Computational Intelligence in Business Analytics

Concepts, Methods, and Tools for Big Data Applications

Les Sztandera

Professor of Computer Information Systems,
Philadelphia University

Computational Intelligence in Business Analytics

This page intentionally left blank

Computational Intelligence in Business Analytics

Concepts, Methods, and Tools for Big Data Applications

Les Sztandera, Ph.D.,
Professor, Computer Information Systems,
School of Business Administration,
Kanbar College of Design,
Engineering and Commerce,
Philadelphia University

Associate Publisher: Amy Neidlinger
Executive Editor: Jeanne Glasser Levine
Operations Specialist: Jodi Kemper
Cover Designer: Alan Clements
Managing Editor: Kristy Hart
Project Editor: Elaine Wiley
Copy Editor: Bart Reed
Proofreader: Debbie Williams
Indexer: Lisa Stumpf
Compositor: Nonie Ratcliff
Manufacturing Buyer: Dan Uhrig

Upper Saddle River, New Jersey 07458

Printed in the United States of America

First Printing June 2014

ISBN-10: 0-13-355208-X
ISBN-13: 978-0-13-355208-9

Pearson Education LTD.
Pearson Education Australia PTY, Limited.
Pearson Education Singapore, Pte. Ltd.
Pearson Education Asia, Ltd.
Pearson Education Canada, Ltd.
Pearson Educación de Mexico, S.A. de C.V.
Pearson Education—Japan
Pearson Education Malaysia, Pte. Ltd.

Library of Congress Control Number: 2014934908

To my family for their unfailing support,
without whose help my passion for computational
intelligence would not have been fully realized.

This page intentionally left blank

Contents

Foreword

"In God we trust. Everyone else bring data."
—W. Edwards Demming

We are at the dawn of a new era. An era that almost crashed and died quietly without many people noticing or paying attention. This era of computational intelligence is about to make all of the accomplishments by the baby boomers obsolete. In this era, computational intelligence paradigms will be embedded into everything—your credit card, the train tracks, the books you read, the stores you frequent, your body! This era will cause massive disruption in businesses throughout the globe. This is an era where new business empires have rapidly emerged—the likes of Google, Facebook, and LinkedIn—and this will be an era where industry giants who are not forward thinking go by the wayside. But by far, a bulk of the industry giants around the globe are in a race to figure out how to capitalize on this maturing technology—to make their businesses not just survive but thrive. They strive to make their businesses "frictionless"—to relentlessly pursue excellence while we are sleeping, to automatically adjust to the ever-changing world without human intervention and delays, and to take into account complexities that our marvelous brain takes into account automatically, but when we try to articulate these complexities we get lost in the process.

Computational intelligence has its roots in artificial intelligence (AI), which started very humbly at Dartmouth College in 1956. In the 60s, artificial intelligence research was well funded in both the United States and United Kingdom. However, by the mid-70s, much of the early excitement and promise of AI had evaporated. Once again in the early 80s, commercial enterprises started investing in research and a new generation of "expert systems" appeared, creating a billion-dollar market. However, by the late 80s the rigidity of those rules-based systems faded into obscurity. Very quietly in the 90s, AI gained new momentum via a foothold in supply chain and logistics, which was being swept up as part of the business process reengineering projects that spanned the globe in an attempt to drive process efficiency and cost savings. The complexities of supply chains with hundreds and thousands of customers, plants, suppliers, and modes of transportation proved to be a very good proving ground for the highly evolved optimization techniques honed through decades of AI research.

From that early launch pad, AI shed its moniker and has taken on many new names—predictive analytics, data mining, advanced analytics, simulation, optimization, just to name a few—to distance itself from the failed promises. From this new vantage, computational intelligence has emerged to deliver concrete results and value for businesses—just as supply chain and logistics optimization did in the early 90s. Many of these early successes have been kept as "trade secrets" by businesses around the globe. Today, after more than 50 years of starts and fits, computational intelligence is ready to deliver on the initial promises of artificial intelligence through a whole new generation of savvy business managers and technologists.

As many of the innovative business executives who made the decision to keep the early computational projects a trade secret know, this technology can deliver amazing business value—business value that can create a *sustainable* competitive differentiation, which is much easier to aspire to but much harder to attain. While the early projects required a lot of validation—because oftentimes the results were so massive that they defied all common wisdom—the new generations—the X-ers, the Millennials, and soon-to-be the Z-ers—trust and believe their technology much more than their own (or anyone else's) intuitions. This new wave of data scientists, leaders, managers, and consumers expect the computer to deal with complexities and make optimal decisions every time. They won't settle for anything less.

This book is the indispensable guide for those of you not familiar with the technology—whether you're a Boomer or part of the new wave. In this book, you'll learn the essence of the technology and understand how to apply it to practical real-world problems. Many of the examples cited contain transferable lessons to many other industries and problems. Embrace the examples and figure out how to make them work in your business and how to improve upon them so they create further value for your business. Then rinse and repeat until you've fully embedded this new wave into the very fabric of your business. That's how you get and keep sustainable competitive differentiation in this new era.

Michele Chambers
Chief Strategy Officer, Revolution Analytics

This page intentionally left blank

Acknowledgments

I would like to express my appreciation to Dr. Stephen Spinelli, Jr., for his vision to transform higher education. Under his leadership, Philadelphia University has become the model for professional university education. As part of the University's bold Strategic Plan, Dr. Spinelli led the way in formalizing the University's signature approach to teaching and learning, Nexus Learning, and establishing the unique College of Design, Engineering and Commerce, which is underpinned by an innovative curriculum based on trans-disciplinary, active, and real-world learning that is infused with the liberal arts—one where data analytics play a crucial role in academic discovery.

This page intentionally left blank

About the Author

Dr. Les Sztandera is Professor of Computer Information Systems in the School of Business Administration at Philadelphia University Kanbar College of Design, Engineering and Commerce. His research interests include data analytics and computational intelligence. He received his Doctor of Philosophy degree from the University of Toledo, Ohio, and his Master of Science from the University of Missouri. Dr. Sztandera has taught undergraduate and master's level courses as well as doctoral seminars in computational intelligence for more than 20 years. He was a recipient of the highest and the most prestigious appointment in the U.S. Fulbright Scholars program in 2003, and served as Fulbright Distinguished Chair at the School of Business and Economics (ISEG) in Lisbon, Portugal, teaching in the MBA and Ph.D. programs there.

Currently, Dr. Sztandera is teaching innovation MBA program courses in Technical Competitive Intelligence and New Product Development, as well as being involved in multidisciplinary industry-sponsored research projects as part of the new curriculum that draws on Philadelphia University's rich history in innovation and design, and long tradition of excellence in teaching and researching integrated product development. The curriculum meets the emerging needs of industry to educate managers, industrial designers, and engineers into more accomplished practitioners in the global product development processes.

Dr. Sztandera draws on his experience in the design and delivery of innovative curricula to facilitate crafting of cross-disciplinary projects. His cross-disciplinary work in encouraging the inclusion of computational intelligence principles and competences in undergraduate education led to the National Science Foundation grant award for Philadelphia University through the Division of Undergraduate Education in 1996. Since joining Philadelphia University in 1993, Dr. Sztandera has been involved in a number of course and curricula developments. He proposed a sequence of courses coupled with laboratory assignments in which students worked on real-life problems. In 2004 he was instrumental in developing a Business Plus program for entering freshmen.

Complementary with his teaching effort, Dr. Sztandera has been involved in a variety of research activities in the field of computational intelligence. His research was funded by the U.S. Department of Defense, U.S. Department of Commerce, National Science Foundation, State Supercomputer Centers, and the American Heart Association, among others. Dr. Sztandera is an eminent scholar, and has a significant publication and teaching record. He has delivered papers, seminars, and workshops, and has published extensively on computational intelligence issues, as well as served as a Chair of International Conferences in the field.

1 Overview

Learning Objectives

- To know how to define data analytics
- To understand the function and uses of data analytics
- To demonstrate an understanding of current theory and research avenues in data analytics
- To understand why computational intelligence occupies a place in big data analytics
- To appreciate areas of ethical sensitivity in data analytics

1.1 Introduction

This book aims to help organizations gain a competitive edge in the marketplace through harnessing the power of computational intelligence approaches. Those approaches—fuzzy sets, artificial neural networks, and genetic algorithms—are at the core of every innovative business, from large corporations to small companies. Businesses that do not leverage computational intelligence will be quickly outperformed by those that do.

