Acquiring expertise in medical radiology through long-term interactions

Alexandre Direne, Marcos Sunyé,
Marcos Castilho, Fabiano Silva,
Luis Bona, Laura García
C3SL – Departamento de Informática – UFPR
Curitiba–PR, 81.531-980, Brazil
alexd@inf.ufpr.br

Donia Scott
Centre for Research in Computing,
The Open University, Walton Hall
Milton Keynes, MK7 6AA, UK
D.Scott@open.ac.uk

Abstract

This paper describes how cognitive and computational concepts can be applied to build interface and learner models for long-term tutorial interactions in medical radiology. Key human-to-human tutorial dialogue factors that typically occur in the different stages of skill acquisition are captured through an empirical study. The results of the study are detailed and linked to the design of RUI, an Intelligent Tutoring System for multiple domains of radiological expertise. A brief discussion and future research directions offer a comparative view of the method and tools.

1. Introduction

It has long been pointed out that one of the main bottlenecks for building adaptive systems is largely due to the lack of knowledge about the user's beliefs [2]. In particular, dynamic learner modelling has populated many research areas of Intelligent Tutoring Systems (ITSs) since updating long-term records accordingly is a major issue. The loss of synchronism between the human mental models and those of the system triggered a full array of related problems, such as the over-general student model [14]. Some have even proposed ways of managing tutorial dialogues while bypassing the intractable problem of student modelling [1].

Dynamic learner modelling differs from the more traditional view of user models of help systems in that the latter tends to consist of a short-term record [10]. Besides, user identity and evaluation is much more severe in learner modelling then in any other short-term user profiling mechanism.

On the one hand, Bull and Kay [2] argue that the main reason for the large and established learner modelling architectures and tools lies in that such systems include well founded cognitive principles for the definition and implementation of such an intricate internal machinery of ITSs. On the other hand, Murray [12] observes that authoring tools still lack appropriate resources for building and integrating the learner model with the other three models: interface, pedagogic and domain. More specifically, despite the apparent success of authoring systems, no one has so far addressed the connection between learner models and interaction models. Nor has any previous work accounted for the dynamic modelling of interface contents based on the underlying pedagogic directives that are appropriate for the long-term development of certain human skills.

The few attempts there have been on authoring tools for learner modelling did not reach a comprehensive insight on the multiple component features of expertise for long-term teaching purposes. One possible exception is the LRDC framework [8], even though it does not account particularly for the relationship between learner and interface features. If ITSs are to accommodate the development of the domain expertise of their users, the learner interface must be informed by detailed models of expertise - *i.e.*, learner and interface features and the way in which they progress over time. This paper describes how cognitive and computational concepts can be applied to build learner and interface models for long-term tutorial interactions.

2. Empirical study

We report here on an empirical study to assess the multiple, evolving skills of trainee radiologists. The study and consultations with expert radiologists were carried out in two schools of medical research [4]. In order to highlight primarily the novice-to-expert differences, we chose a case problem that demanded as much experiential knowledge as possible to reach a diagnosis, while keeping principled knowledge dependencies to a minimum. The case problem involved two often confused classes of abnormality: Ewing sarcoma and osteosarcoma. Only one medical problem, focusing on a confirmed diagnosed case of Ewing sarcoma,

1063-7125/08 $25.00 © 2008 IEEE
DOI 10.1109/CBMS.2008.34

was applied throughout the study to all trainees. The ultimate goal was for each trainee to reach a correct diagnosis; a task that first and second-year trainee radiologists often failed to accomplish.

The expert radiologist conducted tutorial dialogues with the trainees on a one-to-one basis, using a bottom-up teaching approach by allowing the trainees to begin with their own hypothesis (starting with scattered image features). In real tutorial dialogues, experts normally approach the trainee in a top-down fashion [5]. Despite this, we chose the former teaching approach since our focus here is not on optimising teaching styles, and since a bottom-up approach is more likely to reveal the reasoning behind the trainees' judgements. As a supplementary tutoring directive, the expert was asked to interfere fairly often to make trainees externalise their reasoning chain. This was expected to give trainees a better chance to exhibit consistency and completeness aspects of the diagnosis.

