

# Ελληνοποίηση του εγχειριδίου χρήσης για το πακέτο RWeka

Ημερομηνία έκδοσης εγχειριδίου: 19/02/15

Version: 0.4-24

*Περιγραφή:* το παρόν εγχειρίδιο αποτελεί μία σύντομη περιγραφή της διεπαφής της R για το λογισμικό Weka (Version 3.7.12), το οποίο περιέχει μία συλλογή αλγορίθμων μηχανικής μάθησης για ζητήματα εξόρυξη γνώσης υλοποιημένους σε Java, περιλαμβάνοντας παράλληλα εργαλεία για προεπεξεργασία, κατηγοριοποίηση, οπισθοδρόμηση, συσταδοποίηση, εξαγωγή κανόνων συσχέτισης και οπτικοποίηση δεδομένων. Το πακέτο RWeka περιέχει τον κώδικα για την υλοποίηση της διεπαφής, ενώ το Weka jar ανήκει σε ξεχωριστό πακέτο της R, το RWekajars. Για περισσότερες πληροφορίες, οι ενδιαφερόμενοι καλούνται να επισκεφθούν τη σελίδα <http://www.cs.waikato.ac.nz/ml/weka/>.

*Συμβατές εκδόσεις:* R ( $\geq 2.6.0$ )

*Εισαγωγή πακέτων:* RWekajars ( $\geq 3.7.12$ ), rJava ( $\geq 0.6-3$ ), graphics, stats, utils, grid

*Προτεινόμενα πακέτα:* partykit ( $\geq 0.8.0$ ), mlbench, e1071

*Απαιτήσεις Συστήματος:* Java( $\geq 6.0$ )

*Τύπος άδειας:* GPL-2

*Συγγραφείς:* Kurt Hornik [aut, cre], Christian Buchta [ctb], Torsten Hothorn [ctb], Alexandros Karatzoglou [ctb], David Meyer [ctb], Achim Zeileis [ctb]

*Συντήρηση:* Kurt Hornik <Kurt.Hornik@R-project.org>

*Απαίτηση μεταγλώττισης:* όχι

*Αποθετήριο:* CRAN

*Ημερομηνία δημοσίευσης του πακέτου:* 28/01/2015 14:40:41

## Περιεχόμενα

Dot	Δημιουργία DOT Παρουσιάσεων.....	3
evaluate_Weka_classifier	Model Statistics for R/Weka Classifiers.....	4
predict_Weka_classifier	Model Predictions for R/Weka Classifiers .....	7
predict_Weka_clusterer	Class Predictions for R/Weka Classifiers.....	8
read.arff	Read Data from ARFF Files .....	9
Weka_associators	R/Weka_Associators .....	10
Weka_attribute_evaluators	R/Weka Attribute Evaluators .....	12
Weka_Classifiers	R/Weka Classifiers .....	14
Weka_classifier_functions	R/Weka Classifier Functions .....	15
Weka_classifier_lazy	R/Weka Lazy Learners .....	19
Weka_classifier_meta	R/Weka Meta Learners.....	21
Weka_classifier_rules	R/Weka Rule Learners .....	25
Weka_classifier_trees	R/Weka Classifier Trees .....	28
Weka_clusterers	R/Weka Clusterers .....	34
Weka_control	Control Weka Options.....	37
Weka_converters	R/Weka File Loaders and Savers.....	38
Weka_filters	R/Weka Filters.....	40
Weka_interfaces	R/Weka interfaces .....	42
Weka_stemmers	R/Weka Stemmers.....	46
Weka_tokenizers	R/Weka Tokenizers .....	47
WOW	Weka Option Wizard .....	48
WPM	Weka Package Manager.....	49
write.arff	Write Data into ARFF Files .....	51

## Περιγραφή

Συγγραφή σε γλώσσα αναπαράστασης DOT ενός αντικειμένου για επεξεργασία μέσω Graphviz.

## Χρήση

```
write_to_dot(x, con = stdout(), ...)  
## S3 method for class 'Weka_classifier'  
write_to_dot(x, con = stdout(), ...)
```

## Ορίσματα

x	αντικείμενο της R.
con	μία σύνδεση για τη συγγραφή της αναπαράστασης.
...	πρόσθετα ορίσματα για να αποδοθούν από και προς τις μεθόδους.

## Λεπτομέρειες

Το Graphviz ([www.graphviz.org](http://www.graphviz.org)) είναι ένα ανοικτού κώδικα πρόγραμμα για οπτικοποίηση γράφων παρέχοντας διάφορες επιλογές για την απεικόνιση και τη διάταξη των γνωστότερων γράφων, από τις οποίες η dot δημιουργεί ιεραρχικά δομημένες ή πολυεπιπέδες απεικονίσεις κατευθυνόμενων γραφημάτων, και για αυτόν τον λόγο αποτελεί την καταλληλότερη επιλογή για την οπτικοποίηση των διάφορων δένδρων κατηγοριοποίησης.

Χρησιμοποιώντας την dot μορφή, η αντιπροσώπευση του αρχείου 'foo.dot' μπορεί να μετασχηματιστεί σε PostScript ή σε κάποια άλλη μορφή παρουσίασης δεδομένων τύπου γράφων, με τη βοήθεια κάποιας βαριάντας του dot, ως εξής `-Tps foo.dot > foo.ps`.

Ορισμένοι ταξινομητές του Weka (π.χ. οι δενδρικοί εκμαθητές όπως ο J48 και ο M5P) υλοποιούν μία «σχηματίσιμη» διεπαφή παρέχοντας DOT αναπαραστάσεις των κατάλληλων και ταιριασμένων μοντέλων. Για τέτοιους ταξινομητές, η μέθοδος `write_to_dot` παράγει την αναπαράσταση της συγκεκριμένης σύνδεσης.

---

*evaluate\_Weka\_classifier*  
*Model Statistics for R/Weka Classifiers*

---

## Περιγραφή

Υπολογισμός στατιστικών δεικτών για την απόδοση μοντέλων που εκτιμώνται για την περιγραφή των δεδομένων από τους ταξινομητές του Weka.

## Χρήση

```
luate_Weka_classifier(object, newdata = NULL, cost = NULL,  
                        numFolds = 0, complexity = FALSE,  
                        class = FALSE, seed = NULL, ...)
```

## Ορίσματα

object	αντικείμενο ενός Weka_classifier
--------	----------------------------------

newdata	ένα προαιρετικό dataframe <sup>1</sup> στο οποίο αναζητούνται μεταβλητές για την τελική εκτίμηση. Αν παραλειφθεί ή εξισωθεί με το NULL, τα στιγμιότυπα εκπαίδευσης χρησιμοποιούνται αποκλειστικά.
cost	ένας τετραγωνικός πίνακας των συντελεστών ορθής και εσφαλμένης ταξινόμησης.
numFolds	το πλήθος των διαμερίσεων που θα χρησιμοποιηθούν στη διαδικασία του cross-validation.
complexity	επιλογή για την παρουσίαση ή μη στατιστικών δεικτών βασισμένων στην εντροπία.
class	επιλογή για την παρουσίαση ή μη στατιστικών δεικτών για τις κλάσεις.
seed	προαιρετική επιλογή σπόρου για τη διαδικασία του cross-validation.
...	πρόσθετα ορίσματα για να αποδοθούν από και προς τις μεθόδους.

## Λεπτομέρειες

Αυτή η συνάρτηση υπολογίζει και εξάγει μία μη-πλεονάζουσα συλλογή από στατιστικά απόδοσης τα οποία είναι κατάλληλα για την παρουσίαση του εκάστοτε μοντέλου. Εξ ορισμού, τα συγκεκριμένα στατιστικά υπολογίζονται βάση των δεδομένων εκπαίδευσης.

Το παρόν όρισμα ... υποστηρίζει μόνο τη λογική μεταβλητή `normalize` η οποία δίνει την εντολή στο Weka να κανονικοποιήσει τον πίνακα συντελεστών κόστους έτσι ώστε το κόστος κάθε σωστής κατηγοριοποίησης να ισούται με το μηδέν.

Αξίζει να αναφερθεί πως αν η κλάση της μεταβλητές είναι αριθμητική, μόνο ένα υποσύνολο των στατιστικών είναι διαθέσιμα. Τα ορίσματα `complexity` και `class` δεν εφαρμόζονται σε αυτήν την περίπτωση και επομένως αγνοούνται.

---

<sup>1</sup> Dataframe: αποτελεί μία από τις βασικές δομές δεδομένων της R.

## Έξοδος

Ένα αντικείμενο της κλάσης `Weka_classifier_evaluation` αποτελείται από τα ακόλουθα συστατικά στοιχεία:

<code>string</code>	χαρακτήρες, αποτελώντας τη συνένωση της αναπαράστασης των αλφαριθμητικών των στατιστικών απόδοσης.
<code>details</code>	διάνυσμα, περιέχοντας βασικά στατιστικά, όπως το ποσοστό των ορθά κατηγοριοποιημένων στιγμιοτύπων.
<code>detailsComplexity</code>	διάνυσμα, περιέχοντας στατιστικά βασισμένα στην εντροπία (αν επιλεγθεί)
<code>detailsClass</code>	πίνακας, περιλαμβάνοντας στατιστικά για τις κλάσεις, όπως ο δείκτης θετικά κατηγοριοποιημένων στιγμιοτύπων κ.ο.κ. για κάθε διαφορετικό επίπεδο της αντίστοιχης μεταβλητής (αν αυτό επιλεγθεί).
<code>confusionMatrix</code>	πίνακας, παρουσιάζοντας την τελική κατηγοριοποίηση των στιγμιοτύπων σε κάθε κλάση.

## Αναφορές

I. H. Witten and E. Frank (2005). Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition, Morgan Kaufmann, San Francisco.

## Παραδείγματα

```
## Use some example data.  
w <- read.arff(system.file("arff", "weather.nominal.arff",  
                           package = "RWeka"))  
  
## Identify a decision tree.  
m <- J48(play~., data = w)
```

m

```
## Use 10 fold cross-validation.
```

```
e <- evaluate_Weka_classifier(m, cost = matrix(c(0,2,1,0), ncol = 2),  
                              numFolds = 10, complexity = TRUE,  
                              seed = 123, class = TRUE)
```

e

```
summary(e)
```

```
e$details
```

---

*predict\_Weka\_classifier*

*Model Predictions for R/Weka Classifiers*

---

## Περιγραφή

Υπολογισμός εκτιμώμενων τιμών βασισμένων σε ταιριασμένα μοντέλα από ταξινομητές του Weka.

## Χρήση

```
## S3 method for class 'Weka_classifier'  
predict(object, newdata = NULL,  
        type = c("class", "probability"), ...)
```

## Ορίσματα

object	αντικείμενο το οποίο κληρονομεί τα χαρακτηριστικά του από την αντίστοιχη κλάση του εκάστοτε ταξινομητή του Weka.
newdata	ένα προαιρετικό dataframe στο οποίο αναζητούνται μεταβλητές για την τελική εκτίμηση. Αν παραλειφθεί ή εξισωθεί με το NULL, τα στιγμιότυπα εκπαίδευσης χρησιμοποιούνται αποκλειστικά.

type	διάνυσμα αλφαριθμητικών που καθορίζει αν τελικά απαιτείται πρόβλεψη των κλάσεων (αριθμητικές για οπισθοδρόμηση, κατηγορικές για κατηγοριοποίηση) ή αν αρκούν οι πιθανότητες για κάθε κλάση (δυνατότητα αυτής της επιλογής υπάρχει μόνο για τη διαδικασία της κατηγοριοποίησης). Αποτελεί όρισμα που μπορεί να παραλειφθεί.
...	πρόσθετα ορίσματα για να αποδοθούν από και προς τις μεθόδους.

