1 Introduction

The algorithms for association rules extraction were originally developed to find the logical association between variables with the same status. The predictive association rules search the associations between items that characterize a dependent attribute. We are in a supervised learning framework.

Basically, the algorithm is not really modified. Exploration is just limited to itemsets that include the dependent variable. The computation time is then reduced. Two components of Tanagra are dedicated to this task: these are ASSOC SPV and SPV RULE TREE ASSOC. They are available in the ASSOCIATION tab.

Compared to conventional approaches, the components of Tanagra have an additional specificity: we can specify the class value ("dependent variable = value") that we want to predict. The advantage is that we can set precisely the parameters of the algorithm, directly in relation to the characteristics of data. This is crucial for instance when the prior probabilities of the dependent variable values are very different.

We had already described the SPV TREE ASSOC component elsewhere¹. But it was in the context of multivariate characterization of groups of individuals (from a clustering algorithm for instance). We compare it to the GROUP CHARACTERIZATION component. In this tutorial, we will compare the behavior of TREE ASSOC SPV and SPV RULE ASSOC during a prediction task. We will put forward their shared properties, the problems that they can handle, and their differences. SPV ASSOC RULE, which supplies original rule interestingness measures² ("test value" indicator), has the ability to simplify the rule base.

2 Dataset

We use a modified version of the GERMAN CREDIT dataset³. It describes the characteristics of customers. We discretized the quantitative variables. The file is available on line (<u>http://eric.univ-lyon2.fr/~ricco/tanagra/fichiers/credit_assoc.xls</u>).

CLASS is the dependent variable. We want to characterize the "good" customers (CLASS = GOOD). We therefore have two settings to set before the calculations: we indicate that CLASS is the TARGET variable; among the values of CLASS, we want to characterize the GOOD value.

3 Creating a diagram

Importing the dataset. First, we define a new project (FILE / NEW) and we import the dataset. Tanagra can handle directly the Excel file format (XLS). We select the CREDIT_ASSOC.XLS.

Tanagra shows that 17 variables and 1000 examples were loaded.

¹ <u>http://data-mining-tutorials.blogspot.com/2008/11/supervised-association-rules.html</u>

² <u>http://data-mining-tutorials.blogspot.com/2009/02/interestingness-measures-for.html</u>

³ <u>http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)</u>

TANAGRA 1.4.30							
File Diagram Window	Help						
D New							
😅 Open ng dia Cho	pose your dataset and start dow	vnload					
Save							
Save as	Diagram title :				_		
Close	Default title						
Exit	Data mining diagram file na	ame :					
	D:\DataMining\Databases_	_for_mining\da	taset_for_soft_c	lev_and_tompa			
	Dataset (*.txt,*.arff,*.xls) :			<u> </u>			
	D:\DataMining\Databases_	for_mining\da	taset_for_soft_c	lev_and_compa	2		
			🝸 Tanagra				×
			Regarder dans :	assoc mr and spy	v assoc	- G 👌 📂 🖽	-
			0	Nom Date	e de m Tyre	Taille	
			2	Credit assoc.xls	cucina iye		
			Emplacements			The set Freedback and set MA	0.5
	1 1	Comp	recents			Taille : 269 Ko	Crosott Excel
Data visualization	Statistics	Nonparamet				Date de modification : 04/	02/2009 19:05
Feature construction	Feature selection	Regre	Bureau				
PLS	Clustering	Spv le	_				
Spv learning assessment	Scoring	Assoc					
Correlation scatterplot	t 📝 Scatterplot	🚉 View d	Maison				
Export dataset	🖉 Scatterplot with label	🚺 View r					
			Outratu				
			Ordinateur		*		
				Nom du fichier :	credit_assoc.xls	•	Ouvrir
				Types de fichiers :	Excel File (97 & 20	DO) 👻	Annuler

Defining the type of attributes. We insert the DEFINE STATUS component from the shortcut into the tool bar in order to define the type of attributes: CLASS is the TARGET attribute, the others are the INPUT.



We click on the VIEW menu. Here is the output of Tanagra.

