

Research Article

A Robust Scalable Model Using Hybrid Approach for the Detection of the Projected Outliers

¹H. Sadawarti, ²G.S. Kalra and ³Kamal Malik
¹RIMTIET, Punjab Technical University,
²Lovely Professional University, Punjab,
³MMICT and BM, MMU, Mullana, Haryana, India

Abstract: The abnormal and anomalous observations even in the advanced technological era proves to be the biggest jolt to the concerned industry. To reduce and eliminate the outliers from the massive data streams, it is important to accurately highlight them from the higher dimensional data which is itself very challenging. In this study, a Scalable outlier detection model is proposed which is robust enough to resist and detect the projected outliers that are lying at some lower dimensional subspaces. This model exploits the problem of curse of dimensionality which is very frequent in large data streams and massive datasets. Rapid distance and density based approaches are used and then the probability density is measured by Gaussian Mixture Model. Baye's Probability is applied to the final observations so as confirm them as the projected outliers.

Keywords: GMM, KDD, projected outliers

INTRODUCTION

Outlier detection aims at finding out the peculiar, considerably dissimilar, exceptionally and markedly different data items from the massive oceans of the datasets, data streams and databases and hence an important aspects of the data Mining. Detection of outliers is very common and obvious while dealing with lower dimensional data but the major problem of curse of dimensionality arises while dealing with higher dimensional datasets because as soon as the dimensionality of the data goes higher, the prominence of outliers is hidden as the outliers is hidden as the outliers are embedded in some lower dimensional subspaces. As the large and the massive amounts of the DataStream and data are stored in the large databases and datasets, there is a requirement and a need of employing an effective and an efficient methodology for analyzing it to make the information useful that are contained implicitly in the data.

Knowledge Discovery Data or KDD is actually a non trivial process in identifying valid, novel and potentially useful and ultimately understandable knowledge from the data. Usually the research done on KDD includes and focuses on finding out the patterns that are considerably applicable to the portions of the datasets. However, in order to detect the malicious activities like fraud detection, network sensors, spam detection, intrusion detection etc. To find out the anomalous observations among the various data points

is the basic notion to find out the outliers but at the same time it should be crystal clear that it is not always the case, as sometimes they are very much interesting to us, so an occurrence of an event depending upon the situation decides its outlieriness, for example, if a data item has the probability of occurrence is 0.0001 and it happens, it's an outlier but if the data item has the probability of occurrence say, 0.998 and it does not occur, it's again an outlier. So, probability density as well as the likelihood of data item decides it's outlieriness. Similarly, there are many factors and existing methodologies that helps to find out the outliers from the large datasets and databases in lower dimensional subspaces, but as soon as the dimensionality of the data goes higher, the data points or the data items are almost equidistant from each other due to which the higher dimensional outliers or the projected outliers are hidden in the lower dimensional subspaces and they are not highlighted. This problem is usually known as the curse of dimensionality.

One of the solutions to this kind of problem is SPOT i.e., Stream Projected Outlier Detection model (Zhang, 2009) which employs a new window based time model and decaying data summaries to capture statistics from the data streams for outlier detection. Actually to find out the projected outliers from the higher dimensional data is quiet challenging. So, it is utmost important to build the Robust model under the following characteristics or parameters:

- **Scalable:** The system performance should not substantially degrade as the no. of the sequences. Thus there should be an optimized space and Time Complexity.
- **Unique:** The system model should find a unique solution.
- **Comprehensible:** If the system model highlights the set of the projected outliers, it must give the proper explanation that why the particular data objects are termed as anomalous.
- **Robust:** The performance of the model should not be completely dependent on the input provided by the experts. The initial solution should be very rigid and capable enough to handle the adverse conditions.
- **Repeatable:** The model should be capable enough to provide the repeatable solution. Thus the each time when the system is run, it should identify the exactly same amount of outliers.