## Έξοδος

Είτε ένα διάνυσμα με τις κλάσεις είτε ένας πίνακας με τις εκ των υστέρων πιθανότητες επιτυχίας ανά κλάση, όπου οι γραμμές αντιπροσωπεύουν τα διάφορα στιγμιότυπα και οι στήλες τις κλάσεις.

---

*predict\_Weka\_clusterer*  
*Class Predictions for R/Weka Classifiers*

---

## Περιγραφή

Πρόβλεψη των ταυτοτήτων των κλάσεων ή των ιδιοτήτων των μελών τους βασισμένη στις ταιριασμένες συστάδες από το Weka.

## Χρήση

```
## S3 method for class 'Weka_clusterer'  
Predict (object, newdata = NULL,  
        type = c("class_ids", "memberships"), ...)
```

## Ορίσματα

object	αντικείμενο το οποίο κληρονομεί τα χαρακτηριστικά του από την αντίστοιχη κλάση του εκάστοτε ταξινομητή του Weka.
--------	--



newdata	ένα προαιρετικό σύνολο δεδομένων στο οποίο αναζητούνται πιθανές προβλέψεις, οι οποίες μπορούν να βοηθήσουν στην πρόβλεψη των ιδιοτήτων της κάθε κλάσης. Αν παραλειφθεί ή εξισωθεί με το NULL, τα στιγμιότυπα εκπαίδευσης χρησιμοποιούνται αποκλειστικά για την πρόβλεψη των ταυτοτήτων των κλάσεων.
type	διάνυσμα αλφαριθμητικών που καθορίζει αν τελικά θα επιστραφεί η πρόβλεψη των ταυτοτήτων των κλάσεων ή των ιδιοτήτων αυτών . Αποτελεί όρισμα που μπορεί να παραλειφθεί.
...	πρόσθετα ορίσματα για να αποδοθούν από και προς τις μεθόδους.

## Λεπτομέρειες

Η πρόβλεψη των ιδιοτήτων των κλάσεων είναι εφικτή αν και μόνο αν ο συσταδοποιητής του Weka παρέχει μία `distributionForInstance` μέθοδο.

---

*read.arff*

*Read Data from ARFF Files*

---

## Περιγραφή

Διάβασμα δεδομένων από αρχεία τύπου Attribute – Relation – File Format (ARFF) τα οποία είναι συμβατά με το Weka.

## Χρήση

`read.arff(file)`

## Ορίσματα

file	διάνυσμα αλφαριθμητικών που περιέχει το όνομα του αρχείου ARFF από το οποίο αναμένεται να πραγματοποιηθεί η ανάγνωση των δεδομένων ή μία
------	--

σύνδεση, η οποία θα προσπελασθεί αν απαιτείται, και στην περίπτωση αυτή, θα τερματιστεί με την ολοκλήρωση της συνάρτησης.

## Έξοδος

Ένα dataframe το οποίο περιλαμβάνει τα δεδομένα από το ARFF αρχείο.

## Αναφορές

Attribute-Relation File Format <http://weka.wiki.sourceforge.net/ARFF>

## Προτάσεις για παρεμφερείς συναρτήσεις

`write.arff`

## Παραδείγματα

```
read.arff(system.file("arff", "contact-lenses.arff",  
                      package = "RWeka"))
```

---

*Weka\_associators*

*R/Weka\_Associators*

---

## Περιγραφή

Διεπαφές της R για τους αλγορίθμους εκμάθησης κανόνων συσχέτισης του Weka.

## Χρήση

`Apriori(x, control = NULL)`

`Tertius(x, control = NULL)`

## Ορίσματα

x	αντικείμενο της R για το οποίο θα εξαχθούν οι κανόνες συσχέτισης.
control	αντικείμενο της κλάσης <code>Weka_control</code> ή ένα διάνυσμα χαρακτήρων που περιγράφει επιλογές ελέγχου ή εξίσωση του με το <code>NULL</code> (default). Οι διαθέσιμες επιλογές μπορούν να προσπελαθούν είτε on-line χρησιμοποιώντας το <code>Weka Option Wizard (WOW)</code> είτε μέσω του εγγράφου τεκμηρίωσης του <code>Weka</code> .

## Λεπτομέρειες

Η επιλογή `Apriori` υλοποιεί μία βαριάντα του αλγόριθμου `Apriori`, ο οποίος μειώνει επαναληπτικά την παράμετρο της ελάχιστης υποστήριξης μέχρις ότου να βρεθεί το απαιτούμενο πλήθος κανόνων που χαρακτηρίζονται από τη δοθείσα τιμή της ελάχιστης εμπιστοσύνης.

## Έξοδος

Μία λίστα η οποία κληρονομεί τη συμπεριφορά της από την κλάση `Weka_associators`, περιλαμβανομένων διάφορων συστατικών στοιχείων.

associator	μία αναφορά (της κλάσης <code>jobRef</code> ) σε ένα αντικείμενο τύπου <code>Java</code> , η οποία δημιουργείται εφαρμόζοντας τη μέθοδο <code>buildAssociations</code> του <code>Weka</code> , στα στιγμιότυπα εκπαίδευσης χρησιμοποιώντας τις δοθείσες επιλογές.
------------	---

## Σημείωση

Το πακέτο `tertius` απαιτείται να εγκατασταθεί για τη χρήση του `tertius`.

## Αναφορές

R. Agrawal and R. Srikant (1994). Fast algorithms for mining association rules in large databases. Proceedings of the International Conference on Very Large Databases, 478–499. Santiago, Chile:

Morgan Kaufmann, Los Altos, CA.

P. A. Flach and N. Lachiche (1999). Confirmation-guided discovery of first-order rules with Tertius. *Machine Learning*, 42, 61–95.

I. H. Witten and E. Frank (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd Edition, Morgan Kaufmann, San Francisco.

## Παραδείγματα

```
x <- read.arff(system.file("arff", "contact-lenses.arff",
                           package = "RWeka"))

## Apriori with defaults.
Apriori(x)

## Some options: set required number of rules to 20.
Apriori(x, Weka_control(N = 20))

## Not run:

## Requires Weka package 'Tertius' to be installed.
## Tertius with defaults.
Tertius(x)

## Some options: only classification rules (single item in the RHS).
Tertius(x, Weka_control(S = TRUE))

## End(Not run)
```

---

*Weka\_attribute\_evaluators*

---

*R/Weka Attribute Evaluators*

## Περιγραφή

Διεπαφές της R για τους εκτιμητές χαρακτηριστικών του Weka.

## Χρήση

`GainRatioAttributeEval(formula, data, subset, na.action, control = NULL)`

`InfoGainAttributeEval(formula, data, subset, na.action, control = NULL)`

## Ορίσματα

formula	συμβολική περιγραφή του εξεταζόμενου μοντέλου. Αξίζει να τονισθεί πως στην περίπτωση των μη καθοδηγούμενων φίλτρων, η απόκριση μπορεί να παραλειφθεί.
data	προαιρετικό dataframe το οποίο περιέχει της μεταβλητές του μοντέλου
subset το	προαιρετικό διάνυσμα το οποίο καθορίζει ένα υποσύνολο των παρατηρήσεων οποίο θα αξιοποιηθεί στη διαδικασία χτισίματος του μοντέλου.
na.action	συνάρτηση η οποία υποδηλώνει τις ενέργειες που θα εκτελεσθούν σε περίπτωση που τα δεδομένα περιέχουν ελλιπείς τιμές (NAs). Περισσότερες λεπτομέρειες μπορούν να αναζητηθούν στο <code>model.frame</code> .
control	αντικείμενο της κλάσης <code>Weka_control</code> ή ένα διάνυσμα χαρακτήρων που περιγράφει επιλογές ελέγχου ή εξίσωση του με το <code>NULL</code> (default). Οι διαθέσιμες επιλογές μπορούν να προσπελαθούν είτε on-line χρησιμοποιώντας το <code>Weka Option Wizard (WOW)</code> είτε μέσω του εγγράφου τεκμηρίωσης του <code>Weka</code> .

## Λεπτομέρειες

`GainRatioAttributeEval` εκτιμά την αξία ενός χαρακτηριστικού μετρώντας το δείκτη κέρδους σε σχέση με τη κάθε κλάση.

`InfoGainAttributeEval` εκτιμά την αξία ενός χαρακτηριστικού μετρώντας το δείκτη κέρδους σε σχέση με τη κάθε κλάση.

## Έξοδος

Ένα αριθμητικό διάνυσμα με τα αριθμητικά στοιχεία της αξίας του κάθε χαρακτηριστικού να καθορίζεται από τη δεξιά πλευρά του πεδίου formula.

## Παραδείγματα

```
InfoGainAttributeEval(Species ~ . , data = iris)
```

---

*Weka\_Classifiers*

*R/Weka Classifiers*

---

## Περιγραφή

Διεπαφές της R για τους ταξινομητές του Weka.

## Λεπτομέρειες

Οι καθοδηγούμενοι εκμαθητές, π.χ. οι αλγόριθμοι για κατηγοριοποίηση και οπισθοδρόμηση, συνηθίζεται να κατονομάζονται στο περιβάλλον του Weka ως ταξινομητές. (η αριθμητική πρόβλεψη, π.χ. η οπισθοδρόμηση, παρουσιάζεται ως η πρόβλεψη μίας συνεχόμενης κλάσης.)

Οι συναρτήσεις μέσω της διεπαφής της R για το Weka δημιουργούνται μέσω του `make_Weka_classifier` και συνοδεύονται από τις τυπικές παραμέτρους `formula`, `data`, `subset`, `na.action` και `control` (default: `none`), όπου οι τέσσερις πρώτοι παρουσιάζουν τη συνήθη ερμηνεία που έχει επικρατήσει στις συναρτήσεις στατιστικής μοντελοποίησης στην R, και η τελευταία καθορίζει τις επιλογές ελέγχου που επιλέγεται κάθε φορά να υιοθετηθούν από τον εκμαθητή του Weka.

Εξ ορισμού, για τις παραμέτρους του εκάστοτε μοντέλου αρκεί η χρήση των τελεστών '+' και '-' για να επισημανθεί η ενσωμάτωση των μεταβλητών ή όχι, αντίστοιχα.

Περισσότερες πληροφορίες καθώς και χρήσιμες οδηγίες για το πώς αξιοποιείται η επιλογή `na.action` μπορούν αναζητηθούν στο `model.frame`.

Επιπλέον, τα αντικείμενα που δημιουργούνται από αυτές τις διεπαφές πάντοτε κληρονομούν από την κλάση `Weka_classifier`, και περιλαμβάνουν τουλάχιστον τις μεθόδους `print`, `summary` (μέσω του `evaluate_Weka_classifier`) και `predict`.

