WEKA Introduction

WEKA software

WEKA homepage

http://www.cs.waikato.ac.nz/ml/weka/



WEKA software

- Data mining software written in Java (distributed under the GNU Public License).
- Used for research, education, and applications.
- Comprehensive set of data preprocessing tools, learning algorithms and evaluation methods.
- Graphical user interfaces including data visualization.
- Environment for comparing learning algorithms.
- Book and development versions.

ARFF format file

- @relation data-relation-name
- @attribute attribute1 real
 @attribute attribute2 { nominal1, nominal2 }
 @attribute class { class1, class2 }
- @data
 1, nominal1, class1
 2, nominal2, class2
 3, ?, class1
- 4, nominal 1, class 2

WEKA GUI Chooser

 The Weka GUI Chooser (class weka.gui.GUIChooser) provides a starting point for launching Weka's main GUI applications and supporting tools. If one prefers a MDI ("multiple document interface") appearance, then this is provided by an alternative launcher called "Main" (class weka.gui.Main).

 The GUI Chooser consists of four buttons one for each of the four major Weka applications—and four menus.

WEKA GUI Chooser

Weka GUI Chooser	- 0 ×
Program Visualization Tools Help	
	Applications
WEKA	Explorer
The University of Waikato	Experimenter
Waikato Environment for Knowledge Analysis Version 3.6.9	KnowledgeFlow
(c) 1999 - 2013 The University of Waikato Hamilton, New Zealand	Simple CLI

WEKA Experimenter

- An environment for performing experiments and conducting statistical tests between learning schemes.
- The Weka Experiment Environment enables the user to create, run, modify, and analyse experiments in a more convenient manner than is possible when processing the schemes individually. For example, the user can create an experiment that runs several schemes against a series of datasets and then analyse the results to determine if one of the schemes is (statistically) better than the other schemes.

WEKA Experimenter

🕝 Weka Experiment Environment				
Setup Runs Analyse				
Experiment Configuration Mode:	 Simple 	Advanced		
Open	Save	New		
Results Destination				
ARFF file 👻 Filename:		Browse		
Experiment Type	Iteration Control			
Cross-validation	 Number of repetitions: 			
Number of folds:	Oata sets first			
Classification Cassification	 Algorithms first 			
Datasets	Algorithms			
Add new Edit selected Dele	ete selected Add new	Edit selected Delete selected		
Use relative paths				
Up Down	Load options	Save options Up Down		
	Load options			
	Notes			

WEKA Knowledge Flow

- This environment supports essentially the same functions as the Explorer but with a drag-anddrop interface. One advantage is that it supports incremental learning.
- The KnowledgeFlow presents a data-flow inspired interface to WEKA. The user can select WEKA components from a tool bar, place them on a layout canvas and connect them together in order to form a knowledge flow for processing and analyzing data. At present, all of WEKA's classifiers, filters, clusterers, loaders and savers are available in the KnowledgeFlow along with some extra tools.

WEKA Knowledge Flow

🕝 Wel	ka KnowledgeFlow E	nvironment						-		x
	DataSources DataSir	ks Filters Classifie	ers Clusterers Ass	ociations Eval	uation Visualization					k
	DataSources									2
	star i de la constancia	\$	\$					S		
	Arff Loader	C45 Loader	CSV Loader	Database Loader	LibSVM Loader	Serialized InstancesLoader	SVMLight Loader	TextDirectory Loader	XRFF Loader	
Knowle	edge Flow Layout								,	-
										• H
•					III					<u> </u>
Status	Log									
Compo		Parameters		Time	Status					
[Knowle	edgeFlow]			0:0:13	Welcome to the Weka	Knowledge Flow				

WEKA Simple CLI

- Provides a simple command-line interface that allows direct execution of WEKA commands for operating systems that do not provide their own command line interface.
- The Simple CLI provides full access to all Weka classes, i.e., classifiers, filters, clusterers, etc., but without the hassle of the CLASSPATH (it facilitates the one, with which Weka was started).
- It offers a simple Weka shell with separated commandline and output.

