

# A first example of how to extract Frequent Patterns

Solon Ioannou

11/8/2020

## Connect to the database

```
### Define database parameters
cdmdatabaseschema = "main"
resultsdatabaseschema = "main"

connectionDetails <- Eunomia::getEunomiaConnectionDetails()
connection <- connect(connectionDetails)
```

```
## Connecting using SQLite driver
```

```
#on.exit(DatabaseConnector::disconnect(connection)) #Close db connection on error or exit
```

## Define cohort

```
# Define cohort
cohort <- readSql("../data/cohorts/Eunomia_MI_cohort.sql")

renderTranslateExecuteSql(connection, cohort, cdm = "main")
```

```
##      |
## Executing SQL took 0.0446 secs
```

```
sql <- "ALTER TABLE #diagnoses ADD cohort_definition_id INT NOT NULL DEFAULT(1)"

# Execute the script to receive the data
renderTranslateExecuteSql(connection, sql)
```

```
##      |
## Executing SQL took 0.000895 secs
```

```
querySql(connection, "SELECT count(*) FROM diagnoses;")
```

```
##      COUNT(*)
## 1          67
```

Get the data and close the connection

```
# Define covariate settings
TemporalcovariateSettings_eunomia <- createTemporalCovariateSettings(useConditionOccurrence = TRUE,
                                                                    temporalStartDays = seq(-(60*365), -1, by = 1) ,
                                                                    temporalEndDays = seq(-(60*365)+1, 0, by = 1))

# Extract covariates
TemporalcovariateData_eunomia <- getDbCovariateData(connection = connection,
                                                    cdmDatabaseSchema = cdmdatabaseschema,
                                                    cohortDatabaseSchema = resultsdatabaseschema,
                                                    cohortTable = "diagnoses",
                                                    rowIdField = "subject_id",
                                                    covariateSettings = TemporalcovariateSettings_eunomia,
                                                    cohortTableIsTemp = TRUE)

## Sending temp tables to server
## Constructing features on server
## |
## Executing SQL took 40.8 secs
## Fetching data from server
## Fetching data took 0.176 secs

disconnect(connection)
```

Prepare the data

```
getInputFileForFrequentPatterns(covariateDataObject = TemporalcovariateData_eunomia, fileToSave = "example1.txt")

## Extracting temporal data...

## Extracting covariate names...

## Generating input file for frequent pattern mining...

## Input data has been created succesfully and saved in example1.txt
```

Running Analysis

```
frequentPatterns <- runFrequentPatterns(algorithm = "SPADE", inputFile = "example1.txt", outputFile = "output.txt")

## Analysing 67 sequence IDs...

## Running frequent pattern algorithm...

## Running algorithm
```

```
## [1] "java -jar ../inst/java/spmf.jar run SPADE example1.txt results_example1.txt 0.5 true"
```

```
## The command line that has been running is: java -jar ../inst/java/spmf.jar run SPADE example1.txt re
```

```
str(frequentPatterns)
```

```
## 'data.frame': 2469 obs. of 3 variables:  
## $ Sequence: chr "Streptococcal sore throat " "Osteoarthritis " "Acute bronchitis " "Coronary arter  
## $ Count : chr "34" "61" "60" "66" ...  
## $ Support : num 0.507 0.91 0.896 0.985 0.955 ...
```

```
dfWithIDs <- getIdDataFrame("results_example1.txt")  
dim(dfWithIDs)
```

```
## [1] 67 2470
```