

# The SOMA Terror Organization Portal (STOP): social network and analytic tools for the real-time analysis of terror groups

Amy Sliva<sup>†</sup>, V.S. Subrahmanian<sup>†</sup>, Vanina Martinez<sup>†</sup>, and Gerardo I. Simari<sup>†</sup>

<sup>†</sup>{asлива,vs,mvm,gisimari,asлива}@cs.umd.edu, Lab for Computational Cultural Dynamics, Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA

**Abstract** Stochastic Opponent Modeling Agents (SOMA) have been proposed as a paradigm for reasoning about cultural groups, terror groups, and other socio-economic-political-military organizations worldwide. In this paper, we describe the SOMA Terror Organization Portal (STOP). STOP provides a single point of contact through which analysts may access data about terror groups world wide. In order to analyze this data, SOMA provides three major components: the SOMA Extraction Engine (SEE), the SOMA Adversarial Forecast Engine (SAFE), and the SOMA Analyst Network (SANE) that allows analysts to find other analysts doing similar work, share findings with them, and let consensus findings emerge. This paper describes the STOP framework.

## 1 Introduction

Stochastic Opponent Modeling Agents introduced in [1,2,3] were introduced as a paradigm for reasoning about any group  $G$  in the world, irrespective of whether the group is a terror group, a social organization, a political party, a religious group, a militia, or an economic organization. SOMA-rules have been used to encode the behavior of players in the Afghan drug economy [4] as well as various tribes along the Pakistan-Afghanistan border [5], as well as terror groups such as Hezbollah [6]. Continuing studies are expected to track over 300 terror groups by end 2008.

The study of terror groups in the national security community is hampered by many major problems.

1. **Lack of timely, accurate data.** Data about the terror groups is collected manually. This causes any data to be either incomplete or out of date (usually both). One of the best known collections of data about terror groups is the “Minorities at Risk Organizational Behavior” (MAROB) data set [7,8] which tracks approximately 400 properties of ethnopolitical

groups representing 284 ethnic groups worldwide. As mentioned in [5], such manual data collections are incomplete (because the manual coders of this data are only able to process certain numbers of articles in a limited number of languages) and coarse grained (because the variables are classified into categories rather than specific numbers – for instance, the number of deaths caused by a group might be classified as “none”, “low”, “medium” or “high” instead of giving a reasonably accurate estimate). Moreover, they are usually out of date (Minorities at Risk coding is currently complete only until 2003), leaving the data to be “non-actionable”.

2. **Lack of behavioral models.** Analysts today are forced to make up behavioral models in a slow and painstaking manner. The SOMA Extraction engine (SEE), a portion of STOP, extracts behavioral rules automatically from the MAROB data. To date, it has extracted rules on approximately 23 groups with another 2 or 3 groups being analyzed and completed each week.
3. **Lack of a social network.** Analysts are often unaware of what other analysts have found useful. There is no computational mechanism we have seen to date that allows analysts to share their experiences and forecasts for a particular group. In fact, analysts are often in the dark about who else is looking at a certain group – either from the same, or different viewpoint as themselves. The **SOMA Analyst Network (SANE)** provides a social networking framework within which analysts can browse data about the groups of interest, rate the accuracy of the data, browse rules extracted by SEE, rate the rules extracted by SEE, and see the predictions produced by SAFE.
4. **Lack of a forecast engine.** The SOMA Adversarial Forecast Engine (SAFE) uses a rich probabilistic foundation with no unwarranted independence assumptions in order to forecast the most probable sets of actions that a terror group might take in a given situation. Using SAFE, analysts can analyze the current situation and/or hypothesize new situations based on possible actions the US might be contemplating.

**We will not be discussing problem (1) above in much detail in this paper because our CARA architecture described in [9] already addresses many of the problems of dealing with real time data. Components of CARA such as OASYS [10] and T-REX [11] extract valuable data from open sources in real time.**

## 2 SOMA Extraction Engine

The SOMA Extraction Engine (SEE) is intended to derive SOMA-rules automatically from real-time sources such as T-REX [10] and OASYS[11]. These systems automatically extract several kinds of events. They are being tuned to extract social science codes such as those used in the MAROB effort. Till that is complete, the SOMA extraction engine is directly working with the MAROB data to extract rules.

