

» بے نام خدا «

درس : مباحث ویرہ

استاد : محمد احمد زارہ

رشتہ : مهندس / کامپیوٹر

دانشجو : زہرا کبریٰ امل - زہرا صالحی

دانشگاہ : ملی صہارت اکوڑ شدہ میٹاب

Subject:

Day: Month: Year:

در نقشار ذخیره می شود. این امر باعث می شود که عملیات سیستم و تغییر داده ها بسیار مقیاس پذیر باشد.

کاهش هزینه ها: با حذف واسطه ها و تسهیل فرآیندها DLT می تواند به کاهش هزینه ها، افزایش شفافیت و زمان مورد نیاز برای مدیریت داده ها کمک کند.

قابلیت دسترسی: مقیاس پذیر DLT افراد و نهادها این امکان

را می دهد که به راحتی و بدون نیاز به پیچیدگی خاص، داده ها دسترسی

پیدا کنند و به تبادل اطلاعات بپردازند. همچنین این نهادها قابلیت مقیاس

پذیری را کاپی داد و می توانند با افزایش تعداد کاربران، داده ها به راحتی

توسعه داده شود. DLT به دلیل ساختار غیر متمرکز خود و امنیت

یک نهاد متمرکز را کاهش می دهد و این امر به کاهش ریسک های مربوط

Subject

Day Month Year

1 بخش 6 Advanced Topics : 1404/01/20

A: چرا Distributed Ledger Technology (DLT) در

مدیریت داده ها کاربرد دارد؟

چرا Distributed Ledger Technology (DLT) ...

فناوری دفتر کل توزیع شده (DLT) به دلیل ویژگی های خاص و مزایا

متمدد که دارد در مدیریت داده ها کاربرد فراوان پیدا کرده است. در

زیر بر فراز این دلایل اشاره می کنم.

شفافیت و قابل اعتماد بودن: DLT به تعارض شرکت کنندگان در شبکه

این امکان را می دهد که به یک سیستم مشترک از داده ها دسترسی دائم

یابند. این ویژگی باعث افزایش شفافیت و اعتبار اطلاعات می شود

و می تواند به کاهش تقلب و سوءاستفاده کمک کند.

امنیت بالا: اطلاعات در DLT به صورت غیر متمرکز و با استفاده از

به همراه یا سود و بر صحت یک قرارداد امن گفت می کنند.

سازگار با قرارداد های هوشمند، DLT امکان استفاده از قرارداد های

هوشمند (Smart contracts) را فراهم می کند که می توانند به طور

خودکار فرآیند های تجاری و مدیریت را اجرا کنند و افزایش کارایی

و کاهش خطاهای گفت کنند. با توجه به این مزایا DLT می تواند در حوزه

های مختلفی از جمله بانکداری، بیمه و تامین بهداشت و درمان و دیگر صنایع

کاربرد داشته باشد.

Blockchain.B چرا برای ذخیره سازی داده های امن استفاده می شود؟

بلوک چین (Blockchain) به عنوان یک فناوری انتقال پردر ذخیره سازی

داده ها، دلیل و پیشینه های خاص خود و ویژه در زمینه امنیت مورد توجه

قرار گرفته است. در زیر به برخی از دلایل کلیدی که باعث می شود بلوک چین برای

ذخیره سازی داده های امن مناسب باشد، اشاره می کنیم.

Subject

Day Month Year

ما می‌توانیم منتظر از جمله را کنار تا این زمان بمانیم و در حالت وجود
مورد استفاده قرار می‌گیرد.

۵. چرا GANs (Generative Adversarial networks)

در علم داده بیشتر فقه کاربرد دارند؟

و ویژگی‌های خاص منحصربه‌فرد خود را تولید داده‌های جدید و واقع‌گرایانه

در علم داده بیشتر فقه کاربرد دارند زیرا در علم داده‌ها استفاده از داده‌های

GAN داده‌ها به برقراری دلیل اصلی است

تولید داده‌های جدید: GAN‌ها قادرند داده‌های جدید و متنوع را تولید کنند

و واقع‌گرایانه‌تر باشند این ویژگی برای کاربردهای زیادی مانند تولید تصاویر

و صداها و حتی متون بسیار مفید است

تنوعیت داده‌ها: در بسیاری از موارد داده‌های موجود محدود و یا نامتعادل

هستند GAN‌ها قادرند به تولید نمونه‌های متنوع‌تر و برابر تنوع موجود

رمزنگار (Encryption) داده‌ها را به بلوک‌بندی معمولاً با استفاده از کلید خاص رمزنگار محافظت می‌شود. این به معنای این است که تنها افراد که دسترسی مناسب دارند می‌توانند داده‌ها را دست‌رسی پیدا کنند.