TANAGRA 1.4.30 - [Define	status 1]								×	
💇 File Diagram Compo	nent Window	Help						- 8	×	
D 📽 🔲 🙀										
Default title		Attr	tibute	Target	Input	Illustrative				
□	oc.xls)	checking	status	rarget	mput	inusu auve				
Define status 1		duration	status		yes					
Parameters	s	credit his	tory	2	yes	_				
Europe to		creat_ns	scory		yes					
Execute	+>	purpose			yes					
view	1	cavings s	tatue		ves	-				
		employme	nt		vec					
		personal	status		ves					
		other par	rties		ves					
		property	magnitude	_	ves				≡	
		other payment plans			ves					
		housing			ves					
		existing of	credits	-	yes	-				
		job			yes	-				
		own_telep	ohone		yes					
		foreign_w	orker		yes	-				
		class		yes	-	-				
]								Ŧ	
			Cor	nponer	its			-		
Data visualization	Statisti	cs	Nonparam	netric s	tatistic	s In:	stance selection	Feature construction	•	
Feature selection	Regressi	ion	Facto	rial ana	lysis		PLS	Clustering		
Spv learning	Meta-spv le	arning	Spv learn	ing asse	ssmen	t	Scoring	Association		
Correlation scatterplot	🖉 Scatterplo	t with label								
Export dataset	🔛 View datas	set								
Scatterplot	🔥 View multi	ple scatterp	lot							

4 The SPV ASSOC TREE component

The SPV ASSOC TREE component extracts association rules from dataset. The procedure uses internally a search tree but the outputs are rules.

4.1 Choosing the class value

We insert SPV ASSOC TREE into the diagram. We click on the PARAMETERS contextual menu. We must set the class value that we want to characterize. Into the dialog box, we set CLASS VALUE = GOOD.



We validate and we click on the VIEW menu in order to execute the calculations.

TANAGRA 1.4.30 - [Spv A	ssoc Tree 1]								
File Diagram Compo	onent Window Help						-		
I 🚔 🔚 🎎									
Default title			Resul	lts					
Dataset (credit_ass Define status 1	ee 1 "class	es ss" is "good" IF							
N° Antecedent Length Support Confidence Lift									
	1 fc	oreign_worker=yes - checking_status	=no checking	2	0.335 (0.00)	0.879 (0.00)	1.256 (0.00)		
	2 of	ther_parties=none - checking_status	=no checking	2	0.331 (0.00)	0.887 (0.00)	1.268 (0.00)		
	3 cl	necking_status=no checking		1	0.348 (0.00)	0.883 (0.00)	1.262 (0.00)		
		Components							
Data visualization	Statistics	Nonparametric statistics	Instance	selecti	ion Fe	ature constr	uction		
Feature selection	Regression	Factorial analysis	P	LS		Clustering	g		
Spv learning	Meta-spv learning	Spv learning assessment	Sco	oring		Associatio	on		
A priori Asso A priori MR Spv A priori PT 🖶 Spv	oc Outlier Assoc Rule Assoc Tree								

The component generates 3 rules. They are displayed in the lower part of the window. The support, confidence and lift are provided. When the numerical indicators are in parentheses, it means that the indicator was calculated on a test sample, which was not used during the learning phase. Here, we have systematically zero. All observations belong to the learning set.

4.2 Partitioning the dataset in a "train set" and "test set"

According to the approach usually implemented in a supervised learning framework, we can subdivide the dataset in a train set and a test set. The first is used during the extraction of the rules from data; the second is used for the assessment of the rules. We know that the measures computed on the test set give an honest estimate of their interestingness. In order to subdivide the dataset, we click again on the PARAMETERS menu. We set the LEARNING SET RATIO to 0.6 i.e. 60 percent of the dataset are used as a train set, 40 percent as a test set.



We validate this setting and we click on the VIEW menu.

TANAGRA 1.4.30 - [Spv Assoc Tr	ree 1]							x
💇 File Diagram Component	Window Help						_	5 ×
D 📽 🔲 👫								
Default title □-1111 Dataset (credit_assoc.xls)	"class" is "g	ood" IF						*
🖶 🚰 Define status 1	N°	Antecedent	Support	Confidence	Lift			
Spv Assoc Tree 1	1 foreign_wor checking	ker=yes - other_parties=none - ch	ecking_status=no 3		0.343 (0.28)	0.900 (0.86)	1.301 (1.20)	
	2 foreign_wor	ker=yes - checking_status=no che	ecking 2		0.353 (0.31)	0.887 (0.87)	1.282 (1.22)	
	3 other_parti	cking 2		0.352 (0.30)	0.902 (0.86)	1.304 (1.21)		
	4 other_paym	ent_plans=none - housing=own	2		0.428 (0.47)	0.763 (0.78)	1.103 (1.09)	=
	5 checking_st	atus=no checking	1		0.363 (0.33)	0.890 (0.87)	1.286 (1.22)	
								-
		Componer	nts			1		
Data visualization	Statistics	Nonparametric statistics	Instance selection	F	Feature const	truction		
Feature selection	Regression	Factorial analysis	PLS		Clusteri	ng		
Spv learning A	Meta-spv learning	Spv learning assessment	Scoring		Associat	ion		
🔋 A priori 🛛 🕄 A priori P	T Spv Assoc	Rule						
🔋 A priori MR 🔹 Assoc Out	tlier 👌 🗄 Spv Assoc	Tree						
							.,4111141.;	

We note that we obtain more rules without modifying the other settings. It is an artifact. It means that some rules are very near to the support (rule n°1), confidence (rule n°4) and lift threshold values.

We note also that we have more indications about the reliability of the rules now. The measures computed on the test set are less optimistic.

4.3 Ranking the rules

The settings can heavily influence the results. Some important rules can be hidden by inappropriate threshold values. An alternative approach is to set less restrictive settings to obtain more rules and organize them in order to highlight the most informative rules.

The available parameters are:

- SUPPORT defines the minimum support of the extracted rules ;
- CONFIANCE defines the minimum confidence ;
- MAX CARD ITEMSETS defines the maximum length ;
- LIFT defines the minimum lift.

Once the rules extracted, we need to organize so that the most interesting appear first. Tanagra can rank the rules according to one of the numerical criteria above. In this tutorial, we rank them according to the lift criterion.

We click on the PARAMETERS menu.

		Association	tree spv			1
TANAGRA 1.4.30 - [Define status Tanage Diagram Component Image: Spring status Image: Spring status Image: Spring status Image: Spring status Image: Spring status Image: Spring status Image: Spring status Image: Spring status Image: Spring status Image: Spring status Image: Spring status Image: Spring status Image: Spring status Image: Spring status Image: Spring status	1] Window Help Attribute checking_status duration rameters	Param	eters Sorting Support: Confidence : ix card itemsets : Lift : earning set ratio : Class value :	0.10 0.75 4 2 1.1 0.6 good		
Vie	employment personal_status other_parties property_magnitude other_payment_plan	Association	tree spv eters Sorting Sort rules by	OK Cancel	Help	
Data visualization Feature selection Spv learning M BA priori BA priori PT A priori MR SAssoc Outt	Statistics No Regression Sp leta-spv learning Sp Spv Assoc Rule ier E Spv Assoc Tree		no sort rule length rule confidence			construction ustering sociation
				OK Cancel	Help	

We obtain the following results.

File Diagram Component Window Help Default tile Results Class" is "good" IF Rules "class" is "good" IF N Antecedent Length Support Confidence Lift other_payment_plans=none - checking_status=no checking - credit_history=critical/other a cher_payment_plans=none - checking_status=no checking - credit_history=critical/other a other_payment_plans=none - checking_status=no checking - purpose=radio/tv b other_payment_plans=none - checking_status=no checking - purpose=radio/tv checking_status=no checking - status=no checking - on_telephone=yes a other_payment_plans=none - checking_status=no checking - on_telephone=yes a other_payment_plans=none - checking - on_telephone=yes a other_payment_plans=none - checking - on_telephone=yes a other_payment_plans=none - checking - on_telephone=yes b checking_status=no checking - on_telephone=yes c checking_status=no checking - purpose=radio/tv c checking_status=no checking - purpose=radio/tv c checking_status=no checking - status=no checking - status=no c checking_status=no checking - status=no c checking_status=no c	TANAGRA 1.4.30 - [Spv Assoc Tre	ee 1]								X			
Image: Sevent and the status 1 Results Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1 Image: Sevent and the status 1	💇 File Diagram Component	Window Help							_	8 X			
Results Results Results Lefault ible Results Results Class" is "good" IF Note free tatus 1 Lefault ible Class" is "good" IF Note free tatus 1 Image: Class" is "good" IF Note free tatus 1 Image: Class" is "good" IF Note free tatus 1 Image: Class" is "good" IF Note free tatus 1 Image: Class" is "good" IF Note free tatus 1 Image: Class" is "good" IF Note free tatus 1 Image: Class" is "good" IF Image: Class 1 Image: Class 1 <th cols<="" th=""><th>D 🗳 📕 🎎</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></th>	<th>D 🗳 📕 🎎</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>	D 🗳 📕 🎎											
Image: Spr Assoc Tree 1 Rules Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1	Default title			Results						*			
Image: Spr Assoc Tree 1 Kules Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr Assoc Tree 1 Image: Spr A	🖃 🥅 Dataset (credit_assoc.xls)												
Image: Spv Assoc Tree 1 "class" is "good" IF Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Assoc Tree 1 Image: Spv Ison Tree 1 Image: Spv Ison Tree 1	🖹 🚼 Define status 1	Rules	Rules										
N* Antecedent Length Support Confidence Lift 1 other_payment_plans=none - checking_status=no checking - credit_history=critical/other 3 0.135 (0.13) 0.988 (0.96) 1.428 (1.35) 2 housing=own - checking_status=no checking - credit_history=critical/other 3 0.127 (0.11) 0.974 (0.92) 1.409 (1.29) 3 other_payment_plans=none - checking_status=no checking - purpose=radio/tv 3 0.118 (0.11) 0.973 (0.91) 1.406 (1.28) 4 other_payment_plans=none - checking_status=no checking - purpose=radio/tv 3 0.107 (0.01) 0.970 (0.93) 1.402 (1.20) 5 existing_credits=one - checking_status=no checking - own_telephone=yes 3 0.107 (0.07) 0.970 (0.86) 1.428 (1.20) 6 checking_status=no checking - purpose=radio/tv 2 0.123 (0.12) 0.961 (0.92) 1.389 (1.29) Components Components Data visualization Scoring Association Spv learning Spv learning Scoring	E Spy Assoc Tree 1												
N* Antecedent Length Support Confidence Lift 1 other_payment_plans=none - checking_status=no checking - credit_history=critical/other 3 0.135 (0.13) 0.988 (0.96) 1.428 (1.35) 2 housing=own - checking_status=no checking - credit_history=critical/other existing 3 0.127 (0.11) 0.974 (0.92) 1.409 (1.29) 3 other_parties=none - checking_status=no checking - purpose=radio/tv 3 0.118 (0.11) 0.970 (0.93) 1.402 (1.31) 4 other_payment_plans=none - checking_status=no checking - purpose=radio/tv 3 0.107 (0.07) 0.970 (0.83) 1.402 (1.20) 5 existing_credits=one - checking_status=no checking - own_telephone=yes 3 0.107 (0.07) 0.970 (0.81) 1.402 (1.20) 6 checking_status=no checking - own_telephone=yes 3 0.107 (0.07) 0.970 (0.81) 1.402 (1.20) 10 checking_status=no checking - purpose=radio/tv 2 0.123 (0.12) 0.961 (0.92) 1.389 (1.29) 11 other_payment_plans=none - checking_status=no checking - purpose=radio/tv 2 0.123 (0.12) 0.961 (0.92) 1.3402 (1.20) 12 other_payment_plansenoe statistics Instanc		class is go	0a ir										
1 other_payment_plans=none - checking_status=no checking - credit_history=critical/other 3 0.135 (0.13) 0.988 (0.96) 1.428 (1.35) 2 housing=own - checking_status=no checking - credit_history=critical/other existing 3 0.127 (0.11) 0.974 (0.92) 1.409 (1.29) 3 other_parties=none - checking_status=no checking - purpose=radio/tv 3 0.118 (0.11) 0.973 (0.91) 1.402 (1.21) 4 other_payment_plans=none - checking_status=no checking - purpose=radio/tv 3 0.107 (0.10) 0.970 (0.93) 1.402 (1.21) 5 existing_credits=one - checking_status=no checking - own_telephone=yes 3 0.107 (0.07) 0.970 (0.86) 1.402 (1.21) 6 checking_status=no checking - numpose=radio/tv 2 0.123 (0.12) 0.961 (0.92) 1.389 (1.29) Components Components Data visualization Statistics Nonparametric statistics Instance selection Feature construction Feature selection Regression Factorial analysis PLS Clustering Spv learning Meta-spv learning Spv learning assessment Scoring Association Scoring Association Spv learning<		N°	Ar	Support	Confidence	Lift							