A SOMA rule about a group  $G$  has the form

$$\langle \text{Action} \rangle : [L, U] \text{ if } \langle \text{Env-Condition} \rangle$$

Where:

- $\langle \text{Action} \rangle$  is an action that the group took (such as KIDNAP)
- $\langle \text{Env-Condition} \rangle$  is a logical conjunction of elementary conditions on the environmental attributes. An elementary condition associated with the environmental attribute  $A$  is an expression of the form  $A \text{ op value}$  where  $\text{op}$  is in the set  $\{ =, <=, >= \}$ .
- $[L, U]$  is a closed sub-interval of the  $[0, 1]$  interval.

The above rule says that in any year when the  $\langle \text{Env-Condition} \rangle$  is true, there is a probability between  $L$  and  $U$  that the group took the action stated in the rule. The rule below is an example of a rule that we extracted about Hezbollah.

KIDNAP:  $[0.51, 0.55]$  if solicits-external-support & does not advocate democracy.

This rule says that in years when Hezbollah both solicited external support and did not promote democratic institutions, there was a 51 to 55% probability that they engaged in kidnapping as a strategy.

The SEE Algorithm works in accordance with the sketch given below. It takes as input, a group  $G$ , along with a table  $Tab(G)$ , for the group. This table contains one column for each attribute of this group, and one row for each year. The attributes are split into *action* attributes (which describe the actions the group took in a given year) and *environmental* attributes (which describe the environment in which the group functioned during that year – these can include actions that *other* organizations took against group  $G$ ). SEE takes an action name  $a$  and a code  $c$  for that action as input (e.g. KIDNAP may be an action and the code 0 might indicate the code for no kidnappings). Based on these inputs, SEE attempts to find all conditions on the *environmental* attributes that predict  $KIDNAP=0$  (i.e. that kidnappings will not be resorted to by the group) with a high degree of probability. Clearly, we are only interested in conditions which, when true, predict  $KIDNAP=0$  with high probability, and which when false, predict  $KIDNAP=1$  with high probability. In conditional probability terms, we

are interested in finding conditions  $C$  on the environmental attributes such that  $\text{abs}(\mathbf{P}(\text{KIDNAP}=0|C) - \mathbf{P}(\text{KIDNAP}=0|\sim C))$  is high. As usual,  $\mathbf{P}(\text{KIDNAP}=0|C)$  denotes the conditional probability of the value of the KIDNAP attribute being 0 in a given year, given that condition  $C$  holds for that year.

Of course, even when we find such a condition  $C$  above, we would like to ensure adequate *support*, i.e. to ensure that there is a sufficiently large number of years in which both  $C$  and  $\sim C$  are true.

The SEE system implements the algorithm in [2] to ensure that the conditions found satisfy both these requirements.

As of the time of writing this paper (early Dec. 2007), SEE has extracted behavioral models of 23 groups worldwide. The list includes 8 Kurdish groups spanning Iran, Turkey, and Iraq (including groups like the PKK and KDPI), 8 Lebanese groups (including Hezbollah), several groups in Afghanistan, as well as several other Middle Eastern groups.

### 3 SOMA Analyst Network (SANE)

STOP requires an analyst to register and log in prior to accessing the system. Once the user is logged in, she proceeds via four phases.

**Phase 1. Selecting a Group.** In this phase, the user selects a group to study by examining a drop down list of groups that have been processed so far. Figure 1 below shows this screen when the user has selected the PKK as his group of interest.

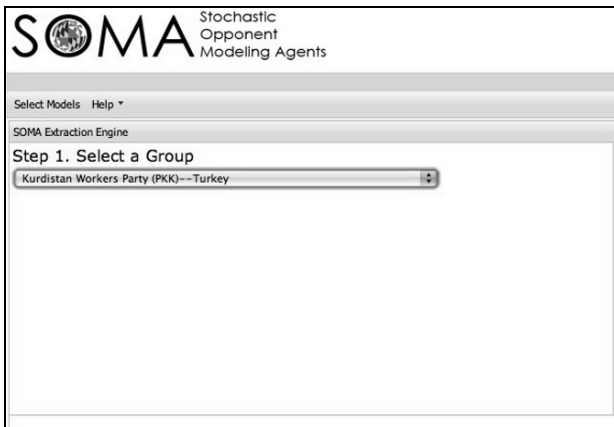


Figure 1: The STOP user has selected the PKK for analysis.