The chapters are designed to provide a foundation upon which one can differentiate one's business through computational intelligence approaches. Thus, the book provides the reader with guidance on how to create acceptable models in a relatively short period of time, and how to arrive at the right innovative decisions before the competition does.

The primary purpose of this book is to facilitate education in the resurgent computational intelligence areas of artificial neural networks, fuzzy sets, and genetic algorithms. The book is written as a text for a course at the graduate or upper-division undergraduate level. It could also be used for short intensive courses of continuing and executive education or as a self-study. No previous knowledge of computational intelligence tools is required to understand the material in this text.

Whereas most (if not all) literature on the topic utilizes statistical software packages, this book urges managers to take advantage of computational intelligence for analysis, exploration, and knowledge generation. As a result, readers are provided with the needed guidance to understand, model, discover, and interpret new patterns and new knowledge from historical evidence and large data sets, and become adept at building powerful models for prediction and classification that do not rely on statistics.

A process based on the exploration of business data with an emphasis on statistical and/or computational intelligence analysis is called Business Analytics. It is used by innovative companies committed to data-driven decision making to gain insights that inform business decisions, and down the road it can be used to automate and optimize business processes. Data obtained through Business Analytics are treated as corporate assets (added value) and are leveraged to gain a competitive edge.

The outcome of Business Analytics depends on data quality (to avoid junk in, junk out), entrepreneurial business analysts who understand the analysis and the business itself, as well as an organizational commitment to data-driven decision making. It should be stressed that entrepreneurial and skillful business analysts are at the core of obtaining competitive advantage for the business. They should operate at every level of their organizations instead of being an elite group of data scientists reporting directly to the executive suite. Those business analysts make the discovered insights actionable, as they discover new knowledge and utilize predictive power of computational intelligence approaches.

Business Analytics itself is used for the following purposes:

- To explore data to find new patterns and relationships (data mining)
- To evaluate and test previous decisions (randomized controlled experiments, multivariate testing)
- To explain why a certain outcome happened (statistical analysis, descriptive analysis)
- To venture into the future (forecast) results (predictive modeling, predictive analytics)

Once the business goal(s) of the analysis is agreed upon, an analysis methodology is selected and data are acquired to support the analysis. Data acquisition often involves extraction from many sources and business systems, data cleaning, dimensionality reduction, feature evaluation, and subsequent integration into a single repository, such as a data mart or a larger data warehouse. Competitive intelligence utilizing intelligent software agents might be used to locate and extract some of the needed data. The analysis is typically performed against a smaller sample set of data to verify its applicability first, and then used on all historical evidence the business has accumulated.

Analytic tools range from spreadsheets with statistical functions to complex data mining and predictive modeling applications. As patterns and relationships in the data are discovered, new questions are asked, new queries are sent, and the analytic process iterates until the business goal is met. This could be optimized by additional technologies such as optimization and/or genetic algorithms.

Deployment of predictive models involves ranking data and information, evaluating their "novelty indexes," and using the ranks to optimize real-time decisions within applications and business processes. One can also utilize Business Analytics at the tactical level of a business pyramid to tackle unforeseen events, and in many cases the decision making could be automated to support real-time inputs. It is its predictive power that makes computational intelligence insights actionable. The finding associated with an action that is deemed reliable, based upon past data, gives the decision maker a high degree of confidence.

Each chapter in this book is followed by a set of exercises, which are intended to enhance the understanding of the material presented in the chapter. The solutions to a selected subset of these exercises are provided in the instructor's manual, which also contains further suggestions for use of the text under various circumstances. The exercises are of varying levels of difficulty. The following rating system is applied to approximately indicate the amount of effort required for solving the exercises:

- **[Level 1]—**Problems at Level 1 are solvable within a day. They test the comprehension and mastery of fundamental concepts. If they relate to the use of software or writing computer code, the programming time is short.
- **[Level 2]—**Solving problems at Level 2 can take days or weeks (e.g., proof of concept programming or implementation). The chapters provide all the information necessary for solving Level 1 and Level 2 problems.

