Abductive Inference for Combat: Using SCARE-S2 to Find High-Value Targets in Afghanistan

Paulo Shakarian
U.S. Army
paulo.shakarian@us.army.mil

Margo K. Nagel, Brittany E. Schuetzle, V. S. Subrahmanian
Dept. of Computer Science
University of Maryland, College Park, MD
mnagel@umd.edu, bschuetz@umiacs.umd.edu, vs@cs.umd.edu

Abstract

Recently, *geospatial abduction* was introduced by the authors in (Shakarian, Subrahmanian, and Sapino 2010) as a way to infer unobserved geographic phenomena from a set of known observations and constraints between the two. In this paper, we introduce the SCARE-S2 software tool which applies geospatial abduction to the environment of Afghanistan. Unlike previous work, where we looked for small weapon caches supporting local attacks, here we look for insurgent high-value targets (HVT's), supporting insurgent operations in two provinces. These HVT's include the locations of insurgent leaders and major supply depots. Applying this method of inference to Afghanistan introduces several practical issues not addressed in previous work. Namely, we are conducting inference in a much larger area ($24,940$ sq km as compared to 675 sq km in previous work), on more varied terrain, and must consider the influence of many local tribes. We address all of these problems and evaluate our software on 6 months of real-world counter-insurgency data. We show that we are able to abduce regions of a relatively small area (on average, under 100 sq km and each containing, on average, 4.8 villages) that are more dense with HVT's ($35\times$ more than the overall area considered).

Introduction

Insurgents operating in Afghanistan require a substantial command-and-control (C2) and logistics support to conduct successful attacks.[1] Military analysts refer to elements that provide C2 and logistics support for large number of insurgent cells as "high-value targets" ("HVTs") as the elimination of these HVTs can have a significant impact on insurgent operations. As a result, NATO and Afghan forces often concentrate on finding these HVTs in an attempt to reduce the level of violence in the country. The insurgents have a limited number of these HVTs that are required to support the activities of lower-level insurgent cells. Additionally, terrain and cultural considerations place constraints on the relationships between an HVT and the lower level insurgent cell it supports. Knowing the locations of the lower-level cells (based on attack data), as well as these constraints (obtained from socio-cultural and terrain data), we wish to abductively infer where the HVTs can be found. This is clearly an instance of a *geospatial abduction* problem originally introduced by the authors in (Shakarian, Subrahmanian, and Sapino 2010) and later extended in (Shakarian and Subrahmanian 2010). We previously applied geospatial abduction to find small weapons hide-sites related to local attacks in Baghdad in (Shakarian, Subrahmanian, and Sapino 2009). However, the environment of Afghanistan provides several challenges that we did not address in the other work. These include the following:

1. In Afghanistan, the influence of multiple tribes affect relationships between areas on the ground. How do we account for this influence?
2. In the two provinces we considered in Afghanistan, the terrain is extremely varied, unlike the more uniform urban terrain of Baghdad. How do we account for this variance in terrain?
3. Unlike our application to Baghdad (25×27 km area), where we could easily discretize the space, our data-set for Afghanistan includes two provinces covering a total area 580×430 km, making discretizing of the space impractical. How do we best represent the space?

We note that using only attack data and socio-cultural information alone will most likely be insufficient to pinpoint a HVT. However, the real-world requirements imposed on the insurgents by logistic and socio-cultural variables should allow a ground commander to significantly reduce the search-space for such targets. Intelligence professionals identify "Named Areas of Interest" or NAIs - regions on the ground where they think HVTs can be located. Then, other intelligence assets, such as unmanned aerial vehicles (UAVs) or tactical human-intelligence (HUMINT) teams can be used in the NAIs to pinpoint targets. (US Army 1994) In a large area, such as a province of Afghanistan, UAVs or HUMINT cannot be used effectively without first determining good NAIs. To address this problem fir the specific case of Afghanistan, we adapted the region-based abduction framework of (Shakarian and Subrahmanian 2010) to our scenario by creating an entirely new piece of software for abductive inference called the SCARE-S2 (Spatio-Cultural Abductive Reasoning Engine System 2). SCARE-S2 abduc-

