SENSORWORLD : A NEW APPROACH TO INCORPORATING LARGE-SCALE SENSOR DATA INTO ENGINEERING LEARNING ENVIRONMENTS

Hanjun Xian, Krishna Madhavan
School of Engineering Education
Purdue University

Abstract

Sensors play a critical role in engineering and science applications. However, most engineering students very rarely have access to large-scale real-world sensor data within the classrooms. Students who major in fields such as environmental engineering are not well prepared for the engineering professions because of the gap between real-world scenarios and scale of the data used within the classrooms. Diverse and non-standard software interfaces to sensors compound this problem significantly. Our goal is to document and make available data from a large variety of real-world sensors to engineering students through the iPhone and iPod Touch. Our project addresses this problem by implementing a middleware framework in the application server and a client on iPhone to facilitate access to sensor data.

The primary research questions that this paper will address are: (1) How can sensor data be incorporated into current engineering learning environments effectively? (2) What are the problems of utilizing large-scale data within the scope of an engineering curriculum? and (3) What are the characteristics of a middleware framework that will allow the inclusion of real-world data sources within the classroom? Currently, we support a total of 1136 sensors from a variety of sources. This dataset contains sensor data of air temperature, water temperature, water level, wind speed, air pressure, precipitation, conductivity, and soil moisture, and is being rapidly expanded to support a large universal set of open sensors.

Success of this project provides a chance to bring practice-oriented education into engineering classrooms. Students will be able to access real-time, real-world sensor data with a single iPhone application. Effective visualization and interface for navigation of sensor data helps engineering students better understand concepts, identify patterns, and discover problems not addressed in the textbooks. Engineering students are likely to be more engaged in the learning process by studying the latest natural phenomenon such as flooding in Atlanta and drought in Texas.

Introduction

Sensors play a critical role in engineering and science applications such as monitoring environmental metrics, controlling industrial processes, and coordinating traffic flow. Inclusion of sensing science (also known as sensor science) and sensor data within engineering classrooms is becoming increasingly beneficial for engineering education. It motivates students to pursue science and engineering disciplines and associated career paths [1]. Further, it makes the teaching in the laboratory more interesting [2] and engaging [3]. Furthermore, sensor science helps prepare students with a foundation of instrumentation technology for the measurement and control of industrial processes [4]. Despite the above efforts to produce a prevailing culture of sensing science, the vast majority of engineering students very rarely have access to a large number of real-world sensors within the classrooms. A lack of effective ways to incorporate large-scale sensor data into engineering curricula retards students' development of problem solving skills in a real-world contexts.

In this paper, we propose a new approach to incorporating sensor data into the engineering learning environment. Our project, SensorWorld, intends to document and make available data from a large variety of real-world sensors to engineering students without requiring expertise in sensor hardware, data acquisition, and data representation. To achieve this goal, we first implement an infrastructure in the application server that constantly acquires sensor data from various sources, normalizes them into a simple format, and offers a standardized data query service. To provide engineering students a more engaging learning environment, we also develop a client on iPhone/iPod Touch to visualize sensor data in a user-friendly and interactive way based on the sensor data infrastructure.

Today's students live in a cyber world and experience an unprecedented number of highly technological tools such as virtual environments and mobile applications. Learning environments that do not incorporate modern technology may drive students away from engineering disciplines [3]. We believe our iPhone/iPod Touch application exemplifies an innovative and engaging cyber-learning environment. In addition, our sensor data infrastructure provides an effective and low-cost solution to deliver large-scale real-world sensor data to students. This paper addresses the following research questions:

(1) How can sensor data be incorporated into current engineering learning environments effectively?
(2) What are the problems of utilizing large-scale data within the scope of an engineering curriculum?
(3) What are the characteristics of a middleware framework that will allow the inclusion of real-world data sources within the classroom?

