Towards Context-Aware, Real Time and Autonomous Decision Making Using Information Aggregation and Network Analytics

Prithiviraj Dasgupta
Department of Computer Science
University of Nebraska at Omaha
Omaha, Nebraska, 68182-0500
Email: pdasgupta@unomaha.edu

Sanjukta Bhowmick
Department of Computer Science
University of Nebraska at Omaha
Omaha, Nebraska, 68182-0500
Email: sbhowmick@unomaha.edu

Abstract*—We consider the problem of real-time, proactive decision making for dynamic and time-critical decision-events where the choices made for multiple, individual decisions over time determine the final decision outcome of an event. We posit that the quality of such individual decisions can be significantly improved if human decision makers are provided with decision aids in the form of dynamically updated information and dependencies between the different decision variables, and the humans affecting those decision variables. In this position paper, we propose the CONRAD (CONtext aware Real-time Adaptive Decision making) system that uses computational techniques from large scale network analysis and game theory-based distributed information aggregation to develop such decision aids. CONRAD's functionalities are implemented through three subsystems - a decision making subsystem that updates and mathematically combines information from different decision variables to predict the outcome of the decision event, a decision assessment subsystem that uses the currently predicted decision outcome to estimate the future decision trajectory and recommends information collection-related actions to the human decision maker, and, a network analysis subsystem that uses those recommended actions to dynamically update the dependencies and correlations between events and people influencing the decision variables. To the best of our knowledge, our work is one of the first attempts towards combining dynamic decision updates and using the predicted decision trajectory as a proactive feedback mechanism to dynamically update the correlations between decision variables so that human decision makers can make more strategically-informed and well-aligned decisions towards the desired outcome of decision events.

I. Introduction

Modern decision making scenarios are characterized by large amounts of data and information that arrive dynamically, over a short period of time, from multiple sources. Processing this data in a time-critical manner to make accurate decisions is an overwhelming task for human decision makers. Over the past few decades several decision making solutions have been proposed to aid human decision makers with tools such as intelligent or automated software that use computational methods and mathematical models of human cognitive processes to make sophisticated decisions on behalf of humans[3], [12]. However, most existing decision support tools provide only limited context awareness of the decision process to the decision maker in rapidly evolving, information-rich and time-critical scenarios. This reduces the efficiency of human decision makers in making accurate decisions, and, consequently, could result in erroneous decision making in critical situations. Therefore, it makes sense to investigate techniques that could alleviate the human decision makers' context awareness by presenting information relevant to the decision making process, precisely and in a timely manner, to the decision maker.

To address this problem, we present the framework of a context-aware, real-time, decision making system called the CONRAD (CONtext aware Real-time Adaptive Decision making) system that focuses on enabling and enhancing the capabilities of human decision makers by developing proactive decision-aids for making high-accuracy, time-critical decisions in complex, data- and information-rich environments. The central research problem that CONRAD proposes to answer is as follows: *Given a set of decision variables in the current decision context along with a set of data sources from which the decision variables can be derived and/or calculated and updated, what is a suitable set of techniques for (i) extracting relevant information, (ii) then using that information to update, correlate and aggregate the decision variables dynamically, and finally , (iii) assessing the quality of the aggregated decision outcome (prediction), so that the divergence between the aggregated decision outcome (predictions) and the desired decision outcome is successively minimized?*

To address this question, we propose to represent a decision event as a collection of decision variables that are affected by the data from the environment. The system dynamically determines the inter-dependencies between these decision variables and also periodically updates them into an aggregated decision outcome (prediction). This aggregated decision outcome is then evaluated with respect to the desired decision outcome, and, depending on the deviation between the actual and desired decision outcomes, actions are recommended to collect additional data/information and discover new data correlations. This information is then used to update the decision variables autonomously and proactively - so that the quality of the aggregated decision outcome successively converges towards the desired outcome. We plan to realize the aforementioned functionalities in CONRAD using three subsytems that are summarized below:

(1) Decision Making Subsystem: The decision making subsystem uses a prediction market-based information aggregation mechanism to update and mathematically combine or aggre-

