} \bar{m}^T \right)
\]
(37)
is obtained where \( \Lambda = \text{diag} \left\{ -\frac{1}{\lambda_1}, \ldots, -\frac{1}{\lambda_M} \right\} \in M_{N \times N}, \chi = \left\{ \chi_{ij} \right\} \in M_{N \times N} \) and identity matrix \( I \in M_{N \times N} \) and \( \bar{m} = (m_1, \ldots, m_N)^T \in R^N \) are already assessed. That is, the correlation cost \( J_{ij} \) is also derived from eq. (37) not eq. (26). Lastly, auxiliary parameter
\[
e_{ij}^{x+1} = e_{ij}^x - \eta_i (K_{ij}^0 - f(\phi^i))
\]
is updated. Thus our proposed faster algorithm for Boltzmann machine learning with constraints is summarized as follows;

Proposed faster algorithm

\[
J = -\frac{1}{\beta} \chi^{-1} + \frac{1}{\beta} \left( \Lambda + \beta^2 \sigma_j^2 (1 - q) I - \frac{2 \beta^2 \sigma_j^2}{N} \bar{m} \bar{m}^T \right)
\]
\[
h_i = \frac{1}{\beta} \tanh^{-1} m_i - \sum_{j=1}^{N} J_{ij} m_j + \beta \sigma_j^2 (1 - q) m_i
\]
\[
m_i = m_i^0 - \frac{1}{NM \beta} \sum_{\mu=1}^{M} e_{i\mu} f'(\phi^i) S_{i\mu}^0
\]
\[
\chi_{ij} = M_{ij}^0 - \frac{1}{NM \beta} \sum_{\mu=1}^{M} e_{i\mu} f'(\phi^i) S_{i\mu}^0 S_{j\mu} - m_i m_j
\]
\[
\mu_j = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{ij}
\]
\[
\sigma_j^2 = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( J_{ij} - \frac{\mu_j}{N} \right)^2
\]
\[
q = \frac{1}{N} \sum_{i=1}^{N} m_i^2
\]
\[
\phi^i = \frac{1}{N} H(S_{i\mu}, \tilde{h}, J)
\]
\[
\Lambda_w = \frac{1}{1 - m_i^2}
\]
\[
e_{ij}^{x+1} = e_{ij}^x - \eta_i (K_{ij}^0 - f(\phi^i))
\]

Note that no constraint case in general form of Boltzmann machine learning represents using \( e_{ij}^x = 0 \).

Fig. 2 The both behaviours of \( \Delta_i \) in the case that correlation cost is known.

Fig. 3 The difference between delta of \( h_i^t \) and \( h_i^{t-1} \) without constraints \( \Delta_i^{0t} \) and delta of \( h_i^t \) and \( h_i^{t-1} \) with constraints \( \Delta_i^{wt} \) in the case that correlation cost is known.

5. Numerical results

5.1 Criteria of Boltzmann machine learning

Above mentioned, we need to compare both approaches of Boltzmann machine learning with and without label constraints. If true cost parameters \( h_i^0 \) and \( J_{ij}^0 \) are given, the validity of both approaches is evaluated using the following criteria,
\[
\Delta_i^{0t} = \sum_{t=1}^{N} \left| h_i^0 - h_i^t \right|
\]
(39)
\[
\Delta_i^{wt} = \sum_{t=1}^{N} \sum_{j=1}^{N} \left| J_{ij}^0 - J_{ij}^t \right|
\]
(40)

However, generally speaking, since true cost parameters are unknown in practical situation, \( \Delta_i^{0t} \) and \( \Delta_i^{wt} \) are not possible to be estimated in principle. Here, we apply the following criteria at estimating step \( t \) instead of \( \Delta_i^{0t} \) and \( \Delta_i^{wt} \),
\[
\Delta_i = \sum_{t=1}^{N} \left| h_i^0 - h_i^{t-1} \right|
\]
(41)
\[
\Delta_j = \sum_{t=1}^{N} \sum_{j=1}^{N} \left| J_{ij}^0 - J_{ij}^{t-1} \right|
\]
(42)

where \( h_i^t \) and \( J_{ij}^t \) are estimated cost parameters at step \( t \). Moreover,
\[
C' = \frac{1}{2M} \sum_{\mu=1}^{M} \left( K_{i\mu}^w - f(\phi_i^t) \right)^2
\]
(43)
is also evaluated where \( \phi_i^t = \frac{1}{N} H(S_{i\mu}, \tilde{h}, J) \) is used. In this paper, since true parameters are assigned as controllable setting, we estimate not only \( \Lambda_w, \Delta_i, C' \) but also \( \Delta_i^{0t} \) and \( \Delta_i^{wt} \).

Hereafter, the setting of numerical experiment in our discussion is as follows; the number of interfaces \( N = 10 \) and the
number of trial product designs \( M = 30 \), single cost \( h_0^i = 1 \) and correlation cost \( J_{ij}^0 = 1/N \) are uniquely assigned. Further we generate trial product designs \( S^i \) using Markow Chain Monte Carlo method and Boltzmann distribution with \( \beta = 4 \). Classification label is two, that is, trial product design is evaluated as good or bad (\( p = 1 \) case) by most-experienced designers, that is, category label function is \( f(u) = 1/(1+e^{\gamma u}) \). Moreover \( \gamma = 2 \) is employed here.

