

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه شهید بهشتی
دانشکده علوم ریاضی
گروه علوم کامپیوتر

پایان نامه کارشناسی ارشد

عنوان

پیش بینی برهم کنش جفت داروها با رویکرد
سیستم های توصیه گر

نگارش

بهاره لویان

استاد راهنما

دکتر چنگیز اصلاحچی

کلیه حقوق اعم از چاپ و تکثیر، نسخه برداری، ترجمه، اقتباس و ... از
این پایان نامه برای دانشگاه شهید بهشتی محفوظ است.

نقل مطالب با ذکر مأخذ آزاد است.

تقدیم بہ تمام عزیزانم،

بہ پاس یک عمر اخلاص، فداکاری و محبتشان...

سپاس گزارى...

شكر و سپاس خدا را كه بزرگترين اميد و ياور در لحظه لحظه زندگيست.
با تقدير و تشكر از زحمات استاد فرهيخته و فرزانه آقاي دكتر چنگيز اصلاحيچى كه همواره راهنما و
راهگشاي من در به نتيجه رسيدن اين پايان نامه بوده اند.

بهاره لويان

۱۳۹۹

نام خانوادگی دانشجو: لویان

نام: بهاره

عنوان: پیش‌بینی برهم‌کنش جفت داروها با رویکرد سیستم‌های توصیه‌گر

استاد راهنما: دکتر چنگیز اصلاحي

مقطع تحصیلی: کارشناسی ارشد رشته: علوم کامپیوتر گرایش: محاسبات علمی

دانشگاه: شهید بهشتی
تاریخ فارغ‌التحصیلی: ۱۳۹۹
علوم ریاضی
تعداد صفحات: ۸۸

واژگان کلیدی: برهم‌کنش دارو-دارو، شباهت دارویی، ادغام شباهت‌های دارویی، انتخاب ویژگی، اعتبارسنجی متقابل، سیستم توصیه‌گر

چکیده

برهم‌کنش دارو-دارو ممکن است باعث بروز واکنش‌های دارویی نامطلوب یا حتی تاثیر دارویی منفی شود. لذا شناسایی برهم‌کنش دارو-دارو قبل از تجویز چند دارو، بسیار مهم است. شناسایی بالینی برهم‌کنش دارو-دارو به هزینه و زمان زیادی نیاز دارد، رویکردهای محاسباتی غربالگری به‌عنوان گزینه جایگزین، روش بسیار ارزان‌تری را برای شناسایی برهم‌کنش‌های بالقوه در مقیاس بزرگ فراهم می‌کنند. با این وجود اکثر آنها فقط پیش‌بینی می‌کنند که آیا یک دارو بر داروی دیگر برهم‌کنش دارد یا خیر، اما از تاثیرات افزایشده (مثبت) و کاهشده (منفی) آنها غافل می‌شوند. این‌گونه برهم‌کنش‌های جامع سه کلاسه به‌طور تصادفی اتفاق نمی‌افتند و برآمده از ویژگی‌های ساختاری گراف برهم‌کنش دارو-دارو است. از طرفی آشکار ساختن چنین رابطه‌هایی بسیار مهم است زیرا به درک و فهم نحوه عملکرد برهم‌کنش دارو-داروهای مرتبه بالا کمک می‌کند و راهنمای مهمی در تجویز نسخه واحد می‌باشد.

در این کار با استفاده از روابط ساختاری ذاتی بین داروها، مجموعه‌ای از برهم‌کنش دارو-دارو جامع را به عنوان یک ماتریس مشخص در نظر گرفته‌ایم و سپس برای پیش‌بینی‌های افزایشده و کاهشده براساس سیستم‌های توصیه‌گر یک مدل جدید ترکیب ویژگی‌ها و یادگیری عمیق به نام SNF-CNN طراحی کرده‌ایم. از ارزیابی نتایج SNF-CNN در پیش‌بینی برهم‌کنش دارو-دارو مقادیر $AUC = 0.9747 \pm 0.0033$ و $AUPR = 0.9666 \pm 0.0045$ برای جفت داروهای کاهشده، مقادیر $AUC = 0.9686 \pm 0.0028$ و $AUPR = 0.8221 \pm 0.0184$ برای جفت داروهای افزایشده و

مقادیر $AUC = 0.9714 \pm 0.0040$ و $AUPR = 0.9480 \pm 0.0083$ برای عدم برهم‌کنش‌ها بدست آمد. که در مقایسه با مقالات برتر در این زمینه، برتری خود را نشان داد. شایان ذکر است این رویکرد جدید نه تنها قادر به پیش‌بینی برهم‌کنش دارو – دارو جامع سه کلاس است بلکه برهم‌کنش دارو – دارو دو کلاس معمولی را نیز پیش‌بینی می‌کند.

منابع اصلی:

1. Shi J-Y, Huang H, Li J-X, Lei P, Zhang Y-N, Dong K, Yiu S-M (2018) TMFUF: a triple matrix factorization-based unified framework for predicting comprehensive drug–drug interactions of new drugs. BMC Bioinform 19(S14):411.
2. Yu H, Mao K-T, Shi J-Y, Huang H, Chen Z, Dong K, Yiu S-M. Predicting and understanding comprehensive drug–drug interactions via seminonnegative matrix factorization. BMC Syst Biol 2018;12:14.
3. Shi J. Mao K. Yu H. et al. Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization. J Cheminform 2019;11: 28.

پیش‌گفتار

برهم‌کنش دارو- دارو یکی از مهمترین موضوعات در زمینه توسعه دارو و سلامت است. برهم‌کنش دارو- دارو به‌عنوان عوارض جانبی ناخواسته ناشی از مصرف هم‌زمان دو یا چند داروی شناخته می‌باشد که می‌تواند تاثیر مصرف دارو را کاهش دهد و یا مسمومیت غیرمنتظره ایجاد کند. در مواردی که پزشک چندین دارو را به‌طور هم‌زمان برای بیمار تجویز کند، ممکن است برهم‌کنش دارو- دارو عوارض جانبی جبران ناپذیری ایجاد کند. تاثیر داروها بر روی یکدیگر ممکن است به بیماری‌های دیگر یا حتی مرگ منجر شود. این عوارض جانبی به‌ویژه در افراد بزرگسال و بیماران سرطانی که روزانه مقدار زیادی دارو مصرف می‌کنند، بسیار قابل توجه است.

تعداد زیادی از تحقیقات حال حاضر در مطالعات بالینی بر روی برهم‌کنش دارو- دارو متمرکز هستند. رویکردهای تجربی و سنتی برای آزمودن و پیدا کردن برهم‌کنش دارو- دارو از روش‌های آزمایشگاهی و بالینی استفاده می‌کنند، که این روند با چالش‌های بسیاری روبروست: هزینه‌ی مالی و زمانی زیادی را می‌طلبد، رفاه حال حیوانات نادیده گرفته می‌شود و تعداد متقاضیان شرکت در طرح‌های آزمایشی کم هستند. برای کمک به کاهش هزینه‌ها و تجزیه و تحلیل تعداد بیشتری از برهم‌کنش‌ها دارو- دارو به روشی جامع و خودکار برای پیش‌بینی نیاز است. رویکردهای محاسباتی، جایگزین امیدوارکننده‌ای برای کشف و تشخیص برهم‌کنش دارو- دارو در مقیاس وسیع هستند که به‌تازگی از طرف مراکز تحقیقاتی و صنعت مورد توجه قرار گرفته‌اند.

یکی از مسائل مهم برای ارائه روش پیش‌بینی برهم‌کنش دارو- دارو، استفاده از داده‌هایی مناسب است که بتواند مقایسه‌ای عادلانه با سایر روش‌ها داشته باشد. در این پایان‌نامه ما از مجموعه داده‌ای شامل دو نوع برهم‌کنش کاهنده و افزایشنده برای بررسی و مقایسه‌ی آن با سایر روش‌ها استفاده کرده‌ایم.

سیستم‌های توصیه‌گر با اولین ظهورشان در زمینه‌ی پالایش گروهی، حوزه تحقیقاتی مهمی را در اواسط دهه ۱۹۹۰ فراهم نمودند. در این پایان‌نامه الگوریتمی برای پیش‌بینی برهم‌کنش دارو- دارو بر مبنای سیستم‌های توصیه‌گر ارائه شده است.

فهرست مطالب

آ	پیشگفتار
ج	فهرست تصاویر
خ	فهرست جداول
۱	۱ مقدمه
۱	۱.۱ دارو
۱	۲.۱ داروشناسی
۲	۱.۲.۱ برهم‌کنش فارماکوکینتیک
۳	۲.۲.۱ برهم‌کنش فارماکودینامیک
۴	۳.۱ پایگاه داده‌های دارو
۴	۱.۳.۱ DrugBank
۵	۲.۳.۱ KEGG
۶	۳.۳.۱ TWOSIDES
۷	۴.۳.۱ FAERS
۷	۵.۳.۱ SIDER
۷	۶.۳.۱ OFFSIDES
۸	۷.۳.۱ PubChem
۸	۴.۱ ویژگی‌های داروها
۹	۱.۴.۱ ساختار شیمیایی

۹	عوارض جانبی	۲۰۴۰۱
۱۰	آنزیم	۳۰۴۰۱
۱۰	منتقل کننده	۴۰۴۰۱
۱۱	پروتئین هدف	۵۰۴۰۱
۱۱	نشانگان	۶۰۴۰۱
۱۱	برهم کنش غذا - دارو	۷۰۴۰۱
۱۱	ویژگی های سه بعدی	۸۰۴۰۱
۱۱	تعریف مسئله	۵۰۱
۱۲	برچسب برهم کنش ها	۱۰۵۰۱
۱۲	ساختار پایان نامه	۶۰۱
۱۵	بررسی تاریخچه، مفاهیم اولیه و الگوریتم های پیش بینی برهم کنش دارو- دارو	۲
۱۵	پیشینه تحقیق	۱۰۲
۱۷	آماده سازی داده	۲۰۲
۱۷	محاسبه شباهت های دارویی	۱۰۲۰۲
۱۸	ترکیب شباهت های شبکه ای	۲۰۲۰۲
۱۹	سامانه های توصیه گر	۳۰۲
۲۳	فیلترینگ همکارانه	۱۰۳۰۲
۲۵	فیلترینگ مبتنی بر محتوا	۲۰۳۰۲
۲۵	فیلترینگ مبتنی بر دانش	۳۰۳۰۲
۲۶	روش های ترکیبی	۴۰۳۰۲
۲۶	یادگیری عمیق	۴۰۲
۲۷	شبکه های عصبی کانولوشن	۱۰۴۰۲
۲۹	ماشین بولتزمن محدود شده	۲۰۴۰۲
۳۰	خودرمنزنگار	۳۰۴۰۲
۳۱	کدگذاری تنک	۴۰۴۰۲
۳۲	الگوریتم های پیشین در مسئله ی پیش بینی سه کلاسه برهم کنش دارو - دارو	۵۰۲

۳۳	چارچوب یکپارچه مبتنی بر تجزیه ماتریس سه کلاسه	۱.۵.۲
۳۵	پیش بینی برهم کنش دارو- دارو بر اساس تجزیه ماتریس غیر منفی	۲.۵.۲
۳۷	تنظیم متعادل تجزیه ماتریس نیمه غیر منفی	۳.۵.۲
۴۱	پیش بینی برهم کنش دارو- داروی با رویکرد سیستم توصیه گر یادگیری عمیق	۳
۴۱	داده ها و ویژگی ها	۱.۳
۴۲	مدل سازی مسئله	۲.۳
۴۳	آماده سازی داده	۳.۳
۴۳	محاسبه ی ماتریس شباهت دارویی	۱.۳.۳
۴۴	ادغام ماتریس های شباهت دارویی	۲.۳.۳
۴۴	تشکیل ماتریس ورودی	۳.۳.۳
۴۴	طراحی سیستم توصیه گر	۴.۳
۴۵	روند انتخاب و آموزش مدل روی برهم کنش های شناخته شده	۱.۴.۳
۴۵	روند اعتبارسنجی جهت انتخاب مدل	۱.۱.۴.۳
۴۶	ارائه ی مدل انتخابی (شبکه عصبی کانولوشن)	۲.۱.۴.۳
۴۹	تشخیص جفت داروهای بدون برهم کنش احتمالی	۲.۴.۳
۵۰	آموزش مدل روی برهم کنش های شناخته شده و ناشناخته	۳.۴.۳
۵۱	روند اعتبارسنجی نتایج مدل پیش بینی برهم کنش داروها	۱.۳.۴.۳
۵۱	مدل نهایی پیش بینی برهم کنش داروها	۲.۳.۴.۳
۵۵	بررسی نتایج و نتیجه گیری	۴
۵۵	روش جامع اعتبارسنجی متقابل k برابری	۱.۴
۵۷	معرفی معیارهای سنجش	۲.۴
۵۹	مقایسه ی نتایج	۳.۴
۶۱	نتیجه گیری و جمع بندی	۴.۴
۶۳	کارهای آینده و مسائل باز	۵
۶۵	واژه نامه فارسی به انگلیسی	

فهرست تصاویر

۱۰.۱	قسمتی از نمایش اطلاعات داروی دیازپام در پایگاه داده‌ی DrugBank	۴
۲۰.۱	نمایش قسمتی از اطلاعات برهم‌کنش داروی دیازپام با دیگر داروها	۵
۳۰.۱	صفحه اول پایگاه داده KEGG	۶
۴۰.۱	صفحه اول پایگاه داده SIDER	۸
۵۰.۱	نمایش نحوه ساخت اثر انگشت دارو	۹
۶۰.۱	ساختار شیمیایی داروی دیازپام	۱۰
۱۰.۲	انواع شباهت بین بردارهای دودویی	۱۸
۲۰.۲	انواع شباهت بین بردارهای دودویی [۴۶]	۱۹
۳۰.۲	شمای کلی فرایند توصیه در سامانه‌های توصیه‌گر	۲۲
۴۰.۲	تقسیم‌بندی سامانه‌های توصیه‌گر [۴۷]	۲۲
۵۰.۲	نمای کلی از روش‌های فیلترینگ همکارانه	۲۳
۶۰.۲	شرح مصور از تجزیه ماتریس	۲۴
۷۰.۲	نمای کلی از روش‌های فیلترینگ مبتنی بر محتوا	۲۵
۸۰.۲	دسته‌بندی روش‌های یادگیری عمیق و چند مثال از شبکه‌های هر دسته [۴۸]	۲۸
۹۰.۲	طرح کلی از معماری یک شبکه عصبی کانولوشن [۴۸]	۲۹
۱۰۰.۲	طرح کلی از معماری یک ماشین بولتزمن محدود شده	۲۹
۱۱۰.۲	فرآیند کلی یک خودرمنزنگار [۴۸]	۳۰
۱۲۰.۲	پیش‌بینی برهم‌کنش TMFUF [۶۸]	۳۳
۱۳۰.۲	ساختار مدل نظارت شده TMFUF [۶۸]	۳۴
۱۴۰.۲	بررسی اجمالی DDINMF [۷۲]	۳۶

۱۵.۲	پیش‌بینی برهم‌کنش جامع در سناریو شروع سرد BRSNMF [۷۱]	۴۰
۱.۳	نحوه‌ی چینش لایه‌های شبکه عصبی تشخیص صفرهای احتمالی	۴۷
۲.۳	وزن‌های قابل یادگیری شبکه‌ی عصبی پیش‌بینی دوکلاسه	۴۸
۳.۳	نمودار توزیع چگالی احتمال کاهنده و افزایشنده	۵۰
۴.۳	چیدمان لایه‌های شبکه عصبی SNF-CNN پیش‌بینی سه‌کلاسه برهم‌کنش	۵۲
۱.۴	موارد صحت و مقادیر تابع خطا برای شماره ایپوک‌های مختلف	۵۶
۲.۴	نمودار صحت و مقادیر تابع خطا برای شماره ایپوک‌های مختلف	۵۷

فهرست جداول

۳۲	مقایسه‌ی روش‌های مختلف شبکه‌های عصبی عمیق [۴۸]	۱.۲
۴۲	جزئیات مجموعه‌ی داده‌های برهم‌کنش [۷۲]	۱.۳
۵۸	حالت‌های ممکن نتایج یک یادگیری ماشین	۱.۴
۶۰	گزارش دسته‌بندی نوع برهم‌کنش	۲.۴
۶۰	گزارش دسته‌بندی حالت سه‌کلاسه برهم‌کنش	۳.۴
۶۱	نتایج الگوریتم SNF-CNN در پیش‌بینی سه‌کلاسه براساس معیارهای AUC و AUPR و بازه‌ی اطمینان آن‌ها	۴.۴
۶۱	مقایسه‌ی نتایج الگوریتم‌های پیش‌بینی سه‌کلاسه براساس معیارهای AUC و AUPR	۵.۴

فصل ۱

مقدمه

در این بخش سعی می‌شود مفاهیم و تعاریف اولیه‌ی مورد نیاز برای ورود به مسئله‌ی اصلی به صورت مختصر تبیین گردد، سپس به تعریف مسئله می‌پردازیم.

۱.۱ دارو

در دانش پزشکی به هر ماده‌ای که برای درمان، تسکین علائم، تشخیص بیماری یا پیش‌گیری از آن به‌کار رود و بر ساختار یا کارکرد ارگانیسم زنده اثر گذارد و پس از ورود به بدن عملکرد بدن را تصحیح کند، دارو^۱ گفته می‌شود. در تعریفی دیگر دارو به ماده‌ای گفته می‌شود که با اثر بر گیرنده‌ای خاص در داخل، خارج یا دیواره‌ی سلول باعث شروع یا مهار عملکردی خاص گردد و قدرت اثر دارو با میزان و تعداد این برهم‌کنش نسبت مستقیم دارد. البته داروهایی که اثر موضعی دارند مانند آنتی‌اسیدها و ضد عفونی‌کننده‌های موضعی و مواد حاجب در این تعریف نمی‌گنجند [۱]. امروزه اطلاعات داروها به طور گسترده در پایگاه‌داده‌های دارو قابل دسترسی می‌باشد که به تعدادی از مهم‌ترین آن‌ها در بخش ۳.۱ اشاره می‌کنیم.

۲.۱ داروشناسی

داروشناسی^۲ دانش بررسی واکنش مواد شیمیایی در سیستم‌های زیستی است. داروشناسی پزشکی^۳ بخشی از داروشناسی که در ارتباط با استفاده از مواد شیمیایی در پیشگیری، تشخیص و درمان بیماری،

¹Drug

²Pharmacology

³Medical Pharmacology

به ویژه در انسان است. دو حوزه اصلی داروشناسی، فارماکودینامیک^۴ و فارماکوکینتیک^۵ هستند. به طور خلاصه فارماکودینامیک مطالعه‌ی اثرات داروها بر بدن و فارماکوکینتیک، اثر و رفتار بدن بر روی داروهاست. به عبارت دیگر فارماکودینامیک به مبحث واکنش‌های دارو در مواجهه با گیرنده‌های زیستی و فارماکوکینتیک به بحث و تعریف جذب، توزیع، متابولیسم و دفع دارو در سیستم‌های زیستی می‌پردازد [۲]. داروشناسی و داروسازی^۶ دو واژه جدا از هم هستند که گاهی به اشتباه به جای هم به کار می‌روند. داروشناسی نوع رفتار دارو و سیستم زیستی (بدن) است ولی داروسازی در رابطه با مواد اولیه، آماده‌سازی، توزیع و مقدار^۷ مصرف داروهای بی‌خطر و موثر است.

۱.۲.۱ برهم‌کنش فارماکوکینتیک

فارماکوکینتیک تأثیر سیستم‌های زیستی بر داروها را نشان می‌دهد. عمده‌ی فرآیندهای دخیل در فارماکوکینتیک جذب، توزیع، متابولیسم و دفع هستند. بیماری‌های مختلف می‌توانند پارامترهای استاندارد فارماکوکینتیک معمول را تغییر دهند. در صورت مشخص شدن پارامترهای فارماکوکینتیک بیمار، تنظیم مقدار مصرف دارو برای یک بیمار خاص قابل محاسبه است [۲]. وقتی دو دارو باهم مصرف می‌شوند، روند جذب، توزیع، متابولیسم یا دفع یکی یا هر دوی آن‌ها ممکن است با رفتار مورد انتظار آن‌ها در حالتی که دارو به تنهایی مصرف می‌شود، متفاوت باشد. چنین تفاوت‌هایی را ناشی از برهم‌کنش‌های فارماکوکینتیک می‌دانیم که عبارتند از:

برهم‌کنش بر اساس جذب:

این امکان وجود دارد که فرآیند جذب^۸ دارو (از دیواره‌ی روده) در صورتی که با داروی دیگری ترکیب شود، کاهش یا افزایش یابد. در فرآیند جذب از دیواره‌ی روده، برخی از داروهای تنگ‌کننده عروق، به محض استفاده (ورود به بدن) با عملکرد موقتی در موضع و با محدود کردن اندازه رگ‌ها باعث کند شدن جذب دیگر داروها می‌شوند. گاهی پزشکان از این پدیده بهره می‌گیرند، برای مثال برخی بی‌حس‌کننده‌های موضعی را با داروهای آلفا-آدرنرژیک (اپی نفرین، نوراپی نفرین، سینفرین) که تنگ‌کننده عروق هستند، برای کند کردن جذب و تداوم اثر دارو در موضع مورد نیاز، به کار می‌برند [۳].

برهم‌کنش بر اساس توزیع:

توزیع^۹ دارو ممکن است تحت تأثیر داروهایی که دارای اثر رقابتی بر جایگاه اتصال پروتئین‌های پلاسما می‌باشند قرار گیرد. به طور مثال داروهای آنتی‌باکتریال سولفونامیدی قابلیت جابجایی داروهایی

⁴Pharmacodynamic

⁵Pharmacokinetic

⁶Pharmacy

⁷Dose

⁸Absorption

⁹Distribution

نظیر متوتروکسات، فنی توئین، سولفونیل اوره‌ها و وارفارین از آلبومین پلاسما را دارند. یک دارو با ایجاد تغییراتی در محیط فیزیکی و در محیط توزیع داروی دیگر، باعث تغییر انتشار داروی دوم می‌شود. برای مثال دیورتیک‌ها می‌توانند از طریق کاهش آب کلی بدن، میزان انتشار آمینوگلیکوزیدها را کاهش دهند و باعث تشدید سمیت دارویی آمینوگلیکوزیدها و لیتیموم بدلیل افزایش غلظت پلاسمایی آنان شوند.

برهم‌کنش بر اساس متابولیسم:

این نوع برهم‌کنش‌ها به‌خوبی اثبات شده‌اند و از اهمیت بالای بالینی برخوردارند. متابولیسم^{۱۰} اکثر داروها در کبد توسط آنزیم سیتوکروم P450 صورت می‌گیرد. متابولیسم یک دارو در صورت تجویز همزمان با داروهایی که اثر القاکنندگی بر این آنزیم‌ها دارند، افزایش می‌یابد و متابولیسم یک دارو در صورت تجویز همزمان با داروهایی که اثر مهارکنندگی بر این آنزیم‌ها دارند، کاهش می‌یابد.

برهم‌کنش براساس دفع:

دفع^{۱۱} بسیاری از داروها از طریق ادرار و سیستم کلیوی صورت می‌گیرد. برای مثال دفع یک دارو در صورت استفاده همزمان با داروهای کاهنده جریان خون کلیوی، کاهش می‌یابد و یا مصرف همزمان یک دارو با داروهای تغییردهنده‌ی PH ادرار، میزان یونیزاسیون آن دارو را کاهش می‌دهد.

۲.۲.۱ برهم‌کنش فارماکودینامیک

ممکن است استفاده همزمان دو دارو منجر به افزایش تاثیرات هر یک از آنها شود یا برعکس ممکن است اثرات یکدیگر را سرکوب کنند. به‌چنین تاثیرات افزایشی یا کاهشی داروها بر روی یکدیگر، برهم‌کنش‌های فارماکودینامیک می‌گوییم که عبارتند از:

برهم‌کنش براساس اثر آنتاگونیسم:

ساده‌ترین نوع برهم‌کنش دارویی است که اغلب قابل پیش‌بینی است. گاهی اوقات جذب یک دارو مانع جذب بعضی داروهای دیگر می‌گردد، این پدیده را آنتاگونیسم یا رقابت‌کنندگی^{۱۲} گویند[۴]. برای مثال داروهای ضد التهابی غیراستروئیدی NSAIDs اثر ضد فشار داروی مهارکننده‌ی ACEIs را از طریق کاهش دفع سدیم، کم می‌کنند.

¹⁰Metabolism

¹¹Elimination

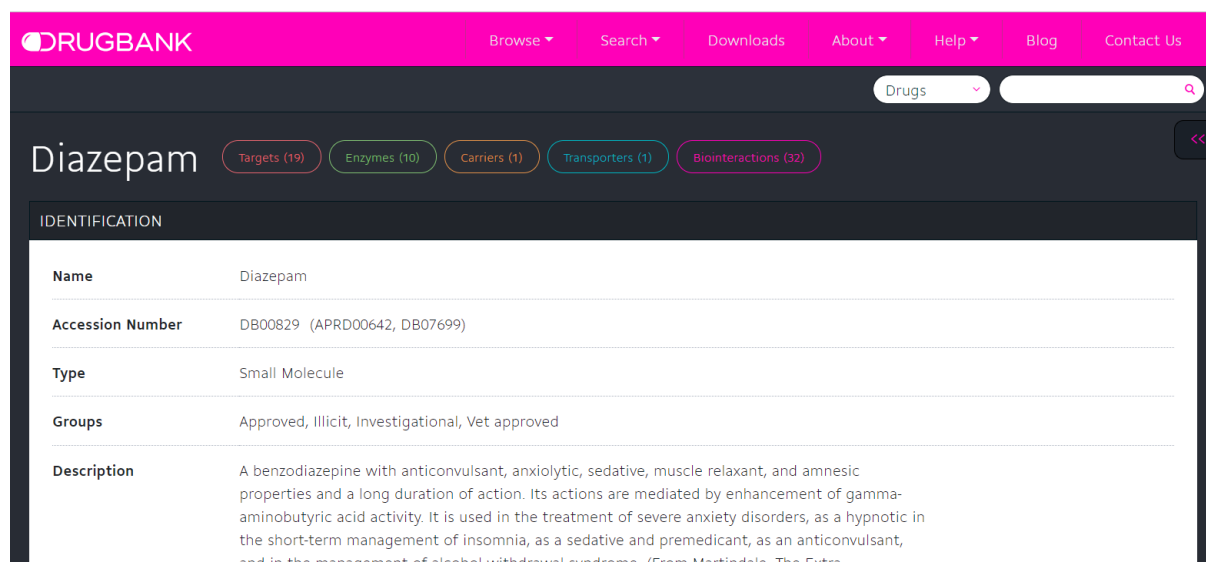
¹²Antagonism

برهم‌کنش براساس اثر سینرژیسم:

گاهی جذب یک دارو باعث تشدید و افزایش شدت جذب داروی دیگر می‌گردد، به این پدیده سینرژیسم یا تشدیدکنندگی^{۱۳} گویند [۵]. جمع‌جبری اثرات دو دارو، ممکن است بر یک گیرنده‌ی خاص عمل نمایند. اگر مجموع اثرات دو دارو با هم از جمع اثر هر دارو به‌تنهایی بیشتر باشد تداخل فوق سینرژیسم یا تشدیدکنندگی می‌باشد.

۳.۱ پایگاه داده‌های دارو

اطلاعات دارویی مورد نیاز برای حل مسائل مختلف مرتبط با دارو، در پایگاه داده‌های معروفی به شرح زیر موجود هستند.



شکل ۱.۱: قسمتی از نمایش اطلاعات داروی دیازپام در پایگاه داده‌ی DrugBank

۱.۳.۱ DrugBank

DrugBank [۹-۶]: یکی از منابع مهم تحقیقاتی بیوانفورماتیک می‌باشد که در آن اطلاعات دارو-هدف^{۱۴}، آنزیم‌های دارو، انتقال دهنده‌های دارو^{۱۵}، عوارض جانبی^{۱۶} دارو، برهم‌کنش‌های شناخته شده‌ی

¹³Synergism

¹⁴Drug-Target

¹⁵Drug Transporters

¹⁶Side Effect

دارو با دیگر داروها و... موجود می‌باشد. در شکل ۱۰.۱ قسمتی از اطلاعات داروی دیازپام^{۱۷} نشان داده شده است و در شکل ۲.۱ قسمتی از اطلاعات برهم‌کنش داروی دیازپام با دیگر داروها نمایش داده شده است.

The screenshot shows the DrugBank website interface. At the top is a pink navigation bar with 'DRUGBANK' and links for Browse, Search, Downloads, About, Help, Blog, and Contact Us. Below this is a search bar with 'Drugs' selected. The main content area is titled 'Drug Interactions' and lists various drug categories: ALL DRUGS, APPROVED, VET APPROVED, NUTRACEUTICAL, ILLICIT, WITHDRAWN, and INVESTIGATIONAL. Under 'ALL DRUGS', there is a sub-section for 'EXPERIMENTAL'. A table displays drug interactions, with columns for 'DRUG' and 'INTERACTION'. The table lists several drugs and their interactions with Diazepam.

DRUG	INTERACTION
(R)-warfarin	The metabolism of (R)-warfarin can be decreased when combined with Diazepam.
(S)-Warfarin	The metabolism of (S)-Warfarin can be decreased when combined with Diazepam.
2,5-Dimethoxy-4-ethylthioamphetamine	The risk or severity of adverse effects can be increased when Diazepam is combined with 2,5-Dimethoxy-4-ethylthioamphetamine.
3-isobutyl-1-methyl-7H-xanthine	The therapeutic efficacy of Diazepam can be decreased when used in combination with 3-isobutyl-1-methyl-7H-xanthine.
3,4-Methylenedioxyamphetamine	The risk or severity of adverse effects can be increased when Diazepam is combined with 3,4-Methylenedioxyamphetamine.
3,5-diiodothyropropionic acid	The metabolism of 3,5-diiodothyropropionic acid can be decreased when combined with Diazepam.

