

KYRIAKOS TSAPPARELLAS: MODELLING THE GAME OF TENNIS OPONENTSKÝ POSUDOK BAKALÁRSKEJ PRÁCE

Vo svojej bakalárskej práci sa autor zaoberá možnosťami štatistického modelovania a predikcie výsledkov tenisových zápasov. Na vzorke výsledkov štyroch veľkých tenisových turnajov odohraných v priebehu roka porovnáva tri modely navrhované v literatúre a následne sám hľadá odpoveď na otázku, či empirické dáta splňujú predpoklady týchto modelov. Jedná sa o sčasti prehľadovú a sčasti originálnu prácu.

Po stručnom úvode do problematiky sú v kapitolách 4-6 predstavené tri rozdielne regresné modely známe z literatúry z ktorých autor vychádza. Pravdepodobnosť víťazstva vyššie nasadeného hráča v zápase je modelovaná pomocou rôznych prediktorov ako sú postavenie hráča v rebríčku, úspešnosť hráča v minulých zápasoch alebo porovnania fyzických dispozícií oboch súperov. Pri modeli 2 autor okrem jednoduchého popisu použitého modelu navrhuje zahrnúť do výpočtu ako regresor nedávnu úspešnosť hráča (t.j. pravdepodobnosť víťazstva v hre - *game*) pri podaní a pri prijíme počas niekoľkých posledných zápasov.

Jadrom práce je kapitola 7, v ktorej sa opúšťa spoločný predpoklad všetkých modelov o tom, že hry v zápase sú nezávislé a rovnako rozdelené náhodné veličiny. Model 2 je rozšírený tak, aby obsahoval nové regresory zachytávajúce rozdiely vo „významnosti“ jednotlivých hier zápasu rovnako ako rozdiely vo výkonnosti hráča vo chvíľach keď je už prakticky rozhodnuté o jeho blízkom víťazstve/prehre. Autor ukazuje, že napriek tomu, že podobné rozšírenia majú svoje opodstatnenie, pre účely predpovedania výsledkov tenisových zápasov je ich prínos zanedbateľný.

V samotnom závere sú všetky modely porovnávané z hľadiska úspešnosti predikcie. Autor nachádza zaujímavú závislosť medzi kvalitou predikcie výsledku zápasu a postavením súperov v rebríčku. Na základe výsledkov analýzy formuluje niekoľko netriviálnych tvrdení a originálnym spôsobom vysvetľuje vyvstávajúce anomálie.

Práca je zaujímavá, dobre štruktúrovaná a nenašiel som v nej žiadne významné faktické chyby. Odkazy na odbornú literatúru sú korektne citované a vlastný prínos autora je v texte náležite vyznačený.

Napriek tomu však čítanie sťažujú pomerne časté gramatické chyby (za všetky napríklad sústavné používanie *donation* namiesto *notation* (str. 10), použitie *upper limit* namiesto *upper bound* (str. 5), striedanie značenia: Us Open, U.S Open (str. 1), US Open, U.S. Open (str. 3)) a občasné nejasné formulácie (str. 34, ... *you must bet when adequate and complete information are possessed, ... even with less information*).

Okrem tohto problému sú aj moje ďalšie pripomienky a odporúčania väčšinou iba formálneho charakteru:

- str. 5: pri výpočte Clopper-Pearsonovho intervalu spoľahlivosti parametru v binomickom rozdelení boli zamenené čísla: pre prvý náhodný výber v skutočnosti platí $p_L \approx 0.603$ a pre druhý náhodný výber $p_U \approx 0.595$.
- str. 15, Table 5.2: podľa tvrdenia na strane 14 vychádzajú Brier-Score nižšie v modeli *Updated method*, zatiaľ čo v grafe sú označené ako nižšie hodnoty *First method*, asi ide o zamenené značenie metód v grafe.
- str. 13, 5. riadok od konca: indexovanie má zrejme byť $k = 1, \dots, 6$ namiesto $k = 0, \dots, 6$.
- str. 32: tvrdenie o tom, že logaritmická transformácia poradia hráča v rebríčku – *ranking* je „skoro iste“ nevhodná pre niektoré skupiny priemerných hráčov je síce zaujímavé, ale usudzovať tak z jedného dátového súboru pri tak slabej evidencii sa mi zdá byť prehnané. Možno by bolo vhodné porovnať modely s logaritmickou a novou transformáciou navrhnutou autorom na ďalšej, nezávislej množine dát.

- str. 34: v práci chýba vysvetlenie, prečo krivky *Net winnings* pri zahrnutí takmer všetkých zápasov (pre hodnoty *Ranking* okolo 200) v grafe 8.5 nekončia v hlbokých stratách tak ako bolo tvrdené na strane 33.
- str. 35: v práci mi chýba detailnejší popis toho, čo presne znamená krivka *Bettors* v grafe 8.6 a ako bola vypočítaná.
- V celej práci sa niekoľkokrát vyskytujú tvrdenia o významnosti rozdielov odhadov (napr. na str. 3 je tvrdenie že pri Wimbledone je hra pri vlastnom podaní významne úspešnejšia ako pri ostatných uvažovaných turnajoch, prípadne na str. 8 je podobné tvrdenie o úspešnosti hry pri vlastnom podaní v porovnaní Grand Slam-ových a ostatných turnajov). Žiadne z týchto tvrdení však nie je podložené formálnym testom. Rovnako tak by mohlo byť zaujímavé formálne testovanie významnosti podmodelov použitých v práci ako podpora pozorovaní o kvalite predikcie (najmä v kapitole 7. pri zavádzaní ďalších parametrov do známeho modelu). Názov kapitoly 8 (Testing the models) tiež nie je vhodný, pretože sa o žiadne testovanie modelov nejedná, skôr iba o zbierku pozorovaní a úvah o kvalite predikcie.
- Odporučil by som dávať väčší dôraz na popisky objektov ako tabuliek a grafov. Na tieto objekty, rovnako ako na rovnice odporúčam odkazovať sa číslom pre zjednodušenie orientácie v práci.
- Zoznam literatúry by mal mať jednotnú formu, v referenciách chýbajú čísla strán. Rovnako aj podoba odkazovania sa na literatúru by mala byť jednotná (v texte sa objavujú odkazy menom autorov, menom a rokom, rovnako ako aj iba číslom v zozname).

Celkove hodnotím prácu ako zaujímavú a pomerne dobre napísanú. Práca obsahuje autorove vlastné výsledky a množstvo netriviálnych pozorovaní. Nepochybujem preto o tom, že by ju bol schopný ďalej rozširovať. Prácu **odporúčam uznať ako bakalársku prácu.**

Mgr. Stanislav Nagy

8. júna 2012, Praha