



CIV8760E - Lab 05

Agenda

→ Presentation of PA4

→ Teamwork time

Creation of the database

GTFS data on [Transit Feed](#).

Each line of the database must correspond to one direction of a service line.

Don't forget to remove the subway from your data!

Create database

Inter-stop time: You're going to have a time between each stop, so you need to choose a descriptive statistic (mode, median, mean, etc.) to describe the inter-stop time for EACH service line direction.

Accessibility: You have one value per stop on the line and one value for the line, so you need to suggest a value to associate with the line based on the available values.

Segmentation of lines

Tanagra

Select your data (including titles) in your Excel file and copy it (avoid accents and spaces).

Create a new text file (".txt"), open it and paste your data.

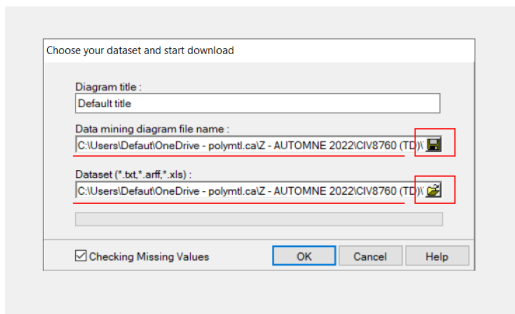
In Tanagra, create a new project file.

Line segmentation

Tanagra

Make sure the file is saved in a known location (under "Data mining diagram file name", you can change the location of the project).

Under "Dataset (*.txt, *.arff, *.xls)", fetch your data file



Segmentation of lines

Tanagra

Define your attributes (*Define status*). Target ? Input ?

Several segmentation methods exist → You've seen k-means in class.

How do you find the right number of classes?

Segmentation of lines

Tanagra

Figure: Within-class homogeneity (WSS) and between-class dispersion (BSS)

Total Sum of Square (TSS) = BSS + WSS

Determining the optimum number of classes: Elbow method or crossing point

Line segmentation

Tanagra

To export data after segmentation, go to the "Data visualization" tab and apply the "Export dataset" function

In the "Export dataset" parameters, you can specify the location and name of the data file (".txt" format).

This text file now contains the class for each line in your database. You can use it with another file in the GTFS data to create your map!

Line segmentation

QGIS

Use the "Points to path" tool with the help of the *shape.txt* file to display STM lines.

Class analysis

Tanagra

You can import in the same way, but this time take the newly created file containing the classes.

A decision tree is a supervised method, so it needs both *input* and *output*!

Which algorithm is used to build the tree? What does it calculate to measure its quality?

Class analysis

Tanagra

Qualitative analysis of the tree and its factors (variables).

Modalities

20 pages maximum

Due December 10, 2023 by 11:59 PM

You can share another type of file (.txt or .csv) if you have calculated your indicators with a tool other than Excel.

Teamwork time

LET'S DO IT!