

Intelligent diagnostic feedback for online multiple-choice questions

R. Guo · D. Palmer-Brown · S. W. Lee · F. F. Cai

Published online: 9 August 2013
© The Author(s) 2013

Abstract When students attempt multiple-choice questions (MCQs) they generate invaluable information which can form the basis for understanding their learning behaviours. In this research, the information is collected and automatically analysed to provide customized, diagnostic feedback to support students' learning. This is achieved within a web-based system, incorporating the snap-drift neural network based analysis of students' responses to MCQs. This paper presents the results of a large trial of the method and the system which demonstrates the effectiveness of the feedback in guiding students towards a better understanding of particular concepts.

Keywords Learning behaviour · Diagnostic feedback · Neural networks · On-line multiple-choice questions

1 Introduction

In recent years, e-learning has become commonplace in higher education. The involvement of intelligent e-learning systems has the potential to make higher education accessible with increasing convenience, efficiency and quality of study. According to the [Hefce National Student Survey \(2007–2010\)](#), in England, only about half of students believe that: (1) feedback on their work has been prompt; (2) feedback on their work has helped to clarify things

R. Guo (✉) · D. Palmer-Brown · S. W. Lee · F. F. Cai
Faculty of Life Sciences and Computing, London Metropolitan University,
Holloway Road, London N7 8DB, UK
e-mail: r.guo1@londonmet.ac.uk; ruisheng_g@hotmail.com

D. Palmer-Brown
e-mail: d.palmer-brown@londonmet.ac.uk

S. W. Lee
e-mail: s.w.lee@uel.ac.uk

F. F. Cai
e-mail: ff.cai@londonmet.ac.uk

they did not understand; (3) they have received detailed comments on their work. These reports reveal that the feedback and its related fields are one of the weakest areas in higher education in England. This research investigates the relative effectiveness of different types of feedback, and how to optimize feedback to facilitate deep learning. It compares and contrasts several methods in order to investigate the effectiveness of using intelligent feedback towards modeling the stages of students' knowledge. The investigation will lead to an understanding of the potential of the on-line diagnostic feedback across different subject areas.

The virtual learning environment (VLE) as an e-learning system based on web which can model personal education by providing virtual access to classes, concepts, key learning point, tests, homework, grades, assessments, and other external resources. It can also provide a method to help students to enhance their learning experiences, and to help teachers manage the gap between teaching and learning. The VLE presented in this paper provides a generic method for intelligent analysis and grouping of student responses that applicable to any area of study. This tool offers important benefits: immediate feedback, significant time-saving evaluating assignments, and consistency in the learning process. The time taken to create the feedback is well spent not only because this feedback can be reused, but also it is made available through the system to large numbers of students.

2 Background and review of previous work

Ma (2006) described intelligent tutoring systems (ITSs) as the milestone of the advanced generation of computer-aided instruction systems, and concluded their key feature as 'the ability to provide a user-adapted presentation of the teaching material'. Rane and Sasikumar (2007) pointed out that to overcome the lack of the presence of a teacher, intelligent tutoring systems attempt to simulate a teacher, who can guide the student's study based on the student's level of knowledge by giving intelligent instructional feedback. Furthermore, according to Blessing et al. (2007), the intense interaction and feedback achieved by intelligent tutoring systems can significantly improve student learning gains. In addition, in Gheorghiu and Vanlehn (2008)'s paper, they also suggested that meaningful, constructive and adaptive feedback is the essential feature of ITSs, and it is such feedback that helps students achieve strong learning gains.

Cullen et al. (2002, cited by Heinze et al. 2007) suggest that feedback is one of the most important indicators of good education. Little (2001) also presents that providing feedback, which can continuously reinforce learning progress and promote learners' attention and engagement, is crucial to effective learning. Besides, Vasilyeva et al. (2008) also state that 'the design of feedback is a critical issue of online assessment development within web-based learning systems'. However, reviewing the education reality, it is surprising that as one of the most important communication methods between teachers and students, feedback was one of the weakest areas according to the Hefce National Student Survey (2007–2010) in the UK. Thus, how to design an optimized intelligent feedback system becomes a major and critical problem to develop a successful e-learning system.

To make the feedback effective and meaningful, a range of quality attributes need to be achieved. Hatziapostolou and Paraskakis (2010) summarized the work by Race (2006), Irons (2008) and Juwah et al. (2004) and suggested that in order to improve learning gains, formative feedback should address as many as possible of the following attributes, including constructive, motivational, personal, manageable, timely and directly related to assessment criteria and learning outcomes. (1) Constructive. Effective feedback should be constructive,

as constructive feedback can lead to more thinking and cognitive learning which in turn improve the student's learning (Bang 2003 and Alessi and Trollip 2001). As Nelson and Schunn (2009) argued, effective feedback should be able to guide a learner 'to change performance in a particular direction rather than just towards or away from a prior behavior'. (2) Motivational. Effective feedback should be motivational to empower and encourage students to learn more, as feedback can affect students' feelings and attitudes towards study, which in turn affect their engagement in the learning process, (Jawah et al. 2004, cited by Hatziapostolou and Paraskakis 2010). (3) Garber (2004) also argued, 'the more personalized the feedback becomes, the more meaning it can have for the individual receiving it. . . and the more likely the individual will be receptive to the feedback. . . If a person does not believe in the reliability or validity of the feedback, it will have little or no benefit'. (4) Manageable. Effective feedback should be detailed enough to ensure that students clearly understand their strengths and weaknesses, and have enough materials to guide them to achieve the learning goals. At the same time, the feedback should not be over-detailed to avoid confusing, and make students can easily interpret it and get the point (Hatziapostolou and Paraskakis 2010). (5) Timely. Feedback should be delivered timely, as students can more easily utilize feedback when they can still remember how they just processed the task (Race 2006), and the reasoning that led to the error is still accessible (Reiser and Kimberg 1992). As Anderson, Boyle and Reiser (1985, cited by Reiser and Kimberg 1992) argued, tutors should provide immediate feedback to students, as 'the learning mechanism for adjusting a faulty rule or forming a new correct rule relies upon the problem situation being active in memory'. (6) Directly related to assessment criteria/learning outcomes. Effective feedback should be directly related to assessment criteria/learning outcomes so that it can explain students' achievement towards the intended learning outcomes, knowledge gaps and specific errors (Hatziapostolou and Paraskakis 2010). Thus, the students can be guided and adjust their effort to achieve the intended leaning outcomes (Race and Brown 2005). Additionally, Clark and Mayer (2009) further argued that learning goal oriented feedback is more effective than performance goal oriented feedback. In another word, feedback should be designed to inform the learners their progress toward achieving a learning goal rather than compare a learner performance with other learners'.

In addition to the required attributes above, many researchers also mentioned that various methods should be used in feedback to ensure better perception of feedback. For example, Özdener and Satar (2009) suggested using animation techniques to achieve better reception and perception of the feedback.

Springay and Clarke (2007) suggested including examples in feedback to achieve better perception of feedback. Multiple choice questions (MCQs) is an effective way to provide students with feedback. The use of multiple-choice questions has been widely studied. A number of advantages can be found in the Epstein et al. (2002), Higgins and Tatham (2003) and Kuechler and Simkin (2003) researches: rapid feedback, automatic evaluation, perceived objectivity, easily computed statistical analysis of test results and the reuse of questions from databases as required, thus saving time for instructors. On the other hand there are also some researches (e.g. Paxton 2000) shows that MCQs have some disadvantages: significant effort is required to construct MCQs, they only evaluate knowledge and recall, and they are unable to test literacy and creativity. Although the MCQs have been primarily used for summative evaluation, they also serve formative assessment purposes. Formative assessment provides students with feedback that highlights areas for further study and indicates the degree of progress. There are many studies investigating the role of different type of feedback in Web-based assessments that report positive results from the use of MCQs in online tests for formative assessments.

Many researches investigating the effect of different types of feedback in web-based assessments showed the positive results of using MCQs in online test for formative assessment (e.g. Epstein et al. 2002; Higgins and Tatham 2003; Kuechler and Simkin 2003; Payne et al. 2007). Higgins and Tatham (2003) studied the use of MCQs in formative assessment in a web-based environment using WebCT for a level one unit on undergraduate law degree. They summed that they could forecast all the possible errors for a question and write a general feedback for this question. However, using this type of feedback, it could be difficult to predict all the possible errors and produce the general feedback for a combination of questions, and it would be impossible for a large test banks (e.g. 3 questions with 5 answers would require 125 answer combinations; 5 questions with 5 answers require 3,125 combinations, etc.).

Payne et al. (2007) assessed the effectiveness of three different forms of feedback (corrective, corrective explanatory, and video feedback) used in e-learning to support students' learning. This type of feedback shows exactly which questions are answered correctly or not, with further corrective explanation and video feedback. Our approach to feedback is different from the above. The intelligent diagnostic feedback we present is concept-oriented instead of question-oriented. The learners are encouraged to review the concepts they misunderstood through the feedback in order to retake the test again and study further. It is important that each category of answers is associated with carefully designed feedback based on the level of understanding and prevalent misconceptions of that category-group of students so that every individual student can reflect on his or her learning level using this diagnostic feedback. In addition, when students retake the test they receive new feedback according to his or her knowledge state, which in turn leads to more self-learning. Moreover, concept-based feedback can also prevent the student from guessing the right answers; if the students do not read the diagnostic feedback carefully, they may not even know which questions were answered incorrectly. In conclusion the focus of our current research is to combine MCQs and formative online assessment using an intelligent agent to analyse the students' response in order to provide diagnostic feedback. To this end, we deploy a neural network

3 Multiple-choice questions online feedback systems (M-OFS)

To analyse the students' answers, and integrate over a number of questions to gain insights into the students' learning needs, a snap-drift neural network (SDNN) approach is proposed. SDNN provides an efficient means of discovering a relatively small and therefore manageable number of groups of similar answers. In the following sections, an e-learning system based on SDNN is described.

