PRÍPADOVÉ USUDZOVANIE PRI DIAGNOSTIKE KARDIOVASKULÁRNYCH OCHORENÍ

Zuzana Tocimáková, školiteľ: prof. Ing. Ján Paralič, PhD., konzultant: Ing. Ľudmila Pusztová

CHI Centrum hospodárskej informatiky

Fakulta elektrotechniky a informatiky
TECHNICKÁ UNIVERZITA V KOŠICIACH

Ciele práce:

Navrhnúť, implementovať a otestovať systém na podporu rozhodovania s prvkami prípadového usudzovania pre proces diagnostiky kardiovaskulárnych ochorení.

Primárnou úlohou navrhovaného systému je poskytnúť jeho používateľovi, teda kardiológovi dáta o najpodobnejších známych prípadoch k aktuálnemu pacientovi a na základe týchto dát mu navrhnúť, či je u daného pacienta vysoké riziko výskytu zúženia koronárnych ciev.

Tieto informácie môže následne kardiológ využiť v procese rozhodovania o tom, či je nevyhnutné poslať tohto pacienta na koronarografické vyšetrenie.

Z pohľadu dolovania v dátach je cieľom vytvoriť klasifikátor na báze algoritmu k-najbližších susedov, ktorý by podľa hodnoty cieľovej triedy najbližších prípadov nachádzajúcich sa v okolí zadaného záznamu klasifikoval do ktorej cieľovej skupiny tento záznam patrí, ale zároveň aj zobrazil nájdené najbližšie prípady.

Dataset:

820 záznamov o pacientoch z VÚSCH, a.s. obsahujúcich 66 atribútov ako:

- vek, pohlavie, krvný tlak, výška, váha, fajčenie, alkohol, diagnózy pacienta aj rodinných príslušníkov, EKG, ...
- kategorický cieľový atribút Nález (0 až 5) – reprezentujúci percentuálne zúženie jednotlivých vetiev koronárnych artérií

Metódy riešenia:

Klasifikačný algoritmus kontrolovaného strojového učenia: *k-najbližších susedov* (k-NN), ktorého podstatou je predpoklad, že záznamy, ktoré sa podľa vybranej metriky pre výpočet vzdialenosti nachádzajú blízko seba, si sú navzájom podobné.

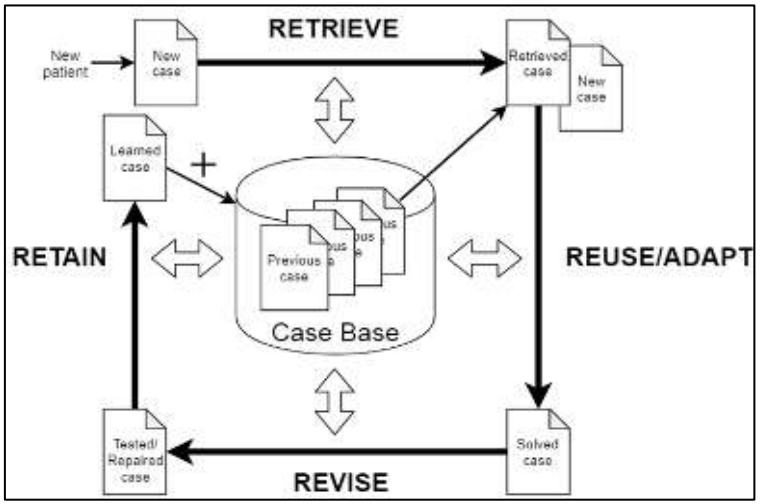
Jeho modifikovaná implementácia ponúka nasledovné možnosti použitia:

- Rôzne spôsoby výpočtu vzdialenosti:
  - Euklidovská vzdialenosť
  - Manhattanská vzdialenosť
  - Mahalanobisova vzdialenosť
  - Maxmin
  - WRF (Weight Ratio Functionality)
- Filtrácia záznamov podľa príkladov z rovnakého uzla/zhluku:
  - Rozhodovací strom
  - Zhlukovanie
- Voľba počtu najbližších susedov – parameter *k*:
  - 1 až 10

Prípadové usudzovanie (CBR):

Využíva princíp analógie. Imituje človeka pri riešení nových problémov, ktorý na to využíva svoje doterajšie skúsenosti a snaží sa ich prispôsobiť aktuálnej situácii. CBR cyklus:

1. **RETRIEVE** - nájdenie najpodobnejších prípadov k novému prípadu
2. **REUSE/ADAPT** - opätovné použitie znalostí z nájdených prípadov na vyriešenie nového prípadu
3. **REVISE** - kontrola navrhovaného riešenia
4. **RETAIN** - uloženie častí riešenia, ktoré môžu byť užitočné v budúcnosti



Experimentálne testovanie:

- Využitie 10-násobnej krížovej validácie.
- Na podmnožine 14 atribútov vybraných pomocou Forward Stepwise Selection.

Metóda výpočtu vzdialenosti	6 cieľových tried		3 cieľové triedy		2 cieľové triedy	
	Najvyššia priemerná presnosť	K susedov	Najvyššia priemerná presnosť	K susedov	Najvyššia priemerná presnosť	K susedov
Euklidovská	32.56%	3	41.31%	1	56.83%	5
Mahalanobisova	**42.20%**	10	50.12%	10	66.95%	10
Maxmin	40.49%	9	**50.73%**	9	**69.27%**	10

Publikácia:

Z. Tocimáková, Ľ. Pusztová, J. Paralič, D. Pella: „Case-Based Reasoning for Support of the Diagnostics of Cardiovascular Diseases“, Studies in Health Technology and Informatics, Digital Personalized Health and Medicine, 2020, s. 537 - 541