

Information Fusion for Environmental Health Assessment

Who?

Hawk Weisman

`hawk@hawkweisman.me`

`http://hawkweisman.me`

From?

Department of Computer Science
Allegheny College

When?

March 6th, 2015

Problem

- Assessing the health of an ecosystem...
 - 1 ...is difficult
 - 2 ...is time-consuming
 - 3 ...requires specially-trained professionals

Problem

- Assessing the health of an ecosystem...
 - 1 ...is difficult
 - 2 ...is time-consuming
 - 3 ...requires specially-trained professionals
- How to **prioritize** regions for assessment?

Solution

- Summarize and classify information from multiple sources into a single score

Solution

- Summarize and classify information from multiple sources into a single score
- 1 Discretize region into cells

Solution

- Summarize and classify information from multiple sources into a single score
 - 1 Discretize region into cells
 - 2 Machine learning:

Solution

- Summarize and classify information from multiple sources into a single score
 - 1 Discretize region into cells
 - 2 Machine learning:
- Assign scores to known healthy and unhealthy sites

Solution

- Summarize and classify information from multiple sources into a single score
- 1 Discretize region into cells
- 2 Machine learning:
 - Assign scores to known healthy and unhealthy sites
 - Train classifier with scores (k -fold cross-validation)

Solution

- Summarize and classify information from multiple sources into a single score
 - 1 Discretize region into cells
 - 2 Machine learning:
 - Assign scores to known healthy and unhealthy sites
 - Train classifier with scores (k -fold cross-validation)
 - 3 Information fusion:
 - Develop error model for sensors

Solution

- Summarize and classify information from multiple sources into a single score
 - 1 Discretize region into cells
 - 2 Machine learning:
 - Assign scores to known healthy and unhealthy sites
 - Train classifier with scores (k -fold cross-validation)
 - 3 Information fusion:
 - Develop error model for sensors
 - Determine value of information in different situations

Solution

- Summarize and classify information from **multiple sources** into a single score

Solution

- Summarize and classify information from **multiple sources** into a single score
 - 1 Inexpensive sensors

Solution

- Summarize and classify information from **multiple sources** into a single score
- 1 ■ **Inexpensive sensors**
- **Soil:** moisture, pH, temperature

Solution

- Summarize and classify information from **multiple sources** into a single score

- 1 **Inexpensive sensors**

- **Soil:** moisture, pH, temperature
- **Water:** pH, dissolved oxygen, temperature

Solution

- Summarize and classify information from **multiple sources** into a single score

- 1 **Inexpensive sensors**

- **Soil:** moisture, pH, temperature
- **Water:** pH, dissolved oxygen, temperature
- **Atomosphere:** gas pollutant sensors

Solution

- Summarize and classify information from **multiple sources** into a single score
 - 1 Inexpensive sensors
 - **Soil:** moisture, pH, temperature
 - **Water:** pH, dissolved oxygen, temperature
 - **Atomosphere:** gas pollutant sensors
 - 2 **Remotely-sensed data**

Solution

- Summarize and classify information from **multiple sources** into a single score

1 Inexpensive sensors

- **Soil:** moisture, pH, temperature
- **Water:** pH, dissolved oxygen, temperature
- **Atomosphere:** gas pollutant sensors

2 Remotely-sensed data

- **Satellite imagery:** visual, IR

Solution

- Summarize and classify information from **multiple sources** into a single score

1 Inexpensive sensors

- **Soil:** moisture, pH, temperature
- **Water:** pH, dissolved oxygen, temperature
- **Atomosphere:** gas pollutant sensors

2 Remotely-sensed data

- **Satellite imagery:** visual, IR
- Aerial photography

Solution

- Summarize and classify information from **multiple sources** into a single score

1 Inexpensive sensors

- **Soil:** moisture, pH, temperature
- **Water:** pH, dissolved oxygen, temperature
- **Atomosphere:** gas pollutant sensors

2 Remotely-sensed data

- **Satellite imagery:** visual, IR
- Aerial photography

3 Past records

Solution

- Summarize and classify information from **multiple sources** into a single score
 - 1 Inexpensive sensors
 - **Soil:** moisture, pH, temperature
 - **Water:** pH, dissolved oxygen, temperature
 - **Atomosphere:** gas pollutant sensors
 - 2 Remotely-sensed data
 - **Satellite imagery:** visual, IR
 - Aerial photography
 - 3 **Past records**
 - Temperature, rainfall, etc

Solution

- Summarize and classify information from multiple sources into **a single score**

Solution

- Summarize and classify information from multiple sources into **a single score**
 - 1 Use classifier to generate health score

Solution

- Summarize and classify information from multiple sources into **a single score**
 - 1 Use classifier to generate health score
 - 2 Output to GIS, produce heat maps

Solution

- Summarize and classify information from multiple sources into **a single score**
 - 1 Use classifier to generate health score
 - 2 Output to GIS, produce heat maps
 - 3 Prioritize least-healthy areas for investigation

Evaluation

- Testing the classifier

Evaluation

- Testing the classifier

- 1 Simulated inputs

Evaluation

■ Testing the classifier

- 1 Simulated inputs
- 2 Real sites with expected scores

Evaluation

■ Testing the classifier

- 1 Simulated inputs
- 2 Real sites with expected scores
- 3 k -fold cross-validation

Evaluation

- Testing the classifier

- 1 Simulated inputs
- 2 Real sites with expected scores
- 3 k -fold cross-validation

- Assessing the Whole System

Evaluation

- Testing the classifier

- 1 Simulated inputs
- 2 Real sites with expected scores
- 3 k -fold cross-validation

- Assessing the Whole System

- 1 Release to environmental health organizations and professionals

Evaluation

- Testing the classifier

- 1 Simulated inputs
- 2 Real sites with expected scores
- 3 k -fold cross-validation

- Assessing the Whole System

- 1 Release to environmental health organizations and professionals
- 2 Collect feedback and assessments

Challenges

Challenges

1

Cost: may be expensive at scale

Challenges

1

Cost: may be expensive at scale

2

Computationally Expensive: running classifier over large data sets

Challenges

- 1 **Cost:** may be expensive at scale
- 2 **Computationally Expensive:** running classifier over large data sets
- 3 **Intedisciplinary:** need input from other fields

Questions?