

Group Identification Based Classification Technique for Aggregated Common Data Used in Digital Forensic Investigation (DFI)

Amit Pratap Singh, Dr. R.C. Jain

Amit Pratap Singh, Research Scholar, Samrat Ashok Technical Institute (SATI), Vidisha, M.P., India, Mob. 7869406534
amitjmail@gmail.com

Abstract— Several files have been examined in the computer system forensic observation, These files made up of unstructured text therefore it is hard to analyze them, this leads to the introduction of the automated techniques, The clustering in any algorithm provides fresh and important information by observing the documents, We have shown this work process to convert the clustering algorithm to the forensic analysis of those computer, We can get this by modifying the algorithm, which differ from the content that is used to make a cluster. Here we deal with the work process for the clustering of the document in the forensic investigation. here we relate it with the K-means algorithm, With the help of some experiment that's results like there is a improvement in the performance ratio of the computers, We have also focus on some other results which are essential for the researchers of forensic computing .This work become active by the application of the MATLAB.

Keywords— Clustering, Forensic Investigation, Weblog, XML, Kmeans.

INTRODUCTION

Clustering [1,2,3] is a common technique for statistical data analysis, which is used in many fields including identifying known pattern, data mining, image synthesis, machine learning and others. Clustering is the process of grouping similar objects into different groups, or more precisely, the partitioning of a data set into subsets, so that the data in each subset according to some defined distance measure. This well-known technique generates the various subsets of instances in a manner, so that the instances present in one subset are similar in some nature and those present in different subset are dissimilar. Thus, these instances are effectually organized by using clustering technique.

In a formal manner clustering is defined as the subset of different cluster as $C = C_1, \dots, C_k$ of S , such that: $S = \bigcup_{i=1}^k C_i$ and $C_i \cap C_j = \emptyset$ for $i \neq j$. Significantly single instance belong to one subset of cluster. The idea of clustering the object is as primitive as when human tried to define the prominent feature of man and object and recognize them by using a type. Thus, this concept was adopted in wide range of scientific discipline: ranges from branch of mathematics to medical science, in particular area this is known by their specific term that generate topology based on analysis. From biological "taxonomies", to medical "syndromes" and genetic "genotypes" to manufacturing "group technology" — the job is similar; determine the classes and assign each individual to its specific class based on evaluating characteristics.

As this method requires grouping of similar instances into single unit therefore it is essential to have some sort of measures that estimate that two instance are similar in some respect or not. Here we discussed the two well known parameter used to estimate the above relationship is: "Distance" and "similarity". A large number of clustering techniques used the distance measure to determine the relationship among the different instances present. The distance parameter between two given instance X_i and X_j is defined as: $d(X_i, X_j)$. In case of similar instance the value of this evaluated distance measure is minimum nearly equal to zero. This distance measure is also called as metric of distance and it must fulfill the two properties given below:

1. inequality of triangle: $d(x_i, x_k) \leq d(x_i, x_j) + d(x_j, x_k) \forall x_i, x_j, x_k \in S$.
2. $d(x_i, x_j) = 0 \Rightarrow x_i = x_j \forall x_i, x_j \in S$.

CLUSTERING

This section provides the brief overview of popular clustering methods [1]. As there is not a general understanding how the "cluster" is accurately defined this was the reason for proposing large number of clustering methods. Thus, there exists large number of clustering algorithm each of which uses different principle of initiation. Firstly the "Farley and Raftery" -1998 classified the clustering methods into two basic group: "hierarchical" and "partitioning" clustering. Followed by this other classification proposed by "Han and Kamber" -- 2001 divides the clustering methods into three basic group such as "density-based", "model-based" and "grid-based" clustering. Esti vill-Castro, 2000 classified another method depending on the principle of initiation.

CLUSTERING METHODS

A) Hierarchical Methods:

Under this method of clustering, cluster is formed by iteratively applied hierarchical approach in top-down or bottom-up fashion. This divides the method in two groups:

- [1] Agglomerative hierarchical clustering — At initial phase of it individual object represent the cluster of its own, but as proceeds the cluster are integrated till the desired structure of cluster is obtained.
- [2] Divisive hierarchical clustering — At the initial phase of this entire object belong to single cluster but as proceeds this segment into sub-cluster, after this, obtained sub-cluster is again segmented into sub-cluster and this continued till the desired structure of cluster is obtained.

