

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Finding an Optimal Combination of Key Training Items Using Genetic Algorithms and Support Vector Machines

¹Chien-Che Huang, ²Ruey-Gwo Chung, ³Rong-Chang Chen,

⁴Tung-Shou Chen, ⁴Tzu-Ning Le, ³Chih-Jung Hsu and ⁴Ying-Chih Tsai

¹Department of Industrial Education and Technology, National Changhua University of Education, Changhua

²Department of International Business Management, Hsiuping Institute of Technology, Taichung

³Department and Graduate School of Logistics

⁴Graduate School of Computer Science and Information Technology,
National Taichung Institute of Technology, No. 129, Sec. 3, Sanmin Rd.,
Taichung, Taiwan, Republic of China

Abstract: The purpose of this study was to find a best combination of key training items. Companies are generally concerned about whether training can increase business performance and want to know what training items are crucial to enhancement of performance. Thus, there is a need to find the key training items. In this study, a combined scheme of Genetic Algorithms (GA) and Support Vector Machines (SVM) is employed to find the optimal combination of the key items. The data used are collected from some small and medium-sized enterprises and are from the database of the Bureau of Employment and Vocational Training (BEVT) in Taiwan. Results from this study show that an optimal combination of key items can be effectively found by using the proposed approach. When companies intend to successfully improve the business performance and cost-efficiently implement training, they can focus on the key training items.

Key words: Genetic algorithm, combinatorial optimization, support vector machine, training, TTQS

INTRODUCTION

To remain competitive, companies need to continuously invest in training their employees. Training can increase productivity, enlarge profits, improve customer satisfaction and have some other benefits. However, training that achieves these results is typically linked to business goals and performance (Department for Business Innovation and Skills, 2009). In addressing this, companies require a strategy for training. They must know what key training items are in top priority for enhancing business performance and should be implemented first and thus can conduct training in a cost-efficient way.

In order to help companies make their training effective and assess the benefits of training investments, some countries establish a training quality management system and provide a guideline standard for companies to follow (Council of Labor Affairs, 2008). Take Taiwan for example, the government set up a Taiwan Training Quality System (TTQS) to help companies evaluate the effectiveness and improve the performance of the training. To evaluate the effectiveness of a company's training system, a TTQS scorecard was developed. There are 21 items (an item may have several sub-items) in the TTQS scorecard for companies. It is difficult for companies to understand how many items and what items are crucial to contribution of the business performance since there are

huge kinds of combination of items. Consequently, an adequate scheme should be employed to address this problem.

The problem mentioned above is a problem of combinatorial optimization, in which the set of feasible solutions is discrete or can be reduced to a discrete one and the goal is to find the best possible solution (Wikipedia, 2009). One of the most challenging problems in combinatorial optimization is to effectively deal with the combinatorial explosion (Daintith, 2004; Gen and Cheng, 1999; Nofelt, 2009). It is reasonable to use a heuristic algorithm to solve this kind of problem. There are numerous heuristic methods employed to find solutions of this problem. Amongst them, Genetic Algorithms (GA) (Coley, 1999; Eiben and Smith, 2003; Fogel, 2006; Gen and Cheng, 1996, 1999; Goldberg, 1989; Holland, 1975; Mitchell, 1996; Winter *et al.*, 1995) has been proven to be very effective to solve some combinatorial optimization problems. Therefore, we will use GA to find an optimal combination in this study.

MATERIALS AND METHODS

Problem description: Because of limited budget and restricted time, companies need to know what key training items can really enhance company performance and thus

should have highest priority to being implemented first. The number of possible combination of training items is generally huge. For example, there are about:

$$\sum_{i=1}^{23} C_i^{23} \sim 8.4 \times 10^6$$

kinds of combination for a dataset with 23 columns. It is difficult for companies to know how many items and what items are crucial to the enhancement of company performance. To find the optimal combination of key training items, real-world data must be collected first. The data include:

- The scores in each training item in the scorecard in the last year
- The business performance in the last year
- Training-related information such as the number of staff responsible for the training in the last year
- The business performance in the following year

The business performance in the last year is compared with that in the following year to see if the business performance is enhanced. Subsequently, companies must be divided into two classes: one has enhanced her company performance and one hasn't. Finally, an approach must be employed to find the best combination of the items that can enhance company performance. After collecting the training-related data, an effective method should be employed to find what combination of training items is the best one to increase the business performance. Since, a company's budget in practice, is limited and the time is constrained, a practical and important issue is how many and what items have the priority to being executed. In this study, we will use an effective combined scheme to solve the problem.

