



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

افزایش یادگیری عمیق در کشف دارو

چکیده

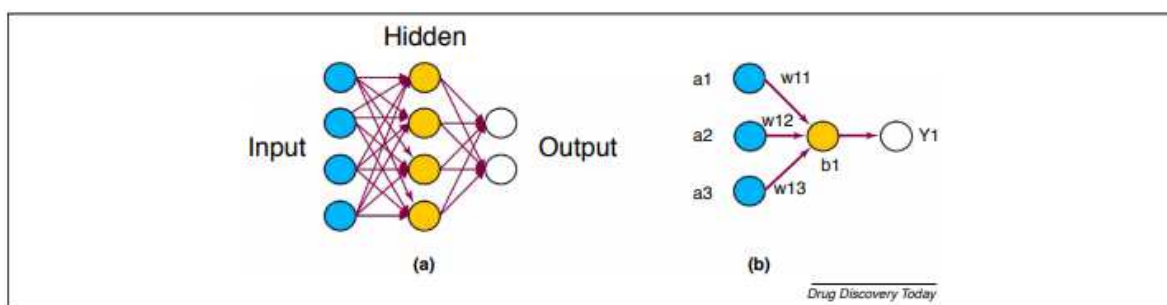
طی دهه گذشته، یادگیری عمیق به موفقیت قابل ملاحظه ای در انواع حیطة های تحقیقاتی هوش مصنوعی دست یافته است. این فناوری که از روی تحقیقات قبلی در زمینه شبکه های عصبی مصنوعی تکامل یافته است، نشان داده شده که عملکرد برتری نسبت به سایر الگوریتم های یادگیری ماشینی در حیطة هایی نظیر تشخیص تصویر و صدا ، پردازش زبان طبیعی در میان سایرین دارند. اولین موج کاربرد یادگیری عمیق در تحقیقات داروشناسی در سالهای اخیر ظهور یافته است و استفاده از آن فراتر از پیشگویی های فعالیت زیستی رفته است و در مطرح سازی مسائل گوناگون در کشفیات دارویی نویدبخش بوده است. مثالهایی برای تحت پوشش قرار دادن پیشگویی فعالیت زیستی، طراحی مولکولی جدید، تحلیل تصویر بیولوژیکی و پیشگویی سنتز مورد بحث قرار خواهد گرفت.

مقدمه

داده های دیجیتالی در همه شکل و اندازه ای به شکل نمایی روبه رشد است. طبق گفته موسسه امنیتی ملی امریکا، اینترنت دارد 1826 پتابایت داده ها را در روز پردازش می کند. در سال 2011، اطلاعات دیجیتالی نه برابر از لحاظ حجمی در عرض صرفا پنج سال رشد کرده است و تا سال 2020، میزان آن در دنیا بنا به انتظار به 35 تریلیون گیگابایت می رسد. تقاضای بالای بررسی و تحلیل داده های کلان باعث تشویق استفاده از الگوریتم های یادگیری ماشینی گرسنه داده مانند یادگیری عمیق DL شده است. DL موفقیت عظیمی را در دامنه وسیعی از کاربردها مانند بازیهای کامپیوتری ، تشخیص گفتاری، بینایی کامپیوتری، پردازش زبان عصبی، ماشین های بدون راننده، در میان سایر موارد کسب کرده اند. منصفانه است که بگوییم DL در حال تغییر زندگی روزمره ماست. در روند فناوری طراز اول منتخب Gartner در سال 2018، فناوریهای AI نماینده با DL در جایگاه ارشد رتبه بندی شدند.

طی دهه گذشته، یک افزایش قابل ملاحظه ای در مقدار فعالیت ترکیب موجود و داده های بیومدیکال به یمن ظهور تکنیک های تجربی جدید مانند HTS و سنتز موازی در میان سایرین وجود آمده است. اینکه چگونه و با چه کارایی داده های شیمی در مقیاس وسیع کاویده شود، یک مسئله حیاتی در کشفیات دارویی می باشد. حجم داده های بزرگتر همراه با فناوری خودکارسازی افزایش یافته باعث افزایش استفاده بیشتر از یادگیری ماشینی می شود. در کنار روشهای معین مانند ماشین های بردار پشتیبان یا SVM، شبکه های عصبی NN و جنگل تصادفی RF که برای ایجاد مدل های QSAR برای مدت زمان طولانی استفاده شده اند، روشهایی مانند عامل گیری ماتریس و DL تازه شروع به کاربرد کرده اند. DL مزیت افزایش مقدار داده ها و افزایش مداوم قدرت کامپیوتر موجود را دارد. یک تفاوت میان اغلب روشهای یادگیری ماشینی دیگر و DL همان قابلیت انعطاف پذیری معماری NN در DL می باشد. معماری های مورد بحث در این مقاله مروری از نوع شبکه های عصبی در هم پیشی CNN، شبکه های عصبی عودکننده RNN، و شبکه های تغذیه رو به جلو کاملاً بهم مرتبط می باشند. NN های تک لایه در مدلسازی QSAR برای مدتهای طولانی استفاده می شده است. و با افزایش اندازه داده ها و قدرت محاسباتی طبیعی است که از شبکه های تغذیه رو به جلو چندلایه برای پیشگویی های فعالیت زیستی استفاده بشود. یک پیشرفت تعجب آور استفاده از RNNها در طراحی های نوین است که نمی شد چند سال پیش آنها پیش بینی نمود. با اتخاذ تجهیزات تصویربرداری سطح بالا، CNNها موفقیت قابل ملاحظه ای را در بینایی کامپیوتری کسب کرده اند و یک انتخاب طبیعی برای پردازش تصویری بیولوژیکی شده اند. عرصه بکارگیری DL در کشف دارو به سرعت با مقالات جدید منتشره تقریباً هر هفته در حال تحول و پیشرفت است.

شکل 1.



اخیرا چندین مقاله مروری روی کاربردهای DL در شیمی محاسباتی و علوم حیاتی منتشر شده است. در اینجا تمرکز ما بیشتر روی کاربردهای DL در کشفیات دارویی بویژه در کموانفورماتیک و حوزه تحلیل تصویر بیولوژیکی است و معماریهای DL را که تا کنون برای کشفیات دارویی استفاده شده است، مورد تاکید قرار می دهد.

اصول یادگیری عمیق

مدل DL یک رده الگوریتم های یادگیری ماشینی است که از شبکه های عصبی مصنوعی ANN با بسیاری لایه های واحدهای پردازش غیرخطی برای نمایش داده های یادگیری استفاده می کند. قدیمی ترین ANN می تواند به سال 1943 بازگردد که در آن زمان آقایان Warren McCulloch و Walter Pitts یک مدل محاسباتی را برای NNها براساس ریاضیات و الگوریتم هایی به نام منطق حدآستانه ابداع کردند. ساختار پایه یک ANN مدرن در شکل 1 نمایش داده شده است و از ساختار مغز انسان الهام گرفته است. سه لایه پایه در یک ANN وجود دارد: لایه ورودی، لایه مخفی و لایه خروجی. بسته به نوع ANN، گره ها که نیز نرون ها نامیده می شوند، در لایه های مجاورت کاملا بهم مرتبط بوده یا تا اندازه ای بهم مرتبط هستند. متغیرهای ورودی توسط گره های ورودی اتخاذ شده و متغیرها از طریق گره های مخفی تغییرشکل می یابند و در پایان متغیرهای خروجی در گره های خروجی محاسبه می شوند. رابطه بینابینی میان مقادیر ورودی و خروجی یک واحد مخفی می تواند در شکل 1b به شکل مثال آورده شود. مقدار خروجی Y_i گره i طبق معادله 1 محاسبه می شود.

$$Y_i = g \left(\sum_j W_{ij} * a_j \right) \quad (1)$$

که در آن a_j اشاره به متغیرهای ورودی دارد، W_{ij} سنسجش گره ورودی j گره i است و تابع g همان تابع فعالسازی می باشد که به طور طبیعی تابع غیرخطی است (برای مثال تابع سیگموئید یا تابع گاوسی) تا ترکیب خطی سیگنال ورودی را از گره های ورودی به یک ارزش خروجی تغییر شکل بدهد. آموزش یک ANN بوسیله اصلاح تکراری مقادیر وزنی در شبکه برای بهینه سازی خطاهای میان مقدار پیشگویی شده و مقدار حقیقی انجام می گیرد که معمولا از طریق روشهای انتشار عقبی است. الگوریتم ANN مدرن طی دهه 60 میلادی تا دهه 80 میلادی ابداع گردید و کاربردهایش از آن زمان پدیدار شده است. ولیکن روش ANN قدیمی

