International Journal of Fuzzy Logic and Intelligent Systems, vol. 6, no. 4, December 2006, pp. 277-281

# Analysis of Client Propensity in Cyber Counseling Using Bayesian Variable Selection

**Su-Young Pi**

**Dept. of Practical Computer, Catholic University of Daegu**

## Abstract

Cyber counseling, one of the most compatible type of consultation for the information society, enables people to reveal their mental agonies and private problems anonymously, since it does not require face-to-face interview between a counsellor and a client. However, there are few cyber counseling centers which provide high quality and trustworthy service, although the number of cyber counseling center has highly increased. Therefore, this paper is intended to enable an appropriate consultation for each client by analyzing client propensity using Bayesian variable selection. Bayesian variable selection is superior to stepwise regression analysis method in finding out a regression model. Stepwise regression analysis method, which has been generally used to analyze individual propensity in linear regression model, is not efficient since it is hard to select a proper model for its own defects. In this paper, based on the case database of current cyber counseling centers in the web, we will analyze clients' propensities using Bayesian variable selection to enable individually target counseling and to activate cyber counseling programs.

**Key words** : Bayesian Variable Selection, Stepwise Regression, Target Cyber Counseling

## 1. Introduction

Cyber counseling is a type of consultation which is conducted in the cyber space to provide solutions to each client's problems and help necessary emotional, behavioral, and psychological development of clients who need assistances. Because of the characteristics of the cyber world, such as usefulness, anonymousness, patency, economic performance, rapidness, etc., cyber counseling has great possibility. Cyber counselling is one of the most compatible type of consultation for the information society.

It enables people to reveal their mental agonies and private problems anonymously, since it does not require face-to-face interview between a counsellor and a client. Cyber counseling has been generally known from 1980's when counselors in United States used computers and had made various academic and career counseling more popular. A consultation using computers was a great innovation in the counseling field and gave birth to the cyber counseling, settling down as a new key of counseling methods in 21st century.

Since clients' desires vary according to the pluralism of society, an emergence of new counseling method, cyber counseling, can be said to be already predicted. The number of cyber counseling institutions which operate cyber counseling centers providing individual and group counseling using a computer are increasing, While the demand of cyber counseling is predicted to be increase, the research on cyber counseling remain insufficient. There are no positive researches that can provide basic information to activate cyber counseling, such as clients' characteristics who care for cyber counseling, types of clients who want a certain kind of information, and how cyber counseling can effect clients' real lives.

In addition, in reality, the notion of cyber counseling is not generally known in spite of the increased number of counseling centers. Also, there are not enough active supports for the development and the diffusion of the cyber counseling and cyber counseling centers which provide high quality and trustworthy service. That is, counseling centers which clients can trust are not enough in spite of highly increased number of cyber counseling centers[1].

Stepwise regression analysis method, which is widely used to analyze individual propensities, is not suitable for selecting an appropriate model since it is hard to regulate variable selection and critical level of significance and cannot select variables properly if there is a high correlation coefficient among independent variable[2,3]. Bayesian variable selection finds subset from Bayesian mixture model when it selects variables. Bayesian variable selection is superior to stepwise regression analysis method in finding out a regression model.

Therefore, in this paper, based on the case database of current cyber counseling centers in the web, we will analyze clients' propensities using Bayesian variable selection to enable individually suitable counselings and to activate cyber counseling programs.

Manuscript received Oct. 20, 2006; revised Dec. 15, 2006.

## 2. Relate Research

### 2.1 The current situation of the cyber counseling

The cyber counseling has formed a new space of understanding and the information exchange by combining computer and communication techniques. Cyber counseling means 'real time or non-real time counseling in the cyber space carried out by a counselor spacially separated from a client to expedite a solution to a problem or the client's growth and development. Cyber counseling is a new way of communication in cyber era, which overcomes some limitations that previous ways of counseling had. Without limitations of the time and the space, cyber counseling provides a positive and active form of counseling and contributes to popularization of counseling.

The cyber counseling is composed of various types. Two major types of cyber counseling are realtime and non-realtime cyber counseling. Realtime cyber counseling is realtime meeting through a computer; for instance, video counseling and chatting counseling. Non-realtime counseling means that the counselor and the client meet each other through e-mail, bulletin board, a slip of paper counseling, or data base counseling. The counseling can be achieved although the client and the counselor do not exist at the same place at the same time.

