Data Mining Framework for Identifying Crime Patterns Using Unsupervised Learning: Focusing on Assault and Theft

Seunghwan Bang
Department of Industrial and Management Engineering
Pohang University of Science & Technology, Pohang, Republic of Korea
Tel: (+82) 54-279-8243, Email: seunghwanb@postech.ac.kr

Minchul Lee
Department of Industrial and Management Engineering
Pohang University of Science & Technology, Pohang, Republic of Korea
Tel: (+82) 54-279-8243, Email: bonoyap@postech.ac.kr

Bongjun Ji
Department of Industrial and Management Engineering
Pohang University of Science & Technology, Pohang, Republic of Korea
Tel: (+82) 54-279-8243, Email: bongjun_ji@postech.ac.kr

Jeongbin Kim
Department of Industrial and Management Engineering
Pohang University of Science & Technology, Pohang, Republic of Korea
Tel: (+82) 54-279-8243, Email: kjb93814@postech.ac.kr

Hyunseop Park
Department of Industrial and Management Engineering
Pohang University of Science & Technology, Pohang, Republic of Korea
Tel: (+82) 54-279-8243, Email: terrywkd2@postech.ac.kr

Hyunbo Cho †
Department of Industrial and Management Engineering
Pohang University of Science & Technology, Pohang, Republic of Korea
Tel: (+82) 54-279-2204, Email: hcho@postech.ac.kr

Abstract. Data mining can be used to identify crime patterns for the purpose of predicting or preventing crime and to help criminal investigators or crime analysts to focus on valuable tasks. Major challenges involved in data mining include identifying the meaning of crime patterns, and analyzing crime data accurately and efficiently. In this paper, we compare the traditional review process and the data mining process of crime reports, and show how data mining is related to existing criminology. We also present a data-driven method than can identify crime patterns and be used to support crime pattern analysis. The framework proposed in this paper can identify crime patterns in categorical datasets by applying a clustering algorithm. Two new non-metric similarity measures are proposed which give high or low weights to non-shared uncommon attributes. In experiments, the method identified common and uncommon patterns in assault and theft data. The patterns can be used to predict, to prevent crime and to help law enforcement personal to identify meaningful crime patterns.

Keywords: data mining, crime pattern, unsupervised learning, clustering, criminology

1. INTRODUCTION

A major challenge facing all law-enforcement and investigative agencies is to analyze the high volume of crime data accurately and efficiently (Chen et al., 2004). Crime pattern analysis for knowledge discovery from criminal events or records is useful to increase the predictive accuracy (Nath, 2006). Also, it is essential to predict crime occurrences or features. Methods of crime pattern analysis have been developed to prevent crime (Chen et al., 2004; Wang, 2014)), and a system to predict crime patterns for preemptive crime prevention has been developed Laleh and M. Abdollahi Azgomi, 2010). Another system has been developed to support police investigations in several crime areas (Brahan et al., 1998).

The process of extracting crime patterns is a topic of interest to criminal investigators, analysts, police, and public prosecutors. Specific information such as characteristics of uncommon crime patterns may be difficult to find, but by providing meaningful insights these patterns will be helpful even if extracting accurate crime patterns is impossible. If a data-driven approach can provide characteristics of common crime patterns or insights that can be used to identify specific patterns, it can assist criminal investigators (Adderley et al., 2007) and reduce the time wasted while reviewing crime reports (Laleh and Abdollahi Azgomi, 2010).

Data mining is a powerful technique that enables extraction of interesting patterns. Various data mining techniques can be used to identify crime patterns, and use of an unsupervised algorithm is a representative technique to extract or identify patterns (Gutierrez-Rodriguez et al., 2015). Clustering one of unsupervised learning is a useful technique to identify groups of data points within a single group that have similar elements, and to distinguish among various different groups. Therefore, clustering can identify groups of criminal events that have similar characteristics. Especially, clustering can be used to identify crime patterns but comparing offenders' profiles, victims' profiles, behavioral characteristics, demography and spatio-temporal information. Therefore, we focus on using a clustering algorithm to group criminal events that have high similarity. Also, we focus on the similarity measures because choice of an appropriate similarity measure must account for the semantics and interrelations among attributes of information (Dos Santos and Zarate, 2015).

This paper introduced the data mining approach in general terms. We would like to show that data-driven approach to identifying crime patterns can reveal meaningful results to crime investigators or analysts. The paper proceeds as follows. In section 2, we review problems related to crime data mining. In section 3 we describe the proposed framework, and in section 4 present the results of experiments. Finally we summarize the whole paper, provide some conclusions with experiment results, and suggest directions for future research.

2. APPLICABILITY OF DATA MINING TO CRIME

Traditionally, crime pattern analysts use criminal event reports, analysts' experience, intuition and knowledge to identify crime patterns [Fig. 1]. To identify crime patterns, the analysts gather all incident reports, then conduct an initial scan of them and remove incidents that are unlikely to ever exhibit any patterns. Then the analysts read the reports and pay careful attention to the Modus Operandi (M.O.), motivation, and behavioral factors. The knowledge obtained can be used to identify new patterns with strong M.O. commonalities or signatures.

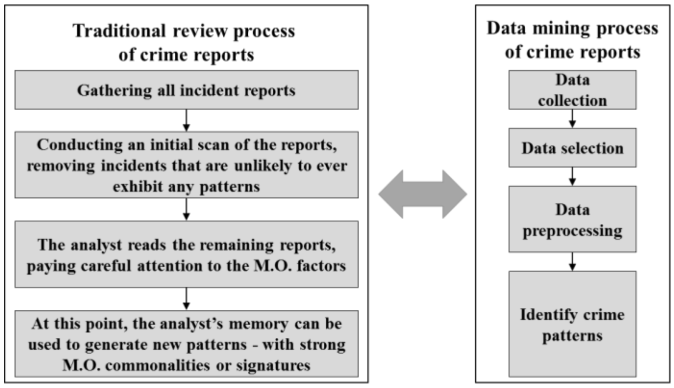

Figure 1: Comparison of Traditional Review Process and Data Mining Process of Crime Reports

The process in data mining looks similar to this preliminary process of examining crime reports. Data mining is a core step of the Knowledge Discovery in Database (KDD) process (Gullo, 2015); "Data mining" and "KDD" are often treated as synonyms (Jiawei and Kamber, 2001). KDD can be divided into five steps: selection, preprocessing, transformation, data mining and interpretation/evaluation. When comparing these five steps with the traditional review process, first step of traditional review process is data collection, which is a preparation before KDD; the second step is selection, processing and transformation for data mining. The third and fourth steps of the traditional review process are pattern identification, which is a core function of data mining.

This comparison demonstrates that the traditional process to identify crime patterns is similar to data mining. The traditional process is a knowledge-based method, whereas data mining is a data-driven method. Therefore, data mining can provide complementary information to the result of the knowledge-based method, and applying data mining to identify crime patterns can be a good way to support crime analysts.

Understanding the offenders' behaviors is important when analyzing crime patterns (Cohen and Felson, 1979;

Brantingham and Brantingham, 1981). According to environmental criminology, criminal behavior is significantly influenced by the environment, and all behavior is a person-situation interaction (Wortley and L. Mazerolle, 2001). To understand criminal events, analysts can use several theories to account for environmental criminology and environmental elements [Table 1]. Routine Activity Theory examines variations in crime rates in terms of social trends. Crime Pattern Theory explains crime patterns at neighborhood and street levels. Awareness Theory considers the behavior of offenders in connection with spatial elements.

Table 1: Environmental elements of each theory

Theory	Environmental Elements
Routine Activity Theory	- Absence of capable guardian - Motivated Offender - Suitable Target
Crime Pattern Theory	- Nodes - Paths - Edges
Awareness Theory	- Victim - Offender - Geo-temporal - Legal

Crime patterns or behaviors of offenders are influenced by environmental elements including demographic and spatio-temporal ones; therefore crime patterns can be defined according to the elements of each theory. If the environmental elements are similar, features of crime are can be said to be similar and to show certain patterns. If data mining techniques use these elements, the result of data mining should be similar to crime patterns identified based on environmental criminology. This linkage between theory and the outcomes of data mining implies that data mining may be a useful tool to identify crime patterns.

