

## تحلیل پتنت با استفاده از داده کاوی 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](mailto:Leila_janlu@yahoo.com)

بابک شیرازی \*

دکتری مهندسی صنایع

استادیار؛ دانشگاه علوم و فنون مازندران؛ [Shirazi\\_b@yahoo.com](mailto:Shirazi_b@yahoo.com)

ایرج مهدوی

دکتری مهندسی صنایع

استاد؛ دانشگاه علوم و فنون مازندران؛ [Irajarash@rediffmail.com](mailto:Irajarash@rediffmail.com)

جواد سلطانزاده

دانشجوی دکتری مدیریت فناوری؛ دانشکده مدیریت و حسابداری  
دانشگاه علامه طباطبائی؛ [Soltanzadeh921@atu.ac.ir](mailto:Soltanzadeh921@atu.ac.ir)

۱۳۹۶/۰۶/۲۶

۱۳۹۵/۱۲/۹

فصلنامه علمی پژوهشی  
پژوهشگاه علوم و فناوری اطلاعات ایران  
شاپا (چاپی) ۲۲۲۳-۲۲۵۱  
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱  
نمایه در SCOPUS، LISA و ISC  
<http://jlist.irandoc.ac.ir>  
دوره XX | شماره X | صص XX-XX  
۱۳XX X

نوع مقاله: پژوهشی

**چکیده:** تجزیه و تحلیل ارتباطات میان فناوری‌ها می‌تواند بینشی برای اتخاذ استراتژی‌های فناورانه و به حداکثر رساندن سود فراهم کند. پتنت‌ها، بخشی جدایی‌ناپذیر از حقوق مالکیت فکری هستند و اطلاعات قابل توجهی درباره فناوری‌های توسعه یافته را دربرمی‌گیرند. با توجه به حجم فزاینده پتنت‌ها، روش داده کاوی 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 براساس تحلیل پتنت است، که وابسته به دانش ذهنی متخصصان یک حوزه خاص فناوری نباشد. برای دستیابی به این هدف یک فرآیند چهار مرحله‌ای شامل جمع‌آوری داده، تشکیل پایگاه داده ساخت یافته، اجرای داده کاوی و شناسایی/تعیین میزان ارتباطات میان فناوری‌ها/ترسیم گرافیکی آن ارائه شده است. فرآیند پیشنهادی با یک مطالعه موردی اجرا و ارزیابی شده است.

**کلیدواژه‌ها:** تحلیل پتنت، داده کاوی، قوانین وابستگی وزن‌دار، خوشه‌های فناوری، کد IPC، فناوری‌های مرتبط.

\* پدیدآور رابط [Shirazi\\_b@yahoo.com](mailto:Shirazi_b@yahoo.com)

به این مقاله به شکل زیر استناد کنید:

**دورن متن:**

(محمودجانلو و دیگران، زودآیند)

**در فهرست منابع:**

محمودجانلو، لیلا، بابک شیرازی، ایرج مهدوی، جواد سلطانزاده. زودآیند. تحلیل پتنت با استفاده از داده کاوی برای شناسایی و تعیین ارتباطات میان فناوری‌ها. پژوهشنامه پردازش و مدیریت اطلاعات. (دسترسی در <http://Jipm.irandoc.ac.ir>)

۱. مقدمه

پتنت‌ها یکی از مهم‌ترین منابع نوآوری و همچنین ابزاری مفید برای تحلیل‌های فنی-اقتصادی به شمار می‌آیند و در مطالعات گسترده‌ای مورد استفاده قرار گرفته‌اند، از جمله: ترسیم منحنی عمر فناوری‌ها، پیش‌بینی فناوری‌های نوظهور، مطالعات توسعه فناوریانه، تجزیه و تحلیل روندهای نوآوری، ارتقای برنامه‌ریزی استراتژیک (ناظمی، شماعی و قدیری ۱۳۹۱، Altuntas and Derehli 2015، ارتقای برنامه‌ریزی استراتژیک (ناظمی، شماعی و قدیری ۱۳۹۱، Altuntas, Turkey and Kusiak 2015a، Li, Zheng and Shi 2014، 2015، Kim and Sungjioo 2015). پتنت‌ها، یکی از بهترین شاخص‌ها برای ارزیابی فعالیت‌های فناوریانه هستند (Mattas and Deepti 2015). بنابراین با تحلیل پتنت می‌توان به اطلاعات ارزشمندی پیرامون روند و مسیر تکامل آن‌ها، فعالان اصلی، دانش‌های نهفته و همچنین ارتباطات میان فناوری‌ها دست یافت. محققان پس از درک اهمیت پتنت در فهم بیشتر آن تلاش‌هایی کردند، به طوری که پس از توافق‌نامه استراسبورگ در سال ۱۹۷۱ و به منظور طبقه‌بندی پتنت‌ها، سیستم بین‌المللی طبقه‌بندی پتنت (IPC) توسعه یافت. با در دسترس بودن این طبقه‌بندی، جست‌وجو و استخراج اطلاعات مرتبط با پتنت‌ها آسان شده‌است. براساس این سیستم، به هر پتنت حداقل یک کد IPC با توجه به دامنه فناوری آن، اختصاص داده شده‌است (علم خواه ۱۳۹۱). هر چند این طبقه‌بندی توانست کاربرد پتنت‌ها و تحلیل برآمده از آن‌ها را افزایش دهد؛ اما، با توجه به حجم فزاینده پتنت‌ها، یک روش تحلیل مناسب برای مجموعه داده‌های بزرگ ضروری به نظر می‌رسد. از این رو، استفاده از روش‌های داده‌کاوی در تحلیل پتنت‌ها پیشنهاد شده‌است. این روش مجموعه‌ای از فنون است که به شخص امکان می‌دهد تا ورای داده‌پردازی معمولی حرکت کند. داده‌کاوی، سازمان‌ها را قادر می‌سازد تا از سرمایه داده‌هایشان به‌درستی بهره‌برداری نمایند و از این ابزار برای پشتیبانی فرآیند تصمیم‌گیری استفاده کنند (غیائی، نظافتی و شکوهمیار ۱۳۹۴). یکی از پرکاربردترین تکنیک‌ها در داده‌کاوی، کاوش قوانین وابستگی است. کاوش قوانین وابستگی یک روش محبوب و متداول داده‌کاوی براساس احتمال شرطی است که ارتباطات میان اهداف مختلف را استخراج می‌کند (Hipp et al. 2015; Kotsiantis and Kanellopoulos 2006; Forsati and Meybodi 2010).

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

<sup>1</sup> International Patent Classification

توسط الگوریتم های پیشنهادی به دست آمده اند.

تمامی پتنت ها دارای اهمیت یکسان نیستند و همچنین سطوح فناوری (در اینجا کدهای IPC) دارای وزن های متفاوت هستند.

بنابراین، الگوریتم های مبتنی بر قوانین وابستگی وزن دار شامل کاوش قوانین وابستگی وزن دار بولین (BOWARM)<sup>1</sup>، کاوش قوانین وابستگی وزن دار (WARM\_2)<sup>2</sup> و سه الگوریتم پیشنهادی، کاوش قوانین وابستگی وزن دار ترکیبی 1، 2 و 3 (CBW\_1, CBW\_2, CBW\_3)<sup>3</sup>، اجرا و نتایج به دست آمده مقایسه شده اند. استناد، یکی از شاخص های مهم در تعیین اهمیت و تاثیر یک پتنت به حساب می آید (منصوری و عصاره ۱۳۹۳). همچنین تعداد زیاد پتنت های تحت پوشش یک کد IPC نشان دهنده ی جذابیت بالای فناوری ارائه شده توسط این کد است (Altuntas, Dereli, and Kusiak 2015b). در الگوریتم های پیشنهادی، وزن و اهمیت متفاوت پتنت ها و همچنین سطوح فناوری به صورت هم زمان و با استفاده از تعداد استنادهای پتنت و تعداد پتنت های موجود در کدهای IPC در نظر گرفته شده است. در این پژوهش برای آماده سازی داده ها به جای بهره گیری از نرم افزارهای تحلیل پتنت که خریداری و استفاده از آنها اغلب مستلزم هزینه ی بسیار است، یک نرم افزار به منظور استخراج اطلاعات طراحی شده و هزینه های این قسمت با وجود خروجی های مشابه به طور قابل توجهی کاهش یافته است. خروجی الگوریتم های مورد استفاده به پرسش هایی از قبیل موارد زیر پاسخ می دهد: فناوری مورد نظر با کدام فناوری ها مرتبط است؟ کدام فناوری ها دارای ارتباطات قوی تری هستند؟ و میزان ارتباطات آنها نسبت به هم چگونه است؟