TTQS scorecard: In order to ensure the benefits of investments in training, a training quality management system should be developed and implemented. During the past years, a number of tools were developed for this purpose. Among them, ISO 10015 Quality Standard (ISO, 1999) offers the best roadmap for organizations that are committed to guarantee a fair amount of ROI (Return on investment) from their organization's training investments. By deploying the ISO 10015 based training quality management system throughout the organization, it turns out to be possible to establish correlations between training and performance.

The government in Taiwan has set up a training quality management system to help companies ensure their training effectiveness. Based on the understanding that human capital (Bassi and van Buren, 1999; Schultz, 1961) is the most important element of productivity in the knowledge economics, the government developed a

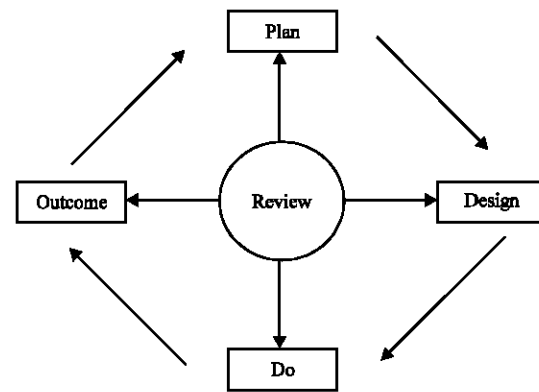


Fig. 1: A schematic diagram of the PDDRO management process

quality management system. The system is called TTQS. The TTQS provides five main management elements, i.e., plan, design, do, review and outcome (PDDRO). Followed by these five elements, a scorecard was developed and there are several items for each element to evaluate the effectiveness of the TTQS. A schematic diagram of the system is depicted in Fig. 1.

Twenty-one items are built up in the TTQS scorecard for companies. Table 1 shows these items. The items can be evaluated with a score of one to five, in which one stands for not implemented, two for the item is partially acknowledged and implemented and there are neither documented records nor evidence, three for the item is implemented according to documented processes, but records or procedures are incomplete, four for the item is implemented according to consistent processes and records and documents are completely maintained and five for the item is implemented, records and documents are completely maintained and analysis of relevant data and continual improvement are made. The minimum scale is 0.5. For companies, there are six grades and the minimum total score for pass is great than 53, as depicted in Table 2.

In addition to the above 21 items, some useful additional information can be added to increase the accuracy of the prediction on business performance. Table 3 lists the additional information used in this study.

The proposed approach: The procedure for finding the best combination of training items is depicted in Fig. 2. Firstly, real-world data of training items are collected. The data are from the Bureau of Employment and Vocational Training (BEVT) in Taiwan. Additionally, some data related to business performance are collected. For example, the yearly turnover (revenue) and the growth rate of turnover. In this study, we use the growth rate of turnover to represent the business performance. The

