

عنوان مقاله: پایگاه داده برداری چیست؟ بررسی کاربردها، نحوه کار و آینده Vector Database

نویسنده مقاله: تیم فنی نیک‌آموز

تاریخ انتشار: ۶ اردیبهشت ۱۴۰۳

منبع: [/https://nikamooz.com/what-is-a-vector-database](https://nikamooz.com/what-is-a-vector-database)

پایگاه داده برداری به دلیل ظهور اپلیکیشن‌های مبتنی بر **یادگیری ماشین** (Machine Learning) و **هوش مصنوعی** (Artificial Intelligence)، توجه زیادی را به خود جلب کرده است. در این مقاله قصد داریم به این پرسش پاسخ دهیم که Vector Database چیست و نحوه کارکرد آن چگونه است. علاوه بر این، عوامل مؤثر در اهمیت این نوع از دیتابیس در **مدل‌های LLM** و هوش مصنوعی را بررسی خواهیم کرد.

داده‌های برداری چیست؟

در حوزه پایگاه داده برداری و یادگیری ماشین، داده برداری (Vector Data) به آرایه‌های چندبعدی از اعداد گفته می‌شود که نشان‌دهنده ویژگی‌ها یا همان Feature های آیتم‌های داده هستند. این بردارها برای نوع‌های مختلف تحلیل، به‌ویژه در مواردی مهم هستند که جستجو مبتنی بر شباهت مدنظر است. به صورت کلی، داده‌های برداری یک نمایش عددی از دیتا به‌شمار می‌روند و در اغلب موارد، از **فرآیند استخراج ویژگی** (Feature Extraction) حاصل می‌شوند. Vector Data به صورت High-Dimension است؛ این یعنی، این نوع از داده‌ها حاوی ابعاد بالایی هستند.

پایگاه داده برداری (Vector Database) چیست؟

پایگاه داده برداری (Vector Database) نوع خاصی از دیتابیس محسوب می‌شود که برای ذخیره‌سازی مؤثر، مدیریت و انجام عملیات روی داده‌های برداری طراحی شده است. در بیشتر موارد، دیتای برداری در یادگیری ماشین و هوش مصنوعی مورد استفاده قرار می‌گیرند و به هدف کار با بردارهای با ابعاد بالا به کار می‌روند.



مزایای پایگاه داده برداری

موارد زیر، کلیدی‌ترین مزیت های Vector Database به حساب می‌آیند:

- **کارآمد در مدیریت داده‌های پیچیده:** پایگاه داده برداری به گونه‌ای طراحی شده است که امکان ذخیره‌سازی و پردازش داده‌ها به صورت مؤثر را دارا است. به همین دلیل، دیتابیس برداری گزینه ایده‌آلی برای مدیریت دیتای چندرسانه‌ای، متن‌ها و در مجموع، دیتاست‌های پیچیده محسوب می‌شود.
- **سرعت:** Vector Database، به‌ویژه در زمان جستجوی تشابهات در فضاهای چندبعدی، سرعت مناسبی را در کوئری‌های مربوط به دیتاست‌های بزرگ ارائه می‌دهند.
- **قابلیت ارتقا:** این پایگاه داده می‌تواند به صورت افقی (Horizontal) اسکال شود.
- **یکپارچگی با هوش مصنوعی و یادگیری ماشین:** Vector Database قابلیت Integrate شدن با مدل‌های هوش مصنوعی، به‌ویژه مدل‌های مربوط به [ایلیکیشن های سیستم های پیشنهاددهنده](#) (Recommendation Systems) و [تشخیص تصویر](#) (Image Recognition) را دارا است.

معایب Vector Database

کاستی‌های پایگاه داده برداری عبارتند از:

- **پیچیدگی:** در مقایسه با [پایگاه های داده رابطه‌ای \(RDBMS\)](#)، تکنولوژی پس‌زمینه پایگاه داده برداری از نظر درک و پیاده‌سازی می‌تواند پیچیده محسوب شود.
- **نیاز به منابع بالا:** برای آن که عملکرد پایگاه برداری به‌خوبی حفظ شود، بالابودن قدرت محاسباتی ضروری است.
- **استانداردسازی محدود:** به‌عنوان یک فناوری نسبتاً جدید، در Vector Database مشابه پایگاه‌های داده سنتی، استانداردسازی بالایی صورت نگرفته است؛ ممکن است این موضوع، به‌طور بالقوه به مشکلات سازگاری منجر شود.

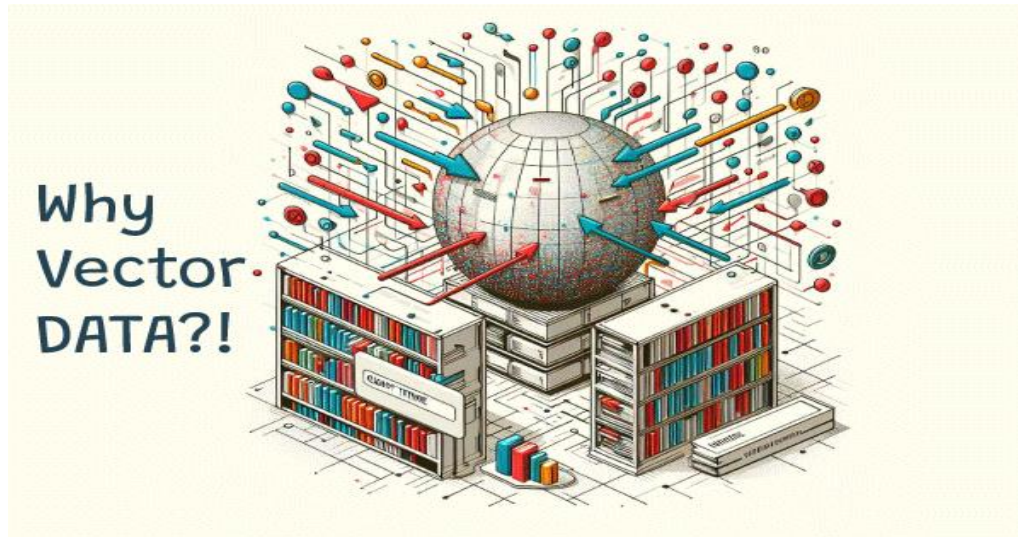
کاربردهای پایگاه داده برداری (Vector Database)

پایگاه داده برداری مانند یک ابزار قدرتمند برای مدیریت کارآمد داده‌های برداری چندبعدی استفاده می‌شود. مهم‌ترین کاربردهای Vector Database به شرح زیر است:

- **جستجوی شباهت:** همان‌طور که پیش‌تر به آن اشاره شد، می‌توان برای بازیابی آیتم‌هایی مشابه با یک کوئری، مانند بازیابی تصویر یا ویدیو، از آن بهره برد.
- **سیستم‌های پیشنهاددهنده:** Engine هایی را تقویت می‌کند که محصول، موسیقی یا فیلم را براساس بردارهای ترجیحات کاربران (User Preference) پیشنهاد می‌دهند.
- **پردازش زبان طبیعی (NLP):** **جستجوی معنایی** (Semantic Search) و بازیابی اطلاعات را در اپلیکیشن‌های مبتنی بر متن تسهیل می‌کند.
- **شناسایی ناهنجاری (Anomaly Detection):** الگوهای غیرمعمول یا **داده های پرت** (Outlier Data) را در مجموعه داده‌ها شناسایی می‌کند که در تشخیص کلاهبرداری و مانیتورینگ سلامت مفید است.
- **بیو انفورماتیک:** در مقایسه ساختارهای ژنتیکی، پروتئینی یا شیمیایی به منظور تحقیق و کشف دارو کمک می‌کند.
- **مدیریت مدل یادگیری ماشین:** بازیابی و مدیریت مدل‌های یادگیری ماشین را براساس بردارهای ویژگی برعهده دارد.
- **کاربردهای جغرافیایی (Geospatial):** داده‌های جغرافیایی را براساس بازنمایی بردار فضایی (Spatial Vector Representations) مدیریت و جستجو خواهد کرد.

چرا پایگاه داده برداری اهمیت دارد؟

پایگاه داده برداری به‌واسطه مزایایی همچون قابلیت‌های خاص آن در رسیدگی به داده‌های برداری با ابعاد بالا، جستجوی کارآمد شباهت‌ها، مقیاس‌پذیری، پردازش بلادرنگ (Real-Time Processing)، امکان یکپارچگی با هوش مصنوعی و ارائه نتایج دقیق‌تر در جستجو، توجه فراوانی را به خود جلب کرده است. Vector Database از پایه‌های محیط‌های دیتامحور و مدرن محسوب می‌شود و به شما این امکان را می‌دهد تا با انجام عملیات پیشرفته روی داده‌ها، بینش‌های ارزشمندی به‌دست آورید.



اجزای اصلی پایگاه داده برداری

به طور معمول، یک پایگاه داده برداری متشکل از چند کامپوننت کلیدی است که در این بخش به آن‌ها می‌پردازیم:

- **ایندکس برداری (Vector Index):** این کامپوننت امکان ایندکس‌گذاری و بازیابی کارآمد بردارهای داده‌ای چندبعدی را فراهم می‌کند.
- **موتور ذخیره‌سازی (Storage Engine):** مدیریت ذخیره‌سازی و بازیابی داده‌های برداری برعهده این بخش است.
- **موتور پرس‌وجو (Query Engine):** پردازش کوئری‌ها را مدیریت می‌کند و به کاربران قابلیت جستجو بر مبنای شباهت و سایر کوئری پیچیده را می‌دهد.
- **لایه مدیریت (Management Layer):** ابزارهای خاصی را برای مدیریت، نظارت و پیکربندی پایگاه داده به کاربر ارائه می‌دهد.

پایگاه داده برداری چگونه کار می‌کند؟

برای راه‌اندازی و استفاده مؤثر از پایگاه داده برداری، اقدامات زیر را دنبال کنید:

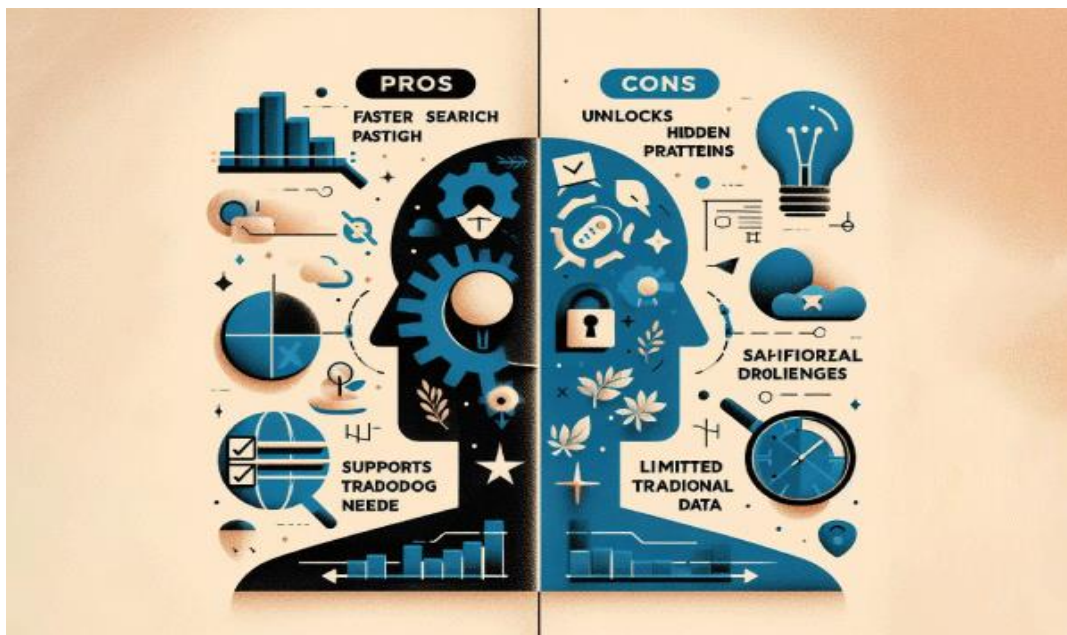
۱. آماده‌سازی داده‌ها و تبدیل آن‌ها به بردارهای عددی متناسب
۲. انتخاب یک Vector Database مناسب براساس نیازمندی‌ها، کارایی و نوع کوئری‌های قابل پشتیبانی
۳. نصب نرم‌افزار دیتابیس و راه‌اندازی آن براساس الزامات محیط عملیاتی و کارایی
۴. انتخاب و پیاده‌سازی استراتژی ایندکس‌گذاری (از جمله KD-Trees و HNSW)
۵. طراحی و توسعه مکانیزم‌های کوئری و تعیین متریک‌های مرتبط با جستجو
۶. یکپارچه‌سازی با جریان‌های کاری و سیستم‌های داده‌ای موجود
۷. بهینه‌سازی و تست
۸. استقرار دیتابیس برداری به Production و برنامه‌ریزی برای Scale کردن
۹. به‌روزرسانی مداوم و نگهداری دیتابیس

آینده پایگاه داده برداری

با افزایش تولید داده‌های غیرساختاریافته به صورت گسترده، کسب‌وکارها به راه‌حل‌های پیشرفته‌ای برای مدیریت داده، مانند Vector Database، نیاز خواهند داشت. ممکن است پیشرفت‌های آتی این نوع از دیتابیس، بر بهبود قابلیت تعامل با پایگاه‌های داده سنتی، افزایش سهولت استفاده و بهینه‌سازی بیشتر عملکرد متمرکز باشد. از سوی دیگر، اگر ادغام این پایگاه داده با سرویس‌های مبتنی بر Cloud فراهم شود، دسترسی به این فناوری قدرتمند را همگانی‌تر می‌کند. در مجموع، به نظر می‌رسد آینده پایگاه داده برداری نویدبخش اتفاق‌ها و پیشرفت‌های متعددی خواهد بود.

