Journal of Theoretical and Applied Information Technology
10th January 2013. Vol. 47 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EMERGING AND PROSPERING TRENDS IN CRIME ANALYSIS AND INVESTIGATION SYSTEMS: A LITERATURE REVIEW

[1]RIZWAN IQBAL, [2]MASRAH AZRIFAH AZMI MURAD, [3]NURFADHLINA BINTI MOHD SHAREF

[1,2,3] Faculty of Computer Science and Information Technology, University Putra Malaysia, Malaysia.

E-mail: [1]mail.rizwaniqbal@yahoo.com, [2]masrah@fsktm.upm.edu.my,[3]fadhlina@fsktm.upm.edu.my

ABSTRACT

Crime analysis and investigation systems play a critical role in battling crime and war against terrorism. This paper is a detailed literature review on current crime analysis and investigation systems developed to date. The reviewed literature is divided into two broad categories, "ontologies" and "clustering and classification". The paper encompasses a detailed discussion on systems belonging to both categories. During literature review, it was discovered that over the years evolving crime systems took inspiration from their predecessors, in terms of functionality and features. Therefore, some interesting trends were identified. These trends gradually prospered with the passage of time, strengthening the capabilities of the evolving systems. This paper discusses these trends in detail, and also anticipates the key features and design for the future generation of crime systems.

Keywords: *Crime Analysis And Investigation Systems, Ontologies, Clustering, Classification, Topic Detection And Tracking*

1. INTRODUCTION

A few decades back crime used to be solved and investigated by crime analysts, detectives and law enforcement officers. The advancement in technology today has made computer science experts to be part of crime investigation teams [1]. The modern day crime analysis and investigation systems make use of sophisticated computer science algorithms and techniques to develop a better and narrow understanding of the committed crimes. These systems not only allow analyzing the past and present committed crimes but also empower the user to predict future crime patterns. Thorough crime analysis of committed crimes and prediction of future crime patterns provide invaluable assistance to the law enforcing agencies, to effectively battle crime and war against terrorism.

There are distinct types of crime analysis and investigation systems developed to date. From the reviewed literature, two distinct categories of crime analysis and investigation systems have been identified. These categories employed different computer science techniques /platforms to carry out crime analysis and investigations tasks. The categories are "ontologies" and "clustering and classification".

Topic Detection and Tracking (TDT) is a research program being carried for last few years. TDT systems can be effectively used to analyze and investigate crime stories. This paper has covered TDT systems under the category of "clustering and classification", as they rely on clustering for grouping similar stories. Evolving systems have taken inspiration from their predecessors in terms of functionality and features. Over the years, there has been a noticeable evolution and transfer of trends in systems within the same category as well as among the categories.

This review also discusses the trends which evolved through the passage of time and how they impacted the design of upcoming systems. The literature review revealed trends including "notion of event", "effective visualization" and "effective user interaction". However the "notion of Event" was found to be one of most critical trends which evolved over time. The "notion of Event" trend was found to gradually develop its worth and strength in both categories of crime analysis and investigations systems, reviewed for this research. Recent and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

evolving crime analysis and investigation systems lay a lot of focus and stress on "notion of event" while designing their systems. This also helps in efficient information retrieval for the user, allowing the user to think of crime as an event, having a set of attributes.

The paper is organized in the following sequence. Next section is related work, which will be discussing some systems related to different categories of crime systems. Then there will be the discussion chapter, discussing the emerging and prospering trends in the field of crime systems. The last section will be conclusions and future work.

2. RELATED WORK

This section discusses about different crime analysis and investigation systems developed over the years. This system will discuss the systems category wise, so it would be appropriate to briefly discuss about the broad definition of each category first and then discuss the systems that belong to each category.

2.1 Systems Using Ontologies

Ontologies have been discussed and debated upon since the times of Plato and Aristotle [4]. It became common in the field of computer since the last two decades. Researchers and scientists have exploited ontologies to bring improvement and efficiency in various areas of computer science and technology. Ontologies have its implementation in the area of internet searching, retrieval of multimedia contents from archives, artificial intelligence, robotics, medicine, location aware services, semantic web and many other areas including crime analysis and investigation.

