بخش اول

۱. دسته بندی های خاص و شاخه هایی که برای بررسی کیهان ما به وجود آمده است را بیان کنید.

پاسخ: علم کیهان شناسی به واسطه ویژگی که دارد هم در بازه گسترده ای از مقیاسهای زمانی یعنی از زمانی در حدود زمان پلانک تا زمان حال که تقریباً 14 میلیارد سال از سن عالم می گذرد و هم در بازه مقیاسی مکانی و طولی گسترده ای از مقیاس های ذرات بنیادی تا ساختارهای بزرگ مقیاس و اندازه کل عالم که البته بزرگتر از طولی که نور از ابتدای کیهان تا زمان حال طی کرده، است (به دلیل انبساط کیهان)، مدلسازی نظری و پدیده شناختی دارد. از این روست که شاخه های متعدد پژوهشی وجود دارند که به موضوع کیهان شناسی مرتبط می شوند. از سویی دیگر اگر به سوالاتی نظیر مبداء کیهان چیست و کجاست؟، اندازه و سن کیهان ما چقدر است؟، شکل کیهان ما چگونه است؟، کیهان ما از چه اجزایی تشکیل شده است؟، منشاء ساختارهایی مانند خوشه‌های کهکشانی، منظومه‌ها چیست؟، چه سرنوشتی در انتظار آینده کیهان است؟، توجه کنید می توانم اشاره کنم که اهم شاخه های کلی در ذیل چتر علم کیهان‌شناسی به صورت زیر است:

- **کیهانشناخت کیهان اولیه (شامل نظریه‌هایی مربوط به شرایط اولیه کیهان، تولید ماده و شکست تقارنهای کیهان اولیه)،**
- **کیهان شناخت تابش زمینه کیهانی (شامل ماهیت فیزیکی فوتونهای اولیه و اثرات سرراهی طی طریق فوتون از زمانی که سن عالم در حدود سیصدهزار سال بیشتر نبوده تا زمان حال)،**
- **کیهانشناخت ساختارهای بزرگ مقیاس در رژیم خطی و البته غیرخطی،**
- **امواج گرانشی (منشاء و چگونگی آشکارسازی، دریچه جدیدی برای کاوش کیهان)،**
- **ماده تاریک و انرژی تاریک.**

۲. بر اساس این دسته بندی ها چه داده هایی و از چه منابعی در حال گرد آوری است. به عنوان نمونه ما داده های CMB، داده های سیاه چاله های دوتایی و ... را مثال میزنیم. تا حد امکان این داده ها جهت راهنمایی بیشتر دسته بندی های مطرح شده در بخش قبلی بیان شود.

پاسخ: قبل از اینکه به صراحت به دسته بندی که شما به دنبال آن هستید اشاره کنم توجه شما را به این نکته جلب می کنم که به طور کلی از امواج الکترومغناطیسی و اثرات ثانویه ای که هر یک از فرایندها و رویدادهای فیزیکی از خود به جا می گذارند به عنوان کاوشگر استفاده می شود. اما با توجه به اینکه اخرین سطحی که به طور مستقیم می توان با دریچه امواج الکترومغناطیسی مشاهده و رصد کرد سطح اخرین پراکندگی یا همان CMB است لذا امیدها برای رصد مستقیم رویدادهای ورای CMB از طریق امواج الکترومغناطیسی نقش بر آب است. اما بعد از آشکارسازی گسترده امواج گرانشی این امیدواری بوجود امده است که با ارتقای فناوری‌ها ساخت آشکارسازهای امواج گرانشی، اکنون شرایط جدیدی مهیا شده که از دریچه اعوجاج فضا زمان، به عنوان روبشگر استفاده شود. از سویی دیگر ارتقای توانمندی شبیه‌سازی رایانه‌ای (Illustris، Millenum) نیز باعث شده که داده‌های تولید شده در شبیه‌سازی‌های بزرگ نیز امروزه به عنوان دسته‌دیگری از داده‌های تصنعی مورد عنایت پژوهشگران قرار گرفته است. با این توضیح می‌توان گفت که داده‌های تابش زمینه کیهانی در نسل های آینده رصدگرها با قدرت تفکیک بالا امکان مطالعه موارادی که پیش از این به عنوان سوزن در انبار کاه بود، فراهم می کند. رصدهای مربوط به کاوش در ساختارهای بزرگ مقیاس که در ادامه پروژه‌هایی نظیر SDSS، eBOSS در فضای مختصات (مساحی های تصویر بردار) و فضای انتقال به سرخ (مساحی‌های طیف سنجی) ، داده‌های مربوط به امواج گرانشی تصادفی زمینه و همچنین امواج گرانشی ناشی از رویدادهای اخترفیزیکی کیهان اخیر مانند ادغام سیاهچاله‌ها، داده‌های مربوط به رویدادهای همگرایی گرانشی در سه گروه ریزهمگرایی، همگرایی ضعیف گرانشی (آماری) و همگرایی گرانشی قوی، داده‌های مربوط به رصد