تفاوت Vector Database با سایر پایگاه‌های داده

برخلاف پایگاه‌های داده متداول که داده‌ها را در قالب سطرها و ستون‌ها ذخیره‌سازی کرده و از زبان SQL برای کوئری‌نویسی استفاده می‌کنند، پایگاه داده برداری از بردارها بهره می‌برد و به منظور جستجو شباهت‌ها بهینه‌سازی شده است. به صورت کلی، پایگاه داده سنتی برای عملیات تراکنشی و کوئری‌های داده‌های ساختاریافته مناسب است، اما در مدیریت داده‌های پیچیده و غیرساختاریافته با مشکل مواجه می‌شود. البته توجه کنید که هر یک از [انواع پایگاه داده](#) کاربردها و مزیت‌های خاص خود را دارا است.



ارتباط پایگاه داده برداری و هوش مصنوعی

پایگاه داده برداری یک تکنولوژی پراهمیت در توسعه و استقرار سریع اپلیکیشن‌های هوش مصنوعی محسوب می‌شود. Vector Database با رسیدگی به چالش‌های محوری مربوط به مدیریت داده‌ها، کارایی مطلوب‌تری را برای سیستم‌های هوش مصنوعی امکان‌پذیر می‌کند. با استناد به رشد مداوم، تقاضا برای سیستم‌های مبتنی بر AI، این احتمال وجود دارد که نقش پایگاه‌های داده برداری در این حوزه‌ها، بیش از پیش پررنگ‌تر شود.

چرا مدل های LLM به پایگاه داده برداری نیاز دارند؟

Large Language Model هایی مشهور مانند [ChatGPT](#) و سایر [ابزارهای هوش مصنوعی](#)، به شدت به پایگاه داده برداری برای مدیریت و انجام کارآمد عملیات با بردارهای چندبعدی وابستگی دارند. این بردارها یکی از اجزای اساسی در پردازش، درک و تولید متن مشابه انسان به شمار می‌روند. درحقیقت، Vector Database نه تنها یک ابزار برای بهبود کارایی در LLM ها است؛ بلکه این اطمینان را ایجاد می‌کند که Task های مدل‌ها به صورت مؤثر اجرا شوند.

مثال پایگاه داده برداری

Pinecone یکی از مثال‌های پایگاه داده برداری در دست استفاده است. Pinecone به صورت خاص برای مدیریت پیچیدگی‌های مربوط به داده‌های برداری و پشتیبانی از عملیات جستجوی شباهت طراحی شده است. Pinecone یک سرویس مدیریت شده است؛ این یعنی Pinecone بخش زیادی از پیچیدگی‌های مربوط به استقرار و مدیریت زیرساخت Vector Database را تسهیل می‌بخشد.

مثال کاربردی از Pinecone استفاده از آن در سیستم پیشنهاددهنده محتوا برای سرویس پخش زنده است. در این سناریو، هر آیتم (مانند فیلم‌ها، سریال‌های تلویزیونی یا ویدئوها، و ترجیحات کاربر) به صورت بردار یا همان Vector نمایش داده می‌شوند. هنگامی که یک کاربر با سرویس تعامل می‌کند، سیستم Pinecone را به منظور یافتن و پیشنهاددهی محتوای مرتبط با ترجیحات کاربر استفاده می‌کند. این موضوع براساس شباهت در فضای برداری انجام می‌شود. در چنین شرایطی، محتوای پیشنهادی شخصی‌سازی شده در زمان واقعی ارائه شده و به دنبال آن، تجربه کاربری بهبود داده می‌شود.

کلام آخر: Vector Database چیست؟

پایگاه داده برداری نوع خاصی از دیتابیس است که برای مدیریت دیتای برداری طراحی شده است. این نوع داده، اغلب در اپلیکیشن‌هایی استفاده می‌شود که جستجوی سریع براساس شباهت و پردازش مؤثر داده‌های پیچیده و چندبعدی موردنیاز است. در این مقاله، به بررسی Vector Database پرداختیم و کاربردها، دلایل اهمیت و ارتباط آن با هوش مصنوعی را بررسی کردیم.