Clustering is a major aspect to detect the projected outliers in a sense that clustering and its anomalies are complement to each other in a way that the observations which are not included in a cluster are treated as noise, exceptions, novel data items w.r.t that cluster. To make the effective and efficient clusters the outliers must be removed from them which may be sometimes regarded as noise. The presence of outliers may have an adverse effect on the reliable clusters so it is desirable to remove the outliers from the prominent clusters but at the same time one cannot say completely that the outliers are the byproduct of clustering although some of the researchers have agreed to it, that's why there are lots of algorithms and research work that is done which emphasis only on clustering but not on the outlier detection and analysis. So, clustering is a very important technique of anomaly detection. The popular clustering algorithms in context of KDD are CLARANS (Ng and Han, 1994) DBSCAN (Ester *et al.*, 1996), BIRCH (Zhang *et al.*, 1996), STING (Wang *et al.*, 1997), WAVECLUSTER (Sheikholeslami *et al.*, 1998), DenCLUE (Hinneburg and Keim, 1998), CLIQUE (Agrawal *et al.*, 1998) are somehow helpful in detecting the anomalous observations or outliers, but their main objective is to do the clustering instead of outlier detection, so, there is a requirement of a model to optimize the outlier detection in an effective and an efficient way from the higher dimensional outliers. In this study we are trying to build a robust scalable clustering model using the hybrid approach and then to evaluate various parameters against it to prove its validity.

LITERATURE REVIEW

There are many outlier detections methodologies already proposed by so many researchers. Distance based outlier detection techniques were first of all

introduced by Knorr and Ng (1998). According to them, "An object p in the data set DS is a $DB(q, dist)$ outlier if at least fraction q of the objects in DS lie at the greater distance than the distance from p ." This definition can be further generalized and used in various statistical outlier detection methodologies. Then Ramaswamy *et al.* (2000) extended the above definition by providing the proper ranking or the outlier score to all the data points. According to the def, given by them " If there are two integers kn and w , an object p is said to be an outlier if less than w objects have higher value for D^k than p , where D^k denotes the distance of the k th nearest neighbor of the object p . The concept of considering the whole neighborhood of the objects to determine the outliers was given by Angiulli *et al.* (2006) in which all the points were ranked based upon the sum of the distances from the k -nearest neighbors rather than considering individually the distance to the k th nearest neighbor. Breunig *et al.* (2000) proposed a very crucial method Local Outlier Factor (LOF) for each data object in a particular data set that quantifies how outlying an object is by indicating the degree of outlierness. Zhang *et al.* (1996) proposed a method named as Local Distance based outlier detection method for detecting outliers. LDOF of an object determines the degree of the detection of an object from its neighborhood. There are many clustering methods like CLARANS, DBSCAN, BIRCH, CURE to detect the outliers. Moreover Aggarwal and Yu (2000) studied the impact of high dimensionality on the distance based outlier detection but most of these approaches are less meaningful in the higher dimensional data. Lazarevic *et al.* (2013) proposed a feature bagging approach to handle higher dimensionality in which various multiple outlier detectors are combined and then built on arandomly selected subset of features. Moreover, the statistical literature (Rousseeuw and Leory, 1987; Rocke and Woodruff, 1996; Atkinson, 1994) extensively includes the various outlier detection methods based on Mahalanobis distance. To provide the robustness, the scattered matrix and locations are used. Rousseeuw and Van Driessen (1999) provide more robustness to the MD based methods feasible for large sample size data. Danuser and Stricker (1998) provided the framework of least squares fitting of multiple parametric models. The classification algorithm for the projected outliers which was developed on the basis of the robust dimensionality reduction technique was introduced by Fidler *et al.* (2006). Time series using an autoregression model was proposed by Takeuchi and Yamanishi (2006). Pearson was first who introduced the case of Gaussian mixture with two components and also focussed various methods to estimate the mixture parameters (Dempster *et al.*, 1977) then provided the theoretical framework of EM algorithm. Hasselblad (1966), Day (1969), Wolfe (1970) and Duda and Hart (1973) also provided the detailed treatment of mixture

densities and EM algorithm. Roeder and Wasserman (1997) provided the detailed illustration of the Bayesian perspective on density estimation using Gaussian Mixtures. Richardson and Green (1997) provided the extension of Bayesian sampling to be cases where the no. of Gaussian components are unknown.

AN OVERVIEW OF THE PROPOSED APPROACH

The proposed method has been made up after concluding the literature survey. It utilizes the basic notion of Distance based outlier detection and then to detect the degree of outlierness, density based outlier detection used. To provide the robustness to the model, Gaussian Mixture Model is used which calculates the probability density function of the data objects and finally Baye's probability is used to confirm the outliers as Projected outliers.

