أخبار سيئة للأخبار الكاذبة: بحثٌ جديد يساعد في مكافحة المعلومات المضلِلة في وسائل التواصل الاجتماعي

أنشومالي شريفاستافا، وهو أستاذ مساعد لعلم الحاسوب في جامعة رايس. حقوق الصورة: Jeff Fitlow/Rice University.

اكتشف باحثو جامعة رايس (Rice University) طريقةً أكثر فاعليّةً لمنع المعلومات المضلِلة من الانتشار عبر الإنترنت، وذلك عن طريق استخدام عوامل تصفيةٍ احتماليةٍ، والتي تُدرَّب باستخدام الذكاء الصنعي.

عبر الإنترنت، في مؤتمر **NeurIPS 2020** حول أنظمة معالجة المعلومات العصبية، وضّح عالم الحاسوب أنشومالي شريفاستافا **Anshumali Shrivastava** وطالب الدراسات العليا تشنوا داي **Zhenwei Dai** النّهج الجديد لمسح وسائل التواصل الاجتماعي في

دراسةٍ قُدِّمت في العاشر من ديسمبر/كانون الأول 2020. تطبّق طريقتهما التعلم الآلي بطريقةٍ أكثر ذكاءً لتحسين أداء مرشحات بلوم **Bloom Filters**، وهي تقنيةٌ مستخدمةٌ على نطاق واسع ابتُكِرت منذ نصف قرن، وهي تُستخدَم في التصفية أو لاختبار انتماء عنصرٍ لمجموعةٍ من عدمه.

باستخدام قواعد بيانات اختبارية للأخبار المزيّفة وفيروسات الحاسوب، أظهر شريفاستافا وداي أن مرشح بلوم الذي يستخدم تقنية التعلم التكيفي **Ada-BF** يتطلّب ذاكرة أقل بنسبة 50% لتحقيق نفس مستوى أداء مرشحات بلوم التي تستخدم التعلّم الآلي.

لشرح نهج التصفية الخاص بهما، استشهدا ببعض البيانات من تويتر، وكشفت شركة التواصل الاجتماعي العملاقة مؤخرًا أن مستخدميها أضافوا نحو 500 مليون تغريدةٍ جديدةٍ يوميًا، وأن التغريدات تظهر عادةً على الإنترنت بعد ثانيةٍ واحدةٍ من نقر المستخدم فوق إرسال.

يصرّح شريفاستافا: "في وقتٍ قريبٍ من الانتخابات الأمريكية، كانوا يتلقون نحو 10,000 تغريدةٍ في الثانية، ومع زمن انتقال مدته ثانية واحدة، فإنّ ذلك يعني نحو ست تغريدات لكل مللي ثانية، إذا كنت تريد تطبيق مرشح يقرأ كل تغريدة ويضع علامةً على تلك التي تحتوي على معلومات معروف أنها مزيفة، فلا يمكن أن تكون آلية الإبلاغ لديك أبطأ من ستة أجزاء من الثانية، أو ستتخلف عن الركب ولن تلحق بالتغريدة أبدًا".

إذا أُرسِلت التغريدات التي وُضِعت علامةٌ عليها من أجل الحصول على مراجعةٍ يدويّةٍ إضافيّة، فمن المهم أيضًا أن يكون معدّل الإيجابيّة الكاذبة منخفضًا. بمعنًى آخر، تحتاج إلى تقليل عدد التغريدات الحقيقية التي أُبلِغ عنها عن طريق الخطأ.

وقال: "إذا كان معدّل الإيجابية الكاذبة الخاص بك منخفضًا مثل 0.1%، فحتى مع ذلك فأنت تبلّغ عن طريق الخطأ عن 10 تغريدات في الثانية، أو أكثر من 800,000 في اليوم، وترسل تلك التغريدات للمراجعة اليدوية، هذا هو بالضبط السبب في أن معظم الأساليب التقليدية للذكاء الصنعي تعجز عن التّحكم في المعلومات المضللة".