## Προτάσεις για παρεμφερείς συναρτήσεις

Οι διαθέσιμες τυπικές συναρτήσεις των συγκεκριμένων διεπαφών που αναφέρθηκαν μέχρι στιγμής καταγράφονται λεπτομερώς στα `Weka_classifier_functions` (συναρτήσεις εκμάθησης για οπισθοδρόμηση και κατηγοριοποίηση), `Weka_classifiers_lazy` (αργοί εκμαθητές), `Weka_classifier_meta` (meta εκμαθητές), `Weka_classifier_rules` (εκμαθητές μέσω κανόνων) και `Weka_classifier_tress` (δενδρικοί εκμαθητές για οπισθοδρόμηση και κατηγοριοποίηση).

---

*Weka\_classifier\_functions*

*R/Weka Classifier Functions*

---

## Περιγραφή

Διεπαφές της R προς τις συναρτήσεις εκμάθησης ταξινομητών και οπισθοδρόμησης του Weka.

## Χρήση

```
LinearRegression(formula, data, subset, na.action,  
                  control = Weka_control(), options = NULL)  
Logistic(formula, data, subset, na.action,  
           control = Weka_control(), options = NULL)  
SMO(formula, data, subset, na.action,  
     control = Weka_control(), options = NULL)
```

## Ορίσματα

<code>formula</code>	συμβολική περιγραφή του εξεταζόμενου μοντέλου το οποίο πρόκειται να προσεγγιστεί.
----------------------	---

data	προαιρετικό dataframe το οποίο περιέχει της μεταβλητές του μοντέλου
subset	προαιρετικό διάνυσμα το οποίο καθορίζει ένα υποσύνολο των παρατηρήσεων το οποίο θα αξιοποιηθεί στη διαδικασία χτισίματος του μοντέλου.
na.action	συνάρτηση η οποία υποδηλώνει τις ενέργειες που θα εκτελεσθούν σε περίπτωση που τα δεδομένα περιέχουν ελλιπείς τιμές (NAs). Περισσότερες λεπτομέρειες μπορούν να αναζητηθούν στο model.frame.
control	αντικείμενο της κλάσης Weka_control ή ένα διάνυσμα χαρακτήρων που περιγράφει επιλογές ελέγχου ή εξίσωση του με το NULL (default). Οι διαθέσιμες επιλογές μπορούν να αναζητηθούν είτε on-line χρησιμοποιώντας το Weka Option Wizard (WOW) είτε μέσω του εγγράφου τεκμηρίωσης του Weka.
options	μία λίστα, που όπως αποκαλύπτει το όνομα της περιέχει περισσότερες επιλογές ή ισούται με το NULL (default). Περισσότερα ακολουθούν στις Λεπτομέρειες.

## Λεπτομέρειες

Υπάρχει η μέθοδος predict για την πρόβλεψη από το ταιριασμένο μοντέλο καθώς και η μέθοδος summary η οποία βασίζεται στο evaluate\_Weka\_classifier.

LinearRegression υλοποιεί κατάλληλα γραμμικά μοντέλα οπισθοδρόμησης, χρησιμοποιώντας το κριτήριο Akaike για την επιλογή μοντέλου.

Logistic υλοποιεί πολυωνυμικά λογιστικά μοντέλα οπισθοδρόμησης βασισμένη στην μέθοδο κορυφογραμμής (le Cessie and van Houwelingen, 1992).

SMO υλοποιεί τον αλγόριθμο βελτιστοποίησης διαδοχικών ελαχίστων του John C. Platt για την εκπαίδευση ενός ταξινομητή τύπου διανυσμάτων υποστήριξης χρησιμοποιώντας πολυωνυμικούς ή RBF πυρήνες. Τα προβλήματα πολλαπλών κλάσεων λύνονται χρησιμοποιώντας κατηγοριοποίηση ανά ζεύγη.

Εξ ορισμού, για τις παραμέτρους του εκάστοτε μοντέλου αρκεί η χρήση των τελεστών '+' και '-' για να επισημανθεί η ενσωμάτωση των μεταβλητών ή όχι, αντίστοιχα.

Το όρισμα options επιδέχεται περαιτέρω παραμετροποίηση. Επί του παρόντος, οι επιλογές model και instances (ή εν μέρει ταιριάσματα αυτών) χρησιμοποιούνται ως εξής: αν οριστούν ως TRUE, το



πλαίσιο του model ή τα αντίστοιχα στιγμιότυπα του Weka, αντίστοιχα, περιλαμβάνονται στο δημιουργηθέν αντικείμενο του ταιριασμένου μοντέλου, πιθανότατα επιταχύνοντας τους μεταγενέστερους υπολογισμούς επί τόπου αντικειμένου αυτού. Εξ ορισμού, καμία από αυτές δεν περιλαμβάνεται.

## Έξοδος

Μία λίστα η οποία κληρονομεί από τις κλάσεις `Weka_functions` και `Weka_classifiers` με τα εξής στοιχεία να περιλαμβάνονται σε αυτή:

<code>classifier</code>	μία αναφορά (της κλάσης <code>jobRef</code> ) σε ένα αντικείμενο τύπου <code>Java</code> , η οποία δημιουργείται εφαρμόζοντας τη μέθοδο <code>buildClassifier</code> του <code>Weka</code> έτσι ώστε να υλοποιηθεί το συγκεκριμένο κάθε φορά μοντέλο χρησιμοποιώντας τις δοθείσες επιλογές ελέγχου.
<code>predictions</code>	ένα αριθμητικό ή κατηγορικό διάνυσμα που περιέχει τις προβλέψεις του μοντέλου για τα στιγμιότυπα της εκπαίδευσης (αποτελεί ουσιαστικά το αποτέλεσμα της κλήσης της μεθόδου <code>classifyInstance</code> του <code>Weka</code> για τον υλοποιημένο ταξινομητή όλων των στιγμιότυπων).
<code>call</code>	η ταιριασμένη κλήση.

## Αναφορές

J. C. Platt (1998). Fast training of Support Vector Machines using Sequential Minimal Optimization. In B. Schoelkopf, C. Burges, and A. Smola (eds.), *Advances in Kernel Methods — Support Vector Learning*. MIT Press.

I. H. Witten and E. Frank (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd Edition, Morgan Kaufmann, San Francisco.

## Προτάσεις για παρεμφερείς συναρτήσεις

`Weka_classifiers`

## Παραδείγματα

## Linear regression:

## Using standard data set 'mtcars'.

```
LinearRegression(mpg ~ ., data = mtcars)
```

## Compare to R:

```
step(lm(mpg ~ ., data = mtcars), trace = 0)
```

## Using standard data set 'chickwts'.

```
LinearRegression(weight ~ feed, data = chickwts)
```

## (Note the interactions!)

## Logistic regression:

## Using standard data set 'infert'.

```
STATUS <- factor( infert$case, labels = c( "control", "case"))
```

```
Logistic(STATUS ~ spontaneous + induced, data = infert)
```

## Compare to R:

```
glm( STATUS ~ spontaneous + induced, data = infert, family = binomial())
```

## parameter (argument '-G') instead of the default polynomial kernel

## (from a question on r-help):

```
SMO(Species ~ ., data = iris,
```

```
    control = Weka_control(
```

```
        K = list( "weka.classifiers.functions.supportVector.RBFKernel", G = 2)))
```

## In fact, by some hidden magic it also "works" to give the "base" name

## of the Weka kernel class:

```
SMO(Species ~ ., data = iris,
```

```
    control = Weka_control (K = list(" RBFKernel" , G = 2)))
```

## Περιγραφή

Διεπαφές της R προς τους αργούς εκμαθητές του Weka.

## Χρήση

```
IBk(formula, data, subset, na.action,  
     control = Weka_control(), options = NULL)  
LBR(formula, data, subset, na.action,  
     control = Weka_control(), options = NULL)
```

## Ορίσματα

formula	συμβολική περιγραφή του εξεταζόμενου μοντέλου το οποίο πρόκειται να προσεγγιστεί.
data	προαιρετικό dataframe το οποίο περιέχει της μεταβλητές του μοντέλου.
subset	προαιρετικό διάνυσμα το οποίο καθορίζει ένα υποσύνολο των παρατηρήσεων το οποίο θα αξιοποιηθεί στη διαδικασία χτισίματος του μοντέλου.
na.action	συνάρτηση η οποία υποδηλώνει τις ενέργειες που θα εκτελεσθούν σε περίπτωση που τα δεδομένα περιέχουν ελλιπείς τιμές (NAs). Περισσότερες λεπτομέρειες μπορούν να αναζητηθούν στο <code>model.frame</code> .
control	αντικείμενο της κλάσης <code>Weka_control</code> που περιέχει τις επιλογές ελέγχου που αναμένεται να περάσουν σαν ορίσματα στον εκάστοτε εκμαθητή του Weka. Οι διαθέσιμες επιλογές μπορούν να αναζητηθούν είτε on-line χρησιμοποιώντας το Weka Option Wizard (WOW) είτε μέσω του εγγράφου τεκμηρίωσης του Weka.
options	μία λίστα, που όπως αποκαλύπτει το όνομα της περιέχει περισσότερες επιλογές ή ισούται με το <code>NULL</code> (default). Περισσότερα ακολουθούν στις Λεπτομέρειες.

## Λεπτομέρειες

Υπάρχει η μέθοδος `predict` για την πρόβλεψη από το ταιριασμένο μοντέλο καθώς και η μέθοδος `summary` η οποία βασίζεται στο `evaluate_Weka_classifier`.

IBk παρέχει τον ταξινομητή των  $k$  – πλησιέστερων γειτόνων. Περισσότερες πληροφορίες παρέχονται από τους Aha & Kibler (1991).

LBR (“Lazy Bayesian Rules”) υλοποιεί μία αργή προσέγγιση εκμάθησης με σκοπό τη μείωση της υπόθεσης του Naïve Bayes περί ανεξαρτησίας των χαρακτηριστικών, όπως χαρακτηριστικά πρότειναν οι Zheng & Webb το 2000.

Εξ ορισμού, για τις παραμέτρους του εκάστοτε μοντέλου αρκεί η χρήση των τελεστών ‘+’ και ‘-’ για να επισημανθεί η ενσωμάτωση των μεταβλητών ή όχι, αντίστοιχα.

Το όρισμα `options` επιδέχεται περαιτέρω παραμετροποίηση. Επί του παρόντος, οι επιλογές `model` και `instances` (ή εν μέρει ταιριάσματα αυτών) χρησιμοποιούνται ως εξής: αν οριστούν ως `TRUE`, το πλαίσιο του `model` ή τα αντίστοιχα στιγμιότυπα του Weka, αντίστοιχα, περιλαμβάνονται στο δημιουργηθέν αντικείμενο του ταιριασμένου μοντέλου, πιθανότατα επιταχύνοντας τους μεταγενέστερους υπολογισμούς επί τόπου αντικειμένου αυτού. Εξ ορισμού, καμία από αυτές δεν περιλαμβάνεται.

