数理解析研究所講究録 1461 巻 2005 年 148-153

Knowledge Discovery from Questionnaire Survey by Nonlinear Optimization

Koji Okuhara[†], Junko Shibata[††], Nobuyuki Ueno[†] and Hiroaki Ishii[†††]

[†] Dept. of Management and Information Sciences, Hiroshima Prefectural University
[††] Dept. of Economics, Kobe Gakuin University
[†††] Graduate School of Information Science and Technology, Osaka University

1. Introduction

In this study we will propose framework of data mining based on statistical modeling. Especially, we focus on knowledge discovery for commerce, marketing and customer satisfaction. In the business scene, questionnaire survey is often used for extracting useful information. It has, however, been difficult to obtain meaningful rules for Kansei modeling, because such sampling data involve much ambiguity. We can list rare data, doubtful data and contradiction data, as main problems which should be careful. From theoretical view point, these problems should be solved in unified way.

As is often the case, it should be thought that uncertainty is contained in data, therefore rules estimated from original data directly are not necessarily reliable. It is widely known that pre-processing to sampling data is important to extract useful information in data mining. Thus, we first apply nonlinear method II of quantification[1] to extract information from data as much as possible. Its optimal mapping can be derived by conditional expectation. We will execute nonlinear method II of quantification by constructing Gaussian mixture model[2]. Gaussian mixture model estimates its parameters by maximizing log-likelihood function for incomplete data. From the result, it is shown that rare sample data can be detected based on occurrence probability. Detection of doubtful data is derived by maximizing entropy[3]. Since maximizing entropy is equivalent to minimizing free energy, clustering for detecting doubtful sample data is executed by melting method which is steepest descent method about free energy. After removing rare data and doubtful data from sampling data, we can try to extract useful rules without contradiction by using dominance-based rough set approach[4] considering ordinal relation. Proposal framework of data mining we will expand can treat sampling data with ambiguity and uncertain in unified theoretical way.

2. Data Mining from Questionnaire Survey

To begin with, I would like to introduce outline of Kansei modeling[5]. Personal image from product to Kansei word is collected by questionnaire survey. To obtain useful product development based on knowledge, this data analysis part is one of important steps. The main content I'm going to present from now on to construct data mining method in Kansei engineering. For example, in case of our purpose is to reveal human image received from front design of car. First, subjets are shown pictures of front design of car. And they answer to reply sheet about Kansei words used for semantic differential scale technique[6]. From these two information both a characteristic of objects we consider and the reply sheet, we can prepare decision table from condition attributes to decision attributes. We will analyze such decision table.

I'll be reminding you about subjects which should be solved for data mining. Main 4 problems will be discussed here such as projection of condition attributes, decision pending for rare data, detection of doubtful data and exclusion of contradiction rules. For example,

when we have such decision table in figure 1, we can obtain these rules describing samples. However, generally speaking, attributes have meaning. If it is assumed to be C1: deluxe, C2: safety, D1: impression, and 1: low, 2: high, then the data C1 and C2 are high is better than they are low for D1 and these two rules are regarded as contradiction rules. In this case, acceptable rules can be obtained by applying dominance-based rough set approach considering ordinal relation. So in this paper, after applying proposal pre-processing, it is assumed that we will adopt dominance-based rough set approach.

Decision table

No.	C_1	C_2	D_1
1	1	2	1
2	2	1	2
3	1	1	2
4	2	2	2

Decision rules

IF $C_1 = 1$ *and* $C_2 = 2$ THEN $D_1 = 1$, → Contradicting rule

IF $C_1 = 2$ *and* $C_2 = 1$ THEN $D_1 = 2$,

IF $C_1 = 1$ *and* $C_2 = 1$ THEN $D_1 = 2$, → Contradicting rule

IF $C_1 = 2$ *and* $C_2 = 2$ THEN $D_1 = 2$.

Figure 1: Acceptable rules obtained by applying dominance-based rough set approach

3. Proposal Framework for Data Mining

In case the condition attributes are nominal scale, or in case the mapping from condition attributes to decision attribute is not simple, it is known that projection of condition attributes to interval works well[7]. In order to derive interval, interval regression model for qualitative data is solved by simplex method conventionally. In this paper, we propose application of Gaussian mixture model to obtain such interval, because we can treat sampling data with ambiguity and uncertain in unified theoretical way.

