International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 11 • Issue 1 pp. 14-21 Sept 2019 - March 2020 www.csjournals.com

Analysis of The Plaint & FIR Support System using Data Mining

Kanchan. K. Binnar, Prof. Santosh Kumar
PG Student, Assistant Professor
Dept of Computer Engg,SITRC, Savitribai Phule ,Pune University,Maharashtra

Abstract: In todays big data era,there is a tremendously wide range of data available.All criminal and justice systems generate a huge amount of data every day.There is a need to organize this data and manage it effectively.Weather to register an employee or to report a theft,going to police station in India is usually an administrative or deal.We define our framework for enabling the smart government vision particularly for the case of criminal justice systems by unifying distinct isolated ICT-based solution and proposed a new algorithm for better analysis of crime data.Text mining is the technique that helps users find useful and important information from a large number of digital text data.

In proposed system we have introduced new clustering algorithm to organize and retrieve the useful information related to criminal and justice systems.The system will generate a digital chargesheet and digital FIR directly by analyzing the complaint/the plaint data given by the victim and perform text mining on the complaint data to display automatically IPC acts and article in the chargesheet and FIR to avoid corruption.Clustering algorithms will help to extract hidden patterns to identify groups and their similarities.This system can be referred by new officers to study how the different cases had solved in history by another officer and also gives a brief history of any criminal in a fraction of second.No need to maintain files or folders for criminal and case histories,everything is digital.

Keywords: Data Mining,Text Mining,Big Data Analytics,Modified K Means,SVM,ICT.

Introduction

In the Big Data era we all in one form or another take part in generating data.Big Data generated can be in the form of structured data,which is generated by different applications and typically stored in columns and rows with well defined schemas.This data can be semi structured,which is generated by event monitors,web feeds,sensors.Semi-structured data generally have meta-data that describes their structure;however this structure does not always exist in rows and columns.Sometimes this data also can be unstructured,typically generated by person in forms such as videos,text documents,email,social media,audio \& images.Along with having a many different data formats Big Data is generated in large volumes at a rapid velocity with no obvious way of defining the variability of the collected data.The data generated from criminal logistics \& justice systems and police stations can also be unstructured data.

The NCRB(National Crime Records Bureau)home affairs ministry of India collects \& stores criminal data & publish reports of crime statistics documents[2].The crime data could be analyzed to find the emerging crime trends at high quality both nationally and Locally .The crime researchers,police department,judicial officials,criminologists in India use NCRBs vast statistical data to analyze & help curbing the crime.To execute crime analysis from huge criminal data,We needs to choose an appropriate scientific field.Data mining introduces drilling or deriving knowledge from historical huge databases.Crime analysis is an attractive area where data mining shows a major role with regards to investigation and forecasting.But the challenges in analysing the crime profiles and policing strategies is becoming more difficult as the crime rate is rising day by day.In this research we use data mining strategies on huge criminal datasets and knowledge gained is valuable and helps police department[2].This analysis helps police to ensure the safety of communities and control the crime.

In India NCRB maintains the crime related data and statistical publications software for analysis of crime info.The activities of criminals have increased in past few decades and the crime rate has also expanded because of good communication systems and transport.Crimes cause terror and also damage our community in several ways.The crime trends rises due to fast developmental activities and increasing population in cities and towns.In India the regional location has a strong impact on criminal activity.The CrimeInfo report of NCRB,India collects,publish and analyze the crime data.The crime zoning and profiling can be modeled with utilization of data mining.In the recent years there's a sharp increase in crime statistics in India.As per the information of the NCRB,cognizable crimes under the IPC have increased from 18,78,291 to 29,49,400 and cognizable crimes under the Local & Special Laws have increased from 32,24,165 to 43,76,697.The data generated from police stations is yet another big data and generated in a large amount which is unstructured.There is a need to store this data in digital format,so as to decrease the time required for writing complaints,making chargesheets etc.but

unfortunately there's no systematic method available to manage and retrieve this data.Our proposed system provided a central database system to connect all police stations and will efficiently manage all of the data used and generated from police stations.so as to maintain the data for further references and also reducing the time to register complaints and make the chargesheets.

