Association Rule Software (ARS) is a basic tool which extracts association rules from attribute-value datasets (categorical or binary attributes). It is distributed with the SIPINA package which includes: a tool for the supervised learning framework, especially the decision tree induction (SIPINA RESEARCH); a tool for the linear regression (REGRESS); and thus, ARS for the association rule mining.

ARS encodes automatically the categorical attributes in dummy variables. If you want use a continuous attributes, you must discretize them before.

This tutorial describes shortly the use of the Association Rule Software (ARS).

## 1. Importing a dataset

First, we must import the dataset. ARS uses its proper file format (.FDM) which is optimised for the I/O processing. But we can also import dataset from other file format, especially tabdelimited text file format. We show here how to import this kind of data file.

We click on the FILE / OPEN and we select the Text Files format into the dialog box. We select the "banque.txt" data file which is available into the "Dataset" subfolder of the installation directory.

🍓 Association rule softwa	are - [Learning set	editor]				
🍪 File Edit Data	🕲 Ouvrir				X	- 8 ×
Selected attribute	Regarder dans :	ipina_datasets	•	← 🗈 💣 📰▼		
	Ca.	Nom	*	Modifié le	Туре	
	Elt-	agaricus_expan	ded.txt	14/12/2003 17:52	Document	
	récents	Autos.txt		09/01/2004 13:25	Document	
		banque.txt		10/10/2003 09:45	Document	
		cigarettes.txt		17/10/2004 09:32	Document	
	Bureau	ionosphere.TXT		03/01/2004 13:00	Document	
		iris.txt		06/03/2004 20:05	Document	
		nursery.txt		28/02/2004 14:40	Document	
	Bibliothèques	vote. IXI		08/11/2006 10:52	Document	
Learning Paramete	Ordinateur					
		•	III		+	
Selected example	Réseau	Nom du fichier :	banque.txt	•	Ouvrir	
		Types de fichiers :	Text file format (*.TXT)	▼ .	Annuler	
	jeaning			Attributes . 1	LXamples . 1	1.
						11.

Then, we specify the organization of the text file (tab delimited, first row corresponds to the name of the attributes).

	100	aitfor	aan	onfont	habit	rouonu	-
	Age	SICIAM	CSP	enranc	nabit	revenu	F
	mature	marie	cadre_moyen	zero	locataire	tranche_2	I
	jeune	marie	cadre_moyen	sup_ou_eg_eg2	locataire	tranche_4	
	jeune	marie	cadre_moyen	zero	locataire	tranche_3	
	mature	marie	cadre_moyen	zero	locataire	tranche_3	1
	jeune	marie	cadre_moyen	sup_ou_eg_eg2	locataire	tranche_3	1
	mature	marie	cadre_moyen	inf_a_2	locataire	tranche_4	
	ancien	marie	cadre_moyen	inf_a_2	locataire	tranche_2	
	mature	marie	cadre_moyen	inf_a_2	locataire	tranche_2	
	ancien	celibataire	cadre_moyen	zero	locataire	tranche_1	
- Specifica	ations		Delimiters				
			Tabs				
First row is name of attributes							
First column is label of examples			🔿 Space				
🔽 Cate	gorical attribute		C Other 🔶	<b>-</b>			

## 2. Specifying the role of attributes

Association Rules mining highlights the co-occurrence between items (values of attributes). So, we must specify the attributes that we want use in our analysis.

We click on the ANALYSIS / SELECT ATTRIBUTES menu. Into the dialog box, the available attributes are showed on the left side. We select those that we want to include into the analysis by clicking on the add **>** button.

🔮 File Edit Data Statistics	Analysis Window Help	_ 8 ×
Selected attributes	Select attributes for analysis     1   2   3   4   csp   enfant   habit   revenu   port_action   demand   accord     10   11   12   13   14   15   16   17	1     demand       Consommi       travaux       consommi       travaux       consommi       travaux       consommi       voiture       consommi       voiture       consommi       voiture       consommi       voiture       consommi       voiture       voiture
Selected examples	18 Clear selection	consommi consommi
	Editing	198

Note: The analysis can be performed on categorical attributes. But, we can use also continuous ones. In this case, they are interpreted as binary variables i.e. the values lower

or equal than 0 equates to the absence of the attribute, the values upper than 0 equates to the presence. We can also discretize the continuous attributes by selecting the appropriate method in the STATISTICS / TRANSFORM / CONTINUOUS ATTRIBUTES menu.

