

人工智能——模式挖掘的作业

孔静——2014K8009929022

October 15, 2016

Contents

1	概 述	1
1.1	问题	1
1.2	要求	1
2	构 造	2
2.1	故事背景	2
2.2	商品种类	2
2.3	购买记录	2
2.4	给定阈值	2
3	计 算	2
3.1	画图	2
3.2	frequent patterns	3
3.3	dominant patterns	3
3.4	maximal frequent patterns	3
3.5	others	3

1 概 述

1.1 问题

设计一个 Transaction Database :

1. 画出它的 lexicographic subset tree
2. 给定 \min_sup , \min_occu 的阈值, 计算出在这些阈值下的 frequent patterns, dominant patterns, maximal frequent patterns

1.2 要求

这个例子中, 必须包含一个 pattern, 使得: 它是 maximal frequent pattern, 但它不是 dominant pattern

2 构造

2.1 故事背景

作为一个曾经的游戏玩家，经常购买游戏皮肤，现以玩家购买游戏皮肤为背景，创建数据。

2.2 商品种类

截止至我不玩英雄联盟，该游戏总共有 655 款皮肤。为简化作业复杂度，简化为五种商品，记为 A,B,C,D,E。

2.3 购买记录

编造了一些数据如下表。

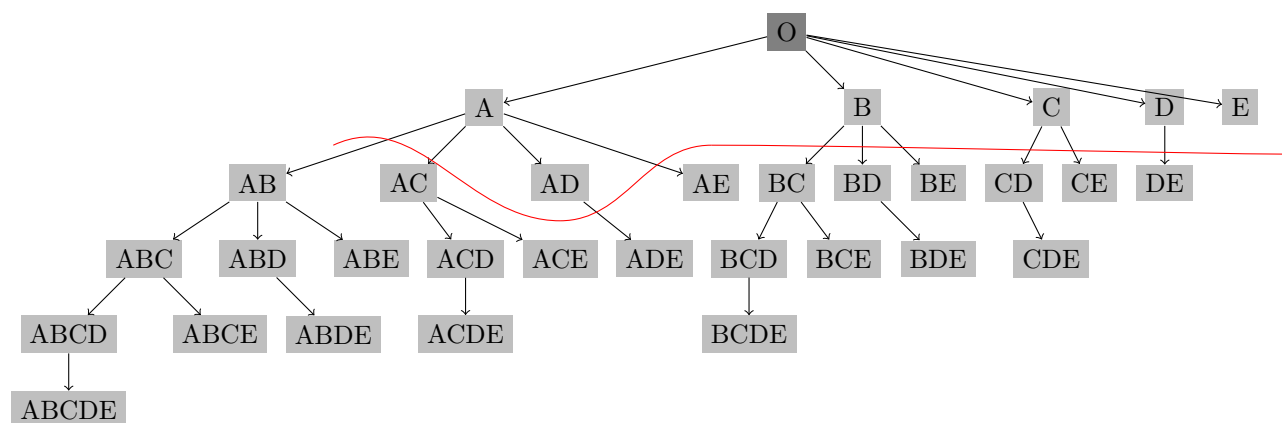
玩家	商品	玩家	商品
1	A B	6	A D
2	A C D	7	E
3	B E	8	A C E
4	A D E	9	B D
5	B C D E	10	A B C D E

2.4 给定阈值

$\min_sup = 0.3$, $\min_occu = 0.5$

3 计算

3.1 画图



3.2 frequent patterns

1	A	B	C	D	E
	0.6	0.5	0.4	0.6	0.6
2	AB	AC	AD	AE	BC
	0.2	0.3	0.4	0.3	0.2
2	BD	BE	CD	CE	DE
	0.3	0.3	0.3	0.3	0.3

结果为：
A,B,C,D,E
AD

3.3 dominant patterns

1	A	B	C	D	E
	$\frac{11}{30}$	0.39	$\frac{67}{240}$	$\frac{127}{360}$	$\frac{157}{360}$
2	AB	AC	AD	AE	BC
	0.7	$\frac{26}{45}$	$\frac{41}{60}$	$\frac{26}{45}$	0.45
2	BD	BE	CD	CE	DE
	$\frac{19}{30}$	$\frac{19}{30}$	$\frac{47}{90}$	$\frac{47}{90}$	$\frac{47}{90}$
3	ABC	ABD	ABE	ACD	ACE
	0.6	0.6	0.6	0.8	0.8
3	ADE	BCD	BCE	BDE	CDE
	0.8	0.675	0.675	0.675	0.6
4	ABCD	ABCE	ABDE	ACDE	BCDE
	0.8	0.8	0.8	0.8	0.9
5	ABCDE				
	1				

结果为：
AB,AC,AD,AE, BD,BE,CD,CE,DE
ABC,ABD,ABE,ACD,ACE, ADE,BCD,BCE,BDE,CDE
ABCD,ABCE,ABDE,ACDE,BCDE
ABCDE

3.4 maximal frequent patterns

结果为：B,C,E,AD

3.5 others

显然，B,C,E 是 maximal frequent patterns，而不是 dominant patterns