The Classification of Invasion Taiwan Typhoon Track

John Chien-Han Tseng, Chien Yu, Jim-Lian Lee, Jau-Fu Wu

Central Weather Bureau, Taipei, Taiwan

Abstract

We aim at understanding invasion Taiwan typhoon tracks by clustering them into nine clusters using the tri-plots and k-means algorithm. Based on the fractal dimension, the tri-plots can classify two large and not equal sizes of the time series datasets. The tri-plots measure three function values which include two self-plots and one cross-plot. The self-plot affords the character of one individual dataset. After regression analysis, the self-plot give us the point in space instead of one sequence of one typhoon track. Collecting all the points of self-plot, the k-means method gives us the seven clusters of the invasion Taiwan typhoon track. The comparison between the traditional classification and new clustering method are given. There are some common views, but the new method gives us more information of retrieving the tracks. In general, we believe the tri-plots with k-means are useful tools for invasion Taiwan typhoon track cluster problem and should benefit more investigation in related community.

Key word: fractal dimension, tri-plots, self-plot, k-means algorithm

1. Introduction

The purpose of this study is aimed at the clustering of invasion Taiwan typhoon track. The tri-plots based on the property of fractal dimension show the potential ability to classify the annual typhoon tracks (Tseng et al., 2010). After the work of annual typhoon track classification, it is worthy to classify the invasion Taiwan typhoon track. Wang (1994) and Hseih et al. (1998) mentioned the different typhoon tracks causing different kinds rainfall patterns and damages, so they gave 9 types or kinds of invasion Taiwan typhoon track. The criterions of the clustering are geometrical shape of typhoon track and the landing area of the typhoon center. Somehow, the previous criterions or the features are too subjective. In this study, we try to compare the traditional subjective clustering and tri-plots clustering. On the other hand, tri-plots are proved successful for finding the similarity between two datasets, so the new clustering based on tri-plots can help to find similar invasion typhoon tracks. This will be helpful for searching and mining typhoon databank information.

The tri-plots（Traina et al., 2001） are one kind of data mining tools to finding the intrinsic patterns between two large and multidimensional datasets. We think the tri-plots are suitable for our research. The tri-plots are composed of two self-plots and one cross-plot. Through comparing two datasets after tri-plots, the two self-plots return the intrinsic characters of the two individual datasets and the cross-plot returns the relation character between two datasets. Taking all invasion Taiwan typhoon events data and comparing two datasets in turn, we can depict the distribution by self-plots. The distances from self-plot can be drawn on the plane and analyzed by k-means clustering algorithm. The clusters using the k-means algorithm can be regarded as one kind of objective classification tool. Once we choose the number of clusters, the k-means can try to cluster the data into several groups.

2. Data

The invasion Taiwan typhoon trajectories data are from typhoon data bank in Central Weather Bureau. The domain covers from (115E, 19N) to (130E, 30N). The data features contain the center of typhoon longitude, latitude, minimum pressure, and wind speed. In this study, we just use the longitude and latitude. The time resolution is about 1-3 hours. The time domain is from 1951 to 2009.

3. Fractal dimension and Tri-plots

Traina et al. (2001) proposed the tri-plots based on fractal dimension to calculate the distribution of distances between two datasets points. The tri-plots include three kinds of plot tools:

1. self-plot (for dataset A),

2. self-plot (for dataset B), and
3. cross-plot (for two datasets).

If A is self similar, the self-plot of A is close to a straight line and the slope is the intrinsic dimension of the set which is correlated with thefractal dimension of the set (Belussi and Faloutsos.,1995). In most situations, the cross-plot and self-plot with $\log(r)$ are not linear, and we can use regression method to determine the regression lines. We use the line slopes to discuss the information between the datasets A and B instead of the original data points (Tseng et al., 2010). In this study, we only use the self-plot to describe the individual typhoon track. According to Tseng et al. (2010), the self-plot consists of the slope and intercept of the regression line which stands for the individual typhoon track. So, for one typhoon track now can be represented as the (slope, intercept), one coordinate on the plane. The more similar two typhoon tracks are, the closer the two points of (slope, intercept) from two typhoon track self-plots. In brief, according to the property of tri-plots, we can conclude to three points: (1) if two datasets are identical, the two slopes of self-plots are same and one regression line overlaps the other; (2) if two datasets have same slopes but different intercepts, that means two datasets have same intrinsic or fractal dimension but different shapes, e.g., line and circle with same fractal dimension and different shapes; (3) the more slopes variations between two datasets, the more differences they have, and we can use this to differentiate these two things.

