تقدم هذه الورقة البحثية NeuSpeech، وهو إطار جديد لترجمة إشارات المخ (MEG) إلى نص مكتوب. وهي تستخدم نموذج شيسبر القوي للتعرف على الكلام واستخدام الترميز العكسي لترجمة الإشارات العصبية إلى كلام مباشرة. يتميز هذا النهج بما يلي:

ميزات:

1. أول طريقة تستخدم الانتباه المتقاطع لترجمة إشارات MEG إلى نص دون الاعتماد على التعليم المسبق أو التدريب الإجباري.

2. تحقيق أداء مرتفع جدًا على مجموعتي بيانات MEG مع BLEU-1 يصل إلى 60.30٪ و 52.89٪ دون تعليم إجباري.

3. يمكن تطبيقها على لغات وأنظمة مختلفة ويمكن تدريبها بشكل مشترك على مجموعات بيانات متعددة.

4. تقدم دراسة شاملة لتأثير التدريب المسبق، وتقسيم البيانات، وزيادة البيانات، وقانون التدرج على الأداء.

سلبيات:

1. لا تزال هناك بعض القيود على قدرة النموذج على توليد جمل جديدة صحيحة تمامًا بسبب قلة البيانات المتاحة.

2. قد تثير مسألة ترجمة أفكار الناس إلى نص مكتوب مخاوف أخلاقية وقانونية إذا أُسيء استخدامها.

الخوارزمية المستخدمة:

يستخدم NeuSpeech نموذج شيسبر للتعرف على الكلام، ويعدل طبقة التراكب الأولى لتتناسب مع إشارات MEG متعددة القنوات. يتم ترميز إشارات MEG الخام مباشرة وإرسالها إلى النموذج الذي يستخدم الانتباه المتقاطع لترجمتها إلى نص بدون تعليم إجباري.

الجديد الذي قدمته:

1. أول محاولة لترجمة إشارات MEG الخام إلى نص مفتوح المفردات بشكل مباشر دون تدريب مسبق أو تعليم إجباري.

2. دراسة شاملة لتأثير العوامل المختلفة مثل التدريب المسبق وزيادة البيانات والتدريب المشترك وحجم النموذج على الأداء.

3. إمكانية التطبيق على لغات وأنظمة مختلفة باستخدام تدريب مشترك للنماذج.

لذا تقدم هذه الورقة نهجًا واعدًا لترجمة إشارات المخ إلى نص بشكل مباشر، مع دراسة شاملة للعوامل المؤثرة على الأداء، لكنها تحتاج إلى مزيد من البيانات والاعتبارات الأخلاقية لتحسين الدقة والتطبيق العملي.

يعتمد NeuSpeech على نموذج شيسبر القوي للتعرف على الكلام، حيث يتم تعديل طبقة التراكب الأولى لاستيعاب إشارات MEG متعددة القنوات. تتلخص الخوارزمية المستخدمة فيما يلي:

1. معالجة البيانات المسبقة:

- تمرير إشارات MEG الخام عبر مرشح ترددي بنطاق (1-60) هرتز لإزالة الضوضاء.

- إعادة أخذ عينات من الإشارات إلى 200 هرتز.

- تنظيف الإشارات باستخدام robust scaler، ثم قصها إلى ±10 وتغيير النطاق إلى (-1, 1).

2. تعديل نموذج شيسبر:

- استبدال طبقة التراكب الأولى بطبقتين للتراكب ذات أبعاد مناسبة لإشارات MEG.

- الطبقة الأولى بحجم نواة 3 للزيادة في أبعاد القنوات.

- الطبقة الثانية بحجم نواة 2 للتقليل من عرض النافذة الزمنية.

3. التدريب:

- إرسال إشارات MEG الخام المعالجة إلى الشبكة المعدلة.

- استخدام انتباه متقاطع للحصول على معلومات دلالية عالمية للجملة.

- تدريب الشبكة بشكل موجه باستخدام الأزواج المطابقة من إشارات MEG والنص الهدف.

- استخدام Adalora للحفاظ على كفاءة الذاكرة عند التدريب على نماذج كبيرة.

4. التوليد:

- بعد التدريب، يتم إرسال إشارات MEG الجديدة إلى النموذج.

- استخدام البحث المشعع لتوليد النص المقابل بدون تدريب إجباري.

أما قواعد البيانات المستخدمة، فهي:

1. GWilliams: مجموعة بيانات MEG باللغة الإنجليزية تحتوي على 23339 زوجًا من إشارات MEG والنص المقابل للتدريب، و2917 للتحقق، و2918 للاختبار.

2. Schoffelen: مجموعة بيانات MEG باللغة الهولندية تحتوي على 8570 زوجًا للتدريب، و1112 للتحقق، و1076 للاختبار.

لاحظ أن عدد الجمل الفريدة محدود جدًا في كلتا المجموعتين (661 و360 على التوالي)، مما يشكل تحديًا للنموذج في التعميم.

يتضمن البحث أيضًا دراسات لتأثير التدريب المسبق، وتقسيم البيانات، وزيادة البيانات، وحجم النموذج على الأداء للحصول على رؤى لتحسين النماذج المستقبلية.