CARA: A CULTURAL ADVERSARIAL REASONING ARCHITECTURE

V. S. Subrahmanian, M. Albanese, V. Martinez, D. Nau, D. Recupero, G. Simari, A. Sliva, O. Udrea, J. Wilkenfeld

University of Maryland
College Park, MD 20742.

Abstract. There is growing need for automated tools to reason about groups in diverse countries. For example, the CDC may wish to understand social patterns that cause diarrheal diseases to spread in Kenya, the World Bank may want to understand how various socio-economic-religious factors may affect the actions of tribes in the Pakistan-Afghanistan border region, while domestic and international security agencies may want to track Kurdish organizations that share characteristics with other organizations that have resorted to violence in the past . In this paper, we describe the CARA architecture for gathering data about different cultural groups, learning the intensity of opinions that those groups have on various topics, and developing a process that supports building/extracting models of behavior of those groups and continuously refining those models through shared, multi-person, learning experiences. The CARA architecture is supported via ongoing applications we have been developing – one in global health care, another in tracking tribes along the Pakistan-Afghanistan border with a view to understanding and eventually reducing the burgeoning drug trade there, and a third focusing on politically active minorities at risk in fragile regions of the world.

I. INTRODUCTION

There is a constant need to reason about diverse cultures all over the world. For example, a World Bank loan aimed at reducing the cultivation of opium along the Pakistan Afghanistan border would greatly benefit from a socio-economic-political-religious model of the behaviors of the tribes in opium producing regions. By building such models and applying them, the World Bank may get a better understanding about how best to target their loan dollars in order to better achieve their goals. Likewise, a health care organization anxious about the spread of diarrheal (or any other disease) in Kenya might wish to understand socio-economic-cultural-environmental aspects of Kenyan society that cause the diseases to spread extensively in some parts of the country and not in others. In almost all cases, the spread of diseases is not due to biological factors alone, but due to a rash of social behaviors, environmental factors, and economic and educational aspects of the disease-stricken community. Kurdish minorities are spread across vast regions of the Middle East – Turkey, Syria, Iraq, and Iran. Our models seeks to pinpoint those organization characteristics that differentiate among those organizations that engage in political action within legitimate frameworks, and those that cross the line and engage in violence and terror, either internally or externally.

In this paper, we introduce an architecture called CARA that we are developing for cultural reasoning about groups whose goals are not always aligned with our goals. CARA itself consists of four innovative components shown in Figure 1. Since CARA is an ongoing project, our current versions of these components have varying degrees of functional capability and integration.

- ***Semantic Web Extraction Engine.*** A set of methods to extract event information and background information about a given group from diverse, heterogeneous text sources and to store these in an annotated semantic web database [1,2]. For example, we may wish to extract background information about the number of occurrences of cholera in the Kwale district of Kenya in 2000.

- ***Opinion Analysis System.*** A set of methods to extract the intensity of opinion that a given group might have on a given topic [3] in order to gauge what the group really cares about. For example, we might want to identify the intensity of opinion in Pakistan about the Guantanamo Bay Koran abuse scandal in order to best understand who is upset, how upset they are, and how best to improve the image of the US in Pakistan.
- ***Stochastic Opponent Modeling Agents.*** A formal, logical-statistical reasoning language within which we can express knowledge about the behaviors of the group of interest, and compile a set of rules in such a language into an "agent". For example, we might want to learn how the "sharecropper" economic class in the Afghan drug economy behaves. How is their salary affected by a switch to an alternative (legal) crop? How is their physical security affected when they switch from poppy cultivation to another crop?
- ***Game Engine.*** A set of algorithms that provide a game environment in which humans can interact with CARA. Our game engine design allows users of a game to learn aspects of the culture on which a game is based. For example, the World Bank official going to Afghanistan must first learn the protocol involved in entering a village rather than just breeze in his Mercedes or Humvee. Moreover, if an expert on Afghanistan were to play such a game, he might be able to provide valuable feedback on the knowledge used in the game.

