

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

DOI: 10.4018/978-1-61520-757-2.ch002

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.

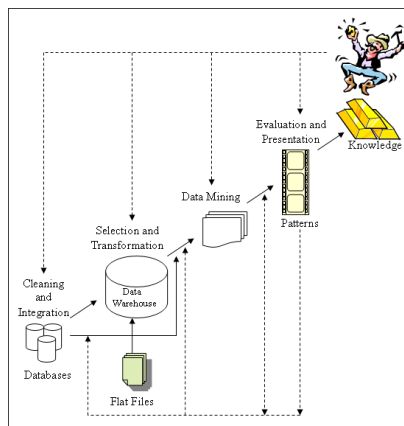
No matter which model is employed, the training data is required to construct the knowledge from mining. The training data may come from the historical records in the database or simply separates the data into two parts. One part of the data is used as the training dataset, and the other part of the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

$$y=a \cdot x+b \quad (1)$$

Figure 1. Data mining as a step in the process of knowledge discovery



16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/cat-swarm-optimization-supported-data/42354

Related Content

Preference-Based Frequent Pattern Mining

Moonjung Cho, Jian Pei, Haixun Wang and Wei Wang (2005). *International Journal of Data Warehousing and Mining* (pp. 56-77).

www.irma-international.org/article/preference-based-frequent-pattern-mining/1759

Data Mining in Aviation Safety Data Analysis

Reima Suomia and Olli Sjöblom (2009). *Social and Political Implications of Data Mining: Knowledge Management in E-Government* (pp. 178-198).

www.irma-international.org/chapter/data-mining-aviation-safety-data/29071

U.S. Federal Data Mining Programs in the Context of the War on Terror: The Congress, Courts, and Concerns for Privacy Protections

Shahid M. Shahidullah (2009). *Social Implications of Data Mining and Information Privacy: Interdisciplinary Frameworks and Solutions* (pp. 151-180).

www.irma-international.org/chapter/federal-data-mining-programs-context/29149

The Charleston Comorbidity Index

Patricia Cerrito (2010). *Text Mining Techniques for Healthcare Provider Quality Determination: Methods for Rank Comparisons* (pp. 153-201).

www.irma-international.org/chapter/charleston-comorbidity-index/36636

On the Usage of Structural Information in Constrained Semi-Supervised Clustering of XML Documents

Eduardo Bezerra, Geraldo Xexéo and Marta Mattoso (2008). *Successes and New Directions in Data Mining* (pp. 67-86).

www.irma-international.org/chapter/usage-structural-information-constrained-semi/29955