## Έξοδος

Μία λίστα η οποία κληρονομεί από τις κλάσεις `Weka_functions` και `Weka_classifiers` με τα εξής στοιχεία να περιλαμβάνονται σε αυτή:

<code>classifier</code>	μία αναφορά (της κλάσης <code>jobRef</code> ) σε ένα αντικείμενο τύπου <code>Java</code> , η οποία δημιουργείται εφαρμόζοντας τη μέθοδο <code>buildClassifier</code> του Weka έτσι ώστε να υλοποιηθεί το συγκεκριμένο κάθε φορά μοντέλο χρησιμοποιώντας τις δοθείσες επιλογές ελέγχου.
<code>predictions</code>	ένα αριθμητικό ή κατηγορικό διάνυσμα που περιέχει τις προβλέψεις του μοντέλου για τα στιγμιότυπα της εκπαίδευσης (αποτελεί ουσιαστικά το αποτέλεσμα της κλήσης της μεθόδου <code>classifyInstance</code> του Weka για τον υλοποιημένο ταξινομητή όλων των στιγμιοτύπων).

call η ταιριασμένη κλήση.

## Προσοχή

LBR απαιτεί το πακέτο `lazyBayesianRules` του Weka έτσι ώστε να εγκατασταθεί επιτυχώς.

## Αναφορές

D. Aha and D. Kibler (1991). Instance-based learning algorithms. *Machine Learning*, 6, 37–66.

Z. Zheng and G. Webb (2000). Lazy learning of Bayesian rules. *Machine Learning*, 41/1, 53–84.

## Προτάσεις για παρεμφερείς συναρτήσεις

`Weka_classifiers`

---

*Weka\_classifier\_meta*

*R/Weka Meta Learners*

---

## Περιγραφή

Διεπαφές της R προς τους meta εκμαθητές του Weka.

## Χρήση

```
AdaBoostM1(formula, data, subset, na.action,  
             control = Weka_control(), options = NULL)
```

```
Bagging(formula, data, subset, na.action,  
         control = Weka_control(), options = NULL)
```

```
LogitBoost(formula, data, subset, na.action,  
            control = Weka_control(), options = NULL)
```

```
MultiBoostAB(formula, data, subset, na.action,
```

```
control = Weka_control(), options = NULL)
```

```
Stacking(formula, data, subset, na.action,  
          control = Weka_control(), options = NULL)
```

```
CostSensitiveClassifier(formula, data, subset, na.action,  
                        control = Weka_control(), options = NULL)
```

## Ορίσματα

formula	συμβολική περιγραφή του εξεταζόμενου μοντέλου το οποίο πρόκειται να προσεγγιστεί.
data	προαιρετικό dataframe το οποίο περιέχει της μεταβλητές του μοντέλου.
subset	προαιρετικό διάνυσμα το οποίο καθορίζει ένα υποσύνολο των παρατηρήσεων το οποίο θα αξιοποιηθεί στη διαδικασία χτισίματος του μοντέλου.
na.action	συνάρτηση η οποία υποδηλώνει τις ενέργειες που θα εκτελεσθούν σε περίπτωση που τα δεδομένα περιέχουν ελλιπείς τιμές (NAs). Περισσότερες λεπτομέρειες μπορούν να αναζητηθούν στο <code>model.frame</code> .
control	αντικείμενο της κλάσης <code>Weka_control</code> που περιέχει τις επιλογές ελέγχου που αναμένεται να περάσουν σαν ορίσματα στον εκάστοτε εκμαθητή του Weka. Οι διαθέσιμες επιλογές μπορούν να αναζητηθούν είτε on-line χρησιμοποιώντας το Weka Option Wizard (WOW) είτε μέσω του εγγράφου τεκμηρίωσης του Weka. Οι βασικοί ταξινομητές μαζί με τη διαθέσιμη διεπαφή R/Weka (δες <code>list_Weka_interfaces</code> ) μπορούν να καθορισθούν (χρησιμοποιώντας την επιλογή “W”) μέσω του ίδιου του ονόματος τους στο χωρίο “base name”, όπως άλλωστε εμφανίζεται και στην καταχώρηση της διεπαφής (δες τα Παραδείγματα) ή στις συναρτήσεις της ίδιας της διεπαφής.
options	μία λίστα, που όπως αποκαλύπτει το όνομα της περιέχει περισσότερες επιλογές ή ισούται με το NULL (default). Περισσότερα ακολουθούν στις Λεπτομέρειες.

## Λεπτομέρειες

Υπάρχει η μέθοδος `predict` για την πρόβλεψη από το ταιριασμένο μοντέλο καθώς και η μέθοδος `summary` η οποία βασίζεται στο `evaluate_Weka_classifier`.

`AdaBoostM1` υλοποιεί τη μέθοδο AdaBoost M1 των Freund & Schapire (1996).

`LogitBoost` πραγματοποιεί την τεχνική του boosting μέσω προσθετικής λογιστικής οπισθοδρόμησης (Friedman, Hastie και Tibshirani, 2000).

`MultiBoostAB` υλοποιεί τον αλγόριθμο MultiBoosting (Webb, 2000), ο οποίος ουσιαστικά αποτελεί μία επέκταση της τεχνικής AdaBoost για το σχηματισμό ομάδων απόφασης, οι οποίοι δύναται να θεωρηθούν ως ένας συνδυασμός του AdaBoost και της τεχνικής “wagging”.

`Stacking` παρέχει τη δυνατότητα για την εφαρμογή της ομώνυμης τεχνικής (Wolpert, 1992).

`CostSensitiveClassifier` τροποποιεί τον βασικό του ταξινομητή σε cost-sensitive ταξινομητή.

Εξ ορισμού, για τις παραμέτρους του εκάστοτε μοντέλου αρκεί η χρήση των τελεστών ‘+’ και ‘-’ για να επισημανθεί η ενσωμάτωση των μεταβλητών ή όχι, αντίστοιχα.

Το όρισμα `options` επιδέχεται περαιτέρω παραμετροποίηση. Επί του παρόντος, οι επιλογές `model` και `instances` (ή εν μέρει ταιριάσματα αυτών) χρησιμοποιούνται ως εξής: αν οριστούν ως `TRUE`, το πλαίσιο του `model` ή τα αντίστοιχα στιγμιότυπα του `Weka`, αντίστοιχα, περιλαμβάνονται στο δημιουργηθέν αντικείμενο του ταιριασμένου μοντέλου, πιθανότατα επιταχύνοντας τους μεταγενέστερους υπολογισμούς επί τόπου αντικειμένου αυτού. Εξ ορισμού, καμία από αυτές δεν περιλαμβάνεται.

## Έξοδος

Μία λίστα η οποία κληρονομεί από τις κλάσεις `Weka_functions` και `Weka_classifiers` με τα εξής στοιχεία να περιλαμβάνονται σε αυτή:

<code>classifier</code>	μία αναφορά (της κλάσης <code>jobRef</code> ) σε ένα αντικείμενο τύπου <code>Java</code> , η οποία δημιουργείται εφαρμόζοντας τη μέθοδο <code>buildClassifier</code> του <code>Weka</code> έτσι ώστε να υλοποιηθεί το συγκεκριμένο κάθε φορά μοντέλο χρησιμοποιώντας τις δοθείσες επιλογές ελέγχου.
-------------------------	---

predictions	ένα αριθμητικό ή κατηγορικό διάνυσμα που περιέχει τις προβλέψεις του μοντέλου για τα στιγμιότυπα της εκπαίδευσης (αποτελεί ουσιαστικά το αποτέλεσμα της κλήσης της μεθόδου <code>classifyInstance</code> του Weka για τον υλοποιημένο ταξινομητή όλων των στιγμιοτύπων).
call	η ταιριασμένη κλήση.

## Προσοχή

multiBoostAB απαιτεί το πακέτο multiBoostAB του Weka έτσι ώστε να εγκατασταθεί επιτυχώς.

## Αναφορές

- L. Breiman (1996). Bagging predictors. *Machine Learning*, 24/2, 123–140.
- Y. Freund and R. E. Schapire (1996). Experiments with a new boosting algorithm. In *Proceedings of the International Conference on Machine Learning*, pages 148–156. Morgan Kaufmann: San Francisco.
- J. H. Friedman, T. Hastie, and R. Tibshirani (2000). Additive logistic regression: A statistical view of boosting. *Annals of Statistics*, 28/2, 337–374.
- G. I. Webb (2000). MultiBoosting: A technique for combining boosting and wagging. *Machine Learning*, 40/2, 159–196.
- I. H. Witten and E. Frank (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd Edition, Morgan Kaufmann, San Francisco.
- D. H. Wolpert (1992). Stacked generalization. *Neural Networks*, 5, 241–259.

## Προτάσεις για παρεμφερείς συναρτήσεις

Weka\_classifiers



## Παραδείγματα

```
## Use AdaBoostM1 with decision stumps.  
m1 <- AdaBoostM1(Species ~ ., data = iris,  
                  control = Weka_control(W = "DecisionStump"))  
  
table(predict(m1), iris$Species)  
summary(m1)      # uses evaluate_Weka_classifier()  
  
## Control options for the base classifiers employed by the meta  
## learners (apart from Stacking) can be given as follows:  
m2 <- AdaBoostM1(Species ~ ., data = iris,  
                  control = Weka_control(W = list(J48, M = 30)))
```

---

*Weka\_classifier\_rules*

*R/Weka Rule Learners*

---

## Περιγραφή

Διεπαφές της R προς τους εκμαθητές κανόνων του Weka.

## Χρήση

```
JRip(formula, data, subset, na.action,  
      control = Weka_control(), options = NULL)  
M5Rules(formula, data, subset, na.action,  
         control = Weka_control(), options = NULL)  
OneR(formula, data, subset, na.action,  
      control = Weka_control(), options = NULL)  
PART(formula, data, subset, na.action,  
     control = Weka_control(), options = NULL)
```

## Ορίσματα

formula	συμβολική περιγραφή του εξεταζόμενου μοντέλου το οποίο πρόκειται να προσεγγιστεί.
data	προαιρετικό dataframe το οποίο περιέχει της μεταβλητές του μοντέλου.
subset	προαιρετικό διάνυσμα το οποίο καθορίζει ένα υποσύνολο των παρατηρήσεων το οποίο θα αξιοποιηθεί στη διαδικασία χτισίματος του μοντέλου.
na.action	συνάρτηση η οποία υποδηλώνει τις ενέργειες που θα εκτελεσθούν σε περίπτωση που τα δεδομένα περιέχουν ελλιπείς τιμές (NAs). Περισσότερες λεπτομέρειες μπορούν να αναζητηθούν στο model.frame.
control	αντικείμενο της κλάσης Weka_control που περιέχει τις επιλογές ελέγχου που αναμένεται να περάσουν σαν ορίσματα στον εκάστοτε εκμαθητή του Weka. Οι διαθέσιμες επιλογές μπορούν να αναζητηθούν είτε on-line χρησιμοποιώντας το Weka Option Wizard (WOW) είτε μέσω του εγγράφου τεκμηρίωσης του Weka.
options	μία λίστα, που όπως αποκαλύπτει το όνομα της περιέχει περισσότερες επιλογές ή ισούται με το NULL (default). Περισσότερα ακολουθούν στις Λεπτομέρειες.

