|  |  |  |
| --- | --- | --- |
|  | Astronomický ústav *Akademie věd České republiky, v. v. i.* |  |

**Čeští vědci objevili téměř tisíc vzácných vesmírných objektů**

**díky spojení lidské a umělé inteligence**

Tisková zpráva ze dne 4. ledna 2021

**Výzkumný tým z Astronomického ústavu Akademie věd České republiky a Fakulty informačních technologií (FIT) ČVUT použil poprvé v astronomii metodu umělé inteligence zvanou aktivní hluboké učení (active deep learning) založenou na interaktivní­m vylepšování předpovědí mnohovrstvé konvoluční neuronové sítě na základě názoru experta. Na rozdíl od běžně požívaných postupů si síť sama požádá člověka o radu v těch případech, kde si je nejméně jistá. Metodu čeští výzkumníci s úspěchem aplikovali na čtyřech milionech spekter z největšího světového archivu spekter pořízeného čínským dalekohledem LAMOST a objevili skoro tisí­c dosud nepopsaných velmi vzácných vesmírných objektů prokazujících se emisní­mi spektrálními čárami. K nim patří vedle horkých hvězd s rychle rotujícími disky i např. nově vznikající hvězdy s formujícím se planetárním systémem či naopak velmi hmotné vyhořívající hvězdy těsně před výbuchem. Klíčovou roli při objevu sehrálo také třináct tisíc spekter pořízených pomocí ondřejovského Perkova dalekohledu o průměru dva metry. Práce byla publikována v prestižním astronomickém časopise Astronomy and Astrophysics a katalog nově objevených hvězd je zveřejněn v celosvětové databázi Vizier.**

Ve vesmíru se vzácně vyskytují objekty skrývající velmi zajímavé fyzikální podmínky a neobvyklé procesy a nemusí jít jen o populární černé díry, za jejichž objev byla v letošním roce udělena Nobelova cena.

Patří sem i horké hvězdy označované jako Be, okolo kterých se vytvořil dosud neobjasněným mechanismem řídký rychle rotující plynový disk, nebo nově vznikající hvězdy zvané T Tauri, ještě zabalené v zárodečné mlhovině, u kterých jsme svědky formování protoplanet z okolního plynu a prachu. Další zajímavou skupinou jsou kataklysmické proměnné hvězdy, ve kterých bílý trpaslík krade hmotu svému souputníkovi tak dlouho, než dojde k mohutnému výbuchu v podobě novy či dokonce supernovy. Vzácné jsou i velmi hmotné (s hmotou několika desítek Sluncí) horké hvězdy s povrchovou teplotou desítek až stovek tisíc stupňů (Slunce má méně než šest tisíc), které jsou označované jako typ WR podle astronomů Wolfa a Rayeta. Tyto hvězdy vyčerpaly zásobu vodíku a spalují uhlík, dusík či kyslík, odfoukly silným hvězdným větrem již většinu hmoty do okolního prostoru a i je čeká poměrně brzy gigantická exploze v podobě supernovy.

Všechny tyto objekty vypadají na snímcích oblohy stejně jako běžné hvězdy a při jejich odhalení nepomůže ani kombinace CCD snímků v několika barevných filtrech, jak nám ji poskytují rozsáhlé obrazové přehlídky oblohy. Spolehlivě je usvědčí jen jejich spektra, ve kterých se nacházejí charakteristické kombinace emisních čar. Každý jistě ví, že duha vznikající na vodních kapkách je vlastně spektrum našeho Slunce, a mnoho lidí si i všimlo, že podobné spektrum vidí i po odrazu na povrchu CD nebo DVD. Pokud ale pustíme na DVD přes úzkou štěrbinu světlo zářivky či pouliční výbojky, uvidíme jen několik oddělených úzkých barevných čar. To jsou právě emisní čáry vzniklé vyzařováním plynu v lampě. Stejné je to u vesmírných objektů. Z jejich spektra dokážeme určit nejen chemické složení, hustotu a teplotu atmosféry, existenci disku či výtrysků hmoty, ale i jak rychle se od nás vzdalují či přibližují, či dokonce zda mají okolo sebe planety. Podobně podle jiného typu emisního spektra poznáme bezpečně kvazary, tedy jádra galaxií mohutně zářící díky superhmotným černým děrám o hmotnosti miliónů až miliard hmot Slunce v jejich nitru.

