AARCTIC: Autonomic Analytics Research for Corrections Technology, Institutional and in the Community

Catherine McFarland, Roy Sterritt
School of Computing and Mathematics
University of Ulster
Northern Ireland
mcfarland-c2@email.ulster.ac.uk,
r.sterritt@ulster.ac.uk

David Bustard, Sally McClean
School of Computing and Information Engineering
University of Ulster
Northern Ireland
dave@infc.ulst.ac.uk,
si.mcclean@ulster.ac.uk

Patricia O'Hagan
Core Systems
Belfast
Northern Ireland
patricia@coresystems.biz

***Abstract*—This work is essentially concerned with Predictive Intelligence for Corrections. The best predictor of future behaviour is past behavior, and is the premise behind predictive analytics. In essence it involves identifying predictors and patterns that can suggest a possible outcome. In human-activity situations prediction can be more difficult due to the inherent fickleness of human behaviour. However, in controlled environments such as correctional facilities a fairly consistent commonality in predictors exists that could be mapped to a computer system. The vision for corrections is to harness all existing electronic data available in a given facility and employ predictive analytics to successfully identify hotspots and pre-empt disturbances and incidents. The research hypothesis behind this research project is to adapt the techniques of data fusion and predictive analytics with the concepts surround big data velocity and autonomics to facilitate near real-time automated predictive intelligence, that being Autonomic Analytics.**

This paper examines predictors of disruptive behaviour followed by the relevant elements of big data analytics, data fusion and predictive analytics. It concludes by considering an area of research contribution utilising autonomics.

***Index Terms*—Component, formatting, style, styling, insert.**

I. Introduction

The vision for corrections in the United States is to harness all available electronic data available in a given facility and employ predictive analytics to successfully identify "hotspots" and pre-empt disturbances and incidents, e.g. riots, assaults, gang fights and contraband.

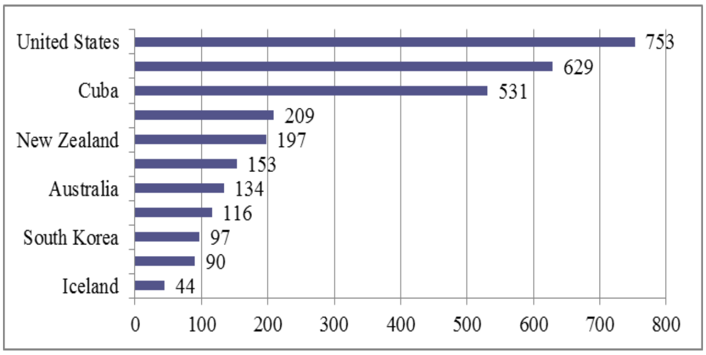
This paper gives a high-level overview of the American corrections environment and the challenges it faces, both politically and relative to the housed inmates. The research has included studies into predictors of disruptive behaviour.

Having considered the more advanced inmate management systems (IMS) and the data they store, the paper then looks at the key concepts and methodologies for data fusion, big data analytics, predictive analytics and finally, autonomics. We consider how these concepts can be amalgamated to aid corrections staff in monitoring and predicting behaviour as well as retrospective investigation upon realization of an incident.

II. The Corrections Environment

The United States of America has the largest prison population per capita in the world [9] (Figure 1). Conversely with the continuing financial environment corrections budgets are continuing to be cut with only 3.5 % of a state's budget being dedicated to corrections [7] while an estimated 1 in 33 adults are supervised in a correctional facility (a total of 1,314,446 inmates were held in state prisons in 2011 alone). [5] This has encouraged the corrections industry to look to automate their systems as much as possible to better manage their growing populations.

Figure 1. Incarceration rate per 100,000 (2006-2009)[1]



Within corrections, intelligence can be defined as the result of collecting and analysing multi-source data within a specific context, with the aim of identifying indicators of unwanted behaviour. These indicators and human-based conclusions can then feed decisions, actions or strategies either preventatively or for retrospective investigation. Valuable intelligence is considered the process of "connecting the dots", identifying

[1] Figures derived from "The High Budgetary Cost of Incarceration", J.Schmitt, K.Warner, S.Gupta

relationships in seemly disparate data or information which is crucial in predicting danger in a dangerous environment.