Now I would like to focus on proposal framework of data mining for Kansei engineering. Generally, it should be thought that uncertainty is contained in data, therefore the rules based on reduct are not necessarily reliable. To solve such problem, we propose framework of data mining for Kansei engineering as figure 2.

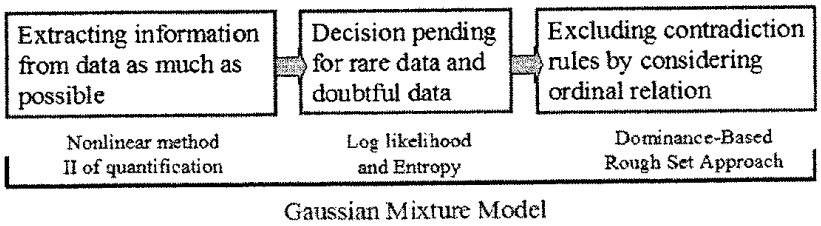

Figure 2: Proposal framework of data mining for Kansei engineering

We first apply nonlinear method II of quantification for extracting information from data as much as possible, secondly, log likelihood and entropy for decision pending for rare data and doubtful data, finally, dominance-based rough set approach for excluding contradiction rules by considering ordinal relation. We treats these processes based on Gaussian mixture model in a unified way.

4. Data Mining from Questionnaire Survey

4. 1 Outline of Nonlinear method II of quantification

Nonlinear method II of quantification[1] is used for extracting information from data as much as possible. Squared error function is minimized for condition attributes $\mathbf{x}_s$ as explanatory variables and projection $\mathbf{y}_s = \Lambda^{-\frac{1}{2}}\mathbf{u}_s$ of condition attributes as explained variables.

$$\begin{aligned}\epsilon^2(\mathbf{x},\mathbf{y}) &= \mathrm{E}[\|\mathbf{y}-\phi(\mathbf{x})\|^2] \\ &= \int\int \|\mathbf{y}-\phi(\mathbf{x})\|^2 p(\mathbf{y}|\mathbf{x})d\mathbf{y}p(\mathbf{x})d\mathbf{x}, \end{aligned} \tag{1}$$

where $\Lambda^{-\frac{1}{2}}$ is defined by

$$\Lambda^{-\frac{1}{2}} = \mathrm{diag}\left[\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \cdots, \frac{1}{\sqrt{\lambda_M}}\right]^{\mathrm{T}} \in \Re^M \times \Re^M. \tag{2}$$

It satisfies

$$\left[\Gamma - p_\Theta p \Theta^{\mathrm{T}}\right] U = P_\Theta U \Lambda \tag{3}$$

where

$$U = [\mathbf{u}_1 \mathbf{u}_2 \cdots \mathbf{u}_s]^{\mathrm{T}} \in \Re^s \times \Re^M, \tag{4}$$

and $\mathbf{u}_s \in \Re^M \times \Re^1$. Γ is vector which consists of joint probability, $[p(i,j)] \in \Re^K \times \Re^1$, and P_Θ is matrix whose ij element consists of occurrence probability, $p(k)\sigma_{ij} \in \Re^K \times \Re^K$, then $p_\Theta = P_\Theta \mathbf{1}_K \in \Re^K \times \Re^1$ and $\mathbf{1}_k = [1, 1, \cdots, 1]^{\mathrm{T}} \in \Re^K \times \Re^1$.

The estimated optimal mapping can be given by this conditional expectation,

$$\hat{\mathbf{y}} = \phi(\mathbf{x}) = \int \mathbf{y} p(\mathbf{y}|\mathbf{x}) d\mathbf{y}. \tag{5}$$

In this paper we apply Gaussian mixture to solve nonlinear method II of quantification.

4. 2 Gaussian Mixture for Nonlinear Regression Analysis

Gaussian mixture[2] for nonlinear regression analysis is used for decision pending for rare data and derivation of interval of rough set analysis. Gaussian mixture is neural networks that estimates a probability density function by adding up some radial basis functions. Normal distribution is used as a radial basis function. Here we describe the dynamics of Gaussian mixture model with K input neurons and 1 output neurons which is used in this study for estimating the interval efficiency[8].

d dimensional sth input vector $\mathbf{Z}_s \in \Re^d$, $(s = 1, 2, \cdots, S)$ is inputted into all input neurons where S denotes a number of sampling data. kth input neuron $(k = 1, 2, \cdots, K)$ has parameter ϕ_k and weight w_k. Parameters ϕ_k represents an average vector and a set of covariance matrix $\{\mathbf{m}_k, \mathbf{\Sigma}_k\}$ where $\mathbf{m}_k = [m_k^1, m_k^2, \cdots, m_k^d]^{\mathrm{T}} \in \Re^d$ and $\mathbf{\Sigma}_k$ is $d \times d$ matrix whose iith element is a variance σ_k^{ii}. $\mathbf{\Sigma}_k$ is a diagonal matrix and a positive define symmetric matrix. $\mathbf{w}$ represents a set $\{w_1, w_2, \cdots, w_K\}$ and ϕ represents a set$\{\phi_1, \phi_2, \cdots, \phi_K\}$. Furthermore, a set of $\mathbf{w}$ and ϕ is expressed with parameter θ.

An output vector of a system can be given by

$$\mathrm{E}[\mathbf{Y}|\mathbf{X}_s] = \int_{\Re^m} \mathbf{y} p(\mathbf{Y}|\mathbf{X}_s, \theta') d\mathbf{y} \tag{6}$$

where the parameter θ' denotes a set of $\mathbf{w}$ and ϕ'. The conditional probability density function in the system is

$$p(\mathbf{Y}|\mathbf{X}_s, \theta') = \sum_{k=1}^{K} \alpha(k) p_k(\mathbf{Y}|\mathbf{X}_s, \phi_k') \tag{7}$$

where

$$\alpha(k) = \frac{p(k) p_k(\mathbf{X}_s|\phi_k')}{\sum_{k=1}^{K} p(k) p_k(\mathbf{X}_s|\phi_k')} \tag{8}$$

and the output vector of a system is rewritten by

$$\begin{aligned} \mathrm{E}[\mathbf{Y}|\mathbf{X}_s] &= \sum_{k=1}^{K} \alpha(k) \int_{\Re^m} \mathbf{y} p_k(\mathbf{Y}|\mathbf{X}_s, \phi_k') d\mathbf{y} \\ &= \sum_{k=1}^{K} \alpha(k) \mathrm{E}_k[\mathbf{Y}|\mathbf{X}_s] \end{aligned} \tag{9}$$

Here let $\{\mathbf{C}_k^x\}^{-1}$ be an inverse matrix of covariance matrix $\mathbf{C}_k^x$, and let a matrix $\mathbf{D}_k$ be

$$\mathbf{D}_k = \mathbf{C}_k^{yx} \{\mathbf{C}_k^x\}^{-1} \tag{10}$$

then we have

$$\mathrm{E}_k[\mathbf{Y}|\mathbf{X}_s] = \mathbf{m}'^y_k + \mathbf{D}_k(\mathbf{x}_s - \mathbf{m}'^x_k) \tag{11}$$

and

$$\mathrm{E}[\mathbf{Y}|\mathbf{X}_s] = \sum_{k=1}^{K} \alpha(k) \left\{\mathbf{m}'^y_k + \mathbf{D}_k(\mathbf{x}_s - \mathbf{m}'^x_k)\right\} \tag{12}$$

Moreover, in the kth input neurons, the conditional probability density function $p_k(\mathbf{Y}|\mathbf{X}_s, \phi'_k)$ obeys a probability density function that is m dimensional normal distribution with a mean vector $\mathrm{E}_k[\mathbf{Y}|\mathbf{X}_s] \in \Re^m$ and a covariance matrix $\mathbf{D}'_k \in \Re^m \times \Re^m$ is given by

$$\mathbf{D}'_k = \mathbf{C}_k^y - \mathbf{D}_k \mathbf{C}_k^x \mathbf{D}_k^{\mathrm{T}} \tag{13}$$

We can derive the conditional probability density function $p(\mathbf{Y}|\mathbf{X}_s, \theta')$ in total system.

