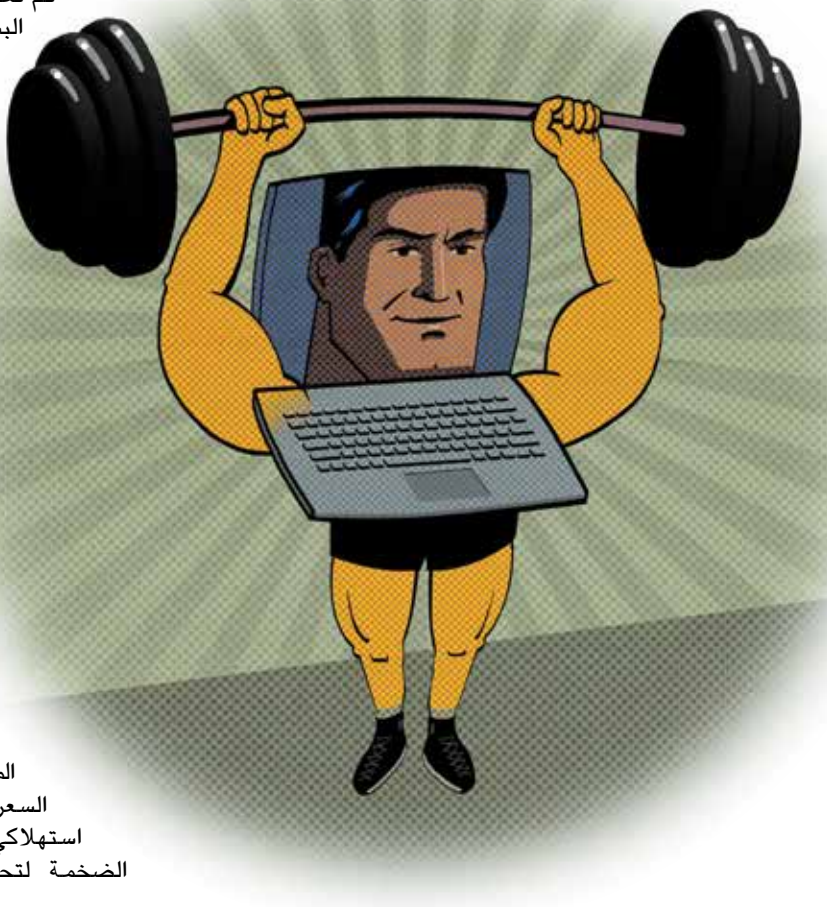
القوة الضخمة للبيانات الضخمة

قوة الحوسبة تدفع التعلم الآلي وتحول الأعمال والتمويل

سانجيف رانجان داس

يمكن أن يصل العالم الآن إلى بيانات أكثر مما كان يمكن تصوره، حتى منذ عقد مضى. وتقوم الشركات بتجميع بيانات جديدة بمعدل أسرع من قدرتها على تنظيمها وفهمها. وعليها الآن اكتشاف كيفية استخدام هذا الكم الضخم من البيانات لاتخاذ قرارات أفضل وتحسين أدائها.

ويسعى هذا المجال الجديد المعني بعلوم البيانات إلى استخلاص المعرفة التطبيقية من البيانات، خاصة البيانات الضخمة، وهي المجموعات الكبيرة جدا من البيانات التي يمكن تحليلها للكشف عن أنماط واتجاهات وارتباطات. وتمتد علوم البيانات من جمع البيانات وتنظيمها إلى تحليلها والحصول على رؤى منها، وفي النهاية إلى التنفيذ العملي لما تم تعلمه. ويتداخل هذا المجال مع كافة الأنشطة البشرية — والاقتصاد، والتمويل والأعمال ليست استثناء.



وتجلب علوم البيانات أدوات التعلم الآلي، وهو نوع من أنواع الذكاء الاصطناعي الذي يعطي للحواسيب القدرة على التعلم بدون برمجة فعلية (دراسة Samuel, 1959). ولدى هذه الأدوات، مقترنة بالكميات الضخمة من البيانات، القدرة على تغيير ساحة إدارة الأعمال وتحليل السياسات الاقتصادية. وتبشر بعض من هذه التغيرات بكثير من الوعود.

