

کشف دانش و کاربرد آن در اینترنت اشياء

کیوان برنا^۱، فرهاد فتحي*^۲، عصمت مؤمنی^۳

مطالعات دانش شناسی

سال پنجم، شماره ۱۷، زمستان ۹۷، ص ۱۲۳ تا ۱۵۵

تاریخ دریافت: ۹۷/۰۷/۳۰

تاریخ پذیرش: ۹۷/۱۰/۱۵

چکیده

اینترنت اشياء، به طور چشمگیری زندگی ما را در آینده ای نزدیک تغییر خواهد داد و بسیاری از ناممکن ها را ممکن خواهد ساخت. حجم عظیم داده ی تولید شده یا گرفته شده توسط تجهیزات اینترنت اشياء، حاوی اطلاعات ارزشمند و قابل استفاده است. با رواج دستگاه های توسعه یافته فناوری بی سیم مانند بلوتوث، شناسایی با فرکانس رادیویی (RFID)، Wi-Fi و خدمات داده بر روی تلفن و همچنین سنسور و محرک و نودهای تعبیه شده در وسایل، شبکه های حسگر بی سیم، اینترنت اشياء مراحل ابتدایی خود را پشت سر گذاشته و در آستانه تبدیل اینترنت ایستای کنونی، به اینترنت کاملاً یکپارچه در آینده است. کشف دانش از طریق داده کاوی و متن کاوی نیز بدون شک نقش زیادی در زمینه هوشمندسازی سیستم ها و در نتیجه ارائه خدمات و محیط مناسب برای ارائه خدمات خواهد داشت. همچنین از روش های داده کاوی برای خوشه بندی تجهیزات در شبکه های حسگر بی سیم و تعیین سرخوشه استفاده بسیاری می شود. در این مقاله به معرفی اینترنت اشياء، معماری، کشف دانش، نقش و کاربرد داده کاوی و متن کاوی در این حوزه پرداخته شده است.

واژه های کلیدی: اینترنت اشياء، داده کاوی، داده های حجیم، شبکه های حسگرهای بی سیم، کشف دانش، متن کاوی

۱. استادیار، گروه علوم کامپیوتر، دانشکده علوم ریاضی و کامپیوتر، دانشگاه خوارزمی، تهران، ایران،

borna@khu.ac.ir

۲. * کارشناس ارشد، مهندسی کامپیوتر (نرم افزار)، موسسه آموزش و ترویج کشاورزی، سازمان تحقیقات، آموزش و

ترویج کشاورزی، تهران، ایران، f.fathi@itvhe.ac.ir

۳. استادیار، گروه علم اطلاعات و دانش شناسی، دانشگاه علامه طباطبائی، momeni.esmat@yahoo.com

مقدمه

عبارت اینترنت اشیاء برای اولین بار توسط کوین اشتون در سال ۱۹۹۹ در متنی با موضوع مدیریت زنجیره‌ای، ارائه شد (اشتون^۱، ۲۰۰۹)؛ اما در دهه گذشته، این تعریف طیف گسترده‌تری از برنامه‌های کاربردی مانند بهداشت و درمان، آب و برق، حمل‌ونقل و غیره را دربرگرفت (ساندماکر و همکاران^۲، ۲۰۱۰). اینترنت اشیاء پدیده‌ای نوظهور است که به سرعت و همراه با ارتباطات بی‌سیم در حال گسترش است. ایده اینترنت اشیاء، اتصال همه‌چیز در جهان، به اینترنت است. اینترنت اشیاء از سه مفهوم اینترنت-گرایی (میان‌افزار)، اشیاء-گرایی (حسگرها)، معناگرایی (دانش) تحقق می‌یابد. ایده اصلی این است که اشیاء یا چیزهایی مانند برچسب‌های شناسایی با فرکانس‌های رادیویی، حسگرها، تلفن‌های موبایل و غیره که الگوی آدرس‌دهی منحصربه‌فردی دارند، می‌توانند با یکدیگر و محیط اطراف خود، برای رسیدن به اهداف مشترک همکاری کنند (آتزوری و همکاران^۳، ۲۰۱۰). هالر^۴ اینترنت اشیاء را این‌گونه معرفی کرده است: «جهانی که اشیاء فیزیکی به‌طور یکپارچه، در شبکه‌های اطلاعاتی ادغام شده‌اند و مکانی است که در آن اشیاء فیزیکی در فرآیندهای کسب‌وکار و اطلاعاتی و با در نظر گرفتن امنیت و حریم خصوصی، در دسترس می‌باشند.» (شن بین و همکاران^۵، ۲۰۱۰). پروفیسور لئو^۶ ایده خود را از دیدگاه تکنولوژی و اقتصاد این‌گونه بیان می‌کند. «از دیدگاه تکنولوژی اینترنت اشیاء ادغامی از شبکه‌های حسگر شامل برچسب‌های شناسایی با فرکانس رادیویی و رایانش فراگیر^۷ است. از دیدگاه اقتصاد، مفهوم جدیدی است که فناوری‌های جدید و برنامه‌های کاربردی مرتبط، محصولات و خدمات، تحقیقات و توسعه، صنعت و بازار را با هم ادغام کرده است» (شن بین و همکاران، ۲۰۱۰).

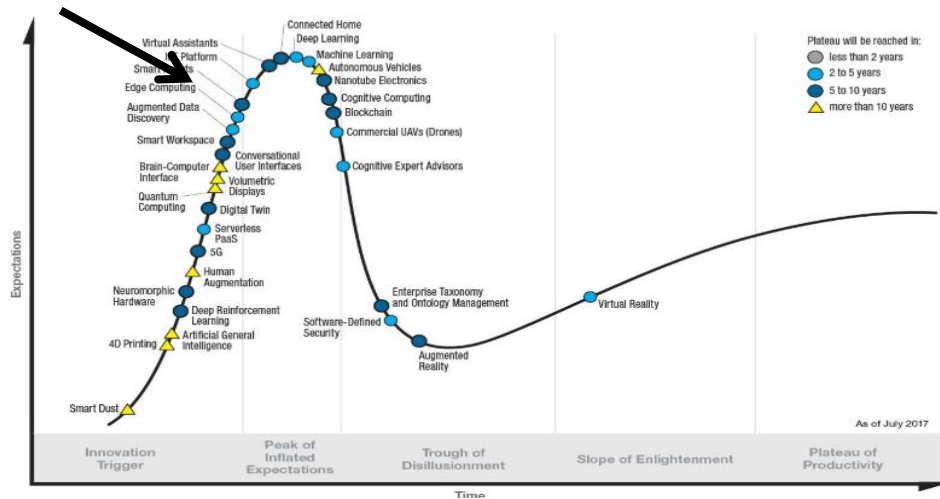
1. Ashton
2. Sundmaeker, et al.
3. Atzori, et al.
4. Haller
5. Shen Bin, et al.
6. Lio
7. ubiquitous networks

ارتباط میدان نزدیک^۱ و شبکه‌های حسگر و فعال‌کننده بی‌سیم^۲ و برجسب‌های شناسایی با فرکانس رادیویی، با هم «اجزای اتمیکی هستند که دنیای واقعی را با دنیای دیجیتال پیوند خواهند داد» (آتزوری و همکاران، ۲۰۱۰). گروه RFID، اینترنت اشیا را این‌گونه تعریف می‌کند: «شبکه جهانی وب از اشیا متصل به هم، با آدرس‌دهی یکتا بر اساس پروتکل‌های ارتباطی استاندارد» (گوبی و همکاران^۳، ۲۰۱۳). بر اساس پروژه‌های تحقیقاتی اروپا درباره اینترنت اشیا: «اشیا» قطعات فعال در کسب و کار، پردازش‌های اطلاعاتی و اجتماعی هستند که قادرند بین خودشان و با محیط اطراف خود به‌وسیله تبادل داده و اطلاع حس شده از محیط، تعامل کرده و ارتباط داشته باشند. در صورت مواجهه با رویداد فیزیکی/حقیقی، واکنشی خودکار و مؤثر از طریق اجرای پردازش‌هایی متناسب با عملیات در حال اجرا و ایجاد خدمات با/ بدون دخالت انسان از خود بروز می‌دهند (سان مارکر و همکاران، ۲۰۱۰) بر اساس ایده فارستر، یک محیط هوشمند، از فناوری‌های اطلاعات و ارتباطات برای ایجاد قطعات زیرساختی اساسی و خدمات مدیریت شهری، آموزشی، سلامت، امنیت عمومی، املاک و مستغلات، حمل و نقل به شکل تعاملی و مؤثر بهره می‌گیرد (بلیسنت^۴، ۲۰۱۰). گوبی و همکاران، (۲۰۱۳)، تعریف کاربر محور خود از اینترنت اشیا را این‌گونه بیان می‌کنند: «اینترنت اشیا در محیط‌های هوشمند عبارت است از ارتباط داخلی دستگاه‌های حسگر و محرک به‌منظور ایجاد امکان به اشتراک‌گذاری اطلاعات از طریق پلت فرم‌هایی با چارچوب یکسان و توسعه عملیات لازم برای توانمندسازی برنامه‌های کاربردی نوآورانه که از طریق وسایل حسگر یکپارچه، تحلیل داده‌های آن‌ها و نمایش اطلاعات در قالب چارچوب مشترک با کمک رایانش ابری ایجاد می‌شوند.»

اینترنت اشیا به‌عنوان یکی از فناوری‌های نوظهور در فناوری اطلاعات بر اساس چرخه گارتنر شناخته شده است (شکل ۱). چرخه گارتنر^۵ روشی برای نمایش ظهور، اتخاذ، بلوغ

1. Near Field Communications(NFC)
2. Wireless Sensor and Actuator Networks (WSAN)
3. Gubby, et al.
4. Belissent
5. <https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/>

و تأثیر آن در برنامه‌های کاربردی فناوری‌های خاص است. پیش‌بینی شده است که اینترنت اشیا ۲-۵ سال برای ارائه و تأثیرگذاری در بازار، به زمان نیاز دارد.



شکل ۱. نمودار گارتنر ۲۰۱۷ از فناوری‌های در حال ظهور

واژه «اینترنت اشیا» به «شبکه گرا بودن» و «اشیا گرا بودن» بودن آن اشاره دارد. «اینترنت اشیا» از نظر معنایی یعنی «یک شبکه جهانی است که اشیا را با آدرس منحصر به فرد، به یکدیگر بر اساس پروتکل‌های ارتباطی متصل می‌کند». این، یعنی تعداد زیادی اشیا ناهمگون به یکدیگر متصل می‌شوند. آدرس دهی منحصر به فرد اشیا و نمایش و مرتب‌سازی تبادل اطلاعات موضوعاتی چالش برانگیز است که «معناگرا بودن» را به عنوان چشم‌انداز سوم آن مطرح می‌کند (آتزوری و همکاران، ۲۰۱۰).