There are several merits in cyber counseling. First, it overcomes time and space restrictions, enabling clients to get an advice in convenient time and place. Clients can save their time and money because they do not need to visit a center or wait for their turns. Second, counsels can be done objectively since clients do not need to disclose their identities. Third, clients and counsellors can deal with sexual, family, and immoral problems because anonymity makes people be open and straight. Forth, clients can decide whether to start or finish the counseling, and whether to disclose information or not. Because of these merits, cyber counseling center has been increasing.

However, there are also some demerits. First, there should be the right technology, equipment, and proper system. Second, there rises a reliability problem, a problem relates to identifications of a counselor and a client. Third, it lacks dynamic and immediate interactions since cyber counsels are fully dependent upon written language. Forth, detailed problems and information are frequently ignored, since it takes more time than face-to-face counseling.

Actually, most of cyber counseling is not operated as a real-time. It is classified into a public counseling and a private counseling. A counsellor or other clients give advice to a client when a client writes his problems on a on-line board in public counseling. That is, public counseling cannot maintain continuous counsels since there is no information about the client who posted problems. The relation between the counselor and client is temporary and superficial. And the possibility of the single-session counseling is high and the counseling is ended easily.

So, we can not achieve the efficiency of the counseling. At least a counselor and a client should trust each other to maintain continuous counsels.

On the other hand, a private counseling guarantee a privacy of counsels. Therefore, Secret counseling can access to a problem more deeply and continuously than open counseling, and also has the advantage of counseling with sufficient time. Secret counseling should be recommended for those clients who want in-depth counseling. Clients do not fully disclose their problems when there is a possibility to be known to the other people. Therefore, private counseling is far more efficient than public counseling in relationship between a counselor and a client and the effect of counsels.

Cyber counsels such as E-mail counseling, counseling using on-line board, providing counsel cases, and internet chatting are generally provided via internet in Korea. The central operating body of counsels are private organizations (52.0%), the country (21%), cities and provinces (15%), universities (8%), religious organizations (3%), and corporations (2%). They operates using unilateral DB (90%), on-line boards (87%), E-mails (79%), interactive DB (25%), and chatting (9%)[1].

### 2.2 The Problems in the operation of cyber counseling

Current cyber counseling is mainly operating based on on-line boards and E-mails, not providing good programs which are widely acknowledged. Most of counseling programs provide only indirect contacts since they are based on on-line boards and E-mails. Counsels which is on a board only gives other clients a impression that there is another person who has a same problem with them, providing a mere feeling of relief. Counseling programs with higher quality are necessary to satisfy clients.

Cyber counseling can be easily discontinued because there is no strong relationship between a counselor and a client than in a face-to-face counseling. There should be proper predetermined counsel frequency, durations, methods, and types. The more information that a counsellor has, the more efficient a counsel to be. Most of counsels are end up with an instant event, since a counsellor does not have information such as why a client want advices, what a client is like, have a client ever taken any kind of advice and what kind of advice a client wants.

Most of cyber counseling centers do not care about whether clients are satisfied or not, finishing a case by sending them an E-mail with possible settlements. Counseling centers should find out how clients feel about result of cyber counseling and should manage clients after the counsels to increase an efficiency.

Counseling centers can check conditions of former clients by sending E-mails in a certain amount of time, and offer another chance of counsels if it is needed. Also counseling centers should provide information that is comply to clients' needs such

as counseling education, special lectures, and books, in order to keep continuous relationships with clients.

## 3. Analyzing client propensity in cyber counseling

### 3.1 Bayesian Variable Selection

Selecting variables is one of the most difficult problems in linear regression analysis. It is the most important to decide an independent variable which represents dependent variable most effectively among many independent variables which explain dependent variables. Most of the researchers and people in charge have been using stepwise regression analysis method in selecting am independent variable. However, it is hard to regulate variable selection and critical level of significance when using stepwise regression analysis method. Also, it cannot select the proper variable when there is a high correlation coefficient among variables.

All subsets regression, which is presented as an alternative plan for stepwise regression analysis method, selects the most appropriate model among all the possible regression models by predetermined criteria. It is known to have higher reliance. However, the use of all subsets regression is restrictive since the amount of calculating increases in geometrical progression as the number of independent variables increases.

Therefore, in this paper, we used Bayesian variable selection as an alternative to widely used stepwise regression analysis. Bayesian variable selection finds subset from Bayesian mixture model when it selects variables. Bayesian variable selection is superior to stepwise regression analysis method in finding out a regression model[4,5].

