

کاربرد هوش مصنوعی در پیش‌بینی واکنش‌ها و طراحی مولکول‌های جدید

مقدمه

در دهه‌های اخیر، پیشرفت فناوری اطلاعات و افزایش توان پردازشی رایانه‌ها تحول قابل توجهی در علوم پایه ایجاد کرده است. شیمی نیز از این تحول تأثیر پذیرفته و روش‌های محاسباتی به بخش مهمی از پژوهش‌های مدرن در زمینه پیش‌بینی خواص مواد، طراحی ترکیبات، بررسی واکنش‌ها و کشف مواد جدید تبدیل شده‌اند. در میان این روش‌ها، هوش مصنوعی، یادگیری ماشین و به‌ویژه یادگیری عمیق جایگاه ویژه‌ای پیدا کرده‌اند. توانایی این فناوری‌ها در پردازش حجم عظیمی از داده‌ها و استخراج الگوهای پیچیده، فرصت‌های جدیدی برای حل مسائل دشوار شیمیایی فراهم کرده است.

یکی از مهم‌ترین کاربردهای هوش مصنوعی در شیمی، پیش‌بینی واکنش‌های شیمیایی است. در یک واکنش آلی، نتیجه واکنش تنها به ساختار مواد اولیه وابسته نیست، بلکه عواملی مانند نوع حلال، دما، فشار، کاتالیزگر، زمان واکنش، غلظت و نسبت مواد نیز می‌توانند بر محصول نهایی و بازده تأثیر بگذارند. بررسی تجربی همه حالت‌های ممکن از نظر زمان و هزینه عملاً امکان‌پذیر نیست. از این رو، مدل‌های هوش مصنوعی می‌توانند با یادگیری از واکنش‌های انجام‌شده، واکنش‌های جدید را پیش‌بینی و گزینه‌های مناسب‌تر را برای آزمایش پیشنهاد کنند.

کاربرد مهم دیگر هوش مصنوعی، طراحی مولکول‌های جدید است. فضای شیمیایی بالقوه بسیار گسترده است و تعداد ترکیبات ممکن از ظرفیت روش‌های تجربی برای بررسی مستقیم آن‌ها بسیار بیشتر است. بنابراین، یکی از مسائل اصلی در شیمی مدرن، پیدا کردن روش‌هایی برای جست‌وجوی هوشمندانه این فضای عظیم است. مدل‌های مولد هوش مصنوعی می‌توانند با توجه به خواص مورد نظر، ساختارهای جدیدی تولید کنند و سپس مدل‌های پیش‌بینی‌کننده، ویژگی‌های این ساختارها را ارزیابی کنند.

این فناوری در شیمی دارویی اهمیت ویژه‌ای دارد، زیرا یک مولکول دارویی مناسب باید مجموعه‌ای از ویژگی‌های مختلف مانند فعالیت زیستی، حلالیت، پایداری، سمیت پایین و قابلیت سنتز مناسب را داشته باشد. بنابراین، هوش مصنوعی می‌تواند به جای بررسی تصادفی تعداد بسیار زیادی مولکول، گزینه‌های امیدوارکننده‌تر را شناسایی کند.

با وجود این پیشرفت‌ها، هوش مصنوعی جایگزین دانش شیمیایی و آزمایش تجربی نیست. کیفیت داده‌های آموزشی، سوگیری موجود در پایگاه‌های داده، محدودیت تعمیم مدل‌ها و تفاوت میان شرایط واقعی آزمایشگاه و داده‌های محاسباتی از مهم‌ترین چالش‌های این حوزه هستند. از این رو، بهترین رویکرد، ترکیب توانایی محاسباتی هوش مصنوعی با دانش و تجربه شیمیدان است.

۱. هوش مصنوعی و یادگیری ماشین در شیمی

هوش مصنوعی به مجموعه‌ای از روش‌های محاسباتی گفته می‌شود که به سامانه‌ها امکان می‌دهند از داده‌ها الگو استخراج کرده و بر اساس آن‌ها پیش‌بینی یا تصمیم‌گیری کنند. یکی از مهم‌ترین شاخه‌های هوش مصنوعی، یادگیری ماشین است.

در روش‌های سنتی، برای حل یک مسئله معمولاً مجموعه‌ای از قواعد توسط متخصص تعریف می‌شود. برای مثال، در یک سیستم مبتنی بر قواعد، شیمیدان می‌تواند مشخص کند که یک گروه عاملی در حضور یک واکنشگر خاص چه تغییری خواهد کرد. این روش در واکنش‌های شناخته‌شده مفید است، اما شیمی آلی بسیار پیچیده‌تر از آن است که بتوان تمام حالت‌های ممکن را با مجموعه‌ای محدود از قواعد توصیف کرد.

در یادگیری ماشین، الگوریتم به جای دریافت تمام قواعد از پیش تعیین‌شده، از نمونه‌های موجود یاد می‌گیرد. اگر مجموعه‌ای شامل هزاران یا میلیون‌ها واکنش شیمیایی در اختیار مدل قرار گیرد، الگوریتم می‌تواند روابط آماری موجود میان واکنش‌دهنده‌ها، شرایط واکنش و محصولات را استخراج کند.

