Selecting Relevant Sensor Providers for Meeting "Your" Quality Information Needs

George Tychogiorgos
Electrical and Electronic Engineering
Imperial College
London SW7 2AZ, UK
g.tychogiorgos@imperial.ac.uk

Chatschik Bisdikian
IBM Research
Thomas J. Watson Research Center
Hawthorne, NY 10514, USA
bisdik@us.ibm.com

***Abstract*—The accelerated, on-demand deployment of sensor networks raises the challenge of selecting the providers supplying the most "relevant" sensory information to a user's needs. This paper considers the spatial relevancy of information provided that accounts for the spatial gradations in the quality of the information desired and, specifically, considers: (*a*) selecting the minimum number of providers that cumulatively maximizes the information relevancy; and (*b*) considering a cost per provider, selecting the subset of providers that cumulatively maximizes the overall information relevancy subject to a budgetary constraint. The performance and robustness of the proposed solutions are studied both analytically and by simulation for a number of provider topologies.**

I. Introduction

Consider the case where, say, a city agency needs to monitor air-quality throughout the area of its authority. The agency is willing to utilize air-quality information of different quality levels, e.g., higher granularity in densely populated regions, and lower granularity at other regions. To collect the needed information, the agency will use a combination of its own sensors and third-party fixed and mobile sensory information providers with whom it would create persistent or transient relations as necessary. The third-party providers could be other city agencies, private operators that, for example, monitor air-quality in public areas (parks, arenas, etc.), fleet operators whose fleet vehicles are equipped (for various reasons) with the necessary sensory devices, and even individuals whose smart-phones are capable of sensing air-quality conditions.

This hypothetical (albeit not improbable) scenario exemplifies a trend where increased deployment and use of sensor networks is ushering a new era of information-rich, fast-paced, pervasive solutions. The emergence of the *Internet of Things* (IoT) [1] and *participatory sensing* [2] will further hasten the rate and ease with which information from tethered, untethered sensors, the Web, etc., will coalesce on demand to support our information needs.

Research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

There are undoubtedly several challenges in realizing the "city agency" scenario, and not all are technological! In this paper we study one of the technological challenges that of dealing with selecting information providers that supply the most *relevant* information for our (i.e., the user's) needs. To this end, we build upon the relevancy metrics in [3] where, assuming compatible types for the sought and provided information, relevancy was measured by "how spatiotemporally close" a piece of information provided is to the information desired. Specifically, we defined and measured spatial relevancy by the degree of overlap between the regions describing the coverage of sensory information from the providers and the region describing the coverage of sensor information desired by a user; we, similarly, defined and measured temporal relevancy. The spatiotemporal properties of information are part of the physical context metadata for the *quality of information* (QoI) [4].

As the number and variety of potential sources of information as well as the number of applications that depend on and search for them increases, the process of selecting the most relevant ones becomes more and more challenging. Furthermore, the fluidity of untethered sources (humans in participatory sensing, sensor-equipped vehicles, etc.) adds to the challenge as an application interested in information from a particular region may need to seek for and bind repeatedly to new(er) relevant sources.

Extending our earlier work, we consider the aforementioned multitude of operational challenges as sensory sources and applications that depend on them increases. Researching this topic we have introduced QoI functions for describing the quality of information of the desired (or provided) information and defined an extended relevancy metric based on the QoI functions. Furthermore, we have looked into the problem of metadata expansion that results from the aggregation of spatiotemporal metadata from different providers and devised finite, expansion-proof metadata descriptors for the QoI functions, using approximation techniques, such as spline surfaces. Due to space limitations, the details of this part of our research can be found in [5]. Nonetheless, we will highlight these as necessary in the next section since they lend support to the following specific contributions in this paper: (*a*) the formulation of optimization problems for selecting relevant providers with our without constraints; and (*b*) the solution algorithms

978-0-7695-4436-6/11 $26.00 © 2011 IEEE
DOI 10.1109/MDM.2011.40

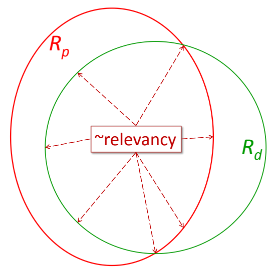

Fig. 1. Spatial properties of desired and provided sensor-originated information–regions are shown as ellipses for illustrative purposes only.

and performance study of these optimization problems.