## Λεπτομέρειες

Υπάρχει η μέθοδος predict για την πρόβλεψη από το ταιριασμένο μοντέλο καθώς και η μέθοδος summary η οποία βασίζεται στο evaluate\_Weka\_classifier.

JRip υλοποιεί έναν προτασιακό εκμαθητή κανόνων, βασισμένος στο “Repeated Incremental Pruning to Produce Error Reduction” (RIPPER), όπως αυτός προτάθηκε από τον Cohen (1995)/

M5Rules παράγει μία λίστα αποφάσεων για προβλήματα οπισθοδρόμησης χρησιμοποιώντας την τεχνική διαίρει και βασίλευε. Σε κάθε επανάληψη υλοποιείται ένα δεδνρικό μοντέλο χρησιμοποιώντας τον M5 και ενσωματώνοντας το “καλύτερο” κάθε φορά φύλλο του δένδρου εντός ενός κανόνα. Περισσότερες πληροφορίες μπορούν να αναζητηθούν στη δουλειά των Hall, Holmes και Frank (1999).

OneR δημιουργεί έναν απλό 1-R ταξινομητή. Περισσότερες πληροφορίες παρέχονται από τον Holte (1993).

PART δημιουργεί μία PART λίστα απόφασης χρησιμοποιώντας την προσέγγιση των Frank και Witten (1998).

Εξ ορισμού, για τις παραμέτρους του εκάστοτε μοντέλου αρκεί η χρήση των τελεστών '+' και '-' για να επισημανθεί η ενσωμάτωση των μεταβλητών ή όχι, αντίστοιχα.

Το όρισμα options επιδέχεται περαιτέρω παραμετροποίηση. Επί του παρόντος, οι επιλογές model και instances (ή εν μέρει ταιριάσματα αυτών) χρησιμοποιούνται ως εξής: αν οριστούν ως TRUE, το πλαίσιο του model ή τα αντίστοιχα στιγμιότυπα του Weka, αντίστοιχα, περιλαμβάνονται στο δημιουργηθέν αντικείμενο του ταιριασμένου μοντέλου, πιθανότατα επιταχύνοντας τους μεταγενέστερους υπολογισμούς επί τόπου αντικειμένου αυτού. Εξ ορισμού, καμία από αυτές δεν περιλαμβάνεται.

## Έξοδος

Μία λίστα η οποία κληρονομεί από τις κλάσεις Weka\_functions και Weka\_classifiers με τα εξής στοιχεία να περιλαμβάνονται σε αυτή:

classifier	μία αναφορά (της κλάσης jobRef) σε ένα αντικείμενο τύπου Java, η οποία δημιουργείται εφαρμόζοντας τη μέθοδο buildClassifier του Weka έτσι ώστε να υλοποιηθεί το συγκεκριμένο κάθε φορά μοντέλο χρησιμοποιώντας τις δοθείσες επιλογές ελέγχου.
predictions	ένα αριθμητικό ή κατηγορικό διάνυσμα που περιέχει τις προβλέψεις του μοντέλου για τα στιγμιότυπα της εκπαίδευσης (αποτελεί ουσιαστικά το αποτέλεσμα της κλήσης της μεθόδου classifyInstance του Weka για τον υλοποιημένο ταξινομητή όλων των στιγμιοτύπων).
call	η ταιριασμένη κλήση.

## Αναφορές

W. W. Cohen (1995). Fast effective rule induction. In A. Prieditis and S. Russell (eds.), Proceedings of the 12th International Conference on Machine Learning, pages 115–123. Morgan Kaufmann.

ISBN 1-55860-377-8. <http://citeseer.ist.psu.edu/cohen95fast.html>

E. Frank and I. H. Witten (1998). Generating accurate rule sets without global optimization. In J. Shavlik (ed.), Machine Learning: Proceedings of the Fifteenth International Conference. Morgan Kaufmann Publishers: San Francisco, CA. <http://www.cs.waikato.ac.nz/~eibe/pubs/ML98-57.ps.gz>

M. Hall, G. Holmes, and E. Frank (1999). Generating rule sets from model trees. Proceedings of the Twelfth Australian Joint Conference on Artificial Intelligence, Sydney, Australia, pages 1–12. Springer-Verlag. <http://citeseer.ist.psu.edu/holmes99generating.html>

R. C. Holte (1993). Very simple classification rules perform well on most commonly used datasets. Machine Learning, 11, 63–91.

I. H. Witten and E. Frank (2005). Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition, Morgan Kaufmann, San Francisco.

## Προτάσεις για παρεμφερείς συναρτήσεις

Weka\_classifiers

## Παραδείγματα

```
M5Rules(mpg ~ ., data = mtcars)
m <- PART(Species ~ ., data = iris)

m
summary(m)
```

---

*Weka\_classifier\_trees*

---

---

*R/Weka Classifier Trees*

---

## Περιγραφή

Διεπαφή της R προς τους ταξινομητές τύπου δένδρου ή οπισθοδρόμησης του Weka.

## Χρήση

```
J48(formula, data, subset, na.action,  
    control = Weka_control(), options = NULL)
```

```
LMT(formula, data, subset, na.action,  
    control = Weka_control(), options = NULL)
```

```
M5P(formula, data, subset, na.action,  
    control = Weka_control(), options = NULL)
```

```
DecisionStump(formula, data, subset, na.action,  
    control = Weka_control(), options = NULL)
```

## Ορίσματα

formula	συμβολική περιγραφή του εξεταζόμενου μοντέλου το οποίο πρόκειται να προσεγγιστεί.
data	προαιρετικό dataframe το οποίο περιέχει της μεταβλητές του μοντέλου.
subset	προαιρετικό διάνυσμα το οποίο καθορίζει ένα υποσύνολο των παρατηρήσεων το οποίο θα αξιοποιηθεί στη διαδικασία χτισίματος του μοντέλου.
na.action	συνάρτηση η οποία υποδηλώνει τις ενέργειες που θα εκτελεσθούν σε περίπτωση που τα δεδομένα περιέχουν ελλιπείς τιμές (NAs). Περισσότερες λεπτομέρειες μπορούν να αναζητηθούν στο <code>model.frame</code> .
control	αντικείμενο της κλάσης <code>Weka_control</code> που περιέχει τις επιλογές ελέγχου που αναμένεται να περάσουν σαν ορίσματα στον εκάστοτε εκμαθητή του Weka. Οι διαθέσιμες επιλογές μπορούν να αναζητηθούν είτε on-line χρησιμοποιώντας το Weka Option Wizard (WOW) είτε μέσω του εγγράφου τεκμηρίωσης του Weka.
options	μία λίστα, που όπως αποκαλύπτει το όνομα της περιέχει περισσότερες επιλογές ή ισούται με το <code>NULL</code> (default). Περισσότερα ακολουθούν στις Λεπτομέρειες.

## Λεπτομέρειες

Υπάρχει η μέθοδος `predict` για την πρόβλεψη από το ταιριασμένο μοντέλο καθώς και η μέθοδος `summary` η οποία βασίζεται στο `evaluate_Weka_classifier`.

Επιπρόσθετα, υπάρχει και μία επιπλέον μέθοδος για σχεδιασμό ταιριασμένων δυαδικών `Weka_trees` που προσφέρεται από το πακέτο `partykit`. Μέσω αυτής της μεθόδου το αντικείμενο `Weka_tree` μετατρέπεται σε ένα αντικείμενο τύπου `party` και έπειτα απλώς καλείται η συνάρτηση σχεδιασμού της κλάσης αυτής (δες `plot.party`)

Δεδομένου πως ο δενδρικός εκμαθητής για ταξινόμηση του `Weka` υλοποιήσει την “Σχεδιάσιμη” διεπαφή (π.χ. όπως αυτή παρέχεται από τη μέθοδο `graph`), η εντολή `write_to_dot` μπορεί αν χρησιμοποιηθεί για να αποκτηθεί η DOT αναπαράσταση για την οπτικοποίηση του δένδρου μέσω του `Graphviz` ή του πακέτου `Rgraphviz`.

`J48` παράγει ένα κλαδευμένο ή μη-κλαδευμένο C4.5 δένδρο απόφασης (Quinlan, 1993).

`LMT` υλοποιεί τα “Logistic Model Trees” (Landwehr, 2003; Landwehr et al., 2005)

`M5P` (όπου το `P` αντιστοιχεί στο ‘prime’) δημιουργεί `M5` δενδρικά μοντέλα χρησιμοποιώντας τον `M5` αλγόριθμο, ο οποίος προτάθηκε από τους Wang & Witten (1997) και βελτίωσε ουσιαστικά τον αρχικό αλγόριθμο του `M5` του Quinlan (1992).

`DecisionStump` υλοποιεί `stump` αποφάσεις (δένδρα με μία απλή διαίρεση), τα οποία συχνά χρησιμοποιούνται ως βασικοί εκμαθητές για τους μετέπειτα μετα-εκμαθητές, όπως ο `Boosting`.

Εξ ορισμού, για τις παραμέτρους του εκάστοτε μοντέλου αρκεί η χρήση των τελεστών ‘+’ και ‘-’ για να επισημανθεί η ενσωμάτωση των μεταβλητών ή όχι, αντίστοιχα.

Το όρισμα `options` επιδέχεται περαιτέρω παραμετροποίηση. Επί του παρόντος, οι επιλογές `model` και `instances` (ή εν μέρει ταιριάσματα αυτών) χρησιμοποιούνται ως εξής: αν οριστούν ως `TRUE`, το πλαίσιο του `model` ή τα αντίστοιχα στιγμιότυπα του `Weka`, αντίστοιχα, περιλαμβάνονται στο δημιουργηθέν αντικείμενο του ταιριασμένου μοντέλου, πιθανότατα επιταχύνοντας τους μεταγενέστερους υπολογισμούς επί τόπου αντικειμένου αυτού. Εξ ορισμού, καμία από αυτές δεν περιλαμβάνεται.

`parse_Weka_digraph` αναλύει τον εκάστοτε γράφο ο οποίος σχετίζεται με έναν δενδρικό ταξινομητή του Weka (και η πρόσβαση σε αυτόν εξυπηρετείται μέσω της μεθόδου `graph()` του Weka), επιστρέφοντας μία απλή λίστα με τους κόμβους και τις ακμές.

## Έξοδος

Μία λίστα η οποία κληρονομεί από τις κλάσεις `Weka_functions` και `Weka_classifiers` με τα εξής στοιχεία να περιλαμβάνονται σε αυτή:

<code>classifier</code>	μία αναφορά (της κλάσης <code>jobRef</code> ) σε ένα αντικείμενο τύπου <code>Java</code> , η οποία δημιουργείται εφαρμόζοντας τη μέθοδο <code>buildClassifier</code> του Weka έτσι ώστε να υλοποιηθεί το συγκεκριμένο κάθε φορά μοντέλο χρησιμοποιώντας τις δοθείσες επιλογές ελέγχου.
<code>predictions</code>	ένα αριθμητικό ή κατηγορικό διάνυσμα που περιέχει τις προβλέψεις του μοντέλου για τα στιγμιότυπα της εκπαίδευσης (αποτελεί ουσιαστικά το αποτέλεσμα της κλήσης της μεθόδου <code>classifyInstance</code> του Weka για τον υλοποιημένο ταξινομητή όλων των στιγμιοτύπων).
<code>call</code>	η ταιριασμένη κλήση.

