A Law Enforcement for Crime Detection (RRL-PMD): Relative Record Linkage based Pattern Mining Algorithm Using Decision Classifier to Identify Crime Rates

M. Ramzan Begam, P. Sengottuvelan

Abstract--- Prediction of crime data become tremendous in automatic resultant process in law enforcement. Due to vast amount of information in crime records are processed automatically through knowledge mining techniques. The term prediction doesn't make the relational terms to identify correct crime key terms. So the classification analysis based on statistical attribute crime weightage leads more complex to analyze. To solve this problem, we propose a Relative Record linkage based pattern mining algorithm using decision classifier to identify crime rates(RRL-PMD). By analyzing the real terms observed from crime case arein relative sentence format, the sentence case similarity measure predicts the crime key terms to form cluster. The record linkage generalize the frequency of count term measure to reduce the dimensionality make links based on relative closeness measure. The subset classifier make decision to categorize the risk analyzed from crime records. The proposed system produce higher efficiency to reduce the redundancy of complexity level make efficient crime analysis.

Keywords--- Crime Analysis, Decision Classifier, Prediction, Cluster, Semantic Analysis, Record Linkage.

1. INTRODUCTION

Crime data analysis become an emerging field of finding crimes to determine law enforcement against crime creating factor. The day to day development, varieties of crimes arrised due to information and knowledge based thinking of criminals .there is no standard process of crime investigation method to be process to analyze the category. This is occurred due to information doesn't make to proper deal from investigators. So the information among the crime patterns become tedious to angles the standard definition against the crime patterns.

The data mining and the investigational approach do various advancements based on the knowledge learning by analyzing the semantic relation among the crime data. The trivial of information extraction on relevance identification give the essential support to find right decisions. The crime pattern analysis is based on the two depending factors: 1) Information similarity measure (ISM) and 2) attribute content similarity measure (ACSM).to identify the potential information from attributes gives only the crime disciplinary. The clustering and classification analysis resembles the crime patterns to find the relativity measure among the attributes.

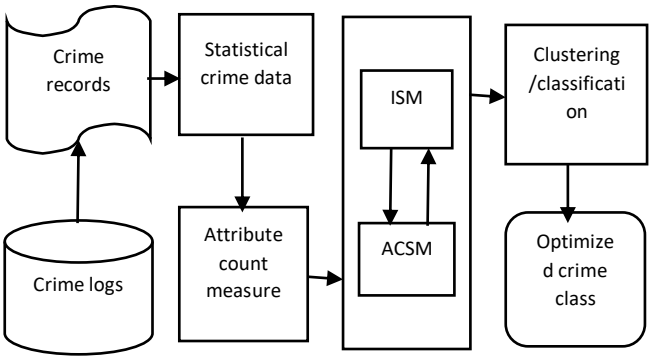

Figure 1: Process of crime pattern analysis

The past crime pattern analyses only consists the crime mapping for produce information from obtained statistical records. The statistical records contains various attributes based on the bio information about the criminals. Most of the crime investigators do only statistical analysis to classify the data. This focusing the attribute based record processing system to measure the count of crimes to produce the ranked result. In such cases the crime involvement maybe vary due to fact of occurring reasons on differential situations.

The crime analyses contains differential task of evaluation whether initially the data obtained from criminals to store in database called crime data logs as shown in figure 1. Further the data logs are organized as crime records to separate the attributes for purpose of crime pattern analysis. To justify the crime patterns to create statistics for law enforcement to make actions. By implementing various terminologies to measure the crime pattern mining algorithm for finding the crime patterns to prevent crimes. The rule mining algorithm participates on crime investigation using association rule mining algorithm, pattern mining, feature classification and clustering and so on. By this appropriate process of crime identification using semantic information similarity measure is the righ choice to predict the crime patterns. The attribute based information analysis needs the multi observation relation between the attribute subset evaluation measures.

By concentrating the crime statement be vary on multi attribute analysis, because the attribute contains statistical analysis of relation measure. So we implement a relative records in crime investigation statements using sentence case relative measure. This sentence case relative measure identify the investigational part of crime statement similar to document evaluation.

Revised Version Manuscript Received on 01 February, 2019.

M. Ramzan Begam, Research Scholar, Bharathiar University, Coimbatore. (e-mail: Prof.ramzanbegam@gmail.com)

Dr.P. Sengottuvelan, Associate Professor and Head, Department of Computer Science, PG Extension Centre, Periyar University, Dharmapuri.

The crime documents contains various relative information related to why the crime was occurred.

The attribute related sentence case similarity measure improve the crime pattern mining process. The features are considered as extracted pattern through the confidence values of sentence case relative measure. The confidence value fix the frequent of crime weightage by class. Further the classification carries the weightage by group by category as class such as high crime risk, moderate, low crime risk. In this paper section 1 contains introduction of crime data analysis, section 2 contains the literature review about dissimilar methods, section 3 contains implementation of the proposed system, section 4 contains the result and discussion followed the conclusion

2. RELATED WORK

A few crime detection techniques are accessible for the fields like fraud, media communications, organize interruption robbery, detections and so on. In any case, the securities can exchange the extortion detection territory is as yet lacking due to criminal records be identified by the classification of risks[1]. Since securities exchange improves the financial advancement of a nation enormously, this field has a fundamental requirement for useful security framework.

Classification" procedures that depend on having "preparing tests" from the misrepresentation and the no fraud bunches based on cluster groups [2]. The PRIDIT procedure does not require a knowledge of which respondents enjoyed misrepresentation with a specific end goal to actualize it.

Perceptions of word co-occurrences and similitude calculations of crimes are regularly utilized as a definite method to speak to the standard settings of words and accomplish a reenactment of semantic word likeness for applications, for example, word or archive clustering and collocation extraction. In spite of the straightforwardness of the fundamental model, it is essential to choose an appropriate hugeness, a similitude measure, and a likeness calculation.