شکل ۲.۱: نمایش قسمتی از اطلاعات برهم‌کنش داروی دیازپام با دیگر داروها

۲.۳.۱ KEGG

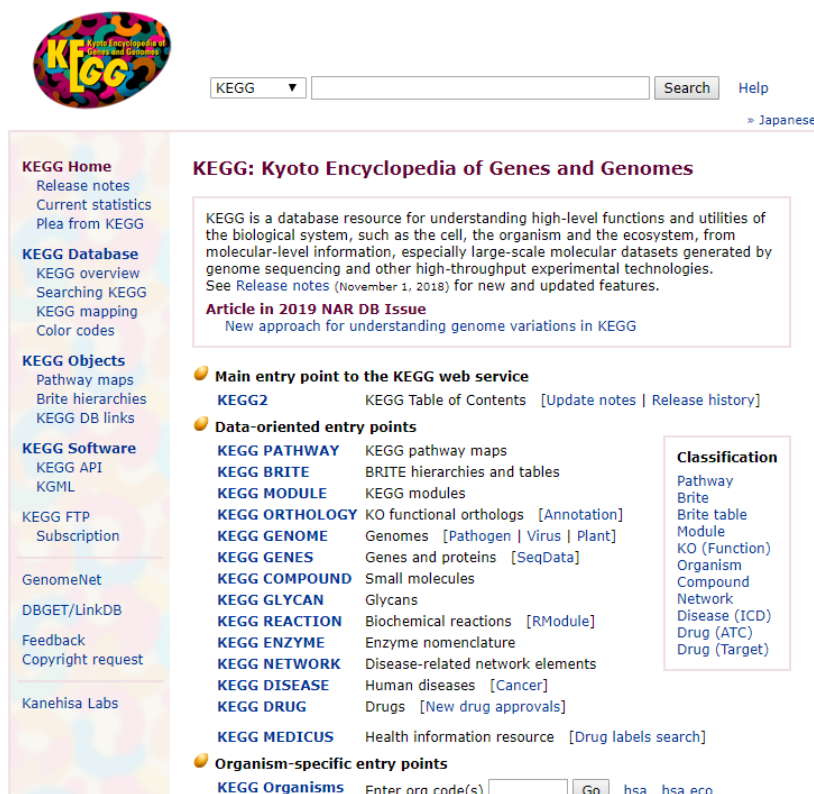
KEGG^{۱۸} [۱۱، ۱۰]: پایگاه داده‌ای مربوط به اطلاعات زیستی است که اولین بار توسط مینورو کانه‌یزا^{۱۹} پروفیسور موسسه‌ی تحقیقات شیمی در دانشگاه کیوتو در سال ۱۹۹۵ هم‌زمان با پروژه ژنوم درست شد. به منظور استفاده از تحلیل کامپیوتری برای تفسیر نتایج حاصل از پروژه ژنوم مینورو شروع به ساخت KEGG Pathway کرد. در این پایگاه داده توسط دانش به‌دست آمده از آزمایش‌های مختلف، نقشه و مسیرهای متابولیکی ترسیم شده‌است و همچنین تعدادی از عملکردهای سلول‌ها و ارگانیزم‌ها به‌صورت جزئی به تصویر کشیده شده‌است. هر نقشه شامل یکسری واکنش بین بیومولکول‌های مختلف می‌باشد که طوری طراحی شده‌است که از آن می‌شود اطلاعاتی در مورد ژن‌ها و پروتئین‌های دخیل در آن یافت. در این پایگاه داده می‌توان مسیرهای مختلف را با هم مقایسه و تحلیل کرد و به‌طور کلی از آن‌ها اطلاعات سودمندی استخراج کرد. بر طبق نظر سازنده، KEGG به صورت یک کامپیوتر ارائه دهنده‌ی سیستم‌های بیولوژیکی می‌باشد که دیگرام‌ها و واحدهای متصل به هم ایجاد می‌کند. دیگرام‌ها شامل ژن‌ها و عملکرد پروتئین‌های مرتبط، واکنش‌های بین مولکولی هستند. همچنین KEGG منبع مناسبی برای مسیرهای

¹⁷Diazepam

¹⁸Kyoto Encyclopedia of Genes & Genomes

¹⁹Minoru Kanehisa

پروتئین^{۲۰} می باشد. در شکل ۳.۱ صفحه‌ی اول این پایگاه داده که یک معرفی از آن می باشد نمایش داده شده است.



شکل ۳.۱: صفحه اول پایگاه داده KEGG

TWOSIDES ۳.۳.۱

TWOSIDES [۱۲]: بانک اطلاعاتی عمومی حاوی عوارض جانبی ناشی از برهم‌کنش دارو-دارو است. این بانک اطلاعاتی حاوی ۸۶۸۲۲۱ ارتباط بین ۵۹۲۲۰ جفت دارو و ۱۳۰۱ عوارض جانبی است. این ارتباطات فقط به مواردی محدود می شوند که به طور واضح به هیچ یک از داروها اختصاص نیافته است (ارتباط‌های تحت پوشش OFFSIDES). این پایگاه داده حاوی ۳۷۸۲۹۱۰ ارتباط مهم دیگر است که، جفت داروها دارای نمره عوارض جانبی بالاتری هستند که با استفاده از نسبت گزارش متناسب^{۲۱}، تعیین می شوند.

²⁰Protein Pathways

²¹Proportional Reporting Ratio (PRR)

FAERS ۴.۳.۱

FAERS^{۲۲}: سیستم گزارش‌دهی رویدادهای ناخواسته‌ی سازمان غذا و دارو پایگاه داده‌ای، حاوی اطلاعات مربوط به عوارض جانبی ارسال شده به سازمان غذا و دارو^{۲۳} است. این پایگاه داده، توسط FDA^{۲۴}، به‌منظور پشتیبانی از برنامه‌ی نظارت بر ایمنی پس از بازاریابی برای داروها و محصولات بیولوژیکی درمانی طراحی شده است. داده‌های استخراج شده از FAERS و TWOSIDES مجموعه داده‌ای است که فقط شامل عوارض جانبی است که به دلیل ترکیب داروها ایجاد می‌شود [۱۴].

SIDER ۵.۳.۱

SIDER [۱۵]: در این پایگاه داده اطلاعات عوارض جانبی داروها و نشانگان دارو وجود دارد. این پایگاه داده حاوی اطلاعاتی در مورد داروهای بازار و واکنش‌های دارویی نامطلوب آن‌ها است. با وجود این‌که SIDER منبع مهمی از عوارض جانبی شناخته شده است، اما اطلاعات آن محدود است [۱۴]. زیرا:

الف) آزمایشات بالینی بر روی جمعیت نسبتاً کمی از بیماران انجام می‌شود و فقط عوارض جانبی متداول با اطمینان بالا در لیست دارویی مشاهده می‌شوند.

ب) از طرفی عارضه‌ی جانبی مشاهده شده در طول آزمایشات بالینی ممکن است اتفاقی باشد و در واقع توسط دارو ایجاد نشده باشد.

لازم به ذکر است، عوارض جانبی که از این پایگاه داده استخراج شده است Labeled Side Effect نامیده می‌شود.

در شکل ۴.۱ نمایی از صفحه‌ی اول پایگاه داده‌ی SIDER مشاهده می‌شود که می‌توان براساس نام دارو یا عارضه‌ی جانبی در این پایگاه داده جستجو نمود.

OFFSIDES ۶.۳.۱

مجموعه‌ای از عوارض جانبی داروها است که حاصل کاوش عوامل مخدوش‌کننده مانند مصرف داروهای همزمان، اطلاعات دموگرافیک بیمار و تاریخچه پزشکی بیمار از پایگاه داده‌ی FAERS است. در این پایگاه داده ۱۳۳۲ دارو و ۱۰۰۹۳ عارضه‌ی جانبی وجود دارد. لازم به ذکر است، عوارض جانبی استخراج شده از OFFSIDES با عنوان Off-Label Side Effect نامگذاری می‌شود [۱۲].

^{۲۲}FDA Adverse Event Reporting System

^{۲۳}Food and Drug Administration (FDA)

^{۲۴}<http://www.fda.gov>

SIDER 4.1 : Side Effect Resource

SIDER contains information on marketed medicines and their recorded adverse drug reactions. The information is extracted from public documents and package inserts. The available information include side effect frequency, drug and side effect classifications as well as links to further information, for example drug-target relations.

Search for drugs or side effects :

type 3 or more characters...

Database statistics

Number of drugs and side effects				
# of SE	# of drugs	# of drug-SE pairs	Pairs with frequency information	
5868	1430	139756	39.9%	

Number of drug-side effect pairs in different frequency ranges					
	frequent (with exact data)	infrequent (with exact data)	rare (with exact data)	postmarketing	total
drug	24562 (23601)	16765 (11426)	11784 (6013)	19265	55730
placebo	7133 (7133)	3294 (3294)	2512 (2512)	0	10748

Version Information

شکل ۴.۱: صفحه اول پایگاه داده SIDER

۷.۳.۱ PubChem

پایگاه داده مربوط به مولکول‌های شیمیایی است که توسط موسسه ملی سلامت در ایالات متحده^{۲۵} ایجاد شده است. این سامانه شامل مولکول‌های کوچکی است که کمتر از ۱۰۰۰ اتم یا ۱۰۰۰ پیوند شیمیایی دارند و اطلاعات ساختار شیمیایی^{۲۶} داروها را در بر دارد [۱۳].

۴.۱ ویژگی‌های داروها

هر دارو می‌تواند به صورت یک بردار ویژگی دودویی، توسط معیارهای مختلفی تعریف شود [۱۶، ۱۷]. قبل از معرفی این ویژگی دو تعریف ارائه می‌شود که در درک بهتر عددی ویژگی‌ها کمک می‌کند.

تعریف ۱.۴.۱. اثرانگشت دارو

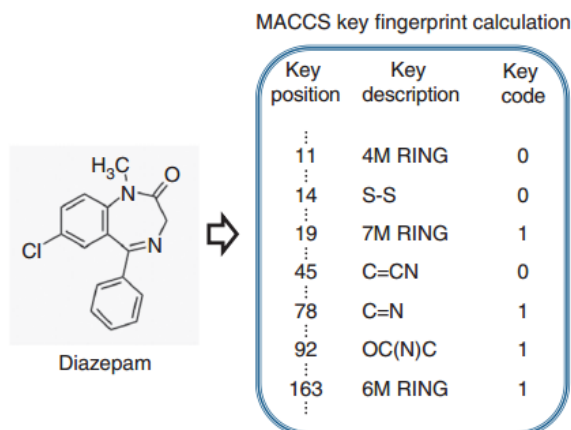
اثر انگشت^{۲۷}، برداری از مقادیر باینری صفر و یک ویژگی‌های یک مولکول است. برای مثال اثر انگشت MACCS key166 نمونه‌ی مشهوری با ۱۶۶ ویژگی زیرساختار است که وجود یا عدم وجود یک

^{۲۵}National Institutes of Health

^{۲۶}Chemical Structure

^{۲۷}Drug Fingerprint

زیرساختار را در این دارو مشخص می‌کند. وجود زیرساختارهایی با کم‌تر از سه اتم اکسیژن در ساختار شیمیایی دارو با یک کد می‌شود. همچنین عدم وجود زیرساختار 4mRing با صفر کد می‌شود که در شکل ۵.۱ به خوبی نمایش داده شده است. به عنوان مثالی دیگر در اثر انگشت عوارض جانبی دارو، وجود یا عدم وجود هر عارضه‌ی جانبی مانند: سردرد، اسهال، حالت تهوع و... برای یک دارو به ترتیب با یک و صفر کد می‌شود.



شکل ۵.۱: نمایش نحوه ساخت اثر انگشت دارو

تعریف ۲.۴.۱. خارج از برچسب

برای توصیف داروی پزشکی به کار می‌رود. در صورتی که دارویی برای درمان بیماری تجویز شود ولی برای درمان آن بیماری از طرف FDA تایید نشده باشد به آن نوع از استفاده، خارج از برچسب^{۲۸} می‌گویند. مثلاً ممکن است دارویی برای معالجه کودک تایید شود اما از آن برای درمان بزرگسالان استفاده شود. نمونه‌ی دیگری از خارج از برچسب در بخش ۲.۴.۱ معرفی می‌شود.

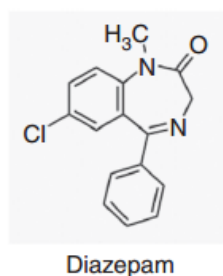
۱.۴.۱ ساختار شیمیایی

۸۸۱ نوع زیرساختار مختلف تعریف شده است که یک دارو ممکن است آن‌ها را داشته یا نداشته باشد. ساختار شیمیایی داروها در پایگاه داده‌های دارو نظیر DrugBank یا PubChem [۱۳] در قالب فایل mol، smiles و sdf وجود دارند. در شکل ۶.۱ ساختار شیمیایی داروی دیازپام نشان داده شده است.

۲.۴.۱ عوارض جانبی

داروها می‌توانند عوارض جانبی ایجاد کنند. برخلاف تصور عموم، که عوارض جانبی دارو را فقط مختص مصرف خودسرانه‌ی دارو می‌دانند، باید دانست که عوارض جانبی بر اثر مصرف همه‌ی انواع داروها ایجاد می‌شوند؛ هم داروهایی که توسط پزشک تجویز شده‌اند و هم داروهایی که بدون نسخه

²⁸Off - Label



شکل ۶.۱: ساختار شیمیایی داروی دیازپام

تهیه می‌شوند. برای مثال، برخی آنتی‌بیوتیک‌ها مثل خانواده‌ی پنی‌سیلین‌ها، می‌توانند در افراد واکنش‌های حساسیتی ایجاد کنند. دانه‌های ریزپوستی^{۲۹}، شایع‌ترین عارضه‌ی واکنش‌های حساسیتی به پنی‌سیلین‌ها و آنافیلاکسی مهم‌ترین یا خطرناک‌ترین عارضه‌ی ثانویه‌ی مصرف پنی‌سیلین‌ها به‌شمار می‌روند. در این جا عوارض جانبی به دو دسته‌ی زیر تقسیم‌بندی می‌شوند:

عوارض جانبی برچسب‌دار^{۳۰}

عوارض جانبی خارج از برچسب^{۳۱}

۳.۴.۱ آنزیم

بسیاری از داروها اثر خود را از طریق انجام واکنش با آنزیم‌ها^{۳۲} اعمال می‌کنند. Cytochrome P450 یک خانواده گسترده از آنزیم‌های هموپروتئینی است که در تمام موجودات زنده وجود دارد و ایزوآنزیم CYP بیشتر متابولیسم داروها را انجام می‌دهد. همچنین فعال‌شدن بسیاری ترکیبات شیمیایی در درون بدن توسط همین ایزوآنزیم‌ها انجام می‌گیرد.

۴.۴.۱ منتقل‌کننده

انتقال دارو یک موضوع حیاتی در پزشکی و درمان است. انتقال داروی کنترل‌شده، دسترسی به دارو را به واسطه جلوگیری از تجزیه در زمان نامناسب، افزایش دریافت دارو، حفظ غلظت دارو در طی درمان به واسطه کنترل سرعت آزادسازی دارو و کاهش عوارض جانبی از طریق هدفمندی دارو به جایگاه و سلول خاص را بهبود می‌بخشد. برای کاهش میزان تجزیه‌ی دارو، جلوگیری از عوارض جانبی مضر، افزایش دسترسی به دارو و تجمع دارو در ناحیه هدف، سیستم‌های متنوع انتقال و هدفمندی دارو در حال پیشرفت است. برای آزادسازی پیوسته‌ی دارو بایستی از پلیمرهایی استفاده کرد که دارو را با سرعتی قابل کنترل

²⁹Skin Rash

³⁰Label Side Effect

³¹Off-Label Side Effect

³²Enzymes

منتشر کنند یا با تجزیه‌ی پلیمر طی زمان آزاد شوند. در بین حامل‌های دارویی می‌توان پلیمرهای محلول، میکروذرات تشکیل شده از پلیمرهای طبیعی و سنتزی غیر محلول و تجزیه پذیر، میکروکپسول‌ها، سلول‌ها، لیپوپروتئین‌ها، لیپوزوم‌ها و میسل‌ها را نام برد. حامل‌ها می‌توانند تجزیه پذیر، القایی (مثلاً حساس به دما)، و حتی با تعبیه آنتی‌بادی خاص علیه ترکیبات ناحیه‌ی مورد نظر هدفمند شوند.

۵.۴.۱ پروتئین هدف

ممکن است دارو به پروتئین‌های مختلفی بچسبد. در طراحی دارو به این نکته توجه می‌شود که از لحاظ هندسی دارو مکمل پروتئین هدف باشد تا بتواند به خوبی به آن متصل شود.

۶.۴.۱ نشانگان

به بیماری‌هایی که دارو برای آنها تجویز می‌شود نشانگان^{۳۳} گفته می‌شود، ممکن است یک دارو در درمان چند نوع بیماری موثر باشد.

۷.۴.۱ برهم‌کنش غذا - دارو

برهم‌کنش غذا - دارو^{۳۴} در اثر واکنش بین دارو و مواد خاص خوردنی یا نوشیدنی به وجود می‌آید. مثلاً آهن و کلسیم موجود در خوراکیها، مکمل‌های ویتامینی و داروهای ضد اسید معده، در معده به آنتی‌بیوتیک‌ها چسبیده و از جذب آنها به داخل خون جلوگیری می‌کنند. به عنوان یک مثال ساده، می‌توان ممنوعیت مصرف آنتی‌بیوتیک خوراکی همراه با شیر را ذکر کرد.

۸.۴.۱ ویژگی‌های سه بعدی

ساختار شیمیایی سه بعدی داروها^{۳۵} اطلاعاتی از نوع اتم‌ها و نحوه‌ی جای‌گیری آنها در فضا ارائه می‌دهد. این ویژگی‌ها با استفاده از نرم‌افزارهای خاصی اندازه‌گیری می‌شوند.

۵.۱ تعریف مسئله

هنگامی که پزشک به‌طور همزمان چند دارو برای یک بیمار تجویز می‌کند، برهم‌کنش دارو-دارو ممکن است عوارض جانبی جبران‌ناپذیری ایجاد کند، به‌طور مثال ممکن است تأثیرات داروها بر روی یکدیگر منجر به بیماری‌های دیگر و یا حتی مرگ شود. این تأثیرات جانبی به‌طور خاص در افراد پیر و

³³Indication

³⁴Food - Drug Interaction

³⁵3-D Drug Chemical Structure

بیماران سرطانی که در روز تعداد زیادی دارو مصرف می‌کنند، چشم‌گیرتر است. عوارض جانبی یک دارو تا حد قابل قبولی در فاز توسعه دارو شناسایی می‌شود ولی عوارض جانبی حاصل از برهم‌کنش دارو-دارو به دلیل دامنه گسترده مسئله به‌ندرت کشف می‌شود. چنین تاثیراتی عمدتاً پس از تایید دارو و ورود دارو به بازار مصرف شناسایی می‌شوند که این خطری جدی برای سلامتی بیماران است. این مسئله از نظر اقتصادی نیز مهم است زیرا پس از شناسایی برهم‌کنش مخرب یک دارو با داروی دیگر، ممکن است آن دارو به طور کل از بازار جمع‌آوری شود. از طرفی به دلیل پرهزینه بودن روش‌های آزمایشگاهی، از مدل‌های پیش‌بینی برهم‌کنش دارو-دارو استفاده می‌کنند. برای این مسئله تاکنون روش‌های متنوع محاسباتی، آماری، یادگیری ماشین و غیره ارائه شده‌است. در این پایان‌نامه روشی نوین برای حل این مسئله ارائه می‌دهیم.

۱.۵.۱ برچسب برهم‌کنش‌ها

به‌طور کلی به برهم‌کنش دارو-دارو یکی از دو برچسب زیر را نسبت می‌دهند:

برچسب مثبت: اگر وجود برهم‌کنش بین دو دارو با شواهد موجود ثابت شده باشد و حداقل در یکی از پایگاه داده‌های دارو ثبت شده باشد.

برچسب منفی: برهم‌کنش دارو-داروهایی در این دسته قرار می‌گیرند که تاکنون شناخته نشده‌اند و در هیچ‌یک از پایگاه داده‌های مربوط به دارو ثبت نشده‌اند. درواقع با گذشت زمان ممکن است شواهدی مبنی بر وجود برهم‌کنش بین دو دارو یافت شود و این برهم‌کنش برچسب مثبت بگیرد، اما تاکنون چنین اطلاعاتی در دست نیست.

در این پایان‌نامه، به برهم‌کنش دارو-دارو یکی از سه برچسب زیر را نسبت می‌دهیم:

برچسب مثبت یک: اگر وجود برهم‌کنش بین دو دارو با شواهد موجود افزایش یافته باشد و حداقل در یکی از پایگاه داده‌های دارو ثبت شده باشد.

برچسب منفی یک: اگر وجود برهم‌کنش بین دو دارو با شواهد موجود کاهش یافته باشد و حداقل در یکی از پایگاه داده‌های دارو ثبت شده باشد.

برچسب صفر: اگر وجود برهم‌کنش بین دو دارو تاکنون شناخته نشده باشد و در هیچ‌یک از پایگاه داده‌های دارو ثبت نشده باشد.

۶.۱ ساختار پایان‌نامه

در ادامه‌ی این پایان‌نامه، ابتدا در فصل دو به بررسی پیشینه‌ی تحقیق و کارهای انجام‌شده بر روی داده‌های سه کلاسه از برهم‌کنش دارو - دارو می‌پردازیم. در مرحله‌ی بعد مفاهیم پیش نیاز برای ارائه روش پیشنهادی خود را معرفی می‌کنیم. این مفاهیم شامل دو بخش آماده‌سازی داده و رویکرد عمده‌ی انتخاب مدل می‌باشد.

در فصل سه، مجموعه داده‌ی به‌کاررفته را شرح خواهیم داد و روش پیشنهادی خود را با شرح مبسوط فرآیند کار برای حل مساله مطرح خواهیم کرد.

در فصل چهار به تحلیل و ارزیابی و بحث روی نتایج خواهیم پرداخت. در ادامه و در فصل پنج روش‌های پیشنهادی و کارهای آتی در جهت بهبود ارائه می‌شوند.

فصل ۲

بررسی تاریخچه، مفاهیم اولیه و الگوریتم‌های پیش‌بینی برهم‌کنش دارو-دارو

۱.۲ پیشینه تحقیق

هنگامی که دو یا چند دارو باهم مصرف می‌شوند، اثرات دارویی یا رفتارهای آن‌ها به‌طور غیرمنتظره‌ای تحت تاثیر یکدیگر قرار می‌گیرند [۱۸]. این نوع تاثیرگذاری به‌عنوان برهم‌کنش دارو-دارو نامیده می‌شود که می‌تواند تاثیر دارویی را کاهش دهد، مسمویت غیرمنتظره را افزایش دهد یا بین داروهای تجویز شده عوارض جانبی ایجاد کند. با افزایش داروهای تایید شده، تعداد برهم‌کنش دارو-داروهای ناشناخته به سرعت در حال افزایش است، به‌طوری‌که در بین داروهای کوچک مولکول تایید شده در DrugBank، به‌طور متوسط از هر یکصد جفت دارو تقریباً حدود پانزده‌تای آن‌ها دارای برهم‌کنش دارو-دارو هستند [۱۹]. برهم‌کنش دارو-داروهای ناشناخته باعث می‌شود بیمارانی که با چندین دارو تحت درمان هستند، در وضعیت ناایمن قرار گیرند [۲۰-۲۳]. همچنین درک برهم‌کنش دارو-دارو اولین قدم برای ترکیب داروهاست که به یکی از راهکارهای امیدوارکننده برای درمان بیماری‌های پیچیده تبدیل می‌شود [۲۴]. بنابراین، غربالگری و تجزیه و تحلیل برهم‌کنش دارو-داروها قبل از تجویز بالینی داروها، یک نیاز فوری است. با این حال، رویکردهای سنتی برای شناسایی برهم‌کنش دارو-دارو (به‌عنوان مثال، در آزمایش Cytochrome P450 [۲۵] یا برهم‌کنش مربوط به انتقال‌دهنده‌ی دارو [۲۶]) با چالش‌هایی روبه‌رو است. از آن جمله می‌توان هزینه‌های بالا، مدت زمانی طولانی، ملاحظات رفاه حیوانات [۱۴]، تعداد بسیار محدود شرکت‌کنندگان در آزمایش و تعداد زیاد ترکیبات دارویی تحت غربالگری آزمایش‌های بالینی را نام برد. تاکنون فقط تعداد کمی از برهم‌کنش دارو-داروها در طول تولید دارو شناسایی شده‌اند (معمولاً در مرحله‌ی آزمایش بالینی). برخی از آن‌ها پس از تایید داروها گزارش شده‌اند و بسیاری از آن‌ها در نظارت بعد از عرضه پیدا شده‌اند.

رویکردهای محاسباتی جایگزین امیدوارکننده‌ای برای کشف برهم‌کنش دارو-داروهای بالقوه^۱ در مقیاس گسترده هستند که بیش‌تر از گذشته، مورد استقبال دانشگاه و صنعت قرار گرفته‌اند [۲۷، ۲۸]. رویکردهای محاسباتی مبتنی بر داده‌کاوی برای تشخیص برهم‌کنش دارو-دارو از منابع مختلف [۱۴] از جمله: متون علمی [۲۹، ۳۰]، سوابق پزشکی الکترونیکی [۳۱] و سیستم گزارش حوادث نامطلوب FDA^۲ استفاده می‌کنند. این رویکردها به شواهد بالینی پس از عرضه به بازار متکی هستند، بنابراین نمی‌توانند هشدارهایی از برهم‌کنش دارو-داروهای بالقوه را قبل از تجویز بالینی دارو ارائه دهند. در مقابل رویکردهای مبتنی بر یادگیری ماشین (رویکرد مبتنی بر شباهت نیو^۳ [۳۲]، مبتنی بر گراف توصیه‌گر^۴ [۱۴]، مبتنی بر طبقه‌بندی^۵ [۳۳]) قادر به ارائه‌ی چنین هشدارهایی با استفاده از ویژگی‌ها و شباهت‌های دارویی، قبل و بعد از ارائه به بازار هستند [۳۴].

در این شیوه‌ها از ویژگی‌های مختلف دارو برای پیش‌بینی برهم‌کنش دارو-دارو استفاده می‌کنند، که عبارت‌اند از: ساختار شیمیایی [۳۲]، هدف [۳۵]، کدهای دسته‌بندی سلسله‌مراتبی^۶ [۳۳] و تاثیرات جانبی [۱۴، ۳۶].

بسیاری از رویکردهای مبتنی بر یادگیری ماشین برای پیش‌بینی دوکلاسه معمولی طراحی شده‌اند که تنها احتمال برهم‌کنش دارو-داروی یک جفت دارو را نشان می‌دهند، اما دو داروی برهم‌کنش پذیرفته‌شده ممکن است رفتارها یا اثرات دارویی خود را در بدن افزایش یا کاهش دهند. به عنوان مثال:

(۱) غلظت سرم Flunisolide (با شناسه‌ی DrugBank، DB00180) وقتی با Mitotane (با شناسه‌ی DrugBank، DB00648) به صورت هم‌زمان مصرف شود، کاهش می‌یابد.

(۲) درحالی‌که غلظت همین سرم وقتی با Roxithromycin (با شناسه‌ی DrugBank، DB00778) هم‌زمان مصرف شود، افزایش می‌یابد.

مورد اول را برهم‌کنش دارو-داروی کاهنده^۷ و دومین مورد را برهم‌کنش دارو-داروی افزایشنده^۸ می‌نامیم و هر دو مورد را به عنوان برهم‌کنش دارو-دارو که شامل تاثیرات دارویی موثر هستند در نظر می‌گیریم. دانستن دقیق این نکته که برهم‌کنش باعث افزایش یا کاهش تاثیر دارو می‌شود، به‌ویژه در هنگام مراقبت از بیمار، تعیین مقدار مصرف دارو، طراحی دارو یا یافتن مقاومت دارو در برابر درمان، بسیار مهم است [۳۷]. بسیاری از رویکردهای موجود هنوز از این خاصیت ساختاری استفاده نکرده‌اند و فقط برای برهم‌کنش دارو-داروهای دوکلاسه معمولی توسعه یافته‌اند. درحالی‌که شناخت نوع برهم‌کنش دارو-دارو، یکی از مهم‌ترین اقدامات برای درمان بیماری‌های پیچیده است [۳۸] و می‌تواند راهنمای پزشکان در تهیه نسخه‌های مطمئن‌تر باشد. در ادامه سه روش برای پرداختن به موارد ذکر شده توضیح

¹Potential

²Event Reporting System

³Naïve Similarity-Based Approach

⁴Network Recommendation-Based

⁵Classification-Based

⁶Hierarchical Classification Codes

⁷Degressive

⁸Enhancive

داده می شود. این سه روش از پیش بینی های سه کلاسه جامع به جای پیش بینی دو کلاسه معمولی استفاده کردند و به بررسی ساختاری داروها در شبکه برهم کنش داروها پرداختند.

در بخش بعدی روش های آماده سازی داده و رویکردهای انتخاب مدل، به کار رفته در مقالات پیشین، فرمول بندی و توضیح داده می شوند. مطالعه ای این بخش کمک شایانی به درک فرآیند به کار رفته در روش پیشنهادی می کند.

۲.۲ آماده سازی داده

در این بخش به بحث در مورد روش های متداول محاسبه ی شباهت های دارویی و ترکیب شباهت های شبکه ای می پردازیم.

۱.۲.۲ محاسبه ی شباهت های دارویی

سه روش متداول محاسبه ی شباهت که در مقالات یادگیری ماشین مانند مقالات [۳۹-۴۱] استفاده می شوند، عبارت اند از: شباهت جاکارد^۹، شباهت کوسینوس^{۱۰} و شباهت گوسین^{۱۱}. این روش ها مقدار شباهت بین دو نمونه داده را منعکس می کنند و به طور گسترده در بیوانفورماتیک مورد استفاده قرار گرفته اند [۴۱-۴۵]. اگر بردارهای ویژگی دارو d_i و d_j را با x_i و x_j نمایش دهیم، سه روش شباهت به شرح زیر تعریف می شوند:

● شباهت جاکارد بین x_i و x_j

$$S_{Jar}(x_i, x_j) = \frac{N_{11}}{N_{1.} + N_{.1} + N_{11}} \quad (۱.۲)$$

که:

N_{11} تعداد کل عناصری است که x_i و x_j هر دو مقدار ۱ گرفته اند.

$N_{1.}$ تعداد کل عناصری است که x_i مقدار ۱ و x_j مقدار ۰ گرفته اند.

$N_{.1}$ تعداد کل عناصری است که x_i مقدار ۰ و x_j مقدار ۱ گرفته اند.

