Authoring Team Tutors in GIFT: An Automated Tool for Alignment of Content to Learning Objectives

Benjamin Bell1, Keith Brawner2, Elliot Robson1, Debbie Brown1, Elaine Kelsey1

Eduworks1, US Army RDECOM-NSRDEC2

INTRODUCTION

The process of authoring Intelligent Tutoring Systems (ITS) in the Generalized Intelligent Framework for Tutoring (GIFT) (Sotillare et al, 2013) is assisted by a continually-evolving collection of authoring tools. These tools can accelerate GIFT development by supporting instructional design tasks like sequencing, feedback, adaptation, and assessment. With growing demand for team tutoring in support of rapidly- evolving Army requirements, GIFT tutors must be able to scale learning to meet team training needs, be capable of incorporating broad content; and offer instructional value for both individual Soldiers and teams (Sottilare et al, 2011; Sottilare et al, 2018; Salas et al, 2015; Sottilare et al, 2017b; Fletcher & Sottilare, 2017). A key need is to help ITS authors efficiently find and maintain relevant content, and to assist authors with discriminating between content supporting individual learning objectives and team learning objectives. Addressing this need efficiently calls for automation that supports the analysis of information and its alignment with learning objectives (LOs) (Bonner et al, 2016).

In this paper we introduce a new authoring aid, to be incorporated within GIFT, to help ITS developers find, organize, and curate resources aligned with desired individual and team learning objectives. ***Ma- chine-Assisted Generation of Instructional Content* (MAGIC)** analyzes source documents and extracts content that aligns with specified learning objectives. MAGIC additionally lends much-needed support for team training development by performing this alignment for both individual and team learning objectives. Building on and extending existing artificial intelligence (AI) and natural language processing (NLP) techniques, MAGIC will streamline content alignment, distinguish between individual and team content, and help extend the reach of GIFT tutoring to meet Army team training demands.

MAGIC will contain three layers: backend algorithms and analytics, services and APIs for integration into the GIFT ecosystem, and integrated end user tools for authors. In this paper we describe our initial focus on developing the underlying techniques used by MAGIC and creating a prototype interface for training developers that presents algorithmic outputs. We will describe our work in extending existing NLP and machine learning (ML) libraries to extract and organize learning objectives and integration of these libraries to create an LO repository. A novel aspect of this work is applying ML models to the discrimination between individual and team LOs. We then describe development of automated methods for aligning excerpts of content with LOs and specific roles within a team.

Our plans include implementing an end-user toolset for integration with GIFT to help authors organize LOs and topics and tag, find, sort, and repurpose content that aligns with given LOs and role-based parameters. Finally, we discuss our longer-term plans for incorporating multimedia resources by applying automated transcription techniques. The work we present will advance the state-of-the-art in applying machine learning and NLP to authoring and development of training and in particular team tutoring, and will extend GIFT by supporting authors in collecting and aligning content with individual and team learning objectives. (Gilbert et al, 2017; McCormack at al, 2018; Sotillare et al, 2017a; Sinatra, 2018)

SCALING TEAM TRAINING

Scaling virtual training for teams to fully address Army needs requires tools and techniques for efficiently creating team tutoring simulations. While GIFT supports several instructional design tasks, finding and organizing content that aligns with desired learning objectives remains a labor-intensive process that takes place outside of GIFT. Achieving scale means that virtual training must span broad content. Maintaining relevance means that virtual training must be readily adaptable as learning needs shift in response to equipment upgrades, changes in tactics, evolving threats, and operations in new theaters.

To benefit teams, authors of team tutoring must navigate complicated content management tasks related to distinguishing content that supports individual skills and content aligned with team skills, as well as trying to identify content associated with specific roles within a team. Creating and maintaining virtual team training systems thus remains costly and time-consuming. To help developers of team training find and tag relevant content more efficiently, automation is needed that supports analysis of content and its alignment with team and individual learning objectives.

MAGIC answers this need by helping training developers find, organize, and curate resources aligned with desired learning objectives. MAGIC analyzes source documents and extracts excerpts of content that aligns with specified learning objectives, and performs this alignment for both individual and team learning objectives. Moreover, MAGIC identifies content associated with specific roles within a team. A schematic depiction of MAGIC is shown in Figure 1.



Figure 1. MAGIC at a glance.

STREAMLINING TRAINING DEVELOPMENT BY MATCHING CONTENT TO OBJECTIVES

MAGIC supports three tasks in a generalized GIFT authoring workflow (Figure 2).

