Hybrid Particle Swarm Optimization Feature Selection for Crime Classification

Syahid Anuar, Ali Selamat, and Roselina Sallehuddin

Faculty of Computing, Universiti Teknologi Malaysia
syah2105@yahoo.com, {aselamat,roselina}@utm.my

Abstract. In this study, we propose a hybrid crime classification model by combining artificial neural network (ANN), particle swarm optimization (PSO) and grey relation analysis (GRA). The objective of this study is to identify the significant features of the specific crimes and to classify the crimes into three different categories. The PSO as the feature selection method, reduce the dimension of datasets by selecting the most significant features. The reduction of the datasets' dimension may reduce the complexity thus shorten the running time of ANN to classify the crime datasets. The GRA is used to rank the selected features of the specific crimes thus visualize the importance of the selected crime's attribute. The experiment is carried out on the Communities and Crime dataset. The result of PSO feature selection will then compare with the other feature selection methods such as evolutionary algorithm (EA) and genetic algorithm (GA). The classification performance for each feature selection method will be evaluated. From our experiments, we found that PSO select less features compare with EA and GA. The classification performance results show that the combination of PSO with ANN produce less error and shorten the running time compare with the combination of EA with ANN and GA with ANN.

Keywords: Crime Classification, Particle Swarm Optimization, Grey Relation Analysis, Artificial Neural Network, Hybrid Artificial Neural Network, Crime Prevention.

1 Introduction

The increasing volume of the crimes had brought serious problems to the community in any country. However, the increase in realization of information technology has open a new door for government to include crime prevention component as a strategy to reduce crime. Classification is one of the data mining technique which can be use to analyze the crime patterns. The data mining approach can help to detect the crime patterns and speed up the process of solving crime [17]. Due to this reasons, research on crime classification has increase because of the potential and effectiveness of the classification in crime prevention programs. Several study on crime classification have been done by several researchers [10, 11, 16]. The crime can be divided into several types and the most common findings at the city level are crimes against property such as

D. Barbucha et al. (eds.), *New Trends in Intelligent Information and Database Systems*, Studies in Computational Intelligence 598, DOI: 10.1007/978-3-319-16211-9_11

burglary, robbery and theft and crime of aggression such as assault, homicides and rape [9]. Each crime's types can be classified into several categories such as low, medium and high. Thus, the objective of this study is to propose a new hybrid classifier to classify the crimes into given categories.

The artificial neural network (ANN) is a model inspired by the connected neurons in the brain. ANN with back-propagation learning algorithm is usually use as a benchmark model for any classifier. However, ANN is a black-box learning approach where it cannot determine automatically the significant input features. To overcome the limitation of ANN in choosing relevant features as input, thus the feature selection is needed. Inspired by bird flocking or fish schooling, particle swarm optimization (PSO) is one of the popular optimization techniques that have the capability to perform the feature selection task [19]. Grey relation analysis (GRA) is a method of analysis proposed in Grey system theory [8]. The GRA also can be used to rank the data according to its importance [3]. Therefore, in this paper, we intend to combine PSO and ANN as a hybrid classifier and using GRA to rank the crime attributes to its importance.

2 Related Works

A general framework for crime data mining has been introduced by Chen et al. [7] as the initial idea of using data mining technique in crime's domain. After that, the data mining techniques in crime was applied by Adderley [2] where in this study, self organising map (SOM) was applied to recognize burglary offences. Nath [17] has proposed data mining technique such as k-means to identify crime pattern using real crime data from a sheriff's office. Ozgul et al. [20] have investigated the sources of crime data mining to forms knowledge discovery that is suitable for which methodologies. Nissan [18] has studied the application of data mining in two forensic fields such as intelligence and investigative tasks.

Classification is one of the data mining techniques. Abu Hana et al. [1] have proposed crime scenes classification using neural network and support vector machine to classify violent crime into two classes which are attack from inside or outside of the scene. Futhermore, Yang et al. [22] have used classification methods such as decision tree, random forest, support vector machine and neural network to predict the relationships between murder victims and the offenders. Kim et al. [13] examine the relationship between weather and crime using decision tree algorithm. Kamiran et al. [12] have proposed discrimination aware classification to predict whether an individual is a crime suspect or not.

