Learning to Analyze Medical Images: A Smart Adaptive Learning Environment for an Ill-Defined Domain

Stuart Johnson[1] and Osmar Zaiane[1]

University of Alberta, Alberta, Canada,
shjohnso@ualberta.ca, zaiane@ualberta.ca

Abstract. While Medical Imaging is a cornerstone of modern medicine, adequately instructing students in medical schools to analyze these images is still lacking in terms of allocated time and diversity in exposure. To help alleviate this problem we propose the creation of a new smart learning environment: the gamification of an online Intelligent Tutoring System focused on medical imaging. This proposed system makes use of modern serious games practices to increase student engagement, and plays on innovative student modelling techniques to provide learner adaptive feedback in a field of study where conventional techniques do not apply.

Keywords: Smart Learning Environments, Intelligent Tutoring Systems, Adaptive Feedback, Gamification, Medical Imaging, Serious Games, Reinforcement Learning

1 Introduction

Analysis of medical images is an important tool in modern medicine. Modern technology has changed how we explore and diagnose issues. Where once it would be necessary to preform a risky exploratory surgery to determine the source of an ailment, now an MRI can peer into the depths of the human body from outside. Beyond simply making diagnosis easier, modern technology has enabled remote diagnosis; an expert no longer needs to be present to provide insight.

These advances, though, come with a downside: the volume and variety of medical images produced has increased by leaps and bounds. With technologies such as Magnetic Resonance Imaging (MRI) becoming common place, the demand for professionals able to analyze them has increased. Unfortunately the current process of training such professionals follows an apprentice model with medical students following a fully practicing doctor in what is known as a practicum, during which they assist their instructor and review their cases. This has proven historically adequate but has some downsides. The students are only exposed to the cases that their instructor receives meaning that rare conditions are not seen, resulting in a less than comprehensive education. In addition, the limited number of cases seen by a doctor mean that the students have a limited ability to practice. Further worsening this is the difficulty of analyzing medical images, resulting in poor rates of detection[8].

A smart learning environment could help with these issues in a variety of ways. By increasing the students access to the cases including rare ones it can improve the breadth of their experience. To induce practice one can use techniques borrowed from serious games [16]. To retain the quality of instruction from tutoring[1] that a student receives during their practicum such an environment can use intelligent tutoring techniques.

To this end we have created a new online smart learning environment, a medical imaging intelligent tutoring system, Shufti. It uses gamification and social techniques to provide the learner with motivation to learn. To increase students access to cases it draws cases form large existing databases of medical images. To retain the benefits of tutoring it makes use of innovative student modelling to provide adaptive learner specific feedback in a domain of study which, due to its structure does not lend itself to normal techniques.

Serious games, games which are used to teach serious topics, are an up and coming innovation in education[14][9]. Gamification has many benefits such as being habit forming [18], resulting in more practice. Completion can transform routine tasks into interesting challenges. Social features can further improve the learning experience by allowing learners to share knowledge amongst one another. Our implementation is explored in Section 3.1.

There are many large medical image databases online from which is it possible to draw exercises. Some examples are the Digital Database for Screening Mammography [5] or the Segmentation Chest Radiograph Database[17]. This enables students to see a broader set of cases than they normally would. These databases also contain rare cases as well, allowing students to increase the depth of the cases that they have seen by emphasizing rare or difficult cases. Our use of these is outlined in the introduction to Section 3

With all of above benefits though it is undesirable to lose the advantages of the practicum system. Namely the benefits of one on one tutoring. To not lose these benefits modern intelligent tutoring system techniques can be used. Such a system can provide leaner specific adaptive feedback to the students, simulating such a tutoring experience. In Shufti's case a innovative form of Reinforcement Learning[15] based student modelling is used to provide advice and feedback to the leaner. This was done due to medical imaging being what is know as an ill-defined domain which is a domain which does not lend itself to normal intelligent tutoring system techniques. The motivation, challenges, methods, and validation of this are outlined in Section 4

With all of these features combined, Shufti would provide a high quality education experience to medical imaging students, and is just about to be released to the public.

2 Contemporary Medical Imaging Learning

The analysis of medical images is an essential tool in the modern medical professional arsenal. Improvements in availability and technology has made them a common part of diagnosing many conditions. They are used by pathologists to analyze biopsies. Oncologists using magnetic resonance imaging (MRI) can see cancer long before it is apparent through other symptoms. X-Ray mammography is a key means for early detection of breast cancer. Dermatologist can detect melanoma using dermatoscopy or even normal photos of skin lesions.

