How Students' Attitudes of Career Design Affect their Academic Achievements

How Students' Attitudes of Career Design Affect their Academic Achievements

How Students' Attitudes of Career Design Affect their Academic Achievements*

○ Zijie Zhang Hidetsugu Suto (Muroran Institute of Technology)

Abstract Muroran Institute of Technology has been conducting questionnaire to ask undergraduates about their attitudes of career design. In this paper, in order to visualize how students' attitudes of career design affect their academic achievements, a new method which combined modified principal component analysis and support vector machines has been proposed.

Keyword: principal component analysis, changing in attitudes, hexagonal binning, support vector machines

1 Introduction

Investigation on career education, concerning student career development as a central theme, that has been a high-priority item throughout the world. The concept of career can be defined briefly as encompassing the educational and vocational strands of life – the unfolding series of educational and vocational experiences and decisions related to a person's productive role in society [1]. In 2012, MEXT (Ministry of Education, Culture, Sports, Science and Technology in Japan) has investigated employment situation of Japan's new graduates [2]. Result of the investigation shows that there were more than 550,000 new graduates, but 22.9% of them have failed to find a stable job. In order to deal with the situation, MEXT has launched personnel training project corresponded to the industry requirements in 2012 [3]. As a part of the project, Muroran Institute of Technology (MuroranIT) and all other 16 universities have been conducting investigations on students' attitudes of career design, and the investigation is introduced in detail in Section 2.1. Concerning the current situation, only usual statistical methods have been used to analyze the data. Thus it is necessary to apply professional data analysis to discover valuable information.

Studies related to people's career design have started from an early stage. In 1972, job characteristics and job dimensions as based on the position analysis questionnaire [4] has been analyzed. Due to principal component analysis (PCA) is a powerful tool capable of reducing dimensions and revealing relationships among data items [5], it is applied for data analysis on questionnaire investigations [6][7]. Meanwhile, we cannot find any cases in which PCA was applied for analyzing dynamically changing data.

For observing distribution of PCA results, hexagonal binning method is employed in the study of this paper. Hexagonal binning plotting used as a high-performance interaction with scatterplot matrices is a powerful approach to exploratory multivariate data analysis. Topics include density representation by gray scale or by symbol area, alternatives to brushing, and animation sequences have been addressed [8]. In 1997, there were researchers who modified hexagonal binning plotting for variable resolution as distribution of points are displayed in two dimensions [9].

In order to find the relationship between students' attitudes of career design and their academic achievements, a new method which combined modified PCA and SVMs has been proposed. As a classification method, support vector machine (SVM) is a global classification model that generates non-overlapping partitions and usually employs all attributes [10].

2 Investigation on Students' Attitudes

2.1 Questionnaire

In the questionnaires mentioned in Section 1, student attitudes towards learning, campus life, and his/her point of view for career, have been investigated. The participants were more than 1,800 undergraduate students from all courses of MuroranIT, and the data was sampled by taking questionnaire once in a year.

The options for each description are set as totally disagree (score as 1), rather disagree (score as 2), agree to some extent (score as 3), and totally agree (score as 4). In the case studies shown in this paper, students who entered the university from 2010 have been chosen as samples, and data collected from a series of three continuous years in 2011 (grade 2), 2012 (grade 3), and 2013 (grade 4) and is used for the analysis.

Table 1: A part of the obtained data.

student ID	Grade	Q1	Q2	Q3	Q4	...	Q43
10024011	2	4	4	4	1	...	4
10024097	2	2	3	2	1	...	4
10024011	3	3	3	4	1	...	4
10024097	3	3	4	4	1	...	1
10024011	4	3	4	4	1	...	4
10024097	4	4	4	3	1	...	3

2.2 Analysis with a Traditional Method

In the analysis shown in this section, the first two dominant principal components are used to project data items onto a two-dimensional coordinate plot. The results of analysis with a traditional method includes data of loadings and scores which is analyzed by different grades. The horizontal axis represents the first principal component, and the vertical axis

*This study has been presented in SICE2015

represents the second principal component. Loadings which are greater than or equal to 0.2 and less than or equal to -0.2 are selected for supporting the interpretation of principal components.

Scores of samples obtained by conducting PCA in the traditional way are plotted in Fig. 1 (A). Color of each sample in the plot indicates the students' grade point average (GPA) data which is ranged from 0.0 to 4.0, and the color turns darker when the sample has a higher GPA. Since groups of samples are different, it is difficult to observe changing in attitudes and discover the variation of distribution for samples with different GPA through a certain period. A new way of applying PCA and an appropriate plotting method are required to fulfil the aim.

2.3 Analysis with the Proposed Method

For observing changing in students' attitudes, a new method has been proposed as shown below.

