Sciences and Techniques of Information Management
Volume 9, Issue 3, 2023

Automatic Inference of Terminology Relationships in the Persian Islamic Sciences Thesaurus using Graph Convolutional Networks (GCNs)

Said Abolhasan Nezamdost
P.hD. Student, Department of Knowledge and Information Science, Kharazmi University, Tehran, Iran.
hasannezamdost@gmail.com

Ali Azimi
Assistant Professor, Department of Knowledge and Information Science, Kharazmi University, Tehran, Iran **(Corresponding author).** azimia@gmail.com

Amir Jalali Bidgoli
Assistant Professor, Department of Computer Engineering, Qom University, Qom, Iran.
jalaly@qom.ac.ir

Nosrat Ali Ashrafi Piyaman
Assistant Professor, Department of Computer Engineering, Kharazmi University, Tehran, Iran.
ashrafi@khu.ac.ir

Abstract

Purpose: The present research aims to develop a model for automatically inferring the relationships between terms in the Thesaurus of Islamic Sciences using Graph Convolutional Networks (GCN). By employing new algorithms in the field of deep learning, the research seeks to enhance the efficiency of information retrieval in the Thesaurus of Islamic Sciences. To enhance accuracy and comprehensiveness, reduce costs, and improve relationships between terms.

Method: The current research employed used of convolutional networks method, networks, is are one of the crucial techniques methods in the field of learning. This method is capable of leveraging from the relationship patterns in the while also focusing on to the characteristics of each node. The dataset under study comprises all the terms from the thesaurus of Islamic sciences generated between 1994 and the early 2022, which are represented as a graph. The vertices represent the terms, and the edges represent the relationships between the terms in the graph. This graph is provided as input to the convolutional network, which then generates a model for the automatic inference of connections. And in order to analyze the obtained outputs, AP and ROC standards have been used.

Findings: The revealed showed the model achieved the average accuracy 75% and a Roc score of 72% obtained for the data. It is noteworthy to accept the results considering that this method was used for the first time in the field of Islamic sciences and thesauruses.

Conclusion: Despite shift in preference opinion thesauri thesauruses to ontologies, the use thesauri remains still of particularly especially in Iran. Compared to previous research, the method used to construct the thesaurus is different, resulting in more reliable outcomes. Consequently, we

Cite this article: Nezamdost, S.A., Azimi, A., Jalali Bidgoli, A. & Ashrafi Piyaman, N.A. (2023). Automatic Inference of Terminology Relationships in the Persian Islamic Sciences Thesaurus using Graph Convolutional Networks (GCNs). Sciences and Techniques of Information Management, 9(3): 75-102. https://doi.org/10.22091/STIM.2023.8958.1912

Received: 2023-06-30 ; **Revised:** 2023-07-19 ; **Accepted:** 2023-07-29 ; **Published online:** 2023-08-01

Article type: Research **Published by:** University of Qom.

can expect improved results for various purposes, such as automatic indexing. New advancements in natural language processing and deep learning also give us hope for improvements in information retrieval and automatic indexing.

Keywords: Thesaurus Relations, Graph Convolutional Networks, Islamic Sciences Thesaurus, Machine Learning.

علوم و فنون مدیریت اطلاعات
دوره ۹، شماره ۳، ۱۴۰۲

استنتاج خودکار روابط بین اصطلاحات در اصطلاح‌نامه فارسی علوم اسلامی با استفاده از شبکه‌های پیچشی گرافی

سید ابوالحسن نظام‌دوست
دانشجوی دکتری، گروه علم اطلاعات و دانش‌شناسی، دانشکده روانشناسی و علوم تربیتی، دانشگاه خوارزمی، تهران، ایران.
hasannezamdost@gmail.com

علی عظیمی
استادیار، گروه علم اطلاعات و دانش‌شناسی، دانشکده روانشناسی و علوم تربیتی، دانشگاه خوارزمی، تهران، ایران (**نویسنده مسئول**). azimia@gmail.com

امیر جلالی بیدگلی
استادیار، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه قم، قم، ایران. jalaly@qom.ac.ir

نصرت‌علی اشرفی‌پیامن
استادیار، گروه مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه خوارزمی، تهران، ایران. ashrafi@khu.ac.ir

چکیده

هدف: پژوهش حاضر درصدد ارائه مدلی برای استنتاج خودکار روابط بین اصطلاحات در اصطلاح‌نامه علوم اسلامی با استفاده از شبکه‌های پیچشی گرافی (GCN) بود، تا با استفاده از الگوریتم‌های جدید در حوزه یادگیری عمیق بتواند از اطلاعات موجود در اصطلاح‌نامه علوم اسلامی، سرعت، دقت و جامعیت را افزایش داده و موجب کاهش هزینه‌ها و در عین‌حال باعث بهبود روابط بین اصطلاحات شود.

روش: پژوهش حاضر، از روش شبکه‌های پیچشی گرافی که یکی از مهم‌ترین روش‌های مطرح در حوزه یادگیری عمیق بوده و قادرند در کنار توجه به ویژگی‌های هر گره، از الگوهای روابط در گراف نیز سود بجویند، استفاده کرده است. دیتاست مورد مطالعه عبارت است از کلیه اصطلاحات اصطلاح‌نامه علوم اسلامی، که از سال ۱۳۷۲ تا ابتدای ۱۴۰۰ تولید شده، و به صورت یک گراف در نظر گرفته شده‌اند. اصطلاحات به عنوان رئوس و ارتباط بین اصطلاحات به‌عنوان یال‌های این گراف هستند و این گراف به عنوان ورودی، به شبکه پیچشی گرافی داده شده و مدلی برای استنتاج خودکار ارتباطات حاصل شده است. به منظور تجزیه و تحلیل خروجی‌های حاصل، از معیارهای AP و Roc استفاده شد.

یافته‌ها: یافته‌های پژوهش نشان داد که میانگین دقت مدل بدست آمده برای داده‌های تست، ۷۵ درصد و همچنین امتیاز Roc حاصل شده، برای داده‌های تست، ۷۲ درصد می‌باشد، و با توجه به اینکه این روش در حوزه علوم اسلامی و

پژوهش حاضر برگرفته از: رساله دکتری با عنوان: **استنتاج خودکار روابط بین اصطلاحات در اصطلاح‌نامه فارسی علوم اسلامی با استفاده از شبکه‌های پیچشی گرافی**، ارائه شده در گروه علم اطلاعات و دانش‌شناسی، دانشکده روانشناسی و علوم تربیتی، دانشگاه خوارزمی، در سال ۱۴۰۱ است.

استناد به این مقاله: نظام‌دوست، س.ا.، عظیمی، ع.، جلالی بیدگلی، ا.، اشرفی‌پیامن، ن.ع. (۱۴۰۲). استنتاج خودکار روابط بین اصطلاحات در اصطلاح‌نامه فارسی علوم اسلامی با استفاده از شبکه‌های پیچشی گرافی. *علوم و فنون مدیریت اطلاعات*، ۹(۳): ۷۵-۱۰۲.
https://doi.org/10.22091/STIM.2023.8958.1912

تاریخ دریافت: ۱۴۰۲/۰۴/۰۹ ؛ **تاریخ اصلاح:** ۱۴۰۲/۰۴/۲۸ ؛ **تاریخ پذیرش:** ۱۴۰۲/۰۵/۰۷ ؛ **تاریخ انتشار آنلاین:** ۱۴۰۲/۰۵/۱۰

نوع مقاله: پژوهشی
ناشر: دانشگاه قم

اصطلاح‌نامه‌ها برای اولین بار مورد استفاده قرار، نتایج قابل قبول است.

نتیجه‌گیری: علی‌رغم چرخش نظر از اصطلاح‌نامه‌ها به هستی‌شناسی‌ها، هنوز هم استفاده از اصطلاح‌نامه‌ها، مخصوصاً در کشور ایران مورد توجه است. در مقایسه با پژوهش‌های قبلی، روش استفاده شده برای ساخت اصطلاح‌نامه، متفاوت بوده و نتایج بدست آمده موجب اطمینان بیشتری است و در نتیجه در اهداف مختلف کاربرد اصطلاح‌نامه ازجمله نمایه‌سازی خودکار، خروجی‌های بهتری بدست آمده است. شیوه‌های جدید در پردازش زبان طبیعی و یادگیری عمیق نیز ما را در بازیابی اطلاعات و نمایه‌سازی خودکار امیدوارتر می‌کند.

کلیدواژه‌ها: روابط اصطلاح‌نامه‌ای، شبکه پیچشی گرافی، اصطلاح‌نامه فارسی علوم اسلامی، یادگیری ماشین، استنتاج خودکار.

۱. مقدمه

یکی از انواع رایج زبان‌های کنترل‌شده، اصطلاح‌نامه[1] است. اصطلاح‌نامه، مجموعه اصطلاحات[2] استاندارد، گزیده شده و نظام‌یافته‌ای است که بین آن‌ها، روابط معنایی و رده‌ای، یا سلسله مراتبی برقرار بوده و توانایی آن را دارد که یک حوزه موضوعی را با تمام جنبه‌های اصلی، فرعی و وابسته به شکل نظام‌یافته و به قصد ذخیره و بازیابی اطلاعات و مدارک و مقاصد جنبی دیگر عرضه کند. هدف اصطلاح‌نامه ایجاد یکدستی در نمایه‌سازی اسناد و سهولت در کاوش اطلاعات است. منظور از یکدستی، ایجاد معیارهایی برای انتخاب واحد از بین چند انتخاب در دسترس است. از این‌رو، وظیفه اصطلاح‌نامه تهیه نقاط دسترسی مؤثری است که از طریق آن‌ها[3] مدارك مرتبط بازیابی می‌شوند. بین هدف و وظیفه اصطلاح‌نامه رابطه نزدیکی وجود دارد، اصطلاح‌نامه زمانی می‌تواند به هدف خود برسد که وظیفه خود را به نحو مؤثر انجام دهد. در غیر این صورت وجود «نقاط مؤثر دسترسی» مفهومی نخواهد داشت (موسایی، ۱۳۸۷).[4]

در یک پایگاه داده، لایه اصطلاح‌نامه، بین زبان طبیعی کاربر با اسناد نمایه شده ارتباطی کنترل شده و منطقی ایجاد کرده، به نحوی که بازیابی اسناد به نحو موثری انجام شود. مهم‌ترین ویژگی اصطلاح‌نامه‌ها مشخص بودن روابط دقیق و تعریف شده بین اصطلاحات است. به عبارت دیگر، باید تمام روابط منطقی و مرتبط با هر اصطلاح، در فرایند تدوین اصطلاح‌نامه دیده شده باشد؛ ولو اینکه اصطلاح تنها دارای یك ارجاع باشد. در غیر این صورت، آنچه اصطلاح‌نامه نامیده می‌شود، تفاوت چندانی با یك فرهنگ لغت تخصصی نخواهد داشت.

