Politecnico di Milano
Data Mining and Text Mining

Driving Style Clustering

http://code.google.com/p/dmdsc/

Filippo Sironi 734456
Matteo Villa 735064

2009

Contents

List of Figures

1 Preprocessing

Before proceeding with the mining phase, we perform a preliminar preprocess step on the data. With the first operation, the nominal attributes representing date and time are converted so the tool can correctly deal with them; then duplicate instances are deleted from the ExampleSet. With the *AttributeFilter* operator we can reduce the dimensionality of the data, dropping those attributes that are not useful for the following activities.

After these first steps the data continue to show missing or inconsistent values; we can easily deal with those situations thanks to the *ExampleFilter* operator, that deletes all the instances whose attributes don't satisfy certain conditions.

Next we apply a *MovingAverage* in order to smooth the values of the timed .

- moving average (MovingAverage) - normalization of numerical attributes (Normalization) - aggregation of multiple instances (MultivariateSeries2WindowExample + AttributeAggregation)