Data-Driven Intelligent Tutoring System for Accelerating Practical Skills Development. A Deep Learning Approach

Robert Marinescu-Muster, Sjoerd de Vries, and Wouter Vollenbroek

Abstract Our data-driven intelligent tutoring system presents promising results in supporting and accelerating the skills acquiring process. For example, mapping of the common latent variables enables the instructors and curricula designers to understand better the relationships between different exercise items and thus to create improved training scenarios. The case study results also reveal significant improvements in accelerating the process of training welders: participants gradually started to improve their welding skills after only 15 trials (approximately 1 hour of training using the system).

Keywords Intelligent tutoring systems · Deep learning · Practical skills mastery · Data-driven education module

1 Introduction

In the last years, the advancements in computing capabilities coupled with increased availability of huge volumes of data and deeper knowledge on computer algorithms, revitalized artificial intelligence (AI) as a key component in various domains, from automotive industry to medicine. In itself, artificial intelligence is an umbrella term describing machines that simulate intelligence by performing large amount of computations at very high speeds. For instance, a self-driving car is basically a computer program which learns about the environment by detecting patterns from large sets of images containing vehicles, people, traffic signs, etc. It is then able to "see and understand" data from sensors by calculating in real time the probabilities that an object on the road fits in one of the learned categories: vehicles, pedestrians, traffic signs, etc.

The ability to uncover patterns from past information and to recognize them in new data made artificial intelligence extremely useful in education. Predicting students

R. Marinescu-Muster (✉) · S. de Vries · W. Vollenbroek
University of Twente, Drienerlolaan 5, 7500AE Enschede, The Netherlands
e-mail: r.marinescu-muster@utwente.nl

Ó. Mealha et al. (eds.), *Ludic, Co-design and Tools Supporting Smart Learning Ecosystems and Smart Education*, Smart Innovation, Systems and Technologies 197,
https://doi.org/10.1007/978-981-15-7383-5_17

achievements, knowledge tracing, automatic evaluation of students' answers in open questions, or intelligent tutoring systems are just a few applications of AI in the education field. In particular, intelligent tutoring systems (ITS) powered by AI have shown impressive results in decreasing training time and improving education quality [1, 2].

Therefore, there was a vast interest in the last three decades for establishing intelligent tutoring systems, designed specifically to support cognitive skills development such as mathematics, physics, and chemistry.

However, little research was done recently regarding the development and implementation of ITS for supporting practical skills mastery. This is in spite of the fact that early studies have shown that the length of learning sessions is reduced, while the quality of skill development is improved considerably when practical training is supported by ITS [3]. For instance, Lesgold et al. [4] showed that 20 h of training using Sherlock system (a ITS for detecting plane breakdowns) was equivalent to 4 years of experience. Other studies in this area showed significant improvements in acquiring rhythm and pitch skills with piano [5], significant decrease of required instructor need and increased learning quality for power plant operator trainees [6], and promising results from improving welding skills by using virtual reality space as a training method [7]. In general, intelligent tutoring systems designed for practical skills development are seen as highly beneficial, especially for addressing the need for highly skilled workforce [3, 8]. The gap in research and development of intelligent tutoring systems for professional practical skills mastery is apparent, most of the research being done in the late nineties and early two thousands. Julian and Smith [9] also noted recently about the lack of research and development of intelligent tutoring systems for medical training, specifically in procedural and practical skills (i.e., training for robotic assisted laparoscopic surgery). We believe that there is a need for revitalizing research on this topic because the current state of the art in artificial intelligence allows for new insights in this field.

This study aims at bridging this gap by proposing a data-driven method using deep learning algorithms for designing an intelligent tutoring system to support practical skills development. The paper is structured as follows: In the next section, we present our approach using recurrent neural networks (a class of deep learning unsupervised artificial intelligence neural networks [10]) for analyzing trainee activity and uncovering relevant patterns of performance. Then, we describe a data-driven monitoring infrastructure of an intelligent tutoring system and outline the results of our approach in a case study of a deep learning training module for shielded metal arc welding (SMAW) intelligent tutoring system-based training. We conclude with a short reflection on future perspectives of such approach.

2 Data-Driven Deep Learning ITS Framework

The main goal of an intelligent tutoring system is to support both students and instructors in achieving higher learning quality. The benefits of ITS as support for the traditional tutoring methods are related mainly to the possibility of continuous student activity data processing for uncovering relevant patterns of performance.

2.1 Data-Driven ITS

In the context of practical skills mastery, the generated data stream is usually high-dimensional and unstructured. Activity data is typically collected through a matrix of various sensors which include motion sensors, cameras, wearables, etc. Processing such amount of data is often one of the common deterrents in developing advanced ITS targeting practical and procedural skills mastery. Additionally, the typical machine learning approach on performance pattern recognition requires supervised machine learning processes in which data needs to be classified *a priori* and all relevant features need to be provided to the system. This is often a tedious process which requires extensive collaboration between instructors, researchers, and ITS developers. Moreover, it also runs the risk of omitting relevant features which describe trainee's learning model.

There are specific requirements and limitations of traditional tutoring methods that make ITS extremely beneficial:

- First, the instruction time is often long because of trial-and-error practice cycle, especially for novices;
- Second, the experiential nature of the practical tasks induces a limitation on the number of trainees an instructor can train at a given time;
- Third, the amount of feedback loops is inversely proportional with the group size,as a result, instructors often focus on one particular group of trainees in detriment of the others (e.g., focusing on the ones who make more errors, either on the ones who make the most progress);
- Fourth, the costs of traditional practical training are extremely high, especially in industrial or medical contexts.

Effective intelligent tutoring systems typically address most of these problems by continuously monitoring the trainee activity, provide instant feedback and correction methods and provide instructors with detailed performance dashboards of their trainees.

The key aspects of effective ITS are personalized feedback resembling one-to-one tutoring and the ability to generate personalized learning paths for optimizing the instruction time and the learning outcome [11, 12]. In order to achieve these, ITS require continuous access to data streams.

2.2 Recurrent Neural Networks

The advent of recurrent neural networks (RNN) offers promising advantages as compared to traditional machine learning algorithms.

First, RNNs are a chained set of artificial neurons in which information is propagated recursively over time. This makes RNNs extremely suitable for analyzing sequential data.

Second, in RNNs, the hidden layers of the network develop recursively on both the system input and on their previous state. This makes RNNs suitable for learning complex patterns that require also a form of memory over time [13], i.e., retaining previous states and either use it or forget it when needed. This is a particular class of RNN named Long Short-Term-Memory (LSTM).

Third, RNNs perform particularly well in fitting continuous, high-dimensional time series (or sequential) data as input and predicting outcomes at later points in time. This renders RNNs as state-of-the art algorithms in speech recognition and translation [14], facial recognition [15], medical diagnostics [16, 17], or knowledge tracing in cognitive skills ITSs [18].

Building on these advantages, we developed a data-driven RNN framework as a basis for an intelligent tutoring system designed to accelerating the instruction time for mastering practical skills. We propose the LSTM variation of RNNs to predict trainees' performance in future trials based on their previous activity.

2.3 Deep Learning Framework

At the core of the proposed framework lies the ability to detect performance patterns from unstructured activity data collected by sensors in past training trials and to predict the performance in the next trial. In general, if the predicted outcome is over a particular threshold (e.g., 80%) in the next trial, the system advises advancing to the next practical exercise. Conversely, if the predicted performance falls below the threshold, then the system advises repeating particular exercises linked to the actual performance.

There are four main outcomes of our proposed model:

1. map the exercise relationships—discovering the relationship between learning items can provide information about the latent structure of learning concepts related to practical skills, allowing clustering them based on direct influence between items;
2. generate a personalized skills training curricula—generating the optimal path for accelerated skill acquiring and stabilization of skills;
3. predict the probability of a failure in the next trial—given that the parameters remain constant, the model predicts the probability of failing to achieve a minimum success ratio in the next trial;

4. develop a personal attribution model of practical skills—each trainee has its own personal style, which becomes apparent from the clustering of the main parameters across trials. This can reveal valuable insights for optimizing the practical skills training path based on the continuous follow-up on the evolution of such clusters. For instance, it can be seen that two parameters are directly influenced. This means that when the trainee makes mistakes in one parameter, the other will also be affected negatively. Such information can therefore provide personalized training advices by the instructor.

