

Abductive Reasoning in 2D Geospatial Problems

Achilleas Koutsoumpas

Abstract Spatial Analysis has been using so far Spatial Reasoning, but it mainly confines itself to spatial statistical analysis of the observed phenomena searching for pattern analysis, geostatistical indices etc. Researchers Shakarian P., Subrahmanian V. S., and Sapino M. L. are the first who examined the possibility of extending Spatial Analysis in finite, discretized, 2D space with the incorporation of Abductive Reasoning, which originates from the cognitive field of Artificial Intelligence, and is related to the analysis of causation of the phenomena under consideration. The new class of Geospatial Problems was named point-based Geospatial Abduction Problems (or point-based GAPs). They primarily focused on a version of GAPs named Improvised Explosive Devices Cache Detection Problem, or IED Cache Detection Problem, and they carried out experiments with real-world data from Baghdad. In this paper a technique which reduces the total computational cost in any version of point-based GAPs will be introduced, and an exact algorithm for the natural optimization problem of point-based GAPs will be presented along with its computational complexity results.

Keywords Abductive reasoning • Geospatial abduction problems • Spatial problems • IED cache detection problem

¹This research has been conducted in the context of the writer's MSc Thesis, which was supervised by Professor Marinos Kavouras, in the field of Geoinformatics at the National Technological University of Athens (NTUA).

A. Koutsoumpas (✉)
Department of Mathematics and Engineering Sciences, Hellenic Military Academy, 166 73,
Vari, Greece
e-mail: axilleas_kouts@yahoo.com

Introduction

Spatial Analysis has been using so far Spatial Reasoning, but it mainly confines itself to spatial statistical analysis of the observed phenomena searching for pattern analysis, geostatistical indices etc. A trivial example is the spatial analysis of the crime phenomenon in a given region, which will result in finding subregions that are statistically characterized by high rates of crime. The most common strategy in response is the increase of policing in those subregions. However, that kind of strategy does not affect the cause of the phenomenon under consideration, i.e. the criminals, but only the observations of the phenomenon. Thus, the crime will either shift or spread out to the remaining space, as the increase of policing in a set of subregions implies the decrease in the rest of the remaining space, which in turn implies that criminals are offered now new opportunities to carry out their illegal activities in those exact under-policing subregions. On the contrary, the best strategy for law enforcement would be the use of a spatial relationship, if such exists, between the locations where the crimes were committed and the locations of the hideouts of the criminals which obviously would pin-point the latter, given the fact that the locations of crime incidents are known to the police, and the criminologists can distinguish the crime signatures, and thus allocate the crimes accordingly. That kind of reasoning can be applied in a wide variety of real-world problems named Geospatial Abduction Problems or GAPs such as the detection of terrorist caches given the location of their attacks, the detection of dwellings of a wild animal given the location of its attacks, the detection of environmental pollutants, etc. [13]. Informally, a GAP is a problem in geospatial space of finding a set of partner locations that best “explain” a given set of locations of observations through a given spatial relationship. The spatial relationship for any GAP, even for instances of the same problem, is unique as it expresses the correlation between partner and observations in the context of a specific domain, in a specific geographical space, and for a specific period of time.

The complexity of human cognition and the difficulty of its simulation by mechanic means are proven by the fact that since the Aristotelian logic was formulated till nowadays various inference rules and kinds of reasoning have been proposed. Informally, reasoning defines a rough theoretical context in which valid inference rules can be applied, whereas inference rules are formally defined mechanisms, algorithms, which produce new valid knowledge [16], or in other words, logical transformations used to produce conclusions from a premise set. Some of the most common inference rules are:

- The Modus Ponens rule which is used in Mathematical Logic, where if one formula implies another one and the first formula is TRUE, then the second one is also TRUE. Formally: $A \rightarrow B, A \vdash B$.
- The Modus Tollens rule, which is also used in Mathematical Logic, where if one formula implies another one and the latter formula is FALSE, then the first formula is also FALSE. Formally: $A \rightarrow B, \neg B \vdash \neg A$.
- The generalized Modus Ponens and Modus Tollens inference rules in the domain of Fuzzy Logic.

- The Generalization rule, where if an attribute is TRUE for some instances of a class, then the attribute is TRUE for any instance of the class.
- The Specification rule, where if an attribute is TRUE for a class, then the attribute is TRUE for any instance of the class.

On the other hand, there are three fundamental kinds of reasoning:

- Deductive Reasoning
- Inductive Reasoning
- Abductive Reasoning

More specifically, Deductive Reasoning infers conclusions of what is assumed, thus given the truth of the assumptions, the conclusion is guaranteed to be TRUE. Assumptions are given in the form of a rule and a fact. For example:

<Rule> All my children are boys.

<Fact> This child is mine.

<Conclusion> This child is a boy.

Inductive Reasoning infers a rule from a set of facts. Even if the set of facts are TRUE and the derived rule seems to be reasonable, this does not ensure the soundness of the rule. It must be noted that this kind of reasoning has nothing in common with the mathematical induction since the latter is a form of Deductive Reasoning. For example:

<Fact 1> This boy is my child.

<Fact 2> This boy is also my child.

<Rule> All my children are boys.

Abductive Reasoning infers a hypothesis given a set of rules, which consists a knowledge basis, and a set of facts, that consists a set of observations. For example:

<Rule> All my children are boys.

<Observation> This child is a boy.

<Hypothesis> This child is mine.

Obviously, there is no such thing as “best” reasoning, since the above kinds of reasoning are not competing against each other. On the contrary, they are complementary since each one’s start point differs, thus unavoidably each one infers a different kind of output.

Despite the fact that the term Abductive Reasoning was proposed by the great American philosopher Peirce in 1890 [6], in order to describe the kind of reasoning used for evaluating explanatory hypotheses for a given set of observations, philosophers and scientists were familiar with the context of the term from the Renaissance at least [1]. Some thinkers have been skeptical that a hypothesis should be accepted merely, but others have argued that this kind of reasoning is a legitimate

part of scientific reasoning [15]. Clearly any feasible explanatory hypothesis, or simply explanation, even if it is supposed to be the inference to the best explanation, is fraught with uncertainty [15] due to the fact that abduction is equivalent to the logical fallacy affirming the consequent, or else in Latin “post hoc ergo propter hoc.” Nevertheless, Abductive Reasoning plays a key role in human cognition. For example when: scientists produce new theories to explain their data, criminologists compose the evidence from criminal activities, mechanics try to find out what problem is responsible for a mechanical breakdown, doctors try to infer which disease explains a patient’s symptom, even when people generate hypotheses to explain the behavior of others [15], they all use abduction. What is left is the formal semantic definition of the term explanation. In philosophy and cognitive science, there have been at least six approaches; explanations have been viewed as deductive arguments, statistical relations, schema applications, analogical comparisons, causal relations and finally, linguistic acts [14]. Geospatial Abduction uses the approach of causal relation in the context of a specific domain. In other words, Geospatial Abduction is used to infer as the best explanation in a specific domain the relation between the locations of the observed phenomena and the partner locations which cause, facilitate, support or are somehow correlated with the observations [13].

Formalism-Technical Preliminaries

To address the Geospatial Abduction for the new class of problems named point-based GAPs, researchers Shakarian, Subrahmanian and Sapino proposed an appropriate formalism [8, 12, 13], which is adopted in this paper, as follows:

Definition 1 (Space). Geospatial universe or universe or simply space S is a finite two-dimensional grid of size $M \times N \in \mathbb{N}^2$, where $M, N \geq 1$ and \mathbb{N} is the set of natural numbers.

Definition 2 (Point). A point p of a finite 2D space S is a representation of a unit square on the grid in the form (x, y) where $x, y \in \mathbb{N}$. Therefore, $S = \{x \mid x \in [0, M)\} \times \{y \mid y \in [0, N)\}$.

In addition, conventionally the coordinates (x, y) in each such unit cell are not assigned to its center, but to its lower left corner, and the point with coordinates $(0,0)$ is the lower left point of the grid. Thus, the space is used to represent a specific region of interest on the ground and the size of the point in ground distance units defines the space resolution. The suitable extents of the space are normally easily defined by the GAP application itself whereas the resolution of the space is defined “arbitrarily” by the application developer or the analyst.

Definition 3 (Observation—Set of Observations). An observation o is a point of a finite 2D space S where a phenomenon under consideration appeared. A set of observations $O = \{o_1, o_2 \dots o_k\}$ is any finite subset of space S where each element o_i is an observation of the phenomenon under consideration.

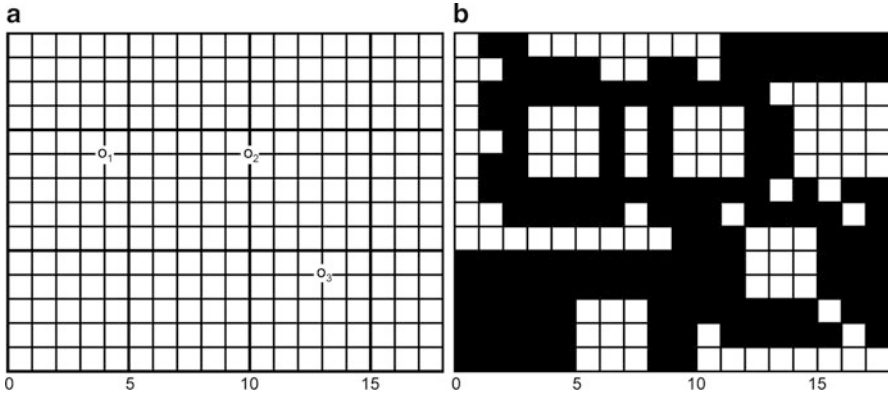


Fig. 1 Examples of a set of observations and a feasibility predicate in space. (a) A set of observations $O = \{o_1, o_2, o_3\}$ in space S . (b) A feasibility predicate in space S

Figure 1a illustrates the above concepts in a space 18×14 . Furthermore, any space S has an associated distance function d that satisfies the following definition.

Definition 4 (Distance Function). A distance function d in metric space (S, d) is a mapping $S \times S \rightarrow \mathbb{R}$, where \mathbb{R} is the set of real numbers, such that for any $x, y \in S$ the following axioms are satisfied:

- $d(x, x) = 0$, coincidence axiom.
- $d(x, y) = d(y, x)$, symmetry axiom.
- $d(x, y) + d(y, z) \geq d(x, z)$, triangle inequality axiom.

There are numerous distance functions that satisfy the above three axioms. The more familiar are the Euclidean distance, where $d_e = [(x_1 - x_2)^2 + (y_1 - y_2)^2]^{1/2}$; the Manhattan distance, where $d_m = |x_1 - x_2| + |y_1 - y_2|$, which are special cases of the Minkowski or L_p -metrics [7]; and the road distance $d_r((x_1, y_1), (x_2, y_2))$ that is defined as the shortest path between two points in a road network.

Definition 5 (Feasibility Predicate). A feasibility predicate $feas$ in a finite 2D space S is a function $feas : S \rightarrow [TRUE, FALSE]$ such that $feas(p) = TRUE$ for any point $p \in S$ if and only if (iff) that point can be included in a feasible solution of the GAP or $feas(p) = FALSE$ otherwise.

It is noted that the feasibility predicate $feas$ is an arbitrary, yet a fixed predicate. It is an arbitrary predicate in the sense that both the semantics for feasibility and its spatial assignment in space are assessed by the analyst. This means that for any space S there are offered 2^n distinct alternative process choices for an analyst, given that $|S| = n$. Thus, two analysts, A and B, even for the same geospatial problem in the space S , will most likely come up to $feas_A \neq feas_B$. At the same time it is a fixed predicate in the mathematical sense, since for any point $p \in S$ is either $feas(p) = TRUE$ or $feas(p) = FALSE$. In terms of computational complexity,

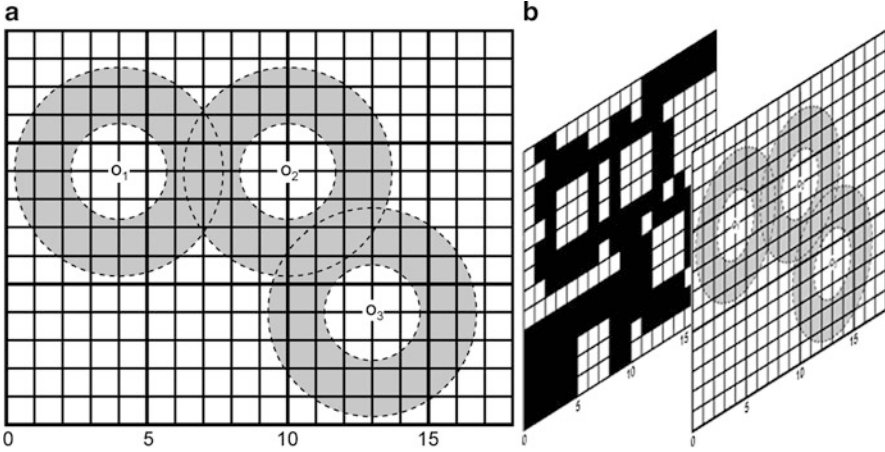


Fig. 2 Examples of geometric restrictions and their logical composition with feasibility predicate. **(a)** Geometric restrictions form $|O|$ concentric rings in space S . **(b)** The (α, β) -explanation is a logical composition in space S

since it is user-defined, it is assumed that it has an $O(1)$ complexity cost. Figure 1b above illustrates in color a feasibility predicate in space S , white indicates that $feas(p) = TRUE$ and black indicates that $feas(p) = FALSE$.

At this point we can define the concept of an explanation in the framework of geospatial abduction. Intuitively, given a set of observations an explanation is a set of points in space such that every point is feasible and such that for every observation there is a point in explanation that satisfies certain geometric, or distance, restrictions. The geometric restrictions, which are set by the analyst or the domain expert, define the minimum and the maximum allowed metric distances between the location of an observation and its partner location. Formally:

Definition 6 ((α, β)-Explanation). Let O be a set of observations in a 2D finite space S , E be a finite set of points where $E \subset S$ and $\alpha, \beta \in \mathbb{R}$ such that $0 \leq \alpha < \beta$, then E is an (α, β) -explanation of O iff:

$$[\forall p \in E | feas(p) = TRUE] \wedge [\forall o \in O \exists p \in E | \alpha \leq d(p, o) \leq \beta]$$

In Fig. 2 are illustrated, for the running example, the geometric restrictions in space (Fig. 2a) and the logical composition of the geometric restrictions and the feasibility predicate (Fig. 2b), given that the size of the grid cell is $500m$, $\alpha = 850m$ and $\beta = 1850m$. The composition produces the set of partner locations for the given set of observations. Any subset of this set that explains all the observations is a possible solution.

The simplest form of point-based GAPs is the decision problem named the Simple (α, β) -Explanation Problem (SEP), where given a finite 2D space S , a finite set of observations O , a feasibility predicate $feas$ and $\alpha, \beta \in \mathbb{R}$, such that

$0 \leq \alpha < \beta$, the question is if there is an (α, β) -explanation for O . Another variant of GAPs is the k-SEP problem, which defines an upper bound in the cardinality of any possible solution. More specifically, it is required that $|E| \leq k$ for a number $k \in \mathbb{N}$, given that $k < |O|$.

Generally, the GAPs are highly non-trivial to solve for four reasons [13]. First of all, the cardinality of the solution is unknown. Secondly, the feasibility predicate defines highly irregular in shape areas for partner locations, thus simple geometric reasoning cannot be applied. Thirdly, in case that the phenomenon under consideration involves a rational adversary who will take any action to evade detection (in terms of Game Theory this is an adaptive adversary), then a simple algorithmic solution would not be sufficient and more sophisticated algorithms must be applied. Finally, the most GAPs are proven to be NP-complete, in terms of computational complexity, or in other words intractable to compute in practice.

Since there are many possible explanations for a given set of observations the truly intriguing variants of point-based GAPs are the ones which try to find an “optimal” explanation according to some cost function χ , where a cost function is a mapping from the set of explanations to non-negative reals.

Definition 7 (Optimal (α, β) -Explanation). Suppose S is a finite 2D space, $O \subset S$ is a finite set of observations, $E \subset S$ is a finite set of points, $\alpha, \beta \in \mathbb{R}$ such that $0 \leq \alpha < \beta$ and χ is a cost function. Then E is said to be an optimal (α, β) -explanation iff E is an (α, β) -explanation for O , and there is no other (α, β) -explanation E' for O such that $\chi(E') < \chi(E)$.

One can easily prove that standard classification problems like k-means [4, 5] are a special case of GAPs by making three simple assumptions: a) $\alpha = 0$, b) $\beta > \max(M, N)$ and c) $\exists p \in S \mid feas(p) = FALSE$ [13].

Normally in philosophy and in science additional requirements are enforced to an explanation, the parsimony requirements. In SEP no such requirement is enforced, since any explanation will suffice. On the other hand, k-SEP and cost-based SEP enforce parsimony requirements. Another parsimony requirement, which expresses the famous principle known as the Occam’s razor: “Pluritas non est ponenda sine necessitate,” is irredundancy. In Geospatial Abduction this means that between two explanations the simplest one must be chosen. The formal definition follows:

Definition 8 (Irredundant (α, β) -Explanation). An (α, β) -explanation E is irredundant iff no strict subset of E is an (α, β) -explanation.

One may think that by enforcing the irredundancy requirement in GAPs the uniqueness of the solution is guaranteed, but this is far from being true. A closer look easily shows that there still exist an exponential number of such solutions, thus an algorithm that simply returns irredundant solutions could produce non-deterministic results [13]. The best one can do is to attempt to find explanations of minimal

cardinality. This requirement poses the Minimal Simple (α, β) -explanation Problem (MINSEP) which is the natural optimization problem associated with k-SEP [8]. Formally:

Definition 9 (MINSEP).

INPUT: a finite 2D space S , a finite set of observations O , a feasibility predicate $feas$ and $\alpha, \beta \in \mathbb{R}$ such that $0 \leq \alpha < \beta$.
 OUTPUT: An (α, β) -explanation E such that $|E| = \text{minimal}$.

Algorithms

In this chapter a technique which reduces the total computational cost in any version of GAPs will be proposed, and an exact algorithm for the natural optimization problem of GAPs, the MINSEP, will be presented.

Even though mathematically for point-based GAPs, it is sufficient to require for the relation between α, β and space S that $\alpha, \beta < \inf\{M, N\}$, this obviously does not maximize the utility of the set of observations. In order to accomplish that the requirement for the relation between α, β and space S has to be rewritten as follows:

$$\forall p(x, y) \in O \text{ then } M, N \mid [x + \beta] \leq M \wedge [y + \beta] \leq N \wedge [x - \beta] \geq 0 \wedge [y - \beta] \geq 0$$

But in reality, in most applications, the analyst faces the exact opposite problem; the size of the given space is bigger than the necessary. Thus, in order to succeed effective computation, space must somehow be reduced to the exact size needed. A technique that reduces the given space S to the absolutely necessary space S' is: given that $x_{\min}, x_{\max}, y_{\min}$ and y_{\max} are the minimum and the maximum values of coordinates for the observation set, then it is easy to show that any explanation, with respect to the geometric restrictions, is strictly in the space S' which equals to the Minimum Bounding Rectangular box (MBR) with coordinates in space S :

$$(X_{\min}, Y_{\min}) = (\lceil x_{\min} - \beta \rceil, \lceil y_{\min} - \beta \rceil)$$

$$(X_{\max}, Y_{\max}) = (\lfloor x_{\max} + \beta \rfloor, \lfloor y_{\max} + \beta \rfloor)$$

The reduction is completed with the parallel shift of the coordination system to the point with coordinates $(\lceil x_{\min} - \beta \rceil, \lceil y_{\min} - \beta \rceil)$, in order for space S' to have coordinates $(0,0)$ in its lower left point of the grid. One can easily show that the size of space $S' = M' \times N'$ is:

$$M' = (\lfloor x_{\max} + \beta \rfloor - \lceil x_{\min} - \beta \rceil), \text{ where } M' \in [1, M]$$

$$N' = (\lfloor y_{\max} + \beta \rfloor - \lceil y_{\min} - \beta \rceil), \text{ where } N' \in [1, N]$$

