

# A Rough-fuzzy C-means Using Information Entropy for Discretized Violent Crimes Data

Chao Yang<sup>†\*</sup>, Shiyuan Che<sup>†</sup>, Xueting Cao<sup>\*</sup>, Yeqing Sun<sup>\*</sup>, Ajith Abraham<sup>‡§</sup>

<sup>†</sup>*School of Information Science and Technology, Dalian Maritime University, Dalian, China*

<sup>\*</sup>*Institute of Environmental Systems Biology, Dalian Maritime University, Dalian, China*

<sup>‡</sup>*Machine Intelligence Research Labs (MIR Labs), WA, USA*

<sup>§</sup>*IT4Innovations, VSB-Technical University of Ostrava, Czech Republic*

*yqsun@dlnu.edu.cn, ajith.abraham@ieee.org*

**Abstract**—This paper presents the factor clustering analysis for violent crimes. The efficiency of Rough-fuzzy C-means algorithm is affected by the numbers of clusters, and not all centroids are beneficial. The analyzing of violent crime data does not need human intervention for impartiality. The information entropy is a helpful tool for resolving those issues. In this paper, a novel discrete Rough-fuzzy C-means based on information entropy algorithm (DRFCMI) is proposed, which can obtain typical conclusions objectively. Experimental results illustrate that our proposed method is efficient.

**Keywords**-Fuzzy C-means, Rough set, Information entropy, Discretization, Violent crimes

## I. INTRODUCTION

Analyzing massive violent crime data using data mining techniques has become an important and meaningful issue for the public security system. This paper developed a framework for mining the data by analyzing the offenders' family background information, psychological factor, the information of MAOA and DRD4 genes and other multi-dimensional factors, with the cooperation from Chongqing Municipal Public Security Bureau and Institute of Environmental Systems Biology, Dalian Maritime University. Mining the characteristics of the data, which can guide police officers to carry out the purposeful education for the crime and trace the influence factors plays a role in governance and crime prevention.

Clustering is a common method used in data mining, which has been widely used in real life. It is a process that divides physical or abstract objects into the collection of similar objects [1]. Dividing fuzzy clustering can get the degree of possibility, which belongs to various categories of sample data. It establishes the possible description of the category [2]. Therefore, it can reflect the real world objectively. Fuzzy C-means is the popular used method.

Classical set theory is promoted by rough set [3] and fuzzy set in dealing with uncertainty and imprecision issues, so they are highly complementary [4]. Lingras [5] introduced rough set theory for clustering. Mitra [6] introduced fuzzy clustering membership to rough C-means and proposed a rough-fuzzy C-means clustering algorithm. Rough C-means

clustering algorithm can deal with uncertain objects and it has been widely applied to bioinformatics and other fields [7]. There are many factors of violent crime in our study and the outcomes of attribute reduction comes from our previous work. Multi-knowledge extraction using swarm - rough algorithm (MESR) [8] has analyzed violent crime data sets. We are facing a problem of discretization of the continuous attributes. A good discretization method not only can speed up the process of running speed, but also achieve better results. This article proposes a discrete rough-fuzzy c-means algorithm based on the information entropy algorithm. The reduced attributes are from our previous work. First, the continuous attributes are discretized. Under different numbers of clusters, the results are acquired by the cluster algorithm. Finally some significative classes are achieved using information entropy. The detailed descriptions of classes are illustrated in this article.

## II. METHODOLOGY

Fuzzy C-means, rough set and information entropy are combined to enhance the ability to deal with the uncertainty, vagueness, and incompleteness. Fuzzy c-means is used for resolving overlapping data. Rough set is good at handling uncertainty and information entropy is used for evaluation criterion.

### A. Information Entropy

Shannon [9] pointed out that entropy was a quantitative measure of unpredictability or information.

The probability of  $x_i$  is  $p_i$  in a probability system of  $n$  objects  $\{x_1, x_2, \dots, x_n\}$ , i.e.  $X = \{x_1, x_2, \dots, x_n\}$ . When event  $x_i$  occurs, its self-information is

$$I(x_i) = \log_2 \frac{1}{p_i} \quad (1)$$

and unit is bit. The average information entropy of the system is

$$H(X) = E[I(X)] = - \sum_{i=1}^n p_i \log_2 p_i \quad (2)$$

The basic role of information is to eliminate the uncertainty of system. Information entropy is a quantitative form of information. If a system is a higher uncertainty, its  $H(X)$  will be higher.

### B. Rough Set

Rough set [3] based upon indiscernability relation can effectively analyze much incomplete and uncertain information. Using the concepts of lower and upper approximations in Rough set theory, implicit knowledge from the information system may be extracted and expressed.

The classical rough set theory is an approximation space, which is a pair  $\langle U, R \rangle$ , where  $U$  is the universe of discourse, and  $R$  is an equivalence relation on  $U$ . Given an arbitrary set  $X \subseteq U$ , it may be impossible to precisely describe  $X$  with the equivalence classes in  $R$ . In this case, one may characterize  $X$  by a pair of lower and upper approximations.

$$\underline{R}(X) = \{x_i \in U : [x_i]_R \subseteq X\} \quad (3)$$

$$\overline{R}(X) = \{x_i \in U : [x_i]_R \cap X \neq \emptyset\} \quad (4)$$

$[x_i]_R$  means the equivalence class  $R$  including  $x_i \in U$ .

The lower approximation  $\underline{R}(X)$  is the set of objects which certainly belong to  $X$ , while the upper approximation  $\overline{R}(X)$  is the set of objects which intersect with  $X$ . There are two basic concepts of  $X$  in the approximation space.

### C. Fuzzy C-Means

Fuzzy C-means (FCM) [10] derives from Hard C-means (HCM) [11]. Because its performance is better than HCM, FCM has become the most well-known and useful method in cluster analysis.

Create a set of  $n$  objects,  $X = \{x_1, \dots, x_j, \dots, x_n\}$ , and a set of  $c$  centroids,  $V = \{v_1, \dots, v_i, \dots, v_c\}$ . The objective function for FCM aims at minimizing Equ. (5).

$$J_{FCM} = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^{m_1} \|x_j - v_i\|^2 \quad (5)$$

where  $\mu_{ij} \in [0, 1]$  is the probabilistic membership of  $x_j$  in the cluster  $i$ , satisfying  $\sum_{i=1}^c \mu_{ij} = 1$ ,  $\sum_{j=1}^n \mu_{ij} > 1$  with the fuzzifier  $m_1 > 1$ , and a distance norm  $\|\cdot\|$ .

The basic steps are outlined as follows:

- 1) Fix number of clusters  $c$ . Choose values for  $m_1$  and threshold  $\varepsilon$ . Set iteration counter  $t$ . Initialize centroids  $v_i (i = 1, \dots, c)$ .
- 2) Update memberships  $\mu_{ij}$  by Equ. (6) for  $c$  clusters and  $n$  objects.

$$\mu_{ij} = \left[ \sum_{k=1}^c \left( \frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{m_1-1}} \right]^{-1} \quad (6)$$

where  $d_{ij}^2 = \|x_j - v_i\|$ .

- 3) Compute centroids  $v_i$  by Equ. (7).

$$v_i = \frac{\sum_{j=1}^n (\mu_{ij})^{m_1} x_j}{\sum_{j=1}^n (\mu_{ij})^{m_1}} \quad (7)$$

- 4) If terminating condition is not met, repeat steps 2 to 4.

## III. DISCRETE ROUGH-FUZZY C-MEANS BASED ON INFORMATION ENTROPY ALGORITHM

Integrating the technologies of fuzzy set and rough set, Lingras and West [5] put forward Rough C-means clustering, which adds a pair of lower and upper approximations to description in each cluster  $\alpha_i$ . Maji et al. [12] points out that the objects in Rough-fuzzy C-means are fuzzified in the boundary  $B(\alpha_i) = \overline{R}(\alpha_i) - \underline{R}(\alpha_i)$  and have a similar influence on the corresponding prototype and cluster in the lower approximation  $C(\alpha_i) = \underline{R}(\alpha_i)$ . It is as shown in Figure 1.

These equations (5) and (7) are respectively modified for the Rough-fuzzy C-Means. The objective function for Rough-fuzzy C-Means is

$$J_{RFCM} = w \times P + w' \times P' \quad (8)$$

$$P = \sum_{i=1}^c \sum_{x_j \in C(\alpha_i)} \|x_j - v_i\|^2$$

$$P' = \sum_{i=1}^c \sum_{x_j \in B(\alpha_i)} \mu_{ij}^{m_1} \|x_j - v_i\|^2$$

Computing centroids  $v_i$  is

$$v_i^{RFCM} = w \times C + w' \times C' \quad (9)$$

$$C = \frac{\sum_{x_j \in C(\alpha_i)} x_j}{|C(\alpha_i)|}$$

$$C' = \frac{\sum_{x_j \in B(\alpha_i)} \mu_{ij}^{m_1} x_j}{\sum_{x_j \in B(\alpha_i)} \mu_{ij}^{m_1}}$$

$\delta$  is calculated in order to determine the approximation regions. In this paper,  $\delta$  is a average of differences.

$$\delta = \frac{1}{n} \sum_{j=1}^n (I_{kj} - I_{ij}) \quad (10)$$

$I_{ij}$  and  $I_{kj}$  represent the self-information of  $u_{ij}$  and  $u_{kj}$  which are the maximum and secondary maximum memberships of  $x_j$ .

$I_{kj} - I_{ij}$  is more sensitive than  $u_{ij} - u_{kj}$  and equivalent to a nonlinear amplifier for the tiny difference. i.e. when  $u_{ij} = \frac{1}{16}$  and  $u_{kj} = \frac{1}{32}$ ,  $u_{ij} - u_{kj} = \frac{1}{32}$ , however,  $I_{kj} - I_{ij} = 1$ .

Information entropy can be used to select points for discretization [13] and ascertain the number of cluters [14].

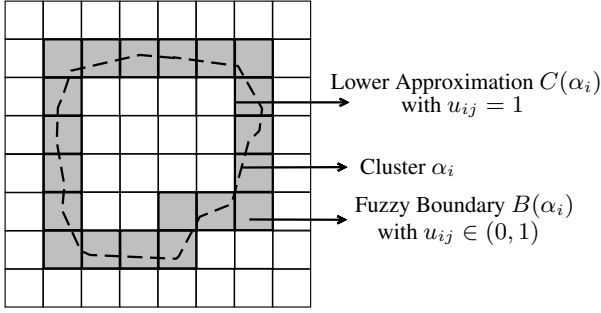


Figure 1. Rough-fuzzy C-means

Under any condition attribute and the decision attribute, the objects will be divided by a candidate breakpoint from the condition attribute. When its information entropy is lower than a threshold, the candidate breakpoint will be a breakpoint.

The number of clusters affects the effectiveness of the cluster algorithm. Under different numbers of clusters, the information entropy is calculated for evaluation of the effect in different cases. The lower value of information entropy tends to provide better clustering results.

Corresponding to the number of clusters  $c$ , the membership matrix  $U_c$  of  $X$  is generated. The information entropy of each object  $x_j$  under  $c$  is computed by Equ. (11).

$$H_{c_j}(x_j) = - \sum_{i=1}^c \mu_{ij} \cdot \log_2 \mu_{ij} \quad (11)$$

The information entropy of  $U_c$  is computed by Equ. (12).

$$H_c(X) = \sum_{j=1}^n H_{c_j}(x_j) \quad (12)$$

Information entropy of these classes will be compute by Equ. (13)

$$H(c_i) = - \sum_{j=1}^n \mu_{ij} \cdot \log_2 \mu_{ij} \quad (13)$$

The pseudo-code for the Discrete Rough-fuzzy C-means Based on Information Entropy Algorithm (DRFCMI) is illustrated in Algorithm 1.

#### IV. EXPERIMENTAL RESULTS OF VIOLENT CRIMES

Our developed algorithm was implemented in the C language and its parameters settings was chosen in accordance with the recommendations of Maji [12] and scope [2, 5] of clustering. The computation environment was an Intel<sup>®</sup> Core<sup>™</sup> i5-3337U CPU @1.80 GHz processor with 4G memory.

---

#### Algorithm 1 DRFCMI Algorithm

---

##### Input:

A decision table, initial centroids  $v_i$ , iterations  $t$ , threshold  $\varepsilon$  and scope  $[c_l, c_r]$  of clustering.

##### Output:

A result of clustering.