موافقت‌ها (Smart Contracts) با استفاده از قرارداد داده‌ها را هوشمند می‌خواند. قوانین و شرایط خاص را به صورت خودکار اجرا می‌کنند و می‌توانند امنیت بیشتر را فراهم کنند.

اجماع (Consensus mechanisms) بلوک‌بندی‌ها از بزرگس‌ها (جماع اجماع) استفاده می‌کنند تا اطمینان حاصل کنند که تمام شبکه‌ها با یکدیگر توافق دارند و تغییرات در داده‌ها به طور مشترک تأیید می‌شوند. این امر مانع از جعل داده‌ها می‌شود. این ویژگی‌ها باعث می‌شود که بلوک‌بندی به عنوان یک فناوری مطمئن و امن برای ذخیره‌سازی داده‌ها شناخته شود و در

Subject _____

Day _____ Month _____ Year _____

غیر قابل تغییر بودن (Immutability) داده‌ها در بلاکچین به این
از ذخیره‌سازی تغییر ناپذیر است. هر بلاک جدید به بلاک‌های قبلی
مستقل می‌شود و هر تغییر در یک بلاک نیاز به تغییر تمام بلاک‌های
بعدی دارد که این کار بسیار دشوار است. این ویژگی به عنوان یک
داده‌ها گنجانده می‌شود.

شفافیت (Transparency) بلاکچین به هم شرکت کنندگان اجازه
در هر تابعی که داده‌ها در سرور داشته باشند. این شفافیت
موجب می‌شود که هر کس بتواند تغییر یا تراکنش قابل مشاهده و بررسی باشد.
توزیع شده (Decentralization) در بلاکچین داده‌ها به صورت
توزیع شده در گره‌ها (Nodes) ذخیره می‌شوند. این موضوع باعث
می‌شود که هیچ نقطه ضعفی وجود نداشته باشد و همه با هم
دشواری‌ها را برطرف می‌کنند.

داده ها کمک کنند که این امر در دیتا دیتا بگیرد ماشین و یادگیری
عمیق مهم است

بهتر کیفیت داده ها: GAN عامر توانا هستند برار تصحیح و بهبود کیفیت

داده ها را موجود داشته تقاضای کم کیفیت یا ناقص استفاده شوند این

کاربرد در زمینه های مانند پزشکی، هنر و طراحی بسیار مفید است.

تصحیح ساز و آفتاب: GAN عامر توانا هستند برار تهیه ساز و آفتاب

مختلف و بهر دستاورد آن ها بر مدل ها مختلف استفاده شده این

دیتا در تحقیق و توسعه کاربرد بسیار متنوع و دیتا بسیار مهم است.

یادگیری و دیتا در میان: GAN های دیتا رقابت بین در شبیه

(عبارت دور و دیتا بیضا تور) مر خواهند و دیتا در میان و دیتا در میان

به چهره در داده ها را با دیتا در میان و دیتا در میان به دیتا در میان به دیتا در میان

دیتا در میان و دیتا در میان و دیتا در میان و دیتا در میان و دیتا در میان

Subject:

Day: Month: Year:

اگرچه در تمام حلقه‌های GAN ها در زمینه‌های مختلف و کاربردهای گوناگون

دارند مانند تولید آثار هنری و موسیقی و طراحی مدل‌های کاربردی و غیره

در طراحی این ابزارها در زمینه‌های مختلف آثار خلق شده و در حال حاضر

تحلیل و تولید داده‌ها و توزیع داده‌ها: GAN ها می‌توانند توزیع داده‌ها را شبیه

سازگار کنند و به ما کمک کنند تا الگوهای مختلف را در داده‌ها شناسایی کنیم

و تحلیل‌های دقیق‌تری انجام دهیم.

