

Using Data Mining Technique to Discover Useful Salinity/Temperature Patterns in ARGO Data

Yo-Ping Huang and Li-Jen Kao



1. Introduction

This research presents a novel data mining technique for finding interesting spatial-temporal patterns in ARGO ocean data. The ARGO data consists of time series measurements for upper ocean salinity/temperature data. Extracting spatial-temporal patterns of interest from ocean variables can be of importance for understanding the relations between ocean salinity/temperature structure and climate variability. Data mining techniques are used to facilitate the automatic extraction and analysis of interesting patterns in huge datasets.



2. Why Data Mining

Traditional Method
Statistical Spatial Analysis

- Need assumptions regarding to statistical independence of spatial distribution.
- Need domain experts to do the analysis work.

Data Mining



- Extracting interesting, implicit, and previously unknown knowledge from large amounts of data.
- No need of domain experts to do the mining work.
- Association rules mining can be applied to spatial-temporal patterns mining.

3. The Challenges of Salinity/Temperature Patterns Mining

- Association rules mining is applied to explore salinity/temperature patterns in the ARGO data.
- We need to find salinity/temperature patterns that incorporate time and location information.
- For example: "If area A salinity rises from 0.05psu to 0.15psu, then area B temperature will rise from 0.4°C to 1.2°C in the next month."
- A salinity/temperature patterns mining framework is developed that is based on traditional association rules mining algorithm.
- Fig. 1 shows the challenges of salinity/temperature patterns mining work.

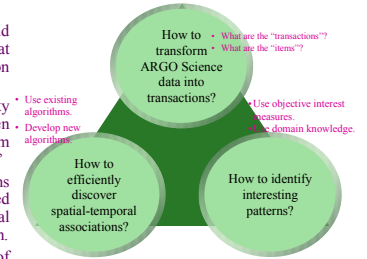


Fig. 1

4. The Research Approach

- A multi-dimensional inter-transaction association rules mining algorithm is developed to find salinity/temperature patterns that incorporate time and location information.
- The proposed mining algorithm is developed based on traditional association rules mining algorithm. However, the traditional association rules mining algorithms are usually time-consuming, especially with huge amount of items.
- In order to promote the performance, the number of spatial features need to be reduced.
- An efficient way to form inter-transactions without increasing mining space is need.
- Fig. 2 shows the mining framework in this research.

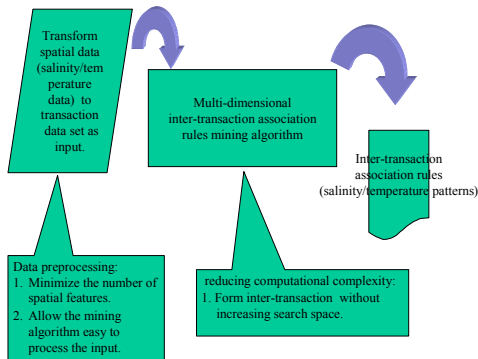
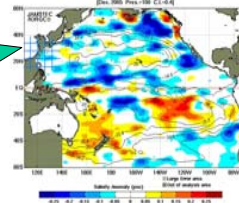


Fig. 2

5. Research Results

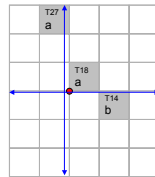
A step by step explanation of how salinity/temperature patterns are discovered.

1) Identify abnormality events in spatial data set.

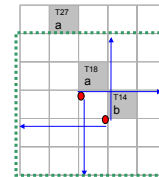


One of the monthly mean temperature data. These data were collected and made freely available by the International Argo Project and the national programmes that contribute to it. (<http://www.argo.ucsd.edu>, <http://argo.jco.monps.org>). Argo is a pilot programme of the Global Ocean Observing System."

3) Form inter-transactions using MBC (Minimum Bounding Cube) and max-span.



1-itemset MBC and max-span=3



2-itemset MBC and max-span=3

2) Treat a single event as a transaction.

- The location and occurring time for each salinity/temperature abnormality need to be recorded.

Tid	Event	Location and time
T1	r	<3, 0, 1>
T2	a	<14, 2, 2>
T3	a	<2, 2, 2>
T4	d	<7, 2, 3>
T5	b	<13, 2, 3>
T6	d	<15, 2, 3>
T7	b	<1, 3, 4>
T8	d	<3, 3, 5>
T9	c	<11, 4, 5>
T10	d	<8, 5, 5>
T11	e	<14, 6, 7>
T12	a	<9, 8, 8>
T13	a	<4, 9, 9>
T14	a	<12, 9, 9>
T15	b	<3, 10, 10>
T16	d	<8, 10, 10>
T17	b	<11, 10, 10>
T18	c	<14, 12, 11>
T19	b	<14, 12, 11>
T20	a	<6, 14, 12>

4) Results and an explanation for one of the 3-event frequent itemsets:

Event explanation:
TDL=temperature dropped little. TDB=temperature dropped much.
SRL=salinity rose little. SRB=salinity rose much.
SDL= salinity dropped little. SDB=salinity dropped much.

2-event frequent itemsets	{A(0,0)<TRL>, A(2,0)<SRL>} {A(0,0)<TRL>, A(2,0,1)<SRL>} {A(0,0)<SDL>, A(0,1,1)<TDL>} {A(0,0)<TRL>, A(0,-1,0)<SRL>} {A(0,0)<TRL>, A(0,-1,1)<SDL>}
3-event frequent itemsets	{A(0,0,0)<TRL>, A(0,-1,1)<SDL>, A(0,-1,2)<TDL>} {A(0,0,0)<TRL>, A(0,-1,0)<TRL>, A(0,-2,1)<SDL>} {A(0,0,0)<TRL>, A(1,0,1)<TRL>, A(3,0,2)<SRL>}

If temperature rose little in north-east ocean of Taiwan, and temperature rose little in east ocean of Taiwan the same time, then salinity will drop little next month.