LSTM Model

The model is designed from the assumption that there are three stages of acquiring practical skills [4]:

0—initial state, no skills are yet acquired. In this stage, the model offers a set of minimum set of assessment trials for both basic skills acquiring and assessment of the initial knowledge state.

1—the trainee acquired skills derived from practice in blocked batches—namely the trainee has developed specific patterns related to particular situations (through repetition). The model can use the data to predict trainee performance on a set of competencies (clustered skills) and to offer a personalized optimal path for accelerating the skill development and sustaining retention and transfer.

2—the trainee is able to perform in novel, complex and difficult situations without dropping performance. Our model can develop a curriculum for skill retention and transfer. For instance, it increases the complexity and difficulty of the tasks for continuing the skill development (i.e., adjusting the thresholds of the parameters, change the exercise type, add more difficult items, etc.) (Fig. 1).

The probability of a failure in the next trial is given by the equation:

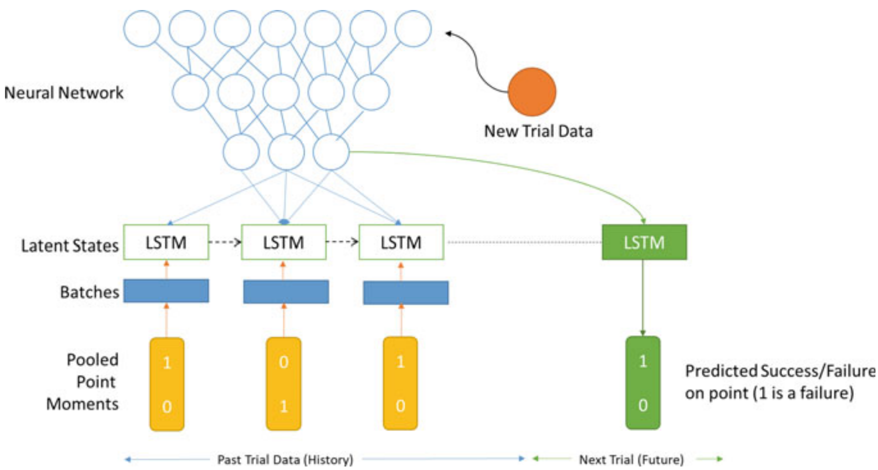

Fig. 1 LSTM adaptive deep learning model

$$P(y|x_{1:n}) = P\big(\text{nnet}_y(\text{pool}\{\text{LSTM}x_{1:n}\})\big) \tag{1}$$

The LSTM cell has three gates which control information flow. First gate is the forget gate—this part uses a sigmoid function to decide whether to keep or not the previous cell state (h_{t-1}). Next, the cell decides which information to store in the cell state using a sigmoid function (the input gate) and then using a tanh function to generate a new vector of values. Then, a pointwise multiplication is applied to update the cell state. Finally, the output gate filters the information by using a sigmoid layer (outputs 1/0) and then pushing the output through a tanh function (the output in $-$1/1). These can be expressed mathematically by the following equations:[1]

Forget Gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_t) \tag{2}$$

Input Gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$\widetilde{C}_t = \tanh\big(W_C \cdot [h_{t-1}, x_t] + b_C\big) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{5}$$

Output Gate:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{6}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{7}$$

There are three main steps of training the LSTM model:

1. **Data preparation**—input data is structured as fixed length vectors, and the values are normalized using MinMax normalization method [19]. Error events are computed based on specific skill performance requirements (e.g., if the observed values reach below an accepted predefined tolerance).
2. **Parameter initial optimization**—stochastic gradient descent optimized for accuracy and loss minimization. Several dropout ratios are applied in successive

[1] Parameter description

W_f = a matrix of learned weights connecting input neurons to hidden layers (h_t);
h_{t-1} = the previous hidden state
x_t = the input vector
t = timestamp
$b_{(.)}$ = scaling factor
C_t = cell state, $\widetilde{C}_t$ = candidate values vector
i_t = input vector

neuron layers to prevent overfitting. To prevent gradient explosion during back-propagation, we propose gradient clipping thresholds for truncating the gradients if they overpass the threshold.

3. **Hyper-parameter optimization—**needs to be performed based on particular architecture of the model and the data set. We recommend at least 128 hidden units and 200 epochs, in line with findings presented by Piech et al. [18].