One test for law implementation and insight organizations is the trouble of investigating vast volumes of data engaged with criminal and document-based repeated activities[4]. Data mining holds the guarantee of making it simple, advantageous, and down to earth to investigate comprehensive databases for associations and clients.

Fuzzy clustering calculation to distinguish concealed crooks from point arrange, which took no utilization of people's earlier character data. We initially developed a nearby doubt count from nodes' neighboring data (node and edge); and after that with worldwide data, we utilized the fuzzy k-implies clustering calculation [5], and made the participation to suspicious gathering as the global doubt degree. Examinations demonstrated it functions admirably on distinguishing proof: known suspects increased relative high qualities and known innocents got relative low qualities. such an approach eases the need for law implementation faculty to filter through uninteresting[6], clear standards with a specific end goal to discover fascinating and significant crime patterns of significance to their locale

Identifying cybercrime can in like manner be troublesome because bustling system movement and continuous online exchanges create a lot of data, just a little bit of which identifies with unlawful exercises. Data mining is a ground-breaking instrument that empowers criminal examiners who may need broad preparing as data investigators to investigate extensive databases rapidly and proficiently.

It finds crime patterns and catches the criminal. In this paper, we show an anomaly based data affiliation strategy. An anomaly score work is characterized to measure the extremeness of observation [8]. The proposed philosophy empowers synchronous representation of the geological degree and length of crime clusters, by which steady and transient space-time crime clusters can be naturally separated [9]. Additionally, the consolidated utilization of the two factual procedures uncovered worldly inter-cluster affiliations demonstrating that transient, on the other hand, showed up in a couple of hotspot districts.

The current data mining detection arrangement of business principles and scorecards, and known misrepresentation coordinating have limitations [10]. To address these restrictions and battle character crime progressively, this projected the crime novelty of related multilayered detection framework supplemented with two extra layers: collective detection (CD) and spike detection (SD). Evaluation of lexical semantic relatedness has numerous applications in NLP, and a wide range of measures are used to identify the crime rates. We assess five of these measures, all of which utilize WordNet as their focal asset, by looking at their execution in identifying and remedying certain word spelling mistakes.

The motivation behind this examination was to give definite confirmation on which law authorization can base upgraded proactive fraud control and counteractive action endeavors. It centers on wholesale fraud offenders [12].Measuring the semantic relation between the crime patterns is straightforwardly but instead are figured employing the set of words exceptionally identified with them, which we call the arrangement of determiner words [13]. Our approach for assessing relatedness has a place with website page tallying based measurement strategies. Utilizing criminal offense records, for instance, we employ cross-connection measures, eigenvalue range investigation [14, 15], and comes about because of irregular framework hypothesis to distinguish spatiotemporal patterns on various scales. Misrepresentation detection, the customer, right now had an unmistakable data store where model scoring was performed [16]. Because of the model score, reports would be questioned against the data distribution center to deliver the cases that were suspect.

This examination additionally gathered another corpus of crime and physically marked them. A machine learning classification system represents the node, edge in light of Naïve Bayes [17] and SVM model in separating nationalities, weapons, and crime area from online crime archives.

To classify clustered crimes in light of occurrence frequency amid various years[18]. Data mining is utilized widely as far as examination, examination, and disclosure of patterns for the occurrence of various crimes[19, 20]. Criminological research would profit by growing the explanatory concentration from private neighborhoods to the system of neighborhoods that people are presented to amid their day by day routine exercises.

3. IMPLEMENTATION OF PROPOSED SYSTEM

Crime data analysis become traditionally to solve automatic process of investigation records from crime detection approach. In present day, crime data analysis concentrates the statistical approach to analyze the crime data to predict the feature crimes. But the traditional problems doesn't solve relational identification of crime problems. So we need a advancement to analyses the crime data to extract the crime features to predict the result. The optimized result are based on the class by risks. To implement a new crime rate analysis for a law enforcement for crime detection called RRL-PMD: Relative Record linkage based pattern mining algorithm using decision classifier to identify crime rates. Mainly our proposal concentrates relative sentence case similarity measure to detect crime risks by frequency of the confidence level.

