Persuasive skill development: On computational surveillance strategies for modeling learning statistics with PLOT Learner

Judith Gottschalk,  
Aalborg University, Nyhavnsgrade 14, 9000 Aalborg  
gottschalk.judith@gmail.com

Nicolai Winther-Nielsen  
Fjellhaug International University College Denmark, Frederiksborggade 1.B.1,  
DK-1360 Copenhagen K  
nwm@dbi.edu

The hypothesis this paper is based on is that computational surveillance within computer-assisted language learning [CALL] systems helps teachers to give their students detailed feedback and enables the computational system to give response to students. We discuss this hypothesis and give as example research work in progress on PeLLITS, short for a Persuasive Intelligent Tutoring System, which is an extension of PLOT Learner. As a result of the data mining it is possible to derive and model learning statistics by using the formal kernel of Item Response Theory [IRT] which has been proposed by Rasch in 1960 and is now used in a paper by Metsämuuronen from 2013. This way we integrate a feedback-loop in our CALL system to achieve a behavior change in the student. Surveillance and tailoring are essential to persuasive learning with our CALL system as this customization supports the application of modern learning theories in which communication and feedback is crucial, following the theory of leaning design formulated by Laurillard (2012). With the implementation of IRT in our CALL system, it is possible to predict the students' learning progress and to give the teacher the opportunity to gently guide student at the right point and with the best feedback. We also discuss the ethical aspects. The computational implementation of this additional module to the online application of PLOT Learner is supervised by Claus Tøndering, developer of PLOT Learner, who is the executive programmer and trains the lead author of this paper in implementing the model developed in this paper.

Keywords: Computer-assisted language learning, surveillance, tailoring, feedback, learning statistics, IRT, learning progress predication, persuasion

1 Introduction

Measuring learning statistics through computational surveillance and derived via data mining algorithms can support persuasive teaching in computer-assisted language learning [CALL] systems. This is the hypothesis this paper deals with. We discuss how monitoring and predicting learning progress is being used in research work in progress on an intelligent tutoring system called Persuasive Language Learning Intelligent Tutoring System [PeLLITS] which is an extension of the database-driven CALL environment PLOT Learner (cf. http://bh.3bmoodle.dk/).

Currently PLOT Learner is being repurposed as an online application, Bible Online Learner (http://pltest.3bmoodle.dk/), developed by programmer Claus Tøndering. In the new project we use logged data on learner performance from this repurposed version of PLOT Learner as input for modeling learner statistics to be used by students, peers and teachers. The ultimate goal of the new project is, however, to emulate the presence of an artificial tutor as a learning supervisor in a virtual and interactive world. The development of PeLLITS is still in the beginning; its architecture has been developed in Gottschalk (2012), where it is also described in length.

The current status of PeLLITS is: It has a web interface for facilitators and learners using PLOT Learner which we call Learning Journey Online. Various data mining algorithms and the formal IRT-based framework used by Metsämuuronen (2013) will enable the facilitator to in
gathering information on the learning progress of a student. This is statistically evaluated to give an overview of the students learning progress. The system can show the learning progress in detail, using graphs and tables. It can give advice on where the student needs further support and where he or she should focus on in grammatical drills, and it can predict the students learning progress as shown in the screen-shot of the Learning Journey Online in figure 1. The surveillance module collects data when learners choose to submit their learning statistics to the system. Based on the learning statistics, the facilitator can give a direct and informed feedback to the learners, and learners can direct their own learning project through the feedback from the system. This way they are able to tailor their own learning course via surveillance. The generated feedback loop will effect a change of behavior towards improved learning skills, which is one of the persuasive functions identified by Fogg (2003).

