A Data Mining Approach to Construct Production Rules in an Educational Game

Fazel Keshtkar, Borham Samei, Brent Morgan, and Arthur C. Graesser

University of Memphis
Institute for Intelligent Systems
Memphis, TN, USA.
{fkshtkar,bsamei,brent.morgan,a-graesser}@memphis.edu

Abstract. One of the most crucial aspects of Intelligent Tutoring Systems in a collaborative serious game is production rules. Given the large number of interactions and conversation between players, it is difficult to follow student questions and reactions in the game environment. Therefore, creating a sophisticated method to construct production rules for handle the students' interactions will boost the performance of the system. In this paper, we propose a state-of-the-art computational approach to automatically generate production rules using co-occurrences of distinct terms from a corpus of students' conversations. Moreover, our model is able to generate additional production rules as new data is available. Finally, we also introduce how to transfer extracted co-occurrences into production rules, and how to build these into the game system.

Keywords: Intelligent Tutoring Systems, Production Rules, Data Mining

1 Introduction

Serious games are increasingly becoming a popular, effective supplement to standard classroom instruction [9]. Some classes of serious games provide microwords [7] that allow players to explore a virtual environment. These simulations have ideal and often simple problems with targeted scaffolding to help users identify important concepts and think critically about them. Multi-party chat is pervasive in recreational games and often crucial to success in multi-player epistemic games [4, 3, 8]. In this paper, we present a method of production rule We employed a computational approach to determine the critical features of multi-party chat in a serious game. We analyzed a corpus of chat conversations and high frequency features along with their co-occurrences. We describe the resulting model below, as well as the process of generating production rules. Finally, we discuss how to utilize this model in the context of a serious game to provide relevant suggestions to a human mentor.

2 Production Rules

A Production Rule consists of a collection of *IF...THEN* rules that together form an information processing model of some task, or range of tasks. Each rule has two parts: a condition part and an action part. Production rules can be represented in various forms [2], e.g.: “IF condition THEN action”, “IF premise THEN conclusion” or on the other hand “IF proposition p_1 and proposition p_2 are true THEN proposition p_3 is true”. In the context of a serious game, for example, it is likely that the players will eventually need help navigating the user interface. Whereas they would normally ask a human mentor to guide them, if a relevant production rule is built in the system, this situation can easily be detected and resolved by the system, saving the resources of the human mentor. The system outlined below must be able to detect the specific facts or features (such as “email” and “check”) to specify relevant conditions and return the appropriate suggestion. As a result, a computational data mining approach helped us to extract these conditions and facts.

2.1 Speech Act Classification

We selected a system for classifying speech acts [5]. Analyzes of a variety of corpora, including chat and multiparty games, have converged on a set of speech act categories that are both theoretically justified and that also can be reliably coded by trained judges [6]. Our classification scheme has 8 broad categories: **Statements** are verbal reports on scientific facts. **Requests** include asking other participants in the conversation to provide information. **Questions** are queries for information from the addressee. **Reactions** are short verbal responses to requests or questions. **Expressive Evaluations** consist of feedback regarding the player’s performance. **MetaStatements** are statements about the communication process. **Greetings** are expressions regarding any party’s entrance to. **Other** represents speech acts which did not fit into the above categories.

2.2 Land Science Game

Urban Science is an epistemic game created by education researchers at the University of Wisconsin-Madison, designed to simulate an urban planning practicum experience [1]. Young people role-play as professional urban planners in an ecologically-rich neighborhood. The players’ primary task is to redesign the city of Lowell, Massachusetts. Players are assigned to one of three planning teams, and interact with team members and a human mentor using a group chat interface [4, 3, 8]. The ”Question” category is likely the most critical speech act when it comes to addressing player problems. We analyzed 26720 unique chat turns across three instances of Land Science data set.

3 Our Approach

In our model, we identify the relevant facts needed to satisfy the conditions in IF ... $THEN$. Based on these facts, we are able to generate suggestions for a human mentor. In our algorithm, facts can have any of the following features: words, tokens, event, status of the game, or patterns of player’s conversation.

Table 1. shows some of tokens that have high co-occurrence

Token 1	Token 2	co-occ	categories	rooms
stakeholders	what	413	Statement Question Request Reaction	7 5 3 4 12
email	maggie	353	Statement Request Question ExpressiveEval	3 2 4 5 6
want	what	306	Statement Request Question Reaction	5 7 3 4 12
meeting	team	281	Statement Request Question ExpressiveEval	4 3 2 10 11
out	what	280	Statement Request Question Reaction	7 4 3 5 2
now	what	262	Statement Question Request Reaction	7 3 2 10 1
find	out	257	Statement Question Reaction Request	5 10 4 3 7
final	proposal	237	Statement Reaction Request Question	12 2 11 3 6
preference	survey	236	Statement Request Reaction Question	6 9 7 5 10
stakeholders	want	229	Statement Request Question Reaction	5 7 4 3 12

Using these facts, we can generate production rules which offer suggestions for a human mentor.

3.1 Computing Co-Occurrences

One of the most important features to build production rules based on a data-mining approach is to determine the co-occurrences of high or even low frequency tokens in the corpus. In the following sections we describe these features and we show how they can be considered as conditions and facts in our production rules. After preprocessing the corpus, we split each utterance into tokens using the OpenNLP tokenizer, a Natural Language Processing Java Library. We used standard stop words to remove unnecessary tokens. We computed the frequency of all remaining tokens in the corpus for each Speech Act category. These tokens are based on Unigram Entropy Cues and Speech Act classification method that developed by [5]. Then, we ranked these frequencies list from high to low order. In addition to token frequency, it is also critical to assess the relevance of each token, as it may be context-specific. We assessed token relevance by computing co-occurrences. Table 1 shows some examples of co-occurrences in our corpus. In Table 1 tokens that have high co-occurrence chance along with the categories and rooms they appeared in. The categories and rooms are ordered by the frequency of the co-occurrence.

3.2 Constructing Production Rules

As we described in previous sections, Production Rules are in forms of IF ... $THEN$ statements. These IF ... $THEN$ statements must obtained by the **Conditions** and the **Facts** to achieve some **Conclusions** or **Actions**. By looking at Table 1, we see the co-occurrences for “Virtual” are: navigation, stakeholder, neighborhood, character, site, during, etc. In our model, we assume that the facts for conditions can be one or more of the co-occurrences for each token.

4 Conclusion and Future Work

In this paper, we discussed the concept of production rules. These rules are $IF...THEN$ statements which contain some conditions (based on relevant facts). When conditions are met, they trigger some system response, such as a suggestion to players from a mentor or intelligent agent. We introduced a state-of-the-art data-mining approach to construct production rules from a corpus of chat conversations. For future work, we plan to use rule based model to generate production rule. This will allow us to fire relevant functions to produce better suggestions. We also plan to analyze more data to construct additional production rules for the Land Science.

Acknowledgments

This work was funded by the National Science Foundation (DRK-12-0918409).

References

1. Bagley, E.: Stop Talking and Type: Mentoring in a Virtual and Face-to-Face Environmental Education Environment. Ph.D. thesis, University of Wisconsin-Madison (2011)
2. Compton, P., J.R.: A philosophical basis for knowledge acquisition. knowledge acquisition 1990; 2: 241-257
3. D. Ketelhut, C. Dede, J.C.B.N.C.B.: Studying situated learning in a multi-user virtual environment. In: Baker, E., Dickieson, J., Wul-feck, W., ONeil, H. (eds.) Assessment of problem solving using simulations, pp. 37-58. Earlbaum, Mahweh (2007)
4. Dieterle, E., Clarke, J.: Multiuser virtual environments for teaching and learning. In: Pagani, M. (ed.) Encyclopedia of multimedia technology and networking (2nd ed). Idea Group, Hershey 2012.
5. Moldovan, C., Rus, V., Graesser, A.C.: Automated speech act classification for online chat. In: Procceding of FLAIRS 2011 (2011)
6. R.G. DAndrade, M.W.: Speech act theory in quantitative research on inter-personal behavior. In: Discourse Processes. 8 (2), 229-259 (1985)
7. Rickel, J., Lesh, N., Rich, C., Sidner, C., Gertner, A.: Building a bridge between intelligent tutoring and collaborative dialogue systems. In: Proceedings of the 10th Int. Conf. on Artificial Intelligence in Education, San Antonio, TX, pp. 592-594, May 2001.
8. Shaffer, D.: How Computer Games Help Children Learn. Palgrave, New York (2007)
9. Shaffer, D., Chesler, N., Arastoopour, G., DAngelo, S.: Nephrotex: Teaching first year students how to think like engineers. Laboratory Improvement (CCLI) PI Conference (2011)