شمع‌های استاندارد یا ابرنواخترها، انفجارهای GRB، طیفهای جذبی و نشری، تابش 21 سانتی‌متری، پروژه کیلومتر مربع، آزمایش‌های رصد مستقیم و غیرمستقیم ماده تاریک، پروژه‌هایی نظیر DES[1]، ZTF[2]، (DES) از جمله مهمترین منابع تولید داده‌های رصدی در حوزه کیهان شناسی به شمار می آیند. اشاره به این نکته نیز ضروری است که برخی از داده‌های رصدی توانایی به نمایش گذاشتن ویژگی‌های زمینه کیهان را دارند و برخی نیز ناظر به اختلالات هستند. به طور کلی مهم‌ترین کمیت‌های مشاهده‌پذیری که تحول زمینه کیهان را نشان می دهند عبارتند از:

چگالی نسبی ماده تاریک	Ω_m
چگالی نسبی نوترینو	Ω_ν
چگالی نسبی ماده باریونی	Ω_b
چگالی نسبی انرژی تاریک	Ω_Λ
چگالی نسبی تابش	Ω_R
کمیت حالت (رابطه بین فشار چگالی انرژی)	w
سن کنونی کیهان	t_0
پارامتر هابل در زمان حال	H_0
پارامتر شتاب منفی در زمان حال	q_0
دمای تابش زمین کیهانی در زمان حال	T_{CMB}

و آنهایی که توصیف‌گر انحراف از اصول کیهان شناختی همگنی و همسانگردی هستند و عبارتند از:

افت‌وخیزچگالی	σ_8
دامنه افت‌وخیزهای تانسوری	A_t
دامنه افت‌وخیزهای اولیه اسکالری	A_s
نمای مقیاسی افت‌وخیزهای اولیه تانسوری	n_t

[1] Dark Energy Survey (https://www.darkenergysurvey.org/)

[2] Zwicky Transient Facility (https://www.ztf.caltech.edu/)

نمای مقیاسی افت‌وخیزهای اولیه اسکالری	n_s
وابستگی نمای مقیاسی افت‌وخیزهای اولیه به مقیاس	$\frac{d\ln n}{d\ln k}$

لذا پروژه‌های رصدی اغلب به نحوی طراحی می‌شوند که بتوانند به صورت مستقیم یا غیرمستقیم اطلاعاتی را در خصوص کمیت‌های پیش‌گفته فراهم کنند.

۳. دو یا سه مورد از تلسکوپ ها و ماهواره های مهمی که اخیرا یا در سال های آینده یا حتی دهه های آینده شروع به کار می کنند را به انتخاب خود به ما معرفی کنید.