Iqbal et al. have proposed a classification methods to applied on crime dataset to predict crime categories. The crime dataset can be downloaded from UCI machine learning repository website [5]. The crime categories which are low, medium and high were added based on attribute name 'Violent Crime Per Pop'. The crime categories need to be added as goals or classes to enable classification algorithms to perform the prediction [10]. However, only statistical techniques such as Naive Bayes and decision tree were considered to classify the crime dataset. Later, the feature selection method and machine learning method have

been applied by Anuar et al. [4] as a continuation of previous work and experimental results show improvement in terms of classification accuracy. To further improve both experiments, we proposed a crime classification model. The classification model consist of feature selection method and classification method. PSO has been selected as feature selection method based on the encouraging experiment result. The GRA will be used to rank the crime's attributes in order of importance and ANN will be used as a classification method. The next section will discuss in detail the proposed crime classification model.

3 Proposed Crime Classification Model

3.1 Particle Swarm Optimization

A population in PSO is called a swarm. The candidate solutions are encoded as particles in the search space. The PSO starts by randomly initialize the population of particles. The whole swarm move in the search space to search for the best solution by updating the position of each particle according to Eqn 1.

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{1}$$

The movement of the current position of the particle i is represented by a vector $x_i = (x_{i1}, x_{i2}, ..., x_{iD})$, where D is the dimension of the search space. The velocity of the particle i is represented as $v_i = (v_{i1}, v_{i2}, ..., v_{iD})$ which is limited by a predefined maximum velocity, $v_{id}^{t} \in [-v_{max}, v_{max}]$. The equation for velocity is given by Eqn 2

$$v_{id}^{t+1} = w * v_{id}^{t} + c_1 * r_{1i} * (P_{id} - x_{id}^{t}) + c_2 * r_{2i} * (P_{gd} - x_{id}^{t}) \tag{2}$$

where t denotes the tth iteration, $d \in D$ denotes the dth dimension in the search space, w is the inertia weight, *c1* and *c2* are acceleration constants or sometimes called the learning rate, r_{1i} and r_{2i} are the random numbers uniformly distributed within the range of [0,1], p_{id} represent the element of the solution for the particle's individual best *pbest*, and p_{gd} is the element of the solution for the particle's global best, *gbest* in the dth dimension. Generally, the PSO algorithm consists of three steps which are repeated until some stopping condition is met [21]:

1. Evaluate the fitness of each particle.
2. Update the individual and the global best fitnesses and positions.
3. Update the velocity and the position of each particle.

For the purpose of the feature selection, the required particle number and the initial coding alphabetic string for each particle is randomly produced. Each particle encoded to imitate a chromosome, like in the genetic algorithm. Each particle was coded to a binary alphabetic string $S = F_1, F_2, ..., F_n, n = 1, 2, ..., m$ where m are the number for features. The bit value 1 represents a selected feature and the bit value 0 represents a non-selected feature. The *gbest* value consider as the fitness value.

3.2 Grey Relation Analysis(GRA)

The grey relation analysis (GRA) is a method of analysis, proposed in the Grey system theory. GRA is a distinct similarity measurement that use data series to obtain grey relational order to describe the relationship between the related series. Therefore, it is significant to use GRA to analyze the relationship between features and the type of crimes. GRA consists of two parts which are reference sequence and compared sequence. The following are the basic steps in the GRA [14]:

1. The type of crimes is set as the reference sequence, $X_0(k), k = 1, 2, 3, ..., M$ where M = number of instances in datasets and compared sequence $X_i(k)$, $i = 1, 2, ..., N$ where N = number of features, is the features after feature selection process. Fig. 1 shows the example of decision matrix for crime dataset.

$$(x_0, x_1, ..., x_N) = \begin{bmatrix} x_0(1) & x_1(1) & ... & x_N(1) \\ x_0(2) & x_1(2) & ... & x_N(2) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ x_0(M) & x_1(M) & ... & x_N(M) \end{bmatrix}$$

Fig. 1. M × N decision matrix [15]

2. Normalization of the sequence to dimensionless form, is calculated according to Eqn 3:

$$x_i(k) = \frac{X_i(k) - X_{min}(k)}{X_{max}(k) - X_{min}(k)} \tag{3}$$

3. Calculate the grey relational coefficient, ξ, between X_0 and X_i is calculated according to Eqn 4:

$$\xi_i(k) = \frac{\min\limits_i \min\limits_k \triangle_i(k) + \rho \max\limits_i \max\limits_k \triangle_i(k)}{\triangle_i(k) + \rho \max\limits_i \max\limits_k \triangle_i(k)} \tag{4}$$

where $\triangle_i(k) = |x_0(k) - x_i(k)|$ is the absolute difference of index k. ρ is the distinguishing coefficient and in most situations takes the value of 0.5 because this value usually offers moderate distinguishing effects and good stability [6].

