

Využití strojového učení pro odhad křivek přežití

Bc. Tomáš Trdla <tomas.trdla@tul.cz>, Ing. Karel Paleček, Ph.D.

Tato práce se zabývá implementací a porovnáním vybraných algoritmů umělé inteligence na odhad křivek přežití. Výchozím modelem je nepoužívanější Coxův model proporcionálních rizik, vůči kterému jsou porovnávány. Jako hodnotící kritéria jsou použity různé metody a přístupy, které dohromady dávají ucelenou představu o kvalitě modelu. Veškeré testy jsou prováděny na reálných anonymizovaných datech. Výsledky prokazují rozdílnou kvalitu modelů na zkoumaných časech přežití a ovlivnění dané kvality velikostí učícího datasetu. Zároveň dokazují, že modely AI jsou schopny dosáhnout přesnějších výsledků než CoxPH model, avšak liší se při rozdílných cílech a datasetech, proto položily základ myšlenky ensemble modelu, která je v rámci této práce teoreticky popsána jako další možné řešení a výzkum.

Klíčová slova: křivky přežití, umělá inteligence, coxův model proporcionálních rizik, neuronová síť, rozhodovací stromové struktury, strojové učení

Úvod

Cílem této práce je prozkoumání schopnosti některých vybraných algoritmů umělé inteligence v oblasti odhadu křivek přežití. Následně pak porovnání kvality a přesnosti daných modelů s běžně používanými modely. Pro experimentální účely byly poskytnuty reálné vzorky dat z oblasti zdravotnictví, konkrétně z databáze transplantací na Ústavu hematologie a krevní transfuze v Praze.

Metodika

Mezi nejznámější a nepoužívanější modely patří semi-parametrický Coxův model proporcionálních rizik [3] (1), který je využíván i ke zkoumání vlivu jednotlivých kovariátů vstupního vektoru na riziko výskytu cílové události. Avšak v mnoha aplikacích je tento přístup příliš zjednodušený, protože předpokládá, že riziková funkce je lineární. Právě tento model má jistou podobnost ve vícenásobné lineární regresi, díky čemuž není z matematické podstaty schopen zmapovat komplexní nelineární vazby v učících datech.

$$\lambda(t|x) = \lambda_0(t) \cdot \exp\left(\sum_{i=1}^n b_i(x_i)\right) \quad (1)$$

Implementovanými modely jsou tedy následující:

1. běžně používaný Coxův model proporcionálních rizik
2. klasická neuronová síť s aktivační funkcí sigmoid ve výstupní vrstvě (cíl je

pravděpodobnost resp. jistota modelu, že se událost projeví)

3. klasifikační neuronová síť s aktivační funkcí softmax ve výstupní vrstvě (rozhodnutí, zda se událost projeví nebo ne)
4. adaptace CoxPH pomocí nahrazení částečně-rizikové funkce neuronovou sítí a místo metody parciální věrohodnosti jako *loss* funkce použita *average-negative-log-likelihood* doplněná o regularizaci L2, což pochází ze studie [1]
5. základní XGBoost model, kde snaha o podobné využití na adaptaci CoxPH jako NN selhala na implementování funkce pro výpočet gradientu kvůli příslušné knihovně
6. Automatic Machine Learning společnosti h2o.ai (cílem bylo využít *state-of-art* algoritmy pro výběr a parametrizaci modelů)
7. Random Survival Forest
8. Gradient Boosting Survival Analysis model, který představuje částečně-rizikovou funkci a je doplněn Nelson-Aalen estimátorem pro odhad základní kumulativní rizikové funkce (stejně jako NN adaptace) což dohromady dává další adaptaci CoxPH modelu

Pro každý model je dále implementována funkce na popis vlivu vstupních proměnných na výsledné riziko vývoje cílové události. Klasický CoxPH model skrývá tento popis v jednotlivých regresních

koeficientech, zatímco AI metody se jeví jako “black-box”, avšak studie [2] předkládá řešení v podobě tzv. “SHAP values”.

