



دانشگاه کاشان
University of Kashan

مجله محاسبات نرم
SOFT COMPUTING JOURNAL
تارنمای مجله: scj.kashanu.ac.ir



ارزش‌گذاری ارجاعات غیرمستقیم در شبکه‌های استنادی با استفاده از تلفیق داده‌ها^{*}

مهدی سالخورده حقیقی^{۱*}، استادیار، فاطمه ساجدی^۲، کارشناسی ارشد
^۱ دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی سجاد، مشهد، ایران.

چکیده

انجام هر فعالیت علمی نیازمند آگاهی از فعالیت‌های مرتبط پیشین است. شبکه‌های استنادی از جمله شبکه‌هایی هستند که هر سند در آن، به‌عنوان حلقه‌ای از یک زنجیره با اسناد قبلی و بعدی خود مقایسه می‌شود و اسنادی که بیشترین تعداد ارجاعات را دریافت کرده‌اند به‌عنوان مؤثرترین اسناد آن حوزه شناخته می‌شوند. تعداد اسنادهای دریافتی هر مدرک نشان‌دهنده جایگاه آن در تحقیقات انجام شده در آن حوزه است. مقالاتی با تعداد اسنادهای دریافتی زیاد، در تحقیقات آن حوزه دارای سهم قابل توجهی هستند. بنابراین ضرورت رتبه‌بندی اسناد با در نظر گرفتن ویژگی‌های مهم آن‌ها و ساختار شبکه‌های استنادی تا حد زیادی احساس می‌شود. یک چالش پیش رو در این بخش، وجود ارجاعات غیرمستقیم و تعیین ویژگی‌های مناسب است که باعث ایجاد پیچیدگی در این رتبه‌بندی می‌شود. در نتیجه، ارائه راهکاری که بتواند تأثیر چنین ارجاعاتی را به شکل مناسبی لحاظ کند، دارای اهمیت است. در این مقاله روشی به‌منظور ارزش‌گذاری بر روی اسناد با استفاده از برخی معیارهای تعریف شده در شبکه‌های استنادی ارائه شده است. ترکیب چندین معیار به‌دست‌آمده از این شبکه‌ها موجب حاصل شدن نتایج بهبودیافته و دقیق‌تری می‌شود. لذا روش پیشنهادی به کمک ترکیب ویژگی‌ها و تلفیق داده‌ها و با بهره‌گیری از روش عملگر میانگین وزنی مرتب‌شده، امتیاز اسناد را محاسبه نموده و رتبه‌بندی آن‌ها را انجام می‌دهد. رتبه‌بندی طبق معیار رتبه‌بندی صفحه‌ها به‌کارگیری ویژگی‌های شبکه‌های استنادی انجام شده است و روش تلفیق داده‌ها با روش‌های دیگر مقایسه شده که نتایج حاکی از عملکرد بهتر روش تلفیق داده‌ها دارد. مقایسه خطای رتبه‌بندی در مقایسه با یک روش مرجع با کاهش متوسط خطای حدود دو درصد در کل آزمایش‌ها نیز این نتایج را تأیید می‌کند.

© ۱۳۹۹ - مجله محاسبات نرم، کلیه حقوق محفوظ است.

اطلاعات مقاله

تاریخچه مقاله:

دریافت ۰۸ مرداد ماه ۱۳۹۸
پذیرش ۰۳ اردیبهشت ماه ۱۳۹۹

کلمات کلیدی:

عملگر میانگین وزنی مرتب‌شده
شبکه‌های استنادی
OWA
تلفیق داده‌ها
رتبه‌بندی اسناد

۱. مقدمه

با توجه به اهمیت روزافزون اطلاعات در جهان و تأثیر آن بر رشد و گسترش ابعاد مختلف جامعه، هر کشوری به دنبال آن

است که برای پیشبرد اهداف علمی پژوهشی و دانشگاهی خود، بتواند با استفاده از امکانات و تسهیلات موجود و از طریق انجام پژوهش‌های گوناگون، میزان و کیفیت تولیدات علمی خود را در دو حیطة کمی و کیفی بهبود داده و همگام با پیشرفت‌های سریع در دانش و فناوری با دیگر کشورها به رقابت علمی بپردازد.

یکی از روش‌های علم‌سنجی در تجزیه و تحلیل تولیدات

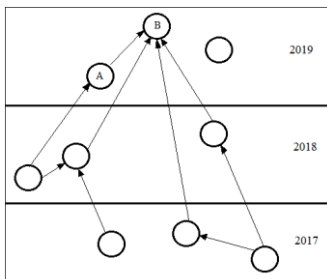
* نوع مقاله: پژوهشی

* نویسنده مسئول

پست‌های الکترونیک: haghghi@sadjad.ac.ir (سالخورده حقیقی)

fa.sajedi254@sadjad.ac.ir (ساجدی)

ارجاع مستقیم در سند A به سند B است. در بسیاری از کاربردها برای ارزش‌گذاری ارجاعات به‌طور معمول از ارجاعات مستقیم استفاده می‌شود، درحالی‌که ارزش‌گذاری دقیق‌تر بر مبنای ارجاعات مستقیم و غیرمستقیم قابل انجام است. شکل (۱) مجموعه اسنادی را نشان می‌دهد که به صورت مستقیم و غیرمستقیم ارجاعی به سند B دارند.



شکل (۱): ارجاع مستقیم و غیرمستقیم به سند B

با توجه به این اهداف، مجموعه داده یک شبکه استنادی با چنین ساختاری برای انجام تحقیق مورد استفاده قرار گرفت. سپس داده‌های این شبکه تحلیل شده و پارامترهای شبکه به منظور به‌کارگیری در محاسبه رتبه سند^۲، محاسبه شده است. مرحله بعد، انتخاب یک روش یادگیری هوشمند به منظور محاسبه رتبه سند با استفاده از پارامترهای محاسبه شده است و در انتها ارزیابی و مقایسه این روش پیشنهادی با نتایج روش مرجع و تحلیل نتایج انجام شده است.

ساختار بخش‌های باقی‌مانده این مقاله بدین صورت است: در بخش دوم، مروری بر کارهای گذشته انجام شده که شامل تحقیقات صورت‌گرفته روی شبکه‌های استنادی و تلفیق داده‌هاست. در بخش سوم، روش پیشنهادی مبتنی بر تلفیق داده‌ها و استفاده از روش OWA شرح داده خواهد شد. همچنین پارامترهای مورد استفاده معرفی و محاسبه می‌شوند. در بخش چهارم، طراحی آزمایش‌ها انجام شده و نتایج مورد مقایسه قرار می‌گیرند.

۲. پیشینه تحقیق

تحقیقات گسترده‌ای در زمینه شبکه‌های اجتماعی و به‌خصوص شبکه‌های استنادی صورت گرفته است. در تحلیل

علمی یک سازمان، بر اساس ارتباط استنادی است. در واقع ارتباطات مختلف بین نویسندگان موجب پیدایش شبکه استنادی می‌شود [۱]. بی‌شک در یک تولید علمی، نویسندگان نیازمند استفاده از پژوهش‌های پیشین هستند و نمی‌توان تنها با اتکا به حافظه، یک مطلب علمی را بیان کرد. لازم است مستندات مطالب به‌روشنی ارائه شود تا با احراز اطمینان از درستی آن‌ها، اعتبار هر نوشته تضمین گردد و بتوان به نتایج آن اعتماد کرد.