Reasoning and inference are two important tasks which are performed on ontologies [5]. The effectiveness and correctness of reasoning and inference is totally dependent on the quality of ontology. In the case of ontology, quality is referred as the correctness in defining the concepts and the relationships among those concepts. In order to carry out complex reasoning it is important for the ontology to be rich in semantics[5]. Ontology creation is not just time consuming but also a complex and tedious job. The maintenance of consistent level among the siblings and coherent relationships between the ancestors and descendants in the taxonomy makes ontology construction a complex and time consuming job [6].

Zhdanova reported about the semantic web portal project which is a semantic meta-portal for managing and accessing information on people [7]. This system focuses on the criticality to update and validate information on individuals. The system effectively manages the current information on individuals, but the drawback is that it does not provide any provision to keep a track on past information. This ignorance is a critical lacking in web portal pertaining information on individuals; it does not keep in to account of any temporal information about the individual.

Another work is constructing ontologies for information of LiveJournal by using FOAF (Friend of a Friend) project [8]. This work allows defining an individual in multiple ways; it constructs a kind of FOAF (Friend of a Friend) linking independent of time. Although this work allows defining an individual in many ways but the use of FOAF restricts to represent any temporal information. Another work from Joint Research Centre (JRC) used ontology to represent relationships between criminals. This system empowers the user to query the knowledge base on the basis of organization, event, place and person[9]. However this work has left the temporal information integration as a future work.

In 2005, Lin & Liang proposed a design for retrieving digital archives from religious text [10]. This work incorporated the notion of event. They created ontology for Bible and demonstrated that how event ontology can be used along a book ontology for efficient retrieval of multimedia contents and verses, on user demand. The results were encouraging, incorporation of the notion of event undoubtedly brought improvement in the overall process of the system.

Another system related to finding and reconstructing information on individuals was revealed in 2007 [3].This work employed the concept of event ontology to store temporal information about individuals. The proposed event ontology defines an event on the basis of "who", "when", "what", "where", and "how". The system stores the current information as well as past information of an individual. The flexibility in the event ontology empowers the user to efficiently retrieve information on individuals based on temporal information.
Evaluation of the system by Han et al. manifests that when a user makes a generic web search about an individual, lot of unrelated material and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

documents pops up making search an irritating and tedious job [3]. Experimental results demonstrated that using event ontology effectively addresses this problem and it can accurately retrieve information on an individual by keeping the temporal variable in mind.

In 2010, Cunhua, Yun, & Zhaoman proposed a system using event ontology, specifically for mining cybercrimes in Chinese web pages [11]. The proposed system used ontology to describe cybercrimes at the level of event, relation and event class. The proposed system operates in 4 steps. Step 1: extracts web text from websites and converting it to clean text (Stop word removal, POS tagging). Step 2 and 3: feature reconstruction is performed on candidate features by feature compression (meaning merging some features using ontology) and feature expansion (meaning append some features using ontology). Step 4: After feature reconstruction data mining is performed including classification, retrieving and associative discovery.

This system uses the event ontology in the feature reconstruction phase [11]. Even, if the document contains an event with missing elements, the event ontology can find a clue for the missing elements using feature compression and feature merging techniques. The paper has discussed about comparing the data mining results with and without the event ontology [11]. The results depicted that using event ontology in the feature reconstruction phase brought considerable improvement in the overall data mining process (using SVM classification for text).

The systems discussed above exploited the platform of ontologies in order to facilitate crime analysis and investigation tasks. The systems discussed above underwent improvements and adopted certain trends over the passage of time. A noticeable trend that prospered over time in this category is the "notion of event". This trend emerged as a vague thought, than gained worth and strength over time.

2.2 Systems Using Clustering And Classification

This category of crime analysis and investigation systems is broadly covered as clustering and classification which comes under the umbrella of data mining. Data mining is defined as the discovery of interesting structure in data, where structure designates patterns, statistical or predictive models of the data, and relationships among parts of the data [12]. In the domain of data mining, clustering is a process of grouping similar data points. Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data.

In 1998, Brown proposed a software data mining framework to catch criminals [13]. In designing his data mining framework Brown kept special focus on spatial data mining. He stressed and justified the importance of spatial data mining over mining other attributes. He analyzed that spatial data mining has the potential to yield immediate benefits for crime analysis. One of the critical benefits of spatial data mining is effective law enforcement resource allocation.

Chen et al. used clustering techniques to identify suspects who conduct crimes in similar ways [14]. In their COPLINK project they demonstrated the use of named-entity extraction, deceptive-identity detection, and criminal-network analysis using their proposed data mining framework. Results showed that the system successfully unveiled target gang members aided by visualizations, allowing viewing number of members in groups and sub groups. A drawback in this system is the dependency that it generates static networks rather than dynamic networks.

Another system related to crime pattern detection was proposed by Nath [1]. This system preferred clustering over techniques like classification (supervised). Nath chose clustering because classification relies on known and solved crimes, lacking the ability to predict future crimes. This research introduced a weighing scheme in clustering, using k-means. It allows applying distinct weights on different attributes dynamically based on the crime types being clustered.

The clusters are displayed along a geo spatial plot. The system gives the flexibility to choose a time span and types of crime on the geography. This selection becomes the input source for data mining. The results are clustered on the basis of weighted attributes [1]. However, weights are assigned by experts; different experts will give distinct weights which make the result subjective and less convincing [15].

Finding similar crime case subsets is a critical but tedious job in crime analysis. It not only provides clues but also improves the efficiency to catch

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

criminals. Ma and Y. Chen 2010 proposed an approach which use weights with IGR (Information gain Ratio) to classify the attributes (classification) and then use it for clustering to find similar case subsets [15]. The proposed approach came up with a variation in classical k-means, it performs clustering based on attribute weights.

Experimental results demonstrated that weighted attributes resulted in higher precision than the non-weighted attributes. On the basis of results, it can be concluded that using weighted attributes can improve the accuracy [15]. Despite the fact that it improves accuracy, limitation is that the user has to set the threshold manually for 'α' which is very critical in compiling the results. Therefore, the approach remains dependent on the user and it requires human intervention for critical decision making.

Malathi and Santhosh Baboo recently proposed a data mining approach employing clustering and classification for predicting crime trends [12]. This system particularly addresses the issue of missing values in the police records. The proposed approach fills the missing values without hampering the clustering process. The approach relies on EM (expectation maximization) clustering which is a deviation of k-means. The specialty of EM (expectation maximization) is that it works well with missing or unidentified values.

The system operates in two phases [12].First phase deals with resolving the issue of missing values followed by the next phase dealing with the prediction of future crime trends. The approach helps to anticipate crime patterns and trends. Prediction of missing values is the strength of this work. However, it is also the weakness as it not known whether the predicted crime trends are reliable or not.

Topic Detection and Tracking (TDT) is a research program being carried for last few years. It investigates methods for automatically organizing news stories based on events discussed in the stories. It can be effectively used to organize crime stories. TDT research includes several evaluation tasks, each of which explores an aspect of the organization of a continuous stream of news. Most research in the area is based on evaluating the performance of algorithms to perform evaluation tasks on the stream of news. Most research on TDT is carried out in the laboratory, isolated from real world users and their interaction. Therefore it lacks the element of user interaction.

In order to remedy the lacking of user interaction in TDT, concept of Interactive Topic Tracking and Detection (iTDT) was introduced. Allan et al introduced Event Organizer, an interactive system that organizes a constantly updating stream of news articles [16]. The organization is based on events discussed in the stories already processed by the system. This system does not only rely on clustering news events, but also provides a user interface which allows the user to view documents with the timeline. The system privileges manual user intervention, user can make corrections by removing stories from clusters and create new clusters.

TDTLighthouse is another system belonging to the category of iTDT [17]. The system shows the result for each search session to the user. A noble feature in the system is that besides displaying a typical list of ranked results, it enables the user to effectively visualize the similarity between documents. The visualization is in the form of floating spheres which are positioned from each other depending on the inter document similarity. The idea of floating sphere visualization effectively aids the user to understand the similarity between the documents in little time [17].

Another system from the same genre is TimeMine [18]. It groups the documents on the basis of their statistical properties. The system display the major events using a timeline based interactive approach, additionally the timeline is also used as an interactive interface to browse the documents. Mohd, Crestani, and Ruthven proposed a system called IEvent, this system took inspiration from systems like Event organizer [16], TDTLighthouse [17], TimeMine [18] and etc. IEvent uses the Named Entity (NE) recognition approach to further enhance the effectiveness of topic tracking task.

This section has discussed some systems from the broad category of "clustering and classification", facilitating crime analysis and investigation tasks. It was identified that a few trends gradually prospered and gained strength with the passage of time, bringing improvements in the systems. The identified trends are "notion of event", "effective visualization" and "user interaction". The identified trends will be discussed later, in the next section.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. DISCUSSION

After reviewing the two categories of crime analysis and investigation systems, some trends were identified. These trends gradually prospered and gained strength with the passage of time, bringing improvements in the systems. Evolving systems have taken inspiration from their predecessors in terms of functionality and features. Over the years, there has been a noticeable evolution and transfer of trends in systems within the same category as well as amongst the categories.

The identified trends are "notion of event", "effective visualization" and "effective user interaction". While going through the literature, it was found that the above trends started as just simply a new idea or a feature for analyzing crime. Later these features proved their significance in analyzing crime and became inspiration for the upcoming generation of systems. This section will individually cover the identified trends in detail, discussing that how these trends have shaped and refined over time.

3.1 Notion Of Event

The notion of event is found to be highly associated with crime [2]. When a crime news is reported or a police report is generated, it is a usually a combination of four "W's" and one "H" related to an event. The four W's stand for “who", "when", "what", "where” and "H" stands for "how" about the event. In the category of ontologies, work by Anna, 2004 was based on maintaining individual information. This system lacked the provision for keeping past information about individuals. John et al, 2005 realized the importance of defining individuals in multiple ways but the use of FOAF (Friend of a Friend) restricted to represent any temporal information. There was a need to overcome this gap, and a mechanism was required to handle temporal information. In 2005,Wennerberg from Joint Research Centre (JRC) permitted the user to query criminal knowledge base on the basis of event [9]. Work by Wennerberg talked about event but still the idea was not that convincing, as the potential of “notion of event" was still not unleashed.

Another work was introduced, using event ontology for retrieving digital archives from religious texts [10]. This work demonstrated the potential of applying event ontology for effective information retrieval. Finally, Han, SY Park, and SB Park clearly proposed event ontology for finding and reconstructing information on individuals [3]. It defined an event on the basis of "who", "when", "what", "where”, and "how". The flexibility in the event ontology empowered the user to efficiently retrieve information on individuals based on temporal information. Later, Cunhua, Yun, and Zhaoman uniquely used ontology to describe cybercrimes at the level of event, relation and event class [11]. This system used event ontology to find clues for missing elements during feature reconstruction phase.

In Parallel, “notion of event" also paved its way in the category of "classification and clustering". Work by Ma and Y. Chen talked about finding similar crime case subsets in the context of events [15]. Similarly, system by Malathi and Santhosh Baboo predicted future crime patterns in terms of events [12]. I-JEN, an interactive crime news retrieval system came to scene recently, validating the importance of events in the domain of crime [2].

