Perspective of Population Characteristics on Mining Offenders specifics

Baldev Singh
P.G. Dept. of Computer Sc. & I.T.
Lyallpur Khalsa College, Jalandhar, Pb. India
bsd.Lkc@gmail.com

Abstract— Population characteristics and processes both are the drivers and results of social and economic development processes and outcomes. A community's population characteristics like age, background, close circle, employment, education, use of drugs, family history, per capita income, locality, housing status etc. are of concern that influences various offences and offence levels in our society. In addition, drinking habits, drugs, per capita income, close circle, status, localities etc. in a community provide clues to causes or solutions to the extent of offence. This paper spotlights on link between population characteristics to offence outcomes. In this paper, emphasis is on to apply data miner on population characteristics relevant data to identify its effects on offences/ misdeeds. On the basis of this analysis, link between population characteristics and wrong doings (offences) be identified and Govt. and its allied agencies, may have programs and policies to pull down the wrong deeds to avoid the adverse effects.

Keywords— Data mining, offenders, population characteristics, offence.

I. INTRODUCTION

Data mining is the technique of extracting meaningful information from large and mostly unorganised data stores. It is the process of performing automated extraction and generating predictive information from large data stores. It enables us to understand the current offence trends and enables to mine the huge data and to take proactive measures to gain maximum results from the same. In recent years, database technology has advanced in stride.

The relevance of data mining is enviable in various fields for knowledge discovery [3]. Implementation of data mining in exploring the offence is equally important. The application of data mining techniques [11] in exploring offence is indispensable to examine, understand and envisage behaviours. In the offence revealing segment, there are many openings to apply data mining, in particular, analysis to detect and analyse, envisage the effects of various parameters to identify the relevant matter for exploring the offences.

This paper in consideration is based on mining the data which is related to the offenders [14]. An excel sheet is used to store the data about offenders and the cases that they are facing. On the other hand, the huge courts data presents tremendous opportunity for data mining applications as well. Judiciary is one of many application areas that the explosive growth of data mining integration is predicted to continue. The most appropriate data set in this research is offenders' data archive.

The goal of using data mining in lawful system is to produce new knowledge base that the users can act upon. This paper revealed the results by way of data mining tools like the discovery of different patterns [5] to be extracted through different data mining tools [12] and visualization techniques that will suggest rational policy. The judiciary will be able to use data mining to access offence pattern [4,13] and adjust resource allotments accordingly. This paper is based on the project study. Various objectives of the findings are:

- Speed up decision making process
- Load balancing
- Reduction in harassment because of speedy system
- Outcome optimisation
- Offence minimization due to fast decisions etc.

II. OFFENDERS DATA CLASSIFICATION

This paper explored the project study that was to analyze the relationship between offence type and its patterns, only transactions relevant to the research were obtained. Since, it made more sense to apply data mining with vast amount of data sets so that the automation capability of data mining can be valued; the expected data set is relatively large. The data set used in the study is of offenders' related data. It was also assumed that the data is confidential. To preserve the confidentiality of data, the data set was sanitized so that the proprietor of the data was kept anonymous and sensitive information was eliminated. Sensitive information in this database is such as offenders-id, offence level, education of offenders, education, age group, family and friend history etc.

Although the data set did not represent complete general data, it is considered complete for case analysis. However, it is worth nothing that, due to the incomplete data set, missing information and limited supporting knowledge and information, the scope of test was limited to a few areas. This is somewhat different from a normal legal engagement [6] where the information limitation is a serious matter that could prevent the data miners from rendering an opinion.

The initial data set contains forty seven tuples (or transactions), with eleven attributes (or columns) where two continuous attributes and remaining were of explanation in nature. Due to the limitation of available data, the area of research was the sample selection step of the "test of controls" phase. The research was focusing on the use of data mining techniques [4,5] to assist judgement job.

III. IMPLEMENTATION OF DATA MINER IN DISCOVERING OUTCOMES

Similar to data mining applications in various other domains, mining justice enforcement data have many challenges. First, partial, erroneous, or inconsistent data can create inconvenience. Second, the exceptional characteristics of offenders' specifics cause problems that are not general in other data mining applications.

It is generally difficult to determine the exact number and grouping of variables [8] that are to be used in the development process. Within this paper, the intention is to model offenders' behaviour to set up consistency and uniformity across offences. To examine this, different sets of variables representing particular behavioural traits [7] are used. Here we apply various population characteristics for the purpose of mining offenders' specifics. First, we must define the INPUT attributes. We add the DEFINE STATUS component in the data miner software menu. We select the INPUT attributes in the following dialog box.

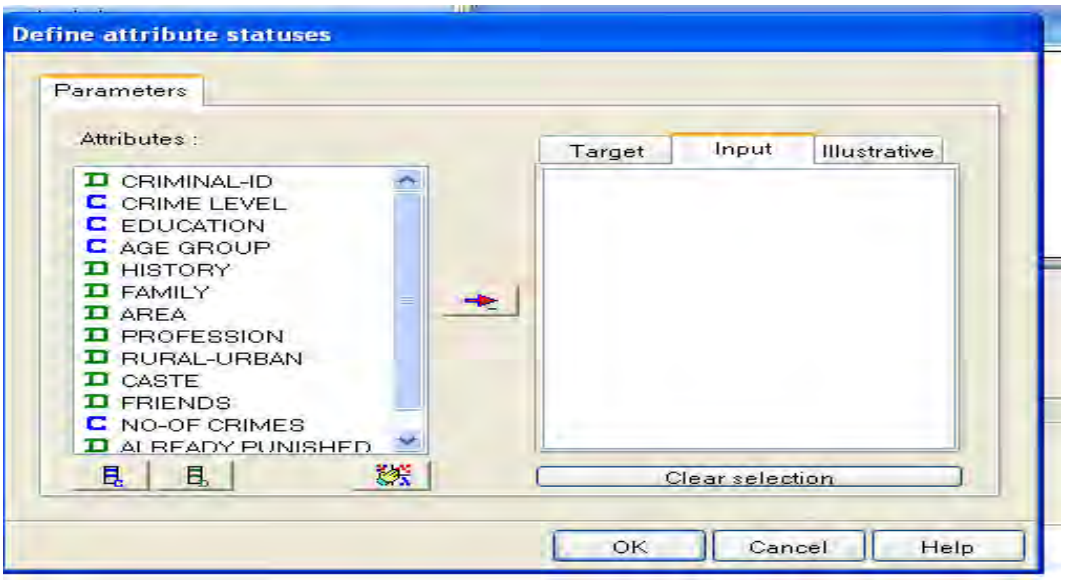

Figure I: Attributes Selection for Data Mining

(i) Using data mining component (Association –A Priori)

Association analysis [10] is the discovery of what are commonly called association rules. A Priori is one of the components of Association. Here we consider four input variables and one target. The input variables used here are: Age group, History, family and Friends(close circle) of the offenders. The target variable is Offence (crime) [12,13] level. It shown that we found that on the basis of these variables (e group, History, family and Friends of the offenders), the offence level of the offenders is affecting. Following table depicted the results of association component of data mining:

A priori 1

Parameters

A-Priori parameters

Support min	0.25
Confidence min	0.75
Max rule length	4
Lift filtering	1.1

Results

ITEMS

Transactions	47

Counting items

All items	
Filtered items	

Counting itemsets

card(itemset) = 2	
card(itemset) = 3	

Rules

Number of rules	

RULES

Number of rules : 7

N°	Antecedent	Consequent	Lift	Support	Confidence
1	"AGE GROUP=25-30"	"HISTORY=not good"	2.043	0.298	1

2	"FAMILY=Normal" - "FRIENDS=Normal"	"HISTORY=not bad"	1.498	0.277	0.765
3	"AGE GROUP=30-35"	"FAMILY=Normal"	1.306	0.383	1
4	"FRIENDS=Bad Company" - "AGE GROUP=30-35"	"FAMILY=Normal"	1.306	0.277	1
5	"HISTORY=not bad" - "AGE GROUP=30-35"	"FAMILY=Normal"	1.306	0.255	1
6	"HISTORY=not bad"	"FAMILY=Normal"	1.251	0.489	0.958
7	"HISTORY=not bad" - "FRIENDS=Normal"	"FAMILY=Normal"	1.212	0.277	0.929

Table-I: A-Priori Result

Here Support and Confidence are the parameters that are used to justify that whether the hypothesis is accepted. Here we assume that the minimum support 0.25 and minimum confidence 0.75 is required to make the hypothesis acceptable. The result shows that family, history and a particular age group affect the offenders' to offence, which is shown in above result.