از ویژگی‌های ابزارهای اینترنت اشیا می‌توان به اندازه کوچک، تعداد مشخص، حافظه کم، استفاده حداقل از انرژی و قابلیت‌های پردازش ویژه آن اشاره کرد که توسط سیستم‌های عامل کوچک که مخصوص این ابزارها ایجاد شده است، راهبری می‌شود. این ابزارها حداقل دارای یکی از توانایی‌های، حس نمودن، فعال‌سازی، ذخیره یا پردازش اطلاعات هستند. پیش‌بینی شده است تا سال ۲۰۵۰ تعداد این ابزارها به ۵۰ میلیارد وسیله برسد. این وسایل به

دو دسته قوی تر که قدرت پردازش و انرژی بیشتری دارند مانند گوشی های هوشمند و دسته ضعیف تر که دارای محدودیت منابع هستند تقسیم می شوند (موسادیق و همکاران^۱، ۲۰۱۸). با پیشرفت اینترنت اشیا، سرویس ها و برنامه های مبتکرانه زیادی ارائه شد که تعدادی از آن ها قابلیت های خوبی را برای عموم مردم به همراه آورند. به عنوان برخی مثال های معروف آن می توان از سیاره هوشمند IBM^۲ و ۵۰ کاربرد برای دنیای هوشمند^۳، استفاده از توانایی اینترنت اشیا در سیستم های هوشمند گرید (یون و یوکسین^۴، ۲۰۱۰) و خانه های هوشمند (رشیدی و همکاران^۵، ۲۰۱۱) نام برد.

از ابزارهای اینترنت اشیا می توان در شهرهای هوشمند برای ایجاد حمل و نقل هوشمند به منظور کاهش بار ترافیک، پیدا کردن نزدیک ترین پارکینگ، پیشگیری از تصادف اشاره کرد. وسایل نقلیه ای که با یکدیگر در بستر شبکه متصل اند و ابزارهای اینترنت اشیا مانند حسگرهای GPS، ابزارهای RFID و دوربین ها، با یکدیگر در بستر فضای ابری، تبادل اطلاعات می کنند. حتی تشخیص ترافیک با کمک تلفن های هوشمند نیز امکان پذیر است. این ابزارها به رانندگان برای رانندگی امن تر کمک می کند، مانند چراغ های هوشمند ترافیک که با کمک حسگرها ترافیک را در چهارراه ها کنترل می کنند یا ابزارهای تشخیص تصادف که به راننده تصادف در مسیر را اطلاع می دهد. از نمونه شهرهای هوشمند می توان به بارسلونا و استکهلم اشاره کرد. تغییر در سبک زندگی و سرگرمی ها، ابزارهایی مانند تبلت ها، ابزارهای پوشیدنی و تلفن های هوشمند که قابلیت حسگر دارند و می توانند به سرعت با یکدیگر ارتباط برقرار کنند. اینترنت اشیا امکانات مختلفی در این شهرها ارائه کرده است از جمله، برای پایش وضعیت سلامت افراد ابزارهای پوشیدنی مختلفی ارائه کرده است که وضعیت سلامتی افراد را ضبط و ارسال، از نمونه توانایی ها در شهر هوشمند است. اطلاعاتی در مورد انرژی وضعیت قند و فشارخون، بررسی میزان استرس که برای مراکز مشخص مانند

1. Musaddiq, et al.
2. <http://www.ibm.com/smarterplanet/us/en/>
3. <http://www.libelium.com/>
4. Yun, & Yuxin, B.
5. Rashidi, et al.

اورژانس ارسال می‌شود. در حوزه سلامت ابزارهایی برای اندازه‌گیری میزان فعالیت افراد، میزان دریافت کالری ارائه کرده است، در حوزه کشاورزی هوشمند، ابزارهایی برای اندازه‌گیری پارامترهای محیطی مانند دما، اطلاعات خاک، میزان رطوبت که بلادرنگ اندازه‌گیری و برای فعال‌کننده ارسال می‌شود. (ستی و سارانگی^۱، ۲۰۱۷).

اجزای اینترنت اشیاء برای ایجاد یک سیستم یکپارچه عبارت‌اند از: الف) سخت‌افزار: شامل حسگرها، محرک‌ها و سخت‌افزارهای ارتباطی تعبیه‌شده؛ ب) میان‌افزار: شامل، منابع ذخیره‌سازی و ابزار ارتباطی برای تحلیل داده و ج) ارائه: روشی آسان برای نمایش و تفسیر داده‌ها که می‌تواند برای برنامه‌های متفاوت طراحی شود (گویی و همکاران، ۲۰۱۳).

تکنولوژی RFID موفقیت بزرگ در زمینه علوم ارتباطات است که طراحی میکروچیپ‌ها را برای ارتباطات داده‌ای بی‌سیم ممکن کرد. آن‌ها در کمک به شناسایی خودکار هر وسیله‌ای که دارای یک بارکد الکترونیکی می‌باشند، به کار می‌روند (ولبورن و همکاران^۲، ۲۰۰۹؛ جولز^۳، ۲۰۰۶).

پیشرفت‌های اخیر در فن‌آوری مدارات یکپارچه کم‌مصرف و ارتباطات بی‌سیم، ساخت وسایل کم‌هزینه و کم‌مصرف را برای استفاده در دستگاه‌های حسگر بی‌سیم فراهم کرده است. ترکیب این عوامل، توانایی به‌کارگیری شبکه‌ی حسگری را ایجاد کرده است که دربرگیرنده‌ی تعداد زیادی حسگرهای هوشمند با توانایی جمع‌آوری، پردازش و تحلیل و انتشار اطلاعات ارزشمند می‌باشند. داده‌های حسگرها در میان نودها به اشتراک گذاشته شده و برای تحلیل به یک سیستم مرکزی یا توزیع شده فرستاده می‌شوند. اجزای شبکه‌های حسگر بی‌سیم شامل موارد زیر است:

الف) سخت‌افزار: نوعاً یک ابزار سخت‌افزاری کوچک است که با نام گره از آن نام می‌برند و شامل واسط حسگر، واحد پردازش، واحد فرستنده و گیرنده و منبع تغذیه است.

1. Sethi & Sarangi
2. Welbourne, et al.
3. Juels

ب) بسته ارتباطی: نودها برای بسیاری از برنامه‌های کاربردی به صورت ادهاک استفاده می‌شوند. نودها با یکدیگر ارتباط دارند و ارتباط آن‌ها با ایستگاه اصلی از طریق یک یا چند هاپ فراهم می‌شود.

ج) میان‌افزار: روشی است برای ترکیب زیرساخت سایبری با معماری سرویس‌گرا^۱ و شبکه‌های حسگر که به منظور دسترسی به منابع اطلاعاتی در حسگرهای ناهمگن استفاده می‌شود. ایده اصلی، جداسازی منابعی است که می‌توانند توسط چند برنامه کاربردی استفاده شود.

د) جمع‌آوری امن داده‌ها: یک روش مطمئن و امن برای جمع‌آوری داده از حسگرها، به منظور توسعه عمر شبکه و نیز اطمینان از معتبر بودن داده‌های جمع‌آوری شده مورد نیاز است. از کارافتادن یک گره یک مشکل رایج در شبکه‌های حسگر بی‌سیم است که توپولوژی شبکه باید قابلیت حل آن را داشته باشد. حفظ و تضمین امنیت در برابر تهاجم احتمالی نیز از اهمیت به‌سزایی برخوردار است (گوش و داش^۲، ۲۰۰۸).

توانایی شناسایی منحصربه‌فرد «اشیاء» برای موفقیت اینترنت اشیا حیاتی است. این امر نه تنها امکان شناسایی انحصاری میلیاردها دستگاه را فراهم می‌کند بلکه، امکان کنترل از راه دور آن‌ها از طریق اینترنت را نیز فراهم می‌آورد. ویژگی‌های مهم ایجاد یک آدرس منحصربه‌فرد عبارت‌اند از: انحصاری بودن، قابلیت اطمینان، پایداری و مقیاس‌پذیری. هر عنصری که به شبکه متصل است و نیز عنصری که قصد اتصال به شبکه را دارند باید توسط شناسه انحصاری، مکان و ویژگی‌های خود شناخته شوند. در حال حاضر IPv4 می‌تواند دستگاه‌های حسگر همکار را از لحاظ جغرافیایی شناسایی کند ولی قادر به شناخت جداگانه آن‌ها نیست. خواص جابه‌جاپذیری اینترنت در IPv6 ممکن است برخی مشکلات در شناسایی این دستگاه‌ها را حل کند اگرچه طبیعت ناهمگن گره‌های حسگر، انواع مختلف داده، عملیات هم‌زمان و مشکل تصادم داده‌های دستگاه‌ها در شبکه، مشکلات را در آینده تشدید خواهد کرد. وسایل ذخیره‌سازی و تحلیل داده‌ها و همچنین برنامه‌های مصورسازی داده‌ها از اجزای دیگر اینترنت اشیا است. (گویی و همکاران، ۲۰۱۳). داده‌های موجود در

1. Service Oriented Architecture (SOA)
2. Ghosh, & Das

اینترنت اشیاء می‌توانند به چند نوع تقسیم‌بندی شود: جریان داده RFID، شناسه‌های یکتای آدرس داده‌های توصیفی، داده‌های مکانی، داده‌های محیطی، داده‌های حسگر و غیره. (کوپر و جیمز^۱، ۲۰۰۹).

با پیشرفت تکنولوژی اینترنت و توسعه شبکه اجتماعی، معقولانه است که نسل جدیدی از اینترنت که اینترنت آینده نامیده می‌شود، به‌زودی ظهور یابد. اینترنت اشیاء نیز باید به همراه اینترنت آینده، طراحی مجدد شوند؛ که آن را اینترنت اشیاء در آینده^۲ می‌نامیم. در اینترنت فعلی برای تبادل محتوا از آدرس IP استفاده می‌شود که با زیاد شدن تعداد وسایل در اینترنت اشیاء مشکل‌ساز خواهد بود، (زورزی و همکاران^۳، ۲۰۱۰) برای حل این مشکل از مدل انتشار محتوای^۴ در معماری اینترنت آینده و برای تغییر روش انتقال اطلاعات بین کاربران در اینترنت استفاده می‌شود. به‌عبارت‌دیگر کاربران باید به‌جای روش انتقال، نگران محتوای اطلاعات خود باشند (بلوم و رولی، ۲۰۰۳).

