

Managing the Pedigree and Quality of Information in Dynamic Information Sharing Environments

Bin Yu, Srikanth Kallurkar, Ganesh Vaidyanathan, and Donald Steiner
Quantum Leap Innovations, Inc.
3 Innovation Way, Suite 100
Newark, DE 19711, USA
{byu, svk, gv, dds}@quantumleap.us

ABSTRACT

The quality of information is crucial for decision making in many mission-critical applications such as battlefield operations and intelligence analysis. However, as the system becomes larger and more diverse, it is becoming increasingly difficult to assess the quality of information from various operators or data sources. In this paper we propose an agent-based approach to managing the quality of information, e.g., its trustworthiness, in network centric information sharing environments, where software agents collaborate with each other to automatically represent and assess the trustworthiness of information from its pedigree within the framework of Dempster-Shafer theory.

Categories and Subject Descriptors

I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence—*Multiagent Systems*

Keywords

Software agents, information sharing, quality of information, trust, pedigree, Dempster-Shafer theory

1. INTRODUCTION

Network centric warfare such as Joint Battlespace Infosphere (JBI) integrates widely dispersed human decision makers with various data sources into a highly dynamic information sharing system [1]. One challenge in JBI-like collaborative information sharing environments is that information provided by different clients, including operators and data sources, is of varying quality, while many mission-critical applications such as battlefield operations and intelligence analysis often require trustworthy and high-quality information [3].

Assessment of information quality can be partially captured by its trustworthiness and pedigree, e.g., source and history [2]. For example, information from reliable sources

usually has higher quality than that provided by unknown or questionable sources. However, in a dynamic network centric environment with diverse communities of interests, it becomes increasingly difficult for human operators to record and analyze the history of derived information and the reliability of various clients [4].

In this paper we propose an agent-based approach to managing the quality of information in a principled way. The goal of this work is to effectively manage the quality of information in a highly dynamic information sharing environment, where the exchanged information may be changed and merged by operators after it is published. Specifically, we describe how pedigree of information can be used to support and enhance the assessment of the quality of information. In our approach, there are multiple client agents (each of which is associated with an operator) and a provenance agent. The client agent collaborates with the provenance agent to track the pedigree of the information and the reliability of each client, e.g., an analyst, a news source or a sensor. Moreover, the client agent reasons about the quality of the derived information in the framework of Dempster-Shafer theory based on the known reliability of each information provider in the pedigree graph and their trust on each object. The proposed approach is currently being evaluated using a simulation testbed for an agent-based information sharing environment.

2. INFORMATION PEDIGREE IN JBI-LIKE ENVIRONMENTS

JBI is a common framework for tactical information sharing and dissemination among commanders in different levels of echelons and operators in the battlefield command and control center [1]. In JBI-like environments, information is stored and made available in the form of information objects. The JBI platform maintains a repository of such information objects. These objects are statements about the real world, such as a SAR sensor report or a recorded UAV video. A JBI information object contains metadata about the information, *information metadata*, as well as the information itself, *information payload* (see Figure 1). The metadata defines a common set of attributes of a JBI object, such as version number, publication time, publisher, pedigree (parents), subject keywords, language, and etc.

There are two kinds of clients in a JBI-like environment: *operators* and *data sources*. The data sources such as sensors and news sources can publish information to the system, but they cannot manipulate information objects once they

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AAMAS'07 May 14–18 2007, Honolulu, Hawai'i, USA.
Copyright 2007 IFAAMAS .

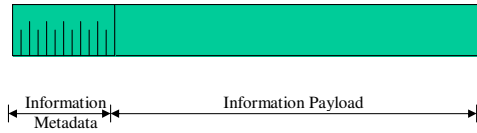


Figure 1: The information metadata and payload in a JBI information object.

are published. Instead, an operator can query, manipulate and merge the published information. The operator can also publish the merged information object and become its owner (the publisher of the object). For example, in the battlefield, a commander may publish a “spot report” based on available intelligence reports, sensor data and other background information such as terrain and weather. Similarly, in the domain of intelligence analysis, an analyst often needs to assess the reports based on his/her understanding of the past and current relevant documents. The analyst will also write reports that reflect his/her critical thinking and prediction about the likely course of action of a specified event. The derived information published in the system can be shared with other analysts with similar interests.

In order to allow users to find detailed information, a derived information object usually contains references to the objects from which it is derived, e.g., its parents. In JBI this “parents” information is recorded in the metadata of an information object. Given the parental information in the metadata of information objects, we can easily build up the pedigree graph for the derived information object. Figure 2 shows an example of pedigree graph for a derived information object *D*. As shown in Figure 2, if the user merges three pieces of information A, B, and C to create a third piece of information D, the parents of D are A, B, and C. The grandparents of D are the union of the parents of A, B, and C which in Figure 2 would include X, Y, and Z.

