

کاربرد هوش مصنوعی در حوزه کلان داده: مروری بر تحقیقات اخیر

محمد رضا وظیفه^۱، سهیلا سارانی^۲

^۱ استادیار دانشگاه آزاد اسلامی واحد زاهدان.

^۲ کارشناسی ارشد مهندسی نرم افزار، دانشگاه آزاد اسلامی واحد زاهدان.

نام نویسنده مسئول:

محمد رضا وظیفه

چکیده

در حالی که داده‌ها به سرعت در محیط دیجیتال جریان پیدا می‌کنند، هوش مصنوعی و کلان داده روش‌های موثری برای کشف دانش از این انبوه داده را فراهم می‌کنند. با این وجود تمایز بین اطلاعات متفاوتی که دائماً در حال جمع‌آوری اند، دشوار است. هوش مصنوعی بیش از شش دهه وجود داشته و در کاربرد‌های مختلف مورد استفاده واقع شده است. در حالیکه کلان داده در سال‌های اخیر مورد توجه بیشتری واقع شده است و امروزه در حوزه تحقیقاتی، استفاده از هوش مصنوعی در کلان داده‌ها از جایگاه ویژه‌ای برخوردار است. این مقاله با بیان ویژگی‌های کلان داده و جایگاه هوش مصنوعی به معرفی هوش تجاری می‌پردازد. در این مقاله همچنین مروری بر تحقیقات اخیر در حوزه هوش مصنوعی و کلان داده‌ها ارائه شده است.

واژگان کلیدی: هوش مصنوعی، کلان داده، هوش تجاری.

مقدمه

عصر دیجیتال در اهتزاز است. به راستی جهان چگونه تغییر خواهد کرد؟ مقدار داده‌هایی که تولید می‌کنیم هر سال دو برابر می‌شود. هر دقیقه صدها هزار جستجو در موتورهای جستجو و پست در شبکه‌های اجتماعی تولید می‌کنیم. این داده‌ها حاوی اطلاعاتی هستند که نشان دهنده افکار و احساسات ما انسان‌ها هستند. به زودی، چیزهایی که در اطراف ما هستند، احتمالاً حتی لباسهای ما نیز با اینترنت متصل خواهند شد. برآورد شده است که طی ۱۰ سال آینده ۱۵۰ میلیارد شبکه حسگر، یعنی حدود ۲۰ برابر بیشتر از افراد روی زمین، وجود خواهد داشت و مقدار داده‌ها هر ۱۲ ساعت دو برابر می‌شوند. در چنین شرایطی بسیاری از شرکت‌ها در حال تلاش برای تبدیل این داده‌ها به پول کلان هستند [1].

در حال حاضر هوش مصنوعی بیش از سایر تکنولوژی‌ها به اتوماسیون کلان داده کمک می‌کند و همواره پیشرفت‌ها در زمینه هوش مصنوعی و کاربرد آن در تجزیه و تحلیل کلان داده در حال توسعه است. الگوریتم‌ها اکنون می‌توانند زبان‌ها و الگوهای دست‌نویس را به اندازه انسان‌ها و حتی برخی بهتر از آن بشناسند. آن‌ها قادر به توصیف محتویات عکس‌ها و فیلم‌ها هستند. امروزه ۷۰٪ از معاملات مالی توسط الگوریتم‌ها انجام می‌شوند. بر این اساس در عرض ۱۰-۲۰ سال آینده حدود نیمی از شغل‌های امروز توسط الگوریتم‌های تهدید می‌شوند [1]. همه چیز هوشمند خواهد شد؛ به زودی ما از گوشی‌های هوشمند، خانه‌های هوشمند، کارخانه‌های هوشمند و شهرهای هوشمند به ملت‌ها و سیاره هوشمند خواهیم رسید و این هوشمندی خطر جدی برای بشریت است، خطری که احتمالاً خطرناک‌تر از سلاح‌های هسته‌ای است. بنابراین هوش مصنوعی و ترکیب آن با کلان داده‌ها از جمله مباحث ویژه‌ای هستند که می‌توانند تعیین‌کننده ادامه راه بشریت باشند.

