



**ÚSTAV INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ  
FF UK V PRAZE**

**Jiří Ivánek**

# **Stručně o rozhodovacích tabulkách a konstrukci rozhodovacích stromů**

**Verze 1.0**

**Praha**

**2014**

# Obsah

<b>1</b>	<b>Základní vlastnosti rozhodovacích tabulek a stromů</b>	<b>2</b>
1.1	Dichotomické rozhodovací tabulky . . . . .	2
1.2	Logické vlastnosti rozhodovacích tabulek . . . . .	4
1.3	Rozhodovací tabulky s pravděpodobnostní distribucí . . . . .	6
1.4	Rozhodovací stromy konzistentní s rozhodovací tabulkou . . . . .	8
1.5	Vztah střední délky rozhodovacího stromu a entropie . . . . .	10
<b>2</b>	<b>Přímé algoritmy konstrukce rozhodovacích stromů</b>	<b>13</b>
2.1	Schéma přímých algoritmů . . . . .	14
2.2	Minimalizace počtu zbývajících možných rozhodnutí . . . . .	16
2.3	Minimalizace maximální zbývající entropie . . . . .	19
2.4	Optimální rozhodovací strom . . . . .	22
<b>3</b>	<b>Konstrukce rozhodovacích stromů z dat</b>	<b>24</b>
3.1	Princip algoritmu TDIDT . . . . .	27
3.2	Příklad výpočtu algoritmem TDIDT . . . . .	27

# Kapitola 1

## Základní vlastnosti rozhodovacích tabulek a stromů

Konstrukce rozhodovacích stromů z dat je zřejmě nejčastěji používanou metodou dobývání znalostí z databází [6]. Seznámíme se s principy vybraných algoritmů konstrukce rozhodovacích stromů z rozhodovacích tabulek a ukážeme jejich aplikaci na matici dat.

V data-miningových systémech (k akademickému použití doporučujeme data-miningový systém Lisp-Miner [5]) je nejčastěji implementována některá varianta výpočetně rychlého přímého algoritmu používajícího k výběru atributu entropii, který ovšem na rozdíl od ostatních uvedených algoritmů nezaručuje nalezení optimálního rozhodovacího stromu. Podrobnější popis metod konstrukce rozhodovacích stromů z dat je dostupný v monografii P. Berky [1]. V našem textu se soustředíme na princip přímého algoritmu založeného na snižování entropie. Ve výkladu vycházíme z prací [1, 2, 3, 4].

### 1.1 Dichotomické rozhodovací tabulky

Pro náš úvodní výklad postačí, když se omezíme na nejjednodušší typ dichotomických rozhodovacích tabulek, které obsahují otázky pouze ve formě booleovských atributů. Tabulka je tvořena souborem rozhodovacích pravidel, v jejichž předpokladu je některá elementární konjunkce booleovských atributů a v závěru je uvedeno rozhodnutí.

## Příklad dichotomické rozhodovací tabulky

	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$	$R_6$	$R_7$
$A$	1	1	–	1	0	0	0
$B$	1	0	0	0	1	0	0
$C$	–	1	1	0	–	1	0
$D$	–	1	0	–	–	1	–
	$a_1$	$a_2$	$a_5$	$a_3$	$a_2$	$a_1$	$a_4$

- **Otázky** (podmínky): booleovské atributy  $A, B, C, D$
- **Rozhodnutí** (akce, závěry):  $a_1, a_2, a_3, a_4, a_5$
- 0, 1 značí **hodnoty** booleovských atributů,  
" – " značí prázdné pole (kdy na hodnotě atributu nezáleží)
- **Rozhodovací pravidla:**  $R_1, R_2, R_3, R_4, R_5, R_6, R_7$

$$R_1: A \wedge B \Rightarrow a_1$$

$$R_2: A \wedge \neg B \wedge C \wedge D \Rightarrow a_2$$

$$R_3: \neg B \wedge C \wedge \neg D \Rightarrow a_5$$

$$R_4: A \wedge \neg B \wedge \neg C \Rightarrow a_3$$

$$R_5: \neg A \wedge B \Rightarrow a_2$$

$$R_6: \neg A \wedge \neg B \wedge C \wedge D \Rightarrow a_1$$

$$R_7: \neg A \wedge \neg B \wedge \neg C \Rightarrow a_4$$

## 1.2 Logické vlastnosti rozhodovacích tabulek

Rozhodovací tabulka je určena k odvození rozhodnutí v případech, které jsou popsány odpověďmi na uvedené otázky rozhodovací tabulky – z hlediska výrokové logiky tedy pravdivostním ohodnocením daných booleovských atributů (výrokových proměnných). Uveďme si stručně některé situace, které mohou v tabulce nastat:

**Překrytí pravidel** nastává, když existuje pravdivostní ohodnocení splňující předpoklady různých pravidel. Přitom může být

- **redundantní** - rozhodnutí v závěrech těchto pravidel jsou stejná,
- **sporné** - rozhodnutí v závěrech těchto pravidel jsou různá.

### Příklad rozhodovací tabulky s překrytím pravidel

	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$	$R'_6$	$R'_7$
$A$	1	1	–	1	0	1	0
$B$	1	0	0	0	1	1	–
$C$	–	1	1	0	–	1	0
$D$	–	1	0	–	–	1	–
	$a_1$	$a_2$	$a_5$	$a_3$	$a_2$	$a_1$	$a_4$

Zde se pravidlo  $R'_6$  překrývá s pravidlem  $R_1$  a jde o redundantní překrytí. Pravidlo  $R'_7$  se překrývá s pravidlem  $R_5$  a jde o sporné překrytí.

Existenci překrytí můžeme zkontrolovat porovnáváním všech dvojic pravidel v tabulce. U každé dvojice zjistíme, zda se v některé otázce předpoklady pravidel explicitně liší, tj. zda v tabulce je v některém řádku u jednoho z pravidel 0 a u druhého 1. Z hlediska výrokové logiky to pak znamená, že konjunkce předpokladů této dvojice pravidel je kontradikcí a žádné pravdivostní ohodnocení ji nemůže splnit. Tento test ukáže, že naše úvodní rozhodovací tabulka neobsahuje překrytí.

Celkově je rozhodovací tabulka je **konzistentní**, jestliže v ní neexistuje sporné překrytí. Poznamenejme, že rozhodovací tabulka bez překrytí je samozřejmě konzistentní.

Další logická vlastnost rozhodovací tabulky je její schopnost určit rozhodnutí v každém případě, který je popsán odpověďmi na otázky uvedené

v tabulce. Rozhodovací tabulka je **úplná**, jestliže každé pravdivostní ohodnocení splňuje předpoklad některého pravidla. Z hlediska výrokové logiky to znamená, že disjunkce předpokladů pravidel v rozhodovací tabulce je **tautologií**.

Vlastnost úplnosti můžeme u rozhodovací tabulky bez překrytí pravidel snadno posoudit rozborem počtu pravdivostních ohodnocení pokrytých jednotlivými pravidly. Ukážeme si to na příkladu úvodní rozhodovací tabulky.

