

Posudek oponenta na doktorskou práci

Cross-entropy based combination of discrete probability distributions for distributed decision making

od Vladimíry Sečkárové

Práce je z oblasti teorie rozhodování a metodologicky se opírá o některé pojmy a postupy teorie informace. Celý přístup k problému je čistě pravděpodobnostní. To znamená, že vstupní data (jakási evidence) jsou dána ve formě souboru pravděpodobnostních rozdělení a očekávaný výstup (odhad) je pak také ve formě pravděpodobnostního rozdělení. Odhadovaná distribuce je tedy jakousi "kombinací" vstupních rozdělení.

Zvolená metoda řešení je celkem intuitivní. Především se očekává, že odhadovaná distribuce je od všech vstupních distribucí stejně vzdálená. Vzdálenost je přitom měřena pomocí Kullback-Leiblerovy (KL) divergence. To pak vede přirozeně na systém lineárních omezení pro hledanou distribuci. Toto vymezení však nemusí vést k jednoznačnému řešení, proto je třeba hledat vhodného reprezentanta. Za tohoto reprezentanta je považována distribuce, která je nejbližší, opět ve smyslu KL divergence, určitému apriornímu rozdělení. Tato nejbližší distribuce má známý exponenciální tvar a tak zbývá jen zvolit vhodné apriorní rozdělení.

Práce je přirozeným způsobem členěna: První kapitola obsahuje základní pojmy a definice. V druhé části je uvedeno řešení v obecné teoretické rovině. Třetí kapitola obsahuje řešení pro Dirichletovo apriorní rozdělení. Zde je metoda dovedena ke konkrétním výpočtům, pro které je Dirichletovo rozdělení vhodné a proto je nadále systematicky využíváno. Ve čtvrté části se autorka vypořádává se situací, kdy vstupní data tak docela neodpovídají předpokladům, tj. jsou k dispozici pouze marginální nebo podmíněná rozdělení, anebo nejsou vůbec v pravděpodobnostním tvaru. V páté kapitole jsou výsledky aplikovány na příklady s reálnými data. V šesté části jsou některé dodatky a porovnání s jinými metodami. V závěrečné sedmé jsou pak úvahy a výhledy pro další výzkum.

Práci bych označil za velmi solidní. Zvolené prostředky jsou zcela adekvátní řešení problematice. Z hlediska teorie nejsou sice výsledky zcela převratné, jsou však dobře zpracovány. Velmi cenná je skutečnost, že teoretické výsledky jsou dotaženy přes výpočetní algoritmy až k úlohám s reálnými daty. Zde je do dizertace vloženo i obrovské množství programátorské práce. Hodna ocenění je též snaha autorky nevyhýbat se problémům s nedokonalými daty. Práce je napsána velmi pečlivě, obsahuje naprosto minimum překlepů a je také psána velmi slušnou angličtinou.

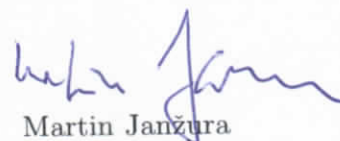
K práci mám jen několik spíše obecných poznámek

- 1) Úvod, který má až filozofický charakter, se mi jeví až příliš "vznešený" vzhledem k spíše praktickému charakteru vlastní práce.

- 2) Možná mohla být rozsáhlejší rešeršní práce a výsledke ještě důkladněji zasazeny do kontextu.
- 4) Kruciálním bodem teoretické části je otázka existence a jednoznačnosti řešení. Náhradní řešení s existencí ve smyslu nejmenších čtverců je pouze naznačeno, nikoli exaktně pojednáno. Nakonec, v klíčovém Proposition 2.4, se existence jednoduše předpokládá.
- 5) Dirichleto apriorní rozdělení je celkem vhodné pro daný problém, Ovšem využití dat pro jeho konstrukci je sice praktické, ale problematické. Apriorní rozdělení je součástí "modelu", který by měl být na datech nezávislý.

Přes zmíněné výhrady autorka prokázala schopnost samostatné vědecké práce a navrhuji udělení titulu PhD.

Praha, 30.června 2015



Martin Janžura