

Primerjava umetne inteligence s sorodnimi tehnologijami

Gregor Burger, Fakulteta za elektrotehniko in Digitalno inovacijsko stičišče Slovenija

Povzetek — Pojem umetne inteligence pogosto zasledimo v medijih, objavah in še marsikje. Pogosto pa se ne zavedamo njenega pomena in prepletanja z drugimi sorodnimi tehnologijami. Nekatere tehnologije in principi so si nasprotni, mnogi pa se medsebojno dopolnjujejo oz. nadgrajujejo. V objavi smo predstavili nekatere pare tehnologij, ki so v laični javnosti pogosto medsebojno zamenjujejo, kar pa ni vedno ustrezno. Osredotočili smo se na umetno inteligenco, velike podatke, podatkovno rudarjenje, strojno učenje, statistik in poslovno inteligenco.

Ključne besede — umetna inteligenca, veliki podatki, podatkovno rudarjenje, strojno učenje, statistika, poslovna inteligenca

Abstract — The concept of artificial intelligence is often found in the media, publications or other places. Often, however, we are unaware of its importance and intertwining with other related technologies. Some technologies and principles are opposite, but some are complementing each other. In this publication, we presented some pairs of technologies that are often mistakenly used by the public. We focused on artificial intelligence, big data, data mining, machine learning, statistics and business intelligence

Keywords — artificial intelligence, big data, data mining, machine learning, statistics, business intelligence.

1. UVOD

V objavi z naslovom Definicije in algoritmi umetne inteligence smo predstavili različne poglede na umetno inteligenco in njene definicije. Dodatno smo se dotaknili še pogosto uporabljenih algoritmov v umetni inteligenci, kot so strojno učenje, globoko učenje in nevronske mreže. Za lažje razumevanje, pred nadaljnjim branjem, priporočamo branje predhodno omenjenega dokumenta. Dan danes pojem umetna inteligenca velikokrat zasledimo v objavah v medijih, spletnih objavah, televiziji, serijah in podobno. Zagotovo se pogosto sploh ne zavedamo, da smo uporabljali ali pa je so bili v povezavi z našimi podatki uporabljene aplikacije umetne inteligence. Meja kaj je umetna inteligenca, kje in v kakšni meri se uporablja je pogosto neukemu opazovalcu zabrisana. V nadaljevanju bomo skušali predstaviti presek umetne inteligence s sorodnimi modernimi tehnologijami in trendi, kot so veliki podatki, podatkovna znanost, poslovna analitika.

Poglej si nekaj primerjav med umetno inteligenco in sorodnimi tehnologijami.

2. UMETNA INTELIGENCA IN VELIKI PODATKI

Umetna inteligenca in veliki podatki (angl. Big data) [1], [2], [3] sta pojma katerima se namenja v poslovnem svetu veliko pozornosti, prav tako pa tudi v akademskih in raziskovalnih ustanovah. Umetna inteligenca in veliki podatki ne deluje eden nasproti drugemu, ampak v sodelovanju enega z drugim, za uresničitev skupne naloge. Kako se umetna inteligenca nasploh razlikuje od velikih podatkov? Velike podatke je mogoče opisati kot surove vhodne podatke, katere je potrebno prečistiti, strukturirati in integrirati, da postanejo uporabni. Medtem, ko umetna inteligenca predstavlja izhodni rezultat, predhodno obdelanih vhodnih podatkov. Umetna inteligenca je način obdelave podatkov, ki omogoča strojem opravljati kognitivne funkcije, npr. ukrepanje na podlagi vhodnih podatkov. Tradicionalni načini obdelave podatkov so zahtevali ročno implementacijo ukazov ukrepanja na podlagi vhodnih podatkov. Takšni sistemi si bili togi in so v primeru nepričakovanih rezultatov potrebovali uporabnikovo posredovanje. Umetna inteligenca pa je sposobna, v mnogih primerih, samostojno prirediti potrebno ukrepanje.

Sistem umetne inteligence je zasnovan za samostojno analizo in razumevanje podatkov na podlagi katerih sistem skuša poiskati rešitev. Sistem se nauči pravilnega ukrepanja od določenih izidih rezultatov, to rešitev pa se zmožen uporabiti tudi v prihodnje, če zazna podoben vhodni problem. Veliki podatki so po svoji zasnovi starejši način obdelave podatkov. Aktivnosti ne izvedejo na podlagi rezultatov, podatke se le analizira. Kot nakazuje že samo ime veliki podatki lahko vsebujejo veliko količino, tudi zelo heterogenih podatkov. Mogoči zajeti podatki so po svoji naravi lahko relacijski, ne relacijski, strukturirani, ne strukturirani, slike, elektronska pošta, podatki senzorjev itd. Primarna uporaba velikih podatkov je predvsem za

pridobivanje vpogleda v same podatke. Umetna inteligenca pa podatke uporabi za učenje, nato pa se na podlagi za rezultatov odloči za ukrepanje.

