Integrating SOM with K-Means in Taiwan TrainQuali System Performance Assessment

WenTsann Lin, ShihWei Shu
Dept. of Industrial Engineering and Management
National Chin-Yi University of Technology
Taichung City 41170, Taiwan

YungTsan Jou, Yihchuan Wu
Dept. of Industrial and System Engineering
Chung-Yuan Christian University
Chung-Li, Taoyuan County 32023, Taiwan
jason_wu1102@yahoo.com.tw

Abstract—**The Taiwanese government has, based on the concept of the UK's Investors in People (IIP), and the Swiss ISO10015, drawn up the so-called Taiwan TrainQuali System (TTQS), a training process including Plan, Design, Do, Review, and Outcome. TTQS provides businesses and training institutions to carry out internal human training by offering a sound system of evaluating the education and training. In this study, confirmative comparison is made by two-stage clustering in "Self-Organizing Maps, SOM, and K-Means" to verify the consistence between the clustering results of all levels of grading. This study thus identifies the relevant factors that affect the effects of TTQS promotion, organizing and analyzing the current conditions regarding the extent of benefits for organizations from introducing the TTQS and the assistance to them in human training. Further, to verify, from the findings, that this training quality system can effectively enhance the human capital in organizations, such that the organizations that introduce and carry out TTQS all are able to construct the human training system that satisfies themselves.**

Keywords- IIP; ISO10015; K-Means; SOM; TTQS

I. INTRODUCTION

The Bureau of Employment and Vocational Training (BEVT) set up the quality assurance system for human training in Taiwan that is relevant to the international training quality to uplift the human training quality and specialized ability as well as international competitive edge for Taiwan.

Starting in 2005, the TTQS specifications began to take shape with the efforts of many experts, aiming at the training of reviewers, developing grading criteria with TTQS score cards and completing the pretest of ten organizations. With the modification by the government, the present changes in foci on and objectives of TTQS development are apparent; for example, the foci on development up to 2009 were uplifting the quality of three types of persons, implementing the website specially for TTQS and integrating marketing and promotion, while the objectives of development were to establish review consistence and standardization, to strengthen digital network platform and to lay the mid-term training quality system plan. Through constant modifications, the government aims to make TTQS ensure of the effects of Taiwanese enterprises' investment in human resources in a more complete manner.

II. RESEARCH METHODS

A. TTQS Indicators

Having a theme of investigating the performance in TTQS promotion, this study will use the 2010 TTQS Review database for analysis. In order to facilitate subsequent employment of related software to make analysis, this study will number the TTQS indicators. The indicators, bases for classification assessment and code list are as shown in Tables 1 below.

B. Two-step Clustering

Cluster analysis is a logical procedure of objective classification based on similarity and difference, having the purpose of facilitating the identification of the similarity between certain things and the division of them into a number of clusters by such feature such that those in a same cluster have high homogeneity and the samples among different clusters have high heterogeneity (Tzeng, 2010) [1].

Self-Organizing Map, SOM, by Kohonen (1990) [2], is applied to graphic recognition by mapping spatial data of high dimension to low dimension. As SOM makes clustering in the concept of nearest neighbor, the resultant clusters are seen with overlapping partitions. The resultant clusters of non-level-type clustering can produce non-overlapping partitions. Abidi & Ong (2000) [3] and Vesanto & Alhoniemi (2000) [4] both proposed the technique of two-stage clustering as the clustering strategy for conducting data mining.

The algorithm of two-stage clustering is as follows:

1. SOM: Kohonen used self-organization to conduct graphic recognition, where vector values in consistence with the graphic elements are used to classify the graphs.

2. K-Means: in terms of economy, simplicity and effectiveness, K-Means is a method worth application. It is a non-layer-type clustering, not subject to the effects of outliers, errors in distance measurement and selection of method of distance calculation. The results of cluster are better if the initial point of the cluster is known. The knowledge of number of clusters ahead of making clusters enables effective process of large amount of numeric data;

this method is suitable for dealing with the characteristics of convex clusters, while ending only fits local optimization.

