PODS: Interpreting Spatial and Temporal Environmental Information[1]

Edoardo Biagioni

esb@hawaii.edu
University of Hawai'i at Mānoa

Abstract

We are building wireless networks of environmental sensors to investigate why endangered species of plants will grow in one area but not in neighboring areas. We collect large amounts of data automatically and repeatedly. This paper describes requirements for presenting this real-time information in ways that allow scientists as well as more casual observers to detect significant patterns in time and space.

1 The Problem

Rare and endangered species of plants are rare and endangered because they only grow in few locations. These locations evidently have some special properties that support the growth of these endangered plants, properties that are not present in adjacent locations. We have started building devices to capture environmental information and transmit it in real-time over the internet. The environmental information we have started to collect includes rainfall, moisture, temperature, sunlight, and wind. Because these plants are in remote areas, we use wireless digital communication to relay the information back to the Internet, and from there to a database. We are also installing digital cameras that can help capture events, such as animal visits, that are not recorded by the other sensors. We call such a grouping of sensors and cameras a *pod*. Pods should be inexpensive so that many can be deployed. To make deployment easier, the network must be self-organizing: any *ad-hoc* layout of pods that allows wireless communication should be sufficient for data transfer back to one or more internet-connected base stations.

All of these sensor pods must be camouflaged in a way that they remain unnoticed by either humans or animals. We have camouflaged pods to look like either rocks or tree branches, and they have gone suitably undisturbed (and presumably unnoticed) over many months in areas much more crowded than our intended areas of deployment.

Ad-hoc wireless sensor networks present interesting issues of routing in wireless networks, of energy management, of data prediction and compression, and of the design of the sensors themselves – for example, we have not yet found a good design for unobtrusive instruments for measuring wind using only low power. In this paper we focus on the issue of presenting relevant data to users of this near-real-time information.

[1]This research was sponsored by the Defense Advanced Research Projects Agency, under the title "A Remote Ecological Micro-Sensor Network.

1.1 Information

One of the reasons for the near-real-time data collection is to allow scientists to take action when situations occur that are in some way interesting or unexpected. For example, a scientist may request pictures or request that special data be collected, or may even visit the site in person to observe for example flowering or the effect of herbivore browsing. The most important data we need to present is therefore the most recent data collected, but we must present it in the context of historical and geographical data.

1.2 Geographical Gradients and Temporal Predictability

In Hawai'i, weather gradients are very sharp, with rainforests and deserts sometimes less than 10 miles apart. We need to present our data to make obvious the geographic and temporal nature of weather gradients. We are therefore not surprised that endangered plants are limited to relatively small areas, but we suffer from a lack of reliable information about the weather in these small areas, since data from collection stations nearby may not provide reliable information about the weather at the location where the species of interest grows. By using a very dense network of relatively inaccurate sensors, we hope to be able to detect correlations among weather phenomena in nearby areas, giving us knowledge that may help protect plants that might need specific weather conditions to survive. Unlike conventional weather displays, we need to focus on the gradients rather than the peaks of phenomena, and on the two dimensions of what the "normal" weather is for a certain area, and on how much the weather differs from "normal" at the current time and in the recent past. If one area is always warmer than another, we want that to be clear in the data presentation. If one area is always warmer than another, but today is cooler, we need to flag this as a significant event likely to be interesting to a human.

We have also started to study the temporal correlation of sensor data. We have placed some sensors near our laboratories on the island of Oahu and have noticed, by repeated observation, that though sometimes the sky is clear in the morning (we observe smooth changes over time in the light measurements), it has never been clear in the afternoon. We need the ability to automatically deduct this information from the data and, to the degree possible, present high-level summaries as well as detailed information.

1.3 Images

Some of our pods are being designed to hold cameras. In contrast to conventional web cameras, these are high-resolution still cameras that can capture many of the details (such as insect visits) that are significant in the life of a plant. With a battery-powered wireless network, there is always a tradeoff between the amount of data transferred and the lifetime of the batteries. We can take more pictures and replace the batteries more frequently, or have fewer pictures and longer battery lifetime. There is sufficient processing power in a pod that, if it is possible to automatically identify phenomena of interest, it will be possible to automatically adjust the frame rate of cameras in different locations to cover a range of weather phenomena. Phenomena such as rain that might be meaningful to a human in the field may not be as significant as wind, for example, which might prevent pollinators from reaching the plants. We have placed preliminary cameras in the field (`http://www.botany.hawaii.edu/pods/`) to begin the process of learning what pictures are significant and worth the energy to transmit them.

2 Human Interfaces

Standard ways of displaying the data include tables, graphs, and maps. These give us the ability to express large amounts of both time-sequence and spatial information simultaneously. We have planned relatively large and dense network with approximately a hundred sensors. We can draw maps showing the position and instantaneous value for each sensor, but this does a poor job of emphasizing gradients and identifying and making visible significant geographical boundaries in the values of sensor measurements. One of our preliminary user interfaces (`http://red2.ics.hawaii.edu/cgi-bin/location`) uses colors to indicate temperature and rainfall on a map (blue for rain, cyan for dry cool, red for dry hot), and in case of rain, uses dot sizes to indicate how recently the rain was detected. We have also been generating daily plots of all our sensor measurements, showing variations during the course of the current day. Some of these variations are predictable – more light during the day, and more variation in light intensity during the afternoon – and some, such as rain, are less predictable.