4. The Traditional Classification and the Tri-plots clustering

We plotted the distribution based on tri-plots and marked with traditional classification criterion in Figure 1. We can see somehow the self-plot grouping the same traditional classification label together, but after all, they are two different methods. In here, we just want to propose the concern that the traditional classification is too subjective and obscure in some situations. Or we can guess if there is a complicated typhoon trajectory, the traditional method can do the classification properly or not. Based on the distribution of Tri-plots, we use k-means algorithm to cluster the typhoon tracks and the result is shown is Figure 2. We give different clusters different colors, and the red squares mean the centers of the different clusters. In the beginning, the k-means algorithm calculates the distances between all data points and the initial guess centers of the clusters, then the algorithm will get the update cluster center. The algorithm will execute the previous procedure iteratively until the k-means algorithm finds the sum of the distances between all member points to the center is minimum and no other points can be added. In the final stage, the k-means help to decide the optimized cluster center. For the invasion typhoon cases, we get seven

clusters based on tri-plots distribution. There is one cluster at the origin point of Figure 2 with nine typhoon cases. Checking these nine cases detail, we find the life time of the typhoon cases is short, and the trajectories are like small circles which the tri-plots cannot decide fractal dimension, or the regression line, the slope and the intercept. Except the origin point cluster in Figure 2, the number of meaningful clusters is six. Somehow, these six clusters can represent the characters of the invasion typhoon tracks. In the left-upper corner, the typhoon tracks are complicated than other clusters. The blue plus and red dot cluster represent the traditional 4, 5, 7, 8 types typhoon tracks which the typhoon came from the south, southeast, southwest of Taiwan. The purple x and green star clusters mean the westward parabola shape typhoon coming from east of Taiwan. Their original traditional classification types are 1, 2, 3, 4 and 6.
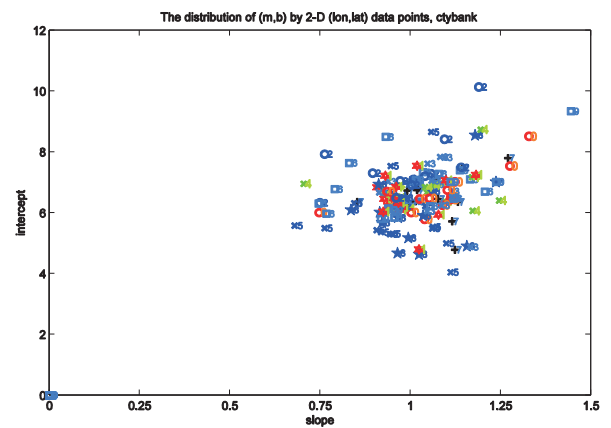

Figure 1: The distribution of the self-plot with traditional classification type number.
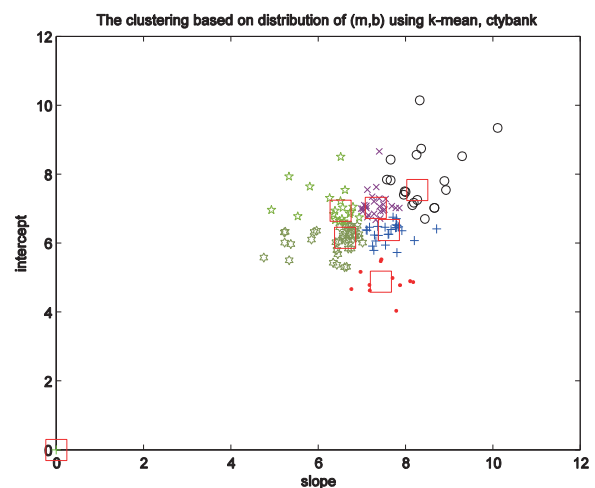

Figure 2: The clusters using k-means algorithm based on self-plot distribution

5. Conclusions and discussions

In this research, we proposed the quantitative and objective tools, tri-plots and k-means cluster, to distinguish the differences between the different yearly invasion Taiwan typhoon tracks events. The purpose is to have more clear definition of classification about invasion Taiwan typhoon. The objective method we proposed can be used for typhoon databank data mining. We can find several similar typhoon tracks efficiently. On the other hand, we do not criticize the traditional classification method, and from some point of view, the traditional method still helps us to realize the typhoon movements and analyze the patterns of rainfall. However, after the typhoon cases increasing dramatically, we should prepare the efficient and effective tool for information retrieval.