تحديد ملامح المستهلكين

إن النمو السريع في تطبيق علم البيانات في الأعمال ليس أمرا مفاجئا نظرا لقوة الاعتبارات الاقتصادية في هذا العلم. ففي السوق التنافسية، يدفع جميع المشترين نفس السعر، وإيرادات البائع تساوي هذا السعر مضروبا في الكمية المباعة. غير أن هناك الكثير من المشترين الذين لديهم استعداد لدفع أكثر من السعر التوازني، ويحتفظ هؤلاء المشترون بفائض استهلاكي يمكن استخراجه باستخدام البيانات الضخمة لتحديد ملامح المستهلكين.

كما أن تحميل المستهلكين أسعار مختلفة استنادا إلى ملامحهم التي تم تحليلها يتيح للشركات إمكانية الحصول على أعلى سعر يكون المستهلك على استعداد لدفعه مقابل منتج معين ويكون لديه القدرة على دفعه. ويكون تحديد أمثل تمييز سعري أو تجزؤ للسوق باستخدام البيانات الضخمة مربحا للغاية. وكانت هذه الممارسة هي العرف السائد في بعض الصناعات، مثل صناعة الطيران، ولكنها تمتد حاليا عبر طيف واسع من المنتجات.

وبالإضافة إلى ذلك، فإن المكاسب من الاستهداف السعري تمكّن الشركات أيضا من تقديم خصومات للمستهلكين الذين لم يكن بوسعهم تحمل السعر التوازني بخلاف ذلك، مما يزيد بالتالي الإيرادات ويوسع قاعدة المستهلكين، ومن المحتمل أن يزيد الرفاهية الاجتماعية. ويعتبر تحديد ملامح المستهلكين باستخدام البيانات الضخمة سببا مهما للتقييمات العالية لشركات مثل فيسبوك وغوغل وأكسيوم، التي تقدم منتجات وخدمات تعتمد على بيانات عملائها.

وفي حين أنه يمكن استخدام البيانات الضخمة لاستغلال المستهلكين، فإنها تغير أيضا الممارسات التجارية بطريقة تساعد هؤلاء المستهلكين أنفسهم. وتستخدم الشركات البيانات التي تولدها التفاعلات بين الناس عبر وسائط التواصل الاجتماعي لفهم سلوكهم الائتماني بشكل أفضل. ويؤدي ربط التاريخ الائتماني السابق للناس بتواجدهم على وسائط التواصل الاجتماعي إلى تحسين نظم قياس الجدارة الائتمانية. كما يسمح للمقرضين بتقديم الائتمان إلى الناس الذين قد لا يحصلون على ائتمان بخلاف ذلك.

وعلى وجه الخصوص، تقضي البيانات الضخمة على التحيزات الناجمة عن القرارات التي يتخذها الناس على أساس معلومات محدودة. وأدى هذا النقص في البيانات الفردية الدقيقة إلى الرفض التمييزي لطلبات الحصول على قروض، وهي ممارسة يعود تاريخها إلى ثلاثينات القرن الماضي. فيرسم مقرضو الرهون العقارية خطوطا حمراء حول مناطق على الخريطة للإشارة إلى أنهم لن يقدموا قروضا فيها بسبب التكوين العنصري أو العرقي لسكانها. وحالت ممارسة التنميط هذه دون منح الائتمان لشرائح كاملة من المجتمع.

وعلى العكس، تلغي البيانات الضخمة التنميط. فيمكن الآن استبدال البيانات الشخصية غير الدقيقة ببيانات أكثر دقة وأكثر صلة بالفرد. ويمكن أن تستغل شركات قياس الجدارة الائتمانية عدم التجانس الذي تكشف عنه التفاعلات بين الناس في وسائط التواصل الاجتماعي، وتدفقات الرسائل النصية، والمدونات الصغيرة، وأنماط بطاقات الائتمان، وبيانات تحديد الملامح، بالإضافة إلى البيانات الديمغرافية النمطية مثل الدخل، والعمر، والموقع (دراسة Wei and others, 2014). وييسر استخدام البيانات الأكثر دقة تصنيف الأفراد بشكل أفضل حسب الجودة الائتمانية.

