# Chapter 2
# Cat Swarm Optimization Supported Data Mining

**Pei-Wei Tsai**
*National Kaohsiung University of Applied Sciences, Taiwan*

**Jeng-Shyang Pan**
*National Kaohsiung University of Applied Sciences, Taiwan*

**Bin-Yih Liao**
*National Kaohsiung University of Applied Sciences, Taiwan*

**Shu-Chuan Chu**
*Cheng Shiu University, Taiwan*

**Mei-Chiao Lai**
*Diwan University, Taiwan*

## ABSTRACT

*This chapter reviews the basic idea and processes in data mining and some algorithms within the field of evolutionary computing. The authors focus on introducing the algorithms of computational intelligence since they are useful tools for solving problems of optimization, data mining, and many kinds of industrial issues. A feasible model of combining computational intelligence with data mining is presented at the end of the chapter with the conclusions.*

## INTRODUCTION

*Data Mining (DM)* is a series of processes, which analyses the data and sieves some useful information or interesting knowledge out from real-world large and complex data sets (Ghosh & Jain, 2005). Various statistics, analysis, and modeling methods are employed to find patterns and relationships in DM. The process of *Knowledge Discovery from Data (KDD)* is the key, which makes the outcome of DM being meaningful. Nevertheless, the fast development of computer science, the database management system (DBMS) and Data Warehouse (DW) pushes the size of the datasets increases forward with an astounding speed. It results in that precisely extracting the knowledge or finding the relationships and patterns become more difficult. Hence, the need for powerful tools to assist DM is clear. To build such

tools for assisting DM, one of the ways is to lead in the evolutionary computing.

In this chapter, the concept of DM is reviewed, and several algorithms in evolutionary computing are presented. In addition, a feasible solution, which combines the intelligent computing and DM, is proposed. Finally, the discussion that concludes the chapter is made.

## CONCEPT OF DATA MINING

Data mining is an essential step in the process of KDD as depicted in Figure 1 (Han & Kamber, 2007). By using data mining as a tool, several models, which are used to simulate the situations in the real world, can be created to describe the relationships between social behaviors in the real world and the patterns observed in the data. These patterns and relationships are very useful for increasing enterprise profit. For instance, the patterns and their relations to the actual shopping behaviors in customers help the store owners understand how to place their goods to make the customers feel more convenient when shopping. Hence, discovering the knowledge, which is hidden in the data, via data mining is very important.

In data mining, we can create *6* kinds of model to assist us to discover knowledge from the data. The models are listed as follows:

1. Regression.
2. Classification.
3. Time-Series Forecasting.
4. Clustering.
5. Association Rule.
6. Sequence Discovery.

No matter which model is employed, the training data is required to construct the knowledge from mining. The training data may comes from the historical records in the database or simply separates the data into two parts. One part of the data is used as the training dataset, and the other part of the data is used as the test dataset.

## Regression

It is a kind of data analysis method used to fit an equation to a dataset. The simplest form of linear regression consists of a linear formula of a straight line in equation (1). After determining the appropriate values of *a* and *b*, the predicted value of *y* can be found based upon a given value of *x*. In more complicated system, such as multi regression, the acceptable number of the input variable is increased, and the model can be constructed by a quadratic equation. (Chapple, n.d.)

$$y=a \cdot x+b \tag{1}$$

*Figure 1. Data mining as a step in the process of knowledge discovery*
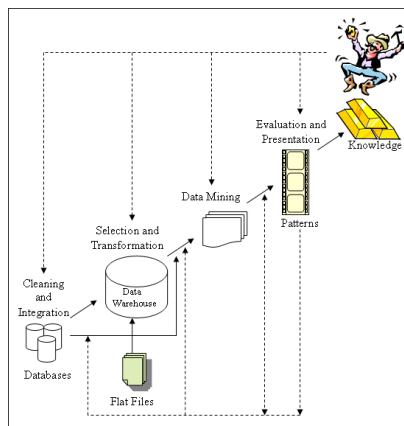
## Related Content

Improving Healthcare with Data-Driven Track-and-Trace Systems
Eldar Sultanowand Alina M. Chircu (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 1229-1246).*
www.irma-international.org/chapter/improving-healthcare-with-data-driven-track-and-trace-systems/150213

The Power of Sampling and Stacking for the PaKDD-2007 Cross-Selling Problem
Paulo J.L. Adeodato, Germano C. Vasconcelos, Adrian L. Arnaud, Rodrigo C.L.V. Cunha, Domingos S.M.P. Monteiroand Rosalvo F.O. Neto (2008). *International Journal of Data Warehousing and Mining (pp. 22-31).*
www.irma-international.org/article/power-sampling-stacking-pakdd-2007/1804

A Survey on Fuzzy Association Rule Mining
Harihar Kalia, Satchidananda Dehuriand Ashish Ghosh (2013). *International Journal of Data Warehousing and Mining (pp. 1-27).*
www.irma-international.org/article/survey-fuzzy-association-rule-mining/75613

Data Warehousing: The 3M Experience
Hugh J. Watson, Barbara H. Wixomand Dale L. Goodhue (2004). *Organizational Data Mining: Leveraging Enterprise Data Resources for Optimal Performance  (pp. 202-216).*
www.irma-international.org/chapter/data-warehousing-experience/27917

Document Classification
 (2021). *Developing a Keyword Extractor and Document Classifier: Emerging Research and Opportunities (pp. 132-136).*
www.irma-international.org/chapter/document-classification/268466