

STROJOVÉ UČENÍ, UMĚLÁ INTELIGENCE A VĚDA O DATECH V BOJI S PANDEMIÍ COVID-19

Jan Hendl

Anotace

V okamžiku rozšíření Covid-19 infekční nemoci se umělá inteligence (UI), strojové učení (ML, machine learning) a věda o datech staly významným pomocníkem v boji proti viru SARS-CoV-2. Metody se využívají při diagnóze, k vývoji nových léků a očkovacích látek, k modelování a předpovědi šíření a k monitorování výskytu nemoci v populaci a v logistice zdravotnictví. Covid-19 pandemie zvýšila snahy členů komunity vědců z oblastí UI, ML a vědy o datech v hledání řešení problémů, které pandemie vyvolala. Objem literatury o aplikacích UI, ML a vědy o datech se stále zvětšuje. V našem příspěvku podáváme přehled hlavních oblastí aplikací a informujeme o literatuře a některých výsledcích snah při zvládnání COVID-19 pandemie.

Klíčová slova

strojové učení, umělá inteligence, věda o datech, COVID-19

1 Úvod

COVID-19 infekční onemocnění bylo deklarováno Světovou zdravotní organizací (WHO) jako pandemie. Nový koronavirus SARS-CoV-2 se šířil od konce roku 2019 rychle a všude, přičemž způsobil poškození zdraví mnoha milionů lidí. Naštěstí také věda reagovala rychle.

Onemocnění COVID-19 zasáhlo celý svět a mnoho lidí včetně vědců se snaží najít prostředky ke zmírnění jeho následků. Gautam et al. [3] zkoumali 196 000 medicínských odborných článků o COVID-19, z nich podrobněji analyzovali 28 904, zjistili, že průměrně se objevilo v jednom týdnu 990 prací, největší počet článků pocházel z Číny. COVID-19 se týkalo 733 systematických přehledů, 480 meta-analýz a 362 klinických randomizovaných pokusů.

V okamžiku úderu Covid-19 pandemie se umělá inteligence (UI, AI, Artificial Intelligence), strojové učení (ML, machine learning) a věda o datech staly významným pomocníkem v boji proti viru SARS-CoV-2 [např. 1, 7, 10, 13, 14].

Islam et al. [10] analyzovali v systematickém přehledu 729 článků s aplikacemi z oblasti UI a strojového učení publikovaných v anglickém jazyku. Z tohoto počtu pocházelo 26% z Čínské republiky, 23,7% z USA a Indie 12,6%. Tyto tři země přispěly podle autorů nejvíce k celkovému počtu článků. Příspěvky se objevily hlavně v časopisech PLUS One, Chaos Solitons Fractals a Journal of Medical Internet Research. Z oblasti UI byly nejvíce citovány práce autorů Huang et al. [8] (150x), Guan et al. [5] (81x), He et al. [6] (78x).

K řešení problémů pandemie se používají nejnovější informační technologie z oblastí umělé inteligence, strojového učení a vědy o datech. Pro základní představu uvádíme stručné vymezení těchto termínů [7].

Umělá inteligence

Existuje několik definic umělé inteligence. Jedna stručně říká, že UI znamená užití počítačů k modelování inteligentního chování s minimální intervencí člověka. UI je považována za inženýrský obor, který implementuje nové koncepty a řešení, které jsou schopny zvládat složité úkoly.

Strojové učení

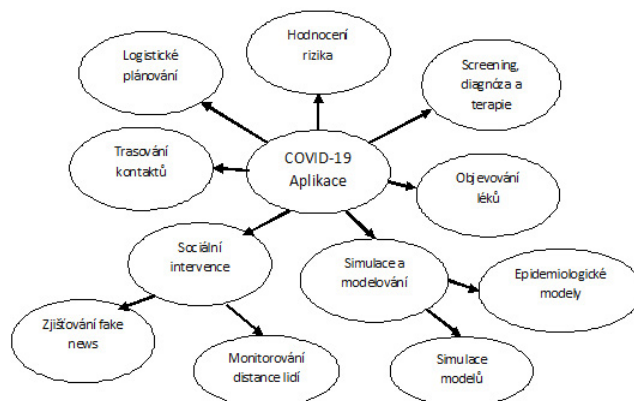
Strojové učení znamená praxi programování učení se z dat. Je spojováno s dobýváním znalostí z dat (DM, Data Mining), DM označuje snahu získat znalosti z velkého množství dat.

