Measuring our environment: Introducing physical sensors

Problem statement

Sensors play a critical role in generating signals that provide us with information about local systems and, in a collective sense, about the world. There is a large amount of data that can be gathered from our surroundings and used to analyze nearly anything?from the flow of people to the quality of public water and air to the stability of a bridge under load.
The recent proliferation of the Internet of Things (IoT) and the growing ubiquity of Smart devices provides the data scientist with an unprecedented perspective of the environment, its constituent elements, and behavior.

Project goal: The goal of this project is (1) to learn that data does not always come from the 'Cloud' and how you can gather meaningful data yourself, and (2) to appreciate what happens when the data does not come out looking perfect, when there is error involved. Data collection is achieved through a process called transduction, wherein sensors act to convert a physical phenomenon into an electrical signal, which, depending on how it is manipulated, then constitutes one or another class of data. The complexity of the resulting data can be tied to the mechanistic basis for transduction, the parameters used in sampling or conditioning the signal, and the inherent stochasticity and interdependence of nature.

Physical Device

You will be given a premade, predominantly single-ended embedded system that features sensors to track temperature, humidity, pressure, light, and motion. The device will not only log the corresponding data against time, but also permit control over data collection parameters.

Sensor Package Specifications:

- Temperature (-20 to $60°C$ ± 0.2)
- Humidity (20 to 95% RH with resolution 0.1%)
- Pressure (300 - 1100 hPa ± 1 hPa)
- Light sensor (cadmium sulfide photoresistor AFEC)
- Motion sensor(s) (passive infrared detector with 150 cm over $100°$ range)
- Date and time (not constantly but in starting points)

Each team is to place their sensor (or two sensors if available) in a spot of their choice inside the room of one of their team member's Harvard Dorm or place of residence if not staying in campus. You will set it to sample time, humidity and temperature, motion, and light in specific time intervals. These sampling rates will be different for each sensor. For motion, for example it makes sense to sample every 1 sec, but for light every 5 min. Think of what the sampling rate for temperature should be (Hint: 30 min?)

Data Resources

You will get the data off the sensor in an SD card. The collection time is **1 month**. Make a csv with the following variables:

- **sensorcode**: "first 3 letters of dorm name+CS109aGroupNumber+"-1" e.g. CAN025-1 (if sensor is not in dorm, letters should be address, e.g. BRA for Brattle str.)
- **temp**
- **hum**
- **press**
- **light**
- **motion**
- **time_string**: deducting this needs minor calculations since we only have the starting time.

High-level project goals

There are a lot of questions that can be answered using the above data. The high-level goal of this project is to predict a) the outside temperature, and b) the time of day or day of the week, using indoor measurements. Below you will find some intermediate questions that you need to answer towards this goal.
Your team will have a meeting with an engineer from SEAS to pick up the sensors and learn to use them. Meeting TBD (most probable time Wed. Oct. 9th at 3:00pm.)

1. The first step is to construct your dataset from all the data you have accumulated for the duration of a **month**. Perform EDA. Ideas: plot the variables against time of day and day of the week, calculate min, max, and mean for each day for the variables for which this makes sense, eg. temp, pressure, and humidity. For measured light what would make sense? Note any extreme values as outliers and suggest ways of dealing with them.
2. Determine what data are important and what data might have noise or interference. For instance, how would you determine the difference between sunlight and artificial light?
3. Plot averaged temperature curves during the day and look for patterns of overheating suggesting a suboptimal usage of the HVAC system and therefore wasted energy.
4. Does light correlate with activity, i.e. do students leave lights on when away?
5. Make a model to predict a) the outside temperature, and b) the time of day or day of the week.
6. If you were to choose a plant to decorate your dorm, which one would you choose based on room conditions?

Authors

Eleni Kaxiras (cs109a Lab Instructor) and Evan Smith (SEAS Active Learning Labs)

References

1. Instructions (coming up)