

Aplikácia na vyhodnocovanie dotazníkov pomocou hĺbkovej analýzy dát

Žilinská univerzita v Žiline

Fakulta riadenia informatiky Katedra informatiky

Autor: Ján Rabčan

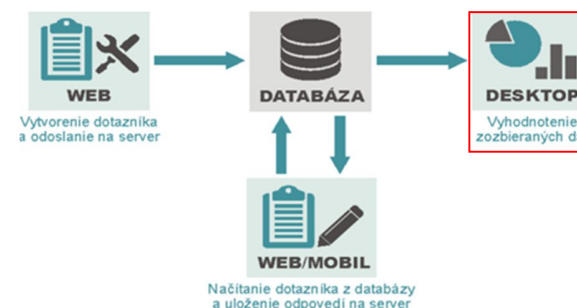
Vedúci: Jozef Kostolný

Motivácia

V práci sa zaoberám tvorbou a návrhom aplikácie pre vyhodnocovanie dotazníkov definovaných pomocou fuzzy premenných. Hlavnou motiváciou bolo vytvoriť systém, ktorý dokáže efektívne klasifikovať nové inštancie dát. Motivácia pre vývoj aplikácie vznikla pri určovaní psychologickéj charakteristiky osoby na základe testov zameraných na podpis osoby a psychologického testu. Cieľom bolo vytvoriť systém, ktorý umožní získať psychologický charakter osoby priamo z podpisu. Celé to fungovalo spôsobom, že užívateľ vyplní množinu testov, z ktorých sa neskôr buduje rozhodovací model. Na základe vzniknutého modelu bude možné klasifikovať nových užívateľov.

Zber dát

Zber dát bol zabezpečený aplikáciami, ktoré vznikli počas inžinierskeho projektu. Nosná časť aplikácie bola tvorba aplikácie, ktorá nazbierané dáta vyhodnotí.

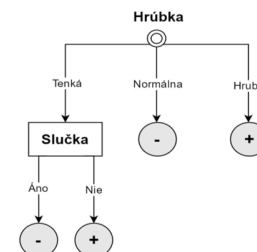


Klasifikačná úloha

Cieľom tejto problematiky je zaradiť nové pozorovanie do konkrétnej podmnožiny, respektíve triedy. Pozorovania predstavujú inštancie (mnohokrát označované ako príklady alebo vzorky) popísané pomocou atribútov. Klasifikačné triedy je možné definovať pomocou numerických atribútov, alebo ich môžeme zadať v slovnej podobe, čiže lingvisticky. Inštancie dát dátových množín v práci sú popísané množinou vstupných atribútov a jedným výstupným atribútom. Inštancie v množine musia mať rovnakú štruktúru, to znamená, že všetky inštancie trénovacej množiny sú popísané rovnakými atribútmi, aj rovnakým počtom atribútov. Cieľom algoritmov je klasifikovať inštanciu do tried popísaných výstupným atribútom na základe hodnôt vstupných atribútov. Tieto algoritmy fungujú v dvoch krokoch. V prvom kroku musia vybudovať klasifikačný mechanizmus, respektíve model. Na budovanie využijú trénovú množinu inštancií. V druhom kroku využívajú vybudovaný model na klasifikovanie nových inštancií. Počas klasifikácie sa model pokúša vyjadriť hodnotu výstupného atribútu klasifikovanej inštancie.

Rozhodovacie stromy

Počas tvorby rozhodovacieho stromu je potrebné zvoliť správny asociačný atribút pre každý vrchol rozhodovacieho stromu a taktiež je potrebné vykonať orezanie (tree-pruning). Orezávanie je možné vykonať dvoma spôsobmi, a to počas tvorby rozhodovacieho stromu, alebo až po jeho vybudovaní. Algoritmy v práci vykonávali orezávanie počas budovania stromov a ako kritéria pre zastavenie rastu stromu v danom vrchole sú minimálna frekvencia výskytov a maximálna úroveň vierohodnosti prijatého rozhodnutia. Algoritmy rozhodovacích stromov využívajú rôzne techniky výberu asociačných atribútov. Všetky algoritmy v práci vyberali algoritmy na základe informačných charakteristík.



Možnosť použitia

Výsledná aplikácia, ktorá počas tvorby práce vznikla nie je obmedzená len na psychologické a podpisové testy. Pomocou aplikácie je možné budovať klasifikačné modely z rôznych dátových množín. Aplikácia taktiež umožňuje rôzne operácie s dátovými množinami.