ارائه راهکاری خودکار بر اساس متن‌کاوی برای شناخت و تحلیل روند تحقیقات حوزه‌های علمی

دو فصلنامه علمی- پژوهشی

مدیریت اطلاعات

دوره ۴، شماره ۱

بهار و تابستان ۱۳۹۷

اشکان خطیر

دانشجوی دکتری مهندسی فناوری اطلاعات، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، تهران، ایران

آزاده محبی

استادیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، تهران، ایران[1]

سهیل گنجه‌فر

استاد مهمان پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، استاد تمام دانشگاه بوعلی سینا، همدان، ایران

چکیده: بررسی روند تحقیقات یک حوزه علمی در بازه‌های زمانی مختلف می‌تواند درک بهتری را برای محققین و سیاست‌گذاران آن حوزه ایجاد نماید تا بتوانند برنامه‌ریزی مناسبی را جهت انجام تحقیقات آتی و تخصیص منابع پژوهشی داشته باشند. یکی از مهم‌ترین رویکردها در تحلیل روند تحقیقات در یک حوزه، بررسی اسناد علمی منتشرشده در آن حوزه با استفاده از روش‌های علم‌سنجی و پیمایش اطلاعات و متون اسناد است؛ بنابراین تحلیل روند، ابزار مناسبی برای محققین و سیاست‌گذاران در اجرای فعالیت‌های آن‌ها است. ازاین‌رو انتخاب روشی مناسب برای تحلیل وضعیت فعلی و پیش‌بینی آن حائز اهمیت است. با توجه به این نکته که دقت و جامعیت تحلیل روند از اهمیت ویژگی‌ای برخوردار است در این پژوهش رویکردی ارائه‌شده که با استفاده از روش‌های متن‌کاوی و اطلاعات کتابشناختی مقالات منتشرشده در یک حوزه، روند پژوهش در آن حوزه موردمطالعه قرار می‌گیرد. در این پژوهش کلمات کلیدی استخراج‌شده از متون با استفاده از یک روش جدید برای محاسبه هم‌رخدادی، خوشه‌بندی می‌شوند. از ویژگی‌های روش پیشنهادی ارائه شاخص جدید در میزان به دست آوردن بلوغ و مرکزیت یک حوزه علمی در تحلیل روند و استفاده از میزان تأثیر و اهمیت کلیدواژه‌های استفاده‌شده در تشخیص حوزه‌های علمی است. برای بررسی و آزمایش روش پیشنهادی، مجموعه‌ای از مقالات حوزه مهندسی مکانیک طی سال‌های ۲۰۱۲ تا ۲۰۱۶ از پایگاه «وب آف ساینس» استخراج‌شده است. با به‌کارگیری شاخص‌های پیشنهادی می‌توان نشان داد که برخی حوزه‌ها به بلوغ خود رسیده‌اند و دیگر روند رو به رشدی ندارند. از طرفی دیگر خوشه‌هایی که با نرخ رشد بالایی در حال رشد هستند و هنوز ازنظر میزان بلوغ در میانه راه قرار دارند، نشان‌دهنده موضوع‌های در حال تکامل هستند.

کلیدواژه‌ها: بلوغ حوزه علمی، تحلیل روند، ماتریس هم‌رخدادی کلمات، مرکزیت حوزه علمی.

۱. نویسنده مسئول: mohebi@irandoc.ac.ir

مقدمه

انجام هر نوع فعالیت علمی و عملی در یک حوزه خاص نیازمند آن است که درک و فهم مناسبی از پژوهش‌های انجام‌شده در آن حوزه و روند تحقیقات آن حاصل شود. به‌طورمعمول سیاست‌گذاران، محققین و پژوهشگران در ابتدای شروع فعالیت خود در یک حوزه علمی، بامطالعه آثار پژوهشگران گذشته سعی بر درک، روند حرکت و وضعیت جاری آن حوزه دارند. بر اساس میزان و عمق مطالعه، درک نسبی از چهارچوب کلی و وضعیت علمی جاری پیدا می‌شود. در راستای این درک، پژوهشگران می‌توانند از آخرین وضعیت پژوهشی علمی یک حوزه درک مناسبی پیداکرده و با در نظر گرفتن روند حرکت آن حوزه، موضوع‌های مناسب پژوهش را برگزینند. سیاست‌گذار نیز با درک مناسب و دریافت نقشه علمی از وضعیت علمی و فناوری مربوطه می‌تواند با دیدی باز آینده علمی آن حوزه را ترسیم کرده و در خصوص آن سیاست‌های لازم را اتخاذ نموده و سرمایه‌گذاری‌های لازم را انجام دهد. ازاین‌رو تحلیل روند یک حوزه، بر روی عملکرد پژوهشگران و سیاست‌گذاران آن حوزه تأثیرگذار خواهد بود. تحلیل روند یک تحلیل فنی است که به‌منظور پیش‌بینی روند و حرکت مقوله‌ای مانند فناوری با استفاده از داده‌های گذشته انجام می‌شود (Wu et al. 2011).

علاوه بر تحلیل روند، شناخت زیرحوزه‌های یک‌رشته علمی و آگاهی از محدوده‌های تحقیقاتی آن بسیار بااهمیت است. تحلیل روند می‌توان در شناخت کرانه‌های یک حوزه علمی و زیرشاخه‌های آن نیز مورداستفاده قرار گیرد (White et al. 2016). تحلیل روند روشی است که به دو منظور تحلیل رفتار گذشته و پیش‌بینی آینده مورداستفاده قرار می‌گیرد. برای انجام تحلیل روند بر اساس نوع داده‌های در دسترس، معمولاً تغییرات یک یا چند شاخص با استفاده از روش‌های ریاضی و آماری در دوره‌های زمانی متفاوت موردمطالعه و تحلیل قرار می‌گیرد و سعی بر آن است که این رفتار استخراج‌شده به آینده تعمیم داده شود (Wu et al. 2011; Porter 1991). یکی از روش‌های تحلیل روند استفاده از روش‌های آماری است (Larsen et al. 1999). برخلاف دیگر روش‌ها که بر اساس نظر متخصصان انجام می‌شود، مزیت روش‌های ریاضی و آماری با استفاده از کامپیوتر و الگوریتم‌های هوشمند تحلیل داده‌ها و اطلاعات است که می‌توانند حجم زیادی از مستندات علمی را بررسی و تحلیل نمایند. تکرارپذیری، مستقل بودن ازنظر اشخاص و دقت این روش‌ها، قابلیت اطمینان آن را در پیش‌بینی بالاتر می‌برد (White et al. 2016).

روش‌های مختلفی برای تحلیل روند یک حوزه علمی با استفاده از رویکرد آماری وجود دارد، طبق مطالعات انجام‌شده این روش‌ها را می‌توان به دودسته کلی تحلیل بر روی اسناد علمی (Kung et al. 2017; Müller et al. 2018) و تحلیل بر روی اختراعات انجام‌شده (Suh and Jeon 2018; Joung and Kim 2017) تقسیم کرد. هرکدام از این دسته‌ها می‌توانند بر اساس اطلاعات کتابشناختی و روش‌های متن‌کاوی مورد ارزیابی قرار گیرند. در تحلیل‌هایی که بر روی اختراعات ثبت‌شده انجام می‌شود، علومی بررسی می‌شوند که به فناوری رسیده باشند؛ بنابراین در این نوع تحلیل‌ها از جامعیت تحلیل کاسته می‌شود. در تحلیل‌هایی که بر روی اسناد علمی انجام می‌شود تمام اسناد مرتبط با یک حوزه (چه علومی که به فناوری رسیده‌اند و چه علومی که هنوز به فناوری نرسیده‌اند) استخراج و موردبررسی قرار داده می‌گیرند. در این نوع تحلیل‌ها، حجم زیاد اسناد ممکن است باعث افزایش پیچیدگی تحلیل شود

(Wu et al. 2011). ازآنجاکه در این مقاله هدف تحلیل روند حوزه‌های علمی است، تحلیل اختراعات که در آن علم تبدیل به فناوری شده، موردبررسی قرار نخواهد گرفت.

تحلیل‌های کتابشناختی نوعی از تحلیل‌های آماری هستند که بر روی اسناد منتشرشده ازجمله مقاله و یا کتاب انجام می‌شود. عموماً این تحلیل‌ها بر روی اطلاعاتی که از این اسناد استخراج می‌شوند انجام می‌شود. این اطلاعات شامل عنوان، چکیده، نویسنده(ها)، زبان سند، نوع سند، کشور تولیدکننده سند، سازمان تولیدکننده سند، استنادها و ... است. در این نوع از تحلیل اطلاعات آماری از داده‌های موجود به‌دست‌آمده و تحلیل‌ها بر روی این اطلاعات انجام می‌شود. در تحلیل‌های متن‌کاوی، اطلاعات و الگوهای پنهان از روی داده‌های متنی استخراج می‌شود (Ozaydin et al. 2017). در تحلیل روند حوزه علمی، تحلیل متن‌کاوی می‌تواند بر روی عنوان، چکیده، استنادها و یا ترکیبی از آن‌ها انجام گیرد.

در این مقاله رویکردی یکپارچه ارائه‌شده که روند رشد یک حوزه علمی را به‌صورت خودکار بررسی و پیش‌بینی کند. به‌منظور افزایش دقت، ترکیبی از روش‌های کتابشناختی و متن‌کاوی به همراه روش‌های جدیدی برای محاسبه شاخص تحلیل روند استفاده و ارائه‌شده است. روش‌های گوناگونی برای تحلیل روند ارائه‌شده است که در برخی از آن‌ها از ماتریس هم‌رخدادی استفاده می‌شود. با استفاده از ماتریس هم‌رخدادی می‌توان کلیدواژه‌های یک حوزه علمی را شناسایی و گروه‌بندی نمود (No and Park 2010; Guo et al. 2017; Wu et al. 2011; Callon et al. 1983).

روش‌های پیشین برای تشکیل ماتریس هم‌رخدادی از مجموع تعداد هم‌رخدادی کلیدواژه‌ها در اسناد مختلف به وجود آمده است؛ بنابراین از تکرار کلیدواژه‌ها که نشان‌دهنده میزان اهمیت کلمات است استفاده‌نشده است. این امر می‌تواند دقت ماتریس را کاهش دهد که در پی آن کاهش دقت در تشخیص روند و تحلیل‌های آتی است. در این پژوهش شاخص جدیدی برای تحلیل روند ارائه می‌شود که امکان در نظر گرفتن میزان هم‌رخدادی واژگان را نیز فراهم می‌کند. از طرفی دیگر، یکی از ابزارها برای تحلیل روند، نمودار استراتژیک است. برای تشکیل این نمودار نیز روش‌های مختلفی استفاده‌شده که در آن‌ها نیز فراوانی کلیدواژه‌های تشکیل‌دهنده خوشه‌ها به‌صورت مستقیم دخیل نشده‌اند. استفاده از تکرار کلیدواژه‌ها می‌تواند در تحلیل‌ها اهمیت حوزه‌ها را بیشتر لحاظ کند که در رویکرد پیشنهادی شاخصی بدین منظور نیز ارائه شده است.

در ادامه، در بخش پیشینه پژوهش، تحقیقات گذشته درباره تحلیل روند بررسی می‌شوند. پس‌ازآن با مشخص کردن اهداف و سؤالات پژوهش و روش پژوهش، رویکرد پیشنهادی تشریح می‌گردد. سپس نحوه عملکرد رویکرد پیشنهادی بر روی مجموعه‌ای از داده‌های علمی یک حوزه موردبررسی قرار می‌گیرد و در انتها نتیجه‌گیری و تحقیقات آتی ارائه می‌گردد.

پیشینه پژوهش

فعالیت‌های بسیاری برای تحلیل روند تحقیقات در یک حوزه انجام‌شده است. در سال ۲۰۰۷ «چاو[1]»، «یانگ[2]» و «جن[3]» روند حرکت فناوری RFID را در بازه زمانی ۱۵ سال بر روی مقاله‌های منتشرشده مجلات SCI[4] انجام داده است. در پژوهش ایشان، اسنادی که در این مقالات در عنوان و یا چکیده خود، واژه RFID و یا Radio Frequency Identification را بکار برده‌اند از پایگاه داده WoS[5] استخراج‌شده‌اند. این پژوهش تنها به تحلیل اطلاعات آماری بسنده کرده و روند حرکتی یک حوزه علمی را بررسی نکرده است؛ بنابراین امکان توانایی در تشخیص میزان بلوغ رشته‌ها و حوزه‌های تازه را ندارد. تحلیل‌های انجام‌شده عبارت‌اند از: میزان همکاری بین‌المللی، نویسندگی، زبانی، نوع سندی و تعداد استنادها.

«چوی[6]» و «پارک[7]» در سال ۲۰۰۹ مسیر حرکت فنّاوری ساختارهای ارگانیک[8] را بررسی کرده است. این تحلیل بر اساس هم‌استنادی اختراعات انجام‌شده است. بر این اساس اسناد با استفاده از روش خوشه‌بندی سلسله‌مراتبی، خوشه‌بندی‌شده‌اند. البته در این تحلیل از کلیدواژه‌ها که در حقیقت ویژگی اصلی تحلیل روند یک حوزه علمی به شمار می‌رود، استفاده‌نشده است. به دلیل عدم بررسی کلیدواژه‌های مقاله‌ها، مفاهیم و محتوای پنهان درون مقاله‌ها نیز بررسی نشده است و امکان کشف میزان بلوغ رشته‌ها به‌صورت دقیق ندارد.

«نو[9]» و «پارک[10]» در سال ۲۰۱۰ نفوذ و روند فناوری نانو را موردبررسی قرار داده است .پژوهش وی با استفاده از تحلیل هم‌استنادی بر روی اختراعات انجام‌شده است. در این پژوهش شاخصی تعریف‌شده که بر اساس آن با استفاده از استنادهای مستقیم میزان نفوذ فناوری‌ها روی یکدیگر نشان داده‌شده‌اند. این تحلیل از سه مرحله جمع‌آوری داده، تشکیل ماتریس هم‌رخدادی استنادی و دسته‌بندی آن‌ها و تحلیل نفوذ تشکیل‌شده است. در پژوهش آن‌ها، از نمایه‌هایی که برای هر اختراع ثبت‌شده، استفاده‌شده و محتوای متن اختراع و کلمات مهمی که در آن نویسنده از آن اختراع استفاده کرده، در نظر گرفته نشده است. از طرفی دیگر میزان دفعات رخداد کلیدواژه‌ها در محاسبه هم‌رخدادی لحاظ نشده است. لازم به ذکر است که تحلیل هم‌استنادی به دلیل عدم توجه به محتوای متن مقاله منجر به چشم‌پوشی از برخی از فناوری‌های نوظهور می‌شود.

1.Chao
2.Yang
3.Jen
4.Science Citation Index
5.Web of Science
6.Choi
7.Park
8.Organic
9.No
10.Park

در سال ۲۰۱۱، «لو[1]» و همکاران تحلیل کتابشناختی را بر روی مقالاتی که در ۵ سال متوالی بر حوزه گرافن[2] در WoS به چاپ رسیده انجام داده است. تحلیل انجام‌شده در دو مرحله استخراج اطلاعات از منابع داده و تحلیل‌های کتابشناختی انجام‌گرفته است. تحلیل‌های انجام‌شده با استفاده از نرم‌افزار تی.دی.ای[3] و تنها بر روی اطلاعات کتابشناختی بوده است؛ بنابراین اطلاعاتی که توسط نویسنده در چکیده و عنوان مستتر است، نادیده گرفته‌شده است. عدم استفاده از متن چکیده و عنوان باعث کاهش دقت در بدست آوردن میزان بلوغ رشته‌ها و حوزه‌هایی است که در متن به‌صورت ضمنی به آن‌ها اشاره‌شده است.

در سال ۲۰۱۱ «وو[4]» و همکاران برای تحلیل روند «اتچینگ[5]» یک روش سیستمی ارائه کرد و برای تحلیل روند از سه روش تحلیل کتابشناختی، تحلیل اختراع و متن‌کاوی استفاده نمود. برای تحلیل اختراع از نرم‌افزار دلفینون[6] و برای متن‌کاوی از نرم‌افزار تیسنگ[7] بهره گرفته‌شده است. در این پژوهش تنها از ۵۰ کلمه پرتکرار اول استفاده‌شده و مابقی کلمات نادیده گرفته‌شده‌اند. فراوانی کلمات هر خوشه با یکدیگر جمع شده و در بازه‌های زمانی یکسانی فراوانی‌های کل خوشه‌ها نمایش داده‌شده است. ازآنجاکه تحلیل متن‌کاوی هم بر روی عنوان و هم بر روی چکیده انجام‌شده، این امکان وجود دارد که کلماتی که در نوآوری مقاله نقشی ندارند و مربوط به فناوری‌های گذشته هستند و برای تعیین ارتباط فناوری‌های پیشین با یک پژوهش در متن چکیده آمده‌اند، در تحلیل‌ها اثر زیادی بگذارند و نتایج تحلیل را منحرف سازند؛ بنابراین استفاده از ۵۰ کلیدواژه نخست و استخراج و استفاده از کلمات پرتکرار حوزه‌های پایه و نادیده دیگر کلیدواژه‌ها، باعث کاهش دقت در تحلیل روند و عدم کشف حوزه‌های جدید خواهد شد.

تحلیل دیگری برای روند سلول‌های بنیادی با استفاده از هم‌رخدادی کلمات و وزن دهی به عنوان‌های موضوعی[8] انجام‌شده است (An and Wu 2011). در این تحلیل که بر مبنای روش نیمه‌خودکار است با استفاده از آنتروپی اطلاعات کلمات بااهمیت استخراج‌شده و با استفاده ازنظر متخصصان حوزه برای استخراج عناوین موضوعی استفاده‌شده است. تحلیل انجام‌شده با استفاده از نمودار استراتژیک صورت گرفته است. از طرفی دیگر میزان دفعات رخداد کلیدواژه‌ها در محاسبه هم‌رخدادی لحاظ نشده است و تنها از دو وزن یک و دو برای عنوان‌های موضوعی استفاده‌شده است. استفاده از متخصصان در انتخاب عناوین موضوعی باعث کندی در استخراج عناوین و جهت‌گیری این عنوان‌ها در تحلیل‌ها می‌شود.

1.Lv
2.Graphene
3.Thomson Data Analyzer (TDA)
4.Wu
5.Etching
6.Delphion
7.Tseng
8.Subject heading

در سال ۲۰۱۳ «هو[1]» و همکاران در پژوهش خود به تحلیل مقالات علم اطلاعات پرداخته است. با استفاده از روش خوشه‌بندی سلسله‌مراتبی و با فرض اینکه هر خوشه یک زیرشاخه‌ای از آن حوزه است و استفاده از نمودار استراتژیک، روند حوزه علم اطلاعات در طول پنج سال موردبررسی قرارگرفته است. در تحلیل درون خوشه‌ای استراتژیک مورداستفاده تعداد ارتباطات بین کلیدواژه‌ها شمارش‌شده که ارزش و وزن خود کلمه صرف‌نظر شده است. در همان سال «صمدی کوچکسرایی[2]»، «حسن‌زاده[3]» و «شکرانه[4]» روند تحقیقات سلول‌های بنیادی در کشور ایران و قاره‌های اروپا و امریکا و در منطقه خاورمیانه را تنها با استفاده از اطلاعات آماری مقالات چاپ‌شده در پاب‌مد[5] موردبررسی قرار داده‌اند. البته در این تحلیل روشی برای بدست آوردن زیرحوزه‌ها و سطح بلوغ رشته‌ها ارائه نشده است.

علاوه بر تحقیقات فوق، با استفاده از تحلیل هم‌رخدادی و نمودار استراتژیک در سال ۲۰۱۵ تحلیل روند در خصوص میزان تمرکز، بلوغ و توسعه تحقیقات بر روی «سیستم‌های توصیه‌کننده» انجام‌شده است (Hu and Zhang 2015). استفاده از تعداد لینک‌ها در نمودار استراتژیک باعث شده که میزان تأثیر وزن کلمات در خوشه در نظر گرفته نشود که این امر باعث صرف‌نظر کردن اهمیت کلیدواژه‌ها و کاهش دقت تحلیل می‌شود. «چن[6]» و همکاران نیز در سال ۲۰۱۶ با استفاده از نرم‌افزار «وی.او.اس ویوئر»[7] تحلیل هم‌رخدادی کلمات و مطالعه روابط بین موضوع‌های پروژه‌های « Management Science and Engineering» در دوره پنج ساله در کشور چین انجام‌شده است. روشی برای انتخاب تعداد کلیدواژه‌های مورد تحلیل و روابط بین خوشه‌ها ارائه نشده و مورد تحلیل قرار نگرفته است. از طرفی دیگر میزان دفعات رخداد کلیدواژه‌ها در محاسبه هم‌رخدادی لحاظ نشده است؛ که لحاظ کردن این تکرار می‌تواند میزان اهمیت و وابستگی دو کلیدواژه را نسبت به یکدیگر نشان دهد.

«چانگ[8]» و همکاران در سال ۲۰۱۷ برای پیدا کردن موضوع‌های تحقیق‌های موردتوجه دانشجویان دندان‌پزشکی از تحلیل هم‌رخدادی کلمات استفاده کرده است. وی این کار را با استفاده از نرم‌افزار «بیب اکسل»[9] و استخراج کلیدواژه‌های با فرکانس بالاتر و تحلیل خوشه‌بندی انجام داده است. کلیدواژه‌های پرتکرار از رابطه ارائه‌شده در مقاله استخراج شدند. در انتها تنها با ترسیم شکلی از خوشه‌بندی کلیدواژه‌های پرتکرار موضوع‌های موردعلاقه نمایش داده‌شده است.

«ژائو[10]» و همکاران نیز در سال ۲۰۱۸ برای تحلیل روند از نمودار استراتژیک و شبکه‌های اجتماعی استفاده کرده است. درروش پیشنهادی ایشان از کلیدواژه‌هایی که بیش از ۲۰ بار تکرار شده‌اند

1.Hu
2.Samadikuchaksaraei
3.Hassanzadeh
4.Shokraneh
5.PubMED
6.Chen
7.VOSviewer
8.Chang
9.BIbExcel
10.Zhao

استفاده‌شده که البته این شیوه انتخاب می‌تواند منجر به نادیده گرفتن کلیدواژه‌های مهم و یا واردکردن کلیدواژه‌هایی غیر مرتبط شود.

همان‌طور که مطالعه تحقیقات گذشته نشان داد روش‌های گوناگونی بر روی حوزه‌های مختلف علمی برای تحلیل روند ارائه‌شده است. عموم این روش‌ها از هم‌رخدادی کلمات برای تحلیل روند و بررسی وضعیت استفاده کرده‌اند. در این روش‌ها پارامترهای آماری ساده از روی فراداده‌های مقالات استخراج‌شده و این پارامترها در بازه‌های زمانی مختلف تحلیل‌شده‌اند. در برخی پژوهش‌های دیگر بعد از بررسی وضعیت آماری با استفاده از تحلیل هم‌رخدادی کلمات (رخداد دو کلمه در یک متن)، خوشه‌بندی و نمودار استراتژیک (میزان ارتباط بین خوشه‌ها و کلمات داخل هر خوشه)، روند حوزه‌ها بررسی‌شده‌اند.

بررسی پژوهش‌های گذشته نشان داد که در مطالعات گذشته، در تشکیل ماتریس هم‌رخدادی تنها رخداد حضور (عدم حضور) دو کلمه در یک متن در نظر گرفته‌شده است و میزان هم‌رخدادی کلمات در یک متن که نشان‌دهنده میزان اهمیت یک کلمه است لحاظ نشده است. به‌عنوان‌مثال اگر دو کلمه هر دو در یک متن یک‌بار تکرار شده باشند یا بیش از یک‌بار، میزان هم‌رخدادی آن‌ها متفاوت نخواهد بود. درصورتی‌که دو کلمه که در یک متن چند بار باهم آمده باشند، ارتباط نزدیک‌تری دارند نسبت به دو کلمه که در همان متن تنها یک‌بار تکرار شده‌اند. توجه به این‌گونه پارامترها می‌تواند میزان اهمیت کلیدواژه‌ها را در تحلیل‌های بعدی ازجمله خوشه‌بندی و بلوغ حوزه‌های علمی، بالاتر ببرد.

از طرفی دیگر در تحلیل‌های انجام‌شده در نمودار استراتژیک به تعداد لینک‌های ارتباطی بین خوشه‌ها و کلمات داخل خوشه توجه شده، لیکن فراوانی تکرار کلمات هر خوشه در میزان چگالی خوشه موردتوجه قرار نگرفته است. در بخش بعدی به بیان روشی خواهیم پرداخت که می‌تواند روش‌های موجود در تحقیقات گذشته را با در نظر گرفتن شاخص‌های جدیدی بهبود دهد.

همان‌طور که بیان شد استفاده از تحلیل روند می‌تواند درک بهتری را برای محققین و سیاست‌گذاران آن حوزه ایجاد نماید تا بتوانند برنامه‌ریزی مناسبی را جهت انجام تحقیقات آتی و تخصیص منابع پژوهشی داشته باشند. روش‌های مختلفی برای تحلیل روند ارائه‌شده که هرکدام نقاط قوت و ضعفی داشتند. هدف از این پژوهش ارائه روشی است تا بتواند با حفظ جامعیت از نقاط قوت روش‌های پیشین استفاده کند. علاوه بر آن سعی شده است در مواردی که خلائی در روش‌های تحلیل روند وجود داشته باشد، با استفاده از روش‌های علمی اثبات‌شده آن خلأها برطرف شود. ازجمله اهداف دیگر این پژوهش ارائه شاخصی برای تحلیل مؤثرتر نمودار استراتژیک است.

سؤال‌هایی که این پژوهش در صدد پاسخ به آنها است عبارتند از:

- شاخص‌های ارزیابی برای تحلیل روند چه هستند و چه مواردی را پوشش می‌دهند؟
- آیا روشی برای برطرف کردن معایب روش‌های پیشین می‌توان ارائه داد؟
- آیا می‌توان شاخص‌هایی ارائه کرد که گستره بیشتری را برای تحلیل روند شامل شود؟

روش‌شناسی پژوهش

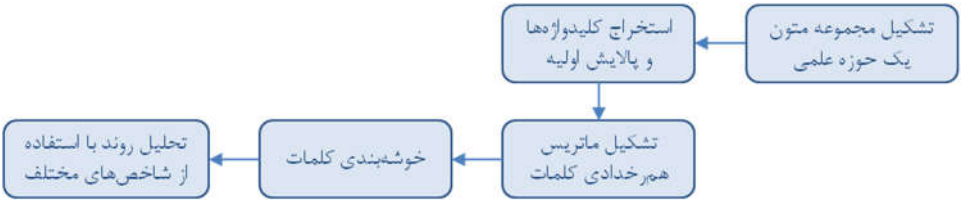
هدف اصلی پژوهش حاضر ارائه روشی است که از طریق آن بتوان روند تحقیقات یک حوزه علمی را بررسی و تحلیل نمود. برای این منظور از روش کتابخانه‌ای استفاده‌شده و تحقیقات گذشته در این خصوص موردمطالعه قرارگرفته است. سپس برمبنای نقاط ضعف و برخی از چالش‌های شناسایی‌شده، روش جدید ارائه‌شده است. بررسی پیشینه تحقیق نشان داده است که یکی از راه‌ها برای تحلیل روند استفاده از روش هم‌رخدادی است. اصول کلی در تحلیل هم‌رخدادی به این صورت است که در ابتدا حوزه‌ای انتخاب می‌شود و اسناد مرتب با آن حوزه استخراج می‌شوند. در ادامه کلیدواژه‌هایی از اسناد استخراج‌شده و پس از جدا کردن کلیدواژه‌های با تکرار بیشتر، ماتریس هم‌رخدادی تشکیل می‌شود. در انتها نیز تحلیل‌هایی بر روی ماتریس هم‌رخدادی انجام می‌شود. در این مقاله بر مبنای تحلیل هم‌رخدادی روش جدیدی پیشنهادشده است که در ادامه گام‌های اصلی این روش به تشریح بیان‌شده است. همچنین برای ارزیابی روش پیشنهادی از روش «ارزیابی تطبیقی[1]» (Vartiainen 2002) استفاده‌شده است. بدین‌صورت که مؤلفه‌هایی که برای روش تحلیل روند وجود دارد، به‌عنوان معیارهای ارزیابی در نظر گرفته می‌شود و در هر گام از تحلیل روند، روش پیشنهادی با رویه‌های رقیب مقایسه می‌شوند و نشان داده می‌شود که روش پیشنهادی از جامعیت و کارایی بهتری برخوردار است.

بنابراین برای ارزیابی نحوه عملکرد روش پیشنهادی، یک حوزه علمی موردبررسی قرارگرفته است. حوزه انتخابی در این مقاله، مجموعه مقالات حوزه مکانیک بوده که در سال‌های ۲۰۱۲ تا ۲۰۱۶ در پایگاه داده WoS ثبت‌شده‌اند.

روش پیشنهادی برای تحلیل روند علمی

روش پیشنهادی در این پژوهش از مراحل مختلفی تشکیل‌شده است. پس از انتخاب حوزه تحقیقاتی موردنظر، مستندات علمی مرتبط با آن حوزه از پایگاه‌های علمی معتبر مانند «وب آف ساینس» استخراج می‌شود. در گام بعدی، با استفاده از روش‌های خودکار تحلیل متن، کلیدواژه‌ها از متن مستندات علمی استخراج می‌گردد. سپس این کلیدواژه‌ها، با کلیدواژه‌های استاندارد اسناد که توسط نمایه‌ساز یا نویسنده سند به هر سند علمی تخصیص‌یافته، ترکیب می‌شود تا فهرست بزرگ‌تری از کلیدواژه‌ها حاصل شود. پس از پالایش فهرست حاصل‌شده، کلیدواژه‌هایی که اهمیت بیشتری دارند، از فهرست استخراج می‌شوند و مجموعه کلیدواژه‌های نهایی را می‌سازند. با استفاده از مجموعه کلیدواژه‌های نهایی ماتریس هم‌رخدادی تشکیل‌شده و خوشه‌بندی صورت می‌گیرد. درنهایت نیز تحلیل روند بر اساس کلیدواژه‌های خوشه‌بندی‌شده، شاخص‌های ارائه‌شده و نمودار استراتژیک، انجام می‌شود. گام‌های اصلی روش پیشنهادی در شکل یک نمایش داده‌شده است. در ادامه هر یک از این گام‌ها تشریح می‌شوند.

1.Comparative Evaluation (CE)



شکل ۱. گام‌های روش پیشنهادی برای تحلیل روند یک حوزه علمی

۱-۱. تشکیل مجموعه متون

برای تحلیل روند یک حوزه علمی، اولین مرحله استخراج و تشکیل مجموعه‌ای از متون مربوط به آن حوزه است. برای این منظور معمولاً با بهره‌گیری از دانش متخصصین آن حوزه و استفاده از طبقه‌بندی موضوعی پایگاه‌های علمی معتبر مانند «وب آف ساینس»، کلیدواژه‌هایی انتخاب می‌شود. سپس با انتخاب محدوده سال‌هایی که هدف تحلیل هستند، اسناد و متون علمی از این پایگاه داده استخراج خواهد شد. مسلماً از بین اطلاعات دریافت شده تنها مقالاتی انتخاب می‌شوند که دارای کلیدواژه نویسنده و نمایه‌ساز، عنوان و چکیده باشند. ازآنجاکه با استفاده از کلیدواژه‌های عنوان و کلیدواژه‌های نویسنده می‌توان به هدف، نوآوری و موضوع اصلی متن پی برده شود (Khatir and Ganjefar 2016)، در گام بعدی کلیدواژه‌های مهم استفاده‌شده در اطلاعات کتابشناختی استخراج خواهند شد.

۱-۲. استخراج کلیدواژه‌ها و پالایش اولیه

برای استخراج کلیدواژه‌ها از یک متن گام‌های زیر انجام می‌شود:

- نرمال‌سازی متن؛
- برچسب‌زنی کلمات و انتخاب اسم‌ها به‌عنوان کلمات کاندید؛
- استخراج کلمات کاندید و ریشه‌یابی آن‌ها؛
- فیلترینگ کلمات بر اساس نوع واژه، تکرار و فهرست کلمات-توقف[1]؛
- تشکیل فهرستی از کلیدواژه‌های استخراج‌شده.

ازآنجاکه احتمال وجود کلیدواژه‌ها در یک سند علمی در عنوان آن بیشتر از چکیده آن است و کلیدواژه‌های موجود در عنوان نماینده نوآوری آن سند علمی است (Khatir and Ganjefar 2016) در این پژوهش پیشنهادشده است که علاوه بر کلیدواژه‌های نویسنده مقاله و نمایه‌سازها از کلیدواژه‌های موجود در عنوان سند علمی نیز استفاده شود. در نرمال‌سازی تمام کلمات ازنظر نگارش و نوع کاراکتر به شکل استانداردی[2] درمی‌آیند که ممکن است قبلاً به آن صورت نبوده‌اند (برای مثال ازنظر نوع ذخیره کاراکترها به‌صورت دیجیتالی[3] کاراکتری در سیستم‌های ذخیره‌سازی با کدهای مختلفی ذخیره‌شده باشند

1.Stop-word List
2.Canonical Form
3.Character Encoding

که در نرمال‌سازی همه کاراکترهای با یک کد خاص ذخیره و نمایش داده می‌شوند). هر سند علمی معمولاً دارای حداقل اطلاعات کتابشناختی زیر است:

- عنوان؛
- چکیده؛
- کلیدواژه‌های نویسنده یا کلیدواژه‌های نمایه‌ساز.

بنابراین اگر D مجموعه کل اسناد باشد، برای هر سند $d_i \in D$ مجموعه‌ای از کلیدواژه‌های آن سند (K_i) استخراج می‌شود که به‌صورت زیر نمایش می‌دهیم:

$$K_i = \{k_1, k_2, \dots, k_{n_i}\}$$

که در آن، هر k_j یک کلیدواژه است و n_i بیانگر تعداد کلیدواژه‌های سند d_i است. بعد از به دست آوردن کلیدواژه‌های کاندید از عناوین اسناد، این لیست با فهرستی از کلیدواژه‌های نویسنده و نمایه‌ساز، ادغام می‌شود. ازآنجاکه تحلیل روند تنها با استفاده از کلیدواژه‌های استخراج‌شده از عنوان و کلیدواژه‌های نویسنده و نمایه‌ساز صورت می‌پذیرد، احتمال رخداد واژه‌هایی که مرتبط با نوآوری سند نیست در لیست کلیدواژه‌ها اندک است.

با تجمیع کلیه کلیدواژه‌های اسناد موجود در D، به مجموعه‌ای از کلیدواژه‌ها دست پیدا می‌کنیم. درنهایت با حذف برخی از کلیدواژه‌های کم تکرار در این مجموعه، مجموعه نهایی کلیدواژه‌ها برای کلیه اسناد D حاصل می‌گردد که آن را K^* می‌نامیم. برای حذف کلیدواژه‌های کم تکرار به‌صورت زیر عمل می‌کنیم:

۱. محاسبه فراوانی کلیدواژه‌ها در کلیه اطلاعات کتابشناختی برای همه اسناد (عناوین، چکیده‌ها و کلیدواژه‌های تخصیص داده‌شده توسط نویسنده و نمایه‌ساز)؛
۲. مرتب کردن کلیدواژه‌ها بر اساس فراوانی به‌صورت نزولی؛
۳. رسم نمودار فراوانی کلیدواژه‌ها؛
۴. انتخاب نقطه شکست (جایی که نرخ کاهش نمودار به‌صورت قابل‌توجهی کم می‌شود) کلیدواژه‌ها
۵. انتخاب کلیدواژه‌هایی که بالاتر از نقطه شکست قرار دارند به‌عنوان کلیدواژه‌های بااهمیت‌تر.

۳-۱. تشکیل ماتریس هم‌رخدادی

ماتریس هم‌رخدادی عموماً در ترسیم نقشه علمی و نقشه دانشی مورداستفاده قرار می‌گیرد و بر اساس رخداد مجموعه‌ای از کلمات تشکیل می‌شود (Leydesdorff and Nerghes 2017; Zhang et al. 2015; He 1999). در ماتریس هم‌رخدادی هر سطر و یا ستون بیانگر میزان هم‌رخدادی یک کلمه با سایر کلمات است. در پژوهش‌های گذشته هم‌رخدادی کلمات عموماً برحسب رخداد دو کلمه در یک مقاله یا در چکیده آن مقاله محاسبه‌شده است (An and Wu 2011; He 1999; Rokaya et al. 2008).

در این پژوهش میزان هم‌رخدادی دو کلمه بر اساس احتمال وقوع دو کلمه در یک متن پیشنهاد می‌شود. برای مؤثر نمودن میزان فراوانی کلمات، احتمال وقوع دو کلمه را به میزان احتمال وقوع هریک از

دو کلمه در مقالات تقسیم کرده‌ایم. این عمل باعث می‌شود تا کلماتی که هم‌رخدادی یکسانی دارند، کلماتی که در اسناد کمتری تکرار شده‌اند نسبت به کلماتی که در اسناد بیشتری تکرار شده‌اند، امتیاز بیشتری داشته باشند و ضریب بالاتری را به خود اختصاص دهند. از طرف دیگر درروش پیشنهادی در محاسبه میزان هم‌رخدادی دو کلمه، تعداد تکرار هر یک از آن دو کلمه، به‌صورت مجزا، در یک سند نیز لحاظ شده است که این امر باعث مؤثر شدن میزان اهمیت یک کلمه در یک سند است. درواقع هم‌رخدادی دو کلمه، در حالتی که دو کلمه بیش از یک‌بار در یک سند تکرار شده باشند نسبت به حالتی که تنها یک‌بار در سند آمده باشند، بیشتر خواهد بود.

