# Context-aware Ubiquitous Data Mining Framework to Predict Malicious Activities

[1]Zahoor ur Rehman*, [2]Muhammad Shahbaz, [3]Muhammad Shaheen, [4]Sajid Mehmood
[1]Ph.D Student, [2]Associate Professor, *Corresponding Author
[1,2]Department of CSE, University of Engineering & Technology Lahore, Pakistan
[1] Al-Khawarizmi Institute of Computer Science, UET Lahore, Pakistan
[1] xahoor@gmail.com, [2] m.shahbaz@uet.edu.pk, [3] shaheen@uet.edu.pk, [4] mayohere@gmail.com

**Abstract:** Advances in wireless sensor networks (WSN), handheld computational and communicational devices, communication systems, sensor stream processing, and data mining techniques, etc. lead us to develop such systems that perform in real-time fashion. These state of the art technologies can be very helpful to predict hazardous and criminal activities in time so that necessary actions can be taken to minimize social, economic and humanitarian loss. Context-aware computing can further increase the overall performance of the systems. A lot of research has been carried out individually in each of WSN, sensor stream processing, data mining, context-aware computing, etc. but very little attention is given to develop systems that utilize all these technologies collectively. In this paper we propose a framework to predict criminal activities and suggest suitable necessary actions on the basis its knowledge-base. The system proposed here gathers information from heterogeneous sensors, fuse that information, and generate real-time alerts to minimize the likelihood of disaster. This model utilizes real-time data from sensors; apply novel context-aware sensor stream association rule mining technique for prediction and decision tree (a machine learning technique) to take necessary action. This model makes ubiquitous data mining process more dependable and improves the reliability of the entire system. [Journal of American Science 2010;6(8):166-171]. (ISSN: 1545-1003).

**Keywords:** Ubiquitous Data Mining; Sensor Streams Mining; Association Rule Mining; Security System; Terror detection

## 1. Introduction

A large number of human fatalities and economic losses around the world daily are due to criminal activities. These activities are alarming in the modern world and thus need due attention. Daily news both from electronic and print media depict considerable numbers of lives lost everyday all over the world due to criminal activities, which disturbs the sustainability of society and makes people afraid to move around in society due to danger. Typically, busier place with increased human activity, such as shopping malls, stadiums,

busy streets, and bus stands, are more prone to criminal activities.

Information Technology can help make society more secure and sustainable. Specifically, Sensor Network Technology monitors environments by collecting and analyzing data from sensors monitoring particular environments.

The aim of this paper is to present the architecture of a Sensor Stream Mining framework that will analyze the real-time contextual information of a particular environment although such information will of course impede the computational and communicative powers of ubiquitous devices. In this paper, the contextual information analyzed in real-time includes persons' movement, actions, physical expressions, as well as current location and the importance of that location. In addition, we also considered personal demographics such as area of residence, criminal record, and security particulars to understand individual behavior and in turn design necessary precautionary measures. In the event of anomalous behavior, alerts and alarms caution security personnel to take necessary actions in a timely manner. These alerts help minimize the chances of disrupting the smooth flow of daily life and make society more sustainable as well as more dependable.

In the next sub-section of this paper we provide an overview of Ubiquitous Data Mining (UDM) and its potential for implementation in real-world scenarios. This emergent technology can play a vital role in predicting individual behavior in a desired location. Next sub-section describes the use of Sensor Stream and Ubiquitous Data Mining. After that next sub-section discusses integrating ubiquitous data mining and contextual information to predict intentions of a person. In Section 2, we present the architectural design of the system, including all sub-modules and interactions among modules. Finally, Section 3 concludes and describes implications for future research.

