

# Deciding on Feedback Polarity and Timing

Stuart Johnson  
Alberta Innovates Centre for Machine Learning  
University of Alberta  
Edmonton, Canada  
shjohnso@ualberta.ca

Osmar Zaiane  
Alberta Innovates Centre for Machine Learning  
University of Alberta  
Edmonton, Canada  
zaiane@cs.ualberta.ca

## ABSTRACT

This paper outlines the feedback creation and assignment techniques used in a mammography focused Intelligent Tutoring System (ITS) Shufti. Shufti's aim is to provide medical students with an improved learning environment, exposing them to a broad range of examples supported by customized feedback and hints driven by an adaptive Reinforcement Learning system and Clustering techniques.

## 1. INTRODUCTION

Shufti is an Intelligent Tutoring System (ITS) which has been designed to help medical students learn the skills they need to master the complexities of producing a medical diagnosis based on relatively poorly defined, low contrast images. ITS's are tutoring systems which approximate human one-on-one tutoring experiences. Shufti takes the form of a web-based computer game in which learners compete with one another to correctly diagnose images. They are presented with mammograms overlaid with grids and are required to identify Regions Of Interest (ROI) by selecting cells in these grids. As students complete each exercise, they are given a score derived from their accuracy in identifying lesions less points for any hints they may have requested.

Effective human tutors play an active role in the learning experience, providing hints and positive and negative feedback in a strategic fashion. Furthermore, they adapt their feedback to suit the learning styles of their students.

For an ITS to produce similar results it must provide comparable forms of interaction, which is non-trivial for an ITS in the field of mammography as it is an ill-defined domain by the definition of Viger et al [1]. Mammography lacks clear domain models, formal theorems, and cognitive models necessary to automatically teach mammogram diagnosis using conventional ITS construction methods[1]. Consequently, Shufti utilizes a variety of means to effectively simulate attributes of a human tutor.

## 2. APPROACHES TO FEEDBACK

Exercises in Shufti are categorized by difficulty level. Students move from one level to the next after accumulating sufficient points on a certain number of mammograms.

For each exercise Shufti, records the task state transitions which comprise of the exercise state and learner's actions during the exercise. Included in this are current and past

states representing their current solution, the last action taken, Shufti's feedback, and the reaction to the feedback by the learner. The *state* is the number of grid cells selected which differ from the exercise solution (i.e. hamming distance). *Actions* are operations such as toggling square selections, certain mouse movements, hint requests, and the submission of an exercise for evaluation. *Reaction* to feedback is whether or not the learner explicitly found the previous feedback helpful.

The polarity of feedback is based on whether it is a positive, encouraging message or a negative, corrective message. The polarity is selected based upon whether the state of the exercise improved or degraded. Degradation or improvement is determined through comparison of current and past hamming distances from the correct state.

Feedback is a critically important part of the effectiveness of a human tutor. To this end, Shufti contains methods for determining the content, polarity and timing of feedback. Shufti employs two feedback control approaches: a clustering-based method and a technique based on Reinforcement Learning.

### 2.1 Clustering-based method

Shufti clusters learners based on their level, points accumulated, the number of requested hints, and the number of exercises they have attempted.

The timing of feedback is governed by a number of different models. *Random* feedback, as its name suggests, occurs randomly. *Timed* feedback is delivered after timed intervals. *After Action* feedback is issued in response to the learner undertaking any action. *Timed After Action* feedback is similar to *After Action* except it is delayed by a specified time. *Random After Action* feedback is similar to *After Action* except it is randomly delivered (it may or may not be issued)

For a given learner, when Shufti has to decide on feedback based on one of the timing models, the potential appreciation of the feedback is assessed based on the task state transitions of similar learners (i.e. learners in the same cluster as the current learner).

If a learner is likely to appreciate a feedback, it is issued. This prediction is based on the likelihood of the reaction to feedback being positive for all available similar records in the task state transition file of all students in the cluster. In the case of a cold start, a random feedback will be issued. This is also performed at random times so as to explore and discover new situations in which feedback may be appropriate.

Clustering allows Shufti to adapt to fit individual learners

or problems. Shufti uses all available data to learn which feedbacks are effective.

## 2.2 RL-based method

Adapting to individual students, though time consuming, is one of the ways in which human tutors offer a superior learning experience. Reinforcement Learning (RL) offers an automated method for an ITS to tune its feedback delivery to individual learners thus approximating a human tutor.