Unfortunately the training of such skills is very difficult, one requires a significant number of cases to learn from to distinguish what are relevant features and anomalies in images. There is significant judgement involved and what is in an image is often unique or different from what a doctor has seen before. To remedy this, medical students follow or shadow an established specialist during what is called a practicum. This though effective, can result in a non comprehensive viewing of the possible conditions as the cases available for instruction are limited to the ones present during the practicum.

2.1 A Medical Imaging Smart Learning Environment

To resolve these problems we have set out to create an online medical imaging smart learning environment. Such a tool has many desirable characteristics[11]. The system can contain many exercises featuring each possible condition, allowing for many hours of practice on even rarely encountered conditions. The system also provides adaptive feedback to the learner much as a real instructor would. The system also encourages the learners to continue to make use of it so as to help with maintaining familiarity.

Our medical imaging Intelligent Tutoring System initially focuses on the analysis of mammograms as there is significant demand for professionals with that skill. The system contains thousands of cases for the students to analyze. It incorporates new and innovative adaptive feedback selection techniques to simulate the presence of a human tutor. It makes use of gamification techniques to encourage more practice by the learners.

3 Shufti

Shufti[7] is an online ITS featuring gamification techniques and an adaptive feedback selection for learners. At the moment, it enables students to practice mammogram analysis on 1827 images revived from a variety of sources including the Digital Database for Screening Mammography [5]. Users are presented with a series of exercises of 4 different types depending on their *level*. These exercises take the following forms, in order of difficulty: Presence Exercises, in which the user is tasked with determining if there is a malignant lesion present in the mammogram presented; Heat Grid Exercises, where an eight by eight grid is overlayed on the mammogram and the student receives information guiding them to the lesion, in the form of a simple hot and cold message; Gird Exercises which also have the eight by eight grid but lack the guidance; and Polygon Exercises, in which the learner is tasked producing the exact outline of any lesions present in the mammogram. Figure 3 shows an example of a Heat Grid Exercise.

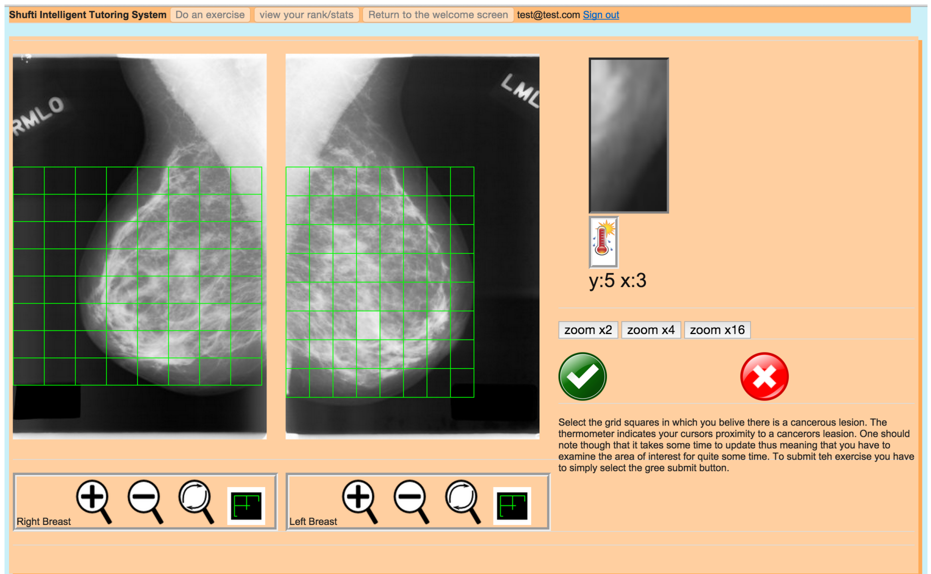

Fig. 1. The User interface for a Dual Image Heat Grid Exercise

3.1 Pedagogical Goals and Gamification

Shufti's Pedagogical goals are to improve the breadth and depth of the knowledge of medical imaging instruction. It is desirable for the learners to see and analyze as many cases as possible. It is also desirable for them to gain as much value out of each viewing as possible. To accomplish this Shufti makes use of modern gamification techniques, to have the learners view more cases than they normally would, while also causing them to focus on their own performance thus leading to deeper practice[9]. Gamification as described in Yee's Pedagogical Gamification[18] is a set of design principles used in the creation of a set of challenges, or in education's case exercises. These design principles are: Displaying Progress, Maximizing Competition, Careful Difficulty Calibration, Providing Diversions, and Employing Narrative Elements. We will now examine each of these principles and how it relates to Shufti's design.

Displaying Progress This display of progress to the learner is an important part of the game experience. By showing how far a player has come they receive encouragement and will continue to play. In the case of a serious game learners would continue to practice encouraged by their growing proficiency[4]. In Shufti displaying progress is accomplished through three means. First they have a prominently displayed score at the completion of an exercises, which is also tracked by a summary graph on the main page, enabling them to view their overall progress. Second as they improve their skills they can ascend to a new level where they are presented with more difficult exercises. Finally they have a chance of seeing the same exercise again, a fact that is revealed to them upon completion of the exercise, this is to enable them directly see their own progress.

Maximizing Competition Humans are naturally competitive, valuing the ability to compare ones self with peers. To succeed against others can be a powerful drive. Shufti optionally provides learners with an anonymized ranking where they can compare themselves with their peers. It also allows learners to see how their peers answered exercises by overlaying colours on the mammogram indicating the frequency that areas were selected as containing malignant lesions, enabling a more focused comparison.

Careful Difficulty Calibration Challenge and the overcoming of challenges is an important part of game design. Too little challenge and the learner would feel the task is trivial and would not practice, too much challenge and the learner would grow tired or frustrated and give up. Shufti calibrates its challenge through the use of the previously mentioned levels. Once they have demonstrated mastery of their current level they ascend to the next level, this continues until they reach the most difficult type of exercise. Each level also has limits on the subtlety and size of the lesion present. This in effect allows Shufti to adapt to the learner and their proficiency providing both a constant source of challenge and a display of the users progress.

Providing Diversions It is also important to break up the learning experience, to reset the learners attention. This can be done though diversions not directly related to the core learning task. In Shufti's case there are many things that learners can do other than directly answering exercises. They have the ability to comment on the exercises they have seen and to see their peers comments. They can see replays of their mouse movements over past exercises to review their actions. They can also annotate the exercises so when they solve them again they can see their past opinions and considerations.

Narrative Elements Many games also provide narrative elements to further entice the learner. Unfortunately the analysis of medical images does not lend itself to this. Thankfully though it is not necessary for a game to have all of these principles to be engaging[18]. A good example of this is chess which is quite engaging but has no story at all.

By using the above principles and techniques Shufti provides an engaging learning experience that is adaptive to the individual leaners skills.

4 Smart Learner Adaptive Feedback

To properly simulate a tutor an ITS must provide advice and encouragement to the learner. This is the foundation of the effectiveness of one-on-one human tutoring. Our proposed system makes use of a new innovate approach to issuing feedbac in domains, where instead of trying to model the domain and infer what a correct action is, the system focuses on managing the learners motivational state and overall correctness. This is done through issuing learner specific custom feedback.Effective feedback can produce large gains in the performance of individual learners [1].

4.1 Feedback and Hints

Feedback is unrequested information given to the learner by the system. This differs from hints in that hints are user requested information. Feedback itself can have a polarity. Positive feedback is encouraging or reinforcing, for example "good job". Negative feedback is corrective or cautionary in nature, for example "You should look again". It is important to note that the polarity of feedback represents what the intended outcome is.

4.2 Intelligent Tutoring System Domains

Unfortunately medical imaging is what is known as an ill-defined domain, which defeats most feedback selection techniques as they rely on having a model of the domain in which they instruct. A domain is a field of knowledge or study. For example mathematics or history. But not all fields have the same properties. Some, such as mathematics, are trivial for an automated system to reason about their structure. Others such as law, are not so easily reasoned about. The latter type of domains are called ill-defined while the former type are called well defined.

4.3 Well-defined Domains

A domain is classified as well-defined if it has all of the following attributes: verifiability, formal theories, a well-defined task structure, clearly defined concepts, and a decomposable task structure[3]. These attributes also enable the use of conventional intelligent tutoring techniques and domain modelling.

Verifiability A well-defined domain must have verifiable answers - such as in arithmetic where there is only one correct answer for a given problem - or, at the very least, the ability to distinguish a correct answer from an incorrect answer. Problems in ill-defined domains lack verifiability in that they commonly do not have a truly correct answer. For example, consider a writing exercise in which the goal is to write an interesting story. It is possible to write an interesting story and it is possible for one to manually determine if a story is interesting, but there is no one correct answer to the exercise. In other words, an ambiguous notion of correctness is one of the possible features of an ill-defined domain, which is in conflict with the idea of verifiability.

Formal Theories A well-defined domain must have a clear-cut set of rules, or formal theorems within which one can reason and solve problems. An example of the use of formal theorems is symbolic logic in which the entire problem space is well-defined and there are rules which can be applied in all situations. In contrast, in image medical diagnosis, there are only vague notions of formal theories. This vagueness and lack of theories exacerbate the creation of domain models which results in ill-defined domains.