Distance based outlier detection method: The modern distance based approach states a simple notion that a data point is said to be an outlier if its locality is sparsely populated (Lazarevic *et al.*, 2013). It is one of the important methods of outlier detection which is distribution free as well as easily applicable for the various types of data. According to distance based approach, Given a dataset X, an object $x \in X$ is a DB(α , δ)- outlier if:

$$|\{x' \in X \mid d(x, x') > \delta\}| \geq \alpha_n \quad (1)$$

Here, $n = |x|$, the no. of objects $\alpha, \delta \in R (0 \leq \alpha \leq 1)$ are parameters There are two drawbacks of the distance based methods:

- Setting the distance threshold δ is quite difficult in practice but setting α is not so difficult because it is always close to 1.
- The lack of the ranking of the outliers as it always gives the binary result of whether the data object is an outlier or not. K^{th} nearest neighbor, i.e., KNN score $q_k^{th} NN(x) = d^k(x; X) - dk(x, X)$ is the distance b/n x and its k^{th} NN in X, but again the two drawbacks of K^{th} nearest neighbor Approach is:
 - **Scalability:** $O(n^2)$ and its solution is the partial computation of the pairwise distances to compute the scores only for the top t outliers.
 - Detection ability and its solution is to introduce the degree of outlierness by switching to density based methods.

Density based outlier detection method: The basic idea of the density based outliers is to calculate and

compare the densities around a point with the densities around it's local neighbors and that related density when computed becomes the outlier score. The general assumption of the density based outlier methods is that around the normal object the density of the data object is similar to the densities around its neighbors whereas the density of outlier is quite different to the density around it's neighbors. LOF or Local outlier Factor is based on pairwise distances and is a very prominent outlier detection method. The local outlier factor (LOF) $q_{LOF}(x)$ is defined as the ratio of the local reachability densities of x and the average of the local reachability densities of the k^{th} nearest neighbors. Consider the following:

The reachability distance is: $Rd(x, x') := \max \{d^k(x', x), d(x, x')\}$.

The local reachability density is:

$$\Delta x = \left(\frac{1}{|N^k(x)|_x} \sum_{x' \in N^k(x)} Rd(x; x') \right)^{-1} \quad (2)$$

The LOF is defined as:

$$LOF(x) = \frac{\left(\frac{1}{|NR(x)|} \sum_{y \in N^k(x)} \Delta y \right)}{\Delta x} \quad (3)$$

In simple and straightforward calculation, $LOF = 1$ implies point is in a cluster i.e., region with homogeneous density around the points and its neighbors and if $LOF \gg 1$: implies point is simply an outlier as mostly the interest is always shown in top n outliers, so LOF saves the time as there is no need to compute LOF for all the data points thereby saving the overall run time.

Gaussian mixture model: GMM is one of the robust techniques to detect the anomalous observations and is generally used for the multimodal forms of the data distributions mixture models and the GMM parameters are estimated through the Expectation Maximization algorithm and it is also effectively applied for the clustering and classification. GMMs were initially used for the structural damage detection, then further it was used with Mahalanobis distance given by Nair and they and proved useful in solving damage detection. GMM model consists of the means, co-variances and a probabilistic assignment of every data point to the Gaussians GMM is the sum of K-Gaussian densities and it has the following forms:

$$p(x) = \sum_{k=1}^K \Pi_k N\left(\frac{x}{\mu_k}, \Sigma_k\right) \quad (4)$$

Every Gaussian component containing the mixture $n(x/\mu_k, \Sigma_k)$ has its own mean μ_k and covariance Σ_k . The parameters Π_k are called mixing coefficients which satisfies:

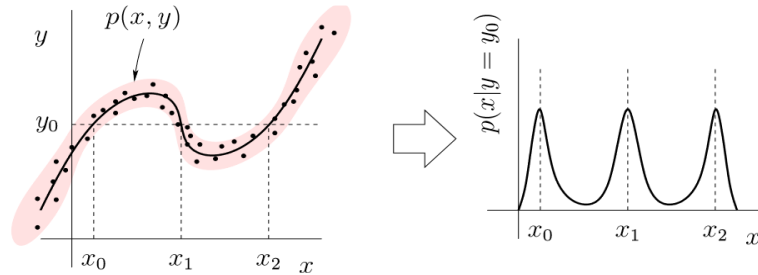


Fig. 1: Modes at conditional distribution

$$\sum_{j=1}^k \prod_k = 1 \text{ and } 0 \leq \prod_k \leq 1 \tag{5}$$

Gaussian Mixture distribution is used to maximize the log of the likelihood function (Dempster *et al.*, 1977) and the EM framework is used which is given by the following steps:

- The parameters of GMM are initialized, i.e. means μ_k , covariance Σ_k and mixing coefficients Π_k
- E-Step(Expectation): In this step, the responsibilities $\gamma(Z_n, k)$ are evaluated using the most recent values of GMM parameters:

$$\gamma(Z_n k) = \frac{\pi_k N(\frac{x_n}{\mu_k}, \Sigma_k)}{\sum_{j=1}^k \pi_j N(\frac{x_n}{\mu_j}, \Sigma_j)} \tag{6}$$

where, $Z_n k$ is an element of k - dimensional binary random variable z for n th learning pattern.

M-Step (Maximization Step): In this step, the weighted means and variances are recomputed:

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(Z_n k) x_n \tag{7}$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(Z_n k) (x_n - \mu_k^{new})(x_n - \mu_k^{new})^T \tag{8}$$

$$\Pi_k^{new} = \frac{N_k}{N} \tag{9}$$

Here, μ_k^{new} is a scalar whereas Σ_k^{new} denotes the d dimensional vectors where $N_k = \sum \gamma(Z_n(k))$ and N is the no of learning patterns. Figure 1 illustrates this: These steps are intended till the convergence.

Baye’s probability: Suppose, $X = (x_1, x_2, x_3, \dots, x_n)$ denotes the set of N observations from d dimensional space. If there are two classes say one of outlier and another of normal class N . Let the series of the projected outliers be $M_0 = \{m_1, m_2, m_3\}$ be the corresponding outlier score of the higher dimensional projected outliers assigned to X . If there is no loss of generality, let us assume that the higher the value of π_i , more likely x_i is an outlier. i.e., $\pi_i = P(O | m_i)$.

The probability that x_i is normal can be computed accordingly by $P(N | m_i) = 1 - \pi_i$. According to it:

$$P\left(\frac{O}{m_i}\right) = \frac{P\left(\frac{m_i}{O}\right)P(O)}{P\left(\frac{m_i}{O}\right)P(O) + P\left(\frac{m_i}{N}\right)P(N)} \tag{10}$$

$$P\left(\frac{O}{m_i}\right) = \frac{1}{1 + \exp(-a_i)} \tag{11}$$

Embedded approach and exact procedure:

- Take the real or synthetic data that is basically used for clustering from UCI repository.
- Make the absolute clusters from it using the K-means clustering algorithm and then the cluster validation problem is resolved from it by choosing or applying the appropriate threshold to it.
- Out of these absolute clusters, firstly the distance based outlier detection technique is applied, i.e., the distance of the data items from the cluster centroid is calculated first and its distance from the k -nearest neighbor is calculated.
- Attribute Relevance Analysis is done so as to find the projected clusters. All the irrelevant attributes contain the noise or the outliers which are also termed as sparse data points whereas the relevant attributes exhibit some cluster structure (Agrawal *et al.*, 1998). Higher the density of the points than it’s surrounding regions highly clustered the region would be. In order to investigate the densely populated region, the sparseness degree has to be computed and hence the density based outliers has to be applied. Then the local outlier factor has to be calculated for the projected clusters with the notion:

LOF = 1 {data item is a normal object}
 LOF >> 1 {data item is simply an outlier}

- After, the distance and density based methodology, apply the Gaussian Mixture Model to the output of both the distance and density based methods in order to calculate the probability density function. To do so, M-Step-Re-estimation of GMM parameters to upto date values of responsibilities $\gamma(Z_n k)$. Free parameters of the Gaussian mixture model consist of the means and covariance matrices of the Gaussian components and the weights indicating the contribution of each Gaussian to the approximation of $P(X | C_j)$. The Log-Likelihood function is given:

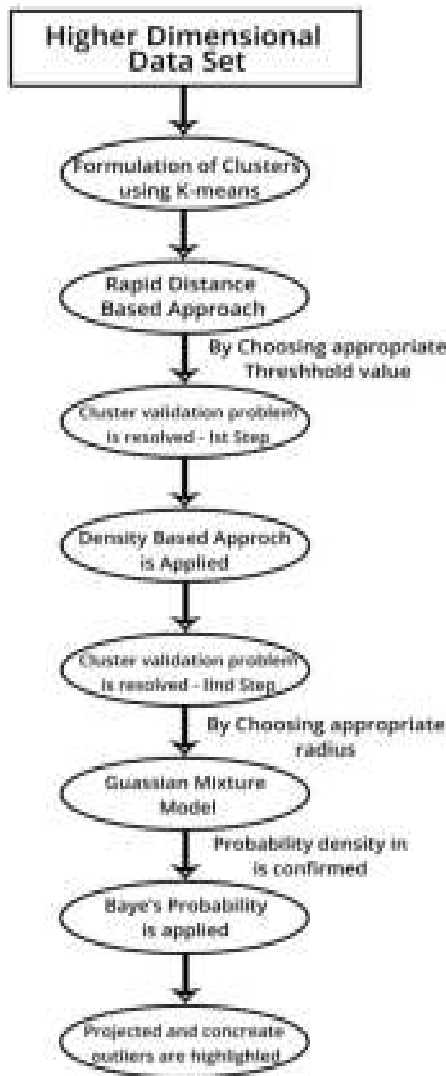


Fig. 2: Flow of the proposed work

$$L_j = \prod_{i=0}^{N_{train}} P\left(\frac{X_i}{C_j}\right) \rightarrow l_n(L_j) = \sum_{i=0}^{N_{train}} l_n\left(P\left(\frac{X_i}{C_j}\right)\right) \quad (12)$$

This step helps us to find the projected outliers.

- Then Baye's Probability is applied to the output obtained from the Gaussian Mixture Model. Baye's probability relates the posterior density functions with the current probability density function to calculate or check the likelihood of a point and if there is a quiet or a little chance of categorizing a point as a normal one or it can be included in the cluster, then that point is pruned out. In this way, the final output obtained will be completely a refined and an exact stream of the projected outliers (Fig. 2).

RESULTS AND DISCUSSION

Initial results: First of all, absolute clusters are formed using K-means clustering algorithm, then the distance based outlier detection is applied to detect the outliers. Then the cluster validation problem is resolved for it by choosing the exact threshold value. This threshold should be exact for the particular clusters so that neither too much outliers are highlighted nor the true outliers should be hidden. Secondly density based approach will exploit the further higher dimensional outliers. Then again the clusters are validated by choosing the appropriate radius value. The final projected outliers are highlighted after estimating probability density functions and conditional probabilities by using Gaussian Mixture model and Baye's Probability respectively. Probability density function performs the relevance attribute analysis and Baye's probability converges the attributes by applying the conditional check to highlight the projected outliers.

This Experiment is done on the real dataset i.e., iris data set taken from UCI Repository, it's a multivariate dataset with 150 instances and 4 attributes. Firstly the work is done on only two attributes sepal length, sepal width and then it is enhanced for another two attributes petal length and petal width to test the scalability of our model. In order to generate the exact accuracy, the absolute clusters are formed by validating the threshold value of the cluster and this threshold value is tested under various different values of radius. Then it's log-

Table 1: No of the Projected outliers and their Elapsed time of various validated clusters

Threshold (1st step cluster validation in %)	Radius value (IInd step cluster validation)	Log-Likelihood	No of iterations	Projected outliers	Time elapsed (in milliseconds)
80	0.200	15.6658	39	94	0.6702764
80	0.230	8.6652	41	54	0.737656
80	0.250	11.9354	72	54	0.670268
80	0.270	2.26815	38	54	0.670107
80	0.285	18.8681	30	47	0.653342
80	0.300	8.81027	37	47	0.711358
90	0.200	17.1259	32	13	0.646975
90	0.230	8.66552	24	13	0.649362
90	0.250	5.23091	100	13	0.669489
90	0.270	3.1248	96	13	0.660074
90	0.285	18.8681	39	13	0.673496
90	0.300	8.65112	43	13	0.821260