کلان داده‌ها

کلان داده یک اصطلاح کلیدی است که برای هر روشی که با پردازش مقدار زیادی از داده‌ها از جمله ضبط، انتقال، ذخیره‌سازی، جمع‌آوری، جستجو، تجزیه و تحلیل، تجسم، امنیت و حریم خصوصی سر و کار دارد، بکار برده می‌شود. اندازه کلان داده‌ها به طور مداوم در حال تغییر است؛ از ترابایت در سال ۲۰۰۵ و پتابایت در سال ۲۰۱۰ به اگزابایت^۱ یا زتابایت^۲ در سال ۲۰۱۷ رسیده است. کلان داده معمولاً با مقدار داده‌هایی که برای پردازش آن در یک زمان قابل تحمل، به فراتر از یک رایانه معمولی نیاز است، تعریف می‌شود [2]. از جمله ویژگی‌های کلان داده که با اصطلاح "3V" نشان داده می‌شوند، عبارتند از [3]:

- حجم^۳: به میزان تولید داده‌ها اشاره دارد. به میزان بهره‌نمایی و تاثیر اجتماعی از داده‌ها
 - تنوع^۴: به انواع مختلف داده‌های تولید شده اشاره دارد. مشاهدات از دیدگاه‌های مختلف، می‌تواند تنوع داده‌ها را به همراه داشته باشد و در نتیجه الگوهای مفید آسان‌تر یافت می‌شوند. به عنوان مثال، شناسایی ببرها با لنز دوربین عادی در یک جنگل ممکن است دشوار باشد. در حالی‌که شناسایی ببرها با عکس مادون قرمز، بسیار ساده‌تر است، زیرا درجه حرارت بدن آنها بسیار بیشتر از پس‌زمینه است. یا به عنوان مثال دیگر اگر جنسیت مشتری اشتباه در سوپرمارکت ثبت شده باشد، ممکن است سوپرمارکت کوپن رژلب را به یک مرد ارسال کند.
 - سرعت^۵: به سرعت تولید داده‌ها اشاره دارد. داده‌ها به صورت مداوم از تعاملات اجتماعی، مانیتور سنسورها و فعالیت‌های تجاری ایجاد می‌شوند، بنابراین سرعت تولید داده‌ها از جمله ویژگی‌های حائز اهمیت کلان داده است.
- علاوه بر سرعت، تنوع و حجم، درستی^۶ (عدم اطمینان و اعتبار)، ۷ چهارم است که در تحقیقات دیگر به آن اشاره شده و با نیاز فزاینده‌ای به اطلاعات بی‌درنگ و سفارشی مورد توجه قرار می‌گیرد [4].

تجزیه و تحلیل کلان داده‌ها در شرکت‌ها همچنان در حال رشد است. نظرسنجی‌ها و مطالعات اخیر، که از خدمات مشاوره، در سال‌های ۲۰۱۵ تا ۲۰۱۷ از ۲۰۰ مدیر عامل فناوری اطلاعات و کسب و کار در سراسر جهان منعکس می‌شود نشان دهنده افزایش استفاده شرکت‌ها از کلان داده‌ها است. نمودار ۱ تصویب داده‌های بزرگ از ۲۰۱۵ تا ۲۰۱۷ را نشان می‌دهد. [5].

¹ Exabyte

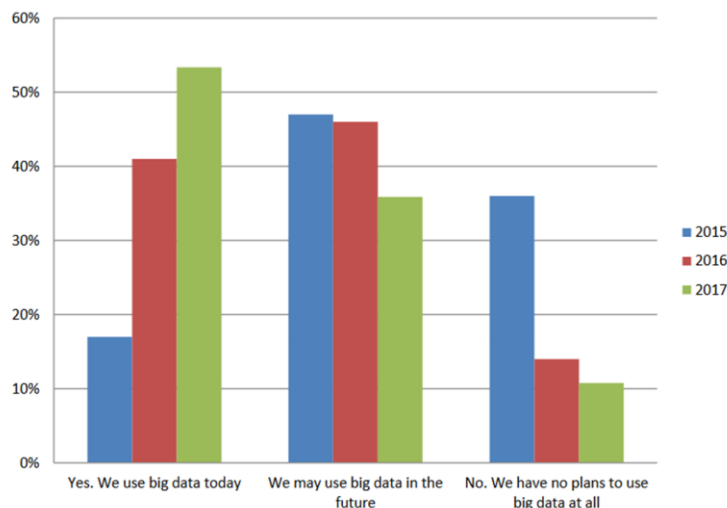
² Zettabyte

³ Volume

⁴ Variety

⁵ Velocity

⁶ Veracity



نمودار ۱: تصویب کلان داده‌ها از ۲۰۱۵ تا ۲۰۱۷ [5]

