The Fifteenth International Symposium on Artificial Life and Robotics 2010 (AROB 15th '10),
B-Con Plaza, Beppu,Oita, Japan, February 4-6, 2010

Parallel Computing Method of Extraction of Frequent Occurrence Pattern of Sea Surface Temperature from Satellite Data

Ayahiko Niimi
Department of Media Architecture
Future University-Hakodate
116-2 Kamedanakano-cho, Hakodate-shi,
Hokkaido 041-8655, JAPAN

Takehiro Yamaguchi
Graduate School of Systems Information Sciences
Future University-Hakodate
116-2 Kamedanakano-cho, Hakodate-shi,
Hokkaido 041-8655, JAPAN

Osamu Konishi
Department of Complex Systems
Future University-Hakodate
116-2 Kamedanakano-cho, Hakodate-shi,
Hokkaido 041-8655, JAPAN

Abstract

In this paper, we propose the method for finding the frequent occurrence patterns and the frequent occurrence time-series change patterns from the observational data of the weather-monitoring satellite. The observational data of the weather-monitoring satellite is temporal and spatial large-scale data. Various use like the forecast of marine resources is thought by analyzing the satellite data. However, there is a problem of the calculation cost to analyze a large amount of data. Then, we propose to use the parallel computation when the frequent occurrence pattern and the frequent occurrence time series change pattern are extracted in this paper.

Keyword:

distributed processing, clustering, frequent occurrence pattern extraction, satellite data, data stream

1 Introduction

In recently network society, the development of information processing technique enables us to collect and utilize massive amount of data, and the data mining is technology of discovery new knowledge and patterns in those data has been paid attention. But those data are changing from moment to moment, and have become new type large scale data. Record of financial and distributional transactions, telecommunications records and network access logs are typical examples, and those data are called data stream. By data stream, it is that the conditions temporally-changed massive amount of data record are generated, cumulative and consumed are looked on as flow of data (stream). In the real world, the requirement that whenever we need information, we want to elicit from those large scale data stream has been growing. At first glance data mining seems to be effective, but data stream has following dynamic properties:

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

and there is a limitation applied data stream to data mining intending static data. Data mining to efficiently deal in large scale data stream, therefore data stream mining technology has been developed [1, 2, 3, 4, 5, 6, 7].

Data from the satellite is a data stream. The satellite data is used for various usages. For instance, the land-cover classification and the forecast that uses the satellite data are researched [8], and the marine information analysis is done [9, 10, 11, 12, 13]. However, the satellite data had been handled up to now as static data. Therefore, there was a problem of taking the computing time to analyze a large amount of data.

In this paper, we propose the method solved by using distributed processing.

In this paper, we propose the method for finding the frequent occurrence patterns and the frequent occurrence time-series change patterns from the observational data of the weather-monitoring satellite. The observational data of the weather-monitoring satellite is temporal and spatial large-scale data. Various use like the forecast of marine resources is thought by analyzing the satellite data. However, there is a problem of the calculation cost to analyze a large amount of data. Then, we propose to use the parallel computation when the frequent occurrence pattern and the frequent occurrence time series change pattern are extracted in this paper.

Our proposal method is the following. First of all, to extract the frequent occurrence pattern from the satellite data, necessary marine information is acquired by using the filter from the satellite data. Next, the extracted marine data are applied clustering to marge similar data, and labeled. As a result, similar data are brought together for data with a spatial extension. The labeling data are re-clustering to the data group, and re-labeled with according to the degree of similarity between labels. As a result, similar data is brought together for data with a time extension. Finally, the frequent event 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 catching the change in the labeling data group. However, it takes the computing time to analyze long-term data. Therefore, we propose to shorten the computing time by the parallel computation of clustering and the frequent occurrence of the time series pattern rule extraction by dividing data and integrating the results. As for clustering and the frequent occurrence of the time series change pattern extraction, the parallel computation is possible by divide data. The more shorten computing time can be expected by division degree, because each algorithm never influences in the parallel calculate.

In this paper, each algorithm examined whether making to the parallel was possible. Because clustering and extracting of change pattern can be applied parallel computing, we cnstructed the system with clustering of marine information and the extraction of the change pattern. In this paper, the frequent occurrence pattern and the frequent occurrence of the time series change pattern of the sea surface temperature have been extracted by using the data of the sea surface temperature with the weather-monitoring satellite for the verification. The extraction result was compared with an existing technique, and shortening at the computing time was confirmed.

2 Details of Satellite Data

We acquired the data of Meteorological Satellite Center as satellite data. We used the observation monthly report of Meteorological Satellite Center to experiment [14]. The observation monthly report can be obtained with CD-ROM. These CD-ROMs contain the Monthly Report of observation data derived from MTSAT-1R and the polar orbital meteorological satellite NOAA. This Monthly Report contains image data observed by the following 4 channels and processed satellite product data from the observation data. 4 channels are IR:Infrared(10.3–11.3μm), VS:Visible(0.55–0.90μm), WV:Water Vapor(6.5–7.0μm) and SW:3.8 micron image(3.5–4.0μm).