The organization of the paper is as follows: Section II highlights the aforementioned additional work in [5]. Section III presents the multi-provider selection problem and solution techniques. Section IV provides the numerical evaluation of our solutions. Finally, Section V summarizes the paper and provides concluding remarks along with related work.

II. Highlighted Work

A. Spatial Relevancy Metrics

In [3], we introduced *spatial relevancy*[1] to represent the degree of spatial overlap that there is between the region for which we seek information for and the coverage of the information we are provided with, assuming compatible types of information, e.g., air-quality, between sought and provided information. With respect to Fig. 1, R_d is the spatial coverage of the desired information and R_p is that of the provided information, then the spatial relevancy $r_s(R_d, R_p)$ was represented in [3] by the area of the overlap normalized by the area of the desired region, $area[R_d \cap R_p]/area[R_d]$.

The relevancy r_s in [3] treated the entire region of overlap uniformly, independently of any QoI expectations desired by the user, or provided by the provider. To accommodate gradation in QoI, we introduced in [5] the *desired QoI function* $q_d(\cdot)$ describing the quality of the desired information for each point related to point $\omega = (x, y)$ in R_d, for example, at point at ω, the desired probability of (an) event detection should be above 95%, or, the air-quality measurement to be within 5% of the actual air-quality level, etc. By convention, we set $q_d(\omega) = 0$ for all points $\omega \notin R_d$. In a similar way, we also introduced the *provided* (or *provider*) *QoI function* $q_p(\cdot)$ defined on the provider coverage set R_p.

Finally, considering the *value function* $v\big(q_p(\cdot); q_d\big)$ that represents the value the sensor-enabled application gains in executing its task when it uses information of quality $q_p(\omega)$ at point ω, while q_d was desired, the relevancy metric becomes:

$$r_s^v(q_d, q_p) = \frac{\int_{R_d \cap R_p} v\big(q_p(\omega); q_d\big)\, d\omega}{\int_{R_d} v\big(q_d(\omega); q_d\big)\, d\omega}. \tag{1}$$

[1]For ease of presentation, and without lack of generality, we focus only on spatial relevancy over two-dimensional regions. Extensions to 3-D (or 4-D) spatiotemporal volumes are possible, albeit at increased levels of notational (and computational) complexity.

B. Spline-based QoI Function Description

To control the potential metadata explosion that may result from the aggregation of providers and their spatial metadata, we introduced in [5] the use of *B-spline surfaces* defined by M parameters for approximating the QoI functions q (either desired or provided) over the respective regions R. With respect to Fig. 2, *B-splines* are used to generate M parameters describing the desired and provided QoI functions q and the corresponding regions R. For each region, three additional points $\{(x_i, y_i); i = 1, 2, 3\}$ may also be used to describe the minimum rectangle containing them. With regard to our "city agency" scenario, the number M of the parameters is assumed to be known to all providers that do business with the city.

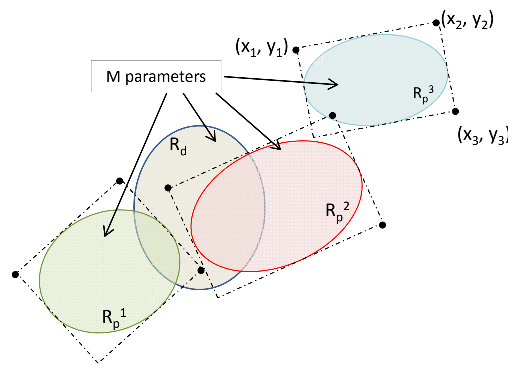

Fig. 2. Example of multiple desired/provided regions R and container rectangles.

While increasing M gives more accurate description of a QoI function, simulation results show the efficiency of the method even for low order approximations; some of these results are shown in Section IV-A. Alternatives to *B-splines* were also considered such as sampling the R regions or quantizing the range of the QoI functions q. Space limitations do not allow us to discuss these other cases any further, but, nonetheless, we have found the spline approximations quite general and effective.

III. Multi-Provider Selection

While it is possible that a single provider may suffice in satisfying an application's needs, it is quite likely that it will not. In this case, it would be desirable to be able to judiciously select a number of providers that cumulatively provide the most relevant information.

For this case, we assume there is an application with q_d and R_d representing its desired QoI function and corresponding region. There is also a set $\mathcal{P}$ of providers of size $|\mathcal{P}| = N$ with q_p^i and R_p^i, $i \in \{1, \ldots, N\}$, the corresponding provider QoI functions and coverage regions. In the following subsections, we consider two cases: (*a*) the no-cost case, where we seek to find the minimum number of providers that satisfy the application needs without any budgetary constraints; (*b*) the cost case, where engaging providers comes at a cost and applications (for example, the aforementioned city agency) have budgetary constraints. In both cases, we will first formulate a model for the problem and then consider a solution for it.

A. Maximum Relevancy with Minimum Providers and No-cost

We start with the case of selecting the minimum number of providers that can cover as much of the desired region as possible while attaining as high quality as possible. To this end, let $\mathbf{I} = [I(1), \ldots, I(N)]$ be the provider *selection indicator* vector, with $I(i) = 1$ if provider i is selected, and $= 0$ otherwise. Additionally, let the aggregate provider region $R_p^{\mathbf{I}}$ be the union of all the selected provider coverage regions, i.e., $R_p^{\mathbf{I}} = \bigcup_{i=1}^N I(i) \cdot R_p^i$, and let $R_p \stackrel{\text{def}}{=} R_p^{\{\mathbf{I=1}\}} = \bigcup_{i=1}^N R_p^i$.

The selection of the appropriate set of providers to maximize the coverage of the desired region with no cost can be modeled by the following optimization problem Π_{nc}:

Problem $\mathbf{\Pi_{nc}}$: For $I(i) \in \{0, 1\}$, $i \in \{1, \ldots, N\}$,

$$\text{minimize} \sum_{i=1}^{N} I(i), \quad \text{such that, } \forall \omega \in R_d \cap R_p :$$

$$(1) \sum_{i:\omega \in R_d \cap R_p^i} I(i) \geq 1; \text{ and} \tag{2}$$

$$(2) \max_{i:\omega \in R_d \cap R_p^i} \big\{ I(i) \cdot q_p^i(\omega) \big\} = \max_{i:\omega \in R_d \cap R_p^i} \big\{ q_p^i(\omega) \big\}.$$

Constraint (1) is a *coverage* constraint that states that for each point $\omega \in R_d$ covered by one or more providers, at least one of them will be selected. Constraint (2) is a *preference* constraint that states that the provider with the highest QoI at a point ω shall be chosen. Note that this model allows the selection of providers that overlap at some points, however, it assures that the best provider at each point is among the selected ones. Therefore, the formulation is implicitly maximizing the aggregated spatial relevancy.

Problem Π_{nc} is a generalization of the *set covering* problem [6] on three dimensions (each 2-D point ω is also associated with quality value $q_d(\omega)$) and for unity costs, which is one of Karp's 21 *NP*-complete problems [7]. Therefore, the Π_{nc} problem is *NP*-complete as well and, hence, there is no polynomial-time algorithm that solves it. The most efficient algorithm solving (approximately) the set covering problem is a *greedy* algorithm. Based on this, we propose a solution to problem Π_{nc} described by Algorithm 1 which, at each iteration, selects the most appropriate subset of providers that maximize the total relevancy with respect to the desired information having QoI function q_d. Thus, the provider that results in the largest increase in the aggregate relevancy is chosen at each iteration and the algorithm terminates when none of the remaining providers can increase the aggregate relevancy further.

More specifically, at each iteration t, the aggregate coverage region $\mathcal{S}$ from the set $\mathcal{F}$ of providers already selected, i.e., $\mathcal{S} = \cup_{k \in \mathcal{F}} R_p^k$, is merged with the new candidate region R_p^i. Then, the relevancy of the aggregated QoI function $q_p^{i,\mathcal{F}}(\omega)$ (see shortly) is calculated for each of the candidate providers i together with those in the set $\mathcal{F}$. Consequently, the provider leading to the highest aggregate relevancy, V^t, is selected until there is no further increase in the total relevancy.

Algorithm 1 – Aggregate Relevancy

1: Initialize: $\mathcal{F} = \emptyset$, $\mathcal{S} = \emptyset$, $\mathcal{P} = \{1, \ldots, N\}$, $t = 1$ and $V^0 = 0$;
2: Set: $\mathcal{F}_i^t = \mathcal{F} \cup \{i\}$, $\mathcal{S}_i^t = \mathcal{S} \cup R_p^i$ for all providers $i \in \mathcal{P}$;
3: Calculate *spatial relevancy*, $r_s^t(i) \stackrel{\text{def}}{=} r_s^t\big(q_d(\omega), q_p^{i,\mathcal{F}}(\omega)\big)$, for all regions $\mathcal{S}_i^t$ using equation (1);
4: Set $k \leftarrow \arg\max_i \big\{ r_s^t(i) \big\}$ and $V^t \leftarrow r_s^t(k)$;
5: **if** $V^t = V^{t-1}$ **then**
6: STOP;
7: **else**
8: Set: $\mathcal{F} \leftarrow \mathcal{F}_k^t$, $\mathcal{S} \leftarrow \mathcal{S}_k^t$; $\mathcal{P} \leftarrow \mathcal{P} \setminus \{k\}$;
9: Go to step 2 with $t \leftarrow t + 1$;
10: **end if**

In step 3 of Algorithm 1, we use the *aggregated* QoI function $q_p^{i,\mathcal{F}}(\omega)$ which represents the collective behavior of the providers already selected, i.e., in the set $\mathcal{F}$, and the new candidate provider i at the point $\omega \in \mathcal{S}$. Specifically, given two providers i and j with q_p^k and R_p^k, $k \in \{i, j\}$, their respective QoI functions and provider regions, their combined QoI function $q_p^{i,j}$ is defined on $R_p^i \cup R_p^j$ where $q_p^{i,j}(\omega) = h\big(q_p^i(\omega), q_p^j(\omega)\big)$; recall that q_p is set to 0 outside its region R_p. The transformation h produces another QoI function from the constituent QoI functions which reflects how the quality of fused information is assessed. For example, if the accuracy of a measurement from provider i at a point ω is 3% and from provider j is 5%, the aggregated quality from the two providers could be the better accuracy of the two, i.e., 3%, i.e., "$h \equiv \max$." We use the latter example h in our numerical results later on, thus for $\omega \in R_p^i \cup R_p^j$, we will use:

$$q_p^{i,j}(\omega) \stackrel{\text{def}}{=} h\big(q_p^i(\omega), q_p^j(\omega)\big) = \max\big\{q_p^i(\omega), q_p^j(\omega)\big\}. \tag{3}$$

Algorithm 1 can be implemented in polynomial time. At each iteration, the algorithm determines the optimal provider to select, but, similarly to how the *greedy* algorithm behaves for the original set covering problem, this may not always lead to the overall optimal solution.

The scenario described in this sub-section did not take into account a possible cost for using the sensory information of a particular provider. Problem formulation Π_{nc} and its solution in Algorithm 1 identify the best subset of providers that maximize the aggregate spatial relevancy of information independently of the cost. Next we consider an additional model formulation that takes this cost into account when choosing the optimal provider set.

B. Maximum Relevancy with Budget Constraints

Since nothing comes for free, sooner or later, the city agency will have to face the realities of budgetary constraints. In this case, suppose the city agency's budget is B and the cost of engaging provider i is c_i, $i = 1, \ldots, N$. Thus, we are now interested in finding the optimal set of providers that will maximize the spatial relevancy of the provided information subject to the budget constraint B. Again, this case can be modeled by a combinatorial optimization problem. Specifically, let again $I(i)$ be the 0-1 indicator variables for selecting provider i and $\mathbf{I}$ the corresponding vector. Thus, the formulation of the optimization problem in this case will be:

Problem $\Pi_{\mathbf{bg}}$: For $I(i) \in \{0,1\}$, $i \in \{1,\dots,N\}$,

$$\text{maximize } r_s\big(q_d, q_p^{\mathbf{I}}\big), \text{ such that } \sum_{i=1}^{N} I(i)\cdot c_i \le B, \quad (4)$$

where $r_s\big(q_d, q_p^{\mathbf{I}}\big)$ is the relevancy of a "super-provider" with a QoI function aggregated from the providers indicated by selection vector $\mathbf{I}$ (see discussion following Algorithm 1) and defined on the coverage region $R_p^{\mathbf{I}}$. We note that in Π_{bg} the increase of the relevancy when adding a specific provider i does not only depend on i alone but on the already selected providers as well. In the case that the providers already selected are offering good enough quality on all points ω in R_p^i, adding provider i may not increase the relevancy attained.

Problem Π_{bg} is a generalization of the 0-1 *knapsack* problem [8] where the value of each item is a function of the items already selected to be included in the knapsack. For example, adding a lighter in the knapsack may reduce (even to zero) the subsequent value of a box of matches. This is captured with the use of $q_p^{\mathbf{I}}$ as a function of the vector $\mathbf{I}$. The 0-1 knapsack problem is an *NP*-hard optimization problem which means that there is no algorithm that finds the optimal solution in polynomial time. The *greedy* algorithm would need to check all 2^N different combinations between the N providers, prune those that do not satisfy the available budget and then choose the combination that maximizes the aggregate relevancy. Instead, we consider the *dynamic programming* type Algorithm 2 that solves the problem in *pseudo-polynomial* time splitting the problem into smaller subproblems, storing their solutions into memory, and, then, using them to calculate the solution of the main problem.

Algorithm 2 iteratively constructs the $N \times B$ matrix $\mathbf{Values}$, whose entry $Values[i,b]$ is the maximum aggregate spatial relevancy of the first i providers for a budget b; the corresponding provider selections reside in the indicator vector $\mathbf{I}_i^b$. The matrix entry $Values[N,B]$ stores the maximum aggregate spatial relevancy of all providers for budget B, which is the optimal solution for Π_{bg}, and the optimal provider selection will reside in the vector $\mathbf{I}_N^B$. As mentioned earlier, Π_{bg} is an extended 0-1 knapsack problem with variable item value. Therefore, lines 6-10 of Algorithm 2 calculate the spatial relevancy of the providers (i.e., the item values) in the specific selection vector $\mathbf{I}$. The relevancy of providers in vectors $\mathbf{I}$ that have already been calculated at earlier iterations are evoked from memory. This has a significant impact in accelerating the algorithm. Moreover, lines 11-18 of the algorithm determine whether selecting a new provider will result in higher aggregate spatial relevancy and, if yes, select the provider.

Algorithm 2 – Budget Constrained Aggregate Relevancy

1: **for** $i = 1$ to N **do**
2:   **for** $b = 0$ to B **do**
3:     **if** $c_i \le b$ **then**
4:       $\mathbf{I} = \mathbf{I}_{i-1}^{b-c_i}$; where: $\mathbf{I}_0^{b-c_i} \stackrel{\text{def}}{=} \mathbf{0}$ and $\mathbf{I}_{i-1}^0 \stackrel{\text{def}}{=} \mathbf{0}$;
5:       $I(i) = 1$;
6:       **if** $r_s\big(q_d, q_p^{\mathbf{I}}\big)$ not calculated **then**
7:         Calculate $r_s\big(q_d, q_p^{\mathbf{I}}\big)$ using (1);
8:       **else**
9:         Get $r_s\big(q_d, q_p^{\mathbf{I}}\big)$ from memory;
10:       **end if**
11:       **if** $r_s\big(q_d, q_p^{\mathbf{I}}\big) > Values[i-1,b]$ **then**
12:         $Values[i,b] = r_s\big(q_d, q_p^{\mathbf{I}}\big)$; $\mathbf{I}_i^b = \mathbf{I}$;
13:       **else**
14:         $Values[i,b] = Values[i-1,b]$; $\mathbf{I}_i^b = \mathbf{I}_{i-1}^b$;
15:       **end if**
16:     **else**
17:       $Values[i,b] = Values[i-1,b]$; $\mathbf{I}_i^b = \mathbf{I}_{i-1}^b$;
18:     **end if**
19:   **end for**
20: **end for**

The dynamic programming algorithm for the 0-1 knapsack problem has complexity of $O(nB)$, where n is the number of items and B the available budget. In the worst case, Algorithm 2 will calculate the spatial relevancy $r_s\big(q_d, q_p^{\mathbf{I}}\big)$ at each iteration, which needs $O(N)$ time. Therefore, the absolutely worst case time complexity of Algorithm 2 is $O(N^2 B)$, where N is the total number of providers. Regarding the memory requirements, in the worst case, it is necessary to store the matrix $\mathbf{Values}$ of size $N \times B$, the relevancy values $r_s(q_d, q_p^{\mathbf{I}})$ for each selection vector $\mathbf{I}$, which are in total $\min\{2^N, N \times B\}$, and the optimal selection vector $\mathbf{I}_i^b$ of size N for the $N \times B$ iterations of the algorithm.

The implementation of the algorithm can be accelerated both in time and memory requirements significantly in two ways. First, instead of examining all N providers the algorithm can be run only for those intersecting with the desired QoI function. The intersection operation will be run only once, at the beginning of the process, and can be implemented in linear time. Then, instead of iterating for all values in the range $[0,B]$, we can calculate the *greatest common divisor* gcd of c_i, $i = 1,\cdots,N$, and B and then run the algorithm in the range $[0, B/gcd]$ with costs c_i/gcd, $i = 1,\cdots,N$.

IV. Numerical Results

The numerical results in this section were derived using a combination of MATLAB-based computations and simulations. We first highlight the effectiveness of the *B-spline* approximation of QoI functions in calculating the spatial relevancy of a single provider and, then, consider the multi-provider case and the performance of the two algorithms presented in Section III. With regard to the value function $v(\cdot,\cdot)$ in (1), we assume that an end-user application gains no benefits if it receives information of higher quality than what it asked for, thus, we set $v(q_p;q_d) = \min\{q_p, q_d\}$, and, hence:

$$r_s(q_d, q_p) = \frac{\int_{R_d \cap R_p} \min\big\{q_p(\omega), q_d(\omega)\big\}\, d\omega}{\int_{R_d} q_d(\omega)\, d\omega}. \qquad (5)$$

A. Single-provider Spatial Relevancy

The objective of the single-provider study is assessing the robustness of the spline-based, finite-size approximation of QoI functions in ordering providers according to their relevancy to a desired QoI function. We show here the results for an *urban* scenario where the desired and the various provider regions R line-up along city streets (the "Manhattan street" topology), see Fig. 3; see [5] for a rural scenario.

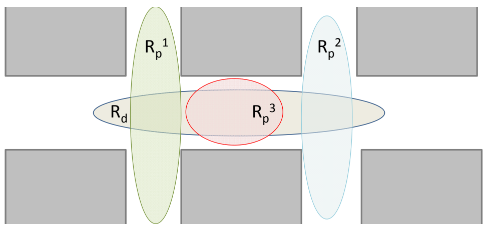

Fig. 3. The urban (Manhattan street) topology case.

We applied the single-provider relevancy method using different values for the number of *B-spline* parameters M. Then, the spatial relevancy metric was calculated through the *B-spline* approximation and was compared against the actual spatial relevancy of the providers using their original QoI functions. We studied: (a) the estimation error as a function of M; and (b) the effect of this error on ordering providers according to their spatial relevancy. Note that, the goodness of the approximation is judged not in absolute terms (which is a comparison over a continuum of values) but rather over an ordering outcome (which is a comparison over a finite set of alternatives).

The analysis results in Fig. 4 illustrate the robustness of the method with regard to this objective. As the top plot shows, the estimation error for the spatial relevancy of each provider is relatively low even when using around $M = 100$ parameters for the QoI function approximation. More importantly, there are no misordering effects even when the spatial relevancy of some providers is almost identical, as in the case of providers 1 and 2. This is indicated in the bottom plot in Fig. 4 by the fact that the red and blue lines do not intersect; an intersection would have meant a change in the relative order of provider relevancy.

Building upon the procedure of calculating spatial relevancy for single providers, in the next subsection we will present the simulation results of the two algorithms proposed for the multi-provider selection problem.

B. Multi-provider Spatial Relevancy

In the case of the multi-provider selection problem with or without the budget constraint, the QoI functions used were mixtures of varying number of Gaussian density functions, randomly scaled and placed on the two-dimensional plane. Fig. 5 shows an example case, where the desired QoI function is colored in blue, and 9 providers are colored in red, cyan and green.

The proposed algorithms are based on pseudopolynomial heuristics to solve *NP-Hard* problems. These algorithms were

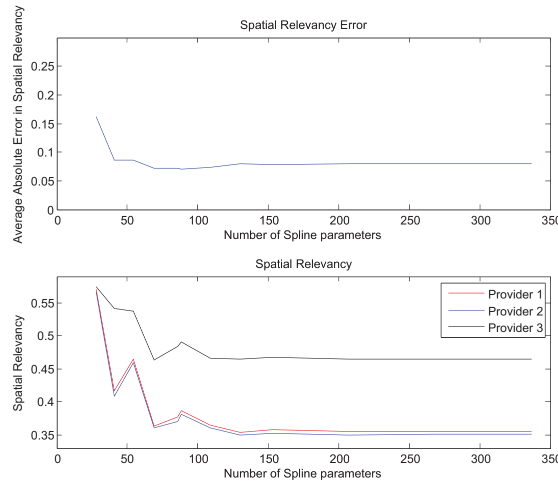

Fig. 4. Spatial relevancy for the urban topology as a function of M.

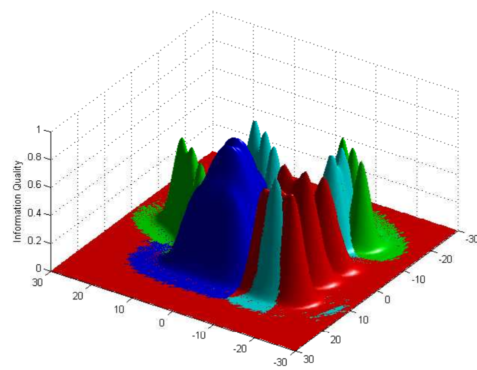

Fig. 5. QoI example functions for the multi-provider case.

adjusted to accommodate our objectives regarding the spatial relevancy of providers. Hence, the objective of our simulation study was the assessment of the algorithm effectiveness in selecting the right providers that satisfy problems Π_{nc} and Π_{bg} in Section III. The assessment is performed by comparing the solutions and execution time of the proposed algorithms against those from the exhaustive search algorithms. For the no-cost case, the latter calculates the spatial relevancy of all $(2^N - 1)$ different combinations between the N providers and the selection of the best one according to (2). For the budget constraint case, the exhaustive search algorithm includes the comparison of all *feasible* combinations, i.e., those with a total cost less than or equal to the budget B, and the selection of the optimal one among them according to (4).

Fig. 6 shows the comparison of the execution time between the proposed algorithms and the exhaustive method in each case. For all cases studied, the solutions that the proposed algorithms arrived at were the same as the ones given by the exhaustive search methods, which of course are the optimal ones. As expected, the execution time of the exhaustive algorithms increases exponentially as the number of providers increases,

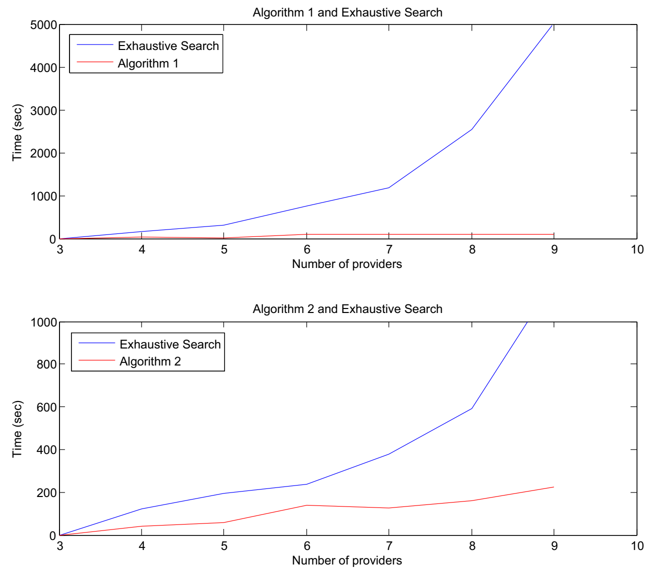

Fig. 6. Execution times comparison between Algorithms 1, 2 and exhaustive search using a 2.4GHz dual core Windows PC with 4GB of RAM.

while it increases almost linearly for algorithms 1 and 2. The execution of the proposed algorithms has also been accelerated by a mechanism of pruning providers not intersecting with the desired QoI function. In such cases, these providers are removed from the rest of the process which results in further reducing the number of combinations examined.

V. Related Work and Concluding Remarks

In this paper, we introduced a novel problem area for sensor networks that of identifying and selecting providers providing *relevant* sensory-information based on their sensing capabilities regarding their coverage and QoI and how they compare to those desired. The focus in this paper has been in the spatial domain for ease of presentation, but temporal extensions are also possible. This problem will become more and more prominent as the number of providers increases and their sensing capabilities change in the spatiotemporal domain, such as when using wireless and mobile sensor networks operating over a multi-administrative domains, e.g., vehicle-mounted sensors, participatory sensors, etc. Within this area we highlighted relevancy metrics based on QoI functions and a finite, expansion-proof technique for metadata based on *B-spines*. Based on these, we have formulated related optimization problems and proposed efficient algorithms for selecting the best collection of providers that are most relevant to our needs given various constraint objectives.

