Information Fusion 18 (2014) 43–61

Contents lists available at SciVerse ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

A fuzzy graph matching approach in intelligence analysis and maintenance of continuous situational awareness

Geoff Gross, Rakesh Nagi, Kedar Sambhoos *

Center for Multisource Information Fusion, Department of Industrial and Systems Engineering, State University of New York at Buffalo, Buffalo, NY, USA

ARTICLE INFO

Article history:
Received 27 October 2011
Received in revised form 17 September 2012
Accepted 6 May 2013
Available online 5 June 2013

Keywords:
Graph matching
Stochastic graphical methods
Fuzzy systems
Situational awareness
Incremental graph matching

ABSTRACT

In intelligence analysis a situation of interest is commonly obscured by the more voluminous amount of unimportant data. This data can be broadly divided into two categories, hard or physical sensor data and soft or human observed data. Soft intelligence data is collected by humans through human interaction, or human intelligence (HUMINT). The value and difficulty in manual processing of these observations due to the volume of available data and cognitive limitations of intelligence analysts necessitate an information fusion approach toward their understanding. The data representation utilized in this work is an attributed graphical format. The uncertainties, size and complexity of the connections within this graph make accurate assessments difficult for the intelligence analyst. While this graphical form is easier to consider for an intelligence analyst than disconnected multi-source human and sensor reports, manual traversal for the purpose of obtaining situation awareness and accurately answering priority information requests (PIRs) is still infeasible. To overcome this difficulty an automated stochastic graph matching approach is developed. This approach consists of three main processes: uncertainty alignment, graph matching result initialization and graph matching result maintenance. Uncertainty alignment associates with raw incoming observations a bias adjusted uncertainty representation representing the true value containing spread of the observation. The graph matching initialization step provides template graph to data graph matches for a newly initialized situation of interest (template graph). Finally, the graph matching result maintenance algorithm continuously updates graph matching results as incoming observations augment the cumulative data graph. Throughout these processes the uncertainties present in the original observations and the template to data graph matches are preserved, ultimately providing an indication of the uncertainties present in the current situation assessment. In addition to providing the technical details of this approach, this paper also provides an extensive numerical testing section which indicates a significant performance improvement of the proposed algorithm over a leading commercial solver.

1. Introduction

Intelligence analysis is a highly complex field which must consider data flowing from a wide array of reporting sources. With the recent paradigm shift in warfare methods to a counterinsurgency (COIN) approach, an increased importance has been placed on swift and accurate intelligence analysis. Accuracy in intelligence analysis requires cohesive intelligence collection for the purpose of obtaining situation awareness or an understanding of the current state of the world. The shift toward more non-kinetic, information based operations requires inventive approaches to handle the ever-increasing volume of useful data.

In the COIN domain the insurgent force attempts to undermine and disrupt political authority through the use of subversive techniques including violence. Recent examples of COIN areas of operation include Afghanistan and the Horn of Africa. The insurgents have some strategic advantages such as their ability to operate in secrecy, provide misinformation and dictate the time and location of armed clashes. Data flows from a wide variety of sources including: hard or physical sensors (e.g. radar, cameras, etc.) and soft or human observations (such as soldiers in the field, informants, and interviews). The combination and exploitation of this data requires specialized processes to ensure timely and reliable consideration of all pertinent data. The heterogeneity and sheer volume of the available data along with the overarching goal of providing domain-wide situational awareness to the commanding officers suggest an information fusion approach should be utilized.

While the consideration of hard data is traditionally well studied within the fusion community, soft or human observations represent a rarely integrated form of data. The importance of this data, particularly within the domain of COIN intelligence analysis

* Corresponding author.
E-mail addresses: gagross@buffalo.edu (G. Gross), nagi@buffalo.edu (R. Nagi), sambhoos@cubrc.org (K. Sambhoos).

1566-2535/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.inffus.2013.05.006

necessitates innovative approaches for its integration into fusion systems. Human observations possess the useful abilities of observing attributes which hard sensors cannot and observing many different attribute types simultaneously. These beneficial characteristics come at the cost of requiring specialized processing methods.

The unique characteristics of soft data which require different processing methods from its hard counterpart include: difficulty in extracting meaning from plain text or raw data, the context dependent nature of observation error characteristics, the methods for representing the uncertainties present and the methods for reasoning over the uncertainties to form conclusions about the state of the world. The focus of this paper is on the methodology for uncertainty representation and consideration.

The remainder of this paper is organized as follows: Section 2 provides background on the problem domain, Section 3 presents the methodology for uncertainty alignment of soft observations, Section 4 details the graph matching procedure utilized to obtain situational awareness, Section 5 explains the incremental methods for preserving situational awareness, Section 6 describes the batching methodology for streaming incoming observations, Section 7 presents an end to end processing example, Section 8 performs numerical testing on the proposed algorithms, Section 9 provides a discussion of the results and Section 10 presents some conclusions.

