

Posudek oponenta na disertační práci Mgr. Vladimíry Sečkárové
CROSS-ENTROPY BASED COMBINATION OF DISCRETE PROBABILITY
DISTRIBUTIONS FOR DISTRIBUTED DECISION MAKING

Oponent: Prof. RNDr. Jana Jurečková, DrSc., MFF UK

Práce je věnována problematice expertních systémů: Je třeba odhadnout rozdělení pravděpodobností diskretní náhodné veličiny X nabývající známých hodnot x_1, \dots, x_n , ale s malou nebo žádnou informací o vektoru jejich pravděpodobností $\mathbf{q} = (q_1, \dots, q_n)$. Odhady tohoto vektoru pravděpodobností jsou získány z několika zdrojů jakožto expertní názory, a výsledný odhad je kombinací těchto subjektivních odhadů, pokud možno optimální vzhledem ke vhodnému kritériu. Každý takový odhad je podmíněnou pravděpodobností; jako podmiňující rozdělení se v dalším textu uvažuje Dirichletovo rozdělení s určitými parametry.

Optimální kombinace expertních odhadů se získává několika způsoby, např. (1) bayesovsky se ztrátou rovnou Kullback-Leiblerově divergenci, nebo (2) přímo minimalizací Kullback-Leiblerovy divergence, pokud máme alespoň přibližný předběžný odhad rozdělení pravděpodobností \mathbf{q} . Vhodnou kombinaci expertních odhadů je možné najít jen pokud tyto splňují určitou podmínku přípustnosti, což se dále předpokládá.

Dirichletovo rozdělení umožňuje větší flexibilitu a matematickou formulaci problému a za určitých podmínek nalezení optimální kombinace. Většinou se neuvažuje preference mezi experty, což umožňuje použít aritmetický průměr jednotlivých podmíněných pravděpodobností, v souladu s principem neuplné informace a maximální entropie.

Kandidátka pronikla do terminologie a způsobu vyjadřování i myšlení, běžné v oblasti expertních systémů, což zřejmě nebylo lehké; lze tedy ocenit, že její práce již vedla k publikacím. Jednotlivé situace a předpoklady ilustrovala na numerických datech; kapitola 5 je celá věnována zajímavým praktickým aplikacím.

V zajímavé kapitole 6 kandidátka použila metod předcházejících kapitol k postupnému upřesňování předpokládaného rozdělení pravděpodobností v dynamickém exponenciálním systému s regresory a k odhadu parametrů systému; tento výsledek byl již publikován.

Práci pokládám za zdařilou; některé partie a praktické aplikace doporučuji k dalšímu publikování. K práci nemám zásadní připomínky; moje poznámky se týkají hlavně neobjasněného značení a chybějících definic některých pojmů.

str. 17, ř. 4 a jinde: Je třeba napsat, že náš náhodný vektor jsou pravděpodobnosti. Často se mluví o odhadu, ale chybí předmět odhad čeho.

str. 18 a dále: Někde uveďte seznam zkratk, které nejsou běžné, např. zkratky BDM.

str. 18, "constraints": Někde je potřeba definici doplnit motivací.

str. 19: Píše se o kooperativním scénáři. Skutečně experti podle něho postupují, a proč ?

str. 20, ř. (-16): Tento odstavec je třeba vysvětlit. Co znamená "satisfying constraints (2.7) in the least-squares sense" ?

str. 20, ř. 1 Aspoň jednou je třeba napsat přesný tvar $\pi(q|p_1, \dots, p_s)$. Bylo by vhodné psát tučně p_1, \dots, p_s jako vektory, jinak je to matoucí.

Toto jsou jen drobné připomínky, práci pokládám za zdařilou a doporučuji uznat jako doktorskou práci na MFF UK.