یکی از متداول‌ترین و معمول‌ترین فنونی که به ارزیابی مقالات علمی در شبکه‌های علمی بر اساس شمارش استنادهای تعلق‌گرفته به مقالات می‌پردازد، تحلیل استنادی یا مطالعه استنادی است. به همین ترتیب، بحث ضرورت استناد به مقالات علمی و همچنین رتبه‌بندی آثار علمی در این حیطه از جمله موضوعات قابل توجه است.

بنابراین با وجود استنادهای بسیار در مقالات علمی، چالش اساسی رتبه‌بندی اسناد، وجود ارجاعات غیرمستقیم به اسناد و در نظر گرفتن ویژگی‌های مهم در شبکه‌های استنادی برای انجام این عمل است. به‌رغم پژوهش‌های بسیاری که در زمینه رتبه‌بندی صفحات و اسناد در شبکه‌های استنادی صورت گرفته، ولی روش‌های تلفیق داده‌ها چندان مورد توجه قرار نگرفته‌اند. در پژوهش حاضر از عملگر میانگین وزن‌دار مرتب‌شده (OWA)^۱ در رتبه‌بندی به‌عنوان یکی از روش‌های قدرتمند در تلفیق داده‌ها استفاده شده است. طبق این گفته‌ها تعیین ویژگی‌های مناسب برای رتبه‌بندی و یافتن بهترین پارامترها به‌منظور طراحی یک سیستم هوشمند که بتواند با کمترین خطا اسناد را رتبه‌بندی کند، همچنین بررسی اثر تلفیق داده‌ها با روش OWA روی رتبه‌بندی، از اصلی‌ترین اهداف این مقاله‌اند.

در روش ارائه‌شده در این تحقیق، شبکه استنادی به صورت گراف جهت‌دار (V, E) در نظر گرفته می‌شود که V مجموعه رأس‌های گراف یا همان اسناد، و E مجموعه کمان‌ها یا همان ارجاعات هستند. کماتی از A به B به صورت (A, B) به معنی

گردید. از معیار یافتن کوتاه‌ترین مسیر ارتباطی^۳ و یافتن مسیر زوج گره‌ها^۴ استفاده شد که با استفاده از یال‌ها مسیرها را مشخص نموده و تعداد گره‌ها را می‌شمارد. از اشکالات روش پیشنهادی آن‌ها، با توجه به معیارهای صورت‌گرفته برای تحلیل شبکه، نتایج بسیار دقیق حاصل نشد؛ زیرا معیارهای مرکزیت مهم‌تری در تحلیل شبکه وجود دارد که دقت تحلیل شبکه را بالا می‌برد.

ماسوکی و همکاران^۵ در [۵] قصد داشتند اعتبار دانشگاه‌های معتبر دنیا را بر طبق مقالاتی که در سال‌های ۲۰۱۰ تا ۲۰۱۴ منتشر شد و استنادهای انجام‌شده به آن‌ها رتبه‌بندی کنند. در این پژوهش، از مجموعه داده‌ای با ۱۸۸۵۳۳ مقاله استفاده کردند. مقالات جمع‌آوری شده در ۵ حوزه مختلف دندان‌پزشکی، دامپزشکی، مالی، تجارت و جراحی دهان است. در این تحقیق، آن‌ها برای محاسبه رتبه سند از معیار in-degree استفاده کردند و رتبه‌بندی را بر اساس رتبه سند به‌دست‌آمده انجام دادند. برای بررسی درستی روش پیشنهادی خود از معیار همبستگی استفاده کردند. همچنین رتبه‌بندی به‌دست‌آمده را با رتبه‌بندی مرکز شانگهای از دانشگاه‌های دنیا^۶ مقایسه کردند. در انتها ضریب همبستگی برای روش پیشنهادی کاهش پیدا کرد و صحت روش اثبات شد. از جمله ضعف‌های این روش می‌توان به استفاده از ویژگی‌های کم برای رتبه‌بندی اشاره کرد. یو لین و همکاران^۷ [۶] روی پیش‌بینی اتصال در شبکه‌های مجازی کار کردند. هدف اصلی استفاده از پیش‌بینی اتصال در شبکه مجازی و تخمین شباهت ارتباط بین دو گره است. روش‌های بسیاری در تحلیل شبکه مجازی وجود دارد، اما روش ارائه‌شده در این مقاله نتایج بهتری دارد. در این مقاله عملگر میانگین وزن‌دار مرتب‌شده بر اساس الگوریتم پیش‌بینی اتصال^۸ برای شبکه مجازی با سه روش مختلف آزمایش شد. در نهایت نتایج روی مجموعه داده شبکه مجازی نشان داد که

این‌گونه شبکه‌ها، از معیارهای متعدد برای رتبه‌بندی مقالات استفاده شده است. در این بخش، برخی مطالعات و تحقیقات صورت‌گرفته در این زمینه‌ها بررسی می‌شود.

در تحقیق انجام‌شده توسط یی و همکاران^۱ [۲] روشی جایگزین برای اندازه‌گیری اهمیت مقاله‌های علمی بر اساس رتبه سند ارائه شده است. در این پژوهش، از مجموعه داده مقالاتی در حوزه علم شیمی در سال‌های ۲۰۰۰ تا ۲۰۰۵ استفاده شد. این مجموعه داده شامل ۲۶۱ مقاله بود. الگوریتم رتبه‌بندی سند برای هر مقاله محاسبه شد و در نهایت، رتبه‌بندی صورت گرفت. نتایج نشان داد آمریکا بیشترین نقش را در تولیدات علمی داشته است اما کشورهای اروپایی بیشترین رتبه را از لحاظ کیفیت و تعداد استنادها دریافت کرده و کشورهای آسیایی کمترین رتبه را به خود اختصاص داده‌اند. ضعف این روش در تعداد کم مجموعه داده و عدم محاسبه وزن‌های معیار رتبه‌بندی بود.

همچنین پائول و همکاران^۲ [۳] در پژوهشی با استفاده از الگوریتمی برای تعیین رتبه سند به ارائه روشی جدید برای رتبه‌بندی نویسندگان در شبکه کتاب‌شناختی چندلایه پرداختند. روش رتبه‌بندی آن‌ها بدین صورت بود که ابتدا مقالات و نویسندگان موجود در مجموعه در طی فرایندهای تکرار شونده رایج در رتبه سند، به ترتیب بر اساس استناد و هم‌تألفی، امتیازدهی و رتبه‌بندی شدند. سپس سهم هر نویسنده در مقالات هم‌تألف به شیوه‌های گوناگون تعیین، و در نهایت رتبه‌بندی نویسندگان انجام شد. نتایج نشان داد همبستگی در رتبه‌های پایین کمتر بوده و روش پیشنهادی این پژوهش در رتبه‌بندی ارجاعات عملکرد مطلوبی داشت. ضعف اندک روش آن‌ها تعداد معیارهای در نظر گرفته شده بود.

در روش ارائه‌شده توسط کریستوفر و همکاران [۴] پژوهشی روی تحلیل شبکه‌های بزرگ و پیچیده انجام شد. در این مقاله برای تحلیل شبکه‌های بزرگ و پیچیده از دو روش استفاده شد که باعث کاهش پیچیدگی و تحلیل آسان شبکه

3. Search Path Link Count (splc)

4. Search Path Node Pair (spnp)

5. Massucci and Docampo

6. The Shanghai Academic Ranking of World Universities (ARWU)

7. Yu, lin, he

8. Link Prediction Ensemble(LPEOWA)

1. Yi, Nan, Liancheng

2. Paul, Kumar, Choudhury & Nandi

روش دمپستر شفر [۱۱]، بی‌زین [۱۲] و OWA [۱۳]. رونالد یاگر در سال ۱۹۸۸ [۱۳] عملگر تجمیع جدیدی تحت عنوان عملگر میانگین وزن‌دار مرتب‌شده (OWA) را برای اولین بار معرفی کرد. این عملگر قالبی واحد برای تصمیم‌گیری در شرایط غیرقطعی ایجاد می‌کند؛ به‌گونه‌ای که در این روش سیاست‌های تصمیم‌گیری مختلف مانند حداقل بیشینه (خوش‌بینانه)، حداکثر کاهش (بدبینانه)، با عملگرهای مختلف و در نتیجه با بردارهای وزن مختلف مشخص می‌شوند. یک مسئله مهم در تئوری عملگر OWA نحوه تعیین وزن‌های این عملگر است. به‌منظور تعیین وزن‌های عملگر OWA نیز روش‌های مختلفی پیشنهاد شده است [۱۴] [۱۵] [۱۶] [۱۷]. این روش‌ها بر مبنای تصمیم‌های خروجی هر بخش، وزن‌ها را تعیین کرده و در فرایند تلفیق از آن‌ها استفاده می‌کنند. در بخش بعدی، جزئیات روش تعیین وزن‌ها در روش پیشنهادی ارائه می‌شود.

۳. روش پیشنهادی

با توجه به هدف پژوهش که تعیین ویژگی‌های مناسب برای رتبه‌بندی و بررسی اثر تلفیق داده‌ها با روش OWA بر روی رتبه‌بندی و ارزش‌گذاری ارجاعات شبکه استنادی است، در انجام آن از تکنیک‌های مختلف تحلیل شبکه برای بررسی و تحلیل داده‌ها استفاده شده است. بنابراین، مجموعه داده‌ای از مقالات در حوزه علم فیزیک مربوط به سال ۲۰۰۶ که توسط نیومن گردآوری شده، مورد استفاده قرار گرفته است. پس از مشخص کردن مجموعه داده، برای تحلیل شبکه استنادی و تعیین پارامترها از نرم‌افزار گفی کمک گرفته شد. نرم‌افزارهای متعددی برای تحلیل شبکه وجود دارد که ما نرم‌افزار گفی را برای این کار انتخاب کردیم. سپس، به کمک این نرم‌افزار پارامترهای شبکه محاسبه شد. پارامترهای شبکه شامل Hub, Modularity-class, Closeness, Betweenness, Authority, Eigencentrality, PageRank می‌باشد. تعاریف و نحوه محاسبه پارامترها در ادامه آمده است.

با در نظر گرفتن مدل گراف از شبکه استنادی مطابق شکل

این روش در پیش‌بینی ارتباطات عملکرد بهتر و دقت بیشتری دارد.

در پژوهش صورت‌گرفته توسط مارا و امروزنژاد [۷] روشی برای کاهش پیچیدگی در تصمیم‌گیری و تحلیل شبکه‌های استنادی ارائه شد. در این مقاله از عملگر وزنی مرتب‌شده به‌منظور ترکیب داده‌ها و برای رتبه‌بندی ارجاعات، استفاده شده است. امروزنژاد مدل وانگ [۷] را با برنامه‌نویسی خطی بهبود داد و در مدل پیشنهادی، حداکثر اختلاف بین جفت وزن‌ها را به حداقل رساند. این مدل برای سیستم‌های رتبه‌بندی بسیار کارایی دارد. در این مقاله، معیارهای مجموعه داده محاسبه شد و مورد تحلیل و بررسی قرار گرفت. جداولی از رتبه‌بندی معیارها نیز ارائه شد. همچنین برای رتبه‌بندی اسناد از عملگر OWA استفاده شد. در نهایت رتبه‌بندی‌های دو روش تحلیلی و این عملگر مورد مقایسه و بحث قرار گرفت.

گنزالز و همکاران [۸] در پژوهش خود برای پوشش مناسب صفحات وب (پیدا کردن سریع سندهای مهم)، الگوریتم خزشی به نام IECA^۱ مبتنی بر چندین ویژگی نظیر درجات خروجی، درجات ورودی، فاصله لگاریتمی و خاصیت ساختاری گراف وب ارائه کردند. برای ارزیابی دقت این الگوریتم از چهار گراف مختلف وب ایران، ایتالیا، انگلستان و دانشگاه برکلی استفاده شد. نتایج آزمایش‌ها حاکی از کارآتر بودن الگوریتم آن‌ها نسبت به سایر الگوریتم‌های بررسی‌شده در این مقاله است.

تلفیق داده‌ها به‌عنوان یکی از روش‌های مؤثر در تصمیم‌گیری در سیستم‌های هوشمند کاربرد دارد [۹]. در مقاله‌ای در سال ۲۰۱۸ [۱۰] مؤلفان از یک عملگر تجمیع فازی به‌عنوان روشی برای تلفیق داده‌ها برای تصمیم‌گیری استفاده کردند. عملگر فازی ارائه‌شده نتایج چند عملگر را تجمیع می‌کند. کاربرد این عملگر در محیط‌هایی است که با اعداد فازی توصیف می‌شوند. روش‌های مختلف دیگری نیز برای تلفیق داده‌ها ارائه شده است که برخی از آن‌ها عبارت‌اند از

مقادیر بالای این معیار بیانگر تمرکز بیشتر اسناد در هر حوزه است که می‌تواند در ارزش‌گذاری ارجاعات موثر باشد.

معیار *betweenness* راهی است برای تشخیص اینکه چه گره‌هایی بین هر دو گره دلخواه *A* و *B* قرار دارند. در نتیجه سه خاصیت مهم با این تعریف به دست می‌آید: ۱. اگر *C* گره‌ای بین *A* و *B* و مخالف *A* باشد، در این صورت *A* بین *B* و *C* قرار ندارد. ۲. اگر *C* بین *A* و *B* باشد و *D* بین *A* و *C*، در این صورت *D* بین *A* و *B* هم هست. ۳. اگر *C* بین *A* و *B* باشد و *D* بین *A* و *C* باشد، *C* بین *D* و *B* هم هست. با توجه به تعریف و خواص ذکر شده، روش‌هایی برای محاسبه آن نیز ارائه شده است [۱۹]. استفاده از این معیار به‌عنوان یک ویژگی برای ارزش‌گذاری ارجاعات می‌تواند تأثیر ارجاعات غیرمستقیم را نیز در محاسبات منظور نماید. نوع تأثیر این معیارها در ارزش‌گذاری در روش پیشنهادی و آزمایش‌های انجام‌شده مشخص می‌شود.

روش پیشنهادی در ارزش‌گذاری اسناد بر مبنای ارجاعات در شبکه‌های استنادی، از معیارهای تعریف‌شده در بالا بر مبنای مدل گراف در این نوع شبکه‌ها استفاده نموده و ارزش‌گذاری را انجام می‌دهد و بر مبنای آن اسناد را رتبه‌بندی می‌نماید. در شکل (۲) مراحل انجام‌شده در سیستم روش پیشنهادی قابل مشاهده است.