Therefore, based on the case database of current cyber counseling centers in the web, we will analyze clients' propensities using Bayesian variable selection to provide fine service to the client in order to make active and continuous counseling programs.

Linear regression analysis model which include all the independent variables can be represented as a simple determinant.

$$Y = X\beta + \varepsilon \tag{1}$$

In this expression, Y is a vector of dependent variable $n\times 1$ and $\varepsilon$ is the error vector of $n\times 1$ which follows $N(0, \sigma^2 I)$. X is a matrix of $n\times(p+1)$.

The first row is about vectors which consists of 1, and p row represent the observation value of independent variables. $\beta=(\beta_0, \cdots, \beta_p)'$ is parameter vector and $\beta_0$ is the parameter to estimate an intercept of regression model. Expression (1) is about selecting a group of independent variables which explain dependent variable Y. Namely, it determines whether $\beta_i$, each element of parameter vector $\beta$, is 0 or not.

Selecting a independent variable which explains dependent variable Y is identical with finding out $\beta_i$ which is not 0, from parameter vector $\beta$. It is defined as $\gamma=(\gamma_0, \cdots, \gamma_p)'$, vector of $(p+1)\times 1$ that satisfies $\gamma_i=0$ if $\beta_i=0$ and $\gamma_i=1$ if $\beta_i\neq 0$.

According to Bayes theorem, prior probability distribution $\pi(\beta_\gamma | \sigma^2, \gamma)$ about joint probability distribution of parameter $\beta_\gamma$, $\sigma$, and $\gamma$ is identical with product of three conditional probabilities $\pi(\beta_\gamma | \sigma^2, \gamma)\pi(\sigma^2|\gamma)\pi(\gamma)$. Hence, prior probability distribution of $\beta_\gamma$ can be represented as expression (2) and posterior probability distribution can be deduced from expression (3).

$$\pi(\beta_\gamma | \sigma^2, \gamma) \sim N(0, c\sigma^2 (X_\gamma' X_\gamma)^{-1}) \tag{2}$$

In this expression, a value of $c$ is a value of random constant and prior probability distribution of $\sigma^2$ follows $\pi(\sigma^2 | \gamma) \propto 1/\sigma^2$, according to the theory of Jeffrey.

$$\begin{aligned} p(\beta_\gamma, \sigma^2, \gamma | Y) &\propto p(Y | \beta_\gamma, \sigma^2, \gamma)\pi(\beta_\gamma, \sigma^2, \gamma) \\ &= p(Y | \beta_\gamma, \sigma^2, \gamma)\pi(\beta_\gamma | \sigma^2, \gamma)\pi(\sigma^2 | \gamma)\pi(\gamma) \end{aligned} \tag{3}$$

From expression (3) above, we can present $p(\gamma|Y)$, posterior probability distribution of $\gamma$, which is the key of selecting variables as expression (4).

$$\begin{aligned} p(\gamma | Y) &\propto p(Y | \gamma)\pi(\gamma) \\ &= [\int_{\sigma_2} [\int_{\beta_\gamma} p(Y | (\beta_\gamma, \sigma^2, \gamma)\pi(\beta_\gamma | \sigma^2, \gamma) d\beta_\gamma] \\ &\quad \pi(\sigma^2 | \gamma) d\sigma^2] \pi(\gamma) \propto (1+c)^{-qr/2} s(\gamma)^{-n/2} \pi(\gamma) \\ &= (1+c)^{-qr/2} s(\gamma)^{-n/2} 2^{-(p+1)} \end{aligned} \tag{4}$$

In this, $q_\gamma = \sum_{i=0}^{p} \gamma_i$

$$s(\gamma) = Y'Y - \frac{c}{1+c} Y' X_\gamma (X_\gamma' X_\gamma)^{-1} X_\gamma' Y$$

We can estimate parameter $\gamma$ from expression (4). That is, one of the method of estimation can be determining which independent variable to choose from $\gamma_{mode}$, the mode of $\gamma$. It is impossible to deduce posterior possibility distribution of $\gamma$ when the value of independent variable is high. Therefore, if dimensions of $\gamma$ is high, the process of estimating parameter $\gamma$ from the distribution in expression (4) can be divided into 2 steps.

