Wekaの 導入と 実行

Wekaのダウンロード

公式ページにアクセスします

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



Machine Lea

Project Software Book

Publicatior

Home

Getting started Requirements Download Documentation FAQ Citing Weka

Further information Datasets Related Projects Miscellaneous Code Other Literature

Developers Development History Subversion Contributors

Various Wekalist stats

Weka 3: Data Mining Software in Java

Weka is a collection of machine learning algorithms for data mining tasks. The directly to a dataset or called from your own Java code. Weka contains tools 1 regression, clustering, association rules, and visualization. It is also well-suite learning schemes.

Weka is open source software issued under the GNU General Public License.

Pentaho's live forum for Weka

The open-source BI software company Pentaho is a major sponsor of Weka dev for interaction among Weka project community members.

The Weka mailing list

Please post Weka-related questions, comments, and bug reports to the **Weka** the **online documentation** first, before posting to the list). There is also the s (Mirrors: **news.gmane.org**, **Nabble**). Please do not email individual members problems.

Also, please have in mind that your message will be sent to several thousand I the **Mailing List Etiquette**. The administrator also removes members from the

Wekaのダウンロード

 左の欄からDownloadを選ぶと、下の方にDeveloper version があるので、使ってるパソコンのbit数に合ったもののJava VM付きのものをダウンロードします

Developer version

This is the trunk of Weka and continues from the stable-3-6 code line. It receives both bug fixes and new features.

Windows x86

Click here to download a self-extracting executable that includes Java VM 1.6 (weka-3-7-7jre.exe; 37.4 MB)

Click **here** to download a self-extracting executable without the Java VM (weka-3-7-7.exe; 22.6 MB)

These executables will install Weka in your Program Menu. Download the second version if you already have Java 1.6 (or later) on your system.

Windows x64

Click here to download a self-extracting executable that includes 64 bit Java VM 1.6 (weka-3-7-7 jre-x64.exe; 37.7 MB)

Click **here** to download a self-extracting executable without the Java VM (weka-3-7-7-x64.exe; 22.6 MB)

These executables will install Weka in your Program Menu. Download the second version if you already have Java 1.6 (or later) on your system.

※上にある3rd ed. Versionは日本語対応ですが不具合があります(日本語ロケールの問題?)

Wekaのインストール

ダウンロードが済んだらexeファイルを実行します



Wekaのインストール

ライセンスに同意します

6	Weka 3.7.7 Setup
Weka	License Agreement Please review the license terms before installing Weka 3.7.7.
Press Page Down to see the	e rest of the agreement.
GNU GENERAL Version 3, 29	PUBLIC LICENSE
Copyright (C) 2007 Free Si Everyone is permitted to co of this license document, b	oftware Foundation, Inc. < <u>http://fsf.org/</u> > opy and distribute verbatim copies out changing it is not allowed.
Preamble	
The GNU General Public Lic software and other kinds o	cense is a free, copyleft license for fworks.
If you accept the terms of t agreement to install Weka 3	he agreement, dick I Agree to continue. You must accept the 3.7.7.
Nullsoft Install System v2.46 –	
	< <u>B</u> ack I <u>Ag</u> ree Cancel

Wekaのインストール

• "全てインストール"にして先に進みます

0	Weka 3.7.7 Setup				
Weka	Choose Components Choose which features of Weka 3.7.7 you want to install.				
Check the components you install. Click Next to continu	want to install and uncheck the co ue.	omponents you don't want to			
Select the type of install:	Full	•			
Or, select the optional components you wish to install:	Associate Files Install JRE	Description Position your mouse over a component to see its description,			
Space required: 78.8MB					
Nullsoft Install System v2,46 -					
	< <u>B</u> ack	Next > Cancel			

Wekaのインストール

• インストール場所を設定します

😚 Weka 3	3.7.7 Setup			
Choose Insta Choose the fo	all Location older in which to install Weka 3.7.7.			
Setup will install Weka 3.7.7 in the following folder. To install in a different folder, click Browse and select another folder. Click Next to continue.				
Destination Folder				
C:¥Program Files¥Weka-3-7	Browse			
Space required: 78.8MB Space available: 834.6GB				
Nullsoft Install System v2.46	< Back Next > Cancel			