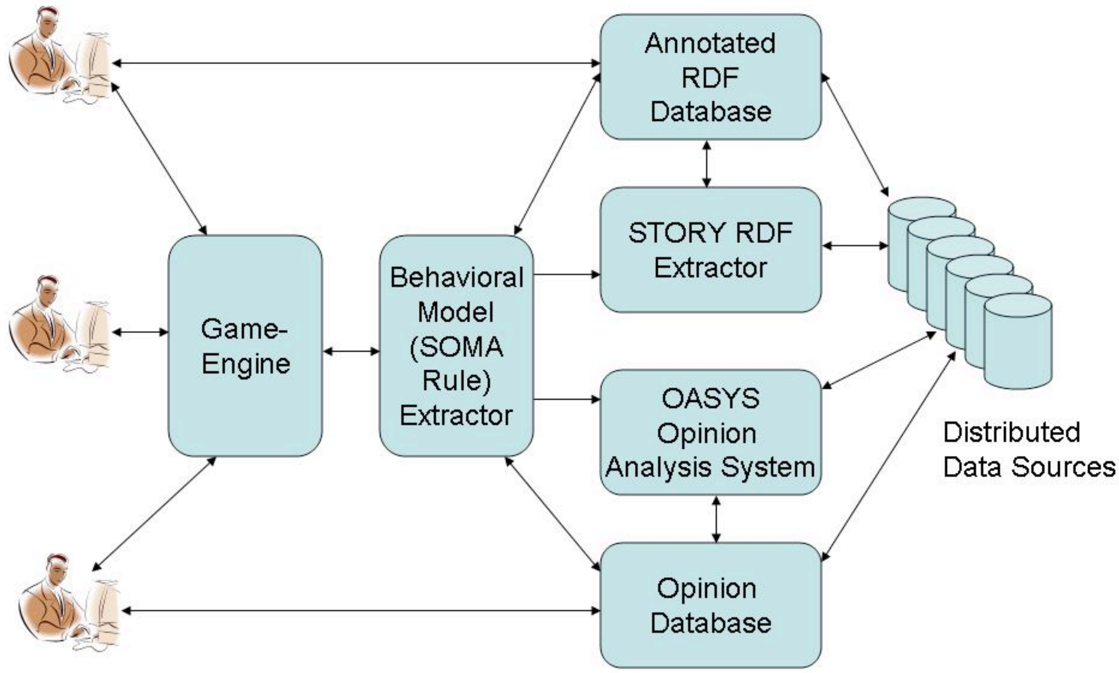

Figure 1. Architecture of the CARA System

We now discuss each of these four components in greater detail below.

II. The Semantic Web Extraction Engine

Our semantic web extraction engine – called STORY [4] – allows us to extract facts about any given person, group, or event from the web. The facts are stored in an extension of the W3C's Resource Description Framework (RDF for short) called *Annotated RDF* [5] database.

Standard RDF largely represents data in the form of *(subject, property, object)* triples. For example, we might extract a fact which says that there were 185 cases of diarrhea in 2000 in clinic number 603. This is naturally represented by the triple (clinic_603, diarrhea_cases, 185). However, a fundamental flaw with this is that representing 4 inter-related items with a triple is impossible – in the above representation, the triple did not state the year "2000". In contrast, annotated RDF allows such triples to be extended with metadata such as the time when the triple is valid, the provenance of the triple (i.e. where it came from), and combinations thereof.

STORY extracts facts by learning rules from annotated sentences. Figure 2 shows STORY's annotation tool applied to the sentence "Over 2 million diarrhea deaths occur annually" taken from a WHO report about Kenya. A parse tree of this sentence can be annotated by a human being as shown in Figure 2. This causes STORY to learn a rule that is somewhat more general than what is expressed in sentence (1). For example, we can accurately extract RDF from *un-annotated sentences* such as *An estimated 2.2 million diarrheal deaths occur annually* and *approximately 585,000 maternal deaths occurred annually*.