[1] Note that throughout this paper, 'attack' refers to an attack conducted by the insurgents against coalition forces.

tively finds regions that can then later be used to cue other intelligence assets to find an HVT. Applying SCARE-S2 to our Afghanistan dataset produced regions with a significantly higher density of HVTs (by a factor of 35), where half of the abduced ground regions (normally of an area less than $100km^2$) would contain at least one HVT. Further, each region produced by SCARE-S2 contained, on average, 4.8 villages - hence searching them is not resource-intensive for many surveillance platforms. Due to the high density of HVTs within the regions, we feel that they could be used for NAIs and aide in combat operations.

This paper is organized as follows. First, we briefly review the region-based abduction framework of (Shakarian and Subrahmanian 2010) and present some extensions we used to address our Afghanistan-specific problem. Then we describe our dataset for Afghanistan. This is followed by a description of our implementation along with our experimental results and discussion.

Region-Based Geospatial Abduction

In this section, we briefly review the framework of (Shakarian and Subrahmanian 2010) - which is not new material. This is followed by our practical, Afghan-specific extensions in the next section (which is new in this paper).

We assume the existence of a real-valued $M \times N$ space $\mathcal{S}$ whose elements are pairs of real numbers from the set $[0, M] \times [0, N]$. An observation is any member of $\mathcal{S}$. We use $\mathcal{O}$ to denote an arbitrary, but fixed, finite set of *observations*. We assume there are real numbers $\alpha \leq \beta$ such that for each observation o , there exists a partner p_o (to be found) whose distance from o is in the interval $[\alpha, \beta]$.[2] Without loss of generality, we also assume that all elements of $\mathcal{O}$ are over β distance away from the edge of $\mathcal{S}$.

Throughout this paper, we assume the existence of a *distance function* d on $\mathcal{S}$ satisfying the usual properties of such distance functions.[3] We now define a region and how they relate to the set of observations. Our intuition is simple - a region *explains* an observation if that region contains a partner point for that observation.

Region/Region Explanation

- A *region* r is a subset of $\mathcal{S}$ such that for any two points $(x, y), (x', y') \in r$, there is sequence a of line segments from (x, y) to (x', y') s.t. no line segment lies outside r.
- A region r **explains** point o in $\mathcal{S}$ iff there exists a point $p \in r$ such that $d(o, p) \in [\alpha, \beta]$.

Note that regions can have any shape and may overlap. Throughout this paper, we assume that checking if some point o is explained by region r can be performed in constant (i.e. $O(1)$) time. This is a reasonable assumption for most regular shaped regions like circles, ellipses and polygons.

Region Explanation Problem (REP)

INPUT: Given a space $\mathcal{S}$, distance interval $[\alpha, \beta]$, set $\mathcal{O}$ of observations, set R of regions, and natural number $k \in [1, |\mathcal{O}|]$.
OUTPUT: Set $R' \subseteq R$, where $|R'| \leq k$ and for each $o \in \mathcal{O}$, there is an $r \in R'$ s.t. r sub-(super-) explains o.

(Shakarian and Subrahmanian 2010), showed this decision problem to be strongly NP-complete, meaning that the optimization version (that seeks to find an explanation of minimal cardinality) cannot be approximated by a fully polynomial-time approximation algorithm unless P==NP. However, the problem also reduces to an instance of set-cover, which means that a solution can be obtained within a reasonable approximation factor $(1 - \lg(f)$, where f is the maximum number of regions associated with any given observation). We have included the algorithm, GREEDY-REP-MC2 from that paper.[4]

```
GREEDY-REP-MC2
INPUT: Set $\mathcal{O}$ of observations, set $R$ of regions
OUTPUT: $R' \subseteq R$
1. Let $\mathbf{O} = \bigcup_{r \in R} \{\mathcal{O}_r\}$
2. For each observation $o \in \mathcal{O}$, let $GRP_o = \{\mathcal{O}_r \in \mathbf{O} | o \in \mathcal{O}_r\}$
3. For each observation $o \in \mathcal{O}$, let $REL_o = \{o' \in \mathcal{O} | o' \in \bigcup_{\mathcal{O}_r \in GRP_o} \mathcal{O}_r\}$ and let $key_o = |REL_o|$
4. Let $\mathcal{O}' = \mathcal{O}$, set $R' = \emptyset$
5. While not $\mathcal{O}' \equiv \emptyset$ loop
   (a) Let $o$ be the element of $\mathcal{O}$ where $key_o$ is minimal.
   (b) Let the element $\mathcal{O}_r$ be the member of $GRP_o$ s.t. $|\mathcal{O}_r \cap \mathcal{O}'|$ is maximized.
   (c) If there are more than one set $\mathcal{O}_r$ that meet the criteria of line 5b, pick the set w. the greatest cardinality.
   (d) $R' = R' \cup r$
   (e) For each $o' \in \mathcal{O}_r \cap \mathcal{O}'$, do the following:
       i. $\mathcal{O}' = \mathcal{O}' - o'$
       ii. For each $o'' \in \mathcal{O}' \cap REL_{o'}$, $key_{o''} - -$
6. Return $R'$
```

Adaptations for Afghanistan. There are two parts of the formalism of region-based abduction that are generally defined – the distance function (d) and the set of regions (R). In the experiments of (Shakarian and Subrahmanian 2010), we used a Euclidean distance function and we generated the regions from the REGION-GEN algorithm of that paper, which discretizes the entire space (hence, making it impractical for use here). Hence, we use d and R as a way to adapt region-based abduction to our Afghanistan scenario and address each of the three concerns outlined in the introduction. Our strategy is to build a special distance function, d_{afgh}, and use this function and the set of observations, $\mathcal{O}$, to generate R.

[2] (Shakarian, Subrahmanian, and Sapino 2010) describes methods to learn α, β automatically from historical data.

[3] $d(x, x) = 0; d(x, y) = d(y, x); d(x, y) + d(y, z) \geq d(x, z)$.

[4] Due to lack of space, we have omitted an example illustrating how this algorithm works. We refer the reader to (Shakarian and Subrahmanian 2010) for such examples.

To address the first concern, that of multiple tribes, assume we have a set of tribes, $\mathcal{T} = \{\mathbf{t}_1, \ldots, \mathbf{t}_m\}$. Based on our data set, we can assume we have the following function $tribes : \mathcal{S} \rightarrow 2^{\mathcal{T}}$ which takes a point in the space and returns a set of tribes. Two points in the space, p_1, p_2, are *tribally-related* iff $tribes(p_1) \cap tribes(p_2) \not\equiv \emptyset$. When we create our distance function, we will do so in a way to enforce this as an additional criterion that there must be at least one tribe that has a presence in the observation and partner location. The idea here is that an HVT must have a tribal-relationship with the lower-level cell conducting the attack, otherwise the two groups may not have a confluence of interest.

To address the second and third concerns, we appeal to the idea that the road networks of Afghanistan binds parts of this varied country together. Such sentiments are echoed in other work such as (Conover 2010). So, for any two villages on the road network (RN, an undirected graph where the vertices are villages) of Afghanistan, we define the function $sp_{RN} : \mathcal{S} \times \mathcal{S} \rightarrow \Re$ to return the shortest distance on the Afghan road network between the two points. Using shortest path on a road network is also useful as our attack and HVT data were all geolocated by village. Hence, we put these concepts together to create our new distance function, d_{afgh}, defined below.

$$d_{afgh}(p, p') = \begin{cases} sp_{RN}(p, p') & \textit{iff } tribes(p) \cap tribes(p') \not\equiv \emptyset \\ \infty & \textit{otherwise} \end{cases}$$

We use this function to generate regions via the algorithm REGION-GEN-AFGH - presented for the first time in this paper. A practical improvement we introduced was in determining the set V_o for each observation. We first determined the set $V_o^{(Euc)}$ V_o computed with a Euclidean distance function on the interval $[0, \beta]$ - as the Euclidean distance function can be calculated much faster than shortest-path. From this set, V_o is determined. It should be noted that the algorithm runs with a complexity $O(K \cdot |\mathcal{O}| \cdot T(RN))$ where K is a constant bound on the number of partners distance β away from a given observation and $T(RN)$ is the time complexity to find the shortest path between two points in RN. Another practical extension we added was to the output of GREEDY-REP-MC2. Any returned region over 1000 sq km was not included in the output. Our intuition here is that a region so large is not useful to an analyst attempting to cue other intelligence assets.

Afghanistan Data Set

Our data-set consisted of HVTs and attack data from the Afghan provinces of Hilmand and Kandahar from January - June 2010 supported by tribal and road network information. Below we provide details of the data-set.

Provincial Data. All provincial data, including boundaries of provinces and districts, road networks, and village locations were provided by (Afghanistan Information Management Services (AIMS)). We considered the Hilmand and Kandahar provinces, which consist of 29 districts. The road-network ($RN = (V, E)$) is an undirected graph of $30,304$ vertices (1604 of which are identified as villages) and $61,064$ edges.

```
REGION-GEN-AFGH
INPUT: Space $\mathcal{S}$, observations $\mathcal{O}$, reals $\alpha, \beta$
OUTPUT: Set of regions $R$
1. Let the road-network, $RN = (V, E)$
2. For each $o \in \mathcal{O}$, find the set $V_o = \{v \in V | d_{afgh}(o, v) \in [\alpha, \beta]$
3. Let $L = \bigcup_{o \in \mathcal{O}} V_o$. For each $p \in L$ let $\mathcal{O}_p$ be the set of observations that can be associated with it.
4. Partition $L$ into subsets, denoted $L_{\mathcal{O}'}$, where $\mathcal{O}' \subseteq \mathcal{O}$ and $p \in L_{\mathcal{O}'}$ iff $\mathcal{O}_p \equiv \mathcal{O}'$.
5. For each $L_{\mathcal{O}'}$, create region $r$ that is the minimum-enclosing rectangle of all elements in $L_{\mathcal{O}'}$. Add $r$ to $R$.
6. Return set $R$.
```

Attack Data. We used a series of 203 attack events in Afghanistan from the (National Counter-Terrorism Center (NCTC)). 103 of these events were from January-April 2010 and used to learn the $[\alpha, \beta]$ distance constraints, while the remaining 100 attacks (May-June 2010) were used as set $\mathcal{O}$ of observations. We actually divided the set of observations into 12 subsets, $\mathcal{O}_1 \subseteq \mathcal{O}_2 \subseteq \ldots \subseteq \mathcal{O}_{12}$, with each subsequent set of observations containing 5 days more worth of attacks than the previous (i.e. $\mathcal{O}_1$ was May 2-6 and $\mathcal{O}_2$ was May 2-11). All attacks in the WITS database were identified by village – corresponding with the AIMS information described earlier.

HVT Data. We collected a total of 78 HVTs based on official reports from (International Security Assistance Force (ISAF) Afghanistan). These reports spanned January-September 2010. We used the reports of January-April 2010 (27 HVTs) to learn the $[\alpha, \beta]$ distance constraints and the remainder for a ground-truth comparison (notice, this time interval is greater than that used for the set of observations, as an associated HVT with an attack may not necessarily have been located in the same time window described earlier). As with the attack data, each HVT was geo-located by the ISAF report with a village, which corresponded to the AIMS information. We manually identified only certain weapons caches and captured/killed enemy personnel as HVTs. Below we present our criteria in Figure 1 - it is based on the combat experience of one of the authors.

Tribal Data. To create the $tribes$ function, we used the tribal data from (Naval Postgraduate School (NPS)) that associated districts in Afghanistan with a set of tribes. All together, there were 23 tribes reported by the NPS.

Distance Constraints. Using the simple algorithm FIND-BOUNDS of (Shakarian, Subrahmanian, and Sapino 2010), which essentially returns an upper and lower distance bound on the shortest distance to an HVT given a set of attacks, determined the $[\alpha, \beta]$ bounds to be $[0.0, 65.88]$ km based on the historical attack and HVT data from January-April 2010.

