Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Computer Science

Arabic Speech Text Summarization Using Deep Learning Techniques

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Doctor of Philosophy in
Computer Science**

By

Hiba Adreese Younis Ibraheem

Supervised by

Asst. Prof. Dr. Yusra Faisal Mohammad Abdelrahim

2024 A.D. **1446 A.H.**

ABSTRACT

Information is considered the oil of the 21th century, with analytics serving as the combustion engine. Because of this, the enormous volume of data that is produced every day is all around us and won't be useful to people until new tools and technologies are created to make it accessible.

Summarization can be classified into extractive and abstractive according to the approach used. The extractive approach extracts the most important sentences based on a specific score while keeping the original meaning of the text, while Abstractive approach is the process of paraphrasing information from a source text in a compact and understandable manner, similar to how humans do. Incoherency out-of-context, or readability are two important problems in extractive approach, so the abstractive approach was adopted in this work.

The dissertation aims to cover Abstractive summarization of Arabic connected speech gap through proposed the intelligent cascade Arabic speech to text summarization system (CASTTS). The cascade system consists of two main integrated stages: automatic Arabic speech recognition(AASR) and abstractive Arabic text summarization(AATS). The self supervised learning and transformer based models were adopted for both stages in order to solve dataset lack, long range dependencies in connected speech sequence, and parallelism.

For AASR stage, several sequential steps were implemented, started with dataset preparation, followed by dataset preprocessing steps to make it adequate for deep learning speech models, then feature extraction and classification was implemented using seven proposed models (three modified pretrained speech models (3MPSM), modified HUBERT with learning rate scheduler(MHLR), in addition to three hybrid models).

In 3MPSM, three separate models (wave2vector2 cross lingual speech representation (wav2vec2xlsr), hidden unit bidirectional encoder representation from transformer (HUBERT), and massive multilingual speech (MMS)) were used for recognition Arabic speech after fine-tuning those models on Arabic part of both common voice and FLEURS benchmark datasets, freezing the feature extractor part, and training the rest of the layers, as well as optimizing a number of hyper parameter values.

The MHLR model was implemented as enhancement of HUBERT model by adding additional layers. It used learning rate scheduler of type ("reduce_lr_on_plateau") to prevent the model from being overfitted.

The three hybrid models were proposed by first: hybridization both wav2vec2xlsr and HUBERT model with adapter layer of MMS model and second: hybridization the HUBERT model with four blocks of transformer encoder to construct a bi-encoder with the original encoder model. These models were fine-tuned on FLEURS benchmark dataset using SWATS (switching from adaptive with moment (ADAM) to stochastic gradient descent (SGD)) optimization technique.

After evaluating the AASR models using word error rate (WER) metric, results showed that HUBERT model with transformer encoder blocks outperformed other proposed models in terms of WER as WER is 23.5. It also showed competitive results when compared on massive multilingual speech model (MMS) pretrained on more datasets and languages. The proposed models also outperformed the previous related studies.

Whereas for the AATS stage, three improved transformer based models (mT5, AraBART, mbert2mbert) were implemented by fine-tuning and evaluating them on both in-domain dataset and out-domain datasets using different evaluation metrics. The results showed that the enhanced transformer based models gave promising results and outperformed other models in previous studies. It also showed that mT5 model outperformed other our study models in terms of BERTScore as BERTScore=0.76 for in-domain dataset followed by 0.75 and 0.69 for AraBART, and mbert2mbert models respectively.

The combination of two stages models of intelligent system was accomplished. The results showed that mT5 model outperformed other models in terms of readability and summary quality when evaluated on FLEURS benchmark dataset despite the shortness of the input text.

وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم علوم الحاسوب

تلخيص نص الكلام العربي باستخدام تقنيات التعلم العميق

اطروحة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة دكتوراه فلسفة في
علوم الحاسوب

من قبل

هبه ادريس يونس ابراهيم

بإشراف
أ.م. د. يسرى فيصل محمد عبد الرحيم

١٤٤٦ هـ ٢٠٢٤ م

الخلاصة

تعتبر المعلومات نفط القرن 21، حيث تعمل التحليلات كمحرك احتراق. ولهذا السبب، فإن الحجم الهائل من البيانات التي يتم إنتاجها كل يوم والموجود في كل مكان حولنا لن يكون مفيدا للأشخاص حتى يتم إنشاء أدوات وتقنيات جديدة لجعلها في متناول الجميع.

يمكن تصنيف التلخيص إلى استخراجي وتجريدي وفقا للمنهجية المستخدمة. يستخرج النهج الاستخراجي أهم الجمل بناء على مقياس معين مع الاحتفاظ بالمعنى الأصلي للنص. في حين أن النهج التجريدي هو عملية إعادة صياغة المعلومات من النص المصدر بطريقة مضغوطة ومفهومة ، على غرار ما يفعله البشر. إن عدم التماسك، خارج السياق أو قابلية القراءة مشكلتان مهمتان في النهج الاستخراجي ، لذلك تم اعتماد النهج التجريدي في هذا العمل.

تهدف الأطروحة إلى تغطية الفجوة البحثية للتلخيص التجريدي للكلام المتصل باللغة العربية من خلال اقتراح نظام ذكي تسلسلي لتلخيص الكلام العربي بعد تحويله إلى نص (CASTTS). يتكون النظام المتتالي من مرحلتين رئيسيتين متكاملتين: التعرف التلقائي على الكلام العربي (AASR)والتلخيص التجريدي للنص العربي (AATS). تم اعتماد نموذج التعلم الذاتي القائم على الإشراف ونموذج المحول لكلا المرحلتين لحل مشكلة نقص مجموعة البيانات، والاعتماديات طويلة المدى في تسلسل الكلام المتصل، والتنفيذ بصورة متوازية.

بالنسبة لمرحلة AASR، تم تنفيذ عدة خطوات متسلسلة بدأت بإعداد مجموعة البيانات ، تلتها خطوات المعالجة المسبقة لمجموعة البيانات لجعلها مناسبة لنماذج الكلام للتعلم العميق، ثم تم استخراج الميزات وتصنيفها باستخدام سبعة نماذج مقترحة (ثلاثة نماذج كلام محسنة مسبقة التدريب أطلق عليها (3MPSM)، ونموذج HUBERT المحسن مع جدولة معدل التعلم(MHLR)، فضلا عن ثلاثة نماذج هجينة). بالنسبة ل 3MPSM، تم استخدام ثلاثة نماذج منفصلة (wav2vec2xlsr، HUBERT، MMS) للتعرف على الكلام العربي بعد الضبط الدقيق لتلك النماذج على الجزء العربي من مجموعات البيانات (common voice, FLEURS) وتجميد جزء استخلاص الميزات

وتدريب باقي الطبقات، فضـــلا عن تحســـين عدد من قيم المعلمات الفائقة. تم تنفيذ نموذج MHLR كتحسين لنموذج HUBERT عن طريق إضافة طبقات إضافية وتم استخدام مجدول معدل التعلم من النوع ("reduce_lr_on_plateau") لمنع النموذج من الإفراط في الموائمـة من خلال تقليـل قيمـة معدل التعلم في حالة عدم حدوث تحسن بخطأ التحقق (validation loss) لعدد معين من الدورات.

كما تم اقتراح ثلاثة نماذج هجينة عن طريق أولا: التهجين لكل من نموذج wav2vec2xlsr ونموذج HUBERT مع الطبقة التكيفية لنموذج MMS وثانيا: تهجين نموذج HUBERT مع أربع كتل من مرمز المحولات لبناء مرمز ثنائي مع نموذج المرمز الأصـــلي. تم ضـــبط هذه النماذج على مجموعة بيانات FLEURS المعيارية واســتخدام تقنية SWATS والتي يتم التبديل فيها بين محســن ADAM ومحسن الSGD .

بعد تقييم نماذج AASR باســـتخدام مقياس معدل خطأ الكلمات(WER) ومقاييس أخرى، أظهرت النتـائج أنـه عنـد تهجين نموذج HUBERT مع 4 كتـل من مرمز المحولات، تفوق على النماذج المقترحة الأخرى من حيث WER حيث أن WER هو0.23 كما أظهر نتائج تنافســـية عند مقارنته بنموذج الكلام متعدد اللغات الضـــخم (MMS)الذي تم تدريبه مســـبقا على عدد كبير من مجموعات البيانات واللغات، فضلا عن تفوق النماذج المقترحة عن الدراسات المتعلقة السابقة.

اما بالنسبة لمرحلة AATS، تم تنفيذ ثلاثة نماذج محسنة (mT5, AraBART, mbert2mbert) عن طريق الضبط الدقيق على مجموعة بيانات ضمن المجال وخارج المجال وتقييمهما باستخدام مقاييس مختلفة. أظهرت النتائج أن النماذج المحسنة المبنية على المحولات أعطت نتائج واعدة وتفوقت على النماذج الأخرى في الدراسات السابقة. كما أظهرت أن نموذج mT5 تفوق على نماذج دراستنا الأخرى من حيث BERTScore حيث بلغ BERTScore = 0.76 لمجموعة البيانات داخل المجال يليه 0.75 و0.69 لنماذج AraBART و mbert2mbert على التوالي.

تم دمج نماذج مرحلتي النظام الذكي، وأظهرت النتائج أن نموذج mT5 تفوق على النماذج الأخرى من حيث قابلية القراءة وجودة الملخص عند تقييمه على مجموعة بيانات FLEURS المعيارية على الرغم من قصر النص المدخل.