روابط در اصطلاح‌نامه به سه دسته تقسیم می‌شوند: روابط هم‌ارز، روابط سلسله مراتبی، و روابط مصداقی. روابط هم‌ارز، روابطی هستند که بین اصطلاحات مترادف و یا شبه مترادف برقرار می‌شود. این روابط همچنین بین اصطلاحاتی که هم‌ارز تعریف شده‌اند، برقرار است. روابط سلسله مراتبی، روابطی است که بین دو اصطلاح برقرار می‌گردد، تا عام‌تر یا خاص‌تر بودن آن دو اصطلاح را نسبت به یکدیگر مشخص کند. به‌کار بردن رابطه سلسله مراتبی از این جهت اهمیت دارد که چنانچه در یك

1. THESAURUS
2. Terms
3. https://vista.ir/w/a/16/i0uki/%D8%A7%D8%B5%D8%B7%D9%84%D8%A7%D8%AD-%D9%86%D8%A7%D9%85%D9%87-%DA%86%DB%8C%D8%B3%D8%AA
4. https://vista.ir/w/a/16/i0uki/%D8%A7%D8%B5%D8%B7%D9%84%D8%A7%D8%AD-%D9%86%D8%A7%D9%85%D9%87-%DA%86%DB%8C%D8%B3%D8%AA

موضوع، اصطلاحات متعددی داشته باشیم، و تمام آن اصطلاحات را با یك اصطلاح برگزیده بیان کنیم، آن اصطلاح برگزیده دربرگیرنده فهرست بزرگی از آن موضوع خواهد بود و چنین اصطلاح برگزیده‌ای برای کاوش، مناسب و مفید نیست. بنابراین، برای بیان موضوعات تحت پوشش موضوع اصلی باید از اصطلاحات خاص‌تری سود برد. در این حالت از ارجاع سلسله مراتبی عام و خاص استفاده می‌کنیم. در روابط مصداقی (موردی)، اصطلاح خاص، مصداقی و موردی از عام است. در یك رابطه مصداقی، کل تنها دارای جز است و نه بیشتر (تندپور، ۱۳۸۳).

در روش‌های سنّتی، روابط بین اصطلاحات به‌صورت دستی تعیین می‌شود. طی مصاحبه‌ای با محمد کریمی (۲۰ مهر ۱۴۰۰)، کارشناس پایگاه‌های اطلاعاتی گروه اشاعه پژوهشکده اطلاعات و مدارك اسلامی، رویه سنتی کار اصطلاح‌نامه علوم اسلامی به این گونه توضیح داده شد که در مرحله اول، پیشنهاددهنده که می‌تواند نمایه‌ساز، کارشناس اصطلاح‌نامه یا هر فرد دیگری باشد، اصطلاح مدنظر را ارائه می‌دهد و در مرحله دوم، پیشنهاد توسط کارشناس اصطلاح‌نامه بررسی می‌شود. در این مرحله ابتدا از نظر صحّت و ارزش‌گذاری محتوا، اصطلاح مورد وارسی قرار گرفته و سپس بررسی می‌شود که آیا اصطلاح پیشنهادی، شأنیّت یک واژه کلیدی را دارد. همچنین در این مرحله، جایگاه واژه در بین علوم مختلف بررسی می‌شود. در این مرحله دو نکته حائز توجه است: اگر واژه از نظر معنایی تک معنایی باشد، جایگاه آن مشخص بوده و سایر حوزه‌ها در صورت استفاده، به آن وابستگی ایجاد می‌کنند و اگر واژه از نظر معنایی، چندمعنایی باشد، باید بررسی دقیق‌تری صورت گیرد تا مشخص شود در کدام حوزه‌های علوم قرار خواهد گرفت. در نهایت، جستجوی روابط یک اصطلاح توسط کارشناس اصطلاح‌نامه انجام می‌شود. زمانی که واژه‌ای وارد این مرحله شد، می‌توان آن را یک اصطلاح در نظر گرفت. در نهایت و در مرحله سوم، اصطلاح پیشنهادی توسط کارشناس اصطلاح‌نامه مورد نظارت و ارزیابی قرار می‌گیرد و در صورت قبول، به ساختار اصطلاح‌نامه اضافه می‌شود و در صورت اختلاف، متناسب با میزان اختلاف، به کارشناس سومی که گروه مشخص می‌کند، ارجاع داده می‌شود تا نظر نهایی را اعلام نماید. در پاره‌ای موارد به جای نفر سوم، موضوع به شورای علمی گروه ارجاع داده می‌شود. یا حتی بعد از نفر سوم نیز مجدد شورای علمی گروه اظهارنظر می‌نماید. به صورت میانگین، شناسایی یک اصطلاح به همراه روابط آن، ۲ ساعت به طول می‌انجامد (کریمی، ۲۰ مهر ۱۴۰۰).

بنابر آنچه ذکر شد، فرایند سنتی زمان‌بر و با صرف هزینه و توان نیروی انسانی مضاعف انجام می‌شود. اگر بتوانیم از داده‌هایی که شامل روابطی بین اصطلاحات است که به وسیله متخصصان حوزه دانشی ایجاد شده و همچنین مدارکی که بر پایه این اصطلاحات نمایه شده‌اند، استفاده کنیم، و

با بهره‌برداری از الگوریتم‌ها و تکنیک‌های مطلوب در پردازش زبان طبیعی و یادگیری ماشین، روابط پیشین و پسین میان اصطلاحات را به صورت خودکار اصلاح و کشف نماییم، آنگاه می‌توان فرایند توسعه و تکامل اصطلاح‌نامه را به خودش واگذار کرد. پیاده‌سازی این ایده علاوه‌بر اینکه هزینه روزآمد نگه داشتن اصطلاح‌نامه را کاهش می‌دهد، توانمندی بهینه‌سازی اصطلاح‌نامه به عنوان یك سامانه خود-اتکاء[1] را نیز در خلال زمان ارتقاء خواهد داد. استنتاج خودکار روابط اصطلاح‌نامه‌ای را می‌توان به مدد فنون پردازش زبان طبیعی و بهره‌گیری از یادگیری ماشینی عملی کرد.

در زمینه تولید و کاربرد اصطلاح‌نامه‌ها، تلاش‌های زیادی انجام شده و نمونه‌های مختلفی تدوین و توسعه داده شده‌اند. از آن جمله می‌توان در سطح بین‌المللی به اصطلاح‌نامه یونسکو[2] (با ویرایش‌های مختلف) در خصوص اطلاعات علمی و فنی، و اصطلاح‌نامه اریک[3] در حوزه تعلیم و تربیت، و در ایران به زبان فارسی، به اصطلاح‌نامه فرهنگی فارسی (اصفا) و نیز اصطلاح‌نامه نظام مبادله اطلاعات (نما) و نیز اصطلاح‌نامه علوم اسلامی اشاره کرد.

اصطلاح‌نامه علوم اسلامی توسط پژوهشکده مدیریت اطلاعات و مدارك اسلامی پژوهشگاه علوم و فرهنگ اسلامی، وابسته به دفتر تبلیغات اسلامی حوزه علمیه قم، توسعه داده شده است. اهداف اصلی اصطلاح‌نامه علوم اسلامی عبارتند از تهیه طرحی از حوزه علوم اسلامی برای نشان دادن روابط منطقی میان مفاهیم و اصطلاحات و ترسیم ساختار آن‌ها، تهیه واژگان کنترل‌شده استاندارد برای حوزه معارف اسلامی، به منظور یکسان‌سازی مدخل‌های نظام ذخیره و بازیابی اطلاعات، تهیه نظام ارجاعات بین اصطلاحات مترادف و شبه مترادف برای کاربرد اصطلاحی واحد از مجموعه مترادفات موجود در شبکه اصطلاحات، تهیه راهنما برای محققان، نمایه‌سازان و استفاده‌کنندگان جهت انتخاب اصطلاح صحیح هنگام جستجوی موضوعات علوم و معارف اسلامی (یعقوب‌نژاد، ۱۳۷۵). از مهم‌ترین ویژگی‌های اصطلاح‌نامه علوم اسلامی، برقرار کردن روابط معنایی میان مفاهیم و اصطلاحات فقهی و غیرفقهی علوم اسلامی است. رابطه هم‌ارز (مترادف)، رابطه سلسله مراتبی (اعم و اخص، کل و جزء، مفهوم و مصداق و...) و وابسته به‌کار گرفته شده‌اند تا مجموع اصطلاحات هر رشته علوم اسلامی را از عام به خاص (حاکم به تابع) و با حفظ همه جنبه‌های اصلی، فرعی و وابسته به شکلی نظام‌دار نشان بدهند. هدف غایی آن است که تمام اصطلاحات حاکم و تابع و زیربخش‌های

1. Autonomous
2. Unesco Thesaurus
3. ERIC Thesaurus

منطقی آن‌ها در یک دید و نگاه کلی به نمایش درآید، به نحوی که جایگاه اصلی و جنبی تمام اصطلاحات در ساختار کلی، به همراه روابطی که با یکدیگر دارند، تعیین شود و هیچ اصطلاحی نباشد، مگر اینکه رابطه‌اش با اصطلاحات دیگر مشخص شده باشد (یعقوب‌نژاد، ۱۳۷۵).