4. Calculate the grey relational grade (GRG) according to Eqn 5:

$$r_i = \frac{1}{n} \sum_{i=1}^{n} \xi_i(k) \tag{5}$$

4 Experiment Setup

4.1 Dataset Preparation

The Communities and Crime dataset is obtained from the UCI Machine Learning Repository. This dataset focus on the communities in United States of America (USA). The data comprises of socio-economic data from the 90 Census, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey and crime data from the 1995 FBI UCR. The dataset consists of 147 attributes and 2215 instances including missing values [5]. Data preparation is essential for successful data classification. Poor quality data typically results in incorrect and unreliable classification results. Thus, the following data preparation mechanisms were carried out:

- The dataset is divided to three type of crimes which are assault, burglaries and murders.
- All the attributes with large number of missing values are remove
- Following the work done by Iqbal et al. [10], the newly added nominal attribute named 'Categories'is created from each type of crime for classification purpose - if the value is less than 25% than the Categories is 'Low'. If the value is equal to or greater than 25% than the Categories is 'Medium'. If the value is equal to or greater than 40% than the Categories is 'High'.
- All the data will be normalize to [0, 1] using the min-max method represented by Eqn 3.

4.2 Parameters Setting

Table 1 shows the parameter setting for GA, EA, and PSO. The parameters setting for all algorithms is based on the best result obtained from several self-experiments.

Table 1. Parameters Setting

PSO	EA	GA
C1:2.0	Crossover rate: 0.6	Crossover rate: 0.6
C2:2.0	Mutation rate: 0.01	Mutation rate: 0.033
Number of particles: 200	Population size: 200	Population size: 200

5 Result and Discussion

5.1 Feature Selection Results

Fig. 2 shows the result of feature selection for each method. The result shows that PSO selects lowest features for assault, burglaries and murders. For assault, PSO selects 4 significance features compare with EA and GA which are 7 and 15. For burglaries, PSO selects only 3 significance features compare with EA which is 7 and GA which is 12. The PSO selects 7 significance features while EA selected 12 and GA selected 8 significance features for the murders.

For all types of crimes, PSO has successfully selected the lowest number of features. Hence, after the feature selection process, GRA will rank the selected features thus visualize the importance of each feature.

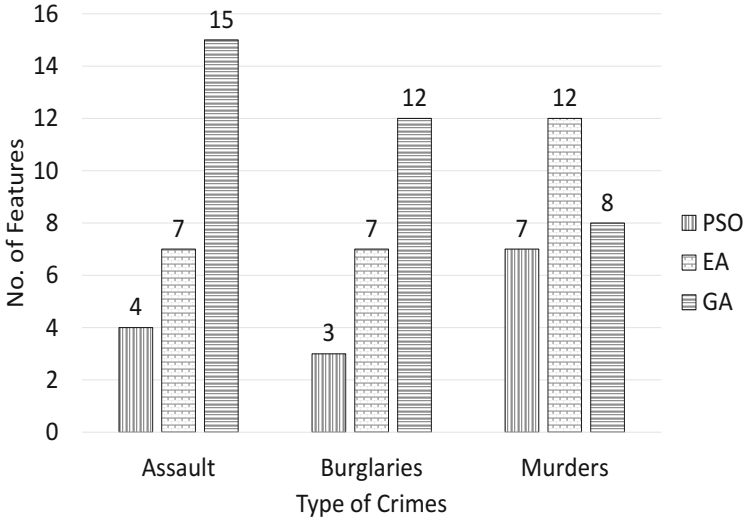

Fig. 2. Feature selection results

Table 2 shows the result of GRA after the feature selection process by PSO. Each selected features are ranked to its importance based on GRG. Higher GRG meaning that the features are the most important. This ranked features may reflect the importance of the attributes of crime and can make it easier for the police to prioritize the work force.

Table 2. The rank of crime features

Type of crimes	Features Number	Features Name	Features Description	GRG	Rank
Assault	4	pctAsian	percentage of population that is of Asian heritage	0.8349	1
	9	pct65up	percentage of population that is 65 and over in age	0.7235	2
	24	pctEmployProfServ	percentage of people 16 and over who are employed in professional services	0.6184	3
	64	pctUsePubTrans	percent of people using public transit for commuting	0.3997	4
Burglaries	49	pctWfarm	percentage of households with farm or self employment	0.8391	1
	63	pctSpeakOnlyEng	percent of people who speak only English	0.3902	2
	64	pctSameCounty-5	percent of people living in the same state	0.3328	3
Murders	4	pctAsian	percentage of population that is of Asian heritage	0.8427	1
	29	pctFemDivorc	percentage of households with farm or self-employment income	0.7713	2
	31	pct2Par	percentage of families (with kids) that are headed by two parents	0.6217	3
	35	pctWorkMom-6	percentage of moms of kids 6 and under in labour force	0.4849	4
	37	pctKidsBornNevrMarr	percentage of kids born to never married	0.3682	7
	41	pctFgnImmig-10	percentage of immigrants who immigrated within last 10 years	0.5399	6
	63	pctSameHouse-5	percent of people living in the same house	0.4208	5