Complaints usually contains a subject or more such as the street lights broken,heating are not hot,dirty and bad environment and traffic congestion and so on.Using inner features and external features of complaints subject words we proposed a method based on clustering to extract subject words of urban complaint data[2].

Our proposed system links this whole system together digitaly for better management of crime data using newly proposed clustering algorithm by overcoming the limitations of K means and modified k means clustering algorithm.It also used to handle all the works in police station from registering a complaint to generating a chargesheet and FIR in a digital way,providing a central database to connect all police stations and provide access rights as per the hierarchy.This system will efficiently manage all the data in police stations related to various cases e.g family matters,accident etc.and save it in a central database for further references.The another purpose of this system is to generate a chargesheet and FIR directly by analyzing the complaint data given by the victim and perform text mining on the complaint data using modified K means algorithm to display automatically article in the digital chargesheet so as to avoid corruption.This system can be reffered by new officers to study how the different cases had solved in history by another officers and also provides a brief history of any criminal in a fraction of second.No need to maintain files or folders for criminal and case histories.everything is digital.There exists no of systems for registration of police complaints which also providing information of different police stations.but there exists no system which automatically selects IPC acts and articles by analysis of complaint data and generates digital FIR.

In recent years techniques of text mining and data mining have been frequently used for analyzing questionnaire and review data.Text mining is placed as the important term in data mining & text mining extracts quality information.Text mining extracts undisclosed data from semistructured and unstructured data.It is the invention by automatically extracting knowledge & information from different written resources.Our proposed system uses text mining method to decide the article on the basis of complaint registered by the victim and use this article in the automatic chargesheet generation.Here we are using proposed k means and SVM(Support Vector Machine) algorithms for analysis of complaint data using Data mining/Text mining.

In this system,a modified algorithm is proposed for retrieving the crime info from centralized database with more usable and more informative format.The data point has been assigned to the suitable class or cluster more remarkably.The proposed algorithm decreases the complexity of the numerical computation.Here we present how to optimize the existing k mean algorithm by applying the modified approach algorithm to the criminal datasets.

K-means is a widely known clustering partitioning approach.K-mean is very useful for applications in which a number of clusters are required,similarly the k-means clustering technique merge the data together taking into account their closeness.Mainly,k-means is an iterative algorithm and performs two mechanisms:first the cluster assignment step and second the group centroid step which also called the move centroid step.The modified algorithm approach is proposed to decrease the complexity of the numerical calculation of existing algorithm.

Literature Survey

In recent years relatively few studies were conceren with complaints data mining for urban areas\& existing work focuses on the automatic identification of Chinese geographical entities \& the classification.LiXue Weietal.[6]proposed a model for recognition of Chinese geographical entities by using divide and conquer strategy to identify geographical entities into recognition of basic location names and construct indicator words.Zhian Dong \& Xueqiang Lv have introduced a system in which through the analysis of the complaint data,they found the title,which may be a good response to the subject of the complaint.So they had extracted the subject words from the title(heading)of complaint document[1].This system first segments the title of complaint data using stop words and extracts candidate subject words and then filter them using overlapping word features,position features and dependency features to get subject words.

Manish Gupta et.al[7] have studied existing system in use by Indian police and introduced a criminal analysis tool,Which is based on interactive queries.The tool helps police departments to control crimes.Rajan Vohra and Priyanka Gera [8]looks at the use of cluster technique in data mining for analysis of the crime patterns.The authors use k means clustering algorithm to help in the process of criminal profiling to primary crime data from Delhi police first information report(FIR)records.The study is useful to identify which type of crime is predominant and find which area categories are most sensitive and show the distribution of each crime type in every area category.

There are however several similar open government initiatives that aim at publishing of public sector information.In recent years initiatives for open criminal and justice have been introduced to increase the accountability and transparency of criminal logistics and justice systems through releasing judicial public sector

data to citizens,enterprises and/or experts.In other areas of the public sector compared to their counterparts,open criminal justice initiatives have been reported less frequently in literature.The CrimeInfo ver 1.0 is a database software developed by NCRB[2],India with the cooperation of the UN system.The crime dataset is developed by applying complexed queries on the CrimeInfo,India database.The dataset derived from NCRB does not consist any class label.It is better to use cluster method to make groups of sets of the raw data.Clusterization is a concept of data mining,groups objects into a category based on same attributes.WEKA software is a utility to analyse the historical dataset with a collection of data mining and machine learning algorithms.The crime dataset is input to the WEKA software to construct cluster zones using K means cluster method.Depending on characteristics K means algorithm groups objects.The groups are created by minimizing sum of squares of distances between data and concerned groups centroid.The cluster output of WEKA software is manually passed as input to MyCustom map[2],an online interactive map tool of Indias maps to create custom India map with the cluster zones of states.The crime datasets developed by applying complex queries on the Crime Info India database.The dataset contain state wise crimes made by female,male and total for the year in India.

