Evaluation Example

A Drift-based Dynamic Ensemble Members Selection using Clustering For Time Series Forecasting

Description:

This package enables a dynamic selection of heteregeneous ensemble base models through a performance drift detection mechanism and ensures ensemble diversity through a second stage selection using clustering that is computed after each drift detection. Predictions of final selected models are combined single prediction using sliding-window averages or stacking.

Illustrative examples

```
path=
#specify path where you have placed the scripts 'packages.R', "functions.R","data-preparation.R" and
#the dataset example

setwd(path)


source("packages.R")
source("functions.R")
source("data-preparation.R")


# example of evaluation


######train the pool of models and generate predictions

data.train=train.reg
data.test=test.reg
n=nrow(data.train)
val.length=10
H=10
data.val=data.train[((n-val.length+1):n), ]
data.train.n=data.train[(1:(n-val.length)), ]
formula=target~.
models=train.models(data.train.n, formula,ker=c("rbfdot" ,"polydot" ,"vanilladot", "laplacedot"))

predictions.table=predict.models(models,data.test,formula, per.arima,ker=c("rbfdot" ,"polydot" ,"vanill


sapply(1:ncol(predictions.table), function(x) rmse(predictions.table[,1],predictions.table[,x]))
```

```
#######compute the drift top best performing base models using the sliding window validation set  of moc

tp1=15 #number of top models
lim=0.1
val.length=20#length of the validation set

updated_selection=topk.model.sel(models,data.train, data.test,val.length, formula, per.arima,ker=c("rbf
#identify models
updated_selection1=updated_selection$models.sel
#identify instant where the drift alarm was triggered
alarm=updated_selection$alarm



# Our main methods: THE TWO MAIN VARIANTS OF THE METHOD: both of them cluster top best performing model
# and cluster them using IMLEC. The main difference is on the combination : one approach use ensemble
#the second one use stacking


##### top model clustering using IMLEC and then aggregation in a sliding window ensemble
# Clusters are updated with each drift detection
H=10
st=Sys.time()
val.length1=10 #number of features for the clusters (predictionss of the models)
tp.cl=compute.cluster.imlec(models,data.train,data.test,val.length1,H, formula, per.arima,ker=c("rbfdot
ed=Sys.time()

pred.tp.imlec=tp.cl$predictions

pos.models=tp.cl$selected_models_id


rmse(data.test$target,pred.tp.imlec)





# top model clustering using IMLEC and then combination using stacking
# Clusters are updated with each drift detection

pos.models=tp.cl[[2]]#see the evaluation extracted from the updated clusters represnetatives

st=Sys.time()
val.length1=50 #increase length of the validation set because it will be used for the training of the mo
pred.st.cl=compute.cl.tp.imlec.stac(models,data.train, data.test,t,val.length1, formula, per.arima,ker=
ed=Sys.time()


#sapply(1:12, function(x) rmse(data.test$target,pred.st.all[[x]]))

pred.top.imlec.st=pred.st.cl[[10]]
```

```
rmse(data.test$target,pred.top.imlec.st)
```

Contact

Any bug report or suggestion please contact me at amal.saadallah@cs.tu-dortmund.de.

If you use our method, please cite the paper "A Drift-based Dynamic Ensemble Members Selection using Clustering for Time Series Forecasting" (ECML2019)