Step 1. Apply PCA on all sample data

All sample data is analyzed by PCA instead of applying PCA with separated data sets. PCs of overall analysis results are decided for plotting.

In step 1, PCA is applied on all sample data without separated into different grades and the loadings and scores are acquired. The principal components used for projecting data items are interpreted and represented in Table 2 by referring descriptions in questionnaire.

Step 2. Divide samples' scores data by groups

Due to plotting method for observing distribution is necessary for discovering distribution variation of samples with different GPA, scores data of samples obtained from PCA is divided into different groups of data sets for being projected into different plots respectively. Thus, in all plots, the same horizontal axes and vertical axes are used.

In step 2, scores data of all samples is divided into grade 2, grade 3, and grade 4. Each data is shown as graphs in Fig. 1 (B). It is not clear that how samples distribute in the two-coordinate space due to overlapping of the points.

Step 3. Apply hexagonal binning plotting method

In order to show distribution more clearly, scores data obtained from the above process is visualized by using hexagonal binning plotting.

At this step, it is required to decide range of each axis. For making the projecting space square, end points of horizontal and vertical axes are set with the same absolute value in plus and minus direction. Scales of axes should be set wider than the end points to include all points and for leaving space at the margin. In order to decide size of a hexagonal cell, the number of cells in a horizontal line is determined by increasing cells' number by 5 starting from 0 for partitioning the horizontal space until the convergence situations can be clearly compared. For comparing the overall convergence situation, the smallest value of the greatest points' number in a cell among the plots is chosen as an index. The value of minimum points' number in a cell among the plots is chosen as another index. In the case study of this paper, the samples' scores obtained by applying PCA in the proposed process are utilized for hexagonal binning plotting.

In step 3, hexagonal binning plotting method is used to support finding information from data distribution. In the plots of Fig. 2, horizontal axis stands for PC1, and vertical axis stands for PC2. Horizontal axis and vertical axis are set with value of range from -15 to 15. The end points differential of horizontal axis is 30 which becomes an appropriate number of cells for partitioning the horizontal region. To compare distribution of scores data of samples with different GPA, scores data of top 100 GPA ranking samples and bottom 100 GPA ranking samples of each grade is plotted in graphs as shown in Fig. 2. The greatest samples' number in a cell of all hexagonal cells in Fig. 2 is greater than 4, except for the plot of grade 3 bottom GPA samples. Hence 4 becomes a index for comparing samples' convergence situation. The minimum number of samples in a cell is 1 that is taken as another index.

Step 4. Apply SVMs

In order to analyze and visualize the changing in attitudes of students with different academic achievements, SVMs have been applied to scores data of samples. Regularization parameter decides the trade off between the margin and the number of mistakes on the training data. Regularization parameter is represented by c.

In step 4, SVMs are applied to the samples' scores data obtained in step 2 to find difference of changing in attitudes between the top GPA students and the bottom GPA samples. Each classification is conducted with small and large parameter of c. Small c allows constraints to be easily ignored (large margin), and large c makes constraints hard to ignore (narrow margin).

3 Results and Discussion

To show the differences of distributions in plots of Fig. 2 among the groups, number of hexagonal cells of each plot is shown in Table 3. It is shown that there are almost the same distribution for both the bottom GPA samples and the top GPA samples during grade 2. From the plot of grade 3, we can see that the top GPA samples starts to converge. In the plot of grade 4, the converging trend of the top GPA samples gets stronger than ever comparing with all other distributions. The overall result shows that distribution of the bottom GPA samples gets splitting as the grades grow, but contrarily distribution of the top GPA samples gathers and forms larger groups. It is also indicated that the grouping trends with similar