همانطور که نمودار ۱ نشان می‌دهد، بررسی نشان داد که تعداد شرکت‌هایی که در سال ۲۰۱۷ در مقایسه با سال‌های گذشته، از کلان داده استفاده می‌کنند، بیشتر است و بیش از ۵۰ درصد از سازمان‌های مورد بررسی از کلان داده استفاده می‌کنند. همچنین بیش از ۳۵ درصد آنها گفته‌اند که احتمالاً این کار را در آینده انجام خواهند داد.

هوش مصنوعی و هوش تجاری

تعریف خاصی از هوش مصنوعی وجود ندارد و اصطلاحات هوش مصنوعی برای اولین بار در دهه ۱۹۵۰ معرفی شد. به طور کلی می‌توان هوش مصنوعی را به عنوان توانایی یک ماشین برای یادگیری از طریق تجربه، تنظیم با ورودی‌های جدید و انجام وظایف مانند انسان معرفی کرد. امروزه هوش مصنوعی با پیشرفت سریع فن‌آوری‌های کلان داده (مانند بهبود قابلیت ذخیره سازی محاسبات و پردازنده‌های فوق سریع) و با قدرت و توانایی دسترسی کلان داده تجدید حیات می‌یابد [6].

هوش تجاری یک اصطلاح چتری است که برای استراتژی‌ها، فن‌آوری‌ها و سیستم‌های اطلاعاتی که توسط شرکت‌ها برای استخراج از داده‌های بزرگ و مختلف بر اساس زنجیره ارزش و دانش مربوط به پشتیبانی از طیف گسترده‌ای از کسب و کار عملیاتی، تاکتیکی و استراتژیک تصمیمات بکار می‌رود [7].

هوش تجاری (BI)^۷ به مجموعه‌ای از تکنیک‌ها و ابزارهایی گفته می‌شود که به تبدیل داده‌های بزرگ^۸ از منابع متفاوت به اطلاعات معنی‌دار کمک می‌کند تا تصمیم‌گیری را پشتیبانی و عملکرد سازمانی را بهبود بخشد. امروزه هوش تجاری به عنوان یک نیروی محرک اصلی برای عملکرد سازمانی محسوب می‌شود و اکثر سازمانها برای تجزیه و تحلیل داده‌ها از هوش تجاری استفاده می‌کنند. از لحاظ فنی، فرایندهای کسب و کار عبارتند از [8]: فرآیند استخراج، تبدیل^۹ و فرایند بار (ETL)^{۱۰}، انبار داده^{۱۱}، پردازش تحلیلی آنلاین (OLAP)^{۱۲}، داده‌کاوی، مدل تصمیم‌گیری و تجسم.

هوش تجاری موجب کاهش پراکندگی اطلاعات، تعامل کاربر، دسترسی آسان به اطلاعات، انتشار اطلاعات به موقع و تصمیم‌گیری در ارتباط با تغییر سازمانی در کسب و کار می‌شود. هوش تجاری بر روی هر دو ضرورت و مدیریت اطلاعات بر تصمیم‌گیری تاثیر می‌گذارد. همچنین مدیریت اطلاعات، اطلاعات ساختاری را برای تصمیم‌گیری فراهم می‌کند [9]. شکل ۱ ارتباط بین هوش تجاری، مدیریت اطلاعات و تصمیم‌گیری را نشان می‌دهد.