بنابراین اگر رویداد مربوط به رخداد کلیدواژه k_i و رویداد مربوط به رخداد کلیدواژه k_j را در نظر بگیریم، آنگاه درایه ماتریس هم‌رخدادی C_{ij} که معادل احتمال رخداد این دو کلیدواژه باهم است، به‌صورت زیر محاسبه می‌شود:

$$C_{ij} = \frac{P(k_i \cap k_j)}{P(k_i \cup k_j)} \quad (۱)$$

که در آن $P(k_i \cup k_j)$ احتمال رخداد هریک از کلیدواژه‌ها در اسناد موردبررسی، و $P(k_i \cap k_j)$ به معنی احتمال رخداد دو کلیدواژه k_i و k_j باهم در یک سند است. احتمال رخداد دو واژه باهم در یک سند، بر اساس تعریف احتمال شرطی به‌صورت زیر است:

$$P(k_i \cap k_j) = P(k_i|k_j) * P(k_j) \quad (۲)$$

احتمال شرطی $P(k_i|k_j)$ به این مفهوم است که با فرض اینکه k_j در یک سند مشاهده‌شده، احتمال اینکه k_i هم در همان سند مشاهده شود، چقدر است. بنابراین اگر D_i مجموعه کلیه اسنادی باشد که k_j را در بر دارد، آنگاه این احتمال شرطی را به‌صورت زیر محاسبه می‌کنیم:

$$P(k_i|k_j) = \frac{\sum_{d \in D_i} \min\left(freq(k_i, d), freq(k_j, d)\right)}{\sum_{d \in D_i}\left(freq(k_j, d)\right)} \quad (۳)$$

منظور از $freq(k_j|d)$ میزان تکرار کلمه k_j در سند d است. برای محاسبه احتمال رخداد یک کلمه به‌شرط رخ دادن کلمه دیگر (یعنی رابطه (۳)) در یک سند، از تقسیم حداقل تعداد تکرار هر دو کلمه بر روی کل کلمات در آن سند استفاده‌شده است. برای مثال اگر کلمه اول، سه بار در سند آمده باشد و کلمه دوم ۴ بار در آن سند آمده باشد و تعداد کل کلمات آن سند ۱۰۰ کلمه باشند نتیجه رابطه (۳) برابر با $\frac{3}{100}$ خواهد بود. این بدان علت است که تعداد هم‌رخدادی دو کلمه هم‌زمان نباید از تکرار هر یک از آن‌ها بیشتر باشد بنابراین در رابطه (۳) حداقل تکرار دو کلمه لحاظ شده است. احتمال رخداد یک کلمه یعنی $P(k_j)$ از رابطه (۴) محاسبه می‌شود که در آن مخرج کسر نشان‌دهنده تعداد کلمات سند است:

$$P(k_j) = \frac{\sum_{i \in D} freq(k_i, d)}{\sum_{j \in Corpus} |D_j|} \quad (۴)$$

از حاصل‌ضرب رابطه (۳) و (۴) می‌توان به این نتیجه رسید که احتمال هم‌رخدادی دو کلمه k_j و k_i به‌صورت زیر است:

$$P(k_i \cap k_j) = \frac{\sum_{i \in D} \min\left(freq(k_i, k_j)\right)}{\sum_{j \in Corpus} |D_j|} \quad (۵)$$

ازآنجاکه مخرج کسر وابسته به کلمات k_j و k_i نیست و برای تمام کسرها یک مقدار دارد می‌توان برای محاسبه احتمال اشتراک دوکلمه‌ای از آن صرف‌نظر کرد؛ بنابراین این احتمال تنها متناسب با صورت کسر به‌صورت رابطه (۵) در نظر گرفته می‌شود:

$$P(k_i \cap k_j) \propto \sum_{d \in D} \min\left(freq(k_i, d), freq(k_j, d)\right) \quad (۶)$$

از طرفی دیگر برای به دست آوردن $P(A \cup B)$ از رابطه (۷) استفاده خواهیم کرد:

$$P(k_i \cup k_j) = P(k_i) + P(k_j) - P(k_i \cap k_j) = \frac{\sum_{i \in D} freq(k_i, d) + \sum_{i \in D} freq(k_j, d) - \sum_{d \in D_i} \min\left(freq(k_i, d), freq(k_j, d)\right)}{\sum_{d \in D_i} \left(freq(k_j, d)\right)} \quad (۷)$$

بنابراین هر درایه ماتریس هم‌رخدادی، یعنی هم‌رخدادی دو کلمه k_j و k_i، به‌صورت زیر محاسبه می‌شود:

$$C_{ij} = \frac{\sum_{i \in D} \min\left(freq(A_i, B_i)\right)}{\sum_{i \in D} freq(k_i|d) + \sum_{i \in D} freq(k_{ij}, d) - \sum_{d \in D_i} \min\left(freq(k_i, d), freq(k_j, d)\right)} \quad (۸)$$

۴-۱. خوشه‌بندی

پس از تشکیل ماتریس هم‌رخدادی، برای تحلیل روند، کلماتی که باهم در ارتباط هستند را با استفاده از خوشه‌بندی در یک مجموعه قرارمی‌دهیم. در این گام از روش‌های خوشه‌بندی مختلفی می‌توان استفاده کرد. در اینجا از یک روش خوشه‌بندی معمول و ساده با عنوان کا-میانه[1] (Jain 2010) استفاده می‌کنیم. از این روش در به دست آوردن روند حوزه‌ها نیز استفاده می‌شود (Tao et al. 2017).

در خوشه‌بندی کلیدواژه‌ها به دسته‌هایی که اعضای آن مشابه یکدیگر هستند، تقسیم می‌شوند. یکی از پارامترهایی که در خوشه‌بندی تأثیر فراوان دارد تعداد خوشه‌ها است. یکی از روش‌ها برای تعیین

1.K-Means

خودکار تعداد خوشه‌ها، بررسی عملکرد خوشه‌بندی بر اساس معیارهای ارزیابی خوشه‌بندی است (Dimitriadou, Dolničar, and Weingessel 2002). برای ارزیابی خوشه‌ها پارامترهای گوناگونی وجود دارد که می‌توان به «شاخص دون»[1]، «شاخص دی‌بی»[2]، «شاخص اس.اس.ای»[3] و «شاخص سیلوئت»[4] اشاره کرد (Rousseeuw 1987; Davies and Bouldin 1979; Dunn 1973). در این پژوهش، برای تعیین خودکار تعداد خوشه‌ها، از معیارهای ارزیابی خوشه‌بندی اشاره‌شده، استفاده می‌کنیم.

برای این منظور خوشه‌بندی را برای تعداد خوشه‌های مختلف (یک تا حداکثر تعداد کلیدواژه‌ها) به‌دفعات انجام می‌دهیم. سپس، عملکرد خوشه‌بندی را بر اساس معیارهای فوق می‌سنجیم. تعداد خوشه مناسب زمانی اتفاق می‌افتد که مقادیر مناسبی برای معیارهای فوق حاصل‌شده باشد.

۵-۱. تحلیل روند

برای تحلیل روند بر اساس نتایج خوشه‌بندی کلمات، دو جنبه ارتباط بین خوشه‌ای (بین‌حوزه‌ای) و تغییرات درون خوشه‌ای (درون‌حوزه‌ای) است. بدین منظور شاخص‌هایی پیشنهادشده است. علاوه بر آن، ازآنجاکه یکی از ابزارهای تحلیل روند استفاده از نمودار استراتژیک است، برای تحلیل‌های فوق نیز از این نمودار کمک گرفته خواهد شد. این نمودار دارای دو بعد محوریت و چگالی است. خوشه‌ها پس از خوشه‌بندی بر روی این نمودار قرار می‌گیرند.

- **شاخص‌های ارزیابی**

درروش پیشنهادی پس از خوشه‌بندی کلمات، شاخص‌های تحلیل روند به دو صورت زیر بکار گرفته می‌شوند:

۱. تحلیل بین خوشه‌ای: ازآنجاکه کلمات قرارگرفته در خوشه‌ها ازنظر معنایی باهم در ارتباط هستند، می‌توان نتیجه گرفت که هر خوشه بیانگر یک زیرحوزه علمی است؛ بنابراین با تحلیل هر یک از خوشه‌ها می‌توان بررسی‌های بیشتری درباره زیرحوزه‌ها انجام داد. برای این منظور می‌توان با شاخص‌های خوشه‌ها را بررسی نمود:

الف. روند رشد زیرحوزه (خوشه): با بررسی روند رشد یک زیرحوزه میزان محبوبیت آن در بین محققین و پژوهشگران بررسی می‌شود و مشخص می‌شود که در طول سال‌های متوالی چه تغییراتی بر روی آن صورت پذیرفته است. رشد زیرحوزه‌ها را می‌توان بر اساس شاخصی با ترکیب میزان تکرار هم‌رخدادی کلمات بین حوزه‌های مختلف و میزان ارتباط آن‌ها موردبررسی قرارگرفته داد.

1.Dunn Index
2.DB Index
3.SSE Index
4.Silhouette Index

ب. نرخ رشد زیرحوزه (خوشه): با استفاده از نرخ رشد، میزان افزایش محبوبیت یک زیرحوزه نسبت به سال‌های گذشته نشان داده می‌شود.

ج. روند رشد هر زیرحوزه به‌صورت نرمال شده: ازآنجاکه در اکثر مواقع در سال‌های متوالی میزان مقالات منتشرشده متفاوت است، با نرمال کردن وزن هر خوشه به تعداد مقالات منتشرشده در آن خوشه در آن سال، میزان توجه (فراوانی کلیدواژه‌ها) به آن خوشه (زیرحوزه) به‌صورت نرمال شده به دست خواهد آمد.

د. زیرحوزه‌های موردتوجه: با استفاده از میزان ارتباط بین زیرحوزه‌ها در نمودار استراتژیک که از طریق هم‌رخدادی کلیدواژه‌ها محاسبه می‌شود، می‌توان زیرحوزه‌هایی که دارای مرکزیت بیشتری نسبت به زیرحوزه‌های دیگر هستند را تشخیص داد. از این روش می‌توان در تشخیص زیرحوزه‌های میان‌رشته‌ای نیز استفاده کرد. در این پژوهش ترکیب میزان هم‌رخدادی دوکلمه‌ای که در یک خوشه قرار ندارند در بدست آوردن میزان توجه به حوزه‌ها لحاظ گردیده است.

۲. درون خوشه‌ای: هرچند که کلمات در هر خوشه ارتباط معنایی با یکدیگر دارند، اما تحلیل کلیدواژه‌ها در هر خوشه کلمات نیز دارای معنی است. تحلیل‌های زیر برای تک کلیدواژه‌ها می‌تواند در نظر گرفته شود:

الف. کلماتی که بیشترین میزان تأثیر را در فراوانی خوشه دارند. این کلمات می‌توانند زیرمجموعه‌های اصلی که در یک زیرحوزه اصلی است را نشان دهند.

کلماتی که بیشترین رشد را در خوشه دارند. این کلمات می‌توانند زیرمجموعه‌هایی که در آن زیرحوزه موردتوجه قرار می‌گیرد را نشان دهند یا به‌اصطلاح حوزه‌های در حال ظهور را نشان می‌دهند.

ب. بلوغ یک زیرحوزه: با استفاده از نمودار استراتژیک می‌توان تا حدودی میزان بلوغ یک زیرحوزه را بدست آورد. این امر با استفاده از میزان اتصالات درونی یک خوشه محاسبه خواهد شد. ازآنجاکه میزان تکرار و استفاده کلمات در میزان بلوغ حوزه‌ها نیز تأثیر دارند و در مطالعات گذشته این مقدار در نظر گرفته نشده، در این پژوهش ترکیب میزان تکرار هم‌رخدادی کلمات و میزان اتصالات درون خوشه برای نشان دادن میزان بلوغ یک زیرحوزه پیشنهاد می‌شود.

ج. کلماتی که بیشترین کاهش فراوانی را دارند به‌منظور تشخیص زیرحوزه‌های قدیمی و در حال انقراض.