V rámci této práce byl zároveň vytvořen vlastní přístup k předzpracování dat (*preprocessing*). Jedním z hlavních důvodů daného přístupu byl problém poskytnutých reálných dat, kdy se po zpracování klasickými postupy velikost (resp. délka) použitelných dat zmenšila o 97,5% (z původních 1385 na 35), zatímco tímto přístupem pouze o 32,5% (z 1385 na 935). Dalším důvodem byla ztráta názvu sloupců, která by učinila zjištění vlivu proměnných na výsledek zbytečným, neboť by byly zobrazeny pouze nic neříkající indexy.

Výsledky a diskuze

Pro hodnocení výsledků kvality a přesnosti jednotlivých modelů byly využito následujících algoritmů: Log-rank, Time-dependent Concordance, Brier Score, Negative binomial log-likelihood, vážení těchto grafů pomocí Kernel Density Estimation z histogramu a nakonec integrace daných grafů pro ucelený přehled. Každý z těchto testů je prováděn na stejných datech a porovnává výsledek modelu vůči skutečnému stavu přežití pacienta. Pro vzájemné porovnávání modelů byla vytvořena třída, která při učení využívá křížovou validaci (*cross validation*) metodou KFold, což zajišťuje učení i testování na všech dostupných datech a díky tomu vyšší statistickou významnost, než pouze jeden test na náhodně vybraných datech.

Na konkrétním příkladu predikce přežití pacienta bylo prokázáno, že adaptace CoxPH pomocí nahrazení částečně rizikové funkce GBSA modelem (regresní rozhodovací stromová struktura učená upravenou *log-partial-likelihood* funkcí) dosahuje lepších výsledků na většině intervalu TTE (*Time-to-event*) časů, především pak na jeho středních hodnotách. Na nízkých časech (do ~150dní) poté nejlépe vychází Random Survival Forest. Následován NN adaptací CoxPH, kde je součástí NN vstupu i simulovaný čas události. Tento model však podává nestabilní výsledky a jeho s reálnými vzorovými daty identická základní riziková funkce (nulová hypotéza Log-rank testu) byla zamítnuta na intervalu 1% i 5%, avšak při změně cílové EOI (*Event-of-interest*) na relaps již platný je. Na vyšších hodnotách TTE vychází nejlépe opět klasický CoxPH, což může být i z důvodu malého počtu událostí

Závěr

Porovnáním několika cílů včetně vlivu velikosti učicího datasetu se došlo k závěru, že nelze striktně říci, že jeden typ modelu je lepší než druhý, neboť pořadí pomyslného žebříčku je proměnné s učicími daty, cílem, ale i jejich velikostí. Avšak ve většině případů docílí AI modely lepších výsledků než standardní CoxPH. Proto byl diskutován návrh ensemble modelu. Ten však nebylo možné realizovat z důvodu omezeného rozsahu práce a může tedy být předmětem další studie v oblasti umělé inteligence a křivek přežití.

Poděkování

Tímto bych chtěl poděkovat MUDr. Janu Vydrovi z ÚHKT za poskytnutí reálných dat a zájmu o toto téma.

Tato práce byla podpořena z projektu Studentské grantové soutěže (SGS) na Technické univerzitě v Liberci v roce 2020.

Reference

- [1] KATZMAN, Jared L., Uri SHAHAM, Alexander CLONINGER, Jonathan BATES, Tingting JIANG a Yuval KLUGER. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology* [online]. 2018, **18**(1) [cit. 2020-05-18]. DOI: 10.1186/s12874-018-0482-1. ISSN 1471-2288. Dostupné z: <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-018-0482-1>
- [2] LUNDBERG, Scott M. a Su-In LEE. A Unified Approach to Interpreting Model Predictions. *Arxiv* [online]. [cit. 2020-05-18]. Dostupné z: <https://arxiv.org/pdf/1705.07874.pdf>
- [3] PRENTICE, Ross L. Introduction to Cox (1972) Regression Models and Life-Tables. *Breakthroughs in Statistics* [online]. New York, NY: Springer New York, 1992, , 519-526 [cit. 2020-05-20]. Springer Series in Statistics. DOI: 10.1007/978-1-4612-4380-9_36. ISBN 978-0-387-94039-7. Dostupné z: http://link.springer.com/10.1007/978-1-4612-4380-9_36