التنبؤ وتحليل المخاطر

تغير التنبؤ الاقتصادي تغيرا هائلا نتيجة أساليب علوم البيانات. ففي التنبؤ التقليدي، لا تتوافر الإحصاءات الرئيسية عن الاقتصاد، مثل التقرير ربع السنوي لإجمالي الناتج المحلي، إلا بعد فترة تأخير طويلة. ويمكن لعلوم البيانات تجنب هذه التأخيرات عن طريق الاعتماد على المعلومات المبلغ عنها بمعدل أكثر تواترا، مثل معدلات البطالة، أو الطلبات الصناعية، أو حتى الشعور السائد إزاء الأخبار، للتنبؤ بالمتغيرات المبلغ عنها بمعدل أقل تواترا.

وتُعرف مجموعة المناهج المشمولة في هذا النشاط باسم «التنبؤ الآني» — وتسمى أيضا بتنبؤ الحاضر — ولكنها تُفهم بشكل أفضل على أنها تنبؤ في الوقت الحقيقي (راجع «ملكة الأرقام»، في عدد مارس ٢٠١٤ من مجلة التمويل والتنمية).

وتحقق علوم البيانات نجاحات أيضا في مجال تحليل المخاطر المالية النظامية. فقد أصبح العالم أكثر ترابطا من أي وقت مضى، ويبشر قياس هذه الروابط برؤية جديدة في عملية صنع القرارات الاقتصادية.

ويعتبر النظر إلى المخاطر النظامية من خلال عدسة الشبكات منهجا قويا. ويستخدم أخصائيو البيانات الآن بيانات غزيرة لبناء صور عن التفاعلات فيما بين البنوك، وشركات التأمين، والسماسرة، وغيرهم. فمن الواضح أن معرفة البنوك الأكثر ترابطا عن غيرها ستكون مفيدة. وينطبق الشيء نفسه على المعلومات المتعلقة بالبنوك الأكثر تأثيرا في السوق، والتي تُقاس باستخدام أسلوب يستند إلى القيم الذاتية. وبعد بناء هذه الشبكات، يمكن أن يقيس أخصائيو البيانات درجة المخاطر في النظام المالي، فضلا عن مساهمة فرادى المؤسسات المالية في المخاطر الإجمالية، مما يوفر للمنظمين طريقة جديدة لتحليل المخاطر النظامية — وإدارتها في نهاية المطاف. راجع Espinosa-Vega and Solé (2010)؛ و IMF (2010)؛ وBurdick and others (2011)؛ وDas (2016).

وتستند هذه المناهج بشدة إلى رياضيات الشبكات الاجتماعية التي تطورت في مجال علم الاجتماع، والتي تطبق على الشبكات الكبيرة جدا باستخدام نماذج متقدمة لعلوم الحاسوب، وتبلغ ذروتها بالاندماج المثمر لعدة مجالات أكاديمية.

أكثر من مجرد كلمات

تعتبر تحليلات النصوص مجالا سريع النمو من مجالات علوم البيانات ومكملا مثيرا للاهتمام للبيانات الكمية في مجال التمويل وعلم الاقتصاد (راجع «وجهان للتغيير»، في هذا العدد من مجلة التمويل والتنمية). فقد ظهرت أعداد كبيرة من التطبيقات التجارية التي تستند إلى التنقيب في النصوص: حيث تستنتج شركات مثل iSentium الشعور بعيد الأجل وقصير الأجل السائد في السوق من وسائط التواصل الاجتماعي باستخدام تويتر؛ وتقدم شبكة StockTwits مؤشرات عن الشعور السائد في السوق من خلال تطبيق على الويب يمكن استخدامه عبر الأجهزة المتنقلة.

تقضي البيانات الضخمة على التحيزات الناجمة عن القرارات المبنية على أساس معلومات محدودة.

ومن الممكن الآن تصنيف شركة ما حسب نتائج الأرباح ربع السنوية التي تعلن عنها في النموذج 10-K، وهو تقرير سنوي عن الأداء المالي للشركة يقدم إلى هيئة الأوراق المالية وعمليات البورصة الأمريكية (SEC). ويقدم سجل للكلمات المتعلقة بالمخاطر في التقارير ربع السنوية نظام تصنيف دقيق للتنبؤ بالأرباح. فالشركات التي تمتلك تقارير ربع سنوية أصعب في القراءة تميل إلى أن تكون قد حققت أرباحا سيئة — ومن المرجح أن يرجع ذلك إلى محاولة هذه الشركات الإبلاغ عن أخبار سيئة باستخدام صيغة ما للتشويش (راجع دراسة Loughran and McDonald, 2014). وباستخدام مقياس قديم لسهولة القراءة، وهو مؤشر Gunning Fog، من السهل إعطاء درجات للتقارير المالية بشأن هذه الخاصية، وينظر المنظمون مثل مكتب الحماية المالية للمستهلكين في وضع معايير بشأن سهولة قراءة التقارير المالية.