3 Data-Driven Infrastructure of a ITS

In the practical skills mastery contexts, data is generated from various sensors in real-time or at a high-speed frequency. This puts immense stress on the computing architecture that forms the backbone of the intelligent tutoring system. Considering also the high-dimensionality and the inherent unstructured nature of the data (i.e., sensors can go offline or new sensors can be added based on the training tasks), the system needs to be designed flexible-first. Our data-driven computing architecture for an ITS in practical skills training is as follows:

- sensor data is stored in a distributed file system as raw data files
- trainees; data and other relevant information are stored in SQL databases
- the data is processed in parallel clusters for speed and resilience
- the deep learning module is a standalone machine which communicates through APIs
- results are processed and then presented in an Web-based environment.

The picture in Fig. 2 illustrates the proposed ITS.

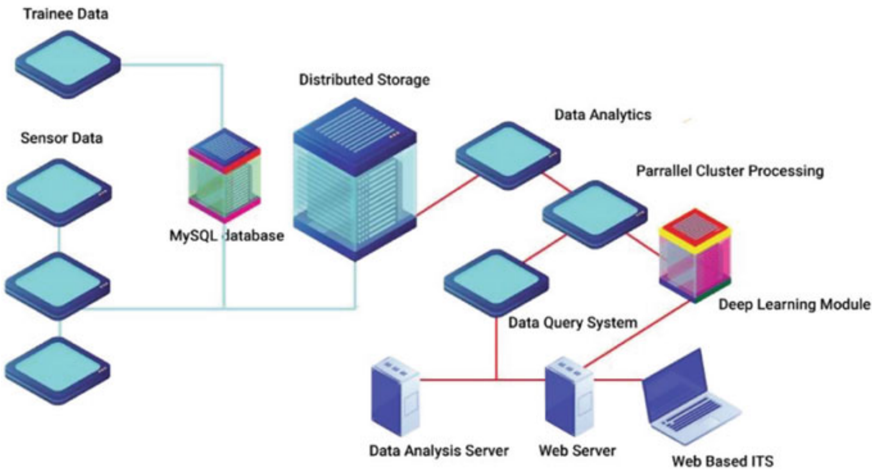

Fig. 2 Computing infrastructure for a data-driven ITS

4 Case Study—the Augmented Welder Data-Driven Deep Learning Intelligent Tutoring Module Application

Standard training method in shielded metal arc welding (SMAW) can be described as a linear learning model: The instructor sets up a batch of exercises in accordance with the level, duration of training and certification required. Typically, exercise items follow a linear path, from easy to difficult, and the welding scenarios are varied similarly. The instructor switches from trainee to trainee to evaluate their performance and provide corrections if needed.

The augmented welder (TAW) is a training tool designed by Institut de Soudure-IS (France) to support standard training methods and to speed up welding skills development for beginner welders and to improve the expertise of advanced professionals [20]. In 2019, we implemented our proposed intelligent tutoring system based on the deep learning framework as a data-driven education module in TAW.

4.1 Method

The goal of the module was threefold: first, to shorten the training time from absolute beginner to novice and advanced welder; second, to enable "hyper-leaps" on the linear learning process (that is, to enable trainees to jump directly to a more complex scenario if they prove sufficient mastery in the current skill level) and to provide advice in this respect; and third, to provide immediate feedback based on the current performance and to enable access to performance metrics to both instructors and trainees.

We used the data from TAW system collected during January–October teaching sessions in 2019 to train our deep learning model. We used a 80–20 training-validation split ratio on all the model variations employed. From the training partition, we kept data from three participants for testing purposes.

In the process of optimization, we included three versions of the model based on the trial windows to predict the performance of the trainee in the next exercise:

(a) using data from last three trials
(b) using data from last trial
(c) using data from all trials.

The model using last trial window as input data performed best in terms of accuracy and loss minimization (Accuracy $= 87\%$. $L = 0.1621$). The model using all trials performed the worst, 67% accuracy and a loss minimization result of 0.4451. One potential explanation is that using all trials data could introduce bias toward the beginner welders (thus being more prone to welding mistakes) because they were over-represented in the dataset (more than 50%). On the other hand, using only the last trial window made possible a more appropriate estimation of the ratio between mistakes and successes. We will explore these aspects in future trials.