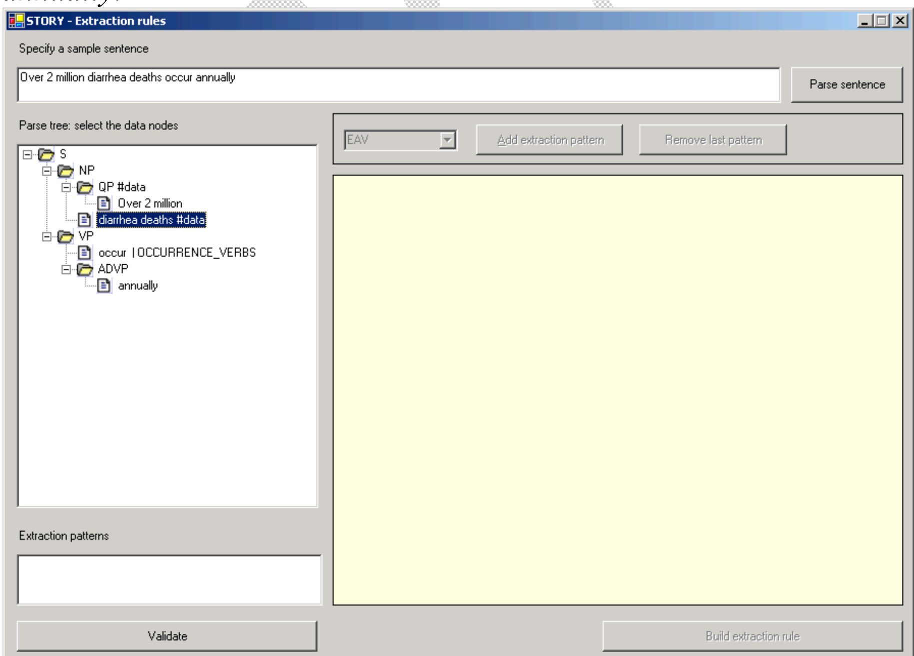

Figure 2. Screenshot of STORY rule learning engine.

The results of STORY's extraction are stored in the annotated RDF database system that we have built. Annotated RDF includes techniques to store RDF data annotated with time points, truth values or levels of certainty, and provenance data and answer atomic and conjunctive queries, as well as aggregate queries on the annotated RDF. Aggregate queries are particularly important because – for example – if we wish to find the total number of diarrhea cases in a given district, we may need to sum over the reports from each clinic in that district.

III. The Opinion Analysis System

Another major task of our system is to gauge the intensity of opinion that a given group or a given individual has on a given topic. For example, returning to our drug example, we may be interested in learning the strength of opinion that Pakistan's President, Gen. Musharraf holds on the drug problem. This, in turn, may give an organization such as the World Bank an early indicator on the best way to approach this problem with the Pakistani leader. When we consider this problem, there are two subproblems:

- First, we need to identify sentences where the Pakistani leader expresses an opinion about the drug industry.
- Second, we need to identify the intensity of opinion expressed in such statements.

Problems of the first kind have been well addressed in the natural language processing community – it is easy to find sentences (e.g. from a web page, a news article, or transcripts of Musharraf's statements) where Musharraf speaks about illegal drugs. Suppose $Rel(d,t)$ is the set of all such "relevant" statements about a given topic t in a given document d.

For any given sentence in $Rel(d,t)$, our OASYS system first finds all adjectives that apply to the topic in question. Thus, in the sentence *"The drug trade is a terrible disaster for the people" said Gen. Musharraf*, the adjective "terrible" applies to the topic "drug trade". It then checks whether such adjectives occur positively or negatively – in the above example, it appears positively. Based on a set of studies we previously conducted [3], every adjective has a given score on a -1 to +1 scale. The more negative an adjective's score, the more negative the adjective; the more positive its score, the more positive the adjective. The score for "terrible" is -0.6. The score $Sc(d,t,s)$ of a sentence s on topic t in document d is now the sum of the applicable adjective scores. If an adjective appears negatively in a sentence as in *The drug trade is not a terrible disaster*, then that negative occurrence gets a score equal to the negation of the adjective score. The score of a document d on topic t is computed as a complex function that merges together the scores of the individual sentences. In this way, we can now assign a score to the level of intensity about a given topic that a document attributes to a given person.

Figure 3 shows us an OASYS screenshot of opinion of Musharraf during the July 5-Sep. 5, 2006 period, as reported in the Pakistani press. This figure shows Musharraf dropping in opinion in August 2006. A political or news analyst might correlate this drop in press coverage with an event such as the unsuccessful US bombing of a Pakistani village while hunting for Al-Qaeda leaders.

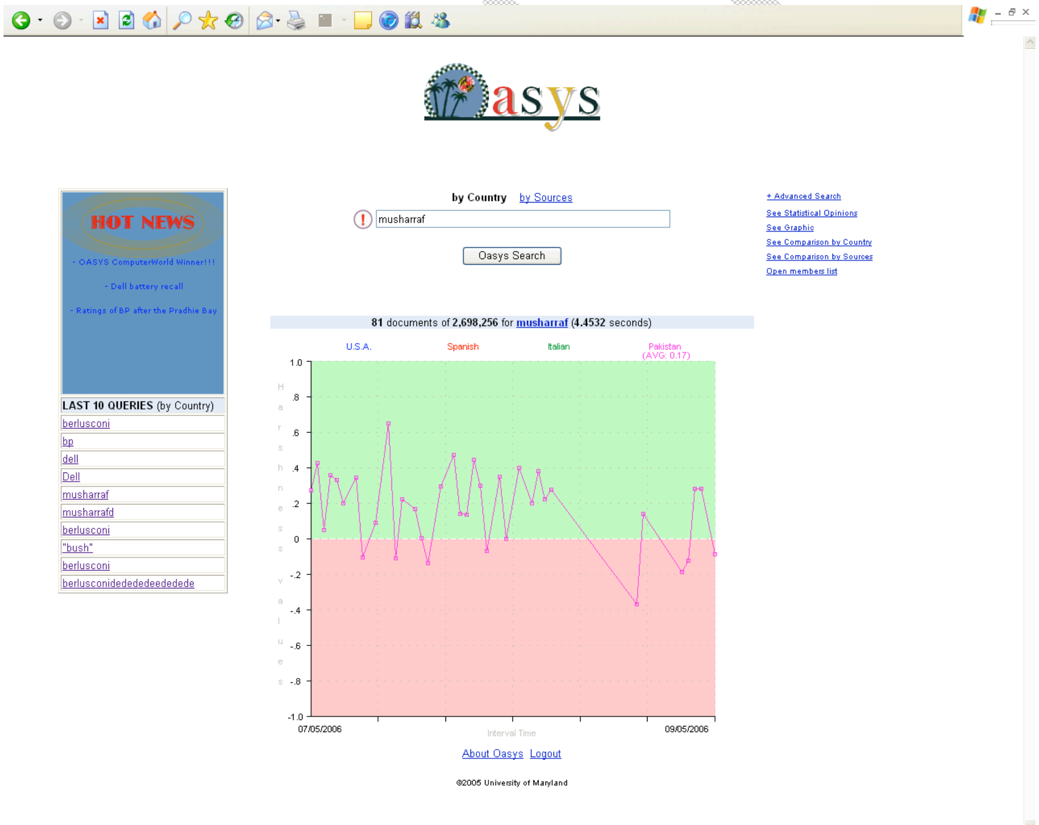

Figure 3. Screenshot of Pakistani press opinions of Pres. Musharraf during a two month period from July 5-Sep. 5, 2006.

IV. Stochastic Opponent Modeling Agents

The CARA architecture also contains the use of a *stochastic opponent modeling agent (SOMA) language*. Within this language, we can express rules about the probability that a given person or group will act a certain way in a given situation. For example, suppose we consider the "sharecroppers" in the Pakistan-Afghan drug scenario. In SOMA, we can write rules of the form "If a farmer has level of debt exceeding *d*, then he will cultivate poppies with over 70% probability". This can be written in SOMA as a rule of the form

Will_cultivate_poppies(F):[0.7,1] - ***if*** *debt-level(F,D)& D > d.*

SOMA also has mechanisms where the consequents of such rules can be more complex. For example, we may say that "If a farmer has level of debt exceeding *d*, then the joint probability that he will cultivate poppies and be hostile to foreign visitors is over 60%". Such a rule can be encoded as:

(hostile_foreigners(F)- and Will_cultivate_poppies(F)):[0.6,1]– ***if*** *debt-level(F,D)& D > d.*

Most interesting of all, given a database of facts about the world and a set of such rules encoding the behavior of either a person or a group, SOMA has developed algorithms to find the most probable set of actions that that person may or may not take. To do this, it uses a possible-worlds model in which a world represents a set of actions that the opponent can take. Figure 4 shows that even when we have to consider over half a million worlds, SOMA can find a quick answer within a few minutes.