Well-Defined Task Structure A well-defined domain must have a regular task structure or, as stated in Fournier-Viger et al. [3], be a "Problem Solving Domain". Task structures other than problem solving are possible, but they are ill-defined. An example of this is an analytic task which requires the learner to make choices based on incomplete or erroneous information and in effect "use their judgement". Another example is problems in which the desired outcome is pre-defined and where novelty is desired such as in writing.

Clearly defined concepts For a domain to be well-defined, its concepts must have a clear definition, without ambiguity. For example, a triangle is a well-defined concept with no ambiguity.This is not always present within a domain. For example, the various styles of art are not well-defined and a work's inclusion in a particular style is based more on consensus of the community and the statements of the artist than on any hard definition. However, clear, unambiguous definitions are necessary if one is to reason about concepts within a domain.

Decomposable task structure For a domain to be well-defined it must be possible to decompose problems within it to a sequence of separate and independent steps. For example, when solving a logic problem the correctness of a given step can be determined from the current state and the desired end state. The correctness of steps taken before the current step have no influence on whether or not the current choice is correct.

4.4 Medical Imaging as an Ill-Defined Domain

Medical Imaging is an ill-defined domain as it lacks some of the attributes of a well defined domain as is argued by Crowley and Medvedeva [2]. It lacks formal theories as it relies on the judgment of the person preforming the analysis. It also does not have clearly defined concepts, for example the line between a cancerous melanoma and a benign one is not clear. Thankfully though it does have some properties which make it possible using unconventional methods to create an ITS without a model of its domain. The domain is verifiable, the task structure is well defined, and the problem is comprised of discreet sub problems. The domain retains verifiability as with human annotation it is possible to create a gold standard of correct answers to medical imaging problems. The task structure is well defined as the task is simply to determine what the correct diagnosis should be. Also the problem is established independent of sub tasks, such as each exercise is self contained.

4.5 Methods

Adapting feedback to individual learners is a non trivial task, learners vary significantly in their preferences for what kind of feedback they receive and the amount. Further complicating this is that as a learners proficiency improves the most effective feedback strategy can shift as shown in Soldato et al. [13]. In which while initially acquiring skills, learners liked many positive reinforcing feedback but once they were more proficient they preferred more negative corrective feedback to point out their remaining misconceptions.

There are many possible tactics with which to provide adaptive feedback, for example in a well defined domain such as mathematics, an ITS can simply indicate the error that the student made by comparing their actions with the steps taken by a computer algebra system. In the case of ill-defined domains, unconventional techniques are required as in Perelman et al.'s Code Hunt [10], a programming ITS, in which how prior students resolved errors is used as the basis for corrective feedback for students.

We propose a new method of providing learner specific adaptive feedback, which is in use in Shufti. This new method makes use of reinforcement learning algorithms[15] to model students to allow the system to determine the most effective feedback strategy for the learner, even in the case that no feedback at all is the best choice. It is described in more detail in Custom Feedback Selection for Intelligent Tutoring Systems in Ill-Defined Domains [6].

Reinforcement Learning is a branch of artificial intelligence which learns how to solve non stationary problems in a real time manner. Tt learns and adapts to the problem even as the problem changes. These attributes are important as learners feedback needs can shift over time and the system must be able to react to changes and novel situations. The high level explanation of Reinforcement Learning is thus: given an environment which provides a reward signal and a state signal, control an agent/actor within that environment such that over time the reward signal is maximized.

To characterize providing feedback to the learner in such an environment, we did the following. The current score, as though the learner had submitted the exercise, is computed. It is then discretized into a state signal with 3 possible values and is used as a state signal for the Reinforcement Learning agent. The discretized score is placed in the ranges of greater that 0.5, less than 0.5 but greater than 0.25 and less than 0.25. These ranges were chosen to efficiently represent the difficulty of the mammogram diagnosis and to make the learning rate of the agent as rapid as possible. The reward signal of the agent is again the learners score but it is not discretized. This reward signal was chosen such that the system would learn the most effective feedbacks as quickly as possible by observing its feedbacks effects during exercises.

The agent chooses from a list of possible feedbacks to issue to the learner. The list is a set of General Statements and, advice for the learner, labeled with a polarity, either positive or negative. Positive feedbacks are reinforcing in nature, advising the learner to continue with their current course of action. Negative feedbacks are correcting in nature, and attempt to dissuade an incorrect course of action.

The Reinforcement Learning agent is limited to positive feedbacks if the learner has been improving their answer and is restricted to negative feedback when the learner has made the exercise state worse.