Věda o datech

Věda o datech (data science) je široký termín zahrnující oblasti jako strojové učení, statistické učení, modelování časových řad, vizualizace dat, expertní systémy a pravděpodobnostní uvažování. Zřejmě se v mnohém překrývá s vymezením umělé inteligence a strojového učení. Znamená interdisciplinární oblast, která využívá vědecké metody, procesy a algoritmy k extrahování znalostí a získání vhledu do daného problému pomocí velkého počtu strukturovaných a nestruturovaných dat.

2 Aplikace informačních technologií při řešení problémů s COVID-19

Výzkumníci využívali metody umělé inteligence, strojového učení, zpracování přirozeného jazyka, statistiky, matematiky, modelování, simulací a další vědecké metody analýzy strukturovaných a nestruturovaných dat [15].



Obrázek 1 – Přehled aplikací vědy o datech při řešení problémů s pandemií COVID-19

Shrnutí oblastí aplikací pokročilých informačních technologií při řešení problémů s pandemií COVID-19 ukazuje obrázek 1 (upraveno podle [11]).

Zdravotní systém se vypořádává se značným tlakem na své zdroje. To vedlo k potřebě rychle odhadovat hladinu rizika u pacienta a příslušně reagovat. To se děje již dnes v případě rakoviny, diabetu nebo kardiovaskulárních onemocnění. Takové opatření je nutné zkoumat a iniciovat také v případě COVID-19. Například údaje jako věk, pohlaví nebo zdravotní stav mohou být využity při odhadu mortality rizika úmrtí. To může být důležité pro výběr na jednotkách intenzivní péče, pokud jsou omezené kapacity.

Diagnostika a screening má podobný význam. Hlavním problémem, kterému se čelí, jsou chyby v řádném screeningu a diagnostice [16]. Například osoby s mírnými příznaky si často nejsou vědomi, že jsou nositeli infekce a nemoci. Kromě testování je klíčovým úkolem vývoj nástrojů pro diagnostiku na dálku pomocí počítačů.

Ve firmách a ve veřejnosti se používají různá testovací schémata. Rychlý rozvoj internetu věcí (IoT, Internet of Things) ústí do masivní explose dat generovaných na těle nesených přístrojů a sensorů. Jejich analýza je velmi důležitá. Zkoumá se možnost na základě zvuku zakašlání nebo míry pohybové aktivity rozpoznat přítomnost onemocnění nebo zvýšení zdravotního rizika. Mnoho studií využívá umělou inteligenci a strojové učení ke zpracování výstupů zobrazovacích technik, k rychlejší a přesnější detekci onemocnění [4].

Pandemie COVID-19 způsobila nedostatek pracovníků v určitých zdravotnických profesích (např. v primární péči). To vedlo ke zkoumání automatizovaných prostředků péče a expertních systémů (nebo k jejich zdokonalování) nebo počítačem podpo-

rovaného vyhodnocování klinických dat. Použití pulsní oximetrie a detekce změn signálu na dálku pomáhá například snižovat mortalitu u pacientů v domácím ošetřování.

Data a umělá inteligence pomáhají vyvíjet nové léky nebo ukázat, že používané léky k jinému účelu představuje užitečnou strategii i při zvládnání onemocnění COVID-19. [12] Oblast medicíny založené na biologických sítích aplikuje techniky vědy o sítích v medicíně pro využití a vývoj léků pomocí počítačových výpočtů. Výpočetní metody mohou zkrátit čas strávený zkoumáním dat, předpovídáním struktury proteinů a genomů. Mohou také pomoci při identifikaci vhodných pacientů pro klinické studie, což je často časově náročná a nákladná část vývoje léčiv.