### 1.1 Ubiquitous Data Mining

Data mining is the process of extracting useful, hidden, and interesting patterns. Data mining technology classify data, identify clusters, and extract associations in the data. When performing data mining tasks, it is important to pay attention to the computational and communicative powers of systems. In conventional data mining systems, data is normally gathered at some central location in the form of a data warehouse to perform data analysis by incorporating statistical techniques and machine learning algorithms. The emergence of wireless and mobile devices has introduced a new dimension and enabled access to a large amount of data located at distributed and remote locations in the form of continuous streams. Ubiquitous Data Mining (UDM) is the process of analyzing data and information retrieved from remote systems on ubiquitous or mobile devices for time-critical applications [2]. Ubiquitous computing and data mining enables users to monitor, retrieve and analyze data from distributed and heterogeneous devices like sensors and mobiles [3, 4, 5, 6]. Continuous increases in the computational power of wireless and portable devices provides an opportunity to intelligibly analyze and monitor data in the form of streams by considering spatial and temporal constraints [7, 8]. The basic techniques for analyzing data and extracting hidden patterns are usually derived from traditional data mining, statistical techniques and machine learning methodologies. However, there is a need to alter those techniques in such a way that they are able to adapt on the bases of available resources. In UDM, resources are constrained in terms of memory, computational power, communication bandwidth, and availability for time-critical applications.

It is important to note that, as compared to traditional data mining, ubiquitous data mining is unable to analyze data in real-time while working in resource constrained environments with the same accuracy and quality of results. It is evident that the quality and accuracy of traditional data mining will be better than that of ubiquitous data mining, since applying the sliding window in a time-series-analysis fashion in combination with one-pass data analyzing algorithms will overcome memory and computational constraints [9, 10].

Handheld and portable devices need ubiquitous data mining software to receive continuous streams of data either from sensors that continuously read/gather data from their environment, or from external sources such as stock exchanges, web click streams, among other things. For time-critical applications, UDM modules analyze this continuous data in close to real-time and transfer the relevant retrieved information to central modules for aggregation or personalization.

Data-intensive applications are starting to appear on PDAs and cell phones such as cell-phone-based patient monitoring systems [21, 17], vehicles and driving monitoring systems [19], and wireless security systems. In the near future, some of the applications to be exercised include monitoring and analyzing data in embedded devices for smart applications, and the use of nano-scale devices for on-board monitoring. Thus, it is necessary to provide support for such applications in terms of advanced data analysis and prediction. Such applications pose various challenges and problems in order to analyze data and apply data mining techniques, which, in this domain, include:

- Developing efficient and effective techniques to analyze data in the form of continuous streams;
- Visualizing results and incorporating results on the screens of hand-held devices;
- Overcoming low bandwidth in wireless networks by introducing algorithms that minimize communication as well as number of hops. One solution is to increase dynamism in data traversing in order to minimize overhead for those devices that are under high computational activities or lacking memory [11, 5].

Further, the problem of limited batteries or battery power is another barrier to fully exploiting the potential of ubiquitous data mining. However, research has explored different methods to optimal utilization of this limited and vital resource to increase the overall lifespan of the network [14, 15].

### 1.2 Related Work

Data mining equips us with a number of techniques that can be used to model crime detection. In traditional data mining, data is gathered at some central location, after which a suitable data mining method is employed to detect useful as well as interesting hidden patterns from that data. Grouping or classifying similar data into clusters and then separating it from dissimilar patterns can help to solve crime problems that otherwise seem unsolvable by normal or manual techniques [16]. Similarly, sensor networks are used to monitor human activities at home or in offices [17].

Context-aware ubiquitous data mining is applied to detect car accidents as well as to warn drivers before the crash occurs. These precautionary alarms are generated on the basis of contextual information gathered from both on-board sensing devices and the environmental sensor network infrastructure [18, 19]. To minimize car accidents, an architectural model is developed by incorporating road

conditions and the psychological factors of drivers [1]. This early breakthrough utilizes contextual information to warn individuals about their driving and provide the necessary information to help them perform better under their current circumstances. It is expected that context awareness will decrease the number of car accidents and in turn lives lost. Detecting divergent patterns and critically analyzing those patterns in detail reveals that anomalous patterns normally cause alarming situations.