WEKA Simple CLI

X 🕝 SimpleCLI Welcome to the WEKA SimpleCLI Enter commands in the textfield at the bottom of the window. Use the up and down arrows to move through previous commands. Command completion for classnames and files is initiated with <Tab>. In order to distinguish between files and classnames, file names must be either absolute or start with '.\' or '~/' (the latter is a shortcut for the home directory). <Alt+BackSpace> is used for deleting the text in the commandline in chunks. > help Command must be one of: java <classname> <args> [> file] break kill cls history exit help <command>

- An environment for exploring data with WEKA (the rest of this documentation deals with this application in more detail).
- At the very top of the window, just below the title bar, is a row of tabs. When the Explorer is first started only the first tab is active; the others are greyed out. This is because it is necessary to open (and potentially pre-process) a data set before starting to explore the data.

🚱 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes Visualize	
Open file Open URL Open DB Genera	rate Undo Edit Save
Filter Choose None	Apply
Current relation Relation: None Instances: None Attributes: None	Selected attribute Name: None Type: None Missing: None Distinct: None Unique: None
Attributes	
All None Invert Pattern	
	✓ Visualize All
Remove	
Status Welcome to the Weka Explorer	Log 💉 O

• The tabs are as follows:

- Preprocess: Choose and modify the data being acted on.
- Classify: Train and test learning schemes that classify or perform regression.
- > Cluster: Learn clusters for the data.
- > Associate: Learn association rules for the data.
- Select attributes: Select the most relevant attributes in the data.
- Visualize: View an interactive 2D plot of the data.

