Theoretical and Empirical Analysis of Crime Data

Manisha Mudgal*, Deepika Punj and Anuradha Pillai

Department of Computer Engineering, JC BOSE UST YMCA Faridabad, Haryana, India
E-mail: mudgal.05.manisha@gmail.com; deepikapunj@gmail.com; anuangra@yahoo.co.in
**Corresponding Author*

Received 30 September 2020; Accepted 31 October 2020;
Publication 17 February 2021

Abstract

Crime is one of the biggest and dominating problems in today's world and it is not only harmful to the person involved but also to the community and government. Due to escalation in crime frequency, there is a need for a system that can detect and predict crimes. This paper describes the summary of the different methods and techniques used to identify, analyze and predict upcoming and present crimes. This paper shows, how data mining techniques can be used to detect and predict crime using association mining rule, k-means clustering, decision tree, artificial neural networks and deep learning methods are also explained. Most of the researches are currently working on forecasting the occurrence of future crime. There is a need for approaches that can work on real-time crime prediction at high speed and accuracy. In this paper, a model has been proposed that can work on real-time crime prediction by recognizing human actions.

Keywords: Crime, data mining, deep learning, KNN, RNN, Gaussian, Naïve Bayes, clustering, classification, decision tree.

Journal of Web Engineering, Vol. 20_1, 113–128.
doi: 10.13052/jwe1540-9589.2016

1 Introduction

With growing population, the crime rate has also escalated at a very high speed in recent years. It has made very difficult for security agencies to detect and predict crime. For the development of smart cities, there is an urgent need for systems that can assist these agencies. In this paper, many approaches have been discussed that can use to make our cities secure.

Criminology is a study of characteristics of criminal and crime identification. Criminology aids the detective agencies and cops in investing about crime and criminal. Government of different countries has also taken steps to develop softwares for solving problems with fast speed. Any kind of research in this field can be beneficial for security agencies.

Data Mining in the 1990s, came into existence and its techniques helped in extracting useful information from datasets. Earlier it was challenging to extract information from big datasets and to find relationships between attributes, but data mining has solved these problems.

Data mining has the power of extracting relevant information from the database. This information can be beneficial for security agencies and police. Many authors have collect 10 to 15 years of previous data. On these datasets,

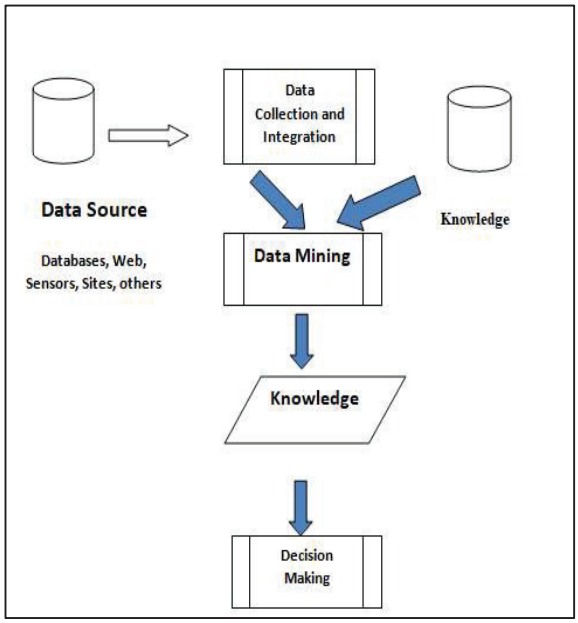


Figure 1 Data mining model.

different Mining techniques and Machine Learning algorithms have been applied to find patterns and to make predictions. With these algorithms, the hotspots of crime can be identified and can do forecasting for future crime rate. Data Mining plays a crucial role in forensic and criminology domain.

Different classification and clustering algorithms can be used to identify crime patterns according to requirements. Depending upon dataset and techniques used accuracy of the result may vary. Some authors have used combined models to improve speed and accuracy.

Smart Cities need some more intelligent ways to improve security in the areas and to develop a city that is not only smart also secure to live. In cities, we have seen CCTV cameras and video surveillance to improve security. For this, there is a supervisor to monitor it. Being a human he/she can perform mistake and can miss detecting suspicious acts. These even create a backlog of videos. For solving these issues, there is a need of the auto-detection system. Many authors have proposed some real time crime activity detection systems.

Auto-Detection systems use the Neural Network, Deep learning algorithms to analyze the streaming of videos. The models can detect crowd movements, objects like knife, gun, and facial features. An alarm system can also be attached to the models for informing the supervisor about the act.

The motivation for this survey paper is to aid a helping hand for researchers who wants to perform their research work in crime analysis and prediction. The survey of this paper will give them insights about procedures for crime analysis and different types of operations that can be performed to produce the desired result.

Figure 2 Camera detecting Gun with a red box.

1.1 Crime Procedure Analysis

Investigation of the crime scene can be a challenging problem for security agencies and police. Every criminal leaves some or other clue while fleeing the crime spot. Identification of those clues can help police in identifying people involved in that crime. 50% of crimes are committed by the same 10% of criminals. The sequence of crime and patterns followed by criminals can be used for analysis. The analyses can be done on the information collected from the crime spot against the previous data and a procedure model can be used for investigating the crime. Predictions can be made useful for tackling crime in advance. For tackling crime, police or security agencies can increase the security in areas where crime can occur or can track the suspicious person.

The data collected for crime prediction or analysis are not uniform for that some preprocessing is always required and in some cases results are also processed. Figure 3 shows how mining is done on crime data for prediction.

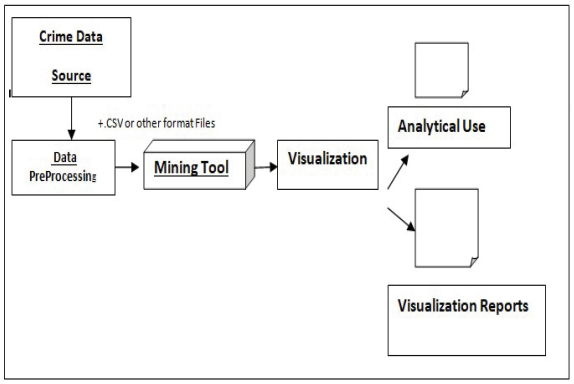


Figure 3 Mining process.

2 Technique Used for Detection of Various Crimes

Genetic Algorithm [12]: Genetic algorithm is used for understanding obliged and unconstrained modifications. In every iteration, it gives points and is based on biological choice process. It is similar to biological algorithms that improve issues like inheritance. They can be used in computer science, mathematics, biology etc. This algorithm is fast, and gives optimal solutions.

Naïve Bayesian [2]: Naive Bayesian gives probability distribution. It is a good classification technique and is based on probability. It calculates the posterior from prior and likelihood.

Fuzzy C-means Algorithm [4]: FCM clustering algorithm was developed by J C Dunn. It is very similar to K-means algorithm. In this algorithm clustering of data is performed and it randomly assigns coefficients to each point in the clusters. FCM with automatic detection of cluster number can enhance the accuracy of detection.

Neural Network [5]: The primary purpose of the Neural Network is to develop algorithms that have the ability to learn and recognize patterns and to generate knowledge out of it. There are input, hidden and output layers. In NN first, inputs are passed to input layer and inputs are then multiplied with combined weight and sum of all inputs is calculated. It is then passed through an activation function and value is compared with threshold. If greater than the threshold, value is passed to the output layer. For getting the appropriate result, the neural network can adjust values through back propagation.

Logistic Regression [7]: It is used to calculate the probability of a class or event. LR can be used to classify whether an image is of car, truck or plane. Every object detected is an image is assigned a probability value between 0 and 1. It is simple and has very low variance and is less prone to over fitting.

Decision Tree [7]: Decision tree has a flowchart like structure and in this there is a root node, internal node represent test or attributes and leaf node represents class label. The classification rule in decision tree is represented by paths from the root to leaf. Tree-based predictive models have high accuracy and have more stability. Tree-based methods like decision tree, random forest etc. are very popularly in solving data science problems.

Region-Based Convolution Neural Network [3]: It is an object detection method. Detection of an object is based on visual information of that object in image. The possible location of object is first computed by network. This proposed region is passed for classification in CNN. But it makes the processing slow and need more space. So, to improve performance, we have Faster-RCNN.