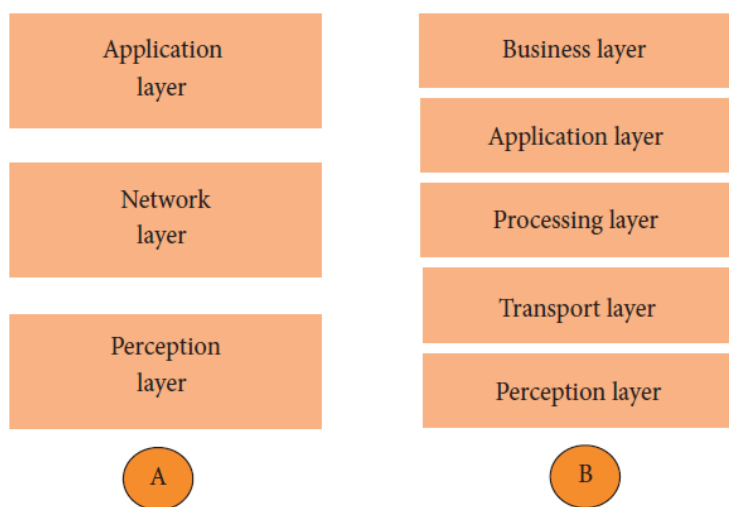
طراحی معماری اینترنت اشیاء از سبک‌های معماری، شبکه و ارتباطات، اشیاء هوشمند، سرویس‌های وب و کاربردها، مدل‌های کسب‌وکار و فرآیندهای مربوطه، پردازش داده مشارکتی، امنیت و غیره تأثیر می‌پذیرد. از دید تکنولوژی، در معماری اینترنت اشیاء باید به توسعه‌پذیری، مقیاس‌پذیری، پیمانه‌پذیر بودن، قابلیت همکاری بین وسایل ناهمگون توجه شود. (ژو و همکاران^۵، ۲۰۱۴)

اتحادیه بین‌المللی ارتباط از راه دور، معماری ۵ لایه را برای اینترنت اشیاء پیشنهاد داده است. شامل: لایه‌های حس‌نمودن، دسترسی، شبکه، میان‌افراز و کاربرد. جیا و همکاران^۶ (۲۰۱۲) و دومینگو^۷ (۲۰۱۲) معماری اینترنت اشیاء سه لایه را پیشنهاد دادند که شامل: لایه ادراک، لایه شبکه، لایه سرویس (یا کاربرد)، (آتزوری و همکاران، ۲۰۱۰) مدل سه لایه‌ای

1. Cooper & James
2. FIOT
3. Zorzi, et al.
4. Name Data Networking (NDN)
5. Jia, et al.
6. Xu, et al.
7. Domingo

را پیشنهاد دادند که شامل لایه کاربرد، لایه شبکه و لایه حسگر بود. لئو و همکاران^۱ (۲۰۱۴)، یک زیرساخت کاربردی از اینترنت اشياء طراحی کردند که شامل لایه فیزیکی، لایه انتقال، لایه میان افزار و لایه کاربرد بود.

معماری پایه ای اینترنت اشياء سه لایه است که در شکل ۲-a نمایش داده شده است که به صورت گسترده، برای تشریح رویکرد اینترنت اشياء به کار می رود لایه ادراکی (که لایه حسگر یا لایه تکنولوژی مرزی^۲ نیز نامیده می شود) پایین ترین لایه است که می تواند لایه فیزیکی یا سخت افزاری نیز در نظر گرفته شود. جمع آوری داده ها در این لایه انجام می شود. لایه میانی، لایه شبکه است که وظیفه اتصال لایه ادراک و لایه کاربردی را دارد؛ بنابراین وسایل می توانند از لایه کاربرد به لایه ادراکی عبور کنند و سیستم ها، برنامه های کاربردی، سرویس ها از لایه کاربرد به لایه ادراکی گذر داده می شوند. لایه کاربرد عموماً قانون مربوط به ارائه خدمات و برنامه های کاربردی است که اطلاعات رسیده شده از طریق دو لایه قبلی را ادغام، تجزیه و تحلیل کند. (پاولراج و همکاران^۳، ۲۰۱۲)



شکل ۲. معماری اینترنت اشياء. (A) معماری سه لایه (B) معماری پنج لایه

در مطالعات جدید، لایه دروازه دسترسی^۱ و لایه میان‌افزار^۲ به معماری سه لایه اضافه شده است، لایه ادراک که لایه سه نیز نامیده می‌شود، وظیفه گرفتن داده را دارد. برخلاف مدل سه لایه اینترنت اشیا که فقط لایه شبکه برای اتصال لایه‌های کاربردی و ادراکی استفاده می‌شوند لایه دروازه دسترسی هم به مدل پنج لایه اضافه شده است که به اتفاق لایه شبکه برای مدیریت ارتباطات در محیط اینترنت اشیا و انتقال پیام‌ها بین اشیا سیستم به کار می‌رود. (پاولراج و همکاران، ۲۰۱۲).

لایه میان‌افزار، لایه دیگری است که در معماری پنج لایه وجود دارد که غالباً برای انعطاف‌پذیری انجمنی بین سخت‌افزارها و برنامه‌های کاربردی به کار می‌رود و در نهایت بالاترین لایه، لایه کاربرد است که تعریف آن، مشابه معماری سه-لایه است (شکل ۲-ب). (ستی و سارنگی، ۲۰۱۷).

معماری سرویس گرا^۳، یک الگوی معماری در طراحی سیستم‌های نرم‌افزاری است که در آن اجزای نرم‌افزار سرویس‌هایی را برای اجزای دیگر، معمولاً بر روی شبکه و از طریق پروتکل ارتباطی، فراهم می‌کنند. اصول سرویس-گرا بودن، بر مبنای استقلال از فروشنده، محصول یا تکنولوژی است. هدف اینترنت اشیا، اتصال وسایل مختلف و ناهمگون در شبکه به یکدیگر است، معماری سرویس گرا می‌تواند برای پشتیبانی از اینترنت اشیا استفاده شود. معماری سرویس گرا با موفقیت در زمینه‌هایی مانند رایانش ابری در شبکه‌های سنسور بی‌سیم و شبکه‌های وسایل سیار^۴ استفاده شده است. معماری سرویس گرا همچنین اجازه می‌دهد که سخت‌افزارها و نرم‌افزارها مجدداً استفاده شوند، زیرا تکنولوژی جدیدی را برای اعمال سرویس خود تحمیل نمی‌کند (ژو و همکاران، ۲۰۱۴). این تکنولوژی اجازه می‌دهد سیستم‌های پیچیده و یکپارچه به برنامه‌هایی با ساختار ساده و اجزای خوب-تعریف‌شده تجزیه شوند (آنزوری و همکاران، ۲۰۱۰).

1. gateway layer
2. middleware
3. service oriented architecture
4. vehicular network

از دید اینترنت اشیا با قابلیت معماری سرویس گرا، معماری چهار لایه اینترنت اشیا در جدول ۱ توسط ژو و همکاران، (۲۰۱۴) پیشنهاد شده است.

جدول ۱. مدل چهار لایه برای اینترنت اشیا

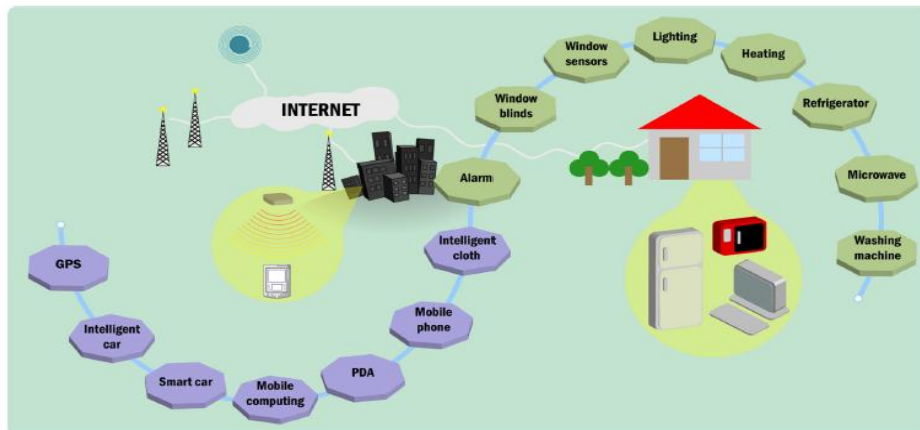
نام لایه	وظیفه لایه
لایه حس نمودن	این لایه مشتمل بر سخت افزارهای (RFID، حسگرها و غیره) برای حس کردن/کنترل دنیای فیزیکی و داده‌های مورد نیاز
لایه شبکه	این لایه پشتیبانی شبکه و انتقال داده در شبکه بی سیم یا باسیم را فراهم می کند.
لایه خدمات	این لایه خدمات را مدیریت و ایجاد می کند. خدمات برای رفع نیازهای کاربران مهیا می شوند.
لایه رابط	این لایه روش‌های تبادل بین کاربران و دیگر برنامه‌ها را فراهم می کند.

اینترنت اشیا و مه داده‌ها (کلان داده‌ها): اینترنت اشیا تعداد بسیار زیادی از حسگرهای شبکه، وسایل و ماشین‌های مختلف را به یکدیگر در دنیای واقعی متصل کرده است. این حسگرها که در زمینه‌های مختلف استفاده می شوند، داده‌های مختلفی را مانند داده‌های محیطی، داده‌های جغرافیایی، داده‌های نجومی و داده‌های منطقی، جمع آوری می کنند. تجهیزات موبایلی امکانات حمل و نقل، امکانات عمومی و لوازم خانگی می توانند از تجهیزات اکتساب داده‌ها باشند که در شکل ۳ نشان داده شده است.

در شهرهای هوشمند داده‌های حجیمی در صنعت، کشاورزی، ترافیک، حمل و نقل، مراقبت پزشکی، ادارات دولتی و خانواده‌ها و غیره تولید می شود، داده‌های حجیم تولید شده در اینترنت اشیا، مشخصه‌های متفاوتی نسبت به داده‌های حجیم معمولی دارد؛ زیرا انواع مختلف داده‌های جمع آوری شده دارای ناهمگونی، تنوع، ویژگی‌های غیر ساخت یافته، نویز و افزونگی بالا می باشند. اگرچه داده‌های کنونی اینترنت اشیا، غالب بر سایر داده‌های حجیم نیست ولی بر اساس پیش‌بینی HP تا سال ۲۰۳۰ تعداد حسگرها به یک میلیارد دستگاه خواهد رسید و داده‌های اینترنت اشیا، بخش مهمی از داده‌های حجیم خواهند شد. اینتل در گزارشی به سه ویژگی داده‌های حجیم اشاره کرده است:

- پایانه‌های فراوان، توده‌های داده را تولید می کنند.
- داده‌های تولید شده به وسیله اینترنت اشیا، معمولاً بی ساختار یا نیمه ساختمند هستند.

- داده‌های اینترنت اشیاء فقط زمانی سودمند هستند که تحلیل شوند.
- معماری شبکه ممکن است به سه لایه تقسیم شود، لایه ادراک، لایه شبکه و لایه کاربرد. در لایه ادراک، جمع‌آوری داده‌ها انجام می‌شود و شامل تجهیزات شبکه‌های حسگر است. لایه شبکه، پاسخگوی انتقال اطلاعات و پردازش است که ارتباط نزدیکی با شبکه حسگر دارد و انتقال راه دور از طریق اینترنت انجام می‌شود. در نهایت لایه کاربرد برنامه‌های کاربردی خاص را پشتیبانی می‌کند.
- بر اساس شاخص‌های اینترنت اشیاء، داده‌های جمع‌آوری شده در اینترنت دارای ویژگی‌های زیر هستند.
- مقیاس بزرگ داده‌ها: در اینترنت اشیاء، تجهیزات جمع‌آوری داده، توزیع شده هستند و ممکن است داده‌های ساده مانند اطلاعات مکان‌ها؛ یا داده‌های چندرسانه‌ای پیچیده مانند نظارت تصویری باشند. با توجه به تقاضاهای تحلیل و پردازش، نه تنها داده‌های فعلی بلکه داده‌های قدیمی که در زمان خاصی اتفاق می‌افتد نیز باید ذخیره شده باشند؛ بنابراین داده‌های تولیدشده در اینترنت اشیاء، داده‌هایی با مقیاس بزرگ شناخته می‌شوند.
- ناهمگونی: به دلیل وجود وسایل مختلف جمع‌آوری داده‌ها، داده‌های مورد نیاز جمع‌آوری شده دارای ساختار مختلف و با ویژگی‌های ناهمگون هستند.
- همبستگی بالا به زمان و فضا: هر وسیله جمع‌آوری داده در یک نقطه جغرافیایی قرار دارد و هر بخش از داده در یک زمان خاص برداشت می‌شود. در تحلیل و پردازش داده‌ها، همبستگی زمان و مکان ویژگی مهم و تأثیرگذاری است.
- داده مؤثر، تنها بخش کوچکی از داده‌های بزرگ است: نویز زیادی ممکن است در طی اکتساب یا انتقال داده‌ها وجود داشته باشد. در میان مجموعه داده‌های به دست آمده به وسیله دستگاه‌های جمع‌آوری داده، تنها مقدار اندکی از داده‌های ناهنجار، با ارزش هستند. برای مثال، در طی جمع‌آوری ویدئوی ترافیک، فقط مقدار اندکی که مربوط به تخلفات رانندگی یا تصادفات می‌شود با ارزش است و بقیه جریان عادی ترافیک را نشان می‌دهند (چن و همکاران، ۲۰۱۴).