Preprocess Cluster Associate Select attributes Open file Open URL Open DB Generate Undo Edit Save Filter Choose None Apply Current relation Relation: dum-data Instances: 3333 Attributes: 21 Name: state Type: Noninal All None Invert Pattern Selected attribute No. Name Invert Pattern No. Name A S5 All None Invert Pattern 1 state S S 2 account length S CA 3 area_code G G 7 number Generate Cass: churn (Nom) Visualize All 6 voice_mail_plan Generate Cass: churn (Nom) Visualize All 11 total_day_charge Istate Generate Generate 12 total_day_charge Istate Generate Cass: churn (Nom) Visualize All Status Code Code Cass Log X 0	🕢 Weka Explorer					
Filter Apply Current relation Relation: churn-data Instances: 3333 Attributes: 21 All None All None Invert Pattern No. Name 1 state 2 account_length 3 area_code 4 phone_number 2 code 6 voice_mail_plan 7 number_scales 9 total_day_ninutes 9 total_day_charge 11 total_day_ninutes 9 total_day_ninutes 12 total_day_ninutes 13 total_day_charge 11 total_day_ninutes 12 total_day_charge 13 total_ava_charge 14 total_day_charge 15 total_day_charge 111 total_ava_charge 12 total_day_charge 13 total_ava_charge 13 total_ava_charge 14 total_ava_charge 15 total_day_charge 16 voice_mail_obar 17 total_ava_charge 18 total_ava_charge 19	Preprocess Classify Cluster Associate Select attributes Visualize					
Choose None Apply Current relation Relation: churn-data Instances: 3333 Attributes: 21 Attributes Name: state Type: Noninal All None Invert Pattern No. Name 1 AK 2 account length 3 AR 3 area_code 7 7 4 phone_number Goo 66 3 area_code 7 74 4 phone_number Goo 66 7 number_ymail_messages Goo 66 9 btata[day_charge Goo 66 11 btata[day_charge Goo 66 11 btata[day_charge Goo 66 12 total_day_charge Goo 66 13 total_day_charge Goo 66 14 btata_day_charge Goo Goo 15 total_day_charge Goo Goo 16 total_day_charge Goo Goo 11 btata_day_charge Goo Goo 12 total_day_charge Goo Goo 13 total_day_charge Goo Goo	Open file Open URL Open DB Genera	ate Undo	Edit Save			
Current relation Selected attribute Relation: durn-data Instances: 3333 Attributes Name: state Type: Nominal All None Invert No. Name Selected attribute No. Name Selected attribute No. Name Output: Selected attribute No. Name None 2 Account_length Selected attribute 2 Account_length Selected attribute 3 Area_code Selected attribute 4 All None 1 AK S2 2 Account_length Selected attribute 3 Area_code Selected attribute 4 AZ Gdd 9 total_day_call Selected attribute 10 total_day_calls Selected attribute 11 total_day_calls Selected attribute 12 total_day_calls Selected attribute 11 total_day_calls Selected attribute 12 total_day_calls Selected attribute	Filter					
Relation: churn-data Instances: 3333 Attributes: All None Invert Pattern No. Name 1 state 2 account_length 3 area_code 4 phone_number 5 international_plan 6 voice_mail_plan 7 number_winal_messages 8 total_day_calls 10 total_day_calls 11 total_eve_calls 12 total_eve_calls 13 total_eve_calls 14 total_eve_calls 15 total_eve_calls 16 total_eve_calls 17 18 total_eve_calls 19 total_eve_calls 10 total_eve_calls 11 total_eve_calls 12 total_eve_calls 13 total_eve_calls <td< th=""><td>Choose None</td><td></td><td>Apply</td></td<>	Choose None		Apply			
Instances: 3333 Attributes: 21 Attributes Attributes Attributes Attributes Attributes Attributes Attributes Attributes Attributes Attributes Attributes No. Label Count 1 AK 52 2 AL 80 3 AR 55 4 AZ 64 5 CA 3 44 6 CO 66 7 CT 74 Class: churn (Nom) Visualize All Visualize All Status	Current relation	Selected attribute				
All None Invert Pattern No. Name 1 AK 52 No. Name 3 AR 55 4 AZ 64 5 GCO 66 7 Contentional plan 6 6 Voice_mail_plan Visualize All 7 number_vmail_messages 6 8 total_day_charge 1 11 total_day_charge 1 12 total_day_charge 1 13 total_exe_calis 1 12 total_exe_calis 1 13 total_exe_charne 1						
All None Invert Pattern No. Name 2 AL 80 1 state 6 64 64 2 account length 7 74 74 3 area_code 7 74 74 4 phone_number 6 CO 66 5 international_plan 7 74 74 6 voice_mail_plan 7 Visualize All 7 number_vmail_messages 6 6 8 total_day_calls 6 6 10 total_day_calls 6 6 12 total_day_calls 6 6 12 total_day_calls 6 6 12 total_day_calls 6 6 13 total_day_calls 6 6 13 total_day_calls 6 6 13 total_day_calls 6 6 14 total_day_calls 6 6 15 total_day_calls 6 6 <t< th=""><th>Attributes</th><th>No. Label</th><th></th></t<>	Attributes	No. Label				
No. Name 1 state 2 account_length 3 area_code 3 area_code 3 area_code 4 AZ 6 CO 6 voice_mail_plan 7 number_ymail_messages 8 total_day_calls 10 total_day_calls 11 total_day_calls 12 total_day_calls 13 total_eve_calls 13 total_eve_calls 13 total_eve_charge	All None Invert Pattern					
No. Name 1 state 2 account_length 2 account_length 3 area_code 3 area_code 4 AZ 6 CO 6 Voice_mail_plan 7 number 7 number 8 total_day_minutes 9 total_day_calls 10 total_day_calls 11 total_eve_minutes 12 total_eve_calls 13 total_eve_calls 13 total_eve_calls 7 Remove Status						
1 state 2 account_length 3 area_code 4 phone_number 5 international_plan 6 voice_mail_plan 7 number_vmail_messages 8 total_day_calls 10 total_day_calls 11 total_eve_minutes 12 total_eve_calls 13 total_eve_calls Status Status	No. Name					
2 account_length 3 area_code 4 phone_number 5 international_plan 6 voice_mail_plan 7 number_vmail_messages 8 total_day_minutes 9 total_day_calls 10 total_que_calls 12 total_eve_minutes 13 total_eve_calls 13 total_eve_calls Status status	1 🗖 state					
4 phone_number 5 international_plan 6 voice_mail_plan 7 number_vmail_messages 8 total_day_minutes 9 total_day_calls 10 total_day_charge 11 total_eve_minutes 12 total_eve_calls 13 total_eve_charge Remove		6 CO	66			
Sinternational_plan Image: Class: churn (Nom) Visualize All 6 voice_mail_plan Image: Class: churn (Nom) Visualize All 7 number_vmail_messages Image: Class: churn (Nom) Visualize All 8 total_day_minutes Image: Class: churn (Nom) Visualize All 9 total_day_calls Image: Class: churn (Nom) Visualize All 10 total_day_charge Image: Class: churn (Nom) Visualize All 11 total_day_charge Image: Class: churn (Nom) Visualize All 12 total_eve_calls Image: Class: churn (Nom) Visualize All Remove Image: Class: churn (Nom) Visualize All Status Image: Class: churn (Nom) Visualize All	3 area_code	7 CT	74 🔻			
S international_plan 6 voice_mail_plan 7 number_vmail_messages 8 total_day_minutes 9 total_day_calls 10 total_day_charge 11 total_eve_minutes 12 total_eve_calls 13 total_eve_calls Status		Class: churn (Nom)	Visualize All			
7 number_vmail_messages 8 total_day_minutes 9 total_day_calls 10 total_day_charge 11 total_eve_minutes 12 total_eve_calls 13 total_eve_charge	5 international_plan					
8 total_day_minutes 9 total_day_calls 10 total_day_charge 11 total_eve_minutes 12 total_eve_calls 13 total_eve_charge Status						
9 total_day_calls 10 total_day_charge 11 total_eve_minutes 12 total_eve_calls 13 total_eve_charge Remove						
10 total_day_charge 11 total_eve_minutes 12 total_eve_calls 13 total_eve_charge Remove						
11 total_eve_minutes 12 total_eve_calls 13 total_eve Remove		1 B. A. B. B.	. In a state			
13 total eve charge Remove	11 total_eve_minutes	_ իստիս _ իվելտե				
Image:						
Status	13 total eve charge					
	Remove					
OK Log X 0	Status					
	ок		Log 💉 ⁰			

