

Hong Kong University of Science and Technology
COMP4211: Machine Learning
Spring 2013

Assignment 2

Due: 17/5/2013, Friday, in tutorial

Given the data set below, please construct the decision tree by using ID3 (that is to apply entropy and gain information to search for the best attribute as the internal classifier). If there are two attributes whose gain information are both the highest, choose the left most one according to the table. Show your calculation steps of each internal classifier.

L0 English				input f	L1 Chinese												
<i>translation</i>	<i>output e</i>	vector representation			positional features			vector representation									
		"go"	"walk"		t-1	t+1	t+2	少 <i>shǎo</i>	路 <i>lù</i>	别 <i>bié</i>	错 <i>cuò</i>	快 <i>kuài</i>	向 <i>xiàng</i>	前 <i>qián</i>	过 <i>guò</i>	去 <i>qù</i>	上 <i>shàng</i>
walk wrong less	"walk"	0	1	"少走错路" <i>shǎo zǒu cuò lù</i>	少 <i>shǎo</i>	错 <i>cuò</i>	路 <i>lù</i>	1	1	0	1	0	0	0	0	0	0
don't go wrong	"go"	1	0	"别走错路" <i>bié zǒu cuò lù</i>	别 <i>bié</i>	错 <i>cuò</i>	路 <i>lù</i>	0	1	1	1	0	0	0	0	0	0
walk fast towards	"walk"	0	1	"快走向前" <i>kuài zǒu xiàng qián</i>	快 <i>kuài</i>	向 <i>xiàng</i>	前 <i>qián</i>	0	0	0	0	1	1	1	0	0	0
don't walk too fast	"walk"	0	1	"别走过快" <i>bié zǒu guò kuài</i>	别 <i>bié</i>	过 <i>guò</i>	快 <i>kuài</i>	0	0	1	0	1	0	0	1	0	0
go fast through	"go"	1	0	"快走过去" <i>kuài zǒu guò qù</i>	快 <i>kuài</i>	过 <i>guò</i>	去 <i>qù</i>	0	0	0	0	1	0	0	1	1	0
go up fast	"go"	1	0	"快走上去" <i>bié zǒu shàng qù</i>	快 <i>kuài</i>	上 <i>shàng</i>	去 <i>qù</i>	0	0	0	0	1	0	0	0	1	1
don't walk there	"walk"	0	1	"别走过去" <i>bié zǒu guò qù</i>	别 <i>bié</i>	过 <i>guò</i>	去 <i>qù</i>	0	0	1	0	1	0	0	1	0	0
go through less	"go"	1	0	"少走过去" <i>shǎo zǒu guò qù</i>	少 <i>shǎo</i>	过 <i>guò</i>	去 <i>qù</i>	1	0	0	0	0	0	0	1	1	0