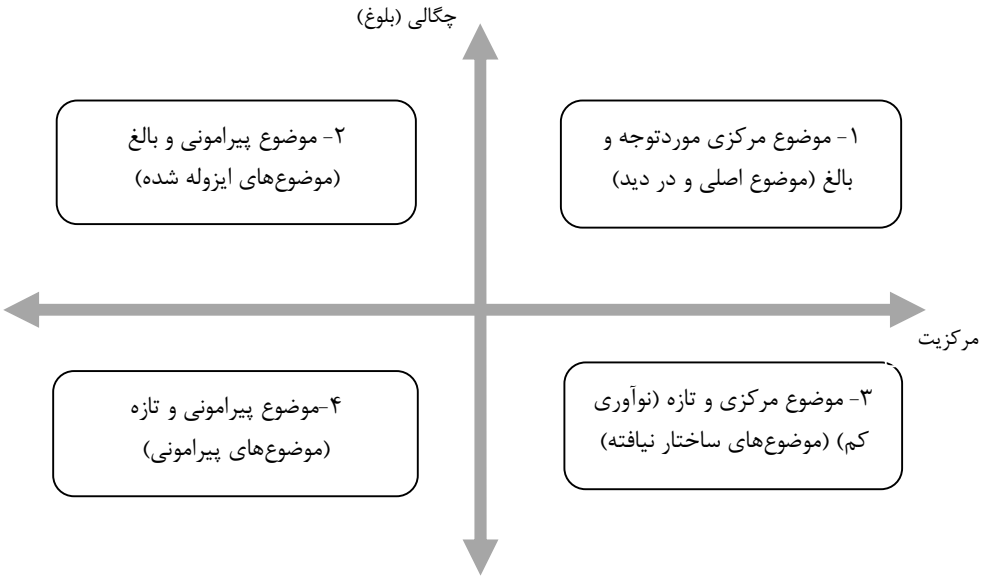
۳. تحلیل روند: مجموع تکرار کلمات (استفاده کلمات) هر زیرحوزه (خوشه) در یک بازه زمانی مشخص، نشان‌دهنده میزان عملکرد آن زیرحوزه در آن بازه زمانی است. با تشکیل هر دسته و محاسبه فراوانی کلمات آن دسته در بازه‌های زمانی معین و متفاوت می‌توان روند رشد و حرکت آن زیرحوزه را در طول زمان تحلیل کرد.

- **نمودار استراتژیک**

با استفاده از نمودار استراتژیک تحلیل ساختار و تغییرات روند یک حوزه تحقیقاتی انجام می‌شود. نمودار استراتژیک اولین بار در سال ۱۹۸۸ توسط «لا[1]» و همکاران ارائه شد. این نمودار به‌صورت کلی روابط درون یک حوزه و ارتباط بین زیرحوزه‌های مختلف آن را نشان می‌دهد. در این نمودار دو بُعد محوریت و چگالی وجود دارد و مکان هر خوشه براساس مقدار این دو معیارها در نمودار مشخص می‌شود (Delecroix and Epstein 2004).

میزان مرکزیت یک خوشه (زیرحوزه) مرجعیت بیشتر آن خوشه است. میزان مرکزیت یک خوشه را می‌توان با استفاده از ریشه دوم مجموع مربعات ارتباطات خروجی یا میانگین مقادیر شش لینک اول محاسبه کرد. امتیازدهی به خوشه‌ها بر اساس مرکزیت آن‌ها می‌تواند نشان دهد که در کل شبکه تحقیق کدام زیرحوزه‌ها مرکزیت بیشتری دارند (Shen, Li, and Gu 2013).

چگالی یا همبستگی درونی یک خوشه، قدرت روابط کلماتی که خوشه را می‌سازند، نشان می‌دهد. هرچه این مقدار بیشتر باشد نشان می‌دهد که این خوشه چه مقدار به بلوغ و تکامل رسیده است. در نمودار استراتژیک محور عمودی نشان‌دهنده میزان چگالی و محور افقی میزان است (An and Wu 2011; Shen, Li, and Gu 2013). در شکل دو ساختار اصلی نمودار استراتژیک نمایش داده‌شده است.



شکل ۲. نمودار استراتژیک

1.Law

در این پژوهش، برای تحلیل روند پیشنهاد می‌شود که پس از خوشه‌بندی، مرکزیت و چگالی هر خوشه محاسبه‌شده و برای تحلیل بهتر بر روی نمودار استراتژیک ترسیم شوند. اندازه چگالی هر خوشه بر اساس رابطه (۹) و مرکزیت خوشه‌ها بر اساس رابطه (۱۰) محاسبه می‌شود. برای بدست آوردن مرکزیت از میانگین مجموع وزن‌های ارتباطات بین کلیدواژه‌های متعلق به خوشه موردنظر با کلیدواژه‌هایی که متعلق به آن خوشه نیستند استفاده‌شده است. برای محاسبه چگالی خوشه از میانگین وزن هر کلیدواژه خوشه و مجموعه وزن‌های ارتباطات کلیدواژه‌های همان خوشه استفاده‌شده است.

$$\text{Density}_{\text{Cluster}_k} = \frac{\sum_{i,j\in\text{Cluster}_k} \text{WeightOfLink}_{ij} + \sum_{i\in\text{Cluster}_k} \text{Frequency}_i}{\sum \text{MemberOfCluster}_k} \quad (۹)$$

$$\text{Centerality}_{\text{Cluster}_k} = \frac{\sum_{\substack{i\in\text{Cluster}_k, \\ j\in\text{Cluster}_m}} \text{WeightOfLink}_{ij}}{\sum \text{NOTMemberOfCluster}_k} \quad (۱۰)$$

مقایسه تطبیقی روش پیشنهادی

در این بخش بر اساس روش ارزیابی تطبیقی، مقایسه‌ای بین روش‌های پیشین با روش پیشنهادی از جنبه‌های مختلف ارائه می‌گردد. همان‌طور که در بخش‌های قبل بیان شد، روش پیشنهادی یک روش جامع خودکار است که با استفاده از بهبود روش‌های پیشین برای تحلیل روند ارائه‌شده است. مطالعات نشان داده است که تحلیل روند از مراحل: استخراج کلیدواژه، انتخاب کلیدواژه، تشکیل ماتریس هم‌رخدادی، خوشه‌بندی، تحلیل روند تشکیل‌شده است. جدول یک مقایسه‌ای بین روش‌های ارائه‌شده پیشین و روش ارائه‌شده در این مقاله را در مراحل مختلف نشان می‌دهد.

جدول ۱. مقایسه بین مراحل مختلف تحلیل روند در روش پیشنهادی و روش‌های پیشین

مراحل اجرا / روش تحلیل روند	استخراج کلیدواژه	انتخاب تعداد کلیدواژه	تشکیل ماتریس هم‌رخدادی	خوشه‌بندی	تحلیل روند
An and Wu 2011	استفاده از کلیدواژه‌های هر سند علمی و نظر متخصص	انتخاب توسط متخصصان	عدم بکار گیری اهمیت کلمات در یک سند	سلسله مراتبی	بر اساس شاخص میزان ارتباط کلیدواژه‌ها
Hu and Zhang 2015	استفاده از کلیدواژه‌های هر سند علمی	انتخاب تعدادی ثابت از فهرست کلیدواژه‌ها	عدم بکار گیری اهمیت کلمات در یک سند	سلسله مراتبی	بر اساس شاخص میزان ارتباط کلیدواژه‌ها

مراحل اجرا / روش تحلیل روند	استخراج کلیدواژه	انتخاب تعداد کلیدواژه	تشکیل ماتریس هم‌رخدادی	خوشه‌بندی	تحلیل روند
Chen et al. 2016	استفاده از کلیدواژه‌های هر سند علمی	انتخاب تعدادی ثابت از فهرست کلیدواژه‌ها	عدم بکار گیری اهمیت کلمات در یک سند	نرم‌افزار «وی او اس ویوئر»	ندارد
Chang et al. 2017	استفاده از کلیدواژه‌های هر سند علمی	به‌صورت پویا	عدم بکار گیری اهمیت کلمات در یک سند	سلسله مراتبی	ندارد
Zhao et al. 2018	استفاده از کلیدواژه‌های هر سند علمی	انتخاب کلیدواژه‌ها با فراوانی بیش از ۲۰	عدم بکار گیری اهمیت کلمات در یک سند	سلسله مراتبی	بر اساس شاخص میزان ارتباط کلیدواژه‌ها
روش پیشنهادی	استفاده از کلیدواژه‌های هر سند علمی و استخراج خودکار کلیدواژه از عنوان مقالات	به‌صورت پویا و بر اساس نمودار فراوانی	بر اساس احتمال رخداد دو کلمه با ترکیب میزان اهمیت کلیدواژه‌ها و میزان هم‌رخدادی	الگوریتم کا-میانه با ترکیب چند شاخص	استفاده از نمودارهای روند، تحلیل نمودار استراتژیک با ارائه شاخص ترکیبی و شاخص میزان اهمیت کلیدواژه‌ها و ارتباط آن‌ها

نتایج تجربی

در این بخش، نتایج به‌کارگیری روش پیشنهادی برای خوشه‌بندی کلیدواژه‌ها و تحلیل‌های پیشنهادشده، جهت تحلیل روند تحقیقات در یک حوزه علمی بر روی مجموعه‌ای از اسناد در آن حوزه نمایش داده‌شده است. در ابتدای این بخش، مجموعه اسناد مورداستفاده تشریح می‌گردد و در ادامه انواع تحلیل‌های پیشنهادشده برای بررسی روند تحقیقات نیز معرفی می‌گردد.

۶-۱. معرفی مجموعه داده اسناد علمی

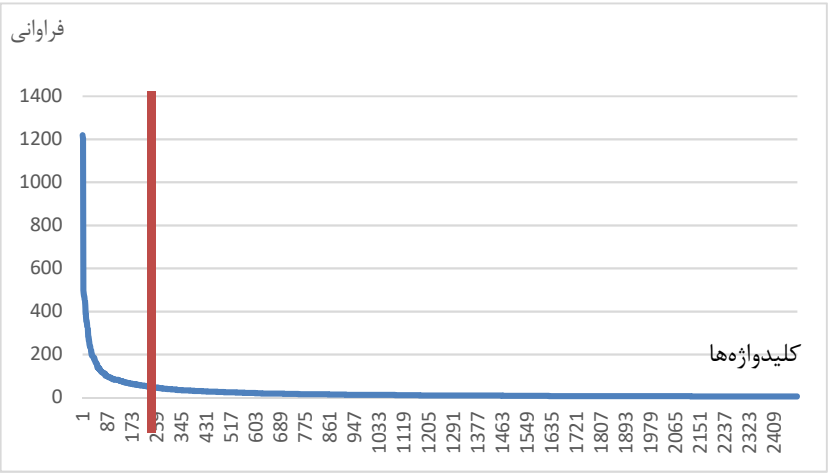
همان‌طور که در بخش‌های قبل بیان شد، برای بررسی بیشتر روش پیشنهادی در این پژوهش روند تحقیقات بر روی مقالات حوزه مهندسی مکانیک منتشرشده در «وب آف ساینس» در سال‌های ۲۰۱۲ تا ۲۰۱۶ موردبررسی قرارگرفته است. بدین منظور، مقاله‌هایی که در دسته‌بندی «وب آف ساینس» در زمره مقاله‌های مهندسی مکانیک وجود داشته، انتخاب‌شده است و اطلاعات «فرا داده‌ای»[1] تمام مقاله‌هایی که را که در سال‌های ۲۰۱۲ تا ۲۰۱۶ به چاپ رسیده‌اند را به‌عنوان مجموعه اسناد موردمطالعه انتخاب کرده‌ایم. بر اساس روش پیشنهادی، از بین اطلاعات فراداده‌ای، از عنوان، چکیده، کلیدواژه‌ها (نویسنده و نمایه‌ساز) و سال انتشار استفاده‌شده است. بر همین اساس تعداد ۴۱۴۵ عدد مقاله استخراج‌شده است. با حذف مقاله‌هایی که عنوان، کلیدواژه و یا سال چاپ ندارند، تعداد مقالات به ۴۰۲۸ کاهش پیداکرده است. در جدول ۱، تعداد مقالات به تفکیک هرسال آمده است.

جدول ۲. توزیع مقالات در هرسال

سال	۲۰۱۲	۲۰۱۳	۲۰۱۴	۲۰۱۵	۲۰۱۶	همه‌سال‌ها
تعداد	۵۰۷	۹۴۱	۵۸۴	۱۲۱۲	۷۸۴	۴۰۲۸

بعد از طی مراحل استخراج کلیدواژه که در بخش قبل بیان شد، کلیدواژه‌های عنوان استخراج‌شده و با فهرست کلیدواژه‌های انتخاب‌شده توسط نویسنده و نمایه‌ساز ترکیب‌شده است. تعداد کلیدواژه‌های استخراج‌شده از عنوان ۵۱۶۹ مورد، کلیدواژه‌های نمایه‌ساز ۷۰۵۱ و کلیدواژه‌های انتخاب‌شده توسط نویسنده ۸۹۴۳ هستند. پس از ترکیب این سه دسته از کلیدواژه‌ها، تعداد کل کلیدواژه‌های کاندید استفاده‌شده به‌صورت یکتا به ۱۷۷۲۲ عدد رسیده است. همان‌طور که گفته شد فراوانی اکثر این کلیدواژه‌ها بسیار اندک است و کلیدواژه‌هایی که فراوانی اندکی دارند تأثیری در تحلیل‌های آتی نخواند داشت. درنتیجه لازم است که کلیدواژه‌های کم تأثیر را حذف کرد. پس از پالایه کردن بر اساس معیارهای بیان‌شده در بخش قبل (مانند حذف کلمات توقف که در عنوان وجود دارند و تعیین مقدار زانو که در شکل سه نشان داده‌شده است)، کلیدواژه‌های نهایی که برای تشکیل ماتریس هم‌رخدادی و خوشه‌بندی بدست آمده‌اند، به ۲۹۳ رسیده است.

1.Meta-Data

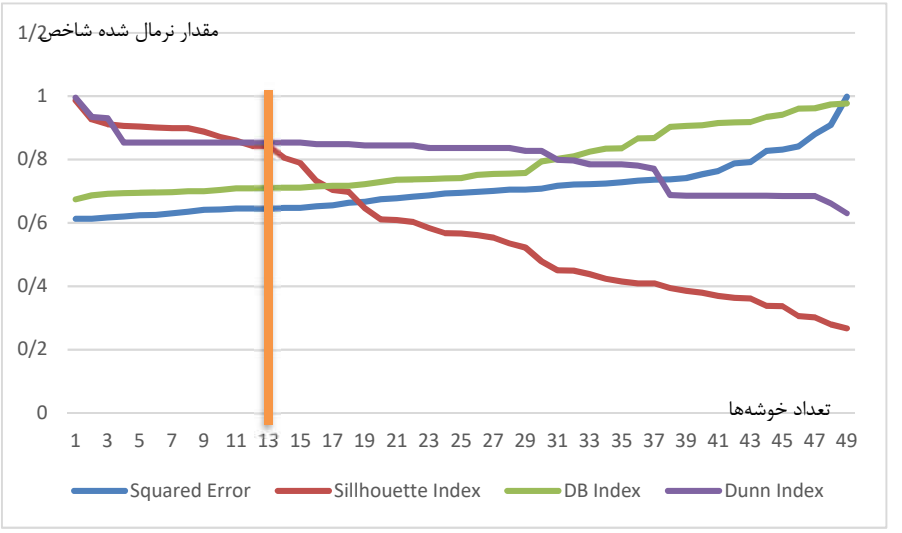


شکل ۳. انتخاب کلیدواژه‌ها بر اساس مقدار زانو در نمودار فراوانی کلیدواژه‌ها

با استفاده از کلیدواژه‌های نهایی و شیوه محاسبه درایه‌های ماتریس هم‌رخدادی و با به‌کارگیری روش خوشه‌بندی کا-میانه، کلمات خوشه‌بندی‌شده‌اند. برای یافتن تعداد بهینه خوشه‌ها از معیارهای «شاخص دون»، «شاخص دی‌بی»، «شاخص اس.اس.ای» و «شاخص سیلوئت» استفاده‌شده است.

برای این منظور خوشه‌بندی را برای تعداد خوشه‌های مختلف (یک تا حداکثر تعداد کلیدواژه‌ها) به‌دفعات انجام داده‌ایم که نتایج، در شکل ۴ چهار نشان داده‌شده است. بر اساس این شکل، تعداد خوشه ۱۳ مقداری است که به ازای آن شاخص‌های ارائه‌شده مقدار مناسبی را دارند. شاخص‌های سیلوئت و دون در حال کاهش سریع (هرچه مقدار این دو شاخص بیشتر باشد خوشه‌بندی بهتر است) و شاخص‌های دی‌بی و اس.اس.ای در حال افزایش (هرچقدر این دو شاخص کمتر باشند، خوشه‌بندی بهتر است) هستند. در شکل ۴ چهار با استفاده از خط عمودی این امر نشان داده‌شده است.

همان‌طور که در بخش قبل بیان شد نمودار استراتژیک به‌منظور تحلیل روند مورداستفاده قرار می‌گیرد. پس از خوشه‌بندی، ۱۳ زیرحوزه بدست آمده از خوشه‌بندی بر اساس میزان بلوغ (چگالی) و بر اساس میزان مرکزیت بر روی این نمودار قرار می‌گیرند. در شکل ۱ شکل پنج خوشه‌های تشکیل‌شده از مرحله قبل بر روی نمودار استراتژیک نمایش داده‌شده است. در این شکل خوشه‌ها با نقاط آبی نشان داده‌شده‌اند و در برچسب تعیین‌شده برای هر خوشه یک و یا چند کلمه پرتکرار آن خوشه برای معرفی آن خوشه، به نمایندگی نوشته‌شده است.

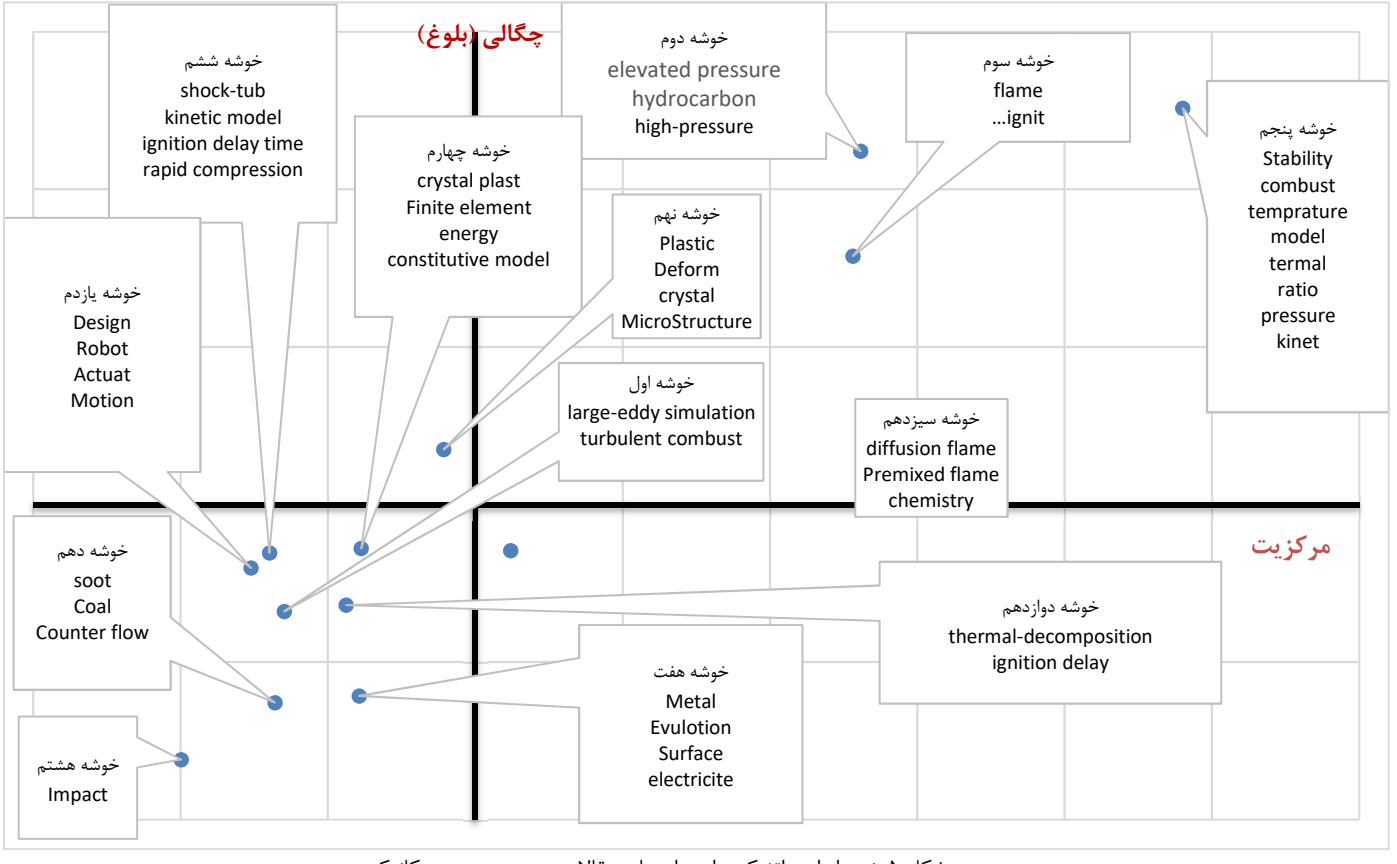


شکل ۴. شاخص‌های خوشه‌بندی

۱-۱. تحلیل درون خوشه

هر یک از خوشه‌های نمایش داده‌شده در شکل پنج نشان‌دهنده یک زیرحوزه تحقیقاتی در حوزه مهندسی مکانیک هستند. هر خوشه می‌تواند از یک و یا چندین کلمه تشکیل‌شده باشد. مجموعه کلمات داخل هر خوشه معرف نوع حوزه آن خوشه هستند. از طرفی دیگر در داخل هر خوشه کلمات بکار گرفته‌شده به‌تنهایی خود نشان‌دهنده زیرمجموعه‌های مورداستفاده در آن خوشه (زیرحوزه) هستند. برای مثال در خوشه شماره نهم زمینه‌های تحقیقاتی کریستال، پلاستیک و ریز ساخت‌ها وجود دارند.

ازآنجاکه هر کلمه در یک خوشه نماینده یک زیرمجموعه کاری می‌تواند باشد، برای بدست آوردن زیرحوزه‌های در حال ظهور می‌بایست روند رشد کلمات آن خوشه را موردبررسی قرار داد. بدین منظور روند کلمات در طی سال‌های متوالی موردبررسی قرار می‌گیرند. کلماتی که در طی سال‌های متوالی روند رو به رشدی داشته‌اند زیرحوزه‌های بالقوه‌ای می‌توانند باشند که در آینده با رشد بیشتر می‌توانند به یک حوزه مستقل تبدیل شوند. برای این منظور بعد از محاسبه نرخ رشد کلمات در سال‌های مذکور کلماتی که بیشترین میانگین نرخ را طی چند سال داشته‌اند به‌عنوان کلمات کاندید برای زیرحوزه‌ای جدید معرفی می‌شود. جدول دو برخی کلماتی را که در سال‌های اخیر نرخ رشد بالایی داشته‌اند، نشان می‌دهد. شکل شش روند رشد کلمات انتخاب‌شده در جدول دو را نشان می‌دهد. همان‌طور که مشاهده می‌شود این کلمات در سال‌های متوالی روند رو به رشدی داشته‌اند که می‌تواند حاکی از به وجود آمدن حوزه‌ای در رابطه با آن‌ها باشند.

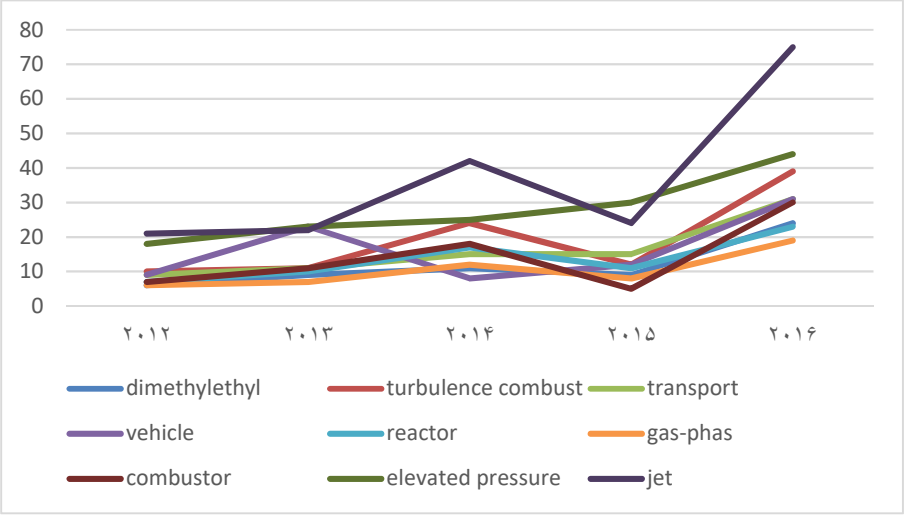


شکل ۵. نمودار استراتژیک برای داده‌های مقالات حوزه مهندسی مکانیک

برای بدست آوردن زیرحوزه‌هایی که درگذر زمان توجه کمتری به آن‌ها شده است، با بدست آوردن نرخ رشد حوزه‌ها در سال‌های متوالی، کلیدواژه‌هایی که نرخ رشد منفی دارند جزء زیرحوزه‌هایی قدیمی که دیگر توجهی به آن‌ها نمی‌شوند قرار خواهند گرفت. برای مثال همان‌طور که در جدول سه مشاهده می‌شود، کلمات motor (موتور)، sensor (حس‌گر)، wave (موج) و gas (گاز) عموماً در سال‌های مختلف نرخ رشدهای منفی داشته‌اند.