5.2 Classification Results

The classification performances are evaluated using several criteria. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) will be used to measure the performance of ANN. Accuracy, precision, recall and f-measure will be used to measure the classification's performance. Table 3 shows the results of RMSE and MAE for each crime. The results for assault indicate that PSO-ANN gives lowest RMSE compare with EA-ANN and GA-ANN. For Burglaries, the PSO-ANN generates lowest RMSE but ANN generate lowest MAE. For murders, the ANN give lowest MAE value.

Table 3. Artificial neural network performance results

Type of crimes	ANN		PSO-NN		EA-NN		GA-NN	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Assault	0.2155	**0.0641**	**0.1792**	0.0647	0.1797	0.0651	0.1830	0.0674
Burglaries	0.1296	**0.0203**	**0.1212**	0.0290	0.1222	0.0283	0.1238	0.0287
Murders	0.2025	**0.0566**	0.1827	0.0671	0.1852	0.0688	0.1827	0.0673

Note: The bold value indicates the lowest error.

Table 4 shows the results of classification performance and processing time for various classification models. These results aim to demonstrate the effectiveness of feature selection methods on various crime datasets.

Table 4. Comparison of classification result and processing time

Type of crimes	Models	Accuracy (%)	Precision (%)	Recall (%)	F-measure	Time(s)
Assault	ANN	92.27	90.60	92.30	91.40	1480
	PSO-ANN	94.99	90.20	95.00	92.50	26
	EA-ANN	94.98	90.20	95.00	92.50	43
	GA-ANN	94.89	90.20	94.90	92.50	108
Burglaries	ANN	97.33	95.50	97.30	96.40	1661
	PSO-ANN	97.74	95.50	97.70	96.60	24
	EA-ANN	97.69	95.50	97.70	96.60	43
	GA-ANN	97.65	95.50	97.70	96.60	74
Murders	ANN	93.45	90.00	93.50	91.70	1504
	PSO-ANN	94.76	89.90	94.80	92.30	44
	EA-ANN	94.76	89.90	94.80	92.30	74
	GA-ANN	94.80	89.90	94.80	92.30	47

The results from Table 4 indicate that PSO-ANN produce lowest running time for each type of crimes. For assault, PSO-ANN give 26 seconds compared with

EA-ANN which give 43 seconds. The GA-ANN give the highest running times which are 108 seconds. The PSO-ANN give 24 seconds for burglaries compare with EA-ANN which give 43 seconds, and the highest 74 seconds for GA-ANN. For murders, the PSO-ANN give 44 seconds compared to EA-ANN which give 74 seconds and GA-ANN which give 47 seconds. The results prove that PSO has produced the lowest number of features thus reduce the running time of ANN. Although the results for accuracy, precision, recall and f-measure are so close, the processing time for each model shows the effectiveness of feature selection. This suggests that by using a small number of features can also produce high accuracy in a short time.

5.3 Conclusions

This paper presents a hybrid application of PSO and GRA in crime classification modelling. The PSO as a feature selection method to identify and obtain the significant features for different type of crimes such as assault, burglaries and murders. The GRA is used to rank the significant features of each type of crimes. The ANN is used to classify each crime into three different categories which are low, medium and high. The experiment result indicates that the PSO select lowest number of features thus shortens the running time of ANN. Moreover, the GRA has ranked the selected features of it important thus visualize the crimes dataset to be more specific. This would help the police to prioritize the important attributes for the specific type of crimes. Finally, the experimental results have proved that PSO-ANN is an acceptable model to analyze crime data.

Acknowledgement. The Universiti Teknologi Malaysia (UTM) and Ministry of Education Malaysia under research Grant 01G72 and Ministry of Science, Technology & Innovations Malaysia, under research Grant 4S062 are hereby acknowledged for some of the facilities that were utilized during the course of this research work.

