

ANALÝZA VÝSLEDKŮ SLEDOVÁNÍ KONDIČNÍCH PŘEDPOKLADŮ SPORTOVNÍHO VÝKONU MLADÝCH KRASOBRUSLAŘŮ POMOCÍ METOD PRO DOBÝVÁNÍ ZNALOSTÍ Z DAT

Jiří Helešic, Jan Górecki

Klíčová slova:

dobývání znalostí z dat, korelace, hlavní komponenty, asociační pravidla, rozhodovací stromy, RapidMiner

Key words:

data mining, correlation, principle components, association rules, decision trees, RapidMiner

Abstrakt

V letech 2001 – 2010 jsme na základě spolupráce s krasobruslařskými oddíly Moravskoslezského kraje a Českým krasobruslařským svazem posuzovali výkony mladých krasobruslařů v testové baterii, která byla zaměřena na projevy výbušné síly dolních končetin. Celkově absolvovalo testování na 280 probandů ve věku 6 – 19 let. Naše výsledky jsme zpracovali analytickými metodami, které spadají pod metodologii nazývanou Dobývání znalostí z dat (angl. Data Mining) jež nám mohou naznačit některé skryté vztahy mezi sledovanými proměnnými a případně potvrdit či vyvrátit předpokládané kauzální vztahy.

Abstract

Between years 2001 – 2010 we have cooperated with figure skating clubs from Moravian-Silesian region and Czech ice skating association. We have been evaluating kinetic performances of young figure skaters in a test battery which was focused on expression of explosive strength of legs. Testing procedure has completed about 280 young figure skaters aged 6 – 19 years. To find eventual hidden relations among monitored variables and to confirm or to overcome assumed causal relations we have used selected analytical methods from Data Mining methodology.

Úvod

Krasobruslení je druh bruslařského zimního sportu, ve kterém je kladen důraz na umělecké ztvárnění piruet, skoků a kroků za hudebního doprovodu. Krasobruslení je sportovní odvětví, spojující sportovní a umělecký projev a řadí se mezi esteticko-koordinační sporty. Do krasobruslařských soutěží zahrnujeme soutěže Jednotlivců (muži a ženy), Sportovní páry, Tance na ledě a Synchronizované krasobruslení. V České republice jsou závodníci rozdělení dle věku do skupin: Nejmladší nováčci a Nováčci (6, 7 let), Žactvo nejmladší (do 8-9 let), Žactvo mladší (do 12 let), Žactvo (do 15 let) a Junioři do 19 let a Senioři.

Specifická pohybová činnost krasobruslaře při volné jízdě má střídavý charakter tělesného zatížení, kde dominantní úlohu hrají pravidly požadované prvky (skoky, piruety a krokové sekvence). Ty mají být provedeny v souladu s požadavky mezinárodní krasobruslařské federace a jejich posuzování odráží i „výbušnost“ provedení.

V naší studii jsme se soustředili na oblast dynamicko-silových charakteristik předpokladů sportovního výkonu krasobruslaře jednotlivce. Sledováním úspěšných krasobruslařů v kategorii seniorů bylo nezpochybnitelně prokázáno, že ač je výkon krasobruslaře

multifaktoriální, je především založen na realizaci rychlostně silového potenciálu svalstva dolních končetin.

Naše výsledky ukazují sledování téměř tří set mladých krasobruslařů ve věku 6 – 19 let. Jedná se o mladé sportovce zařazené do „projektu talentované mládeže“ Českého krasobruslařského svazu a členy krasobruslařských oddílů z Karviné, Třince, Ostravy, Frýdku Místku, Orlové a Havířova ve stejných věkových kategoriích.

1. Sportovní výkon jako výsledek adaptace

Lidský organismus je ovlivňován nejrůznějšími činiteli z vnějšího a vnitřního prostředí. Reakce organismu na vlivy prostředí se obvykle dělí na bezprostřední reakce a adaptace. Bezprostřední reakce chápeme jako rychlé odpovědi na zpravidla jednorázový podnět, které se uskutečňují prostřednictvím vzrušivých soustav, které jsou k těmto funkcím předem připraveny. Pro adaptaci nejsou naproti tomu zprostředkující soustavy předem připravené, a proto dochází vlivem opakovaných podnětů, ať přerušovaných či kontinuálních, k postupné přestavbě příslušných orgánů a jejich funkcí [11]. Adaptace je dynamická schopnost organismu vyrovnávat se s vlivy prostředí. Využití poznatků o mechanismech adaptace organismu v tréninkovém procesu je východiskem pro zvyšování sportovní výkonnosti jedince.

Nezbytným funkčním základem sportovní výkonnosti je výkonnostní potenciál, který chápeme jako aktuální sumu funkčních předpokladů k pohybové činnosti. Adaptace v tomto užším pojetí na tréninkové a soutěžní zatížení se nejvýrazněji projevuje v práci systémů zajišťujících energetickou stránku výkonnosti [6].

2. Měření a testování

Terénní diagnostické prostředky musí splňovat požadavky pro testování pohybových schopností jako nepřímě měřitelných teoretických vlastností, na které usuzujeme z empirických přímo pozorovatelných výsledků pohybových výkonů v testech, především s ohledem na objektivitu, spolehlivost a obsahovou validitu. Rozhodli jsme se pro výběr prostředků již standardizovaných. Standardizovaných terénních testů, které testují pohybové výkony rychlostního charakteru a dynamické síly dolních končetin, je celá řada. Jsou to především různé varianty skoků a dynamometrie. Rychlost lokomoce testujeme především pomocí běhů a bruslení na různou vzdálenost.

Prostorové nároky na použití terénních diagnostických prostředků byly dalším výběrovým kritériem. Vzhledem k našemu záměru sledovat vybrané skupiny krasobruslařů v průběhu celoroční přípravy byly voleny takové prostředky, které lze provádět v uzavřených prostorách zimního stadionu. K dispozici se nabízí využití ochozů zimních stadiónů, které však limitují délku běžeckých úseků.

Využili jsme našich zkušeností z testování hráčů ledního hokeje, jež jsme prováděli v letech 1996 – 2001. Při detailnějším výběru testů jsme se řídili těmito dalšími skutečnostmi:

- a) délka trvání zatížení v jednotlivém testu do 15 s., nad tento interval zatížení se jedná o různé formy sledování silové a rychlostní vytrvalosti, u skoků je to do úrovně 16 opakování skoků [12],
- b) ze „skokanských“ testů (různé varianty víceskoků) jsme vybrali trojskok na pravé (levé) noze a tradiční test „skok daleký z místa“, tento především pro svou

jednoduchost a také proto, že v minulosti byl u této věkové kategorie sportovců dlouhou dobu užíván, a je tedy možné porovnání našich výsledků,

- c) rychlostní schopnosti se tradičně měří běhy na určitou vzdálenost. Pominuli jsme oblast rychlosti reakční a zaměřili se na akční rychlost komplexních pohybových aktů - běžeckou rychlost. Na základě předchozích zkušeností jsme pro krasobruslaře vybrali pro naše podmínky vyhovující testy – běh na 15 m s pevným startem a bruslení na 15 m s pevným startem. Pro naši práci se nám podařilo zajistit technické prostředky Jumper a FiTRO Speed check od firmy FiTRONiC s.r.o. Bratislava. Tyto jsme použili pro měření výbušné síly dolních končetin a lokomoční rychlosti běhu a jízdy na bruslích.

