Friendly e-tutor for Natural Deduction

4 authors, including:

K. Broda
Imperial College London
117 PUBLICATIONS   **1,195** CITATIONS

SEE PROFILE

Jiefei Ma
Imperial College London
28 PUBLICATIONS   **365** CITATIONS

SEE PROFILE

Friendly e-tutor for Natural Deduction

Krysia Broda, Jiefei Ma, Gabrielle Sinnadurai, Alex Summers
Department of Computing, Imperial College London
{*kb,jm103,apgs,ajs300m*} *@doc.ic.ac.uk*

Abstract

Pandora is a tool to support the learning of first order natural deduction. It includes a help window, an interactive context sensitive tutorial known as the "e-tutor" and facilities to save, reload and export to LATEX. Every attempt to apply a natural deduction rule is met with either success or a helpful error message, providing the student with instant feedback. This paper describes the e-tutor and our experiences of using the tool in teaching.
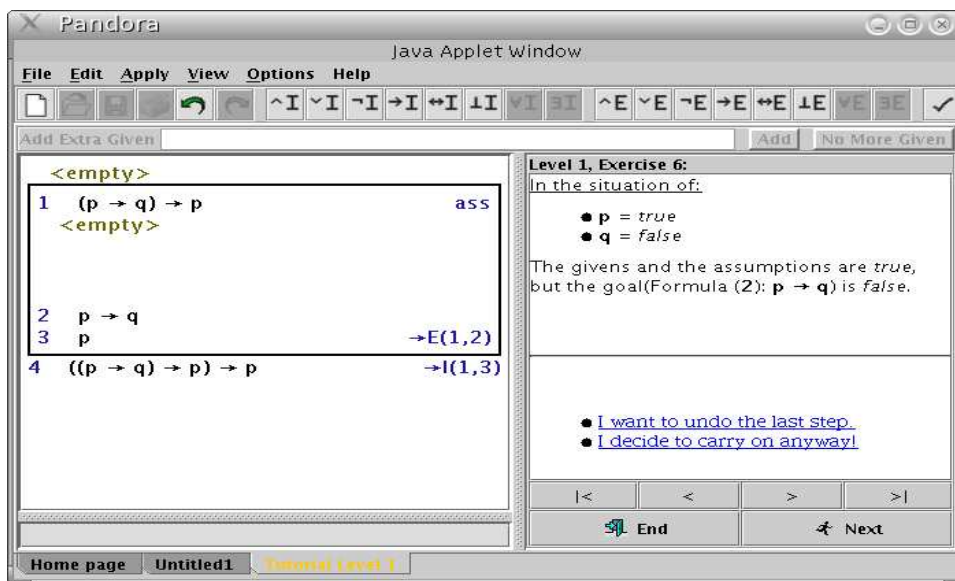
Keywords: natural deduction, Fitch box proof, first order logic, predicate logic, teaching, learning, e-learning

1. WHAT IS PANDORA?

Our experience in teaching Natural Deduction to first year undergraduates in the Department of Computing at Imperial College since 1991 confirms the folklore that students find formal proof arcane and difficult, even when used informally in discrete mathematics proofs. *Pandora* (Proof Assistant for Natural Deduction using Organised Rectangular Areas) is a learning support tool designed to guide the construction of natural deduction proofs. It was conceived in 1996 and is based on the Fitch-style "proof box notation" for natural deduction as described in [1]. There have been three versions of Pandora, all of which have been implemented by undergraduates as individual or group projects. Version 1 was written in Tcl/Tk and made available to students but was not robust enough for use in lectures. Version 2 was written in early Java; it was robust enough for class use and students were encouraged to use it. Version3, written in Java also, is the current version and is described in this paper. It has a context sensitive tutorial, a help facility and various facilities for saving, loading and printing proofs and has been extensively used in class and by students for coursework and tests. An applet version is available for use via the Pandora web site [5] and we hope to make it available for download in the near future.

Pandora provides learning support to guide students in their construction of a natural deduction proof of a conclusion or goal from given premises. It allows the user to reason "forwards", that is, from one or more given formulas to deduce another formula using one of the rules, and to reason "backwards", that is, to reduce one of the current goals to one or more subgoals from which the current goal can be deduced using one of the rules. The rules are all the usual introduction and elimination rules of first order natural deduction with equality plus a few derived rules found useful by more advanced users.

