

Čeští vědci objevili téměř tisíc vzácných vesmírných objektů díky spojení lidské a umělé inteligence

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 před několika dny udělena Nobelova cena.

Patří sem i horké hvězdy označované jako Be okolo kterých vznikl řídký plynový disk nebo nově vznikající hvězdy zvané T Tauri, ještě zabalené v zárodečné mlhovině, u kterých můžeme pozorovat vznik protoplanet formujících se z okolního plynoprachového disku. Další zajímavou skupinou jsou kataklysmické proměnné, ve kterých bílý trpaslík krade hmotu svému souputníkovi tak dlouho, než dojde k mohutnému výbuchu v podobě novy či dokonce supernovy. Velmi vzácné jsou i hmotné (s hmotou desítek Sluncí) horké hvězdy s povrchovou teplotou desítek až stovek tisíc stupňů (Slunce má méně než 6000), které jsou označované jako typ WR podle pánů Wolfa a Rayeta. Tyto hvězdy vyčerpaly zásobu vodíku a spalují uhlík dusík či kyslík, odhodily 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 moc nepomůže při jejich odhalení ani kombinace snímků v několika barevných filtrech.

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 se ale podíváte přes úzkou šterbinu na zářivku či pouliční výbojku, uvidíte několik oddělených úzkých barevných čar. To jsou právě emisní čáry vzniklé vyzařováním plynu v lampě. Stejně je to u vesmírných objektů. Z jejich spektra dokážeme určit nejen chemické složení, hustotu a teplotu atmosféry, existenci disku či výtrysků hmoty, ale i zda jsou součástí vícehvězdného systému nebo zda mají okolo sebe planety. Podobně podle jiného typu emisního spektra poznáme kvazar či tzv. aktivní galaktické jádro, tedy vlastně onu v úvodu zmíněnou černou díru.

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 dlouhá čára s různými úzkými kopci (naše emisní čáry) a údolími (čáry absorpční). 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 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 pořizeno několik tisíc spekter. Díky tomuto pokroku dnes existují archívy čítající deset a více milionů spekter, z nichž naprostou většinu lidské oko dosud nevidělo a většina informací o nich je výsledkem automatických algoritmů aplikovaných při jejich zpracování z původních CCD snímků.

Největším spektrální přehlídkou je v současnosti archív čínského šestimetrového dalekohledu LAMOST disponujícího čtyřmi tisíci vlákny umístěvanými pomocí dvojnásobného počtu mikromotorů v zorném poli o průměru deseti měsíčních úplňků. LAMOST nese jméno Guo Shoujingův dalekohled po významném středověkém čínském astronomovi, geografovi, vodním 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 na jen pár set vláken na základě priorit vědeckých programů a zbytek pozic vybere automatický algoritmus na základě rozsáhlých hvězdných katalogů. Ten vlákna umísťuje tak, aby si jejich ramena či motorky nepřekážely, a přitom aby vybrané hvězdy byly dostatečně jasné pro pořízení dobrého spektra. O drtivé většině cílů tak dosud není známo více než v katalogích (obvykle jasnost a barva). Je zde proto velká šance najít zde dosud neznámé exotické objekty popsané výše. Navíc nová data přibývají tempem několika milionů za rok. Stačí "jen" aby si všechna spektra v archívu expert detailně prohlédl.

Je jasné, že to není prakticky možné. I kdyby expert okomentoval jedno spektrum každých 10 sekund, trvala by mu klasifikace archívu LAMOSTu s 10 milióny spekter přes 3 roky. Proto se dnes vkládají velké naděje do použití umělé inteligence. V současnosti hlavně do hlubokých (tj. mnohovýšších) neuronových konvolučních sítí s úspěchem používaných v samoříditelných autech, mobilních hlasových asistentech, strojovém překladu, personalizaci reklamy či při rozpoznávání zvuků, písma, a tváří. Díky své podstatě, odvozené z mechanismu 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.