FiTRO speed check představuje zařízení na registraci a vyhodnocení rychlosti, zrychlení a dráhy při přímém pohybu do 30 m. Na rozdíl od měření pomocí fotobuněk je možné průběžné snímání uvedených parametrů.

Jumper je jednoduchý výskokový ergometr pracující na principu měření doby opory a letu s přesností na tisícinu sekundy. Při měření opakovaných výskoků umožňuje zprostředkované měření výšky výskoku, výkonu v aktivní fázi odrazu, průměrného výkonu, zrychlení v aktivní fázi odrazu a rychlosti vzletu v závěrečné fázi odrazu.

Na základě uvedených kritérií byly vybrány následující *diagnostické prostředky*:

1. Opakované vertikální skoky na zařízení JUMPER firmy FiTRONiC, s.r.o. Bratislava - test výbušné síly dolních končetin dle metodiky: [9]. Hodnotili jsme: h - výšku výskoku, tc - dobu opory, P - výkon v aktivní části odrazu, h/tc - koeficient efektivity odrazu. Délka trvání testu 3 x 10 s. Zaznamenávány byly průměrné hodnoty zjištěné ve třech nejúspěšnějších výskokových cyklech.
2. Skok daleký z místa (dále SDM) - test výbušné síly dolních končetin (horizontální směr) - metodika dle: [10]. Zaznamenávali jsme delší ze dvou měřených pokusů.
3. Víceskok na jedné noze, trojskok na pravé (TP) a levé noze (TL) - jde o test výbušné síly dolních končetin (odrazové síly v horizontálním směru) - metodika dle: [9]. Zaznamenávali jsme nejdelší ze tří pokusů na každé noze.
4. Běh na 15 m z vysokého startu - test běžecké rychlosti a akcelerace rychlosti, měřeno na zařízení FiTRO speed check firmy FiTRONiC, s.r.o. Bratislava. Zaznamenávali jsme celkový dosažený čas (t15_běh), průměrnou rychlost v 1., 2., 3. s. běhu (a1s_běh, a2s_běh, a3s_běh).
5. Bruslení na 15 m z vysokého startu - test bruslařské rychlosti a akcelerace rychlosti, měřeno na zařízení FiTRO speed check firmy FiTRONiC, s.r.o. Bratislava. Zaznamenávali jsme celkový dosažený čas, průměrnou rychlost v 1., 2., 3. s. bruslení (a1s_led, a2s_led, a3s_led).

Testování mimo ledovou plochu trvá 45 minut. Celková doba zatížení jednoho sportovce je 60 až 70 s. Testované osoby absolvují testovou baterii v určeném pořadí, mezi jednotlivými testy je dostatečný odpočinek (5 min.) k plné regeneraci sil.

Protože byly použity standardizované diagnostické prostředky s výjimkou měření běhu a bruslení na 20 m pomocí zařízení FiTRO speed check, byla ověřována pouze spolehlivost testových výsledků v těchto položkách testové baterie. Ověřování spolehlivosti proběhlo u skupiny hokejistů v roce 1999. Pro krasobruslaře jsme zvolili kratší vzdálenost.

Spolehlivost výsledků v testech běhu a bruslení na 20 m byla ověřována metodikou test - retest, tedy stability v čase a je vyjádřena pomocí Pearsonova součinného koeficientu párové korelace. Výsledky ukazuje Tabulka 1. Testy byly opakovány s intervalem odpočinku 10 min.

Tabulka 1: Koeficienty spolehlivosti testů 20 m běh a bruslení měřené pomocí FiTRO Speed check

Testový výsledek	Test – retest běh (r_{xx})	Test – retest bruslení (r_{xx})
20 m celkový čas	0,985	0,989
Prům. rychlost 1.s.	0,966	0,968
Prům. rychlost 2.s.	0,980	0,978
Prům. rychlost 3.s.	0,989	0,991

Vzhledem k tomu, že pro individuální diagnostiku je postačující reliabilita $r_{xx} \geq 0,85$ [8], naše výsledky vykazují vysokou reliabilitu – stabilitu v čase a můžeme tyto testy použít pro posouzení individuálních diferencí v našem výzkumu. V ostatních testových položkách jsme vycházeli z údajů o spolehlivosti testů citovaných v [7], [8], [9] a [10] a dodrželi jsme předepsané protokoly měření.

Dalšími sledovanými proměnnými jsou věk, výška a váha jedince a % tuku měřeného pomocí měření kožních řas kaliperem dle metodiky Evy Blahušové.

3. Analýza dat

Pro podrobnou analýzu výsledků sledování kondičních předpokladů sportovního výkonu mladých krasobruslařů jsme použili analytické metody, které spadají pod metodologii nazývanou Dobývání znalostí z dat (angl. Data Mining). Obecně tyto metody umožňují získávání netriviálních skrytých a potencionálně užitečných informací z dat [4]. Z celé řady metod, které lze najít v literatuře [3], [4] a které lze použít v oblastech vědy a výzkumu (naš případ), v komerční sféře (v Business Intelligence jako podpoře e-commerce systémů) nebo například pro podporu řízení složitých automatizovaných průmyslových zařízení (např. expertní systémy pro řízení čističek odpadních vod), jsme vybrali metody podle několika kritérií: 1) metoda musela umožnit použít jako svůj vstup naše data, 2) metoda dávala srozumitelné a dobře interpretovatelné výsledky, 3) metoda byla implementována v systému pro dobývání znalostí z dat RapidMiner [5]. Z mnoha metod a analýz, které jsme nad daty provedli, jsou potom v článku popsány ty, které přinesly zajímavé výsledky. Nejdříve jsou tedy popsány výsledky základní statistické analýzy dat (průměry, odchylky a výpočet korelační matice) a dále jsou popsány výsledky analýz metodou hlavních komponent, metodou pro hledání asociačních pravidel a metodou pro tvorbu rozhodovacích stromů. V závěru je pak shrnut celkový přínos práce, který zahrnuje jak ověření obsahové validity baterie testů, tak nalezení závislostí v datech, které a) správně ohodnotí dynamicko-silové vlastnosti dolních končetin u každého jedince (jak již trénujícího, tak nově příchozího), b) umožní formovat trénink jedince na základě průběžných výsledků testů, c) umožní zjednodušit hodnocení výkonnosti jedince pomocí hodnot jenom malé podmnožiny testů.

4. Popis dat a atributů

Data obsahovala údaje o 287 jedincích. Každý jedinec byl popsán na základě 20 atributů, které jsme rozdělili vzhledem k charakteru do 5 skupin (viz. popis diagnostických příspěvků výše). Tyto skupiny atributů jsou:

Základní údaje (dále v textu označeno jako *Zákl*): Pohlaví (z, m), Věk (roky), Výška (cm), Váha (kg), %Tuku

Charakteristika skoku do dálky (*Dálka*): SDM (cm), TP (cm), TL (cm)

Charakteristika rozběhu na „suché“ podlaze na 15m (*Běh*): t15_běh (s), a1s_běh (m/s), a2s_běh (m/s), a3s_běh (m/s)

Charakteristika jízdy na bruslích na 15m (*Běh_led*): t15_led (s), a1s_led (m/s), a2s_led (m/s), a3s_led (m/s)

Charakteristika skoku snožmo do výšky (Výška): t_opory (ms), P (W/kg), h (cm), h/tc (bezrozměrná)

V datech se vyskytly také u některých atributů chybějící údaje – s tímto faktem bylo potřeba počítat při analýzách jednotlivými metodami.

