

Crime Analysis Based on Association Rules Using Apriori Algorithm

Mehmet Sevri, Hacer Karacan, and M. Ali Akcayol

Abstract—Data mining is an important technique used in many fields with the purpose of acquiring valuable information from big data. This study aims to reveal the relations between the attributes of independent criminal records. The NIBRS database, which includes criminal records in USA that are recorded in 2013, is used in this study. The association rules created by the Apriori algorithm have been used to extract the relationships between features of criminal records. The experimental results show that the association rules created by the Apriori algorithm are useful for criminal analysis. This study reveals the relationships between the attributes of different criminal records and allows the authorities to establish relationships between new and old incidents.

Index Terms—Crime analysis, association rules, apriori, offender profiling.

I. INTRODUCTION

While crime is a concept with various definitions, generally accepted definition of crime consists of actions which are outside of society's moral values, condemned by the society and requiring legal punishment by a government [1]. Concept of crime is considered to be inside the study fields of different sciences such as sociology, psychology and criminology. Furthermore, advantages of technology and analyzing the components of crime with the support of computer sciences have become important in solving the crime and detecting the offenders [2], [3]. Because criminal cases have become more complicated than they were before, using new and up-to-date approaches in the analysis of cases and detailed detection of similarities between crime records constitute an important problem for criminology. Utilizing big data approach and data mining techniques, big data stacks that include previously recorded crimes can be analyzed in order to reveal the associations and relations between the attributes of records. These approaches can provide important benefits in creating offender profiles, solving criminal cases and taking necessary precautions before the criminal cases [4]. Fig. 1 shows the numbers of convicts sent to correctional facilities in Turkey during the years 2011-2015. The graphic in the figure can be interpreted to say that crime rate increases every year.

Crime analysis has many important effects in actions such as offender detection, attribute extraction from the cases, directing legal investigations, revealing crime trends and

making the decisions of authorities easier. In the event of a case; comparing it to previous cases, creating a offender profile, predicting the hidden aspects of the case can be critically important while solving the case [5]-[7].

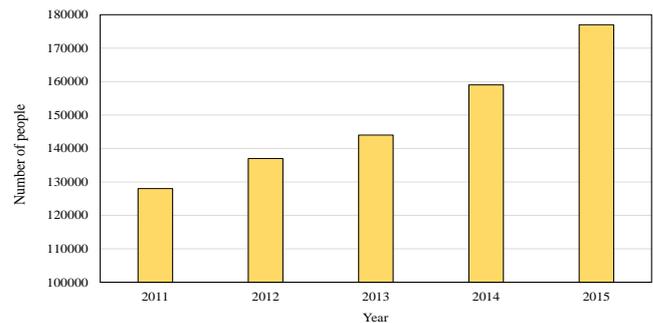


Fig. 1. Numbers of convicts sent to correctional facilities in Turkey.

Although a criminal case is usually instantaneous and unpredictable, if crimes of the same kind are committed by offender(s) of the same profile in a similar way and inside certain geographical zones, there is a possible prevention. Crime analysis can be used to put forward many relations and deductions about crime records containing offender's characteristics, victim's characteristics, case's characteristics, crime weapons, and crime's type of conduct, geographical distribution and cause-effect relations. There are different studies on crime analysis and creating offender profiles. Crime analysis researches based on data mining techniques occasionally use clustering, classification, association rules, sequential pattern mining methods and in addition entity extraction, decision trees, support vector machines, naive bayes, artificial neural networks and social media analysis methods are also used [8], [9]. Some studies about crime analysis and creating offender profiles are as below.

The software developed by Strano uses crime database in order to create offender profile in independent crime cases by utilizing artificial neural networks and data mining techniques. The importance of Strano's study is that it is not only based on the detection of similarities to past cases like traditional methods but also taking into account the database and criminal evidences to create a biographic, motivational and psychological profile for the unknown offender [10].