Epidemiologické modely a simulace poskytují možnost získat představu, jak se šíří epidemie, což je důležité pro plánování populačních opatření a logistiky. V epidemiologických modelech se populace rozděluje na propojené části a vývoj v čase se modeluje diferenciálními rovnicemi (např. se využívá model SIR). Užitím Bayesovských hierarchických modelů se takové modely kombinují a sestavují jako globální modely průběhu celé pandemie.

Simulační modely mají širokou aplikovatelnost. Všechny otázky nemohou být zodpovězeny epidemiologickými modely. Používají se při rozhodování se vztahem k přenosu nemoci (karanténní opatření nebo sociální odstup), při uvažování kapacit počtu lůžek, jednotek kritické péče, nutného personálu, při rozhodování o limitech pro přijetí a propouštění pacientů a o minimalizaci dopadu na ostatní pacienty. Currie et al. [2] podávají podrobný přehled o tom, jak modely a simulace mohou pomoci snížit dopad COVID-19.

Online šetření nebo sledování mobilů pomáhají při rychlém odhalování infikovaných jedinců, při zavádění karanténních opatření a omezování shromažďování lidí. Taková opatření se vyrovnávají s kompromisy mezi zajištěním veřejného zdravotnictví a soukromým právem. Extrahují se data ze sociálních médií, z osobních přenosných zařízení. Další informace se získávají měřením a analýzou dat ekologického znečištění, dat o vodních zdrojích a ovzduší.

Šíření zavádějících zpráv může narušit strategie v oblasti veřejného zdraví [7]. Provádějí se výzkumy, jak odlišit zavádějící zprávy od vědecky podporovaných zpráv. Informace o těchto akcích obsahuje také Wikipedie.

Pandemie COVID-19 způsobila vznik zvýšené potřeby vhodné organizovat logistiku zdravotní péče. Týká se to například zajištění testovacích souprav, speciálních lůžek a ventilátorů, ochrany ošetřujícího personálu nebo převedení části personálu na jiné pracoviště. I zde se uplatňují techniky umělé inteligence. Pomáhají také při optimalizaci řízení dodavatelského řetězce pro zásobování zdravotních center. U lze použít i k preventivní alokaci zdrojů. Datoví výzkumníci také pomáhají při odhalování nezvyklého chování zákazníků v prodejních řetězcích [10].

3 Závěr

Od okamžiku objevení nového viru SARS-CoV-2 vědci a pracovníci zdravotnického sektoru po celém světě hledají metody boje proti vzniklé pandemii. Týká se to screeningu a predikce, sledování kontaktů, předpovídání průběhu nemoci a vývoje vakcín nebo léků s přesnějšími a účinnějšími metodami.

Strojové učení a postupy umělé inteligence jsou slibnými metodami, které využívají poskytovatelé zdravotní péče. Náš příspěvek se zabývá studii, které shrnují v přehledu zmíněné pokročilé technologie při aplikaci výzkumných výsledků a při vývoji algoritmů asistence v reálných problémech.

Umělá inteligence a ML mohou výrazně zlepšit léčbu, diagnostiku, screening a predikci, předpověď, vyhledávání kontaktů a vývoj léků/vakcín v kontextu pandemie Covid-19.

Závěrem poznamenejme, že u nadějných výsledků je také nutné uvažovat jejich omezení. Mnohdy jsou k vývoji postupů UI a strojového učení zapotřebí velké objemy dat o vysoké kvalitě. Výzkumníci se setkávají se značnými limity v tomto směru. Například elektronické záznamy jsou často segregovány podle nemocnic, regionálně nebo podle států. Scházejí platné standardy a větší mezinárodní spolupráce. Také záleží na rychlosti v přístupu k novým datům. Data jsou často po jejich analýze zastaralá a s nimi i získané výsledky. Proto mnoho modelů není nasazeno v praxi tak, aby se ověřilo jejich fungování v reálném světě.