Dle kritérií na výběr metod pro analýzu dat (viz. výše) jsme zvolili tyto metody:

1. *Korelační matice* – většina atributů jsou reálné atributy, má tedy smysl zjišťovat u nich stupeň lineární závislosti (vyjádřený koeficientem korelace)
2. *Hlavní komponenty* – opět, většina atributů jsou reálné atributy, je možno tedy očekávat skryté faktory měřené původními atributy
3. *Asociační pravidla* – pro hledání závislostí mezi hodnotami atributů
4. *Rozhodovací stromy* – pro detailnější hledání závislosti hodnoty atributu skoku do výšky h, který má pro každého krasobruslaře zásadní význam, na ostatních attributech

Analýzy byly provedeny v Data-miningovém systému RapidMiner.

5. Základní statistika

Základní statistické údaje o datovém souboru rozděleném dle jednotlivých věkových kategorií jsou uvedeny v Tabulka 2. Z hodnot průměrů a směrodatné odchylky není zřejmá žádná výrazná chyba v datech.

Tabulka 2: Průměry +/- směrodatné odchylky výsledků jednotlivých testových položek dle věkových kategorií

	chlapci do 9 let	dívky do 9 let	chlapci 10 - 12 let	dívky 10 - 12 let	chlapci 13 - 15 let	dívky 13 - 15 let	chlapci nad 16 let	dívky nad 16 let
počet	20	67	43	118	9	24	4	3
Výška	134.200 +/- 6.346	131.582 +/- 6.858	142.571 +/- 6.337	142.229 +/- 7.034	157.333 +/- 4.472	155.500 +/- 11.128	172 +/- 8.756	163.333 +/- 4.933
Váha	29.950 +/- 3.502	28.261 +/- 3.634	35.155 +/- 3.093	34.462 +/- 4.814	44.222 +/- 3.993	44.042 +/- 7.371	63.250 +/- 12.447	56.333 +/- 6.658
%tuku	5 +/- 1.856	12.481 +/- 2.805	5.010 +/- 1.755	12.225 +/- 2.343	5.011 +/- 1.204	11.959 +/- 2.661	5.600 +/- 1.655	17.100 +/- 4.950
SDM	163.550 +/- 19.605	161.746 +/- 17.695	188.024 +/- 15.461	183.508 +/- 16.355	209.333 +/- 11.467	204.583 +/- 15.712	243.750 +/- 18.644	207.667 +/- 11.240
TP	450.750 +/- 60.570	445.164 +/- 65.600	541.881 +/- 52.130	530.178 +/- 54.049	629.556 +/- 44.727	601.542 +/- 59.637	736.250 +/- 67.129	628.333 +/- 30.551
TL	448.750 +/- 53.246	443.239 +/- 68.259	544.119 +/- 51.137	524.492 +/- 56.516	625.778 +/- 38.091	586.833 +/- 61.088	720 +/- 70.119	619.333 +/- 20.648
t_opory	175 +/- 22.095	171.582 +/- 20.746	172.857 +/- 18.740	173.674 +/- 19.424	178 +/- 20.488	170.375 +/- 13.990	183.500 +/- 18.230	177.667 +/- 26.502
P (W/kg)	34.160 +/- 8.167	36.509 +/- 6.927	41.295 +/- 6.715	42.636 +/- 6.974	43.422 +/- 9.835	46.392 +/- 7.127	51.400 +/- 3.771	46.600 +/- 4.882
h	22.045 +/- 4.577	23.503 +/- 4.198	27.812 +/- 3.770	28.408 +/- 4.111	29.756 +/- 4.304	30.817 +/- 3.689	37.600 +/- 3.249	31.400 +/- 0.656
h/tc	125.175 +/- 36.216	134.303 +/- 31.502	156.536 +/- 31.274	160.383 +/- 31.997	165.400 +/- 45.940	177.542 +/- 29.555	195.250 +/- 17.720	175.667 +/- 25.183
t15_běh	3.257 +/- 0.377	3.305 +/- 0.328	3.233 +/- 0.381	3.180 +/- 0.357	3.037 +/- 0.254	3.065 +/- 0.276	2.863 +/- 0.162	3.220 +/- 0.161
a1s_běh	2.663 +/- 0.644	2.630 +/- 0.596	2.692 +/- 0.523	2.749 +/- 0.573	2.900 +/- 0.451	2.922 +/- 0.495	3.180 +/- 0.443	2.423 +/- 0.742
a2s_běh	4.858 +/- 0.726	4.790 +/- 0.667	4.936 +/- 0.762	5.027 +/- 0.727	5.214 +/- 0.517	5.241 +/- 0.698	5.745 +/- 0.404	4.890 +/- 0.225
a3s_běh	5.801 +/- 0.798	5.655 +/- 0.634	5.859 +/- 0.850	5.972 +/- 0.787	6.264 +/- 0.640	6.285 +/- 0.775	6.877 +/- 0.396	6.067 +/- 0.112
t15_brus	3.490 +/- 0.354	3.587 +/- 0.359	3.328 +/- 0.355	3.405 +/- 0.430	3.255 +/- 0.395	3.273 +/- 0.332	2.898 +/- 0.142	3.323 +/- 0.058
a1s_brus	2.361 +/- 0.600	2.277 +/- 0.536	2.465 +/- 0.528	2.429 +/- 0.661	2.467 +/- 0.678	2.679 +/- 0.510	3.265 +/- 0.398	2.805 +/- 0.049
a2s_brus	4.329 +/- 0.551	4.185 +/- 0.662	4.625 +/- 0.676	4.513 +/- 0.755	4.793 +/- 0.659	4.847 +/- 0.763	5.585 +/- 0.309	4.580 +/- 0.071
a3s_brus	5.272 +/- 0.491	5.097 +/- 0.677	5.663 +/- 0.673	5.484 +/- 0.826	5.856 +/- 0.711	5.780 +/- 0.796	6.735 +/- 0.467	5.570 +/- 0.071

6. Korelační matice

V korelační matici jsou spočteny míry korelace mezi jednotlivými atributy (atribut pohlaví binarizován v předzpracování dat). Korelační koeficienty jsou počítány mezi všemi atributy (tedy i mezi atributy s chybějícími údaji – v případě chybějícího údaje je příspěvek ke koeficientu korelace ignorován).

V obrázku (Obrázek 3: Korelační matice) lze vidět vysokou míru korelace (lineární závislosti) mezi atributy uvnitř jednotlivých podskupin atributů (obdélníky přerušovanou čarou). Očekávaná podobnost (závislost) mezi atributy uvnitř podskupin se tedy potvrdila. U podskupin atributů Běh a Běh_Led lze dokonce vidět lineární závislost nejen uvnitř podskupin samotných, ale i mezi atributy z jedné a druhé skupiny (atributy obou skupin navzájem korelují – obdélník přerušovanou čarou vpravo dole). Výjimečné jsou (v první podskupině atributů Zákl – levý horní obdélník přerušovanou čarou) atributy Pohlaví a %Tuku, které korelují jen mezi sebou navzájem (elipsa). Detailnější vztah mezi těmito atributy se ukáže pomocí asociačních pravidel. Další výjimečný atribut je t-opory (obdélník plnou čarou), který převážně s ničím nekoreluje. Je to dáno tím, že atribut t-opory (čas opory při odrazu) popisuje vlastnost, která je zdá se pro každého krasobruslaře charakteristická a nedá se pozitivně příliš ovlivnit tréninkem – lépe řečeno špatně cíleným tréninkem se dá tato vlastnost zhoršit, kdežto dobrým tréninkem lze vlastnost spíše jen udržet a popř. velmi mírně korigovat.

Výpočet korelační matice nám tedy odhalil množství lineárních závislostí mezi jednotlivými atributy, které se dále projevují i ve výsledcích dalších analýz. Dále výsledek získaný výpočtem korelační matice ukazuje, že záměry konstrukce baterie (= sady měření vlastností) testů byly splněny. Šlo o to, vytvořit sadu testů, kde celá sada bude zachycovat dynamicko-silové vlastnosti měřených jedinců a v této sadě bude vždy nějaká skupina atributů popisovat

jednotlivé typy těchto dynamicko-silových vlastností (lokomoce (= pohyb vpřed), vertikální zdvih těžiště, kombinace obou pohybů).

7. Hlavní komponenty

Při analýze dat touto metodou (popsána v [3]) bylo z datového souboru vyjmuto 26 záznamů s chybějícími údaji. Jedná se o údaje o rychlosti bruslení, které v době letní přípravy nebyly zjišťovány. Dále byl binarizován atribut pohlaví.

Na Obrázku 4 je první část výsledku metody „vlastní čísla korelační matice“. První čtyři komponenty mají 85%ní podíl na velikosti stopy matice vlastních čísel (elipsa). Dalšími komponentami se dále nebudeme zabývat neb jejich podíl na celkovém rozptylu dat je již nevýrazný.

Na dalším obrázku (Obrázek 5: Vlastní vektory) lze vidět první čtyři komponenty a jejich vztah k původním atributům. První komponenta PC1 (48%) se dá charakterizovat jako faktor popisující celkový výkon - který bychom mohli nazvat silový impulz odrazu, další součástí této komponenty jsou základní somatické údaje a věk jedince. Tuto komponentu ovlivňují s poměrně vyrovnaným podílem všechny atributy až na atributy odpovídající elipsám kresleným přerušovanou čarou (tedy pohlaví, %tuku a t_opory), jež ovlivňují tuto komponentu zanedbatelně.

Druhá komponenta PC2 (19%) má téměř identické složení atributů jako PC1 s jediným rozdílem, že tato komponenta zahrnuje i atribut doby opory (t_opory), který nám pomůže vysvětlit celkový rozptyl výsledků z téměř 70%. Z tohoto faktu lze usuzovat, že atribut t_opory, který se v analýze korelační matice jevil jako lineárně nezávislý, má významný vliv na celkovém rozptylu výsledků. Dá se tedy říci, že PC1 a PC2 tvoří hlavní dvojici komponent (skrytých faktorů), které popisují celkový výkon a základní somatické údaje o jedinci.

Další dvě komponenty PC3 (10%) a PC4 (8%) si jsou taktéž mezi sebou podobné. Tyto komponenty zahrnují atributy, pohlaví, % tuku, jež nebyly zahrnuty v PC1 a PC2 a atributy vertikálního výskoku (v elipsách kreslených plnou čarou). Lze usoudit, že tyto komponenty vysvětlují celkový rozptyl výsledků skrytými faktory, které souvisí s vlivem % tuku a pohlaví na vertikální výskok. Ukazuje se tedy, že naše výsledky potvrzují závislost odrazových schopností na pohlaví jedince a negativní ovlivnění výsledků odrazu nadbytečnou tukovou tkání.

8. Asociační pravidla

Ve dvou předchozích metodách jsme se dozvěděli, zda existují lineární závislosti mezi jednotlivými reálnými atributy, popř. zda v datech existují nějaké skryté faktory (komponenty), jejichž počet je sice výrazně menší než počet původních atributů, avšak stále zachycují většinu rozptylu v datech. Zatím jsme se však stále nedozvěděli nic o vztahu určitých hodnot (intervalů hodnot) jedné skupiny atributů k hodnotám (intervalům hodnot) druhé skupiny atributů. Pro zjištění a zachycení takovýchto vztahů je vhodná metoda pro získávání asociačních pravidel.

Pro metodu (popsána v [1]), která získávala z dat asociační pravidla, bylo nutné předzpracování dat. Byla provedena diskretizace všech reálných atributů do 5 ekvifrekvenčních intervalů. Dále byl nutný převod všech kategoriálních atributů na binární.

Jako kvantifikátor byla zvolena fundovaná implikace.

Experimentálně byla metoda nastavena takto:

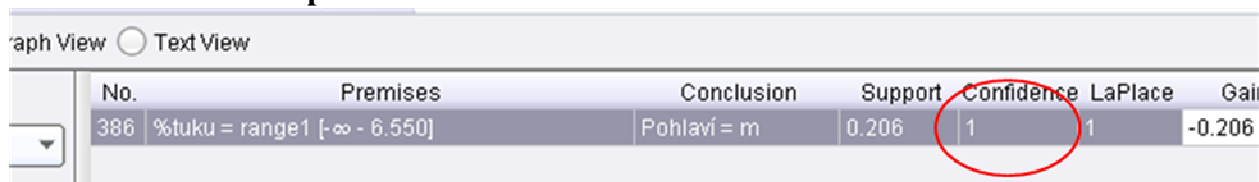
- minimální podpora – 8%
- minimální spolehlivost – 90% - bylo možné nastavit takto vysoko, jelikož data jsou získána přesným měřením objektivní skutečnosti (neobsahují nějaké subjektivní veličiny, jako např. u dotazníků)

Metoda pracovala i s atributy s chybějícími údaji.

8.1. Výsledky analýzy metodou pro hledání asociačních pravidel

V první části analýzy jsme zvolili do množiny atributů {ancetendent, sukcedent} atributy z podskupin Zákl, Výška, Dálka.

Obrázek 1: Asociační pravidla 1



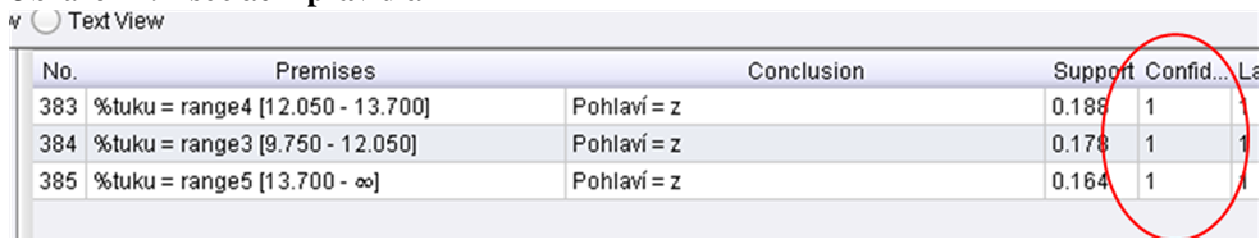
No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain
386	%tuku = range1 [-∞ - 6.550]	Pohlaví = m	0.206	1	1	-0.206

Objevilo se nám pravidlo (viz. Obrázek 1: Asociační pravidla 1)

Málo tučný => muž

Toto pravidlo jsme očekávali vzhledem k vysoké korelaci atributů %tuku a Pohlaví, nicméně nás překvapilo, že hodnota spolehlivosti (definice v [1]) tohoto pravidla je 100% (viz. elipsa výše).