Similarly, researchers have also experimented with monitoring patients who have been in emergencies and putting their information onto mobile devices on a contextual basis [17, 21]. The relevant patient information facilitates in the treatment process as it is displayed on the handheld devices of the attending practitioner or nurses.

To detect human behavior in real time, unsupervised feature extraction is incorporated and human activities are made discrete. To preserve privacy, the analysis of human activities is performed locally on small-scale devices, considering factors such as location visited, goods purchased, tasks performed and routes adopted for traveling, etc., so that any unusual behavior can be detected by applying link analysis techniques [20].

To date, no research has attempted to combine contextual information, personal profiles, and location-based data in order to detect society's vulnerability, nor has ubiquitous data mining had the power to enable prediction processes in (close to) real-time. Therefore, combining heterogeneous sensor streams and predicting human intentions in real-time enables us to make society more secure and reliable

### 1.3 Context-aware UDM to predict uneven situations

A situation is the state of affairs of an entity, and discussing both situation and awareness collectively requires knowing about the state of affairs of an entity. Context-awareness includes information about the location, person and the environment in which the person is currently located, among other things. Compared with others in the same environment, information about a person can detect his or her behavior in the current environment, as well as any sort of discrepancy in movement. Information about the environment enables us to determine the importance of the location as well as the likelihood of criminal activities regarding the importance of location based on busyness, for example, in crowded locations such as bus stops, airports, and railway stations. Security personnel have PDAs capable of applying UDM techniques that analyze readings on contextual bases from sensors distributed around the environment. Based on such contextual information, the unusual behavior of each individual is determined.

An automatic alert informs security personals about the expected intentions of a specific person and suggests a set of actions appropriate for necessary countermeasures. Information about the person including age, race and colour, place of residence, current residence, history, purchasing habits and recent purchases, and traveling history may help determine the likelihood that a person will engage in malicious activities. Similarly, information about place or location include importance of location, activities performed there, number of people present, number of key persons or prominent folk present, and criticality of place. Combined, these things make up the scenario, and can help detect vulnerability to risks. Recognizing such situations in a manner that is timely, reliable, and cost-effective is challenging. Similarly, it is also very difficult to select and perform appropriate countermeasures in a timely and reliable manner.

## 2. Proposed Methodology

There is a great diversity of engineering and safety in life and society security areas. However, research and development have not yet considered the advantages of combining situation and environment awareness, personal bio-data and history, and criticality of location to protect society from criminal activities. To consider this shortcoming, we propose a conceptual model that employs ubiquitous data mining for situation detection in real-time. This model uses context-awareness and proposes countermeasures for the event detected. Information about these events and countermeasures will be propagated to security personnel through hand-held devices. The conceptual architecture is presented in Fig. 1.

Our approach is novel in that it integrates personal behavior, contextual information and environmental factors, historic data about criminal activities, and personal profile or bio-data. Similarly, it is also innovative in that it uses personal behavior and history with ubiquitous data mining to detect and predict the possibility of anomalous events. These factors have yet to be used collectively to assess risk factors and predict a person's intentions in a given location.

Information about the person's location and environment is acquired to measure the risk of criminal activities from each individual. To achieve this objective, we incorporate a wide range of technological devices such as sensors to gather contextual information about movement in terms of directions as

well as frequency. We define movement pattern at some specific location by observing the movement of the majority of the people in that vicinity at a given time.

In real-time, we analyze our data on PDA'S or hand-held devices available to security personnel in order to assess the risk of criminal activities. Using past patterns of criminal activity, historic data, personal profiles, and current contextual information, the model detects odd events. For example, if a person had been involved in some criminal activity, has been identified as suspicious, or his/her current contextual information clusters him in a criminal category, the model alerts security personnel to take necessary countermeasures.

Initially, we may train the model on past/historic data of criminal activities, or we may use simulated data. Subsequently, our database will continue to grow, as the number of criminal activities around the world are substantial. Patterns incorporated from past incidents provide further insight into the causes of particular incidents. One of our research goals is to extract the most important and influential factors in detecting criminal activities. In the future, if any combination of these factors is found, security personnel are alerted about the criticality of the situation before a criminal event occurs.