2 housing=own - checking_status=no checking - credit_history=critical/other existing 3 0.127 (0.11) 0.974 (0.92) 1.409 (1.25 3 other_parties=none - checking_status=no checking - purpose=radio/tv 3 0.118 (0.11) 0.973 (0.91) 1.406 (1.25 4 other_payment_plans=none - checking_status=no checking - purpose=radio/tv 3 0.107 (0.10) 0.970 (0.93) 1.402 (1.31 5 existing_credits=one - checking_status=no checking - purpose=radio/tv 3 0.107 (0.07) 0.970 (0.86) 1.402 (1.20 6 checking_status=no checking - purpose=radio/tv 2 0.123 (0.12) 0.961 (0.92) 1.389 (1.29 Components Data visualization Statistics Nonparametric statistics Instance selection Regression Factorial analysis PLS Clustering Spv learning Spv learning assessment Scoring Association Statistion		1 other_payme existing	nt_plans=none - checking_status		0.135 (0.13)	0.988 (0.96)	1.428 (1.35)						
3 other_parties=none - checking_status=no checking - purpose=radio/tv 3 0.118 (0.11) 0.973 (0.91) 1.406 (1.26 4 other_payment_plans=none - checking_status=no checking - purpose=radio/tv 3 0.107 (0.01) 0.970 (0.93) 1.402 (1.20 5 existing_credits=one - checking_status=no checking - own_telephone=yes 3 0.107 (0.07) 0.970 (0.86) 1.402 (1.20 6 checking_status=no checking - purpose=radio/tv 2 0.123 (0.12) 0.961 (0.92) 1.389 (1.29 Components Data visualization Statistics Nonparametric statistics Instance selection Feature construction Feature selection Regression Factorial analysis PLS Clustering Spv learning Meta-spv learning Spv learning assessment Scoring Association Spv learning Spv learning		2 housing=own	 checking_status=no checking - 	credit_history=critical/other	existing 3		0.127 (0.11)	0.974 (0.92)	1.409 (1.29)				
4 other_payment_plans=none - checking - status=no checking - purpose=radio/tv 3 0.107 (0.010) 0.970 (0.93) 1.402 (1.31) 5 existing_credits=one - checking_status=no checking - own_telephone=yes 3 0.107 (0.07) 0.970 (0.93) 1.402 (1.20) 6 checking_status=no checking - purpose=radio/tv 2 0.123 (0.12) 0.961 (0.92) 1.389 (1.29) Components Data visualization Statistics Nonparametric statistics Instance selection Feature construction Feature selection Regression Factorial analysis PLS Clustering Spv learning Meta-spv learning Spv learning assessment Scoring Association Clustering Spv learning Meta-spv learning		3 other_parties	s=none - checking_status=no che	ecking - purpose=radio/tv	3		0.118 (0.11)	0.973 (0.91)	1.406 (1.28				
5 existing_credits=one - checking_status=no checking - own_telephone=yes 3 0.107 (0.07) 0.970 (0.86) 1.402 (1.20 (0.92) 1.389 (1.29 (0.92) 1.389 (4 other_payme	nt_plans=none - checking_status	s=no checking - purpose=radio/1	:v 3		0.107 (0.10)	0.970 (0.93)	1.402 (1.31				
6 checking_status=no checking - purpose=radio/tv 2 0.123 (0.12) 0.961 (0.92) 1.389 (1.25 Components Components Data visualization Statistics Nonparametric statistics Instance selection Feature construction Feature selection Regression Factorial analysis PLS Clustering Spv learning Meta-spv learning Spv learning assessment Scoring Association Clustering Spv learning		5 existing_crea	dits=one - checking_status=no ch	necking - own_telephone=yes	3		0.107 (0.07)	0.970 (0.86)	1.402 (1.20				
Components Data visualization Statistics Nonparametric statistics Instance selection Feature construction Feature selection Regression Factorial analysis PLS Clustering Spv learning Meta-spv learning Spv learning assessment Scoring Association		6 checking_sta	tus=no checking - purpose=radio	o/tv	2		0.123 (0.12)	0.961 (0.92)	1.389 (1.29	-			
Data visualization Statistics Nonparametric statistics Instance selection Feature construction Feature selection Regression Factorial analysis PLS Clustering Spv learning Meta-spv learning Spv learning assessment Scoring Association				Components			1						
Regression Factorial analysis PLS Clustering Spv learning Meta-spv learning Spv learning assessment Scoring Association <td>Data visualization</td> <td>Statistics</td> <td>Nonparametric statistics</td> <td>Instance selection</td> <td>Feature construct</td> <td>on</td> <td>Feature</td> <td>e selection</td> <td></td> <td></td>	Data visualization	Statistics	Nonparametric statistics	Instance selection	Feature construct	on	Feature	e selection					
Spv learning assessment Scoring Association	Regression F	Factorial analysis	orial analysis PLS Clustering Spv learning Meta-spv lea										
	Spv learning assessment	Scoring	Association										
I 25 A priori II 25 A priori PT II Spv Assoc Rule	🕄 A priori 🛛 🕄 A priori PT	Spv Assoc Ru	ıle										
🕄 A priori MR 🛛 📜 Assoc Outlier 🗧 Spv Assoc Tree	🕄 A priori MR 🛛 🕄 Assoc Outl	lier 🗄 Spv Assoc Tr	ee										
								at attack	latar addition addi	dati .etil			

