مقدمه

در سال‌های اخیر شاهد رشد بسیار زیادی در شبکه‌های اجتماعی بوده‌ایم و مطالعات زیادی روی این شبکه‌ها انجام گرفته است. داده‌های شبکه‌های اجتماعی یکی از ابزار محبوب برای مدل کردن رابطه و رفتار افراد و جامعه یا گروهی که در آن عضو هستن بشمار می‌رود. این داده‌ها معمولا به صورت گرافی نمایش داده می‌شود که در آن گره‌ها افراد و لبه‌ها روابط بین این افراد می‌باشد.

همچنین یادگیری ماشین مدت زیادی است که در علوم کامپیوتر جایگاه خود را پیدا کرده و به عنوان ابزار قدرمتندی برای کمک به انسان در زمینه‌های مختلف محسوب می‌شود و ماشین را بیش از پیش در خدمت انسان در آورده است.

در روش‌های اولیه یادگیری ماشین، عمدتا از داده‌ها و متن‌های خام در زمینه‌ی یادگیری استفاده می‌شد. اخیرا از طرف برخی شرکت‌ها و موسسات بزرگ، همچون گوگل، ای‌بی‌ام، مایکروسافت و … پایگاه‌های دانشی معرفی شده است که انجام راه‌کار‌های مختلف یادگیری ماشین را ساده‌تر و کاربردی‌تر کرده است. در این پایگاه‌های دانش اطلاعات مورد نیاز برای عملیات‌های مختلف یادگیری به صورت منظم و نیمه‌منظم موجود است و دغدغه‌ی نرمال کردن و رفع خطا و استخراج حقایق را به مقدار زیادی کم کرده است. این پایگاه‌های دانش عمدتا به صورت یک شبکه از موجودیت‌ها و روابط بین آن‌ها که می‌توان آن را به صورت یک گراف داده نمایش داده به این صورت که گره‌ها موجودیت‌ها و یال‌های بین گره‌ها نشانگر روابط بین آن‌ها باشند، که این یال‌ها می‌توانند از یک نوع باشند یعنی گراف نشانگر یه شبکه تک-رابطه‌ای باشد (همگون) یا هر یال با یک برچسب، نشانگر نوع رابطه باشد و شبکه نشانگر یک گراف چند رابطه‌ای (ناهمگون) باشد.

یکی از مسائل بنیادی و اساسی در یادگیری ماشین روی شبکه‌های اجتماعی، پیش‌بینی پیوند در شبکه‌های همگون و ناهمگون است به این معنی که از اطلاعات موجود در گراف دانش استفاده کرده و وجود یا عدم وجود یک یال را پیش‌بینی کرد، یا به عبارت دیگر مساله‌ی پیش‌بینی پیوند این است که نمایی از یک شبکه به ما داده می‌شود و ما مایل هستیم که بدانیم در آینده نزدیک، احتمالا چه تراکنش‌هایی میان اعضای فعلی شبکه روی می‌دهد و یا اینکه کدام یک از تراکنش‌های موجود را از دست می‌دهیم.

این راهکار در زمینه‌های مختلف یادگیری ماشین مورد استفاده قرار می‌گیرد و کاربرد گسترده‌ای در زندگی انسان پیدا کرده است. برای مثال از این راه‌کار در سیستم‌های توصیه‌گر در فروشگاه‌های اینترنتی، سیستم‌های تشخیص پزشکی، جواب‌گویی به سوال و … استفاده می‌شود.

اکثر مطالعات انجام شده در این زمینه روی شبکه‌های تک‌رابطه‌ای بوده است. به این معنا که روابط بین موجودیت‌ها از یک نوع است و این روابط به صورت دوتایی‌های مرتب مورد استفاده قرار می‌گیرند. برای مثال اگر در یک شبکه اجتماعی رابطه را دوستی بین افراد در نظر بگیریم یال‌های گراف شبکه به صورت «الف، ب» خواهد بود به این معنی که شخص الف با شخص ب رابطه دوستی دارد.