As uneven event detection at critical locations require quick and robust process following appropriate actions therefore in our ubiquitous framework three levels are employed. Firstly experts at monitoring centers continuously check the streaming data from the locations under observation and alerts the security personals at the location but all this process is automated. Secondly uneven events can also be detected by analyzing the streaming data collected and stored in the data centers for futuristic use. This helps to improve performance of the emergency detection system. Thirdly motion detectors and video cameras installed at location transmit abnormal motions to the devices of security personals in the same vicinity as well as transmitting those to the monitoring centers. An early action can be taken by the security staff before receiving an alert from the monitoring centers hence improving performance.

Before taking any action, the probability of an unusual event must be calculated. If the results are above a specified threshold, security personnel are alerted. On the other hand, if the results are lower than the specified threshold, but within some predetermined range, then that person will be placed under close observation to avoid the possibility of vulnerability. The success and failure rates of the alerts generated, as well as the resulting predictions will be analyzed and the main repository database will be updated. This process of continuous learning improves system performance and future decisions become more accurate and reliable.

To save energy and communication cost, data sampling is performed. Only that data is recorded which is predicted as malicious and propagated towards main repository database for futuristic use. It will decrease network traffic and utilization of communication channel is expected to optimize. This mechanism increase overall lifetime as well as performance of network.

To convert this proposed model into practical shape, several issues need consideration and catering those factors is necessary to obtain expected results as well as accuracy. Few factors are:

- **Data:** multidimensional data from heterogeneous sensors arrive in the form on continuous streams at a high rate. To deal with such a substantial amount of data in real-time is a complex challenge. Moreover, classifying this continuous stream of data by incorporating a predictive model requires access to historic data about criminal events. Normally, data about a location where criminal action has occurred is available, as is information about the person who committed the crime. However, personal information about history of the criminal is often unavailable. Similarly, the movements and actions performed before committing a crime are not available as there is no existing system to record such information. To obtain such data, a simulator is ideal at initial stages; later, continuous learning processes enable models to become more realistic.
- **Analysis:** As lightweight algorithms have already been developed which perform well in resource constraint environment but there is a need to optimize those algorithms and cater those so that data obtained from integration of heterogeneous streams can be analyzed and process in effective manner [12, 13].

**Human Rights and Legal Issues:** Unfortunately, public organizations, constitutional rights, and basic rules of independence are core hindrances in the implementation of this model. To deploy this model, constitutional shelter as well as acceptance by citizens is required, given that such a system will be perceived an invasion of their privacy due to modern ethics, which do not allow viewing an individual’s personal details without sufficient and appealing proof.

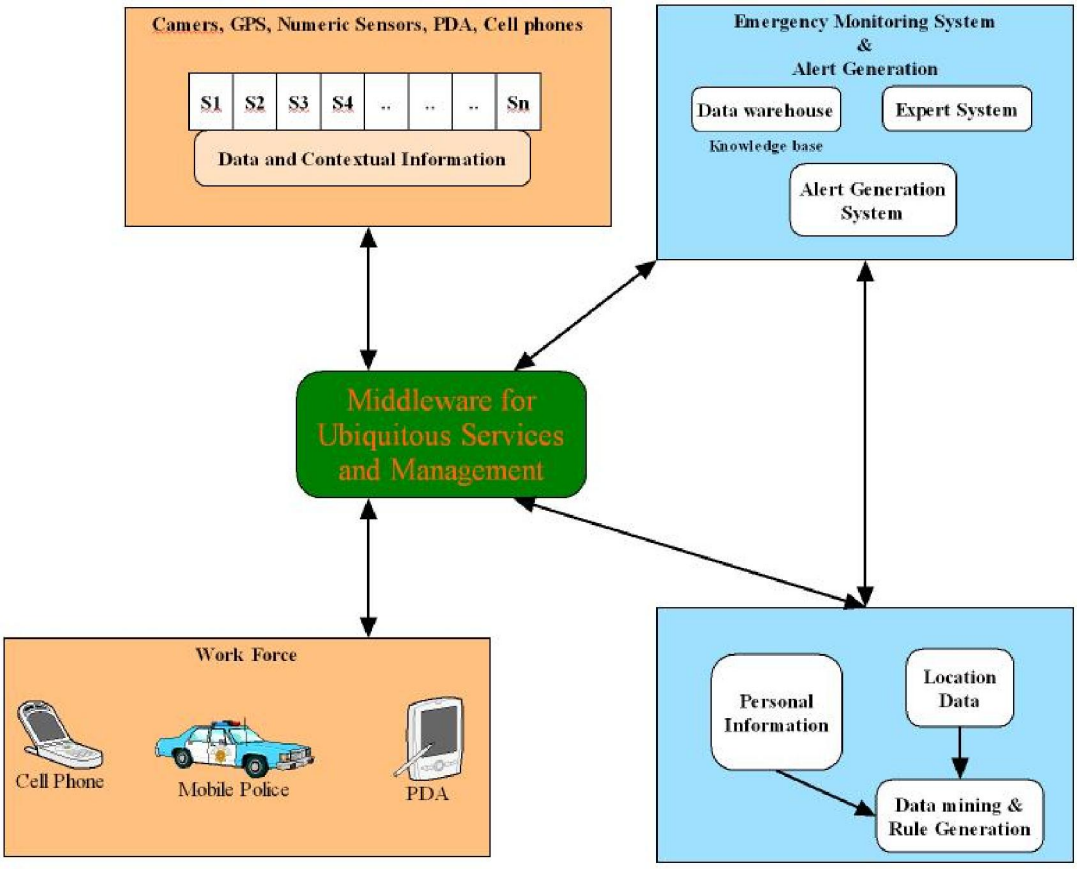


Fig. 1 Ubiquitous Data Mining and Context-awareness to Predict Uneven Events

## 3. Conclusion and Future Work

There is large toll of human fatalities and economic losses on daily bases due to criminal activities all over the world. Securities and law enforcing agencies have taken several measures to reduce causalities and financial losses to make society sustainable, dependable and more reliable. The methods applied so far have not caused some significant improvement in this regard. Therefore some advanced and alternate method to develop a sustainable society by incorporating state of the art technology is proposed. This paper presented a novel approach to detect risk by analyzing factors including movements of persons, their track record, historic data and location. Other potential factors can also be found and mined to improve overall performance and accuracy of the system. Using this model enable security personals to predict intensions of people in some specific vicinity and take necessary actions in timely manner. It is expected that this technique will bring considerable improvement in the reliability of society.

As a next step, we are planning for laboratory implementation of this system and will propose necessary actions in case of detected risks. After successful completion of laboratory implementation, real life deployment will be considered with necessary changes.

## References

[1] Krishnaswamy S. Seng Wai Loke. Andry Rakotonirainy. Osnat Horovitz. Mohamed Medhat Gaber. Towards Situation- Awareness and Ubiquitous Data Mining for Road Safety: Rationale And Architecture For A Compelling Application. Proceedings of Conference on Intelligent Vehicles and Road Infrastructure, Melbourne, Victoria 2005.

[2] Hsu Colton J. Payne D. Bhatnagar N. Woratschek C R. Data Mining Trends and Developments: The Key Data Mining Technologies and Applications for the 21st Century. In The Proceedings of ISECON 2002.

[3] Chen R. Sivakumar K. Kargupta H. An Approach to Online Bayesian Learning from Multiple Data Streams. Workshop on Ubiquitous Data Mining for Mobile and Distributed Environments, Held in Conjunction with Joint 12th European Conference on Machine Learning (ECML'01) and 5th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'01), September 3-7, 2001.

