Online Clustering of Radar Pulses Using Data Mining Algorithm

Zeinab Liaghat[1, a], Koorush Ziarati [2, b] and Neda Froutan[3, c]
[1]Department of Computer Science and Engineering and Information Technology Shiraz University, Shiraz, Iran
[2]Department of Computer Science and Engineering and Information Technology Shiraz University, Shiraz, Iran
[3]Department of Computer Science and Engineering and Information Technology Shiraz University, Shiraz, Iran
[a]zliaghat2010@gmail.com, [b]ziarati@cse.shirazu.ac.ir, [c]nfroutan.neda@gmail.com

Keywords: Data mining, data stream, clustering, electronic warfare

Abstract. Growing military systems and arnaments, together with the applying electronic equipments and making some other activities to deal with such systems leads to the development of a specific field in defensive sciences which is called electronic warfare. The electronic warfare includes discovery, stopping and neutralizing the effects of enemies' electronic telecommunication devices and electromagnetic systems.
One of the most important systems in electronic warfare is Electronic Support Measure (ESM). ESM system can detect and analyze active emitters in environment. This system receives emitted signals from the active radars. The task of ESM's information processing unit is to separate pulses of each source from interleaved pulses and to specify each radar parameters distinctly.
In this paper separation of interleaved radar pulses is investigated and a new method for pulse multi-parameter de-interleaving is presented.

I. INTRODUCTION

Recently, data stream model has attracted much attention for its applicability to various types of data such as sequence of clicks, web pages, sensor data, satellite data and radar data. Data stream is data that is produced in high volume so that there is not the possibility of storing all data in one place and data processing in limited time is needed. These data are produced in many applications. Data stream model requires classifications, clustering, discovering iterative algorithms and summarizing in the important processes and it faces particular challenges.
There are many active radars in an environment; the ESM receives the signals emitted from individual radars. It measures parameters of each pulse including pulse arrival time, pulse radio frequency, pulse arrival angle, pulse width and pulse amplitude and separates radars interleaved pulses with the help of these parameters. Finally, each emitter source can be identified with its parameters. The task of ESM's signal processing unit is to determine pulses of each radar from interleaved pulses stream and to specify radar parameters separately. Clustering methods are used to separate pulses. Clustering of monopulse parameters is one of the important subjects in the pulses separation. Usually in conventional multi-dimensional de-interleaving techniques, two or three pulse descriptor words are used in the separation process.
In this paper, a new method is presented for clustering the interleaved radar pulses. This algorithm de-interleaves the signals using mono pulse parameters such as direction-of-arrival (DOA), radio frequency (RF) and pulse width (PW).
The reminder of this paper is organized as follows. Section II reviews the prior works related to our research. In Section III, the basic concepts of the K-means clustering algorithm are described. Description of pulse parameters and ESM system are given in Section IV. In Section V we introduce our algorithm for clustering radar pulse streams. Section VI discusses about the experimental results. Finally, Section VII concludes this work.

II. RELATED WORK

De-interleaving of radar pulses is an essential part of an electronic warfare system and has attracted attention of many researchers recently.
First time, Mardia[10] and Milojevic and Popovic[11] proposed their method for pulse clustering. Their approach used histogramming of TOA parameter. Among the known algorithms for pulse clustering, we can name CDIF, SDIF and TOA folding.
As for popularity of neural networks, some researchers used it for the purpose of de-interleaving radar pulse. For example Granger et al. compared four neural networks on radar pulses named fuzzy adaptive resonance theory (FA), fuzzy min-max clustering (FMMC), self-organizing feature mapping (SOFM), and integrated adaptive fuzzy clustering (IAFC). Their comparison was based on quality of clustering, convergence time and time complexity. They found self-organizing feature and fuzzy adaptive resonance theory as the best candidates. SOFM obtained the best quality while FA had the faster run times.
Recently, Ataa and Abdullah raised a received pulse separation system with combining fuzzy adaptive resonance theory network clustering of RF and AOA features[12].
Liu et al. also proposed two algorithms which were applied on data with high dimension.
It is noteworthy that our proposed approach differs from these algorithms in the fact that we use PDW information while they do not use it. Our algorithm is based on k-means clustering algorithm.