شکل ۳. نمایش وسایل مختلف اکتساب داده‌های اینترنت اشياء

هر وسیله در اینترنت اشياء ممکن است یک سیل داده بسازد که حاوی اطلاعات مختلف و ارزشمند است؛ بنابراین اینکه چطور با این داده‌ها رفتار کنیم و چگونه از اطلاعات سودمند آن‌ها استفاده کنیم موضوع چالش‌برانگیزی است که این سال‌ها موضوع بحث است. به علت ویژگی‌های ذکر شده، نوع داده‌های اینترنت اشياء در دسته مه داده‌ها^۱ معرفی می‌شوند (برکویچ و لیائو^۲، ۲۰۱۲). بارانیوک^۳، (۲۰۱۲) عنوان می‌کند که مشکل اینترنت اشياء، از معماری مربوط به حسگرها، به پردازش داده‌ها، ارتباط میان اشياء و حجم محل ذخیره‌سازی داده‌های این اشياء، انتقال یافته است.

برخی روش‌های قدیمی که برای تحلیل و بررسی مه داده‌ها استفاده می‌شوند، عبارت‌اند از: استفاده از روش‌های نمونه‌برداری تصادفی^۴ (ان جی و هان^۵، ۲۰۰۲) چگالش داده^۶ (زانگ و همکاران^۷، ۱۹۹۶)، تقسیم و حل (گاها و همکاران^۸، ۲۰۰۳)، یادگیری افزایشی^۹ (حمودا و

1. big data
2. Berkovich & Liao
3. Baraniuk
4. random sampling
5. NG & Han
6. data condensation
7. Zhang, et al.
8. Guha, et al.
9. incremental learning

کامل^۱، (۲۰۰۴). روش عملیاتی برای حل داده‌های حجیم حسگرها، تحلیل مؤلفه‌های مهم^۲ داده، به‌جای تحلیل تمامی داده‌ها (دینگ و هی^۳، ۲۰۰۴). روش دیگر کاهش الگوهاست^۴. در روش PCA از کاهش فیلدها و در روش PR از کاهش الگوها استفاده می‌شود (ژو و همکاران، ۲۰۱۴).

کشف دانش^۵ و کاربرد آن در اینترنت اشیا: توسعه محتوا و ساختار علم، اساسی‌ترین ویژگی مربوط به بعد شناختی هر علمی است (موئد، گلنزل و اشموج، ۲۰۰۴). برای درک این شناخت باید به استخراج دانش مفید حوزه‌های علمی دست زد. در تعریفی که توربن، مکلین و وترب^۶ (۱۹۹۹) از کشف یا استخراج دانش دارند این است که آن را فرآیند استخراج دانش مفید از میان انبوهی از داده‌ها دانسته‌اند. کشف یا استخراج دانش به دو صورت است: استخراج دانش از پایگاه داده یا داده‌کاوی؛ استخراج دانش از متن‌ها یا متن‌کاوی (هتو، نارنبرگر و پاب^۷، ۲۰۰۵). لازم به ذکر است که پایگاه داده‌ها به‌صورت ساخت یافته^۸، برای تحلیل مقادیر زیادی از داده‌ها و برنامه‌های پردازش خودکار، طراحی شده‌اند؛ درحالی‌که «متن» بی‌شکل، بدون ساختار^۹ و برای خواندن مردم نوشته شده است و علیرغم اینکه تنظیم آن مشکل است، رایج‌ترین وسیله برای تبادل رسمی اطلاعات است (تیلور^{۱۰}، ۲۰۰۶).

داده‌کاوی. در تعریف داده‌کاوی می‌توان بیان کرد که داده‌کاوی یک حوزه میان‌رشته‌ای است که حوزه‌های مختلفی همچون پایگاه داده، آمار، یادگیری ماشینی^{۱۱} و سایر زمینه‌های مرتبط را باهم تلفیق می‌کند (هان و کمبر^{۱۲}، ۲۰۰۶ نقل از رضانی، علیپور حافظی و مؤمنی،

1. Hammouda & Kamel
2. principal component analysis
3. Ding & He
4. pattern reduction
5. knowledge discovery
6. Turban, Mclean & Wetherbe
7. Hotho, Nürnbergger & Paaß
8. structured
9. unstructured
10. Taylor
11. Machine learning
12. Han & Kamber

۱۳۹۳). روش‌های اصلی داده‌کاوی دو دسته هستند: توصیفی^۱ و پیش‌بینانه^۲. وظایف توصیفی، خواص عمومی داده را مشخص می‌کنند. هدف از توصیف، یافتن الگوهایی در مورد داده‌هاست که برای انسان قابل تفسیر باشد. وظایف پیش‌بینانه، به‌منظور پیش‌بینی رفتارهای آینده آن‌ها استفاده می‌شود. منظور از پیش‌بینی به‌کارگیری چند متغیر یا فیلد در پایگاه داده برای پیش‌بینی مقادیر آینده یا ناشناخته دیگر متغیرهای موردعلاقه است (غضنفری، علیزاده و تیمورپور، ۱۳۸۷ نقل از رضوانی، علیپور حافظی و مؤمنی، ۱۳۹۳).

در ادامه نمونه‌هایی از کاربرد داده‌کاوی را، بررسی می‌کنیم.

- داده‌کاوی در تجارت الکترونیک: با کمک داده‌کاوی، الگوهای خاص از بین تراکنش‌های خرید در گذشته، کشف می‌شود و از نتایج حاصله می‌توان برای برنامه‌ریزی و ایجاد سیستم‌های توزیع مؤثر، استفاده کرد (میشرا و همکاران^۳، ۲۰۱۳). رفتارهای کاربران در وب، برای یافتن شباهت‌ها و الگوهای رفتاری گشت‌وگذار، نظارت و تحلیل می‌شود تا شباهت‌ها و الگوی رفتاری فرد، برای وبگردی مشخص شده و سپس وب‌طوری برنامه‌ریزی می‌شود که فرد مطالب موردنیاز خود را بیشتر ملاقات کند (هیر و چی^۴، ۲۰۰۱).

تجارت می‌تواند با داده‌کاوی بر روی ردیابی مسیر حرکت انسان‌ها نیز بهبود یابد. برای مثال می‌توان آگهی تبلیغاتی را در مکان‌هایی که بازدیدکننده بیشتری دارد، نصب کرد. ردیابی داده‌ها را می‌توان با اطلاعاتی همچون جریان وجوه نقد و امور کسب‌وکار تلفیق کرد (پان و همکاران^۵، ۲۰۱۳).

- داده‌کاوی در صنعت: داده‌کاوی کاربردهای زیادی در صنایعی همچون خرده‌فروشی، بانکداری با ارتباطات از راه دور دارد. روش‌های طبقه‌بندی و خوشه‌بندی در این حوزه‌ها اعمال می‌شوند (الگنسی و الراگال^۶، ۲۰۱۴)، یکی از عوامل کلیدی برای سازمان‌های بیمه و بانکداری ارزیابی و بررسی اعتبار وام‌گیرندگان است. چندین روش داده‌کاوی برای حل

1. descriptive
2. predictive
3. Mishra, et al.
4. Heer & Chi
5. Pan, et al.
6. Elgendy & Elragal

مشکل اعتبارسنجی افراد وجود دارد. (اچسیه و هانگ^۱، ۲۰۱۰؛ کمبال و همکاران^۲، ۲۰۱۳). خرده‌فروشان اطلاعات مشتری و تراکنش‌های مربوطه را جمع‌آوری و از این اطلاعات برای بالا بردن دقت تولید برحسب تقاضا، بهینه‌سازی مجموعه، پیشنهاد محصول و رتبه‌بندی موزعین و تولیدکنندگان استفاده می‌کنند (لئو و همکاران، ۲۰۱۴؛ مآب و همکاران^۳، ۲۰۱۴) محققان در این زمینه از روش‌های SVM (دو و همکاران^۴، ۲۰۱۳)، رگرسیون ماشین بردار (لو و ونگ^۵، ۲۰۱۳) یا مدل باس^۶ (لی و همکاران^۷، ۲۰۱۳) برای پیش‌بینی تقاضا استفاده می‌کنند.

- مراقبت‌های بهداشتی: اینترنت اشیاء فرصت‌های جدیدی در زمینه بهبود مراقبت‌های بهداشتی فراهم کرده است. با استفاده از امکانات شناسایی در همه‌جا، حس نمودن و قابلیت‌های ارتباطی همه اشیاء اعم از مردم، تجهیزات، داروها و غیره در سیستم‌های مراقبت‌های بهداشتی می‌توانند به دقت رهگیری و نظارت شوند. همه اطلاعات مربوط به مراقبت‌های بهداشتی اعم از تدارکات، تشخیص، درمان، دارو، مدیریت، مالی و حتی فعالیت روزانه می‌تواند جمع‌آوری، مدیریت و به شکل کارآمد، به اشتراک گذاشته شود. برای مثال ضربان قلب بیمار می‌تواند به کمک حسگرها در زمان‌های مشخص جمع‌آوری و به دفتر پزشکی ارسال شود. با استفاده از وسایل محاسباتی نظیر لپ‌تاپ، تلفن موبایل، تبلت و غیره؛ و دسترسی به اینترنت سیار Wifi, 3G, LTE و غیره، خدمات مربوط به مراقبت‌های بهداشتی می‌تواند مبتنی بر اینترنت اشیاء و به شکل سیار و شخصی ارائه شود. استفاده از خدمات اینترنت همراه می‌تواند به توسعه استفاده از اینترنت اشیاء در منزل کمک کند (ژو و همکاران، ۲۰۱۴). در زمینه مراقبت‌های بهداشتی، داده‌کاوی بسیار محبوب است و استفاده از آن به‌طور فزاینده‌ای افزایش یافته است (چن و همکاران^۸، ۲۰۱۰؛ لئو و همکاران، ۲۰۱۴؛

1. Hsieh & Hung
2. Kambal, et al.
3. Maaß, et al.
4. Du, et al.
5. Lu & Wang
6. Boss model
7. Lee, et al.
8. Chen, et al.