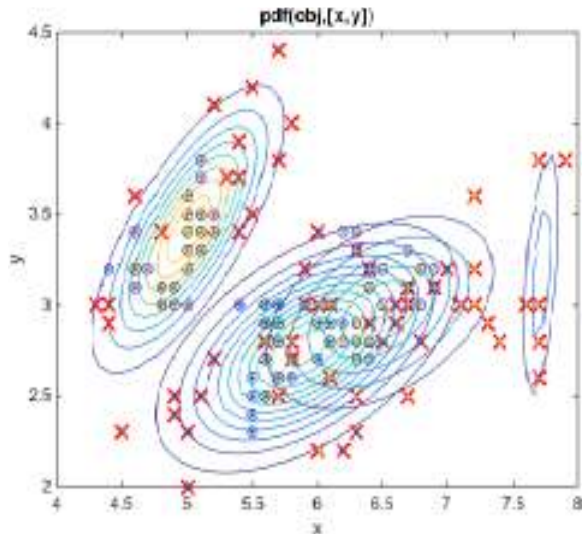


Fig. 3: Projected outliers at 80% threshold with radius 0.200

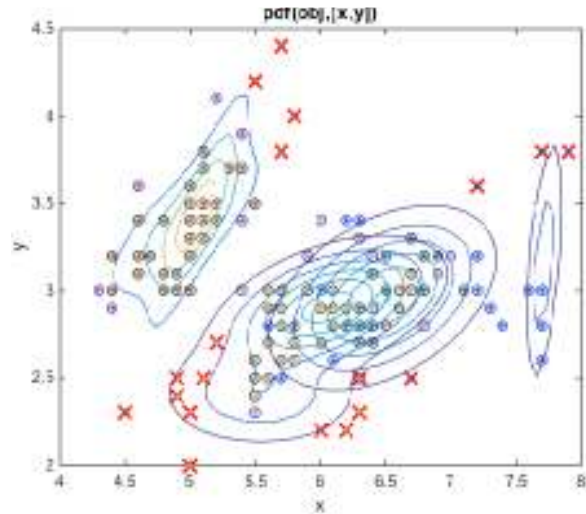


Fig. 6: Projected outliers at 90% threshold with radius 0.300

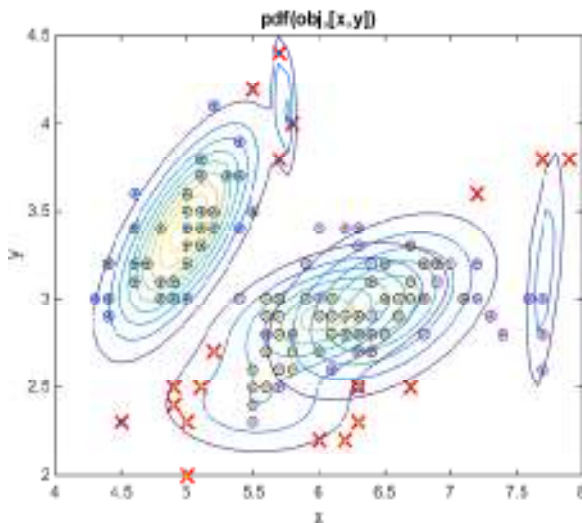


Fig. 4: Projected outliers at 80% threshold with radius 0.300

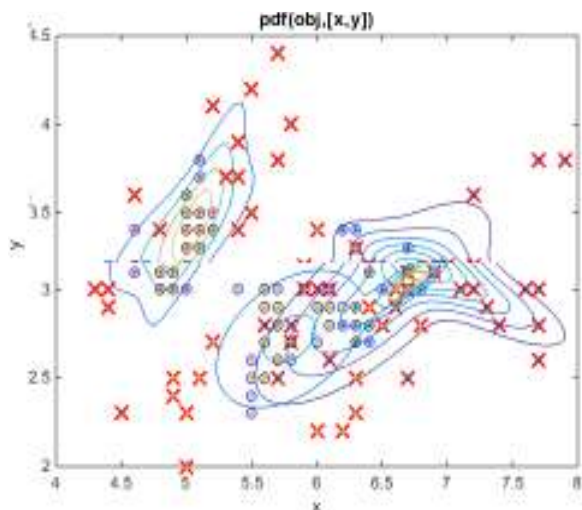


Fig. 5: Projected outliers at 90% threshold with radius 0.200

likelihood value is estimated using the expectation step of Gaussian Mixture and finally the no of projected outliers and the time elapsed in executing them is shown in the Table 1. Consider the Fig. 3 and 4 obtained after executing the model in matlab software which shows both the clusters and outliers are also highlighted. In these two figures the clusters are shown at a proper threshold value 80% and 90% at their extreme radius points so as to resolve the cluster validity problem and to obtain the absolute clusters. The attribute relevance analysis is done which helps in converging the large no of data points within the clusters which are marked as circles in Fig. 3 and 4 and the outliers of high projection are marked as small crosses.

Performance evaluation: Our main motive is to construct the scalable model that should be Robust enough to handle the large no. of outliers. The various parameters are calculated for the above shown results (Fig. 5 and 6).

Scalability: When the scalability is enhanced, for further more dimensions even then the performance will not degrade and the following results will be shown as in the fig. This fig shows the results when two more attributes are included in the data set. Now, the projected outliers are also enhanced at both 80 and 90% threshold value. The performance is not degraded as time elapsed is also comparatively very less.

Uniqueness: In Table 1 and 2 every time when the model is executed, it provides unique no of outliers. At different threshold and different radius, a varying no of outliers with varying radius are shown.

Time complexity and space complexity: The Time elapsed by the projected outliers is calculated to be less

Table 2: No. of the projected outliers and their elapsed time of various validated clusters with enhanced dimensions

Cases	Threshold (%)	Radius value	Projected outliers	Time in milliseconds
1	80	0.200	118	1.161
2	90	0.200	113	1.158
3	80	0.230	71	1.145
4	90	0.230	62	1.133
5	80	0.250	71	1.140
6	90	0.250	62	1.132
7	80	0.270	71	1.139
8	90	0.270	62	1.131
9	80	0.285	65	1.135
10	90	0.285	53	1.128
11	80	0.300	65	1.132
12	90	0.300	51	1.127

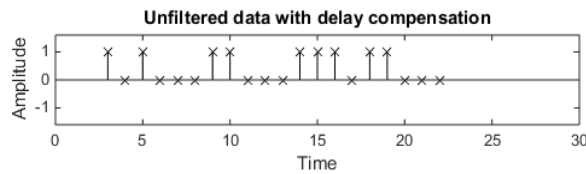


Fig. 7: Comparison of time with amplitude

than order of $O(n^2)$ and the CPU burst time and the self time is calculate to be 0.48 and 0.16 milliseconds. Individual time elapsed for all the cases are shown in Fig. 7.

Usually, in the higher dimensional data, when the dimensionality is enhanced, the time is increased and it degrades the system's performance upto a large extent, but in our case, when the dimensionality is enhanced although the time is increased (Table 2) but still the performance is not degraded as shown in Fig. 7. The unfiltered data after delay composition will remain as it is and as it is not affecting the amplitude and hence it will not affect the system's overall performance.

Curse of dimensionality: This is the problem where the higher dimensional or projected outliers are hidden in the lower dimensional subspaces and hence their prominence is overlooked which has a major impact at the overall performance of the system. In order to resolve this problem, we have found out the individual probability densities of the clusters using Gaussian mixture and to confirm their outlierness, conditional probability is used using Bayes Probability. Hence, the projected outliers that are hidden in lower dimensional subspaces are highlighted and become more prominent.

Robustness: The enhancement of the scalability does not compromise with the quality and quantity of the outliers; moreover, it does not enhance the time and the space complexity of the outliers, so the overall performance of the system remains unaffected, hence the system is Robust under the severe conditions.

CONCLUSION

In this study, we proposed a robust and a scalable model using hybrid approach and clustering. After

constructing the clusters using K-means, different other methodologies like Distance Based, Density Based, Gaussian Mixture and ultimately Bayes probability is applied. At varying radius and threshold the quantity and the quality of the outliers is checked and the model is finally tested under various parameters like Scalability, Robustness, Time Complexity etc at initial position and when the dimensionality is enhanced, the performance is no impacted.

REFERENCES

- Aggarwal, C.C. and P.S. Yu, 2000. Finding generalized projected clusters in higher dimensional spaces. Proceeding of the ACM SIGMOD International Conference on Management of Data, pp: 70-81.
- Agrawal, R., J. Gehrke, D. Gunopulos and P. Raghavan, 1998. Automatic subspace clustering of high dimensional data for data mining applications. Proceeding of the ACM SIGMOD International Conference on Management of Data, Seattle, WA, pp: 94-105.
- Angiulli, F., S. Basta and C. Pizzuti, 2006. Distance-based detection and prediction of outliers. IEEE T. Knowl. Data En., 18(2): 145-160.
- Atkinson, A.C., 1994. Fast very robust methods for the detection of multiple outliers. J. Am. Stat. Assoc., 89(428): 1329-1339.
- Breunig, M.M., H.P. Kriegel, R.T. Ng and J. Sander, 2000. LOF: Identifying density-based local outliers. Proceeding of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD '00). ACM, New York, USA, pp: 93-104.
- Danuser, G. and M. Stricker, 1998. Parametric model fitting: From inlier characterization to outlier detection. IEEE T. Pattern Anal., 20(3): 263-280.
- Day, N.E., 1969. Estimating the components of a mixture of normal distributions. Biometrika, 56(3): 463-474.
- Dempster, A.P., N.M. Liard and D.B. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. B Met., 39: 1-38.
- Duda, R.O. and P.E. Hart, 1973. Pattern Classification and Scene Analysis. Wiley, New York.

- Ester, M., H.P. Kriegel, J. Sander and X. Xu, 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceeding of the 2nd International Conference on Knowledge Discovery and Data Mining. AAAI Press, Portland, OR, pp: 226-231.
- Fidler, S., D. Skocaj and A. Leonardis, 2006. Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. *IEEE T. Pattern Anal.*, 28(3): 337-350.
- Hasselblad, V., 1966. Estimation of parameters for a mixture of normal distributions. *Technometrics*, 8(3): 431-444.
- Hinneburg, A. and D.A. Keim, 1998. An efficient approach to clustering in large multimedia databases with noise. Proceeding of the 4th International Conference on Knowledge Discovery and Data Mining. New York City, pp: 58-65.
- Knorr, E.M. and R.T. Ng, 1998. Algorithms for mining distance based outliers in large datasets. Proceeding of the 24th International Conference on Very Large Data Bases (VLDB, 1998), pp: 392-403.
- Lazarevic, A., L. Ertöz, A. Ozgur, J. Srivastava and V. Kumar, 2013. A comparative study of anomaly detection schemes in network intrusion detection. Proceeding of the 3rd SIAM Conference on Data Mining.
- Ng, R.T. and J. Han, 1994. Efficient and effective clustering methods for spatial data mining. Proceeding of the 20th International Conference on Very Large Data Bases. Morgan Kaufmann Publishers, Santiago, Chile, San Francisco, CA, pp: 144-155.
- Ramaswamy, S., R. Rastogi and K. Shim, 2000. Efficient algorithms for mining outliers from large data sets. Proceeding of the ACM SIGMOD International Conference on Management of Data (SIGMOD '00), pp: 427-438.
- Richardson, S. and P.J. Green, 1997. On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Stat. Soc. B*, 59: 731-792.
- Rocke, D.M. and D.L. Woodruff, 1996. Identification of outliers in multivariate data. *J. Am. Stat. Assoc.*, 91(435): 1047-1061.
- Roeder, K. and L. Wasserman, 1997. Practical Bayesian density estimation using mixtures of normals. *J. Am. Stat. Assoc.*, 92(439): 894-902.
- Rousseeuw, P.J. and A.M. Leory, 1987. Robust Regression and Outlier Detection. John Wiley and Sons, NY.
- Rousseeuw, P.J. and K. Van Driessen, 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3): 212-223.
- Sheikholeslami, G., S. Chatterjee and A. Zhang, 1998. WaveCluster: A multi-resolution clustering approach for very large spatial databases. Proceeding of the 24th VLDB Conferences. New York, USA, pp: 428-439.
- Takeuchi, J. and K. Yamanishi, 2006. A unifying framework for detecting outliers and change points from time series. *IEEE T. Knowl. Data En.*, 18(4): 482-492.
- Wang, W., J. Yang and R. Muntz, 1997. STING: A statistical information grid approach to spatial data mining. Proceedings of the 23th International Conference on Very Large Data Bases (VLDB). Morgan Kaufmann Publishers, Athens, Greece, San Francisco, CA, pp: 186-195.
- Wolfe, J.H., 1970. Pattern clustering by multivariate mixture analysis. *Multivar. Behav. Res.*, 5(3): 329-350.
- Zhang, J., 2009. Towards outlier detection for high-dimensional data streams using projected outlier analysis strategy. Ph.D. Thesis, Dalhousie University.
- Zhang, T., R. Ramakrishnan and M. Linvy, 1996. BIRCH: An efficient data clustering method for very large databases. Proceeding of the ACM SIGMOD International Conference on Management of Data. ACM Press, New York, pp: 103-114.