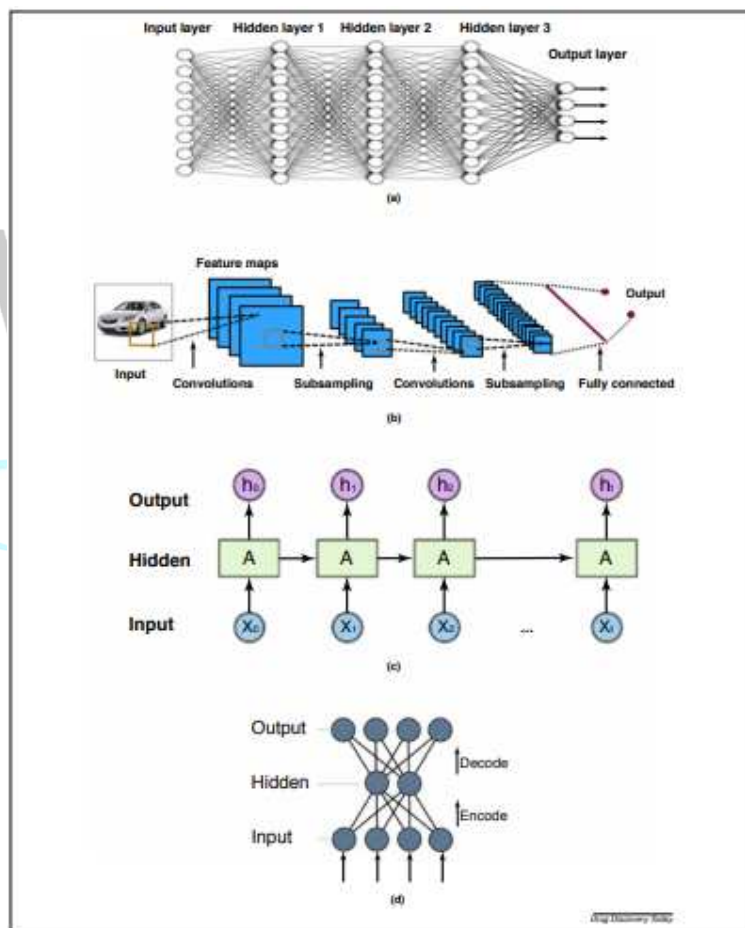
از مسائلی مانند برازش بیش از حد، کاهش گرادیان ها، در میان سایر مسائل رنج برده و تا حد زیادی با سایر الگوریتم های یادگیری ماشینی نظیر SVM و RF جایگزین شده است. پیشرفت اخیر DL به مدل ANN یک تولد دوباره داده است. تفاوت اصلی میان DL و ANN قدیمی مقیاس و پیچیدگی NNها می باشد. DL از تعداد زیادی لایه های مخفی استفاده می کند در صورتیکه ANNهای سنتی به طور طبیعی می تواند تنها یک یا دو لایه مخفی را به یمن محدودیت سخت افزاری کامپیوتری در روزهای اول مورد رسیدگی قرار دهد. DL می تواند قدرت استفاده از گره های بسیار بیشتری را در هر لایه به یمن وجود سخت افزار CPU و GPU قوی تر داشته باشد. همچنین بسیاری بهبودهای الگوریتمی در DL برای مثال با استفاده از روشهای حذف و DropConnect برای مطرح سازی مسائل بیش برازش، با بکارگیری واحد خطی اصلاح کننده یا ReLU برای اجتناب از ناپدیدسازی گرادیان ها و ورود لایه های معمول و ذخیره سازی شده به عنوان معماری های شبکه تازه برای قابلیت استفاده از تعداد زیادی متغیرهای ورودی وجود دارند. اغلب بسته های نرم افزاری DL از نوع منبع باز هستند. TensorFlow، Caffe، PyTorch، Keras و Theano در میان مشهورترین بسته های نرم افزاری DL استفاده شده در جامعه علوم داده ها می باشند. در اینجا به اختصار به معرفی چندین معماری NN مشهور استفاده شده در DL می پردازیم (شکل 2). اول از همه شبکه عصبی عمیق کاملاً به هم مرتبط DNN می باشد که حاوی چندین لایه متعدد مخفی است و هر لایه متشکل از صدها واحد پردازش غیرخطی می باشد (شکل 2a). DNNها می توانند تعداد زیادی ویژگی های ورودی را اتخاذ کنند و نرونها در لایه های مختلف یک DNN می توانند به طور خودکار ویژگی هایی را در سطوح سلسله مراتبی مختلف استخراج کنند.

نوع مشهور دیگر NN همان CNN می باشد که به طور وسیعی برای تشخیص تصویر (شکل 2b) استفاده می شود. معمولاً حاوی چندین لایه درهم پیچشی و لایه های زیرنمونه سازی است. لایه درهم پیچشی شامل یک مجموعه فیلترها می باشد (یا هسته ها) که دارای میدان گیرنده کوچک و پارامترهای قابل یادگیری است. طی عبور رو به جلو، هر فیلتر از عرض و طول به حجم ورودی درهم پیچیده شده و محصول نقطه ای را بین کل فیلتر و میدان گیرنده اش در حجم ورودی محاسبه کرده و یک نقشه ویژگی دو بعدی را از آن فیلتر تولید می کند. لایه زیرنمونه گیری برای کاهش اندازه نقشه های ویژگی استفاده می شود. در پایان نقشه های ویژگی روی لایه های کاملاً بهم مرتبط به صورت مسلسل وار قرار می گیرند که در آن نرونها در لایه های مجاور همگی

درست مثل یک ANN قدیمی بهم مرتبط می شوند و یک مقدار خروجی نهایی را بدست می دهند. به یمن اشتراک همان پارامترها برای هر فیلتر، یک CNN تا حد زیادی باعث کاهش تعداد پارامترهای آزاد یادگیری شده می شود با این حساب باعث کاهش حافظه مصرفی و افزایش سرعت یادگیری می شود. این مدل از سایر انواع الگوریتم های یادگیری ماشینی در تشخیص تصویر بهتر عمل می کند.

یک نوع دیگر از یک ANN (شکل 2c) همان RNN می باشد. برخلاف NN های تغذیه رو به جلو، این مدل اجازه ارتباط در میان نرون ها در لایه مخفی یکسانی را برای تشکیل یک چرخه هدایت شده می دهد.

شکل 2 .



RNN ها می توانند داده های متوالی را به شکل ویژگی های ورودی بگیرند، که برای کارهای وابسته به زمان مانند مدل سازی زبان خیلی مناسب می باشد. با استفاده از یک فناوری به نام حافظه کوتاه مدت طولانی یا LSTM، مدل های RNN می توانند باعث کاهش مسئله گرادیان رنگ پریدگی بشوند.

چهارمین معماری ANN که در شکل 2d نشان داده شده است کدگذاری کننده خودکار یا AE نام دارد. یک AE یک NN است که برای یادگیری بدون نظارت استفاده شده است. این کدگذاری کننده شامل یک بخش کدگذاری کننده است که یک NN برای تغییر شکل اطلاعات دریافتی از لایه ورودی به یک تعداد محدودی واحدهای مخفی و بعد مزدوج سازی یک کدگشاینده NN با لایه خروجی دارای همان تعداد گره های لایه ورودی می باشد. به جای پیشگویی برچسب های نمونه های ورودی، هدف کدگشاینده NN بازسازی ورودی های خودش از یک تعداد کمتری واحدهای مخفی می باشد. معمولاً، هدف از AE برای کاهش بعدگرایی غیرخطی می باشد. اخیراً، مفهوم AE بطور وسیع تری برای یادگیری مدل‌های زیاده از روی داده ها استفاده شده است. در ذیل، ما نشان داده ایم که چگونه این فناوریهای DL در تحقیقات کشف دارو بکار گرفته می شود.

کاربرد یادگیری عمیق در خصوصیات ترکیب و پیشگویی فعالیت

روشهای یادگیری ماشین شامل ANN در پیشگویی فعالیت ترکیب برای مدت زمان طولانی بکار بسته شده است. به طور طبیعی، روشهای DL اتخاذ شده اند تا مسائل پیشگویی فعالیت را در وهله اول مطرح سازند. وقتی ترکیبات با همان تعداد توصیفگرهای مولکولی ارائه می شود، روش مستقیم رو به جلو همان استفاده از DNN کاملاً مرتبط برای ساختن مدلها می باشد. Dahl و همکارانش یک مدل DNN را روی پایگاه داده چالش Merck Kaggle با استفاده از تعداد زیادی توصیفگرهای توپولوژیکی دو بعدی بکار گرفتند. و DNN عملکرد اندکی بهتر را در 13 تا از 15 هدف کل نسبت به روش RF استاندارد نشان داده است. برخی از یادگیری های کلیدی از این مطالعه عبارتند از: (1) DNNها می توانند هزاران توصیفگر را بدون نیاز به انتخاب ویژگی مورد رسیدگی قرار دهند. (2) حذف می تواند از مسئله برازش بیش از حد معروف که توسط یک ANN سنتی مواجه شده است، اجتناب ورزد. (3) هایپرپارامتر (تعداد لایه ها، تعداد گره ها در هر لایه، نوع توابع فعالسازی و غیره). بهینه سازی می تواند عملکرد DNN را به حداکثر خودش برساند. (4) مدل‌های DNN چندکاره بهتر از مدل‌های تک کاره عمل می کنند. Mayr و همکارانش گزارش کردند که مدل‌های DNN چندکاره آنها چالش Tox21 را روی یک پایگاه داده ها متشکل از 12 هزار ترکیب برای 12 سنجش سمیت سطح بالا پشت سر گذاشته اند. مشابه معماری Dahl، تابع حذف و فعالسازی ReLU در DNN استفاده گردید و آموزش مدل به طور موازی روی ماشین های GPU انجام گرفت. آنها از مجموعه ویژگی بزرگی با توصیفگرهای استاتیک (توصیفگرهای 3