3 Literature Survey

Researchers have used a variety of techniques to do analyses, prediction of crime and for identifying weapons like gun, knife. Researchers have also developed models to identify crime hotspots. All these models have been used according to requirements. Few such papers have been discussed below:

Table 1 List of papers which have discussed crime analysis and prediction

Researches Already Done	Data Set Used	Strength	Weakness	Technique Used	Results
M. Nieto et al. (2018) [1]	CERTH/ITI	Project works on real time online as well as offline video analysis	Working process need to be explained in more detail, graphs are less.	Semantic Analysis, Perspective-based detection, Optical ?ow based tracking.	In this paper authors have proposed a system that can work on real time videos as well as on offline videos for crime detection. For this CERTH/ITI data set has been used and videos analysis has been done. In last results have also been shown with 75% to 90% accuracy.
Ricardo Resende de Mendonça [2]	Twitter Data	Good approach to used twitter coded data for crime prediction	Ontology are not auto updated.	Semantic Web, Ontology Machine Learning Algorithms: Support Vector Machine, Artificial Neural Networks, Naive Bayes, Random Forest	Criminals use online social platforms for planning and execution of criminal activities. Semantic web provides computer interpretable models. Ontology based framework for intention classification is proposed to describe the relations and make inference to determine weights of messages that are suspicious. Machine learning techniques are used for automatic classification of posts according to proposed framework for intention classification. In original phrases Random Forest performed best. In deciphered phrases all shown average result.

(*Continued*)

Table 1 Continued

Researches Already Done	Data Set Used	Strength	Weakness	Technique Used	Results
Julio Suarez-Paez [3]	Colombian National Police	Low computation cost	Accuracy is only upto 70%	Region Based Convolution Network (R-CNN) CNN	In this author has used real time video analysis for analysis of crime activities. CNN model (AlexNet) is used for training and detection of knife, bladed weapons, fire. Author has proposed a Region Proposal Network to propose whether a certain image region is of interest or not. Final detection work is done by R-CNN that uses proposed model result and also (AlexNet) CNN to detect region of interest. Accuracy of this model is 70%.
B. Sivana-galeela [4]	Indian Crime data	Well explained	Accuracy rates are not shown	Fuzzy Mean Clustering Algorithm	Crime is something which cannot be tolerated by anyone. Author has tried to identify crime areas and most frequency occurred crimes. Instead of normal clustering author has used Fuzzy C Mean clustering Algorithm.
Tamanna Siddiqui [5]	Social Media Platform data	Sentiment analysis well explained	No crime prediction and model creation	K-Means Clustering, Naïve Bayes Classifier, Support Vector Machine, Decision Tree, Neural Networks, Association Rules, Sentiment Analysis, Topic Modeling	Now a day's online platforms are mostly used for planning criminal activities. In this paper author has tried to detect criminal activities on online social networks. Author has used text mining algorithms for extracting texts and then text analysis is done using Sentiment Analysis, Topic Modeling.

Mohammad Nakib [6]	Images of gun, knife, blood from different sources.	Accuracy is high.	Only predicting that crime occurred or not.	Convolutional Neural Network(CNN), Rectified Linear Unit (ReLu)	With the rapid growth in Computer Vision. Now we can predict Crime scene without human intervention. In this paper author has used ReLU, CNN to detect suspicious objects like knife, gun, and blood from an image. This detection of object can help in predicting whether crime has occurred or not and where it happened. This model gives 90.2% accurate result.
Alkesh Bharati [7]	Chicago Crime data	Crime visualization is done using many different parameters.	Accuracy rate is little low.	KNN Classification Logistic Regression Decision Trees Random Forest Support Vector Machine Bayesian methods	It will use machine learning techniques for crime prediction of Chicago data set. For prediction KNN classification and other algorithms will be used. The algorithm which have better accuracy will used for training of dataset. For data visualization different graphical representations are used.
Umadevi V Naval-gund [8]	Trained with gun and Knife dataset	It works on real time video analysis with auto SMS facility to make the security authorities aware	It does not checks for crime motions like hitting.	CNN, Faster RCNN, RCNN, VGCNet19, GoogleNet Inception V3	In this author has proposed a crime intention detection system that can automatically detect criminal activities and suspicious objects like knife, gun through CCTV cameras. For pretraining VGGNET19 has been used as its computationa time is less than GoogleNet V3 model. Fast RCNN algorithm is used to draw boxes around suspicious objects like gun. SMS sending system has been implemented for alerting supervisor.

