Table of Contents

[مروری روش‌ها موجود و کار‌های قبلی 9](#_Toc484911596)

[مقدمه 9](#_Toc484911597)

[روش‌های مبتنی بر ویژگی‌های پنهان 10](#_Toc484911598)

[تقسیم‌بندی داده‌های پایگاه دانش 12](#_Toc484911599)

[مراحل آموزش مدل 14](#_Toc484911600)

[روش‌های موجود 17](#_Toc484911601)

[۲- روش‌های دوخطی – رسکال 18](#_Toc484911602)

[۳- روش‌های ادراک چندلایه‌ای 23](#_Toc484911603)

[۳- روش‌های فاصله‌ی پنهان 28](#_Toc484911604)

* مقدمه
  + ضرورت و هدف
  + تعریف مساله
  + روش‌های انجام پژوهش
  + ساختار پایان‌نامه
* پیش‌بینی پیوند در شبکه‌های ناهمگون
  + شبکه‌های ناهمگون
  + پیش‌بینی پیوند
  + چالش‌ها
* مروری بر مطالعات انجام شده
  + مقدمه
  + نحوه‌ی آموزش
  + روش‌های مبتنی بر ویژگی‌های پنهان
  + نتایج
* عملکرد روش‌های موجود
  + مقیاس‌پذیری
  + نتایج روش‌ها
  + نقاط قوت و ضعف
* قوانین انجمنی
  + مقدمه
  + معیار اطمینان
  + الگوریتم
  + قوانین هدف
* استفاده از قوانین انجمنی
  + نحوه‌ی بهبود در مدل‌ها
* آزمایشات و نتیجه گیری و کار‌های آینده
  + آزمایشات
* نتیجه‌گیری
  + نتیجه‌گیری
  + کار‌های آینده
* مراجع
* پیوست‌ها
* مقدمه
  + ضرورت و هدف
  + تعریف مساله
  + روش‌های انجام پژوهش
  + ساختار پایان‌نامه
* مفاهیم پایه
  + شبکه‌های همگون و ناهمگون
  + پیش‌بینی پیوند
  + فرض جهان باز و بسته
  + SGD
* پیشینه‌ی پژوهش
  + روش‌های کلی آموزش
  + روش‌های مبتنی بر ویژگی‌های پنهان
    - RESCAL
    - NTN
    - SE
    - TransE
    - TransH
  + قوانین انجمنی
    - مقدمه
    - معیار اطمینان
    - الگوریتم
    - قوانین هدف
  + جمع‌بندی
* روش پیشنهادی
  + چالش‌ها
  + عملکرد روش‌های موجود
    - مقیاس‌پذیری
    - نتایج روش‌ها
    - نقاط قوت و ضعف
  + استفاده از قوانین انجمنی
    - نحوه‌ی بهبود در مدل‌ها
* ارزیابی
* نتیجه‌گیری
  + نتیجه‌گیری
  + کار‌های آینده

مقدمه

تعریف مساله و هدف و ضرورت

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. پیش‌بینی وقوع پیوند‌ها، یک مساله اساسی و بنیادین در شبکه‌های اجتماعی می‌باشد. در موضوع پیش‌بینی پیوند، 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.

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.

در سال‌های اخیر پایگاه‌های دانشی ایجاد شده‌اند که داده‌ها در آن به صورت چند رابطه‌ای ذخیره شده‌اند و اطلاعات بیشتری از یک رابطه دوتایی بلی یا خیر به ما می‌دهند. ما در این تحقیق از داده‌های چند رابطه‌ای استفاده می‌کنیم. منظور از داده‌های چند رابطه‌ای گراف جهت‌داری است متشکل از موجودیت‌ها و روابط بین آن‌ها که بصورت «مبداh، رابطهr، مقصدt» نمایش داده می‌شود، به این معنی که یک رابطه r بین موجودیت‌های h وt 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. شناخت پایگاه‌دانش و دریافت اطلاعات از آن نیز یکی از مسايل پیشبینی‌پیوند روی داده‌های چند رابطه‌ای است.

پیشبینی پیوند مقیاس‌پذیر روی داده‌های چند رابطه‌ای عمدتا با استفاده از ویژگی‌های پنهان صورت می‌گیرد. مشکل این روش‌ها تعداد بسیار زیاد پارامترها در فاز یادگیری مدل است به صورتی که انجام عملیات یادگیری مدل را غیرممکن می‌کند. کاهش تعداد پارامترهای مساله که در برخی روش‌ها پیشنهاد شده نیز برخی از الگوی‌های موجود در پایگاه دانش را در نظر نمی‌گیرند و باعث کاهش دقت مدل خواهند شد. در این تحقیق قصد داریم که بدون افزایش تعداد پارامتر‌های روش‌های الگوی پنهان شده و با استفاده از قوانین انجمنی موجود در پایگاه‌دانش عمل‌کرد این روش‌ها را بهبود بخشیم.

در این تحقیق قصد داریم که با کار روی داده‌های چند رابطه‌ای به سؤالاتی که از پایگاه‌دانش می‌شود پاسخ دهیم، به این صورت که یک مدل از روی داده‌های پایگاه دانش ساخته و آموزش داده می‌شود و پس از آن سؤالات به گونه‌ای که مقصد یا مبدأ آن مجهول باشد از مدل ایجاد شده پرسیده می‌شود. برای نمونه در مثال بالا اگر قسمت کشور مجهول باشد از پایگاه دانش پرسیده می‌شود «تهران، واقع در، ؟» و انتظار می‌رود که مدل آموزش دیده پاسخ سؤال را با دقت نسبتاً خوبی بیابد.

