

Preface

The ideas introduced in this book explore the relationships among big data, machine learning and granular computing. In many studies, machine learning has been considered as a powerful tool of big data processing. The relationship between big data and machine learning is very similar to the relationship between resources and human learning. In this context, people can learn from resources to deal with new matters. Similarly, machines can learn from big data to resolve new problems. However, due to the vast and rapid increase in the size of data, learning tasks have become increasingly more complex. In this context, traditional machine learning has been too shallow to deal with big data sufficiently, so granular computing concepts are used in this book to advance machine learning towards the shift from shallow learning to deep learning (in its broader sense).

The focus of this book is on the development and evaluation of granular computing based machine learning approaches in terms of classification accuracy. In this context, the authors consider traditional machine learning to be of single-granularity and the proposal of granular computing based machine learning is aimed at turning single-granularity learning into multi-granularity learning. In particular, the authors proposed the following transformations: (a) supervised learning to semi-supervised learning, (b) heuristic learning to semi-heuristic learning, (c) single-task learning to multi-task learning, (d) discriminative learning to generative learning and (e) random data partitioning to semi-random data partitioning. In addition, the authors also explore how to achieve in-depth evaluation of attribute-value pairs towards induction of high-quality rules, in the setting of multi-granularity learning.

Multi-granularity learning is not only a scientific proposal to address issues of traditional machine learning, but also an indication of philosophical inspiration from real-life problems. For example, it is assumed in traditional machine learning that different classes are mutually exclusive and each instance is clear-cut. However, this assumption does not always hold in reality, e.g. the same movie may belong to more than one category or the same book may belong to more than one

subject. This indicates that there could be specific relationships between different classes, such as mutual exclusion, correlation and mutual independence, which has inspired the authors to adopt the concept of relationships between granules. In this context, each class is viewed as a granule, and generative learning needs to be adopted instead of discriminative learning, if the relationship between classes is not mutual exclusion.

On the other hand, in the context of computer science, correlation is also referred to as association, which is considered as a horizontal relationship between classes. Also, different classes may involve hierarchical relationships such as inheritance and aggregation. In other words, one class can be specialized or decomposed into several sub-classes and several classes may be generalized or aggregated into a super-class. In this book, the authors define that classes that involve a horizontal relationship are viewed as granules located in the same level of granularity and that classes that involve hierarchical relationships are viewed as granules located in different levels of granularity.

Based on the above definition, if all the classes are in the same level of granularity, the learning task could simply be undertaken in the setting of single-granularity learning, unless the task is complex. In particular, it would be considered the case of single-task learning, if the learning task is just aimed to discriminate one class from the other classes towards classifying a clear-cut instance. Otherwise, multi-task learning needs to be undertaken, if the learning task is aimed at judging independently on each class in terms of the membership or non-membership of an instance to the class. In contrast, if these classes are in different levels of granularity, it would usually be necessary to undertake multi-granularity learning.

Multi-granularity learning is also inspired philosophically from the examples of human learning and military training. In the context of human learning, a student typically needs to take a course that involves modules in different levels of a degree. In other words, a student needs to learn and pass all the basic modules in one level, such that the student can progress to the next level towards learning and passing more advanced modules. In the setting of granular computing, each module of a course is viewed as a granule and each level of a degree is viewed as a level of granularity.

In the context of military training, a soldier normally needs to be with a particular unit at one of the military levels, such as squad, platoon, company, battalion, regiment, division and corps. In other words, military training could be organized at different scales, which needs to consider units of different levels as basic units, e.g. a company can be considered as a basic unit in a small scale of training, whereas a regiment needs to be considered as a basic unit in a large scale of training. In the setting of granular computing, each unit is viewed as a granule and each military level is viewed as a level of granularity.

For each of the transformations discussed in this book (e.g. supervised to semi-supervised learning, heuristic to semi-heuristic learning), the granular framework is clearly outlined. In addition, two case studies are presented, one on biomedical data and one on sentiment analysis.

Cardiff, UK
Portsmouth, UK

Han Liu
Mihaela Cocea

Granular Computing Based Machine Learning
A Big Data Processing Approach

Liu, H.; Cocea, M.

2018, XV, 113 p. 27 illus., 19 illus. in color., Hardcover

ISBN: 978-3-319-70057-1