Sensor Data in Engineering Learning Environments

The major challenge of incorporating large-scale real-world sensor data into engineering learning environment is the infrastructure, knowledge, and equipment prerequisite for utilizing sensor data. Typically speaking, engineering students must at first have fundamental knowledge of sensors and data acquisition in order to understand what the sensors measure, how sensors transmit data, and how to acquire the sensor data prior to learning how to represent the data and conduct statistical analyses using the data. Only after they learn how to represent and analyze a large problem space do they have an opportunity to derive applications from sensor data. To support courses and labs, institutions must provide infrastructure such as sensors and data acquisition cards. This dependency chain implies that incorporating sensor data into engineering classrooms requires significant effort from engineering students, faculty, and institutions.

Efforts have been undertaken to teach fundamental concepts related to sensors and data acquisition in engineering classrooms. Rochester Institute of Technology [5] initiated a course in wireless sensor network security to provide undergraduate students with experience in this emerging area. The SMEAGOL project [6] aimed to teach students not only programming and hardware design but also basic teamwork and project management skills in the field of wireless sensor networks. The above specialized courses and labs are indispensable for those engineering students majoring in electrical engineering, those involved in projects studying wireless sensor networks, and students who are expected to demonstrate knowledge of how sensors and sensor networks work. Similarly, it is also reasonable to offer courses on data acquisition for students specializing in electronic engineering [7,8] and biomedical engineering [9]. Even some students in other engineering disciplines such as agricultural engineering are also required to take fundamental courses in sensors and data acquisition [10] so as to gain the advantage of incorporating sensor data into their projects. Ideally, if engineering students and faculty are more concerned about sensor data rather than sensor hardware and data acquisition, they would be relieved from the time-consuming process of developing an infrastructure and focus more on the use of sensor data for engineering and science. New courses using sensors are a financial burden due to expenses related to acquisition of sensors and more importantly data cards. Often budgetary limitations constrain the scale of sensors and sensor data used in the courses, thereby creating a rift between the classroom and what is state-of-the-art in the industry.

Second, a lot of research has addressed the problem of knowledge visualization as it relates to providing students with better understandings of the learning content. Reisslein et al. [11] innovated a computer-based instructional module with equation-based and graph-based representations of the learning content. The comparison of the two indicated that the graphical representation produced significantly higher performance among seniors in electrical engineering. Jenkins et al. [12] reported a similar result that computer-aided multimedia module had a significant effect on course performance in an introductory mechanics courses. Arruarte et al. [13] organized the learning resources in the course *Introduction to Artificial Intelligence* through concept maps and revealed that the visual representation had a positive impact on inventory searching and increased satisfaction among students. Graphing techniques in the course on data acquisition only address small-scale data applicable to a limited number of equipments. Little attention has been paid to the use of visualization to convey large-scale data to students.

Methodology

In response to the difficulties identified in the previous section, we develop a new, effective way for incorporating large-scale real-world sensor data into engineering learning environments. First, we propose a new approach to simplifying the process of sensor data acquisition. Our implementation transparently provides engineering students, faculty, and other users with sensor data using a single unified format that does not require user expertise in sensor hardware and data acquisition. After we guarantee the accessibility and homogeneity of large-scale real-world sensor data, we focus on providing an effective sensor data representation by developing multiple visualizations and an interactive interface to help users better understand and process sensor data. We discuss how we implement a client on iPhone/iPod Touch to allow a simple, interactive, and effective access to the data stored on our application server.

Data acquisition of large-scale real-world sensors

As we mentioned above, introducing sensors and data acquisition often results in high expenses on equipments and a heavy burden on offering extra courses to teach fundamental sensing science and data acquisition. The scale of sensors and sensor data is often constrained by this resource limitation. To increase the scale of sensor data while lowering the cost, engineering educators may access existing and public available sensors owned by government, research groups, and other organizations. Most of these sensors are already working as part of a large-scale sensor network and thus accessing third-party sensors can save money of purchasing and deploying our own. The sensor owners often manage the data acquisition process and provide free access to the sensor data. Therefore, teaching fundamental sensing science and data acquisition is no longer a prerequisite for accessing sensor data. Although ordinary users are typically not permitted to modify any sensor parameters and thus lose hands-on experience in how sensors work physically, this approach complements the learning experience of those engineering students who are not so interested in the physical layer by allowing them to work on a more complex and practical problem. What’s more important, it opens a new opportunity for a wider range of engineering students without fundamental knowledge in sensing science allowing them to focus more on their individual engineering or science goals.

Different sensor data providers follow discrepant, many times non-standard and non-overlapping propriety data policies, publishing protocols, and software interfaces. Sensor data acquisition from these providers often requires expertise in programming and network protocols. Even if sensor data is sometimes accessible in the form of downloadable Excel files or XML files, the costly data downloading and converting process limits the scale of sensor data being accessed. Data of the same type also varies in unit of measurement and precision. When we expanded the scope of our project to obtain data from multiple sources, data integration became an obstacle in incorporating large-scale real-world sensor data into engineering learning environments.

To eliminate the heterogeneity in data format and data acquisition, we develop a middleware framework to transparently provide a single standard interface to the sensor data. Figure 1 shows the architecture and workflow of the

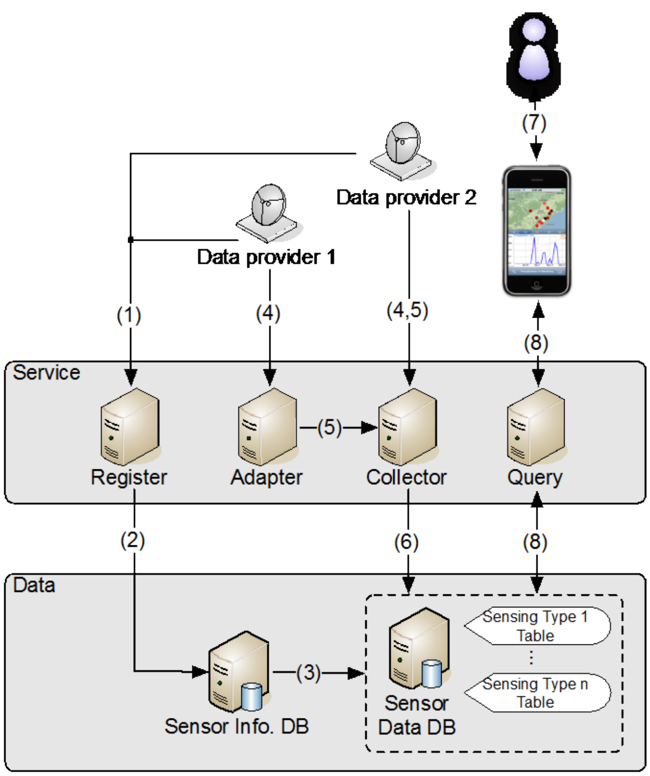

Figure 1: Architecture of the middleware framework.

middleware in our project. There are two layers in the middleware framework: the data layer and the service layer. The data layer manages the storage and low-level input/output manipulation of sensor information and sensor data. Sensor information refers to sensor-dependent characteristics such as location and the environmental metrics a sensor measures. Sensor information is pre-determined and rarely changed by the sensor owner. For engineering students and faculty, this set of read-only properties contains the essential descriptions that help judge whether the sensor fits their educational goals. For example, engineering students studying the correlation of a carbon footprint and sea level may select water-level sensors along the coastline. Sensor data refers to the data value being measured such as precipitation or air temperature. The service layer includes modules that are responsible for registering new sensors, adapting and collecting sensor data, and processing client requests. Manipulation of sensor information and sensor data can only be accomplished by making requests to the service layer.

First, we discover public available sensors and register them to the register module on the service layer, as shown in steps (1) to (3). Then we establish a network connection to the data provider and acquire data periodically. After obtaining the sensor data, the middleware unifies sensor data of diverse formats, as shown in step (4) and (5). Our middleware implementation encapsulates the entire data acquisition process and provides a simple interface for engineering students to query. To access sensor data, an engineering student only needs to send out a JSON-RPC request to the application server specifying constraints such as sensor name and time period. The application server will respond with an array of timestamp-value pairs representing sensor data within that period of time, as shown in step (7) and (8). Figure 2 shows an example of a JSON-RPC request for monthly precipitation data from the sensor *#49862* beginning from April 24, 2009. The query module in the middleware responds with sensor data packet in Table 1. JSON-RPC is a remote procedure call protocol similar to XML-RPC and SOAP but is encoded in a much simpler way.

```
{“method”: “SensorData Query”,
  “params”: {“SensorID”:49862,
             “AggregateType”:”sum”,
             “sDate”: “2009-4-24 00:00:00”,
             “eDate”: “2009-5-23 23:59:59”,
             “unit”: 86400},
  “id”: 132}
```

Figure 2: JSON-RPC request for monthly data.

Table 1: Sensor data packet.

SensorID	**Data array**
49862	[<2009-4-24, 0>,<2009-4-25, 0>, <2009-4-26, 0>,<2009-4-27, 0>, <2009-4-28, 0>,<2009-4-29, 0>, <2009-4-30, 0>,<2009-5-1, 0>, <2009-5-2, 0>,<2009-5-3, 0>, <2009-5-4, 17.8>, <2009-4-25, 35.2>, …, <*timestamp*$_n$, *value*$_n$>, … <2009-5-23, 0>]

Readiness of large-scale sensor data and a full-fledged query implementation allows engineering students and faculty to access sensor data of a certain type, at a specified location, during a specific period of time. It enables emergence of new models of educational practices in engineering classrooms. For example, students can apply statistical analyses to study the impact of a forest fire on the environment; students can inspect data of deep-sea sensors to document the effect of global warming over the past decade; students can design and validate their traffic control systems based on the real-world traffic flow data. In the process of implementing the middleware, we notice three major challenges:

(1) Wide inclusion of sensors and sensor providers.

To guarantee the broadest inclusion of sensors in the world, we must adopt a standard and prevalent specification for sensor information description. Such a specification must have sufficient descriptive power to document most of the existing sensors deployed in the world. In the meantime, we intend to tailor the specifications to only describe sensor data without considering other parameters such as transmission rate and hardware model number. To achieve this goal, we derive a sensor description template from Sensor Model Language (SensorML) [14] to characterize the output data of a sensor. Only the outputs and location sections in SensorML specification pertain to this project because they contain properties that sufficiently describe the output data of a sensor. These properties include: sensor name, sensing type, location, unit of measurement, data type, value domain, time domain, and textual description of a sensor.

(2) Sensor data integration.

In contrast to the specifications for sensor information that must contain sensor-dependent information, the unified format of sensor data must contain only the essence of all sensor data without any sensor-dependent information. This ensures seamless data integration from sensors across multiple data providers. For example, if two data providers both publish precipitation data but follow different data structures and precision, the middleware must be able to convert all this data into a single format. Sensor identifier, timestamp, and value are three indispensable elements regardless of what the sensor measures, who owns the sensor, or how the sensor data is originally structured. As a result, the unified format of sensor data in this project is modeled as a triple <*Sensor*, *Timestamp*, *Value*>. In Figure 1, the adapter module is used to transform data into this unified format. Then the collector module saves the sensor data to the corresponding data table. Sensor data of the same sensing type will be collected into one table to optimize the query execution.

(3) Query processing in a large problem space.

The large data space presented by the project proves difficult for engineering students because students have no good concepts of where to start the problem solving process. Even given a query portal, it is still not clear what can be queried and which sensor should be selected. To provide more guidance for first-time users, the query module in our middleware supports a composite query of geological information, sensing types, sensors, and sensor data. For example, given a geological rectangular area, our application server returns the available sensing types in the area. Then if a sensing type is further specified in the request, the query module can return the distribution of sensors of this type available in the area. Given a specific sensor and a period of time, it returns an array of timestamp-value pairs representing the data collected by this sensor within the specified period of time.

Another critical problem in handling a large problem space is the undesirable processing time. It is often infeasible to send all data back to engineering students when they request an extremely high volume of data. The Quality of Service (QoS) will experience a significant decline because of the costly processing and transferring time and unaffordable bandwidth consumption. To solve this problem, the query module supports user-specified aggregation type and aggregation unit. Aggregation type indicates how aggregate values are calculated. Our project currently supports four preset aggregation types: *sum*, *average*, *max*, and *min*. Aggregation unit defines the unit of time within which one aggregate value will be produced to represent all data collected during this time unit. In the example shown in Figure 2, although the sensor collects data every five minutes, the client requests the query module to group monthly data into a list of daily (86400 seconds) data. Within each day, all sensor data will be aggregated to calculate only one sum value.