Like this, the application of data mining techniques to analyze crime patterns in the review process is reasonable in terms of environment criminology. However, even though the data mining techniques can be applied to crime pattern analysis, any procedures for data mining process have not been shown. Therefore, not simply apply only to try analytical techniques, systematic analysis procedure is needed

3. PROPOSED FRAMEWORK

The procedure from data collection to data analysis for crime pattern analysis is required. In this study, we propose an appropriate data mining framework of the crime pattern analysis procedure.

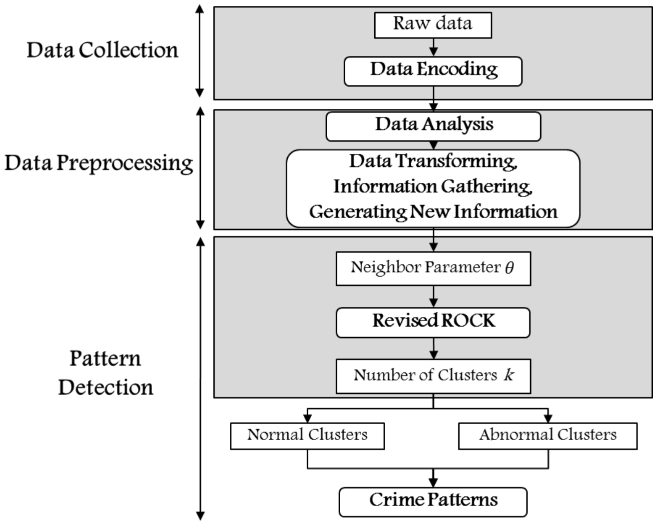

Figure 2: Data Mining Framework for Identifying Crime Patterns

3.1 Data Collection

Data collection is important in any type of study. If t he data are valueless or inaccurate, the algorithms derive u seless or inaccurate information as a result, no matter how good the algorithms are. Generally, records of criminal eve nts are written in narrative form. Thus, the records must b e converted to a form that is suitable for mining. Also, va riables and data type (e.g., nominal, ordinal, and interval) must be defined appropriately for collecting data without l oss of information. If the number of variables is unnecessa rily high, it can cause collinearity between variables. How ever, if the number of variables is not sufficient, it can ca use the loss of information. Therefore, it is important to d etermine the right level of the variables.

3.2 Data Preprocessing

Understanding the nature of data and doing proper data analysis should take precedence over data preprocessing. Crime data have the following characteristics.

1) Multiple data sources (diversity of criminal event writers): Different types, regions and intuitions of people reflected in the criminal records. To collect all data without loss, similar variables can be used even if they are not independent of each other.

2) Small amount of data compared to the number of variables: The right to personal privacy means that some information is difficult to obtain, while various variables.

3) Missing attribute values: People write records manually, so essential information can be omitted.

4) Noisy or irrelevant data: The choice of which useful information to record is not easy. Thus, unrelated information can be included.
5) Non-metric data: Criminal events have different attributes. Time, place and such attributes are always recorded, but attributes such as behavior patterns are different for each event.

Appropriate data preprocessing techniques should be chosen, depending on the purpose of analysis and the nature of the data. Data filtering, ordering and editing can be used to make decision about what to do with noisy or irrelevant data and with missing values. Principal component analysis is suitable for selecting appropriate attributes by checking the dependency among variables in a set of data attributes. It can be used to solve problems caused by dependency due to use of multiple data sources.