پاسخ:

- **پروژه CMB-S4 که نسل آینده رصدگر تابش زمینه کیهانی است**
- **پروژه SKA که بزرگترین تلاش بین‌االمللی برای ساخت تلسکوپ‌های رادیویی**
- **پروژه Advanced Ligo و پروژه LISA که برای رصد امواج گرانشی است**
- **پروژه های مبتنی بر رصد تپ اخترها برای رصد امواج گرانشی**
- **پروژه جیمز وب(JWST) جهت امواج در طیف فرو سرخ**
- **تلسکوپ بزرگ نقشه‌برداری سینوپتیک (LSST) جهت بررسی ابرنواخترها، بررسی دقیقتر ماهیت انرژی تاریک و ...**

۴. در علم کیهان شناسی ما مدل استاندارد کیهان شناسی را داریم. در مورد دلایل پذیرفتن و یا نپذیرفتن آن چند مورد را ذکر کنید و جامعه علمی امروز چه مسیر را در پیش گرفته است.

پاسخ: برای دستیابی به یک معرفت علمی از چارچوبی که در آن از ابزار ریاضی برای مدل‌سازی جهت توصیف رفتار طبعیت، استفاده می‌شود. مدلی در این رهیافت بهتر (نه لزوماً واقعی‌تر) است که به طور کلی ساده‌تر و با اندازه‌گیری‌ها، مشاهدات و رصدها سازگارتر باشد. بنابراین مادامی که پیش‌بینی‌های قابل مشاهده و معیارهایی که بتوان تمایزی برای آنها قایل شد، ارایه نشوند کماکان همه مدلها قابل قبول بوده و هیچ ارجحیتی نسبت به هم ندارند. برای پاسخ به سوالاتی در خصوص طبیعت و ماهیت آن، تحول و احتمالاً پیش‌بینی وضعیت آینده آن، مدلی بنا می‌شود. در این مدل باید مشاهده‌پذیرها و اندازه‌پذیرها تعیین شوند. از سویی دیگر باید با کمک ابزارهای مناسب به رصد و مشاهده پرداخت و مقادیر مشاهده‌پذیرها را تعیین کرد. در چهارچوب مدل کردن داده‌ها یعنی مقایسه نظریه و مشاهده، سازگاری مدل با رصد سنجیده می‌شود. برای بررسی تمام جوانب و ایده‌های جدید نیازمند انجام شبیه‌سازی نیز هستیم که البته این کار خود بازخوردهایی را برای مشاهدات، اندازه‌گیرها و رصدهای آینده و همچنین نظریه‌پردازی به دنبال خواهد داشت. علی‌الاصول یک نظریه موفق کیهان‌شناختی باید تا آنجا که می‌تواند به سؤالاتی از قبیل: الف: مبداء کیهان ما چیست و کجاست؟ ب: کیهان ما از چه اجزایی تشکیل و ساخته شده است؟ ج: اندازه و سن کیهان ما چقدر است؟د: شکل کیهان ما چگونه است؟ه: منشاء ساختارهایی نظیر منظومه‌ها، کهکشان‌ها و خوشه‌هایی که امروزه مشاهده می‌کنیم چیست؟و: آینده عالم ما چگونه است؟ و سایر سؤالاتی از این دست، پاسخی سازگار و عقلانی ارایه دهد. با این مقدمه، مدل استانداردکیهانشناسی با حداقل پارامترهای

آزاد در مقایسه با سایر نظریه‌ها، توانسته، پیش‌بینی‌ها و همچنین سازگاری‌هایی بسیار چشمگیر، از رصدهای کیهان اخیر و کیهان اولیه ارایه دهد. البته با ارتقای دقت اندازه‌گیری، اکنون مواردی از ناهمخوانی به عنوان ناهنجاری پدید آمده است. این شرایط البته در دوره‌های دیگری نیز در حوزه کیهانشناسی و همچنین سایر حوزه‌های علمی نیز پدید آمده و منجر به باز شدن دریچه‌ای جدید برای فهم ما از جهان طبیعت شده است.

