The Role of Blackboard-based Reasoning and Visual Analytics in RESIN's Predictive Analysis

Dingxiang Liu[1], Jia Yue[1], Xiaoyu Wang[2], Anita Raja[1], William Ribarsky[2]
[1]Department of Software and Information Systems
[2]Department of Computer Science
The University of North Carolina at Charlotte
{dliu, jyue, xwang25, anraja, ribarsky }@uncc.edu

Abstract

Knowledge gathering and investigative tasks in open environments can be very complex because the problem-solving context is constantly evolving, and the data may be incomplete, unreliable and/or conflicting. This paper significantly extends our previous work on a mixed-initiative agent by making it capable of assisting humans in foraging task analysis using AI blackboard-based reasoning, visualizations and a mix-initiative user interface. The agent is equipped with the ability to adapt its processing to available resources, deadlines and its current problem-solving context.

1. Introduction

Due to the significant increase in collected data and the increased complexity of the reasoning process itself, performing investigative analytical tasks has become more challenging. These tasks typically involve identifying and tracking multiple hypotheses; gathering evidence to validate the correct hypotheses and eliminating the incorrect ones. They also require the assistance from interactive visualizations, which enable analysts to explore and pre-process large amounts of data. More importantly, the analysis tasks are often time critical and need to adopt appropriate approaches, which vary from straightforward methods to comprehensive investigations.

One critical task is to predict missing or unknown information about current events based on trends from the past. The prediction process could be influenced by the varying viewpoints of stakeholders and internal biases of the news stories and sources of data used for the analysis, which lead to high levels of uncertainty in the analysis domain.

To facilitate the task-solving process, we are interested in building an automated reasoning agent, RESIN, which will determine predictions about a single event based on information from a single viewpoint. RESIN stands for, a RESource bounded INformation gathering agent for visual analytics. RESIN extends our previous work on TIBOR [1]. It emphasizes the blackboard reasoning and mixed-initiative reasoning aspects of our agent architecture that will assist investigative analysts in performing viewpoint-based predictive analysis. RESIN leverages sequential decision making [2] and an AI blackboard system [3] to support hypothesis tracking and validation in a highly uncertain environment. Providing clear explanation in support of the decision making process is critical to gain and maintain the analyst's trust in the system. We use an AI blackboard to achieve this goal, which maintains a clear evidential path for supporting and contradicting information while allowing for explicit modeling of concurrent top-down and bottom-up processing. RESIN has the capability to pass information between analysts and itself during the problem-solving process by leveraging an interactive visual analytics tool. Moreover, RESIN provides ways for the user to interact with its problem-solving process or even control it at every step through a rich user interface. By using RESIN, investigative analysts can have the access to automated support for their decision making, the capability for finding non-myopic alternate solution paths and a tool to investigate outliers.

RESIN consists of an AI Blackboard [3], a TÆMS [4] task structure library, a Markov Decision Process (MDP) [2] solver and heterogeneous knowledge sources (KSs), as shown in Figure 1. The AI Blackboard contains reasoning results from processing existing information, which includes raw data, various problem-solving states, partial solutions and current goals; the TÆMS is an abstraction of the low-level execution model and captures uncertainty in outcome distributions, while the MDP is a probabilistic model, which captures the essence of sequential processes and is used to compute policies that identify, track, and plan to resolve confidence values associated with blackboard objects. The KSs [3] are independent specialist computational modules that contain the domain knowledge needed to solve a problem. The steps in Figure 1 illustrate the control flow of this predictive analysis process, which involves handling several issues: choosing the appropriate set of databases, analyzing the high dimensional data; generating the type of decision tree to extract and represent the data; determining appropriate interactive visualizations for these data; performing reasoning processes; and generating final solutions.

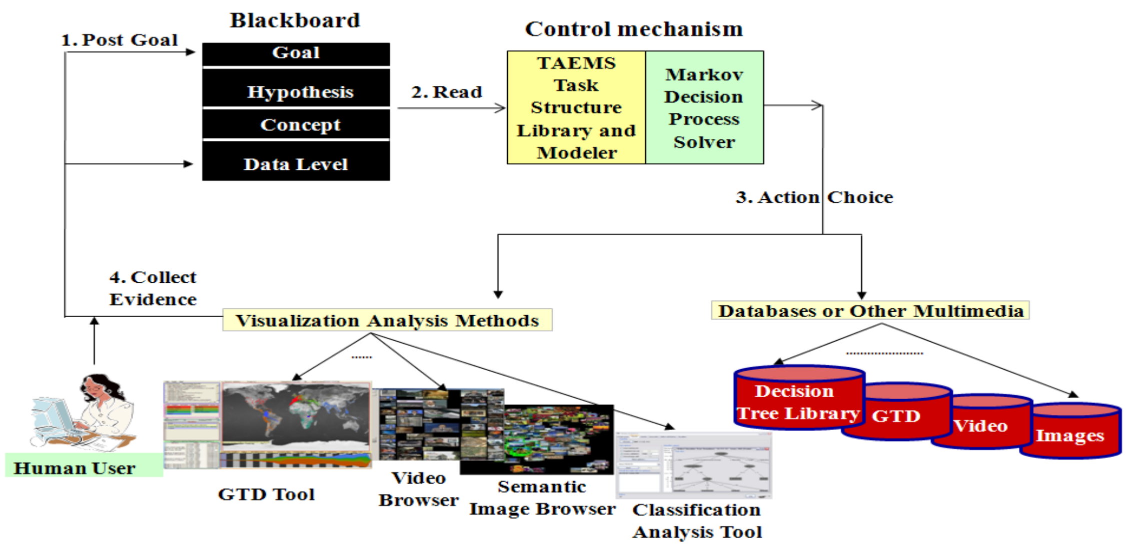

Figure 1. RESIN's Control Flow

In this paper, we apply RESIN to the Global Terrorism Database (GTD) [5] to perform predictive analysis tasks. Using machine-learning classification techniques (as discussed in section 2), blackboard-based reasoning and the GTD Visualization Tool [6], we are able to make predictions based on existing historical data.

2. RESIN's Architecture for Predictive Analysis

In this section, we provide a description of the prediction process that we use in order to determine which terrorist group is likely to be responsible for a particular incident. Our overall approach to this problem is to determine which KSs to be triggered based on the resource and deadline constrains. The KSs will try to match the input event to past events in order to predict the group name.

KSs, AI blackboard and control mechanism are three main components of RESIN's architecture (Figure 1). RESIN employs a set of KSs, including the C4.5 [7] algorithm, the GTD, and an investigative visual analytics system built on the GTD. At appropriate times defined by RESIN's reasoning process, the knowledge source takes relevant information from the blackboard and makes a contribution towards solving problem with its specialized knowledge.

C4.5 is a classical machine learning algorithm introduced by Quinlan for inducing classification models from data. RESIN is integrated with the knowledge source of C4.5 based on Weka [10] in order to automatically access and preprocess global terrorism data.

The GTD [5], developed by a team of researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), contains terrorist activities occurred all over the world between 1970 and 1997. The GTD tool [6] is a visual analytics approach that will visually provide investigators knowledge about terrorist activities and their relationships with its four highly coordinated views (corresponding to Who, What, When, Where). Among all the views that this tool provides, there are two views that are significant in helping RESIN's reasoning process: *MapView* and *TemporalView*. *MapView* provides straightforward geospatial information to depict terrorists' incidents while *TemporalView* reveals their temporal trends and patterns, as well as the relative growth and decline among the patterns over time. The categorical information of input tuple (for example of Figure 2) can be mapped into the two views of the GTD tool with high interactivity, which plays an important role in the foraging analysis of our mixed-initiative agent. More details about how the KSs are used and how the confidence value is computed are presented in our technical report [9].

The AI blackboard [3] data structure is a global shared repository containing problems, elementary data, a set of partial solutions, contributed information, and other data, which is available to all KSs, functions as a multilevel database and serves as a communication medium. It is the kernel of the agent which contains four different levels: Goal, Hypothesis, Concept and Data, in order of decreasing granularity. The information at a given level is derived from the level(s) below it, and it in turn supports the hypothesis at higher levels.

The control mechanism makes runtime decisions about the problem-solving process, specifically for a given resource (e.g. time) constraints, which will determine the databases and tools that need to be accessed. We have modeled this control mechanism using TÆMS-based [4] uncertainty reasoning and a MDP-based [2] sequential decision making for RESIN. For the prediction problem, the TÆMS task structure contains the various choices of visualization tools, classification algorithm, database, and levels of user interaction relevant to the particular query. Then the TÆMS task structure is translated into a MDP solver by initializing a state set, identifying the possible actions to determine the optimal action choices, and expanding each possible outcome which is characterized by discrete quality, cost and duration values, as described in our previous paper [1].