### Rozbor rozhodovacích pravidel z příkladu

Počet pravdivostních ohodnocení výrokových proměnných  $A, B, C, D$  pokrytých (nepřekrývajícími se) pravidly:

$$2^2 \text{ pravidlem } R_1: A \wedge B \Rightarrow a_1$$

$$2^0 \text{ pravidlem } R_2: A \wedge \neg B \wedge C \wedge D \Rightarrow a_2$$

$$2^1 \text{ pravidlem } R_3: \neg B \wedge C \wedge \neg D \Rightarrow a_5$$

$$2^1 \text{ pravidlem } R_4: A \wedge \neg B \wedge \neg C \Rightarrow a_3$$

$$2^2 \text{ pravidlem } R_5: \neg A \wedge B \Rightarrow a_2$$

$$2^0 \text{ pravidlem } R_6: \neg A \wedge \neg B \wedge C \wedge D \Rightarrow a_1$$

$$2^1 \text{ pravidlem } R_7: \neg A \wedge \neg B \wedge \neg C \Rightarrow a_4$$

Pravidla se nepřekrývají, takže tabulka je jistě konzistentní. Celkový počet pokrytých pravdivostních ohodnocení výrokových proměnných  $A, B, C, D$  je tedy 16, tj. je pokryto všech  $2^4$  pravdivostních ohodnocení. Rozhodovací tabulka z příkladu je tedy úplná.

Můžeme se o tom přesvědčit také tak, že pomocí ekvivalentních úprav formulí výrokové logiky ukážeme, že disjunkce předpokladů pravidel

$$(A \wedge B) \vee (A \wedge \neg B \wedge C \wedge D) \vee (\neg B \wedge C \wedge \neg D) \vee (A \wedge \neg B \wedge \neg C) \vee \\ \vee (\neg A \wedge B) \vee (\neg A \wedge \neg B \wedge C \wedge D) \vee (\neg A \wedge \neg B \wedge \neg C)$$

je tautologií.

## 1.3 Rozhodovací tabulky s pravděpodobnostní distribucí

V typických aplikacích mají případy, které jsou popsány odpověďmi na otázky rozhodovací tabulky, různou frekvenci výskytu, což lze vyjádřit pomocí zadání pravděpodobnosti jednotlivých kombinací odpovědí. Pravidla rozhodovací tabulky pak můžeme doplnit o pravděpodobnosti, že budou splněny jejich předpoklady. Součet těchto pravděpodobností musí být u úplné tabulky bez překrytí roven 1. Budeme to požadovat i u neúplné tabulky bez překrytí (očekáváme, že nepokryté kombinace odpovědí se nemohou vyskytnout, takže mají nulovou pravděpodobnost). Rozdělení pravděpodobnosti v rámci předpokladů jednotlivých pravidel budeme chápat jako rovnoměrné – pravděpodobnosti pravidel se rozdělí rovnoměrně mezi pokryté kombinace odpovědí. Můžeme tak přiřadit pravděpodobnosti pravidlům v **rozvinuté rozhodovací tabulce**, v níž každé pravidlo původní tabulky nahradíme souborem pravidel podle pokrytých kombinací odpovědí – viz příklad.

### Příklad rozhodovací tabulky s pravděpodobnostní distribucí

$P$	0.2	0.1	0.4	0.3	0
$Q_1$	–	0	0	1	1
$Q_2$	0	0	1	–	1
$Q_3$	0	1	–	1	0
$A$	$a_1$	$a_2$	$a_2$	$a_3$	

### Rozvinutá rozhodovací tabulka

$P$	0.1	0.1	0.1	0.2	0.2	0.15	0.15	0
$Q_1$	0	1	0	0	0	1	1	1
$Q_2$	0	0	0	1	1	0	1	1
$Q_3$	0	0	1	0	1	1	1	0
$A$	$a_1$	$a_1$	$a_2$	$a_2$	$a_2$	$a_3$	$a_3$	

### Entropie prostoru možných rozhodnutí

Z rozhodovací tabulky s pravděpodobnostní distribucí můžeme též určit pravděpodobnost  $P(a_i)$  jednotlivých rozhodnutí  $a_1, \dots, a_m$  jako součet pravděpodobností pravi-

del se závěrem  $a_i$ . Odtud se určí **entropie prostoru možných rozhodnutí**:

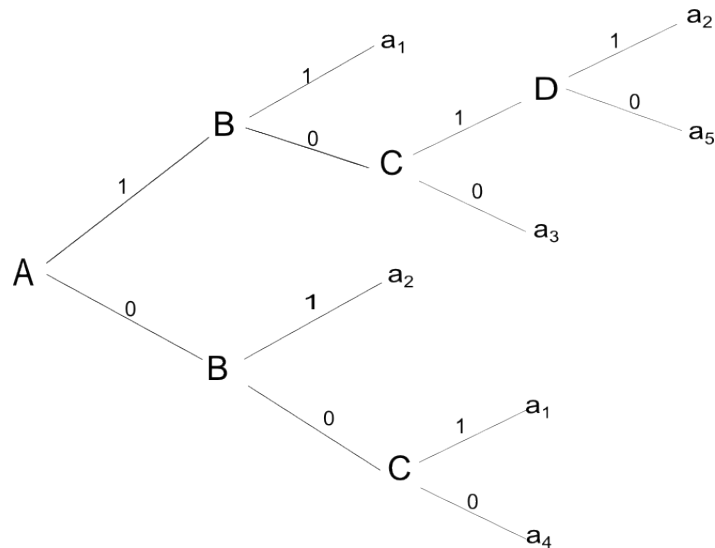
$$H(A, P) = - \sum_{i=1}^m P(a_i) \cdot \log_2 P(a_i)$$

Orientační tabulka pro výpočet entropie je umístěna do přílohy textu.

V našem příkladu rozhodovací tabulky s pravděpodobnostní distribucí určíme  $P(a_1) = 0.3$ ,  $P(a_2) = 0.4$ ,  $P(a_3) = 0.3$ , takže entropie prostoru možných rozhodnutí je

$$H(A, P) = -0.3 \log_2 0.3 - 0.4 \log_2 0.4 - 0.3 \log_2 0.3 = 0.52 + 0.53 + 0.52 = 1.57$$





Obrázek 1.1: Příklad rozhodovacího stromu

## 1.4 Rozhodovací stromy konzistentní s rozhodovací tabulkou

V teorii grafů je strom obecně definován jako souvislý graf neobsahující cykly.

**Rozhodovací strom** (viz Obr. 1.1) bude strom s orientovanými hranami, v němž

- jeden vrchol je **kořen** stromu, do nějž nevstupují žádné hrany, do ostatních uzlů vstupuje právě jedna hrana,
- **terminální** uzly (listy stromu) jsou ty vrcholy, z nichž žádná hrana nevystupuje,
- **neterminální** uzlům jsou **přiřazeny otázky**  $Q_1, \dots, Q_n$  tak, že na žádné cestě z kořene se nesmí žádná otázka opakovat
- z každého neterminálního uzlu vycházejí orientované hrany (**větve**) označené jednotlivými **možnými odpověďmi** na přiřazenou otázku,
- **terminálním** uzlům jsou **přiřazena rozhodnutí**.

Je zřejmé, že do každého uzlu rozhodovacího stromu vede jediná cesta z jeho kořene (její **délka**, tzv. **úroveň** uzlu znamená počet otázek potřebných k dosažení uzlu).