● شباهت کوسینوس بین x_i و x_j

$$S_{Cos}(x_i, x_j) = \frac{x_i \cdot x_j}{\|x_i\|_2 \|x_j\|_2} \quad (۲.۲)$$

که $\| \cdot \|_2$ نرم اقلیدسی است و $x_i \cdot x_j$ ضرب داخلی دو بردار را نمایش می دهد.

⁹ Jacard Similarity

¹⁰ Cosine Similarity

¹¹ Gaussian Similarity

● شباهت گوسین بین x_i و x_j

$$S_{Gau}(x_i, x_j) = \exp(-\sigma \|x_i - x_j\|_2^2) \quad (3.2)$$

که $\sigma (> 0)$ پارامتر پهنای باند است و به صورت $\sigma = \frac{1}{\sum_{i=1}^m \frac{|x_i|}{m}}$ تنظیم می‌شود.

سایر روش‌های محاسبه شباهت بین دو دارو در شکل ۱.۲ نمایش داده شده است.

Measure	Equations	Description
Hamming	$S = d - x^t y + \bar{x}^t \bar{y} = \sum_{i=1}^d s_i$ where $s_i = \begin{cases} 1 & \text{if } x_i = y_i \\ 0 & \text{otherwise} \end{cases}$	Hamming similarity is the number of identical bits in two binary vectors
Sokal-Michener	$S = \frac{x^t y + \bar{x}^t \bar{y}}{d}$	Normalized Hamming
Rogers-Tanimoto	$S = \frac{x^t y + \bar{x}^t \bar{y}}{x^t y + \bar{x}^t \bar{y} + 2x^t \bar{y} + 2\bar{x}^t y}$	Alternative normalized hamming
Azzo	$S = \sum_{i=1}^d s_i$ where $s_i = \begin{cases} 1 & \text{if } x_i = y_i = 1 \\ \sigma & \text{if } x_i = y_i = 0 \\ 0 & \text{otherwise} \end{cases}$	Variable contribution factor for zero matches $1 \leq \sigma \ll \infty$
Inner-Product	$S = x^t y = \sum_{i=1}^d s_i$ where $s_i = \begin{cases} 1 & \text{if } x_i = y_i = 1 \\ 0 & \text{otherwise} \end{cases}$	Consider positive matches
Russell-Rao	$S = \frac{x^t y}{d}$	A normalization on inner product
Jaccard	$S = \frac{x^t y}{x^t y + \bar{x}^t \bar{y} - x^t y}$	A normalization on inner product
Dice	$S = \frac{x^t y}{2x^t y + \bar{x}^t \bar{y} - x^t y}$	A normalization on inner product
Kulzinsky	$S = \frac{x^t y}{x^t \bar{y} + \bar{x}^t y}$	A normalization on inner product
Tanimoto	$S = \frac{x^t y}{x^t x + \bar{y}^t \bar{y} - x^t y}$	A normalization on inner product
Correlation	$S = \frac{x^t y \times \bar{x}^t \bar{y} - x^t \bar{y} \times \bar{x}^t y}{\sqrt{(x^t \bar{y} + \bar{x}^t y)(\bar{x}^t y + x^t \bar{y})(x^t y + \bar{x}^t \bar{y})(\bar{x}^t \bar{y} + x^t y)}}$	A multiplicative form of positive and negative matches
Yule	$S = \frac{x^t y \times \bar{x}^t \bar{y} - x^t \bar{y} \times \bar{x}^t y}{x^t y \times \bar{x}^t \bar{y} + \bar{x}^t y \times x^t \bar{y}}$	A multiplicative form of positive and negative matches

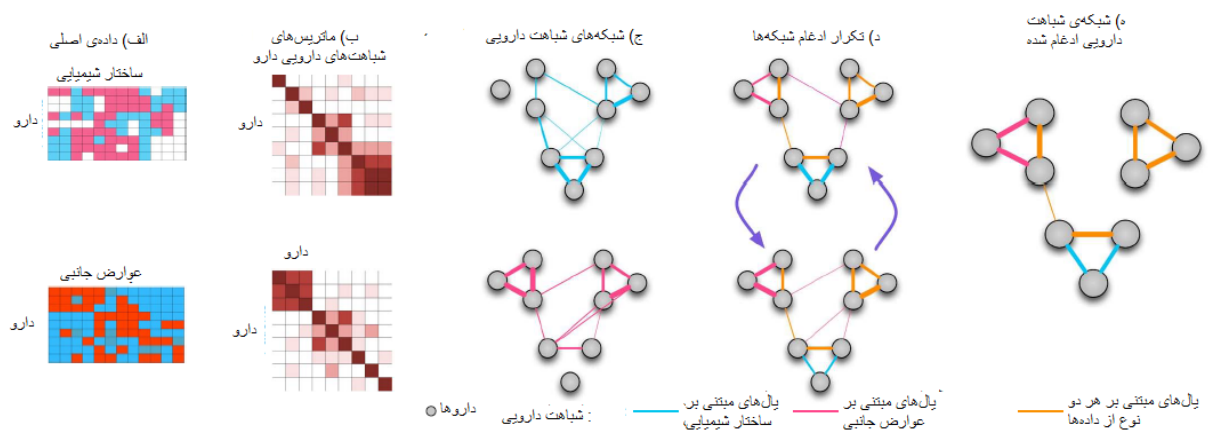
شکل ۱.۲: انواع شباهت بین بردارهای دودویی

۲.۲.۲ ترکیب شباهت‌های شبکه‌ای

ترکیب شباهت‌های شبکه‌ای^{۱۲} [۴۶] (SNF) یک روش محاسباتی جدید برای ترکیب داده‌ها است. این روش انواع مختلفی از ویژگی‌ها (مانند ساختار شیمیایی، عوارض جانبی، آنزیم، منتقل‌کننده و غیره) را با مجموعه‌ای معینی از نمونه‌ها (به عنوان مثال داروها) ترکیب می‌کند. در این روش ابتدا برای هر یک از انواع داده‌ها یک شباهت نمونه ایجاد می‌شود و سپس با استفاده از یک روش جدید ترکیب شبکه، شبکه‌ها یکپارچه می‌شوند. کار در فضای شبکه به SNF اجازه می‌دهد تا از برخورد با مقیاس‌های مختلف، ارزیابی و نويز در انواع مختلف داده جلوگیری کند. روش ترکیب غیرخطی داده‌ها به SNF اجازه می‌دهد تا هم از اطلاعات رایج و هم از اطلاعات مکمل در انواع مختلف داده استفاده برد. شکل ۲.۲ مثالی از مراحل SNF است.

¹²Similarity Network Fusion

¹³Bias



شکل ۲.۲: انواع شباهت بین بردارهای دودویی [۴۶]

الف) نمایش نمونه‌ای از ویژگی ساختار شیمیایی و ویژگی عوارض جانبی خارج از برچسب برای همان داروها. ب) ماتریس تشابه دارو-دارو برای هر نوع ویژگی. ج) شبکه‌های شباهت دارو-دارو، معادل داده‌های دارو-دارو. داروها توسط گره‌ها نمایان می‌شوند و شباهت‌های جفتی داروها توسط یال‌ها نشان داده می‌شود. د) تلفیق شبکه توسط SNF به‌طور مکرر، هر یک از شبکه‌ها را با اطلاعاتی از شبکه‌های دیگر به روز می‌کند و آن‌ها را در هر مرحله شبیه‌تر می‌کند. ه) ترکیب شبکه به‌صورت تکراری منجر به همگرایی می‌شود. رنگ یال نشان می‌دهد که کدام نوع داده به شباهت داده‌شده کمک کرده است.

انتخاب مدل

در این بخش توضیح مختصری از تاریخچه‌ی سامانه‌های توصیه‌گر و یادگیری عمیق داده می‌شود و سپس به بررسی تقسیم‌بندی آنها می‌پردازیم.

۳.۲ سامانه‌های توصیه‌گر

به کمک کامپیوتر جوامع امروزی دستخوش تغییرات سریعی شده‌اند. بخش قابل توجهی از زندگی اجتماعی را کامپیوتر و اینترنت در برگرفته است. یکی از موضوعات جدید مورد بحث حال حاضر دنیای کامپیوتر، سامانه‌های توصیه‌گر می‌باشد.

سامانه‌های توصیه‌گر^{۱۴} با اولین ظهورشان در زمینه فیلترینگ همکارانه^{۱۵}، حوزه تحقیقاتی مهمی را در اواسط دهه ۱۹۹۰ فراهم نمودند. در دهه‌های اخیر دو بخش صنعت و دانشگاه دستاوردهای جدیدی در زمینه سامانه‌های توصیه‌گر توسعه داده‌اند، با این وجود علاقه‌مندی به این بخش هنوز در سطح بالایی

¹⁴Recommender Systems

¹⁵Collaborative Filtering

است. تحقیقات این حوزه غنی بوده و نیاز مبرم به برنامه‌های کاربردی وجود دارد. این برنامه‌های کاربردی به‌منظور کمک به کاربران که با حجم زیادی از اطلاعات مواجه هستند و به‌منظور شخصی‌سازی اطلاعات توسعه داده شدند. به‌عنوان مثال از چنین برنامه‌هایی، می‌توان به سامانه‌های توصیه‌گر کتاب، سی‌دی و دیگر محصولات سایت آمازون^{۱۶} و پیشنهاد فیلم توسط شرکت موی‌لنز^{۱۷} اشاره کرد.

تقریباً در اواسط دهه نود، مطالعه بر روی سامانه‌های توصیه‌گر به‌عنوان یک شاخه مستقل در تحقیقات مطرح شد. علت این توجه خاص، ابراز تمایل محققین، به‌استفاده از سامانه‌های توصیه‌گر برای حل مسئله جستجو در حجم فراوان اطلاعات بود. ظرفیت کامپیوترها در فراهم آوردن توصیه‌ها تقریباً از همان اوایل تاریخچه کامپیوترها شناخته شد. گراندی^{۱۸}، کتابدار کامپیوتری گامی اولیه به سمت سامانه‌های توصیه‌گر خودکار برداشت. این کتابدار یک توصیه‌گر نسبتاً ساده و اولیه بود که کاربران را به قالب‌هایی براساس مصاحبه کوتاه با استفاده از اطلاعات مستقیم‌کدشده^{۱۹} درباره‌ی سلیقه کتاب قالب‌های مختلف گروه‌بندی می‌کرد تا توصیه‌ها را تولید کنند. این کار ورود اولیه به فضای سامانه‌های توصیه‌گر قلمداد می‌شود.

در اوایل دهه نود میلادی، تکنیک پالایش مشارکتی^{۲۰} به‌عنوان راه‌حلی برای مدیریت فضای اطلاعات بسیار زیاد آنلاین به‌وجود آمد. تپستری^{۲۱} یک سامانه پالایش مشارکتی دستی بود. این سامانه به کاربر اجازه انجام پرس‌وجو برای آیتم‌های موجود در حوزه اطلاعاتی مانند ایمیل بر اساس عقاید و اقدامات دیگر کاربران را می‌داد. این‌کار مستلزم تلاش از طرف کاربران بود و به آنها اجازه کنترل واکنش‌های خوانندگان قبلی در قسمتی از مکاتبات را می‌داد تا میزان ارتباط با آنها را تعیین کند. خیلی زود بعد از سامانه‌های خودکار پالایش مشارکتی، مکان‌یابی خودکار عقاید مرتبط و تجمع آنها برای دادن توصیه مطرح شد. گروپ لنز^{۲۲} از این تکنیک برای تعیین مقاله‌های یوزنت^{۲۳} که احتمال دارد مورد علاقه کاربر خاصی باشد استفاده کرد. کاربران تنها به نمره‌دهی یا دیگر اقدامات قابل مشاهده می‌پرداختند. سامانه این فعالیت‌ها را با نمره‌ها یا اقدامات دیگر کاربران ترکیب می‌کرد تا نتایج شخصی‌شده تولید کند. با این سامانه‌ها، کاربران برای دریافت پیشنهادات، قادرند بدون آنکه بدانند کاربران یا آیتم‌های دیگر سامانه چه چیزهایی هستند، اطلاعات مستقیمی از عقاید دیگر کاربران بدست آورند.

طی این دوره، سامانه‌های توصیه‌گر و پالایش مشارکتی تبدیل به موضوع مورد علاقه محققین حوزه‌های تعاملات انسان-رایانه، یادگیری ماشین و بازیابی اطلاعات شدند. این علاقه منجر به ایجاد سامانه توصیه‌گر در زمینه‌های مختلف شد. که از آن جمله می‌توان به Ringo برای موسیقی، توصیه‌گر ویدیو BellCore برای فیلم‌ها و Jester برای لطیفه‌ها اشاره کرد. خارج از دنیای رایانه و درحوزه بازاریابی از

¹⁶ Amazon

¹⁷ MovieLens

¹⁸ Grundy

¹⁹ Hard-coded

²⁰ Collaborative Refinement

²¹ Tapestry

²² Group Lens

²³ Usenet

توصیه‌گرها، به دلیل توانایی آنها در افزایش فروش و بهبود تجربه مشتریان، استفاده کرده‌اند.

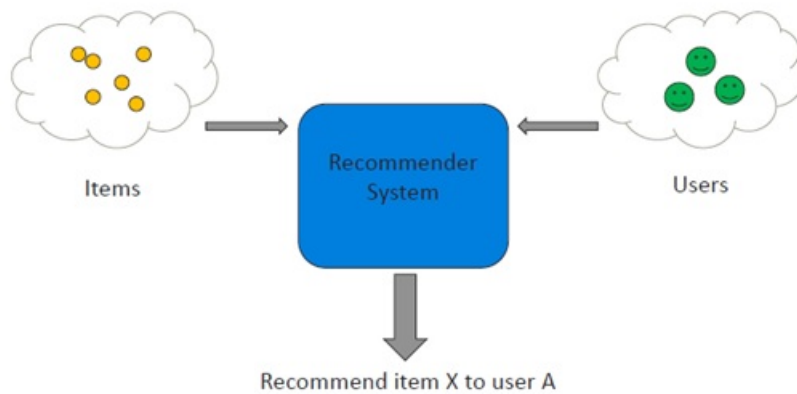
در اواخر دهه نود میلادی، پیاده‌سازی‌های تجاری فناوری توصیه‌گرها شروع به ظهور کردند. شاید معروف‌ترین کاربرد فناوری سامانه‌های توصیه‌گر وبسایت Amazon.com باشد. براساس تاریخچه خرید، تاریخچه بازدید و موردی که کاربر در حال مشاهده است، وبسایت به کاربر مواردی را برای خرید توصیه می‌کند. از زمان به‌کارگیری فناوری توصیه‌گر توسط آمازون، در بسیاری از سیستم‌های تجارت الکترونیک و آنلاین، سامانه‌های توصیه‌گر (که اغلب بر اساس پالایش مشارکتی هستند)، ایجاد شده است. انگیزه مهم انجام این کار افزایش حجم فروش است زیرا ممکن است اگر کالایی به مشتریان پیشنهاد شود آن را بخرند و در غیر این صورت ممکن است هرگز متوجه آن کالا نشده و آنرا نخرند. شرکت‌های بسیاری مانند NetPerceptions و Strands به دلیل فراهم کردن فناوری و خدمات توصیه به خرده‌فروشان آنلاین بوجود آمده‌اند.

زمانی که Netflix جایزه Netflix Prize را در سال ۲۰۰۶ به منظور بهبود بخشیدن وضعیت توصیه‌های فیلمش برقرار کرد، تحقیق بر روی الگوریتم‌های سامانه‌های توصیه‌گر توجه بسیاری را به خودش جلب کرد. هدف این رقابت ساختن الگوریتم توصیه‌گری بود که بتواند الگوریتم CineMatch که متعلق به خود Netflix بود را با ۱۰٪ بهبود در آزمایشات آفلاین شکست دهد. این امر موجب ایجاد اقدامات گسترده‌ای در محیط دانشگاهی و سایر علاقمندان شد. جایزه یک میلیون دلاری ارزشی که فروشندگان برای دقت توصیه‌ها قائل هستند را نشان می‌دهد [۴۷]. به‌طور کلی دلیل استفاده از سامانه‌های توصیه‌گر و فیلترینگ اطلاعات، تبدیل اطلاعات موجود در مورد کاربران و اولویت‌های آنها جهت پیش‌بینی علاقه آنها در آینده می‌باشد. در واقع در این سامانه از بین موارد مختلف بر اساس اینکه چه موردی برای کاربر جذاب است، مجموعه‌ی بزرگی از موارد مورد علاقه کاربران جمع‌آوری می‌شود. وبسایت‌های زیادی مانند Amazon، CDnow، Ebay و غیره شروع به استفاده از سامانه‌های توصیه‌گر برای اهداف فروش الکترونیکی کردند. این وبسایت‌ها با توجه به اطلاعات و علایق مشتری به مشتری محصول مناسب سلیقه او را پیشنهاد می‌دهند که سبب افزایش رابطه میان وبسایت‌ها و مشتریان می‌شود. این باور وجود دارد که با ارائه توصیه‌های با کیفیت به مشتری و جلب رضایت او، وفاداری مشتری بهبود می‌یابد.

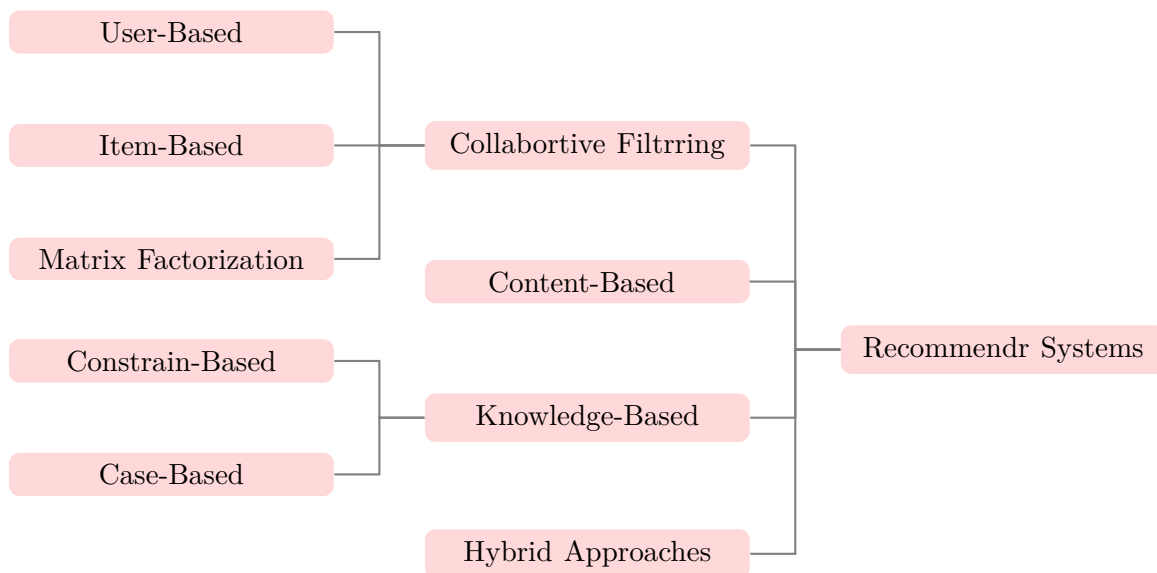
پیشرفت اینترنت باعث شده است اطلاعات آنلاین زیادی قابل دسترسی باشد. سامانه‌های توصیه‌گر می‌توانند مشکل انباشت اطلاعات را حل کنند. هدف این سامانه‌ها شناسایی سلیقه‌های کاربران و فیلتر کردن داده‌های نامناسب از نظر آنهاست. نتیجه استفاده از سامانه‌های توصیه‌گر صرفه‌جویی زمانی جستجوهای کاربران است. در شکل ۳.۲ شمای کلی فرایند توصیه در سامانه‌های توصیه‌گر را مشاهده می‌کنید.

تقسیم‌بندی سامانه‌های توصیه‌گر


























در کتاب‌های مختلف تقسیم‌بندی‌های مختلفی برای سامانه‌های توصیه‌گر وجود دارد. معمول‌ترین تقسیم‌بندی سامانه‌های توصیه‌گر را می‌توانید در شکل ۴.۲ مشاهده کنید.



شکل ۳.۲: شمای کلی فرایند توصیه در سامانه‌های توصیه‌گر



شکل ۴.۲: تقسیم‌بندی سامانه‌های توصیه‌گر [۴۷]

شکل ۵.۲: نمای کلی از روش‌های فیلترینگ همکارانه

۱.۳.۲ فیلترینگ همکارانه

در روش فیلترینگ همکارانه^{۲۴} پیشنهادات براساس شباهت‌های رفتاری و الگوهای عملکردی کاربرانی که شباهت‌های رفتاری و الگوهای مشابهی با کاربر فعلی در گذشته داشته‌اند، ارائه می‌شود. شاید تعریف آن کمی پیچیده باشد، به‌طور ساده روش فیلترینگ همکارانه بر این فرض استوار است که کاربرانی با نظرهای مشابه درباره یک مورد (منظور فیلم، عکس، موزیک یا هر چیز دیگری است که توصیه می‌شود) دارند، درباره موارد دیگر هم نظرات مشابه دارند. نمای کلی عملکرد این روش را در شکل ۵.۲ مشاهده کنید. فیلترینگ همکارانه شامل سه بخش است که به بررسی آن‌ها می‌پردازیم:

- مبتنی بر کاربر^{۲۵}: در این روش افراد هم سلیقه با توجه به امتیازاتی که به موارد داده‌اند دسته‌بندی می‌شوند. از آنجا که کاربرانی با سلیقه شبیه به فرد مورد نظر بود و از موارد یکسان خوششان آمده بود، پس آن مورد را به فرد مورد نظر نیز پیشنهاد می‌دهند.

- مبتنی بر آیتم^{۲۶}: در این روش وقتی یک کاربر یک مورد را می‌بیند، سامانه به‌صورت خودکار به دنبال آن می‌گردد که کاربرانی که قبلاً این مورد را دیده‌اند بعد از آن چه موارد دیگری را دیده بودند. سپس سامانه آن موارد را به کاربر پیشنهاد می‌دهد. مثلاً در سایت آمازون اگر یک گوشی بخرید بعد از آن، سامانه توصیه‌گر قاب گوشی و محافظ صفحه نمایشگر و چیزهایی مربوط به گوشی توصیه می‌کند.

- تجزیه ماتریس^{۲۷}: روش‌های مبتنی بر کاربر و مبتنی بر آیتم ساده هستند، اما معمولاً روش‌های تجزیه ماتریس بیشتر تاثیرگذار هستند. زیرا این روش‌ها، ویژگی‌های پنهانی که در ارتباط میان

²⁴Collaborative Filtering

²⁵User-Based

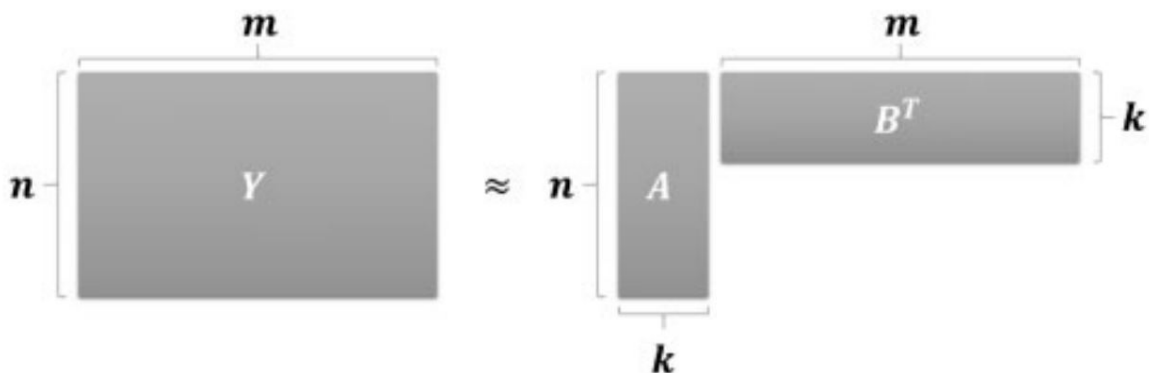
²⁶Item-Based

²⁷Matrix Factorization

کاربران و موارد وجود دارد را کشف می‌کنند. از این روش برای پیش‌بینی امتیازها در فیلترینگ همکارانه استفاده می‌کنند. روش تجزیه ماتریس یک ماتریس ورودی را می‌گیرد و تلاش می‌کند دو ماتریس دیگر را به گونه‌ای بیابد که در صورت ضرب شدن در هم، ماتریس ورودی را تقریب بزنند. الگوریتم تجزیه ماتریس با مثال معروف توصیه‌ی کالاهای دیده نشده به کاربر معرفی می‌شود. در این حالت ماتریسی به نام Y حاوی بازخوردهای کاربران بر روی کالاها می‌باشد که شامل m کاربر و n کالا می‌باشد. مطابق شکل ۶.۲، برای پیش‌بینی بازخورد کاربران روی کالاهای دیده نشده ماتریس $Y \in R^{(m \times n)}$ به دو ماتریس $A \in R^{(m \times k)}$ و $B \in R^{(k \times n)}$ تقسیم می‌شود به طوری که $AB^T \approx Y$. پارامتر قابل تنظیم k ، مربوط به تعداد ویژگی‌های نهفته A و B است که $k \ll m \times n$.

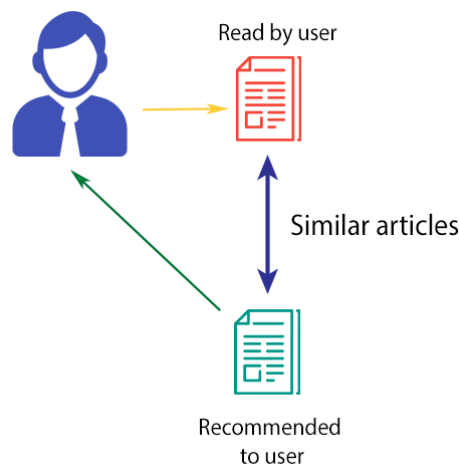
تجزیه ماتریس ویژگی‌های پنهان کاربران و کالاها را به شیوه‌ی بدون سرپرست مشخص می‌کند، که برای فیلترکردن مشارکتی مفید است. به عنوان مثال، اگر بردارهای پنهان دو کاربر شبیه باشند، احتمال بالایی دارد که این کاربرها بازخوردهای یکسانی را به اشتراک بگذارند. به این ترتیب احتمال بازخورد مشابه روی کالاهای یکسان یا کالاهای مشابه بالاتر خواهد بود.

هدف یافتن بازخوردهای ناموجود در ماتریس Y است که تجزیه ماتریس می‌تواند به عنوان تکنیک تکمیل‌کننده‌ی ماتریسی مورد استفاده قرار گیرد. تصویری از تجزیه ماتریسی در شکل ۶.۲ نمایش داده شده است.



شکل ۶.۲: شرح مصور از تجزیه ماتریس

هدف این است که دو ماتریس ویژگی‌های پنهان A و B را به گونه‌ای بیابیم که ماتریس بازخورد Y از ضرب این دو ماتریس بهتر بازسازی شود [۶۷].



شکل ۷.۲: نمای کلی از روش‌های فیلترینگ مبتنی بر محتوا

۲.۳.۲ فیلترینگ مبتنی بر محتوا

فیلترینگ مبتنی بر محتوا^{۲۸} از اطلاعات متنی و سابقه‌های کاربر برای تطابق موارد استفاده کرده و گزینه‌های مناسب را پیشنهاد می‌دهد. دانش اصلی که در این تکنیک استفاده می‌شود شامل اطلاعات مواردی مثل ژانر، سال تولید و ... است که براساس این اطلاعات، مواردی که شبیه به هم هستند توصیه می‌شوند. در شکل ۷.۲ نحوه کلی عملکرد این روش را مشاهده می‌کنید. نمونه عمده این روش در وبسایت آمازون برای پیشنهاد کتاب‌ها بر اساس کلمات کلیدی کتاب‌های مشابه که سابقاً توسط کاربر خریداری شده استفاده می‌شود.

۳.۳.۲ فیلترینگ مبتنی بر دانش

دو روش توصیه‌گر قبلی محدودیت‌هایی دارند. محدودیت‌های استفاده از فیلترینگ همکارانه و فیلترینگ مبتنی بر محتوا را می‌توانید در قالب مثال بعدی ببینید. معمولاً افراد به‌صورت مکرر ماشین، خانه و یا کامپیوتر نمی‌خرند. در این سناریوها یک سامانه فیلترینگ همکارانه خوب عمل نمی‌کند چون تعداد امتیازهای موجود بسیار کم است. علاوه بر این، محدوده زمانی نقش بسیار مهمی ایفا می‌کند. برای مثال، امتیازهایی که پنج سال قبل به یک کامپیوتر داده شده است به طور قطع برای توصیه مبتنی بر محتوا مناسب نیست.

روش فیلترینگ مبتنی بر دانش^{۲۹} چالش‌های ذکرشده را حل می‌کند. سامانه‌های توصیه‌گر مبتنی بر دانش، نسل جدیدی از سامانه‌های توصیه‌گر هستند که مبتنی بر دانش موجود در رابطه با کاربران و موارد هستند. چنین سامانه‌هایی، پیشنهادات خود را بر پایه تفسیر و استنباط خود از سلاقی و نیازهای کاربر

²⁸Content-Based

²⁹Knowledge-Based

ارائه می‌دهند و از دیدگاه تئوری نسبت به سایر روش‌های ذکر شده از دقت و کیفیت بیشتری برخوردار هستند. برای پیاده‌سازی چنین سامانه‌هایی نیاز به یک بستر و ساختار مبتنی بر دانش وجود دارد. در این گونه سامانه‌های توصیه‌گر موارد اولیه مورد استفاده برای تولید لیستی از پیشنهادها، دانش سامانه درمورد مشتری و کالا است. سامانه‌های مبتنی بر دانش از روش‌های مختلفی که برای تحلیل دانش قابل استفاده هستند بهره می‌برند. روش‌های رایج در الگوریتم‌های ژنتیک^{۳۰}، فازی^{۳۱}، شبکه‌های عصبی^{۳۲} و ... از جمله آنها است. همچنین، در این سامانه‌ها از درخت‌های تصمیم^{۳۳}، استدلال نمونه محور^{۳۴} و ... نیز می‌توان استفاده کرد. روش مبتنی بر دانش خود به دو روش مبتنی بر محدودیت^{۳۵} و مبتنی بر مورد^{۳۶} تقسیم می‌شود. هر دو روش از لحاظ فرایند توصیه یکی هستند یعنی اول یک کاربر باید بطور دقیق درخواست خود را بگوید سپس سامانه تلاش کند که یک راه‌حل تشخیص بدهد. سامانه حتی می‌تواند توضیح کوتاهی برای دلیل توصیه مورد بدهد. اما این دو روش از لحاظ تهیه دانش باهم تفاوت دارند. روش مبتنی بر مورد، آیت‌های مشابه را با استفاده از میزان تشابه^{۳۷} توصیه می‌کند اما روش مبتنی بر محدودیت، با توجه به قانون‌های توصیه‌ای که از قبل به طور صریح تعبیه شده فرایند توصیه را انجام می‌دهد.