In normal mode Pandora just responds to the student's actions by either applying the rule they request or giving appropriate error messages. The student's strategy is not checked and if a student applies a rule correctly but unwisely, turning an achievable goal into an unprovable goal they are not warned. When an exercise is attempted in the e-tutor the level of support is greatly enhanced. Hints and explanations are provided, a warning and counter-model are given if the student applies a rule which creates an unprovable goal, and the friendly e-tutor will simply do a step for the student if asked.

In a recent contribution to an international congress on tools for teaching logic [2] we presented the basic functionality of the current version of Pandora and briefly described the e-tutor and our evaluation of the tool. Here we describe the e-tutor in more detail and further discuss our experience of using Pandora in teaching.

2. THE E-TUTOR

On loading Pandora, the user is presented with the Pandora home page, which offers various options, including loading a previously stored proof, starting a new proof or starting an e-tutorial. When the e-tutorial is started, four tutorials of propositional exercises are available. The first consists of fixed exercises and is useful in teaching when we demonstrate the tool in class or run a laboratory session in which all the students are required to attempt the same exercise. The other three, known as "easy", "medium" and "hard", consist of sets of five exercises which are randomly selected at run time by the e-tutor.

We will explain the use of the e-tutor by following a hypothetical student's attempt to derive Pierce's law $((p \to q) \to p) \to p$ which is one of the fixed tutorial exercises. In the e-tutor the view is split into a left window in which the proof is constructed and a right window in which the e-tutor offers advice. Initially the proof window just contains the goal $((p \to q) \to p) \to p$ and the advice window just contains the remark "I am always here to help you" and a clickable link "I need some advice". Suppose our student clicks the link. The tutor suggests "Arrow introduction rule" and the suggestion is itself a clickable link to an explanation of the rule. On clicking this link the student is given an explanation of the usual use of the rule and what will happen if it is applied in this particular case and another clickable link "How do I do it?". On clicking this link the student is told exactly what to click on to apply the rule and given two more clickable links. In this case the advice is to select the goal, $((p \to q) \to p) \to p$, and click on the $\to\mathcal{I}$ button. The links are "Do the steps for me!" and "Show the advice again". When our lazy student selects the former the proof window is updated as shown in the proof state below and the advice window goes back to showing the remark "I am always here to help you" and a clickable link "I need some advice". The box in the new proof state shows the scope of the assumption $(p \to q) \to p$ which is discharged by the $\to\mathcal{I}$ rule and the conclusion $((p \to q) \to p) \to p$ has the justification $\to\mathcal{I}(1,2)$. The new goal is p.

$$
\begin{array}{ll}
 & < empty > \\
\boxed{\begin{array}{ll}
1 & (p \to q) \to p \quad \text{ass} \\
 & < empty > \\
2 & p
\end{array}} \\
3 & ((p \to q) \to p) \to p \quad \to\mathcal{I}(1,2)
\end{array}
$$

Our student can again ask for advice and this time will be offered the choice of "Law of Excluded Middle" or "Proof by Contradiction". Again these are clickable links to explanations of the rules and on selecting them further offers of links "How do I do it?" and "Do the steps for me!" would

be supplied as above. But suppose our overconfident student reasons that the goal p could be proved by $\rightarrow\mathcal{E}$ using $(p \rightarrow q) \rightarrow p$ if $p \rightarrow q$ can be proved and so decides to ignore the advice and instead applies the $\rightarrow\mathcal{E}$ rule backwards by clicking on the goal p then the $\rightarrow\mathcal{E}$ button then the assumption $(p \rightarrow q) \rightarrow p$. The friendly e-tutor allows the rule to be applied in the proof window but responds with a warning in the advice window "Be careful, one or more of your goals are not provable!" and offers the links "Why?", "Undo last step" and "I want to carry on anyway". If the student asks why, the tutor provides a counter-example truth value assignment as shown in the screenshot on the first page. If the student decides to carry on anyway the tutor reports that "No advice is available at the moment because the goal is not provable".

Suppose our student wisely decides to undo the incorrect step and apply Excluded Middle (EM) as advised instead. This is easily achieved - click on $< empty >$ then the EM button then enter p in the text box - so our hypothetical student is gaining confidence and continues without asking for more advice to apply $\vee\mathcal{E}$. The proof state below shows the result of these steps.

$$< empty >$$

1	$(p \rightarrow q) \rightarrow p$	ass
2	$p \vee \neg p$	EM

$< empty >$

3	p	ass	5	$\neg p$	ass
	$< empty >$			$< empty >$	
4	p		6	p	

7	p	$\vee\mathcal{E}(2,3,4,5,6)$
8	$((p \rightarrow q) \rightarrow p) \rightarrow p$	$\rightarrow\mathcal{I}(1,7)$

The proof now has two goals so if we ask the tutor for advice it first asks us "Which goal would you like advice for?" then continues as outlined above. The left-hand box can be completed using the tick rule, which allows the goal p to be proved by selecting p (line 3) from the current context. The box will turn grey, to indicate that part of the proof is complete. The e-tutor just sits quietly presenting the link "I need some advice" unless the student either clicks that link or makes a move which results in an unprovable goal.

In the right-hand box, the student can now apply $\rightarrow\mathcal{E}$ backwards in the way they tried before without the tutor needing to intervene because, now that we have the assumption $\neg p$, the goal $p \rightarrow q$ is provable. Applying the $\rightarrow\mathcal{I}$ rule to the resulting goal $p \rightarrow q$ gives a new assumption p, so now we have both p and $\neg p$ in scope so we can deduce $\bot$ by the $\neg$ elimination rule. To do this in Pandora we select the $< empty >$ line in the left box, then click the $\neg\mathcal{E}$ button, then click on the two formulas which give the contradiction. The resulting proof state is shown below.

$$< empty >$$

1	$(p \rightarrow q) \rightarrow p$	ass
2	$p \vee \neg p$	EM

$< empty >$

3	p	ass	5	$\neg p$	ass

6	p	ass
7	$\bot$	$\neg\mathcal{E}(5,6)$
	$< empty >$	
8	q	

9	$p \rightarrow q$	$\rightarrow\mathcal{I}(6,8)$
4	p	$\checkmark(3)$	10	p	$\rightarrow\mathcal{E}(1,9)$
11	p			$\vee\mathcal{E}(2,3,4,5,10)$
12	$((p \rightarrow q) \rightarrow p) \rightarrow p$			$\rightarrow\mathcal{I}(1,11)$

Finally, we can use $\bot\mathcal{E}$ backwards followed by the tick rule again to note that we have $\bot$ and we need to show $\bot$. The proof is now complete. Pandora removes all the empty lines and greys

out the proof and is intelligent enough to remove the last application of the tick rule which is not needed in the final proof. Note that the line numbers and references to them in the justifications were consistently updated as the proof emerged. Note also that in the completed proof every line has a justification.

$$
\begin{array}{lll}
1 & (p \to q) \to p & \text{ass} \\
2 & p \lor \neg p & \text{EM} \\
\end{array}
$$

1	$(p \to q) \to p$	ass
2	$p \lor \neg p$	EM
	3 p ass 5 $\neg p$ ass	
	6 p ass	
	7 $\perp$ $\neg\mathcal{E}(5,6)$	
	8 q $\perp\mathcal{E}(7)$	
	9 $p \to q$ $\to\mathcal{I}(6,8)$	
	4 p $\checkmark(3)$ 10 p $\to\mathcal{E}(1,9)$	
11	p $\lor\mathcal{E}(2,3,4,5,10)$	
12	$((p \to q) \to p) \to p$ $\to\mathcal{I}(1,11)$	

3. USING PANDORA IN TEACHING

We have two different cohorts of students learning logic in their first year. One group (JMC) is small (20-25 students) and studying for a joint Mathematics and Computer Science degree, whereas the other class (COMP) is large (over 100 students) and studying for a Computing degree. The treatment in respect of using Pandora was different for the two cohorts.

The JMC group is taught by an enthusiastic lecturer, has enthusiastic tutors, and uses Pandora in an integrated way to teach natural deduction. Initially, the propositional natural deduction rules are presented and hand written examples of proofs are given. Only after they have seen several proofs and tried a few on paper themselves is Pandora introduced. Pandora is demonstrated using several of the same examples so the students can focus on how to drive Pandora rather than on how to prove the theorems. Over the next few weeks, in the lab and tutorials, we give the students many exercises, some assessed, and introduce them to the first order and equality rules. The course finishes with a "driving test" consisting of ten problems which the students have one hour to attempt under exam conditions. Last year the test problems were:

1. $\vdash (p \to (q \to r)) \to ((p \to q) \to (p \to r))$
2. $\neg p \lor q \vdash (p \to q)$
3. $\neg(p \to q) \vdash p \land \neg q$
4. $p \to (q \to (r \lor s)), \neg(r \lor t), (s \land q) \to t \vdash q \to \neg p$
5. $\forall x \neg p(x) \vdash \neg \exists x p(x)$
6. $\neg \exists x p(x) \vdash \forall x \neg p(x)$
7. $\neg \forall x p(x) \vdash \exists x \neg p(x)$
8. $\neg \exists x \exists y (\neg x = y) \vdash \forall x \forall y (p(x,y) \to p(y,x))$
9. $\forall x (x = a \lor x = b), \neg(a = b), g(a) = b, \forall x \forall y (g(x) = g(y) \to x = y) \vdash g(b) = a$
10. $\forall x \forall y \forall z (r(x,y) \land r(y,z) \to r(x,z)), \forall x (r(x,a) \lor r(x,b)), r(a,b) \vdash \exists y \forall x r(x,y)$

Pandora has been enhanced with various features to facilitate automatic marking. For assessed coursework we provide the students with exercise "skeletons". These include the given premises and conclusion together with a "magic number" which encodes any restrictions we care to impose on the rules they are allowed to use. For example, this feature allows us to disable certain derived rules (eg the PC rule) or to disable all the predicate rules which may be a distraction when they are proving a propositional assertion. The magic number also encodes the current state of the proof so the students cannot take the number from one proof and put it on another. The students download the skeletons from our web based continuous assessment tracking system and the first time they save a proof (usually when they have completed it!) their identity is also coded into the magic number. Whenever the proof is saved subsequently, the coded identity is unchanged so the first saver can be checked against the submitter of the proof. We are pleased to say that for the two cohorts we have checked there was no evidence of "work sharing" at all even though the

students did not know about this feature. The electronically gathered proofs from both the driving test and assessed coursework are checked for correctness and converted to LaTeXby "text to text" command line programs included in the Pandora package. We can thus produce a report for each student and a summary of results for their tutors with minimal human intervention.

The experience of the larger class (COMP) has been variable over the last three years. The first year Pandora version 3 was available its use was not encouraged, with the result that few students tried it. In the second year there was light encouragement and a demo, and more students tried it out, some using it to do their coursework. Only in the third year of its availability was Pandora seriously encouraged, but there was still no driving test and no requirement to use it for assessed coursework.

For both cohorts, all exercises done throughout the term, including the driving test for the JMC, are essentially formative. The main summative assessment is an end of year written examination.

4. EVALUATION

We have put considerable effort into evaluating Pandora and used three methods.

Firstly, we asked the students what they think of it, both verbally and using anonymous feedback forms. The feedback is generally encouraging and students say and write that they enjoy using Pandora and find it useful. The JMC classes, who learned Pandora thoroughly, enjoyed using Pandora more and gave more encouraging feedback than the COMP cohort. Moreover, the feedback from the COMP cohort improved as Pandora was more actively encouraged. We concluded that perhaps students tell us what they think we want to hear! Students also made useful criticisms: for example the first release of the current version only had the facility to undo the last rule application but the facility to repeatedly undo steps was added by popular demand.

Secondly, we compared performance on the written exam by the two cohorts. This did not give the clearcut advantage to Pandora users that we hoped for and in fact there was little difference in terms of marks between the two cohorts. There did, however, seem to be a difference in style, namely that those who used Pandora were much more at home with using rules backwards and did not make "arbitrary" assumptions which they had no hope of discharging whereas the cohort who did not use Pandora mainly reasoned forwards and frequently made arbitrary assumptions. We had feared that Pandora users may find it hard to adapt to writing proofs by hand but it turned out that the users were more precise syntactically in their hand written proofs than were the non-users.