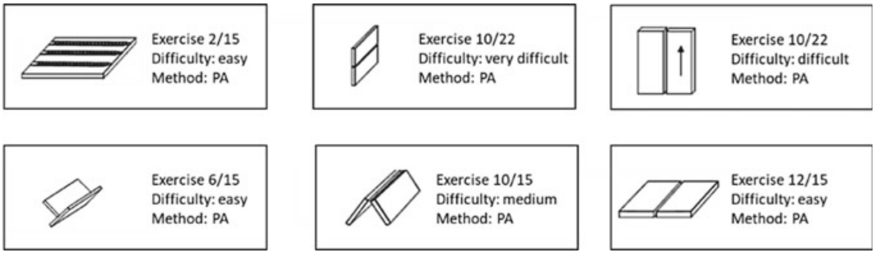

Fig. 3 Welding exercises used during trials

4.2 Dataset

The dataset for implementing the model consisted in welding results from 20 trainees (17 male trainees, and three female trainees) in 536 trials originating from six welding exercises. The exercises are presented in Fig. 3.

In total, there were around 31 h of trial data. Most of the participants used right hand as the dominant ($N = 14$ participants) and were beginners ($N = 11$), followed by five experts and four advanced welders. The average height of the trainees was 182 cm (SD = 4.1013 cm.), and their average weight was around 79 kg (SD = 6.6353 kg). Each welding session generated on average approximately 450 data points (at a frequency of 2 s) covering 11 different sensors (angles, speed, arc-height, etc.). The total data pool for training the deep learning model consisted in 139.620 valid observations.

4.3 Results

The deep learning data-driven module using last trial data window performed substantially better than the others presented in the method section ($\text{AUC}_{\text{LastTrial}} = 0.87$, $\text{AUC}_{\text{LastThreeTrials}} = 0.78$ and $\text{AUC}_{\text{AllTrials}} = 0.67$).[2]

The exercise relationship map is providing an overview of the influence between the exercises, moderated by the exercise difficulty and number of trials available in the database. The map is realized by the model from the relationships inferred from the available data. It can be seen that the most difficult exercises are having the most influence on the others and that exercise 10/15 have a considerable influence on the more difficult exercises. The model predicts that performing well in exercise 10/15 (medium difficulty) will likely result in a good performance on exercise 10/22 which is more difficult. Hence, it can advise the trainee to move to the more difficult exercise if sufficient mastery is attained instead of repeating the same exercise. The

[2] AUC = area under the receiver operating characteristic curve (ROC), representing the area under the discretized curve of precision versus recall values (estimating the probability of a binary outcome). More detailed explanations are available in [22].

model calculates the likelihood that exercise T influences exercise W, in the sense that performance on exercise T will influence the performance on exercise W (because of shared latent attributes). The exercise relationship mapped by the model is presented in Fig. 4.

The results of the model are available online to both trainees and instructors. This allows the trainees to investigate their performance and reply particular low performance moments. The instructors can evaluate the trainees on all or partial trials, can rate the performance, and has the possibility to evaluate the quality of the predictions of the deep learning ITS module. An excerpt of the data-driven feedback dashboard can be seen in Fig. 5. The upper right part of the figure illustrates the trial success predictions and the optimal training path for this trainee (Fig. 5).