۴.۳.۲ روش‌های ترکیبی

روش‌های ترکیبی^{۳۸}، همانطور که اسم آنها پیداست ترکیبی از روش‌های قبلی است که سعی کرده با ترکیب روش‌ها از مزیت هریک از روش‌ها استفاده کند و محدودیت‌های آن‌ها را پوشش دهد.

۴.۲ یادگیری عمیق

الگوریتم‌های یادگیری عمیق^{۳۹} زیرمجموعه‌ای از الگوریتم‌های یادگیری ماشین هستند که هدف آنها کشف چندین سطح از بازنمودهای توزیع شده^{۴۰} از داده ورودی است. اخیراً الگوریتم‌های یادگیری عمیق زیادی برای حل مسائل هوش مصنوعی سنتی ارائه شده‌اند. در این بخش مروری بر بعضی از الگوریتم‌های به‌روز این مبحث خواهیم داشت. در ابتدا خلاصه‌ای از چندین روش یادگیری عمیق متفاوت و پیشرفت‌های اخیر آنها ارائه می‌شود و سپس به‌صورت خلاصه کاربردهای هریک در زمینه‌های مختلف توضیح داده

³⁰Genetic Algorithms

³¹Fuzzy

³²Artificial Neural Networks

³³Design Tree

³⁴Case-Based Reasoning

³⁵Constrain-Based

³⁶Case-Based

³⁷Similarity Measure

³⁸Hybrid Approaches

³⁹Deep Learning

⁴⁰Distributed Representation

می‌شود. در انتهای این بخش نیز گرایش‌ها و مشکلات آینده در طراحی و آموزش شبکه‌های عصبی عمیق به صورت خلاصه ارائه می‌شود. طی سالهای اخیر، یادگیری عمیق به صورت گسترده مورد مطالعه قرار گرفته است و به همین دلیل، تعداد زیادی از روش‌های مرتبط با آن به وجود آمده است. به طور کلی، این روش‌ها بر اساس روش پایه‌ای که از آن مشتق شده‌اند به چهار دسته مختلف تقسیم می‌شوند [۴۸].

دسته‌بندی روش‌های یادگیری عمیق و کارهای انجام شده با هر یک از این روش‌ها در شکل ۸.۲ مشاهده می‌کنید.

۱.۴.۲ شبکه‌های عصبی کانولوشن

شبکه‌های عصبی کانولوشن^{۴۱} یکی از مهم‌ترین روش‌های یادگیری عمیق هستند که در آن‌ها چندین لایه با روشی قدرتمند آموزش می‌بینند. این روش بسیار کارآمد بوده و یکی از رایج‌ترین روش‌ها در کاربردهای مختلف است. تصویر کلی یک معماری شبکه عصبی کانولوشن در شکل ۹.۲ نمایش داده شده است. به طور کلی، شبکه CNN از سه لایه اصلی تشکیل می‌شود که عبارتند از: لایه کانولوشن^{۴۲}، لایه ادغام^{۴۳} و لایه‌ی تماماً متصل^{۴۴}. لایه‌های مختلف وظایف مختلفی را انجام می‌دهد. در شکل ۹.۲ معماری کلی شبکه عصبی کانولوشن برای دسته‌بندی تصاویر به صورت لایه به لایه نمایش داده شده است [۴۴]. در هر شبکه عصبی کانولوشن دو مرحله برای آموزش وجود دارد. مرحله تغذیه رو به جلو^{۴۵} و مرحله بازگشت به عقب^{۴۶}. در مرحله اول تصویر ورودی به شبکه داده می‌شود و این عمل چیزی جز ضرب نقطه‌ای بین ورودی و پارامترهای هر نورون^{۴۷} و نهایتاً اعمال عملیات کانولوشن در هر لایه نیست، سپس خروجی شبکه محاسبه می‌شود. در اینجا به منظور تنظیم پارامترهای شبکه و یا به عبارت دیگر همان آموزش شبکه، از نتیجه خروجی جهت محاسبه میزان خطای شبکه استفاده می‌شود. برای اینکار خروجی شبکه را با استفاده از تابع خطا^{۴۸} با پاسخ صحیح مقایسه و میزان خطا محاسبه می‌شود. در مرحله بعدی براساس میزان خطای محاسبه شده، مرحله بازگشت به عقب آغاز می‌شود. در این مرحله گرادینت^{۴۹} هر پارامتر با توجه به قاعده زنجیره‌ای^{۵۰} محاسبه می‌شود و تمامی پارامترها با توجه به تاثیری که بر خطای ایجاد شده در شبکه دارند تغییر پیدا می‌کنند. بعد از بروزرسانی شدن پارامترها مرحله بعدی تغذیه رو به جلو شروع می‌شود. بعد از تکرار تعداد مناسبی از این مراحل، آموزش شبکه پایان می‌یابد.

⁴¹Convolutional Neural Networks(CNN)

⁴²Convolution

⁴³Pooling Layer

⁴⁴Fully Connected

⁴⁵Feed Forward

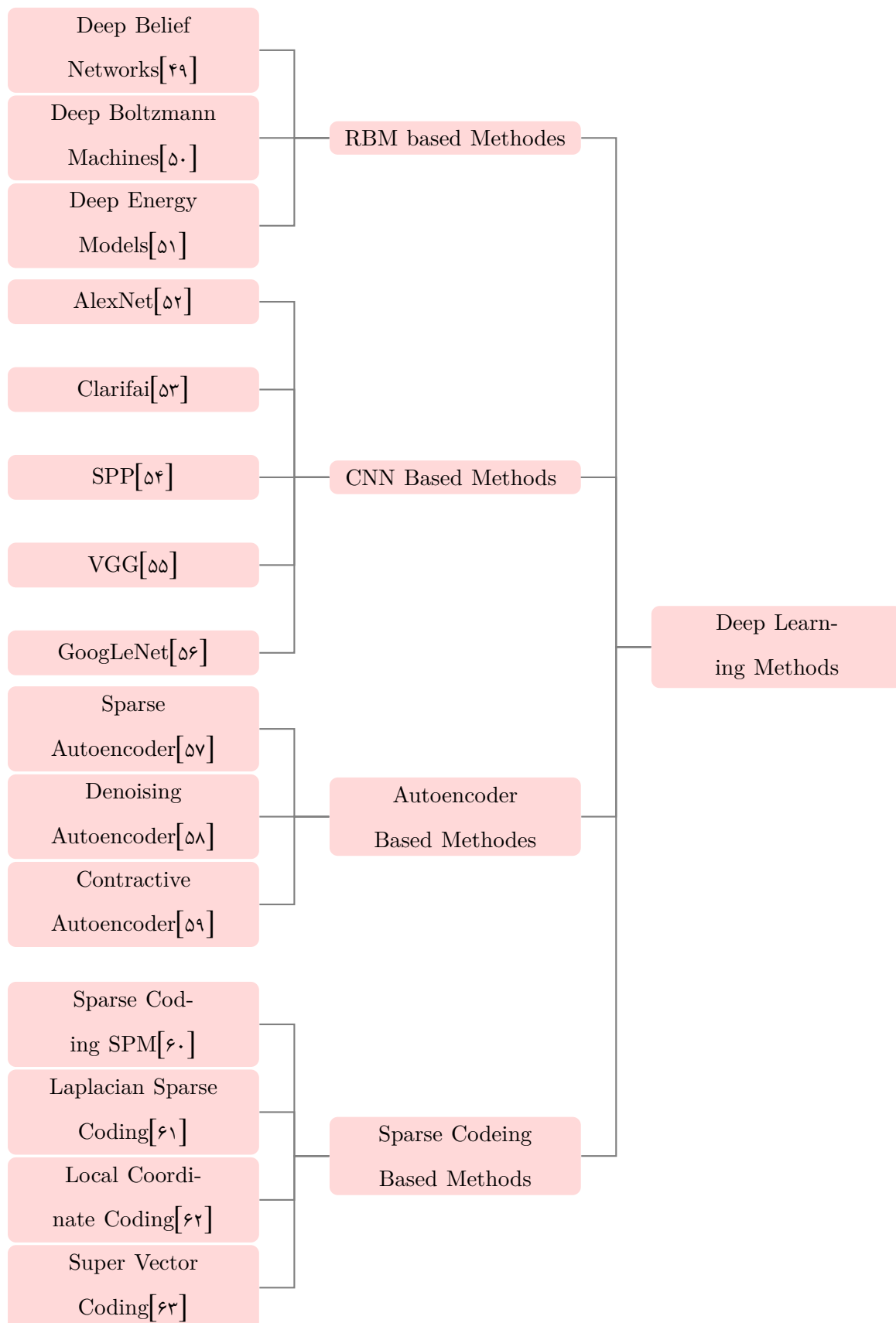
⁴⁶Back Propagation

⁴⁷Neurons

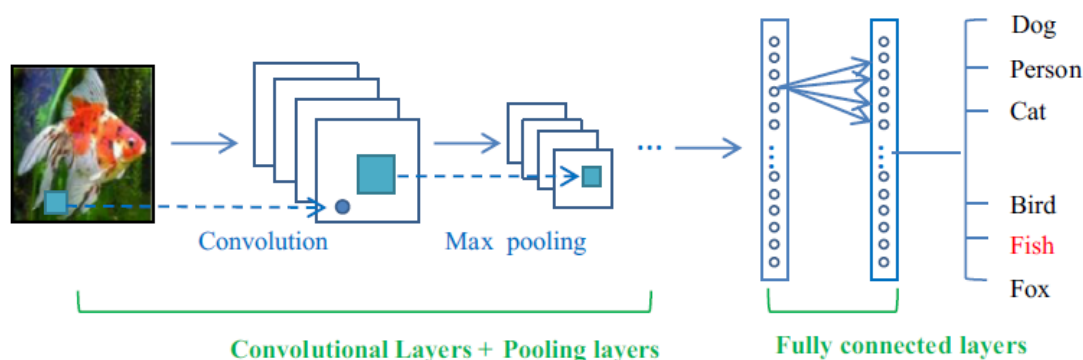
⁴⁸Loss Function

⁴⁹Gradient

⁵⁰Chain Rule



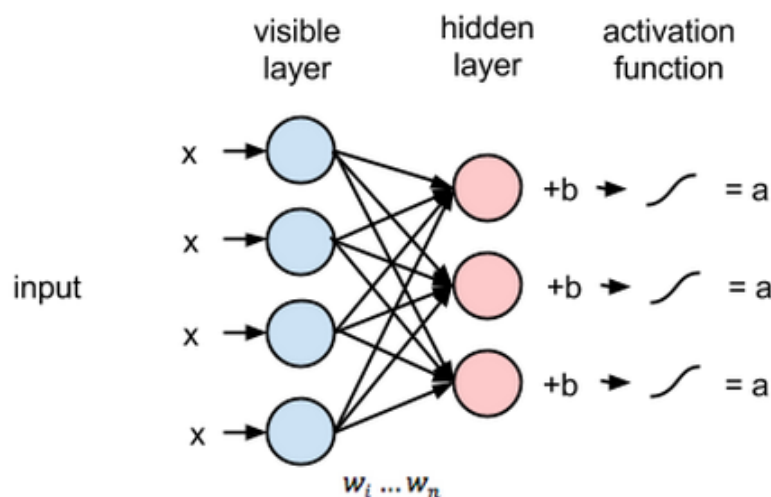
شکل ۸.۲: دسته‌بندی روش‌های یادگیری عمیق و چند مثال از شبکه‌های هر دسته [۴۸]



شکل ۹.۲: طرح کلی از معماری یک شبکه عصبی کانولوشن [۴۸]

۲.۴.۲ ماشین بولتزمن محدود شده

از لحاظ ساختاری، ماشین بولتزمن محدود شده^{۵۱} یک شبکه عصبی مولد و تصادفی^{۵۲} همراه با محدودیت است. منظور از محدودیت آن است که واحدهای قابل مشاهده^{۵۳} و واحدهای پنهان^{۵۴} یک گراف دو بخشی^{۵۵} تشکیل دهند. واضح است که به لایه‌ی شامل واحدهای قابل مشاهده، لایه‌ی قابل مشاهده و به لایه‌ی شامل واحدهای پنهان، لایه پنهان می‌گویند و بر طبق محدودیت RBM واحدهای یک لایه با هم هیچ‌گونه ارتباطی ندارند. این محدودیت باعث می‌شود الگوریتم‌های آموزش کارآمدتر باشند [۶۵].



شکل ۱۰.۲: طرح کلی از معماری یک ماشین بولتزمن محدود شده

^{۵۱}Restricted Boltzmann Machines(RBM)

^{۵۲}Generative Stochastic Neural Network

^{۵۳}Visible Units

^{۵۴}Hidden Units

^{۵۵}Bipartite Graph

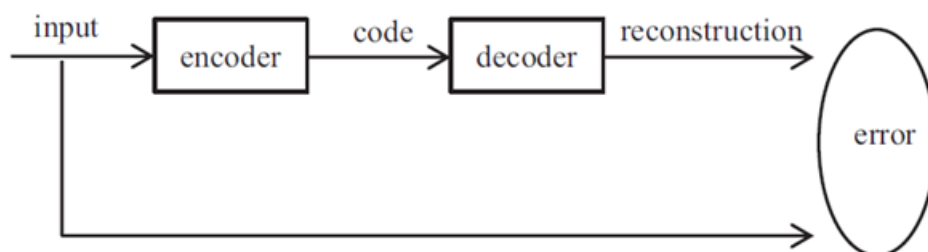
مطابق شکل ۱۰.۲ هدف RBM بازسازی تا حد امکان دقیق ورودی‌ها است. در مرحله‌ی تغذیه رو به جلو، ورودی‌ها توسط وزن‌ها و ارباب‌ها اصلاح و برای فعال‌کردن لایه‌ی پنهان استفاده می‌شوند. از آنجایی که این مدل یک گراف دو بخشی است، واحدهای پنهان H و واحدهای قابل مشاهده V به صورت مشروط مستقل‌اند. بنابراین در معادله زیر:

$$P(HV_1) = P(H_1V_1)P(H_2V_2)...P(H_nV_n) \quad (۴.۲)$$

هم H و هم V_1 توزیع بولتزمن را برآورده می‌کنند. با ورودی V_1 می‌توانیم H را از طریق $P(HV_1)$ به دست آوریم. به همین ترتیب می‌توانیم مقدار V_2 را از طریق $P(HV_2)$ به دست آوریم. با تنظیم پارامترها، اختلاف بین V_1 و V_2 را به حداقل می‌رسانیم در نتیجه H حاصل، به عنوان ویژگی خوبی از V عمل می‌کند. در مقاله [۶۵] توضیح مبسوطی در این زمینه وجود دارد و راه عملی برای آموزش RBM توضیح داده شده است.

۲.۴.۲ خودرمنگار

خودرمنگار^{۵۶} نوع خاصی از شبکه عصبی مصنوعی است که برای رمزگذاری کردن^{۵۷} بهینه یادگیری مورد استفاده قرار می‌گیرد. در ساده‌ترین حالت شامل یک رمزگذار^{۵۸} و یک رمزگشا^{۵۹} به همراه فقط یک لایه پنهان می‌باشد. ورودی به رمزگذار داده شده و خروجی از رمزگشا استخراج می‌شود. به جای آموزش شبکه و پیش‌بینی مقدار هدف Y به ازای ورودی X ، یک خودرمنگار آموزش می‌بیند تا ورودی X خود را بازسازی کند. بنابراین بردارهای خروجی همان ابعاد بردار ورودی را خواهند داشت. فرآیند کلی یک



شکل ۱۱.۲: فرآیند کلی یک خودرمنگار [۴۸]

خودرمنگار در شکل ۱۱.۲ نشان داده شده است. در حین این فرآیند، خودرمنگار با کمینه‌سازی خطای نوسازی^{۶۰} بهینه می‌شود. کد متناظر همان ویژگی پیش‌بینی شده است. به طور کلی یک لایه‌ی تنها، قادر به دریافت ویژگی‌های متمایز از داده خام نیست. در حال حاضر محققان معمولاً از خودرمنگار عمیق استفاده می‌کنند که کد یادگرفته شده از خودرمنگار قبلی را به خودرمنگار بعدی جهت به انجام رساندن کار ارسال می‌کند.

⁵⁶ Autoencoder

⁵⁷ Encode

⁵⁸ Encoder

⁵⁹ Decoder

⁶⁰ Reconstruction Error

۴.۲.۲ کدگذاری تنک

کدگذاری تنک^{۶۱} نوعی روش بدون سرپرست است که به منظور یادگیری مجموعه خیلی کامل^{۶۲} از توابع پایه برای توصیف داده ورودی مورد استفاده قرار می‌گیرد [۶۶]. این روش مزیت‌های متعددی دارد به عنوان مثال می‌توان به این موارد اشاره کرد:

(۱) این روش قادر به بازسازی بهتر توصیه‌گر است زیرا از چندین پایه و ارتباط بین توصیه‌گرهای دارای پایه‌های مشترک مشابه، استفاده می‌کند.

(۲) پراکندگی این اجازه را می‌دهد که خصوصیت‌های برجسته داده ورودی دریافت شوند.

(۳) از آنجا که این روش همسو با سیستم بینایی بیولوژیکی است، ویژگی‌های تنک ورودی‌ها نیز برای یادگیری مفید هستند.

(۴) الگوها با ویژگی‌های تنک از لحاظ خطی بیشتر قابل جداسازی هستند.

برای مقایسه بین این چهار دسته از روش‌های یادگیری عمیق و بدست آوردن درک درستی از آنها، خلاصه‌ای از فواید و مشکلات هریک با توجه به خصوصیات متنوع آنها در جدول ۱۰.۲ آمده است. در این جدول نه خصوصیت کلی وجود دارد. که در آن تعمیم‌پذیری^{۶۳} اشاره به موثر بودن روش مورد نظر برای سایر ورودی‌ها مانند تصویر، صدا و... و کاربردهای متنوع آنها مثل تشخیص صدا^{۶۴}، تشخیص تصویر^{۶۵} و... دارد. یادگیری بدون سرپرست^{۶۶} اشاره به توانایی یادگیری خودکار مدل یادگیری عمیق با داده‌های ورودی بدون برچسب و بدون تفسیر نظارتی دارد. یادگیری ویژگی^{۶۷} مجموعه‌ای از روش‌هاست که به سیستم امکان اکتشاف خودکار ارائه‌های مورد نیاز برای تشخیص ویژگی یا دسته‌بندی بر اساس داده‌های خام را می‌دهد. زمان واقعی آموزش^{۶۸} و زمان واقعی پیش‌بینی^{۶۹} به ترتیب به کارایی فرآیندهای یادگیری و پیش‌بینی اشاره دارند. ادراک بیولوژیکی^{۷۰} و توجیه نظری^{۷۱} نشانگر، شالوده بیولوژیکی یا مبانی نظری روش مورد نظر است. تغییرناپذیری^{۷۲} اشاره به مقاومت روش مورد نظر در مقابل تبدیلاتی مانند دوران، تعجاس و تقارن دارد. مجموعه آموزشی کوچک^{۷۳} اشاره به توانایی یادگیری مدل عمیق با استفاده از تعداد کمی نمونه دارد. این نکته حائز اهمیت است که جدول ۱۰.۲ فقط یافته‌های کلی کنونی را نمایش می‌دهد و صبحتی از فرصت‌های آینده و یا نمونه‌های اختصاصی نمی‌کند.

⁶¹Sparse Coding

⁶²Over-Complete

⁶³Generalization

⁶⁴Speech Recognition

⁶⁵Visual Recognition

⁶⁶Unsupervised Learning

⁶⁷Feature Learning

⁶⁸Real-time Training

⁶⁹Real-time Prediction

⁷⁰Biological Understanding

⁷¹Theoretical Justification

⁷²Invariance

⁷³Small Training Set

Sparse coding	AutoEncoder	RBM	CNN	خصوصیات
بله	بله	بله	بله	تعمیم‌پذیری
بله	بله	بله	خیر	یادگیری بدون سرپرست
خیر	بله*	بله*	بله	یادگیری ویژگی
بله	بله	خیر	خیر	زمان واقعی آموزش
بله	بله	بله	بله	زمان واقعی پیش‌بینی
بله	خیر	خیر	خیر	ادراک بیولوژیکی
بله	بله	بله	بله*	توجه نظری
بله	خیر	خیر	بله*	تغییرناپذیری
بله	بله	بله*	بله*	مجموعه آموزشی کوچک

جدول ۱.۲: مقایسه‌ی روش‌های مختلف شبکه‌های عصبی عمیق [۴۸]

توجه: «بله» اشاره به این دارد که دسته در این خصوصیت به‌خوبی عمل می‌کند. در غیراینصورت با «خیر» علامت‌گذاری می‌شود. «بله*» اشاره به توانایی ابتدایی و یا ضعیف دارد.

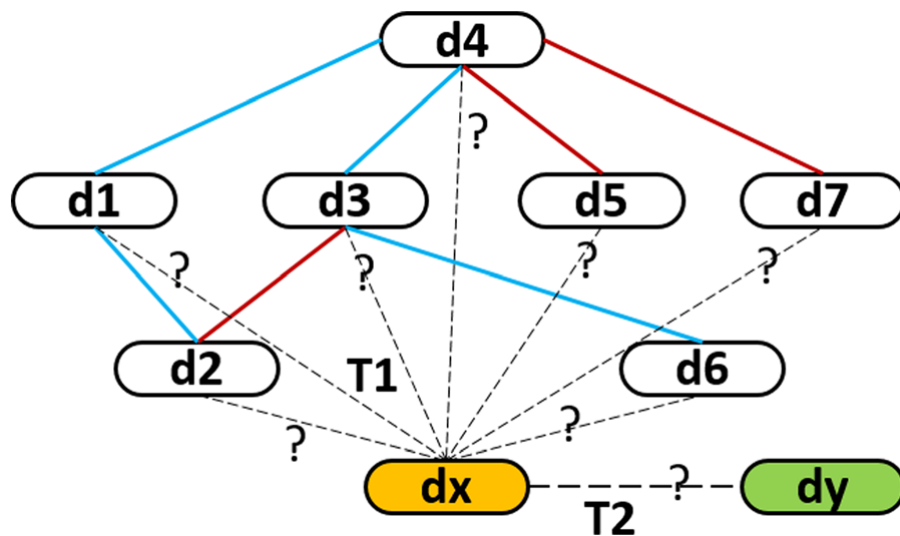
۵.۲ الگوریتم‌های پیشین در مسئله‌ی پیش‌بینی سه‌کلاسه برهم‌کنش دارو-دارو

در این بخش الگوریتم‌های ارائه شده برای پیش‌بینی سه‌کلاسه برهم‌کنش دارو-دارو معرفی و نحوه‌ی عملکرد آنها به‌صورت مختصر شرح داده می‌شود. هر سه الگوریتم معرفی شده، از روش‌های تجزیه ماتریس استفاده می‌کنند. روش تجزیه ماتریس یک مدل توصیه‌گر است که در بخش ۱.۳.۲ توضیح داده شده است. این روش با کمی تغییر یک راه‌حل مناسب برای مسئله‌ی پیش‌بینی برهم‌کنش دارو-دارو است که مثال‌هایی از آن در ادامه‌ی این بخش توضیح داده می‌شود.

لازم به ذکر است که روش پیشنهادی این تحقیق، یک توصیه‌گر مبتنی بر شبکه عصبی عمیق است و از نظر ساختاری هیچ مشابهتی با روش‌های تجزیه ماتریس ندارد. تنها دلیل ذکر این روش‌ها محدود بودن مقالاتی است که در کار خود از داده‌های سه‌کلاسه استفاده کرده‌اند (نگارنده تا زمان نگارش، این سه مقاله را یافته‌است). در فصل ۴ روند ارزیابی الگوریتم ارائه شده و نتیجه‌ی حاصل با نتایج این مقالات مقایسه می‌شود.

۱۰.۵.۲ چارچوب یکپارچه مبتنی بر تجزیه ماتریس سه‌کلاسه

ایده اصلی مدل چارچوب یکپارچه مبتنی بر تجزیه ماتریس سه‌کلاسه^{۷۴}، ایجاد ارتباط بین ویژگی‌های داروهای شناخته شده با ساختار گراف برهم‌کنش‌های دارو-دارو است [۶۸]. چنین ارتباطی را به صورت رگرسیون کمترین مربعات جزئی^{۷۵} مدل می‌کند که به عنوان یک مدل تجزیه ماتریس سه‌کلاسه (TMF) معرفی شده است. برای آموزش مدل رگرسیون از ویژگی عوارض جانبی خارج از برجسب داروها استفاده می‌شود. براساس ادعای نویسندگان این روش اولین روشی است که چارچوب یکپارچه (UF) ارائه می‌دهد. این چارچوب قادر به پیش‌بینی برهم‌کنش‌های دارویی برای هر دو مدل مسئله‌ی دوکلاسه و سه‌کلاسه است.



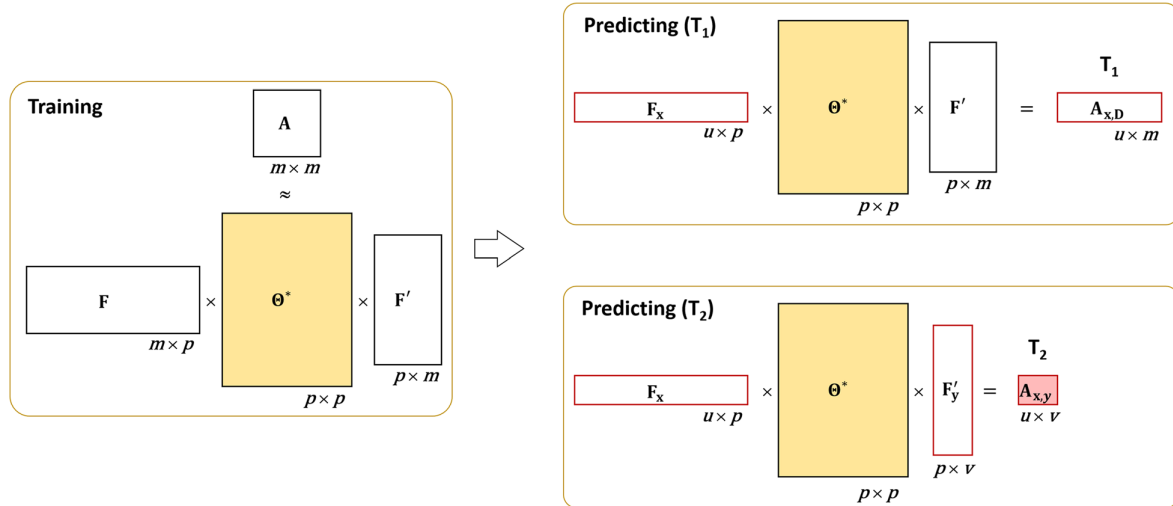
شکل ۱۲.۲: پیش‌بینی برهم‌کنش TMFUF [۶۸]

با توجه به شکل ۱۲.۲ در گراف برهم‌کنش، داروها به عنوان گره معرفی می‌شوند که داروهای شناخته شده از d_1 تا d_7 شماره‌گذاری شده‌اند. برهم‌کنش بین داروها با یال‌ها نشان داده می‌شوند و برهم‌کنش افزایشدهنده و کاهشدهنده به ترتیب توسط یال‌های قرمز و آبی مشخص شده‌اند. d_x و d_y دو داروی جدید هستند که به صورت گره‌های مجزا و به ترتیب با برجسب‌های زرد و سبز مشخص می‌شوند. دو نوع پیش‌بینی، که با T_1 و T_2 برجسب گذاری شده‌اند، با خطوط نقطه‌چین نشان داده می‌شوند. T_1 احتمال برهم‌کنش دارویی d_x با داروهای شناخته شده را پیش‌بینی می‌کند، در حالی که T_2 احتمال برهم‌کنش دارویی d_x با داروهای جدید d_y را پیش‌بینی می‌کند. هر دو آنها پیش‌بینی می‌کنند که برهم‌کنش‌های بالقوه باعث افزایش یا کاهش رفتارهای دارویی می‌شوند.

با توجه به شکل ۱۳.۲ در مرحله آموزش، ماتریس مجاورت برهم‌کنش‌ها A و ماتریس ویژگی داروها F ،

⁷⁴Triple Matrix Factorization-Based Unified Framework (TMFUF)

⁷⁵Partial Least Square Regression



شکل ۱۳.۲: ساختار مدل نظارت شده TMFUF [۶۸]

شکل سمت چپ روند آموزش و دو شکل سمت راست دو حالت پیش‌بینی را برای مدل TMFUF نشان می‌دهند. این مدل برهم‌کنش‌های بین داروهای جدید و داروهای شناخته شده را پیش‌بینی می‌کند (بالا راست). همچنین برهم‌کنش‌های بین داروهای جدید با داروهای جدید را پیش‌بینی می‌کند (پایین راست).

ماتریس تصویرمقارن Θ ، با استفاده از یک مدل رگرسیون Bi-Linear^{۷۶}، به صورت زیر ساخته می‌شود.

$$A = F \cdot \Theta \cdot F' \quad (۵.۲)$$

از آنجا که مقدار Θ را نمی‌توان مستقیماً با معکوس‌گیری معادله ۵.۲ به دست آورد، بنابراین روش جدیدی برای حل آن پیشنهاد شد. در روش پیشنهادی فرض می‌شود فضای ویژگی‌های نهان کم بعدی وجود دارد که در آن هر دارو توسط یک بردار نهان نشان داده می‌شود و ماتریس حاصل ضرب داخلی بین داروها با ماتریس برهم‌کنش داروها ارتباط مستقیم دارد. به علاوه فرض می‌شود ویژگی‌های نهفته داروها با ویژگی مشاهده شده آنها مرتبط است. حل Θ به شرح زیر است.

$$A_d^* = \operatorname{argmin} \| A - A_d \cdot A_d' \|_F,$$

$$B^* = \operatorname{argmin} \| A_d^* - F \cdot B \|_F \quad (۶.۲)$$

$$\Theta^* = B^* \cdot (B^*)'$$

A_d ماتریس برهم‌کنش پنهان است که هر سطر آن نشان‌دهنده بردار ویژگی یک دارو در فضای نهفته است

^{۷۶} رگرسیون دو خطی شیب یک خط رگرسیون برای Y نسبت به X است که به عنوان یک تابع خطی از متغیر سوم Z تغییر می‌کند.