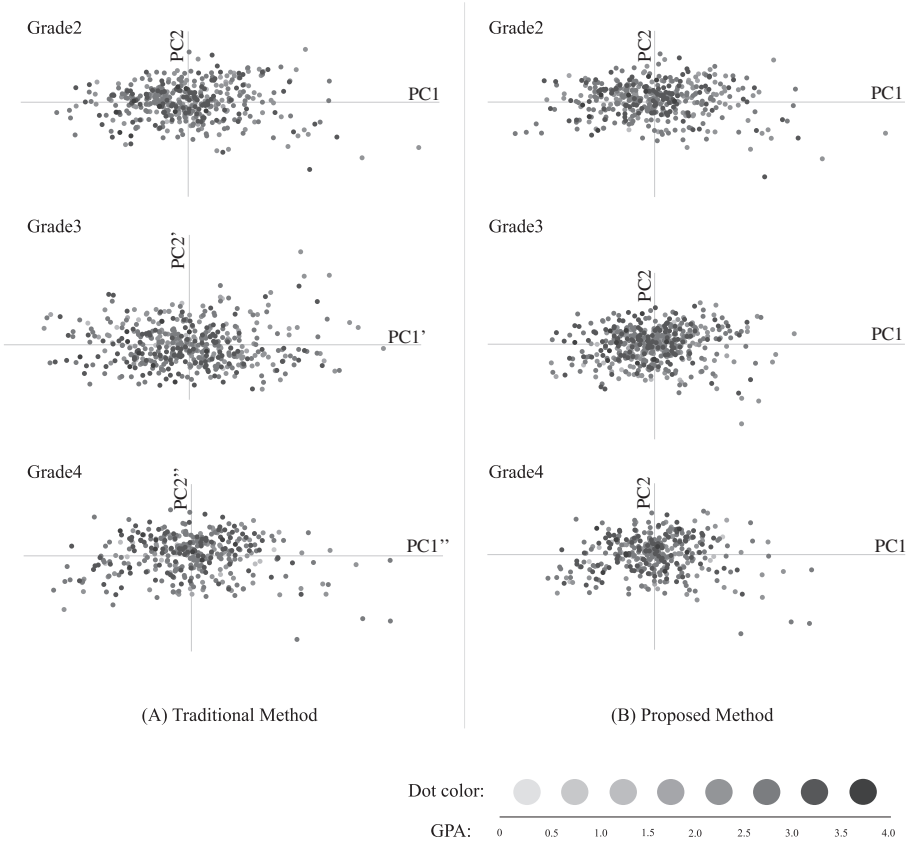

Fig. 1: Scores plots comparison between the traditional method and the proposed method.

Table 2: Interpretation of PCs depending on loadings of PCA results by using the proposed method.

Sample	Interpretation of PCs by understanding the selected descriptions		
		+ (in plus direction)	- (in minus direction)
All samples	PC1	attitude of self-advancement for employment goes weak	attitude of self-advancement for employment goes strong
	PC2	attitude of just enjoying university life	attitude of achieving the goal by actions

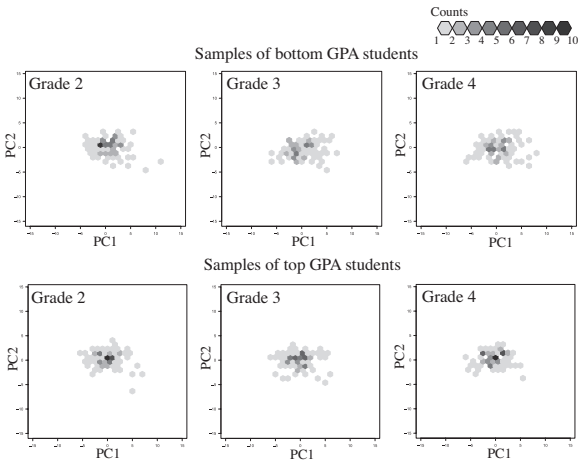

Fig. 2: Hexagonal binning on results of PCA.

Table 3: Number of hexagonal cells

		Bottom GPA samples	Top GPA samples
Grade2	Counts ≥ 4	4	4
	Counts $= 1$	37	39
Grade3	Counts ≥ 4	2	6
	Counts $= 1$	40	40
Grade4	Counts ≥ 4	3	4
	Counts $= 1$	41	31

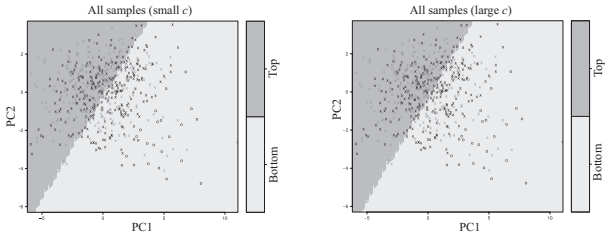

Fig. 3: SVMs plots for all samples.

attitudes of career design grows concerning distribute of the top GPA students, and distribution of the bottom GPA students' attitudes keeps splitting.

The graphs in Fig. 3 and Fig. 4 show the results obtained from the step 4. The graphs in Fig. 3 shows the results from all samples, and the graphs in Fig. 4 shows the results from each grade. In each figure, 'x' indicates support vector, and 'o' indicates data sample. Dimensions of SVMs results are fixed as PC1 and PC2 which are acquired in the step 2. Meanings of the axes are shown in Table 2.