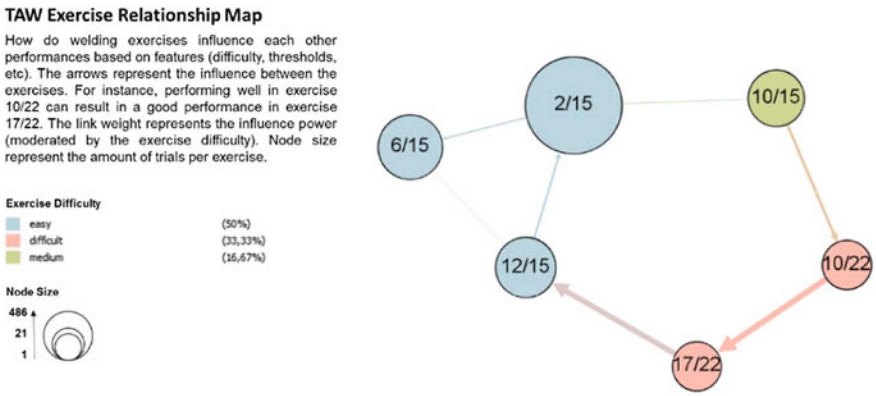

Fig. 4 Exercise relationship map

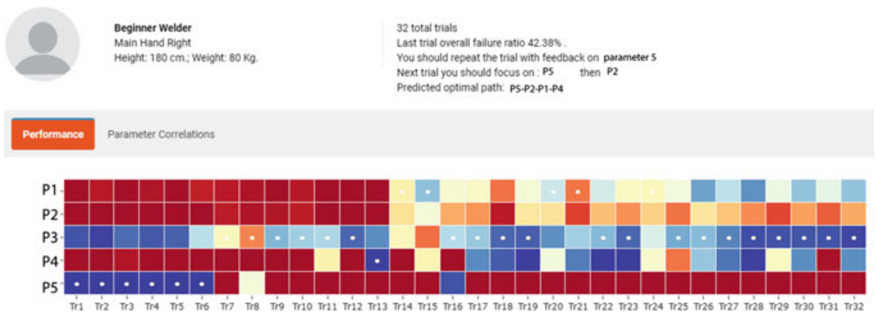

Fig. 5 Data-driven ITS feedback dashboard—trainee view. In the picture, red represents higher mistake ratio, while blue means lower. The training expertise gain is visible in the chart by the changing of colors on the majority of the parameters monitored (P1, P3, P4, and moderately improved on P2) after Trial 13

5 Discussion

In this paper, we introduced a full-fledged deep learning framework for developing intelligent tutoring systems in practical skills mastery, together with a case study presentation of a pilot implementation. We built our approach to address the need for developing intelligent tutoring systems for practical skills mastery, which are extremely important in training and continuous professionalization of skilled professionals.

The system presents some very promising results. First, the mapping of the common latent variables enables the instructors and curricula designers to understand better the relationships between different exercise items and thus to create improved training scenarios. Second, the system allows for continuous data collection and is designed with flexibility at the heart. Last but not least, the data-driven intelligent tutoring system based on deep learning algorithms removed the need for labeled data. As such, there is no need for the instructors to pre-evaluate and label the trials used to train the models. Instead, the LSTM is able to work with any input that can be expressed as a sequential vector.

One limitation of the deep learning framework is the need for substantially large amount of training data. Therefore, it can be suitable in setups which have the possibility of generating large volumes of data (e.g., industry, medical) but not in small groups. We should also consider the interpretability of such recurrent neural network models as an important limitation, especially in the context of predicting human skills acquisition based on past performance where specific temporal events are important. One way to address this is to use a combination of attention and post-analysis methods as suggested in [21].

The implementation of artificial intelligence in developing intelligent tutoring systems for practical skills mastery leaves many directions for future research and development. For instance, further research can focus on different industrial processes which require similar training facilities. Similarly, another direction can explore the impact of such tools in developing continuous professional education programs, for accelerating the mobility from beginner to advanced users.

We continue the collaboration with Institut de Soudure for deepening our understanding of the proposed framework and for intensive validation of the model.

6 Conclusion

Our data-driven Intelligent Tutoring System presents promising results in supporting and accelerating the skills acquiring process. For example, mapping of the common latent variables enables the instructors and curricula designers to understand better the relationships between different exercise items and thus to create improved training scenarios. The case-study results also reveal significant improvements in accelerating the process of training welders: participants gradually started to improve their welding skills after only 15 trials (apprx. 1 hour of training using the system). The data-driven module is a critical module on the TAW platform and is expected to open advanced opportunities for welding training and professional welding. We consider it as an example of a “data-driven skills training platform.” The design is expected also to be applied in other professional skill set domains. Our innovative analytics approach, the distinction between learning, design, and quality analytics, has proven to be applicable and practical.