The problem-solving process is initiated when the Human User posts a goal on RESIN's AI Blackboard and this action triggers the RESIN agent (Step 1). In this paper, the goal is defined as an input vector with partial information about a single terrorist incident (Figure 2). The task in Figure 2 has six categories: TYPE, ENTITY, REGION, YEAR, NKILL, and WEAPON as initial inputs. Each category has a different number of possible values, for example, TYPE (e.g. assassination, bombing, facility attack) contains different types of attacking methods, while ENTITY represents different attack targets, such as 'Political Party', 'US Police/Military' and so on. And the goal is to predict the group name within a given deadline based on this partial information. In Step 2, the TÆMS task structure modeler generates an appropriate task structure and translates it to the MDP solver for action assessment. Using dynamic programming, the MDP solver computes the optimal policy based on resources constraints (e.g. deadline) and generates the best action, which will trigger appropriate methods to perform predictive analysis (Step 3). Through a built-in user interface, the RESIN agent enables the user to interact with the visual analytics tools supporting the mixed-initiative problem solving process, to validate the initial RESIN results and to post results back to the AI blackboard (Step 4). These initial results could include several hypotheses that could be possible solutions to the problem. Using these visualization results as well as previous analysis results, the blackboard will then propagate the evidence information and verify a specific hypothesis with an associated confidence value.

TYPE: Assassination
WEAPON: Explosives
ENTITY: Political Party
YEAR: 1992
REGION: Middle East/North Africa
NKILL: 2
GNAME: ?

Figure 2. Partial Terrorist Incident Description

We show that using the control flow described above, RESIN can largely enhance the accuracy of results in solving prediction problems; with the integration of visual analytics tools, RESIN will provide the capability for the user to manually perform tasks or even override the agent's suggestions, interactively.

3. Experiments

In this section, we describe initial experiments to assess the effectiveness of RESIN's blackboard-based reasoning mechanism, with the goal to determine unknown group name (GNAME) based on key tuples, as shown in Figure 2.The TÆMS [4] task structure modeler generates a task structure that models problem-solving patterns based on the input tuples. The top-level task is *Predictive-Analysis*, which is decomposed into two subtasks, *Classification-Algorithm* and *Visualization-Analysis*. The *Classification-Algorithm* will determine the data classification algorithm C4.5 and *Visualization-Analysis* will trigger the appropriate data visualization tool. To justify the importance of user interaction in a mixed-initiative agent, the RESIN's task structure also provides user interaction options: *Map-View-Interaction-Option* and *Temporal-View-Interaction-Option*. The information from these two views will equip the human user with clues to estimate their confidence values and to adjust the predicted results. Each of these methods is characterized by quality and duration distribution.

This experiment is based on a training set of 2700 incidents selected from the GTD and for each task we use the same ten incidents from a test set, with different deadlines from 30 to 70 (ranging from a very tight to a loose deadline). There are three users involved in the experiments with the access to GTD tool. User will determine the confidence values towards initial predictions with values from -0.9(strongly disagree and dispute the result) to 0.9(strongly agree and accept the result) through interactions with *MapView* and *TemporalView*.

We compare the predictive performance of the MDP policy and a Deterministic Schedule (DS) for task structures under different deadlines. DS is a deterministic process scheduler that builds a static schedule with the highest possible quality. The DS in this experiment is: {*HQ-Model-Option; Classification-Analysis; HQ-Map-View-Option; HQ-Map-View-Interaction-Option; HQ-Temporal-View-Option; HQ-Temporal-View-Interaction-Option*}.

Compared with a traditional DS, our MDP policy shows a significant improvement in assisting the users to predict the correct group name. Shown in both Figure 3 and Figure 4, we provide detailed comparisons on both cases with correct predictions and incorrect ones. Both charts clearly show the dynamic policy that the MDP provides allows users to get more correct probability results and fewer incorrect ones than if they use a DS. For instance, there was 10.27%, 6.4%, and 10.04%

improvement in performance for the deadline of 40, 50, and 60 with the correct prediction respectively. The t-test values (0.000976, 0.016675, and 0.002888) are less than 0.05, which shows that performance of MDP policy is statistically significantly different from the performance of DS. Therefore, RESIN agent is able to assist analysts to make better responses especially on task deadlines 40 to 60 with MDP policy.

Noticeably, there is not much difference for deadline 30 and 70 since they are highly constrained and loosely constrained problem. On a tight deadline (30), MDP solver cannot generate a policy better than the DS, while on a loose deadline (70), the DS gets enough time to complete all methods, just as the MDP policy could. Overall, the MDP policy outperforms the DS throughout our entire task set.