[4] Krishnaswamy S. Loke S W. Zaslavsky A. Supporting the Optimisation of Distributed Data Mining by Predicting Application Run Times. Proceedings of the Fourth International Conference on Enterprise Information Systems (ICEIS 2002) Spain. 2002, pp. 374-381.

[5] Kargupta H. Park B. Pittie S. Liu L. Kushraj D. Sarkar K. MobiMine: Monitoring the Stock Market from a PDA. SIGKDD Explorations, January 2002. pp. 37-46.

[6] Zaki, M J. Interactive and Anytime Data Mining. SIGKDD Explorations. January 2002.

[7] Krishnaswamy S. Loke S W. Zaslavsky A. Towards Anytime Anywhere Data Mining Services. Proceedings of the Australasian Data Mining Workshop (ADM02), Held in conjunction with the 15th Australian Joint Conference on Artificial Intelligence (AI02), Canberra, Australia, 3rd December 2002. pp. 47-56.

[8] Kargupta H. Park B. Pittie S. Liu L. Kushraj D. Sarkar K. MobiMine: Monitoring the Stock Market from a PDA," ACM SIGKDD Explorations. ACM Press. January 2002. pp. 37-46.

[9] Loh P.K.K. Long S.H. Pan Y. An Efficient and Reliable Routing Protocol for Wireless Sensor Network. Sixth IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks. June 2005. pp. 512-516.

[10] Nan Jiang. Le Gruenwald. Research Issues in Data Stream Association Rule Mining. ACM SIGMOD Record. 2006. pp. 14-19.

[11] Kargupta H. Park B. Fourier Spectrum-based Approach to Aggregate and Visualize Decision Trees for Mobile Applications. Workshop on Ubiquitous Data Mining for Mobile and Distributed Environments, Held in Conjunction with Joint 12th European Conference on Machine Learning (ECML'01) and 5th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'01). September 3-7 2001.

[12] Gaber M. M. Krishnaswamy S. Zaslavsky A. Cost-Efficient Mining Techniques for Data Streams. Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation. 2004. pp.109-114.

[13] Gaber M. M. Zaslavsky A. Krishnaswamy S. A Cost-Efficient Model for Ubiquitous Data Stream Mining. SIGMOD Record. 2005. pp. 18-26.

[14] Puccinelli D. Haenggi M. Wireless sensor networks: applications and challenges of ubiquitous sensing. Circuits and Systems Magazine IEEE. 2005. pp. 19-31.

[15] Zhuang L. Q. K. Goh M. Zhang J. B. The Wireless Sensor Networks for Factory Automation: Issues and challenges. IEEE Conference on Emerging Technologies and Factory Automation. 2007. pp.141-148.

[16] Shyam Varan Nath. Crime Pattern Detection Using Data Mining. Oracle Corporation. 2006.

[17] Kumar S. Fei Hu. Mark Lifson. Yang Xiao. Ubiquitous computing for remote cardiac patient monitoring: a survey. International Journal of Telemedicine and Applications.2008.

[18] Flora Dilys Salim. Krishnaswamy S. Seng Wai Loke. Andry Rakotonirainy. Context-Aware Ubiquitous Data Mining Based Agent Model for Intersection Safety. Springer Berlin / Heidelberg. 2005. pp. 61-70.

[19] Krishnaswamy S. Seng Wai Loke. Andry Rakotonirainy. Osnat Horovitz. Gaber M. M. Towards Situation-Awareness and Ubiquitous Data Mining for Road Safety: Rationale and Architecture for a Compelling Application. Proceedings of Conference on Intelligent Vehicles and Road Infrastructure. February 2005.

[20] Hunter J. Colley M. Feature Extraction from Sensor Data Streams for Real-time Human Behaviour Recognition. Springer-Verlag Berlin Heidelberg. 2007. pp.115-126.

[21] Hsu-Yang Kung. Mei-Hsien Lin. Chi-Yu Hsu. Chia-Ni Liu. Context-Aware Emergency Remedy System Based on Pervasive Computing. Springer Berlin/ Heidelberg. 2005. pp. 775-784.

5/1/2010