Imata pa umetna inteligenca in veliki podatki tudi skupne točke, se medsebojno dopolnjujeta. To pa je tudi eden izmed razlogov zakaj se ju pogosto medsebojno zamenjuje. Algoritmi umetne inteligence za svoje delovanje in učenje potrebujejo podatke. Večji kot je razpoložljiv nabor podatkov za učenje, posledično boljši je rezultat z manjšo napako. Podatki je pred uporabo v algoritmičnih umetne inteligence potrebno prečistiti in strukturirati, drugače ne bomo prejeli željenih rezultatov. Čeprav sistem umetne inteligence izvede postopek učenja na osnovnem učnem podatkovnem setu, se sistem umetne inteligence skozi svoj življenjski cikel neprenehoma uči. Za to pa so potrebni vedno novi in novi podatki.

Z razvojem vzporednega procesiranja je postal razvoj in uporabnost umetne inteligence in tudi velikih podatkov hitrejši, bolj dostopen in učinkovit. Ne smemo pa pozabiti, da za uspešno delovanje sistemov umetne inteligence potrebujemo veliko količine podatkov, ki pa so vedno bolj na voljo.

3. PODATKOVNO RUDARJENJE IN STROJNO UČENJE

Podatkovno rudarjenje (angl. Data mining) in strojno učenje (angl. Machine learning) so tehnike oz. algoritmi, ki jih uporablja umetna inteligenca za svoje delovanje [4], [5]. Umetna inteligenca za iskanje rešitev uporablja tehnike rudarjenja podatkov, ki tako služijo kot osnova delovanja umetne inteligence. Strojno učenje pa obravnava široko področje algoritmov za učenje sistemov umetne inteligence. Samo strojno učenje ima globoke povezave s področjem statistike, a ga ne obravnavamo kot takšno.

Čeprav obstaja med podatkovnim rudarjenjem in stojnim učenjem veliko podobnosti, pa je med njima kar nekaj bistvenih razlik. Podatkovno rudarjenje uporabljamo za iskanje različnih atributov podatkovnih setov, kakšne so relacije med atributi, pogosto s tehnikami vizualizacije. Z analizo želimo pokazati odvisnosti med vsaj dvema atributoma in napovedati izid rezultata. Strojno učenje pa uporabljamo za izdelavo napovedi izidov kot so ocene cen ali ocena trajanja poteka dogodka. Algoritem se skozi svoje izvajanje uči in podaja vedno boljši rezultat. Funkcijsko je podatkovno rudarjenje tehnika, ki z analizo podatkov podaja koristne informacije. Strojno učenje pa je postopek uporabe kompleksnih algoritmov za iteracije strojne

obdelave podatkov. Stojno učenje ima večjo praktično uporabo s podajanjem ocen za produkte, čas trajanje, optimizacije delovnih procesov, podatkovno rudarjenje pa se pogosteje uporablja v raziskovalne namene in analizo. Različna sta tudi koncepta delovanja, pri čemer je osnovni koncept podatkovnega rudarjenja podajanje koristnih informacij z iskanjem trendov in vzorce, koncept strojnega učenja pa je učenje na obstoječih podatkih z namenom izboljšanja svojega delovanja. Zavedati pa se moramo, da strojno učenje za svoje delovanje uporablja tudi tehnike podatkovnega rudarjenja. Ena pglavitnih razlik med obema tehnikama je že način izvajanja operacij. Postopki rudarjenja podatkov se izvajajo glede na zahtevo po analizi podatkov, postopek strojnega učenja pa na iterativni osnovi, ob predpostavki, da so nam vedno na voljo novi podatki za obdelavo. Iterativni način dela omogoča strojnemu učenju, da se sistem samostojno uči in prilagaja glede na vhodne podatke. Podatkovno rudarjenje pa potrebuje človeški poseg za izvajanje prilagoditev in zamenjavo tehnik rudarjenja. Po ocenah podaja strojno učenje točnejše rezultate kot jih podaja podatkovno rudarjenje, saj je prvi postopek avtomatiziran, drugi se izvaja ročno. Podatkovno rudarjenje za analizo potrebuje izdelane modele. Proces podatkovnega rudarjenja zavzema podatkovno bazo, jedro podatkovnega rudarjenja (angl. data mining engine) in postopek evalvacije vzorcev. Stojno učenje pa je izvedeno z algoritmi strojnega učenja, nevronske mreže, globokega učenja in podobno. Točnost podatkovnega rudarjenja je odvisna od kakovosti in točnosti vhodnih podatkov. Pridobljeni podatki s postopkom podatkovnega rudarjenja se uporabijo v algoritmičnih stojnega učenja, ki posledično podaja točnejše rezultate. V splošnem, se po ocenah strokovnjakov, algoritmi stojnega učenja izkazujejo z večjo točnostjo kot podatkovno rudarjenje. Ima pa stojno učenje tudi eno pglavitno slabost v primerjavi s podatkovnim rudarjenjem. Podatki, ki jih je mogoče uporabiti morajo biti prilagojeni formatom podatkov podprtih v algoritmičnih stojnega učenja. Podatkovno rudarjenje pa ima spodobnost obdelave in pretvorbe večjega števila različnih formatov podatkov. Zato se uporablja podatkovno rudarjenje za pripravo podatkov, podatki se nato pretvorijo v ustrezen format podatkov, ki jih je nato mogoče uporabiti pri strojnem učenju.