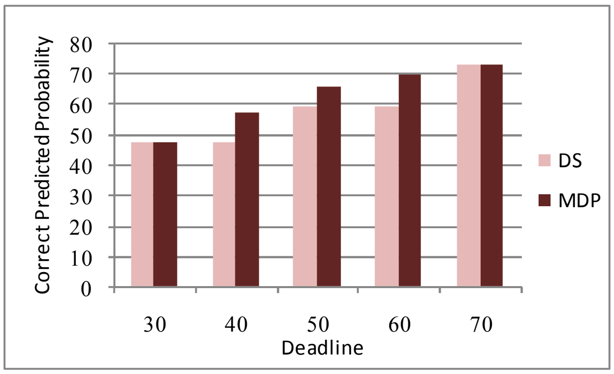

Figure 3. Comparison of correct predicted probability under different deadlines

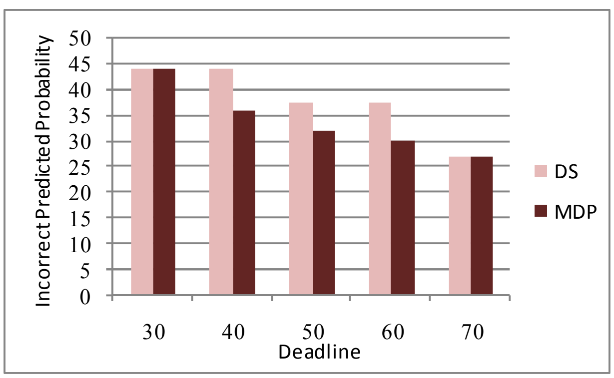

Figure 4. Comparison of incorrect predicted probability under different deadlines

4. Conclusion and Future Work

We have described a complex reasoning agent RESIN for predicting unknown or missing information in the GTD. We have identified abstract representations of the tasks to assist in the automated analysis as well as integrated the agent with the visualization tool, classification analysis tool and GTD as well as measurement for associated confidence value [9].

RESIN is a good start but there are still some interesting areas on which we want to work in the future. We plan to extend RESIN so that it will facilitate an analyst's problem-solving process by determining predictions about an event from multiple and conflicting viewpoints. Also, we attempt to employ current time-series analysis with forecasting technologies [8] to analyze past known events and predict future trends.

5. Acknowledgement

This work was sponsored by the National Visualization and Analytics Center (NVAC) under the auspices of the Southeastern Regional Visualization and Analytics Center. NVAC is a U.S. Department of Homeland Security Program led by Pacific Northwest National Laboratory. We also thank the anonymous reviewers for their helpful comments.

References

[1] Liu, D., Raja, A., and Vaidyanath, J., *TIBOR: A Resource-bounded Information Foraging Agent for Visual Analytics*, Proceedings of 2007 IEEE/ WIC/ ACM International Conference on Intelligent Agent Technology (IAT 2007), Silicon Valley, CA, November 2007, pp. 349-355.

[2] Bertesekas, D. and Tsitsiklis, J., *Neuro-Dynamic Programming*, Athena scientific, Belmont, MA, 2006.

[3] Corkill, D., *Blackboard Systems* AI Expert, 1991, 6(9):pp.40-47.

[4] Decker, K. and Lesser, V., *Quantitative modeling of complex environments*, International Journal of Intelligent Systems in Accounting, Finance, and Management, December 1993, 2(4): pp.215-234.

[5] LaFree, G. and Dugan, L., *Global Terrorism Database, 1970 – 1997* [Computer file]. ICPSR04586-v1. College Park, MD: University of Maryland [producer], 2006. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007-04-04.

[6] Wang, X., Miller, E., Smarick, K., Ribarsky, W., and Chang, R., *Investigative Visual Analysis of Global Terrorism,* Journal of Computer Graphics Forum, Eurovis ,2008.

[7] Quinlan, J. R., *C4.5: Programs for Machine learning*. Morgan Kaufman, 1993.

[8] Zeng, Z., Yan, H., and Alan M. N. Fu, *Time-series prediction based on pattern classification*. AI in Engineering, 2001, 15(1):pp.61-69.

[9] Liu, D., Yue, J., Wang, X., Raja, A., and Ribarsky, W., *The Role of Blackboard-based Reasoning and Visual Analytics in RESIN's Predictive Analysis*, Charlotte Visualization center Technical Report CVC-UNCC-08-29, July 2008.

[10] Witten, I. and Frank, E., *Data Mining: Practical Machine learning tools and techniques*, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.