The challenge faced by corrections is that despite the volume of data available there are few mechanisms to identify these relationships and so predict problems and events, or even to gather this data (which is often split among multiple vendor systems), into one cohesive, clean pool suitable for automatic analytics.

Current market research has shown, that there are only a small number of intelligence applications available to the corrections industry. The majority of these focus on data gathering and graphical representation but rely heavily on user interaction to gleam "intelligence" from the data presented. Those that offer predictions through analytics have either been custom designs with a long pilot phase or come from law enforcement and as such are far removed from being a perfect fit for corrections.

Offense	Number	Per cent
Total	**1,362,028**	**100**
Violent	**725,000**	**53.23**
Murder	166,700	12.24
Manslaughter	21,500	1.58
Rape	70,200	5.15
Other sexual assault	90,600	6.65
Robbery	185,800	13.64
Assault	146,800	10.78
Other violent	43,400	3.19
Property	**249,500**	**18.32**
Burglary	130,000	9.54
Larceny	45,900	3.37
Motor vehicle theft	15,000	1.10
Fraud	30,800	2.26
Other property	27,700	2.03
Drug	**237,000**	**17.40**
Public-order	**142,500**	**10.46**
Other/unspecified	**7,900**	**0.58**

TABLE 1. ESTIMATED NUMBER OF PRISONERS UNDER STATE JURISDICTION BY OFFENSE FOR 2010[2]

Current research highlights some key demographic characteristics [8]. Of those the following would typically be recorded in an inmate management system: age, race, criminal history, conviction type, length of incarceration and gang affiliation. Prison gangs are considered particularly high risk. As gang culture becomes more prevalent on American streets so it grows within their prisons. Worrall et al. [10] found a strong correlation between gang membership and inmate-on-inmate violence. In particular they noted that the extent of violence rose with the degree of gang integration which supported Schenk's findings.

Though not listed by Schenk a study from the Journal of Criminal Justice [6] found that inmates with a history of drug use were more likely to commit rule violations than those without, though he felt further research is required in this area. Again, Jiang agreed with Schenk that more powerful indicators may result from correlating their indicators with data on age, criminal history and incarceration history, as do Baskin and Sommers [1]. In their consideration of the impact of inmates with mental health problems (half the inmate population [6], on violence against the self, other inmates and property, they also felt that a combination of factors would prove more conclusive. This has in turn been put into practice within California's Inmate Classification System [4].

Offense	Number	Per cent
Total	**190,641**	**100**
Violent	**15,000**	**7.87**
Homicide	2,900	1.52
Robbery	8,300	4.35
Other violent	3,800	1.99
Property	**10,300**	**5.40**
Burglary	400	0.21
Fraud	7,500	3.93
Other property	2,400	1.26
Drug	**99,300**	**52.09**
Public-order	**65,000**	**34.10**
Immigration	20,200	10.60
Weapons	29,200	15.32
Other	15,600	8.18
Other/unspecified	**1,100**	**0.58**

TABLE 2. ESTIMATED NUMBER OF PRISONERS UNDER FEDERAL JURISDICTION BY OFFENSE FOR 2010[1]

Finally, consideration should be given to prison procedures. Brierie [2] identified an increase in violence whenever there was a reduction in support staff, e.g. teachers, counselors, but also an increase in violence when there was a greater number of custody staff. His work also noted that slow or biased responses to grievances lodged by inmates created a sense of powerlessness that increased frustration and hence violence - a phenomenon that is graphically highlighted in the 1980 New Mexico Prison Riot [3] – the inmate's need for fairness.

[2] Figures taken from U.S. Department of Justice Bulletin "Prisoners in 2011", E.A. Carson, W.J. Sabol

III. Technologies Required for Predictive Analytics

To enable the corrections industry to automate their intelligence gathering and production will rely on some key technologies around handling, monitoring/sensing, analyzing and acting on multi-vendor heterogeneous data. To facilitate the mining of multi-vendor heterogeneous data a combination of big data analytics, data fusion and autonomic concepts will need to be considered along with what the research has determined to be reliable predictors of disruptive or violent behaviour within correctional institutions. This section considers each of these areas as it pertains to the vision of this research.

