禪

تصورات اشتباه در ارتباط با مدلهای دادهای

با 10 باور نادرست در ارتباط با بزرگ دادهها آشنا شوید

برای آنکه اطمینان حاصل کنید برنامههای سازمان شما در ارتباط با بزرگ دادهها در مسیر درستی قرار گرفتهاند یا نه بهتر است با 10 باور اشتباه و رایجی که در ارتباط با این فناوری وجود دارند آشنا شوید. به طور مثال، یکی از عجیبترین جملاتی که در ارتباط با بزرگ دادهها اغلب میشنویم این است که اگر مقدار کمی از دادهها خوب هستند، پس حجم بسیار بالایی از دادهها ایدهآل خواهد بود . به نظر شما این حرف تا چه اندازه میتواند درست باشد؟ برای آنکه به این پرسش پاسخ روشنی دهیم، اجازه دهید از یک مثال ساده استفاده کنیم . گفتن این جمله درست مصداق این است که بگوییم در یک تابستان گرم یک نسیم خنک حس خوبی را به وجود میآورد، اما آیا یک گردباد نیز همان حس را به شما منتقل خواهد کرد؟

این مطلب مقدمه مجموعه مقالات پرونده ویژه «دادههای بزرگ؛ فردای بزرگتر» شماره **197** ماهنامه شبکه است. علاقهمندان میتوانند کل این پرونده ویژه را از روی سایت شبکه دانلود کنند .

مطلب پیشنهادی


داده‌های بزرگ: فردای بزرگ‌تر
BIG DATA

□□□□□□□□□□□□□□□ □□ □□□□□□□ □□□□□□□□□ □□ □□□□□ □□□□ □□□□□ □□□□ .□□□□ □□□□□ □□ □□□ □□□ □□□□
□□□□□□□□□□ □□□□□□ □□□□□□□ □□□□□□□□□ □□□□.» (□□□ 1)

data
iku
TOP PREDICTION ALGORITHMS
TYPE
NAME
DESCRIPTION
ADVANTAGES
DISADVANTAGES
Linear
Linear regression
The "best fit" line through all data points. Predictions are numerical.
Easy to understand -- you clearly see what the biggest drivers of the model are.
X Sometimes too simple to capture complex relationships between variables.
X Tendency for the model to "overfit".
Logistic regression
The adaptation of linear regression to problems of classification (e.g., yes/no questions, groups, etc.)
Also easy to understand.
X Sometimes too simple to capture complex relationships between variables.
X Tendency for the model to "overfit".
Tree-based
Decision tree
A graph that uses a branching method to match all possible outcomes of a decision.
Easy to understand and implement.
X Not often used on its own for prediction because it's also often too simple and not powerful enough for complex data.
Random Forest
Takes the average of many decision trees, each of which is made with a sample of the data. Each tree is weaker than a full decision tree, but by combining them we get better overall performance.
A sort of "wisdom of the crowd". Tends to result in very high quality models. Fast to train.
X Can be slow to output predictions relative to other algorithms.
X Not easy to understand predictions.
Gradient Boosting
Uses even weaker decision trees, that are increasingly focused on "hard" examples.
High-performing.
X A small change in the feature set or training set can create radical changes in the model.
X Not easy to understand predictions.
Neural networks
Neural networks
Mimics the behavior of the brain. Neural networks are interconnected neurons that pass messages to each other. Deep learning uses several layers of neural networks put one after the other.
Can handle extremely complex tasks - no other algorithm comes close in image recognition.
X Very, very slow to train, because they have so many layers. Require a lot of power.
X Almost impossible to understand predictions.
©2017 Dataiku, Inc. | www.dataiku.com | contact@dataiku.com | @dataiku

-1 □□□

□□□□□□□□□□□ □□ □ □□□□□ □□□□ □□□□ □□□□□□ □□□□□□□□□□□ □□ □□□□□□□□□□□□ □□□□□□□□□□ □□□□□□
.□□□□□ □□□□ □□□□ □□□□□□□ □□ □□□□□□□ □ □□□□□ □□□□□□□□

-4

-5

6- شما نمیتوانید برنامههای مربوط به بزرگ دادهها را روی زیرساختهای مجازی اجرا کنید