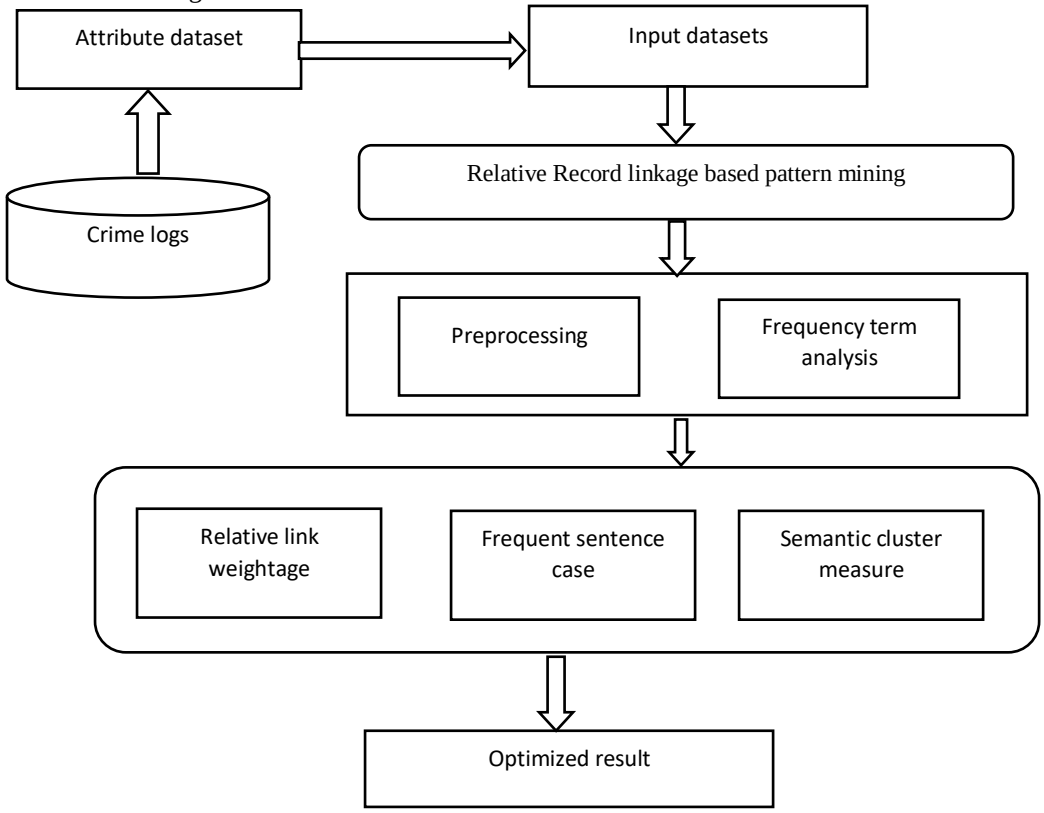

Figure 2: Shows the proposed architecture diagram

The collection of crime dataset which the provided the real enquiry data observed from criminal. The dataset contains bio details considered as attributes like name, crime id, crime type, crime count, crime reason, etc. figure 2 shows the proposed implementation of architecture. The crime analysis begins preprocess to reduce the task reduction for preparing data to predict the crime patterns. Initially the record linkage projects the coincidence match of relative measures among the crime rates. The crime counts are highly considered as the risk factor of attribute. The attribute based relative measure have the scores for depending sentence level counter measure. In secondary think, frequency measures between the attributes are measures by confidence value which the crime patterns are repeatedly occurred. The proposed system improves the Cluster efficiency by measure the semantic relational measure. The frequent terms sentence has crime related observations. Through this observation sematic similarity document case are analyzed by giving the weighted terms. Finally the subspace clusters formed the group by relevance category approach based on the risk.

The cluster ensembles the similarity groups to point the centroid weightage based key terms. The similarity of cluster groups are formed by relative closeness measure between the crime attributes. The subset features identify the attribute pattern by statistical analysis as considered as object similar to other relative crime records. Cluster finalize the category of groups by identify the risks by class.

The main objective of this crime analysis is is proceeding the sentence case crime statement record evaluation or crime document analysis, specifically the relational analysis of crime pattern violates the law enforce support to make decisions against crimes. The information of statement analysis has the meaningful data which is sentimental, behavioral fact of crimes which is originates the realistic crimes or even related to situational related crime occurrence .the disseminated features of identified crimes are categorized to identify the level of risk level.

The classes are categorized by decision classifiers using the maximum states confidence value.

3.1 Crime Data preparation

Crime data analysis contains high volume of data in the form of high dimensional to pointing various attributes. Due to the collection of vast information the noise become raised. (I.e., irrelevant information, data mismatch, empty fields, other irrigated signs. etc.). The collection of records contains the attribute which is redundant form to select the crime patterns. The preprocessing integrates data cleansing, empty record removal filling the nearest missing value to reduce the burden of the further proposed implementation. The preprocessing enhance to reduce the dataset. Also this sentence case observer is used to remove the stop words to reduce the length of the sentence. This points the count terms of crime patterns are extracted from the records for frequency relation identification.

Algorithm

Input: collective crime datasetCrmds
Output: Noise free crime dataset Ps
Step 1 start process to compute collective crime set
ProcessStart
Observe all the record at primary crmd1, crmd2….crmdn
Step 2attribute observation by collective record set Amdi

Compute each record set
For I=0Amd$_i$ fromCrmds i=n
Attribute value V = Extract crime content from Adi.
Check null ==empty:
Remove→crd;
Originate order of crime id ←crds
Step 3: compute the sentence in the relative record
Identify the attribute term matches to crime patterns
Word pattern selective set Ss= $(\sum_{n=1}^{size(Ds)}$ Text€Di) ×Splitby(.)
Observe the sentence selective setRs←selective word (SL)

Step 3 compute each sentence
Forselective word SL from Rs
Identifykey terms from crime set
Crime set CTi = ∑bag of Terms
For each crime term Ct from Rs
If Rs€Stop wordSet removal then
Key termkt←Rs
Else
Cleaning fact for Stemming non keywords
Speech term to reduction.
End
End
Step 4: compute the max bag of crime terms returns count value Ps
Ps←kt++;
End
End
Stop

The above algorithm prepares the crime data for further relative record linkage identification process. This reduce the dimensionality of the crime datasets which is redundant to make easy process. This process reduces the time complexity of crime pattern investigation because of the least selected feature used to make efficient crime analysis. This find the behavioral of data originate the main key terms.

3.2 Relative record linkage

In this stage non homogenous value of crime data be analyzed through record linkage concept. This process the individual contents of crime sentence variables, attributes, tokens are relatively identified by links of connectivity between the communal records. Because the frequent crimes are associated based on repeated count. These crime variables has identity links to clearance the relational connectivity between the crime records. This reduce the non-relational terms as well cleansing way and to combine distortional records as same attribute relational variables to reduce the non-relational terms.

