Proc. 21th Behavior Representation in Modeling & Simulation (BRIMS) Conference, Extended Abstract
Amelia Island, FL, March 12-15, 2012, pp. 271-218

Behavior Modeling and Classification via Probabilistic Context Free Grammars

Sahin Cem Geyik, Boleslaw K. Szymanski
Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180
geyiks@cs.rpi.edu, szymansk@cs.rpi.edu

Keywords:
Team behavior/interaction/coordination, Pattern recognition, Socio-cultural modeling

Social networks analysis includes examining the actions of entities in a social setting. These actions can be either interactions between entities (e.g. talking, exchanging items etc.), or actions which do not include interactions, but nevertheless are happening in a social context, hence are influenced by social relations. Such actions often contain behavioral patterns that are specific to the actions involved. It is important to understand such patterns to be able to model social environments reliably.

In this work, we introduce a novel method for modeling and classifying behavior of nodes in a social network using Probabilistic Context Free Grammars (PCFGs). Informally, PCFGs are regular context free grammars (consisting of START symbol, terminals, nonterminals and production rules), augmented with probabilities assigned to the production rules to denote how likely each rule is to be used when producing a sentence from this grammar. Given a set of action sequences, a PCFG can be automatically constructed [Geyik, 2009] to derive probabilities for these actions and concisely represent behavioral patterns based on the input data. This PCFG can either be used to predict or classify future behaviors, or to understand the relationships between these patterns via manual inspection possible thanks to conciseness of the PCFG representation.

To evaluate our proposed methodology, we present the results from processing the Mission Survival Corpus (MSC-1) dataset [Pianesi, 2008] collected by Project FBK (Fondazione Bruno Kessler). This dataset basically contains time-stamped annotations of 11 meetings of people deciding on how to proceed in a disaster scenario. The annotations include the social role label and task role label of each meeting attendant as well as an indication of who speaks at the time-stamp. Social roles (supporter, protagonist, attacker, and neutral) mainly represent the attendant's attitude towards the group's function while the task roles (giver, seeker, orienteer, neutral) represent the individual's function and technical skills. We examine the task roles taken (as a sequence) by a meeting attendant while undertaking a social role (socio label), or vice versa (we call this metric "*Which Roles Go Together*"). In our previous work [Geyik, 2010], we presented the constructed grammars and provided our interpretation of them based on manual examination that led us to understanding the underlying properties of this application domain. As a simple example, the grammar constructed for the *attacker* social role within the MSC-1 dataset is shown in Figure 1.

START → N0 M1 (0.18) M1 (0.18) M1 N0 (0.55) N0 (0.09)	N0 → n (1.0) - - - - - - - - - - M1 → s (0.5) | g (0.5)

Figure 1. A Sample Grammar for the Attacker Socio-Label (n-neutral, s-seeker, g-giver)

An inspection of this grammar shows that for this dataset, seeker (terminal s) and giver (terminal g) have equal probability (as defined by the right hand side of the production defining M1) for the attacker social role (i.e. an attacker is equally likely to ask questions as it is to provide facts). Such domain specific property can be easily observed due to the concise representation achieved by the PCFG modeling.

As an extension to our previous efforts, we present here the classification results of the PCFG modeling on the MSC-1 dataset. For this purpose, we utilize the aforementioned metric, "*Which Roles Go Together*", to classify the roles of meeting attendants. We first provide the separation capability of the PCFGs between roles when we utilize the whole dataset; and then, we separate the dataset into training and test partitions and evaluate the classification ability of PCFGs on the unseen data.

Table 1 presents the classification results for the social roles in MSC-1 dataset. First, four grammars (for each of four social roles) are constructed automatically, and then we try to parse the sequence that is to be classified by these four grammars. Similar to a Naive Bayesian Classifier, the classification metric depends on the sequence's production probability given a grammar, and the prior probability of the class itself (according to how frequently the role is observed in the training data). To account for unseen sequences in the training data, we inserted a smoothing nonterminal into each grammar to accept all sequences (but with a very small probability if it was not in the training data).