• Selecting a Classifier

At the top of the classify section is the Classifier box. This box has a text field that gives the name of the currently selected classifier, and its options. The Choose button allows you to choose one of the classifiers that are available in WEKA.

There are four test modes available:

- > Use training set: The classifier is evaluated on how well it predicts the class of the instances it was trained on.
- Supplied test set: The classifier is evaluated on how well it predicts the class of a set of instances loaded from a file. Clicking the Set... Button brings up a dialog allowing you to choose the file to test on.
- Cross-validation: The classifier is evaluated by cross-validation, using the number of folds that are entered in the Folds text field.
- Percentage split: The classifier is evaluated on how well it predicts a certain percentage of the data which is held out for testing. The amount of data held out depends on the value entered in the % field.
- Once the classifier, test options and class have all been set, the learning process is started by clicking on the Start button. While the classifier is busy being trained, the little bird moves around.

Preprocess Classify Cluster Associate Select attributes Visualize Classifier Classifiers <liclassifiers< li=""> Classifiers Classifiers</liclassifiers<>	📀 Weka Explorer		
MSP NBTree RandomForest RandomTree REPTree SimpleCart Filter Remove filter Close	Preprocess Classify Cluster Associate Select attributes Classifier Classifier dassifiers dassifiers	r output	
	J48graft LADTree LMT MSP NBTree RandomForest RandomTree REPTree SimpleCart Filter Remove filter Close		Log x0

Weka Explorer							_ 0 <mark>_ X</mark>
	Select attributes Visua	liza				_	
Classifier	Select attributes Visua	alize					
Choose J48 -U -M 2							
Test options	Classifier output						
O Use training set	Junuary	-					
							_
Supplied test set Set	Correctly Clas Incorrectly Cl			3068 265		92.0492 7.9508	-
Oross-validation Folds 10	Kappa statisti		liscances	0.66	39	7.9500	•
Percentage split % 66	Mean absolute			0.09			
More options	Root mean squa	red error		0.26	5		
Plore options	Relative absol	ute error		38.32	15 %		
	Root relative			75.28	04 %		
(Nom) churn 👻	Total Number o	f Instance	3	3333			
Start Stop	=== Detailed A	ccuracy By	Class ===				
Result list (right-click for options)							
17:08:50 - trees.J48				Precision			
		0.671	0.037				0.829
	Weighted Avg.		0.287				
	=== Confusion Matrix ===						
a b < classified as							
	324 159 a = True						
	106 2744 b = False						
	•						4
Status							
OK						Log	x 0

- When training is complete, several things happen. The Classifier output area to the right of the display is filled with text describing the results of training and testing. A new entry appears in the Result list box.
- The text in the Classifier output area is split into several sections:
 - Run information: A list of information giving the learning scheme options, relation name, instances, attributes and test mode that were involved in the process.
 - Classifier model (full training set): A textual representation of the classification model that was produced on the full training data.
 - Summary: A list of statistics summarizing how accurately the classifier was able to predict the true class of the instances under the chosen test mode.
 - Detailed Accuracy By Class: A more detailed per-class break down of the classifier's prediction accuracy.
 - Confusion Matrix: Shows how many instances have been assigned to each class. Elements show the number of test examples whose actual class is the row and whose predicted class is the column.

WEKA Manual

More can be found in WEKA Manual:

http://www.cs.waikato.ac.nz/ml/weka/docu mentation.html

