Information Fusion 10 (2009) 51–69

Contents lists available at ScienceDirect

Information Fusion

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

Force deployment analysis with generalized grammar

Daniel McMichael *, Geoff Jarrad, Simon Williams, Michael Kennett

CSIRO ICT Centre, Locked Bag 2, Glen Osmond, 5064 South Australia, Australia

ARTICLE INFO

Article history:
Received 16 January 2006
Received in revised form 10 August 2007
Accepted 22 May 2008
Available online 3 June 2008

Keywords:
Information fusion
Generalized grammar
GFCCG
Chomsky hierarchy
Situation assessment
Parsing

ABSTRACT

We analyse the key algorithms of data and information fusion from a linguistic point of view, and show that they fall into two paradigms: the primarily syntactic, and the primarily semantic. We propose an alternative *grammatical paradigm* which exploits the ability of grammar to combine syntactic inference with semantic representation. We generalize the concept of formal generative grammar to include multiple rule classes each having a *topology* and a *base vocabulary*. A generalized Chomsky hierarchy is defined. Analysing fusion algorithms in terms of grammatical representations, we find that most (including multiple hypothesis tracking) can be expressed in terms of conventional regular grammars. Situation analysis, however, is commonly attempted using first order predicate logic, which while expressive, is recursively enumerable and so scales badly.

We argue that the core issue in situation assessment is *force deployment assessment,* the extraction and scoring of hypotheses of the *force deployment history*, each of which is a multiresolution account of the activities, groupings and interactions of force components. The force deployment history represents these relationships at multiple levels of granularity and is expressed over time and space. We provide a grammatical approach for inferring such histories, and show that they can be estimated accurately and scalably. We employ a generalized context-free grammar incorporating both sequence and multiset productions. Elaborating [D. McMichael, G. Jarrad, S. Williams, M. Kennett, Grammatical methods for situation and threat analysis, in: Proceedings of The 8th International Conference on Information Fusion, Philadelphia, PA, 2005], a Generalized Functional Combinatory Categorial Grammar (GFCCG) is described that is both generalized and semantically functional (in that the semantics can be calculated directly from the syntax using a small number of rules). Force deployment modelling and parsing is demonstrated in naval and air defence scenarios. Simulation studies indicate that the method robustly handles the errors introduced by trackers under noisy cluttered conditions. The empirical time complexity of batch force deployment parsing is better than $O(N^{1.5})$, where N is the number of track segments.

Force deployment assessments are required in real-time, and we have developed an incremental parser that keeps up with real-time data, and fulfils at Level 2 in the JDL fusion hierarchy the role that trackers fulfil at Level 1.

1. Introduction

The Joint Directors of Laboratories (JDL) model for data and information fusion [2] has evolved [3,4] to encompass four levels, each describing a function required for the extraction of information from sensor data for the military command and control process. Level 0 involves signal processing, Level 1 concerns the extraction of objects such as combatants, vehicles and installations, Level 2 extracts the interrelationships between such objects, and Level 3 performs impact assessment. This paper is concerned with Level 2: *situation analysis*.

1.1. The requirement

While there is no universally recognised and precise definition of situation assessment [5], there is some consensus that it at least involves the following aspects: (i) counter-surveillance measures, (ii) the environment (terrain, weather, etc.), (iii) socio-policitical background (permissible attrition rates, etc.), (iv) the disposition, deployment and location of forces, and (v) the structure and dynamics of the command systems. In this paper, we focus on real-time determination of *force deployment and location* to provide a coherent appreciation of the battle describing the relationships between the active force components.

The classical view of the situation assessment is that it is a batch process leading to an appreciation of the prospective battle. The approach we present is different: the situation assessment process provides a picture of both the present situation and its

---

* Corresponding author. Tel./fax: +61 88339 7434.

E-mail addresses: Daniel.McMichael@semanticsciences.org (D. McMichael), Geoff.Jarrad@semanticsciences.org (G. Jarrad), Simon.Williams@semanticsciences.org (S. Williams), Michael.Kennett@csiro.au (M. Kennett).

1566-2535/$ - see front matter Crown Copyright © 2008 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.inffus.2008.05.008

history, which can be updated in real time as new information is received.

The consumers of situation assessments are commanders and automated components within the command and control system. Experience at Level 1 has shown that commanders like to be given a clear picture of battle activity that they can explore and query. Irrelevant information should be excluded; and where possible, the information should be unequivocal. A force deployment analysis system should be able to answer such questions as:

- what is the best interpretation of the situation now?
- what are the other credible interpretations?
- how has the situation developed over time?
- what is this force component doing now? what has it been doing?
- which force components are engaged on what activities? where are they?

From the point of view of an individual force component, a force deployment assessment describes its composition, its activities and its interactions with other components. The overall force deployment assessment describes the evolution of the composition, activities and interactions of all components of all forces.