A. Data Fusion

Though there is no universally agreed definition of an information fusion system (IFS) it is agreed that to be considered an IFS it must receive information from a number of different sources. Essentially, the purpose of an IFS is to have more or better information as a result of the fusion that existed before. [20]: hence its value in a predictive intelligence system.

In their comprehensive look at data fusion Bleiholder and Naumann [12] consider the importance of conciseness and completeness within the integrating information system. That is, ensuring all valued data is represented in the smallest number of variables and objects. By doing this we create a concise data set that will more easily satisfy the velocity requirement of big data.

Nilsson and Ziemke [20] see value to be gained through the increase in data sources. This can also increase confidence, accuracy and robustness in the data through cross-referencing the same data from different sources. Allen's "mixed-initiative interaction" [11] could be considered in creating an IFS to maintain the human in the loop.

The JDL model [16] is the most popular data fusion concept model. Its military based focus on data rather than the framework has proved restrictive. Kokar et al. [19] presented a more recent abstract framework that offers sufficient generality to capture multiple fusion types, e.g. data fusion, feature fusion and decision fusion. Unlike the JDL model it can express both data and processing.

Current research in data fusion is considering areas such as the human-centered data fusion paradigm [17] relating back to mixed-initiative, and the notion of "certainty about uncertainty" [18]. Here researchers are attempting to use reliability co-efficients to define a second level of uncertainty. Relating back to the fundamental principle of data fusion (conciseness and completeness), is research into developing approaches that can determine the reliability and credibility of the data to establish a degree of confidence [16]. Such a notion could then feed through to the degree of confidence in the ultimate analytic process. NATO's STANAG 2022 standardization agreement is considered a significant work in this area that researchers are hoping to extend [15].

Pichon et al. [21] consider performance in regard to the information's "relevance and truthfulness". A particularly interesting aspect of their work is that it can be applied "to all domains where information sources are intelligence agents able to lie". While semantic matching is yet beyond the scope of a predictive intelligence system, an approach that could identify inconsistencies in human narrative could prove valuable in future systems. Khaleghi et al. [21] consider information rate as a much neglected area of research. Based on the close relationship between data fusion and big data the exclusion of "velocity" should make it an area of focus in future work.

Multi-sensor fusion is the use of data fusion to combine data from multiple sensors and any related database information to create more meaningful insight. In 1997 Hall and Llinas [15] saw real-time multi-sensor fusion as a realistic goal for data fusion through the advancements in technology. A recent paper by Guivant et al [14] is an example of this fulfillment through the use of real-time fusion in autonomous 3D mapping. This study employed real-time fusion from three sources. The work included consideration of timestamps and how latencies could be handled to avoid negative effect on the 3D image synthesis.

B. Big Data Analytics

Big data has evolved due to the increased digitization of data and the advancement of analytics technologies. It aspires to garner intelligence from data and translate it into a business advantage [28], or in the context of this work to offer an "intelligence" advantage to security staff. It is characterized by the 3 Vs.: volume, velocity and variety [26]. Later work has extended the 3V model to include veracity – the certainty or reliability placed on certain types of data; and complexity [25].

Data fusion can be used to increase big data reliability through the combination of less reliable sources to create a singular, more accurate data set [29], but Bollier [22] warns that combining data from multiple sources could magnify existing issues with the data.

In their 2012 study into the corporate use of big data Schroek et al. [29] found organizations were using big data "to target customer-centric outcomes, tap into internal data and build a better information ecosystem". They found that 63% of their 1144 respondents were creating competitive advantage through leveraging big data and analytics. Their findings were similar to those of Le Valle et al. [27], who found an understanding of how to leverage analytics and access to data two common obstacles in big data analytics.

Boyd and Crawford [23] point out that research into big data forces accessibility challenges. Those granted permission to large data companies may feel obliged to ask questions whose answers are favorable to the company. While those without are forced to use synthetic, over cleansed data which may be biased towards their hypothesis [30].

Fisher at al. [24] defines big data analytics as "a workflow that distils terabytes of low-value data…down to…a single bit of high-value data". They offer as consideration of future work drip-feeding analysis results to users to speed up turnaround rather than waiting for the full completed analysis. Kaisler et al. [25] feel that further research is required into the storage, management and processing of big data. They see predictive analytics as the new normal in processing big data, but recognize that some analytics may not scale to the anticipated zettabytes. They predict that big data will continue to get bigger and our need for it will grow as we move towards advanced analytics, systems that learn.

C. Predictive Analytics

"The future of Data Mining lies in Predictive Analytics" [38]. As early as 1994 predictive analysis has been used to fight crime [31]. The Predictive Analysis System used a model based approach to predict drug trafficking events using tree node comparison. More recently advanced analytics have been used in a number of areas for criminal investigation [40]. A combination of human observations and mining of electronic data is used to produce network diagrams to highlight persons of interest. Date grids and clusters were used to show patterns, with a combination of structural and temporal data. Visualization was a key element.

Boulos et al. [37] used Technosocial Predictive Analytics (TPA) to mine social web data to gain insights into the collective health of a given population. TPA is currently focused on the public health sector but it could be considered for the corrections environment as it also deals with population prediction. Business analytics [34] uses quantitative and qualitative data analysis including predictive models. The analytics is used to analyses business process performance against a Balanced Scoreboard and to re-create the processes for optimum performance.

Elsewhere multi-agent data mining has been used to identify subtle changes in neonates that may predict the onset of some condition ahead of blood cultures [33]. Historical temporal data is cross correlated with current data streams. Rule association mining is proposed as a means to define and test rules. A domain expert tests each rule set and, if satisfied, passes them to the rule generating agent for processing.

Pratt et al. [39] proposes the use of fuzzy cognitive maps and cellular automata to model insurgency. The research only reached proof of concept due to real-time data access issues. Some comparisons can be drawn between insurgency and prison riots. They are attempting simulations to prove the concept, but recognize experts needed to design scenarios to test the true predictive capabilities. PA is commonly used by financial institutions to predict the performance of new services through simulations [35]. Social analytics was used in Twitter to predict election results with the use of sentiment analysis [32]. Fülöp et al. [36] combined Complex Event Processing (CEP) and PA. They raise the valid concern that taking action before the predicted event will ruin the data set. Further exploratory research is required.

D. Autonomics

Autonomic computing (AC) is a concept created by IBM in 2001 [43]. The computing complexity crisis led to the need for self-managing systems, a concept derived from the biological autonomic nervous system. The level of interconnectivity and diversity in today's heterogeneous systems as we move towards pervasive computing has made traditional management no longer viable.

Kephart and Chess [42] saw the evolution of AC as moving from supporting the decision making process via data preparation to advising on those decisions. As confidence grew AC would make and act on low-level and eventually high-level decisions. The ultimate goal would be for AC to be so seamless and inherent that we forget it's there.

By 2010 Dobson et al. [41] felt that IBM's original vision remained unfulfilled. Further work is needed on combining the individual problem solutions into a broader engineering perspective. Considerations could be given to the capabilities of autonomics in the popular and growing research fields of big data analytics and data fusion. Potential exists for evolving autonomic frameworks such as the Unified Fault Management Architecture as defined by Sterritt et al. [44] to move towards autonomic systems engineering.

IV. Conclusions

An increasing amount of human activity is being recorded electronically, leaving a trail through the analysis of which conclusions and actions can be derived. This is particularly the case in the US Corrections Systems. With increasing prisoner numbers and costs, and decreasing funds and staff, advanced automation is a much needed way forward. The automation of predictive intelligence is ultimately at the heart of the next generation IMS. In this paper we have outlined the key enabling technologies we are researching to provide an autonomic analytics solution for corrections.

