Heuristics in Analytics

A Practical Perspective of What Influences Our Analytical World

CARLOS ANDRE REIS PINHEIRO
FIONA McNEILL

WILEY

Contents

Wiley & SAS Business Series

The Wiley & SAS Business Series presents books that help senior-level managers with their critical management decisions.

Titles in the Wiley & SAS Business Series include:

Activity-Based Management for Financial Institutions: Driving Bottom-Line Results by Brent Bahnub

Big Data Analytics: Turning Big Data into Big Money by Frank Ohlhorst

Branded! How Retailers Engage Consumers with Social Media and Mobility by Bernie Brennan and Lori Schafer

Business Analytics for Customer Intelligence by Gert Laursen

Business Analytics for Managers: Taking Business Intelligence beyond Reporting by Gert Laursen and Jesper Thorlund

The Business Forecasting Deal: Exposing Bad Practices and Providing Practical Solutions by Michael Gilliland

Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy by Olivia Parr Rud

CIO Best Practices: Enabling Strategic Value with Information Technology, second edition by Joe Stenzel

Connecting Organizational Silos: Taking Knowledge Flow Management to the Next Level with Social Media by Frank Leistner

Credit Risk Assessment: The New Lending System for Borrowers, Lenders, and Investors by Clark Abrahams and Mingyuan Zhang

Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring by Naeem Siddiqi

The Data Asset: How Smart Companies Govern Their Data for Business Success by Tony Fisher

Delivering Business Analytics: Practical Guidelines for Best Practice by Evan Stubbs

Demand-Driven Forecasting: A Structured Approach to Forecasting, Second Edition by Charles Chase

Demand-Driven Inventory Optimization and Replenishment: Creating a More Efficient Supply Chain by Robert A. Davis

The Executive's Guide to Enterprise Social Media Strategy: How Social Networks Are Radically Transforming Your Business by David Thomas and Mike Barlow

Economic and Business Forecasting: Analyzing and Interpreting Econometric Results by John Silvia, Azhar Iqbal, Kaylyn Swankoski, Sarah Watt, and Sam Bullard

Executive's Guide to Solvency II by David Buckham, Jason Wahl, and Stuart Rose

Fair Lending Compliance: Intelligence and Implications for Credit Risk Management by Clark R. Abrahams and Mingyuan Zhang

Foreign Currency Financial Reporting from Euros to Yen to Yuan: A Guide to Fundamental Concepts and Practical Applications by Robert Rowan

Health Analytics: Gaining the Insights to Transform Health Care by Jason Burke

Heuristics in Analytics: A Practical Perspective of What Influences Our Analytical World by Carlos Andre Reis Pinheiro and Fiona McNeill

Human Capital Analytics: How to Harness the Potential of Your Organization's Greatest Asset by Gene Pease, Boyce Byerly, and Jac Fitz-enz

Information Revolution: Using the Information Evolution Model to Grow Your Business by Jim Davis, Gloria J. Miller, and Allan Russell

Manufacturing Best Practices: Optimizing Productivity and Product Quality by Bobby Hull

Marketing Automation: Practical Steps to More Effective Direct Marketing by Jeff LeSueur

Mastering Organizational Knowledge Flow: How to Make Knowledge Sharing Work by Frank Leistner

The New Know: Innovation Powered by Analytics by Thornton May

Performance Management: Integrating Strategy Execution, Methodologies, Risk, and Analytics by Gary Cokins

Retail Analytics: The Secret Weapon by Emmett Cox

Social Network Analysis in Telecommunications by Carlos Andre Reis Pinheiro

Statistical Thinking: Improving Business Performance, second edition by Roger W. Hoerl and Ronald D. Snee

Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics by Bill Franks

Too Big to Ignore: The Business Case for Big Data by Phil Simon

The Value of Business Analytics: Identifying the Path to Profitability by Evan Stubbs

Visual Six Sigma: Making Data Analysis Lean by Ian Cox, Marie A. Gaudard, Philip J. Ramsey, Mia L. Stephens, and Leo Wright

Win with Advanced Business Analytics: Creating Business Value from Your Data by Jean Paul Isson and Jesse Harriott

For more information on any of the above titles, please visit www.wiley.com.