3.1 Snap-drift neural networks (SDNNs)

The learning process involves a combination of fast, convergent, minimalist learning (snap) and overall feature averaging (drift) to capture both precise sub-features in the data and more general holistic features. Snap and drift learning phases are combined within a learning system that toggles its learning style between the two modes. On presentation of input data patterns at the input layer, the distributed SDNN (dSDNN) learns to group them according to their features using snap-drift (Palmer-Brown and Jayne 2011). The neurons whose weight vectors result in them receiving the highest activations are adapted. Weights are normalised weights so that in effect only the angle of the weight vector is adapted, meaning that a

recognised feature is based on a particular ratio of values, rather than absolute values. The output winning neurons from dSDNN act as input data to the selection SDNN (sSDNN) module within performs feature grouping and this layer is also subject to snap-drift learning.

The learning process is unlike error minimisation and maximum likelihood for SDNN toggles its learning mode to find a set of sub-features and average feature in the data and uses them to group the data into categories. Each weight vector is bounded by snap and drift: snapping gives the angle of the minimum values (on all dimensions) and drifting gives the average angle of the patterns grouped under the neuron. Snap creates a feature common to all the patterns in the group and gives a high probability of rapid (in terms of epochs) convergence (both snap and drift are convergent, but snap is faster). Drifting, which uses learning vector quantization (LVQ), tilts the vector towards the centroid angle of the group and ensures that an average, generalised feature is included in the final vector. The angular range of the pattern-group membership depends on the proximity of neighbouring groups (natural competition), and can be controlled by adjusting a threshold on the weighted sum of inputs to the neurons.

3.2 Training neural network

The e-learning snap-drift neural network (ESDNN) is trained with the students' responses to questions on a particular topic in a course. The responses are obtained from the previous cohorts of students. Before training, each of the responses from the students is encoded into binary form in preparation for presentation as input patterns for ESDNN. Table 1 shows examples of a possible format of questions for five possible answers and some encoded responses. This version of ESDNN is a simplified unsupervised version of the snap-drift algorithm (Palmer-Brown and Jayne 2011) as shown in Fig. 1.

During training, on presentation of an input pattern at the input layer, the dSDNN will learn to group the input patterns according to their general features. In this case, 5 F12 nodes,

Table 1 Example of input patterns for ESDNN

Codification	A:00001	B:00010	C:00100	D:01000	N/A:00000
Response	Recorded response				
[C, D, B, A]	[0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1]				
[E, B]	[1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0]				
[D, A, A]	[0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1]				
[A, C, D, B, A]	[0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 10, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1]				

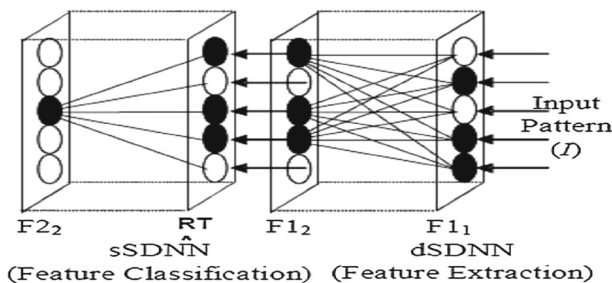


Fig. 1 E-learning SDNN architecture

whose weight prototypes best match the current input pattern, with the highest net input are used as the input data to the sSDNN module for feature classification. In the sSDNN module, a quality assurance threshold is introduced. If the net input of an sSDNN node is above the threshold, the output node is accepted as the winner; otherwise a new uncommitted output node will be selected as the new winner and initialised with the current input pattern. For example, for one group, every response might have in common the answer C to question 2, the answer D to question 3, the answer A to question 5, the answer A to question 6, the answer B to question 8, and the answer A to question 10. The other answers to the other questions will vary within the group, but the group is formed by the neural network based on the commonality between the answers to some of the questions (four of them in that case). From one group to another, the precise number of common responses varies in theory between 1 and X, where X is the number of questions. In this experiment, where there are 10 questions in 1st English trial, the groups had between 5 and 8 (Trial 1) common answers. More details of the steps that occur in ESDNN and the ESDNN learning algorithm are given in (Palmer-Brown and Jayne 2011). The training relies upon having representative training data. The number of responses required to train the system so that it can generate the states of knowledge varies from one domain to another. When new responses create new groups, more training data is required. Once new responses stop creating new groups, it is because those new responses are similar to previous responses, and sufficient responses to train the system reliably are already available. The number of groups formed depends on the variation in student responses.

3.3 How the system guides learning

The feedback is designed by academics so that it does not identify which questions were incorrectly answered. The academics are presented with the groups in the form of templates of student responses. Table 2 shows some examples of the group formed. For example, “A/D B mix” represents group 1 characterized by all the students answering A or D to question 1, B to question 2, and mixed answers to question 3. Hence, the educator can easily see the common mistakes in the groups of the student answers highlighted by the tool. The feedback texts are associated with each of the pattern groupings and are composed to address misconceptions that may have caused the incorrect answers common to that pattern group. The student responses, recorded in the database, can be used for monitoring the progress of the students and for identifying misunderstood concepts that can be addressed in subsequent face-to-face sessions. The collected data can be also used to analyze how the feedback influences the learning of individual students by following a particular student’s progress over time and observing how that student’s answers change after reading the feedback. Student responses can also be used to retrain the neural network and see whether refined groupings are created, which can be used by the educator to improve the feedback. Once designed, MCQs and feedbacks can be reused for subsequent cohorts of students.

