

Designing a Model Based on Cumulative Citation to Identify and Analyze Scientific Changes in the Field of Data Quality

**Ahmad Khalili Jafarabad¹, Amir Manian², Mohammad Fathian³,
Nader Naghshineh⁴**

Abstract: Identification and tracking scientific changes is critical for scientific policy makers. This research proposed a model for identification and tracking changes in Data Quality research area. We used cumulative citation network in order to find research communities and their changes during the time. The proposed model can be applied in other scientific disciplines. It can also shows all types of scientific changes including birth, growth, merging and death. In order to verify the model in Data Quality area, we selected all papers that is published between 1970 and 2009 that covers more than 7000 papers. It is shown that Data Quality research area is studied in different disciplines. According to the results, there is 82 percent correlation between number of citations and the growth of Data Quality communities that shows the importance of citation for community survival and growth.

Key words: *Citation network, Data quality, Science tracking, Scientometrics.*

-
1. PhD Candidate in IT, Faculty of Management, University of Tehran, Tehran, Iran
 2. Associate Prof. in IT, Faculty of Management, University of Tehran, Tehran, Iran
 3. Prof. of Industrial Engineering, University of Science and Industry, Tehran, Iran
 4. Assistant Prof., Dep. of Information and Knowledge, University of Tehran, Tehran, Iran
-

Submitted: 12 / February / 2017

Accepted: 14 / June / 2017

Corresponding Author: Ahmad Khalili

Email: ahmad.khalili@ut.ac.ir

طراحی مدل مبتنی بر ارجاعات تجمعی به منظور بررسی کشف و تحلیل تغییرات علمی در حوزه کیفیت داده

احمد خلیلی جعفرآباد^۱، امیر مانیان^۲، محمد فتحیان^۳، نادر نقشینه^۴

چکیده: شناخت حوزه‌های علمی و بررسی تغییرات آن از اهمیت بسزایی برخوردار است. در این مقاله با ارائه مدلی مبتنی بر ارجاعات تجمعی زمان‌مبنا، تغییرات علمی حوزه کیفیت داده بررسی و تحلیل شده است. از ویژگی‌های این مدل، قابلیت انطباق بر سایر حوزه‌ها و همچنین قابلیت نمایش کلیه تغییرات علمی از جمله تولد، ادغام، شکست یا مرگ یک حوزه علمی است. گام‌های اجرایی این مدل به ترتیب عبارت‌اند از: آماده‌سازی داده، استخراج مدل ارجاعات زمان‌مبنا، کشف مجامع علمی، اتصال مجامع علمی، استخراج مدل تغییرات، تحلیل تغییرات علمی. به منظور بررسی کیفیت مدل ارائه شده، کلیه مقالات به چاپ رسیده بین سال‌های ۱۹۷۰ تا ۲۰۰۹ که بیش از ۷۰۰۰ مقاله را شامل می‌شود، بررسی شدند. بر اساس مدل مبتنی بر ارجاعات تجمعی زمان‌مبنا، مشخص شد حوزه کیفیت داده در حوزه‌های مختلف در حال مطالعه است. همچنین بر اساس بررسی انجام شده، بین تعداد ارجاعات و رشد حوزه کیفیت داده، همبستگی ۸۲ درصدی وجود دارد که نشان‌دهنده اهمیت ارجاعات برای زنده ماندن و رشد حوزه‌های علمی است.

واژه‌های کلیدی: تحلیل ارجاعات، تحلیل تغییرات علمی، علم‌سنجی، کیفیت داده.

۱. دانشجوی دکتری مدیریت فناوری اطلاعات، دانشکده مدیریت دانشگاه تهران، تهران، ایران

۲. دانشیار مدیریت فناوری اطلاعات، دانشکده مدیریت دانشگاه تهران، تهران، ایران

۳. استاد مهندسی صنایع، دانشگاه علم و صنعت، تهران، ایران

۴. استادیار گروه آموزشی علم اطلاعات و دانش‌شناسی، دانشگاه تهران، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹۵/۱۱/۲۴