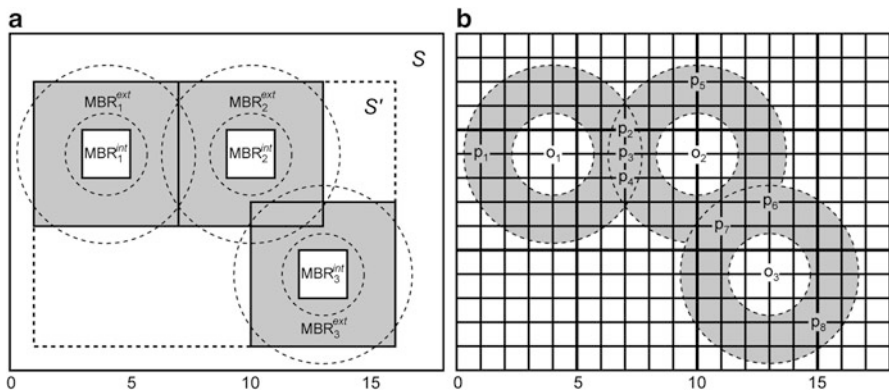


Fig. 3 Examples of the reduction for the given space, the MBRs for the set of observations and the set of partner locations. (a) The reduction of space S to space S' and the MBRs of the set of observations. (b) The set of partner locations for the (1.7, 3.7)-explanation

Practically, for the running example since $\beta = 1850/500 = 3.7$ metric units in space S , then the coordinates of space S' are calculated as follows and the result is illustrated in Fig. 3a.

$$(X_{\min}, Y_{\min}) = (\lceil 4 - 3.7 \rceil, \lceil 4 - 3.7 \rceil) = (\lceil 0.3 \rceil, \lceil 0.3 \rceil) = (1, 1)$$

$$(X_{\max}, Y_{\max}) = (\lfloor 13 + 3.7 \rfloor, \lfloor 9 + 3.7 \rfloor) = (\lfloor 16.7 \rfloor, \lfloor 12.7 \rfloor) = (16, 12)$$

In order to achieve more effective computation the use of MBRs can be applied in the set of observations as well. Instead of searching for partner locations generally in space S' for any observation o_i , the search can be focused in an appropriate subspace $S_i \subseteq S'$. Suppose that MBR_i^{ex} is an MBR which contains any point that is no more than β metric units away from the observation o_i and that MBR_i^{in} is an MBR which contains any point that is at least α' metric units away from the observation o_i , where $\alpha' \leq \alpha$ such that $\alpha' = \min d_e(o_i, p_j) \mid [d(o_i, p_j) = \alpha]$. Obviously, if the distance function used in space is the Euclidean distance d_e , then $\alpha' = \alpha$. Now, it is easy to see that space S_i is defined as $S_i = MBR_i^{\text{ex}} - MBR_i^{\text{in}}$. Trivially it can be proven that both MBR_i^{ex} and MBR_i^{in} are squares, where geometrically the first one equals to a square in which a circle of radius β is inscribed, while the latter equals to a square inscribed in a circle of radius α' . The MBR_i^{ex} in space S is defined by the coordinates:

$$(x_{\min(i)}, y_{\min(i)}) = (\lceil x_i - \beta \rceil, \lceil y_i - \beta \rceil)$$

$$(x_{\max(i)}, y_{\max(i)}) = (\lfloor x_i + \beta \rfloor, \lfloor y_i + \beta \rfloor)$$

Whereas the MBR_i^{in} in space S is defined by the coordinates:

$$(x_{\min(i)}, y_{\min(i)}) = \left(\left[x_i - \frac{\sqrt{2}}{2} \cdot \alpha' \right], \left[y_i - \frac{\sqrt{2}}{2} \cdot \alpha' \right] \right)$$

$$(x_{\max(i)}, y_{\max(i)}) = \left(\left[x_i + \frac{\sqrt{2}}{2} \cdot \alpha' \right], \left[y_i + \frac{\sqrt{2}}{2} \cdot \alpha' \right] \right)$$

For the running example, and if it is supposed that $d = d_e$, then it is easily computed that the MBR_1^{ex} is defined in space S by the pair of points $(x_{\min(1)}, y_{\min(1)}) = (1, 6)$ and $(x_{\max(1)}, y_{\max(1)}) = (7, 12)$. In the same way, the MBR_2^{ex} is defined by the pair $(7, 6)$ and $(13, 12)$, the MBR_3^{ex} by the pair $(10, 1)$ and $(16, 7)$, whereas MBR_1^{in} by the pair $(3, 8)$ and $(5, 10)$, the MBR_2^{in} by the pair $(9, 8)$ and $(11, 10)$, and finally the MBR_3^{in} by the pair of points $(12, 3)$ and $(14, 5)$. These results are illustrated in Fig. 3a. Additionally in Fig. 3b is illustrated the $(1.7, 3.7)$ -explanation of maximum cardinality, or in other words the set of partner locations, in space S for the running example.