Figure 2. Primary authoring process tasks supported by MAGIC.

Using the MAGIC prototype UI, a training developer provides a list of learning objectives and selects the target library (or corpus of documents) to be analyzed as shown in Figure 3. For our initial demonstration of the MAGIC algorithms, we drew learning objectives from battle drills in the Maneuver domain; for the library we used the Central Army Registry (CAR) and the Milgaming portal’s Training Support Packages (TSPs) to create a collection of over 1,200 documents.

Figure 3: Selecting LOs and corpus documents to configure a content analysis.

MAGIC then generates a collection of text excerpts from across the selected documents, each tagged by the learning objectives, individual or team types, and team roles the excerpt aligns with. In the current demonstration interface, these results may be viewed, filtered, and compared with human rater results when available (Figure 4). In future work, the toolset will offer more flexible export packaging options designed to integrate into GIFT repository search and authoring components using the MAGIC API.



Figure 4. Filtering content by LO, task type, and role

MACHINE LEARNING: THE MAGIC BEHIND MAGIC

MAGIC uses ML and NLP techniques to train algorithms that associate content with learning objectives, tag content as having individual or team relevance, and associate content with specific team roles when applicable. We developed three sets of ML models for our initial research and testing: (1) *unsupervised general* models trained using Wikipedia and the New York Times Annotated Corpus to map concepts; (2) *unsupervised domain-specific* models trained with military-sourced documents to define domain-specific concepts; (3) *supervised, domain-specific* models trained with human-tagged data from a team of instruc- tional designers and subject-matter experts to enhance outcomes.

In the case of the battle drill use cases, we manually created learning objectives (LOs) outlined as hierar- chical task procedures, based on original document text, and manually tagged content with task type and role as depicted in Figure 5. The manually tagged LOs were used to train the ML algorithms for task type and role detection.



|  |  |  |
| --- | --- | --- |
|  |  |  |

Figure 5. Example learning objectives for a battle drill.

To create the tagged data we used a team of three human raters with instructional design, research, and military backgrounds, led by an expert in instructional design. Raters were trained on the rating task, which included scoring relevance of sections of content to a learning objective and tagging with individu- al/team and team role identifiers. The resulting tagged data set consists of 3,132 tagged items and was segmented into two corpora: one for training the supervised learning models, and one for evaluating performance of all three ML model sets. The average interrater reliability (n=3) was 81.6% for text selection and extraction, 87.8% for distinguishing team and individual content, and 78% for identifying team roles.

NOVEL SOLUTIONS

A challenge MAGIC addresses is matching content excerpts to a learning objective (typically a short text string) rather than to a topic (typically supported by larger amounts of descriptive text). To address this difficulty, we extended existing work in word embedding approaches (e.g. Word2Vec, GLoVe) (Mikolov et al, 2013; Pennington et al, 2014) , to develop a new technique we refer to as *concept embedding*. The approach first involves parsing an input corpus of documents to detect entities and relations as short phrases (rather than as individual words) using TensorFlow- or SyntaxNet-style dependency parsing along with traditional ontological approaches (Goldberg & Levy, 2014). In the next step, we build corpus models using the resulting dependency trees as the input into distinct entity and relation embedding models, where ‘concepts’ are defined as tight clusters of phrases in the resulting vector spaces (Levy & Goldberg, 2014). By mapping entities and relations separately, and then linking them through a combined (modified W2V-SG) model, we are able to instantiate concepts as tight clusters of phrases that exist in the resulting entity and relation vector spaces. For example, this approach might instantiate the concept “*Santa Claus*” as associated with “*Jolly Old St. Nick*” and “*the fat man in the red suit*.” (Li et al, 2016, Shalaby et al, 2018),

This concept embedding approach gives MAGIC the ability to extract a richer description of meaning from very short text strings (namely, learning objectives). In our use case, the approach is applied in multiple steps to perform excerpt extraction:

* Extract entities and relations from the LOs
* Generate an *embedding space*
* Map entities to concepts
* Use any available context to disambiguate between concepts
* Map documents to the concept space (both concept and topic levels)
* Match concepts in each LO to concepts in the corpus
* Rank results based on match to both entity-concepts & relation-concepts of given LO

In order to discriminate between individual or team LO types, we applied a hybrid ML approach that was combined with syntactic-semantic patterns (Kelsey et al, 2017). On the ML side, we first extracted the semantic and syntactic features and tested using Naïve Bayes and Support Vector Machine (SVM) classification techniques which produced similar results. However, these two approaches were more accurate and required less training data than either a Convolutional Neural Network (CNN) or Recurrent Neural Network (RNN) implementation. On the Syntactic-Semantic side, we extracted combined syntac- tic-semantic features using SyntaxNet with TensorFlow, and then matched using the pattern library. We achieved the best results by applying both the ML and Syntactic-Semantic Pattern approaches and then using context-specific heuristics (where ‘context’ is derived from features of the source document and larger source text) to resolve any disagreements when selecting the team or individual label.