به بیان دقیق تر، هدف این پژوهش، یافتن روشی پایدار جهت استخراج فناوری های مرتبط براساس تحلیل پتنت است، که راهنمایی برای مدیران به منظور شناسایی حوزه های فناوری، ارتباط میان فناوری ها و کاهش خطرات ناشی از سرمایه گذاری باشد. برای دستیابی به این هدف، روش پژوهش در قالب یک فرآیند چهار مرحله ای ارائه شده است. مرحله نخست، جمع آوری داده، شامل جمع آوری اسناد پتنت از پایگاه داده پتنت با توجه به فناوری های مورد نظر. در مرحله دوم، با تشکیل پایگاه داده ساخت یافته، اطلاعات مورد نیاز از پتنت های بازبایی شده به دست می آیند. برای پیاده سازی فرآیند ذکر شده، یک سیستم نرم افزاری با استفاده از زبان برنامه نویسی PHP طراحی و

<sup>1</sup> Boolean weighted association rule mining

<sup>2</sup> Weighted association rules mining\_2

<sup>3</sup> Combination of BOWARM & WARM\_2

اجرا شده است. زیرا، استفاده از نرم افزارهای رایج تحلیل پتنت مستلزم هزینه‌ی نسبتاً بالایی است. مرحله سوم، اجرای داده کاوی است. در واقع، برای شناسایی و تعیین میزان ارتباطات میان فناوری-های مختلف، الگوریتم‌های قوانین وابستگی وزن دار پیشنهاد شده‌اند و سه الگوریتم ابتکاری نیز ارائه شده است. در الگوریتم‌های پیشنهادی اهمیت نابرابر پتنت‌ها و نیز سطوح فناوری از نظر تاثیر فناوریانه و اهمیت تجاری به صورت همزمان در نظر گرفته شده است. مرحله چهارم، ترسیم گرافیکی، ارتباطات به دست آمده به صورت گرافیکی نمایش داده می‌شوند تا تجسم بهتری برای کاربران فراهم شده و تحلیل و ارزیابی نتایج بدست آمده تسهیل شود.

انجام این پژوهش برای مدیران و تصمیم گیران عرصه فناوری به ویژه فناوری‌های نوین مفید خواهد بود؛ زیرا، مستندات تحلیل پتنت اطلاعات ارزشمندی ارائه می‌دهند، از جمله: اجتناب از دوباره کاری (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 سرمایه گذاری، شناسایی اختراعاتی آزمایش شده به عنوان رویدادهای جدید، بررسی پژوهش‌های گذشته، پیدا کردن راه حل برای مشکلات که این یافته‌ها در کنار هم، افق دید بازاری را برای مدیران و کارکنان تحقیق و توسعه فراهم می‌آورد. ادامه مقاله به شرح زیر است. مبانی نظری و مطالعات پیشینی ارتباطات میان فناوری‌ها و کاربردهای داده کاوی در تحلیل پتنت در بخش دوم بررسی می‌شوند. 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. نتیجه گیری، پیشنهادات آتی و محدودیت‌های پژوهش در بخش ششم ارائه شده است.

## ۲. مبانی نظری و مطالعات پیشینه پژوهش

### تحلیل پتنت با استفاده از داده کاوی

محققان رویکردهای داده کاوی از قبیل قوانین وابستگی، کاوش متنی، خوشه بندی، طبقه بندی و تحلیل سری های زمانی را برای پتنت‌ها به کار گرفته‌اند ( Wu et al. 2011; Lee et al. 2009; ) و مطالعات (Trappey et al. 2011; Dereli and Durmusoglu 2009; Hidalgo and Gabaly 2012). محدودی از قوانین وابستگی برای تحلیل اسناد پتنت استفاده کرده اند. «هاشلر»<sup>۱</sup> و همکاران<sup>۱</sup> قوانین

<sup>1</sup> Hashler et al

وابستگی را به منظور تحلیل پتنت به کار گرفتند. آن‌ها مجموعه‌ای شامل قوانین وابستگی ارائه دادند که زیرساخت‌های اساسی را برای ایجاد و تغییر مجموعه داده‌های ورودی و تجزیه و تحلیل مجموعه داده‌ها و قوانین فراهم می‌کند (Hahsler, Bettina and Kurt 2005). مجیدفر و همکارانش با استفاده از تجزیه و تحلیل مورفولوژیک پتنت‌ها، فرآیندی چهار مرحله‌ای به صورت مدول‌های نرم‌افزاری مستقل داده کاوی به منظور استخراج فناوری‌های جدید و بالقوه ارائه دادند و به این نتیجه رسیدند که روش پیشنهادی، پتانسیل پردازش سیستماتیک داده‌های موجود در متون ساخت-نیافته پتنت‌ها و استفاده از این داده‌ها در پیش‌بینی حوزه‌های جدید فناوری را داراست (Majidfar, Majidfar and Shadpoor 2007). «جون»<sup>1</sup> 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 توسعه نیافته) را پیش‌بینی کرد (Jun 2011a). «کیم» و همکاران کاوش قوانین وابستگی، فرآیند 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» و همکاران از کاوش قوانین وابستگی و سری‌های زمانی برای پیش‌بینی روند بیوفناوری<sup>2</sup> استفاده کردند (Jun, Park and Jang 2012). «وانگ» و همکاران<sup>3</sup> 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 حوزه وزنی خاص، به هر آیتی نسبت داده شوند (Wang, Jiong and Philip 2000). آلتونتاس و همکاران در سال ۲۰۱۵ داده‌های پتنت را توسط نرم افزار 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 داده شده براساس قضاوت خبرگان بوده است. در این پژوهش، الگوریتم قوانین

<sup>1</sup> Jun

<sup>2</sup> Biotechnology

<sup>3</sup> Wang et al

وابستگی وزن‌داری که اهمیت نابرابر پتنت‌ها و همچنین سطوح فناوری را به صورت توامان در نظر گرفته و وابسته به دانش ذهنی متخصصان یک حوزه‌ی خاص فناوری نباشد، برای یافتن ارتباطات میان فناوری‌ها و شناسایی خوشه‌های شامل فناوری‌های مرتبط به کار گرفته شده‌است.

### ارتباطات میان فناوری‌ها

مرور مبانی نظری نشان می‌دهد، برای بررسی ارتباطات میان فناوری‌های مختلف معمولاً از تکنیک تحلیل اثر متقابل (CIA)<sup>۱</sup> استفاده شده‌است. تحلیل اثر متقابل با استفاده از یک ماتریس، روابط علی و معلولی بین وقایع را نشان می‌دهد و شاخص اثر متقابل (A,B) به صورت احتمال شرطی  $P(B|A)$  تعریف می‌شود. جزئیات تکنیک تحلیل اثر متقابل را می‌توان در پژوهش «دالکی» (Dalkey 1971) مطالعه کرد. «جئونگ» و «کیم»<sup>۲</sup> در سال ۱۹۹۷ یک مدل اثر متقابل براساس منطق فازی، برای تجزیه و تحلیل کیفی فناوری ارائه کردند (Jeong and Kim 1997). «یون» و «پارک»<sup>۳</sup> در سال ۲۰۰۴ با استفاده از یکپارچه‌سازی متن کاوی و تحلیل شبکه به توضیح فرآیند کلی توسعه شبکه پتنت و نیز ارائه شاخص‌های جدید مانند مرکزیت فناوری، دوره‌ی فناوری و خوشه‌های کلیدواژه فناوری جهت تحلیل کمی عمیق پرداخته و به این نتیجه رسیدند که روش پیشنهادی می‌تواند در تحلیل به روزرسانی روند فناوری‌های پیشرفته و شناسایی راه‌های جدید برای توسعه محصول به کار گرفته شود (Yoon and Park 2004).

«چوی» و همکاران<sup>۴</sup> در سال ۲۰۰۷ به جای تکیه بر درک‌های بصری و کیفی که در مدل‌های متعارف تحلیل اثر متقابل استفاده می‌شود، یک متدولوژی مبتنی بر پتنت برای اندازه‌گیری اثرات متقابل به کار گرفتند. آن‌ها داده‌ها را از اداره پتنت و علائم تجاری ایالات متحده (USPTO)<sup>۵</sup> با استفاده از یک کد برنامه‌نویسی شده، جمع‌آوری و یک شبکه اثر متقابل برای تحلیل روابط پیچیده بین فناوری‌ها ارائه کردند (Choi, Kim and Park 2007). «لی» و همکاران<sup>۶</sup> در سال ۲۰۰۹ رویکرد فرآیند تحلیل شبکه‌ای (ANP)<sup>۷</sup> را برای یافتن اثر متقابل بین فناوری‌ها پیشنهاد کردند (Lee

<sup>1</sup> Cross impact analysis

<sup>2</sup> Jeong and Kim

<sup>3</sup> Yoon and Park

<sup>4</sup> Choi et al

<sup>5</sup> United States Patent and Trademark Office

<sup>6</sup> Li et al

<sup>7</sup> Analytical Network Process

(et al, 2009). «تورلیوچتر» و همکارانش<sup>۱</sup> در سال ۲۰۱۰ در پژوهشی با استفاده از رویکرد تحلیل اثرمتقابل کمی، روشی برای شفاف تر کردن اثر فناوری ها و تخمین اثر متقابل فناوری ها پیشنهاد کردند (Thorleuchter, Poel and Prinzie 2007). «نو» و «پارک»<sup>۲</sup> در سال ۲۰۱۰ از اعداد استناد جهت توسعه مسیرهای ادغام فناوری برای نانوبیوفناوری<sup>۳</sup> استفاده کردند (No and Park 2006). «کیم» و همکاران<sup>۴</sup> در سال ۲۰۱۱ با استفاده از تحلیل استنادی پتنت، تحلیل هم طبقه ای پتنت<sup>۵</sup> و تحلیل استنادی سری های زمانی به بررسی ارتباطات فناورانه میان خدمات مبتنی بر فناوری و فناوری های اطلاعات و ارتباطات (ICT) پرداختند (Kim.C, Kim.S and Kim 2011).

### ضرورت و اهداف پژوهش

اطلاعات به دست آمده از تحلیل پتنت در مطالعات مرتبط با توسعه فناورانه، شناسایی روندهای فناورانه، شناسایی پتانسیل بالقوه فناوری ها و تصمیم گیری های مدیریتی به منظور شناسایی حوزه های فناورانه جذاب برای سرمایه گذاری به کار می رود. از آنجایی که، حجم این داده ها بسیار زیاد بوده و نیز در حال افزایش است، استفاده از تکنیک های داده کاوی در تحلیل اطلاعات پتنت پیشنهاد شده است. روش داده کاوی پیشنهادی چارچوبی برای یافتن ارتباطات میان فناوری های مختلف ارائه می دهد و می تواند توسط مدیران و کارکنان سازمان ها به منظور کاهش ریسک مورد استفاده قرار گیرد. از پرکاربردترین تکنیک ها در داده کاوی، الگوریتم قوانین وابستگی است. از آنجایی که، قوانین وابستگی سنتی قادر به در نظر گرفتن اهمیت نابرابر پتنت ها و سطوح مختلف فناوری نیستند، این پژوهش ضمن بهره گیری از الگوریتم های قوانین وابستگی وزن دار، الگوریتم های جدیدی، توسعه یافته از الگوریتم های وزن دار رایج، ارائه می دهد.

در نتیجه، با استفاده از الگوریتم های پیشنهادی، سازمان ها می توانند شرکت های تولیدکننده فناوری را مورد بررسی قرار داده و دریابند که این شرکت ها در حوزه های فناوری مورد نظر، به دنبال چه هستند. با توجه به قابلیت تصویرسازی، درک و تحلیل ارتباطات کشف شده تسهیل شده

<sup>1</sup> Thorleuchter et al

<sup>2</sup> No and Park

<sup>3</sup> Nano Biotechnology

<sup>4</sup> Kim et al

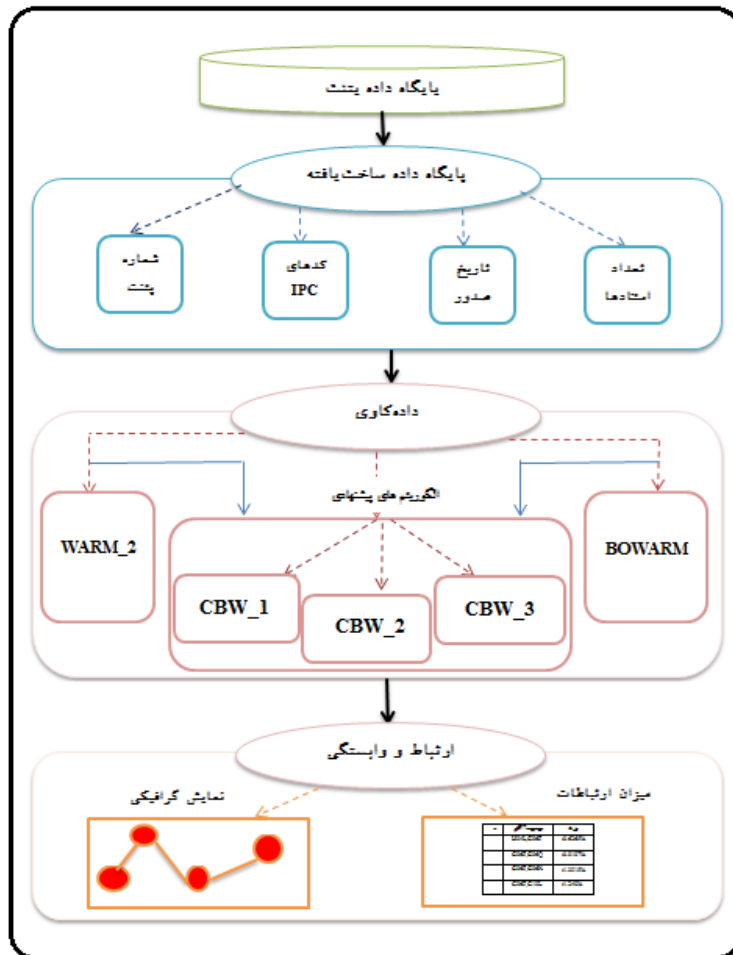
<sup>5</sup> Co-classification

<sup>6</sup> Information and Communication Technologies

است. در نتیجه، مدیران با شناخت دقیق تری از فناوری‌ها می‌توانند برای سرمایه‌گذاری‌های جدید تصمیم‌گیری نمایند.

### ۳. مراحل پژوهش

در این مقاله هر کد IPC به عنوان یک آیتم، هر سند پتنت به عنوان یک تراکنش در نظر گرفته شده‌است. 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 شده‌است.



شکل ۱. چارچوب کلی پژوهش



همان طور که در شکل ۱ دیده می شود، ابتدا اطلاعات پتنت شامل شماره پتنت، کدهای IPC، تاریخ صدور و تعداد استنادها از پایگاه داده پتنت به دست آورده می شود، سپس قوانین وابستگی وزن دار مبتنی بر الگوریتم های داده کاوی برای یافتن وابستگی های میان فناوری ها استفاده شده است. مراحل پژوهش به شرح زیر می باشد:

### آماده سازی داده ها

اسناد پتنت مرتبط با فناوری تئوری پایگاه داده و کاربردهای آن (DTA)<sup>۱</sup> از پایگاه داده پتنت آمریکا (USPTO) موجود در سایت «فری پتنتس آنلاین»<sup>۲</sup> با استفاده از معادله زیر استخراج شده اند.

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).

به این ترتیب ۱۰۰۲۷۰ پتنت بازیابی شدند. در این مرحله با استفاده از زبان برنامه نویسی PHP، یک نرم افزار هوشمند (ربات) برای پردازش داده ها طراحی شده است. این نرم افزار داده های پتنت را بررسی کرده و اطلاعات مورد نیاز (شماره پتنت، تاریخ صدور، تعداد استنادها، کدهای IPC) را تهیه و ذخیره سازی می کند. خروجی این مرحله، در قالب یک پایگاه داده ساخت یافته به عنوان ورودی مرحله داده کاوی در نظر گرفته شده است. در این مطالعه ۸۸۰ پتنت با توجه به مفروضات در نظر گرفته شده، بازیابی شدند. اما تنها ۲۲۹ پتنت پس از حذف پتنت های متعلق به یک کلاس IPC یا نامرتب به DTA در نظر گرفته شدند. اگر پتنتی متعلق به یک کلاس IPC باشد، در نظر گرفته نمی شود، زیرا برای تعیین وابستگی میان فناوری ها، حداقل دو کد IPC مورد نیاز است. پتنت های بازیابی شده در این مطالعه پیش از ۲۰ جولای ۲۰۱۲ صادر شده اند. گاهی اوقات باید از صدور پتنت ها زمانی سپری شود (برای مثال، حدود ۴ سال) تا ارجاعات و استنادهای اولیه انجام شود (Gay et al, 2005). در جدول ۱ برشی از پایگاه داده ساخت یافته ارائه شده است.

<sup>1</sup> Database theory and its applications

<sup>2</sup> www.freepatentsonline.com

جدول ۱. پایگاه داده ساخت یافته

شماره پتنت	تاریخ ثبت پتنت	تعداد اسنادها	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

#### داده کاوی با الگوریتم‌های قوانین وابستگی وزن دار

پس از آماده‌سازی داده‌ها، فرآیند داده کاوی و اجرای الگوریتم‌های قوانین وابستگی آغاز می‌شود. داده کاوی 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 برنامه نویسی را فراهم می‌کند. در ادامه الگوریتم‌های قوانین وابستگی وزن دار مورد استفاده، در قالب یک مثال ارائه شده است.

#### ۱. کاوش قوانین وابستگی وزن دار بولین (BOWARM)

قوانین وابستگی وزن دار بولین نخستین بار توسط «لی» و «هو»<sup>۱</sup> در سال ۲۰۱۱ و به منظور ارائه چارچوبی جدید برای در نظر گرفتن وزن‌ها در الگوریتم قوانین وابستگی معرفی شد (Li and Hu 2011). این الگوریتم با در نظر گرفتن وزن آیت‌ها، ارزش قوانین را محاسبه می‌کند. دو تعریف کلیدی در این الگوریتم عبارت‌انداز:

وزن مجموعه آیت هر تراکنش (ITW)<sup>۲</sup>: مجموع وزن همه‌ی آیت‌های موجود در مجموعه

<sup>۱</sup> Li and Hu

<sup>۲</sup> Itemset transaction weight

آیتم تقسیم بر تعداد کل آیتم‌ها (Li and Hu 2011).  
پشتیبانی وزن دار (WS<sup>1</sup>): مجموع ITW تمام پتنت‌های شامل مجموعه آیتم مورد نظر تقسیم بر  
تعداد کل پتنت‌ها (Li and Hu 2011).  
چارچوب پیشنهادی با بهره‌گیری از الگوریتم BOWARM به شرح زیر می‌باشد:

#### الف) محاسبه وزن سطوح فناوری (آیتم‌ها)

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

جدول ۲. لیست پتنت‌ها (تراکش‌ها)

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 می‌باشد. 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، چهار پتنت وجود دارند که تحت کد G06F طبقه‌بندی شده‌اند (US1,US2,US3,US4). برای محاسبه وزن نرمال شده هر کد IPC، تعداد پتنت‌های موجود در کد IPC مورد نظر بر تعداد کل پتنت‌های بازیابی شده تقسیم می‌شود، برای مثال (0.51=191/377). بنابراین، وزن اختصاص یافته به کد G06F، 0.51 است. وزن‌ها بین 0 تا 1 تغییر می‌کنند و مجموع آن‌ها یک است. در جدول ۳ یک نمونه پایگاه داده آیتم وزن دار نشان داده شده است.

<sup>1</sup> Weighted support

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

ردیف	آیتم	تعداد پنت‌ها	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 محاسبه وزن هر آیتم، وزن مجموعه آیتم‌های محتمل (ITW) محاسبه می‌شود. به عنوان مثال، ITW برای مجموعه آیتم (G06F, H04N) با توجه به جدول ۳ با استفاده از معادله ۱ محاسبه می‌شود:

$$ITW(G06F, H04N) = \frac{\{(G06F) + (H04N)\}}{2} = \frac{0.51 + 0.05}{2} = 0.28 \quad (1)$$

#### ب) محاسبه ارزش قوانین وابستگی یا میزان ارتباط در خوشه‌های فناوری

پس از محاسبه وزن هر مجموعه آیتم، برای به دست آوردن ارزش قوانین وابستگی، وزن مجموعه آیتم در تعداد پنت‌های شامل مجموعه آیتم ضرب شده و بر تعداد کل پنت‌ها تقسیم می‌شود. قدرت وابستگی بین مجموعه آیتم‌ها در این الگوریتم با استفاده از ارزش WS شناخته می‌شود. برای مجموعه آیتم (G06F, H04N) با توجه به جدول ۲، محاسبات به شرح معادله ۲ می‌باشد:

$$WS(G06F, H04N) = \frac{\{(ITW(G06F, H04N)) \times \text{تعداد کل پتنت های شامل مجموعه آیتم}\}}{\text{تعداد کل پتنت ها}} = \frac{0.28 \times 2}{5} \quad (2)$$

$$= 0.112$$

به طور خلاصه، چارچوب الگوریتم BOWARM به شرح زیر می باشد:

۱. بازیابی پتنت ها از پایگاه داده پتنت
۲. محاسبه وزن هر یک از آیتم ها ( تمام کدهای IPC مربوط به پتنت های بازیابی شده)
۳. یافتن تمامی مجموعه آیتم های ممکن
۴. محاسبه ارزش ITW برای هر مجموعه آیتم
۵. محاسبه ارزش WS برای هر مجموعه آیتم
۶. مرتب سازی مجموعه آیتم ها به ترتیب نزولی بودن ارزش WS

#### کاوش قوانین وابستگی وزن دار (WARM\_2)

الگوریتم WARM\_2 از دیگر الگوریتم های پر کاربرد قوانین وابستگی وزن دار می باشد که از الگوریتم WARM، که توسط «تائو» و همکاران<sup>۱</sup> در سال ۲۰۰۳ معرفی شده (Tao et al. 2003)، توسعه یافته است (Altuntas, Dereli, and Kusiak 2015b). این الگوریتم با در نظر گرفتن وزن تراکنش ها، مقدار قوانین وابستگی را محاسبه می کند. دو تعریف کلیدی در این الگوریتم عبارتند از:

وزن تراکنش (TW<sub>2</sub>): تعداد کل اسنادهای پتنت، تقسیم بر تعداد سال های سپری شده از تاریخ صدور پتنت مورد نظر (Altuntas, Dereli, and Kusiak 2015b).

مجموعه آیتم وزن دار (WI<sub>2</sub>): مجموع وزن تمام پتنت های شامل مجموعه آیتم مورد نظر (Altuntas, Dereli, and Kusiak 2015b). چارچوب پیشنهادی با بهره گیری از الگوریتم WARM\_2 به شرح زیر می باشد:

<sup>1</sup> Tao et al

<sup>2</sup> Transaction weight

<sup>3</sup> Weighted itemset

الف) محاسبه وزن پتنت‌ها

جدول ۴ یک پایگاه داده پتنت وزن دار را نمایش می‌دهد.

جدول ۴. پایگاه داده پتنت وزن دار

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

\*تاریخ صدور پتنت از ۲۰۱۶ کم می‌شود.

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

به این ترتیب، وزن اختصاص داده شده به هر پتنت بین صفر تا یک و مجموع وزن‌ها برابر یک خواهد بود. به عنوان مثال، برای پتنت US1 ارائه شده در جدول ۴، تعداد سال‌های سپری شده چهار است (۲۰۱۲ - ۲۰۱۶) و میانگین تعداد استنادهای پتنت در یک سال ۳۰ است (۱۲۰/۴). همچنین، وزن نرمال شده آن برابر ۰,۲۴۰ خواهد بود (۳۰/۱۲۵).

### ب) محاسبه میزان ارتباط در خوشه‌های فناوری

در این گام، برای به دست آوردن ارزش قونین وابستگی، مجموع وزن تمام پتنت‌های شامل مجموعه آیتم را به دست می‌آوریم. به عنوان مثال، برای مجموعه آیتم (G06F,H04N) با توجه به جدول ۴ و ۲:

$$\begin{aligned} WI - 2(G06F, H04N) & \\ = \{ \{G06F, H04N\} \text{مجموع وزن تمامی پتنت های شامل مجموعه آیتم} \} & \quad (1) \\ = \{TW_{US3} + TW_{US4}\} = 0.096 + 0.200 = 0.296 & \end{aligned}$$

به طور خلاصه، مراحل الگوریتم WARM\_2 به شرح زیر می‌باشد:

۱. بازیابی پتنت‌ها از پایگاه داده پتنت
۲. محاسبه ارزش TW\_2 برای هر تراکنش (پتنت)
۳. یافتن تمامی مجموعه آیتم‌های ممکن
۴. محاسبه ارزش WI\_2 برای هر مجموعه آیتم
۵. مرتب‌سازی همه مجموعه آیتم‌ها به ترتیب نزولی بودن 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 براساس تعداد استناد پتنت‌های شامل مجموعه آیتم موردنظر محاسبه می‌شود. تعاریف مورد استفاده در الگوریتم ترکیبی ۱ به شرح زیر می‌باشد:

وزن مجموعه آیتم (ITW): مجموع وزن همه ی آیتم‌های موجود در مجموعه آیتم تقسیم بر تعداد کل آیتم‌ها.

پشتیبانی وزن‌دار (WS): مجموع ITW تمام پتنت‌های شامل مجموعه آیتم موردنظر تقسیم بر تعداد کل پتنت‌ها.

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

الف) محاسبه وزن سطوح فناوری (آیتم‌ها)

برای محاسبه وزن آیتم‌ها ابتدا وزن هر پتنت با تقسیم تعداد اسنادهای پتنت بر تعداد سال‌های گذشته از صدور آن به دست آورده شده و سپس نرمالسازی می‌شود. در مرحله بعد، وزن پتنت-های شامل آیتم (کد IPC) مورد نظر به عنوان 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 مجموعه آیتم‌های محتمل نیز محاسبه می‌شود. به عنوان مثال، برای مجموعه آیتم (G06F,H04N) باتوجه به جدول ۵ محاسبات به شرح معادله ۴ خواهد بود:

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

ب) محاسبه میزان ارتباط میان فناوری‌های مختلف

برای به دست آوردن ارزش یک قانون وابستگی مشابه روش BOWARM، وزن مجموعه آیتم مورد نظر در تعداد پتنت‌های شامل مجموعه آیتم ضرب شده و بر تعداد کل پتنت‌ها تقسیم می‌شود. به عنوان مثال، برای مجموعه آیتم (G06F,H04N) باتوجه به جدول ۲ در معادله ۵ خواهیم



داشت:

$$WS(G06F, H04N) = \frac{\{ \text{تعداد کل پتنت های شامل مجموعه آیتم} \times (ITW(G06F, H04N)) \}}{\text{تعداد کل پتنت ها}} = \frac{0.516 \times 2}{5} \quad (5)$$

$$= 0.2064$$

به طور خلاصه، چارچوب الگوریتم CBW\_1 به شرح زیر می باشد:

۱. بازیابی پتنت ها از پایگاه داده پتنت
۲. محاسبه وزن هریک از آیتم ها
۳. یافتن تمامی مجموعه آیتم های محتمل
۴. محاسبه ارزش ITW برای هر مجموعه آیتم
۵. محاسبه ارزش WS برای هر مجموعه آیتم
۶. مرتب سازی مجموعه آیتم ها به ترتیب نزولی بودن میزان WS

#### کاوش قوانین وابستگی وزن دار ترکیبی ۲ (CBW\_2)

الگوریتم ترکیبی ۲ با در نظر گرفتن وزن تراکنش ها، ارزش قوانین وابستگی را محاسبه می کند. در الگوریتم حاضر وزن تراکنش ها بر اساس تعداد استادهای پتنت و نیز وزن مجموعه آیتم مورد نظر محاسبه می شود. تعاریف مورد استفاده در الگوریتم ترکیبی ۲ به شرح زیر می باشد:

وزن تراکنش (TW): وزن استادی پتنت (شامل مجموعه آیتم) ضرب در وزن مجموعه آیتم مورد نظر.

مجموعه آیتم وزن دار (WI): مجموع وزن تمام پتنت های شامل مجموعه آیتم مورد نظر.

چارچوب الگوریتم ترکیبی ۲ به شرح زیر می باشد:

#### ۱- محاسبه وزن پتنت ها

وزن هر آیتم با تقسیم تعداد پتنت های موجود در آیتم بر تعداد کل پتنت ها محاسبه شده و وزن هر مجموعه آیتم با تقسیم مجموع وزن آیتم های موجود در مجموعه آیتم بر تعداد کل آیتم های موجود به دست می آید. برای محاسبه وزن تراکنش ها، ابتدا، وزن استادی پتنت های شامل

مجموعه آیتم با تقسیم تعداد اسنادهای پتنت بر تعداد سال‌های سپری شده از صدور آن محاسبه می‌شود و سپس، مطابق جدول ۴ نرمالسازی می‌شود. در نهایت، در وزن مجموعه آیتم موردنظر ضرب می‌شود. به عنوان مثال، برای مجموعه آیتم (G06F,H04N) پس از محاسبه وزن هر آیتم (مطابق جدول ۳)، در معادله ۶ خواهیم داشت:

$$ITW(G06F, H04N) = \frac{\{(G06F) + (H04N)\}}{2} = \frac{0.51 + 0.05}{2} = 0.28 \quad (6)$$

باتوجه به جداول ۲، ۳ و ۴ وزن پتنت‌ها به ترتیب زیر خواهد بود:

جدول ۶. 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- محاسبه میزان ارتباط میان فناوری‌های مختلف

پس از محاسبه وزن هر پتنت شامل مجموعه آیتم موردنظر، مجموع وزن تمام پتنت‌های شامل مجموعه آیتم مورد نظر را برای به دست آوردن ارزش یک قانون وابستگی مشابه روش WARM\_2 به دست می‌آید. به عنوان مثال، میزان ارتباط میان مجموعه آیتم (G06F,H04N) باتوجه به جدول ۶ بصورت زیر در معادله ۷ بدست می‌آید:

$$\begin{aligned} WI(G06F, H04N) &= \{\text{مجموع وزن تمام های پتنت شامل مجموعه آیتم}(G06F, H04N)\} \\ &= \{TW_{US3} + TW_{US4}\} = 0.02688 + 0.056 = 0.0828 \end{aligned} \quad (7)$$

به طور خلاصه، مراحل الگوریتم CBW\_2 به شرح زیر می باشد:

۱. بازیابی پتنت‌ها از پایگاه داده پتنت
۲. محاسبه وزن هر یک از آیتم‌ها
۳. یافتن تمامی مجموعه آیتم‌های محتمل
۴. محاسبه ارزش ITW برای هر مجموعه آیتم
۵. محاسبه وزن استنادی هر پتنت
۶. محاسبه ارزش TW برای هر تراکنش شامل مجموعه آیتم مورد نظر
۷. محاسبه ارزش WI برای هر مجموعه آیتم
۸. مرتب‌سازی مجموعه آیتم‌ها به ترتیب نزولی بودن میزان WI

#### کاوش قوانین وابستگی وزن دار ترکیبی ۳ (CBW\_3)

الگوریتم پیشنهادی ترکیبی ۳ به منظور محاسبه‌ی 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 به عنوان معیار وزن دهی آیتم‌ها و از تعداد استنادهای پتنت به عنوان معیار وزن دهی تراکنش‌ها استفاده شده است. تعاریف مورد استفاده در الگوریتم ترکیبی ۳ به شرح زیر می باشد:

وزن مجموعه آیتم هر تراکنش (ITW): مجموع وزن همه‌ی آیتم‌های موجود در مجموعه آیتم تقسیم بر تعداد کل آیتم‌ها.  
وزن تراکنش (TW\_2): تعداد کل استنادهای پتنت، تقسیم بر تعداد سال‌های سپری شده از تاریخ صدور پتنت مورد نظر.

پشتیبانی وزن دار (WS<sub>Total</sub>): ترکیب وزن مجموعه آیتم و پشتیبانی وزن دار.

چارچوب الگوریتم پیشنهادی به شرح زیر می باشد:

#### الف) محاسبه 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 محاسبه وزن آیت‌ها مشابه آنچه در الگوریتم BOWARM توضیح داده شد و روش محاسبه وزن پتنت‌ها مشابه الگوریتم WARM\_2 است. به عنوان مثال، برای مجموعه آیت‌ها (G06F, H04N) پس از محاسبه وزن هر آیت، با توجه به جدول ۳ در معادله ۸ خواهیم داشت:

$$ITW(G06F, H04N) = \frac{\{(G06F) + (H04N)\}}{2} = \frac{0.51 + 0.05}{2} = 0.28 \quad (8)$$

برای به دست 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 نرمال کردن، میانگین تعداد استنادها برای هر پتنت بر میانگین کلی تعداد استناد پتنت‌ها تقسیم می‌شود.

#### ب) محاسبه میزان ارتباط در خوشه‌های فناوری

پس از محاسبه وزن آیت‌ها و پتنت‌ها برای به دست آوردن مقدار پشتیبانی وزن‌دار، ارزش قوانین وابستگی حاصل از الگوریتم BOWARM با ارزش قوانین حاصل از الگوریتم WARM\_2 بصورت معادله ۹ ترکیب می‌شود.

$$WS_{Total} = WS(G06F, H04N) + WI_2(G06F, H04N) = 0.112 + 0.296 = 0.408 \quad (9)$$

به طور خلاصه، مراحل الگوریتم CBW\_3 به شرح زیر می‌باشد:

۱. بازیابی پتنت‌ها از پایگاه داده پتنت
۲. محاسبه میزان TW\_2 برای هر پتنت و وزن هریک از آیت‌ها
۳. یافتن تمامی مجموعه آیت‌های ممکن
۴. محاسبه میزان ITW برای هر مجموعه آیت
۵. محاسبه میزان WS<sub>Total</sub> برای هر مجموعه آیت
۶. مرتب‌سازی مجموعه آیت‌ها به ترتیب نزولی بودن میزان WS<sub>Total</sub>

#### ۴. نتایج پژوهش

از میان پتنت‌های بررسی شده، ۲۲۹ پتنت مطلوب و ۳۹ کد IPC منحصر به فرد بدست آمد. آخرین پتنت در نظر گرفته شده در این مطالعه، US8214319 می‌باشد که در تاریخ ۳ جولای ۲۰۱۲ صادر شده است. بیشترین ارتباط و وابستگی متعلق به کد G06F، پردازش داده دیجیتال الکتریکی<sup>۱</sup>، می‌باشد. 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 توسط 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، به عنوان مثال در زمینه طراحی، بر ۳۶ فناوری دیگر تاثیر می‌گذارد و بالعکس. بنابراین، فناوری G06F می‌تواند به عنوان مهم‌ترین فناوری در میان ۳۹ فناوری شناسایی شده در نظر گرفته شود. فناوری‌های درگیر در G01R، G01D، G01J، G10L، G06E، E21B، H03M، G09G، G11C، G09C، A61B، G01S، G21C تنها یک ارتباط دارند و آن ارتباط نیز با فناوری G06F است. بدین ترتیب اگر فناوری G06F توسعه داده شود، این امکان وجود دارد که فناوری‌های مذکور نیز به دلیل وابستگی میان آن‌ها در آینده توسعه داده شوند.

---

<sup>1</sup> Electric digital data processing

جدول ۷. ماتریس ارتباطات میان خوشه های دوتایی

	G06F	H04L	G06Q	H04Q	H04N	H04K	H04W	H04J	G01N	G06G	G06E	G06K	G06T	G06N	G07F	G08B	G09B	G10L	G11B	G09G	G07C	G01R	G01D	G01J	G11C	G09C	A61B	B64F	G01S	G21C	مجموع		
H04L	۱																															۱	
H04Q			۱																													۲	
H03M																																۱	
H04M		۱		۱																												۴	
H04N		۱		۱								۱	۱																			۱۱	
H04K																																۲	
H04W			۱										۱																			۳	
H04J																																۱	
C12Q									۱	۱																						۲	
C07D																			۱			۱										۳	
C40B									۱																							۲	
E21B																																۱	
G01N										۱		۱																	۱			۶	
G01C																														۱		۲	
G05B																		۱														۲	
G06F			۱							۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۲۱	
H04B												۱																				۱	
G06Q											۱	۱	۱	۱	۱																	۱۱	
G06G			۱																													۲	
G06K													۱	۱															۱			۴	
G06T														۱																		۴	
G06N																																۱	
G07F																						۱										۱	
G09B																																۲	
G11B																																	۱
G07C																																	۱
مجموع	۱۵	۲	۸	۳	۱	۱	۳	۳	۲	۳	۴	۵	۶	۴	۲	۳	۳	۷	۳	۳	۳	۳	۳	۲	۳	۳	۲	۲	۴	۲	۱	۱۰۷	

وابستگی می‌تواند در خوشه‌های دو، سه، چهار و پنج‌تایی رخ دهد. همچنین، حداکثر اندازه ممکن یک خوشه فناوری برابر است با حداکثر تعداد آیتم‌های (کدهای IPC) موجود در یک تراکنش (پنت) از بین پنت‌های بازیابی شده. (در اینجا پنج بود). در این پژوهش، وابستگی در یک خوشه فناوری به این معناست که این فناوری‌ها بر یکدیگر اثر می‌گذارند. به عنوان مثال، برای خوشه دوتایی (G06F, G06Q)، در صورت توسعه فناوری G06F، این احتمال وجود دارد که در آینده فناوری G06Q به دلیل وابستگی میان آن‌ها توسعه داده شود و نیز سرمایه‌گذاری در فناوری G06F، موجب افزایش احتمال سرمایه‌گذاری در فناوری G06Q در آینده خواهد شد. خوشه 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، ترکیب اول از خوشه‌های دو تا چهارتایی و جدول ۱۱ خوشه‌های پنج‌تایی به دست آمده از الگوریتم‌های مورد استفاده را نشان می‌دهند. قدرت وابستگی میان مجموعه 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 خوشه دوتایی برای همه‌ی روش‌ها مشابه هستند. خوشه‌های باقیمانده برای روش‌های مختلف توالی‌های متفاوت دارند. در جدول ۹، هشت مجموعه آیتم اول برای سطح ۳ (خوشه سه‌تایی) در روش‌های BOWARM و CBW\_1 یکسان هستند و همچنین، چهار مجموعه 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. براساس جدول ۱۰، تمامی روش‌ها نتایج مشابهی را برای اولین مجموعه آیتم در سطح ۴ تولید کردند. ده مجموعه آیتم اول برای سطح ۴ که توسط روش‌های BOWARM و CBW\_2 تولید شده‌اند، مشابه هستند و نیز ۴ مجموعه آیتم اول برای روش‌های WARM\_2 و CBW\_3 یکسان هستند. همان‌طور که در جدول ۱۱ نشان داده شده‌است، روش‌های BOWARM و CBW\_1 توالی یکسان برای سطح ۵ مجموعه آیتم تولید می‌کنند و الگوریتم‌های WARM\_2 و CBW\_2 و CBW\_3 توالی‌های مشابه برای سطح ۵ مجموعه آیتم فراهم کرده‌اند. در این مطالعه موردی، علت یکسان شدن توالی مجموعه آیتم‌ها در برخی سطوح، شدت ارتباط و وابستگی مجموعه آیتم‌ها است. در واقع، مجموعه آیتم‌های به شدت مرتبط، با وجود روش‌های مختلف، توالی‌های یکسان تولید می‌کنند.

جدول (۸): برشی از خوشه های دوتایی (اعداد بر حسب درصد می باشند)

BOWARM		WARM_2		CBW_1		CBW_2		CBW_3	
مجموعه آیتم	WS	مجموعه آیتم	WI	مجموعه آیتم	WS	مجموعه آیتم	WI	مجموعه آیتم	WS <sub>TOTAL</sub>
H04L,G06F	۰,۵۰۸۹	H04L,G06F	۳,۱۱۷۶	H04L,G06F	۱,۳۵۸۱	H04L,G06F	۰,۷۲۶۶۸	H04L,G06F	۳,۶۲۶۵
G06F,G06Q	۰,۴۳۴۶	G06F,G06Q	۲,۵۷۶	G06F,G06Q	۱,۱۳۳۴	G06F,G06Q	۰,۵۶۹۸۴	G06F,G06Q	۳,۰۱۰۷
G06F,G06K	۰,۲۴۸۱۲	G06F,G10L	۱,۱۵۷۹	G06F,G06K	۰,۶۱۰۰۱	G06F,G06K	۰,۲۳۲۰۱	G06F,G06K	۱,۳۹۱۵
G06F,G06N	۰,۲۲۶۲۴	G06F,G06K	۱,۱۴۳۳	G06F,G06N	۰,۵۲۵۳۸	G06F,G10L	۰,۲۱۵۰۳	G06F,G10L	۱,۲۴۸
H04N,G06F	۰,۱۳۴۷۶	G06F,G09B	۱,۰۴۴۱	H04N,G06F	۰,۳۲۸۴۸	G06F,G09B	۰,۱۸۹۹۲	G06F,G09B	۱,۰۹۹۷
G06F,G06T	۰,۱۰۰۱۱	G09B,G10L	۰,۹۷۴۸۲	G06F,G06T	۰,۲۴۵۳۶	G06F,G06T	۰,۰۸۸۹۴۳	G09B,G10L	۰,۹۷۸۸۱
G06F,G10L	۰,۰۸۹۱۳	H04L,G06Q	۰,۷۱۵۵۶	G06F,G10L	۰,۲۳۸۵۴	G06F,G06N	۰,۰۸۸۶۳۹	H04L,G06Q	۰,۷۳۸۵۲
G06F,G06G	۰,۰۶۳۸۶	G06F,G06T	۰,۴۶۵۵۷	G01N,G06F	۰,۱۵۹۸۶	H04N,G06F	۰,۰۸۴۷۲	G06F,G06N	۰,۶۷۱۰۶
G01N,G06F	۰,۰۶۳۵۴	G06F,G06N	۰,۴۴۴۸۲	G06F,G06G	۰,۱۵۷۸	H04L,G06Q	۰,۰۷۵۲۱	H04N,G06F	۰,۵۷۴۰۴
G06F,G09B	۰,۰۵۵۶	H04N,G06F	۰,۴۳۹۲۹	G06F,G09B	۰,۱۵۱۲۳	G01N,G06F	۰,۰۷۰۸۳	G06F,G06T	۰,۵۶۵۶



جدول (۹): برشی از خوشه های سه تایی (اعداد بر حسب درصد می باشند)

BOWARM		WARM-2		CBW_1		CBW_2		CBW_3	
مجموعه آیتم	WS	مجموعه آیتم	WI	مجموعه آیتم	WS	مجموعه آیتم	WI	مجموعه 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	H04L,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	۰,۳۳۲۴۸
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



X | شماره ۱۳xx | دوره xx | X

جدول (۱۰): برشی از خوشه های چهارتایی (اعداد بر حسب درصد می باشند)

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

طبقه‌بندی بین‌المللی پتنت‌های مورد استفاده در اداره 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، مخلوط کردن عملیات مختلف شکل دادن، چاپ و نشر، حمل‌ونقل، ترابری، فناوری ساختارهای ذره بینی، نانوفناوری. بخش ج: شیمی متالورژی، زیربخش‌ها: شیمی، متالورژی، بخش د: منسوجات، کاغذ، زیربخش‌ها: منسوجات یا سایر مواد قابل انعطاف که به صورت 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 هسته ای و بخش ح: برق.

همان‌طور که مشاهده می‌شود، انواع مختلفی از پتنت‌ها وجود دارد. در این پژوهش، به منظور 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.

## ۵. بحث و ارزیابی نتایج

در این پژوهش، برای سنجش صحت و اعتبار خروجی‌های به دست آمده، از روش قیاسی استفاده شده است. همان‌طور که در جدول ۸ دیده می‌شود، اولین و دومین خوشه دوتایی برای همه روش‌ها مشابه هستند. برای محاسبه میزان تشابه میان خروجی‌های روش‌های به دست آمده از شاخص ژاکار استفاده می‌کنیم. شاخص تشابه ژاکار راهی برای مقایسه دو جامعه با نگاهی به اعضای هر دو جمعیت (تعداد کل اعضای مجزا) و اعضای حاضر در هر دو جمعیت (تعداد اعضای مشترک) می‌باشد. شاخص ژاکار با استفاده از رابطه ۶ محاسبه می‌شود. میزان تشابه در خوشه‌های دو تا پنج تایی در جدول ۱۲ ارائه شده است.

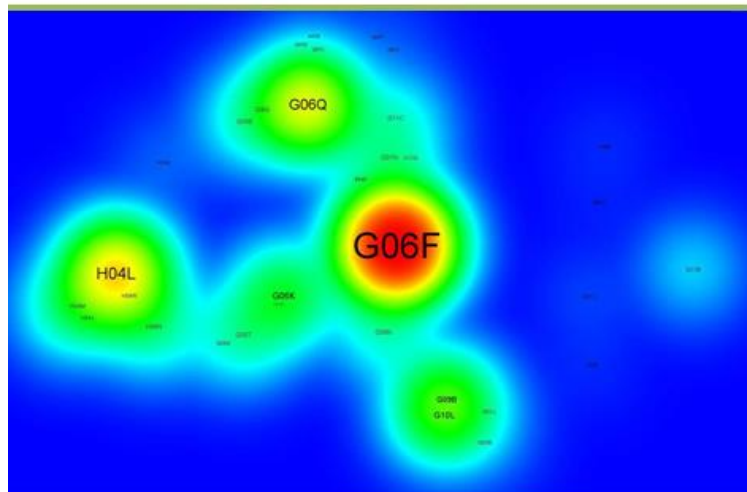
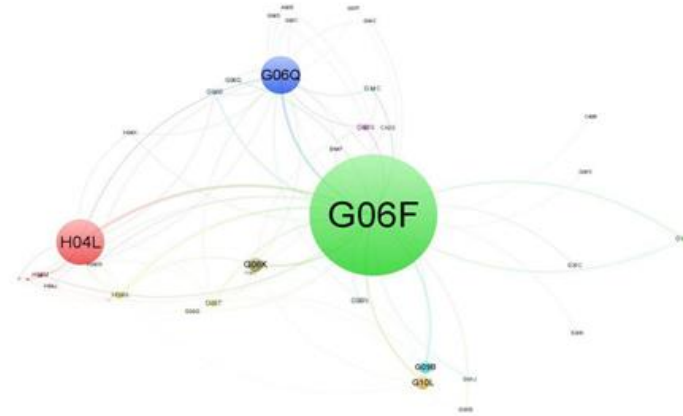
$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

جدول (۱۲): میزان تشابه روش های پیشنهادی در خوشه های ۲ تا ۵ تایی (برحسب درصد)

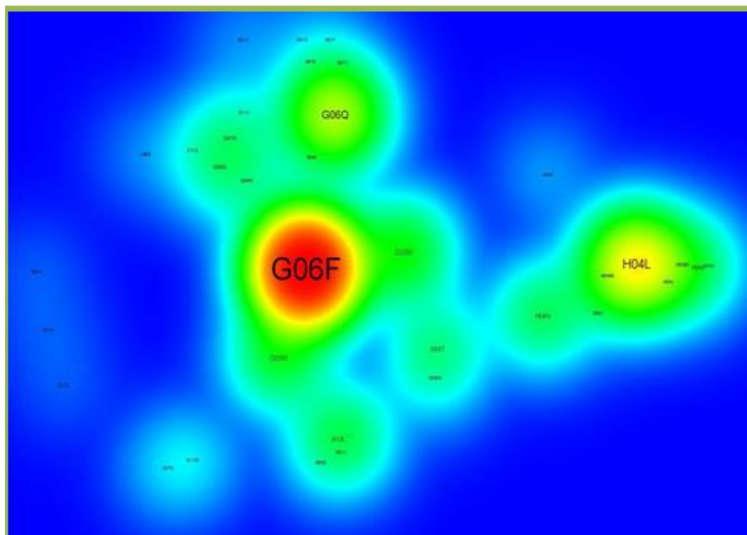
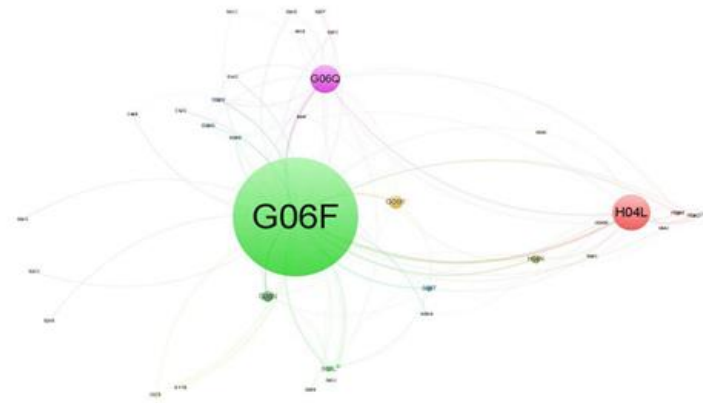
سطح	روش ها	CBW_2	CBW_3
۲	CBW_1	۹۵	۹۶
	CBW_2	-	۹۲
۳	CBW_1	۹۴	۱۰۰
	CBW_2	-	۹۴
۴	CBW_1	۹۵	۱۰۰
	CBW_2	-	۱۰۰
۵	CBW_1	۱۰۰	۱۰۰
	CBW_2	-	۱۰۰

باتوجه به نتایج به دست آمده میزان مشابهت روش ها بسیار بالاست. تفاوت های موجود اغلب در اولویت بندی مجموعه آیت های مشتق شده و به دلیل تفاوت در وزن دهی هریک از روش ها است. بنابراین، از تکرار مجموعه آیت ها در الگوریتم های مختلف، می توان صحت و اعتبار وابستگی های به دست آمده را نتیجه گرفت.

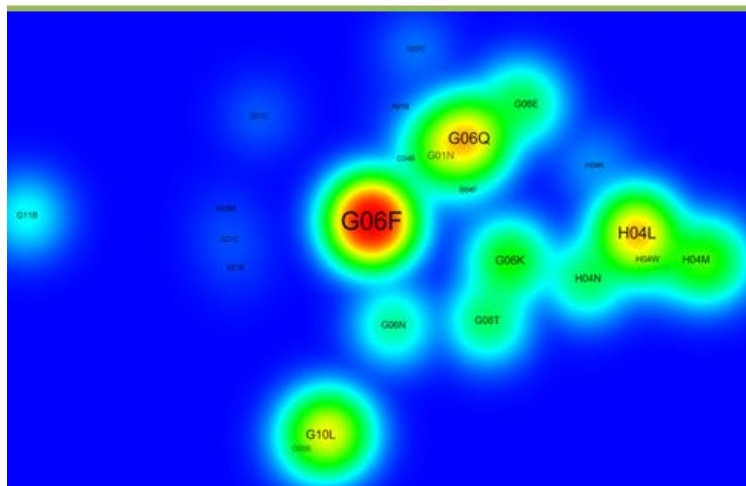
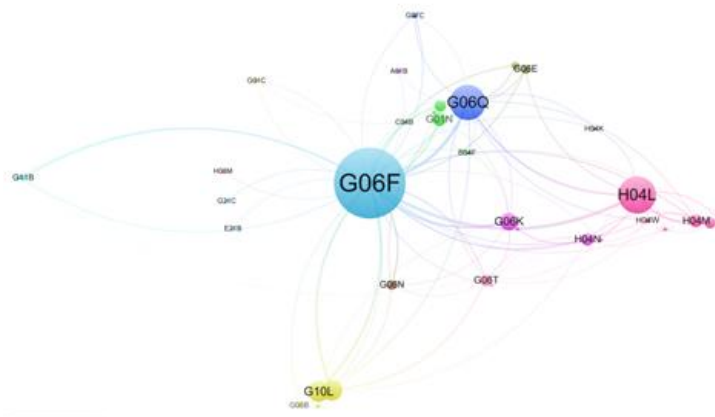
یکی از محدودیت هایی که در مقالات پیشین ذکر شده، عدم امکان تجسم وابستگی ها و ارتباطات بدست آمده برای کاربر است (Altuntas, Dereli and Kusiak 2015b). تجسم وابستگی های به دست آمده به درک بهتر قوانین استخراج شده کمک می کند. در این پژوهش، ارتباطات استخراج شده، براساس الگوریتم های پیشنهادی به صورت گراف و با استفاده از نرم افزار VOS Viewer در شکل های ۲، ۳ و ۴ ترسیم شده اند. تصویرسازی اطلاعات می تواند نقش مهمی در اکتشاف، شناسایی و تفسیر قوانین وابستگی به منظور توانمند ساختن کاربران در درک داده هایی با ابعاد بالا و مقیاس بزرگ ایفا کند (Sekhvat and Hoeber 2013).



شکل ۲. فاصله و ناحیه تحت پوشش کدهای IPC در الگوریتم های پیشنهادی CBW\_1



شکل ۳. فاصله و ناحیه تحت پوشش کدهای IPC در الگوریتم های پیشنهادی CBW\_2



شکل ۴. فاصله و ناحیه تحت پوشش کدهای IPC در الگوریتم های پیشنهادی CBW\_3

## ۶. نتیجه گیری

در این مطالعه، از قوانین وابستگی وزن دار که از الگوریتم های محبوب داده کاوی است، برای شناسایی و ارزیابی خوشه های شامل فناوری های مرتبط براساس تحلیل پتنت استفاده شده است. در این پژوهش پنج الگوریتم مبتنی بر قوانین وابستگی وزن دار به کار گرفته شده است (BOWARM, WARM\_2, CBW\_1, CBW\_2, CBW\_3). خروجی این الگوریتم ها، فناوری های مرتبط با فناوری مورد نظر و همچنین میزان ارتباطات آن ها را تعیین می نماید.

الگوریتم‌های پیشنهادی ارتباطات میان فناوری‌ها را با در نظر گرفتن اهمیت نابرابر فناوری‌ها و همچنین پتنت‌ها، بصورت همزمان، به دست می‌آورند. نتایج نشان می‌دهد که روش‌های پیشنهادی، به طور کلی، توالی‌های یکسان در یک یا چند مجموعه آیتم اول در تمامی سطوح فراهم می‌کنند و مجموعه آیتم‌های باقیمانده نیز مشابه هستند و تنها اولویت‌بندی‌های متفاوتی دارند. تفاوت در اولویت‌بندی‌ها به دلیل تفاوت در نوع و معیار وزن‌دهی می‌باشد. کارمندان و مدیران می‌توانند از روش پیشنهادی به منظور یافتن ارتباطات میان فناوری‌های مورد نظر استفاده کنند. کارآفرینان ممکن است شرکت‌هایی که فناوری‌های مرتبط با آنها را تولید می‌کنند، شناسایی کرده و دریابند که این شرکت‌ها در حوزه فناوری مورد نظر در حال انجام چه کاری هستند. سرمایه‌گذاران، همچنین سیاست‌گذاران می‌توانند اطلاعات مربوط به فناوری‌های جایگزین مرتبط با فناوری‌های ساخت خود را با استفاده از روش‌های پیشنهادی به دست بیاورند. یک پروژه سرمایه‌گذاری ممکن است با بیش از یک فناوری منفرد مرتبط باشد و نیز ممکن است موفقیت و پایداری فناوری سرمایه‌گذاری شده را در آینده نشان دهد. پیشنهادات آتی: ارقام اضافی کدهای پتنت می‌تواند برای یافتن ارتباطات، در نظر گرفته شود. تحلیل ادعاهای پتنت با استفاده از قوانین وابستگی می‌تواند انجام شود. روش‌های دیگر مانند الگوریتم‌های مبتنی بر منطق فازی می‌توانند برای تحلیل پتنت به منظور پیش‌بینی ارتباطات میان تکنولوژی‌ها به کار گرفته شوند. محدودیت‌های روش پیشنهادی: آماده‌سازی و پاکسازی داده‌ها آسان نیست. نرم‌افزاری که داده‌های مورد نظر را استخراج کند و پردازش‌های لازم را انجام دهد، مورد نیاز است. تعدادی از مجموعه آیتم‌ها تاحدودی بزرگ هستند که این امر موجب افزایش تلاش محاسباتی می‌شود.

همچنین، همان‌طور که گفته شد، از مزایای راهبردی تحلیل پتنت در فرایندهای مدیریتی می‌توان به توانایی شناسایی حوزه‌های تکنولوژیکی برای سرمایه‌گذاری اشاره کرد. پیش‌بینی حوزه‌های تکنولوژی در حال ظهور برای مدیران دشوار است. در نتیجه، سرمایه‌گذاری روی تکنولوژی‌های جدید ریسک بالایی به همراه دارد. برای کاهش خطرات این سرمایه‌گذاری، مدیران می‌توانند با بهره‌گیری از تحلیل و ارزیابی پتنت با استفاده از تکنیک‌های مختلف داده-کاوی پیش‌بینی دقیق‌تری از حوزه‌های تکنولوژی مناسب برای سرمایه‌گذاری داشته‌باشند. 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 معرفی شده در این پژوهش می‌تواند نقشه راهی در حوزه‌های تکنولوژیکی برای مدیران و شرکت‌ها فراهم کند. به علاوه، 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

- علم خواه، حسن. ۱۳۹۱. پتنت (حق اختراع) مطالعه، جست‌وجو در منابع بین‌المللی و تحلیل. موسسه دارایی‌های فکری و فناوری مدرس.
- غیاثی، فرناز، نوید نظافتی، و سجاد شکوهیار. ۱۳۹۴. خوشه بندی کاربران داده های دریایی با استفاده از تکنیک داده کاوی. پژوهشنامه پردازش و مدیریت اطلاعات ۳۰(۴): ۱۰۲۵-۴۹.
- منصوری، علی، و فریده عصاره. ۱۳۹۳. بررسی ارزش منابع دانش بر اساس عامل اهمیت: («نگاه به آینده» و «نگاه به گذشته») مطالعه موردی پروانه های ثبت اختراع. پژوهشنامه پردازش و مدیریت اطلاعات ۲۹(۳): ۶۵۷-۸۳.
- ناظمی، امیر، علی شماعتی، و روح‌الله قدیری. ۱۳۹۲. شناسایی حوزه‌های نوظهور براساس تحلیل چرخه عمر فناوری به عنوان نمونه فناوری زیر دریایی های بدون سرنشین. فصلنامه مدیریت توسعه فناوری ۱(۳): ۲۹-۴۷.
- Altuntas, Serkan., and Turkey 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: 1003-12.
- Altuntas, S., T. Dereli, and A. Kusiak. 2015a. Forecasting technology success based on patent data. *Technological Forecasting and Social Change* 96: 202-14.
- Altuntas, S., T. Dereli, and A. Kusiak. 2015b. Analysis of patent documents with weighted association rules. *Technological Forecasting and Social Change* 92: 249-62.
- 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): 1296-314.
- Dalkey, Norman. C. 1971. An elementary cross-impact model. *Technological Forecasting and Social Change* 3, 341-51.
- Dereli, Turkey., 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): 263-72.

- 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): 1316-30.
- 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): 339-50.
- 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): 203-14.
- 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.
- Jun, Sunghae. 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): 539-46.
- 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): 12559-64.
- Kim, C., 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): 1948-52.
- Kim, Jeeun., and Sungjoo Lee. 2015. Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO. *Technological Forecasting and Social Change* 92: 332-45.
- 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.
- Lee, 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-

86.

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

Li, S., D. Zheng, and J. Shi. 2014. Technology forecasting of green energy using patent analysis and simulation modeling. *Advanced Materials Research* 875: 1619–24.

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*: 217-21.

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.

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):1037-50.

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*: 270-74.

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): 3386–407.

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.

## Patent Analysis by Data Mining for identifying and determining relationships among technologies

### Leila MahmudJanlu

Ms In innovation and technologies management, Mazandaran University of Science and Technology, [Leila\\_janlu@yahoo.com](mailto:Leila_janlu@yahoo.com)

### Babak Shirazi\* (Corresponding author)

Assistant Professor, Mazandaran University of Science and Technology, [Shirazi\\_b@yahoo.com](mailto:Shirazi_b@yahoo.com)

### Iraj Mahdavi

Professor, Mazandaran University of Science and Technology, [Irajarash@rediffmail.com](mailto:Irajarash@rediffmail.com)

### Javad Soltanzadeh

PhD Candidate of technology management, Faculty of Management and Accounting, Allameh Tabataba'i University, [Soltanzadeh921@atu.ac.ir](mailto:Soltanzadeh921@atu.ac.ir)

**ABSTRACT:** Analyzing technologies relationships can provide insight into technological strategies and maximizing the profits. Patents are integral parts of intellectual property rights including 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/theirs 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.

### لیلا محمودجانلو

متولد سال ۱۳۶۸، دارای مدرک تحصیلی کارشناسی ارشد رشته مدیریت نوآوری و فناوری از دانشگاه علوم و فنون مازندران است. وی دانش آموخته کارشناسی مهندسی صنایع دانشگاه صنعتی نوشیروانی بابل است. مباحث مرتبط با داده کاوی، برنامه نویسی، تحلیل پتنت، سازماندهی دانش از جمله علایق پژوهشی وی است.



### بابک شیرازی

متولد سال ۱۳۵۷، دارای مدرک تحصیلی دکتری مهندسی صنایع در حوزه تولید با فناوری اتوماتیک می باشد. ایشان هم اکنون استادیار گروه مهندسی صنایع دانشگاه علوم و فنون مازندران می باشد. فناوری های نوین مهندسی تولید، سیستم های اطلاعاتی پیشرفته و آینده پژوهی فناوری از جمله علایق پژوهشی وی است.



### ایرج مهدوی

استاد تمام گروه مهندسی صنایع دانشگاه علوم و فنون مازندران می باشد. دوره دکتری را در زمینه مهندسی تولید در کشور هند گذراندند. ایشان بیش از ۳۰۰ مقاله علمی در زمینه های تولید سلولی، زنجیره تامین، برنامه ریزی تولید، شبکه های فازی، مدیریت دیجیتال و مدیریت عملیات هوشمند در مجلات و کنفرانس های بین المللی منتشر کرده اند. عضویت در هیات تحریریه ۵ مجله معتبر بین المللی، پژوهش گر برتر کشور در گروه فنی و مهندسی، استاد نمونه کشوری از دیگر افتخارات ایشان است.



### جواد سلطانزاده

متولد ۱۳۶۴، دانشجوی دکتری مدیریت تکنولوژی دانشگاه علامه طباطبائی است. وی دانش آموخته کارشناسی مهندسی صنایع دانشگاه علم و صنعت ایران و کارشناسی ارشد مدیریت تکنولوژی دانشگاه علامه طباطبائی است. مدیریت فناوری و نوآوری، سیاست-گذاری نوآوری و نظام نوآوری از حوزه علایق پژوهشی وی است.