5.2 The case that correlation cost is known and single cost is not known

Firstly, we discuss the case that correlation cost \( J_{ij} \) is known and single cost \( h_i \) is not known, that is, we can assess only three criteria \( \Delta h_i \), \( \Delta^0 h_i \) and \( C_i \), since \( \Delta^1 h_i = \Delta^0 h_i = 0 \) is obtained. The behaviours of three criteria of both Boltzmann machine learning algorithms with and without constraints are shown in from figure 2 to figure 7. From figure 2 and figure 3, it turns out that delta of \( h^i \) and \( h^{i-1} \) without constraints \( \Delta h_i \) is not smaller than delta of \( h^i \) and \( h^{i-1} \) with constraints \( \Delta^0 h_i \). On the other hand, from figure 4 and figure 5, it turns out that delta of \( h^i \) and \( h^{i-1} \) without constraints \( \Delta h_i \) is not larger than delta of \( h^i \) and \( h^{i-1} \) with constraints \( \Delta^0 h_i \). The finding of our proposed method is more close to true parameter than the one of previous method. Thus from figure 6 and figure 7, since \( C_i \) is not smaller than \( C_i \), the effectiveness of our proposed method is confirmed in this case.

5.3 The case that correlation cost and single cost are not known

Next, we discuss the case that correlation cost \( J_{ij} \) and single cost \( h_i \) are not known, that is, we can assess five criteria \( \Delta h_i \), \( \Delta^0 h_i \), \( \Delta^1 h_i \), \( \Delta^0 h_i \) and \( C_i \). The behaviours of five criteria of both Boltzmann machine learning algorithms with and without constraints are shown in from figure 8 to figure 17. In similar way, the effectiveness of our proposed method is also confirmed in this case.

6. Conclusion

In this paper, in order to estimate a set of parameters used in the optimal design model, we construct an estimation model that can calculate mathematically the coefficient parameters, which are very hard to be fixed accurately by tacit knowledge of designers. A set of simultaneous equations that represent the re-
lations between product designs and correlative category labels has been built to carry out the parameter estimation. Moreover Boltzmann machine algorithm with Thouless-Anderson-Palmer equation is used here to perform the parameter estimation, where the number of simultaneous equations may be less than the number of parameters. Numerical experiments are provided to show the efficiency and utility of the algorithm. The results showed that training error is decreasing to 0 slowly and stably, so it can be foresaw that parameters are estimated approximately. As the future works, firstly it is necessary to verify the proposed approach based on practical data that come from designers. Secondly, though we propose an approach to solve parameter estimation of product design mathematically, but some methods be interest on that can give direct definition of parameter. Last but not least, the comparative study of existed product design evaluation methods and the proposed product design based on product architecture is also important.
A novel product design model has been proposed in order to develop a new product. However, the optimal solution is obtained by the given coefficient parameters, which are very difficult to be fixed accurately by tacit knowledge of designers. So we built a set of simultaneous equations that represent the relations between product designs and correlative category labels to carry out the parameter estimation. And a Boltzmann machine algorithm with Thouless-Anderson-Palmer equation is used here to perform the parameter estimation, where the number of simultaneous equations less than the number of parameters. Numerical experiments are provided to show the efficiency and utility of the algorithm. The results showed that training error is decreasing to 0 slowly and stably, so it can be foresaw that parameters are estimated approximately.

As the future works, firstly it is necessary to verify the proposed approach based on practical data that come from designers. Secondly, though we propose an approach to solve parameter estimation of product design mathematically, but some methods which can give direct definition of parameters are still interest us. Last but not least, the comparative study of existed methods to carry out the parameter estimation. And a Boltzmann machine algorithm with Thouless-Anderson-Palmer equation is used here to perform the parameter estimation, where the number of simultaneous equations less than the number of parameters. Numerical experiments are provided to show the efficiency and utility of the algorithm. The results showed that training error is decreasing to 0 slowly and stably, so it can be foresaw that parameters are estimated approximately.

References

[1] Ulrich, K. and Tung, K.,(1991). Fundamentals of Product Modularity, Issues Design Manufacturing/Integration, DE-39.
[2] Pimmler, T. U. and Epping, S. D., 1994. Integration Analysis of Product Decompositions, Proceeding of ASME 6th International conference on Design Theory and Methodology, DE 68.
[3] O’Grady, P., 1999. The Age of Modularity: Using the New World of Modular Products to Revolutionize Your Corporation, New York: Wiley.
[4] Kusiak, A. and Huang, C. C., 1996. Development of Modular Products, IIE Transactions, Components, Packaging, and Manufacturing Technology, A, 19(4), 523-538.
[5] He, D. W. and Kusiak, A., 1997. Design of Assembly Systems for Modular Products, IIE Transactions. Robotics and Automation, 13(5), 646-655.
[6] Clark, K. B. and Fujimoto, T., 1990. The Power of Product Integrity, Harvard Business Review,68(6), 107-118.
[7] Kaku, I., Jiang, D-X. and Shinzato, T., 2010. A New Product Design Approach through Integral Architecture. ICIC Express letters, 5(3), 1709-1716.
[8] Jiang, D-X., Shinzato, T., Hoshino, M., Zhang, R-Q. and Kaku, I., 2014. A Branch and Bound Approach for Solving Quadratic Product Design Model and its Linear Programming Transformation, 65(1), 9-16.
[9] Ackley, D. H., Hinton, G. E. and Sejnowski, T. J., 1985. A Learning Algorithm for Boltzmann Machines, Cognitive Science, 9(1), 147-169.
[10] Bishop, C. M., 2006. Pattern Recognition and Machine Learning, Springer.
[11] Plefka, T., 1982, Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model, Journal of Physcs A, 15(6), 1971-1978.
[12] Nishimori, H. and Ortiz, G., 2011. Elements of Phase Transitions and Critical Phenomena, Oxford University Press.