

Oponentský posudek disertační práce

Název práce: Using fuzzy integral as an Aggregation Operator in Dynamic Classifier Systems

Autor práce: Ing. David Štefka, FJFI ČVUT Praha

Oponent: Doc. RNDr. Martin Štěpnička, Ph.D., ÚVAFM OU Ostrava

Posuzovaná práce se zabývá velmi aktuálním tématem kombinačních metod a technik v oblasti klasifikací. Důležitým a pozitivně hodnoceným aspektem posuzované práce je fakt, že v ní nebyla studována jedna či dvě partikulární otázky z dané oblasti, ale že se autor pokusil analyzovat přístup ke kombinačním technikám velmi poctivě od začátku až do konce celého procesu výstavby kombinačních technik. Obhajovaná práce tak poskytuje kompletní analýzu pro mnoho dalších autorů, kteří mohou v poměrně přehledně napsané práci vždy nalézt konkrétní požadovanou část výzkumu a konzultovat ji se svým vlastním výzkumem.

Práce pochopitelně má své nedostatky. Z hlediska studovaného oboru bych preferoval práci více formálního matematického rázu. Konkrétně mám například na mysli zavedení mnoha pojmů pouze v textu popř. za pomoci matematického vztahu, nicméně bez formální definice, což není ku prospěchu jasnosti, jednoznačnosti a čitelnosti textu. Tento nešvar se vyskytuje především v kapitolách 3 a 4. Názorným příkladem je strana 57, kde je nejprve zaveden symbol $U(s)$ značící podmnožinu validační množiny $\mathcal{V}$ závislou na parametru s . Tento symbol není zaveden pomocí formální definice a název množiny 'restricted validation set' je uveden pouze velmi nenápadně ve volném textu níže. Dále je níže zavedena míra 'Similarity to the Oracle Confidence Measure' $\mathcal{F}^{SOR}$ opět bez formální definice pouze pomocí volného textu a matematického vztahu (4.4). Ovšem vztah (4.4) používá symbol $\mathcal{V}$ nikoliv ve významu celé validační množiny, ale pouze ve významu zúžené 'restricted validation set'. Nicméně tento fakt se čtenář dozví opět pouze z volného textu. Formální zavedení všech pojmů a striktní dodržování značení by předložené disertační práci v těchto případech kvalitativně pomohlo.

Nastavení experimentu v kapitole 4.4 lze vytknout fakt, že u malých datových množin bylo zvoleno nastavení $\mathcal{V}^1 = \mathcal{V}^2$, tedy v podstatě shodnost validační a testovací množiny. Výsledky jsou v takovém případě neprůkazné a vhodnější by bylo postup ověřovat pouze na dostatečně velkých datových množinách, u kterých taková volba není nutná.

Poněkud nejasným místem je Definice 14 v podkapitole 5.1. V této definici autor vychází ze třech prostorů typu $([0, 1], \Delta)$, kde Δ značí t-konormu, přičemž není jasné, o jaké prostory se má z matematického hlediska jednat. Algebraicky se jedná pouze o komutativní monoidy. Použití takových struktur se jeví redundantní a multiplikativní operace mohla být zdefinována pouze jako binární operace na $[0, 1]$ splňující podmínky 3 a 4. Definice dále obsahuje podmínku (č. 2), která zamezuje použití t-norem s děliteli nuly jakožto multiplikativních operací v definovaném systému. Ačkoliv se jedná o definici přejatou a nikoliv autorovu původní, měla by práce obsahovat zdůvodnění, proč bylo této definice využito, a to obzvláště vzhledem k tomu, že např. pro zobecněné Sugenyovy integrály se běžně používá t-norem s děliteli nuly.

Práce je psaná velmi dobrou angličtinou a výskyt chyb či překlepů je minimalizován. Seznam nepodstatných technických a formálních chyb či drobných

prohřešků proti zvyklostem uvádím níže. Jedná se o chyby kosmetického rázu, na které není třeba při obhajobě jednotlivě reagovat.

Připomínky technické a formální:

- Na straně 15 je 'easy to hande' zřejmě místo 'easy to handle'.
- Podkapitola 2.2.6. nese nepatrně zavádějící název **Fuzzy Logic Methods**, její obsah nepojednává o metodách založených na fuzzy logice v užším slova smyslu.
- Z hlediska zvyklostí i lepší čitelnosti by bylo vhodné sázet definice základním písmem a definovaný pojem kurzívou, nikoliv naopak.
- V Definici 2 jsou použity různé symboly γ_i a μ_i pro totéž.
- Obrázek 3.1. používá symbol τ místo zavedeného symbolu Γ .
- Vztahy (3.8), resp. (3.9), jsou totožné se vztahy (3.1), resp. (3.2). Vztahy se číslují právě proto, aby se na ně dalo odkázat a nemusely se opakovat.
- Při definování 'Two-best Margin Confidence Measure' $\kappa_{\phi}^{(TBM)}$ je použita permutace $\langle \cdot \rangle$ opačně (klesající a nikoliv rostoucí) než ve zbytku práce, což není ku prospěchu čitelnosti.
- V práci jsou použity symboly $|\cdot|$ pro absolutní hodnoty i pro kardinality množin a to i v jednom a totéž vztahu, viz (4.4).
- Postrádám definice t-normy a t-konormy. Představení pouze čtyř základních v kapitole 5.1 není šťastné, navíc z formy představení není neobeznámenému čtenáři jasné, která z operací je t-normou či t-konormou.
- Na straně 72 je hovořeno o definičních oborech $[0, 1]$, přičemž tyto intervaly zde vystupují jako obory hodnot.
- Na straně 91 je uvedeno 'weakness of I-ISFM if' místo '... of I-ISFM is'.
- Na straně 104 je uvedeno, že dle Obrázku 6.5 bylo nejlepších výsledků dosaženo při použití Choquetova integrálu na základě I-ISFM míry těsně následováno stejným integrálem na základě G-ISFM míry. Nicméně z Obrázku 6.5 vyplývá opak.

Závěrem je třeba podtrhnout, že autor se v této práci skutečně zabývá tématem, které je v dnešní době aktivně studováno mnoha autory, a že autor bezpochyby přinesl do dané problematiky nové zajímavé myšlenky a prokázal dostatek erudice. Finální experimentální část působí také velice přesvědčivě. Vědecký přínos práce je tedy bezesporný. Lze konstatovat, že autor splnil cíle disertační práce a prokázal schopnost samostatné vědecké práce.

Na základě výše uvedeného hodnocení práce, kterou považuji za velmi zdařilou, ji doporučuji k obhajobě.