Chau et al. have developed a new approach by unifying crime data based on police reports and dictionary search based, rule based, statistics based and machine learning based methods. By applying the inputs derived from this approach on the artificial neural network based entity extraction model, they tried to detect the usable objects mentioned in the reports. By applying 36 drug crime related reports randomly chosen from Phoenix police department's reports on this model, they have correctly predicted the relation between offenders and their drugs with 76% and 85.4% precision and 73.4% and

Manuscript received February 10, 2017; revised May 6, 2017.

M. Sevri is with the Informatics Institute, Gazi University Ankara, Turkey (e-mail: mehmetsevri@gazi.edu.tr).

H. Karacan and M. A. Akcayol are with the Computer Engineering Department, Gazi University, Ankara, Turkey (e-mail: hkaracan@gazi.edu.tr, akcayol@gazi.edu.tr).

77.9% recall, respectively [11].

Lin and Brown’s study aims to detect serial offenders by applying association rules technique on the dataset about theft crime. Their study uses Apriori algorithm and it is seen that association rules create fine results on crime data. By their study, it is detected that true prediction rate increases in proportion to the increase in the amount of data in the dataset [12].

In this study, by using crime dataset which includes several different types of crime that are based on real data, the relations between attributes of criminal cases are exposed by association rules. Apriori algorithm is used for creating association rules. The purpose of this study is to predict the unknown characteristics of a specific case such as offender profile, crime weapon, victim profile and geographical zone by taking into consideration the known characteristics of past criminal cases. When the literature is reviewed, most crime profiling studies are carried out for only one crime and criminal profiles. With this study, a large number of different crime events were examined together and it was shown that the strong links between the features can be revealed with the help of technology.

II. CREATING OFFENDER PROFILE USING APRIORI ALGORITHM

A. Dataset and Its Specification

TABLE I: ATTRIBUTES USED IN THE STUDY AND EXAMPLE CASES

Attributes	Case - 1	Case -2
State	CO	KS
Population Group	Cities 2,500-9,999	Cities 25,000-49,999
Incident Date Hour	15	22
Ucr Offense	Burglary/Breaking and Entering	Simple Assault
Offender(S) Suspected Of Using	Not applicable	Not applicable
Location Type	Residence/home	Residence/home
Weapon / Force	Undetermined	Personal Weapons (hands, feet, teeth, etc.)
Type Property Loss	Stolen/Etc.	NA LT 3 records
Property Description	Computer hardware/software	NA LT 3 records
Sex Of Victim	Female	Male
Race Of Victim	White	Black or African American
Ethnicity Of Victim	Unknown/Missing/DNR	Not Hispanic or Latino
Relationship Vic To Of	Not applicable	Victim was Boyfriend/Girlfriend
Age Of Offender	25	35
Sex Of Offender	Female	Female
Race Of Offender	White	Black or African American

In this study, NIBRS (National Incident-Based Reporting

System – 2013) dataset is used, which includes real crime cases in the USA in 2013 recorded by the FBI [13]. The dataset includes 4,998,574 crime cases that belong to 48 different types of crime, which have occurred independently. In the creation of the dataset, missing attributes of cases are shown by negative values and the missing data is disregarded in the scope of this study. Attributes that belong to more than one victim and more than one offender are given in the dataset.

By performing frequency analysis, it is detected that more than 90% of the cases have single victim and single offender. Thus, only the attributes of the first victim and the first offender are used. For the detection of frequent itemsets and the creation of association rules, 16 most significant attributes of a crime case are used. Table I shows the attributes used and their exemplary values.

B. Association Rule Mining and Apriori Algorithm

Association rule mining is an unsupervised data mining technique which is used for the detection of frequent itemsets from the attributes of items in the dataset and revealing the relations between these items. It can be used to predict the unknown attributes of possible future cases by analyzing past cases [14], [15]. The most known field of association rule mining usage is market basket analysis which is a technique frequently used by large corporations to detect the items commonly purchased together and to discover customers’ shopping behaviours. As a result, marketing and promotion strategies for the products can be developed [16].

Apriori algorithm is used for determining frequent itemsets according to minimum support value and setting relationship rules of these items that provide the confidence value. Support value is obtained by dividing the amount of records in which A and B attribute values occur together in a dataset to the total number of records. Support value, showing all transactions in dataset T, is calculated as shown in Equation (1). To create a frequent itemset, the itemset must have a support value that is equal to or greater than the minimum support value. Apriori algorithm works bottom-up, therefore the first pass of the algorithm scans the whole dataset to define the frequent itemsets. In each following pass, n-frequent itemset providing minimum support value is created out of n-1 sets that are created in the previous pass. Confidence value is the ratio of the amount of records in which A and B occur together in the dataset to the amount of records in which A occurs. Confidence value is calculated as shown in Equation (2). An association rule is stated as $A \Rightarrow B$ and the association rules which have minimum support and minimum confidence values are accepted as strong rules [17].

$$Support(A) = \frac{|t \in T; A \subseteq t|}{|T|} \quad (1)$$

$$Confidence(A \Rightarrow B) = \frac{Support(A \cup B)}{Support(A)} \quad (2)$$

C. Method and Materials

The proposed study consists of five steps mainly as

follows: data pre-processing, data encoding, creating transactions in the dataset, creating frequent itemsets which provide minimum support value and creating association rules which provide minimum confidence value out of the created itemsets. Fig. 2 demonstrates the flow chart of the developed application.

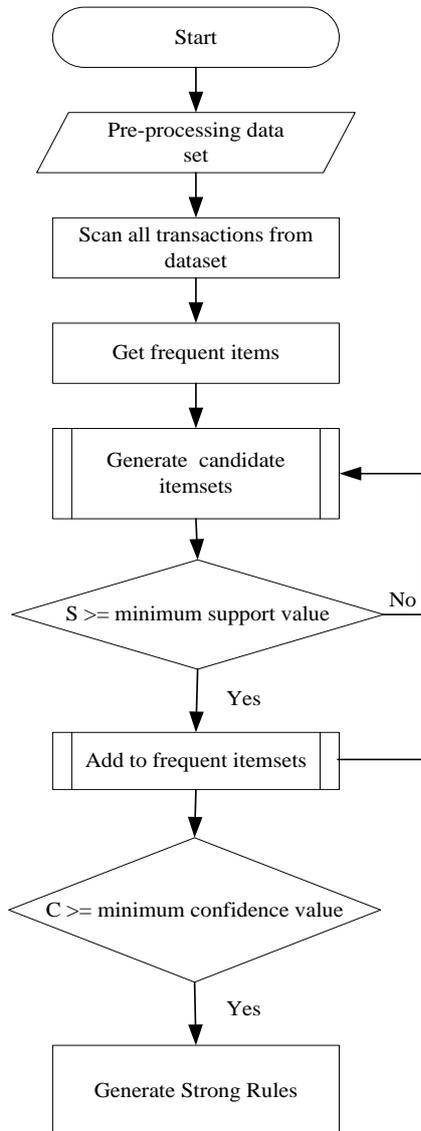


Fig. 2. Flow chart of the application.

The steps in the proposed study are explained in detail below:

- In the data pre-processing step, the attributes in the dataset are examined by frequency analysis method, the insignificant attributes are reduced and the nominal values are converted into numerical values for program optimization.
- To provide legible output in the developed application, the values of the attributes in the dataset have been coded.
- In the first phase of the Apriori algorithm, all transactions included in the dataset are created.
- Items that provide the minimum support value in the dataset have been identified.
- Frequent itemsets have been created from items that provide the minimum support value.

- Using these frequent itemsets, association rules are created that provide the minimum confidence value.

In this study, SPSS is used in the analysis of the dataset and Python programming language is used for applying the Apriori algorithm. Main reason for preferring this programming language is its flexibility and speed when performing big data and calculation operations.

III. RESULTS

In the developed application, minimum support value is set as 0.05 and minimum confidence value is set as 0.6. Association rule mining is performed on nearly 5 million crime cases and as a result, 300 frequent itemsets and 368 strong relations are defined. In addition to the most general association rules created in this study for all crime types, more privatized association rules are possible to be created according to crime type, offender profile, victim profile, geographical location and other attributes.

TABLE II: EXAMPLES OF STRONG ASSOCIATION RULES CREATED

Pre-Rule	Post-Rule	Confidence
Sex of Victim: Female , Race of Offender: White , Ethnicity of Victim: Not Hispanic or Latino	Race of Victim: White , Location Type: Residence/home	0.674
Race of Victim: White , Location Type: Residence/home , Race of Offender: White	Sex of Offender: Male	0.727
Weapon / Force: Personal Weapons (hands, feet, teeth, etc.) , Race of Offender: White	Race of Victim: White , Ucr Offense: Simple Assault	0.780
Sex of Victim: Female , Race of Victim: White , Location Type: Residence/home , Sex of Offender: Male	Race of Offender: White	0.833
Race of Victim: White , Sex of Offender: Female	Race of Offender: White	0.852
Race of Offender: White , Weapon / Force: Personal Weapons (hands, feet, teeth, etc.)	Race of Victim: White	0.917
Sex of Victim: Female , Ethnicity of Victim: Not Hispanic or Latino , Location Type: Residence/home , Race of Offender: White	Race of Victim: White	0.954

Table II shows some of the strong rules which are created by performing association rule mining via Apriori algorithm on 48 different types of crime. These strong association rules are described below:

1. If sex of victim is female and ethnicity of victim is neither Hispanic nor Latino and race of offender is White, then race of victim is White too and crime scene is home.
2. If races of both the victim and the offender is White and crime scene is home, then sex of victim is male.

3. If weapon/force is a personal weapon (hands, feet, teeth, etc.) and race of offender is White, then race of victim is White and type of offense is simple assault.
4. If sex of victim is female and race of victim is White and crime scene is home and sex of offender is male, then race of offender is White too.
5. If race of victim is White and sex of offender is female, then race of offender is White too.
6. If race of offender is White and weapon/force is a personal weapon (hands, feet, teeth, etc.), then race of victim is White too.
7. If sex of victim is female and she is neither Hispanic nor Latino and crime scene is home and race of offender is White, then race of victim is White too.

IV. CONCLUSIONS AND RECOMMENDATIONS

The rapid increase in crime cases has led to the development of new crime analysis techniques. One of the common crime analysis techniques used by law enforcement agencies today is the use of data mining techniques to create an offender profile, a victim profile and unknown characteristics of a newly occurring criminal case, which is committed by similar offender(s) in similar past criminal cases. In this study, strong association rules about crime cases are created by using association rule mining, which is one of data mining techniques. The NIBRS crime database, which consists of 48 different crime types and contains almost 5 million different crime incidents, is used in the study. Strong association rules about the attributes of records in NIBRS have been created via the developed program. This study is important in terms of using real crime data and covering a large number of criminal cases. Another highlight of the work is presenting a way for judicial authorities and law enforcement officers to figure out the undiscovered characteristics of similar criminal cases in the light of criminal investigations they have resolved in the past. This study differs from the others in the field for the reasons such as including many types of crime, examining multiple number of criminal records, using real criminal records and creating profiles in relation with the offender, victim and all the attributes of a case. As a result of the study, it is contemplated that the creation of crime databases has utmost importance for both scientific and forensic reasons. Therefore, creating such databases and enabling them to be accessed by researchers in countries will be beneficial for the enlightenment of criminal investigations and the solution of cases with unknown perpetrators.

REFERENCES

- [1] U. Alacaptan, "Suçun unsurları," *Ankara Üniversitesi Yayınları*, vol. 372, pp. 2-3, 1975.
- [2] M. Orakçı and B. Cıylan, "Prediction of unknown elements in rape cases: Machine learning approach in crime analysis," *International Proceedings of The International Conference on Research in Education & Science*, Muğla/Bodrum, May 19-22, 2016.
- [3] R. F. Sokullu-Akıncı, *Kriminoloji*, Beta Basım Yayım, İstanbul, 2016, ch. 12.
- [4] Tuik, "Ceza infaz kurumu istatistikleri, 2014," *TUİK Haber Bülteni*, vol. 73-2009. 2008, 2014.

- [5] M. Orakçı, B. Cıylan, İ. Kök, and M. Sevri, "Suç analizinde veri madenciliği teknikleri ve makine öğrenmesi algoritmalarının kullanılması," presented at XVIII. Akademik Bilişim Conference, Aydın, February 3-5, 2016.
- [6] D. E. Brown, "The regional crime analysis program (RECAP): A framework for mining data to catch criminals," *International Proceedings of IEEE International Conference on Systems and Cybernetics*, pp. 2848-2853, 1998.
- [7] H. Takçı and Ş. Hayta, "Suç veri madenciliği yardımıyla hırsızlık suçları hakkında kural çıkarımı," *International Proceedings of Elektrik - Elektronik - Bilgisayar ve Biyomedikal Mühendisliği Symposium*, pp. 694-699, Bursa, 2014.
- [8] J. Hosseinkhani, M. Koochakzadeh, S. Keikhaee, and Y. H. Amin, "Detecting suspicion information on the web using crime data mining techniques," *International Journal of Advanced Computer Science and Information Technology (IJACSIT)*, vol. 3, no. 1, pp. 32-41, 2014.
- [9] H. Hassani, X. Huang, E. S. Silva, and M. Ghodsi, "A review of data mining applications in crime. Statistical analysis and data mining," *The ASA Data Science Journal*, vol. 9, no. 3, pp. 139-154, 2016.
- [10] M. Strano, "A neural network applied to criminal psychological profiling: An Italian initiative," *International Journal of Offender Therapy and Comparative Criminology*, vol. 48, no. 4, pp. 495-503, 2004.
- [11] M. Chau, J. J. Xu, and H. Chen, "Extracting meaningful entities from police narrative reports", in *Proc. of the 2002 Annual National Conference on Digital Government Research*, Los Angeles, CA, pp. 1-5, 2002.
- [12] S. Lin, D. E. Brown, "An outlier-based data association method for linking criminal incidents." *Decision Support Systems*, vol. 41, no. 3, pp. 604-615, 2016.
- [13] National Archive of Criminal Justice Data, "National incident-based reporting system, 2013: Extract files," *Inter-university Consortium for Political and Social Research*, 2013.
- [14] K. Leung and C. Leckie, "Unsupervised anomaly detection in network intrusion detection using clusters," in *Proc. of the Twenty-eighth Australasian Computer Society Conference on Computer Science*, pp. 333-342, vol. 38, 2005.
- [15] T. Hastie, R. Tibshirani, and J. Friedman, *Unsupervised Learning. In: The elements of statistical learning*, Springer, New York, 2009, pp. 485-585.
- [16] M. Dhanabhakyan and M. Punithavalli, "A survey on data mining algorithm for market basket analysis," *Global Journal of Computer Science and Technology*, vol. 11, no. 11, 2011
- [17] B. Doğan, B. Erol, and A. Buldu, "Sigortacılık sektöründe müşteri ilişkileri yönetimi için birliktelik kuralı kullanılması," *Marmara Fen Bilimleri Dergisi*, vol. 3, pp. 105-114, 2014.



Mehmet Sevri is currently a Ph.D candidate in information systems and researcher assistant at Gazi University. He has received his bachelor's degree in Computer Engineering from Karadeniz Technical University and master's degree in Information Systems from Gazi University, Turkey in 2011 and 2016 respectively. Sevri's area of research interests are cyber security, data mining.



Hacer Karacan is currently an associate professor in the Department of Computer Engineering at Gazi University. She has received her Msc and Phd degree in Informatics Institute from Middle East Technical University (METU), Turkey in 2007. Karacan's area of research interests are data mining, software engineering.



M. Ali Akcayol is currently a full professor in the Department of Computer Engineering at Gazi University. He has received his MSC and Ph.D degrees in Graduate School of Natural and Applied Sciences from Gazi University, Turkey in 1998 and 2002 respectively. Akcayol's area of research interests are big data analytics, web mining, mobile wireless systems, artificial intelligence, hybrid intelligent systems.