

Progress Report 3

Jayveersinh Raj, Makar Shevchenko, Nikolay Pavlenko

April 11th, 2023

1 Current Progress

Dear comrades in software engineering, we have a momentous announcement to make. Our Cross-lingual NLP project has achieved a groundbreaking milestone that will change the world of language processing forever.

Thanks to the dedication and hard work of our talented NLP engineers, we have implemented an incredibly robust and powerful DistilBERT model that can process a staggering 60 sentences per second (sentences themselves consist of 128 tokens - a count that is significantly larger than is normally used in human speech). This unprecedented performance speed is a testament to the innovation and expertise of our team.

Once it develops the ability to effectively process and translate multiple languages, including low-resource ones, our Cross-lingual NLP project is going to decisively break down language barriers and making information more accessible to people around the world. This is a crucial step towards creating a more connected and inclusive global community.

We are immensely proud of the progress we have made and the potential impact of our project. This achievement would not have been possible without the tireless efforts of our team and the support of their friends and family.

Together, we are forging a new era of language processing and communication. Let us continue to push the boundaries of what is possible and work towards a world where language is no longer a barrier to understanding and progress, and information can flow freely between all corners of the globe.

2 Team Member Contribution

- **Jayveersinh Raj:** work on different pre-trained models for language identification;

- **Makar Shevchenko:** implementation of DistilBERT, work on interpretable metrics for our pipelines;
- **Nikolay Pavlenko:** composition of intermediary report.

3 Plan for the Next Three Weeks

In the future three weeks we intend to implement metrics evaluation, results of which can be easily interpreted and used on the final project presentation and defense, as well as continue the work towards zero-shot translation for different languages. As a baseline we will be sticking to Latin ones, as they are more easily identified with pre-trained models.

4 Github Link

Link:https://github.com/SyrexMinus/cross_lingual_nlp