The study involved as subjects, sixteen junior doctors of varying levels as trainee radiologists: three in their first-year, six second-years, and seven third-years. We carried out a comprehensive analysis of the sixteen transcriptions of dialogues to identify the component features (skills) of expertise. Fifteen expertise features were observed from the dialogues. A more detailed description of evidence for each one is presented in a technical report [6].

3. Expertise features

This section presents details about the expertise features and their evidence found throughout the transcriptions of the dialogues. After the definition, a brief explanation about the teaching method is given for each feature of expertise in order to clarify the context under which the transcriptions have been annotated and analised.

All the data from the teaching approach came from extensive consultations with experts that actually teach radiology for a long time. As mentioned before, the teaching sessions have been carried out by the tutor on a one-to-one (tutor-learner) basis, focusing on the same case for all the trainees. It is a case of Ewing sarcoma, affecting a thirteen year old patient. The case is often confused with an osteosarcoma and requires differential diagnosis to be solved. The available images about the case are of two types: (a) conventional X-ray - two scans, a frontal and a bi-lateral one, covering the knee, calf and shins of both legs; (b) computerised Tomography (CT) - a set of scans covering the legs from the ankle up to the fist.

Due to the text size allowed, the following subsections present just two of the fifteen indicators divided by the expertise features introduced previously (a complete description of the indicators can be found in an early technical report [6]). In all the figures, "T" stands for tutor, "J1" stands for a junior doctor in the first year of training in radiology, "J2" a second-year and "J3" a third-year.

T: . . . But you have already detected the misalignment and the discontinuity of the cortical, even including an alteration of the bone axis. Don't you think this could suggest that there is a fracture somewhere there?

J2: No.

T: But the cortical extends all over here <POINTS TO SCAN>, gets interrupted, and follows this way. There is indeed a line of fracture here. What types of fracture do you know and how would you classify this one?

J2: Gosh! It is a fracture . . .

T: What types of fracture do you know?

Figure 1. Dialogue between tutor and a J2

3.1. Searching for barely visible features

In medical radiology, there is a widespread belief about the observation of subtle visual features that has been referred to in the literature as "you only see what you are looking for" [15]. As part of their experiential knowledge, experts are good at spontaneously looking for features that are not clearly/obviously visible. In fact, experts expect to see such features as a consequence of the major abnormality. This happens even when the features are faded to the most extreme conditions, in which case, the full skills of an advanced-level radiologist allow him or her to work out the diagnosis through top-down analysis first (from the major abnormality to the image findings). In other words, after they know what to expect from an initial hypothesis as to the pathology, experts individually search for the "already expected" visual features, according to their radiological signs (from the principles of radiology). This procedure very often ends with the confirmation of the diagnosis but, in a few difficult cases when an expected feature is not found, it may be necessary for the expert to change the initial hypothesis into a new pathology.

The training of this skill is based on the selection of cases that include such barely visual elements, which must also be relevant to the diagnosis. The imaging technique must also be considered as an important factor behind the idea of "barely visible" since a change of technique to a more appropriate one may be an important decision to be taught, particularly in cases where a feature becomes enhanced.

Figure 1 shows a fragment of a dialogue with a J2 who becomes very surprised to be pointed at a fracture (quite an important feature) that he had not noticed initially, even

T: . . . Do you think the tumor is also of the soft tissue?

J1: . . . I do not think so.

T: . . . How do you explain this alteration of the fibula? Do you notice that it is bent opposite to the lesion?

J1: Yes. . . . Perhaps there is also a problem affecting the soft tissue.

T: . . . What can you say about the trabeculae, density and width of the cortical of the femur?

J2: I think that there is a certain degree of osteopeny.

T: . . . and the osteopeny would be secondary in relation to what?

Figure 2. Dialogue between tutor and a J1

after being questioned about the possibility of finding one. Also, in both Figures 2 and 3, the questions posed by the tutor were similar but the replies were very different. The superiority of the J3's reply is quite clear in relation to that of the J1, or even in relation with that of the J2 (the J3 found the fracture without any help from the tutor).