بعدی، 2 بعدی، توکسیکوفورهای از قبل تعیین شده) و نیز توصیفگرهای اثرانگشت ارتباطی بسط یافته و با روش پویا ایجاد شده ECFP برای اینکه DNN بتواند استنتاج خودویژگی را طی آموزش بنماید، استفاده کرده اند. جالبتر اینکه، تحلیل رابطه آماری برای مدل‌های DNN به طور انحصاری با استفاده از ECFP انجام گرفته است، و زیرساختارها به طور معنی داری با توکسی فورهای معین در هر لایه مخفی که می تواند شناسایی شود همراه است. این نتایج معیار نشان دهنده مزیت یک DNN چندکاره در مقایسه با یک DNN تک کاره و روشهای یادگیری ماشینی مرسوم بوده است.

اخیراً، برخی مطالعات معیار دیگر حمایت و نتیجه گیری بیشتری را منتشر کرده اند. Ramsundar و همکارانش یک مطالعه سیستماتیک را برای ساختن DNNهای چندکاره ساخته اند و عملکرد آنها را با مدل‌های DNN تک کاره مقایسه کرده اند. نتایج آنها نشان می دهد که مدل‌های چندکاره به طور ثابتی بهتر از مدل‌های تک کاره و مدل‌های RF می باشد. Koutsoukas و همکارانش یک مدل DNN را با برخی روشهای یادگیری ماشینی که به طور متداول استفاده می شود نظیر SVM، RF و در میان بقیه موارد روی هفت پایگاه داده انتخابی از ChEMBL مقایسه کرده اند. مشخص گردید که DNNها به لحاظ آماری بهتر از سایر روشهای یادگیری ماشینی عمل می کنند (با p value کمتر از 0.01 براساس تست آماری ویلکوکسون). Lenselink و همکارانش مطالعه معیار دیگری را برای مقایسه DNN با روشهای یادگیری ماشینی معمول RF و SVM و بیزی نوین و روشهای رگرسیون منطقی با در نظر گرفتن توصیفگرهای پروتئین (یعنی مطالعه پروتئوکوموتریک PCM) گزارش کردند. آنها روی عملکرد انواع مدل‌های طبقه بندی مختلف در یک پایگاه داده متشکل از 314767 تعاملات هدف-ترکیب شیمیایی تحقیقاتی انجام دادند. مدل DNN به بهترین مدل از لحاظ BEDROC تبدیل شد (تمیز تقویت شده طبق روش Boltzmann برای مشخصات عملگری گیرنده)، و مشخص گردید که اجرای چندکاره و PCM عملکرد را روی DNNهای تک کاره بهبود می دهد.

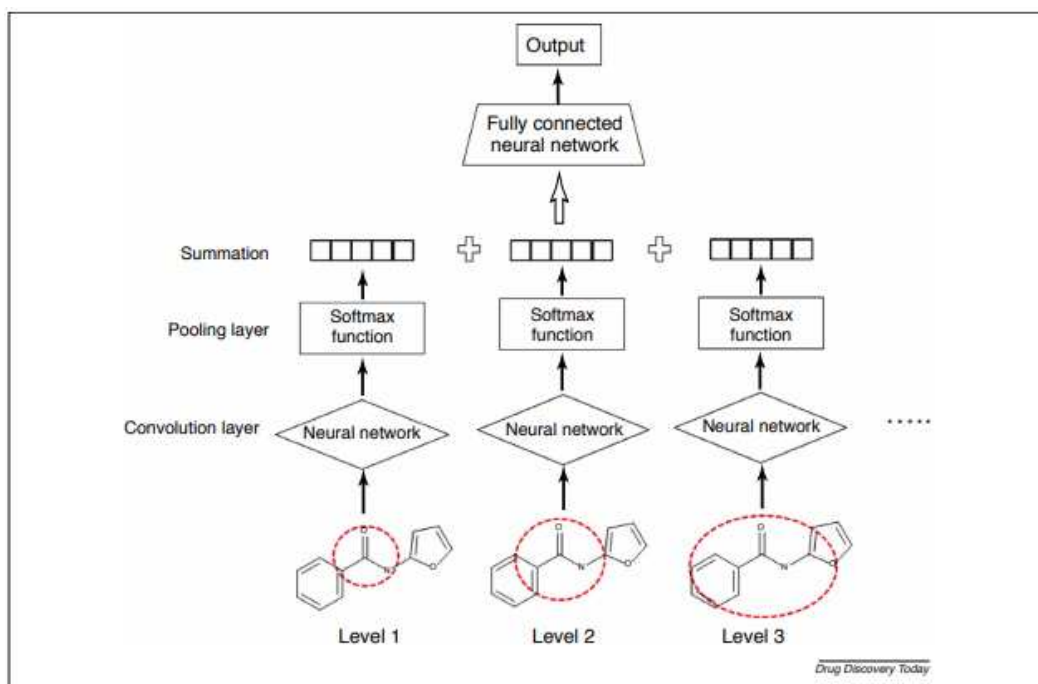
در کنار مطالعات معیار DNN، Subramanian و همکارانش یک مطالعه را با استفاده از DNN با توصیفگرهای توپولوژیکی دوبعدی برای ایجاد یک مدل فعالیت BACE پیشگویی گزارش کردند و به یک صحت طبقه بندی معادل 0.82 و خطای استاندارد برابر با pic50 تقریباً برابر با 0.53 روی مجموعه روایی سازی دست یافتند. Aliper و همکارانش مدل‌های DNN را برای پیشگویی خواص داروشناختی داروها و برای تغییر هدف دارو برای

اجرای داده های بین پوستی از پروژه LINCS و نیز اطلاعات مسیریابی ساخته اند. نشان داده شده است که با استفاده از مسیر و اطلاعات در سطح ژنی، مدل‌های DNN به صحت بالایی در پیشگویی موارد مصرف داروها دست یافته اند و با این حساب می توانند برای تغییر هدف داروها بکار گرفته شوند.

اقداماتی نیز در استفاده از یادگیری نمایش صورت گرفته است (یعنی قادر کردن NNها برای یادگیری مستقیم از ساختار مولکولی به جای استفاده از توصیفگرهای مولکولی از قبل تعریف شده). این ایده ابتدا توسط Merkwirth و همکارانش در 2005 بررسی شده است. چندین سال بعد، دو روش مختلف برای مطرح سازی این مسئله ایجاد شده است. Lusci و همکارانش یک روشی را گزارش کردند که یک نوع RNN را بکار گرفته است که UGRNN نام دارد که ابتدا باعث تغییر شکل ساختارهای مولکولی به شکل بردارهایی با همان طول نمایش مولکولی یکسان شده و بعد آنها را به لایه NN کاملاً مرتبطی برای ساخت مدل منتقل می کند. مقادیر بیت در بردارها از پایگاه داده یاد گرفته می شود. روش UGRNN نشان داده شده است که قادر به ایجاد مدل‌های قابل حل پیشگویانه است که صحت قابل مقایسه ای با مدل‌های ساخته شده با توصیفگرهای مولکولی دارد. Xu و همکارانش همان روش را برای مدلسازی صدمه کبدی القایی با دارو یا DILI بکار بستند. مدل‌های DL براساس 475 دارو ساخته شده و روی یک پایگاه داده خارجی با 198 دارو روایی سازی گردیدند. بهترین مدل یک AUC با مقدار 0.955 را حاصل کرد که از صحت مدل‌های DILI قبلاً گزارش شده بیشتر است.

