

Strojové učení

(Olga Štěpánková, Petr Berka, Filip Železný)

Všechny živé organismy se snaží využívat pro dosažení svých cílů všeho, co charakterizuje jejich životní prostředí. Pro poznání těchto charakteristik využívají zkušenosti, které postupně získávají během svých aktivit. K tomu slouží **učení**, tedy proces transformující zkušenosti na znalosti, informace, dovednosti či hodnoty, které přispívají k tomu, aby organismus mohl lépe nebo snadněji dosahovat stanovených cílů. V digitálním prostředí získávají zkušenosti i počítačové systémy, a to ve formě dat, se kterými pracují nebo která generují. Různé počítačové systémy realizující vzájemně blízké úlohy sbírají své „zkušenosti“ ve formě souborů dat a ty mohou velmi přirozeně sdílet! Strojové učení se snaží tuto situaci využít a hledá i aplikuje algoritmy či statistické modelování pro zpracování takto získaných „zkušeností“ do formy, která umožní počítačovému systému příslušnou úlohu efektivně řešit jen na základě využití shromážděných zkušeností (trénovacích data), tedy bez přesných instrukcí, jak postupovat.

Úlohy, které strojové učení řeší, se liší podle vlastností souborů trénovacích dat. Když trénovací data obsahují ke každému vstupu i údaj o požadovaném výstupu, hovoří se o učení s učitelem. Pokud trénovací data reprezentují jen typické vzorky možných vstupů bez dalšího upřesnění, je cílem strojového učení identifikovat v těchto datech zatím nepopsanou charakteristickou strukturu (odpovídající např. typickým shlukům) a hovoří se o učení bez učitele. V dynamickém prostředí se uplatňuje posilované učení, při kterém systém postupně shromažďuje zkušenosti bez přesné informace o požadovaném výstupu, ale pouze doplněné o zpětnovazební údaj, který informuje o dílčí odměně či pokutě za provedený krok.

Strojové učení již teď nabízí řadu metod, které lze aplikovat nezávisle na oblasti, která generuje zkoumaná data. Výsledkem učení s učitelem často bývají rozhodovací stromy, soubory rozhodovacích pravidel nebo různé typy neuronových sítí. Při učení bez učitele se vedle např. shlukování a bayesovských metod uplatňují dokonce i neuronové sítě. Pro vývoj dalších metod strojového učení se často stávají zdrojem inspirace úlohy v nějaké nezvyklé aplikační oblasti. Strojové učení proto úzce souvisí i s dalšími oblastmi jako je data science, zpracování řeči a přirozeného jazyka, robotika či rozpoznávání. Je přirozené, že k tématu „strojové učení“ se hlásí takřka všechna pracoviště v ČR, která se věnují umělé inteligenci.

Přehled témat strojového učení, kterým se věnují nejvýznamnější pracoviště v ČR:

- **UI AV ČR:** Studium vlastností neuronových sítí z pohledu jejich aplikovatelnosti na úlohy separace dat, regresní analýzy a extrakce funkčních závislostí z dat. Návrh modelů neuronových sítí a jejich optimalizačních strategií a algoritmů. Aplikace neuronových sítí na řešení klasifikačních úloh z oblasti vysokých energií, rozpoznávání vzorů a predikce časových řad.
- **UTIA AV ČR:** Návrh invariantního příznakového popisu objektů. Aplikace v medicíně.

- **CVUT-CIIRC:** Posilované učení a učení z analogií. Použití učení pro zefektivnění logického odvozování. Rozvoj neuronových sítí, zpracování signálů a předzpracování dat pro dobývání znalostí z dat, detekce anomálií. Vizualizace vícerozměrných a multimodálních dat. Aplikace v úlohách optimalizace, rozpoznávání, učení robotů a v medicíně.
- **CVUT-FEL:** Statistické učení včetně hlubokých neuronových sítí, symbolické učení včetně metod založených na logických reprezentacích, induktivní logické programování, relační učení, kombinace statistických a symbolických metod. Aplikace ve strojovém vnímání a rozpoznávání obrazu, v bioinformatice a robotice.
- **MU-FI:** Vývoj spolehlivých algoritmů pro segmentaci, kvantifikaci a sledování buněk. Registrace, segmentace, extrakce rysů a klasifikace. Vizualizace multimodálních dat. Aplikace strojového učení při zpracování textu v přirozeném jazyce a pro analýzu rozsáhlých výukových dat. Učení nad relačními daty. Detekce anomálií.
- **OU-ÚVAFM:** Fuzzy přístupy ke strojovému učení, strojové učení popisu chování dynamického systému v přirozeném jazyce.
- **TUL-ITE:** Aplikace strojového učení při zpracování textu v přirozeném jazyce, automatické rozpoznávání a přepis řeči.
- **UK-MFF:** Použití metod on-line strojového učení pro tvorbu obsahu pro 3D počítačovou grafiku. Strojové učení při zpracování přirozeného jazyka. Dobývání znalostí z dat.
- **UPOL:** Výzkum nových metod objevování faktorů v binárních, ordinálních a obecných relačních datech; snižování dimenzionality relačních dat; formální pojmy a kognitivní psychologie; klasifikace binárních a ordinálních dat.
- **VŠB:** Vývoj SW prostředí vhodného pro paralelní spouštění řady úloh strojového učení – využito pro vývoj predikčních modelů.
- **VŠE:** Vývoj systémů pro hledání asociačních a rozhodovacích pravidel. Vysvětlitelné strojové učení. Aplikace strojového učení v medicíně, byznysu a veřejné správě
- **VUT-FIT:** Pokročilé modely neuronových sítí a analýza časových řad či rozsáhlé grafové úlohy na tocích dat. Detekce objektů v obrazech a v kolaborativní robotice. Neuronové sítě pro extrakci parametrů z řeči a multi-lingvální trénování pro rozpoznávání řeči
- **VUT-CEITEC:** Systém optického rozpoznávání znaků využívající proprietární SW
- **ZČU:** Učení pro rozpoznávání a syntézu mluvené nebo znakové řeči. Intuitivní učení (programování) robotů, pokročilé metody učení robotů operátory výroby (bez nutnosti znalosti standardních přístupů k programování robotů)

Významné výsledky

- **UI AV ČR:** Návrh multikriteriálního evolučního algoritmu pro generování matoucích vzorů, který nevyžaduje přístup k parametrům modelu a je aplikovatelný na jakýkoli model strojového učení. (P. Vidnerová, R. Neruda. Vulnerability of Machine Learning Models to Adversarial Examples. Proceedings ITAT 2016, 187-194. ISBN 978-1-5370-1674-0. ISSN 1613-0073. ;
- **CVUT-CIIRC:** Systém rCoP pro automatické dokazování velmi inovativně používá posilované učení pro průběžné zefektivňování svého výkonu a díky tomu pravidelně

vyhrává světové soutěže ve své disciplíně. (C. Kaliszyk, J. Urban, H. Michalewski, M. Olsak: Reinforcement Learning of Theorem Proving. NeurIPS 2018: 8836-8847)

- **CVUT-FEL:** Systém pro hluboké relační učení. (Sourek G., Aschenbrenner V., Zelezny F., Schockaert S., Kuzelka. O: Lifted Relational Neural Networks: Efficient Learning of Latent Relational Structures. Journal of Artificial Intelligence Research 62:69-100, 2018)
- **VŠB:** vytvořila model, který předpovídá pro off-grid systém (daný výčtem zapojených zdrojů) kvalitu vyráběné elektrické energie. Výsledky umožňují hledat co nejefektivnější využití nestandardních zdrojů v souladu s požadavky odběratele. (T Vantuch, S Misak, T Jezovicz, T Burianek, V Snasel: The Power Quality Forecasting Model for Off-Grid System Supported by Multi-objective Optimization, IEEE Trans.on Indust. Electronics, 64(12): 9507-9516, 2017)

Nejvýznamnější projekty

- **MFF-UK** aplikuje metody strojového učení pro tvorbu 3D počítačové grafiky s cílem zjednodušit její používání, a tak jí otevřít cestu do nových odvětví a průmyslových aplikací. „High-fidelity appearance fabrication“, GAČR, 19-07626S, 2019-2021
- **VŠE: LISp-Miner** – data miningový systém pro výzkum a aplikace metody GUHA; zahrnuje procedury pro hledání bohatě strukturovaných asociačních pravidel, akčních pravidel a podmínek, za kterých kontingenční tabulky nebo histogramy splňují zadaná kritéria zajímavosti.
- **VUT-BUT:** SW nástroj **KALDI** pro dolování informací z řeči, který umožňuje identifikaci řečníka i jazyka, rozpoznávání jazyka a detekci klíčových slov.

Vize rozvoje a příspěvku k ekonomice.

Metody strojového učení se mohou uplatnit všude, kde jde o vývoj nebo návrh systému pro podporu rozhodování a rozpoznávání. Na zajímavé příklady aplikací upozorňuje nejen popis významných výsledků a projektů, ale i přehled témat řešených na různých pracovištích v ČR.

Aplikační výstupy zajímavé pro Svaz průmyslu a dopravy a další partnery

UI AV ČR

- Expertní dovednosti při implementaci metod strojového učení a deep learningu pro řešení separačních a regresních úloh; pokročilé metody pro předzpracování dat a extrakce významných (a odvozených) proměnných.
- Modelování růstu krystalů pomocí umělých neuronových sítí a gaussovských procesů
- Patentované řešení metody extrémně rychlého výpočtu odezvy neuronové sítě s přepínacími jednotkami pro časově kritickou on-line filtraci či separaci dat (Úřad průmyslového vlastnictví Praha, patent č. 306533, 2016).

UTIA AV ČR

- Původní velmi efektivní metody pro invariantní rozpoznávání objektů na těžce poškozených snímcích.

CVUT-FEL

- Spolupráce s Toyota Motor Europe na vývoji asistivních technologií a autonomního automobilu

CVUT-CIIRC

- Použití strojového učení v kombinaci se strojovým viděním pro odhad kvality a jízdních vlastností vozovky a pro rozpoznání okolí nezávisle na sezonních vlivech.
- Použití hlubokých NN pro identifikaci neobvyklého chování řidiče (spolupráce s VW)

OU-ÚVAFM

- Automatický systém pro čtení registračních značek projíždějících automobilů z digitálních obrazů zaznamenaných v reálném provozu. Realizace na základě smlouvy s fy CGI IT Czech Republic, s.r.o. Systém je v provozu od r. 2017.

VUT-CEITEC

- Projekt H2020 AUTODRIVE (2017 – 2020) zkoumá architektury vhodné pro systémy tvořené elektronickými součástkami, které jsou zabezpečeny proti chybě nebo jsou schopny detekovat, když se dostanou do chybového stavu. Cílem výzkumu je vyvinout architektury, které zajistí plně bezpečný provoz samořiditelných vozidel v plně automatizovaném režimu.
- Nástroje a metody zpracování videa a obrazu pro zvýšení efektivity operací bezpečnostních a záchranných složek (VRASSEO, projekt MV ČR)
- Využití hlubokých neuronových sítí pro neanotovatelné či pouze zčásti anotovatelné objemné data-sets v dopravních a průmyslových úlohách vizuální inspekce.

TUL

- Spolupráce s firmou Newton Technologies na vývoji technologií pro automatický přepis mluvené řeči

VUT-FIT

- Systém pro rozpoznávání emocí, kognitivní zátěže a psychického stavu z videa, nositelné elektroniky a řeči (možná vhodné i pro řidiče).

ZČU

- Asistent pro automatické couvání s dvounápravovým vlekem (Digiteq Automotive s.r.o., Volkswagen Group)