Computers & Industrial Engineering 80 (2015) 171–180

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

A data mining approach for training evaluation in simulation-based training ☆

Juite Wang [a,*], Yung-I Lin [b], Shi-You Hou [a,1]

[a] Graduate Institute of Technology Management, National Chung Hsing University, 250 KuoKuang Road, Taichung 402, Taiwan, ROC
[b] Aeronautical Systems Research Division, National Chung-Shan Institute of Science and Technology, P.O. Box 90008-11-21, Taichung, Taiwan, ROC

ARTICLE INFO

Article history:
Received 25 October 2013
Received in revised form 17 July 2014
Accepted 8 December 2014
Available online 13 December 2014

Keywords:
Training evaluation
Simulation
Multimedia systems
Data mining

ABSTRACT

With the significant evolution of computer technologies, simulation has become a more realistic and effective experiential learning tool to assist in organizational training. Although simulation-based training can improve the effectiveness of training for company employees, there are still many management challenges that need to be overcome. This paper develops a hybrid framework that integrates data mining techniques with the simulation-based training to improve the effectiveness of training evaluation. The concept of confidence-based learning is applied to assess trainees' learning outcomes from the two dimensions of knowledge/skill level and confidence level. Data mining techniques are used to analyze trainees' profiles and data generated from simulation-based training for evaluating trainees' performance and their learning behaviors. The proposed methodology is illustrated with an example of a real case of simulation-based infantry marksmanship training in Taiwan. The results show that the proposed methodology can accurately evaluate trainees' performance and their learning behaviors and can discover latent knowledge for improving trainees' learning outcomes.

1. Introduction

Due to the increasingly complex and changing business environment, enterprise employees not only must possess required professional knowledge and skills, but also need to flexibly adapt their knowledge for use in the changing environment. To develop this adaptive expertise, trainees should be active participants in the learning process and learning should occur in a meaningful or relevant context (Bell & Kozlowski, 2002).

With the significant evolution of computer technologies, simulation has become a more realistic and effective experiential learning tool to assist in organizational training (Bell, Kanar, & Kozlowski, 2008). Simulation is defined as "an artificial environment that is carefully created to manage individuals' experiences of reality" (Bell et al., 2008). Simulation-based training (SBT), therefore, is "the ability to augment, replace, create, and/or manage a learner's actual experience with the world by providing realistic content and embedded instructional features" (Cannon-Bowers & Bowers, 2009). It is highly flexible in terms of place and time of training, which can be used to reduce or eliminate variable costs in traditional training. In addition, SBT can also provide the following advantages (Bell et al., 2008; Cannon-Bowers & Bowers, 2009): safer conditions than real-life situations, minimal influence from external factors, and more opportunities to repeatedly practice rare situations. It has been found that SBT is already used in academic and industrial applications, such as health care (Issenberg, Gordon, Gordon, Safford, & Hart, 2001; Issenberg, McGaghie, Petrusa, Lee, & Scalese, 2005; McGaghie, Issenberg, Petrusa, & Scalese, 2010; Salas, Wilson, Burke, & Priest, 2005), business education (Salas, Wildman, & Piccolo, 2009), pedestrian traffic (Usher & Strawderman, 2010), and disaster prevention (Summerhill et al., 2008).

Although SBT can improve the effectiveness of training for company employees, there are still many management challenges that need to be overcome (Bell et al., 2008; McGaghie et al., 2010). For example, several studies have indicated that the applications of SBT have produced mixed results and have not successfully and effectively grasped the advantages of SBT (Bell et al., 2008; Salas & Cannon-Bowers, 2001). Cannon-Bowers and Bowers (2009) also noted that past simulation-based education efforts have put too much effort on specific technological training systems and too little on training needs. Since SBT has been widely applied in the health care industry, several success factors have been identified in simulation-based medical education (McGaghie et al., 2010). One

☆ This manuscript was processed by Area Editor Paul Savory.

* Corresponding author. Tel.: +886 4 22840547x605; fax: +886 4 22859480.
E-mail addresses: rdwang@nchu.edu.tw (J. Wang), robin.bravo@msa.hinet.net (Yung-I Lin), jo760823@hotmail.com (S.-Y. Hou).
[1] Tel.: +886 4 22840547.

http://dx.doi.org/10.1016/j.cie.2014.12.008
0360-8352/

of the challenges for the development of SBT is how to effectively evaluate training performance and its subsequent impacts (Bell et al., 2008; McGaghie et al., 2010; Salas et al., 2005).

Data mining is the process of exploration and analysis of large quantities of data in order to discover meaningful patterns and rules that can improve business decision making (Berry & Linoff, 2004). It has gradually become an important tool for modern business to transform data into business intelligence and achieve competitive advantage. Accordingly, this paper proposes a hybrid framework that integrates data mining techniques with simulation-based training to improve the effectiveness of training evaluation. The concept of confidence-based learning (CBL) is applied to assess trainees' learning outcomes from the two dimensions of knowledge/skill level and confidence level. Data mining techniques are used to analyze trainees' profiles and data generated from SBT for evaluating trainees' performance and their learning behaviors. The proposed methodology is illustrated with an example of a real case of simulation-based rifle shooting training. From the experimental results, we show that the proposed methodology can accurately evaluate trainees' performance and their learning behaviors and can discover latent knowledge for improving trainees' learning outcomes.

Since SBT usually collects a large amount of data from training sessions, integrating the data mining techniques may be helpful for instructors and trainees to discover useful patterns or rules that can provide immediate feedback for trainees to improve their performance. However, data mining has been rarely used in the field of simulation-based training and this study aims to fill this research gap.

The paper is organized as follows: Section 2 presents the fundamentals of training evaluation and data mining techniques; the proposed methodology is presented in Section 3; the case study and computational experiment are illustrated in Section 4; Section 5 concludes the paper.

2. Fundamentals

2.1. Training evaluation and simulation-based training

Training evaluation is a process that compares the cost of training with the intended learning outcomes assessed in terms of improved performance by trainees (Buckley & Cape, 1990). It can help managers determine if a training program has achieved the desired results and diagnose the strengths and weaknesses of a program for needed improvements. According to Spitzer (1999), training evaluation can turn training into a powerful force for improvement of a business, for both the organization and the people in it.

