Chapter 33
A Data Mining Framework for Forest Fire Mapping

Ahmed Toujani
Silvopastoral Institute of Tabarka, Tunisia & LTSIRS Laboratory, University of Tunis El Manar, Tunisia

Hammadi Achour
Silvopastoral Institute of Tabarka, Tunisia

ABSTRACT

Forest fires constitute the major reasons for the loss of biodiversity and degradation of ecosystems. Locally, forest fires are one of the major natural risks in the Kroumirie mountains, northwestern Tunisia. In these massifs, fires occur frequently, and this requires understanding the complex biophysical parameters of this phenomenon. The special attention of the research is paid to the spatial forecasting of forest fires. Different types of classical frequent itemset algorithms have been tested and employed to reveal forest fire patterns that relate the spatial parameters with the probability of fire occurrence. Extracted frequent patterns are then being aggregated through a defined measurement of pertinence. The forest fire risk zone maps are then generated, resulting in extracted spatial patterns. The experiments showed that, the integration of these patterns into GIS could be advantageous to determine risky places and able to produce good prediction accuracy.

INTRODUCTION

Due to the advances in scientific data collection and in the artificial intelligence methods, we are faced with an exponentially growing amount of data which are too large and complex to be processed and analyzed through traditional methods. Therefore, the need of transforming data into useful knowledge has contributed to the birth and the increase of Knowledge Discovery in Databases (KDD) and Data Mining (DM) (Frawley et al., 1992). KDD is a rapidly growing research area at the intersection of several disciplines including databases, artificial intelligence, statistics, pattern recognition, machine learning, and data visualization (Zytkow et al., 2002), with the aim of extracting knowledge. Data Mining is the key step in the knowledge discovery process. In general, Data Mining comprises a large set of algorithms

DOI: 10.4018/978-1-5225-5191-1.ch033

and techniques that serve to extract implicit and potentially useful knowledge, and may provide support to the process of decision-making.

Similarly, the spatialization of information systems, the potential of knowledge that contains geospatial data, and the inadequacy of traditional data mining methods have contributed to the naissance of Geographic Knowledge Discovery (GKD). The GKD is a sub branch of KDD which aims to discover implicit and effective knowledge in geospatial database (Miller et al., 2001).

Indeed, spatial information processing technology like remote sensing, GIS, and GPS (Global Positioning System) which is combined with the need of the market, has led to an accumulation of more and more of geospatial data. Unfortunately, due to the complexity of the data, their nature and the relationships between them, the use of traditional data mining techniques proved impossible to produce consistent results (Koperski et al., 1996; Shekhar et al., 2004). The inconsistency of certain results is mainly due to the characteristics of geospatial entities that Miller et al. (2001) summarized as follows: the spatial dependence and the heterogeneity, the diversity of data types and the complexity of spatio-temporal objects.

Two approaches have been proposed (Rinzivillo et al., 2008) to deal with this peculiarity of geospatial data. The first approach is to develop tools, algorithms and methods that are able to dynamically treat the correlations between the geospatial entities. The major drawback of this approach is that it is more focused on the modeling phase (Klösgen et al., 2002) while the other steps of the GKD process are as important as the modeling step. The second approach is to use the traditional data mining tools with being necessary to extract beforehand the conventional data types (string, numeric, boolean, etc.) and the relationships that can exist between these data. This approach is quite useful because it is simple to be implemented and it allows us to exploit the existing tools that have already proven their effectiveness (Klösgen et al., 2002). Just like at the KDD, where the data mining is the main phase, the spatial data mining remains one of the major phases of GKD (Miller et al., 2001), this is the application of techniques and methods to find interesting patterns in objects and events distributed in geographic space and across time (Miller & Han, 2001). The aim of this chapter is to show how to structure data mining methodology for improving spatial geographic risk assessments. We describe in detail the proposed methodology and we show how it can be effectively applied for the for the assessment of fire risk potential of Tunisian forests.

BACKGROUND

Spatial data mining techniques can assist in comprehension and analysis of spatial data, and in exploration of relationships among spatial and non-spatial variables. More specifically, spatial data mining techniques encompass visual analysis, spatial query, characterization and generalization, classification, spatial regression and clustering analysis, detection of spatial and non-spatial association rules besides a wide variety of other fields (Tang & McDonald, 2002).

Conservation planning, precision agriculture, deforestation prevention, resource discovery and various other areas can benefit largely from the extraction of patterns of interest and rules from the spatial data sets like the GIS-data (the data of geographic information system (GIS)) and the remotely sensed imagery (Guo & Mennis, 2009). Spatial Pattern Mining (SPM) is greatly exploited in ecosystem modeling, disaster prevention, forest fire evaluation and other analogous fields. The proposed approach aims to prevent of forest fires with utilizing the spatial data and pattern mining techniques.