This integration or segmentation of cluster is performed based on some measure of similarity, this measure be selected in a way so that it optimized the result produced. Furtherance the method of hierarchical clustering techniques is divided based on the evaluated measure of similarity.

- Single-link clustering- the nearest neighboring technique — consider the distance between the cluster; this considered distance is equitable to the smallest distance from any member of one cluster or another. This similarity measure is defined as- the similar parameter between the clusters is equitable to the largest similarity parameter from the any member of one cluster to another. (Proposed by “Sneath and Sokal”-- 1973).
- Complete-link clustering- furthest neighboring technique - considers the distance between the clusters; this considered distance is equitable to the largest distance from any member of one cluster or another. (proposed by King, 1967).
- Average-link clustering (minimum variance method) - consider the distance between the clusters; this considered distance is equitable to the average of distance from any member of one cluster or another. This category of clustering method suggested by “Ward”, 1963 and “Murtagh”, 1984).

Fig. 1. Partitioning Methods :

This group of clustering i.e Partitioning clustering [2] moves the location of instances from one cluster to another, initiate this with the initial partition. This type of clustering techniques need to pre-defined the number of cluster by the user. In this method of clustering to obtain the global optimization this comprehensibly enumerates all the available partition. As this is not the feasible way so some greedy measures are adopted in fashion of repeated optimization. Commonly here a relocation method is implied to move the location of cluster from one to another. Below here gives the brief overview of some of the method of partitioning cluster.

- [2] Error Minimization Algorithms. This category of clustering is well suited to use in the situation having isolated and compact cluster, this method of clustering is frequently applicable. The basic principle behind this strategy is to determine the structure of clusters which assist in minimizing error criteria the “distance” of each instance to its value of representation. One of the popular criteria is “Sum of Squared Error” (SSE), this estimate square of the total Euclidian distance of instances to its representative values. The SSE method utilizes heuristic approach and to obtain global optimization this comprehensibly enumerates all the available partition which is one of the time consuming job and give us an approximate solution. One simple, typically applied method of squared error criteria is K- means clustering method. This k means clustering method partition or divide the dataset into K clusters such as C_1, C_2, \dots, C_k and this different cluster is represent by center cluster or the mean value of cluster. To determine the center of the cluster the mean of all instances present in the cluster is evaluated.

C) Density-based Methods:

This class of clustering i.e density-based [3] followed the assumption that the data points belongs to the cluster drawn by some probabilistic distribution-“Banfield and Raftery”-- 1993. Thus, the entire distribution of data is result of the various distributions. The basic objective of this technique is to determine the cluster and its distribution function. This density based methods are suitable to recognize the clusters of any arbitrary shapes, not essentially a convex $x_i, x_j \in C_k$.

Here the concept utilized is to grow as long as the shape of the cluster until the density (which referred as the number of data-points within the cluster) factor in neighborhood reaches beyond the threshold value. Explicitly the neighbor within the defined radius comprises minimum number of data objects. If the local maxima of density are used to identify the different clusters then it is called as “mode seeking”.

The Density based spatial cluster with application of noise popularly recognized as DBSCAN method is suitable to present an arbitrary shape clusters, effectually applied for large size spatial databases. In this method searching of cluster is proceeds by searching in the database, neighborhood of individual object which comprises large number of object than the minimum value. AUTOCLASS another algorithm that has major spread root of it, and cover large number of distribution such as Bernoulli, Guassian, and normal log distribution –“Cheeseman and Stutz”,-- 1996”. Some other popular method of this class includes: SNOB “Wallace and Dowe”,-- 1994) , MCLUST “Farley and Raftery”-- 1998). In jain -1999 proposed method of clustering might also utilize some non-parametric techniques, as example in input object space of multi-dimensional histogram, finding the bins comprising the higher count value.

D) **Grid-based Methods:**

In Grid based method a grid structure is obtained by segmenting the space into finite number of cell; all the executable operation of clustering are performed over this grid structure. The speedier processing of this method is one major advantage of it.

