Knowledge Discovery and Data Visualization:
Theories and Perspectives

Kijpokin Kasemsap, Suan Sunandha Rajabhat University, Bangkok, Thailand

ABSTRACT

This article reviews the literature in the search for the theories and perspectives of knowledge discovery and data visualization. The literature review highlights the overview of knowledge discovery; Knowledge Discovery in Databases (KDD); Knowledge Discovery in Textual Databases (KDT); the overview of data visualization; the significant perspectives on data visualization; data visualization and big data; and data visualization and statistical literacy. Knowledge discovery is the process of searching for hidden knowledge in the massive amounts of data that individuals are technically capable of generating and storing. Data visualization is an easy way to convey concepts in a universal manner. Organizations, that utilize knowledge discovery and data visualization, are more likely to find both knowledge and information they need when they need them. The findings present valuable insights and further understanding of the way in which knowledge discovery and data visualization efforts should be focused.

INTRODUCTION

Knowledge discovery is a data-intensive process that allows for finding patterns, trends, and models (Congiusta, Talia, & Trunfio, 2008) in a broad range of sources, such as relational databases, images, and texts (Wanderley, Tacla, Barthès, & Paraiso, 2015). Current database technology involves processing a large volume of data in order to discover new knowledge. Many methods concerning knowledge discovery have been proposed to find useful information and to discover useful knowledge, such as decision trees, Bayesian belief networks, evidence theory, rough set theory, fuzzy set theory, kNN (k-nearest-neighborhood) classifier, neural networks, and support vector machines (Wu, McGinnity, Prasad, & Bell, 2009).

Data visualization is a collection of techniques translating data from its numeric format to the graphic presentation for visual impact (Zhu, Hoon, & Teow, 2015). Data visualization is the modern technology that provides pictorial descriptions of results from computations (Leng, Rhyne, & Sharrock, 2012) and is closely connected with data analysis approaches (Mehlan et al., 2013). Data visualization transforms data into information, and visualization converts information into picture

forms and creates graphic images to express the meaning (Ursyn, 2015). New technologies offered by modern browsers greatly increase the capabilities for data visualization on the Web 2.0 (Mwalongo, Krone, Reina, & Ertl, 2016).

This article is based on a literature review of knowledge discovery and data visualization. The extensive literature of knowledge discovery and data visualization provides a contribution to practitioners and researchers by explaining the advanced issues of knowledge discovery and data visualization in the digital age.

THEORETICAL AND PRACTICAL REVIEW OF THE LITERATURE

The literature review indicates the overview of knowledge discovery; Knowledge Discovery in Databases (KDD); Knowledge Discovery in Textual Databases (KDT); the overview of data visualization; the significant perspectives on data visualization; data visualization and big data; and data visualization and statistical literacy.

Overview of Knowledge Discovery

Efficient knowledge discovery can be achieved by introducing domain knowledge of experts to the creation of business action (Yada, 2004). There is an urgent need for effective methods to extract unknown and unexpected information from data sets of unprecedentedly large size (e.g., millions of observations), high dimensionality (e.g., hundreds of variables), and complexity (e.g., heterogeneous data sources, space-time dynamics, multivariate connections, spatial relations, and interactions) (Mehenni, 2017).

Machine learning is an approach of knowledge discovery, whereby autonomous algorithms are utilized for acquiring knowledge toward enhancing the organization of the knowledge obtained (Tecuci & Kodratoff, 1995). One important step in applying machine learning technique is to decide an effective knowledge representation scheme for both the training data and the knowledge extracted (Chen, Khoo, Chong, & Yin, 2014). Knowledge representation involves the modeling of knowledge in the explicit schemes that promote the acquisition, learning, manipulation, and application of knowledge (Chen et al., 2014). These schemes include mathematical expressions, predicate calculus, conceptual graphs, frames, scripts, objects, semantic networks and production rules (Chong, Chen, & Leong, 2009).

Brain is the soft convoluted mass of nervous tissue within the skull of vertebrates that is the commanding and coordinating center of the nervous system and the seat of thought, memory, and emotion (Kasemsap, 2017a). The selection of brain tissue is a challenging task because it is difficult to distinguish neural tissue from other neighboring structures (Uher & Burget, 2012). Segmentation is often very important for the evaluation of medical treatment and its success for some diseases, such as for expressing the ratio between brain volume and the affected part of the brain. Extracted brain image represented by voxels can be used to measure and monitor the brain volume in time or for other medical purposes. In addition, the relationship between how brain activities respond and learning can be enhanced by brain segmentation technology with the support of cloud computing (Chang, 2014a). Regarding knowledge discovery in the health care industry, the visualized three-dimensional (3D) model can help a surgeon to perform an operation more accurately (Uher & Burget, 2012).

The Cloud Computing Business Framework (CCBF) is proposed to help organizations gain good Cloud design, deployment, migration, and services (Chang, Walters, & Wills, 2013). The IT Infrastructure Library (ITIL) Version 3 (ITIL V3) service framework provides the best practice in IT service management (Hanna et al., 2009). In addition, ITIL V3 offers an advanced framework for

12 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/article/knowledge-discovery-and-data-visualization/182757

Related Content

Congestion Management Using Hybrid Particle Swarm Optimization Technique

Sujatha Balaramanand N. Kamaraj (2012). *Innovations and Developments of Swarm Intelligence Applications (pp. 165-181).*

www.irma-international.org/chapter/congestion-management-using-hybrid-particle/65812

An Experimental Performance Comparison for Indexing Mobile Objects on the Plane

S.Sioutas, G. Papaloukopoulos, K. Tsichlasand Y. Manolopoulos (2012). *Intelligent and Knowledge-Based Computing for Business and Organizational Advancements (pp. 340-358).*

www.irma-international.org/chapter/experimental-performance-comparison-indexing-mobile/65802

An Adaptive Intrusion Detection Scheme for Cloud Computing

Nurudeen Mahmud Ibrahimand Anazida Zainal (2019). *International Journal of Swarm Intelligence Research (pp. 53-70).*

www.irma-international.org/article/an-adaptive-intrusion-detection-scheme-for-cloud-computing/234699

Unit Commitment by Evolving Ant Colony Optimization

K. Vaisakhand L. R. Srinivas (2012). *Innovations and Developments of Swarm Intelligence Applications (pp. 207-218).*

www.irma-international.org/chapter/unit-commitment-evolving-ant-colony/65814

Recommending Turmeric Variety for Higher Production Using Interval-Valued Fuzzy Soft Set Model and PSO

R. K. Mohantyand B. K. Tripathy (2021). *International Journal of Swarm Intelligence Research (pp. 94-110).*

www.irma-international.org/article/recommending-turmeric-variety-for-higher-production-using-interval-valued-fuzzy-soft-set-model-and-pso/276787