Table 2 Example of answer groups

Group number	Group formed		
	Question 1	Question 2	Question 3
1	A/D	B	Mix
2	C	D	D
3	Mix	A	A/B/D

4 Experimental environment

In order to evaluate the performance and effectiveness of this novel e-learning system, abundant target-oriented testing needs to be carried out in different fields. Furthermore, we also aim to enhance this system to overcome its deficiencies during practical applications. Thus, this study is composed of three main parts. Firstly, we evaluated the system by collecting and analysing a large number of testing data reflecting the students' learning gains by using this system as well as the survey and interview data reflecting the students' satisfaction and attitudes towards this system. Secondly, the investigation will lead to an understanding of the potential of the on-line diagnostic feedback approach across different subject areas. Furthermore, this research should also produce guidelines for the design principles of on-line MCQs in the context of diagnostic feedback learning environments. Thirdly, we managed to enhance the existing system according to the evaluation data. The details of this experiment which are conducted to assess the use of M-OFS during academic year 2010–2011 are shown in next section.

4.1 Data collection and feedback generation

Meaning group have emerged and formed with small number of student with training data, not computational intensive scalability with large number of data. For the first trial presented in Sect. 5, the ESDNN was trained with the responses for 10 MCQs on English Grammar obtained from previous cohort of students. After training, appropriate feedback text was written by academics for each of the group of students' responses that address the conceptual errors implicit in combinations of incorrect answers. During the trial, a current cohort of students was asked to provide responses on the same questions, they were given the feedback on the combination of incorrect answers and their responses recorded in the database. The feedback texts are composed around the pattern groupings and are aimed at misconceptions that may have caused the incorrect answers common within the pattern group.

An example of a typical response to the questions below is (1 < D > 2 < B > 3 < A > 4 < B > 5 < B/D > 6 < A > 7 < A > 8 < C > 9 < A/C > 10 < B/D >)

- _____ no cause for alarm, the old man went back to his bedroom.
A. There was B. Since C. Being **D. There being**
- Even as a girl, _____ to be her life, and theater audiences were to be her best teachers.
A. performing by Melissa were.
B. it was known that Melissa's performances were
C. knowing that Melissa's performances were
D. Melissa knew that performing was
- Agriculture is the country's chief source of wealth, wheat _____ by far the biggest cereal crop.
A. is B. been C. be D. being
- This company has now introduced a policy _____ pay rises are related performance at work.
A. which **B. where** C. whether D. What
- She managed to save _____ she could out of her wages to help her brother.
A. how little money **B. so little money**
C. such little money D. what little money
- He left orders that nothing _____ touched until the police arrived here.
A. should be B. ought to be C. must be D. would be

7. As it turned out to be a small house party, we _____ so formally.
A. need not have dressed up B. must not have dressed up
 C. did not need to dress up D. must not dress up
8. I _____ the party much more if there hadn't been quite such a crowd of people there.
 A. would enjoy B. will have enjoyed
C. would have enjoyed D. will be enjoying
9. There ought to be less anxiety over the perceived risk of mountain climbing than _____ in the public mind today.
 A. exists B. exist **C. existing** D. to exist
10. Fat cannot change into muscle _____ muscle changes into fat.
 A. any more than **B. no more than** C. no less than D. much more than

This is classified into Group 6, which generates the following feedback:

Group 6 Feedback 1 Four points should be stressed. First, the logical subject of the adverbial phrase should agree with that of the main clause. Second, two verbs in a sentence need a conjunction. E. G. I am a teacher but you are a student. Third, the usage of various noun clauses should be familiar with. E. G. The news came that he died. (“that” does not serve as any part of the clause.) Fourth, some fixed structures in the comparative form should be memorized. E. G. not so...as...

In total data of six trials were collected (3 English trials, 2 Math trials, 1 Plagiarism awareness trial). In this paper, it presents the details of data collection of 2nd English trial. The data for training is collected from three previous year's MCQs tests. For these three tests, 94 students' answers were used to training. The trials data were collected during academic year 2010–2011. The data of two separate MCQ paper tests and final examination results were gathered. 83 students entered the survey and 16 students were randomly selected for interview. The states of knowledge of students were achieved by using ESDNN.

4.2 English experiments

To investigate and evaluate how the M-OFS guide and support students to learn, three English experiments were under taken by level 2 and level 3 students at JinQiao University (JQU) and Kunming Technology University (KTU) in China during the academic year 2010–2011. The 1st experiment is introduced below.

