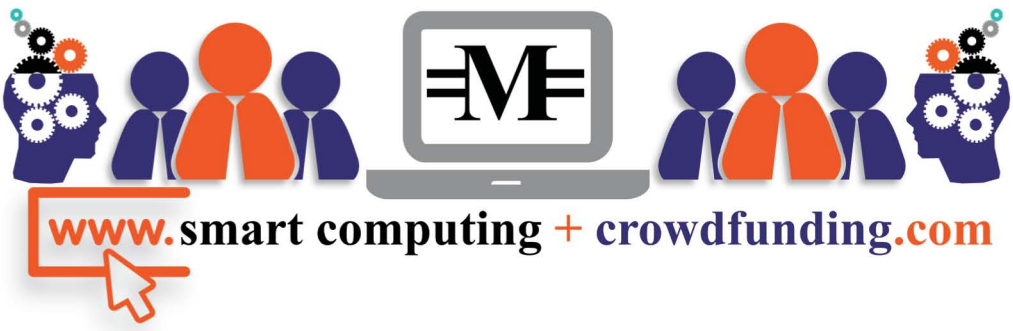
SMART COMPUTING APPLICATIONS IN CROWDFUNDING



Bo Xing and Tshilidzi Marwala

CRC Press
Taylor & Francis Group
A SCIENCE PUBLISHERS BOOK

Smart Computing Applications in Crowdfunding

Bo Xing
Institute for Intelligent Systems
University of Johannesburg, Johannesburg
South Africa

and

Tshilidzi Marwala
Vice Chancellor and Principal
University of Johannesburg, Johannesburg
South Africa

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A SCIENCE PUBLISHERS BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20180807

International Standard Book Number-13: 978-1-138-57771-8 (Hardback)

Foreword

In 2008, the world economy was brought to its knees by the worst financial crisis since the Great Depression. Many countries around the globe, if not every one of them, were impacted in one way or another, some suffered far more severely than others did. All this created a welcome breeding ground for the concept of crowdfunding: a cooperative, Internet-facilitated solution of apportioning funds and resources, across geographical confines, directly to the projects of interest, circumnavigating and complementing the incumbent financial services institutions. Unsurprisingly, crowdfunding is rapidly becoming part of the world's advancement towards a shared and digital economy. If 2008 was signalling a great increase in crowdfunding related activities, 2013 was the year crowdfunding started receiving global recognition and drawing the attention of various bodies, such as the established financial services industry, economists, politicians, and enterprises. The impression of the scale of the crowdfunding market has since attracted numerous large institutions to jump on the bandwagon.

Simultaneously, the scientific research underbuilding the crowdfunding phenomenon has been gaining momentum. Since 2010, crowdfunding relevant topics have been gradually studied from multiple perspectives across distinct disciplines such as psychology, social science, information and communication technology, economics, computer science, engineering, and entrepreneurship. But the establishment of an elaborated research domain is still far away from us. In 2013, the corresponding academic discussion on crowdfunding was still very rare. Only thereafter, some publications had begun to emerge here and there, and several conferences selected crowdfunding as a side topic, but the scope was nearly negligible from an academic viewpoint.

This is what *Smart Computing Applications in Crowdfunding* is about. In this book, Bo and Tshilidzi offer their readers a new angle from which to view crowdfunding, i.e., via less model-based smart computing algorithms which have their roots in engineering, in computer science, and in informatics. Though less mathematically rigorous to some extent, these intelligent algorithms, often nature-inspired, can cope with numerous real-world hard problems efficiently. Thus, the marriage of smart computing

and crowdfunding has the potential to spark novel methods, ways, and means of understanding crowdfunding that worth further dissemination and continuous development.

Overall, this book is a welcome addition to the literature of crowdfunding, artificial intelligence, granular computing, and beyond. I wish all of you lots of joy in reading this exciting work. So, read on.

Ben Shenglin, Ph.D.
Dean and Professor, Academy of Internet Finance
Zhejiang University, China
Director of International Monetary Institute
Renmin University of China

December 2017

Preface

Finance has a great impact on real economy's development. From a historical perspective, financial innovation is often coupled with technological advancement. Among various innovations in the financial sector, crowdfunding is a burgeoning and dynamic industry, in which a diverse variety of business models are incorporated. It is, thus, imperative to conduct a comprehensive exploration of the crowdfunding landscape. Meanwhile, the uprising of smart computing-enabled artificial intelligence is also broadly witnessed, which has inspired (or even forced) numerous sectors (including the financial sector) all over the world to react. Motivated by these two noteworthy phenomena, this book covers the key players and critical issues typically encountered in the crowdfunding domain from the smart computing perspective. Under each player, the application of smart computing technique(s) towards the representative issues is elaborated. It is hoped that this book, *Smart Computing Applications in Crowdfunding*, could be a timely publication that may meet the requirements of a wide spectrum of readerships.

The book consists of eleven chapters which are organized into seven parts. The interrelationship of chapters and sections is illustrated in Fig. P.1 (next page).

Acknowledgments

We would like to thank the University of Johannesburg for contributing to the writing of this book. We dedicate this book to the schools that gave us the foundation to always seek excellence in everything we do: The University of Cambridge and the University of Johannesburg.

Bo Xing, D.Ing.
Tshilidzi Marwala, Ph.D.
Johannesburg, South Africa

January 2018

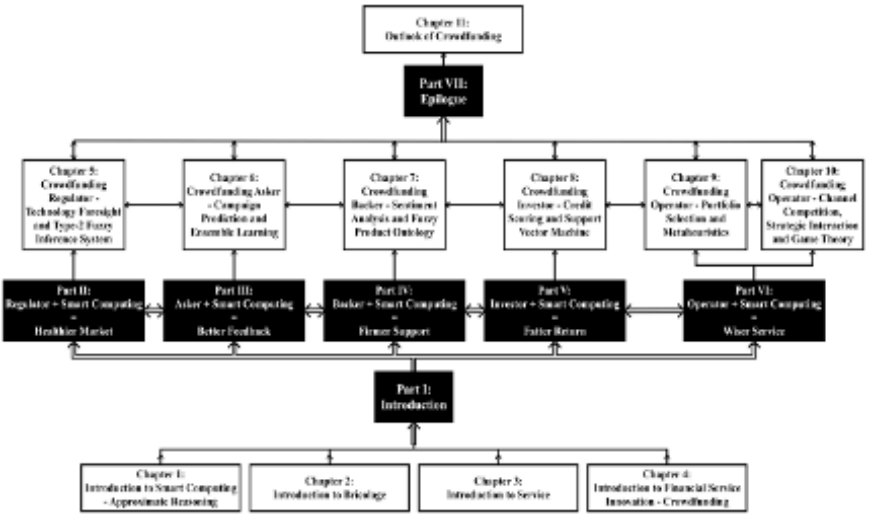


Figure P.1: Interrelationship among different chapters of the book.

Contents

List of Abbreviations

A number of key terms frequently used within this book are defined as below:

1IR	First Industrial Revolution
2IR	Second Industrial Revolution
3IR	Third Industrial Revolution
4IR	Fourth Industrial Revolution
ABM	Adaptive Basis-function Model
AI	Artificial Intelligence
AML	Anti-Money Laundering
ANN	Artificial Neural Network
AoN	All or Nothing
APAC	Asia-Pacific
ATMs	Automated Teller Machines
BMA	Bayes Model Averaging
CA	Correspondence Analysis
CaaS	Computing-as-a-Service
CART	Classification and Regression Tree
CF-IOF	Concept Frequency-Inverse Opinion Frequency
CSA	Canadian Securities Regulatory Authorities
DaaS	Data-as-a-Service
DAO	Decentralized Autonomous Organization
EaaS	Education-as-a-Service
ECOC	Error-Correcting Output Code
ELM	Extreme Learning Machine
EPO	European Patent Office

ESMA	European Securities and Markets Authority
EU	European Union
FINMA	Swiss Financial Market Supervisory Authority
FMCA	The New Zealand's Financial Markets Conduct Act
FSE	Fuzzy Synthetic Evaluation
FSVM	Fuzzy Support Vector Machine
GA	Genetic Algorithm
GT2FIS	General Type-2 Fuzzy Inference System
HC	Hill Climbing
HSI	Hamiltonian of Strategic Interaction
IaaS	Infrastructure-as-a-Service
IAIS	International Association of Insurance Supervisors
ICO	Initial Coin Offering
IFSs	Intuitionistic Fuzzy Sets
IOT	Internet of Things
IPO	Initial Public Offering
IT2FIS	Interval Type-2 Fuzzy Inference System
IVFSs	Interval-Valued Fuzzy Sets
IVIFSs	Interval-Valued Intuitionistic Fuzzy Sets
JOBS Act	Jumpstart Our Business Startups Act
KIA	Keep It All
KL	Kullback-Leibler
KYC	Know Your Customer
LDA	Latent Dirichlet Allocation
LIBSVM	Library for Support Vector Machine
LS-FSVM	Least Squares Fuzzy Support Vector Machine
MCA	Multiple Correspondence Analysis
NLP	Natural Language Processing
NPD	New Product Development
P2P	Peer-to-Peer
PaaS	Platform-as-a-Service
PBoC	People's Bank of China
PC	Personal Computer

PCA	Principal Component Analysis
PFM	Personal Finance Management
PSO	Particle Swarm Optimization
QDT	Quantum Decision Theory
SA	Simulated Annealing
SaaS	Software-as-a-Service
SEC	Securities and Exchange Commission
SMEs	Small and Medium Enterprises
SVM	Support Vector Machine
T1FIS	Type-1 Fuzzy Inference System
T2FIS	Type-2 Fuzzy Inference System
TFM	Technology Foresight Methods
UK	United Kingdom
US	United States
VC	Virtual Currency
WoM	Word-of-Mouth
WoS/K	Web of Science/Knowledge
XaaS	X-as-a-Service

Part I
Introduction

CHAPTER 1

Introduction to Smart Computing Approximate Reasoning

1.1 The Necessity of Computing in Practice: A Brief Reminder

The main aim of this chapter is to introduce smart computing to anyone who intends to apply the corresponding approaches to the interested practical problems. In view of this aim, we begin with simplifying why computations are generally required in practice. Then, we explain the uncertain aspects associated with various practical applications. This will bring us the main theme of this chapter, smart computing.