نوع دیگر این روش به نام مدل‌های در هم پیچش گراف می باشد. ایده اساسی مشابه با روش UGRNN می باشد، که NNها را برای تولید خودکار یک بردار توصیف مولکولی بکار می گیرد و مقادیر بردار با آموزش NNها یادگرفته می شود. با الهام از روش اثرانگشت حلقوی مورگان، Duvenaud و همکارانش روش اثرانگشت عصبی را به عنوان یکی از اولین اقدامات در ایجاد مدل در هم پیچش گراف مطرح کردند. جریان کاری این روش را می توان در شکل 3 مشاهده کرد. اول اینکه، ساختار مولکولی دو بعدی برای تشکیل یک ماتریس حالت خوانده می شود، که حاوی اتم و اطلاعات پیوند اتمی برای هر اتم است (که براساس پیوندهای متصل به اتم می باشد). ماتریس حالت سپس از طریق عملیات در هم پیچشی از طریق NN تک لایه برای تولید یک بردار طول ثابت همانند نمایش مولکولی عبور می کند. عملیات در هم پیچش می تواند در سطوح مختلف با در نظرگیری نقش اتم های مجاور اجرا بشود که معادل اثرانگشت های حلقوی در سطوح مجاور مختلف می باشد.

شکل 3 .



بردارهای ایجاد شده از عملیات درهم پیچش مختلف ابتدا از طریق تحولات نرم افزاری ماکزیمم عبور کرده و بعد به شکل بردار نهایی برای ترکیب جمع بندی شده است که اطلاعات سطح مولکولی کدگذاری کننده اثرانگشتی عصبی است. اثرانگشت های عصبی از طریق لایه دیگر کاملاً مرتبط NN عبور کرده تا ایجاد خروجی نهایی بشود. مقادیر بیت در اثر انگشت عصبی از طریق آموزش یادگیری شده و قابل تمایز است. در سه مورد تست Duvenaud، نتایج بهتری با استفاده از اثرانگشت هایی به جای اثرانگشت های Morgan کسب شده است و از همه مهمتر، ساختارهای فرعی اثرگذار در مدل درهم پیچش گراف می تواند برای تفسیر این مدل به رویت درآید. مزیت مدل در هم پیچش گراف آنست که توصیفگرها به طور خودکار طی آموزش تولید شده و نیازی به توصیفگر مولکولی از قبل تعریف شده ندارند. چنین توصیفگری یک توصیفگر عمومی نیست، ولیکن خاص کار است و کاملاً قابل تمایز است و با این حساب می تواند احتمالاً باعث پیشگویی بهتری بشود. سایر روشهای درهم پیچش گراف مولکولی توسط Kearnes و همکارانش، Xu و همکارانش، Li و همکارانش و Coley و همکارانش برای بسط روش Duvenaud گزارش شده است. اخیراً، محققان گوگل چندین الگوریتم درهم پیچش گراف موجود را مجدداً تنظیم فرمول کرده اند تا یک چارچوب به نام شبکه عصبی انتقال پیام یا MPNN بدست آید و از MPNNها برای پیشگویی خصوصیات شیمیایی کوانتوم استفاده کرده اند.

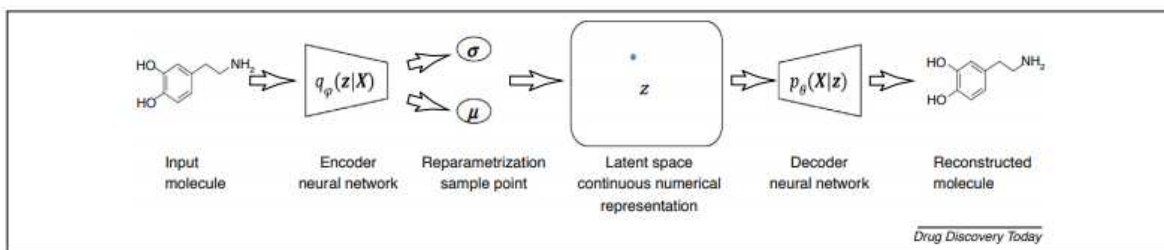
در کنار روشهای یادگیری نمایش مبتنی بر گراف، روشهای DL براساس انواع دیگر نمایشات مولکولی نیز بررسی شده است. Bjerrum از یک رشته SMILES به عنوان ورودی RNNهای LSTM استفاده کردند تا مدل‌های پیشگویی را بدون نیاز به ایجاد توصیفگرهای مولکولی بسازند. جالب تر اینکه، مشاهده گردید ترکیب پایگاه داده با استفاده از رشته های متعدد SMILES برای نمایش همان ترکیب باعث کسب نتایج بهتر نسبت به استفاده از SMILES اساسی شده است. Goh و همکارانش یک CNN را روی تصاویر ترسیمات دوبعدی مولکولها بکار بستند و به نتایج قابل مقایسه ی شگفت انگیزی برای مدل‌های DNN آموزش دیده روی ECFP دست یافتند. بعلاوه، وقتی تصاویر با برخی اطلاعات شیمیایی پایه ترکیب می شوند، عملکرد مدل بیشتر بهبود می یابد. قابلیت یادگیری نمایشات از ساختارها مستقیماً بدون استفاده از هر گونه توصیفگر ساختاری از قبل تعریف شده ای یک DL متمایزکننده ویژگی مهم از سایر روشهای یادگیری ماشینی است و اساساً باعث می شود که انتخاب ویژگی سنتی و عملیات کاهش غیرضروری بشود.

طراحی جدید از طریق یادگیری عمیق

کاربرد جالب دیگر DL در کموانفورماتیک همان ایجاد ساختارهای شیمیایی جدید از طریق NN ها می باشد. Gomez-Bombarelli و همکارانش یک روش جدیدی را با استفاده از کدگذاری کننده خودکار متغیر یا VAE برای تولید ساختارهای شیمیایی مطرح کردند (شکل 4).

TarjomeFa.Com

شکل 4



اولین مرحله استفاده از VAE برای انجام یادگیری بدون نظارت برای نقشه گذاری ساختارهای شیمیایی است (رشته های SMILES) در پایگاه داده های ZINC به فضای پنهان می باشد. وقتی یادگیری VAE انجام شد، بردار پنهان در فضای نهفته یک نمایش مداوم ساختار مولکولی شده و می تواند به طور معکوسی به یک رشته SMILES از طریق VAE آموزش دیده تغییر شکل بدهد. ایجاد یک ساختار جدید با خواص دلخواه می تواند با

جستجوی راه حل های پنهان بهینه در فضای پنهان پیوسته از طریق روش بهینه سازی (برای مثال بهینه سازی بیزی) و بعد با کدگشایی راه حل های پنهان جستجو شده به SMILES تحقق یابد. به پیروی از کار تحقیقاتی Gomez-Bombarelli ، آقایان Kadurin و همکارانش از VAE به عنوان یک ژنراتور توصیفگر مولکولی همراه با شبکه زایای متنوع GAN که یک معماری NN خاص است برای ایجاد ساختارهای جدیدی استفاده کرده اند که بنا به ادعا خواص ضدسرطانی خاص و نویدبخشی داشته است. Blaschke و همکارانش از روش VAE برای ایجاد ساختارهای تازه با فعالیت پیشگویی شده علیه گیرنده نوع دوم دوپامین استفاده کردند.

RNNها در حیطه پردازش زبان طبیعی خیلی موفق بوده اند. Segler و همکارانش و نیز Yuan و همکارانش مطالعات خود را با استفاده از RNNها برای تولید ساختارهای شیمیایی تازه گزارش کردند. بعد از آموزش RNN روی تعداد بزرگی از رشته های SMILES، روش RNN به طرز تعجب آوری به خوبی برای ایجاد رشته های SMILES معتبر جدید کار کرده است که در مجموعه آموزشی گنجانده نشده است (شکل 5). RNN به لحاظ ساختاری SMILES های معتبر را با یادگیری توزیع احتمالات زیربنایی کاراکترها در یک رشته SMILES می نویسد و در این مورد RNN می تواند یک مدل زایا برای ساختارهای مولکولی در نظر گرفته شود. Segler و همکارانش همچنین از احتمال استفاده از RNNها برای تولید کتابخانه های خاص هدف با اولین آموزش یک مدل قبلی کلی و بعد یک مدل متمرکز و تنظیم شده ظریف از طریق یادگیری انتقالی روی مجموعه کوچکی از ترکیبات فعال خاص هدف بهره برداری کردند. در یک تحلیل گذشته نگر برای تست روی دو هدف فعال انتی بیوتیکی، مدل های متمرکز آنها قادر به تولید 18 درصد فعال های حقیقی نادیده برای استافیلوکوک ارئوس و 28 درصد باری Plasmodium falciparum شده است.

Jaques و همکارانش یک فناوری یادگیری تقویتی به نام یادگیری Q عمیق را همراه با یک RNN برای ایجاد SMILES با خصوصیات مولکولی دلخواه نظیر cLogP و QED شبیه به دارو را بکار بستند. ولیکن روش آنها نیازمند یک تابع جایزه است که قوانین دست نویس را برای جریمه سازی انواع نامطلوب ساختارها اضافه کند که در غیراینصورت منجر به بهره برداری از نتایج پاداش دهی در مولکول های ساده غیرواقعی گرایانه می شود. برای غلبه بر موانع کار، Olivecrona یک شیوه یادگیری تقویت مبتنی بر سیاستگزاری را برای تنظیم RNNهای از قبل آموزش دیده برای تولید مولکول هایی مطرح کرده که دارای خواص معین تعریف شده کاربر می باشند. در یک

مثال تست برای تنظیم مدل برای تولید ترکیبات پیشگویی شده برای فعالیت علیه گیرنده دوپامین نوع دو، این مدل باعث ایجاد ساختارهایی شده است که بیش از 95 درصد بنا به پیشگویی فعال بوده و شامل مولکولهای فعال تایید شده آزمایشگاهی است که در مدل زایا و در مدل پیشگویی فعالیت وجود نداشته است.

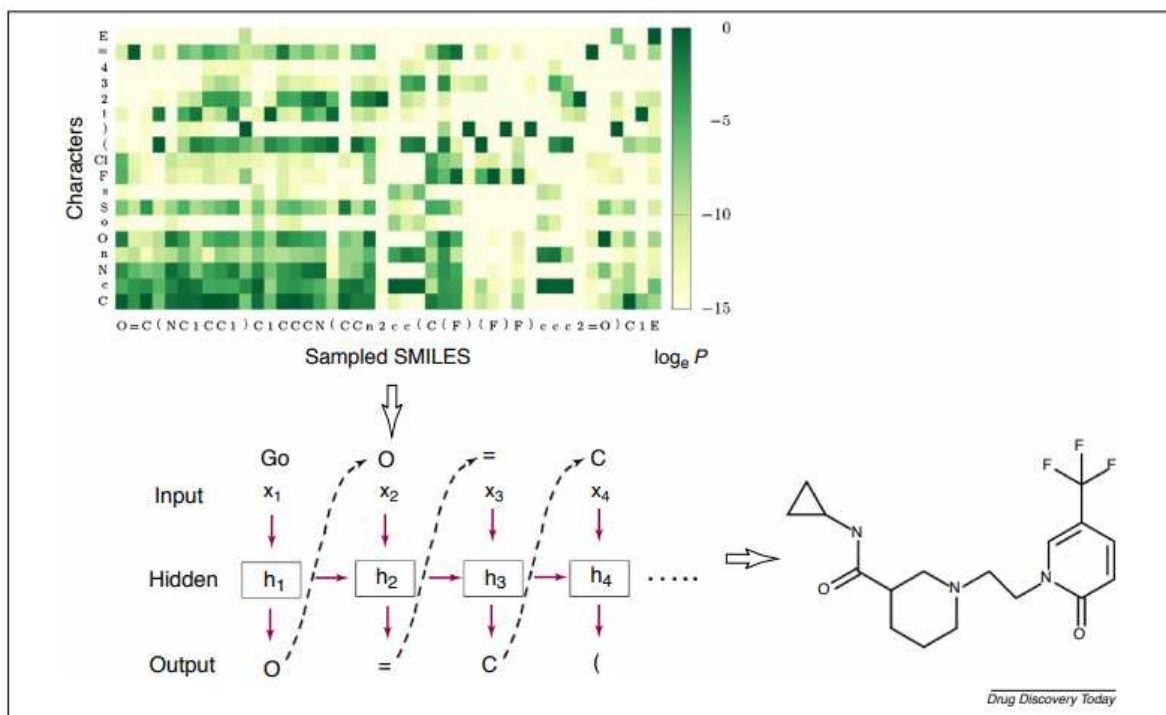
روشهای توضیح داده شده قبلی نشان دهنده قابلیت هایی مانند راههای دیگر برای روشهای مبتنی بر قانون قدیمی برای طراحی جدید می باشد. ولیکن مشخص شده است که GANها و روشهای یادگیری تقویت شده مستعد ریزش حالت هستند (یعنی مدلها تنها تولید یک راه حل منفرد یا خانواده کوچکی از راه حل های کوچک می نمایند). این امر در تحقیقات اخیر در زمینه تولید ساختار با استفاده از ابزار DL مورد تاکید قرار گرفته است. اقدامات قابل ملاحظه ای طرف مطرح سازی این مسئله شده است.

کاربرد یادگیری عمیق در پیشگویی واکنش ها و آنالیز رتروسنتتیک

پیشگویی های سنتتیک دارای یک سابقه طولانی مدت می باشد که به زمان روشهای مبتنی بر قانون در دهه 60 میلادی باز می گردد. همین اواخر برخی نتایج نویدبخش در پیشگویی واکنش با استفاده از روشهای DL گزارش گردیدند. هرچند هیچ مقایسه آشکاری با سایر روشهای یادگیری ماشینی وجود نداشته است، نتایج نشان داده اند که DL می تواند به عملکرد موازی یا برتر یا روشهای مبتنی بر قانون دست یابد. به لحاظ شماتیک، دو نوع مسئله می تواند با یادگیری ماشینی مطرح بشود از جمله DL در انفورماتیک واکنش می باشد. یک نوع پیشگویی واکنش رو به جلو است، که در آن محصولات با یک مجموعه واکنشگر معین می شود، و نوع دیگری پیشگویی رتروسنتتیک است که با یک محصول نهایی معین مراحل واکنش که تولید محصول می کنند پیشگویی می شوند. Coley و همکارانش از NN برای رتبه بندی محصولات کاندیدا برای مجموعه ای از واکنش های مبتنی بر مجموعه آموزشی با 15 هزار واکنش از پنتت های امریکا استفاده کرده اند. واکنش ها به الگوها و مدلهای آموزش دیده تقسیم بندی شده اند که به درستی به الگوها و مدل آموزشی با اختصاص صحیح رتبه محصول اصلی یک در 71.8%، رتبه برابر سه یا کمتر در 86.7% و رتبه برابر 5 یا کمتر در 90.8% درصد موارد طبقه بندی شدند. برای غلبه بر مسائل کارایی و پوشش که با مسائل پیشگویی واکنش مبتنی بر الگو مواجه است، یک روش عاری از الگو در یک مطالعه پیگیری با گروه تحقیقاتی یکسانی مطرح گردیده است. آنها از شبکه تفاوت Weisfeiler-Lehman برای امتیازدهی واکنش های مورد نظر تولید شده استفاده کردند و

عملکرد برتر در مقایسه با روشهای مبتنی بر الگوی واکنش حاصل آمد. Segler و همکارانش از تعداد 3.5 میلیون واکنش به عنوان مجموعه آموزشی برای DNN استفاده کردند. یک صحت بالا تراز بالای 97 درصدی برای پیشگویی واکنش و 95 درصدی برای آنالیز رتروسنتتیک حاصل شد. در مطالعه دیگری، آنها شبکه های سیاستگزاری و جستجوی درخت مونته کارلو را برای پیشگویی رتروسنتتیک با استفاده از یک مجموعه آموزشی شامل 12 میلیون واکنش از متون علمی ترکیب کردند.

شکل 5.



سیستم آنها می تواند دوبار بر برنامه های رتروسنتتیک مولکولها را نسبت به روش های مبتنی بر قانون حل نماید. Liu و همکارانش از مدل های توالی-به-توالی عصبی برای پیشگویی رتروسنتتیک استفاده کرده اند. آنها از 50 هزار واکنش بدست آمده از پتنت های آمریکایی برای آموزش شبکه و صحت مشابه بدست آمده برای روشهای مبتنی بر قانون استفاده کرده اند.

کاربرد شبکه های عصبی معمول برای پیشگویی تعاملات لیگاند-پروتئین

ارزیابی تعامل بین یک پروتئین و یک لیگاند بخش مهم برنامه باراندازی مولکولی است و مقدار زیادی توابع امتیازدهی یا براساس میدان نیرو یا دانش مربوط به ساختار کمپلکس پروتئین-لیگاند ساخته شد. با الهام از

موفقیت CNN ها در آنالیز تصویر، چندین مطالعه اخیرا در بکارگیری CNN برای امتیازدهی تعاملات پروتئین-لیگاند منتشر شده است. یک مثال نمونه همان تحقیقات انجام شده توسط Ragoza و همکارانش می باشد. ساختارهای پروتئین-لیگاند به شبکه ای با رزولاسیونی برابر با 0.5 آنگسترام قسمت بندی گردید. این شبکه روی هر وجهش 24 آنگسترام بود که مرکز آن روی محل باندسازی قرار داشت. هر اتم هم با یک عملکرد توصیف شده بود و تراکم اتم ها روی شبکه برای تشکیل ماتریس ورودی ایجاد گردید. مدلهای چندلایه CNN تعریف شدند و با استفاده از چارچوب Caffe DL آموزش داده شدند. امتیازدهی CNN از روش AutoDock Vina روی پایگاه داده با پیشگویی موضع با هدف بینابینی بهتر عمل کرده است، ولیکن عملکرد بدتری برای رتبه بندی درون هدفی مواضع داشت. سایر مطالعاتی که از CNN ها یا DNN ها استفاده کرده بودند نیز منتشر شده اند. هر چند برخی نتایج تشویق کننده با شبکه های معمولی بدست آمده است، روشن نیست که آنها در مقایسه با توابع امتیازدهی که در حال حاضر استفاده می شوند، بطور همیشگی باعث بهبود نتایج می شوند یا خیر.

