

EXERCISE 2 : PREPROCESSING ON DATASETS USING WEGA TOOL

Procedure

Performing the following Pre-processing process on predefined datasets like weather.numeric.arff and weather.nominal.arff

a) Loading the Data

- Open the WEKA Explorer.
- Navigate to the "Preprocess" tab.
- Use the "Open" button to load a predefined dataset. Specify the directory and select a dataset such as "weather.numeric.arff," or "weather.nominal.arff."
- Inspect the left panel in the "Preprocess" tab to confirm the recognized attributes. Verify that the base relation and current working relation are initially the same.
- Click on any attribute in the left panel to view basic statistics.
For categorical attributes, check the frequency of each attribute value.
For continuous attributes, note statistics such as min, max, mean, standard deviation, etc.
- Select an attribute (e.g., "temperature" in "weather.numeric.arff") and visualize its basic statistics.
- Choose another attribute (e.g., "play" in "weather.nominal.arff") and visualize its basic statistics.
- Explore the visualization in the right bottom panel, which is a form of cross-tabulation across two attributes.
For example, understand the default cross-tabulation of "humidity" with "play."
Using the drop-down list, select another attribute for cross-tabulation (e.g., "outlook") and observe the changes.

b) Selecting or Filtering Attributes

- Load a dataset with multiple attributes (e.g., "labor.arff" or "weather.nominal.arff").
- Navigate to the "Filter" panel and explore various attribute filters available.
- Implement the "weka.filters.unsupervised.attribute.Remove" filter to remove a specific attribute. Mention the correct attribute name to remove (e.g., "windy").
- Apply the "weka.filters.unsupervised.attribute.Add" filter to add a new attribute with appropriate values. Specify the correct attribute name and values.
- Try the "weka.filters.unsupervised.attribute.Normalize" filter to normalize numeric attributes. Specify the correct numeric attribute names.
- Experiment with the "weka.filters.supervised.attribute.AttributeSelection" filter to select relevant attributes. Specify the correct attribute selection options.
- Save the modified dataset after each filter operation.

c) Discretization

Numeric to Nominal Discretization:

- Load the "weather.numeric.arff" dataset.
- Go to the "Preprocess" tab.
- In the "Filter" section, click "Choose" and select "weka.filters.unsupervised.attribute.Discretize."

- Click the box next to "Choose" to configure the Discretize filter.
- Select the numeric attributes you want to discretize (e.g., "temperature," "humidity").
- Set the number of bins or configure other options.
- Click "OK" to close the configuration.
- Click "Apply" to apply the Discretize filter.
- Save the new dataset as "weather-nominal-discretized.arff."

Nominal to Numeric Discretization:

- Open WEKA Explorer.
- Load the "weather.nominal.arff" dataset.
- Go to the "Preprocess" tab.
- In the "Filter" section, click "Choose" and select "weka.filters.unsupervised.attribute.Discretize."
- Click the box next to "Choose" to configure the Discretize filter.
- Select the nominal attributes you want to discretize (e.g., "temperature," "humidity").
- Set the number of bins or configure other options.
- Click "OK" to close the configuration.
- Click "Apply" to apply the Discretize filter.
- Save the new dataset as "weather-numeric-discretized.arff."

d) Handling Missing Values

- Load "weather.numeric.arff."
- Check for missing values in attributes.
- Introduce missing values in "outlook" (Nominal) and "temperature" (Numeric).
- Note max count label in "outlook" and mean of "temperature."
- Apply "ReplaceMissingValues" filter.
- Examine data for replacement.
- Edit "weather.numeric.arff" with a text editor, replace some data with '?' and save as "weather-numeric-missing.arff."
- Load "weather-numeric-missing.arff" in WEKA, observe data.

Video References:

1. <https://youtu.be/x4Z4JolZW6c>
2. <https://youtu.be/1Nk0NbDNm4o>
3. <https://youtu.be/yk0jR1wotwg>
4. <https://youtu.be/f56XrUqzlhv>
5. <https://youtu.be/yk0jR1wotwg>
6. <https://youtu.be/Ys3x2E9WZXQ>