E) **Soft-computing Methods:**

Here we present the effectiveness of different soft computing technique in the processing of clusters.

- **Fuzzy Clustering.** A primitive method of clustering generates the subset of clusters or partition; in this generated partition the instance belong to any one of the cluster based on its evaluated characteristics. Thus the obtained structure of cluster is hard cluster and isolated from one another. The method of fuzzy clustering (presented Hopper-2005) broadens this idea and proposed the schema of soft clustering. In this class of clustering each pattern associated with the cluster have some membership function associated with it. Thus cluster comprises as the fuzzy set of pattern. If the value of associated membership is greater than pattern exhibit higher confidence to cluster. By utilizing the threshold value of the membership function then the hard clustering is obtained from fuzzy partition. One well known algorithm of this class is the fuzzy c-means (FCM). However this algorithm proved effective then the hard clustering method in eliminating local minima, but for square error they converge to point of local minima. The sensitive issue that taken into consideration here is the design of the membership function; various parameter for selecting includes here are depending on similarity measure and centriods of cluster. Thus Fuzzy c-means algorithm is presented in generalized manner using different objective functions. Another algorithm that proposed for recognizing the circular and elliptical boundaries is fuzzy c-shell technique.
- **A developmental approach to clustering:** For finding the solution of optimization problem, this technique proves a hypothetical commonly used method. Since clustering problem can be defined as an optimization problem, so the proper way is to use the developmental approach. Here the utilized concept is evolutionary operator, and the obtainable population of cluster is integrated to have an optimal structure. Chromosomes referred as the encoded candidate clusters. And some well known operators used for development are selection, crossover, and mutation. The fitness function evaluated on a chromosome is an indication of the survival of the chromosome in the next obtainable generation. The Genetic Algorithm is one of the popular methods used as an evolutionary approach in clustering technique. All the obtainable structure of clusters is associated with the fitness value. This parameter is an indication of the cluster structure in a way the structure which comprises higher value of fitness is better cluster structure. However the fitness function has inverse relationship with square error value. Thus structure which associated with small value of square error in turn comprises the higher value of fitness function.
- **Simulated Annealing:** This is another hypothetical search technique commonly used in the clustering; the useful of this technique is lies in the property of avoiding the point of local minima. For attaining this, it required to have new solution of some probability for latter iterations of degrading quality (evaluated by some function of criteria). This probability of new solution is controlled by a recognized parameters such as temperature, which commonly specified in terms of a starting (first iteration) and final temperature value. The impact of controllable parameter on the obtainable performance of the algorithm is analyzed by Selim and Al-Sulten- in 1991. This method is statistically guaranteed to find the global optimal solution.

DIGITAL FORENSIC INVESTIGATION

Due to this wide sources of data the use of forensic science is not only limited to investigation purpose but it spread the root of usage for example violation of implicit rule, rebuilding the security event of any system, easily troubleshoot the operational interruption, and enable to have a recovery from any accidental breakup or damage. Practically speaking individual organization has the ability to perform the forensic investigation. The absence of such ability lack the organizations to identify what event actually take place within the system or network, for example disclosure of sensitive or critical data. This work illustrates the usefulness of digital forensic along with the procedure of its and its useful in adopting in different condition. The different rule and regulation of other organization is such that the digital forensic assist in providing security of information within it.

DIGITAL FORENSIC

The term “data” is simply represents as the digital piece of information which can be processed in number of ways. The expansion of computers for professional and personal use and the pervasiveness of networking have fueled the need for tools that can record and analyze the rapid rise of data from large number of sources. As instance data can be preserved or transmitted by using computer systems (e.g., desktops, laptops, servers), networking equipment (e.g., firewalls, routers), computing peripherals (i.e., printers), personal digital assistants (PDA), CDs, DVDs, removable hard drives, backup tapes, flash memory, thumb drives, and jump drives. Many consumer electronic devices (e.g., cell phones, video game consoles, digital audio players, and digital video recorders) can also be used to store data. With such a huge source of data aggregation provoke the requirement of fined tune forensic tools and methods. This has also been caused by the realization these tools and methods of forensic has wide range of applications, for example investigating crimes, violation of implicit rule, rebuilding the security event of any system, easily troubleshoot the operational interruption, and enable to have a recovery from any accidental breakup or damage.