For the third evaluation method we electronically recorded detailed logs of the students' use of Pandora; essentially we recorded every "click" they make so that we can see in detail how they actually use it. This allowed us to see which rules were applied, whether undo, help or the tutorial were used, and so on. We computed proportions of attempted rule applications that were correct and counted error rates for various types of errors. Some of the results from analysing the logs came as a disappointment in that they showed that the help and tutorial facilities were little used. The logs also showed a surprisingly high failure rate in students' attempts to apply the rules. A small number of students had virtually no failures but many had almost as many failures as successes. Analysis of the logs showed that many students were not selecting the $< empty >$ line or a goal line before applying a rule. Comparing the logs for the driving test with those for the previous work we were pleased to observe that, with experience, the proportion of failed rule applications decreased. The logs yielded detailed information about the common errors made by students for each rule but, perhaps surprisingly, not many general problems could be diagnosed. The one fact that really jumped out of the data was that students frequently tried to apply the $\neg\mathcal{I}$ rule backwards to a formula which was not a negation, whereas it was comparatively rare for them to try applying $\rightarrow\mathcal{I}$ backwards to a formula which was not an implication. We believe they were confusing the $\neg\mathcal{I}$ rule with the derived PC rule.

Overall the evaluation has taught us that Pandora is well liked and considered useful by the students, to the extent that they no longer find formal proof arcane and difficult. To improve the learning outcomes we have modified our teaching in a number of ways including:

- giving an advertisement for the tutorial and help as part of the initial demonstration
- when demonstrating the individual rules emphasising that either the $< empty >$ or goal line needs to be selected **before** clicking the rule application button
- explaining how to avoid what the logs show to be the common pitfalls in applying the rules.

5. RELATED AND FUTURE WORK

Pandora is quite similar in appearance to another tool for teaching natural deduction using Fitch-style boxes developed at Carnegie Mellon by Wilfried Sieg [6, 9]. A similar tool for building Gentzen Type tree proofs has also been developed by Rein Prank [7] though his interface looks rather different as it does not use boxes. A fairly comprehensive list of such educational logic software is available on Hans Van Ditmarsch's web site [4]. To our knowledge, our use of an e-tutor is novel, and no-one else has analysed student's use of the tools quantitatively using logs.

We have two plans for future extensions to the tool. The first is to extend it to allow unification and the use of free variables in proofs. This will enable a demonstration of how the execution of logic programs can be viewed as a special case of natural deduction in contrast to the more usual explanation in terms of resolution. The second, and more significant extension, is to incorporate reasoning about programs by adding new rules to implement the rules of Hoare logic. We have prototyped the idea using the Isabelle generic theorem prover [3] and will simplify the implementation of the first version by providing a "Trust Me" rule which students can apply for arithmetic steps. Applications of this rule will be accepted by the tool if some simple semantic plausibiliy checks are passed.

REFERENCES

[1] K. Broda, S. Eisenbach, H. Khoshnevisan and S. Vickers. *Reasoned Programming.* Prentice-Hall, 1994.

[2] K. Broda, J. Ma, G. Sinnadurai and A. Summers. Pandora: Natural Deduction made Easy, in [8], pp 11-14.

[3] `http://www.cl.cam.ac.uk/Research/HVG/Isabelle/`

[4] `http://www.cs.otago.ac.nz/staffpriv/hans/logiccourseware.html`

[5] `http://www.doc.ic.ac.uk/pandora/`

[6] `http://www.phil.cmu.edu/projects/apros/`

[7] R. Prank. Trying to cover the exercises with reasonable software, in [8], pp 149-152.

[8] *Proceedings of the second international congress on tools for teaching logic*, Eds: Maria Manzano, Belen Perez Lancho and Ana Gil, Salamanca, 2006, ISBN: 84-690-0348-8.

[9] W. Sieg. STRATEGIC THINKING: web-based and dynamically tutored, in [8], pp 199-200.