روش پیشنهادی با عنوان رتبه‌بندی سند مبتنی بر تلفیق داده‌ها یا به اختصار *DFDR*^۱ رتبه‌بندی اسناد را در سه مرحله اصلی انجام می‌دهد. مرحله اول محاسبه پارامترهای صفحات و تولید مجموعه داده‌های آموزش است. مرحله دوم استفاده از تخمین گره‌های آموزش‌پذیر و آموزش آن‌ها برای تخمین رتبه هر سند است. در مرحله سوم، تلفیق نتایج تخمین گره‌ها انجام شده و نتیجه نهایی که رتبه سند است، به دست می‌آید.

در مرحله اول، پارامترهای مختلف ذکر شده در بالا برای هریک از اسناد مجموعه داده مورد آزمایش محاسبه می‌شود. رتبه‌بندی مبنای در نظر گرفته شده برای این اسناد رتبه‌ای است که نرم‌افزار گفی به آن اختصاص می‌دهد. در مرحله بعد،

(۱)، چند معیار کلیدی در این شبکه‌ها تعریف می‌شود که در روش ارائه شده در این تحقیق استفاده شده‌اند. از جمله مهم‌ترین معیارهای محاسبه شده بر مبنای این مدل از شبکه عبارت‌اند از: *hub*، *authority*، *betweenness*، *closeness*، *modularity*، *in-degree* و *out-degree*. در توضیح این پارامترها ذکر چند نکته ضروری است. معیارهای *hub* و *authority* نشانگرهایی برای ساختار شبکه هستند. آن‌ها وارده‌های بردار ویژه گراف وزن‌دار بدون جهتی هستند که از گراف جهت‌دار اولیه به دست آمده‌اند [۷]. برای هر گره از گراف اولیه مانند *A*، وزن این دو مقدار بر مبنای وزن آن‌ها در گره‌هایی که *A* به آن‌ها ارجاعی دارد یا آن‌ها به *A* ارجاع دارند محاسبه می‌شود. بنابراین برای گراف اولیه به صورت $G(V, E)$ ، که *E* مجموعه یال‌ها و *V* مجموعه رأس‌های گراف است، وزن معیار *hub* و *authority* (*HW* و *AW*) برای گره *A* با رابطه‌های (۱) و (۲) محاسبه می‌شود.

$$HW(A) = \sum_{\forall B \in \{(A, B) \in E\}} AW(B) \quad (1)$$

$$AW(A) = \sum_{\forall B \in \{(B, A) \in E\}} HW(B) \quad (2)$$

معیار *in-degree* برای سند *A* تعداد اسنادی است که به سند *A* به‌طور مستقیم ارجاع داده‌اند. معیار *out-degree* تعداد اسناد ارجاع‌شده توسط *A* می‌باشد. معیار *closeness* (*CN*) سهولت دسترسی به یک سند را نشان می‌دهد. مقدار آن عکس میانگین فاصله گره *A* و هر گره دیگر در گراف *G* است که با رابطه (۳) محاسبه می‌شود.

$$CN(A) = (n - 1) / \sum_{\forall B \in V - \{A\}} d(A, B) \quad (3)$$

در رابطه (۳) مقدار *n* تعداد کل اسناد و نماد *d* نشان‌دهنده فاصله بین دو گره مورد نظر است. معیار *modularity* قابلیت تقسیم شدن گراف به ماژول‌های کوچک‌تر را نشان می‌دهد. به‌منظور تعیین این قابلیت، روشی برای شناسایی ماژول‌ها مورد نیاز است. هر ماژول شامل گره‌هایی است که ارتباطات بین آن‌ها زیاد و فشرده بوده درحالی‌که ارتباطات بین ماژول‌ها کم و پراکنده است. از این معیار برای خوشه‌بندی گره‌ها و تشخیص جوامع در کاربردهای مختلف استفاده می‌شود [۱۸].

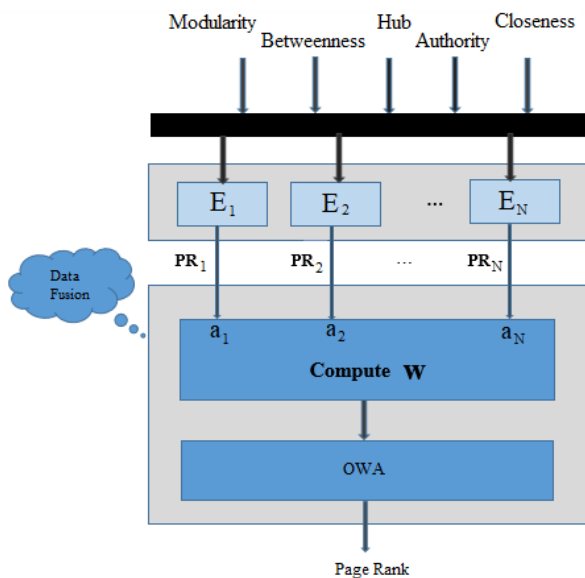
$$f(a_1, a_2, \dots, a_n) = \sum_{j=1}^n w_j b_j = w_1 b_1 + w_2 b_2 + \dots + w_n b_n \quad (4)$$

$$w_j \in [0,1]$$

$$\sum_{j=1}^n w_j = 1$$

که در آن، b_j برابر است با j امین عنصر بزرگ در میان مقادیر a_i . چالش اصلی این روش به دست آوردن وزن‌هاست. در این روش، راه‌حل‌های متعددی برای تعیین وزن‌های بردار W ارائه شده است. این روش‌ها بر دو مفهوم اصلی این روش که توسط یاگر معرفی شده‌اند تأکید دارند. اولین مفهوم *orness* است که با رابطه (۵) تعریف شده است:

$$orness(\vec{w}) = \frac{1}{(n-1)} \sum_{i=1}^n w_i (n-i) \quad (5)$$



شکل (۲): ساختار کلی سیستم پیشنهادی

رابطه (۵) میزان «یایی» یا میزان شباهت عملگر *OWA* را به عملگر «OR» نشان می‌دهد. همچنین معیار *dispersion* با رابطه (۶) تعریف می‌شود:

$$dispersion(\vec{w}) = - \sum_{i=1}^n w_i \ln(w_i) \quad (6)$$

برای تعیین وزن‌های *OWA*، از معیار *dispersion* نیز استفاده می‌شود. در رابطه (۶) مشاهده می‌شود که مقدار لگاریتم اگر $w_j = \frac{1}{n}$ باشد حداکثر است و حداقل آن صفر است در زمانی که یکی از $w_i = 1$ باشد و بقیه صفر باشند.

پارامترهای محاسبه‌شده برای هر سند به‌عنوان ویژگی‌ها و رتبه داده‌شده به آن به‌عنوان خروجی، در مجموع یک نمونه داده ورودی سیستم پیشنهادی را تشکیل می‌دهند. در نتیجه مجموعه داده مورد نیاز برای آزمایش‌ها با استفاده از گراف شبکه استنادی این‌گونه ایجاد می‌شود.