و با توجه به متقارن بودن A ، A_d را با الگوریتم تجزیه به مقدار ویژه به دست می‌آوریم. A_d^* فضای پنهان را با تجزیه ماتریس بازتاب می‌دهد. B ماتریس ضرایب رگرسیون است که رابطه‌ای بین فضای ویژگی‌های مشاهده شده و فضای ویژگی پنهان داروها ایجاد می‌کند. برای به دست آوردن آن از رگرسیون کم‌ترین مربعات جزئی استفاده می‌شود.

بعد از Θ^* ، مدل پیش‌بینی T^1 و T^2 به شرح زیر استخراج می‌شود.

$$\begin{aligned} A_{x,D} &= F_x \cdot \Theta^* \cdot F'_y, \\ A_{x,y} &= F_x \cdot \Theta^* \cdot F'_y \end{aligned} \quad (۷.۲)$$

در مدل کردن سه‌کلاسه، تغییرات داروشناسی ناشی از برهم‌کنش‌ها ثبت می‌شود که پیش‌بینی این مدل با ارزش‌تر و از نظر داروشناسی معتبرتر است. علاوه بر این، راه حلی یکپارچه برای پیش‌بینی برهم‌کنش دارویی در دو سناریو ارائه می‌دهد. در سناریوی اول توانایی مدل برای بازگردانی برهم‌کنش بین داروهای شناخته شده سنجیده می‌شود. در سناریوی دوم توانایی مدل برای بازگردانی برهم‌کنش بین داروهای شناخته شده و داروی جدید محک می‌خورد. بدیهی است که داروهای جدید با داروهای قبلی هیچ برهم‌کنشی ندارند و اصطلاحاً اطلاعات کمتری از آن‌ها وجود دارد. این مدل در هر دو سناریوی مسئله‌ی سه‌کلاسه موظف است نوع برهم‌کنش را پیش‌بینی کند. برهم‌کنش‌های احتمالی جفت داروها افزاینده یا کاهنده هستند. مزیت دیگر روش TMFUF در توانایی نشان دادن وجود تاثیر معنادار جفت ویژگی‌های دارویی (عوارض جانبی خارج از برچسب هر دو دارو) در پیش‌بینی برهم‌کنش دارویی است.

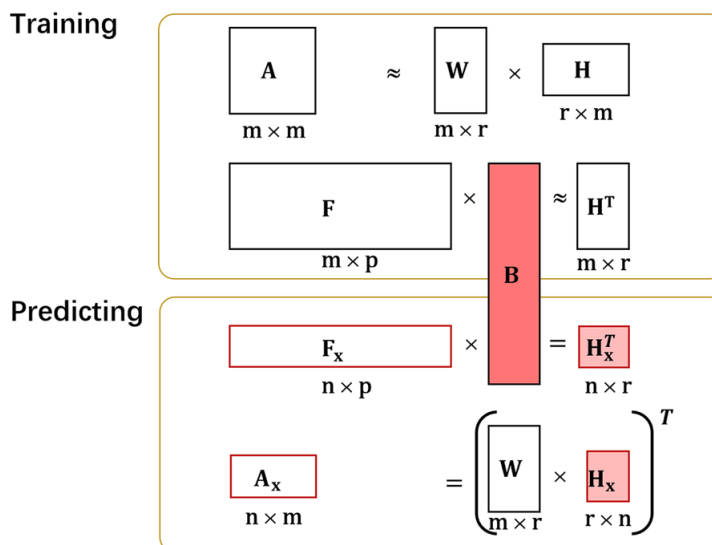
۲.۵.۲ پیش‌بینی برهم‌کنش دارو-دارو بر اساس تجزیه ماتریس غیرمنفی

در روش پیش‌بینی برهم‌کنش دارو-دارو بر اساس تجزیه ماتریس غیرمنفی^{۷۷} (DDINMF) یک مدل نظارت شده برای انجام پیش‌بینی برهم‌کنش دارو-دارو ایجاد می‌شود. [۷۲]

در این روش برهم‌کنش‌ها بر اساس خواص ساختار شبکه‌ی دارویی و در روند تجزیه ماتریس آموزش داده می‌شوند. سپس برای ایجاد رابطه بین ویژگی‌های دارویی (ساختار شیمیایی و عوارض جانبی) و گراف برهم‌کنش داروها از رگرسیون کم‌ترین مربعات جزئی استفاده می‌شود. این ارتباط در فرآیند پیش‌بینی برهم‌کنش بین داروی جدید و داروهای موجود به کار می‌رود. با توجه به شکل ۱۴.۲ مدل DDINMF شامل دو مرحله است.

۱) مرحله‌ی آموزش: در این مرحله ماتریس مجاورت ساخته شده بین داروهای شناخته شده به دو ماتریس پایه W و ماتریس ویژگی پنهان H توسط $A_{m \times m} \approx W_{m \times r} \times H_{r \times m}$ تجزیه می‌شود. در اینجا r اندازه‌ی بعد فضای پنهان است. سپس رابطه‌ای بین ماتریس ویژگی ورودی F و ماتریس ویژگی پنهان H توسط رگرسیون به صورت $(H^T)_{m \times r} = F_{m \times P} \times B_{P \times r}$ مدل می‌شود که B ماتریس ضرایب رگرسیون است.

⁷⁷Drug-Drug Interactions Via Semi- Non Negative Matrix Factorization



شکل ۱۴.۲: بررسی اجمالی DDINMF [۷۲]

۲) مرحله‌ی پیش‌بینی: در این مرحله، ماتریس ضرایب B ماتریس ویژگی ورودی داروهای جدید F_x را به ماتریس ویژگی پنهان جدید به صورت $H_x^T = F_x \times B$ می‌نگارد^{۷۸}. سپس با استفاده از ماتریس ویژگی پنهان جدید، پیش‌بینی بین داروهای جدید و داروهای شناخته شده توسط $A_x = (W \times H_x)^T$ ایجاد می‌شود.

رویکرد ارائه شده شامل دو روش تجزیه ماتریس غیر منفی^{۷۹} (NMF) و تجزیه ماتریس نیمه-غیر منفی^{۸۰} (Semi-NMF) است.

مدل NMF از قید غیرمنفی بودن هر دو ماتریس حاصل از تجزیه استفاده می‌کند تا ماتریس اصلی را تقریب بزند. تابع هدف و تابع بروزرسانی H و W به شرح زیر است.

$$\min ||X - WH||_F^2 = \sum_{i,j} (x_{ij} - \sum_{k=1}^r w_{ik} h_{kj})^2, W, H \geq 0 \quad (۸.۲)$$

$$w_{ik} \leftarrow w_{ik} \frac{(XH^T)_{ik}}{(WHH^T)_{ik}}$$

$$h_{kj} \leftarrow h_{kj} \frac{(X^TW)_{kj}}{(H^TW^TW)_{kj}} \quad (۹.۲)$$

که در آن w_{ik} مراکز خوشه در شبکه‌ی دارویی و h_{kj} شاخص^{۸۱} خوشه یا عضویت نمونه‌ها در شبکه را مشخص می‌کند.

⁷⁸Map

⁷⁹None Negative Matrix Factorization

⁸⁰Semi- None Negative Matrix Factorization

⁸¹Indicator

مدل NMF محدودیت قوی نامنفی بودن را دارد، لذا نمی‌تواند برای حل مسئله‌ی سه‌کلاسه به‌صورت مستقیم استفاده شود. پس لازم است هربار جداگانه روی پیش‌بینی برهم‌کنش‌های افزایشنده و کاهشنده کار کند. اما مدل Semi-NMF با استفاده از قید غیرمنفی بودن تنها یکی از ماتریس‌های فضای نهان این امکان را دارد که برای آموزش و پیش‌بینی مسئله‌ی سه‌کلاسه به‌کار رود. روش Semi-NMF داروها را خوشه‌بندی کرده و اطلاعات و ویژگی‌های بیشتری از داروها را، در کنار پیش‌بینی برهم‌کنش دارویی، ارائه می‌دهد. در این روش نویسندگان برای پیش‌بینی برهم‌کنش داروهای جدید، از الگوریتم متفاوتی برای به حداقل رساندن تابع هدف استفاده می‌کنند.

$$W \leftarrow XH^\dagger$$

$$H \leftarrow H \odot \sqrt{\frac{(W^T X)^{pos} + [(W^T W)^{neg} H]_{ik}}{(W^T X)^{neg} + [(W^T W)^{pos} H]_{ik}}} \quad (۱۰.۲)$$

که در آن H^\dagger ماتریس شبه وارون H است. A^{pos} و A^{neg} به ترتیب ماتریس‌هایی هستند که مولفه‌های منفی و مثبت ماتریس A را با صفر جایگزین کرده‌اند:

$$\forall i, j, A_{ij}^{pos} = \frac{|A_{ij}| + A_{ij}}{2}, A_{ij}^{neg} = \frac{|A_{ij}| - A_{ij}}{2} \quad (۱۱.۲)$$

این رویکرد قادر به پیش‌بینی برهم‌کنش‌های دوکلاسه معمولی و برهم‌کنش‌های جامع سه‌کلاسه است. از همه مهم‌تر، باید این نکته‌ی کلیدی را در نظر داشته باشیم که شبکه‌ی دوکلاسه برهم‌کنش‌های دارو-دارو به هیچ‌وجه حاوی اطلاعاتی در مورد برهم‌کنش‌های افزایشنده و کاهشنده نیست، اما بررسی درجه‌ی رئوس در شبکه‌ی سه‌کلاسه جامع اطلاعات مفیدی از انواع برهم‌کنش‌ها ارائه می‌دهد که می‌تواند به درک بهتر برهم‌کنش‌های دارویی بیانجامد. نحوه‌ی چینش برهم‌کنش‌ها در شبکه سه‌کلاسه جامع نشان می‌دهد بروز علائم ناشی از برهم‌کنش‌های افزایشنده و کاهشنده تصادفی نیست و از قواعد و قوانین خاصی پیروی می‌کند.

گراف سه‌کلاسه‌ی جامع برهم‌کنش یک شبکه با ساختاری متعادل^{۸۲} است. نظریه‌ی تعادل ضعیف^{۸۳} بیان می‌کند که گره‌های شبکه را می‌توان در k خوشه قرار داد، به‌طوری که اکثر یال‌های درون هر خوشه افزایشنده و اکثر یال‌های بین خوشه‌ها کاهشنده باشند [۷۰].

۳.۵.۲ تنظیم متعادل تجزیه ماتریس نیمه غیر منفی

روش موسوم به تنظیم متعادل تجزیه ماتریس نیمه غیر منفی [۷۱]^{۸۴} (BRSNMF) تلاشی برای تقویت روش Semi-NMF پیشین است و علاوه بر آن ساختار اساسی تعادل ضعیف شبکه جامع سه‌کلاسه را نشان می‌دهد. این مدل از روابط ساختاری برای حل دو مسئله استفاده می‌کند:

^{۸۲}Structural Balance

^{۸۳}Weakly Balance Theorem

^{۸۴}Balance Regularized Semi-Nonnegative Matrix Factorization

(۱) کشف انجمن^{۸۵}‌های دارویی و تقویت توان انجمن‌یابی^{۸۶}

(۲) پیش‌بینی جامع برهم‌کنش‌ها.

طبق پیش‌بینی نویسندگان مقاله، شبکه‌ی برهم‌کنش دارویی را می‌توان به انجمن‌ها و گروه‌هایی تقسیم‌بندی کرد. به‌شکلی بیشتر یال (برهم‌کنش)‌های درون انجمن دارویی افزاینده و بیشتر یال‌های بین انجمنی کاهنده باشند. در این راستا دو معیار تنظیم گراف، Gr_{\setminus} برای درون انجمن‌ها و Gr_{\setminus} برای بین انجمن‌ها به‌صورت زیر معرفی شدند. همچنین، این دو معیار تنظیم از به وجود آمدن خوشه‌هایی با گره‌های کم جلوگیری می‌کنند.

$$\begin{aligned} Gr_{\setminus} &= \min \sum_{c=\setminus}^k \frac{h_c^T A^- h_c}{h_c^T h_c}, \\ Gr_{\setminus} &= \min \sum_{c=\setminus}^k \frac{h_c^T L^+ h_c}{h_c^T h_c} \end{aligned} \quad (۱۲.۲)$$

که در آن h_c ، A^- ، L^+ ستون برداری در H است و $L^+ = D^+ - A^+$ است که در آن D^+ ماتریس قطری A^+ است و داریم $a_{ij}^+ = \frac{|a_{ij}| + a_{ij}}{2}$ ، $a_{ij}^- = \frac{|a_{ij}| - a_{ij}}{2}$.

با ترکیب Gr_{\setminus} و Gr_{\setminus} و پس از ساده‌سازی داریم:

$$Gr \equiv \max_{tr}(H^T(\sigma I - \eta(A^- + L^+))H) = \max_{tr}(G) \quad (۱۳.۲)$$

که در آن $\sigma, \eta > 0$ اندازه خوشه‌ها را کنترل می‌کند.

همچنین، معیار تنظیم Sr برای کنترل پراکندگی H به‌صورت زیر طوری تعریف می‌شود که گره‌های دارو در شبکه برهم‌کنش به کمترین انجمن‌های ممکن تعلق داشته باشند:

$$Sr = \sum_j^m \|h_j\|_1^2 = tr(H^T H) = tr(s) \quad (۱۴.۲)$$

که در آن 1 ماتریس $k \times k$ یی است که تمام درآیه‌های آن یک است.

پس از یکپارچه‌سازی معیارهای تنظیم به ماتریس درجه پایین، تابع هدف BRSNMF به شرح زیر است.

$$\min \|A - WH^T\|_F^2 + \alpha \cdot tr(S) - \beta \cdot tr(G), h_{ij} \geq 0, \forall i, j \in [1, \dots, m] \quad (۱۵.۲)$$

^{۸۵}Community

^{۸۶}Community Detection

برای حل و بروزرسانی از قوانین زیر استفاده می‌شود.

$$\begin{aligned} W &\leftarrow AH(H^T H)^{-1} \\ H &\leftarrow H \odot (N \div D)^{\frac{1}{2}} \end{aligned} \quad (۱۶.۲)$$

$$N = (A^T W)_+ + (H W^T W)_- + \beta \eta (L^+ H)_- + \beta \sigma H$$

$$D = (A^T W)_- + (H W^T W)_+ + \alpha H I + \beta \eta A^- H + \beta \eta (L^+ H)_+$$

که عملگرهای X_+ و X_- بصورت $X_+ = \frac{|A| + A}{2}$ و $X_- = \frac{|A| - A}{2}$ تعریف می‌شوند. این تغییرات دو خصوصیت زیر را به نتایج حاصل از اجرای روش تحمیل کرد.

(۱) ارتباطات بین انجمنی بیشتر از نوع کاهنده باشد که کمک کند دوره‌های متعادل در درون انجمن‌ها تشکیل شده و دوره‌های نامتعادل در بین انجمن‌ها باشد.

(۲) از ایجاد خوشه‌هایی با تعداد گره‌های کم جلوگیری شود که باعث شد نسبت به Semi-NMF انجمن‌های خیلی چاق یا خیلی لاغر تشکیل نشود.

اتخاذ چارچوب مسئله شروع سرد^{۸۷} [۷۲] در رویکرد مبتنی بر BRSNMF در سناریو پیش‌بینی برهم‌کنش برای داروی جدید، شامل دو مرحله به شرح زیر است که در شکل ۱۵.۲b نشان داده شده است.

(۱) در مرحله آموزش، رویکرد تجزیه ماتریس $A_{m \times m} \approx W_{m \times k} \times (H_{m \times k})^T$ توسط BRSNMF و رگرسیون خطی $H_{m \times k} = F_{m \times p} \times B_{p \times k}$ توسط رگرسیون کمترین مربعات جزئی به دست آمد.

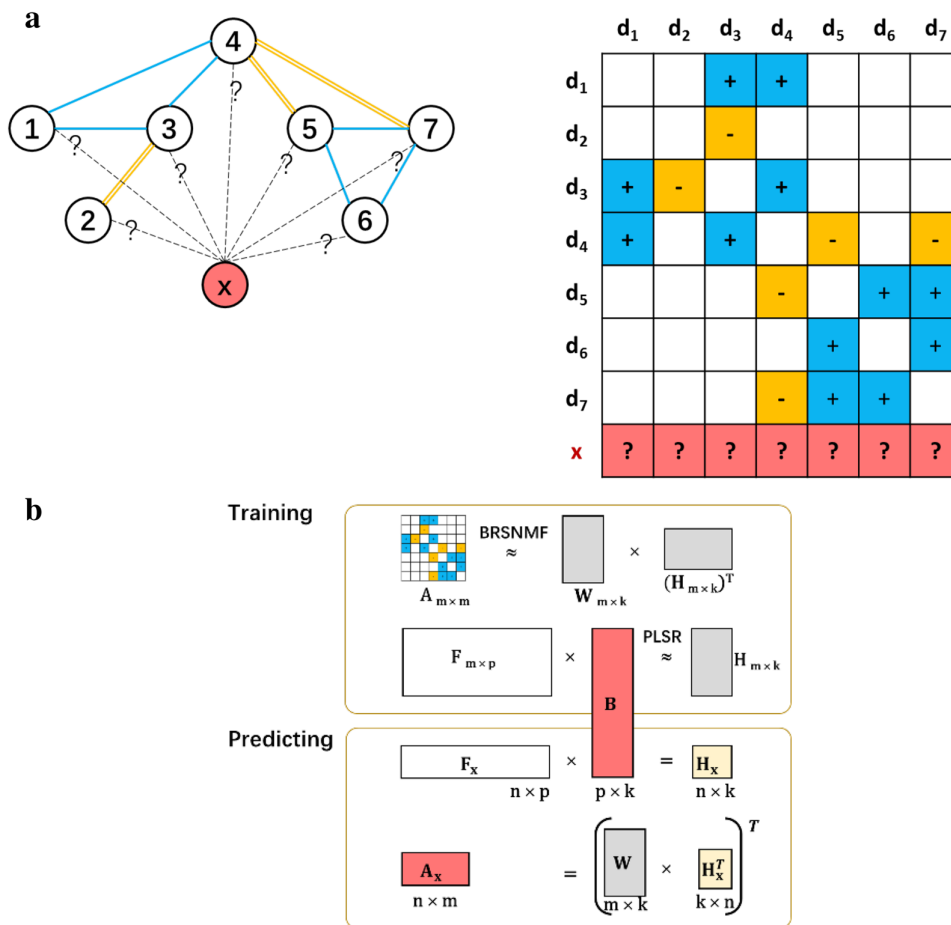
(۲) در مرحله پیش‌بینی، ماتریس ضرایب آموزش داده شده $B_{p \times K}$ در ابتدا ماتریس ویژگی F_x را به فضای پنهان $n \times k$ به صورت $H_x = F_x \times B$ می‌نگارد. سپس برهم‌کنش پیش‌بینی شده $n \times m$ بین داروهای جدید و داروهای تایید شده توسط $A_x = H_x W^T = (F_x B) W^T$ محاسبه می‌شود. در این مرحله از ویژگی پروتئین‌های مرتبط به دارو^{۸۸} استفاده می‌کند.

نتایج نشان می‌دهند که BRSNMF انجمن‌های دارویی ایجاد می‌کند که دارای اندازه‌های مناسب‌تری هستند، خاصیت تعادلی ضعیف بیشتر و اهمیت دارویی دارند.

در این فصل سعی شد پیش‌نیازهای محاسباتی و الگوریتم‌های سامانه‌های توصیه‌گر و یادگیری عمیق معرفی شوند. همچنین الگوریتم‌های مرتبط که قبلاً روی این مسئله کار کرده بودند معرفی و بررسی شدند. در ادامه و در فصل بعد روند آماده‌سازی داده برای ورودی دادن به روش و همچنین پیاده‌سازی روش بیان می‌شود.

⁸⁷Cold Start

⁸⁸Drug Binding Protein(DBP)



شکل ۱۵۲: پیش‌بینی برهم‌کنش جامع در سناریو شروع سرد BRSNMF [۷۱]

فصل ۳

پیش‌بینی برهم‌کنش دارو- داروی با رویکرد سیستم توصیه‌گر یادگیری عمیق

در این فصل، ابتدا داده‌ها و ویژگی‌ها معرفی می‌شوند. سپس الگوریتم جدیدی مبتنی بر (۱) ادغام شباهت‌های دارویی و (۲) سیستم‌های توصیه‌گر یادگیری عمیق برای پیش‌بینی برهم‌کنش دارو - دارو در حالت جامع سه کلاسه ارائه می‌شود. الگوریتم مذکور SNF-CNN نامیده می‌شود که مخفف Predict- ing Comprehensive Drug - Drug Interaction via Similarity Network Fusion and Convolutional Neural Networks است. در این الگوریتم ابتدا روند آماده‌سازی داده‌ها توضیح داده می‌شود و سپس سیستم توصیه‌گری طراحی و بروی برهم‌کنش‌های افزایشنده و کاهشنده آموزش داده می‌شود که جفت داروهای بدون برهم‌کنش را، با احتمال بالا تشخیص می‌دهد. در ادامه سیستم توصیه‌گر قبلی که مبتنی بر شبکه‌ی عصبی کانولوشن است، بروی داده‌های برهم‌کنش افزایشنده و کاهشنده و بدون برهم‌کنش (تشخیص داده‌شده در مرحله‌ی قبل) آموزش داده می‌شود. نتایج در روند اعتبارسنجی متقابل ۱۰ برابری^۱ ثبت و در فصل چهار تحلیل می‌شود.

۱.۳ داده‌ها و ویژگی‌ها

در این پژوهش از مجموعه داده‌ای که در مقاله [۷۲] ارائه شده است، استفاده می‌شود. این مجموعه شامل ۵۶۸ داروی کوچک مولکول تایید شده است که هر دارو حداقل یک برهم‌کنش با دیگر داروهای مجموعه دارد. در مجموع برهم‌کنش‌های بین این ۵۶۸ دارو شامل ۲۱۳۵۱ برهم‌کنش است، که متشکل از ۱۶۷۵۷ برهم‌کنش افزایشنده و ۴۵۹۴ برهم‌کنش کاهشنده می‌باشد.

علاوه بر این، هر دارو با دو بردار ویژگی ۸۸۱ بعدی ساختار شیمیایی F_{str} از PubChem [۱۳] و ۹۱۴۹

^۱10-fold CV

بعدی عوارض جانبی خارج از برچسب F_{se} از OFFSIDES [۱۲] نشان داده می‌شود. برای جزئیات بیشتر به بخش ۲.۳ مراجعه کنید.

داروها و برهم‌کنش‌ها تشکیل گراف می‌دهند. در این گراف گره‌ها دارو و یال‌ها برهم‌کنش هستند. جزئیات مجموعه‌ی داده‌های برهم‌کنش در جدول ۱.۳ ذکر شده است.

مقدار	درجه‌ی D-DDI	مقدار	درجه‌ی E-DDI	مقدار	درجه	مقدار	خصوصیت
۱۸/۱۶	میانگین	۰۰/۵۹	میانگین	۱۸/۷۵	میانگین	۵۶۸	تعداد دارو
۰۰/۸	میانه	۰۰/۴۵	میانه	۵۰/۶۱	میانه	۲۱۳۵۱	تعداد DDI
۲۰۶	بیشینه	۲۳۰	بیشینه	۲۹۶	بیشینه	۱۶۷۵۷	تعداد E-DDI
۰	کمینه	۰	کمینه	۱	کمینه	۴۵۹۴	تعداد D-DDI

جدول ۱.۳: جزئیات مجموعه‌ی داده‌های برهم‌کنش [۷۲]

۲.۳ مدل‌سازی مسئله

$D = \{d_i\}, i = 1, 2, 3, \dots, m$ را به عنوان مجموعه‌ای از m داروی تایید شده دارای برهم‌کنش تعریف می‌کنیم. برهم‌کنش بین m داروی تایید شده را به صورت یک ماتریس مجاورت متقارن $A_{m \times m} = \{a_{ij}\}^2$ در نظر می‌گیریم که m تعداد داروها می‌باشد و a_{ij} هر درآیه از ماتریس A است. برای مقادیر a_{ij} دو دیدگاه متفاوت وجود دارد:

الف) دیدگاه دوکلاسه $a_{ij} = 1$: اگر داروهای i و j اثر متقابل شناخته شده داشته باشند و در غیر این صورت $a_{ij} = 0$. در این دیدگاه تاثیرهای شناخته شده همگی در یک کلاس و با برچسب یک نمایش داده می‌شوند و حالت‌های عدم تاثیر و حالت‌های ناشناخته با برچسب صفر نشان داده می‌شوند.

ب) دیدگاه سه‌کلاسه $a_{ij} = 1$: اگر داروهای i و j اثر متقابل شناخته شده افزاینده داشته باشند و $a_{ij} = -1$ اگر داروهای i و j اثر متقابل شناخته شده کاهنده داشته باشند و در غیر این صورت $a_{ij} = 0$. در این دیدگاه تاثیرات شناخته شده، خود به دو کلاس با برچسب‌های متفاوت تقسیم می‌شوند و همچنان حالت‌های عدم تاثیر و حالت‌های ناشناخته با برچسب صفر نشان داده می‌شوند.

دیدگاه استفاده شده در این پژوهش از نوع سه‌کلاسه است.

علاوه بر این، هر داروی d_i در D به شکل بردار ویژگی p -بعدی $f_i = [f_{i1}, f_{i2}, \dots, f_{ik}, \dots, f_{ip}]$ نمایش داده می‌شود که $f_k = 1$ بیانگر مشاهده‌ی ساختار شیمیایی دارو یا وقوع عارضه جانبی خارج از برچسب،

²Symmetric Adjacency Matrix

³Binary Approach

⁴Triple Approach

⁵Feature Vector

در k امین بعد بردار است و در غیر این صورت $f_k = 0$ می باشد. از آنجا که هر دارو دارای دو بردار ویژگی ساختار شیمیایی و عوارض جانبی خارج از برجسب می باشد، دو ماتریس ویژگی $F_{m \times p}$ تعریف می شود. به قسمی که F_{str} ماتریس ویژگی ساختار شیمیایی و F_{se} ماتریس ویژگی عوارض جانبی خارج از برجسب هستند. به عنوان توضیح کوتاه، داروی شناخته شده/ تایید شده به داروهای موجود در D گفته می شود و داروی جدید دارویی است، که هیچ برهم کنش مشخصی با هیچ یک از داروهای D ندارد. در این پایان نامه روی داروهای جدید غربالگری انجام می دهیم و پیش بینی می کنیم چقدر احتمال دارد یک داروی جدید با یکی یا بیشتر از داروهای شناخته شده برهم کنش داشته باشد.

مدل SNF-CNN

در بخش های ۳.۳ و ۴.۳ به ترتیب روند آماده سازی داده جهت ورودی دادن به مدل های یادگیری ماشین توضیح داده شده است. در ادامه روند انتخاب و اعمال مدل توصیه گر روی داده ها شرح داده می شود. همان طور که در بخش ۳.۳ توضیح داده خواهد شد، روند استفاده از ویژگی ها و شباهت های دارویی به نحوی است که مستقل از نوع آن ها می باشد. لذا مدل و روال آماده سازی داده بر روی ویژگی ها با شباهت های متنوع قابل اعمال و قابل تکرار است. همچنین اگر تعداد ویژگی ها بیش از یک نوع باشد با استفاده از روش SNF، ادغام شده و برای ورودی به ماشین آماده می شود. این ویژگی خاص امکان استفاده ی مجدد روش ارائه شده را برای حالت ها و داده ها مختلف فراهم می کند. همچنین با توجه به این خصوصیت می توان مدل را برای مسائل مشابه در حوزه ی توصیه گر ها مورد استفاده قرار داد.

۳.۳ آماده سازی داده

از آنجا که داروهای جدید گره های جدا شده در شبکه برهم کنش هستند، نمی توانیم برهم کنش احتمالی آنها را تنها با اطلاعات توپولوژیکی استنباط کنیم. بنابراین، اطلاعات اضافی آنها (به عنوان مثال ساختار شیمیایی یا عوارض جانبی) مورد نیاز است، که از نظر یادگیری ماشین به آنها ویژگی دارویی گفته می شود. ابتدا ویژگی ها را مبتنی بر مدل خود آماده می کنیم و سپس یک مدل یادگیری عمیق از پیش بینی برهم کنش را آموزش می دهیم.

۱.۳.۳ محاسبه ی ماتریس شباهت دارویی

همان طور که در بخش ۲.۳ مشاهده شد، مقادیر ماتریس ویژگی ها گسسته اند و همچنین ابعاد ماتریس ها زیادند (ساختار شیمیایی ۸۸۱ بعد و عوارض جانبی خارج از برجسب ۹۱۴۹ بعد). از طرفی الگوریتم های یادگیری ماشین با داده های گسسته ی با ابعاد بالا، خوب کار نمی کنند و نتایج مطلوبی به دست نمی آورند. لذا در ابتدا با استفاده از روش کسینوس مشروح در بخش ۱.۲.۲ ماتریس های شباهت دارویی مبتنی بر

ساختار شیمیایی و عوارض جانبی خارج از برجسب محاسبه می‌شوند. این ماتریس‌ها به ترتیب S_{str} و S_{se} هستند که ابعاد این دو ماتریس $m \times m$ است. هر درآیه از این ماتریس مقداری بین صفر و یک دارد و هر $S_{i,j}$ از ماتریس‌های شباهت، مقدار شباهت داروی d_i و d_j است.