## Αναφορές

N. Landwehr (2003). Logistic Model Trees. Master's thesis, Institute for Computer Science, University of Freiburg, Germany. [http://www.cs.uni-potsdam.de/ml/landwehr/diploma\\_thesis.pdf](http://www.cs.uni-potsdam.de/ml/landwehr/diploma_thesis.pdf)

N. Landwehr, M. Hall, and E. Frank (2005). Logistic Model Trees. *Machine Learning*, 59, 161–205.

R. Quinlan (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA.

R. Quinlan (1992). Learning with continuous classes. *Proceedings of the Australian Joint Conference on Artificial Intelligence*, 343–348. World Scientific, Singapore.

Y. Wang and I. H. Witten (1997). Induction of model trees for predicting continuous classes. *Proceedings of the European Conference on Machine Learning*. University of Economics, Faculty of

Informatics and Statistics, Prague.

I. H. Witten and E. Frank (2005). Data Mining: Practical Machine Learning Tools and Techniques.  
2nd Edition, Morgan Kaufmann, San Francisco.

## Προτάσεις για παρεμφερείς συναρτήσεις

Weka\_classifiers

## Παραδείγματα

```
m1 <- J48(Species ~ ., data = iris)

## print and summary
m1
summary(m1) # calls evaluate_Weka_classifier()

table(iris$Species, predict(m1)) # by hand

## visualization
## use partykit package
if(require("partykit", quietly = TRUE)) plot(m1)

## or Graphviz
write_to_dot(m1)

## or Rgraphviz
## Not run:
library("Rgraphviz")
ff <- tempfile()

write_to_dot(m1, ff)
plot(agraph(ff))

## End(Not run)
## Using some Weka data sets ...
```



```

## J48
DF2 <- read.arff(system.file("arff", "contact-lenses.arff",
                             package = "RWeka"))

m2 <- J48(`contact-lenses` ~ ., data = DF2)
m2
table(DF2$`contact-lenses`, predict(m2))
if(require("partykit", quietly = TRUE)) plot(m2)

## M5P
DF3 <- read.arff(system.file("arff", "cpu.arff", package = "RWeka"))
m3 <- M5P(class ~ ., data = DF3)
m3
if(require("partykit", quietly = TRUE)) plot(m3)

## Logistic Model Tree.
DF4 <- read.arff(system.file("arff", "weather.arff", package = "RWeka"))
m4 <- LMT(play ~ ., data = DF4)
m4
table(DF4$play, predict(m4))

## Larger scale example.
if(require("mlbench", quietly = TRUE)
  && require("partykit", quietly = TRUE)) {

## Predict diabetes status for Pima Indian women

data("PimaIndiansDiabetes", package = "mlbench")

## Fit J48 tree with reduced error pruning
      m5 <- J48(diabetes ~ ., data = PimaIndiansDiabetes,
                control = Weka_control(R = TRUE))
    }
plot(m5)
## (Make sure that the plotting device is big enough for the tree.)

```

## Περιγραφή

Διεπαφή της R προς τους αλγορίθμους συσταδοποίησης του Weka.

## Χρήση

`Cobweb(x, control = NULL)`

`FarthestFirst(x, control = NULL)`

`SimpleKMeans(x, control = NULL)`

`XMeans(x, control = NULL)`

`DBScan(x, control = NULL)`

## Ορίσματα

x	ένα αντικείμενο της R, τα δεδομένα του οποίου απαιτούνται να συσταδοποιηθούν.
control	αντικείμενο της κλάσης <code>Weka_control</code> που περιέχει τις επιλογές ελέγχου που αναμένεται να περάσουν σαν ορίσματα στον εκάστοτε εκμαθητή του Weka. Οι διαθέσιμες επιλογές μπορούν να αναζητηθούν είτε on-line χρησιμοποιώντας το Weka Option Wizard (WOW) είτε μέσω του εγγράφου τεκμηρίωσης του Weka.

## Λεπτομέρειες

Υπάρχει η μέθοδος `predict` για την πρόβλεψη των ταυτοτήτων των κλάσεων ή των ιδιοτήτων των μελών τους δια μέσου των ταιριασμένων συστάδων.

Coweb υλοποιεί τους αλγορίθμους συσταδοποίησης Cobweb (Fisher, 1987) και Classit (Gennari et al., 1989).

FartherFirst υλοποιεί τον αλγόριθμο “farthest first traversal algorithm” των Hochbaum και Shmoys, ο οποίος λειτουργεί ως ένας ταχύς, απλός προσεγγιστικός συσταδοποιητής, ο οποίος λαμβάνει χώρα στο δημιουργηθέν μοντέλο μετά την εφαρμογή του απλού k-means.

SimpleKMeans υλοποιεί συσταδοποίηση δια μέσου του αλγορίθμου k-means.

Xmeans υλοποιεί μία επέκταση του k-means μέσω του “Improve-Structure part”, καθορίζοντας αυτόματα το πλήθος των συστάδων.

DBScan υλοποιεί τον “density-based clustering algorithm” των Ester, Kriegel, Sander και Xu. Αξίζει να αναφερθεί πως σε αυτήν την περίπτωση, τα σημεία τα οποία θεωρούνται ως θόρυβος θεωρούνται ως ελλιπείς τιμές (NAs).

## Έξοδος

Μία λίστα η οποία κληρονομεί από την κλάση `Weka_clusterers` με τα εξής στοιχεία να περιλαμβάνονται σε αυτή:

<code>clusterer</code>	μία αναφορά (της κλάσης <code>jobRef</code> ) σε ένα αντικείμενο τύπου <code>Java</code> , η οποία δημιουργείται εφαρμόζοντας τη μέθοδο <code>buildClassifier</code> του <code>Weka</code> έτσι ώστε να υλοποιηθεί το συγκεκριμένο κάθε φορά μοντέλο χρησιμοποιώντας τις δοθείσες επιλογές ελέγχου.
<code>class_ids</code>	ένα διάνυσμα ακεραίων που υποδεικνύει την κλάση που τελικώς κάθε στιγμιότυπο έχει τοποθετηθεί (αποτελεί ουσιαστικά το αποτέλεσμα της κλήσης της μεθόδου <code>clusterInstance</code> του <code>Weka</code> για τον υλοποιημένο συσταδοποιητή όλων των στιγμιοτύπων).

## Προσοχή

XMeans απαιτεί το πακέτο `XMeans` έτσι ώστε να εγκατασταθεί επιτυχώς.

DBScan απαιτεί το πακέτο `optics_dbScan` έτσι ώστε να εγκατασταθεί επιτυχώς.

## Αναφορές

M. Ester, H.-P. Kriegel, J. Sander, and X. Xu (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96), Portland, OR, 226–231. AAAI Press.

D. H. Fisher (1987). Knowledge acquisition via incremental conceptual clustering. Machine Learning, 2/2, 139–172.

J. Gennari, P. Langley, and D. H. Fisher (1989). Models of incremental concept formation. Artificial Intelligence, 40, 11–62.

D. S. Hochbaum and D. B. Shmoys (1985). A best possible heuristic for the k-center problem, Mathematics of Operations Research, 10(2), 180–184.

D. Pelleg and A. W. Moore (2006). X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In: Seventeenth International Conference on Machine Learning, 727–734. Morgan Kaufmann.

I. H. Witten and E. Frank (2005). Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition, Morgan Kaufmann, San Francisco.

## Παραδείγματα

```
cl1 <- SimpleKMeans(iris[, -5], Weka_control(N = 3))
```

```
cl1
```

```
table(predict(cl1), iris$Species)
```

```
## Not run:
```

```
## Requires Weka package 'XMeans' to be installed.
```

```
## Use XMeans with a KDTree.
```

```
cl2 <- XMeans(iris[, -5],
```

```
             c("-L", 3, "-H", 7, "-use-kdtree",
```

```
             "-K", "weka.core.neighboursearch.KDTree -P"))
```

```
cl2
```

```
table(predict(cl2), iris$Species)
```

```
## End(Not run)
```

## Περιγραφή

Σύνολο επιλογών ελέγχου για τους εκμαθητές του Weka.

## Χρήση

`Weka_control(...)`

## Ορίσματα

... ορίσματα για να αποδοθούν οι διάφορες επιλογές ελέγχου. Περισσότερες πληροφορίες παρέχονται στις Λεπτομέρειες και τα Παραδείγματα.

## Λεπτομέρειες

Οι διαθέσιμες επιλογές για οποιονδήποτε εκμαθητή του Weka, λόγω χάριν για τον `foo()`, μπορούν να ερωτηθούν δια μέσου της εντολής `WOW(foo)` και να ορισθούν σχετικά εύκολα μέσω των επιλογών του `Weka_control()`. Ενδεικτικό παράδειγμα ακολουθεί στην αντίστοιχη παράγραφο.

Ενναλλακτικός τρόπος είναι η χρήση λιστών για τις επιλογές πολλαπλών ορισμάτων, όπως για παράδειγμα αναφέρεται στο έγγραφο τεκμηρίωσης του SMO.

## Έξοδος

Μία λίστα της κλάσης `Weka_control` η οποία μπορεί να μετατραπεί σε μορφή χαρακτήρων για να δοθεί σαν όρισμα στο Weka.

## Προτάσεις για παρεμφερείς συναρτήσεις

WOW

## Παραδείγματα

```
## Query J4.8 options:
```

```
WOW("J48")
```

```
## Learn J4.8 tree on iris data with default settings:
```

```
J48(Species ~ ., data = iris)
```

```
## Learn J4.8 tree with reduced error pruning (-R) and
```

```
## minimum number of instances set to 5 (-M 5):
```

```
J48(Species ~ ., data = iris, control = Weka_control(R = TRUE, M = 5))
```

---

[\*Weka\\_converters\*](#)

[\*R/Weka File Loaders and Savers\*](#)

---

## Περιγραφή

Διεπαφές της R προς συναρτήσεις φόρτωσης και αποθήκευσης αρχείων του Weka.

## Χρήση

```
C45Loader(file)
```

```
XRFFLoader(file)
```

```
C45Saver(x, file, control = NULL)
```

```
XRFFSaver(x, file, control = NULL)
```

## Ορίσματα

file	ένα μη κενό αλφαριθμητικό που περιέχει αποκλειστικά χαρακτήρες, οι οποίοι καθορίζουν το έγγραφο προς ανάγνωση ή προς εγγραφή.
x	αποτελεί τα δεδομένα προς εγγραφή, και είθισται να δίνεται σε μορφή πίνακα ή dataframe. Αν όχι εκτελείται αναγκαστικά προσπάθεια μετατροπής σε μορφή dataframe.
control	αντικείμενο της κλάσης Weka_control ή ένα διάνυσμα χαρακτήρων που περιέχει τις επιλογές ελέγχου ή ισούται με το NULL (default). Οι διαθέσιμες επιλογές μπορούν να αναζητηθούν είτε on-line χρησιμοποιώντας το Weka Option Wizard (WOW) είτε μέσω του εγγράφου τεκμηρίωσης του Weka.