Table 1: TTQS scorecard for companies

Training element	Evaluation (training) Item
Plan	1 Defining <ul style="list-style-type: none"> • Disclosure of the vision, mission and strategy of organization • Defining business objectives and organizational needs • Defining training policy • Defining types or areas of core training 2 Documentation <ul style="list-style-type: none"> • TTQS and documented quality manual • Application of competence analysis related to training processes 3 Conformance <ul style="list-style-type: none"> • Conformance of training plan to achievement of business objectives 4 Competence <ul style="list-style-type: none"> • Competence of personnel performing work affecting training quality • Coordination of training-related competence
Design	5 Criteria for selecting training products or services 6 Participation process of the interested parties 7 Combining training to needs and objectives 8 Systematic design of training plans 9 Standardization of purchasing procedures for training-related products and services
Do	10 Level of conformance to the training plan <ul style="list-style-type: none"> • Conformance of trainee selection • Conformance of teaching material selection • Conformance of instructor selection • Conformance of training method selection • Conversion of learning accomplishments to work environment 11 Records and information system <ul style="list-style-type: none"> • Classification and documentation of training records • Level of MIS implementation
Review	12 Evaluation report and periodic integrated analysis 13 Monitoring and corrective actions <ul style="list-style-type: none"> • Monitor of execution process • Corrective actions
Outcome	14 Diversity and integrity of training outcome evaluation <ul style="list-style-type: none"> • Level 1: Response evaluation • Level 2: Learning evaluation • Level 3: Behavior evaluation • Level 4: Outcome evaluation 15 Working performance of trained employees 16 Organizational diffusion effect of training 17 Special training performance <ul style="list-style-type: none"> • Financial aspect, skill aspect and social aspect 18 Miscellaneous

Table 2: Grades of assessment in the TTQS (for companies)

Grades	Scores
Fail	≤53
Pass	53.5-63
Bronze award	63.5-74
Silver award	74.5-85
Golden award	85.5-96
Benchmark	≥96.5

Table 3: Additional information for predicting business performance

Item No.	Items
61	Capital
75	The number of staff performing training
76	Years of performing training

human resource management practices of a company can really influence its yearly turnover (Huselid, 1995). Thus, we will use the turnover growth rate as an indicator of business performance.

The analytical tools we used in this study are GA (Coley, 1999; Gen and Cheng, 1996, 1999; Goldberg, 1989; Holland, 1975; Mitchell, 1996; Winter *et al.*, 1995) and Support Vector Machines (SVM) (Burges, 1998; Corinna

and Vladimir, 1995; Liang *et al.*, 2008; Rüping, 2000; Vapnik, 1995). The GA has been developed and applied successfully to the problems of combinatorial optimization. It provides a number of feasible solutions, facilitating the decision under the environments of multiple-objectives and allowing the management to select the best alternative. The SVM was developed by Vapnik (1995) and has been presented with sound theoretical justifications to provide a good generalization performance compared to other algorithms (Burges, 1998). It has already been successfully used for a wide variety of problems, like pattern recognition (Burges, 1998), credit card fraud detection (Chen *et al.*, 2004, 2006a) and more. In this study, GA is employed to find the optimal combination that has the best prediction accuracy, which is obtained by using SVM. The details of steps are as:

Step 1: Encoding and initialization: The encoding of the chromosome is illustrated in Fig. 3. The binary encoding method is employed. The value of a gene is either 1 or 0,

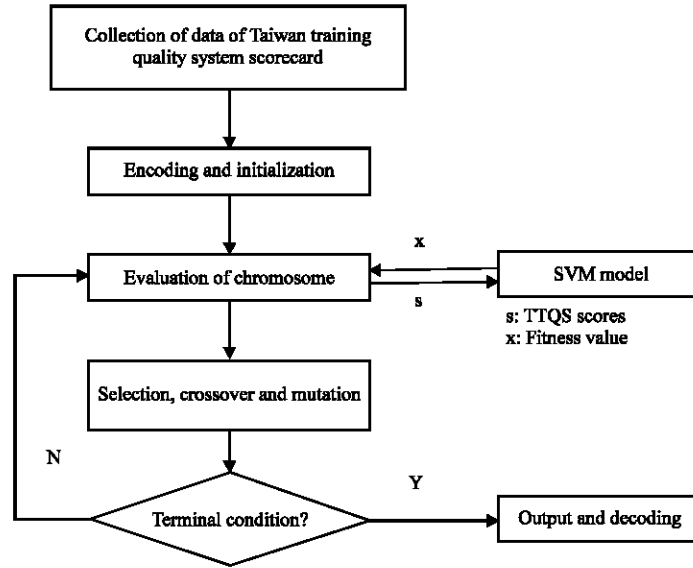


Fig. 2: The flowchart for finding the optimal optimization

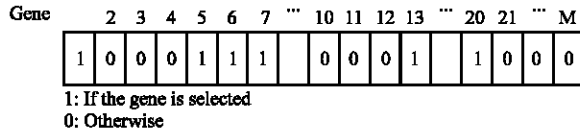


Fig. 3: Representation of chromosome

where 1 stands for the gene is selected and otherwise, 0 is given. A gene represents an item. There are M genes in total, including 20 items (item 18 Miscellaneous is excluded) and additional data items such as capital, the number of staff responsible for the training and more.