روش پیشنهادی با استفاده از یک روش تلفیق داده‌ها و مجموعه داده ایجادشده، سعی در رتبه‌بندی اسناد بر مبنای ارجاعات مستقیم و غیرمستقیم آن‌ها دارد. میزان مناسب بودن رتبه‌بندی ارائه‌شده توسط روش پیشنهادی، با انجام مقایسه با رتبه‌بندی پایه ایجادشده با نرم‌افزار گفی تعیین می‌شود. روش مقایسه در بخش آزمایش‌ها توضیح داده خواهد شد. ساختار تخمین‌گرهای مورد استفاده در این روش می‌تواند شبکه عصبی، فازی، یا هر نوع تخمین‌گر دیگر یا ترکیب آن‌ها باشد. در روش ارائه‌شده در این تحقیق، از شبکه‌های عصبی با یک یا دو لایه مخفی و تعداد نرون‌های متفاوت استفاده شده است. تفاوت در تعداد لایه‌ها و تعداد نرون‌ها می‌تواند گوناگونی تخمین‌گرها را تضمین کند که برای تلفیق داده‌ها ضروری است [۲۰]. در سیستم پیشنهادی، از کل مجموعه داده، ۷۰ درصد به‌صورت تصادفی برای آموزش انتخاب شد و بقیه برای آزمایش در نظر گرفته شد. در مرحله آموزش، هریک از تخمین‌گرها آموزش داده شده و خروجی آن‌ها به‌عنوان ورودی بخش تلفیق داده‌ها استفاده می‌شود. شکل (۲) ساختار سیستم پیشنهادی را نشان می‌دهد.

در شکل (۲)، N تعداد تخمین‌گرهاست. ورودی هریک از آن‌ها ویژگی‌های محاسبه‌شده برای هر سند است و خروجی هریک رتبه محاسبه‌شده برای آن خواهد بود. در مرحله تلفیق داده‌ها، خروجی‌های این تخمین‌گرها با روش *OWA* تلفیق شده تا رتبه نهایی محاسبه شود.

روش میانگین وزن‌دار مرتب‌شده (*OWA*) توسط یاگر معرفی شد [۱۳]. این روش یک نگاشت n بعدی به‌صورت $f: R^n \rightarrow [0,1]$ همراه با بردار وزنی $W = (w_1, \dots, w_n)$ و یک مجموعه از داده‌های ورودی $A = (a_1, \dots, a_n)$ است. نگاشت f با رابطه (۴) تعریف می‌شود.

تعداد داده‌های آزمایش است.

۱.۳. محاسبه وزن‌های عملگر

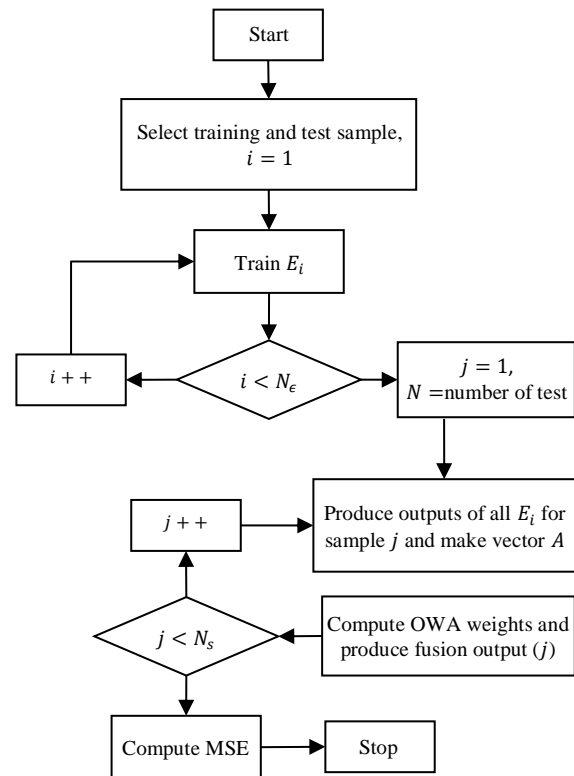
در روش پیشنهادی، هدف اصلی بالا بردن معیار dispersion در تصمیم‌گیری با در نظر گرفتن مقدار مشخصی برای orness و به حداقل رساندن خطاست. برای رسیدن به این اهداف از مفهوم اعداد فازی استفاده می‌شود. در روش پیشنهادی که در شکل (۲) مشاهده می‌شود، با استفاده از داده‌های ورودی، هر تخمین‌گر مقداری را در خروجی به‌عنوان رتبه‌ی صفحه تولید می‌کند. با توجه به وجود عدم قطعیت در مقادیر محاسبه‌شده، استفاده از این مقادیر به‌عنوان مقادیر غیرقطعی در بخش تلفیق باید با استفاده از توصیف غیرقطعی آن‌ها انجام شود. لذا برای توصیف غیرقطعی این مقادیر از مفهوم اعداد فازی استفاده شده است. این توصیف با استفاده از توابع عضویت مختلف قابل انجام است که در این روش از تابع مثلثی استفاده شده است. شکل (۴) چگونگی تعریف و استفاده از آن را نشان می‌دهد.

در این روش، خروجی هر تخمین‌گر مقداری است که به‌عنوان رتبه سند در نظر گرفته می‌شود. این مقدار به‌عنوان یک عدد فازی در نظر گرفته می‌شود، به‌طوری که برای مقایسه مقادیر تولیدشده توسط تخمین‌گرها، مقادیر به‌صورت فازی با یکدیگر مقایسه می‌شوند. برای اعداد فازی استفاده‌شده در این روش، تابع عضویت مثلثی در نظر گرفته شده است. البته برای انجام آزمایش‌ها و مقایسه نتایج، می‌توان از توابع عضویت دیگری نیز استفاده کرد.

در روش ارائه‌شده، اگر تعداد تخمین‌گرها N_E باشد و مینیمم و ماکزیمم مقادیر تولیدشده E_{min} و E_{max} باشد، برای خروجی تخمین‌گر i (E_i) عدد فازی به‌صورت مثلث شکل (۴) که با رنگ مشکی مشخص شده تعریف می‌شود. اگر تعداد خروجی‌های بیشتر از E_i با N_{ih} و تعداد خروجی‌های کمتر از آن با N_{il} مشخص شود، رأس هر مثلث بر مبنای رابطه (۷) جابه‌جا می‌شود.

$$\hat{E}_i = \begin{cases} E_i + d & \text{if } N_{ih} > N_{il} \\ E_i - d & \text{if } N_{il} > N_{ih} \end{cases} \quad (7)$$

این معیار میزان تأثیر هر یک از مقادیر ورودی را در نتیجه خروجی مشخص می‌کند. بنابراین حداکثرسازی این معیار با در نظر گرفتن میزان مشخصی برای orness می‌تواند از بایاس شدن خروجی به‌سمت یکی از مقادیر ورودی جلوگیری و نتایج مطلوب‌تری تولید کند. علاوه بر این، وزن‌ها باید طوری مشخص شوند که خطای کلی تصمیم‌گیری به حداقل برسد. در روش پیشنهادی، هدف بالا بردن معیار dispersion و به حداقل رساندن خطا با در نظر گرفتن میزان orness می‌باشد. در ادامه روش محاسبه وزن‌ها توضیح داده خواهد شد. مراحل اجرای روش پیشنهادی در شکل (۳) مشاهده می‌شود.