RL is a class of machine learning techniques which resolve problems of mapping situations to actions in order to maximize or minimize a metric[2]. RL allows Shufti to adapt to individual students, learning the most effective times to issue feedback, thus avoiding preset timing models.

An RL system can be thought of as two components; an agent and an environment within which it acts. The environment provides state data and a reward signal to the agent which in turn attempts to maximize the total reward over time. The agent makes use of methods such as Temporal-Difference Learning[2], or Monte Carlo Methods[2] to determine the most long term rewarding action to take in any given state.

Shufti's environment offers task state transitions as state information to the agent. The reward signal is determined by the following formula with the agent seeking to minimize it. It should be noted that when we refer to penalties in the coming paragraph we refer to penalties applied to the agent not to the learner.

$$P = \sigma * count(\tau) + \omega * count(f) - \alpha * score$$

Where  $P$  is the total penalty assessed to the agent,  $\sigma$  is the penalty assigned over time,  $count(\tau)$  is the total time passed,  $\omega$  is the feedback penalty,  $count(f)$  is the total number of feedbacks given by the agent,  $\alpha$  is the reward per score point earned by the user, and  $score$  is the total score that the user is assigned for the exercise. Time taken is penalized to encourage the agent to give feedback as a means to hasten the answering of the question. Penalties are also given to the agent each time it gives out a feedback in order to produce strategic feedback selection and timing. In other words, this allows the agent to strike a balance between helping the learner and allowing self driven action. The rate of feedback can be controlled by varying  $\sigma$  and  $\omega$  with it increasing with  $\sigma$  and decreasing as  $\omega$  increases. Such variation of the values can either be done automatically so as to simulate the withdrawal of support of a human tutor, or can be done manually by an instructor as part of a larger lesson plan.

This RL-based method not only offers Shufti the ability to control the timing, polarity and content of feedback, but also the ability to adapt to individual learners, thus more closely mimicking human tutors. The downside of this method is the need to understand an effective tutoring strategy for each learner, unlike the previously discussed clustering method which takes advantage of information from many learners in order to adapt.

## 3. COMPETITION

One of the key limitations in traditional training of medical students in imaging analysis is the amount of cases students are exposed to. There are two ways in which Shufti addresses this issue: first of all, Shufti has a very extensive

selection of exercises covering a wide range of scenarios unlikely to be seen during a student's short rotation in a radiology department and, secondly, Shufti uses competitive techniques learned from gaming to incite students to expose themselves to as broad a range of scenarios as possible, deepening their knowledge in the field.

Competitive practices in learning have been shown to produce significant improvements in learner performance [3]. To foster competition, Shufti adopts practices from competitive sports and gaming. Learners are not assigned scores based on any one single measure but instead on a composite of measures designed to work with hints from Shufti.

In total, the scores are created based on problem difficulty, answer accuracy, time spent answering the exercise, and learner requested hints. Learners are presented with a wide variety of means to see how they rank next to their peers. In addition to typical public leader boards (commonly used with popular on-line games), Shufti presents performance distribution curves. A learner's overall ranking in Shufti is based on the sum of all scores they have received from all exercises, encouraging them to attempt a large number of exercises.

## 4. HINTS

Hints in Shufti are user-requested, optional pieces of information which aid in solving exercises. They differ from feedback in both how they are issued to the learner and their content. Feedback, for example, takes the form of general statements such as, "Good job!", whereas hints are more direct such as suggesting a general area in which an ROI may be located.

Users are presented with a set of possible hints to request, each hint being labelled with a description of what kind of information the user will receive, along with a specific score penalty which will be applied should the user accept the hint.

Hint penalties ensure the user does not try to improve their score through excessively requesting hints - a phenomenon known as gaming the system [4]. Hint penalties may also have the interesting effect that learners will strategically select the minimum number of hints necessary for them to answer an exercise correctly. Additionally, this causes students to think strategically about which hints they might want to accept, thus broadening their understanding of the diagnostic process.

## 5. REFERENCES

- [1] P. Fournier-Viger, R. Nkambou, and E. M. Nguifo. Building intelligent tutoring systems for ill-defined domains. *Studies in Computational Intelligence*, 308, 2010.
- [2] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.
- [3] S.-F. Lam, P.-S. Yim, J. S. F. Law, and R. W. Y. Cheung. The effects of competition on achievement motivation in Chinese classrooms. *British Journal of Educational Psychology*, 74(2):281-296, June 2004.
- [4] R.S. Baker, A.T. Corbett, K.R. Koedinger, A.Z. Wagner, Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System". *ACM Computer-Human Interaction*, 383-390, 2004