**Phase 2. Selecting an Action.** Once the user selects a group, a list of actions that this group has engaged in during the period of study (typically the last 25 years or a bit less, depending on how long the group has been in existence) pops up. The user can select one or more actions to study. Figure 2 shows this screen when the user has selected the action “Arms trafficking.” Note that there is a number next to each action. This number shows the number of rules we have for the group for this action. For instance, Figure 2 shows that we have 28 rules about “Armed Attacks” for the PKK.



Figure 1 The STOP user has selected the action Arms trafficking. There are 10,873 rules about "Arms trafficking" in the system.

**Phase 3. Selecting Conditions.** Once the user selects an action(s), the system tries to find all rules extracted by SEE about the group dealing with the action(s) selected in Phase 2. Phase 3 allows the user to further focus on those SOMA rules that have certain antecedents. Figure 3 shows what happens when we continue looking at the PKK using the action “Arms trafficking” and we select the condition “Does not advocate wealth redistribution”. This selection basically says: Find all rules about the PKK that SEE has derived which deal with the action *arms trafficking* where one of the antece-

dents of the rule is *does not advocate wealth redistribution*. Once this query has been formulated and submitted, the system computes a web page about the PKK.

Step 3. Select up to 3 Conditions

Condition 1: Does not advocate wealth redistribution \* (2125)

AND

Condition 2:

AND

Condition 3:

Figure 2 The STOP user has selected "Does not advocate wealth redistribution as the antecedent in their rules about Arms trafficking.

Figure 4 shows the result of the above inquiry by an analyst. The panel on Figure 4 is divided up into multiple parts.

**Recent News.** The “Recent News” panel on the result window shows recent news about the PKK. Recent news is obtained by accessing online news wires to present the analyst with the most recent news on an organization, irrespective of whether it deals with the query or not.

**Extracted SOMA Rules.** The “Extracted SOMA Rules” panel shows all the SOMA rules extracted for the analyst’s inquiry. Our default is to order these in descending order of probability (though other options are available to be discussed below). For instance, the first rule shown in Figure 4 tells us that there is an 88.23% probability that the PKK engages in the *arms trafficking* action when:

- They do not advocate wealth redistribution and
- They receive foreign state support and
- There is no intra-organizational conflict.

The bottom right of the screen shows that there were a total of 82 SOMA rules we derived for this query and that the first 25 are shown on this screen. The user can browse all these 25 rules on this screen or move to another screen by transitioning to another page (see bottom left of the panel for how to move from one page to another).

**Basic Information about the Group.** The bottom of the page also shows that STOP can access other sources such as Wikipedia in order to present the user some background information about the group. Sources other than Wikipedia are also planned for access in coming months

**Access to other Systems.** Other systems such as T-REX [10] can provide valuable information about activities that different groups around the world might be

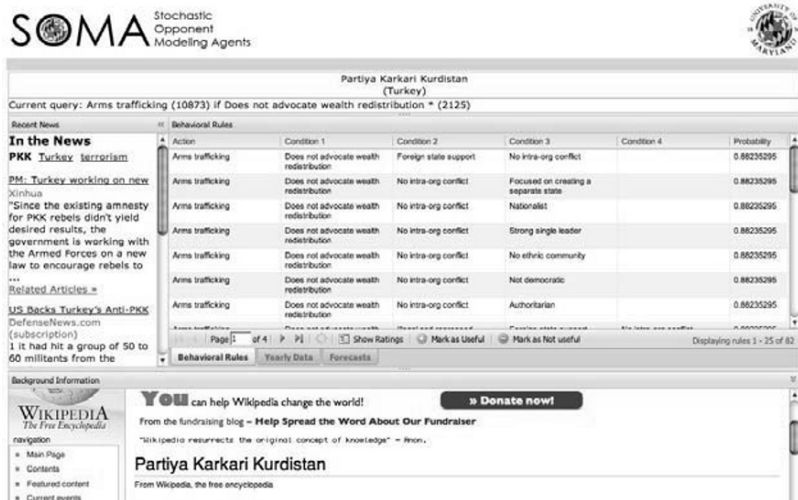


Figure 3. The results from a query in the STOP system

engaging in. For example, the T-REX system can automatically extract which violent activities groups are involved in. T-REX processes around 150K pages each day from approximately 145 news sources worldwide.