- 1: Each continuous condition attribute  $co$  and the decision attribute  $d$  will be a new decision table;
  - 2: Compute  $H(X/d)$  using Equ (2) and  $p_v = \frac{|d_v|}{|X|}$ .  $d_v$  represents any value of  $d$ . The threshold  $H = H(X/d)$ ;
  - 3: The candidate breakpoint set  $B_r$  and the breakpoint set  $B'$ . The value of  $co$  is sorted. Each value in  $B_r$  is the average from two consecutive numbers of these sorted value and the initial  $B'$  is null;
  - 4:  $X$  is divided into  $m$  portions,  $X = \{X_1, X_2, \dots, X_m\}$ , by these breakpoints from  $B'$ ;
  - 5: Compute  $H(b, X) = H^{X_1}(b) + H^{X_2}(b) + \dots + H^{X_m}(b)$  about all breakpoints ( $b \in B_r$ ).  $H^{X_w}(b)$  ( $w = 1, 2, \dots, m$ ) is the information Entropy of  $X_w$  which is divided by  $b$ . Find the minimum  $H(b', X)$  of  $H(b, X)$  and the corresponding breakpoint  $b'$ ;
  - 6: **if**  $H(b', X) < H$  **then**
  - 7:  $B_r$  deletes  $b'$  and  $B'$  adds  $b'$ , then  $H = H(b', X)$ ;
  - 8: **end if**
  - 9: Untill the sort of  $d_v$  in any  $X_w$  is 1 and  $B'$  under  $co$  is got, otherwise repeat steps 4 to 9.
  - 10: A discreted decision table is obtained by all  $B'$ .
  - 11: **for**  $c = c_l$  to  $c_r$  **do**
  - 12: Calculate  $\mu_{ij}$  using Equ. (6).
  - 13:  $u_{ij}$  and  $u_{kj}$  are the highest and the second highest memberships of  $x_j$ , respectively. Calculate  $I_{ij}$  and  $I_{kj}$  using Equ. (1).
  - 14: Calculate  $\delta$  using Equ. (10).
  - 15: **if**  $(I_{kj} - I_{ij}) > \delta$  **then**
  - 16:  $x_j \in C(\alpha_i)$ ;
  - 17: **else**
  - 18:  $x_j \in B(\alpha_i)$  and  $x_j \in B(\alpha_k)$ .
  - 19: **end if**
  - 20: Update new centroids as per Equ. (9).
  - 21: Compute objective function by Equ. (8).
  - 22: Repeat steps 12 to 21 by incrementing  $t$ , untill  $|J_{RFCM}(t) - J_{RFCM}(t-1)| < \varepsilon$ .
  - 23: Under the number of clusters  $c$ , the information entropy of  $U_c$  and each class are respectively calculated by Eqs. (12) and (13).
  - 24: **end for**
-

Table I  
RESULTS OF DISCRETIZATION

	N	P	MAOA	DRD4	Jail time	Brith place
numbers of value	24	21	5	10	64	5
numbers after Discretization	4	3	3	5	3	4

### A. Discretization and Result

Based on previous work, the reduced attributes are N-value, P-value, MAOA, DRD4, jail time and birth place and crime type. The reduced attributes except crime type are continuous. The crime type is the decision attribute. The results of discretization is shown in Table I. The information entropy  $H_c(X)$  in all cluster numbers and obviously low value in  $H(c_i)$  are recorded in Table II. And The explanation of the corresponding centroids in  $H(c_i)$  are shown in Table III.

### B. Discussion

According to our data and results, the following rules are extracted from Table III

- All N-values are normal, however, all P-values are higher-than-normal;
- DRD4 is 4 in these classes. There are two types of MAOA: 3 and 3.5 or 4. The jail time of the former is longer;
- These prisoners from central and north area are typicality. i.e.  $H(4)$  has been found by MESR: P-value is abnormal; DRD4 is 4; Long penalty and north area; Crime type is intentional injury;
- The penalty of intentional injury is harsh. The jail time is very long and even life imprisonment;
- P-value is abnormal. MAOA is 3. These prisoners from central area were sentenced to death for murder.

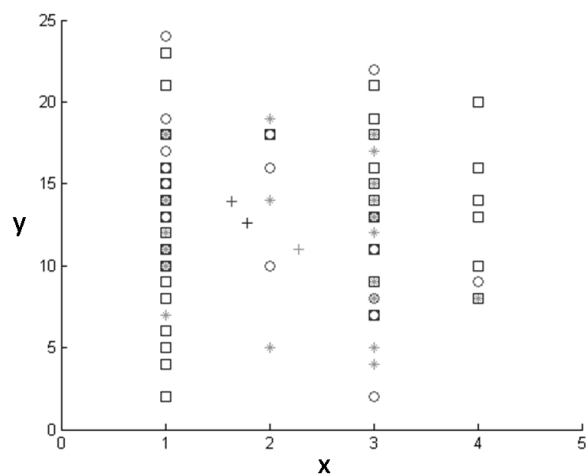
Some interesting results are collected from the available statistics:

- The rate of higher-middle P-value is great in Fig. 2, especial intentional injury ( $x = 3$ ) and rape ( $x = 4$ );
- Most of N-value is normal in Fig. 3 whatever the jail time and crime type are;

The factors for occurrence of crimes do not depend on single variable. If the high-P people are contributed together by some factors, a crime will be probably committed.