نتایج بررسی‌ها حاکی از این بود که در بافت و زمینه پژوهش حاضر، نمونه پژوهشی کاملاً مشابه با پژوهش حاضر یافت نشد. پس از بررسی و مطالعه دقیق و در نظر گرفتن وجوه شمول در مطالعه، در نهایت پژوهش‌هایی که به‌نحوی با پژوهش حاضر از نظر موضوعی هم‌راستا بودند، شناسایی شده و در ادامه ارائه می‌گردند. این تحقیقات در زمینه‌های مرتبط با اصطلاح‌نامه‌ها و مباحث پیشرفته یادگیری عمیق هستند، شامل:

- تحقیقاتی که برای ساخت اصطلاح‌نامه در محیط‌های مختلف، روش‌های متفاوتی ارائه کرده‌اند. پژوهش‌های ناکایاما، هارات و نیشیو[1] (۲۰۰۷)، ایتو[2] و همکاران (۲۰۰۸)، تسنگ[3] (۲۰۰۲) گواه چنین پژوهش‌هایی هستند.
- تحقیقاتی که از اصطلاح‌نامه برای اهداف مختلف، ازجمله نمایه‌سازی خودکار، ساخت هستی‌شناسی[4] یا یافتن شباهت معنایی بهره برده‌اند، مانند پژوهش‌های ایوانز[5] و همکاران (۱۹۹۱)، جارماش و شپاکوویچ[6] (۲۰۰۳).
- تحقیقاتی که از مفاهیم «پردازش زبان طبیعی» یا «ان.ال.پی.»[7] در بازیابی اطلاعات و نمایه‌سازی خودکار استفاده کرده‌اند، مانند پژوهش‌های جینگ و کرافت[8] (۱۹۹۴)، اسمیتون[9] (۱۹۹۹)، استوکز[10] و همکاران (۲۰۰۸).
- تحقیقاتی که معطوف به مباحث پیشرفته یادگیری عمیق، مدل‌های پیش‌بین در شبکه‌های اجتماعی و روابط استنادی هستند، مانند پژوهش مک‌کوی[11] و همکاران (۲۰۲۱).

1. Nakayama, Harat & Nishio
2. Ito
3. Tseng
4. Ontology
5. Evans
6. Jarmasz and Szpakowicz
7. N.L.P.
8. Jing and Croft
9. Smeaton
10. Stokes
11. McCoy

از آنجایی که حدود یک یا دو دهه از عمر اصطلاح‌نامه‌های تدوین شده در ایران می‌گذرد، در سال‌های اخیر نیاز به روزآمدسازی آن‌ها کاملاً احساس شده و در همین راستا پژوهش‌های مبتنی بر روش‌هایی نوین مدنظر قرار گرفته‌اند (رجبی، حسینی بهشتی و صدیقی، ۱۳۹۸). بنابراین، تولید اصطلاح‌نامه‌ها و ارتقای روابط بین اصطلاحات، ازجمله مباحث مطرح در این زمینه بوده که همواره تحت تاثیر روش‌ها و تکنیک‌ها قرار گرفته است، مانند تلاش‌هایی که چانسانم[1] و همکاران (۲۰۲۱) برای اصطلاح‌نامه دیجیتالی داشته‌اند و یا هارکین[2] (۲۰۲۲) که با تکیه بر داده‌های پیوندی، درصدد ایجاد اصطلاح‌نامه بودند. در داخل کشور توجه به ساخت اصطلاح‌نامه با ابزارهای ماشینی، کمتر مورد توجه بوده و حداقل بررسی‌های پژوهشگر به نتایج خاصی نیانجامید. در پژوهش‌های مرور شده در این خصوص، استخراج اصطلاحات با استفاده از TF-IDF و سایر روش‌های سنتی مدنظر پژوهشگران قرار داشته که طبق نتایج بدست آمده از پژوهش‌ها، معمولاً روش‌های سنتی مطرح عملیاتی نبوده و نتوانسته‌اند اصطلاحاتی را که در متن نیامده، شناسایی نمایند، مانند آنچه در پژوهش ناکایاما، هارات و نیشیو (۲۰۰۷) بررسی شده است. در مجموع، مرور ادبیات پژوهش نشان می‌دهد که پژوهشگران به‌طور فزاینده‌ای بر استفاده و کاربرد شبکه‌های پیچشی گراف مشغول به فعالیت هستند و بروندادهای علمی فوق در زمینه‌های موضوعی گوناگون قابل رویت است.

از زمان توسعه GCN در سال ۲۰۱۶ میلادی، موارد استفاده از آن در مسائل مختلف، به شدت افزایش پیدا کرده و توانسته است برای یافتن روابط، در کاربردهای متفاوتی موفق و قوی ظاهر شود. برای مثال، پیش‌بینی روابط در علوم مرتبط با سلامت، تبلیغات هوشمند در صفحات اجتماعی، پیشنهاد دوستی‌های جدید در صفحات اجتماعی نظیر فیسبوک و اینستاگرام، شبکه‌های جریانی[3] فیلم نظیر نتفلیکس[4] و فیلیمو و غیره، تنها بخش کوچکی از کاربردهای این حوزه هستند. به طور مشخص، پیدا کردن روابط در رشته داروسازی پیشرفت‌های اساسی به‌همراه داشته است.

استفاده از مدل‌های مبتنی بر شبکه‌های پیچشی گراف، در زمینه‌های موضوعی گوناگون از حوزه پزشکی و درمان (جیانگ[5] و همکاران، ۲۰۲۰؛ یوآنیدیس[6] و همکاران، ۲۰۲۰؛ ساکای[7] و همکاران،

1. Chansan
2. Harkin
3. Stream (پخش آنلاین فیلم)
4. Netflix
5. Jiang
6. Ioannidis
7. Sakai

۲۰۲۱؛ لی[1] و همکاران، ۲۰۲۱؛ مک‌کوی[2] و همکاران، ۲۰۲۱؛ آنّو[3] و همکاران، ۲۰۲۲؛ وِن[4] و همکاران، ۲۰۲۲)، رفتار و احساسات (یو[5] و همکاران، ۲۰۲۰؛ ژو[6] و همکاران، ۲۰۲۰)، جغرافیای شهری و ترافیک (هووانگ[7] و همکاران، ۲۰۲۲؛ ژانگ و همکاران، ۲۰۱۸)، فناوری (دوو[8]، وان و شِن[9]، ۲۰۲۲؛ دینگ[10] و همکاران، ۲۰۲۲؛ فن[11]، سان[12] و روزین[13]، ۲۰۲۱) موجب ارتقای مدل‌های موجود و یا ایجاد مدلی نو با عملکردی مطلوب‌تر شده است. کریمی و همکاران (۲۰۲۰) در پژوهشی با عنوان «مدل‌های مولد عمیق مبتنی بر شبکه، برای طراحی ترکیبات دارویی، به عنوان مجموعه‌های گرافی»، حتی یک مرحله جلوتر رفته و با تخمین یک گراف با یال‌های وزن‌دار، ترکیبی از چند دارو را برای بیماری‌های جدید پیشنهاد کرده‌اند.

پیدا کردن روابط با استفاده از گراف، در کاربردهای ساده‌تری نیز موفق بوده است. از آن جمله می‌توان به شبکه‌های توزیع برق اشاره کرد که تلاش می‌کنند مدارهایی را بین نیروگاه‌های تولید و مصرف‌کننده، به صورت هوشمند برقرار نمایند. برای این موضوع می‌توان به تحقیق حسن‌زاده و همکاران (۲۰۱۹) با عنوان «رمزگذار خودکار تغییر گراف نیمه‌ضمنی»، و ژانگ و همکاران (۲۰۱۸) با عنوان «پیش‌بینی پیوند براساس شبکه‌های عصبی نمودار» اشاره کرد.

با پیاده‌سازی این مدل ماشینی، علاوه‌بر اینکه توانایی کاهش هزینه روزآمد نگه‌داشتن اصطلاح‌نامه را ایجاد کرده‌ایم، توانمندی بهینه‌سازی اصطلاح‌نامه به عنوان یك سامانه خود-اتکاء را نیز در خلال زمان ارتقاء می‌دهیم. امروزه شاهد بهره‌گیری از پردازش زبان طبیعی و یادگیری ماشینی جهت ارتقای مدل‌ها در زمینه‌های مختلف هستیم، اما به نظر می‌رسد علی‌رغم توجه ویژه به پردازش زبان طبیعی و یادگیری ماشینی، استفاده از این ابزارها در حوزه توسعه اصطلاح‌نامه‌ها و حتی هستی‌شناسی‌ها، آنطور

1. Lee
2. McCoy
3. Anno
4. Van
5. You
6. Zhou
7. Huang
8. Doo
9. Shen
10. Ding
11. Fan
12. San
13. Rosin

که باید مورد توجه قرار نگرفته است. با توجه به محدودیت‌های ساخت دستی اصطلاح‌نامه، همواره تلاش‌هایی در زمینه کشف روش‌ها و راه‌های نوین، به‌خصوص در فعالیت ماشینی و ابزاری، خصوصاً در محیط‌هایی با داده‌های حجیم، مورد توجه بوده است. در این پژوهش، داده‌های موجود در اصطلاح‌نامه، به صورت یک گراف در نظر گرفته می‌شوند. اصطلاحات به‌عنوان رئوس، و ارتباط بین اصطلاحات به‌عنوان یال‌های این گراف خواهند بود و این گراف به عنوان ورودی، به شبکه پیچشی گرافی[1] داده خواهد شد، تا مدلی برای استنتاج خودکار ارتباطات حاصل شود.

۲. مبانی نظری

پردازش زبان طبیعی شاخه‌ای از هوش مصنوعی است که بر تعامل بین دانش داده‌ها و زبان انسان تمرکز دارد و به ماشین‌ها توانایی خواندن، درک و استخراج معانی از زبان انسان را می‌دهد (ویلکس، ۱۹۹۶). امروزه این شاخه از هوش مصنوعی به دلیل پیشرفت‌های زیاد در دسترسی به داده‌ها و افزایش قدرت محاسباتی، می‌تواند به انسان در انجام کارهای زیادی کمک کند و زمینه‌های کاربردی آن، روزانه افزایش می‌یابد. به طور مثال:

- برنامه‌های ترجمه ماشینی از قبیل ماشین ترجمه گوگل،[2]
- پردازشگرهای متن مانند مایکروسافت Word و گرامر که از پردازش زبان طبیعی، برای بررسی دقت گرامری متن استفاده می‌کنند،
- برنامه‌های دستیار شخصی مانند Siri، Cortana، Alexa که رابط‌های صوتی هوشمند هستند و از پردازش زبان طبیعی برای پاسخ دادن به پیام‌های صوتی استفاده می‌کنند و هر کاری را مانند پیدا کردن یک مغازه خاص، پیش‌بینی آب و هوا، و یافتن بهترین مسیر انجام می‌دهند.