بخش دوم

۵. چه مجموعه داده ای را داده بزرگ(Big Data) می نمایم. آیا این مفهوم با گذشت زمان تغییر می کند؟

پاسخ: به نظر می رسد که از جمله اولین مراکزی که عبارت Data Big را بکار برده است NASA در سال ۱۹۹۷ است. کلان داده دارای سه ویژگی مهم است یعنی کلان بودن از حیث اندازه (Volume)، کلان بودن از حیث داشتن آهنگ تولید بسیار چشمگیر (rate Production or Velocity) و کلان بودن از حیث تنوع زیاد (Variaty). گاهی نیز کلان داده را VVV در نظر می‌گیرند. البته در حوزه کیهان‌شناسی ممکن است یک یا دو ویژگی مذکور به طور همزمان برای یک دسته داده که ادعا می‌کنیم جزو کلان داده محسوب می‌شوند، وجود نداشته باشند. از منظر فیزیک نیز می‌توان اضافه کرد که این مفهوم را یک بار دیگر در جایی مانند ترمودینامیک تجربه کرده‌ایم. آنجایی که رفتارهای مقیاس ماکروسکوپیک الزاماً در حد ترمودینامیک اهمیت پیدا می‌کند. در خصوص وابستگی مفهوم کلان داده به زمان، نظر من بر این است که از آنجا که کلان داده خود محصول است و نه معلول لذا، هر گونه تغییری می‌تواند داشته باشد. از این منظر، باید گفت تصور از کلان داده در علم کیهان‌شناسی با آنچیزی که در حدود ۳۰ تا ۴۰ سال پیش وجود داشت با تغییرات بنیادینی مواجه شده است. این تغییرات بنیادین نه صرفاً به دلیل حجم داده بلکه تغییر نگرش مواجه با آنها برای استخراج علم، است. اکنون پارادایم علم داده‌محور نیز بیش از پیش در کیهانشناسی رسوخ پیدا کرده است. این درحالی است که در حدود ۱۰ سال پیش گفتگو از این رویکرد با برچسب گفتمان غیرعلمی مواجه بود. لذا به نظر من مفهوم ان با گذشت زمان برای فیزیکدانان از این جهت عوض شده که نگریستن به داده‌ها فقط به مثابه ابزاری به منظور مقید کردن کمیتهای مدلهای نظری نیست و جستجو گری برای بکاربستن روشهای محاسباتی مناسب برای استخراج معرفت علمی نیز شروع شده است. البته اگر به کلان داده از منظر صرف ماهیت کلان داده توجه شود در آن صورت قول دیگری نیز می‌تواند درست باشد و آن اینکه ویژگی کلان داده به مثابه خواص ماکروسکوپیک سیستم است و تا زمانی که حواس ما تغییر نکند، مفاهیمی که توسط کمیت‌های ماکروسکوپیک حمل می‌شود تغییر نمی‌کند. اما در صورتی که حواس ما تغییر کند، مفهوم کمیت ماکروسکوپیک و در نتیجه مفهوم کلان داده می‌تواند تغییر کند.

۶. چالش های کار با داده های بزرگ چیست؟

پاسخ: چالش‌ها را به دو بخش سخت‌افزاری و نرم‌افزاری تقسیم‌بندی می‌کنم. در بخش سخت‌افزاری همانطور که از اسمش پیداست عمده چالش در انتقال داده‌ها و همچنین محدودیت‌هایی نظیر سروکله زدن با داده توسط سخت‌افزارهای قابل دسترس است و در نتیجه نیاز به صرف زمان زیاد است. اما در بخش نرم‌افزاری موضوع کمی پیچیده‌تر است. ابزارهای داده‌کاهی، تمیز کردن داده‌ها تا روشهای تحلیل و بررسی و الگوریتم‌ها، لازم است تغییرات بنیادی داشته باشد. روشهای مرسوم تحلیل داده‌ها باعث می‌شود که از ظرفیت‌های و پتانسیل‌های نهفته در انها به طور کامل استفاده نشود.

۷. راه حال های این چالش ها چیست؟

پاسخ: در بخش سخت‌افزاری خب جواب واضح است، اما به نظر من در بخش نرم‌افزاری لازم است به ابزارهای و الگوریتم‌های جدید تحلیل آنها مجهز شد.

