Scholars Journal of Engineering and Technology (SJET) **ISSN 2321-435X (Online)**
Sch. J. Eng. Tech., 2015; 3(1A):33-36 **ISSN 2347-9523 (Print)**

(An International Publisher for Academic and Scientific Resources)
www.saspublisher.com

Research Article

Applying Clustering Technique into the Earthquake Analysis at Taiwan

I-ChingYang[1], Chun-Nan Lin[2], Wen-Kuei Hung[3], Jann-Yenq Liu[4], Kun-Lin Hsieh[5*]
[1,4,5]National Taitung University, Taitung County, Taiwan
[2]Fortune Institute of Technology, Kaohsiung County, Taiwan
[3]Taitung County Government, Taiwan

***Corresponding author**
Kun-Lin Hsieh
Email: klhsieh2644@gmail.com

Abstract: As we known, earthquake (EQs) prediction had been mentioned as an important issue to many countries. In this paper, we proposed an approach to find out the events of large earthquakes from the time series data of GPS total electron content (GPS TEC) by using clustering analysis. To search for possible precursors before earthquake occurring in a large area, the GPS TEC of the global ionosphere map (GIM), which is routinely published in a 2hr time interval for monitoring global ionospheric weather, reported at the Center for Orbit Determination in Europe (CODE) is used in this study. And, an illustrative case at Taiwan is also used to denote the feasiblilty of the proposed approach.
Keywords: clustering analysis, earthquake, ionospheric anomalies.

INTRODUCTION

Large earthquakes are often accompanied by signals of different nature, like as electric, electromagnetic, or luminous, and have been observed by many scientists [1-2]. Since the 1980s, seismo-ionospheric phenomena have received considerable discussions[3-4] and the ionospheric anomalies associated with earthquakes (EQs) have been investigated[5-6]. Recently, scientists analyzing data recorded from numerous ground-based receivers of the global positioning system (GPS) have observed ionospheric disturbances of the total electron content (TEC) triggered by seismic surface waves [7-9] and by tsunami waves[10]. In particular, Liu et al.[11] have reported that the ionospheric TEC derived from ground-based receivers of the GPS can be used to observe pre-earthquake ionospheric anomalies (PEIAs), and have showed that the reduction of TEC caused prior to occurrence of EQs. Furthermore, some studies also observe similar anomalous reduction features of the ionospheric GPS TEC appearing in the afternoon and evening periods within day 1-6 before M≥6.0 earthquakes12-14].

In this paper, in order to investigate the possibility of the EQs prediction, we would like to find out the events of large EQs from a lot data by using the method of clustering analysis on the time series of GPS TEC.

To search for possible precursors before earthquake occurring in a large area, the GPS TEC of the global ionosphere map (GIM), which is routinely published in a 2hr time interval for monitoring global ionospheric weather, reported at the Center for Orbit Determination in Europe (CODE) [15] is used in this study. The spatial resolution of the GIM on the ±87.5°N latitude and ±180°E longitude are 2.5° and 5°, therefore, each map consists of 5040 (=70×72) grid points. Here, we consider 7 time series of 11d periods within one M > 5.5 earthquake in Taiwan, including 1-10 day before and the day of the earthquake(Table 1), and 7 without any M > 4.5 earthquakes in Taiwan(Table 2). All time series of these 14 periods are located at a given grid point (120°E, 22.5°N), which is the one of those two nearest grid points to Taiwan.

Table 1: 7 M > 5.5 earthquakes

year	month	day	hour	Magnitude scales (M)	Longitude (°E)	Latitude (°N)	Label
2011	3	20	16	5.8	121.38	22.44	A110320
2010	11	21	20	6.1	121.69	23.85	A101121
2010	7	25	11	5.7	120.69	22.84	A100725
2010	3	4	8	6.4	120.71	22.97	A100304
2009	11	5	17	6.2	120.72	23.79	A091105
2007	8	9	8	5.7	121.08	22.65	A070809
2006	4	1	18	6.2	121.08	22.88	A060401

Table 2: 7 11d periods without any M > 4.5 earthquakes

year	beginning		terminal		Label
	month	day	month	day	
2009	1	9	1	19	B090119
2009	4	20	4	30	B090430
2009	9	14	9	24	B090924
2008	1	31	2	10	B080210
2008	3	10	3	20	B080320
2007	6	4	6	14	B070614
2007	9	24	10	3	B071003

MATERIAL AND METHODS

The purpose of clustering algorithm is to group several sub-group data sets in a large data set, and finds out the similarity of data points as possible in each sub-group after analyzed the characteristics of these data. Based on the result, we can detect the useful strategy, and understand, enhance, or revise the strategy planning and implementation. Clustering techniques can be classified two types: hard clustering and fuzzy clustering (also referred to as soft clustering); the primary difference is the relationship between data point and sub-group [16]. In hard clustering, data is divided into one sub-group, and keeps to belong to this sub-group. However, in fuzzy clustering, data can belong to more than one sub-group, each data has the membership degree for each sub-group, based on the membership degree, we can observe the relationship among these data [17-18]. We use hierarchical clustering analysis and fuzzy c-means to analyze the data set in this study. About the details of two clustering methods include the following:

Hierarchical clustering analysis

Hierarchical clustering analysis (HCA) is a method of clustering analysis which can build a hierarchy relationship for data clusters in data mining. The two general strategies of HCA are used: agglomerative and divisive; the former is a bottom up approach, and depends on the merged concept; the latter is a top down approach, and conforms the concept of split. No matter which strategy, they are determined in the greedy method, and the result can be presented in hierarchy diagram.

The metric (mathematics) is used to measure the similarity among a large data set in hierarchical clustering. The commonly metric is Euclidean distance:

$$Ed = \sqrt{\sum_{i=1}^{n} (x_{i1} - x_{i2})}^{2} \qquad (1)$$

where x_{i1} = the ith dimension coordinates of 1st data point.

x_{i2} = the ith dimension coordinates of 2nd data point.

In addition, we adopt the Ward's method of agglomerative method in HCA. Ward's method is based on the minimum variance to form the within-cluster sum of squares is minimized which presents the high similarity of data in a sub-group.

Fuzzy c-means

Fuzzy c-means (FCM) is one of widely used fuzzy cluster algorithm which is different from hard clustering that employs hard partitioning [17-20]. Fuzzy partitioning is employed by FCM such that a data point can belong to all groups with different membership degrees between 0 and 1 [16]. The algorithm is based on the following steps [16, 20]:

Step1: select *c* data points as the initial representatives.
Step2: the membership matrix is randomly initialized.
Step3: calculate central data point.
Step4: compute similarity between central data point and other data points. If the improvement over previous iterations is below a threshold, the process can be stopped.
Step5: compute a new membership matrix and got to Step2.

RESULTS AND DISCUSSION

Figure 1 shows the result of HCA. We can know that the 14 data are grouped to 5 sub-clusters. Especially, the 1st and 4th sub-clusters are only label B (B080210, B090924, B090119) data and label A (A060401, A100304, A101121) data. Although the other sub-clusters cannot be grouped clearly, we still feel the phenomenon of clustering in the ionospheric TEC data. In other words, we know the method of HCA or the concept of clustering can help us to observe the phenomenon of earthquake from ionospheric TEC data, actually, these data have the clustering phenomenon.

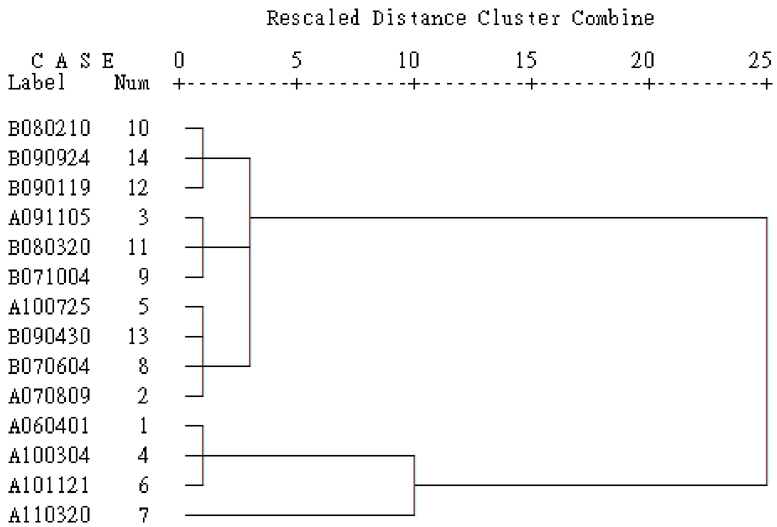

Fig-1: The result of HCA

We also use FCM to analyze the ionospheric TEC data. The result of FCM is shown in Table 3. After comparing the results, the same phenomena are observed for us from HCA and FCM analyses.

Table-3: The result of FCM

sub-cluster	data
1st	A060401(1), A100304(4), A101121(6)
2nd	B080210(10), B090119(12), B090924(14)
3rd	A110320(7)
4th	A091105(3), B071004(9), B080320(11)
5th	A070809(2), A100725(5), B070604(8), B090430(13)

Based on the result of fuzzy membership function degree (see Table 4), we know that the data A060401, A100304, and A101121 which have high fuzzy membership degree (>0.5) belong to 1st sub-cluster. In the 2nd sub-cluster, B080210 and B090119 have high fuzzy membership degree (>0.5) which belong to 2nd sub-cluster. B090924 is also belonged to 2nd sub-cluster, although its fuzzy membership degree is not more than 0.5, 0.495737 closes to 0.5. A110320 is the special data, it has the highest fuzzy membership degree (>0.99) in the 3rd sub-cluster, and also the 3rd sub-cluster has only it. In 4th and 5th sub-clusters, 7 data are belonged to them. They are included label A and B data which are almost 50%. Their fuzzy membership degrees also present the phenomenon, such as A070809 (0.3832664).

Table-4: The membership function degree of FCM

cluster \ data	A060401	A070809	A091105	A100304	A100725	A101121	A110320
1st	0.727356	0.062182	0.046444	0.641093	0.061321	0.722466	3.25E-05
2nd	0.050561	0.298702	0.10121	0.063505	0.154553	0.04852	1.18E-05
3rd	0.029647	0.012769	0.007042	0.024669	0.010907	0.045832	0.999923
4th	0.105846	0.243082	0.716482	0.160034	0.152412	0.101686	1.73E-05
5th	0.08659	0.383264	0.128821	0.110699	0.620807	0.081496	1.56E-05

cluster \ data	B070604	B071004	B080210	B080320	B090119	B090430	B090924
1st	0.127365	0.13451	0.0363	0.179916	0.039072	0.054202	0.044826
2nd	0.125054	0.14234	0.685211	0.139634	0.690981	0.136853	0.495737
3rd	0.015765	0.019712	0.007701	0.020238	0.009584	0.009201	0.008642
4th	0.255851	0.484427	0.121652	0.465101	0.125334	0.17394	0.181314
5th	0.475965	0.21901	0.149136	0.195111	0.135029	0.625804	0.269481

To compare HCA and FCM, we can observe the result is consistent when the situation of cluster number is 5. The 1st sub-cluster is included data A060401, A100304, and A101121. The 2nd sub-cluster is included data B080210, B090119, and B090924. Data A110320 is the 3rd sub-cluster. The data such as, A091105, B071004, and B080320 is clustered into the 4th sub-cluster. The 5th sub-cluster is included data A070809, A100725, B070604, and B090430. We know the ionospheric TEC data have the phenomenon of clustering based on these results, and between the results of hard and fuzzy clustering algorithms are consistent. Although the 3rd sub-cluster only included one data, the 1st, 2nd, and 3rd sub-cluster has the pure situation.

CONCLUSIONS

Based on a series of surveys of Liu et al., it is confirmed that large EQs will occur TEC anomalies. In this study, we attempted to distinguish from the time series of GPS TEC within one M > 5.5 earthquake and without any M > 4.5 one in Taiwan. Although our results reveal that the sub-cluster of GPS TEC in which the time series either label A or label B are belonged only is presence, other sub-cluster of GPS TEC contains the time series label A and label B both is existence also. It can’t put to sort out the time series of GPS TEC with large EQs though by using HCA, because many factors which can cause TEC anomaly don’t be removed. So, we introduce another method of data mining, FCM, and consider the possibility of regulating the members of sub-clusters. But, the clustering result of FCM is similar to HCA. In FCM, we are conscious that the fuzzy membership function degree in FCM maybe could be the index of probability about identifying EQs occurred. Certainly, it need more studies about the relation between the fuzzy membership function degree in FCM and the probability about identifying EQs occurred. Predict the earthquake should be based on the concept of probability, not directly say yes or no. So, the perspective of probability should be considered in the future work, especially, we can apply the fuzzy membership function degree.

REFERENCES

1. Bolt BA; Earthquake (Freeman: New York), 1999.
2. Freund F; Time-resolved study of charge generation and propagation in igneous rocks, J. Geophys. Res.,2000; 105:11001-11019.
3. Hayakawa M, Molchanov OA; Seismo Electromagnetics: Lithosphere-Atmosphere-Ionosphere Coupling, Terra Sci, 2002; 477.
4. Pulinets S, Boyarchuk K; Ionospheric Precursors of Earthquakes (Springer: New York), 2004.
5. Liu JY, Chen YI, Pulinets SA, Tsai YB, Chuo YJ; Seismo-ionospheric signatures prior to M _ 6.0 Taiwan earthquakes, Geophys. Res. Lett.,2000; 27:3113-3116.
6. Molchanov OA, Hayakawa M; Subionospheric VLF signal perturbations possibly related to earthquakes. J. Geophys. Res, 1998; 103-8 :17489-17504.
7. Calais E, Minster JB; GPS detection of ionospheric perturbations following the January 1994, Northridge earthquake, Geophys. Res. Lett., 1995;22:1045–1048.
8. Afraimovich EL, Perevalova NP, Plotnikov AV, Uralov AM; The shock-acoustic waves generated by the earthquakes, Annales Geophysicae, 2001; 19:395-409.
9. Ducic V, Artru I, Lognonne P; Ionospheric remote sensing of the Denali Earthquake Rayleigh surface wave, Geophys. Res. Lett., 2003; 30-18:.
10. Artru J, Ducic V, Kanamori H, Lognonné P, Murakami M; Ionospheric detection of gravity waves induced by tsunamis, Geophys. J. Int., 2005; 160: 840–848.
11. Liu JY, Chen YI, Chuo YJ, Tsai HF; Variations of ionospheric total electron content during the Chi-Chi earthquake, Geophys. Res. Lett., 2001; 28:1381-1386.
12. Liu J Y, Chuo YJ, Shan SJ, Tsai YB, Chen SI, Pulinets SA, Yu SB; Pre-earthquake ionospheric anomalies registered by continuous GPS TEC measurements, Ann. Geophys., 2011; 22:1585-1593.
13. Liu JY, Chen YI, Chen CH, Liu CY, Chen CY, Nishinashi M, Li J Z, XIA Y Q, Oyama K I, Hattori K, Lin CH; Seismo-ionospheric GPS TEC anomalies observed before the 12 May 2008 Mw7.9 Wenchuan Earthquake, Journal of Geophysical Research, 2009;114:A04320
14. Zhao B, Yu T, Wang M, Wan W, Lei J, Liu L, Ning B; Is an unusual large enhancement of ionospheric electron density linked with the 2008 great Wenchuan earthquake?, J. Geophys. Res., 2008; 113:A11304
15. Schaer S; Mapping and Predicting the Earth’s Ionosphere Using the Global Positioning System, 59(1999), Geod. Geophys. Arb. Schweiz.Inst. für Geod. und Photogramm, Zurich, Switzerland.
16. Lin CN, Tsai CF, Roan J; Personal photo browsing and retrieval by clustering techniques: Effectiveness and efficiency evaluation, Online Infor. Rev., 2008; 32(6):759-772.
17. Bezdek JC; Pattern recognition with fuzzy objective function algorithms (Plenum Press, New York), 1981.
18. Zimmermann JJ; Fuzzy Set Theory and Its Applications (Kluwer Academi, Boston), 1991.
19. Dunn JC; A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, Journal of Cyber., 1973;3(3) :32-57.
20. Jang JS, Sun CT, Mizutani E; Neuro-fuzzy and soft computing(Prentice Hall, New Jersey) 1997.