Abychom mohli formulovat vztah rozhodovacích stromů a tabulek, zavedeme nejdříve pojem rozhodovací funkce. Uvažujme prostor možných otázek  $Q_1, \dots, Q_n$  a možných rozhodnutí  $a_1, \dots, a_m$ . **Rozhodovací funkce**  $d$  přiřazuje každé kombinaci odpovědí  $r_1, \dots, r_n$  na otázky  $Q_1, \dots, Q_n$  některé rozhodnutí z množiny rozhodnutí  $a_1, \dots, a_m$ .

Rozhodovací funkce  $d$  je **konzistentní** s danou rozhodovací tabulkou, jestliže platí: kdykoli kombinace odpovědí  $r_1, \dots, r_n$  na otázky  $Q_1, \dots, Q_n$  splňuje předpoklad některého pravidla z rozhodovací tabulky, pak rozhodnutí určené rozhodovací funkcí  $d$  je stejné jako rozhodnutí určené tímto pravidlem.

Je-li rozhodovací tabulka konzistentní, pak existuje s ní konzistentní rozhodovací funkce - pokud kombinace odpovědí  $r_1, \dots, r_n$  na otázky  $Q_1, \dots, Q_n$  splňuje předpoklad některého pravidla z rozhodovací tabulky, pak hodnotu rozhodovací funkce  $d$  definujeme jako rozhodnutí určené tímto pravidlem; v ostatních případech definujeme hodnotu rozhodovací funkce  $d$  libovolně. Je-li navíc rozhodovací tabulka úplná, pak je rozhodovací funkce takto definována jednoznačně.

Rozhodovací strom jednoznačně určuje **rozhodovací funkci**:

- pro každou kombinaci odpovědí  $r_1, \dots, r_n$  na otázky  $Q_1, \dots, Q_n$  najdeme cestu (existuje právě jedna) z kořene stromu do terminálního uzlu danou v každém neterminálním uzlu s přiřazenou otázkou  $Q_i$  větví označenou odpovědí  $r_i$ ,
- hodnotou rozhodovací funkce je pak **rozhodnutí přiřazené tomuto terminálnímu uzlu**.

Uvažujme nyní rozhodovací strom, který má přiřazený otázky, odpovědi a rozhodnutí z dané rozhodovací tabulky. Potom můžeme srovnávat rozhodovací funkci určenou rozhodovacím stromem s danou rozhodovací tabulkou.

Rozhodovací strom je **konzistentní s rozhodovací tabulkou**, je-li jím určená rozhodovací funkce konzistentní s rozhodovací tabulkou. Rozhodovací strom na obr. 1.1 je například konzistentní s rozhodovací tabulkou ze začátku této kapitoly.

Úlohou je **sestrojit co nejlepší rozhodovací strom** konzistentní s danou rozhodovací tabulkou. Cílem je rozhodovací strom, který by měl **co nejmenší počet otázek nutných k rozhodnutí**.

To lze vyjádřit jako **minimalizaci maximální délky cest** z kořene stromu do terminálních uzlů; u rozhodovací tabulky s pravděpodobnostní distribucí pak jako **minimalizaci střední délky stromu**, kterou budeme dále definovat. Za **optimální rozhodovací strom** pak budeme považovat rozhodovací strom, který má minimální střední délku mezi všemi rozhodovacími stromy konzistentními s danou rozhodovací tabulkou s pravděpodobnostní distribucí.

## 1.5 Vztah střední délky rozhodovacího stromu a entropie

Uvažujme rozhodovací tabulku s pravděpodobnostní distribucí a rozhodovací strom s ní konzistentní. Pak můžeme určit pravděpodobnost  $P(u)$  dosažení jeho uzlu  $u$  tak, že sečteme z rozvinuté rozhodovací tabulky pravděpodobnosti těch pravidel, jejichž předpoklad je konzistentní s odpověďmi na jednoznačné cestě z kořene  $u_0$  do uzlu  $u$ .

**Vlastnosti** pravděpodobnosti dosažení uzlů pak jsou:

- pravděpodobnost kořene:  $P(u_0) = 1$
- pravděpodobnost uzlu  $u$  se dělí mezi jeho následníky  $v \in N(u)$ :  

$$P(u) = \sum_{v \in N(u)} P(v)$$
- pravděpodobnosti dosažení terminálních uzlů  $w \in W$  tvoří distribuci:  

$$\sum_{w \in W} P(w) = 1$$

**Střední délka rozhodovacího stromu**  $d_P(T)$  je pak vážený průměr délek  $d(w)$  cest z kořene do terminálních uzlů  $w \in W$ :

$$d_P(T) = \sum_{w \in W} P(w) \cdot d(w)$$

Z pravděpodobností dosažení terminálních uzlů  $w \in W$  můžeme též určit **entropii rozhodovacího stromu**  $H(W, P)$ :

$$H(W, P) = - \sum_{w \in W} P(w) \cdot \log_2 P(w)$$

Pro všechny rozhodovací stromy konzistentní s danou rozhodovací tabulkou platí

$$H(A, P) \leq H(W, P),$$

kde  $H(A, P)$  je **entropie prostoru možných rozhodnutí**, kterou určíme přímo z rozhodovací tabulky tak, že sečteme pravděpodobnosti pravidel vedoucích ke každému rozhodnutí  $a_i$  a na výslednou distribuci  $P(a_i)$  použijeme vzorec entropie:

$$H(A, P) = - \sum_{i=1}^m P(a_i) \cdot \log_2 P(a_i)$$

Důležité je, že pro střední délku  $d_P(T)$  dichotomického rozhodovacího stromu platí:

$$H(W, P) \leq d_P(T)$$

a tedy

$$H(A, P) \leq d_P(T).$$

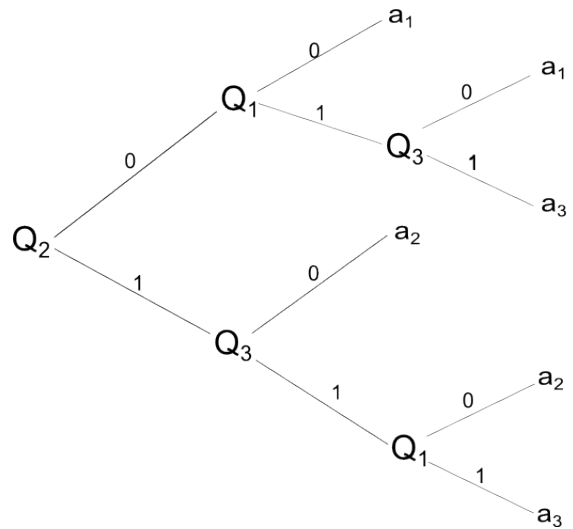
Entropie je tudíž dolním odhadem střední délky rozhodovacího stromu. Tento vztah budeme využívat v různých algoritmech konstrukce rozhodovacích stromů.

## Příklad výpočtu střední délky a entropie rozhodovacího stromu

Již dříve jsme spočítali entropii  $H(A, P) = 1.57$  prostoru možných rozhodnutí v našem příkladu rozhodovací tabulky s pravděpodobnostní distribucí:

$P$	0.1	0.1	0.1	0.2	0.2	0.15	0.15	0
$Q_1$	0	1	0	0	0	1	1	1
$Q_2$	0	0	0	1	1	0	1	1
$Q_3$	0	0	1	0	1	1	1	0
$A$	$a_1$	$a_1$	$a_1$	$a_2$	$a_2$	$a_3$	$a_3$	

Určíme střední délku rozhodovacího stromu na Obr. 1.2:



Obrázek 1.2: Příklad rozhodovacího stromu

$w$	$P(w)$	$d(w)$	$P(w).d(w)$
$(Q_2, 0) - (Q_1, 0) - a_1$	0.2	2	0.4
$(Q_2, 0) - (Q_1, 1) - (Q_3, 0) - a_1$	0.1	3	0.3
$(Q_2, 0) - (Q_1, 1) - (Q_3, 1) - a_3$	0.15	3	0.45
$(Q_2, 1) - (Q_3, 0) - a_2$	0.2	2	0.4
$(Q_2, 1) - (Q_3, 1) - (Q_1, 0) - a_2$	0.2	3	0.6
$(Q_2, 1) - (Q_3, 1) - (Q_1, 1) - a_3$	0.15	3	0.45
$d_P(T) =$			2.6

a vypočteme jeho entropii:

$w$	$P(w)$	$-\log_2 P(w)$	$-P(w). \log_2 P(w)$
$(Q_2, 0) - (Q_1, 0) - a_1$	0.2	2.32	0.46
$(Q_2, 0) - (Q_1, 1) - (Q_3, 0) - a_1$	0.1	3.32	0.33
$(Q_2, 0) - (Q_1, 1) - (Q_3, 1) - a_3$	0.15	2.74	0.41
$(Q_2, 1) - (Q_3, 0) - a_2$	0.2	2.32	0.46
$(Q_2, 1) - (Q_3, 1) - (Q_1, 0) - a_2$	0.2	2.32	0.46
$(Q_2, 1) - (Q_3, 1) - (Q_1, 1) - a_3$	0.15	2.74	0.41
$H(W, P) =$			2.53

## Kapitola 2

# Přímé algoritmy konstrukce rozhodovacích stromů

Připomeňme, že se zabýváme úlohou **sestrojit co nejlepší rozhodovací strom** konzistentní s danou rozhodovací tabulkou. Cílem je rozhodovací strom, který by měl **co nejmenší počet otázek nutných k rozhodnutí**.

Přímé algoritmy konstruují rozhodovací strom od jeho kořene, tedy "shora dolů". Nejdříve se vybere počáteční otázka, která bude přiřazena kořenu stromu. Strom se pak rozvine o větve vycházející z kořene a korespondující jednotlivým možným odpovědím na otázku přiřazenou kořenu. V nově vzniklých uzlech se postup opakuje, dokud není v jednotlivých větvích dosaženo jednoznačného rozhodnutí. Předností přímých algoritmů je rychlost výpočtu a nízké nároky na paměť. Nevýhodou je, že není zaručena optimalita sestrojeného rozhodovacího stromu. Výsledkem aplikace přímého algoritmu je v každém případě rozhodovací strom konzistentní s danou rozhodovací tabulkou, který je z hlediska použitého kritéria "suboptimální". Algoritmus budeme demonstrovat na příkladu převzatém z [2].

## 2.1 Schéma přímých algoritmů

V postupně se rozvíjejícím rozhodovacím stromu:

Pro každý **prozatím terminální uzel**  $u$  zúžíme rozhodovací tabulku na ta pravidla, s nimiž je konzistentní kombinace odpovědí na cestě z kořene stromu do  $u$ , a provedeme:

1. **Test terminality uzlu:**

Pokud zúžená tabulka neobsahuje různá rozhodnutí, označíme uzel  $u$  za terminální a přiřadíme mu konzistentní rozhodnutí. Jinak:

2. **Výběr otázky:**

Z otázek, které ještě nebyly na cestě z kořene stromu do  $u$  použity, vybereme podle daného kritéria nejvhodnější.

3. **Rozvinutí stromu:**

Z uzlu  $u$  přidáme větve odpovídající možným odpovědím na vybranou otázku.

### Kritéria výběru otázky

Pokud vycházíme z rozhodovací tabulky bez pravděpodobnostní distribuce, bude hlavním cílem zmenšování prostoru možných rozhodnutí. Vybrána bude otázka **minimalizující počet zbývajících možných rozhodnutí:**

Pro každou přípustnou otázku  $Q$  a možnou odpověď  $r$  určíme počet zbývajících rozhodnutí  $m(Q, r)$ . Kritériem bude maximum z těchto hodnot:

$$m(Q) = \max_r m(Q, r)$$

Pokud máme u rozhodovací tabulky k dispozici pravděpodobnostní distribuci, pak můžeme velikost prostoru možných rozhodnutí charakterizovat pomocí jeho entropie a použít pro výběr otázky **princip snižování entropie:**

Pro každou přípustnou otázku  $Q$  a možnou odpověď  $r$  určíme entropii zbývajících rozhodnutí  $H(Q, r)$ . Nejprve určíme pravděpodobnost dosažení uzlu  $P(Q, r)$  a pravděpodobnost každého rozhodnutí  $P(Q, r, a_i)$  v tomto uzlu. Odtud vypočteme **podmíněné pravděpodobnosti** všech zbývajících rozhodnutí

$$P(a_i|Q, r) = \frac{P(Q, r, a_i)}{P(Q, r)}$$

**Entropie zbývajících rozhodnutí**  $H(Q, r)$  je pak dána vzorcem:

$$H(Q, r) = - \sum P(a_i|Q, r) \cdot \log_2 P(a_i|Q, r)$$

Kritériem pro posouzení vhodnosti otázky  $Q$  bude **maximální zbývající entropie** při libovolné odpovědi  $r$ :

$$m_H(Q) = \max_r H(Q, r)$$

nebo **střední entropie**:

$$\bar{H}(Q) = \sum_r P(Q, r) \cdot H(Q, r)$$

### **Příklad rozhodovací tabulky pro konstrukci rozhodovacího stromu přímým algoritmem**

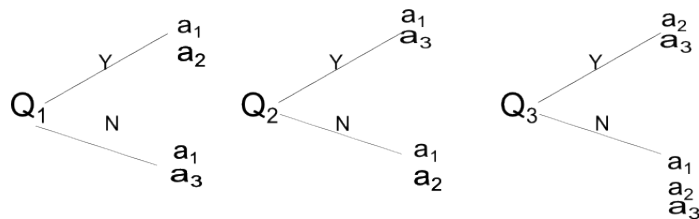
$P$	0.15	0.2	0.05	0.25	0.2	0.15
$Q_1$	$Y$	$N$	$Y$	$Y$	$N$	$N$
$Q_2$	$Y$	$Y$	$N$	$N$	$Y$	$N$
$Q_3$	$N$	$N$	$Y$	$N$	$Y$	$N$
$A$	$a_1$	$a_1$	$a_2$	$a_2$	$a_3$	$a_3$

Z dané rozhodovací tabulky zjistíme rozložení pravděpodobnosti  $P(a_i)$  možných rozhodnutí tak, že sečteme pravděpodobnosti pravidel vedoucích ke každému rozhodnutí  $a_i$ . Odtud vypočteme entropii prostoru možných rozhodnutí:

$$H(A, P) = - \sum_{i=1}^m P(a_i) \cdot \log_2 P(a_i)$$

	$P(a_i)$	$-\log_2 P(a_i)$	$-P(a_i) \cdot \log_2 P(a_i)$
$a_1$	0.35	1.51	0.53
$a_2$	0.3	1.74	0.52
$a_3$	0.35	1.51	0.53
$H(A, P)$			1.58





Obrázek 2.1: Výběr otázky pro kořen stromu

## 2.2 Minimalizace počtu zbývajících možných rozhodnutí

### 1. Výběr otázky pro kořen stromu:

Pro každou otázku  $Q_i$  určíme maximální počet zbývajících rozhodnutí při obou možných odpovědích  $Y, N$  - viz Obr. 2.1.

