Authoring adaptive tutoring systems for complex visual skills

Alexandre Direne, Luis Bona, Marcos Sunyé, Marcos Castilho, Fabiano Silva, Laura García
C3SL - Departamento de Informática
UFPR – Universidade Federal do Paraná
Curitiba, 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— **In the world of Intelligent Tutoring Systems research, the connection between learner models and interaction models has been largely ignored. Similarly, previous works have not accounted for the dynamic representation of interface contents based on the underlying pedagogic directives. The paper describes an approach to the authoring of training contents aimed at intermediate-level learners of medical Radiology. It is argued that, in certain dialogues designed for supporting visual diagnosis, it is worthwhile employing domain-general teaching mechanisms and reuse them in various domain-specific situations. An empirical study has been carried out with a corpus of human-to-human tutorial dialogues to identify the component features of expertise in medical Radiology. The results of the study form the basis for implementing interface and learner models that guide long-term tutorial interactions through an intelligent shell called RUI. Conclusions and future research directions are also described briefly.**

Keywords-Intelligent tutoring systems; authoring tools; learner models; interface models; visual expertise.

I. Introduction

This paper reports on the research steps developed to extend an existing Intelligent Tutoring System (ITS) to convey learner and interface models for long-term tutorial interactions. In domains of visual diagnosis, like medical Radiology, training methods take three to four years of practice [1]. Therefore, it is important that any computer-based system used to aid the learning process be capable of tracking users' performance and misconceptions over long periods of time. But this is still a largely unexplored field [2] that tends to rely on empirical studies of human-to-human technical dialogues [3] to accomplish consistency, completeness, granularity and many other factors that compose pedagogic knowledge bases [4].

Recent accounts of ITS architectures attribute a key role to interface design [5], claiming that knowledge should be communicated in different modalities (e.g., by applying a variety of tutorial styles and their corresponding presentation formats). This can be achieved by including cognitive principles in the design and behaviour of interface objects. Yet, there has been little effort in detailing the nature and content of these objects. It is indeed observed that the usability of graphical user interfaces relies to a large extent on the spatial arrangement of their visual elements [6], but the efficacy of any pedagogic directive depends critically on the meaningful connection between the interface objects and long-term learner models.

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 [7], even though it does not account particularly for the relationship between learner and interface objects.

Parallel to the field of learner modelling, the efforts progressed in computer-interface modelling to capture the semantics of the real world from users more naturally as well as to provide such users with adequate feedback. More specifically, Artificial Intelligence and Human Computer Interaction researchers have proposed several symbolic formalisms for external knowledge representation. One such method, WYSIWYM [8], explicitly tries to deal with complex conceptual input by means of a multi-language, feedback text generator to provide an authoring framework for defining knowledge bases. The method and the tools are organised around the idea of "what you see is what you meant" as the user is presented with natural language text to convey the intended meaning of standardised graphic input.

Except for some syntactic differences of knowledge descriptions, previous formalisms and tools offer similar notions for representing world concepts, their features and their classification. Because of that, just using an existing knowledge representation formalism is not sufficient to facilitate a human author when creating the interface of a complex, multi-model knowledge-based system, like an ITS. Indeed, the interface authoring process is a cyclic, time-consuming task [9]. It is often influenced by uncertainty in the way authors decide between one knowledge structure or another to represent an interaction model; i.e., several different representations may reflect the same model. Additionally, interface modelling is also characterised by an intuitive and empirical methodology. As a result, the quality of the representations obtained tends to be dependent upon the author's experience and insight in the domain as well as in interface design techniques.

However, even when authors manage to produce reasonable interface resources by using any modelling formalism and authoring tool, it does not follow that they will be able to keep such representations appropriate throughout long-term interactions [10]. This can be a problem in ITSs, for

978-0-7695-3711-5/09 $25.00 © 2009 IEEE
DOI 10.1109/ICALT.2009.166

example, since training systems are often meant to be used for long-term expertise development purposes rather than just for brief interactive help-like tasks. As the learner becomes more skilled in the subject domain, the system interface is expected to change dynamically, according to his or her evolving capacities, to convey new training needs. This raises the need for highly interactive, multimodal, adaptive systems that so far are not found in the literature.