In future, we will keep going to research the detail information of the different clusters. The more comparisons will be done. The relations between the members and the clusters centers will be checked. One of the final goals is to establish one multi-class classifier with the rainfall estimating system.

5. References

Belussi, A. and C. Faloutsos, 1995: Estimating the selectivity of spatial queries using the 'correlation' fractal dimension. In *the 21-th conference of Very Large Data Bases*, 299-310.

Bishop, C. M, 2006: *Pattern recognition and machine learning*. Springer press.

Camargo, S, J., A. W. Robertson, S. J. Gaffney, and P. Smyth, 2007a: Cluster analysis of typhoon tracks. Part I: general properties. *Journal of Climate*, **20**, 3635-3653.

Camargo, S, J., A. W. Robertson, S. J. Gaffney, and P. Smyth, 2007b: Cluster analysis of typhoon tracks. Part II: large circulation and ENSO. *Journal of Climate*, **20**, 3654-3676.

Chan, J. C. L. and W. M. Gray, 1982: Tropical cyclone movement and surrounding flow relationships. *Monthly Weather Review*, **100**, 1354-1374 .

Emanuel, K, 2005: *Divine wind: the history and science of hurricanes*. Oxford university press.

Gaffney, S. and P. Smyth. Trajectory clustering with mixtures of regression models. In *the International Conference on Knowledge Discovery and Data Mining*, pages 63-72, 1999.

Harr, P. A. and R. L. Elsberry, 1991: Tropical cyclone track characteristics as a function of large-scale circulation anomalies. *Monthly Weather Review*, **119**, 1448-1468.

Hseih, S.-L., S.-T. Wang, M.-D. Cheng, and T.-C. Yen, 1998: *Tropical cyclone tracks over Taiwan and it vicinity*. Central Weather Bureau, 413pp.

Hsieh, W., 2009: *Machine learning methods in the environmental sciences*. Cambridge university press, Cambridge, 349pp.

Lee, J.-G., J. Han, and K.-Y. Whang, 2007: Trajectory clustering: A partition-and-group framework. *International Conference on Management of Data*, 593-640.

Lin, H, 2009: *Trajectory based on behavior analysis for verification and recognition*. National Taiwan University of Science and Technology, master thesis, Taipei.

Miller J., P. B. Weichman, and M. C. Cross, 1992: Statistical mechanics, Euler's equation, and Jupiter's Red Spot. *Physics Review*, **A45**, 2328–2359.

Risi, C, 2004: *Statistical synthesis of tropical cyclone tracks in a risk evaluation perspective*. Massachusetts Institute of Technology, internship report, Cambridge.

Strogatz, S. H, 2001: *Nonlinear dynamics and chaos: with applications to Physics, Biology, Chemistry, and engineering*. Westview press.

Tenebaum, J. B., V. d. Silva, and J. C. Langford, 2000: A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**, 2319-2323.

Tseng, J. C.-H., H.-K. Pao, and C. Faloutsos, 2010: The typhoon track classification using tri-plots and Markov chain. *International joint Conference on* knowledge discovery*, knowledge engineering and knowledge management*, Valencia, Spain, 364-369.

Traina, A., C. Traina, S. Papadimitriou, and C. Faloutsos, 2001: Tri-Plots: Scalable tools for multi dimensional data. In *the International Conference on Knowledge Discovery and Data Mining*, 184-193.

Vlachos, M. D. Gunopulos, and G. Kollios, 2002: Discovering similar multidimensional trajectories. In *the International Conference on Data Engineering*, 673-684.

Wang, S.-T., 1994: *Typhoon over Taiwan*. Sinotech Foundation for Research and Development of Engineering Sciences and Technologies. 139pp. (in Chinese)

Webster, P. J., G. J. Holland, J. A. Curry, and H.-R. Chang, 2005: Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science*, **309**, 1844-1846.