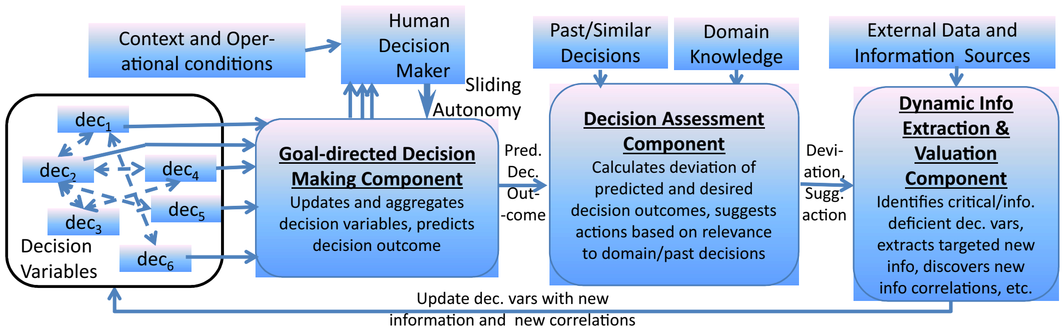

Fig. 1. Different components of the CONRAD system that integrate the decision making and network analytics aspects.

gate information from different decision variables and predict the outcome of the decision event.
(2) Decision Assessment Subsystem: The decision assessment subsystem uses the currently predicted decision outcome from the decision making subsystem along with relevant domain knowledge from past decisions made in similar domains to predict the decision trajectory and recommends information collection-related actions to the human decision maker. Machine learning and AI-based planning techniques are used to implement the functionalities of the decision assessment subsystem.
(3) Dynamic Information Extraction and Valuation Subsystem: The dynamic information extraction and valuation subsystem uses the actions recommended by the decision assessment subsystem to model and dynamically update the dependencies and correlations between events and people that influence the decision variables using metrics and techniques from large scale network analysis. The different components of CONRAD and their main functionalities are given in Figure 1 and discussed in the following sections.

II. Goal-Directed Decision Making

The main research question addressed in CONRAD's goal directed decision making subsystem is how to design a suitable set of computation techniques to dynamically update the different decision variables in the current decision context and combine or aggregate them into a single, global decision outcome. The decision variables are extracted from the environment's information by CONRAD's information extraction component, discussed in Section IV. We propose to perform the update and aggregation of the individual decision variables using an information aggregation technique inspired by prediction markets. Prediction market based information aggregation [14] has been recently shown to be a reasonably accurate means of predicting the outcome (usually binary or discrete valued outcome) of an event that is going to happen in the future. In our previous research, we have developed several successful techniques for multi-agent based prediction markets [8], [9] where the market's trading operations are performed by automated software agents. In prediction markets, information is collected from people, news sources, etc. in the form of bids, using either virtual or real money, on the possible outcome (binary-valued) of a future event. These bids are aggregated and the aggregated value represents the people's prediction of the event's outcome. A schematic of CONRAD's goal-directed decision making component is shown in Figure 2. To explain our approach, we use a few mathematical notions - let $\{dec_i^t\}$ denote the set of individual decision variables of the current decision making context at time step t, $AggDec^t$ denote the aggregated decision from aggregating $\{dec_i^t\}$ at time step t, where $dec_i^t, AggDec^t \in [0,1]$. With this formulation, each dec_i^t can be interpreted as a probabilistic confidence or belief of the decision variable; likewise for $AggDec^t$.

Dynamic decision variable updates. To enable dynamic updates of the decision variables, we associate each dec_i^t with a decision making (or belief update) agent a_i; a_i is responsible for updating the value of dec_i^t at time step t. Agent a_i performs this update using the following belief update formula:

$$dec_t^i = bel_i(dec_i^{t-1}, dec_{-i}^{t-1}, AggDec^{t-1}),$$

where $bel_i(.)$ is the belief update function used by agent i, dec_i^{t-1} is the value of dec_i during time step $t-1$, dec_{-i}^{t-1} is the set of decision variables from time step $t-1$, excluding dec_i^{t-1} itself, that are correlated with dec_i^t during time step t and $AggDec^{t-1}$ is the value of the aggregated decision outcome during time step $t-1$. The decision maker agent also ensures that decision variables that have already converged to their optimal or best value are not updated. The decision maker agent uses the intelligence, from reviewing the current context, to identify only those decision variables that need updating.

Aggregating decision variables. At the next step, the individual decision variables are combined into an aggregate or predicted decision outcome by the aggregator agent. A market-based aggregation mechanism provides a suitable way to combine information from multiple sources (e.g., multiple decision variables updated by the decision maker agents) into a single aggregated decision outcome value using a technique called a scoring rule [7].

III. Decision Assessment

The objective of CONRAD's decision assessment component is to determine how well the current aggregated (predicted) decision outcome is aligned with the desired decision outcome and to recommend actions related to future information collections that could potentially improve the convergence of the predicted decision outcome towards the decision outcome. A schematic of the decision assessment component is shown in Figure 3. Because we have represented decision outcomes as probability distributions (belief values), statistical divergence metrics such as the Kullback-Leibler (KL) divergence can be used to predict the future decision trajectory - Some

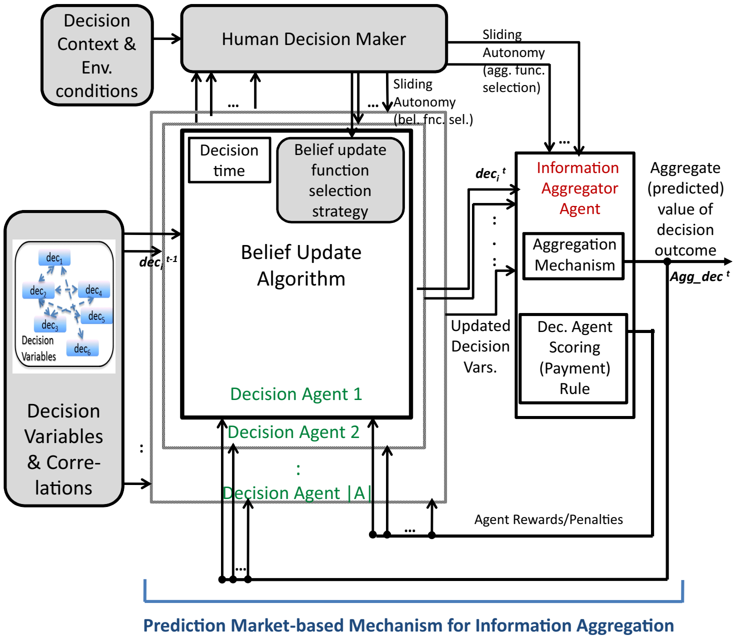

Fig. 2. The core of the prediction market based information aggregation technique used in the decision making subsystem of CONRAD.

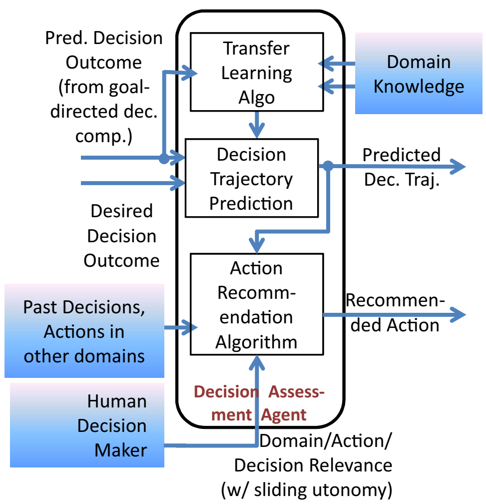

Fig. 3. Different components within the decision assessment subsystem of CONRAD.

well-known decision trajectories can be constructed from past decisions and then Bayesian inference can be used to classify the current decision trajectory into one of the trajectory types. The historical aggregated decisions can be further refined with domain knowledge to reflect the changes in the situation since the decisions were aggregated. The decision assessment subsystem also suggests actions related to future information collection to the human decision maker and to CONRAD's information extraction/evaluation component using AI-based planning techniques such as MDPs and POMDPs [12]. The outcome of the action recommendation algorithm would be a probabilistic distribution over recommended actions from which an action can be picked strategically by CONRAD's Information Extraction and Valuation component.