To the best of our knowledge this is the first endeavor in the area of QoI and information relevancy in sensor networks, and we are aware of no prior work that directly relates to our research in this paper. There is, however, prior literature that inspired and influenced our research. Specifically, supplementing our own cited work on QoI, [9] discusses quality metadata describing geospatial information. Ref. [10] provides an extensive review of the models for spatio-temporal information databases and related queries. Ref. [11] considers summarizing 2-D shapes via a bounded number of parameters. These shapes could correspond to our regions R and, thus, the proposed approach in [11] could serve as an alternative to our *B-spline* approach. We do not discount the latter approach and could have been used in our paper as well. However, given that we ultimately pursue a comparison and selection of relevant providers using QoI criteria as well, we found the use of the *B-spline* approach more flexible. Ref. [12] provides a survey of coverage in sensor networks, but it identifies no study regarding coverage comparisons between sought and provided information or associating coverage with QoI gradations. Finally, our inspiration in using splines comes from [13] which considers the increase in the number of the time-decaying security metadata of documents produced by the combination of constituent documents.

Future work in this novel area, may include the study of the various architectural aspects related to QoI function advertisements (what, when and how to advertise), as well as the consideration of time-varying QoI functions that could result by system impediments, such as loss of sensors, and fluidity of sources, such as when considering mobile sensors and participatory sensing.

References

[1] N. Gershenfeld, R. Krikorian, and D. Cohen, "The Internet of Things," *Scientific American*, vol. 291, no. 4, pp. 76–81, October 2004.
[2] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," in *World Sensor Web Workshop (in ACM Sensys'06)*, Boulder, CO, USA, October 31 2006.
[3] C. Bisdikian, J. Branch, K. K. Leung, and R. I. Young, "A letter soup for the quality of information in sensor networks," in *IEEE Information Quality and Quality of Service (IQ2S'09) Workshop (in IEEE PerCom'09)*, Galveston, TX, USA, Mar. 2009.
[4] C. Bisdikian, L. M. Kaplan, M. B. Srivastava, D. J. Thornley, D. Verma, and R. I. Young, "Building principles for a quality of information specification for sensor information," in *12th Intl Conf. on Information Fusion (FUSION'09)*, Seattle, WA, USA, July 2009.
[5] G. Tychogiorgos and C. Bisdikian, "Selecting relevant sensor providers for meeting "your" quality of information needs," *IBM Res. Rep. RC25116*, Dec. 10 2010.
[6] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. MIT Press, 2009, pp. 85–103.
[7] R. Karp, *Reducibility Among Combinatorial Problems*. Plenum Press, 1972, pp. 85–103.
[8] S. Martello and P. Toth, *Knapsack Problems: Algorithms and Computer Implementations*. John Wiley & Sons, 1990.
[9] R. Devillers, Y. Bdard, and R. Jeansoulin, "Multidimensional management of geospatial data quality information for its dynamic use within gis," *Photogrammetric Engineering & Remote Sensing*, vol. 71, no. 2, pp. 205–215, February 2005.
[10] N. Pelekis, B. Theodoulidis, I. Kopanakis, and Y. Theodoridis, "Literature review of spatio-temporal database models," *The Knowledge Engineering Review*, vol. 19, no. 3, pp. 235–274, September 2004.
[11] J. Hershbergery, N. Shrivastava, and S. Suriz, "Summarizing spatial data streams using clusterhulls," in *8th Wksp on Algorithm Engineering and Experiments (ALENEX'06)*, Miami, FL, USA, Jan. 2006.
[12] A. Ghosh and S. K. Das, "Coverage and connectivity issues in wireless sensor networks: A survey," *Pervasive and Mobile Computing*, vol. 4, no. 3, pp. 303–334, June 2008.
[13] M. Srivatsa, D. Agrawal, and S. Reidt, "A metadata calculus for secure information sharing," in *16th ACM Conference on Computer and Communications Security (CCS'09)*, Chicago, IL, USA, Nov. 2009.