Several training evaluation models have been developed in the literature (Eseryel, 2002; Holton, Bates, & Ruona, 2000; Moore, Green, & Gallis, 2009). Kirkpatrick's model (Kirkpatrick, 1998) is the most well-known and frequently used model for measuring the effectiveness of training programs in terms of reactions, learning, behavior, and results. The first level measures the immediate reactions of trainees towards training programs (e.g., enjoyment, perceived usefulness, and perceived difficulty). The second level measures the extent to which learning has occurred, where learning is conceived in terms of knowledge, skill, and/or attitude. Further levels measure whether job performance or organizational results have been changed as a result of training (e.g., turnover, volume of activity, cost-cutting, or quality indicators). Regardless of its popularity, Kirkpatrick's model continues to be criticized by researchers for issues, such as liability (Alliger & Janak, 1989) and limited variables and outcome measures (Santos & Stuart, 2003). For example, Phillips (1997, 2003) further developed a framework to compute the return on investment of training. However, challenges still remain in evaluating the effects that the training programs produce in the workspace and in the organization. Since knowledge retention, behavior changes, and organizational impacts resulting from training can only be apparent over time, behavioral and organizational criteria are difficult to measure. However, they are still necessary for training evaluation, because if the desired changes in attitude and behavior do not occur, then the training program is a failure.

SBT has strong potential to create a highly realistic training environment and allow trainees more active participation in the training process. Trainees are expected to act as if they are in a *real* situation. SBT also allows for repeated practice and the quest for excellence through error correction, feedbacks, and debriefing. These help trainees to develop expertise and the necessity for retention of these skills and behavior patterns (Issenberg et al., 2005). Instructors are able to rate trainees' behavior and give feedback to trainees for improving their performance.

Some research has shown that SBT could not only modify trainees' behavioral patterns but also increase their self-efficacy, promoting transfer of training to the workspace (McGaghie et al., 2010). It is well established that self-efficacy enhances learning outcomes and performance (Stevens & Gist, 1997). Similarly, Bruno (1993) also proposed the methodology of confidence-based learning (CBL) which can be used in a learning/training program to measure a trainee's knowledge quality by determining both the correctness of the trainee's knowledge/skill and his/her confidence in that knowledge/skill (see Fig. 1). Once the knowledge/skill correctness and confidence levels have been identified, CBL can identify the learning behavior of a trainee into categories: 'uninformed', 'misinformed', 'doubt', and 'mastery'. Then the instructor can diagnose the learning behavior of a trainee and provide useful feedback to improve the trainee's learning performance. Hunt (2003) also showed that the retention of newly learned knowledge is systematically related to the confidence level people have about the correctness of knowledge. A similar concept was also developed by Jeffries (2005), who included knowledge/skill performance and self-confidence in a simulation-based learning model for nursing. In addition, Yen, Ho, Chen, Chou, and Chen (2010) proposed a confidence-weighting computerized adaptive testing model that provided a more interactive testing environment by focusing on the examinees' confidence in their responses. Their results showed their model yielded ability estimates that were higher and better correlated to examinees' performance in English learning.

2.2. Data mining

The goal of data mining is to extract meaningful patterns and rules from a data set and transform it into an understandable structure for further use (Han & Kamber, 2006; Witten, Frank, & Hall, 2011). Data mining involves various techniques including statistics, neural networks, decision trees, genetic algorithms, and

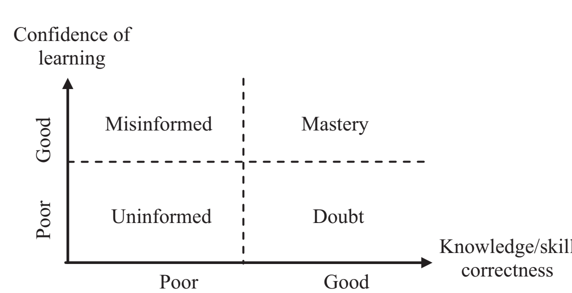

Fig. 1. Classification of learning behavior.

visualization techniques. It has been applied in many fields, such as design (Kwon, Omitaomu, & Wang, 2008), manufacturing (Ferreiro, Sierra, Irigoien, & Gorritxategi, 2011), health care (Lim, 2013), customer relationship management (Chen, Fan, & Sun, 2012), and failure detection and prediction (Magro & Pinceti, 2009). However, it has been applied rarely in simulation-based training.

In general, data mining tasks can be classified into the following types: concept description, association rules, classification/prediction, cluster analysis, outlier analysis, and evolution analysis. Concept description is used to describe individual classes and concepts in summarized, concise, and precise terms. Association rules analysis is used to mining frequent patterns leading to the discovery of interesting associations and correlations within data. Classification/prediction is the process of finding a model that describes and distinguishes data classes, in order to use the model to predict the class of data objects whose class label is unknown. Whereas classification models predict categorical (discrete, unordered) labels, prediction models forecast continuous-valued functions. Unlike classification/prediction, which analyzes class-labeled data objects, cluster analysis groups data objects based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity. Outlier analysis can be used to determine outlier data that do not comply with the general behavior. Evolution analysis describes and models regularities or trends for objects whose behavior changes over time.

3. Proposed data mining approach

This paper develops a data mining approach to improve the effectiveness of training evaluation in SBT. Since confidence is a good indicator for knowledge retention (Hunt, 2003), this research assesses trainees' learning outcomes and behaviors based on the concept of confidence-based learning, which measures a learner's knowledge quality using the correctness of the learner's knowledge/skill and confidence in that area of knowledge/skill. It has been identified that measuring knowledge and confidence is a better predictor of performance than measuring knowledge alone. The proposed methodology consists of the following five phases (see Fig. 2): problem definition, data understanding and preparation, model building, model evaluation and analysis, and model deployment.

3.1. Problem definition

This initial step focuses on understanding the objectives and training requirements for an organization. It is necessary to know the background of target trainees, the content of tasks to be learned, and the context of training in order to find the best way to measure, analyze, and improve the training performance. Domain knowledge accumulation is important in order to understand the nature of the problem and substantially improve data mining effectiveness. It is important to work closely with domain experts to define the project objectives and the requirements from a business perspective. The project objective can then be more successfully translated into a data mining problem definition.

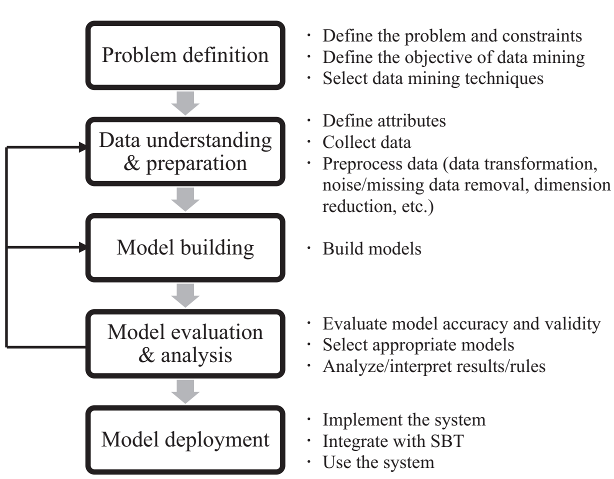

Fig. 2. The data mining methodology.

3.2. Data understanding and preparation

Collecting the right data is the basis of data mining. It is important to conduct a literature review of the problem domain and consult expert opinions in order to find and select for analysis the important data attributes that may influence trainees' learning outcomes.