Tanagra

We obtain 327 rules. The most interesting according to the lift are in the upper part of the table. The values into the brackets, computed on the test sample, give an honest estimate of the rule performance.

5 SPV ASSOC RULE component

SPV ASSOC RULE extracts also predictive association rules. It is very similar to the previous component, but: (a) the computations are organized differently; (b) it provides more interestingness measure; (c) it can simplify the rule base.

5.1 Setting the parameters

We insert SPV ASSOC RULE into the diagram. We click on the PARAMETERS menu. We specify the class value (CLASS = GOOD).

TANAGRA 1.4.30 - [Spv Assoc	Tree 1]		
🕎 File Diagram Component	t Window Help		_ 8 ×
🗅 📽 🖬 🎎		Spv Assoc Rule Parameters	
Default title		Parameters	
Define status 1	Rules		
→ E Spv Assoc Tree 1	1	Support: 0.33	
Spv Assoc Rule 1	"class" is "good"	Confidence : 0.75	
Execut	te pr_payment_plans=	Max card itemsets : 4	idence Lift 8 (0.96 1.428 (1.35)
View	2 credit_history=cretica 3 other_parties=none - c	Learning set ratio : 1 Repetition : 1	4(0.92 1.409(1.29) 3(0.91 1.406(1.28)
Data visualization	Statistics Nonpa	Class value : good	ction
Feature selection Spv learning	Regression Fac Meta-spv learning Spv learning	VT Boundary Filtering : 2	
🕄 A priori 🕄 A priori	PT Spv Assoc Rule		
A priori MR Assoc O	Dutlier 🗄 Spv Assoc Tree	OK Cancel Help	
			- 14 4.4 - 14 4.4 - 14 4.4 4.4

Other parameters are provided. They are in relation to the new measures described on line (<u>http://data-mining-tutorials.blogspot.com/2009/02/interestingness-measures-for.html</u>). There is especially the test value. These parameters are:

- REPETITION defines the number of replication during the Monte Carlo procedure;
- VT Boundary Filtering defines the minimum test value for the extracted rules. We use the Z (HYP) measure for the comparison.

We click on VIEW.

Tanagra

TANAGRA 1.4.30 - [Spv As	ssoc Rule 1]												, 🗆 🗙
💇 File Diagram Compo	nent Window I	Help											_ 8 ×
D 📽 🔲 🙀													
Default title													
Dataset (credit_asso Define status 1	oc.xls) ee 1 ile 1	Rules e	evaluatio	rules									
		N° An	técédent	Conséquent	n	n[A]	n[C]	n[A^C]	Support	Confiance	Lift	Leverage	Importar
		1 "checki checkir	ing_status=no ng"	"class=good"	1000	394	700	348	0.3480	0.8832	1.2618	0.0722	0.4191
All rules Rules evaluation													
		N° An	técédent	Conséquent	n	n[A]	n[C]	n[A^C]	Support	Confiance	Lift	Leverage	Importan
		1 "checki checki	ing_status=no ng"	"class=good"	1000	394	700	348	0.3480	0.8832	1.2618	0.0722	0.4191
		"checki 2 checkii "other_	ing_status=no ng" - _parties=none"	"class=good"	1000	373	700	331	0.3310	0.8874	1.2677	0.0699	0.4107
		"checki 3 checkii "foreigi	ing_status=no ng" - n_worker=yes"	"class=good"	1000	381	700	335	0.3350	0.8793	1.2561	0.0683	0.3995
				Component	ts								
Data visualization	Statistic	s	Nonparame	tric statistics		Inst	tance	e select	ion	Feature	constr	uction	
Feature selection	Regressio	ession Factorial analy					F	PLS		Clustering			
Spv learning	Meta-spv lea	arning	Spv learning	g assessment			Sc	oring		Ass	ociatio	n	
Correlation scatterplot	📴 Export data	aset	🖉 Scatt	erplot		ii	Sca	tterplot	with lab	el 🔛	/iew da	taset	Ĕ.:
													Þ

