ENFORCING CRIME FACTOR ANALYSIS THROUGH OPTIMIZED MULTI CLASS SVM CLASSIFICATION

Dr.J.Chockalingam
M.Sc., M.B.A., M.Phil., Ph.D.,
Associate Professor/Research Supervisor
Computer Science Department
Khadir Mohideen College,
Adirampattinam-India

B.Venkatesan
M.C.A., M.Phil., (Ph.D.),
Lecturer, Govt. Arts College(A),
Kumbakonam / Research Scholar
Computer Science Department
Khadir Mohideen College,Adirampattinam-India

Abstract: In recent world, the crime factors are increasing day by day. Some of the preventive measures are taken to identify and avoid the crimes by their concern departments. There are huge amount of heterogeneous unstructured data are collected about the crime factors and the criminals. That the gathered unstructured data should be gathered, elicited and classified for analyzing the crime factors and find out the predictive measures for the prevention of crime activities. Many classification techniques are used in the analysis of crime factors and make predictive analysis through it. However the effective techniques are used for the analysis of the crime factors, this is potentially strong based on the dataset taken as an input or in training phase. In existing, for gathering the relevant data for the analysis is obtained through the Hierarchal Clustering. It is most efficient clustering technique but it works well in the structured text data. To obtain the effective data clustering in heterogeneous unstructured data, the K Means Clustering has been proposed. Next to clustering, the classification technique is required to classify the properties of the crime factors through the parameters in unstructured data using Cluster Based Support Vector Machine Classification. The main contribution of this proposed work is to create a novel mechanism for obtain the predictive analysis for prevention of crime factors from the various crime classes. The predictive analysis can be obtained through the continuous process of K means clustering and Multi Class SVM Classification. To prove the effectiveness of the proposed mechanism capable of deals with multiple domains, SVM Classification algorithm can be optimized through this algorithm. The experiments have been illustrating the efficiency of the mechanism with the dataset from UCI repository and data from government of India.

Keywords: K Means Clustering, SVM, Ant Colony Optimization Algorithm

1. INTRODUCTION

Data mining techniques [2] are plays a major role in the analyzing of crime factors to reduce and take counter measures to the crime. Involving with the intention to reduce the crime factors, the data should be analyzed to improve the results. Thus the purpose of taking counter steps to the unpredicted cases, taking counter steps for the crime those having more affects in the society. The important thing is to gather the data factors to obtain the data cluster through the clustering techniques in data mining.

Why the intention required for the enforcing crime factor analysis? The answer for the question is as follows:
1. Crime data are not properly gathered for the analysis.
2. Crime factors should be analyzed for counter measures such as special intentions, creating special operations to avoid or reduce the factors.
3. In the department of police, the various sub departments are there for the various crimes, the crime should not be occurs in that range of department, the police team should be transferred to the place of that increased respective crime.

These are main reason for the analysis about the crime factors. The two main type of analysis are taken place in this research. The analysis based on the crime data to obtain the severity of the crime and number of occurrences of the crime and its causes. The second type of analysis is to obtain the experts from the team of police those who are taking more counters in that case of crimes and taken efforts for the crimes.

In existing, the traditional machine learning algorithms for classification can be done through the Naïve Bayes Classification [9] and Support Vector Machine Classification [2][3]. In general, these classification techniques are done through the clustered data items from the clustering technique in data mining, that the clustering techniques are Hierarchal clustering and K Means clustering algorithm. Thus the improved results can be done through the Cluster Based Support Vector Machine Algorithm. These algorithms provide the efficient classified results for the crime factors.

However they provide the efficient results for the crime factor analysis. The improved optimization is required to fulfill the standard of obtaining analysis for various kinds of crime factors. The crime factors are not divided and separated into classes. Thus the crime factors are clustered into different classes. Then the algorithm should satisfy the parameters for the analysis needed in this multiple classes. The Multi Class Support Vector Machine Algorithm should provide the improved and better performance results through the analysis of crime factors data in multiple domains. Thus the classification can be optimized.