III. CLUSTERING AND K-MEANS ALGORITHM

Nowadays, clustering is considered as one of the most common tasks in data mining. Clustering methods are used to group data based on their similarities. Analysis is done only according to similarity and without any knowledge. In this paper, our objective is separation of radar signals and clustering active radars in environment. One of the well-known clustering algorithms in data mining is K-means.
Our proposed method is based on K-means clustering algorithm. In this paper, we improved the K-means algorithm so that it is to be effective to cluster radar data stream.
K-means is one of the simplest non-supervised learning algorithms that solves well-known clustering problems. Despite the simplicity of this method, it is a base for many other clustering algorithms such as fuzzy clustering. K-means is a simple algorithm that has been adapted to many problem domains. This method is exclusive and flat. The most common algorithm for k-means uses an iterative refinement technique.
Generally, the algorithm follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) given in advanced. The main idea is to define k centroids, one for each cluster. Because different location of these points causes different result, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, we need to re-calculate k new centroids as centers of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. This procedure is done iteratively until no more changes are done. In other words centroids do not move any more.
Although it can be proved that the algorithm will always terminate, there is no guarantee that it will converge to the global optimum, and the result may depend on the initial clusters. In other words each time the algorithm runs with different starting points, its result may not to be same as previous result. Generally, this simple version of algorithm has some weaknesses listed below.

- Number of clusters must be specified before, but in many applications, it is not clear.
- The results produced depend on the initial values for the means, and it frequently happens that suboptimal partitions are found. The standard solution is to try a number of different starting points.
- There is no general theoretical solution to calculate initial values for the means.
- It can happen that the set of samples closest to one cluster (mi) is empty, so that mi cannot be updated.
- K-means does not change the concept of data effectively.

- Data should be fully available at first step.
- The same result are not produced in each iteration of running algorithm
- The concept of data changes over time

IV. ELECTRONIC SUPPORT MEASURE SYSTEM

Electronic Support Measure (ESM) system is one of the important systems in electronic warfare (EW). This system receives combined signals from the active radars available in the environment. The parameters of signals are extracted and features of radars are determined separately. These parameters will discuss in following subsections.

A. parameters description

For every pulse the measured parameters are packed into a structure called a pulse descriptor word (PDW). PDWs often include information of pulse amplitude (PA), radio frequency (RF), pulse width (PW), time of arrival (TOA), direction of arrival (DOA) and pulse repetition interval (PRI). Fig. 1 illustrates some of these parameters. Clustering is done with these radar parameters. The proposed algorithm extracts features of each radar separately.

1) Radio Frequency (RF)

Transmitted pulses are sinusoidal signals which are radio frequencies. In other words, frequency is the sinusoidal signal of a particular pulse. It is a parameter that determines how often the sinusoidal signal goes through a cycle. The range of RF is summarized in Table 1.

2) Pulse Width (PW)

The duration of time for which a pulse is being transmitted is called pulse width.

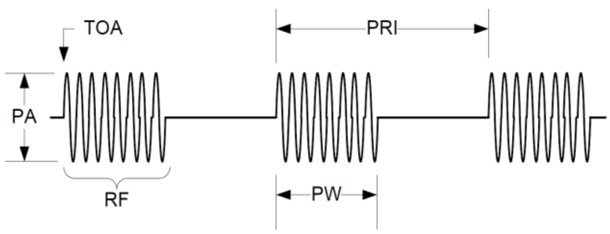

Figure 1: Pulse parameters

Table 1: Range of pulse parameters

Parameter	Range
Frequency	DC - 32 GHz
Pulse Width	0.05- 204 μs
DOA	360 degrees

3) Direction Of Arrival (DOA)

The arrival direction of signal to ESM system's receiver is the other one radar pulse parameters. This parameter just depends on physical location of radar. The pulse arrival direction can be used to separate radars pulses and to specify position of objects with accuracy about 10 degrees.

