

Auto Data Feedback and Noise Reduction Advisor : Paul Scerri **RISS** Yi Yang

Introduction

The cooperative watercraft is autonomous environmental monitoring robot with little human interaction.

Environmental Monitoring

Sample Analysis





Problem:

Give timely feedback to our data processing intermediate.

Filter data noise from the raw data, removing corrupt data points and useless out-of-river data points.

Method

About information feedback:

Adding an email sending function in android app immediately after exiting.

About mutated data filtering:

Self-adapted DBSCAN **Clustering Algorithm**

• Self-generate corresponding input parameters of DBSCAN algorithm to different data sets.

Run DBSCAN on data set

Time-related Algorithm • Find a better W matrix which can mostly smooth noise without compromising normal data

► X-revised data matrix Y-raw data matrix W-modifying matrix

 $\arg\min\left(\sum_{i}\left\|\vec{x_{i}}-\vec{y_{i}}\right\|^{2}+\sum_{i}w_{i,j}\left\|\vec{x_{i}}-\vec{x_{j}}\right\|^{2}\right)$

Results

Auto email sending. After app quits, it grabs latest log file and sends data as an email attachment. An extra button is also added to allow testers to report anytime they want.



1

2

-5

◄ figure on the left shows the interface of airboat application.

► figure on the right shows the e-mails received once app exits or report button is clicked.

🗌 🕁 📙 me	file from airboat -
🗌 🕁 📙 me	file from airboat -
🗌 🕁 📙 me	file from airboat -
🗌 🕁 📙 me	file from airboat
🗆 🕁 📙 me	file from airboat
🗆 🕁 📙 me	file from airboat
🗆 🕁 📄 me	file from airboat

EC value filtering. It can automatically grab the email attachment and get the log file as input file. Then, It can filter the log file and get the time and EC(eletrical conductivity) value. Theoretically, EC value remains 0 when the boat is out of water, and jumps to certain value once it is put into water.



With self-adapted DBSCAN, the noise can be filtered and also the data set is clustered into two or three clusters. Each color is a cluster, as shown in figure 1.2, 2.2. Especially if the in-river data is much more than the out-of-river data, this algorithm can even filter the data head and tail, which is shown in figure3.2.

The vertical lines are the groud truth of time when the boat was taken in or out of the river.



▲ With time-related algorithm, noise can be revised to a regular value which pretty much like the values of its normal neighbors with modifying W matrix(see above). Up to now, a W matrix which contains the information of mutation of noise has been being tested. Looking forward to the final result.





Conclusion

About information feedback: The application successfully sends an email once it quits the program. About corrupt data filtering:

Clustering algorithm works well on the EC value data set. It can cluster the data set into three clusters. It is limited to data slowly rising which should be viewed as normal data. Time-related algorithm can predict noise to a protential value instead of roughly removing it. So continuous information of data set can be saved. Significance:

Manual data feedback and noise reduction work can be avoided.

Future Work

Go deep in time-related algorithm: A better W matrix is needed to adjust noise to normal level. Deep filtering for the in-river data Filter the noise which cannot easily be visibly distinguished.



