كيف تترجم سماعات غوغل اللاسلكية "بيكسل بادز" اللغات؟



حقوق الصورة: شاترستوك Shutterstock

في عام 1978 ظهر أحد أعمال دوغلاس آدمز **Douglas Adams** بعنوان "دليل المسافر إلى المجرة" **Hitchhiker's Guide to The Galaxy**، الذي بدأ كمسلسل إذاعي في قناة **BBC** ثم تحول إلى سلسلة روايات ثم إلى فيلم سينمائي والآن هو أيقونة ثقافية؛ حيث إنها تحتوي تنبؤات عديدة بالكثير من التقنيات الحديثة ومن بينها السمكة المتكلمة **Babel Fish** وهي كائن أصفر صغير يدخل إلى الأذن البشرية ويتغذى على طاقة الدماغ وكان قادراً على الترجمة من وإلى أي لغة.

والآن، يبدو أن شركة غوغل **Google** العملاقة طورت نسختها الخاصة من السمكة المتكلمة، والتي تعرف باسم بيكسل بادز **Pixel**

Buds، وتستفيد سماعات الأذن اللاسلكية هذه في عملها من مساعد غوغل الرقمي **Google Assistant**، وهو تطبيق ذكي يفهم أوامر المستخدم ويستطيع التحدث وكذلك تقديم المساعدة له، وأهم إمكانياته هي دعم مترجم غوغل **Google Translate** الذي يعرف عنه بأنه يترجم 40 لغة مختلفة، وتكلفة هذه التقنية المثيرة الاهتمام هي أقل من 200$.

إذاً كيف تعمل هذه التقنية؟

تشمل ترجمة الكلام آنياً سلسلة من التقنيات المختلفة التي تطورت بشكل كبير وسريع خلال السنوات القليلة الماضية. تلك السلسلة، من بداية المدخلات وحتى المخرجات، هي كالتالي:

تجهيز الدخل **Input conditioning**: تستقبل السماعات التشويش والضجيج في الخلفية، حيث إنها تسجل بفاعلية مزيجاً من صوت المستخدم وأصوات أخرى، مرحلة إزالة الضجيج تزيل الأصوات غير المرغوب بها في الخلفية بينما يستخدم كاشف الصوت (**Voice activity detector (VAD**) لتشغيل النظام فقط عند تكلم الشخص الصحيح (أي إنه لا يُشغل النظام استجابةً لصوت شخص يقف خلف المستخدم ويقول بصوت عال: "أوكي غوغل"*)، وتستخدم حساسات اللمس من أجل زيادة دقة عمل **VAD**.

التعرف على اللغة **Language identification (LID)**: يستخدم هذا النظام تعلم الآلة **machine learning** لتحديد ماهية اللغة المتكلمة في غضون ثوان، وهذا مهم لأن جميع المراحل التالية مرتبطة باللغة، ومن أجل التعرف على اللغة فإن السمات اللفظية غير كافية وحدها للتمييز بين اللغات، فمثلاً بعض اللغات كالأوكرانية والروسية، أو الأردو والهندية هي فعلياً متطابقة في الوحدات الصوتية أو ما يسمى بالمقاطع اللفظية **phonemes**، لذلك كان من الأفضل تطوير تمثيلات صوتية مختلفة جديدة.

التعرف الآلي على الكلام **ASR**: **Automatic speech recognition (ASR)** يستخدم نموذجاً صوتياً لتحويل الكلام المسجل إلى سلسلة من المقاطع اللفظية ثم تُستخدم نمذجة اللغة لتحويل هذه المعلومات اللفظية إلى كلمات، يستخدم هذا النظام قواعد نحو الكلام بالإضافة إلى السياق والاحتمالات وكذلك قاموس النطق، ومن خلال ذلك يستطيع ملء فجوات المعلومات المفقودة وتصحيح المقاطع الصوتية التي تم التعرف عليها بشكل خاطئ من أجل استنتاج تمثيل نصي لما قاله المستخدم.

معالجة اللغة الطبيعية **Natural language processing NLP**: ينجز الترجمة الآلية من لغة إلى أخرى، لايقتصر هذا الأمر ببساطة على استبدال أسماء أو أفعال من لغة بمثيلاتها في لغة أخرى، وإنما يتضمن فك رموز معنى الكلام المدخل، ومن ثم إعادة تركيب ذلك المعنى لنحصل على كلام المخرج بلغة مختلفة، متضمناً كل الفوارق البسيطة والالتباسات التي تُصعّب علينا تعلم اللغات الأخرى.

توليف الكلام أو تحويل النص إلى كلام **Speech synthesis or text-to-speech (TTS)**: وهو تقريباً عكس التعرف الآلي إلى الصوت، حيث يتضمن ذلك تشكيل كلام يبدو طبيعياً للسامع من سلسلة من الكلمات (أو المعلومات الصوتية). استخدمت الأنظمة القديمة التوليف الجمعي الذي يعني فعلياً دمج العديد من التسجيلات القصيرة لشخص يتكلم مقاطع صوتية مختلفة في تسلسل صحيح. أما الأنظمة الأحدث فهي تستخدم نماذج الكلام الإحصائية المعقدة **complex statistical speech models** من أجل إعادة تشكيل صوت محكي يبدو طبيعياً للسامع.