Výzkumníci z oblastí vědy o datech, strojového učení a umělé inteligence ve spolupráci se zdravotnickými odborníky pomáhají při zvládnání problémů způsobených pandemií COVID-19. Objem tak vznikající literatury v tomto kontextu je značný. Je zapotřebí pro informační potřeby připravovat přehledy. Náš příspěvek měl poukázat na některé literární zdroje z této oblasti.

Literatura

- [1.] Ahuja, A.S. et al. *Artificial Intelligence and COVID-19: A Multidisciplinary Approach*. *Integr. Med. Res.* 2020, 9, 100434.
- [2.] Currie, C. S. M. et al. *How simulation modelling can help reduce the impact of COVID-19*. *Journal of Simulation*, 14(2), 83-97. DOI: 10.1080/17477778.2020.1751570
- [3.] Gautam, P. et al. *COVID-19: A bibliometric analysis and insights*. *International Journal of Mathematical, Engineering and Management Sciences* 5(6), 1156–1169, 2020. DOI: 10.33889/IJMEMS.2020.5.6.088 2020.
- [4.] Ghaderzadeh, M., Asadi, F. *Deep learning in detection and diagnosis of Covid-19 using radiology modalities: A systematic review*. *arXiv* 2020, arXiv:2012.11577.
- [5.] Guan, W. et al. *Clinical characteristics of 2019 novel coronavirus infection in China*, *New England Journal of Medicine*, 2020, DOI: 10.1056/NEJMoa2002032
- [6.] He, K. et al. *Momentum Contrast for Unsupervised Visual Representation Learning*. 2020 arXiv:1911.05722v3
- [7.] Hendl, J. *Big data*. *Grada* 2021.
- [8.] Huang, C. et al. *Clinical features of patients infected with novel coronavirus in Wuhan, China*. *Lancet*, 2020 Feb 15;395(10223):497-506, DOI: 10.1016/S0140-6736(20)30183-5
- [9.] Chang, A.C. *Artificial intelligence and COVID-19: Present state and future vision*. *Intelligence-Based Medicine*, 3-4 (2020) 100012, 2020.
- [10.] Islam, M..M., Poly, T.N. et al. *Application of Artificial Intelligence in COVID-19, Pandemic: Bibliometric Analysis*. *Healthcare*, 9, 441. DOI: 10.3390/healthcare9040441, 2021,
- [11.] Latif, S. et al. *Leveraging Data Science To Combat COVID-19: A Comprehensive Review*. *IEEE Transactions on Artificial Intelligence*, DOI: 110.1109/TAI.2020.3020521, 2020.
- [12.] Mak, K.K., M. R. Pichika, M.R. *Artificial intelligence in drug development: present status and future prospects*. *Drug discovery today*, 24(3) 773–780, 2020.
- [13.] Naseem, M. et al. *Exploring the Potential of Artificial Intelligence and Machine Learning to Combat COVID-19 and Existing Opportunities for LMIC: A Scoping Review*. *Journal of Primary Care & Community Health*, 11, s. 1–11, 2020.
- [14.] Nguyen, T.T. *Artificial intelligence in the battle against coronavirus (COVID-19): A survey and future research directions*. *arXiv* 2020, arXiv:2008.07343.
- [15.] Syeda, H.B. et al. *Role of Machine Learning Techniques to Tackle the COVID-19 Crisis: Systematic Review*. *JMIR Med Inform*, 2021, 9(1), e23811, s. 1-19, 2021.
- [16.] Vinod, D.N., Prabakaran, S.R.S. *Data science and the role of Artificial Intelligence in achieving the fast diagnosis of Covid-19*. *Chaos, Solitons and Fractals*, 140 (2020) 110182, 2020.

Kontakt

Prof. Jan Hendl
LF1 UK
U Hadovky 8
160 00 Praha 6
[e-mail: jan.hendl@ruk.cuni.cz](mailto:jan.hendl@ruk.cuni.cz)