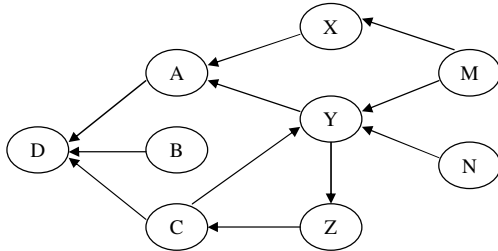


Figure 2: An example of the pedigree graph for information object *D*.

However, for information derived from many sources, a brute force method that annotates and retrieves the entire pedigree information with a tracking log may be combinatorially expensive in storage requirements and in processing time. We have developed a preliminary framework for scalable pedigree information representation and exploitation, in which we provide a suitable representation of pedigree data, *information chromosome* or *VChrome*, that can be configured to provide cost-apportioned provenance informa-

tion given an operator’s requirement [4].

Specifically, the operators can explicitly configure the scope of provenance in the pedigree graph within certain storage and time bounds. For example, in Figure 2, an operator might only want to access the upstream nodes of 2 hops away from object *D*, e.g., *A*, *B*, *C*, *X*, *Y*, and *Z*. The motivation here is that the size of the annotations may grow exponentially with the scope of a pedigree graph, while the utility of these annotations may diminish quickly along the path from the derived information object.

3. OUR APPROACH

In JBI-like environments data sources often provide ambiguous, uncertain and even contradictory information. Moreover, some of the quality of source information as well as the pedigree metadata may be unknown due to lacking knowledge or security concerns. For these two reasons, we use the Dempster-Shafer theory as the underlying computational framework to capture conflicting subjective judgements such as trust and distrust towards an information object as well as the ignorance about the pedigree and quality of an information object [6].

Specifically, we extend the existing JBI model in two ways,

- **Trustworthiness of an object** Evidential intervals (EI) $[l, h]$ in Dempster-Shafer theory are used to model the degree of trust and distrust on information objects, where l and h are the lower and higher bounds of the interval and $0 \leq l \leq h \leq 1$.
- **Reliability of a client** The reliability of each client s_i is captured as a scalar r_i in $[0, 1]$, which is dynamically adjusted based on the feedback from operators for derived information objects. The value of r_i is close to zero if the entity or source s_i is unreliable.

In order to facilitate the interactions between operators and the system, each operator is associated with a software agent, called client agent. Moreover, there is a provenance agent in the system. The provenance agent assesses the metadata of each information object and reliability of each client. The provenance agent provides the following two services to all client agents: (1) retrieve, track and update the reliability of clients; (2) reason about the trustworthiness of a derived information object. The client agent collaborates with the provenance agent to annotate the calculated evidential interval of a derived information object in its metadata when the client s_i publishes the object into the repository.

We now introduce the key concepts of the Dempster-Shafer approach. Let T mean that the given agent considers a given object to be trustworthy. A frame of discernment $\{T, \neg T\}$ is the set of propositions under consideration.

DEFINITION 1. Let Θ be a frame of discernment. A basic probability assignment (bpa) is a function $m : 2^\Theta \mapsto [0, 1]$ where (1) $m(\phi) = 0$ (ϕ is the empty set), and (2) $\sum_{A \subseteq \Theta} m(\hat{A}) = 1$.

An evidential interval $[l, u]$ in Dempster-Shafer theory is a subinterval of $[0, 1]$. The lower bound, called the support or belief, indicates the degree to which the evidence supports the propositional statement in question, while the upper bound, called the plausibility, indicates the degree to

which the evidence fails to refute the proposition, i.e., the degree to which it remains plausible. Examples of these intervals are described in Figure 3, where the shaded area describes the upper and lower bounds of each interval.

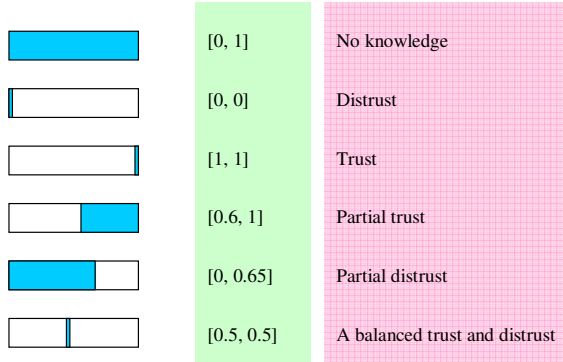


Figure 3: Examples of evidential intervals in Dempster-Shafer theory.

The following definition gives the relationship between an evidential interval and a belief function.