Experimental Results

Setup. Our implementation of SCARE-S2, runs on a Lenovo T400 ThinkPad laptop equipped with an Intel Core 2 Duo T9400 processor operating at 2.53 GHz and 4.0 GB of RAM. The computer was running Windows Vista 64-

- Cache HVTs:
  - The cache contains 3x or more mortar rounds
  - The cache contains mortar tubes
  - The cache contains 3x or more rockets
  - The cache contains 10x or more grenades
  - The cache contains 5x or more RPG launchers
  - The cache contains 20x or more RPG rounds
  - The cache contains 15x or more AK-47's (or other similar rifles)
  - The cache contains 3x or more land-mines
  - The cache contains ”rooms” full of communications equipment (or ”rooms” full of any type of equipment)
  - The cache contains a DsHK or any other anti-aircraft weapon (including any number of Stinger missiles)
  - The cache contains 5x or more RPK machine guns (or similar capable systems such as M60, M249, etc.)
  - The cache contains 5x or more sniper rifles (such as a Dragunov)
- Personnel HVTs:
  - Reported listed individual as an insurgent “commander”
  - Reported listed individual as an insurgent “sub-commander”
  - Reported listed individual as an insurgent “planner”

Figure 1: HVT criteria.

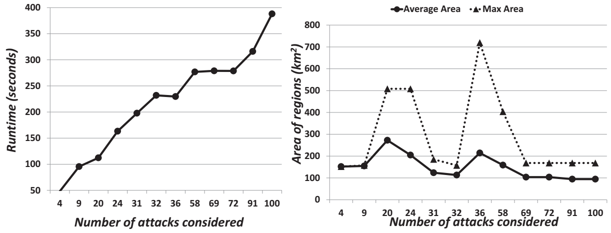

Figure 2: Number of attacks vs. runtime (average over 10 trials) and average region area.

bit Business edition with Service Pack 1 installed. This modest hardware setup was selected as deploying units to Afghanistan are typically equipped with Windows-based laptop systems. Isolated command posts, with limited connectivity to a network due to terrain restrictions may only have access to this limited computation power.

We implemented SCARE in approximately 4000 lines of Java code. Java Runtime Environment (JRE) Version 6 Update 14 was used. The software was developed with Eclipse version 3.4.0. We used the JGraphT library version 0.81 to implement the Fibonacci heap and the graph structure. Additionally, BBN OpenMap was used for some of the geospatial methods. We also added the capability to output KML files so that the results could be viewed in Google Earth - we used Google Earth 4.3.7284.3916 (beta) operating in DirectX mode. Experimental results were also collected in CSV-formatted spreadsheets.

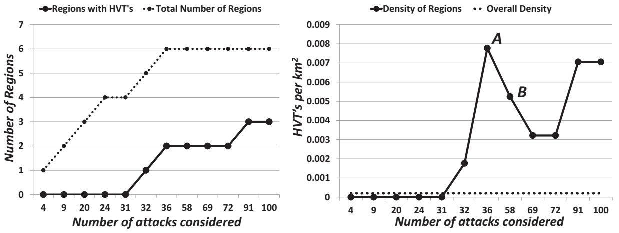

Figure 3: Number of attacks vs. number of regions and HVT density.

Runtime Experiments. We examined runtime of the algorithm by running the algorithm on each of the 12 subsets of observations described earlier. We observed two things: that the relationship between runtime and number of attacks was linear and that the runtime of REGION-GEN-AFGH dominated the runtime of GREEDY-REP-MC2 (which was negligible). This is primarily the result of the calculation of the shortest path. As stated earlier, this relationship is linear, so our result depicted in Figure 2 is unsurprising.

Area of Regions. As with (Shakarian and Subrahmanian 2010), we examine the average area of the regions. In general, smaller regions are preferred and as set $\mathcal{O}$ grows, the regions should become smaller. In each of the 12 trials, there was never more than one region over 200 sq km, and as set $\mathcal{O}$ increased, the average area approach 100 sq km – this is exactly what we are looking for. We plot the average and maximum areas in Figure 2. Note that a few spikes in average area are directly related to spikes in maximum area from a few outliers produced on some runs. Note that only a third of our runs produced a region over 200 sq km. Although even 100 sq km may seem like a large area, we must consider the density of villages - which is what we are attempting to locate. The overall density of villages for the entire area considered was 0.0064.[5] By the nature of how the regions are generated, they inherently have more villages. We observed that when we considered the entire set of attacks, no region contained more than 8 villages, with an average village density of 4.8 villages per region. As such is the case, we feel that the regions produced by SCARE-S2 will be helpful in directing intelligence, surveillance, and reconnaissance (ISR) assets.

HVTs Enclosed by Regions. In Figure 3, we plot the number of regions returned by each run, as well as the number of regions that enclose at least one HVT from the ground-truth set. Although the number of regions increase with the number of attacks (from 1 to 6), the number of regions enclosing an HVT also increase (from 0 to 3). While we should expect that solutions with more regions to enclose more HVTs, we must also recall that the regions become smaller with each run. Further, we also examined HVT density (number of HVTs divided by total area of all regions), which also

[5] In the newest version, SCARE-S2 also runs the geospatial abduction algorithm of (Shakarian, Subrahmanian, and Sapino 2010) which abduces points (villages, in this case). Hence, the output now not only included regions, but villages of interest as well - which allows us to further reduce the search-space for HVTs.

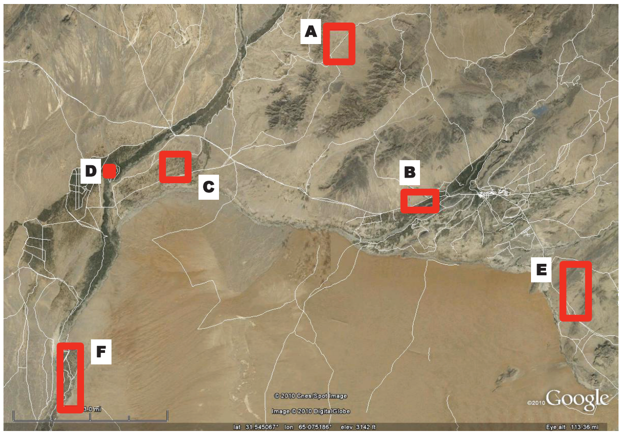

Figure 4: Regions returned after considering attacks from May-June 2010.

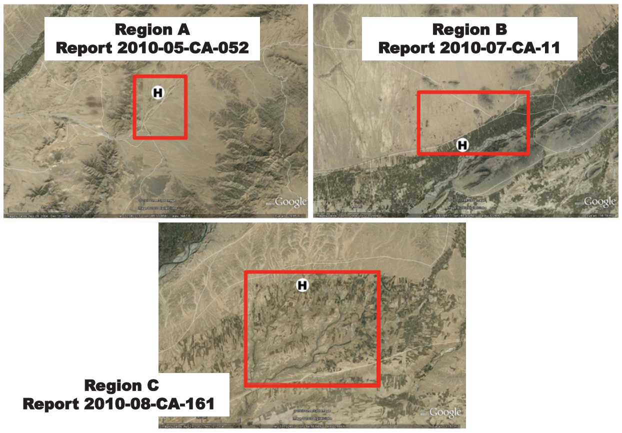

Figure 5: Close-up view of regions A-C with actual HVTs plotted.

increased with each run (note we had two outliers, identified in Figure 3 as points A and B. In these two runs, the software returned larger regions of size 719.68 sq km and 403.34 sq km that enclosed a large urban area where many HVTs were found. Eliminating these regions from the solution would eliminate these artificial spikes in density). When we considered the entire two months of observed attacks, the HVT density in the regions was over 35 times greater than the overall HVT density in the provinces. We remind the reader that the the regions are meant to be used as Named Areas of Interest (NAIs) for use by intelligence personnel. These NAIs would then be used to cue other intelligence assets (for example, a UAV or a HUMINT team) to conduct a more fine-grained search (hence, avoiding a search in a larger area). Therefore, despite only half the returned SCARE regions containing NAIs, the small size of the regions, along with the high density of HVTs, make them invaluable for the intelligence process.