Wekaのインストール

スタートメニュー・ショートカットの設定をします

Weka 3.7.7 Setup Image: Control of the setup
Choose Start Menu Folder Choose a Start Menu folder for the Weka 3.7.7 shortcuts.
Select the Start Menu folder in which you would like to create the program's shortcuts. You can also enter a name to create a new folder.
7-Zip Accessibility Accessories Administrative Tools Android SDK Tools AVG AVG 2012 Avira Buffalo CANON IMAGE GATEWAY 無料会員登錄 MX870 Canon MX870 series
Do not create shortcuts Nullsoft Install System v2,46 < <u>Back</u> Install Cancel

Javaのインストール

インストール途中にJavaのインストールが始まります。

đ	Java セットアップ - よ	うこそ	X
(java			ORACLE
Java(TM) へようこそ			
Java は素晴らしい Java コンラ ンから役に立つユーティリティー 験を真に実現します。	テンツの世界への安全なアクセス やエンターテイメントに到るまで、	、を提供します。 ビジ Java はお客様の・	ジネスソリューショ インターネット作
インストールプロセスにおいて値 いては、http://java.com/data	国人情報が収集されることはあり すをご覧ください)ません。収集する1	青華服の言羊糸田につ
Java をインストールするには「~	インストール」をクリックしてくださし	۱,	
「インストール先フォルダの変」	Ð	キャンセル(C)	インストール(<u>1</u>) >

Javaのインストール

インストールが終わったらウィンドウを閉じま

谩	Java セットアップ - 完了	X
Java ⁻		ORACLE
Java のインストールに	成功しました。	
最新機能や改善された ンロードされます。これ?	セキュリティーを提供するために、Java の更新版 を変更する場合は、http://java.com/autoupdate を	が自動的にダウ ご覧ください
		開いる

Wekaのインストール

インストールが終わったらNextを押して先に 進みます

7	Weka 3.7.7 Setup
Weka	Installation Complete Setup was completed successfully.
Completed	
Output folder: C:¥Prog Execute: RunJREInstall Delete file: C:¥Program Created uninstaller: C:¥ Output folder: C:¥User Create shortcut: C:¥Us Create shortcut: C:¥Us Create shortcut: C:¥Us Create shortcut: C:¥Us Create shortcut: C:¥Us Create shortcut: C:¥Us	am Files¥Weka-3-7 er.bat Files¥Weka-3-7¥RunJREInstaller.bat Program Files¥Weka-3-7¥uninstall.exe s¥NewServer¥AppData¥Roaming¥Microsoft¥Windows¥Star ers¥NewServer¥AppData¥Roaming¥Microsoft¥Windows¥Star ers¥NewServer¥AppData¥Roaming¥Microsoft¥Windows¥Star ers¥NewServer¥AppData¥Roaming¥Microsoft¥Windows¥Star
Nullsoft Install System v2,46	< <u>B</u> ack <u>Next</u> > Cancel

Wekaのインストール

Start Wekaにチェックを入れてFinishを押します。うっかり閉じた場合は(C:¥Program Files¥Weka-3-7にあるWekaa3.7を実行します)



 Wekaを実行すると警告が出ますが、チェック を入れて了解を押します



- これがWekaの起動画面です。
- Explorerを押します



この画面からデータを開いたりデータマイニン
 グをしたりします

Weka Explorer Preprocess Classify	Cluster Associate	Select attributes V	(isualize			
Open file	Open URL	Open DB	Gene	rate Undo	Edit	Save
Filter Choose None						Apply
Current relation Relation: None Instances: None		Attributes Sum of weights	∷ None ∷ None	Selected attribute Name: None Missing: None	Distinct: None	Type: None Unique: None
Attributes	None	Invert Patt	ern			
						✓ Visualize All
	Remove					
tatus Velcome to the Weka I	Explorer					Log 💉 X

データセットの作成

例としてbounce ballのデータセットを作成し、
 データマイニングをしてみましょう

Ball	Size	Color	Weight	Rubber?	Result(Bounce?)
1	Small	Green	Light	Yes	Yes
2	Small	Blue	Medium	No	No
3	Medium	Red	Medium	No	No
4	Small	Red	Medium	Yes	Yes
5	Large	Green	Heavy	Yes	Yes
6	Medium	Blue	Heavy	Yes	No
7	Medium	Green	Heavy	Yes	No
8	Small	Red	Light	No	No

データセットの書式

• Wekaではarff形式のファイルが推奨されている



※dataフォルダ内のサンプルも参考にしてもよい ※Wekaでは大文字小文字を区別します ※csv形式のファイルも読み込めますが、推奨はされません

データセットの読み込み

• Preprocessタブを開き、Open fileを押します

0	Weka E	xplorer		
Preprocess Classify Cluster	Associate Select attributes Visualize			
Open file Oper	n URL Open DB Gener	rate Undo	Edit	Save
Filter Open a set	t of instances from a file			Apply
Current relation		Selected attribute		
Relation: None Instances: None	Attributes: None Sum of weights: None	Name: None Missing: None	Distinct: None	Type: None Unique: None
Attributes	-			
All None	Invert Pattern			
				▼ Visualize All
	Remove			
	T WHICH'S			
Status Welcome to the Weka Explore	ïr			Log 💉 V

データセットの読み込み

いま、作成したBounceBallのデータセットを読み込みます

☑ 開<		×
参照:	🔑 weka	🔹 🤌 📂 🎞 -
最近使った項 目 デスクトップ	 01 02 03 04 05 06 ai data - ショートカット 	Invoke options dialog Note: Some file formats offer additional options which can be customized when invoking the options dialog.
ער איז די אינאבא די בטעב-א-	 ball.arff old_ball.arff 	
く ネッドワーク	レ ファイル名: ball.arff ファイルのタイプ: Arff data files (*.arff)	■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■

データセットの読み込み

 書式が間違っている場合は、以下の様なエ ラーが出るので、間違っている部分を修正し ましょう

Load Ins	stances
	File 'C:¥Users¥admin¥Desktop¥weka¥ball.arff' not recognised as an 'Arff data files' file. Reason: nominal value not declared in header, read Token[Yes], line 10
	OK Use Converter



WCKd	Explorer	
Preprocess Classify Cluster Associate Select attributes Visualize		
Open file	nerate Undo (Edit Save
Filter		
Choose None		Apply
Current relation Relation: weather Attributes: 5 Instances: 14 Sum of weights: 14	Selected attribute Name: outlook Missing: 0 (0%) Dist	Type: Nominal inct: 3 Unique: 0 (0%)
Attributes	No. Label	Count Weight
	1 sunny	5 5.0
All None Invert Pattern	2 overcast	4 4.0
	3 rainy	5 5.0
3 humidity 4 windy 5 play	Class: play (Nom)	Visualize All
Remove		
ok		Log 💉 🛛



Classifyタブを開き、Chooseを押します

0	Weka Explorer	
Preprocess Classify Cluster Associate Select	attributes Visualize	
Choose ZeroR		
Test options	Classifier output	
🔘 Use training set		
O Supplied test set Set		
Cross-validation Folds		
O Percentage split % 66		
More options		
(Nom) play		
Basult list (vielt slick for estima)		
Result list (right-click for options)		
Status		
ОК	Log	🚽 🛷 × 0

アルゴリズムの指定

 アルゴリズムの一覧が表示されるので、treesフォル ダの中にあるRandom Treeを選択します

Weka Explorer						
Preprocess Classify Cluster Associate Select attribute	s Visualize					
Classifier						
🔰 🕪 weka						
🔚 📄 퉲 classifiers						
👘 連 bayes	er output					
🗈 🧶 functions						
lazy						
the second se						
DecisionStump						
→ J48						
(+ LMT						
₩5P						
RandomForest						
RandomTree						
REPTree						
Close						