The first step is to select initial value of $\gamma$, $\gamma^{[0]} = \gamma_1^{[0]}, \cdots, \gamma_p^{[0]}$. We select each value of $\gamma_i^{[0]}$ randomly from binominal distribution, $p(\gamma_i = 1) = p(\gamma_i = 0) = 1/2$. The second step is to sample each value of $\gamma_i (i=0, 1, \cdots, p)$ from conditional probability distribution $p(\gamma_i | Y, \gamma_{j\neq i})$, randomly. Based on the expression (4), conditional probability

distribution $p(\gamma_i | Y, \gamma_{j\neq i})$ can be represented as an expression (5).

$$p(\gamma_i | Y, \gamma_{j\neq i}) \propto p(Y|\gamma)\pi(\gamma_i) \propto (1+c)^{-qr/2} s(\gamma)^{-n/2} 2^{-1} \quad (5)$$

### 3.2 Research objects

We analyzed client propensity based on a database of a current cyber counseling center that is operation present. 600 people were sampled from consultation cases and 415 people who satisfied with consultations to some degree were processed as objects. Membership function of triangle form was utilized for the client's preference towards counsellors.

Four evaluation items were used for counselor evaluation, and membership function was applied to each of them for the counselor evaluation. The evaluation results are expressed as follows:

$$g = w_1 \times q_1 + w_2 \times q_2 + \cdots + w_k \times q_k = \sum_{i=1}^{k} w_i \bullet q_i \quad (6)$$

Here, $g$ means the result of the total evaluation value for four evaluation items for which clients evaluated. $w_i$ means the weight for each item, $q_i$ means the evaluation for each item. The different weight for each item and the introduction of optimism index $\lambda$ make flexible counselor evaluation possible, since they allow different values added according to the client's total satisfaction value[6]. The evaluation results are expressed as follows:

$$fg = \lambda g_1^{(\alpha)} + (1-\lambda) g_3^{(\alpha)} \quad (7)$$

We sampled some variables that effect clients' preference towards counsellors and analyzed to what propensities preferences are varies. We compared and evaluated 415 cases with stepwise regression analysis method and Bayesian variable selection. First of all, we divided 415 case data into 300 and 115 cases, and estimated parameters using stepwise regression analysis method and variable selection and tested a prediction of model to 115 cases with established model. We applied 3 models from 115 samples that is not used in estimation, in order to overcome adversities of evaluating within samples used in estimation.

The analysis is based on clients' basic personal information from the clients' profile database in a counsel database and the counsel cases in which clients satisfied with the result to some degree. We selected preference variables towards counsellors from the counsel database, and variables such as gender(x1), age(x2), academic background(x3), regional differences(x4), blood type(x5), income(x6), the number of visit(x7), use of internet(x8), and previous experience in counseling(x9) from client profile database. Table 1 shows the result of stepwise regression analysis with 9 independent variables with setting up variable accept significance and variable reject significance as 0.05.

Table 1. Stepwise regression analysis

| parameter | estimator (standard deviation) | p value |
|---|---|---|
| $\beta_1$ | 4.88(0.48) | 0.001 |
| $\beta_2$ | 5.22(0.50) | 0.000 |
| $\beta_3$ | 4.43(1.78) | 0.001 |
| $\beta_4$ | 4.61(1.46) | 0.002 |
| $\beta_5$ | 2.99(1.38) | 0.046 |
| $\beta_7$ | 3.45(1.20) | 0.030 |

6 parameters which is estimated in table 1 was statistically significant in the significance level $\alpha = 0.05$ . To make a comparative study, we applied Bayesian variable selection for the same data. Table 2 shows the 3 most frequent vectors which occur when we randomly sampled 1000 vector $\gamma$ from posterior probability distribution of $\gamma$ .

Table 2. Bayesian variable selection

| selected variables | Frequency |
|---|---|
| x1, x2, x3, x7 | 230 |
| x1, x2,x3 | 190 |
| x1, x2, x3, x4, x7 | 158 |

A model which has 4 variables is generated when selecting the 3 most frequent vectors as table 2 represents. That is, preference of counsellors differs according to the clients' propensities. Variables which effect clients' preference towards counsellors are gender, age, academic backgrounds, regional differences, blood type, and the number of visit in stepwise regression analysis. Also variables such as gender, age, academic backgrounds, and the number of visit effect the preference in Bayesian variable selection.