Heuristics in Analytics

A Practical Perspective of What Influences Our Analytical World

Carlos Andre Reis Pinheiro
Fiona McNeill

WILEY

Cover image: ©iStockphoto / loops7

Cover design: Wiley

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Reis Pinheiro, Carlos Andre, 1940-

Heuristics in analytics : a practical perspective of what influences our analytical world / Carlos Andre Reis Pinheiro, Fiona McNeill.

1 online resource. — (Wiley & SAS business series)

Includes bibliographical references and index.

Description based on print version record and CIP data provided by publisher; resource not viewed.

ISBN 978-1-118-42022-5 (pdf) — ISBN 978-1-118-41674-7 (epub) — ISBN 978-1-118-34760-7 (cloth)

1. Management—Statistical methods. 2. Decision making—Statistical methods. 3. Business planning—Statistical methods. 4. Heuristic algorithms. 5. System analysis. I. Title.

HD30.215

658.4'033—dc23

2013043850

No matter how heuristic our world is, we always have some certainties in life.
This book is dedicated to our certainties, family. Regardless how much we try,
and how many mistakes we make, nothing would be possible without family.
This book is totally dedicated to Daniele, Lucas, and Maitê.
And Jerry, Glo, and Emily.

Preface

Suppose you have a magical black box. You put a huge amount of data into it, from everywhere you can find. And say that each of these inputs are in different formats and each format, in turn, represents several distinct time periods, frequencies, and seasons, as well as different characteristics. Imagine that you can then push a button and voilà! Out pops new knowledge from that hodgepodge of input. You now know the likelihood that the next event will occur, you can predict behavior; you have quantified risk, the propensity to act, and well-defined emerging patterns. And all this magically produced output would be ready to use—in the format needed to take action. This is amazing, isn't? You just fill some box with lots of different stuff and get results that are easy to understand and that can immediately be put into use.

That black box is *analytics*. And it isn't a brand-new discipline but is heavily used in most industries these days. Analytics combines mathematics, statistics, probability, and heuristics theories. So now, after this brief description, we all understand that analytics is easy, right?

Mathematical disciplines, including statistics, probability, and others, have always been assigned to practical applications in some form or another, including those of business. Sometimes the relationship between the method and the application is quite clear, and sometimes it is not. As human beings, we have used trade since we started to relate to each other. Because of a need to interact we'd exchange goods. Mathematics was, and still is, a wonderful tool to support exchanges and business purposes—used to address issues from the simple to the complex.

The word *mathematics* comes from the Greek word *máthema*, which means learning, study, science. This word has always related to describing quantities, to depicting structures of distinct types and shapes, and to portraying spaces. Pattern recognition is forever an elusive topic for mathematicians. Recognizing patterns, whatever it is they represent and from whatever source they emerge, people hope to better understand the past, and as a rule, understand the past to foresee the future.

History (the study of past social and human perspectives) is often used as a way to understand former events, learn from them to explain the present, and possibly help understand and even predict the future. We document how societies, governments, groups of people, ghettos, all clusters of some sort, behave prior to a particular event. By understanding and describing these sorts of scenarios (like documenting them), historians try to explain events that took place to provide context of what happens in the future.

Historians and mathematicians are quite similar that way. For mathematicians, past events and their associated facts are gathered together and analyzed in order to describe some set of data. If the attributes describing those past events are not too large in terms of size or scope, a mental correlation might be done.

This is what social and human researchers often do. Given some attributes along with a description of a particular scenario in relation to a past event, they correlate distinct aspects of the event itself to explain what happened. And, based on the description and the strength of correlation between event attributes associated with the past event, they also try to foresee what might happen next.