Input: preprocessed dataset Ps
Output: Relative weightage Rds set
Start
Step 1: Initialize the processed set Ps
For each record (read ←crime id)
Step 2:For each attribute Ca from Ps
Identify the count terms Ct value
Max term value (max variable count > confidence)
Rearrange the crime Id
Create links between term crime id→cd1,cd2,..
Step 3: compute the key terms sentence kt
For each crime record kt from Ps
For each crime term Ktfrom Ps
If Kt€ max term then
Identify max count relation with other attributes.
Relation Set Rs = ∑(Concepts€Kt)+Ps.
Compute Number of attribute relations it has.
Kts= ∑Relations€Gi similar term
Compute the max count attribute value
mval = ∑record Links(Kt)<-∑mval(Kt)
Create link Kt identified relative record links
Relative link Rt←kt+ca;
End
End
End

Step 5: compute the conceptual record links→CRL.
For each record←RL
CRL = = ∑Concept (Links (Kt))) € ∑Concept (Ca)!=Ps
Compute relative record→RLK
RLK = $(Kt + Ca) + NIL$
Add to record link set CRL←RLK;
End
End
Stop.

The above algorithm reduce the frequent intervals between the connected attribute links that are related to each other. The maximal crime have count reference to the repeated crimes. So distinct of frequency be evaluated to originate the data.

3.3 Frequent sentence case pattern extraction

The repetitive terms are easy to incident of crime occurrence the relativity depends number of times the key terms are occurred.

The repeated pattern have the max confidence value to extract the feature for redundant crime occurrences. To search some relationships by using the frequency occurrence of crime incidents.

Then, they analyzed the result to produce max confidence frequency risk by category, which can be used to perceive pattern based on the behavior of criminal**.**

Algorithm

Input: crime frequent set Pi, Pattern set Ops.
Output: Crime terms Pi.
Step1: initialize PI to null.
Step2: for each item I from SI
Identify the patterns where I is present.
PS = number of crime transactions
For each pattern Pi from PS
Compute count of Pi = max (count $\sum TS(i) == Pi$)
End
Step 3 Compute support = Count/size(Ts).
Add to MinMaxcount MM = $\sum MM(i) + Support$
If count>(then
Add I to frequent crime set PI = $\sum PI(i) + I$
End.
End.
Step 4: stop.

The crime factor depends the key terms from the attribute value measure by repeat count terms to find the frequency of repeat terms the record linkage utilize the total crime occurrence by number of different crime involved by various level.

For example the different crime records represent a single person have involved in multi attribute reference terms as shown below.

Table 1: Crime data terms

Crime id	Crime terms identification	Record linkage statement
1	{theft, murder, robbery}	Person involved theft, murder, robbery
2	{theft, murder}	Person involved theft, murder
3	{Rapist, robbery}	Person involved rapist , robbery
4	{Murder, rapist}	Person involved murder, rapist

The above table 1 shows the crime terms contains the count term record contains the linkage to other terms of crime relational sets to covers the frequent crime occurrence of groups on the different involvement of crimes.

The frequency of the crime records contains number of count occurrences have the different involved terms to be access.

Table 2: Frequency confidence level in record linkage

Crime terms (CT)	Record linkage (RL)	Max confidence value(Mcv)	Sentence level key terms	Crime Frequency level (%)	Crime risk(CR)
All crime	1, 2, 3, 4	4	person involved all crimes	100	High
theft	1, 2	2	Person involved repeat theft counts	50	moderate
murder	1, 2, 4	3	Person involved other crime with murder	75	high
rapist	3, 4	2	Person involved rapist with robbery and murder	50	Moderate
robbery	1, 3	2	Peron involved theft robbery	50	Moderate
theft, murder	1, 2	2	Person involved theft murder with robbery	50	Moderate
theft, robbery	1	1	Person involved either theft or robbery	25	Low
murder, rapist	4	1	The person involved rapist or murder	25	Low
murder, robbery	1	1	Person involved murder or robbery.	25	Low

The above table 2 shows the frequency of crime terms with related occurrence of crime in different stages. Based on the person involvement the risk factor analyze the involvement of confidence record linkage level to find the frequency by utilizing risk by category of access. Relational terms are analyzed using the sentence level of crime terms average to the repentance of pattern occurred in max confidence state.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

At the end the crime risk factor fix the risk to group the semantic relational access.

3.4 Semantic pattern relational closeness measure

In this section, the semantic sentence case relational terms are taken from the attribute case reason. This crime reason contain the real entity of documentation holds the crime information. The sentence contains the behavioral part of crime involvement plans, purpose, persons involved etc. all the terms are relatively closed to the crime rates form risks. Using the semantic similarity based on singular vector decomposition (SVD) to find the relative closeness of crime pattern. The crime key terms are analyzed with extracted keyword related on crime objectives. This analyses keyword relation between two terms based on the frequency level of patterns.

Algorithm:

Input: frequent processed record crime set:

Output: optimized relational class

Step1: Compute the frequent crime set processed records

Start

Observe the term crime record set Ts.

Sematic ontology observation Ts→O.

Step 2: Compute the crime set

For each crime record semantic indexTi from Ts

Observe bound measure in semantic relation SMr = Nc/Tn.

Non represented class→Nc = Ti representation of number of class T

Term point of crime set →Tn- ontology representation of terms

End observation

Step 3 compute the category of semantic observation

Representing For each class C

Process the closeness measure by semantic relation Scm

$$=\int \frac{sbm}{Numberoftermspresentinotherclass}$$

Step 4.Relational process can be identified by the following equation,

$$RT\ (t) = \frac{term\ to\ serching\ the\ data\ in\ document}{total\ number\ of\ documents\ in\ dataset}$$,*(t-> in terms of time frequency)

Step 5: Defending to crime set identification record set

Relational semantic text level

$$SRT = \frac{\log\left(\sum crime\ relation\ text(RT)\right)}{total\ crime\ in\ relation\ text}$$

Step 6: sematic relational identification from semantic measure.

$$RI = \frac{\mathrm{RT(t)} \times SRT}{Total\ time\ by\ document}$$

End

Select the closure measure at top rate class C = O(Max(Scm))