Step 2: Evaluation of the fitness of each chromosome:

Evaluation is performed based on the fitness of the chromosome. The fitness is defined as the prediction accuracy of the growth rate of yearly turnover. The growth rate is classified into two classes:

$$y_i = \begin{cases} 1 & \text{if the turnover growth rate of company } i \geq 20\% \\ -1 & \text{otherwise} \end{cases}$$

The growth rate is predicted using SVM. The classification tool we used in this study is LIBSVM (Chang and Lin, 2009). The kernel used is Radial Basis Function (RBF). There are two parameters for this function. Optimal parameters are first found using the tool provided by LIBSVM. The 10-fold cross validation is employed to investigate the prediction accuracy. The chromosome that has a better fitness value will have more chance to be retained to the next generation, depending on the elitism strategy.

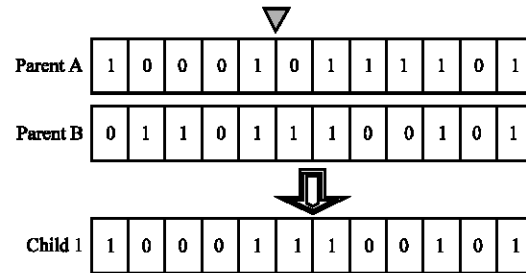


Fig. 4: Crossover operation

Step 3: Using Roulette Wheel Selection method to select parent chromosomes: The selection is performed according to the probability $P(x_i)$ defined as:

$$P(x_i) = \frac{x_i}{\sum_{j=1} x_j} \quad (1)$$

where, x_i stands for the fitness value of i th chromosome. A chromosome with a higher fitness value has more chance to be selected. In this study, the elitism strategy is employed. An assigned percentage of chromosomes which have best fitness values are directly retained to the next generation.

Step 4: Using crossover to generate a child chromosome as offspring: The crossover operation is shown in Fig. 4. A single-point crossover method is employed. A location in parent A and in parent B is randomly selected and their genes are combined to generate child 1, while the others are put together to produce the other child.

Step 5: Mutating to explore more solution space: The mutation is done using the single-point mutation. An assigned percentage of chromosomes are chosen to be mutated to generate the offspring.

Step 6: Replacement: Comparison of the fitness values in the population is made and the least fit offspring is replaced. A steady-state method is used, i.e., the number of chromosomes is fixed in the population.

Step 7: Termination condition: Stop, if the termination condition is satisfied. Otherwise, go to step 2. The termination condition is the pre-assigned generation number. When the generation is attained, the program is stopped and the result is outputted.

RESULTS AND DISCUSSION

The system we developed in this study has a user friendly interface. The GA program is coded by Java. The SVM employed in this study is LIBSVM (Chang and Lin, 2009). To facilitate the input of the collected data, four different input formats are allowed to use: LIBSVM, CLC (Chen *et al.*, 2006b), KNN (k-nearest neighbors algorithm) and Mysvm (Rüping, 2000). The system is illustrated in Fig. 5. In the main menu, the population size, the generation number, the crossover rate and the mutation rate can be easily entered. In addition, an elitism strategy can be implemented and a percentage of best chromosomes can be directly retained to the next generation.

When the program is executed, the prediction accuracy is calculated and outputted, as shown in Fig. 6.

Real-world data of training items are collected from 85 companies. The data are from the Bureau of Employment and Vocational Training (BEVT) in Taiwan. Additionally, some data related to business performance are also collected from the national database. The influences of genetic parameters are first examined to ensure that an optimal solution can be obtained. The variation of the fitness value with the generation number is observed and the experimental results show that the generation number is 200 can give a good and stable solution. Consequently, the generation number is set to be 200 in the following experiments. The population size is also changed and the results illustrate that 50 is enough to get best solutions. The influences of crossover rate and mutation rate are also tested. These two parameters have little influence when the generation number is 200 and the population size is 50.