II. Empirical Study

In our investigation, we choose the domains of complex visual categorisation and diagnosis, particularly those of medical Radiology. Medical diagnosis is a hard task, especially when combined with the perceptual abilities required for the recognition of detailed visual patterns. Despite this, expert radiologists (unlike novice trainees) can accurately identify major abnormalities in just a few seconds [11]. The study and the consultations with expert radiologists and junior doctors (trainees) have been carried out in two schools of medical research.

A. Aims

The aims of the study are two-fold:

- to confirm previously identified capacities of expertise reported in the literature;
- to determine any additional capacities not yet described or formalised in previous research.

  Briefly, this was done in accordance to a wide degree of freedom in terms of means and formats that, in addition to natural language, were utilised as inputs and outputs during the discussions. We conducted and recorded a series of training sessions between an expert radiologist and junior doctors. The design of the study was informed by extensive consultations with expert radiologists to determine the scope of abnormalities to train on, the most appropriate teaching approach to use and the subjects to be involved.

B. Methodology

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, was applied throughout the empirical 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 [10]. 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, 16 junior doctors of varying levels as trainee radiologists: three in their first-year, six second-years, and seven third-years.

C. Results

We carried out a comprehensive analysis of the 16 transcriptions of dialogues to identify the component features (skills) of expertise. Fifteen expertise features (briefly listed below) were observed from the dialogues. A more detailed description of evidence for each one is presented in a technical report [12]. Only the first six of these have been identified in previous research works [13]. The label given to each expertise feature reflects, in fact, the capacity of:

- mapping 2D-3D structures;
- quickly diagnosing cases;
- providing differential diagnosis;
- pinpointing discrimination features;
- recalling peculiar abnormalities;
- searching for barely visible features;
- viewing disease evolution;
- expressing conclusive justifications;
- explicitly reporting on the logical relations among features;
- accurately detecting disproportional features;
- consistently assessing symbolic relations;
- reporting the complete set of abnormal features in a scan;
- applying technical vocabulary;
- inferring totally invisible features;
- structuring the reasoning (and the reports).

III. Human-to-human Dialogue Analysis

This section presents a more detailed discussion about one of the fifteen expertise indicators (mapping 2D-3D structures) and its evidence found throughout the transcriptions of the dialogues. In the Figures 1, 2 and 3, "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. After the definition, a brief explanation about the teaching method is given in order to clarify the context under which the transcriptions have been annotated and analysed.

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.

As part of their principled (not experiential) knowledge, medical radiologists learn the anatomy of the human body in detail, during their undergraduate programme. However, mapping from 2D image features (e.g., of an X-ray) to 3D mental structure is a skill that needs to be developed by the trainee radiologist. In fact, anatomical abnormalities are easily identified by experts in 2D images because they distort and project 3D mental elements in many different ways to make them best fit onto the bi-dimensional space. Previous research shows how accurate this mapping can be for experts [7] when they were asked to sketch the stylised 3D structure of what they saw on an X-ray. On the other hand, for novices or even intermediate-level trainees, the sketches displayed a wide variety of mistakes, ranging from those caused by a displaced view of the projected 2D structure to those influenced by the wrong hypothesis as to the pathology.

J1: There seems to be an alteration of the bone higher up there.
T: Right. There is indeed an alteration of the bone. What area of the bone is affected?
J1: The head of the bone close to the joint <STOPS>
T: Right, very close to the joint. Can you identify the epiphysis, the metaphysis and the diaphysis?
J1: The epiphysis is this little head here, isn't it? It <REFERS TO THE LESION> is close to the head <STOPS-CONTINUES> it is enlarged.
T: What is the enlarged part here?
J1: The diaphysis.
T: No. It is the metaphysis. Is metaphysis white or black on X-rays?
J1: White.
T: Metaphysis is black on X-rays because it has got bone inside ...

Figure 1. Dialogue between tutor and a J1.

The training of this ability is based on the constant demand for discussing about the shapes and formats of 3D entities associated with 2D image regions. This is usually carried out by means of geometric conventions labelled with standard terms, but can also be done with the help of hand drawings.

During the empirical study, the need to discuss 2D/3D mapping details emerged several times for most trainees. Intuitively, it seemed even more intensively exercised by the tutor when interviewing second years (J2's). Figure 2 shows a dialogue between the tutor and a second-year (the tutor invites the trainee to label the bone production pattern which is only achieved by exclusion of possibilities) while in Figure 3 the tutor approaches a third-year (who spontaneously reported the feature, although, in a way that was not completely accurate). The dialogue in Figure 1 shows the tutor first checking principles of radiology (by asking about the J1's knowledge of anatomy for a much simpler 3D element if compared with the periosteal reaction asked to the J2 and the J3) and then testing the actual experiential knowledge of the 2D/3D mapping function. Although the J1 knows the principles of anatomy, she does not show enough experience to build a reply in this type of analysis.

T: ··· Depending on how aggressive the lesion is the periosteal reaction changes, creating varying patterns. In general, from the type of pattern of the periosteal reaction you can identify the process. For instance, is there a periosteal reaction at all in osteomyelitis?
J2: I do not know.
T: It is indeed possible ... However, periosteal reactions like onion peal, Codman's angle, spikelets or divergent spikelets tend to appear in more aggressive pathologies. In this case, you've already detected the periosteal reaction. How would you classify it?
J2: By exclusion, perhaps as Codman's angle.
T: Fine. Did you try to match the shape? And here <POINTS TO SCAN> what type of periosteal reaction is it?
J2: Would it be of another type?
T: Yes. There can be more than one ...

Figure 2. Dialogue between tutor and a J2.

J3: I can see a lithic lesion with sclerotic areas destroying the cortical of the bone with ... also the rupture of the cortical associated with the periosteal reaction of at least two types: Codman's angle and sun-rays.
T: Quite right. However, in this case we actually have Codman's angle and parallel spikelets in this region here. <POINTS TO SCAN> Do you agree?
J3: I do.

Figure 3. Dialogue between tutor and a J3.

IV. Authoring and Tutoring Tools

The RUI framework [10] 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. The authoring method is integrated with the model of teaching in the sense that courseware authors can externalise their expertise by using high-level formalisms (e.g., visual images, names of visual features and graphical overlays) which are similar to those presented to learners for them to acquire expertise. This integration is achieved by implemented prototype software tools that support the general method with domain-specific knowledge.

RUI's interface objects are called ITWs (Intelligent Tutoring Widgets). With them, domain experts can easily create long-term, system-active and system-passive tutorial interactions by directly manipulating such objects. This is achieved with a visual programming authoring tool which offers access to the more internal knowledge structures of an ITS (domain knowledge, learner model and pedagogical directives) through the definition of the interface objects. Figure 4 shows a snapshot of the authoring tool.

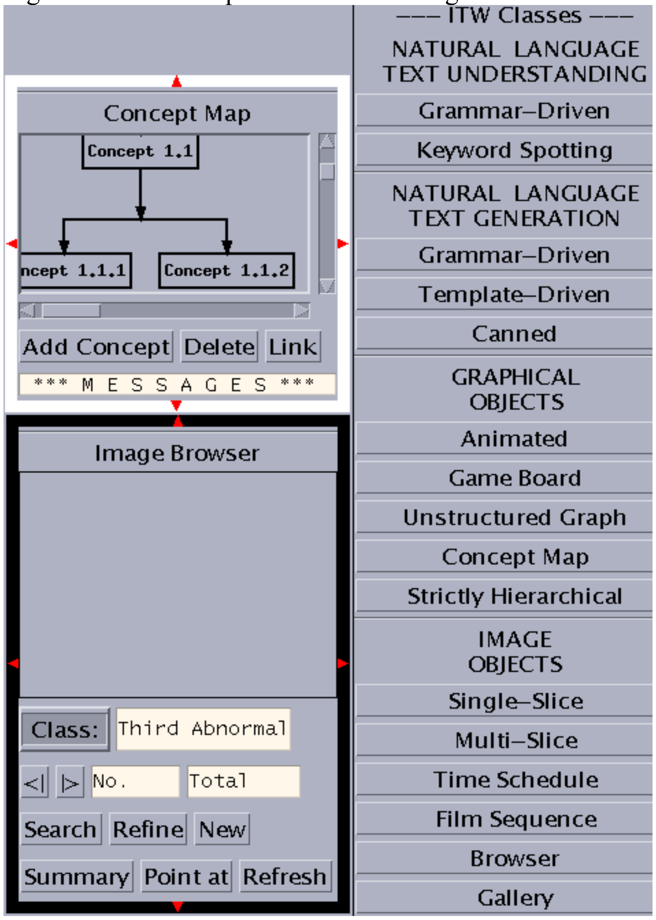

Figure 4. Snapshot of the authoring tool

The ITS shell is supported by a model of dialogue interpretation which is mainly controlled by domain-independent knowledge. More important, the dialogue interpreter is also capable of incorporating the effect of domain-specific teaching directives aimed at the long-term consistency of tutorial interactions.

A dual interface for integrating free exploration with guided tutorials is proposed here as a solution for bridging the gap between Lesgold's [7] two types of knowledge in visual domains: principled and experiential. We view the acquisition of principled knowledge by students through the interaction with an image database browser (a system-passive interface) while the acquisition of experiential knowledge is expected to occur through a system-active, ITS-like interaction mode. Figure 5 shows the ITS shell dual interface loaded with a knowledge base for teaching about meningiomas (a brain lesion).

V. Consistency and Completeness

One of the most basic constraints imposed on formal languages is consistency. Natural languages, on the other hand, lack such a quality for their sentences usually include alternate meanings of words (ambiguity) and little ordering of ideas. Besides that, as humans, we are also capable of reasoning with inaccurate and incomplete knowledge. However, consistency and completeness can become very important requirements if pseudo- natural language is to be used by computer tools like RUI. These two important characteristics, although extensively developed for the construction of knowledge bases, do not seem to be approached by existing authoring frameworks for tutorial dialogues. The only possible exceptions is Huang's logic of attention [14], in which consistency rules can be designed to avoid implausible knowledge states while delivering instruction.

Figure 5. Snapshot of the ITS shell's interface

A. Consistency of tutorial dialogues

To enforce consistency of more elaborate tutorial dialogues which evolve on top of teaching actions, domain-specific teaching rules are integral processing ingredients. Through the conditional clauses of such rules, RUI computes which body of action is appropriate to be executed at the moment in order to propagate diagnostic information. Such information carries, in textual form, the reflex of an expert's experience organised in well-defined knowledge units (production rules) located in specific nodes of a graph structure of anatomic components. After being refined during the design phase, the feedback provided to learners through these rules tends to display rich information along with how well they are doing over time or how contradictory their statements are.

Conditional clauses of teaching rules are logic expressions which have their truth values checked against the long-term overlay learner model as well as against the current sentence input by the learner. RUI then submits the conditional clause of candidate rules to its theorem prover for determining truth values. This logic proof mechanism is specifically adapted to work with the corresponding rule actions in the sense that, whenever existing interrelationships are violated by the learner, some action is performed towards bringing the learner's beliefs to a consistent state. In effect, it is actually this correspondence mechanism that carries the responsibility of linking the underlying formal proof of the logic expressions with the natural appearance of surface text, giving us the illusion that consistency resides in the interface language when in fact it emerges form a deeper structure.

B. *Completeness of tutorial dialogues*

To account for completeness of tutorial sessions, RUI once more makes use of the overlay learner model, but in a rather differential fashion as used in ACE and PIXIE [4]. For this purpose, every time a new input is supplied to the system, the current state of a learner's knowledge is compared with the whole symbolic description of the image being mastered. Based on that, and once more applying meta-level knowledge, RUI computes and reorders the remaining abnormal features that still need appropriate discussion. If domain-specific rules are not found in this completion process, default teaching actions are used instead to feed the question-slot of the interaction, thus giving continuity to the dialogue. Therefore, we say that default actions are to dialogue completeness as domain-specific rules are to dialogue consistency.

An interesting effect resulting from the above discussion is that, after a learner has seen a reasonable number of example images belonging to the same class of abnormality, it is expected that all the relevant abnormal features related to the class will be learned. Some example images of a class might display very little in common with others which are more frequently seen but, after a large stock has been inspected, the learner is expected to have inductively acquired the prototypical view of the class in question. Although we do not make use of explicit statistical knowledge, this is not in disagreement with more recent theories of concept formation [7] which states that an idealised or abstracted view of a class is based on a combination of concepts assimilated from many typical features of instances of the class. However, no single instance of a class may exhibit all the idealised features.

VI. Conclusion

The method outlined provides a formal approach for structuring domain knowledge and mechanisms for tutoring about symbolic image descriptions of abnormal features guided by a general model of dialogue interpretation. Conceptual and computations aspects of the method and software tools have been discussed, highlighting solutions for specific problems like dealing with long-term learner beliefs and enforcing consistency of tutorial dialogues. The authoring and tutoring prototype tools are implemented and allow the development of a range of ITSs for complex visual domains.

To substantiate generality claims about RUI's ITWs, we have combining the evaluation of the software tools with the definition of knowledge bases for four domains of visual expertise: (1) chest X-rays, (2) MR-scans of brain lesions, (3) MR-scans of cerebellum tumors, and (4) CT scans of aortic aneurisms in the abdomen. The empirical studies have concentrated on the potential of human-to-human dialogues to reveal the capacities of expertise in visual domains. The positive results obtained so far suggest the suitability of the framework for the development of novel ITSs based on RUI's ITWs - i.e., a deeper integration of interface and learner models to guide future research.

Acknowledgment

This research and development project was partly sponsored by Fundação Araucária/PR and by CNPq Brazil. Travelling expenses for presenting this paper at the conference have been funded by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

References

[1] Sharples, M., Jeffery, N, Boulay, B. Teather, D. Teather, B. Boulay, G. Structured computer-based training in the interpretation of neuroradiological images. International Journal of Medical Informatics, 136(3):263-280, 2002.

[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] S. Ainsworth. DefT - A conceptual framework for considering learning with multiple representations. Learning and Instruction, 16: (3), 183-198, 2006.

[4] D. H. Sleeman. PIXIE: a shell for developing intelligent tutoring systems. In R. Lawler and M. Yazdani, editors, AI and Education: Learning Environments and Intelligent Tutoring Systems. Ablex Publishing, 1987.

[5] A. Collins, P. Neville, K. Bielaczyc. The role of different media in designing learning environments. IJAIED, 11(1): 144-162, 2000.

[6] A. Ravenscroft, R. M. Pilkington. Investigation by design: developing dialogue models to support reasoning and conceptual change. International Journal of Artificial Intelligence in Education, 11: 273-298, 2000.

[7] S. Katz, A. M. Lesgold, E. Hughes, D. Peters, G. Eggan, M. Gordin, L. Greenberg. Sherlock 2: An intelligent tutoring system built on the lrdc framework. In C. P. Bloom and R. B. Loftin, eds, Facilitating the development and use of interactive learning environments. Lawrence Erlbaum, 1998.

[8] R. Power and D. Scott. Multilingual authoring using feedback texts. In Proceedings of the 17th International Conference on Computational Linguistics (COLING/ACL-98), pp 1053-1059, 1998.

[9] I. Arroyo, T. Murray, B. Woof, and C. Beal. Inferring hidden learning variables from student help seeking behavior. Proceed. of the 7th International Conference on Intelligent Tutoring Systems (ITS-2004), 782-784, 2004.

[10] A. Direne, M. Sunye, M. Castilho, F. Silva, L. Bona, L. Garcia, and D. Scott. Acquiring expertise in medical radiology through long-term interactions. Proc. of the 21th IEEE International Symposium on Computer-Based Medical Systems (CBMS2008). IEEE Press, 403-408, 2008.

[11] M. Myles-Worsley, W. A. Johnston. The influence of expertise on x-ray image processing. Journal of Experimental Psychology: Learning Memory and Cognition, 14(3): 553-557, 1988.

[12] A. Direne, 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.

[13] D. Cury, A. Direne, and N. Omar. Modelos baseados em oráculos para a aprendizagem de conceitos visuais. Revista Brasileira de Informática na Educação, 2(1):18-29, 1998.

[14] X. Huang. Inconsistent Beliefs, Attention, and Student Modelling. Journal of Artificial Intelligence in Education, 3(4): 417-428, 1992.