From Fig. 3, we can see that the top GPA samples take positions mostly in the fourth quadrant,

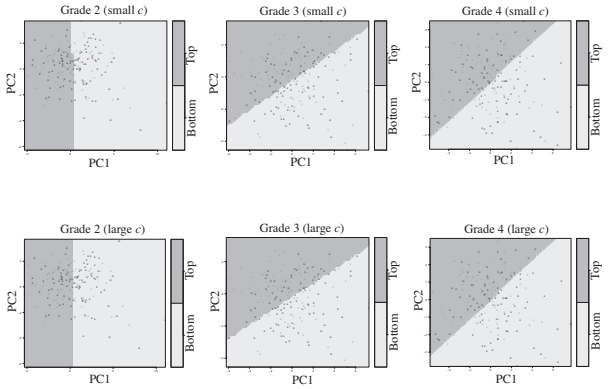

Fig. 4: SVMs plots for each grade.

which means that the top GPA students are more motivated by self-advancement for employment and enjoying university lives than the bottom GPA samples.

From Fig. 4, we can see that the top GPA samples of grade 2 distribute at the third and fourth quadrant of the two-coordinate space. It could be explained that the top GPA samples from grade 2 more tends to focus on self-advancement for employment than the bottom GPA samples. Concerning results of SVMs of grade 3 and grade 4 students, changing in attitudes occurred and the majority of the top GPA samples takes the fourth quadrant and it means that most of them think more about self-advancement for employment and have the attitudes of enjoying their studying lives. Comparatively, the bottom GPA students care more about taking actions for their compulsory missions.

The reason for the results could be that bottom GPA students set their goals as passing all the lectures instead of getting employment. For those top GPA students, they may have enough spare time to think about their careers. Otherwise, it is assumed that the top GPA students could enjoy their university lives more than the bottom GPA students during grade 3 and grade 4, because top GPA students could have received satisfying scores at an early stage.

From Fig. 3 and Fig. 4, we can see that the bottom GPA samples occupy larger space than the top GPA samples, which means that the attitudes of the top GPA students are more similar with each other. The same trend has also been discovered from using hexagonal binning plotting on PCA results of the proposed way. It is assumed that top GPA students could have similar goal which is to be get employed, and bottom GPA students could have their own goals and take actions under different attitudes.

4 Conclusion

For pursuing the relationship between students' attitudes of career design and their academic achievements, a new method, in which modified PCA and SVMs has been proposed. By applying the proposed method, the same principal components are used for projecting the samples' scores of each grade. Therefore, it is possible to compare and discuss changing in attitudes through a certain period.

Further analysis could be applied to explore valuable information from the results obtained by using the proposed method. If students from different courses are taken into consideration, the employers and job contents will be also different. With the purpose of supporting career design of all students in different majors, it is supposed that the proposed method should be applied on samples from different departments to obtain an objective analysis result and understanding of students' attitudes. Based on the analysis, educational guidance and training could be schemed. In addition, individual changing history that is merged with several records obtained in a determined period is supposed to be studied for investigating individual career development process to support self-assessment. Through conducting self-assessment as a part of the career education, student can make career decisions properly and start planning for the next stage of life.

Acknowledgement

This work was supported by Grants-in-Aid for Scientific Research from Japan Society for the Promotion of Science (NO. 15K00486, 26350013).

Reference

1) P. Dale, J. Roth, and R. Noeth: Nationwide Study of Student Career Development: Summary of Results, ACT Research Report(1973)

2) MEXT: Survey data summary for educational institutes, Basic investigation on educational institutes(2012)

3) MEXT: Educational reform corresponded to industry requirements, Support on educational reform for universities(2012)

4) EJ McCormick, PR Jeanneret, and RC Mecham: A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ) Journal of Applied Psychology, Vol.(**56**)-4, 347(1972)

5) S. Wold, K. Esbensen, and P.Geladi: Principal Component Analysis, Chemometrics and Intelligent Laboratory Systems, Vol.(**2**)-1, 37/52(1987)

6) R. Launois, J. Reboul-Marty, and B. Henry: Construction and validation of a quality of life questionnaire in Chronic Lower Limb Venous Insufficiency (CIVIQ), Quality of Life Research, Vol.(**5**)-6, 539/554(1996)

7) J. Wardle, C. A. Guthrie, S. Sanderson, and L. Rapoport: Development of the Children Eating Behaviour Questionnaire, Journal of Child Psychology and Psychiatry, Vol.(**42**-7, 963/970(2001)

8) D. B. Carr, R. J. Littlefield, W. L. Nicholson, and J. S. Littlefield: Scatterplot Matrix Techniques for Large N, Journal of the American Statistical Association, Vol.(**82**)-398, 424/436(1987)

9) C. Huang, J. A. McDonald, and W. Stuezle: Variable resolution bivariate plots, Journal of Computational and Graphical Statistics, Vol.(**6**)-4, 383/396(1997)

10) C. Cortes and V. N. Vapnik: Support-Vector Networks Machine Learning, 20, 273/297(1995)