Research and Innovation Action (RIA) H2020 – 957017

Stream Learning for Multilingual Knowledge Transfer

https://selma-project.eu

D3.3 Initial release of post-editing and user feedback capabilities

Work Package	3
Responsible Partner	LIA
Author	Yannick Estève
Contributors	Antoine Caubrière, Jarod Duret, Tilo Himmelsbach, Gaëlle Laperrière, Salima Mdhaffar, Tugtekin Turan, Marcely Zanon Boito
Reviewer	Tugtekin Turan
Version	V2.5
Contractual Date	31 March 2022
Delivery Date	~~31 March 2022~~, ~~25 May 2022~~, 18 November 2022
Dissemination Level	Public

Version History

Version	Date	Description
0.1	15.03.2022	Initial Table of Contents (ToC)
0.2	29.03.2022	Main input from partners
0.3	30.03.2022	Merging
0.4	31.03.2022	Internal Review version
0.5	31.03.2022	Finalization
1.0	31.03.2022	Publishable version
1.5	17.05.2022	Unmerged deliverables
2.0	25.05.2022	Publishable version
2.1	11.11.2022	Corrected footer + section reference
2.2	16.11.2022	New Introduction
2.5	18.11.2022	Resubmission

Executive Summary

This initial deliverable describes the first release of software components developed within WP3 in particular relating to post-editing and user feedback capabilities. This document will be followed by the interim and final releases later in the project.

SELMA's approach to speech and language processing is targeting both low and high resourced languages

Table of Contents

1. Introduction

In this document, we present our first software that allows us to inject linguistic information from text into an end-to-end neural ASR model. To our knowledge, this is the first software that offers this possibility for such technology.

As this stage, the state of the software is not ready to be deployed as a component in the SELMA platforms, it is still a research prototype.

2. Released Software

2.1 Toward the Injection of User Feedback and Linguistic Information

As in Section 2.5 of D3.2, named entity recognition from speech consists in recognizing words from speech, detecting word sequences that support a named entity, and categorizing this entity. End-to-end neural approaches suffer from the lack of paired audio and textual data with a named entity annotation. An end-to-end model for named entity recognition from speech without paired training data has been built in the framework of the SELMA project. This success opens new perspectives in order to update the linguistic information into a pre-trained end-to-end ASR model. Such linguistic information could be obtained from user feedback or textual daily news.

Our approach is based on the use of an external model (named Text-to-ASR-Embeddings model) to generate a sequence of vectorial representations from text, similar to the ASR hidden representations. A NER-S module is then trained using these representations as input and the annotated existent textual data as output.

This ASR model is based on the software material shared here: https://github.com/SELMA-project/LIA_speech/tree/main/asr.

The Text-to-ASR-Embeddings model is used to mimic the 80-dimensional embeddings of ASR. It is based on the Tacotron2 neural architecture designed for speech synthesis (text-to-speech). The NER module consists of a BiLSTM model composed of 5 BiLSTM layers, with

512 dimensions each. Experiments carried out on the QUAERO corpus show that this approach is very promising. A paper has just been submitted to INTERSPEECH 2022[1].

This novel contribution is a step to the first direction planned in the scientific document of the SELMA project. Thanks to this success, we will be able to investigate novel ways to generate massive amounts of training data for the post-editing task.

3. Future Plan

Starting from the first success on the injection of textual data into the training of speech to text end-to-end models, automatic post-editing task will be first addressed as an “on-the-fly“ fine-tuning approach of the implicit language model embedded in the end-to-end neural architecture. An experimental protocol will be implemented in order to confirm the benefits of such an approach.

In addition, the user feedback entered into the GUI will also be taken into account, and coupled to entity information provided by T2.1 and T2.2.

When the use of these new research contributions will be validated in the framework of the SELMA use cases, the software will be consolidated in order to be deployed into the SELMA platforms.

[1] https://interspeech2022.org