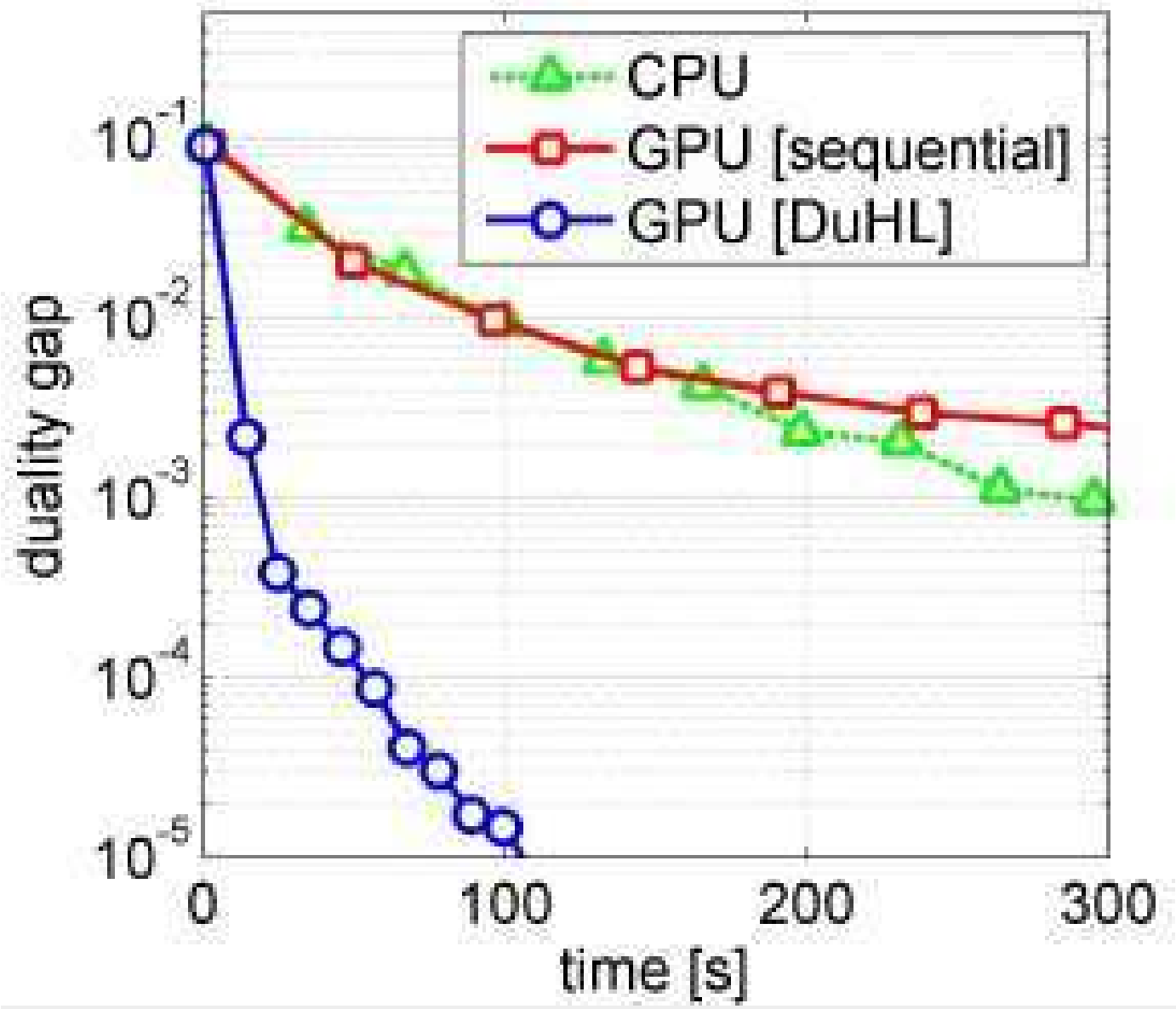
تطوير نماذج تعلم آلي أسرع من سابقاتها بعشرة أضعاف



حقوق الصورة: مدونة بحوث آي بي إم IBM Blog Research

طوَّر فريق الباحثين في شركة آي بي إم **IBM** بالتعاون مع علماء مدرسة لوزان للفنون التطبيقية الاتحادية **EPFL** نهجًا سريعًا لمجموعات بيانات التدريب الكبيرة، إذ يمكن معالجة مجموعة البيانات التدريبية ذات الـ **30** غيغابايت في أقل من دقيقة باستخدام وحدة معالجة الرسوميات (**GPU**)، وزيادة سرعة أساليب تدريب الذاكرات المحدودة عشرة أضعاف. ففي عام **2017** قُدِّمت النتائج التي استخدمت وحدة معالجة الرسوميات بكامل إمكاناتها في مؤتمر خطط التنفيذ الوطنية في لونج بيتش **LongBeach**، كاليفورنيا **California**.

يُعدُّ التدريب على نموذج التعلُّم الآلي في مجموعة البيانات ذات نطاق التيرابايت مشكلةً مشتركةً وصعبةً، فإذا كنت محظوظًا قد يكون لديك خادم مع ذاكرة كافية لتناسب جميع البيانات، ولكن سيستغرق التدريب وقتًا طويلًا جدًّا. ومن الممكن أن تستغرق هذه المسألة بضع ساعات، أو بضعة أيَّام، أو حتَّى أسابيع.

نالت الأجهزة المتخصِّصة مثل وحدة معالجة الرسوميات قبولًا متزايدًا في الكثير من المجالات، لتسريع كثافة حجم الأعمال، إلَّا إنَّه من الصعب توسيع ذلك النطاق، بسبب أعباء عمل البيانات الكبيرة. ومن أجل الاستفادة من قوَّة حوسبة وحدة معالجة الرسوميات، فإنَّنا

بحاجة لتخزين البيانات داخل ذاكرتها، من أجل الوصول إليها ومعالجتها. ومع ذلك، تملك وحدات المعالجة المركزية قدرة ذاكرة محدودة (حاليًّا **16** غيغابايت)، وهذا ليس عمليًّا لبياناتٍ كبيرةٍ جدًّا.

ويوجد حلٌّ واحدٌ ومباشرٌ لهذه المشكلة، وهو معالجة البيانات بصورة متتالية في وحدة معالجة الرسوميات على شكل دفعات، ولذلك تُقسَّم البيانات إلى أجزاء **16** غيغابايت، وتُحمَّل هذه الأجزاء على ذاكرة وحدة معالجة الرسوميات بالتتابع.

للأسف فإنَّه من المكلف نقل البيانات من وحدة معالجة الرسوميات وإليها، ويمكن أن يُشكِّل الوقت المستغرق لنقل البيانات منها إلى وحدة المعالجة المركزية نفقات عامَّة كبيرة. وفي الواقع إنَّ النفقات العامَّة كبيرةٌ جدًّا، لدرجة أنَّها قد تفوق فائدة استخدام وحدة معالجة الرسوميات في المرتبة الأولى.

انطلق فريقنا لإنشاء تقنيةٍ تُحدِّدُ الجزء الأصغر حجمًا من البيانات الأكثر أهميَّة لخوارزمية التدريب في أيِّ وقتٍ معطى. وبالنسبة إلى معظم مجموعات البيانات ذات الفائدة، فإنَّ أهميَّة كلٍّ من مجموعة البيانات في خوارزمية التدريب غير موحَّدة، وكذلك التغيُّرات التي تحصل في أثناء عملية التدريب. وبذلك يمكننا أن نتعلَّم نموذجنا بسرعة أكبر، بمعالجة نقاط البيانات في ترتيب صحيح.

حقوق الفيديو: مدونة بحوث IBM

على سبيل المثال، يمكن تخيُّل كيفية عمل خوارزمية التدريب للتمييز بين صور الكلاب والقطط، إذ يمكن لخوارزمية واحدة تمييز آذان القطط التي عادةً ما تكون أصغر من آذان الكلاب، فهي تحتفظ بالمعلومات، كذلك يمكنها تخطي استعراض هذه الميزات، وبذلك تصبح أسرع وأسرع في النهاية.

وهذا هو سبب تنوع مجموعات البيانات، وهو أمرٌ بالغ الأهمية، لأنَّ ذلك يكشف عن ميزات إضافية لن تنعكس على نماذجنا التي سنتعلمها فحسب. فإذا كان الطفل ينظر للخارج فقط والسماء دائمة الزرقة، فلن يتعلَّم أنَّه سيحلُّ الظلام في الليل أو أنَّه ستتشكَّل ظلال رمادية من

الغيوم، وهنا الشيء نفسه.

ويتحقَّق ذلك من طريق استخلاص رؤى نظريَّة جديدة بشأن كميَّة المعلومات التي يمكن أن تسهمَ بها عيِّنات التدريب الفردي في تقدم خوارزمية التعلم، ويعتمد هذا التدبير بشدَّة على مفهوم تراخيص الفجوات الازدواجية، ويتكيَّف بسرعةٍ فائقةٍ في الوقت الراهن مع خوارزمية التدريب، وتتغيَّر أهمية كلِّ نقطة من البيانات مع تقدم الخوارزمية.

وُضِعَت هذه النظرية موضع التنفيذ، وطوِّرت نماذج عناصر تدريب جديدة لتعلم الآلة على منصات حساب غير متجانسة، ودعوها باسم التعلُّم المتعدِّد المبني على الفرق بين الحلول المختلفة واختصارًا ديول **DuHL**، فضلًا عن تطبيقٍ يتضمَّن وحدات معالجة الرسوميَّات، ويمكن أن يُطبَّق هذا البرنامج على مسرِّعات ذاكرة أخرى (على سبيل المثال يمكن للنظام استخدام مصفوفة البوَّابات المنطقيَّة القابلة للبرمجة **FPGAs** بدلًا من وحدة معالجة الرسومات) والكثير من التطبيقات، بما في ذلك مجموعات البيانات الكبيرة من وسائل التواصل الاجتماعيَّة والتسويق عبر الانترنت، التي يمكن استخدامها لتنبؤ الإعلانات التي تظهر للمستخدمين. وتشمل التطبيقات الإضافية العثور على أنماط في بيانات الاتصال وكشف الاحتيال.

ونستخدم ديول في تدريب تطبيق واسع النطاق يدعم الأجهزة الناقلة الممتدة وإصدارات من **30** غيغابايت لقاعدة بيانات **ImageNet**. وقد استخدمنا وحدة معالجة رسومات من مجموعة إنفيديا كوادرو إم **NVIDIA Quadro M4000 4000** مع ذاكرة بسعة **8** غيغابايت، وأصبح بإمكاننا أن نرى النظام الذي يستخدم التجميع المتسلسل، الذي يؤدِّي في الواقع دورًا أسوأ من وحدة المعالجة المركزية نفسها، في حين أنَّ النهج الجديد يحقِّق سرعة أكبر بعشرة أضعاف من وحدة المعالجة المركزية.

والهدف التالي لهذا العمل هو تقديم ديول كخدمة سحابية، ففي البيئة السحابية عادةً ما تُرسل الموارد مثل وحدة معالجة الرسومات فواتير على مدار الساعة. ولذلك، إذا كان بإمكان المرء تدريب نموذج تعلم آلي في ساعة واحدة بدلًا من **10** ساعات، فهذا يترجم مباشرة إلى توفير تكلفة كبيرة جدًّا. ونتوقَّع أن يكون هذا الأمر ذا قيمة كبيرة بالنسبة إلى الباحثين والمطورين وعلماء البيانات الذين يحتاجون إلى تدريب نماذج تعلم آليّ على نطاقٍ واسعٍ.

هذا البحث هو جزء من بحث شركة آي بي إم في سبيل تطوير التعليم العميق الموزع (**DDL**)، والبرمجيات، والخوارزميات التي تقوم بأتمتة وتحسين مهام الحوسبة الكبيرة والمعقدة عبر مئات من مسرعات وحدة معالجة الرسومات المرفقة بعشرات الخوادم.

- التاريخ: 2018-06-22
- التصنيف: تكنولوجيا

#الخوارزميات الحاسوبية #نقل البيانات #شركة IBM

المصطلحات

- **الغاز (Gas)**: أحد الحالات الأساسية الثلاث للمادة. في هذه الحالة تتحرك الذرات، أو الجزيئات، أو الأيونات بحُريّة، فلا ترتبط

مع بعضها البعض. وفي علم الفلك، تُشير هذه الكلمة عادةً إلى الهيدروجين أو الهيليوم. المصدر: ناسا

المصادر

- Phys
- الصورة

المساهمون

- ترجمة
  - ولاء الحايك
- مُراجعة
  - حنان مشقوق
- تحرير
  - رأفت فياض
  - كرار رحيم حبيب
- تصميم
  - أحمد أزميزم
- نشر
  - كرم الحلبي