Although the methods may differ, this isn't far removed from what mathematicians do. We may use different terms (in fact, that is one of the plaguing realities of every discipline), but we analysts do the same thing. Mathematics is just another way to map what human beings do, think, understand, act, explain, and so on. What is a mathematical formula in the end? It is a way to describe the future by understanding past observations (well, at least one branch of mathematics does this). We used to do this in the past, when mathematics was used to count things for trade. Today, we examine scenarios with specific conditions and particular constraints, calculating outcomes for use in operations and decisions.

As with any other discipline, mathematics has evolved over time. In addition to counting physical objects, early uses of mathematics involved quantifying time, defining values for trade (bartering of goods lead to the use of currency), measuring land, measuring goods, and many other areas. As time passed, arithmetic, algebra, and geometry were used for taxation and financial calculations, for building and construction, and for astronomy. As uses continued to evolve, so did mathematical theory. Mathematics became a tool to support different scientific disciplines such as physics, chemistry, and biology, among others. And it is likely that mathematics will be used as a tool for disciplines that haven't even been defined yet.

In order to play this important role and support science, mathematics must be quite rigorous. One of the rigors of mathematics is proofs. Mathematical proof is a method that turns theorems into axioms by following a particular set of laws, rules, constraints, meanings, and reasons. The level of rigor needed to prove theorems has varied over time, changed in different cultures, and certainly extended to satisfy distinct scenarios for political, economic, and social applications.

But a crucial truth of mathematical proofs is that they are often simply a heuristic process. The procedures used to prove a particular theorem are often derived by trial and error. Heuristic characteristics might be present in the entire theorem scenario, or just at the inception. But in reality, heuristics is often involved in one or more stages of proof. Heuristics can support the definition of the proof from simple observation of an event (therefore helping to recognize a pattern), or heuristics may govern the entire mathematical proof by exception or induction.

Most mathematical models have some limits in relation to the set of equations that portray a particular scenario. These limits indicate that the equations work properly given specific conditions and particular constraints. Consider standard conditions for temperature and pressure in chemistry. Some formulas work pretty well if temperature and pressure are in a specific range of values, otherwise, they don't. Is the formula wrong? Or could it be that at other times in history, the formula was valid, but not now? The answer is no, on both counts. It just means

that this particular formula works pretty well under some specific conditions, but not others. Remember that formulas are built to model a particular event, based on a set of observations and, therefore, this formula will work fine when the constraints of the model scenario are true.

This is a perfect example of why mathematical models can describe a particular scenario using one or more equations. Although these equations may properly depict a specific scenario, the methods work properly if, and usually just if, a particular set of conditions are satisfied. There are boundary conditions limiting accuracy of equations, and therefore, a model simply represents some specific scenario. These boundary conditions are sometimes just a specific range of possible values assigned to constants or variables.

Think about physics. The classical mechanical physics of Isaac Newton describe regular movements considering regular bodies (not too small and not too big), traveling at regular speeds (not too slow and not too fast). However, once we start to consider very high speeds, such as the speed of light, the classical Newtonian theories no longer describe the movements. This doesn't mean the formulas are wrong; it just means that the particular scenario Newton wished to describe using those formulas requires a particular set of condition and constraints. To describe regular bodies at very high velocity we need to use Einstein's formulas. Einstein's theory of relativity (and his corresponding formulas) describes the movement of regular bodies at very high speeds. His theory doesn't conflict with Newton's; it simply explains a different scenario. Eventually, very small bodies at very high speeds also needed a distinct theory to describe their movement, and quantum physics was born. As we continue to learn, and delve even more specifically into areas of physics (and the authors would argue, any discipline), the need for different methods will continue—a story like this one never ends.