جدول ۳. برخی کلمات با بیشترین نرخ رشد

کلمه	نرخ رشد از سال ۲۰۱۲ تا ۲۰۱۶	نرخ رشد از سال ۲۰۱۳ تا ۲۰۱۶	نرخ رشد از سال ۲۰۱۴ تا ۲۰۱۶	نرخ رشد از سال ۲۰۱۵ تا ۲۰۱۶
dimethylethyl	۹	۷,۵	۶,۵	۷,۵
turbulence combust	۱۴,۵	۱۴	۷,۵	۱۳,۵
transport	۱۱	۱۰	۸	۸
vehicle	۱۱	۴	۱۱,۵	۹,۵
reactor	۸	۶,۵	۳	۶
gas-phas	۶,۵	۶	۳,۵	۵,۵
combustor	۱۱,۵	۹,۵	۶	۱۲,۵
elevatedpressur	۱۳	۱۰,۵	۹,۵	۷
jet	۲۷	۲۶,۵	۱۶,۵	۲۵,۵



شکل ۶. کلمات کاندید که در بیشترین نرخ‌های رشد بوده‌اند

جدول ۴. نرخ کاهش توجه به برخی کلیدواژه‌ها (زیرحوزه‌ها)

نرخ رشد / کلمه	سال ۲۰۱۲ تا سال ۲۰۱۶	سال ۲۰۱۳ تا سال ۲۰۱۶	سال ۲۰۱۴ تا سال ۲۰۱۶	سال ۲۰۱۵ تا سال ۲۰۱۶
diffusion	-۷	-۵	-۳۵	-۳۲
chemic	-۷	-۴	۰	-۱۱
motor	-۷	-۱۱	-۴	-۱۵
wave	-۱۰	۱۰	-۱۲	۴
decomposition	-۳۲	۰	-۹	۱
gas	-۲۶	۳	۰	-۴۱
sensor	-۶	-۱۴	-۱	-۱۷

یکی دیگر از پارامترهای موردبررسی در تحلیل روند، میزان بلوغ یک حوزه است. میزان بلوغ یک حوزه یکی از پارامترهایی است که در نمودار استراتژیک مورداستفاده قرارگرفته است. برای محاسبه میزان بلوغ یک حوزه هم‌رخدادی کلیدواژه‌هایی که در یک حوزه مورداستفاده قرارگرفته، باهم جمع شده و برای یکسان‌سازی حوزه‌های مختلف، این مقدار بر تعداد لینک‌های موجود در آن حوزه تقسیم می‌شود. با استفاده از رابطه (۹) و (۱۰) مقادیر مربوط به مرکزیت و چگالی خوشه‌ها محاسبه‌شده‌اند.

جدول ۵. مقادیر چگالی و مرکزیت خوشه‌ها

شماره خوشه	مرکزیت خوشه‌ها نرمال شده بر اساس لینک‌ها	چگالی خوشه‌ها نرمال شده بر اساس لینک‌ها
۱	۱,۱۸	۴,۸۰
۲	۵,۰۹	۱۹,۴۰
۳	۵,۰۳	۱۶,۰۸
۴	۱,۷	۶,۸۰
۵	۷,۲۷	۲۰,۷۷
۶	۱,۰۷	۶,۶۵
۷	۱,۶۸	۲,۱۲
۸	۰,۴۸	۰,۱۰
۹	۲,۲۶	۹,۹۴
۱۰	۱,۱۱	۱,۹۱
۱۱	۰,۹۵	۶,۱۸
۱۲	۱,۵۹	۵,۰۰
۱۳	۲,۷۱	۶,۷۳

۲-۱. تحلیل برون خوشه‌ای

همان‌طور که در بخش قبل بیان شد هر خوشه نماینده یک حوزه است که آن را در بازه‌های زمانی تحلیل می‌شود. ازآنجاکه میزان انتشار مقالات در سال‌های مختلف متفاوت است، برای ارزیابی، وزن محاسبه‌شده به خوشه‌ها بر اساس تعداد اسناد نرمال شده است. **Error! Reference source not found.**پنج نشان می‌دهد که در طی سال‌های متوالی روند رشد هر خوشه در حوزه مهندسی مکانیک چگونه بوده است.

جدول ۶. روند رشد خوشه‌ها در سال‌های متوالی بر اساس میزان کلیدواژه‌های هر خوشه در حوزه مهندسی مکانیک

شماره خوشه / سال انتشار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳
۲۰۱۲	۳۷	۲۱	۲۳۹	۳۰	۳۴۹	۲۶	۵۶	۱	۶۶	۶۷	۷۲	۵	۳۳
۲۰۱۳	۲۷	۲۵	۲۴۳	۲۷	۳۳۶	۲۷	۶۰	۱	۵۴	۶۷	۵۵	۳	۳۲
۲۰۱۴	۲۶	۲۱	۱۸۷	۳۶	۳۲۵	۳۱	۶۳	۲	۸۰	۴۵	۹۴	۶	۳۰
۲۰۱۵	۳۷	۲۳	۲۴۰	۲۹	۳۳۹	۳۱	۵۷	۲	۴۵	۶۱	۷۶	۳	۳۴
۲۰۱۶	۲۵	۲۰	۱۶۹	۴۹	۳۲۵	۲۸	۶۶	۱۱	۸۳	۴۵	۱۰۵	۳	۲۷

بامطالعه روند تغییر کلیدواژه‌ها با توجه به **Error! Reference source not found.** می‌توان دریافت که برخی از حوزه‌ها در سال‌های مختلف روند حرکتی تغییرات متفاوتی داشته‌اند. برخی از آن‌ها با افزایش و برخی با کاهش اندازه تکرار کلیدواژه‌های خود روبرو بوده‌اند. برای درک بهتر هر زیرحوزه، نرخ تغییر در اندازه خوشه در سال‌های متوالی نیز محاسبه‌شده که نتایج آن در جدول ۵ آمده است. بر اساس اطلاعات این جدول، برخی خوشه‌ها مانند خوشه چهار و خوشه یازده همیشه نسبت به سال‌های مختلف افزایش داشته‌اند درحالی‌که برخی دیگر مانند خوشه اول و دوم در سال ۲۰۱۶ نسبت به سال‌های قبل از آن نرخ کاهشی را در میزان اندازه خود داشته‌اند.

جدول ۵. نرخ تغییرات خوشه‌ها نسبت به سال‌های متفاوت

شماره خوشه / نرخ تغییرات	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳
نسبت به سال ۲۰۱۲	-۳۲	-۵	-۲۹	۳۸	-۷	۶	۱۴	-۷	۲۱	-۳۳	۳۱	-۳۷	-۱۶
نسبت به سال ۲۰۱۳	-۶	-۱۶	-۳۰	۴۵	-۳	۳	۹	۲۰	۳۵	-۳۳	۴۷	-۸	-۱۳
نسبت به سال ۲۰۱۴	-۲	-۲	-۷	۲۵	۰	-۱۰	۴	-۱۷	۳	۰	۱۰	-۵۷	-۷
نسبت به سال ۲۰۱۵	-۳۲	-۱۲	-۲۹	۴۰	-۴	-۱۰	۱۳	-۱۹	۴۶	-۲۵	۲۸	-۶	-۲۱
میانگین نرخ تغییرات ۴ سال	**-۱۸**	**-۹**	**-۲۴**	**۳۷**	**-۴**	**-۳**	**۱۰**	**-۶**	**۲۶**	**-۲۳**	**۲۹**	**-۲۷**	**-۱۴**

از جدول ۵شش می‌توان نتیجه گرفت که خوشه‌هایی که در سال‌های متوالی روند رو به افزایشی داشته‌اند زمینه‌های تحقیقاتی هستند که توجه محققین به آن‌ها در حال افزایش است. از طرف دیگر در زیرحوزه‌هایی که در نمودار استراتژیک در سمت راست محور افقی قرار دارند در مرکز توجه تحقیقات در میان زیرحوزه‌های دیگر هستند. شکل پنج نشان می‌دهد که حوزه‌های دو، سه و پنج جزء زیرحوزه‌هایی هستند که در مرکز ارتباط با زیرحوزه‌های دیگر قرار دارند. علاوه بر این، ازآنجاکه این زیرحوزه‌ها مقدار چگالی بیشتری را از آن خودکرده‌اند می‌توان این برداشت را انجام داد که به بلوغ خود نیز رسیده‌اند و فعالیت تحقیقاتی بیشتری بر روی آن‌ها انجام نخواهد شد. این زیرحوزه‌ها را می‌توان از جدول ۵شش نیز شناسایی کرد که نرخ رشد آن‌ها نسبت به سال‌های گذشته منفی است.

شکل پنج نشان می‌دهد که زیرحوزه‌هایی چهار، هفت، نه و یازده نسبت به زیرحوزه‌های موجود دارای میزان مرکزیت و بلوغ متوسطی هستند و از طرفی جدول ۵شش نشان می‌دهد که این زیرحوزه‌ها نرخ رشد مثبتی نیز دارند، بنابراین انتظار می‌رود که این زیرحوزه‌ها در سال‌های آینده با توجه به نرخ رشد مثبت خود به سمت بلوغ بیشتر و یا به مرکزیت و توجه بالاتری حرکت کنند. ازآنجاکه خوشه ۱۳ مرکزیت بالاتر از متوسط دارد و در حال حاضر دارای نرخ رشد منفی در چند سال گذشته دارد می‌توان نتیجه گرفت که پیش‌ازاین، زیرحوزه شماره ۱۳ موردتوجه پژوهشگران بوده و پیش از اینکه به بلوغ برسد جذابیت خود را ازدست‌داده و از توجه پژوهشگران دور شده است. از طرفی دیگر زیرحوزه‌های باقی‌مانده (حوزه‌های یک، شش، هشت، ده و دوازده) نسبت به بقیه حوزه‌ها مرکزیت و بلوغ کمتری دارند و با نرخ رشد منفی که در پیش‌گرفته‌اند، این انتظار می‌رود که در سال‌های آینده دیگر فعالیت پژوهشی چشم‌گیری بر روی آن‌ها انجام نشود و در حاشیه قرارگرفته و یا از بین بروند.

نتیجه‌گیری و جمع‌بندی

در این پژوهش با در نظر گرفتن نقاط قوت و ضعف روش‌های تحلیل روند پیشین، روشی برای تحلیل روند ارائه گردیده است. روش ارائه‌شده با استفاده از ارزیابی تطبیقی موردبررسی قرارگرفته شده است. از طرفی دیگر برای بررسی بیشتر روش ارائه‌شده از مجموعه مقالات حوزه مکانیک در سال‌های ۲۰۱۲ تا ۲۰۱۶ مورداستفاده قرارگرفته شده است. مطالعه انجام‌شده نشان داد یکی از روش‌های مهم و اصلی برای تحلیل روند استفاده از هم‌رخدادی کلمات است (Guo et al. 2017; Callon et al. 1983). در تحلیل هم‌رخدادی با استفاده از کلیدواژه‌هایی که از اسناد علمی استخراج‌شده است که تکرار بیشتری نسبت به بقیه دارند، ماتریس هم‌رخدادی تشکیل می‌شود. با استفاده از این ماتریس خوشه‌بندی انجام‌شده و تحلیل‌ها بر روی خوشه‌ها انجام خواهند شد.

مطالعه نشان داد در روش‌های پیشین تحلیل روند کلیدواژه‌ها به‌صورت ثابت توسط سیستم و بدون تحلیل مفهومی واژگان انتخاب می‌شوند (برای مثال (Wu et al. 2011)) و یا در انتخاب آن‌ها از متخصصین استفاده می‌شود (برای مثال (An and Wu 2011)). درصورتی‌که در روش ارائه‌شده این انتخاب به‌صورت پویا و خودکار انجام‌شده است. از طرفی دیگر مطالعه نشان داد که در پژوهش‌های پیشین برای تحلیل روند با استفاده از ماتریس هم‌رخدادی، تأثیر میزان هم‌رخدادی کلیدواژه‌ها در هر

مقاله در نظر گرفته نشده است و تنها به هم‌رخدادی دو کلیدواژه در یک مقاله بسنده کرده‌اند (برای مثال (No and Park 2010)). ازآنجاکه تعداد دفعات تکرار یک کلمه در سند علمی نشان‌دهنده میزان اهمیت آن کلمه است (Rose et al. 2010)، در این پژوهش با در نظر گرفتن تأثیر میزان هم‌رخدادی کلیدواژه‌ها در یک مقاله، میزان اهمیت آن‌ها در ماتریس هم‌رخدادی اعمال‌شده و به‌تبع آن این تغییر در خوشه‌بندی و تحلیل روند نیز دیده‌شده است.