Then each estimator of parameter Θ is acquired by maximizing the logarithmic function

$$L(\theta) = \sum_{s=1}^{S} \log p(\mathbf{Z}_s|\theta) \tag{14}$$

We can obtain these estimators by using iterative calculations derived by Expectation Maximization algorithm[9].

$$w_k^{(t+1)} = \frac{1}{S} \sum_{s=1}^{S} h_k^{(t)}(\mathbf{Z}_s) \tag{15}$$

$$m_k^{(t+1)} = \frac{\sum_{s=1}^{S} \mathbf{Z}_s h_k^{(t)}(\mathbf{Z}_s)}{\sum_{s=1}^{S} h_k^{(t)}(\mathbf{Z}_s)} \tag{16}$$

$$\mathbf{\Sigma}_k^{(t+1)} = \frac{\sum_{s=1}^{S} (\mathbf{Z}_s - \mathbf{m}_k^{(t)})(\mathbf{Z}_s - \mathbf{m}_k^{(t)})^{\mathrm{T}} h_k^{(t)}(\mathbf{Z}_s)}{\sum_{s=1}^{S} h_k^{(t)}(\mathbf{Z}_s)} \tag{17}$$

where

$$h_k^{(t)}(\mathbf{Z}_s) = \frac{w_k^{(t)} N_d(\mathbf{Z}_s, \phi^{(t)})}{\sum_{s=1}^{S} w_k^{(t)} N_d(\mathbf{Z}_s, \phi^{(t)})} \tag{18}$$

Above update rule about the parameter θ is derived from condition maximizing a conditional expectation of log-likelihood function

$$\begin{aligned} Q(\theta|\theta^{(t)}) &= \mathrm{E}[L(\theta, k)|\mathbf{Z}_s, \theta^{(t)}] \\ &= \sum_{s=1}^{S} \sum_{k=1}^{K} h_k^{(t)}(\mathbf{Z}_s) \log p(\mathbf{Z}_s, k|\theta) \end{aligned} \tag{19}$$

We can consider a steepest descent method to acquire the parameter θ maximizing log-likelihood function $L(\theta)$, however a number of learning might be changed by learning coefficient. Therefore, we apply EM algorithm to proposed system as learning algorithm. The reason making a covariance matrix $\mathbf{\Sigma}_k$ a $\mathrm{diag}(\sigma_k^{ii})$ is to avoid to unstability from non-diagonal element. If we apply the nonlinear regressive analysis to the sampling data,

then we can give the following ranges $S[s_1, s_2]$ as the interval efficiency (see figure 3).

$$s_1 = \sum_{k=1}^{k} \alpha_k^x \mathrm{m}_k^y - 3\sqrt{\sum_{k=1}^{k} (\alpha_k^x)^2 \mathrm{C}_k^y} \tag{20}$$

$$s_2 = \sum_{k=1}^{k} \alpha_k^x \mathrm{m}_k^y + 3\sqrt{\sum_{k=1}^{k} (\alpha_k^x)^2 \mathrm{C}_k^y} \tag{21}$$

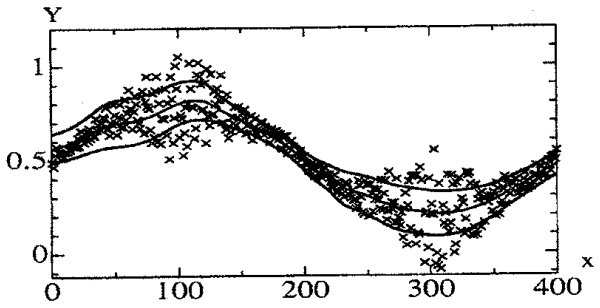

Figure 3: Conceptual figure of interval by nonlinear estimation

We can try to extract useful rules without contradiction by using dominance-based rough set approach[10] considering ordinal relation and interval.

4. 3 Detection of doubtful data based on entropy for rule selection

Detection of doubtful data based on entropy for rule selection is given by maximizing entropy

$$S(\mathbf{x}) = -\int p(\mathbf{y}|\mathbf{x}) \log p(\mathbf{y}|\mathbf{x}) d\mathbf{y} \tag{22}$$

under the following conditions.

$$\epsilon^2(\mathbf{x}, \mathbf{y}) = \int \|\mathbf{y} - \phi(\mathbf{x})\|^2 p(\mathbf{y}|\mathbf{x}) d\mathbf{y} \tag{23}$$

is constant and

$$\int p(\mathbf{y}|\mathbf{x}) d\mathbf{y} = 1. \tag{24}$$

Optimal solution is given by

$$p(\mathbf{y}|\mathbf{x}) = Z_\beta^{-1} \exp^{-\beta\|\mathbf{y}-\phi(\mathbf{x})\|^2}, \tag{25}$$

where

$$Z_\beta(\mathbf{x}) = \int \exp^{-\beta\|\mathbf{y}-\phi(\mathbf{x})\|^2} d\mathbf{y}. \tag{26}$$

Since maximizing entropy is equivalent to minimizing free energy

$$S_\beta(\mathbf{x}) = -F_\beta(\mathbf{x}) + \beta\epsilon^2(\mathbf{x}), \tag{27}$$

where

$$F_\beta(\mathbf{x}) = -\frac{1}{\beta} \log Z_\beta. \tag{28}$$

Optimal solution which satisfies

$$\frac{\partial F_\beta(\mathbf{x})}{\partial \mathbf{y}} = 0 \tag{29}$$

becomes also the conditional expectation.

$$\hat{\mathbf{y}} = \phi(\mathbf{x}) = \int \mathbf{y} p(\mathbf{y}|\mathbf{x}) d\mathbf{y}. \tag{30}$$

Melting[3] for detection of doubtful data is executed by applying steepest descent method.

$$\hat{\mathbf{y}} \Leftarrow \hat{\mathbf{y}} - \int \|\mathbf{y} - \phi(\mathbf{x})\| p(\mathbf{y}|\mathbf{x}) d\mathbf{y}. \tag{31}$$

In this study we will propose framework of data mining based on statistical modeling. From theoretical view point, these problems should be solved in unified way. The conditional expectation works important role.

$$\hat{\mathbf{y}} = \int \mathbf{y} p(\mathbf{y}|\mathbf{x}) d\mathbf{y} \qquad \text{Eqs. (5), (12) and (30).} \tag{32}$$

5. Conclusion

We proposed application of Gaussian mixture to nonlinear method II of quantification for extracting information from data as much as possible, and to melting for pending decision for rare data and doubtful data. We can apply dominance-based rough approximation to rule selection for excepting contradicting rules after removing rare data and doubtful data. Established framework of data mining can unify and deal with nonlinear method II of quantification and melting through Gaussian mixture for extracting rules.

References

[1] N. Otsu. (1975). Nonlinear discriminant analysis as a natural extension of the linear case. Behaviormetrika, 2, 45-59.

[2] R. L. Streit and T. E. Luginbuhl. (1994). Maximum likelihood training of probabilistic neural networks. IEEE Trans. NN, 5, 3, 764-783.

[3] Y. Wong. (1993). Clustering data by melting. Neural Computation, 5, 89-105.

[4] S. Greco, and R. Slowinski. (1999). Rough approximation of a preference relation by dominance relations. European Journal of Operational Research, 117, 63-83.

[5] N. Mori. (2001). Rough Set and Kansei Engineering. Jornal of Japan Society for Fuzzy Theory and Systems, 13, 6, 52-59.

[6] C. E. Osgood, G. J. Suci and P. H. Tannenbaum. (1957). The measurement of meaning. University of Illilnois Press.

[7] K. Sugihara, H. Ishii and H. Tanaka. (2003). New Approach to Conjoint Analysis Based on Rough Sets. Jornal of Japan Society for Fuzzy Theory and Intelligent Informatics, 15, 4, 59-65.

[8] K. Okuhara, H. Ishii and M. Uchida. (2003). Support of Decision Making by Data Mining Using Neural System. The Institute of Electronics, Information and Communication Engineers, J86-DII, 4, 535-542.

[9] A. P. Dempster, N. M. Laird and D. B. Rubin. (1977). Maximum-likelihood from incomplete data via the EM algorithm. J. Royal Statist. Soc. Ser. B, 39, 1-38.

[10] K. Okuhara, Y. Matsubara, K. Sugihara and H. Tanaka. (2004). Rule Selection by Rough Set Considering Ordinality in Attributes for Kansei Evaluation. The Institute of Electronics, Information and Communication Engineers, J87-A, 7, 1045-1053.