The main contributions of the proposed work for crime factor analysis is

1. To obtain the clusters from the heterogeneous unstructured crime data collected in the various departments
2. To classify the clusters with the requirements using Multi Class Support Vector Machine Classification Algorithm.

II. RELATED WORK

R.Anusha and Dr.N.Rama [4] were proposed new multi kernel hybrid support vector machine data classification algorithm to obtain the results for classify the data. To measure the accuracy in classification of data collected for obtaining the website quality and behavior pattern analysis in online shopping websites from Neural Network feed forward back propagation algorithm.

Himani and Panchal [5] were presented a survey about the Support Vector Machine Algorithm and discuss about the statistical learning problems such as spam filtering, text classification, face and object recognition, handwriting analysis and countless others. They defined the SVM is a robust tool for regression and classification in noisy and complex multiple domains.

Julio Lopez et.al. [6] were proposed an maximum margin approach for multi class support vector machines. To obtain the optimization in the traditional SVM, they proposed the multi class SVM and extend this through the kernel based classification.

H.Liu et.al. [7] were presented an article to analyze the violent crime information systems containing the generic factors. They also combine the essentials of rough set theory and particle swarm optimization with fuzzy logic to obtain an algorithm and methodology to extract the multi domain data knowledge from information systems.

In reference paper [8], the authors were proposed mechanism with Genetic Algorithm for analysis and prediction of crimes by means of clustering and classification. Tahani Almaine et.al. [9] discussed an survey shows how to use the crime factor analysis using Decision tree classifier and Naïve Bayes Classifier in order to predict the potential crime types.

III. BACKGROUND KNOWLEDGE

A. *Support Vector Machine Classification*

Support Vector Machine Classification Algorithm [1] [2] is developed practically by Vapnik and Chervonenkis in 1963.
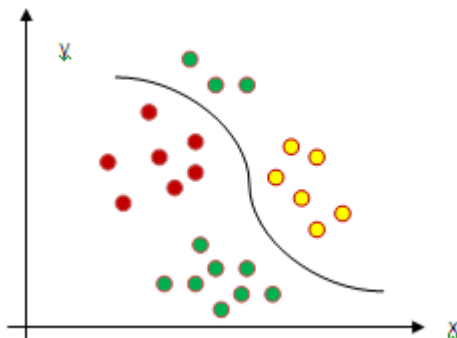

Fig 3.1 Multi Class - SVM Classification

Later 1992, Vapnik, Bernhard and Guyon proposed with the novel suggestion to produce the nonlinear classifiers by submitting the kernel trick to high maximum margin hyperplanes. The next iteration with evolution in 1995 by Cortes and Vapnik called as current standard incarnation i.e., Soft Margin Linear SVM.

The SVM based classification [3] [5] is more descriptive and potential in nature, because of its performance of analysis and its efficiency does not depends directly on the dimensions of entity classified. The SVM is robust, powerful and computationally high cost because it is based on convex quadratic programming. SVM can also be reveals the non-linear decision functions by the projection of heterogeneous input test data into a training data in a high dimensional feature space using kernel functions and also obtains the linear classification problem in that training feature space.

The SVM is majorly divided into two types.
1. Linear SVM classifier
2. Non-Linear SVM classifier.

In linear SVM classifier model, the training data samples simply plotted in the hyperplane space. The apparent gap represents the separation of data points. It can be divided into two classes and obtains a straight hyperplane called as a maximum margin hyperplane.

In Non-Linear SVM classifier, the training data samples are generally detached up to a quantity of point. Here the kernel trick applies to maximum margin hyperplanes for regression analysis. In this most powerful data analysis obtained for multiple domains of data.

B. *Multi Class SVM Classification*

SVM only originally created with the proper binary classification i.e., it have only two domains or classes. It can be complete to multi dimensional because of in need of multiple domains or classes. There are two main approaches were developed for multi class SVM [10] [11]. First is to build and merge all more than a few binary classifiers and another one is considering all data in single optimization. It is mainly used for the optimization of large amount of data.

In this Multi Class SVM, it has the parameters definition for the various domains of data. For example, we take sentiment analysis in data process. If the review data obtained for the domain Hotel Review Mining, it works with the parameters such as cost, sophistication and service but it does not work for the other domain like Product Review Mining, because it works with the parameters such as cost, quality and size. Now the Multi-Class SVM works for the mining for the various domains. It supports as based on the class iin classification.

IV. PROPOSED ARCHITECTURE

The novel mechanism implementation has been done based on the architecture illustrated as figure 4.1. The crime data from the various heterogeneous data sources from the police departments. The multi class of the crime data can be analyzed and collected through the data retrieval process. The collected crime can be preprocessed with the following processes such as data cleaning,
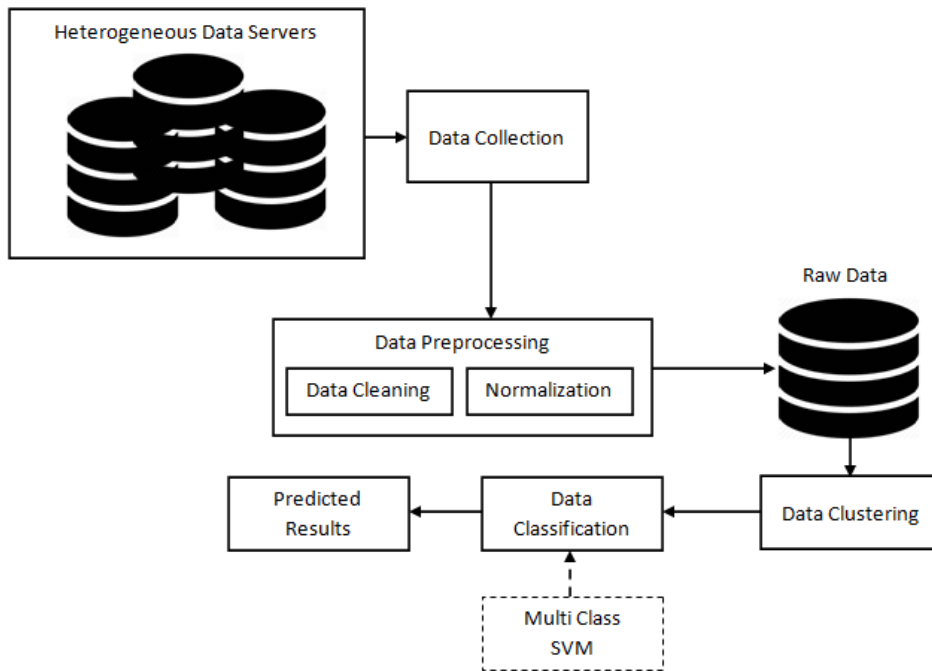
Figure 4.1. Architecture for Proposed Work

data transformation, data scaling and data normalization. The raw data can be obtained from the preprocessing steps. The clustering can be performed through the K-Means Clustering to obtain the data clusters and classify the data results through the Multi Class Support Vector Machine Algorithm to predict the classified results needed.

Thus the algorithm for the clustered Multi Class SVM can be done as follows:

Algorithm : CB – Multi Class SVM Framaework

Input : E={$e_1,e_2,e_3,\ldots,e_n$}entities
        k- number of clusters
Ouput: C={$c_1,c_2,\ldots C_n$} cluster sets
        CL={$cl_1,cl_2,\ldots,cl_n$}
        R={R1,R2,…,Rn}classification set

Algorithm:
 obtain the entities E and no. of clusters
 foreach c_i Є C do
 c_i←e_i Є E
 end
 foreach e_i Є E
 cl(e_i) ← E
 end
 set ch=false; itr=0;
 repeat
 foreach e_i Є E do
 UpdateClusters(c_i)
 end

 foreach e_i Є E do
 minDist ← argminDistance(E);
 if(minDist ≠ cl(E) then
 ch=true;
 end
 end
 until ch=true and itr<k
 while there are interruption points do

Find a candidate
candidateSV=candidateSV ∪ candidate
if any $\alpha_p < 0$ £ additions of S then
candidaeSV = candidataSV / P
repeat until such points pruned
end if
end while