As quoted by Reed and Dongarra (2015), the English chemist Humphrey Davy once wrote, about two hundred years ago, "*Nothing tends so much to the advancement of knowledge as the application of a new instrument. The native intellectual powers of men in different times are not so much the causes of the different success of their labors, as the peculiar nature of the means and artificial resources in their possession*". Such observation is no less true nowadays that the competitive advantages can be obtained by someone who has the most powerful scientific tools at hand. In 2013, the Nobel Prize in chemistry was shared among three chemists for their remarkable achievements in computational modelling. Actually, computer models simulating real life have turned to be a crucial element for many advancements achieved in numerous domains today, ranging from describing high-energy particle accelerators' advantages (Hamada, 2017), mighty astronomical equipment (e.g., Hubble Space Telescope) (Chen, 2015a), DNA sequencing (Sung, 2017), to agent-based computational economics or artificial economics (LeBaron, 2000; Tesfatsion, 2003; Martinez-Jaramillo, 2007; Marwala, 2013d; Xing et al., 2011; Xing et al., 2012a; Xing et al., 2012b; Marwala, 2013c; Xing et al., 2014; Chen, 2008; Hamill and Gilbert, 2016), and agent-based manufacturing environments (Xing, 2016a; Xing and Gao, 2014gg; Xing and Gao, 2014kk; Xing and Gao, 2014qq; Xing et al., 2014; Xing et

al., 2011; Xing and Gao, 2014a). The scientific instruments are becoming increasingly powerful and continuously advancing human knowledge. Here, computing is largely needed inside each one of these instruments in order to facilitate functions such as sensor controlling, data processing, wireless communication, and many more.

In the scientific domain, researchers tend to have many types of tools; however, many of them are often constrained within a specific niche. On the contrary, computing is not just a science augmenter but rather something with inherent computational modelling and data analytics capabilities that can be potentially applied to all patches of science and engineering landscape (Ceruzzi, 2012; Gustafsson, 2011).

1.1.1 Practical Problems Generalization and Classification

In order to grasp the necessity of computations in practice, we need to recall what types of practical issues we would like to resolve. Broadly speaking, we can classify the majority of these problems into the following categories:

- **Learning:** We are curious about what is happening around us; in particular, we want to acquire different quantities' numerical values.
- **Predicting:** Upon having these values at hand, we are interested in forecasting the future status of our surroundings along the time axis.
- **Controlling:** By roughly knowing the subsequent states of our environment, we are eager to figure out what changes we can make, if possible, so that the desired future outputs can be obtained.

It should be noted that, more often than not, practical problems involve addressing all three types of tasks. In fact, according to Kreinovich (2008), this categorization can lead us to perceive the differences between science and engineering disciplines.

- **Science Discipline:** The tasks related to learning and predicting world states are usually treated as science.
- **Engineering Discipline:** The tasks of identifying a proper control strategy can be generally regarded as engineering.

A crowdfunding example for explaining such differences can be stated as follows:

1) The problems of measuring fundraising amount at different time intervals and predicting how a particular fundraising campaign will evolve over time belong to the science discipline.
2) The problems of finding the best means to control this development trajectory (e.g., adding new rewards, introducing new investment

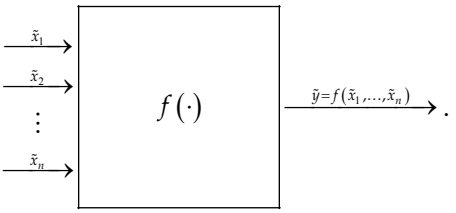
bundles, etc.) so that an asker/borrower's funding goal can be better achieved fall into the engineering discipline.

Without exaggeration, computations are required almost everywhere in dealing with both science and engineering problems.

1.1.1.1 Learning the Current State of the World

Suppose some state is characterized by different quantities (denoted by y), the essence of learning is thus to obtain these quantities' numerical values.

- **Measurable Situations:** Sometimes, these values can be easily acquired by directly measuring y. For instance, when we plan to know the current state of a crowdfunding campaign, we can measure its many factors, e.g., fundraising goal, funding raised so far, project start and close date, the number of backers, etc.
- **Unmeasurable Situations:** However, there are numerous quantities that are of our interest but are difficult to measure or are simply unmeasurable, e.g., crowdfunding backers' opinion, evaluation, attitude, emotion, and mood. Under this circumstance, a comprise can be done via the following steps (Kreinovich, 2008):
  1) *Step 1*: Get some relative easier-to-measure quantities (denoted by $x_1,...,x_n$) measured; and then
  2) *Step 2*: Get y estimated according to the measured values (represented by $\tilde{x}_i$) of these auxiliary quantities x_i.
- **How to Compute?** In order to accomplish the above Step 2, i.e., using the approximation of x_i to get an estimate of y, the relationship between y and $x_1,...,x_n$ has to be identified. Here, some algorithm, indicated by $f(x_1,...,x_n)$, is needed to fulfil such task of transforming the values of x_i into an estimate for y. This process can be expressed using Eq. 1.1 (Kreinovich, 2008):



1.1

The complexity of the algorithm $f(x_1,...,x_n)$ varies from situation to situation. In any case, a certain amount of computations are needed to perform learning tasks.

1.1.1.2 Predicting the Future State of the World

Since the current status of the environment is characterized by a set of quantities (denoted by $y_1,...,y_m$), as soon as the values of these quantities are obtained somehow, we can set out to predict their future values. In order to do so, we get to know how the future value z is determined by the current values $y_1,...,y_m$. Likewise, an algorithm, indicated by $g(y_1,...,y_m)$, is further needed to transform the acquired values of y_i into an estimate for z. This is, for example, how a socio-economic system (with no less than 100 million agents involved) is predicted: such simulations and predictions need a lot of computations, therefore, they have to be run on high performance computing equipment.

If we examine the learning and prediction tasks through the lens of computation, some similarities can be identified as follows (Kreinovich, 2008):

- **First**, both problems begin their process by estimating $\tilde{x}_1,...,\tilde{x}_n$ for the quantities of $x_1,...,x_n$; and
- **Second**, a specifically selected algorithm $f(\cdot)$ is always applied to these estimates. The resultant of this operation is an estimate $\tilde{y} = f(\tilde{x}_1,...,\tilde{x}_n)$ for the desired quantity y. In practice, this process shared by both problem classes is often referred to as data processing.

1.1.1.3 Controlling the Desirable State of the World

When the current status is knowable to some extent, and the subsequent future possibilities are partially predictable, we intuitively encounter the third task that is figuring out a means to get guaranteed desirable results. Under this category, two subclasses can be further grouped (Kreinovich, 2008):

- **Constraint Satisfaction:** Practical problems tend to have many constraints, and satisfying all those restrictions (maybe more than one possible design alternative) is often what we want. The goal of this sub-class is, thus, to find any one of these alternatives, and no preferences are imposed.
- **Function Optimization:** For this sub-class there exists an evident preference between different alternative designs (denoted by x). This is why an objective function $F(x)$ is often introduced in order to represent such preference: the larger the value of $F(x)$, the more preferable are design alternatives x. Therefore, optimization means that we would like to find the largest objective function value so that the most preferable design alternative x can be pinpointed.

Both sub-classes inevitably require a great amount of computations (Du and Ko, 2014).

1.1.2 Computational Science and Engineering

Broadly speaking, the main duty of scientists and engineers is to comprehend, develop, or optimize all sorts of 'systems'. Here, the term 'system' represents the object of interest, either a living system (e.g., stem cell), or an artificial technological system (e.g., crowdfunding). Suppose we did not have complex systems like computer processors (Harris and Harris, 2013; Comer, 2017), wind turbines (Hau, 2013), supply chain (Xing et al., 2010b; Xing et al., 2010a; Xing and Gao, 2015c; Xing and Gao, 2015a; Gao et al., 2013a), layout (Xing et al., 2010e), clustering (Xing et al., 2010d), load dispatch (Xing, 2015d), smartphones (Woyke, 2014; Xing and Marwala, 2018f), operating systems (Holcombe and Holcombe, 2012; Xing and Marwala, 2018g), remanufacturing system (Xing and Gao, 2014a; Xing et al., 2010a; Xing and Gao, 2015b), reconfigurable manufacturing system (Xing et al., 2006a; Xing et al., 2009; Xing et al., 2006b), design automation system (Xing and Marwala, 2018e), and robots (Xing and Marwala, 2018n; Xing and Marwala, 2018k; Xing and Marwala, 2018a; Marwala and Hurwitz, 2017; Xing, 2016e) in our life, then engineers/scientists would also not exist.

- **Model:** The underlying reason for the existence of both scientists and engineers is the complexity associated with natural systems and man-made systems. In general, to deal with such complexity, a common practice employed by engineers or scientists is simplification. This means, if something is complicated, it should be made simpler, but not too simple. To put it more formally, a viable simplified system description (i.e., model) is needed in order for engineers and scientists or anyone else to learn about complex systems.
- **Mathematical Model:** In the literature, there are many definitions of a mathematical model available. According to Velten (2009), a more general version can be given as follows: A mathematical model can be represented by a triplet (*System*, *Question*, *Statement*) where a system is denoted by *System*, a question related to the interested system is indicated by *Question*, and *Statement* stands for a set of mathematical statements given by Eq. 1.2 (Velten, 2009):

$$Statements = \{\sum_1, \sum_2, \ldots, \sum_n\}. \tag{1.2}$$

The process of this problem solving scheme can be seen in Fig. 1.1.

1) *Real-world*: Initially, we have a system (*System*) under consideration and a question (*Question*) related to this system.
2) *Mathematical universe*: This hemisphere consists of a set of mathematical statements (*Statement*) together with a possible problem solution (denoted by A^*).

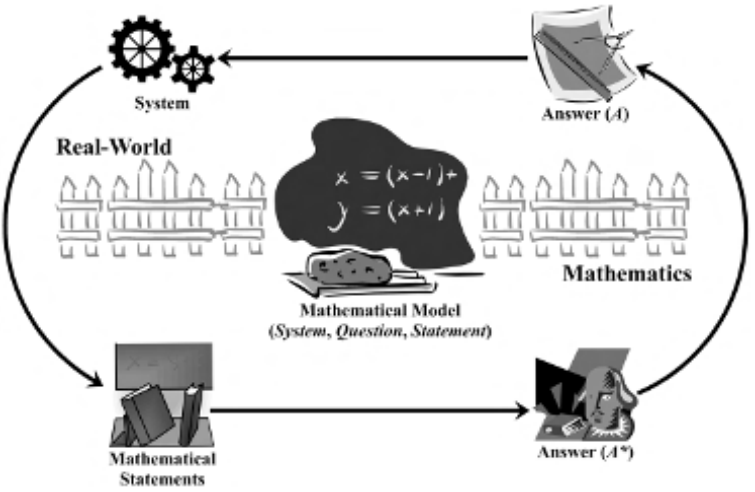


Figure 1.1: Problem solving process.

3) *Bridge*: These two hemispheres are bridged by a mathematical model (*System*, *Question*, *Statement*) which can translate the real-world problem into mathematical terms and interpret the obtained solutions in real-world language.

- **Computational Model:** In practice, there is a clear distinction between the following two tasks: (1) Formulating a mathematical model, and (2) Solving the resulting mathematical problem. The former can be done by non-mathematicians, while the latter is often tackled by someone with mathematical expertise (with the aid of advanced software in many cases). When the exact mathematical model of a problem is hard to obtain, an experimental way to find the appropriate solution is very time- and cost-consuming. In this regard, the complex mathematical models' subtleties can be illuminated by computational modelling (Shiflet and Shiflet, 2014).

1.1.3 Key Concepts

In the realm of modern computing, one often talks about computer science, mathematics, logic, and statistics (Paule, 2013). These foundational views can be given a clear technological meaning in the context of smart computing that has an aim of writing algorithms (i.e., mathematical and statistical notions as a base) and using computers for long calculations and verifications. For the rest of this section, we will describe various key concepts for bridging different research endeavours.

1.1.3.1 Sets

Generally speaking, a set refers to a collection of objects (or elements) that share the same properties or satisfy certain equations (Garnier and Taylor, 2002; O'Regan, 2013). In practice, it is a fundamental building block that gives a place to all needed domains in modern societies. For example, in the design of intelligent systems, objects (concrete or abstract) appear naturally (i.e., either definitely in or definitely out of the set) when considering constraints, uncertainties, and design specifications. Furthermore, due to fact that the object of a set need not be real, sets are the most appropriate language to specify several system performances (Veazie, 2017). For instance, we can define a set in asserting an imaginary domain (e.g., the domain of attraction), a conceptual domain (e.g., the domain of credit prediction), or even a specific domain (e.g., the error domain of a proposed algorithm). Accordingly, sets do not only serve as the terms of formulation, but also play a key role in constructing problem solutions (Blanchini and Miani, 2015).

In general, there are four types of sets (Pedrycz, 2013), namely crisp sets (e.g., yes/no, dichotomies), fuzzy sets (e.g., partial memberships), rough sets (e.g., lower and upper boundary), and shadowed sets (e.g., uncertainty regions). A comparative view of these sets is illustrated in Fig. 1.2.

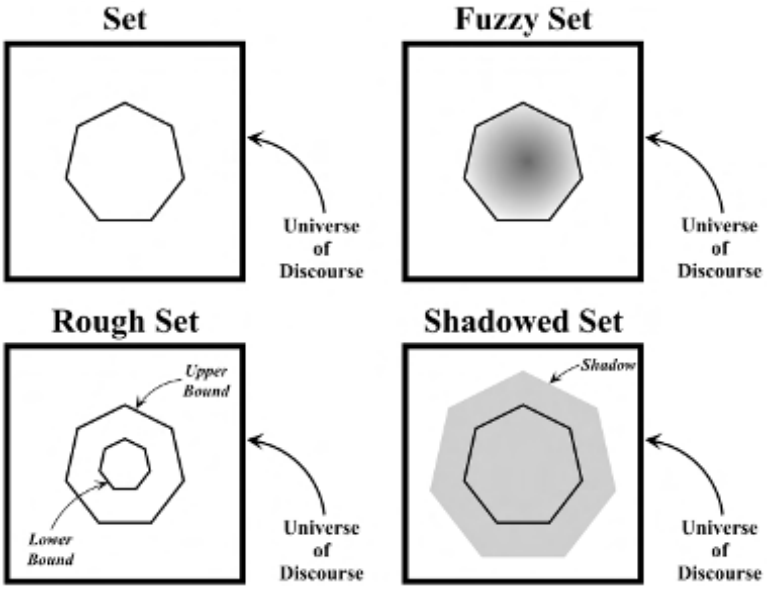


Figure 1.2: A comparative view of sets.

- **Classical (Crisp) Sets:** Classical, two-valued logic (e.g., qualified & unqualified, or true & false) is a basis of traditional mathematics and, in particular, of the crisp set theory. In other words, crisp set theory enables all the objects under consideration to be deterministically classified into two disjoint classes: belonging to a set, or not (Bělohlávek et al., 2017). Its membership function can be defined using Eq. 1.3 (Klir and Yuan, 1995):

$$X_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}. \tag{1.3}$$

  However, although every crisp set is defined by a "sharp" predicate, not every predicate is good enough and thus there is no a perfect classification of a crisp set to which it refers to (Pykacz, 2015; Trillas and Eciolaza, 2015). For example, in many real applications of data mining and image/natural language processing, datasets often contain a large number of features. In these cases, obtaining higher classification accuracy is neither possible nor necessary. Therefore, this practical need leads to the introduction of fuzzy set notions.

- **Fuzzy Sets:** The first publication in fuzzy set theory, written by Zadeh (1965), is over 50 years old. As its name implies, fuzzy set theory refers to a theory of graded concepts, in which an element x of given fuzzy set $\tilde{A}$ has various levels of membership, ranging from 0 (full non-membership) to 1 (full membership) (Zimmermann, 1992). Essentially, this idea intends to imitate the process of human brain solving complex problems by using mathematics (Zadeh, 1965). In general, fuzzy sets can be written in the form of Eq. 1.4 (Zimmermann, 1992):

$$\mu_{\tilde{A}} : x \mapsto \mu_{\tilde{A}}(x) \in [0,1]. \tag{1.4}$$

  where $\mu_{\tilde{A}}(x)$ is called membership function or grade of membership.

  It aims to provide a natural way to deal with problems in which imprecise or fuzzy predictions, relations, criteria and phenomena exist. Applications of this theory can be found in areas such as manufacturing (Azadegan et al., 2011; Xing et al., 2010c; Xing and Gao, 2014ii; Xing and Gao, 2014mm), decision policies (Xing, 2016c; Salles et al., 2016; Bosma et al., 2011), and data analysis (Petry and Zhao, 2009; Petkovic, 2014). More information can be found in (Zimmermann, 2001).

  In the literature, fuzzy sets also have three extended versions, namely interval-valued fuzzy sets (IVFSs) (Gozalczany, 1987; Turksen, 1996;

Wang and Li, 1998), intuitionistic fuzzy sets (IFSs) (Atanassov, 1986), and a hybrid of IVFSs and IFSs which is called interval-valued intuitionistic fuzzy sets (IVIFSs) (Atanassov and Gargov, 1989).

1) *Interval-valued fuzzy sets (IVFSs)*: According to Zhang et al. (2009), in many cases, an objective procedure is unfortunately not available in terms of identifying the crisp membership degrees for elements in a fuzzy set. The IVFSs emerged based on these observations. For instance, due to the inherent uncertainties associated with an expert's knowledge, describing the degree of belief (or termed as a membership function's values) in the form of a crisp number (e.g., ranging from 0 to 10) tends to be very hard. More often than not, only a very raw estimation (say between 4 and 6) is obtainable which in turn gives us the values of a membership function falling within an interval of [0.4, 0.6]. More formally, IVFSs belong to type-2 fuzzy sets, in which a fuzzy set A is over a referential set U, i.e., $A: U \to FS([0, 1])$. More details can be found in (Celik et al., 2015; Mendel, 2017).
2) *Intuitionistic fuzzy sets (IFSs)*: In IFSs, a membership function (denoted by μ) and a non-membership function (indicated by ν) are jointly introduced, in which $\mu + \nu \leq 1$. This formulation relaxes the originally enforced condition $\nu = 1 - \mu$ found in classical fuzzy set theory (Zhang et al., 2009). The basic definition of an IFS in a universe of discourse can be given by Eq. 1.5 (Atanassov, 1986):

$$A = \left\{ \left\langle x, \mu_A(x), \nu_A(x) \right\rangle \middle| x \in X \right\}. \tag{1.5}$$

where $X = \{x_1, x_2, x_3, \cdots, x_n\}$; and $\mu_A(x): X \to [0, 1]$ and $\nu_A(x): X \to [0, 1]$ refer to membership degree and non-membership degree, respectively, under the condition of $0 \leq \mu_A(x) + \nu_A(x) \leq 1$. Meanwhile, an intuitionistic index of x to A, termed as the hesitancy degree of the element $\pi_A(x)$, can be defined by Eq. 1.6 (Atanassov, 1986):

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x). \tag{1.6}$$

where $\pi_A(x) \in [0, 1]$, $\forall x \in X$. In practice, thanks to non-membership function, IFSs can be used to address the hesitancy that is often caused by information impression (Song et al., 2015). Further discussions in this regard can be found in (Li, 2014).
3) *Interval-valued intuitionistic fuzzy sets (IVIFSs)*: Finally, the IVIFSs act as a further generalization of IFSs in which unit intervals $[\mu_1, \mu_2]$ are employed for membership and non-membership values rather than exact numerical values (Atanassov and Gargov, 1989).

Nowadays, these variations have been widely used in various domains, such as social networking (Chen et al., in press), group decision making (Zhang and Xu, 2015; Chen, 2015b), and technology evaluation (Dereli and Altun, 2013).

- **Rough Sets:** Rough set theory, proposed by Pawlak (1982), serves as a novel mathematical tool for dealing with inconsistency problems (Zhang et al., 2016). A powerful principle underpinning the rough set theory is that hidden patterns in data cannot always be disclosed by precise measurements (Anderson et al., 2000). Accordingly, rough set theory is often regarded as a fundamental concept for artificial intelligence (AI) and cognitive sciences, such as image processing (Sen and Pal, 2009), machine learning (Henry, 2006), data mining (Bae et al., 2010; Fan and Zhong, 2012; Nelwamondo and Marwala, 2007), and knowledge discovery (Ali et al., 2015; Hassan and Tazaki, 2003; Marwala and Lagazio, 2011b).

  In a rough set, the set X is typically approximated via information extracted from B and then constructing the following terms, namely, lower approximation set, upper approximation set, and boundary region by using Eq. 1.7 (Pawlak, 2002):

$$\begin{aligned} \underline{B}X &= \left\{x \middle| [x]_B \subseteq X\right\} \\ \overline{B}X &= \left\{x \middle| [x]_B \cap X \neq \varnothing\right\}. \\ BN_B(x) &= \overline{B}X - \underline{B}X \end{aligned} \tag{1.7}$$

  where the lower and the upper approximations are denoted by $\underline{B}X$ and $\overline{B}X$, respectively, and $BN_B(x)$ represents the B – boundary region of rough set X.

  When a target set involves uncertainty or imprecision, one can use rough set to define such a set approximately via some definable sets (Pawlak and Skowron, 2007b). More specifically, any pair of precise sets can be divided into two parts: (1) The first part is called the lower approximating sub-set including all surely belonged objects, and (2) The second part is called the upper approximating sub-set containing all objects that possibly belong to the set. For those objects that cannot be classified into either upper- or lower sub-set, one can deposit them in the boundary region of a rough set (Pawlak, 2002). More in-depth discussions can be found in (Pawlak and Skowron, 2007a; Polkowski, 2002; Peters and Skowron, 2014; Peters and Skowron, 2016; Peters et al., 2014).

  When compared with fuzzy set theory, we can have the following observations: On the one hand, although extracting rules from data is made possible by rough set theory, a smooth extrapolation across

cases is still not allowed. On the other hand, fuzzy set outperforms rough set in terms of smooth extrapolation, but it needs a set of rules to get its process started (Anderson et al., 2000). Based on this, two new models of fuzzy-rough hybridization, i.e., rough fuzzy sets and fuzzy rough sets, are becoming increasingly popular. More theoretical backgrounds can be found in (Hinde and Yang, 2009; Cock et al., 2007; Dubois and Prade, 1990; Morsi and Yakout, 1998; Nanda and Majumdar, 1992; Radzikowska and Kerre, 2002; Nguyen et al., 2014; Lu et al., 2016; Yang and Hinde, 2010) and the corresponding applications can be found in decision making (Anderson et al., 2000), feature selection (Kuncheva, 1992), data mining (Nilavu and Sivakumar, 2015; Srinivasan et al., 2001), to name just a few.

- **Shadowed Sets:** According to Pedrycz (1998), the concept of shadowed sets was proposed to close the gap between fuzzy sets and rough sets. In general, a shadowed set (denoted by S) in a universe of discourse (represented by U) stands for a set-valued mapping $S : U \rightarrow \{0,[0, 1],1\}$ and thus has following properties given by Eq. 1.8 (Pedrycz, 1998):

$$\begin{aligned} Core(S) &= \{x \in U : S(x) = 1\} \\ Sh(S) &= \{x \in U : S(x) = [0,1]\} \\ Supp(S) &= cl\{x \in U : S(x) \neq 0\} \end{aligned} \quad . \tag{1.8}$$

  where a shadowed set's core is denoted by $Core(S)$, in which all objects are fully definable; $Sh(S)$ represents the shadow which incorporates uncertainty and imprecision, and a shadowed set's support is indicated by $Supp(S)$, in which all elements are incompatible with the rules defined by S.

  In general, one can view shadowed sets as an expression of fuzzy sets in a three-way approximation, i.e., {0,1,[0,1]} (Yao et al., 2017), though conceptually, shadowed sets and rough sets are more close to each other, even though their theoretical foundations are indeed very different (Zhou et al., 2011). More specifically, the rough set theory's key concepts (i.e., negative region, lower bound, and boundary region) are associated with shadowed sets' three-logical values (i.e., 0, 1, and [0,1] which correspond to excluded, included, and uncertain characteristics, respectively) (Pedrycz, 2009; Zhou et al., 2011). Further discussions on this matter can be found in (Pedrycz, 1998; Grzegorzewski, 2013; Pedrycz, 2005).

1.1.3.2 Logic

Logic is about reasoning, going from the knowledge assimilation (i.e., premises) to making deductions based on this knowledge (i.e., a

conclusion) (Gensler, 2010). In computer science, most theories operate in accordance with a logic system, typically Boolean logic. Typically, a Boolean logic is a set, which consists of at least two distinct special elements 0 and 1, respectively. In its simplest form, any outcome is governed and calculated by following two main laws: (1) The law of contradiction, the possibility that p and $-p$ can co-exist at the same time is zero, hence one side of a contradiction has to be invalid (or false); and (2) The law of the excluded third, nothing can be found between to be and not to be (Moller and Struth, 2013).

Fuzzy logic provides another foundational view for reasoning based on uncertain statements. Generally speaking, this concept was inspired by two notable human abilities: (1) The ability of reasoning and decision-making under the situations like imprecision, information incompleteness, uncertainty, and partiality of truth; (2) The ability to accomplish various perception-based physical/mental tasks with no accurate measurements and computations involved (Pedrycz et al., 2008). In essence, fuzzy logic acts as a novel theory of inference by introducing linguistic IF-THEN rules (Zadeh, 1975a; Zadeh, 1975b; Zadeh, 1975c). The main application domains of fuzzy logic include decision making (Lin and Chen, 2004; Patel and Marwala, 2006), control (Raber, 1994), healthcare (Barro and Marín, 2002; Massad et al., 2008), and image processing (Caponetti and Castellano, 2017).

1.1.3.3 Probability Theory

Probability can be loosely defined as "the frequency of occurrence" of an outcome. Typically, the probability is between 0 (i.e., the event cannot occur) and 1 (i.e., the event is guaranteed to occur). In other words, probability theory is about the study of chance. To put probability on firm mathematical ground, we need to first introduce the concept of randomness, which is the central issue in this domain.

- **Randomness:** In general, one can view randomness as a kind of objective uncertainty associated with random variables (Ross, 2014; Schinazi, 2012). Broadly speaking, randomness can be categorized into two classes, namely classical and chaotic randomness (Plotnitsky, 2016; Clegg, 2013).

  1) *Classical randomness*: A classical randomness can be defined as a meaningfully compressed category (i.e., obeying rules) which includes the overall information of a collection of random objects (Clegg, 2013). In other words, the classical randomness depends not on an individual event that carries discernible information but instead on a probability distribution (e.g., cumulative distribution functions) in which we can designate a special pattern or objective

from the whole population (Denker and Woyczyriski, 1998). For example, to measure the randomness in a gambling game, we must not care too much about any particular individual event, but should focus on how often an event comes up (i.e., how different outcomes are distributed). Other examples containing classical randomness also include lottery and stock market.

2) *Chaotic randomness*: On the other hand, when the effect of randomness arises in dynamic systems with only deterministic and well-controlled ingredients, we call this phenomenon chaotic randomness (Denker and Woyczyriski, 1998). Examples in this category include the ball's trajectory on the pinball table, the Brazilian butterfly effect, and the iterations of quadratic maps.

- **Probability Models:** Based on our understanding of randomness, we can now turn to a formal mathematical setting for analysing probability models. In general, any probabilistic model must include the following components: (1) A sample space (denoted by Ω) for an experiment which refers to the set of all possible outcomes of a random process; (2) An event (indicated by B) which represents a subset of the sample space; and (3) A probability function, represented by $\Pr(B)$, which satisfies three properties defined by Eq. 1.9 (Johnson, 2018):

$$\begin{array}{ll} \text{Requirement 1.} & 0 \le \Pr(B) \le 1, \text{ for each event } B \text{ in } \Omega. \\ \text{Requirement 2.} & \Pr(\Omega) = 1 \\ \text{Requirement 3.} & \text{If } A \text{ and } B \text{ are mutually exclusive events in } \Omega, \text{ then} \\ & \Pr(A \cup B) = \Pr(A) + \Pr(B) \end{array} \quad . \quad 1.9$$

- **Probability Rules:** Bayes' theorem is one of probability theory that defines a relation between certain conditional probabilities (Morin, 2016).

1) *Simple form*: Typically, the 'simple form' of Bayes' theorem is defined by Eq. 1.10 (Morin, 2016):

$$\Pr(A|Z) = \frac{\Pr(Z|A) \cdot \Pr(A)}{\Pr(Z)}. \quad 1.10$$

2) *Explicit form*: While the 'explicit form' can be given by Eq. 1.11 (Morin, 2016):

$$\Pr(A|Z) = \frac{\Pr(Z|A) \cdot \Pr(A)}{\Pr(Z|A) \cdot \Pr(A) + \Pr(Z|\text{not } A) \cdot \Pr(\text{not } A)}. \quad 1.11$$

where the first part of denominator, i.e., $\Pr(Z|A)\cdot\Pr(A)$, refers to the combination of the accuracy of the test and the historical data of the problem; while the second part of denominator, i.e., $\Pr(Z|\text{not } A)\cdot\Pr(\text{not } A)$, is still related to the accuracy of the test, denoted by $\Pr(Z|\text{not } A)$, together with new information, represented by $\Pr(\text{not } A)$.

3) *General form*: And the 'general form' of Bayes theorem is given by Eq. 1.12 (Morin, 2016):

$$\Pr\left(A_k|Z\right)=\frac{\Pr\left(Z|A_k\right)\cdot\Pr\left(A_k\right)}{\sum_i \Pr\left(Z|A_i\right)\cdot\Pr\left(A_i\right)}. \tag{1.12}$$

where A_i stands for a complete and mutually exclusive set of events; $\Pr(A_i)$ denotes the prior probabilities; $\Pr(Z|A_i)$ represents the conditional probability, and $\Pr(A_k|Z)$ indicates the posterior probability.

In practice, Bayes' rule finds application in a wide variety of domains, such as pattern recognition (Marwala, 2007a), militarized interstate dispute (Marwala and Lagazio, 2011a), and finite element model updating (Marwala et al., 2017).

1.1.3.4 Possibility Theory

Through the above discussion, we learn that probability theory provides an acceptable concept of quantitative chance. While in the literature, there is another method called possibility theory which focuses more on the analysis of different types of uncertainty other than chance (Nguyen and Walker, 2006). More specifically, possibility theory is used to deal with problems that have a non-probabilistic character. One example in this regard is to combine possibility theory with linguistic variable concept (found in fuzzy set theory) for the purpose of offering a unified formal framework. Under such framework, a formal management of information with inherent imprecision, vagueness, ambiguity, and uncertainty (Kraft and Colvin, 2017). A crude comparison among possibility theory, probability theory and Boolean algebra was performed by Zimmermann (1992) from different aspects and the results are illustrated in Fig. 1.3.

In practice, applied researchers have learned possibility theory through the lens of measure theory, since it can provide a relatively easy translation between mathematical theory and real-world problems. From a historical perspective, possibility theory is based on two semi-continuous generalized measure theories, i.e., possibility measures and necessity measures (Bělohlávek et al., 2017). The former estimates the feasibility

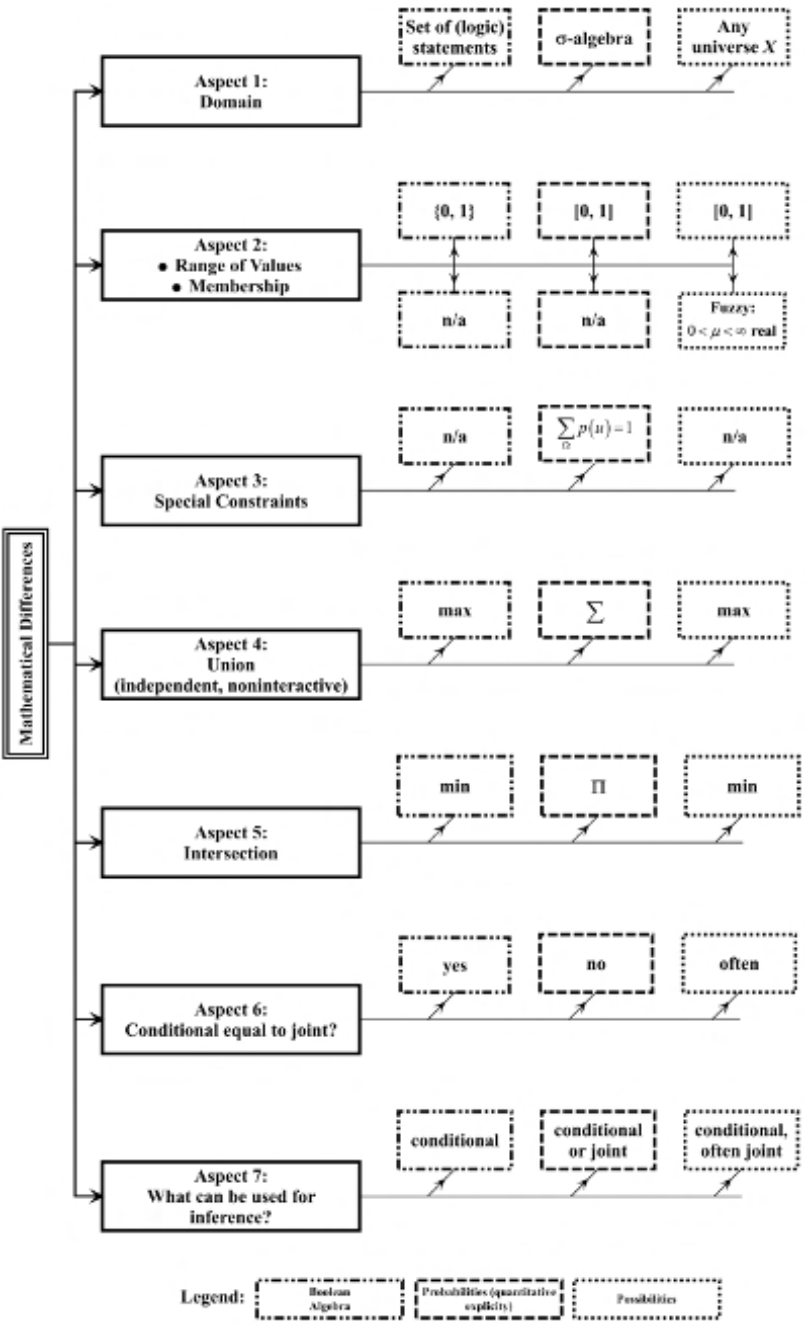


Figure 1.3: Mathematical differences among three areas: Boolean algebra, probabilities, and possibilities.

degrees of alternative options, while the latter indicates priorities (Dubois et al., 1996).

Suppose we have a universe of discourse (denoted by U), then a possibility measure (represented by Pos) is a set function Pos: $P(X) \rightarrow [0,1]$ that meets the properties given by Eq. 1.13 (Zimmermann, 1992; Bělohlávek et al., 2017):

$$\begin{array}{ll} \text{Property 1.} & \text{Pos}(\varnothing) = 0,\ \text{Pos}(U) = 1 \\ \text{Property 2.} & A \subseteq B \Rightarrow \text{Pos}(A) \le \text{Pos}(B) \\ \text{Property 3.} & \text{Pos}\left(\bigcup_{i \in I} A_i\right) = \sup_{i \in I} \text{Pos}(A_i) \end{array} . \tag{1.13}$$

Given a Pos, the necessity measure (denoted by Nec) can be defined by Eq. 1.14 (Zimmermann, 1992; Bělohlávek et al., 2017):

$$\text{Nec}(A) = 1 - \text{Pos}\left(\overline{A}\right). \tag{1.14}$$

If $\tilde{A}$ is a fuzzy set in the universe U and π_x denotes a possibility distribution (associated with a variable X that takes value from set U), then the possibility measure can be defined by Eq. 1.15 (Zimmermann, 1992; Bělohlávek et al., 2017):

$$\begin{aligned} \text{Pos}\left\{X \text{ is } \tilde{A}\right\} &\triangleq \pi\left(\tilde{A}\right) \\ &\triangleq \sup_{u \in U} \min\left\{\mu_{\tilde{A}}(u), \pi_x(u)\right\} \end{aligned} . \tag{1.15}$$

In a similar vein, the necessity measure can also be given by Eq. 1.16 (Bělohlávek et al., 2017):

$$\text{Nec}_F(A) = 1 - \text{Pos}_F\left(\overline{A}\right). \tag{1.16}$$

where F represents a fuzzy set defined on U.

1.1.3.5 Interval Analysis

In general, an interval can be denoted as real numbers with brackets or parentheses, based on whether the end points are included or not. For instance, if $a > b$, then $[b, a]$, $[b, a)$, $(b, a]$, and (b, a) are sets of numbers x that satisfy the conditions given by Eq. 1.17 (Hausdorff, 1962):

$$\begin{array}{l} b \le x \le a \\ b \le x < a \\ b < x \le a \\ b < x < a \end{array} . \tag{1.17}$$

The purpose of interval analysis is to address numerical errors emerging from computation (Moore et al., 2009). In other words, this technique is designed to automatically provide rigorous bounds on all potential errors and uncertainties (Hansen and Walster, 2004). Interested readers should refer to Chakraverty (2014) for more information.

A fuzzy interval typically represents a fuzzy set of real numbers in which the membership function is characterized by unimodal and upper-semi continuous features (Kacprzyk and Pedrycz, 2015). Accordingly, the calculus of fuzzy intervals is an extended version of interval arithmetic built on a possibilistic counterpart of a random variable's computation. For instance, in order to obtain the addition of two fuzzy intervals (denoted by A and B, respectively), one must calculate the membership function of $A \oplus B$ as the possibility degree via the possibility distribution, i.e., $\min(\mu_A(x), \mu_B(y))$, as given by Eq. 1.18 (Kacprzyk and Pedrycz, 2015):

$$\mu_{A\oplus B}(z) = \Pi\left(\left\{(x,y) : x + y = z\right\}\right). \tag{1.18}$$

Further discussions about the fuzzy interval can be found in (Dubois et al., 2000; Nguyen et al., 2012).

1.1.3.6 Category Theory

Briefly, a category stands for a labelled directed graph, in which the nodes are termed as objects and the labelled directed edges are called morphisms (Barendregt, 2013). From a conceptual point of view, category theory is a mathematical structure that can be used to formalize high-level concepts (e.g., sets, rings and groups). More in-depth explanations can be found in (Roman, 2017; Awodey, 2006).

1.2 The Unavoidability of Uncertainty in Reality: A Quick Retrospection

The kind of information that we get every day is likely to be as follows: Join us for dinner and the meeting at 7.00 tomorrow. Of course, the uncertainty of such information might lead us into trouble: (1) The first part of the statement is vague with respect to what time the common dinner will be held; and (2) The second part of the statement is ambiguous regarding whether the meeting time is in A.M. or P.M.

In the literature, uncertainty is defined as a measure of the users' understanding of the difference between the information carried by the proposition corresponding to certain phenomena (Galbraith, 1973). Some scholars pointed out that uncertainty is unavoidable and, thus, worth the scrutiny. For example, Faber (2012) examined engineering decision

problems that are subject to uncertainty. Kraft and Colvin (2017) treated information retrieval as an uncertain problem and used fuzzy logic to deal with it. Other examples also include learning with uncertainty (Wang and Zhai, 2017), artificial intelligence with uncertainty (Li and Du, 2017), and economic models under uncertainty (Aliyev, 2014).

Traditionally, measures of uncertainty were only related to classical set theory and probability theory (Wierman, 1999; Friedlob and Schleifer, 1999). However, this unique connection is now challenged by many. Among them, several researchers proposed that uncertainty is a multi-dimensional concept (Zadeh, 1965; Pawlak, 1982; Zimmermann, 2001). As a result, a unified perspective on the recent studies regarding uncertainty calls for other theories which can discover different properties of those incomplete and imprecise data, such as fuzziness, roughness, and indefiniteness (Wang and Zhai, 2017; Friedlob and Schleifer, 1999). For the rest of this section, we will briefly discuss some of these properties together with the corresponding mathematical foundations.

1.2.1 Fuzziness

Fuzziness is a kind of mathematical way to represent cognitive uncertainty, such as ambiguity and vagueness, in measurements and natural language expressions, imperfectly in experts' thoughts and knowledge, and absence of concepts' boundaries (Colubi and Gonzalez-Rodriguez, 2015; Coppi et al., 2006; Seising, 2008; Zhang, 1998; Klir, 1987). In practice, those imprecise informations are modelled by fuzzy sets based on their associated degrees of membership (Zadeh, 1965; Nguyen and Walker, 2006). Some useful definitions are discussed as follows:

- **Fuzzy Sets:** Suppose we have a non-empty set denoted by $X = \{x_1, x_2, x_3, \cdots, x_n\}$. A fuzzy set (represented by $\tilde{A}$) in X stands for a set of ordered pairs given by Eq. 1.19 (Zimmermann, 1992):

$$\tilde{A} = \left\{ \left(x, \mu_{\tilde{A}}(x) \right) \middle| x \in X \right\}. \tag{1.19}$$

  where the membership degree of x in $\tilde{A}$ is denoted by $\mu_{\tilde{A}}(x)$, and each element x in X maps to a real number belonging to the interval of [0,1].

  1) A support of a fuzzy set $\tilde{A}$, denoted by $S(\tilde{A})$, stands for the crisp set of all $x \in X$ that satisfy $\mu_{\tilde{A}} > 0$.
  2) In classical (crisp) set theory, α-cut means that the degree for elements that belong to a fuzzy set is at least α as given by Eq. 1.20 (Zimmermann, 1992):

$$A_\alpha = \left\{ x \in X \middle| \mu_{\tilde{A}}(x) \geq \alpha \right\}. \tag{1.20}$$

3) A fuzzy set is convex, if for all elements $x_1, x_2 \in X$ and $\lambda \in [0,1]$, the relationship defined by Eq. 1.21 (Zimmermann, 1992) hold:

$$\mu_{\tilde{A}}\left(\lambda x_1 + (1-\lambda)x_2\right) \geq \min\left(\mu_{\tilde{A}}(x_1), \mu_{\tilde{A}}(x_2)\right). \tag{1.21}$$

4) For fuzzy sets, the basic set-theoretic operations are defined by Eq. 1.22 (Zimmermann, 1992):

$$\begin{cases} \text{Fuzzy Union}: & \mu_{\tilde{A}\cup\tilde{B}}(x) = \max\left\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\right\},\ x \in X \\ \text{Fuzzy Intersection}: & \mu_{\tilde{A}\cap\tilde{B}}(x) = \min\left\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\right\},\ x \in X \\ \text{Fuzzy Complement}: & \mu_{\notin\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x),\ x \in X \end{cases}. \tag{1.22}$$

- **Fuzzy Numbers**: A fuzzy number is a fuzzy quantity that stands for a generalization of a real number (Nguyen and Walker, 2006). Typically, there are two types of fuzzy numbers: (1) Triangular fuzzy numbers, and (2) Trapezoidal fuzzy numbers. In practice, triangular fuzzy numbers are the most employed type, in which the fuzzy numbers are characterized by a triangular shape. The general formulation of fuzzy numbers can be given by Eq. 1.23 (Novák et al., 2016):

$$\mu_{\tilde{A}}(x) = \begin{cases} 0 & \text{if } x < a \text{ or } x > c \\ \dfrac{x-a}{b-a} & \text{if } a \leq x \leq b \\ \dfrac{x-c}{b-c} & \text{if } b < x \leq c \\ 1 & \text{if } x = b \end{cases}. \tag{1.23}$$

where $a < b < c$.

1.2.2 Roughness

In rough set theory, roughness stands for the uncertainty associated with a target concept that results from its boundary region (Pawlak, 1991). In other words, the uncertainty of rough sets is expressed by means of approximations. Some useful concepts are reviewed as follows:

- **Information System Framework:** The starting point of any rough set is a dataset that is called an information table or information system. Suppose $S = \langle U, A, V, f \rangle$ is introduced to represent an information system, we can have the following (Zhang et al., 2016; Pawlak, 1982):

1) U and A are finite non-empty sets;
2) The former stands for the universe of objects;
3) While the latter indicates the attribute set;
4) $V = \bigcup_{a \in A} V_a$, where V_a denotes the set of values of attribute a; and
5) $f: A \to V$ denotes a description function.

- **Indiscernible Relation:** Given $\forall B \subseteq A$, there is an associated indiscernibility (or equivalence) relation based on U as defined by Eq. 1.24 (Zhang et al., 2016; Pawlak, 1982):

$$Ind(B) = \left\{ (x, y) \in U^2, \forall_{a \in B} \left(a(x) = a(y) \right) \right\}. \tag{1.24}$$

Accordingly, the equivalence class of an object ($x \in U$) is denoted by $[x]_{Ind(B)}$ or simply $[x]$, and the pair $(U, [x]_{Ind(B)})$ is termed as the approximation space.

- **Lower- and Upper-Approximation Sets:** For a sub-set $X \subseteq U$, its lower- and upper-approximation sets are given by Eqs. 1.25 and 1.26 (Zhang et al., 2016; Pawlak, 1982):

$$Appr_{lower}(X) = \left\{ x \in U \middle| [x] \subseteq X \right\}. \tag{1.25}$$

$$Appr_{upper}(X) = \left\{ x \in U \middle| [x] \cap X \neq \varnothing \right\}. \tag{1.26}$$

where $Appr_{lower}(X)$ stands for an object ($x \in U$) certainly belonging to $X \subseteq U$, while $Appr_{upper}(X)$ denotes an object $x \in U$ possibly belonging to $X \subseteq U$. Furthermore, the set $BND(X) = Appr_{upper}(X) - Appr_{lower}(X)$ represents the boundary region of X.

1) *Rough sets*: If and only if $Appr_{upper}(X) \neq Appr_{lower}(X)$, X can be regarded as a rough set (Zhang et al., 2016; Pawlak, 1982).
2) *Roughness of rough sets*: The roughness of set X is defined by Eq. 1.27 (Zhang et al., 2016; Pawlak, 1982):

$$Roughness(X) = 1 - \frac{\left| Appr_{lower}(X) \right|}{\left| Appr_{upper}(X) \right|} = \frac{\left| Appr_{upper}(X) - Appr_{lower}(X) \right|}{\left| Appr_{upper}(X) \right|}. \tag{1.27}$$

1.2.3 Indefiniteness

Indefiniteness is a specific concept for the uncertainty of meaning between two or more unclear objects, situations, and/or problems (Wang and Zhai, 2017). Typically, it is associated with a possibility distribution that was

proposed by Zadeh (1978). In general, such possibility distribution π can be defined by Eq. 1.28 (Wang and Zhai, 2017; Zimmermann, 1992):

$$\pi(A) = \sup_{x \in A} f(x). \tag{1.28}$$

where $A \subset X$, and X represents a classical set, i.e., $X = \{x_1, x_2, x_3, \cdots, x_n\}$. If and only if $\max_{x \in X} \pi(x) = 1$, the possibility distribution π can be treated as a normalized possibility distribution.

Based on these definitions, the measure of indefiniteness can be given by Eq. 1.29 (Wang and Zhai, 2017):

$$g(\pi) = \sum_{i=1}^{n} \left(\pi_i^* - \pi_{i+1}^*\right) \ln i \,. \tag{1.29}$$

where $\pi = \{\pi(x) | x \in X\}$ denotes a normalized possibility distribution, π^* represents the possibility distribution's permutation that satisfies $\pi_i^* \geq \pi_{i+1}^*$, for $i = 1,2,\cdots,n$. Since possibility theory focuses more on imprecision and vagueness of linguistic meanings, in order to elicit knowledge from those words or sentences, fuzzy judgements rather than probabilistic values are required.

Suppose we have a fuzzy set ($\tilde{F}$) of universe U and the associated membership function, denoted by $\mu_{\tilde{F}}(u)$, the assignment of the values of variable u to X can be given by Eq. 1.30 (Zimmermann, 1992):

$$X = u : \mu_{\tilde{F}}(u). \tag{1.30}$$

Accordingly, the fuzzy membership function of possibility distribution $\tilde{\pi}$ can be given by Eq. 1.31 (Zimmermann, 1992):

$$\tilde{\pi}_x = \mu_{\tilde{F}}. \tag{1.31}$$

Given $i = 2$, the measure of fuzzy indefiniteness is obtainable by using Eq. 1.32 (Wang and Zhai, 2017):

$$g(\tilde{\pi}) = \begin{cases} \left(\dfrac{\mu_1}{\mu_2}\right) \times \ln 2 & \text{if } 0 \leq \mu_1 < \mu_2 \\ \ln 2 & \text{if } \mu_1 = \mu_2 \\ \left(\dfrac{\mu_2}{\mu_1}\right) \times \ln 2 & \text{others} \end{cases}. \tag{1.32}$$

1.2.4 Relations Underlying the Uncertainties

Since the types of uncertainties vary a lot, different methodologies offer us varied alternatives, listed below.

- **Option 1:** Based on probability theory, randomness measures the uncertainty that the target objects have definite boundaries;
- **Option 2:** Built on fuzzy set theory, fuzziness measures the uncertainty emerging from vagueness;
- **Option 3:** According to rough set concept, roughness measures the uncertainty with respect to imprecise and incomplete information; and
- **Option 4:** With the aid of possibility theory, indefiniteness measures the uncertainty associated with non-specificity of information.

In addition, by combining these fundamental methods with other technologies (Zhang et al., 2016), a large number of new toolkits have become available, such as probabilistic fuzzy sets (Jiang et al., 2017; Kentel and Aral, 2004; Fialho et al., 2016), fuzzy/rough sets with neural networks (Boutalis et al., 2014; Raveendranathan, 2014; Ding et al., 2014), neighbourhood rough sets (Pal et al., 2012), rough/fuzzy sets with clustering algorithms (Zhou et al., 2011; Baraldi and Blonda, 1999a; Baraldi and Blonda, 1999b), fuzzy sets with game theory (Li, 2014; Jiménez-Losada, 2017), rough/fuzzy sets with soft sets (Sun and Ma, 2014; Feng et al., 2010; Das et al., 2017), and fuzzy/rough sets with different smart computing techniques (Xing and Gao, 2014b; Kolokotsa, 2007; Mardani et al., 2015; Mitra and Hayashi, 2000; Kubler et al., 2016; Castillo and Melin, 2015).

1.2.5 Measure Theory

In the domain of uncertainty, the introduction of suitable measures, for comparing different information contents carried by distinct uncertainties (e.g., randomness, fuzziness, roughness, and indefiniteness), is ranked as the most attractive topic. Indeed, measure theory stays at the center of addressing uncertainty. In the literature, various types of measures (bearing distinct properties) have been proposed, such as distance, correlation, divergence, entropy and similarity. Amongst them, entropy and similarity are the most frequently used methods for studying the uncertainty/certainty information carried by fuzzy sets. According to Liu (1992), these two measures represent the complementary information towards each other, i.e., certainty (similarity) and uncertainty with respect to the corresponding crisp set.

1.2.5.1 Entropy Measurement for Uncertainty—Shannon's Entropy for Classical System

Typically, the measure used for quantifying the uncertainty associated with random variables is called entropy (Wang and Zhai, 2017; Cover

and Thomas, 1991; Mitzenmacher and Upfal, 2017). Although this term originated from thermodynamics and was initially used to measure the disorder in a system, it is now widely used in information and communication systems to measure the uncertainty regarding the information content of the system (Zadeh, 1965; De Luca and Termini, 1972), i.e., determining the variation degree of the probability distribution.

Basically, Shannon's entropy emphasizes the measurement of the average uncertainty in bits corresponding to the prediction of a random experiment's outcomes, that is, entropy allows us to learn the distribution function's shape. From the recipient viewpoint, the amount of missed information is known.

- **Discrete Entropy:** In general, entropy relies on a probabilistic description of an event. Suppose we have a discrete random variable (denoted by X) which consists of n instances, represented by $\{X_i$, for $i = 1,2,...,n\}$, the probability mass function of X can thus be given by Eq. 1.33 (Wang and Zhai, 2017):

$$p(X_i) = \Pr(X = X_i). \tag{1.33}$$

Based on this formulation, the entropy in bits of a discrete random variable X can then be defined by Eq. 1.34 (Cover and Thomas, 1991; Li and Du, 2017; Wang and Zhai, 2017):

$$H(X) = -\sum_{i=1}^{n} p(X_i)\log_2 p(X_i). \tag{1.34}$$

If we have two random systems (denoted by X and Y), then the joint entropy of these two systems can be defined by Eq. 1.35 (Wang and Zhai, 2017):

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{i=1}^{n} p(X_i,Y_i)\log_2 p(X_i,Y_i). \tag{1.35}$$

Similarly, we can also compute the conditional entropy $H(X|Y)$ by using Eq. 1.36 (Wang and Zhai, 2017):

$$\begin{aligned} H(X|Y) &= -\sum_{i=1}^{n} p(X_i) H(Y|X = X_i) \\ &= -\sum_{i=1}^{n} p(X_i)\sum_{i=1}^{n} p(Y_i|X_i)\log_2 p(Y_i|X_i). \\ &= -\sum_{i=1}^{n}\sum_{i=1}^{n} p(X_i,Y_i)\log_2 p(Y_i|X_i) \end{aligned} \tag{1.36}$$

where $H(X|Y)$ is, in general, not equal to $H(Y|X)$; while $H(X) - H(X|Y)$ is always equal to $H(Y) - H(Y|X)$.

Since the mutual information can quantify the closeness degree of two random variables, suppose X and Y are discrete random variables, when it comes to measure the relevance of X and Y, we can further define the mutual information by using Eqs. 1.37–1.39 (Wang and Zhai, 2017; Michalowicz et al., 2014):

$$I(X;Y) = H(X) - H(X|Y). \tag{1.37}$$

$$I(X;Y) = H(Y) - H(Y|X). \tag{1.38}$$

$$I(X;Y) = H(X) + H(Y) - H(XY). \tag{1.39}$$

- **Differential Entropy:** The concept of entropy for continuous distribution is called differential entropy. Suppose X represents a continuous random variable with probability density function, denoted by $p_X(x)$, the differential entropy can be defined by Eq. 1.40 (Michalowicz et al., 2014):

$$h_X = -\int_S p_X(x)\log_2\left(p_X(x)\right)dx\,. \tag{1.40}$$

where $S = \{x|p_x(x) > 0\}$ denotes the support set of X. While the value of discrete entropy is always non-negative, differential entropy may take any value between ∞ and $-\infty$.

For continuous random variables X and Y, their joint differential entropy, conditional differential entropy, and the mutual information can be further defined by Eqs. 1.41–1.43, respectively (Michalowicz et al., 2014):

$$h_{XY} = \int\int p_{XY}(x,y)\log_2\left(\frac{1}{p_{XY}(x,y)}\right)dxdy\,. \tag{1.41}$$

$$h_{Y|X} = \int\int p_{XY}(x,y)\log_2\left(\frac{1}{p_Y(y|x)}\right)dxdy\,. \tag{1.42}$$

$$I(X;Y) = \int\int p_{XY}(x,y)\log_2\left(\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}\right)dxdy\,. \tag{1.43}$$

- **Maximum Entropy Estimation Method:** Assume that the density function $p_X(x)$ is unknown, but we know a number of related constraints, such as mean and variance. The maximum entropy estimate of the unknown probability density function is the one that can maximize the entropy, subject to given constraints (Theodoridis and Koutroumbas, 2009).

1.2.5.2 Entropy Measurement for Uncertainty—Fuzzy Entropy

A fuzzy set's entropy is to calculate the degree of fuzziness on such fuzzy set (Zadeh, 1968). Among different measures, one measure considered by Zadeh (1968) can be expressed by using Eq. 1.44 (Zimmermann, 1992; Wang and Zhai, 2017):

$$H\left(\tilde{A}\right) = -\sum_{i=1}^{n} \mu_{\tilde{A}}\left(x_i\right) p_i \log_2 p_i \,. \tag{1.44}$$

In addition, another fuzzy entropy considered by De Luca and Termini (1972) can be defined by using Eq. 1.45 (Zimmermann, 1992; Wang and Zhai, 2017; De Luca and Termini, 1972):

$$H\left(\tilde{A}\right) = -K\sum_{i=1}^{n}\left[\mu_{\tilde{A}}\left(x_i\right)\log_2 \mu_{\tilde{A}}\left(x_i\right) + \left(1-\mu_{\tilde{A}}\left(x_i\right)\right)\log_2\left(1-\mu_{\tilde{A}}\left(x_i\right)\right)\right]. \tag{1.45}$$

where K represents a positive constant.

Fuzzy entropy is quite different from the classical Shannon's entropy (Jumarie, 1992). The former handles vagueness and ambiguity uncertainties, while the latter deals with randomness uncertainty (i.e., probabilistic). Typically, there are three types of fuzzy entropy, i.e., interval-valued fuzzy entropy, intuitionistic fuzzy entropy, and interval-valued intuitionistic fuzzy entropy (Bustince and Burillo, 1996; Wei et al., 2011; Zhang et al., 2014). Nowadays, those entropy measures have been utilized in dealing with fuzzy systems, such as image processing (Naidu et al., in press), fuzzy decision-making systems (Shi and Yuan, 2015), and fuzzy software testing (Kumar et al., 2012).

1.2.5.3 Entropy Measurement for Uncertainty—Rough Entropy

In order to get the a set's knowledge incompleteness quantified, rough entropy was introduced by Beaubouef et al. (1998). In general, it can be formulated according to the uncertainty in granulation and the set's definability (Sen and Pal, 2009). For an information system, denoted by $S = (U, A)$ where $X \subseteq U$, rough entropy of X can be described by using Eq. 1.46 (Liang, 2011):

$$E_A(X) = -\rho_A(X)\left(\sum_{i=1}^{m}\frac{|R_i|}{|U|}\log_2\frac{|R_i|}{|U|}\right). \quad 1.46$$

where $p_A(X)$ represents the rough degree of X. In a similar vein, rough entropy of A can be given by Eq. 1.47 (Liang, 2011).

$$E_r(A) = -\sum_{i=1}^{m}\frac{|R_i|}{|U|}\log_2\frac{1}{|R_i|}. \quad 1.47$$

The relationship between rough entropy and Shannon's entropy can, thus, be defined by Eq. 1.48 (Liang, 2011):

$$E_r(A) + H(A) = \log_2|U|. \quad 1.48$$

1.2.5.4 *Similarity Measurement for Uncertainty—General Similarity Measure*

The entropy measures are introduced in order to address how much uncertainty is associated with non-deterministic phenomena. Yet, what is the data certainty with respect to the deterministic data? In addition, some of the non-deterministic phenomena are expressed in natural language, e.g., pretty large, about 100 km, and quite close. Furthermore, human perception also has inherent uncertainty, which is different from other uncertainties, e.g., the degree of membership value, and group/interval number. In light of this observation, similarity measure was proposed as an alternative (Song et al., 2015; van Eck and Waltman, 2009; Candan and Li, 2001; Khorshidi and Nikfalazar, 2017). Its prominent application domains include pattern recognition (Chen and Chang, 2015; Papacostas et al., 2013; Zeng et al., 2016; Chen et al., 2016; Nguyen, 2016) and decision making (Ye, 2014; Li et al., 2015; Chen, 2015c; Luukka, 2011).

In fact, the study of similarity is an established domain in various research branches of mathematics, such as topology and approximation theory. Typically, a distance function is used in order to identify the similarity between two instances (e.g., patterns, images, and reasoning) quantified.

1.2.5.5 *Similarity Measurement for Uncertainty—Fuzzy Similarity Measure*

Literally, fuzzy similarity measure refers to the calculation of the similarity (or proximity) relationships between fuzzy sets. Zadeh (1971) offered a more formal definition, in which fuzzy similarity measure is considered as the classical equivalence notion's multivalued generalization.

Typically, the methods of fuzzy similarity measures can be broadly categorised into three groups: Set-theoretic, proximity-based and logic-based (Cross and Sudkamp, 2002). Like the fuzzy entropy, fuzzy similarity measure also includes the following three variants:

- Similarity measure for interval-valued fuzzy sets (Chen and Chen, 2009; Ye and Du, in press; Chen, 2015c),
- Similarity measure for intuitionistic fuzzy sets (Li and Cheng, 2002; Liang and Shi, 2003; Farhadinia, 2014; Baccour et al., 2013), and
- Similarity measure for interval-valued intuitionistic fuzzy sets (Xu, 2007).

Interested readers should refer to (Pappis and Karacapilidis, 1993; Xing, 2017b; Wang, 1997; Zwick et al., 1987) for further discussions. In fact, fuzzy similarity measure can be regarded as the dual concept of fuzzy entropy. Accordingly, the relationships between similarity measures and entropy measures have been intensively investigated in the literature (Li et al., 2012; Zhang et al., 2014; Zeng and Li, 2006; Deng et al., 2015; Zeng and Guo, 2008).

1.3 Smart Computing

Recently, the First International Conference on Smart Computing and Informatics was successfully held on 3–4 March 2017 with the aim of offering a unified platform that can incorporate multi-disciplinary and the state-of-the-art research in terms of designing smart computing and information systems. According to (Satapathy et al., 2018), the theme of the conference was to focus on a diverse variety of innovation schemes in system science, artificial intelligence, and sustainable development that can be applied to offer solutions to various problems encountered in society, environment and industries. The scope of smart computing and informatics also consists of the deployment of emerging computational and knowledge transfer methodologies, as well as optimization techniques in distinct disciplines across science, technology, and engineering.

With the engineered system (e.g., financial system) becoming more and more advanced and sophisticated, the associated analysis and synthesis tasks are ever-increasingly demanding. This book borrows the 'smart computing' concept from the literature for the purpose of addressing these issues. Essentially, smart computing comprises different tools developed in other disciplines such as system theory, optimization theory, and computational intelligence. In principle, the process of analysing and synthesizing complex systems via smart computing concept includes: (1) Obtaining each possible action or feasible solution through analysis, (2) Evaluating the obtained outcomes against a certain scale of value or

desirability, and (3) Determining the most desired action or optimum solution according to the selected criterion of a system's decision-based goals. This process can be illustrated in Fig. 1.4.

Figure 1.4: Process of smart computing for optimal decision-making.

The development of smart computing has experienced several stages, ranging from the static optimization approaches that are unfortunately incapable of handling or guaranteeing global solutions, through the recent multiple objective optimization methodologies (e.g., particle swarm optimization and ant colony optimization), to the emerging adaptive dynamic stochastic techniques. The evolutionary footprint of smart computing is depicted in Fig. 1.5.

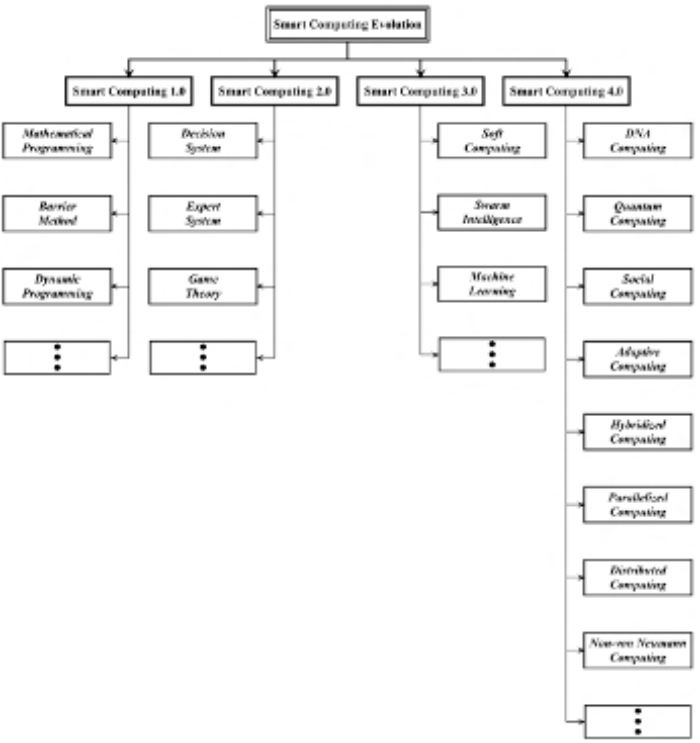


Figure 1.5: Smart computing evolution paradigm.

Regarding the exemplary examples under each category, interested readers should refer to (Suri, 2017; Hageback, 2017; Fister and Fister Jr, 2015; Indiveri, 2015; Polkowski and Artiemjew, 2015; Xing and Gao, 2014b; Chen et al., 2007; McGeoch, 2014; Yu, 2017; Hager and Wellein, 2011; Jeannot and Žilinskas, 2014; Xing, 2017a; Hurwitz et al., 2015; Marinescu, 2013; Xing et al., 2013b; Xing and Gao, 2014m; Xing and Gao, 2014x; Xing and Gao, 2014q; Xing and Marwala, 2018l; Xing, 2015c; Xing, 2014; Xing and Gao, 2014aa; Xing and Gao, 2014bb; Xing and Gao, 2014c; Xing and Gao, 2014cc; Xing and Gao, 2014d; Xing and Gao, 2014dd; Xing and Gao, 2014e; Xing and Gao, 2014f; Xing and Gao, 2014g; Xing and Gao, 2014h; Xing and Gao, 2014i; Xing and Gao, 2014j; Xing and Gao, 2014k; Xing and Gao, 2014l; Xing and Gao, 2014n; Xing and Gao, 2014o; Xing and Gao, 2014p; Xing and Gao, 2014r; Xing and Gao, 2014s; Xing and Gao, 2014t; Xing and Gao, 2014u; Xing and Gao, 2014v; Xing and Gao, 2014w; Xing and Gao, 2014y; Xing and Gao, 2014z; Xing and Gao, 2014ff; Marwala, 2007b; Marwala, 2013a; Marwala and Lagazio, 2011a; Marwala, 2010a; Marwala, 2012a; Marwala, 2014a).

1.4 Data Analytics

Imagine you stroll through any neighbourhood today, when you approach a building, the front door can slide open automatically. When you enter an empty room, a light can flick on by itself. When you jump up and down, a thermostat can trigger the air conditioner that compensates for the gradually warming air around you. When you roam at will, various motion-sensing surveillance cameras can slowly turn to keep you tracked. All these automated electromechanical gadgets work tedious (or even dangerous) jobs that were once performed by human beings.

At the heart of this story is the emergence of data and the corresponding analytics as the means to facilitate our lives. In fact, today's data is being produced so fast that the whole volume tends to be very large. According to one estimation, by the end of 2020 the total amount of data in the world annually will exceed 44 trillion gigabytes (IBM, 2014). Undoubtedly, data brings many benefits to us. However, in the meantime, there are many challenges associated with handling and making the sense of those data (Marwala, 2009; Marwala, 2015).

The basic idea behind data analytics is to collect raw data and convert them into meaningful information that is essential for making decisions. Unfortunately, extracting valuable information from the raw data is not as easy as it may sound. It is difficult to know where to start with the handling of data. To address these challenges, individuals and/or companies must resort to the process of data analytics, which typically includes three phases: Pre-processing, analytics, and post-processing

(Myatt and Johnson, 2014; Verbeke et al., 2018; EMC Education Services, 2015). A detailed overview of this process is depicted in Fig. 1.6.

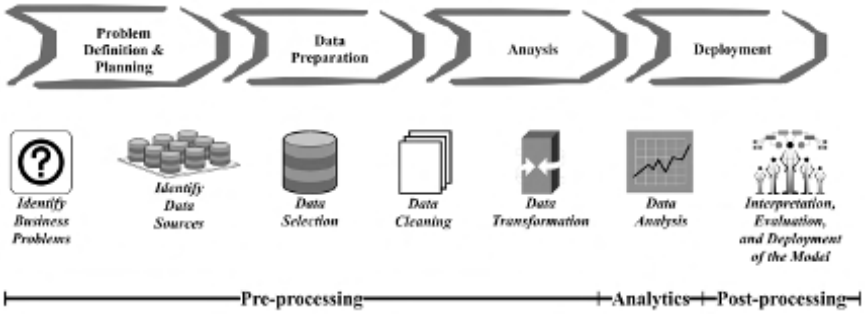


Figure 1.6: General data analytics scheme.

1.4.1 Define Critical Problem and Generate Working Plan

In the first step, a clear definition of the critical problem being addressed needs to be obtained, together with the generation of a feasible working plan. To achieve this, one has to take the following factors into account (Myatt and Johnson, 2014): (1) Outline possible deliverables, (2) Identify success causes, (3) Understand useful resources and their constraints, (4) Assemble a suitable team, (5) Come up with a working plan, and (6) Perform an analysis regarding costs and benefits.

1.4.2 Identify Data Sources

After defining the problem, we need to identify the data sources in order to support the project. In reality, there are numerous sources with various representations and formats (related to the core problem) that must be figured out. The golden rule here is that the more data there is available, the better the results.

1.4.3 Select and Explore the Data

Before a formal data analysis is conducted, a preliminary selection and exploration process towards the collected datasets should be considered. The objective of this step is to help the data analytics teams familiarize themselves with the interested data, i.e., how many cases are covered in the data table, what variables are included and what general hypotheses the data are likely to support.

- **Structured Data Selection and Exploration:** It is commonly agreed that the selection and exploration of structured data can be performed in a relatively controlled manner, e.g., accessing and combining

data tables, and summarizing the data by using some descriptive/inferential statistics. The main objective is to improve the reliability of data, thereby ensuring that no substantial different values should be obtained from the repetition of measurement.

1) *First*: The starting point is normally a data table (or called dataset), which consists of the measured or collected data values expressed in the form of numbers or texts. One of the most common ways to deal with de-normalized source data tables is to merge them into a spreadsheet, where the raw data is outlined as rows and columns, denoting observations and variables, respectively. Based on scale, we can classify variables into four classes: Nominal-, ordinal-, interval-, and ratio-scale. Meanwhile, according to the roles they play in the mathematical models, variables can be broadly categorized into two categories, i.e., independent variables and response variables (Myatt and Johnson, 2014).
2) *Second*: Next, in order to get some initial insights with respect to a specific characteristic, we need to summarize the data. Among others, the most commonly reported characteristics for a particular variable include central tendencies, frequency distribution patterns, and something about the real-world estimations deduced from sub-sets of data (e.g., initial hypotheses to test the data) (Myatt and Johnson, 2014). To this end, several descriptive and/or inferential statistical approaches are needed, such as mode, median, mean, variances, standard deviations, confidence intervals, and hypothesis tests. Further discussions on the topic can be found in (Rumsey, 2010; Weiss, 2017; Ott and Longnecker, 2016). In addition, to make data better visualized (e.g., showing trends, outliers, and relationships among data variables), several graphical methods (e.g., bar charts, scatter plots, and box plots) can be employed (Myatt and Johnson, 2014).

- **Unstructured Data Selection and Exploration:** According to IBM's estimation (IBM, 2014), there are about 2.5 quintillion bytes of data being generated every day and among them, 90% are contributed by new technologies (e.g., smart phones and Internet of things) which are characterized by features such as semi-structured, quasi-structured and unstructured (Dobre and Xhafa, 2014; Sivarajah et al., 2017; Wessler, 2013). In general, the structured data format enjoys good predictability, high machine readability and may often be utilized as input to many other applications; while for semi-structured, quasi-structured and/or unstructured type of data, no data tables are ready to be used (Zhai and Massung, 2016; EMC Education Services, 2015), thus the operation involving computationally generated different attributes relevant to the target problem needs to be performed