Dá se říci, že spektrum je vlastně pro vesmírný objekt takový nezaměnitelný otisk prstu. Astronomové si ho nejraději prohlížejí v podobě grafu závislosti intenzity záření na vlnové délce (čili barvě) a vypadá jako vodorovná či trochu nakloněná čára s různými kopci (emisní čáry) a údolími (absorbční čáry). Přestože se desítky let vyvíjejí automatické algoritmy, jak objekty podle podoby spekter zařazovat do různých tříd, fungují dobře jen pro běžné hvězdy, které umíme i celkem dobře matematicky modelovat. Ale na výše zmíněné exotické případy to neplatí. Tam hraje stále zkušené oko astrofyzika klíčovou roli.

V minulosti velký dalekohled pořizoval spektrum jednoho objektu několik hodin na fotografickou desku. Takto byl také před téměř sto lety Edwinem Hubblem objeven rudý posuv ve spektrech vzdálených galaxií a bylo tím prokázáno rozpínání vesmíru. Moderní spektrografy obsahují tisíce optických vláken, která robotický podavač automaticky umístí podle fotografie hvězdného pole na jednotlivé objekty a během desítek minut je naráz pořízeno několik tisíc spekter.

Díky tomuto pokroku dnes existují archívy čítající deset i více milionů spekter, z nichž naprostou většinu lidské oko dosud nevidělo, a většina informací o nich je pouze výsledkem aplikace automatických algoritmů.

**Největší spektrální přehlídkou je v současnosti archív** **čínského šestimetrového dalekohledu LAMOST** disponujícího čtyřmi tisíci vlákny umísťovanými pomocí dvojnásobného počtu mikromotorů v zorném poli o průměru deseti měsíčních úplňků. LAMOST nese jméno Guo Shoujingův dalekohled po významném středověkém čínském astronomovi, geografovi, inženýrovi a vědeckém poradci Kublajchána. Jeho znalosti a přesné přístroje obdivoval i Marco Polo během svých pobytů. Málo se ví, že Guo Shoujing vytvořil přesný kalendář na základě stanovení délky oběhu Země s chybou pouhých 26 sekund, i to, že dosáhl podobné přesnosti měření pohybu planet jako tři století po něm Tycho Brahe.

Při plánování pozorování s LAMOSTem se na konkrétní cílové objekty v zorném poli umístí jen pár set vláken na základě priorit vědeckých programů a zbytek pozic vybere automatický algoritmus na základě rozsáhlých hvězdných katalogů. Ten vlákna umisťuje tak, aby si jejich ramena či motorky nepřekážely, a přitom tak, aby vybrané hvězdy byly dostatečně jasné pro pořízení dobrého spektra. O drtivé většině cílů, na které jsou umístěna vlákna, proto dosud není známo více než v oněch katalozích (obvykle poloha a jasnost v několika filtrech). Je proto velká šance najít zde dosud neznámé exotické objekty popsané výše. Navíc nová spektra přibývají tempem několika milionů za rok. Stačí "jen" aby si všechna spektra v archívu detailně prohlédl zkušený astrofyzik.

Je ale zřejmé, že to není prakticky možné. I kdyby expert označil (zařadil do příslušné třídy) jedno spektrum každých 10 sekund, trvala by mu klasifikace současného veřejného archívu LAMOSTu s devíti milióny spekter skoro tři roky. Proto se vkládají velké naděje do použití umělé inteligence. V současnosti hlavně do hlubokých (tj. mnohovrstvých) neuronových konvolučních sítí s úspěchem používaných v samořiditelných autech, mobilních hlasových asistentech, strojovém překladu, personalizaci reklamy či při rozpoznávání zvuků, písma a tváří. Díky své podstatě, odvozené z procesu analýzy obrazu v lidském mozku, jsou velmi dobré ve vnímání drobných detailů stejně jako hlavních rysů celkového obrazu. Měly by tedy dobře poznat i rozdíly ve spektrech stejně jako zkušený astrofyzik.

Přes svoje nesporné přednosti však mají jednu zásadní nevýhodu. Vyžadují velké množství, běžně desítky či stovky tisíc, člověkem označených příkladů, na kterých se musí učit, aby pak poznaly podobný vzor v milionech neznámých. Tomuto postupu se říká učení s učitelem (supervised learning).