Discussion. We shall now consider our final run of the algorithm, where we considered the entire set of 100 attacks from May-June 2010. This run produced the most regions enclosing HVTs, the greatest HVT density (discounting spikes A and B), and the smallest average region area.

This trial of the software produced 6 regions, labeled A-F, shown in Figure 4. Half of them enclosed an HVT. However, there were other ISAF reports that did not include village information. We did not consider these additional reports in any part of our experiments. However, all three regions returned by this experiment that did not enclose an HVT were located in districts where an HVT was reported (with no village information). For region D, there were 11 such reports, for region F, there were 4 such reports, and for district E there were was one such report. Let us now consider the HVTs found within regions A-C, depicted in Figure 5. Region A (with an area of 102.5 sq km) encloses the village of Bahram in the Ghorak district of Kandahar. According to ISAF PAO report 2010-05-CA-052, on May 5, 2010, a combined ISAF-Afghan force captured a Taliban commander in this village, who was responsible for several improvised explosive device (IED) attacks as well as movement of foreign fighters in the country. He also had a cache that included automatic rifles and heroin. Region B (with an area of 72.0 sq km) encloses the village of Makuan, in the Zhari district of Kandahar. According to ISAF PAO report 2010-07-CA-11, on July 18, 2010, a combined Afghan-ISAF force conducted a raid on a compound where a Taliban weapons facilitator was believed to reside. The unit received fire from insurgents, and returned fire killing several of them. As they approached the compound, they found several IED's placed to guard the facility. The compound was found to be a IED factory and a bunker system that contained munitions. Region C (with an area of 71.0 sq km) encloses the village Kharotan in the Nahri Sarraj district of Hilmand. ISAF PAO report 2010-08-CA-161 describes how ISAF forces detained the Taliban deputy-commander of the Lashkar Gah district there on August 14, 2010.

Related Work and Conclusions

Recently, there has been some work dealing with analytical and computational methods for reducing the IED threat in a counter-insurgency environment. (Marks 2009) uses dynamic programming scheme to determine optimal path on a network to conduct route-clearing operations. (Curtin 2009) explores the use of linear-referencing to associate IED events to certain parts of a road network. In (Li et al. 2009), the authors introduce the PITS system for predicting IED events based on geographic features and other non-geographic event (such as time). (Benigni and Furrer 2008) the authors look to quantify the IED threat of a given route at specific times of day. We would like to point out that all of this work deals with the either the prediction of IED attacks or avoiding potential locations of IED attacks – not locating HVTs (enemy personnel or logistics sites). Additionally, our search for HVTs is at a much larger scale – we are considering whole provinces of a very large area. Hence, the neutralization of the HVTs associated with an area of this size has a greater effect on the battlefield. To our knowledge, this paper introduces the first computational method for finding HVTs on the counter-insurgency battlefield. As stated above, this works build on the concept of *geospatial abduction* introduced in (Shakarian, Subrahmanian, and Sapino 2009; 2010; Shakarian and Subrahmanian 2010). However, none of these papers consider applying geospatial abduction to the Afghanistan scenario as presented here, or the special considerations already discussed. Geospatial abduction is a form of abductive inference, first introduced in (Peirce 1955). Two major existing theories of abduction include logic-based abduction (Eiter and Gottlob 1995) and set-covering abduction (Bylander et al. 1991). Geospatial abduction is related to set-covering abduction (which has been extensively explored in its application to medical diagnosis in (Peng and Reggia 1990)) as it reduces to an instance of set-cover. Some instances of other problems such as facility location and clustering can actually be encoded in a geospatial abduction, but a reduction in the opposite direction is not possible (see (Shakarian, Subrahmanian, and Sapino 2010) for a detailed discussion on this comparison).