۸. چند مورد از مجموعه داده های بزرگ در علم نجوم و کیهان شناسی را برای ما به همراه حجم داده بیان کنید.

پاسخ: داده‌های تابش زمینه‌کیهانی که در نتایج منتشر شده توسط ماهواره پلانک با داده‌هایی بالغ بر 50 میلیون داده در نه باند فرکانسی در بخش افت و خیز دمایی و بخش دیگری نیز به قطبش این تابش مرتبط است. در خصوص مساحی هایی ساختارها نیز می توان به مساحی آسمان دیجیتال اسلون (SDSS) اشاره کرد که در هر برونداد از نتایج خود، حجم داده‌ای از مرتبه صد ترابایت ارائه می‌کند. برای مثال در شبه‌سازی‌های ساختارهای بزرگ کیهانی نظیر Millennium ، Illustris، حجم داده‌های خام از مرتبه ۲۵۰ ترابایت است.(در شبیه‌سازی‌ها معمولا حجم داده مطرح نیست، چرا که شبیه‌سازی در سرورهای از قبل تعریف شده مورد استفاده قرار می‌گیرد.)
در پروژه یک میلیارد دلاری کیلومتر مربع (SKA) پیش بینی می شود که بالغ بر 400 گیگابایت داده در هر ثانیه تولید شود این در حالی است که در پروژه شتابدهنده سرن به طور متوسط تولید 25 گیگابایت داده در هر ثانیه انجام شده است.
در جدول زیر نیز حجم داده های چند پروژه معروف را می توانید مشاهده کنید:

Sky Survey Projects	Data Volume
DPOSS (The Palomar Digital Sky Survey)	3 TB
2MASS (The Two Micron All-Sky Survey)	10 TB
GBT (Green Bank Telescope)	20 PB
GALEX (The Galaxy Evolution Explorer)	30 TB
SDSS (The Sloan Digital Sky Survey)	40 TB
SkyMapper Southern Sky Survey	500 TB
PanSTARRS (The Panoramic Survey Telescope and Rapid Response System)	~ 40 PB expected
LSST (The Large Synoptic Survey Telescope)	~ 200 PB expected
SKA (The Square Kilometer Array)	~ 4.6 EB expected

منبع: Cui, Chenzhou, et al. "Astronomy research in big-data era." *Chinese Science Bulletin* 60.5-6 (2015): 445-449.

بخش سوم

۹. آیا دسته بندی خاصی برای تحلیل داده میتوانید بیان کنید؟

پاسخ: از نقطه‌نظر اهداف می توان گفت تحلیل داده‌ها به دسته تحلیل پسینی (Posterior analysis)، استفاده از داده در شبیه‌سازی‌ها و همچنین کاربست داده‌ها در علم داده‌محور، تقسیم بندی می‌شود.

۱۰. خارج از داده های مشاهده(observe) به چه دلیل سراغ شبیه سازی میرویم؟

پاسخ: همانطور که در سوالات قبل اشاره کردم، زمانبر و هزینه‌بر بودن انجام آزمایشها از یک سو و همچنین دغدغه برای تعیین خط‌مشی بهینه برای انجام ازمایشها و احتراز هر چه بیشتر از اتخاذ رویکرد آزمون و خطا از جمله دلایل توجه به انجام شبیه‌سازی‌ها به شمار می‌رود. همچنین انجام شبیه‌سازی می‌تواند بازخوردی برای مدلسازی و نظریه‌پردازی نیز در پی داشته باشد.