3.3 Crime pattern Identification

The main objective of this phase is to use the preprocessed data to identify crime patterns. The most preprocessed data are in categorical form because criminal events are recorded in narrative form, and the data are non-metric. Thus, we adopted a categorical clustering algorithm that uses non-metric similarity to identify crime patterns. When identifying crime patterns, a non-fixed number k of clusters can be more useful than using a fixed number k. In contrast, because a type of crime can be considered as a single pattern of crime, specific patterns can be derived by dividing this type of crime. In this case, hierarchical clustering algorithms are more suitable than non-hierarchical clustering algorithms.

ROCK (Guha et al., 1999) is a robust hierarchical clustering algorithm based on the notion of links. Also, it is suitable for analysis of categorical data, and it uses non-metric similarity measure. For non-metric data, a possible definition is the Jaccard coefficient (Duda and Hart, 1973)

$$sim(E_i, E_j) = \frac{|E_i \cap E_j|}{|E_i \cup E_j|} \quad (1)$$

that describes the similarity between the two events E_i and E_j. However, the Jaccard coefficient assigns the same importance to all attributes, and therefore manner cannot capture the importance of each categorical attribute. If criminal events can be classified based on specific attributes, these attributes are important factors to identify crime patterns. Therefore, a similarity measure should consider the importance of attributes.

Several similarity measures consider the importance of attributes (Lin, 1998; Sparck Jones, 1972; Goodall, 1966). These measures give the weight according to the values. However, identification of crime patterns requires a non-metric similarity measure that weights attributes. QROCK (a quick version of the ROCK algorithm for clustering of categorical data) uses a non-metric similarity measure that assigns weights to attributes (Dutta, 2005). The main idea behind the similarity measure is that the distance between two events on an attribute that has two values should differ from an attribute that has more than two values. However, this similarity measure uses predefined attribute values to assign importance to each categorical attribute; consequently, this measure cannot easily reflect the characteristics of the collected records.

In contrast, if a similarity measure considers the frequency of collected records, the task of representing the characteristics of attributes becomes easy. The main idea behind a similarity measure is that if two records differ on a common attribute, then the similarity between them is higher than between two other records that differ on an uncommon attribute.

For this, we proposed two weighted similarity measures as follows.

Similarity measure 1:

$$sim(E_i, E_j) = \frac{|E_i \cap E_j|}{|E_i \cap E_j| + \sum_{k \notin E_i \cap E_j} \left(1 - \frac{n_k}{N}\right)} \quad (2)$$

Similarity measure 2:

$$sim(E_i, E_j) = \frac{|E_i \cap E_j|}{|E_i \cap E_j| + 2\sum_{k \notin E_i \cap E_j} \left(1 - \frac{n_k}{N}\right)} \quad (3)$$

where n_k is the number of records having kth attribute, and N is the total number of records.

Similarity measure 1 gives weights to attributes that are not part of the intersection. Similarity measure 2, it gives more weights on attributes than does similarity measure 1. Similarity measure 1 is always higher than Jaccard coefficient, and similarity measure 2 is higher than Jaccard coefficient if n_k > N/2, and lower than Jaccard coefficient otherwise. We used goodness measure $g(C_i, C_j)$ for merging clusters C_i, C_j (Dutta, 2005):

Goodness: $g(C_I, C_j)$

$$= \frac{link(C_i, C_j)}{[(n_i + n_j)^{1+2f(\theta)} - n_i^{1+2f(\theta)} - n_i^{1+2f(\theta)}]} \quad (4)$$

where link(C_i, C_j) is the sum of cross links between cluster tuples in C_i and C_j, and $f(\theta) = (1+\theta)/(1-\theta)$, n_i and n_j are the sizes of the clusters C_i and C_j.

