Preface

The ideas introduced in this book explore the relationships among rule-based systems, machine learning and big data. Rule-based systems are seen as a special type of expert systems, which can be built by using expert knowledge or learning from real data. From this point of view, the design of rule-based systems can be divided into expert-based design and data-based design. In the present big data era, the latter approach of design, which typically follows machine learning, has been increasingly popular for building rule-based systems. In the context of machine learning, a special type of learning approach is referred to as inductive learning, which typically involves the generation of rules in the form of either a decision tree or a set of if-then rules. The rules generated through the adoption of the inductive learning approach compose a rule-based system.

The focus of this book is on the development and evaluation of rule-based systems in terms of accuracy, efficiency and interpretability. In particular, a unified framework for building rule-based systems, which consists of the operations of rule generation, rule simplification and rule representation, is presented. Each of these operations is detailed using specific methods or techniques. In addition, this book also presents some ensemble learning frameworks for building ensemble rule-based systems. Each of these frameworks involves a specific way of collaborations between different learning algorithms. All theories mentioned above are designed to address the issues relating to overfitting of training data, which arise with most learning algorithms and make predictive models perform well on training data but poorly on test data.

Machine learning does not only have a scientific perspective but also a philosophical one. This implies that machine learning is philosophically similar to human learning. In fact, machine learning is inspired by human learning in order to

simulate the process of learning in computer software. In other words, the name of machine learning indicates that machines are capable of learning. However, people in other fields have criticized the capability of machine learning by saying that machines are neither able to learn nor outperform people intellectually. The argument is that machines are invented by people and their performance is totally dependent on the design and implementation by engineers and programmers. It is true that machines are controlled by programs in executing instructions. However, if a program is an implementation of a learning method, then the machine will execute the program to learn something. On the other hand, if a machine is thought to be never superior to people, this will imply that in human learning students would never be superior to their teachers. This is not really true, especially if a student has the strong capability to learn independently without being taught. Therefore, this should also be valid in machine learning if a good learning method is embedded in the machine.

In recent years, data mining and machine learning have been used as alternative terms in the same research area. However, the authors consider this as a misconception. According to them, data mining and machine learning are different in both philosophical and practical aspects.

In terms of philosophical aspects, data mining is similar to human research tasks and machine learning is similar to human learning tasks. From this point of view, the difference between data mining and machine leaning is similar to the difference between human research and learning. In particular, data mining, which acts as a researcher, aims to discover something new from unknown properties, whereas machine learning, which acts as a learner, aims to learn something new from known properties.

In terms of practical aspects, although both data mining and machine learning involve data processing, the data processed by the former needs to be primary, whereas the data processed by the latter needs to be secondary. In particular, in data mining tasks, the data has some patterns which are previously unknown and the aim is to discover the new patterns from the data. In contrast, in machine learning tasks, the data has some patterns which are known in general but are not known to the machine and the aim is to make the machine learn the patterns from the data. On the other hand, data mining is aimed at knowledge discovery, which means that the model built is used in a white box manner to extract the knowledge which is discovered from the data and is communicated to people. In contrast, machine learning is aimed at predictive modelling, which means that the model built is used in a black box manner to make predictions on unseen instances.

The scientific development of the theories introduced in this book is philosophically inspired by three main theories—namely information theory, system theory and control theory. In the context of machine learning, information theory generally relates to transformation from data to information/knowledge. In the context of system theory, a machine learning framework can be seen as a learning

system which consists of different modules including data collection, data pre-processing, training, testing and deployment. In addition, single rule-based systems are seen as systems, each of which typically consists of a set of rules and could also be a subsystem of an ensemble rule-based system by means of a system of systems. In the context of control theory, learning tasks need to be controlled effectively and efficiently, especially due to the presence of big data.

<div align="right">
Han Liu
Alexander Gegov
Mihaela Cocea
</div>

![Springer logo]

Rule Based Systems for Big Data
A Machine Learning Approach
Liu, H.; Gegov, A.; Cocea, M.
2016, XIII, 121 p. 38 illus., 5 illus. in color., Hardcover