۱۱. آیا شبیه سازی ها می‌توانند ما را گمراه کنند؟

دقت کنید که چه موقع می توان گفت که قادر خواهیم بود شبیه‌سازی انجام دهیم؟ به واقع در شبیه‌سازی اغلب طبیعت و محیط اطراف ما منشاء تولید مسئله‌ها هستند. نظریه‌های در بستر مدلها تدوین می‌شوند. روشها توسط الگوریتم‌ها پیاده سازی می‌شوند و نهایتاً برنامه‌های رایانه‌ای بخش آخر تولید یک شبیه‌سازی را تشکیل می‌دهد. با عنایت به این توصیف، به روشنی درمی‌یابیم که در شبیه‌سازی‌های امکان تولید تصنعی فرآیندهایی که در طبیعت رخ می‌دهند وجود دارد. اما ایا هر نتیجه‌ای که در شبیه‌سازی به دست می‌آید، قابل اعتماد است؟ پاسخ این سوال به این نکته ظریف برمی‌گردد که رایانه هرچیزی که ما به آن داده‌ایم اجرا می‌کند و نه آنچیزی را انتظار داریم. لذا حداقل دو منشاء وجود دارد که باعث می‌شود نتایج بدست آمده با قطعیت مورد استناد نباشد. اول اینکه مدلهای نظری که استخراج شده اند ممکن است تمام انچه را که در طبیعت رخ می‌دهد را پوشش ندهد. در نتیجه، پیش‌بینی‌های حاصل از شبیه‌سازی ممکن است دچار خدشه باشد. دوم اینکه الگوریتم‌های بکار گرفته شده اساسا خود دارای محدودیت‌های ذاتی هستند و این محدودیت‌ها ممکن است در برخی اوقات با اهمیت باشند.

۱۲. رابطه علم داده، هوش مصنوعی، یادگیری ماشین و شبکه های عصبی به چه شکل است؟ اگر نیاز هست خیلی خلاصه برای هر کدام مثالی بزنید.

پاسخ: هوش مصنوعی مفهومی نظری با هدف توسعه ماشین برای انجام وظایفی است که به طور معمول برای اجرای آن‌ها نیاز به هوش انسانی است. با این تعریف، هوش مصنوعی مفهومی کلی است که در خیلی از حوزه‌ها کاربرد داشته و نیازمندی‌های فراوانی از قبیل شبکه، استخراج داده، داده‌های حجیم و غیره برای توسعه آن وجود دارد. یادگیری ماشین بخش مهمی از فرآیند توسعه در هوش مصنوعی است. شبکه‌های عصبی یکی از روش‌های مرسوم در یادگیری ماشین است. روش‌های دیگری مانند روش بردار پشتیبان، درخت تصیمم و غیره نیز در یادگیری ماشین کاربرد دارند و در برخی موارد حتی استفاده از آن‌ها بهینه‌تر از استفاده از شبکه عصبی است. از طرفی در علم داده در واقع تاکید ما بر استخراج دانش از داده است. این موضوع به صورت ویژه وقتی که با داده‌های بزرگ یا داده‌های با خطای زیاد، داده‌های غیرساختارمند و مانند آن مواجه هستیم اهمیت پیدا می‌کند. یکی از ابزارهایی که برای جستجو در داده‌ها و استخراج علم و دانش بکار گرفته می‌شود الگوریتم‌های توسعه یافته در حوزه یادگیری ماشین است. در این حالت ماشین وظیفه دارد تا بر اساس در خواست ما، روابط معنی‌دار را از داده‌های پیچیده استخراج کند. ما درخواست خود را بر روی الگوریتم‌هایی یادگیری ماشین اعمال می‌کنیم. برای مثال دسته‌بندی داده‌ها طبق ویژگی‌هایی که برای ما اهمیت دارد می‌تواند یک درخواست باشد. یا تخمین نتیجه در صورتی که ویژگی‌های مشخصی در داده به وقوع بپیوندند. در نهایت می توان گفت که شبکه های عصبی زیر مجموعه ی یادگیری ماشین و یادگیری ماشین نیز زیر مجموعه هوش مصنوعی قرار می گیرد.

۱۳. هنگام بررسی داده ها مثلا برای مسائلی مانند طبقه بندی ابزارهای آماری و ابزارهای یادگیری ماشین و یادگیری عمیق را در دسترس داریم. چالش های استفاده از هر کدام را بیان کنید.