<table>
<thead>
<tr>
<th>Filename</th>
<th>Start at</th>
<th>Duration (min/sec)</th>
<th>Seconds per right</th>
<th>Correct</th>
<th>Wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>VocabularyFrequency1-718-FlUCl3et</td>
<td>2012-12-18 15:48</td>
<td>2:57</td>
<td>48.6</td>
<td>16</td>
<td>51</td>
</tr>
<tr>
<td>VocabularyFrequency1-220ToDtn73et</td>
<td>2012-12-18 15:48</td>
<td>50:14</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>VocabularyFrequency1-718-FlUCl3et</td>
<td>2012-12-18 14:01</td>
<td>25:21</td>
<td>0.8</td>
<td>73</td>
<td>183</td>
</tr>
<tr>
<td>VocabularyFrequency1-718-FlUCl3et</td>
<td>2012-12-17 13:59</td>
<td>33:51</td>
<td>16.4</td>
<td>124</td>
<td>232</td>
</tr>
</tbody>
</table>

Quick overview on student Christian Thomsen

You won 347.92 PLOTpoints
You lost 98.07 PLOTpoints
This results in 249.85 PLOTpoints after this training sessions!
You took 13.92 seconds on the average to give a right answer!

Right now the best advice is to work on these three activities
1. VocabularyFrequency1-718-FlUCl3et Here 68.63% of the exercises caused problems
2. Midterm_01_Nouns3et Here 62.92% of the exercises caused problems
3. BHTest101_Translation_20min3et Here 49.4% of the exercises caused problems
To Individuell analysis

Figure 1: Screen-shot from Learning Journey Online

The system is inspired by experimentation in a course taught for the EuroPLOT project by Nicolai Winther-Nielsen at Fjellhaug International University College Denmark in Copenhagen. During this course he developed and tested the 12 introductory sessions which are delivered as open learning for EuroPLOT (http://bh.3bmoodle.dk/course/view.php?id=2 login as guest). In this classroom students used PLOTLearner and gave feedback on their learner experience in oral discussions during class as well as in e-mails, and the students forwarded their learning statistics for two months during this course. Both EuroPLOT’s PC version of PLOTLearner and a new online application, Bible Online Learner, can manage upload to a server, and our project uses these uploaded data as input for the Learning Journey Online. Without this solution the teacher has to analyze the learning statistics by hand when he or she wants to provide feedback for the learner. This is both a time consuming task and it depends to a high degree on the teacher’s previous experience of how students usually develop their language skills when making specific errors. The course design of the lecture using PLOTLearner was inspired by the idea that the learners are self-directed.

The research questions the current paper deals with are: 1) How does a persuasive approach to learning statistics look like? 2) What are the formal requirements on a statistics module which is designed to support tailored learning in a CALL system via surveillance?

Choosing a formal approach to the prediction of learning progress which employs surveillance naturally raises ethical concerns because it is possible to predict learning progress in learners. The paper discusses the ethical dimensions of surveillance in persuasive learning.

The paper is organized as follows: Section 2 discusses how surveillance can be persuasive and argues
that a feedback-loop which is created via surveillance and interaction with the CALL and the teacher will have a persuasive effect on the learner. Also ethical aspects are discussed in this section. In section 3 the formal framework for a statistics module based on Metsämuuronen is laid out.

2 How surveillance can be persuasive

The principles of persuasion that inspires our surveillance module was set out by Fogg (2003) in his introduction to Persuasive Technology. A decade of research into the role of the computer as a persuader suggested its potential for “adaptive education and training products that tailor motivational approaches to match each individual learner” (Fogg 2003:246). The PLOTLearner project has focused on how exactly technology can persuade learners and change the way they act and learn (cf. Winther-Nielsen ms b).

The development of the surveillance module aims on the one hand to be accessed by teachers and learners by a simple web interface which will enable them to keep track of the students' learning progress and on the other hand to assist the student in self-monitoring his or her own learning progress.

We define surveillance according to the theory of Persuasive Technology: “One party monitors the other party to modify the behavior in a specific way” (Fogg 2003: 46), and this is a common technique to persuade people to change their behavior. The approach chosen is uncovered monitoring; it persuades the learners in their learning by given instant and corrective feedback relative to the practice. As it has been explained in Laurillard (2012) and been exemplified in Gottschalk (2012), feedback is what most users of software for computer-assisted language learning desire. Surveillance is a way to provide this kind of feedback via detailed learning statistics, because it supports both the teachers and the CALL system in giving detailed feedback on the students' progress.

The persuasive architecture which Winther-Nielsen (ms a: 7) has developed for PLOTLearner is based on Fogg (2003), proposing only a slight modification of the architecture of a persuasive tool. Winther-Nielsen differentiates three levels in the development of ability and motivation through computer-assisted language learning. The first and simplest method to enhance ability is reduction which is well-known from quizzes satisfying the learner’s basic need for reviewing knowledge and memorization. A more sophisticated approach is tunneling. It proceeds in predefined learning progression. Such a predetermined sequence enforces automated results in mastery however it also limits freedom. The most persuasive activation of ability is achieved however via tailoring.

