Using Data Mining to Automate ADDIE

Fritz Ray
Eduworks Corporation
136 SW Washington Ave. STE 203
Corvallis, OR 97333
+1 (541) 753-0844 x308
fritz.ray@eduworks.com

Keith Brawner
U.S. Army Research Laboratory
HRED-STTC
Orlando, FL 32826
+1 (407) 380-4648
keith.w.brawner@us.army.mil

Robby Robson
Eduworks Corporation
136 SW Washington Ave. STE 203
Corvallis, OR 97333
+1 (541) 753-0844 x304
robby.robson@eduworks.com

Abstract:

The goal of this work is to transform informational and instructional content into adaptive and personalized training experiences. We have developed semi-automated methods to do this that parallel the traditional "ADDIE" (Analysis, Design, Development, Implementation, and Evaluation) process. The source content can include documents, presentations and manuals and existing courseware. The techniques use artificial intelligence (AI), data mining, and natural language processing and generally belong to the discipline of "educational data mining." This poster/demo demonstrates the processes and discusses the algorithms used.

1. PROBLEM STATEMENT

Today's digital environment is rich with learning content, but much of it is purely didactic in nature. This content includes manuals and presentations not intended for instructional purposes and e-learning that consists of presentations and lectures with multiple choice questions. As online learning replaces instructor-led training in corporations, government agencies, and educational institutions [10], its effectiveness can be improved by transforming this wealth of didactic content into more interactive and adaptive learning experiences [5].

Here, we address aspects this transformation problem in the context of multiple research and commercial projects. A large portion of the work we report here comes from a U.S. Army Small Business Innovation Research (SBIR) project called *Tools for the Rapid Generation of Expert Models*, or TRADEM, that applies data mining to (a) deconstruct existing content at a deep and granular level and (b) reconstruct it in a form that can be used to create adaptive intelligent tutoring systems. This process automates many steps in the "ADDIE" (Analysis, Design, Development, Implementation, and Evaluation) process [1] commonly used to develop instructional content.

1.1 Motivation

The three primary benefits of applying EDM to automate a process such as ADDIE are cost, speed, and the effectiveness of the training produced.

Data about e-learning development [3] shows that about 40% of the cost involves analysis and design tasks, which includes the expensive activity of engaging with subject matter experts. Using EDM to extract the domain analyses and instructional designs from existing content is more cost-effective than going through the entire ADDIE process each time instruction is developed. For example, in Army Civilian Affairs training using TRADEM, simulations provide experiential learning on how to conduct civilian affairs in current, real-world situations. The content changes frequently, requiring continual repetition of the ADDIE process. As a result, manual processes are too slow and too expensive, but automated the generation of up to date domain models, concept and skill maps, and instructional content allows the Army Civilian Affairs Corps to rapidly deploy new training in response to a real and changing world.

Providing highly effective training also drives the development of TRADEM. Classroom instruction and most existing e-learning falls far short of the effect sizes that have been shown to be achieved with *intelligent tutoring systems* [4; 13]. While TRADEM can be used to develop and implement many different types of learning environments, our work has focused on producing intelligent tutors.

2. DESCRIPTION OF TRADEM

ADDIE's design step consists of determining learning objectives, sequencing instruction, and writing assessments. TRADEM automates this step via an assisted full workflow solution.

Workflow: First, TRADEM extracts a topic map from a user-input corpus of content. This topic map visualizes a set of topics that cover the core topics present in the input corpus. For each node (topic) in the concept map, TRADEM then selects the pieces (granules) of the initial input corpus most associated with that topic. Next, TRADEM builds an assessment for each topic based on the granules associated with that topic. On demand, these assessments are then exported in an intelligent tutoring format for use in instruction.

Topic Generation: TRADEM ingests a corpus of content and performs a front-end analysis that results in a concept map consisting of a directed tree of topics. The topics are extracted from the corpus using topic-detection techniques [9] that are applied as described in [12]. The number of topics generated optimizes coverage of the input corpus, but the user can alter the number of topics based on pedagogical needs. This is necessary in real-world applications. For example, the user may wish to match a list of topics that appear in standardized curricula.

To determine topic relationships and order, TRADEM calculates a relation strength for each pair of topics, creating a graph with relationship strengths between all topics. This fully connected graph is transformed into a directed tree spanning all topics by inferring directionality using a precedence metric and tree selection algorithm based on aggregate relationship strength. We interpret this tree as representing optimal learner paths between any two topics in the input corpus. This mirrors the way classical instructional designs progress through a subject, including intermediate learning objectives leading to a terminal learning target [12].

Content Granules: To identify the pieces of the input corpus most closely aligned with each topic in the topic tree, TRADEM decomposes the input corpus into *granules* of content. For standard text, these are paragraphs, while slides and bulleted lists may end up as single or multiple granules. A sentence parsing algorithm is used to selecting which sentences in the granule are best suited to generate assessment questions using assessment generation techniques based on the work of Mitkov, Ha, Heilman and Smith [8; 11]. These techniques produce template forms that can be transformed into essay or multiple choice questions using a manual process. Additionally, each granule is automatically tagged with suggested relevant instructional types. For example, the *Generalized Intelligent Framework for Tutoring* (GIFT) includes an *Engine for Macro- and Micro-Adaptive Pedagogy* (EMMAP), that recognizes four pedagogical strategies: Rule, Example, Recall, or Practice [5]. This allows granules associated with each topic to be selected by an intelligent tutor based on its pedagogical needs. Thus, extracted topics are associated with meaningful chunks of corpus content that become the basis for real instruction driven by an intelligent tutoring framework.

Tutor Implementation: The target intelligent tutor we currently produce is dialogue-based tutor that we call T-Tutor. It is described in more detail in [2]. T-Tutor engages the learner in conversation in one panel and displays content in another. A chat bot powered by ChatScript [14] gives T-tutor the capability to engage in human-modeled conversation. Student responses are evaluated against target responses using ChatScript's innate functionality and standard semantic analysis techniques like those used by the AutoTutor family of tutors [6; 7].

T-Tutor uses GIFT as its core adaptivity engine[5]. GIFT guides the learner through a topic sequence from the extracted topic tree and guides learner-level pedagogy by adaptively selecting granules based on pedagogical need and learner state. In order to provide a dynamic link between the analyzed content and specific intelligent tutor, TRADEM generates on-demand JSON files that encodes all of the information needed for an intelligent tutor to adaptively and interactively implement a pedagogical plan.

Evaluation: In our Topic Detection, we use AI and data mining techniques to extract topics and sequencing data. This results in an *a priori* model based implied by the source materials. Our goal in evaluation is to use actual learning outcomes to update this model. To this end, TRADEM enables the delivery system to report observed assessment results, with each result mapped to one or more learning outcomes or topics. Once data is gathered, it will be processed to determine the goodness-of-fit between the observed data, the existing topic model, and other potential variants of this model.

3. THE BIGGER PICTURE

The processes we have described brings data mining and practices into the realms of training and education to improve speed, quality, and flexibility of content production. In addition, this approach allows for direct comparison between pedagogical approaches. TRADEM's automated methods produce standardized machine-readable data with testable topic models analyzed based on observed learning outcomes. In other words, we can mine data generated by users and determine how well a given model fits the data. Markov modeling and structural equation modeling can be used to *infer* learning effects if the model or pedagogy is changed, and will immediately update tutors constructed from the models. In other words, the use of EDM to automate ADDIE creates standardized data structures upon which e-learning content is built which, in turn, enables EDM to be used to improve the structures and make the e-learning more effective.

4. ACKNOWLEDGEMENTS

The work reported here was supported in part by Small Business Innovation Research grants and contracts from the National Science Foundation and the Army Research Laboratory.

5. REFERENCES

[1] BRANSON, R.K., RAYNER, G.T., COX, J.L., FURMAN, J.P., and KING, F., 1975. Interservice procedures for instructional systems development. Executive summary and model Florida State Univesrity, Tallahassee.

[2] BROWN, D., MARTIN, E., RAY, F., and ROBSON, R., 2014. Using GIFT as an adaptation engine for a dialogue-based tutor. In *GIFT Symposium 2* Army Research Lab, Carnegie Mellon University (to appear).

[3] CHAPMAN, B., 2010. How long does it take to create learning.

[4] DODDS, P. and FLETCHER, J.D., 2004. Opportunities for New "Smart" Learning Environments Enabled by Next-Generation Web Capabilities. *Journal of Educational Multimedia and Hypermedia 13*, 4, 391-404.

[5] GOLDBERG, B., BRAWNER, K., SOTTILARE, R., TARR, R., BILLINGS, D.R., and MALONE, N., 2012. Use of Evidence-based Strategies to Enhance the Extensibility of Adaptive Tutoring Technologies. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* NTSA.

[6] GRAESSER, A., CHIPMAN, P., HAYNES, B., and OLNY, A., 2005. AutoTutor: An Intelligent Tutoring System with Mixed-Initiative Dialogue. In *IEEE Transactions on Education* IEEE, 612-618.

[7] GRAESSER, A., PENUMATSA, P., VENTURA, M., CAI, Z., and HU, X., 2007. Using LSA in AutoTutor: Learning through mixed-initiative dialogue in natural language.

[8] HEILMAN, M. and SMITH, N.A., 2009. Question generation via overgenerating transformations and ranking Carnegie Mellon University, Pittsburgh, PA.

[9] HORNIK, K. and GRÜN, B., 2011. topicmodels: An R package for fitting topic models. *Journal of Statistical Software 40*, 13, 1-30.

[10] LYKINS, L., DIXON, A., MILLER, L., MANDZUK, C., FRANKEL, D., MCDONALD, A., and KELLY, M., 2013. Informal Learning: The Social Revolution. In *White Papers* American Society for Training and Development, Alexandria, VA.

[11] MITKOV, R. and HA, L., 2003. Computer-Aided Generation of Multiple-Choice Tests. In *HLT-NAACL Workshop on Building Educational Applications Using Natural Language Processing*, Edmonton, Canada, 17-22.

[12] ROBSON, R., RAY, F., and CAI, Z., 2013. Transforming Content into Dialogue-based Intelligent Tutors. In *The Interservice/Industry Training, Simulation & Education Conference* National Training and Simulation Association, Orlando, FL.

[13] VANLEHN, K., 2011. The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist 46*, 4, 197-221.

[14] WILCOX, B., 2013. ChatScript SourceForge.