Despite the fact that the SEP problem has a proven computational complexity in PTIME [8, 12, 13], its outputs are not reliable because any (α, β) -explanation E , independently of its cardinality, suffices. Thus, an algorithm for SEP would produce, with high probability, an explanation that is redundant. On the other hand, real-world data from Baghdad for the IED Cache Detection Problem, which were experimentally tested in order to show that the Geospatial Abduction framework is viable, proved that the cardinality of the solution plays a key role in the accuracy [8, 12, 13]. But shifting from a SEP problem to a k-SEP problem (or in other words the GAP variant which imposes cardinality constraints to the acceptable explanations) is not cost-free, since it automatically implies NP-completeness. So, since an exact solution to k-SEP takes exponential time (in k), it follows that any efficient algorithm would be nothing more than a fundamental trade-off between accuracy and computational cost. In the end, among the three approximation algorithms that were tested, the algorithm which was chosen to be implemented in SCARE (Social-Cultural Abductive Reasoning Engine) was not the fastest one, but the algorithm that consistently returned the solution of the smallest cardinality [8, 12, 13].

The algorithm NAIVE-MINSEP-EXACT below includes the technique that was presented above, which reduces the total computational cost, and provides an exact solution for MINSEP by leveraging the NAIVE-KSEP-EXACT algorithm [8].

It must be noted that the NAIVE-MINSEP-EXACT algorithm is flexible enough to include the domain's expert assessment, if such exists, of the expected cardinality of the solution, k . If this kind of information is unavailable or its use bears high risk, then a simple modification will be needed in the body of the algorithm. Just a simple replacement in the 5th and 6th step of the number k with the cardinality of the set of observations $|O|$ suffices. Furthermore, the search for the size of the minimal size explanation is performed by applying the technique "divide and conquer," which guarantees that no more than $\log_2 k$ (respectively $\log_2 |O|$) trials will be needed at the most [2].

Algorithm 1 NAIVE-MINSEP-EXACT

INPUT: Space S , set of observations O , feasibility predicate $feas$, distance function $d \in \mathbb{R}$, $\alpha, \beta \in \mathbb{R}$ such that $0 \leq \alpha < \beta$ and $k \in \mathbb{N}$.

OUTPUT: Set $E \subseteq S$ such that $|E| \leq k$ and $E = \text{minimal}$.

1. Define distance α' and space S' .
 2. Let M be a matrix array of pointers to binary string $\{0, 1\}^{|O|}$. M is of the same dimensions as S' .
Each element in M is initialized to NULL. For a given $p \in S'$, $M[p]$ is the place in the array.
 3. Let L be a list of pointers to binary strings. L is initialized as NULL.
 4. For each $o_i(x_i, y_i) \in O$ do the following:
 - a. Determine all points $p_j(x_j, y_j) \in S'$ such that $(x_j, y_j) \in [[x_i - \beta], [x_i + \beta]] \wedge [[y_i - \beta], [y_i + \beta]]$ and $(x_j, y_j) \notin [[x_i - \frac{\sqrt{2}}{2} \cdot \alpha'], [x_i + \frac{\sqrt{2}}{2} \cdot \alpha']] \wedge [[y_i - \frac{\sqrt{2}}{2} \cdot \alpha'], [y_i + \frac{\sqrt{2}}{2} \cdot \alpha']]$ and $feas(p_j) = TRUE$.
 - b. For each of these points $p_j(x_j, y_j)$:
If $M[p] = NULL$, then
 - (1) Initialize a new array where only bit i is set to 1.
 - (2) Add a pointer to $M[p]$ in L .
 Else, set bit i of the existing array to 1.
 5. Let $n \in \mathbb{N}^*$ such that $n \leq k$. For any n elements of L (actually the n elements pointed to by elements of L), we shall designate $l_1, l_2 \dots l_j \dots l_n$ as the elements. The i -th bit of element $l_j(i)$ will be referred as $l_j(i)$.
 6. Exhaustively generate all possible combinations of n elements, for every $n \leq k$, starting from $n = \frac{k}{2}$ and by applying the technique “divide and conquer” until one such combination of minimal n is found where $\forall i \in [1, |O|] \sum_{j=1}^n (l_j(i)) > 0$.
 7. If no such combination is found, then
return NO
Else, return the first combination that was found.
-

Complexity

In this section the computational complexity of the NAIVE-MINSEP-EXACT algorithm is studied by leveraging the complexity results of the NAIVE-KSEP-EXACT algorithm [8, 12, 13]. It is noted that k-SEP has a proven NP-completeness [13] via a reduction from the well known NP-complete problem Geometric Covering by Discs (GCD) [3].