Since this research applies the theory of CBL for training evaluation, data attributes are selected from the two aspects: knowledge/skill correctness and confidence in that knowledge/skill. For the dimension of knowledge correctness, the trainee's knowledge/skill that has been learned from training can be assessed in different scenarios given by the training simulator. Some measurable performance metrics, such as, test scores and overall answering accuracy, can be used to evaluate the correctness of a trainee's knowledge/skill. This data can be collected either from the training simulator or a questionnaire survey from learners. For the dimension of knowledge confidence, a questionnaire survey alone is used to collect the confidence-level data from trainees, because it is more subjective in nature. The confidence level can be assessed according to linguistic terms, such as: "I am sure", "I am partially sure", and "I am not sure" (Hunt, 2003). In addition, data related to trainee characteristics (e.g., age, gender, learning experience) is also collected because it may contain important factors that influence their learning outcomes (Cannon-Bowers & Bowers, 2009; Salas & Cannon-Bowers, 2001).

Before using data mining techniques to build models for analysis, it is important to cleanse the selected data and to transform it by joining and aggregation so that it is suitable for data mining analysis. It is also necessary to remove data that contains noise or is incomplete (Han & Kamber, 2006).

3.3. Model building

This phase applies the data mining techniques (Han & Kamber, 2006) to construct models for evaluating trainees' learning outcomes and their learning behaviors based on the concept of CBL. The present problem can be structured as a classification problem, which is used to predict the class of data objects whose class label is unknown. For example, a classification model could be used to identify a trainee's performance as "fail", "pass", or "excellent". In general, there are two types of classification: supervised classification and unsupervised classification. Supervised classification analyzes a dataset in which the class assignments are known and intends to find relationships between values of the predictors and values of the target variable. These relationships are summarized in a model, which can then be applied to a different data set in which the class assignments are unknown. Conversely, in unsupervised classification, the set of possible classes is not known, which is useful for analyzing problems with little or no pre-existing knowledge.

This paper applies supervised classification techniques for assessing trainees' learning outcomes because the rating assignments for trainees are usually known and can be collected from the simulation-based training. Three supervised classification techniques, decision trees (Quinlan, 1993b), artificial neural networks (ANNs) (Haykin, 1999; Wong, Bodnovich, & Selvi, 1997), and logistic regression (Hosmer & Lemeshow, 2000), are used for model building. The constructed models are evaluated and the one with the best prediction performance is selected for model deployment. Furthermore, this research applies cluster analysis, which is a type of unsupervised classification, to identify learning behaviors of individual trainees based on the CBL framework (see Fig. 1) because trainees' learning behaviors are not known prior to the training. Similarly, after trainees' learning behaviors have been identified, supervised learning techniques are used for building a model to predict trainees' learning behaviors. Each trainee can be assigned to one of the knowledge quadrants in the CBL framework, so the instructors can have a better understanding of trainees' learning behavior and provide feedback to trainees to improve their learning outcomes.

The following briefly introduces the theoretical backgrounds of these techniques. Interested readers can refer to the popular textbook by Han and Kamber (2006) for an introduction to data mining techniques.

(1) Decision tree analysis

Decision tree analysis is one of the commonly used techniques for supervised classification learning. This technique comprises of the construction of decision trees from a class-labeled training dataset. Let us denote the dependent variable by y (target variable) and the independent variables (predictors) by $x_1,x_2,\ldots,x_n$. A decision tree is a flow-chart-like tree structure, where each internal node (non-leaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The decision tree learning recursively splits the training dataset into subsets based on an attribute value test. This process is repeated on each derived subset, until the subset at a node has all the same value of the target variable (y), or when splitting no longer adds value to the predictions.

Many decision tree algorithms, such as ID3, C4.5, C5, and CART, have been developed in the literature (Breiman, Friedman, Olshen, & Stone, 1984a; Quinlan, 1986, 1993a). In this study, we chose to use the C5 and CART algorithms as our decision tree methods, because both algorithms can handle categorical and continuous predictors. In addition, since using a full-grown tree based on the training data may result in complete over-fitting of the data, validation data is used to prune the tree that has been over-grown using the training data.

The advantages of decision tree methods are their robust resistance to outliers and the presence of IF–THEN rules, which make it easier to understand and find important influential factors. However, due to the implicit linear structure during tree generation, if the problem structure is nonlinear, it is likely to have lower performance than methods such as ANNs (Lim, Loh, & Shih, 2000).

(2) Back-propagation neural networks (BPNN)

An artificial neural network (ANN) has at least two components: the processing units called neurons and the connections between them. Every connection has a weight parameter associated with it. Each neuron receives stimulus from the neighboring neurons connected to it, processes the information, and produces an output. A variety of neural network structures have been developed in the literature. A multilayer neural network with the back-propagation (BP) algorithm is the most commonly used network configuration (Haykin, 1999).

A multilayer feed-forward neural network typically consists of an input layer, one or more hidden layers, and an output layer. It is fully connected in that each unit provides input to each unit in the next forward layer. The BP algorithm propagates each input data object in the training dataset forward through the input layer, through hidden layers, to the output layer. The associated output value is calculated based on the current state of connection weights (initially, the weight will be random) and a nonlinear activation function. An example of a logistic activation function is:

$$Output_j = g\left(\theta_j + \sum_{i=1}^{p} w_{ij}x_i\right) = \frac{1}{1+e^{-\left(\theta_j+\sum_{i=1}^{p} w_{ij}x_i\right)}}, \tag{1}$$

where x_i is an input value from node i, θ_j is the bias of node j that controls the level of contributions of node j, and w_{ij} is the weight on the connection from node i to node j. The output value of a node in the output layer is then compared with the associated target output to compute the error for this input data object. Then, a gradient steepest descent approach is used to propagate this error back through the network adjusting the weights so as to minimize the sum-of-squares error. The whole process is repeated for each data object of the training dataset, until the overall error value drops below some pre-determined threshold.

The most prominent advantage of BP neural networks is their high tolerance to noise data and their ability to capture highly complicated relationships between the predictors and a target variable. However, it is not easy to determine the appropriate network size (e.g., number of hidden layers, number of nodes in each hidden layer) for a given problem complexity. The network size depends on the degree of nonlinearity and dimensionality of the given problem. Too many hidden neurons may result in over-fitting of the neural network, while fewer hidden neurons may not be able to accurately learn the problem behavior. Therefore, the adaptive method has been developed to determine the appropriate number of neurons by adding/deleting neurons as needed during training process (Hirose, Yamashita, & Hijiya, 1991; Rivals & Personnaz, 2003). Furthermore, another weakness is in providing insight into the structure of the relationship, which can be improved using the decision tree approach.

This paper will apply both the traditional BPNN with a fixed topology and the adaptive BPNN method for model building.