یادگیری عمیق نیز زیرمجموعه‌ای از یادگیری ماشین است که از شبکه‌های عصبی چندلایه استفاده می‌کند. این شبکه‌ها می‌توانند روابط بسیار پیچیده میان ورودی و خروجی را یاد بگیرند و به همین دلیل در سال‌های اخیر در مسائل شیمیایی کاربرد گسترده‌ای پیدا کرده‌اند.

به‌طور کلی، مهم‌ترین کاربردهای هوش مصنوعی در شیمی شامل پیش‌بینی خواص مولکولی، پیش‌بینی محصولات واکنش، طراحی شرایط واکنش، رتروسنسز، طراحی مولکولی *de novo*، بهینه‌سازی ساختار و کشف مواد جدید است.

۲. نمایش مولکول‌ها برای مدل‌های هوش مصنوعی

پیش از آنکه یک مدل هوش مصنوعی بتواند مولکول یا واکنش شیمیایی را تحلیل کند، ساختار شیمیایی باید به فرم قابل پردازش برای رایانه تبدیل شود. یکی از روش‌های متداول، نمایش مولکول با استفاده از SMILES است.

SMILES یک نمایش خطی از ساختار مولکولی است که اتم‌ها و پیوندهای آن‌ها را به‌صورت یک رشته متنی نشان می‌دهد. استفاده از SMILES این امکان را فراهم می‌کند که مولکول‌ها مانند توالی‌های متنی به مدل‌هایی مانند Transformer داده شوند.

روش مهم دیگر، نمایش مولکول به شکل گراف است. در گراف مولکولی، اتم‌ها به‌عنوان گره و پیوندها به‌عنوان یال تعریف می‌شوند. برای هر اتم می‌توان ویژگی‌هایی مانند نوع عنصر، بار، هیبریداسیون و تعداد پیوندها را مشخص کرد. برای پیوندها نیز می‌توان نوع پیوند، آروماتیک بودن و اطلاعات استریوشیمیایی را وارد مدل کرد.

این نوع نمایش پایه اصلی شبکه‌های عصبی گرافی یا GNN است. از آنجا که ساختار مولکولی ذاتاً یک گراف است، این روش می‌تواند اطلاعات توپولوژیکی را به شکل مناسبی در اختیار مدل قرار دهد.

علاوه بر این دو روش، می‌توان مولکول‌ها را با استفاده از Fingerprintها، بردارهای ویژگی یا نمایش‌های نهفته نیز توصیف کرد. انتخاب نمایش مناسب اهمیت زیادی دارد، زیرا اطلاعاتی که مدل می‌تواند یاد بگیرد تا حد زیادی به نحوه نمایش داده وابسته است.

۳. Transformer و کاربرد آن در پیش‌بینی واکنش

Transformer یکی از مهم‌ترین معماری‌های یادگیری عمیق در سال‌های اخیر است. این معماری در ابتدا برای پردازش زبان طبیعی توسعه یافت، اما قابلیت آن در مدل‌سازی وابستگی میان عناصر یک توالی باعث شد به حوزه شیمی نیز وارد شود.

مبنای اصلی Transformer سازوکار Attention یا توجه است. مدل با استفاده از Attention می‌تواند اهمیت بخش‌های مختلف ورودی را هنگام تولید خروجی تعیین کند. برخلاف شبکه‌های بازگشتی که داده‌ها را به‌صورت ترتیبی پردازش می‌کنند، Transformer قادر است روابط میان قسمت‌های مختلف یک توالی را به‌صورت موازی بررسی کند.

در شیمی می‌توان یک واکنش را به‌صورت یک مسئله Sequence-to-Sequence در نظر گرفت. در این حالت، نمایش SMILES مواد اولیه و واکنشگرها به مدل داده می‌شود و مدل تلاش می‌کند رشته SMILES محصول را تولید کند.

Molecular Transformer یکی از شناخته‌شده‌ترین نمونه‌های این کاربرد است Schwaller و همکاران نشان دادند که این معماری می‌تواند برای پیش‌بینی واکنش‌های شیمیایی استفاده شود و در یک مجموعه داده متداول، دقت Top-1 حدود ۹۰٫۴ درصد و دقت Top-2 حدود ۹۳٫۷ درصد به دست آورد. (Schwaller et al., ۲۰۱۹)

اهمیت این نتیجه در آن است که مدل توانست بسیاری از الگوهای واکنش را مستقیماً از داده‌ها یاد بگیرد. بنابراین، نیازی نبود تمام قواعد مربوط به واکنش‌های مختلف به‌صورت دستی در سیستم تعریف شوند.

Transformer علاوه بر پیش‌بینی محصول می‌تواند در رتروسنسز نیز مورد استفاده قرار گیرد. در این حالت، به جای پیش‌بینی محصول از واکنش‌دهنده‌ها، مدل از مولکول هدف شروع کرده و پیش‌سازهای احتمالی را پیشنهاد می‌دهد.

