Behavior Mining of Spatial Objects with Data Field

WANG Shuliang  WU Juebo  CHENG Feng  JIN Hong  ZENG Shi

Abstract  The advanced data mining technologies and the large quantities of remotely sensed imagery provide a data mining opportunity with high potential for useful results. Extracting interesting patterns and rules from data sets composed of images and associated ground data can be of importance in object identification, community planning, resource discovery and other areas. In this paper, a data field is presented to express the observed spatial objects and conduct behavior mining on them. First, most of the important aspects are discussed on behavior mining and its implications for the future of data mining. Furthermore, an ideal framework of the behavior mining system is proposed in the network environment. Second, the model of behavior mining is given on the observed spatial objects, including the objects described by the first feature data field and the main feature data field by means of the potential function. Finally, a case study about object identification in public is given and analyzed. The experimental results show that the new model is feasible in behavior mining.

Keywords  behavior mining; data field; spatial object identification; spatial data mining

Introduction

Data mining is the nontrivial extraction of implicit, previously unknown, potentially useful and ultimately understood patterns from data. It is the science of extracting useful information from large data sets or databases [1, 2]. With the application and development of modern science, techniques, micro-sensors and macro-sensors, the tremendous amounts of spatial and nonspatial data have been stored in large spatial databases. The traditional data mining cannot match the requirement, so Li et al. presented the concept of “Spatial Data Mining and Knowledge Discovery (SDMKD)” and gave an overview about SDMKD [3]. However, since people from different backgrounds need different information and knowledge, more and more methods and algorithms from other disciplines are introduced into data mining, which has made a great progress in recent years.

Behavior mining is a branch of data mining that focuses on finding behavioral patterns from large common data, so as to do object’s behavior identification, classification, prediction. By now, most of the contributions on behavior mining are limited in the field of web application but fewer in other fields. A new mechanism, which consists of preprocessing phase, two-layer pattern discovering phase, and pattern explanation phase, was proposed to discover unknown patterns on mining network behavior [4]. To
mine high impact exceptional behavior patterns, Cao et al. developed several types of high impact exceptional behavior patterns that were associated with either positive or negative impact, and frequent behavior patterns that led to both positive and negative impact[5]. Based on behavior mining and machine learning techniques, Mukkamala et al. described the results concerning the classification capability of unsupervised and supervised machine learning techniques in detecting intrusions using network audit trails [6]. Kuo et al. provided an approach to discover frequent learning patterns and divided a sequence database into three sets to perform real-time learning behavior mining[7,8]. Although behavior mining has been applied within a few areas, it cannot be found in the field of spatial data. In this paper, a model of behavior mining is proposed on the observed spatial objects, which aims to object identification, peculiarity recognition and so on.

The rest of the paper is organized as follows. In Section 1, the major principles and techniques are presented, including the brief introduction to data field and the definition of behavior mining of spatial objects. A typical behavior mining system is also described in the network environment. In Section 2, a model of behavior mining of spatial objects is given based on the data field and gives a detailed discussion on it. A case study on object identification and prediction is given in Section 3, while the conclusion and future work is given in Section 4.

1 Principles and techniques

1.1 Data field

Data field is proposed in behavior mining of spatial objects as the key technique used in this paper. This section gives a concise description of the main ideas on the data field that will be used in the following sections.

The main idea of data field originated from physical fields, which is used to describe how the data essentials are diffused from the discourse of sample to the discourse of universe. It is given to express the power of an item in the discourse of universe by means of potential function as the physical field does.

The description of potential function based on the assumption that all observed data in the number universe will radiate their data energies and be influenced by others simultaneously. Taking into account that the obtained spatial data are comparatively incomplete, it is necessary for the observed data to radiate their data energies from the sample space to their parent space [9].

Data field is often classified into two different categories by their behavior under the symmetry transformations of space-time, i.e. scalar field and vector field. In this paper we choose scalar field with its potential function to express properties. For a single data field created by sample \( A \), the potential of a point \( x_1 \) in the number universe can be computed by:

\[
\varphi(x) = m \times e\left(-\frac{|x-x_1|}{\sigma}\right)
\]

where, \( \|x-x_1\| \) is the distance between \( A \) and \( x_1 \), \( m \) \((m \geq 0)\) denotes the power of \( A \) and \( \sigma \) indicates influential factors. Usually, \( \|x-x_1\| \) is Euclidean norm.

The influential factors \( \sigma \), e.g. radiation brightness, radiation gene, data amount, space between the neighbour isopotential, grid density of Descartes coordinate, and so on, all make their contributions to the data field.

In most cases, there is more than one sampling point in the number universe. In order to obtain the power of one point under these circumstances, all energies from every sample should be concerned. Because of overlap, the potential of each point in the number universe is the sum of all data potentials. Referring to potential Eq. (1), the potential can be calculated by:

\[
\varphi(x) = \sum_{i=1}^{n} \left( m_i \times e\left(-\frac{|x-x_i|}{\sigma}\right) \right)
\]

where, the sum is over all the sample points. The relevant symbols are explained in Eq. (1).

In a similar way to the distribution of scalar physical field, the equipotential line or surface can be utilized to describe the spatial distribution of potential function in the low-dimensional potential field. More specifically, given a potential value \( \Psi \), the corresponding equipotential line or surface can be obtained, that is, according to the set of potential values \( \{\Psi_1, \Psi_2, \cdots\} \) which satisfy \( \varphi(x) = \Psi \), a series of equipotential lines or surfaces can be the spatial distribution