(*Continued*)

Table 1 Continued

Researches Already Done	Data Set Used	Strength	Weakness	Technique Used	Results
Sharmila Chack-ravarthy [9]	Work on videos	Can detect knife, gun and even detect hitting actions.	Accuracy Rate is not shown.	Deep Convolutional Neural Network, Recurrent Neural Network, Hybrid Deep Learning Algorithm	For quicker and accurate detection of criminal activities author has developed a model for forecasting using Neural Network with Hybrid Deep learning algorithm. This will analyze the video data.
Suhong Kim [10]	Crime data of Vancouver	Many graphs are used for data visualization	Accuracy rate is very low.	Machine Learning Algorithms	Crime is a problem that affects us socially as well as economically. In this paper author has collected Vancouver Crime data of the last 15 years. In this author has used Machine learning predictive models and boosted decision tree with accuracy between 39% to 44%.

4 Result and Discussion

Most of the current solutions work on forecasting the occurrence of future crime. These solutions work on historical data and try to predict the type of crime that can occur on which place in future. The main issue with these solutions is that they work on the future and not on present crime prediction. So, there is a need for a system that can work on real-time criminal activities prediction and can warn us about the mishap. For smart cities, along with CCTV cameras, some technologies are also required that can auto monitor these videos and can notify if some suspicious activity is occurring.

The Proposed system works on an approach that can track human actions. This approach analyzes the videos and extracts features. The features of motion like hitting, robbery etc. are presented by Human Motion Tracking approach using Gaussian Mixture Model and Kalman filter method. Other features of video are based on visual characteristics of the frame. These features are extracted using Recurrent Neural Network model and Gated Recurrent Unit. For better recognition of criminal actions, this hybrid model can play an important role.

Human Motion Tracking: It is very necessary to have good feature extractor in order to improve the performance of Grated RNN in video classification. The main aim of feature extractor is to reduce the data size and improve the performance by decreasing the time of computation. In this hybrid approach Gaussian Mixed Model and Kalman Filter are used to filter the suspicious acts by bounding a box around the moving person in each frame of video and this then used by Grated RNN.

Gaussian Mixture Model and Kalman Filter: GMM is a probabilistic model. This model assumes that all the points of data are generated from a mixture that contains a finite number of Gaussian distributions. It can be considered as generalized K means clustering. Background modeling is essential to detect the person with some motion in a dynamic scene. GMM is very efficient for detecting human motion. Kalman filter is used to estimates the position of human along with its velocity and acceleration of that moving person. It is also called prior state estimate.

The primary purpose of using the Gaussian Mixture Model and Kalman filter is to decrease the complexity by creating a boundary box around the person moving in the suspicious frame.

Recurrent Neural Network: It is a class of ANN that is Artificial Neural Network. In this connections between the nodes form a graph which is directed and it exhibits dynamic behavior. RNN have variable number of RNN and hidden units. A memory cell of gated recurrent is added to reduce

the variable number of RNN and hidden unit parameters. This solves the problem of gradient vanishing. A reset gate and update gate in Gated RNN simplifies video classification and human action recognition problem. In this reset gate combines the new input frame with previous and update has control on keeping the information from previous. They are best for sequential data. In proposed model GRU acts as central element for action prediction of human. AS shown in Figure 4 the GRNN model can be split into three parts:

1. The input data part
2. Sequence Modeling of GRU
3. Predictive Module.

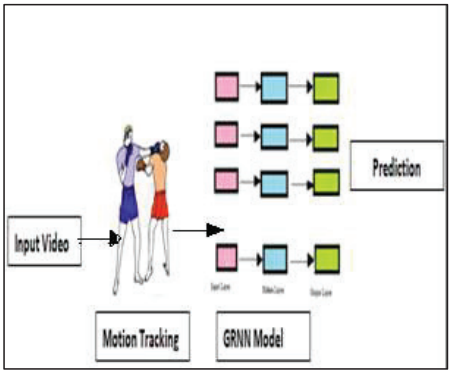


Figure 4 Proposed human action prediction model.

The experimental test of this approach shows very good results. That means this hybrid approach can be used for human action recognition.

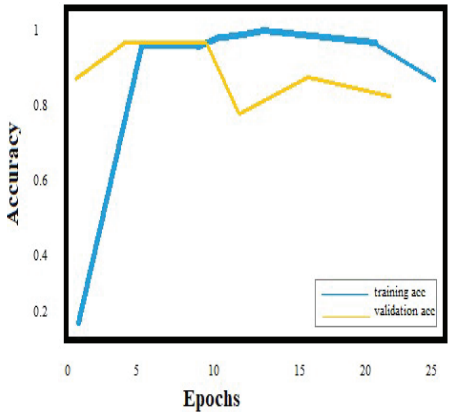


Figure 5 Accuracy in training and validation phase.

5 Conclusion

Crime rate is increasing day by day and it has become one of the most challenging problem. There is a need for a system that can detect and predict these activities. Many models have been already developed to reduce crime level but still more can be done to improve their accuracy and speed. In this paper, study of 10 papers related to crime prediction has been conducted and in last a new approach has been proposed that works on real time crime prediction by recognizing human action. The Proposed methods works on Grated Recurrent Neural Network Model.

References

[1] M. Nieto, L. Varona, O. Senderos, P. Leskovsky, and J. Garcia Real-time video analytics for petty crime detection. IEEE (2018).

[2] Ricardo Resende de Mendonça, Daniel Felix de Brito, Ferrucio de Franco Rosa, Júlio Cesar dos Reis and RodrigoBonacin, "A Framework for Detecting Intentions of Criminal Acts in Social Media: A Case Study on Twitter", Information Technology: New Generations (ITNG 2019).

[3] Julio Suarez Paez, MayraSalcedo Gonzalez, Alfonso Climente, ManuelEsteve, Jon Ander Gómez, Carlos Enrique Palau, Israel Pérez-Llopis, "A Novel Low Processing Time System for Criminal Activities Detection Applied to Command and Control Citizen Security Centers", Advanced Topics in Systems Safety and Security (2019).

[4] B. Sivanagaleela, S. Rajesh, "Crime Analysis and Prediction Using Fuzzy C-Means Algorithm", International Conference on Trends in Electronics and Informatics, IEEE 2019.

[5] Tamanna Siddiqui, Abdullah Yahya Abdullah Amer, Najeeb Ahmad Khan, "Criminal Activity Detection in Social Network by Text Mining: Comprehensive Analysis", International Conference on Information Systems and Computer Networks (ISCON), IEEE (2019).

[6] Mahmud, Nafiz, et al. "Crimecast: A crime prediction and strategy direction service." 19th International Conference on. IEEE, 2016.

[7] Mohammad Nakib, Rozin Tanvir Khan, Md. Sakibul Hasan, Jia Uddin, "Crime Scene Prediction by Detecting Threatening Objects Using Convolutional Neural Network", International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), IEEE 2018.

[8] Alkesh Bharati, Dr. Sarvanaguru RA, "Crime Prediction and Analysis Using Machine Learning", International Research Journal of Engineering and Technology (IRJET), (2018).

[9] Umadevi V. Navalgund, Priyadharshini K., "Crime Intention Detection System Using Deep Learning", International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET) 2018, IEEE (2019).

[10] S. Chackravarthy, S. Schmitt and L. Yang, "Intelligent Crime Anomaly Detection in Smart Cities Using Deep Learning," in 4th International Conference on Collaboration and Internet Computing (CIC), IEEE (2018).

[11] Suhong Kim, Param Joshi, Parminder Singh Kalsi, Pooya Taheri, "Crime Analysis Through Machine Learning", 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE (2018).

[12] Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav, "Crime Pattern Detection, Analysis & Prediction", IEEE 2018.

[13] Adewale Opeoluwa Ogunde, Gabriel Opeyemi Ogunleye, Luwaleke Oreoluwa, "A Decision Tree Algorithm Based System for Predicting Crime in the University", Machine Learning Research, Science Publishing Group 2017.

[14] Zhang Q, Yuan P, Zhou Q, Yang Z., "Mixed spatial-temporal characteristics based Crime Hot Spots Prediction". In Computer Supported Cooperative Work in Design (CSCWD, 2016 May 4 (pp. 97–101). IEEE (2016).

[15] S. Sivaranjani, Dr. S. Sivakumari, Aasha. M, "Crime prediction and forecasting in Tamilnadu using clustering approaches", Crime prediction and forecasting in Tamilnadu using clustering approaches, IEEE (2016).

[16] Cesario, Cesario E, Catlett C, Talia D. Forecasting Crimes Using Autoregressive Models. In Dependable, Autonomic and Secure Computing, 2016 IEEE 14th Intl C 2016 Aug 8 (pp. 795–802). IEEE (2016).

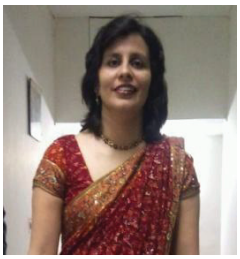
[17] Yang L. Classifiers selection for ensemble learning based on accuracy and diversity. Elisvier. 2011 Jan1;15:4266–70.

[18] Zeng X, Wong DF, Chao LS. Constructing better classifier ensemble based on weighted accuracy and diversity measure. The Scientific World Journal. 2014 Jan28;2014.

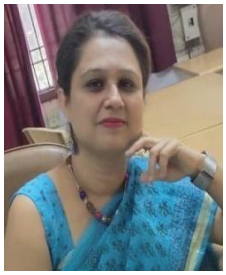
[19] Hassan MF, Abdel-QaderI. Performance Analysis of Majority Vote Combiner for Multiple Classifier Systems. In Machine Learning and Applications (ICMLA), 2015, IEEE.
[20] Sathyadevan, Shiju, and Surya Gangadharan. "Crime analysis and prediction using data mining." Networks & Soft Computing (ICNSC), 2014 First International Conference on. IEEE, 2014.
[21] Bogomolov A, Lepri B, Staiano J, Oliver N, Pianesi F, Pentland A. Once upon a crime: towards crime prediction from demographics and mobile data. 2014 Nov 12 (pp. 427–434). ACM.
[22] Tayebi MA, Ester M, Glässer U, Brantingham PL. Crimetracer: Activity space based crime location prediction. In Advances in Social Networks Analysis and Mining (ASONAM), IEEE/ACM 2014 Aug 17 (pp. 472–480). IEEE.
[23] Babakura A, Sulaiman MN, Yusuf MA. Improved method of classification algorithms for crime prediction. In Biometrics and Security Technologies (ISBAST), 2014 Aug 26 (pp. 250–255). IEEE.
[24] Retnowardhani, Retnowardhani A, Triana YS. Classify interval range of crime forecasting for crime prevention decision making. In Knowledge, Information and Creativity Support Systems (KICSS), 2016 Nov 10 (pp. 1–6). IEEE.
[25] Yu CH, Ward MW, Morabito M, Ding W. Crime forecasting using data mining techniques. In Data Mining Workshops (ICDMW), 2011 Dec 11 (pp. 779–786). IEEE.
[26] Iqbal R, Murad MA, Mustapha A, Panahy PH, Khanahmadliravi N. An experimental study of classification algorithms for crime prediction. Indian Journal of Science and Technology. 2013 Mar1;6(3):4219.

Biographies

Manisha Mudgal is a PHD scholar in Department of Computer Engineering at JC BOSE University of Science and Technology YMCA, Faridabad, India. She has done her M. Tech from M D University Haryana, India. She has successfully published 5 papers in Reputed National and International Journals. Her subjects of interest include Data Mining, Information Retrieval, and Machine Learning.



Deepika Punj is working as Assistant Professor in Department of Computer Engineering at JC BOSE University of Science and Technology YMCA, Faridabad, India. She has done Ph.D in Computer Engineering. She is having 14 years of experience in teaching. She has published more than 25 papers in Reputed National and International Journals. Her research interests include Data Mining, Deep Learning, Machine Learning and Internet Technologies.



Anuradha Pillai is an Associate Professor in the Department of Computer Engineering, JC Bose University of Science and Technology, YMCA, Faridabad, Haryana, India. She received Ph.D. in Computer Engineering from Maharishi Dayanand University, Rohtak. She published more than 60 papers in reputed international journals and successfully guided 4 PhD students. Her subjects of interest include Data Mining, Information Retrieval, Hidden web, Web Mining and Social Networks.