Framework for legal logistics shown in fig 1 was proposed by Niels Netten[3].In his framework he introduces framework for allowing the smart government vision by unifying distinct isolated ICT-based solutions,specificaly for criminal logistics systems.framework named as Legal Logistics supports much better working of a legal systems so as to streamline the innovations in these legal systems.The framework targets the handling of all important data which is generated from the ICT-based solutions.For the Dutch criminal justice system.

System Architecture / System Overview

Analysis of The Plaint and FIR Support System Using Data Mining consist of architecture as shown in fig 3.Whenever information is entered about crimes and complaints,it is stored directly into the central dataset after preprocessing.This central dataset also contains the collected information from NCRB.If any user wants to access a particular information,he can access it easily from centralized data set.Data is provided to the user by using proposed algorithm in more efficient way.

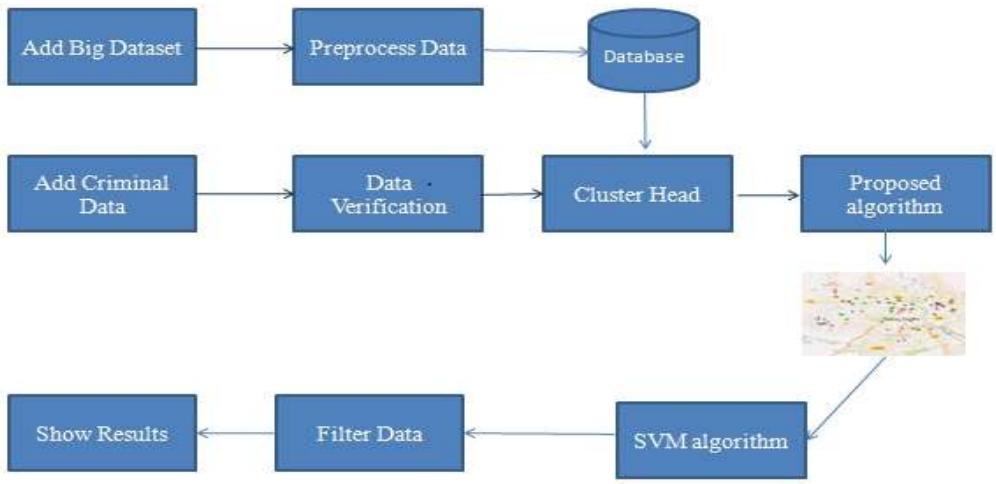

Fig.1.System Architecture

Using this algorithm system can easily filter data needed to the user from large amount of data inside the dataset and finally results sent to the user.While storing the data,it can be stored in a random manner,but while retrieving the data system needs to recognize the requested data from huge datasets.hence we are using modified algorithm to filter the data,classify it and store it in a systematic way.eg.an person is handling a case related to child kidnapping and he is interested in accessing the data which is related to solved kidnapping cases only for his reference.Then using this system the person can get the detailed information regarding the similar type of kidnapping cases for his study,simply by entering the type of information he required.Modified proposed algorithm used specificaly to reduce the time required for the analysis and retrieval of data.Proposed algorithm will overcome the limitations of K means and modifeid K means clustering algorithms and will provide better performance as compared to them.

The existing systems for criminal pattern identification had used basik k means algorithm for filtering of data,which has some limitations like complexity in numerical calculatins.Such obstacls of K means algorithm are removed by using proposed algorithm.In many fields of machine learning K means algorithm has been broadly

used.It does not require more computation.However when applying K-means in the large dataset there comes some obstacle and limitations.To reduce the complexity of the numerical computation while applying the k-mean algorithm,The modified algorithm approach is designed.The modified algorithm provides an alternative way of computing and calculating the frequency of each data point in segments and dividing the entire space into more than a few segments.To calculate the frequency of the input vector in every part and to pick the highest k frequency section partition of the space is performed.