In the first experiment, data was collected from 148 students taking English language courses whom were randomly separated into two groups. The experimental group of 83 students used M-OFS, and the control group of 65 students received the same training but without using M-OFS. The system trial includes 10 MCQs with 4 potential answers, related to English grammar. The duration of this trial is flexible. When students were using M-OFS, they were encouraged to answer the MCQs (submit their answers) as many times as they wish until they got all the correct answers or gave up (students were not given answers or how many answers were correct in their feedback, except that they answered all correct answers). Two MCQ paper tests with different questions from system trials were applied to 116 students, and 83 students participated in both paper test and system trial. 83 students completed survey after second paper test. This system trial, paper test and survey were completed in practice lessons in a computer room at JQU.

5 Empirical study

This section discusses the results from the first experiment in order to evaluate the effectiveness of using M-OFS to support students' deep learning.

The survey and interview were conducted after the system trial. 83 (100%) students conducted the survey. 16 (19%) students were randomly chosen for interview. For the survey, 71.1% students are satisfied with using system. 84.4% students think the feedback is what they need. Using M-OFS to learn were positively evaluated by students, illustrate that the hypotheses H1 is supported. 90.4% students would like to use the system again. 92.8% students would like to recommend the system to a friend or classmate in the future. 81.9% students have never used similar system before. For interviews, most students (94%) feel this system is useful and helps them to improve their knowledge, it indicates the hypotheses H1 is supported as well; moreover, 69% students want the exact answers in the feedback in the end. Students also want a picture of their learning process which can point out their weakness and a suggestion of how to improve their English. Some students feel that if they tried many times but cannot find the correct answer, they will lose patience in the end.

5.1 Experiment and result

148 students are involved in the first experiment. 116 students completed the separate MCQs paper test before and after using the system. 83 students participated in system trial, and separate MCQs paper tests. For system trials, a total of 1,118 answers/attempts were submitted and a total 2,143 min were spent by 83 participants. All of the students submitted their answers at least once. The maximum number of attempts was 106 times and the minimum was 1. The average attempts for each student is 13.5 times. The average time spent by each student is 25.8 min and the average time of each attempt is 1.92 min. Two students (2.4%) spent more than 60 min. 35 (42.2%) students spent more than the average time. No students achieved the all correct answers at the beginning, 55 (66.3%) students increased their scores by an average of 12.8%, whilst one student increased his score by 70%. In this trial, with 10 questions and 4 possible answers, there are more than one million possible combinations of answers, thus the students are unlikely to make improvement by guessing answers; hence, the results show the feedback had a positive impact which partially supports hypotheses H2. For separate MCQs paper tests, the average score before system trial is 51.6%, and the average score after system trial is 59.1%. One student increased his score by 40%. 74% students increased their scores. In this test, the students were not given any answers or feedback between first (before system trial) and second (after system trial) test; furthermore, the first trial were applied 3h before the system trial and the second test were conducted 30min after system trial; hence, the students are only learnt by using M-OFS but not any other ways; thus the results above are confident, therefore partially supporting hypotheses H3. In addition, this result also can partially support hypotheses H2. For final examination, both the experimental group and the control group enter the same 4 days final examination. The experimental group got 79.52% and control group got 71.3% in English grammar module. This result confirms the hypotheses H4; furthermore, it also supports hypotheses H2.

5.2 Some group behavioural characteristics

Previous work has made an initial investigation of the behavioural characteristics of students during their learning interaction with a diagnostic feedback system [Alemán et al. \(2011\)](#).

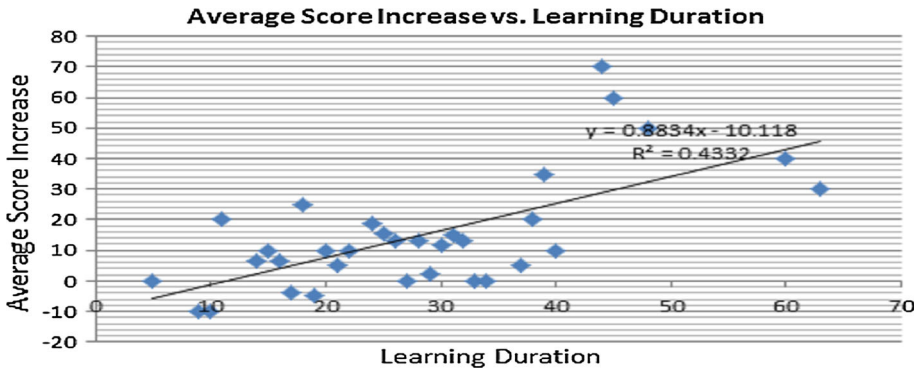


Fig. 2 Average score increased versus learning duration

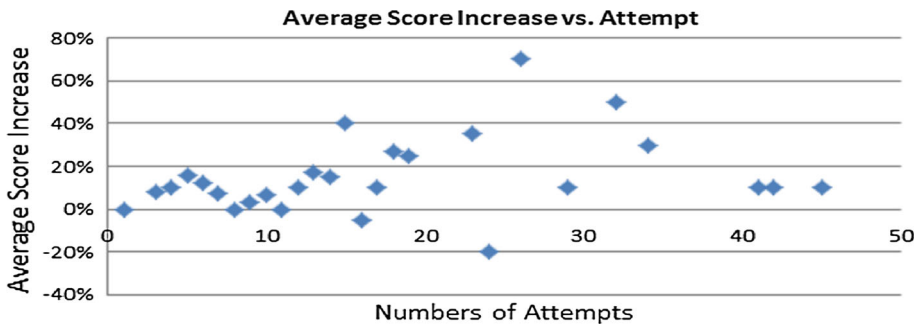


Fig. 3 Average score increase versus attempt

In order to explore the characteristics of students, and relate these to student responses and performance in the tests, five behavioural variables were analysed: the number of attempts (submissions), the average score changed between attempts, the average score at the end of trial, the amount of time spent to make each attempt, and the learning duration. Figure 2 illustrates a learning behaviour of this group of students by analysing the relationship between average scores increased and learning duration. Each blue point represents average scores increased of all students used the same learning time, and its coordinate of *x*-axis represents student’s learning duration, and its coordinate of *y*-axis represents average scores increased. It can be achieved from this figure that average scores increased when students spent more time on studying from the system.

Figure 3 illustrates a learning behaviour of this group of students by analysing the relationship between average scores increased and number of attempts. Each blue point represents average scores increased of all students did the same number of attempts, and its coordinate of *x*-axis represents number of students’ attempts, and its coordinate of *y*-axis represents average scores increased. It can be achieved from this figure that average scores increased when students did more attempts before peak, and average scores no longer increased when students did 26 attempts, and average scores decreased by doing more attempts after peak.

Long Learning Duration (time spent on learning ≥ 44 min) and Medium Attempts ($35 \geq$ number of attempt ≥ 18) are consistently associated with good score increases, especially the combination of Long Learning Duration and Medium Attempts, and hence represent successful learning strategies amongst the students (Tables 3, 4, 5, 6).

Table 3 *Short learning duration*: time spent on learning < 44 min; *Long learning duration*: time spent on learning \geq 44 min

Behavioral group	Average score change (%)	Average score in the end (%)	Number of students
Short learning duration	9.2	50.6	77
Long learning duration	46.7	95	6

Table 4 *Many attempts*: number of attempt > 35, *Medium attempts*: $35 \geq$ number of attempt \geq 18, *Few attempts*: number of attempt < 18

Behavioral group	Average score change (%)	Average score in the end (%)	Number of students
Many attempts	8.8	51.2	65
Medium attempts	31.5	72.3	13
Few attempts	15	57.5	4

Table 5 *Slow attempts*: average time spent on each attempt \geq 4 min; *Rapid attempts*: average time spent on each attempt < 4 min

Behavioral group	Average score change (%)	Average score in the end (%)	Number of students
Slow attempt	12.96	53.52	54
Rapid attempt	11.72	54.48	29

Table 6 *Few rapid attempts*: average time spent on each attempt < 4 min and number of attempt < 18; *Many rapid attempts*: number of attempt >35 and average time spent on each attempt < 4 min; *Slow Few attempts*: average time spent on each attempt \geq 4 and number of attempt < 18; *Rapid medium attempts*: $35 \geq$ number of attempt \geq 18 and average time spent on each attempt < 4 min

Behavioral group	Average score change (%)	Average score in the end (%)	Number of students
Few rapid attempts	6	48.4	45
Many rapid attempts	15	57.5	4
Slow few attempts	16	54.5	20
Rapid medium attempts	31.5	72.3	13

Figure 4 illustrates the behaviour of students in terms of what might be called knowledge states. These states correspond to the student responses triggered by patterns of student answer. In other words, a state of knowledge captures some commonality in a set of questions responses. For example, if there are several students who give the same answer (correct or incorrect) to two or more of the questions, snap-drift will form a group associated with one particular output neuron to include all such cases. That is an over simplification, because some of those cases may be pulled in to other 'stronger' groups, but that would also be characterized by a common feature amongst the group of responses. Figure 3 shows the knowledge state transitions. Each time a student gives a new set of answers, having received some feedback associated with their previous state, which in turn is based on their last answers, they are