شکل (۳): مراحل اجرای روش پیشنهادی

در شکل (۳) مجموعاً سه مرحله کلی وجود دارد: مرحله اول آموزش تخمین‌گرها با داده‌های آموزش تولیدشده است. در مرحله دوم، خروجی تخمین‌گرها برای هر داده آزمایش به دست آمده و پس از محاسبه وزن‌ها رتبه سند با تلفیق این مقادیر تولید می‌شود. در مرحله آخر، خطای رتبه‌بندی نسبت به رتبه‌بندی مرجع محاسبه خواهد شد. در این شکل، N_s

j در این گراف وجود دارد. برای این مجموعه داده، که در قالب گراف ارائه شده است، معیارهای تعریف‌شده در بخش ۳ شامل authority, hub, out-degree, in-degree, modularity, closeness, between-ness و degree محاسبه شده و مجموعه داده جدیدی تولید می‌شود. ویژگی‌های این مجموعه داده جدید این معیارها هستند و رتبه محاسبه‌شده برای هر نمونه نیز خروجی آن در نظر گرفته شده است. از این مجموعه داده جدید، ۷۰ درصد از نمونه‌ها برای آموزش و بقیه نمونه‌ها برای آزمایش در نظر گرفته شده‌اند. اگرچه میزان مناسب بودن رتبه اسناد در کاربردهای مختلف به نوع کاربرد بستگی دارد، برای انجام مقایسه‌ها و تحلیل نتایج، در این آزمایش‌ها رتبه‌بندی انجام‌شده توسط روش پیشنهادی با رتبه‌بندی مرجع که توسط نرم‌افزار گفنی تولید شده مقایسه می‌گردد و به‌عنوان یک معیار مقایسه دیگر برای تعیین خطای روش پیشنهادی با روش مرجع، از معیار میانگین مربعات خطا^۲ استفاده شده است.

با توجه به اینکه از نرم‌افزار گفنی نیز برای مقایسه‌ها استفاده شده است، روش محاسبه رتبه صفحه توسط آن به‌طور خلاصه تشریح می‌شود. محاسبه رتبه صفحه به‌طور کلی توسط رابطه (۱۲) محاسبه می‌شود.

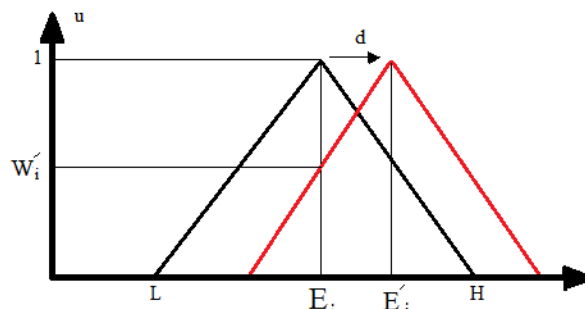
$$PR(A) = (1 - d) + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right) \quad (12)$$

در این رابطه، $PR(A)$ رتبه صفحه A و $PR(T_i)$ رتبه صفحه T_i و $C(T_i)$ تعداد اتصالات خارج‌شده از صفحه T_i می‌باشد. مقدار d نیز یک فاکتور محدودکننده است که $0 < d < 1$ می‌باشد. با توجه به این رابطه، رتبه صفحه بر مبنای مجموع رتبه تمام صفحه‌های متصل شده به آن (واردشده به آن) تقسیم بر تعداد اتصالات خارج‌شده از هر یک از آن‌ها (خارج‌شده از هر یک از آن‌ها) محاسبه می‌شود.

آزمایش‌های متعددی روی ترکیب معیارهای مختلف در این بخش برای به‌دست‌آوردن رتبه سند انجام شده است. در جدول (۱) نتیجه رتبه‌بندی برای ۱۰ رتبه برتر که توسط روش

مقدار d با رابطه (۸) محاسبه می‌شود. در شکل (۴) عدد فازی به‌دست‌آمده به توجه به رابطه (۷) با رنگ قرمز نشان داده شده که با فرض $N_h > N_l$ رسم شده است.

$$d = \frac{|N_{ih} - N_{il}|}{N_{ih} + N_{il}} (E_{max} - E_{min}) \quad (8)$$



شکل (۴): محاسبه وزن‌ها با اعداد فازی

در شکل (۴) مقادیر L و H از رابطه (۹) و (۱۰) به دست می‌آیند.

$$L = E_i - (E_{max} - E_{min})/N_E \quad (9)$$

$$L = E_i + (E_{max} - E_{min})/N_E \quad (10)$$

در نتیجه، وزن W_i مطابق شکل (۴) به دست می‌آید. به‌منظور برقراری رابطه (۴)، این وزن‌ها می‌بایست با رابطه (۱۱) نرمال شوند.

$$w_i = \frac{W_i}{\sum_{i=1}^{N_E} W_i} \quad (11)$$

۴. آزمایش‌ها و تحلیل نتایج

به‌منظور انجام آزمایش‌های این بخش، از مجموعه داده‌ای شامل مقاله‌هایی در حوزه علم فیزیک استفاده شده است که دانشگاه استنفورد تهیه کرده است [۲۱]. این مجموعه داده یک شبکه هم‌تألیف^۱ است. ساختار شبکه‌های هم‌تألیفی از گره‌های نویسندگان و روابط بین گره‌ها تشکیل شده است. هم‌تألیفی رسمی‌ترین جلوه همکاری علمی، فکری و اجتماعی نویسندگان در تولید پژوهش‌های علمی است. ساختار این مجموعه داده به‌صورت گراف جهت‌داری است که با استفاده از ۳۴۵۴۶ گره و ۴۲۱۵۷۸ یال تعریف شده است. در این گراف، اگر سند i به سند j ارجاع دهد، کماتی از گره i به گره

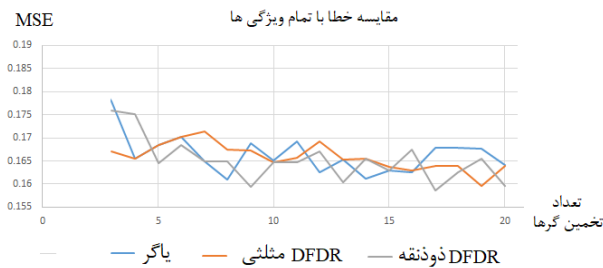
پیشنهادی و روش مرجع انجام شده، مشاهده می‌شود.

جدول (۱): مقایسه رتبه بندی ۱۰ مقاله برتر روش پیشنهادی و روش مرجع

رتبه	روش پیشنهادی		روش مرجع	
	رتبه صفحه	شناسه	رتبه صفحه	شناسه
۱	۰/۴۱۱	۷۸	۰/۴۱۲	۷۸
۲	۰/۳۷۳	۳۳	۰/۳۷۲	۳۳
۳	۰/۳۲۷	۸۶	۰/۲۷۶	۳۴
۴	۰/۲۷۸	۲۸۱	۰/۲۳۵	۲۸۱
۵	۰/۲۲۹	۲۹۴	۰/۲۳۴	۲۹۴
۶	۰/۲۰۹	۴۱	۰/۲۲۷	۲۱۶
۷	۰/۲۰۷	۹	۰/۲۱۴	۵۴
۸	۰/۱۹۹	۹۶	۰/۱۹۶	۹۶
۹	۰/۱۹۲	۲۳۸	۰/۱۹۵	۱۵۰
۱۰	۰/۱۸۹	۴۶	۰/۱۹۳	۴۶

شده است. با توجه به اینکه روش پیشنهادی از مفهوم اعداد فازی برای تلفیق استفاده کرده است و این اعداد با تابع عضویت مثلثی تعریف شده‌اند، در این مقایسه‌ها، روش پیشنهادی با توابع عضویت مثلثی و دوزنقه استفاده شده است. علاوه بر آن، مقایسه با روش یاگر نیز صورت گرفته است.

