

بازبینی

Predicting protein–protein interactions based  
only on sequences information

داریوش حسن پور آده

۹۳۰۸۱۶۴

PPI ها مرکزی برای اکثر فرآیندهای بیولوژیکی می باشند. اگر چه تلاشی برای پیش بینی PPI ها و توسعه شبکه های متعامل پروتئینی انجام شده است اما این روش ها به دلیل کمبود اطلاعات پروتئین ها محدود است. در این مقاله، یک روش برای پیش بینی PPI تنها با استفاده از اطلاعات توالی پروتئین ها ارائه شده است. این روش بر اساس یک الگوریتم یادگیری SVM همراه با یک تابع کرنل و یکی از ویژگی های سه گانه متقارن برای توصیف اسیدهای آمینه توسعه داده می شود.

کار جدیدی که در این مقاله انجام می شود این است که پیش بینی های PPI ها را فقط بر اساس توالی پروتئین ها در می آورند. و همچنین خلاصه می کند ویژگی های جفت پروتئین های بر اساس این کلاس بندی های آمینواسیدهایی که کرده است. سپس از یک تابع کرنل استفاده کرده است. همچنین آورده است که برای پیش بینی PPI بر اساس توالی، یکی از چالش های محاسباتی اصلی که وجود دارد این است که یک راه مناسب برای توصیف اطلاعات مهم PPI پیدا کنیم. برای حل این مشکل یک توصیف کننده Conjoint Triad معرفی شده است که خصوصیات یک آمینواسید و دو همسایه کناری را در نظر می گیرد و به عنوان یک واحد ۳ تایی ارائه می دهد. که این روش می تواند کلاس ها را با توجه به آمینواسیدها متفاوت کند - که راه حل اول مقاله را شامل می شود.

مقاله می آید با استفاده از یک SVM و یک تابع کرنل و یک توصیف کننده به عنوان راه حل اول استفاده می کند. اولین کاری که می کند آنزیم ها را در ۷ عدد کلاستر، دسته بندی می کند. از آنجایی که ویژگی ها و توالی پروتئین ها خیلی آنزیم دارند؛ هر خوشه را به همراه دو همسایه چپ و راست آن را در نظر می گیرد و این واحد ۳ تایی را به عنوان ویژگی کلی در نظر می گیرد. فضای برداری با  $(V, F)$  نشان داده که هر عضو مجموعه  $V$  در واقع همان ۳ تایی بدست آمده اند، و  $F$  نیز تکرار آن ۳ تایی ها در توالی پروتئین می باشد. در مورد تعداد ویژگی ها نیز می توان گفت که از آنجایی که آن ۳ تایی ها در کنار هم قرار گرفته اند و هر کدام از آنها ۷ عدد مقدار داشتند پس در کل  $7 \times 7 \times 7$  ویژگی داریم.

حالا اگر طول پروتئین خیلی طولانی باشد مقدار  $F$  که تکرارهایشان را نشان می داد، خیلی زیاد می شد در نتیجه نرمال سازی روی  $F$  انجام داده است.

در مورد مزایای روش ارائه شده می توان گفت که از S-Kernel کرنل استفاده کرده است آمده در روش ارائه شده را با دیگر کرنل ها مقایسه کرده و نشان داده است که S-Kernel بهتر از دیگر کرنل ها نتیجه داده است. و همچنین چون از SVM استفاده کرده است چون شامل دو کلاس پارامتر ظرفیت و نوع کرنل می باشد که این پارامترها را با استفاده از روش هایی پیدا کرده است که همین عمل پیدا کردن پارامترهای مناسب باعث شده است که روش بهتر از دیگر روش ها جواب دهد. همچنین طبق مطالب گفته شده در مقاله می شود الگوریتم ارائه شده به انواع شبکه ها بسط داد - شبکه های تک هسته ای، چند هسته ای و شبکه های متقاطع<sup>۱</sup>.