An Adaptive Learning Environment for Statistics

Dagobert Soergel[1]

University of Buffalo, United States[1]

Abstract

This paper describes the design of an adaptive learning environment using statistics as the experimental domain. Providing better ways to learn statistics in higher education would improve statistics knowledge among graduates and make offering statistics courses cheaper. Statistics learning can demonstrate benefits of adaptation to (1) the subject domain in which the learner will apply statistics and (2) the mathematical knowledge of the learner. The system will create individualized learning paths optimally adapted to the learner's goals. existing knowledge, cognitive abilities, learning style, and circumstances by recommending the next step throughout a learning interaction look at another example suitable for this learner, solve another practice problem, or move on to the next concept. These recommendations draw on an extensive knowledge base using a combination of artificial intelligence and machine learning, of knowledge-based reasoning, and of learning analytics. The knowledge base includes statistical concepts and their relationships; learning objectives; many minimal presentation chunks ⊔ concept explanations, examples, problems to solve, questions ⊔ that can be sequenced in a learning path and that are indexed by concepts covered, learning objectives, prerequisites, difficulty, etc.; learners with rich profiles, including learner characteristics, history of progress through the system (including performance on tests after completing a learning path), and feedback on what presentations they liked. The learning analytics approach uses this data to derive even better learner profiles and conclusions about what learning materials suit a learner (and learners with similar profiles) to make more accurate predictions of the best next step in a learning path. The knowledge base can also be used to power a statistical advisor to recommend statistical methods and things the user should learn to apply these methods properly. Our hypothesis is that system following our design produces better learning outcomes; we plan to test this hypothesis.

Keywords: *Individualized learning, Learning path recommendations, knowledge base of learning units, Learner characteristics, Learning analytics, Artificial intelligence*

1. Introduction

We propose the design of learning support systems that adapt to the learner in selecting and presenting materials and activities. The paper is situated in the larger context of using information technology (IT), including artificial intelligence (AI) and machine learning (ML) to improve learning outcomes [1], offering each learner an experience that is optimally adapted to the learners goals, existing knowledge, cognitive abilities, perceptual abilities, learning style, and circumstance]. With ever more things to learn, we need more efficient learning. This is a special case of intelligent personalization using AI/ML to make life easier; systems learn about individuals from many sources, creating a profile to adapt ads, services, information, learning materials to each individuals needs and preferences.

Statistics learning in higher education is a good domain for building a prototype. Improving statistics learning promises large impact by improving statistics knowledge while lowering costs. Better knowledge of statistics leads to better research, better understanding and evaluation of research results, and better application of statistics to problem solving and decision making in business, government, and every day life. Statistics is a good subject to demonstrate benefits of adaptation, in this case to: (1) the discipline in which the learner will apply statistics and (2) the mathematical knowledge the learner brings to the table.

Support for learning statistics – or any other subject –requires an extensive database / knowledge base. This KB can also support an *online statistics adviser* for students, faculty, and practioners. After receiving advice, the user can turn to the learning environment to learn what is required to fully understand and implement the advice – just-in-time learning. The system can support users in understanding literature they found. It can support life-long learning. An adaptive learning environment is a highly complex recommender system, creating a learning path by recommending the next step throughout a learning interaction. These recommendations use a combination of AI and ML, of knowledge-based reasoning and of learning analytics / educational data mining.

1 The knowledge-based approach uses complex inferences using knowledge of statistical concepts and their relationships, of characteristics of learning materials, and of characteristics of learners.

2 The learning analytics approach uses a massive amount of data on the progress of learners through the system and explicit learner feedback to derive even better profiles of individual learners and more evidence of what learning materials suit a learner (and learners with similar profiles) to make more accurate predictions of the best next step in a learning sequence – look at another example suitable for this learner, solve another practice problem, or move on to the next concept.