Step 7: Compute the synonyms relations

$$\sum (\mathrm{Vi}, \mathrm{Vj}) = . \frac{(distance(\mathrm{Vi},\mathrm{Vj}) + \beta(Seq) * \mu * (d(Vi) + d(Vj))}{N(Vi,Vj) * 2 * seq * \max\left(d(Vi) - d(Vj)\right)}$$

Step 8: Non elements are observed Ne = $\sum Terms(Ts) \nexists O(c)$

Stop

Where d (Vi)and d (Vj) respectively express the hierarchy of Viand Vj corresponding to the node of ontology tree. Distance (Vi , Vj) is the weight-value sum of all edges on the shortest path between Vi and Vj also N(Vi,Vj) is the number sum of edges on the shortest path between Vi and Vj. Seq is the maximum depth of ontology tree, β is an adjustable parameter.

3.5 Cluster class by evaluation

This clustering algorithms in data mining are equivalent to the task of identifying groups of records that are similar between themselves but different from the rest of the data. In our case some of these clusters will be useful for identifying a crime spree committed by one or same group of suspects. Given this information, the next challenge is to find the variables providing the best clustering. These clusters will then be presented to the detectives to drill down using their domain expertise. The confidence score identify the frequency of the crime evaluation by the category.

Input : crime Term set Ts, classify request term O

Output : Semantic Closeness measure; frequent relative scoreFrs, crime risk cr

Start

Step 1: Readcrime Term set ts.

Step 2: Observe the semantic ontology term O←Ts.

For each frequent term Ti from selectivesetTs

If match case crime weightage ws←ti

Step 3: Create class by semantic measure

For(crime state a→ws)

Group class by category class C

End for

End

Step 4: Compute Cluster by class evaluation

For each class C

Crime cluster Cs→Relational category similarity

Cs← cs1, cs2…

End

Return Cs class by cluster group

Stop

Finally the risk factors are crimes are arranged by class by reference .by the confidential mean of relative score evaluation. The classification points the confidence level to retail the crime factors by order to which the intelligence to make crime scores support to the investigator.

4. RESULT AND DISCUSSION

The crime accuracy are tested with collected criminal datasets from investigational departments which is in the form of real high dimensional dataset. The relative record linkage concept is applied to find the somatic measure among crime records to cluster group by risk factor evaluation. The preprocessed data set is input progress for testing the sensitivity specificity and clustering accuracy. The proposed RRL-PMD produce higher efficient test result compared to the other dissimilar methods. The crime case reasons are evaluated through visual studio frame work .net 4.0 by testing the various crime records. The implementation enhance the crime case analyses with higher efficiency to identify class by risk of evaluation. The table given below shows the values and parameter processed in test case analysis.

Table 4.1: Details of a crime Dataset

Parameter	Value
Framework	visual.net framework 4.0
Number of crime Records	30000
Number of Crime attributes with relative documents	10
Number of class by category	High risk, low risk, medium ,moderate

The above table 4.1 describes the test case parameters with values processed to categorize the class of crime risks. The evaluation of crime records are finalized through the cluster accuracy by comparing other methods with least time complexity, false classification, the evaluated cluster are classified by risk by group of records done by semantic relational key terms.

The proportion of negatives which are correctly identified is measured by Specificity. In order to find these metrics, we first compute some of the terms like, True positive, True negative, False negative and False positive based on the crime patterns.

$$\text{Sensitivity} = TP/(TP + FN);$$

$$\text{Specificity} = TN/(TN + FP);$$

$$\text{Accuracy} = (TN + TP)/(TN + TP + FN + FP)$$

The performance testing resultants are figure out by representation points that,

Cluster evaluation accuracy (cs)

$$=\sum_{k=0}^{k=n} \times \frac{total\ crime\ class\ representation(Cds)}{Ttotl\ crime\ records(Tr)}$$

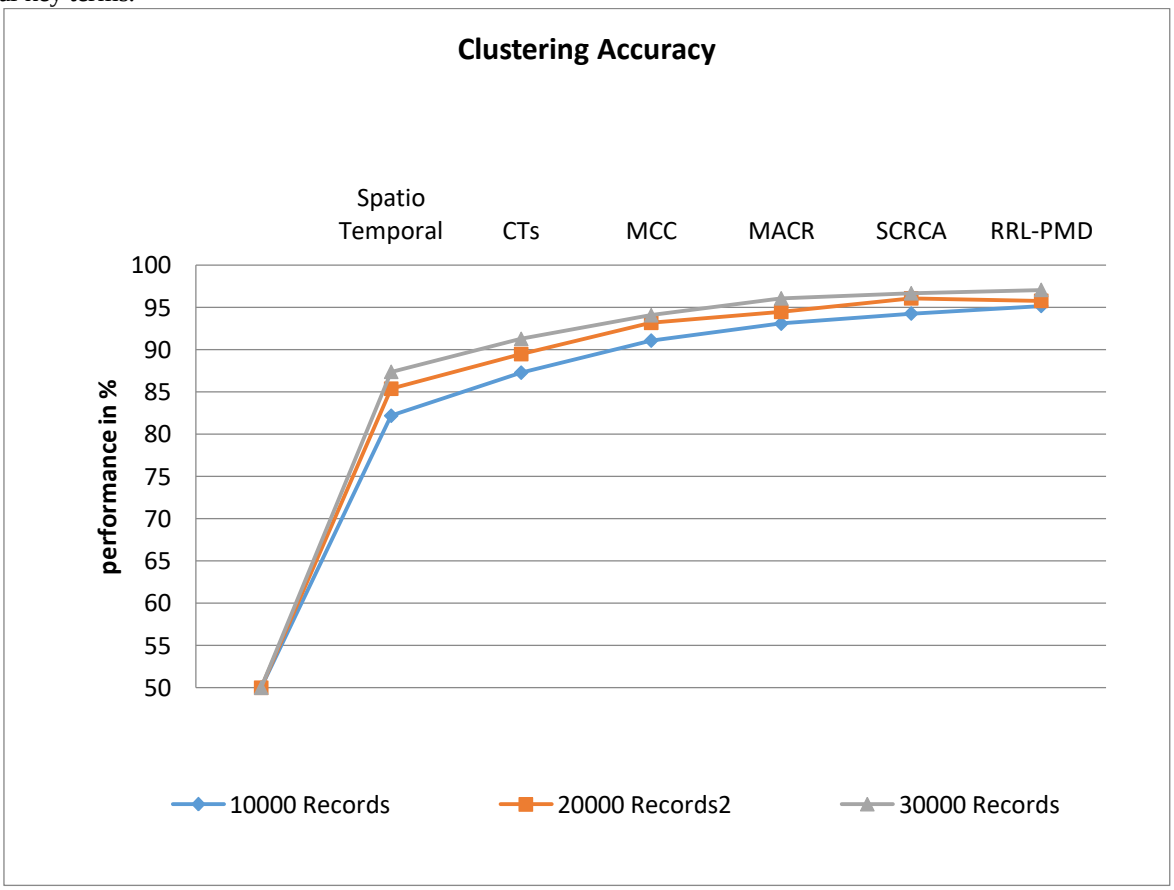

Figure 4.3: Comparison of Clustering Accuracy

Figure 4.3, shows the clustering accuracy comparison which they produce higher performance revised other methods. This improvement related to task finding of feature based cluster analyses that doesn't availed by other methods. So the performance be relatively high than other methods.

Table 4.2: Comparison of clustering accuracy

Methods/number of records	Spatiotemporal	Cts	MCC	MACR	SCRCA	RRL-PMD
10000 records	82.2	87.3	91.1	93.1	94.3	95.2
20000 records	85.4	89.5	93.2	94.5	95.3	95.8
30000 records	87.4	91.3	94.1	96.1	96.7	97.1

Table 4.2, reviews the execution of crime pattern clustering accuracy SCRCA has produced 10000records as 94.3%, 20000records as 95.3% and 30000records as 96.7 % shows that the RRL-PMD proposed approach has produced higher clustering accuracy 97.1 %.

The similar datasets are ignored as unclassified clusters be considered as false classification, the false classification is calculated by

False classification Ratio (Fcr)

$$=\sum_{k=0}^{k=n} \times \frac{clustering\ Accuracy(cs)}{Total\ no\ of\ failed\ cluster\ rate(Fr)}$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

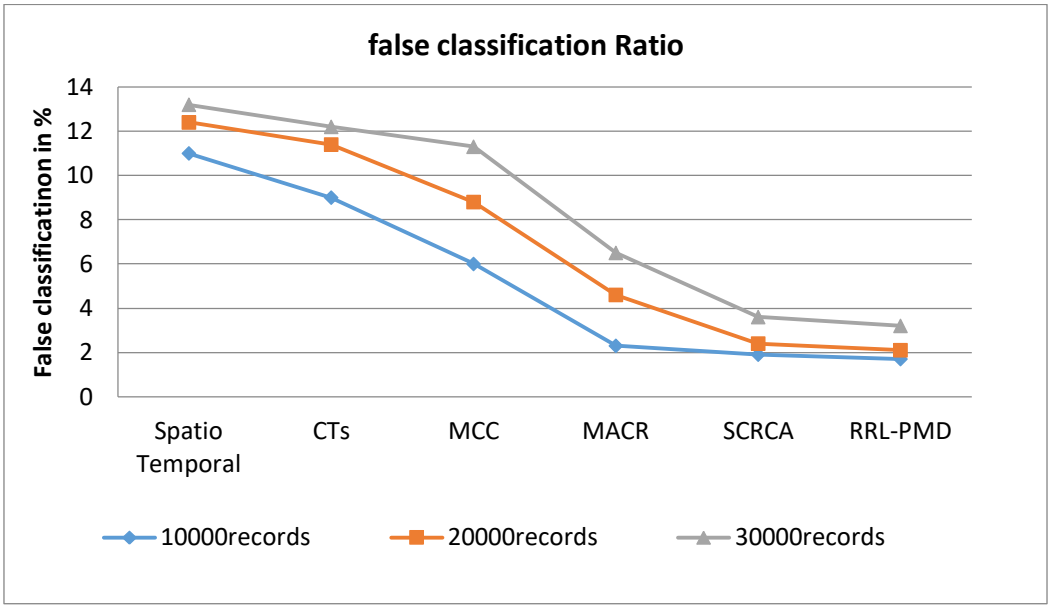

Figure 4.4: Comparison of false classification

The Figure 4.4, reviews the extraction of false rate using various representation of techniques the the implementation new system produce higher efficient compared to other dissimilar methods.

Table 4.3: Comparison of false classification

Methods/number of records	Spatiotemporal	Cts	MCC	MACR	SCRCA	RRL-PMD
10000 records	11	9	6	2.3	1.9	1.7
20000 records	12.4	89.5	8.8	4.6	2.4	2.1
30000 records	13.2	12.2	11.3	6.5	3.6	3.2

The Table 4.3, shows the comparison of false classification ratio produced SCRCA has 10000records as 1.9%, 20000records as 2.4% and 30000records as 3.6 % shows that the RRL-PMD proposed approach produces less false classification ratio 3.2 well.