The rules are displayed in 2 separate parts: **ALL RULES** shows all the extracted rules, there are 3 here; **FILTERED RULES** shows the simplified rule base i.e. *after the elimination of redundant rules*.

Indeed, we note that the rules n°2 and n°3 do not give more information compared to the first rule (n°1). The computation is based only on a logical criterion. We assume that all the rules have the same weight.

5.2 Extracting more rules

As above, we can generate more rules by modifying some parameters. The simplification module allows highlighting the most important information afterwards.

We click on the PARAMETERS menu. We set LEARNING SET RATIO = 0.6 and SUPPORT = 0.1

Tanagra

TANAGRA 1.4.30 - [Spv Assoc Rule 1]				
Tile Diagram Component Window	Help			_ 8 ×
	Sp	v Assoc Rule Parameters		
Default title Dataset (credit_assoc.xls)	Filtered	Parameters		•
Spv Assoc Tree 1		Support :	0.1	
Execute	"checking"	Max card itemsets : Lift :	4	8 0.0722 0.4191
	All rules	Learning set ratio : Repetition :	1	
	L	Class value :	good 🔻	
Data visualization Statis Feature selection Regres Spv learning Meta-spv	tics Nong ision F learning Spv (VT Boundary Filtering :	2	truction ng ion
Correlation scatterplot Export d	ataset 🕼		OK Cancel Help	Jataset

We validate and we click on the VIEW menu. The indicators computed on the test set are displayed in the second part of the table.

TANAGRA 1.4.30 - [Spv Asso	oc Rule 1]												- 🗆 🗙
🝸 File Diagram Compone	nt Window I	Help											_ & ×
D 📽 🖪 🏭													
Default title Default title Default title Dataset (credit_assoc. Define status 1 Define status 1 Spv Assoc Tree Spv Assoc Rule	.xls) - 1 1	Rules RULES Filtered = 4 rules Rules evaluation									H		
		"credit 1 existin "other	_history=critical/other g " - _parties=none"	"class=good"	n 600	n[A] 162	n[C] 415	n[A C] 137	0.2283	0.8457	Lift 1.2227	0.0416	0.2870
		"credit 2 existin "other_	_history=critical/other g " - _payment_plans=none"	"class=good"	600	147	415	129	0.2150	0.8776	1.2687	0.0455	0.3293
		3 "credit existin	on=lo_1_year" - :_history=critical/other g "	"class=good"	600	69	415	66	0.1100	0.9565	1.3829	0.0305	0.3752
		4 ^{"check} checki	ing_status=no ng"	"class=good"	600	245	415	218	0.3633	0.8898	1.2865	0.0809	0.4722
		•											4
	Charles of		C	omponents									
Data visualization	Statistic	5	Nonparametric stati	sucs I	nstai	nce si	elect -	ion	Feat	Churchenel	lction		
Feature selection	Kegressio	ssion Factorial analysis		s		PLS				Clustering			
Correlation scatterplot	Export data	aset	ient	<u>1</u> 29	catte	erplot	t with la	bel	Associatio	taset	🤃 View multi		
•													Þ
													đ

```
Tanagra
```

The complete rule base contains 81 rules. We have less than SPV TREE ASSOC because VT BOUNDARY FILTERING also limits the number of rule: a rule is accepted if and only if [Z (HYP)> VT BOUNDARY FILTERING]. If we set VT BOUNDARY FILTERING 0, we obtained 327 rules, like to SPV TREE ASSOC.

After removing the redundant rules, we have only 4 rules. The interpretation of the results is simplified.

6 Conclusion

In this tutorial, we presented two components of Tanagra for the extraction of predictive association rules. They differ in the strategy used to overcome the problem of the abundance of rules inherent to the extraction algorithm: SPV TREE ASSOC offers the possibility to organize the rules according to a numerical criterion chosen by the user; SPV RULE ASSOC uses a simplification procedure.