پاسخ: اول از همه باید تا آنجا که امکان دارد نسبت به داده‌ها شهود کافی یافت. بر اساس تجربه شخصی خودم، ابتدا سعی میکنم که تا آنجا که امکان دارد داده‌ها را به تصویر تبدیل کنم تا با آنها ارتباط عینی ایجاد شود. در مرحله بعد باید به این سوال پاسخ داد که آیا صرفا طبقه‌بندی برای کار ما کفایت می‌کند؟ و مسأله را باید با نگاه صرف نتیجه‌محوری نگریست؟ اگر پاسخ این گونه سوالات مثبت است خب روشی باید اتخاذ شود که

هم با دقت و هم با سرعت بالا بتواند به نحو قابل اعتمادی طبقه‌بندی را انجام دهد. اگر دسته داده‌های مورد بررسی را بتوان به عنوان یک کلان داده شناخت، استفاده از روشهای یادگیری ماشین به مراتب می‌تواند اثربخش باشد. یادگیری عمیق نیز الزاما به دسته‌داده‌های بسیار بزرگ برای فرآیند یادگیری نیاز است. هرچند که در تحقیقاتی نشان داده شده است که با اصلاح الگوریتم می‌توان تا حدی بر این مشکل چیره شد. در بکاربردن ابزارهای آماری، به نظرم مهمترین چالش اماده سازی داده و مولفه‌هایی که در داده وجود دارد می‌تواند گمراه کننده باشد. برای مثال فرض کنید می‌خواهیم ویژگی‌های غیربدیهی را از سری داده‌ای استخراج کنیم. اگر این داده با یک روند تناوبی همراه باشد و این روند تناوبی به خوبی شناخته شده باشد در آن صورت امکان مخفی شدن ویژگی غیربدیهی در حضور این روند تناوبی به شدت وجود دارد. همچنین اغلب روشهای آماری ابداع شده با فرض پیوستگی داده‌ها توسعه یافته‌اند، این درحالی است که در آزمایشها و رصدها داده‌های گسسته و اغلب نامنظم ثبت می‌شوند، لذا باید این نوع حالت‌ها مورد توجه قرار گیرد.

۱۴. در بحث کاربرد یادگیری ماشین و یادگیری عمیق چه کاربردها یا مقالات مهمی در زمینه کیهان شناسی و اختر فیزیک صورت گرفته است و نقش جامعه علمی ایران چگونه بوده؟

در حوزه بین‌المللی در ابتدای گسترش ادبیات یادگیری ماشین، بسیاری از کیهانشناسان با موضع‌گیری اغلب منفی در قبال این رویکرد و اتخاذ این استراتژی که یادگیری ماشین کمکی به افزایش بصیرت علمی ما نمی‌کند و یا در بهترین وضعیت اذعان داشتند که اطلاعات مستخرج از یادگیری ماشین چیزی بیشتر از روشهای مبتنی بر مدلسازی داده‌های بیزی نیست. اما گسترش کاربست این الگوریتم‌ها و مواجه با دستاوردهای شگرف آن هم از حیث دقت و سرعت و کاربست آمارهای اهمیت بردارهای ویژگی که پیش از آن به سادگی فراهم نبود، کمکم جامعه علمی کیهانشناسی به این موضوع رغبت نشان داد و به سرعت استفاده از این رویکرد فراگیر شد. از جمله مقالات مهم در این حوزه می توان به

Carleo, Giuseppe, et al. "Machine learning and the physical sciences." Reviews of Modern Physics 91.4 (2019): 045002

Ball, Nicholas M., and Robert J. Brunner. "Data mining and machine learning in astronomy." International Journal of Modern Physics D 19.07 (2010): 1049-1106.

Ball, Nicholas M., and Robert J. Brunner. "Data mining and machine learning in astronomy." International Journal of Modern Physics D 19.07 (2010): 1049-1106.

Scargle, Jeffrey D., Kamal M. Ali, and Ashok N. Srivastava. Advances in machine learning and data mining for astronomy. Ed. Michael J. Way. Boca Raton, FL: CRC Press, 2012.

Cuoco, Elena, et al. "Enhancing gravitational-wave science with machine learning." Machine Learning: Science and Technology 2.1 (2020): 011002.