The technology for constructing such analyses needs to be able to model the composition of force components, what they do, and how they interact. It is likely that the history of the objects extracted at Level 1 will have many possible interpretations. Force deployment analysis technology should be able to model such interpretations as hypotheses and to assign levels of belief to each of them, given observations and prior knowledge. It should provide efficient scalable algorithms for extracting the most likely hypotheses and for predicting the situation in the future. Implemented systems should be able to be programmed with expert knowledge, to learn from training data and to adapt to experience. It should be possible for domain experts to easily understand the information they provide. Finally, they should be cheap to configure and deploy, and be amenable to distributed and parallel implementation.

1.2. Previous work

Early work on situation analysis, including the development of the situation calculus [6] and its application to military situation analysis [7] sought to generate a knowledge base that could represent a changing world. The system designer provides models in the form of rule sets, and some external process contributes facts to the knowledge base in real time. A reasoning system responds to queries emanating from human operators and other system components. A number of such systems have been built [8–10].

A constant problem with logical reasoning technology is the need to provide sufficient rules to allow reasoning and constrain search. One approach is to provide coherently created modular ontologies that represent aspects of the application domain expressed in such languages as UML and OWL [9,11–13]. To support this process, Little and Rogova [14] have proposed a hierarchy of situation ontologies to facilitate modular construction.

The situation representation, and the queries upon it, are constrained by the limitations of the logic employed and by the scope of the ontology. First order logics are Turing-complete, but their expressive richness comes at the cost of decidability[1], computational complexity[2], uncertainty of computation time, and uncertainty over whether there is sufficient knowledge in the system to resolve novel real-time queries. However, their flexibility does allow rich interfaces to language and the ability to reason on knowledge itself [15]. Eliciting knowledge is expensive; but gradually better tools are becoming available. The problem remains that pure logical reasoning does not provide any mechanism for the resolution of genuine ambiguities. However, inductive methodologies, such as case-based reasoning, can be so equipped and have been applied to fusion [16].

Graphical models and Bayesian networks are able to represent uncertainty within systems of variables with dependency relationships [17]. Even though inference under such models scales badly [18–20] they can be useful when their internal variables are already bound to evidence in the outside world. Bayesian networks have been taken up within the data and information fusion community [21,22] where mechanisms for composing networks from *modules* [23–26] have eased their application. However, these approaches tend to require manual binding of network variables to external evidence. In situations where the objects and associations are transitory, manual binding techniques require significant configuration effort at the time of use.

A mechanism for assigning probabilities to sentences in propositional logic was first provided by Nilsson in 1986 [27]. The extension of probabilistic reasoning to first order logic has developed gradually [27–33], and recently Laskey [34,35] has provided a comprehensive approach. While the development of these tools is a major achievement, inference using such approaches will scale even worse than the observed exponential scalability of non-probabilistic theorem proving systems [36].

Modelling the behaviour of targets has been tackled using dynamic Bayesian networks [37], which have been modularised [38], and efficient implementations have been developed [39–43], together with appropriate learning algorithms [44]. Applications of hidden Markov (HMM) modelling techniques are still under development, even though these cannot model long-range dependencies without large parametric sensitivities [45]. Higher-order HMMs, which attempt to capture longer range relationships tend to work poorly in comparison to context-free grammars [46] because implementations tend to have vastly too many parameters for satisfactory estimation from realistically-sized data sets. Saul and Jordan [47] have sought to alleviate this problem by modelling the high-order conditional distributions with finite mixtures of low-order components. Context-free grammars are likely to scale still better because they are composable and therefore may be able to cover a large part of the high-order state space with a relatively small number of rules.

Maupin and Jousselme [48] and Steinberg [49] have proposed belief-based logical frameworks for Level 2 analysis. Belief-based techniques for template matching under uncertainty have been developed by Yu et al. [50,51], and a belief-based target aggregation algorithm is summarised by Bakert and Losiewicz [52].

To summarise: on one hand, unconstrained first-order predicate logic provides great flexibility — more than is required to provide force deployment assessments. As a result, tools that use it can be richly featured, but may scale badly. (Specifically, first order binary predicate logic belongs to complexity class RE, and is therefore *intractable*). While probabilistic logic offers better robustness and is able to provide useful marginal and conditional distributions, it scales even worse than deterministic logic. Maximisation over distributions (e.g. to find the best assessment) scales worse still [53]. There are, however, effective and scalable techniques for solving the temporal or the force aggregation sub-problems, but it appears that hitherto there has been no scalable solution for the whole.

1.3. The case for grammar

In our analysis of the requirements of situation analysis, we suggested that a force deployment assessment is a spatio-temporal model that specifies the evolution of force structure and the activ-

---

[1] If the language has at least one predicate of valence at least 2 that is not equality it is undecidable. However, if a well-formed formula is true it is possible to construct an algorithm that will halt to verify its truth.

[2] If the language has at least one predicate of valence at least 2 that is not equality it belongs to complexity class RE.