مهندسی صنایع و مدیریت شریف، (تابستان ۱۳۹۹)
دوره‌ی ۱-۳۶، شماره‌ی ۱/۱، ص. ۷۳-۸۵

Original Article

اولویت‌بندی شاخص‌های کیفیت داده‌های پژوهشی بر پایه‌ی روش‌های تصمیم‌گیری با معیارهای چندگانه

هدیه کرم نیای فر (دانشجوی کارشناسی ارشد)
دانشکده‌ی مهندسی صنایع ـ مدیریت سیستم و بهره‌وری، دانشگاه آزاد اسلامی، واحد علوم تحقیقات
محمدجواد ارشادی * (استادیار)
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
فرزاد موحدی سبحانی (استادیار)
دانشکده‌ی فنی و مهندسی، دانشگاه آزاد اسلامی، واحد علوم تحقیقات

امروزه داده‌های پژوهشی که در حین پژوهش تولید و اشاعه داده می‌شوند، نقشی تأثیرگذار در بالندگی اقتصاد دانش‌بنیان بر عهده دارند. در این پژوهش، به منظور ارزیابی کیفیت داده‌های پژوهشی، معیارهای اصلی تأثیرگذار بر روی کیفیت داده‌ها در بعدهای ذاتی، دسترس‌پذیری، متنی و نمایشگری دسته‌بندی شد. سپس با استفاده از مقایسه‌ی زوجی و روش دیمتل فازی بر مبنای ANP فازی به تعیین ارتباطات درونی و بیرونی عوامل و زیرعوامل پرداخته شد. کیفیت ذاتی داده‌ها نه تنها دقت و عینیت را در بردارد، بلکه باورپذیری و اعتبار عمومی را نیز شامل می‌شود. دسترس‌پذیری خود شامل ابعاد فرعی دسترسی و امنیت است. کیفیت متنی داده‌ها شامل ارزش افزوده داشتن، مرتبط بودن، زمان‌مند بودن، کامل بودن و کافی بودن داده‌هاست. قابلیت نمایشگری داده شامل چارچوب ارائه و معنای داده‌ها (تفسیرپذیری و درک‌پذیری) است. نتایج نشان داد در دسترس بودن، امنیت و سهولت درک تأثیرگذارترین عوامل مؤثر بر کیفیت داده‌های پژوهشی محسوب می‌شوند. طراحی لایه‌های امنیتی در سامانه‌ی ثبت داده‌های پژوهشی و نیز طراحی فیلدهای پایگاه داده به منظور افزایش سهولت درک کاربران از جمله پیشنهادهای کاربردی این پژوهش هستند.

واژگان کلیدی: سامانه‌های اطلاعاتی، کیفیت داده‌ها، دیمتل فازی، ANP فازی، FDANP.

hediehkaramnia@gmail.com
mjershadi@gmail.com
farzad1348@yahoo.com

۱. مقدمه

یک محصول اطلاعاتی همانند محصولات فیزیکی (مثلاً یک اتومبیل) دارای ابعادی است که مرتبط با کیفیت آن، اطلاعاتی را به ما می‌دهد و کیفیت آن را مشخص می‌کند. کیفیت سامانه‌های اطلاعاتی از اجزای اصلی و حیاتی مهمی چون فنّاوری، داده‌ها و افراد درگیر در ارائه‌ی اطلاعات و خدمات تشکیل شده‌اند. همچنین خدمات متنوعی که بشر در حوزه‌های مختلف فرهنگی، اجتماعی و اقتصادی دریافت می‌کند، مبتنی بر سامانه‌های اطلاعاتی است. از سوی دیگر در تمام سامانه‌های اطلاعاتی پیچیدگی ساختار نرم‌افزار و دشواری یکپارچه ساختن نرم‌افزارها با یکدیگر یکی از مشکلات اساسی است که ارائه‌ی خدمات از سوی سامانه‌های اطلاعاتی را با دشواری روبه‌رو ساخته است. سامانه‌های اطلاعاتی تحقیقاتی یکی از اساسی‌ترین بخش‌های سامانه‌های اطلاعاتی‌اند که به جمع‌آوری، کنترل و اشاعه‌ی تحقیقات علمی می‌پردازند و با توجه به اهمیت این موضوع لازم است تا از سطح کیفیت مطلوبی برخوردار باشند. داده‌ها یک دارایی اطلاعاتی اساسی هستند و تمرکز اصلی امنیت سیستم‌های اطلاعاتی، وابسته به داده‌هاست.[۱] از این‌رو امروزه کیفیت داده‌ها و فراداده‌هایی که به کار گرفته می‌شوند در فرایند آن کسب‌وکار تأثیر حائز اهمیتی می‌گذارند و کیفیت داده نقش بسیار مهمی را در بسیاری از سازمان‌ها ایفا می‌کند. داده‌ها در سازمان‌ها نیز منبع حیاتی برای حمایت از فرایند کسب‌وکار و تصمیم‌گیری مدیریتی هستند.[۲] کیفیت داده با عنوان شایستگی کارکرد مطرح شده است. برای بالا بردن کیفیت داده‌ها و اندازه‌گیری آن روش‌های گوناگونی مطرح شده است و از این‌رو برای بالا بردن کیفیت داده‌ها در سازمان‌ها این روش‌ها توسعه یافته‌اند.[۳] امروزه با پیشرفت فنّاوری اطلاعات با حجم زیادی از داده‌ها مواجه هستیم. بدین ترتیب با افزایش حجم داده‌ها از نظر اندازه و دامنه در سال‌های اخیر مدیریت آنها پیچیده‌تر شده و

* نویسنده مسئول
تاریخ: دریافت ۱۳۹۷/۱۰/۱۶، اصلاحیه ۱۳۹۸/۴/۹، پذیرش ۱۳۹۸/۵/۲.
DOI:10.24200/J65.2019.52158.1934

خطر کاهش کیفیت داده‌ها رو به افزایش است. داده‌های نامرغوب نیز می‌توانند بر قابلیت سیستم و همچنین تضمین عملکرد عملیاتی زیان وارد کنند.[۴] از سوی دیگر داده‌های نامرغوب می‌توانند اثرات اجتماعی و اقتصادی قابل توجهی بگذارند.[۵] همچنین بر رضایت مشتریان و روحیه‌ی کارکنان و اطمینان میان سازمان‌ها تأثیرگذار است.[۱] در سیستم‌های اطلاعات شبکه‌یی، اگر اطلاعات ورودی و کیفیت فرایندها کنترل نشده باشد، می‌تواند بر روی کیفیت کلی داده‌هایی که در سیستم اطلاعات وجود دارد، تأثیر بگذارد و به مرور زمان کیفیت داده‌ها را کاهش دهد. به کمک روش‌های تحلیل ریشه می‌توان به بهبود سامانه‌های اطلاعاتی پژوهشی و داده‌های حاصل از آن پرداخت.[۶]

همان طورکه گفته شد، عامل‌های بسیاری بر فرایندهای تصمیم‌گیری و نتایج آنها تأثیرگذار است و داده‌های نامرغوب یکی از آنهاست، که باعث تصمیمات ضعیف می‌شود. مسئله‌ی کیفیت داده[۱]، به دلیل وجود حجم بالای داده‌های در دسترس، یک مسئله‌ی مهم در سازمان‌هاست. تحقیقات حاکی از آن است که تصمیم‌گیرندگانی که با DQ آشنایی داشته‌اند از آن بیشتر برای حل در تصمیم‌گیری استفاده کرده‌اند و همچنین دقت بالاتری در تصمیمات خود داشته‌اند.[۷]

همان‌گونه که از بررسی پیشینه‌ی پژوهش برمی‌آید در کیفیت داده‌ها از دیدگاه نظری و به ویژه از دیدگاه کاربردی پژوهش‌های زیادی صورت نگرفته است. در این تحقیق به منظور وزن‌دهی و اولویت‌بندی عوامل و زیرعوامل تأثیرگذار بر کیفیت فراداده‌های سامانه‌ی ملی ثبت پایان‌نامه‌های داخل کشور (به عنوان یک نمونه‌ی کاربردی) در پژوهشگاه علوم و فنّاوری اطلاعات ایران (ایرانداک)، که مطالعه‌ی موردی تحقیق حاضر است، از روش دیمتل فازی برپایه‌ی ANP فازی استفاده شده است.

۲. مبانی نظری تحقیق

۱.۲. کیفیت داده

کیفیت داده‌ها خود داده‌یی است که برای مصرف‌کنندگان داده مناسب است. علاوه بر این ابعاد کیفیت داده‌ها مجموعه‌یی از ویژگی‌ها که نشان‌دهنده‌ی یک جنبه یا ساختار DQ است، در نظر گرفته می‌شود.[۵] کیفیت داده‌ها یک مفهوم چندبعدی است.[۸] از این‌رو برای درک بهتر، اندازه‌گیری و بهبود کیفیت داده، ابعادی مشخص شده است.[۳] در این تحقیق از ابعاد کیفیت داده که توسط وانگ و استرانگ در سال ۱۹۹۶ عنوان شده است، استفاده شد. این ابعاد که شامل ۱۵ بعد هستند در ۴ دسته گروه‌بندی شده‌اند. این دسته‌بندی به صورت زیر است:[۹]

۱. ذاتی؛

۲. دسترس‌پذیری؛

۳. متنی؛

۴. نمایشگری.

ابعاد نمایش سازگار و جامعیت، دو بعد مهم در میان ابعاد ۱۵ گانه‌ی کیفیت داده‌ها به حساب می‌آیند و از شروط حیاتی برای زیربنای فنّاوری اطلاعات و منبع کسب‌وکار بزرگ هستند. همچنین دارایی بسیار مهمی برای دست‌یابی به مزایای رقابتی بلند مدت محسوب می‌شوند.[۱۰]

۲.۲. ابعاد کیفیت داده‌ها

برای اندازه‌گیری ابعاد کیفیت داده‌ها، ابزارها یا معیارهایی مشخص شده است، که این معیارها یا به صورت ذهنی[۲] یا به صورت عینی[۳] هستند. معیارهای ذهنی بر اساس نظرات و تجربه‌های کاربران داده که می‌توانند مدیران یا متخصصان داده باشند، مشخص می‌شوند. سنجش ذهنی، معمولاً با استفاده از مصاحبه‌هایی که مربوط به داده‌ها یا پرسش‌نامه‌هاست، صورت می‌پذیرد. که منعکس‌کننده‌ی نیازها و تجربه‌های ذی‌نفعان از جمله گردآورندگان، نگهبانان و مصرف‌کنندگان داده می‌باشد. از طرف دیگر سنجش عینی بر مبنای فرمول‌های ریاضی است که برای اندازه‌گیری یک مجموعه داده استفاده می‌شود.[۳]

همین طور کیفیت‌داده بستگی زیادی به سازمان‌دهی سیستم اطلاعاتی و چگونگی پردازش داده دارد و اندازه‌گیری و بهبود کیفیت داده در سازمان‌ها وظیفه‌یی پیچیده است از این‌رو برای بررسی کیفیت فراداده‌ها، از ابعاد مدل استاندارد توسط وانگ و استرانگ استفاده شده است. این ابعاد عبارت‌اند از: دقت، عینیت، باورپذیری، معتبربودن، مربوط بودن، ارزش افزوده، به‌روز بودن، جامعیت، مقدار داده‌ها، تفسیرپذیری، سهولت درک، نمایش مختصر، نمایش سازگار، دردسترس بودن و امنیت. که به طور مشخص در ۴ دسته با نام‌های ذاتی، متنی، نمایشگری و دسترس‌پذیری قرار می‌گیرند.(جدول ۱)[۹]

جدول ۱. ابعاد کیفیت داده‌ها.

نام بعد	
ذاتی [۱]	دقت [۲]
	عینیت [۳]
	باورپذیری [۴]
	معتبر بودن [۵]
دسترس‌پذیری [۶]	در دسترس بودن [۷]
	امنیت [۸]
متنی [۹]	مربوط بودن [۱۰]
	ارزش افزوده [۱۱]
	به روز بودن [۱۲]
	جامعیت [۱۳]
	مقدار داده‌ها [۱۴]
نمایشگری ۱۵	تفسیر پذیری [۱۶]
	سهولت درک ۱۷
	نمایش مختصر [۱۸]
	نمایش سازگار [۱۹]

[۱] intrinsic	[۲] accuracy
[۳] objectivity	[۴] believability
[۵] reputation	[۶] accessibility
[۷] access	[۸] security
[۹] contextual	[۱۰] relevancy
[۱۱] value added	[۱۲] timelines
[۱۳] completeness	[۱۴] amount of data
[۱۵] representational	[۱۶] interpretability
[۱۷] ease of understanding	[۱۸] concise representation
	[۱۹] consistent representation

۳.۲. کیفیت داده در سیستم‌های اطلاعاتی

روش‌های کیفیت داده توسط نوع سیستم‌های اطلاعاتی که به فعالیت‌های ارزیابی و بهبود مربوط می‌شوند نیز تحت تأثیر قرار می‌گیرند. این متون مفهوم معماری سیستم‌های اطلاعاتی (معماری IS [۴]) را برای توصیف مدل هماهنگی که توسط سیستم اطلاعاتی شرکت پشتیبانی شده است، فراهم می‌کنند.[۱۱]

معماری انواع سیستم‌های اطلاعاتی بر اساس درجه‌ی داده، پردازش و ادغام مدیریت پشتیبانی شده توسط یک سیستم فنی، متمایز شده‌اند. همان طورکه درجه‌ی ادغام، پردازش و مدیریت داده کاهش می‌یابد، ارزیابی کیفیت داده و روش‌های بهبود که اعمال می‌شوند، پیچیده‌تر می‌شوند. هم‌زمان، ارزیابی و بهبود کیفیت داده‌ها هم پیچیده‌تر می‌شود. انواع سیستم‌های اطلاعاتی زیر را می‌توان بر اساس درجه‌ی ادغام آنها متمایز کرد:

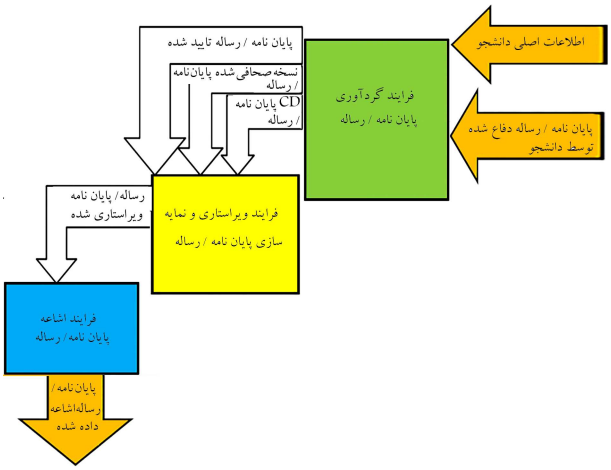
- در یک سیستم اطلاعات یکپارچه در حقیقت نرم‌افزارها تک‌لایه‌یی هستند که دسترسی به داده‌ها را ارائه نمی‌دهند. اگرچه معمولاً داده‌ها در پایگاه داده‌یی مرتب می‌شوند که قابل جستجو باشند، اما نرم‌افزارهای جداگانه را به اشتراک نمی‌گذارند. این امر می‌تواند موجب تکرار داده‌ها شود، که احتمالاً بر تمام ابعاد کیفیت نیز مؤثر است.
- یک انبار داده مجموعه‌یی متمرکز از اطلاعات بازیابی شده از پایگاه داده‌های متعدد است. انبار داده به صورت دوره‌یی با داده‌های به‌روز از پایگاه داده‌ی اصلی به وسیله‌ی استخراج روندهای خودکار و هماهنگی داده‌ها، بهنگام‌سازی می‌شود. داده‌ها از لحاظ فیزیکی نیز یکپارچه می‌شوند؛ چون آنها با توجه به طرح، ادغام و در نهایت هنگام ذخیره در انبار داده‌ها مجدداً فرمت‌دهی می‌شوند.
- سیستم اطلاعات توزیعی، مجموعه‌یی از ماژول‌های نرم‌افزاری پشتیبانی شده توسط گردش کار است. نرم‌افزارها معمولاً به ردیف‌هایی مانند نمایش‌ها، منطق برنامه، مدیریت داده‌ها و ویژگی‌های دسترسی به داده‌های خروجی تقسیم می‌شود. در این حالت داده‌ها را می‌توان در پایگاه داده‌های مختلف ذخیره کرد، اما قابلیت همکاری توسط ادغام منطقی طرح آنها تضمین شده می‌ماند.
- سیستم اطلاعات تعاونی (CIS) [۵] را می‌توان به عنوان یک سیستم اطلاعاتی در مقیاس بزرگ نیز تعریف کرد که سیستم‌های مختلف در سازمان‌های مختلف و مستقل را برای به اشتراک‌گذاری اهداف مشترک به صورت داخلی به هم وصل می‌کند.[۱۲] هماهنگی با دیگر سیستم‌های اطلاعاتی نیازبه توانایی تبادل اطلاعات دارد. در سیستم‌های اطلاعات تعاونی، داده‌های منطقی یکپارچه نیست؛ زیرا در پایگاه داده‌های جداگانه با توجه به طرح‌های مختلف ذخیره می‌شوند. بااین‌حال، نرم‌افزارها روند انتقال و تبادل داده‌ها را ترکیب می‌کنند تا بتوانند موجب همکاری میان فرایندهای رایج شوند. به عبارت دیگر، ادغام در سطح فرایند تحقق می‌یابد.
- در متون علمی برای نشان دادن هر نوع اطلاعات اتخاذ شده از فنّاوری‌های وب، از اصطلاح سیستم اطلاعاتی وب (WIS) [۶] استفاده می‌شود.[۱۳] از دیدگاه فنی سیستم اطلاعاتی وب یک نرم‌افزار خدمت‌گیرنده/خدمت‌دهنده است. چنین سیستمی معمولاً از داده‌های ساختاریافته، نیمه‌ساختاریافته و ساختارنیافته استفاده می‌کند و توسط ابزارهای توسعه و مدیریت بر اساس روش‌های خاص برای هر نوع داده پشتیبانی می‌شود.
- در یک سیستم اطلاعاتی همتا به همتا ($P2P$) [۷]، تمایزی بین کاربر و سرور وجود ندارد. این سیستم توسط مجموعه‌یی از گره‌های یکسان تشکیل شده است که داده‌ها و خدمات نرم‌افزاری را به منظور برآوردن نیازهای کاربران به اشتراک می‌گذارد. سیستم‌های $P2P$ توسط چند خاصیت مشخص می‌شوند:
  - نداشتن هماهنگی مرکزی؛
  - نداشتن پایگاه داده مرکزی؛
  - همتا نداشتن در دیدگاه جامع سیستم؛
  - همتایان آن‌ها خودمختارند و می‌توانند به صورت پویا به سیستم وصل یا از سیستم قطع شوند. با این‌حال، همتایان معمولاً روش‌های مدیریت مشترک را به اشتراک می‌گذارند.

۴.۲. روش‌های مدیریت کیفیت داده و اطلاعات

جدول ۲ فهرست روش‌های مورد توجه در این پژوهش را توسط کلمات اختصاری همراه با نام کامل روش و مرجع اصلی نشان می‌دهد. همچنین از علائم به صورت مخفف برای اشاره به هر روش در ادامه‌ی این پژوهش استفاده خواهد شد.

۵.۲. سامانه‌ی ثبت

سامانه‌های گردآوری و ثبت، سازمان‌دهی و اشاعهٔ اطلاعات پایان‌نامه‌ها/رساله‌های دانش‌آموختگان داخل کشور یکی از مهم‌ترین سامانه‌های اطلاعاتی تحقیقاتی هستند که رسالت اصلی آن حفظ و اشاعه‌ی پایان‌نامه‌ها و رساله‌های داخل کشور است و روزانه بازدیدکنندگان فراوانی دارد.[۱۳] از آن‌جا که پایان‌نامه‌ها و رساله‌های اشاعه داده شده در سامانه‌ی گنج [۸] پایه و اساس پژوهش خیل کثیری از پژوهشگران در مقطع کارشناسی ارشد و دکتری قرار خواهد گرفت، ارتقای کیفیت این سامانه تأثیری انکارناپذیر در افزایش کیفیت پژوهش‌های صورت گرفته از سوی پژوهشگران خواهد داشت و ضرورت انجام یک تحقیق جامع برروی دسته‌بندی مشکلات کیفی این سامانه را دو چندان می‌کند. نمودار جریان ورودی و خروجی‌ها در سامانه‌های اطلاعاتی گردآوری و ثبت، سازمان‌دهی و اشاعه‌ی اطلاعات پارسا (پایان‌نامه/رساله) در شکل زیر درج شده است. مطابق با (شکل ۱) فرایندهای اصلی شامل فراهم‌آوری، سازمان‌دهی و تحلیل اطلاعات واشاعه‌ی پایان نامه/رساله است.[۲۹]



شکل ۱. ورودی و خروجی سامانه‌های گردآوری و ثبت، سازمان‌دهی و اشاعه‌ی اطلاعات پارسا.

جدول ۲. مدل‌های مختلف کیفیت داده‌ها.

مخفف روش	نام اصلی	مراجع
TDQM	مدیریت کیفیت داده (Total Data Quality Management)	Wang ۱۹۹۸ [۱۴] Kunz et al. ۲۰۱۹ [۱۵]
DWQ	روش شناسی کیفیت انبار داه (The Data Warehouse Quality Methodology)	Jeusfeld et al/ ۱۹۹۸ [۱۶] DiTria et al. ۲۰۱۸ [۱۷]
TIQM	مدیریت کیفیت اطلاعات کلی	English ۱۹۹۹ [۱۸]، Vaziri et al. ۲۰۱۷ [۱۵]
AIMQ	روشی برای ارزیابی کیفیت اطلاعات	Lee et al. ۲۰۰۲ [۱۹]، Kunz et al. ۲۰۱۹ [۱۵]
CIHI	روش مؤسسه کانادایی برای اطلاعات سلامت	Leeb et al/ ۲۰۰۵ [۲۰]
DQA	ارزیابی کیفیت داده	Pipino et al. ۲۰۰۲ [۲۱] Bacal et al. ۲۰۱۹ [۲۲]
IQM	اندازه‌گیری کیفیت اطلاعات	Eppler&Munzenmaier ۲۰۰۲ [۲۳] Kunz et al/ ۲۰۱۹ [۱۵]
ISTAT	روش ISTAT	falorisi et al. ۲۰۰۳ [۲۴]
AMEQ	روش اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول	Su and Jin ۲۰۰۴ [۲۵] Kunz et al. ۲۰۱۹ [۱۵]
DaQuinCISData	کیفیت در سیستم‌های اطلاعات تعاونی	Scannapieco et al. ۲۰۰۴ [۲۶] Kunz et al. ۲۰۱۹ [۱۵]
QAFD	روشی برای ارزیابی کیفیت داده‌های مالی	De Amicis and Batini ۲۰۰۴ [۲۷] Kunz et al. ۲۰۱۹ [۱۵]
CDQ	روشی جامع برای مدیریت کیفیت داده	Batini and Scannapieco ۲۰۰۶ [۲۸]

- فرایند فراهم‌آوری اطلاعات

فرایند فراهم‌آوری اطلاعات شامل تهیه‌ی مدارک و ثبت اطلاعات است که در حین اجرای آنها، عملیات کنترل کیفی اجرا خواهد شد تا از صحت و بدون مشکل بودن اطلاعات اطمینان حاصل شود.

- فرایند تجزیه و تحلیل اطلاعات

این فرایند شامل ویرایش اطلاعات دریافتی و نمایه‌سازی اطلاعات است. باتوجه به این‌که اطلاعات وارد شده در این بخش قبلاً توسط کاربران ثبت شده است، لازم است تمام نواقص در این مرحله اصلاح شود. همچنین کلیدواژه‌های مؤلف بررسی و کنترل می‌شود و در صورت لزوم اصلاح صورت می‌پذیرد.

- فرایند اشاعه‌ی اطلاعات تحقیقاتی

با توجه به این که در این بخش از اطلاعات خروجی بخش‌های قبلی استفاده می‌شود، در صورت وجود و شناسایی مشکلات کیفی این اطلاعات دوباره به واحد تجزیه و تحلیل اطلاعات بازگردانده می‌شود.

در ادامه، به توضیح ورودی و خروجی‌ها به ترتیب توالی مطابق آن‌چه که در (شکل ۱) مشاهده می‌شود، می‌پردازیم:

ورودی‌های واحد فراهم‌آوری پارسا شامل موارد زیر است:

۱. اطلاعات اصلی دانشجو؛

۲. پیشنهاده‌ی (پروپوزال) دانشجو؛

جدول ۳. روند پژوهش‌های پیشین در اندازه‌گیری کیفیت داده.

پژوهشگران	سال	عنوان پژوهش	نوآوری پژوهش
وانگ و استرانگ[۳۰]	۱۹۹۶	Beyond accuracy: what data quality means to data consumers?	برای اولین بار ارائه ی مدلی برای ارزیابی کیفیت داده
باتینی و همکاران[۳۱]	۲۰۰۹	Methodologies for data quality assessment and improvement	مروری جامع بر انواع مدل‌های کیفیت داده
خسروانجم و همکاران[۲]	۲۰۱۱	Using fuzzy AHP for evaluating the dimensions of data quality	تعیین اهمیت شاخص‌های کیفیت داده به کمک روش AHP
موگس و همکاران[۷]	۲۰۱۶	Determinig the use of data quality metadata (DQM) for decision making purposes and its impact on decision otucomes-an explorateory study	ارزیابی جایگاه کیفیت داده در تصمیم‌گیری‌های حیاتی سازمان
وزیری و همکاران[۳]	۲۰۱۷	Measuring data qualitiy with weighted metrics	اندازه‌گیری کیفیت داده با وزن‌دهی به برخی داده‌ها

۳. رساله یا پایان‌نامه‌ی دانشجو

خروجی این بخش که ورودی واحد سازمان‌دهی و تحلیل اطلاعات است شامل ۳ بخش زیر است:

۱. پارسای تأیید شده؛

۲. نسخه‌ی صحافی شده‌ی پارسا؛

۳. سی‌دی پارسا.

خروجی این بخش که ورودی فرایند اشاعه‌ی پارساست، مدارک نمایه‌سازی و ویراستاری شده است. در انتها خروجی فرایند اشاعه‌ی پارسا، مدارک اشاعه داده شده است. از آن‌جا که داده‌های پژوهشی در فرایندهای سه‌گانه‌ی توضیح داده شده جریان دارد و در انتها مورد استفاده‌ی میلیون‌ها پژوهشگر قرار می‌گیرد، کیفیت این داده‌ها مسئله‌ی مهمی به حساب می‌آید که لازم است از جنبه‌های مختلف و بر پایه‌ی مدل‌های استاندارد ارزیابی شود.

۲.۶. جمع‌بندی پژوهش‌های پیشین و بیان نوآوری پژوهش حاضر

همان طور که پیشتر به آن اشاره شد، کیفیت داده‌های پژوهشی و ارزیابی آنها از اهمیت ویژه‌یی برخوردار است. در این راستا، سامانه‌ی ثبت پایان‌نامه‌ها/رساله‌های داخل کشور، مهم‌ترین پایگاه اطلاعاتی ملی ثبت و اشاعه‌ی پژوهش‌های کشور، در قالب یک مطالعه‌ی موردی انتخاب شد. از آن‌جا که ارزیابی کیفیت داده‌ها جنبه‌ها و ابعاد مختلفی دارد، مدل‌های استاندارد و بین‌المللی این حوزه بررسی شد (جدول ۲). مسئله‌ی اصلی پژوهش کنونی تعیین مهم‌ترین ابعاد و معیارهای ارزیابی کیفیت در داده‌های پژوهشی است که بتوان با تمرکز بر آنها در سامانه‌ی ثبت به بهبود فرایند پژوهش و توسعه‌ی اقتصاد دانش‌بنیان کمک کرد. علاوه بر مقالات گوناگونی که به معرفی مدل‌های کیفیت داده پرداخته‌اند (جدول ۲)، پژوهش‌هایی نیز به طور ویژه به چگونگی وزن‌دهی و اولویت‌بندی ابعاد کیفیت داده توجه کرده‌اند (جدول ۳). روند تاریخی اصلی این پژوهش‌ها به صورت زیر است: پس از توسعه‌ی اولیه مدل‌های کیفیت داده توسط وانگ و استرانگ[۳۰] مدل‌ها و روش‌های گوناگونی ارائه شدند که باتینی و همکاران[۳۱] در پژوهش خود به طورکامل به معرفی آنها پرداختند. پس از گسترش استفاده از مدل‌های کیفیت داده خسروانجم و همکاران[۲] در پژوهش خود به موضوع وزن‌دهی به شاخص‌های کیفیت داده پرداختند و از روش AHP [۹] را استفاده کردند. موگس و همکاران[۷] در پژوهش خود به نقش تصمیم‌گیری و برهم‌کنش آن با کیفیت داده اشاره کردند. همچنین وزیری و همکاران[۳] به تعیین وزن برخی داده‌های یک پایگاه در مقایسه با سایر داده‌های آن در ارزیابی کیفیت داده‌های یک پایگاه پرداخت و روابطی را در این خصوص ارائه کرد. در پژوهش حاضر با توجه به اثر متقابل شاخص‌های کیفیت داده (که در پژوهش‌های پیشین مورد توجه قرار نگرفته است) از روش FDANP به تعیین اثرگذاری ابعاد و اولویت‌دهی آنها در ارزیابی کیفیت داده‌های پژوهش پرداخته شده است. ابتدا با استفاده از روش دیمتل فازی به بررسی روابط میان عوامل و زیرعوامل مورد بحث می‌پردازیم و پرسش موجود درمورد چگونگی روابط این عوامل را پاسخ می‌دهیم؛ که کدام‌یک از عوامل تأثیرگذار و کدام عوامل تأثیرپذیر هستند. سپس با ترکیب این روش با ANP فازی به وزن عوامل و درجه اهمیت آنها می‌پردازیم و عوامل را رتبه‌بندی می‌کنیم. با استفاده از رتبه‌بندی به دست آمده می‌توان عوامل وزین‌تر و دارای اهمیت بالاتر را شناسایی و در جهت بالا بردن کیفیت سامانه به آنها پرداخت.

۳. فرایند روش دیمتل فازی

پروفسور لطفی‌زاده در سال ۱۹۶۵، با معرفی نظریه‌ی مجموعه‌های فازی، مقدمات مدل‌سازی اطلاعات نادقیق را فراهم کرد. در این تحقیق از اعداد فازی مثلثی استفاده شده است. دلیل استفاده از اعداد مثلثی، محاسبات ساده و قابل فهم آن است و همچنین ثابت شده است که در مسائلی که اطلاعات آن ذهنی و نادقیق است، استفاده از اعداد مثلثی مؤثراست. یک عدد فازی مثلثی می‌تواند به صورت سه‌تایی (a^l، a^m، a^u) نمایش داده شود. تابع عضویت عدد فازی مثلثی $\tilde{A}$ به صورت شکل ۲ است.

$$\mu_{\tilde{A}} = \begin{cases} 0 & x < a^l, x > a^u \\ \frac{x-a^l}{a^m-a^l} & a^l \le x \le a^m \\ \frac{a^u-x}{a^u-a^l} & a^l \le x \le a^m \end{cases}$$

دیمتل فازی یک مدل ساختاری برای تحلیل روابط علی و معلولی میان عوامل پیچیده، است. روش دیمتل توانایی ایجاد نقشه‌ی ساختاری از سیستم بر پایه‌ی روابط درونی، بر اساس تأثیرگذاری و تأثیرپذیری عامل‌ها را داراست.[۳۳]

جهت غربال‌گری عوامل مؤثر بر کیفیت داده‌ها که پیش‌تر به آن اشاره شد، از شیوه‌ی امتیازدهی استفاده شده است. بدین منظور از هریک از ۱۰ خبره خواسته شد که میزان اهمیت هر زیرعامل را براساس طیف لیکرت (۹ و ۱،۳،۵،۷) مشخص کند. سپس میانگین نظر خبرگان برای هر زیرعامل محاسبه شد و زیرعواملی که میانگین آنها پایین‌تر از عدد ۵ (امتیاز متوسط) بود، کنار گذاشته شد. بدین ترتیب ۱۴ زیرعامل از زیرعوامل فوق که بیشترین تأثیر را بر کیفیت فراداده‌های سامانه‌ی ملی ثبت پایان‌نامه‌های داخل کشور در پژوهشگاه علوم و فنّاوری اطلاعات ایران (ایرانداک) داشته‌اند، شناسایی شد.

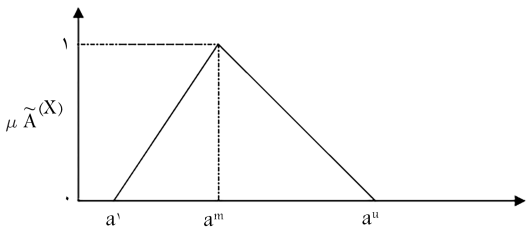
سپس گام‌های روش دیمتل فازی که در ادامه به آن خواهیم پرداخت، پیموده شد.

گام ۱. تشکیل گروه تصمیم‌گیری که در این مقاله یک گروه ۱۰ نفره متشکل از خبرگان در زمینه‌ی کیفیت داده‌هاست که در ایرانداک مشغول به خدمت‌اند.

گام ۲. غربال‌گری

در این تحقیق به منظور وزن‌دهی و اولویت‌بندی عوامل و زیرعوامل تأثیرگذار بر کیفیت فراداده‌های سامانه‌ی ملی ثبت پایان‌نامه‌های داخل کشور در پژوهشگاه علوم و فنّاوری اطلاعات ایران (ایرانداک) از روش دیمتل فازی بر مبنای ANP فازی استفاده شده است. بدین منظور برای مقایسه‌ی زیرعوامل اصلی از ۵ عبارت کلامی استفاده شده است که طیف فازی و معادل کلامی آن در جدول ۴ نشان داده شده است.

$$\mu_{\tilde{A}}(X) = \begin{cases} 0, & x < a^1 \\ (a - x^1)/(a^m - a^1), & a^1 \le x \le a^m \\ (a^u - x)/(a^u - a^m), & a^1 \le x \le a^m \\ 0, & x > a^u \end{cases}$$



شکل ۲. تابع عضویت و نمودار عدد فازی مثلثی.[۳۲]

جدول ۴. متغیرهای زبانی و اعداد فازی متناظر.[۳۲]

متغیر زبانی	معادل قطعی	اعداد فازی مثلثی متناظر
بی‌تأثیر	۰	(۰،۰،۰/۲۵)
تأثیر خیلی کم	۱	(۰، ۰، ۰/۲۵)
تأثیر کم	۲	(۰، ۰/۲۵، ۰/۵)
تأثیر زیاد	۳	(۰/۲۵، ۰/۵، ۰/۷۵)
تأثیر خیلی زیاد	۴	(۰/۷۵، ۱، ۱)

گام ۳. محاسبه‌ی ماتریس ارتباط مستقیم (D)

در این گام از پاسخ‌دهندگان خواسته شد تا میزان تأثیرگذاری معیار i بر معیار j را با استفاده از جدول ۴، نشان دهند. برای بررسی معیارها از نظر ۱۰ خبره استفاده شد. در این ماتریس‌ها، $\tilde{x}_{ij} = (l_{ij}, m_{ij}, u_{ij})$ اعداد فازی مثلثی هستند. برای در نظرگرفتن نظر همه‌ی خبرگان طبق رابطه‌ی ۱، از آنها میانگین حسابی گرفته می‌شود.

$$\tilde{z} = \frac{\tilde{x}^1 \oplus \tilde{x}^2 \oplus \tilde{x}^3 \oplus \ldots \oplus x^p}{p} \qquad (۱)$$

گام ۴. نرمال‌سازی ماتریس ارتباط مستقیم

سپس ماتریس میانگین را نرمال می‌کنیم و آن را ماتریس H می‌نامیم. برای نرمال‌سازی ماتریس به دست آمده از رابطه‌های ۲ و ۳ استفاده می‌کنیم.

$$\tilde{H}_{ij} = r \times \tilde{z}_{ij} = (r \times l'_{ij}, r \times m'_{ij}, r \times u'_{ij} = (l''_{ij}, m''_{ij}, u''_{ij}) \qquad (۲)$$

که r از رابطه‌ی زیر به دست می‌آید:

$$r = min \left\{ \frac{1}{max_{1 \le i \le n} \sum_{j=1}^{n} a_{ij}}, \frac{1}{max_{1 \le i \le n} \sum_{i=1}^{n} a_{ij}} \right\} \qquad (۳)$$

بنابراین تک‌تک عناصر جدول ماتریس ارتباط مستقیم را در (r) ضرب می‌کنیم.

گام ۵. محاسبه‌ی ماتریس ارتباط کامل زیرعامل‌ها ($d_i + r_j$)

بعد از محاسبه‌ی ماتریس‌های بالا، ماتریس روابط کل فازی با توجه به روابط ۴ تا ۷ به دست می‌آید.

$$T = \lim_{k \to +\infty} \left(\tilde{H}^1 \oplus \tilde{H}^2 \oplus \ldots \oplus \tilde{H}^k \right) \qquad (۴)$$

که هر درایه‌ی آن عدد فازی به صورت $\tilde{t}_{ij} = \left(l^t_{ij}, m^t_{ij}, u^t_{ij}\right)$ است و به صورت زیر محاسبه می‌شود:

$$\left[l^t_{ij}\right] = H_l \times (I - H_l)^{-1} \qquad (۵)$$

$$\left[m^t_{ij}\right] = H_m \times (I - H_m)^{-1} \qquad (۶)$$

$$\left[u^t_{ij}\right] = H_u \times (I - H_u)^{-1} \qquad (۷)$$

در این فرمول‌ها I ماتریس یکه و H_l، H_m و H_u هرکدام ماتریس $n \times n$ هستند، که درایه‌های آن را به ترتیب عدد پایین، عدد میانی و عدد بالاییِ اعداد فازیِ مثلثیِ ماتریسِ H تشکیل می‌دهد.

گام ۶. محاسبه‌ی ماتریس ارتباط کامل عوامل اصلی

نخست باید ماتریس T_D را از ماتریس ارتباط کامل زیر عامل‌ها T_C استخراج کرد.

هر درایه‌ی ماتریس T_D را اگر t_{ij} بدانیم، هر $n \times 1$ از میانگین هر T_C^{ij} حاصل می‌شود.

گام ۷. محاسبه‌ی شدت و جهت تأثیر

مطابق با رابطه‌های ۸ و ۹ میزان شاخص‌های r_i و C_j را محاسبه می‌کنیم. شاخص r_i بیانگر مجموع سطر iام و شاخص C_j بیانگر مجموع ستون jام از ماتریس T_C با توجه به بعد مربوطه است. به همین صورت میزان شاخص $\tilde{R}$

و $\tilde{D}$ را محاسبه می‌کنیم. شاخص R_I بیان‌گر مجموع سطر iام و شاخص C_j بیان‌گر مجموع ستون jام از ماتریس T_D است، برای ترسیم و تحلیل نمودار به ۲ شاخص شدت اثرگذاری و اثرپذیری و جهت تأثیر نیاز داریم که با استفاده از r_i و c_j به دست می‌آیند. برای هر $j = i$ خواهیم داشت:

$$\widetilde{D} = \left(\widetilde{D}_i\right)_{n\times 1} = \left[\sum_{j=1}^{n} \widetilde{T}_{ij}\right]_{n\times 1} \quad (8)$$

$$\widetilde{R} = \left(\widetilde{R}_i\right)_{1\times n} = \left[\sum_{j=1}^{n} \widetilde{T}_{ij}\right]_{1\times n} \quad (9)$$

که $\widetilde{D}$ و $\widetilde{R}$ به ترتیب ماتریس $n \times 1$ و $1 \times n$ هستند.

در مرحله‌ی بعدی میزان اهمیت شاخص‌ها $(\widetilde{D_i} + \widetilde{R_I})$ و رابطه‌ی بین معیارها $(\widetilde{D_i} - \widetilde{R_i})$ مشخص می‌شود. اگر $\widetilde{D}_i - \widetilde{R}_i > 0$، معیار مربوطه اثرگذار و اگر $\widetilde{D}_i - \widetilde{R}_i < 0$، معیار مربوطه اثرپذیر است.

$d_i - r_j$ شدت اثرگذاری و اثرپذیری (به عبارت دیگر هر چه مقدار T_D^{∞} عاملی بیشتر باشد، آن عامل تعامل بیشتری با سایر عوامل سیستم دارد.)

$d_i - r_j$ جهت تأثیرگذاری یا تأثیرپذیری را نشان می‌دهد. (بدین صورت که اگر $d_i - r_j > 0$ باشد معیار مربوطه اثرگذار و اگر $d_i - r_j < 0$ باشد معیار مربوطه اثرپذیر است.)

با توجه به مقادیر محاسبه شده، مقدار شاخص $ri + dj$ و $r_i - d_j$ را برای معیارها و همچنین شاخص $\widetilde{D}_i + \widetilde{R}_I$ و $\widetilde{D}_i - \widetilde{R}_I$ را برای ابعاد به دست می‌آوریم و سپس با استفاده از رابطه‌ی ۱۰ فازی زدایی می‌کنیم:

$$\mathrm{B} = \frac{(\mathrm{a}_1 + \mathrm{a}_3 + 2 \times \mathrm{a}_2)}{4} \quad (10)$$

گام ۸. ترسیم نقشه‌ی روابط شبکه (NRM)

برای تعیین نقشه‌ی روابط شبکه (NRM)، پس از به دست آوردن ماتریس‌های T_C و T_D باید $D-R$ و $d-r$ ماتریس‌های عوامل اصلی و زیرعوامل را محاسبه کنیم. عواملی که $D-R$ بیشتری دارند، روی عواملی که $D-R$ کمتری دارند، تأثیرگذارند و بر عکس عواملی که $D-R$ کمتری دارند از عواملی که $D-R$ بیشتری دارند، تأثیر می‌پذیرند.

گام ۹. نرمال‌سازی ماتریس ارتباط کامل ابعاد (T_D^{∞})

با توجه به رابطه‌ی ۱۱ به نرمال‌سازی ماتریس T_D اقدام می‌کنیم. به این طریق که مجموع هر سطر از ماتریس T_D را با توجه به بعد مربوطه محاسبه، سپس عنصر هر سطر را بر مجموع عناصر همان سطر تقسیم و در پایان جای سطر و ستون را عوض می‌کنیم.

$$T_D = \begin{bmatrix} t_{11}^{D_{11}} & \dots & t_{11}^{D_{1j}} & \dots & t_{11}^{D_{1m}} \\ \vdots & & \vdots & & \vdots \\ t_{i1}^{D_{i1}} & \dots & t_{ij}^{D_{ij}} & \dots & t_{im}^{D_{im}} \\ \vdots & & \vdots & & \vdots \\ t_{m1}^{D_{m1}} & \dots & t_{mj}^{D_{mj}} & \dots & t_{mm}^{D_{mm}} \end{bmatrix}$$

$$\rightarrow \quad \begin{aligned} d_1 &= \sum_{j=1}^{m} t_{1j}^{D_{1h}} \\ d_i &= \sum_{j=1}^{m} t_{ij}^{D_{ih}} \quad , i = 1, 2, \dots, m \\ d_m &= \sum_{j=1}^{m} t_{mj}^{D_{mh}} \end{aligned}$$

$$T_C^{\infty 11} = \begin{bmatrix} t_{11}^{D_{11}}/d_1 & \dots & t_{11}^{D_{1j}}/d_1 & \dots & t_{11}^{D_{1m}}/d_1 \\ \vdots & & \vdots & & \vdots \\ t_{i1}^{D_{i1}}/d_i & \dots & t_{ij}^{D_{ij}}/d_i & \dots & t_{im}^{D_{im}}/d_i \\ \vdots & & \vdots & & \vdots \\ t_{m1}^{D_{m1}}/d_m & \dots & t_{mj}^{D_{mj}}/d_m & \dots & t_{mm}^{D_{mm}}/d_m \end{bmatrix}$$

$$= \begin{bmatrix} t_D^{\alpha_{11}} & \dots & t_D^{\alpha_{1j}} & \dots & t_D^{\alpha_{1m}} \\ \vdots & & \vdots & & \vdots \\ t_D^{\alpha_{i1}} & \dots & t_D^{\alpha_{ij}} & \dots & t_D^{\alpha_{im}} \\ \vdots & & \vdots & & \vdots \\ t_D^{\alpha_{m1}} & \dots & t_D^{\alpha_{mj}} & \dots & t_D^{\alpha_{mm}} \end{bmatrix} \quad (11)$$

گام ۱۰. نرمال‌سازی ماتریس ارتباط کامل معیارها (T_C^{∞}) و تشکیل ابرماتریس ناموزون

ماتریس T_C را با استفاده از روابط زیر نرمال می‌کنیم؛ به این طریق که در این گام مجموع هر سطر T_C^{ij} را با توجه به بعد مربوطه محاسبه و سپس در هر T_C^{ij}، هر عنصر بر مجموع عناصر سطر مربوط به خود تقسیم می‌شود. برای مثال اگر هر T_C^{∞} را شامل مجموعه‌یی از $T_C^{\infty ij}$ بدانیم، $T_C^{\infty 11}$ از نرمال‌سازی T_C^{11} به دست می‌آید. با انتقال ماتریس T_C^{∞}، ابرماتریس ناموزون به دست می‌آید.

$$d_{ci}^{11} = \sum_{j=1}^{m_1} t_{cij}^{11}, i = 1, 2, \dots, m_1 \quad (12)$$

$$T_C^{\infty 11} = \begin{bmatrix} t_{c11}^{11}/d_{c1}^{11} & \dots & t_{c1j}^{11}/d_{c1}^{11} & \dots & t_{c1m_1}^{11}/d_{c1}^{11} \\ \vdots & & \vdots & & \vdots \\ t_{ci1}^{11}/d_{ci}^{11} & \dots & t_{cij}^{11}/d_{ci}^{11} & \dots & t_{cim_1}^{11}/d_{ci}^{11} \\ \vdots & & \vdots & & \vdots \\ t_{cm_11}^{11}/d_{cm_1}^{11} & \dots & t_{cm_1j}^{11}/d_{cm_1}^{11} & \dots & t_{cm_1m_1}^{11}/d_{cm_1}^{11} \end{bmatrix}$$

$$= \begin{bmatrix} t_{c11}^{\infty 11} & \dots & t_{c1j}^{\infty 11} & \dots & t_{c1m_1}^{\infty 11} \\ \vdots & & \vdots & & \vdots \\ t_{ci1}^{\infty 11} & \dots & t_{cij}^{\infty 11} & \dots & t_{cim_1}^{\infty 11} \\ \vdots & & \vdots & & \vdots \\ t_{cm11}^{\infty 11} & \dots & t_{cm1j}^{\infty 11} & \dots & t_{cm1m_1}^{\infty 11} \end{bmatrix} \quad (13)$$

گام ۱۱. تشکیل ابرماتریس موزون

در این مرحله ماتریس T_D^{∞} را در ماتریس $W(T_C^{\infty})$ ضرب می‌کنیم. به این طریق که هر $t_D^{\infty ij}$ را در W_{ij} نظیر ضرب می‌کنیم.

گام ۱۲. محدود کردن ابرماتریس موزون

مطابق با رابطه‌ی ۱۴، ابرماتریس موزون آن قدر به توان (متوالی اعداد فرد) رسانده می‌شود تا تمام اعداد هر سطر هم‌گرا شوند (جداول ۵ و ۶).

جدول ۵. محاسبه‌ی شدت و جهت تأثیر عوامل اصلی.

وضعیت عامل	(Di-Ri) def	(Di+Ri) def	Di-Ri	Di+Ri	Rj	Di
تاثیرپذیر	-۰.۰٤	۱,٦٨٩٩	(-۰,۰۰۹۷،-۰,۰۲۰۳،-۰,۰۱۰۹٦)	(۰,۲۹۷۱،۰,۸۸۹۳،٤,٦٨٤)	(۰,۱۵۳٤،۰,٤۵٤۸،۲,۳۹٦۸)	(۰,۱٤۳۷،۰,٤۳٤۵،۲,۲۸۷۲)
تاثیرگذار	۰,۰۱۳٦	۱,٦٤۵۷	(۰,۰۰۰۷،۰,۰۰۷۱،۰,۰۳۹۳)	(۰,۲۷٦۹،۰,۸۵۳۵،٤,۵۹۸۹)	(۰,۱۳۸۱،۰,٤۲۳۲،۲,۲۷۹۸)	(۰,۱۳۸۸،۰,٤۳۰۳،۲,۳۱۹۱)
تاثیرگذار	۰,۰٤۱۰۱	۱,۳۵۳۵	(۰,۰۱۳۱،۰,۰۲۵۳،۰,۰۹٦۵)	(۰,۲۰۱۱،۰,٦٤۹۹،۳,۹۱۲۹)	(۰,۰۹٤۰,۳۱۲۳،۱,۹۰۸۲)	(۰,۱۰۷۱،۰,۳۳۷٦،۲,۰۰٤۷)
تاثیر پذیر	-۰,۰۱۳٦	۱,۷۳۰۹	(-۰,۰۰٤۱،-۰,۰۱۲۱،-۰,۰۲٦۲)	(۰,۳۰٦۷،۰,۹۱۵۷،٤,۷۸۵٦)	(۰,۱۵۵٤،۰,٤٦۳۹،۲,٤۰۵۹)	(۰,۱۵۱۳،۰,٤۵۱۸،۲,۳۷۹۷)

جدول ۶. محاسبه‌ی شدت و جهت تأثیر زیرعوامل.

وضعیت زیرعامل	(Di-Ri) def	(Di+Ri) def	Di-Ri	Di+Ri	Ri	Di
تاثیرگذار	۰,۰۵٤٦	۱,۹۰٦۵	(۰,۰۲۲۸،۰,۰٤۵،۰,۱۰۵٦)	(۰,۳۹۱،۱,۰٤۱۲،۵,۱۵۲٤)	(۰,۱۸٤۱،۰,٤۹۸۱،۲,۵۲۳٤)	(۰,۲۰٦۹،۰,۵٤۳۱،۲,٦۲۹)
تاثیر پذیر	-۰,۱۷۱۸	۱,۸٦۳۳	(-۰,۰۹٦۲،-۰,۱۵۲۳،-۰,۲۸٦۳)	(۰,۳۹۵٤،۱,۰۳۲۳،٤,۹۹۳۱)	(۰,۲٤۵۸،۰,۵۹۲۳،۲,٦۳۹۷)	(۰,۱٤۹٦،۰,٤٤٤،۲,۳۵۳٤)
تاثیرگذار	۰,۰۵۸٦	۱,٦۵٤	(۰,۰۲۷٤،۰,۰۵۳۹،۰,۰۹۹۱)	(۰,۳۲۳۲،۰,۸۷۳۹،٤,۵٤٤۹)	(۰,۱٤۷۹،۰,٤۱۰۲,۲۲۲۹)	(۰,۱۷۵۳،۰,٤٦۳۹،۲,۳۲۲)
تاثیرگذار	۰,۰۵۸٦	۱,۷۲۰۳	(۰,۰٤٦۰،۰,۰۵۳٤،۰,۰۸۱٦)	(۰,۲۹۱،۰,۸۹۵،٤,۸)	(۰,۱۲۲۵،۰,٤۲۰۰۸،۲,۳۵۹۲)	(۰,۱٦۸۵،۰,٤۷٤۲،۲,٤٤۰۸)
تاثیرگذار	۰,۰٤۱۱	۱,۷۷۳۸	(۰,۰۵٤۵،۰,۰٤۷٤،۰,۰۱٤۹)	(۰,۲٦۷۹،۰,۹۰۵،۰,۲۷۳)	(۰,۱۰٦۷،۰,٤۲٦۳،۲,۵۰٦۲)	(۰,۱٦۱۲،۰,٤۷۳۷،۲,۵۲۱۱)
تاثیرگذار	۰,۰۰۱۹	۱,٤۸۵۱	(-۰,۰۰۸۷،-۰,۰۰۲٦،۰,۰۲۱۳)	(۰,۲۱۹۷،۰,۷۳۰٦،٤,۲۵۹۵)	(۰,۱۱٤۲،۰,۳٦٦٦،۲,۱۱۹۱)	(۰,۱۰۵۵،۰,۳٦٤،۲,۱٤۰٤)
تاثیر پذیر	-۰,۰٤٦۲	۱,۵۳۸۹	(-۰,۰۲۲٤،-۰,۰۲۷٦،-۰,۱۰۷)	(۰,۲۳۱،۰,۷٦۹۲،٤,۳۸٦۲)	(۰,۱۲٦۷،۰,۳۹۸٤،۲,۲٤٦٦)	(۰,۱۰٤۳،۰,۳۷۰۸،۲,۱۳۹٦)
تاثیرگذار	۰,۰۰۳۳	۱,۷٦٦٦	(-۰,۰۲۳٤،-۰,۰۱۷۲،۰,۰۷۰۸)	(۰,۳۰۱٦،۰,۹۱۸۸،٤,۹۲۷۲)	(۰,۱٦۲۵،۰,٤٦۸۵،۲,٤۲۸۲)	(۰,۱۳۹۱،۰,٤۵۰۸،۲,٤۹۹)
تاثیرگذار	۰,۳۳	۰,٦۱٤۸	(۰,۰۲٤٤،۰,۰۲۸۸،۰,۰۵)	(۰,۱۱۱۸،۰,۲۹۱۸،۱,۷٦۳۸)	(۰,۰٤۳۷،۰,۱۳۱۵،۰,۸۵٦۹)	(۰,۰٦۸۱،۰,۱٦۰۳،۰,۹۰٦۹)
تاثیر پذیر	-۰,۰۳۳	۰,۵۸٦۱	(-۰,۰۲٤٤،-۰,۰۲۸۸،-۰,۰۵)	(۰,۱۱۰۲،۰,۲۷۸٦،۱,٦۷۷)	(۰,۰٦۷۳،۰,۱۵۳۷،۰,۸٦۳۵)	(۰,۰٤۲۹،۰,۱۲٤۹،۰,۸۱۳۵)
تاثیرگذار	۰,۰۱۰۲	۱,۸۵۷۹	(۰,۰۱٤٦،۰,۰۱۷۱،-۰,۰۰۷۹)	(۰,۳٤۰٦،۰,۹۸۳۳،۵,۱۲٤۵)	(۰,۱٦۳،۰,٤۸۳۱،۲,۵٦٦۲)	(۰,۱۷۷٦،۰,۵۰۰۲،۲,۵۵۸۳)
تاثیر پذیر	-۰,۰۲۲۱	۱,۸۲٤۸	(-۰,۰۱۷۳،-۰,۰۲۸۲،-۰,۰۱٤۸)	(۰,۳۳۳۳،۰,۹۵،۵,۰٦٦)	(۰,۱۷۵۳،۰,٤۸۹۱،۲,۵٤۰٤)	(۰,۱۵۸،۰,٤٦۰۹،۲,۵۲۵٦)
تاثیرگذار	۰,۰۱۳	۱,۷۸۰۵	(۰,۰۰۸٦،۰,۰۱٦۲،۰,۰۱۱۱)	(۰,۳۱۳۸،۰,۹۳۲٤،۹٤۳۹)	(۰,۱۵۲٦،۰,٤۵۸،۲,٦٦٤)	(۰,۱٦۱۲،۰,٤۷٤۲،۲,٤۷۴۵)
تاثیر پذیر	-۰,۰۰۱۱	۱,۸۵۲	(-۰,۰۰۵۹،-۰,۰۰۵۱،۰,۰۱۱٦)	(۰,۳۵٤۱،۱,۰۰۲۹،۵,۰٤۸)	(۰,۱۸۰،۰,۵۰٤،۲,۵۱۸۲)	(۰,۱۷٤۱،۰,٤۹۸۹،۲,۵۲۹۸)

جدول ۷. ماتریس ارتباط کامل عوامل اصلی (T_D) فازی‌زدایی شده.

	نمایشگری	متنی	دسترس‌پذیری	ذاتی
نمایشگری	۰/۲۲۳۳	۰/۲۰۹۷	۰/۱۶۳۸	۰/۲۲۸۳
متنی	۰/۲۲۴۷	۰/۲۰۵۲	۰/۱۷۱۷	۰/۲۲۸۱
دسترس‌پذیری	۰/۱۸۱۴	۰/۱۷۸	۰/۱۵۰۱	۰/۱۸۷۳
ذاتی	۰/۲۳۵۷	۰/۲۲۳۳	۰/۱۷۱۲	۰/۲۲۸۶

کمتری دارند تأثیرگذار است و برعکس عاملی که $D-R$ کمتری دارد از عاملی که $D-R$ بیشتری دارد، تأثیر می‌پذیرد. در نهایت روابط شبکه در نمودارها ترسیم شده است (شکل‌های ۳ تا ۷).

۴. رتبه‌بندی عوامل و زیرعوامل مؤثر بر کیفیت فراداده‌های سامانه‌ی ملی ثبت پایان‌نامه‌های داخل کشور

در نهایت رتبه‌بندی عوامل و زیرعوامل مؤثر برکیفیت فراداده‌های سامانه‌ی ملی ثبت پایان‌نامه‌های داخل کشور در شکل‌های ۸ و ۹ نشان داده شده است.

۵. نتایج آمار استنباطی و پاسخ به اهداف و سؤالات تحقیق

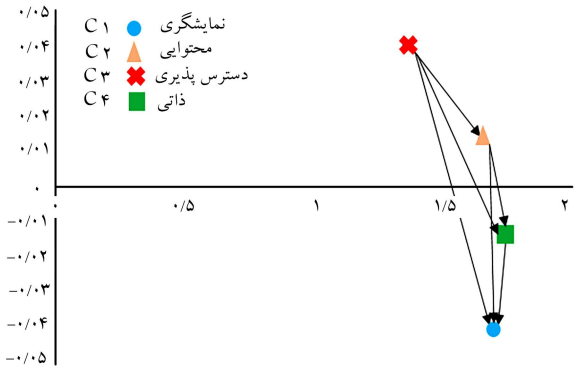
۱.۵. تعیین زیر عوامل مؤثر برکیفیت داده‌های سامانه‌ی ملی ثبت پایان نامه‌ها در داخل کشور

در این تحقیق از دو پرسش‌نامه استفاده شده است که نتیجه‌ی بررسی و تجزیه و تحلیل پرسش‌نامه‌ی شماره یک که توسط خبرگان پاسخ داده شده است، نشان می‌دهد که

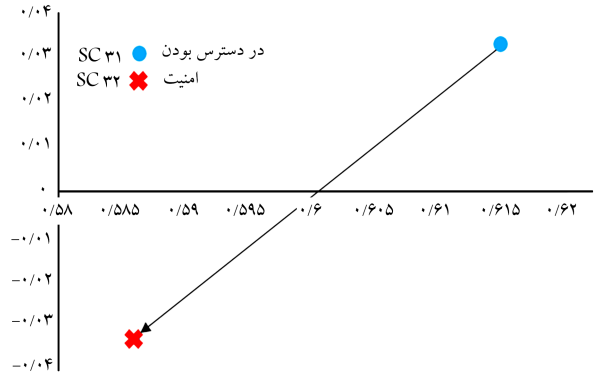
$$\lim_{Z\to\infty}(W^{\propto l})^Z,\ \lim_{Z\to\infty}(W^{\propto m})^Z,\ \lim_{Z\to\infty}(W^{\propto u})^Z \qquad (۱۴)$$

در ماتریس نهایی همگرا شده اعداد نشان‌دهنده‌ی وزن نسبی زیرعوامل تحقیق‌اند که بر اساس آن می‌توان عوامل را با توجه به میزان اهمیتشان اولویت‌بندی کرد.

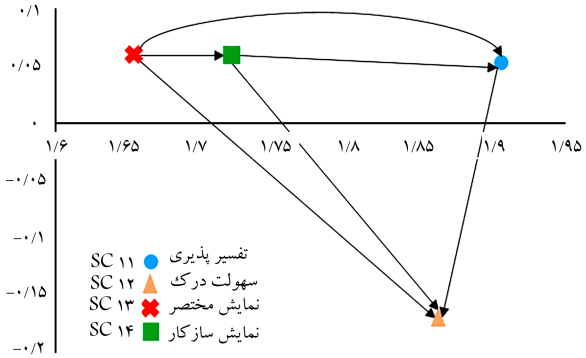
در ابتدا از پاسخ‌دهندگان خواسته شد تا میزان تأثیرگذاری زیرعامل i بر زیرعامل j را با استفاده (جدول ۴) نشان دهند. سپس از آنها میانگین حسابی گرفته شد و ماتریس ارتباط مستقیم ساخته شد. پس از پیمودن مراحل روش دیمتل فازی برای ترسیم روابط شبکه، ماتریس ارتباط کامل زیرعوامل و سپس به وسیله‌ی آن عوامل اصلی محاسبه شد (جداول ۷ و ۸). و در نهایت با فازی‌زدایی آن اعداد مثبت و منفی مشخص شد. به منظور تعیین نقشه‌ی روابط شبکه (NRM)، پس از به دست آوردن ماتریس های T_C و T_D باید $D-R$ و $d-r$ ماتریس‌های عوامل اصلی و زیرعوامل را محاسبه کنیم. عاملی که $D-R$ بیشتری دارد، روی عاملی که $D-R$



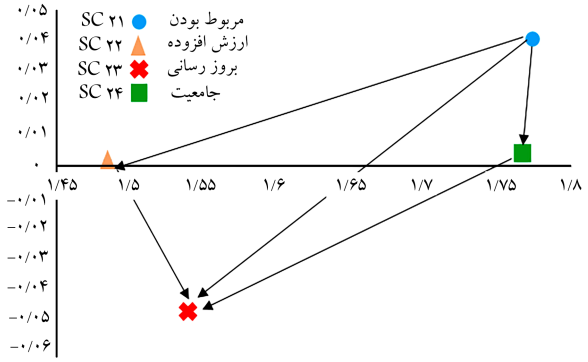
شکل ۳. روابط درون شبکه مربوط به عوامل اصلی.



شکل ۵. روابط درون شبکه مربوط به عامل دسترس‌پذیری.



شکل ۴. روابط درون شبکه مربوط به عامل نمایشگری.



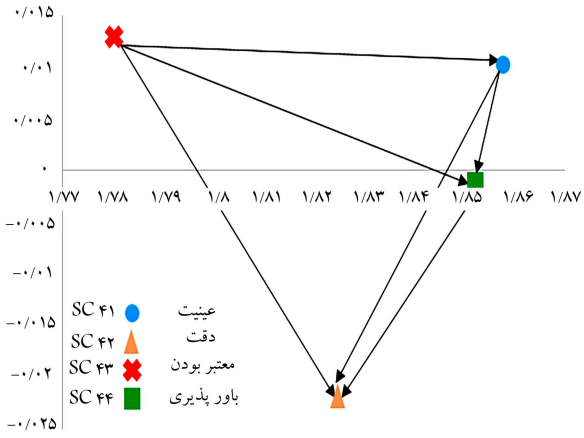
شکل ۶. روابط درون شبکه مربوط به عامل متنی.

جدول ۸. ماتریس ارتباط کامل زیرعوامل (T_c) فازی‌زدایی شده.

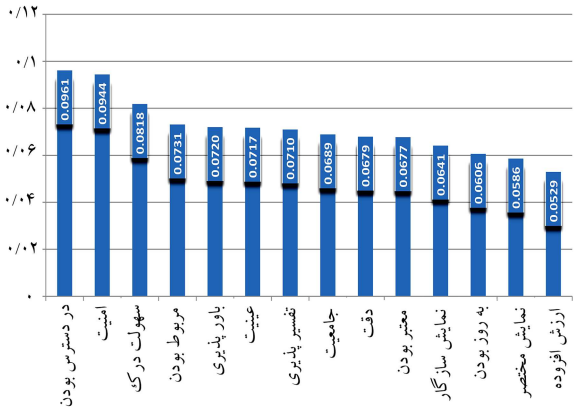
	$SC11$	$SC12$	$SC13$	$SC14$	$SC21$	$SC22$	$SC23$	$SC24$	$SC31$	$SC32$	$SC41$	$SC42$	$SC43$	$SC44$
$SC11$	۰٫۲۱	۲۹۷۷.۰	۰٫۲۳۴۲	۰٫۲۳۸۶	۰٫۲۶۲۷	۰٫۲۰۷۶	۰٫۲۱۳۶	۰٫۲۵۴۷	۰٫۱۸۲۴	۰٫۱۷۵۹	۰٫۲۷۰۱	۰٫۲۵۲۶	۰٫۲۲۶۸	۰٫۲۵۹۷
$SC12$	۰٫۲۴۲۷	۰٫۱۹۷۸	۰٫۲۰۲۹	۰٫۲۰۲۳	۰٫۲۳/۱۸	۰٫۱۷۰۴	۰٫۱۷۸۲	۰٫۲۹۷۴	۰٫۱۶۷۵	۰٫۱۴۶۴	۰٫۲۳۷۵	۰٫۲۱۲۸	۰٫۲۰۹۹	۰٫۲۳۷۵
$SC13$	۰٫۲۳۴۲	۰٫۲۵۷۴	۰٫۱۵۵۳	۰٫۲۰۹۴	۰٫۲۳۰۶	۰٫۱۶۳۱	۰٫۱۶۱۵	۰٫۲۱۶۳	۰٫۱۷۰۹	۰٫۱۵۰۵	۰٫۲۰۱۲	۰٫۲۰۷۱	۰٫۱۸۹۳	۰٫۲۱۴۸
$SC14$	۰٫۲۳۹	۰٫۲۶۴۷	۰٫۲۰۵۲	۰٫۱۸۰۵	۰٫۲۵۳۶	۰٫۱۸۳۱	۰٫۱۸۹۸	۰٫۲۳۰۶	۰٫۱۵۶۶	۰٫۱۵۹۶	۰٫۲۳۰۵	۰٫۲۳۲۷	۰٫۲۲۷۲	۰٫۲۴۲۳
$SC21$	۰٫۲۵۶۹	۰٫۲۷۴۷	۰٫۲۱۶۶	۰٫۲۴۷۷	۰٫۲۱۲	۰٫۲۰۹۱	۰٫۲۲۳۵	۰٫۲۶۲۸	۰٫۱۷۵۱	۰٫۱۸۰۴	۰٫۲۵۶	۰٫۲۵۸۶	۰٫۲۴۷۹	۰٫۲۶۰۹
$SC22$	۰٫۱۹۷۴	۰٫۲۱۰۶	۰٫۱۷۵۵	۰٫۱۹۲۲	۰٫۲۰۳۳	۰٫۱۴۰۶	۰٫۱۹۲۹	۰٫۲۰۶۶	۰٫۱۵۱۴	۰٫۱۴۷۱	۰٫۲۰۹۴	۰٫۲۱۰۴	۰٫۲۰۱	۰٫۲۰۲۷
$SC23$	۰٫۲۰۳	۰٫۲۲۶۷	۰٫۱۸۱۴	۰٫۱۸۴۸	۰٫۲۰۰۴	۰٫۱۸۳۴	۰٫۱۵۱۴	۰٫۲۱۱۱	۰٫۱۶۲۱	۰٫۱۶۸	۰٫۲۰۶۹	۰٫۱۹۷۷	۰٫۲۰۲۶	۰٫۱۹۸۸
$SC24$	۰٫۲۶۸۵	۰٫۲۹۱۶	۰٫۲۲۱۴	۰٫۲۴۶۱	۰٫۲۵۰۷	۰٫۲۰۸۶	۰٫۲۲۴۶	۰٫۲۰۱۱	۰٫۱۹۵۵	۰٫۱۸۹۶	۰٫۲۷۱۲	۰٫۲۳۳۵	۰٫۲۳۶۵	۰٫۲۵۵۸
$SC31$	۰٫۲۰۵۵	۰٫۲۲۰۷	۰٫۱۷۱۱	۰٫۱۸۹۹	۰٫۱۹۲۱	۰٫۱۶۶۴	۰٫۱۸۹۷	۰٫۱۹۹۲	۰٫۱۳	۰٫۱۹۳۹	۰٫۲۰۹۸	۰٫۱۹۶۵	۰٫۱۹۲۹	۰٫۱۹۴
$SC32$	۰٫۱۶۲۶	۰٫۱۷۶۱	۰٫۱۵۹	۰٫۱۶۵۶	۰٫۱۸۸	۰٫۱۴۳۹	۰٫۱۷۶۲	۰٫۱۶۸۵	۰٫۱۶۰۹	۰٫۱۱۵۶	۰٫۱۷۲۹	۰٫۱۶۹۶	۰٫۱۹۸۶	۰٫۱۶۴۱
$SC41$	۰٫۲۴۷۶	۰٫۲۶۸۷	۰٫۲۰۳۴	۰٫۲۲۶۳	۰٫۲۵۶۸	۰٫۱۸۷	۰٫۲۲۴۱	۰٫۲۳۱	۰٫۱۸	۰٫۱۷۰۱	۰٫۱۹۹۶	۰٫۲۴۸۸	۰٫۲۳۵۱	۰٫۲۵۰۶
$SC42$	۰٫۲۵۳۵	۰٫۲۷۲۴	۰٫۲۰۳۸	۰٫۲۳۷۶	۰٫۲۴۸۵	۰٫۲۰۱۷	۰٫۲۱۵۴	۰٫۲۲۹۹	۰٫۱۷۱۸	۰٫۱۷۷۳	۰٫۲۳۴۱	۰٫۱۹۳۱	۰٫۲۳۰۹	۰٫۲۴۳۲
$SC43$	۰٫۲۲۸۷	۰٫۲۵۰۱	۰٫۱۹۴۳	۰٫۲۲۰۱	۰٫۲۵۳۴	۰٫۱۹۶۵	۰٫۲۱۱۷	۰٫۲۲۷۷	۰٫۱۵۹۶	۰٫۱۸۴۳	۰٫۲۳۸۶	۰٫۲۳۸۲	۰٫۱۸۳۶	۰٫۲۳۶۴
$SC44$	۰٫۲۴۶۴	۰٫۲۷۱۶	۰٫۲۱۰۵	۰٫۲۳۵۵	۰٫۲۵۴۷	۰٫۱۹۸۹	۰٫۲۰۶۳	۰٫۲۲۸۵	۰٫۱۶۹۷	۰٫۱۵۶۲	۰٫۲۵۱۵	۰٫۲۴۳۴	۰٫۲۳۴۲	۰٫۱۹۶۳

جدول ۹. وزن و رتبه‌ی عوامل و زیرعوامل مؤثر بر کیفیت فراداده‌های سامانه‌ی ملی ثبت پایان‌نامه‌های داخل کشور.

عوامل اصلی	زیرعوامل	علائم اختصاری	وزن نسبی زیرعوامل و الویت آنها		وزن نهایی زیرعوامل و الویت آنها	
نمایشگری (۰/۲۷۵۵) (۲)	تقسیم‌پذیری	$SC11$	۰/۲۵۷۷۱۳	۲	۰/۰۷۱	۷
	درک سهولت	$SC12$	۰/۲۹۶۹۱۵	۱	۰/۰۸۱۸	۳
	نمایش مختصر	$SC13$	۰/۲۱۲۷۰۴	۴	۰/۰۵۸۶	۱۳
	نمایش سازگار	$SC14$	۰/۲۳۲۶۶۸	۳	۰/۰۶۴۱	۱۱
متنی (۰/۲۵۵) (۳)	مربوط بودن	$SC21$	۰/۲۸۶۱۰۶	۱	۰/۰۷۳۱	۴
	ارزش افزوده	$SC22$	۰/۲۰۷۰۴۵	۴	۰/۰۵۲۹	۱۴
	به‌روز بودن	$SC23$	۰/۲۳۷۱۸۲	۳	۰/۰۶۰۶	۱۲
	جامعیت	$SC24$	۰/۲۶۹۶۶۷	۲	۰/۰۶۸۹	۸
دسترس‌پذیری (۰/۱۹۰۵) (۴)	در دسترس بودن	$SC31$	۰/۵۰۴۴۶۲	۱	۰/۰۹۶۱	۱
	امنیت	$SC32$	۰/۴۹۵۵۳۸	۲	۰/۰۹۴۴	۲
ذاتی (۰/۲۷۹۳) (۱)	عینیت	$SC41$	۰/۲۵۶۷۱۳	۲	۰/۰۷۱۷	۶
	دقت	$SC42$	۰/۲۴۳۱۰۸	۳	۰/۰۶۷۹	۹
	معتبر بودن	$SC43$	۰/۲۴۲۳۹۲	۴	۰/۰۶۷۷	۱۰
	باورپذیری	$SC44$	۰/۲۵۷۷۸۷	۱	۰/۰۷۲	۵



شکل ۷. روابط درون شبکه مربوط به عامل ذاتی.



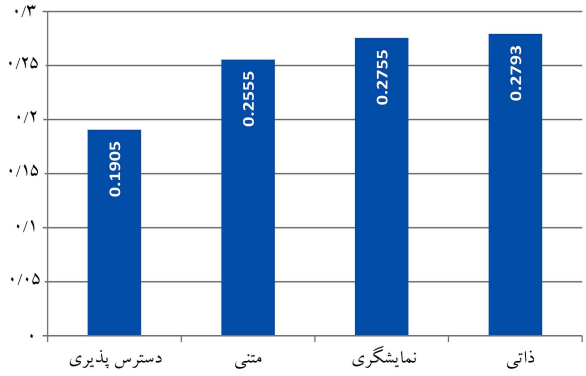
شکل ۸. رتبه‌بندی نهایی زیرعوامل مؤثر بر کیفیت فراداده‌های سامانه‌ی ملی ثبت پایان‌نامه‌های داخل کشور.

بیشتر زیرعوامل مؤثر بر کیفیت فراداده‌های سامانه‌ی ملی ثبت پایان‌نامه‌های داخل کشور، به غیر از زیرعامل مقدار داده‌ها، دارای تأثیر قابل توجه‌اند (جدول ۹).

۲.۵. نتایج انجام روش ترکیبی دیمتل فازی بر مبنای ANP فازی (FDANP)

با انجام روش دیمتل فازی برای تعیین ارتباطات درونی و بیرونی عوامل و زیرعوامل، تعیین سطوح مختلف مدل، مشخص کردن شدت تأثیرگذاری و تأثیرپذیری، رسم نمودار شدت ارتباطات، که با تجزیه و تحلیل پرسش‌نامه‌های مقایسات زوجی توزیع شده بین خبرگان به دست آمد، نتیجه‌ی انجام روش به شرح زیر است:

- با توجه به پاسخ خبرگان به مقایسات زوجی بین زیرعوامل نسبت به هم مشخص شد که از دیدگاه خبرگان عوامل متنی و دسترس‌پذیری با شدت تأثیر ۰/۱۳۶،



شکل ۹. رتبه‌بندی نهایی عوامل مؤثر بر کیفیت فراداده‌های سامانه‌ی ملی ثبت پایان‌نامه‌های داخل کشور.

۰٫۰۴۰۱ تأثیرگذار و عوامل نمایشگری و ذاتی به ترتیب هر یک با مقادیر ۰٫۰۴- و ۰٫۱۳۶- تأثیرپذیرند.

- در میان زیرعوامل نمایشگری، زیرعوامل تفسیرپذیری، نمایش مختصر و نمایش سازگار، تأثیرگذار روی سایر زیرعوامل عامل نمایشگری بوده‌اند و زیرعامل سهولت درک تأثیرپذیر است.
- در میان زیرعوامل عامل متنی، زیرعوامل مربوط بودن، ارزش افزوده و جامعیت زیرعوامل تأثیرگذار و زیرعامل به‌روز بودن تأثیرپذیر است.
- در میان زیرعوامل عامل دسترس‌پذیری، زیرعامل در دسترس بودن تأثیرگذار و زیرعامل امنیت تأثیرپذیر است.
- در میان زیرعوامل عامل ذاتی، زیرعوامل عینیت و معتبر بودن تأثیرگذار و زیرعوامل دقت و باورپذیری تأثیرپذیرند.
- با توجه به روش انتخاب شده در رتبه‌بندی عوامل و زیرعوامل مورد نظر در این تحقیق که از روش تحلیل شبکه انجام می‌گیرد، پس از رسم نمودار شدت ارتباطات بین عوامل و زیرعوامل، محاسبات لازم بر روی عوامل و زیرعوامل تحقیق برای دست‌یابی به هدف مورد نظر تحقیق انجام گرفته است که نتایج حاصل به شرح زیر است:
- نتایج تحلیل پرسش‌نامه‌های پاسخ داده شده توسط خبرگان نشان می‌دهد که عامل ذاتی با وزن ۰٫۲۷۹۳ وزین‌ترین (ارجح‌ترین) عامل و عامل دسترس‌پذیری با وزن ۰٫۱۹۰۵ کم‌وزن‌ترین عامل قلم‌داد شده‌اند.
- همچنین با تحلیل پرسش‌نامه‌های مقایسات زوجی زیر شاخص‌ها مشخص شده است که از نظر خبرگان، زیرعامل در دسترس بودن با وزن ۰٫۰۹۶۱ با اهمیت‌ترین یا وزین‌ترین زیرعامل و همین طور زیرعامل ارزش افزوده با وزن ۰٫۰۵۲۹ کم‌اهمیت‌ترین زیرعامل معرفی شده‌اند.

با توجه به نتیجه‌ی نهایی حاصل از تجزیه و تحلیل پرسش‌نامه‌ها رتبه‌بندی نهایی عوامل و زیرعوامل مؤثر بر کیفیت فراداده‌های سامانه‌ی ملی ثبت پایان‌نامه/رساله‌های داخل کشور به شرح زیر خواهد بود (جدول ۱۰).

جدول ۱۰. نتیجه‌ی نهایی رتبه‌بندی ابعاد کیفیت داده.

رتبه	شاخص
اول	در دسترس بودن
دوم	امنیت
سوم	سهولت درک
چهارم	مربوط بودن
پنجم	باورپذیری
ششم	عینیت
هفتم	تفسیرپذیری
هشتم	جامعیت
نهم	دقت
دهم	معتبر بودن
یازدهم	نمایش سازگار
دوازدهم	به روز بودن
سیزدهم	نمایش مختصر
چهاردهم	ارزش افزوده

۶. نتیجه‌گیری

در پژوهش حاضر به اولویت‌بندی و تعیین درجه‌ی اهمیت ابعاد و شاخص‌های کیفیت داده‌های پژوهشی در سامانه‌ی ملی ثبت پایان‌نامه/رساله‌های دانش‌آموختگان کل کشور پرداخته شد. با توجه به اثرگذاری معیارهای کیفیت داده بر روی یکدیگر که در برخی پژوهش‌های پیشین مانند خسروانجم و همکاران[۲] و باتینی و اسکاناپیکا[۲۸] مورد اشاره قرار گرفته است، به‌منظور ارزیابی نوع اثرگذاری و اثرپذیری و نیز میزان سطح وابستگی معیارها از روش ترکیبی دیمتل و ANP استفاده شده است. پس از بررسی نتایج پژوهش این‌گونه می‌توان بیان کرد که دردسترس بودن تأثیرگذارترین عامل بر کیفیت فراداده‌های سامانه‌ی ملی ثبت پایان‌نامه‌های داخل کشور به حساب می‌آید. امنیت، سهولت درک و مرتبط بودن در اولویت‌های بعدی قرار دارند. از این‌رو در فرایند کنترل و بهبود کیفیت داده‌های پایان‌نامه/رساله‌ها که در پژوهشگاه علوم و فنّاوری اطلاعات ایران توسط کارشناسان مختلف در جریان است، لازم است به این ویژگی‌ها در داده‌ها توجه بیشتری کرد. پایش‌های کیفی صورت گرفته در فرایند کنترل کیفیت داده باید بر پایه‌ی این معیارهای کیفی استوار شود و طرح‌های کیفیت تدوین گردد. در ادامه به برخی اقدامات کاربردی به منظور بهبود کیفیت داده‌های پژوهشی با تمرکز بر ابعاد بیان شده خواهیم پرداخت.

۷. پیشنهادها

پیشنهادها در دو بخش رهنمودهایی برای بهبود کیفیت داده‌های پژوهشی و نیز پیشنهادهایی برای پژوهش‌های آتی ارائه می‌شود.

در چارچوب بهبود کیفیت داده‌های پژوهشی بر پایه‌ی معیارهای دارای بالاترین اولویت راه‌کارهای زیر را می‌توان ارائه کرد.

۱. **در دسترس بودن:** طراحی لایه‌های امنیتی در سامانه‌ی ثبت (که به طور مستقیم با کاربران در ارتباط است) باید به گونه‌یی باشد که امکان دسترسی به اطلاعات شخصی دانشجویان تحصیلات تکمیلی برای افراد غیرمجاز مقدور نباشد.

۲. **امنیت:** داشتن رمز عبور و رایانامه که هر کاربر فقط برای دسترسی به محیط کاربری خود در سامانه‌ی ثبت پایان‌نامه استفاده می‌کند.

۳. **سهولت درک:** فراهم کردن زیرساخت لازم که تمام فیلدهای موجود در سامانه‌ی ثبت به راحتی برای کاربر قابل تشخیص و قابل فهم باشد.

۴. **مرتبط بودن:** تمام اطلاعاتی که باید در سامانه‌ی ثبت درج شود به صورت فیلدهایی در این سامانه تعبیه شده است. به منظور بهبود میزان مرتبط بودن داده‌های هر رکورد، کلیه‌ی نیازمندی‌ها در ویرایش‌های آتی این سامانه شناسایی شود تا در صورت نیاز فیلدها اضافه یا کم شوند. مثلاً اگر در آینده قرار است بانک اطلاعاتی پژوهشگران تقویت و تکمیل شود آدرس Google Scholar هر نویسنده یا استاد وی در فرایند ثبت مدرک اخذ شود (یک فیلد اضافه شود).

پژوهشگران در پژوهش‌های آتی می‌توانند رویکردهای زیر را به کار برند.

۱. استفاده از رویکردهای برنامه‌ریزی راهبردی به منظور تدوین راهبردهای کیفیت داده.

۲. توسعه‌ی روابط ریاضی به منظور ارزیابی دقیق هریک از ابعاد کیفیت داده.

۳. استفاده از انواع نمودارهای کنترل (به ویژه نمودارهای چندمتغیره) به منظور پایش آماری شاخص‌های کیفیت داده.

پانوشت‌ها

1. data quality(DQ)
2. subjective
3. objective
4. information system
5. cooperation information system
6. Web Information System
7. Peer to Peer
8. پایگاه اطلاعات پایان‌نامه‌ها و رساله‌های دانش‌آموختگان داخل کشورکه توسط پژوهشگاه علوم و فنّاوری اطلاعات ایران (ایرانداک) مدیریت می‌شود.
9. analytic hierarchy process

منابع (References)

1. Dhillon, G. and Backhouse, J. "Technical opinion: information system security management in the new millennium", *Communications of the ACM. Jul*, **1;43**(7) (2000).
2. Khosroanjom, D., Ahmadzade, M., Niknafs, A. and et al."Using fuzzy AHP for evaluating the dimensions of data quality", *International Journal of Business Information Systems*, **8**(3), pp. 269-285 (2011).
3. Vaziri, R., Mohsenzadeh, M. and Habibi J. "Measuring data quality with weighted metrics", *Total Quality Management & Business Excellence*, Jun **1**, 1-3 (2017).
4. Redman, TC. "The impact of poor data quality on the typical enterprise", *Communications of the ACM.* Feb **1;41**(2), pp. 79-82 (1998).
5. Wu,W. W. "Segmenting critical factors for successful knowledge management implementation using the fuzzy DEMATEL method", *Applied Soft Computin*, **12**, pp. 527-535 (2012).
6. Ershadi, MJ, Aiasi, R. and Kazemi, S. "Root cause analysis in quality problem solving of research information systems: a case study", *International Journal of Productivity and Quality Management*, **24**(2), pp. 284-99 (2018).
7. Moges, HT., Van Vlasselaer, V., Lemahieu, W. and et al. "Determining the use of data quality metadata (DQM) for decision making purposes and its impact on decision outcomes—An exploratory study", *Decision Support Systems*, Mar **31;83**, pp. 32-46 (2016)
8. Pipin, LL., Lee, YW. and Wang, RY. "Data quality assessment", *Communications of the ACM*, Apr **1;45**(4), pp. 211-8 (2002).
9. Wang , RY. and Strong, DM. "Beyond accuracy: What data quality means to data consumers?" *Journal of management information systems*, Mar **1;12**(4), pp. 5-33 (1996).
10. Kwon, O., Lee, N. and Shin, B. "Data quality management, data usage experience and acquisition intention of big data analytics", *International Journal of Information Management*, Jun **30;34**(3), pp. 387-94 (2014).
11. Zachman, J. "The zachman framework for enterprise architecture", *Virginia: Zachman Framework Associates* (2006).
12. De Michelis, G., Dubois, E., Jarke, M. and et al. "Cooperative information systems: a manifesto", *Cooperative Information Systems: Trends & Directions*, pp. 315-165 (1997).
13. Isakowitz, T., Bieber, M. and Vitali, F. "Web information systems", *Communications of the ACM*, Jul **1;41**(7), pp. 78-80 (1998).
14. Wang RY. "A product perspective on total data quality management", *Communications of the ACM*, Feb **1;41**(2), pp. 58-65 (1998).
15. Kunz, M., Puchta, A., Groll, S. and et al. "Attribute quality management for dynamic identity and access management", *Journal of information security and applications*, Feb **1; 44**, pp. 64-79 (2019).
16. Jeusfeld, MA., Quix, C. and Jarke, M. "Design and analysis of quality information for data warehouses", *In International Conference on Conceptual Modeling* Nov 16 (pp. 349-362) Spring, Beelin, Heidelberg (1998).
17. Di Tria, F., Lefons, E. and Tangorra, F. "A framework for evaluating design methodologies for big data warehouses: measurement of the design process", *International Journal of Data Warehousing and Mining (IJDWM)*, Jan **1; 14**(1), pp. 15-39 (2018).
18. English, LP. *Improving data warehouse and business information quality: methods for reducing costs and increasing profits*, New York: Wiley, (Apr 9 1999).
19. Lee, YW., Strong, DM. Kahn, BK. and et al. "AIMQ: a methodology for information quality assessment", *Information & management*, Dec **31;40**(2), pp. 133-46 (2002).
20. Leep, K., Morris, K. and Kasman, N. "CIHI survey: dying of cancer in canada's acute care facilities", *Healthcare Quarterly*, May **15;8**(3) (2005).
21. Pipino, LL., Lee, YW. and Wang, RY. "Data quality assessment", *Communications of the ACM*, Apr **1;45**(4), pp. 211-8 (2002).
22. Bacal, V, Choudhry, AJ., Baier, K. and et al. "A validation of hysterectomy procedural codes in the Canadian institutes for health information discharge abstract database", *Journal of Obstetrics and Gynecology Canada*, May **1;41**(5), pp. 708-9 (2019).
23. Eppler, MJ. and Muenzenmayer, P. "Measuring information quality in the web context: a survey of state-of-the-art instruments and an application methodology", *InIQ*, Nov , pp. 187-196 (2002).
24. Falorsi, PD., Pallara, S., Pavone, A. and et al. "Improving the quality of toponymic data in the italian public administration", *InProceedings of the ICDT*, **3** (2003).
25. Su, Z. and Jin, Z. "A methodology for information quality assessment in the designing and manufacturing processes of mechanical products", *InInformation Quality Management: Theory and Applications*, pp. 190-220 IGI Global (2007).
26. Scannapieco,M., Virgillito, A., Marchetti, C. and et al. "The DaQuinCIS architecture: a platform for exchanging and improving data quality in cooperative information systems", *Information systems*, Oct **1;29**(7), pp. 551-82 (2004).

27. Amicis, FD. "A methodology for data quality assessment on financial data", *Studies in Communication Sciences*, **4**(2), pp. 115-37 (2004).

28. Batini, C. and Scannapieca, M. "Data quality dimensions", *Data Quality: Concepts, Methodologies and Techniques*, pp. 19-49 (2006).

29. Ershadi, MJ. and Aiassi, R. "A Model for Quality Assurance on Acquisition and Registration", *Processing, and Dissemination of Theses and Dissertations Systems.*

30. Wang, RY. and Strong, DM. "Beyond accuracy: What data quality means to data consumers?", *Journal of management information systems*, Mar **1;12**(4), pp. 5-33 (1996).

31. Batini, C., Cappiello, C., Francalanci, C. and et al. "Methodologies for data quality assessment and improvement", *ACM computing surveys (CSUR)*, Jul **1;41**(3), pp. 16 (2009).

32. Shavandi, H."Fuzzy sets theory and application in industrial engineering and management", *Tehran: The expansion of basic science*, **97** (In Persian)(2016).

33. Mavi, RK. and Standing, C. "Critical success factors of sustainable project management in construction: A fuzzy DEMATEL-ANP approach", *Journal of Cleaner Production*, Sep bf 1;194, pp. 751-65 (2018).