Civilian Analogs of Army Tasks: Supporting Pedagogical Storytelling Across Domains

Andrew Gordon, Mark Core, Sin-Hwa Kang, Catherine Wang, and Christopher Wienberg
ICT, University of Southern California, 12015 Waterfront Drive, Los Angeles, CA 90094
Email: gordon@ict.usc.edu, core@ict.usc.edu, kang@ict.usc.edu, catherhw@usc.edu, cwienberg@ict.usc.edu

Abstract: Storytelling is the most basic means by which people learn from the experiences of others. Advances in educational technologies offer new opportunities and experiences for learners, but risk losing the natural forms of pedagogical storytelling afforded by face-to-face teacher-student discussion. In this paper, we present a technology-supported solution to the problem of curating and algorithmically delivering relevant stories to learners in computer-based learning environments. Our approach is to mine public weblogs for textual narratives related to specific activity contexts, both inside and outside the domain of the target skillset. These stories are then linked directly to task representations in the learner model of an intelligent tutoring system, and delivered to learners along with other tutoring guidance. We demonstrate our approach to curating stories by creating collections of narratives that are analogous to tactical tasks of the U.S. Army, and evaluate the difficulty of incorporating these stories into intelligent tutoring systems.

Introduction

In George Lucas's classic 1977 film, *Star Wars*, protagonist Luke Skywalker considers the rebels' plan to destroy the enemy's Death Star super-weapon by dropping a bomb down a ventilation shaft: "I used to bull's-eye womp rats in my T-16 back home. They're not much bigger than two meters." Here Luke draws an analogy between a military task (precision bombing) and a civilian pastime (womp rat hunting), which he uses to persuade others of the technical feasibility of the plan. With this dialogue, George Lucas establishes that a young moisture farmer from the planet Tatooine could believably pilot a rebel fighter in space combat, appealing to the audience's commonsense intuitions about the applicability of experiences across domains.

The science of transfer learning, however, is often counter to these commonsense intuitions. In their survey of research in educational psychology on transfer learning, Day and Goldstone (2012) note that the spontaneous transfer of solutions across situations is difficult for learners. Much of this difficulty may be attributed to the challenges that learners have in recognizing that past situations have solutions that are relevant to current problems. In an often-repeated result, Gick and Holyoak (1983) found that subjects perform poorly on a classic insight problem, even when directly analogous problems and solutions were presented immediately prior. However, explicitly noting the relevance of the previous case to the current problem led to dramatic improvements, implicating retrieval and correspondence as the cognitive challenges in the application of experience to new situations. In human-to-human tutoring, the tutor often performs these two cognitive tasks, retrieving stories from their own past experiences and explaining the correspondence between these experiences and learners' impasses. As researchers seek to model effective tutoring in intelligent tutoring systems, we must ask: Can the process of pedagogical storytelling be automated as well?

In this paper, we present a technology-supported solution to the problem of curating and algorithmically delivering pedagogically relevant stories from inside and outside of the target skill domain, for use in intelligent tutoring systems. Our solution is to harvest collections of stories that are relevant to specific task domains from public weblogs, and to provide system developers with the tools needed to connect these stories to the skill representations that drive intelligent tutoring behaviors. We describe an application of our approach, the Civilian Analogs of Army Tasks, consisting of a hundred collections of personal stories from public weblogs that are analogously related to tactical tasks performed by the United States Army. We describe the technical methodology used to curate these collections, and an approach to automated pedagogical storytelling in intelligent tutoring systems that links individual stories to explicit task models. Finally, we describe an online experiment to assess people's abilities to draw connections between stories and tasks, comparing stories both inside and outside the task domain.