4. PODATKOVNO RUDARJENJE, STOJNO UČENJE IN GLOBOKO UČENJE

Kot je bilo podrobno opisano že v predhodnem poglavju ima strojno učenje možnost učenja na

podatkovnih setih in napovedovanja izvidov glede na vhodne podatke. Globoko učenje je podskupina algoritmov stojnega učenja [6], [7]. Globoko učenje posnema človeški postopek obdelave podatkov. Postopek je podoben kot pri človeških možganih, ki zaznavajo različne vzorce s primerjavo že predhodno poznanih vzorcev. Globoko učenje ima možnost prepoznave atributov surovih podatkov avtomatsko, stojno učenje pa potrebuje ustrezno pred pripravljene podatke. Globoko učenje je tesno povezano z nevronskimi mrežami, ki posnemajo delovanje človeških možganov. Podatkovno rudarjenje, pa kot prav tako predhodno že opisano, odkriva skrite vzorce in pravila v velikih količinah podatkov.

5. UMETNA INTELIGENCA IN STATISTIKA

Stojno učenje je podskupina umetne inteligence, pri čemer se sistem samostojno uči za podajanje rezultata. Posplošeno rečeno je strojno učenje le učenje na podatkih z algoritmi. Statistika pa je veja matematike, za odkrivanje vzorcev in korelacij z matematičnimi orodji [8], [9], [10]. Če izpostavimo nekaj ključnih razlik med umetno inteligenco, natančneje stojnim učenjem in statistiko. Strojno učenje je podskupina umetne inteligence, ki je ukvarja z napovedovanjem rezultatov na podlagi vhodnih podatkov samostojno, brez človeškega posredovanja. Statistika pa je veja matematike, ki temelji na statističnih vrednostih in statistični obdelavi podatkov. Strojno učenje je ena izmed panog podatkovne znanosti, pri čemer so znanja statistike osnova za zasnovo modelov strojnega učenja. Za zasnovo modela strojnega učenja je potrebno načrtovati kateri atributi bodo uporabljeni, kakšni bodo rezultati in podobno. Potrebno pa je predvideti tudi korelacije med neodvisnimi spremenljivkami in podatkovnimi točkami. Čeprav so razlike med konceptoma stojnega učenja in statistiko, pa sta oba koncepta neločljivo povezana. Brez statističnega znanja ni mogoče ustvariti modela, ki ga nato za svoje napovedovanje uporablja strojno učenje. Statistična znanja uporabljamo tudi po izdelavi modelov stojnega učenja. S statistiko merimo oz. ocenjujemo rezultate in napovedi. Obravnavati pa moramo tudi pojav resnično pozitivnih, lažni negativnih, resnično negativnih in lažno pozitivnih rezultatov. Strojno učenje in statistika skupaj podajata vpogled v podatke, ki omogoča povečanje prodaje, povečanje vpliva ali deleža trga.

6. POSLOVNA INTELIGENCA IN STOJNO UČENJE

Poslovna inteligenca [11] (angl. Business Intelligence - BI) je eno izmed področij podatkovne analitike, ki je množično vpeta v poslovne procese podjetij. Poslovna inteligenca se usmerja na iskanje uspešnih rešitev poslovne strategije, če naštejemo le nekaj primerov, ki jih obravnava. Med drugim se to zajem, analiza in razlaga podatkov in sprejemanje odločitev. Moderni pristopi poslovne inteligence omogočajo uporabnikom samostojno kreiranje vsebin in analiz, brez občutne podpore IT oddelka podjetja. Delovanje in koncept strojnega učenja pa je bilo zadostno opisano že v predhodnih poglavjih.