In addition, the CONVEX system [12] under development at the University of Maryland provides an *alternative to SOMA* for forecasting. In a future version of STOP, we anticipate providing access to CONVEX through the SOMA Terror Organization Portal. We also hope to provide access to a wide variety of third party systems as well, but this is not the case at this point.

SANE contains facilities for social networking amongst national security analysts and researchers working on counter-terrorism efforts worldwide. The kind of social networking supported by SANE is shown at the bottom on Figure 4.

- **Mark as Useful.** Analysts can mark certain rules as “Useful”. When an analyst marks certain rules as useful, a tuple gets stored in our SOMA Utility Database underlying STOP. This tuple identifies the group, the id of the rule, and the userid of the user who found the rule to be useful. In addition, the

analyst may enter a comment about why he or she found the rule useful. This too is stored. Additional metadata about the transaction is also stored.

- **Mark as Not Useful.** Just as in the preceding case, analysts can mark rules as “Not Useful” as well.
- **Show Ratings.** **SANE** typically shows rules to an analyst in descending order of probability. However the design of **SANE** also allows a “Show Ratings” capability where the analyst can see ratings that other analysts have given to rules. We expect, in the very near future, to allow users to query the SOMA Utility Database so that he can see which rules have been marked as interesting or not interesting by colleagues’ whose opinion he respects.

The “Yearly Data” tab towards the bottom left of the screen shows yearly data (i.e. the actual Minorities at Risk Organizational Behavior codes for a given year). The **SANE** design allows an analyst to pull these up, mark them as useful or not, and also insert comments on the codes. These features, which are in the process of being incorporated into the system, will, upon completion, allow further social networking activities along the lines mentioned above.

## 4 SOMA Analyst Forecast Engine (SAFE)

The “Forecast” tab towards the bottom left of Figure 4 allows an analyst to invoke the SOMA Analyst Forecast Engine (SAFE). The basic idea in SAFE is to allow an analyst to hypothesize a “state”  $S$  of the world and to ask the system what *set* of actions the group  $G$  he is studying will engage in.

SAFE uses the technology described in [2,3] in order to achieve this. Based on the state  $S$  that the analyst specifies, SAFE focuses on those rules SEE has extracted about group  $G$  that are applicable w.r.t. situation  $S$ . In other words, SAFE focuses on the set  $\{ r \mid r \text{ is a rule SEE has extracted about group } G \text{ and } S \text{ satisfies the antecedent of rule } r \}$ . This set of rules is called the set of *applicable rules for } G \text{ w.r.t. state } S*, denoted  $App(G, S)$ .

SAFE then sets up a set  $LC(G, S)$  of linear constraints associated with  $G$  and  $S$  as described in [2,3]. Each set of actions that group  $G$  can engage in is termed a “world” (or a possible world). The variables in  $LC(G, S)$  correspond to the (as yet unknown) probabilities of these worlds. By solving one linear program for each world using the algorithms in [2,3], SAFE is able to find the  $k$  most probable worlds.

Thus, an analyst can use SAFE in order to hypothesize a state  $S$  and to see what the  $k$  most probable worlds are. The algorithms underlying SAFE have been fully implemented – at the time of writing of this article, they are being “plugged into” SAFE to support system usage.



We are currently inserting Social Networking features into the SAFE component. When an analyst chooses to do so, he can save a state he has hypothesized, save a prediction he has hypothesized, choose whether he agrees with the forecast or not, and comment on the prediction as well. All these are saved in a *Forecast and Comment* database. Other analysts who are authorized to do so may comment on those same predictions as well. The FAC database will provide a query facility to other users who can see which forecasts when authorized to do so.

## 5 Conclusions

In this paper, we have described the SOMA Terror Organization Portal (STOP) – a facility that national security analysts can use in order to understand terror threats worldwide. STOP provides a single, Internet or Intranet accessible (password protected) site within which national security analysts can study certain groups. Not only does STOP provide tools they might use, it also provides a valuable social networking capability that allows analysts to often create and expand a network of experts on a given topic. It hardly needs to be said that STOP allows them to leverage this network so that different points of view can be incorporated into their analytic task before a final recommendation is made.