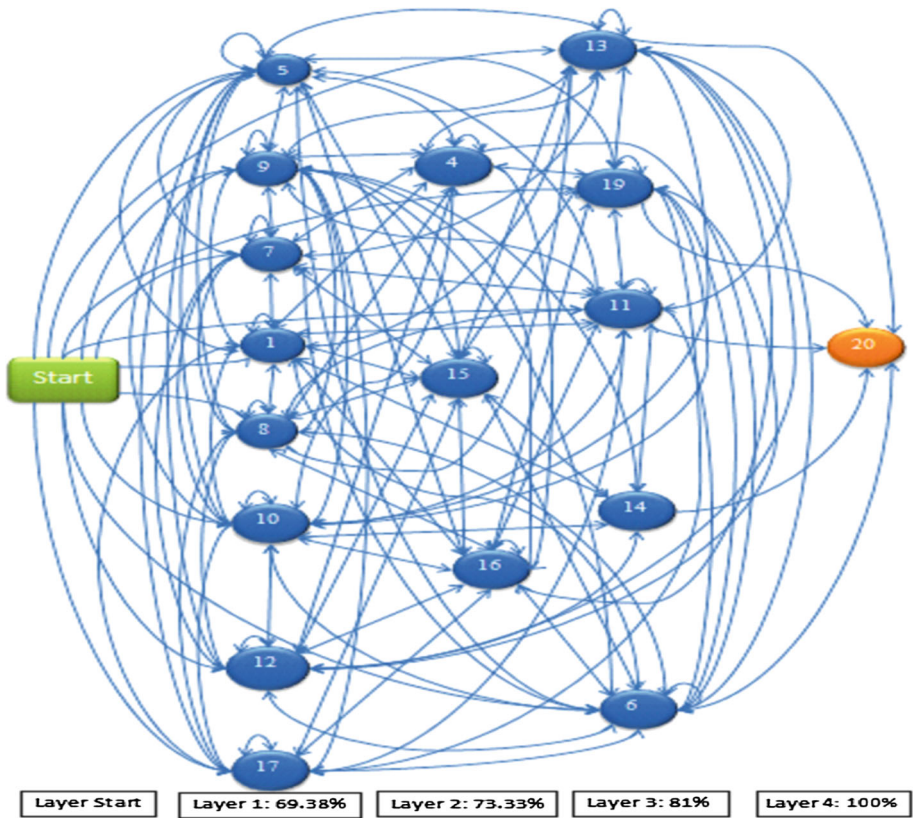


Fig. 4 Knowledge state transitions

reclassified into a new (or the same) state, and thereby receive new (or the same) feedback. The tendency is to undergo a state transition immediately or after a second attempt or several attempts. A justification for calling the states ‘states of knowledge’ is also to be found in their self-organization into the layers. There are 5 layers: Start, Layer 1, Layer 2, Layer 3, and Layer 4. A student on state 5, for example goes via one of the states in the further layer such as state 4 or 13, before reaching the ‘state of perfect knowledge’ (state 20) which represents correct answers to all questions. On average, and unsurprisingly, the state-layer projecting onto state 20 (states 12) is associated with more correct answers than the states in the previous layer. This is true of state 13 which projects onto state 20, and it is also true of state 4 although this state does not project onto state 20. The states in the middle layer (layer 2) do not connect to start layer or layer 4. The average score of each layer is increased from start to state 20 (Average scores are 69.38% at beginning level (layer 1), 73.33% at middle layer (layer 2), 81% at advanced level (layer 3), and 100% at state 20). Students often circulate within layers before proceeding to the next layer. They may also return to a previous layer, but that is less common.

6 Summary

Six trials across three totally different subject areas have been carried out in three universities in two countries, the UK and China: 3 English trials, 2 Mathematics trials, and 1 Java

Table 7 *Short learning duration many attempts*: time spent on learning < 44 min and number of attempt > 35; *Short learning duration medium attempts*: time spent on learning < 44 min and 35 \geq number of attempt \geq 18; *Short learning duration few attempts*: time spent on learning < 44 min and number of attempt < 18; *Long learning duration few attempts*: time spent on learning \geq 44 min and number of attempt < 18; *Long learning duration medium attempts*: time spent on learning \geq 44 min and 35 \geq number of attempt \geq 18

Behavioral group	Average score change (%)	Average score in the end (%)	Number of students
Short learning duration few attempts	7.9	48.4	62
Short learning duration many attempts	16	57.5	4
Short learning duration medium attempts	23	64	10
Long learning duration few attempts	33.3	90	3
Long learning duration medium attempts	60	100	3

Programming trial. The English trials are very successful with 500 students participated. A large volume of data has been captured during the trials.

The students participating in this second trial are from KTU which is one of the top universities in China. The entrance requirements of KTU are much higher than those of JQU, which means that students studying in KTU generally have a better academic background than students in JQU. The results of the second English system trials show that the average score is increased by 21.2% in the end. The results of the separate MCQ paper test indicated the average mark of the group is increased by 7.6%. In addition, the final examination, the average mark of the experimental group is 7.5% higher than the control group. Student surveys show that 87.8% students are satisfied with our e-learning system and 89.2% students feel that the intelligent diagnostic feedback is beneficial to enhance their learning experience. Furthermore, according to all of the results, the M-OFS system demonstrates that it is working well for both less qualified and better qualified students at two universities JQU and KTU. The results of group behaviours in Table 7 show that students in the group “Long learning duration medium attempts” achieved much better learning effects than others. In comparison with the 1st trial, the students in the group “Long learning duration medium attempts” achieved better learning effects than others. This has indicated that the students in the 2nd trial who have better academic background prefer more reflective thinking and independent study from the feedbacks, whilst, the students in the 1st trial who have weaker academic background prefer more frequent and direct feedbacks as part of their learning process.

