IJARCCE

aISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

A Survey on Identification of Dipoles in Climate Data

Ramireddy Thushara Bharathi[1], J. Velmurugan[2], Dr. M. Venkatesan[3]

M.Tech Scholar, Department of Computer Science and Engineering,

Sri Venkateswara College of Engineering and Technology, Chittoor, India[1]

Associate Professor, Department of Computer Science and Engineering,

Sri Venkateswara College of Engineering and Technology, Chittoor[2]

Research Scholar in VIT University, Vellore, India[2]

Associate Professor, School of Computing Science & Engineering, VIT University, Vellore, India[3]

Abstract: In this paper, the concept of climate with regarding to the dipoles formed and its types are illustrated. The concept of using various methodologies such as agglomerative methods in finding the dynamic dipoles and also various algorithms used to find the dynamic dipoles under parameters considered are also illustrated. This mainly involves data regarding the discovery of dipoles in climate data using various data mining techniques.

Keywords: Dipoles, Climate Indices, Agglomerative Method, Clustering Methods.

I. INTRODUCTION

a. Climate

- Climate can be defined as the prevalent weather conditions of an area, determined by latitude, position relative to oceans or continents.
- This can also be defined as the prevalent weather conditions of certain region including temperature, rainfall and pressure.
- The distance between climates of two places on the globe can be determined by Teleconnections [17].
- These teleconnections are detected by equivalent parameters of climate.

b. Dipole

- Dipoles [9] are termed as a pair of poles of same magnitude but of opposite sign with a finite distance apart representing distance between anomalies of two regions.
- When size of regions increases and become too large then the climate phenomenon gets disappeared and can be defined as a class of teleconnections characterized by the anomalies of opposite polarities at two locations at the same time.
- Dipoles are used in correlating the data. These are the pair of regions whose locations have strong negative correlation while locating in other region and strong positive correlation while locating in same region.

c. Types of Dipoles

- The dipoles are of different kinds based on the characteristics and locations.
  Based on the characteristics the dipoles can be classified as pressure dipoles that capture recurring and persistent large scale patterns of pressure and temperature dipoles that defines the sudden fluctuations occurred in the atmospheric temperature.
- Based on the locations the dipoles can be classified as Indian Ocean Dipole [10, 14] that defines the difference in sea level temperature between the western pole and the eastern pole and Arctic Dipole [24] that characterizes high pressure on arctic regions of North America and low pressure on the Eurasia region.
- Among all the kinds of dipoles [18] these are the two major kinds of dipoles that are often used as Indian Ocean Dipole and Arctic Dipole.

d. Parameters

- The dipoles are found for various parameters as Pressure, Temperature, Rainfall, Droughts, Tropical cyclones and Hurricanes.
- These extreme weather events generally depend on the seasonality, high variability and autocorrelation.

e. Climate Indices

- The climate Indices are used to study the impact of climate phenomenon on the weather events.
- These define a calculated value to describe the state and change in climate system. There are many climate Indices [21] as,

North Atlantic Oscillation (NAO), the foremost climate Indices of atmosphere defined that depends on the sea level pressure (SLP). This is one of the best Pressure Dipoles.

El Niño-Southern Oscillation (ENSO) [3], the global coupled ocean atmosphere phenomenon used to define the precipitation and temperature anomalies [13]. This also defines reduced hurricane activities.

Antarctic Oscillation (AAO) and Arctic Oscillation (AO) are used in calculating the Sea Level Pressure anomalies by Empirical Orthogonal Function (EOF) [20] analysis.