Analysis of The Plaint and FIR Support System using Data Mining consist of centralized database.When a complaint comes,it is directly entered into the system in the police station by police officer into the form provided by the software for complaint registration.then text mining or pattern matching is applied to the data of complaint by using SVM algorithm and the related IPC criminal act and article is selected automatically and as per the information FIR is generated,which is digital FIR.Also when higher authority wants to manage the lower level police departments.They have access to all the data and has a separate login.Similarly whenever victim who has registered a complaint in police station wants to know the current status of his case,he/she can check it online using our system.Proposed system manages all the data in police stations related to various cases e.g family matters,accident,kidnapping etc and save it in a central database for further references.

The another purpose is to generate a chargesheet and FIR directly by analyzing the complaint.data given by the victim and system performs text mining on that data to display automatically article in the chargesheet so as to avoid corruption.This task may include the language translation too,if text is entered in Marathi or Hindi language.Language translation is also done automatically by system whenever needed.This system can be reffered by new officers to study how the different cases had solved in history by another officers and also provides a brief history of any criminal or case in a fraction of second.No need to maintain files or folders for criminal and case histories.everything is digital.Text mining is the exploration of interesting knowledge in text documents.

Proposed Algorithm

Input:Dataset of N points,Desired number of k clusters

Output: N points grouped into k clusters

Phase1: Finding Initial centroids}

- Input the dataset and value of k$>=$2.
- the data point set into k$*$k segments /*k vertically and k horizontally*/
- For each dimension
- Calculate the maximum and minimum value of data points.
- Calculate the width $(Rg) rg= (min+max)/k$
- reach $(n/r*f)$
- Calculate the frequency of data points for each segment.
- Choose k highest frequency segments.
- For each segment i = 1 to k
- For each point j in the segment i
- Calculate the distance of point j with origin
- Sort these distances in matrix D in ascending order.
- Select the middle point distance.
- The co-ordinates corresponding to the distance in 12 is selected as initial center for the ith cluster.
- In matrix C these k co-ordinates are stored which represents the initial centroids.

Phase2: Assigning points to the cluster
Repeat
For each data point p = 1 to N
For each cluster j = 1 to k

- Calculate distance between point p and cluster centroid cj of Cj
- Assign p to min{d(p,cj)}where j=[1,k].

- Check the termination condition of the algorithm if Satisfied

Exit
Else

- Calculate the new centroids of Go to step 1.

For each point j in the segment i
{
Calculate the distance of point j with origin
}

Phase3: Clusters Generation

The appropriate number of clusters is then found using following steps.

- If $Sk-2 < Sk$ and $Sk > Sk+2$ then run phase 1 and phase 2 using $k+1$ and $k-1$ and corresponding $Sk+1$ and $Sk-1$ are found. The maximum of the three $Sk-1, Sk, Sk+1$ then determines the value of k as appropriate number of clusters

- Else If $Sk+2 > Sk$ and $Sk+2 > Sk-2$ then run phase 1 and phase 2 using $k+1, k+3, k+4$ and corresponding $Sk+1, Sk+3$ and $Sk+4$ are found. The k values corresponding to maximum of the $Sk+1, Sk+2, Sk+3, Sk+4$ is returned.

- Else If $Sk+2 < Sk-2$ and $Sk < Sk-2$ then run phase 1 and phase 2 using $k-1, k-2, k-3, k-4$ and corresponding $Sk-1, Sk-2, Sk-3$ and $Sk-4$ are found.The k values corresponding to maximum of the $Sk-1, Sk-2, Sk-3, Sk-4$ is returned.

Stop
When the best value of k is found,the algorithm terminates itself ,This value of k shows appropriate number of clusters for a given data set.

Performance

Modified approach K-mean is having better performance in comparison to standard K means algorithm[6] and our Proposed algorithm is having much better performance in comparison to standard K means algorithm.This section shows the actual results obtained by Analysis of the plaint and FIR support system using data mining.