یکی از محدودیت‌های Transformer این است که SMILES یک نمایش متنی از مولکول است و تمام ویژگی‌های فضایی و سه‌بعدی ساختار را به‌صورت مستقیم نشان نمی‌دهد. همچنین یک مولکول می‌تواند چندین نمایش SMILES معتبر داشته باشد. بنابراین، آموزش مدل با نمایش‌های متنوع می‌تواند به افزایش مقاومت مدل کمک کند.

۴. شبکه‌های عصبی گرافی یا GNN

GNN‌ها از مهم‌ترین روش‌های یادگیری عمیق برای پردازش ساختار مولکولی هستند. دلیل اصلی این موضوع آن است که مولکول‌ها به‌طور طبیعی ساختار گرافی دارند.

در یک GNN، هر اتم به‌عنوان یک گره و هر پیوند به‌عنوان یک یال در نظر گرفته می‌شود. سپس مدل از طریق فرایندی به نام Message Passing اطلاعات میان اتم‌های مجاور را منتقل می‌کند.

در ساده‌ترین حالت، نمایش یک اتم در لایه بعدی تابعی از ویژگی‌های خودش و ویژگی‌های اتم‌های همسایه است. با تکرار این فرایند، مدل می‌تواند اطلاعات مربوط به محیط شیمیایی گسترده‌تر هر اتم را استخراج کند.

برای مثال، اگر هدف پیش‌بینی حالیت باشد، مدل می‌تواند اطلاعات مربوط به نوع اتم‌ها، پیوندها، قطبیت و ساختار کلی مولکول را به یک نمایش عددی تبدیل کند و سپس از آن برای پیش‌بینی استفاده کند.

GNN‌ها در پیش‌بینی خواص فیزیکی و شیمیایی، سمیت، فعالیت زیستی، انرژی و واکنش‌پذیری کاربرد دارند. همچنین می‌توان از آن‌ها برای تشخیص مرکز واکنش استفاده کرد.

یکی از محدودیت‌های GNN‌ها این است که اطلاعات سه‌بعدی مولکول به‌طور کامل در گراف ساده قابل مشاهده نیست. دو مولکول یا دو کانفورماسیون می‌توانند توپولوژی یکسان اما هندسه فضایی متفاوت داشته باشند. از این رو، مدل‌های سه‌بعدی و Equivariant GNN‌ها توسعه یافته‌اند تا اطلاعات هندسی و تقارن‌های فضایی را نیز در نظر بگیرند.

۵. شبکه‌های مولد تخصصی یا GAN

GAN یا Generative Adversarial Network یکی از معماری‌های مهم در حوزه مدل‌های مولد است. این معماری از دو شبکه تشکیل می‌شود: Generator و Discriminator.

Generator وظیفه تولید نمونه‌های جدید را بر عهده دارد و Discriminator تلاش می‌کند نمونه‌های واقعی را از نمونه‌های تولیدشده تشخیص دهد. آموزش این دو شبکه به صورت رقابتی انجام می‌شود.

در طراحی مولکولی، Generator می‌تواند ساختارهای جدید ایجاد کند و Discriminator بررسی کند که آیا این ساختارها به توزیع مولکول‌های واقعی شباهت دارند یا خیر.

یکی از مشکلات اساسی استفاده از GAN در شیمی، گسسته بودن ساختار مولکولی است. در تصویر می‌توان مقادیر پیکسل‌ها را به صورت پیوسته تغییر داد، اما در یک مولکول، اضافه یا حذف کردن یک اتم یا تغییر یک پیوند می‌تواند باعث ایجاد ساختاری از نظر شیمیایی نامعتبر شود.

برای حل این مشکل، معماری‌هایی برای تولید گراف‌های مولکولی توسعه یافته‌اند. در این روش‌ها، Generator می‌تواند گراف‌هایی از اتم‌ها و پیوندها تولید کند.

یکی از مشکلات شناخته‌شده GAN ها Mode Collapse است. در این حالت، مدل به جای تولید تنوع گسترده‌ای از مولکول‌ها، تعداد محدودی ساختار مشابه تولید می‌کند. این موضوع در طراحی مولکولی مشکل مهمی محسوب می‌شود، زیرا یکی از اهداف اصلی مدل مولد، جست‌وجوی بخش وسیعی از فضای شیمیایی است.

با وجود این محدودیت‌ها، GAN ها نقش مهمی در توسعه طراحی مولکولی مولد داشته‌اند و زمینه را برای مدل‌های مولد پیشرفته‌تر فراهم کرده‌اند.

۶. خودرمزگذار متغیر یا VAE

Variational Autoencoder یا VAE روشی دیگر برای تولید مولکول‌های جدید است. در VAE، مولکول ابتدا توسط Encoder به فضای نهفته منتقل می‌شود و سپس Decoder تلاش می‌کند ساختار مولکولی را از این نمایش بازسازی کند.

تفاوت اصلی VAE با Autoencoder معمولی در این است که فضای نهفته به صورت یک توزیع احتمالی مدل می‌شود. بنابراین می‌توان نقاط جدیدی از فضای نهفته انتخاب کرد و با استفاده از Decoder مولکول‌های جدید ساخت.