Engaging user interface for accessing large-scale sensor data.

Our middleware simplifies the process of sensor data acquisition using a simple JSON-RPC request and response. However, availability of a large amount of sensor data does not directly contribute to effective usage of sensor data. We must implement appropriate visualizations to help engineering students better understand the sensor data. In this section, we first share our experience in developing multiple visualizations for better data representation. Then we describe our interactive interface design on iPhone/iPod Touch.

Sensor data visualization primarily involves rendering sensor distribution on the map and visualizing <*Sensor*, *Timestamp*, *Value*> triples in line charts and tables. In Figure 3(b), red dots on the map denote the space distribution of only precipitation sensors available in the area. The darkness indicates different amount of rainfall during the period of Apr 24, 2009 to May 23, 2009. The line chart shows the time distribution of how precipitation fluctuated within a month, corresponding to the data in Table 1. The visualizations utilizing a map and a line chart help understand time-space distributional characteristics of environmental metrics. Users can also select more precise data views by looking at a table view that lists all time-value pairs numerically.

User interaction with our client allows manipulation of geological area by panning and zooming on the map, selection of multiple sensing types from a candidate list (Figure 3(a)), choosing a sensor by clicking on the map, switching the length of observation time frame to a year, a month, a week, a day, or four hours, turning to previous/next time frame by panning right/left on the line chart, and picking a specific value from a sensor at a certain time (Figure 3(b)). If the user operations require new data, a JSON-RPC request is generated and sent to the query module on the middleware. Once new data is returned, the interface will be redrawn to reflect the changes.

(a) **(b)**

Figure 3: (a) Selecting sensing types from a list;
(b) Visualizing precipitation data of a sensor in Bamberg, SC, USA

We target our applications to mobile devices because of the encouraging and ready-to-use nature of mobile applications. We choose iPhone and iPod Touch because it provides higher computing capacity, a larger display, and more innovative input peripherals and techniques than other mobile devices. Data Logger [15] is an iPhone application that aims to visualize data from sensors owned by users and allows data sharing across multiple users. Since data accuracy is neither guaranteed by the sensor owner nor by Data Logger, the appropriateness of this service for educational purposes is questionable.

Implications

Availability of large-scale real-world sensor data provides a chance to innovate practice-oriented engineering education. For engineering students needing hands-on experience in sensor hardware and data acquisition, they have an opportunity to handle practical problems of higher complexity and are more aware of the social context where the sensors are situated. For example, students can compare their sensor deployment plan with the sensor network in industry to gain more real-world experience. For engineering students who develop applications concerning sensor data, they can focus more on their specialized projects rather than being distracted by fundamental problems of data acquisition. For example, students in environmental engineering can study the environmental impact of the forest fire in Los Angeles, without having to learn how sensors work and how to acquire sensor data. Engineering students showing interest in sensing science will have a chance to gain an intuitive sense of sensor data. We offer a client on iPhone/iPod Touch to give students an engaging experience. For example, a high school student may discover the basic theory of tides by observing sensor data of water level.

Conclusion

In this paper, we propose a new approach to incorporating large-scale real-world sensor data into engineering learning environments. We implement a middleware framework in the application server to simplify the process of sensor management and data acquisition. Introducing data from a large variety of real-world sensors becomes possible and affordable. Currently, we support a total of 1136 sensors from a variety of sources such as SCDOT [16], NOAA CO-OPS [17], and Intelligent River [18]. These datasets include various measures of air temperature, water temperature, water level, wind speed, air pressure, precipitation, conductivity, and soil moisture. We also develop a client on iPhone/iPod Touch to provide intuitive visualization and effective navigation of sensor data.

The SensorWorld project opens a new opportunity to bring practice-oriented education into engineering classrooms. By sending requests to our application server, students will be able to access real-world sensor data and develop their own applications based on the data. Engineering students can also navigate learn to navigate extremely large datasets using a single iPhone/iPod Touch application.

Acknowledgement

This project is supported through National Science Foundation Grant EEC-0956819.

Bibliography

1. Arsenault, J., et al. *Integration of sensors into secondary school classrooms*. in *Frontiers in Education, 2005. FIE '05. Proceedings 35th Annual Conference*. 2005.