⁷ Business Intelligence (BI)

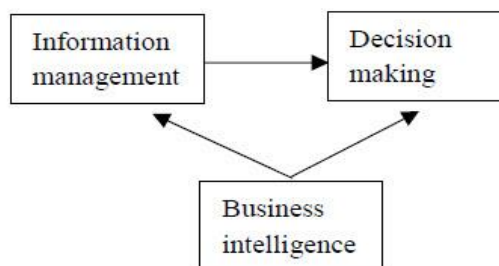
⁸ Big data

⁹ Transformation

¹⁰ Extract & Transform & Load (ETL)

¹¹ Data warehouse

¹² Online analytical processing (OLAP)



شکل ۱: ارتباط بین هوش تجاری، مدیریت اطلاعات و تصمیم‌گیری [9]

مروری بر کاربردهای اخیر هوش مصنوعی در کلان داده‌ها

امروزه هوش مصنوعی از مراقبت‌های بهداشتی تا جنگ کاربرد دارد. به مردم در انتخاب موسیقی و کتاب‌ها کمک می‌کند. در پیگیری رزومه، تعیین اعتبار حساب، حتی کیفیت عکس‌های تلفن همراه و به طور خلاصه، در تصمیمات موثر بر زندگی فرد می‌تواند مورد استفاده واقع شود. استفاده از هوش مصنوعی در کلان داده‌ها نقش مهمی در تعیین اولویت‌های مصرف‌کننده ایفا می‌کنند. بطوری که با دستیابی به نتایج معنی‌دار، داده‌های درست به دست می‌آیند. سیستم‌های هوش مصنوعی، به ویژه آنهایی که به تکنولوژی ماشین مجهز می‌شوند، می‌توانند نتایج قابل توجهی از طریق حذف سریع مجموعه‌های داده‌های بزرگ به دست آورند و این منجر به تعیین تغییرات ساختاری در مدل‌های رفتار مصرف‌کننده و استراتژی‌های بازاریابی می‌شود [10]. در ادامه برخی از تحقیقاتی که اخیراً در حوزه هوش مصنوعی و کلان داده، مرور می‌شوند.

تو و دیگران [11] در سال ۲۰۱۹ مقاله‌ای با عنوان یک روش ذخیره‌سازی خوشه‌بینه برای کلان داده‌های زمان واقعی در اینترنت اشیا ارائه دادند. در این روش ابتدا، وزن بلوک‌های داده در هر دوره دسترسی تاریخی^{۱۳} با محاسبه موقع دسترسی داده‌ها محاسبه می‌شود و فرکانس‌های دسترسی بلوک داده در دوره بعدی به وسیله وزن پیش‌بینی می‌شود. در مرحله بعدی، برای بهبود بهره‌وری دسترسی به داده‌ها و استفاده از منابع و همچنین کاهش هزینه‌های ذخیره‌سازی کپی، کپی بلوک داده به صورت پویا تنظیم و ذخیره می‌شود. در نهایت، نتایج تجربی نشان می‌دهد که هزینه ذخیره‌سازی روش پیشنهادی ۷۰٪ کمتر از روش‌های سنتی است، به این معنی که روش پیشنهادی به طور موثر سرعت دسترسی به داده‌ها را بهبود می‌بخشد، فضای ذخیره‌سازی را کاهش می‌دهد و بار ذخیره‌سازی را متعادل می‌کند.

گارسیا گیل^{۱۴} و دیگران [12] در سال ۲۰۱۹ مقاله‌ای با عنوان فعال کردن داده هوشمند: فیلترینگ نویز در طبقه بندی کلان داده‌ها ارائه دادند در این مقاله بر تاثیر نویز در کیفیت داده‌ها در مسئله طبقه بندی تاکید شده و دو روش پیش پردازش کلان داده‌ها برای حذف نویز پیشنهاد شده است؛ یک فیلتر موثر همگن و یک فیلتر موثر ناهمگن با تاکید خاص در مقیاس پذیری و ویژگی‌های عملکرد آنها. نتایج به دست آمده نشان می‌دهد که این پیشنهادات متخصص را قادر می‌سازد تا به طور موثر یک مجموعه داده هوشمند از هر مسئله طبقه بندی کلان داده به دست آورد.