	$m(Q_i, Y)$	$m(Q_i, N)$	$m(Q_i)$
$Q_1$	2	2	2
$Q_2$	2	2	2
$Q_3$	2	3	3

Minima z  $m(Q)$  se nabývá u otázek  $Q_1, Q_2$ ; do kořene stromu vybereme např.  $Q_1$  a budeme pokračovat s výběrem otázek pro následující uzly.

### 2. Výběr otázky pro uzel daný cestou $(Q_1, Y)$

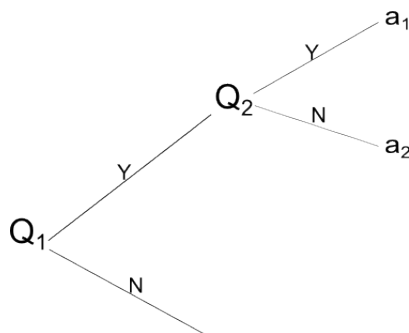
U otázek  $Q_2$  a  $Q_3$ , které ještě nebyly použity, opět určíme maximální počet zbývajících rozhodnutí při obou možných odpovědích  $Y, N$  (a odpovědi  $Y$  na otázku  $Q_1$ ):

	$m(Q_i, Y)$	$m(Q_i, N)$	$m(Q_i)$
$Q_2$	1	1	1
$Q_3$	1	2	2

Vybereme samozřejmě otázku  $Q_2$ , která dokonce již vede k terminálním uzlům - viz Obr. 2.2.

### 3. Výběr otázky pro uzel daný cestou $(Q_1, N)$

U otázek  $Q_2$  a  $Q_3$ , které ještě nebyly použity, opět určíme maximální



Obrázek 2.2: Částečně rozvinutý strom

počet zbývajících rozhodnutí při obou možných odpovědích  $Y, N$  (a odpovědi  $N$  na otázku  $Q_1$ ):

	$m(Q_i, Y)$	$m(Q_i, N)$	$m(Q_i)$
$Q_2$	2	1	2
$Q_3$	1	2	2

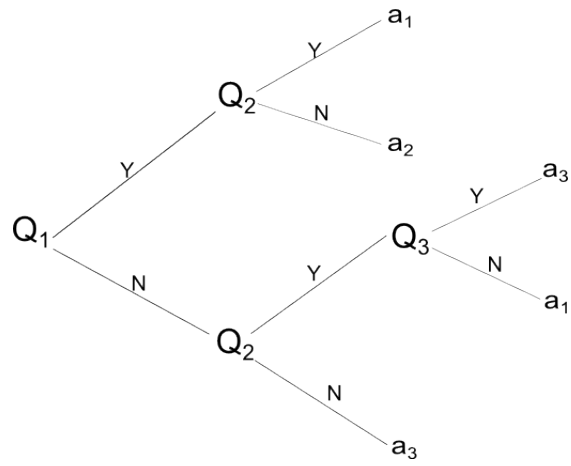
Obě otázky mají stejnou hodnotu kritéria, vybereme např. otázku  $Q_2$ .

4. V částečně rozvinutém stromu je třeba ještě doplnit otázku v uzlu na cestě dané odpověďmi  $(Q_1, N) - (Q_2, Y)$ , neboť zde jsou stále dvě možná rozhodnutí. Zbývá jedině otázka  $Q_3$ .

Výsledný strom pro danou rozhodovací tabulku při použití kritéria minimalizace počtu zbývajících možných rozhodnutí  $m(Q)$  je na Obr. 2.3.

V této konstrukci rozhodovacího stromu jsme nevyužívali pravděpodobnostní distribuci, můžeme si však vypočítat jeho střední délku  $d_P(T)$  jako vážený průměr délek cest do terminálních uzlů  $w$  (násobíme je pravděpodobností jejich dosažení  $P(w)$ ):

$w$	$P(w)$	$d(w)$	$P(w).d(w)$
$(Q_1, Y) - (Q_2, Y) - a_1$	0.15	2	0.3
$(Q_1, Y) - (Q_2, N) - a_2$	0.3	2	0.6
$(Q_1, N) - (Q_2, Y) - (Q_3, Y) - a_3$	0.2	3	0.6
$(Q_1, N) - (Q_2, Y) - (Q_3, N) - a_1$	0.2	3	0.6
$(Q_1, N) - (Q_2, N) - a_3$	0.15	2	0.3
			2.4



Obrázek 2.3: Rozhodovací strom sestrožený přímým algoritmem s výběrem otázky podle kritéria minimalizace počtu zbývajících možných rozhodnutí  $m(Q)$

Pro dosažení rozhodnutí je tedy třeba podle tohoto rozhodovacího stromu položit v průměru  $d_P(T) = 2.4$  otázek.

## 2.3 Minimalizace maximální zbývající entropie

### 1. Výběr otázky pro kořen stromu:

Pro každou otázku  $Q_i$  určíme entropii zbývajících rozhodnutí při obou možných odpovědích  $Y, N$ .

	$P(a_i, Q_1, Y)$	$P(a_i Q_1, Y)$	$-P(a_i Q_1, Y) \cdot \log_2 P(a_i Q_1, Y)$
$a_1$	0.15	0.33	0.53
$a_2$	0.3	0.67	0.39
$a_3$	0		
$P(Q_1, Y)$	0.45	$H(Q_1, Y)$	0.92

	$P(a_i, Q_1, N)$	$P(a_i Q_1, N)$	$-P(a_i Q_1, N) \cdot \log_2 P(a_i Q_1, N)$
$a_1$	0.2	0.36	0.53
$a_2$	0		
$a_3$	0.35	0.64	0.41
$P(Q_1, N)$	0.55	$H(Q_1, N)$	0.95

	$P(a_i, Q_2, Y)$	$P(a_i Q_2, Y)$	$-P(a_i Q_2, Y) \cdot \log_2 P(a_i Q_2, Y)$
$a_1$	0.35	0.64	0.41
$a_2$	0		
$a_3$	0.2	0.36	0.53
$P(Q_2, Y)$	0.55	$H(Q_2, Y)$	0.95

	$P(a_i, Q_2, N)$	$P(a_i Q_2, N)$	$-P(a_i Q_2, N) \cdot \log_2 P(a_i Q_2, N)$
$a_1$	0		
$a_2$	0.3	0.67	0.39
$a_3$	0.15	0.33	0.53
$P(Q_2, N)$	0.45	$H(Q_2, N)$	0.92

	$P(a_i, Q_3, Y)$	$P(a_i Q_3, Y)$	$-P(a_i Q_3, Y) \cdot \log_2 P(a_i Q_3, Y)$
$a_1$	0		
$a_2$	0.05	0.2	0.46
$a_3$	0.2	0.8	0.26
$P(Q_3, Y)$	0.25	$H(Q_3, Y)$	0.72

	$P(a_i, Q_3, N)$	$P(a_i Q_3, N)$	$-P(a_i Q_3, N) \cdot \log_2 P(a_i Q_3, N)$
$a_1$	0.35	0.47	0.51
$a_2$	0.25	0.33	0.53
$a_3$	0.15	0.2	0.46
$P(Q_3, N)$	0.75	$H(Q_3, N)$	1.5

	$H(Q_i, Y)$	$H(Q_i, N)$	$m_H(Q_i)$
$Q_1$	0.92	0.95	0.95
$Q_2$	0.95	0.93	0.95
$Q_3$	0.72	1.5	1.5

Minima z  $m_H(Q)$  se nabývá u otázek  $Q_1, Q_2$  – do kořene stromu vybereme např.  $Q_1$  a budeme pokračovat s výběrem otázek pro následující uzly.