(ii) Using data mining component (Group Characterization)

Group Characterization is a comparative descriptive statistics, which is used to characterize groups [1,9] defined by discrete attributes. Following are the results produced on Offence data. On the basis of inputs, the target is characterized as High, Middle and Low on the basis of Inputs. Conclusions reached after analysing are:

- Group persons likely to come from the same age.
- Group persons likely to cooperate and come together for required skills to commit offence.
- Group persons are high likely coming from the same area.
- Group persons are likely to operate in the same areas.

Group characterization 1

Parameters

Normalization : 1

Results

OFFENCE LEVEL=High				OFFENCE LEVEL=Middle			
Examples	[55.3 %] 26			Examples	[29.8 %] 14		
Att - Desc	Test value	Group	Overral	Att - Desc	Test value	Group	Overral
Continuous attributes : Mean (StdDev)				Continuous attributes : Mean (StdDev)			
Discrete attributes : [Recall] Accuracy				Discrete attributes : [Recall] Accuracy			
AGE GROUP=30-35	0.32	[77.8 %] 53.8 %	38.30%	AGE GROUP=35-40	0.63	[63.6 %] 50.0 %	23.40%
AGE GROUP=20-25	0.12	[100.0 %] 3.8 %	2.10%	AGE GROUP=25-30	0.13	[35.7 %] 35.7 %	29.80%
AGE GROUP=25-30	0.11	[64.3 %] 34.6 %	29.80%	AGE GROUP=20-25	-0.15	[0.0 %] 0.0 %	2.10%
AGE GROUP=40-50	-0.1	[33.3 %] 3.8 %	6.40%	AGE GROUP=40-50	-0.26	[0.0 %] 0.0 %	6.40%
AGE GROUP=35-40	-0.46	[9.1 %] 3.8 %	23.40%	AGE GROUP=30-35	-0.49	[11.1 %] 14.3 %	38.30%

Description of "OFFENCE LEVEL"

OFFENCE LEVEL=Low			
Examples			[14.9 %] 7
Att - Desc	Test value	Group	Overral
Continuous attributes : Mean (StdDev)			
Discrete attributes : [Recall] Accuracy			
AGE GROUP=40-50	0.91	[66.7 %] 28.6 %	6.40%
AGE GROUP=35-40	0.46	[27.3 %] 42.9 %	23.40%
AGE GROUP=20-25	-0.15	[0.0 %] 0.0 %	2.10%
AGE GROUP=30-35	-0.2	[11.1 %] 28.6 %	38.30%

AGE GROUP=25-30	-0.65	[0.0 %]	0.0 %	29.80%

Table II: Result Based on Group Characterization

(iii)Supervised Learning Assessment – Decision List

Supervised Learning is used when the data samples have known outcomes that the user wants to predict. This Knowledge based system [2, 15] gave the following results on the basis of the inputs given. This analysis explored that offenders likely to come from the same age, likely to cooperate and come together for required skills to commit offence, likely to come from the same area.

Supervised Learning 1 (Decision List)

Parameters

Parameters	
Rule evaluation measure	Shannon entropy
Significance level	0.1
Min support of rule	10

Results

Classifier performances

Error rate	0.2979

Values prediction

Value	Recall	1-Precision
High	1	0.2973
Middle	0.5	0.3
Low	0	1

Confusion matrix

	High	Middle	Low	Sum
High	26	0	0	26
Middle	7	7	0	14
Low	4	3	0	7
Sum	37	10	0	47

Classifier characteristics

Data description

Target attribute	OFFENCE LEVEL (3 values)

descriptors

9

Number of rules = 2

Knowledge-based system

Antecedent	Consequent	Distribution
IF AREA in [Border]	OFFENCE LEVEL in [High]	(12; 0; 0)
ELSE IF AGE GROUP in [35-40]	OFFENCE LEVEL in [Middle]	(0; 7; 3)
ELSE (DEFAULT RULE)	OFFENCE LEVEL in [High]	(14; 7; 4)

Table III: Result based on Decision List

IV. CONCLUSION

It has been shown that offenders' information is mined for analysis purpose. The main data mining techniques used in this work are A-priori, Group Characterization and Decision List. Its results have explored various outcomes. One of them showed that offence (crime) was affected by various input parameters such as family background, history of the offender, area, close circle, education level etc. The other thing which was derived in this work is offence level classification. Offence level classification has shown the level of offence that was affected by the various parameters like close circle of the offender, his history, family, education etc. With the use of proper data mining tools on the data source, various queries can be built for statistical analysis or summary information with different case management systems. In addition, it is able to provide ability to execute queries, without having to understand the data specific to a judicial system at district state or at national level. It is able to provide the ability to find a person known to the particular judicial system for all cases in which he was involved.

By way of knowledge discovery of data mining tools, judicial system is able to take appropriate decisions in advance to control the system. As no data warehouse is available so far in the judicial system, hence implementation will take time to implement properly after availability of data warehouse. However, if the benches have vast quantities of data to be mined and the users are familiar with the nature of transactions and expected error patterns, then data mining does provide an efficient means to surface interesting matters. However, it is still a long way from possessing "Artificial Intelligence" to fully automate the mining tasks in judicial system. The future research using data mining is possible however for that purpose data warehouse and limited data surveillance is must.

REFERENCES

[1] Dahbur, K. and Muscarello, T. (2003). Classification system for serial criminal patterns, Artificial Intelligence and Law 11(4) :51–269.

[2] J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] Chen, H., Zeng, D., Atabakhsh, H., Wyzga, W., and Schroeder, J. COPLINK: Managing law enforcement data and knowledge, Communications of the ACM, 46(1), 28-34 (2003).

[4] Klerks, P. The network paradigm applied to criminal organizations: theoretical nitpicking or a relevant doctrine for investigators? Recent developments in the Netherlands. Connections, 24, No. 3, 53-65 (2001).

[5] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann publications, pp. 1-39, 2006.

[6] J. Mena, "Investigative Data Mining for Security and Criminal Detection", Butterworth Heinemann Press, pp. 15-16, 2003.

[7] D.E. Brown, "The regional crime analysis program (RECAP): A Frame work for mining data to catch criminals," in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vol. 3, pp. 2848-2853, 1998.

[8] J.S. de Bruin, T.K. Cocx, W.A. Kosters, J. Laros and J.N. Kok, "Data mining approaches to criminal career analysis," in Proceedings of the Sixth International Conference on Data Mining (ICDM'06), pp.171-177, 2006.

[9] S.V. Nath, "Crime pattern detection using data mining," in Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 41-44,2006.

[10] Brown, D.E and Hagen, S., "Data association methods with application to law enforcement." Decision Support Systems, 34, 2000, 369-378.
[11] Faith Ozgul, Claus Atzenbeck, AhmetCelik, Zeki, Erdem, "Incorporating data Sources and Methodologies for Crime Data Mining," IEEE proceedings, 2011.
[12] Adderley, R. (2004), The use of data mining techniques in operational crime fighting in 2nd Symposium on Intelligence and Security nformatics, ISI- 2004. Tucson, AZ, USA. 3073, pp. 418–425.
[13] Adibi, J., P. Pantel, et al. (2005), 'Report Link Discovery: Issues, Approaches and Applications', (KDD-2005 Workshop - LinkKDD-2005), IGKDD Explorations 7(2), pp. 123-125.
[14] Chen, H., Chung, W., et al. (2004), Crime data mining: a general framework and some examples, in Computer 37(4), pp. 50-56.
[15] Chen, H., J. Schroeder, et al. (2002), 'COPLINK Connect: information and knowledge management for law Enforcement', Decision Support Systems 34, pp. 271-285.