Obrázek 2: Asociační pravidla 2



No.	Premises	Conclusion	Support	Confid...	La
383	%tuku = range4 [12.050 - 13.700]	Pohlaví = z	0.188	1	
384	%tuku = range3 [9.750 - 12.050]	Pohlaví = z	0.178	1	
385	%tuku = range5 [13.700 - ∞]	Pohlaví = z	0.164	1	

Podobně se objevilo pravidlo (viz. Obrázek 2: Asociační pravidla 2)

Tučný => žena

Opět nás překvapilo, že pravidlo platí se 100%ní spolehlivostí (viz. elipsa výše). Je zde potvrzeno, že již u žákovských kategorií je procento tuku žen a dívek významně vyšší než u chlapců. A výsledky měření jsou velmi homogenní v závislosti na pohlaví jedince. Pochopitelně je zřejmý tlak trenérů na co nejmenší podíl tuku na hmotnosti jedince, jež představuje zbytečnou zátěž při realizaci sportovního výkonu. Naše výsledky ukazují, že převážná část probandů vykazuje dlouhodobě procento tuku na úrovni: chlapci do 7%, děvčata v intervalu 9,7 – 20,6, což výrazně nižší podíl než u běžné populace.

Dodejme jen, že každý reálný atribut byl rozdělen do pěti ekvifrekvenčních intervalů a na obrázcích výše vidíme, že pro atribut %tuku asociační pravidla hovoří o čtyřech z těchto pěti intervalů (range1, 3, 4, 5). Chybí pravidlo pro interval range2 [6.55 – 9.75] - důvod je takový,

že hodnoty %tuku v tomto intervalu jsou na rozmezí mezi malými a velkými hodnotami, a těmto hodnotám odpovídají jak muži (ti více tuční), tak ženy (ty méně tučné), tudíž nelze s dostatečnou spolehlivostí na tomto intervalu tvrdit něco o závislosti %tuku a jednoho druhu pohlaví (jinak řečeno, v počtu objektů odpovídajících tomuto intervalu výrazně nepřevažují ani muži, ani ženy).

9. Rozhodovací stromy

Pro krasobruslaře je zásadní dostatečná výška výskoku, jelikož ta jim umožní provést všechny požadované akrobatické prvky při jízdě. Zajímalo nás tedy, jestli a jak souvisí hodnota tohoto atributu s ostatními atributy. Abychom se dozvěděli něco více o vztahu výšky skoku k ostatním atributům, použili jsme právě analýzu pomocí rozhodovacích stromů, která je pro tyto účely velmi vhodná

Analýza pomocí rozhodovacích stromů (metoda popsána v [2]):

- byla provedena na základě atributů Pohlaví, Váha, %tuku, SDM, TP, TL, t_opory, h, t15_běh, t15_brus, als_běh, als_brus – jsou to zástupci jednotlivých podskupin atributů a byly vybrány na základě výsledků analýzy korelační matice, na základě naší znalosti reálného charakteru těchto atributů a na základě výsledků experimentů provedených pomocí metody pro tvorbu rozhodovacích stromů
- byly v ní použity vybrané atributy bez předzpracování, pouze byl diskretizován klasifikační atribut d 5 ekvifrekvenčních intervalů
- byly zpracovány i atributy s chybějícími údaji
- jako klasifikační atribut byla vybrána výška skoku h, jelikož hodnota tohoto atributu je pro krasobruslaře zásadní

Pokud jsme však pro tvorbu stromu použili všechny objekty z dat (analyzovali jsme celý vzorek dat najednou), dozvěděli jsme se pouze to, že těžší (a tedy i silnější a zřejmě i starší) jedinci skácou převážně vysoko a lehčí jedinci nízko, což není příliš zajímavý výsledek. Abychom tedy omezili vliv váhy na výšku skoku, rozdělili jsme krasobruslaře do několika skupin. Vzhledem k tomu, že váha u takto mladých jedinců je silně závislá na věku, rozdělili jsme tedy jedince do několika věkových kategorií, které se běžně při tréninku používají.

Použili jsme tedy kategorie:

- 1) do 9 let včetně (20 chlapců, 67 dívek),
- 2) 10 až 12 let (43 chlapců, 118 dívek),
- 3) 13 až 15 let (9 chlapců, 24 dívek),
- 4) 16 let a více (4 chlapců, 3 dívky).

Při takovémto rozdělení se potom objevilo množství vztahů mezi atributy, o kterých nám předchozí analýzy nic neřekly (např. netriviální vztah výšky výskoku h a %tuku, který jaksi tušíme, nicméně zatím se nikde neobjevil (např. korelace těchto dvou atributů je velmi nízká, téměř nulová)). Pozn.: Věkovou kategorii 4) jsme vzhledem malému zastoupení v datech vypustili.

9.1. Věková kategorie do 9 let

Zajímavé je (viz. Obrázek 6: Věková kategorie do 9 let), že kořenový atribut rozhodovacího stromu je pohlaví, který dělí strom na dvě větve odpovídající mužům a ženám (v této věkové kategorii spíše chlapcům a dívkám) – poprvé se tedy ukazuje vztah Pohlaví a výšky skoku.

Pro chlapce je pak rozhodující SDM, který je dělí na lepší a na horší. Zarážející je však to, že ti lepší z těch lepších mají $als_b\ddot{e}h < 3,162$ (tmavá elipsa vlevo). Vypadá to tedy tak, že rychlost běhu po 1 vteřině má na výšku výskoku u chlapců přesně opačný vliv, než bychom čekali - prozatím tuto skutečnost neumíme vysvětlit - snad by bylo možné uvažovat o rozdílném zvládnutí dynamického stereotypu jednorázového skoku (výskoku) a běhu. V dalším sledování se na tuto záležitost zaměříme.

U žen rozhoduje nejdříve váha a potom:

- 1) TP pro ty nejtěžší dívky (tmavá elipsa kreslená přerušovanou čarou),
- 2) %tuku pro méně těžké dívky (světlá elipsa),
- 3) SDM pro nejlehčí dívky (světlá elipsa kreslená přerušovanou čarou) – pro malé SDM skáčou dívky jen nízko.

9.2. Věková kategorie 10 až 12 let

Pro tuto početně nejvíce zastoupenou kategorii jsme vygenerovali poměrně košatý strom, který jsem musel rozdělit na dvě části (viz. Obrázek 7: Věková kategorie 10 až 12 let - větev muži a Obrázek 8: Věková kategorie 10 až 12 let - větev ženy). Kořenový atribut byl Pohlaví, horní obrázek tedy odpovídá chlapcům a obrázek na další stránce odpovídá dívkám.

U chlapců se objevila skupina o počtu 7, která je těžká a SDM má vysoký (tmavá elipsa). Pro lehčí chlapce (pravá větev) se ukazuje, že ti lepší z nich mají vysoké TL (světlá elipsa). Pro ženy v této věkové kategorii lze tvrdit pouze to, že ty těžší (a zřejmě tedy silnější) skáčou nejvýše (elipsa). Zbýlá pravidla (odpovídající jednotlivým listům) mají buď příliš nízkou podporu, nebo příliš nízkou spolehlivost, aby se dalo na jejich základě něco tvrdit.

