



پیش‌بینی کاربردهای جدید برای داروهای قدیمی

رضا حسن‌زاده^۱، زهرا ابراهیم‌زاده^۲، چنگیز اصلاح‌چی^{۳*}

^۱دانشکده‌ی فناوری‌های نوین، دانشگاه محقق اردبیلی، اردبیل، ایران

^۲پژوهشکده علوم زیستی، پژوهشگاه دانش‌های بنیادی، تهران، ایران

^۳گروه علوم کامپیوتر، دانشکده علوم ریاضی، دانشگاه شهید بهشتی، تهران، ایران

چکیده در سال‌های اخیر با وجود پیشرفت قابل توجه علم داروسازی، داروهای طراحی شده جایگاه خود در پزشکی و بازار را از دست می‌دهند و مهم‌ترین علت آن، عوارضی است که بعد از مصرف دارو نمایان می‌شود. همچنین حدود ۹۰ درصد داروها در مرحله اول کارآزمایی بالینی با شکست مواجه می‌شوند و شرکت‌های تولیدی را متحمل خسارت‌های فراوانی می‌کنند. هزینه‌ی تولید داروهای جدید میلیون‌ها دلار است و برای این‌که یک داروی جدید جایگاه خود را در بازار پزشکی پیدا کند حدود ۱۵ سال زمان نیاز دارد. به همین دلیل، دانشمندان به دنبال پیدا کردن کاربردهای جدید برای داروهای هستند که قبلاً امتحان خود را هم از نظر عوارض و هم از نظر بازار دارو پس دادند. مهم‌ترین مرحله‌ی این کار، پیش‌بینی تعامل بین داروها و هدف‌های دارویی یا پروتئین‌ها است. تاکنون چندین الگوریتم برای پیش‌بینی تعامل بین داروها و پروتئین‌ها ارائه شده‌اند. در این مقاله الگوریتمی جدید برای این مسأله ارائه می‌کنیم و برای بررسی کارایی و دقت آن، از نتایج بهترین الگوریتم موجود در این زمینه یعنی الگوریتم *DT - Hybrid* استفاده می‌کنیم. هر دو الگوریتم روی داده‌های پایگاه اطلاعاتی *DrugBank* اجرا می‌شوند و نشان می‌دهیم که الگوریتم پیشنهادی ما در مقایسه با الگوریتم *DT - Hybrid* نتایج بهتری تولید می‌کند.

واژه‌های کلیدی: دارو، پروتئین، پیشگویی، سیستم پیشنهادی و گراف دوپوششی

۱ مقدمه

زمان و هزینه دو عامل مهم در صنعت داروسازی است. به طور میانگین، حدود پانزده سال زمان و هشتصد میلیون دلار هزینه نیاز است تا یک دارو بتواند در این سالها جایگاه خود را در داروخانهها پیدا کند (آدامز و براکتر ۲۰۰۶). به علاوه حدود ۹۰ درصد داروها، در همان مراحل نخست کارآزمایی بالینی با شکست مواجه می‌شوند (کرافتز ۱۹۹۸). مهم‌ترین دلایل شکست خوردن یک دارو، ناکارآمدی و عوارض آن هستند. همین این عوامل دست به دست هم می‌دهند تا دانشمندان به دنبال راهی جدید برای پیدا کردن داروهای مناسب برای بیماران باشند. اخیراً دیدگاهی که مورد توجه قرار گرفته است، پیدا کردن کاربردهای جدید برای داروهای است که قبلاً طراحی شده‌اند. یکی از دلایل اهمیت این دیدگاه، بحث هزینه و زمان است. چون این داروها قبلاً طراحی شده‌اند لذا زمان و هزینه‌های مربوط به طراحی خود به خود حذف می‌شوند. به علاوه اکثر عوارض این داروها تاکنون شناسایی شده‌اند و این اطلاعات کمک شایانی به داروساز می‌کنند. اخیراً شرکت‌های بزرگ داروسازی دنیا نظیر شرکت‌های فایزر آمریکا، نوآرتیز سوئیس و بایر آلمان بخش‌های تحقیقاتی برای این منظور راه‌اندازی کرده‌اند و آینده علم داروسازی را در این دیدگاه جدید می‌دانند. اولین و مهم‌ترین مرحله از دیدگاه فوق، شناسایی یا به طور دقیق‌تر پیش‌بینی شامل‌های مفیدی بین داروها و هدف داروها (پروتئین‌ها) است. روش‌های تجربی برای این کار هم هزینه‌بر و هم زمان‌بر هستند و فقط زمانی می‌توانند مورد استفاده قرار بگیرند که تعداد هدف‌ها و داروها محدود باشند. ولی در مقاله، روش‌های محاسباتی توانایی کار با حجم عظیمی از داروها و هدف‌ها را دارند. این کار معمولاً با استفاده از اطلاعات مربوط به شامل‌هایی که تاکنون شناسایی شده‌اند انجام می‌پذیرد. پایگاه داده‌هایی وجود دارند که این اطلاعات را شامل می‌شوند. مشهورترین بانک اطلاعاتی در این زمینه DrugBank است (ویسارت و ناگس ۲۰۰۶). این پایگاه داده اطلاعات مفصلی در رابطه با داروهای که تاکنون طراحی شده‌اند، ساختار شیمیایی آنها، شامل‌هایی که با هدف‌های دارویی دارند و ساختار توالی هدف‌های دارویی را شامل می‌شود و همین روش‌هایی که برای پیش‌بینی شامل بین دارو و هدف ارائه می‌شوند از این بانک اطلاعاتی برای ارزیابی روش خود استفاده می‌کنند.

تاکنون چندین الگوریتم برای پیش‌بینی شامل بین دارو و هدف معرفی شده‌اند که اکثر آنها مبتنی بر روش‌های آماری، یادگیری ماشین، روش‌های سادرات دیفرانسیل و غیره هستند (یاساتسکی و ازاکی ۲۰۰۸)، هی و ژانگ (۲۰۱۰)، پرلمن و گاتلیب (۲۰۱۱)، فن و لیو (۲۰۱۲)، می و وو (۲۰۱۳). در این میان الگوریتمی به نام $DT-Hybrid$ آلیسو و پلویرتی (۲۰۱۳) وجود دارد که در مقایسه با همین روش‌ها عملکرد قابل توجهی دارد. این الگوریتم مبتنی بر سامانه‌های توصیه‌گر بهبود یافته الگوریتم NBI است (زورن ۲۰۰۷). سامانه‌های توصیه‌گر بیشتر در تجارت الکترونیک کاربرد دارند و برای فائق آمدن به مشکلات ناشی از حجم زیاد و رو به رشد اطلاعات ارائه شده‌اند. این سامانه‌ها، به کاربر با تحلیل رفتارهای پیشین کالا را پیشنهاد می‌دهد و کمک می‌کند تا در میان حجم عظیم اطلاعات سریع‌تر به هدف خود نزدیک شود. الگوریتم NBI داده‌ها را به یک گراف دو بخشی تبدیل کرده و با استفاده از دو مرحله تکنیک تخصیص منابع به هر کاربر نیستی از پیشنهادها ارائه می‌دهد که این لیست از نظر امتیاز به صورت نزولی مرتب شده است. الگوریتم $DT-Hybrid$ با اضافه کردن تشابهات ساختاری بین داروها و تشابهات توالی بین پروتئین‌ها به گراف دو بخشی فوق، برای هر دارو نیستی نزولی از پروتئین‌هایی پیشنهاد می‌دهد که قبلاً شامل بین دارو و این پروتئین‌ها شناخته نشده است (برای جزئیات بیشتر مراجعه شود به آلیسو و پلویرتی ۲۰۱۳). ما در اینجا الگوریتم جدیدی ارائه می‌دهیم که مشابه $DT-Hybrid$ برای هر دارو نیستی از پیشنهادها ارائه می‌دهد. برای ارزیابی کارایی الگوریتم پیشنهادی، نتایج آنرا پس از اعمال روی داده‌های DrugBank با نتایج الگوریتم $DT-Hybrid$ مقایسه خواهیم کرد. دلیل اول که الگوریتم پیشنهادی را تنها با $DT-Hybrid$ مقایسه می‌کنیم صرفاً کمبود فضا و دلیل دوم این است که قبلاً همین الگوریتم‌های مشهور موجود را با $DT-Hybrid$ مقایسه کردیم و مشاهده شد است که $DT-Hybrid$ نسبت به بقیه الگوریتم‌ها عملکرد بهتری دارد.

در بخش بعدی الگوریتم پیشنهادی معرفی خواهد شد و سپس در بخش آخر، نتایج این الگوریتم و الگوریتم $DT-Hybrid$ روی داده‌های

DrugBank مقایسه خواهد شد. برای تحلیل نتایج دو الگوریتم ابتدا از معیارهای دقت و یادآوری معرفی شده در *آلایو و پاورتی* (۲۰۱۳) استفاده خواهیم کرد. با یک مثال نشان خواهیم داد که این معیارها به تنهایی قادر به نشان دادن تفاوت بین دو لیست پیشنهادی مختلف نیستند. برای حل این مشکل، دو معیار جدید معرفی خواهیم کرد و سپس نتایج هر دو الگوریتم را با همی این چهار معیار بررسی خواهیم کرد. نشان خواهیم داد که نتایج الگوریتم پیشنهادی در مقایسه با الگوریتم *DT - Hybrid* به مراتب بهتر است.

۲ الگوریتم پیشنهادی

فرض کنید $D = \{d_1, d_2, \dots, d_m\}$ و $T = \{p_1, p_2, \dots, p_n\}$ به ترتیب مجموعه‌ی داروها و هدف داروها (پروتئین‌ها) باشند. شبکمی شامل بی داروها و پروتئین‌ها توسط یک گراف دوپیکشی مانند $G(D, T, E)$ نمایش داده می‌شود که در آن $E \subseteq \{p_i, d_j\}$ اگر داری p_i با پروتئین d_j تعامل داشته باشد. بنابراین ماتریس مجاورت گراف مورد نظر به صورت $A = \{a_{ij}\}_{m \times n}$ است که در آن $a_{ij} = 1$ اگر داری d_j و پروتئین p_i به هم وصل باشند و در غیر این صورت $a_{ij} = 0$. برای هر داری d_j پروتئین‌هایی را که d_j با آنها در تعامل است یا مجموعه $I(d_j)$ نمایش می‌دهیم. هدف ما ارائه‌ی الگوریتمی است که به داری d_j پروتئین‌هایی از مجموعه‌ی $T \setminus I(d_j)$ پیشنهاد دهد. راهبرد الگوریتم به این صورت است که داروهای مشابه، هدف‌های مشابهی داشته باشند. الگوریتم از دو معیار شباهت بین داروها و محبوبیت پروتئین‌ها بین داروها استفاده می‌کند تا بهترین پیشنهادها را برای داروها ارائه دهد. قبل از تعریف این دو معیار نیاز به چند تعریف و سادگناری داریم.

فرض کنید $S^d = \{s_{ij}^d\}_{n \times n}$ که در آن s_{ij}^d شباهت توالی میان پروتئین‌های p_i و p_j است و روش *اسمیت-اسمیت* و *واترمن* (۱۹۸۱) به دست آمده است. از این ماتریس برای معرفی جدید از شباهت بین داروها استفاده خواهیم کرد. ماتریس شباهت ساختاری بین داروها را با $S^o = \{s_{ij}^o\}_{m \times m}$ نشان می‌دهیم که s_{ij}^o شباهت ساختاری میان داروها d_i و d_j است و از روش *SIMCOMP* به دست آمده است *هاتوری و اکتو* (۲۰۰۳). معیار شباهتی که ما برای داروهای d_i و d_j تعریف می‌کنیم به صورت زیر است:

$$DS(d_i, d_j) = s_{ij}^o \left(\sum_{p \in I(d_i)} s_{ip}^d + \sum_{q \in I(d_j)} s_{jq}^d \right).$$

این معیار علاوه بر شباهت ساختاری بین داروها، اطلاعاتی را که در گراف دوپیکشی برای داروها موجود است در بر می‌گیرد. در واقع داروهای مشابه در نظر می‌گیرند که هم ساختار مشابهی داشته باشند و هم با پروتئین‌های مشابهی در تعامل باشند. فرض کنید

$$F_i = \{d_k \in D : I(d_i) \cap I(d_k) \neq \emptyset\},$$

$$D(j) = \{d_k \in D : p_j \in I(d_k)\},$$

$$F_i(j) = \{d_k \in F_i : p_j \in I(d_k)\}.$$

حال معیار محبوبیت پروتئین p_j برای داری d_i را به صورت زیر تعریف می‌کنیم:

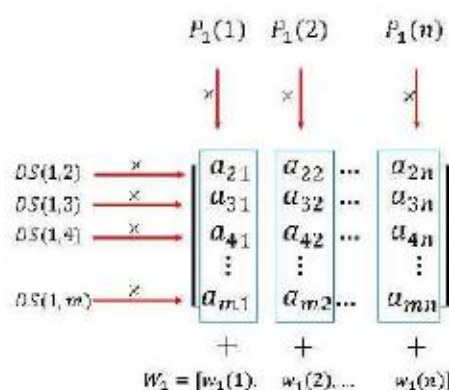
$$P_i(j) = \frac{|F_i(j)|/|F_i|}{|D(j)|/|D|}.$$

در واقع، این عدد با بررسی داروهای هم‌هدف با d_i و مقایسه‌ی آنها با کل داروها، میزان محبوبیت پروتئین p_j برای پیشنهاد شدن به داری d_i را نشان می‌دهد.

حالت فرض کنید می‌خواهیم برای d_k لیستی از پروتئین‌های عضو $T \setminus d_k$ پیشنهاد کنیم. ابتدا سطر مربوط به این دارو را از ماتریس A حذف می‌کنیم. سپس به هر سطر i از این ماتریس، $DS(i, j)$ و به هر ستون k مقدار $P_i(k)$ را ضرب می‌کنیم. ماتریس به دست آمده را ماتریس امتیاز داروی d_k می‌نامیم. در این ماتریس جمع هر ستون، امتیاز پروتئین متناظر با این ستون برای پیشنهاد شدن به داروی d_k است. پس از جمع کردن تمامی ستون‌ها، برداری به صورت $W_i = [w_i(j)]_{1 \times n}$ حاصل می‌شود که در آن

$$w_i(j) = \sum_{\substack{k \in \{1, \dots, m\} \\ k \neq d_k}} (P_i(j) \cdot a_{ik} \cdot DS(i, k)).$$

پس از مرتب کردن بردار W_i به صورت نزولی و حذف پروتئین‌های عضو d_k لیست پیشنهادی برای داروی d_k به دست می‌آید. روند ضرب دو معیار تعریف شده در سطر و ستون ماتریس سجاورته به دست آوردن ماتریس امتیاز برای داروی d_k در شکل ۱ نشان داده شده است.



شکل ۱. نحوه‌ی به دست آوردن ماتریس امتیاز بر داروی d_k

بحث و نتیجه‌گیری

۳ نتایج

۱.۳ معرفی داده‌ها

برای ارزیابی الگوریتم، از بانک اطلاعاتی DrugBank استفاده می‌کنیم. **ویهارت و ناگس (۲۰۰۶)**. این پایگاه داده، چهار نوع اصلی هدف‌های دارویی را دسته‌بندی کرده و داروهای را هم‌اکنون یا هر دسته به صورت تجربی ثابت شده است مشخص کرده است. اطلاعات مربوط به DrugBank در جدول ۱ آمده است.

۲.۳ مقایسه

دقت و کارایی الگوریتم را با مقایسه‌ی نتایج الگوریتم ما و الگوریتم $DT - Hybrid$ **آلیمر و پاورتی (۲۰۱۳)** روی داده‌های بانک اطلاعاتی DrugBank ارزیابی می‌کنیم. برای این کار، یال‌هایی را از گراف متناظر با هر مجموعه‌ی داده از DrugBank حذف می‌کنیم و پس از آن

جدول ۱: مشخصات مربوط به بانک اطلاعاتی DrugBank

	Proteins	Drugs
Nuclear Receptors	25	53
GPCRs	94	222
Ion Channels	203	209
Enzymes	663	444
Complete DrugBank	1481	1663

بررسی می‌کنیم که کدام یک از الگوریتم‌ها در بازسازی یال‌های حذف شده بهتر عمل می‌کنند. حذف هر یال معادل با حذف پروتئین هدف یکی از داروها و بازسازی آن معادل با پیشنهاد آن پروتئین به دارویی مورد نظر است. برای حذف یال‌ها از دو روش استفاده خواهیم کرد. در روش اول که به *fold cross validation* - ۱۰ مشهور است ابتدا مجموعه داده را به سه قسمت تقسیم می‌کنیم. سپس در هر مرحله یکی از قسمت‌ها را به عنوان مجموعه‌ی اعتبار سنجی یا آزمایش و نه قسمت باقیمانده را به عنوان مجموعه‌ی آموزشی در نظر می‌گیریم. این عمل را ۳۰ بار انجام می‌دهیم. در روش دوم که عیب *jackknife validation* است ابتدا از هر دارو یک پروتئین را حذف می‌کنیم به شرطی که اطلاعات مربوط به هیچ دارویی از بین نرود، یعنی در نهایت هر دارویی حداقل یک پروتئین داشته باشد. سپس بررسی می‌کنیم که کدام یک از الگوریتم‌ها در بازسازی یال‌های حذف شده بهتر عمل می‌کنند. پس از آن، از هر دارو دو یال حذف می‌کنیم (اگر امکانش وجود داشته باشد) و الگوریتم‌ها را اعمال می‌کنیم و این عمل حذف را تا هر تعدادی که می‌توانیم ادامه می‌دهیم. برای ارزیابی الگوریتم‌ها در هر دو روش فوق از چند معیار استفاده می‌کنیم. قول از معرفی این معیارها، نیاز به چند نمادگذاری داریم که در جدول زیر آمده‌اند. اولین معیارهایی که برای ارزیابی الگوریتم‌ها استفاده می‌کنیم معیارهای یادآوری

جدول ۲: نمادگذاری

$DT(d_i) = \{T_1^d, T_2^d, \dots, T_n^d\}$	پروتئین‌های حذف شده از طریق d_i
$L(d_i) = \{T_1, T_2, \dots, T_n\}$	لیست پیشنهادی برای طریق d_i
t	طول لیست پیشنهادی
$L_t(d_i) = \{T_1, T_2, \dots, T_t\}$	لیست پیشنهادی برای طریق d_i با طول t
M	مجموعه آزمایش (مجموعه داروهایی که حداقل یک پروتئین حذف شده دارند)

R_t و P_t هستند که در مقالهی *آلیسو و پاورتی (۲۰۱۳)* هم از آنها استفاده شده است. این معیارها به صورت زیر تعریف می‌شوند.

$$R_t = \frac{\sum_{d_i \in M} R_t(d_i)}{|M|} \quad P_t = \frac{\sum_{d_i \in M} P_t(d_i)}{|M|}$$

که در آنها

$$R_t(d_i) = \frac{|DT(d_i) \cap L_t(d_i)|}{|DT(d_i)|}, \quad P_t(d_i) = \frac{|DT(d_i) \cap L_t(d_i)|}{t}$$

معیار دقت بیان می‌کند که از پروتئین‌های پیشنهاد شده در لیست به طول t چند پروتئین به درستی پیشنهاد شده‌اند و معیار یادآوری بیان می‌کند که چند درصد از پروتئین‌های حذف شده در لیست اصلی به درستی برگردانده شده‌اند. علی‌رغم شناخته شده بودن این دو معیار، اما ترتیب پیشنهاد دادن پروتئین‌ها در آنها لحاظ نمی‌شود. به عنوان مثال، اگر از دارویی t پروتئین‌های $\{p_1, p_2, p_3\}$ حذف شده باشند و دو الگوریتم به ترتیب لیست‌های $\{p_1, p_2, p_3\}$ و $\{p_3, p_1, p_2\}$ را پیشنهاد داده باشند در آن صورت هم معیار دقت و هم معیار یادآوری برای هر دو الگوریتم برابر

با ۰.۶۶ خواهد بود در صورتی که الگوریتم اول دو پروتئین صحیح را در مکان‌های اول و دوم پیشنهاد داده است و باید دقت بهتری در مقایسه با الگوریتم دوم داشته باشد. برای نشان دادن این تفاوت ما معیارهای بهبود یافته دقت و یادآوری را با اختصاص دادن وزن به هر مکان از لیست پیشنهادی معرفی می‌کنیم. طبقاً مکان اول بالاترین وزن یعنی t را دارد، مکان دوم $t-1$ و به همین ترتیب مکان آخر وزن یک خواهد داشت. فرض کنید $\{T_{i_1}, T_{i_2}, \dots, T_{i_t}\} = DT(d_i) \cap L(d_i)$ که $1 \leq i_1 \leq i_2 \leq \dots \leq i_t \leq k$. معیارهای جدید R'_t را به صورت زیر تعریف می‌کنیم:

$$R'_t = \frac{\sum_{d_i \in M} R'_t(d_i)}{|M|} \quad P'_t = \frac{\sum_{d_i \in M} P'_t(d_i)}{|M|},$$

که در آن‌ها

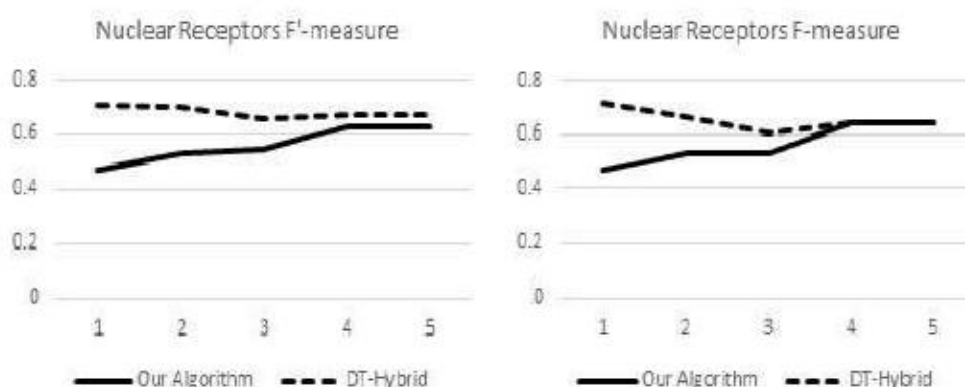
$$R'_t(d_i) = \frac{|DT(d_i) \cap L(d_i)|}{\min(|DT(d_i)|, t)},$$

$$P'_t(d_i) = \frac{(t+1-i_1) + (t+1-i_2) + \dots + (t+1-i_t)}{t + (t-1) + \dots + 1}.$$

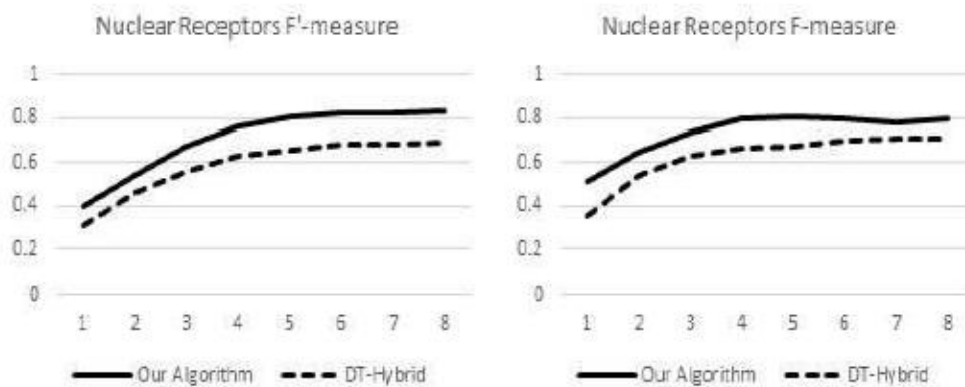
با این تعاریف، معیار دقت برای لیست‌های پیشنهادی دو الگوریتم ذکر شده در مثال قبلی به ترتیب برابر با ۰.۸۲ و ۰.۵ خواهد بود و به وضوح تفاوت لیست‌های پیشنهادی با معیار جدید نشان داده می‌شود در صورتی که معیارهای معمولی قادر به نشان دادن این تفاوت نبودند. دو معیار دیگری که از آن‌ها استفاده خواهیم کرد میانگین توافقی معیارهای دقت و یادآوری معمولی (F -measure) و میانگین توافقی معیارهای دقت و یادآوری بهبود یافته (F' -measure) هستند. میانگین توافقی برای دو پارامتر x و y به صورت زیر تعریف می‌شود:

$$F\text{-measure} = \frac{2xy}{x+y}.$$

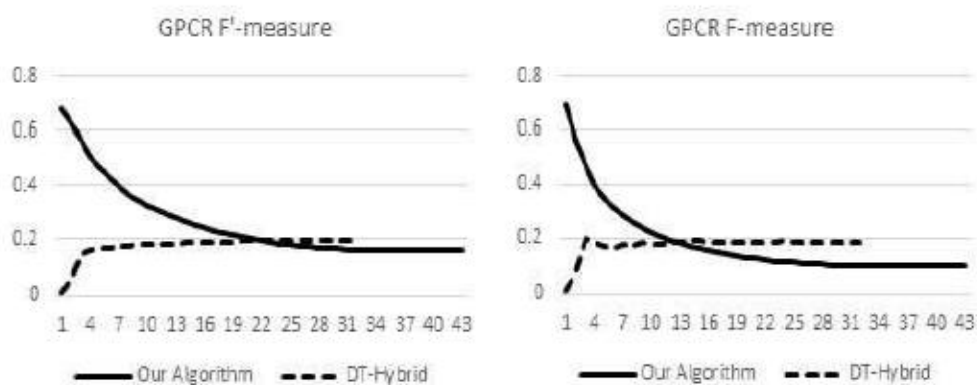
برای به دست آوردن نتایج و مقایسه آن‌ها، هر دو الگوریتم را بر روی داده‌های جدول ۱ و به هر دو روش 10-fold و $Jackknife\text{-like}$ اجرا کردیم و نتایج به دست آمده در شکل‌های ۲ تا ۱۱ آمده است. به دلیل زیاد بودن شکل‌ها و این‌که میانگین توافقی هر دو معیار دقت و یادآوری را با هم نشان می‌دهد در اینجا فقط نتایج مربوط به معیارهای F -measure و F' -measure نشان داده شده است. برای روش 10-fold فقط در دلیلی *Nuclear Receptors* معیارهای دقت، دقت بهبود یافته، یادآوری و یادآوری بهبود یافته برای الگوریتم *DT-Hybrid* بهتر از الگوریتم ما بود و در بقیه داده‌ها نتایج الگوریتم ما بهتر بود. در روش *Jackknife-like* در تمامی داده‌ها همه معیارها برای الگوریتم ما در مقایسه با الگوریتم *DT-Hybrid* بالاتر بود و این نشان می‌دهد که الگوریتم ما پیشنهادهای بهتری برای طوره‌ها ارائه می‌کند.



شکل ۲. نتایج مربوط به روش 10-fold برای دلیلی *Nuclear Receptors*.



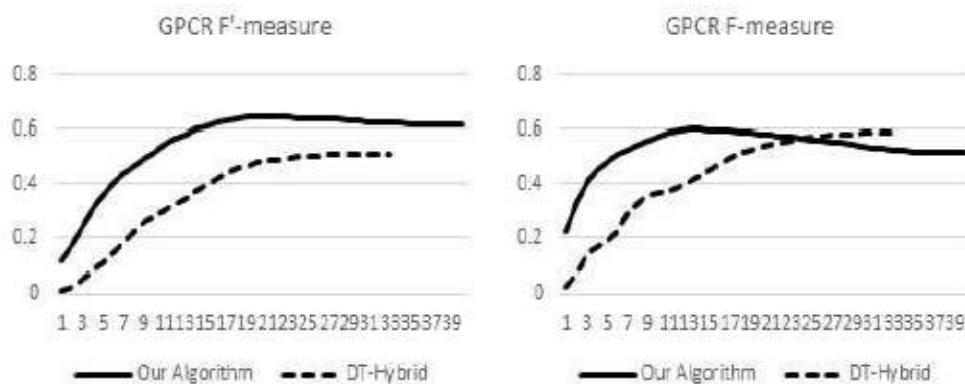
شکل ۱۳: نتایج مربوط به روش *Jackknife - hite* برای دانه‌ی *Nuclear Receptors*.



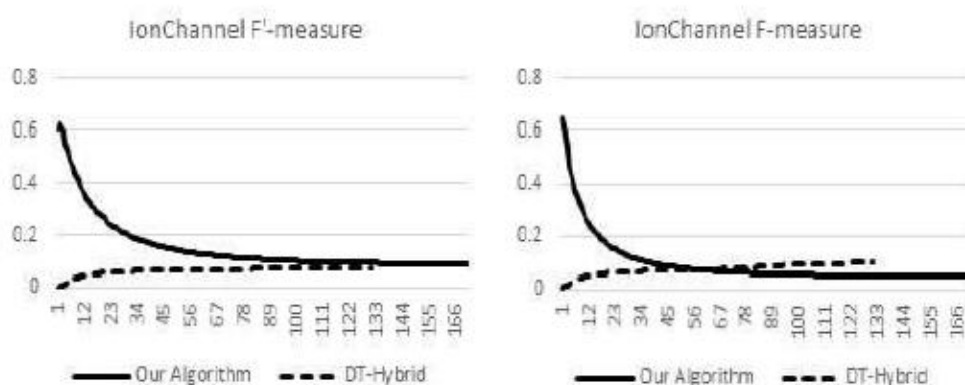
شکل ۱۴: نتایج مربوط به روش *fold* - *f* برای دانه‌ی *GPCR*.

مراجع

- Adams, C. P. and Brantner, V. V. (2006), Estimating the cost of new drug development: is it really \$802 million?, *Health Affairs* 25, 2, 420-428.
- Krantz, A. (1998), Diversification of the drug discovery process., *Nature biotechnology* 16, 13, 1294-1294.
- Wisbart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassansali, M., Stofbard, P., Cheng, Z. and Woolsey, J. (2006), DrugBank: a comprehensive resource for in silico drug discovery and exploration., *Nucleic acids research* 34, suppl 1, D668-D672.
- Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W. and Kanehisa, M. (2008), Prediction of drug-target interaction networks from the integration of chemical and genomic spaces., *Bioinformatics* 24, 13, i232-i240.



شکل ۵: نتایج مربوط به روش *Jachknife-like* برای دانه‌های GPCR.



شکل ۶: نتایج مربوط به روش *fold* - ۱۰ برای دانه‌های Ionchannel.

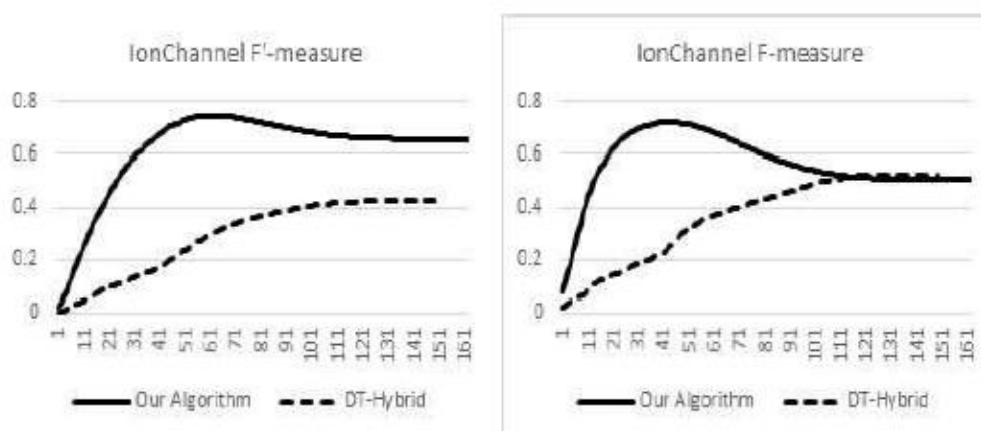
He, Z., Zhang, J., Shi, X. H., Hu, L. L., Kong, X., Cai, Y. D. and Chou, K. C. (2010), Predicting drug-target interaction networks based on functional groups and biological features., *PLoS one* 5, 3, e9603.

Perdman, L., Gottlieb, A., Atias, N., Ruppin, E. and Sharan, R. (2011), Combining drug and gene similarity measures for drug-target elucidation., *Journal of computational biology* 18, 2, 133-145.

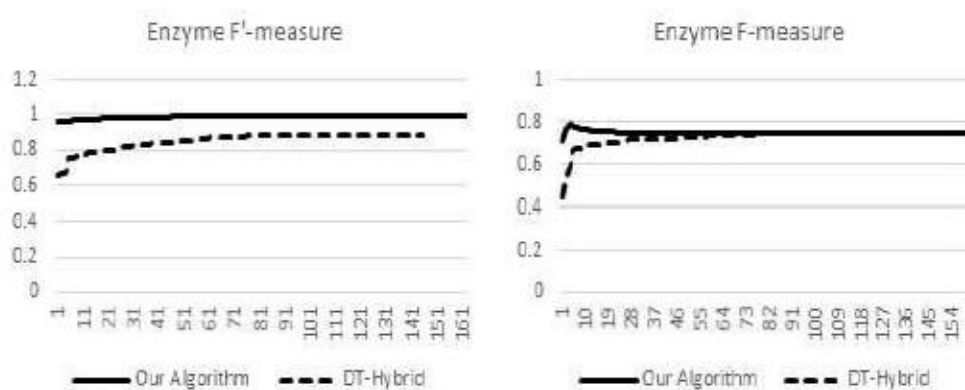
Chen, X., Liu, M. X. and Yan, G. Y. (2012), Drug-target interaction prediction by random walk on the heterogeneous network., *Molecular BioSystems* 8, 7, 1970-1978.

Mei, J. P., Kwok, C. K., Yang, P., Li, X. L. and Zheng, J. (2013), Drug-target interaction prediction by learning from local information and neighbors., *Bioinformatics* 29, 2, 238-245.

Alaimo, S., Pulvirenti, A., Giugno, R. and Ferro, A. (2013), Drug-target interaction prediction through domain-tuned network-based inference., *Bioinformatics* 29, 16, 2004-2008.



شکل ۷، تابع مربوط به روش *Jackknife-like* برای دانه‌های *Ionchannel*.

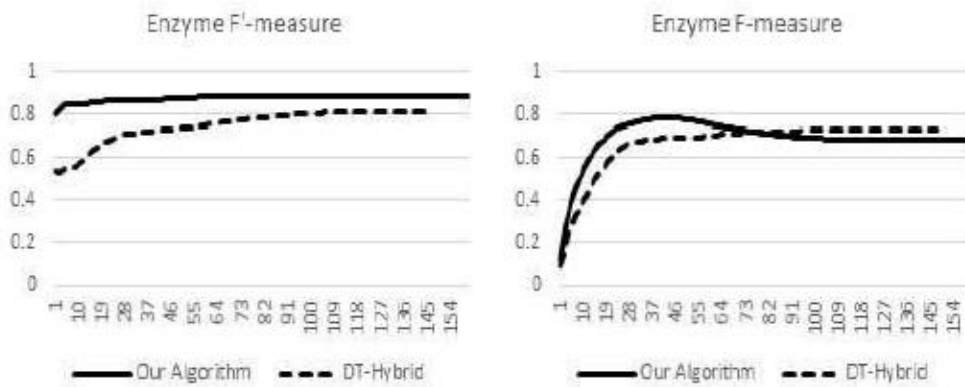


شکل ۸، تابع مربوط به روش *fold* - ۱۰ برای دانه‌های *Enzymes*.

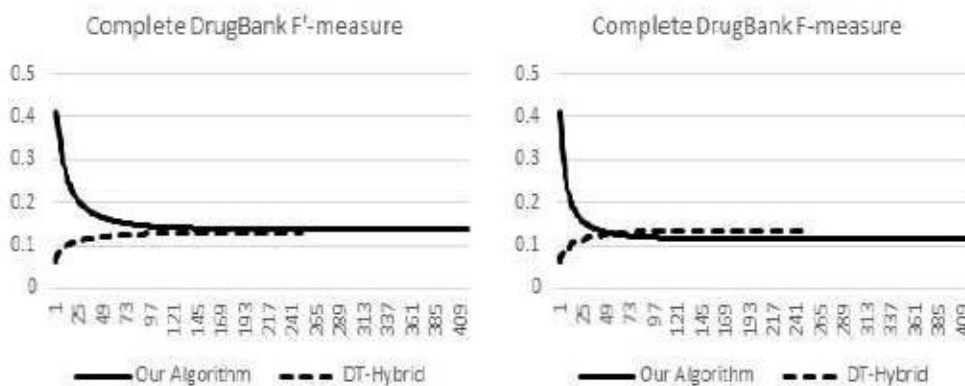
Zhou, T., Ren, J., Medo, M. and Zhang, Y. C. (2007), Bipartite network projection and personal recommendation., *Physical Review E* 76, 4, 046115.

Smith, T. F. and Waterman, M. S. (1981), Identification of common molecular subsequences., *Journal of molecular biology* 147, 1, 195-197.

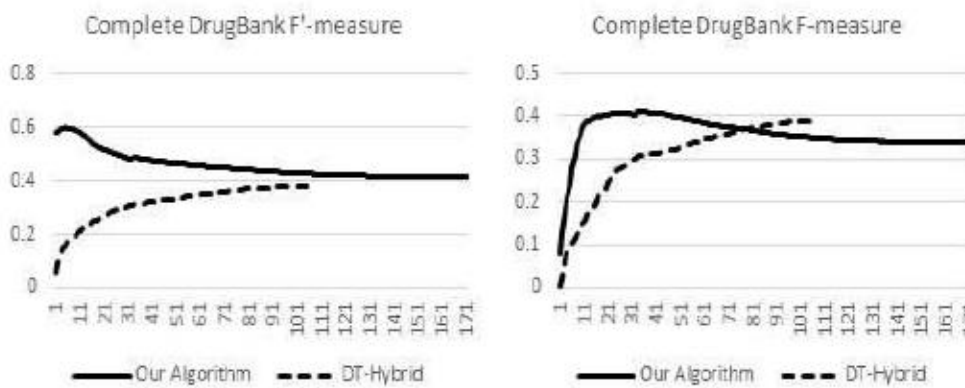
Hattori, M., Okuno, Y., Goto, S. and Kanehisa, M. (2003), Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways., *Journal of the American Chemical Society* 125, 39, 11853-11865.



شکل ۹. نتایج مربوط به روش *Jackknife-like* برای دانه‌های *Enzymes*.



شکل ۱۰. نتایج مربوط به روش *fold* - ۱۰ برای دانه‌های *Complete DrugBank*.



شکل ۱۱. نتایج مربوط به روش *Jackknife-like* برای دانه‌های *Complete DrugBank*.