The proposed similarity measures are calculated in several steps [Fig. 3]. The clustering algorithm begins by computing neighbor lists using a proposed similarity measure and θ. Then, it computes the number of links between data points. A local heap $q(C_i)$ and a global heap Q for each clusterC_j. are maintained during the execution of the algorithm. The local heap $q(C_i)$ contains each C_i with max $g(C_i, C_j)$ and the global heap Q contains each cluster C_i with max $g(C_i, C_j)$. The merging process is continued until a specified number of cluster k remain or the number of links between the clusters is zero.

```
Clustering Algorithm:
Input: A set of D of data-points
    The similarity threshold θ
begin
    compute nbrlist[i] for every point i ∈ D using θ
    compute link[i, j] for all i, j ∈ D
    for each x ∈ D
        build local heap q[x]
    build global heap Q
while size(Q) > k do {
    u = extract_max(Q)
    v = max(q[u])
    delete(Q, v)
    w = merge(u, v)
    for each x ∈ q[u] ∪ q[v] {
        link[x, w] = link[x, u] + link[x, v]
        delete(q[x], u)
        delete(q[x], v)
        insert(q[x], w, g[x, w])
        insert(q[w], x, g[x, w])
        update(Q, x, q[x])
    }
    insert(Q, w, q[w])
}
end
```

Figure 3: Clustering algorithm

A similarity measure and θ determine the number of links and hence affect the final set of clusters (Goodall, 1966). In the absence of a specific value of k, the algorithm terminates naturally when the number of links is zero, and the point when the algorithm terminates naturally is differ depending on the similarity measure and θ. Thus, the most appropriate similarity measure and θ must be determined.

After the algorithm terminates, the number of clusters should be determined. To define the number of clusters, the number of remain clusters and goodness change are used. By considering the goodness change, the range of the appropriate number of clusters could be determined. When the clustering algorithm is naturally terminated, we represent the result of clustering as two types; common patterns and uncommon patterns. Common patterns refer to well grouped records that indicate representative crime patterns; uncommon patterns refer to uncommon records or outliers. However, uncommon patterns can be regarded specific crime patterns and should not be ignored.

4. EXPERIMENTS

In this section we examine our framework with real crime data that are records of crime that occurred in Seoul, Republic of Korea. To examine the framework, we collected crime data, and then preprocessed it considering characteristics of crime. The resulting dataset was used to identify crime patterns.

Data were collected from reports and documentaries of the Seoul Probation Office; data include information about 1,384 offenders under probation and parole in 2013, who had committed assault (n = 629) or theft (n = 755). The original data from the Seoul Probation Office were based on official reports and documentaries such as pre-sentencing investigation reports and judgment papers from courts. Therefore, key variables for data mining were selected by consulting five criminologists, and the process of selecting variables was implemented using standard definitional operationalization and consistent criteria for coding the variables. Crucial variables for analysis include fundamental demographic variables, targeted victim selection process, and criminal behavior.

After gathering and integrating data in the database, the crime records still contain missing values, unnecessary value and records recorded in inappropriate form. Before data preparation, cleansing and transformation, we must understand the nature of data and to identify the important attributes. If new or modified features are required according to the result of analysis, we defined new features based on criminology. Some of the variables considered, such as information related to the victim selection process and criminal behavior were not always included in records. We enter ‘No’ to distinguish non-existent from incomplete values. However, when using data mining to identify crime patterns, these values have no meaning and should be excluded. Thus, we deleted them and used only meaningful information.

The goal of this paper is to show how to identify crime patterns. The most important two components in the framework are similarity measures and neighbor parameter θ. In this section, we present the results of clustering algorithm using the proposed similarity measures. ROCK is the original algorithm; ROCK 1 used similarity measure 1, and ROCK 2.

When the clustering algorithm was naturally terminated, we defined crime patterns. The number of clusters k was defined using the total number of records and the level at which the process naturally terminated. A naturally terminated level increases by one when two clusters are joined. When clustering algorithm is naturally terminated and final k = 1, then naturally terminated level = total number of records +1. For example, if the total number of records is 622, the clustering algorithms is naturally terminated when level = 623. There was no significant change in the goodness before the clustering algorithm was naturally terminated, we defined the number of clusters k as

$$k = \text{Total number of records} - \mathit{Round\ off}\ (\text{natually terminated level} \times 0.9\ )\ \mathit{to\ the\ nearest\ one}\ + 2 \qquad (5)$$

5. DISCUSSION

The final clusters can divided into common and uncommon patterns. We defined a cluster as uncommon if the number of records assigned to it was one, and as common otherwise. For both assault [Table 2] ad theft [Table 3] the numbers of uncommon clusters increased as θ increased, and the number of common clusters was high at low and high θ, but low at intermediate θ.

At a given θ, theft generally had fewer uncommon patterns than did assault. This difference means that criminal events of theft are more similar than events of assault, and the crime patterns of theft are more consistent than are those of assault. In addition, ROCK generated more uncommon clusters than did ROCK 1 but similar with did ROCK 2. These results indicate that ROCK 1 is an appropriate clustering algorithm to increase overall similarity by giving weights to uncommon attributes, and that ROCK 2 is an appropriate clustering algorithm to decrease overall similarity and increase difference between uncommon attributes that are not included in the intersection. In other words, these results mean that ROCK 1 is appropriate to identifying common patterns, and ROCK 2 is appropriate to identifying uncommon patterns.

One of the aims of this study was to determine how well uncommon attributes clustered. Therefore, we evaluated clustering algorithms by considering the frequency of clusters that contained uncommon attributes rather than by evaluating the clustering accuracy. We selected the most five uncommon attributes of assault and theft. To ensure that the clustering results were obtained using the same conditions, we set $\theta = 0.3$, because all clustering algorithms terminated naturally when all records were clustered in one cluster and the proportions of normal clusters and uncommon clusters were distributed reasonably. We set the total number of clusters $k = 38$.

The uncommon attributes of assault were ‘Meal or Drinking Before Crime (Att. 1)’, ‘Play or Amusement Before Crime (Att. 2)’, ‘Intoxication of Offender (Att. 3)’, ‘Intoxication of Victims (Att. 4)’ and ‘Enticing (Att. 5)’. The uncommon attributes of theft were ‘Meal or Drinking Before Crime’, ‘Play or Amusement Before Crime’, ‘Intoxication of Offender’, ‘Intoxication of Victims’ and ‘The Method of Approach or Trail toward Victim (Att. 6)’.

Considering both assault [Table 4] and theft [Table 5], ROCK 1 identified more common patterns than did ROCK and ROCK 2, which identified similar numbers of common patterns. Uncommon patterns include only one observation, so the number of observations classified into normal patterns can be determined by subtracting the number of uncommon patterns from the number of observations (in parentheses) of that offense. A low number of uncommon patterns means that uncommon attributes are concentrated. These results mean that the proposed similarity measure 2 is an appropriate measure to group uncommon attributes.

6. CONCLUSIONS

The main contributions of this paper are the presentation of applicability of data mining for crime data, and the proposal of data mining method with similarity measures to identify crime patterns for categorical and non-metric data. We showed the similarity between the traditional review process and data mining process, and the meaning of data mining results. In addition, we defined the characteristics of crime data for crime pattern identification. Finally, we identified common patterns and outlier patterns to support review process of crime reports.

In the experiments, we identified common patterns and uncommon patterns. When comparing the clustering results of Jaccard coefficient and the two proposed similarity measures, proposed similarity measure 1 increased the overall similarities compared to Jaccard coefficient by giving the weights to uncommon attributes. Also, proposed similarity measure 2 gives the weights to uncommon attributes, it decreases overall similarities compared to the Jaccard coefficient. We use similarity measures 1 and 2 to identify crime patterns of assault and theft. These similarity measures weight uncommon attributes differently, and neither is consistently appropriate for all crime data. However, similarity measure 2 is more appropriate than similarity measure 1 at grouping uncommon attributes. When comparing assault and theft, crime records of theft are more similar than records of assault. This is consistent in criminology that theft is a premeditated crime and that assault is a spontaneous one.

In general, the task of identifying appropriate crime patterns in an unsupervised manner is difficult, because the appropriate level at which to define a “crime patterns” is not easily defined; i.e., the level of neighbor parameter θ and the number of clusters k cannot be defined easily. Normal clusters can be regarded as common patterns, and abnormal clusters can be regarded as uncommon patterns or outliers. However, in the experiments we could not distinguish abnormal clusters from outliers. To clarify the difference, the appropriate levels of θ and k should be defined, but this is not an easy task.