وخلصت الدراسات إلى أن مجرد طول التقرير ربع السنوي يكفي للكشف عن الأخبار السيئة (التقارير الطويلة تشير إلى انخفاض الأرباح)، ومرة أخرى لأن التشويش يرتبط باستخدام الكلام الزائد؛ وقياسا بذلك، فإن حجم الملف وحده الخاص بتقارير الشركات التي ترفع على موقع الإنترنت الخاص بهيئة الأوراق المالية وعمليات البورصة الأمريكية يعطي مؤشرا على أداء الأرباح ربع السنوي. ومن المتوقع تحقيق المزيد من النجاح في إطار هذا المجال سريع التطور.

وينطوي مجال جديد يُعرف باسم «تحليلات الأخبار» على التنقيب في الأخبار للبحث عن بيانات. وتتزايد الخدمات التي تقدمها شركات مثل RavenPack، وتتراوح ما بين إعطاء درجات للشعور السائد في السوق وتحليلات التنبؤ بالتجارة إلى التنبؤ الاقتصادي الكلي. وتنقب شركة RavenPack في كميات ضخمة من البيانات غير المنظمة الواردة في الأخبار ووسائط التواصل الاجتماعي وتحولها إلى بيانات ومؤشرات دقيقة لدعم الشركات المالية في إدارة الأصول، وصنع السوق، وإدارة المخاطر والامتثال.

خلصت الدراسات إلى أن مجرد طول التقرير ربع السنوي يكفي للكشف عن الأخبار السيئة.

وضمن هذه الفئة، يعتبر تحليل تدفق الأخبار مثيرا للاهتمام بشكل خاص. فتنقب صناديق التحوط في آلاف من رسائل الأخبار يوميا بهدف تحديد أهم خمسة أو عشرة موضوعات ثم تتتبع التطور في نسبة الموضوعات من يوم لآخر للكشف عن تغيرات في أوضاع السوق يمكن اتخاذ قرارات تداول على أساسها. وقد يكون التحليل المماثل لذلك مفيدا لصناع السياسات والمنظمين، مثل محافظي البنوك المركزية. وعلى سبيل المثال، قد يكون الوقت قد حان لإعادة النظر في سعر الفائدة الرسمي عندما تتغير نسبة موضوعات معينة تُناقش في الأخبار (مثل التضخم، أو أسعار الصرف أو النمو) بشكل مفاجئ.

ويبدأ تحليل الموضوعات بإعداد جدول ضخم لتكرار الكلمات، يُعرف باسم «مصفوفة مصطلحات الوثائق»، ويعد فهرس لآلاف المقالات الإخبارية. وترد المصطلحات (الكلمات) في صفوف الجدول، ويمثل كل مقال أخبار عمودا في الجدول. ويمكن أن تكشف هذه المصفوفة الضخمة عن موضوعات من خلال تحليل رياضي للارتباط بين الكلمات وبين الوثائق. وتتم فهرسة مجموعات الكلمات والموضوعات المكتشفة من خلال التعلم الآلي مثل فهرسة الدلالات الكامنة وتخصيص ديريتشليت الكامن (LDA). ويسفر تحليل تخصيص ديريتشليت الكامن عن مجموعة من الموضوعات وقوائم الكلمات التي تظهر في هذه الموضوعات.

وتعتبر مناهج إعداد النماذج هذه فنية للغاية لمناقشتها هنا، ولكنها في الواقع مجرد تقنيات إحصائية تكشف عن مبدأ تجميع الكلمات الرئيسية الواردة في مجموعة من الوثائق (على سبيل المثال، في تدفق الأخبار). ومن المرجح أن صناع السياسات سيستخدمون هذه الأدلة اللغوية على نطاق واسع في عملية صنع القرارات السياسية، وعلى سبيل المثال في إعادة تعريف رسالة حملة سياسية.

الذكاء الاصطناعي والمستقبل

أصبحت أجهزة الحاسوب أقوى مما كانت عليه في أي وقت مضى، وحفزت قدرتها على تجهيز كميات ضخمة من البيانات العمل في مجال الذكاء الاصطناعي. وقد ثبت أن فئة جديدة من الخوارزميات تعرف باسم «شبكات التعلم المتعمق»، المستوحاة من الشبكات العصبية البيولوجية، لديها قوة هائلة في محاكاة الطريقة التي يعمل بها المخ البشري، وقدمت أمثلة ناجحة عديدة على الذكاء الاصطناعي.

