学術情報処理研究 No.1 1997 pp.51-54

On Entrance Examination Analysis

Akihiro Hashimoto and Hisao Miyano*

Information Processing Center, Chiba University
1-33 Yayoi, Inage, Chiba 263
hash@rrose.ipc.chiba-u.ac.jp

* Faculty of Letters, Chiba University
1-33 Yayoi, Inage, Chiba 263
miyano@cogsci.l.chiba-u.ac.jp

Abstract

We provide a method to determine an optimal composite score which can be used for the selection of applicants for admission to a university. Our method has an advantage that it can consider a realistic condition that any change of weight vector may exert unfavorable effects to univeristy education.

Introduction

Problems related to university entrance examination apparently are critical for every university. In fact, any change in entrance examination often exerts serious effects on university education. Hence, many endeavors to improve entrance examination have been cautiously administered from various points of view.

In our country, the selection of applicants for admission to university is based on a composite score of two examinations. One is a common first-stage examination called ``The National Center Test for University Admissions'' (First-stage Examination; FE), and the other is a second-stage examination (Second Examination; SE) administrated by each university. The composite score is normally a weighted sum of marked scores for these two examinations, and the weights for the composite score are determined by each university.

In addition to receiving the above two examinations, the applicants are requested to show their high-school report (HR). However, the high-school marks usually are not considered explicitly for the selection of applicants, perhaps because there exists no decisive method to adopt HR effectively.

We here propose a new method to determine a composite score which depends not only on FE and SE but also on HR. The method is based on the Convex Cone Method (Hashimoto, Miyano, & Taguri, 1997), and gives an optimal composite score in a sense that the correlation of the composite score and university records is maximized under given appropriate constraints.

The Convex Cone Method

The Convex Cone Method developed by Hashimoto et al. (1997) gives an exact solution of the following optimization problem.

(**Problem 1**). Find $\mathbf{x} \in R^q$ which maximizes

$$R(\mathbf{x}) = \mathbf{x}'\mathbf{b}/(\|\mathbf{x}\| \cdot \|\mathbf{b}\|), \qquad (1)$$

subject to the constraint

$$(\mathbf{x}-\mathbf{a})'\mathbf{S}(\mathbf{x}-\mathbf{a}) \le r^2, \qquad (2)$$

where $\mathbf{S} \in R^{q\times q}$ is a symmetric positive definite matrix, $\mathbf{a}, \mathbf{b} \in R^q$ are nonzero vectors and $\|\cdot\|$ is a Euclidean norm.

In case of entrance examination, $\mathbf{x}$ corresponds to unknown weight vector of which elements are the weights for a composite score, and $\mathbf{b}$ is the vector corresponding to average university records. And the constraint means that the optimal weight should not be much different from corresponding current weight vector $\mathbf{a}$.

Formulation of our problem

Let $\mathbf{y} \in R^n$ be a given vector of which elements denote the average university records of selected applicants where *n* is the number of the selected, and $\mathbf{X}_0 \in R^{n\times p_0}, \mathbf{X}_1 \in R^{n\times p_1}$ and $\mathbf{X}_2 \in R^{n\times p_2}$ be matrices whose elements denote the scores for HR, FE and SE respectively, where p_0, p_1 and p_2 are the numbers of subjects included in HR, FE and SE. For each of these matrices, we introduce unknown weight vectors $\mathbf{w}_i \in R^{p_i}$ for $\mathbf{X}_i$, *i=0,1,2,* to define a composite score $\mathbf{z}$ by

$$\mathbf{z} = \sum_{i=0}^{2} \mathbf{X}_i\mathbf{w}_i = \mathbf{X}\mathbf{w} \in R^n, \qquad (3)$$

where $\mathbf{X} = (\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2) \in R^{n\times(p_0+p_1+p_2)}$, and $\mathbf{w} = (\mathbf{w}_0', \mathbf{w}_1', \mathbf{w}_2')' \in R^{p_0+p_1+p_2}$.