9.3. Věková kategorie 13 až 15 let

Kořenovým atributem (viz. Obrázek 9: Věková kategorie 13 až 15 let) je opět pohlaví. Chlapci, u nichž je difference mezi trojskokem na pravé a levé noze (atributy TP, TL) minimální, dosahují i lepších výsledků ve vertikálním výskoku. Lze tedy usoudit, že pro silový impuls svalstva dolních končetin ve vertikálním směru je také důležitá vyrovnanost odrazových parametrů obou nohou. Zaznamenali jsme i výsledky z nichž plyne, že pokud je rozdíl mezi trojskoky na pravé a levé noze větší než 50 cm (výrazná dominance „odrazové“ nohy), projevuje se to negativně na výsledku vertikálního výskoku.

U dívek byl kořenovým atributem trojskok na pravé noze. Pokud navíc ve skupině těch lepších ($TP > 609,5$) mají děvčata nižší váhu než 48,5 kg a nízké procento tuku ($\% \text{ tuku} \leq 11,05$), pak dosahují nejlepších výsledků ve vertikálním výskoku (viz. žlutá elipsa).

Závěr

Sledování ukazatelů výbušné síly dolních končetin může napomoci trenérům při práci s pro krasobruslení talentovanou mládeží (viz. testování pro ČKS - projekt talentované mládeže). Dále má umožnit hledání efektivní kombinace struktury pohybových předpokladů pro co nejlepší provedení sportovního výkonu a v neposlední řadě může napovědět, které z tréninkových prostředků vyvolávají žádané adaptační změny a kdy lze očekávat jejich kladný efekt na sportovní výkon jedince. Naše analýzy naznačují, že záměry konstrukce baterie testů byly splněny. Tato baterie reprezentuje dynamicko-silové vlastnosti měřených jedinců a v této sadě sledovaných proměnných bude vždy nějaká skupina atributů, jež se vztahují k: popisu pohybu probanda v horizontálním směru (lokomoce na čas), popisu pohybu

ve vertikálním směru (opakované výskoky) a skoků na vzdálenost, jež nám reprezentují schopnosti vyvinout značný silový impuls odrazem snožmo i opakovaně z jedné nohy.

Analýza hlavních komponent naznačuje, že 85% rozptylu výsledků lze vysvětlit pomocí 4 hlavních komponent. První je logicky komponenta vztahující se k silovému impulsu nespecifického odrazu ve skoku i lokomoci. Druhá komponenta zahrnuje kromě již zmíněných i dobu opory při odrazu. Tyto dvě se podílí ze 70% na rozptylu výsledků kde relativně nezávislý atribut doby opory má dle našich výsledků významný podíl na vysvětlení celkového rozptylu. Další komponenty zahrnují vliv pohlaví, i když v námi sledované skupině mladých krasobruslařů je relativně malý a je zde naznačen negativní vliv vyššího procenta tuku u dívek.

Výsledky analýzy pomocí metody asociačních pravidel překvapivě ukazují na přímou závislost % tukové tkáně na pohlaví i u prepubescentů.

Při analýze metodou rozhodovacích stromů jsme se pokusili eliminovat vliv věku probandů na výšku výskoku rozdělením výsledků dle věkových kategorií. Použitá analýza naznačila odlišné vztahy výšky výskoku a ostatních sledovaných proměnných v závislosti na věkové kategorii. U všech věkových kategorií je však vidět stejné základní rozdělení do kategorií dle výšky výskoku podle vždy stejného kořenového atributu pohlaví. Nejhrubší rozdělení probandů – vzhledem k přesnosti klasifikace (viz. [2]) na listech stromu – je tedy provedeno pomocí atributu pohlaví, které je potom dále zpřesňováno rozdělením podle dalších atributů (viz. výše). Ukázal se tedy jasný (i když netriviální a na věku závislý) vztah mezi atributy pohlaví a výška výskoku.

Tento článek byl napsán v návaznosti na projekt SGS/24/2010 - Využití BI a BPM pro podporu efektivního řízení.

Literatura:

- [1] AGRAWAL, R., IMIELINSKI, T., SWAMI, A. *Mining Association Rules between Sets of Items in Large Databases*. In: Proceedings of the 1993 ACM SIGMOD Conference on management of data, Washington DC, USA, May 1993, s. 207-216.
- [2] BERKA, P. *Dobývání znalostí z databází*. Academia Praha. ISBN 80-200-1062-9.
- [3] BOSCO, C., LUHTANEN, P., KOMI, PV. *A simple method for measurement of mechanical power in jumping*. Eur. J. Appl. Physiol. 1983, vol. 50, P. 273 –282.
- [4] Bunc, V. *Biokybernetický přístup k hodnocení reakce organismu na tělesné zatížení*. Praha, Univerzita Karlova 1990.
- [5] DOVALIL, J. ET AL. *Výkon a trénink ve sportu*. Praha, Olympia 2002.
- [6] GROSSER, M., STARISCHKA, S. *Konditions-tests. Theorie und Praxis aller Sportarten*. BLV sportwissen, München 1981.
- [7] HAMAR, D., TSCHAN, H. *Výskokový ergometer – princíp a možnosti uplatnenia v diagnostike trénovanosti a tréningovej praxi*. Bratislava, FTVŠ UK 1991.
- [8] HAVLÍČKOVÁ, L. A KOL. *Fyziologie tělesné zátěže I*. Karolinum, Praha 1994, s. 92 – 99.
- [9] KOVÁŘ, R. A KOL. *Manuál pro hodnocení úrovně základní motorické výkonnosti a vybraných charakteristik tělesné stavby školních dětí a mládeže ve věku od 6 do 20 roků*. Těl. Vých. Mlád. 59, 1993, s. 3-63.
- [10] QUINLAN, J. R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [11] RapidMiner: User Manual [on-line]. URL:
http://ovh.dl.sourceforge.net/project/rapidminer/1.%20RapidMiner/5.0/rapidminer-5.0-manual-english_v1.0.pdf
- [12] ŠARMANOVÁ, J. *Datové sklady a dolování znalostí z nich*, učební text, Ostrava 2003. ISBN 80-248-0302-X.

Klasifikace JEL: C38, Z00

PaedDr. Jiří Helešic, Ph.D.