پایگاه داده های معیار درون کموانفورماتیک

پیشرفتهای سریع انجام شده در حوزه تشخیص تصویری می تواند نه تنها به ظهور الگوریتم تازه نسبت داده شود بلکه به وجود پایگاه داده های اساسی و بزرگ هم مربوط می شود. پایگاه داده استانداردسازی شده به جامعه امکان می دهد تا معیاربندی یا ارزیابی راحت روشهای یادگیری ماشینی توسعه یافته را انجام دهند. الگوریتم سالیانه ImageNet Large Scale Visual Recognition Competition یا ILSVRC شاهد تولد بسیاری معماری های CNN تاثیرگذار بوده است.

هر چند چندین پایگاه داده کموانفورماتیک منبع باز در دست می باشد، اثر آنها روی توسعه روش یادگیری ماشینی همچنان به یمن اندازه محدود آن پایگاه داده ها، فقدان شیوه های گوناگون تقسیم بندی مجموعه های آموزشی و تست و از همه مهمتر کمبود پلتفرم ارزیابی استاندارد برای الگوریتم های جدید مطرح شده از محدودیت برخوردار است. با الهام از WordNet و ImageNet، Wu و همکارانش به معرفی پایگاه داده های MoleculeNet با اداره تعدادی مجموعه های گوناگون از جمله اصول مکانیک کوانتوم، پایگاه داده های فیزیوشیمیایی، بیوفیزیک، و فیزیولوژیکی و گسترش مجموعه نرم افزاری برای اجرای چندین نمایش مولکولی معلوم و الگوریتم های یادگیری ماشینی پرداختند. MoleculeNet طبق بسته نرم افزاری منبع باز

DeepChem ساخته شده و دسترسی آسانی را به برخی الگوریتم های DL موجود در DeepChem فراهم کرده است. این امر تا حد زیادی باعث تسهیل مقایسه و ایجاد الگوریتم های یادگیری ماشینی تازه در آینده شده است.

کاربرد یادگیری عمیق در تحلیل تصویربرداری بیولوژیکی

در پروسه کشف دارو، تصویربرداری بیولوژیکی و تحلیل تصویری وسیعا در مراحل مختلف از R&D پیش بالینی آزمونهای بالینی استفاده شده است. تصویربرداری باعث می شود که دانشمندان فنوتیپ ها و رفتارهای میزبان (انسان و حیوان)، اندامها، بافت ها و سلولها و محتوای درون سلولی را ببینند. از طریق آنالیز تصویر دیجیتالی، بیولوژی پنهان و آسیب شناسی زیرین و نیز مکانیسم عملکرد دارو آشکار می شود. مثالهایی از مودالیتته تصویربرداری همان تصاویر میکروسکوپی نشان دار و غیرنشان دار فلورسنت ، توموگرافی کامپیوتری CT اسکن، توموگرافی صدور پوزیترون PET، MRI ، تصویربرداری آسیب شناختی بافتی و تصویربرداری توده طیف سنجی MSI می باشند. روش DL نیز راهی به سوی موفقیت در تحلیل تصویر بیولوژیکی باز کرده است و بسیاری مطالعات یک عملکرد عالی را در مقایسه با طبقه بندی کننده های کلاسیک گزارش کرده اند.

برای تصاویر میکروسکوپی، CNNها برای قسمت بندی و تعیین نوع بیشتر سلولهای نشاندار فلورسانت منفرد و نیز تصاویر بدون برجسب از میکروسکوپ افتراقی فازی استفاده شده اند. سایر کارهای پرکار قبلی از تنظیمات بالینی، مانند رهگیری سلول و شمارش کلنی ها، را می توان با استفاده از DL خودکارسازی نمود. تصاویر حاصله از آسیب شناسی بافتی معمولا ماهیت پیچیده ای در مقایسه با تصاویر نشاندار فلورسنت به دلیل ریخت شناسی غنی سلولی دارند. مع ذلک، در سطح سلولی، قسمت بندی و طبقه بندی سلولهای فردی در بافت سینه و کولون رنگ امیزی شده با رنگ هماتوکسیلین و ائوزین H&E بدست آمده است. در سطح ناحیه بافتی، مناطق تومور در بافت سینه رنگ امیزی شده H&E از طریق DL شناسایی شده بود، در صورتیکه دستجات بیشتر لکوسیت و بافت چربی می تواند نیز تشخیص داده شود. فراتر از قسمت بندی تصویر پایه، DL در واقع برای تشخیص بافت آسیب شناسی با رنگ امیزی H&E و بافت رنگ امیزی با ایمنوهیستوشیمی استفاده شده است.

بکارگیری DL نیز برای تحلیل تصویربرداری CT، MRT و PET بکار رفته است. در کنار کاربرد مشهور قطعه بندی تصویر و طبقه بندی آن، استفاده هایش نیز در بازیابی تصویر مبتنی بر محتوا نشان داده شده است و گزارش شده است که روشهای DL از روشهای مشهور ISOMAP و Elastic Net هم بهتر است.

برای MSI نوظهور، مشابه با کاربرد DL در آسیب شناسی بافتی، تعیین نمونه تومور می تواند با روش MSI بازجذب/یونیزاسیون لیزری با کمک ماتریس رزولاسیون بالا یا MALDI اجرا بشود. بادرنگیری اینکه MSI می تواند اطلاعات متابولیکی یک بافت را رویت کند، مناطق درونی یک تومور با ناهمگنی متابولیکی حاصل از MSI ی یونیزاسیون پراکنش الکترونی بازجذب یا DESI می تواند قبلا از طریق DL شناسایی گردد. نهایتاً، در یک ناحیه تصویربرداری غیرمعمول: سیتومتری جریان، DL باعث طبقه بندی سلولی در زمان واقعی برای کاربردهای خیلی سطح بالا می شود. آموزش DNNها برای تصویربرداری کاری زمانبر است و مستلزم پردازش پیچیده GPU می باشد. بعلاوه، در زمینه غربالگری تصویربرداری سطح بالا، مجموعه آموزشی با کیفیت خوب نادر است. از اینرو خصوصیات تصویر آموزش دیده از صحنه های طبیعی و سایر پایگاه های داده قرض گرفته می شود تا قطعه بندی تصویر بیولوژیکی و دسته بندی آن انجام شود و عملکردهای مستحکمی هم گزارش گردیده است.

پیشرفت آتی یادگیری عمیق در کشف دارو

روشهای یادگیری ماشینی و DL نیاز به پایگاه داده های بزرگ برای آموزش دارد، ولیکن، مغز انسان توانایی یادگیری از طریق تنها چند مثال را دارد. بنابراین اینکه چگونه با تنها مقدار اندکی داده های موجود یادگیری انجام بشود یکی از داغ ترین مباحث یادگیری ماشینی است. یک مثال DL از بهره برداری از داده های کمکی برای بهبود مدل با تنها چند نقطه داده همان شبکه های یادگیری می باشد که به عنوان نوعی از یادگیری با یک عکس مطرح شده است. نتایج بهبودیافته زمانی بدست آمد که داده های کمکی اضافه گردید. روشهایی مانند یادگیری با یک عکس مربوط به کشف دارو می باشد، که شیمی دانان پزشکی اغلب روی اهداف تازه با داده های محدود موجود کار می کنند. Altae-Tran و همکارانش از روش LSTM روی پایگاه داده انفومارتیک های شیمیایی برای ساخت مدلهایی با مجموعه آموزشی خیلی کوچک استفاده کردند و نتایج نویدبخشی گزارش گردید. همین اواخر، یک نوع جدید معماری در DL استفاده گردید: یعنی شبکه های عصبی با حافظه افزایشی. اولین نسخه همان ماشین Turing عصبی است. این معماری به نحو معنی داری با یک کامپیوتر عصبی افتراقی

DNC بهبود یافته است. DNCها برای چندین مسئله نظیر سیستم های پاسخ گویی به سوال و یافتن کوتاهترین مسیر در گراف ها بکار بسته شده اند. ولیکن این معماری های پیشرفته تر تا کنون در کشف دارو بکار نرفته بودند.

