

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	2
----------	---

Contents

1 Preprocessing	4
-----------------	---

List of Figures

1 Preprocessing

Before proceeding with the mining phase, we have to perform a preliminar preprocessing step on the data, in order to eliminate the noise due to the sampling of real 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 next 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 let us delete all the instances whose attributes don't satisfy certain conditions.

Next we apply the *MovingAverage* and *Normalization* operators, in order to smooth the values of the timed series.

Finally we can aggregate multiple data instances into a single one, trying to reduce the overall dimension of the dataset; with the operators *Multivariate-Series2WindowExample* and *AttributeAggregation* a set of data examples can be replaced with a single instance containing the mean of the previous values.