پردازش زبان طبیعی مستلزم استفاده از الگوریتم‌هایی برای شناسایی و استخراج کردن قوانین مربوط به زبان طبیعی انسان‌ها است، به طوری که داده‌های زبانی بدون ساختار و قائده به شکلی تبدیل شوند که کامپیوترها بتوانند آن را درک نمایند.

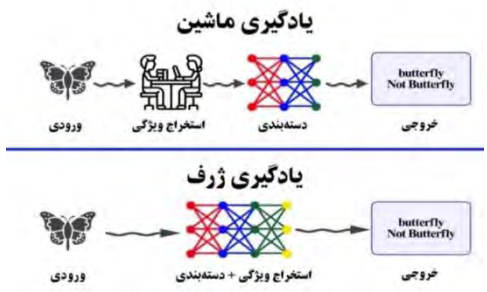
یادگیری ماشین[3] شاخه‌ای از هوش مصنوعی است که با استفاده از یک‌سری الگوریتم‌های استنتاج‌گرا به ماشین کمک می‌کند که قواعدی را فراتر از قواعد محتوایی و یا صوری آشکار کشف، و براساس آنها عمل کند. یادگیری ماشینی مطالعه علمی الگوریتم‌ها و مدل‌های آماری مورد استفاده

1. Graph Convolutional Networks
2. https://blog.google/products/translate/ten-years-of-google-translate
3. Machine learning

سیستم‌های کامپیوتری است که به جای استفاده از دستورالعمل‌های واضح، از الگوها و استنباط برای انجام وظایف سود می‌برند (گوریل[1] و همکاران، ۲۰۱۹). یادگیری ماشین، استخراج یک مدل از روی داده‌های رقومی است. مدل به دست آمده می‌تواند برای پیش‌بینی و یا به منظور استخراج دانش از داده‌ها به‌کار آید. تکنولوژی‌های ML اثرات اجتماعی عظیمی را در محدوده وسیعی از کاربردها مثل بینایی ماشین، پردازش صوت، فهم زبان طبیعی، علوم عصبی، سلامتی و اینترنت اشیاء به وجود آورده است. پیدایش کلان‌داده، گرایش‌های زیادی را در ML ایجاد کرده است (باغبانی، ۱۳۹۶). یادگیری ماشینی به سه گروه یادگیری با نظارت[2] (وظیفه‌محور)، یادگیری بدون نظارت[3] (داده‌محور) و یادگیری تقویت شده[4] (آزمون و خطا) تقسیم می‌شود. یادگیری با نظارت، زمانی اتفاق می‌افتد که شما با استفاده از داده‌های برچسب‌گذاری شده، به یک ماشین آموزش می‌دهید. به بیان دیگر، در این نوع یادگیری، داده‌ها از قبل با پاسخ‌های درست یا همان نتیجه برچسب‌گذاری شده‌اند. در یادگیری بدون نظارت، ماشین با استفاده از داده‌هایی آموزش می‌بیند که هیچ‌گونه برچسب‌گذاری بر روی آن‌ها انجام نشده است. در این روش به الگوریتم یادگیری گفته نمی‌شود که داده‌ها نمایانگر چه چیزی هستند. در یادگیری تقویت‌شده نیز مانند یادگیری بدون نظارت، داده‌های مورد استفاده برای یادگیری، برچسب‌گذاری نمی‌شوند. زمانی که پرسشی برای داده‌ها مطرح شد، نتیجه آن درجه‌بندی می‌شود.

یادگیری عمیق[5] نیز شاخه‌ای از بحث یادگیری ماشین و هوش مصنوعی است، که در تلاش برای مدل کردن مفاهیم انتزاعی سطح بالا، با استفاده از یادگیری در سطوح و لایه‌های مختلف می‌باشد که این فرایند با استفاده از یک گراف عمیق که دارای چندین لایه پردازشی متشکل از چندلایه تبدیلات خطی و غیرخطی است، انجام می‌گیرد. امروزه با ظهور پردازنده‌های گرافیکی، فرایند یادگیری شبکه‌های ژرف، بسیار سریع‌تر شده است. نتایج پژوهش‌های مختلف نشان می‌دهد که یادگیری عمیق در مقایسه با روش‌های یادگیری ماشین کلاسیک، عملکرد بسیار بالایی دارند و می‌توانند برای بسیاری از کارها مورد استفاده قرار بگیرند. بزرگ‌ترین مزیت یادگیری ژرف را می‌توان استخراج و یادگیری خودکار ویژگی‌ها دانست که در شکل (۱) نیز به آن اشاره شده است. در شکل (۱) تفاوت یادگیری ماشین و عمیق یا ژرف نمایش داده شده است.

1. Gowril
2. Supervised
3. Unsupervised
4. Reinforced
5. Deep Learning



شکل ۱- مقایسه یادگیری ماشین و یادگیری ژرف

شبکه‌های عصبی[1] مجموعه‌ای از نورون‌ها هستند که از الگوریتم‌های منحصربه فردی پیروی می‌کنند. این مجموعه که از مغز انسان الگوبرداری و الهام گرفته شده است، با هدف شناسایی الگوها طراحی می‌شوند و مورد استفاده قرار می‌گیرند. به طور کلی می‌توان گفت که شبکه عصبی شامل الگوریتم‌هایی برای یادگیری ماشین است که منجربه طبقه‌بندی کردن داده‌های ورودی و ارائه خروجی مطلوب می‌گردد. به همین دلیل است که می‌توان شبکه‌های عصبی را به عنوان جزئی از فرایند یادگیری ماشین در نظر گرفت. در یادگیری ماشین، یک شبکه عصبی پیچشی،[2] نوعی از شبکه عصبی مصنوعی است که در آن نورون‌ها، به نواحی روی هم افتاده در یک ناحیه دیداری واکنش نشان می‌دهند. این نوع شبکه‌ها از فرایندهای بیولوژیکی الهام گرفته شده‌اند و گونه‌هایی از شبکه‌های پرسپترون چند لایه‌اند که با طراحی خاصی از حداقل میزان پیش‌پردازش بهره می‌برند (حسن‌پور متی کلایی و همکاران، ۱۳۹۵).

یک گراف، نمایشی تصویری از مجموعه اشیائی است که با هم ارتباط دارند. هر یک از این اشیاء را «رأس» یا گره می‌نامند. رأس‌ها نیز از طریق «یال»ها یا لبه‌ها با هم مرتبط هستند. گراف G را به‌صورت G=(V,E) تعریف می‌کنند که در آن، V مجموعه رئوس و E مجموعه یال‌های آن می‌باشند. اصطلاح شبکه‌های عصبی گرافی[3] با بهره‌گیری از رویکرد یادگیری عمیق، در تجزیه و تحلیل گراف‌ها مورد استفاده قرار می‌گیرد. از آنجایی‌که GNNها به طور طبیعی توانایی سازگاری با دیتاست‌های دنیای واقعی با ساختار گرافی را دارند، در حوزه‌های مختلفی ازجمله تبلیغات اینترنتی، پیش‌بینی ترافیک جاده، طراحی دارو و غیره مورد استفاده قرار می‌گیرند. شبکه‌های پیچشی گرافی بر پایه شبکه‌های

1. Nueral Network
2. ConvNet یا CNN به اختصار
3. Graph Neural Networks

عصبی پیچشی شکل گرفته‌اند و دوباره آن‌ها را برای گراف‌ها تعریف می‌کنند. هدف چارچوب‌های پیچشی این است که همانند تصاویر، اطلاعات همسایه را برای رئوس گراف نیز ذخیره کنند.

شبکه‌های پیچش گرافی، یک نوع از شبکه‌های عصبی گرافی است و عمل پیچش را از داده‌های مشبک[1] به داده‌های گرافی تعمیم می‌دهد. ایده اصلی این است که نمایش یک گره براساس تجمیع[2] ویژگی‌های آن گره و همسایه‌هایش محاسبه شود. پیچش گرافی[3] یک سیگنال x با فیلتر $\boldsymbol{g_\theta} \in R^n$ n به صورت زیر تعریف می‌شود:

$$\boldsymbol{x} *_G \boldsymbol{g_\theta} = U((U^T\boldsymbol{x}) \odot (U^T\boldsymbol{g_\theta})), \quad (۱)$$

که در آن $U = u_0 ; u_1 ; \dots u_{n-1}] \in \mathbb{R}^{n \times n}$ ماتریس بردارهای ویژه گراف لاپلاسین است. U^Tx تبدیل فوریه گرافی سیگنال x است و $\odot$ ضرب عنصر به عنصر است. تفاوت اصلی بسیاری از شبکه‌های عصبی گرافی در فیلتر مورد استفاده آنها است. بر این اساس، Chebnet از فیلتر چندجمله‌ای چبیشف ماتریس قطری مقادیر ویژه به صورت زیر استفاده می‌کند:

$$\boldsymbol{g_\theta} = \sum_{i=0}^{k} \theta i Ti(\Lambda) \quad (۲)$$

که در آن، $\tilde{\boldsymbol{\Lambda}} = \frac{2\Lambda}{\lambda max} - I_n$، Λ ماتریس قطری مقادیر ویژه بوده و بزرگ‌ترین مقدار ویژه λmax است. به این ترتیب فیلتر سیگنال x به صورت زیر خواهد بود:

$$\boldsymbol{x} *_G \boldsymbol{g_\theta} = \sum_{i=0}^{k} \theta i Ti(L)\, xi \quad (۳)$$

که در آن، $L = \frac{2l}{\lambda max} - I_n$ و L ماتریس لاپلاسین است. در مقابل GCN، K را برابر ۱، $\lambda max = 2$ و $\boldsymbol{\theta} = \boldsymbol{\theta}_0 = -\boldsymbol{\theta}_1$ در نظر می‌گیرد. بر این اساس، فرمول GCN به صورت زیر خواهد بود:

$$\boldsymbol{H^{(l+1)}} = \boldsymbol{f(H^{(l)}, A)} = \boldsymbol{g}\,(\boldsymbol{D}^{\frac{1}{2}}\tilde{\boldsymbol{A}}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{H^{(l)}}\boldsymbol{\theta}^{(l)}), \quad (۴)$$

که در آن، $\boldsymbol{A}$ ماتریس مجاورت گراف بوده که ماتریس همانی به آن افزوده شده است. علت این عملیات ایجاد خودحلقه است که در عمل سبب می‌شود در هنگام ایجاد بازنمود، از ویژگی‌های گره هدف نیز استفاده شود. $\boldsymbol{H^{(l)}}$ حاصل فعال‌سازی لایه l، g تابع ReLU و $\boldsymbol{D}_{ii} = \sum_i Aij$ است.

۳. روش پژوهش

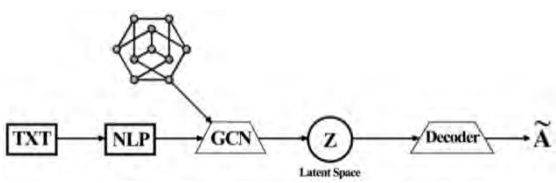
در پژوهش حاضر، ابزار مورد استفاده، روش گراف است که یکی از روش‌های پرکاربرد در حوزه

1. Grid data
2. Aggregation
3. Graph convolution

یادگیری ماشین می‌باشد. برای کشف روابط و راستی‌آزمایی روش بکار گرفته شده، از داده‌های موجود در اصطلاح‌نامه علوم اسلامی به‌صورت عینی استفاده شد. داده‌های مورد استفاده از اصطلاح‌نامه علوم اسلامی[1] که از سال ۱۳۷۲ تا ابتدای سال ۱۴۰۰ توسط متخصصان موضوعی ایجاد و نگهداری می‌شوند، گرفته شده است.

مسأله استنتاج روابط بین اصطلاحات را می‌توان به مسأله پیش‌بینی لینک در یک گراف کاهش داد. مسأله پیش‌بینی لینک از مسائل قدیمی شناخته شده است که روش‌های مختلفی برای حل آن‌ها تاکنون ارائه شده است. در این حالت اصطلاحات به عنوان رئوس گراف و روابط بین اصطلاحات به عنوان یال‌های بین رئوس مدل می‌شوند. سپس توسط یک رویکرد یادگیری ماشینی و با دادن بخشی از این گراف، تلاش می‌شود ماشین الگوهای لازم برای وجود یا عدم وجود یال بین رئوس را شناسایی و یاد بگیرد.

شبکه‌های پیچشی گرافی یکی از مهم‌ترین روش‌های مطرح در حوزه یادگیری عمیق بوده و قادر هستند که در کنار توجه به ویژگی‌های هر گره، از الگوهای روابط در گراف نیز سود بجویند. به این منظور اصطلاحات و مستند حاوی نمایه‌های آنها توسط یک مدل تعبیه‌ساز، به بردارهای معنایی تبدیل خواهند شد، که در آن نزدیکی دو بردار به معنای نزدیکی مفاهیم دو متن ورودی می‌باشد. این بردارها مقادیر اولیه رئوس گراف را تشکیل خواهند داد. در هر مرحله از آموزش مدل، با در نظر گرفتن دو بردار ورودی و همچنین تجمیع بردار همسایگان، هر گره تلاش خواهد کرد وجود یا عدم وجود یال بین آن دو گره را پیش‌بینی کند. مدل به واسط روش انتشار رو به عقب آموزش داده شده، تا تابع هزینه آن کمینه شود. مراحل انجام کار در روش گراف در شکل (۲) به تصویر کشیده شده است.



شکل ۲- مراحل انجام کار در گراف

شبکه‌های عصبی گرافی نخستین بار در سال ۲۰۰۴ میلادی مطرح شدند (کرامتفر[2] و همکاران، ۲۰۲۱). شبکه‌های عصبی گرافی دسته‌ای از شبکه‌های عصبی دارای ساختار گراف هستند. ساختار

1. https://thesaurus.isca.ac.ir
2. Honorable

یک شبکه عصبی گرافی مبتنی بر یک گراف $g(v,\varepsilon)$ است که از مجموعه رئوس v و مجموعه یال‌های ε تشکیل شده است. این شبکه یک گراف دریافت کرده که در آن هر گره با یک بردار ویژگی x_i همراه بوده و هر یال نیز می‌تواند چنین باشد. ساختار گراف، به‌روزرسانی‌های گذر پیام که در رابطه (۱) مشخص شده را تعیین کرده که به نوبت انجام شده تا نمایش‌های نهایی گره‌ها یال‌ها حاصل شود:

$$h_{(i,j)} = f_{edge}(h_i, h_j, x_{(i,j)}) \qquad (۱)$$

$$h'_i = f_{node}(h_i, \sum_{j\in Ni} h_{(j,i)}, x_i)$$

N_i مجموعه همسایه‌های i با یک یال وارد شونده به آن و h_i بازنمود گره i است. f_{edge} و f_{node} معمولاً شبکه‌های عصبی ساده چندلایه هستند که اتصال آرگومان‌های تابع را به عنوان ورودی دریافت می‌کنند. می‌توان چندین لایه از این دست را روی هم قرار داد و نمایش سطح بالاتری از گره‌ها به دست آورد. این نوع از شبکه عصبی گرافی توسط گیلمر[1] و همکاران (۲۰۱۷) ارئه شد و می‌توان آن را به عنوان مدل عام‌تری از بسیاری از انواع شبکه‌های عصبی گرافی نظیر شبکه پیچشی گرافی دانست (Kipf, 2020). در شبکه‌های پیچشی گرافی، عمل پیچش از داده‌های مشبک،[2] به داده‌های گرافی تعمیم داده می‌شود. ایده اصلی این است که نمایش یک گره براساس تجمیع[3] ویژگی‌های آن گره و همسایه‌هایش محاسبه شود.

مدل‌های ساخته شده با هریک از الگوریتم‌ها، باید با روش مناسبی مورد ارزیابی واقع شوند، تا دقت و عملکرد مدل موردنظر سنجیده شود و بتوان مدل‌ها را براساس این معیارها مورد مقایسه قرار داد. کلاس‌های پیش‌بینی شده بر روی مدل و کلاس واقعی داده‌ها را می‌توان در جدول (۱) که به آن ماتریس درهم‌ریختگی می‌گویند، مشاهده کرد.

جدول ۱- ماتریس درهم‌ریختگی

		برچسب پیش‌بینی شده	
		مثبت	منفی
برچسب شناخته شده	مثبت	TP	FN
	منفی	FP	TN

هریک از مقادیر موجود در جدول (۱) این‌گونه تفسیر می‌شوند:

TP[4]: این مفهوم به معنای تعداد موارد دسته‌بندی شده مثبت صحیح هست.

1. Gilmer
2. Grid data
3. Aggregation
4. True Positive

TN[۱]**:** این مفهوم به معنای تعداد موارد دسته‌بندی شده منفی صحیح هست.

FP[۲]**:** این مفهوم به معنای تعداد نمونه‌هایی است که به صورت اشتباه به عنوان داده‌های مثبت دسته‌بندی شده‌اند.

FN[۳]**:** این مفهوم به معنای تعداد نمونه‌هایی است که به صورت اشتباه به عنوان داده‌های منفی دسته‌بندی شده‌اند.

Rate TP[٤]**:** این مفهوم به معنای تعداد پیش‌بینی صحیح هر کلاس نسبت به تعداد نمونه‌های همان کلاس است.

Rate FP[٥]**:** این مفهوم برابر است با نسبت تعداد نمونه‌هایی که منفی بوده و اشتباهاً مثبت برچسب‌گذاری شده، به کل نمونه‌هایی که مدل یادگیری آنها را منفی برچسب‌گذاری نموده است، که از طریق فرمول (۵) محاسبه می‌شود.

$$FP = \frac{FP}{FP+TN} \qquad (۵)$$

جامعیت:[٦] نسبت میزان نمونه‌هایی که به درستی جزء نمونه‌های مثبت دسته‌بندی شده‌اند، به کل نمونه‌های مثبت واقعی است، که محاسبه آن به شکل فرمول (۶) است.

$$Recall\ (sencivity) = Hit\ Rat = \frac{TP}{TP+FN} * 100 \qquad (۶)$$

مانعیت:[۷] نسبت میزان نمونه‌هایی که به درستی جزء نمونه‌های مثبت دسته‌بندی شده‌اند، به کل نمونه‌هایی که ماشین یادگیری به عنوان مثبت در نظر گرفته است، که نحوه محاسبه آن به شکل فرمول (۷) است:

$$Precision = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} * 100 \qquad (۷)$$

دقت:[۸] دقت مجموعه را معین می‌کند، و طبق فرمول (۸) محاسبه می‌شود.

$$Acuuracy = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP}+\mathrm{TN}+\boldsymbol{FP+FN}} * 100 \qquad (۸)$$

1. True Negative
2. False Positive
3. False Negative
4. True Positive Rate
5. False Positive Rate
6. Recall
7. Percision
8. Accurecy

این فاکتور، یک فاکتور بسیار مهم در سنجش کارایی سیستم به‌شمار می‌آید.

نمره F[1]: میانگین هارمونیک جامعیت و مانعیت است که طبق فرمول (۹) محاسبه می‌شود:

$$F - score = \frac{2*Recall*Precision}{Recall+Precision} * 100 \qquad (۹)$$

معیار AP[2]: معیار میانگین دقت، یکی از مهم‌ترین معیارهایی است که برای سنجش دقت مدل‌های شبکه‌های عصبی پیچشی گرافی استفاده می‌شود. AP برابر با میانگین Precision برای Recallهای مختلف بین ۰ و ۱ است. برای محاسبه Average Precision، ابتدا نمودار PR کمی درون‌یابی و نرم می‌شود. سپس میانگین مقدار Precision به ازای مقادیر recall محاسبه می‌گردد. معیاری برای خلاصه کردن عملکرد یک دنباله رتبه‌بندی شده از نتایج است، که با در نظر گرفتن میانگین مقادیر precision مربوط برای هر نتیجه، محاسبه می‌شود.

