The Sixteenth International Symposium on Artificial Life and Robotics 2011 (AROB 16th '11),
B-Con Plaza, Beppu,Oita, Japan, January 27-29, 2011

# Two-Level Time-Series Clustering for Satellite Data Analysis

Ayahiko Niimi
*Faculty of Systems Information Science*
*Future University Hakodate*
*116-2 Kamedanakano-cho, Hakodate-shi,*
*Hokkaido 041-8655, JAPAN*
*(niimi@fun.ac.jp)*

Takehiro Yamaguchi
*Graduate School of Systems Information Sciences*
*Future University Hakodate*
*116-2 Kamedanakano-cho, Hakodate-shi,*
*Hokkaido 041-8655, JAPAN*
*(g2109046@fun.ac.jp)*

Osamu Konishi
*Faculty of Systems Information Science*
*Future University Hakodate*
*116-2 Kamedanakano-cho, Hakodate-shi,*
*Hokkaido 041-8655, JAPAN*
*(okonishi@fun.ac.jp)*

***Abstract:*** In this paper, we propose a method for finding the frequent occurrence patterns and the frequent occurrence time-series change patterns from the observational data of a weather-monitoring satellite. The observational data of the weather-monitoring satellite are temporal and spatial large-scale data. However, to analyze this large amount of data incurs a high calculation cost. Therefore, we propose parallel computation when the frequent occurrence pattern and the frequent occurrence time-series change pattern are extracted at the Artificial Life and Robotics conference (AROB) 2010. In this paper, we apply the proposed system to Moderate Resolution Imaging Spectroradiometer (MODIS) data and discuss its results.

***Keywords:*** distributed processing, clustering, frequent occurrence pattern extraction, satellite data, data stream

## I. INTRODUCTION

In our network society, the development of information processing enables us to collect and utilize massive amounts of data; and data mining has gained attention as a technology to discover new knowledge and patterns. But those data are changing continuous, and new types of large-scale data have emerged. For example, records of financial and distributional transactions, telecommunications records, and network access logs are typical data streams. The term data stream suggests that the temporally changing, massive amounts of data records that are generated, accumulated, and consumed are looked on as flow of data (stream). In the real world, the requirement has been growing to elicit information from those large data streams whenever we need information. At first glance, data mining seems to be effective; but a data stream has the following dynamic properties:

1. massive amounts of data are
2. coming over a high-speed stream,
3. temporally changing, and
4. continue to arrive permanently;

and data mining is intended for static data, not a stream. Therefore, data stream mining technology has been developed to deal efficiently with large-scale data streams [1, 2, 3, 4, 5, 6, 7].

An example of a data stream for this technology is data from satellites. Satellite data are used for various purposes, such as land-cover classification and forecasts [8] and marine information analysis [9, 10, 11, 12, 13]. However, satellite data treated up to now as static data. Therefore, much computing time was required to analyze a large amount of data. In this paper, we propose a method to solve this problem by using distributed processing.

At the Artificial Life and Robotics conference (AROB) 2010, we propose a method for finding the frequent occurrence patterns and the frequent occurrence time-series change patterns from the observational data of a weather-monitoring satellite [14]. The observational data of the weather-monitoring satellites are temporal and spatial large-scale data. Various uses are possible such as forecasting marine resources by analyzing satellite data. However, there is an issue with respect to the calculation cost to analyze a large amount of data. Thus, we propose to use parallel computation when the frequent occurrence pattern and the frequent occurrence time-series change pattern are extracted in this paper.

Our proposed method is as follows. First, to extract the frequent occurrence pattern from satellite data, the necessary marine information is acquired by using the filter from satellite data. Next, the extracted marine data undergo clustering to merge similar data and are labeled. As a result, similar data are brought together for data with a spatial extension. The labeling data are re-clustered to the data group and re-labeled according to the degree of similarity between labels. As a result, similar data are brought together for data with a time extension. Finally, frequent events are extracted as the frequent occurrence pattern from the labeling data. Moreover, the frequent occurrence of the time-series pattern can be extracted as rules by detecing the change in the labeling data group. However, it takes computing time to analyze long-term data. Therefore, by dividing data and integrating the results, we propose to shorten the computing time by parallel computation of clustering and the frequent occurrence of the time-series pattern rule extraction. As for clustering and the frequent occurrence of the time-series change pattern extraction, parallel computation is possible by dividing data. The shorter computing time can be expected by division degree because each algorithm never influences the parallel calculation. Each algorithm was examined with regard to whether it was possible to it make parallel. Because clustering and extracting change patterns can be applied in parallel computing, we constructed the system with clustering of marine information and the extraction of the change pattern.

At AROB 2010, the frequent occurrence pattern and the frequent occurrence of the time-series change pattern of the sea surface temperature are extracted by using the sea surface temperature data, with the weather-monitoring satellite providing verification. In this paper, we use Moderate Resolution Imaging Spectroradiometer (MODIS) data, which includes the sea surface temperature (SST) and the concentration of chlorophyll-a (chl-a) to verify our proposed system and discuss its results.

## II. THE PROPOSED METHOD

At AROB 2010, we propose a method for finding the frequent occurrence patterns and the frequent occurrence time-series change patterns from the observational data of a weather-monitoring satellite [14].

Fig. 1 shows a flowchart of the proposed system.

The flow of the algorithm is shown below.

1. First, to extract the frequent occurrence pattern from satellite data, necessary marine information is acquired by using the filter from satellite data.
2. Next, the extracted marine data undergo clustering to merge similar data and are labeled.
3. The labeling data are re-clustered to the data group and re-labeled according to the degree of similarity between the labels.
4. Finally, the frequent events are extracted as the frequent occurrence pattern from the data.

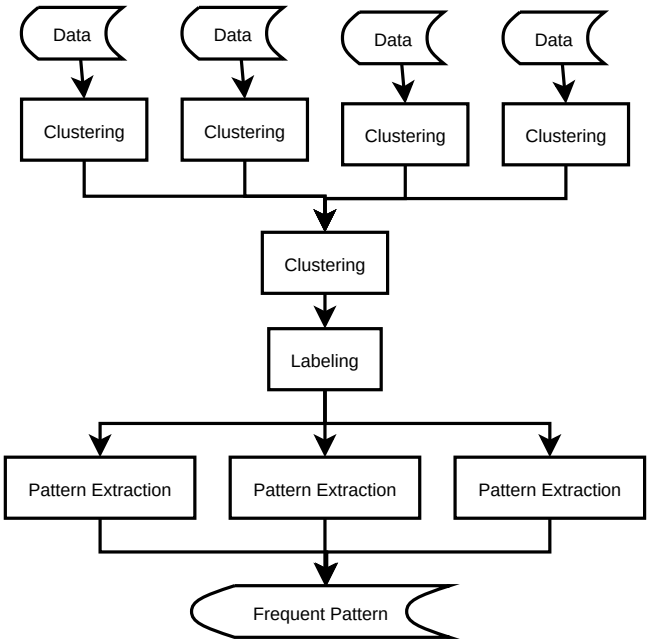


Fig. 1: Flowchart of proposed system

In the first clustering, similar data are brought together for data with a spatial extension. In the second clustering, similar data are brought together for data with a time extension. Moreover, the frequent occurrence of the Time-series pattern can be extracted as rules by detecting the change in the labeling data group. However, it takes computing time to analyze long-term data. Therefore, by dividing data and integrating the results, we propose to shorten the computing time by parallel computation of clustering and the frequent occurrence of the time-series pattern rule extraction. As for clustering and the frequent occurrence of the time-series change pattern extraction, parallel computation is possible by dividing data. The shorter computing time can be expected by division degree because each algorithm never influences the parallel calculation. Each algorithm is examined with regard to whether it was possible to make parallel. Because clustering and extracting of change patterns can be applied in parallel computing, we constructed the system with clustering of marine information and the extraction of the change pattern.

## III. DETAILS OF SATELLITE DATA

At AROB 2010, we acquire the data of Meteorological Satellite Center as satellite data. We use the observation monthly report of the Meteorological Satellite Cen-

ter for the experiment [15]. The observation monthly report can be obtained on CD-ROM. These CD-ROMs contain the monthly report of observation data derived from Multi-functional Transport Satellite (MTSAT-1R) and the polar orbital meteorological satellite from National Oceanic and Atmospheric Administration (NOAA). For the problem of forecasting the hot spot of marine products using satellite data, the sea surface temperature, the chlorophyll, and the flow of the ocean are often targeted. Because only the sea surface temperature data are included in the observational data, sea surface temperature is used in the experiment. Ten-day mean sea surface temperature consists of grid points arrayed every one degree latitude and longitude covering the area from 50 degrees North to 50 degrees South and 90 degrees East to 170 degrees West. We use data from March 2009 to May 2009 because the regression equation for SST was updated on March 1, 2009.

In this paper, we use MODIS data, which includes Sea Surgace Temperature(SST) and the concentration of Chlorophyl-a(Chl-a) to verify our proposed system. MODIS is a key instrument aboard the Terra (Earth Observing System (EOS) AM) and Aqua (EOS PM) satellites. Terra ' s orbit around the Earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. Terra MODIS and Aqua MODIS view the entire Earth 's surface every 1 to 2 days, acquiring data in 36 spectral bands, or groups of wavelengths [16, 17, 18, 19]. We use MODIS data from the Academic Frontier Promotion Center of Tokyo University of Information Sciences.

We used the publicly available data of SST and chl-a. The public data contain 30 days composite data, 5 days composite, and daily composite by PNG format, FLAT format, and HDF format. We convert from HDF data to CSV data by HDF utility commands. We use Hakodate-era data from January 2005 to September 2010.

## IV. EXPERIMENTS

In this paper, we try to extract the frequent occurrence pattern and the frequent occurrence of the time-series change pattern using SST data and chl-a with the weather-monitoring satellite for verification.

We use SST and chl-a data from MODIS data. We use linear interpolation with the data at a time before and after the missing value. The input data use nine attributes in which the data of eight neighborhoods are added to the data of a certain point. We already described the data in detail in section III..

We use the InTrigger platform of the information explosion project [20]. InTrigger is a distributed platform for information technology research for the Information Explosion Era. It is a cluster of clusters distributed across Japan. Weka is used for clustering [21].

The first step clusters data; and the effect of the distributed surrounding is examined. The result is being reasoned now.

## V. CONCLUSIONS

We proposed a method for finding the frequent occurrence patterns and the frequent occurrence time-series change patterns from the observational data of a weather-monitoring satellite at AROB 2010. The observational data of the weather-monitoring satellite are temporal and spatial large-scale data. Various uses are possible such as forecasting marine resources by analyzing satellite data. However, there is a problem with respect to calculation cost to analyze a large amount of data. Therefore, we proposed parallel computation when the frequent occurrence pattern and the frequent occurrence time-series change pattern are extracted. In this paper, we applied the proposed system to MODIS data, which includes the sea surface temperature and the concentration of chlorophyl-a to verify our proposed system.

## REFERENCES

[1] Martin H. C. L., Zhang, N., Anil, K. J. (2004), Nonlinear Manifold Learning For Data Stream. In Proc. SIAM International Conference for Data Mining, pp.34–44

[2] Jain, A., Zhang, Z., Chang, E. Y. (2006), Adaptive non-linear clustering in data streams. CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management, Arlington, Virginia, USA, pp.122–131

[3] Graf, H. P., Cosatto, E., Bottou, L., Durdanovic, I., Vapnik, V. (2005), Parallel support vector machines: The cascade svm. In Advances in Neural Information Processing Systems, pp.521–528

[4] Zhang, Y., Jin, X. (2006), An Automatic Construction and Organization Strategy for Ensemble Learning on Data Streams. SIGMOD Record, Vol.35, No.3, pp.28–33

[5] Wang, H., Fan, W., Yu, P. S., Han, J. (2003), Mining Concept-Drifting Data Streams Using Ensemble Classifiers. SIGKDD'03, pp.226–235

[6] Yamaguchi, T., Niimi, A. (2009), Community Graph Sequence with Sequence Data of Network Structured Data. 5th International Workshop on Computational Intelligence & Applications (IWCIA2009), Hiroshima, Japan, pp.196–201

[7] Minegishi, T., Ise, M., Niimi, A., Konishi, O. (2009), Extension of Decision Tree Algorithm for

Stream Data Mining Using Real Data. 5th International Workshop on Computational Intelligence & Applications (IWCIA2009), Hiroshima, Japan, pp.208–212

[8] Yamaguchi, T., Noguchi, Y., Ichimura, T., Mackin, K.J. (2009), Applying Cluster Ensemble to Adaptive Tree Structured Clustering. 5th International Workshop on Computational Intelligence & Applications (IWCIA2009), Hiroshima, Japan, pp.186–191

[9] Mustapha, M. A., Saitoh, S. (2008), Observations of sea ice interannual varitions and spring bloom occurrences at the Japanese scallop farming area in the Okhotsk Sea using satellite imageries. Estuarine, Coastal and Shelf Science, 77, pp.577–588

[10] Zainuddin, M., Kiyofuji, H., Saitoh, K., Saitoh, S. (2006), Using multi-sensor satellite remote sensing and catch data to detect ocean hot spots for albacore (Thunnus alalunga) in the northwestern North Pacific. Deep-Sea Research II, 53, pp.419–431

[11] Iida, T., Saitoh, S. (2007), Temporal and spatial variability of chlorophyll concentrations in the Bering Sea using empirical orthogonal function (EOF) analysis of remote sensing data. Deep-Sea Research II, 54, pp.2657–2671

[12] Radiarta, I. N., Saitoh, S. (2008), Satellite-derived measurements of spatial and temporal chlorophyll-a variability in Funka Bay, southwestern Hokkaido, Japan. Estuarine, Coastal and Shelf Science, 79, pp.400–408

[13] Zainuddin, M., Saitoh, K., Saitoh, S. (2008), Albacore (Thunnus alalunga) fishing ground in relation to oceanographic conditions in the western North Pacific Ocean using remotely sensed satellite data. Fisheries Oceanography, Vol.17, No.2, pp.61–73

[14] Niimi, A., Yamaguchi, T., Konishi, O. (2010), Parallel Computing Method of Extraction of Frequent Occurrence Pattern of Sea Surface Temperature from Satellite Data. International Symposium on Artificial Life and Robotics (AROB 15th '10), Beppu, Oita, Japan: 4 pages (in CD-ROM)

[15] Meteorological Satellite Center Monthly Report, Meteorological Satellite Center.

[16] MODIS Website, http://modis.gsfc.nasa.gov/

[17] MODIS Near Real Time Data, Japan Aerospace Exploration Agency, http://kuroshio.eorc.jaxa.jp/ADEOS/mod_nrt/

[18] Earth Observation Research Center (EORC), Japan Aerospace Exploration Agency, http://www.eorc.jaxa.jp/index.php

[19] Academic Frontier Promotion Center of Tokyo University of Information Sciences. http://www.frontier.tuis.ac.jp/modis/frontier/index.html

[20] InTrigger, https://www.intrigger.jp/wiki/index.php/InTrigger

[21] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, H. Ian. (2009), The WEKA Data Mining Software. SIGKDD Explorations, Volume 11, Issue 1