

WEKA Waikato Environment for Knowledge Analysis

Attribute Selection

Prof. Pietro Ducange





Attribute Selection (1)

- Attribute selection involves searching through all possible combinations of attributes in the data to find which subset of attributes works best for prediction.
- Attribute selection consists basically of two different types of algorithms:
- evaluator determines the merit of single attributes or subsets of attributes
- **search algorithm** the search heuristic





Attribute Selection (2)

🕢 Weka Explorer	
Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Attribute Evaluator weka ChiSquaredAttributeEval ClassifierSubsetEval ConsistencySubsetEval CostSensitiveSubsetEval CostSensitiveSubsetEval CostSensitiveSubsetEval CostSensitiveSubsetEval GainRatioAttributeEval FilteredAttributeEval GainRatioAttributeEval InfoGainAttributeEval OneRAttributeEval PrincipalComponents ReliefFAttributeEval SymmetricalUncertAttributeEval SymmetricalUncertAttributeEval WrapperSubsetEval Filter Remove filter Close 	Weka Explorer Image: Construction of the second
Filter Remove filter Close	Close
Status OK	Status OK





Attribute Selection: First Example (1)

Weka Explorer Preprocess Classify Cluster Attribute Evaluator Choose CfsSubsetEval Search Method Choose BestFirst -D 1 -N 5 Attribute Selection Mode Attribute Selection Mode Attribute Selection Folds 10	Evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them
Start Stop Result list (right-click for options) Status	Searches the space of attribute subsets by greedy hill climbing augmented with a backtracking facility. Forward or Backward Search can be selected

in in



=== Attribute Selection on all input data ===

Search Method:

Best first. Start set: no attributes Search direction: forward Stale search after 5 node expansions Total number of subsets evaluated: 12 Merit of best subset found: 0.887

Attribute Subset Evaluator (supervised, Class (nominal): 5 class): CFS Subset Evaluator Including locally predictive attributes

Selected attributes: 3,4 : 2 petallength petalwidth





Attribute Selection: Second Example (1) 🛃 Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Evaluates the worth of an attribute by Attribute Evaluator measuring the information gain InfoGainAttributeEval Choose. with respect to the class Search Method Choose. Ranker -T -1.79769313486231575308 -N -: Attribute Selection Mode Attribute selection output Ose full training set === Run information === Folds 10 Cross-validation Evaluator: weka.attributeSelection.InfoGainAttri Search:weka.attributeSelection.Ranker -T -1.7976931 Relation: iris (Nom) class Instances: 150 Attributes: 5 Start sepallength -Result list (right-click for options) eenelwidth Ranks attributes by their individual evaluations. A selection threshold can be fixed





Attribute Selection: Second Example (2)

=== Attribute Selection on all input data ===

Search Method: Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 5 class): Information Gain Ranking Filter

Ranked attributes: 1.418 3 petallength 1.378 4 petalwidth 0.698 1 sepallength 0.376 2 sepalwidth

Selected attributes: 3,4,1,2 : 4





Attribute Selection: Wrapper Method (2)

Preprocess Classify Cluster Associate Select attribute Evaluator Choose WrapperSubsetEval -B weka.classifiers.refees.ZeroR	ct attributes Visualize Forecast	Evaluates attributes by	the worth of a set of using a specific classi
earch Method Choose BestFirst -D 1 -N 5			
Attribute Selection Mode Seed 1	t		
(Nom) class ¢ Start Stop		a qui CanaricObia tEditor	
esult list (right-click for options)	weka.attributeSelection.WrapperSubsetEval About WrapperSubsetEval: Evaluates attribute sets by using a learning s	cheme.	More Capabilities
	IRClassValue classifier Choose J	48 -C 0.25 -M 2	
(doNotCheckCapabilities False		\$
	evaluationMeasure Default: accu folds 5	racy (discrete class); RMSE (numeric	class)
	seed 1 threshold 0.01		
AITA	Open Save	. ОК	Cancel

