### Wekaを実行する 練習その1

静岡大学大学院理工学研究科 阿部 秀尚 hidenao@ks.cs.inf.shizuoka.ac.jp http://panda.cs.inf.shizuoka.ac.jp/~hidenao/

#### 練習問題

・ 設問 1:用語説明

 - データマイニングでの用語の定義を確認する

 ・ 設問 2:決定木を実行してみる

 - 基本的で結果が明示的なモデルで得られる決定木を実行してみる

#### 設問1

- 下記の用語を簡潔にわかりやす (説明く ださい)
  - ・ データマイニング
  - 属性
  - ・ クラス
  - インスタンス
  - 学習 (される)モデル

#### 設問2

- Wekaに付属するデータセット(任意のデー タセット)に対して決定木アルゴリズムを適 用する
  - データセットの仕様 (名前,属性数[名義属性: 数値属性],インスタンス数,クラス数,問題設定 (クラスは何を意味しているか))
  - 結果の画面 (テキスト,図)
  - 結果の説明・考察

## 練習問題の解答

#### 設問1の回答

• データマイニング

- – 蓄えられたデータから有用なパターンや知識を見つけ
   出すこと(何が 有用」かはデータマイニングが行われ
   る場面によって決まる)
- 属性
  - 観測される各事例を様々な側面から特徴づけするための変数
  - 名義値が与えられる名義属性と数値が与えられる数 値属性の2種類が一般的に扱われる
  - 説明変数とも呼ばれる

#### 設問1の回答

- ・クラス
  - ある視点をもってインスタンスを分類した際につけられたラベル
  - 目標概念や目的変数と呼ばれる
- インスタンス
  - 各属性に具体的な値(数値,名義値)与えたもの
  - 観測される1事例が1インスタンスに相当する
- 学習されるモデル
  - 所与のデータセットから得られるタスクに必要な情報 (分類タスクならば、分類ルールや決定木)

# 練習問題 設問2の解答 !!Wekaのヴァージョンは3-2-3です!!!

#### 属性数を数える

#### Base relationのところにAttributes(属性数)として出ている 後でクラスの分を引く

Preprocess Class	sify Cluster Assoc	iate Select attribu	tes Visualize		
Open file	Open URL	Open DB	Apply Filters	Replace	Save
Base relation Relation: weather Instances: 14	Attributes	s: 5	Working relation Relation: weather Instances: 14	Attri	ibutes: 5
Attributes in base rela	ation		Filters		
All	None	Invert	AddFilter - N unnamed -	C 0	A
No. 1 vottook 2 vtempera 3 vhumidity 4 vindy	Name ture			Delete	
5 🗹 play			Attribute info for base i Name: outlook Missing: 0 (0%)	elation Distinct: 3	Type: Nominal Unique: 0 (0%)
			Label		Count
			sunny	5	
			overcast	4	
			Tanty	10	

名義属性を数える

Open file									
		Open URI		Open DB	Apply Filters	Replace	Save		
ase relation Relation: weather Instances: 14 Attributes: 5 Itributes in base relation					Working relation Relation: weather				
Attributes in base relation					ilters				
All Nono Imort			Iment	ddFilter -N unnamed -	CO	Add			
				Type	Name: outlook Missing: 0 (0%) Label wercast	Distinct: 3 nalならば 属性の分			
og 2:11:32: email: v 2:11:32: Started 2:12:05: Base re 2:12:05: Working	wekasupp on 木曜日 elation is r g relation i	ort@cs.wa , 3 7月 200 iow weathe is now wea	kato.ac.nz 3 r (14 instanc ther (14 insta	es) ances)					

数値属性を数える

Preprocess	Classify	Cluster	Associate	Select attributes	Visualize		
Open file.	•	Open UR	L	Open DB	Apply Filters	Replace	Save
ase relation Relation: we Instances: 14	eather	А	attributes: 5	rv	Vorking relation Relation: weather Instances: 14	Attribute	es: 5
ttributes in ba	se relation				ilters		
011		Nana		A	ddFilter -N unnamed	-C 0	Ad
International International	a <u>y</u>	•		Type	Name: outlook Missing: 0 (0%) Label Vercast	Distinct: 3	Type: Nominal Onique: 0 (0%)
				n	anny	5	
na							
79 2:11:32: email 2:11:32: Starte 2:12:05: Base	l: wekasupp ed on 木曜E relation is i	oort@cs.wa ],37月200 now weathe	ikato.ac.nz 13 er (14 instanci	es)			

#### 問題設定を調べる

- \*.arffファイルをMedowやWordで開き、ドキュ メントを読む
- ドキュメントが無い場合、属性名とクラス名
   や実行結果から推測する
- 本来のデータマイニングでは、問題(クラス)
   を設定するのはとても重要な仕事!!