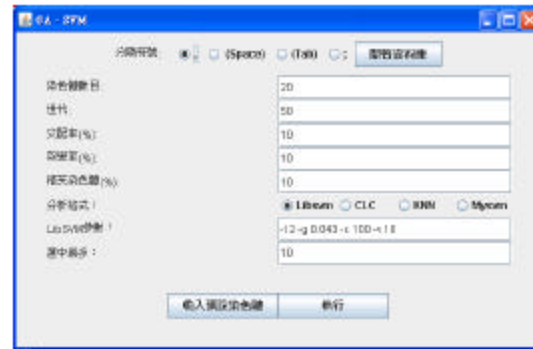


Fig. 5: The GA-SVM system

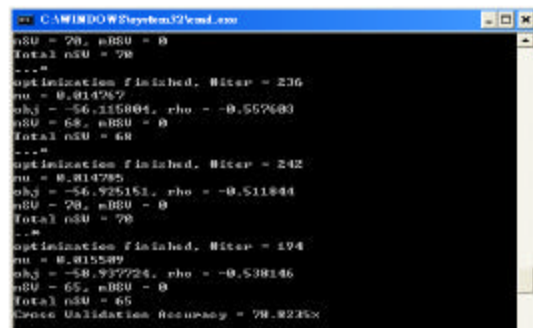


Fig. 6: The result of prediction accuracy by SVM

Table 4: Optimal combinations of key training items with different k

The No. of key items k	Evaluation (Training) items							
2	14	16						
3	2a	10e	16					
4	1d	2a	10e	14				
5	1d	2a	10e	14	16			
6	1d	2a	10e	14	15	16		
7	1d	2a	3	10e	14	15	16	

In practice, companies may not have much time and money to perform all the training items. There is, therefore, a need of finding some key training items that are in top priority to performing. The number of the key items k can be set in the GA-SVM program. Table 4 showed the experimental results. When k = 5, for example, the key items are 1d, 2a, 10e, 14 and 16 (Table 1), indicating that defining types or areas of core training, TTQS and documented quality manual, conversion of learning accomplishments to work environment, diversity and integrity of training outcome evaluation, organizational diffusion effect of training are crucial and should be executed first. If a company thinks five items to be too much, they can choose just three or four items by setting k = 3 or 4. The results demonstrate that the proposed approach can effectively and efficiently find the optimal combination of key training items for companies.

CONCLUSIONS

The purpose of this study is to apply a combined scheme of Genetic Algorithms (GA) and classification algorithms to find the best combination of key items for training. It is of most concern for many enterprises to understand if training can promote employees' competence and attain the organizational goal. Enterprises are also worried about if the implementation of training can boost management performance. In this study, the evaluation items in the Taiwan Training Quality System scorecard as well as the turnover growth rate are employed to classify 85 companies in the manufacturing industry in Taiwan. In addition, the key items for increasing the turnover growth rate of an enterprise are found using the combined scheme.

Results from the experiments show that an optimal combination of key items can be easily and effectively found. When enterprises aim to successfully improve the business performance and cost-efficiently implement training, they can concentrate on the key items of training. Further studies are recommended to use the combined scheme developed in this study to solve the other kinds of combinatorial optimization problems, such as key items for selecting potential companies in stock market. Moreover, other indicators of business performance can be employed to examine the correlation between the training items and business performance.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to the Bureau of Employment and Vocational Training (BEVT) in Taiwan. Appreciation is also extended to Miss Lindy Chen for her assist in the course of this study. This study was supported by the National Science Council under grant No. 97-2221-E-025-012.

REFERENCES

- Bassi, L.J. and M.E. van Buren, 1999. Valuing investment in intellectual capital. *Int. J. Technol. Manage.*, 18: 414-432.
- Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. *Data Min. Knowledge Discovery*, 2: 121-167.
- Chang, C.C. and C.J. Lin, 2009. LIBSVM A library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- Chen, R.C., M.L. Chiu, Y.L. Huang and L.T. Chen, 2004. Detecting Credit Card Fraud by Using Questionnaire-Responded Transaction Model Based on Support Vector Machines. In: *Intelligent Data Engineering and Automated Learning-IDEAL 2004*, Yang, Z.R. *et al.* (Eds.). LNCS. 3177, Springer-Verlag, Berlin, Heidelberg, ISBN: 978-3-540-22881-3, pp: 800-806.
- Chen, R.C., T.S. Chen and C.C. Lin, 2006a. A new binary support vector system for increasing detection rate of credit card fraud. *Int. J. Pattern Recognition Artificial Intel.*, 20: 227-239.
- Chen, T.S., C.C. Lin, Y.H. Chiu, H.L. Lin and R.C. Chen, 2006b. A New Binary Classifier: Clustering-Launched Classification. In: *Computational Intelligence*, Huang, D.S., K. Li and G.W. Irwin (Eds.). LNAI. 4114, Springer-Verlag, Berlin, Heidelberg, ISBN: 978-3-540-37274-5, pp: 278-283.
- Coley, D.A., 1999. *An Introduction to Genetic Algorithms for Scientists and Engineers*. 1st Edn., World Scientific Press, Singapore.
- Corinna, C. and V. Vladimir, 1995. Support vector networks. *Machine Learning*, 20: 273-297.
- Council of Labor Affairs, 2008. Quality assurance mechanism for vocational training. http://www.cla.gov.tw/cgi-bin/siteMaker/SM_theme?page=48f40081.
- Daintith, J., 2004. Combinatorial explosion a dictionary of computing. *Encyclopedia.com*. Jan. 25 2010. <http://www.encyclopedia.com/doc/1O11-combinatorialexplosion.html>.
- Department for Business Innovation and Skills, 2009. Fit the training to your needs. <http://www.businesslink.gov.uk/bdotg/action/detail?type=RESOURCES&itemId=1074453113>.
- Eiben, A.E. and J.E. Smith, 2003. *Introduction to Evolutionary Computing*. 1st Edn., Springer, New York, ISBN: 3-540-40184-9.
- Fogel, D.B., 2006. *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. 3rd Edn., IEEE Press, WileyBlackwell, Piscataway, ISBN-10: 0471669512. pp: 296.
- Gen, M. and R. Cheng, 1996. *Genetic Algorithms and Engineering Design*. 1st Edn., Wiley, New York.
- Gen, M. and R. Cheng, 1999. *Genetic Algorithms and Engineering Optimization*. 1st Edn., Wiley, New York.
- Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. 1st Edn., Addison-Wesley, New York, USA., ISBN: 0201157675.

- Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems*. 1st Edn., University of Michigan Press, Ann Arbor, Michigan, ISBN: 0472084607.
- Huselid, M.A., 1995. The impact of human resource management practices on turnover, productivity and corporate financial performance. *Acad. Manage. J.*, 38: 635-672.
- ISO, 1999. ISO 10015:1999: Quality management-guidelines for training. International Organization for Standardization, pp: 14. http://www.techstreet.com/cgi-bin/detail?doc_no=ISO%7C10015_1999_and_product_id=47998.
- Liang, X., R.C. Chen and X. Guo, 2008. Pruning support vector machines without altering performances. *IEEE Trans. Neural Networks*, 19: 1792-1803.
- Mitchell, M., 1996. *An Introduction to Genetic Algorithms*. 1st Edn., Massachusetts Institute of Technology, A Bradford Book, Cambridge, Boston, ISBN: 0262133164.
- Nofelt, P., 2009. Combinatorial explosion. http://pespmc1.vub.ac.be/ASC/COMBIN_EXPLO.html.
- Rüping, S., 2000. *mySVM-Manual*. Computer Science Department, AI Unit University of Dortmund, Dortmund, Germany.
- Schultz, T., 1961. Investment in human capital. *Am. Econ. Rev.*, 51: 1-17.
- Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory*. 1st Edn., Springer-Verlag, New York, ISBN: 0-387-94559-8.
- Wikipedia, 2009. *Discrete Optimization*. Elsevier, New York.
- Winter, G., J. Periaux and M. Galan, 1995. *Genetic Algorithms in Engineering and Computer Science*. 1st Edn., Wiley, New York.