T: . . . Assess the soft tissue?

J3: By comparison with the other leg, I can see that the alteration of the soft tissue is not prominent.

T: What can you say about the bone conditions, axis, shape . . . ?

J3: Drawing both the tibial and femural lines of the affected member, and comparing it with the normal one, I can see a slight variation of the inferior part . . . and a slight misalignment of the proximal portion of the metaphysis . . .

T: Right. Carry on.

J3: . . . there is also lowering of bone density . . . and the presence of structural disorder with bone material production . . . the presence of a fracture . . .

T: Correct. . . .

Figure 3. Dialogue between tutor and a J3

3.2. Selecting discrimination features

Unlike the ability of providing differential diagnosis, selecting discrimination features is related to the skill of assessing and choosing the relevant set of features that distinguishedly allows the categorisation of the case under one class of abnormality only. However, like differential diagnosis, the case under analysis can still be a difficult one, easily confusable with cases of other similar diseases. In the modern theories of classification found in Cognitive Psychology [7], such difficult cases correspond to the "less typical" members of a category. The reason for greater difficulty in classification is due to the fact that such less typical members share fewer (more scarce) features with the more typical (or representative) members of the category. This also results in misclassification since the case plausibly becomes an equally atypical member of others categories.

T: . . . The clinical data in this case includes pain on the calf and shins in the last two months.

J2: Just that? So, the pain can be due to the fracture only. The patient may have started this process much earlier without any pain. Do you have the exact date of the fracture?

T: No. Do you think this fracture is . . .

J2: I do not know if a fracture like this would calcify in two months . . .

T: In a normal bone!? But there is another thing here which suggests that this is a much longer process. As you have already classified the case as a tumor, do you think it makes a difference if you find out that the patient feels pain for two or six months in order to determine the type of tumor?

J2: No.

T: Fine. Also, there is no history of trauma.

J2: Could it be a history of stress?

T: No. But would you still consider, at all, the possibility of a trauma that led to an infection and then osteomyelitis?

J2: No, not any more, since the patient would have looked for assistance if it all had started from a trauma. . . .

Figure 4. Dialogue between tutor and a J2

Training for this skill can be covered by an archive of borderline cases [14]. Additionally, such cases should be constantly compared with counter-examples (as well as with normal cases) in an attempt to facilitate inductive learning. Available evidence from the cited work suggest that intermediate-level trainees (J2s) are those who need more help with discrimination features, given the difficulty of the cases that they are expected to solve.

Figures 4 and 5 show two different second-year junior doctors reacting in face of the need to judge discrimination features. In Figure 4, the difficulty was related with the exact type of the fracture (which was identified quite easily by the learner – but not so easily by J1s), whether it derived from trauma, stress or even from normal effort. Even under the hypothesis of a tumoral process, the learner failed

T: Can you tell me another feature that deserves analysis in relation to the soft tissue.

J2: They would be the anatomical lines. There are lines that could – <STOPS>

T: What do these lines consist of? <POINTS TO THE SCAN>

J2: This could be fat.

T: Right. Do they show up clearly on an X-ray?

J2: On an X-ray they would be darker.

T: Correct. Black lines, showing up between soft tissue layers. . . .

J2: We should try to determine the trajectory of such lines. They may be shifted or even deleted in an inflammatory process or in an oedema. In other cases, they may depart from the the bone plain due to a stroke, for instance.

T: Do tumors generally shift or delete the fat lines?

J2: I think a tumor tends to shift the fat lines.

T: Right. On the other hand, an inflammatory process tends delete the fat lines. So far, we have discussed the soft tissue in general terms. And in this specific case, how is it?

J2: In this case, I think it is normal. At least from what I can see . . .

T: What if you compare with the other leg.

J2: In relation to the other leg, I cannot define any enlargement of the soft tissue. In addition, it seems that the density of the leg I consider sound is slightly higher than that of the affected leg.

T: What is the influence of this difference?

J2: It is simply due to the imaging technique.

T: Let us carry on a bit further about that . . .

Figure 5. Dialogue between tutor and a J2

to consider the (normal effort) fracture as a relevant feature for confirming the diagnosis. In Figure 5, the difficulty was concerned with the detection and importance of the fat lines to discriminate between an inflammatory and a tumoral process. Furthermore, in this case, the tutor also had to recall principles of radiology, like the visual effect of radiodensity.

The trouble in dealing with discrimination features was indeed clear with all six J2s in our transcriptions. At the same time, the dialogue of Figure 6 with a J3 suggests a more thorough control of discrimination features, which also tended to happen with all the other six J3s.

T: . . . Now, describe the X-ray according to the classical sequence of tissue-bone analysis.

J3: Firstly, the imaging technique is adequate. I can see the bone . . . I can clearly see the fat lines in the muscle plain and . . .

T: Is it important to check whether the fat lines are preserved?

J3: Yes, because this will give me an idea of the infiltration of inflammatory or tumoral processes.

T: Is it common to find a neoplasic processes that delete the fat lines?

J3: It is more common for inflammatory processes to do that than the tumoral ones . . .

T: Correct. What else is common in a neoplasic process that affects the fat lines?

J3: They get dislocated.

T: Correct. And how is the soft tissue in this X-ray?

Figure 6. Dialogue between tutor and a J3

4. Interface-oriented authoring

A substantial part of the theoretical findings has been implemented in an existing shell, called RUI [5]. It consists of (1) an authoring language and tools for managing the complexity of ITS design, and (2) a domain-independent model of dialogue interpretation, integrated with the tools, for controlling adaptive tutorial interactions. Figure 7 shows a snapshot of the ITS shell interface loaded with a knowledge base for teaching about aortic aneurysm.

Before RUI, no domain-independent method for ITS authoring has focused specifically on the representation of visual concepts. Computer tutors for complex visual concepts differ from other tutoring systems in that the skills to be communicated to learners are closely linked to the interpretation of image patterns as a primary task. Such systems must therefore include facilities for learners to manipulate and display large stocks of visual images.

Likewise, design methods and tools for producing these tutors must provide experts with mechanisms for creating and assigning high-level, symbolic descriptions to such images as well as for defining pedagogic directives on how to teach their content.

Ongoing work in RUI is exploring and implementing a number of classes of interface objects, called ITWs (Intelligent Tutoring Widgets), so that domain experts can create long-term, system-active and system-passive tutorial interactions by directly manipulating such objects. This is achieved with the visual programming authoring tool which offers access to the more internal knowledge structures of

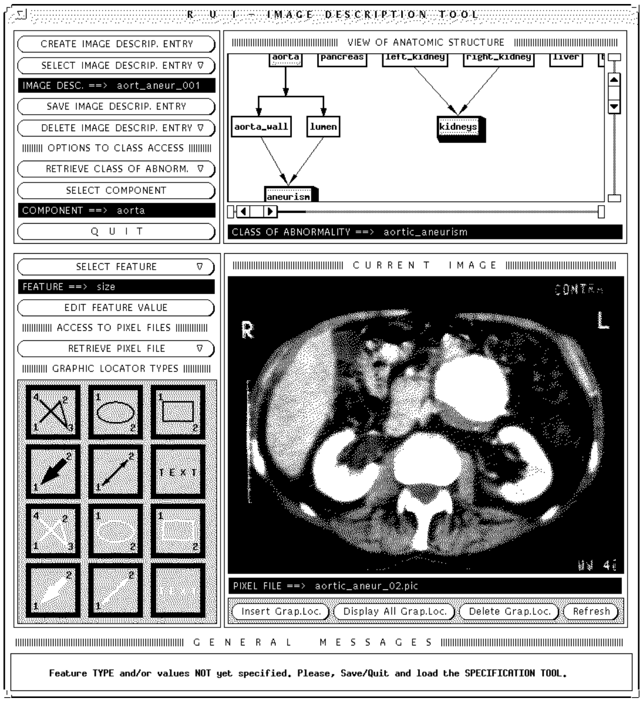

Figure 7. Snapshot of the ITS shell's interface

an ITS (domain knowledge, learner model and pedagogical directives) through the definition of the interface module.

5. Discussion

Because of the exploratory nature of most existing image browsing systems for radiology, they only permit learners to acquire diagnostic relationships inductively, rather than explicitly discussing these relationships during the interaction. However, knowledge acquired by induction is fragile [14] and, as a result, learners could be expected to show misconceptions even after using such systems. One source of diagnostic error is based upon the possibility of learners inferring facts about images which "fit in" a conceptual feature space, but in practice do not occur. That is, as instruction proceeds, learners' beliefs are expected to become "over-general", needing explicit (guided) interventions to avoid the problem.

To provide explicit advice on building diagnoses and thus help the development of expertise, RUI engages learners in Socratic-like dialogues. Socratic dialogues are meant to be rich dialogues, often based on natural language, where the computer poses a problem to the learner and both, together, "discuss" it. The learner may ask back another question for clarifying some facts he or she believes to be relevant. The computer can acknowledge correct assertions, send warnings about *inconsistent* points or even, if necessary, alter the course of the learner's work. The three most relevant characteristics of Socratic dialogues for the current research are:

1. There is only one source of belief, *i.e* the knowledge communication process is unidirectional and flows from the tutor to the learner. This is in accordance with the idea of the more *experienced* element in the conversation (tutor) guiding the less experienced one (learner).

2. The hypothetico-deductive approach of the conversation which consists of the learner extracting general principles from available evidence, found in image features, about the case under discussion. The learner is driven towards forming an initial hypothesis as to the case-problem, examine the validity of the hypothesis and search for further evidence that can support or contradict the hypothesis. The tutor can address a learner's incorrect response by treating it as a hypothesis and showing the consequences of the hypothesis. This approach becomes important to help learners when assessing classes of abnormalities, in a *data-driven* (forward) fashion where the system is capable of proposing hypotheses when solving diagnostic problems in a given domain of expertise.

3. There is fluidity of the dialogue. The tutor examines the steps in the learner's reasoning to make sure that the following question is an appropriate one to be asked, providing for continuity and consistency of the dialogue. This is particularly appropriate when learners have to develop their ability in gathering evidence for producing more complete diagnoses.

Despite the difficulties of implementation, these three characteristics make Socratic dialogues a suitable means for helping learners extend their skills in an *explicit* and *consistent* way. Many computer tutors that adopt Socratic dialogues for teaching in a variety of subject domains, such as the the Radiology tutor [13], the MENO-TUTOR [12], and Huang's approach [9] seem to use Socratic dialogues in equivalent forms.

RUI offers only a Socratic-like interface to simulate tutorial interactions. Although such a method of teaching does not provide for wider variations of teaching styles, if authored appropriately, it is capable of imitating quite elegant human-like tutorial behaviour exhibited by more experienced teachers. Currently, RUI does not allow the learner to ask questions to it but compensates the deficiency by incorporating techniques of differential diagnosis, similarly to ICON [15] and GUIDON [3] or even to tutors of non-medical domains like PROUST and its derivatives [11]. Such techniques allow the tutor to determine how a

RUI: Give a diagnosis for this case.
LEARNER: The brain is affected by the lesion. The lesion size is medium.
RUI: Quite right, there is a lesion in the brain. The lesion is of type meningioma. The lesion appears to be large. Which parts of the brain are affected by the lesion?
LEARNER: The cortical grey is affected by the lesion. · · ·
RUI: Indeed, the cortical grey appears to be affected by the lesion. · · · Do you notice any other indication of meningioma?
LEARNER: The lesion is necrotic.
RUI: That is not a feature of meningioma. It can be associated with glioma, for example. The image opposite shows a case of glioma where the lesion is necrotic. What do you think the arrows above indicate?

Figure 8. Dialogue with RUI about MR-scans

learner's response departs from the scope of the main class of abnormality and provides support for the tutor to construct an appropriate critique. Lines 13-18 of Figure 8 illustrate the present discussion. This is achieved through RUI's meta-level knowledge, enhanced with the activity of domain-specific teaching rules, allowing the system to monitor which abnormal features need special attention in the discussion while keeping focus on relevant information (see model of dialogue interpretation for details).

6. Conclusion

Finally, in the context of ITS evolution, although impressive results have been accomplished around prototype tools, most ITS authoring tools and tutoring architectures still remain as proposals. Aimed at different tutoring philosophies, these domain-independent systems place emphases on different aspects of the instructional process but invariably overlook interface-oriented authoring issues related to long-term learner models.

We are developing a facility for allowing an ITS created with RUI to have its interface easily changed (not only during the authoring phase but also dynamically, during a tutorial session). This will be based on information derived from knowledge of the learner model as well as the pedagogic model. However, new computational concepts are needed for authoring dynamic and versatile interface elements for ITS which can cope with the complex phenomenon of knowledge communication (from expert to machine and from machine to learner) and with the role of different media when designing learning tools.

References

[1] I. Arroyo, T. Murray, B. Woolf, and C. Beal. Inferring hidden learning variables from student help seeking behavior. In *Proc. of the 7th International Conference on Intelligent Tutoring Systems (ITS2004)*, pages 782–784. Springer, 2004.

[2] S. Bull and J. Kay. Student models that invite the learner in: The SMILI open learner modelling framework. *International Journal of Artificial Intelligence in Education*, 17(2):89–120, 2007.

[3] W. J. Clancey, editor. *Knowledge-based Tutoring: the GUIDON Program*. MIT Press, 1987.

[4] D. Cury, A. Direne, and N. Omar. Modelos baseados em e oráculos para a aprendizagem de conceitos visuais. *Revista Brasileira de Informática na Educação*, 2:18–29, 1998.

[5] A. Direne. Authoring intelligent systems for teaching visual concepts. *International Journal of Artificial Intelligence in Education*, 8(1):71–78, 1997.

[6] A. I. Direne and D. Scott. Identifying the component features of expertise in domains of complex visual recognition. Technical Report ITRI-01-23, Information Technology Research Institute, University of Brighton – UK, 2001.

[7] M. Estep, editor. *Self-Organizing Natural Intelligence: Issues of Knowing, Meaning, and Complexity*. Springer, New York, 2006.

[8] P. Gott and A. Lesgold. Competence in the workplace: How cognitive performance models and situated instruction can accelerate skill acquisition. In R. Glaser, editor, *Advances in instructional Psychology*. Lawrence Erlbaum, 2000.

[9] X. Huang. Inconsistent beliefs, attention, and student modelling. *Journal of Artificial Intelligence in Education*, 3(4):417–428, 1992.

[10] S. Lam, J. Pan, D. Sleeman, and W. Vasconcelos. A fine-grained approach to resolving unsatisfiable ontologies. In *Proc. of the IEEE/WIC/ACM International Conference on Web Intelligence (WI-2006)*, 2006.

[11] K. Luchini, C. Quintana, and E. Soloway. Design guidelines for learner-centered handheld tools. In *Proc. of the ACM Conference on Human Factors in Computing Systems (CHI2004)*, pages 24–29, 2004.

[12] T. Murray, B. Woolf, and D. Marshall. Lessons learned from authoring for inquiry learning: A tale of authoring tool evolution. In *Proc. of the 7th International Conference on Intelligent Tutoring Systems (ITS2004)*, pages 782–784. Springer, 2004.

[13] M. Sharples, N. Jeffery, B. du Boulay, B. Teather, D. Teather, and G. du Boulay. Structured computer-based training in the interpretation of neuroradiological images. *International Journal of Medical Informatics*, 60(3):263–280, 2000.

[14] M. Sharples, N. Jeffery, B. du Boulay, B. Teather, D. Teather, and G. du Boulay. Socio-cognitive engineering: a methodology for the design of human-centred technology. *European Journal of Operational Research*, 136(2):310–323, 2002.

[15] H. A. Swett. Computers: Power tool for imaging diagnosis. *Diagnostic Imaging International*, July/August:29–37, 1992.