## Λεπτομέρειες

C45Loader και C45Saver χρησιμοποιούν τη μορφή που έχει υιοθετήσει και ο C4.5 αλγόριθμος, όπου τα δεδομένα αποθηκεύονται σε δύο ξεχωριστά αρχεία, τα '.names' και τα '.data' αντίστοιχα.

XRFFLoader και XRFFSaver χειρίζονται XRFF αρχεία (eXtensible attribute – Relation File Format, το οποίο αποτελεί μία επέκταση, βασισμένη στο XML format, των αρχείων που χειρίζεται εξ ορισμού το Weka, τα Attribute – Relation File Format).

## Έξοδος

Εξ ορισμού παράγονται NULL μεταβλητές για τους savers.

Ένα dataframe το οποίο περιέχει τα δεδομένα του προς ανάγνωση αρχείου για τους loaders.

## Προτάσεις για παρεμφερείς συναρτήσεις

read.arff , write.arff

## Περιγραφή

Διεπαφές της R προς τα φίλτρα του Weka.

## Χρήση

`Normalize(formula, data, subset, na.action, control = NULL)`

`Discretize(formula, data, subset, na.action, control = NULL)`

## Ορίσματα

<code>formula</code>	συμβολική περιγραφή του εξεταζόμενου μοντέλου το οποίο πρόκειται να προσεγγιστεί. Πρέπει να αναφερθεί πως στην περίπτωση μη καθοδηγούμενων φίλτρων, η απόκριση μπορεί να παραλειφθεί.
<code>data</code>	προαιρετικό dataframe το οποίο περιέχει της μεταβλητές του μοντέλου.
<code>subset</code>	προαιρετικό διάνυσμα το οποίο καθορίζει ένα υποσύνολο των παρατηρήσεων το οποίο θα αξιοποιηθεί στη διαδικασία χτισίματος του μοντέλου.
<code>na.action</code>	συνάρτηση η οποία υποδηλώνει τις ενέργειες που θα εκτελεσθούν σε περίπτωση που τα δεδομένα περιέχουν ελλιπείς τιμές (NAs). Περισσότερες λεπτομέρειες μπορούν να αναζητηθούν στο <code>model.frame</code> .
<code>control</code>	αντικείμενο της κλάσης <code>Weka_control</code> που περιέχει τις επιλογές ελέγχου που αναμένεται να περάσουν σαν ορίσματα στον εκάστοτε εκμαθητή του Weka. Οι διαθέσιμες επιλογές μπορούν να αναζητηθούν είτε on-line χρησιμοποιώντας το Weka Option Wizard (WOW) είτε μέσω του εγγράφου τεκμηρίωσης του Weka.



## Λεπτομέρειες

Normalize υλοποιεί ένα μη καθοδηγούμενο φίλτρο, το οποίο κανονικοποιεί όλα τα στιγμιότυπα ενός συνόλου δεδομένων, δεδομένης μίας νόρμας. Για το σκοπό αυτό, μόνο οι αριθμητικές τιμές λαμβάνονται υπόψη, ενώ το χαρακτηριστικό που περιλαμβάνει την κλάση αγνοείται.

Discretize υλοποιεί ένα καθοδηγούμενο φίλτρο για τα στιγμιότυπα το οποίο διακριτοποιεί ένα εύρος αριθμητικών χαρακτηριστικών εντός του συνόλου δεδομένων σε ονομαστικά χαρακτηριστικά με συγκεκριμένες πλέον τιμές. Η μέθοδος διακριτοποίησης που χρησιμοποιείται εξ ορισμού είναι των Fayyad & Irani's MDL.

Σημαντική επισήμανση είναι πως οι παραπάνω μέθοδοι αγνοούν τα ονομαστικά χαρακτηριστικά, π.χ. οι μεταβλητές της κλάσης factor.

## Έξοδος

Ένα απλό dataframe.

## Αναφορές

U. M. Fayyad and K. B. Irani (1993). Multi-interval discretization of continuous-valued attributes for classification learning. Thirteenth International Joint Conference on Artificial Intelligence, 1022–1027. Morgan Kaufmann.

I. H. Witten and E. Frank (2005). Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition, Morgan Kaufmann, San Francisco.

## Παραδείγματα

```
## Using a Weka data set ...  
w <- read.arff(system.file("arff", "weather.arff",  
                           package = "RWeka"))  
  
## Normalize (response irrelevant)  
m1 <- Normalize(~., data = w)
```

m1

```
## Discretize
```

```
m2 <- Discretize(play ~., data = w)
```

m2

---

*Weka\_interfaces*

*R/Weka\_interfaces*

---

## Περιγραφή

Δημιουργία μίας R διεπαφής είτε για έναν υπάρχον εκμαθητή του Weka, είτε για κάποιον εκτιμητή ή κάποιο φίλτρο χαρακτηριστικών, είτε και για την εμφάνιση των διαθέσιμων σχετικών διεπαφών.

## Χρήση

```
make_Weka_associator(name, class = NULL, init = NULL)
```

```
make_Weka_attribute_evaluator(name, class = NULL, init = NULL)
```

```
make_Weka_classifier(name, class = NULL, handlers = list(),  
                      init = NULL)
```

```
make_Weka_clusterer(name, class = NULL, init = NULL)
```

```
make_Weka_filter(name, class = NULL, init = NULL)
```

```
list_Weka_interfaces()
```

```
make_Weka_package_loader(p)
```

## Ορίσματα

name	ένα αλφαριθμητικό όρισμα με χαρακτήρες το οποίο προσδιορίζει το όνομα της κλάσης του εξειδικευμένου εκμαθητή ή φίλτρου του Weka μέσω της σύμβασης JNI.
class	εξ ορισμού ισούται με το NULL. Διαφορετικά αποτελεί ένα διάνυσμα χαρακτήρων το οποίο καθορίζει τα ονόματα των κλάσεων για τα R αντικείμενα, τα οποία επιστρέφονται κάθε φορά από τις συναρτήσεις διεπαφής, από τις οποίες οφείλουν να κληρονομούν επιπρόσθετα χαρακτηριστικά και ιδιότητες πέραν από τις τυπικές κάθε φορά (έτσι ώστε να αναπαριστούν συσχετίσεις, ταξινομητές και συσταδοποιητές)
handlers	μία ονομαστική λίστα που περιέχει ειδικές συναρτήσεις χειρισμού. Περισσότερες πληροφορίες στην παράγραφο Λεπτομέρειες.
init	εξ ορισμού ισούται με το NULL. Διαφορετικά αποτελεί μία συνάρτηση χωρίς ορίσματα, η οποία καλείται όταν η διεπαφή χρησιμοποιείται για τη δημιουργία εκμαθητών ή φίλτρων. Τέλος, μπορεί να χρησιμοποιηθεί για ερωτοαπαντήσεις μέσω των διαθέσιμων επιλογών του WOW. Τυπικά χρησιμοποιείται για τη φόρτωση των πακέτων του Weka όταν επιτρέπεται η λειτουργία διεπαφών σε αυτά.
p	ένα αλφαριθμητικό χαρακτήρα το οποίο καθορίζει το πακέτο του Weka το οποίο απαιτείται να φορτωθεί μέσω του WPM.

## Λεπτομέρειες

make\_Weka\_associator και make\_Weka\_clusterer δημιουργούν μία συνάρτηση της R που παρέχει μία διεπαφή είτε σε έναν εκμαθητή συσχετίσεων του Weka είτε σε ένα συσταδοποιητή του Weka, αντίστοιχα. Αυτή η συνάρτηση διεπαφής έχει τυπικές παραμέτρους το x και το control = NULL, παρουσιάζοντας με αυτόν τον τρόπο τα στιγμιότυπα εκπαίδευσης και τις επιλογές ελέγχου που θα χρησιμοποιηθούν. Τα αντικείμενα που δημιουργούνται μέσω αυτών των συναρτήσεων διεπαφής κληρονομούν πάντοτε από τις κλάσεις Weka\_associator και Weka\_clusterer, αντίστοιχα, και μάλιστα

διακατέχονται τουλάχιστον από κατάλληλες print μεθόδους. Οι προσαρμοσμένοι συσταδοποιητές παρέχουν επιπλέον και μία μέθοδο predict.

`make_Weka_classifier` δημιουργεί μία συνάρτηση διεπαφής της R για έναν οποιονδήποτε ταξινομητή του Weka, που δέχεται σαν ορίσματα τα `formula`, `data`, `subset`, `na.action`, `control` (default: none), όπου τα πρώτα τέσσερα κατά σειρά ορίσματα επιδέχονται της συνηθισμένης ερμηνείας όπως όλες οι συναρτήσεις στατιστικής μοντελοποίησης της R, ενώ η τελευταία καθορίζει τις επιλογές ελέγχου που χρησιμοποιούνται από έναν εκμαθητή του Weka. Τα αντικείμενα που δημιουργούνται μέσω αυτών των συναρτήσεων διεπαφής κληρονομούν πάντοτε από τις κλάσεις `Weka_classifier`, αντίστοιχα, και μάλιστα διακατέχονται τουλάχιστον από κατάλληλες print και predict μεθόδους.

`make_Weka_filter` δημιουργεί μία συνάρτηση της R που παρέχει μία διεπαφή σε ένα φίλτρο του Weka, που δέχεται σαν ορίσματα τα `formula`, `data`, `subset`, `na.action`, `control` (default: none), όπου τα πρώτα τέσσερα κατά σειρά ορίσματα επιδέχονται της συνηθισμένης ερμηνείας όπως όλες οι συναρτήσεις στατιστικής μοντελοποίησης της R, ενώ η τελευταία καθορίζει τις επιλογές ελέγχου που χρησιμοποιούνται από έναν εκμαθητή του Weka. Προσοχή πως αν το φίλτρο είναι μη καθοδηγούμενο, η μεταβλητή απόκρισης μπορεί και να παραλειφθεί. Τα αντικείμενα που δημιουργούνται μέσω αυτών των συναρτήσεων διεπαφής ανήκουν (επί του παρόντος) πάντοτε στην κλάση `data.frame`.

`make_Weka_attribute_evaluator` δημιουργεί μία συνάρτηση διεπαφής για μία κλάση εκτίμησης χαρακτηριστικών του Weka, η οποία υλοποιεί την διεπαφή `AttributeEvaluator`, η οποία δέχεται ορίσματα όμοια με αυτά των συναρτήσεων διεπαφής για τους ταξινομητές.

Συγκεκριμένες πτυχές των συναρτήσεων διεπαφής μπορούν να παραμετροποιηθούν και να προσαρμοστούν με τη χρήση κατάλληλων χειρισμών (handlers). Προς το παρόν, μόνο οι χειρισμοί `control` χρησιμοποιούνται (συναρτήσεις που αποτελούν στοιχείο των `control` ως λίστες των χειρισμών) για την επεξεργασία των δοθέντων ορισμάτων ελέγχου πριν αυτές δοθούν σαν είσοδος στους διάφορους εκτιμητές του Weka. Αυτή η τακτική ακολουθείται για παράδειγμα από τους μετα-εκπαιδευτές έτσι ώστε να επιτρέψουν τον καθορισμό των καταγεγραμμένων βασικών εκμαθητών μέσω των τυπικών ονομάτων τους (παρά μέσω των πλήρως καταγεγραμμένων ονομάτων των Weka/Java κλάσεων που τους διακατέχουν).