(3) Logistic regression

Logistic regression is a probabilistic statistical classification model that extends the ideas of linear regression to the situation where the dependent variable is categorical (Hosmer & Lemeshow, 2000). A categorical variable is used to divide the dataset into several classes. The purpose of the logistic regression model is to measure the relationship between a categorical dependent variable y and a number of independent variables, $x_1, x_2,\ldots,x_n$, by using probability scores as the predicted values of the dependent variable. The standard formulation of a logistic regression model is as follows:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_nx_n, \tag{2}$$

where p is the probability of belonging to a class specified by y and $\beta_1,\beta_2,\ldots,\beta_n$ are coefficients. The function on the left-side is called the *logit*, which is the natural log of the odds. Therefore Eq. (2) takes the *logit* as the dependent variable and models it as a linear function of the n predictors. The regression coefficients are usually estimated using the maximum likelihood estimation.

(4) Cluster analysis

Cluster analysis is the task of grouping a set of data objects in such a way that data objects in the same group are more similar to each other than to those in other groups. There are several clustering methods, such as iterative partitioning methods and hierarchical methods, which have been developed in the literature. The selection of the clustering algorithm appears to be critical to the successful use of cluster analysis. Though empirical studies of the performance of clustering algorithms suggest that the iterative partitioning methods, such as k-means algorithm, are preferable to the hierarchical methods, iterative partitioning methods require prior specification of the number of clusters desired and a good starting point (Punj & Stewart, 1983).

Therefore, this research uses the two-step clustering approach, where Ward's minimum variance method is first used to get some sense of the possible number of clusters and their centroids, and then the k-means method is applied to place all the data objects (Punj & Stewart, 1983). Ward's method, at each step, finds the pair of clusters that leads to the minimum increase in total within-cluster variance after merging, where the total within-cluster variance is defined as:

$$ESS = \sum_{j=1}^{k}\left(\sum_{i=1}^{n_j} X_{ij}^2 - \frac{1}{n_j}\left(\sum_{i=1}^{n_j} X_{ij}\right)^2\right), \tag{3}$$

where k is the total number of cluster, X_{ij} is the data object i assigned to cluster j, and n_j is the number of data objects assigned to cluster j.

According to a candidate number of clusters and their centroids determined by Ward's method, the second stage uses the k-means method which aims to partition n data objects into k clusters in which each data object x_j belongs to the cluster $S_i \in S$ $(1 \leqslant i \leqslant k)$ with the nearest mean so as to minimize the within-cluster sum of squares:

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2, \tag{4}$$

where μ_i is the mean of data objects in S_i.

3.4. Model evaluation and analysis

This step reviews and evaluates the constructed model, and the accuracy rate is used to evaluate the performance of the classification methods. If the model does not satisfy expectations, then the model is rebuilt by changing its parameters until optimal values are achieved. When the model obtains satisfactory results, it can be used for training evaluation, such as predicting trainees' performance ratings or extracting rules that can determine outstanding trainees.

3.5. Model deployment

Model deployment is to actually use the models created in the previous step for improving the performance of practical simulation-based training. For example, an organization may want to deploy a trained model or set of models (e.g., ANNs, decision trees) to quickly identify trainees who have a high probability of being failed. Then instructors can provide useful feedback to the trainees for improving their performance.

4. Empirical study: simulation-based infantry Marksmanship training in Taiwan

4.1. Background

With advances in computer technology, optical technology, and mechanical design, Taiwan' military has developed an assault rifle indoor shooting simulator for infantry marksmanship training. The entire system design intends to emulate the practice of different shooting scenarios (e.g., different weather conditions) so decision makers can enhance the effectiveness of training, while also reducing the required time, costs and potential risks. This study was done in cooperation with a military training unit that was planning to adopt an assault rifle shooting simulator for infantry marksmanship training.

The regular infantry marksmanship training session includes three phases: sight-in phase, 175M-practice phase, and 175M-final phase. In the sight-in phase, the trainee is allowed three rounds with three shots each to check and adjust the zero of the weapon. In the 175M-practice phase, each trainee has six shots without time limit to exercise his/her shooting knowledge and skills. Finally, the 175M-final phase examines the shooting performance of each trainee, who has to finish 6 shots within 30 s. The shooting score is proportional to the number of hits on the camouflage target and is calculated as follows: 30 points for one hit, 50 points for two hits, 60 points for three hits, 70 points for four hits, 90 points for five hits, and 100 points for six hits. According to the scores obtained, each trainee is further divided into three performance levels: "fail" (0, 30, 50), "pass" (60, 70), and "excellent" (90, 100).

4.2. Problem definition

Though using the rifle marksmanship simulator can improve the effectiveness of training while reducing the required time, costs and potential risks, the Taiwanese military still sought to more fully take advantage of SBT to enhance training evaluation so that it can further help trainees improve their learning outcomes. The proposed data mining approach could fulfill this need, because it is much easier and faster to collect training data from the rifle simulator-based training. An instructor can utilize the data collected during training sessions, analyze training data for assessing the trainee's learning outcomes, and give timely feedback to trainees and instructors for improving trainees' shooting performance. This can simplify the time and effort of instructors in providing recommendations for improving trainees' shooting skill and enhancing the training effectiveness. The following three issues are addressed in this case study:

- Assessing the shooting performance and learning behavior of a trainee based on the concept of CBL.
- Understanding the patterns of outstanding trainees.
- Providing recommendations back to instructors and trainees for improving trainees' learning outcomes.

4.3. Data understanding and preparation

Since this research assesses trainees' learning outcomes from the two perspectives of knowledge/skill correctness and confidence in that knowledge/skill, it is necessary to define data attributes that can be used to measure trainees' performance from both dimensions and design questionnaires for collecting this data.

The questionnaire was developed based on a literature review, a marksmanship training manual, and opinions from the shooting instructors. Due to the diverse backgrounds of trainees, the questionnaire had to be designed for easy understanding and answering by respondents. After the original questionnaire design was completed, instructors responsible for the overall training lessons reviewed the questionnaire to ensure that all relevant questions were included.

The required data for this study includes the following three parts:

(1) *Basic profile:* basic information of a trainee including the following nine items: identification number, age, height, weight, body mass index (BMI), academy background, sight, military service years, and shooting training times attended.

(2) *CBL items:* the retention of newly acquired knowledge for a trainee is significantly related to the trainee's confidence in that knowledge (Hunt, 2003). Therefore, we include question items that enable trainees to recall past basic training experience and self-assess their confidence in the shooting knowledge and skills in which they have already been trained, including the following five items: proficiency of eight shooting principles, shooting confidence, proficiency in gun use, concentration degree, and shooting posture. A three-point scale was used in the questionnaire to measure the confidence of knowledge learned by the trainee: "I am sure", "I am partially sure", and "I am not sure".

(3) *Marksmanship data collected from the SBT sessions:* the data collected here is used to assess the correctness of the knowledge/skill learned in the training lessons. The collected data includes gun holding position, coordinates of shot holes, shooting scores, and performance levels for three training phases. According to the coordinate data of shot holes, the shooting precision and accuracy can be calculated. Let (x_i, y_i) be the coordinate of a shot point, where $i = 1–6$. The shooting precision is defined by the average distance of any two shot points:

$$\alpha = \frac{\sum_{i=1,j=i+1}^{6} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{6}, \tag{5}$$

where a larger value of α indicates lower precision. Shooting accuracy is defined by the average distance from every shot point to the camouflage target center (x_t, y_t):

$$\beta = \frac{\sum_{i=1}^{6} \sqrt{(x_t - x_i)^2 + (y_t - y_i)^2}}{6}, \tag{6}$$

where a larger value indicates lower accuracy. The following six items are used for analysis: gun holding position, shooting precision and accuracy and performance levels for the sight-in and 175M-practice phases, and performance level for the 175M-final phase.

The questionnaire survey was conducted on the infantry training base for 2 months. A pre-test was performed prior to actual data collection both for item validation and to avoid possible ambiguity. This was conducted on five shooting instructors who received a printed version of the questionnaire and were asked to comment on the listed items, wording, and format. As a result, a few overlapping items were removed and some items were modified to make their wording as precise as possible.

Before the shooting training session began, the research team distributed the questionnaires to respondents and explained its content in detail to ensure that each respondent clearly understand all question items. After each run of a training session, the research team gathered completed questionnaires from the trainees and also collected the target papers recording locations of shooting holes for all trainees.

After removing questionnaires with missing data and unclear answers, a total of 289 (95.7%) questionnaires and shooting records were collected. The required data were processed and formatted to fit a format compatible with data mining functions.

4.4. Model building

4.4.1. Shooting performance assessment

Based on the collected data, this phase applied supervised classification techniques to assess trainees' shooting performance, including two decision tree methods (C5 and CART), two BPNN approaches (traditional and adaptive methods), and the logistic regression approach. The original data set contained twenty-three variables which might contain redundant and irrelevant information. Feature selection methods, such as likelihood ratio, Pearson's Chi-square, one-way ANOVA *F*-test, Cramer's *V*, and Lambda can be used to select important variables that most contribute to prediction accuracy (Guyon & Elisseeff, 2003; Saeys, Inza, & Larranaga, 2007). Since our predictors contained categorical and continuous variables and can obtain a better prediction accuracy in the preliminary computational experiment, one-way ANOVA *F*-test was applied in this research to identify important variables to target variables (i.e., "shooting performance level") from the dataset. The importance value of each variable was then calculated as $(1 - \gamma)$, where γ is the p value of the appropriate test of association between the candidate predictor and the target variable. The variables with importance greater than 0.95 were selected and a total of the top eleven data attributes were used for model building, where three confidence-related data attributes, including confidence about shooting posture, proficiency of the eight shooting principles, and proficiency in gun use, were included (see Table 1). The data mining software SPSS PASW modeler 13.0 on the PC platform was used for building models in this study.

Next, the preliminary computational experiments were conducted to identify suitable parameter values for building models with the five supervised classification methods. In the training of C5.0 decision tree method, the information gain criteria and the boosting method (number of trials = 10) were applied to grow the initial decision tree for improving its accuracy rate. To avoid the overfitting the training data and poorly generalizing to new samples, global pruning was applied with pruning severity within the range of 70–85 and the minimum records per child branch within the range of 5–10 (Quinlan, 1992). The final parameter values were determined in order to minimize the misclassification error in the subsequent validation phase. For the CART decision tree method, both the Twoing and the Gini indices (Breiman, Friedman, Olshen, & Stone, 1984b) were tested for growing the initial tree and the Gini index was selected with for model building, because of its superior performance. The minimum change in impurity to stop the tree growth was set to 0.0001. The standard error rule was used to prune the tree based on the validation dataset to avoid the overfitting problem, while the standard error multiplier was set within the range of 1.0–2.5 for improving validation accuracy.

For the training of back-propagation neural networks, the number of input nodes, hidden nodes (only one hidden layer), and output nodes were set to 27, 21, and 3, respectively. The number of hidden nodes was determined based on the rule of thumb that

Table 1
Feature selection for shooting performance assessment.

Rank	Attribute	Importance
1	175M practice shooting score	1.000
2	Precision at the practice phase	1.000
3	Accuracy at the practice phase	1.000
4	Accuracy at the sight-in phase	0.999
5	Number of training lessons	0.976
6	Confidence about shooting posture	0.976
7	Academy degree	0.966
8	Confidence about proficiency of the eight shooting principles	0.966
9	Confidence about proficiency in gun use	0.965
10	Precision at the sight-in phase	0.959
11	BMI	0.956
12	Confidence about shooting performance	0.867
13	Right sight	0.866
14	Gun holding position	0.711
15	Military service duration	0.682
16	Left sight	0.589
17	Confidence about concentration	0.46

Table 2
Performance comparison of different classification methods in the shooting performance assessment.

	Accuracy	Decision tree		BP neural network		Logistic regression (%)
		CART (%)	C5.0 (%)	Traditional (%)	Adaptive (%)	
Training	Mean	69.98	88.12	95.16	85.04	62.86
	SD	2.76	10.72	0.55	1.68	8.89
Validation	Mean	54.30	57.08	95.17	91.04	55.37
	SD	7.61	7.72	4.05	4.07	11.74

Table 3
Agglomeration schedule for Ward's minimum variance method.

Stage	Agglomeration coefficients	Percentage of changes (%)	Number of clusters
279	26.75	–	10
280	31.07	16.1	9
281	35.27	13.5	8
282	50.82	44.1	7
283	56.58	11.3	6
284	74.44	31.6	5
285	92.16	23.8	4
286	196.63	113.4	3
287	521.5	165.2	2
288	736.4	41.2	1

the number of hidden neurons should be 2/3 the size of the input layer, plus the size of the output layer. Momentum to avoid getting trapped in a local minimum was set to 0.9, while the learning rate which controls how much the weights were adjusted at each update was initially set to 0.3, decreased to 0.01, then rested to 0.1 and decreased to 0.01 for every 30 training cycle until the stopping criteria (training period = 5 min) was reached. For the adaptive BPNN, the number of hidden layers, hidden nodes and learning parameters were adjusted automatically during the training process. For the logistic regression, multi-nominal procedure and main effects option including all the variables specified with no interaction terms were selected.

Finally, the exclusive computational experiments were conducted to validate the models built by the above five supervised classification methods. The k-fold cross-validation method was used to estimate the performance of a predictive model and k was set to 10. Its advantage is that all the examples in the dataset are eventually used for both training and validation. A 10-fold partition of the dataset was created, where, for each of 10 experiments, 9 folds were used for training to build a predictive model and then the remaining one was used for testing model validity. The performance comparison results of different classification methods are shown in Table 2. Among the five classification methods, the BPNN models performed better than decision tree methods and logistic regression, whose average prediction accuracies were all below 60% in validation. In contrast, both BPNN models had much higher average prediction accuracies in validation that were greater than 90%. The traditional model could achieve the best performance with a 95.17% prediction accuracy with standard deviation equal to 4.05%. Therefore, the traditional BPNN model is recommended for use in trainees' shooting performance assessment.

4.4.2. Learning behavior classification

This paper applies cluster analysis to determine the learning behavior of a trainee based on confidence-based learning. The purpose of cluster analysis is to assign a set of data objects into several clusters, so the data objects in the same cluster are similar. This research applies the two-stage clustering analysis method, Ward's minimum variance method (Ward, 1963) followed by the k-means method (Hartigan & Wong, 1979), as proposed by Punj and Stewart

Table 4
Descriptive statistics of 4 clusters.

Variables	Statistics	Behavior			
		Uninformed	Misinformed	Doubt	Mastery
Basic profile					
Number of training lessons	Mean	3.347	3.405	3.689	4.266
	SD	1.927	1.992	1.924	1.978
Academy degree	Mean	2.541	2.633	2.767	2.719
	SD	0.558	0.616	0.794	0.826
BMI	Mean	22.111	22.387	22.694	23.079
	SD	2.936	2.628	2.801	2.836
Correctness of knowledge/skills					
175M practice shooting score	Mean	20.153	20.270	77.344	83.944
	SD	23.147	24.887	26.186	17.001
Accuracy at the practice phase	Mean	8.838	8.784	5.876	5.547
	SD	1.250	1.162	1.725	1.552
Precision at the practice phase	Mean	11.466	11.029	6.927	6.243
	SD	2.752	2.283	2.911	2.533
Accuracy at the sight-in phase	Mean	5.710	2.690	2.204	1.945
	SD	3.635	1.412	1.121	0.853
Precision at the sight-in phase	Mean	6.135	2.033	1.970	1.898
	SD	6.396	1.365	1.115	1.245
Confidence of learning					
Shooting posture	Mean	1.486	2.265	1.967	2.719
	SD	0.559	0.488	0.350	0.453
Proficiency of eight shooting principles	Mean	1.811	2.469	2.133	2.812
	SD	0.616	0.522	0.502	0.393
Proficiency in gun use	Mean	1.351	2.122	1.911	2.297
	SD	0.484	0.437	0.323	0.494

(1983). Table 3 presents the agglomeration schedule which provides a solution for every possible number of cluster from 1 to 10 and the corresponding agglomeration coefficients. This coefficient is the squared Euclidean distance between the two cases of clusters being combined. The coefficient can be used as a stopping rule that evaluates the changes in the coefficient at each stage of the hierarchical process. Large coefficients indicate that two very different clusters are being merged. The researcher looks for large increases in the value for determining the optimal number of clusters.

From Table 3 we can see that the noticeable jumps in percentage increase occur when going from 4 to 3 clusters. Therefore, the promising number of clusters is 4. Then the *k*-means method is applied to improve the quality of each clusters using the number of clusters as 4. The descriptive statistic for each cluster is displayed in Table 4. The major characteristics of each cluster are evaluated and a class name is assigned on each cluster based on the CBL framework in Fig. 1.

- Cluster 1: Uninformed

This cluster contained 37 trainees and had the smallest number of training lessons attended (3.347) among the four clusters. In the confidence of learning dimension, the confidence levels for shooting posture, proficiency in the eight shooting principles, and proficiency in gun use on shooting were the worst among the four clusters. In addition, for the correctness of knowledge/skill dimension, their shooting precision and accuracy at the sight-in and practice phases were the worst, and the average shooting score in the practice phase was only 20.153. In summary, trainees in this cluster had not learned the knowledge/skill from previous training lessons and they also lacked confidence in their knowledge/skill. Therefore they can be classified as ‘uninformed’.

- Cluster 2: Misinformed

This cluster contained 98 trainees, with an average of 3.405 training lessons attended, which is slightly more than the ‘uninformed’ cluster. For the confidence of learning dimension, the confidence levels on related attributes were worse than the ‘mastery’ cluster, but better than the other two clusters. Although the trainees in this cluster had enough confidence in their knowledge/skill, for the correctness of knowledge/skill dimension, the shooting precision and accuracy were slightly better than those in the ‘uninformed’ cluster, and the average shooting scores in the practice phase was only 20.27. In short, the trainees in this cluster had good confidence in their knowledge/skill, but the correctness of their knowledge/skill is poor. Thus, they can be classified as ‘misinformed’.

- Cluster 3: Doubt

This cluster contained 90 trainees, with an average of 3.689 training lessons attended, which is slightly less than the ‘mastery’ cluster. For the confidence of learning dimension, the confidence-related attributes were slightly better than the ‘uninformed’ cluster. For the correctness of knowledge/skill dimension, the shooting precision and accuracy at the sight-in and practice phases were worse than those in the ‘mastery’ cluster, but better than the other two clusters. Furthermore, the average shooting scores in the practice phase was 77.344. In short, the trainees in this cluster had poor confidence in their knowledge/skill, but they had good correctness of knowledge/skill, which was only worse than those in the ‘mastery’ cluster. Thus, they can be classified as “doubt”.

- Cluster 4: Mastery

This cluster contained 64 trainees, with an average of 4.266 training lessons attended, the most among the four clusters. In both confidence of learning and correctness of knowledge/skill dimensions, this cluster outperformed the other three clusters. In addition, the average shooting scores in the practice phase was 83.944. In short, the trainees in this cluster not only were confident in their knowledge/skill but also had excellent knowledge/skill correctness. Thus, they can be classified as ‘mastery’.

In addition, the supervised classification methods were used to predict trainee learning behavior based on the data collected from shooting training sessions. A new categorical attribute, knowledge quadrant, was added to the data set, and its value ranged within the 1-to-4 scale representing ‘uninformed’, ‘misinformed’, ‘doubt’, and ‘mastery’ categories. A one-way ANOVA *F*-test, was also used to select important variables that most contribute to prediction accuracy of behavior classification. The variables with importance greater than 0.95 were selected and a total of the top nine data attributes were used for model building (see Table 5). Similar to the shooting performance assessment, preliminary experiments were conducted to select suitable parameter values for model building. Then, 10-fold cross-validation was used to validate the performance of a predictive model. The prediction results are shown in Table 6. Though all five methods could obtain validation performance greater than 76.8% accuracy, only the BPNN approach had greater than 95% accuracy. The adaptive BPNN approach is

Table 5
Feature selection for shooting behavior classification.