Figure 2: Persuasive architecture (Winther-Nielsen ms a: 7)
persuasive technology the training is adjusted to the learners’ knowledge level, age, learning style, progression, goals and other highly individual parameters which are related to vocational needs.

It is tailoring which is supported by the PeLLITS technology. Detailed feedback based on learning statistics will make it possible to adjust the whole learning process according to the statistics results and the specific needs of the student. By choice of the student, the teacher receives a detailed learning profile of this learner’s previous practice. Based on this profile the teacher can adjust his or her teaching according to the individual needs and goals of the learners. The teacher can intervene and provide the student with feedback and advice. Since the student gets feedback by the facilitator based on the learning statistics the student can as a self-directed learner focus on the teacher’s advice and adjust her or his learning according to it.

The other branch of Winther-Nielsen's architecture is motivation with similar functions and set up in a parallel track to focus on increasingly persuasive feedback. In the traditional curriculum, the first and often only motivational factor is conditioning which is caused by fear of failing the exam, but rewarded by the hope of acquiring a certificate. This is at the core of learning technology with simple exercise functions and instant conditioned feedback. At the next more sophisticated level of motivation surveillance will offer learners the choice to voluntarily share their outcomes with teachers, fellow learners, or in an open community. Using logged data on activity, an offer to receive more instructional help could motivate learners to improve on their learning processes. Furthermore, they could be given the opportunity to attain a certain status as a good gamer in a competition with other learners or to become the best helper in collaboration among peers. However, Winther-Nielsen (ms a) assumes that the ultimate persuasive system will be based on self-monitoring as in this case self-directed learners experience the freedom of being able to actively explore the learning environment and plot their own personal course through the learning content. In this way they are persuaded to plan their individual learning journey based on a visualization of progress and they can respond to the right kind of corrective feedback. Hence, the ultimate language learning system envisioned by Winther-Nielsen uses artificial intelligence and natural language processing to record the individual’s processes and outcomes and measure performance on language learning tasks. This system automatically adapts to individual learner differences and learning processes through tailoring and it primes the student for learning tasks thanks to surveillance and self-monitoring.

It is predominately the addition of a further feedback element to the learning circle which makes the system persuasive. In the current version of PeLLITS this is surveillance, but in the long run PeLLITS will be enhanced and developed into a self-monitoring system using artificial intelligence as envisioned in Gottschalk (2012).

While Fogg's approach to persuasion has inspired work on PLOTLearner and PeLLITS, the pedagogical approach for both is mainly based on the robust theory of design for learning which has been introduced in Laurillard (2012). It defines what makes the system persuasive. Laurillard's framework models the interaction between learners, facilitators and the learning environment. Laurillard's model of conversational theory is used to visualize the processes involved in learning from an external environment of a teacher (cf. Winther-Nielsen ms b). Laurillard's approach from is reproduced in figure 3, adding the surveillance part to her system.
In Laurillard's (2012: 60) approach learners use their personal goals and the current organization of learning spaces to select a desired practice which will generate learning actions on the external environment. As Laurillard explains, learners use actions which are modeled by the teacher or even results of own actions to modulate and build practice and capability. If a teacher is present in the learning cycle, there is also the possibility to learn through direct communication. Explanations and comments from the teacher or the learning environment enable the learner to develop a conceptual organization of the learning environment. This is shown by the elements (CT) and (FT).

It is specifically this part of the learning cycle described by Laurillard which is supported by the surveillance module. The reason is that surveillance activates a feedback circle which results in a behavior change in the student as he or she can adjust the learning progress according to this particular aim. This aspect of the system makes the statistics agent a tool for persuasive learning.

However, the use of surveillance in CALL raises serious concerns of ethic nature. The role of surveillance in persuasion has been discussed at length in Fogg (2003) and in Jespersen et al. (2007). Both note that surveillance makes it possible to change behavior, however this only works if the students know that they are monitored. This surveillance element might raise memories of the Panopticum introduced by Bentham and Bozovic (1995) where prisoners are permanently monitored. The Panopticum is a prison that in its center has a tower from which the guards can permanently monitor the prisoners but do not necessarily do so. This results in a deprivation of privacy and forces the prisoners to behave appropriately.

While the system was intended as an enhancement of the conditions of detention because no physical punishment is used, Foucault (1975) later criticized this approach severely and this critique is also followed by Jespersen at al (2007). The essential problem is that the Bethamian system simply redefines power in an inhuman way as the prisoners in the Panopticum lack any privacy due to the constant surveillance (cf. Jespersen at al. 2007). However, one can argue from a utilitarian point-of-view that the Panopticum has a good economic balance and does not physically hurt prisoners, and there is a better alternative and ethically more valid. The Panopticum brings wealth to a big number of people while its negative effects are moderate. This would also be acceptable to Foucault (1975) who assumes that the microwords of power cannot be changed, but one needs to deal with them.

A similar argumentation is also true for the statistics module, although with the difference in stance towards the ethical status of surveillance mentioned by Fogg (2003) and Jespersen et al. (2007): As...
long as an intended outcome is friendly and supportive for the user of the software, in that it positively enhances objectively measurable processes relevant to the user of the learning software, it can be regarded as ethical. This implies that if the surveillance technology is uncovered and it supports learners in acquiring desirable skills in a target language the goal can be regarded as friendly and therefore ethical. In the end the students immolate their privacy, but receive feedback which enhances their learning. From the utilitarian position, this is positive and regarded as ethical.

3 Formal framework

To develop a formal statistics framework for PeLLITS Judith Gottschalk is now using Item Response Theory [IRT] following the earlier helpful work by Metsämuuronen (2013). The main purpose of IRT is to test people and to predict the probability of a testee's response by establishing the position of an individual testee along the line of some latent dimension. Since IRT is used in an educational environment, the latent trait is often called *ability* which is set into relation to the difficulty of an item (Partchev 2004: 5).

A first step towards applying a formal statistics approach with IRT for PLOT Learner is the identification of items which can be tested with PeLLITS. These items include the morphosyntax of the learning object dealt with in PeLLITS.

For each student an SQL database receives and stores the following: data on the student identity, the test question answered by the student (this is what is called the item in what follows), whether the student has given the correct answer, and how much time it took the student to answer the question. Whether an answer is correct is determined by the truth values true or false. If a student does not respond to a question, the answer is interpreted as false in this statistical framework. This is the information the statistics module uses to determine statistical information on both the student and the item.

For the statistical module a dichotomous Rasch-Model is used. This is an IRT model which predicts the probability of a learner giving a certain response to some item. Within this model it is assumed that learners have a different level of ability and items have different levels of difficulty (Partchev 2004: 9).

The simplest IRT model for a dichotomous Rasch-Model does only possess one item parameter. It is the item response function, which gives the probability of a correct response given the single item parameter $b_i$ and the individual ability level $\theta_j$. The item-response function which contains one parameter is given in (1) below. It is also known as one-parameter logistic function.

\[
P_{ij}(\theta_j, b_i) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)}.
\]

The crucial part of the formula is the expression $\exp(\theta_j - b_i)$ which predicts the probability of a correct response from the interaction between the ability $\theta_j$ of an individual and the item parameter $b_i$. The latter parameter is called the location parameter, or the difficulty parameter (cf. Patchev 2004: 11).

The question arising from this is how the ability and the difficulty of an item are determined in this model: These are parameters, which are estimated based on different likelihood methods. The likelihood method, for the determination of the ability parameter, also known as person parameter is the marginal maximum likelihood [MML] procedure. It is used to estimate the distribution of means and standard derivations of the latent ability of each students; this is represented by the $\theta$-parameter in each updated version of the learning statistics right after each test is executed (cf. Metsämuuronen 2013: 17). For the estimation of the difficulty Metsämuuronen used the conditional maximum likelihood method [CLM] because this reduces the imprecision which would otherwise result from
the person parameter which is also used for the estimation of the item parameter. This is the β-parameter.