Proposition. *The complexity of the NAIVE-MINSEP-EXACT algorithm is:*

$$O\left(\frac{\log_2 k}{(k-1)!} \cdot (\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|)^{(k+1)}\right)$$

Proof. Obviously, the computations $S', \alpha' \in \text{PTIME}$. In addition the cost to set up the matrix array of pointers M is $O(1)$ and not the size of the matrix, since all the pointers in M are initially null, thus there is no need to iterate every element in M , and only lists in M can be initialized as needed. \square

Since each observation $o_i \in O$ has at most $\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil$ partner points, which are located in space $S_i = MBR_i^{\text{ex}} - MBR_i^{\text{in}}$, then list L has at most $\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|$ elements. Thus, there are $\binom{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|}{n}$ iterations taking place at most in the sixth step, which are as many as the n -tuples in the list L . Each iteration has a complexity cost of $n \cdot |O|$, as there must be compared $|O|$ bits of each n -bit string. Hence, the combined cost of computations equals to: $\binom{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|}{n} \cdot n \cdot |O|$.

Additionally, since we are looking for the minimal $n \in \mathbb{N}^*$, where $n \leq k$, and we use the “divide and conquer” technique we will need at the most $\log_2 k$ trials, thus $\log_2 k$ times the combined cost of computations. It is noted that if we have not adopted the “divide and conquer” technique, then the computational cost would be $\sum_{n=1}^k \binom{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|}{n} \cdot n \cdot |O|$. Next, we are searching for an upper bound for the computational cost of the algorithm’s 6th step.

Lemma. $\forall n \in \mathbb{N}^*$, where $n < k$, can be proven that:

$$\binom{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|}{n} \cdot n \cdot |O| < \binom{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|}{k} \cdot k \cdot |O|$$

Proof. Let $A = \lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|$, then it must be proven that for any $n < k$:

$$\begin{aligned} \binom{A}{n} \cdot n \cdot |O| &< \binom{A}{k} \cdot k \cdot |O| \Leftrightarrow \\ \frac{A!}{n! \cdot (A-n)!} \cdot n \cdot |O| &< \frac{A!}{k! \cdot (A-k)!} \cdot k \cdot |O| \Leftrightarrow \\ \frac{1}{(n-1)! \cdot (A-n)!} &< \frac{1}{(k-1)! \cdot (A-k)!} \Leftrightarrow \\ (k-1)! \cdot (A-k)! &< (n-1)! \cdot (A-n)! \end{aligned} \quad (1)$$

Since $n < k$, where $n \in \mathbb{N}^*$, then there exists $t \in \mathbb{N}^*$ such that:

$$n = k - t \quad (2)$$

From (1), (2) \Rightarrow

$$(k-1)! \cdot (A-k)! < (k-(t+1))! \cdot (A-(k-t))! \Leftrightarrow$$

$$(k-(t+1))! \cdot (k-(t+1)+1) \cdot (k-(t+1)+2) \dots \cdot (k-2) \cdot (k-1) \cdot (A-k)!$$

$$\begin{aligned}
 &< (k - (t + 1))! \cdot (A - k)! \cdot (A - k + 1) \cdot (A - k + 2) \dots \cdot (A - (k - t)) \Leftrightarrow \\
 &(k - t) \cdot (k - t + 1) \dots \cdot (k - 2) \cdot (k - 1) < (A - k + 1) \cdot (A - k + 2) \dots \cdot (A - (k - t)) \quad (3)
 \end{aligned}$$

In the products of the inequality $k - 1$ terms, which are naturals, appear at the most. Obviously, the inequality holds if each term of the right product is larger from the corresponding term of the left product, or that each one of the following inequalities holds:

$$k - t < A - k + 1 \tag{4.1}$$

$$k - t + 1 < A - k + 2 \Leftrightarrow k - t < A - k + 1 \tag{4.2}$$

.....

$$k - 2 < A - k + t - 1 \Leftrightarrow k - t < A - k + 1 \tag{4.k-2}$$

$$k - 1 < A - k + t \Leftrightarrow k - t < A - k + 1 \tag{4.k-1}$$

It is observed that each one of the above inequalities (4.1)...(4.k-1) collapses to the same expression:

$$2 \cdot k < A + t + 1 \tag{5}$$

Hence, since $(t + 1) \in \mathbb{N}^*$ and $(t + 1) \geq 2$ (directly from the definition of t) it suffices, from (5), to prove that for any $n < k$, where $n \in \mathbb{N}^*$:

$$\begin{aligned}
 2 \cdot k < A &\Leftrightarrow 2 \cdot k < \lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O| \Leftrightarrow \\
 k < \frac{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil}{2} \cdot |O| &\tag{6}
 \end{aligned}$$

But

$$\frac{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil}{2} > \frac{\pi}{2} > 1 \tag{7}$$

And

$$k < |O| \tag{8}$$

From (7),(8) the inequality in (6) holds. Thus, it is proven that for every $n \in \mathbb{N}^*$, where $n < k$, the following inequality holds:

$$\left(\frac{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|}{n} \right) \cdot n \cdot |O| < \left(\frac{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|}{k} \right) \cdot k \cdot |O|$$

□

Now, the proof of the proposition can be continued. Since the above lemma is proven, then trivially it can be proven that $\forall n \in \mathbb{N}^*$, where $n \leq k$, the following inequality stands:

$$\binom{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|}{n} \cdot n \cdot |O| \leq \binom{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|}{k} \cdot k \cdot |O|$$