در مجموع GAN ها به دلیل قابلیت‌های بی‌شماره‌ای که دارند به ابزار

قدرتمندی در علم داده و یادگیری ماشین تبدیل شده‌اند و به محققان و توسعه‌دهندگان

در زمینه‌های مختلف امکان‌پذیر می‌دهد که ابزارهای نوآورانه‌ای برای مسائل مختلف

ارائه دهند.

D: t-SNE و PCA به تفاوت دارند؟

هر دو تکنیک‌های کاهش ابعاد هستند که برای تجزیه و تحلیل داده‌ها

Subject:

Day: Month: Year:

استفاده می شود و آموزش ها و اعداد آن ها متفاوت است و در ادامه

تفاوت های کلیتر بین این دو روش را بررسی می کنیم.

1. روش اولی:

PCA

هدف اصلی PCA کاهش ابعاد داده ها با حفظ واریانس (توزیع) داده ها است

این روش به دنبال پیدا کردن تعداد کم از ویژگی ها (کمترین تفاوت ها) است

که بیشترین واریانس را در داده ها توزیع می دهند و معمولاً برابر بیشترین

داده ها و کاهش ابعاد قبل از استفاده از روش های دیگر مانند

استفاده می شود.

$t-SNE$

بیشتر برای تجسم داده ها و ابعاد پایین (معمولاً $t-SNE$ با 2 بعد) (بهر)

استفاده می شود این روش به دنبال حفظ ساختار محلی داده ها (بین فاصله ها

و ارتباط بین نقاط نزدیک به هم) است و معمولاً برابر کشف خوشه ها و خوشه

منور داده ها مناسب $t-SNE$ است.

Subject:

Day: Month: Year:

PCA

2. روش های خاص برای:

از ترکیب خطی داده ها استفاده می کنند و به صورت PCA و یا خطی

خاصیت مفاد بیرونی و هر دو را در هر دو ویژگی تاثیر می گذارند یا کوواریانس

داده ها کار می کنند این روش به عنوان یک سری همگرا و اگر مقدار $t-SNE$

بسیار به ویژگی برابر داده های نزدیک

$t-SNE$

از یک رویکرد مفاد فرار برای نگاشت نقاط در فضای ابعاد $t-SNE$ بالا

فضای ابعاد پایین استفاده می کنند این روش از توزیع های احتمالی

برای مدل ساز و مقایسه تفاوت های بین ابعاد استفاده می کنند

معقول زمان برتر و نیاز به تنظیم پارامترهای فاصله $t-SNE$

PCAplexity دارد

3. منابع

PCA:

نتایج PCA بر توانایی و امتزاج بیشتر متغیرها و یا کمبودها در ترکیب های

Subject

Day Month Year

E: چرا UMAP برابر Dimensionality Reduction استفاده می شود؟

چون تکنیک خودرشد برای کاهش ابعاد است که دایره و بیضی را به هم می پیوندد

تحلیل داده ها بسیار مورد استفاده قرار می گیرد UMAP قادر است دایره

را به دایره تبدیل کند و ابعاد را حفظ کند

محاسبات از صاف شدن محاسبه UMAP به عنوان یک امر ساده است که محاسبه

محاسبه داده ها را حفظ می کند این بدان معناست که نقاط نزدیک به هم

بسیار از یکدیگر جدا می شوند و یکدیگر را نزدیک می کنند

مدل ساز صاف کننده ها: مدل ساز صاف کننده UMAP قادر است تصاویر

های مختلف را به یک مدل ساز تبدیل کند این در واقع به تحلیل این تصاویر

است و هر یک از آن تصاویر را به یک مدل ساز تبدیل می کند

سرعت و کارایی: UMAP به طور معمول سریع تر از تکنیک های دیگر است

t-SNE مدل می کشد و نیز به هم می پیوندد که با مجموعه های داده بزرگ

Subject:

Day: Month: Year:

میشود. این امر به دلیل استفاده از الگوریتم های کارآمدتر در UMAP است.

انطباق پذیری UMAP: می تواند با انواع مختلف داده ها و توزیع های نامنظم و اختلال در داده ها و همچنین توان آن را برای نیازهای خاص تنظیم کرد. حجم نمونه: نتیجه های UMAP به صورت گسترده قابل تجزیه و تحلیل و تفسیر هستند که این امر به پژوهشگران و تحلیلگران داده کمک می کند تا الگوها و ارتباطات موجود در داده ها را به راحتی مشاهده و تفسیر کنند.

تطبیق پذیری: UMAP می تواند به عنوان یک ابزار بین داده های مختلف از عمل های یادگیری ماشین یا به عنوان یک تکنیک تجزیه و تحلیل مستقل استفاده شود. به طور کلی UMAP به دلیل توانایی هایش در حفظ ساختار داده ها و کارایی های بالا به یکی از محبوب ترین روش های یادگیری ماشین تبدیل شده است.

Subject

Date Month Year

NetworkX: F
چرا برای تحلیل گراف ها کاربرد دارد؟

یکی کتابخانه معروف در زبان برنامه نویسی پایتون NetworkX

است که برای تحلیل و کار با گراف ها و شبکه ها طراحی شده است. این کتابخانه

به دلیل ویژگی ها و قابلیت ها، مقبول گشته و داده ها را به سادگی و به روش های

متعدد عمل و محاسبه می دهد. قابلیت های زیادی دارد و به سادگی می تواند برای

کارهای مختلف گراف ها کاربرد دارد. در این کتابخانه NetworkX می توانیم

ساختار و کار با گراف ها را به سادگی و به روش های مختلف انجام دهیم و

نتایج آن را به سادگی و به روش های مختلف نمایش دهیم. این کتابخانه

به سادگی و به روش های مختلف می تواند برای کارهای مختلف استفاده شود.

در این کتابخانه می توانیم به سادگی و به روش های مختلف گراف ها را

ساخت و به سادگی و به روش های مختلف آن را نمایش دهیم. این کتابخانه

به سادگی و به روش های مختلف می تواند برای کارهای مختلف استفاده شود.

در این کتابخانه می توانیم به سادگی و به روش های مختلف گراف ها را

ساخت و به سادگی و به روش های مختلف آن را نمایش دهیم. این کتابخانه

به سادگی و به روش های مختلف می تواند برای کارهای مختلف استفاده شود.

Subject:

Day Month Year

نوع در نوع کتاب ها: NetworkX از نوع مختلف گراف ها را جست و آرد
بدون جهت وزن دارد و ... بیشتر از مرکز گراف این امر امکان تحلیل
انواع مختلف شبکه ها را فراهم می آورد

قابلیت تجزیه و تحلیل بیشتر شده: NetworkX ابزار و هاست متنوعی
برای تحلیل ساختار و ویژگی های گراف ها مانند معادله مرکزیت
خوشه بندی و کوتاه ترین مسیرها و غیره را دارد.

قابلیت ادغام با دیگر کتابخانه ها: NetworkX به راحتی با کتابخانه
های پایتون مانند matplotlib برای ترسیم گراف ها و Pandas برای
مدیریت داده ها ادغام می شود.

مستند سازی و جامعه فعال: این کتابخانه دارای مستندات جامع و جامعه کاربری
فعال است که به کاربران کمک می کند تا با هر مشکلی که در حال
کنند و از منابع موجود بهره ببرند.



Subject

Day Month Year

استیمار از الگوریتم های پیچیده: `Networkx` شامل پیاده سازی

استیمار از الگوریتم های معروف و فیلترهای مختلف است که به عنوان این

امکان دهنده و فیلترهای مختلف است که به عنوان این

`NLTK: G` چرا برای پردازش زبان طبیعی (NLP) استفاده می شود

`NLTK (Natural Language Toolkit)` مجموعه

فیلترهای پردازش زبان طبیعی است که برای پردازش زبان طبیعی

`NLTK` پردازش شده است و دلیل استفاده از (NLP) پردازش زبان طبیعی

مجموعه زیر است.

