Tutorial overview

While developing TANAGRA, the underlying objective was to give access to a lot of data mining methods, and not to manage with the numerous formats of dataset files (anyway, it is more the purpose of a commercial software).

TANAGRA loads data from text files with tab separator, built in the following way:

- 1st line : names of attributes
- next lines : values of the attributes for the sample (one line for each record).
- N.B: TANAGRA does not manage missing data.

In this tutorial, we show the steps for data importation from an Excel spreadsheet towards TANAGRA. We choose Excel as the original source of data, because it is generally easy to copy the data you're working on, and paste it in an Excel sheet.

Saving an Excel sheet to TANAGRA format

- 1 Launch Excel.
- 2 Enter the example data as in the following screenshot.

M	licrosoft Excel	- Classeur1				_ D ×
	Eichier Edition	Affichage Inse	ertion Forma <u>t</u>	<u>O</u> utils <u>D</u> onnées	Fe <u>n</u> être <u>?</u>	<u>_8×</u>
	68	👌 🖻 🖬 🖞	🕈 n - 🍓	Σ f* 😨)
	E16 💌	=		C No constant		
	A	B	С	D	E	F 🗖
1	outlook	temperature	humidity	windly	play	
2	sunny	85	85	FALSE	no	
3	sunny	80	90	TRUE	no	
4	overcast	83	86	FALSE	yes	
5	rainy	70	96	FALSE	yes	
6	rainy	68	80	FALSE	yes	
7	rainy	65	70	TRUE	no	
8	overcast	64	65	TRUE	yes	
9	sunny	72	95	FALSE	no	
10	sunny	69	70	FALSE	yes	
11	rainy	75	80	FALSE	yes	
12	sunny	75	70	TRUE	yes	
13	overcast	72	90	TRUE	yes	
14	overcast	81	75	FALSE	yes	100
15	rainy	71	91	TRUE	no	
16						
17						-
	Feuil1	/ Feuil2 / Feuil3	3/			
Prê	t				NUM	

3 – In the *File* menu, choose the "*Save as…*" command. A dialog box opens, in the "Type of file" list, choose « Text (separator: tabulation) ». Enter « weather.txt » as the name of the file.

Enregistrer sous									? ×
Enregistrer <u>d</u> ans :	Tutorials		•	4 E 🔇	X	🔢 👻 Outi	<u>s</u> •		
Historique									2
Mes documents									N
B ureau									
Favoris									
Dossiers Web	Nom du <u>fi</u> chier : <u>T</u> ype de fichier :	weather.txt	ur: tabulatio	on) (*.txt) (*	.txt)	-		inregistrer Annuler	
		,							

4 - Click on the Save button.

5-A dialog box advertises you that, in textfile format, only one sheet can be saved at the same time. Click the OK button.

Microsoft	Excel
~	Le type de fichier sélectionné ne prend pas en charge les classeurs qui contiennent plusieurs feuilles.
	 Pour enregistrer seulement la feuille active, cliquez sur OK. Pour enregistrer toutes les feuilles, procédez feuille après feuille en leur attribuant des noms de fichiers différents ou bien sélectionnez un type de fichier qui prenne en charge plusieurs feuilles.
	OK Annuler

6 – Another dialog box tells you that « weather.txt » can contain data that are incompatible with the text file format. Click on Yes.

Microsoft	Excel
	weather.txt peut contenir des informations non compatibles avec Texte (séparateur: tabulation) (*.txt) . Voulez-vous conserver le format du classeur ?
į	 Cliquez sur Oui pour conserver le format. Les fonctionnalités non compatibles seront perdues. Cliquez sur Non pour conserver ces caractéristiques. Enregistrez ensuite une copie de votre document dans le format de fichier Excel le plus récent Cliquez sur Aide pour vérifier les pertes possibles.

7 – The data are ready to be imported into TANAGRA. Close Excel.

8 – You can visualize the file you've just saved under any text editor. Here is this file « weather.txt », opened in the Windows notepad.

<u>F</u> ichier	<u>E</u> dition <u>I</u>	<u>R</u> echerche	2				
outlool	k tempe	rature	humidit	ty	windly	play	A.
sunny	85	85	FALSE	no			
sunny	80	90	TRUE	no			
overcas		83	86	FALSE	yes		
rainy	70	96	FALSE	yes			
rainy	68	80	FALSE	yes			
rainy	65	70	TRUE	no			
overcas	st	64	65	TRUE	yes		
sunny	72	95	FALSE	no			
sunny	69	70	FALSE	yes			
rainy	75	80	FALSE	yes			
sunny	75	70	TRUE	yes			
overcas	st	72	90	TRUE	yes		
overcas		81	75	FALSE	yes		
rainy	71	91	TRUE	no			
							· · ·

Importing your data in TANAGRA

- 1 Choose "File/New..." in the main menu of TANAGRA.
- 2 Enter a title for the diagram: « TANAGRA : Importing Data ».

3 – Enter the name of the associated file in which you will save your work (« TANAGRA_ImportingData.bdm »). Before click on this button to run through the hard disk and place yourself in the directory « ...\TANAGRA\Tutorials ».

4 – Click on the icon to seek the file you have created "weather.txt".

Tanagra : In	iporting data
Data mining	diagram file name :
C:\Program	Files\Tanagra\Tutorials\TanagraImportingData.bdm 🛛 🛽
Dotooot (tau	t file) ·
Dataset (tex	

5 – Validate with OK to start data importation.

🎋 Tanagra - [Dataset (wea 🏧 <u>F</u> ile <u>D</u> iagram Componer						_ [] > _ [] >	
D 🛩 🖬 🍇							
Tanagra : Im	porting data			Dataset	(weather.txt)	A	
Dataset (weather.t)		Parameters					
	Database : C:\Program Files\Tanagra\Tutorials\weather.txt						
		Results					
		5 attribute(s) 14 example(s					
		Attribute	Category	Informations			
		outlook	Discrete	3 values			
		temperature	Continue	-			
		humidity	Continue	-			
		windly	Discrete	2 values			
		play	Discrete	2 values			
		Execution tim Created at 08		:11:03			
		Comp	onents				
Data visualization	Descriptive stats	Instance	e selectior	n Feat	ture construction		
Feature selection	Regression	Factori	al analysis		Clustering		
Spv learning	Meta-spv learning	Spv learni	ng assesme	ent	Association		
Correlation scatterplot							
Export dataset	🗓 View multiple scatte	erplot					
🖉 Scatterplot							

A new diagram is created, based on the file « weather.txt ». You can see the description of its contents in the right frame.

Important precisions on diagrams saving formats

While creating a new diagram, we have used the extension "BDM" for the name of the diagram file (« TANAGRA_ImportingData.bdm »).

BDM (Binary Diagram) is the extension for a TANAGRA file containing, besides the diagram description, the imported dataset (here the data from the file « weather.txt »). So we loose the association to the original data textfile, which can be modified or deleted without impacting the work we done under TANAGRA.

TDM (Text Diagram), the other TANAGRA file format for diagrams, is the one used if, besides the diagram description, you only want to save a **reference** to the data file. So that if you make any change in the data file (here « weather.txt »), you will have different results the next time you execute the sequences of operations in TANAGRA! But be careful not to delete the source file if you want to be able to reopen the diagram.