And so, we are not that different from historians. As mathematicians, statisticians, data miners, data scientists, analysts (whatever name we call ourselves), we too take into consideration a set of attributes used to represent/describe a particular scenario, and analyze the available data to try to describe and explain patterns or predict (which is, as a matter of fact, a particular type of pattern).

Once upon a time, a mathematician named Edward Lorenz was quite focused on predicting weather. At that time, this sort of work was largely based on educated guesses and heuristics. Is it so different today? Maybe (guesses are still fundamental and the heuristic process is undoubtedly still present). Back in Edward Lorenz's day, weather prediction included assumptions, observations, and lots of guesswork—in spite of the scientific instruments available at the time. When computers came onto the scene, Lorenz foresaw the possibility to combine mathematics with meteorology. He started to build a computer mathematical model, using differential equations, to forecast changes in temperature and pressure. Lorenz had created a dozen differential equations and managed to run some simulations and estimate virtual weather conditions. This was certainly a dramatic improvement over historical guesswork. During the winter of 1961, Edward Lorenz was examining a sequence of numbers. At the time, computers used a total of six decimals for all observations. However, in order to save space

walk. It explains all notions of pubs and Guinness. If a drunkard is at the end of the night and there is a combination of pubs each with Guinness, what is the fastest way to consume . . . sorry, we digress. The important thing here is that there are events in business, science, and many other environments where knowledge about the past doesn't in fact give us appropriate knowledge about the future.

The drunkard's walk phenomenon refers to the element of randomness assigned to this particular process. And if we decide to be totally comprehensive, this randomness is assigned to most processes we experience in life. It is similar to optimization problems. We always want to consider the problem as linear, but if we look deep inside it, it is most often a nonlinear problem. But this is par for the course. In order to make it easier to solve problems, we accept approximations. And these approximations don't significantly change the results. Given that, we feel better and more confident in using them.

So, after all that, analytics is quite easy, isn't it?

Sorry, but that magical box still doesn't exist. Analytics isn't that easy after all. The truth is that analytics can be quite hard, precisely because there are heuristic processes at work. Taking into account the heuristic factors of analytics, wherein trial and error processes are used to discover useful knowledge, or when past events don't lead to future forecasts, and when small noises in the initial conditions can dramatically change final results, we can easily assume that analytics is not easy at all.

In the development of analytical models we need to consider how heuristic the business scenario is. We need to examine past occurrences to understand if the scenario is governed by stochastic process or not. We also have to properly define initial conditions for the model being created.

For each situation there is a particular technique employed. Sometimes, even for the same situation, considering exactly the same scenario, there is more than one technique to be used. Comparison and assessment processes are needed in order to select the model representation of the available data. Data mining models are based on the patterns in data, and so accuracy and performance are highly influenced by the data being studied.

We have a plethora of techniques, methods, and types of analytical models to choose from for a specific business objective. On the other hand, we have a huge amount of data to process. The only way to move forward is by using the proper tool or set of tools. Analytics is no longer mathematics, statistics, or probability. Analytics has become computer science. Analytics involves all those theories (math and stats, etc.) via the right computer program. The appropriate set of computer programs that are grounded in theory, but that can also handle huge amounts of data is the new pen and paper.

This book focuses on analytic processes and how they fit into our heuristic world.

In spite of the strong heuristic characteristic inherent in analytical processes, this book emphasizes the need to have the proper tools and approach to analytical problem solving. Randomness has a valid impact on virtually everything we do, on everything that happens, and on business events. Nevertheless, this

fact makes it even more important to put in place analytics to assess the randomness, and to understand business events, marketplace scenarios, and behavior.

