

Résumé

La reconnaissance automatique de la parole est une branche de l'intelligence artificielle qui permet la transcription du langage naturel humain exprimé par la parole. Un système de reconnaissance de la parole prend en entrée un signal audio capturé par un microphone pour produire une sortie sous forme de texte qui sera par la suite traité selon le besoin.

Dans le cadre de ce projet, nous avons conçu et implémenté *ASeR-System*, un système de reconnaissance de la parole pour la langue arabe. Pour la conception de ce système, nous avons opté pour une approche End-to-End en utilisant des techniques d'apprentissage profond. Nous avons proposé différentes architectures de modèles à savoir le modèle de base, le modèle à couches convolutionnelles et ce en utilisant deux types d'encodage, l'encodage basé mots et l'encodage basé caractères. L'apprentissage a été effectué sur le corpus QCRI contenant 1200 heures de dialogue des programmes d'Al-Jazeera TV intégrant plus d'une centaine de locuteurs.

Par la suite nous avons développé une application qui intègre la reconnaissance de la parole à un système de questions-réponses. L'application permet à un utilisateur de poser une question oralement, ensuite ASeR-System la transformant en texte, avant d'utiliser un système de questions-réponses pour obtenir la réponse à la question.

Abstract

Automatic speech recognition is a branch of artificial intelligence that allows the transcription of human natural language expressed by speech. A speech recognition system inputs an audio signal captured by a microphone to produce a text output which will subsequently be processed as needed.

As part of this project, we designed and implemented ASeR-System, a speech recognition system for the Arabic language. For the design of this system, we went for an End-to-End approach using deep learning techniques. We have put in place different model architectures namely the baseline model and the convolutional layers based model while using for each, two types of encoding, word-based encoding and character-based encoding. Learning has been done on the QCRI corpus which consists of 1200 hours of Al-Jazeera TV programmes that feature more than a hundred speakers.

Subsequently we developed an application that uses speech recognition with a question-answering system. The application allows a user to ask a question orally, then ASeR-System transforms it into text, before using the question-answering to get the answer to the question.

ملخص

التعرف التلقائي على الكلام هو فرع من الذكاء الاصطناعي الذي يسمح بنسخ اللغة (أي التمثيل النصي للغة) الطبيعية للإنسان التي يعبر عنها بالكلام أو تحت التعبير الصوتي. يقوم نظام التعرف على الكلام بإدخال إشارة صوتية يتم التقاطها بواسطة ميكروفون لإنتاج نص يتم معالجته لاحقاً حسب الحاجة.

كجزء من هذا المشروع ، قمنا بتصميم وتنفيذ ASeR-System ، وهو نظام التعرف على الكلام للغة العربية. لتصميم هذا النظام ، اخترنا أسلوب End-to-End باستخدام تقنيات التعلم العميق. لقد اقترحنا تصميمات مختلفة للنماذج وهي النموذج الأساسي ونموذج الطبقة التلافيفية وهذا باستخدام نوعين من الترميز ، الترميز القائم على الكلمات والترميز القائم على الأحرف. وقد تم تنفيذ التعلم على مدونة QCRI التي تحتوي على 1200 ساعة من الحوارات من برامج قناة الجزيرة التلفزيونية والتي شارك فيها أكثر من مائة متحدث.

بعد ذلك قمنا بتطوير تطبيق يدمج التعرف على الكلام مع نظام الأسئلة والأجوبة. يسمح التطبيق للمستخدم بطرح سؤال شفهيًا ، ثم يقوم ASeR-System بتحويله إلى نص ، قبل استخدام نظام الأسئلة والأجوبة للحصول على إجابة على السؤال.