Vedoucí katedry tělovýchovných a sportovních aktivit
Obchodně-podnikatelská fakulta v Karviné
Slezská univerzita
Univerzitní náměstí 1943, Karviná
helesic@opf.slu.cz

Ing. Jan Górecki

Odborný asistent Katedry informatiky
Obchodně-podnikatelská fakulta v Karviné
Slezská univerzita
Univerzitní náměstí 1943, Karviná
gorecki@opf.slu.cz

Obrázek 3: Korelační matice

Correlation Matrix (Correlation Matrix)																				
<input checked="" type="radio"/> Table View <input type="radio"/> Pairwise Table <input type="radio"/> Plot View																				
Attributes	Pohlaví	Věk	Výška	Váha	%tuku	SDM	TP	TL	t_opory	P (W/kg)	h	h/tc	t15_běh	a1s_běh	a2s_běh	a3s_běh	t15_brus	a1s_brus	a2s_brus	a3s_brus
Pohlaví	1	-0.110	-0.121	-0.140	0.802	-0.148	-0.138	-0.167	-0.044	0.055	0.013	0.037	0.014	-0.007	-0.011	-0.021	0.127	-0.055	-0.112	-0.148
Věk	-0.110	1	0.808	0.822	-0.130	0.744	0.783	0.757	0.050	0.503	0.642	0.475	-0.295	0.192	0.272	0.343	-0.347	0.269	0.362	0.380
Výška	-0.121	0.808	1	0.915	-0.103	0.745	0.774	0.758	0.045	0.431	0.556	0.415	-0.335	0.210	0.291	0.380	-0.384	0.318	0.409	0.431
Váha	-0.140	0.822	0.915	1	-0.033	0.708	0.722	0.698	0.080	0.386	0.522	0.366	-0.291	0.196	0.256	0.344	-0.357	0.298	0.377	0.410
%tuku	0.802	-0.130	-0.103	-0.033	1	-0.273	-0.243	-0.274	0.030	-0.120	-0.153	-0.129	0.059	-0.025	-0.055	-0.057	0.093	0.008	-0.051	-0.066
SDM	-0.148	0.744	0.745	0.708	-0.273	1	0.904	0.905	0.015	0.601	0.748	0.575	-0.467	0.300	0.432	0.499	-0.475	0.335	0.437	0.471
TP	-0.138	0.783	0.774	0.722	-0.243	0.904	1	0.956	-0.003	0.674	0.812	0.653	-0.490	0.305	0.457	0.535	-0.498	0.352	0.468	0.501
TL	-0.167	0.757	0.758	0.698	-0.274	0.905	0.956	1	-0.008	0.663	0.801	0.639	-0.471	0.291	0.435	0.513	-0.480	0.338	0.460	0.493
t_opory	-0.044	0.050	0.045	0.080	0.030	0.015	-0.003	-0.008	1	-0.441	-0.117	-0.543	-0.106	0.099	0.069	0.083	-0.155	0.139	0.143	0.154
P (W/kg)	0.055	0.503	0.431	0.386	-0.120	0.601	0.674	0.663	-0.441	1	0.877	0.949	-0.243	0.134	0.242	0.303	-0.164	0.109	0.169	0.170
h	0.013	0.642	0.556	0.522	-0.153	0.748	0.812	0.801	-0.117	0.877	1	0.850	-0.373	0.242	0.368	0.437	-0.307	0.234	0.312	0.322
h/tc	0.037	0.475	0.415	0.366	-0.129	0.575	0.653	0.639	-0.543	0.949	0.850	1	-0.266	-0.168	-0.272	-0.324	-0.179	-0.119	-0.188	-0.186
t15_běh	0.014	-0.295	-0.335	-0.291	0.059	-0.467	-0.490	-0.471	-0.106	-0.243	-0.373	-0.265	1	-0.793	-0.906	-0.912	0.861	-0.674	-0.765	-0.784
a1s_běh	-0.007	0.192	0.210	0.196	-0.025	0.300	0.305	0.291	0.099	0.134	0.242	0.158	-0.793	1	0.832	0.747	-0.655	0.646	0.676	0.666
a2s_běh	-0.011	0.272	0.291	0.256	-0.055	0.432	0.457	0.435	0.069	0.242	0.368	0.272	-0.906	0.832	1	0.949	-0.808	0.722	0.821	0.833
a3s_běh	-0.021	0.343	0.380	0.344	-0.057	0.499	0.535	0.513	0.083	0.303	0.437	0.324	-0.912	0.747	0.949	1	-0.829	0.714	0.827	0.846
t15_brus	0.127	-0.347	-0.384	-0.357	0.093	-0.475	-0.498	-0.480	-0.155	-0.164	-0.307	-0.179	0.861	-0.655	-0.808	-0.829	1	-0.850	-0.896	-0.911
a1s_brus	-0.055	0.269	0.318	0.298	0.008	0.335	0.352	0.338	0.139	0.109	0.234	0.119	-0.674	0.646	0.722	0.714	-0.850	1	0.881	0.837
a2s_brus	-0.112	0.362	0.409	0.377	-0.051	0.437	0.468	0.460	0.143	0.169	0.312	0.188	-0.765	0.676	0.821	0.827	-0.896	0.881	1	0.965
a3s_brus	-0.148	0.380	0.431	0.410	-0.066	0.471	0.501	0.493	0.154	0.170	0.322	0.186	-0.784	0.666	0.833	0.846	-0.911	0.837	0.965	1

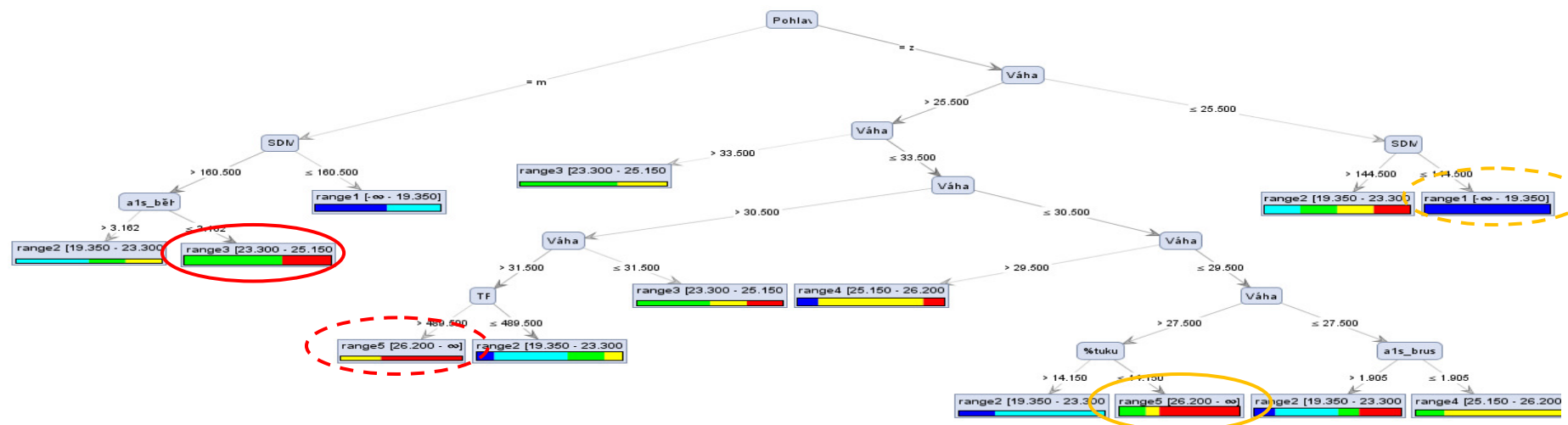
Obrázek 4: Vlastní čísla

<input checked="" type="radio"/> Eigenvalues <input type="radio"/> Eigenvectors <input type="radio"/> Cumulative Variance Plot			
Component	Standard Deviation	Proportion of Variance	Cumulative Variance
PC 1	3.101	0.481	0.481
PC 2	1.962	0.192	0.673
PC 3	1.410	0.099	0.773
PC 4	1.277	0.082	0.854
PC 5	0.846	0.036	0.890
PC 6	0.751	0.028	0.918
PC 7	0.580	0.017	0.935
PC 8	0.492	0.012	0.947
PC 9	0.458	0.010	0.958
PC 10	0.411	0.008	0.966
PC 11	0.395	0.008	0.974
PC 12	0.343	0.006	0.980
PC 13	0.329	0.005	0.985
PC 14	0.250	0.003	0.988
PC 15	0.236	0.003	0.991
PC 16	0.214	0.002	0.993
PC 17	0.202	0.002	0.995
PC 18	0.192	0.002	0.997
PC 19	0.184	0.002	0.999
PC 20	0.152	0.001	1.000