Competitive marketplaces are compelled to use analytics. To perform better, organizations need to understand not just the market and those within it but how consumers behave when they have options. By understanding customer behavior, companies are able to take business actions in an attempt to encourage different behavior—like enhancing service usage, improving product adoption, extending customer lifetimes, reducing insolvency issues, and more. The effectiveness odds improve when companies act in the marketplace based on the knowledge generated by analytical models. Why? Because they are placing their bets more wisely—investing in customer-focused strategies that they have some evidence will work, and measured expectations regarding what the results are likely to be. Increasing customer spending and service adoption and reducing losses from fraud, for example, change the portfolio of behavior of the customer base. If it didn't, then the strategies employed wouldn't be effective. So if you are effective, then the behavior of customers will change, as will the associated data describing them. The only real constant, in fact, is change.

This book is in praise of analytics, no matter how heuristic our world is.

Carlos Andre Reis Pinheiro and Fiona McNeill

July, 2013

Acknowledgments

I started writing this book in 2011, first in Dublin as a visiting researcher, and then with Oi Communication in Brazil. During the commute from Niteroi to Rio de Janeiro (usually by bus or ferry), I read three books. Actually I read more than three, but these three impacted the desire to write this book. The first one was *The New Know: Innovation Powered by Analytics*.[1] Thornton May brilliantly explored the power of analytics to solve business problems, allowing companies to innovate. The book really gives us a sense of how important analytics is and how crucial it is to companies nowadays. However, as most of the books do, it describes what goes right. As with a formula, if we properly select the variables, we may get the expected result. Unfortunately, analytics is not a science. It is based on science, but it is also an art. Maybe analytics is really about the art of how to use science to solve problems. But anyway, this book was an inspiration. So, my first acknowledgment is to Thornton May.

The second book was *Blink: The Power of Thinking without Thinking*, by Malcolm Gladwell.[2] This book sticks to you. I missed my bus stop many times because of it. It shows how often we make decisions without thinking. Let me rephrase that: How often we make decisions thinking we are thinking about these decisions. In a flash, when required, we capture information in our brain, we gather it appropriately, weighing the pros and cons, and just then—we make the decision. Sometimes right, sometimes wrong. But the interesting thing in here is all these happen in a blink of an eye. This book showed me how heuristic the process of thinking truly is, particularly when we consider time. We also learn from the process itself and eventually we make better decisions. And I thought to myself, these books are seemingly opposite, but they aren't, not really. They are complements. Analytics really helps solve most of the business problems we currently face. But the process to solve these problems sometimes is far from being exact, as it is in science (actually in many times science is not like this as well). Very often we do analytics thinking without thinking. And most of the times, it works. So my second acknowledgment is to Malcolm Gladwell.

Finally, the third book that was a catalyst for this book, and has driven many studies in Heuristics, is *The Drunkard's Walk: How Randomness Rules Our Lives*, by Leonard Mlodinow.[3] In this book Mlodinow explores a series of random events that touch our lives on a daily basis. He describes historical mathematic stories, including experiments covering probability, demographic and economic data, statistics, logic, regression, and prediction. The book illustrates how our lives are profoundly impacted by change and randomness, and how everything from wine rating and corporate success to school grades and political polls are less reliable than we believe.

The sum of how analytics can solve real problems, how important it is to believe in your feelings and in your way of thinking, and given that, in the end, everything is biased by chance—found in these three books formed the essence

of this work.

Unfortunately, I never met any of these brilliant authors. But even so, I also want to thank the people I actually do know. And for that, I'll do this chronologically. I begin with my former boss, Bruno Martins, who created an effective environment for me and the other members of the Oi team, for being brave in embracing the idea of an Analytics Lab at Oi. Also thanks to my peers Herivelton Pereira, Jair Carvalho, Flávio Sebastião, Rodrigo Telles, and Rosane Lisboa. Other Oi colleagues who helped create the encouraging atmosphere that began this endeavor are Bruno Kitsuta, Ana Sidreira, Daniella Alfaya, Vitor Cordeiro, Nivaldo de Moraes, Adriana Ferreira, Michelle dos Santos, Larissa Borges, and Ana Paula Cohen.

Other industry colleagues that deserve special mention include Luis Francisco de Campos Henriques from Brasil Telecom and Karl Langan from SAS Ireland.

