Homeland Security Data Mining and Link Analysis

Bhavani Thuraisingham  
The MITRE Corporation, USA

INTRODUCTION

Data mining is the process of posing queries to large quantities of data and extracting information often previously unknown using mathematical, statistical, and machine-learning techniques. Data mining has many applications in a number of areas, including marketing and sales, medicine, law, manufacturing, and, more recently, homeland security. Using data mining, one can uncover hidden dependencies between terrorist groups as well as possibly predict terrorist events based on past experience. One particular data-mining technique that is being investigated a great deal for homeland security is link analysis, where links are drawn between various nodes, possibly detecting some hidden links.

This article provides an overview of the various developments in data-mining applications in homeland security. The organization of this article is as follows. First, we provide some background on data mining and the various threats. Then, we discuss the applications of data mining and link analysis for homeland security. Privacy considerations are discussed next as part of future trends. The article is then concluded.

BACKGROUND

We provide background information on both data mining and security threats.

Data Mining

Data mining is the process of posing various queries and extracting useful information, patterns, and trends often previously unknown from large quantities of data possibly stored in databases. Essentially, for many organizations, the goals of data mining include improving marketing capabilities, detecting abnormal patterns, and predicting the future, based on past experiences and current trends. There is clearly a need for this technology. There are large amounts of current and historical data being stored. Therefore, as databases become larger, it becomes increasingly difficult to support decision making. In addition, the data could be from multiple sources and multiple domains. There is a clear need to analyze the data to support planning and other functions of an enterprise.

Some of the data-mining techniques include those based on statistical reasoning techniques, inductive logic programming, machine learning, fuzzy sets, and neural networks, among others. The data-mining problems include classification (finding rules to partition data into groups), association (finding rules to make associations between data), and sequencing (finding rules to order data). Essentially, one arrives at some hypotheses, which is the information extracted from examples and patterns observed. These patterns are observed from posing a series of queries; each query may depend on the responses obtained to the previous queries posed.

Data mining is an integration of multiple technologies. These include data management, such as database management, data warehousing, statistics, machine learning, decision support, and others, such as visualization and parallel computing. There is a series of steps involved in data mining. These include getting the data organized for mining, determining the desired outcomes to mining, selecting tools for mining, carrying out the mining, pruning the results so that only the useful ones are considered further, taking actions from the mining, and evaluating the actions to determine benefits. There are various types of data mining. By this we do not mean the actual techniques used to mine the data, but what the outcomes will be. These outcomes also have been referred to as data-mining tasks. These include clustering, classification anomaly detection, and forming associations.

While several developments have been made, there also are many challenges. For example, due to the large volumes of data, how can the algorithms determine which technique to select and what type of data mining to do? Furthermore, the data may be incomplete and/or inaccurate. At times, there may be redundant information, and at times, there may not be sufficient information. It is also desirable to have data-mining tools that can switch to multiple techniques and support multiple outcomes. Some of the current trends in data mining include mining Web data, mining distributed and heterogeneous databases, and privacy-preserving data mining, where one ensures that one can get useful results from mining and at the same time maintain the privacy of
Related Content

Fuzzy Information and Data Analysis
Reinhard Viertl (2005). Encyclopedia of Data Warehousing and Mining (pp. 519-522).
www.igi-global.com/chapter/fuzzy-information-data-analysis/10652?camid=4v1a

Entity Resolution on Single Relation
(2014). Innovative Techniques and Applications of Entity Resolution (pp. 87-122).
www.igi-global.com/chapter/entity-resolution-on-single-relation/103245?camid=4v1a

Two Rough Set Approaches to Mining Hop Extraction Data
www.igi-global.com/chapter/two-rough-set-approaches-mining/7682?camid=4v1a

Evolution of Data Cube Computational Approaches
Rebecca Boon-Noi Tan (2005). Encyclopedia of Data Warehousing and Mining (pp. 469-476).
www.igi-global.com/chapter/evolution-data-cube-computational-approaches/10643?camid=4v1a