## 2. Výběr otázky pro uzel daný cestou $(Q_1, Y)$

U otázek  $Q_2$  a  $Q_3$ , které ještě nebyly použity, máme určit maximální entropii zbývajících rozhodnutí při obou možných odpovědích  $Y, N$  (a odpovědi  $Y$  na otázku  $Q_1$ ) a pak vybrat minimální. Avšak otázka  $Q_2$  již vede k terminálním uzlům (viz Obr. 2.2), takže entropie zbývajících rozhodnutí je u ní nulová a  $Q_2$  bude jistě vybrána. Výpočet entropie u  $Q_3$  tak již nemusíme provádět.

## 3. Výběr otázky pro uzel daný cestou $(Q_1, N)$

U otázek  $Q_2$  a  $Q_3$ , které ještě nebyly použity, opět určíme maximální entropii zbývajících rozhodnutí při obou možných odpovědích  $Y, N$  (a odpovědi  $N$  na otázku  $Q_1$ ):

	$P(a_i, Q_1, N; Q_2, Y)$	$P(a_i Q_1, N; Q_2, Y)$	$-P \cdot \log_2 P$
$a_1$	0.2	0.5	0.5
$a_2$	0		
$a_3$	0.2	0.5	0.5
$P(Q_1, N; Q_2, Y)$	0.4	$H(Q_1, N; Q_2, Y)$	1

	$P(a_i, Q_1, N; Q_2, N)$	$P(a_i Q_1, N; Q_2, N)$	$-P \cdot \log_2 P$
$a_1$	0		
$a_2$	0		
$a_3$	0.15	1	0
$P(Q_1, N; Q_2, N)$	0.15	$H(Q_1, N; Q_2, N)$	0

	$P(a_i, Q_1, N; Q_3, Y)$	$P(a_i Q_1, N; Q_3, Y)$	$-P \cdot \log_2 P$
$a_1$	0		
$a_2$	0		
$a_3$	0.2	1	0
$P(Q_1, N; Q_3, Y)$	0.2	$H(Q_1, N; Q_3, Y)$	0

	$P(a_i, Q_1, N; Q_3, N)$	$P(a_i Q_1, N; Q_3, N)$	$-P \cdot \log_2 P$
$a_1$	0.2	0.57	0.46
$a_2$	0		
$a_3$	0.15	0.43	0.52
$P(Q_1, N; Q_3, N)$	0.35	$H(Q_1, N; Q_3, N)$	0.98

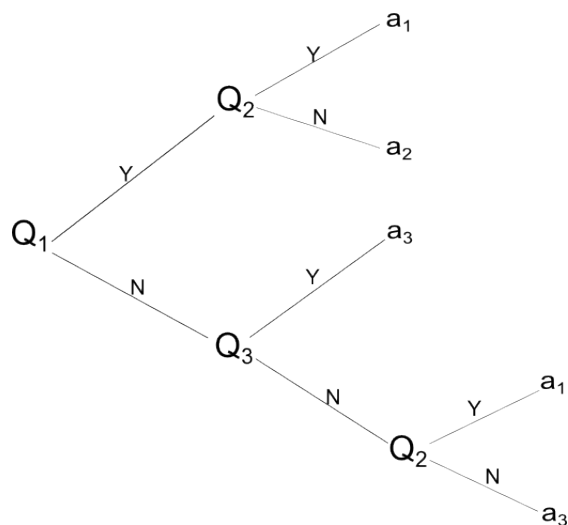
	$H(Q_1, N; Q_i, Y)$	$H(Q_1, N; Q_i, N)$	$m_H(Q_1, N; Q_i)$
$Q_2$	1	0	1
$Q_3$	0	0.98	0.98

Podle kritéria minimalizace maximální zbývající entropie vybereme otázku  $Q_3$ .

4. V částečně rozvinutém stromu je třeba ještě doplnit otázku v uzlu na cestě dané odpověďmi  $(Q_1, N) - (Q_3, N)$ , neboť zde jsou stále dvě možná rozhodnutí. Zbývá jedině otázka  $Q_2$ .

Výsledný strom  $T_H$  pro danou rozhodovací tabulku při použití kritéria minimalizace maximální zbývající entropie  $m_H(Q)$  je na Obr. 2.4.

V této konstrukci rozhodovacího stromu jsme využili pravděpodobnostní distribuci, což vede k lepšímu stromu z hlediska střední délky  $d_P(T)$ , kterou opět vypočteme jako vážený průměr délek cest do terminálních uzlů  $w$  (násobíme je pravděpodobností jejich dosažení  $P(w)$ ):



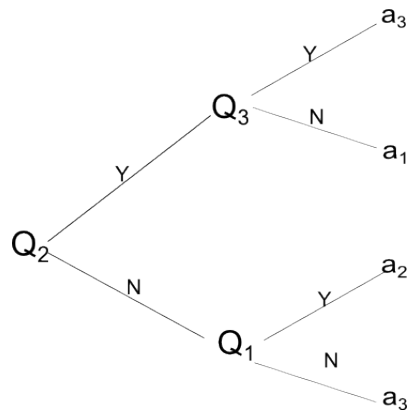
Obrázek 2.4: Rozhodovací strom sestrojený přímým algoritmem s výběrem otázky podle kritéria minimalizace maximální zbývající entropie  $m_H(Q)$

$w$	$P(w)$	$d(w)$	$P(w).d(w)$
$(Q_1, Y) - (Q_2, Y) - a_1$	0.15	2	0.3
$(Q_1, Y) - (Q_2, N) - a_2$	0.3	2	0.6
$(Q_1, N) - (Q_3, Y) - a_3$	0.2	2	0.4
$(Q_1, N) - (Q_3, N) - (Q_2, Y) - a_1$	0.2	3	0.6
$(Q_1, N) - (Q_3, N) - (Q_2, N) - a_3$	0.15	3	0.45
$d_P(T_H) =$			2.35

Pro dosažení rozhodnutí je tedy třeba podle tohoto rozhodovacího stromu  $T_H$  položit v průměru  $d_P(T_H) = 2.35$  otázek.

## 2.4 Optimální rozhodovací strom

Předností přímých algoritmů je rychlost výpočtu a nízké nároky na paměť. Nevýhodou je, že není zaručena optimalita sestrojeného rozhodovacího stromu. Kombinatoricky náročná úloha nalezení optimálního rozhodovacího stromu s minimální střední délkou má řadu principiálně odlišných algoritmických řešení:



Obrázek 2.5: Optimální rozhodovací strom  $T_O$  sestrojený algoritmem větví a mezí

- algoritmus větví a mezí používající entropii k výpočtu dolní meze,
- algoritmus dynamického programování,
- algoritmus  $AO^*$  prohledávání AND/OR grafu řízeného heuristikou.

Uvedené algoritmy nebudeme v tomto textu popisovat. Pro zajímavost je však na Obr. 2.5 uveden optimální rozhodovací strom sestrojený algoritmem větví a mezí pro náš příklad rozhodovací tabulky. Jeho střední délka  $d_P(T_O) = 2$  je tedy minimální. Optimální rozhodovací strom nás tedy dovede k dosažení rozhodnutí v každé situaci již po zodpovězení dvou otázek.



## Kapitola 3

# Konstrukce rozhodovacích stromů z dat

V této kapitole si ukážeme, jak lze konstrukci rozhodovacích stromů z rozhodovacích tabulek aplikovat při dobývání znalostí z databází. Uvažujme například ilustrativní matici dat, která zachycuje výsledky vyšetření pacientů a stanovení jejich diagnózy  $a_1$ ,  $a_2$ , nebo  $a_3$ . Vyšetření spočívá ve třech testech  $Q_1$ ,  $Q_2$ ,  $Q_3$  s pozitivním  $Y$  nebo negativním  $N$  závěrem.

Cílem dobývání znalostí z těchto dat bude zjistit, jak na základě těchto dat efektivně postupovat při stanovování diagnózy (kterým testem začít, jak dále pokračovat). Je zřejmé, že tento cíl vede k úloze konstrukce co nejlepšího rozhodovacího stromu odpovídajícího daným datům.

Pokud jsou data konzistentní v tom smyslu, že neobsahují dva řádky se stejnými hodnotami testů a různými diagnózami, můžeme je shrnout v rozvinuté rozhodovací tabulce - v našem příkladě to bude tabulka s pravděpodobnostní distribucí, v níž rozhodovací pravidla a jejich pravděpodobnosti jsou získány z daných dat.

Matice dat				
pacient	$Q_1$	$Q_2$	$Q_3$	diagnóza
k1	$Y$	$Y$	$N$	$a_1$
k2	$Y$	$Y$	$N$	$a_1$
k3	$Y$	$Y$	$N$	$a_1$
k4	$N$	$Y$	$N$	$a_1$
k5	$N$	$Y$	$N$	$a_1$
k6	$N$	$Y$	$N$	$a_1$
k7	$N$	$Y$	$N$	$a_1$
k8	$Y$	$N$	$Y$	$a_2$
k9	$Y$	$N$	$N$	$a_2$
k10	$Y$	$N$	$N$	$a_2$
k11	$Y$	$N$	$N$	$a_2$
k12	$Y$	$N$	$N$	$a_2$
k13	$Y$	$N$	$N$	$a_2$
k14	$N$	$Y$	$Y$	$a_3$
k15	$N$	$Y$	$Y$	$a_3$
k16	$N$	$Y$	$Y$	$a_3$
k17	$N$	$Y$	$Y$	$a_3$
k18	$N$	$N$	$N$	$a_3$
k19	$N$	$N$	$N$	$a_3$
k20	$N$	$N$	$N$	$a_3$

$P$	0.15	0.2	0.05	0.25	0.2	0.15
$Q_1$	$Y$	$N$	$Y$	$Y$	$N$	$N$
$Q_2$	$Y$	$Y$	$N$	$N$	$Y$	$N$
$Q_3$	$N$	$N$	$Y$	$N$	$Y$	$N$
$A$	$a_1$	$a_1$	$a_2$	$a_2$	$a_3$	$a_3$

Rozhodovací stromy konzistentní s touto rozhodovací tabulkou tedy budou odpovídat daným datům. Pro jejich konstrukci tak můžeme použít pojmy a algoritmy popsané v předchozích kapitolách.

Připomeňme si rozhodovací strom  $T_H$  sestrojený v minulé kapitole pro tuto rozhodovací tabulku přímým algoritmem při použití kritéria minimalizace maximální zbývající entropie  $m_H(Q)$  (Obr. 2.4). Tento strom tedy

odpovídá datům z našeho ilustrativního příkladu. Pro dosažení diagnózy je třeba podle tohoto rozhodovacího stromu  $T_H$  provést v průměru  $d_P(T_H) = 2.35$  testů. Pokud bychom měli dokonce k dispozici optimální rozhodovací strom (Obr. 2.5), pak bychom pro dosažení diagnózy podle něj potřebovali pouze dva testy.

### 3.1 Princip algoritmu TDIDT

Nejčastěji používaný algoritmus konstrukce rozhodovacích stromů z dat je označován zkratkou TDIDT **Top Down Induction of Decision Tree**. Jak vyplývá už z názvu, jde o přímý algoritmus konstrukce rozhodovacích stromů, v němž jsou potřebné pravděpodobnosti získávány z daných dat.

**Kritérium výběru otázky** v dosaženém uzlu  $u$  je založeno na výpočtu entropie. Pro každou přípustnou otázku  $Q$  a možnou odpověď  $r$  určíme frekvenci odpovědi  $P(Q, r)$  a frekvenci každého rozhodnutí  $P(Q, r, a_i)$  v datech zúžených na část odpovídající kombinaci odpovědí na cestě z kořene stromu do  $u$ . Odtud vypočteme **podmíněné pravděpodobnosti a entropii** zbývajících rozhodnutí

$$P(a_i|Q, r) = \frac{P(Q, r, a_i)}{P(Q, r)}$$

**Výpočet entropie zbývajících rozhodnutí:**

$$H(Q, r) = - \sum_{a_i} P(a_i|Q, r) \cdot \log_2 P(a_i|Q, r)$$

Kritériem je nejčastěji **střední entropie**:

$$\bar{H}(Q) = \sum_r P(Q, r) \cdot H(Q, r)$$

### 3.2 Příklad výpočtu algoritmem TDIDT

Konstrukci rozhodovacího stromu z dat algoritmem TDIDT si ukážeme na ilustrativním příkladě převzatém z [1], kde lze nalézt další podrobnosti. Matice dat v tomto případě zachycuje atributy klientů banky (příjem vysoký či nízký; konto vysoké, střední či nízké; pohlaví; nezaměstnanost) spolu s

údajem o splácení jim poskytnutých úvěrů ( $A$  znamená řádné splácení,  $N$  znamená problémy se splácením). Konstrukce rozhodovacího stromu může bance přinést znalosti o tom, kterým klientům poskytnout úvěr v budoucnu podle jejich charakteristiky.

Matice dat					
klient	příjem	konto	pohlaví	nezaměstnaný	úvěr
k1	v	v	ž	n	A
k2	v	v	m	n	A
k3	n	n	m	n	N
k4	n	v	ž	a	A
k5	n	v	m	a	A
k6	n	n	ž	a	N
k7	v	n	m	n	A
k8	v	n	ž	a	A
k9	n	s	m	a	N
k10	v	s	ž	n	A
k11	n	s	ž	a	N
k12	n	s	m	n	A

## Konstrukce rozhodovacího stromu z příkladu

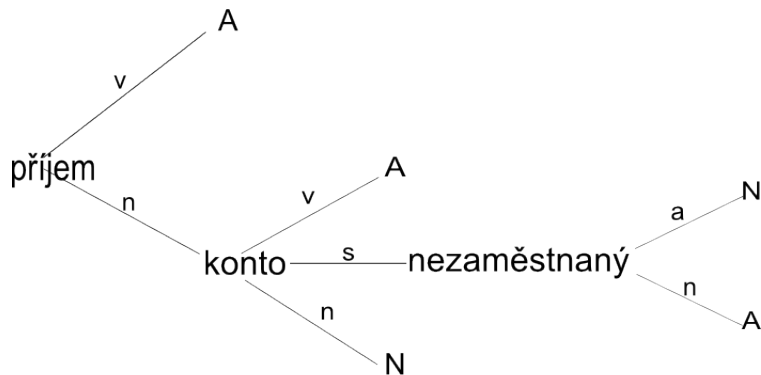
1. Prozkoumáme možné atributy pro umístění do kořene stromu:

- $\bar{H}(\text{příjem}) = 0.57$
- $\bar{H}(\text{konto}) = 0.92$
- $\bar{H}(\text{pohlaví}) = 0.92$
- $\bar{H}(\text{nezaměstnaný}) = 0.83$

Nejmenší střední entropii má atribut "příjem" - kořen stromu

2. Umístíme do kořene atribut "příjem" a prozkoumáme větve odpovídající hodnotám atributu v kořeni stromu:

- "příjem(v)" je terminální s rozhodnutím "úvěr(A)"
- "příjem(n)" je neterminální – určíme střední entropie zbývajících atributů v datech zúžených na případy "příjem(n)":  
 $\bar{H}(\text{konto}) = 0.39$ ,  $\bar{H}(\text{pohlaví}) = 0.97$ ,  $\bar{H}(\text{nezaměstnaný}) = 0.98$   
 Nejmenší střední entropii pro "příjem(n)" má atribut "konto"



Obrázek 3.1: Rozhodovací strom zkonstruovaný z dat algoritmem TDIDT

- Rozvětvíme podle hodnot atributu "konto" v uzlu "příjem(n)":
  - "příjem(n) ∧ "konto(v)" je terminální s "úvěr(A)"
  - "příjem(n) ∧ "konto(n)" je terminální s "úvěr(N)"
  - "příjem(n) ∧ "konto(s)" je neterminální
    - určíme střední entropie zbývajících atributů v datech zúžených na případy "příjem(n) ∧ "konto(s)"
    - $\bar{H}(\text{pohlaví}) = 0.67, \bar{H}(\text{nezaměstnaný}) = 0$
    - Vybereme tedy atribut "nezaměstnaný":
    - hodnota "a" - "úvěr(N)", hodnota "n" - "úvěr(A)"

Výsledný strom je na obr. 3.1. Závěrem analýzy je pro banku zjištění, že hlavním atributem pro rozhodování o poskytnutí úvěru je výše příjmu klienta, na druhé straně na atributu pohlaví klienta nezáleží.

Konstrukce rozhodovacího stromu z dat algoritmem TDIDT je pro akademické účely dostupná v dataminingovém systému LISp-Miner [5].

# Literatura

- [1] Berka, P.: *Dobývání znalostí z databází*. Academia, Praha 2003, 366 p.
- [2] Ganapathy, S., Rajaraman, V.: Information Theory Applied to the Conversion of Decision Tables to Computer Programs. *Communications of the ACM* 16, 1973, pp. 532-539
- [3] Jiroušek, R.: *Metody reprezentace a zpracování znalostí v umělé inteligenci*. VŠE, 1995, 103 str.
- [4] Martelli, A., Montanari, U.: Optimizing Decision Trees through Heuristically Guided Search. *Communications of the ACM* 21, 1978, pp. 1025-1039
- [5] Šimůnek, M.: *Systém LISp-Miner: akademický systém pro dobývání znalostí z databází: historie vývoje a popis ovládání*. Praha: Oeconomica, 2010. ISBN 978-80-245-1699-8.
- [6] Wu, X. a kol.: Top 10 Algorithms in Data Mining. *Knowledge and Information Systems*, 2008, 14, s. 1 – 37

$p$	$-\log_2 p$	$-p \log_2 p$	$p$	$-\log_2 p$	$-p \log_2 p$
0,01	6,6439	0,07	0,26	1,9434	0,51
0,02	5,6439	0,11	0,27	1,8890	0,51
0,03	5,0589	0,15	0,28	1,8365	0,51
0,04	4,6439	0,19	0,29	1,7859	0,52
0,05	4,3219	0,22	0,3	1,7370	0,52
0,06	4,0589	0,24	0,31	1,6897	0,52
0,07	3,8365	0,27	0,32	1,6439	0,53
0,08	3,6439	0,29	0,33	1,5995	0,53
0,09	3,4739	0,31	0,34	1,5564	0,53
0,1	3,3219	0,33	0,35	1,5146	0,53
0,11	3,1844	0,35	0,36	1,4739	0,53
0,12	3,0589	0,37	0,37	1,4344	0,53
0,13	2,9434	0,38	0,38	1,3959	0,53
0,14	2,8365	0,40	0,39	1,3585	0,53
0,15	2,7370	0,41	0,4	1,3219	0,53
0,16	2,6439	0,42	0,41	1,2863	0,53
0,17	2,5564	0,43	0,42	1,2515	0,53
0,18	2,4739	0,45	0,43	1,2176	0,52
0,19	2,3959	0,46	0,44	1,1844	0,52
0,2	2,3219	0,46	0,45	1,1520	0,52
0,21	2,2515	0,47	0,46	1,1203	0,52
0,22	2,1844	0,48	0,47	1,0893	0,51
0,23	2,1203	0,49	0,48	1,0589	0,51
0,24	2,0589	0,49	0,49	1,0291	0,50
0,25	2,0000	0,50	0,5	1,0000	0,50

$p$	$-\log_2 p$	$-p \log_2 p$	$p$	$-\log_2 p$	$-p \log_2 p$
0,51	0,9714	0,50	0,76	0,3959	0,30
0,52	0,9434	0,49	0,77	0,3771	0,29
0,53	0,9159	0,49	0,78	0,3585	0,28
0,54	0,8890	0,48	0,79	0,3401	0,27
0,55	0,8625	0,47	0,8	0,3219	0,26
0,56	0,8365	0,47	0,81	0,3040	0,25
0,57	0,8110	0,46	0,82	0,2863	0,23
0,58	0,7859	0,46	0,83	0,2688	0,22
0,59	0,7612	0,45	0,84	0,2515	0,21
0,6	0,7370	0,44	0,85	0,2345	0,20
0,61	0,7131	0,44	0,86	0,2176	0,19
0,62	0,6897	0,43	0,87	0,2009	0,17
0,63	0,6666	0,42	0,88	0,1844	0,16
0,64	0,6439	0,41	0,89	0,1681	0,15
0,65	0,6215	0,40	0,9	0,1520	0,14
0,66	0,5995	0,40	0,91	0,1361	0,12
0,67	0,5778	0,39	0,92	0,1203	0,11
0,68	0,5564	0,38	0,93	0,1047	0,10
0,69	0,5353	0,37	0,94	0,0893	0,08
0,7	0,5146	0,36	0,95	0,0740	0,07
0,71	0,4941	0,35	0,96	0,0589	0,06
0,72	0,4739	0,34	0,97	0,0439	0,04
0,73	0,4540	0,33	0,98	0,0291	0,03
0,74	0,4344	0,32	0,99	0,0145	0,01
0,75	0,4150	0,31	1	0,0000	0,00