

Posudek oponenta bakalářské práce

Název práce: Feature network regularized SVM omics classifier

Autor bakalářské práce: Filip Masri

Posudek vypracoval: Ing. Vojtěch Franc, Ph.D.

Předložená práce studuje možnost integrace apriorní znalosti o interagujících příznacích do metody Support Vector Machines s cílem zvýšit přesnost a interpretovatelnost klasifikačního modelu. Práce je psaná srozumitelnou angličtinou. Problém a navržené řešení je dostatečně přesně popsáno. Hlavní cíle zadání byly podle všeho splněny. Z hlediska obsahu a použité metodologie bych práci vytkl následující nedostatky:

- Použitá “omická” data a řešené klasifikační úlohy jsou popsány jen velmi stručně. Bohužel není uveden ani počet příkladů, což je stěžejní parametr při posuzování efektivity navržené metody. Dále není hned zřejmé, proč by data měla být lineárně separovatelná, jak vyžadují použité formulace SVM.
- Poměrně rozsáhlá část experimentů je věnována porovnání přesnosti klasifikátorů naučených řešením buď primární a nebo duální úlohy. Z teorie je známo, že řešení obou úloh je v tomto případě shodné, což je i pěkně vidět z výsledků. Není tedy úplně jasné, jaký byl účel těchto experimentů.
- Grafy na stranách 44 až 46 nejsou popsány a některé (str. 44, 46) ani čitelné.
- Některé části textu obsahují popis obecně známých metod (např. sekce 2.1, 2.1.1) nebo až zbytečně detailní popis implementace (např. sekce 5.1 a 5.2). Tyto sekce by šlo lehko vypustit anebo přesunout do přílohy.
- Zadání práce specifikuje, že kvadratické optimalizační úlohy spojené s učením SVM klasifikátorů mají být řešeny pomocí knihovny CPLEX. Vzhledem k existenci mnohem efektivnějších optimalizačních metod pro tento typ úloh, by v práci mohla být uvedena motivace pro tuto netypickou volbu.
- V textu se vyskytuje poměrně dost překlepů a i několik technických omylů, například:
 - str. 9, “this criterion is not differentiable, thus impossible to convert into dual form”. To není pravda, protože minimalizaci absolutní hodnoty (tj. nediferencovatelné číslí kriteria) lze snadno převést na minimalizaci lineární funkce s omezením.
 - Konstanta $\beta > 0$ v kritériu úlohy (2.12) je nadbytečná, protože její volbou se mění jen hodnota kriteria, ale řešení zůstává stejné.
 - U definice optimalizační úlohy (2.13) chybí omezující podmínky.
 - Regularizační konstanty λ, β jsou podle všeho laděny na testovacích datech a tudíž je přesnost klasifikace vychýlena k nižším hodnotám (Tabulka 8.18 a 8.19).
 - Obsah některých odstavců je dosti nejasný. Např. sekce 4.2.5 popisující barierové metody a nebo sekce 9.2, ve které autor diskutuje stabilitu řešení se mi nepodařilo pochopit.

Doporučuji předloženou bakalářskou práci k obhajobě a hodnotím ji známkou **B-velmi dobře**.

V Praze 14. června 2015

Ing. Vojtěch Franc, Ph.D.
Centrum strojového vnímání
Katedra kybernetiky, FEL ČVUT
Karlovo nám. 13, 121 35 Praha 2