2. The learning support system

Table 1. The learning support system in a nutshell

Purpose:	Learning statistics
Inputs:	Learner profile (knowledge, ability, learning style, etc.). Learning goal
Output:	Personalized learning path, a sequence of learning materials. tasks, and activities that lead from existing knowledge to the learning goal. Presented adaptively, depending on the learner pofile and how the learner progresses through each step.
Extension:	Statistics Adviser User can immediately learn what is necessary to carry out the advice.

At the heart of the system is a database of learning units, presentation chunks, and many other entities [2]. At the lowest level are PresentationChunks (text, visualization, simulation) explaining or illustrating a concept; examples; assignments. The same content is presented in different ways and at different difficulty levels so that the chunk best suited to a given learner can be selected. As importantly, the database contains learner profiles and histories. See Figure 1 for some examples from a hypothetical database.

For the system to be effective in both inferencing and learning analytics, it must store the many different types of information required, so these must be modeled in the highly complex structure of this database, best in a detailed entity-relationship (E-R) schema. Table 2 shows a few sample entity types

Table 2. Entity type examples

KnowledgeItem	
. ConceptBroad	Abstract concept, method, named entity
. KnowledgeChunk	One or more statements of fact, conjecture, prescription, etc. conneting two or more ConceptBroad, identified as KC1, KC2, …
LearningUnit	A structured set of PresentationChunks. Also external lecture etc.
PresentationChunk	A unit that can be presented to a learner.Identified as PC1, PC2, …
PresentationChunkFormat	Sample values: WrittenText, AudioText, Image, Visualization, Simulation
PresentationChunkFunction	Possible values include Exposition, Example, TestQuestion, Answer
EntertainmentValue	Scale 0 - 4, used to select a PresentationChunk for a reluctant learner.
Person	Persons can play many roles (learner, instructor, author, ...)
PersonCharacteristic (LearnerCharateristics)	Just some examples: VisualThinker, WritingAbility, EyeSight, Hearing.

Conceptually, information is stored in the system in statements that connect two or more entities using **relationship types** as patterns, for example

PC5	*<dealsWith>*	ConceptBroad t-Test
PC5	*<hasFunction>*	Example
PC5	*<hasApplicationDomain>*	ConceptBroad Agriculture
PC5	*<hasEntertainmentValue>*	3
PC11	*<hasFunction>*	Exposition
PC18	*<hasSameContentAs>*	PC11
KnowledgeItem267	*<hasPrerequisite>*	KnowledgeItem113
Person256	*<hasPersonCharacteristic>*	(EyeSight, 0) (Person256 is blind)

<iEexampleFor>
<isAnalogyFor>

The complete E-R schema, a complete listing of all entity types and relationship types would take some time to develop. Below is an informal description of key information typese.

Figure 1. A glimpse at the database (PC = PresentationChunk)

Objective 1	Statistics-topic: t-test	easy
Able to apply the t-test. Know when to use.		

Objective 2	Statistics-topic: t-test	hard
Able to explain mathematics of the t-test		

Person 1	Appl::Agriculture	Math poor
Completed Units …		
Obj. 1 Outcome B		
Xiao Wang		

Person 2	Appl: Medicine	Math good
Completed Units …		
Obj. 1 Outcome A	Obj. 2 Outcome A	
Hans Kohli		

PC 436	Stat-topic: t-test	Function: Definition	Format: Written text	Difficulty: Medium
Prerequisites: Understanding of inferential statistics, hypothesis testing, statistical significance				

A t-test determines if there is a significant difference between means of a given variable, such asmath test score, blood pressure, crop yield. We can compare the mean of one group and a given value, or the means of two independent groups, or pre-test and post-test results in the same group of people.

PC 578	Stat-topic: t-test	Function: Example	Appl: Agriculture	Format: Written text

Effect of fertilizer mix on wheat yield **under medium-dry conditions**
Each sample consists of 10 plots which are the same n every respect except one gets fertilizer mix 1 and the other mix 2. At the end of the season, wheat yield is measured for each plot and the mean is computed for each sample. We want to know whether any difference in the means could be due to chance without any effect of the fertilizer mixx used or whether the difference between means is so large that its occurrence without an effect of the fertilzer mix used would be hiighly unlikely.

