

## Research and Application of Hidden Markov Model in Data Mining

Zhang Youzhi

School of Computer and Information  
Anqing Teachers College  
Anqing, China  
E-mail: zyztcboy@163.com

**Abstract**—Data mining is to extract large amounts of data or to mine the information from which they need to find important data model, such as business decision-making, knowledge base, scientific and medical research has made great contributions. As users of data mining for the accuracy and precision required, hidden Markov models are widely used in data mining. This paper describes a hidden Markov model in the field of data mining applications, and with hidden Markov model we will resolve some issues.

**Keywords**—data mining; hidden Markov models; information extraction

### I. INTRODUCTION

Data Mining (Data Mining, DM), also known as knowledge discovery in databases (Knowledge Discovery in Database, KDD), is from the early 90s the rise of a database technology. Data mining from a large number of incomplete, noisy, ambiguous, the practical application of random data, extracting implicit in them, people do not know in advance, but is potentially useful information and knowledge. Data mining is a multidisciplinary cross-product of a combination of database, artificial intelligence, statistics, machine learning and visualization techniques found useful by the new rules and new concepts, improve the data owner a lot of raw data on the deep-seated understanding, awareness and address the "data rich, knowledge poor" problem, has broad application prospects[1-3].

This paper introduces the basic concepts of data mining and technology, and hidden Markov model described in the data mining application.

### II. THE CONCEPT OF DATA MINING

Data Mining (Data Mining, DM), also known as knowledge discovery in databases (Knowledge Discover in Database, KDD), refers to data from the database revealed a large number of implicit, previously unknown and potentially non-trivial value of information process. Data mining is mainly based on artificial intelligence, machine learning, pattern recognition, statistical, database, visualization technology, and is the field of artificial intelligence and database research hot spots.

Data mining contains the following:

Data cleaning: remove noise or inconsistent data; data integration: a variety of data sources can be combined; Data selection: From a database search and analysis tasks related data; Data transformation: data transformation or unified into a form suitable for mining, such as through summary or aggregation operations; Data Mining: Basic steps to use intelligent methods to extract data model; model assessment: According to some interest measure to

identify the knowledge that the real interesting pattern; Knowledge Representation: Use visualization and knowledge representation techniques, to provide users of the knowledge.

A typical data mining system has the following main components:

Database, data warehouse or other repository: This is one or a group of databases, data warehouses, spreadsheets, or other type of information repository, the data in data cleaning and integration. Data warehouse to provide data analysis capabilities, known as OLAP (Online Analytical Processing).

Database or data warehouse server: a request based on user data mining, database or data warehouse server for data extraction.

Knowledge Base: also known as domain knowledge, to guide the search, or the results of the assessment model interestingness. Such knowledge may include the concept hierarchy for the attribute or attribute value organized into different abstraction layers. Other examples of domain knowledge are interested in degree limit or threshold and metadata (for example, describes the data from multiple heterogeneous data sources).

Data mining engine: This is a basic part of the data mining system, a group of functional modules, for characterization, association, classification, clustering analysis, and Evolution and deviation analysis[4].

Data Mining involves the integration of multidisciplinary technologies, including database technology, statistics, machine learning, high performance computing, pattern recognition, neural networks, data visualization, information retrieval, image and signal processing and spatial data analysis. If a system can only retrieve data or information, including a large database to find out together interpretation of value or to answer queries, it should be classified as database systems or information retrieval system, or deductive database system. Data mining can extract interesting knowledge from databases, rule, or high-level information, and can be observed from different angles or visit. Discovered knowledge can be applied to decision-making, process control, information management, query processing and so on. Therefore, the data mining industry is the information database system that is one of the most important front, the information industry, and the most promising interdisciplinary.

Data mining is used to specify the data mining task to find the model type. General data mining tasks can be divided into two categories: description and prediction. Descriptive data mining tasks characterize the general properties. Predictive data mining tasks in the current inferred to predict.