اظهارات نتیجه گیری

یادگیری ماشینی از اواخر دهه 90 میلادی در کشف دارو کاربرد داشته است و خود را وسیله مفیدی در کشف دارو به اثبات رسانیده است. یک بسط تازه از جعبه ابزار یادگیری ماشینی شامل DL می باشد. در مقایسه با سایر روشها، DL یک معماری خیلی قابل انعطاف تر است از اینرو این امکان هست که یک معماری NN سازگار برای یک مسئله خاص بوجود آید. یک معایب این روش آنست که DL به طور کلی مستلزم مجموعه آموزشی خیلی بزرگی است. یک سوال مرتبط آنست که آیا DL نسبت به سایر روشهای یادگیری ماشینی برتری دارد یا خیر؟ ما بر این باوریم که همچنان خیلی زود است که هیچ گونه نتیجه گیری محکمی بنماییم، و نتایج تا کنون نشان داده است که روش DL برای کارهای خاصی مانند تحلیل تصویری برتری دارد و برای طراحی مولکولی و پیشگویی های واکنش مولکولی جدید خیلی مفید می باشد. برای کارهایی مانند توصیفگرهای ورودی ساختارمند، DL ظاهرا دست کم در سطح یکسانی با سایر روشهای دیگر عمل می کند. مرتبط ترین مثال پیشگویی فعالیت زیستی است که در آن DL ظاهرا به عملکرد بهتری در کل از طریق یادگیری چندوظیفه ای دست می یابد. ولیکن سایر روشهای یادگیری ماشینی نیز بهبود یافته اند. یک مثال همان روش XGBoost می باشد که بر رقابت Kaggle برای داده های ورودی ساختارمند پس از ابداعش غلبه کرده است. با اینحساب، در عمل انتخاب روش استفاده شده در پیشگویی فعالیت زیستی ممکن است بستگی به روشی داشته باشد که مدلساز با آن بیشترین آشنایی را دارد. اگر روشهای یادگیری ماشینی تقریبا به همان صحت دست یابند، محدوده آنچه می تواند با مدل یادگیری ماشینی حاصل آید می تواند بستگی به عدم یقین آزمایشی برای داده ها و اندازه بانک داده داشته باشد تا الگوریتم ویژه مورد استفاده.

References

- 1 National Security Agency statement. Available at: <https://www.nsa.gov/news-leaders/press-room/statements/2013-08-09-the-nsa-story.shtml>
- 2 Gantz, J. and Reinsel, D. (2011) Extracting value from chaos. Available at: <https://www.emc.com/collateral/analyst-reports/ids-extracting-value-from-chaos-ar.pdf>
- 3 Gantz, J. and Reinsel, D. (2010) The digital universe decade – are you ready? Available at: <https://www.emc.com/collateral/analyst-reports/ids-digital-universe-are-you-ready.pdf>
- 4 Howard, J. (2013) The business impact of deep learning. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135
- 5 Top Strategic Technology Trends for 2018. Available at: <http://www.gartner.com/technology/research/top-10-technology-trends/>
- 6 Papadatos, G. et al. (2015) Activity, assay and target data curation and quality in the ChEMBL database. *J. Comput. Aided Mol. Des.* 29, 885–896
- 7 Kim, S. et al. (2016) PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1202–D1213
- 8 Gibson, M.E. et al. (2016) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* 44, D1045–D1053
- 9 Cortes, C. and Vapnik, V. (1995) Support-vector networks. *Mach. Learn.* 20, 273–297
- 10 Salt, D.W. et al. (1992) The use of artificial neural networks in QSAR. *Peptid. Sci.* 36, 161–170
- 11 Ho, T.K. (1998) The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 832–844
- 12 Ammad-Ud-Din, M. et al. (2016) Drug response prediction by inferring pathway-response associations with kernelized Bayesian matrix factorization. *Bioinformatics* 32, 1455–1463
- 13 Ching, T. et al. (2017) Opportunities and obstacles for deep learning in biology and medicine. *bioRxiv* <http://dx.doi.org/10.1101/142760>
- 14 Goh, G.B. et al. (2017) Deep learning for computational chemistry. *J. Comput. Chem.* 38, 1291–1307
- 15 Gawehn, E. et al. (2016) Deep learning in drug discovery. *Mol. Inf.* 35, 3–14
- 16 Mamoshina, P. et al. (2016) Applications of deep learning in biomedicine. *Mol. Pharm.* 13, 1445–1454
- 17 Ekins, S. (2016) The next era: deep learning in pharmaceutical research. *Pharm. Res.* 33, 2594–2603
- 18 Baskin, L.I. et al. (2016) A renaissance of neural networks in drug discovery. *Expert Opin. Drug Discov.* 11, 795–799
- 19 McCulloch, W.S. and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biol.* 52, 115–133
- 20 Dreyfus, S.E. (1973) The computational solution of optimal control problems with time lag. *IEEE Trans. Autom. Control* 18, 383–385
- 21 Srivastava, N. et al. (2014) Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958
- 22 Wan, L. et al. (2013) Regularization of neural networks using DeepConnect. In *Proceedings of the 30th International Conference on Machine Learning*, (Vol. 28) (Sanjoy, D. and David, M., eds) in pp. 1058–1066, PMLR
- 23 Nair, V. and Hinton, G.E. (2010) Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Omnipress, pp. 807–814
- 24 TensorFlow™. Available at: <https://www.tensorflow.org/>
- 25 Caffe. Available at: <http://caffe.berkeleyvision.org/>
- 26 PYTORCH. Available at: <http://pytorch.org>
- 27 Keras. Available at: <https://keras.io/>
- 28 Theano. Available at: <http://deeplearning.net/software/theano/>
- 29 Lee, H. et al. (2011) Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Commun. ACM* 54, 95–103
- 30 Szegedy, C. et al. (2015) Going deeper with convolutions. *CVPR*, IEEE Computer Society pp. 1–9
- 31 Fernández, S. et al. (2007) An application of recurrent neural networks to discriminative keyword spotting. In *Proceedings of the 17th International Conference on Artificial Neural Networks*, Springer-Verlag, pp. 220–229
- 32 Hochreiter, S. and Schmidhuber, J. (1997) Long short-term memory. *Neural Comput.* 9, 1735–1780
- 33 Bengio, Y. (2009) Learning deep architectures for AI. *Found. Trends Mach. Learn.* 2, 1–127
- 34 Kingma, D.P. and Welling, M. (2013) Auto-encoding variational bayes. *ArXiv* 1312.6114
- 35 Ma, J. et al. (2015) Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* 55, 263–274
- 36 Mayr, A. et al. (2016) DeepTox: toxicity prediction using deep learning. *Front. Environ. Sci.* <http://dx.doi.org/10.3389/fenvs.2015.00080>
- 37 Dahl, G.E. et al. (2014) Multi-task neural networks for QSAR predictions. *ArXiv* arXiv:1406.1231
- 38 Ramsundar, B. et al. (2017) Is multitask deep learning practical for pharma? *J. Chem. Inf. Model.* 57, 2068–2076
- 39 Koutsoukas, A. et al. (2017) Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. *J. Cheminformatics* 9, 42
- 40 Gaulton, A. et al. (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100–1107
- 41 Lesislak, E.B. et al. (2017) Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set. *J. Cheminformatics* 9, 45
- 42 Subramanian, G. et al. (2016) Computational modeling of beta-secretase 1 (BACE-1) inhibitors using ligand based approaches. *J. Chem. Inf. Model.* 56, 1936–1949
- 43 Aliper, A. et al. (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharm.* 13, 2524–2530
- 44 NIH LINCS program. Available at: <http://www.lincsproject.org/LINCS/>
- 45 Merkwith, C. and Lengauer, T. (2005) Automatic generation of complementary descriptors with molecular graph networks. *J. Chem. Inf. Model.* 45, 1159–1168
- 46 Lenci, A. et al. (2013) Deep architectures and deep learning in cheminformatics: the prediction of aqueous solubility for drug-like molecules. *J. Chem. Inf. Model.* 53, 1563–1575
- 47 Xu, Y. et al. (2013) Deep learning for drug-induced liver injury. *J. Chem. Inf. Model.* 55, 2085–2093
- 48 Morgan, H.L. (1965) The generation of a unique machine description for chemical structures—a technique developed at Chemical Abstracts Service. *J. Chem. Doc.* 5, 107–113
- 49 Duvenaud, D. et al. (2015) Convolutional networks on graphs for learning molecular fingerprints. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, MIT Press, pp. 2224–2232
- 50 Kearnes, S. et al. (2016) Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided Mol. Des.* 30, 595–608
- 51 Xu, Y. et al. (2017) Deep learning based regression and multiclass models for acute oral toxicity prediction with automatic chemical feature extraction. *J. Chem. Inf. Model.* 57, 2672–2685
- 52 Li, J. et al. (2017) Learning graph-level representation for drug discovery. *ArXiv* arXiv:1709.03741
- 53 Culey, C.W. et al. (2017) Convolutional embedding of attributed molecular graphs for physical property prediction. *J. Chem. Inf. Model.* 57, 1757–1772
- 54 Güllmer, J. et al. (2017) Neural message passing for quantum chemistry. *ArXiv* arXiv:1704.01212
- 55 Li, Y. et al. (2015) Gated graph sequence neural networks. *ArXiv* arXiv:1511.05493
- 56 Kipf, T.N. and Welling, M. (2016) Semi-supervised classification with graph convolutional networks. *ArXiv* arXiv:1609.02907
- 57 Bjerrum, E.J. (2017) SMILES enumeration as data augmentation for neural network modeling of molecules. *ArXiv* arXiv:1703.07076
- 58 Goh, G.B. et al. (2017) Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed QSAR/QSPR models. *ArXiv* arXiv:1706.06689
- 59 Goh, G.B. et al. (2017) How much chemistry does a deep neural network need to know to make accurate predictions? *ArXiv* arXiv:1710.02238