این ویژگی برای بهینه‌سازی مولکولی اهمیت زیادی دارد. فرض کنیم تعدادی مولکول دارای خواص مطلوب هستند VAE. می‌تواند ساختار آن‌ها را در فضای نهفته قرار دهد و سپس الگوریتم می‌تواند در این فضا به دنبال نواحی مناسب‌تر بگردد.

در طراحی دارو، می‌توان این فضای نهفته را با یک مدل پیش‌بینی‌کننده ترکیب کرد. مدل پیش‌بینی‌کننده می‌تواند فعالیت زیستی یا سایر خواص مولکول را تخمین بزند و سپس الگوریتم بهینه‌سازی تلاش کند نقاطی را پیدا کند که عملکرد مطلوب‌تری دارند.

مشکل اصلی VAE نیز تولید ساختارهای شیمیایی نامعتبر است. اگر معماری مدل نتواند محدودیت‌های شیمیایی را به‌درستی لحاظ کند، ممکن است مولکول‌هایی تولید شوند که از نظر ظرفیت پیوندی یا ساختار شیمیایی امکان‌پذیر نیستند.

به همین دلیل، VAE‌های مولکولی معمولاً با نمایش‌های گرافی، محدودیت‌های شیمیایی و روش‌های اعتبارسنجی ساختار ترکیب می‌شوند.

۷. مدل‌های Diffusion در طراحی مولکولی

مدل‌های Diffusion یا انتشار از جدیدترین خانواده‌های مدل‌های مولد هستند و در سال‌های اخیر توجه زیادی در طراحی مولکولی به خود جلب کرده‌اند.

در فرایند انتشار، ابتدا مقدار مشخصی نویز به داده اصلی اضافه می‌شود و این فرایند در چندین مرحله ادامه پیدا می‌کند تا داده اولیه به توزیعی تقریباً تصادفی تبدیل شود. سپس مدل یاد می‌گیرد این فرایند را معکوس کند؛ یعنی از یک داده نویزی شروع کرده و مرحله به مرحله ساختار منظم را بازسازی کند.

به صورت مفهومی، فرایند به دو بخش تقسیم می‌شود:

مولکول واقعی $\leftarrow$ افزودن تدریجی نویز $\leftarrow$ توزیع نویزی

و در جهت معکوس:

توزیع نویزی $\leftarrow$ حذف تدریجی نویز $\leftarrow$ مولکول جدید

مدل یاد می‌گیرد که در هر مرحله چگونه نویز را حذف کند. در نهایت می‌تواند از یک نمونه تصادفی، ساختار مولکولی جدید تولید کند.

یکی از مزایای مهم Diffusion Models قابلیت تولید نمونه‌های متنوع و امکان شرطی‌سازی فرایند تولید است. برای مثال، می‌توان مدل را به سمت تولید مولکول‌هایی با ویژگی مشخص هدایت کرد.

این روش در طراحی سه‌بعدی مولکول اهمیت زیادی دارد. مدل‌های سه‌بعدی می‌توانند مختصات فضایی اتم‌ها را نیز در نظر بگیرند و به همین دلیل برای طراحی لیگاندها و بررسی برهم‌کنش‌های مولکولی مناسب‌تر هستند.

مدل‌های انتشار همچنین می‌توانند برای تولید ساختارهای کریستالی و طراحی مواد جدید مورد استفاده قرار گیرند.

با این حال، این مدل‌ها از نظر محاسباتی پیچیده هستند و فرایند تولید معمولاً شامل مراحل متعددی است. همچنین، تولید یک ساختار معتبر از نظر مدل لزوماً به معنی امکان‌پذیر بودن سنتز آن نیست.

۸. مقایسه روش‌های هوش مصنوعی

هر یک از معماری‌های ذکرشده نقاط قوت و ضعف خاص خود را دارند. Transformer، به‌خصوص برای داده‌های توالی‌محور مانند SMILES و پیش‌بینی واکنش‌ها مناسب است. GNN، برای داده‌هایی که ساختار گراف‌ی مولکول اهمیت دارد، انتخاب طبیعی‌تری محسوب می‌شود.

GAN و VAE عمدتاً در نسل‌های اولیه طراحی مولکولی مولد اهمیت زیادی داشتند. VAE، به دلیل ایجاد فضای نهفته قابل جست‌وجو، برای بهینه‌سازی ساختاری مناسب است، در حالی که GAN بر تولید نمونه‌هایی مشابه توزیع داده‌های واقعی تمرکز دارد.

Diffusion Models در سال‌های اخیر به دلیل قابلیت تولید ساختارهای پیچیده و سه‌بعدی مورد توجه قرار گرفته‌اند.

در عمل، این معماری‌ها الزاماً جایگزین یکدیگر نیستند. می‌توان از آن‌ها به‌صورت ترکیبی استفاده کرد. برای مثال، یک سیستم می‌تواند از GNN برای استخراج ویژگی‌های ساختاری مولکول، Transformer برای پردازش اطلاعات واکنش و Diffusion Model برای تولید ساختارهای جدید استفاده کند.