In view of the specific test design, it is necessary to equate the tests as pointed by Metsämuuronen. In our context the problem of different levels of proficiency for expert and novice level students does not affect the teacher surveillance tool because all students use PLOT Learner from the very beginning of learning Biblical Hebrew. If, however, the tool is used by intermediate or advanced learners and intended to store a broad number of learning profiles, the factor of comparing expert- and novice level students will play an important role.

For the current statistics database design is even more important that the test length can differ, given the different topics covered in the course of the week. It is important that the test scores are comparable and therefore we introduce vertical equation. By doing this, test scores from different tests of different lengths and administered in different time frames can still be compared statistically. Vertical equating was used for IRT modeling in the same way as Metsämuuronen:

\[ \text{Vertical equating was used for IRT modeling in the same way as Metsämuuronen:} \]

\[ \ldots \text{the scores were fitted to the same scale on the basis of characteristics of IRT models, which assume that a learner’s latent level of ability (\( \theta \)) and the difficulty level of an item (\( \beta \)) are identical, when certain preconditions are met (see Wright, 1968). The latent ability of each learner can be determined in the same scale for every test as long as linked items connect the test versions. Because of the small number of students in the experimental group (N = 13), the only recommendable model for estimating latent ability was the one-parameter model (that is, the Rasch model). The estimation was carried out using the OPLM program \ldots} \] (Metsämuuronen 2013: 17).

Accordingly, Judith Gottschalk uses the different equations, the CLM and MML introduced above for a Rasch-Model. It is used in the same way as Metsämuuronen does and is executed in the following order: In a first step the structure of the linked items is defined. The values of the difficulty of the linked items are exactly the same in each version, the difficulty levels of all other items can be calibrated according to the same scale as the one used for linked items. In a second step the CLM is used in order to estimate the difficulty level of each item which is represented by the \( \beta \)-parameter. The third step uses the MML to calculate the \( \theta \)-parameter. In the final step the \( \theta \)-parameter of the scores is estimated for each version. To do this, the means and the derivations of the distributions of the \( \beta \) and the \( \theta \) are used. This estimation results in a unique latent value which is measured on a common scale for each observation of the scores in all versions. Metsämuuronen points out that the success of equating depends on three different factors which are explained.

The proficiency level of the students should be reflected by the linked items. The items should represent a sufficient range of ability and they should be neither too easy nor too difficult. In the intervention, the linked items for the next test are selected based on previous tests. Metsämuuronen points out that this enables the system to discriminate items that are considered not too difficult or too easy. The items should cover the different content areas. This selection generates a short test embedded within the main test. It is necessary that in this short test the linked items are selected in a way which represents different content areas as widely as possible. In the equation the stable parameters depend on the sample. All students were tested at the beginning of the intervention to obtain a population which is as large as possible. In this way stable item parameters are acquired:

\[ \ldots \text{However, from the viewpoint of the population, the parameters for items measured only in the EG are unstable. Also, due to the small population in the intervention, the values for item “difficulty parameter” depended considerably on those students who participated in the test. Thus it was important to get all the possible test papers from the test-takers – even day after the test. Although the item parameters are somewhat vague, the results are much more accurate than if only classical metrics (the proportion of correct answers) were used in comparison.} \] (Metsämuuronen 2013: 17)

It will be a task for future research to develop an approach to IRT which also includes the response
time in the analysis of learning statistics. The reason why it has not been used here is that the so-called coefficient of variation which is usually used to display learning progress based on response time is quite controversial. This is a further argument in favour of using the IRT approach of Metsämuuronen (2013) to formally model a statistics approach for measuring learning progress in PeLLITS and to add surveillance to the software.

4 Conclusion

Surveillance combined with tailoring is at the core of the architecture of the new persuasive tutoring system PeLLITS which we have outlined here. The system generates a feedback-loop which can either be given by a teacher or a CALL system and the response motivates learners to change their behavior as an effect of the persuasive technology. Adding a formal approach based on IRT will even make it possible to develop a framework which is so informative that it can predict learning progress, enhancing the motivation for behavior change even more.

Acknowledgements

This research was funded by the Education, Audiovisual and Culture Executive Agency (EACEA) of the European Commission through the Lifelong Learning Program with grant #511633. We would like to thank Claus Tøndering for his supervision on the implementation of the statistics module described in this paper. We would also like to thank Christian Højgaard for discussions on the design of the statistics module.

References