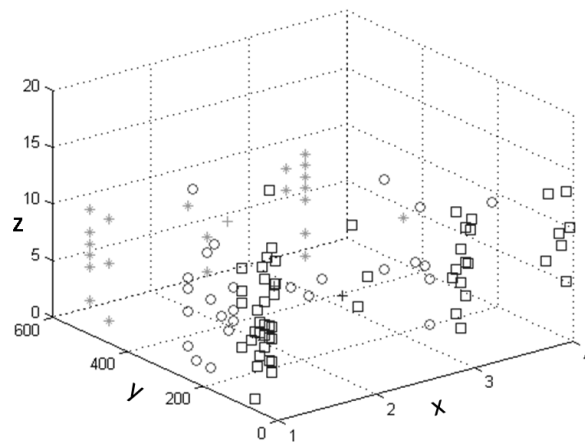
## V. CONCLUSIONS

The sensitivity of violent crime data requires reducing human intervention in the analysis process. As an objective evaluation criterion, information entropy is applied to discretization, forming rough set of Rough-fuzzy C-means and screening typical conclusions in this paper. As every class is quantized by information entropy, these clusters can be understood by analyzing information. The information



x axis: crime type  
y axis: P-value

Figure 2. The relationship diagram of crime type and P-value in  $c = 3$



x axis: crime type  
y axis: jail time  
z axis: N-value

Figure 3. The relationship diagram of crime type, jail time and N-value in  $c = 3$

Table II  
INFORMATION ENTROPY

Number of Clusters	2	3	4	5
$H_c(X)$	23.1266	33.6615	45.4087	45.8876
$H(c_i)$	7.9930	5.2849	2.9647	3.1149 and 3.4313

Table III  
THE EXPLANATION OF THESE CENTROIDS IN LOW INFORMATION ENTROPY

attributes	N	P	MAOA	DRD4	Jail time	Brith place	Crime type
$H(2)$	low	middle	3.5 or 4	4	middle	central area	intentional injury
$H(3)$	low	middle	3.5 or 4	4	middle	central area	intentional injury
$H(4)$	low	middle	3	4	long	north area	intentional injury
$H(5_1)$	low	middle	3	4	long	north area	intentional injury
$H(5_2)$	low	middle	3	4	death	central area	murder

entropy is a mathematical expectation of a information system. The low information entropy means that the class is more probably achieved. Although we try to analyze violent crimes from a computational point of view, some conclusions need further validation by biologists. For parameter values and references detailed to relative theories and principles, the reader may consult [13], [12].

#### ACKNOWLEDGMENT

The authors would like to thank Jiang Jiao, Shuai Zhou, Xiaolin Ding, Meng Zhang and Hongbo Liu for their scientific collaboration in this research work. This work is supported partly by the Fundamental Research Funds for the Central Universities (Grant No. 2012TD027) and the Nature Science Foundation of Chongqing of China (Grant No. 2009BA5074). Ajith Abraham acknowledges the support of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 funded by Structural Funds of the European Union and state budget of the Czech Republic.

#### REFERENCES

- [1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan kaufmann, 2006.
- [2] E. H. Ruspini, "A new approach to clustering," *Information and control*, vol. 15, no. 1, pp. 22–32, 1969.
- [3] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.
- [4] G. Wang, Y. Yao, and H. Yu, "A survey on rough set theory and applications," *Chinese Journal of Computers*, vol. 32, no. 7, pp. 1229–1246, 2009.
- [5] P. Lingras and C. West, "Interval set clustering of web users with rough k-means," *Journal of Intelligent Information Systems*, vol. 23, no. 1, pp. 5–16, 2004.
- [6] S. Mitra, H. Banka, and W. Pedrycz, "Rough-fuzzy collaborative clustering," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 4, pp. 795–805, 2006.
- [7] J. Guo, D. Miao, and J. Zhou, "Shadowed sets based threshold selection in rough clustering," *Computer Science*, vol. 38, no. 10, pp. 209–210, 2011.
- [8] C. Yang, H. Liu, Y. Sun, and A. Abraham, "Multi-knowledge extraction from violent crime datasets using swarm rough algorithm," in *Proceedings of 12th International Conference on Hybrid Intelligent Systems (HIS)*. IEEE, 2012, pp. 560–565.
- [9] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.
- [10] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, 1981.
- [11] J. T. Tou and R. C. Gonzalez, *Pattern recognition principles*. Addison-Wesley, 1974.
- [12] P. Maji and S. K. Pal, "Rough set based generalized fuzzy c-means algorithm and quantitative indices," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 6, pp. 1529–1540, 2007.
- [13] H. Xie, H. Cheng, and D. Niu, "Discretization of continuous attributes in rough set theory based on information entropy," *Chinese Journal of Computers*, vol. 28, no. 9, pp. 1570–1573, 2005.
- [14] J. Li, T. Meng, and X. Zhang, "Research of fault fuzzy criterion method based on information entropy- fuzzy clustering and rough set theory," *Computer Measurement & Control*, vol. 19, no. 5, pp. 1036–1039, 2011.