والتعلم المتعمق هو منهجية إحصائية تستخدم شبكات عصبية اصطناعية بهدف الربط بين عدد كبير من متغيرات المدخلات ومتغيرات المخرجات، أي عملية تحديد للأنماط. ويتم تحليل المعلومات من خلال شبكة من الخلايا العصبية القائمة على السيليكون والبرمجيات. وتستخدم البيانات لتعزيز الروابط بين هذه الخلايا العصبية مثلما يتعلم البشر من التجارب مع مرور الوقت. وتنقسم أسباب النجاح المذهل للتعلم المتعمق إلى شقين: توافر كميات ضخمة من البيانات لتتعلم منها الآلات والنمو المطرد في قوة الحوسبة، المدفوع بالتطور في رقاقات الحاسوب ذات الأغراض الخاصة والمتعلقة بتطبيقات التعلم المتعمق.

ويعطى التعلم المتعمق قوة دفع لجزء كبير من التكنولوجيا المعاصرة التي بدأ العالم يعتبرها أمرا مسلما به، مثل الترجمة الآلية، والسيارات ذاتية القيادة، والتعرف على الصور والتوسيم. ومن المرجح أن تغير هذه الفئة من التكنولوجيا علم الاقتصاد والسياسات في القريب العاجل. وتستخدمها بالفعل هيئات التصنيف الائتماني لإعداد تقارير بدون تدخل بشري. وقد تقدم الشبكات العصبية الكبيرة للتعلم المتعمق قريبا تنبؤات وتحدد العلاقات بين المتغيرات الاقتصادية بطريقة أفضل من الأساليب الإحصائية القياسية.

ومن الصعب التنبؤ بمجالات العلوم غير المثيرة التي ستشهد أكبر نمو في استخدام التعلم الآلي، ولكن هذا العصر الجديد قد وصل بالتأكيد. وكما أشار كاتب الخيال العلمي ويليام جيبسون، «المستقبل موجود هنا بالفعل؛ إلا أنه غير موزع توزيعا عادلا.» ■

سانجيف رانجان داس أستاذ في كلية ليفي لإدارة الأعمال في جامعة سانتا كلارا.

المراجع:

Billio, Monica, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon, 2012, "Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors," Journal of Financial Economics, *Vol. 104, No. 3, pp. 535–59.*

Burdick, Douglas, Mauricio A. Hernandez, Howard Ho, Georgia Koutrika, Rajasekar Krishnamurthy, Lucian Popa, Ioana Stanoi, Shivakumar Vaithyanathan, and Sanjiv Das, 2011, "Extracting, Linking and Integrating Data from Public Sources: A Financial Case Study," IEEE Data Engineering Bulletin, *Vol. 34, No. 3, pp. 60–7.*

Das, Sanjiv, 2016, "Matrix Metrics: Network-Based Systemic Risk Scoring," Journal of Alternative Investments, *Vol. 18, No. 4, pp. 33–51.*

Espinosa-Vega, Marco A., and Juan Solé, 2010, "Cross-Border Financial Surveillance: A Network Perspective," IMF Working Paper 10/105 (Washington: International Monetary Fund).

International Monetary Fund (IMF), 2010, "Systemic Risk and the Redesign of Financial Regulation," Global Financial Stability Report, *Chapter 2 (Washington, April).*

Lin, Mingfen, Nagpurnanand Prabhala, and Siva Viswanathan, 2013, "Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending," Management Science, *Vol. 59, No. 1, pp. 17–35.*

Loughran, Tim, and Bill McDonald, 2014, "Measuring Readability in Financial Disclosures," Journal of Finance, *Vol. 69, No. 4, pp. 1643–71.*

Samuel, A.L., 1959, "Some Studies in Machine Learning Using the Game of Checkers," IBM Journal of Research and Development, *Vol. 3, No. 3, pp. 210–29.*

Wei, Yanhao, Pinar Yildirim, Christophe Van den Bulte, and Chrysanthos Dellarocas, 2015, "Credit Scoring with Social Data," Marketing Science, *Vol. 352, pp. 234–58.*