SmartFly: Predict probability of delay of scheduled flights

Cindy Lamm

17:22, Sunday 18th January, 2015

Load prepared data from the previous step "Prepare Data For Prediction"

```
rm(list=ls())   #clear memory
load("../04_prepare_data_for_prediction/rfPredictData.RData")
```

and load the estimated model:

```
load("../03_train_model/delayRf.RData")
```

```
library(randomForest)
# formula is_delayed ~ . - id will take care for not using id for prediction
predictionDelayRf <- predict(delayRf, rfPredictData, type="prob")
```

Save the prediction result in R format:

```
save(predictionDelayRf, file="../05_predict_delay_proba/rfPrediction.RData")
```

Write the prediction result as requested into csv:

```
o <- order(predictionDelayRf[,2], decreasing=TRUE)
head(predictionDelayRf[o,])

##        on_time delayed
## 27859    0.100   0.900
## 40700    0.104   0.896
## 292940   0.106   0.894
## 499920   0.108   0.892
## 551857   0.110   0.890
## 16455    0.112   0.888

idIdx <- which(names(rfPredictData) == "id")
idsOrderedByDelayProba <- rfPredictData[o,idIdx]
write.table(idsOrderedByDelayProba, file="../../out/problem1.csv",
            quote=FALSE, col.names=FALSE, row.names=FALSE)
```

The end!