For the evaluation of weight vectors, we consider a correlation coefficient between $\mathbf{y}$ and $\mathbf{z}$;

$$R(\mathbf{z}) = \frac{\mathbf{y}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{z}}{\sqrt{\mathbf{y}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{y}}\sqrt{\mathbf{z}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{z}}}, \qquad (4)$$

where $\mathbf{Q}_n = \mathbf{1}_n\mathbf{1}_n'/n \in R^{n\times n}$ and $\mathbf{1}_n = (1, ..., 1)' \in R^n$.

The correlation coefficient may be more appropriate for the evaluation of weight vectors than a squared distance between $\mathbf{y}$ and $\mathbf{z}$, since subjects administered in entrance examination have not always equal full marks and the squared distance is affected by the different choice of full marks.

Considering the normalization of the weight vectors and the condition that weight vector much different from a current weight vector may be unfavorable, we assume the following two constraints;

$$\begin{aligned} \mathbf{1}_p'\mathbf{w} &= 1, & (5)\\ (\mathbf{w}-\mathbf{w}_0)'\mathbf{D}(\mathbf{w}-\mathbf{w}_0) &\le r^2, & (6) \end{aligned}$$

where $p = p_0 + p_1 + p_2, \mathbf{1}_p = (1, ..., 1)' \in R^p, \mathbf{w}_0 \in R^p$ is a current weight vector, and $\mathbf{D} \in R^{p\times p}$ is a symmetric positive definite matrix. In addition to these two constraints, without loss of generality we can assume

$$\begin{aligned} \mathbf{1}_p'\mathbf{w}_0 &= 1, & (7)\\ \mathrm{rank}(\mathbf{X}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{X}) &= p, & (8)\\ \mathbf{y}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{y} &\neq 0, & (9)\\ \mathbf{y}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{X} &\neq \mathbf{0}, & (10)\\ n &> p. & (11) \end{aligned}$$

Eventually our problem can be formulated as a optimization problem under nonlinear constraints:

(**Problem 0**). Find $\mathbf{w} \in R^p$ which maximizes $|R(\mathbf{w})|$,

$$R(\mathbf{w}) = \frac{\mathbf{y}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{X}\mathbf{w}}{\sqrt{\mathbf{y}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{y}}\sqrt{\mathbf{w}'\mathbf{X}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{X}\mathbf{w}}}, \qquad (12)$$

subject to the constraints,

$$\begin{aligned} \mathbf{1}_p'\mathbf{w} &= 1, & (13)\\ (\mathbf{w}-\mathbf{w}_0)'\mathbf{D}(\mathbf{w}-\mathbf{w}_0) &\le r^2. & (14) \end{aligned}$$

Numerical solution and an example

Numerical solution of our problem, **problem 0**, can be obtained by using the Convex Cone Method (Hashimoto et al. 1997). Of course the Convex Cone Method can not be applied to our current problem directly, since our problem includes the constraint $\mathbf{1}_p'\mathbf{w} = 1$ in addition to a quadratic constraint. After tedious considerations on several cases defined by the given constraints, we however can assert that our problem can be numerically solved by the Convex Cone Method.

We here give a numerical example to illustrate our method. Analyzed data were artificial but partly borrowed from real entrance examination data. The following shows a summary of the analyzed data: *n=100,* $p_0 = p_1 = p_2 = 1, \mathbf{D} = \mathbf{I}_3, \mathbf{w}_0' = (0, 1/2, 1/2), \sqrt{\mathbf{y}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{y}} = 103.64,$

$$\begin{aligned} \mathbf{X}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{X} &= \begin{pmatrix} 1740.44 & 1078.91 & 1116.02 \\ 1078.91 & 17454.43 & 355.30 \\ 1116.02 & 355.30 & 28282.89 \end{pmatrix}, & (15)\\ \mathbf{y}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{X} &= \begin{pmatrix} 1536.61 & 894.00 & 2763.97 \end{pmatrix}. & (16) \end{aligned}$$

Figure 1 shows relative changes of $R(\mathbf{w}^*)$ with different r^2 where $\mathbf{w}^*$ is a solution of **problem 0**; that is, in order to compare the optimal weight $\mathbf{w}^*$ with the current weight $\mathbf{w}_0$, we calculated the efficiency ratio defined by

$$\mathrm{Eff} = \frac{R(\mathbf{w}^*)}{R(\mathbf{w}_0)}. \quad (17)$$

Note that $R(\mathbf{w}_0) = 0.16$ and $R(\mathbf{w}_0) = 0.16$.

Concluding remarks

In this paper, we proposed a method to determine an optimal composite score which can be used for the selection of applicants for admission to university. Our method has an advantage that it can consider a realistic condition that any drastic change of weight vector may exert unfavorable effects to university education.

While we may say that our method is effective on the whole, there still remains some problems to be solved. The most important one is a problem related to the estimation of $\mathbf{y}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{y}, \mathbf{X}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{X}$ and $\mathbf{y}'(\mathbf{I}-\mathbf{Q}_n)\mathbf{X}$. These values should be estimated from the data of selected applicants. We however have no decisive method to estimate these values.

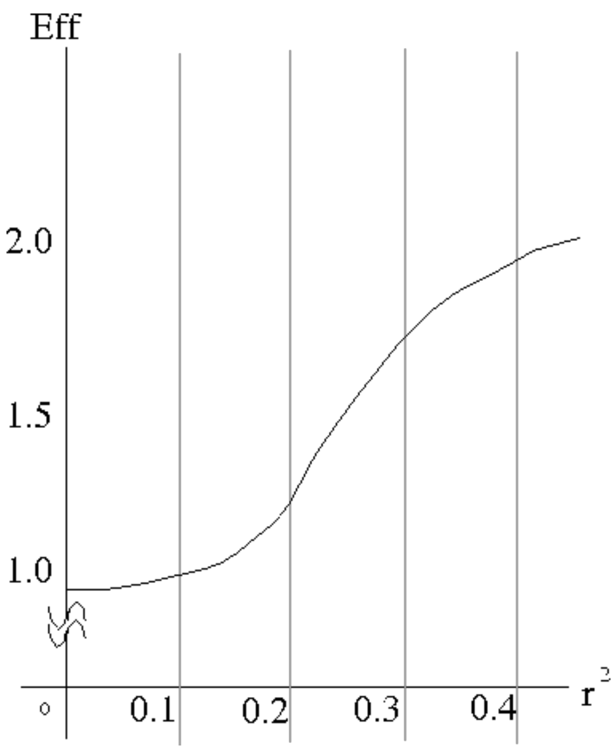

Figure 1: Relative efficiency Ratio

References

[1]

Bock, M. E. (1982). Employing vague inequality information in the estimation of normal mean vectors (Estimators that shrink to closed convex polyhedra). In S. S. Gupta & J. O. Berger (Eds.), *Statistical Decision Theory and Related Topics-III* (Vol.1, pp. 169-193). Academic Press: New York.

[2]

Dengupta, D., & Sen, P. K. (1991). Shrinkage estimation in restricted parameter space. *Sankhya, Ser. A, Vol. 53,* pp. 389-411.

[3]

Frommer, A., & Mayer, G. (1990). On the r-order of Newton-like methods for enclosing solutions of nonlinear equations. *SIAM Journal on Numerical Analysis, Vol. 27, No. 1*.

[4]

Golub, G. H., & Van Loan, C. F. (1989). *Matrix Computation, 2nd Edition.* John Hopkins Univ. Press: London.

[5]

Hashimoto, A., Miyano, H., & Taguri, M. (1997). Maximization of Correlation Coefficient under Quadratic Constraint. In the 6th China & Japan S. S. (to appear)

[6]

Schneider, R. (1993). *Convex Bodies: The Brown-Minkowski Theory.* Cambridge Univ. Press: Cambridge.

[7]

Takane, Y., Kiers, H. A., & DeLeeuw, J. (1995). Component analysis with different set of constraints on different dimensions. *Psychometrika, Vol. 60, No. 2*, pp. 259-280.

[8]

Yano, K., Ouchi, S. and Taguri, M. (1990). Daigaku nyushi ni okeru senbatsu bairitsu ni tuiteno kentou (Investigation on the multiplier of preliminary selection in entrance examination to universities). The Japanese Journal of Behaviormetrics, Vol. 17, pp. 25-33.