We consider these interfaces preliminary because, although they do a good job of presenting small slices of the data, they fail to present the big picture. The remainder of this section sketches some characteristics of user interfaces that we have envisioned but have not yet realized.

The first characteristic is flexibility and configurability. A scientist interested in the solar radiation in a given area should be able to focus in on the data of interest, ignoring irrelevant data. This is a challenge of choosing which dimensions of the data to display. Not only do we have the three axes of space and the fourth dimension of time, we have one dimension for each type of sensor in our pods. Correlations between any set of dimensions may be of interest to scientists and ecologists.

In contrast, students doing a project for high school might want to focus on year-to-year data, ignoring more ephemeral phenomena. This is the second characteristic, and it may be described as the ability to "zoom" into specific sets or summaries of the data. Another description of this characteristic is the range and precision of a query on the sensor data. The precision might be spatial, temporal, and over ranges of sensor values.

In general, users might want to query the database using both different dimensions and different range and precision for each dimension. We are aware of much work having been done in many related ares, for example as described by Swan and others (1998), by Larson (2001), or even for the presentation of weather information, for example NOAA (2001). We look forward to adapting these ideas to our application as much as possible.

For our high-resolution camera images, we are planning to use large displays to provide specialists with close-up long-term study of the species (*phenology*), and non-specialists with immersive experiences of special environments such as rainforests and drylands and such weather phenomena as rainshowers and droughts.

3 Machine Learning

The ultimate user interface, for this and for other problems, is one that can respond to natural-language queries or can directly infer the user's interest and display the appropriate data in response. We do not expect useful interfaces of this kind to be available within the lifetime of this project.

We are aware of and intend to investigate the applicability of existing machine-learning algorithms such as presented by Mitchell (1997) to do data predictions. Successful data prediction would contribute to our project in two ways. By only sending data that does not match the prediction, we can substantially reduce the amount of data traffic and hence the power consumption of the pods. In addition, by having a model that predicts the expected set of readings from the sensors, it becomes much easier to identify unusual situations, whether it is a shift, weakening, or reversal in a gradient, or a long-term trend that can be evaluated by a human to determine its significance.

Some of the machine learning algorithms require substantial computing power to derive a model of the data. Since the data is available in the laboratory, these computations can be done on powerful computers. The computed models can then be distributed back to the pods, and used to both reduce the data flow and to clearly signal unexpected events. We are aware that for our application, small variations from the model are usually not significant, so the machine learning algorithms need only approximately predict the actual measurements.

4 Related Work

Many groups are working on distributed sensor networks for a variety of purposes. There is a conference track (Dasarathy, 2001) devoted to sensor fusion architectures, algorithms, and applications. Some of the characteristics that appear to be specific to our application include the overall predictability of the data (we are measuring weather and plants, rather than animals, troops, or industrial processes), and the strong interest in detecting phenomena (combinations of measurements and measurement patterns in space and time) for which we have no *a priori* knowledge.

We mentioned above some of the related work in human computer interfaces for information retrieval and for distributed sensor data and weather data. Weather data is usually presented with different symbols and sometimes colors on a two-dimensional map. These symbols usually represent different values for temperature, pressure, and wind at different locations at a given time. Recently animations have been used to display the evolution of weather systems over time. It appears to the author that in general, weather data is represented in such a way that a user searching for a specific item of data can easily locate that item, but a user looking for unknown correlations is unlikely to find unexpected patterns or unusual combinations of data. In addition, it is very hard to represent time sequences of data meaningfully using these techniques – the animations tend to provide relatively little information compared to the charts, in part because the movies focus more on the "big picture" and less on the details, and in part because it is hard for humans to focus on the details in a moving picture.

Information retrieval does accomodate searches for patterns and items that are not known *a priori*. Unlike in most of IR, where spatial and temporal relations are just one more dimension of the retrieval, we expect to have user interfaces designed to focus users' attention on significant temporal and spatial correlations.

5 Future Work

As is common in research generally and specifically in the design of human-computer interfaces, we expect to keep improving our interfaces through a succession of prototypes, and make gradual progress towards a satisfactory interface for our sensor data. The initial focus of our work is on deploying the sensor nets themselves, but once the sensors are in place, we will need to have suitable means of determining what is happening in the field, if anything, and for focusing our attention on areas which are likely to prove interesting. We are interested in making all our data available to both scientists and to individuals interested in self- or group-study, including students at the high school and university levels.

6 Acknowledgements

The PODS project would not be possible without my co-principal investigators, Kim Bridges and Brian Chee.

References

Belur Dasarathy (2001), program chair. Sensor Fusion: Architectures, Algorithms, and Applications V One of the tracks in the AeroSense conference.

Ray Larson (2001). `http://sherlock.berkeley.edu/geo_ir/PART1.html`, retrieved April 10, 2001.

Tom Mitchell (1997). Machine Learning, McGraw Hill, 1997, ISBN 0070428077.

NOAA (2001). `http://iwin.nws.noaa.gov/iwin/main.html`, retrieved April 12, 2001.

Russell Swan, James Allan, Don Byrd (1998). Evaluating a Visual Retrieval Interface: AspInquery at TREC-6 Position paper for the CHI '98 Workshop on Innovation and Evaluation in Information Exploration Interfaces (Los Angeles, April 1998).