*"Pokud takové označené vzory nemáte, nebo jich je málo, nemůžete tyto nástroje umělé inteligence použít"*, říká spoluautor článku Ondřej Podsztavek, doktorand na FIT ČVUT. *Bylo jasné, že archív LAMOSTu skrývá mnoho tajemného, ale zpočátku jsme v něm neměli vůbec žádné případy emisních hvězd.* Problém vypadal beznadějně.

*„Mojí velkou touhou, jako astrofyzika pracujícího přes třicet let s dalekohledem o průměru 2 metry na stelárním oddělení Astronomického ústavu Akademie věd České republiky v Ondřejově, bylo přimět umělou inteligenci, aby objevila ve vesmíru něco nového, co ještě neznáme, alespoň dosud neznámé Be hvězdy, kterými se na Ondřejově zabýváme přes půl století"*, říká Petr Škoda, který mimo práce na observatoři v Ondřejově působí také na FIT ČVUT. Zde vede již témeř deset let bakalářské a diplomové práce založené na použití nejmodernějších softwarových technologií a umělé inteligence při řešení náročných astronomických problémů. Pro tento obor se v poslední době používá termín **astroinformatika** a ve světě začínají na významných univerzitách vznikat specializovaná výzkumná centra sdružující astronomy, softwarové inženýry a odborníky na strojové učení.

*"Malý astroinformatický tým jsme před dvěma lety dali dohromady i na FIT ČVUT v rámci naší výzkumné skupiny HPC a velká data v rámci velkého grantu MŠMT nazvaného RCI, Výzkumné centrum informatiky, na kterém se podílíme spolu s kolegy z Fakulty elektrotechnické a Fakulty jaderné a fyzikálně inženýrské ČVUT. Loni jsme na grant přijali z Oxfordské univerzity i kolegu Karla Adámka, jehož software umožnil zachytit velmi vzácné rychlé rádiové zablesky (FRB)"*, poznamenává Pavel Tvrdík, spoluautor práce a vedoucí výzkumné skupiny. *"Zmíněná publikace je prvním důležitým výstupem našeho astroinformatického týmu"*, dodává.

Pro hledání emisních spekter v LAMOST archívu bylo třeba překonat dvě zásadní překážky. Ukázat síti spektra známých emisních objektů, jako by je pořídil LAMOST, a donutit síť, aby fungovala i při zlomku označených vzorků, než běžně vyžaduje.

Tým astroinformatiků po mnoha neúspěšných experimentech nakonec našel unikátní řešení.

Pomohl tomu archív dvoumetrového Perkova dalekohledu v Ondřejově, který obsahuje několik tisíc CCD spekter oněch vzácných objektů s emisními čárami. Spektrograf Perkova dalekohledu na rozdíl od LAMOSTu ale pořizuje spektra s větším spektrálním rozlišením, tedy vidí jemnější detaily spektrálních čar, ale v kratším úseku spektra. Proto byla nejprve tato spektra uměle rozostřena, jak by ony objekty viděl LAMOST. To, že to funguje, se podařilo ověřit, když se na základě stejných hvězdných souřadnic našly tři objekty pozorované jak v Ondřejově, tak v Číně. Takto uměle vytvořených spekter však bylo jen 13 000, což je jako počáteční trénovací vzorek pro hlubokou neuronovou síť pro tento problém stále málo.

Tato druhá překážka byla překonána pomocí málo známé **metody aktivního učení**. Ta byla v informatické literatuře často s úspěchem použita na případy, kdy je jen několik málo známých vzorů, avšak vždy byla zkoušena s velmi jednoduchými algoritmy strojového učení, jako jsou např. rozhodovací stromy. *"Ačkoli jsme intenzivně hledali v astronomické literatuře i ptali se kolegů ve světě, zdá se, že jsme první, kdo použil aktivní učení ve spojení s hlubokými neuronovými sítěmi v astronomii"*, říká Petr Škoda. Ondřej Podsztavek vysvětluje: *"Principem je, že neuronová síť si v každém opakovaném kroku vybere ta spektra, kde si je svojí předpovědí nejméně jistá. Tato předloží expertovi (což může být nejen člověk ale i jiný algoritmus), který předpověď sítě (např. zda patří předložené spektrum Be hvězdě) potvrdí či naopak vyvrátí. Toto se provede na malém vzorku například sta spekter. Jakmile jsou expertem označená, jsou přidána do trénovací množiny, na které je síť znovu učena. Takto se do ní postupně dostává stále více velmi těžko rozhodnutelných případů. To se opakuje do okamžiku, kdy se již síť perfektně trefuje.*

Tímto poměrně komplikovaným způsobem se podařilo najít ve čtyřech milionech spekter ze starší veřejně dostupné verze archívu dalekohledu LAMOST přes 4000 objektů s emisními čarami. Při následném porovnávání se známými astronomickými databázemi se ukázalo, že se o většině z nich skutečně ví, že jsou nějaké zvláštní (např. že se rychle mění jejich jasnost, nebo např. nějaká družice zjistila, že vyzařují rentgenové záření), u části se ukázalo, že jsou to již známé a správně zatříděné emisní objekty. Kromě toho ale zbylo **skoro tisíc objektů** dosud v literatuře detailně nepopsaných (objevují se jen záznamy o jejich poloze a jasnosti). Jejich seznam je jako součást článku k dispozici dalším výzkumníkům prostřednictvím hojně používané celosvětové databáze astronomických katalogů Vizier. Při shromažďování informací o nich dotazováním celosvětového systému propojených astronomických archivů a databází, pro který se používá název Virtuální observatoř, bylo objeveno i **několik podivných objektů, které neumíme zařadit** ani po důkladné analýze dat. *"Tyto se pokoušíme sledovat 2m Perkovým dalekohledem, ale některé jsou už moc slabé. Budeme muset žádat čas na něčem mnohem větším,"* povzdychne si Petr Škoda, který má na mysli velké dalekohledy např. na Kanarských ostrovech, Havaji nebo v Chile.

*"Naše metoda aktivního hlubokého učení nastiňuje budoucí směr používání umělé inteligence v astronomii a možná i jiných vědách"*, říká Petr Škoda a pokračuje: *"Stroj bude pracovat v těsné spolupráci s člověkem, bude za něj dělat rutinní práci, ale jakmile si nebude jistý, obrátí se o radu. Na této koncepci teď začínáme v našem týmu pracovat. Časem by takové řešení mohlo připomínat dialog mezi velitelským můstkem lodi Enterprise a palubním počítačem v seriálu Star Trek při průzkumu Galaxie..* A dodává:

*"Během našeho výzkumu jsme dospěli k přesvědčení že člověk bude vždy potřeba aby pomohl počítači se složitým rozhodováním, zejména v případech, kdy bude k dispozici málo dat. Zkušený expert bude vždy (třeba i podvědomně) rozhodovat na základě více indicí než kolik bylo ve vstupních datech pro algoritmus strojového učení. V tom tkví síla lidské intuice a je dobrá zpráva v době, kdy jsou v mediích často prezentovány obavy, že umělá inteligence brzy nahradí člověka. Mimochodem v odborné literatuře se pro toho experta v aktivním učení používá poetický termín Orákulum, čili věštec, věštba, odkazující na slavnou Pythii v Delfách, ke které přicházeli lidé z celého Řecka aby si potvrdili či vyvrátili svá závažná rozhodnutí"*

Odkaz na publikaci:

Active deep learning method for the discovery of objects of interest in large spectroscopic surveys, 2020, Astronomy and Astrophysics, 643, A122; doi: 10.1051/0004-6361/201936090" - Škoda P., Podsztavek O., Tvrdík P.

https://www.aanda.org/articles/aa/full\_html/2020/11/aa36090-19/aa36090-19.html

katalog objektů: http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/643/A122

Spolupracující subjekt:

Fakulta informačních technologií ČVUT, Praha

Kontakt:

RNDr. Petr Škoda, CSc. – Astronomický ústav AV ČR, Stelární oddělení, telefon 323 620 361, 608927409, email: skoda@asu.cas.cz



Jeden z nalezených objektů Složité emisní spektrum je vyzařováno z podivné symetrické trojice červenýc hvězd uprostřed obrázku. Snímek je z archivu projektu PanSTARRS.