Rank	Attribute	Importance
1	175M practice shooting score	1.000
2	Precision at the practice phase	1.000
3	Accuracy at the practice phase	1.000
4	Accuracy at the sight-in phase	1.000
5	Number of training lessons	1.000
6	Confidence about shooting posture	1.000
7	Confidence about proficiency of the eight shooting principles	1.000
8	Confidence about proficiency in gun use	1.000
9	Precision at the sight-in phase	0.968
10	BMI	0.709
11	Academy degree	0.464

Table 6
Performance comparison of different classification methods in shooting behavior classification.

	Accuracy	Decision tree		BP Neural net		Logistic regression (%)
		CART (%)	C5.0 (%)	Traditional (%)	Adaptive (%)	
Training	Mean	88.16	98.66	95.16	98.27	91.54
	SD	1.58	0.86	0.48	0.27	3.84
Validation	Mean	76.80	83.75	95.15	98.28	78.53
	SD	7.70	5.13	4.37	2.44	9.05

Table 7
Sample rules for classifying trainees' learning behavior

Rule no.	Knowledge/skill dimension			Confidence dimension			Behavior[a]	Confidence level
	175 Practice shooting score	Sight-in precision[b]	Training times	Shooting posture	Eight shooting principles	Proficiency in gun use		
1	(0,30]			L			1	0.933
2	(0,30]			M		L	1	0.717
3	(30,60]					L	1	0.889
4	(0,30]			M		M–H	2	0.871
5	(0,30]			H			2	0.815
6	(30,60]	M	⩾6		L–M	M	3	0.857
7	(60,75]					L–M	3	0.863
8	(75,100]			L–M			3	0.852
9	(75,100]			H			4	0.929

[a] 1: "uninformed", 2: "misinformed", 3: "doubt", 4: "mastery".
[b] Precision (cm): High: $\alpha < 3.8$, Medium: $3.8 \leqslant \alpha < 10.1$, Low: $\alpha \geqslant 10.1$.

recommended for system deployment, because its accuracy rate was 98.28%.

Though the ANN approach produced high accuracy rate for predicting a trainee's shooting performance, it is difficult to interpret the knowledge inside it. Thus, we applied the decision tree approach to discover useful patterns to improve trainee's trainee learning behavior which might result in better shooting performance. The obtained decision rules are summarized in Table 7. For example, Rule 1 shows that if a trainee's 175-meter practice shooting score is less than 30, and s/he has a low confidence about shooting posture, then the trainee is 'uninformed' with confidence degree 93.3%. In addition, Rule 4 shows that if a trainee's 175-meter practice shooting score is less than 30, but s/he has a medium confidence about shooting posture and even has a medium-to-high confidence about proficiency in gun use, then the trainee is 'misinformed' with confidence degree 87.1%. Further, it is interesting to note that the shooting score in the practice training is the dominant factor for shooting behavior classification in the correctness of knowledge/skill, while the confidence about shooting posture followed by confidence about proficiency in gun use and the eight shooting principles are the important influential factors in the confidence of learning dimension. Therefore, in addition to a good shooting score in the practice training, a "mastery" trainee should have high confidence in shooting posture. These rules can be used as a guideline to educate trainees to improve their shooting skills and confidence.

4.5. Model deployment

The training evaluation system was developed based on the models identified in the previous stages and implemented with Microsoft Visual Basic on the PC platform. The system framework is shown in Fig. 3. The simulation-based training process can be divided into the following three steps:

Step 1: Simulation-based training lessons

At first, the instructor asks each trainee to fill in the questionnaire containing the basic profile and confidence related questions. Following three stages of training lessons, the simulator replicates various scenarios and events and gathers shooting data of individual trainees during the training lessons. Then the shooting data and questionnaire data are fed into the training evaluation system.

Step 2: Diagnosis

The training evaluation system analyzes shooting data from the simulator and the trainees' profile data, and assesses trainees'

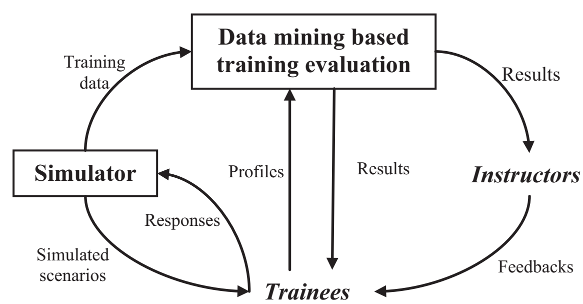

Fig. 3. System framework of the assault rifle shooting SBT.

shooting performance and their learning behavior. The system then produces recommendations for the instructors and trainees.

Step 3: Prescription

According to the quadrant where an individual trainee is, the knowledge gap of each trainee can be identified and appropriate feedbacks can be given for trainees. The instructor can provide more solid advice based on his experience and recommendations from the system.

4.6. Discussion

Depending on the behavior of an individual trainee, we suggest the following guidelines to improve the learning outcomes. For the 'uninformed' trainees, it appears that the trainees cannot fully understand and absorb shooting knowledge to improve their skills. Therefore, they will lose their confidence gradually, because of their low shooting scores. In this situation, the instructors should gather the 'uninformed' trainees into a group, instruct them carefully with more patience and attention to enhance their confidence, and ask them to practice diligently to progressively improve their basic shooting knowledge and skills.

For the 'misinformed' trainees who are over-confident, instructors can adjust their mindset, find their weaknesses on shooting knowledge/skill, and improve their shooting knowledge/skill by repetitive practice during the training lesson. On the other hand, for the 'doubt' trainees, the instructors should take the initiative to acknowledge their correct knowledge/skill during training lessons in order to improve their self-confidence. Then their shooting precision and accuracy can be gradually improved to increase their shooting stability. Finally, for a 'mastery' trainee, instructors still can refine their knowledge/skill and confidence toward a higher expert level, such as sniper. Furthermore, the 'mastery' trainee can be a role model for the other three types of trainees. Instructors may empower the "mastery" trainees to assist in training less capable trainees, thereby consolidating their knowledge, skills, and self-efficacy.

5. Conclusions

This paper integrated a data mining approach with the theory of confidence-based learning to improve training evaluation for SBT. Data mining techniques were used to analyze data generated from SBT to assess trainees' training performance and learning behaviors. The proposed methodology was illustrated with an example of a real case of Taiwan's infantry marksmanship training. The results show that the proposed methodology can accurately evaluate trainees' performance and their learning behaviors, and discover latent knowledge for improving trainees' learning outcomes. Future research will integrate expert systems with the developed training evaluation system that can provide more comprehensive feedbacks to trainees for improving their learning outcomes. In addition, the developed data mining methodology can be applied to other application domains, such as: management education.

Acknowledgements

This research is supported by Grant No. XW98166 P152 PE from the National Chung-Shan Institute of Science & Technology of the Republic of China. We also want to thank anonymous referees for their constructive comments on this paper.