## 3. Launching the analysis

To launch the analysis, we click on the ANALYSIS / START ANALYSIS menu. A dialog box appears, asking the learning parameters. Standard settings for association rule mining are proposed:

- Support: the minimum support for the extracted rules.
- Confidence: the minimum confidence of the rules.
- Max Rule Length: the maximum length (the cardinality of the itemset) of the rules.
- Max Consequent Length: the maximum length of the consequent of the rules.

We note that ARS can extract rules with a consequent having more than one item. Of course, the length of the consequent is lower than the length of the rule.

🍪 Association rule software - [Lea	arning set ed	itor]	1200	1.2					
👶 File Edit Data Statistics	Analysis	Windo	ow Help						_ 8 ×
C 🗳 🔒									
Selected attributes		Age	sitfam	csp	enfant	habit	revenu	port_action	demand 🔺
Age	1	mat	Association Rule P	arameters		an opposite the	- X )	non	consomm
sitfam	2	jeur	Association nulle n	arameters	-	Concession of the local division of the loca		oui	travaux
CSP enfant	3	jeur	Learning Measure	es				oui	consomm
habit	4	mat						non	travaux
revenu	5	jeur				0.1		non	consomm
port_action demand	6	mat	Support	::		0.1		oui	voiture
accord	7	anc	Confida			0.0		non	consomm
	8	mat	Connue	nce.		0.0		oui	voiture
	9	anc	May rul	e lenath :		4 🛃		non	consomm
	10	mat					non	consomm	
	11	mat	Max co	nsequent leng	jth:	1 🛃		non	voiture
Learning Parameters	12	jeur						non	consomm
	13	anc						non	voiture
	14	mat				. 1		non	voiture
	15	mat	<u> </u>	Ok j	Cano	el		oui	voiture
	16	and						oui	voiture
	17	ancie	n celibataire	cadre_moye	rzero	locataire	tranche_1	non	voiture
	18	matur	re marie	cadre_moye	rinf_a_2	locataire	tranche_2	non	consomm
Selected examples	19	matur	re separe	employe	inf_a_2	locataire	tranche_3	oui	consommi 🚽
198 examples - learning set	<b>₹</b>	Longia	n lealibataire	londro movo	7070	lecataira	Itrancho 1	loui	4
	Editing N	IEW.FI	DM			Attribu	utes : 9	Examples : 1	98 _//
								_	

At the same time, we observe that, because we do not set explicitly the instances to treat during the analysis, ARS uses automatically all the available instances (198 cases).

We click on the OK button to start the learning process. The rules are displayed into a new visualization window.

<u>^</u>		_										
🚭 Analysis 🛛 Rule managen	nent	Win	idow Help								- 6	
🗅 🚔 📕												
Selected attributes		<b>3</b> I										
Age sitfam	٦ E	Id	Antecedent	Consequent	Length	Support	Confi	Recall	F-mea	Lift	Convi	T
csp	•	1	revenu=tranche_1 & sitfam=celibataire	Age=ancien	3	0.1010	0.8000	0.2703	0.4040	2.1405	2.8125	-
enfant	Ξ.	2	habit=locataire & revenu=tranche_1 & sitfam=celibataire	Age=ancien	4	0.1010	0.8333	0.2703	0.4082	2.2297	3.2500	
revenu		3	demand=voiture	accord=oui	2	0.2980	0.8939	0.4126	0.5646	1.2378	2.3800	
port_action	- I.	4	port_action=oui	accord=oui	2	0.4495	0.8318	0.6224	0.7120	1.1517	1.6063	
demand	-	5	revenu=tranche_3	accord=oui	2	0.1970	0.8478	0.2727	0.4127	1.1739	1.6800	
Learning Parameters		6	Age=ancien & demand=voiture	accord=oui	3	0.1313	0.8387	0.1818	0.2989	1.1613	1.5400	
Min support = 0.1000		7	Age=mature & demand=voiture	accord=oui	3	0.1313	0.9630	0.1818	0.3059	1.3333	4.0600	
Min confidence = 0.8000 Max rule length = 4		8	Age=mature & enfant=inf_a_2	accord=oui	3	0.1768	0.8537	0.2448	0.3804	1.1820	1.7200	
Max cons. length = 1		9	Age=mature & port_action=oui	accord=oui	3	0.2172	0.9149	0.3007	0.4526	1.2668	2.7440	
		10	Age=mature & revenu=tranche_2	accord=oui	3	0.1818	0.8000	0.2517	0.3830	1.1077	1.3160	
		11	Age=mature & sitfam=marie	accord=oui	3	0.1970	0.8478	0.2727	0.4127	1.1739	1.6800	
Selected examples		12	csp=cadre_moyen & demand=voiture	accord=oui	3	0.2525	0.8929	0.3497	0.5025	1.2363	2.3200	
198 examples - learning set		₹			·-						,	
D examples - idle	ľ		Frequent items	ets = 655, Rules = 6	546, Calculat	ion time = 1	16 ms.					_