The average time taken to calculate the dataset cluster evaluation is calculated by,

Time complexity (Tc) $=\sum_{k=0}^{k=n} \times \frac{clustering\ Accuracy(cs)+false\ classification(Fcr)}{Time\ taken(Ts)}$

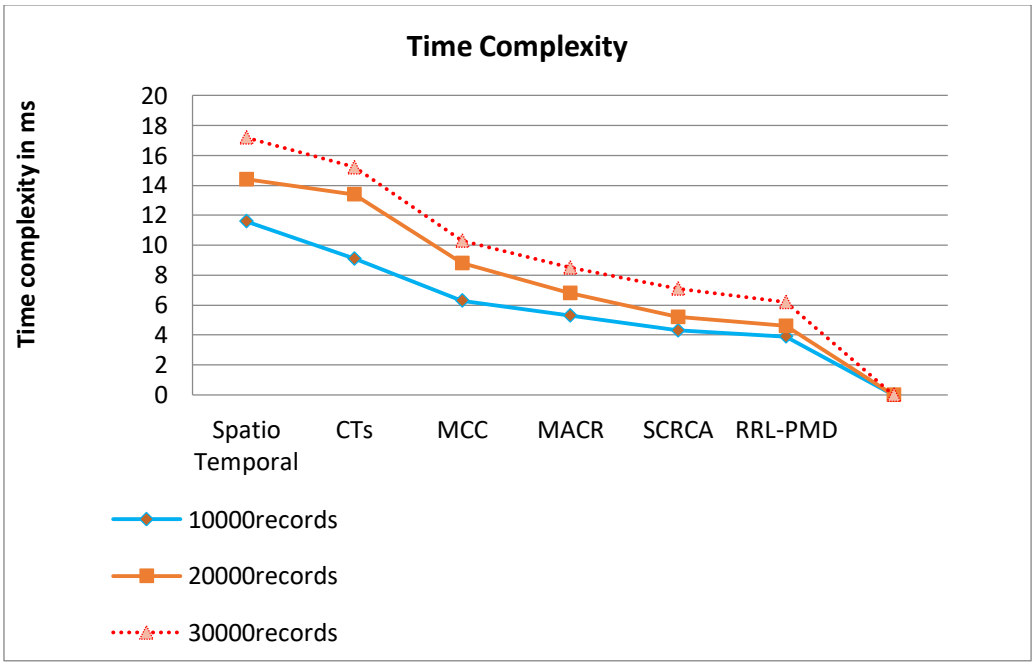

Figure 4.5: Comparison of time complexity

The Figure 4.5, shows the execution rate to time evaluation by finding the crime records. The proposed execution has lower time taken to complete the execution compared to the other system.

Table 4.4: Comparison of time complexity

Methods/number of records	Spatiotemporal	Cts	MCC	MACR	SCRCA	RRL-PMD
10000 records	11.6	9.1	6.3	5.3	4.1	3.9
20000 records	14.4	13.4	8.8	6.6	5.2	4.6
30000 records	17.2	15.2	10.3	8.5	7.3	6.2

The Table 4.4, reviews the execution of time complexitye valuation SCRC A produced 10000records as 4.1(ms), 20000records as 5.2(ms) and 30000records as 7.3(ms) shows that the RRL-PMD proposed approach has produced less time complexity up to 6.2 % well.

Frequent occurrence (Fc)

$$=\sum_{k=0}^{k=n}\times\frac{repeated\ clusters(Rs)+irrelavant\ clusters(Irc)}{Total\ number\ of\ clusters}$$

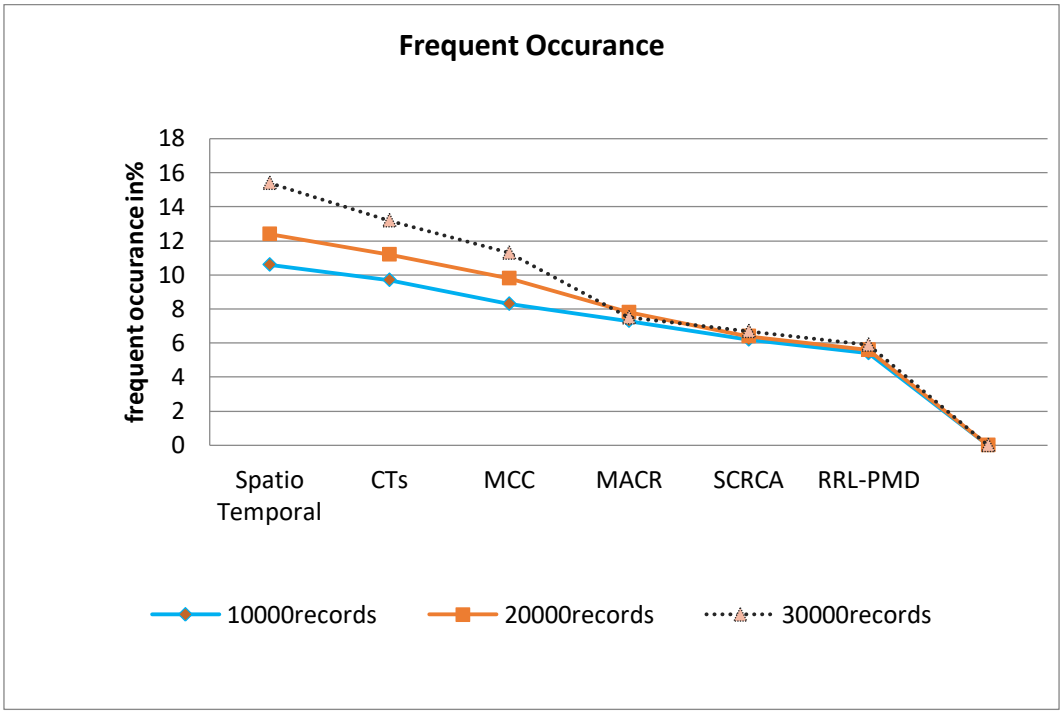

Figure 4.6: Comparison of the frequent occurrence

The Figure 4.6, shows the repeated execution of evaluation frequent terms from crime records identified different methods the represented crime novelty reduce the performance of cluster evaluation. The projected implementation had higher rate of lower frequent observation then other methods.

