Massimo Buscema
William J. Tastle *Editors*

Intelligent Data Mining in Law Enforcement Analytics

New Neural Networks Applied to Real Problems

Springer

Intelligent Data Mining in Law Enforcement Analytics

Massimo Buscema • William J. Tastle
Editors

Intelligent Data Mining in Law Enforcement Analytics

New Neural Networks Applied to Real Problems

Springer

Editors
Massimo Buscema
Semeion Research Centre of Sciences of Communication
Rome
Italy

William J. Tastle
Ithaca College
NY
USA

ISBN 978-94-007-4913-9 ISBN 978-94-007-4914-6 (eBook)
DOI 10.1007/978-94-007-4914-6
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012953015

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Assembling the contents of an academic book dealing with some new technology or a sophisticated advancement is the task given over to the academic researcher who typically embraces the challenge with dedication and purpose for it is what makes us unique among our brethren. Libraries are filled with esoteric research that is the product of excellent minds, research that is so arcane and possibly cryptic that it might remain on the shelves for potentially centuries until an application succeeds in being brought forth by some other, equally sophisticated and talented, individuals who have the rare talent of merging new-found knowledge with practical application.

Such is not the case with this academic text for it was immediately observed that this method of data analysis and mining could be brought to bear in helping to solve some very complex problems that have plagued the law enforcement community since the advent of the database and its concomitant assortment of management systems. The easy applications, that is to say, the most trivial but definitely useful, were quickly subsumed *by the law enforcement community and began a movement to digitise all past and present case data for easy access and management; for the last few decades, their expectation has usually been met with varying degrees of success.*

However, the databases that were built over many years, or decades in some cases, still contained unknown and undiscovered knowledge, but no one knew of its existence until a meeting that occurred with an official of one of the most respected law enforcement agencies in the work, the London Metropolitan Police force (also known as New Scotland Yard), and the principal researcher of one of the most prestigious research institutes in Italy, Semeion Research Center of the Sciences of Communications of Rome.

It is to Sergeant Geoff Monaghan of New Scotland Yard that this book is dedicated for it was he who first taught us about the complex world of crime analysis. Sergeant Monaghan inspired us and motivated Semeion towards the adventure of crime analytics. It was his vision to "see" that knowledge was trapped in huge databases and needed some very sophisticated methods to extract it and make it understandable to the "front line" of police. Over the past 3 years, Semeion has worked closely with Sgt Monaghan, and this book explains, in detail, the successes and methods used to extract this unknown knowledge. From here, extraction of knowledge from other databases can become commonplace, as long as there *exist other talented visionaries in other disciplines who are willing to take the risk in creating knowledge.*

—Semeion and its staff

Preface

This book was written specifically for the law enforcement community although it is applicable to any organisation/institution possessing a database of activity it seeks to analyse for unknown and undiscovered knowledge. This is typically called data mining and the purpose is to extract useful knowledge. Generally, most organisations typically use structured query language (SQL) to query their database. While this does give information, one must know exactly the questions to ask in order to gather a response, and any question raised by means of a query will have an answer if and only if the answer is already present in the database. This kind of information is called *blatant* for it is conspicuous as opposed to hidden. Unfortunately, the knowledge hidden within databases requires some very sophisticated methods in order to coax it out.

The extraction of only blatant information from a database is too limiting given the demands for useful information in the complex society of the twenty-first century. We need to creatively explore a database to extract its hidden information, that is, the underlying information which produces the structure by which the evident information becomes obvious and available for query. In short, the hidden information is responsible for the blatant information making sense. This special meta-information is hidden and trapped in the blatant information. This hidden information is the condition of existence for the blatant information in the same way that the Kantian "noumenon" is the condition for the perception of the phenomenon. Hidden information is like the sea waves, while the blatant information, explicitly coded in a database, a similar to the foam of the waves. For most forms of analysis, hidden information is considered "noise". But it is within this noise that the genetic code of process, that from which this noise is derived, is encrypted. Our challenge is to successfully decrypt the genetic code; such a decryption is explained in this book.

We name this search for the hidden information trapped in the database intelligent data mining (IDM), and we think that the most advanced artificial adaptive algorithms are able to understand which part of the so-called noise is the treble clef of any database music.

The sophistication of the criminal element is exceptional. Drug cartels and terrorist organisations have the financial strength to purchase, or muscle to coerce, brilliant individuals to work for them, and it is egregious for any law enforcement organisation to underestimate the cleverness of those groups. It is argued that the best we can hope to do is minimise the distance between what they do and how we protect against them. To do so requires us to embrace the maxim *scientia est potentia.* This is Latin for "knowledge is power" and is attributed to Sir Francis Bacon, though it first appeared in the 1658 book, *De Homine*, by his secretary Thomas Hobbes. In order to extract knowledge, one must first have information, and to get information one must have data. There is another word used to describe the extraction of knowledge from data: *semeion*. Its origin is from the Greek, and it means the extraction of a large amount of knowledge from a small amount of data given a prepared mind and the spirit of discovery. Not only can remarkable information be gathered from a database, we show in this book how to harness that information to produce knowledge that can be brought to bear on the criminal element in our efforts to defeat them.

