CBR System goals

The main goal of the CBR engine is to predict a response based on the library of previous recorded cases. This particular CBR engine was developed for providing a continuous output for an unseen vector of features. In particular, this CBR system was developed and tested with predicting house prices based on 79 numerical and nominal features.

Set up and demo

All the installation procedure is explained in detail in
`[extracted_folder]/Documentation/Documentation.html`

Functionality

The

Most of the functions, variables, inputs and outputs of functions of the system are described in the report and the documentation file
`[extracted_folder]/Documentation/Documentation.html`

Here we present how to use some functions, whose functionality is important but was not already explained in sufficient detail.

Distance functions

For the retrieval stage, 3 different distance metrics with or without feature weighing have been implemented: Euclidean “EUCL”, Euclidean Weighted “EUCL+W”, Manhattan “MANH”, Manhattan Weighted “MANH+W” and Eixample “LEIX”, as described in the accompanying paper. In order to select the required distance measure, provide dist_meas variable in CBR_retrieve as one of the above. For example:

```
CB_ret = CBR_retrieve(case_base=CB,
                      new_case=new_case_demo_X,
                      new_label=new_case_demo_y,
                      k=2,
                      dist_meas="EUCL+W")
```

Retention strategies

Three options for retentions strategy have been provided, namely Always, Never and Regular. Refer to main report for explanation about each of it.

In order to define which strategy the CBR engine must use, this can be done in by supplying strategy variable in CBR_Retain function.

```
retained = CBR_Retain(CB_ret, strategy="Regular")
```

Creating Eixample map

In the data preprocessing stage it is necessary to produce Eixample Map in order to use the Eixample distance. See the main report for details about what it is and its usage. Note: for the data provided, the map has already been created.

To create a new Eixample Map, run

```
python make_LEixample_map.py path/to/input/file.csv
path/to/output/file.csv
```

from the console. The program asks to provide information about each feature. Use keyboard to enter values as requested. The process consists of the following steps:

Step 1. Data type. Type in the corresponding value as per below and press enter

```
-1: Skip
0: Numerical
1: Catergorical ordered
2: Catergorical un-ordered
```

Step 2. Data binning ranges for numerical data. Initially the data is made of 1 bin, ranging from minimum to maximum of the data. By entering cut-off points, additional bins are created. E.g. feature has values from 0 to 100. Entering 25,50,75 creates 4 bins.

- or -

Step 2. Provide the rank of lineal data in ascending order

```
Attribute 18 - ExterQual has the following values
0 : Ex
1 : Fa
2 : Gd
3 : TA
Type ids of these attributes in the ascending order
1,3,2,0
```

Once completed the function generates an “Eixample map” .csv, which can then be used for Retrieval using Eixample distance.