Civilian Analogs of Army Tasks

Officers, enlisted soldiers, and government civilians in the United States Army are routinely tasked with missions that are far more diverse than those depicted in Hollywood war movies. Along with the skills of armed combat, Army personnel in recent deployments have conducted bilateral negotiations with local government officials, provided preventative dentistry services, constructed educational facilities, contracted covert informants, produced mass media programming, and razed fields of illicit agricultural crops. The need to execute tasks such as these often arises during deployments based on changing conditions in the operational

environment. As a consequence, soldiers often are faced with tasks for which they have received little institutional training. The United States Army, like other very large organizations that must be highly adaptable in conducting operations, alleviates some of the need for more training by capitalizing on the wisdom of the organization itself, i.e. the experiences of other soldiers who have previously executed similar tasks. To enable learning from these experiences, effective knowledge management continues to rise as a major priority within the Army. Sharing knowledge in online discussion forums is one solution, such as those integrated into the organization's Army Knowledge Online (AKO) web portal. An example is the popular Company Command web forum (Dixon et al., 2005), where company commanders effectively bridge the gap between people who have experiences to share and those that can most benefit from receiving them.

Our belief is that organizational learning inside the Army could be improved by looking outwards, expanding the pool of experiences from which it draws to include those from analogous civilian domains. The vast majority of tasks assigned to soldiers are clearly analogous with one or more activities that are skillfully executed by people outside of the Army. Some analogies are extremely close. School construction in the Army requires nearly exactly the same skills as those employed by professional construction workers. Razing fields of crops in Afghanistan has a lot in common with razing fields of crops in Arkansas. Other analogies share fewer surface features but share deep features. Bilateral negotiations between soldiers and government officials pose many of the same concerns found in international business negotiations. The production of mass media material in Army psychological operations is not unlike the work done by brand managers at advertising agencies. The stories told by these construction workers, crop farmers, business executives, and advertising agents could be useful to soldiers in the acquisition of Army skills.

To investigate the broader applicability of civilian experiences to Army tasks, we conducted an analysis of a large-coverage taxonomy of tasks assigned to Army units. The Army Universal Task List: Field Manual 7-15 (United States, 2012) lists official designations for hundreds of tasks that are assigned to Army units at the tactical level (as opposed to the operational or strategic level), to include tasks such as Purify Water (ART 4.1.3.11.1) and Support Famine Prevention and Emergency Food Relief Programs (ART 7.3.3.5). For each of these tasks, we brainstormed the activities in the civilian world that were most analogous, e.g. working at a water treatment facility and charitable food drives. Although many Army tasks lacked obvious civilian analogs, we found that correspondences could be made at higher levels of abstraction. We created a notional list of one hundred and two civilian activities that were broadly relevant to Army tasks, organized into eight high level Army concerns. Table 1 lists these categories, with examples of civilian tasks that we saw as a potentially relevant source of stories from large numbers of civilian practitioners.

High-level Army category	Example civilian tasks
Training	Going to shooting ranges, Backpacking, and Triathlons
Education	Tutoring, Taking exams, Study abroad
Leadership	Coaching sports teams, Refereeing sporting events
Battle Command	Firefighting, Wedding planning, Poker
Stability Operations	Prison operations, Political campaigns, Fundraising campaigns
Support Operations	Charitable food drives, Health examinations, Emergency room care
Signals Intelligence	Hackathons, Computer repair, Fireworks
Logistics	Emergency evacuations, Class field trips, Truck driving

Table 1. Eight high-level Army concerns and examples of analogous civilian tasks

Retrieving Narratives From Public Weblogs

Millions of people chronicle the events in their personal lives online in public weblogs, creating opportunities to automatically amass large collections of stories about specific civilian activities. However, Swanson (2011) estimated that only 5.4% of all non-spam English-language weblog posts are personal stories, defined as non-fictional narrative discourse that describes a specific series of causally related events in the past, spanning a period of time of minutes, hours, or days, where the storyteller or a close associate is among the participants. Using supervised machine learning techniques, Swanson constructed a text classifier to identify these personal stories in streams of weblog posts (precision 59.1%, recall 41.4%), which was subsequently used by Gordon et al. (2012) to construct a story extraction pipeline from data provided by a commercial weblog aggregator. Running since January of 2010, this pipeline has collected over 30 million stories thus far.

This story repository included numerous narratives for each of the 102 activities that we identified as civilian analogs of Army tasks. In order to curate collections of stories for each activity, we used two prototype search technologies. The first, *StoryUpgrade* (Gordon et al., 2012), is an activity search tool that incrementally builds a statistical topic model for use as a textual search query through the use of relevance feedback provided by the user. In this approach, users begin a search for stories by authoring a paragraph-sized "boring story" of the desired activity: a fictional past-tense narrative describing the activity that includes as much of the domain-specific vocabulary as possible, but avoids specific terminology unrelated to the activity, e.g. proper names of

people and places. Terms in this initial query are weighted and elaborated by hand-annotating the relevance of top search results, which are then used to iteratively query the collection until an adequate number of relevant stories are identified, or when the top search results consistently show irrelevant stories. The second search tool we used, *PhotoFall* (Wienberg & Gordon, 2012), capitalizes on the finding that 82% of photographs in narrative posts were taken during the course of events described in the surrounding narrative text. The PhotoFall search tool exploits this close connection between photographs and the narrative text to provide search users with a fast relevance feedback mechanism. Photographs from the top one thousand results of the current query are extracted and shown to the search user as a proxy for the full posts, allowing the user to quickly guess their relevance from the image alone. This feedback is again used to iteratively weight and elaborate the terms in a textual query, improving the learned topic model for the activity for both retrieval tools. In both of these search tools, story collections are created only from stories judged as relevant, a process that overcomes the less-than-perfect precision of both the learned topic model and the initial story classifier.

Using these search tools, we found that it required roughly three person-hours to curate modest-sized story collections for each of our 102 activities. On average, we identified 14.25 relevant stories for each activity, and a total of 1454 narratives of civilian analogs of Army tasks.

Integration Into an Intelligent Tutoring System

We sought to integrate pedagogical storytelling into technology-based immersive training environments where learners acquire skills through deliberate practice, and where instructional strategies are partially automated in software through Intelligent Tutoring Systems (ITS). The central problem of pedagogical storytelling in these environments is to deliver just the right story at just the right time in the course of a specific learner's acquisition of target skills. As a first attempt to integrate pedagogical storytelling into immersive training environments, our approach was to piggyback on top of user-modeling technologies used by an ITS to critique and provide feedback for a learner's performance.

Our integration approach is best suited for ITSs that model learners by aligning their behavior in an immersive training environment with an explicit task model of expert performance of a skill. This ITS design is seen in the BiLAT training system, a game-based environment for practicing negotiation skills in a cross-cultural context (Kim et al., 2009). Built for training U.S. Army personnel, learners in BiLAT prepare for and execute face-to-face negotiations with virtual characters in order to improve conditions in a fictional Iraqi city. Using a menu-based dialogue system, learners select conversational moves in an evolving dialogue context. The associated ITS in BiLAT (Lane et al., 2013) compares these selections to an explicit task model of expert performance, formulated through a cognitive task analysis of the skills of expert negotiators. This explicit task model is organized as a hierarchy of required steps, optional steps, rules of thumb, and actions to be avoided in expert performance of negotiation skills in a cross-cultural context. In developing BiLAT's ITS, every possible dialogue action in BiLAT was associated positively or negatively with one or more items in this hierarchy. When using BiLAT, a learner's abilities are assessed by aggregating evidence throughout the interaction, enabling the ITS to provide targeted feedback and hints during and after each training session.

Pedagogical storytelling can be integrated into an ITS of this sort by linking stories to specific items in the explicit task model. When the learner demonstrates poor performance on a particular task, the ITS can present one or more of the stories linked to the task, determined by the nature of the performance problem. In the BiLAT domain, there are three specific causes of poor performance that storytelling may address: 1) Learners may ignore the advice given about negotiation and cultural awareness thinking they know better. 2) The advice is necessarily general and real world problems require determining specific words to say. 3) The domain is inherently unpredictable; correct actions do not always bring success, and incorrect actions do not necessarily hurt performance. To enable an ITS to select the most appropriate story for a learner, we developed a taxonomy of relations between task-model elements and stories (Table 2). To address motivational issues, an ITS could use stories labeled with *evidence, warning* and/or *motivation* relations to convince the learner to heed the advice given. Stories labeled with *explanation* could teach learners causal relationships to help them adapt to the variety in real world situations. *Background* stories would contain concrete examples allowing learners to practice matching general advice to specific situations. Stories about *exceptions* could promote adaptive thinking to deal with real world uncertainty and help prevent misconceptions.

Relation name	Relation to task
Evidence	An example of correct behavior resulting in a positive outcome.
Warning	An example of incorrect behavior resulting in a negative outcome.
Explanation	An explanation of causal information underlying correct behaviors.
Exception	An exception to the rules due to luck or special circumstances.
Motivation	Personal and/or emotional motivations for performing the correct behavior.
Background	A real world example.

Table 2. Task model / story relationship taxonomy.

Evaluation of Authoring Feasibility

The ITS-based pedagogical storytelling approach described in the previous section requires that instructional designers can reliably make the relational connections between stories and task model elements. Given a collection of stories, instructional designers must read and comprehend each one, identify the elements of the task model to which they are related (if any), and characterize the relationship between these elements and the stories. Our hypothesis was that instructional designers are capable of performing these tasks, but we were concerned that stories from analogous domains might be more difficult than those directly related to the training domain. To investigate these concerns, we conducted an experiment to assess people's ability to link stories to task models, comparing performance on within-domain stories to analogous-domain stories via measures of inter-coder agreement and self-reported confidence.

For the purpose of this experiment, we selected the training domain of bilateral negotiations as conducted by personnel in the U.S. Army. This choice allowed us to use the explicit task model from the BiLAT ITS as a basis for our experimental task. The complete task model of BiLAT is a large collection of actions and rules of thumb organized into a hierarchy defining different levels of granularity. We focused on the top-level of this hierarchy and identified seven crucial cross-cultural negotiation skills (Table 3).

Skill	Description
Gather background information	Gather background information about your meeting partners and their possible actions as well as your possible actions and their impacts.
Respect culture of partners	As possible, learn and follow customs of your meeting partners, and follow their lead (if they are hosting).
Present yourself well	Given the culture of your meeting partners and the situation, present yourself in the best light. For example, some cultures value the concept of "face" and you should do nothing to lose face or cause your partners to lose face.
Use caution with interpreters	If you have the option to choose an interpreter, you may want to use one that you are already comfortable with, or who is a relevant specialist. Otherwise, be careful of bad translations.
Master small talk	Avoid sensitive topics, focus instead on neutral topics, or topics of interest to your meeting partners.
Build relationships	As culturally appropriate, get to know your meeting partners and build relationships.
Use a win/win negotiation strategy	Where possible, avoid a win/lose strategy. The negotiation should support a long term relationship. The idea is that everyone is able to present their interests and reach a middle ground which everyone is comfortable with.

Table 3. Simplified list of skills for bilateral negotiations

Using the *StoryUpgrade* and *PhotoFall* weblog story search tools described above, we gathered eight nonfiction stories related to the skills in Table 3. Four of these stories (*domain stories*) were narrations of experiences of soldiers during recent military deployments in Afghanistan (i.e., negotiations and culture and language issues). The four remaining stories (*analogous stories*) were from the analogous domain of businessmen relating experiences in foreign countries (i.e., buying and selling in marketplaces, dinner table etiquette, corporate culture). For the purpose of our experiment, each of these eight stories was abridged to a single page of text. In several cases, we wrote an introductory paragraph for the story that provided context to the readers that would normally be presented directly on the weblogs from which these stories were gathered.

As a proxy for a population of training developers, we recruited 202 college-educated Americans to participate as subjects in a 35-minute online experiment. Using a within-subjects experimental design with randomized ordering of trials, we tasked subjects to read each of the eight stories, identify the most relevant skill from Table 3, and identify the relation from Table 2 that best describes the nature of this relevance. For each selection task, subjects were given the option to select "None of the above," and were additionally asked to judge their confidence in their selection on a 7-point Likert scale.

To compare inter-coder reliability, we calculated Krippendorff's alpha on the nominal data from this study. Table 4 shows low levels of agreement overall among subjects for both skill and relation types. Agreement is better for domain stories than for analogous stories, and better for skill selection than for relation selection.

	Domain stories			Analogous stories		
	Alpha	*Do*	*De*	*Alpha*	*Do*	*De*
Skill	.448	.133	.241	.156	.192	.228
Relation	.260	.181	.245	.132	.213	.245

Table 4. Krippendorff's Alpha for Skill Annotation and Relation Annotation, with Observed Disagreement (*Do*) and Expected Disagreement (*De*)

To compare confidence scores, we analyzed the data with the Repeated Measures ANOVA, a variation of ANOVA used when the same subjects participate in all conditions of an experiment. Table 5 shows significantly higher confidence when linking domain stories versus analogous stories, both for skill selection and relation selection.

	Domain stories		Analogous stories		F	η^2	p
	M	*SD*	*M*	*SD*			
Skill	5.59	.89	5.05	.99	157.14	.44	<.001
Relation	5.15	1.07	4.91	1.12	27.71	.12	<.001

Table 5. Effect of the type of a story on the level of subjects' confidence for their decisions regarding the *Skill* and the *Relation.*

Overall, the results of this evaluation indicate to us that instructional designers will, indeed, find it more difficult to link stories from analogous domains to task models than those from within the domain of the task. As a consequence, we expect that utilizing the story collections in the Civilian Analogs of Army Tasks will place additional burdens on training developers over the use of stories from within the domain of the training application. The additional challenges of cross-domain story linking should be considered when developing authoring tools for intelligent tutoring systems and assessing learning effectiveness.

Conclusions

In this paper we have explored how technology can support new forms of pedagogical storytelling, where the narrated experiences of practitioners across diverse skill domains can be found and delivered to learners within the contexts of immersive training environments. First, we have shown that stories related to specific activities and skill sets can be harvested from public Internet weblogs, and efficiently retrieved using text retrieval technologies that incorporate relevance feedback. Second, we have shown that analogies can be drawn between activities narrated in public weblogs and a broad range of skillsets that are the focus of training in a large organization, the U.S. Army. Third, we provide a mechanism by which authors of intelligent tutoring systems for immersive training environments can automate the delivery of stories to trainees, linking specific stories to specific items in an explicit task model. Fourth, we conducted an evaluation of the feasibility of this authoring process, finding promising initial results, but also finding that stories from analogous domains may pose additional difficulties for instructional designers.

Ultimately, the technologies and methodologies described in this paper will only be useful if they facilitate learning. What is needed is additional research to evaluate the training effectiveness of pedagogical storytelling in immersive training environments, both for within-domain and analogous-domain narratives.

References

Day, S., and Goldstone, R. (2012) The Import of Knowledge Export: Connecting Findings and Theories of Transfer of Learning. *Educational Psychologist 47*(3):153-176.

Dixon, N., Allen, N., Burgess, T., and Kilner, P. (2005) *Company Command*. West Point, NY: Center for the Advancement of Leader Development & Organizational Learning.

Gick, M. and Holyoak, K. (1983) Schema induction and analogical transfer. *Cognitive Psychology 15*:1-38.

Gordon, A., Wienberg, C., and Sood, S. (2012) Different Strokes of Different Folks: Searching for Health Narratives in Weblogs. ASE/IEEE International Conference on Social Computing, Amsterdam.

Kim, J., Hill, R.W., Durlach, P.J., Lane, H.C., Forbell, E., Core, M., Marsella, S., Pynadath, D., and Hart, J. (2009). BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context. International Journal of Artificial Intelligence in Education, 19(3), 289-308.

Lane, H.C., Hays, M.J., Core, M.G., & Auerbach, D. (2013). Learning intercultural communication skills with virtual humans: Feedback and fidelity. *Journal of Educational Psychology 105*(4):1026-1035.

Swanson, R. (2011) Enabling open domain interactive storytelling using a data-drive case-based approach. Dissertation. University of Southern California.

United States (2012) *The Army Universal Task List: Field manual 7-15*, Washington, DC: Headquarters, Dept. of the Army.

Wienberg, C. and Gordon, A. (2012) PhotoFall: Discovering Weblog Stories Through Photographs. ACM Conference on Information and Knowledge Management (CIKM-2012), Maui, HI.

Acknowledgments

The projects or efforts depicted were or are sponsored by the U. S. Army. The content or information presented does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.