Πέρα από τη δημιουργία συναρτήσεων διεπαφής, οι διεπαφές καταχωρούνται (υπό το όνομα των κλάσεων διεπαφής του Weka), γεγονός το οποίο επιτρέπει στο Weka Option Wizard (WOW) να παρέχει

σχετικά εύκολα την πρόσβαση πληροφοριών on-line σχετικά με τις διαθέσιμες επιλογές ελέγχου για την εκάστοτε διεπαφή.

`list_Weka_interfaces` περιλαμβάνουν σε λίστες τις διαθέσιμες διεπαφές.

Τέλος, το `make_Weka_package_loader` παράγει εντός του εαυτού του συνδέσμους για τη φόρτωση των απαιτούμενων, αλλά και των ήδη διαθέσιμων και εγκατεστημένων πακέτων του Weka.

Από τα προηγούμενα, προκύπτει άμεσα η ανάγκη καταχώρισης νέων διεπαφών, πέρα από τα ήδη υπάρχοντα πακέτα του RWeka που παρέχονται εξ ορισμού.

## Αναφορές

K. Hornik, C. Buchta, and A. Zeileis (2009). Open-source machine learning: R meets Weka. *Computational Statistics*, 24/2, 225–232.

## Παραδείγματα

```
## Create an interface to Weka's Naive Bayes classifier.
NB <- make_Weka_classifier("weka/classifiers/bayes/NaiveBayes")

## Note that this has a very useful print method:
NB

## And we can use the Weka Option Wizard for finding out more:
WOW(NB)

## And actually use the interface ...
if(require("e1071", quietly = TRUE) &&

require("mlbench", quietly = TRUE)) {

    data("HouseVotes84", package = "mlbench")
    model <- NB(Class ~ ., data = HouseVotes84)
```

```

predict(model, HouseVotes84[1:10, -1])
predict(model, HouseVotes84[1:10, -1], type = "prob")
}

## (Compare this to David Meyer's naiveBayes() in package 'e1071'.)

```

---

*Weka\_stemmers*

*R/Weka Stemmers*

---

## Περιγραφή

Διεπαφές της R προς τους stemmers του Weka.

## Χρήση

```

IteratedLovinsStemmer(x, control = NULL)
LovinsStemmer(x, control = NULL)

```

## Ορίσματα

x	ένα διάνυσμα χαρακτήρων που περιέχει λέξεις προς
control	αντικείμενο της κλάσης <i>Weka_control</i> ή ένα διάνυσμα χαρακτήρων που περιέχει τις επιλογές ελέγχου ή ισούται με το NULL (default). Οι διαθέσιμες επιλογές μπορούν να αναζητηθούν είτε on-line χρησιμοποιώντας το Weka Option Wizard (WOW) είτε μέσω του εγγράφου τεκμηρίωσης του Weka.

## Έξοδος

Ένα διάνυσμα χαρακτήρων με τις λέξεις.

## Αναφορές

J. B. Lovins (1968), Development of a stemming algorithm. *Mechanical Translation and Computational Linguistics* 11, 22–31.

---

*Weka\_tokenizers*

*R/Weka Tokenizers*

---

## Περιγραφή

Διεπαφές της R προς τα tokenizers (διαχωριστές λέξεων) του Weka.

## Χρήση

AlphabeticTokenizer(x, control = NULL)

NGramTokenizer(x, control = NULL)

WordTokenizer(x, control = NULL)

## Ορίσματα

x	ένα διάνυσμα χαρακτήρων που περιέχει λέξεις προς
control	αντικείμενο της κλάσης <i>Weka_control</i> ή ένα διάνυσμα χαρακτήρων που περιέχει τις επιλογές ελέγχου ή ισούται με το NULL (default). Οι διαθέσιμες επιλογές μπορούν να αναζητηθούν είτε on-line χρησιμοποιώντας το Weka Option Wizard (WOW) είτε μέσω του εγγράφου τεκμηρίωσης του Weka.

## Έξοδος

Ένα διάνυσμα χαρακτήρων με τις λέξεις.

## Περιγραφή

Παροχή on-line πληροφοριών σχετικά με τις διαθέσιμες επιλογές ελέγχου που αφορούν τους εκμαθητές ή τα φίλτρα του Weka και τις διεπαφές της R.

## Χρήση

WOW(x)

## Ορίσματα

x ένα αλφαριθμητικό όρισμα με χαρακτήρες το οποίο προσδιορίζει είτε το πλήρως εξειδικευμένο όνομα της κλάσης κάποιου εκμαθητή ή φίλτρου του Weka μέσω της σύμβασης JNI, είτε το όνομα μίας διαθέσιμης διεπαφής της R, είτε τελικά ένα αντικείμενο το οποίο δημιουργείται από την εφαρμογή των αντίστοιχων διεπαφών για το χτίσιμο ενός συσχετιστή, ταξινομητή, συσταδοποιητή ή φίλτρο.

## Λεπτομέρειες

Περισσότερες πληροφορίες για τις διαθέσιμες συναρτήσεις διεπαφής παρέχονται στο `list_Weka_interfaces`.

## Αναφορές

K. Hornik, C. Buchta, and A. Zeileis (2009). Open-source machine learning: R meets Weka. *Computational Statistics*, 24/2, 225–232.

## Παραδείγματα

## The name of an "existing" (registered) interface.



WOW("J48")

## The name of some Weka class (not necessarily in the interface

## registry):

WOW("weka/classifiers/bayes/NaiveBayes")

---

*WPM*

*Weka Package Manager*

---

## Περιγραφή

Διαχείριση των πακέτων του Weka.

## Χρήση

WPM(cmd, ...)

## Ορίσματα

cmd	ένα αλφαριθμητικό χαρακτήρων το οποίο προσδιορίζει τις ενέργειες προς εκτέλεση. Υποχρεωτικά πρέπει να επιλεγθεί ένα εκ των “refresh-cache”, “list-packages”, “package-info”, “install-package”, “remove-package” ή “load-package” (ή μία αποκλειστική συντόμευση αυτών)
-----	---

...	αλφαριθμητικά χαρακτήρων τα οποία παρέχουν περισσότερα ορίσματα τα οποία απαιτούνται για τις ενέργειες που χρειάζεται κάθε φορά να εκτελεστούν. Περισσότερες οδηγίες ακολουθούν στην παράγραφο Λεπτομέρειες.
-----	--

## Λεπτομέρειες

Τα διαθέσιμα και αντίστοιχα τα επιπρόσθετα ορίσματα που προσφέρονται, περιγράφονται παρακάτω:

“refresh-cache” ανανεώνει το αντίγραφο της προσωρινής μνήμης του πακέτου των μετα-δεδομένων από το κεντρικό αποθετήριο πακέτων.

“list-packages” τυπώνει πληροφορίες (τον αριθμό έκδοσης και μικρές περιγραφές) σχετικές με τα πακέτα όπως αυτά ορίζονται από μία επιπρόσθετη λέξη κλειδί, η οποία πρέπει να είναι μια εκ των “all” (όλα τα πακέτα που αναγνωρίζει το σύστημα), “installed” (όλα τα πακέτα που έχουν εγκατασταθεί τοπικά), «available” (όλα τα γνωστά πακέτα που δεν έχουν εγκατασταθεί τοπικά) ή μία αποκλειστική συντόμηση των προηγούμενων.

“package-info” τυπώνει πληροφορίες (μετα-δεδομένα) σχετικά με ένα πακέτο. Απαιτεί δύο επιπρόσθετα αλφαριθμητικά χαρακτήρων ως ορίσματα: μία λέξη κλειδί και το όνομα του πακέτου. Ως λέξη κλειδί μπορεί να χρησιμοποιηθεί το “repository” (τυπώνει πληροφορίες από το αποθετήριο), το “installed” (τυπώνει πληροφορίες για την εγκατεστημένη έκδοση) ή μία αποκλειστική συντόμηση των προηγούμενων.

“install-package” εγκαθιστά ένα πακέτο, με το όνομα αυτού να ορίζεται από ένα πρόσθετο αλφαριθμητικό χαρακτήρων. (εναλλακτική επιλογή αποτελεί η χρήση ενός μονοπατιού για το αρχείο ή ενός URL προς ένα συμπιεσμένο αρχείο.)

“remove-package” αφαιρεί ένα (εγκατεστημένο) πακέτο.

“load-package” φορτώνει το δοθέν πακέτο προσθέτοντας παράλληλα τα αρχεία jar που το συνοδεύουν στο Java classpath.

## Προσοχή

Το Weka αποθηκεύει τα πακέτα και τις πληροφορίες αυτών στον αρχικό κατάλογο του, όπως αυτός ορίζεται από την τιμή της μεταβλητής environment WEKA\_HOME. Εάν αυτή δεν έχει οριστεί, ο υποκατάλογος ‘wekafiles’ του αρχικού καταλόγου του χρήστη χρησιμοποιείται για αυτόν το σκοπό. Στην περίπτωση ο αρχικός κατάλογος δεν έχει δημιουργηθεί ακόμη, η εντολή WPM() θα χρησιμοποιήσει στη θέση αυτών ένα προσωρινό κατάλογο εντός του τρέχων καταλόγου της R. Για να μην δημιουργηθούν προβλήματα με πολλαπλές θέσεις του αρχικού καταλόγου, ο κάθε χρήστης καλείται να δημιουργήσει ένα κύριο αρχικό κατάλογο για το Weka προτού χρησιμοποιήσει την εντολή WPM().

## Παραδείγματα

```
## Not run:  
## Start by building/refreshing the cache.  
WPM("refresh-cache")  
  
## Show the packages installed locally.  
WPM("list-packages", "installed")  
  
## Show the packages available from the central Weka package  
## repository and not installed locally.  
WPM("list-packages", "available")  
  
## Show repository information about package XMeans.  
WPM("package-info", "repository", "XMeans")  
  
## End(Not run)
```

---

*write.arff*

*Write Data into ARFF Files*

---

## Περιγραφή

Εγγραφή δεδομένων εντός των Attribute-Relation File Format αρχείων του Weka.

## Χρήση

```
write.arff(x, file, eol = "\n")
```

## Ορίσματα

x                    τα προς εγγραφή δεδομένα, κατά προτίμηση σε μορφή πίνακα ή dataframe. Αν δεν εισέλθουν με τις παραπάνω μορφές, τότε πραγματοποιείται αναγκαστική μετατροπή σε dataframe.

file	είτε αλφαριθμητικό χαρακτήρων ονομάζοντας ένα αρχείο ή μία σύνδεση. Τα σύμβολα “ ” υποδεικνύουν την έξοδο σε συγκεκριμένη σύνδεση εξόδου.
eol	αποτελεί τον/τους χαρακτήρα/ες που επιλέγονται να τυπώνονται μετά το τέλος κάθε γραμμής ή στήλης κάθε φορά.

## **Αναφορές**

Attribute-Relation File Format <http://weka.wiki.sourceforge.net/ARFF>

## **Προτάσεις για παρεμφερείς συναρτήσεις**

read.arff

## **Παραδείγματα**

```
write.arff(iris, file = " ")
```