شبكةٌ عصبيّةٌ تُحاكي نبرة الصوت البشريّة

باختصار

في العام الماضي، نشرت ديبمايند **DeepMind** تفاصيلًا حول شبكة ويفنت **WaveNet**، وهي شبكةٌ عصبونيّةٌ عميقةٌ قادرةٌ على إنتاج خطابٍ اصطناعيٍّ مشابهٍ للحقيقيّ. الآن، نُقِّح النظام بما فيه الكفاية ليعمل مع تطبيق جوجل أسستنت **Google Assistant** في كلّ المنصات.

دروسٌ في الإلقاء

في السنة الماضية، نشرت شركة ديبمايند معلوماتٍ عن ويفنت، وهو عبارةٌ عن شبكةٍ عصبونيّةٍ عميقةٍ، تُستخدَم لتصنيع محادثاتٍ كلاميّةٍ بشريّةٍ، وقد طُرحت نسخةٌ محسّنةٌ من هذه التقنية لاستخدامها بجانب تطبيق جوجل أسستنت.

نظام توليف الكلام، المعروف أيضًا باسم "نصٌ إلى صوتٍ" اختصارًا (**TTS**)، عادةً ما يَستخدم واحدةً من اثنتين من هذه التقنيات.

يشمل **TTS** تجميع مقاطعَ من تسجيلاتٍ صوتيّةٍ من ممثلٍ صوتيٍّ، والعيب في هذا أن عليك تبديل المكتبات الصوتية في حال حدثت تغييرات أو ترقية.

تقنيةٌ أخرى، وهي **TTS** المعتمدة على المعاملات **parametric TTS**، تستخدم مجموعةً من المعاملات لإنتاج خطابٍ حاسوبيٍّ، إلا أن هذا الكلام يبدو في بعض الأحيان غير طبيعيٍّ أو روبوتيٍّ.

من ناحيةٍ أخرى، تُنتج ويفنيت أشكالًا موجيّةً من نقطة الصفر معتمدةً على نظامٍ مطوّرٍ باستخدام شبكةٍ عصبونيةٍ التفافيّةٍ **convolutional neural network**.

في البداية، استُخدِم عددٌ كبيرٌ من عيّنات الكلام لتدريب المنصة على توليف الأصوات، مع الأخذ بعين الاعتبار أيّ الأشكال الموجيّة تبدو واقعية وأيّها لا. فأعطى هذا مولّدَ الكلام المقدرةَ على توليّد نغمة صوتٍ طبيعيّةٍ، وتفاصيلَ أخرى أيضًا كحركة الشفاه.

اعتمادًا على العينات المدخلة للنظام، فإنه ستُطوَّر "لكنةٌ" فريدةٌ من نوعها، مما يعني إمكانية استخدامها لإنشاء أيّ عددٍ من الأصوات المتميزة إذا أُدخل لها مجموعةٌ مختلفةٌ من البيانات.

لسانٌ فصيح

في الحقيقة، كان العائق الأكبر لتطور ويفنت أنها تطلّبت في البداية قدرة معالجةٍ كبيرةٍ، كما أنها كانت عمليّةً بطيئة، حيث كانت تستغرق ثانيةً واحدةً لتوليد 0.02 ثانية من الصوت.

بعد تحسين النظام على مدى الاثني عشر شهرًا المنصرمة، قام مهندسو شركة ديب مايند بتحسين شبكة ويفنت إلى درجةٍ بإمكانها إنتاج موجةٍ صوتيّةٍ تدوم ثانيةً واحدةً في 50 ميلي ثانية فقط؛ أي أسرع ب 1,000 مرّةٍ من الأصل.

وأكثر من ذلك، زادت جودة العينة من 8 بت إلى 16 بت، وهذا ما أدى لحصوله على درجةٍ عاليةٍ في الاختبار الذي أجراه مستمعون بشريون.

وتعني هذه التحسينات إمكانية دمج النظام في المنتجات الاستهلاكية، مثل مساعد غوغل.

حقوق الصورة: Deepmind

تُستخدم ويفنت الآن لتوليد الأصوات الأمريكية واليابانية لمساعد جوجل عبر جميع المنصات. ونظرًا لقدرة النظام على خلْق أصواتٍ متخصّصة استنادًا إلى كلِّ العينات المُدخَلة، ينبغي أن تكون جوجل قادرةً على استخدام ويفنت لتركيب محادثةٍ بشريّة واقعيّةٍ للغاتٍ ولهجاتٍ أخرى.

أصبحت واجهات الصوت **Voice interfaces** واسعةَ الانتشار في جميع أشكال الحوسبة، ولكن الطبيعة المتكلّفة لبعض الكلام الصناعيّ قد جعلت بعض الأشخاص يتوقفون عن استخدامها. إلا أن جهود ديب مايند لتحسين هذه التقنية يمكن أن تؤدي إلى اعتمادها في كثيرٍ من المجالات، لذا ستعمل بالتأكيد على تحسين التجربة الحالية.

- التاريخ: 2017-11-27
- التصنيف: تكنولوجيا

#الصوت #جوجل #توليف الكلام

المصادر

- futurism

المساهمون

- ترجمة
  - رؤى سلامة
- مُراجعة
  - علي مرعي
- تحرير
  - حسن شوفان
  - رأفت فياض
- تصميم
  - أحمد أزميزم
- نشر
  - ريم فاخر