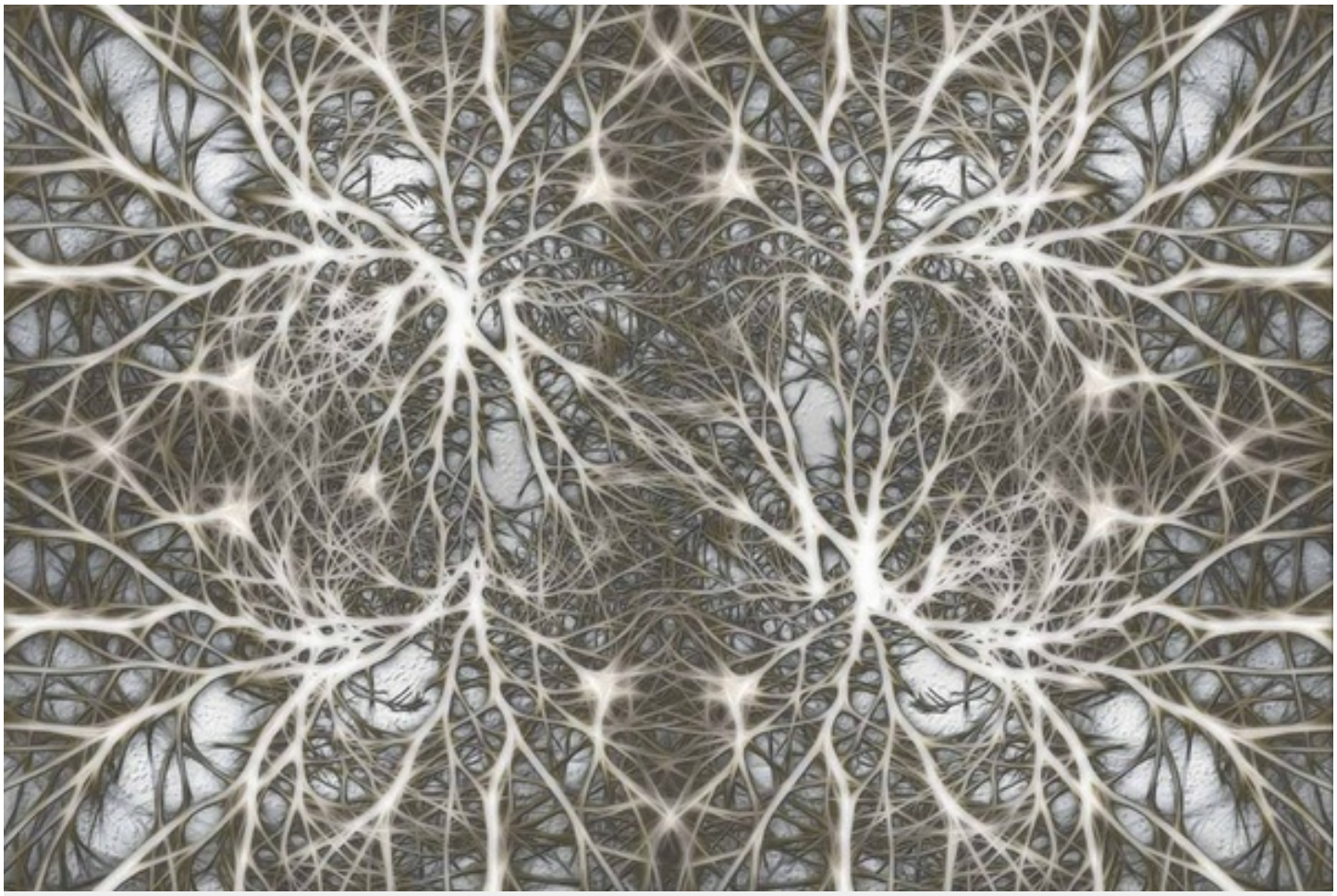
ماذا تعرف عن خوارزمية البحث عن المعمارية العصبية الأفضل؟

تقدم هذه المقالة عرضاً مجملاً عن خوارزمية البحث عن المعمارية العصبية الأفضل، ومدى علاقتها بطريقة تحسين اختيار البارمترات الأساسية.