- 60 Gómez-Bombarelli, R. *et al.* (2016) Automatic chemical design using a data-driven continuous representation of molecules. *ArXiv* arXiv:1610.02415
- 61 Kadurin, A. *et al.* (2017) drugGAN: an advanced generative adversarial autoencoder model for *de novo* generation of new molecules with desired molecular properties *in silico*. *Mol. Pharm.* 14, 3098–3104
- 62 Goodfellow, I.J. *et al.* (2014) Generative adversarial networks. *ArXiv* arXiv:1406.2661
- 63 Blaschke, T. *et al.* (2017) Application of generative autoencoder in *de novo* molecular design. *Mol. Inf.* <http://dx.doi.org/10.1002/minf.201700123>
- 64 Segler, M.H.S. *et al.* (2018) Generating focussed molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* 4, 120–131
- 65 Yuan, W. *et al.* (2017) Chemical space-mimicry for drug discovery. *J. Chem. Inf. Model.* 57, 875–882
- 66 Jaques, N. *et al.* (2016) Sequence Tutor: conservative fine-tuning of sequence generation models with KL-control. *ArXiv* arXiv:1611.02796
- 67 Leo, A. *et al.* (1971) Partition coefficients and their uses. *Chem. Rev.* 71, 523–616
- 68 Bickerton, G.R. *et al.* (2012) Quantifying the chemical beauty of drugs. *Nat. Chem.* 4, 90–98
- 69 Olivettone, M. *et al.* (2017) Molecular *de-novo* design through deep reinforcement learning. *J. Cheminformatics* 9, 48
- 70 Benlienda, M. (2017) ChemGAN challenge for drug discovery: can AI reproduce natural chemical diversity? *ArXiv* arXiv:1708.08227
- 71 Metz, L. *et al.* (2016) Unrolled generative adversarial networks. *ArXiv* arXiv:1611.02163
- 72 Untertiner, T. *et al.* (2017) Coupled GANs: provably optimal Nash equilibria via potential fields. *ArXiv* arXiv:1708.08819
- 73 Corey, E.J. and Wipke, W.T. (1969) Computer-assisted design of complex organic syntheses. *Science* 166, 178–192
- 74 Coley, C.W. *et al.* (2017) Prediction of organic reaction outcomes using machine learning. *ACS Cent. Sci.* 3, 434–443
- 75 Jin, W. *et al.* (2017) Predicting organic reaction outcomes with Weisfeiler–Lehman network. *ArXiv* arXiv:1709.04555
- 76 Segler, M.H.S. and Waller, M.P. (2017) Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chemistry* 23, 5966–5971
- 77 Segler, M.H.S. *et al.* (2017) Learning to plan chemical syntheses. *ArXiv* arXiv:1708.04202
- 78 Liu, B. *et al.* (2017) Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Central Science* 3, 1103–1113
- 79 Pagadala, N.S. *et al.* (2017) Software for molecular docking: a review. *Biophys. Rev.* 9, 91–102
- 80 Ragosa, M. *et al.* (2017) Protein–ligand scoring with convolutional neural networks. *J. Chem. Inf. Model.* 57, 942–957
- 81 Trott, O. and Ohon, A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461
- 82 Dunbar, J.B., Jr *et al.* (2011) CSAR benchmark exercise of 2010: selection of the protein–ligand complexes. *J. Chem. Inf. Model.* 51, 2036–2046
- 83 Gómez, J. *et al.* (2017) Atomic convolutional networks for predicting protein–ligand binding affinity. *ArXiv* arXiv:1703.10603
- 84 Wallach, I. *et al.* (2015) AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *ArXiv* arXiv:1510.02855
- 85 Pereira, J.C. *et al.* (2016) Boosting docking-based virtual screening with deep learning. *J. Chem. Inf. Model.* 56, 2495–2506
- 86 Russakovsky, O. *et al.* (2015) ImageNet large scale visual recognition challenge. *Int. J. Comput. Vision* 115, 211–252
- 87 Mysinger, M.M. *et al.* (2012) Directory of Useful Decoys, Enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* 55, 6582–6594
- 88 Sun, J. *et al.* (2017) ExCAPE-DB: an integrated large scale dataset facilitating Big Data analysis in chemogenomics. *J. Cheminformatics* 9, 17
- 89 Miller, G.A. (1995) WordNet: a lexical database for English. *Commun. ACM* 38, 39–41
- 90 Li, F.-F. *et al.* (2009) ImageNet: constructing a large-scale image database. *J. Vision* 9, 1057
- 91 Wu, Z. *et al.* (2017) MoleculeNet: a benchmark for molecular machine learning. *ArXiv* arXiv:1703.00564
- 92 DeepChem package. Available at: <https://deepchem.io/>
- 93 Angermueller, C. *et al.* (2016) Deep learning for computational biology. *Mol. Syst. Biol.* 12, 878
- 94 Kraus, O.Z. *et al.* (2015) Classifying and segmenting microscopy images using convolutional multiple instance learning. *ArXiv* arXiv:1511.05286
- 95 Ronneberger, O. *et al.* (2015) U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 234–241
- 96 Ning, Y. *et al.* (2005) Toward automatic phenotyping of developing embryos from videos. *IEEE Trans. Image Process.* 14, 1360–1371
- 97 Ferrari, A. *et al.* (2015) Bacterial colony counting by convolutional neural networks. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, IEEE. pp. 7458–7461
- 98 Cireřan, D.C. *et al.* (2013) Mitosis detection in breast cancer histology images with deep neural networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 411–418
- 99 Sirinukunwattana, K. *et al.* (2016) Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans. Med. Imaging* 35, 1196–1206
- 100 Xu, Y. *et al.* (2014) Deep learning of feature representation with multiple instance learning for medical image analysis. *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference*, IEEE. pp. 1626–1630
- 101 Turkki, R. *et al.* (2016) Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples. *J. Pathol. Inf.* 7, 38
- 102 Vandenberghe, M.E. *et al.* (2017) Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer. *Sci. Rep.* <http://dx.doi.org/10.1038/srep45938>
- 103 Litjens, G. *et al.* (2016) Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. Rep.* 6, 26286
- 104 Bar, Y. *et al.* (2015) Deep learning with non-medical training used for chest pathology identification. *Proc. SPIE Vol. 9414* pp. 94140V
- 105 Cheng, J.-Z. *et al.* (2016) Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci. Rep.* 6, 24454
- 106 Cha, K.H. *et al.* (2016) Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets. *Med. Phys.* 43, 1882–1896
- 107 Averbil, M. *et al.* (2016) A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med. Image Anal.* 30, 108–119
- 108 Li, R. *et al.* (2014) Deep learning based imaging data completion for improved brain disease diagnosis. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 305–312
- 109 Liu, S. *et al.* (2014) High-level feature based PET image retrieval with deep learning architecture. *J. Nucl. Med.* 55 (Suppl. 1), 2028
- 110 Behrmann, J. *et al.* (2017) Deep learning for tumor classification in imaging mass spectrometry. *ArXiv* arXiv:1705.01015
- 111 Inglese, P. *et al.* (2017) Deep learning and 3D-DHSI imaging reveal the hidden metabolic heterogeneity of cancer. *Chem. Sci.* 8, 3500–3511
- 112 Chen, C.L. *et al.* (2016) Deep learning in label-free cell classification. *Sci. Rep.* 6, 21471
- 113 Zhang, W. *et al.* (2016) Deep model based transfer and multi-task learning for biological image analysis. *IEEE Trans. Big Data* <http://dx.doi.org/10.1109/BIODATA.2016.2573280>
- 114 Vinyals, O. *et al.* (2016) Matching networks for one shot learning. *ArXiv* arXiv:1606.04080
- 115 Allae-Tran, H. *et al.* (2017) Low data drug discovery with one-shot learning. *ACS Cent. Sci.* 3, 283–293
- 116 Graves, A. *et al.* (2016) Hybrid computing using a neural network with dynamic external memory. *Nature* 538, 471–476
- 117 Chen, T. and Guestrin, C. (2016) XGBoost: a scalable tree boosting system. *ArXiv* 1603. arXiv:1603.02754
- 118 A Kaggle Master Explains Gradient Boosting. Available at: <http://blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting/>

برای خرید فرمت ورد این ترجمه، بدون واتر مارک، اینجا کلیک نمایید.

این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

لیست مقالات ترجمه شده ✓

لیست مقالات ترجمه شده رایگان ✓

لیست جدیدترین مقالات انگلیسی ISI ✓

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی