ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 9(17), DOI: 10.17485/ijst/2016/v9i17/92994, May 2016* ISSN (Online) : 0974-5645

Detection of Crimes using Unsupervised Learning Techniques

R. Bulli Babu*, G. Snehal and P. Aditya Satya Kiran

Department of Electronics and Computer Engineering, K. L. University, Guntur - 522502, Andhra Pradesh, India;
babuklu123@kluniversity.in, snehalganguri25595@gmail.com, adityakrn7@gmail.com

Abstract

Objectives: The main objective of this paper is to solve the criminal problems with in less amount of time. There are many methods to do so but this paper concentrates in solve the easily and reduce the time in solving the case. **Methods:** To solve the criminal cases with in less time there are many methods but here we used clustering technique. Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). When a case is enrolled into the data base before if there is any case similar to it then we can solve the case easily by doing the same procedure. **Findings:** Before they used to file a case on FIR. But now a day, they are using data bases to file a case. By getting any new case they are comparing the new case with the older case so that it will be easy to find the suspect as it takes less time to solve the case. Before they used for other techniques like classification etc. But in my findings and research work clustering is simple, more accurate and takes less time to solve the case easily. In clustering techniques also we have different type of algorithm, but in this paper we are using the k-means algorithm and expectation – maximization algorithm. We are using these techniques because these two techniques come under the partition algorithm. Partition algorithm is one of the best method to solve crimes and to find the similar data and group it. K-means algorithm is done by partitioning data into groups based on their means. K-means algorithm has an extension called expectation - maximization algorithm here we partition the data based on their parameters. **Applications:** This system can be used for the Indian crime departments for reducing the crime and solving the crimes with less time. This technique can be used to solve the crimes with in less time.

Keywords: Clustering, Data Mining, Expectation– Maximization, K-Means, Unsupervised

1. Introduction

Data collection and storing that data during the past decades is difficult and referring it for new crime is also difficult as we have to refer all the crimes from the starting which crime can be similar to it. Document analysis can be difficult to solve a crime. If there are more documents to solve the crimes we have to study more documents and refer to them and understand them is difficult[1]. To solve this problem we use computer forensics we can solve the crimes fast and it is fast growing field where it can easily be examine the evidence. In case if there is any damage to the computer we can recover the stored data there will be no loss of information. By using this digital content it is easy to solve the crime. Even this data can be hidden where others can't see this data. This can be done by using the user name and password[2]. By this only the officers related to that investigation can see the data. We can also know identify that who logged in to the criminal data. Because of this it reduces the manual effort, time, redundancy and to solve the crimes easily. By using the digital device we can store large amount of data.

There are some methods already presented by different researchers to analyze the multiple documents. Existing methods is DFI propose multi-level search approach, it gives the accurate results and also produce the evidence related to the current investigation. The drawback of these methods has no provision for end user. Here the end user has to search the data relevant to that task or group the data on given subject. The DIF system takes the

* *Author for correspondence*

input as text file in unstructured format. This data will be converted to the structured form by using different data mining techniques. There are many clustering algorithms to group the relevant data can be used to analysis the crime. Such methods are used for the data analysis which has less or no prior information about the input data. All digital data produce applications of end results[3]. Data sets are made up of unlabeled sets or classes of data which is identified as unknown initially. Even if we consider the availability of labeled dataset there is no certainty that classes that are available of labeled dataset in input dataset or next raw dataset which is being collected through different computers or related to different investigations[4]. The unstructured data sample can be of different sources. To provide an efficient solution for such heterogeneous data, we use clustering techniques. These clustering techniques are used to find the related document by using patterns[5]. This algorithm improves the performance by end users. The method of this cluster is to group the data related to each other with some similarities which we define it as cluster. Similarly we have different type of clusters grouping with some similarities[6]. The investigators can easily find the related document from which cluster they required. By this we can examine other documents with each cluster. By this we can easily solve the difficult tasks by analyzing the documents easily and it also saves the time compared to earlier.

In the recent investigation we have studied the work done on different clustering algorithms such as k-means, single link, complete link, average link with different digital forensic datasets in[7]. In[8], author presented the methodology for the clustering algorithms which were used for the forensic analysis of data/evidence in the criminal cases are being investigated by detectives. In this paper we are analyzing the partition algorithm for the criminal data.

2. Approach used by Clustering Algorithms

2.1 Types of Clustering

Clustering is an unsupervised task without having a priori knowledge by discovering groups of similar documents. There are two types of categories in clustering algorithms, they are the partitional algorithm and the hierarchical algorithm. K-Means algorithm and the link clustering they come under these two categories. K-Means and hierarchical clustering have many comparisons. In hierarchical clustering the size of data increases as the computational expansive, since to merge small clusters and D_D similarity matrix by using the certain link functions[9]. By comparing with them K-Means is faster. It updates the centroid clusters with each iteration and reallocates each document by its nearest centroid by this we can say that it is an iterative algorithm. Comparison of K-Means and hierarchical algorithm[10].

2.1.1 K-Means Algorithm

K-Means clustering investigation plans to partition n perceptions into k bunches during which each perception includes a place with the bunch with the nearest mean.

Algorithm

For partitioning the K-Means algorithm, where the mean value the objects in the cluster is represented by each cluster.

Input: Number of groups.

- Place K focuses into the space represented by the data that are being grouped. These focuses represent the group centroids initially.
- Cluster that has nearest centroid assign every data to it.
- When the objects are assigned, recalculate the position of the k centroids.
- Until the centroids has no move repeat the steps of 2 and 3. This type of partitioning of data from the centroid to be minimized can be calculated.

output: An arrangement of k groups. To locate the mean qualities K-Means calculation is a base for all other grouping.

The K-Means algorithm does not find the corresponding to the global objective function minimum, to find the most optimal configuration. We can reduce the effect by running the K-Means for multiple times.

2.1.2 Expectation-Maximization Algorithm

Expectation-Maximization is a type of model based clustering method. Expectation-Maximization calculation can be utilized to discover the parameter gauges for every cluster[11]. It is an expansion of K-Means algorithm.

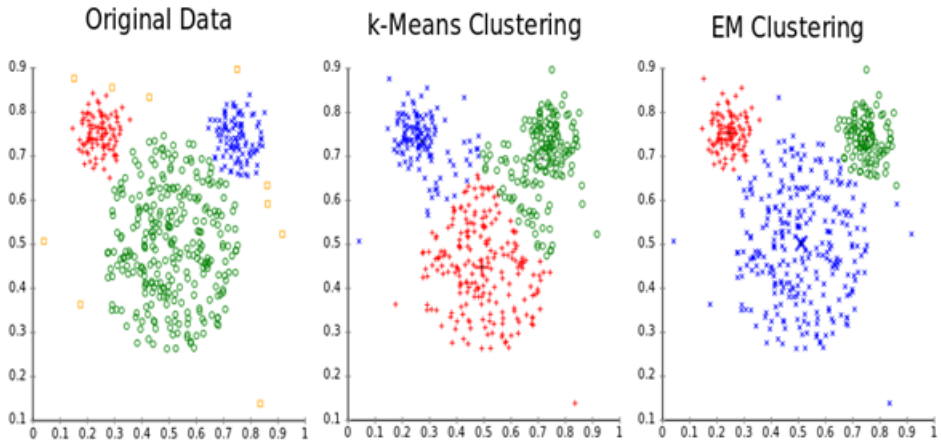

Figure 1. Results of the K-means and expectation-maximization algorithms.

Algorithm

- Find the initial input by finding the initial centroid.
- By using the cosine distance formula or any other distance formula calculate the distance between each centroid and each data point.
- As per the probability of membership of a data pint to a particular cluster assign the weights for each combination of data point and cluster.
- Repeat
- Which has highest weight reassign each data point to the cluster i.e., highest probability.
- If a data point belongs to more than one cluster with the same probability, depending on the minimum distance, (re)assign the data point to the cluster.
- Renovate the cluster means foe every iteration until clustering converges.

5. Results and Analysis

In Figure 1, in the original data there are groups which are not exactly clustered, some data is not grouped here. In K-Means clustering the data is grouped based on the algorithm, here it finds the nearest centroid and groups the data according to the nearest centroid. In Expectation-Maximization is the extension of K-Means here the data is expected and it is grouped to the exactly nearest cluster.

6. Conclusion

In crime data clustering techniques plays a vital role to investigate the crime and it helps for solving the unsolved crimes easily. By grouping the data with similar objects we can easily solve the unsolved crimes. For finding similarity objects partitioning clustering algorithm is one of the finest method. It is observed that finding similar words and collect them in a single cluster which helps in crime analysis. This paper deals with the study of clustering techniques and affinity measures in crime data.

7. References

1. da Cruz Nassif LF, Hruschka ER. Document clustering for forensic analysis: An approach for improving computer inspection. IEEE Transactions on Information Forensics and Security. 2013 Jan; 8(1):46–54.
2. Gupta M, Chandra. B, Gupta MP. Crime data mining for Indian police information system; 2007. p. 1–10.
3. Hussain KZ, Durairaj M, Farzana GRJ. Application of data mining techniques for analyzing violent criminal behavior by simulation model. IJCSITS. 2012; 2(1):1–5.
4. Malathi A, Baboo SS. Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters. Global Journal. 2011 May; 21(1):1–6.
5. Akhter MI, Ahamad MG. Detecting telecommunication fraud using neural networks through data mining. International Journal of Scientific and Engineering Research. 2012; 3(3):601–6.
6. Rizwan I, Masrah AAM, Aida M, Payam HSP, Nasim K. An experimental study of classification algorithms for crime prediction. Indian Journal of Science and Technology. 2013 Mar; 6(3):4219–25.
7. Cheng H, Hua KA, Vu K. Constrained locally weighted clustering. Journal Proceedings of the VLDB Endowment. 2008; 1(1):90–101.
8. Stoffel K, Cotofrei P, Han D. Fuzzy methods for forensic data analysis. European Journal of Scientific Research. 2014 Aug; 2(8):1–5.

9. Prakash A, Chandrasekar C. An optimized multiple semi-hidden markov model for credit card fraud detection. Indian Journal of Science and Technology. 2015 Jan; 8(2):165–71.
10. Rizwan I, Masrah AAM, Aida M, Payam HSP, Nasim K. An experimental study of classificationalgorithms for crime prediction. Indian Journal of Science and Technology. 2013 Mar; 6(3):4219–25.
11. Ramageri BM. Data mining techniques and applications. Indian Journal of Computer Science and Engineering. 1(4):301–5. Available from: http://www.ijcse.com/docs/IJCSE10-01-04-51.pdf