

سلام، موضوع ارائه‌ی من «یادگیری عمیق در علوم داده‌های زیست‌پزشکی» هست.
در این مقاله، نویسنده یعنی Pierre Baldi بررسی کرده که چطور یادگیری عمیق به ابزاری قدرتمند برای تحلیل داده‌های زیستی و پزشکی تبدیل شده.

از دهه‌ی ۱۹۸۰، شبکه‌های عصبی کم کم رشد کردند تا به نسل جدیدی از الگوریتم‌ها برسن که امروز بهشون می‌گیم یادگیری عمیق (Deep Learning).
ترکیب سه عامل باعث جهش بزرگ این حوزه شد:
۱. در دسترس بودن داده‌های عظیم (Big Data)
۲. افزایش قدرت محاسباتی GPU (و رایانش ابری)
۳. ابزارهای نرم‌افزاری متن‌باز مثل TensorFlow.

بخش اول: داده‌های زیست‌پزشکی

داده‌های زیست‌پزشکی خیلی متنوع‌اند از مولکول‌ها و ژن‌ها گرفته تا تصاویر پزشکی و پرونده‌های سلامت بیماران.

اما این داده‌ها مشکلات خاصی دارن:

- بعضی خیلی زیادن (مثل داده‌های ژنومی)،
- بعضی خیلی کم و محرمانه‌ان (مثل داده‌های بالینی)،
- و بعضی اصلاً ساختار ثابتی ندارن (مثل گراف‌های مولکولی).

بنابراین، مدل‌های یادگیری عمیق باید انعطاف‌پذیر باشن و بتونن با داده‌های مختلف کار کنن.
برای داده‌های کم، از روش‌هایی مثل «Dropout»، «افزایش داده‌ها» یا «یادگیری نیمه‌نظرارتی» استفاده می‌شه تا مدل بهتر آموزش ببینه.

بخش دوم: معماری‌ها و الگوریتم‌ها

مدل‌های یادگیری عمیق انواع مختلفی دارند:

- شبکه‌های پیش‌خور(Feedforward) : برای طبقه‌بندی‌های ساده
- شبکه‌های بازگشتی(RNN, LSTM) : برای داده‌های ترتیبی مثل DNA یا سیگنال‌ها
- شبکه‌های کانولوشنی(CNN) : برای تصاویر
- خودرمندگزارها(Autoencoders) : برای کاهش ابعاد و حذف نویز
- شبکه‌های مولد(GAN) : برای ساخت داده‌های مصنوعی

مدل‌ها از طریق الگوریتم (پس‌انتشار خطای Backpropagation) آموزش می‌بینند و با «نزول گرادیان» خطای را کم می‌کنند.

به محض اینکه شبکه چند لایه‌ی پنهان داشته باشد، بهش می‌گیم شبکه‌ی عمیق.

بخش سوم: کاربرد در شیمی‌انفورماتیک

در شیمی‌انفورماتیک، هدف اینه که ویژگی‌های مولکول‌ها و واکنش‌های شیمیایی پیش‌بینی بشه. برای این کار، مولکول‌ها به شکل گراف یا رشته‌های SMILES به شبکه داده می‌شن. مدل‌های CNN و RNN یاد می‌گیرن که مثلاً:

• یک مولکول سمی هست یا نه،

• انحلال‌پذیری یا میل ترکیبی‌اش چقدر،

• و در چه واکنش‌هایی شرکت می‌کنه.

اما چون داده‌های تجربی در شیمی محدوده، هنوز این مدل‌ها جای پیشرفت دارند.

بخش چهارم: پروتئومیکس

در این بخش، مدل‌ها برای پیش‌بینی ساختار و عملکرد پروتئین‌ها به کار می‌رن.
از توالی اسیدهای آمینه (Sequence) می‌خوان بفهمن:

- ساختار ثانویه (هلیکس یا ورقه) چیه،
- کدوم آمینواسیدها در فضا به هم نزدیکان (Contact Map)،
- و پروتئین چه کاری انجام می‌دهد.

مدل‌های بازگشتی دوطرفه (BRNN) و CNN توانستن دقیق پیش‌بینی ساختار ثانویه را تا حدود ۹۵٪ افزایش بدن.

همچنانی از شبکه‌ها برای پیش‌بینی ویژگی‌های عملکردی مثل محل پیتید سیگنال، جایگاه اتصال و جهش‌های تأثیرگذار استفاده می‌شه.

بخش پنجم: ژنومیکس و تنسکریپتو میکس

در ژنومیکس، داده‌ها توالی RNA و DNA هستن، و یادگیری عمیق در اینجا فوق العاده موفق بوده.

مدل‌هایی مثل:

• DeepBind: تشخیص جایگاه‌های اتصال فاکتورهای رونویسی

• DeepSEA و Basset: پیش‌بینی اثر جهش‌ها و دسترسی کروماتین

این مدل‌ها مستقیماً از توالی خام DNA الگوهای تنظیمی واقعی یاد می‌گیرن.

در RNA هم از LSTM برای تشخیص جایگاه‌های اسپلیسینگ و از خودرمزگذارها برای تحلیل داده‌های RNA-Seq استفاده می‌شه.

در آینده، ترکیب داده‌های ژنومی، اپی‌ژنتیکی و بیانی با مدل‌های چندوجهی، مسیر پزشکی شخصی‌سازی شده رو باز می‌کنه.

بخش ششم: تصویربرداری زیست‌پزشکی

یادگیری عمیق در تحلیل تصویرهای پزشکی شاید موفق‌ترین کاربردش باشه.
از CNN برای کارهایی مثل:

- تشخیص تومور یا ضایعه
- بخشندی اندامها و بافت‌ها مثل (U-Net)
- بازسازی تصویر CT یا MRI با دز پایین
- و تولید داده‌های مصنوعی با GAN استفاده می‌شه.

مثلاً مدل‌ها می‌تونن سرطان پوست یا ضایعات ریوی رو با دقیقیتی برابر یا بیشتر از پزشک شناسایی کنن.

اما چالش‌هایی مثل کمبود داده، تفاوت بین دستگاه‌ها، و نیاز به تفسیر پذیری همچنان وجود داره.

بخش هفتم: سلامت و مراقبت‌های پزشکی

در پزشکی بالینی، یادگیری عمیق وارد پرونده‌های سلامت الکترونیکی (EHRs) شده.
مدل‌های LSTM می‌تونن:

- خطر بستری مجدد یا مرگ بیماران ICU رو پیش‌بینی کنن،
- احتمال بروز بیماری‌ها رو تخمین بزنن،
- یا حتی درمان مناسب پیشنهاد بدن.

در پزشکی شخصی، مدل‌ها با ترکیب داده‌های ژنتیکی، بالینی و تصویری،
می‌تونن درمان‌ها رو برای هر بیمار شخصی‌سازی کنن.
این دقیقاً هدف پزشکی دقیق (Precision Medicine) هست.

بخش هشتم: مسئله‌ی جعبه‌سیاه

یکی از چالش‌های بزرگ اینه که مدل‌های عمیق خیلی دقیق ولی غیرقابل توضیح‌ان.
پزشکان باید بدونن چرا مدل گفته (این بیمار سرطان دارد).
به همین خاطر، پژوهش‌ها به سمت هوش مصنوعی قابل توضیح (Explainable AI) رفته‌اند.
روش‌هایی مثل:

- SHAP و LIME برای تعیین اهمیت ویژگی‌ها
- نقشه‌های حرارتی در (CNN) دیدن ناحیه‌ی تمرکز مدل در تصویر)
- و مکانیزم توجه (Attention) در مدل‌های ترتیبی

هدف نهایی اینه که مدل‌ها هم دقیق باشن، هم شفاف و قابل اعتماد.

جمع‌بندی و نتیجه‌گیری

یادگیری عمیق امروز تقریباً در تمام لایه‌های زیست‌پزشکی نفوذ کرده:

- از سطح مولکولی و ژنومی تا
- سطح بافتی و تصویری و در نهایت
- سطح بالینی و تصمیم‌گیری پزشکی.

به گفته‌ی نویسنده، رابطه‌ی میان مغز انسان و یادگیری عمیق دوطرفه است:

(مغز الهام‌بخش یادگیری عمیق بوده، و حالا یادگیری عمیق به ما کمک می‌کند مغز را بهتر بفهمیم).

در نهایت، یادگیری عمیق مسیر را برای پزشکی هوشمند، داده‌محور و شخصی‌سازی‌شده باز کرده است.