

Rozhodovací stromy pro klasifikaci dat

V předložené diplomové práci se autor zabývá problematikou klasifikace pomocí rozhodovacích stromů, konkrétně autor studuje vlastnosti metod C4.5, CART a SDT. Kromě teoretického popisu zmíněných metod je součástí diplomové práce také implementace těchto metod v jazyce Java.

Text práce je členěn do 6 kapitol: po úvodní kapitole autor zavádí v Kapitole 2 nutný formalismus. V Kapitole 3 popisuje autor modely C4.5 a CART, včetně různých metod ořezávání stromu. Obsahem Kapitoly 4 je popis klasifikačních stromů SDT, které využívají teorie fuzzy množin. V páté kapitole jsou popsány experimenty, které autor provedl na dvou datových množinách, a jejich výsledky. Také jsou zde popsány teoretické výsledky, kterých autor dosáhl. Kapitola 6 shrnuje a uzavírá text práce.

Z textu práce je zřejmé, že autor se velmi dobře seznámil se studovanou problematikou a v oblasti klasifikace pomocí rozhodovacích stromů se dobře orientuje, což potvrzuje i výběr relevantní literatury. Formální a jazyková úroveň práce je dobrá a práce neobsahuje mnoho překlepů a gramatických chyb. Na druhou stranu v teoretickém popisu jsou místy drobné nepřesnosti (např. definování vstupních a výstupních uzlů v neorientovaném grafu v Def. 2.2.1). Autor také občas předbíhá a používá intuitivně nějaký pojem, který korektně definuje až později – to může mít vliv na srozumitelnost textu. Při popisu experimentů autor používá místy nejednotné značení; vždy je však zřejmé, co má na mysli.

Pro softwarovou implementaci popsaných algoritmů (která je obsahem příloženého CD) použil autor dobré podpůrné knihovny. Vlastní kód je kvalitní a relativně dobře zdokumentovaný. Autor používá správným způsobem principy OOP. Oceňuji také integraci algoritmů do dataminingového frameworku Weka. Negativně bych hodnotil skutečnost, že některé komentáře jsou v češtině, navíc u některých tříd a rozhraní dokumentace zcela chybí a uvnitř metod by mohlo být také více komentářů.

Autor prokázal schopnost nejenom nastudovat teoretický popis algoritmů, ale také kvalitním způsobem algoritmy implementovat. Velmi pozitivně hodnotím skutečnost, že práce obsahuje kromě experimentálního porovnání studovaných metod také výsledky v oblasti teorie (především důkaz toho, že v SDT stromech se běžně používá pro optimalizaci stromu Levenberg-Marquardtova metoda nekorektním způsobem).

I přes uvedené nedostatky hodnotím celkově práci **velmi kladně**. Autor zadání práce **jednoznačně splnil**, a proto doporučuji práci **uznat** jako diplomovou.

V Praze, dne 17. 5. 2008

ing. David Štefka
Ústav informatiky AV ČR
stefka@insophy.cz