The motivation for this book came out of a cooperative venture with the London Metropolitan Police, well known by its metonym Scotland Yard, and the Semeion Research Center of Rome. In a correspondence from the Assistant Commissioner Tarique Ghaffur of the London Specialist Crime Directorate to the Italian Minister of University Education and Research, the basis for successful cooperation is clearly established:

> From the outset, I [Assistant Commissioner Ghaffur] want to emphasise that the Central Drug Trafficking Database (CDTD) Project is an important element of the Specialist Crime Directorate's (SCD) intelligence strategy and I'm delighted to tell you that the project is going very well. Moreover, the CDTD, which has been designed by Semeion in accordance with specifications laid down by my officers, is working very well. One of the most exciting aspects of this project is the idea of using Artificial Adaptive Systems (AAS) to analyse drug trafficking data. I readily acknowledge that this component is totally dependent on the founder and Director of Semeion, Professor Massimo Buscema, in view of his extensive and pioneering work in the field of artificial intelligence. I know my officers hold Professor Buscema in high regard and I would like to place on record my thanks to him and his colleagues at Semeion, particularly Dr Stefano Terzi, for helping to make our partnership a success.

Operationally, Semeion created a database structure that permitted both the use of traditional SQL queries and analysis using adaptive neural network technology. The outcomes, from the Metropolitan Police perspective, are detailed in the letter:

> By way of background, the CDTD is the first of its kind and has been designed to enable the SCD to produce reliable and objective data to help the MPS and its partners to: (a) assess the extent of the problem in London, and (b) devise appropriate responses to tackle the problem. The information will, in the main, be drawn from 4,500 drug trafficking reports recorded by the MPS in 2004. The reports will be scrutinised and the information validated by specially trained Data Entry Operators (DEOs). Where necessary, additional information will be obtained from the Forensic Science Service, the Police National Computer and a number of other databases. The refined data will then be entered onto the CDTD and new records created. Each record comprises around 500 fields. Subsequent analyses will shed new light on the structure of drug markets in London, how organised criminal networks

> shape and influence these markets, and the effectiveness of police tactics (e.g. stop and search, test purchases and controlled deliveries). Data gleaned from drug seizures – unit prices, purity, and chemical profiling – will also be analysed. The project will also highlight operational successes as well as noting the deficiencies in the recording and investigation of drug trafficking crimes. In sum, the CDTD will produce high-quality intelligence, which will be tailored to the varying needs of decision makers in the MPS from the strategic to the tactical levels.

The last point to address is that of the complementary relationship between traditional statistics and neural network technology. While statistics definitely plays an important role in data analysis, there are other methods that provide an entirely different view of the system under investigation. The London Metropolitan Police recognised the limitations of traditional statistics and sought to apply artificial adaptive systems (AAS) to their analysis.

> During their research, the Project Team will be using conventional statistical programmes. But in order to process the vast volumes of data generated and recognising that comprehensive analyses cannot be done without highly advanced data processing capabilities, the team also wants to use AAS. To this end, SCD has contracted Semeion to design the database management system and to analyse the data using AAS developed by Professor Buscema and his colleagues. Although AAS have been applied to various areas of research, we believe that this is the first time that they have been used to analyse drug trafficking crimes (or indeed any other type of crime) on this scale and in this detail.

The theory, methods and applications described in this book can be utilised by any police agency or modified to fit the needs of any business or organisation seeking to extract knowledge from a database. Non-profit organisations will find that donor/membership databases contain knowledge that could be utilised to enhance fundraising and membership drives. Military databases are typically huge and contain hundreds, if not thousands, of variables. A wealth of unknown knowledge may well be contained within those databases if only these new methods presented in this book were applied to them. Medical databases will benefit from the identification of hidden knowledge and could give scientists valuable insights into novel directions for research. Financial institutions have data on every customer, loan, stock portfolio, etc., and there is new knowledge to be gleaned from an analysis using these very sophisticated methods.

