ID Tree generation and visualization with Weka (run example in Weka)

Identification Trees

Last, but not least, an ID tree is a decision tree in which all possible divisions is created by training the tree against a list of known data. The purpose of an ID tree is to take a set of sample data, classify the data and construct a series of test to classify an unknown object based on like properties.



Refined Rules



How to get and start Weka

• Go to this page

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

Download Weka from

<u>Weka-3.7</u>

Weka-download .exe file

- Start Weka
- Press Explorer



Weka Setup

• After download, please run exe file.



• Agree the license

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

• Select the type of install "Full" and go next.

()	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 contin	u want to install and uncheck the com ue.	nponents 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			

• Set install directory.

🕞 Weka 3.7.7 Setup			
Choose Install Location Choose the folder 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 Dext > Cancel			

• Set to add start menu or not, to create shortcut or not.

)	Weka 3.3	7.7 Setup		
Weka	Choose Start N Choose a Start N	lenu Folder 1enu folder for th	e Weka 3.7.7 s	hortcuts.
Select the Start Menu fo can also enter a name to Weka 3.7.7	lder in which you wou o create a new folder.	ld like to create th	ne program's sh	ortcuts. You
7-Zip Accessibility Accessories Administrative Tools Android SDK Tools AVG AVG 2012 Avira Buffalo CANON iMAGE GATEWA Canon MX870 series	YY 無料会員登録 MX8	370		< III
Do not create shortco iullsoft Install System v2.4	uts 6	< <u>B</u> ack	<u>I</u> nstall	Cancel

JRE

• Before installing Weka, JRE's installation starts.

i ²	Java セットアップ - よ	うこそ	X
(S) 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。	シ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1	ネスソリューショ インターネット体
インストールプロセスにおいて個 いては、http://java.com/data	目人情報が収集されることはあり をご覧ください)ません。収集する†	青幸服の言羊糸田(こつ
Java をインストールするには「ィ	(ンストール」をクリックしてくださし) ₀	
「インストール先フォルダの変更		キャンセル(の)	インストール(<u>1</u>) >

• After JRE's install, close the window.



• After installing JRE, Weka installation starts.

0	Weka 3.7.7 Setup
S Weka	Installation Complete Setup was completed successfully.
Completed	
Output folder: C:¥P Execute: RunJREIns Delete file: C:¥Prog Created uninstaller: Output folder: C:¥U Create shortcut: C: Create shortcut: C: Create shortcut: C: Create shortcut: C: Create shortcut: C:	rogram Files¥Weka-3-7 staller.bat ram Files¥Weka-3-7¥RunJREInstaller.bat C:¥Program Files¥Weka-3-7¥uninstall.exe Isers¥NewServer¥AppData¥Roaming¥Microsoft¥Windows¥Star ¥Users¥NewServer¥AppData¥Roaming¥Microsoft¥Windows¥Star ¥Users¥NewServer¥AppData¥Roaming¥Microsoft¥Windows¥Star ¥Users¥NewServer¥AppData¥Roaming¥Microsoft¥Windows¥Star
Nullsoft Install System v2	2,46 < <u>B</u> ack Next > Cancel

• Click Finish button after checking Start Weka.



Start Weka

• At the first time, the warning like below is appeared. Please check and click ok.

-	Weka GUIChooser
	Weka has a package manager that you can use to install many learning schemes and tools. The package manager can be found under the "Tools" menu. Do not show this message again
	了解

Use Explorer

- This is the Weka GUI.
- To start data mining, click Explorer.



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					
Status Welcome to the Weka	e Explorer					Log ×

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

ball.csv \rightarrow ball.arff open ball.csv in openFile and save .arff format

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 Open URL	Open DB Gener	rate Unde	Edit	Save
Filter Open a set of instan Choose None	ces from a file			Apply
Current relation		Selected attribute		
Relation: None Instances: None	Attributes: None Sum of weights: None	Name: None Missing: None	Distinct: None	Туре: None Unique: None
Attributes				
All None	Invert Pattern			
		[▼ Visualize All
				Visualize Thi
Remove]			
I L INCINOVE				
Status Welcome to the Weka Explorer				Log × 0

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	🔹 🤌 📂 🎞 •
9	01	Invoke options dialog
最近使った項目	03 04	Note: Some file formats offer additional
	0 5	options which can be customized when invoking the options dialog.
デスクトップ	ai	
マイドキョベット		
	Old_ball.artt	
レー ター		
	ファイル名: ball artf	
ネットワーク	ファイルのタイプ: Arff data files (*arff)	· Ⅲ八 • 取消

ball.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	itances
8	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



Classifyタブを開き、Chooseを押します

0	Weka Explorer	
Preprocess Classify Cluster Associate Select a	attributes Visualize	
Choose ZeroR		
Test options	Classifier output	
🔘 Use training set		
Supplied test set Set		
Cross-validation Folds		
Percentage split % 66		
More options		
(Nom) play		
Beauty list (sinks aligh for antions)		
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

O 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	atrix ===						
	a b < clas	sified as						E
	$0.5 \mid b = No$							
								-
	•		III					•

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-validatio	n Folds 10	Root mean squared	i error		0.19	25			
Percentage soli	it % 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)	66.66	67 %			
(Nom) Bounce	-	Total Number of I	instances		3				
Start	Stop	=== Detailed Accu	iracy By	Class ===					
Result list (right-cli	ick for options)		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	Ţ
10:58:03 - trees.	View in main window	w	1	0	1	1	1	1	1
	View in separate wi	ndow	5	0	1	1	1	1	
	view in separate wi	luow	5	0	1	1	1	1	
	Save result buffer		ŕ	о - С	-	-	1	-	
	Delete result buffer		rix ===						
	Load model		fied as						
	Save model								
	Re-evaluate model	on current test set							
	De eveluthie seedel								
	Re-apply this model	s configuration							
	Visualize classifier e	rrors							Þ.
Status	Visualize tree								
ок	Visualize margin cur	ve					Log	-00	~
	Vieualize threehold of								

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



Predict new data

• Add unknown data in the dataset. Unknown data is presented in ?.

@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,Ne,Ne↓ Small,Blue,Light,Yes,?←

Add setting

• Click More Options and change Output predictions to PlainText from null.

Choose RandomTree -K 0 -M 1.0 -S	Classifier evaluation options
Test options O Use training set	V Output model
⊙ Supplied test set Set	☑ Output per-class stats
Cross-validation Folds 10	Output entropy evaluation measures
More options	Output confusion matrix
(Nom) Bounce	Store predictions for visualization
	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

Result of prediction

• Start mining, the unknown data is predicted "Yes".

=== Predictions on training set ===

inst#	actual	predicted	error	pre	diction
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	2:10	2:No		1	
9	1:?	1:Yes		1	

Ball	Size	Color	Weight	Rubber?	Result (Bounces?)
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

Figure C1: Identification Tree Training Data

Work in class:

Implement and generating ID tree on the bouncing ball dataset, ball.arff using Weka.

(correctly generate ball.arff file and then use Weka)





Home Work

1. Please read the article <u>decision-tree-article</u> (no submission)