IV. Information Extraction and Valuation

The key to efficient decision making is to ensure that the available information is dynamically updated and important correlations in data are accurately captured. To achieve these objectives, CONRAD will perform the following operations in real time;

Extract Decision Variables from Raw Data. The data extraction tool of CONRAD extracts data from different heterogeneous and potentially changing sources and filters decision variables - id of the data creator, the data creation time, and a list of key fields such as demography, topic of discussion, etc. We will use Semantic Technology and represent the list of fields through an ontology based language such as OWL. Our goal is to create a database similar to DBpedia (dbpedia.org/About) that will allow users to submit queries with multiple conditions and identify entities that fulfill those queries. The correlations between the decision variables are modelled as networks (or graphs). The vertices in the network represent the entities and the edges represent the correlations. Using this collected data, the information component performs the following subtasks: *(a) Creating Multilevel Networks.* A network is created from the processed data as follows - one field in the dataset is identified as the entity variable and other selected field(s) as the relation variable(s). Each vertex of the network represents a unique instance of the entity variable (here each entry is the name) and two entities are connected if they satisfy certain relations between the relation variables (for example, ids with age difference of five years or less are connected). The connectivity patterns of the networks can with the time stamp changes. Networks based on the same entity variable can be further combined to a multilevel network. This enables us to unearth obscure information that is not immediately relevant from only one database. *(b) Real Time Analysis of Networks.* The analysis of the networks provides insights to the characteristics of the data. Some of the common analysis objectives in CONRAD include (i) detecting communities to identify tightly connected groups of vertices [10] and (ii) computing centrality metrics, core numbers and driver nodes to determine the influential people (or data) [11]. We plan to extend these analysis by including the semantics of the networks. The edges in the network will be annotated by their semantic values (i.e. age, demography, etc). We can therefore refine communities obtained from the initial vertex based method combining entities that have similar semantic values in their links.

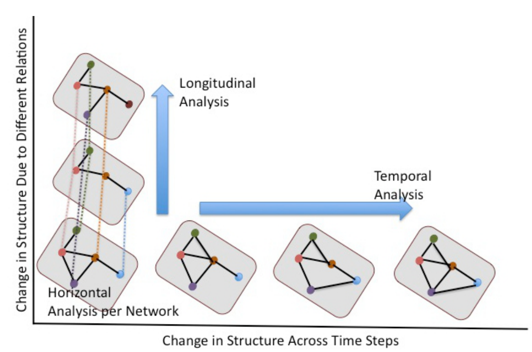

Fig. 4. Analysis of network models over three levels; vertices of the same color represent the same entity

CONRAD performs network analysis operations at three levels, as illustrated in Figure 4. The first is the horizontal level that analyzes each entity network. The second is the longitudinal level where the analysis is conducted across levels (the networks at each level have the same entity variable, but the relational variable(s) and therefore the structure is different). The third level is the temporal level where we track the changes to the network structure across different time steps [2]. CONRAD will implement parallel algorithms and approximate methods to perform the analysis in real time [1], [13]). The information network is connected to the decision variables by matching the component networks, each representing a decision variable to the appropriate decision making agent. For example, if the agent's decision is to deliver supplies to disaster stricken areas, then the agent has to obtain information from networks whose entity variable is the location as well as from the network whose entity variable is the demography.

Identifying Critical Decision Variables. Identify important decision variables that can predict future events will enable users to maintain the correct decision trajectory. The critical variables are the ones whose corresponding networks guarantee that the analysis results are accurate under various perturbations to the data and are sensitive to changes in the data. CONRAD evaluates the reliability of the network models based on *well-posedness* and the *sensitivity* with respect to the analysis objective. To the best of our knowledge, our work is one of the first instances that a network analysis toolkit will include a component to compute the accuracy of the data. *Well-posedness* is a measure of whether the analysis objective, is feasible for a given network. To compute well-posedness of a network, CONRAD computes the number of solutions that the network has for a given analysis function. For community detection, this can be computed by changing the vertex ordering, and then taking the consensus of the communities obtained at each ordering to find the well-posed subgraph [4]. This computation can be extended to the overlapping communities as well. For centrality metrics and core numbers, we are interested in only the high valued ones. To determine whether a network is well-posed, the centrality values for each vertex is first evaluated and then the size of the set of 'high-valued' vertices is checked. If this set is very large, then none of the vertices will be distinctively important. *Sensitivity* measures whether a small change to the input produces a commensurate change in the results. To compute the sensitivity of network analysis, CONRAD uses models of small perturbations (or noise) to the network and metrics to evaluate this noise [6]. After evaluating all the networks based on these well-posedness and sensitivity, the system will retain only the ones that produce accurate results and are sensitive to changes in data. The entity and relation variables of these networks are the critical variables and will produce reliable data patterns that can be used for prediction.