Many individuals, beyond the chapter authors, were involved in the production of this book, and we gratefully acknowledge their contribution:

- Dr Giulia Massini, computer scientist and deputy director of the Semeion Research Center of Sciences of Communication, Rome, Italy
- Dr Guido Maurelli, Semeion Research Center of Sciences of Communication, Rome, Italy
- Marco Intraligi, computer data processing expert, Semeion Research Center of Sciences of Communication, Rome, Italy
- Dr Stefano Terzi, computer scientist, formerly Semeion Research Center, Rome, Italy
- Dr Leslie A King, consultant, formerly head of the Drugs Intelligence Unit, Forensic Science Service, London, UK, and former advisor to the Department

of Health, England, and the European Monitoring Centre for Drugs and Drugs Addiction, Lisbon, Portugal
- Paul Richards, formerly CDTD project manager and inspector, Drugs Directorate, New Scotland Yard, Metropolitan Police Service, London
- Ms Mandeep Kaur Bajwa, Ms Zoe nee Beard and Mr Adam Stevens, formally data entry operators, CDTD project, New Scotland Yard, Metropolitan Police Service, London
- Mr Dean Ames, formerly forensic scientist, Drugs Intelligence Unit, Forensic Science Service, London, UK

A special acknowledgement to:

- Tarique Ghaffur, CBE, QPM, formerly assistant commissioner, Central Operations, New Scotland Yard, Metropolitan Police Service, London
- Andy Baker, deputy director of the Serious Organised Crime Agency, London, UK, and formerly commander, Specialist Crime Directorate, New Scotland Yard, Metropolitan Police Service, London
- Dr Stanley "Sholmo" Einstein, Jerusalem, Israel
- Professor (Emeritus) John G D Grieve CBE, QPM, John Grieve Centre for Policing and Community Safety, Faculty of Social Sciences and Humanities, London Metropolitan University, London, UK, and formerly deputy assistant commissioner, Specialist Crime Directorate, New Scotland Yard, Metropolitan Police Service, London
- Paul Hoare, formerly detective superintendent, Drugs Directorate, New Scotland Yard, Metropolitan Police Service, London

Paolo Massimo Buscema
General Director, Semeion Research Center, Rome, Italy
and
Professor, University of Colorado at Denver
Dept of Mathematical and Statistical Sciences
m.buscema@semeion.it
www.semeion.it

William J. Tastle
Professor, Ithaca College, New York, USA
Research Fellow, Semeion
tastle@ithaca.edu
faculty.ithaca.edu/tastle

Philip Bean
Department of Social Sciences
Midlands Centre for Criminology and Criminal Justice
University of Loughborough
Loughborough, Leicestershire LE11 3TU, UK

Teresa Nemitz
Department of Social Sciences
Midlands Centre for Criminology and Criminal Justice
University of Loughborough
Loughborough, Leicestershire LE11 3TU, UK

All images in this book are depicted in black and white and, consequently, the color detail is lost. The color images can be viewed at http://www.semeion.it or in the electronic publication of the book on www.springerlink.com.

Contents

Chapter 1
Introduction to Artificial Networks and Law Enforcement Analytics

William J. Tastle

> The word "semeion" is derived from the Greek and defined as a knowledge extraction process that utilizes a small amount of data to achieve a large quantity of knowledge given a prepared mind and the spirit of discovery.

"Intelligence, in itself, does not make up part of the specific attributes of crime, but when it is present, it increases danger immeasurably and causes it to become organized crime." This phrase appears in the book both as an affirmation of the difficulty of the problem as well as a point of departure for finding solutions because what is intelligent is not causal, and what is not causal is foreseeable. Since intelligence must be fought with intelligence, the forces of order require a good dose of "civil intelligence" to fight the "uncivil intelligence" of criminals.

The book is a thorough description and summary of the means currently available to the law enforcement investigators to utilize artificial intelligence in making criminal behavior (both individual and collective) foreseeable and for assisting their investigative capacities. Concretely, there are five cognitive activities carried out by an investigator: (1) the classification of criminal situations; (2) the spatial visualization of where the events occurred; (3) a prediction of how the events developed; (4) the construction of somatic, sociological, and psychological profiles; and finally (5) hypotheses regarding links between events, persons, and clues. Yet, *all* five cognitive activities can be explained (and often in more than one way) by adaptive artificial systems, furnishing a second opinion regarding the analysis of criminal events.

The artificial adaptive systems are efficacious for two reasons: in the first place, they keep in "mind" all the data; the human mind, in contrast, must make a selection from among the various pieces of data before being able to reason; in the

W.J. Tastle (✉)
School of Business, Ithaca College, New York, USA

Semeion Research Center, Rome, Italy
e-mail: tastle@ithaca.edu

M. Buscema and W.J. Tastle (eds.), *Intelligent Data Mining in Law Enforcement Analytics: New Neural Networks Applied to Real Problems*, DOI 10.1007/978-94-007-4914-6_1,

second place, they successfully confront complex phenomena in which there are no relational links and before which the human mind, in its tendency to simplify, finds itself in difficulty because it tends to reason precisely in terms of such relations.