Acknowledgment

Catherine McFarland's PhD research is funded through a CAST Award (Co-operative Awards in Science and Technology) by the Northern Ireland Department for Employment and Learning (DEL http://www.delni.gov.uk/) and Core Systems (www.coresystems.biz).

.

REFERENCES

[1] Baskin, Deborah R., Ira Sommers, and Henry J. Steadman. "Assessing the impact of psychiatric impairment on prison violence." Journal of Criminal Justice 19.3 (1991): 271-280.

[2] Bierie, D.M. (2012) Procedural Justice and Prison Violence: Examining Complaints Among Federal Inmates (2000–2007). .

[3] Bingaman, J. (1980) Report of the Attorney General on the February 2 and 3, 1980 Riot at the Penitentiary of New Mexico. Santa Fe: Office of the Attorney General, .

[4] FERNANDEZ, K.E. & Neiman, M. (1998) California's inmate classification system: Predicting inmate misconduct. The Prison Journal, 78(4), 406-422.

[5] Glaze, L.E. & James, D.J. (2006) Mental Health Problems of Prison and Jail Inmates.

[6] Jiang, Shanhe. "Impact of drug use on inmate misconduct: A multilevel analysis." Journal of Criminal Justice 33.2 (2005): 153-163.

[7] McGarry, P. 2010, The Continuing Fiscal Crisis in Corrections: Setting a New Course, Center on Sentencing and Corrections, VERA Institute of Justice, New York.

[8] Schenk, A.M. & Fremouw, W.J. (2012) Individual characteristics related to prison violence: A critical review of the literature. Aggression and Violent Behavior, 17(5), 430-442.

[9] Schmitt, J., Warner, K. & Gupta, S. (2010) The high budgetary cost of incarceration. Washington, DC: Center for Economic and Policy Research.http://www.cepr.net/documents/publications/incarceration-2010-06.pdf, .

[10] Worrall, J.L. & Morris, R.G. (2012) Prison gang integration

[11] Allen, J., Guinn, C. & Horvtz, E. (1999) Mixed-initiative interaction. Intelligent Systems and their Applications, IEEE, 14(5), 14-23.

[12] Bleiholder, J. & Naumann, F. (2008) Data fusion. ACM Computing Surveys (CSUR), 41(1), 1.

[13] Cholvy, L. & Nimier, V. (2004) Information Evaluation: discussion about STANAG 2022 recommendations, .

[14] Guivant, J.E., Marden, S. & Pereida, K. 2012, "Distributed multi sensor data fusion for autonomous 3D mapping", Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on, pp. 1.

[15] Hall, D.L. & Llinas, J. (1997) An introduction to multisensor data fusion. Proceedings of the IEEE, 85(1), 6-23.

[16] Hall, D.D.L. & McMullen, S.A.H. (2004) Mathematical Techniques in Multisensor Data Fusion 2nd Ed. Artech House Publishers.

[17] Hall, D.L., McNeese, M.D., Hellar, D.B., Panulla, B.J. & Shumaker, W. 2009, "A cyber infrastructure for evaluating the performance of human centered fusion", Information Fusion, 2009. FUSION'09. 12th International Conference onIEEE, , pp. 1257.

[18] Khaleghi, B., Khamis, A., Karray, F.O. & Razavi, S.N. (2011) Multisensor data fusion: A review of the state-of-the-art. Information Fusion, .

[19] Kokar, M.M., Tomasik, J.A. & Weyman, J. (2004) Formalizing classes of information fusion systems. Information Fusion, 5(3), 189-202.

[20] Nilsson, M. & Ziemke, T. 2007, "Information fusion: a decision support perspective", Information Fusion, 2007 10th International Conference onIEEE, , pp. 1.

[21] Pichon, F., Dubois, D. & Denœux, T. (2012) Relevance and truthfulness in information correction and fusion. International Journal of Approximate Reasoning, 53(2), 159-175.

[22] Bollier, D. (2010) 'The promise and peril of big data', [Online] Available at: http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf (11 July 2011).

[23] Boyd, D. and Crawford, K. (2011) Six Provocations for Big Data. A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011

[24] Fisher, D., DeLine, R., Czerwinski, M. & Drucker, S. (2012) Interactions with big data analytics. interactions, 19(3), 50-59.

[25] Kaisler, S., Armour, F., Espinosa, J.A. & Money, W. 2013, "Big Data: Issues and Challenges Moving Forward", System Sciences (HICSS), 2013 46th Hawaii International Conference onIEEE, , pp. 995.

[26] Laney, D. (2001) 3-D Data Management: Controlling Data Volume. Velocity and Variety", META Group Original Research Note, .

[27] LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S. & Kruschwitz, N. (2011) Big data, analytics and the path from insights to value. MIT Sloan Management Review, 52(2), 21-32.

[28] McAfee, A. & Brynjolfsson, E. (2012) Big data: the management revolution. Harvard business review, 1.

[29] Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D. & Tufang, P. 2012, Analytics: The real-world use of big data, IBM Global Business Services, New York.

[30] Weikum, G. 2013, , Big Data should be Interesting Data! [Homepage of ACM SIGMOD Blog], [Online]. Available: http://wp.sigmod.org/?p=786March] .

[31] Abramson, M., Bennett, S., Brooks, W., Hofmann, E., Krause, P. & Temin, A. 1994, "Predictive Analysis System: a case study of AI techniques for counternarcotics", Artificial Intelligence for Applications, 1994., Proceedings of the Tenth Conference on, pp. 134.

[32] Bermingham, A. & Smeaton, A.F. (2011) On using Twitter to monitor political sentiment and predict election results.

[33] Bjering, H. & McGregor, C. 2010, "A multidimensional temporal abstractive data mining framework", Proceedings of the Fourth Australasian Workshop on Health Informatics and Knowledge Management-Volume 108Australian Computer Society, Inc., , pp. 29.

[34] Bronzo, M., de Resende, Paulo Tarso Vilela, de Oliveira, Marcos Paulo Valadares, McCormack, K.P., de Sousa, P.R. & Ferreira, R.L. (2013) Improving performance aligning business analytics with process orientation. International Journal of Information Management, 33(2), 300-307.

[35] Castro, C. (2011) The Power of Aggressive Analytics. Marketing, .

[36] Fülöp, L.J., Beszédes, Á., Tóth, G., Demeter, H., Vidács, L. & Farkas, L. 2012, "Predictive complex event processing: a conceptual framework for combining complex event processing and predictive analytics", Proceedings of the Fifth Balkan Conference in InformaticsACM, , pp. 26.

[37] K. Boulos, M.N., Sanfilippo, A.P., Corley, C.D. & Wheeler, S. (2010) Social Web mining and exploitation for serious applications: Technosocial Predictive Analytics and related

technologies for public health, environmental and national security surveillance. Computer methods and programs in biomedicine, 100(1), 16-23.

[38] Mishra, N. & Silakari, S. Predictive Analytics: A Survey, Trends, Applications, Opportunities & Challenges.

[39] Pratt, S.F., Giabbanelli, P.J., Jackson, P. & Mago, V.K. 2012, "Rebel with many causes: A computational model of insurgency", Intelligence and Security Informatics (ISI), 2012 IEEE International Conference onIEEE, , pp. 90.

[40] Westphal, C. (2008) Data Mining for Intelligence, Fraud & Criminal Detection: Advanced Analytics & Information Sharing Technologies. CRC Press.

[41] Dobson, S., Sterritt, R., Nixon, P. & Hinchey, M. (2010) Fulfilling the vision of autonomic computing. Computer, 43(1), 35-41.

[42] Kephart, J. 2011, "Autonomic computing: the first decade", Proceedings of the 8th ACM international conference on Autonomic computing-ICAC, pp. 1.

[43] Kephart, J.O. & Chess, D.M. (2003) The vision of autonomic computing. Computer, 36(1), 41-50.

[44] Sterritt, R., Gunning, D., Meban, A. & Henning, P. 2004, "Exploring autonomic options in an unified fault management architecture through reflex reactions via pulse monitoring", Engineering of Computer-Based Systems, 2004. Proceedings. 11th IEEE International Conference and Workshop on the, pp. 449. 89.