References

1. Abu Hana, R.O., Freitas, C.O., Oliveira, L.S., Bortolozzi, F.: Crime scene classification. In: Proceedings of the 2008 ACM Symposium on Applied computing, pp. 419–423. ACM Press (2008)
2. Adderley, R.: The use of data mining techniques in operational crime fighting. In: Chen, H., Moore, R., Zeng, D.D., Leavitt, J. (eds.) ISI 2004. LNCS, vol. 3073, pp. 418–425. Springer, Heidelberg (2004)
3. Alwee, R., Shamsuddin, S.M.H., Sallehuddin, R.S.: Economic indicators selection for crime rates forecasting using cooperative feature selection. In: Proceeding of the 20th National Symposium on Mathematical Sciences Research in Mathematical Sciences: A Catalyst for Creativity and Innovation, vol. 1522, pp. 1221–1231. AIP Publishing (2013)
4. Anuar, M.S., Selamat, A., Sallehuddin, R.: Particle swarm optimization feature selection for violent crime classification. In: Advanced Approaches to Intelligent Information and Database Systems, pp. 97–105. Springer (2014)

5. Bache, K., Lichman, M.: UCI machine learning repository (2013), http://archive.ics.uci.edu/ml
6. Chang, T.C., Lin, S.J.: Grey relation analysis of carbon dioxide emissions from industrial production and energy uses in taiwan. Journal of Environmental Management 56(4), 247–257 (1999)
7. Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y., Chau, M.: Crime data mining: a general framework and some examples. Computer 37(4), 50–56 (2004)
8. Deng, J.-L.: Introduction to grey system theory. The Journal of Grey System 1(1), 1–24 (1989)
9. Gorr, W., Olligschlaeger, A., Thompson, Y.: Assessment of crime forecasting accuracy for deployment of police. International Journal of Forecasting (2000)
10. Iqbal, R., Murad, M.A.A., Mustapha, A., Panahy, S., Hassany, P., Khanahmadliravi, N.: An experimental study of classification algorithms for crime prediction. Indian Journal of Science & Technology 6(3) (2013)
11. Theresa, M.J., Raj, V.J.: Fuzzy based genetic neural networks for the classification of murder cases using trapezoidal and lagrange interpolation membership functions. Applied Soft Computing 13(1), 743–754 (2013)
12. Kamiran, F., Karim, A., Verwer, S., Goudriaan, H.: Classifying socially sensitive data without discrimination: an analysis of a crime suspect dataset. In: 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW), pp. 370–377. IEEE (2012)
13. Kim, J.-M., Ahn, H.-K., Lee, D.-H.: A study on the occurrence of crimes due to climate changes using decision tree. In: IT Convergence and Security 2012, pp. 1027–1036. Springer (2013)
14. Kung, C.-Y., Yan, T.-M., Chuang, S.-C., Wang, J.-R.: Applying grey relational analysis to assess the relationship among service quality customer satisfaction and customer loyalty. In: 2006 IEEE Conference on Cybernetics and Intelligent Systems, pp. 1–5. IEEE (2006)
15. Lu, J., Chen, P., Shen, J., Liang, Z., Yang, H.: Study on the prediction of gas content based on grey relational analysis and bp neural network. In: Proceedings of The Eighth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), 2013, pp. 677–685. Springer (2013)
16. Nasridinov, A., Ihm, S.-Y., Park, Y.-H.: A decision tree-based classification model for crime prediction. In: Information Technology Convergence, pp. 531–538. Springer (2013)
17. Nath, S.V.: Web Intelligence and Intelligent Agent Technology Workshops. In: 2006 IEEE/WIC/ACM International Conference on. WI-IAT 2006 Workshops, pp. 41–44. IEEE (2006)
18. Nissan, E.: An overview of data mining for combating crime. Applied Artificial Intelligence 26(8), 760–786 (2012)
19. Omar, N., Shahizan, M., Othman, o.b.: Particle swarm optimization feature selection for classification of survival analysis in cancer. International Journal of Innovative Computing 2(1) (2013)
20. Ozgul, F., Atzenbeck, C., Çelik, A., Erdem, Z.: Incorporating data sources and methodologies for crime data mining. In: 2011 IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 176–180. IEEE (2011)
21. Shi, Y., Eberhart, R.: A modified particle swarm optimizer. In: The 1998 IEEE International Conference on Evolutionary Computation Proceedings, IEEE World Congress on Computational Intelligence, pp. 69–73. IEEE (1998)
22. Yang, R., Olafsson, S.: Classification for predicting offender affiliation with murder victims. Expert Systems with Applications 38(11), 13518–13526 (2011)