3.2 Effective Visualizations

Visualizations have always played a critical role in analyzing and investigating crime. It allows the user to analyze huge amount of data in a couple of seconds, revealing trends and patterns which could have never been realized otherwise. During literature review, it was identified that over the year’s crime investigation systems adopted different visualization techniques to represent crime data. The main focus for adopting new visualizations over time was to make sure that user is able to develop the best understanding of the data in minimum possible time.

The availability of "document timeline" visualization allows exploring documents in a group on the basis of their occurrence, with respect to time. The visualization feature of "document timeline" evolved as a critical need to browse similar documents within the same time span. This feature helps to analyze similar documents easily, especially when there is a requirement to compare attributes within similar documents. Cunhua, Yun, and Zhaoman introduced "document timeline" to display the results of web crime mining [10].

Nath [1] and H.chen et al [paper 11] developed systems for crime data mining. Their systems performed data mining functions, assisted with visualizations. The visualizations helped the user to visualize data mining results, but lacking of a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

"document timeline" limit them to further explore the inter similarity between the crime records. Ma and Y. Chen introduced a document timeline in his work, related to finding similar crime case subsets [14]. They required a more detailed and in depth analysis of crime data, to further compare the attributes of similar crime records. Providing a document timeline privileged the user to have deeper insight of similar documents, enabling the user to realize hidden nature of trends [14].

Systems belonging to the domain of Topic Detection and Tracking (TDT) heavily relied on "document timeline". TDT systems [15], [17], [18] effectively used "document timeline", enabling the users to explore similar crime stories on a timeline. Systems like, Event Organizer intuitively used document timeline not only as an interactive interface to browse similar stories, but also make corrections in the clusters at the same time [15]. Undoubtedly, using document timeline provides ease to the user in analyzing similar documents.

Another kind of visualization in TDTLighthouse [16] was found to be effective [18]. This system provides visualization in the form of floating spheres, which are positioned from each other depending on the inter document similarity. Such visualizations are very effective in understanding inter document similarity, in almost no time.

3.3 Effective User Interaction

In systems related to crime, user interaction plays a critical role. The thing that made user interaction more critical for crime domain is the sensitivity of the data being analyzed by the user, on the basis of which critical decisions are drawn. Mohd, Crestani, and Ruthven discussed about this factor in their work [18]. The experimental results from their work validated the idea of effective user interaction [18].

In 2011, Ali et al. came up with the idea of applying Usability Engineering lifecycle in the designing phase of a crime system [19]. The system applied the Usability Engineering lifecycle to assure that usability is fully induced in the system design [19]. In order to comply with the Usability Engineering lifecycle, qualitative analysis was performed using surveys and questionnaires. This System is still in under construction and not operational yet [19].

4. CONCLUSION AND FUTURE WORK

Over the years, computer science has facilitated the area of crime analysis and investigation to carry out distinct crime analysis and investigation tasks. Improvements and innovations emerged in all the categories of crime systems. From existing literature in the field, this research identified some trends which gradually prospered and gained strength with the passage of time. These trends started as simply a new idea or a feature for analyzing crime. Later these features proved their significance in analyzing crime, becoming inspiration for the upcoming generation of systems. The identified trends are "notion of event", "effective visualization" and "effective user interaction".

After a thorough review, it can be anticipated that the key features and designs for the future generation of crime systems will be deeply influenced by the identified trends. Future crime systems will be heavily relying on the notion of event, as an event is identified by event triggers, and is associated with participants, time, location and others parameters. It is a larger semantic unit compared with a concept.

Additionally, future generation of crime systems will be focusing critically on adopting and exploring new visualization techniques. New visualization techniques are expected to be more effective for the user, meaning that they should be able to deliver better insight of data in minimum possible time. Adopting newer visualization techniques will allow the users to develop a clear and deeper understanding of the crime data being analyzed.