ويقول شريفاستافا إن تويتر لا يكشف عن طرقه لتصفية التغريدات، لكن يُعتقد أنهم يستخدمون مرشح بلوم، وهي تقنية ذاكرة منخفضة اختُرِعت في عام 1970، لمعرفة ما إذا كان عنصر بيانات معين جزءًا من مجموعة معروفة من العناصر، على سبيل المثال يُختبَر جزءٌ من كودٍ برمجيٍّ للحاسوب فيما إذا كان جزءًا من قاعدة بيانات فيروسات الحاسوب المعروفة. عامل تصفية بلوم مضمون للعثور على جميع الأكواد التي تطابق قاعدة البيانات، لكنه يسجل بعض الإيجابيات الكاذبة أيضًا.

ويضيف شريفاستافا: "لنفترض أنك تعرّفت على جزءٍ من المعلومات الخاطئة، وتريد التأكد من عدم نشرها في التغريدات، يسمح لك عامل تصفية بلوم بفحص التغريدات بسرعةٍ كبيرةٍ في جزءٍ من المليون من الثانية أو أقل. إذا قالت إن تغريدةً ما نظيفة، وإنها لا تتطابق مع أي شيء في قاعدة بياناتك الخاصة بالمعلومات الخاطئة، فهذا مضمون بنسبة 100%. لذلك لا توجد فرصةٌ للموافقة على تغريدةٍ تحتوي على معلوماتٍ خاطئةٍ معروفةٍ، لكن لن يكون مرشح بلوم فعّالًا طول الوقت في تحديد التغريدات غير الضارة".

خلال السنوات الثلاث الماضية، قدّم الباحثون خططًا مختلفةً لاستخدام التعلّم الآلي لزيادة قدرة فلاتر بلوم وتحسين كفاءتها. يمكن تدريب برنامج التعرف على اللغة للتعرف على معظم التغريدات والموافقة عليها، ما يقلل من الحجم الذي يجب معالجته باستخدام مرشح بلوم. يمكن أن يؤدي استخدام مصنفات التعلم الآلي إلى تقليل مقدار النفقات الحسابية اللازمة لتصفية البيانات، ما يسمح للشركات بمعالجة المزيد من المعلومات في وقتٍ أقل باستخدام نفس الموارد.

قال داي: "عندما يستخدم الناس نماذج التعلم الآلي اليوم، فإنهم يضيّعون الكثير من المعلومات المفيدة التي تأتي من نموذج التعلم الآلي".

تتمثّل الطريقة النموذجية التي يتحدث عنها البحث هي في تعيين حد للتسامح، وإرسال كل ما يقل عن هذا الحد إلى مرشح بلوم، فإذا كان حد التسامح يتعين عندما تكون درجة الثقة في المعلومات 85%، فهذا يعني أن المعلومات التي يعتبرها المصنف آمنة بمستوى ثقة 80% تتلقى نفس مستوى التدقيق التي تتلقاه المعلومات ذات 10%.

قال داي: "على الرغم من أننا لا نستطيع الاعتماد بشكلٍ كاملٍ على مصنف التعلم الآلي، فإنه لا يزال يوفر لنا معلوماتٍ قيمةً يمكن أن تقلل من كمية موارد مرشح بلوم، ما فعلناه هو تطبيق هذه الموارد بشكلٍ احتمالي، إذ نقدم المزيد من الموارد عندما يكون المصنف واثقًا من المعلومات بنسبة 10% فقط، مقابل موارد أقل قليلًا عندما يكون واثقًا بنسبة 20% وهلم جرًا، إذ نأخذ النطاق الكامل للمصنف ونحسنه باستخدام مجموعةٍ كاملةٍ من المصادر التي يمكن تخصيصها من عامل تصفية بلوم".

وقال شريفاستافا إن انخفاض حاجة **Ada-BF** إلى الذاكرة يترجم مباشرةً إلى زيادة قدرة أنظمة التصفية على العمل لحظيًا.

وأضاف: "نحن بحاجة إلى نصف المساحة فقط، لذا يمكننا بشكل أساسي معالجة ضعف كمية المعلومات باستخدام نفس المصدر".

- التاريخ: 2020-12-28
- التصنيف: تكنولوجيا

#الذكاء الاصطناعي #وسائل التواصل الاجتماعي

المصادر

- techxplore.com

المساهمون

- ترجمة
  - محمد السيد عبده
- مُراجعة
  - سارة صالح
- تحرير
  - رأفت فياض
- تصميم
  - روان زيدان
- نشر
  - احمد صلاح