- **[Level 3]—**Problems at Level 3 are even harder, and their solutions can take several weeks or even months (e.g., semester-long projects). Many of these exercises are related to innovative avenues of current research.
- **[Level 4]—**Problems at Level 4 concern open research questions and could be topics of graduate theses or dissertations. Solving Level 3 and Level 4 exercises typically requires doing further literature searches and/or conducting extensive experiments.

It is recommended that the reader do the Level 1 and Level 2 exercises, and tackle at least some of the problems at Levels 3 and 4. Carefully working through Level 1 and Level 2 problems will reward the reader with a thorough understanding of the material of the chapters, and solving Level 3 and Level 4 exercises could turn a reader into an innovator!

1.2 A Need for Computational Intelligence in Business Analytics

There exists a need for a resource that can be drawn upon by an innovative organization that seeks to gain a competitive edge in the marketplace through harnessing the power of computational intelligence approaches. This book provides managers with the tools, knowledge, and strategies to differentiate their business and to successfully add value to their organization. The driving force behind this differentiation lies in utilization of computational intelligence approaches, such as fuzzy sets, artificial neural networks, and genetic algorithms. Those approaches are much more adequate for dealing with uncertainty and the complexity of today's organizations than the statistical analysis and tools currently utilized.

1.3. Differentiating Your Business Through Computational Intelligence

This book aims to provide a thorough introduction to the main issues associated with the design and implementation of computational intelligence tools that could add value to an organization.

In general, computer systems are not good at knowing what to do: every action the system performs must be explicitly anticipated, planned for, and coded by the programmers. If a computer system encounters a situation that its programmers did not anticipate, then the situation usually results in a system crash. For the most part, business managers accept that computers satisfy a purely computing (number crunching) role. For many applications (such as payroll processing), it is entirely acceptable. However, for an increasingly large number of business processes, managers require computer systems that could decide for themselves what they need to do in order to satisfy both their design objectives and tackle a given business problem. Such computer systems utilize computational intelligence tools. They must operate robustly in rapidly changing and often unpredictable environments. Thus, to some extent, these computer systems are *anthropomorphic* and utilize the power of computational intelligence techniques, such as fuzzy sets, artificial neural networks, and genetic algorithms.

The techniques are outlined in the book, and examples for suggested implementations are provided. It should be noted that these approaches are much more adequate for dealing with multiple data (both structured and unstructured) as well as with uncertainty and the complexity of today's organizations than the statistical analysis and tools most commonly used. Computational intelligence tools provide actionable insights for decision making in addition to their capability to explain the past.

For example, these tools (fuzzy sets, artificial neural networks, and genetic algorithms) could address problems and tackle tasks of product mechanical design within a framework of integrated design. The mechanical design process is usually divided in several subproblems from engineering and programming points of view. The fuzzy sets approach allows for the comparison of different stakeholders' points of view to one another and the final solution through a global compromise. This approach allows the mechanical design to be distributed in parallel tasks. The genetic algorithms and artificial neural networks tools could be used in order to encapsulate the data and analysis of each engineering design point of view. So, they allow for multiple constraints (factors of materials selection, reliability, performance, safety, and environmental impacts) to be incorporated into the mechanical design. Genetic algorithms and artificial neural networks also allow for exploration of new mechanical design solutions, thus fostering innovation.

Even more autonomous behavior of computer systems utilizing computational intelligence tools is expected in scientific applications. For example, when a space probe makes its long flight from Earth to outer planets in the solar system and beyond, a large ground crew of scientists is usually required to continually track its progress and decide how to deal with unexpected eventualities. This is not practical and very costly. For these reasons, organizations like the National Aeronautics and Space Administration (NASA) have been experimenting with making space probes more autonomous—giving them richer decision making capabilities and responsibilities, in part by utilizing computational intelligence tools.

Computational intelligence poses both a challenge and an opportunity for many businesses and organizations, and data scientists and researchers are at the forefront of learning how to leverage these changes for business impact.

I hope you enjoy learning about computational intelligence and its tools and how the industry is adapting to today's environment.

Exercises

1. [Level 1] Business internal databases are repositories of data and information gathered by a company, typically during the course of business transactions. They could be augmented by external, secondary data sources. Companies gather information about customers when they purchase a product, inquire about a service, or have a product serviced. Business internal databases are used by the companies to strengthen their relationships with customers and for direct marketing. Those databases could become quite large over time, and dealing with the vast quantity of data poses quite a challenge for decision makers. The managers are often further hindered by the relational nature of a database.

   Computational intelligence software helps managers make sense of the enormous mass of information contained in non-relational data warehouses. The software is capable of creating new knowledge that is actionable for decision makers, as it does not just have the ability to explain the past, but also possesses predictive power. Micromarketing refers to using a differentiated marketing mix for specific customer segments, sometimes fine-tuned for the individual shopper. Data analytics make micromarketing drive sales and business profits.

   Like many businesses nowadays, Target has utilized data analytics to micromarket—that is, to target each consumer with promotional materials designated for that individual. *The New York Times* reported on those practices on February 16, 2012 [Duhigg, 2012]. According to the report, titled "How Companies Learn Your Secrets," Target collects data on every person who shops at its department stores, assigning a unique code known as customer ID. If the consumer uses a credit card or coupon, fills out a survey, mails in a refund form, calls the customer center, or opens an e-mail, Target links those interactions with the customer ID. The company also collects demographic information, such as age, gender, ZIP Code, marital status, number of children, place of residence, and income, and has the ability to purchase additional data. Like many other innovative businesses, Target utilizes data analytics in its marketing research. One avenue of these analyses is to relate to the customers during major life events, such as having a baby, graduating from college, moving from state to state (or coast to coast), and so on.

How does Target identify specific customer segments, such as pregnant female consumers, for example? Target data scientists use the store's baby registry to identify consumers who have used it, and then backtrack to find out what products they had bought early in their pregnancy. The researchers discovered unique patterns that became clusters, and subsequently classes of consumers, and associated them with products. It appeared that women in their first 20 weeks of pregnancy purchased supplements such as calcium, magnesium, and zinc, and subsequently bought a lot of unscented lotions in their second trimester. In their third trimester, they purchased washcloths, hand sanitizers, soap, and cotton balls. Overall, Target data scientists identified 25 products, and came up with a "Pregnancy Prediction Score." This is just one of the metrics Target uses. It applies those metrics to all customers, and those who score high enough are contacted. In this particular "Pregnancy Prediction Score" case, the high-scoring customers are assumed to be pregnant, and receive targeted promotions on products Target predicts they will need. Reportedly, Target sales on mom-to-be and baby products have increased since the data analytics tools were applied.

Similarly, many of the company's online customers' data browsing habits are collected and analyzed. Do you think these data analytics practices are ethical?

1.1. [Level 1] Provide documented examples of data analytics usage by businesses. How many of these companies gained a competitive edge in the marketplace through harnessing the power of computational intelligence approaches?

2. If you do an Internet search on the phrase "faith-based businesses," the results direct you to companies that pursue a religious agenda. But, according to Fast Company [Safian, 2014], there is another kind of faith in business nowadays: the belief that a product or service can perform a radical industry makeover, completely change consumer habits, challenge economic assumptions, and enable city, county, and state officials to be proactive about health care, weather, and traffic emergencies.

With $452 million distributed in 2013, Bloomberg Philanthropies is among the largest philanthropic foundations in the United States. It differentiates itself by utilizing innovative and sophisticated data-driven solutions in its business processes. As a result, the foundation has been extraordinarily effective, as it positions itself for maximum impact [Safian, 2014].

2.1. [Level 2] Examine the data analytics tools used by Bloomberg and write a comprehensive report on them. Identify avenues of possible use of computational intelligence to further Bloomberg's philanthropic pursuits.

3. Fierce competition, time-to-market pressure, and an increasing demand for product differentiation call for more sophisticated, yet rapid product design. Businesses are increasingly seeking more efficient ways to integrate consumers' preferences into the product design process. Taking advantage of techniques from the field of computational intelligence, it is possible to construct systems that can computationally design products with specified desirable consumer characteristics.

3.1. [Level 3] Conduct research on techniques from the field of computational intelligence that deal with optimal design of products. Write a case study to illustrate the application of the computational intelligence approaches in your industry.

3.2. [Level 4] Propose a research project for your Ph.D. dissertation that will support the construction and deployment of a sophisticated, computational intelligence model to design products that your business unit is involved with.

This page intentionally left blank

Index

G

H

I

J-K

L

M

N

O

P

Q-R

S

T

U-V

W-X-Y-Z