E

Embedding Bayesian Networks in Sensor Grids

Juan E. Vargas
University of South Carolina, USA

INTRODUCTION

In their simplest form, sensors are transducers that convert physical phenomena into electrical signals. By combining recent innovations in wireless technology, distributed computing, and transducer design, grids of sensors equipped with wireless communication can monitor large geographical areas. However, just getting the data is not enough. In order to react intelligently to the dynamics of the physical world, advances at the lower end of the computing spectrum are needed to endow sensor grids with some degree of intelligence at the sensor and the network levels. Integrating sensory data into representations conducive to intelligent decision making requires significant effort. By discovering relationships between seemingly unrelated data, efficient knowledge representations, known as Bayesian networks, can be constructed to endow sensor grids with the needed intelligence to support decision making under conditions of uncertainty. Because sensors have limited computational capabilities, methods are needed to reduce the complexity involved in Bayesian network inference. This paper discusses methods that simplify the calculation of probabilities in Bayesian networks and perform probabilistic inference with such a small footprint that the algorithms can be encoded in small computing devices, such as those used in wireless sensors and in personal digital assistants (PDAs).

BACKGROUND

Recent innovations in wireless development, distributed computing, and sensing design have resulted in energy-efficient sensor architectures with some computing capabilities. By spreading a number of smart sensors across a geographical area, *wireless ad-hoc sensor grids* can be configured as complex monitoring systems that can react intelligently to changes in the physical world. Wireless sensor architectures such as the University of California Berkeley's Motes (Pister, Kahn, & Boser, 1999) are being increasingly used in a variety of fields to

1. Monitor information about enemy movements, explosions, and other phenomena of interest (Vargas & Wu, 2003)
2. Monitor chemical, biological, radiological, nuclear, and explosive (CBRNE) attacks and materials
3. Monitor environmental changes in forests, oceans, and so forth
4. Monitor vehicle traffic
5. Provide security in shopping malls, parking garages, and other facilities
6. Monitor parking lots to determine which spots are occupied and which are free

Typically, sensor networks produce vast amounts of data, which may arrive at any time and may contain noise, missing values, or other types of uncertainty. Therefore, a theoretically sound framework is needed to construct virtual worlds populated by decision-making agents that may receive data from sensing agents or other decision-making entities. Specific aspects of the real world could be naturally distributed and mapped to devices acting as data collectors, data integrators, data analysts, effectors, and so forth. The data collectors would be devices equipped with sensors (optical, acoustical, radars, etc.) to monitor the environment or to provide domain-specific variables relevant to the overall decision-making process. The data analysts would be engaged in low-level data filtering or in high-level decision making. In all cases, a single principle should integrate these activities. Bayesian theory is the preferred methodology, because it offers a good balance between the need for separation at the data source level and the integrative needs at the analysis level (Stone, Lawrence, Barlow, & Corwin, 1999). Another major advantage of Bayesian theory is that data from different measurement spaces can be fused into a rich, unified, representation that supports inference under conditions of uncertainty and incompleteness. Bayesian theory offers the following additional advantages:

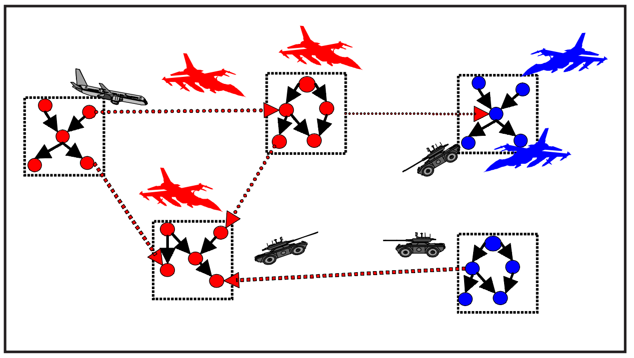
- Robust operational behavior: Multisensor data fusion has an increased robustness when compared to single-sensor data fusion. When one sensor becomes unavailable or is inoperative, other sensors can provide information about the environment.
- Extended spatial and temporal coverage: Some parts of the environment may not be accessible to some sensors due to range limitations. This occurs especially when the environment being monitored is

vast. In such scenarios, multiple sensors that are mounted at different locations can maximize the regions of scanning. Multisensor data fusion provides increased temporal coverage, as some sensors can provide information when others cannot.
- Increased confidence: Single target location can be confirmed by more than one sensor, which increases the confidence in target detection.
- Reduced ambiguity: Joint information from multiple sensors can reduce the set of beliefs about data.
- Decreased costs: Multiple, inexpensive sensors can replace expensive single-sensor architectures at a significant reduction of cost.
- Improved detection: Integrating measurements from multiple sensors can reduce the signal-to-noise ratio, which ensures improved detection.

Bayesian-based or entropy-based algorithms can the used to construct efficient data structures–known as *Bayesian networks*–to represent the relations and the uncertainties in the domain (Pearl, 1988). After the Bayesian networks are created, they can act as hyperdimensional knowledge representations that can be used for probabilistic inference. In situations when data is not as rich, the knowledge representations can still be created from statements of causality and independence formulated by expert opinions. Under this framework, activities such as vehicle control, maneuvering, and scheduling could be planned, and the effectiveness of those plans could be evaluated online as the actions of the plans are executed.

To illustrate these ideas, consider a domain composed of threats, assets, and grids of sensors. Although unmanned vehicles loaded with sensors might be able to detect potential targets and provide data to guide the distribution of assets, integrating and transforming those data into meaningful information that is amenable for intelligent decisions is very demanding. The data must be filtered, the relationships between seemingly unrelated data sets must be determined, and knowledge representations must be created to support wise and timely decisions, because conditions of uncertain and incomplete information are the norm, not the exception. Therefore, a solution is to endow sensors with embedded local and global intelligence, as shown in Figure 1. In the figure, friendly airplanes and tanks, in red, use Bayesian networks (BNs) to make decisions. The BNs are illustrated as red boxes containing graphs. Each node in the graphs corresponds to a variable in the domain. The data for each variable may come from sensors spread in the battlefield. The red nodes are variables related to the friendly resources, and the blue variables to the enemy resources. The dotted red arrows connecting the BNs represent wireless communication between the BNs.

Figure 1. A domain scenario



The goal is to recognize situations locally and globally, identify the available options, and make global and local decisions quickly in order to reduce or eliminate the threats and optimize the use of assets. The task is difficult due to the dynamics and uncertainties in the domain. Threats may change in many ways, targets may move, enemy forces may identify the sensing capabilities and eliminate them, and so forth. In most cases, sensing information will contain noise and most likely will be inaccurate, unreliable, and uncertain. These constraints suggest a distributed, bottom-up approach to match the natural dynamics and uncertainties of the problem. Thus, at the core of this problem, a theoretical framework that effectively balances local and global conditions is needed. Distributed Bayesian networks offer that balance (Xiang, 2002; Valtorta, Kim, & Vomlel, 2002).

MAIN THRUST

Embedding Bayesian Networks in Sensor Grids

Recent advances in the theory of Bayesian network inference (Darwiche, 2003; Castillo, Gutierrez, & Hadi, 1996; Utete, 1998) have resulted in algorithms that can perform probabilistic inference on very small-scale computing devices that are comparable to commercially available PDAs. The algorithms can encode, in real-time, families of polynomial equations representing queries of the type $p(e|h)$ involving sets of variables local to the device and its neighbors.

Using the knowledge representations locally encoded into these devices, larger, distributed systems can be interconnected. The devices can assess their local conditions given local observations and engage with other devices in the system to gain better understanding of the global situation, to obtain more assets, or to convey

Related Content

Addressing Challenges in Data Analytics: A Comprehensive Review and Proposed Solutions
Lakshmi Haritha Medidaand Kumar (2024). *Critical Approaches to Data Engineering Systems and Analysis (pp. 16-33).*
www.irma-international.org/chapter/addressing-challenges-in-data-analytics/343880

Clustering in the Identification of Space Models
Maribel Yasmina Santos, Adriano Moreiraand Sofia Carneiro (2005). *Encyclopedia of Data Warehousing and Mining (pp. 165-171).*
www.irma-international.org/chapter/clustering-identification-space-models/10586

Business Processes
David Sundaramand Victor Portougal (2005). *Encyclopedia of Data Warehousing and Mining (pp. 118-123).*
www.irma-international.org/chapter/business-processes/10577

Discovering Frequent Embedded Subtree Patterns from Large Databases of Unordered Labeled Trees
Yongqiao Xiao, Jenq-Foung Yaoand Guizhen Yang (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 3235-3251).*
www.irma-international.org/chapter/discovering-frequent-embedded-subtree-patterns/7832

Heterogeneous Gene Data for Classifying Tumors
Benny Yiu-ming Fungand Vincent To-yee Ng (2005). *Encyclopedia of Data Warehousing and Mining (pp. 550-554).*
www.irma-international.org/chapter/heterogeneous-gene-data-classifying-tumors/10658