لا تنتبه إلا للكلمات المهمة.. نظامٌ جديد لتعلم اللغة



يتمكن باحثو معهد ماساتشوستس للتكنولوجيا (إم آي تي) من تطوير نظام باستطاعته الحد من طاقته المستهلَكة، ومن الوقت الذي يتطلبه لتحليل الكلمات التي يستقبلها، وتوليد النص. حقوق الصورة: Jose-Luis Olivares, MIT

قد تكون لغة الإنسان غير فعالة، بعض الكلمات حيوية وضرورية، وبعضها.. مستهلَكة.

هلّا أعدت قراءة السطر الأول مرةً أخرى؟ إذا فعلت، ستجد أننا يمكن بكلمتين فقط، "لغة" و"غير فعالة"، أن نفهم مغزى الجملة كاملًا. ومن هنا، شكلت الكلمات المفتاحية أساسَ فكرةِ أداةٍ جديدة تُستخدم في معالجة اللغات الطبيعية (**Natural Language Processing NLP**)، وهي أداة آلية الانتباه **Attention mechanism** تولي اهتمامها للكلمات المفتاحية، بدلًا من أن تعامل كل الكلمات بنفس القدر

من الأهمية، ما أعطى نتائج أفضل في مهام معالجة اللغات الطبيعية مثل تحديد المشاعر الإيجابية والسلبية، والتنبؤ بأي الكلمات ستأتي تاليًا في الجملة.

غالبًا ما تأتي دقة آلية الانتباه على حساب السرعة وقدرة العمليات الحوسبية، كما أنها تعمل ببطء على المعالجات المصممة لأغراض عامة كالتي نجدها في أجهزتنا الشخصية، ولهذا صمم باحثو معهد ماساتشوستس للتكنولوجيا حاسوبًا خاصًا متضمنًا المعالجات والبرمجيات المطلوبة لتشغيل أداة آلية الانتباه، أطلقوا عليه اسم سباتن **SpAtten**، وقد أتاح هذا الحاسوب كفاءةً أعلى في معالجة اللغات الطبيعية مع تقليل قدرة الحوسبة.

يقول هانروي وانغ **Hanrui Wang**، طالب الدكتوراه في قسم الهندسة الكهربية وعلوم الحاسب، والمؤلف الرئيسي لورقة البحث: "إن نظامنا يشبه الآلية التي يتعامل بها الدماغ البشري مع اللغة، قد نقرأ بسرعة كبيرة ونركز فقط على الكلمات ذات الأهمية، وهذه نفسها فكرة عمل سباتن".

مثلت آلية الانتباه منذ طرحها لأول مرة عام 2015 التطورَ الأحدثَ في معالجة اللغة الطبيعية وأُدمجت في نماذج حديثة لها مثل **Google's BERT** و**OpenAI's GPT-3**، ويكمن الابتكار في آلية الانتباه في "الانتقائية"، فهي تملك القدرة على تخمين أيٍّ من الكلمات أو العبارات هو الأهم، بناء على مقارنةٍ تُجريها مع ما لديها من معلومات من تدريبات سابقة خضعت لها هذه الآلية. ولكن، على الرغم من السعي السريع لاعتماد هذه الآلية، فهي لا تأتي بدون تكلفة.

تتطلب نماذج معالجة اللغات الحية كمًّا هائلًا من قدرة الحاسوب، ويرجع سببُ هذا جزئيًا إلى الذاكرة التخزينية الضخمة التي تتطلبها آلية الانتباه، حتى أن وانغ يصف هذه المشكلة بعنق الزجاجة بالنسبة لتكنولوجيا معالجة اللغات الحية. ويشير إلى عدم وجود أجهزة مؤهَّلة لتشغيل هذه التكنولوجيا، فالمعالِجات العامة مثل وحدة المعالجة المركزية **Central Processing Unit (CPU)** ووحدة معالجة الرسوميات **Graphics Processing Unit (GPU)** عاجزةٌ عن مسايرة آلية عملها. وستزداد المشكلة سوءًا مع زيادة تعقيد نماذج معالجة اللغات الطبيعية أكثر، خصوصًا مع العبارات الطويلة. يقر وانغ: "نحتاج إلى تحسينات في هذه الخوارزمية، وكذلك إلى أجهزة خاصة لتشغيل هذه التكنولوجيا التي تتطور أكثر".

وبخصوص جهاز سباتن، فإن أحد الملامح الرئيسية في برمجته المتطورة هو استخدامه لتقنية "التشذيب المتسلسل" أي استبعاد البيانات غير الأساسية من حساباته، فما إن تلتقط آليةُ الانتباه الكلماتِ الرئيسية (والتي يسمونها "الرموز")، حتى يسارع سباتن بمحو الرموز الأخرى غير المهمة وإلغاء الحسابات وحركة البيانات التي كانت ستنشأ عنها. تتضمن آلية الانتباه أيضًا فروعًا حوسبيةً متعددة (تُسمى الرؤوس) تشبه الرموز، وهي تُحَدد الرؤوس غير المهمة وتُستبعد. وبعد هذه العملية، لا تدخل الرموز والرؤوس غير المهمة في حسابات الخوارزميات، ما يقلل الحمل على الذاكرة التخزينية والقدرة الحوسبية.