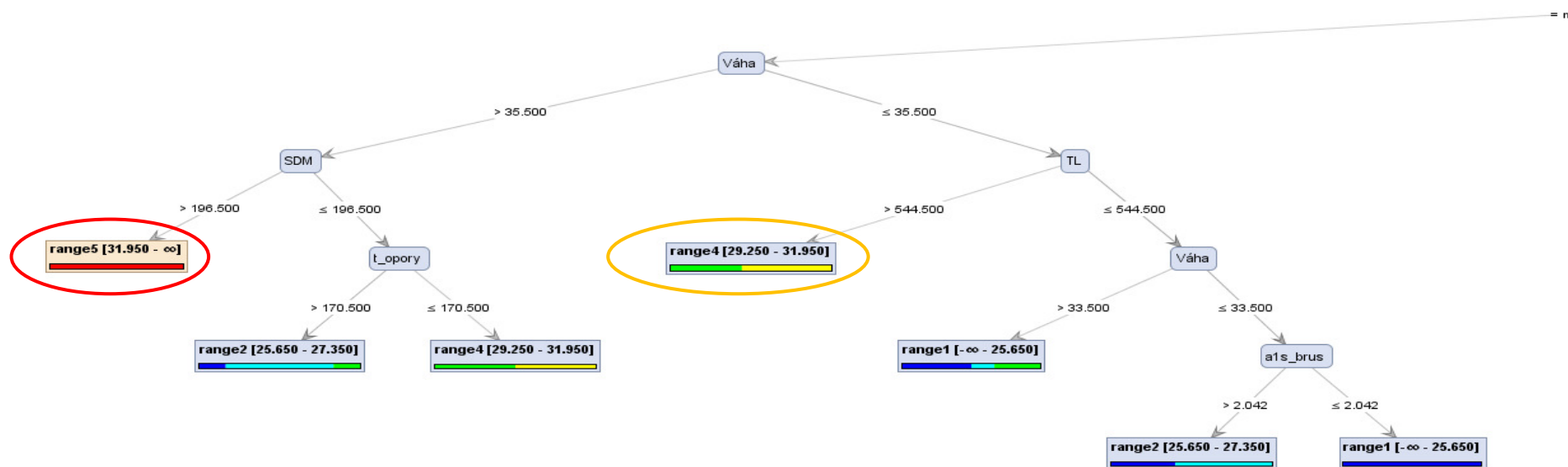
Obrázek 5: Vlastní vektory

<input type="radio"/> Eigenvalues <input checked="" type="radio"/> Eigenvectors <input type="radio"/> Cumulative Variance Plot				
Attribute	PC 1	PC 2	PC 3	PC 4
Pohlaví	-0.050	0.048	-0.520	0.452
Věk	0.228	-0.214	0.161	0.235
Výška	0.232	-0.178	0.192	0.267
Váha	0.222	-0.172	0.220	0.310
%tuku	-0.051	0.089	-0.428	0.549
SDM	0.262	-0.199	0.115	0.862
TP	0.277	-0.210	0.063	0.046
TL	0.271	-0.211	0.089	0.025
t_opory	0.012	0.194	0.364	0.404
P (W/kg)	0.174	-0.320	-0.311	-0.167
h	0.228	-0.276	-0.169	-0.012
h/tc	0.174	-0.306	-0.336	-0.225
t15_běh	-0.257	-0.223	0.082	0.047
a1s_běh	0.213	0.244	-0.093	-0.063
a2s_běh	0.259	0.237	-0.116	-0.084
a3s_běh	0.272	0.202	-0.096	-0.047
t15_brus	-0.260	-0.240	-0.005	0.063
a1s_brus	0.225	0.262	-0.037	-0.024
a2s_brus	0.260	0.243	-0.010	-0.041
a3s_brus	0.265	0.233	0.015	-0.039

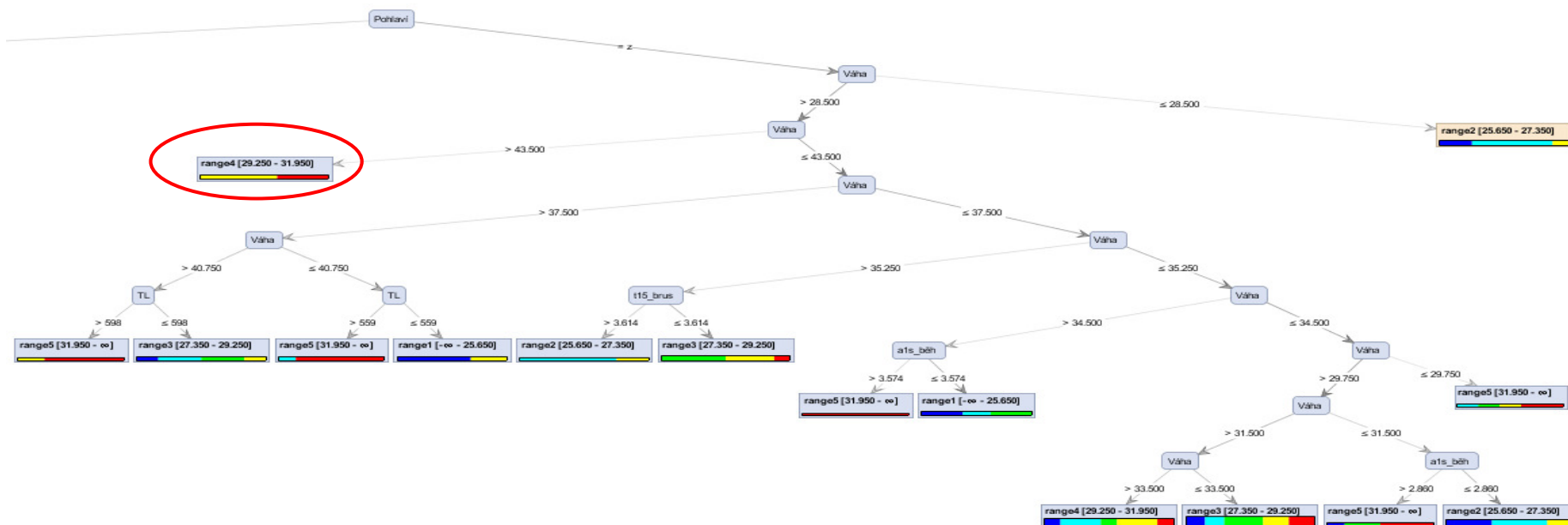
Obrázek 6: Věková kategorie do 9 let



Obrázek 7: Věková kategorie 10 až 12 let - větev muži



Obrázek 8: Věková kategorie 10 až 12 let - větev ženy



Obrázek 9: Věková kategorie 13 až 15 let