Clusterization of data is done by selecting clustering option in main menu,proposed algorithm is used for clustering,give the attributes for clustering like number of clusters etc.Run the clustering by pressing start button.The output file is shown with generated clusters and CrimeInfo dataset.It shows the classification of dataset into no of clusters like crime against society,crime against person,crime against property etc.It shows systematic classification of crime data using our proposed algorithm.This system also provides the application for generating the Digital Chargesheet from the analysis of crime data info.User just have to enter FIR no and all the information related to FIR 1 is displayed automatically in textboxes.After click on generate FIR ditital FIR is get produced and pdf generated.

Graph in fig.2 shows difference between original k-means,modified approach and proposed algorithm.When comparing execution times of proposed algorithm in milliseconds with other algorithms,it is observed that our proposed algorithm takes much less time than the k-Means and modified k means and for the larger datasets such as dataset of 500 points;the proposed algorithm also outperforms the previous algorithms.

International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 11 • Issue 1 pp. 14-21 Sept 2019 - March 2020 www.csjournals.com

Fig.2 Performance of k means, Modified k means and Proposed algorithm

Graph in fig.3 shows the actual performance obtained by proposed system over the existing system using k means clustering algorithm.From this graph we can say that time required for analysis of crime info using proposed algorithm is very less as compared to the time required for existing clustering algorithm.x axis shows the type of algorithm while y axis represents time required to execute the algorithm.

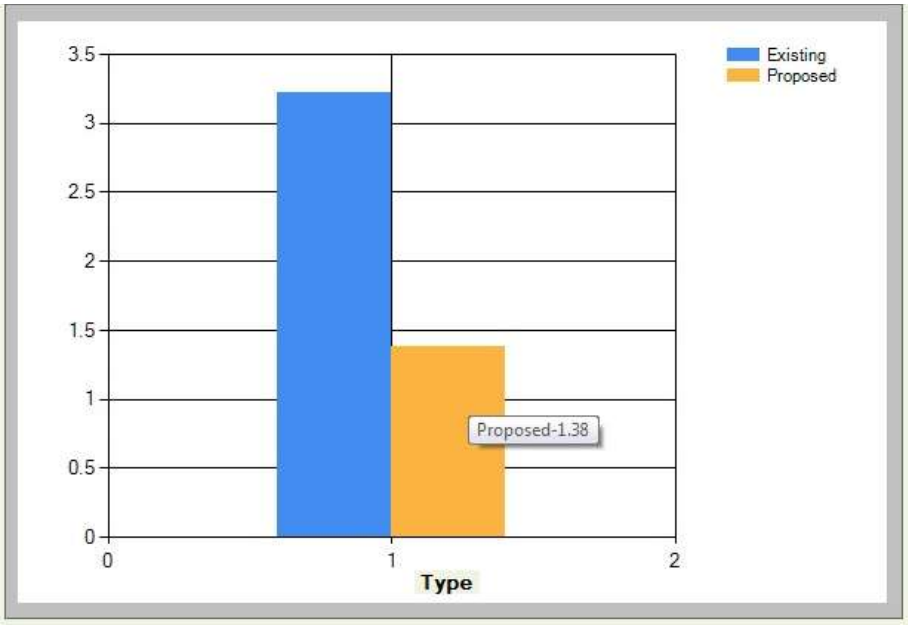

Fig.3 Performance of Existing approach & proposed algorithm

Fig 4 represents graph for the classification of crime data for each year eg.2015,2016,2017 in the form of crime against society,crime against person and others.You have to simply select year and then the type of crime from the dropdown list provided to get the results for perticular year.In our application this graph is generated auotomaticaly by analysing dataset and results of classification processes.