چن و همکاران^۱، ۲۰۱۳؛ لئو و همکاران^۲، ۲۰۱۳؛ چن^۳، ۲۰۱۴). در حوزه مراقبت‌های بهداشتی تجزیه و تحلیل‌های مربوط به داده‌های پرت، خوشه‌بندی و قوانین انجمنی می‌تواند اعمال شود (سان و ردی^۴، ۲۰۱۳؛ کینکاد^۵، ۱۹۹۸). اطلاعات مربوط به درمان می‌تواند برای کم کردن هزینه‌ها و استفاده از داروهای بهتر استفاده شود (برلی و زوپان^۶، ۲۰۰۸؛ لئو و همکاران^۷، ۲۰۱۳) همچنین داده‌کاوی برای تشخیص بیماران پرهزینه (سیلور و همکاران^۸، ۲۰۰۱) به کار می‌رود و بر حجم بزرگ داده‌ها مشتمل بر میلیون‌ها، نسخه، عملیات و مراحل درمان برای تشخیص الگوهای نامعمول و کشف تقلب اعمال می‌شود (کوه و تان^۹، ۲۰۱۱).

- داده‌کاوی در مدیریت شهری: در زمینه سرویس‌های عمومی، داده‌کاوی برای کشف نیازهای عمومی و بالا بردن کارایی تصمیم‌گیری‌های مدیریت شهری و در سیستم‌های خدمات خودکار، برای کاهش مخاطرات اجتماعی استفاده می‌شود. الگوریتم‌های طبقه‌بندی، خوشه‌بندی، تحلیل سری‌های زمانی در مورد حل این مسائل به کار می‌روند. از مزایای دولت الکترونیک می‌توان به کیفیت خدمات دولتی، صرفه‌جویی در هزینه، مشارکت سیاسی گسترده و ارائه سیاست‌ها و برنامه‌های مؤثر اشاره کرد (هلبگ و همکاران^{۱۰}، ۲۰۰۹؛ ون و همکاران^{۱۱}، ۲۰۱۲) و همچنین باعث افزایش ارتباط میان شهروندان و سازمان‌های دولتی و بهبود اعتماد سیاسی می‌شود (چادویک و می^{۱۱}، ۲۰۰۳). سیستم مدیریت حوادث شهری از روش‌های داده‌کاوی برای تهیه ارزیابی جامع از تأثیر بلایای طبیعی بر محصولات کشاورزی و دسته‌بندی مناطق آسیب‌دیده، کمک به دولت برای آماده‌سازی در مقابل بلایا و تخصیص منابع استفاده می‌کند (پنگ و همکاران^{۱۲}، ۲۰۱۱). تحلیل داده‌ها و درک اینکه

1. Chen, et al.
2. Liu, et al.
3. Chen et al.
4. Sun & Reddy
5. Kincade
6. Bellazzi & Zupan
7. Silver, et al.
8. Koh & Tan
9. Helbig, et al.
10. Wan, et al.
11. Chadwick & May
12. Peng, et al.

کدام یک از معیارهای زندگی و خدمات شهری منجر به تصمیم خروج از شهرنشینی می‌شود. پژوهشگران می‌توانند پیش‌بینی کنند که کدام یک از ساکنان، احتمالاً از شهر مهاجرت می‌کنند (چن^۱، ۲۰۱۳). یکی از چالش‌های دولت و مأموران اجرای قانون این است که داده‌های مربوط به افزایش جرم و جنایت را تحلیل کنند (چن و همکاران^۲، ۲۰۰۴) محققان تکنیک داده‌کاوی مکانی را برای پیدا کردن قوانین انجمنی، میان نقاط جرم و جنایت خیز و مکان زندگی پیشنهاد می‌کنند (هیوانگ^۳، ۲۰۱۳). برخی از محققان دیگر از الگوریتم خوشه‌بندی K-Means برای پیدا کردن الگوهای جرم و جنایت و تکنیک‌های نیمه نظارت‌شده یادگیری برای کشف دانش و کمک به افزایش دقت در پیش‌بینی‌ها استفاده کردند (شیام^۴، ۲۰۰۶). همچنین داده‌کاوی برای تشخیص تغییر هویت، از طریق تحلیل اطلاعات مردم از جمله اسم، آدرس، تاریخ تولد و شماره امنیت اجتماعی (ونگ و همکاران^۵، ۲۰۰۴) و الگوهای ساختاری شناسایی نشده از طریق شبکه‌های جنایی (چن و همکاران^۶، ۲۰۰۷) به کار رود. در سیستم حمل‌ونقل، داده‌کاوی برای پالایش نقشه بر اساس جی.پی.اس و پیشنهاد مسیر مسافرت و محل مسافرت و پیدا کردن مکان‌های جالب برای سفر استفاده می‌شود (چن و همکاران، ۲۰۱۵).

- داده‌کاوی بر روی تجهیزات اینترنت اشیاء: وسعت زیاد مناطق پوشش‌دهنده، عمر کوتاه باطری حسگرها، احتمال خراب‌شدن نودها در طول کار و تعداد زیاد حسگرها مسائلی هستند که در بیشتر کاربردهای حسگرها با آن مواجه هستیم. طراحی و عملیاتی کردن این چنین شبکه بزرگی، نیاز به معماری مقیاس‌پذیر و استراتژی مدیریت دارد. در مجموع در چنین محیط‌هایی با محدودیت انرژی مواجه هستیم و امکان شارژ باطری حسگرها وجود ندارد؛ بنابراین نیاز به الگوریتم‌های حساس به مصرف انرژی^۷ داریم که بتوانند عمر حسگرها را

1. Chen
2. Chen, et al.
3. Huang
4. Shyam
5. Wang, et al.
6. Chen, et al.
7. energy-aware

افزایش دهد (عباسی و یونس^۱، ۲۰۰۷). در الگوریتم‌های خوشه‌بندی سنتی از روش‌های جستجو برای پیدا کردن نتایج خوشه‌بندی استفاده می‌شود (مانند k-means)، در الگوریتم‌های فرا ابتکاری جدید، از روش‌های تصادفی برای تخمین نتایج خوشه‌بندی استفاده می‌شود (بلوم و رولی، ۲۰۰۳). این روش مخصوصاً برای داده‌های حجیم و پیچیده شانس بیشتری برای یافتن نتایج مطلوب‌تر دارند. در خوشه‌بندی اینترنت اشیاء دقت خوشه‌بندی تنها هدف نیست. دو هدف عمده دیگر، به ترتیب، ساخت خوشه‌های مناسب بر اساس شرایط جدید و ارضای وضعیت‌های مختلف مسئله است. زمینه دیگر در زمینه خوشه‌بندی توزیع شده (یونس و فهمی^۲، ۲۰۰۴)، طولانی کردن عمر حسگرهای شبکه بی‌سیم است. از آنجا که حسگرهای بی‌سیم یکی از وسایل پر کاربرد در اینترنت اشیاء است الگوریتم‌های خوشه‌بندی زیادی در این زمینه توسعه یافته‌اند. خوشه‌بندی بهترین روش برای ادغام قابلیت‌های شناسایی، حس و تحریک این دستگاه‌ها است. یکی از الگوریتم‌های مشهور در زمینه کاهش مصرف انرژی این دستگاه‌ها سلسله‌مراتب خوشه‌بندی مصرف کم انرژی^۳ است. روش دیگر خوشه‌بندی، ساخت درخت پوشای بیشینه و استفاده از یک الگوریتم حریصانه برای تنظیم دقیق پارتیشن و تنظیم تعداد خوشه‌ها برای کمینه کردن فاصله مجموع بین نودهای حسگر و سرخوشه‌های حسگرها است. (عباسی و یونس، ۲۰۰۷).

برای افزایش سرعت انتقال داده در حسگرهای بی‌سیم، الگوریتم خوشه‌بندی دومرحله‌ای^۴ برای کاهش میانگین فاصله بین آن‌ها و اضافه کردن یک لینک کمکی برای تمام حسگرها استفاده می‌شود (چوی و همکاران^۵، ۲۰۰۴) برای پیشگیری از مشکل افزایش بیش از حد پرس‌وجو، الگوهای شناسه منحصر به فرد^۶ برای شناسایی اجزای اینترنت اشیاء به کار می‌رود. لی و همکاران^۷ (۲۰۰۵) الگوی طبقه‌بندی مبتنی بر درخت دودویی را برای حل

1. Abbasi, & Younis
2. Younis & Fahmy
3. Low Energy Adaptive Clustering Hierarchy (LEACH)
4. two phase clustering
5. Choi, et al.
6. unique item identifier
7. Lee, et al.

این مسئله، پیشنهاد دادند که با کمک درخت دودویی و بررسی تعدادی عدد در ابتدای بیت‌ها، قادر است، نوع وسیله را سریعاً تشخیص دهد. دو بررسی جامع درباره قوانین انجمنی و الگوهای ترتیبی انجام شده است (هان و همکاران^۱، ۲۰۰۳؛ ژائو و بومیک^۲، ۲۰۰۳) چندین مطالعه مانند داده کاوی زمانی^۳ (گالوشکا و همکاران^۴، ۲۰۰۶)، حفظ حریم خصوصی در اینترنت اشیاء، استخراج میزان استفاده از وب، بیوانفورماتیک، تشخیص نفوذ (چن و رن^۵، ۲۰۱۲) در این باره انجام شده است.

متن کاوی تکیه‌اش روی پیدا کردن دانش جدید^۶ از متن است (معمولاً دانشی که به‌طور ضمنی در سندهای متنی است) درحالی‌که بازایی اطلاعات، سندهایی که بیشترین ارتباط را دارند، می‌یابد. هدف اولیه متن کاوی، بازایی اطلاعات از متون ساخت‌نیافته و همچنین ارائه دانش به‌صورت خالص برای کاربران در یک‌شکل چکیده است (آنانیودو و مک‌نوآت^۷، ۲۰۰۶ نقل از رضانی، علیپور حافظی و مؤمنی، ۱۳۹۳). متن کاوی، قادر ساختن استفاده‌کنندگان برای جمع‌آوری، ذخیره، تفسیر و کشف دانش موردنیاز برای پژوهش و آموزش مؤثر و نظام‌مند است. در کل می‌توان بیان داشت، هدف متن کاوی، کشف اطلاعات از قبل ناشناخته است که هنوز کسی نمی‌داند و بنابراین مستند نشده است (دورری، گرستل و سیفرت^۸، ۱۹۹۹ نقل از رضانی، علیپور حافظی و مؤمنی، ۱۳۹۳). متن کاوی شامل سه فعالیت بزرگ است؛ «بازایی اطلاعات» که بازایی متون مربوط به سؤال استفاده‌کنندگان است؛ «خلاصه اطلاعات» که شناختن و استخراج نکات ریز متون که مربوط به سؤال هستند؛ و «داده کاوی» که رابطه مستقیم یا غیرمستقیم بین قسمت‌های اطلاعات استخراجی از متون را می‌یابد (توماس، مک‌نوآت و آنانیودو^۹، ۲۰۱۱ نقل از رضانی، علیپور حافظی و مؤمنی، ۱۳۹۳).

1. Han, et al.
2. Zhao & Bhowmick
3. temporal data mining
4. Galushka, et al.
5. Chen & Ren
6. new knowledge
7. Ananiadou & McNaught
8. Dorre, Gerstl & Seiffert
9. Thomas, McNaught & Ananiadou

متن کاوی کاربردهای متعددی دارد؛ از جمله، بررسی روندهای علمی با خوشه‌بندی و دسته‌بندی مفهومی متون علمی و صفحات وبی و غیره. دورری و همکارانش بر این باورند، از آنجا که حدود ۸۵ تا ۹۰ درصد از داده‌های شرکت‌ها و سازمان‌ها با روش‌های متعارف داده کاوی قابل کشف دانشی نیستند و از اسناد متنی تشکیل یافته‌اند؛ چه بسا متن کاوی وظیفه پیچیده تری نسبت به داده کاوی سنتی دارد. چون مستلزم برخورد کردن با اطلاعات متنی غیر ساختاریافته است که ذاتاً مبهم هستند (دورری، گرسنل و سیفرت، ۱۹۹۹؛ هتو، نارنبرگر و پاب، ۲۰۰۵).

علاوه بر این، از ابزارهای متن کاوی برای کمک به محققان در زمینه‌های مختلف استفاده می‌شود. به عنوان مثال، یک اخترشناس که علائم اشعه ایکس را در ناحیه‌ای از فضا کشف می‌کند ممکن است تمایل داشته باشد که در پیشینه‌های برخط جستجو کند تا ببیند که آیا تاکنون هیچ نوع اشعه مادون قرمزی در همان ناحیه کشف شده است یا خیر؟ یک زیست‌شناس که دارای فهرستی از ۱۰۰ ژن است که در مقالات مختلف شناسایی شده‌اند، ممکن است تمایل داشته باشد که به سراغ پژوهش‌های منتشرشده موجود برود و در جستجوی مقالاتی باشد که درباره عملکرد این ژن‌ها توضیح داده باشند (شارما، ۲۰۰۵).

نتیجه‌ای که از تعاریف فوق حاصل می‌گردد، این است که «متن کاوی»، شاخه‌ای از علم داده کاوی است که جهت کشف دانش ضمنی در فایل‌های متنی غیرساختارمند، به صورت خودکار و با استفاده از فنون وام گرفته‌شده از دانش داده کاوی عمل می‌کند.

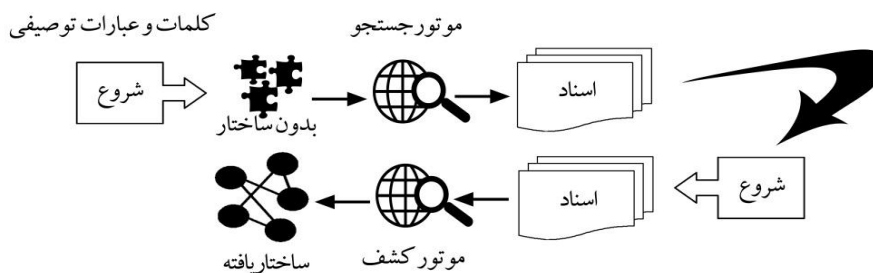
فرآیند متن کاوی: یکی از دلایل استفاده از روش‌های داده کاوی بر روی اسناد متنی ایجاد ساختار در آن‌ها است. ایجاد ساختار می‌تواند دسترسی کاربر را به یک مجموعه از اسناد به طور چشمگیری ساده کند. از ساختارهای دسترسی معروف می‌توان به فهرست‌های کتابخانه‌ای و نمایه‌های کتاب اشاره کرد؛ اما نمایه‌های طراحی شده به صورت دستی مسئله‌ای زمان‌بر است و این نمایه‌ها معمولاً به روز نیستند و برای انتشارات جدید و یا اطلاعات زودبه‌زود در حال تغییر هستند، مانند اطلاعات روی اینترنت، قابل استفاده نیستند. در واقع،

روش متن کاوی اغلب همان روش‌های داده‌کاوی هستند که با تغییراتی برای متون استفاده می‌شوند (کانتاردزیک^۱، ۲۰۱۱).

در شیوه‌های سنتی پردازش اطلاعات غیرساختاریافته، فرد یا افرادی در نظر گرفته می‌شدند تا با مطالعه داده‌های متنی به تحلیل آن‌ها بپردازند. متن کاوی فرآیندی است که به کمک آن اطلاعات ارزشمند و مفید نهفته در داده‌های غیرساختاریافته، استخراج می‌شود. در متن کاوی تحلیل محتوای داده‌های غیرساختاریافته بر اساس آنالیز کمی متن صورت می‌پذیرد. چنانچه اطلاعات کمی با سیستم‌های هوشمند مورد بررسی و تحلیل قرار گیرد، می‌توان تا اندازه‌ای به بررسی کیفی متون پرداخت و یک سیستم (Man- Machine) Interactive ایجاد نمود (هیرست، ۱۹۹۹).

فرآیند متن کاوی (شکل ۶) شامل مراحل زیر است (کریمی، ۱۳۸۶):

- آماده‌سازی داده‌های متنی؛
- نادیده گرفتن کلمات بدون ارزش اطلاعاتی^۲؛
- ریشه‌یابی کلمات^۳؛
- جستجوی کلمات مهم و کلیدی متن؛
- آنالیز و تحلیل متن‌های مختلف بر اساس کلمات کلیدی (همانند یافتن متن‌های مشابه، خوشه‌بندی کردن متن‌ها، تعیین موضوع متون و غیره) که از آن می‌توان به‌عنوان لایه کاربردی فرآیند متن کاوی نام برد.



شکل ۶. فرآیند متن کاوی

1. Kantardzic
2. stop word
3. stemming

در آماده‌سازی داده‌های متنی، چگونگی ذخیره‌سازی داده‌های غیرساختاریافته، نمایه‌سازی آن‌ها، تبدیل قالب آن‌ها به قالب‌های قابل‌استفاده در نرم‌افزارها موردبررسی قرار می‌گیرد. نادیده گرفتن کلمات بدون محتوای با ارزش اطلاعاتی همانند: از، به، با تا و غیره (که معمولاً بیشترین میزان تکرار در متن را به خود اختصاص می‌دهند) در مرحله دوم موردبررسی قرار می‌گیرد. از آنجا که در متن کاوی به دنبال کلمات با ارزش اطلاعاتی هستیم، این دسته از کلمات را کنار خواهیم گذاشت. ریشه‌یابی کلمات در مرحله سوم موردبررسی قرار می‌گیرد. کلماتی مانند «دانش»، «دانشمند»، «دانایی» و غیره کلمات هم‌ریشه به حساب می‌آیند. در بررسی یک متن توسط فرآیند متن کاوی به تعیین ریشه و کلمات هم‌ریشه پرداخته می‌شود تا در بررسی آماری کلمات متن، این دسته از کلمات در کنار هم موردبررسی قرار گیرد. تعیین کلمات کلیدی و مهم متن، مهم‌ترین بخش فرآیند متن کاوی است و دقت تمامی کاربردهای متن کاوی تا حدود زیادی به دقت در تعیین کلمات کلیدی بستگی خواهد داشت. برای تعیین کلمات کلیدی و مهم متن، الگوریتم‌های مختلفی بر اساس تکنیک‌های ریاضی ارائه شده است. روش‌های «تی.اف-آی.دی.اف»، «تکرار کلمات»، «تکرار کلمات نرمال شده»، «وزن‌دهی بر اساس جستجوی کاربر» و سایر الگوریتم‌های مشتق شده از این الگوریتم‌ها در تعیین کلمات کلیدی و مهم به کار گرفته می‌شوند. وجه اشتراک تمامی این الگوریتم‌ها آن است که تعیین کلمات کلیدی یک متن بر اساس آنالیز یک مجموعه متون مرتبط باهم تعیین می‌گردد. کاربردهای مختلفی را می‌توان از فرآیند متن کاوی انتظار داشت. بر اساس کلمات کلیدی می‌توان با الگوریتم‌های خوشه‌بندی «k میانگین»، «سلسله‌مراتبی» و یا سایر الگوریتم‌ها به خوشه‌بندی متون پرداخت. با استفاده از خوشه‌بندی متون می‌توان متن‌های مشابه را شناسایی و دسته‌بندی نمود. با تعیین واژه‌های کلیدی و استفاده از تکنیک‌های مختلف وزن‌دهی به جملات و کلمات متن می‌توان یک چکیده از متن تهیه و ارائه نمود (زعفریان، ۱۳۸۵).

چالش‌های پژوهشی در زمینه اینترنت اشیا:

- اینترنت اشیا شامل نودهای بسیار زیادی خواهد بود که هر کدام از این‌ها محتوایی تولید می‌کنند که باید بدون توجه به مکان انتقال به دست کاربر مجاز رسانده شوند. در حال حاضر

هر نود با IPv4 و توسط ۴ بایت شناسایی می‌شود که می‌دانیم تعداد باقیمانده آن به‌زودی به پایان می‌رسد. IPv6 می‌تواند ۱۰^{۳۸} وسیله را پوشش دهد که نگرانی در این مورد را کاهش می‌دهد با این وجود اشیاء متحرک مسئله مهمی است که باید برای آن راهکاری پیشنهاد شود. در گذشته راهکارهایی در زمینه مدیریت جابجایی اشیاء پیشنهاد شده است (آکیلدیز و همکاران^۱، ۲۰۰۴). ولی صحت این راهکارها باید اثبات شود چون برخی از این‌ها مشکلاتی در زمینه مقیاس‌پذیری و سازگاری با محیط‌های ناهمگون دارند (آتروری و همکاران، ۲۰۱۰).

- هدف اصلی لایه انتقال، تضمین قابلیت اطمینان و کنترل ازدحام است. در اینترنت سنتی، پروتکل TCP در لایه انتقال برای قابلیت اطمینان ارتباطات استفاده می‌شد (لاکشمین و مادهو^۲، ۱۹۹۷). واضح است که به دلایل زیر TCP برای اینترنت اشیاء ناکافی است و باید مفهوم جدیدی برای لایه انتقال ارائه شود:

۱. تنظیم ارتباط: در TCP در سه مرحله^۳ انجام می‌شود که این عمل در بسیاری از اشیاء اینترنت اشیاء که حجم اطلاعات کمی را منتقل می‌کنند، لازم نیست. علاوه بر این حمل داده‌های اضافی در مرحله تنظیم ارتباط باعث هدر رفت انرژی می‌شود.

۲. کنترل ازدحام: TCP مسئول کنترل ازدحام پایانه‌هاست. از آنجا که بیشتر وسایل در اینترنت اشیاء به شکل بی‌سیم خواهند بود و این محیط برای TCP چالش‌برانگیز است بنابراین این مسئله ممکن است باعث بروز مشکلاتی در کارایی سیستم شود (دمیرکل و همکاران^۴، ۲۰۰۶). به‌علاوه چون حجم داده‌های تبادل شده از یک ایستگاه بسیار ناچیز است. کنترل ازدحام TCP بی‌فایده خواهد بود.

۳. بافر نمودن داده‌ها: TCP نیاز به ذخیره داده‌ها در حافظه بافر در مبدأ و مقصد دارد. مدیریت این بافرها در مصرف انرژی باطری‌های اشیاء بسیار پرهزینه است. در نتیجه TCP نمی‌تواند به‌صورت کارا این انتقال را برای اینترنت اشیاء کنترل کند و در حال حاضر

1. Akyildiz, et al.
2. Lakshman, & Madhow
3. three way handshakes
4. Demirkol, et al.

جایگزینی برای آن ارائه نشده است و نیاز به گسترش پژوهش‌ها در این زمینه است (آتزوری و همکاران، ۲۰۱۰)

- در بسیاری از پژوهش‌ها، گزارش‌ها، سخنرانی‌ها، مسائل مربوط به اینترنت اشیا، داده‌های حجیم، گرید هوشمند، محاسبات ابری را با یکدیگر قرار می‌دهند؛ زیرا این موضوعات و محصولات آن‌ها ممکن است به صورت معناداری بر زندگی ما تأثیر بگذارد، این‌ها موضوعات بسیار نزدیک و جدانشدنی از یکدیگر می‌باشند. داده‌های حجیم که معمولاً از طریق اینترنت اشیا به دست می‌آیند و شبکه‌های هوشمند، محاسبات ابری به منابع قدرتمند برای انجام محاسبات و ذخیره‌سازی داده‌های حجیم نیاز دارند، در نتیجه اینکه چگونه آن‌ها را با اینترنت اشیا ادغام نماییم از موضوعات پژوهش‌های آینده است (دمیرکل و همکاران، ۲۰۰۶).

- کارهای پژوهشی مختلفی و موفق‌تری در این زمینه، درباره معماری از دیدگاه‌های مختلف با محوریت شبکه‌های حسگر بی‌سیم، اینترنت اشیا، رایانش ابری انجام شده است، معماری دیگری با محوریت کاربرانی که قصد استفاده از این داده‌ها و وسایل را برای توسعه نرم‌افزارهای جدید دارند، پیشنهاد شود (گویی و همکاران، ۲۰۱۳).

- پروتکل‌های پیاده‌سازی شده در حسگرهای اینترنت اشیا نقش کلیدی در تحقق کامل نقش آن‌ها در ارتباط با جهان خارج دارند. برای کارایی این سیستم‌ها، داشتن پروتکل MAC با کارآمدی انرژی و پروتکل‌های مسیریابی مناسب، بسیار حیاتی است. پروتکل‌های MAC متفاوتی برای دامنه‌های مختلف پیشنهاد شده است، مانند طرح‌های پروتکل بدون برخورد (TDMA)، پروتکل با بهره‌وری ترافیک کم (CSMA)، پروتکل بدون برخورد ولی نیازمند مداربندی در نود (FDMA). هیچ‌کدام از آن‌ها به‌عنوان یک استاندارد شناخته نشده‌اند و با «اشیا» مختلف این سناریو به‌هم‌ریخته‌تر می‌شود که نیاز به پژوهش‌های بیشتر دارد (گویی و همکاران، ۲۰۱۳).

- یک حسگر به دلایل مختلفی ممکن است خراب شود که در این صورت شبکه باید خود را با شرایط موجود وفق دهد و اجازه مسیریابی از مسیرهای مختلف را بدهد. پروتکل‌های چندپایه در شبکه‌های موبایلی ادهاک و شبکه‌های حسگر بی‌سیم استفاده می‌شوند. این پروتکل‌ها بر اساس نوع استفاده در برنامه‌ها، عموماً به سه دسته: داده محور، مبتنی بر مکان و

سلسله مراتبی تقسیم می‌شوند. ملاحظه اصلی این پروتکل‌ها مصرف انرژی است. در مورد اینترنت اشیا، باید توجه نمود که تعداد هاپ‌هایی که در سناریوی چندهاپه می‌توان استفاده کرد، محدود است. باوجود این محدودیت، ایجاد پروتکل‌های مسیریابی موجود باید با حداقل تغییرات، اجرایی شوند (گوبی و همکاران، ۲۰۱۳).

- امنیت در اینترنت اشیا به دلیل حجم کوچک تجهیزات استفاده شده، دسترسی فیزیکی به حسگرها، فعال‌کننده‌ها و اشیا و همچنین استفاده سیستم‌ها از بستر ارتباطی بی‌سیم مهم است. مسئله امنیت اهمیت بیشتری می‌یابد. چون وجود خطاهای دائمی یا موقت عادی است و خطا یعنی آسیب‌پذیری که می‌تواند مورد سوءاستفاده مهاجمان قرار گیرد. برنامه‌های اینترنت اشیا باید با توجه به حضور حملات امنیتی به کار خود به‌درستی ادامه داده و این حملات را به‌خوبی دفع کنند. ممکن است نیاز به دانلود کدهای جدیدی باشد که این خود یکی از زمینه‌های پژوهشی است. همچنین سیستم وقتی برای اولین بار مستقر می‌شود باید خود را با حملات پیش‌بینی نشده انطباق دهد؛ بنابراین سیستم‌ها باید قابلیت بالایی برای شناسایی حملات داشته باشند و در صورت شناسایی حمله واکنش مناسب توسط خود دستگاه برای دفع حمله اتفاق بیفتد. برای دفع اثر یک حمله باید شناسایی حمله، تشخیص نوع حمله، انجام عمل متقابل و ترمیم انجام شود ولی باید ظرفیت پائین دستگاه‌ها در نظر گرفته شود. بسیاری از راه‌حل‌های کنونی نیاز به محاسبات پیچیده و حافظه بالا دارند؛ بنابراین راه‌حل‌ها برای اینترنت اشیا در این زمینه جزو چالش‌های پژوهشی است (ستی و همکاران، ۲۰۱۷).

- از دید حریم خصوصی: از آنجا که برخی فناوری‌ها، از دوربین مداربسته برای تشخیص و شناخت موقعیت‌ها استفاده می‌کنند، نگرانی حریم خصوصی کاربران، بیشتر مردم را ناراحت ساخته است. نگرانی مشابه در زمینه اعمال تکنیک‌های داده‌کاوی برای تحلیل داده‌های اینترنت اشیا است. برای مثال شرکت‌ها امروزه می‌توانند به‌راحتی داده‌های مختلف مشتریان را از منابع و وسایل مختلف جمع‌آوری کنند و از تکنیک‌های داده‌کاوی برای پیدا کردن اطلاعات برای افزایش فروش استفاده کنند ولی واقعیت این است بسیاری از مشتریان

علاقه‌ای ندارند که حریم خصوصی‌شان، مانند رفتار خریدشان، جمع‌آوری شود (تی سای و همکاران^۱، ۲۰۱۴).

- پشتیبانی بلادرنگ: وسایل اینترنت اشياء به برنامه‌های کاربردی مختلفی نیاز دارند که برخی از آن‌ها به عملکرد بلادرنگ سیستم نیاز دارند. به طوری که حسگر و انجام عمل مناسب بلادرنگ انجام شود؛ بنابراین سیستم‌عاملی که این چالش را مرتفع کند از دیگر زمینه‌های پژوهشی است (موسادیق و همکاران، ۲۰۱۸).

- مدیریت بافر شبکه: بافر شبکه از ابزارهای مهم عمل‌کننده‌های اینترنت اشياء است که برای اختصاص حداقل میزان حافظه محدود برای ارسال بسته‌ها در شبکه از موضوعات پژوهش است (موسادیق و همکاران، ۲۰۱۸).

بحث و نتیجه‌گیری

شکی نیست که مه داده‌ها گرایش دنیای آینده است، زیرا قریب به تمام وسایل، اطلاعات خود را حداقل یک‌بار بر روی اینترنت آپلود می‌کنند که برای استفاده از فناوری‌های داده‌کاوی، باید نمونه‌برداری، فشرده‌سازی، یادگیری افزایشی، فیلتر کردن را در مورد آن‌ها به کار گیریم. در این زمینه استفاده از متدهای فازی و چندهدفه، پتانسیل لازم را برای اضافه کردن قدرت تصمیم‌گیری در حسگرها دارند. برای تصمیم‌گیری در زمینه نقش مردم و اشياء، فناوری‌های داده‌کاوی باید با فناوری‌های اینترنت اشياء در این زمینه تلفیق شوند. داده‌کاوی شامل کشف الگوهای بدیع، جالب و بالقوه کاربردی از داده‌ها و اعمال الگوریتم‌ها برای استخراج اطلاعات مخفی از آن‌هاست. در این مقاله ضمن معرفی اینترنت اشياء، درباره تجهیزات، معماری، رابطه آن با داده‌های حجیم و کاربردهای داده‌کاوی بر روی آن‌ها بحث شده است. موضوعاتی مانند ارتباط بین وسایل، رایانش ابری، شبکه‌های اجتماعی، تکنیک‌های مختلف داده‌کاوی، مشکل انرژی در وسایل، استفاده از گرید هوشمند و هوشمندسازی سیستم‌ها می‌تواند از موضوعات پژوهشی در زمینه اینترنت اشياء باشد.

1. Tsai, et al.

منابع

- رمضانی، هادی، علیپور حافظی، مهدی و مؤمنی، عصمت. (۱۳۹۳). نقشه‌های علمی: فنون و روش‌ها. *ترویج علم*، ۵(۶)، ۵۳-۸۴.
- غفارزادگان، مریم. (۱۳۹۲). کشف ساختار درونی مطالعات خلاقیت به روش متن‌کاوی. پایان‌نامه کارشناسی ارشد علم اطلاعات و دانش‌شناسی، دانشگاه علامه طباطبائی، تهران.

References

- Abbasi, A. A., & Younis, M. (2007). A survey on clustering algorithms for wireless sensor networks. *Computer communications*, 30(14-15), 2826-2841.
- Ashton, K. (2009). That 'internet of things' thing. *RFID journal*, 22(7), 97-114.
- Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15), 2787-2805.
- Baraniuk, R. G. (2011). More is less: signal processing and the data deluge. *Science*, 331(6018), 717-719.
- Bélissent, J. (2010). Getting clever about smart cities: New opportunities require new business models. *Cambridge, Massachusetts, USA*.
- Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: current issues and guidelines. *International journal of medical informatics*, 77(2), 81-97.
- Berkovich, S., & Liao, D. (2012, July). On clusterization of big data streams. In *Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications* (p. 26). ACM.
- Bin, S., Yuan, L., & Xiaoyi, W. (2010, April). Research on data mining models for the internet of things. In *2010 International Conference on Image Analysis and Signal Processing* (pp. 127-132). IEEE.
- Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM computing surveys (CSUR)*, 35(3), 268-308.
- Chadwick, A., & May, C. (2003). Interaction between States and Citizens in the Age of the Internet: "e-Government" in the United States, Britain, and the European Union. *Governance*, 16(2), 271-300.
- Chen, H., Chung, W., Qin, Y., Chau, M., Xu, J. J., Wang, G., ... & Atabakhsh, H. (2003, May). Crime data mining: an overview and case studies. In *Proceedings of the 2003 annual national conference on Digital*

- government research* (pp. 1-5). Digital Government Society of North America.
- Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: a general framework and some examples. *computer*.
- Chen, L., & Ren, G. (2012). The research of data mining technology of privacy preserving in sharing platform of internet of things. In *Internet of Things* (pp. 481-485). Springer, Berlin, Heidelberg.
- Chen, M. (2013). Towards smart city: M2M communications with software agent intelligence. *Multimedia Tools and Applications*, 67(1), 167-178.
- Chen, M., Gonzalez, S., Leung, V., Zhang, Q., & Li, M. (2010). A 2G-RFID-based e-healthcare system. *IEEE Wireless Communications*, 17(1), 37-43.
- Chen, M., Ma, Y., Wang, J., Mau, D. O., & Song, E. (2013, October). Enabling comfortable sports therapy for patient: a novel lightweight durable and portable ECG monitoring system. In *2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013)* (pp. 271-273). IEEE.
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications*, 19(2), 171-209.
- Choi, W., Shah, P., & Das, S. K. (2004, August). A framework for energy-saving data gathering using two-phase clustering in wireless sensor networks. In *The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004.* (pp. 203-212). IEEE.
- Da Xu, L., He, W., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on industrial informatics*, 10(4), 2233-2243.
- Demirkol, I., Ersoy, C., & Alagoz, F. (2006). MAC protocols for wireless sensor networks: a survey. *IEEE Communications Magazine*, 44(4), 115-121.
- Ding, C., & He, X. (2004, July). K-means clustering via principal component analysis. In *Proceedings of the twenty-first international conference on Machine learning* (p. 29). ACM.
- Domingo, M. C. (2012). An overview of the Internet of Things for people with disabilities. *Journal of Network and Computer Applications*, 35(2), 584-596.
- Du, X. F., Leung, S. C., Zhang, J. L., & Lai, K. K. (2013). Demand forecasting of perishable farm products using support vector machine. *International journal of systems Science*, 44(3), 556-567.
- Elgendy, N., & Elragal, A. (2014, July). Big data analytics: a literature review paper. In *Industrial Conference on Data Mining* (pp. 214-227). Springer, Cham.

- Galushka, M., Patterson, D., & Rooney, N. (2006). Temporal data mining for smart homes. In *Designing Smart Homes* (pp. 85-108). Springer, Berlin, Heidelberg.
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7), 1645-1660.
- Guha, S., Meyerson, A., Mishra, N., Motwani, R., & O'Callaghan, L. (2003). Clustering data streams: Theory and practice. *IEEE transactions on knowledge and data engineering*, 15(3), 515-528.
- Hammouda, K. M., & Kamel, M. S. (2004). Efficient phrase-based document indexing for web document clustering. *IEEE Transactions on knowledge and data engineering*, 16(10), 1279-1296.
- Han, J., Cheng, H., Xin, D., & Yan, X. (2007). Frequent pattern mining: current status and future directions. *Data mining and knowledge discovery*, 15(1), 55-86.
- Heer, J., & Chi, E. H. (2001, April). Identification of web user traffic composition using multi-modal clustering and information scent. In *Proc. of the Workshop on Web Mining, SIAM Conference on Data Mining* (pp. 51-58).
- Helbig, N., Gil-García, J. R., & Ferro, E. (2009). Understanding the complexity of electronic government: Implications from the digital divide literature. *Government Information Quarterly*, 26(1), 89-97.
- Hsieh, N. C., & Hung, L. P. (2010). A data driven ensemble classifier for credit scoring analysis. *Expert systems with Applications*, 37(1), 534-545.
- Peng, Y., Zhang, Y., Tang, Y., & Li, S. (2011). An incident information management framework based on data integration, data mining, and multi-criteria decision making. *Decision Support Systems*, 51(2), 316-327.
- Jia, X., Feng, Q., Fan, T., & Lei, Q. (2012, April). RFID technology and its applications in Internet of Things (IoT). In *2012 2nd international conference on consumer electronics, communications and networks (CECNet)* (pp. 1282-1285). IEEE.
- Juels, A. (2006). RFID security and privacy: A research survey. *IEEE journal on selected areas in communications*, 24(2), 381-394.
- Kambal, E., Osman, I., Taha, M., Mohammed, N., & Mohammed, S. (2013, August). Credit scoring using data mining techniques with particular reference to Sudanese banks. In *2013 INTERNATIONAL CONFERENCE ON COMPUTING, ELECTRICAL AND ELECTRONIC ENGINEERING (ICCEEE)* (pp. 378-383). IEEE.
- Kincade, K. (1998). Data mining: digging for healthcare gold. *Insurance & Technology*, 23(2), 2-7.

- Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. *Journal of healthcare information management*, 19(2), 65.
- Lakshman, T. V., & Madhow, U. (1997). The performance of TCP/IP for networks with high bandwidth-delay products and random loss. *IEEE/ACM transactions on networking*, 5(3), 336-350.
- Lee, H., Kim, S. G., Park, H. W., & Kang, P. (2014). Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach. *Technological Forecasting and Social Change*, 86, 49-64.
- Lee, Y. H., Kim, H. J., Roh, B. H., Yoo, S. W., & Oh, Y. C. (2005, December). Tree-based classification algorithm for heterogeneous unique item ID schemes. In *International Conference on Embedded and Ubiquitous Computing* (pp. 1078-1087). Springer, Berlin, Heidelberg.
- Liu, C. H., Yang, B., & Liu, T. (2014). Efficient naming, addressing and profile services in Internet-of-Things sensory environments. *Ad Hoc Networks*, 18, 85-101.
- Liu, J., Pan, J., Wang, Y., Lin, D., Shen, D., Yang, H., ... & Cao, X. (2013). Component analysis of Chinese medicine and advances in fuming-washing therapy for knee osteoarthritis via unsupervised data mining methods. *Journal of Traditional Chinese Medicine*, 33(5), 686-691.
- Liu, J., Wan, J., He, S., & Zhang, Y. (2014). E-healthcare supported by big data. *ZTE Communications*, 12(3), 46-52.
- Liu, J., Wang, Q., Wan, J., Xiong, J., & Zeng, B. (2013). Towards Key Issues of Disaster Aid based on Wireless Body Area Networks. *KSII Transactions on Internet & Information Systems*, 7(5).
- Liu, Q., Wan, J., & Zhou, K. (2014). Cloud manufacturing service system for industrial-cluster-oriented application. *Journal of Internet Technology*, 15(3), 373-380.
- Lu, C. J., & Wang, Y. W. (2010). Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting. *International Journal of Production Economics*, 128(2), 603-613.
- Maaß, D., Spruit, M., & de Waal, P. (2014). Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniques. *Decision Analytics*, 1(1), 4.
- Moed, H. F., Glänzel, W., & Schmoch, U. (Ed.) (2004). *Handbook of quantitative science and technology research*. The use of publication and patent statistics in studies of S&T systems. Dordrecht (the Netherlands): Kluwer Academic Publishers.
- Musaddiq, A., Zikria, Y. B., Hahm, O., Yu, H., Bashir, A. K., & Kim, S. W. (2018). A survey on resource management in IoT operating systems. *IEEE Access*, 6, 8459-8482.

- Nath, S. V. (2006, December). Crime pattern detection using data mining. In *2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops* (pp. 41-44). IEEE.
- Ng, R. T., & Han, J. (2002). CLARANS: A method for clustering objects for spatial data mining. *IEEE Transactions on Knowledge & Data Engineering*, (5), 1003-1016.
- Padhy, N., Mishra, D., & Panigrahi, R. (2012). The survey of data mining applications and feature scope. *arXiv preprint arXiv:1211.5723*.
- Pan, G., Qi, G., Zhang, W., Li, S., Wu, Z., & Yang, L. T. (2013). Trace analysis and mining for smart cities: issues, methods, and applications. *IEEE Communications Magazine*, 51(6), 120-126.
- Paulraj, D., Swamynathan, S., & Madhaiyan, M. (2012). Process model-based atomic service discovery and composition of composite semantic web services using web ontology language for services (OWL-S). *Enterprise Information Systems*, 6(4), 445-471.
- Peng, Y., Zhang, Y., Tang, Y., & Li, S. (2011). An incident information management framework based on data
- Rashidi, P., Cook, D. J., Holder, L. B., & Schmitter-Edgecombe, M. (2010). Discovering activities to recognize and track in a smart environment. *IEEE transactions on knowledge and data engineering*, 23(4), 527-539.
- Sethi, P., & Sarangi, S. R. (2017). Internet of things: architectures, protocols, and applications. *Journal of Electrical and Computer Engineering*, 2017.
- Silver, M., Sakata, T., Su, H. C., Herman, C., Dolins, S. B., & O Shea, M. J. (2001). Case study: how to apply data mining techniques in a healthcare data warehouse. *Journal of healthcare information management*, 15(2), 155-164.
- Sundmaeker, H., Guillemin, P., Friess, P., & Woelfflé, S. (2010). Vision and challenges for realising the Internet of Things. *Cluster of European Research Projects on the Internet of Things, European Commission*, 3(3), 34-36.
- Taylor, P. (2006). *From patient data medical knowledge: the principles & practice of health informatics*. Massachusetts: Black Well.
- Thornton, D., Mueller, R. M., Schoutsen, P., & Van Hillegerberg, J. (2013). Predicting healthcare fraud in medicaid: a multidimensional data model and analysis techniques for fraud detection. *Procedia technology*, 9, 1252-1264.
- Tsai, C. W., Lai, C. F., Chiang, M. C., & Yang, L. T. (2013). Data mining for internet of things: A survey. *IEEE Communications Surveys & Tutorials*, 16(1), 77-97.

- Turban, E., Mclean, E., & Wetherbe, J. (1999). *Information technology for management: making connections for strategic advantage*. New York: Willey & sons.
- Wan, J., Li, D., Zou, C., & Zhou, K. (2012, October). M2M communications for smart city: An event-based architecture. In *2012 IEEE 12th International Conference on Computer and Information Technology* (pp. 895-900). IEEE.
- Wang, G., Chen, H., & Atabakhsh, H. (2004). Automatically detecting deceptive criminal identities.
- Welbourne, E., Battle, L., Cole, G., Gould, K., Rector, K., Raymer, S., ... & Borriello, G. (2009). Building the internet of things using RFID: the RFID ecosystem experience. *IEEE Internet computing*, 13(3), 48-55.
- Xie, J. L. (2004). *Mobility Management in Next Generation All-IP Based Wireless Systems* (Doctoral dissertation, Georgia Institute of Technology).
- Yin, H., Sun, Y., Cui, B., Hu, Z., & Chen, L. (2013, August). LCARS: a location-content-aware recommender system. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 221-229). ACM.
- Younis, O., & Fahmy, S. (2004, March). Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach. In *IEEE INFOCOM 2004* (Vol. 1). IEEE.
- Yun, M., & Yuxin, B. (2010, June). Research on the architecture and key technology of Internet of Things (IoT) applied on smart grid. In *2010 International Conference on Advances in Energy Engineering* (pp. 69-72). IEEE.
- Zhang, T., Ramakrishnan, R., & Livny, M. (1996, June). BIRCH: an efficient data clustering method for very large databases. In *ACM Sigmod Record* (Vol. 25, No. 2, pp. 103-114). ACM.
- Zhao, Q., & Bhowmick, S. S. (2003). Sequential pattern mining: A survey. *ITechnical Report CAIS Nanyang Technological University Singapore*, 1, 26.
- Zorzi, M., Gluhak, A., Lange, S., & Bassi, A. (2010). From today's intranet of things to a future internet of things: a wireless-and mobility-related view. *IEEE Wireless communications*, 17(6), 44-51.