تاریخ پذیرش نهایی مقاله: ۱۳۹۶/۰۳/۲۴

نویسنده مسئول مقاله: احمد خلیلی جعفرآباد

E-mail: ahmad.khalili@ut.ac.ir

مقدمه

امروزه جهان به دلیل محدودیت رشد ناشی از تمرکز بر استراتژی‌های مبتنی بر نیروی کار و سرمایه، به دنبال حرکت به سمت اقتصاد دانش‌بنیان و ایجاد مزیت رقابتی با اتکا به دانش و اطلاعات است. در این اقتصاد، دانش و تکنولوژی مؤلفه‌های اصلی پیشرفت در نظر گرفته می‌شوند (لی، ۲۰۰۸). به همین منظور، شرکت‌ها و مؤسسه‌ها به دنبال کشف حوزه‌های علمی نوین و شناسایی حوزه‌های تحقیقاتی نوظهور هستند. به دلیل رقابت کمتر و امکان استفاده از موقعیت که در نتیجه کمبود تعداد رقبا و به بلوغ نرسیدن صنعت است، شناخت این حوزه‌ها یک فرصت محسوب می‌شود (کازنس و دیگران، ۲۰۱۰). با توجه به اینکه از یک سو تحلیل داده و علوم مرتبط با آن در حال تبدیل شدن به یکی از استراتژی‌های اصلی تمام صنایع است و از سوی دیگر، ایجاد مزیت رقابتی به واسطه فناوری اطلاعات هزینه بسیار کمتری دارد، نیاز است روی این بخش سرمایه‌گذاری جدی انجام شود. یکی از موضوعات مهم در این حوزه، بحث کیفیت داده‌هاست؛ چراکه داده خوب و با کیفیت، مبنای رسیدن به نتایج مطلوب از سیستم‌های اطلاعاتی است. به بیانی می‌توان گفت سیستم اطلاعاتی همراه با داده بی کیفیت، هر چند طراحی و پیاده‌سازی مناسبی داشته باشد، نمی‌تواند تصمیم‌گیران و کاربران سیستم‌های اطلاعاتی را به نتایج مطلوب مد نظر برساند. بنابراین، تمرکز بر موضوع کیفیت داده در سیستم‌های اطلاعاتی امری اجتناب‌ناپذیر است. از دید حوزه‌های تحقیقاتی، کیفیت داده به عنوان موضوع میان‌رشته‌ای بین مدیریت، آمار و علوم کامپیوتر مطرح است (مسلا و دیگران، ۲۰۰۲) و به همین دلیل شناخت روندهای اصلی و تغییرات این حوزه اهمیت زیادی دارد.

پیش از ورود به مسائل مرتبط با شناخت حوزه‌های تحقیقاتی نوین، نیاز است سازوکارهای تغییرات علمی و شیوه‌های تحول در علم کاوش شود. اندیشمندان بسیاری نظریه‌هایی ارائه کرده‌اند که یکی از بهترین نظریه‌ها را می‌توان نظریه کوهن دانست (کوهن و هاوکینز، ۱۹۶۳). نظریه کوهن یکی از نظریه‌های موجود در زمینه شیوه توسعه علم است. بر اساس نظریه کوهن، پیشرفت علم به وسیله ایده‌های انقلابی اتفاق می‌افتد که به صورت ناگهانی سیمای یک حوزه علمی را تغییر می‌دهد. کوهن معتقد است که بسیاری از تغییرات از طریق توسعه‌های جزئی اتفاق می‌افتد، اما ایده‌هایی هستند که موجب چرخش ناگهانی و پویا در تمرکز علم می‌شوند. کشف DNA یا نظریه نسبیت انیشتین را می‌توان از بزرگ‌ترین مثال‌های این اتفاق دانست. البته باید توجه داشت که این انقلاب‌ها منحصر به این چند نظریه نیست و در سطوح مختلف و در علوم گوناگون می‌توان نمونه‌های مشابهی را جست‌وجو کرد (مدلاک براون، ۲۰۱۴). نکته دیگری که باید به آن توجه شود، تفاوت مفهوم پیش‌بینی و اکتشاف است. در این مقاله پیش از ورود به

بحث پیش‌بینی، ابتدا باید مختصری درباره حوزه‌ای که در این پژوهش مطالعه می‌شود، توضیح داده شود. مسائل اکتشافی بیشتر رویکرد گذشته‌نگر دارند و به دنبال کشف یک پدیده یا یک مدل از میان مشاهدات یا داده‌های موجود هستند. برای نمونه، کشف یک الگو در میان انبوه داده‌ها به منظور ارائه توضیحی درباره عملکرد یک پدیده، از جمله این روش‌هاست. اما مدل‌های پیش‌بینی به دنبال پیش‌بینی آینده رفتاری پدیده با توجه به وضعیت فعلی آن هستند که برای آنها نیز دو رویکرد کمی و کیفی وجود دارد (کاجیکاوا، یوشیکاوا، تاکدا و ماتسوشیما، ۲۰۰۸).

این پژوهش به دنبال ارائه مدلی به منظور کشف سازوکارهای تغییرات علمی در طول زمان و مدل‌سازی با استفاده از داده‌های موجود در مقاله‌های چاپ‌شده است. به همین منظور با استفاده از اطلاعات کتاب‌شناختی مرتبط با مبحث کیفیت داده که از WOS استخراج شده است، به مدل‌سازی تغییرات علمی اقدام شده است.

پیشینه پژوهش

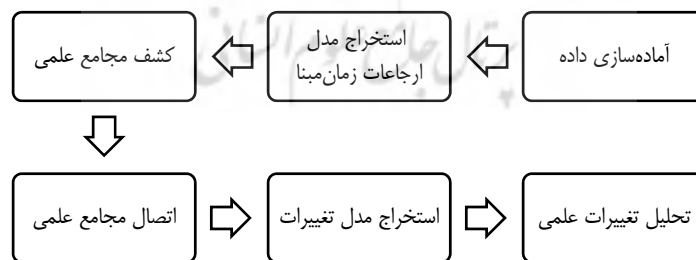
برای تبیین تغییرات در حوزه‌های علمی تا کنون مطالعات زیادی انجام شده که بیشتر در خصوص ارائه مدل‌هایی برای اکتشاف حوزه‌های تحقیقاتی و تکنولوژی‌های نوظهور است. همان‌طور که در جدول ۱ نیز مشخص شده است، بیشتر این دست از مقالات بر کشف حوزه‌های جدید و پیش‌بینی آنها تمرکز کرده‌اند. این نوع تحلیل با ارائه مدل‌های مبتنی بر بیلیومتری در حوزه علم‌سنجی ظهور کرد. برای مثال می‌توان به شاخص‌های پرایس که در سال ۱۹۷۰ ارائه شد (پرایس، ۱۹۷۰) اشاره کرد. نمونه دیگر، شاخص‌های عمومی مربوط به درصد مقالات یا منابع در یک بازه زمانی مشخص، میانگین یا میانه عمر است که به شاخص‌های نیمه عمر شهرت دارند، یا شاخص فوریت که در سال ۱۹۷۲ ارائه شده است (گارفیلد، ۱۹۷۲). در ادامه این روند، محققان به بررسی شبکه‌های علمی و ارجاعات پرداختند که آغاز این روند نیز مطالعات انجام شده در شبکه ارجاعات است (اسمال، ۱۹۷۷). پس از این مرحله، مطالعات بسیاری با استفاده از رویکردهای مختلف در این حوزه انجام گرفت که در جدول ۱ به طور خلاصه به آنها اشاره شده است. نکته مهم در مطالعات مندرج در جدول ۱، این است که در خصوص تغییرات حوزه‌های علمی مطالعات کمی انجام شده است. شناسایی تولد، رشد، ادغام یا مرگ یک حوزه علمی از این جهت حائز اهمیت است که امکان شناخت چرخه عمر یک حوزه و همچنین پیش‌بینی رفتار یک حوزه علمی را برای تحلیل‌گران فراهم می‌آورد. همچنین این موضوع می‌تواند به شناخت بهتر چرخه‌های علمی منجر شود. به همین منظور این مقاله به دنبال ارائه مدلی به منظور نشان دادن سازوکارهای تغییرات علمی است که از مدل پیشنهادی برای تعیین تغییرات حوزه کیفیت داده استفاده می‌کند.

جدول ۱. مرور ادبیات بررسی تغییرات علمی

منبع	نوع تحلیل	موضوع و نوآوری
اسمال، بویاک و کلاوان (۲۰۱۴)	تحلیل خوشه‌های شبکه ارجاعات	ارائه روشی مبتنی بر مفهوم «نخ علم» برای تعیین حوزه‌های نوظهور در اطلاعات با حجم بالا
بویاک، کلاوان، اسمال و اونگار (۲۰۱۴)	تحلیل خوشه‌های شبکه ارجاعات	ارائه مدلی برای بررسی ظهور حوزه‌های تحقیقاتی و مدل دیگر برای تعیین سازوکار تغییر خوشه‌ها از طریق ادغام و شکست
الکساندر، جیس، نیومن، پورتر و روزنر (۲۰۱۲)	-	تبیین مفهوم نوظهوری از دیدگاه‌های مختلف و سطح‌بندی مفهوم نوظهوری در سیستم و زیرسیستم‌ها
گو، وینگارت و بورنر (۲۰۱۱)	روند تعداد مقالات	تبیین سازوکار ظهور حوزه تحقیقاتی نوظهور و ارائه مدلی برای نوظهوری
گازنس و دیگران (۲۰۱۰)	-	ارائه شاخص‌های رشد تکنولوژی بر اساس دیدگاه‌های مختلف
لی (۲۰۰۸)	تحلیل شاخص‌های شبکه	ارائه مدل تشخیص حوزه‌های تحقیقاتی نوظهور در بحث امنیت اطلاعات بر اساس ساختار شبکه
چن (۲۰۰۴)	تحلیل روند	دسته‌بندی انواع نکات کلیدی در شبکه و بررسی روند تغییرات زمانی
تو و سنگ (۲۰۱۲)	-	ارائه مدلی بر اساس مفهوم بدیع بودن تحقیقات و ارائه شاخص‌های بدیع بودن تحقیقات
وانگ، چنگ و لو (۲۰۱۴)	تحلیل شبکه متون	ارائه مدلی برای تحلیل حوزه‌های تحقیقاتی بر اساس شبکه پویای لغات

فرایند اجرایی پژوهش

برای دستیابی به هدف اصلی پژوهش مبنی بر ارائه مدلی به‌منظور کشف تغییرات علمی، رویه نشان‌داده شده در شکل ۱ طی شده است. همان‌طور که در شکل نیز مشخص است، این پژوهش دارای شش گام اساسی است که در ادامه به‌ترتیب هریک تشریح می‌شود.



شکل ۱. روند گام به گام پیاده‌سازی مدل پیشنهادی

آماده‌سازی اطلاعات

برای آماده‌سازی اطلاعات، تمام مقاله‌های نمایه شده در WOS که با حوزه کیفیت داده و نیز، کامپیوتر و سیستم‌های اطلاعاتی ارتباط داشتند، استخراج شدند. به منظور استخراج کلیدواژه‌های مناسب، از طبقه‌بندی استخراج شده برای این حوزه استفاده شده است (صدیق، یگانه و ایندولسکا، ۲۰۱۱). بدین ترتیب، مشخص شد افزون بر ۷۰۰۰ طی سال‌های ۱۹۷۰ تا ۲۰۰۹ به چاپ رسیده است.

استخراج مدل ارجاعات زمان‌مبنا

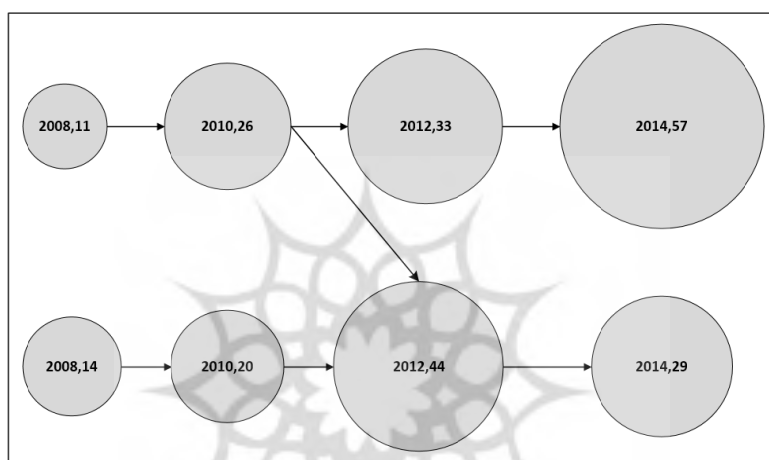
یکی از پیچیدگی‌های مدل‌سازی تغییرات علمی، ارائه مدلی به منظور بررسی تغییرات حوزه‌های علمی است. در این مقاله از مدل ارجاعات تجمعی برای مدل‌سازی استفاده شده است. برای ساخت این مدل، تمام مقاله‌های چاپ شده از سال ۱۹۷۰ تا ۲۰۰۹ جمع‌آوری شدند تا مدل ارجاعات مستقیم ارائه شود. برای مثال، شبکه‌های مجزا در سال‌های ۱۹۹۰-۱۹۷۰ و ۱۹۷۲-۱۹۷۰ ساخته شده است. برای ساخت مدل ارجاعات، هر یک از مقالات را به صورت یک گره در نظر گرفتیم و اگر مقاله‌ای به مقاله دیگری ارجاع داده بود، بین دو گره یک یال ایجاد کردیم.

کشف مجامع علمی

برای کشف مجامع علمی روش‌های گوناگونی وجود دارد که در این مقاله از مدل لواین استفاده شده (بلوندل، گویلام، لامبویت و لفر، ۲۰۰۸) که از جمله مدل‌های دقیق و سریع برای کشف اجتماع در شبکه‌های پیچیده است. این روش که بر مبنای ماژولاریتی در شبکه است، از جمله روش‌های مناسب و سریع برای کشف مجامع به شمار می‌رود. در این بخش برای هر دوره زمانی یک بار مدل لواین به اجرا درآمد. الگوریتم لواین با رویکرد دستیابی به سرعت زیاد برای تشخیص اجتماعات، می‌کوشد تا با استفاده از سنجه ماژولاریتی، گره‌ها را در پایین‌ترین سطح - که هر یک از آنها در یک اجتماع قرار دارند - در اجتماعات کوچک اولیه با هم ادغام کند. آنگاه در لایه بالاتر، اجتماعات کوچک را به عنوان گره در نظر گرفته و فرایند بالا را روی آنها انجام می‌دهد. این الگوریتم در مرحله بعد، اجتماعات ممکن میان اجتماعات جدید را می‌یابد و این کار تا زمانی که تشکیل اجتماع دیگری امکان‌پذیر نباشد، ادامه می‌یابد. بدین ترتیب یک ساختار سلسله‌مراتبی ایجاد می‌شود و به دلیل کاهش اندازه مسئله در هر لایه، سرعت الگوریتم برای تشکیل اجتماعات نهایی به شدت افزایش می‌یابد.

اتصال مجامع علمی

یکی از پیچیدگی‌های این مدل ارائه راهکاری به منظور کشف تغییرات حوزه‌های علمی است. راهکار پیشنهادی برای این بخش به این صورت است که در هر دوره زمانی t ارتباط بین تمام مجامع به مجامع دوره $t-1$ ارزیابی می‌شود و بین مجامع با شباهت بیش از مقدار α ارتباطی به وجود می‌آید که خروجی این بخش در قالب شکل ۲ به نمایش گذاشته شده است.



شکل ۲. مدل مفهومی نمایش روندهای استخراج شده

برای محاسبه شباهت بین مجامع نیز از معیار جاکارد استفاده شده است که به شکل اشتراک بین دو مجموعه نسبت به اجتماع دو مجموعه محاسبه شده است. مثلاً برای محاسبه شباهت بین دو مجموعه A و B باید از رابطه ۱ استفاده کرد.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad \text{رابطه ۱}$$

تحلیل تغییرات علمی

در این گام با پیاده‌سازی مدل ارائه شده روی داده‌های مربوط به مقالات کیفیت داده، به بررسی سازوکارهای تغییرات علمی در این حوزه اقدام شده است. خروجی این مرحله شامل تحلیل شیوه تغییرات علمی و تحلیل آماری این تغییرات در حوزه کیفیت داده خواهد بود.

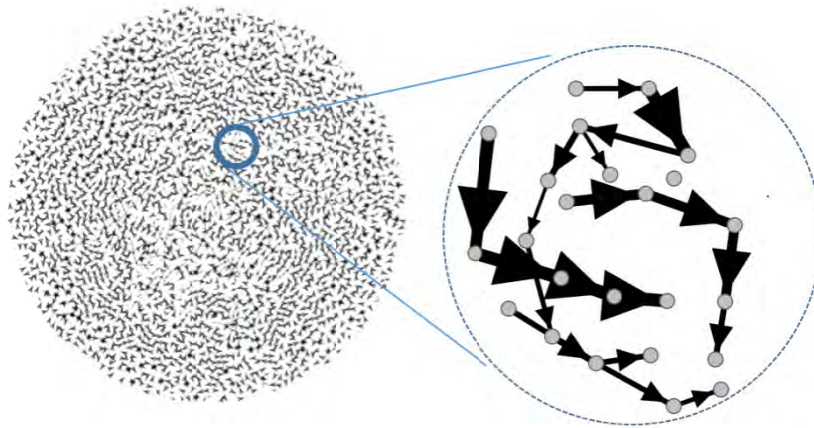
یافته‌های پژوهش

برای پیاده‌سازی مدل، ابتدا تمام مقالات مرتبط با حوزه کیفیت داده که در مجموع افزون بر ۷۰۰۰ مقاله بود، از سایت WOS جمع‌آوری شد. پس از این مرحله، شبکه ارجاعات مقالات به شکل تجمعی و با بازه ۲ سال شکل گرفت. علت انتخاب ۲ سال این است که بر اساس مطالعات انجام شده، طول عمر ارجاعات بین ۳ تا ۵ سال برآورد شده است و این به معنای تحول در شبکه ارجاعات در این مدت است. بنابراین با انتخاب بازه دو ساله، قبل از تحول جدی در یک حوزه که امکان پیش‌بینی را سلب خواهد کرد، شبکه بار دیگر ارزیابی خواهد شد. پس از این مرحله برای هر بازه زمانی، یک بار مدل لواین به اجرا درآمد. برای مشخص شدن مدل استفاده شده، در جدول ۲ تعداد مقالات هر بازه، تعداد گره‌های شبکه ارجاعات و تعداد مجامع کشف شده در هر بازه برای سال‌های ۱۹۷۰ تا ۱۹۹۹ درج شده است.

جدول ۲. اجتماعات شناسایی شده در سال‌های مختلف

بازه زمانی	تعداد گره‌های شبکه ارجاعات	تعداد اجتماعات شناسایی شده
۱۹۷۰ تا ۱۹۸۹	۳۸۵	۲۱
۱۹۷۰ تا ۱۹۹۱	۱۳۴۶	۵۶
۱۹۷۰ تا ۱۹۹۳	۴۳۵۲	۱۱۸
۱۹۷۰ تا ۱۹۹۵	۸۷۰۷	۱۴۷
۱۹۷۰ تا ۱۹۹۷	۱۲۰۹۹	۲۰۷
۱۹۷۰ تا ۱۹۹۹	۱۶۵۳۶	۲۸۰

گفتنی است، به دلیل تعداد کم مقالات در سال‌های ۱۹۷۰ تا ۱۹۸۹، این دوره یک بازه واحد در نظر گرفته شده است. پس از تعیین مجامع در هر بازه زمانی، یک زنجیره از اجتماعات بر اساس میزان شباهت هر زوج از مجامع ایجاد می‌شود. به منظور بصری‌سازی این روند از مدل گراف استفاده شده است که شکل ۳ خروجی آن را نمایش می‌دهد. شایان ذکر است که در شکل ۳ هر گره معرف یک حوزه علمی و اندازه آن نشان‌دهنده تعداد اعضای عضو در آن اجتماع است. همان‌طور که در این شکل نیز مشخص است، این نمودار می‌تواند روند تغییرات مجامع مختلف را در طول زمان مشخص کند. باید توجه داشت که بر اساس مدل ارائه شده تمام رفتارها و تغییرات یک حوزه علمی قابل رصد و ارزیابی است. همچنین در جدول ۳ نمونه‌ای از رفتارهای مختلف یک حوزه علمی و اکتشاف آن توسط مدل طراحی شده مشاهده می‌شود.

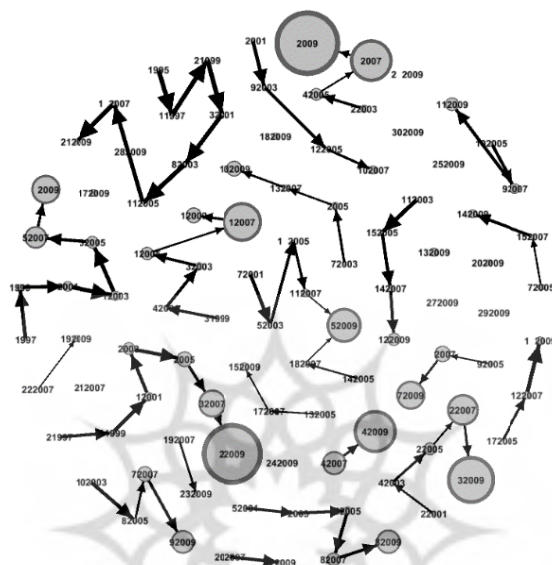


شکل ۳. مدل ارجاعات تجمعی حوزه کیفیت داده

جدول ۳. انواع روندهای قابل شناسایی توسط مدل پیشنهادی

نام	توضیحات	نمونه کشف شده
تولد	یکی از مجامع که در دوره قبلی وجود نداشته و در دوره جدید ایجاد شده است.	652003
رکود	تغییرات محسوسی در یک اجتماع اتفاق نیفتاده و اندازه اجتماع ثابت است.	
رشد	رشد تعداد اعضای یک مجمع که حاصل تولید مقاله در آن حوزه است.	
افول	مهاجرت اعضا از یک اجتماع برای ایجاد اجتماع جدید.	
ادغام	ادغام دو یا چند اجتماع با یکدیگر برای ساخت اجتماع بزرگتر.	
تجزیه	شکستن یک اجتماع بزرگ به دو یا چند اجتماع کوچکتر.	
مرگ	از بین رفتن یک اجتماع در یک بازه زمانی	

به منظور ارزیابی مدل ارائه شده، ابتدا ۱۰۰ اجتماع بزرگ پیش از سال ۲۰۰۹ بر اساس تعداد اعضا انتخاب شد و روابط با شباهت کمتر از ۰/۲ حذف شدند. در نتیجه مدل کلی به صورت زیر ساده سازی شد که نشان دهنده روندهای اصلی این حوزه است.



شکل ۴. روندهای علمی برتر حوزه کیفیت داده

همان طور که در شکل ۴ نیز مشخص است، ۲۰ روند اصلی با عمر بیش از دو سال، وجود دارد. برای بررسی این روندها ویژگی‌های مختلفی نظیر تعداد اعضا، میانگین تعداد ارجاعات و میانگین رشد هر حوزه و نوع روند مشخص شده است. اگر تعداد اعضای یک اجتماع در دوره t به صورت $f_c(t)$ در نظر گرفته شود، برای محاسبه میانگین رشد از رابطه ۲ که میانگین حسابی رشد تعداد اعضا است، استفاده می‌شود.

$$\text{رابطه ۲} \quad \frac{\sum_{t=1}^n (f_c(t) - f_c(t-1))}{n}$$

برای به دست آوردن میانگین تعداد ارجاعات، کلیه ارجاعات مقالات از WOS استخراج شدند و میانگین تعداد ارجاعات برای هر اجتماع محاسبه شد. برای تعیین موضوعات هر حوزه نیز، مقالات برتر هر حوزه بررسی و موضوعات آنها فهرست شدند. مقالات برتر مقالاتی با بیش از ۷۰ ارجاع در نظر گرفته شده است. در جدول ۴ روندها و ویژگی‌های مرتبط با هر یک به تفکیک ارائه شده است.

جدول ۴. تحلیل روندهای علمی کیفیت داده

نام روند	موضوع	میانگین تعداد عضو	میانگین رشد	میانگین ارجاع مقالات	نوع روند
THR1	استراتژی، مدیریت	۱۱۸۴/۱۲	۳۶/۰۶۷	۳۶۰۲/۶	خطی - صعودی
THR2	ارزیابی کیفیت داده	۵۹۶/۸۸	۱۲/۸۵۸	۲۳۸۳/۱	خطی - صعودی
THR3	عوامل موفقیت و رضایت کاربران	۹۹۳/۳	۲۶/۰۶۳	۶۳۳۲/۸	خطی - صعودی
THR4	رفتار سازمانی	۷۱۴/۴۱	۱۱/۴۶۸	۴۸۱۸	خطی - صعودی
THR5	سازگاری منابع داده‌ای	۷۶۰	۲۱/۷۹۵	۳۸۴/۵	خطی - صعودی
THR6	ارتباط بین داده‌ها	۶۸۸/۶۲	۱۹/۰۸۲	۷۳۳	خطی - صعودی
THR7	بانک اطلاعاتی (تکرار، زمینه داده)	۲۳۰۲	۲۹/۱۸۰	۱۵۸۱/۵	خطی - صعودی
THR8	ارتباط بین داده‌ها در بیوانفورماتیک	۹۸۸/۲۵	۲۴/۱۲۴	۱۰۲۵۱/۴	خطی - صعودی
THR9	کیفیت در داده‌کاوی	۱۰۵۳/۵	۳۱/۷۰۵	۲۲۴۰	خطی - صعودی
THR10	کیفیت داده در DNA	۵۹۸/۲۵	۱۳/۴۱۸	۱۴۵۲	خطی - صعودی
THR11	داده‌های عظیم (کلان‌داده)	۶۶۹/۵	۲/۶۸۴	۱۹۶	خطی - صعودی
THR12	سیستم‌های اطلاعاتی سازمانی	۵۷۴	۱۱/۲۲۶	۴۱۹	خطی - صعودی
THR13	ارتباط اجزا در سیستم‌های پزشکی	۶۳۳/۷۵	۳/۷۸۵	۱۴۱۵	خطی - ساکن
THR14	کیفیت در ارجاعات و علم‌سنجی	۶۳۴/۱۶	۵/۱۴۹	۸۶۰/۵	خطی - ساکن
THR15	کیفیت داده و ژنتیک	۷۵۲/۵	۱۵/۰۴۶	۹۵۲/۶	خطی - صعودی
THR16	کیفیت در تیم‌سازی	۴۶۶/۶۶	-۶/۶۱۴	۱۱۰۰/۲	خطی - نزولی
THR17	کیفیت در داده‌کاوی	۶۱۸/۳۳	۱۵/۶۴۵	۸۸۸/۲	خطی - صعودی
THR18	کیفیت داده در DNA	۳۳۰	۱/۰۴۷	۸/۱	خطی - ساکن
THR19	داده‌های عظیم (کلان‌داده)	۷۳۶/۸۷	۲۸/۹۴۸	۲۴۰۸/۲	ادغامی - صعودی
THR20	سیستم‌های اطلاعاتی سازمانی	۶۹۷	۲۴/۴۶۴	۷۱۰/۵	خطی - صعودی

همان‌طور که در جدول ۴ مشخص شده است، تمام روندهای موجود در این حوزه علمی با استفاده از مدل ارائه‌شده قابل تشخیص است و روندهای حوزه کیفیت داده نیز مشاهده می‌شود. در این جدول حوزه‌ها به سه گروه صعودی، ساکن و نزولی و شیوه تغییر آنها نیز به سه گروه خطی، ادغامی و شکست دسته‌بندی شده است.

نکته مهم این است که موضوعات موجود در روندها ممکن است با یکدیگر تشابه زیادی داشته باشند و در حوزه‌های مختلفی ارائه شوند. برای نمونه، موضوع کیفیت داده در حوزه‌های مختلفی بررسی شد که در روندهای استخراج‌شده در جدول مشاهده می‌شود. در روندهای اصلی یک روند ادغامی وجود دارد که مربوط به حوزه انبار داده است. همچنین بسیار واضح است که

روندهای ساکن تعداد ارجاعات بسیار کمی دارند. بر اساس بررسی انجام شده به‌طور میانگین، بین رشد تعداد ارجاعات و رشد یک اجتماع علمی، همبستگی بالای ۸۲ درصد وجود دارد که نشان می‌دهد زنده بودن یک حوزه علمی تا حد بسیار زیادی وابسته به ارجاعات بین اعضای یک حوزه علمی است.

نتیجه‌گیری و پیشنهادها

در این مقاله مدلی مبتنی بر ارجاعات تجمعی برای تعیین تغییر و تحولات در حوزه‌های علمی ارائه شده است. به‌منظور آزمایش توانایی مدل در نشان دادن تغییرات علمی، مقالات حوزه کیفیت داده بررسی شد و روندهای اصلی این حوزه بر اساس مدل ارائه شده به‌دست آمد. بر اساس نتایج، بین رشد علمی و تعداد ارجاعات، همبستگی مثبت بالایی وجود دارد که نشان می‌دهد رشد و ادامه حیات یک حوزه علمی تا حد زیادی وابسته به تعداد ارجاعات است. نکته دیگری که می‌توان از مدل ارائه‌شده برداشت کرد، قابلیت آن برای کشف حوزه‌های جدید است. برای نمونه حوزه مرتبط با کیفیت داده‌های عظیم، حوزه‌ای است که نسبت به سایر حوزه‌ها عمر کوتاه‌تری دارد و در روندهای رو به رشد قابل مشاهده است.

پیشنهاد می‌شود برای ادامه این مسیر از سایر مدل‌های تشخیص اجتماعات در حوزه‌های علمی استفاده شود. همچنین می‌توان از مدل‌های دیگری به‌منظور استخراج دقیق‌تر مقالات حوزه کیفیت داده استفاده کرد. و دست آخر، می‌توان با استفاده از تحلیل سایر مشخصه‌های مرتبط با اجتماعات علمی، به شناسایی متغیرهای اصلی که می‌تواند به پیش‌بینی درست حوزه‌های رو به رشد منجر شود، اقدام کرد.

منابع

- Alexander, J., Chase, J., Newman, N., Porter, A. & Roessner, J. D. (2012). *Emergence as a conceptual framework for understanding scientific and technological progress*. Paper presented at the Technology Management for Emerging Technologies (PICMET), 2012 Proceedings of PICMET'12.
- Blondel, V.D., Guillaume, J.L., Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), 10008.
- Boyack, K. W., Klavans, R., Small, H. & Ungar, L. (2014). Characterizing the emergence of two nanotechnology topics using a contemporaneous global micro-model of science. *Journal of Engineering and Technology Management*, 32, 147-159.

- Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5303-5310.
- Cozzens, S., Gatchair, S., Kang, J., Kim, K.-S., Lee, H. J., Ordóñez, G. & Porter, A. (2010). Emerging technologies: quantitative identification and measurement. *Technology Analysis & Strategic Management*, 22(3), 361-376.
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation *Science* 178.4060 (1972): 471-479.
- Guo, H., Weingart, S. & Börner, K. (2011). Mixed-indicators model for identifying emerging research areas. *Scientometrics*, 89(1), 421-435.
- Kajikawa, Y., Yoshikawa, J., Takeda, Y. & Matsushima, K. (2008). Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy. *Technological Forecasting and Social Change*, 75(6), 771-782.
- Kuhn, T. S. & Hawkins, D. (1963). The structure of scientific revolutions. *American Journal of Physics*, 31(7), 554-555.
- Lee, W. H. (2008). How to identify emerging research fields using scientometrics: An example in the field of Information Security. *Scientometrics*, 76(3), 503-525.
- Madlock-Brown, C. R. (2014). A framework for emerging topic detection in biomedicine. Diss. THE UNIVERSITY OF IOWA
- Mecella, M., Scannapieco, M., Virgillito, A., Baldoni, R., Catarci, T. & Batini, C. (2002). Managing data quality in cooperative information systems. *Paper presented at the OTM Confederated International Conferences, On the Move to Meaningful Internet Systems. CoopIS, DOA, and ODBASE. OTM 2002. Lecture Notes in Computer Science*, vol 2519. Springer, Berlin, Heidelberg.
- Price, D. J. (1970). *Citation measures of hard science, soft science, technology, and nonscience*. Communication among scientists and engineers, 3-22 .
- Sadiq, S., Yeganeh, N. K., & Indulska, M. (2011). 20 years of data quality research: themes, trends and synergies. *Paper presented at the Proceedings of the Twenty-Second Australasian Database Conference*, 115, 153-162.
- Small, H., Boyack, K. W. & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy*, 43(8), 1450-1467 .
- Small, H. G. (1977). A co-citation model of a scientific specialty: A longitudinal study of collagen research. *Social studies of science*, 7(2), 139-166 .
- Tu, Y.-N. & Seng, J.-L. (2012). Indices of novelty for emerging topic detection. *Information processing & management*, 48(2), 303-325.
- Wang, X., Cheng, Q. & Lu, W. (2014). Analyzing evolution of research topics with NEViewer: a new method based on dynamic co-word networks. *Scientometrics*, 101(2), 1253-1271.