Southern Oscillation Index (SOI) is termed as extensions for El Niño events in Pacific Ocean. This is also one of the best Climate Indices in calculating the Pressure Dipoles.
Pacific Decadal Oscillation (PDO) is used to display warm or cool surface waters in the Pacific Ocean.
Interdecadal Pacific Oscillation (IPO) is used to display the similar Sea Level Temperature and Seal Level Pressure of PDO analysis in particular time period.
Palmer Drought Severity Index (PDSI) [22] and Crop Moisture Index (CMI) are the indices estimating the relative dryness and wetness affecting water sensitive economies through the available temperature and precipitation data.

f. Data Mining Techniques

- There are many techniques used in data mining as classification, Association, Regression, Cluster Analysis, Outlier Analysis and so on. Among these cluster analysis technique [19] is here mostly used.
- Clustering [7] belongs to unsupervised learning Techniques which is used to depict the creation of clusters for considered data.
- Clustering methods [23] can be classified into various categories as,
  - Partitioning Methods
  - Hierarchical Methods
  - Density - based methods
  - Grid - based methods
  - Model - based methods

II. METHODOLOGIES USED

There are many different methodologies used in detection of dipoles. They include algorithms for classifying and clustering the data into communities and finding the distances in between them.

a. Shared Nearest Neighbor Approach
The SNN approach [2] is used to find the similarity between individual data points. This identifies the size, shape and densities of dimensional data sets. This initially finds the nearest neighbors for each and then finds the similarity in between two different points. This avoids the cluster to be affected from the noisy and outliers and generates reliable clusters.

b. A1 Algorithm
This algorithm is used to in finding the dipoles through Nearest Neighbor approach. This collects the correlated locations and these related ones are said to be as a part of dipole region defined by the two regions. After finding dipole the edges of that will be removed making the graph sparse when it is below the threshold value specified. Finally the dipoles get obtained from the two of the related buckets of data obtained by both the correlation values.

c. Community Based Approach
This is used for partitioning the network into smaller nodes and then to cluster appropriately for correlated regions using correlation thresholds [6]. There are many kinds of community based algorithms [11] as

1) Modularity Optimization Algorithm :
This has been introduced by Newman and Girvin. It measures connectivity of communities with same degree distribution. Based on this modularity there are few algorithms to be considered as

- FastGreedy [8], given by Newman is based on the greedy optimization using hierarchical approach [12]. It merges the communities until only one remains with all nodes.
- Louvain [4], given by Blondel is based on the greedy optimization using two – phase hierarchical (hybrid) approach. In first phase it identifies communities and in the second phase it builds new network to the identified communities of first phase. This continues till it obtains only one community.
- Eigenvector, given by Newman is based on the spectral algorithm (matrix representation of networks) and is termed as classic graph partitioning approach using divisive method [12].
- Spinglass, given by Reichardt and Bornholdt is based on simulated annealing optimization technique (used to solve unconstrained optimization).

2) Random Walk - Based Algorithm:

- Walktrap [5], given by Pons and Latapy is based on random walks using hierarchical agglomerative method which differs with the technique used by the FastGreedy. This uses node-to-node distance measure in identifying the communities. The main concept here is if two nodes are in same community then the probability of third node is located in the same community.
- MarkovCluster, given by Stijn van Dongen is based on random walks. This uses two techniques as Expansion, used in calculating random walks and Inflation, used to highlight the IntraCluster walks and to demote InterCluster walks.

3) Community Structure Algorithm or Information Based Algorithm:
This uses the tools derived from information theory in giving the partition of network for obtaining best community structure [1, 15]. This encompasses two kinds of algorithms as defined.

- Infomod, given by Rosvall and Bergstorm is based on simplified representation of a network. This follows simulated annealing optimization technique by using minimum description length principle.
- Infomap, given by Rosvall and Bergstorm is based on compact representation using simulated annealing. The community structure is given by two taxonomies one to distinguish communities in network and the other to distinguish nodes in a community.

III. COMPARISON

Among all the methods and algorithms discussed the Shared Nearest Neighbor (SNN) just finds the neighbors with similarities in between the two points and A1 results in finding the correlated locations which results in

contiguous regions due to autocorrelation. Whereas, the Community Based Approach is the best amongst because this partitions the network into smaller nodes and the dipole search here takes place form the smaller regions by making them robust by generating fewer dipoles and removing spurious connections through clustering into graphs [16].

IV. FUTURE WORK

The future work may involve in detecting dipoles for extreme weather events of various parameters. This also includes improvement in detection of the dipoles for the same polarities with respective to the different locations corresponding in the same period or year.

V. CONCLUSION

Thus this literature survey encompasses the various parameters and the kind of climate indices to be used. This includes the methodologies used in detecting the dipoles in an effective way. This also includes the comparison in between the different kinds of algorithms along with their functionalities that can be used in detection of the dipoles.

REFERENCES

1. On Accuracy of Community Structure Discovery Algorithms by Gunce Keziban Orman, Vincent Labatut, Hocine Cherifi of Galatasary University and University of Burgandy.
2. An approach towards the Shared Nearest Neighbor Clustering Algorithm by Hetal Bharat Bhavsar of Bharmsinh Desai University and Anjali Ganesh Jivani of M. S. University.
3. Steinbach, M., Tan, P., Kumar, V., Potter, C and Klooster, S. Data mining for the discovery of ocean climate indices. In Mining Scienti_c Datasets Work- shop, 2nd Annual SIAM International Conference on Data Mining, 2002.
4. Vincent D. Blondel, Jean- Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre, Fast Unfolding of Communities in large Networks.
5. Computing Communities in Large Networks Using Random Walks by Pascal Pons and Matthieu Latapy of LIAFA University.
6. Donges, J.F., Zou, Y., Marwan, N. Complex networks in climate dynamics. In European Physical Journal Special Topics,174 (1), pp, 157-179, 2006.
7. Ertoz, L., Steinbach, M., Kumar. V. A new shared nearest neighbor clustering algorithm and its applications in Workshop on clustering High Dimensional Data and its Appilications , SIAM Datamining, 2002.
8. Aaron Clauset, M. E. J. Newmann and Cristopher Moore, Finding Community Structure in Very Large Networks.
9. Data Guided Discovery of Dynamic Climate Dipoles by Jaya Kawale, Stefan Liess, Arjun Kumar, Michael Steinbach, Auroop Ganguly.
10. Comments on “Dipoles, Temperature Gradients, and Tropical Climate Anomalies” by Toshio Yamagata, A. Rao, Hameed N. Saji.
11. Community Detection in Complex Networks by keziban Orman, B.S.
12. Data Clustering Theory, Algorithms, and Applications Guojun Gan York University Toronto, Ontario, Canada Chaoqun Ma Hunan University Changsha, Hunan, People’s Republic of China Jianhong Wu York University Toronto, Ontario, Canada.
13. Topology and Predictability of El Nino and La Nina Networks by Anastaios A. Tsonics and Kyle Swanson f Department of mathematical Sciences, University of Wisconsisn – Milwaukee, USA.
14. Comments on “Dipoles, Temperature Gradients and Tropical Climate Anomalies” by Toshio Yamagata, Swadhin Behera and Hameed N. Sar.
15. Models for Community Detection Using Data Mining by Sanjeev Kumar, Prabhat Kumar and M. P. Singh.
16. Graph Clustering, a survey paper by Satu Elisa Schaeffer in Computer Science Review journal, Laboratory for theoretical Computer Science, Helsinki University of Technology TKK, Finland in the year 2007.
17. GEOPHYSICAL RESEARCH LETTERS, VOL. 32, L15710, doi: 10.1029/2005GL022709, 2005.
18. http://www.highbeam.com/doc/1P3-3653666771.html
19. http://www.tutorialspoint.com/data_mining/dm_cluster_analysis. html
20. http://eros.eas.gatech.edu/eas-8803/lectures/EOFs/eofprimer.pdf
21. http://www.wmo.int/pages/themes/climate/significant_natural_climate_fluctuations.php
22. https://climatedataguide.ucar.edu/climate-data/palmer-drought-severity-index-pdsi
23. http://www.ise.bgu.ac.il/faculty/liorr/hbchap15.pdf
24. http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-14-00507.1