Data Mining Techniques in Agricultural and Environmental Sciences

Altannar Chinchuluun, University of Florida, USA

Petros Xanthopoulos, University of Florida, USA

Vera Tomaino, University of Florida, USA and University Magna Græcia of Catanzaro, Italy

P.M. Pardalos, University of Florida, USA

ABSTRACT

Data mining techniques are largely used in different sectors of the economy and they increasingly are playing an important role in agriculture and environment-related areas. This paper aims to show our vision on the importance of knowing and efficiently using data mining and machine learning-related techniques for knowledge discovery in the field of agriculture and environment. Efforts for searching hidden patterns in data are not a recent phenomenon. History shows that extensive observations on data have helped discover empirical laws in different fields of research. Therefore, it is important to provide researchers in agriculture and environmental-related areas with the most advanced knowledge discovery techniques. Data mining is the process of extracting important and useful information from large sets of data. This information can be converted into useful knowledge that could help to better understand the problem in study and to better predict future developments. The paper presents the state of the art in data mining and knowledge discovery techniques and provides discussions for future directions.

Keywords: Agriculture; Artificial Neural Networks; Data Mining; kNearest Neighbor; k-Means; Optimization; Support Vector Machines

Keywords: Agricultural and Environmental Sciences, Data Mining Techniques

INTRODUCTION

The problem of searching for patterns in data is a fundamental one and has a long and successful history. There are many examples in different research areas that extensive observations of data has led to discovering empirical laws. As

DOI: 10.4018/jaeis.2010101302

an example, the attentiv astronomical observations undertaken by several astronoms allowed Kepler to discover the laws of planetary motion (Bishop, 2006).

Over the years, several techniques have been developed to discover hidden patterns in data and these efforts led to the creation of a rigourous discipline known as data mining or knowledge discovery. Data mining is the process of finding useful patterns or correlations amongst data. These patterns, associations, or relationships between data can provide information about the problem in study and information can then be transformed into knowledge. The idea of using information hidden into relationships amongst data inspired researchers to apply these techniques for predicting future trends (Mucherino, Papajorgi, & Pardalos, 2009). Data mining techniques are developed from mainly three areas: statistics, artificial intelligence and machine learning. Although the roots of data mining may seem different, but essentially they aim the same target: discover a relationship that more or less maps measurements in one part of a data set to measurements in another, linked part of the data set (Pyle, 2003).

Regardless of the method used, the goal of data mining techniques is to split data in different categories, each of them representing some feature of interest the data may have (Mucherino et al., 2009). Thus, fundamental for the success of a data mining technique is the ability to group available data in disjoint categories, where each category contains data with similar properties. The similarity between data is usually measured using a distance function; similar data should belong to the same group or cluster. Therefore, the success of a data mining technique depends on the adequate definition of a suitable distance between data samples.

As the similarity between data samples is measured using a distance function, often it

occurs that this distance needs to be optimal. Thus, many data mining techniques led to the formulation of a global optimization problem (Mucherino et al., 2009).

Data mining techniques can be grouped in three categories as shown in Figure 1.

Statistical Methods

Statistical methods such as Principal Component Analysis (PCA) and regression techniques are commonly used as simple methods for finding patterns in sets of data. PCA is a useful statistical technique that has found application in fields such as image compression, and is a common technique for finding patterns in big data sets. PCA helps identifying patterns in data, and expressing the data in such a way as to emphasize their similarities and differences. Since patterns in large data sets can be hard to find, because graphical representation of the data is not available, PCA is a powerful tool for analyzing data.

The main advantage of PCA is that once patterns in the data are identified data can be represented as components ordered by their relevance and it is possible then to discard components of low level of relevance without loss of important information and thus, reducing the complexity of the problem. In many cases, dimension reduction makes it possible to represent data graphically that enormously facilitates the understanding of discovered patterns.



Figure 1. A schematic representation of the classification of the data mining techniques

13 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: <u>www.igi-</u> <u>global.com/article/data-mining-techniques-agricultural-</u> <u>environmental/39026</u>

Related Content

Matrix Converter: A Solution for Electric Drives and Control Applications

Megha Vyasand Shripati Vyas (2022). *Optimal Planning of Smart Grid With Renewable Energy Resources (pp. 219-244).* www.irma-international.org/chapter/matrix-converter/293180

What Lessons Can Be Learned for the Agroecological Transition From the Use of Social Media in Preventive Medicine?

Vincent Soulignac, François Pinet, Mathilde Bodeletand Hélène Gross (2023). International Journal of Agricultural and Environmental Information Systems (pp. 1-28).

www.irma-international.org/article/what-lessons-can-be-learned-for-the-agroecologicaltransition-from-the-use-of-social-media-in-preventive-medicine/316936

Seeking Cities of the Future: Techno-City Visions From the 1960s to the Present

A.Hilal lavarone (2022). International Journal of Digital Innovation in the Built Environment (pp. 1-11).

www.irma-international.org/article/seeking-cities-of-the-future/306257

Rectangular Ribbons and Generalized Topological Relations

Brahim Lejdeland Okba Kazar (2016). *International Journal of Agricultural and Environmental Information Systems (pp. 70-88).*

www.irma-international.org/article/rectangular-ribbons-and-generalized-topologicalrelations/158096

Bioremediation of Environmental Pollutants

Akash Akash, Vinay Mohan Pathak, Neelesh Babuand Navneet Navneet (2022). Research Anthology on Emerging Techniques in Environmental Remediation (pp. 110-134).

www.irma-international.org/chapter/bioremediation-of-environmental-pollutants/291230