اولین مجموعه نتایج آزمایش‌ها در شکل (۵) مشاهده می‌شود. در این شکل، محور افقی تعداد تخمین‌گرهایی است که در مرحله تلفیق استفاده شده‌اند. محور عمودی میزان خطای محاسبه‌شده با رابطه (۱۳) می‌باشد که صد برابر شده‌اند. در این آزمایش از تمام ویژگی‌های تعریف‌شده در بالا استفاده شده است.



شکل (۵): مقایسه خطای روش‌های تلفیق با استفاده از تمام ویژگی‌ها

همان گونه که در شکل (۵) مشاهده می‌شود، افزایش تعداد تخمین‌گرها تا ۹ عدد می‌تواند تأثیر مثبتی بر کاهش خطا در هر سه روش داشته باشد. افزایش تعداد آن‌ها بعد از این تأثیر چندان مثبتی را نشان نمی‌دهد. همچنین مشاهده می‌شود که روش پیشنهادی در هر دو حالت نسبت به روش یاگر به‌طور کلی با کاهش خطا همراه است.

در آزمایش‌های بعدی علاوه بر ویژگی‌های پایه که در in-degree, out-degree هستند، به‌منظور بررسی تأثیر معیارهای مختلف بر میزان خطا و مقایسه روش‌های تلفیق مختلف، از دو معیار closeness و hub نیز استفاده شده است. نتیجه این مقایسه در شکل (۶) مشاهده می‌شود.

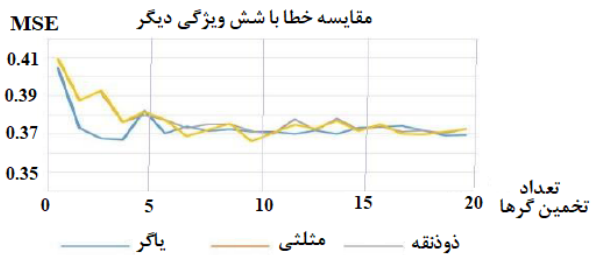
همان گونه که در شکل (۶) مشاهده می‌شود، میزان خطای هر سه روش نسبت به شکل (۵) که از همه ویژگی‌ها استفاده کرده، بیشتر است. همچنین بعد از به‌کارگیری تعداد پنج تخمین‌گر، افزایش تعداد آن‌ها تأثیر چندانی بر کاهش خطا نداشته است. آزمایش بعدی تأثیر افزودن دو ویژگی دیگر را

همان گونه که در جدول (۱) مشاهده می‌شود، مقاله‌های رتبه اول و دوم در هر دو روش مشابه تشخیص داده شده‌اند؛ اگرچه مقدار عددی رتبه سند محاسبه شده کمی اختلاف دارد. در رتبه‌های ۳ تا ۱۰ تفاوت در رتبه‌بندی در برخی موارد وجود دارد. در رتبه‌بندی ۱۰ مقاله برتر اگر یکسان نبودن مقالات هم‌رتبه را به‌عنوان خطا در نظر بگیریم، ۴ مورد خطا در این رتبه‌بندی نسبت به روش مرجع وجود دارد. البته نتیجه‌گیری دیگر می‌تواند با مقایسه اعداد محاسبه‌شده برای رتبه سند انجام شود و این مقدار به‌عنوان خطا در نظر گرفته شود. در این صورت، میانگین مربعات خطا (MSE) برای مقدار محاسبه‌شده برای رتبه سند برای تمام اسناد با رابطه (۱۳) به دست می‌آید.

$$MSE = \frac{1}{N} \sum_{i=1}^N (PR_r - PR_p)^2 \quad (13)$$

در رابطه (۱۳)، مقدار N تعداد اسناد و PR_r رتبه سند محاسبه‌شده توسط روش مرجع و PR_p رتبه محاسبه‌شده توسط روش پیشنهادی است. این مقدار برای ۱۰ سند برتر برابر است با $0.5748e^{-3}$. این عدد می‌تواند معیاری برای میزان اختلاف در عملکرد دو روش باشد.

در مجموعه دیگری از آزمایش‌ها تأثیر تعداد تخمین‌گرها بر عملکرد الگوریتم تلفیق بررسی می‌شود. همچنین مقایسه روش پیشنهادی با روش‌های دیگر نیز در این آزمایش‌ها انجام



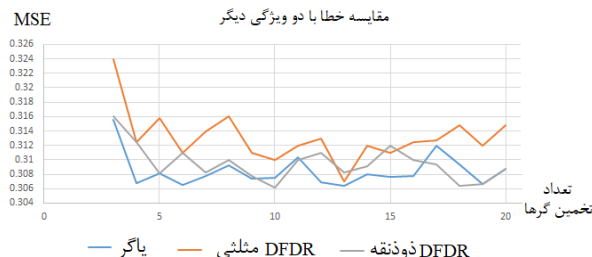
شکل (۸): مقایسه خطای روش‌های تلفیق با استفاده از شش ویژگی دیگر

با توجه به آزمایش‌های انجام‌شده و بررسی نتایج به‌دست‌آمده، ملاحظه می‌شود که افزودن ویژگی به مجموعه داده در آزمایش‌ها لزوماً باعث بهبود قابل توجه در نتایج نمی‌شود و برخی ویژگی‌ها تأثیر بیشتری در بهبود کارایی دارند. این مورد با مقایسه شکل (۵) که تمام ویژگی‌ها را مورد استفاده قرار داده و شکل‌های (۶) و (۷) که به ترتیب، دو و چهار ویژگی اضافه را استفاده کرده‌اند قابل مشاهده است. همچنین افزایش تعداد تخمین‌گرها تا حدی می‌تواند بهبود قابل توجه در کارایی را به دنبال داشته باشد و از آن مرحله به بعد می‌تواند باعث یکنواخت‌تر شدن عملکرد سیستم گردد. آزمایش‌های انجام‌شده‌ای که از رابطه (۱۳) برای محاسبه خطا استفاده کرده‌اند، با هدف نمایش کارایی روش پیشنهادی و مقایسه آن با روش‌های دیگر تلفیق طراحی شده‌اند. اگرچه رتبه‌بندی اسناد تا حد زیادی به کاربرد نیز وابسته است، در نتایج آزمایشی که در جدول (۱) نمایش داده شده، رتبه‌بندی با روش پیشنهادی و روش پایه مقایسه شده است. به این ترتیب در این آزمایش‌ها دو روش برای مقایسه رتبه‌بندی‌ها معرفی شده است که می‌تواند به منظور تحلیل هر روش ارائه‌شده برای این منظور مورد استفاده قرار گیرد.

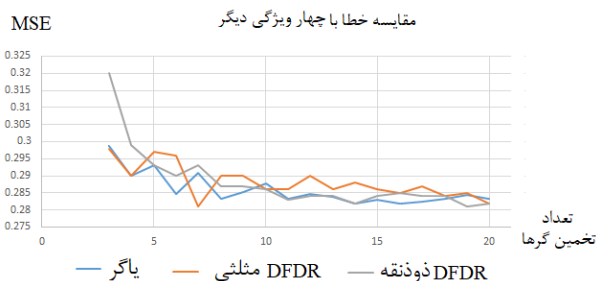
۵. نتیجه‌گیری

در این مقاله، یک روش برای رتبه‌بندی اسناد در یک شبکه استنادی با استفاده از تلفیق داده‌ها ارائه شده است. این روش مبتنی بر تعریف اعداد فازی بوده و از مجموعه‌ای تخمین‌گر برای این منظور استفاده کرده است. مجموعه داده استفاده‌شده شامل مجموعه‌ای از اسناد است که هر سند به مجموعه‌ای از اسناد دیگر ارجاعاتی دارد و خود نیز می‌تواند توسط

بررسی می‌کند. در این آزمایش ویژگی‌های authority, modularity نیز به ویژگی‌های آزمایش شکل (۶) افزوده شده است. نتیجه این آزمایش که با استفاده از این چهار ویژگی انجام شده، در شکل (۷) مشاهده می‌شود.