TABLE I. BASES FOR CLASSIFICATION EVALUATION

Code	Level	Performance level	Executive results
8	Platinum medal	Benchmarking	Eligible for example of success in the industry
7	Gold medal	Excellent	Has really carried out and should continue to improve
6	Silver medal	Fair	Has carried out as appropriate and should make critical improvements
5	Bronze medal	A little disastisfacory	Has carried out partially and should make major improvements
4	Threshold	Poor	Obvious evidence to executive results has yet to be seen

Data source: Sorted by this study

III. EMPIRICAL MEASUREMENT AND ANALYSIS

This study uses all the businesses and training institutions in Taiwan (2010) [5] that have applied for "TTQS review service" to introduce the train quality system as research subjects.

After the analysis by Back-propagation neural networks model [6], comparison to verify is made by using two-stage clustering, which is to divide the data into a number of clusters such that data in a same cluster have high similarity; also, verification is conducted in the review results by the TTQS reviewers to compare with the reviews in the database for consistence. As such, the initial solutions obtained from the non-supervised SOM network in the first stage can be rendered into Table 2, the distribution of numbers of composing SOM clusters. These are further plotted into below Fig. 1. In the drawing for the distribution of SOM clusters, being the numbers of initial clusters in the first stage, the numbers are 9 in both "Business version" and "Training institution version".

To know the correct numbers of records in each cluster, it is possible to verify from the "distribution chart" the correct numbers of records in each cluster and the proportion of each cluster to the whole. Sorting results in Table 3, which indicates the distribution of composing numbers of K-Means clusters. As this study observes by "color normalization" with each cluster containing the records in the number of 100, that includes the ratio of the number of fields selected by overlapping; hence, normalization allows better understanding when observing the relationship between the clusters and the set fields by overlapping.

Now the initial numbers of clusters are substituted into K-Means before making step-two clustering of the nodes in the mapping drawing with K-Means of different concept of distances. A sorting results in Table 4 and Table 5, which is further plotted into Fig. 2, the distribution charts of two-stage clustering model by SOM－K-Means, in which the X-axis represents the initial numbers of clusters from step-one SOM and the Y-axis represents the numbers of clusters of the initial numbers of clusters when substituted into K-Means. Now diagonal plotting is made to display the distribution chart for inspecting the status of the two-stage clustering, whereby (nine) can be determined as the number

of clusters by two-stage clustering in the distribution chart of SOM－K-Means clusters in "Business version" and (nine) in "Training institution version". Consequently, the approximate number of records contained in each cluster can be obtained.
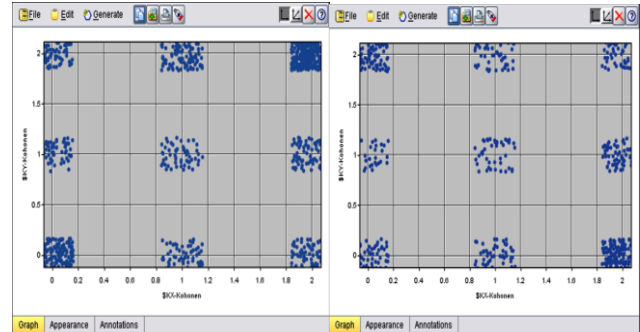

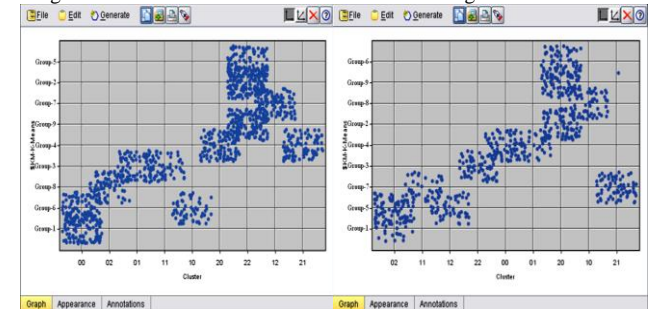Fig. 1 SOM Clusters-Business version and Training institution version