7 Conclusion and future work

This research develops a novel method for using snap-drift in a diagnostic tool to provide feedback. The neural network diagnostic feedback approach in MCQ has been systemically applied to large cohorts of students and evaluated across a range of different subject areas. The neural network discovers groups of similar answers that represent different knowledge states of the students. The feedback targets the level of knowledge of individuals, and their misconceptions, guiding them toward a greater understanding of particular concepts. The results of the experiment show that an improvement in the learning process can be obtained by using M-OFS. In future work, it is intended to compare the effects of the feedback in new trials to the effects of other

types of feedback already reported in the literature. Another promising avenue for further investigation is the extension of the tool to support knowledge state transition diagram construction and statistical data collection, which could help instructors to analyse the difficulty of the MCQs and to track students though the developmental stages of their learning.

References

- Alemán JLF, Palmer-Brown D, Jayne C (2011) Effects of response-driven feedback in computer science learning. *IEEE Trans Educ* 54:501–508
- Alessi M, Trollip S (2001) *Multimedia for learning: methods and development*, 3rd edn. Allyn and Bacon, London
- Bang P (2003) Engaging the learner—how to author for best feedback. In: Felix U (ed) *Language learning online: towards best practice*. Swets & Zeitlinger, The Netherlands
- Blessing S, Gilbert S, Ourada S, Ritter S (2007) Lowering the bar for creating model-tracing intelligent tutoring systems. In: Luckin R et al. (eds) *Artificial intelligence in education*. IOS Press, Amsterdam
- Clark RC, Mayer RE (2009) Instructional strategies for directive learning environments. In: Silber KH, Foshay WR (eds) *Handbook of improving performance in the workplace: instructional design and training delivery*. Pfeiffer, San Francisco
- Epstein ML, Lazarus AD, Calvano TB, Mathews KA, Hendel RA, Epstein BB, Brosvic GM (2002) Immediate feedback assessment technique promotes learning and corrects inaccurate first response. *Psychol Record* 52(2):187–201
- Garber PR (2004) *Giving and receiving performance feedback*. HRD Press, Canada
- Gheorghiu R, Vanlehn K (2008) XTutor: an intelligent tutor system for science and math based on excel. In: Woolf BP et al. (eds) *Intelligent tutoring systems: 9th international conference, ITS 2008*. Springer, Germany
- HatziaPOSTOLOU T, Paraskakis I (2010) Enhancing the impact of formative feedback on student learning through an online feedback system. *Electron J E-Learn* 8(2):111–122. Available online at www.ejel.org
- Hefce national student survey (2007–2010) HEFCE, London, UK. Online at: <http://www.hefce.ac.uk/whatwedo/lt/publicinfo/nationalstudentsurvey/>
- Heinze A, Procter C, Scott B (2007) Use of conversation theory to underpin blended learning. *Int J Teach Case Stud* 1(1&2):108–120
- Higgins E, Tatham L (2003) Exploring the potential of multiple-choice questions in assessment. *Learn. Teach. Action* 2(1). Assessment, Winter 2003
- Kuechler WL, Simkin MG (2003) How well do multiple choice tests evaluate student understanding in computer programming classes? *J Inf Syst Educ* 14(4):389–399
- Little B (2001) Achieving high performance through e-learning. *Ind Commer Train* 33(6):203–206
- Ma Z (2006) *Web-based intelligent e-learning systems*. Information Science Publishing, USA
- Nelson MM, Schunn CD (2009) The nature of feedback: how different types of peer feedback affect writing performance. *Instr Sci* 37(4):375–401
- Özdener N, Satar HM (2009) Effectiveness of various oral feedback techniques in CALL vocabulary learning materials. *Egitim Arastirmalari—Eurasian J Educ Res* 34:75–96
- Palmer-Brown D, Jayne C (2011) Snap-drift neural network for self-organisation and sequence learning. *Neural Netw* 24(8):897–905
- Paxton M (2000) A linguistic perspective on multiple-choice questioning assessment and evaluation. *Assess Eval High Educ* 25(2):109–110
- Payne A, Brinkman, Wilson F (2007) Towards effective feedback in e-learning packages: The design of a package to support literature searching, referencing and avoiding plagiarism. In: *Proceedings of HCI2007 workshop: design and use and experience of e-learning systems*, pp 71–75
- Race P, Brown S (2005) *500 tips for tutors*, 2nd edn. Routledge Falmer, London
- Race P (2006) *The lecturer's toolkit—a practical guide to assessment*. Learning and teaching, 3rd edn. Routledge, London
- Rane A, Sasikumar M (2007) A constructive learning framework for language tutoring. In: Iskander M (ed) *Innovations in e-learning, instruction technology, assessment, and engineering education*. Springer, The Netherlands
- Reiser BJ, Kimberg DY et al (1992) Knowledge representation and explanation. In: Larkin JH, Chabay RW (eds) *Computer assisted instruction and intelligent tutoring systems*. Lawrence Erlbaum Associates, New Jersey

- Springgay S, Clarke A (2007) Mid-course feedback on faculty teaching: a pilot project. In: Darling LF, Erickson GL, et al (eds) *Collective improvisation in a teacher education community*, Chapter 13. Springer, The Netherlands, pp.171–185
- Vasilyeva E, Pechenizkiy M, Bra PD (2008) Adaptation of elaborated feedback in e-learning. In: Nejd W, Kay J, Pu P (eds) *Adaptive hypermedia and adaptive web-based systems*. Springer, German