References

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria. *Personnel Psychology, 42*(2), 331–342.
Bell, B. S., Kanar, A. M., & Kozlowski, S. W. J. (2008). Current issues and future directions in simulation-based training in North America. *International Journal of Human Resource Management, 19*(8), 1416–1434.
Bell, B. S., & Kozlowski, S. W. J. (2002). Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology, 55*, 267–307.
Berry, M., & Linoff, G. (2004). *Data mining techniques: For marketing, sales, and customer relationship management* (2nd ed.). New York: John Wiley & Sons.
Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984a). *Classification and regression trees*. Belmont, CA: Wadsworth.
Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984b). *Classification and regression tree*. CA, U.S.A.: Wadsworth.
Bruno, J. (1993). Using testing to provide feedback to support instruction: A reexamination of the role of assessment in educational organizations. In D. Leclercq & J. Bruno (Eds.). *Item banking: Interactive testing and self-assessment* (Vol. 112, pp. 190–209). Berlin Heidelberg: Springer.
Buckley, R., & Cape, J. (1990). *The theory and practice of training*. London: Kogan Page.
Cannon-Bowers, J. A., & Bowers, C. A. (2009). Synthetic learning environments: On developing a science of simulation, games, and virtual worlds for training. In S. W. J. Kozlowski & E. Salas (Eds.). *Learning, training, and development in organizations* (pp. 229–261). Routledge.
Chen, Z. Y., Fan, Z. P., & Sun, M. H. (2012). A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research, 223*(2), 461–472.
Eseryel, D. (2002). Approaches to evaluation of training: Theory & practice. *Educational Technology & Society, 5*(2), 93–98.
Ferreiro, S., Sierra, B., Irigoien, I., & Gorritxategi, E. (2011). Data mining for quality control: Burr detection in the drilling process. *Computers & Industrial Engineering, 60*(4), 801–810.
Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182.
Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques* (2nd ed.). Morgan Kaufmann.
Hartigan, J. A., & Wong, M. A. (1979). A K-means clustering algorithm. *Applied Statistics, 28*(1), 100–108.
Haykin, S. (1999). *Neural networks: A comprehensive foundation*. Prentice Hall.
Hirose, Y., Yamashita, K., & Hijiya, S. (1991). Back-propagation algorithm which varies the number of hidden units. *Neural Networks, 4*, 61–66.
Holton, E. F., III, Bates, R. A., & Ruona, W. E. A. (2000). Development of a generalized learning transfer system inventory. *Human Resource Development Quarterly, 11*(4), 333–360.
Hosmer, D., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.
Hunt, D. P. (2003). The concept of knowledge and how to measure it. *Journal of Intellectual Capital, 4*(1), 100–113.
Issenberg, S. B., Gordon, M. S., Gordon, D. L., Safford, R. E., & Hart, I. R. (2001). Simulation and new learning technologies. *Medical Teacher, 23*(1), 16–23.
Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Lee, G. D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Medical Teacher, 27*(1), 10–28.
Jeffries, P. (2005). A framework for designing, implementing, and evaluation simulations used as a teaching strategies in nursing. *Nursing Education Perspectives, 26*(2), 96–103.
Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels* (2nd ed.). San Francisco: Berrett-Khoeler.
Kwon, Y., Omitaomu, O. A., & Wang, G. N. (2008). Data mining approaches for modeling complex electronic circuit design activities. *Computers & Industrial Engineering, 54*(2), 229–241.
Lim, T. C. (2013). Review of data mining methodologies for healthcare applications. *Journal of Medical Imaging and Health Informatics, 3*(2), 288–293.
Lim, T. S., Loh, W. Y., & Shih, Y. S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning, 40*(3), 203–228.
Magro, M. C., & Pinceti, P. (2009). A confirmation technique for predictive maintenance using the rough set theory. *Computers & Industrial Engineering, 56*(4), 1319–1327.
McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Education, 44*(1), 50–53.
Moore, D., Green, J., & Gallis, H. (2009). Achieving desired results and improved outcomes: Integrating planning and assessment throughout learning activities. *Journal of Continuing Education in the Health Professions, 29*(1), 1–15.
Phillips, J. J. (1997). *Handbook of training evaluation and measurement methods*. Boston, MA: Butterworth-Heinemann.
Phillips, J. J. (2003). *Return on investment in training and performance improvement programs*. Routledge.
Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research, 20*(2), 134–148.
Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*, 81–106.
Quinlan, R. J. (1992). *C4.5: Programs for machine learning*. Morgan Kaufmann.
Quinlan, J. R. (1993a). *C4.5: Programs for machine learning*. San Francisco: Morgan Kaufmann.
Quinlan, R. (1993b). *C4.5: Programs for machine learning*. Los Altos, CA: Morgan Kaufmann.
Rivals, I., & Personnaz, L. (2003). Neural-network construction and selection in nonlinear modeling. *IEEE Transactions on Neural Networks, 14*, 804–819.
Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics, 23*(19), 2507–2517.
Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*(1), 471–499.
Salas, E., Wildman, J. L., & Piccolo, R. F. (2009). Using simulation-based training to enhance management education. *Academy of Management Learning & Education, 8*(4), 559–573.
Salas, E., Wilson, K. A., Burke, C. S., & Priest, H. A. (2005). Using simulation-based training to improve patient safety: What does it take. *Journal on Quality and Patient Safety, 31*(7), 363–371.
Santos, A., & Stuart, M. (2003). Employee perceptions and their influence on training effectiveness. *Human Resource Management, 13*(1), 27–45.
Spitzer, D. R. (1999). Embracing evaluation. *Training, 36*(6), 42–47.
Stevens, C. K., & Gist, M. E. (1997). Effects of selfefficacy and goal-orientation training on negotiation skill maintenance: What are the mechanisms? *Personnel Psychology, 50*(4), 955–978.
Summerhill, E. M., Mathew, M. C., Stipho, S., Artenstein, A. W., Jagminas, L., Russo-Magno, P. M., et al. (2008). A simulation-based biodefense and disaster preparedness curriculum for internal medicine residents. *Medical Teacher, 30*(6), 145–151.
Usher, J. M., & Strawderman, L. (2010). Simulating operational behaviors of pedestrian navigation. *Computers & Industrial Engineering, 59*(4), 736–747.
Ward, J. H. Jr., (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*, 236–244.
Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Morgan Kaufmann.
Wong, B. K., Bodnovich, T. A., & Selvi, Y. (1997). Neural network applications in business: A review and analysis of the literature (1988–1995). *Decision Support Systems, 19*(4), 301–320.
Yen, Y. C., Ho, R. G., Chen, L. J., Chou, K. Y., & Chen, Y. L. (2010). Development and evaluation of a confidence-weighting computerized adaptive testing. *Educational Technology & Society, 13*(3), 163–176.