خلية عصبية

يوفر مفهوم التعليم العميق (**deep learning**) الفرصة لتجاوز المشاكل التي ترافق أحياناً تنفيذ ما يسمى بهندسة الخصائص (**feature engineering**) وذلك من خلال تعلّم التمثيل المقترن بالنماذج الإحصائية ذات الأسلوب المتكامل (**end-to-end fashion**).
ومع ذلك، فإنّ تصميم معماريات الشبكات العصبية نفسها عادةً ما يتم على أيدي الخبراء بطريقة تنضوي على الكثير من العمل المضني والمتخصص، وعليه فقد تم اعتبار خوارزمية البحث عن المعمارية العصبية الأفضل (**Neural Architecture Search**) (ويرمز لها اختصاراً بـ **NAS**) على أنها الوسيلة الأمثل للتقليل من هذه الصعوبات من خلال البحث، وبشكل تلقائي عن المعمارية الأفضل لنموذج الشبكة العصبية فيما لو أردنا مقارنتها مع التصاميم اليدوية.

ولكن، وبما أنّ التطور في هذا المجال يسير بسرعة فائقة، سواءً فيما يتعلق بالمجال البحثي أو في الضجة المثارة حوله، فقد يكون من الصعب الحصول على إجابات عن مجموعة من الأسئلة الجوهرية:
ما هي بالضبط تقنية NAS؟
وهل تختلف اختلافاً جوهرياً عن تقنية التعلم الآلي المؤتمت ("**AutoML**" **Automated Machine Learning**) أو تقنية تحسين اختيار البارمترات الأساسية ("**HO**" **hypermeter optimization**)؟
وهل أساليب NAS المتخصصة تعمل بالفعل؟
ألا تعتبر هذه الاساليب المتخصصة مكلفة (من ناحية الحوسبة)؟
وهل من الأفضل استخدام هذه الأساليب المتخصصة؟

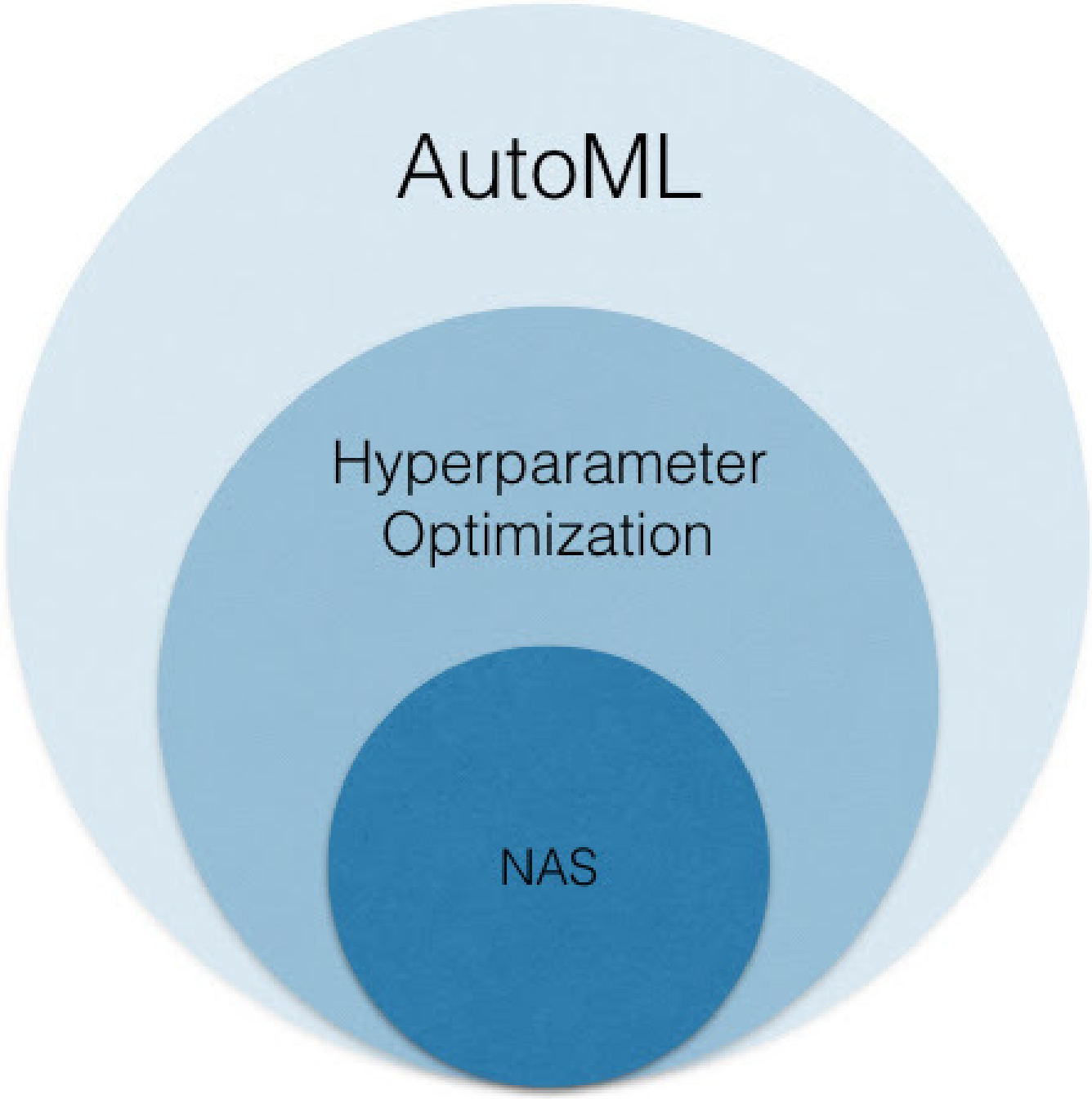
سنحاول في هذه المقالة الإجابة عن هذه الأسئلة من خلال إلقاء الضوء على عدد من النقاط المهمة:

● يوجد انقسام غير مُبرّر حول عمل كل من خوارزمية **NAS** وتقنية **HO**، في الواقع، فإنّ خوارزمية **NAS** هي جزءٌ من النوع الثاني. فضلاً عن ذلك، فإن أساليب **NAS** المتخصصة في حقيقتها ليست مؤتمتة بشكل كامل، لأنها تعتمد على البنى التي يقدمها البشر باعتبارها نقطة الإنطلاق.

● في الوقت الذي يُعتبر فيه اكتشاف وتوليف البنى المختلفة للشبكات العصبية خطوةً مهمة جداً نحو تطوير تطبيقات كفؤة جداً للتعلم العميق، فإنّ أساليب **NAS** المتخصصة ليست جاهزة بعد بشكل كامل، فهي تتسبب ببعض المشاكل الخوارزمية والحوسبية المعقدة مقارنةً مع خوارزميات **HO** عالية الجودة (ومنها **ASHA**) ومن غير أن تُظهر أداءً جيداً باستخدام الطرق الأساسية لتقييم الأداء.

● مع ذلك، فإنّ أساليب **NAS** المتخصصة قد أعطت بالفعل نتائج متميزة في السنوات القليلة الماضية من حيث دقة النتائج، وقلة تكاليف الحوسبة، وحجم البنية المعمارية، واستطاعت في النهاية التفوق على الأداء البشري فيما يتعلق بتصميم العمارات العصبية.

من أجل تمهيد الطريق، لِنقم أولاً بمناقشة كيف أن خوارزمية **NAS** تتوائم داخل المظلة الأكبر المتثملة بتقنية **AutoML**.



خوارزمية

NAS هي جزء من تقنية **Hypermeter Optimization** التي بدورها هي جزء من تقنية **AutoML**.