۲.۳.۳ ادغام ماتریس‌های شباهت دارویی

در این مرحله از روش ترکیب شباهت شبکه‌ای که در بخش ۲.۲.۲ توضیح داده شد، استفاده می‌شود. با استفاده از روش مذکور ماتریس‌های شباهت ساختار شیمیایی و عوارض جانبی خارج از برجسب داروها با یکدیگر ادغام شد. خروجی این ادغام ماتریس شباهت جدید S_{snf} است که ابعاد آن ۵۶۸×۵۶۸ است و هر درآیه‌ی آن مقداری بین صفر و یک دارد. برای ترکیب شباهت شبکه‌ای از بسته‌ی SNFPy، پیاده‌سازی شده در پایتون^۶ موجود در آدرس [۷۳]، استفاده شده است.

۳.۳.۳ تشکیل ماتریس ورودی

در این مرحله ماتریسی با ۱۱۳۹ ستون، شامل ستون‌هایی به شرح زیر تشکیل می‌شود:

(۱) جفت داروها: نام داروی i ام و نام داروی j ام.

(۲) نوع تاثیر: کاهنده (۱-)، افزایشنده (۱+) و نامشخص (۰).

(۳) بردار شباهت ۵۶۸ تایی داروی i ام از ماتریس S_{snf}

(۴) بردار شباهت ۵۶۸ تایی داروی j ام از ماتریس S_{snf}

۵۶۸ دارو داریم و برهم‌کنش هر دارو با خودش بی‌معنی است. از طرفی جفت داروهای (d_i, d_j) و (d_j, d_i) دوگان هم بوده و حضور هر دوی آن‌ها در داده موجب افزایش داده‌های آموزش می‌شود که متعاقباً توان ماشین را در پیش‌بینی بهتر افزایش می‌دهد. در نتیجه ماتریس حاصل ۳۲۲۰۵۶ نمونه یا ردیف داده دارد. باتوجه به توضیحات ارائه‌شده، ماتریس B با ابعاد ۱۱۳۹×۳۲۲۰۵۶ جهت ورودی دادن به ماشین توصیه‌گر، تشکیل می‌شود.

۴.۳ طراحی سیستم توصیه‌گر

در مراحل قبل داده برای ورود به هر ماشین یادگیرنده از جمله ماشین‌های یادگیری عمیق آماده شد. اما قبل از ارائه‌ی مدل و ورود داده به ماشین باید یک نکته‌ی مهم را در نظر گرفت. همان‌طور که در بیان مسئله ذکر شد، چه در این تحقیق و چه در دیگر تحقیقات، داده‌های مثبت یک یا منفی یک برجسب‌های مشخص و معینی هستند. درحالی که برجسب صفر به هیچ عنوان به معنی عدم وجود برهم‌کنش در بین یک جفت دارو نیست، بلکه بیان می‌کند که برای این جفت دارو هنوز برهم‌کنشی یافت نشده است. در

^۶Python

ادامه روشی برای تشخیص جفت داروهای بدون برهم‌کنش ارائه می‌دهیم. سپس از این جفت داروها به عنوان داده‌هایی با برچسب صفر در آموزش بعدی استفاده می‌شود.

۱.۴.۳ روند انتخاب و آموزش مدل روی برهم‌کنش‌های شناخته‌شده

برای حل این مسئله نیاز است مدلی ارائه شود که عدم برهم‌کنش را با دقت و اطمینان خوبی تشخیص دهد. داده‌ها و تحقیقات قبلی از داده‌های دوکلاسه استفاده می‌کردند که دراصل تنها دسته‌ی مثبت (وجود برهم‌کنش) قابل اعتماد است و نمی‌توان به دسته‌ی منفی (عدم وجود برهم‌کنش) اعتماد کرد. در اینصورت ارائه‌ی مدل با هدف تشخیص صفرهای واقعی‌تر با استفاده از داده‌های دوکلاسه امری سخت و کمتر قابل اعتماد است. از طرفی داده‌های سه‌کلاسه این نوید را می‌دهند که با افزایش دسته و تقسیم بندی انواع برهم‌کنش، مشخصات عدم برهم‌کنش به شکل بهتری توسط مدل‌ها بازنمایی شود.

باتوجه به توضیحات داده‌شده و با وجود استفاده از داده‌های سه‌کلاسه این تحقیق سعی می‌کند ابتدا مدلی مبتنی بر یادگیری عمیق ارائه دهد که عدم برهم‌کنش احتمالی جفت داروها را پیش‌بینی کند. بدیهی است، دقت بالا در تشخیص این صفرها می‌تواند در ارائه‌ی مدل سه‌کلاسه دقیق‌تر و قابل اعتمادتر کمک شایانی کند.

۱.۱.۴.۳ روند اعتبارسنجی جهت انتخاب مدل

سطرهایی از ماتریس B که شامل برهم‌کنش‌های مثبت یک و منفی یک هستند، جدا می‌شود. ماتریس جدید شامل ۴۲۷۰۲ جفت دارو با تاثیر کاهنده و افزایشنده است. این داده برای آموزش و یافتن مدلی مناسب‌تر استفاده شد تا از بین تعداد زیادی از مدل‌ها با ساختارهای شبکه‌ای مختلف مدل قوی‌تری پیدا و استفاده شود. مدل نهایی شبکه‌ی عصبی عمیقی بود که از لایه‌های کانولوشن و تماماً متصل بهره می‌برد.

ویژگی‌های همه برهم‌کنش‌ها (مثبت یک و منفی یک) شامل ۱۱۳۶ ویژگی است. در ابتدا این ویژگی‌ها را به ۱۰ دسته مساوی تقسیم می‌کنیم. سپس در یک حلقه ۱۰ تایی هر بار، یک دسته را به عنوان مجموعه ارزیابی و ۹ دسته دیگر را به عنوان مجموعه داده آموزش در نظر می‌گیریم.

مدل‌های مختلفی را انتخاب کرده و در روند اعتبارسنجی مذکور با ۹۰ درصد داده‌ها، مدل را آموزش می‌دهیم. سپس بر روی ۱۰ درصد باقیمانده داده‌ها، ارزیابی مدل را انجام می‌دهیم. در روند جداسازی جفت داروهای دوگان در نظر گرفته شده است. از آنجا که جفت داروهای (d_i, d_j) و (d_j, d_i) از نظر زیستی فرقی با هم ندارند پس در جداسازی داده‌ی آموزش و ارزیابی همواره یک جفت دارو و دوگان آن لزوماً در یک گروه یکسان قرار می‌گیرند. این امر از شبهه‌ی تقلب ماشین جلوگیری می‌کند.

در روند انتخاب مدل، از مدل‌های یادگیری ماشین کلاسیک نظیر جنگل تصادفی^۷، ماشین بردار

^۷Random Forest

پشتیبان^۸، برازش بخت^۹، شبکه‌ی پرسپترون چند لایه^{۱۰}، شبکه‌ی حافظه طولانی کوتاه مدت^{۱۱}، شبکه‌ی عصبی کانولوشن یک بعدی و دو بعدی، شبکه عصبی مصنوعی^{۱۲}، خودرمنزنگار و ترکیبی از آنها استفاده کردیم.

۲.۱.۴.۳ ارائه‌ی مدل انتخابی (شبکه عصبی کانولوشن)

بعد از آزمودن ساختارهای مختلف، شبکه‌ی عصبی عمیق مورد استفاده نهایی را همانند ساختار شکل ۱.۳ مدل‌سازی کردیم. این شبکه دارای سه لایه کانولوشن دو بعدی و در امتداد آن سه لایه‌ی تماماً متصل کانولوشن می‌باشد. لایه‌ی آخر دو خروجی برای پیش‌بینی تاثیر کاهنده و یا افزایشنده دارد. لایه‌های کانولوشن از فیلترهای مربعی با ابعاد ۴ و گام ۱^{۱۳} استفاده می‌کنند. همچنین هر لایه‌ی کانولوشن دارای یک تابع فعال‌سازی^{۱۴} Relu است. تعداد فیلترهای کانولوشن به ترتیب ۱۲۸، ۳۲ و ۸ می‌باشد. لایه‌های تماماً متصل به ترتیب ۶۴، ۱۶ و ۲ گره دارند، دو لایه‌ی اول دارای فعال‌سازی Relu بوده و لایه‌ی آخر با ۲ گره دارای فعال‌سازی Sigmoid می‌باشد. لایه‌های کانولوشن با استفاده از لایه‌ی هموارکننده^{۱۵} به لایه‌های تماماً متصل مرتبط می‌شود. وظیفه‌ی این لایه تغییر شکل مستطیلی دو بعدی به حالت برداری یک بعدی است. خروجی این لایه ورودی لایه‌ی اول تماماً متصل می‌باشد. همچنین بین لایه‌های تماماً متصل ۶۴ و ۱۶ گره‌ای از یک لایه‌ی Dropout با مقدار دور ریخت ۲/۰ استفاده می‌شود. مقدار ۲/۰ بیان می‌کند که شبکه در این لایه بصورت تصادفی بیست درصد ویژگی‌ها را در نظر نمی‌گیرد. این لایه بدین منظور استفاده می‌شود که از بیش‌برازش^{۱۶} مدل جلوگیری کند و مدل را وادار می‌کند تا تعداد ویژگی‌های بیشتر و با اعتماد بالاتری را جهت پیش‌بینی استخراج و مورد استفاده قرار دهد و در صورت حذف تعدادی از آنها توان پیش‌بینی الگوریتم افت نکند و متکی به چند ویژگی خاص نباشد.

در بررسی‌ها مشاهده شد لایه‌های کانولوشن دوبعدی بهتر از نوع یک بعدی آنها کار می‌کنند، زیرا در این حالت فیلترها می‌توانند شباهت‌های دارویی بیشتری را در هنگام پیمایش زیر نظر گرفته و این امکان وجود دارد ویژگی‌های قدرتمندتری را استخراج کنند. لذا بردارهای ویژگی^{۱۱۳۶} بعدی به ماتریس‌هایی با ابعاد ۱۶×۷۱ تغییر فرم داده می‌شوند. در شکل ۲.۳ تعداد وزن‌های قابل یادگیری هر لایه مشخص شده است. همچنین تعداد کل وزن‌ها که نشان دهنده‌ی میزان پیچیدگی کلی مدل است محاسبه شده است.

^۸Support Vector Machine(SVM)

^۹Logistic Regression

^{۱۰}Multilayer Perceptron

^{۱۱}Long Short-Term Memory(LSTM)

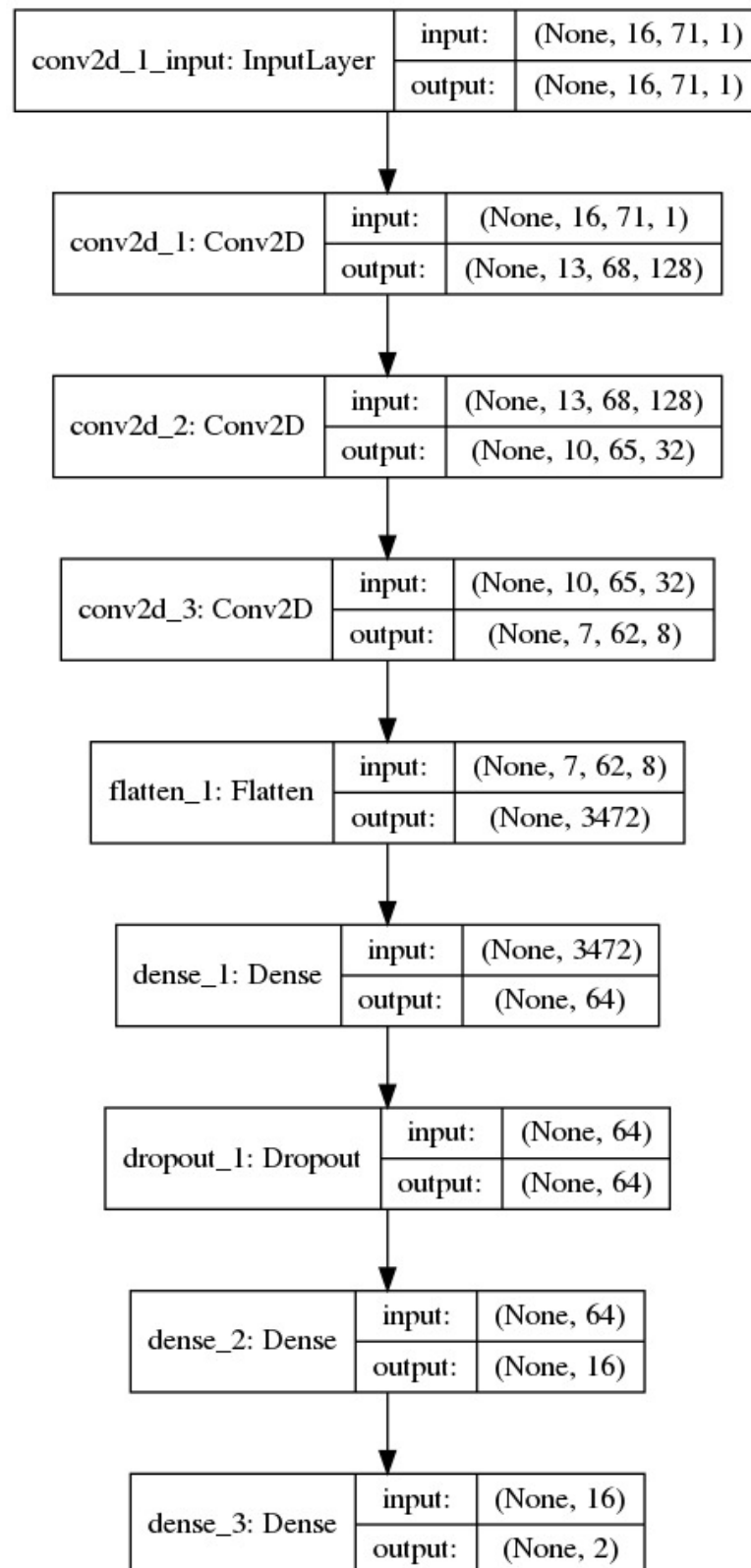
^{۱۲}Artificial Neural Network

^{۱۳}Stride

^{۱۴}Activation Function

^{۱۵}Flatten

^{۱۶}Overfit



شکل ۴.۳: نحوه‌ی چینش لایه‌های شبکه عصبی تشخیص صفرهای احتمالی

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 13, 68, 128)	2176
conv2d_2 (Conv2D)	(None, 10, 65, 32)	65568
conv2d_3 (Conv2D)	(None, 7, 62, 8)	4104
flatten_1 (Flatten)	(None, 3472)	0
dense_1 (Dense)	(None, 64)	222272
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 16)	1040
dense_3 (Dense)	(None, 2)	34
Total params: 295,194		
Trainable params: 295,194		
Non-trainable params: 0		

شکل ۲۰۳: وزن‌های قابل یادگیری شبکه‌ی عصبی پیش‌بینی دوکلاسه

در ساخت شبکه عصبی کانولوشن از تنظیمات زیر استفاده می‌شود:

(۱) برای پیاده‌سازی شبکه عصبی از بسته‌های Tensorflow [۷۴] (نسخه‌ی ۰.۱۴.۱) و Keras [۷۵] (نسخه‌ی ۰.۲.۲) استفاده شد.

(۲) از تابع بهینه‌سازی ADAM استفاده شد.

(۳) تابع خطا Categorical-cross entropy در نظر گرفته شد.

(۴) تعداد اپوک ۱۷ ها ۵ در نظر گرفته شد.

(۵) نرخ یادگیری ۰.۰۰۰۱/۰^{۱۸} استفاده شد.

در نظر داشته باشید فرایامترهای^{۱۹} این شبکه بهینه نشده است و پارامترهای مشخص شده لزوماً در بهترین حالت خود نیستند. برای عدم بهینه‌سازی فرایامترها دو دلیل وجود دارد:

(۱) اجتناب از بیش‌برازش مدل^{۲۰}: در صورت تغییر فرایامترها به بهترین مقادیر، انتظار می‌رود، مدل نتایج بهتری روی داده‌ی حاضر بگیرد، اما تضمینی وجود ندارد ویژگی‌های استخراج شده توسط مدل،

¹⁷Epoch

¹⁸Lerning Rate

¹⁹Hyperparameter

²⁰Overfitting

معنی‌دار بوده و در صورت استفاده از مدل بر روی داروهای جدید به‌خوبی عمل کند. در این صورت اصطلاحاً مدل بیش‌برازش شده و نکته‌ی منفی برای مدل خواهد بود.

(۲) حفظ استواری^{۲۱}: فرایپارامترهای بهینه برای داده‌ی حاضر نتایج بهتری می‌دهند اما ممکن است در آینده از شباهت‌های دارویی متفاوتی استفاده شود یا داده‌های جدیدی جمع‌آوری شود و نتایج حاضر تکرار نشود. در این صورت مدل قوت و مستحکمی خود را از دست داده و مقبولیتی در جامعه‌ی داروسازی و داروشناسی نخواهد داشت.

در نهایت نتایج مدل پیشنهادی را در روال اعتبارسنجی ۱۰ برابری از سه دیدگاه بررسی می‌کنیم:

(۱) دقت مدل: مدل در روال اعتبارسنجی ۱۰ برابری برای تشخیص برهم‌کنش‌های کاهنده $AUC = 0.97$ و $AUPR = 0.93$ بدست آمد و برای تشخیص برهم‌کنش‌های افزایشنده $AUC = 0.97$ و $AUPR = 0.99$ بدست آمد. این نتایج حاکی از دقت و توان تشخیص بالای مدل می‌باشد.

(۲) واریانس نتایج: بازه‌ی اطمینان برای مقادیر گزارش شده با ضریب اطمینان بالای ۹۵ درصد باریک و نزدیک به هم بوده است. به شکلی که از چهار مقدار از سه مقدار گزارش شده کمتر از ۰.۰۲ بوده و فقط برای تشخیص کاهنده مقدار $AUPR$ در بازه‌ی به اضافه و منهای ۰.۰۵ بوده است. با توجه به مقدار کم واریانس بدست آمده از مدل کاملاً مشخص است که مدل ارایه شده استوار می‌باشد.

(۳) توانایی تفکیک‌پذیری مدل: با رسم نمودار توزیع احتمالی خروجی ماشین مطابق شکل ۳.۳ مشخص است که مقادیر ۱ و ۱- به‌خوبی از هم جدا شده‌اند و توزیع احتمال کاهنده و افزایشنده اشتراک کمی دارند.

شبه کد ۱ روند انتخاب مدل را مرحله به مرحله نشان می‌دهد.

الگوریتم ۱ شبه کد انتخاب مدل

ورودی: ویژگی‌های جفت داروهای +۱ و -۱
خروجی: مدل تشخیص دهنده +۱ و -۱

۱: اعتبارسنجی متقابل ۱۰ برابری را روی ویژگی‌های جفت داروهای +۱ و -۱ اعمال کن.

۲: مدل مناسب را انتخاب کن.

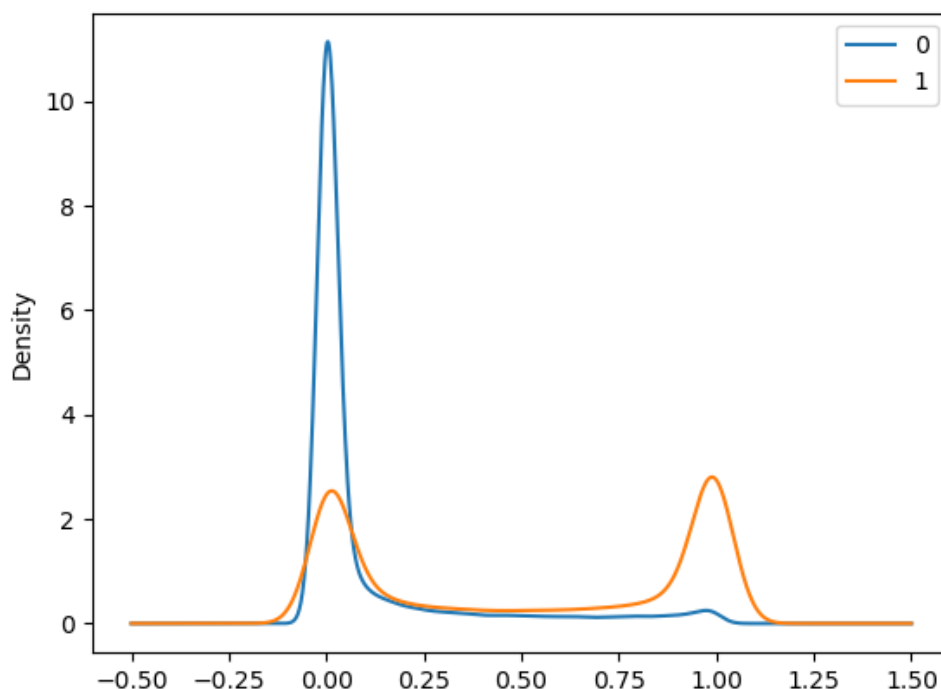
۳: نتایج مدل را در روال اعتبارسنجی متقابل ۱۰ برابری بررسی کن.

۴: در صورت ارضای شرط ۳ مدل انتخابی را برگردان در غیر این صورت به مرحله ۲ بازگرد.

۲.۴.۳ تشخیص جفت داروهای بدون برهم‌کنش احتمالی

در مرحله‌ی قبل مدلی با دقت بالا، قوی و مستحکم ارائه شد که می‌توانست تأثیرات افزاینده و کاهنده جفت داروها را به‌خوبی مدل کرده و تشخیص دهد. لذا این مدل به شرح زیر توانایی تشخیص

²¹Robust



شکل ۳.۳: نمودار توزیع چگالی احتمالی کاهنده و افزایشنده در این شکل ۰ همان برجسب +۱ بوده و ۱ همان -۱ است.

غیربرهم‌کنش‌ها (صفرهای واقعی) را دارد. اگر با احتمال کمی جفت دارویی نامزد برهم‌کنش باشند، آنگاه آن جفت دارو، به احتمال زیاد، صفرهای واقعی هستند.

با توجه به این فرض، از مدل برای انجام پیش‌بینی روی کلیه‌ی جفت داروهای ناشناخته (صفرها) استفاده شد. جفت داروهای ناشناخته شامل ۲۷۰۰۰۰ جفت دارو می‌باشد. در خروجی مدل، جفت داروهایی که با احتمال کمتر از ۰/۴ افزایشده و با احتمال کمتر از ۰/۴ کاهنده بودند را به عنوان جفت داروهای بدون برهم‌کنش در نظر می‌گیریم. از بین داده‌هایی با برجسب ناشناخته حدود ۶۵۰۰۰ جفت دارو شرایط مذکور را داشتند. این دسته از جفت داروها نامزد عدم برهم‌کنش هستند. با توجه به دقت بالای مدل، واریانس کم نتایج و توانایی تفکیک‌پذیری بالای مدل، جفت‌های مذکور را به عنوان جفت داروهای فاقد برهم‌کنش در نظر می‌گیریم.

۳.۴.۳ آموزش مدل روی برهم‌کنش‌های شناخته‌شده و ناشناخته

در این بخش از داده‌های شناخته‌شده و نامزدهای بالقوه برای عدم برهم‌کنش در جهت تشکیل مجموعه داده استفاده می‌شود. در این نگارش از این پس، جفت داروهای نامزد عدم برهم‌کنش را با عبارت صفر

واقعی جایگزین کرده و به کار می‌بریم. همچنین برای مدل نهایی نیز از مدل توصیه‌گر ارائه شده در بخش ۲.۱.۴.۳ استفاده می‌شود. در ادامه روند کار به شکل مبسوط آمده است.

۱.۳.۴.۳ روند اعتبارسنجی نتایج مدل پیش‌بینی برهم‌کنش داروها

ابتدا سطرهایی از ماتریس B که شامل برهم‌کنش‌های مثبت یک و منفی یک هستند مطابق روند مشروح در بخش ۱.۱.۴.۳ جدا شده و در ۱۰ دسته قرار می‌گیرند. سپس به صورت تصادفی از میان ۶۵۰۰ جفت داروی نامزد بدون برهم‌کنش، ۳۰۰۰ جفت دارو انتخاب شد. در جفت داروی انتخابی، حتما باید خود جفت و دوگان آن نامزد بدون برهم‌کنش باشند. این گروه از صفرها به صورت تصادفی به ۱۰ دسته تقسیم می‌شوند. به گونه‌ای که دسته‌ی هر جفت دارو و دوگان یکسان باشد. سپس ۱۰ دسته‌ی صفرها با ۱۰ دسته‌ی از قبل آماده‌ی +۱ و -۱ ها ادغام می‌شوند.

حال مجموعه داده‌ای شامل تقریباً ۷۲۷۰۲ جفت دارو به دسته‌ی نسبتا برابر تقسیم شده و آماده‌ی استفاده برای روند آموزش و ارزیابی مدل نهایی توصیه‌گر می‌باشد.

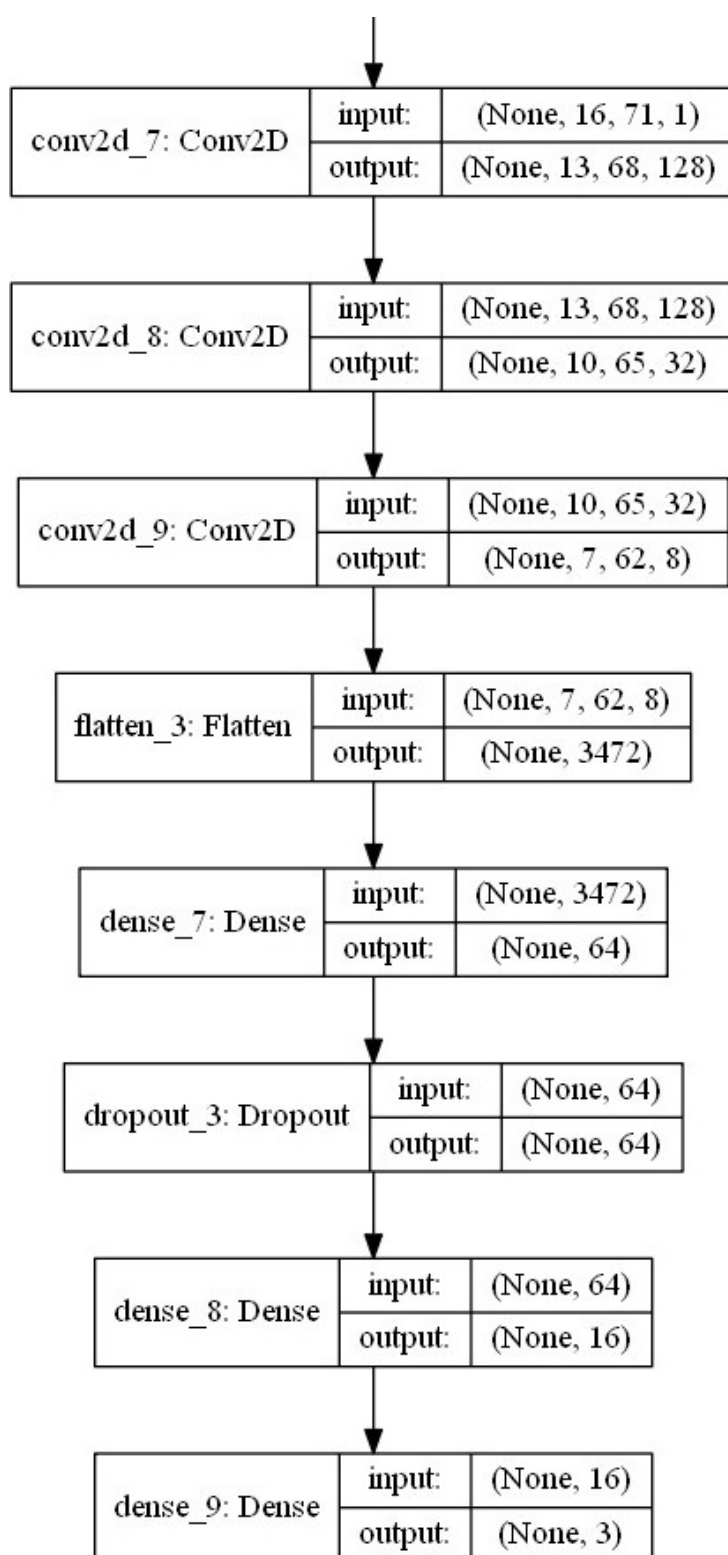
۲.۳.۴.۳ مدل نهایی پیش‌بینی برهم‌کنش داروها

در قسمت قبل طریقه‌ی آماده‌سازی داده برای روال اعتبارسنجی مدل شرح داده شد. در این مرحله مدل نهایی ارائه می‌شود. مدل نهایی تقریباً همان مدل بیان شده در بخش ۲.۱.۴.۳ است. یعنی شبکه سه لایه‌ی کانولوشن با تعداد فیلترهای کانولوشن ۱۲۸، ۳۲ و ۸ دارد. سپس همانند قبل از لایه‌های کانولوشن سه لایه تماما متصل استفاده شد. با این تفاوت که تعداد گره‌ها از ۱۶، ۶۴ و ۲ به ترتیب به ۱۶، ۶۴ و ۳ گره در هر لایه تغییر کرد. مشخصا این مدل سه خروجی احتمالی برای سه حالت برهم‌کنش افزایشده، عدم برهم‌کنش و برهم‌کنش کاهنده می‌دهد. همچنین تعداد ایپوک‌ها ۹ در نظر گرفته شد. روند انتخاب ایپوک در بخش ۱.۴ شرح داده شده است. مدل شبکه‌ی عصبی عمیق جهت پیش‌بینی برهم‌کنش در شکل ۴.۳ نشان داده شده است. در این مرحله از انتخاب مدل جدید خودداری شد، زیرا:

(۱) توان این مدل برای تشخیص نسبتا دقیق برهم‌کنش افزایشده و کاهنده در آخر بخش ۲.۱.۴.۳ به اثبات رسیده است.

(۲) صفرهای واقعی استفاده شده در این بخش پیشنهادی هستند و توسط آزمایشگاه داروشناسی تایید نشده‌اند و تا زمان نگارش این نوشتار پایگاه جامعی برای موارد عدم برهم‌کنش به صورت عمومی منتشر نشده است. لذا اگر روند انتخاب مدل مجددا انجام شود، ممکن است مدلی انتخاب و استفاده شود که از نظر کاربرد در دنیای واقعی معتبر نبوده و مورد قبول قرار نگیرد.