زمانی که بزرگ دادهها بهشکل عمومی مورد توجه مردم قرار گرفت، نزدیک به ده سال پیش بود که تقریباً مترادف با ظهور آپاچی هادوپ بود. جاستین موری از شرکت VMware در 12 می 2017 میلادی در مقالهای تحت عنوان Inside Big Data به این موضوع اشاره کرد و اعلام داشت که این واژه اکنون با فناوریهای رایج از (MongoDB، Apache Cassandra) NoSQL گرفته تا آپاچی اسپارک احاطه شده است. منتقدان در گذشته درباره عملکرد هادوپ روی ماشینهای مجازی پرسشهایی را مطرح کرده بودند، اما موری به این موضوع اشاره میکند که عملکرد هادوپ روی ماشینهای مجازی با عملکرد هادوپ روی ماشینهای فیزیکی قابل مقایسه است. موری همچنین به این نکته اشاره کرده است که بسیاری بر این باورند که ویژگیهای اساسی VM به SAN (سرنام Storage Area Network) نیاز دارند، اما این موضوع به هیچ عنوان صحت ندارد. با این حال، فروشندگان اغلب توصیه میکنند از ذخیرهسازهای DAS استفاده شود، بهواسطه آنکه عملکرد بهتری دارند و هزینههای پایینتری را تحمیل میکنند.

مطلب پیشنهادی

INTERNET
OF THINGS

رویای دیروز، واقعیت امروز

اینترنت اشیا با کلان دادهها عجین شده است

7- یادگیری ماشینی مترادف با هوش مصنوعی است

شکاف میان الگوریتمی که قادر است الگوها را در حجم عظیمی از دادهها شناسایی کند و الگوریتمی که قادر است یک نتیجه منطقی را بر اساس الگوهای دادهای نشان دهد، بسیار زیاد است. وینیت جی از ITProPortal در مقالهای که در ارتباط با یادگیری ماشینی منتشر کرده بود، به این موضوع اشاره کرد که یادگیری ماشینی از آمارهای تفسیری برای تولید مدلهای پیشبینیکننده استفاده میکند. این فناوری در پسزمینه الگوریتمهایی قرار دارد که پیشبینی میکنند یک مصرفکننده بر مبنای تاریخچه خریدهای خود در گذشته احتمال دارد چه محصولی را خریداری کند یا به چه آهنگهایی بر مبنای تاریخچه گذشته خود گوش فرا دهد. آن چنان که الگوریتمها بهسمت هوشمندی پیش میروند، ممکن است از دستیابی به اهداف هوش مصنوعی که بهمنظور الگوبرداری از تصمیمات انسانی بود دور شوند. پیشبینیهای مبتنی بر آمار فاقد استدلال، قضاوت و تخیل انسانی هستند. حتی پیشرفتهای حاصل شده در سامانههای هوش مصنوعی امروزی همچون واتسون آیبیام نیز قادر نیستند در بسیاری از موارد بینشی که دانشمندان علم داده مطرح میکنند را ارائه دهند.

مطلب پیشنهادی

بزرگ دادهها بازیگر اصلی فناوریهای فردا

[illegible] -8

[illegible] NewVantage Partners [illegible] (2 [illegible]) [illegible] 95 [illegible] 48.4 [illegible] 5 [illegible] 15 [illegible]

-2 [illegible] **NewVantage Partners** [illegible]

196 [illegible] **130** [illegible] **163** [illegible] **116** [illegible] **138** [illegible] **118** [illegible]

[illegible]



[redacted]

[redacted]

9- [redacted]

[redacted]
Robert Half Technology Salary Guide [redacted]

10- [redacted]

[redacted] NewVantage Partners [redacted]

[redacted]:
[redacted]
[redacted]:
15:10 - 05/10/1396
[redacted]:
[redacted] - big data - [redacted] - [redacted]

[redacted]:
https://www.shabakeh-mag.com/cover-story/10810/%D8%A8%D8%A7-10-%D8%A8%D8%A7%D9%88%D8%B1-%D9%86%D8%A7%D8%AF%D8%B1%D8%B3%D8%AA-%D8%AF%D8%B1-%D8%A7%D8%B1%D8%AA%D8%A8%D8%A7%D8%B7-%D8%A8%D8%A7-%D8%A8%D8%B2%D8%B1%DA%AF-%D8%AF%D8%A7%D8%AF%D9%87%E2%80%8C%D9%87%D8%A7-%D8%A2%D8%B4%D9%86%D8%A7-%D8%B4%D9%88%DB%8C%D8%AF