Vafaei Sadr, A., et al. "Cosmic string detection with tree-based machine learning." Monthly Notices of the Royal Astronomical Society 478.1 (2018): 1132-1140.

در ایران نیز همراه با پیشرفت‌هایی که در دنیا انجام شده، نیز فعالیت‌های پژوهشی متعددی با تکیه بر کاربست‌های یادگیری ماشین در کیهان‌شناسی و اخترفیزیک انجام شده و در حال انجام است. برای مثال می‌توان به موضوعات زیر اشاره کرد:

- جستجو برای یافتن ردپای نواقص توپولوژیک کیهان اولیه

[1] Vafaei Sadr, A., et al. "Cosmic string detection with tree-based machine learning." Monthly Notices of the Royal Astronomical Society 478.1 (2018): 1132-1140.

[2] M Torki, H Hajizadeh, M Farhang, A Vafaei Sadr, S M S Movahed, Planck limits on cosmic string tension using machine learning, Monthly Notices of the Royal Astronomical Society, Volume 509, Issue 2, January 2022, Pages 2169–2179

[3] Movahed, Seyed Mohammad Sadegh, Alireza Vafaei Sadr, and Marzieh Farhang. "Searching cosmic strings network in the CMB." International Journal of Modern Physics D 28.16 (2019): 2040006.

- جستجو برای یافتن ردپای امواج گرانشی تصادفی زمینه

[1] پویان گودرزی، پایان نامه دوره کارشناسی ارشد، دانشکده فیزیک دانشگاه شهید بهشتی، 1400.

سخن پایانی با دانشجویان علاقه مند به این حوزه؟

به نظرم من ما به عنوان فیزیک‌پیشه باید به توجه به مزیت‌های نسبی که در مقایسه با سایر محققان علوم کامپیوتر، علوم ریاضی، مهندسی کامپیوتر، داریم با موضوع یادگیری ماشین و تحلیل داده‌های بزرگ، مواجه شویم. به طور مشخص استفاده از بردار ویژگی (Feature vector)، و سهم اثر هر یک در رسیدن به نتیجه می تواند برای فیزیک پیشگان مهم و حیاتی باشد. لذا می توانیم بر اساس دانشی که از معیارها و ابزارهای آماری و همچنین جبر عملگری (کوانتوم) داریم، الگوریتم‌های بهینه‌تری را ساخت. به طور مشخص موضوع یادگیری ماشین بر اساس رویکردهای کوانتومی، می تواند در آینده نزدیک نقش مهمی در الگوریتمهای رایانه ای ایفا می کند. تحیلی توپولوژیک داده‌ها، خصوصاً انجا که خواص سراسری داده‌های برای ما اهمیت دارد این روش می تواند مفید واقع شود.

در سایت Kaggle نیز به طور مستمر چالشهایی مبتنی بر رهیافت‌های محاسباتی و داده‌های کلان با توجه به کیهانشناسی و اخترفیزیک ارایه می‌شود که شرکت در آنها تجربه کار با داده‌های تصنعی در این خصوص را به ارمغان خواهد داشت.

مراجع و وبگاه‌های زیر زیر می‌توانند برای موضوعات پیش‌گفته مفید واقع شوند

[1] Ntampaka, Michelle, et al. "The role of machine learning in the next decade of cosmology." arXiv preprint arXiv:1902.10159 (2019).

[2] Ishida, Emille EO. "Machine learning and the future of supernova cosmology." Nature Astronomy 3.8 (2019): 680-682.

[3] https://cosmostatistics-initiative.org/(وبگاهی برای معرفی رویکردهای داده‌محور در حوزه کیهانشناسی و اخترفیزیک)

[4] Carleo, Giuseppe, et al. "Machine learning and the physical sciences." Reviews of Modern Physics 91.4 (2019): 045002.

[5] Cui, Chenzhou, et al. "Astronomy research in big-data era." Chinese Science Bulletin 60.5-6 (2015): 445-449.

با تشکر

موحد

۹ دی ماه ۱۴۰۰