Přes svoje nesporné přednosti však mají jednu zásadní nevýhodu. Vyžadují velké množství, běžně desítky či stovky tisíc, člověkem okomentovaných 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 komentované vzory nemáte, nebo jich je málo, nemůžete tyto velmi propracované a komerčně masivně propagované nástroje umělé inteligence použít", říká spoluautor článku Ondřej Podsztavek, doktorand na FIT ČVUT. Bylo jasné, že archiv LAMOSTu skrývá mnoho tajemného, ale zpočátku jsme neměli vůbec žádné případy emisních hvězd z LAMOSTu. Problém vypadal beznadějně. 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 komentovaných vzorků než běžně vyžaduje.

Mojí velkou touhou jako astrofyzika pracujícího přes třicet let s 2m dalekohledem na stelárním oddělení Astronomického ústavu 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 Ondřejově také téměř deset let vede na FIT ČVUT bakalářské a diplomové práce založené na použití nejmodernějších IT 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 nedávno dali dohromady i na FIT ČVUT v rámci naší pracovní skupiny superpočítačových výpočtů díky velkému grantu MŠMT nazvanému RCI-Výzkumné centrum pro informatiku, na kterém se podílíme spolu s kolegy s FEL ČVUT", poznamenává Pavel Tvrdlík, spoluautor práce a vedoucí zmíněného grantu na FIT ČVUT. "Zmíněná publikace je jeho prvním konkrétním výstupem", dodává.

Astroinformatický tým po mnoha neúspěšných experimentech nakonec našel unikátní řešení.

Pomohl archiv 2m Perkova dalekohledu v Ondřejově, který obsahuje několik tisíc CCD spekter oněch vzácných objektů s emisními čarami. Spektrograf 2m dalekohledu ale pořizuje spektra s větším spektrálním rozlišením, tedy vidí jemnější detaily spektrálních čar, ale v kratším úseku spektra. Proto byla 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 3 objekty pozorované jak na Ondřejově tak v Číně. Těchto uměle vytvořených spekter však bylo jen 13 000, což je jako počáteční trénovací vzorek pro hlubokou neuronovou síť 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 se neučí vše, co ji předložíte, ale v každém mnohokrát opakovaném kroku si vybere ty příklady, kde si nejméně svoji volbou jistá. Ty předloží expertovi (což může být nejen člověk ale i jiný algoritmus) pro kterého se používá poetický termín Orákulum - věstec, věštba odkazující na slavnou Pýthii v Delfách, ke které přicházeli lidé z celého Řecka si potvrdit či vyvrátit závažná rozhodnutí. Orákulum předpověď sítě (např. patří předložené spektrum Be hvězdě?) potvrdí či naopak vyvrátí. Toto se provede na malém vzorku třeba sto spekter.

Jakmile jsou takto okomentována, jsou přidána do trénovací množiny, na které je znovu síť učena. Takto postupně se do ní dostává stále více velmi těžko rozhodnutelných případů. Toto se opakuje do okamžiku, kdy již se síť perfektně trefuje. Občas kvalitu její získané inteligence zkontrolujeme na sadě příkladů, které známe, ale které ona nikdy neviděla."

Tímto poměrně komplikovaným způsobem (znázorněným na obrázku) se podařilo najít ve 4 milionech spekter ze starší, veřejně dostupné verze LAMOST archivu, přes 4000 objektů s emisními čarami. Při následném porovnávání se známými astronomickými databázemi se ukázalo, že skutečně o většině z nich se ví, že jsou nějaké zvláštní (např že se rychle mění jejich jasnost, nebo např. nějaká družice zjistila, že vyzařují rentgenové záření). Kromě toho ale zbylo skoro tisíc objektů dosud v literatuře detailně nepopsaných (objevují se jen záznamy o jejich poloze a jasnosti). Jejich seznam je 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 nečem mnohem větším." Povzdychne si Petr Škoda, který má na mysli velké dalekohledy např. v Chile.

"Naše metoda aktivního hlubokého učení tak nastiňuje budoucí směr používání umělé inteligence v astronomii a možná i jiných vědách", říká Petr Škoda. "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", zasní se.

reference

Spolupracující subjekt:

"Fakulta informačních technologií ČVUT Praha"

Kontaktní osoba (jméno, telefon , email):

"Petr Škoda, 323 620361, skoda@asu.cas.cz"

Publikace podepírající výsledek:

"Škoda P., Podsztavek O., Tvrdík P.

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"