PC 579	Stat-topic: t-test	Function: Example	Appl: Medicine	Format: Written text

Effect of Amlodipine on lowering blood pressure
For a group of people with specified characteristics test the effectiveness of Amlodipine vs a placebo in lowering blood pressure. A pool of people meeting the criteria for the group is gathered by asking for volunteers. Two random samples are drawn from the pool. One is given Amlodipine (the experimental sample), the other is given a placebo (the control sample). For each participant the decrease of blood pressure from the beginning to the end of the trial is measured. For each sample, compute the mean decrease. Could the difference of the means be due to chance without any effect of Amlodipine or is the difference is so large that its occurrence without an Amlodipine effect would be highly unlikely?

LU 681	Statistics-topic: t-test
Objective 1 Apply t-test, know conditions	
Parameter A. Math level (from low to high).	
Parameter B. Application domain	
Prerequisites: PC 436 AND (PC 578 OR PC 579), …	

A learning unit is a structured set of PCs. This LU is customizable by the two parameters, e.g. for a learner with low math level and application domain Agriculture the LU will use example PC 578 and a corresponding data set.

LU 682	Statistics-topic: t-test
Objective 2 Be able to explain t-test	
Parameter A. Math level (>= med.	
Parameter B. Application domain	
Prerequisites …	

See comments for LU 681

PC 735	Stat-topic: t-test	Function: Exam question	Appl: Medicine	Format: Written text

Here are the results of a clinical trial of prescribing Bactrim vs. Penicillin for Staphylococcus UTI. Success measured by number of days to relief symptoms. Find out whether there is a significant difference.

Find the best learning path for Xiao Wang and for Hans Kohli

2.1. Database content: PresentationChunks

Information given

- Statistical topic covered
- Application domain (e.g. Agriculture, Medicine, Education) (for examples, assignments)
- Function in the learning process (Introduction, Prerequistes, Definition, Example, Rationale of a statistical test, Conditions for using a statistical test, "Cookbook" prescription of how to carry out a statistical test. testExercise, Assignment, TestQuestion, etc.)
- Format (Media type)
- DifficultyLevel
- EntertainmentValue (would be high for a cartoon illustrating a statistical concept)
- Learner characteristics for which suitable
- Links to other PresentationChunks. Links can be embedded in the text. The type of link is indicated, such as *elaboration* or *moreFormalExposition* or *definition* (for a term in the text).
- Detailed data on interactions of learners with the PresentationChunk and learning outcomes These data include time spent on the chunk (or even more granular, sections of the chunk, performance in small tests, questions time spent on a snippet, performance in small tests and end-of-a larger-unit test, learner frustration (if system is able to detect), questions the learner asked, explicit feedback from the learner.

Sources for the information

PresentationChunks can be extracted from statistical textbooks or, for more advanced content, journal articles, or they can be harvested from the Web, but copyright is serious problem and a barrier. Both processes should be automated to the extent possible, a very challenging task The information about each PresentationChunk could be produced by automatically by analyzing text or images (challenging) or by a human editor (much effort). This information should be constantly updated / modified by machine learning/

2.2. Database content: Persons (learners) – A profile with much data

Information given

- Knowledge of statistics and other mathematics (e.g., calculus)
- Interest in / knowledge of a subject domain (e.g., Education)
- Other interests
- Abilities (cognitive, artistic, sensory, manipulative)
- Learning style
- Attention span
- Tolerance for frustration
- Emotions, attitudes, self-esteem
- Detailed data om interactions with PresentationChunks and learning outcomes

(for convenience, these data are stores both with the PresentationChunk and the Person)

Sources for the information (as available, with leaner's or parent's permission, as required)