Fig. 2 SOM－K-Means Clusters-Business version (Left) and Training institution version (Right)

Fig. 2 above reveal the distribution of the distribution chart obtained by SOM－K-Means and further allows the observation of the degree of consistence. The analytical results in "Business version" are described as follows:

In respect to SOM clusters "00, 02 and 01" and K-Means "Clusters 1, 6 and 8" falling in this area, because their composition includes "Gold" (22 organizations) which all fall in this area, it means the high consistence of the grading for the level of "Gold" by the reviewers. It also suggests that organizations under review who fall in this area have higher performance in TTQS promotion, with the indication that the organizations under review with "Silver", "Bronze" and "Threshold" that fall in this area all have higher levels of performance in execution than the corresponding grades and that their performances are higher than those of all the businesses who exercise TTQS.

TABLE II. COMPOSITION TABLE OF SOM CLUSTERS

Business version						Training institution version					
$KX-Koh	$KY-Koh	Grade	Quant.	%	Total	$KX-Koh	$KY-Koh	Grade	Quant.	%	Total
0	0	Gold	22	11.58		0	0	Bronze	26	36.62	71
0	0	Silver	88	46.32	190	0	0	Threshold	45	63.38	
0	0	Bronze	80	42.11		0	1	Bronze	39	100	39
0	1	Bronze	82	100	82	0	2	Gold	12	11.11	
0	2	Bronze	30	34.09	88	0	2	Silver	36	33.33	108
0	2	Threshold	58	65.91		0	2	Bronze	60	55.56	
1	0	Bronze	62	100	62	1	0	Threshold	59	100	59
1	1	Bronze	28	49.12	57	1	1	Bronze	28	63.64	44
1	1	Threshold	29	50.88		1	1	Threshold	16	36.36	
1	2	Threshold	105	100	105	1	2	Bronze	52	100	52
2	0	Bronze	61	58.65	104	2	0	Threshold	225	100	225
2	0	Threshold	43	41.35		2	1	Threshold	72	100	72
2	1	Threshold	95	100	95	2	2	Bronze	46	65.71	70
2	2	Threshold	404	100	404	2	2	Threshold	24	34.29	

Data source: Sorted by this study

TABLE III. COMPOSITION TABLE OF K-MEANS CLUSTERS

Business version						Training institution version					
Cluster	Code	Grade	Quant.	%	Total	Cluster	Code	Grade	Quant.	%	Total
1	7	Gold	22	23.16		1	7	Gold	12	40.00	30
1	6	Silver	72	75.79	95	1	6	Silver	18	60.00	
1	5	Bronze	1	1.05							
2	4	Threshold	157	100	157	2	4	Threshold	86	100	86
3	5	Bronze	110	79.14	139	3	5	Bronze	46	65.71	70
3	4	Threshold	29	20.86		3	4	Threshold	24	34.29	
4	5	Bronze	61	30.65	199	4	5	Bronze	65	59.09	110
4	4	Threshold	138	69.35		4	4	Threshold	45	40.91	
5	4	Threshold	107	100	107	5	6	Silver	18	13.85	130
						5	5	Bronze	112	86.15	
6	6	Silver	16	10.19	157	6	4	Threshold	61	100	61
6	5	Bronze	141	89.81							
7	4	Threshold	105	100	105	7	5	Bronze	28	24.35	115
						7	4	Threshold	87	75.65	
8	5	Bronze	30	34.09	88	8	4	Threshold	59	100	59
8	4	Threshold	58	65.91							
9	4	Threshold	140	100	140	9	4	Threshold	79	100	79

Data source: Sorted by this study

2. In respect to SOM clusters "11, 10 and 20" and K-Means "Clusters 3, 4 and 9" falling in this area, because the composition of this area includes "Bronze" and "Threshold", it means that the organizations under review who fall in this area are seen to be inferior to those with "Gold" and "Silver" in terms of the performance in the operations of TTQS promotion, such as the extent to which indicators are carried out and the comprehensiveness, but they are not inferior to most organizations under review with "Threshold"; they thus present performance in TTQS promotion equal to the overall performance by all organizations.

