ICAAMM-2016

Featured Based Pattern Analysis using Machine Learning and Artificial Intelligence Techniques for Multiple Featured Dataset

Annaram Soujanya\textsuperscript{a*}, O. Subhash Chander Goud\textsuperscript{b}, Sai Prasad. K\textsuperscript{c}, G Prabhakar Reddy\textsuperscript{d}

\textsuperscript{a} Asst Prof, IARE, JNTUH, Dundigal, Hyderabad, Telangana, India.
\textsuperscript{b} Asst Prof, NIZAM College, OU, Hyderabad, Telangana, India
\textsuperscript{c} Asst Prof, MLRIT, JNTUH, Dundigal, Hyderabad, Telangana, India
\textsuperscript{d} Asst Prof, MLRIT, JNTUH, Dundigal, Hyderabad, Telangana, India

Abstract

Data mining is a process of extracting patterns from a large datasets. We are trying to uncover the data bonded features which are hard to visualize and if many feature exists for the data then it becomes difficult to analyze the data. The main aim of the scheme of work is to classify the features and over which try to classify the data sets. Machine learning is one of the techniques of Artificial Intelligence which is used for extracting valuable knowledge from large data base. Machine Learning is also used for extracting patterns, models in data. In this paper we are trying to group the data based on multi-dimensional feature classification. Clustering process makes the similar features to form into one group and or else multiple groups, here in we try to group the features which are similar and form multiple groups. The US schooling data is in the form of flat files. Classification process is performed on the raw data. Classification is performed according to hierarchical clustering. Filtration process is used in order to obtain non-zero values. The saturation points are generated by performing clustering. Based on the clusters obtained the patterns can be extracted. Attribute based classification and hierarchical clustering is performed on the data. The attributes obtained are named as income and expenses. By combining the income and expenses attributes patterns can be identified. By using some combination of the two attributes some patterns have been obtain. By performing clustering on all the combinations of each attribute we can identify the patterns. Then according to this process we find the patterns generated by the feature classification or clustering process against the data classification or clustering. This new methodology tries to give the co-relation between the data and properties of a data set and how they behave.

© 2017 Elsevier Ltd. All rights reserved.
Selection and Peer-review under responsibility of the Committee Members of International Conference on Advancements in Aeromechanical Materials for Manufacturing (ICAAMM-2016).

Keywords: Classification, DBSCAN Clustering, Multi-dimensional clustering, US Schooling data, Pattern Extraction.
1. Introduction

Text clustering is one of the techniques used for obtaining meaning data from large amount database. There are some traditional clustering algorithms which uses some of the approach like BOW (Bag of Words) approach. WordNet and based lexical chains semantics approaches are used for solving documents clustering. It removes ambiguity, overcomes the high dimensionality of the clusters [1]. The process of mining is extracting meaning data from huge amount of spatial data is known as Spatial Data mining. One of the famous density based clustering algorithm is DBSCAN which performs clustering by increasing high density area and finds any kind of shape of clustering [13]. The two parameters used in DBSCAN clustering are epsilon (eps) and MinPts. The clustering starts from unvested arbitrary starting point. A cluster is formed only when number of neighbor are greater equal to MinPts. After cluster is formed the starting point is marked as visited and neighbor added into that particular cluster. This process is repeated until all the neighbors are visited. If the number of MinPts is less than neighbor then that point is marked as Noise. Clustering are mostly used for identify classes in spatial database [2]. K-means clustering is used in various fields like image and audio compression, system modeling and neutral network structures [3]. Sequential pattern mining, maximum pattern mining are some of the pattern mining [4]. Among them most of the methods are proposed for developing efficient mining algorithms in order to find particular patterns within a reasonable and acceptable time [5].

1. Methodology

In our paper, we have taken the Public School System Finances database for performing clustering and classification process. The Public School System Finances database consists of fifteen thousand records of various schools. The data contains the financial information like revenue, expenditure, debt and assets for public elementary-secondary school collected as a part of the Census of Governments. Classification process is performed by removing the zero values from the 15869 records. The below table consists of various school databases like STATE, IS, SUPID, NAME, 19H, 61V….W61. Table 1 shows the US schooling database.

<table>
<thead>
<tr>
<th>STATE</th>
<th>ID</th>
<th>SUID</th>
<th>NAME</th>
<th>V33</th>
<th>TOTAL REV</th>
<th>61V</th>
<th>66W</th>
<th>W01</th>
<th>W31</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>015001001</td>
<td>503</td>
<td>ALTAUGA CO SCH DIST</td>
<td>7137</td>
<td>21039</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>0</td>
</tr>
<tr>
<td>01</td>
<td>015002001</td>
<td>403</td>
<td>BALDWIN CO SCH DIST</td>
<td>17983</td>
<td>67403</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>0</td>
</tr>
<tr>
<td>01</td>
<td>015003001</td>
<td>303</td>
<td>BARBOUR CO SCH DIST</td>
<td>2255</td>
<td>8554</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>0</td>
</tr>
<tr>
<td>01</td>
<td>015003002</td>
<td>103</td>
<td>EUFAULA CTY SCH DIST</td>
<td>2997</td>
<td>10906</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>0</td>
</tr>
<tr>
<td>01</td>
<td>015004001</td>
<td>203</td>
<td>BIBB CO SCH DIST</td>
<td>3564</td>
<td>11548</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>0</td>
</tr>
<tr>
<td>01</td>
<td>015005001</td>
<td>103</td>
<td>BLOUNT CO SCH DIST</td>
<td>6045</td>
<td>21990</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>0</td>
</tr>
<tr>
<td>01</td>
<td>015005002</td>
<td>903</td>
<td>ONEONTA CTY SCH DIST</td>
<td>1105</td>
<td>3674</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>0</td>
</tr>
<tr>
<td>01</td>
<td>015006001</td>
<td>003</td>
<td>BULLOCK CO SCH DIST</td>
<td>2023</td>
<td>7361</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>0</td>
</tr>
<tr>
<td>01</td>
<td>015007001</td>
<td>903</td>
<td>BUTLER CO SCH DIST</td>
<td>4476</td>
<td>14780</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>0</td>
</tr>
</tbody>
</table>
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات