

بزرگ داده‌ها؛ فرصت‌های طلایی یا زمینه‌سازهای تبعیض بزرگ داده‌ها می‌توانند به سلاح‌های مخرب و خطرناک تبدیل شوند



برای آن دسته از افرادی که از طریق حل مشکلات گوناگون زندگی خود را می‌گذرانند، ظهور بزرگ داده‌ها همانند یک فرصت طلایی است. دانشمندان علم داده‌ها و برنامه‌نویسان می‌توانند با واکنشی داده‌های انسانی، آن‌ها را در مواردی به کار بگیرند که تا یک دهه پیش در حد فکر و خیال بود. اما در بین این همه شور و هیجان به تدریج شاهد جنبه‌های نگران‌کننده هستیم. مواردی وجود دارد که نشان می‌دهد الگوریتم‌های منظم و زیبا و مدل‌های پیش‌بینی ما کمبودها و نقایص مهمی دارد. اما چه کسی زمینه‌ساز بروز این کمبودها شده است؟ کسی که این فناوری را به وجود آورده است؛ عامل انسانی.

این مطلب یکی از 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



دانلود کنید: پرونده ویژه داده‌های بزرگ؛ فردای بزرگ‌تر

برای آن دسته از افرادی که از طریق حل مشکلات گوناگون زندگی خود را می‌گذرانند، ظهور بزرگ داده‌ها همانند یک فرصت طلایی است. دانشمندان علم داده‌ها و برنامه‌نویسان می‌توانند با واکنشی داده‌های انسانی، آن‌ها را در مواردی به کار بگیرند که تا یک دهه پیش در حد فکر و خیال بود. اما در بین این همه شور و هیجان به تدریج شاهد جنبه‌های نگران‌کننده هستیم. مواردی وجود دارد که نشان می‌دهد الگوریتم‌های منظم و زیبا و مدل‌های پیش‌بینی ما کمبودها و نقایص مهمی دارد. اما چه کسی زمینه‌ساز بروز این کمبودها شده است؟ کسی که این فناوری را به وجود آورده است؛ عامل انسانی. به عنوان مثال، به خبری که باعث رسوایی گوگل شد توجه کنید. پژوهش‌های به عمل آمده نشان داد که این شرکت ترجیح می‌دهد تبلیغات مرتبط با مشاغل پردرآمد را به جای آنکه برای زنان نشان دهد بیشتر برای مردان به نمایش بگذارد. واقعیت این است که ما اغلب جنبه‌های تجاری و اقتصادی بزرگ داده‌ها و الگوریتم‌ها را مورد بررسی قرار می‌دهیم، حال آنکه بزرگ داده‌ها و الگوریتم‌ها این قابلیت را دارند تا زمینه‌ساز تبعیضات مختلفی شوند.

کتی اونیل ریاضی‌دان و دانشمند علم داده در رابطه با این مدل‌های تبعیضی که روز به روز در حال گسترش هستند از یک واژه پرمعنا استفاده می‌کند: سلاح‌های کشتار جمعی. وی در کتاب جدید خود که به همین نام منتشر شده است، روش‌های مورد استفاده در الگوریتم‌ها را که عموماً باعث وخامت بی‌عدالتی و نابرابری می‌شوند به دقت شرح داده است. چندی پیش وی در یکی از جلسات زنده فیسبوک شرکت کرد و به سؤالات کاربران و شرکت‌کنندگان پاسخ داد. وی در این مصاحبه درخصوص اینکه برنامه‌نویسان و [دانشمندان داده](#) چه کارهایی می‌توانند انجام دهند تا مدل‌های آن‌ها بیشتر از صدمه زدن، باعث بهتر شدن اوضاع شود صحبت کرد. در این مقاله به اختصار چند بخش مهم از صحبت‌های خانم اونیل را برای شما آماده کرده‌ایم.

مطلب پیشنهادی



چالش بزرگ بعدی اکوسیستم داده‌محور چیست؟
انتقال سریع و هوشمندانه داده‌ها، معدن طلای عصر جدید

1- نشانه‌های سلاح‌های ویرانگر مبتنی بر ریاضیات را بشناسید

این مسئله اهمیت زیادی دارد. اساس وجود سلاح‌های ویرانگر مبتنی بر محاسبات/ ریاضیات، استفاده از آن‌ها برای شناخت نقاط بحرانی در زندگی افراد است. در حال حاضر، از الگوریتم‌هایی استفاده می‌شود که رزومه‌های کاری را مرتب‌سازی می‌کنند، به طور خودکار شیفت کاری نیروهای خدمات‌رسان را زمان‌بندی می‌کنند، هزینه بیمه و نرخ سود وام بانکی را یا حتی اینکه یک مجرم چه مدت زمانی را به‌عنوان مدت محکومیت خود در زندان سپری خواهد کرد، تعیین می‌کنند. از آنجا که نتایج این الگوریتم‌ها می‌تواند اثرات بحرانی روی زندگی میلیون‌ها نفر داشته باشد، پتانسیل خراب‌کاری‌های گسترده در آن‌ها وجود دارد.

*** این سلاح‌ها کاملاً سری و توصیف‌ناپذیر هستند**



شکل 1-
کتی
اونیل
ریاضی‌دان
و
دانشمند
علم داده

کسانی که از WMD ضربه می‌خورند معمولاً نمی‌توانند منبع این ضربه را درک کنند یا حتی متوجه نمی‌شوند از چه جایی ضربه خورده‌اند. روشی که پشت این مسئله قرار دارد «راز تجارت» است. با اینکه برخی از شرکت‌ها اعتقاد دارند این کار باعث جلوگیری از یادگیری قوانین و نحوه بازی سیستم می‌شود، اما نبود شفافیت باعث می‌شود تا راهی برای بررسی دقیق این موضوع نباشد. ولی سیستم یادگیری ماشین توانسته است این مشکل را یک قدم به

جلو ببرد، زیرا از یک طرف ابزار بسیار قدرتمندی برای یافتن روابط است و از طرف دیگر، در اکثر مواقع حتی برای سازندگان آن هم حکم جعبه سیاه را دارد.

* مخرب و ویران‌کننده هستند

به گفته اونیل، سلاح‌های مبتنی بر ریاضیات این توانایی را دارند که واقعیت مخصوص خود را بسازند و سپس، از همان واقعیت برای توجیه مدلشان استفاده کنند. به عنوان مثال، الگوریتمی که افراد ضعیف از نظر مالی را با وام‌های سنگین هدف قرار می‌دهد، با ساخت یک لوپ فیدبک، کار را برای خارج شدن این افراد از زیر بار بدهی سخت‌تر می‌کند. به طور مشابه، مدلی که برچسب اعتیاد بر کسی می‌زند که برای نخستین بار از مواد مخدر استفاده کرده است، احتمال ریسک را بالا می‌برد. زیرا این فرد در منطقه‌ای رشد کرده است که پتانسیل جرم و جنایت در آن بالا است. اگر این ریسک به محکومیت طولانی‌مدت زندان منجر شود، ارتباطات فرد خاطی با جامعه کمتر می‌شود و زمانی که از زندان آزاد شود، فرصت‌های شغلی کمی برای وی وجود دارد.

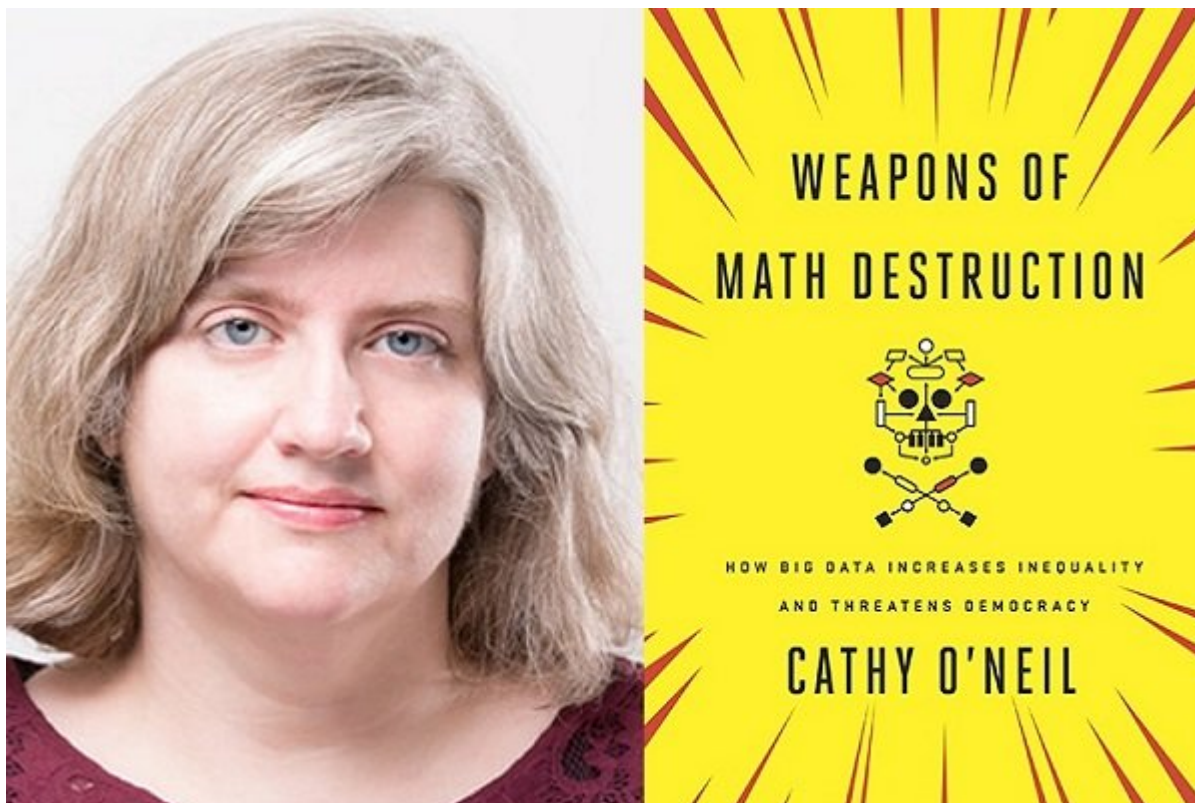
مطلب پیشنهادی



برای حرکت به سوی SDDC چه اقداماتی لازم است؟
پنج استراتژی برای موفقیت مرکز داده نرم‌افزار محور

2- به این مسئله توجه کنید که هیچ چیزی تحت عنوان «الگوریتم هدف» وجود ندارد

یکی از جنبه‌هایی که باعث جذابیت بالای [بزرگ داده‌ها](#) می‌شود این فرضیه است که این فناوری باعث از بین رفتن فردیت انسان‌ها و تبعیض در بین آن‌ها می‌شود. اما این طرز فکر غلط است. به گفته اونیل، مدل‌ها و الگوریتم‌های پیش‌بینی کننده در واقع نظریات و عقایدی هستند که در دل ریاضیات گنجانده شده‌اند. الگوریتم‌ها با یک دستور مشخص توسط انسان‌ها نوشته شده‌اند. تعریف اینکه کدام الگوریتم موفق است، شبیه قضاوت ارزش‌ها است و آنچه برای سازندگان الگوریتم موفقیت محسوب می‌شود، همیشه برای عموم جامعه خوب و مفید نیست. برای همین لازم است دانشمندان علم داده‌ها به تصویر بزرگ‌تری نگاه کنند. آن‌ها باید این پرسش را برای خود مطرح کنند که چه کسانی در الگوریتم من پیروز خواهند شد و از همه مهم‌تر اینکه چه اتفاقی برای بازندگان می‌افتد؟



شکل 2- کتاب
Weapon of Math Destruction
به خوبی نشان می‌دهد چگونه بزرگ داده‌ها می‌توانند زمینه‌ساز فشار مضاعف بر قشر ضعیف جامعه شوند.

مطلب پیشنهادی



رویای دیروز، واقعیت امروز،
پیش‌پردازش‌ها و تبدیل داده‌ها، رمز موفقیت اینترنت اشیا

3- به داده‌هایی که مصرف می‌کنید توجه داشته باشد

یکی دیگر از دلایلی که الگوریتم‌ها به آن اندازه که فکر می‌کنیم قابل اعتماد و اطمینان نیستند این است که: **داده‌هایی** که مصرف می‌کنند اغلب از دنیایی می‌آید که به شدت مملو از نابرابری، تعصب و تبعیض است. شاید آمار جرایم این طور نشان ندهد ولی به عنوان مثال، به مکانیسم سیستم عدالت کیفری امریکا نگاه کنید. این سیستم همواره در طول تاریخ اقلیت‌ها را ناعادلانه هدف قرار داده است. این جانب‌داری در داده‌ها مربوط به جرایم به خوبی مشخص است. محققان می‌دانند که میزان مصرف ماریجوانا در بین سفیدپوستان و رنگین‌پوستان تقریباً مشابه هم است، اما احتمال دستگیری جوانان رنگین‌پوست به دلیل حمل ماریجوانا بیشتر از جوانان سفیدپوست است. این اختلاف و تضاد رفتاری به سیستم نژادپرستی و حضور دائم نیروهای پلیس در مناطق رنگین‌پوست نشین مربوط می‌شود. ما به عنوان یک جامعه تصمیم گرفته‌ایم تا تبعیض نژادی، جنسیتی، تمایلات جنسی، معلولیت و ناتوانی جسمی را کنار بگذاریم و خوشبختانه، بسیاری از دانشمندان داده هنگام استفاده از این خصوصیات برای گروه‌بندی افراد و یا مدل رفتاری یا دقت زیادی عمل می‌کنند. اما داده دنیای واقعی اغلب گویای چیز دیگری است.



بزرگ داده‌ها بازیگر اصلی فناوری‌های فردا
بزرگ داده‌ها چه هستند و چرا باید به آن‌ها اهمیت دهیم؟

4- در رابطه با چیزی که مدل‌سازی می‌کنید عدالت به خرج دهید

رفتار انسان خیلی پیچیده، به هم ریخته و نامنظم است و به همین دلیل، ارزیابی مستقیم این خصوصیات که سعی در مدل‌سازی آن‌ها داریم در واقع وجود ندارد. از همین رو، دانشمندان داده‌ها اغلب روی متغیرهای دیگری تکیه دارند که از نظر آن‌ها با چیزهایی که سعی در ارزیابی دارند در ارتباط هستند. به‌عنوان مثال، شرکت‌های بیمه خودرو از امتیازات راننده به‌عنوان راهی برای تعیین میزان اطمینان‌پذیری وی استفاده می‌کنند. شاید با یک نظر اجمالی، این طور به نظر برسد که راننده‌ای که همیشه قبض‌های خود را در موعد مشخص پرداخت می‌کند، مسئولیت‌پذیری بالاتری دارد. اما Consumer Reports اخیراً به نتایجی دست یافته است که خیلی عجیب به نظر می‌رسد. افرادی که امتیاز منفی کمی دارند و سابقه پاک‌ی در زمینه رانندگی دارند بیشتر از رانندگانی که امتیاز منفی دارند از طرف شرکت‌های بیمه جریمه می‌شوند.

افرادی که امتیاز منفی کمی دارند و سابقه پاک‌ی در زمینه رانندگی دارند بیشتر از رانندگانی که امتیاز منفی دارند از طرف شرکت‌های بیمه جریمه می‌شوند

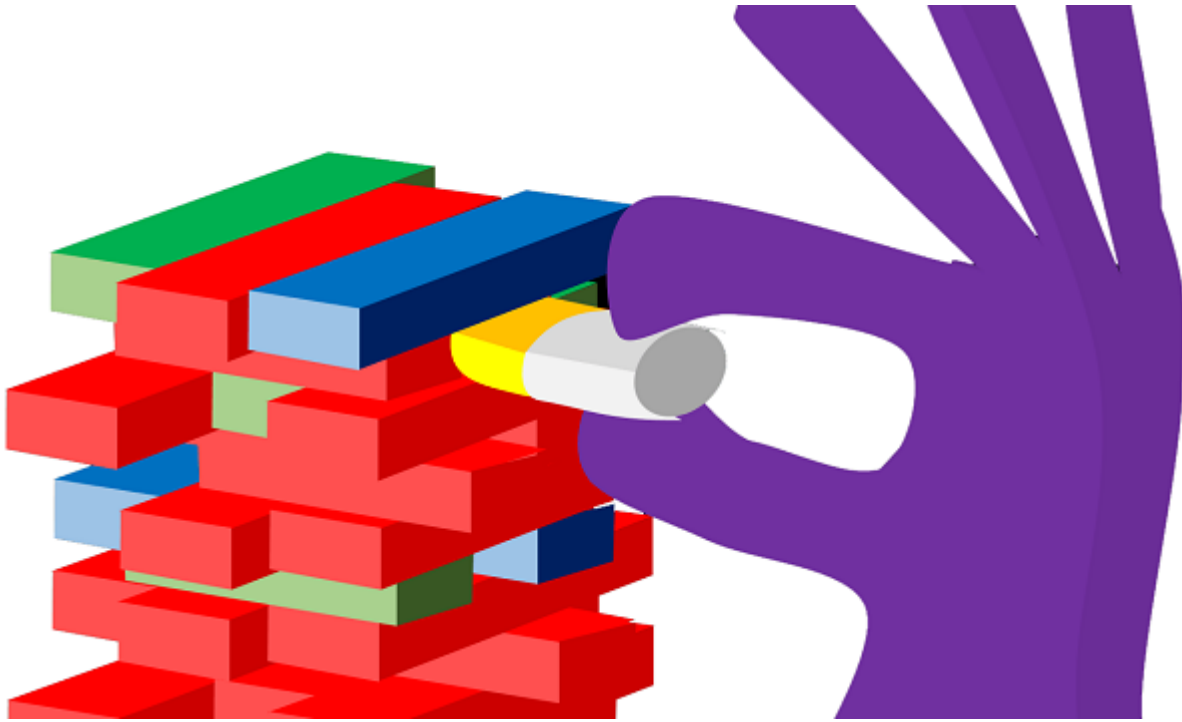
از طرفی، اونیل معتقد است حتماً دلیل پنهانی وجود دارد که شرکت‌های بیمه همچنان از امتیاز رانندگان در مدل‌های خود استفاده می‌کنند: در واقع، این نمونه بارزی از ارزیابی مستقیم آسیب‌پذیری مالی است. رانندگانی که امتیاز منفی کمی دارند بیشتر دنبال نرخ‌های پایین‌تر هستند، ولی فردی که به‌شدت به دنبال بیمه کردن خودروی خود است حاضر است مبلغ بیشتری هم برای آن پرداخت کند.



روای دیروز، واقعیت امروز
پیش‌پردازش‌ها و تبدیل داده‌ها، رمز موفقیت اینترنت اشیا

5 - فرضیات خود را از نظر سیستمی ارزیابی و آزمایش کنید

این احتمال وجود دارد که حتی الگوریتم‌هایی که به‌خوبی نوشته شده‌اند فرضیات غلطی در خود داشته باشند. این قضیه درباره تلاش برای حفظ امنیت جوامع از طریق بازداشت افرادی که پتانسیل خشونت و ارتکاب جرم را دارند و آزاد کردن افرادی که رفتار خوبی از خود نشان داده‌اند به‌خوبی قابل بررسی است. کسانی که موافق این سیستم هستند اعتقاد دارند که این کار باعث کاهش جمعیت زندان‌ها و برقراری بهتر عدالت در سیستم کیفری می‌شود. اما اونیل می‌گوید به محض اینکه کسی به زندان می‌افتد مانند یک جعبه سیاه با وی رفتار می‌شود و دیگر سؤالی پرسیده نمی‌شود.



اما روش غول‌های آنلاین دنیا مانند Amazon.com عکس آن است. در واقع، یادگیری، آموزش و آزمایش کردن در مدل کسب و کار آن‌ها در نظر گرفته شده است. شرکت آمازون یک آزمایشگاه تخصصی داده دارد که محققان در آن به طور پیوسته در حال بازبینی و ارزیابی چند باره تمام جنبه‌های تجارب مشتریان خود هستند. آن‌ها با این کار سعی دارند در طول فرآیند درخواست کالا توسط مشتری تا تحویل آن، هر چیزی که باعث سردرگمی و کلافگی مشتری می‌شود را پیدا کنند. این فیدبک باعث یادگیری دائمی آمازون می‌شود و در نتیجه سود کسب و کار آنلاین آن‌ها بالاتر می‌رود. اونیل می‌گوید اگر واقعاً می‌خواستیم سیستم عدالت کیفی را بهینه‌سازی کنیم، باید دائم آزمایش‌های کنترل شده‌ای را انجام می‌دادیم. وقتی کسی پشت میله‌های زندان می‌رود و در کنار سایر مجرمان قرار می‌گیرد، احتمال ارتکاب جرم دوباره در وی بالا می‌رود یا پایین می‌آید؟ سلول انفرادی چه مزایایی دارد؟ هزینه درمان فردی که از اختلالات ذهنی رنج می‌برد در برابر بازداشت چند باره او چقدر است؟

در پایان باید گفت برای اینکه سرعتی برابر با توسعه فناوری داشته باشیم، نیاز به قوانین و استانداردهای صنعتی در کنار سطح مشخصی از شفاف‌سازی شرکت‌ها درخصوص نحوه استفاده آن‌ها از داده‌ها داریم. در این بین، برنامه‌نویسان نقش مهمی را ایفا می‌کنند و وظیفه سنگینی بر دوش آن‌ها است. اولین گام حیاتی در این زمینه کسب آگاهی درباره این مسئله است. شاید بهترین کار عمل کردن به تعهداتی باشد که در سال 2008 و در آستانه بحرانی مالی جهانی مطرح شد:

- به خاطر خواهم داشت که من دنیا را نساخته‌ام و این دنیا توقعات من را برآورده نمی‌کند.
- برای همین و به منظور ارزیابی ارزش‌ها از مدل‌ها استفاده می‌کنم و همیشه تحت تأثیر ریاضیات نخواهم بود.
- هیچ‌گاه واقعیت را فدای زیبایی نمی‌کنم، مگر اینکه دلیل انجام آن را توضیح دهم.
- هیچ‌گاه به افرادی که از مدل من استفاده می‌کنند، اطمینان نادرست درخصوص میزان دقت آن نمی‌دهم.
- درک می‌کنم که کار من می‌تواند تأثیر شگرفی بر روی جامعه و اقتصاد داشته باشد که بیشتر آن‌ها فراتر از قدرت درک من است.

منبع:

[Spectrum](#)

[ieee](#)

تاریخ انتشار:

03 دی 1396

نشانی منبع:

<https://www.shabakeh-mag.com/cover-story/10804/%D8%A8%D8%B2%D8%B1%DA%AF-%D8%AF%D8%A7%D8%AF%D9%87%E2%80%8C%D9%87%D8%A7-%D9%85%DB%8C%E2%80%8C%D8%AA%D9%88%D8%A7%D9%86%D9%86%D8%AF-%D8%A8%D9%87-%D8%B3%D9%84%D8%A7%D8%AD%E2%80%8C%D9%87%D8%A7%DB%8C-%D9%85%D8%AE%D8%B1%D8%A8-%D9%88-%D8%AE%D8%B7%D8%B1%D9%86%D8%A7%DA%A9-%D8%AA%D8%A8%D8%AF%DB%8C%D9%84-%D8%B4%D9%88%D9%86%D8%AF>