ابزارهای متنوع: `NLTK` شامل مجموعه ای از ابزارهای متنوع

متن پردازش متن، تجزیه و تحلیل، فیلترهای متن و غیره

ها، تحلیل آماری و غیره است.

داده های آموزشی: این مجموعه شامل مجموعه های داده و متن

Subject:

Date: Month: Year:

حیاض مثال: NLTK دارای کتابخانه بزرگی و فعال است که کارهای
چهر توانسته تغییراتی که پیش می‌رفتند و در حل مشکلات خود به
چرا دارند.

به علاوه NLTK که ابزار قدرتمند و مفید برای پردازش زبان
طبیعی است که به عنوان یک سیستم و متن‌ها و واژه‌ها و کلمات و کلمات
در زمینه‌های مختلف NLP کار کنند.

H: برای scrapy و web scraping کاربرد دارد.

یک فریم‌ورک محبوب و قدرتمند برای وب اسکراپینگ scrapy در
زبان برنامه‌نویسی پایتون است و به (web scraping) دلائل
توسعه کاربرد دارد.

سازگار با سیستم‌های scrapy به علاوه خاص برای جمع‌آوری

داده‌ها از وب سایت‌ها طراحی شده است و قابلیت مدیریت و درخواست‌ها

Subject

Day

Month

Year

مدیریت فایل ها و شناسایی اکثر عناصر HTML را دارد.

مدیریت درخواست ها: این فریم ورک به صورت خودکار درخواست ها را

مدیریت می کند و می تواند به طور هم زمان چندین درخواست را ارسال

کند. این امر باعث افزایش سرعت می شود و در نتیجه کار می شود.

بمقایسه از XPath و CSS selectors: scrapy و XPath

و CSS selectors به مقایسه می کنند که به کار بردن این امکان نامرئی

باشد و اکثر عناصر مورد نظر خود را در صفحات وب پیدا کنند.

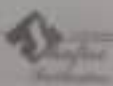
پلاگین ها و دستورات: scrapy می تواند با دستورات دیگر

شده را به فرمت های مختلف مانند CSV، JSON و XML ذخیره

کند و این امکان را برای کار بردن فراهم می آورد و به راحتی داده ها را

گرفته.

بمقایسه از scrapy: crawling قابلیت ها را پیشرفته تر



Subject

Day Month Year

پیدا کردن (crawling) در وب سایت ها داده ها و به علاوه
که می تواند به طور خودکار یک بار ارسال کند و داده ها را از صفحات
مستقیم جمع آوری کند.

مستندات و جامع حال: scrapy داده از صفحات جامع و یکجا جمع
م فعال از توهم دهستان است که به کار بردن کمک می کند تا مشکلات
خود را حل کند و از تجربیات دیگران بهره مند شود.

1: Beautifulsoup برای کار استفاده می شود؟

یک کتابخانه پایتون که برای تجزیه Beautifulsoup استفاده می شود

از HTML و XML و تحلیل و پردازش کتابخانه به ویژه در زمینه

وب استرینگ ها به کار می رود و در زیر به فرایند کار هر امر

اشاره می کنیم Beautifulsoup

تجزیه و تحلیل HTML, Beautifulsoup, XML می تواند به

Subject:

Day: Month: Year:

راحت صفحات صفحات و با را تجزیه و تحلیل کند و به شما اجازه دهد
مردود تا به نام مختلف مانند کتابخانه کلاسیک و کتابخانه ها (درست
پیدا کنید.

امتیاز داده های با استفاده از BeautifulSoup می تواند داده های
فایل را صفحات و با را امتیاز دهد مانند متن، آیکان ها و تصاویر و غیره.
پیدا کردن درخت DOM: این کتابخانه به شما امکان می دهد تا به سادگی
در ساختار و درخت HTML یا XML پیدا کنید و عناصر مورد نظر خود را
پیدا کنید.

مدیریت صفحه BeautifulSoup به خوبی با HTML ها و ناظر
و نام فایل نام می گذارد و می تواند صفحه های را به دست آورد.

ادغام با دیگر کتابخانه ها BeautifulSoup با کتابخانه های
دیگر مانند Requests برای ارسال درخواست های HTTP به
نام و به سادگی به نام سایت ها و دریافت کنند.