پایگاه‌های دانشی که اخیرا معرفی شده‌اند عمدتا داده‌ها را به صورت داده‌های چند رابطه‌ای ذخیره می‌کنند و اطلاعات بیشتری از یک رابطه دوتایی بلی یا خیر به ما می‌دهند. منظور از داده‌های چند رابطه‌ای گراف جهت‌داری است متشکل از موجودیت‌ها و روابط بین آن‌ها که بصورت «مبدا h، رابطه r، مقصد t» نمایش داده می‌شود، به این معنی که یک رابطه r بین موجودیت‌های h و t وجود دارد. برای مثال سه‌تایی «تهران، واقع در، ایران» این اطلاع را به ما می‌دهد که استان تهران داخل کشور ایران قرار دارد. در این نوع پایگاه دانش هم انواع مختلف موجودیت وجود دارد و هم انواع مختلف رابطه بین موجودیت‌ها. پایگاه‌های دانشی مانند Google Knowledge Graph، FreeBase و WordNet وجود دارند که شامل تعداد زیادی نمونه چندرابطه‌ای می‌باشند و تعداد زیادی موجودیت و روابط بین آن‌ها را می‌توان در آن‌ها یافت و از آن برای یادگیری مدل استفاده کرد.

تعریف مساله

روش‌های مختلفی برای حل مساله‌ی پیش‌بینی پیوند در پایگاه‌های دانش ناهمگون ارائه شده است که از رویکرد‌های مختلفی سعی به حل این مساله می‌کنند از جمله روش‌های آماری، روش‌های ویژگی‌های پنهان، روش‌های ویژگی‌های گراف و …، تمرکز ما در این مقاله بر روی روش‌های پیش‌بینی پیوند مبتنی بر ویژگی‌های پنهان خواهد بود.

در روش‌های مبتنی بر ویژگی‌های پنهان، با استفاده از ویژگی‌هایی که در موجودیت‌ها و روابط بین آن‌ها وجود دارد سعی می‌شود میزان ارتباط بین موجودیت‌ها را تشخیص دهیم و به این صورت وجود یک پیوند را تایید یا رد کنیم. برای مثال اگر دو شخص با هم همکار هستند، به احتمال زیادی ویژگی‌های مشترکی دارند، مثلا هر دو اهل یک شهر هستند، هر دو در یک رشته‌ی دانشگاهی تحصیل کرده‌اند، خصوصیات اخلاقی یکسانی دارند و …. در رابطه‌ی همکار بودن به هیچ یک از این ويژگی‌ها به طور مستقیم اشاره نشده است و این ویژگی‌ها به طور ضمنی در این موجودیت‌ها قرار دارند که با استفاده از آن می‌توانیم وجود یا عدم وجود رابطه‌ی همکار بودن را حدس بزنیم. پس هر موجودیت می‌تواند تعداد زیادی ویژگی پنهان داشته باشد که رابطه‌ها به این ویژگی‌ها وزن می‌دهند، مثلا در رابطه‌ی همکاری احتمال اینکه رشته‌ی تحصیلی دو شخص در همکار شدن آن‌ها تاثیر گذار باشد بیشتر از ویژگی رنگ پوست دو شخص است، پس وزن ویژگی رشته‌ي تحصیلی در این رابطه بیشتر از ویژگی رنگ پوست است.

روش‌های زیادی برای حل مساله‌ی پیش‌بینی پیوند مبتنی بر ویژگی‌های پنهان ارائه شده است. همه‌ی این روش‌ها از یک روال ثابتی برای حل مساله استفاده می‌کنند و هر کدام با نوآوری‌هایی که داشته‌اند بهبود‌هایی در نتایج بدست‌آمده حاصل کرده‌اند. این روش‌ها در بخش + به طور کامل معرفی خواهند شد.

در این پژوهش قصد داریم که با استفاده از قوانین انجمنی موجود در پایگاه‌های دانش به دو هدف برسیم،

1. بهبود نتایج در روش‌های معرفی شده‌ی موجود: در ادامه پنج روش از روش‌ها معروف مبتنی بر ویژگی‌های پنهان که به حل مساله‌ی پیش‌بینی پیوند پرداخته‌اند را انتخاب می‌کنیم و در بخش + نشان می‌دهیم که استفاده از قوانین انجمنی در فرایند آموزش این پنج روش باعث بهبود در نتایج این روش‌ها خواهد شد.
2. ارائه‌ی چارچوبی برای سنجش نقاط ضعف و قوت روش‌های ارائه شده: همچنین با دسته‌بندی کردن قوانین انجمنی در بخش + نشان می‌دهیم که با اعمال جداگانه‌ی دسته‌های مختلف قوانین انجمنی می‌توانیم میزان قدرت و ضعف روش‌ها را در قانون‌های مختلف بررسی کنیم که با برطرف کردن نقاط ضعف می‌توان روش‌های قوی‌تری ارائه کرد.

روش‌های انجام پژوهش

برای دست‌یابی به اهدافی که در بخش قبل مطرح شد، همانطور که اشاره شد از قوانین انجمنی استفاده خواهیم کرد. در روش‌های معرفی شده فقط از روابط موجود در پایگاه‌دانش در فرایند آموزش استفاده شده است. مثلا اگر در پایگاه‌دانش داشته باشیم که «علی، پدر، حسن» به این معنی که علی پدر حسن است می‌توانیم به این نتیجه برسیم که «حسن، فرزند، علی» و از آن در فرایند آموزش استفاده کنیم اما در این روش‌ها این روابط مورد استفاده قرار نگرفته اند. در این پژوهش قصد داریم که اینگونه قوانین را به کمک روش‌هایی که در بخش + توضیح داده می‌شوند از پایگاه دانش استخراج کنیم و به کمک معیار‌های کیفیت که در بخش + معرفی می‌شوند قوانینی که کیفیت مناسب دارند را انتخاب کنیم و در فرایند آموزش از آن‌ها استفاده کنیم، نحوه‌ی استفاده از این قوانین در مدل‌های موجود در بخش + توضیح داده شده است.

کاربرد‌های پژوهش

مسائل پیش‌بینی پیوند در زمینه‌های زیادی قابل بکار‌گیری هستند و همین امر باعث شده است که در سال‌های گذشته بسیار مورد توجه و تحقیق قرار بگیرند. به عنوان مثال در ادامه تعدادی از این کاربردها را مطرح خواهیم کرد:

* پیشبینی پیوند‌های احتمالی در شبکه‌های اجتماعی، برای مثال در شبکه‌های اجتماعی بین کاربران و مطالب ثبت شده، نظرات و … بررسی شود که روابطی مانند دوستی، پسندیدن و نپسندیدن، روابط فامیلی و… وجود دارد یا خیر
* استفاده یه عنوان سیستم‌های توصیه‌گر، برای مثال کاربران و کالاها یا اشیاء موجودیت‌ها هستند و خریدن، امتیاز دادن، بررسی کردن و … رابطه‌ها هستند که می‌توان از بررسی این روابط و موجودیت‌ها اطلاعاتی نظیر کاربر xکالای y را خواهد خرید یا خیر یا اینکه کالای x به تعداد بالا فروش خواهد رفت یا خیر بدست آورد.
* استفاده در سیستم‌های تشخیص پزشکی، برای مثال شبکه‌ی بین بیماران، بیماری‌ها، دارو‌ها و ... را در نظر بگیرید، با بررسی دقیق این شبکه می‌توان علل و درمان‌های بیماری‌ها را به کمک ماشین بدست آورد.

علاوه بر کاربردهایی که در بالا برای پیش‌بینی پیوند اشاره شد، از روش ارائه شده در این پژوهش می‌توان برای کامل‌تر کردن پایگاه‌های داده نیز استفاده کرد، به این صورت که روابطی که از قوانین انجمنی استخراج شده از پایگاه‌دانش بدست می‌آید و کیفیت لازم را دارد را نیز به پایگاه دانش اضافه کنیم و آن را کامل‌تر کنیم.

ساختار پایان‌نامه

متن