DEFINITION 2. Formally, given a basic probability assignment m , the degree of belief and plausibility in an evidential interval $[l, u]$ can be defined as (1) $l = m(\{T\})$, (2) $u = m(\{T\}) + m(\{T, \neg T\})$.

A subset \hat{A} of a frame Θ is called a *focal element* of a belief function Bel over Θ if $m(\hat{A}) > 0$. Given two belief functions over the same frame of discernment but based on distinct bodies of evidence, *Dempster's rule of combination* enables us to compute a new belief function based on the combined evidence.

DEFINITION 3. (*Dempster's rule of combination*) Let Bel_1 and Bel_2 be belief functions over Θ , with basic probability assignments m_1 and m_2 , and focal elements $\hat{A}_1, \dots, \hat{A}_k$, and $\hat{B}_1, \dots, \hat{B}_l$, respectively. Suppose

$$\sum_{i,j, \hat{A}_i \cap \hat{B}_j = \phi} m_1(\hat{A}_i) m_2(\hat{B}_j) < 1$$

Then the function $m : 2^\Theta \mapsto [0, 1]$ that is defined by $m(\phi) = 0$, and

$$m(\hat{A}) = \frac{\sum_{i,j, \hat{A}_i \cap \hat{B}_j = \hat{A}} m_1(\hat{A}_i) m_2(\hat{B}_j)}{1 - \sum_{i,j, \hat{A}_i \cap \hat{B}_j = \phi} m_1(\hat{A}_i) m_2(\hat{B}_j)} \quad (1)$$

for all non-empty $\hat{A} \subset \Theta$ is a basic probability assignment [5].

Bel , the belief function given by m , is called the *orthogonal sum* of Bel_1 and Bel_2 . It is written $\text{Bel} = \text{Bel}_1 \oplus \text{Bel}_2$.

Obviously, information from reliable clients usually has higher quality than the one provided by unreliable sources. Given each client's belief in the information object and the client's reliability r_i , the agent's belief in the derived object can be computed in many different ways. For simplicity, here we only consider the effects of its direct parents in the pedigree graph on its trustworthiness.

Suppose d is derived from a list of information objects $\{d_1, d_2, \dots, d_L\}$ by client s and the provenance agent will estimate the trustworthiness of object d . We now show how trustworthiness of object d can be determined from its belief m_i on the corresponding objects d_i , where m_i is converted from the evidential interval of d_i and $d_i \in \{d_1, d_2, \dots, d_L\}$.

DEFINITION 4. Suppose d is derived from a list of objects d_1, d_2, \dots, d_L by client s , the belief of object d can be defined as $m = m_1 \oplus m_2 \oplus \dots \oplus m_L$, which m_i is the belief function corresponding to the trustworthiness of the information object d_i .

The motivation behind Definition 4 is that the degree of trust is higher for an object when it is derived from trustworthy information objects. However, when there are two or more information objects with conflicting belief, the value of $m(\{T\})$ and $m(\{\neg T\})$ of the derived information object will mostly contribute to the ignorance $m(\{T, \neg T\})$.

4. CONCLUSION

In this paper we present a computational approach to managing the pedigree and quality of information in dynamic information sharing environments in the framework of Dempster-Shafer theory. Our approach enables operators to easily identify trustworthy information in dynamic environments. In the future work we will empirically evaluate the effectiveness of our approach within a simulated agent-based information sharing environment. We also plan to develop techniques for capturing and integrating the feedback from users to better assess the trustworthiness of information objects and the reliability of different clients.

Acknowledgements

This research has been sponsored in part by AFRL under Contract No. FA8750-06-C-0052. We are indebted to the anonymous reviewers for their helpful comments.

5. REFERENCES

- [1] Joint battlespace infosphere (JBI). <http://www.rl.af.mil/programs/jbi/>.
- [2] Marion G. Ceruti, Subrata Das, Adam Ashenfelter, Gary Raven, Richard Brooks, Moises Sudit, Genshe Chen, and Edward Wright. Pedigree information for enhanced situation and threat assessment. In *Proceedings of the Ninth International Conference on Information Fusion*, 2006.
- [3] M. Edwin Johnson and K. C. Chang. Quality of information for data fusion in net centric publish and subscribe architectures. In *Proceedings of the Eighth International Conference on Information Fusion*, 2005.
- [4] Srikanth Kallurkar, Apperson Johnson, Donald Steiner, and Robert Flo. AIMS: Agent-based information management system. In *Proceedings of AAAI Fall Symposia on Semantic Web for Collaborative Knowledge Acquisition*, 2006.
- [5] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.
- [6] Bin Yu and Munindar P. Singh. An evidential model of distributed reputation management. In *Proceedings of First International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 294–301, 2002.