

Popisné modely špecifických skupín pacientov vzhľadom na ich kardiovaskulárne riziko

Ing. Adam Pavlíček, školiteľ: prof. Ing. Ján Paralič, PhD., konzultanti: Ing. Zuzana Pella, PhD., Ing. Oliver Lohaj



CHI
Centrum hospodárskej
informatiky



Fakulta elektrotechniky
a informatiky

1. Ciele práce

Táto diplomová práca sa zaoberá problematikou **aplikovania popisných modelov** na špecifickú skupinu pacientov Východoslovenskú ústavu srdcových a cievnych chorôb. Pri predkladanej diplomovej práci možno definovať hlavný cieľ ako návrh, implementáciu a následne overenie aplikácie pre podporu výskumu kardiológov na báze popisných modelov.

2. Charakteristika aplikácie

Pri realizácii, resp. implementácii softvérového riešenia bol využitý **programovací jazyk R** a taktiež knižnica **R Shiny**, ktorá je vhodná pre tvorbu webových aplikácií. Samotný softvér pozostáva zo **štyroch hlavných častí**, ktorých význam bližšie charakterizujeme v nasledujúcich podkapitolách.

2.1. Načítanie dát

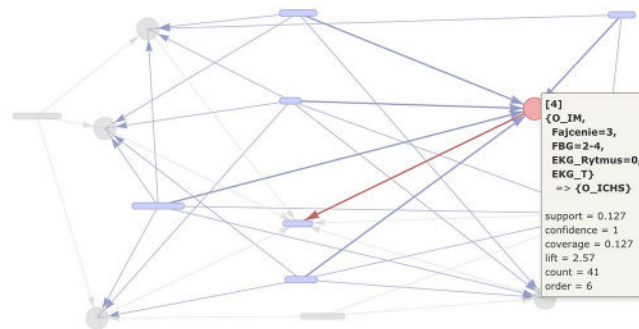
Aplikácia disponuje funkcionalitou, vďaka ktorej má používateľ možnosť načítať a následne pracovať so svojou **vlastnou dátovou množinou**. Nie je tak limitovaný len na prácu s dátami o špecifickej skupine pacientov daného ústavu. Pri načítaní svojej vlastnej dátovej množiny má používateľ taktiež možnosť voľby spôsobu **úpravy numerických atribútov na kategorický tvar**, a to pomocou diskretizácie alebo priameho pretypovania.

2.2. Hľadanie asociačných pravidiel

Pre hľadanie asociačných pravidiel sme využili algoritmus **Apriori**. Používateľ má možnosť prispôbiť obsah vstupnej dátovej množiny, ale aj definovať hraničné hodnoty vstupných parametrov daného modelu, akými sú minimálna **podpora**, **spôľahlivosť** či **dĺžka ľavej strany pravidiel**. Taktiež vie obmedziť výskyt zvolených atribútov na pravej strane asociačných pravidiel.

Medzi výstupy je možno zaradiť nájdené **asociačné pravidlá prezentované vo forme tabuľky** s možnosťou filtrovania či editácie daných záznamov, ale taktiež **interaktívne vizualizácie**, vďaka ktorým je možné viac a

lepšie pochopiť obsah získaných výsledkov. Konkrétne je možné hovoriť o bodovom a sieťovom grafe, ktoré sú používateľovi k dispozícii.



Obrázok 1: Ukážka vizualizácie vzťahov medzi hodnotami atribútov a asociačnými pravidlami

2.3. Aplikovanie metód zhlukovania

Pri zhlukovaní má používateľ na výber medzi algoritmi **PAM**, **k-Means** a taktiež **hierarchickým zhlukovaním**. Pre výpočet vzdialenosti medzi jednotlivými záznamami sme využili tzv. **Gowerovu vzdialenosť**, avšak následne bola pridaná aj možnosť výberu **Euklidovskej** či **Manhattanskej vzdialenosti**.

Rovnako ako v predchádzajúcom prípade, používateľ si môže prispôbiť obsah vstupnej dátovej množiny či vykonať analýzy pre odhad optimálneho počtu zhlukov, a to pomocou tzv. **metódy siluety** alebo tzv. **metódy lakt'a**. Výstupom je k dispozícii opäť niekoľko. Konkrétne sa jedná o výsledky jednotlivých **validačných kritérií**, vizualizáciu rozloženia dátových bodov v jednotlivých zhlukoch, výslednú dátovú množinu s informáciou o tom, ktorý záznam zo vstupnej dátovej množiny do akého zhluku patrí, či informáciu o tom, ktoré záznamy sú si v jednotlivých zhlukoch **najviac**, resp. **najmenej podobné**.

2.4. Asociačné pravidlá a zhlukovanie

V prípade, že je používateľ spokojný s výsledkami zhlukovania, má možnosť pre ním zvolené zhluky nájsť

asociačné pravidlá, ktoré odpovedajú zvoleným hodnotám daných metrik, a teda minimálnej **podpore**, **spôľahlivosti** či **dĺžke ľavej strany pravidiel**. V tejto časti sme opäť využil algoritmus **Apriori**, pričom výstupy sú identické s tými, ktoré sme definovali v časti **2.2**.

3. Používateľská štúdia

Používateľskej štúdie sa zúčastnilo **sedem respondentov**, ktorými boli lekári pôsobiaci na rôznych oddeleniach Univerzitnej nemocnice Louisa Pasteura v Košiciach a Východoslovenského ústavu srdcových a cievnych chorôb.

Samotné testovanie prebiehalo v dvoch fázach. V prvej z nich testujúci **vykonali úlohy** pokrývajúce funkcionalitu hlavných častí implementovaného softvéru. Scenáre jednotlivých úloh boli navrhnuté v spolupráci s doktorom špecializovaného odboru kardiológie. V druhej fáze respondenti **ohodnotili použiteľnosť aplikácie**, a to na základe metódy, ktorá je známa pod názvom **System usability scale**. Tá pozostáva z desiatich preddefinovaných otázok, pričom pri výbere odpovedí, resp. hodnotenia je aplikovaná **Likertova škála**. Priemerné skóre, ktoré bolo vyjadrené na základe definovaného postupu a pravidiel pre vyhodnotenie metódy **System usability scale** je **74.64**, čo taktiež možno interpretovať ako hodnotenie známku **B**. Na základe výberu odpovedí pri jednotlivých otázkach je taktiež možné konštatovať, že:

- všetci respondenti úplne súhlasia, resp. čiastočne súhlasia s tým, že daný systém **je konzistentný**,
- šesť respondentov úplne súhlasí, resp. čiastočne súhlasí s tým, že mnoho ľudí by sa aplikáciu **naučilo používať veľmi rýchlo**,
- päť respondentov úplne súhlasí, resp. čiastočne súhlasí s tým, že tento systém bol **jednoduchý na použitie**.

4. Odkaz na prácu

Pavlíček, A.: Popisné modely špecifických skupín pacientov vzhľadom na ich kardiovaskulárne riziko, diplomová práca, Technická univerzita v Košiciach, 2022, Košice, Slovenská republika