It is useful to have an initial learner profile as complete as possible because this enables good adaptation from the very beginning. Some sources that can be used:

- System-administered pretest and initial interview
- Facebook and such (if the learner agrees to be friends with the system) (also source for updates)
- Transcripts. Syllabi of other courses taken in the past or concurrently

The learner profile must be constantly updated / modified through machine leering and perhaps feedback from the learner, for example interviews at the completion of one chapter**.**

2.3. Further observations on the proposed system. Support for learner-directed inquiry

- Find out what KnowledgeItems are especially difficult and find ways to make then easier to learn.
- Use analogies and metaphors to help learners understand.
- The system suggests learning paths, but it also supports learner-directed inquiry through many links between PresentationChunks, offering help when the learner gets lost or navigates to a chunk for which she is not prepared, and visual displays, such as concept maps, strand maps [3], concept

hierarchies (Fig. 2}, the equivalent of a book table of contents and chapter outlines, for the learner to browse and navigate with the option to drill down to PresentationChunks at any time.

Figure 2. Snippet from a hierarchy of statistical tests

t-test
. one-sample t test against a known mean
. two-sample t-test
. . two independent samples t-test
. . paired samples t-test
Distinctions applicable to several tests, including the t-test
, One-tailed vs two-tailed test
. sample variances equal vs unequal

2.4. Challenges in building and maintaing the proposed system

- Computer assistance in segmenting text into chunks with metadata.
- Automatically classify a learning unit by its suitability for different kinds of learners. Use. naturallanguage processing (NLP) methods that can determine difficulty of a text beyond readability
- Derive learner characteristics from their interaction with the system.
- Advance recommender systems to use more data more intelligently.
- Find suitable analogies and metaphors automatically
- Solve copyright problems, perhaps through a system of micro-payments that depend on the length of a presentation chunk and the frequency with which it was accessed.
- Address the tension between personalization and privacy, following applicable privacy laws and letting users opt-in to more collection and use of their personal data.

3. Some thoughts on evaluation

Logging detailed data on learner interaction with the system provides rich data for quantitative and qualitative analysis of learner performance. Some of the measures require special data collection using an IRB-approved protocol. Ultimately we need to be able to compare our learning environment with, for example, traditional courses with respect to leaning time and learning outcomes, but his is notoriously difficult. Evaluation can use

process measures. such as

- number of clarifying questions from learners,
- number of backtracks,
- Time to complete learning path,
- Number of questions that take the concepts learned further

learning outcomes, such as

- understanding of statistical concepts,
- knowing when to apply what test,
- performance on statistics problems, attitude towards statistics,
- students' confidence in their own ability.
- later use of statistics in research)

4. Future work

We plan to prepare a small learning unit that volunteer students in a regular statistics courses could take instead of attending a lecture and do a qualitative comparison how students dide.

5. Conclusions

Building a learning suppport system as proposed here requires a large investment, but this investment will pay off. The proposed approach to individualized education and innovation in intelligent tutoring makes learning more efficient (learners need less time), more effective, and cheaper. This is responsive to much criticism leveled at higher education.

Acknowledgements

I thank Sunha Kim, University at Buffalo, for very helpful discussions on this project.

References

[1] Soergel, D. (2018). Innovative education enabled by Knowledge Organization and IT: Goal directed, flexible, individualized, collaborative. FOE 8, Florence, 2018-06-28/29. Proceedings . libreriauniversitaria. it Edizioni. 2018. 688 p. ISBN: 8833590208. p. 273-278
[2] Soergel, D.; Iturralde Escudero, .F. (2018). Toward a Universal Document Model for Active Knowledge. ICKM 2018. https://digital.library.unt.edu/ark:/67531/metadc1393843/?q=soergel
[3] Atlas of Science Literacy. Mapping k-12 science learning" Volumes 1 and 2, 2001/2007, http://www.project2061.org/publications/atlas