評価方法の指定

 初期設定ではCross-validation(交差検定)に なっていますが、Use training setに変更しま す(全てを学習データとする)

Test options				
💿 Use training set				
Supplied test set	Set			
Cross-validation	Folds	10		
🔘 Percentage split	%	66		
More options				

Startを押すと解析を実行します

Result list (right-click for options)									
13:35:47 - trees.RandomTree		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	I	
13:38:51 - trees.RandomTree		1	0	1	1	1	1	()	
13:39:07 - trees.RandomTree		1	0	1	1	1	1	1	
	Weighted Avg.	1	0	1	1	1	1	(
	=== Confusion M a b < clas 3 0 a = Yes 0 5 b = No	atrix === sified as						E	
	•							•	

結果が右側に表示されたら、結果のリストを 右クリックし、Visualize treeを選択します

Cross-validation Folds 10	Root mean squared error			0.19	25			
Percentage colit K 66	Relative absolute error		23.33	33 %				
	Root relative squ	ared err	or	40.01	63 %			
More options	Coverage of cases	(0.95 1	evel)	100	8			
	Mean rel. region size (0.95 level)			66.6667 %				
(Nom) Bounce 🔹	Total Number of I	Instances		3				
Start Stop	=== Detailed Accu	aracy By	Class ===					
Result list (right-click for options)								
10:58:03 - trees.	1	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	I
view in main windo	w	p p	0	1	1	1	1	-
View in separate wi	ndow	1	0	1	1	1	1	-
Save result buffer		1	0	1	1	1	1	1
Delete result buffer		rix ===						
Lond model	fied as							
Load model								
Save model								
Re-evaluate model	on current test set							
Re-apply this mode	l's configuration							
								b.
Visualize classifier e	rrors							
Status Visualize tree								
OK Visualize margin cu	rve					Log	-	2
Visualize threshold	000/0						-	_

データのbounceにおける決定木が生成され ました

RandomTree

Size = Small
| Rubber = Yes : Yes (2/0)
| Rubber = No : No (2/0)
Size = Medium : No (3/0)
Size = Large : Yes (1/0)

Size of the tree : 6



新しいデータの予測

データセットに、不明なデータを追加します。 不明なデータは?と書く

@data↓ Small,Green,Light,Yes,Yes↓ Small,Blue,Medium,No,No↓ Medium,Red,Medium,No,No↓ Small,Red,Medium,Yes,Yes↓ Large,Green,Heavy,Yes,Yes↓ Medium,Blue,Heavy,Yes,No↓ Medium,Green,Heavy,Yes,No↓ Small,Red,Light,No,No↓ Small,Blue,Light,Yes,?←



 More Optionsを開き、Output predictionsを PlainTextに変えます

Choose RandomTree -K 0 -M 1.0 -S	Classifier evaluation options				
Test options	☑ Output model				
Ouse training set	✓ Output per-class stats				
Cross-validation Folds 10	Output entropy evaluation measures				
More options	Output confusion matrix				
	Store predictions for visualization				
(Nom) Bounce	Output predictions Choose PlainText				
Result list (right-click for options)	Cost-sensitive evaluation Set				
13:35:47 - trees.RandomTree	Random seed for XVal / % Split 1				
	Preserve order for % Split				
	Output source code WekaClassifier				
	ок				

予測結果

マイニングを実行すると、不明な部分がYesと 予測されました

=== Predictions on training set ===

inst#	actual	predicted	error	pre	diction	1
1	1:Yes	1:Yes		1		
2	2:No	2:No		1		
3	2:No	2:No		1		
4	1:Yes	1:Yes		1		
5	1:Yes	1:Yes		1		
6	2:No	2:No		1		
7	2:No	2:No		1		
8	2:No	2:No		1	1	
9	1:?	1:Yes		1		