As limitations, we could not show that the proposed framework groups records into correct clusters because the records of criminal events do not have crime pattern categories. Also, we proposed two similarity measures and we gave the number 2 as a weight in similarity measure 2. However, we could not guarantee that the number 2 is an optimal.

In future study, the problem of distinguishing abnormal clusters and from outliers should be considered. Therefore, a method that incorporates crime analysts’ knowledge to distinguish abnormal clusters and outliers is needed. In addition, a more effort needs to find an optimal weight for the similarity measure 2 Also, to use the crime patterns for crime prediction and prevention, we should be able to determine the characteristics of crime patterns in detail.

ACKNOWLEDGEMENT

This research was supported by the IT R&D program of MSIP/NIPA [10047146, Real-time Crime Prediction and Prevention System based on Entropy-Filtering Predictive Analytics of Integrated Information such as Crime-Inducing Environment, Behavior Pattern, and Psychological Information].

Table 2: Numbers of common and uncommon patterns depending on θ (assault)

	ROCK		ROCK 1		ROCK 2	
	Common	Uncommon	Common	Uncommon	Common	Uncommon
0.1	63	0	63	0	63	0
0.2	63	0	63	0	2	61
0.3	22	41	48	15	18	45
0.4	1	62	1	62	5	58
0.5	1	63	1	62	7	62
0.6	8	62	1	62	39	59
0.7	47	100	4	69	59	178
0.8	65	293	48	130	76	354
0.9	47	493	66	407	44	505

Table 3: Numbers of common and uncommon patterns depending on θ (theft)

	ROCK		ROCK 1		ROCK 2	
	Common	Uncommon	Common	Uncommon	Common	Uncommon
0.1	74	0	74	0	74	0
0.2	74	0	74	0	70	4
0.3	31	43	62	12	15	59
0.4	1	73	1	73	1	73
0.5	1	73	1	73	1	79
0.6	1	78	1	73	6	96
0.7	8	105	1	85	41	129
0.8	59	203	13	123	82	253
0.9	87	450	77	290	86	466

Table 4: Frequency of patterns containing uncommon attributes (assault), (Number in parenthesis in attribute number column means the total number of patterns containing the attribute)

	ROCK		ROCK 1		ROCK 2	
	Common	Uncommon	Common	Uncommon	Common	Uncommon
Att. 1 (54)	6	2	16	0	6	3
Att. 2 (32)	4	3	15	2	7	3
Att. 3 (122)	6	7	18	3	9	6
Att. 4 (67)	5	4	15	2	6	7
Att. 5 (58)	7	3	16	4	6	4

Table 5: Frequency of patterns containing uncommon attributes. (theft), (Number in parenthesis in attribute number column means the total number of patterns containing the attribute)

	ROCK		ROCK 1		ROCK 2	
	Common	Uncommon	Common	Uncommon	Common	Uncommon
Att. 1 (31)	6	6	14	1	3	10
Att. 2 (28)	7	4	11	1	3	9
Att. 3 (32)	5	6	16	2	3	10
Att. 4 (64)	6	6	18	1	3	10
Att. 6 (34)	4	5	16	1	3	8

REFERENCES

Borg, A., Boldt, M., Lavesson, N., Melander, U., & Boeva, V. (2014). Detecting serial residential burglaries using clustering. *Expert Systems with Applications*, 41, 5252-5266.

Gutierrez-Rodríguez, A. E., Martínez-Trinidad, J. F., García-Borroto, M., and Carrasco-Ochoa, J. A. (2015). Mining patterns for clustering on numerical datasets using unsupervised decision trees. *Knowledge-Based Systems*, 82, 70-79.