با توجه به دلایل بالا مدل توصیه‌گر صفرها، با تغییر تعداد خروجی‌ها از ۲ به ۳ و تغییر داده‌های ورودی برای پیش‌بینی برهم‌کنش‌های دارویی به کار می‌رود. فرآیند کلی روش پیشنهادی SNF-CNN در قالب شبه کد ۲ ارائه شده است که شامل مراحل آماده‌سازی، انتخاب مدل، تشخیص صفر واقعی و ارائه‌ی توصیه‌گر جامع می‌باشد.



شکل ۴.۳: چیدمان لایه‌های شبکه عصبی SNF-CNN پیش‌بینی سه‌کلاسه برهم‌کنش
عدم برهم‌کنش (۰)، برهم‌کنش کاهنده (۱-) و برهم‌کنش افزایشنده (۱+)

الگوریتم ۲ شبه کد روال کلی الگوریتم SNF-CNN

ورودی: ویژگی‌های جفت داروها $(+1, -1, 0)$ و $(0, +1, -1)$ واقعی
 خروجی: مدل تشخیص دهنده‌ی نوع برهم‌کنش و عدم برهم‌کنش $(+1, -1, 0)$

- ۱: ماتریس‌های شباهت دارویی را با استفاده از روش کسینوس محاسبه کن.
- ۲: ماتریس‌های شباهت دارویی را با استفاده از روش ترکیب شباهت شبکه‌ای ادغام کن.
- ۳: ماتریس ورودی مدل را تشکیل بده.
- ۴: مدل مناسب برهم‌کنش‌های شناخته شده را انتخاب کن و آموزش بده.
- ۵: صفرهای احتمالی را با استفاده از مرحله ۴ پیش‌بینی کن.
- ۶: مدل مناسب برهم‌کنش‌های شناخته شده و صفرهای مرحله ۵ را انتخاب کن و آموزش بده.
- ۷: روی جفت داروهای ناشناخته پیش‌بینی انجام بده.

خلاصه‌ی فصل سوم و نگاهی به فصل چهار

در این فصل روندی جامع برای آماده‌سازی داده ارائه شد که می‌تواند بر روی ویژگی‌های مختلف اعمال شود و به انواع ماشین‌های یادگیرنده ورودی داده شود. سپس مدل توصیه‌گری ارائه شد که نه تنها توانایی تشخیص نوع برهم‌کنش (افزاینده و کاهنده) را دارد بلکه می‌تواند جفت داروهایی را که به احتمال قوی باهم برهم‌کنش ندارند را نیز تشخیص دهد. روش ارائه شده به نام SNF-CNN نام‌گذاری شد. مدل توصیه‌گر سه‌کلاسه جامع در روند اعتبارسنجی آموزش و ارزیابی شد که نتایج آن در فصل چهار گزارش می‌شود. سپس برای اعتبارسنجی داروشناسی مدل با تمام جفت داروهای شناخته‌شده و صفرهای واقعی آموزش داده شد و روی تمام صفرها پیش‌بینی انجام شد و نتایج پیش‌بینی در پایگاه داده‌ی دارویی جست‌وجو شد که نتیجه اعتبارسنجی داروشناسی را نیز در فصل چهار ذکر شده و بررسی می‌شود.

فصل ۴

بررسی نتایج و نتیجه‌گیری

در این فصل به بررسی نتایج رویکرد ارائه شده و مقایسه‌ی آن با سایر روش‌های پیش‌بینی خواهیم پرداخت و در آخر نتیجه‌گیری و جمع‌بندی کلی ارائه خواهیم کرد.

۱.۴ روش جامع اعتبارسنجی متقابل k برابری

روش جامع اعتبارسنجی متقابل k برابری^۱ رویکردی تثبیت شده برای اعتبارسنجی قدرت الگوریتم‌ها در یادگیری ماشین است. برای نشان دادن این واقعیت که داروهای جدید چه نوع برهم‌کنشی دارند و برای جلوگیری از پیش‌بینی بیش از حد خوش‌بینانه، اعتبارسنجی متقابل باید به‌طور دقیق طراحی شود. برای جفت داروهایی که هیچ برهم‌کنشی شناخته شده‌ای ندارند، اعتبارسنجی متقابل سعی در پیش‌بینی نوع برهم‌کنشی جدید بین آنها و داروهای دارای برهم‌کنش شناخته شده دارد. تولید نمونه‌های آموزش و آزمایش به شرح زیر است:

کل مجموعه داده‌ها به k قسمت مساوی تقسیم می‌شوند. $k-1$ قسمت به عنوان مجموعه داده‌های آموزشی استفاده می‌شود و براساس آن مدل ساخته می‌شود و با یک قسمت باقی‌مانده عملیات ارزیابی انجام می‌شود. فرآیند مزبور به تعداد k مرتبه تکرار خواهد شد، به گونه‌ای که از هر کدام از k قسمت تنها یکبار برای ارزیابی استفاده شده و در هر مرتبه یک دقت^۲ برای مدل ساخته شده، محاسبه می‌شود. در این روش ارزیابی دقت نهایی دسته‌بند^۳ برابر با میانگین k دقت محاسبه شده خواهد بود. معمول‌ترین مقداری که در متون علمی برای k در نظر گرفته می‌شود برابر با ۵ یا ۱۰ می‌باشد. بدیهی است هر چه مقدار k بزرگتر شود، دقت محاسبه شده برای دسته‌بند قابل اعتمادتر بوده و دانش حاصل شده جامع‌تر خواهد بود و البته افزایش زمان ارزیابی دسته‌بند نیز افزایش می‌یابد که مهم‌ترین مشکل محسوب می‌شود. هر

¹k-Fold Cross Validation (K-fold CV)

²Precision

³Classifier

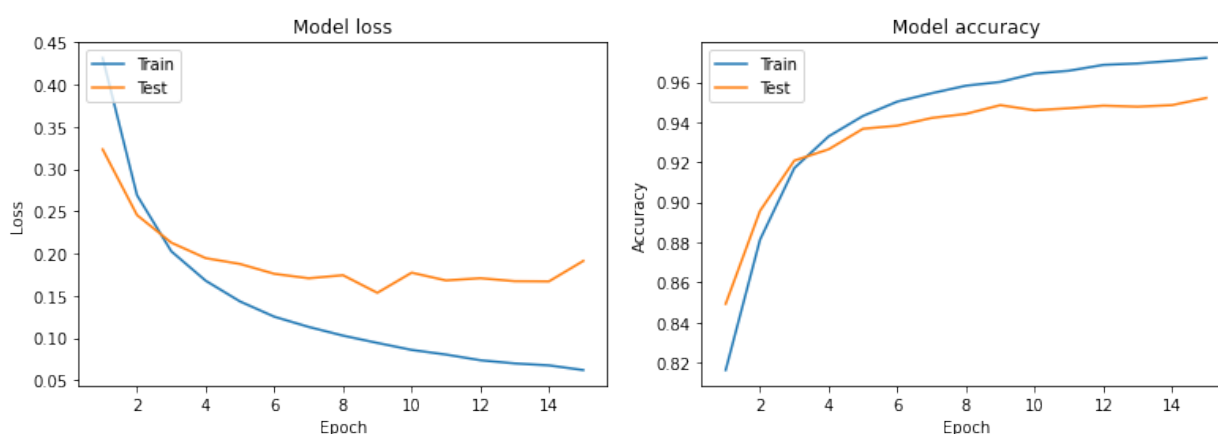
تنظیم و هر مجموعه داده، اعتبارسنجی مختص خود را دارد. در این پایان نامه با توجه به نوع مسئله و روش‌های به‌کارگرفته شده، از دو اعتبارسنجی متقابل ۱۰ برابری متفاوت به منظور تقسیم کردن داده‌ها به دو مجموعه ارزیابی و آموزش در نظر گرفته شد که عبارتند از:

حالت اول: اعتبارسنجی متقابل ۱۰ برابری بدون برهم‌کنش‌های ناشناخته

در این حالت ۹۰ درصد از برهم‌کنش‌های مثبت و منفی را به‌طور تصادفی^۴ انتخاب می‌کنیم. برای مجموعه‌ی ارزیابی نیز ۱۰ درصد باقی‌مانده از برهم‌کنش‌های مثبت و منفی در نظر می‌گیریم.

در روال اعتبارسنجی حالت اول مدل انتخاب شده و بعضی پارامترهای اساسی، همچون تعداد ایپوک، تعیین مقدار شد. شکل ۱۰.۴ روند آموزش مدل انتخابی را نشان می‌دهد. طبق انتظار صحت مدل روی داده‌های آموزش صعودی اکید است اما برای داده‌های اعتبارسنجی بعد از ایپوک ۵ فراز و فرودهایی دیده می‌شود.

در نمودار تابع خطا، تا پایان ایپوک ۵، با افزایش ایپوک‌ها، مقدار تابع خطای مدل روی داده‌های آموزش و روی داده‌های اعتبارسنجی کم می‌شود. بعد از ایپوک ۵ روند مدل روی داده‌های آموزشی ادامه می‌یابد اما روند بر روی داده‌های اعتبارسنجی برعکس می‌شود. به معنای دیگر بیش‌برازش رخ می‌دهد. لذا براساس نمودارها تعداد ایپوک مناسب در این مرحله ۵ در نظر گرفته شد.



شکل ۱۰.۴: موارد صحت و مقادیر تابع خطا برای شماره ایپوک‌های مختلف

در شکل راست صحت مدل روی داده‌های آموزش و اعتبارسنجی در طی ۱۵ ایپوک نشان داده می‌شود و در شکل چپ مقادیر تابع خطا در ایپوک‌های مختلف مشاهده می‌شود.

حالت دوم: اعتبارسنجی متقابل ۱۰ برابری همراه با برهم‌کنش‌های ناشناخته

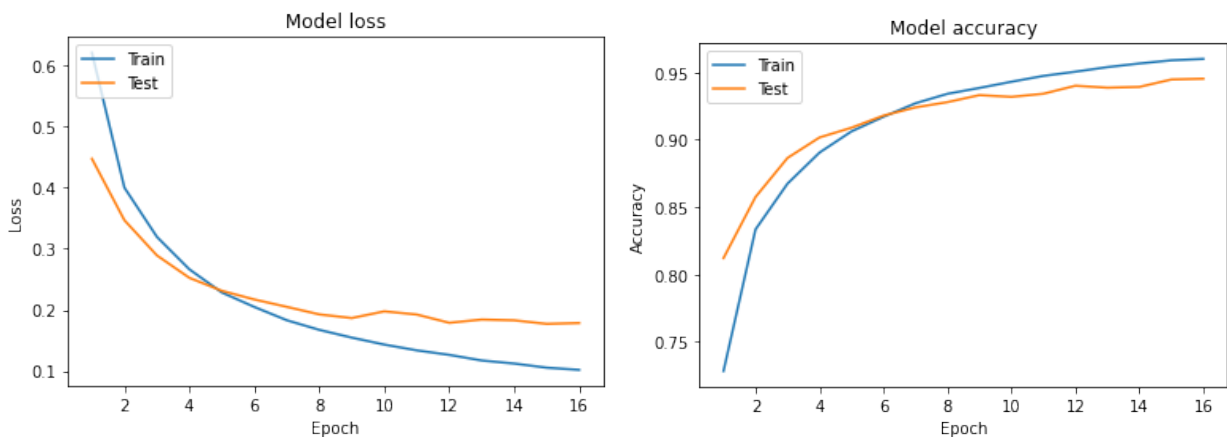
در این حالت مجموعه همه برهم‌کنش‌ها (مثبت، منفی، صفرهای مرحله اول) را به ۱۰ دسته مساوی تقسیم می‌کنیم. یک دسته را به عنوان مجموعه ارزیابی و ۹ دسته دیگر را به عنوان مجموعه داده‌ی آموزش

⁴Random

در نظر می‌گیریم. همه صفرهای مرحله‌ی قبل را هم به ۱۰ قسمت تقسیم کرده و به نسبت ۱ به ۹ هم به مجموعه‌ی ارزیابی و هم به مجموعه‌ی آموزش اضافه می‌کنیم.

در روال اعتبارسنجی حالت دوم مدل پیشین با کم‌ترین تغییرات برای پیش‌بینی سه‌کلاس آموزش داده شد. علاوه‌براین، پارامتر اساسی تعداد ایپوک، تعیین مقدار شد. شکل ۲۰۴ روند آموزش را نشان می‌دهد. روند صحت مدل روی داده‌های آموزش با افزایش ایپوک‌ها صعودی اکید است اما برای داده‌های اعتبارسنجی، مدل بعد از ایپوک ۹ مقدار صحت ثابت و کمی افول می‌کند.

در نمودار تابع خطا، تا پایان ایپوک ۹، با افزایش ایپوک‌ها، مقدار تابع خطای مدل روی داده‌های آموزش و روی داده‌های اعتبارسنجی کم می‌شود. بعد از ایپوک ۹ روند مدل روی داده‌های آموزشی ادامه می‌یابد اما روند بر روی داده‌های اعتبارسنجی گاهی بالا و گاهی پایین می‌رود. بدان معناست که احتمال بیش‌برازش بعد از ایپوک ۹ وجود دارد. لذا براساس نمودارها تعداد ایپوک مناسب برای مدل پیش‌بینی سه‌کلاس ۹ می‌باشد.



شکل ۲۰۴: نمودار صحت و مقادیر تابع خطا برای شماره ایپوک‌های مختلف

در شکل راست صحت مدل روی داده‌های آموزش و اعتبارسنجی در طی ۱۶ ایپوک نشان داده می‌شود و در شکل چپ مقادیر تابع خطا در ایپوک‌های مختلف مشاهده می‌شود.

۲۰۴ معرفی معیارهای سنجش

به منظور مقایسه عملکرد روش خود با سایر روش‌های موجود، از چهار معیار سنجش، $F_{measure}$ ، صحت^۵، مساحت زیر نمودار منحنی مشخصه‌ی عملکرد سیستم (AUC)^۶ و مساحت زیر نمودار بازخوانی-دقت (AUPR)^۷ استفاده شده است. برای تعریف این معیارها ابتدا چهار معیار شمارشی

^۵Accuracy

^۶Area Under Roc Curve

^۷Area Under Precision-Recall Curve

در این زمینه را در جدول ۱۰۴ معرفی کرده‌ایم.

پیش‌بینی برهم‌کنش مثبت	پیش‌بینی برهم‌کنش منفی	
مثبت درست	منفی نادرست	در واقعیت برهم‌کنش مثبت
مثبت نادرست	منفی درست	در واقعیت برهم‌کنش منفی

جدول ۱۰۴: حالت‌های ممکن نتایج یک یادگیری ماشین

با استفاده از آن‌ها چهار معیار سنجش به ترتیب زیر تعریف می‌شوند:

صحت: نسبت نتایج واقعی (مثبت درست و منفی درست) به تمام موارد مورد بررسی است.

$$\text{صحت} = \frac{\text{تعداد منفی‌های درست} + \text{تعداد مثبت‌های درست}}{\text{تعداد کل نمونه‌ها}}$$

همان‌طور که از رابطه بالا مشخص است، حاصل جمع تعداد مثبت‌های درست و تعداد منفی‌های درست، نشانگر تعداد نمونه‌هایی است که توسط سیستم به درستی تشخیص داده شده‌اند. مشکل استفاده از معیار صحت، این است که این معیار در زمانی که داده‌ها نامتوازن هستند، معیار مناسبی نیست. زیرا در این حالت، رده‌های را که در بین داده‌ها بیشترین آرا را دارد (رده اکثریت) را به تمام داده‌ها نسبت می‌دهد.

$F_{measure}$: یک میانگین هارمونیک^۸ میان پارامترهای فراخوانی و دقت می‌باشد و هدف اصلی پیشنهاد کردن این معیار می‌باشد. $F_{measure}$ بر اساس رابطه‌ی زیر محاسبه می‌شود.

$$F_{measure} = \frac{2 \times (\text{دقت} + \text{فراخوانی})}{\text{دقت} + \text{فراخوانی}}$$

هر کدام از معیارهای فراخوانی^۹ و دقت نیز طبق رابطه‌های زیر محاسبه می‌شوند.

$$\text{دقت} = \frac{\text{تعداد مثبت‌های درست}}{\text{تعداد مثبت‌های درست} + \text{تعداد مثبت‌های نادرست}}$$

$$\text{فراخوانی} = \frac{\text{تعداد مثبت‌های درست}}{\text{تعداد مثبت‌های نادرست} + \text{تعداد مثبت‌های درست}}$$

مساحت زیر نمودار بازخوانی-دقت: نمودار بازخوانی-دقت یک منحنی دو بعدی است که در آن دقت روی محور عرض‌ها و به‌طور مشابه فراخوانی روی محور طول‌ها رسم می‌شوند. به بیان دیگر منحنی

^۸Harmonic Average

^۹Recall

دقت- فراخوانی مصالحه نسبی میان دقت و فراخوانی در آستانه‌های متفاوت را نشان می‌دهد. سطح زیر این نمودار یکی از مهم‌ترین معیارهای ارزیابی مدل می‌باشد.

مساحت زیر نمودار منحنی مشخصه‌ی عملکرد سیستم: منحنی‌های مشخصه‌ی عملکرد سیستم، منحنی‌های دو بعدی هستند که در آن‌ها نرخ تشخیص صحیح دسته‌ی مثبت (TPR) ^{۱۰} روی محور عرض‌ها و به‌طور مشابه نرخ تشخیص غلط دسته‌ی منفی (FPR) ^{۱۱} روی محور طول‌ها رسم می‌شوند. به‌بیان دیگر یک منحنی مشخصه عملکرد سیستم مصالحه نسبی میان سودها و هزینه‌ها را نشان می‌دهد. فرمول‌های نرخ تشخیص غلط دسته‌ی منفی و نرخ تشخیص صحیح دسته‌ی مثبت که در زیر آورده شده‌است، ماهیت این عناصر را روشن می‌کنند.

$$\text{نرخ تشخیص صحیح دسته مثبت} = \frac{\text{تعداد مثبت های درست}}{\text{تعداد مثبت های درست} + \text{تعداد منفی های نادرست}}$$

$$\text{نرخ تشخیص صحیح دسته منفی} = \frac{\text{تعداد مثبت های نادرست}}{\text{تعداد مثبت های نادرست} + \text{تعداد منفی های درست}}$$

۳.۴ مقایسه‌ی نتایج

طی روال اعتبارسنجی مبسوط در بخش ۱.۴ مدل تشخیص نوع برهم‌کنش انتخاب و آموزش داده شد. سپس با تشخیص عدم برهم‌کنش‌های محتمل‌تر مدل سه‌کلاسه نهایی ارائه شد. در ادامه برای بررسی قابلیت اطمینان، استواری و کارایی، مدل SNF-CNN در روال اعتبارسنجی بر روی داده آزموده شد. نتایج SNF-CNN و سایر روش‌ها که برای مقایسه در نظر گرفته شده‌اند، در این بخش ارائه شده و بررسی می‌شوند. قبل از مقایسه‌ی روش‌های مختلف، نمونه‌ای از نتایج اجرای شبکه‌ی عصبی برای تشخیص نوع برهم‌کنش کاهنده و افزایشنده ارائه می‌شود.

جدول ۲.۴ مثالی از نتایج اجرای مدل است که توانایی مدل از جهت دقت، بازخوانی و F_{score} در تشخیص نوع برهم‌کنش‌ها نشان می‌دهد. براساس جدول ۲.۴ دقت مدل در تشخیص برهم‌کنش افزایشنده و کاهنده ۹۵ و ۹۴ درصد است درحالی که فراخوانی به ترتیب ۹۹ و ۸۳ درصد می‌باشد. مقدار $F_{measure}$ نیز ۹۷ و ۸۸ درصد می‌باشد که توانایی بالاتر مدل در تشخیص برهم‌کنش‌های افزایشنده از تعداد بالاتر این نوع برهم‌کنش‌ها می‌آید. نسبت برهم‌کنش افزایشنده به کاهنده تقریباً ۴ به ۱ است.

همچنین در جدول ۳.۴ گزارش دسته‌بندی برای حالت سه‌کلاسه نشان داده می‌شود. در این اجرا دقت مدل برای تشخیص برهم‌کنش افزایشنده، بدون برهم‌کنش و برهم‌کنش کاهنده به ترتیب ۹۵، ۹۶ و ۸۸ درصد است. فراخوانی به ترتیب ۹۷، ۹۵ و ۸۴ درصد است و در نهایت $F_{measure}$ ۹۶، ۹۶ و ۸۶ درصد می‌باشد.

¹⁰True Positive Rate

¹¹False Positive Rate

تعداد	F_{score}	فراخوانی	دقت	
۸۵۰	۸۸/۰	۸۳/۰	۹۴/۰	کاهنده
۳۰۵۲	۹۷/۰	۹۹/۰	۹۵/۰	افزاینده
۳۹۰۲	۹۵/۰			صحت
۳۹۰۲	۹۳/۰	۹۱/۰	۹۵/۰	Macro Avg
۳۹۰۲	۹۵/۰	۹۵/۰	۹۵/۰	Weighted Avg

جدول ۲.۴: گزارش دسته‌بندی نوع برهم‌کنش

تعداد	F_{score}	فراخوانی	دقت	
۸۵۰	۸۶/۰	۸۴/۰	۸۸/۰	کاهنده
۳۰۰۰	۹۶/۰	۹۵/۰	۹۶/۰	عدم برهم‌کنش
۳۰۵۲	۹۶/۰	۹۷/۰	۹۵/۰	افزاینده
۶۹۰۲	۹۵/۰			صحت
۶۹۰۲	۹۳/۰	۹۲/۰	۹۳/۰	Macro Avg
۶۹۰۲	۹۵/۰	۹۵/۰	۹۵/۰	Weighted Avg

جدول ۳.۴: گزارش دسته‌بندی حالت سه‌کلاسه برهم‌کنش

مشاهده شد، توان مدل در حالت سه‌کلاسه مقدار کمی نسبت به دوکلاسه کاهش می‌یابد که می‌تواند دو دلیل داشته باشد.

(۱) مسئله سه‌کلاسه از دوکلاسه سخت‌تر است.

(۲) صفرها یا عدم برهم‌کنش‌ها لزوماً واقعی نبوده و از نظر داروشناسی تایید شده نیستند پس احتمال وجود مقداری اختلال می‌رود.

لذا به دلایل فوق الذکر مقداری کاهش توان مدل در تشخیص سه‌کلاسه دور از انتظار نبود.

از آنجا که الگوریتم‌های پیشین در پیش‌بینی سه‌تایی برهم‌کنش دارو-دارو از AUC و AUPR بهره می‌برند، لذا نتایج الگوریتم ارائه شده براساس این دو معیار و برای تشخیص برهم‌کنش افزاینده، عدم برهم‌کنش و برهم‌کنش کاهنده مطابق جدول ۴.۴ می‌باشد. همچنین در جدول برای الگوریتم ارائه شده این تحقیق بازه‌ی بالا و پایین با اطمینان ۹۵ درصد گزارش شده‌است که نشان می‌دهد، نتایج الگوریتم در روال اعتبارسنجی ۱۰ تایی تغییرات کمی داشته و الگوریتم ارائه‌شده استوار بوده و قابل اعتماد می‌باشد.

در جدول ۵.۴ نتایج الگوریتم SNF-CNN برای سه حالت میانگین گرفته شده و در جدول با دیگر الگوریتم‌های سه‌کلاسه موجود مقایسه شده است. طبق جدول ۵.۴ الگوریتم ارائه شده اختلاف بالایی نسبت به دیگر الگوریتم‌های برتر مسئله‌ی سه‌تایی، داشته و توانسته الگوریتم‌های دیگر را به چالش بکشد.

$AUPR$	AUC	
0.9666 ± 0.0045	0.9747 ± 0.0033	افزاینده
0.8221 ± 0.0184	0.9686 ± 0.0028	کاهنده
0.9480 ± 0.0083	0.9714 ± 0.0040	عدم برهم‌کنش

جدول ۴.۴: نتایج الگوریتم SNF-CNN در پیش‌بینی سه‌کلاسه براساس معیارهای AUC و $AUPR$ و بازه‌ی اطمینان آن‌ها

$AUPR$	AUC	
۹۱۲/۰	۹۷۱/۰	$SNF-CNN$
۳۴۶/۰	۶۴۵/۰	$BRSNMF [71]$
۵۷۹/۰	۷۹۶/۰	$Semi-NMF [72]$
۵۲۶/۰	۸۴۲/۰	$TMFUF [68]$

جدول ۵.۴: مقایسه‌ی نتایج الگوریتم‌های پیش‌بینی سه‌کلاسه براساس معیارهای AUC و $AUPR$

۴.۴ نتیجه‌گیری و جمع‌بندی

در این رساله، با مدل‌سازی مسئله‌ی پیش‌بینی برهم‌کنش دارو-دارو و بهره‌گیری از رویکردهای مسئله‌ی سیستم‌های توصیه‌گر، روشی نوین با نتایج به مراتب بهتر نسبت به کارهای گذشته ارائه دادیم. با بررسی چهار معیار سنجش، صحت، مساحت زیر نمودار منحنی مشخصه‌ی عملکرد سیستم، مساحت زیر نمودار فراخوانی-دقت و $F_{measure}$ دیدیم که روش ما نتایج دقیق‌تری نسبت به سایر روش‌های موجود ارائه می‌کند. همچنین برای بررسی میزان قدرت روش SNF-CNN در شناسایی برهم‌کنش‌های ناشناخته، تعدادی از برهم‌کنش‌های جدید که مدل آن‌ها را پیش‌بینی کرده است را مورد تحقیق قرار داده‌ایم.

کارهای آینده و مسائل باز

رویکردهای محاسباتی موجود قادر به ارائه‌ی برهم‌کنش‌های بالقوه در مقیاس بزرگ قبل از استفاده از داروها در بازار هستند. با این حال نمی‌توانند برهم‌کنش‌های جامع، شامل برهم‌کنش‌های افزاینده و کاهنده را پیش‌بینی کنند. دانستن نوع برهم‌کنش افزاینده یا کاهنده یک جفت دارو، از این که فقط بدانیم یک جفت دارو برهم‌کنش دارند، مفیدتر است. اکثر رویکردهای موجود بدون در نظر گرفتن تغییرات دارویی ناشی از برهم‌کنش، فقط پیش‌بینی دودویی را گزارش می‌دهند. علاوه بر این، بروز برهم‌کنش‌های افزاینده و کاهنده تصادفی نیست، اما هیچ یک از رویکردهای موجود از این خاصیت ذاتی مهم برهم‌کنش‌ها در هنگام درمان بیماری‌های پیچیده (شامل درمان با سه یا چند دارو)، مورد بررسی و استفاده قرار نمی‌دهد.

در این پایان‌نامه پس از ارائه‌ی شبکه‌ی جامع برهم‌کنش‌ها، از ساختار سیستم‌های توصیه‌گر برای طراحی یک الگوریتم جدید استفاده کردیم. اگرچه پیش‌بینی به‌دست آمده توسط الگوریتم جدید الهام بخش است، اما عملکرد کلی هنوز هم می‌تواند بهبود یابد. به‌همین دلیل برهم‌کنش‌های نادرست پیش‌بینی شده را بررسی می‌کنیم. پس از بررسی آنها به‌صورت موردی، و در جهت اثبات عملی الگوریتم به بررسی عملکرد پیش‌بینی الگوریتم در آخرین نسخه‌ی پایگاه داده‌ی Drug Bank پرداختیم. مشاهدات و بررسی‌ها منجر به کشف دو علت برای پیش‌بینی‌های نادرست شد.

(۱) تعدادی جفت داروی مثبت کاذب که در نسخه‌ی چهار Drug Bank به‌طور دقیق به‌عنوان برهم‌کنش برچسب‌گذاری شده‌اند اما در نسخه فعلی به‌درستی به‌عنوان غیر برهم‌کنش شناخته می‌شوند. به‌عنوان مثال، در نسخه قدیمی Drug Bank ثبت شده است که Apraclonidine (داروی تقلید سمپاتیک^۱ مورد استفاده در درمان آب‌سیاه^۲) در استفاده هم‌زمان با Alprenolol و Bevantolol فعالیت‌های انسداد دهلیزی^۳ را افزایش می‌دهد، درحالی‌که نسخه جدید آن را حذف می‌کند.

(۲) جفت داروهای منفی کاذب که در نسخه‌ی چهار Drug Bank به اشتباه به‌عنوان غیر برهم‌کنش

¹Sympathomimetic

²Glaucoma

³Atrioventricular Blocking

برچسب‌گذاری شده‌اند اما در نسخه فعلی به‌عنوان برهم‌کنش شناخته می‌شوند.

به‌عنوان مثال، جفت داروی Valrubicin و Cyclosporine و همچنین جفت داروی Ergocalciferol و Calcitriol. نسخه جدید Drug Bank گزارش می‌کند، Valrubicin (داروی درمان سرطان مثانه) باعث افزایش فعالیت داروی نفروتوکسیک^۴ Cyclosporine (داروی سرکوب‌کننده سیستم ایمنی بدن با عمل خاص روی T-lymphocytes) می‌شود. درمان ترکیبی Calcitriol و Ergocalciferol ریسک یا شدت عوارض جانبی را در درمان چند دارویی افزایش می‌دهد.

همچنین مشاهدات بیشتر نشان داد بعضی برهم‌کنش‌های موجود در نسخه‌ی چهارم در نسخه پنج دیگر برهم‌کنش نیستند. مشاهدات ما را مقاله‌ی [۷۱] تایید می‌کند. پیش‌بینی می‌شود در آینده، مدل‌های پیش‌بینی موجود با جمع‌آوری مجموعه داده‌های بهتر و بیش‌تر بهبود یابند. مجموعه داده‌های جدیدتر تعداد کمتری از هر دو نوع جفت داروهای مثبت و منفی کاذب جمع‌آوری خواهند کرد. برای کارهای بعدی پیشنهاد می‌شود مجموعه داده همواره از آخرین نسخه‌ی Drug Bank جمع‌آوری شود.

داده‌ی سه‌تایی تلاشی برای بهبود در بیان و حل مسئله نسبت به حالت دودویی است. با این وجود داده‌های سه‌تایی نیز به اندازه‌ی کافی معناداری زیستی ندارد و اطلاعات زیستی محدودی ارائه می‌دهد. بدین معنی که پیش‌بینی نوع برهم‌کنش می‌تواند مفید باشد اما مشخص نمی‌شود که این برهم‌کنش در چه مرحله‌ای از مراحل فارماکوکنتیک یا فارماکودینامیک رخ داده است. لذا پیشنهاد می‌شود داده‌هایی با برچسب کاهنده و افزایشنده از هریک از مراحل فارماکوکنتیک و فارماکودینامیک جمع‌آوری شود. در این حالت مدل‌های معنی‌دارتری از نظر داروشناسی و هم‌یادگیری ماشین آموزش داده می‌شوند. مدل‌های حاصل برای داروشناس و داروساز اهمیت بالاتری داشته و قابل استفاده‌تر خواهد بود.