لو^{۱۵} [13] در سال ۲۰۱۹ مقاله‌ای با عنوان تجزیه و تحلیل کلان داده‌ها برای شناسایی غیرقانونی زباله ساخت و ساز: مطالعه موردی هنگ کنگ ارائه داد. این مقاله با هدف بهبود مدیریت دفع زباله‌های ساختمانی در هنگ کنگ، از روش شناسایی مجموعه‌های غیرمستقیم تخلیه از طریق استخراج داده‌های موجود که شامل بیش از ۹ میلیون پرونده دفع زباله در سال‌های ۲۰۱۱ تا ۲۰۱۷ استفاده می‌کند. در این روش با استفاده از شاخص‌های رفتاری و به روزآوری تحلیل‌های کلان داده‌ها، رانندگان احتمالی تخریب غیرقانونی شناسایی شدند. این مقاله شناسایی رفتار غیرمستقیم تخریب کمک می‌کند و به جامعه پژوهشی جهانی به درک ارزش کلان داده‌ها، به خصوص برای مبارزه با جرم و جنایت شهری کمک می‌کند.

خزیمطلق و دیگران [14] در سال ۲۰۱۸ مقاله‌ای با عنوان تحلیل پوششی داده و کلان داده ارائه دادند. از جمله ویژگی‌های این منبع مطالعاتی عبارت است از:

- یک چارچوب جدید برای مقابله با تجزیه و تحلیل پوششی داده‌ها با مقیاس وسیع پیشنهاد شده است.
- روش ارائه شده زمان‌های محاسباتی را برای اندازه‌گیری میزان عملکرد مجموعه‌های داده بزرگ کاهش می‌دهد.

¹³ Historical

¹⁴ García-Gil

¹⁵ Lu

- روش ارائه شده، سریع ترین روش تجزیه و تحلیل اطلاعات موجود برای کلان داده است.
 - ارائه چارچوب انتخابی برای تحلیل پوششی داده در مقیاس وسیع.
 - پیاده سازی روش آسان است.
- سناسو^{۱۶} [15] در سال ۲۰۱۷ مقاله ای با عنوان انبار داده با تکنولوژی کلان داده برای آموزش عالی ارائه داد. در این مقاله مقایسه ای از انبار داده سنتی و انبار داده مدرن ارائه شده است. جدول ۱ مقایسه‌ای از مشخصات انبار داده سنتی و انبار داده مدرن ارائه می‌دهد.

جدول ۱: مقایسه مشخصات انبار داده سنتی و انبار داده مدرن [15]

مشخصات	انبار داده سنتی	انبار داده مدرن
هدف	- تجمیع داده‌های جمع آوری شده برای یک حوزه خاص کسب و کار - پشتیبانی از فرایند تصمیم گیری	- پردازش ساختار یافته، نیمه ساختار یافته و داده های بدون ساختار از منابع متنوع با حجم اطلاعات بیش از توانایی ابزارهای سنتی برای - ذخیره، مدیریت و تجزیه و تحلیل آنها.
منبع داده	معمولا تراکنش‌ها در نظر گرفته می‌شوند و از پایگاه‌های داده عملیاتی استفاده می‌شود.	منابع مختلف با انواع داده های (اجتماعی، رسانه، حسگرها، بلاگ، تصویر، صدا)
حوزه	ساختارهای یکپارچه داده برای پشتیبانی از اطلاعات کسب و کار (BI) و پردازش تحلیل آنلاین (OLAP)	تجزیه و تحلیل و کشف دانش از حجم زیادی از داده با حجم، سرعت، تنوع و صحت مشخص
معماری	با تمرکز بر فرآیندهای استخراج، تبدیل و بارگذاری (ETL). رویکرد ستاره ای راه حل مناسب برای معماری است.	معماری به مسئله بستگی دارد و هنوز هیچ معماری مرجع یا استاندارد وجود ندارد.
تکنولوژی	تکنولوژی به بلوغ رسیده و ابزارهای تست شده در اپلیکیشن های رایگاه و مجاز زیادی در دسترس است.	فناوری هنوز در حال رشد است Hadoop یکی از نرم افزارهای منبع باز است که برای ذخیره سازی و پردازش داده های توزیع شده از مجموعه داده های بزرگ استفاده می‌کند.
کاربر نهایی	تحلیلگران و یا مدیران ارشد کسب و کارها که نیاز به دانش خاصی از فناوری ها یا اکتشاف داده ها ندارند.	دانشمندان داده با دانش در فن آوری ها، الگوریتم ها، ریاضیات و آمار