### 決定木を実行する(1)

#### Classifyの タグを選択

👺 Weka Knowledge Explorer	
Preprocess Classify Cluster	Associate Select attributes Visualize
Classifier	
ZeroR	
Test options	Classifier output
O Use training set	
O Supplied test set Set	
Cross-validation Folds 10	
O Percentage split % 66	
Mara antiana	
(Nom) play 👻	
Start Stop	
Result list	
Log	
22:11:32: email: wekasupport@cs.w	aikato.ac.nz
22:12:05: Base relation is now weath	ier (14 instances)
22:12:05: Working relation is now we	ather (14 instances)

### 決定木を実行する(2)

#### Classifierの下のテキスト(ZeroRと出ているところ)をクリック



### 決定木を実行する(3)

#### 新しく出てきたウィンドウのプルダウンメニューをクリック

🐦 Weka GUI Chooser		the second se
Maikato	Emironment for	
Knowle		
(c) (	100 Drawraacaa Claasify Clump Accessing Calent attributes Misuraling	
Univers	sity Classifier	
Ne	w ZeroR weka.classifiers.ZeroR v	
	Toct aptions	
	Weka.classifiers.DecisionStump	
	weka.classifiers.HyperPipes	
	Supplied test set Set weka.classifiers.IB1	
	Cross-validation Folds 1 weka.classifiers.IBk	
	Percentage split % 6 weka.classifiers.id3	
and the second se	More options	
GUI	(Nom) play	
Simple CLI	Exc Start Stop	
	Result list	
	pr	
Acrobat 5.0 1.V		
🐴 👘 🚺	Yis I I I I I I I I I I I I I I I I I I I	
cs6002 - No	pr l	
Tiger へのシ Ka.		
	22:11:32: email: wekasupport@cs.waikato.ac.nz	
	22:11:32: Started on 木曜日, 3 7月 2003	
Media	22:12:05: Working relation is now weather (14 instances)	
	Status	
	OK	×0
OK712SJT02 NetBea 3.	ns 5 <b>m</b> agna	
		ј 🗛 般 😂 🔂 🔀 🕫 кана
<b>昭</b> スタート   🙈 😭	) 🕜 📀 🛛 😋 weka-3-2-3 📕 🗃 ローカル ディス 🛯 🔤 コマンド プロン 🕨 Weka GUI Ch 🛚 🍪 Weka Knowle 🛛 📴 Microsoft Po	🏀 weka.gui.G 🔗 🗐 🥒 22:25

### 決定木を実行する(4)

#### 決定木 (J48.j48 )をクリック



### 決定木を実行する(5)

#### 今回はパラメータの調整は行わず,単に"OK"をクリックする

۲۲ ۴キュメント OK712	A Cata MeadowNT.exe@OREAD Buffers Files Tools Edit Search Mule Help	
	Weka Knowledge Explorer           Preprocess         Cluster         Resociate         Select attributes         Visualize           Classifier         Select attributes         Visualize         Visualize	
	J48 - C 0.25 - M 2 weka.classifiers.j48.J48	
	Test options binarySplits False	
マイ ネットワーク JCI	confidenceFactor 0.25	
S S	Supplied test set	
こみ箱 IA2IS	O Percentage split %	
	More options reducedErrorPruning False	
	(Nom) play	
Explorer N	Start Start Start	
	Result list unpruned False	
Adobe Acrobat 5.0		
	Open Save OK Cancel	
Tiger ∧@ົົン	, Log	
- Conta	( 22:11:32: email: wekasupport@cs.waikato.ac.nz 2	<b>A</b>
setup Wind Media	2 22:12:05: Base relation is now weather (14 instances) 2 22:12:05: Working relation is now weather (14 instances)	
1 1 10	Status	
OK712SJ102 NetBean	ок	×0
3.5		
all the set		

### 決定木を実行する(6)

#### 性能(正解率など)の評価の仕方を設定する 今回はこのままで良い (別の評価法の場合は,理由を説明)

🌺 Weka Knowledge I	Explorer				_	
Preprocess Cl	assify Cluster	Associate	Select attributes	Visualize		
Classifier						
J49-C 0.25-M 2						_
Test options		Classifier out	tput			
O Use training s	et					
O Supplied test	set Set					1
Cross-validati	on Folds 10					
O Percentage s						
Mare or	tions					
More of	Juons					
(Nom) play	•					
Start	Stop					
Result list						

### 決定木を実行する(7)

#### クラスを決定

#### 今回はこのままで良い

#### ('(Nom)"となっている名義属性をクラスにすることが可能)

Preprocess 0	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier						
J48-C 0.25-M 2						
Test options			Classifier ou	tput		
O Use training	set					
O Supplied tes	t set	Set				
Cross valida	tion Eo	Ide 10				
	uun ru					
	shur	% <u>00</u>				
More o	ptions					
(Mana) alors			$\mathbf{b}$			
(NOM) play						
Start	S	top				
Result list						

### 決定木を実行する(8)

#### Startをクリック

Transpoor	Associate Coloct attributes Micualize	
reprocess classify cluster	Associate Select attributes visualize	
lassmer		
J48 -C 0.25 -M 2		
est options	-Classifier output	
Use training set		-
O Supplied test set Set		
Cross validation Folds 10		
O Percentage split % 66		
More options		
Nom) plav 👻		
Start Stop		
esuit list		