Attribute Selection: Wrapper Method (2)

=== Attribute Selection on all input data ===

Search Method: Best first. Start set: no attributes Search direction: forward Stale search after 5 node expansions Total number of subsets evaluated: 11

Merit of best subset found: 0.947

Attribute Subset Evaluator (supervised, Class (nominal): 5 class): Wrapper Subset Evaluator Learning scheme: weka.classifiers.trees.J48 Scheme options: -C 0.25 -M 2 Subset evaluation: classification accuracy Number of folds for accuracy estimation: 5

Selected attributes: 4 : 1 petalwidth





N SUD

1343

Attribute Selection as a Filter

00	Weka Explorer
Preprocess Classify Cluste	r Associate Select attributes Visualize
Open file Open URL Open DB	Generate Undo Edit Save
weka uteS	Selected attribute
MultiFilter	Name: sepallength Type: Numeric Missing: 0 (0%) Distinct: 35 Unique: 9 (6%)
supervised attribute	Statistic Value Minimum 4.3
AddClassification atte	rn Maximum 7.9 Mean 5.843 StdDev 0.828
 NominalToBinary PLSFilter instance unsupervised 	Class: class (Nom) Visualize All
	30 34 28 25 16 10
Remove	
Filter Remove filter Close	4.3 6.1
De la companya	10

SUP SUP

Attribute Selection as a Filter (setting parameters)

Reprocess Classify Cluster Associate Select attributes Visualize open file Open URL Open DB Generate Undo er Choose AttributeSelection - E "weka.attributeSelection.CfsSubsetEval " - S "weka.attributeSelection.BestFirst - D 1 - N 5" rrent relation Relation: iris weka.filters.supervised.attributeSelection More attributes A supervised attribute filter that can be used to select More attributes. Capabilities evaluator Choose 1 sepailength search Choose ffssubsetEval 3 petallength search Choose more 4 petallength search Choose ffssubsetEval S class Open Save OK Cancel Weka.attributeSelection.cfsSubsetEval More Genabilities gabilities Isolation indexes the worth of a subset of attributes by Capabilities Gapabilities Isolation Indexes the worth of a subset of attributes by Capabilities gapabilities Isolation Indexes Indexes Gapabi	0 0	Weka Explorer	
Open file Open URL Open DB Generate Undo ser Choose AttributeSelection -E "weka.attributeSelection.CfsSubsetEval" - S "weka.attributeSelection.BestFirst -D 1 -N 5" The set of the se		Preprocess Classify Cluster Associate Select attribute	s Visualize
er Choose AttributeSelection - E "weka.attributeSelection.cfsSubsetEval " - S "weka.attributeSelection.BestFirst - D 1 - N 5" Trent relation: iris Stances: 150 All None Augervised attribute filter that can be used to select A gupervised attribute filter that can be used to select A gupervised attributes Papawidth 3 opeallength 4 opealwidth 5 oclass Open Save OK Cancel About CfsSubsetEval Abou	Open file Open URL	Open DB Generate	Undo