When identifying an appropriate team role for an excerpt, we determined that the link to LOs/competency frameworks can provide important role implications as well as provide a predefined list of possible roles. Our approach was to expand each role into a Concept using the Concept Embedding Model, and then to apply a similar matching approach. We continue to take steps to improve results with role assignment by using human-labelled data to detect discourse and semantic-syntactic markers for a list of common domain-specific roles. The application of a supervised learning layer using human-tagged samples is expected to further enhance MAGIC outcomes, with a goal of achieving accurate extractions and tag selections more often than the human raters.

PRELIMINARY RESULTS

To provide early metrics of MAGIC’s performance, we used the second set of labeled data as a test set. Both the training and test sets comprised approximately 5,000 comparisons of a text excerpt to a learning objective, and each task was completed by the three independent raters. Interrater reliability was 81.6%.

Figure 6. Preliminary results for each of MAGIC’s ML models.

The results (Figure 6) demonstrate the algorithms performing slightly below human performance when using only the domain-general unsupervised model, at or near human performance when adding the unsupervised domain-specific model, and slightly above human performance when adding the supervised domain-specific model.

CONCLUSIONS AND FUTURE WORK

With preliminary results already meeting human-rater levels of reliability using the combined unsuper- vised general and domain specific models, and with the addition of a supervised domain-specific model performing better than the human raters, the MAGIC approach is showing promising results and a path for continued enhancement. Based on these early findings, we see the potential for automated content discovery using LO auto-alignment and text extraction will result in faster, scalable team training devel- opment processes. Integration of MAGIC services into the GIFT authoring workflows will propel reuse of training materials, while helping training developers overcome the challenges of distinguishing content supporting team or individual learning and aligning content with specific team roles.

Our next steps in the MAGIC project will include creating a supervised domain-specific model for assigning team roles; incorporating non-text content (such as metadata or automated transcriptions); designing a MAGIC services API; testing and evaluation of MAGIC with authors of team training simulations; and the integration of MAGIC services with Army-selected authoring/CMS/LMS tools.

REFERENCES

Bonner, Desmond; Gilbert, Stephen B.; Dorneich, Michael C.; Winer, Eliot; Sinatra, Anne M.; Slavina, Anna; MacAllister, Anastacia; and Holub, Joseph, (2016) "The Challenges of Building Intelligent Tutoring Sys- tems for Teams". *Industrial and Manufacturing Systems Engineering Conference Proceedings and Posters*. 23. Retrieved from <http://lib.dr.iastate.edu/imse_conf/23>

Fletcher, J. D., & Sottilare, R. A. (2017). Shared mental models in support of adaptive instruction of collective tasks using GIFT. *International Journal of Artificial Intelligence in Education*. [https://doi.org/10.1007/s40593-](https://doi.org/10.1007/s40593-017-0147-y) [017-0147-y.](https://doi.org/10.1007/s40593-017-0147-y)

Gilbert, S., Slavina, A., Dorneich, M., Sinatra, A., Bonner, D., Johnston, J., Holub, J., MacAllister, A., and Winer, E. (2017). Creating a team tutor using GIFT. *International Journal of Artificial Intelligence in Educa- tion.* [https://doi.org/10.1007/s40593-017-0151-2.](https://doi.org/10.1007/s40593-017-0151-2)

Goldberg, Y., & Levy, O. (2014). word2vec Explained: deriving Mikolov et al.'s negative-sampling word- embedding method. arXiv preprint arXiv:1402.3722.

Kelsey, E. Goetschalckx, R. Robson, E., Ray, F. & Robson, R. Automated Tools for Question Generation. United States Patent Application 20180260472

Levy, O., & Goldberg, Y. (2014). Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Volume 2: Short Papers) (Vol. 2, pp. 302-308).

Li, C., Wang, H., Zhang, Z., Sun, A., & Ma, Z. (2016, July). Topic modeling for short texts with auxiliary word embeddings. In *Proceedings of the 39th International ACM SIGIR conference on Research and Develop- ment in Information Retrieval* (pp. 165-174). ACM.