نتیجه گیری

کلان داده ها، الگوریتم های پیشرفته، افزایش توان محاسباتی و ذخیره سازی، امروزه سیستم های هوش مصنوعی را به یک عنصر تعبیه شده در سیستم های دیجیتال تبدیل کرده است و به طور خاص تاثیر زیادی بر تصمیم گیری های انسانی دارند. اگر چه تکنولوژی های جدید مبتنی بر هوش مصنوعی می توانند پایان انسان باشند اما با همکاری درست بین انسان و ماشین نباید نگران تهدید ناشی از هوش مصنوعی باشیم. در نتیجه، نیاز به تلاش بیشتر محققان برای بررسی و درک پیامدهای هوش مصنوعی برای تصمیم گیری و کمک به پیشرفت نظری و موفقیت عملی خصوصا در حوزه کلان داده وجود دارد.

منابع و مراجع

- [1] D. Helbing (ed.), (2019), Will Democracy Survive Big Data and Artificial Intelligence?, © Springer International Publishing AG, part of Springer Nature 2019, https://doi.org/10.1007/978-3-319-90869-4_7
- [2] Li Da Xu & Lian Duan, (2018), Big data for cyber physical systems in industry 4.0: a survey, Enterprise Information Systems, DOI: 10.1080/17517575.2018.1442934
- [3] De Mauro, A., M. Greco, and M. Grimaldi. 2016. "A Formal Definition of Big Data Based on Its Essential Features.", Library Review 65 (3): 122–135. doi:10.1108/LR-06-2015-0061
- [4] Ali Reza Alaei, Susanne Becken and Bela Stantic, (2017), Sentiment Analysis in Tourism: Capitalizing on Big Data, Journal of Travel Research, DOI: 10.1177/0047287517747753
- [5] Watson, H. J. (2019). Update Tutorial: Big Data Analytics: Concepts, Technology, and Applications. Communications of the Association for Information Systems, 44, pp-pp. <https://doi.org/10.17705/1CAIS.04421>
- [6] Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. International Journal of Information Management, 48, 63–71
- [7] Darren Edge, Jonathan Larson, Christopher White, (2018), Bringing AI to BI: Enabling Visual Analytics of Unstructured Data in a Modern Business Intelligence Platform, ACM 978-1-4503-5621-3/18/04...\$15.00, <https://doi.org/10.1145/3170427.3174367>
- [8] Tanaporn Panrungsri, Esther Sangiamkul, (2017), Business Intelligence Model for Disaster Management: A Case Study in Phuket, Thailand, WiPe Paper – Operational applications and perspectives, Proceedings of the 14th ISCRAM Conference – Albi, France, pages:727-738
- [9] Shah J. Miah, Shahadat KhanShaheb Ali,(2017), Aalysis of interaction between Business in Telligence and SMES: Learn from each other, JISTEM - Journal of Information Systems and Technology Management, Vol. 14, No. 2, May/Aug., 2017 pp. 151-168
- [10] M. Sigala et al. (eds.), Big Data and Innovation in Tourism, Travel, and Hospitality, © Springer Nature Singapore Pte Ltd. 2019, https://doi.org/10.1007/978-981-13-6339-9_13
- [11] Li Tu Shuai Liu, Yan Wang, Chi Zhang , Ping Li, (2019), An optimized cluster storage method for real-time big data in Internet of Things, © Springer Science+Business Media, LLC, part of Springer Nature 2019, The Journal of Supercomputing ,<https://doi.org/10.1007/s11227-019-02773-1>
- [12] Diego García-Gil, Julián Luengo , Salvador García , Francisco Herrera, (2019), Enabling Smart Data: Noise filtering in Big Data classification,Information Sciences 479 (2019) 135–152.
- [13] Weisheng Lu, (2019), Big data analytics to identify illegal construction waste dumping: A Hong Kong study, Resources, Conservation & Recycling 141 (2019) 264–272.
- [14] Dariush Khezrimotlagh , Joe Zhu , Wade D. Cook , Mehdi Toloo , Data Envelopment Analysis and Big Data, European Journal of Operational Research (2018), doi: <https://doi.org/10.1016/j.ejor.2018.10.044>
- [15] Leo Willyanto Santoso, (2017), Yulia, Data Warehouse with Big Data Technology for Higher Education, 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia , Procedia Computer Science 124 (2017) 93–99.