Fig. 4 Graph showing performance of two algorithms to compute the most probable world associated with SOMA rules.

SOMA-rules can be readily learned from past data as follows. For example, suppose we have a single relational database ***D*** of debt levels of farmers and which plants they cultivate. Any selection condition *C* on debt levels will split the database into two parts: D^+ is the part that satisfies *C*, while D^- is the part that doesn't. Suppose $\boldsymbol{D}_{\text{poppy}}$ is the part of ***D*** that contains tuples where the plant being cultivated is poppy. In this case, we want to find a *C* that maximizes some linear combination of both recall (i.e. the ratio of number of tuples – or rows - in the intersection of $\boldsymbol{D}_{\text{poppy}}$ and $\boldsymbol{D}^+$ to the number of tuples in $\boldsymbol{D}_{\text{poppy}}$) and precision (i.e. the ratio of number of tuples in the intersection of $\boldsymbol{D}_{\text{poppy}}$ and $\boldsymbol{D}^+$ to the number of tuples in $\boldsymbol{D}^+$). This can be readily done with an appropriate binary-search like strategy on the *debt-level* field.

We have developed the SOMA language and developed experimental code to assess the performance of the algorithms used by SOMA – a prototype interpreter is under development.

V. Game Engine

The final component of the CARA architecture is a game engine - we have a preliminary design of the game engine in place and an application built based on the design principles, but are still building the game engine.

For example, suppose we want to build a game telling a World Bank official what cultural norms are expected when entering an Afghan village. The same game engine, however, should be usable when building an application about how to enter a Kenyan or Kurdish village, even though these societies have dramatically different customs.

In both cases, the game consists of a set of game states. Each game state offers the player an array of information and a set of possible moves he can make. In our Afghan village game, for example, the game state consists of a multimedia (audio, video, image) presentation. The objects to be shown in a given game state (e.g. audio clip, video clip, text bar, etc.) can be obtained by querying a database and selecting a presentation method for the answer returned to the query. For example, if the answer returned is a video clip, the presentation method might be 30 frames a second, while if the answer returned is a body of text, the presentation method might be to have scrolling text at the rate of 10 lines per minute. Each game state might also pose a question to the player with a set of answers to the question. Independently of which answer is correct, the game proceeds to a new state (similarly defined) that shows the player another body of information – perhaps real, perhaps hypothetical – and invites him to do something else. This continues till the end of the game when the player can be told what he did right and what he did wrong. Figure 5 shows a sample screen shot from our entering an Afghan village game.

Figure 5. Screenshot of a CARA game.

VI. Conclusions

There is growing need for reasoning about how diverse cultural, political, industrial and other organizations make decisions. Past work on adversarial reasoning in AI has focused primarily on games such as Chess, Bridge, Clue, and Poker, where the rules of the game are well articulated and where the state is well defined. Reasoning about real world adversaries can be viewed as a complex game-tree search problem, but it is difficult to know the "rules" of the game, the state space is huge, and the state is often largely unknown. Often, even the variables constituting the state are unclear, let alone their values. Past adversarial reasoning work in AI can be a great asset even in real-world cultural reasoning, but a major problem is to identify the adversary's objectives and payoffs and the rules that the adversary adheres to, and determine how best to "play" the adversary given our knowledge of his behavioral rules. Accomplishing this is one of the goals of our ongoing work [5].

The CARA architecture provides a paradigm to articulate the components that go into the behavioral model of the adversary, even if it is at a reasonably coarse level. Moreover, CARA provides methods to understand what the adversary might be thinking about through the opinion analysis component, and methods to extract knowledge about the socio-economic-political aspects of the adversary's context. It provides methods to infer behavioral models of the opponent and to determine what best the opponent might do. From there, classical adversarial reasoning methods might be used to reason about how we should act in order to elicit the response we desire from the adversary. Nonetheless, CARA is just the tip of the iceberg – much work remains to be done on approximating how adversaries in the real world will behave.

Acknowledgements. This work is funded in part by AFOSR grant FA95500510298.

REFERENCES

1. T. Berners-Lee, J. Hendler and O. Lassila, The Semantic Web:When the Internet gets smart, *Scientific American,* May 2001.
2. O. Udrea, D. Reforgiato and V.S. Subrahmanian. Annotated RDF. *Proc. 2006 European Semantic Web Conference*, Budva, Montenegro, June 2006, Springer Lecture Noties in Computer Science, Vol. 4011, pages 487—501.
3. C. Cesarano, B. Dorr, A. Picariello, D. Reforgiato, A. Sagoff, and V.S. Subrahmanian. The OASYS Opinion Analysis System, Proc. 2006 AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs, Stanford, CA, March 2006.
4. M. Albanese, C. Cesarano, M. Fayzullin, A. Picariello and V.S. Subrahmanian. "Story Creation from Heterogeneous Data Sources", accepted for publication in *Multimedia Tools and Applications* Journal, May 2005.
5. T.-C. Au and D. Nau. Accident or intention: That is the question (in the iterated prisoner's dilemma). *International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2006.

About the Authors:

V.S. Subrahmanian is Professor of Computer Science and Director of the University of Maryland Institute for Advanced Computer Studies. He has worked extensively in both databases and AI. His focus in on diverse types of reasoning about heterogeneous, real-world information sources including relational, object oriented, flat file, textual, image and audio, news and blog sources and developed several systems, some of which have won major awards. He is interested in the applications of these techniques to help reduce conflict and war. He has published over 130 papers and has won numerous awards.

Massimiliano Albanese received the Laurea degree in Computer Science and Engineering (2002) and the Ph.D. in Computer Science and Engineering (2005), both from the University of Napoli, Italy. In August 2006 he joined the University of Maryland Institute for Advanced Computer Studies (UMIACS) as a postdoctoral research assistant. His main research interests lie in the area of Knowledge Extraction and Management, Image and Video Databases.

María Vanina Martinez received the Licenciatura degree in Computer Science (2005) from the Universidad Nacional del Sur university, Argentina. She is now a Ph.D. student at the University of Maryland, College Park, under the supervision of Dr. VS Subrahmanian. Her main research interests lie in the area of Intelligent Agents, Interaction Protocols for Multi-agent sytems, Agent Communication Languages, Reasoning Under Uncertainty, and Opponent Modelling.

Dana Nau is a Professor of both Computer Science and Systems Research, and a Fellow of the American Association for Artificial Intelligence (AAAI). He has about 300 refereed technical publications on automated planning, game-tree search, and other topics. He has received an NSF Presidential Young Investigator award, several best-paper awards, and several prizes for the performance of software systems.

Diego Reforgiato received the bachelor degree in Computer Science in 2001 from University of Catania and the Ph.D. in Computer Science in 2004 from the University of Napoli Federico II. In April 2005 he joined the University of Maryland Institute for Advanced Computer Studies (UMIACS) as a postdoctoral research assistant. His main research interests lie in the area of Opinion Analysis, Knowledge Extraction and Management, Clustering and Graph Matching.

Gerardo I. Simari received the Licenciatura in Computer Science degree (2002) and the Master in Computer Science degree (2004), both from the Universidad Nacional del Sur university, Argentina. He is currently a Ph.D. student at the University of Maryland College Park, under the supervision of Dr. VS Subrahmanian. Among his research interests are Intelligent Agents and Multi-agent Systems, Reasoning Under Uncertainty, Game Theory and Decision Theory, and Opponent Modelling.

Amy Sliva is a Ph.D. student in Computer Science at the University of Maryland. She received a Bachelor of Science in Computer Science from Georgetown University in 2005. She is currently involved in artificial intelligence research in cultural modeling and adversarial agents, and is also interested in data mining and automated inference of agent models.

Octavian Udrea is a PhD student at the University of Maryland College Park. His interests include knowledge representation, probabilistic reasoning and activity analysis in video and temporal databases. He has received his B.S. degree in Computer Science from the School of Computer Science and Engineering, Politehnica University of Bucharest, Romania in 2003.

Jonathan Wilkenfeld is Professor in the Department of Government and Politics and Director of the Center for International Development and Conflict Management at the University of Maryland. He is also Co-Principal Investigator for the Center for the Study of Terrorism and Responses to Terrorism. Wilkenfeld is the author of eight books and almost 100 articles and chapters, most focusing on international conflict and crisis, foreign policy decision making, experimental and simulation techniques in political science and, most recently, the role of third-party mediation in international disputes.