فاز نهایی در تحلیل روند یک حوزه استفاده از ابزارها و شاخص‌هایی برای تحلیل است. مطالعه نشان می‌دهد که به‌غیراز استفاده از اطلاعات آماری ساده (برای نمونه) برای تحلیل‌های پیشرفته‌تر از نمودار استراتژیک استفاده‌شده است (برای نمونه (Hu and Zhang 2015)). نمودارهای استراتژیک از میزان ارتباط بین خوشه‌ها و ارتباط بین کلیدواژه‌های درون خوشه‌ای برای محاسبه میزان بلوغ و مرکزیت استفاده‌شده است و میزان تکرار کلیدواژه‌ها در این نمودارها در نظر گرفته نشده است (برای نمونه (An and Wu 2011; Hu and Zhang 2015)). برای محاسبه میزان بلوغ و مرکزیت یک حوزه علاوه بر در نظر گرفتن میزان ارتباطات کلیدواژه‌های یک خوشه از میزان هم‌رخدادی کلیدواژه‌های آن خوشه نیز بهره گرفته‌شده است. این مقادیر درنهایت برحسب تعداد اتصالات درون خوشه‌ای و برون خوشه‌ای نرمال شده‌اند. از دیگر ویژگی‌های روش ارائه‌شده ارائه شاخص برای تحلیل رفتاری زیرحوزه‌ها و تحلیل روند درون زیرحوزه‌ها اشاره کرد. این تحلیل روند با استفاده از روش‌های متن‌کاوی و استفاده از اطلاعات کتابشناختی حاصل‌شده است.

برای بهبود طرح پیشنهادی تحلیل روند، در کارهای آتی برای بهبود تحلیل روند می‌توان از لیست-توقف تخصصی یک حوزه برای پالایش کلمات استفاده کرد. در روش پیشنهادی فعلی طرح لیست توقف عمومی استفاده‌شده است و از اسناد مورد تحلیل لیست توقف تخصصی حوزه تشکیل‌شده است بنابراین می‌توان از لیست‌های توقف استانداردی که با اسناد با حجم بیشتر تشکیل‌شده‌اند نیز بهره گرفت. از سویی دیگر برای تشکیل خوشه نیز می‌توان از خوشه‌بندی‌های متفاوت استفاده و آن‌ها را با یکدیگر مقایسه نمود. روش‌های رده‌بندی وجود دارند که با استفاده از ارتباط معنایی یا ارتباط موضوعی بین کلمات آن‌ها را در رده‌های مختلفی قرار دهند (برای مثال استفاده از اصطلاح‌نامه‌های تخصصی در کنار خوشه‌بندی کلمات این رده‌بندی انجام شود). افزایش دامنه‌های تحقیق و افزایش بازه زمانی می‌تواند باعث افزایش دقت در تحلیل روند و به وجود آمدن موضوع‌های میان‌رشته‌ای و نشان دادن بلوغ حوزه‌های دیگر شود.

فهرست منابع

An, Xin Ying, and Qing Qiang Wu. 2011. "Co-word analysis of the trends in stem cells field based on subject heading weighting." Scientometrics 88 (1):133-144.

Callon, Michel, Jean-Pierre Courtial, William A Turner, and Serge Bauin. 1983. "From translations to problematic networks: An introduction to co-word analysis." Information (International Social Science Council) 22 (2):191-235.

Chang, Xing, Xin Zhou, Linzhi Luo, Chengjia Yang, Hui Pan, and Shuyang Zhang. 2017. "Hotspots in research on the measurement of medical students' clinical competence from 2012-2016 based on co-word analysis." BMC medical education 17 (1):162.

Chao, Chia-Chen, Jiann-Min Yang, and Wen-Yuan Jen. 2007. "Determining technology trends and forecasts of RFID by a historical review and bibliometric analysis from 1991 to 2005." Technovation 27 (5):268-279.

Chen, Xiuwen, Jianming Chen, Dengsheng Wu, Yongjia Xie, and Jing Li. 2016. "Mapping the research trends by co-word analysis based on keywords from funded project." Procedia Computer Science 91:547-555.

Choi, Changwoo, and Yongtae Park. 2009. "Monitoring the organic structure of technology based on the patent development paths." Technological Forecasting and Social Change 76 (6):754-768.

Davies, David L, and Donald W Bouldin. 19" .۷۹A cluster separation measure." IEEE transactions on pattern analysis and machine intelligence (2):224-227.

Delecroix, Bertrand, and R Epstein. 2004. "Co-word analysis for the non-scientific information example of Reuters Business Briefings." Data Science Journal 3:80-87.

Dimitriadou, Evgenia, Sara Dolničar, and Andreas Weingessel. 2002. "An examination of indexes for determining the number of clusters in binary data sets." Psychometrika 67 (1):137-159.

Dunn, Joseph C. 1973. "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters".

Guo, Daoyan, Hong Chen, Ruyin Long, Hui Lu, and Qianyi Long. 2017. "A co-word analysis of organizational constraints for maintaining sustainability." Sustainability 9 (10):1928.

He, Qin. 1999. "Knowledge discovery through co-word analysis." Library trends 48 (1):133-133.

Hu, Chang-Ping, Ji-Ming Hu, Sheng-Li Deng, and Yong Liu. 2013. "A co-word analysis of library and information science in China." Scientometrics 97 (2):369-382.

Hu, Jiming, and Yin Zhang. 2015. "Research patterns and trends of Recommendation System in China using co-word analysis." Information processing & management 51 (4):329-339.

Jain, Anil K. 2010. "Data clustering: 50 years beyond K-means." Pattern recognition letters 31 (8):651-666.

Joung, Junegak, and Kwangsoo Kim. 2017. "Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data." Technological Forecasting and Social Change 114:281-292.

Khatir, Ashkan, and Soheil Ganjefar. 2016. "The Analysis of the Distribution and Focus of Keywords in Theses and Dissertations: The Compliance with Descriptors, Title, and Abstract." Journal of Information Processing and Management.

Kung, Yen-Ying, Shinn-Jang Hwang, Tsai-Feng Li, Seong-Gyu Ko, Ching-Wen Huang, and Fang-Pey Chen. 2017. "Trends in global acupuncture publications: An analysis of the Web of Science database from 1988 to 2015." Journal of the Chinese Medical Association 80 (8):521-525.

Larsen, SE, B Kronvang, J Windolf, and LM Svendsen. 1999. "Trends in diffuse nutrient concentrations and loading in Denmark: statistical trend analysis of stream monitoring data." Water science and technology 39 (12):197-205.

Law, John, Serge Bauin, J Courtial, and John Whittaker. 1988. "Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification." Scientometrics 14 (3-4):251-264.

Leydesdorff, Loet, and Adina Nerghes. 2017. "Co-word maps and topic modeling: A comparison using small and medium-sized corpora (N< 1,000)." Journal of the Association for Information Science and Technology 68 (4):1024-1035.

Lv, Peng Hui, Gui-Fang Wang, Yong Wan, Jia Liu, Qing Liu, and Fei-cheng Ma. 2011. "Bibliometric trend analysis on global graphene research." Scientometrics 88 (2):399-419.

Müller, Andre Matthias, Carol A Maher, Corneel Vandelanotte, Melanie Hingle, Anouk Middelweerd, Michael L Lopez, Ann DeSmet, Camille E Short, Nicole Nathan, and Melinda J Hutchesson. 2018. "Physical Activity, Sedentary Behavior, and Diet-Related eHealth and mHealth Research: Bibliometric Analysis." Journal of medical Internet research 20 (4):e122.

No, Hyun Joung, and Yongtae Park. 2010. "Trajectory patterns of technology fusion: Trend analysis and taxonomical grouping in nanobiotechnology." Technological Forecasting and Social Change 77 (1):63-75.

Ozaydin, Bunyamin, Ferhat Zengul, Nurettin Oner, and Dursun Delen. 2017. "Text-mining analysis of mHealth research." mHealth 3 (12.(

Porter, Alan L. 1991 .Forecasting and management of technology. Vol. 18: John Wiley & Sons.

Rokaya, Mahmoud, Elsayed Atlam, Masao Fuketa, Tshering C Dorji, and Jun-ichi Aoe. 2008. "Ranking of field association terms using co-word analysis." Information processing & management 44 (2):738-755.

Rose, Stuart, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. "Automatic keyword extraction from individual documents." Text Mining: Applications and Theory:1-20.

Rousseeuw, Peter J. 1987. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." Journal of computational and applied mathematics 20:53-65.

Samadi kuchaksaraei, Ali, Hafez Mohammad hassanzadeh, and Farhad Shokraneh. 2013. "A bibliometric trend analysis of stem cells and regenerative medicine research output in Iran: comparison with the global research output".

Shen, Jun, Xiaoxia Li, and Zusha Gu. 2013. "Strategic Diagram Analysis Based on Knowledge Network." 2012 First National Conference for Engineering Sciences (FNCES 2012.(

Suh, Yongyoon, and Jeonghwan Jeon. 2018. "Monitoring patterns of open innovation using the patent-based brokerage analysis." Technological Forecasting and Social Change.

Tao, Hongzhi, Jianfeng Li, Tao Luo, and Cong Wang. 2017. "Research on topics trends based on weighted K-means." Electronics Information and Emergency Communication (ICEIEC), 2017 7th IEEE International Conference on.

Vartiainen, Pirkko. 2002. "On the principles of comparative evaluation." Evaluation 8 (3):359-371.

White, George O, Orhun Guldiken, Thomas A Hemphill, Wu He, and Mehdi Sharifi Khoobdeh. 2016. "Trends in International Strategic Management Research From 2000 to 2013: text mining and bibliometric analyses." Management International Review 56 (1):35-65.

Wu, Feng-Shang, Chun-Chi Hsu, Pei-Chun Lee, and Hsin-Ning Su. 2011. "A systematic approach for integrated trend analysis—The case of etching." Technological Forecasting and Social Change 78 (3):386-407.

Zhang, Wei, Qingpu Zhang, Bo Yu, and Limei Zhao. 2015. "Knowledge map of creativity research based on keywords network and co-word analysis, 1992–2011." Quality & Quantity 49 (3):1023-1038.

Zhao, Fangkun, Bei Shi, Ruixin Liu, Wenkai Zhou, Dong Shi, and Jinsong Zhang. 2018. "Theme trends and knowledge structure on choroidal neovascularization :a quantitative and co-word analysis." BMC ophthalmology 18 (1):86.

A new automatic approach for research trend analysis based on scientific text mining

Ashkan Khatir
PhD Candidate in Information Technology Engineering, Iranian Research Institute for Information Scince and Technology (IranDoc), Tehran, Iran
Azadeh Mohebi
Assistant Professor, Iranian Research Institute for Information Science and Technology (IranDoc), Tehran, Iran[1]
Soheil Ganjefar
Visiting Professor in Iranian Research Institute for Information Science and Technology (IranDoc) and Professor, Buali Sina University, Hamedan, Iran

Abstract: Research trend analysis for a specific research area (through different time frames) can lead to a better understanding for researchers in that area, and for policy makers in research for contributing in assigning research funds and policies. An important and practical approach to analyzing research trends is to study and review research data and publications using scientometrics and research document processing. In this research, we have proposed a text mining approach for analysing research publications in a specific area, in order to analyze and identify important research topics. In this paper, we propose a method to analysis a scientific trend. The proposed approach is based on clustering keywords through a new co-occurrence matrix. A new metric is also adopted to identify centrality and density (maturity) of a specific research area and identify keywords and contributory topics. For achieving this, we proposed a trend analysis method using co-word matrix with and for deeper analysis we use clustering and strategic diagram with propose indices. In order to test and evaluate the proposed method, we use comparative evaluation method. In addition, for more analysis, we have selected research publications in a specific period of time (2012-2016) in Mechanical Engineering, which are extracted from WoS (Web of Sciences) database. We have applied the proposed metrics to evaluate research trends and identify contributory areas and topics in the selected documents. The comparative evaluation shows an improvement in proposed trend analysis method.

Keywords: Centrality of a research area, Co-occurrence matrix, Maturity of a research area, Text mining, Trend analysis.

1.Corresponding Author: mohebi@irandoc.ac.ir