V. IMPLEMENTATION RESULTS:

The implementation based on the above illustration has been done using the dataset UCI Data Repository or real time clustered datasets from the Indian Government from the web data source https://data.gov.in/catalog/crime-india-2015

From the web data source, we collect the data for the all types of crime such as Value of Property Stolen & recovered and percentage recovery during 2015, Disposal of Persons arrested under IPC by Police during 2015 and Disposal of Persons Arrested under Special and Local laws crimes by courts during 2015.
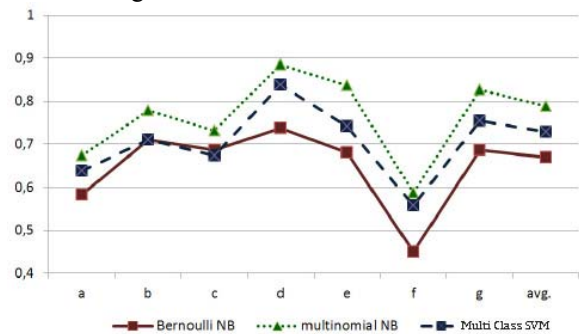

Figure 5.1 Comparison of algorithms

The above comparison results based on the identified results from the Weka Tool displayed as above figure 5.1. The results compared with the traditional algorithms such as Bernouli Naïve Bayes and Multinomial Naïve Bayes with the proposed algorithm Multi Class Support Vector Machine.

VI. CONCLUSION

Thus the cluster based Multi Class Support Vector Machine Classification has been performed for the analysis of crime data factors. In this proposed work, the data can be clustered through the K Means Clustering Algorithm. From the clustered data the multi class domains should be analyzed to make decisions for counter measures in the crimes.

REFERENCES

[1] Han. J, Kamber.J, and Pei.M, "Data Mining: Concepts and Techniques", Elsevier 2nd Edition.

[2] Chih-Wei Hsu, Chih-Chug Chang and Chih-Jeh Lin, "A Practical Guide to Support Vector Machine Classification", Dept. of Computer Sci., National Taiwan Univ., Taipei, 106, Taiwan, http://www.cse.ntu.edu.tw/~cjin 2007.

[3] N.Chirstianini and J.Shawe-Taylor, "An Introduction to Support Vector Machine", Cambridge University, University Press, 2000.

[4] R.Anusha and Dr.N.Rama, "A New Multi Kernel Hybrid Support Vector Machine Data Classification Algorithm", Intl. Journal of Scientific and Research Publications, Vol. 6, No.12, December 2016.

[5] Himani Bhavsar and Mahesh H.Panchal, "A review on Support Vector Machine for Data Classification", Intl. Journal of Advanced Research in Computer Engineering and Technology, Vol.1, No.10, December 2012.

[6] Julio Lopez, Sebastin Maldonado, M Carrasco, "A Novel Multi-Class SVM Model using second – order cone constraints", Journal in Artificial Intelligence, Vol.44, No.2, March 2016.

[7] H.Liu, Chao Yang, Meng Zang and Yeqing Sun, "A Computational Intelligence Approach to Multi-factor Analysis of Violent Crime Information Systems", Article in Enterprise Information systems, Vol.11, No.2, February 2017, pp.161-184.

[8] R Kiani, S Mahdavi and A Keshavarzi, "Analysis and Prediction of Crimes by Clustering and Classification", Intl. Journal of advanced research in Artificial Intelligence, vol.4, no.8, 2015, pp.11-17.

[9] Tahani Almaine, Rsha Mirza and Elizabeth Lor, "Crime Prediction based on Crime Types and using spatial and temporal criminal hotspots", Intl. Journal of Data Mining and Knowledge Process, vol.5, No.4, July 2015.

[10] Hetl Bhavasar and Amit Ganatra, "Variations of Support Vector Machine Classification Technique:A Survey", Int. journal of Advanced Computer Research, Vol.2, No.4, December 2012.

[11] Chih-Wei Hsu and Chih Jen Lin, "A Comparision of Methods for Multi-Class Support Vector Machines", Dept. of Computer Sci., National Taiwan Univ., Taipei, 106, Taiwan.