وبعد أن تعرفنا على التقنيات الخمسة في هذه السلسلة، لنرى كيف سيُطبق هذا النظام عملياً في الترجمة بين اللغات، كاللغتين الصينية والإنكليزية مثلاً.

بمجرد جهوزيتها للترجمة، فإن السماعات اللاسلكية تسجل العبارة المنطوقة حيث تستخدم **VAD** لتحديد بداية و نهاية الكلام، يمكن إزالة ضجيج الخلفية جزئياً عن طريق السماعات نفسها، أو عند نقل التسجيل إلى الهاتف الذكي عن طريق البلوتوث، وبعد ذلك يُضغط التسجيل ليصبح كمية أصغر بكثير من البيانات، ويحول بعدها عبر **WiFi** أو **3G** أو **4G** إلى مخدمات الكلام التابعة لغوغل.

مخدمات غوغل، التي تعمل كسحابة **Cloud**، تستقبل التسجيل وتفك الضغط ثم تحدد باستخدام تقنية التعرف على اللغة **LID** ما إذا كان الكلام محكياً باللغة الصينية أو الإنكليزية.

يمرر بعدها التسجيل إلى مرحلة التعرف الآلي على الكلام **ASR** الخاص باللغة الصينية، ثم إلى مرحلة معالجة اللغات الطبيعية **NLP** حيث يركب المترجم الآلي مخططاً للكلمات من الصينية إلى الإنكليزية، وأخيراً فإن خرج هذه المرحلة يرسل إلى برمجيات تحويل النص إلى كلام **TTS** باللغة الإنكليزية، وينتج تسجيل مضغوط يعاد إرساله عكسياً ليشغل بواسطة السماعات.

قد تبدو تلك كمراحل كثيرة جداً من الاتصال، ولكن العملية في الواقع تحدث فقط في غضون بضع ثوان، وهذا شيء مهم للعمل لأن المعالج الخاص بالسماعات ليس فعالاً بالقدر الكافي للقيام بعمليات الترجمة بنفسه، وأيضاً حجم الذاكرة لا يكفي ليخزن نماذج اللغات والنماذج الصوتية، حتى إن تضمنت السماعات معالجاً فعالاً وذاكرة كافية فإن عمليات المعالجة الحاسوبية المعقدة سوف تستنزف كامل البطارية الخاصة بالسماعات خلال ثوان معدودة.

وعلاوة على ذلك، فإن الشركات التي تمتلك مثل هذه المنتجات (مثل غوغل، آي فلاي تيك **iFlytek** وآي بي إم **IBM**) تعتمد على

التحسينات المستمرة لتصحيح وتعديل وتطوير نماذج ترجمتها، حيث يعد تعديل النموذج أكثر سهولة في المخدمات السحابية الخاصة بتلك الشركات، بينما يكون ذلك صعباً في حال وجوده ضمن السماعات.

من المؤكد أن الراحل دوغلاس آدامز كان سيعتقد أن التكنولوجيا المستخدمة في آلات الترجمة الواقعية هذه مذهلة، وهي كذلك بالفعل. ولكن علماء ومهندسي الحاسب لن يتوقفوا عند هذه النقطة، يمكن أن الخطوة التالية في تكنولوجيا الحوسبة المفعلة بالصوت مستوحاة من جهاز خيالي آخر، مثل الحاسوب الذكي في فيلم الرجل الحديدي **Iron Man** المسمى جارفيس **J.A.R.V.I.S** (**Just Another Rather Very Intelligent System** – مجرد نظام ذكي جداً آخر) في سلسلة أفلام مارفل **Marvel**.

هذا النظام سيتخطى حدود الترجمة وسيستطيع التحدث معنا ومعرفة ماذا نشعر وماذا نفكر وتوقع احتياجاتنا.

* أوكي غوغل **OK Google** هو الأمر الصوتي المستخدم لتفعيل خدمات غوغل التي تعمل بالصوت.

- التاريخ: 2018-02-17
- التصنيف: تكنولوجيا

#اللاسلكية #سماعات غوغل #الترجمة #بيكسل بادز

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.
- **المجرة (galaxy)**: عبارة عن أحد مكونات كوننا. تتكون المجرة من الغاز وعدد كبير (في العادة، أكثر من مليون) من النجوم التي ترتبط مع بعضها البعض، بوساطة قوة الجاذبية. و عندما تبدأ الكلمة بحرف كبير، تُشير Galaxy إلى مجرتنا درب التبانة.
المصدر: ناسا

المصادر

- techxplore

المساهمون

- ترجمة
  - لايا البشلاوي
- مُراجعة
  - فرح درويش
- تحرير
  - ليلاس قزيز
  - محمد شوبك
- تصميم
  - رنيم ديب

- صوت
  - ابتسام الخيال
- نشر
  - ريم فاخر