3. In respect of the SOM clusters "22, 12 and 21" and K-Means "Clusters 7, 2 and 5" falling in this area, because the composition in this area is solely "Threshold", it suggests that the organizations under review who fall in this area are relatively weak in the cognition and promotion of TTQS such that their levels of performance are lower than that by all TTQS promoting organizations.

TABLE IV. COMPOSING TABLE OF SOM - K-MEANS CLUSTERS

Business version					
SOM cluster	K-Means cluster	Grade	Quantity	%	Total
00	Cluster 1	Gold	22	23.16%	
00	Cluster 1	Silver	72	75.79%	95
00	Cluster 1	Bronze	1	1.05%	
22	Cluster 2	Threshold	157	100%	157
01	Cluster 3	Bronze	82	59.00%	
11	Cluster 3	Bronze	28	20.14%	139
11	Cluster 3	Threshold	29	20.86%	
20	Cluster 4	Bronze	61	30.65%	
20	Cluster 4	Threshold	43	21.61%	199
21	Cluster 4	Threshold	95	47.74%	
22	Cluster 5	Threshold	107	100.00%	107
00	Cluster 6	Silver	16	6.35%	
00	Cluster 6	Bronze	79	31.35%	252
10	Cluster 6	Bronze	157	62.30%	
12	Cluster 7	Threshold	105	100.00%	105
02	Cluster 8	Bronze	30	34.09%	88
02	Cluster 8	Threshold	58	65.91%	
22	Cluster 9	Threshold	140	100.00%	140

Data source: Sorted by this study

TABLE V. COMPOSING TABLE OF SOM - K-MEANS CLUSTERS

Training institution version					
SOM cluster	K-Means cluster	Grade	Quantity	%	Total
02	Cluster 1	Gold	12	100.00%	12
20	Cluster 2	Threshold	86	100.00%	86
22	Cluster 3	Bronze	46	65.71%	70
22	Cluster 3	Threshold	24	34.29%	
00	Cluster 4	Bronze	26	23.64%	
00	Cluster 4	Threshold	45	40.91%	110
01	Cluster 4	Bronze	39	35.45%	
02	Cluster 5	Silver	18	13.85%	
02	Cluster 5	Bronze	60	46.15%	130
12	Cluster 5	Bronze	52	40.00%	
20	Cluster 6	Threshold	60	98.36%	61
21	Cluster 6	Threshold	1	1.64%	
11	Cluster 7	Bronze	28	24.35%	
11	Cluster 7	Threshold	16	13.91%	115
21	Cluster 7	Threshold	71	61.74%	
10	Cluster 8	Threshold	59	100.00%	59
20	Cluster 9	Threshold	79	100.00%	79

Data source: Sorted by this study

The analytic results in "Training institution version" are described as follows:

1. In respect of the SOM clusters "02, 11 and 12" and K-Means "Clusters 1, 5 and 7" who fall in this area, because the composition of this area includes "Gold" (12 organizations), all of them falling in this area, it suggests high consistence of the grading for the level of "Gold" by the reviewers, also indicating that the organizations under review with "Silver", "Bronze" and "Threshold" that fall in this area all have higher levels of performance in execution than the corresponding grades and that their performances are higher than those of all the businesses who exercise TTQS

2. In respect of the SOM clusters "22, 00 and 01" and K-Means "Clusters 3, 4 and 2" who fall in this area, because the composition of this area includes "Bronze" and "Threshold", the organizations under review who fall in this area are seen to be inferior to those with "Gold" and "Silver" in terms of the performance in the operations of TTQS promotion, such as the extent to which indicators are carried out and the comprehensiveness, but they are not inferior to most organizations under review with "Threshold"; they thus exhibit performance in TTQS promotion equal to the overall performance by all organizations.