A frequently cited definition for Digital Forensic Science is that of the Digital Forensic Research Workshop (DFRWS) of 2001.

‘The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations (DFRWS, 2001)” [4]

RELATED WORK

Sarac, E and Ozel,S.A[5]: In this paper authors present firefly Algorithm brought into existence by Xin-She Yang in 2007-2008 at Cambridge University, this method has coined its idea from the nature and behavior of fireflies. The assumption introduced in this method are:

As fireflies are unisexual so they an attraction to each other despite of their sex behavior.

This attractive feature of fireflies is directly dependable on their brightness and these both factor have impact of distance on it with an increase in the distance they decrease. Since if there are two fireflies with different brightness, then the less brighter fireflies will move closer to more brighter, in case if brightness are equal then there movement are random.

We evaluate the brightness by the measure of the objective function, and we define a the attraction measure β with a distance factor r as

$$\beta = \beta_0 e^{-\gamma r^2}$$

The motion of firefly I which get attracted by the brighter firefly j is given by the term as follows:

$$X_i^{t+1} = X_i^t + \beta_0 e^{-\gamma r^2} (X_j^t - X_i^t) + \alpha_t \epsilon_i^t$$

Another factor indicates an attraction. The third factor is randomization α_t being the randomization parameter, and ϵ_i^t indicate a random number vector evaluate by a Gaussian or uniform dispersion at an interval of time t . In case if $\beta_0 = 0$, it considers a usual tour. On other part, if $\gamma = 0$, it curtails to an alternative of particle swarm optimization.

Aliakbary, S. Abolhassani, H. , Rahmani, H. and Nobakht, B.[6]: Social tagging is a process in which many users add metadata to a shared content. Through the past few years, the popularity of social tagging has grown on the Web. In this method we investigated the use of social tags for Web page classification: adding new Web pages to an existing Web directory. A Web directory is a general human-edited directory of Web pages. It classifies a collection of pages into a wide range of hierarchical categories. The problem with manual construction and maintenance of Web directories is the significant need of time and effort by human experts. Our proposed method is based on applying different automatic approaches of using social tags for extending Web directories with new URLs [7].

Gowri. R and Lavanya, R.[7]: In this paper author describe a brief survey about the existing approaches in web services composition. The main research areas in web services are related to discovery, security, and composition. Among all these areas, web

services composition turns out to be a challenging one, because within the service-oriented computing domain, Web service composition is an effective recognition to satisfy the hastily changing requirements of business. Therefore, the Web service composition has unfolded broadly in the research side. However, the current attempts to classify Web service composition is not appropriate to the objectives. This paper proposes a novel classification matrix for Web service composition that distinguishes between the context and technology dimension. The context dimension is aimed at analyzing the QoS influence on the effort of Web service composition, while the technology dimension focuses on the technique influence on the effort. Finally, this paper provides a suggestion to improve the quality of service selection those participates in the composition process with Cskylene approach using agents [6-8].

Jinbeom Kang and Joongmin Choi [8]: This paper present various analysis are actively study on mining web data from the numerous available data on WWW. As the web page are not fully structures so it become difficult to determine from it's the informative block, methods which provide the useful data extraction from the useless data such as advertisements is more important. Commonly a web page has many different blocks in which it include data and structural facts. In this proposed method we introduce a web page classification in form of blocks by constructing a Tree Alignment model that indicate the HTML feature and a vector model that represents a feature of blocks. The different websites have their personal templates and blocks may be linked by category nevertheless they are placed at similar position in the web browser or have a similarity in their structure. Thus, by constructing the single classifier it becomes difficult to classify a block accurately. To overcome this problem in our proposed method we use the multiple classifier one for each training data set and classification method succeed by combining all of them. [7-9]

Kovacevic, M. Diligenti, M. Gori and M. Milutinovic, V. [9]: In this paper author have studied that with the fast development of Internet, the Web has become the largest information source for people. Searching information through Internet becomes a more and more popular activity. However, useful information is often accompanied by a large amount of noises. Almost all web pages on the Internet contain noises irrelevant to the main content, such as navigation bar, copyright information, survey or feedback questionnaire etc. These noises affect the efficiency of algorithms for web page classification, clustering, information extraction and searching although they could be useful for other purposes, such as to ease browsing the web pages. It is important to distinguish the informative blocks from the noisy blocks. In this method, we use block to denote the semantic part of the web page, and informative block to denote the main content. Web page information extraction aims to identify and extract relevant data from web pages and put the extracted data into some particular format so that the extracted information can be easily used by some software applications such as Search Engine.

PROCESSING FORENSIC METHODOLOGY

One major aim of forensic is to develop a better understanding of event by recognize and analyses the facts which are related to that event. As known, forensics may be needed in many different situations, such as evidence collection for legal proceedings and internal disciplinary actions, and dealing with malevolent incidents and problem of abnormal operation. Regardless of the need, forensics should be performed using the four-phase process shown in Figure 2. The exact details of these steps may vary based on the specific need for forensics; the organization's policies, guidelines, and procedures should indicate any variations from the standard procedure. This section describes the basic phases of the forensic process: collection, examination, analysis, and reporting [10].

As already explain in section 6. 6 the details of individual phases of forensic process a brief description is taken here: In collection phase the data is recognized, label, and aggregated from all possible sources such as digital file of different category, operating system, and traffic over network and other, while retaining its integrity. In examination phase the tools and the procedure which are well suited for the data are adopted to retrieve the information of particular interest thus retaining the integrity of data. This examination phase uses the combination of automated or the manual tool for retrieval. This phase analyzes the ramification obtained from the examination phase by using the justifiable methods, how impetus the derived information to answer the question is depends on the collection and examination phase. Last phase is report the outcome of analysis such as illustrate the used action, describe the usage of tools and procedure adopted, evaluate the other action requires to be executed (e.g., examination of other available data source, protect the recognized vulnerabilities, enhance the existing security measure), and confront the recommendation should be adopted to enhance the tools, procedure, rule of forensic process. Only after this forensic process the data obtained is transformed into evidence. As shown at the bottom of Figure 1, the digital forensic is useful methodology that clearly transformed data into a valid evidence or proof, this valid proof can even be used to present in court or can be used for solving internal conflict .



Figure 1: Digital Forensic Processes

Specifically, the initial transformation take place when aggregated data go through the examination phase and retrieve data from media and changed into a format that is easily processed by the tools of forensic. And latter, transformation occurs when retrieved data go through the analysis phase. Thus, by having both transformations the data is converted into a useful proof or evidence, as similar to the process of transforming knowledge to action—utilizing the result produced by the analysis in some way to last phase of forensic process i.e reporting. For example, it could be used as evidence to help prosecute specific individual, actionable information to help stop or mitigate some activity or knowledge in the generation of new leads for a case.

AGGREGATED COMMON DATA BASED GROUP IDENTIFICATION BASED CLULSTERING FOR FORENSIC DIGITAL INVESTIGATION (ACD-GIC-FDI)

In this session, it provide the details and description of proposed method. This illustrate the use of this proposed work (acd-gic-fdi) and how it works for the datasets. To maintain the forensic digital investigation through classification result while reducing the memory required is the goal of this work. The methodology for evaluating the above objective of the proposed technique is also discussed in this session. The proposed method of classification is discuss in algorithm mentioned bellow (fig 2).

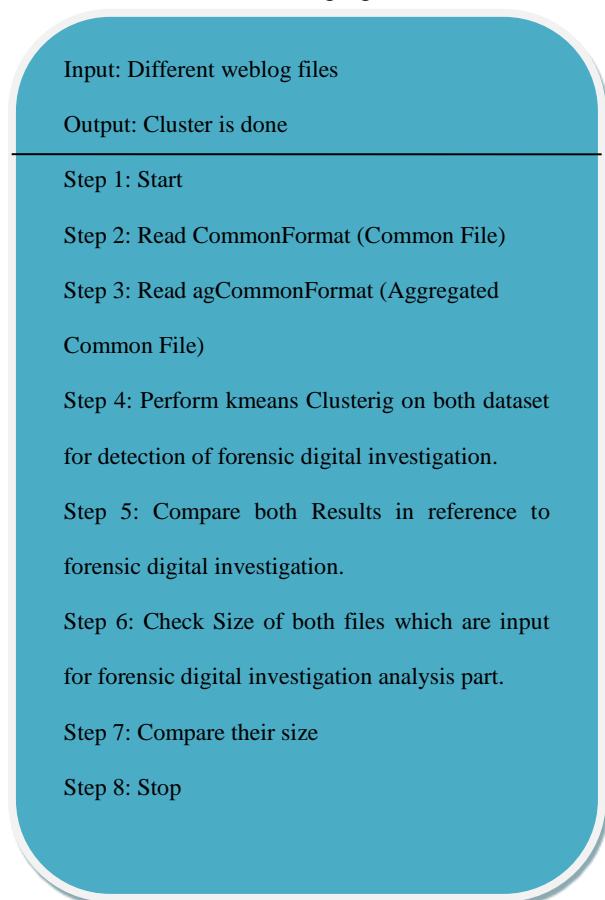


Fig 2: Proposed k-means based Forensic Detection Method

RESULT AND ANALYSIS

Model	Pentium i3 CPU
RAM	4 GB
64 Bit Operating System	
Windows 8	

TABLE 1: System Configuration

The system used for execution of the proposed method for forensic digital investigation by classification is as follows:

In order to evaluate the classification and its efficiency of the various data records through this propose method, two web log datasets are taken into consideration.

TABLE I. NASA weblog file[11]

TABLE II. comdotzone weblog file [12]

The input common XML file contains total 90860 records which gets classified by proposed method. The structure of the CommonFormat file is show in figure 3.

This XML file does not appear to have any style information associated with it. The document tree is shown below:

```

<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<logInfo>
  <info>
    <date>01/Jul/1995</date>
    <time>00:00:01</time>
    <ip>199.72.81.55</ip>
    <url>none</url>
  </info>
  <info>
    <date>01/Jul/1995</date>
    <time>00:00:05</time>
    <ip>none</ip>
    <url>unicomp.unicomp.net</url>
  </info>
  <info>
    <date>01/Jul/1995</date>
    <time>00:00:09</time>
    <ip>199.128.110.21</ip>
    <url>none</url>
  </info>
  <info>
    <date>01/Jul/1995</date>
    <time>00:00:11</time>
    <ip>none</ip>
  </info>
</logInfo>
    
```

Figure 3 show the structure of input record set ‘CommonFormat ‘.

The output file ‘agCommonFormat’ contains total only 3 records after proposed method. Figure 4 show the structure with records of output record set. Here the aggregated file ‘agCommonFormat’ by Date (expanded nodes) is shown.

This XML file does not appear to have any style information associated with it. The document tree is shown below:

```

<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<date date="01/Jul/1995">
  <ip time="00:00:01">199.72.81.55</ip>
  <ip time="00:00:05">none</ip>
  <ip time="00:00:09">199.128.110.21</ip>
  <ip time="00:00:11">none</ip>
  <ip time="00:00:11">199.128.110.21</ip>
  <ip time="00:00:12">none</ip>
  <ip time="00:00:13">205.212.115.186</ip>
  <ip time="00:00:13">none</ip>
  <ip time="00:00:13">129.04.144.152</ip>
  <ip time="00:00:14">none</ip>
  <ip time="00:00:14">none</ip>
  <ip time="00:00:14">none</ip>
  <ip time="00:00:15">none</ip>
  <ip time="00:00:15">none</ip>
  <ip time="00:00:15">none</ip>
  <ip time="00:00:17">129.04.144.152</ip>
  <ip time="00:00:17">199.128.110.21</ip>
  <ip time="00:00:18">none</ip>
  <ip time="00:00:18">none</ip>
  <ip time="00:00:18">none</ip>
  <ip time="00:00:18">none</ip>
  <ip time="00:00:24">205.189.154.54</ip>
  <ip time="00:00:25">none</ip>
  <ip time="00:00:27">none</ip>
  <ip time="00:00:28">205.189.154.54</ip>
</date>
    
```

Figure 4 show the structure with records of output record set ‘agCommonFormat’ aggregated file by Date (expanded nodes). Aggregated file ‘agCommonFormat’ by Date (collapsed nodes) – having 3 lines of 3 dates and it is shown in figure 5.

This XML file does not appear to have any style information associated with it. The document tree is shown below:

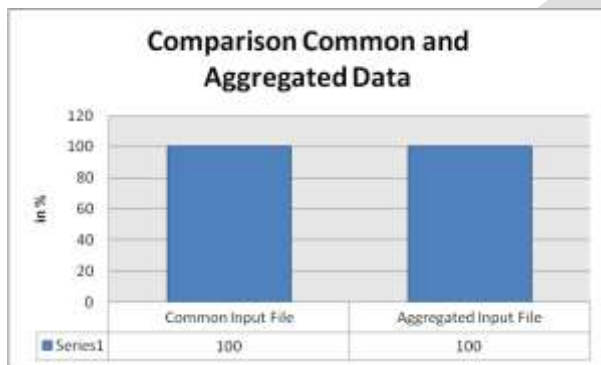
```

    <?xml version="1.0" encoding="UTF-8" ?>
    <Info>
    <Date date="01/Jul/1995" ...></Date>
    <Date date="01/Jul/1995" ...></Date>
    <Date date="01/Sept/2013" ...></Date>
    </Info>
    
```

Figure 5: Aggregated file ‘agCommonFormat’ by Date (collapsed nodes)

Data Size	Common Input File	Aggregated Input File
Full (46040 Records)	100	100

TABLE 2: Forensic Digital Investigation by Clustering method, results are based on ‘agCommonFormat’ and ‘agCommonFormat’ files.



Graph 1: Show the comparative classification results before and after aggregation method on common file.

CONCLUSION

This work investigates that forensic digital investigation by the means of clustering. Forensic investigation is responsible for finding or detecting the illegal or abnormal activities. This works is dedicated to find forensic activities by the k-means algorithm specifically. Session 8 of the paper clearly show the efficiency of the proposed work. Even after applying the forensic detection of aggregated common format data, results would not get changed. Graph 1 and Table 2 show this thing very clearly. The whole implementation is done in MATLAB simulation tool.

REFERENCES:

- [1] Anil Jain and Richard Dubes, “Algorithm for clustering method”, library of Congress Cataloging in publication.
- [2] Jerzy Stefanowski, “Data Mining – Clustering”, Institute of Computing Sciences Poznan University of Technology Poznan, Poland Lecture 7 SE Master Course 2008/2009.
- [3] T. Soni Madhulatha, “AN OVERVIEW ON CLUSTERING METHODS”, IOSR Journal of Engineering Apr. 2012.
- [4] Jeremy Leighton John, “Digital Forensic and Preservation”, Digital Preservation Coalition 2012 and Jeremy Leighton John 2012.
- [5] Esra Saraç and Selma Ayşe Özel, "Web Page Classification Using Firefly Optimization", IEEE Innovations in Intelligent Systems and Applications (INISTA) 19-21 June 2013).

- [6] Sadegh Aliakbary, Hassan Abolhassani, Hossein Rahmani and Behrooz Nobakh,"Web Page Classification Using Social Tags",IEEE International Conference on Computational Science and Engineering Volume4 29-31 Aug. 2009.
- [7] Gowri, R.and Lavanya, R.,"A novel classification of web service composition and optimization approach using skyline algorithm integrated with agents", IEEE Computational Intelligence and Computing Research (ICCIC),26-28 Dec. 2013)
- [8] Jinbeom Kang and Joongmin Choi,"Block Classification of a Web Page by Using a Combination of Multiple Classifiers", IEEE Networked Computing and Advanced Information Management Volume2,2-4 Sept.2008.
- [9] Milos Kovacevic, Michelangelo Diligenti, Marco Gori2 and Veljko Milutinovic,"Recognition of Common Areas in a Web Page Using Visual Information: a possible in a page classification," application in a page classification ", IEEE Data Mining, 2002. ICDM 2003. Proceedings 2002).
- [10] Amarpreet S. Arora, Susheel Ch. Bhatt, Anamika Pant, "Forensics Computing-Technology to Combat Cybercrime", International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 7, July 2012.
- [11] <http://ita.ee.lbl.gov/html/contrib/NASA-HTTP.html>.
- [12] www.comdotzone.com