۹. پیش‌بینی شرایط واکنش و بازده

پیش‌بینی محصول تنها یک بخش از مسئله واکنش شیمیایی است. حتی اگر محصول صحیح پیش‌بینی شود، هنوز باید مشخص شود که واکنش تحت چه شرایطی انجام خواهد شد.

دما، فشار، نوع حلال، غلظت، مقدار کاتالیزگر، زمان و سرعت افزودن مواد می‌توانند نتیجه واکنش را تغییر دهند. یادگیری ماشین می‌تواند با استفاده از داده‌های تجربی موجود، ارتباط میان شرایط واکنش و بازده را مدل‌سازی کند.

یکی از روش‌های مهم برای این کار، بهینه‌سازی بیزی یا Bayesian Optimization است. در این رویکرد، مدل بر اساس نتایج آزمایش‌های قبلی، شرایطی را انتخاب می‌کند که احتمال می‌رود بازده را افزایش دهند یا اطلاعات جدید و ارزشمندی برای مدل فراهم کنند.

این روش می‌تواند تعداد آزمایش‌های لازم برای یافتن شرایط بهینه را کاهش دهد.

با این حال، بازده واکنش متغیری پیچیده است. نحوه انجام آزمایش، خلوص مواد، تجهیزات و روش جداسازی می‌تواند روی مقدار گزارش‌شده بازده اثر بگذارد. بنابراین، داده‌های بازده ممکن است دارای عدم قطعیت قابل توجهی باشند.

۱۰. رتروسنتز و برنامه‌ریزی مسیر سنتزی

رتروسنتز یکی از کاربردهای مهم هوش مصنوعی در شیمی آلی است. در این روش، از مولکول هدف شروع شده و ساختار آن به پیش‌سازهای ساده‌تر تقسیم می‌شود.

هوش مصنوعی می‌تواند پیوندهایی را که احتمالاً در مسیر سنتزی مناسب هستند شناسایی کند و پیش‌سازهای احتمالی را پیشنهاد دهد.

اگر چندین مسیر مختلف وجود داشته باشد، سیستم می‌تواند آن‌ها را بر اساس معیارهایی مانند تعداد مراحل، احتمال موفقیت، در دسترس بودن مواد اولیه و پیچیدگی واکنش رتبه‌بندی کند.

نکته مهم این است که کوتاه‌ترین مسیر همیشه بهترین مسیر نیست. یک مسیر ممکن است تعداد مراحل کمی داشته باشد، اما نیازمند مواد اولیه گران یا شرایط بسیار حساس باشد. از این رو، سیستم‌های هوش مصنوعی باید معیارهای عملیاتی، اقتصادی و زیست‌محیطی را نیز در نظر بگیرند.

این موضوع در تولید ترکیبات دارویی و مواد شیمیایی با ارزش بالا اهمیت ویژه‌ای دارد.

۱۱. طراحی *de novo* مولکول‌های جدید

طراحی *de novo* به معنای ایجاد ساختارهای مولکولی جدید بدون محدود شدن به ترکیبات موجود است. در این روش، مدل مولد تلاش می‌کند ساختارهایی تولید کند که خواص مورد نظر را داشته باشند.

برای مثال، در طراحی یک ترکیب دارویی ممکن است هدف تولید مولکولی با فعالیت زیستی بالا، حلالیت مناسب، سمیت پایین و قابلیت سنتز مناسب باشد.

این مسئله را می‌توان به شکل یک مسئله بهینه‌سازی چندهدفه تعریف کرد. یک مولکول ممکن است در یک ویژگی عملکرد بسیار خوبی داشته باشد اما در ویژگی دیگری ضعیف باشد.

برای مثال، افزایش آب‌گریزی یک مولکول ممکن است برهم‌کنش آن با یک هدف زیستی را افزایش دهد، اما هم‌زمان حلالیت آن در آب را کاهش دهد. بنابراین طراحی مولکولی نیازمند ایجاد تعادل میان چندین ویژگی است.

مدل‌های مولد جدید تلاش می‌کنند این محدودیت‌ها را به‌صورت هم‌زمان در فرایند تولید وارد کنند. (Du et al., ۲۰۲۴)

۱۲. کاربرد هوش مصنوعی در کشف دارو

شیمی دارویی یکی از حوزه‌هایی است که هوش مصنوعی در آن کاربرد گسترده‌ای دارد. کشف دارو فرایندی پیچیده است که شامل شناسایی هدف، طراحی ترکیبات اولیه، غربالگری، بهینه‌سازی ساختار و ارزیابی ویژگی‌های دارویی است.

هوش مصنوعی می‌تواند در بسیاری از این مراحل نقش داشته باشد. برای مثال، مدل‌های پیش‌بینی‌کننده می‌توانند ویژگی‌هایی مانند حلالیت، چربی‌دوستی، نفوذپذیری و سمیت احتمالی را تخمین بزنند.

سپس مدل مولد می‌تواند ساختارهایی را پیشنهاد کند که احتمالاً در محدوده مطلوب این ویژگی‌ها قرار دارند.

این روش باعث می‌شود تعداد مولکول‌هایی که باید در آزمایشگاه سنتز و بررسی شوند کاهش پیدا کند.

با این حال، فعالیت زیستی یک مولکول تنها تابع ساختار آن نیست. محیط زیستی، متابولیسم، اتصال به پروتئین‌های مختلف و عوامل فیزیولوژیک می‌توانند رفتار واقعی ترکیب را تغییر دهند. بنابراین، پیش‌بینی محاسباتی باید با آزمون‌های تجربی تکمیل شود.

۱۳. اهمیت داده‌های شیمیایی

داده را می‌توان زیرساخت اصلی هوش مصنوعی در شیمی دانست. حتی پیچیده‌ترین مدل‌ها بدون داده مناسب نمی‌توانند عملکرد قابل اعتمادی داشته باشند.

داده‌های شیمیایی از مقالات، پتنت‌ها، پایگاه‌های داده واکنش، نتایج آزمایشگاهی و محاسبات نظری به دست می‌آیند. با این حال، این داده‌ها ممکن است ناقص یا دارای خطا باشند.

برای مثال، شرایط واکنش ممکن است به‌طور کامل گزارش نشده باشد یا یک ماده در منابع مختلف با نام‌ها و فرمت‌های متفاوت ثبت شده باشد.

یکی دیگر از مشکلات، سوگیری انتشار است. واکنش‌های موفق بیشتر از واکنش‌های ناموفق در ادبیات علمی ثبت می‌شوند. در نتیجه، مجموعه داده ممکن است تصویری غیرواقعی از فضای واکنش‌ها ایجاد کند.

به همین دلیل، پیش‌پردازش داده، استانداردسازی ساختارها، حذف داده‌های تکراری، شناسایی داده‌های غیرقابل اعتماد و ثبت کامل شرایط آزمایش اهمیت بسیار زیادی دارد.

۱۴. تعمیم‌پذیری و عدم قطعیت

یک مدل زمانی واقعاً مفید است که بتواند روی داده‌های جدید نیز عملکرد مناسبی داشته باشد. اگر مدل تنها روی واکنش‌هایی که بسیار شبیه داده‌های آموزشی هستند عملکرد خوبی داشته باشد، کاربرد آن محدود خواهد بود.

این مشکل در طراحی مولکول‌های کاملاً جدید اهمیت بیشتری دارد. ممکن است مدل ساختاری را پیشنهاد کند که در داده آموزشی نمونه مشابهی برای آن وجود نداشته باشد.

در چنین شرایطی، تخمین عدم قطعیت اهمیت زیادی دارد. مدل باید بتواند مشخص کند که کدام پیش‌بینی‌ها با اطمینان بالا و کدام پیش‌بینی‌ها با عدم قطعیت زیاد انجام شده‌اند.

برای مثال، پیش‌بینی موفقیت یک واکنش با احتمال ۹۸ درصد با پیش‌بینی موفقیت ۵۵ درصد از نظر تصمیم‌گیری آزمایشگاهی کاملاً متفاوت است.

استفاده از روش‌های تخمین عدم قطعیت می‌تواند به شیمیدان کمک کند منابع آزمایشگاهی را روی گزینه‌هایی متمرکز کند که احتمال موفقیت بیشتری دارند.

۱۵. تفسیرپذیری مدل‌های هوش مصنوعی

یکی از چالش‌های مهم هوش مصنوعی در شیمی، مسئله تفسیرپذیری است. بسیاری از مدل‌های عمیق مانند جعبه سیاه عمل می‌کنند؛ یعنی خروجی مناسبی ارائه می‌دهند، اما دلیل دقیق تصمیم آن‌ها همیشه مشخص نیست.

در علوم شیمی، این موضوع اهمیت زیادی دارد. اگر مدل بگوید یک واکنش احتمال موفقیت بالایی دارد، شیمیدان ممکن است بخواهد بداند چه ویژگی ساختاری یا چه الگوی واکنشی باعث این پیش‌بینی شده است.

روش‌های Explainable AI می‌توانند به شناسایی عوامل مؤثر بر پیش‌بینی کمک کنند.

تفسیرپذیری علاوه بر افزایش اعتماد به مدل، می‌تواند به کشف دانش جدید نیز کمک کند. اگر یک مدل الگویی را در مجموعه بزرگی از داده‌ها شناسایی کند که قبلاً مورد توجه قرار نگرفته است، شیمیدان می‌تواند آن الگو را به‌عنوان یک فرضیه علمی بررسی کند.

بنابراین، مدل هوش مصنوعی در بهترین حالت تنها یک پیش‌بینی‌کننده نیست؛ بلکه می‌تواند ابزاری برای تولید فرضیه‌های جدید شیمیایی نیز باشد.

۱۶. ترکیب هوش مصنوعی با شیمی محاسباتی

هوش مصنوعی می‌تواند در کنار روش‌های سنتی شیمی محاسباتی نیز مورد استفاده قرار گیرد. روش‌هایی مانند محاسبات مکانیک کوانتومی، Molecular Dynamics و Docking می‌توانند اطلاعات دقیقی درباره ساختار و انرژی مولکول‌ها تولید کنند.

اما انجام این محاسبات برای تعداد بسیار زیادی مولکول ممکن است زمان‌بر باشد. در چنین شرایطی می‌توان از یادگیری ماشین برای ایجاد یک مدل جانشین یا Surrogate Model استفاده کرد.

در این رویکرد، ابتدا تعدادی محاسبه دقیق انجام می‌شود. سپس مدل یادگیری ماشین با استفاده از نتایج موجود آموزش می‌بیند و می‌تواند ویژگی‌های سایر مولکول‌ها را با هزینه محاسباتی کمتر پیش‌بینی کند.

این روش به‌خصوص برای غربالگری مجازی بسیار مفید است. تعداد زیادی مولکول ابتدا با مدل سریع بررسی می‌شوند و سپس تنها گزینه‌های منتخب تحت محاسبات دقیق‌تر قرار می‌گیرند.

۱۷. آزمایشگاه‌های خودکار و حلقه بسته

یکی از چشم‌اندازهای مهم آینده، ترکیب هوش مصنوعی با ربات‌های آزمایشگاهی است.

در روش سنتی، شیمیدان آزمایش را طراحی می‌کند، آزمایش انجام می‌شود، نتیجه تحلیل می‌شود و بر اساس آن آزمایش بعدی طراحی می‌گردد.

در یک سیستم خودکار، این فرایند می‌تواند به شکل یک حلقه بسته انجام شود:

طراحی محاسباتی → انتخاب آزمایش → اجرای رباتیک → تحلیل نتیجه → آموزش مجدد مدل → طراحی آزمایش بعدی

این روش می‌تواند سرعت بهینه‌سازی واکنش‌ها و کشف ترکیبات جدید را افزایش دهد.

برای مثال، اگر هدف پیدا کردن بهترین شرایط یک واکنش باشد، مدل می‌تواند بر اساس نتایج آزمایش‌های قبلی شرایط بعدی را پیشنهاد کند. ربات آزمایش را انجام می‌دهد و نتیجه به مدل بازگردانده می‌شود.

در این حالت، مدل به‌صورت پیوسته از نتایج واقعی یاد می‌گیرد و فضای آزمایش را هوشمندانه‌تر جست‌وجو می‌کند.

۱۸. محدودیت‌های کاربرد هوش مصنوعی در شیمی

با وجود مزایای متعدد، استفاده از هوش مصنوعی در شیمی با محدودیت‌های مهمی همراه است.

نخستین محدودیت، کیفیت داده است. اگر داده‌های آموزشی دارای خطا باشند، مدل نیز می‌تواند الگوهای نادرست یاد بگیرد.

دومین مسئله، تعمیم‌پذیری است. عملکرد مدل روی داده‌های خارج از محدوده آموزشی ممکن است به‌شدت کاهش پیدا کند.

سومین چالش، قابلیت تفسیر است. مدل‌های پیچیده ممکن است پیش‌بینی‌های دقیق اما دشوار برای توضیح ارائه دهند.

چهارمین مسئله، قابلیت سنتز ساختارهای تولیدشده است. ممکن است یک مدل مولکولی تولید کند که از نظر ریاضی معتبر باشد، اما مسیر سنتزی عملی برای آن وجود نداشته باشد یا سنتز آن بسیار پرهزینه باشد.

پنجمین محدودیت، تفاوت میان داده محاسباتی و واقعیت آزمایشگاهی است. شرایط واقعی آزمایش تحت تأثیر عواملی قرار می‌گیرد که ممکن است در داده آموزشی مدل وجود نداشته باشند.

به همین دلیل، معیارهایی مانند Synthetic Accessibility، هزینه مواد اولیه، تعداد مراحل سنتز و پایداری شرایط واکنش باید در طراحی سامانه‌های هوش مصنوعی مورد توجه قرار گیرند.

۱۹. آینده هوش مصنوعی در شیمی

آینده هوش مصنوعی در شیمی احتمالاً به سمت مدل‌های چندمنظوره و ترکیبی حرکت خواهد کرد. این مدل‌ها می‌توانند به‌طور هم‌زمان اطلاعات مربوط به ساختار مولکولی، واکنش، خواص فیزیکی، هندسه سه‌بعدی و داده‌های زیستی را پردازش کنند.

یکی از مسیرهای مهم، توسعه مدل‌هایی است که بتوانند از مرحله طراحی مولکول تا مرحله سنتز را به‌صورت یکپارچه مدیریت کنند. چنین سیستمی می‌تواند یک ساختار جدید تولید کند، خواص آن را پیش‌بینی کند، قابلیت سنتز آن را ارزیابی کند و سپس بهترین مسیر سنتزی را پیشنهاد دهد.

در مرحله بعد، آزمایش می‌تواند توسط یک سامانه رباتیک انجام شود و نتایج برای اصلاح مدل مورد استفاده قرار گیرد.

چنین ساختاری در نهایت می‌تواند یک چرخه خودتقویت‌شونده ایجاد کند که در آن هر آزمایش جدید اطلاعات بیشتری برای تصمیم‌گیری‌های بعدی فراهم می‌کند.

با این حال، توسعه چنین سیستم‌هایی نیازمند استانداردهای داده‌ها، ثبت دقیق شرایط آزمایش، روش‌های ارزیابی معتبر، زیرساخت محاسباتی و همکاری میان متخصصان شیمی، علوم داده و مهندسی است.

نتیجه‌گیری

هوش مصنوعی در حال تبدیل شدن به یکی از ابزارهای مهم شیمی مدرن است و کاربرد آن از پیش‌بینی ساده خواص مولکولی فراتر رفته است. امروزه روش‌های مختلف یادگیری ماشین و یادگیری عمیق می‌توانند در پیش‌بینی محصولات واکنش، پیش‌بینی شرایط واکنش، طراحی مسیرهای سنتزی، رتروسنتز و تولید مولکول‌های جدید مورد استفاده قرار گیرند.

در میان معماری‌های موجود، Transformer به دلیل توانایی در پردازش توالی‌های SMILES و مدل‌سازی روابط پیچیده، در پیش‌بینی واکنش‌ها بسیار مؤثر بوده است. Molecular Transformer نمونه مهمی از این رویکرد محسوب می‌شود (Schwaller et al., ۲۰۱۹). در مقابل، GNN‌ها به دلیل نمایش طبیعی مولکول به شکل گراف، برای پیش‌بینی خواص و تحلیل ساختار مولکولی مناسب هستند.

GAN‌ها و VAE‌ها مسیر مهمی را برای طراحی مولکولی مولد ایجاد کردند. GAN‌ها با استفاده از رقابت میان Generator و Discriminator قادر به تولید ساختارهای جدید هستند. در حالی که VAE‌ها با ایجاد فضای نهفته، امکان جست‌وجو و بهینه‌سازی مولکولی را فراهم می‌کنند. در سال‌های اخیر نیز Diffusion Models به‌عنوان نسل جدیدی از مدل‌های مولد، به‌خصوص در طراحی ساختارهای پیچیده و سه‌بعدی، مورد توجه قرار گرفته‌اند.

با وجود این پیشرفت‌ها، نباید خروجی هوش مصنوعی را با حقیقت تجربی یکسان دانست. یک مدل می‌تواند احتمال موفقیت یک واکنش را بالا ارزیابی کند، اما تنها آزمایش واقعی می‌تواند موفقیت آن را تأیید کند. همچنین مولکولی که از نظر محاسباتی دارای ویژگی‌های مطلوب است، الزاماً در آزمایشگاه قابل سنتز یا از نظر زیستی مؤثر نیست.

بنابراین، آینده شیمی مبتنی بر هوش مصنوعی به جای حذف شیمیدان، به سمت همکاری میان شیمیدان و سامانه‌های هوشمند حرکت خواهد کرد. هوش مصنوعی می‌تواند فضای بسیار گسترده واکنش‌ها و مولکول‌های ممکن را جست‌وجو و اولویت‌بندی کند، در حالی که شیمیدان می‌تواند با استفاده از دانش مکانیزمی، تجربه آزمایشگاهی و قضاوت علمی، نتایج را ارزیابی کند.

در نهایت، ترکیب هوش مصنوعی، شیمی محاسباتی و آزمایش‌های خودکار می‌تواند فرایند کشف مولکول‌ها و واکنش‌های جدید را سریع‌تر و هدفمندتر کند. موفقیت این رویکرد، بیش از آنکه صرفاً به ساخت مدل‌های بزرگ‌تر وابسته باشد، به کیفیت داده، تفسیرپذیری، اعتبارسنجی تجربی و ادغام دانش شیمیایی با روش‌های محاسباتی بستگی خواهد داشت.

منابع

Du, Y., Jamasb, A. R., Guo, J., Fu, T., Harris, C., Wang, Y., Duan, C., Liò, P., Schwaller, P., & Blundell, T. L. (۲۰۲۴). Machine learning-aided generative molecular design. *Nature Machine Intelligence*, 6, ۵۸۹–۶۰۴.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (۲۰۱۴). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27.

Ho, J., Jain, A., & Abbeel, P. (۲۰۲۰). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33.

Kingma, D. P., & Welling, M. (۲۰۱۴). Auto-encoding variational Bayes. *International Conference on Learning Representations*.

Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Hunter, C. A., Bekas, C., & Lee, A. A. (۲۰۱۹). Molecular Transformer: A model for uncertainty-calibrated chemical reaction prediction. *ACS Central Science*, 5(۹), ۱۵۷۲–۱۵۸۳. <https://doi.org/10.1021/acscentsci.9b005۷6>

Schweidtmann, A. M., et al. (۲۰۲۴). Machine learning-guided strategies for reaction conditions design and optimization. *Beilstein Journal of Organic Chemistry*, 20, ۲۴۷۶–۲۴۹۲.

Zhang, L., & Liu, Z. (۲۰۲۴). Exploring chemical reaction space with machine learning models: Representation and feature perspective. *Journal of Chemical Information and Modeling*.