^۴Nephrotoxic

واژه‌نامه فارسی به انگلیسی

Drug Fingerprint	اثر انگشت دارو
Biological Understanding	ادراک بیولوژیکی
Bias	اریب
Potential	احتمالی
Case-Based Reasonong	استدلال نمونه محور
Cross Validation	اعتبارسنجی
Enhancive	افزاینده
Genetic Algorithm	الگوریتم ژنتیک
Drug Transporter	انتقال دهنده دارو
Atrioventricular Blocking	انسداد دهلیزی
Glaucoma	آب‌سیاه
Drug Enzyme	آنزیم دارو
Back Propagation	بازگشت به عقب
Distributed Representation	بازنمود توزیع شده
Potential	بالتقوه
Logistic Regression	برازش بخت
Feature Vector	بردار ویژگی
Food - Drug Interaction	برهم‌کنش غذا - دارو
Over Fiting	بیش‌برازش
Collaborative Refinement	پالایش مشارکتی
Web Database	پایگاه داده واقع بر گستره شبکه وب
Python	پایتون

Drug Binding Protein.....	پروتئین مرتبط به دارو.....
Loss Function.....	تابع خطا.....
Activation Function.....	تابع فعال‌ساز.....
Matrix Factorization.....	تجزیه ماتریس.....
None Negative Matrix Factorization.....	تجزیه ماتریس غیر منفی.....
Semi- None Negative Matrix Factorization.....	تجزیه ماتریس نیمه- غیر منفی.....
Similarity Network Fusion.....	ترکیب شباهت‌های شبکه‌ای.....
Visual Recognition.....	تشخیص تصویر.....
Speech Recognition.....	تشخیص صدا.....
Synergism.....	تشدیدکنندگی.....
Random.....	تصادفی.....
Structural Balance.....	تعادل ساختاری.....
Generalization.....	تعمیم‌پذیری.....
Feed Forward.....	تغذیه رو به جلو.....
Invariance.....	تغییرناپذیری.....
Fully Connected.....	تماما متصل.....
Theoretical Justification.....	توجیه نظری.....
Distribution.....	توزیع.....
Absorption.....	جذب.....
Random Forest.....	جنگل تصادفی.....
Off - Label.....	خارج از برچسب.....
Reconstruction Error.....	خطای نوسازی.....
Autoencoder.....	خودرمزنگار.....
Clustering.....	خوشه بندی.....
Over-Complete.....	خیلی کامل.....
Drug.....	دارو.....
Drug-Target.....	دارو- هدف.....
Pharmacy.....	داروسازی.....
Pharmacology.....	داروشناسی.....
Medical Pharmacology.....	داروشناسی پزشکی.....

New Drug	داروی جدید
Skin Rash	دانه‌های ریز پوستی
Design Tree	درخت تصمیم
Classifier	دسته‌بند
Elimination	دفع
Precision	دقت
Diazepam	دiazepam
Antagonism	رقابت‌کنندگی
Partial Least Square Regression	رگرسیون کمترین مربعات جزئی
Encoder	رمزگذار
Encode	رمزگذاری کردن
Decoder	رمزگشا
k-Fold Cross Validation (K-CV)	روش جامع اعتبارسنجی متقابل k برابری
Hybrid Approaches	روش‌های ترکیبی
Naïve Similarity-Based Approach	رویکرد مبتنی بر شباهت نیو
Classification-Based	رویکرد مبتنی بر طبقه‌بندی
Network Recommendation-Based	رویکرد مبتنی بر گراف توصیه‌گر
Real-time Training	زمان واقعی آموزش
Real-time Prediction	زمان واقعی پیش‌بینی
Chemical Structure	ساختار شیمیایی
Food and Drug Administration (FDA)	سازمان مواد غذایی و دارویی
Recommender System	سامانه توصیه‌گر
Similarity	شباهت
Jacard Similarity	شباهت جاکارد
Cosine Similarity	شباهت کوسینوس
Gaussian Similarity	شباهت گوسین
Artificial Neural Network	شبکه عصبی
Convolutional Neural Network (CNN)	شبکه عصبی کانولوشن
Multilayer Perceptron	شبکه‌ی پرسپترون چند لایه
Long Short-Term Memory (LSTM)	شبکه‌ی حافظه طولانی کوتاه مدت

Generative Stochastic Neural Network.....	شبکه‌ی عصبی مولد تصادفی
Cold Start.....	شروع سرد
Accuracy.....	صحت
Side Effect.....	عوارض جانبی
Label Side Effect.....	عوارض جانبی برچسب‌دار
Off-Label Side Effect.....	عوارض جانبی خارج از برچسب
Pharmacodynamic.....	فارماکودینامیک
Pharmacokinetic.....	فارماکوکینتیک
Fuzzy.....	فازی
Hyperparameter.....	فروپارامتر
Recall.....	فراخوانی
Collaborative Filtering.....	فیلترینگ همکارانه
Chain Rule.....	قانونه زنجیره‌ای
Robust.....	قدرت و استحکام
Convolution.....	کانولوشن
Degressive.....	کاهنده
Sparse Coding.....	کد نویسی پراکنده
Hierarchical Classification Codes.....	کدهای دسته‌بندی سلسله‌مراتبی
Stride.....	گام
Gradient.....	گرادیان
Bipartite Graph.....	گراف دو بخشی
Pooling Layer.....	لایه‌ی ادغام
Adjacency Matrix.....	ماتریس مجاورت
Support Vector Machine(SVM).....	ماشین بردار پشتیبان
Restricted Boltzmann Machines.....	ماشین بولتزمن محدود شده
Item-Based.....	مبتنی بر آیتم
Knowledge-Based.....	مبتنی بر دانش
User-Based.....	مبتنی بر کاربر
Content-Based.....	مبتنی بر محتوا
Constrain-Based.....	مبتنی بر محدودیت

Case-Based.....	مبتنی بر مورد
Metabolism.....	متابولیسم
Symmetric.....	متقارن
Small Training Set.....	مجموعه آموزشی کوچک
Area Under Precision-Recall Curve (AUPR).....	مساحت زیر نمودار بازخوانی-صحت
Area Under Roc Curve (AUC).....	مساحت زیر نمودار منحنی مشخصه‌ی عملکرد سیستم
Hard-Coded.....	مستقیم‌کدشده
Protein Pathways.....	مسیرهای پروتئین
Dose.....	مقدار
Sympathomimetic.....	مقلد سمپاتیک
Harmonic Average.....	میانگین هارمونیک
Similarity Measure.....	معیار شباهت
True Positive Rate (TPR).....	نرخ تشخیص صحیح دسته‌ی مثبت
False Positive Rate (FPR).....	نرخ تشخیص غلط دسته‌ی منفی
Lerning Rate.....	نرخ یادگیری
Indication.....	نشانگان دارو
Nephrotoxic.....	نفروتوکسیک
Neuron.....	نورون
Flatten.....	هموارکننده
Hidden Unit.....	واحد پنهان
Visible Unit.....	واحد قابل مشاهده
3-D Drug Chemical Structure.....	ویژگی‌های سه بعدی داروها
Unsupervised Learning.....	یادگیری بدون سرپرست
Deep Learning.....	یادگیری عمیق
Feature Learning.....	یادگیری ویژگی
Usenet.....	یوزنت

واژه‌نامه انگلیسی به فارسی

ویژگی‌های سه بعدی دارو	3-D Drug Chemical Structure
جذب	Absorption
صحت	Accuracy
تابع فعال‌ساز	Activation Function
ماتریس مجاورت	Adjacency Matrix
رقابت‌کنندگی	Antagonism
مساحت زیر نمودار بازخوانی- صحت	Area Under Precision-Recall Curve (AUPR)
مساحت زیر نمودار منحنی مشخصه‌ی عملکرد سیستم	Area Under Roc Curve (AUC)
شبکه عصبی	Artificial Neural Network
انسداد دهلیزی	Atrioventricular Blocking
خودرمزنگار	Autoencoder
بازگشت به عقب	Back Propagation
اریب	Bias
ادراک بیولوژیکی	Biological Understanding
گراف دو بخشی	Bipartite Graph
مبتنی بر مورد	Case-Based
استدلال نمونه محور	Case-Based Reasonong
قائده زنجیره‌ای	Chain Rule
ساختار شیمیایی	Chemical Structure
رویکرد مبتنی بر طبقه‌بندی	Classification-Based
دسته‌بند	Classifier
خوشه‌بندی	Clustering

شروع سرد	Cold Start
فیلترینگ همکارانه	Collaborative Filtering
پالایش مشارکتی	Collaborative Refinement
مبتنی بر محدودیت	Constrain-Based
مبتنی بر محتوا	Content-Based
کانولوشن	Convolution
شبکه عصبی کانولوشن	Convolutional Neural Network(CNN)
شباهت کوسینوس	Cosine Similarity
اعتبارسنجی	Cross Validation
رمزگشا	Decoder
یادگیری عمیق	Deep learning
کاهنده	Degrassive
درخت تصمیم	Design Tree
دiazepam	Diazepam
بازنمودهای توزیع شده	Distributed Representation
توزیع	Distribution
مقدار	Dose
دارو	Drug
انتقال دهنده دارو	Drug Transporter
پروتئین مرتبط به دارو	Drug Binding Protein
آنزیم دارو	Drug Enzyme
اثر انگشت دارو	Drug Fingerprint
دارو-هدف	Drug-Target
دفع	Elimination
رمزگذاری کردن	Encode
رمزگذار	Encoder
افزاینده	Enhanceive
نرخ تشخیص غلط دسته‌ی منفی	False Positive Rate (FPR)
یادگیری ویژگی	Feature Learning
بردار ویژگی	Feature Vector

تغذیه رو به جلو	Feed Forward
هموار کننده	Flatten
برهم‌کنش غذا - دارو	Food - Drug Interaction
سازمان مواد غذایی و دارویی	Food and Drug Administration (FDA)
تماما متصل	Fully Connected
فازی	Fuzzy
شباهت گوسین	Gaussian Similarity
تعمیم‌پذیری	Generalization
شبکه‌ی عصبی مولد تصادفی	Generative Stochastic Neural Network
الگوریتم ژنتیک	Genetic Algorithm
آب‌سیاه	Glaucoma
گرادیان	Gradient
مستقیم‌کد شده	Hard-Coded
میانگین هارمونیک	Harmonic Average
واحد پنهان	Hidden Unit
کد دسته‌بندی سلسله‌مراتبی	Hierarchical Classification Code
روش ترکیبی	Hybrid Approach
فرایپارامتر	Hyperparameter
نشانگان دارو	Indication
تغییرناپذیری	Invariance
مبتنی بر آیتم	Item-Based
شباهت جاکارد	Jacard Similarity
روش جامع اعتبارسنجی متقابل k برابری	k-Fold Cross Validation (K-CV)
مبتنی بر دانش	Knowledge-Based
عوارض جانبی برچسب‌دار	Label Side Effect
نرخ یادگیری	Learning Rate
برازش بخت	Logistic Regression
شبکه‌ی حافظه طولانی کوتاه مدت	Long Short-Term Memory (LSTM)
تابع خطا	Loss Function
تجزیه ماتریس	Matrix Factorization

داروشناسی پزشکی	Medical Pharmacology
متابولیسم	Metabolism
شبکه‌ی پرسپترون چند لایه	Multilayer Perceptron
رویکرد مبتنی بر شباهت نیو	Naïve Similarity-Based Approach
نفروتوکسیک	Nephrotoxic
رویکرد مبتنی بر گراف توصیه‌گر	Network Recommendation-Based
نورون	Neuron
داروی جدید	New Drug
تجزیه ماتریس غیر منفی	None Negative Matrix Factorization
خارج از برچسب	Off - Label
عوارض جانبی خارج از برچسب	Off-Label Side Effect
بیش‌برازش	Over Fiting
خیلی کامل	Over-Complete
رگرسیون کمترین مربعات جزئی	Partial Least Square Regression
فارماکودینامیک	Pharmacodynamic
فارماکوکینتیک	Pharmacokinetic
داروشناسی	Pharmacology
داروسازی	Pharmacy
لایه‌ی ادغام	Pooling Layer
بالقوه	Potential
دقت	Precision
مسیر پروتئین	Protein Pathway
پایتون	Python
تصادفی	Random
جنگل تصادفی	Random Forest
زمان واقعی پیش‌بینی	Real-time Prediction
زمان واقعی آموزش	Real-time Training
فراخوانی	Recall
سامانه توصیه‌گر	Recommender System
خطای نوسازی	Reconstruction Error

ماشین بولتزمن محدود شده	Restricted Boltzmann Machines
قدرت و استحکام	Robust
تجزیه ماتریس نیمه- غیر منفی	Semi- Non Negative Matrix Factorization
عوارض جانبی	Side Effect
شباهت	Similarity
معیار تشابه	Similarity Measure
ترکیب شباهت‌های شبکه‌ای	Similarity Network Fusion
دانه‌های ریز پوستی	Skin Rash
مجموعه آموزشی کوچک	Small Training Set
کد نویسی پراکنده	Sparse Coding
تشخیص صدا	Speech Recognition
گام	Stride
تعادل ساختاری	Structural Balance
ماشین بردار پشتیبان	Support Vector Machine(SVM)
متقارن	Symmetric
مقلد سمپاتیک	Sympathomimetic
تشدیدکنندگی	Synergism
توجیه نظری	Theoretical Justification
نرخ تشخیص صحیح دسته‌ی مثبت	True Positive Rate (TPR)
یادگیری بدون سرپرست	Unsupervised Learning
یوزنت	Usenet
مبتنی بر کاربر	User-Based
واحد قابل مشاهده	Visible Unit
تشخیص تصویر	Visual Recognition
پایگاه داده واقع بر گستره شبکه وب	Web Database

مراجع

- [1] BG Katzung, AJ Trevor. *Basic & Clinical Pharmacology 14th Eddition*. academia.edu, 2018:3.
- [2] Katzung, BG and Kruidering-Hall, M Katzung & Trevor's pharmacology: examination & board review. 12e *New York, NY: McGraw-Hill Education, 2019*.
- [3] Rescigno, Aldo. Foundations of pharmacokinetics. *Springer Science & Business Media, 2003*.
- [4] Kenakin, T. Pharmacologic analysis of drug±receptor interaction 3rd ed. *New York: Raven Press, 1997*.
- [5] Chou, Ting-Chao. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacological reviews, 58(3):621–681, 2006*.
- [6] Law, Vivian, Knox, Craig, Djoumbou, Yannick, Jewison, Tim, Guo, An Chi, Liu, Yifeng, Maciejewski, Adam, Arndt, David, Wilson, Michael, Neveu, Vanessa, et al. Drugbank 4.0: shedding new light on drug metabolism. *Nucleic acids research, 42(D1):D1091–D1097, 2013*.
- [7] Wishart, David S, Knox, Craig, Guo, An Chi, Cheng, Dean, Shrivastava, Savita, Tzur, Dan, Gautam, Bijaya, and Hassanali, Murtaza. Drugbank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research, 36(suppl_1):D901–D906, 2007*.
- [8] Knox, Craig, Law, Vivian, Jewison, Timothy, Liu, Philip, Ly, Son, Frolkis, Alex, Pon, Allison, Banco, Kelly, Mak, Christine, Neveu, Vanessa, et al. Drugbank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic acids research, 39(suppl_1):D1035–D1041, 2010*.
- [9] Wishart, David S, Knox, Craig, Guo, An Chi, Shrivastava, Savita, Hassanali, Murtaza, Stothard, Paul, Chang, Zhan, and Woolsey, Jennifer. Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research, 34(suppl_1):D668–D672, 2006*.

- [10] Ogata, Hiroyuki, Goto, Susumu, Sato, Kazushige, Fujibuchi, Wataru, Bono, Hidemasa, and Kanehisa, Minoru. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 27(1):29–34, 1999.
- [11] Kanehisa, Minoru and Goto, Susumu. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.
- [12] Tatonetti, N. P., Ye, P. P., Daneshjou, R. & Altman, R. B. Data-driven prediction of drug effects and interactions. *Sci Transl Med*, 4, 125ra31; doi: 10.1126/scitranslmed.3003377 (2012).
- [13] Y. Wang, J. Xiao, T.O. Suzek, J. Zhang, J. Wang, S.H. Bryant, PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res* 2009; 37.
- [14] Zhang P, Wang F, Hu J, Sorrentino R. Label propagation prediction of drugdrug interactions based on clinical side effects. *Sci Rep*. 2015;5:12339.
- [15] Kuhn, Michael, Letunic, Ivica, Jensen, Lars Juhl, and Bork, Peer. The sider database of drugs and side effects. *Nucleic acids research*, 44(D1):D1075–D1079, 2015.
- [16] Cheng, Feixiong and Zhao, Zhongming. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *Journal of the American Medical Informatics Association*, 21(e2):e278–e286, 2014.
- [17] Zhang, Wen, Chen, Yanlin, Liu, Feng, Luo, Fei, Tian, Gang, and Li, Xiaohong. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC bioinformatics*, 18(1):18, 2017.
- [18] Wienkers LC, Heath TG. Predicting in vivo drug interactions from in vitro drug discovery data. *Nat Rev Drug Discov*, 4(10):825–833, 2005.
- [19] Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res*, 42(Database issue):D1091–D1097, 2014
- [20] Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T, Hallisey R, Ives J, Laird N, Laffel G et al Systems analysis of adverse drug events. ADE Prevention Study Group. *Jama* 1995;274(1):35–43.
- [21] Businaro R. Why we need an efficient and careful pharmacovigilance. *J Pharmacovigil*, 2013;1:4.

- [22] Karbownik A, Szałek E, Sobańska K, Grabowski T, Wolc A, Grześkowiak E Pharmacokinetic drug–drug interaction between erlotinib and paracetamol: a potential risk for clinical practice. *Eur J Pharm Sci*, 2017;102:55–62.
- [23] Mulroy E, Highton J, Jordan S. Giant cell arteritis treatment failure resulting from probable steroid/antiepileptic drug–drug interaction. *N Z Med J*, 2017;130(1450):102–104.
- [24] Zhao XM, Iskar M, Zeller G, Kuhn M, van Noort V, Bork P. Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Comput Biol*. 2011;7(12):e1002323
- [25] Veith H, Southall N, Huang R, James T, Fayne D, Artemenko N, Shen M, Inglese J, Austin CP, Lloyd DG, et al. Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries. *Nat Biotechnol*. 2009;27(11):1050–5.
- [26] Huang SM, Temple R, Throckmorton DC, Lesko LJ. Drug interaction studies: study design, data analysis, and implications for dosing and labeling. *Clin Pharmacol Ther*. 2007;81(2):298–304.
- [27] Wiśniowska B, Polak S. The role of interaction model in simulation of drug interactions and QT prolongation. *Curr Pharm Anal*, 2008;4(1):1–19.
- [28] Zhou D, Bui K, Sostek M, Al-Huniti N. Simulation and prediction of the drug-drug interaction potential of Naloxegol by physiologically based pharmacokinetic modeling. *CPT Pharmacometrics Syst Pharmacol*, 2016;5(5):250–7.
- [29] Bui QC, Sloot PMA, van Mulligen EM, Kors JA. A novel feature-based approach to extract drug-drug interactions from biomedical text. *Bioinformatics*, 2014;30(23):3365–71.
- [30] Zhang Y, Wu HY, Xu J, Wang J, Soysal E, Li L, Xu H. Leveraging syntactic and semantic graph kernels to extract pharmacokinetic drug drug interactions from biomedical literature. *BMC Syst Biol*, 2016;10 Suppl 3:67.
- [31] Duke JD, Han X, Wang ZP, Subhadarshini A, Karnik SD, Li XC, Hall SD, Jin Y, Callaghan JT, Overhage MJ, et al. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, 2008;24(13):i232–40.
- [32] Vilar S, Uriarte E, Santana L, Lorberbaum T, Hripesak G, Friedman C, Tatonetti NP. Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nat Protoc*. 2014;9(9):2147–63.

- [33] Cheng F, Zhao Z. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *J Am Med Inform Assoc.* 2014;21(e2):e278–86.
- [34] Pahikkala T, Airola A, Pietilä S, et al. Toward more realistic drug–target interaction predictions. *Brief Bioinform.* 2015;16: 325–37.
- [35] Luo H, Zhang P, Huang H, Huang J, Kao E, Shi L, He L, Yang L. DDI-CPI, a server that predicts drug-drug interactions through implementing the chemicalprotein interactome. *Nucleic Acids Res.* 2014;42(Web Server issue):46–52.
- [36] Shi J-Y, Huang H, Li J-X, Lei P, Zhang Y-N, Yiu S-M. In: IWBBIO, editor. Predicting comprehensive drug-drug interactions for new drugs via triple matrix factorization. *Spain: Lecture Notes in Computer Science: Bioinformatics and Biomedical Engineering, Springer;* 2017. p. 108–17.
- [37] Koch-Weser J. Serum drug concentrations in clinical perspective. *Ther Drug Monit.* 1981;3(1):3–16
- [38] Cokol M, Kuru N, Bica E, Larkins-Ford J, Aldridge BB. Efficient measurement and factorization of high-order drug interactions in Mycobacterium tuberculosis. *Sci Adv.* 3(10):e1701881
- [39] F. Wang, C. Zhang. Label Propagation through Linear Neighborhoods, *IEEE Trans. Knowl. Data Eng.*, 20 (1) (2008) 55–67.
- [40] W. Zhang, Y.L. Chen, S.K. Tu, F. Liu, Q.L. Qu, Drug side effect prediction through linear neighborhoods and multiple data source integration, *Ieee Int. C Bioinform.*, (2016) 427–434.
- [41] W. Zhang, X. Yue, Y.L. Chen, W.R. Lin, B.L. Li, F. Liu, X.H. Li. Predicting drug disease associations based on the known association bipartite network, *Ieee Int. Conf. Bioinform. Biomed., (Bibm) 2017 (2017) 503–509.*
- [42] W. Zhang, Y. Chen, D. Li, Drug-target interaction prediction through label propagation with linear neighborhood information, *Molecules*, 22 (12) (2017) 2056.
- [43] W. Zhang, X. Yue, F. Liu, Y. Chen, S. Tu, X. Zhang, A unified frame of predicting side effects of drugs by using linear neighborhood similarity, *BMC Syst. Biol.*, 11 (6) (2017) 101.
- [44] W. Zhang, X. Yue, F. Huang, R. Liu, Y. Chen, C. Ruan, Predicting drug-disease associations and their therapeutic function based on the drug-disease association bipartite network, *Methods*, (2018).

- [45] W. Zhang, X. Yue, W. Lin, W. Wu, R. Liu, F. Huang, F. Liu, Predicting drug-disease associations by using similarity constrained matrix factorization, *BMC Bioinf*, 19 (1) (2018) 233.
- [46] Wang, B. et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat. methods*, 11, 333 (2014).
- [47] Ekstrand, Michael D., Riedl, John T., and Konstan, Joseph A. Collaborative filtering recommender systems. *Foundations and Trends® in Human-Computer Interaction*, 4(2):81–173, 2011.
- [48] Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. *Neurocomputing.*, 2016 Apr 26;187:27-48
- [49] G. Hinton, S. Osindero, Y. W. Teh. A fast learning algorithm for deep belief nets. *Neural Comput*, 18(7)(2006)1527–1554.
- [50] R. Salakhutdinov, G. E. Hinton. Deep boltzmann machines. *Proceedings of the AISTATS*, 2009.
- [51] J. Ngiam, Z. Chen, P. W. Koh, et al.. Learning deep energy models. *Proceedings of the ICML*, 2011.
- [52] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Proceedings of the NIPS*, 2012.
- [53] M. D. Zeiler, R. Fergus. Visualizing and understanding convolutional neural networks. *Proceedings of the ECCV*, 2014.
- [54] K. He, X. Zhang, S. Ren, et al.. Spatial pyramid pooling in deep convolutional networks for visual recognition. *Proceedings of the ECCV*, 2014.
- [55] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition. *Proceedings of the ICLR*, 2015.
- [56] C. Szegedy, W. Liu, Y. Jia, et al. Going deeper with convolutions. *Proceedings of the CVPR*, 2015.
- [57] C. Poultney, S. Chopra, Y. L. Cun. Efficient learning of sparse representations with an energy-based model. *Proceedings of the NIPS*, 2006.
- [58] P. Vincent, H. Larochelle, Y. Bengio, et al., Extracting and composing robust features with denoising auto encoders. *Proceedings of the ICML*, 2008.
- [59] S. Rifai, P. Vincent, X. Muller, et al., Contractive auto-encoders: explicit invariance during feature extraction. *Proceedings of the ICML*, 2011.

- [60] J. Yang, K. Yu, Y. Gong, et al., Linear spatial pyramid matching using sparse coding for image classification. *Proceedings of the CVPR, 2009*.
- [61] S. Gao, I. W. Tsang, L. T. Chia, et al., Local features are not lonely–Laplacian sparse coding for image classification. *Proceedings of the CVPR, 2010*.
- [62] K. Yu, T. Zhang, Y. Gong, Nonlinear learning using local coordinate coding. *Proceedings of the NIPS, 2009*.
- [63] X. Zhou, K. Yu, T. Zhang, et al., Image classification using super-vector coding of local image descriptors. *Proceedings of the ECCV, 2010*.
- [64] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks. *Proceedings of the NIPS, 2012*.
- [65] Hinton GE, Sejnowski TJ. Learning and relearning in Boltzmann machines. *Explorations in the microstructure of cognition, 1986 Jan 3;1(282-317):2*.
- [66] B.A. Olshausen, D.J. Field. Sparse coding with an over complete basis set: a strategy employed by V1? *Vis Res, 1997; 3311–3325*.
- [67] Ezzat, A.; Wu, M.; Li, X.-L.; Kwoh, C.-K. Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. *Brief. Bioinform*, 2018, bby002-bby002.
- [68] Shi J-Y, Huang H, Li J-X, Lei P, Zhang Y-N, Dong K, Yiu S-M. TMFUF: a triple matrix factorization-based unified framework for predicting comprehensive drug–drug interactions of new drugs. *BMC Bioinform*, 2018, 19(S14):411.
- [69] Yu H, Mao K-T, Shi J-Y, Huang H, Chen Z, Dong K, Yiu S-M. Predicting and understanding comprehensive drug–drug interactions via seminonnegative matrix factorization. *BMC Syst Biol*, 2018;12:14.
- [70] Davis JA. Clustering and structural balance in graphs. *Hum Relat*, 1967; 20(2):181–187.
- [71] Shi J. Mao K.Yu H. et al. Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization. *J Cheminform*, 2019;11: 28.
- [72] Yu H, Mao K-T, Shi J-Y, Huang H, Chen Z, Dong K, Yiu S-M. Predicting and understanding comprehensive drug–drug interactions via seminonnegative matrix factorization. *BMC Syst Biol*, 2018;12(s1):14.

-
- [73] SNFPy: <https://pypi.org/project/snfpypy/>, Accessed 31 July 2020.
- [74] Tensorflow 1.14.0: <https://pypi.org/project/tensorflow/1.14.0/>, Accessed 29 August 2020.
- [75] KERAS: <https://pypi.org/project/Keras/2.2.5/>, Accessed 29 August 2020.

Surname: Levian

Name: Bahareh

Title: Drug-Drug Interaction Prediction by Recommender System Approach

Supervisor: Prof. Changiz Eslahchi

Degree: Master of Science

Subject: Department of Computer Science

Field: Scientific Calculations

Shahid Beheshti University

Faculty of Mathematical Sciences

Date: 2020

Number of pages: 88

Keywords: Drug-Drug Interaction, Drug Similarity, Drug Similarity Integration, Feature Selection, Recommender System

Abstract

Drug-drug interactions (DDIs) always cause unexpected and even adverse drug reactions. It is important to identify DDIs before drugs are used in the market. However, preclinical identification of DDIs requires much money and time. Computational approaches have exhibited their abilities to predict potential DDIs on a large scale by utilizing premarket drug properties. Nevertheless, most of them only predict whether or not one drug interacts with another, but neglect their enhancive (positive) and depressive (negative) changes of pharmacological effects. Moreover, these comprehensive DDIs do not occur at random, and derived from the structural features of the graph of DDIs. Revealing such a relationship is very important, because it is able to help understand how DDIs occur. Both the prediction of comprehensive DDIs and the discovery of structural relationship among them play an important guidance when making a co-prescription.

In this work, treating a set of comprehensive DDIs as a signed network, we design a novel model (SNF-CNN) for the prediction of enhancive and depressive DDIs based on similarity network fusion and convolutional neural networks. SNF-CNN achieves the depressive DDI prediction ($AUC = 0.9747 \pm 0.0033$ and $AUPR = 0.9666 \pm 0.0045$), enhancive DDI prediction ($AUC = 0.9686 \pm 0.0028$ and $AUPR = 0.8221 \pm 0.0184$) and the Unknown DDI prediction ($AUC = 0.9714 \pm 0.0040$ and $AUPR = 0.9480 \pm 0.0083$). Compared with two

state-of-the-art approaches, SNF-CNN shows its superiority. Finally, this new approach is not only able to predict comprehensive DDI, but also predicts conventional DDI.

Contact: Dr. Changiz Eslahchi

The main references of the thesis are:

1. Shi J-Y, Huang H, Li J-X, Lei P, Zhang Y-N, Dong K, Yiu S-M (2018) TMFUF: a triple matrix factorization-based unified framework for predicting comprehensive drug–drug interactions of new drugs. BMC Bioinform 19(S14):411.
2. Yu H, Mao K-T, Shi J-Y, Huang H, Chen Z, Dong K, Yiu S-M. Predicting and understanding comprehensive drug–drug interactions via seminonnegative matrix factorization. BMC Syst Biol 2018;12:14.
3. Shi J, Mao K, Yu H, et al. Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization. J Cheminform 2019;11: 28.



Shahid Beheshti University
Faculty of Mathematical Sciences
Department of Computer Science

M. Sc. Thesis

Drug-Drug Interaction Prediction by Recommender System Approach

by

Bahareh Levian

Supervisor

Prof. Changiz Eslahchi

2020