تركز تقنية AutoML في عملها على أتمتة جميع نواحي سير عمل منظومة التعلّم الآلي (**Machine Learning**) (يرمز له اختصاراً بـ **ML**) من أجل زيادة كفاءة المنظومة وإضفاء الطابع الديمقراطي على تعلم الآلة وبما يمكّن غير المختصين من تطبيق تقنيات التعلم الآلي في مجالات عملهم ودون عناء.
فبينما تعمل ال **AutoML** على أتمتة عدد كبير من العمليات المرتبطة بما يعرف بـ **ETL** (ويقصد بها استخراج وتحويل وتحميل البيانات)، وتدريب وتطوير النماذج، فإنّ المشكلة التي تُعنى بها تقنية **HO** تعتبر هي محور اهتمام تقنية **AutoML**، تتعلق هذه المشكلة بكيفية تهيئة الإعدادت الداخلية التي تتحكم بسلوك نموذج (أو خوارزمية) التعلم الآلي من أجل الوصول إلى نموذج تنبؤي (**predictive**

model) عالي الجودة.

فعلى سبيل المثال، يتطلب استخدام نموذج تحليل الانحدار (**regression model**) تحديد قيمة مصطلح التسوية (**regularization term**)، ويتطلب استخدام نماذج الغابات العشوائية (**random forest**) تحديد الحد الأقصى لعمق الشجرة والحد الأدنى لعدد العينات لكل ورقة، بينما يتطلب تدريب أي نموذج انحدار متدرج عشوائي (**stochastic gradient descent**) تهيئة حجم خطوة مناسب.
في المقابل، فإنّ التعامل مع الشبكات العصبية يتطلب أيضاً تهيئة قيم العديد من البارمترات الأساسية، وهذا يشمل: (1) اختيار طريقة تحسين مناسبة جنباً إلى جنب مع مجموعة البارامترات الأساسية ذات الصلة، و(2) تحديد قيمة التوقف وغيرها من بارامترات التسوية الأساسية، وإذا لزم الأمر (3) توليف البارمترات التي تتحكم في بنية الشبكة (ومنها، عدد الطبقات المخفية، وعدد المرشحات الملتفّة).

قد يوحي الكلام بخصوص تقنية **NAS** بأنها مشكلة جديدة تماماً، إلا أنّ المثال الأخير في أعلاه يدل على وجود علاقة وثيقة بين تقنية **HO** وخوارزمية **NAS**، فبينما مساحة البحث المستخدم من قبل **NAS** بشكل عام هو أكبر ولديه القدرة على السيطرة على نواحٍ مختلفة من معمارية الشبكة العصبية، إلا أن المشكلة الأساسية هي نفسها المشكلة التي تتناولها تقنية **HO** ألا وهي محاولة ضبط الإعدادات داخل مساحة البحث بشكل يضمن تقديم أفضل أداء من أجل تحقيق الهدف المنشود، ومن هنا ننظر إلى المشاكل التي تتناولها خوارزمية **NAS** على أنها مشاكل فرعية تقع ضمن عمل تقنية **HO**.

مع هذا، فإن خوارزمية **NAS** هي بلا شك اتجاه بحثي مثير وجدير بالدراسة، إذ إنه يركز على مشكلة فرعية متخصصة وبما يعطي الفرصة لاستخدام بنية إضافية لوضع حلول مناسبة، كما هو الحال مع العديد من أساليب **NAS** المتخصصة.

في الجزء التالي من المقالة، سنقدم نظرة عامة عن خوارزمية **NAS** ونتعمق أكثر في أوجه التشابه والاختلاف بين طريقة عملها وطريقة عمل تقنية **HO**.

نظرة عامة على تقنية **NAS**

الأجزاء التي تتألف منها خوارزمية NAS

ازداد الاهتمام بهذه التقنية بعد أن استخدم مجموعة من الباحثين (زوف **Zoph** وآخرون) طريقة التعليم المعزز (**reinforcement learning**) من أجل إعداد تصاميم حديثة كان الهدف منها في ذلك الوقت التعرف على الصور والنمذجة اللغوية، إلا أن هؤلاء الباحثين، وكما هو الحال مع المناهج الأولى المتخصصة من **NAS**، كانوا بحاجة إلى استخدام كمية هائلة من الطاقة الحسابية (على سبيل المثال، احتاجوا إلى استخدام المئات من وحدات معالجة الرسومات (**GPUs**) وبحاجة إلى العمل للآلاف من أيام وحدة معالجة الرسومات) مما يجعلها طريقة غير متاحة للتطبيق إلا من قبل شركات كبيرة مثل شركة غوغل **Google**.
أما الأساليب الأحدث، فإنها تعتمد على استخدام طرق متعددة من إعادة الاستخدام لتقليل التكلفة الحاسوبية بشكل جذري، بينما في المقابل يتم تطوير طرق حديثة من قبل المجتمع البحثي لمعالجة هذه المشكلة.

في الجزء التالي، سوف نتعمق قليلاً في قرارات التصميم الأساسية المرتبطة بأساليب **NAS** المتخصصة.

المكونات الثلاثة الرئيسية هي:

1) مساحة البحث: يتعلق هذا المكون بمجموعة أبنية الشبكات العصبية المُتوقَّع النظر فيها، يتم تصميم مساحات البحث هذه لتناسب تطبيقات معينة – على سبيل المثال: يوجد مساحة بحث خاصة بالشبكات الملتفة (**convolutional networks**) لتحقيق مهام تتعلق بالرؤية الحاسوبية، كما يوجد مساحة بحث خاصة بالشبكات العصبية المتكررة (**recurrent networks**) لتحقيق مهام تتعلق بالنمذجة اللغوية.
وعليه، فإنّ أساليب **NAS** المتخصصة ليست مؤتمتة بشكل كامل، حيث يعتمد تصميم تلك المساحات البحثية بشكل أساس على البنى المصممة من قبل البشر كنقطة انطلاق، ومع هذا لا يزال هناك العديد من القرارات المُنتَظَرَة بخصوص المعمارية، في الحقيقة، غالباً ما يتجاوز عدد الأبنية المحتملة المتعلقة بتلك المساحات البحثية 10 مرفوع للأس 10.

2) طريقة التحسين: هذا المكون مسؤول عن تحديد آلية اكتشاف مساحة البحث من أجل العثور على البنية المناسبة، بينما طريقة البحث الأساسية هنا هي البحث العشوائي، يتم في المقابل تطوير أساليب تكيفية مختلفة ـ مثل التعليم المعزز، البحث التطوري (**evolutionary search**)، التحسين القائم على التدرج (**gradient-based optimization**)، والتحسين البايزي (**Bayesian optimization**).
في حين تختلف هذه الطرق التكيفية المختلفة من حيث طريقة تحديد أي البنى يتم اختيارها لغرض التقييم، فإن جميعها يسعى من أجل تحييز البحث نحو البنى التي من المرجح أن يكون أداءها جيداً، زمن غير المستغرب أن يكون لجميع هذه الطرق ما يقابلها من مهمات يكون المسؤول عن تنفيذها تقنية **HO**.

3) طريقة التقييم: هذا المكون هو المسؤول عن تقييم كفاءة كل بنية يتم أخذها بعين الاعتبار من خلال طريقة التحسين، يعتبر التدريب الكامل للشبكة العصبية هو الخيار الأبسط لكنه الأكثر تكلفة من ناحية الحوسبة، الطريقة البديلة للتدريب الكامل هو أن يتم استخدام التدريب الجزئي للشبكة العصبية، وهي طريقة مشابهة لأسلوب "التوقف في وقت مبكر" التي تستخدمها بعض تقنيات **HO** مثل طريقة **ASHA**.
تم أيضاً تصميم بعض طرق التقييم ذات العلاقة بعمل خوارزمية **NAS** مثل تغيير شكل الشبكة، مشاركة الأوزان، والشبكات الفائقة بهدف استخدام هيكلية الشبكات العصبية من أجل إيجاد قيم جودة أرخص.
عادةً ما تكون الأساليب المختلفة للتدريب الجزئي أقل كلفة بكثير من التدريب الكامل، في حين أن أساليب التقييم الخاصة بخوارزمية **NAS** هي أقل كلفة بمرتين أو ثلاث مرات من التدريب الكامل.

الجدير بالذكر هنا أنّ هذه المكونات الثلاثة هي نفس المتطلبات الخاصة بأساليب تحسين اختيار البارمترات الأساسية (**hypermeter optimization**) المعروفة، لقد اعتمد المجتمع العلمي، ومن أجل تقييم أداء طرق البحث المختلفة، على مجموعة من البيانات والطرق الأساسية.
سنقوم بناء عليه فيما يلي باستخدام هذه المعايير لعمل مقارنة بين:
1) البنى التي تم تصميمها من قبل البشر والتي تم تحسينها باستخدام أي من الطرق المختلفة لتقنية **HO**،
2) البنى المعتمدة على خوارزمية **NAS** والتي يمكن التعرف عليها عبر الأساليب المتخصصة التابعة لهذه التقنية.
(تركز خوارزمية **NAS** على مشكلة التعرفة على البنى المختلفة، ومع هذا فهي بحاجة إلى خطوة ثانوية من أجل تحسين الخصائص الأساسية غير المتعلقة بالبنية التي تحددها، تُظهِر النتائج التي حصلنا عليها القيم الخاصة بخطأ الاختبار بعد إتمام كلتا الخطوتين).

النماذج المصممة بخوارزمية **NAS** في مقابل النماذج المصممة يدوياً

الطريقتان الأكثر شيوعاً والتي تستخدم من أجل تقييم أداء الطرق المتعددة لتقنية **NAS** هي: (1) تصميم البنى باستخدام الشبكات العصبية الملتفة **CNN** وتقييمها باستخدام مجموعة بيانات **CIFAR-10**، و(2) تصميم البنى باستخدام الشبكات العصبية الراجعة **RNN**

وتقييمها باستخدام مجموعة بيانات ("**PennTree Bank** "**PTB**).

الجدول أدناه يعرض قيم خطأ الاختبار للمعماريات المختلفة باستخدام مجموعة بيانات **CIFAR-10**:

جدول رقم 1: خطأ الاختبار باستخدام مجموعة بيانات **CIFAR-10** والبنى المعمارية الرئيسية سواء ذات التصميم البشري أو التي تم تصميمها باستخدام أحد أساليب **NAS** المتعددة وبمختلف طرق البحث والتقييم، لاحظ أن كل البنى المعمارية قد تم تحسينها عبر طرق تحسين اختيار البارمترات الأولية.

فيما يتعلق بتقييم الأداء باستخدام مجموعة بيانات **CIFAR-10**، نلاحظ أنّ الأداء كان أفضل بالنسبة للطرق التي تم تصميمها باستخدام أساليب **NAS** المتخصصة والتدريب الكامل بالمقارنة مع الطرق التي تم تصميمها بشكل يدوي، ومع ذلك فهي مكلفة من ناحية الحوسبة وتحتاج إلى 1000 يوم عمل من أيام وحدة معالجة الرسوميات.
وعلى الرغم من أن الطرق التي تستخدم التدريب الجزئي أو أي من طرق التقييم الخاصة بتقنية **NAS** لها كلفة أقل من ناحية الحوسبة (401 يوم من أيام وحدة معالجة الرسوميات، على التوالي)، لكن تم التفوق عليها بواسطة الطرق المُصممة يدوياً، وكما يظهر في الجدول رقم 1 بشكل خاص، فإن البنى المعمارية المختلفة لتقنية **NAS** استخدمت بارمترات أقل بكثير من تلك التي استخدمتها التصاميم البشرية، مما يفتح الباب للتطبيقات التي تتطلب ذاكرة أكبر وسرعة استجابة أعلى.

جدول رقم 2 يظهر اضطراب الاختبار ولمختلف البنى المعمارية وباستخدام مجموعة بيانات **PTB**.

جدول رقم 2: اضطراب الاختبار باستخدام مجموعة بيانات **PTB** والبنى المعمارية الرئيسية سواء ذات التصميم البشري أو التي تم تصميمها باستخدام أحد أساليب **NAS** المتعددة وبمختلف طرق البحث والتقييم، لاحظ أن كل البنى المعمارية قد تم تحسينها عبر طرق تحسين اختيار البارمترات الأولية.

أما أساليب **NAS** المتخصصة فلم تكن نتائجها بالمستوى المطلوب بالمقارنة مع البنى المعمارية المعمولة يدوياً، بناء على هذا المعيار، فإنه من المدهش أن تقدّم أساليب التقييم الرخيصة (فيما يتصل بالجانب الحسابي) أداءً افضل بالمقارنة مع الطرق التي اعتمدت التدريب الكامل، فمن المحتمل أن يكون السبب وراء هذا راجع إلى التحسينات التي لحقت تدريب مجموعة بيانات **LSTMs** منذ أن نشر زوف وزملاءه بحثهم في العام **2016**.

هل أصبحت أساليب **NAS** المتخصصة جاهزة للتبني على نطاق واسع؟

ليس بعد! لكي نكون واضحين، فإن استكشاف البنى المعمارية المختلفة وإجراء تحسينات كبيرة على عملية اختيار البارمترات الأساسية يبقيان عنصرين هامين في حياة أي منظومة تعلم عميق. مع هذا، وفي ضوء النتائج البحثية التي تم التوصل إليها حتى الآن (كما هو موضح أعلاه) فإننا نعتقد أنه في الوقت الذي تُظهِر فيه أساليب **NAS** المتخصصة نتائج واعدة باستخدام المعيارين المذكورين أعلاه، فإنها ليست جاهزة بعد للعمل الجاد وذلك للأسباب التالية:

1) بما أن البنى المعمارية المصممة يدوياً والتي تم توليفها بشكل جيد قادرة على المنافسة مع أساليب **NAS** القابلة للحوسبة فيما لو استخدمنا مجموعة بيانات **CIFAR-10**، كما أنها تتفوق على أساليب **NAS** المتخصصة فيما لو استخدمنا مجموعة بيانات **PTB**، فنحن

نعتقد أنه تم إنفاق الموارد بشكل أفضل نحو تحسين اختيار البارمترات الأولية الخاصة بالمعماريات المصممة يدوياً.

2) معظم أساليب **NAS** المتخصصة هي مختصة إلى حدٍ ما بمساحة بحثية معينة، كما أنها بحاجة إلى إعادة تدريب أو إعادة تجهيز لكل مساحة بحث جديد، بالإضافة إلى ذلك تعاني بعض تلك الأساليب من مشكلة قوة التحمّل (**robustness**) حيث يصبح من الصعب القيام بتدريبها.

تعيق هذه المشاكل في الوقت الحالي تطبيق أساليب **NAS** المتخصصة الموجودة حالياً من أجل تنفيذ المهام المختلفة.

- التاريخ: 2019-04-18
- التصنيف: تكنولوجيا

#الشبكات العصبية #التعلم الآلي #خوارزمية #المعمارية العصبية #التعليم العميق

المصطلحات

- **تعليم الآلة (machine learning)**: تعلم الآلة هو أحد أنواع الذكاء الاصطناعي، يمكّن التطبيقات البرمجية من التنبؤ بنتائج أكثر دقة دون برمجتها بشكل صريح. ويتم ذلك عن طريق بناء خوارزميات تتلقى بيانات الإدخال وتستخدم التحليل الإحصائي للتنبؤ بقيمة المخرجات ضمن نطاق مقبول.
- **التعلم العميق (deep learning)**: هو أحد ميزات الذكاء الاصطناعي التي تَعنى بمحاكاة نهج التعلم الذي يستخدمه البشر للحصول على أنواع معينة من المعرفة، كما يمكن اعتباره وسيلة لأتمتة التحليلات التنبؤية.
- **الأمثلة (optimization)**: هي اختيار العنصر الأفضل، بالنسبة لمعيارٍ معين، من مجموعة من البدائل المحتملة.

المصادر

- O'REILLY
- الصورة

المساهمون

- ترجمة
  - محمد زهير الطائي
- مُراجعة
  - فرح درويش
- تحرير
  - أحمد كنينة
- تصميم

- حسن ديب

- نشر
  - آلان حسن