Integrating Decision Making and Data Extraction. The final objective of the information extraction and valuation component in CONRAD is to use the recommendations from the decision assessment algorithm to update the information networks. Based on the recommended actions, the information component tries to extract 'meaningful information' from 'raw data'. The main operations of this process are (i) improving the data gathering mechanism, (ii) improving the quality of the networks and (iii) improving accuracy of the analysis.

Data Gathering. The data gathering operation can be improved by adding more varied sources of information. For example, we can enrich information about possible disasters, by including information of past hurricanes and earthquakes, in addition to tracking the current disaster through news sources, and social network sites.

Adaptive Refining of Data Data is generally gathered 'wholesale', without specifically considering the subsequent use of the information. In the network modeling stage, the system refines the data by filtering the initial network based on certain combinatorial properties. For example if the agents' focus is on finding clusters of similar entities, then a chordal graph based filtering that will retain only the tightly connected components in the network is used. Conversely if the decisions are to be based on centrality metrics, then filtering to reduce the low weight edges is more effective [5].

V. Conclusion

In this position paper, we have proposed CONRAD, a real-time, proactive decision aiding tool that leverages the advantages of game theory, machine learning and network analysis. Each of the individual components proposed for CONRAD have been shown to be successful in their respective domains and we posit that combining them will further enhance the decision making capabilities.

References

[1] D. Bader, S. Kintali, K. Madduri and M. Mihail. Approximating Betweenness Centrality. *The 5th Workshop on Algorithms and Models for the Web-Graph* (WAW2007), 4863, 134, 2007.

[2] S. Bansal, S. Bhowmick and P. Paymal. Fast Community Detection For Dynamic Complex Networks. *Communications in Computer and Information Science* Vol. 116, 196-207, 2010.

[3] F. Burstein and C. Holsapple (editors), "Handbook on Decision Support Systems 1: Basic Themes,", International Handbooks on Information Systems, Springer, 2008.

[4] T. Chakrabory, S. Srinivasan, N. Ganguly, S. Bhowmick, A. Mukherjee. Constant Communities in Complex Networks. *Nature Scientific Reports* 3, Article number: 1825 doi:10.1038/srep01825 2013.

[5] K. Dempsey, S. Bhowmick and H. Ali. Function-preserving Filters for Sampling in Biological Networks. *Procedia Computer Science*,Vol. 9, 587-595, 2012.

[6] A. Flaxman and A. M. Frieze. The diameter of randomly perturbed digraphs and some applications. APPROX-RANDOM, 2004.

[7] R. Hanson, Logarithmic Market scoring rules for Modular Combinatorial Information Aggregation. *Journal of Prediction Markets*, 1(1):3-15, 2007.

[8] J. Jumadinova, and P. Dasgupta. Distributed Decision Making in a Multi-agent Prediction Market using a Partially Observable Stochastic Game. *Proceedings of the 13th International Conference on E-Commerce (ICEC). Liverpool, UK,*:21:1-21:10, 2011

[9] J. Jumadinova and P. Dasgupta, A Multi-Agent System for Analyzing the Effect of Information on Prediction Markets. *Intl. J. of Intell. Systems* 26(5):383-409, 2011.

[10] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E.* 69, 2004.

[11] Y.-Y. Liu and J.-J. Slotine and A.-L. Barabasi. Controllability of complex networks. *Nature* 473, 167173. 2011. doi:10.1038/nature10011

[12] S. Russell and P. Norvig. Artificial Intelligence: A Modern Approach, Prentice Hall, 2009.

[13] V. Ufimtsev and S. Bhowmick, Application of Group Testing in Identifying High Betweenness Centrality Vertices in Complex Networks. *11th Workshop on Machine Learning with Graphs, KDD* 2013.

[14] J. Wolfers and E. Zitzewitz. Prediction Markets. *Journal of Economic Perspectives*, 18(2):107-126, 2004.