Buczak, A. L. and Gifford, C. M. (2010), Fuzzy association rule mining for community crime pattern discovery, *ACM SIGKDD Workshop on Intelligence and Security* Informatics, 1-10

Ordonez, C. (2011), Data set preprocessing and transformation in a database system, *Intelligent Data Analysis*, 15, 613-631

Lin, D. (1998), An information-theoretic definition of similarity, *Proceedings of the 15th International Conference on Machine Learning*, 296-304

Goodall, D.W. (1966), A new similarity index based on probability, *Biometrics*, 882-907

Famili, F., Shen, W.M., Weber, R., and Simoudis, E. (1997), Data Pre-processing and Intelligent Data Analysis, *International Journal on Intelligent Data Analysis*, 1, 1-28

Gullo, F. (2015), From patterns in data to knowledge discovery: What data mining can do, *Physics Procedia*, 62, 18-22

Oatley, G. and Eward, B. (2011), Data mining and crime analysis, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1, 147-153

Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y., and Chau, M. (2004), Crime data mining: A general framework and some examples, *IEEE Computer Society*, 37, 50-56

Jiawei, H. and Kamber, M. (2001), Data mining: Concepts and techniques, 3rd ed. Morgan Kaufmann, *San Francisco, CA*

Lee, I. and Phillips, P. (2008), Urban crime analysis through areal categorized multivariate associations mining , *Applied Artificial Intelligence*, 22, 483-499

Brahan, J.W., Lam, K.P., Chan, H., and Leung, W. (1998), AICAMS: Artificial intelligence crime analysis and management system, *Knowledge-Based Systems*, 11, 355-361

Dahbur, K. and Muscarello, T. (2013), Classification system for serial criminal Patterns, *Artificial Intelligence and Law*, 11, 251-269

Sparck Jones, K. (1972), A statistical interpretation of term specificity and its application in retrieval, *Journal of Documentation*, 28, 11-21

Cohen, L. and Felson, M. (1979), Social change and crime rate trends: A route activity approach, *American Sociological Review*, 44, 588-608

Craglia, M., Haining, R., and Wiles, P. (2000), A comparative evaluation of approaches to urban crime pattern analysis, *Urban Studies*, 37, 711-729

Dutta, M., Mahanta, A.K., and Pujar, A.K. (2005), QROCK: A Quick version of the ROCK algorithm for clustering of categorical data, *Pattern Recognition Letters*, 26, 2364-2373

Laleh, N. and Abdollahi, Azgomi. M. (2010), A hybrid fraud scoring and spike detection techniques in streaming data, *Intelligent Data Analysis*, 14, 773-800

Brantingham, P.J. and Brantingham, P.L. (1981), Environmental criminology, *Beverly Hills, CA*,

Bsoul, Q., Salim, J., and Zakaria, L.Q. (2013), An intelligent document clustering approach to detect crime patterns, *Procedia Technology*, 11, 1181-1187

Adderley, R., Townsley, M. and Bond, J. (2007), Use of data mining techniques to model crime scene investigator performance, *Knowledge-Based Systems*, 20(2), 170-176

Wortley, R. and Mazerolle, L. (2001), Environmental Criminology and Crime Analysis 2rd ed. *Routledge, London and New York*

Duda, R.O. and Hart. P.E. (1973), Pattern classification and scene analysis, *Wiley, New York*

Guha, S., Rastogi, R., and Shim, K. (1999), ROCK: A robust clustering algorithm for categorical attributes, *Proceedings of 15th International Conference on Data Engineering*, 512-521

Zhang, S., Zhang, C., and Yang, Q., Data preparation for data mining. (2003), *Applied Artificial Intelligence*, 17, 375-381

Nath, S.V. (2006), Crime pattern detection using data mining, IEEE/WIC/ACM *International Conference on Web Intelligence and Intelligent Agent Technology Workshops*, 41-44

Wang, T., Rudin, C., Wagner, D., and Sevieri, R. (2014), Learning to detect patterns of crime, *Machine Learning and Knowledge Discovery in Databases*

Dos Santos, T.R. and Zarate, L.E. (2015), Categorical data clustering: What similarity measure to recommend? *Expert Systems with Applications*,42, 1247-1260