نتایج این مساله کاربرد‌های زیادی خواهد داشت 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.
* کامل‌تر کردن پایگاه‌های داده، از نتایج این تحقیق می‌توان استفاده کرد و با اضافه کردن رابطه‌های انجمنی بدست آمده به پایگاه‌داده آن را کامل‌تر کرد.

روش‌ها:

داده‌هایی که قصد کار روی آن‌ها را داریم معمولاً ابعاد بالایی خواهند داشت. برای مثال پایگاه دانش Freebase شامل یک میلیون نوع موجودیت و ۲۵هزار نوع رابطه و ۱۷ میلیون نمونه برای آموزش و آزمون مدل است. از آنجایی که کار روی داده‌هایی با این تعداد ابعاد بسیار سخت و تقریباً غیر ممکن است باید از روش‌های کاهش ابعاد استفاده کنیم. در این تحقیق سعی می‌کنیم که این داده‌ها را به فضای برداری آورده و مساله را به یک مساله کمینه‌سازی تبدیل کنیم. این داده‌ها را به صورت یک تنسور[[1]](#footnote-1) فرض می‌کنیم که برای هر رابطه یک فضای دوبعدی برداری برای موجودیت‌های مختلف در نظر می‌گیریم، همچنین برای هر رابطه یک بردار بین موجودیت‌های دخیل در آن رابطه خواهیم داشت. سپس تلاش می‌کنیم که از این بردار ها در یک مساله کمینه سازی استفاده کنیم. برای مثال اگه سه‌تایی «head, relation, tail» را داشته باشیم که هرکدام به یک بردار نگاشت شده‌اند انتظار داریم مجموع بردار head و relation به بردار tail برسد پس باید مقدار زیر را کمینه کنیم:

head + relation – tail => 0

برای نمایش این قانون از تخمین مرتبه پایین استفاده می‌کنیم به صورت:

fr(h,t) = ||h + r – t||22

در صورتی که تابع f برای سه‌تایی «h, r, t» مقداری نزدیک به صفر داشته باشد این رابطه برقرار است و درغیر این صورت این رابطه برقرار نیست.

در تحقیقاتی که قبلاً روی این مدل داده‌ها انجام شده است فقط از حقایقی که به صورت مستقیم داده شده استفاده شده است. برای مثال اگر در حقایق موجود در داده‌ها داشته باشیم «الف، مادر، ب» و «الف، همسر، ج» فقط از این دو حقیقت استفاده می‌شود و از این حقیقت که «ج، پدر، ب» استفاده نمی‌شود. در این تحقیق تصمیم داریم از قوانین انجمنی بین داده‌ها نیز استفاده کرده و حقایق ضمنی که بین داده‌ها هست را نیز در آموزش مدل استفاده کنیم.

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 رو 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 راهکارهای یادگیری ماشین، 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. 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.

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 است.

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.

داده‌های چندرابطه‌ای به یک گراف جهت‌دار گفته می‌شود که در آن گره‌ها نشانگر موجودیت‌ها و یال‌ها نشانگر رابطه‌ی بین موجودیت‌ها است، که هر رابطه به صورت یک سه‌تایی (t، r، h)[[2]](#footnote-2) نمایش داده می‌شود به این معنی که موجودیت h رابطه‌ی به نام r دارد با موجودیت t. [TransE] مدل‌های بر اساس داده‌های چند رابطه‌ای نقش اساسی‌ای در بسیاری از زمینه‌های یادگیری ماشین بازی می‌کنند. برای مثال:

* در آنالیز شبکه‌های اجتماعی که در آن اعضای شبکه‌ی اجتماعی گره‌های گراف (موجودیت‌ها) و روابط دوستی بین اعضا یال‌ها (رابطه‌ها) را تشکیل می‌دهد.
* در سیستم‌های توصیه‌گر مانند یک فروشگاه آنلاین که در آن کاربران و کالا‌ها موجودیت‌ها و پیوند‌هایی مانند خرید کردن، نظر دادن، امتیاز دادن، جستجو کردن رابطه‌ها را در گراف داده تشکیل می‌دهند.

# مرور روش‌ها موجود و کار‌های قبلی

نتایج روش‌های موجود

جمع‌بندی

در این فصل ابتدا معرفی مختصری از قوانین انجمنی داشتیم و در ادامه معیار‌های پشتیبان، پوشش سر، اطمینان استاندارد و اطمینان PCA را معرفی کردیم که به کمک این معیارها میزان کیفیت و قابل اعتماد بودن این قانون‌ها را می‌سنجیم. در ادامه در بخش + نحوه‌ی استخراج قوانین انجمنی از پایگاه‌های دانش معرفی شده را مطرح کردیم و الگوریتم این روش را توضیح دادیم. پس از آن قوانین هدف که قصد داریم از آن‌ها برای بهبود روش‌ها استفاده کنیم را تعریف کردیم و در پایان توضیح دادیم که از هریک از این قوانین چگونه برای بهبود مدل‌های مبتنی بر روش‌های پنهان استفاده کنیم.

آزمایش‌ها و نتیجه‌گیری

کارهای آینده

1. Tensor [↑](#footnote-ref-1)
2. مشخص کننده (head, relation, tail) می‌باشد که به صورت (subject, predicate, object) هم تعریف می‌شود. [↑](#footnote-ref-2)