2. Background

Intelligence analysts are faced with an ever expanding collection of available data sources as collection capabilities increase and senor costs continue to decrease. These disparate sources often contain references to common entities while containing varying levels of uncertainty. The analyst is faced with the cognitively challenging task of associating these pieces of information while providing situational awareness to the commanding officers. Among the many cognitive challenges facing intelligence analysts are the large volume of available data, high degree of heterogeneity, highly complex connections [1] and the uncertainties present in the observed data. As Heuer points out [2], “the mind is poorly ‘wired’ to deal with (uncertainty)”. To help overcome these cognitive challenges an automated system can be developed to trim irrelevant data, perform data alignment to a common format and identify complex connections while presenting the uncertainties of the data and assessments through a uniform representation. The system design presented here is built on the information fusion paradigm.

The information fusion approach is a multilevel methodology for combining large quantities of data with the goal of obtaining situation and threat assessments. Data fusion accomplishes these goals through the combination and exploitation of the competencies of multiple sources of information. The typical data fusion levels as defined by the Joint Directors of Laboratories (JDL) fusion process model are provided in [3].

Although fusion processing systems do not necessarily require all fusion levels and the tasks are not necessarily performed in numerical order, the multilevel description presents a common lexicon for the discussion of fusion systems. The focus of this paper is on the Level 0 source preprocessing task of uncertainty alignment (Section 3) and the Level 2 situation assessment task performed through graph matching (Sections 4 and 5). The Level 1 task of object assessment is not discussed here since this natural language processing task is covered by a separate effort [4–6].

3. Uncertainty alignment

With observations taking many different forms and coming from many different sources, it is unreasonable to expect the methods for initially representing the uncertainties related to these observations to be common. One easily made distinction between uncertain representations is between representations of qualitative and quantitative linguistic terms. For example, an observation of “a few people” would require a different native uncertainty representation than an observation of the same group of people represented by “three people”. The numerosity of people described qualitatively by “a few” is best represented by a possibilistic uncertainty function while the quantity of people represented quantitatively by “three” is best represented by a probabilistic uncertainty function. A discussion of these differing uncertainty theories is provided subsequently.

3.1. Uncertainty representation: probability and possibility theory

Uncertainties can be broadly categorized into two categories, aleatory or statistical uncertainty and epistemic or systematic uncertainty. Aleatory uncertainty is uncertainty due to natural system variability and is commonly represented probabilistically. Epistemic uncertainty characterizes the uncertainty due to a lack of knowledge. Epistemic uncertainty is typical of linguistic observations due to their vague and highly context dependent nature. Possibility theory has been suggested as a method for handling this type of uncertainty [7]. A description of these uncertainty frameworks follows.

Probability theory is a well-studied method for representing uncertainties which has been utilized for hundreds of years. Probability theory is the most commonly used uncertainty theory in fusion systems. The strength of probability theory is the ability to draw clear cut conclusions about differences in probability density functions. This strength comes with the requirement that valid statistical data is available for the uncertain attribute of interest. Unfortunately, in the domain of linguistic observation this data is often unavailable or is of limited statistical validity.

Linguistic observations commonly contain qualitative language which is not easily integrated into a probabilistic framework. To counteract this limitation, possibilistic representations are utilized for these observations. Possibility theory has been successfully utilized in a diverse range of fields, including: law, logic controllers, medicine, computational linguistics and the social sciences [8].

Possibility theory is a more recently developed uncertainty theory which attempts to describe the degree to which an element is possibly a part of a set. Possibility theory is built on the fuzzy set, a set where membership is not a strict binary relationship (membership or non-membership) [9]. Instead, membership in a fuzzy set is defined by a membership function, allowing intermediary membership values. The theory of possibility and the fuzzy set is both intuitively appealing and backed by an axiomatic development [10]. From an intuitive perspective, linguistic observation of fuzzy terms such as “reddish” clearly display a lower membership degree to the concept “red” than an explicit observation of “red”. The measure theoretic development and relationship of fuzzy sets and possibility theory is analogous to the link between classical sets and probability theory.

The less strict set membership requirement in possibility theory leads to some less restrictive defining axioms [11]. For example, the value of the integration over the uncertainty function is not required to be equal to 1 (as it is in probability theory). Possibility theory is better suited for representing vague observations characterized by imprecision and qualitative language and has been shown to mirror humans psychological reasoning about uncertain concepts.

Kochen [12] describes people as “estimators” who assign a degree of belief to a fuzzy concept and reflect this degree of belief in their numerical representation of the concept. Other examples of the natural psychological appeal of fuzzy sets is provided by