The division between negative and positive feedback is important as it prevents the agent from issuing feedback which does not match the current situation for example, issuing encouraging, positive, feedback when the learner has made a mistake. At all times it is possible for the reinforcement learning agent to choose to issue no feedback and, in fact, that is a common choice.

Each learner has their own agent which, over time, learns their preferences when it comes to feedback, which ones are effective in improving their performance and which ones make the learner act incorrectly. This results in the system eventually issuing only the feedbacks the learner likes and which are effective in helping them learn.

4.6 In Vitro evaluation of feedback system

To provide some validation of the effectiveness of this feedbacks system we have created a learner simulator. This is to provide us with some idea as to its effectiveness and to provide a form of theoretical validation as argued for in Self's Theoretical Foundations for Intelligent Tutoring Systems[12]. At a high level a set of simulated learners are tasked with answering exercise within Shufti. These simulated learners produce realistic answers to the exercises and over a simulated course slowly improve. Each learner also has preferences related to feedback, represented by a given feedback improving or worsening their performance. Each learners preferences for feedback are unique.

To use this as validation we then compared runs with feedback being possible to ones where it was not and we saw some significant gains in the simulated learners performance. With simulated learners without feedback achieving a course average of approximately 60% and ones with feedback achieving 77%.

With the validation in hand it is our hope that this system when placed online will become a useful tool for teaching the analysis of mammograms.

5 Conclusion

Medical Imaging is a large and growing field with significant utility. Unfortunately it is also difficult to train students in. The variety and difficulty of the cases results in a less than satisfactory outcome. Shufti uses gamification of learning, an extremely large exercise set, and adaptive feedback to remedy this.

References

1. B. S. Bloom. The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring, 1984.
2. R Crowley and O Medvedeva. SlideTutor: A model-tracing Intelligent Tutoring System for teaching microscopic diagnosis. *Artificial Intelligence in ...*, 2003.
3. Philippe Fournier-Viger, Roger Nkambou, and Engelbert Mephu Nguifo. Building Intelligent Tutoring Systems for Ill-Defined Domains. In *Advances in Intelligent Tutoring Systems*, volume 308, chapter 5, pages 81–101. 2010.
4. James Paul Gee. Deep Learning Properties of Good Digital Games How Far Can They Go? *Theories and Mechanisms: Serious Games for Learning*, pages 63–80, 2009.
5. M Heath, K Bowyer, D Kopans, R Moore, and P Kegelmeyer. The digital database for screening mammography. *Proceedings of the Fifth International Workshop on Digital Mammography*, pages 212–218, 2001.
6. Stuart Johnson. *Custom Feedback Selection for Intelligent Tutoring Systems in Ill-Defined Domains*. Msc, Univerity Of Alberta, 2016.
7. Stuart Johnson and Osmar R Zaïane. Deciding on Feedback Polarity and Timing. In *Education Data Mining*, 2012.
8. Alvin I. Mushlin, Ruth W. Kouides, and David E. Shapiro. Estimating the accuracy of screening mammography: A meta-analysis. *American Journal of Preventive Medicine*, 14(2):143–153, 1998.
9. Harold F. O'Neil, Eva L. Baker, and Ray S. Perez. *Using Games and Simulations for Teaching and Assessment*. Routledge, 2016.
10. Daniel Perelman and Dan Grossman. Test-Driven Synthesis for Automated Feedback for Introductory Computer Science Assignments Categories and Subject Descriptors. 2014.
11. Jorge G Ruiz, Michael J Mintzer, and Rosanne M Leipzig. The impact of E-learning in medical education. *Academic medicine : journal of the Association of American Medical Colleges*, 81(3):207–12, mar 2006.
12. John Self. Theoretical Foundations for Intelligent Tutoring Systems. *Journal of Artificial Intelligence in Education*, 1(45):3–14, 1990.
13. Teresa Del Soldato, Istituto Tecnologie, Didattiche Cnr, and Benedict Du Boulay. Implementation of motivational tactics in tutoring systems. *Journal of Artificial Intelligence in Education*, 6:337–378, 1995.
14. Tarja Susi, Mikael Johannesson, and Per Backlund. Serious Games An Overview. *Elearning*, 73(10):28, 2007.
15. R S Sutton and A G Barto. Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, 9:1054–1054, 1998.
16. Mary Ulicsak. Games in Education: Serious Games. *A FutureLab Literature Review*, page 139, 2010.
17. B. Van Ginneken, M.B. Stegmann, and Loog M. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Medical Image Analysis*, 10(1):19–40, 2006.
18. Kevin Yee. *Pedagogical Gamification: Principles of Video Games That Can Enhance Teaching*, volume 32. 2013.