The work arises from collaboration between the Semeion Research Center and New Scotland Yard in a joint project and contains the multiple articles typical of a group effort. It passes from an analysis of the drug-trafficking situation in London by a Scotland Yard investigator to the description of how a reliable database was crafted, arriving finally at the practical application of various algorithms upon criminal events registered in the database. The specialists in artificial intelligence at Semeion have not only supervised the construction of a database by Scotland Yard, as well offered their capacity in applying it, but have also created original algorithms for the precise purpose of making crime predictable.

The results obtained, in width and depth, can be considered the basis for the construction of an "artificial investigator," an integrated support system that functions on the various levels of the police organization: strategic, directive, and operative. On each level, specific types of software support must be furnished; for the upper levels, what is needed is the capacity for prediction on a wide scale and for making a synthesis of all the facts available and, for lower levels, facility and speed of usage.

The value of the ideas and methods presented in this book goes beyond the area strictly linked to crime, to which reference is made in order to have a concrete field of application. A reading of the book can thus be useful, not only for those concerned with investigation or specialists in the algorithms of artificial intelligence but also for those who work in the vast field of the social sciences.

1.1 Navigating the Book

One can approach the reading of this book in the traditional way, from front to back. However, the theoretical chapters are more mathematically demanding than the application-oriented chapters, so there are some tracts one could use to guide their reading. The first track is the theoretical chapters tract and consists of Chaps. 2, 3, 4, 7, 8, 9, 11, 13, 15, 17, and 19. For those individuals already adept at neural network methodologies, these chapters capture the details of the algorithms and provide the basis for which similar algorithms can be created.

Those individuals who seek only to understand how adaptive neural networks can be applied to law enforcement problems can focus their attention on Chaps. 5, 6, 10, 12, 14, 16, and 18.

The last two tracks involve a merger of both the theoretical and applied chapters. To adequately be able to interface with software engineers/programmers who might be creating specialized programs for use in your facility, the reader can take two approaches, each of which is a combination of theoretical and applied chapters. First involves Chaps. 2, 3, 4, 5, 6, 8, 9, 10, and 15, 16, 17, 18, 19 and is a focus on theory augmented with certain applied chapters that bring life to the theoretical chapters.

The other track involves Chaps. 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, and 18, 19. These chapters combine selected theoretical chapters with applied chapters to give one the tools and understanding needed to customize the algorithms to specific needs.

1.2 The Chapters

To properly prepare you to maximize your understanding of some of the very complex methods presented, the following chapters are brief summaries to give you an opportunity to experience the entire flavor of the book and perhaps direct your attention to some specific areas of interest. The actual usage of these algorithms is rather complex and would require the services of someone who is knowledgeable in both mathematics and a skilled programming to develop an interface for use by a particular police agency. On the other hand, an interested enforcement agency could simply seek advice from the director and staff of Semeion.

Chapter 2 ("Law Enforcement and Artificial Intelligence") is a high-level description of the motivation for the work in enforcement analytics. It establishes a justification for the use of adaptive neural networks, briefly explains how artificial learning occurs, and explains why traditional mathematics and methods of analysis with which we have become comfortable are no longer able to serve our needs as they did in the past. The criminal element is very smart, and they have the funds to do smart and innovative things to thwart the efforts of the "traditional" law enforcement community. When enforcement ceases to be creative, those on the other side profit at our expense. As enforcement becomes increasingly creative in its war on crime, the other side must expend increasing resources on higher levels of creativity, and the cycle will continue until one side surrenders. It is unlikely that those who unlawfully profit will be the first to capitulate.

Chapter 3 ("The General Philosophy of Artificial Adaptive Systems") describes the philosophy of the artificial adaptive system and compares it with our natural language. Some parallels are striking. The artificial sciences create models of reality, but how well they approximate the "real world" determines their effectiveness and usefulness. At the conclusion of this chapter, one will have a clear understanding of expectations from using this technology, an appreciation for the complexities involved, and the need to continue forward with a mind open to unexpected and unknown potential. The word "algorithm" is used almost continuously throughout the book. It is a very common word and can be interpreted as a simple "set of steps" used to attain an answer. Each step is very precise and acted upon by a computer as one line of instructional code. Once the computer has completed the running of the algorithm, an answer is provided to the user either in the form of screen output or sometimes as a hard-copy report. Most programs today use the screen as the mechanism for displaying output.

Chapter 4 ("A Brief Introduction to Evolutionary Algorithms and the Genetic Doping Algorithm") is an introduction to evolutionary algorithms, a commonly used method by which solutions to problems that might otherwise be impossible to solve are solved. One such method is that of the genetic algorithm. One of its strengths is its ability to solve problems in a relatively short time period that would otherwise not be solvable with the fastest computers working since the beginning of time. Such problems might be called NP or NP-hard problems, meaning that the time required for a computer to solve them is very, very long. On the downside, the answer provided by the genetic algorithm may not be optimal, but it is an "adequate" answer. To ensure an optimal solution, a computer would have to complete an examination of every possible solution, then select from the list the single winner. That solution would be optimal, but what if it took a supercomputer working at maximum speed a few years to deliver that answer; is it reasonable to expect one to wait that long a period of time for the exact answer? Or if an answer could be provided in a relatively short time period of a few hours (or minutes) and is "close" to optimal be acceptable? These approximate solutions are found to be quite useful and do provide for confident decision-making.

Sometimes evolutionary algorithms are based on what is called heuristics, or rules of thumb. They are guidelines for solutions that work; there are no mathematical proofs of their effectiveness; they just work well. Consequently, methods incorporating heuristics are deemed to be "weak." The word is unfortunate for it conveys a sense of inaccuracy or approximation, but it is, in fact, responsible for some excellent solutions. These weaker methods use less domain knowledge and are not oriented toward specific targets. In law enforcement analytics, the existence of such methods has been shown to be very advantageous. The chapters up through four are an excellent review of the operations of the genetic algorithm, and these are well known in the AI field. Chapter 4 presents a new genetic algorithm that is much more effective, the genetic doping algorithm (GenD). The word "dope" is unfortunate for it congeries up images of narcotics or a stupid person, but it actually means information gotten from a particularly reliable source. In this case, the reliable source is the data, and the effort is to extract maximal information from it.

GenD analyzes the data as though it were a tribe of individuals in which not everyone engages in crossover. To anecdotally explain, individuals in the tribe who are old or weak do not necessarily marry someone from that tribe (crossover does not occur in all genes); the fitness score (a calculated value that determines the ordering of the individuals in the tribe) is calculated on the basis of vulnerability and connectivity, and instead of dealing with the separate genes as individuals, GenD transforms the dataset into a dynamic structure and attempts to more closely mimic a genotype. A detailed and easy-to-read explanation of the differences between traditional genetic algorithms and GenD is given.

Chapter 5 ("Artificial Adaptive Systems in Data Visualization: Proactive Data") addresses the issue of the visualization of data modeled by artificial adaptive systems and one relatively easy visualization if that of the tree structure. A tree is a graph that contains a root, trunk, and leaves given a suitable imagination. Essentially, it is a diagram in which each point is connected to another point but without any circuits or loops anywhere in the graph. Thus, one can move from one point (called a vertex or node) to another following the lines (called edges or arcs) that connect the nodes and never come back over one's previous track. The structure is very important to understanding some very complex datasets. One of the ways it simplifies visualization is in its "dimensionality."

To see the visual representation of one single variable, we need only to plot a point on the *x*-axis of a graph, say a variable with a value of 12. At 12 units from the origin, we can place a dot to represent that variable. If we expand to two variables, say variable A has a value of 12 and variable B has a value of 3, then we can visualize this by placing a point in the XY coordinate plan that is located at the intersection of $X = 12$ and $Y = 3$. Similarly, we can add another variable to the mix, say variable $C = 2$, but visualization becomes somewhat more of a challenge for we must create a three-dimensional diagram on a sheet of paper (or computer screen). This can be easily done, and now we can see a point in position $X = 12$, $Y = 3$, and $Z = 2$ where X, Y, and Z are the *x*-axis, *y*-axis, and *z*-axis. So we have gone from one dimension, a line, to two dimensions, a plane, to three dimensions, a cube. Suppose we now add a fourth variable, or a fifth, or a 100th variable to the dataset. Visualization becomes a challenge to "see" the structure of the answer. Tree structures are one way by which many dimensions can be reduced to representation in two or three dimensions. While it takes some practice getting used to correctly reading and interpreting the graphs, the outcome is well worth the effort.

This chapter makes it clear that when one has a mass of data, possibly collected over years and on which SQL queries have been repeatedly made to the point that one might not think there is any more information that can be gleaned from further mining, it is the artificial neural network set of tools that come into play to explain the interactions and relationships existent among the data. The rules that connect the various sets of data within the database will very likely be fuzzy and dynamic. As the data submitted to the ANN are updated, it will adjust its "rules" in accordance, integrating the old data with the new, permitting us to correctly generalize new, dirty, incomplete, or future data.

Chapter 6 ("The Metropolitan Police Service Central Drug-Trafficking Database: Evidence of Need") discloses the problems inherent in large database systems, the errors that are entered into it by nontrained or only partially trained data input operators, the inconsistencies in the data that further thwart efforts to glean useful information using traditional methods, and the absence of a recognition that correct DB input, though time consuming, can be an ardent partner in the identification of relationships and the generation of profiles as a definite source of help and assistance to the enforcement community. It becomes apparent that the police, local, national, and international, have at their disposal access to information that could revolutionize the ways in which their jobs are performed, if only they had the knowledge, foresight, funding, and incentive to utilize it.

Chapter 7 ("Supervised Artificial Neural Networks: Back Propagation Neural Networks") becomes technical with a description of one of the most basic neural networks, that of the back propagation network. To understand it requires first a familiarity with the feedforward backpropagation artificial neural network (FF BP ANN). The first half of this chapter is a relatively low-level introduction to the theory FF BP, but it does get into some more challenging mathematics in the second half. If one has a background in calculus and differential equations, the math will be easy to follow. If not, one can simply accept that the mathematics are correct and read “around” the equations. In this manner, one can learn the theory and get a basic understanding how it works. This is probably the pivotal chapter for all the remaining algorithms; most everything else builds on this content.

Chapter 8 (“Preprocessing Tools for Nonlinear Datasets”) addresses the most difficult, and arguably the most important, problem in artificial neural networks, the training and testing of the network to ensure the best possible outcome. ANNs must first be “trained” to understand the data and establish the relationships among the variables, and it is a task that the algorithm must do itself. In the classical sense, the dataset would simply be randomly partitioned into two or more subsets, and one subset would be used to train the network, another to test it, and finally one subset on which to actually run the network. There are problems inherent in this method, especially when the database is extremely large, as is typically the case with enforcement DBs, and when the data is “noisy.” Noise is the existence of data that does not have any strong relationships with other variables. If a network is overtrained, the noise is incorporated as if it was strongly tied to other variables, and hence, new evaluated data would consider the noise to be an important part of the system. This would yield an incorrect interpretation of the data. Noise must be eliminated so that the network is properly trained. This chapter discusses how best to perform that action.

One way of eliminating noise, or at least reducing its impact, is addressed by two new algorithms called the training and testing algorithm (T&T) and the training and testing reverse algorithm (T&Tr). These are preprocessing systems that permit procedures to be far more effective in training, testing, and validating ANN models. This chapter presents the concept and mathematics of the algorithm and then illustrates their effectiveness with an example.

Chapter 9 (“Metaclassifiers”) describes methods by which data can be classified. There are many methods which purport to classify data, and each one performs the classification in a different manner and typically with differing results. The variation in outcome can be explained by saying that the different mathematics associated with each method views the data from various different perspectives, assigning data to classifications that can be, and usually are, different. A metaclassifier, however, is a method by which the results of these individual classifiers are considered as input to an ANN that forms the classifications based on the differing views and perspectives of the individual ANNs. In short, the different perspectives of the individual ANNs are brought together to produce a single, superior classification taking into account the various algorithms that produce certain views of the data.

Chapter 10 (“Auto-identification of a Drug Seller Utilizing a Specialized Supervised Neural Network”) is a comprehensive illustration of the application of pattern recognition on a law enforcement database of drug-related data using the metaclassification algorithm discussed in the previous chapter. This chapter is more accessible to the nontechnician and gives an exciting, and detailed, description of how the metaclassifier can be used to identify unknown relationships.

Chapter 11 (“Visualization and Clustering of Self-organized Maps”) describes a type of neural network that has been around for some 30 odd years, the self-organizing map. The main significance of this type of ANN is that it can take high-dimensional data and produce a diagram (map) that displays it in one or two dimensions. In short, humans can visualize interactions when displayed in one, two, or three dimensions, but not four or more dimensions. Data composed of only one variable can “see” a point on an *x*-axis diagram; data composed of two variables can be displayed on an *x*-*y*-axis diagram; data composed of three variables can be displayed on an *x*-*y*-*z*-axis diagram, and the visualization stops here. We simply cannot visualize diagrams in four or more dimensions, and that is where the self-organizing map comes into play. It has the ability of analyzing data in an unsupervised way (without any preconceived indication of the number of patterns present in the data) and placing the resulting analysis in a one- or two-dimensional diagram. While some information might be lost in the translation, it is more than made up with the insights that one can glean from the resulting diagram.

This type of ANN is continued in Chap. 12 (“Self-organized Maps: Identifying Nonlinear Relationships in Massive Drug Enforcement Databases”) with its use in the analysis of a massive drug enforcement database collected over time by the Scotland Yard Metropolitan Police. Throughout this chapter, the theory of the self-organizing map, as presented in Chap. 12, is explained in substantial detail ending with many visualizations of the drug industry in London. The results yield a “profile” that can be used by law enforcement agencies to target their investigations, monitoring, etc. Since the profile is the result of a mathematical algorithm, an argument that a particular ethnic group is being targeted can and should be dismissed, for the data speak for themselves.