میانگین دقت، در مسائلی کاربرد دارد که داده‌های نابرابر استفاده می‌شوند و معیار ACCURACY در این نوع مسائل خوب نیست و نمی‌تواند مدل را به خوبی ارزیابی نماید. در جایی که داده‌ها متوازن باشند، معیار ACCURACY کاربرد دارد. در مسائل دوکلاسه، در واقع Precision تعداد درست‌های گفته شده مدل، به تعدادی که در مجموع مدل پیش‌بینی کرده است، می‌باشد. در نتیجه، در صورتی که با داده‌های نامتوازن سروکار داریم، معیار Precision بهتر از معیار ACCURACY است.

در پژوهش حاضر، تعداد داده‌های موجود نامتناسب بوده و تعداد صفرهای ماتریس گراف موجود و در واقع تعداد حالت‌هایی که یال ندارند، زیاد است. در اینجا Precision، تعداد رابطه‌های درست پیش‌بینی شده توسط مدل، تقسیم بر تعداد کلی که مدل از رابطه‌ها پیش‌بینی می‌کند، است. اما در پژوهش موجود، خروجی مدل، بودن یا نبودن رابطه را پیش‌بینی نمی‌کند، بلکه یک احتمال از وجود یک رابطه را در خروجی می‌دهد و در اینجا مفهومی به نام THRESHOLD وجود دارد که میزان درصد موردنظر در احتمال ارائه شده توسط مدل، برای وجود رابطه می‌باشد. در نهایت در معیار AP، مقادیر احتمال، یک درصد یک درصد افزایش پیدا کرده و به‌ازای هر کدام Precision محاسبه و در نهایت میانگین مقادیر Precision به عنوان مقدار AP بدست می‌آید.

معیار ROC: ROC یك اصطلاح در جنگ جهانی دوم برای ارزیابی عملکرد رادار است. امتیاز این معیار اطلاعاتی را در مورد اینکه یك مدل، کار خود در جداسازی موارد را به خوبی انجام می‌دهد، به ما ارائه می‌دهد. در واقع ROC مبادله دو مفهوم حساسیت و خاصیت می‌باشد. منحنی ROC توسط

1. F-score
2. Average Precision

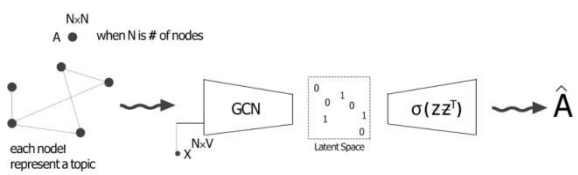
ترسیم نسبت یا «نرخ مثبت صحیح»[1] که به اختصار TPR نامیده می‌شود، برحسب «نرخ مثبت کاذب»[2] با نام اختصاری FPR، ایجاد می‌شود. TPR مشخص می‌کند که به چه نسبتی پیش‌بینی صحیح صورت گرفته است. یعنی تعداد پیش‌بینی‌های صحیح بر تعداد نتایج مثبت واقعی تقسیم شده و نرخ پیش‌بینی صحیح مثبت محاسبه می‌شود. از طرف دیگر، FPR نشانگر تعداد شناسایی‌های مثبت از میان مشاهدات منفی است. این نسبت نیز به عنوان نرخ مثبت کاذب در معیار ROC به‌کار می‌رود. در ادامه نحوه محاسبه Sensitivity و Specificity در فرمول (۱۰ و ۱۱) آمده است:

$$\text{Sensitivity} = (\text{TPR}) = TP / (TP+FN) \quad (۱۰)$$

$$\text{Specificity} = (\text{TNR}) = TN / (TN+FP) \quad (۱۱)$$

در نهایت، امتیازی که ROC نشان می‌دهد، به این معنی است که یک مدل چقدر احتمال دارد بتواند دو کلاس را از هم تشخیص و تمیز دهد.

روش کار بدین صورت خواهد بود که ابتدا داده‌ها برای اینکه بتوانند در گراف مورد استفاده قرار گیرند، تمیز می‌شوند. همچنین متون با استفاده از الگوریتم‌های پردازش زبان طبیعی، به بردار تبدیل شده و گراف بر روی آن‌ها اعمال شده و خروجی‌های بدست آمده در این مرحله، به عنوان ورودی‌های الگوریتم شبکه‌های پیچشی گرافی[3] استفاده خواهند شد، سپس در یك فضای نهفته[4]، یك خروجی گراف حاصل خواهد شد که خروجی نهایی بدست آمده، مورد بررسی و ارزیابی قرار گرفته و در پایان مدل بدست آمده، تحلیل می‌شود. در شکل (۳) مراحل کلی پژوهش به نمایش گذاشته شده است:



شکل ۳- مراحل اجرایی پژوهش

X is Attributes matrix
A is adjacency matrix
N is # of nodes
V is vector dimention of attributes
each node is related to multiple documents
we can provid a vector whith length V for each node

1. True Positive Rate
2. False Positive Rate
3. GCN
4. Latent Space

۴. یافته‌ها

یک شبکه عصبی شامل یک لایه ورودی و یک لایه خروجی و تعدادی لایه مخفی است. از طریق لایه ورودی، داده‌ها وارد شبکه عصبی شده و به وسیله لایه خروجی، خروجی تولید می‌گردد. در میان این دو لایه، تعدادی لایه دیگر وجود دارند که باعث افزایش پیچیدگی شبکه و بالا رفتن دقت مدل می‌شوند که به آن‌ها لایه‌های مخفی[1] گفته می‌شود. مدل موجود، مدل نیمه نظارت شده[2] بوده که مدلی است که نه‌تنها از داده‌های دارای برچسب، بلکه از داده‌های بدون برچسب نیز استفاده می‌کند. در پژوهش حاضر، مدل موجود علاوه‌بر استفاده از داده‌های ورودی، قابلیت پیش‌بینی تعدادی از یال‌هایی که می‌توانند بین دو رأس وجود داشته باشند را دارد و از آن‌ها هم در روند آموزش استفاده شده است. در واقع مدل موجود، قابلیت پیش‌بینی تعداد یال‌هایی که از قبل وجود نداشته‌اند و در دیتاست ما نبوده‌اند را دارد و می‌تواند روابط معنایی را پیش‌بینی کند.

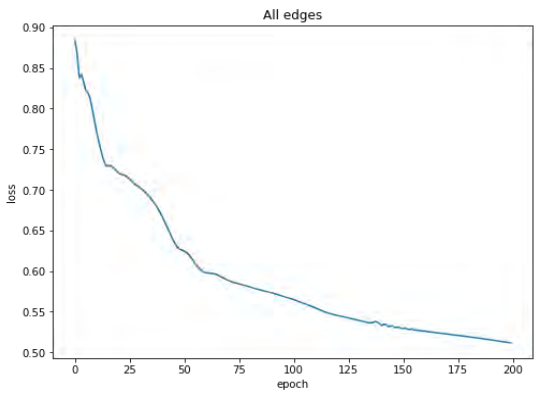
در ابتدا پارامترهایی نظیر learning rate, epoch تعداد نورون‌های لایه‌های پنهان[3] و مقدار dropout تعیین شده و مقادیر دیتا به سه بخش تست[4]، آموزش[5] و ارزیابی[6] تقسیم شد. ۱۰ درصد از داده‌ها به عنوان داده‌های تست و ۵ درصد از داده‌ها به عنوان داده‌های ارزیابی و ۸۵ درصد از داده‌ها برای آموزش[7] مورد استفاده قرار گرفته‌اند. مدل را با استفاده از کلاس‌ها و توابع موجود در پایتون ساخته و سپس روال آموزش انجام شد و با استفاده از متریک‌های average_precision_score و ROC مدل ارزیابی شد. لازم به ذکر است که روند آموزش به دو صورت انجام شده است. ابتدا بدون در نظر گرفتن نوع روابط معنایی و صرفاً وجود یک ارتباط معنایی، مدل آموزش داده و ارزیابی انجام شد. در مرحله دوم به تفکیک، هر نوع ارتباط مورد بررسی قرار گرفت و روابط به صورت خاص به مدل آموزش داده شده و ارزیابی انجام شد. تست‌ها با مقدار learning rate = 0.01 و تعداد epoch = 200 انجام شده است. به طور کلی در بررسی روابط، به طور منحصربه فرد دقت مدل کاهش پیدا کرده است.

1. Hidden
2. Semi supervised
3. Hidden
4. Test
5. Train
6. Validation
7. Train

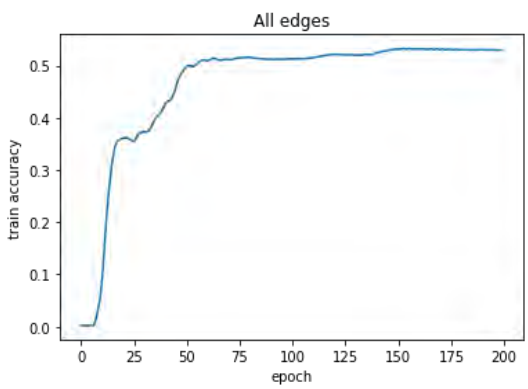
جدول ۲- نتایج آزمایشات

	node number	edge number	epoch	learning rate	test AP Score	test ROC Score	train Accuracy
تمام ارتباط‌ها	1282	1364	200	0.01	0.75	0.72	0.52
هم‌ارزی	145	78	200	0.01	0.55	0.42	0.52
سلسله مراتبی	745	612	200	0.01	0.75	0.69	0.539
ر.ک	37	20	200	0.01	0.83	0.75	0.67
نیز.ر.ک	136	101	200	0.01	0.87	0.83	0.54
وابستگی	805	610	200	0.01	0.68	0.65	0.54

تعداد رئوس در گراف ۱۲۸۲ و تعداد یال‌ها ۱۳۶۴، در حالتی که تمام یال‌ها در نظر گرفته شده‌اند، می‌باشد. همان‌طور که در نمودار (۱) مشخص است، مقدار loss در روند آموزش کاهش پیدا کرده است.



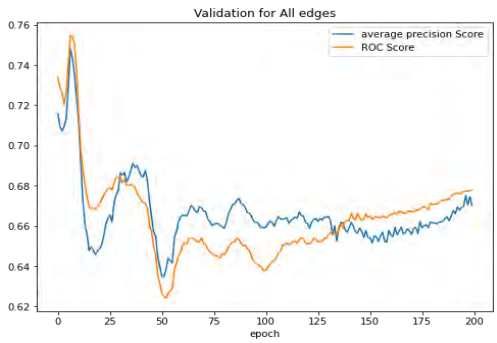
نمودار ۱ – مقدار loss



نمودار ۲ – مقدار دقت

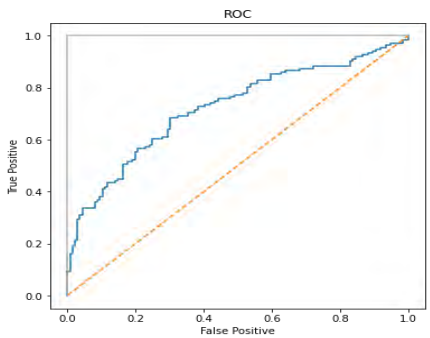
با توجه به نمودار (۲)، مقدار دقت مدل در روند آموزش افزایش پیدا کرده و از مقدار صفر به دقت حدود ۵۰ درصد افزایش یافته است.

نمودار (۳)، مقادیر ROC و average precision برای داده‌های validation در روند آموزش را به تصویر کشیده است. براساس این نمودار، ابتدا مقادیر در ۲۰ epoch اول بسیار بالا هستند و در ادامه آموزش، مقادیر کاهش پیدا کرده و نهایتاً مقدار ROC به صورت متناوب افزایش داشته است.



نمودار ۳- مقادیر ROC و average precision

نمودار (۴)، نتایج مقدار ROC روی داده‌های تست را نمایش می‌دهد. مقدار ROC در اینجا ۷۲% می‌باشد.



نمودار ٤ – مقادیرROC

۵. نتیجه‌گیری

در پژوهش حاضر با استفاده از داده‌هایی که شامل روابطی بین اصطلاحات است که به وسیله متخصصان حوزه دانشی ایجاد شده و همچنین مدارکی که بر پایه این اصطلاحات نمایه شده‌اند، و نیز با بهره‌برداری از الگوریتم‌ها و تکنیک‌های مطلوب در پردازش زبان طبیعی و یادگیری ماشین، روابط

پیشین و پسین میان اصطلاحات را به صورت خودکار اصلاح و کشف شد و به نوعی فرایند توسعه و تکامل اصطلاح‌نامه به خودش واگذار گردید.

نتایج بدست آمده حاکی از میانگین دقت ۷۵ درصد در کل روابط و از بین انواع ارتباط‌ها، میانگین دقت در نوع رابطه سلسه نیز. «ر.ک.» ۸۷ درصد، ر.ک ۸۳ درصد، سلسله مراتبی ۷۵ درصد، وابستگی ۶۸ درصد و هم‌ارزی ۵۵ درصد می‌باشد. همچنین در معیار راک مقدار در کل روابط ۷۲ درصد بوده و از بین انواع ارتباط‌ها، مقدار «راک» در نوع رابطه سلسه نیز «ر.ک.» ۸۳ درصد، «ر.ک.» ۷۵ درصد، سلسله مراتبی ۶۹ درصد، وابستگی ۶۵ درصد و هم‌ارزی ۴۲ درصد است. این اعداد با توجه به استفاده از الگوریتم‌های یادگیری عمیق برای اولین بار در حوزه علوم انسانی و اصطلاح‌نامه‌ها، بسیار خوب می‌باشد و حتی در تعیین ارتباط کلی بین اصطلاحات، می‌تواند در مرحله قبل از تحلیل انسانی، به عنوان یک سیستم پیشنهاددهنده، آغازکننده خوبی در کشف ارتباط باشد.

از آنجا که حدود یک یا دو دهه از عمر اصطلاح‌نامه‌های تدوین‌شده در ایران می‌گذرد، در سال‌های اخیر نیاز به روزآمدسازی آن‌ها کاملاً احساس شده و می‌بایستی از شیوه‌های نو و جدیدی که در عرصه‌های دیگری استفاده شده و نتایج خوبی را به همراه داشته، در فرایند توسعه اصطلاح‌نامه‌ها بهره برد. بنابراین، تولید اصطلاح‌نامه‌ها و ارتقای روابط بین اصطلاحات ازجمله مباحث مطرح در این زمینه بوده، که همواره تحت تاثیر روش‌ها و تکنیک‌ها قرار گرفته است. در داخل کشور، توجه به ساخت اصطلاح‌نامه با ابزارهای ماشینی، کمتر مورد توجه بوده و حداقل بررسی‌های پژوهش حاضر به نتایج خاصی نیانجامید. در پژوهش‌های پیشین، استخراج اصطلاحات با استفاده از TF-IDF و سایر روش‌های سنتی، مدنظر پژوهشگران قرار داشته که طبق نتایج بدست آمده از پژوهش‌ها، معمولاً روش‌های سنتی مطرح عملیاتی نبوده و نتوانسته‌اند اصطلاحاتی را که در متن نیامده، شناسایی نمایند؛ مانند آنچه در پژوهش ناکایاما، هارات و نیشیو (۲۰۰۷) آمده است. تمرکز پژوهش‌های انجام شده بیشتر بر تکنیک TF-IDF بوده است.

علی‌رغم چرخش نظر از اصطلاح‌نامه‌ها به هستی‌شناسی‌ها، هنوز هم استفاده از اصطلاح‌نامه‌ها، مخصوصاً در کشور ایران مورد توجه است. اصطلاح‌نامه علوم اسلامی، نظر به جایگاه و اهمیتی که مجموعه علوم اسلامی در ساخت و توسعه چارچوب فکری و عملی نظام جمهوری اسلامی دارد، به آن توجه ویژه‌ای شده است و از طرفی، مهم‌ترین موضوع در تدوین و توسعه هر اصطلاح‌نامه، استنتاج روابط بین اصطلاحات است و خودکارسازی این فرایند، سهم موثری در بقاء و توسعه اصطلاح‌نامه‌ها دارد. افزایش سرعت، کاهش هزینه‌های مربوط به نیروی انسانی، افزایش دقت و جامعیت، در نهایت

باعث افزایش پوشش مستندات و بهبود فرایند بازیابی اطلاعات خواهد شد.

در انتهای عملیات خودکارسازی، مطابقت فرآورده ماشین با تزاروس توسعه داده شده به وسیله انسان، کیفیت را تضمین می‌کند. همچنین در بررسی‌ها، پژوهش‌هایی مشاهده شد که ساخت هستی‌شناسی با استفاده از اصطلاح‌نامه ویکی پدیا را مدنظر قرار داده و این نشان‌دهنده کاربردهای اصطلاح‌نامه‌ها در ساخت هستی‌شناسی‌ها می‌باشد. از سوی دیگر، مسأله پیش‌بینی یال در گراف، از مسائل شناخته شده است که اخیراً با رویکردهای نوین هوش مصنوعی شامل یادگیری عمیق، پیشرفت‌های بسیاری داشته است. شبکه‌های پیچشی گرافی که ابزار اصلی مورد استفاده در این پژوهش بود، از جدیدترین روش‌های هوش مصنوعی و موضوعات داغ این حوزه است که طبق بررسی‌های انجام شده، تاکنون در توسعه اصطلاح‌نامه یا هستی‌شناسی‌ها مورد استفاده قرار نگرفته بود و برای اولین بار از منظر روش‌شناسی، یک رویکرد جدید در این حوزه مطرح شده است. استنتاج خودکار روابط اصطلاح نامه‌ای توانست منجربه افزایش روابط بین اصطلاحات و در نتیجه بهبود عملکرد آن و بازیابی موثرتر شود.

پژوهش حاضر درصدد یافتن مدل ماشینی برای استنتاج خودکار روابط بین اصطلاحات بودیم. همان‌طور که بحث شد، کشف ارتباطات بین اصطلاح‌نامه، فرآیند مهمی قلمداد می‌شود. نتایج بدست آمده از آزمایشات در چارچوب معیارهای استاندارد ارزیابی و با توجه به اینکه این روش اولین بار است که در این حوزه مورد استفاده قرار می‌گیرد، حاکی از آن است که ماشین توانایی استنتاج خودکار روابط را دارد و می‌تواند در این مبحث، توسعه اصطلاح‌نامه‌ها را تسریع نماید. در مقایسه با پژوهش‌های قبلی، روش استفاده شده برای ساخت اصطلاح‌نامه متفاوت بوده و نتایج بدست آمده، موجب اطمینان بیشتری است و در نتیجه در اهداف مختلف کاربرد اصطلاح‌نامه ازجمله نمایه‌سازی خودکار، خروجی‌های بهتری خواهیم داشت. شیوه‌های جدید در پردازش زبان طبیعی و یادگیری عمیق نیز ما را در بازیابی اطلاعات و نمایه‌سازی خودکار امیدوارتر می‌کند. امروزه با ظهور «یادگیری عمیق» و الگوریتم‌های مبتنی بر آن مانند «شبکه عصبی پیچشی»، «شبکه‌های پیچشی گراف»، «شبکه‌های توجه‌گرافی»، «شبکه‌های فضازمانی گراف»، و غیره ساختارگرافی و نوپدیدی آن و همچنین استفاده از الگوریتم‌های فوق، محبوبیت زیادی یافته‌اند. در مجموع، باید گفت که با استفاده از شیوه‌های نوین می‌توان نتایج خوبی را در حوزه اصطلاح‌نامه‌ها رقم زد و از ماشین برای تسریع و دقیق کردن آنها استفاده کرد و در نتیجه از دانش‌های حاصل موجود در بانک‌های اطلاعاتی بهره برده و در واقع مدیریت دانش اتفاق افتاده است.

۶. تقدیر و تشکر

از پژوهشکده مدیریت اطلاعات و مدارک اسلامی، پژوهشگاه فرهنگ و علوم اسلامی وابسته به دفتر تبلیغات اسلامی حوزه علمیه قم، جهت در دسترس قرار دادن اطلاعات اصطلاح‌نامه علوم اسلامی، تشکر و قدردانی می‌شود.

منـابـع

باغبانی، ش. (۱۳۹۶). *تکنیک‌ها و روش‌های یادگیری ماشین روی کلان‌داده*. در: اصفهان: کنفرانس ملی فناوری‌های نوین در مهندسی برق و کامپیوتر.

تندپور، ا. (۱۳۸۳). اصطلاحنامه: ساختار و شکل. *علوم اطلاع‌رسانی*، ۱۹(۱-۲).

حسن‌پور متی کلایی، س.ح.؛ سعادتی، ر. (۱۳۹۵). *مروری بر آخرین تغییرات و بروزآوری‌ها در شبکه عصبی کانولوشن*. در: کاشان: سومین کنفرانس ملی مهندسی برق و کامپیوتر سیستم‌های توزیع شده و شبکه‌های هوشمند.

رجبی، ت.؛ حسینی بهشتی، م.س.؛ صدیقی، م. (۱۳۹۸). روزآمدسازی و توسعهٔ اصطلاح‌نامه‌های علمی و فنی ایرانداک. *مدیریت اطلاعات*، ۵(۱): ۹۹-۱۱۸.

کریمی، م. (۲۰ مهر ۱۴۰۰). *رویه سنّتی کار اصطلاح‌نامه علوم اسلامی*. مصاحبه‌کننده: ا. نظام دوست.

موسایی، ع.ا. (۱۳۸۷). *اصطلاحنامه* چیست؟ قابل دسترس در: https://vista.ir/w/a/16/i0uki

یعقوب‌نژاد، م.ه. (۱۳۷۵). *درآمدی بر مبانی اصطلاحنامه علوم اسلامی*. قم: بوستان کتاب.

References

Anno, S., Hirakawa, T., Sugita, S. & Yasumoto, S. (2022). A graph convolutional network for predicting COVID-19 dynamics in 190 regions/countries. *Frontiers in public health*, No.10.

Baghbani, Sh. (2017). *Techniques and methods of machine learning on big data*. In: Isfahan: National Conference of New Technologies in Electrical and Computer Engineering. [in persian]

Chansanam, W., Kwiecien, K., Buranarach, M. & Tuamsuk, K. (2021). A Digital Thesaurus of Ethnic Groups in the Mekong River Basin. *Informatics-Basel*, 8(3): 50.

Ding, Y., Zhang, Z.L., Zhao, X.F., Hong, D.F., Cai, W., Yu, C.G., Yang, N.J. & Cai, W.W. (2022). Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification. *Neurocomputing*, 501: 246-257.

Du, X.W., Wan, L. & Shen, G. (2022). An Improved Graph Convolution Network for Robust Image Retrieval. *Neural Processing Letters*, Early Access, 13 NOV 2022. https://doi.org/10.1007/s11063-022-11083-2.

Evans, D.A., Ginther-Webster, K., Hart, M., Lefferts, R.G. & Monarch, I. (1991). *Automatic indexing using selective NLP and first-order thesauri*. RIAO Conference.

Fan, L., Sun, X. & Rosin, P.L. (2021). *Siamese Graph Convolution Network for Face Sketch Recognition: An application using Graph structure for face photo-sketch recognition*. In: 2020 25th International Conference on Pattern Recognition (ICPR): 8008-8014.

Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O. & Dahl, G.E. (2017). Neural Message Passing for Quantum Chemistry. *ArXiv*, abs/1704.01212.

Gowril, G., Devi, R., Sethuraman, K. & Phil, M. (2019). Machine learning. *International Journal of Research and Analytical Reviews*, 6(2).

Harkin, T. (2022). Creating a Linked Data thesaurus for Irish traditional music. *AI & Society*, 37(3): 967-974.

Hasanpour Mati kalai, S.H.; Saadati, R. (2016). *An overview of the latest changes and updates in convolutional neural network*. In: Kashan: 3rd National Conference on Electrical and

Computer Engineering, Distributed Systems and Smart Networks. [in persian]

Hasanzadeh, A., Hajiramezanali, E., Duffield, N.G., Narayanan, K.R., Zhou, M. & Qian, X. (2019). Semi-Implicit Graph Variational Auto-Encoders. *ArXiv*, abs/1908.07078.

Huang, X.H., Ye, Y.M., Ding, W.H., Yang, X.F. & Xiong, L.Y. (2022). Multi-mode dynamic residual graph convolution network for traffic flow prediction. *Information Sciences*, 609: 548-564.

Ioannidis, V.N., Zheng, D. & Karypis, G. (2020). *Few-shot link prediction via graph neural networks for Covid-19 drug-repurposing*. arXiv, 1-6.

Ito, M., Nakayama, K., Hara, T. & Nishio, S. (2008). *Association thesaurus construction methods based on link co-occurrence analysis for wikipedia*. CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management (pp. 817-826). New York, NY, USA: ACM.

Jarmasz, M. & Szpakowicz, S. (2003). *Roget's thesaurus and semantic similarity.* Paper presented at the meeting of the Conference on Recent Advances in Natural Language Processing: 212-219.

Jiang, H., Cao, P., Xu, M., Yang, J. & Zaiane, O. (2020). Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction. *Computers in biology and medicine*, 127: 104096.

Jing, Y. & Croft, W.B. (1994). *An association thesaurus for information retrieval*. *Technical Report UMASS-CS-94-17*. University of Massachusetts.

Karimi, M. (2021). *The traditional work procedure of the thesaurus of Islamic sciences*. Interviewer: A. Nexsmdost. [in persian]

Karimi, M., Hasanzadeh, A. & Shen, Y. (2020). Network-principled deep generative models for designing drug combinations as graph sets. *Bioinformatics, 36(Supplement_1)*: i445-i454.

Keramatfar, A., Rafie, M. & Amirkhani, H. (2021). *Graph Neural Networks: a bibliometrics overview. Preprint*. IEEE Access. Available Online at: https://www.researchgate.net/profile/Mohadeseh-Rafie/publication/353953475_Graph_Neural_Networks_a_bibliometrics_overview/links/611bb34e169a1a0103082d34/Graph-Neural-Networks-a-bibliometrics-overview.pdf

Kipf, T. (2020). *Deep learning with graph-structured representations. Doctoral thesis.* Amsterdam Machine Learning lab (IVI, FNWI). Available at: https://dare.uva.nl/personal/pure/en/publications/deep-learning-with-graphstructured-representations(1b63b965-24c4-4bcd-aabb-b849056fa76d). html

Li, L., Zhu, H.G., Wen, L.B., Lan, W.Z. & Yang, Z.K. (2021). *An Approach of Combining Convolution Neural Network and Graph Convolution Network to Predict the Progression of Myopia*. Neural Processing Letters. https://doi.org/10.1007/s11063-021-10576-w

McCoy, K., Gudapati, S., He, L., Horlander, E., Kartchner, D. & et al. (2021). Biomedical Text Link Prediction for Drug Discovery: A Case Study with COVID-19. *Pharmaceutics*, 13(6): 794.

Musaei, A.A. (2008). *What is a thesaurus?* Available at: https://vista.ir/w/a/16/i0uki [in persian]

Nakayama, K., Hara, T. & Nishio, S. (2007). *A Thesaurus Construction Method from Large ScaleWeb Dictionaries*. In: 21st International Conference on Advanced Information Networking and Applications (AINA '07): 932-939.

Rajabi, T., Hosseini Beheshti, M.S. & Sediqi, M. (2019). Updating and developing scientific and

technical thesauruses of Irandak. *Information Management*, 5(1): 99-118. [in persian]

Sakai, M., Nagayasu, K., Shibui, N., Andoh, C., Takayama, K., Shirakawa, H. & Kaneko, S. (2021). Prediction of pharmacological activities from chemical structures with graph convolutional neural networks. *Scientific reports*, 11(1): 525.

Smeaton, A.F. (1999). *Using NLP or NLP Resources for Information Retrieval Tasks*. In: Strzalkowski, T. (eds.), Natural Language Information Retrieval. Text, Speech and Language Technology, Vol.7. Springer, Dordrecht.

Stokes, N., Li, Y., Moffat, A. & Rong, J. (2008). An empirical study of the effects of NLP components on Geographic IR performance. *International Journal of Geographical Information Science*, 22(3): 1-14.

Tandpour, A. (2004). Thesaurus: structure and form. *Information sciences*, 19(1-2). [in persian]

Tseng, Y.H. (2002). Automatic thesaurus generation for Chinese documents. *Journal of the American Society for Information Science and Technology*, 53: 1130-1138.

Wen, G., Cao, P., Bao, H., Yang, W., Zheng, T. & Zaiane, O. (2022). MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. *Computers in biology and medicine*, 142.

Wilks, Y. (1996). Natural language processing. *Commun. ACM*, 39(1): 60–62. https://doi.org/10.1145/234173.234180

Yaqub-nejad, M.H. (1996). *An introduction to the basics of the thesaurus of Islamic sciences*. Qom: Bostan Ketab. [in persian]

Yu, K., Jiang, H., Li, T., Han, S. & Wu, X.F. (2020). Data Fusion Oriented Graph Convolution Network Model for Rumor Detection. *IEEE Transactions on Network and Service Management*, 17(4): 2171-2181.

Zhang, M. & Chen, Y. (2018). Link prediction based on graph neural networks. *Advances in Neural Information Processing Systems*, 31: 5165-5175.

Zhou, J., Huang, J.X., Hu, Q.V. & He, L. (2020). SK-GCN: Modeling Syntax and Knowledge via Graph Convolutional Network for aspect-level sentiment classification. *Knowledge-Based Systems*, 205: 106292.