McCormack, R., Kilcullen, T., Sinatra, A.M., Brown, T., & Beaubian, J. (2018). Scenarios for Training Teamwork Skills in Virtual Environments with GIFT. In Proceedings of the 6th Annual *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6)*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Salas, E. , Shuffler, M. L., Thayer, A. L., Bedwell, W. L. and Lazzara, E. H. (2015), Understanding and Improving Teamwork in Organizations: A Scientifically Based Practical Guide. *Human Resource Management*, 54: 599-622. do[i:10.1002/hrm.21628](https://doi.org/10.1002/hrm.21628)

Shalaby, W., Zadrozny, W., & Jin, H. (2018). Beyond word embeddings: learning entity and concept representations from large scale knowledge bases. *Information Retrieval Journal*, 1-18.

Sinatra, A.M. (2018). Team Models in the Generalized Intelligent Framework for Tutoring: 2018 Update. In Proceedings of the 6th Annual *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6)*.

Sottilare, R., Holden, H., Brawner, K., & Goldberg, B. (2011). Challenges and emerging concepts in the develop- ment of adaptive, computer-based tutoring systems for team training. In *Proceedings of the Interserv- ice/Industry Training Systems & Education Conference*, Orlando, December 2011.

Sottilare, R., Holden, H., Goldberg, B., & Brawner, K. (2013). The Generalized Intelligent Framework for Tutoring (GIFT). In Best, C., Galanis, G., Kerry, J. and Sottilare, R. (Eds.) *Fundamental Issues in Defence Simula- tion & Training*. Ashgate Publishing.

Sottilare, R., Brawner, K., Sinatra, A. & Johnston, J. (2017a). An updated concept for a generalized intelligent framework for tutoring (GIFT). Orlando: US Army research laboratory. May

2017. [https://doi.org/10.13140/RG.2.2.12941.54244.](https://doi.org/10.13140/RG.2.2.12941.54244)

Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2017b). Towards a design process for adaptive instruction of Teams: A Meta-Analysis*. International Journal of Artificial Intelligence in Education*. [https://doi.org/10.1007/s40593-017-0146-z.](https://doi.org/10.1007/s40593-017-0146-z)

Sottilare, R., Graesser, A., Hu, X., and Sinatra, A. (Eds.). (2018). *Design Recommendations for Intelligent Tutoring Systems: Volume 6 - Team Tutoring*. Orlando, FL: U.S. Army Research Laboratory. ISBN 978-0-9977257- 4-2. Available at: <https://gifttutoring.org/documents/>and on Google Play.

ABOUT THE AUTHORS

**Dr. Benjamin Bell** is the Principal Investigator on the MAGIC project and president of Eduworks. He has been been leading funded research in education, training and simulation for over twenty years, with an emphasis on defense and national security applications. He holds a Ph.D. from Northwestern University in Artificial Intelligence, Master’s degrees from Embry Riddle and Drexel University, and a Bachelor’s degree from the University of Pennsylvania.

**Dr. Keith Brawner** is a senior researcher for the U. S. Army Combat Capability Development Command Soldier Center at the Simulation and Training Technology Center (CCDC-SC-STTC), and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). He has 13 years of experience within U.S. Army and Navy acquisition, development, and research agencies. He holds a Masters and PhD degree in Computer Engineering with a focus on Intelligent Systems and Machine Learning from the University of Central Florida. His current efforts are on artificial intelligence for the Synthetic Training Environment Simulation and Network Compression. He manages research in adaptive training, semi/fully automated user tools for adaptive training content, and architectural programs towards next-generation training.

**Elliot Robson** leads research and development for Eduworks corporation as General Manager. He has fifteen years of experience in training technology and has served as PI on projects using GIFT.

**Debbie Brown** acts as Learning Engineer, Instructional Designer, and Software Developer at Eduworks, and has been active in eLearning research and development for 20 years in academic, workforce, and government/military applications. Since joining Eduworks, she has served on multiple projects using GIFT.

**Elaine Kelsey** is a research engineer at Eduworks, focusing on applications of natural language processing and machine learning in topic and concept detection and automated text generation. She designed and led the develop- ment of Eduworks’ automated question generation algorithm, and is currently developing solutions in human- interpretable machine learning. She has multiple Bachelors and Masters degrees in computer science, linguistics, and molecular biology, and is currently working towards a PhD in Computational Linguistics.