4) Pulse Amplitude (PA)

Pulse amplitude represents the amount received power by the antenna of ESM system. The received power depends on the sending power, the target distance, frequency and receiver and transmitter antennas.

5) Pulse Repetition Intervals (PRI)

This parameter has an important role in the analysis of radar data and represents the time period among two similar pulses of one radar.

6) Time Of Arrival (TOA)

Time of arrival of radar is the most accurate parameter. Unlike other pulse parameters, which alone can be used in the analysis of radar data, TOA values must be processed as a group. Several values of the TOA can be used to obtain the pulse repetition interval of radar that is one of the radar‘s important parameters.
The Introduced method clusters received pulses according to the best extracted features same as K-means. The similar pulses are grouped in same cluster.

V. THE PROPOSED ALGORITHM

In this section, we discuss the details of our method in order to reduce the number of features and accelerate the algorithm, we ran about 18 feature selection algorithms from Weka which is data mining software in Java. For example bagging algorithm, linear and additive regression, rule decision table, m5 rules and etc. The results showed that the correlation between some of parameters is very high. these are DOA, RF and PW parameters. So our algorithm just uses these three information of PDW for de-interleaving of radar signals.
For feature selection, we use Weka and run several of its algorithms on our data. According to their result, DOA, Freq and PW parameters are best features among six existed parameters.
In the radar data stream, we cannot save all data because of large amount of data. Furthermore, since we do not access all data specifically in this application, we cannot run the K-means algorithm on the data stream.
So based on the time of arrival of pulse, we have to use and process them simultaneously and omit them afterwards. One of the receiving Data stream methods is dividing data into multiple windows. So we will separate the sequentially received data, window by window. During the operation on windows, the entered new data will be saved in the next windows. After the operation on a previous window, it will be done on the next one as well. This cycle of operation will continue to the end.
As it was mentioned before, the objective in analyzing radar data is detection of active radars in the environment and presenting the last characteristics of each active radar.
The proposed algorithm to cluster the radar data stream is described step by step as below:

A. Separate data into windows:

In this algorithm data is given to the algorithm by its order of arrival times, and the cluster related to each data will be specified at that time. Because of particular application of this algorithm, the number of the clusters are not predefined and this number accompanied by the amount of sensitivity on the degree parameter, frequency, pulse width and the window size, will be given to the algorithm as its inputs.
When turning the system on, a sequence of data stream from the active radars is received, and we have to be able to cluster them.
Some of these radars may turn off during the time, so there is the necessity to eliminate the clustering produced for this radar. Moreover, it is possible that a radar is turned on at this time and it is the data which enters the system, so there is still the necessity to be able to produce a new cluster along the program running.
It should be considered that in some radars, their parameters are changing continuously like the airplane's radar which has motion.
The processes of production, elimination and updating clusters are described in following sub sections.

B. new cluster creation

In the start, algorithm receives the first window. The first received sample from that window, is considered as the first cluster, and the reminder of received data are compared with the lately produced clusters based on chronological order. If there is any similarity between those three mentioned parameters, they will be included in that cluster; otherwise a new cluster will be produced with parameters of incoming data.
In data stream, it should be considered that the old data is less valuable over time and new data becomes more valuable. Nevertheless, in radar data, the old data give us some information and helps us in recognizing clusters. By the way, due to the change in data over time, not sufficient

memory, and all defects of data stream; we just save the last three data of each cluster, but this number can be changed in the program input.

C. Cluster updating

In order to reduce the effect of old data, only the three last data are saved. By entering new data into the cluster, the first saved data in the cluster is deleted just like the FIFO linked list. Afterwards the centroid of cluster will be updated.

D. eliminating the cluster

In order to eliminate old clusters for which that data belonging to those clusters has not seen recently, we will set a life time for cluster. After a determined period of time, cluster life time is decreased and the cluster will be deleted when its life time reaches zero.

E. Framework of the algorithm

Data are received from antenna in the form of a sequential stream and are stored until the window is thoroughly loaded. Data stored in the window are called one by one. According to the delivered parameters, a comparison is done between the existing clusters and the receiving data. If the data is consistent with the cluster, that data will be assigned to that cluster and the cluster information will be updated. Otherwise, a new cluster will be created. Finally, life time of each cluster is set. Then the next data will be selected on the process will be done on it. This process is continued until the end of each window and then the next window will be received.

VI. THE EXPERIMENTAL RESULTS

We ran the introduced algorithm in several stages on two different datasets. In each stage, percentage of noise and data complexity is changed. We partition the executed stages into two parts based on the input type: static data and dynamic data. Static data part means that all received radar parameters are assumed as constant while in dynamic data part, data is varied between 10% and 15% depending on the scale determined by the radar. We used simulator software to produce the radar data stream. The obtained results by running the proposed algorithm on these data are presented in the following subsections.

A. The static data sets

In this section, we ran the algorithm three times with different number of clusters including 3, 5 and 10 clusters. The characteristics of experimental data are shown in Table 2. The results in each stage are described as below:

1) First stage: *the static data with three clusters*

In this experiment, we executed algorithm on first three data of two datasets in static part with the radars' features shown in Table 2. Our algorithm predicted the number of clusters and source of each instance correctly. In other words, algorithm clusters data exactly

2) Second stage: *the static data with five clusters*

By entering the data statically with the first five clusters of two datasets, the proposed algorithm has predicted the number of clusters and source of each instance correctly. This shows that the data are clustered correctly.

3) Third stage: *the static data with ten clusters*

The proposed algorithm calculates the number of clusters and source of each instance correctly when data set 1 is used as input. But by entering data set 2 as input, algorithm predicted only 91% of clusters correctly due to complexity of data.

Because of overlapping two incoming signals from two different radars, this case is unavoidable in both static and dynamic data.

Table 2: SECOND STATIC DATA SET

No	DOA	DOA sd	PRF	PRI Type	J/S	PPP	PW(ns)	PW sd	Freq
1	30	0	1000	Stable	0	0	1000	0	9500
2	160	0	2000	Stable	0	0	700	0	8500
3	100	0	3000	Stable	0	0	880	0	5800
4	270	0	4040	Stable	0	0	930	0	9000
5	330	0	1500	Stable	0	0	390	0	7500
6	180	0	2400	Stable	0	0	500	0	9500
7	210	0	5000	Stable	0	0	1100	0	5000
8	75	0	2600	Stable	0	0	820	0	8000
9	140	0	1800	Stable	0	0	660	0	9500
10	245	0	3500	Stable	0	0	1900	0	9500

B. The dynamic data sets

In this section, the program is executed using dynamic data which their characteristics are the same Table 2. Observed result of each step is presented in Fig. 2.

1) **First step:** *Dynamic data with three clusters*

Applying the data stream generated by the first three radars of the datasets, our algorithm predicts 100% of resources correctly.

2) **Second step**: *Dynamic data with five clusters*

In this step, the data streams generated by the first five radars of the datasets were applied. Again our algorithm predicts 100% of resources and number of clusters correctly.

3) **Third step:** *Dynamic data with ten clusters*

When the whole data stream generated by the ten radars of the first dataset was applied, our proposed algorithm predicts 92% of resources correctly; and when the whole data stream generated by the ten radars of the second data set was applied, our algorithm predicts 86% of resources correctly. This error is due to outliers.

VII. CONCLUSION

Our proposed algorithm is able to classify static data streams appropriately. For dynamic data streams, an accuracy parameter should be considered to determine sensitivity of the program. Initializing this parameter properly, the program can identify the number of clusters better. However, because of the dynamic nature of data stream and the possibility of gathering similar data from different radars, it is possible that some of the data to be located in wrong cluster temporarily. This problem would be solved as soon as arrival of the next data. Considering life time for clusters eliminates the effect of outliers.

As mentioned earlier, there exist some kinds of noise which cannot be determined and therefore cannot be avoided.

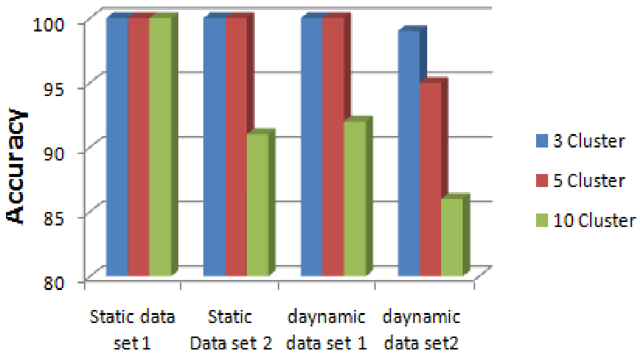

Figure 2: Results obtained from static data and dynamic sets

VIII. FUTURE WORKS

In this work, three different attributes are investigated during clustering. According to existing references, other attributes such as PRI and TOA can affect clustering considerably. In order to achieve appropriate results, the input window size should be properly chosen. For this reason, a large amount of real data is needed.

REFERENCES

[1] S. Guha , A. Meyerson, N. Mishra, and R. Motwani, “Clustering data streams: theory and practice,” Knowledge and data engineering, IEEE transactions on, 2003.

[2] Z. Liaghat, A.H. Rasekh, k. Ziarati,”Radar Data Stream Clustering” ,the Fifth 2011 International Conference on Data Mining, Dec 2011.

[3] J. L. Lee, J.P.Y. Lingjie, L. Zhi-Quan, and L. Wong, “Online Clustering Algorithms for Radar Emitter Classification,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, August 2005.

[4] Q. Guo, X. Zhang, and Z. Li, “SVC and K-Means and Type-Entropy Based De-Interleaving/Recognition System of Radar Pulses,” Information Acquisition, IEEE International Conference, February 2007.

[5] Q. Guo, X. Zhang, and Zheng Li, “Cascade Coupling and Support Vector Clustering Based Novel Sorting Method of Radar Pulses,” Circuits and Systems, 49th IEEE International Midwest Symposium, August 2006.

[6] J.H. Chang, and W.S. Lee, “estWin: Online data stream mining of recent frequent itemsets by sliding window method,” Journal of Information Science, Vol. 31(2), April 2005.

[7] R. Agrawal, and R. Srikant, “Fast algorithms for mining association rules,” in Proceedings of the 20th International Conference on Very Large Databases, 2004.

[8] Q. Guo, and X. Zhang, “A Novel Joint De Interleaving/Recognition System of Radar Pulse Sequence,” Proceedings of the First International Conference on Innovative Computing, Information and Control, Vol. 3, 2006.

[9] J.H. Chang, and W.S. Lee, “Online data stream mining of recent frequent itemsets by sliding window method,” Journal of Information Science, Vol. 31, pp.420-432, October 2005.

[10] H. Mardia, ”New techniques for the deinterleaving of repetitive sequences,” in Radar and Signal Processing, IEEE Proceedings, Vol. 136, August 1989, pp.149-154.

[11] D. Milojevic and B. Popovic, “Improved algorithm for the deinterleaving of radar pulses,” in Radar and Signal Processing, IEEE Proceedings, Vol. 139, February 1992, pp. 98-104.

[12] E. Granger, Y. Savaria, P. Lavoie, and M.A. Cantin, “A comparison of self organizing neural networks for fast clustering of radar pulses," Signal Processing, Vol. 64, pp. 249-269, 1998.

[13] A. Ataa and S. Abdullah, “Deinterleaving of radar signals and prf identicational gorithms," Vol. 1, October 2007, pp. 340-347.

[14] J. Liu, J. P. Lee, L. Li, Z.Q. Luo, and K. M. Wong, “Online clustering algorithms for radar emitter classification,” IEEE transactions on pattern analysis and machine intelligence, Vol. 27, August 2005, pp.1185-1196.

[15] J.H. Chang, and W.S. Lee, “Finding recent frequent itemsets adaptively over online data streams,” Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining, 2003.

[16] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, “Dynamic itemset counting and implication rules fot market basket data,” Proceedings of the ACM SIGMOD International Conference on Management of Data, 2004