Attributesetettion - E weka.attributesetettion.esstrist = 0 1 + 8 3 rent relation Relation: iris stances: 150 ributes Ali Name About About About About About Capabilities evaluator Choose CfsSubsetEval Sepalkingth	Choose Astribute Selection 5 "webs attribute Selection	ortion Of Subset Fuel " S "wells attribute Selection PartFirst D 1 N F"	
Name All Name 1 epailength 2 3 petallength 3 2 epailength 3 2 3 etallength 3 3 etallength 3 3 etallength 3 2 etallength 3 2 etallength 3 2 etallength 3 etallength 3 etallength 3 etallength 3 2 etallength 3 2 etallength 3 etallength search Choose BestFirst -D 1 -N 5 Open Save OK cancel Veka.attributeSelection.CfsSubsetEval kabout CfsSubsetEval : Evaluates the worth of a subset of attributes by considering the individual predictive ability of each hem. locallyPredictive True <td>Attributeselection - E weka.attributesele</td> <td>ction.clssubseteval -3 weka.attributeselection.bestrifst -D 1 -N 5</td> <td></td>	Attributeselection - E weka.attributesele	ction.clssubseteval -3 weka.attributeselection.bestrifst -D 1 -N 5	
ributes All None All None Asupervised attribute filter that can be used to select Asupervised attributes. Asupervised attributes. Capabilities evaluator Choose CfsSubsetEval search Choose BestFirst -D 1 -N 5 Open Save OK Cancel About CfsSubsetEval Subset of attributes by Capabilities Evaluates the worth of a subset of attributes by Capabilities Evaluates the worth of a subset of attributes by Capabilities LocallyPredictive True them. IocallyPredictive True them. Capabilities	Relation: iris Instances: 150	weka.gui.GenericObjectEditor weka.filters.supervised.attributeAttributeSelection	
Name 1 2 sepallength 2 sepallength 3 petallength 4 petallength 5 class Open Save OK Cancel Image: Sepallength Image: Save 6 weka.gui.GenericObjectEditor weka.attributeSelection.CfsSubsetEval More About CfsSubsetEval Katuates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. IocallyPredictive True missingSeparate False Open Save OK	All None	About A supervised attribute filter that can be used to select Attributes. Capabilities	
Open Save OK Cancel weka.gui.GenericObjectEditor weka.attributeSelection.CfsSubsetEval About CfsSubsetEval : More Evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. 34 locallyPredictive True ‡ missingSeparate False ‡	5. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class	evaluator Choose CfsSubsetEval search Choose BestFirst -D 1 -N 5	
weka.attributeSelection.CfsSubsetEval About CfsSubsetEval : More Evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Capabilities IocallyPredictive True ‡ missingSeparate False ‡		Open Save OK Cancel	
IocallyPredictive True + missingSeparate False + Open Save OK Cancel		weka.attributeSelection.CfsSubsetEval About CfsSubsetEval : Evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Capabilities	34
Remove Open Save OK Cancel		IocallyPredictive True ‡ missingSeparate False \$	
	Remo	Open Save OK Cancel	
	Y Z	11	