Between the Roles of	Separation Accuracy	2-Fold Cross Val.	10-Fold Cross Val.
Protagonist and Supporter	69.7 %	65.4 %	65.3 %
Neutral and Supporter	71.4 %	69.2 %	66.6 %
Neutral and Protagonist	63.5 %	63.6 %	60.3 %
Attacker and Supporter	94.3 %	92.6 %	93.7 %
Attacker and Protagonist	96.1 %	95.4 %	95.8 %
Attacker and Neutral	97.3 %	97 %	96.9 %
Neutral, Protagonist, and Supporter	51.7 %	51.1 %	48 %
Attacker, Protagonist and Supporter	68.2 %	63.7 %	63.7 %
Attacker, Neutral, and Supporter	70.1 %	67.8 %	65.2 %
Attacker, Neutral, and Protagonist	62.6 %	62.5 %	59.2 %
Attacker, Neutral, Protagonist, and Supporter	51.1 %	50.4 %	47.3 %

Table 1. Separation and Classification Results for the Social Roles in MSC-1 Dataset

From the above table, it can easily be observed that the attacker social role is easily distinguished from the rest of the social roles, while the others are harder to distinguish according to their action role sequences. We also see that the 10 and 2-fold cross validation results are not much worse than testing on the training data itself, which demonstrates the classification ability of the PCFG model on unseen data.

Table 2 presents the classification results for the task roles in MSC-1 dataset. From the table, it can be observed that the task role classes neutral and giver are easy to distinguish from the task role classes seeker and orienteer in terms of their social role taking, but the classification performs worse between the classes in these two groups (i.e. neutral is difficult to distinguish from giver, and seeker is difficult to distinguish from orienteer).

Future work directions in this area include various other application domains within social network context where the PCFGs can be used for prediction and classification. Examples include role recognition in automatic analysis of videotapes of clandestine meetings or the monitoring and classification of normal vs. abnormal behavior of people in other security applications. Furthermore, a question that is interesting for future work direction is whether a certain social stance is distinguishable for different application domains (i.e. the question “Is attacking behavior always as distinct as it is for the MSC-1 dataset?”).

Between the Roles of	Separation Accuracy	2-Fold Cross Val.	10-Fold Cross Val.
Orienteer and Seeker	63.6 %	66.9 %	67.7 %
Neutral and Seeker	88 %	87.8 %	86.7 %
Neutral and Orienteer	84.4 %	82.1 %	81.3 %
Giver and Seeker	87.9 %	87.2 %	87.4 %
Giver and Orienteer	83.8 %	81.9 %	81.8 %
Giver and Neutral	59.2 %	57.2 %	55.5 %
Neutral, Orienteer and Seeker	75.8 %	73.5 %	72.4 %
Giver, Orienteer and Seeker	75.3 %	72.9 %	73.1 %
Giver, Neutral, and Seeker	55.3 %	53.4 %	51.8 %
Giver, Neutral, and Orienteer	54.1 %	51.5 %	49.6 %
Giver, Neutral, Orienteer, and Seeker	50.8 %	48.4 %	46.6 %

Table 2. Separation and Classification Results for the Task Roles in MSC-1 Dataset

This work was supported in part by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053 and by the Office of Naval Research Grant No. N00014-09-1-0607.

References

Geyik, S. C., Szymanski, B. (2009). Event Recognition in Sensor Networks by Means of Grammatical Inference. *In Proc. IEEE INFOCOM*, 900-908.

Pianesi, F., Zancanaro, M., Not, E., Leonardi, C., Falcon, V., Lepri, B. (2008). Multimodal Support to Group Dynamics. *Personal and Ubiquitous Computing, Vol. 12, Iss. 3*, 181-195.

Geyik, S. C., J. Xie, J., Szymanski, B. (2010). Behavior Modeling with Probabilistic Context Free Grammars. *In Proc. Int. Conf. on Information Fusion*.

Author Biographies

SAHIN CEM GEYIK is currently a Ph.D Candidate in the Computer Science Department, at Rensselaer Polytechnic Institute (RPI), Troy, NY. He received his BS degree in Computer Engineering from Bogazici University in Istanbul, Turkey, in 2007.

BOLESLAW K. SZYMANSKI is the Claire and Roland Schmitt Distinguished Professor of Computer Science and the Founding Director of the Center for Pervasive Computing and Networking at Rensselaer Polytechnic Institute (RPI), Troy, NY. He received his Ph.D. degree in computer science from the National Academy of Sciences, Warsaw, Poland, in 1976.