وللحدّ أكثر من شَغل الذاكرة، طور الباحثون تقنيةً تُسمى "التكميم التدريجي"، تسمح للخوارزمية بتخزين البيانات في أحجامٍ أصغر لتشغل حيزًا أقل ما يمكن من الذاكرة. فتُستخدم للجمل البسيطة إصداراتٌ مخفّضة الدقة، ما يشغل حيزًا صغيرًا من الذاكرة، في حين تُستخدم إصداراتٍ عالية الدقة للجمل المعقدة. يشبه الأمر أن تَعتبر عبارة "**cmptr progm**" إصدارًا مخفّضَ الدقة لعبارة "**computer program**".

وبجانب هذه التطورات في الخوارزميات، طور الباحثون بنية الجهاز التي تشغل سباتن وآلية الانتباه، ويحتوي تصميمهم لهذه البينية على درجة عالية من "الموازاة" أي إمكانية إجراء العديد من عمليات المعالجة في نفس الوقت، وهذا مفيد للغاية إذ أن آلية الانتباه تُحلل كل كلمة في الجملة على حدة. إذ يمكّن هذا التصميم سباتن من ترتيب الرموز والرؤوس حسب أهميتها (من أجل إجراء التقليمات اللازمة)

في وقت قصير، وبشكل عام فإن بنية سباتن وبرمجته تتحدان معًا لمحو البيانات غير الأساسية وغير الفعالة، للتركيز فقط على المهام اللازمة لإكمال الهدف المطلوب.

يمكننا أن نلتقط الفلسفة الكامنة وراء هذا النظام من اسمه، إذ أنه دمج كلمتي الاهتمام المتفرق **sparse attention**، كما أشار الباحثون في ورقتهم البحثية أن سباتن تأتي على قافية **Spartan** التي تعني البسيط والمقتصِد، ويعلق وانغ: "يشبه بالضبط تقنيتنا: جعل العبارات أكثر اختصارًا".

صنع الباحثون محاكاةً لتصميم بنية سباتن (حيث أنه لم يُصنع بالكامل بعد) واختبروها في منافسة مع المعالِجات العامة، فوُجد سباتن أسرعَ بـ100 مرة من منافسه التالي له، وهو **TITAN Xp GPU**، وليس هذا فحسب، بل إن سباتن وُجد أكثر كفاءةً في استهلاك الطاقة من منافسيه بـ1000 مرة، ما يشير إلى قدرته على المساعدة في تقليل احتياجات الكهرباء الكبيرة لنظام معالجة اللغات الحية.

كما دمج الباحثون سباتن في عملهم السابق للتحقق من صحة فلسفتهم القائلة بأن الأفضل لبنية الجهاز وبرمجته أن يُصنعا معًا. لقد بنوا نموذجًا متخصصًا لمعالجة اللغات الطبيعية من أجل سباتن، باستخدام محوّل البنية **Hardware-Aware Transformer (HAT)**، وحققوا سرعةً أعلى بمرتين من النموذج الأعمّ.

يعتقد الباحثون أن سباتن سيكون نافعًا للشركات التي تستخدم نماذج معالجة اللغات الطبيعية في معظم أعمال الذكاء الصنعي الخاص بها. يقول وانغ: "تتمثل رؤيتنا للمستقبل في أن الخوارزميات والأجهزة الجديدة التي تزيل الإسهاب في اللغة ستقلل من الكلفة وتوفر ميزانية الطاقة لأعمال معالجة اللغات الطبيعية".

وعلى الجانب الآخر، يمكن لسباتن أن يجلب تكنولوجيا معالجة اللغات الطبيعية إلى الأجهزة الصغيرة والشخصية، ويقول وانغ: "يمكننا أن نحسّن عمر البطارية للهاتف المحمول أو لأجهزة إنترنت الأشياء **Internet of Things (IoT)** (وهي الأجهزة التي تتصل بالإنترنت مثل التلفاز ومكبرات الصوت الذكية وغيره)، وهذا في غاية الأهمية لأنه في المستقبل ستتفاعل العديد من أجهزة إنترنت الأشياء مع الإنسان عبر الحديث الصوتيّ واللغة، ولهذا ستكون تكنولوجيا معالجة اللغة الطبيعية هي أول ما نحتاج".

يؤمن وانغ أن تركيز سباتن على إزالة الإسهاب والتكرار في النص اللغوي هو الطريق للتقدم في تكنولوجيا معالجة اللغات الطبيعية ويتابع قائلًا: "إن الأدمغة البشرية تُنشَّط بالكلمات المفتاحية، ونماذج معالجة اللغات الطبيعية التي تتبنى هذه الآلية ستكون واعدةً في المستقبل. ليست كل الكلمات متساويةً، انتبهوا فقط للمهم منها".

- التاريخ: 2021-03-15
- التصنيف: تكنولوجيا

#معهد ماساتشوستس للتكنولوجيا #اللغة

المصطلحات

- **إنترنت الأشياء (Internet of Things (IoT))**: هو نظام أجهزة الحوسبة المترابطة والآلات الميكانيكية والرقمية والأشياء والحيوانات أو الأشخاص التي يتم تزويدها بمعرِّفات فريدة unique identifiers لتصبح لها القدرة على نقل البيانات عبر شبكة ما دون الحاجة إلى التفاعل (إنسان لإنسان) أو (إنسان وحاسوب).

المصادر

- techxplore.com

المساهمون

- ترجمة
  - سماء محمد
- مُراجعة
  - سما أحمد
- تحرير
  - رأفت فياض
- تصميم
  - فاطمة العموري
- نشر
  - احمد صلاح