References

1. Gutierrez, F., Atkinson, J.: Adaptive feedback selection for intelligent tutoring systems. Expert Syst. Appl. **38**(5), 6146–6152 (2011). https://doi.org/10.1016/j.eswa.2010.11.058
2. Ma, W., Adesope, O.O., Nesbit, J.C., Liu, Q.: Intelligent tutoring systems and learning outcomes: a meta-analysis, (2014). https://doi.org/10.1037/a0037123.supp
3. Frasson, C., Aïmeur, E.: Designing a multi-strategic intelligent tutoring system for training in industry. Comput. Ind. **37**(2), 153–167 (1998). https://doi.org/10.1016/s0166-3615(98)00091-8
4. Lesgold, A., Lajoie, S., Bunzo, M., Eggan, G.: Sherlock: a coached practice environment for an electronics troubleshooting job. (1988)
5. Chan, L.M.Y., Jones, A.C., Scanlon, E., Joiner, R.: The use of ICT to support the development of practical music skills through acquiring keyboard skills: a classroom based study. Comput. Educ. **46**(4), 391–406 (2006). https://doi.org/10.1016/j.compedu.2004.08.007
6. Gutierrez, J., Elopriaga, J.A., Fernandez-Castro, I., Vadillo, J.A., Diaz-Ilarraza, A.: Intelligent tutoring systems for training of operators for thermal power plants. Artif. Intell. Eng. **12**(3), 205–212 (1998). https://doi.org/10.1016/S0954-1810(97)00015-0
7. Sakata, S., Mizuno, S.: Proposal of a welding skill training system using VR technology. In: International Workshop on Advanced Image Technology (IWAIT), vol. 11049, p. 146 (2019). https://doi.org/10.1117/12.2521636
8. Wessner, C.W., Howell, T.R.: Educating and training a high-tech workforce, pp. 217–276. Springer, Cham (2020)
9. Julian, D., Smith, R.: Developing an intelligent tutoring system for robotic-assisted surgery instruction. In: The International Journal of Medical Robotics and Computer Assisted Surgery, vol. 15, no. 6, Dec. 2019. https://doi.org/10.1002/rcs.2037
10. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**(6088), 533–536 (1986). https://doi.org/10.1038/323533a0
11. Phobun, P., Vicheanpanya, J.: Adaptive intelligent tutoring systems for e-learning systems. Procedia Soc. Behav. Sci. **2**(2), 4064–4069 (2010). https://doi.org/10.1016/j.sbspro.2010.03.641
12. Kulik, J.A., Fletcher, J.D.: Effectiveness of intelligent tutoring systems. Rev. Educ. Res. **86**(1), 42–78 (2016). https://doi.org/10.3102/0034654315581420
13. Williams, R.J., Zipser, D.: A learning algorithm for continually running fully recurrent neural networks. Neural Comput. **1**(2), 270–280 (1989). https://doi.org/10.1162/neco.1989.1.2.270
14. Passricha, V., Kumar Aggarwal, R.: Convolutional neural networks for raw speech recognition. In: From Natural to Artificial Intelligence—Algorithms and Applications, IntechOpen (2018)
15. Jain, D.K., Shamsolmoali, P., Sehdev, P.: Extended deep neural network for facial emotion recognition. Pattern Recogn. Lett. **120**, 69–74 (2019). https://doi.org/10.1016/j.patrec.2019.01.008
16. Blom, M., Nobile, N., Suen, C.Y., Xi, P., Goubran, R., Shu, C.: Cardiac murmur classification in phonocardiograms using deep recurrent-convolutional neural networks. In: Frontiers in Pattern Recognition and Artificial Intelligence, World Scientific, pp. 189–209 (2019)

17. Roy, I., Kiral-Kornek, I., Harrer, S.: Chrononet: A deep recurrent neural network for abnormal EEG identification. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11526 LNAI, pp. 47–56 (2019). https://doi.org/10.1007/978-3-030-21642-9_8
18. Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L.J., Sohl-Dickstein, J.: Deep knowledge tracing. In: Advances in neural information processing systems, pp. 505–513 (2015)
19. Gopal, S., Patro, K., Kumar Sahu, K.: Normalization: a preprocessing stage, arXiv preprint arXiv: 1503.06462 (2015)
20. The Augmented Welder Digital Industry. A powerful tool to speed up training and improve practice in welding (2019). Available online at https://www.eitdigital.eu/fileadmin/files/2018/factsheets/digital-industry/TheAugmentedWelder-Factsheet.pdf. Accessed on 22-01-2020
21. Guo, T., Lin, T., Antulov-Fantulin, N.: Exploring interpretable lstm neural networks over multi-variable data, arXiv preprint arXiv: 1905.12034 (2019)
22. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006). https://doi.org/10.1016/J.PATREC.2005.10.010