شکل (۶): مقایسه خطای روش‌های تلفیق با استفاده از دو ویژگی دیگر



شکل (۷): مقایسه خطای روش‌های تلفیق با استفاده از چهار ویژگی دیگر

با مقایسه شکل (۶) و (۷) مشاهده می‌شود که با افزودن این دو ویژگی، بهبود اندکی در سیستم و کاهش خطا صورت گرفته است. همچنین با به‌کارگیری تعداد شش تخمین‌گر حداقل خطا به دست آمده و افزودن تعداد تخمین‌گرها باعث همگرا شدن عملکرد این روش‌ها شده است.

همچنین در آزمایش بعد که نتایج آن در شکل (۸) مشاهده می‌شود، میانگین خطای هرکدام از تخمین‌گرها به کمک ترکیب ۶ ویژگی نشان داده شده است. اگرچه در آزمایش‌ها و نمودارهای گذشته اثر ترکیب داده‌ها با معیارهای مختلف بررسی شد، حتی با اضافه کردن یک ویژگی جدید دیگر به معیارهای قبلی می‌توان سیر کاهشی و سپس حالت پایدار میانگین خطا را مشاهده کرد. مطابق الگوهای قبلی نیز تا تعداد ۹ تخمین‌گر همچنان کاهش قابل توجهی در خطا مشاهده می‌شود و از آن به بعد، روش‌ها تا حدی همگرا می‌شوند.

مجموعه داده‌های مناسب و کارهای محدود انجام شده در این حوزه برای انجام آزمایش‌ها و مقایسه‌ها و وابسته بودن رتبه‌بندی به نوع کاربرد، از یک مجموعه داده برای آزمایش‌ها استفاده شده است. در کارهای آینده پیشنهاد می‌شود مجموعه داده‌های جامع‌تری برای آزمایش‌ها در حوزه‌های مختلف تهیه شود که امکان استخراج ویژگی‌های متنوع‌تر و جامع‌تری را فراهم کند. این ویژگی‌ها می‌تواند شامل ویژگی‌های نویسندگان و اطلاعات پروفایل آن‌ها و موارد دیگر باشد. همچنین به کارگیری روش‌های دیگر تلفیق داده‌ها و تأثیر آن‌ها می‌تواند مفید باشد.

مجموعه‌ای از اسناد دیگر مورد ارجاع قرار گیرد. در روش ارائه شده، معیارهایی تعریف شده‌اند که می‌توانند بر مبنای ارجاعات مستقیم و غیرمستقیم اسناد را ارزش‌گذاری کنند. به‌منظور نمایش قابلیت روش پیشنهادی و امکان مقایسه، دو روش معرفی شده است. یکی با مقایسه رتبه‌بندی انجام شده توسط روش پیشنهادی و مبنای مقایسه که رتبه‌بندی انجام شده توسط گفنی است. در روش دیگر، میزان خطا در رتبه‌بندی این دو روش بر مبنای میانگین مربعات خطاست. البته این مورد اخیر به‌منظور انجام تحلیل‌ها روی روش پیشنهادی ارائه شده است. با توجه به محدودیت‌های موجود در دسترسی به

مراجع

- [1] Newman M., "Coauthorship networks and patterns of scientific collaboration", *proceedings of the national academy of sciences*, vol. 101, pp. 5200-5205, 2004.
- [2] Nan M., Jiancheng G. and Yi Z., "Bringing PageRank to the citation analysis", *information processing and management*, vol. 44, pp. 800-810, 2008.
- [3] Paul P., Kumar V., Chudhury P. and Nandy S., "Temporal analysis of author ranking using citation-collaboration network", in *communication systems and networks*, pp. 1-6, 2015.
- [4] Christopher D., Pateman B. and Alaster K., "A citation network analysis of lithic microwear reaserch", vol. 7, pp. 54-62, 2018.
- [5] Massucci F. A. and Docampo D., "Measuring the academic reputation through citation networks via pagerank", *informetrics*, vol. 13, pp. 185-201, 2019.
- [6] Yo-lin H., James N. and Xi-Zhao W., "Owa operator based link prediction ensemble for social network", *expert systems with applications*, vol. 42, no. 1, pp. 21-50, 2015.
- [7] Marriana M. and Emrouznejad A., "The value of indirect ties in citation networks: SNA analysis with OWA operator weights", *Inf science*, vol. 115, 2015.
- [8] Gonzalez D. G., "Combined social networks and data envelopment analysis for ranking", *European Journal of Operational Research*, 2017.
- [9] Murali K., Rama Koti Reddy D. V. and Mulaveesala R., "Application of image fusion for the IR images in frequency modulated thermal wave imaging for Non Destructive Testing (NDT)", *Materials Today: Proceedings*, vol. 5, pp. 544-549, 2018.
- [10] Merigo J. M., Gil-Lafuente A. M., Yu D. and Llopis-Albert C., "Fuzzy decision making in complex frameworks with generalized aggregation operators", *Applied Soft Computing*, 2018.
- [11] Kuncheva L., *Combining Pattern Classifiers*, John Wiley & Sons, 2014.
- [12] Taylor C. N. and Bishop A. N., "Homogeneous Functionals and Bayesian Data Fusion with Unknown Correlation", *Information Fusion*, 2018.
- [13] Yager R., "An ordered weighted Averaging Aggregation operators in multicriteria Decision making", *inform sci*, vol. 31, pp. 107-139, 1998.
- [14] Emrouznejad A. and Amin G., "Improving minimax disparity model to determine the OWA operator weights", *Information science*, vol. 180, pp. 1477-1485, 2010.
- [15] Xu Z., "An Overview of Methods for Determining OWA Weights", *International Journal of Intelligent Syatems*, vol. 20, pp. 843-865, 2005.
- [16] Ligeng W. and Sifeng L., "A Model to Determine Owa Weights and Its Application in Energy Technology Evaluation", *international journal of inteeligent systems*, vol. 30, pp. 798-806, 2015.
- [17] Fuller R. and Majlender P., "An analytic approach for obtaining maximal entropy OWA operator weights", *Fuzzy sets and systems*, vol. 136, pp. 203-215, 2003.
- [18] Ailem M., Role F. and Nadif M., "Graph Modularity Maximization as an Effective Method for Co-clustering Text Data", *Knowledge-Based Systems*, 2016.
- [19] Changat M., Narasimha-Shenoi P. G. and Seethakuttyamma G., "Betweenness in graphs: A short survey on shortest and induced path betweenness", *AKCE International Journal of Graphs and Combinatorics*, 2018.
- [20] Salkhordeh Haghighi M., Vahedian A. and Sadoghi Yazdi H., "Creating and measuring diversity in multiple classifier systems using support vector data description", *Applied Soft Computing*, vol. 11, no. 8, pp. 4931-4942, 2011.
- [21] Gehrke J., Ginsparg P. and Kleinberg J. M., "High-energy physics citation network", Stanford University, 2003.
- [22] Wang Y. and Parkan C., "A minimax disparity approach for obtaining OWA operator weights", *Information science*, vol. 175, pp. 3356-3363, 2005.