Table 4.5: Comparison of frequent occurrence

Methods/number of records	Spatio temporal	Cts	MCC	MACR	SCRCA	RRL-PMD
10000 records	10.6	9.7	8.3	7.3	6.1	5.4
20000 records	12.4	11.2	9.8	7.8	6.4	5.6
30000 records	15.4	13.2	11.3	7.5	6.7	5.9

The Table 4.5, reviews the dissimilar analysis frequent occurrence SCRCA produced 10000records as 6.1%, 20000records as 6.4% and 30000records as 6.7 % shows that the RRL-PMD proposed approach has produced less frequent occurrence up to 5.9% well.

5. CONCLUSION

The proposed relative record linkage based crime detection identify the effective crime rates which is process the crime patterns from identified key term extraction. The crime problematic factors leads non identical terms of relational analysis. The proposed pattern mining decisions makes the cluster evaluation accurate by reducing the feature to handle the classification. They conclude the proposed make maximum redundant features to handle the crime data. The semantic analysis find the frequency between the terms of crime rate occurring the repeated terms with maximum confidence value to make decision. Finally the relational semantic score fix the centroid key terms to form clusters to make decision to split by the category of risk. The proposed implementation produce higher efficient result of accuracy 97.2 % well with least time complexity 6.2 milliseconds. This support real-time investigation of mankind approach in crime analysis.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

REFERENCES

1. R. Bolton and D. Hand, "Unsupervised Profiling Methods for Fraud Detection," Statistical Science, vol. 17, no. 3, pp. 235-255, 2001.
2. P. Brockett, R. Derrig, L. Golden, A. Levine, and M. Alpert, "Fraud Classification Using Principal Component Analysis of RIDITs," The J. Risk and Insurance, vol. 69, no. 3, pp. 341-371, 2002,
3. S. Bordag, "A comparison of co-occurrence and similarity measures as simulations of context," in CICLing, 2008.
4. Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer JieXu, Gang Wang, RongZheng, HomaAtabakhsh, "Crime Data Mining: An Overview and Case Studies", AI Lab, University of Arizona, proceedings National Conference on Digital Government Research, 2003, available at: http://ai.bpa.arizona.edu/
5. Fan, C., Xiao, K., Xiu, B., &Lv, G. (2014, August). A fuzzy clustering algorithm to detect criminals without prior information. In Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on (pp. 238-243). IEEE.
6. BUCZAK, A. L., AND GIFFORD, C. M. Fuzzy association rule mining for community crime pattern discovery. In ACM SIGKDD Workshop on Intelligence and Security Informatics (2010), ACM, p. 2.
7. Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer JieXu, Gang Wang, RongZheng, HomaAtabakhsh, "Crime Data Mining: A General Framework and Some Examples", IEEE Computer Society April 2004.
8. Lin, S., & Brown, D. E. (2006). An outlier-based data association method for linking criminal incidents. Decision Support Systems, 41(3), 604-615.
9. NAKAYA, T., AND YANO, K. Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. Transactions in GIS 14, 3 (2010), 223–239.
10. Clifton Phua, Member, IEEE, Kate Smith-Miles, Senior Member, IEEE, Vincent Lee, and Ross Gayler, "Resilient Identity Crime Detection", IEEE transactions on knowledge and data engineering ,vol.24 no.3 year 2012
11. A. Budanitsky and G. Hirst, "Evaluating wordnet-based measures of lexical semantic relatedness," Comput. Linguist., vol. 32, no. 1, pp. 13– 47, March 2006.
12. G. Gordon, D. Rebovich, K. Choo, and J. Gordon, "Identity Fraud Trends and Patterns: Building a Data-Based Foundation for Proactive Enforcement," Center for Identity Management and Information Protection, Utica College, 2007.
13. M. A. Salahli, "An approach for measuring semantic relatedness between words via related terms," Mathematical and Computational Applications, vol. 14, no. 1, pp. 55–63, April 2009
14. TOOLE, J. L., EAGLE, N., AND PLOTKIN, J. B. Spatiotemporal correlations in criminal offense records. ACM Transactions on Intelligent Systems and Technology (TIST) 2, 4 (2011)
15. X. Li and M. Juhola, "Crime and its social context: Analysis using self-organising map," Proceedings of the European Conference on Intelligence and Security Informatics, pp. 121-124, 2013
16. AniruddhaKshirsagar, Lalit Dole, "Recognizing the theft of identity using data mining" ,International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 4, April 2014
17. K. R. HafedhShabat, Nazila Omar, "Named entity recognition in crime using machine learning approach," in Information Retrieval Technology, December 2014, pp. 280–288
18. Kiani, R., Mahdavi, S., &Keshavarzi, A. (2015). Analysis and Prediction of Crimes by Clustering and Classification. Analysis,
19. X. Li, H. Joutsijoki, J. Laurikkala, M. Siermala, and M. Juhola, "Crime vs. demographic factors revisited: Applications of data mining methods," Webology, Vol. 12, No. 1, Article 132, 2015.
20. GRAIF, C., GLADFELTER, A. S., AND MATTHEWS, S. A. Urban poverty and neighborhood effects on crime: Incorporating spatial and network perspectives. Sociology Compass 8, 9 (2014), 1140–1155.
21. Bin Pei, Xiuzhen Wang, Fenmei Wang, Parallelization of FP-growth Algorithm for Mining Probabilistic Numerical Data based on MapReduce,2016 9th International Symposium on Computational Intelligence and Design, 2473-3547/16 $31.00 © 2016 IEEE.