In order to overcome adversities of evaluating within samples used in estimation indirectly, we evaluated 115 samples that are not used in the estimation by using stepwise regression analysis and Bayesian variable selection. The result is in table 3. We used 2 measurement to evaluate the model's prediction and errors. One was mean squared error and the other was mean absolute error which are shown in expression (8).

$$MSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 / n$$

$$MAE = \sum_{i=1}^{n} |Y_i - \hat{Y}_i| / n \quad (8)$$

Table 3. Comparison in samples which is not used in estimation

| Model | Selected variables | MSE | MAE |
|---|---|---|---|
| all subsets regression analysis | x1, x2, x3, x4, x5, x6, x7, x8, x9 | 191.7 | 9.1 |
| Stepwise regression analysis | x1, x2, x3, x4, x5, x7 | 194.5 | 9.8 |
| Bayesian variable selection | x1, x2, x3, x7 | 179.6 | 7.9 |

As table 3 represents, Bayesian variable selection is proven to be the best model and stepwise regression analysis the worst in both criteria. As a result of the processing, it is proven that clients' preference towards counsellors differs according to gender, age, academic backgrounds, and the number of visit. After studying clients' propensities with neural network, we can enhance efficiency of counseling programs by recommending a counselor which is suitable for a client.

## 4. Conclusions

Due to the rapid growth of the internet, off-line counselling is changing to the on-line cyber counseling. However, in spite of the increased number of counseling centers, there are not enough cyber counseling centers which provide good and suitable services for each client's propensities. Most of the researchers and people in charge have been using stepwise regression analysis method when analyzing individual propensities. However, it is hard to regulate actual variable selection and critical level of significance when using stepwise regression analysis method. Also, it cannot select the proper variable when there is a high correlation coefficient among variables.

Therefore, this paper used Bayesian variable selection as an alternative to widely used stepwise regression analysis. Bayesian variable selection finds subsets from Bayesian mixture model when it selects variables and is superior to stepwise regression analysis method in finding out a regression model.

We analyzed client propensity based on a database of currently operating cyber counseling centers. As a result of using both stepwise regression model and Bayesian variable selection, it turned out that the model selected by Bayesian variable selection is more reliable.

This paper intended to find out the best model to analyze client propensity precisely, enabling an individually suitable counseling and to activate continuous cyber counseling by providing good services. Currently, cyber counseling centers give advice to clients and finish it, not interacting each other or receiving some feedbacks from clients. Studies on developing systematic programs that can offer further assistances according to a client's propensity should be conducted.

## References


[1] Im eun mi, kim ji eun, "*Cyber Counseling Operation Report*," Korea Institute Youth Counseling, 1999.

[2] Draper, N . and Smith, H, "*Applied Regression Analysis*", second edition. John Wiley, New York, 1981.

[3] Miller, A., "*selection of Subsets and Regression Vari*ables", Journal of Royal Statistical Society, A, 147, 389 - 429, 1984.

[4] Smith, M. and Kohn, R., "*Nonparameter Regression Using Bayesian Variable Selection*", Journal of Econometrics, 75. 317-343, 1996.

[5] Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.b. *Bayesian Data Analysis*, Chapman & Hall, 1995

[6] Pi Su Young, "Design of Target Cyber Counselling using Counseling Assistance Agent", *International Journal of Fuzzy Logical and Intelligent Systems*, Vol. 4, no. 3, pp. 311-315, 2004.

[7] David Heckerman, *A Tutorial on Learning Bayesian Networks*, Technical Report MSR-TR-95-06, 1995.

[8] Jensen, F. V., *Bayesian Networks and Decision Graphs*, Springer Verlag, 2001.

[9] Berk, K. ,*Comparing Subset Regression Procedures*, Technometrics, 20, 1, 1-6, 1978

[10] Fabio Gagliardi Cozman, "Generalizing Variable Elimination in Bayesian Networks", *Workshop on probabilistic Reasoning in Artificial Intelligence*, Atibaia, Brazil, 2000

[11] Choi suk yeong, Baek Hyun ki, "Design and Implementation of XML-based Cyber Counseling System Supporting Counseling Analysis Information," *Korean information education learned society*, Vol. 7, No. 3,, pp.341-352, 2003.



**Su Young Pi**

Su-Young Pi received the Ph.D. degrees in computer engineering Catholic University of Daegu 2000. She is an full-time lecturer at the department of Practical Computer in Catholic University of Daegu, in Korea since 2000. Her current research interests include intelligent agent, Bayesian network, fuzzy theory, neural network and application.

Phone : +82-53-850-3699