Lastly, the element of effective user interaction will be dealt more seriously, compared to old systems. Experimental results have shown that effective user interaction is an important factor in evaluating the overall efficiency and performance of a system. Upcoming systems will be taking advantage of new techniques and methodologies to induce usability to the system design, enabling the users to have the best possible interaction experience. It is expected that future systems will be better facilitators for crime analysis and investigation tasks, compared to current systems.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFRENCES:

[1] S. V. Nath, "Crime Pattern Detection Using Data Mining," in 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2006, vol. 1, no. 954, pp. 41– 44.

[2] I. Horrocks, "Semantic web: the story so far," in W4A '07 Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A), 2007.

[3] M. Sabou, C. Wroe, C. Goble, and G. Mishne, "Learning domain ontologies for web service descriptions: an experiment in bioinformatics," WWW '05 Proceedings of the 14th international conference on World Wide Web, pp. 190–198, 2005.

[4] H. Yang and J. Callan, "Ontology generation for large email collections," in Proceedings of the 2008 international conference on Digital government research (dg.o '08), 2008, pp. 254–261.

[5] A. V. Zhdanova, "People ' s Community Semantic Web Portal and Metaportal," 2004.

[6] J. C. Paolillo, S. Mercure, and E. Wright, "The Social Semantics of LiveJournal FOAF: Structure and Change from 2004 to 2005," Proceedings of the ISWC 2005 Workshop on Semantic Network Analysis, vol. 72, no. 7, pp. 69–80, 2005.

[7] P. O. Wennerberg, "Ontology Based Knowledge Discovery in Social Networks," European CommissionJRC, 2005.

[8] H. Lin and J. M. Liang, "Event-based Ontology design for retrieving digital archives on human religious self-help consulting," in 2005 IEEE International Conference on e-Technology, e-Commerce and e- Service, 2005, no. 1, pp. 522–527.

[9] Y. Han, S. Park, and S. Park, "Reconstruction of people information based on an event ontology," in International Conference on Natural Language Processing and Knowledge Engineering, 2007, pp. 446–451.

[10] L. Cunhua, H. Yun, and Z. Zhaoman, "An event ontology construction approach to web crime mining," Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), vol. 5, no. Fskd, pp. 2441–2445, 2010.

[11] A. Malathi and S. Santhosh Baboo, "An Intelligent Analysis of Crime Data for Law Enforcement Using Data Mining," Data Engineering, no. 1, pp. 77–83, 2011.

[12] D. Brown, "The regional crime analysis program (RECAP): a framework for mining data to catch criminals," in 1998 IEEE International Conference onSystems, Man, and Cybernetics, 1998, no. 3, pp. 2848 – 2853.

[13] H. Chen, W. Chung, J. J. Xu, G. Wang, and Y. Qin, "Crime data mining: a general framework and some examples," Computer (New York), vol. 37, pp. 50–56, 2004.

[14] L. Ma and Y. Chen, "AK-Modes: A weighted clustering algorithm for finding similar case subsets," in 2010 International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2010, pp. 218 – 223.

[15] J. Allan, S. Harding, D. Fisher, A. Bolivar, S. Guzman-Lara, and P. Amstutz, "Taking topic detection from evaluation to practice," in 38th Annual Hawaii international Conference on System Sciences (Hicss 2005), 2005, vol. 4, pp. 1–10.

[16] A. Leuski and J. Allan, "Lighthouse: Showing the way to relevant information," IEEE Symposium on Information Visualization 2000 (InfoVis'00), pp. 125–129, 2000.

[17] R. Swan and J. Allan, "Automatic generation of overview timelines," in SIGIR conference on Research and development in information retrieval. SIGIR '00, 2000, pp. 49–56.

[18] M. Mohd, F. Crestani, and I. Ruthven, "Design of an Interface for Interactive Topic Detection," Flexible Query Answering Systems, Lecture Notes in Computer Science, vol. 5822, pp. 227–238, 2009.

[19] N. M. Ali, M. Mohd, H. Lee, and A. Smeaton, "i-JEN: visual interactive Malaysia crime news retrieval system," in Proceedings of the Second international conference on Visual informatics: sustaining research and innovations, 2011, pp. 284–294.