

Posudek oponenta bakalářské práce

Jméno autora práce: David Milec

Studijní program: Open Informatics

Specializace: Computer and Information Science

Název bakalářského projektu: Identification of Cycling Segments in Raw GPS Trajectories

Posudek vypracoval oponent práce: Radomír Černocho, MSc.

Posuzovaná práce se zabývá přiměřeně obtížným tématem ve zpracování GPS záznamů. Práci považuji za nadprůměrnou zejména ve dvou oblastech. První je obsáhlá rešerše relevantní literatury. Práce korektně cituje 19 zdrojů, z nichž většina je v textu práce shrnuta, zmíněny jsou výhody i nevýhody konkurenčních přístupů. Druhým silným pilířem je jazyková a formální úprava. Autor prokázal dobrou znalost angličtiny, což spolu s přehledným členěním kapitol, podkapitol i jednotlivých odstavců činí text srozumitelným a snadno čitelným – byť některé pasáže by zasloužily delší popis (rozdíl point-wise a trajectory-wise měření, přítomnost či absence štítků v DPNK datasetu, ...).

Moje hlavní výhrada se týká křížové validace. Pro naučení modelu je sice správně oddělena trénovací a testovací množina, ne však pro ladění meta-parametrů, jako je počet neuronů v síti, použitý kernel v SVM a počet stromů v náhodných lesech. Metodicky správný postup by zahrnoval rozdělení datasetu na trénovací, validační a testovací množinu, kde optimalizace meta-parametru sleduje chybu na validační množině, ale kvalita modelu se vykazuje podle testovací chyby. Tuto vadu ale nepovažuji za zásadní, protože dle uvedených tabulek by se přesnost modelu nezhoršila více než o desetiny, maximálně jednotky procent.

Shrnutí: Práce splnila všechny body zadání adekvátním způsobem. Vyniká jazykovou stránkou, srozumitelností a obsáhlou rešerší relevantní literatury. Výhrada týkající se krosvalidace nezpochybňuje dosažené výsledky zásadním způsobem.

Předloženou práci hodnotím klasifikačním stupněm **B – velmi dobře**.

Otázky k obhajobě:

1. Proč se segmentace používá pouze na data v testovací množině a ne v trénovací?
2. Je možné použít nově navržený příznak „proximity to bicycle infrastructure“ i v konkurenčních algoritmech? Doporučil byste takový experiment v dalším výzkumu?

Datum: 5. 6. 2017

Podpis: