Patent Analysis by Data Mining for Identifying and Determining Relationships among Technologies

Leila MahmudJanlu
Master Student of Innovation and Technologies Management; Mazandaran University of Science and Technology;
leila_janlu@yahoo.com

Babak Shirazi
Assistant Professor; Mazandaran University of Science and Technology;
Corresponding Author shirazi_b@yahoo.com

Iraj Mahdavi
Professor; Mazandaran University of Science and Technology;
irajarash@rediffmail.com

Javad Soltanzadeh
PhD Candidate of technology management; Faculty of Management and Accounting; Allameh Tabataba'i University;
soltanzadeh921@atu.ac.ir

Iranian Journal of Information Processing and Management

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 33 | No. 4 | pp. 1575-1610
Summer 2018

Received: 27, Feb. 2017 **Accepted: 17, Sep. 2017**

Abstract: Analyzing technologies relationships can provide insight into technological strategies and maximizing the profits. Patents are integral parts of intellectual property rights and include significant information about developed technologies. Due to the increasing amount of patents, data mining method is proposed for patent analysis. Thus, weighted association rules have been used for patent analysis; meanwhile unequal importance of patents and levels of technology have been considered in terms of technological impact and commercial importance. The aim of this study is to find a sustainable method for extracting technologies relationships based on patent analysis being independent of subjective knowledge of experts in a particular field of technology. To achieve this goal, a four-step process including data collection, structured database creation, data mining implementation, and technologies relationship identification/determination and their graphical representation are proposed. The proposed process has been implemented and evaluated within a case study.

Keywords: Patent Analysis, Data Mining, Weighted Association Rules, Technology Clusters, IPC Code, Related Technologies

تحلیل پتنت با استفاده از داده‌کاوی برای شناسایی و تعیین ارتباطات میان فناوری‌ها

لیلا محمودجانلو
کارشناسی ارشد مهندسی صنایع؛ گرایش مدیریت نوآوری و فناوری؛ دانشگاه علوم و فنون مازندران؛
leila_janlu@yahoo.com

بابک شیرازی
دکتری مهندسی صنایع؛ استادیار؛
دانشگاه علوم و فنون مازندران؛
پدیدآور رابط shirazi_b@yahoo.com

ایرج مهدوی
دکتری مهندسی صنایع؛ استاد؛ دانشگاه علوم و فنون مازندران irajarash@rediffmail.com

جواد سلطان‌زاده
دانشجوی دکتری مدیریت فناوری؛
دانشکدهٔ مدیریت و حسابداری؛ دانشگاه علامه طباطبائی soltanzadeh921@atu.ac.ir

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۳۹۵/۱۲/۱۹ **پذیرش:** ۱۳۹۶/۰۶/۲۶ مقاله برای اصلاح به مدت ۲۶ روز نزد پدیدآوران بوده است.

چکیده: تجزیه و تحلیل ارتباطات میان فناوری‌ها می‌تواند بینشی برای اتخاذ استراتژی‌های فناورانه و به حداکثررساندن سود فراهم کند. پتنت‌ها، بخشی جدایی‌ناپذیر از حقوق مالکیت فکری هستند و اطلاعات قابل توجهی دربارهٔ فناوری‌های توسعه‌یافته را دربرمی‌گیرند. با توجه به حجم فزایندهٔ پتنت‌ها، روش داده‌کاوی برای تحلیل پتنت پیشنهاد شده است. در روش پیشنهادی، از قوانین وابستگی وزن‌دار برای تحلیل پتنت استفاده می‌شود و اهمیت نابرابر پتنت‌ها و سطوح فناوری از نظر اهمیت تجاری و تأثیر فناورانه در نظر گرفته شده است. هدف این پژوهش، یافتن روشی پایدار برای شناسایی ارتباطات میان فناوری‌ها بر اساس تحلیل پتنت است، که وابسته به دانش ذهنی متخصصان یک حوزهٔ خاص فناوری نباشد. برای دستیابی به این هدف یک فرایند چهار مرحله‌ای شامل جمع‌آوری داده، تشکیل پایگاه دادهٔ ساخت‌یافته، اجرای داده‌کاوی و شناسایی/ تعیین میزان ارتباطات میان فناوری‌ها و ترسیم گرافیکی آن ارائه شده است. فرایند پیشنهادی با یک مطالعهٔ موردی اجرا و ارزیابی شده است.

کلیدواژه‌ها: تحلیل پتنت، داده‌کاوی، قوانین وابستگی وزن‌دار، خوشه‌های فناوری، کد IPC، فناوری‌های مرتبط

فصلنامه | علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۳۳ | شمارهٔ ٤ | ص‌ص ۱۵۷۵-۱۶۱۰
تابستان ۱۳۹۷

۱. مقدمه

پتنت‌ها یکی از مهم‌ترین منابع نوآوری و همچنین، ابزاری مفید برای تحلیل‌های فنی-اقتصادی به‌شمار آمده و در مطالعات گسترده‌ای مورد استفاده قرار گرفته‌اند؛ از جمله: ترسیم منحنی عمر فناوری‌ها، پیش‌بینی فناوری‌های نوظهور، مطالعات توسعهٔ فناورانه، تجزیه و تحلیل روندهای نوآوری، ارتقای برنامه‌ریزی استراتژیک (ناظمی، شماعی و قدیری ۱۳۹۲، Altuntas and Dereli 2015; Li, Zheng and Shi 2014; Altuntas, Turkay and Kusiak 2015a; Kim and Lee 2015). پتنت‌ها، یکی از بهترین شاخص‌ها برای ارزیابی فعالیت‌های فناورانه هستند (Mattas and Deepti 2015). بنابراین، با تحلیل پتنت می‌توان به اطلاعات ارزشمندی پیرامون روند و مسیر تکامل آن‌ها، فعالان اصلی و دانش‌های نهفتهٔ مرتبط و همچنین، ارتباطات میان فناوری‌ها دست یافت. محققان پس از درک اهمیت پتنت، برای فهم بیشتر آن تلاش‌هایی کردند، به‌طوری که پس از «توافقنامهٔ استراسبورگ» در سال ۱۹۷۱ و به‌منظور طبقه‌بندی پتنت‌ها، «سیستم بین‌المللی طبقه‌بندی پتنت»[1] توسعه یافت. با در دسترس بودن این طبقه‌بندی، جست‌وجو و استخراج اطلاعات مرتبط با پتنت‌ها آسان شده است. بر اساس این سیستم، به هر پتنت با توجه به دامنهٔ فناوری آن حداقل یک کد IPC اختصاص داده شده است (علم‌خواه ۱۳۹۱). هرچند این طبقه‌بندی توانست کاربرد پتنت‌ها و تحلیل برآمده از آن‌ها را افزایش دهد، اما با توجه به حجم فزایندهٔ پتنت‌ها، یک روش تحلیل مناسب برای مجموعهٔ داده‌های بزرگ ضروری به نظر می‌رسد. از این رو، استفاده از روش‌های داده‌کاوی در تحلیل پتنت‌ها پیشنهاد شده است. این روش، مجموعه‌ای از فنونی است که به شخص امکان می‌دهد فراتر از داده‌پردازی معمولی حرکت کند. داده‌کاوی، سازمان‌ها را قادر می‌سازد که از سرمایهٔ داده‌های خود به‌درستی بهره‌برداری نمایند و از این ابزار برای پشتیبانی فرایند تصمیم‌گیری استفاده کنند (غیاثی، نظافتی و شکوهیار ۱۳۹۴). یکی از پرکاربردترین تکنیک‌ها در داده‌کاوی، کاوش قوانین وابستگی است. کاوش قوانین وابستگی یک روش محبوب و متداول داده‌کاوی بر اساس احتمال شرطی است که ارتباطات میان اهداف مختلف را استخراج می‌کند (Hipp, Güntzer and Nakhaeizadeh 2000; Kotsiantis and Kanellopoulos 2006; Forsati and Meybodi 2010).

1. International Patent Classification (IPC)

در روش پیشنهادی با استفاده از اطلاعات پتنت وابستگی میان فناوری‌ها شناسایی شده، میزان ارتباط در خوشه‌های فناوری تعیین و در نهایت، به‌صورت گرافیکی نمایش داده می‌شود. در این پژوهش خوشه‌های فناوری نشان‌دهندهٔ گروه‌های دو یا چندتایی فناوری‌های مرتبط هستند که توسط الگوریتم‌های پیشنهادی به‌دست آمده‌اند.

تمامی پتنت‌ها دارای اهمیت یکسان نیستند و سطوح فناوری (در اینجا کدهای IPC) نیز دارای وزن‌های متفاوت هستند. بنابراین، الگوریتم‌های مبتنی بر قوانین وابستگی وزن‌دار شامل کاوش قوانین وابستگی وزن‌دار بولین (BOWARM)[1]، کاوش قوانین وابستگی وزن‌دار (WARM_2)[2] و سه الگوریتم پیشنهادی کاوش قوانین وابستگی وزن‌دار ترکیبی ۱، ۲ و ۳ (CBW_1, CBW_2, CBW_3)[3]، اجرا و نتایج به‌دست آمده مقایسه شده‌اند. استناد، یکی از شاخص‌های مهم در تعیین اهمیت و تأثیر یک پتنت به حساب می‌آید (منصوری و عصاره ۱۳۹۳). همچنین، تعداد زیاد پتنت‌های تحت پوشش یک کد IPC، نشان‌دهندهٔ جذابیت بالای فناوری ارائه‌شده توسط این کد است (Altuntas, Dereli, and Kusiak 2015b). در الگوریتم‌های پیشنهادی، وزن و اهمیت متفاوت پتنت‌ها و همچنین، سطوح فناوری به‌صورت همزمان و با استفاده از تعداد استنادهای پتنت و تعداد پتنت‌های موجود در کدهای IPC در نظر گرفته شده است. در این پژوهش برای آماده‌سازی داده‌ها به جای بهره‌گیری از نرم‌افزارهای تحلیل پتنت که خریداری و استفاده از آن‌ها اغلب مستلزم هزینهٔ بسیار است، یک نرم‌افزار به‌منظور استخراج اطلاعات طراحی شده و هزینه‌های این قسمت با وجود خروجی‌های مشابه به‌طور قابل توجهی کاهش یافته است. خروجی الگوریتم‌های مورد استفاده به پرسش‌هایی از قبیل موارد زیر پاسخ می‌دهد: فناوری مورد نظر با کدام فناوری‌ها مرتبط است؟ کدام فناوری‌ها دارای ارتباطات قوی‌تری هستند؟ و میزان ارتباطات آن‌ها نسبت به هم چگونه است؟

به بیان دقیق‌تر، هدف این پژوهش، یافتن روشی پایدار جهت استخراج فناوری‌های مرتبط بر اساس تحلیل پتنت است که راهنمایی برای مدیران به‌منظور شناسایی حوزه‌های فناوری، ارتباط میان فناوری‌ها و کاهش خطرات ناشی از سرمایه‌گذاری باشد. برای دستیابی به این هدف، روش پژوهش در قالب یک فرایند چهار مرحله‌ای ارائه شده است.

1. BOolean Weighted Association Rule Mining (BOWARM) 2. Weighted Association Rules Mining_2 (WARM_2)
3. Combination of BOWARM & WARM_2 (CBW)

مرحلهٔ نخست، جمع‌آوری داده، شامل جمع‌آوری اسناد پتنت از پایگاه دادهٔ پتنت با توجه به فناوری‌های مورد نظر. در مرحلهٔ دوم، با تشکیل پایگاه دادهٔ ساخت‌یافته، اطلاعات مورد نیاز از پتنت‌های بازیابی‌شده به‌دست می‌آیند. برای پیاده‌سازی فرایند ذکرشده، یک سیستم نرم‌افزاری با استفاده از زبان برنامه‌نویسی PHP طراحی و اجرا شده است، زیرا استفاده از نرم‌افزارهای رایج تحلیل پتنت مستلزم هزینهٔ نسبتاً بالایی است. مرحلهٔ سوم، اجرای داده‌کاوی است. در واقع، برای شناسایی و تعیین میزان ارتباطات میان فناوری‌های مختلف، الگوریتم‌های قوانین وابستگی وزن‌دار پیشنهاد شده‌اند و سه الگوریتم ابتکاری نیز ارائه شده است. در الگوریتم‌های پیشنهادی اهمیت نابرابر پتنت‌ها و نیز سطوح فناوری از نظر تأثیر فناورانه و اهمیت تجاری به‌صورت همزمان در نظر گرفته شده است. مرحلهٔ چهارم، ترسیم گرافیکی که در آن ارتباطات به‌دست آمده به‌صورت گرافیکی نمایش داده می‌شوند تا تجسم بهتری برای کاربران فراهم شده و تحلیل و ارزیابی نتایج به‌دست آمده تسهیل شود.

انجام این پژوهش برای مدیران و تصمیم‌گیران عرصهٔ فناوری، به‌ویژه فناوری‌های نوین مفید خواهد بود، زیرا مستندات تحلیل پتنت اطلاعات ارزشمندی ارائه می‌دهند؛ از جمله: اجتناب از دوباره‌کاری (کاهش هزینه‌های پژوهشی)، انجام پژوهش در یک سطح دانش بالاتر، استفاده از راه جدید، آگاهی از روندهای فنی و تجاری فناوری در دیگر کشورها و تعیین زمینه‌های مطمئن برای سرمایه‌گذاری، شناسایی اختراع‌های آزمایش‌شده به‌عنوان رویدادهای جدید، بررسی پژوهش‌های گذشته، پیداکردن راه حل برای مشکلات که این یافته‌ها در کنار هم افق دید بازتری را برای مدیران و کارکنان تحقیق و توسعه فراهم می‌آورد. ادامهٔ مقاله به شرح زیر است. مبانی نظری و مطالعات پیشینهٔ ارتباطات میان فناوری‌ها و کاربردهای داده‌کاوی در تحلیل پتنت در بخش دوم بررسی می‌شود. مراحل پژوهش در بخش سوم و نتایج پژوهش در بخش چهارم شرح داده شده است. بحث و ارزیابی نتایج نیز در بخش پنجم خلاصه شده است. نتیجه‌گیری، پیشنهادات آتی و محدودیت‌های پژوهش در بخش ششم ارائه شده‌اند.

مبانی نظری و مطالعات پیشینهٔ پژوهش

تحلیل پتنت با استفاده از داده‌کاوی

محققان، رویکردهای داده‌کاوی از قبیل قوانین وابستگی، کاوش متنی، خوشه‌بندی، طبقه‌بندی و تحلیل سری‌های زمانی را برای پتنت‌ها به‌کار گرفته‌اند (Wu et al. 2011; Lee et al. 2009; Trappey et al. 2011; Dereli and Durmusoglu 2009; Hidalgo and Gabaly 2012). مطالعات محدودی از قوانین وابستگی برای تحلیل اسناد پتنت استفاده کرده‌اند. «هاشلر، بتینا و کورت» قوانین وابستگی را به‌منظور تحلیل پتنت به کار گرفتند. آن‌ها مجموعه‌ای شامل قوانین وابستگی ارائه دادند که زیرساخت‌های اساسی را برای ایجاد و تغییر مجموعهٔ داده‌های ورودی و تجزیه و تحلیل مجموعهٔ داده‌ها و قوانین فراهم می‌کند (Hahsler, Bettina and Kurt 2005). «مجیدفر، مجیدفر و شادپور» با استفاده از تجزیه و تحلیل مورفولوژیک پتنت‌ها، فرایندی چهار مرحله‌ای به‌صورت مدول‌های نرم‌افزاری مستقل داده‌کاوی به‌منظور استخراج فناوری‌های جدید و بالقوه ارائه دادند و به این نتیجه رسیدند که روش پیشنهادی، پتانسیل پردازش سیستماتیک داده‌های موجود در متون ساخت‌نیافتهٔ پتنت‌ها و استفاده از این داده‌ها در پیش‌بینی حوزه‌های جدید فناوری را داراست (Majidfar, Majidfar and Shadpoor 2007). «جون» با استفاده از نقشه‌ها و قوانین وابستگی روشی برای تحلیل پتنت و کدهای IPC پیشنهاد کرد. با بهره‌گیری از چارچوب پیشنهادی می‌توان نواحی فناوری خالی (ضروری ولی توسعه‌نیافته) را پیش‌بینی کرد (Jun 2011a). «کیم» و همکاران کاوش قوانین وابستگی، فرایند تحلیل شبکه‌ای و تحلیل اثر متقابل را برای تعیین تعاملات میان فناوری‌ها به کار گرفتند. احتمال‌های شرطی میان فناوری‌ها با استفاده از کاوش قوانین وابستگی تعیین می‌شود. سپس، رویکرد تحلیل شبکه‌ای فناوری‌ها را اولویت‌بندی می‌کند (Kim et al. 2011). «جون، پارک و جانگ» از کاوش قوانین وابستگی و سری‌های زمانی برای پیش‌بینی روند بیوفناوری[1] استفاده کردند (Jun, Park and Jang 2012). «وانگ، یانگ و فیلیپ» یک متدولوژی کاوش کارا برای قوانین وابستگی وزن‌دار پیشنهاد کردند. این نظریه مبتنی بر این واقعیت است که مشخصه‌های عددی می‌توانند با توجه به قضاوت خبرگان در یک حوزهٔ وزنی خاص به هر آیتمی نسبت داده شوند (Wang, Yang and Philip 2000).

1. biotechnology

«آلتونتاس، درلی و کوسیاک» در سال ۲۰۱۵ داده‌های پتنت را توسط نرم‌افزار AcclaimIP استخراج کرده و از الگوریتم‌های قوانین وابستگی وزن‌دار (BWARM, WARM_2, BOWARM) برای تحلیل مدارک پتنت استفاده کردند. در الگوریتم‌های BWARM و BOWARM اهمیت (وزن) نابرابر سطوح تکنولوژی و در الگوریتم WARM_2 اهمیت (وزن) نابرابر پتنت‌ها در نظر گرفته شده است (Altuntas, Dereli and Kusiak 2015b).

مطالعات موردی مذکور اغلب اهمیت یکسانی برای پتنت‌ها و فناوری‌ها در نظر گرفته‌اند و یا وزن اختصاص داده‌شده بر اساس قضاوت خبرگان بوده است. در این پژوهش، الگوریتم قوانین وابستگی وزن‌داری که اهمیت نابرابر پتنت‌ها و همچنین، سطوح فناوری را به‌صورت توأم در نظر گرفته و وابسته به دانش ذهنی متخصصان یک حوزهٔ خاص فناوری نباشد، برای یافتن ارتباطات میان فناوری‌ها و شناسایی خوشه‌های شامل فناوری‌های مرتبط به‌کار گرفته شده است.

ارتباطات میان فناوری‌ها

مرور مبانی نظری نشان می‌دهد که برای بررسی ارتباطات میان فناوری‌های مختلف معمولاً از تکنیک تحلیل اثر متقابل[1] استفاده شده است. تحلیل اثر متقابل با استفاده از یک ماتریس، روابط علّی و معلولی بین وقایع را نشان می‌دهد و شاخص اثر متقابل (A, B) به‌صورت احتمال شرطی $P(B|A)$ تعریف می‌شود. جزئیات تکنیک تحلیل اثر متقابل را می‌توان در پژوهش (Dalkey 1971) مطالعه کرد. «جئونگ و کیم» در سال ۱۹۹۷ یک مدل اثر متقابل بر اساس منطق فازی برای تجزیه و تحلیل کیفی فناوری ارائه کردند (Jeong and Kim 1997). «یون و پارک» در سال ۲۰۰۴ با استفاده از یکپارچه‌سازی متن‌کاوی و تحلیل شبکه به توضیح فرایند کلی توسعهٔ شبکهٔ پتنت و نیز ارائهٔ شاخص‌های جدید مانند مرکزیت فناوری، دورهٔ فناوری و خوشه‌های کلیدواژهٔ فناوری جهت تحلیل کمی عمیق پرداخته و به این نتیجه رسیدند که روش پیشنهادی می‌تواند در تحلیل به‌روزرسانی روند فناوری‌های پیشرفته و شناسایی راه‌های جدید برای توسعهٔ محصول به‌کار گرفته شود (Yoon and Park 2004).

«چوی، کیم و پارک» در سال ۲۰۰۷، به جای تکیه بر درک‌های بصری و کیفی که

1. Cross Impact Analysis (CIA)

در مدل‌های متعارف تحلیل اثر متقابل استفاده می‌شود، یک متدولوژی مبتنی بر پتنت برای اندازه‌گیری اثرات متقابل به‌کار گرفتند. آن‌ها داده‌ها را از «ادارهٔ پتنت و علائم تجاری ایالات متحده»[1] با استفاده از یک کد برنامه‌نویسی‌شده جمع‌آوری و یک شبکهٔ اثر متقابل برای تحلیل روابط پیچیده بین فناوری‌ها ارائه کردند (Choi, Kim and Park 2007). «لی» و همکاران در سال ۲۰۰۹ رویکرد فرایند تحلیل شبکه‌ای[2] را برای یافتن اثر متقابل بین فناوری‌ها پیشنهاد کردند (Lee et al. 2009). «تورلیوچتر، پل و پرینزی» در سال ۲۰۱۰ در پژوهشی با استفاده از رویکرد تحلیل اثر متقابل کمّی، روشی برای شفاف‌ترکردن اثر فناوری‌ها و تخمین اثر متقابل فناوری‌ها پیشنهاد کردند (Thorleuchter, Poel and Prinzie 2007). «نو و پارک» در سال ۲۰۱۰ از اعداد استناد جهت توسعهٔ مسیرهای ادغام فناوری برای نانوبیوفناوری[3] استفاده کردند (No and Park 2006). «کیم، کیم و کیم» در سال ۲۰۱۱ با استفاده از تحلیل استنادی پتنت، تحلیل هم‌طبقه‌ای پتنت[4] و تحلیل استنادی سری‌های زمانی به بررسی ارتباطات فناورانه میان خدمات مبتنی بر فناوری و فناوری‌های اطلاعات و ارتباطات پرداختند (Kim, Kim and Kim 2011).

ضرورت و اهداف پژوهش

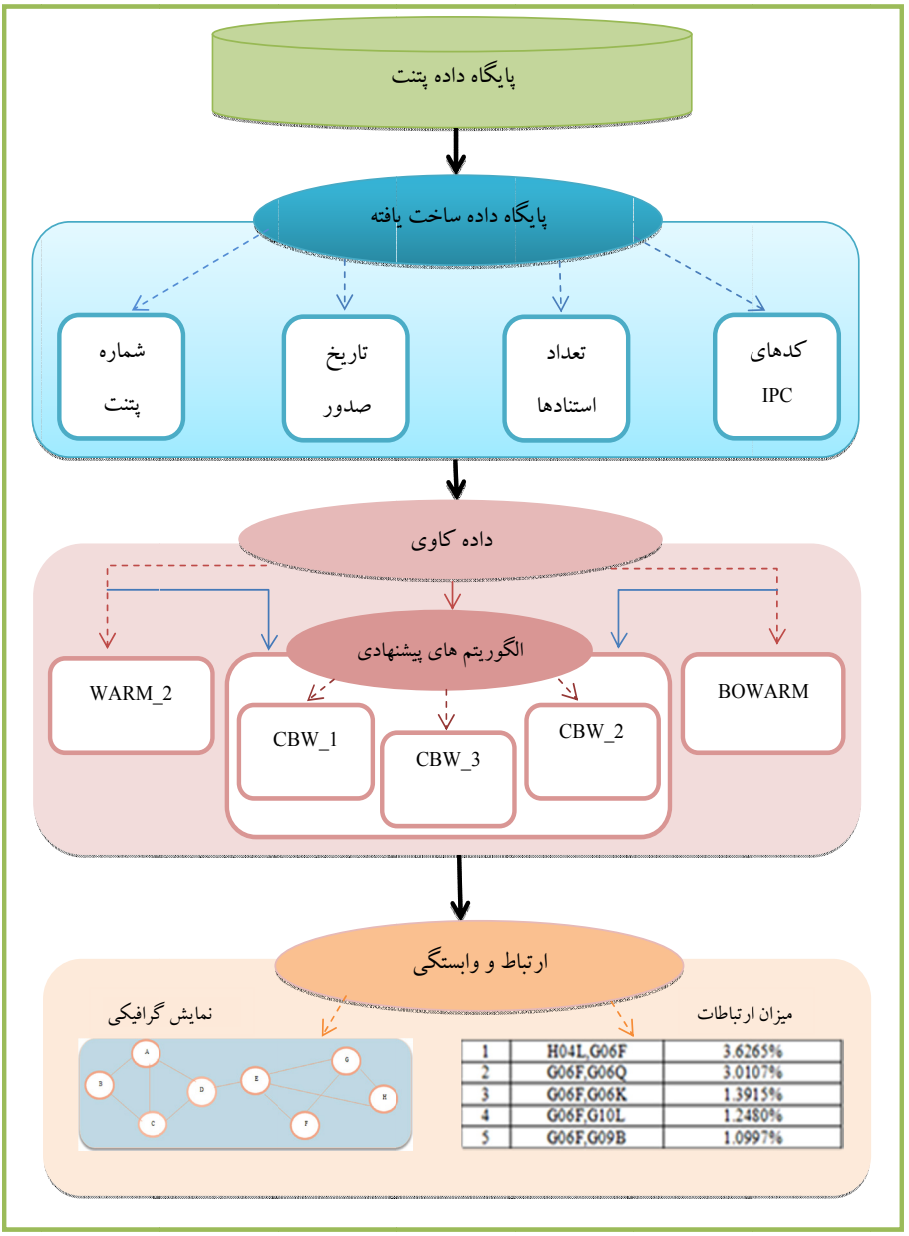
اطلاعات به‌دست آمده از تحلیل پتنت در مطالعات مرتبط با توسعهٔ فناورانه، شناسایی روندهای فناورانه، شناسایی پتانسیل بالقوهٔ فناوری‌ها و تصمیم‌گیری‌های مدیریتی به‌منظور شناسایی حوزه‌های فناورانهٔ جذاب برای سرمایه‌گذاری به‌کار می‌رود. از آنجا که حجم این داده‌ها بسیار زیاد بوده و نیز در حال افزایش است، استفاده از تکنیک‌های داده‌کاوی در تحلیل اطلاعات پتنت پیشنهاد شده است. روش داده‌کاوی پیشنهادی چارچوبی برای یافتن ارتباطات میان فناوری‌های مختلف ارائه می‌دهد و می‌تواند توسط مدیران و کارکنان سازمان‌ها به‌منظور کاهش ریسک مورد استفاده قرار گیرد. از پرکاربردترین تکنیک‌ها در داده‌کاوی، الگوریتم قوانین وابستگی است. چون قوانین وابستگی سنتی قادر به در نظر گرفتن اهمیت نابرابر پتنت‌ها و سطوح مختلف فناوری نیستند، این پژوهش ضمن بهره‌گیری از الگوریتم‌های قوانین وابستگی وزن‌دار، الگوریتم‌های جدیدی توسعه‌یافته از الگوریتم های وزن‌دار رایج، ارائه می‌دهد.

1. United States Patent and Trademark Office (USPTO)
2. Analytical Network Process (ANP)
3. nano biotechnology
4. co-classification

در نتیجه، با استفاده از الگوریتم‌های پیشنهادی، سازمان‌ها می‌توانند شرکت‌های تولیدکنندهٔ فناوری را مورد بررسی قرار داده و دریابند که این شرکت‌ها در حوزه‌های فناوری مورد نظر به‌دنبال چه هستند. درک و تحلیل ارتباطات کشف‌شده با توجه به قابلیت تصویرسازی، تسهیل شده است. در نتیجه، مدیران با شناخت دقیق‌تری از فناوری‌ها می‌توانند برای سرمایه‌گذاری‌های جدید تصمیم‌گیری نمایند.

۲. مراحل پژوهش

در این مقاله هر کد IPC به‌عنوان یک آیتم، هر سند پتنت به‌عنوان یک تراکنش در نظر گرفته شده است. برای تحلیل پتنت از چهار رقم اول کدهای IPC استفاده شده است (Jun 2011b; Altuntas, Dereli and Kusiak 2015b). روش پیشنهادی در شکل ۱، شرح داده شده است.



شکل ۱. چارچوب کلی پژوهش

همان‌طور که در شکل ۱، دیده می‌شود، ابتدا اطلاعات پتنت شامل شمارهٔ پتنت، کدهای IPC، تاریخ صدور و تعداد استنادها از پایگاه دادهٔ پتنت به‌دست آورده می‌شود.

سپس، قوانین وابستگی وزن‌دار مبتنی بر الگوریتم‌های داده‌کاوی برای یافتن وابستگی‌های میان فناوری‌ها استفاده شده است. مراحل پژوهش به شرح زیر است:

آماده‌سازی داده‌ها

اسناد پتنت مرتبط با فناوری تئوری پایگاه داده و کاربردهای آن (DTA[1]) از پایگاه دادهٔ پتنت آمریکا (USPTO) موجود در سایت «فری پتنتس آنلاین»[2] با استفاده از کوئری[3] زیر استخراج شده‌اند.

Title: databases AND (multimedia OR spatial OR active OR biomedical OR classification OR ranking OR clustering OR mining OR visualization OR integration OR management OR ubiquitous OR mobile OR extraction OR modeling OR provenance OR quality OR security OR privacy OR streaming OR heterogeneous OR warehousing OR OLAP OR parallel OR distributed OR query OR processing OR reliability OR semantic OR metadata OR analysis OR temporal OR services) (Thorleuchter, Poel and Prinzie 2007). (Jun 2011b).

به این ترتیب، ۱۰۰۲۷۰ پتنت بازیابی شدند. در این مرحله با استفاده از زبان برنامه‌نویسی PHP، یک نرم‌افزار هوشمند (ربات) برای پردازش داده‌ها طراحی شد. این نرم‌افزار داده‌های پتنت را بررسی کرده و اطلاعات مورد نیاز (شمارهٔ پتنت، تاریخ صدور، تعداد استنادها، کدهای IPC) را تهیه و ذخیره‌سازی می‌کند. خروجی این مرحله در قالب یک پایگاه دادهٔ ساخت‌یافته به‌عنوان ورودی مرحلهٔ داده‌کاوی در نظر گرفته شده است. در این مطالعه ۸۸۰ پتنت با توجه به مفروضات در نظر گرفته‌شده بازیابی شدند. اما تنها ۲۲۹ پتنت پس از حذف پتنت‌های متعلق به یک کلاس IPC یا نامرتبط به DTA در نظر گرفته شدند. اگر پتنتی متعلق به یک کلاس IPC باشد، در نظر گرفته نمی‌شود؛ زیرا برای تعیین وابستگی میان فناوری‌ها، حداقل دو کد IPC مورد نیاز است. پتنت‌های بازیابی‌شده در این مطالعه پیش از ۲۰ جولای ۲۰۱۲ صادر شده‌اند. گاهی اوقات باید از صدور پتنت‌ها زمانی سپری شود (برای مثال، حدود ۴ سال) تا ارجاعات و استنادهای اولیه انجام گیرد (Gay et al. 2005). در جدول ۱، برشی از پایگاه دادهٔ ساخت‌یافته ارائه شده است.

1. Database theory and its applications (DTA)
2. www.freepatentsonline.com
3. Query

جدول ۱. پایگاه دادهٔ ساخت‌یافته

شماره پتنت	تاریخ ثبت پتنت	تعداد استنادها	کدهای IPC
US1	۲۰۱۲/۰۷/۱۷	۸	G06F, H04L
US2	۲۰۱۲/۰۷/۱۷	۱۱	G06F, G06N
US3	۲۰۱۲/۰۷/۳	۱	G06F, G06N
US4	۲۰۱۲/۰۶/۲۶	۵	G06F, G06N
US5	۲۰۱۲/۰۶/۱۹	۱	G06F, G06N, G10L
US6	۲۰۱۲/۰۶/۱۲	۹	G10L, G06F
US7	۲۰۱۲/۰۵/۲۹	۳	G06Q, H04M, G06F
US8	۲۰۱۲/۰۵/۲۹	۶	G06K, H04B
US9	۲۰۱۲/۰۵/۱۵	۱۱	G06F, H04J, H04L
US10	۲۰۱۲/۰۵/۱	۲	G06E, G06F, G06G

داده‌کاوی با الگوریتم‌های قوانین وابستگی وزن‌دار

پس از آماده‌سازی داده‌ها، فرایند داده‌کاوی و اجرای الگوریتم‌های قوانین وابستگی آغاز می‌شود. داده‌کاوی با استفاده از نرم‌افزار «متلب» انجام شده است. این نرم‌افزار امکان انجام آسان عملیات ماتریسی، محاسباتی و توابعی، استفاده از الگوریتم‌های مختلف و همچنین، امکان ارتباط آسان با زبان‌های مختلف برنامه‌نویسی را فراهم می‌کند. در ادامه، الگوریتم‌های قوانین وابستگی وزن‌دار مورد استفاده در قالب یک مثال ارائه شده است.

کاوش قوانین وابستگی وزن‌دار بولین (BOWARM)

قوانین وابستگی وزن‌دار بولین نخستین بار توسط «لی و هو» در سال ۲۰۱۱ و به‌منظور ارائهٔ چارچوبی جدید برای در نظر گرفتن وزن‌ها در الگوریتم قوانین وابستگی معرفی شد (Li and Hu 2011). این الگوریتم با در نظر گرفتن وزن آیتم‌ها، ارزش قوانین را محاسبه می‌کند. دو تعریف کلیدی در این الگوریتم عبارت‌اند از:

وزن مجموعهٔ آیتم هر تراکنش (ITW)[1]: مجموع وزن همهٔ آیتم‌های موجود در مجموعهٔ

1. Itemset Transaction Weight (ITW)

آیتم تقسیم بر تعداد کل آیتم‌ها (Li and Hu 2011).

پشتیبانی وزن‌دار[1]: (WS) مجموع ITW تمام پتنت‌های شامل مجموعهٔ آیتم مورد نظر تقسیم بر تعداد کل پتنت‌ها (Li and Hu 2011).

چارچوب پیشنهادی «آلتونتاس، درلی و کوسیاک» با بهره‌گیری از الگوریتم BOWARM به شرح زیر است: (Altuntas, Dereli and Kusiak 2015b).

الف) محاسبهٔ وزن سطوح فناوری (آیتم‌ها)

لیست پتنت‌ها و کدهای IPC در جدول ۲، ارائه شده است.

جدول ۲. لیست پتنت‌ها (تراکنش‌ها)

شماره پتنت	کدهای IPC
US1	G06F, G06Q
US2	G06F, G06G, C12Q, G01N
US3	G06F, H04L, H04N
US4	G06F, G11B, H04N
US5	H04Q, G08B, H04L

پنج پتنت مرتبط با فناوری فرضی X وجود دارد (US1, US2, US3, US4, US5). هر پتنت شامل بیش از یک کد IPC مرتبط با فناوری X است. همان‌طور که در جدول ۲، نشان داده شده، ۱۰ کد آیتم (کد طبقه‌بندی) وجود دارد. تعداد پتنت‌های موجود در هر کد IPC را محاسبه می‌کنیم. برای مثال، ۴ پتنت وجود دارند که تحت کد G06F طبقه‌بندی شده‌اند (US1, US2, US3, US4). برای محاسبهٔ وزن نرمال‌شدهٔ هر کد IPC، تعداد پتنت‌های موجود در کد IPC مورد نظر بر تعداد کل پتنت‌های بازیابی‌شده تقسیم می‌شود. برای مثال، (۰/۵۱=۱۹۱/۳۷۷). بنابراین، وزن اختصاص‌یافته به کد G06F، ۰/۵۱ است. وزن‌ها بین ۰ تا ۱ تغییر می‌کنند و مجموع آن‌ها ۱ است. در جدول ۳، یک نمونه پایگاه داده آیتم وزن‌دار نشان داده شده است.

1. Weighted Support (WS)

جدول ۳. وزن آیتم‌ها (کدهای IPC) در الگوریتم BOWARM

ردیف	آیتم	تعداد پتنت‌ها	وزن کدهای IPC
۱	G06F	۱۹۱	۰/۵۱
۲	G06Q	۵۱	۰/۱۴
۳	G06G	۹	۰/۰۲
۴	C12Q	۲	۰/۰۱
۵	G01N	۸	۰/۰۲
۶	H04L	۶۴	۰/۱۷
۷	H04N	۲۰	۰/۰۵
۸	G11B	۵	۰/۰۱
۹	H04Q	۲۲	۰/۰۶
۱۰	G08B	۲	۰/۰۱
	مجموع	۳۷۷	۱

پس از محاسبهٔ وزن هر آیتم، وزن مجموعهٔ آیتم‌های محتمل (ITW) محاسبه می‌شود. به‌عنوان مثال، ITW برای مجموعهٔ آیتم (G06F,H04N) با توجه به جدول ۳، با استفاده از معادلهٔ ۱ محاسبه می‌شود:

$$ITW\ (G06F, H04N) = \frac{\{(G06F) + (H04N)\}}{2} = \frac{0.51 + 0.05}{2} = 0.28 \quad (۱)$$

ب) محاسبهٔ ارزش قوانین وابستگی یا میزان ارتباط در خوشه‌های فناوری

پس از محاسبهٔ وزن هر مجموعهٔ آیتم، برای به‌دست آوردن ارزش قوانین وابستگی، وزن مجموعهٔ آیتم در تعداد پتنت‌های شامل مجموعهٔ آیتم ضرب شده و بر تعداد کل پتنت‌ها تقسیم می‌شود. قدرت وابستگی بین مجموعهٔ آیتم‌ها در این الگوریتم با استفاده از ارزش WS شناخته می‌شود. برای مجموعهٔ آیتم (G06F,H04N) با توجه به جدول ۲، محاسبات به شرح معادلهٔ ۲ است:

$$WS(G06F, H04N) = \frac{\left\{\left(ITW(G06F, H04N)\right) \times \text{تعداد کل پتنت‌های شامل مجموعه آیتم}\right\}}{\text{تعداد کل پتنت‌ها}} = \frac{0.28 \times 2}{5} = 0.112 \quad (۲)$$

به‌طور خلاصه، چارچوب الگوریتم BOWARM به شرح زیر است:

۱. بازیابی پتنت‌ها از پایگاه دادهٔ پتنت؛
۲. محاسبهٔ وزن هر یک از آیتم‌ها (تمام کدهای IPC مربوط به پتنت‌های بازیابی‌شده)؛
۳. یافتن تمامی مجموعهٔ آیتم‌های ممکن؛
۴. محاسبهٔ ارزش ITW برای هر مجموعهٔ آیتم؛
۵. محاسبهٔ ارزش WS برای هر مجموعهٔ آیتم؛
۶. مرتب‌سازی مجموعهٔ آیتم‌ها به‌ترتیب نزولی‌بودن ارزش WS.

کاوش قوانین وابستگی وزن‌دار (WARM_2)

الگوریتم WARM_2 از دیگر الگوریتم‌های پرکاربرد قوانین وابستگی وزن‌دار است که از الگوریتم WARM، که توسط «تائو، مورتاق و فرید» در سال ۲۰۰۳ معرفی شده، توسعه یافته است (Tao, Murtagh and Farid 2003 نقل در Altuntas, Dereli, and Kusiak 2015b). این الگوریتم با در نظر گرفتن وزن تراکنش‌ها، مقدار قوانین وابستگی را محاسبه می‌کند. دو تعریف کلیدی در این الگوریتم عبارت‌اند از:

وزن تراکنش (TW_2[1]) :تعداد کل استنادهای پتنت تقسیم بر تعداد سال‌های سپری‌شده از تاریخ صدور پتنت مورد نظر (Altuntas, Dereli and Kusiak 2015b).

مجموعهٔ آیتم وزن‌دار (WI_2[2]): مجموع وزن تمام پتنت‌های شامل مجموعهٔ آیتم مورد نظر (Altuntas, Dereli and Kusiak 2015b). چارچوب پیشنهادی «آلتونتاس، درلی و کوسیاک» با بهره‌گیری از الگوریتم WARM_2 به شرح زیر است: (Altuntas, Dereli and Kusiak 2015b).

الف) محاسبهٔ وزن پتنت‌ها

جدول ۴، یک پایگاه دادهٔ پتنت وزن‌دار را نمایش می‌دهد.

1. Transaction Weight (TW) 2. Weighted Itemset (WI)

جدول ٤. پایگاه داده پتنت وزن‌دار

گام ۳	گام ۲	گام ۱	داده‌های بازیابی‌شده از پایگاه دادهٔ پتنت		
وزن پتنت‌ها (TW)	میانگین تعداد استنادها	تعداد سال‌های سپری‌شده	تاریخ صدور*	تعداد استنادها	پتنت
۰/۲۴	۳۰	۴	۲۰۱۲	۱۲۰	US1
۰/۲	۲۵	۹	۲۰۰۷	۲۲۵	US2
۰/۰۹۶	۱۲	۱۲	۲۰۰۴	۱۴۴	US3
۰/۲	۲۵	۱۸	۱۹۹۸	۴۵۰	US4
۰/۲۶۴	۳۳	۱۶	۲۰۰۰	۵۲۸	US5
۱	۱۲۵		مجموع		

*تاریخ صدور پتنت از ۲۰۱۶ کم می‌شود.

برای محاسبهٔ وزن پتنت‌ها از استنادهای پتنت استفاده شده است. پتنت‌هایی که به‌تازگی صادر شده‌اند، به‌طور طبیعی تعداد کمی استناد خواهند داشت. همان‌طور که در بخش آماده‌سازی داده‌ها ذکر شد، پتنت‌های بازیابی‌شده در این مطالعه پیش از ۲۰ جولای ۲۰۱۲ صادر شده‌اند تا ارجاعات و استنادهای اولیه انجام شود. همچنین، برای حذف تأثیر زمان، تعداد کل استنادهای هر پتنت را بر تعداد سال‌هایی که از صدور آن می‌گذرد، تقسیم می‌کنیم تا میانگین تعداد استنادها برای یک سال به‌دست آید. در نهایت، برای نرمال‌سازی، میانگین تعداد استنادها برای هر پتنت بر میانگین کلی تعداد استناد پتنت‌ها تقسیم می‌شود.

به این ترتیب، وزن اختصاص داده‌شده به هر پتنت بین ۰ تا ۱ و مجموع وزن‌ها برابر ۱ خواهد بود. به‌عنوان مثال، برای پتنت US1 ارائه‌شده در جدول ۴، تعداد سال‌های سپری شده ۴ است (۲۰۱۲ - ۲۰۱۶) و میانگین تعداد استنادهای پتنت در یک سال ۳۰ است (۱۲۰/۴). همچنین، وزن نرمال‌شدهٔ آن برابر ۰/۲۴۰ خواهد بود (۳۰/۱۲۵).

ب) محاسبهٔ میزان ارتباط در خوشه‌های فناوری

در این گام، برای به‌دست آوردن ارزش قوانین وابستگی، مجموع وزن تمام پتنت‌های شامل مجموعهٔ آیتم را به‌دست می‌آوریم. به‌عنوان مثال، برای مجموعهٔ آیتم (G06F, H04N) با توجه به جدول ۲ و ۴:

$$WI - 2(G06F, H04N)$$
$$= \{(G06F, H04N)\{\text{های شامل مجموعهٔ آیتم‌مجموع وزن تمامی پتنت}\} \quad (۱)$$
$$= \{TW_{US3} + TW_{US4}\} = 0.096 + 0.200 = 0.296$$

به‌طور خلاصه، مراحل الگوریتم WARM_2 به شرح زیر است:

۱. بازیابی پتنت‌ها از پایگاه داده پتنت؛
۲. محاسبهٔ ارزش TW_2 برای هر تراکنش (پتنت)؛
۳. یافتن تمامی مجموعهٔ آیتم‌های ممکن؛
۴. محاسبهٔ ارزش WI_2 برای هر مجموعهٔ آیتم؛
۵. مرتب‌سازی همهٔ مجموعهٔ آیتم‌ها به‌ترتیب نزولی‌بودن وزن.

کاوش قوانین وابستگی وزن‌دار ترکیبی ۱ (CBW_1)

الگوریتم ترکیبی ۱، با در نظر گرفتن وزن آیتم‌ها، ارزش قوانین وابستگی را محاسبه می‌کند. در الگوریتم حاضر وزن آیتم‌ها بر اساس تعداد استناد پتنت‌های شامل مجموعهٔ آیتم مورد نظر محاسبه می‌شود. تعاریف مورد استفاده در الگوریتم ترکیبی ۱، به شرح زیر است:

وزن مجموعهٔ آیتم (ITW) :مجموع وزن همهٔ آیتم‌های موجود در مجموعهٔ آیتم تقسیم بر تعداد کل آیتم‌ها.

پشتیبانی وزن‌دار (WS): مجموع ITW تمام پتنت‌های شامل مجموعهٔ آیتم مورد نظر تقسیم بر تعداد کل پتنت‌ها.

چارچوب پیشنهادی با بهره‌گیری از الگوریتم ترکیبی ۱، به شرح زیر است:

الف) محاسبهٔ وزن سطوح فناوری (آیتم‌ها)

برای محاسبهٔ وزن آیتم‌ها ابتدا وزن هر پتنت با تقسیم تعداد استنادهای پتنت بر تعداد سال‌های گذشته از صدور آن به‌دست آورده شده و سپس، نرمال‌سازی می‌شود. در مرحلهٔ بعد، وزن پتنت‌های شامل آیتم (کد IPC) مورد نظر به‌عنوان وزن آیتم جمع می‌شوند که در جدول ۵، نشان داده شده است.

جدول ۵. وزن آیتم در الگوریتم ترکیبی ۱(CBW_1)

ردیف	آیتم	پتنت‌های شامل آیتم	وزن کدهای IPC
۱	G06F	US1, US2, US3, US4	۰/۷۳۶ (۰/۲۴۰+۰/۲۰۰+۰/۰۹۶+۰/۲۰۰)
۲	G06Q	US1	۰/۲۴۰
۳	G06G	US2	۰/۲۰۰
۴	C12Q	US2	۰/۲۰۰
۵	G01N	US2	۰/۲۰۰
۶	H04L	US3, US5	۰/۳۶۰
۷	H04N	US3, US4	۰/۲۹۶
۸	G11B	US4	۰/۲۰۰
۹	H04Q	US5	۰/۲۶۴
۱۰	G08B	US5	۰/۲۶۴

سپس، وزن مجموعهٔ آیتم‌های محتمل نیز محاسبه می‌شود. به‌عنوان مثال، برای مجموعهٔ آیتم (G06F, H04N) با توجه به جدول ۵، محاسبات به شرح معادلهٔ ۴ خواهد بود:

$$ITW(G06F, H04N) = \frac{\{(G06F) + (H04N)\}}{2} = \frac{0.736 + 0.296}{2} = 0.516 \qquad (۴)$$

ب) محاسبهٔ میزان ارتباط میان فناوری‌های مختلف

برای به‌دست آوردن ارزش یک قانون وابستگی مشابه روش BOWARM، وزن مجموعهٔ آیتم مورد نظر در تعداد پتنت‌های شامل مجموعهٔ آیتم ضرب شده و بر تعداد کل پتنت‌ها تقسیم می‌شود. به‌عنوان مثال، برای مجموعهٔ آیتم (G06F, H04N) با توجه به جدول ۲، در معادلهٔ ۵ خواهیم داشت:

$$WS(G06F, H04N) = \frac{\left\{\left(ITW(G06F, H04N)\right) \times \text{تعداد کل پتنت‌های شامل مجموعهٔ آیتم}\right\}}{\text{تعداد کل پتنت‌ها}} = \frac{0.516 \times 2}{5} = 0.2064 \qquad (۵)$$

به‌طور خلاصه، چارچوب الگوریتم CBW_1 به شرح زیر است:

۱. بازیابی پتنت‌ها از پایگاه دادهٔ پتنت؛

۲. محاسبهٔ وزن هر یک از آیتم‌ها؛
۳. یافتن تمامی مجموعهٔ آیتم‌های محتمل؛
۴. محاسبهٔ ارزش ITW برای هر مجموعهٔ آیتم؛
۵. محاسبهٔ ارزش WS برای هر مجموعهٔ آیتم؛
۶. مرتب‌سازی مجموعهٔ آیتم‌ها به‌ترتیب نزولی بودن میزان WS.

کاوش قوانین وابستگی وزن‌دار ترکیبی۲ (CBW_2)

الگوریتم ترکیبی ۲، با در نظر گرفتن وزن تراکنش‌ها، ارزش قوانین وابستگی را محاسبه می‌کند. در الگوریتم حاضر وزن تراکنش‌ها بر اساس تعداد استنادهای پتنت و نیز وزن مجموعهٔ آیتم مورد نظر محاسبه می‌شود. تعاریف مورد استفاده در الگوریتم ترکیبی ۲، به شرح زیر است:

وزن تراکنش (TW) :وزن استنادی پتنت (شامل مجموعهٔ آیتم) ضرب در وزن مجموعهٔ آیتم مورد نظر.

مجموعهٔ آیتم وزن‌دار (WI): مجموع وزن تمام پتنت‌های شامل مجموعهٔ آیتم مورد نظر.

چارچوب الگوریتم ترکیبی ۲ به شرح زیر می‌باشد:

۱. محاسبهٔ وزن پتنت‌ها

وزن هر آیتم با تقسیم تعداد پتنت‌های موجود در آیتم بر تعداد کل پتنت‌ها محاسبه شده و وزن هر مجموعهٔ آیتم با تقسیم مجموع وزن آیتم‌های موجود در مجموعهٔ آیتم بر تعداد کل آیتم‌های موجود به‌دست می‌آید. برای محاسبهٔ وزن تراکنش‌ها، ابتدا وزن استنادی پتنت‌های شامل مجموعهٔ آیتم با تقسیم تعداد استنادهای پتنت بر تعداد سال‌های سپری‌شده از صدور آن محاسبه می‌شود و سپس، مطابق جدول ۴، نرمال‌سازی می‌شود. در نهایت، در وزن مجموعهٔ آیتم مورد نظر ضرب می‌شود. به‌عنوان مثال، برای مجموعهٔ آیتم (G06F, H04N) پس از محاسبهٔ وزن هر آیتم (مطابق جدول ۳)، در معادلهٔ ۶ خواهیم داشت:

$$ITW(G06F, H04N) = \frac{\{(G06F) + (H04N)\}}{2} = \frac{0.51 + 0.05}{2} = 0.28 \qquad (۶)$$

با توجه به جداول ۲، ۳ و ۴ وزن پتنت‌ها به‌ترتیب زیر خواهد بود:

جدول ٦. وزن پتنت‌ها در الگوریتم ترکیبی ۲ (CBW_2)

شماره پتنت	کدهای IPC	وزن استنادی پتنت‌ها	وزن مجموعه آیتم (G06F,H04N) ITW(G06F,H04N)	وزن تراکنش (TW)
US1	G06F, G06Q	۰/۲۴۰	۰/۲۸	–
US2	G06F, G06G, C12Q, G01N	۰/۲۰۰	۰/۲۸	–
US3	G06F, H04N, H04L	۰/۰۹۶	۰/۲۸	۰/۰۲۶۸۸(۰/۲۸*۰/۰۹۶)
US4	G06F, H04N, G11B	۰/۲۰۰	۰/۲۸	۰/۰۵۶۰۰
US5	H04Q, G08B, H04L	۰/۲۶۴	۰/۲۸	–

۲. محاسبهٔ میزان ارتباط میان فناوری‌های مختلف

پس از محاسبهٔ وزن هر پتنت شامل مجموعهٔ آیتم مورد نظر، مجموع وزن‌های تمامی این پتنت‌ها برای محاسبهٔ ارزش یک قانون وابستگی (مشابه روش WARM_ 2) به‌دست آورده می‌شود. به‌عنوان مثال، میزان ارتباط میان مجموعهٔ آیتم (G06F, H04N) با توجه به جدول ۶، به‌صورت زیر در معادلهٔ ۷ به‌دست می‌آید:

$$WI(G06F,H04N) = \{\text{مجموع وزن تمام پتنت‌های شامل مجموعه آیتم}(G06F,H04N)\} = \{TW_{US3} + TW_{US4}\} = 0.02688 + 0.056 = 0.0828 \quad (۷)$$

به‌طور خلاصه، مراحل الگوریتم CBW_2 به شرح زیر است:

۱. بازیابی پتنت‌ها از پایگاه دادهٔ پتنت؛
۲. محاسبهٔ وزن هر یک از آیتم‌ها؛
۳. یافتن تمامی مجموعهٔ آیتم‌های محتمل؛
۴. محاسبهٔ ارزش ITW برای هر مجموعهٔ آیتم؛
۵. محاسبهٔ وزن استنادی هر پتنت؛
۶. محاسبهٔ ارزش TW برای هر تراکنش شامل مجموعهٔ آیتم مورد نظر؛
۷. محاسبهٔ ارزش WI برای هر مجموعهٔ آیتم؛
۸. مرتب‌سازی مجموعهٔ آیتم‌ها به‌ترتیب نزولی‌بودن میزان WI.

کاوش قوانین وابستگی وزن‌دار ترکیبی ۳ (CBW_3)

الگوریتم پیشنهادی ترکیبی ۳، به‌منظور محاسبهٔ میزان ارتباط میان فناوری‌های مختلف با توجه به وزن و اهمیت نابرابر کدهای IPC و همچنین پتنت‌ها ارائه شده است. در این الگوریتم از تعداد پتنت‌های هر کد IPC به‌عنوان معیار وزن‌دهی آیتم‌ها و از تعداد استنادهای پتنت به‌عنوان معیار وزن‌دهی تراکنش‌ها استفاده شده است. تعاریف مورد استفاده در الگوریتم ترکیبی ۳، به شرح زیر است:

وزن مجموعهٔ آیتم هر تراکنش (ITW): مجموع وزن همهٔ آیتم‌های موجود در مجموعهٔ آیتم تقسیم بر تعداد کل آیتم‌ها.

وزن تراکنش (TW_2): تعداد کل استنادهای پتنت، تقسیم بر تعداد سال‌های سپری‌شده از تاریخ صدور پتنت مورد نظر.

پشتیبانی وزن‌دار (WS_{Total}): ترکیب وزن مجموعهٔ آیتم و پشتیبانی وزن‌دار.

چارچوب الگوریتم پیشنهادی به شرح زیر است:

الف) محاسبهٔ وزن پتنت‌ها و وزن سطوح فناوری (آیتم‌ها)

این الگوریتم با در نظر گرفتن وزن کدهای IPC (آیتم‌ها) و همچنین، وزن پتنت‌ها (تراکنش‌ها) ارزش قوانین وابستگی را محاسبه می‌کند. روش محاسبهٔ وزن آیتم‌ها مشابه آنچه در الگوریتم BOWARM توضیح داده شد و روش محاسبهٔ وزن پتنت‌ها مشابه الگوریتم WARM_2 است. به‌عنوان مثال، برای مجموعهٔ آیتم (G06F, H04N) پس از محاسبهٔ وزن هر آیتم، با توجه به جدول ۳، در معادلهٔ ۸ خواهیم داشت:

$$ITW(G06F, H04N) = \frac{\{(G06F) + (H04N)\}}{2} = \frac{0.51 + 0.05}{2} = 0.28 \qquad (۸)$$

برای به‌دست آوردن وزن پتنت‌ها نیز مطابق جدول ۴، تعداد کل استنادهای هر پتنت بر تعداد سال‌هایی که از صدور آن سپری شده تقسیم می‌شود. سپس، برای نرمال‌کردن، میانگین تعداد استنادها برای هر پتنت بر میانگین کلی تعداد استناد پتنت‌ها تقسیم می‌شود.

ب) محاسبهٔ میزان ارتباط در خوشه‌های فناوری

پس از محاسبهٔ وزن آیتم‌ها و پتنت‌ها برای به‌دست آوردن مقدار پشتیبانی وزن‌دار، ارزش قوانین وابستگی حاصل از الگوریتم BOWARM با ارزش قوانین حاصل از الگوریتم

WARM_2 به‌صورت معادلهٔ ۹ ترکیب می‌شود.

$$WS_{Total} = WS(G06F, H04N) + WI_2(G06F, H04N = 0.112 + 0.296 = 0.408 \quad (۹)$$

به‌طور خلاصه، مراحل الگوریتم CBW_3 به شرح زیر است:

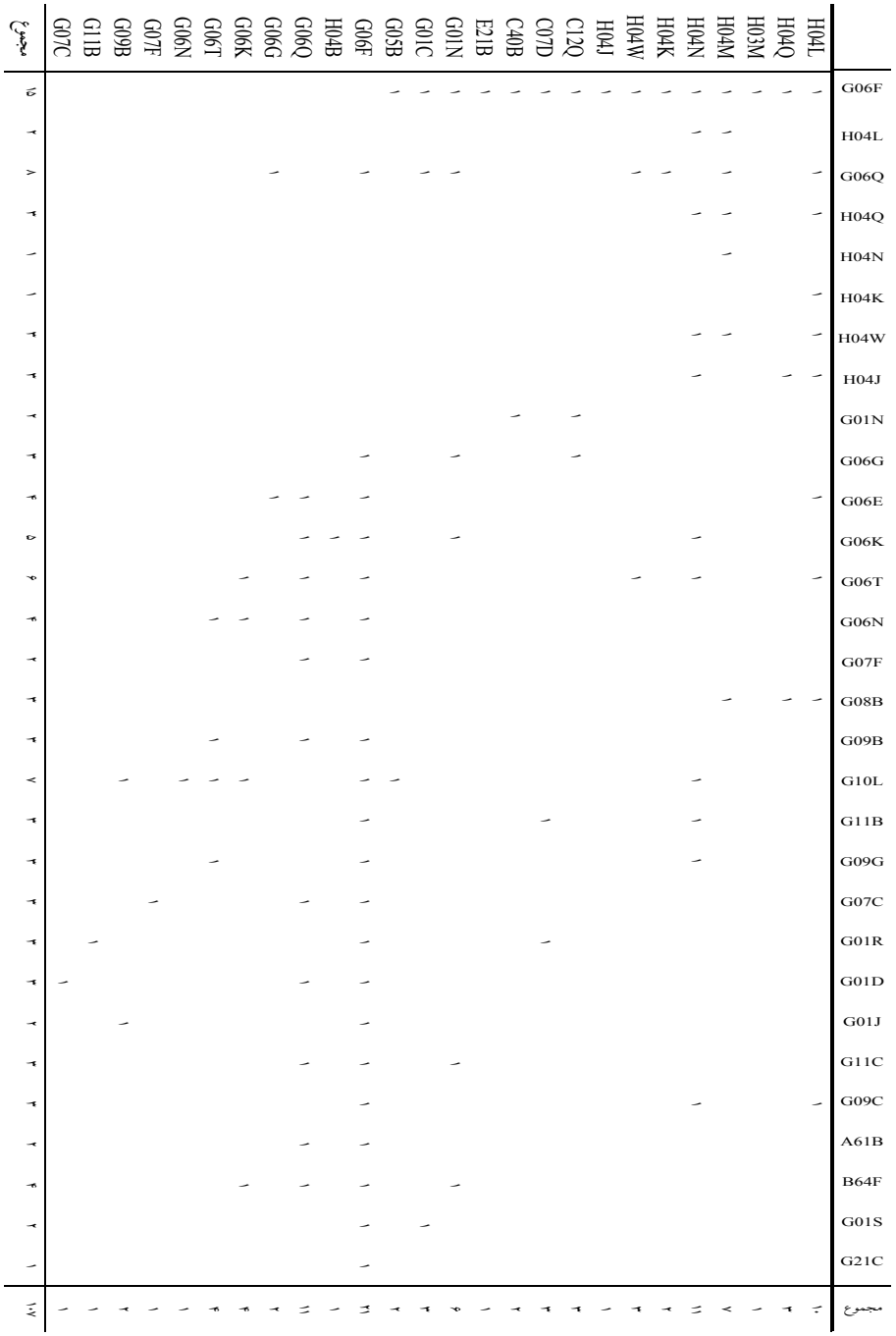
۱. بازیابی پتنت‌ها از پایگاه دادهٔ پتنت؛
۲. محاسبهٔ میزان TW_2 برای هر پتنت و وزن هر یک از آیتم‌ها؛
۳. یافتن تمامی مجموعهٔ آیتم‌های ممکن؛
۴. محاسبهٔ میزان ITW برای هر مجموعهٔ آیتم؛
۵. محاسبهٔ میزان WS_{Total} برای هر مجموعهٔ آیتم؛
۶. مرتب‌سازی مجموعهٔ آیتم‌ها به‌ترتیب نزولی‌بودن میزان WS_{Total}.

۵. نتایج پژوهش

از میان پتنت‌های بررسی‌شده، ۲۲۹ پتنت مطلوب و ۳۹ کد IPC منحصربه‌فرد به‌دست آمد. آخرین پتنت در نظر گرفته‌شده در این مطالعه، US8214319 است که در تاریخ ۳ جولای ۲۰۱۲ صادر شده است. بیشترین ارتباط و وابستگی متعلق به کد G06F، پردازش دادهٔ دیجیتال الکتریکی[1] است. برای درک دقیق‌تر، سطح ارتباطات دوتایی در قالب ماتریس (جدول ۷) نمایش داده می‌شود. جدول ۷، تمامی ارتباطات میان آیتم‌های موجود در ۲۲۹ پتنت را ارائه می‌دهد. وجود ارتباط میان سطوح مختلف فناوری با عدد ۱، نشان داده شده است. در میان ۱۰۷ ارتباط موجود در میان خوشه‌های دوتایی، فناوری G06F با ۳۶ کلاس مختلف فناوری ارتباط دارد. به بیان دقیق‌تر، هر تغییری در فناوری G06F، به‌عنوان مثال، در زمینهٔ طراحی، بر ۳۶ فناوری دیگر تأثیر می‌گذارد و بالعکس. بنابراین، فناوری G06F می‌تواند به‌عنوان مهم‌ترین فناوری در میان ۳۹ فناوری شناسایی‌شده در نظر گرفته شود. فناوری‌های درگیر در H03M، E21B، G06E، G10L، G01J، G01D، G01R، G09G، G11C، G09C، A61B، G01S، G21C تنها یک ارتباط دارند و آن ارتباط نیز با فناوری G06F است. بدین‌ترتیب، اگر فناوری G06F توسعه داده شود، این امکان وجود دارد که فناوری‌های مذکور نیز به‌دلیل وابستگی میان آن‌ها در آینده توسعه داده شوند.

1. electric digital data processing

جدول ۷. ماتریس ارتباطات میان خوشه‌های دوتایی



وابستگی می‌تواند در خوشه‌های دو، سه، چهار و پنج‌تایی رخ دهد. همچنین، حداکثر اندازهٔ ممکن یک خوشهٔ فناوری برابر است با حداکثر تعداد آیتم‌های (کدهای IPC) موجود در یک تراکنش (پتنت) از بین پتنت‌های بازیابی‌شده. (در اینجا پنج بود). در این پژوهش، وابستگی در یک خوشهٔ فناوری به این معناست که این فناوری‌ها بر یکدیگر اثر می‌گذارند. به‌عنوان مثال، برای خوشه دوتایی (G06F, G06Q)، در صورت توسعهٔ فناوری G06F، این احتمال وجود دارد که در آینده فناوری G06Q به‌دلیل وابستگی میان آن‌ها توسعه داده شود و نیز سرمایه‌گذاری در فناوری G06F، موجب افزایش احتمال سرمایه‌گذاری در فناوری G06Q در آینده خواهد شد. خوشهٔ فناوری (G06F, G06Q, H04L) نیز به چنین رابطه‌ای میان سه فناوری دلالت می‌کند و خوشه‌های چهار و پنج‌تایی نیز به همین ترتیب خواهند بود. جداول ۸ و ۹ و ۱۰، ۱۰ ترکیب اول از خوشه‌های دو تا چهارتایی و جدول ۱۱، خوشه‌های پنج‌تایی به‌دست آمده از الگوریتم‌های مورد استفاده را نشان می‌دهند. قدرت وابستگی میان مجموعهٔ آیتم‌ها در الگوریتم BOWARM و CBW_1 با استفاده از ارزش WS و در WARM_2 و CBW_2 توسط WI و در الگوریتم CBW_3 توسط WS_{Total} سنجیده می‌شود.

همان‌طور که در جدول ۸، دیده می‌شود، اولین و دومین خوشهٔ دوتایی برای همهٔ روش‌ها مشابه هستند. خوشه‌های باقی‌مانده برای روش‌های مختلف توالی‌های متفاوت دارند. در جدول ۹، هشت مجموعهٔ آیتم اول برای سطح ۳ (خوشهٔ سه‌تایی) در روش‌های BOWARM و CBW_1 یکسان هستند و همچنین، چهار مجموعهٔ آیتم اول برای سطح ۳ در روش‌های WARM_2 و CBW_2 و CBW_3 یکسان هستند، هرچند، این مجموعه آیتم‌ها قدرت وابستگی متفاوتی دارند. بر اساس جدول ۱۰، تمامی روش‌ها نتایج مشابهی را برای اولین مجموعهٔ آیتم در سطح ۴ تولید کردند. ده مجموعهٔ آیتم اول برای سطح ۴ که توسط روش‌های BOWARM و CBW_2 تولید شده‌اند، مشابه هستند و نیز ۴ مجموعهٔ آیتم اول برای روش‌های WARM_2 و CBW_3 یکسان هستند. همان‌طور که در جدول ۱۱، نشان داده شده، روش‌های BOWARM و CBW_1 توالی یکسان برای سطح ۵ مجموعهٔ آیتم تولید می‌کنند و الگوریتم‌های WARM_2 و CBW_2 و CBW_3 توالی‌های مشابه برای سطح ۵ مجموعهٔ آیتم فراهم کرده‌اند. در این مطالعهٔ موردی، علت یکسان‌شدن توالی مجموعهٔ آیتم‌ها در برخی سطوح، شدت ارتباط و وابستگی مجموعهٔ آیتم‌هاست. در واقع، مجموعهٔ آیتم‌های به شدت مرتبط، با وجود روش‌های مختلف، توالی‌های یکسان تولید می‌کنند.

جدول ۸. برشی از خوشه‌های دوتایی (اعداد بر حسب درصد هستند)

CBW_3		CBW_2		CBW_1		WARM_2		BOWARM	
WS_{TOTAL}	مجموعهٔ آیتم	WI	مجموعهٔ آیتم	WS	مجموعهٔ آیتم	WI	مجموعهٔ آیتم	WS	مجموعهٔ آیتم
۳/۶۲۶۵	H04L, G06F	۰/۷۲۶۶۸	H04L, G06F	۱/۳۵۸۱	H04L, G06F	۳/۱۱۷۶	H04L, G06F	۰/۵۰۸۹	H04L, G06F
۳/۰۱۰۷	G06F, G06Q	۰/۵۶۹۸۴	G06F, G06Q	۱/۱۳۳۴	G06F, G06Q	۲/۵۷۶	G06F, G06Q	۰/۴۳۴۶	G06F, G06Q
۱/۳۹۱۵	G06F, G06K	۰/۲۳۲۰۱	G06F, G06K	۰/۶۱۰۰۱	G06F, G06K	۱/۱۵۷۹	G06F, G10L	۰/۲۴۸۱۲	G06F, G06K
۱/۲۴۸	G06F, G10L	۰/۲۱۵۰۳	G06F, G10L	۰/۵۲۵۳۸	G06F, G06N	۱/۱۴۳۳	G06F, G06K	۰/۲۲۶۲۴	G06F, G06N
۱/۰۹۹۷	G06F, G09B	۰/۱۸۹۹۲	G06F, G09B	۰/۳۲۸۴۸	G06F, G09B	۱/۰۴۴۱	H04N, G06F	۰/۱۳۴۷۶	H04N, G06F
۰/۹۷۸۸۱	G09B, G10L	۰/۰۸۸۹۴۳	G06F, G06T	۰/۲۴۵۳۶	G06F, G06T	۰/۹۷۴۸۲	G09B, G10L	۰/۱۰۰۱۱	G06F, G06T
۰/۷۳۸۵۲	H04L, G06Q	۰/۰۸۸۶۳۹	G06F, G06N	۰/۲۳۸۵۴	G06F, G10L	۰/۷۱۵۵۶	H04L, G06Q	۰/۰۸۹۱۳	G06F, G10L
۰/۶۷۱۰۶	G06F, G06N	۰/۰۸۴۷۲	H04N, G06F	۰/۱۵۹۸۶	G01N, G06F	۰/۴۶۵۵۷	G06F, G06T	۰/۰۶۳۸۶	G06F, G06G
۰/۵۷۴۰۴	H04N, G06F	۰/۰۷۵۲۱	H04L, G06Q	۰/۱۵۷۸	G06F, G06G	۰/۴۴۸۲	G06F, G06N	۰/۰۶۳۵۴	G01N, G06F
۰/۵۶۵۶	G06F, G06T	۰/۰۷۰۸۳	G01N, G06F	۰/۱۵۱۲۳	G06F, G09B	۰/۴۳۹۲۹	H04N, G06F	۰/۰۵۵۶	G06F, G09B

جدول ۹. برشی از خوشه‌های سه‌تایی (اعداد بر حسب درصد هستند)

BOWARM		WARM-2		CBW_1		CBW_2		CBW_3	
مجموعهٔ آیتم	WS	مجموعهٔ آیتم	WI	مجموعهٔ آیتم	WS	مجموعهٔ آیتم	WI	مجموعهٔ آیتم	WS TOTAL
H04N, H04L, G06F	۰/۰۳۶۵۹	G06F, G09B, G10L	۰/۹۷۴۸۲	H04N, H04L, G06F	۰/۰۹۵۲	G06F, G09B, G10L	۰/۱۲۵۳۴	G06F, G09B, G10L	۱/۰۰۲۹
H04L, G06F, G06Q	۰/۰۳۲۵۷۱.	H04L, G06F, G06Q	۰/۶۷۱۴۶	H04`L, G06F, G06Q	۰/۰۸۸۵۷۹	H04L, G06F, G06Q	۰/۱۲۵۲۱	H04L, G06F, G06Q	۰/۷۰۴۰۳
G06F, G09B, G10L	۰/۰۲۸۰۷۴	G01N, G06F, G06Q	۰/۳۵۰۴۶	G06F, G09B, G10L	۰/۰۸۰۵۲۴	G01N, G06F, G06Q	۰/۰۵۳۳۹۲	G01N, G06F, G06Q	۰/۳۷۰۴۲
G06F, G06G, G06E	۰/۰۲۷۵۴۲	G01N, G06F, G11C	۰/۳۱۹۵	G06F, G06G, G06E	۰/۰۶۷۹۵۶	G06F, G06Q, G11C	۰/۰۴۷۵۰۶	G06F, G06Q, G11C	۰/۳۳۳۲۴۸
H04N, G06F, G06K	۰/۰۲۷۷۵۹	G01N, G06Q, G11C	۰/۳۱۹۵	H04N, G06F, G06K	۰/۰۶۱۸۲۳	G01N, G06F, G11C	۰/۰۳۹۱۳۴	G01N, G06F, G11C	۰/۳۳۰۲
G01N, G06F, G06Q	۰/۰۱۹۹۵۸	G06F, G06Q, G11C	۰/۳۱۹۵	G01N, G06F, G06Q	۰/۰۵۲۰۷۴	H04M, H04L, G06F	۰/۰۳۰۷۱۹	G01N, G06Q, G11C	۰/۳۲۲۷۵
G06F, G06K, G06T	۰/۰۱۹۱۶	H04M, H04L, G06F	۰/۱۸۱۹۹	G06F, G06K, G06T	۰/۰۴۶۲۰۲	H04N, G06F, G06K	۰/۰۲۶۰۴۹	H04N, G06F, G06K	۰/۲۰۲۴
H04M, H04L, G06F	۰/۰۱۴۷۴۲	H04N, G06F, G06K	۰/۱۷۶۶۴	H04M, H04L, G06F	۰/۰۳۸۲۷۸	H04L, G06F, G06E	۰/۰۲۴۰۴۵	H04M, H04L, G06F	۰/۱۹۶۷۳
G06F, G06Q, G06N	۰/۰۱۴۳۱۷	H04L, G06F, G06E	۰/۱۵۰۶	H04M, G06F, G06Q	۰/۰۳۵۶۴۴	G06F, G06Q, G06E	۰/۰۲۲۸۵۲	H04L, G06F, G06E	۰/۱۵۷۵۷
H04M, G06F, G06Q	۰/۰۱۴۰۵	H04L, G06Q, G06E	۰/۱۵۰۶	G06F, G06Q, G06N	۰/۰۳۵۰۱۵	G06F, G06G, G06Q	۰/۰۱۸۲۷۲	G06F, G06Q, G06E	۰/۱۵۷۲۳

پژوهش‌نامه پردازش و مدیریت اطلاعات
تابستان ۱۳۹۷ | دورهٔ ۳۲ | شمارهٔ ٤

جدول ۱۰. برشی از خوشه‌های چهارتایی (اعداد بر حسب درصد هستند)

CBW_3		CBW_2		CBW_1		WARM_2		BOWARM	
WS_{TOTAL}	مجموعه آیتم	WI	مجموعه آیتم	WS	مجموعه آیتم	WI	مجموعه آیتم	WS	مجموعه آیتم
۰/۳۲۹۹۵۶	G01N, G06F, G06Q, G11C	۰/۰۳۶۷۹۸	G01N, G06F, G06Q, G11C	۰/۰۲۶۷۳۵	G01N, G06F, G06Q, G11C	۰/۳۱۹۵	G01N, G06F, G06Q, G11C	۰/۰۱۰۰۵۹	G01N, G06F, G06Q, G11C
۰/۱۵۶۸۵	H04L, G06F, G06Q, G06E	۰/۰۲۱۵۴۴	H04L, G06F, G06Q, G06E	۰/۰۱۹۶۸۳	C07D, G06F, G11B, G01R	۰/۱۵۰۶	H04L, G06F, G06Q, G06E	۰/۰۰۷۹۸۳	C07D, G06F, G11B, G01R
۰/۰۴۶۶۹۸	H04M, H04L, H04Q, G08B	۰/۰۰۳۹۷۶	G01N, G06F, G06Q, G06K	۰/۰۱۷۰۸۹	H04M, H04L, H04Q, G08B	۰/۰۳۹۹۳۲	H04M, H04L, H04Q, G08B	۰/۰۰۶۷۶۵	H04M, H04L, H04Q, G08B
۰/۰۳۶۵۷۳	G01N, G06F, G06Q, G06K	۰/۰۰۳۸۷۷	G06F, G06Q, G06K, B64F	۰/۰۱۶۹۳۹	H04L, G06F, G06Q, G06E	۰/۰۳۰۹۶۵	G01N, G06F, G06Q, G06K	۰/۰۰۶۲۴۶	H04L, G06F, G06Q, G06E
۰/۰۳۶۴۳۳	G06F, G06Q, G06K, B64F	۰/۰۰۳۵۵۲	G01N, G06F, G06Q, B64F	۰/۰۱۵۰۵۳	H04M, H04N, H04L, G06F	۰/۰۳۰۹۶۵	G01N, G06F, G06Q, B64F	۰/۰۰۵۹۲۷	H04M, H04N, H04L, G06F
۰/۰۳۵۹۶۴	G01N, G06F, G06Q, B64F	۰/۰۰۳۲۶۹	G01N, G06F, G06K, B64F	۰/۰۱۴۵۵۶	H04M, H04L, H04W, G06F	۰/۰۳۹۹۳۲	G01N, G06F, G06K, B64F	۰/۰۰۵۶۸۸	H04M, H04L, H04W, G06F
۰/۰۳۵۵۷۵	G01N, G06F, G06K, B64F	۰/۰۰۲۸۸۵	H04N, G06F, G06K, G06T	۰/۰۱۴۴۸۲	H04N, H04L, H04W, G06F	۰/۰۳۹۹۳۲	G01N, G06Q, G06K, B64F	۰/۰۰۵۶۴۸	H04N, H04L, H04W, G06F
۰/۰۳۴۷۹۶	C07D, G06F, G11B, G01R	۰/۰۰۲۷۴۸	G06F, G06K, G06T, G06N	۰/۰۱۴۳۴۶	G01N, G06F, G06Q, G06K	۰/۰۳۹۹۳۲	G06F, G06Q, G06K, B64F	۰/۰۰۵۶۰۸	G01N, G06F, G06Q, G06K
۰/۰۳۲۸۷۱	G01N, G06Q, G06K, B64F	۰/۰۰۲۴۵۰	C07D, G06F, G11B, G01R	۰/۰۱۴۲۷۸	H04N, H04L, G06F, G09C	۰/۰۲۶۸۱۳	C07D, G06F, G11B, G01R	۰/۰۰۵۵۰۸	H04N, H04L, G06F, G09C
۰/۰۲۹۴۸۶	H04N, G06F, G06K, G06T	۰/۰۰۲۰۶۲	H04M, H04L, H04Q, G08B	۰/۰۱۳۹۵۴	G06F, G06Q, G06K, B64F	۰/۰۲۴۲۷۹	H04N, G06F, G06K, G06T	۰/۰۰۵۴۶۸	G06F, G06Q, G06K, B64F

طبقه‌بندی بین‌المللی پتنت‌های مورد استفاده در« ادارهٔ ثبت اختراعات ایران» به شرح زیر ارائه شده است:

بخش الف) نیازهای انسانی. زیربخش‌ها: کشاورزی، مواد غذایی تنباکو، اقلام شخصی یا خانگی، سلامتی و بهداشت، سرگرمی و تفریحات.

بخش ب) انجام عملیات مختلف، حمل‌ونقل و ترابری. زیربخش‌ها: جداسازی، مخلوط‌کردن عملیات مختلف شکل دادن، چاپ و نشر، حمل‌ونقل، ترابری، فناوری ساختارهای ذره‌بینی، نانوفناوری.

بخش ج) شیمی متالورژی. زیربخش‌ها: شیمی، متالورژی.

بخش د) منسوجات، کاغذ. زیربخش‌ها: منسوجات یا سایر مواد قابل انعطاف که به‌صورت دیگری ارائه نمی‌شوند، کاغذ.

بخش ه) ساختمان‌ها و بناهای ثابت. زیربخش‌ها: ساختمان، حفاری زمین و صخره، استخراج معدن.

بخش و) مهندسی مکانیک، روشنایی، گرمایش و اسلحه‌ها و انفجار. زیربخش‌ها: انواع موتور یا پمپ، مهندسی به مفهوم عام آن، روشنایی، گرمایش، انواع اسلحه، انفجار.

بخش ز) فیزیک. زیربخش‌ها: وسایل، صنعت هسته‌ای.

بخش ح) برق.

همان‌طور که مشاهده می‌شود، انواع مختلفی از پتنت‌ها وجود دارد. در این پژوهش، به‌منظور اطمینان از کارایی مدل روی انواع پتنت، بدون توجه به دسته‌بندی‌های رایج، از انواع مختلف پتنت در حوزه‌های متفاوت استفاده شده که نتایج مغایرت خاصی را در روند اجرایی مدل نشان نداد. در نتیجه، این پژوهش روی انواع مختلف پتنت‌ها قابل اجرا خواهد بود.

٦. بحث و ارزیابی نتایج

در این پژوهش، برای سنجش صحت و اعتبار خروجی‌های به‌دست آمده، از روش قیاسی استفاده شده است. همان‌طور که در جدول ۸ دیده می‌شود، اولین و دومین خوشهٔ

دوتایی برای همهٔ روش‌ها مشابه هستند. برای محاسبهٔ میزان تشابه میان خروجی‌های روش‌های به‌دست آمده از شاخص «ژاکار» استفاده می‌کنیم. شاخص تشابه «ژاکار» راهی برای مقایسهٔ دو جامعه با نگاهی به اعضای هر دو جمعیت (تعداد کل اعضای مجزا) و اعضای حاضر در هر دو جمعیت (تعداد اعضای مشترک) است. شاخص «ژاکار» با استفاده از رابطهٔ ۶ محاسبه می‌شود. میزان تشابه در خوشه‌های دو تا پنج‌تایی در جدول ۱۲، ارائه شده است.

$$J(A,B)=\frac{|A\cap B|}{|A\cup B|} \qquad (۲)$$

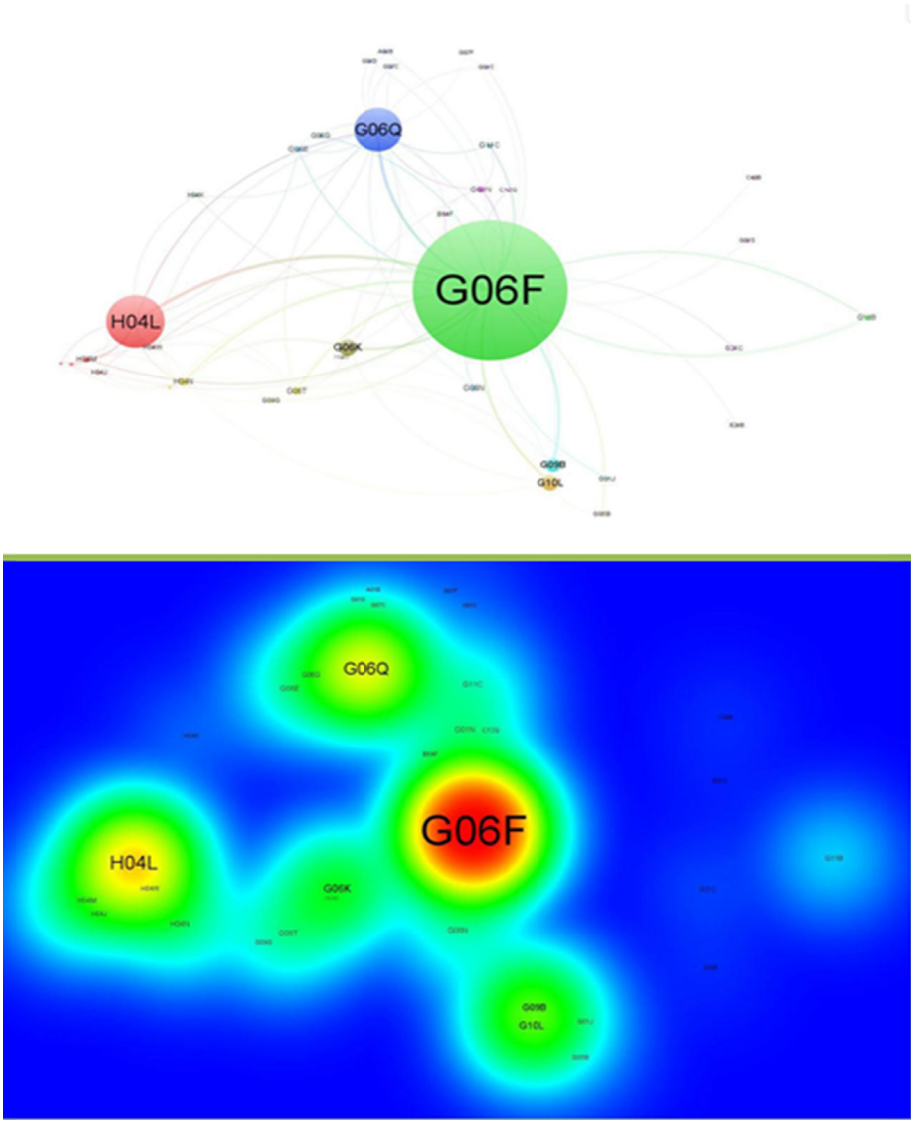
جدول ۱۲. میزان تشابه روش‌های پیشنهادی در خوشه‌های ۲ تا ۵تایی (بر حسب درصد)

سطح	روش‌ها	CBW_2	CBW_3
۲	CBW_1	۹۵	۹۶
	CBW_2	–	۹۲
۳	CBW_1	۹۴	۱۰۰
	CBW_2	–	۹۴
۴	CBW_1	۹۵	۱۰۰
	CBW_2	–	۱۰۰
۵	CBW_1	۱۰۰	۱۰۰
	CBW_2	–	۱۰۰

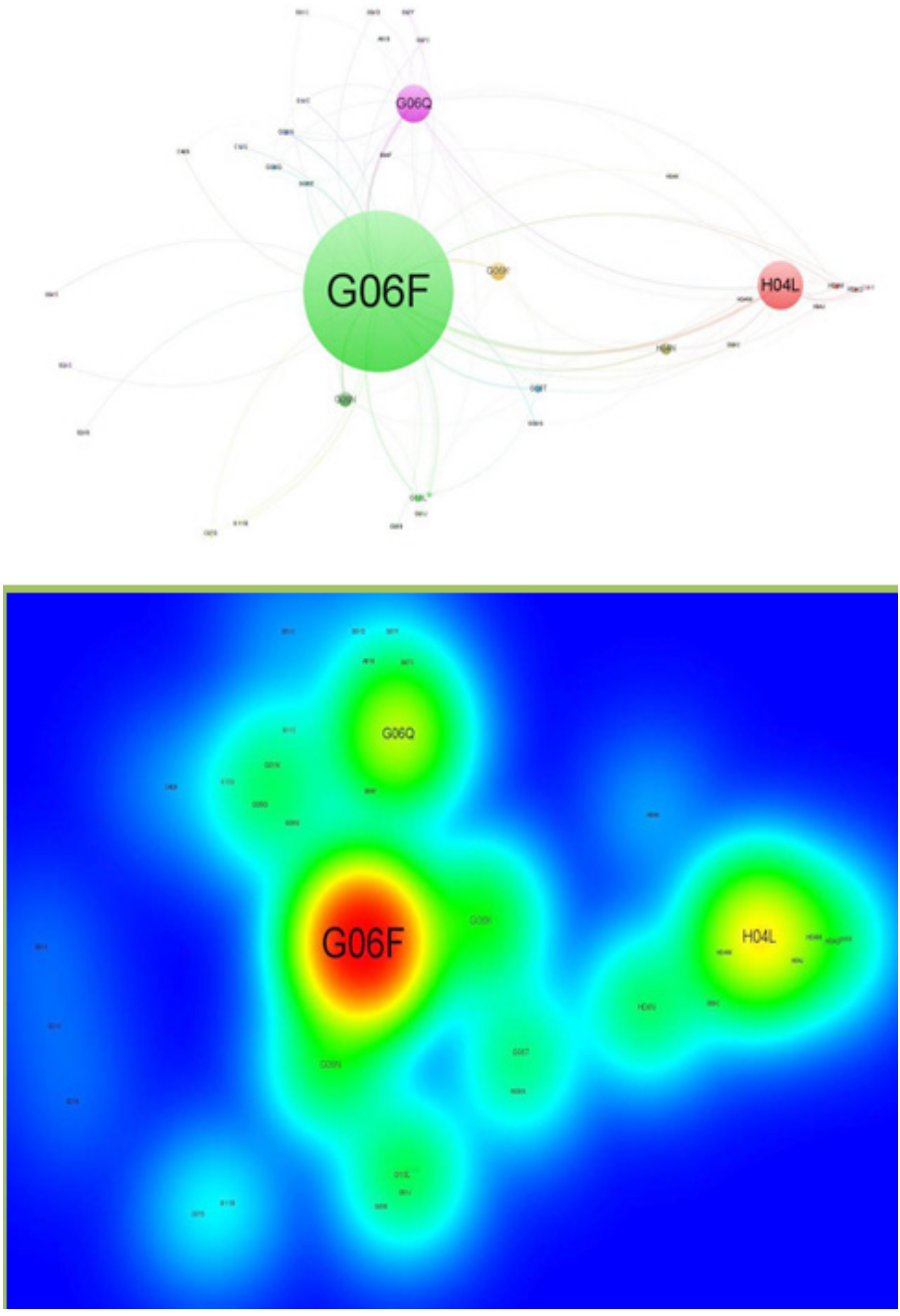
با توجه به نتایج به‌دست آمده، میزان مشابهت روش‌ها بسیار بالاست. تفاوت‌های موجود اغلب در اولویت‌بندی مجموعهٔ آیتم‌های مشتق شده و به‌دلیل تفاوت در وزن‌دهی هر یک از روش‌هاست. بنابراین، از تکرار مجموعهٔ آیتم‌ها در الگوریتم‌های مختلف می‌توان صحت و اعتبار وابستگی‌های به‌دست آمده را نتیجه گرفت.

یکی از محدودیت‌هایی که در مقالات پیشین ذکر شده، عدم امکان تجسم وابستگی‌ها و ارتباطات به‌دست آمده برای کاربر است (Altuntas, Dereli and Kusiak 2015b). تجسم وابستگی‌های به‌دست آمده به درک بهتر قوانین استخراج‌شده کمک می‌کند. در این پژوهش، ارتباطات استخراج‌شده بر اساس الگوریتم‌های پیشنهادی به‌صورت گراف و

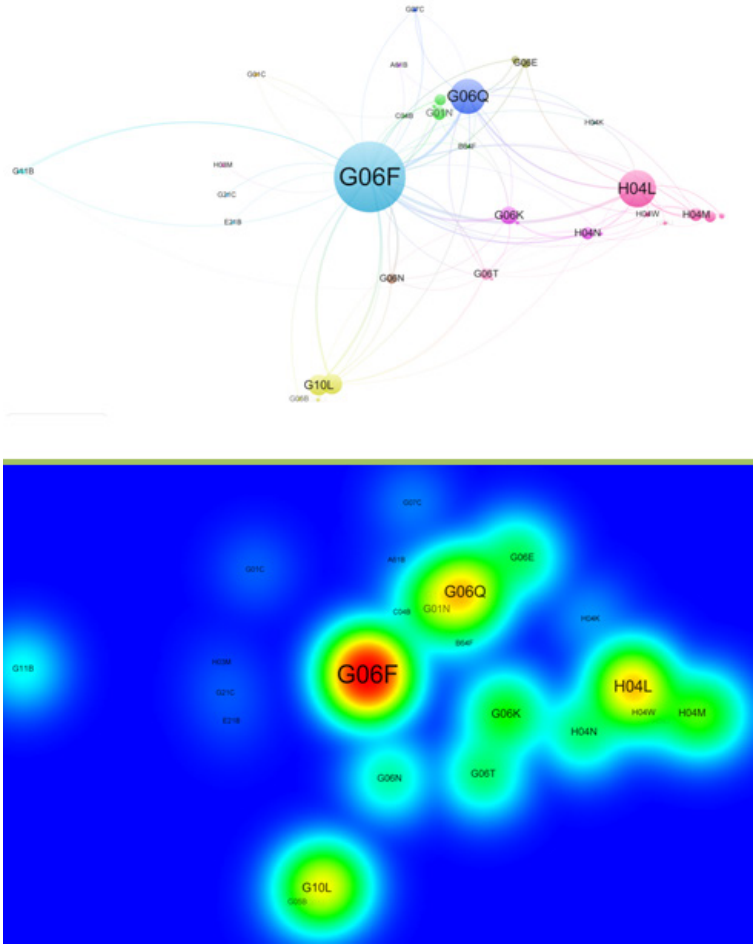
با استفاده از نرم‌افزار VOS Viewer در شکل‌های ۲، ۳ و ۴ ترسیم شده‌اند. تصویرسازی اطلاعات می‌تواند نقش مهمی در اکتشاف، شناسایی و تفسیر قوانین وابستگی به‌منظور توانمندساختن کاربران در درک داده‌هایی با ابعاد بالا و مقیاس بزرگ ایفا کند (Sekhavat and Hoeber 2013).



شکل ۲. فاصله و ناحیهٔ تحت پوشش کدهای IPC در الگوریتم‌های پیشنهادی CBW_1



شکل ۳. فاصله و ناحیهٔ تحت پوشش کدهای IPC در الگوریتم‌های پیشنهادی CBW_2



شکل ٤. فاصله و ناحیهٔ تحت پوشش کدهای IPC در الگوریتم‌های پیشنهادی CBW_3

٦. نتیجه‌گیری

در این مطالعه، از قوانین وابستگی وزن‌دار که از الگوریتم‌های محبوب داده‌کاوی است، برای شناسایی و ارزیابی خوشه‌های شامل فناوری‌های مرتبط بر اساس تحلیل پتنت استفاده شده است. در این پژوهش پنج الگوریتم مبتنی بر قوانین وابستگی وزن‌دار به‌کار گرفته شده است (BOWARM, WARM_2, CBW_1, CBW_2, CBW_3). خروجی این الگوریتم‌ها، فناوری‌های مرتبط با فناوری مورد نظر و همچنین، میزان ارتباطات آن‌ها را

تعییـن می‌نمایـد.

الگوریتم‌هـای پیشـنهادی ارتباطـات میـان فناوری‌هـا را بـا در نظـر گرفتـن اهمیـت نابرابـر فناوری‌هـا و همچنیـن پتنت‌هـا، به‌صـورت همزمـان به‌دسـت می‌آورنـد. نتایـج نشـان می‌دهـد کـه روش‌هـای پیشـنهادی، به‌طـور کلـی، توالی‌هـای یکسـان در یـک یـا چنـد مجموعـهٔ آیتـم اول در تمامـی سـطوح فراهـم می‌کننـد و مجموعـهٔ آیتم‌هـای باقی‌مانـده نیـز مشـابه هسـتند و تنهـا اولویت‌بندی‌هـای متفاوتـی دارنـد. تفـاوت در اولویت‌بندی‌هـا به‌دلیـل تفـاوت در نـوع و معیـار وزن‌دهـی اسـت. کارمنـدان و مدیـران می‌تواننـد از روش پیشـنهادی به‌منظـور یافتـن ارتباطـات میـان فناوری‌هـای مـورد نظـر اسـتفاده کننـد. کارآفرینـان ممکـن اسـت شـرکت‌هایی را کـه فناوری‌هـای مرتبـط بـا آن‌هـا را تولیـد می‌کننـد، شناسـایی کـرده و دریابنـد کـه ایـن شـرکت‌ها در حـوزهٔ فنـاوری مـورد نظـر در حـال انجـام چـه کاری هسـتند. سـرمایه‌گذاران، و همچنیـن سیاسـت‌گذاران می‌تواننـد اطلاعـات مربـوط بـه فناوری‌هـای جایگزیـن مرتبـط بـا فناوری‌هـای سـاخت خـود را بـا اسـتفاده از روش‌هـای پیشـنهادی به‌دسـت بیاورنـد. یـک پـروژهٔ سـرمایه‌گذاری ممکـن اسـت بـا بیـش از یـک فنـاوری منفـرد مرتبـط باشـد و نیـز ممکـن اسـت موفقیـت و پایـداری فنـاوری سرمایه‌گذاری‌شـده را در آینـده نشـان دهـد.

۷. پیشنهادات آتی

ارقـام اضافـی کدهـای پتنـت می‌توانـد بـرای یافتـن ارتباطـات در نظـر گرفتـه شـود. تحلیـل ادعاهـای پتنـت بـا اسـتفاده از قوانیـن وابسـتگی می‌توانـد انجـام شـود. روش‌هـای دیگـر ماننـد الگوریتم‌هـای مبتنـی بـر منطـق فـازی می‌تواننـد بـرای تحلیـل پتنـت به‌منظـور پیش‌بینـی ارتباطـات میـان تکنولوژی‌هـا بـه‌کار گرفتـه شـوند.

محدودیت‌هـای روش پیشـنهادی: آماده‌سـازی و پاکسـازی داده‌هـا آسـان نیسـت. نرم‌افـزاری کـه داده‌هـای مـورد نظـر را اسـتخراج کنـد و پردازش‌هـای لازم را انجـام دهـد، مـورد نیـاز اسـت. تعـدادی از مجموعـهٔ آیتم‌هـا تـا حـدودی بـزرگ هسـتند کـه ایـن امـر موجـب افزایـش تـلاش محاسـباتی می‌شـود.

همچنیـن، همان‌طـور کـه گفتـه شـد، از مزایـای راهبـردی تحلیـل پتنـت در فرایندهـای مدیریتـی می‌تـوان بـه توانایـی شناسـایی حوزه‌هـای تکنولوژیکـی بـرای سـرمایه‌گذاری اشـاره کـرد. پیش‌بینـی حوزه‌هـای تکنولـوژی در حـال ظهـور بـرای مدیـران دشـوار اسـت. در نتیجـه، سـرمایه‌گذاری روی تکنولوژی‌هـای جدیـد ریسـک بالایـی بـه همـراه دارد. بـرای کاهـش خطرات

این سرمایه‌گذاری، مدیران می‌توانند با بهره‌گیری از تحلیل و ارزیابی پتنت با استفاده از تکنیک‌های مختلف داده‌کاوی، پیش‌بینی دقیق‌تری از حوزه‌های تکنولوژی مناسب برای سرمایه‌گذاری داشته باشند. همچنین، با شناسایی خانواده و خوشه‌های تکنولوژی می‌توان وابستگی میان تکنولوژی‌ها را شناسایی کرده و ترجیحات سرمایه‌گذاری روی تکنولوژی‌های موجود در یک خوشه را درک نمود. بهره‌گیری از تکنیک‌های معرفی‌شده در این پژوهش می‌تواند نقشهٔ راهی در حوزه‌های تکنولوژیکی برای مدیران و شرکت‌ها فراهم کند. به‌علاوه، این اطلاعات در سطوح مختلف سازمانی و همچنین، در حوزه‌های مختلف فعالیت‌های پژوهشی مانند فعالیت‌های سیاست‌گذاری، پشتیبانی، زیربنایی و اجرای دستاوردهای پژوهشی کاربرد دارد.

فهرست منابع

علم‌خواه، حسن. ۱۳۹۱. *پتنت (حق اختراع)؛ مطالعه، جست‌وجو در منابع بین‌المللی و تحلیل*. تهران: مؤسسهٔ دارایی‌های فکری و فناوری مدرس.

غیاثی، فرناز، نوید نظافتی، و سجاد شکوهیار. ۱۳۹۴. خوشه‌بندی کاربران داده‌های دریایی با استفاده از تکنیک داده‌کاوی. *پژوهشنامهٔ پردازش و مدیریت اطلاعات* ۳۰ (۴): ۱۰۲۵-۱۰۴۹.

منصوری، علی، و فریده عصاره. ۱۳۹۳. بررسی ارزش منابع دانش بر اساس عامل اهمیت: («نگاه به آینده» و «نگاه به گذشته») مطالعهٔ موردی پروانه‌های ثبت اختراع. *پژوهشنامه پردازش و مدیریت اطلاعات* ۲۹ (۳): ۶۵۷-۶۸۴.

ناظمی، امیر، علی شماعی، و روح‌الله قدیری. ۱۳۹۲. شناسایی حوزه‌های نوظهور بر اساس تحلیل چرخهٔ عمر فناوری به‌عنوان نمونه فناوری زیردریایی‌های بدون سرنشین. *فصلنامه مدیریت توسعه فناوری* ۱ (۳): ۲۹-۴۸.

Altuntas, Serkan., and Turkay Dereli. 2015. A novel approach based on DEMATEL method and patent citation analysis for prioritizing a portfolio of investment projects. *Expert Systems with Applications* 42: 1003-1012.

_____, and A. Kusiak. 2015a. Forecasting technology success based on patent data. *Technological Forecasting and Social Change* 96: 202-214.

_____. 2015b. Analysis of patent documents with weighted association rules. *Technological Forecasting and Social Change* 92: 249-262.

Choi, C., S. Kim, and Y. Park. 2007. A patent-based cross impact analysis for quantitative estimation of technological impact: The case of information and communication technology. *Technological Forecasting and Social Change* 74 (8): 1296-1314.

Dalkey, Norman. C. 1971. An elementary cross-impact model. *Technological Forecasting and Social Change* 3: 341–351.

Dereli, Turkay, and Alptekin Durmusoglu. 2009. Classifying technology patents to identify trends:

applying a fuzzy-based clustering approach in the Turkish textile industry. *Technology in Society* 31 (3): 263–272.

Forsati, Rana., and Mohammad Reza Meybodi. 2010. Effective page recommendation algorithms based on distributed learning automata and weighted association rules. *Expert Systems with Applications* 37 (2): 1316-1330.

Gay, C., C. L. Bas, P. Patel, and K. Touach. 2005. The determinants of patent citations: an empirical analysis of French and British patents in the US. *Economics of Innovation and New Technology* 14 (5): 339-350.

Hahsler, M., G. Bettina, and H. Kurt. 2005. A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software* 14 (15): 1-25.

Hidalgo, Antonio., and Samuel Gabaly. 2012. Use of prediction methods for patent and trademark applications in Spain. *World Patent Information* 34 (1): 19–29.

Hipp, J., U. Güntzer, and G.R. Nakhaeizadeh. 2000. Algorithms for association rule mining–a general survey and comparison. *ACM sigkdd explorations newsletter* 2 (1): 58-64.

Jeong, Gi Ho, and Soung Hie Kim. 1997. A qualitative cross-impact approach to find the key technology. *Technological Forecasting and Social Change* 55 (3): 203-214.

Jun, Sunghae. 2011a. A forecasting model for technological trend using unsupervised learning. In *Database Theory and Application, Bio-Science and Bio-Technology*. Springer Berlin Heidelberg 51-60.

_____. 2011b. IPC Code Analysis of Patent Documents Using Association Rules and Maps–Patent Analysis of Database Technology. *In Database Theory and Application, Bio-Science and Bio-Technology*. 21-30. Springer Berlin Heidelberg.

Jun, S., S. S. Park, and D. S. Jang. 2012. Patent management for Technology Forecasting: A case study of the Bio-Industry. *Journal of Intellectual Properties Rights* 17 (6): 539-546.

Kim, C., H. Lee, H. Seol, and C. Lee. 2011. Identifying core technologies based on technological cross-impacts: An association rule mining (ARM) and analytic network process (ANP) approach. *Expert Systems with Applications* 38 (10): 12559-12564.

_____, S. Kim, and M. S. Kim. 2011. Identifying relationships between technology-based services and ICTs: a patent analysis approach. World Academy of Science, Engineering and Technology. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering* 5 (12): 1948-1952.

Kim, Jeeeun., and Sungjoo Lee. 2015. Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO. *Technological Forecasting and Social Change* 92: 332-345.

Kotsiantis, Sotiris., and Dimitris Kanellopoulos. 2006. Association rules mining: A recent overview. *GESTS International Transactions on Computer Science and Engineering* 32 (1): 71-82.

Lee, H., C. Kim, H. Cho, and Y. Park. 2009. An ANP-based technology network for identification of core technologies: A case of telecommunication technologies. *Expert Systems with Applications* 36 (1): 894-908.

_____, S., B. Yoon, C. Lee, and J. Park. 2009. Business planning based on technological capabilities: patent analysis for technology-driven road mapping. *Technological Forecasting and Social Change* 76 (6): 769–786.

Li, Guang-yuan, and Qin-bin Hu. 2011. A Framework for Weighted Association Rule Mining from Boolean and Fuzzy Data.In *International Conference on Internet Technology and Applications (iTAP)*: 1-4. IEEE. Wuhan: China.

_____, S., D. Zheng, and J. Shi. 2014. Technology forecasting of green energy using patent analysis and simulation modeling. *Advanced Materials Research* 875: 1619–1624.

Majidfar, F., F. Majidfar, and M. Tafazoli Shadpoor. 2007. T Four Steps Process Development and Software Designing for Knowledge Discovery in Patent Databases and Technology Forecasting. First Iran Data Mining Conference, Tehran: Iran.

Mattas, Nisha., and Mehrotra. Deepti. 2015. Comparing data mining techniques for mining patents. ACCT '15 Proceedings of the 2015 *Fifth International Conference on Advanced Computing & Communication Technologies*: 217-221. Haryana: India.

Noa, Hyun Joung and Yongtae Park. 2006. Trajectory patterns of technology fusion: Trend analysis and taxonomical grouping in nanobiotechnology. *Technological Forecasting and Social Change* 77 (1): 63-75.

Sekhavat, Yoones A., and Orland Hoeber. 2013. Visualizing association rules using linked matrix, graph, and detail views. *International Journal of Intelligence Science* 3 (1): 34-49.

Tao, F., F., Murtagh, and M., Farid. 2003. Weighted association rule mining using weighted support and significance framework. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*: 661-666. ACM. Washington, D.C: USA.

Thorleuchter, D., D. V. D. Poel, and A. Prinzie. 2007. A compared R&D-based and patent-based cross impact analysis for identifying relationships between technologies. *Technological forecasting and social change* 77 (7):1037-1050.

Trappey, C. V., H. Y. Wua, F. T. Dutta, and A. J. C. Trappey. 2011. Using patent data for technology forecasting: China RFID patent analysis. *Advanced Engineering Informatics* 25 (1): 53-64.

Wang, W., J. Yang, and S. Y. Philip. 2000. Efficient mining of weighted association rules (WAR). In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*: 270-274. Boston, Massachusetts: USA.

Wu, F. S., C. C. Hsu, P. C. Lee, and H. N. Su. 2011. A systematic approach for integrated trend analysis—The case of etching. *Technological Forecasting and Social Change* 78 (3): 386–407.

Yoon, Byungun., and Yongtae Park. 2004. A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research* 15 (1): 37-50.

لیلا محمودجانلو

متولد سال ۱۳۶۸، دارای مدرک تحصیلی کارشناسی ارشد رشتهٔ مدیریت نوآوری و فناوری از دانشگاه علوم و فنون مازندران است. وی دانش‌آموختهٔ کارشناسی مهندسی صنایع دانشگاه صنعتی نوشیروانی بابل است.

مباحث مرتبط با داده‌کاوی، برنامه‌نویسی، تحلیل پتنت، سازماندهی دانش از جمله علایق پژوهشی وی است.

بابک شیرازی

متولد سال۱۳۵۷، دارای مدرک تحصیلی دکتری مهندسی صنایع در حوزهٔ تولید با فناوری اتوماتیک است. ایشان هم‌اکنون استادیار گروه مهندسی صنایع دانشگاه علوم و فنون مازندران است.

فناوری‌های نوین مهندسی تولید، سیستم‌های اطلاعاتی پیشرفته و آینده‌پژوهی فناوری از جمله علایق پژوهشی وی است.

ایرج مهدوی

استاد تمام گروه مهندسی صنایع دانشگاه علوم و فنون مازندران است. وی دورهٔ دکتری را در زمینهٔ مهندسی تولید در کشور هند سپری کرده است. ایشان بیش از ۳۰۰ مقالهٔ علمی در زمینه‌های تولید سلولی، زنجیرهٔ تأمین، برنامه‌ریزی تولید، شبکه‌های فازی، مدیریت دیجیتال و مدیریت عملیات هوشمند در مجلات و کنفرانس‌های بین‌المللی منتشر کرده‌اند. عضویت در هیئت تحریریهٔ ۵ مجله معتبر بین‌المللی، پژوهشگر برتر کشور در گروه فنی و مهندسی، استاد نمونهٔ کشوری از دیگر افتخارات ایشان است.

جواد سلطان‌زاده

متولد سال ۱۳۶۴، دانشجوی دکتری مدیریت تکنولوژی دانشگاه علامه طباطبائی است. وی دانش‌آموختهٔ کارشناسی مهندسی صنایع دانشگاه علم و صنعت ایران و کارشناسی ارشد مدیریت تکنولوژی دانشگاه علامه طباطبائی است.

مدیریت فناوری و نوآوری، سیاست‌گذاری نوآوری و نظام نوآوری از علایق پژوهشی وی است.