We have the description of the rule (antecedent  $\rightarrow$  consequent). We have also some standard indicators which measure the interestingness of the rules. We can sort the rules according these various criteria e.g. according the lift criterion in the screenshot bellow.

Association rule software -	[Asso	ociatio	on rule viewer (Beta = 1.00)]	and the second division of the second divisio			1.0	1	23		Ξ Σ	3
🍅 Analysis 🛛 Rule managem	ent	Win	dow Help								- 8	×
D 🚔 🖬												
Selected attributes		er e										
Age		Id	Antecedent	Consequent	Length	Support	Confi	Recall	F-mea	Lift 🔻	Convi	$\overline{}$
csp		632	Age=ancien & habit=locataire & revenu=tranche_1	sitfam=celibataire	4	0.1010	0.9091	0.3226	0.4762	2.9032	5.4800	
enfant	Ξ	630	Age=ancien & revenu=tranche_1	sitfam=celibataire	3	0.1010	0.9091	0.3226	0.4762	2.9032	5.4800	
revenu		633	csp=cadre_moyen & enfant=zero & revenu=tranche_1	sitfam=celibataire	4	0.1162	0.8519	0.3710	0.5169	2.7204	3.9730	
port_action		634	enfant=zero & habit=locataire & revenu=tranche_1	sitfam=celibataire	4	0.1162	0.8519	0.3710	0.5169	2.7204	3.9730	
demand	*	631	enfant=zero & revenu=tranche_1	sitfam=celibataire	3	0.1212	0.8276	0.3871	0.5275	2.6429	3.5392	
Learning Parameters		2	habit=locataire & revenu=tranche_1 & sitfam=celibataire	Age=ancien	4	0.1010	0.8333	0.2703	0.4082	2.2297	3.2500	
Min support = 0.1000		1	revenu=tranche_1 & sitfam=celibataire	Age=ancien	3	0.1010	0.8000	0.2703	0.4040	2.1405	2.8125	
Min confidence = 0.8000 Max rule length = 4		345	accord=non & habit=locataire & sitfam=celibataire	enfant=zero	4	0.1010	1.0000	0.1869	0.3150	1.8505	10.1200	
Max cons. length = 1		332	accord=non & sitfam=celibataire	enfant=zero	3	0.1061	1.0000	0.1963	0.3281	1.8505	10.5800	
1		636	revenu=tranche_4	sitfam=marie	2	0.1010	1.0000	0.1802	0.3053	1.7838	9.6800	
1		339	revenu=tranche_1 & sitfam=celibataire	enfant=zero	3	0.1212	0.9600	0.2243	0.3636	1.7764	6.2100	
Selected examples		360	habit=locataire & revenu=tranche_1 & sitfam=celibataire	enfant=zero	4	0.1162	0.9583	0.2150	0.3511	1.7734	5.9800	-
198 examples - learning set		•	· · · · · · · · · · · · · · · · · · ·								+	
0 examples - idle	Î		Frequent itemse	ts = 655, Rules = 64	6, Calculat	ion time = '	16 ms.					_
												//

We can copy the rules into the clipboard (to paste them into a spreadsheet for instance) or save them in an output file (text file format) [RULE MANAGEMENT menu].

## 4. Conclusion

This short introduction describes how to use ARS in an association rules mining process.