3. In respect of the SOM clusters "20, 10 and 21" and K-Means "Clusters 8, 9 and 6" falling in this area, because the composition in this area is solely "Threshold", it suggests that the organizations under review who fall in this area are relatively weak in the cognition and promotion of TTQS such that their levels of performance are lower than that by all TTQS promoting organizations.

The analytic results of the above-stated "verification and comparison by two-stage clustering" indicate that TTQS is able to assist the Taiwanese businesses and training institutions as a whole in the uplifting of train quality and human capital to a high extent and that by the consistence of the TTQS framework with the training system which the organization requires, the organization can incorporate a sound training system of high quality to further advance toward the achievement of the goal of continuous growth, self-breakthrough and sustainable operation.

IV. CONCLUSIONS

The object of this study is to examine whether the TTQS evaluation system can bring effective and appropriate benefits to the human capital and education and training systems of businesses and training institutions as a whole in Taiwan, that is, to analyze the TTQS "review mechanism". Hence, courtesy of BEVT, the 2010 TTQS Review database was mined and the businesses and training institutions who have introduced TTQS were observed to see whether the indicators and specifications have all been internalized and carried out.

Research results are described below based on the analytic results from the research process built herein:

1. "Two-stage clustering" tests and compares the clustering results and verifies the consistence between grades for clusters. In the concept of similarity measurement by cluster analysis, the two-stage clustering of "SOM+K-Means" defined herein is employed to work on the 2010 TTQS Review database. First, data are clustered

by SOM to obtain the initial solution (numbers of clusters) before the initial numbers of clusters are substituted in K-Means to make step-two clustering on the nodes on the mapping chart. Such clustering results are used to verify the degree of consistence between the review operations by TTQS reviewers.

2. The "two-stage clustering" reveals that in the results in both "Business version" and "Training institution version", when the TTQS reviewers grade an organization for the levels of "Gold" and "Silver", it also suggests consistence in the grading with these two levels. In contrast, as with the grading for "Bronze" and "Threshold", less consistence is seen. The cause may be that the organizations that are graded for these two levels are less able to grasp the TTQS spirit and the requirements of the indicators in the introduction and execution of TTQS, such that they embody the TTQS with lower degree of completeness and thoroughness, which receives lower levels of grading from the reviewers.

Taiwan TrainQuali System has been able to positively and fast benefit the businesses and training institutions in Taiwan, assisting in the uplift of human capital and in the building of education and training system, further to draw the nation's attention to human resources and education and training, bringing up the nation's strength of human capital to be internationally competitive. The research findings are also provided as reference in continued exercise of TTQS in future for businesses and training institutions.

REFERENCES

[1] S.-T. Tzeng, "A Studying of Energy Performance Indicators In National Cheng Kung University," master thesis at the Graduate Institute of Resources Engineering, National Cheng Kung University, 2010.

[2] T. Kohonen, "The Self-Organizing Map. Proceedings of the IEEE," 1990, pp.1464-1480.

[3] S.S. Abidi, and J. A Ong, "Data Mining Strategy for Inductive Data Clustering: Asynergy Between Self-Organising Neural Networks and K-Means Clustering Techniques," IEEE Transactions On Neural Networks, 2000, pp.568-573.

[4] J. Vesanto, and E. Alhoniemi, "Clustering of the Self-Organization Map," IEEE Transactions On Neural Networks, 2000, pp.568-600.

[5] TTQS Project Office, TTQS Development strategy planning for 2010 to 2012, 2009.

[6] Y.-C. Yeh, "Application and practice of neural network models," 8th ed., Taipei: Scholars Books, 2003.