To date, STOP allows analysts the ability to examine approximately 23 terror groups from about ten countries ranging from Morocco all the way to Afghanistan. Users can see rules extracted automatically about these groups (over 14,000 for Hezbollah; over 20,000 for the PKK), browse them, experiment with them, and mark them as useful or not. They can use these markings to build consensus (or at least identify different camps) about a given topic, and explore the pros and cons of alternative views – all without having to move from their desk.

As the powerful real-time capabilities of the T-REX Information Extractor become available to STOP, the need for manually coded information currently used by T-REX will disappear. We expect this transition to start in mid-2008.

**Acknowledgements** The authors gratefully acknowledge funding support for this work provided by the Air Force Office of Scientific Research through the Laboratory for Computational Cultural Dynamics (LCCD) grants AFOSR grants FA95500610405 and FA95500510298, the Army Research Office under grant DAAD190310202, and the National Science Foundation under grant 0540216. Any opinions, findings or recommendations in this document are those of the authors and do not necessarily reflect the views of sponsors.

## References

1. Gerardo I. Simari, Amy Sliva, Dana S. Nau, V. S. Subrahmanian: A stochastic language for modeling opponent agents. Proc. AAMAS 2006, pages 244-246.
2. S. Khuller, V. Martinez, D. Nau, G. Simari, A. Sliva, V.S. Subrahmanian. Action Probabilistic Logic Programs, accepted for publication in *Annals of Mathematics and Artificial Intelligence*, 2007.
3. S. Khuller, V. Martinez, D. Nau, G. Simari, A. Sliva, V.S. Subrahmanian. Finding Most Probable Worlds of Logic Programs, Proc. 2007 Intl. Conf. on Scalable Uncertainty Management, Springer Lecture Notes in Computer Science Vol. 3442, pages 45-59, 2007.
4. A. Sliva, V. Martinez, G. Simari and V.S. Subrahmanian. SOMA Models of the Behaviors of Stakeholders in the Afghan Drug Economy: A Preliminary Report, Proc. 2007 Intl. Conf. on Computational Cultural Dynamics, pages 78-86, AAAI Press.
5. V.S. Subrahmanian. Cultural Modeling in Real-Time, *Science*, Vol. 317, Nr. 5844, pages 1509-1510.
6. A. Mannes, M. Michael, A. Pate, A. Sliva, V.S. Subrahmanian and J. Wilkenfeld. Stochastic Opponent Modeling Agents: A Case Study with Hezbollah, Proc. Intl. Conf. on Social and Behavioral Processing, Sedona, Arizona, April 2008, Springer Verlag Lectures Notes in Computer Science, to appear.
7. V. Asal, C. Johnson, J. Wilkenfeld. "Ethnopolitical Violence and Terrorism in the Middle East" in J.J. Hewitt, J. Wilkenfeld, T.R. Gurr (eds) *Peace and Conflict 2008*. Boulder, CO: Paradigm Publishers. 2007.
8. Minorities at Risk Project. College Park, MD: Center for International Development and Conflict Management. 2005 Retrieved from <http://www.cidcm.umd.edu/mar/>.
9. V.S. Subrahmanian, M. Albanese, V. Martinez, D. Nau, D. Reforgiato, G. I. Simari, A. Sliva and J. Wilkenfeld. CARA: A Cultural Reasoning Architecture, IEEE Intelligent Systems, Vol. 22, Nr. 2, pages 12-16, March/April 2007.
10. M. Albanese and V.S. Subrahmanian. T-REX: A System for Automated Cultural Information Extraction, Proc. 2007 Intl. Conf. on Computational Cultural Dynamics, pages 2-8, AAAI Press, Aug. 2007.
11. C. Cesarano, B. Dorr, A. Picariello, D. Reforgiato, A. Sagoff and V.S. Subrahmanian. Opinion Analysis in Document Databases, Proc. AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs, Stanford, CA, March 2006.
12. V. Martinez, G. I. Simari, A. Sliva and V.S. Subrahmanian. CONVEX: Context Vectors as a Paradigm for Learning Group Behaviors based on Similarity, in preparation.



<http://www.springer.com/978-0-387-77671-2>

Social Computing, Behavioral Modeling, and Prediction

Liu, H.; Salerno, J.; Young, M.J. (Eds.)

2008, XVIII, 264 p., Hardcover

ISBN: 978-0-387-77671-2