Oglejmo si primerjavo med poslovno inteligenco in stojnim učenjem. Strojno učenje kot uporabi vhodne podatke za treniranje modela, na podlagi katerih pripravi napovedi izidov. Poslovna inteligenca pa se bolj naslanja na tehnologije, aplikacije in principe, ki omogočajo sprejemanje strateških odločitev v podjetju. Tudi poslovna inteligenca potrebuje za svoje delovanje podatke, pri čemer imamo v misli velike podatke. Velike podatke je potrebno vizualizirati, da prinesejo dodano vrednost. Poslovna inteligenca se običajno uporablja v dva namena. Prvi je vodenje podjetja in drugih je izvedba spremembe poslovanja oz. delovanja podjetja.

Poslovna analitika se usmerja na analizo podatkov, te funkcionalnosti strojno učenje nima. Analiza podatkov predvideva uspešnost poslovnih strategij na podlagi podatkov in ne občutkov vodilnih kadrov podjetja. Ob tem ponuja podjetjem uporabno orodje za uporabo podatkov poslovanja podjetja v za rast podjetja primernem načinu. Odvisnost poslovne analitike in strojnega učenja temelji na kakovosti, dostopnosti in točnosti podatkov. Večja kot je njihova dostopnost, točnost in kakovost, boljši bodo končni rezultati.

7. ZAKLJUČEK

Predstavili smo nekaj primerjav med umetno inteligenco in sorodnimi tehnologija. Pogosto se namreč opaža nepravilno razumevanje in napačna uporaba strokovnih izrazov, algoritmov in principov pri opisovanju umetne inteligence. Kot je razvidno marsikatero napačno poimenovanje izhaja iz nerazumevanje relacij med posameznimi tehnologijami. Nekatere so podmnožice druge, na primer umetna inteligenca ter trojno učenje. Ali pa se tehnologiji oz. procesa dopolnjujeta kot je to v primeru velikih podatkov in velikih podatkov. Podali smo le osnovne opise in razlage delovanja, za

podrobnejše razumevanje pa priporočamo poglobljeno preučevanje posameznih orodij, algoritmov, principov in podobnega.

LITERATURA

- [1] <https://bigdatapath.wordpress.com/2020/11/02/big-data-vs-artificial-intelligence/> (prvi dostop: 16. 4 . 2021)
- [2] <https://www.datamation.com/big-data/big-data-vs-artificial-intelligence> (prvi dostop: 16. 4 . 2021)
- [3] <https://enterpriseproject.com/article/2019/10/how-big-data-and-ai-work-together> (prvi dostop: 16. 4 . 2021)
- [4] <https://www.softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/> (prvi dostop: 5. 5. 2021)
- [5] <https://www.forbes.com/sites/forbesagencycouncil/2018/08/01/do-you-know-the-difference-between-data-analytics-and-ai-machine-learning/> (prvi dostop: 5. 5. 2021)
- [6] <https://www.softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/> (prvi dostop: 26.04.2021)
- [7] <https://www.mathworks.com/discovery/deep-learning.html> (prvi dostop: 26.04.2021)
- [8] <https://onlinestats.canr.udel.edu/machine-learning-vs-statistics/> (prvi dostop: 26.04.2021)
- [9] <https://www.datasciencecentral.com/profiles/blogs/machine-learning-vs-traditional-statistics-different-philosophi-1> (prvi dostop: 26.04.2021)
- [10] <https://www.educba.com/machine-learning-vs-statistics/> (prvi dostop: 26.04.2021)
- [11] <https://www.educba.com/business-intelligence-vs-machine-learning/?source=leftnav> (prvi dostop: 27. 4.2021)

Gregor Burger je raziskovalec na Fakulteti za elektrotehniko Univerze v Ljubljani in sodelavec Digitalnega inovacijskega stičišča Slovenije . Njegova raziskovalna zanimanja vključujejo telemedicino, m-zdravje, raziskave uporabniške izkušnje in uporabnosti, zasnove uporabniških vmesnikov in aplikacij navidezne, obogatene in mešane resničnosti ter sledenju pogledov uporabnikov (eye-tracking). Trenutno je vključen v številne projekte razvoja intuitivnih uporabniških vmesnikov in uporabniške izkušnje. Je tudi član mednarodne elektrotehniške organizacije IEEE.