2. Mahonen, P., E. Meshkova, and J. Riihijarvi. *A novel partnership of a school and a university: using the work of university students to enhance science teaching and to foster interest to technology in K-12 schools*. in *Frontiers in Education Conference, 2008. FIE 2008. 38th Annual*. 2008.

3. Sobhan, S., et al., *Modern Sensing and Computerized Data Acquisition Technology in High School Physics Labs.* 2007. **23**: p. 902-909.

4. Kashlan, A.E.l. and S.E.l. Kashlan, *Instrumentation and measurement course for non-electrical engineering students.* 2006. **43**: p. 173-184.

5. Mishra, S., Y. Pan, and T. Oh. *Undergraduate Research in Wireless Sensor Security Course*. in *Proceedings of the 13th Colloquium for Information Systems Security Education*. 2009. Seattle, WA.

6. Meshkova, E., J. Riihijarvi, and P. Mahonen. *Teaching embedded systems with active learning: The SMEAGOL approach*. in *Frontiers in Education Conference, 2008. FIE 2008. 38th Annual*. 2008.

7. Li, S. and A.A. Khan, *Applying IT Tools to a Laboratory Course for Measurement, Analysis, and Design of Electric and Electronic Circuits.* Education, IEEE Transactions on, 2005. **48**(3): p. 520-530.

8. Perles, et al., *Design and Application of a Data Acquisition Card Simulator to Electronic Engineering Studies.* 2008. **24**: p. 825-832.

9. Nagurney, L.S. *The evolution of a bioinstrumentation course*. in *Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE '07. 37th Annual*. 2007.

10. Sumali, H., *Developing a Laboratory Course in Sensors and Data Acquisition for Agricultural Engineering.* 2006. **22**: p. 1231-1240.

11. Reisslein, J., P. Seeling, and M. Reisslein, *Computer-Based Instruction on Multimedia Networking Fundamentals: Equational Versus Graphical Representation.* Education, IEEE Transactions on, 2005. **48**(3): p. 438-447.

12. Jenkins, H. and J. Burtner, *An Approach to Develop and Measure Engineering Visualization in an Introductory Mechanics Course using Computer-Aided Learning Modules.* 2007. **23**: p. 150-161.

13. Arruarte, A., U. Rueda, and J.A. Elorriaga. *Organizing the learning resources related to the subject Introduction to Artificial Intelligence through Concept Maps*. in *Frontiers in Education Conference, 2008. FIE 2008. 38th Annual*. 2008.

14. *Sensor Model Language (SensorML)*. Available from: http://vast.nsstc.uah.edu/SensorML.

15. *Pachube.apps*. 2010; Available from: http://apps.pachube.com/datalogger/.

16. *South Carolina Department of Transportation*. Available from: http://www.dot.state.sc.us.

17. *National Oceanic and Atmospheric Adminstration CO-OPS*. Available from: http://opendap.co-ops.nos.noaa.gov/.

18. *Intelligent River*. Available from: http://www.intelligentriver.org/.

Biographical Information

Hanjun Xian is a Ph.D student in the School of Engineeirng Education at Purdue University. He focuses on the role of ultra large-scale data mining and visualization in engineering education research. Hanjun holds a MS and BS in computer science and engineering.

Krishna Madhavan is an Assistant Professor in the School of Engineering Education at Purdue University. Prior to his arrival at Purdue, he was an Assistant Professor with a joint appointment in the School of Computing and the Department of Engineering and Science Education at Clemson University. Dr. Madhavan also served as a Research Scientist at the Rosen Center for Advanced Computing, Information Technology at Purdue University where he led the education and the educational technology effort for the NSF-funded Network for Computational Nanotechnology (NCN). He was the Chair of the IEEE/ACM Supercomputing Education Program 2006 and was the curriculum director for the Supercomputing Education Program 2005. In January 2008, he was awarded the NSF CAREER award for work on learner-centric, adaptive cyber-tools and cyber-environments. He was one of 49 faculty members selected as the nation's top engineering educators and researchers by the US National Academy of Engineering to the Frontiers in Engineering Education symposium.