Chapter 13 (“Theory of Constraint Satisfaction Neural Networks”) is a description of the constraint satisfaction neural network (CS ANN). Problems typically have some constraints that limit a decision, and we have this situation regularly occurring. For example, a search of a database for the owner of a particular car whose license begins with ABC is a constraint imposed on the solution. A search for a male whose height is between 5′6″ and 5′8″ (167.6 and 172.7 cm) and weight is 220 lb (100 kg) is a constraint problem. Thus, the constraint satisfaction ANN involves finding a solution given the imposition of a series of conditions on it.

Chapter 14 (“Application of the Constraint Satisfaction Network”) is an extension of the previous chapter and describes the application of the CS ANN on a dataset composed of 144 variables on 1,120 cases involving drug trafficking within the boroughs of London. The examples show the level of detail that can be derived from data using this method of analysis, and the results are graphically shown in tree diagrams for which the interpretation of which is also provided. Law enforcement officers will get a very good understanding as to the kinds of information, and the depiction of the results of the analysis, that may be available in databases. There is a richness of information that very likely has not been mined, and the methods described here should excite the reader as to possible results.

Chapter 15 (“Auto-contractive Maps, H Function, and the Maximally Regular Graph: A New Methodology for Data Mining”) describes an original artificial neural network called the auto-contractive map (AutoCM). This specialized analytical method produces a graphical image that displays the overall relationships that exist among the variables at the most fundamental level of system construction. That is, a set of variables that constitute some system under investigation possess some degree of connectivity among all the variables. To draw a graph in which each point is connected to every other point does, in fact, represent the overall system, though it also gives absolutely no information as to the structure of the underlying, basic structure. We seek to understand the structure but at the point when the individual variables have enough of a relationship with each other to form an initial linkage. Thus, the AutoCM shows a graph in which each point (called a node or vertex) is connected to another node but without the creation of any loops or circuits in the graph. This is called a minimal spanning tree (MST). The benefit of this kind of graph is that it permits us to see how one variable is related to another from a hierarchical perspective. For example, if we possess a large number of records on individuals and we seek to understand how these people are related (in the mathematical sense) to each other, the MST will give us exactly that information. It is this ANN that can be used, for example, to determine the structure of a drug network: who are the pushers, who supplies the pushers, who supplies the suppliers, etc.

An embellishment of this method is the inclusion of the maximal regular graph, the linking of individual nodes that show the strongest degree of "relationship," which creates circuits in parts of the MST. These circuits typically end up creating a perfect (or almost perfect) subgraph. An example of the AutoCM is given using the characters from West Side Story, the Jets, and the Sharks.

Chapter 16 ("Analysis of a Complex Dataset Using the Combined MST and Auto-contractive Map") uses the auto-contractive map ANN described in Chap. 15 to analyze the drug activity in London. The main value in this chapter is the interpretation of the various graphs, taking the visual representation of the mathematical relationships and putting them into words that can yield some action. Anyone seeking to use the AutoCM ANN should spend the time necessary to master this chapter.

Chapter 17 ("Auto-contractive Maps and Minimal Spanning Tree: Organization of Complex Datasets on Criminal Behavior to Aid in the Deduction of Network Connectivity") is a comprehensive example of using adaptive neural networks to solve interesting and important law enforcement-related problems and should be read very carefully. Essentially, let us assume that we have a very large database composed of data on individuals that include gender, age, nationality, where arrested, number of previous convictions, arrests and offenses, type of drugs seized, behavior of the individual, and much more. Missing from the database is any information about associations with others, and that is the item of interest we seek to discover: Which individuals belong to the same "gang" or are involved in the same drug-trafficking "circle"? Using the auto-contractive map and minimal spanning tree methods, this chapter shows how mathematics can answer this question. For law enforcement personnel seeking to get some handle on gang activity or conspiracy activity, guidance toward a solution can be found here.

The illustration of the usage of this ANN technology to solve interesting problems continues with Chap. 18 ("Data Mining Using Nonlinear Auto-associative Artificial Neural Networks: The Arrestee Dataset"). Using the drug database, this chapter explores how to apply nonlinear auto-associative systems (nonsupervised ANNs) to data analysis. The results of the analysis are presented in various MSTs and the graphs interpreted. While fascinating and informative graphs are created and discussed, these structures are the results of the application of mathematical algorithms to records of data, and it must be emphasized that the analysis depends directly on the quality of the data entry and that the results should be interpreted as a point of departure for anyone using these methods for investigative purposes.

The final Chap. 19 ("Artificial Adaptive System for Parallel Querying of Multiple Databases") addresses the interesting problem of analyzing multiple databases that do not possess a similar data structure. An analogy is made with different wineries in the same community. Each winery produces its own special wines, and each wine has its own set of characteristics, but they all come from the same geographical area. Hence, once could want to understand the complex interactions that occur among the different wineries. Similarly, different police organizations may have their data stored in databases whose data structures, or the kinds of data placed in the fields that make up the records, differ across the organizations.

1.3 Collaborative Opportunities

The Semeion Research Center in Rome, Italy, the research facility responsible for the discovery and development of all the methods described in this book, regularly works with companies and organizations who seek to implement some of the methods described here in their own organization but do not have the funding necessary to create their own systems. Arrangements can be made with Semeion to utilize their software under a licensing arrangement. Also, it is not uncommon for organizations to pass their database directly to Semeion where the analysis occurs and a detailed report is provided. The analysis is completed at Semeion and the results are explained, in whatever detail is necessary to ensure proper understanding of the results, to the organizational officials.

Chapter 2
Law Enforcement and Artificial Intelligence

Paolo Massimo Buscema

2.1 Data and Methods

The original purpose of several of the chapters in this book was to acquaint the Metropolitan Police Service (MPS) of London, England, with the interim findings arising from the analyses of data drawn from drug trafficking crimes recorded by the MPS in 2004. Since then, the purpose has evolved into a book that would provide new and exciting guidance as to the possibilities of numerous nontrivial neural network applications to the many different fields of law enforcement. Law enforcement agencies typically collect huge quantities of data and might remain underutilized for numerous reasons, that is, insufficient time for investigators to mine the data, insufficient resources to engage in creative and exploratory methods of analytically gathering useful information from the data, or simply of not being aware of the various forms by which data can be utilized to create new insights into how mathematics, computer science, and systems analysis can be brought to bear on critical problems. These analyses were undertaken by the Semeion Research Center of Sciences of Communication (hereafter *Semeion*) using an array of artificial adaptive systems. As a prerequisite to the testing, training, and analytical phases, the report also discusses the importance of:

- Collecting, organizing, and validating data
- The necessity of designing and constructing a database "fit for purpose"

The report further points out that having a wealth of well-organized data and a sophisticated database is of little use unless the analytical tools used are underpinned by robust mathematics. Without good mathematics, there is a high risk that any analyses of data will generate contradictory and arbitrary gobbledegook.

P.M. Buscema (✉)
Semeion Research Center of Sciences of Communication, Via Sersale 117, Rome, Italy
e-mail: m.buscema@semeion.it

M. Buscema and W.J. Tastle (eds.), *Intelligent Data Mining in Law Enforcement Analytics: New Neural Networks Applied to Real Problems*, DOI 10.1007/978-94-007-4914-6_2,

Data are read by tools, and suitable tools are provided by mathematics. The basic statistics through which often complex database are read are not sufficient to understand the secrets that data freeze in the form of numbers. Frequency tables, means, variances, two-variable correlations, etc., are excellent tools for "getting an idea" of the data that we have before us. But these basic analyses do not reveal anything about the complex relations concealed in an important database. And the reason is simple: in any database, any individual datum interacts in parallel with all the others, and it is only this global "many-to-many" interaction that generates the meaning of each individual datum (Grossi et al. 2007; Dzeroski and Lavrac 2001).

Even more complex statistics sometimes might miss the subtle associations that constitute the framework of any database. Every linear multivariate analysis, in fact, links data through simplified relations. But most natural, biological, and cultural phenomena do not follow plain "cause-effect" relationships. Any structure that has feedback, from a thermostat to human brain circuits, often violates in some way the dynamics of these statistics. This means that in many databases nonsignificant linear relationships may conceal the keystone of the entire system. To be unaware of them may mean not to understand at all how that system will actually respond to our actions.

A database is in fact like a living system frozen in one or more moments of one's life. To understand it means reactivating the interactions between each of its individual data and every other datum. A mathematics that is suitable for activating this reanimation process must consider each individual datum as an agent whose aim may consist of either of the following:

1. Negotiating the value of its relationship with all the others, in order to maintain its own identity, that is, its original numerical value
2. Defining its new identity, in order to maintain its relationships with the others, if these are binding

To operate on a database in this way means understanding the history that has produced it and the future on which it is focused.

Analyses of this type cannot be done through statistical tools that review all the data just once to establish simplified relationships that may entirely miss the mark. A more complex mathematics is required, able to consider each datum as an agent that develops its linear and nonlinear relationships with all the others over time and in parallel. A mathematics operating in a manner similar to that of operation of the human brain is required, that is, locally, in parallel and through continuous feedback between basic units.

Artificial adaptive systems are the mathematical tools suitable for these types of analysis.

These are systems whose:

1. Basic units interact in parallel and locally with one another
2. Units, therefore, negotiate their interactions over time and according to highly nonlinear logics