In the problem of forecasting the hot spot of marine products using the satellite data, the sea surface temperature, the chlorophyl, and the flow of the ocean, etc. are often targeted. Because only the data of the sea surface temperature is included in the observational data, sea surface temperature was used to experiment.

Ten-day mean Sea Surface Temperature (SST) 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. SST is derived from brightness temperatures of infrared split-window channels (IR1 and IR2) and satellite zenith angle using multiple regression equation. In the table, SSTs are expressed in 0.1 degree centigrade by integer form (multiplied by 10) of three digits, e.g., 12.3 degree centigrade is expressed as 123. The marks "/" and "." show "land" and "no valid data" respectively. SST data are provided by text data in tabular form generated in Meteorological Satellite Center. These data are recorded in ASCII code and HTML format.

We use data from 2009-March to 2009-May, because the regression equation for Sea Surface Temperature is updated on March 1, 2009.

3 The Proposed Method

In this paper, we propose the method for finding the frequent occurrence patterns and the frequent occurrence time-series change patterns from the observational data of the weather-monitoring satellite.

Fig. 1 shows a flowchart of the proposed system.

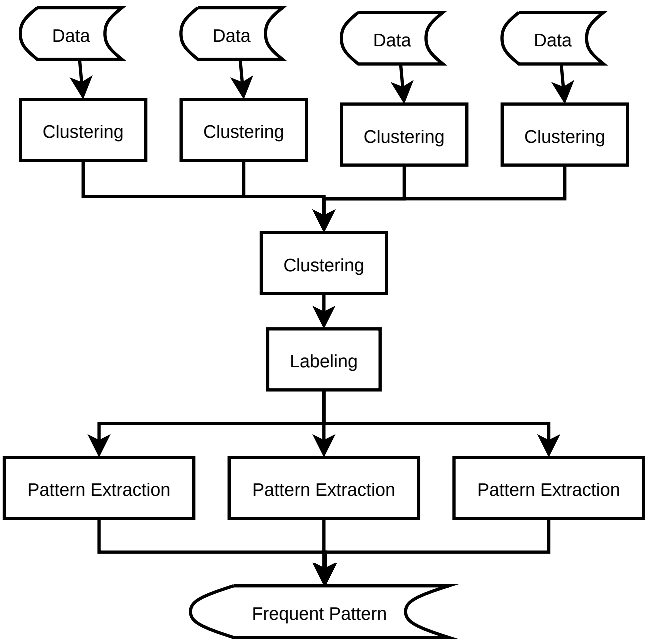

Figure 1: Frowchart of proposed system

The flow of the algorithm is shown below.

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

In first clustering, similar data are brought together for data with a spatial extension. In second clustering, similar data is brought together for data with a time extension. Moreover, the frequent occurrence of the time series pattern can be extracted as rules by catching the change in the labeling data group. However, it takes the computing time to analyze long-term data. Therefore, we propose to shorten the computing time by the parallel computation of clustering and the frequent occurrence of the time series pattern rule extraction by dividing data and integrating the results. As for clustering and the frequent occurrence of the time series change pattern extraction, the parallel computation is possible by divide data. The more shorten computing time can be expected by division degree, because each algorithm never influences in the parallel calculate.

In this paper, each algorithm examined whether making to the parallel was possible. Because clustering and extracting of change pattern can be applied parallel computing, we cnstructed the system with clustering of marine information and the extraction of the change pattern.

4 Experiments

In this paper, the frequent occurrence pattern and the frequent occurrence of the time series change pattern of the sea surface temperature have been extracted by using the data of the sea surface temperature with the weather-monitoring satellite for the verification.

The data used by the experiment is the sea surface temperature of the observation monthly report of Meteorological Satellite Center. We use linear interpolation with the data at time before and after for the missing value. The input data used 9 attributes in which the data of 8 neighborhood are added to data of a certain point. We already described detailed data in section 2.

We use InTrigger platform of the information explosion project [15]. InTrigger is a distributed platform for information technology research for Information Explosion Era. It is a cluster of clusters distributed across Japan.

In the experiment, we applied the first clustering, and it was confirmed to be able to shorten the computing time by the proposal technique.

5 Conclusions

In this paper, we propose the method for finding the frequent occurrence patterns and the frequent occurrence time-series change patterns from the observational data of the weather-monitoring satellite. The observational data of the weather-monitoring satellite is temporal and spatial large-scale data. Various use like the forecast of marine resources is thought by analyzing the satellite data. However, there is a problem of the calculation cost to analyze a large amount of data. Then, we propose to use the parallel computation when the frequent occurrence pattern and the frequent occurrence time series change pattern are extracted in this paper.

It is scheduled to conduct the evaluation experiment of the entire system including second-clustering and pattern extraction, and to experiment to data that contains the density of the chlorophyl in the future.

References

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

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

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

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

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

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

[7] Minegishi, T., Ise, M., Niimi, A., Konishi, O.: 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 (2009).

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

[9] Mustapha, M. A., Saitoh, S.: 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 (2008).

[10] Zainuddin, M., Kiyofuji, H., Saitoh, K., Saitoh, S.: 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 (2006).

[11] Iida, T., Saitoh, S.: 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 (2007).

[12] Radiarta, I N., Saitoh, S.: 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 (2008).

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

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

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