In this paper we introduced a piece of software called "SCARE-S2" that applies geospatial abduction to the environment of Afghanistan. Unlike previous work, where we looked for small weapon caches supporting local attacks, here we looked for insurgent high-value targets (HVTs), supporting insurgent operations in two provinces. These HVTs included the locations of insurgent leaders and major supply depots. Applying this method of inference to Afghanistan introduced several practical issues not addressed in previous work. Namely, we are conducting inference in a much larger area ($24,940$ sq km as compared to 675 sq km in previous work), on more varied terrain, and must consider the influence of many local tribes. We address all of these problems and evaluate our software on 6 months of real-world counter-insurgency data. We show that we are able to abduce regions of a relatively small area (on average, under 100 sq km, containing, in average, 4.8 villages) that are more dense with HVTs ($35\times$ more than the overall area considered). There are other possible uses of geospatial abduction, including counter-drug, police, and naturalist uses. In our lab, we are also collecting data concerning illegal mining operations in Africa and are considering geospatial abduction as a possible tool to explore this international problem. Some of these have been described as examples in work such as (Shakarian, Subrahmanian, and Sapino 2010; Shakarian and Subrahmanian 2010), but not explored from an implementation standpoint. Such future studies would highlight other practical issues to consider when applying geospatial abduction to real-world problems, as was done in this paper for the Afghan scenario. One such practical extension we are considering is the use social network data to relate observations and partners (as we did in this work with tribal data), which could aide in predictions.

Acknowledgments

Paulo Shakarian is funded under the US Army ACS/West Point Instructor (EECS) program. Some of the authors of this paper were funded in part by AFOSR grant FA95500610405 and ARO grants W911NF0910206 and W911NF0910525.

References

Afghanistan Information Management Services (AIMS). GIS / Mapping Services, http://http://www.aims.org.af/.

Benigni, M., and Furrer, R. 2008. Periodic spatio-temporal improvised explosive device attack pattern analysis. Technical report, Golden, CO.

Bylander, T.; Allemang, D.; Tanner, M. C.; and Josephson, J. R. 1991. The Computational Complexity of Abduction.

Conover, T. 2010. *The Routes of Man: How Roads Are Changing the World and the Way We Live Today*. Knopf.

Curtin, K. 2009. Linear Referencing for Network Analysis of IED. In *AFCEA-GMU Symposium*.

Eiter, T., and Gottlob, G. 1995. The complexity of logic-based abduction. *J. ACM* 42(1):3–42.

International Security Assistance Force (ISAF) Afghanistan. Press Releases, http://www.isaf.nato.int/article/isaf-releases/index.php.

Li, H.; Muoz-Avila, H.; Bramsen, D.; Hogg, C.; and Alonso, R. 2009. Spatial event prediction by combining value function approximation and case-based reasoning. In McGinty, L., and Wilson, D., eds., *Case-Based Reasoning Research and Development*, volume 5650 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. 465–478.

Marks, C. 2009. Optimization-Based Routing and Scheduling of IED-Detection Assets in Contemporary Military Operations. Master's thesis, Massachusetts Institute of Technology.

National Counter-Terrorism Center (NCTC). Worldwide Incident Tracking System (WITS), https://wits.nctc.gov/.

Naval Postgraduate School (NPS). Program for Culture and Conflict Studies, http://www.nps.edu/programs/ccs/.

Peirce, C. S. 1955. *Philosophical writings of Peirce, selected and edited with an introd. by Justus Buchler*. Dover Publications New York,.

Peng, Y., and Reggia, J. A. 1990. *Abductive inference models for diagnostic problem-solving*. New York, NY, USA: Springer-Verlag New York, Inc.

Shakarian, P., and Subrahmanian, V. 2010. Region-based Geospatial Abduction with Counter-IED Applications. In Wiil, U. K., ed., *Counterterrorism and Open Source Intelligence (to appear)*. Springer.

Shakarian, P.; Subrahmanian, V.; and Sapino, M. L. 2009. SCARE: A Case Study with Baghdad. In *Proceedings of the Third International Conference on Computational Cultural Dynamics*. AAAI.

Shakarian, P.; Subrahmanian, V.; and Sapino, M. L. 2010. Gaps: Geospatial abduction problems. *ACM Transactions on Intelligent Systems and Technology (to appear)*.

US Army. 1994. *Intelligence Preparation of the Battlefiled (US Army Field Manual)*, FM 34-130 edition.