Attribute Selection as a Filter (results)

Preprocess Classify Cluster A	ssociate Select attributes Visualize]
Open file Open URL Open DB Gene	rate Undo Ed	it Save
Filter		
Choose AttributeSelection -E "weka.attributeSelection.CfsSub	osetEval " – S "weka.attributeSelection.Bes	stFirst -D 1 -N 5" Apply
Current relation Relation: iris-weka.filters.supervised.attribute.AttributeSelect	Selected attribute Name: petallength	Type: Numeric
Instances: 150 Attributes: 3	Missing: 0 (0%) Distinct: 43	Unique: 10 (7%)
Attributes	Minimum 1	
All None Invert Pattern	Maximum 6.9 Mean 3.7 StelDay 1.7	59
No. Name	1.7	04
2 petalwidth		
3 Class		
	Class: class (Nom)	\$ Visualize All
	50	13
		47
	34	
		16
	3	
Remove		
	1 3.95	6.9
otatus OK		Log 💦 x 0

Now we can switch to the classify module and perform a cross validation analysis....

...Is it a correct way to act?





Classification and Attribute Selection (1)

🕢 Weka Explorer		
Preprocess Classify Cluster Associate S	Select attributes Visualize	
Classifier		Dimensionality of training and
Choose AttributeSelectedClassifi	Fier -E "weka.attributeSelection.CfsSubsetEval" -S "weka.attributeSelection.BestFirst -D 1 -N 5" -W weka.classifiers.tre	s.148 CO. test data is reduced by
<u> </u>		attribute selection
Test options	Classifier output	
C Use training set		before being passed
C Supplied test set Set	sepalwi ^{d+b}	on to a classifier
	mean weka.gui.genericobjeccedicor	
C Cross-validation Folds 10	std. weka.classifiers.meta.AttributeSelectedClassifier	
Percentage split % 66	About	
More options	Dimensionality of training and test data is reduced by More	
	petalle attribute selection before being passed on to a classifier.	
(Nom) class	mean	
	std.	
Start Stop	weigh classifier Choose 348-C 0.25-M 2	
	preci debug Fake	
12:20:51 Januar IRk		
12:20:51 - Id2y.IDK 12:29:08 - Id2y.IBk	mean evaluator Choose CfsSubsetEval	
12:31:38 - bayes.NaiveBayes	std.	
	weigh search Choose BestFirst -D 1 -N 5	
	preci	
	Open Save OK Cancel	

Time taken to build model: 0 seconds



Classification and Attribute Selection (2)

Selected attributes: 3,4 : 2 petallength petalwidth

=== Evaluation on test split === === Summary ===

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	





Command Line Attribute Selection: An Example

Generating new training and test files

java weka.filters.supervised.attribute.AttributeSelection \

- -E "weka.attributeSelection.CfsSubsetEval " \
- -S "weka.attributeSelection.BestFirst -D 1 -N 5" \
- -b \
- -i <Training.arff> \
- -o <TrainingSel.arff> \
- -r <Test.arff> \
- -s <TestSel.arff>
- Generating and testing a classifier

java weka.classifiers.trees.J48 -t TrainingSel.arff –T TestSel.arff





Classification and Filtering Data (1)

💌 Weka Explorer		
Preprocess Classify Cluster Associate S	ielect attributes Visualize	
Classifier		
Choose FilteredClassifier - E "weka.	.filters. MultiFilter - F \"weka. filters. supervised.instance. Resample - B 0.0 - 5 1 - 7 50.0\" - F \"weka. filters. supervised.attribute. Discretize - R first-last\"" - W weka.classifiers.trees. J48 C 0.2	5-M 2
Test options	Classifier output	
C Use training set	🖉 weka gui Geperic Object Editor	
C Supplied test set Set	webs describer meta Elitered Classifier	ased exclusively
C Cross-validation Folds 10		
	on the training data and te	est instances
(• Percentage split % job	A Class for running an arbitrary classifier on data that has been More will be processed by well be processed by	ov the
More options	Capabilities	
		eir structure
(Nom) class	e classifier Choose J48 -C 0.25 -M 2)
Start Stop		
	debug False	
Result list (right-click for options)	filter Choose MultiFilter - F "weka, filters, supervised, instance, Resample -B 0.0 -:	
12:28:51 - lazy.IBk		
12:29:06 - Id2y.10k 12:31:38 - bayes.NaiveBayes	n Open Save OK Cancel	
16:01:29 - meta.AttributeSelectedClassifier		
	@attribute pe 🥡weka.qui.GenericObjectEditor	
	Gattribute pe weka.filters.MultiFilter	
	@attribute cl:	
	Applies several litters successively. More	
	Capabilities Chosen Personelle B 0 0 - S 1 - 7 5 - 0 dd	
	Classifier No	
	J48 pruned tri debug False	
	filters 2 weka.filters.Filter	
	petalwidth <=	
	petalwidth > 1 Open Save OK Cancel	
	petalwidt	
Æ	petallength <= 4.9: Iris-versicolor (48.0/1.0)	
	<pre>petallengtn > 4.9 petallength <= 1 5: Tris-virginica (3.0)</pre>	
	petalwidth > 1.5: Iris-versicolor (3.0/1.0)	
	$1 \text{natalwidth} > 1 ? \cdot \text{Trig-wirgining} (A6 0/1 0) \qquad 16$	
	10	

Exercise II

Perform the classification by using the three different meta classifiers (select a classification algorithm and three different attribute selection methods) and the Pima Diabets dataset with a 5-fold cross validation.

Which is the best attribute selection method?