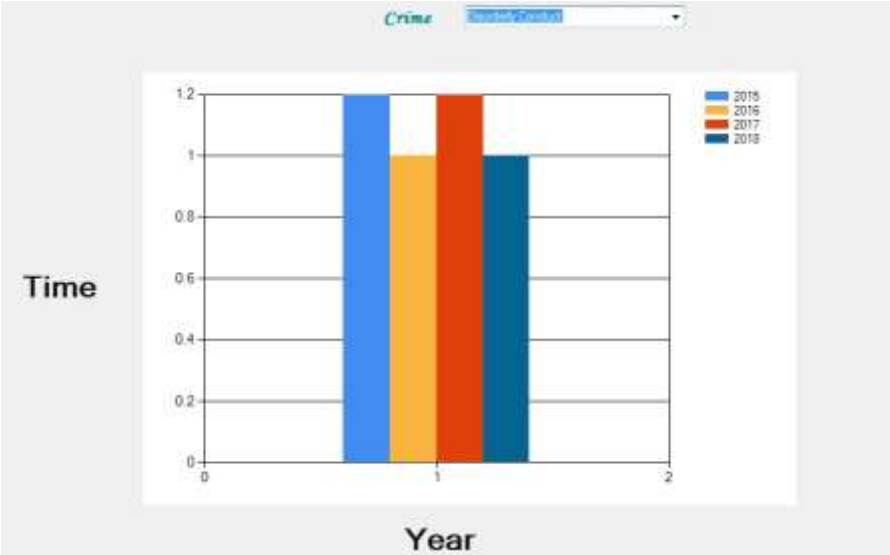

Fig. Graph analysis for classification of crime data per year

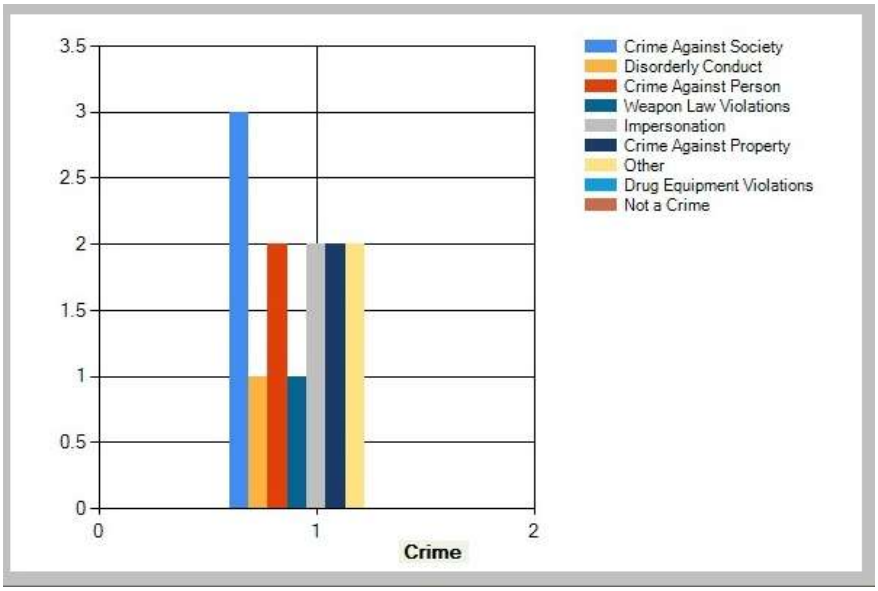

Fig. Bar graph for classification of crime data

Fig 5 shows the classification of crime data using Bar Graph.This Bar graph is generated in our application on the basis of dataset and classification results.In this case crime against society is very large as compared to crime against person and other type of crimes.You can classify crime into no of sub types and generate classification of crimes for better inevestigation and for the help of judicial processes also.This analysis is very helpful for investigating similar type of crimes in future.Using proposed system this results can be generated in very less time as compared to other clustering algorithms.

Conclusion

Criminology is a sensitive area where proficient clustering approaches of data mining plays important role for crime analysts.We have focused on criminal analysis by introducing modified clustering algorithm.We have presented our framework for enabling the smart government vision in this system,particularly for the case of criminal justice systems by unifying distinct isolated ICT-based solution and using modified algorithm.Well-functioning of a legal system is supported by this framework named as Legal Logistics in order to integrate the innovations in these legal systems and also reduces the time required for analysis of data using proposed

algorithm.The system will also generate a chargesheet and FIR directly by analyzing the complaint/the plaint data given by the victim and perform text mining on the complaint data to display automatically IPC acts and article in the chargesheet and FIR to avoid corruption.

Acknowledgement

I would sincerely like to thank our Professor Santosh Kumar,Department of Computer Engineering,SITRC.Nasik for his encouragement ,guidance and the interest shown in this project by timely suggestions in this work.His expert suggestions \& scholarly feedback had greatly enhanced the effectiveness of this work.Thank You.

References

[1]. Suresh Babu Changalasetty,Lalitha Saroja Thota1{"Cluster based Zoning of Crime Info".}978-1-5090-5814-3/17/31.00©2017 IEEE

[2]. Xueqiang Lv,Zhian Dong,Xuewei Li and Xueqiang Lv{'Subject extraction method of urban complaint data'}978-1-5386-3120-1/17 31.00 © 2017 IEEE

[3]. Mortaza S.Bargh,Niels Netten,Susan van den Braak,Sunil Choenni and Frans Leeuw {'On Enabling Smart Government: A Legal Logistics Framework for Future Criminal Justice Systems'}dg.o '16, June 08-10, 2016, Shanghai,China © 2016 ACM.ISBN 978-1-4503-4339-8/16/06.

[4]. Wataru Sunayama,Tomoya Matsumoto,Yuji Hatanaka and Kazunori Ogohara {"Data Analysis Support by Combining Data Mining and Text Mining".}978-0-7695-6178-3/17 31.00 © 2017 IEEE.

[5]. N.Kurinjivendhan\& Dr.K.Thangadurai{'Modified K-Means Algorithm and Genetic Approach for Cluster Optimization'.} University of Computer Studies, Mandalay Myanmar,978-1-5090-5507-4/17/©2017 IEEEICIS 2017.

[6]. Lv Xueqiang,Li Xuewei,Liu Kehui{"Automatic Recognition of Chinese Location Entity[M]".}Natural Language Processing and Chinese Computing379-391 2014

[7]. B.Chandra,Manish Gupta and M.P.Gupta,{"Crime Data Mining for Indian Police Information System".}Computer society of India,2007

[8]. Rajan Vohra\& Priyanka Gera{"City Crime Profiling Using Cluster Analysis".}International Journal of Computer Science and Information Technologies IJCSIT,Vol.5(4),2014

[9]. KazuhikoTsuda,Yoko Kobayashi{"Comparison of K-means and Modified K-mean algorithms.for Large Data-set".}Volume 1,No.3,2012,IJCCN,International Journal of Computing,Communications and Networking.

[10]. Prof.Vinayak Shinde,Akshata Raut{"Effective Methods and Techniques in Text Mining".}International Conference On Emanations in Modern Technology and Engineering (ICEMTE-2017) ISSN: 2321-8169Volume:5 Issue:3.

[11]. Akshata Raut,C.Prof.Vinayak Shinde{"Effective Methods and Techniques in Text Mining.".} International Conference On Emanations in Modern Technology and Engineering (ICEMTE-2017).

[12]. Gabriel Murray,Giuseppe Carenini{"Methods for Mining and Summarizing Text Conversations.".}Department of Computer Science,University of British Columbia Vancouver, B.C., Canada 2012.

[13]. X.He and C.Ding{"K-means clustering via principal component analysis".}ICML 04,pages 29,New York,NY,USA,2004.ACM.

[14]. S.Goel and D.Pandove{"A comprehensive study on clustering approaches for big data mining."}(ICECS),2015 2nd International Conference on,Feb 2015.

[15]. Cheng-Hung Lin and Wei-Chen Liu {"Research On Big Data Management And Analysis Method Of Multiplatform Avionics System".} 2017 IEEE International Conference on Applied System InnovationIEEE-ICASI 2017 Meen,PriorLam(Eds)

[16]. Sunil Choenni,Jan van Dijk,Frans Leeuw{"Analyzing a Complaint Database by Means of Genetic-Based Data Mining Algorithm".}ICAIL-2009 Barcelona,Spain.Copyright 2009 ACM 1-60558-597.

[17]. Yuta Sano,Tsunenori Mine{"Extraction of Current Actual Status and Demand Expressions from Complaint Reports".}iiWAS 16,November 28-30,2016,Singapore 2016 ACM. ISBN 978-1-4503-4807-2/16/11.

[18]. \bibitem[18]{} Ana G.Maguitman,Rocío.B.Hubert,Marcos A.Malamud, {"CitymisVis: a Tool for the Visual Analysis and Exploration of Citizen Requests and Complaints".}ICEGOV '17,March 07-09,2017,New Delhi,AA,India© 2017 ACM.ISBN 978-1-4503-4825\bibitem J.Satheesh Kumar and Revatthy Krishnamurthy{"Survey of data mining