Investigation of Lazy Classification in Data Mining using WEKA tool

Paramjeet Kaur¹, Poonam Rani²

¹Department of Computer Science and Applications, Chaudhary Devi Lal University, Sirsa, Haryana, India
²Assistant Professor, Department of Computer Science and Applications, Chaudhary Devi Lal University, Sirsa, Haryana, India

ABSTRACT

Lazy classification is would allow data domain of complex nature that cannot be properly explained by various learning algorithms. In this research we are getting Correlation coefficient error, mean absolute error, root mean square error, relative absolute error, root relative square error analysis using WEKA. Calculation of mean absolute error, root mean square error, relative absolute error, root relative square error analysis using WEKA would be made for KSTAR, LWL, and IBK. Comparative analysis would be made of all these three lazy classifiers. Here we would take real dataset of advance handsets. In this reading there are mobile name, screen size, CPU speed, number of Sims, Ram size and pixel. It is found from research that there is minimum error in case of KSTAR.

Keywords: Lazy Learning, Data mining, KSTAR, LWL, IBK

I. INTRODUCTION

Removal of data is not just process that are likely to play out information mining similar incorporates other some procedures like Data Integration, Data Cleaning, Data Mining, Pattern Evaluation, Data Transformation, and Data Presentation.

Purpose of data mining is extracting vital information from huge databases or data warehouses. Many Data mining applications have used for commercial and scientific sides. This kind of study forcefully explains Data Mining system into scientific side. Here Scientific data mining differentiates itself and explores that nature of datasets is various from present market concentrated data mining applications.

Market Analysis and Management

Planned below are many area of market where data mining is used

1. Cross Market Analysis:- Mining performs association / correlations between product sales.
2. Target Marketing:- Data mining helps to find clusters of model customers who share same characteristics such as interests, spending habits, income, etc.
3. Provides summary information:- Data mining provides us various multidimensional summary reports.
II. PROBLEM STATEMENT

Lazy learning is considered as learning processes where process of generalization ahead of training information is delayed until a query for processing is sent to system. It is just opposite to eager learning in which system is expected to generalize training information just before queries are received. Case-based reasoning is a major benefit of lazy learning method as its target function is exist locally just like k-nearest neighbour algorithm.

That is why lazy learning systems could solve several problems. It could also deal modification in problem domain.

The limitation in case of lazy learning consists of huge need of space in order to store complete training dataset. Second limitation is that lazy learning methods are slower to calculate. However, it is used in case of a faster training phase.

III. OBJECTIVES OF RESEARCH

There are several classifiers in existence. In this research discussion for lazy classifier is made. Three classifiers KSTAR, LWL, IBK are considered here and comparisons of efficiency of such classifiers are made. Objective of research is as follow:-

1. To take a dataset of handset in order to make comparative analysis of different lazy classifiers.
2. Apply classifiers KSTAR, LWL, IBK on data set.
3. To calculate Relative Absolute Error, Absolute Error, Root Mean Square Error and relative square error.
4. To make comparative analysis of error in case of different classifiers.
5. To make conclusion of efficient of lazy classifier among them.

IV. RESULT ANALYSIS

Reading data set from WEKA tool

Here we would take real dataset of advance handsets. In this reading there are mobile name, screen size, CPU speed, number of Sims, Ram size and pixel.

Getting mean absolute error, root mean square error, relative absolute error, root relative square error analysis

The calculation of mean absolute error, root mean square error, relative absolute error, root relative square error analysis using WEKA would be made for KSTAR, LWL and IBK. Comparative analysis would be made of all these three lazy classifiers.

Getting mean absolute error, root mean square error, relative absolute error, root relative square error analysis using WEKA (process and screen shot).
Data analysis:-

1. **Correlation coefficient error**:-

   **Table 1.** Data Set of Correlation coefficient error

<table>
<thead>
<tr>
<th>Correlation coefficient analysis</th>
<th>INSTANCES</th>
<th>KSTAR</th>
<th>IBK</th>
<th>LWL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>0.1651</td>
<td>0.3423</td>
<td>0.6366</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.0282</td>
<td>0.1052</td>
<td>-0.2164</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>-</td>
<td>0.1003</td>
<td>-0.2161</td>
</tr>
</tbody>
</table>

Here K-STAR has minimum Correlation coefficient error it is 0.0521 in case of K-STAR 0.1003 in case of IBK and -0.2161 in case of LWL.

2. **Mean absolute error**:-

   **Table 2.** Mean absolute error

<table>
<thead>
<tr>
<th>Mean absolute error</th>
<th>INSTANCE</th>
<th>KSTAR</th>
<th>IBK</th>
<th>LWL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>1.8671</td>
<td>1.6</td>
<td>1.2388</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>4.6405</td>
<td>5.55</td>
<td>5.1444</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>3.6696</td>
<td>4.6</td>
<td>4.0446</td>
</tr>
</tbody>
</table>

Here K-STAR has minimum mean absolute error it is 3.6696 in case of K-STAR 4.6 in case of IBK and 4.0446 in case of LWL.

3. **Root mean squared error**:-

   **Dataset:-**

   **Table 3.** Root Mean squared error

<table>
<thead>
<tr>
<th>Root mean square error</th>
<th>INSTANCES</th>
<th>KSTAR</th>
<th>IBK</th>
<th>LWL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>1.8671</td>
<td>1.6</td>
<td>1.2388</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>4.6405</td>
<td>5.55</td>
<td>5.1444</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>3.6696</td>
<td>4.6</td>
<td>4.0446</td>
</tr>
</tbody>
</table>

Here K-STAR has minimum root mean squared error it is 3.6696 in case of K-STAR 4.6 in case of IBK and 4.0446 in case of LWL.
4. **Relative absolute error:**

   **Table 4. Relative absolute error**

<table>
<thead>
<tr>
<th>INSTANCE</th>
<th>KSTAR</th>
<th>IBK</th>
<th>LWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>105.03%</td>
<td>9000%</td>
<td>69.68%</td>
</tr>
<tr>
<td>20</td>
<td>6.62%</td>
<td>79.16%</td>
<td>73.38%</td>
</tr>
<tr>
<td>30</td>
<td>72.01%</td>
<td>90.26%</td>
<td>79.36%</td>
</tr>
</tbody>
</table>

   Here K-STAR has minimum relative absolute error it is 72.01 in case of K-STAR 90.26 in case of IBK and 79.36 in case of LWL.

5. **Root relative squared error:**

   **Table 5. Root relative squared error**

<table>
<thead>
<tr>
<th>INSTANCE</th>
<th>KSTAR</th>
<th>IBK</th>
<th>LWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>111.20%</td>
<td>95.46%</td>
<td>73.86%</td>
</tr>
<tr>
<td>20</td>
<td>99.34%</td>
<td>102.60%</td>
<td>98.81%</td>
</tr>
<tr>
<td>30</td>
<td>99.97%</td>
<td>103.21%</td>
<td>98.91%</td>
</tr>
</tbody>
</table>

   Here K-STAR has minimum root relative squared error it is 99.97 in case of K-STAR 103.21 in case of IBK and 98.91 in case of LWL.

V. **CONCLUSION**

Lazy classifiers are case based reasoning technique so they help in reducing redundancy. Case based reasoning techniques are do not store redundant information in base if test case already existing in base. It simply returns class value of existing replica case as a solution.

Lazy learning model supports incremental learning procedure. Model is open to new case or addition of knowledge every times. It is according to human mode of learning.

Lazy classifiers provide greater quality of solutions. In this research we have expressed that error are minimum in case of KSTAR as hand data set increases as compared to different lazy classifiers. Lazy classification is would allow data domain of complex nature that cannot be properly explained by various learning algorithms. It could lead explore CBR structure further in future research work oriented towards image testing, video sequences testing, and different forms of Information.
VI. REFERENCES