Thus, an upper bound for the computational cost of the algorithm's 6th step is:

$$\begin{aligned} \log_2 k \cdot \binom{\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|}{k} \cdot k \cdot |O| &= \log_2 k \cdot \frac{(\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|)!}{k! \cdot (\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O| - k)!} \cdot k \cdot |O| \\ &= \log_2 k \cdot \frac{(\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O| - k)! \cdot (\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O| - (k-1)) \dots (\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|)}{(k-1)! \cdot (\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O| - k)!} \cdot |O| \\ &< O\left(\frac{\log_2 k}{(k-1)!} \cdot (\lceil \pi \cdot (\beta^2 - \alpha^2) \rceil \cdot |O|)^{(k+1)}\right) \end{aligned}$$

It must be noted that, if number k is unavailable, then it suffices to replace k with $|O|$ in the above expression. As this term dominates the complexity of the other steps, this will be the complexity of the NAIVE-MINSEP-EXACT algorithm. ■

Conclusions

In this paper a technique which reduces the total computational cost in any version of point-based Geospatial Abduction Problems (point-based GAPs for short) was introduced, and an exact algorithm for the natural optimization problem of point-based GAPs, named MINSEP, was presented along with its computational complexity results.

Even though the use of GAPs has already been extended from point-based explanations to region-based explanations [9, 10, 13] and Geospatial Abduction has even been examined as a game between adaptive adversaries (in terms of game theory) [11], the field is very promising for future work. First of all, the extension of Geospatial Abduction must be explored in dimensions greater than two [8]. In particular, priority should be given to the 3D surface of earth, which is produced by geospatial digital data such as Digital Elevation Model (DEM), Digital Terrain Model (DTM) etc. Secondly, the possibility of replacing the deterministic feasibility predicate with a probability distribution function (PDF) must be explored, or analogously the replacement of the deterministic α , β distances with a PDF based on distance [8]. Finally, Geospatial Abduction should be able to capture topological relations between each observation and the set of partner locations, for example, [8] an observation cannot have a partner across a body of water.

References

1. Blake, R. M., Ducasse, C. J., Madden, E. H., 1960. Theories of scientific method: The renaissance through the nineteenth century, Reprint 1989, ISBN 2881243517, Gordon and Breach Science Publishers S.A.
2. Cormen, H. T., Leiserson, E. C., Rivest, L. R., Stein, C., 2002. Introduction to Algorithms, Second Edition, ISBN 0262032937, MIT Press.
3. Johnson, D. S., 1982. The np-completeness column: An ongoing guide, Journal of Algorithms, Volume 3(2), pp. 182 – 195.
4. Lloyd, S., 1982. Least Squares Quantization in PCM, IEEE Transactions on Information Theory, Volume IT-28(2), pp. 129 – 137.
5. MacQueen, B. J., 1967. Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, Volume 1, pp. 281 – 297.
6. Peirce, C. S., 1931–1958. Collected Papers, Cambridge, MA: Harvard University Press.
7. Preparata, P. F., Shamos, I. M., 1985. Computational Geometry: An Introduction, ISBN 03879611313, Springer-Verlag, New York.
8. Shakarian, P., Subrahmanian, V. S., Sapino, M. L., 2009. SCARE: A Case Study with Baghdad, Proc.2009 Intl. Conf. on Computational Cultural Dynamics, Dec. 2009, Nau D. and Mannes A., AAAI Press.
9. Shakarian, P., Subrahmanian, V. S., 2011. Region-based Geospatial Abduction with CounterIED Applications, accepted for publication in Wiil, U.K. (ed.) Counterterrorism and Open Source Intelligence, Springer Verlag Lecture Notes on Social Networks, to appear.
10. Shakarian, P., Nagel, M., Schuetzle, B., Subrahmanian, V. S., 2011. Abductive Inference for Combat: Using SCARE-S2 to Find High-Value Targets in Afghanistan, in Proc. 2011 Intl. Conf. on Innovative Applications of Artificial Intelligence, Aug. 2011, AAAI Press.
11. Shakarian, P., Dickerson, P. J., Subrahmanian, S. V., 2012a. Adversarial Geospatial Abduction Problems ACM Transactions on Intelligent Systems and Technology (TIST) Volume 3 (2), Article No. 34.
12. Shakarian, P., Subrahmanian, V. S., Sapino, M. L., 2012b. GAPS: Geospatial Abduction Problems, ACM Transactions on Intelligent Systems and Technology (TIST), Volume 3(1), to appear.
13. Shakarian, P., Subrahmanian, V. S., 2012. GeoSpatial Abduction: Principles and Practice, ISBN 9781461417934, Springer.
14. Thagard, P., 1992. Conceptual revolutions, ISBN 9780691024905, Princeton, Princeton University Press.
15. Thagard, P., 2007. Abductive Inference: From Philosophical Analysis to Neural Mechanisms, in Inductive Reasoning: Experimental, Developmental and Computational Approaches by Freeny A. and Heit E., ISBN-9780521856485, Cambridge University Press, pp. 226 – 247.
16. Vlahavas, I., Bassiliades, N., Kefalas, P., Kokkoras, F., Sakellariou, I., 2011. Artificial Intelligence, 3rd Edition, ISBN 978-960-8396-64-7, University of Macedonia Press, Greece.