### 決定木の実行結果(1)

🥵 Weka Knowledge Explorer		
Preprocess Classify Cluster Classifier	Associate Select attributes 結果の先頭には訓練データセ	ッと
J48-C 0.25-M 2	ー 中仁 トラナー しの 桂 お が 主	= th z
Test options          Use training set         Supplied test set         Cross-validation         Percentage split         More options	Classic Lui ut UCATAOTHADICATAOTHADICA	
Result list	Test mode: 10-fold cross-validation	
18:06:12 - j48.J48	=== Classifier model (full training set) === J48 pruned tree 	
Log 18:06:02: Base relation is now weath 18:06:02: Working relation is now we 18:06:12: Started weka.classifiers.j48 18:06:12: Finished weka.classifiers.j-	er (14 instances) ather (14 instances) 3.J48 48.J48	
OK		€ ×0

### 決定木の実行結果(2)

🌺 Weka Knowledge Explorer	<b>二 訓結データセットから得られる</b>
Preprocess Classify Cluster	Associate Associate Visualize C 717555 01 60
Classifier	注定大の実元
J48 -C 0.25 -M 2	
Test options         Use training set         Supplied test set       Set         Cross-validation       Folds       10         Percentage split       %       66         More options       66         (Nom) play          Start       Stop         Result list       18:06:12 - j48.J48	Classifier outpy (Lassifier model (full training set) === J48 pruned tree 
Log 18:06:02: Base relation is now weather 18:06:02: Working relation is now weather 18:06:12: Started weka.classifiers.j48 18:06:12: Finished weka.classifiers.j4 Status OK	Size of the tree : 8

### 決定木の実行結果(3)



#### 決定木の実行結果の処理 (テキスト) 1.Result listのj48.J48を右クリック

	🌺 Weka Knowledge Exp	plorer					
	Preprocess Clas	ssify Cluste	r Associate	Select attributes	Visualize		
	Classifier						
	J48 -C 0.25 -M 2						
2 .Save result buffer ≉	J48 - C 0.25 - M 2 Test options Use training set Supplied test se Cross-validation Percentage split More option (Nom) play Start Result list 22:50:13 - j48.J48	et Set n Folds 10 t % 66 ons Stop View View Save Visua Visua	Classifier ou Mean absolu Root mean s Relative ab Root relati Total Numbe === Detaile TP Rate F 0.778 0.2 === Confusi a b < in main window in separate wind result buffer lize classifer err lize tree	Itput te error solute error ve squared error or of Instances d Accuracy By Clas P Rate Precision 0.8 0.636 0.222 0.333 on Matrix === classified as sow	0.3571 0.5214 76.9231 X 108.734 X 14 ss === n Recall F-Measure 0.778 0.7 0.2 0.25	Class yes no	
	22:12:05: Base relati 22:12:05: Working re 22:50:13: Started we	elation i ka.clas Visua	lize margin curv lize threshold cu	ne urve ▶			
フリップし、加末を休1子	22:50:14: Finished w	veka.cla Visua	lize cost curve	•			
	Status OK						

#### 決定木の実行結果の処理 図:その1) 1.Result listのj48.J48を右クリック

	😹 Weka Knowledge Explorer										
1	Preprocess	Classify	Cluster	Associate	Selec	t attributes	Visualize				
	Classifier										
	J48 -C 0.25 -M 2										
	Test options Use training set Supplied test set Cross-validation Folds 10 Percentage split % 66 More options (Nom) play Start Stop Result list 22:50:43, i48, 148		Set Ids 10 % 66 top	Classifier output         Mean absolute error       0.3571         Root mean squared error       0.5214         Relative absolute error       76.9231 %         Root relative squared error       108.734 %         Total Number of Instances       14         === Detailed Accuracy By Class ===       TP         TP Rate       FP Rate       Precision         0.778       0.8       0.636       0.778       0.7         0.2       0.222       0.333       0.2       0.25       no         === Confusion Matrix ===       a.b.       <== classified as							
	View in View in		main window separate window								
	Savere										
2.Visualize treeを クリック	Log Visualize 22:12:05: Base relation is n 22:12:05: Working relation i 22:50:13: Started weka.clas 22:50:14: Finished weka.cla			e classifier er e tree e margin curv e threshold ci e cost curve	rors ∧e urve ⊧ ⊧						
	Status OK										×0



### 決定木の実行結果の処理 (図:その2)

#### Alt+PrintScreenを押してから, 木の表示されているウィンドウをクリックする



#### 結果の説明・考察のポイント

- 決定木の枝分かれがどのような条件で起こっているのか
- 決定木をかけることにより、データセットから何を得たのか
- 決定木を使って何ができるのか

#### 注意事項

- この練習問題は静岡大学情報学部 学部 共通専門科目 知識システム論」の中で利 用した課題を元にしています.
- 用語の定義は一部の定義にすぎないかも しれません、より詳しい解説はWeb上のリ ソースや書籍を参照してください。

#### 2003年7月30日 阿部秀尚