A special thanks to Fiona McNeill who honored me by coauthoring this book. Fiona gave me new perspectives for my own ideas, organizing them, putting them into a more understandable shape, and then creating a perfect atmosphere for sharing innovative thoughts about analytics. Because of her, this is a brand new book, quite different than the one I started. And I am so proud of this one!

■ ■ ■

Indeed Carlos, it has been my honor working with you! Frankly, when I was first introduced to some of the initial concepts of the text, Carlos won me over with butterflies and beer. His openness, responsiveness, and dedication, even as he moved across the globe, were inspirational. I'd like to thank Carlos for his gracious sharing of time, thoughts, and talents in coauthoring this book. I'm pretty proud of it too.

I'd like to also thank Buffie Silva because, if it wasn't for her, I'd never have had the opportunity to work on this project. She got the ball rolling, all the necessary wheels on the bus, cheering along the way. I sincerely appreciate your faith.

■ ■ ■

In addition to the inspiration I was gifted from Carlos, I'd like to acknowledge four key people who, over the years, inspired me and encouraged me to go outside the box, teaching me things in my career that I drew on to complete this book. So, to Tony Lea, Greg Breen, Joel Grosh, and Don Goodland—I extend my sincerest gratitude. Your words, deeds, and mentorship continue to echo within me.

■ ■ ■

And finally, to Stacey Hamilton, who had the challenge of reviewing the book. Stacey you are more than a reviewer, guiding us along the exhausting process of writing the book (and poking us from time to time too). Your expertise as editor and patience as a colleague made this book real. Thank you so much Stacey (once again)!

1 May, Thorton. *The New Know: Innovation Powered by Analytics* Hoboken, New Jersey, John Wiley & Sons, 2009.

2 Gladwell, M. *Blink: The Power of Thinking without Thinking*., New York, Little Brown and Company, Time Warner Book Group, 2005.

3 Mlodinow, Leonard. *The Drunkard's Walk: How Randomness Rules Our Lives*. New York, Vintage Books, 2009.

Chapter 1

Introduction

Analytics is used to address many different types of business problems. It is used to understand customer behavior and how consumers may be adopting new products and services. It is used to describe different marketplace scenarios and their impacts. It is also used to decipher competitor's movements and patterns. And it is used for predicting potential future revenue, detecting risk, uncovering fraud, minimizing bad debt, optimizing operational processes, and more. Analytics is used in all of these business applications.

In most cases, and in particular in customer scenarios, there are many factors that cause misunderstanding of what is currently happing within a market, or even what is happing with a specific customer within a particular market.

It is always important to bear in mind that consumers present different types of behaviors and in accordance with the market they are interacting with. As a customer, I can be very aggressive in terms of purchasing high-tech products, often buying cutting-edge gadgets. However, I am quite conservative in terms of investing, putting my money into low-risk accounts. There is no one, overall general behavior for any customer. We each behave in different ways depending upon the situation in which we find ourselves. Essentially, we wear different hats, having distinct behaviors that are observable—each in relation to the distinct roles we play. And sometimes, even in similar scenarios, we may play different roles and exhibit different behaviors, depending on the other scenario actors that are involved.

All analytical models, whether they are supervised (classification), semi-supervised, or unsupervised (segmentation), take into consideration most of the structured information that companies currently hold in their databases. They include information about customer characteristics, the products and services that they offer, and how customers interact with them. They include financial inputs such as credit rating, payment history, late payment durations, and so forth. All of this information, however, describes only a limited part of the end-consumer's behavior. In other words, we really can't say too much about an individual customer's profile, but we can describe how they behave in a single scenario, like when using the company's products and services.

You could say that, based on my historical data, I am an aggressive buyer of high-tech gadgets. And it is just as possible to state, based on my buying history, that I work hard to purchase high-tech products in advance. But this behavior doesn't replicate to other situations, like my conservative financial investment

Index