International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

National Center for Scientific and Technical Research, Abdelmalek Essadi University, Tetouan, Morocco

Chaali Kaoutar[1], Lachhab Yasser[2], Sefiani Maha[3]

[1]Accredited Senior Lecturer, Abdelmalek Esaadi
[2,3]Phd Student, Abdelmalek Esaadi

Abstract:
Over the same period, thanks to deep neural networks, particularly in machine translation tasks that fall under natural language process (NLP), there have been major advances. Context is not something that lies well with traditional systems based on statistical models or tranlation memory and they also struggle with the level of fluency in translation. Through the invention of their Transformer models, Vaswani et al. This has also made great strides in improving both precision and performance [28] from 2017.
In this paper, we explore how to adapt pre-trained Transformer models into low-resource languages. For lower-engaged languages, we want to introduce hybrid methods in machine translation systems using multi-task learning and also transfer-learning with different metrics. To address this, we exploit the high-resource languages for training and introduce representation reduction techniques to be able to better handle some low-resourced cases.
Our results show that multi-task training improves BLEU scores by a large margin, especially for more scarce languages such as Swahili and Amharic. Our results exemplify the power of combining effective multi-task training with transfer learning in low-resource language translation.
Keywords : Natural Language Processing (NLP), Machine Translation , Transformer Models ,Low-Resource Languages ,Multi-Task Learning , Transfer Learning, Vocabulary Reduction , Neural Networks , BLEU Score , Language Models .

Introduction :

The study of deep neural networks had a clear impact in the advancement of Natural Language Processing (NLP, KerMet translation), enabling better translations machine-translation wise. Fluency and contextual adaptation, for example, were often a pain point for traditional translation systems that employed statistical models or translation memory. In light of this, it was introduced by Vaswani et al. the Transformer networks. Since then, a horizon of the inaccuracy and underperformance era is over with 2017.
In this article, we solve the problem that in low-resource languages, when using Transformer model for mostly translation application cases and hope to make performance as good it is possible. In this work, we aim to take a step towards making BERT more data-efficient for low-resource languages by investigating hybrid and multi-task methods.

Literature Review :

The earliest methods of machine translation were rule-based algorithms and word dictionaries. Then, this

was followed by statistical models like the IBM Model 1 through model5 which used probabilistic word alignments for translation.
Neural networks (LSTM, GRU) Translation memory models improved the translation quality by taking long-term context into account. The main problem was that these models found it difficult to effectively process longer sequences.
Since then, the Transformer model and its attention mechanism has set a new standard in NLP due to their effective parallel computation of sequences. It is better many variations of recurrent models, and faster to train. Many recent studies including BERT (Devlin et al., 2018) and GPT (Radford et al., 2019), have pointed out the usefulness of supervising context-aware learning by large textual corpora. However, these models scale poorly for languages with little to no data.

Methodology:
In this work we present a model using pre-trained Transformer with multi-task training. We aim to boost machine translation for less-resourced languages. Our method combines:
Multi-task learning: We merged training data of the high-resource languages (English, Spanish) and low-resources ones (like Swahili, Amharic), while making use of a single model that can predict multiple translations.
Transfer learning: The model is first pre-trained on a large multilingual corpus (e.g., OpenSubtitles) and then fine-tuned towards target languages.
Lowered Vocabulary: To make it more efficient in small word count languages, a vocabulary reduction technique was used by reducing easy-learn words from the reduced-vocabulary lists to decrease the language of interest efficiency,targeting better content measurement as usual for context capturing in Low-resourced languages.
Testing was done on standard datasets suc91 as WMT (Workshop for Machine Translation), and compared against baseline models like LSTM, or traditional Transformer models.

Results and Discussion:
Results We observe a dramatic performance increase for the Transformer model when multi-task training is used. The multi-task model scored 32.5 BLEU for Swahili, compared to Transformer (28.7) and LSTM models (25.4). The same goes for Amharic, the improved model gave a BLEU score of 30.1 whereas other solutions are performing lower than this one.
Essentially, the hypothesis that transfer learning as well as multi-task training enables model to efficiently utilize information learnt from high-resource languages for better understanding and translation of underrepresented ones is confirmed by our results.
Contrary to Zoph et al. (ijcai) In contrast with Luong et al. (2016) that used attention-based neural networks for translation between closely related languages, our method provides more adaptability among the systems of distant linguistic characteristics. The vocabulary reduction technique, too, was essential to the naturalness of sentence composition in different alphabets or languages even if—and maybe especially when—algorithms will work with tens and hundreds of characters as space.

Conclusion:
In this work, we showed that levering a dual training and transfer learning based optimization results in much better machine translation models compared to baseline Transformers especially for low resource

language pairs. It points to the need for custom-tailored neural network architectures in multilingual settings, particularly when data is scarce.

We plan to investigate model compression techniques in the future for faster inference suitable on real-time systems. On the other hand, more analysis on bias specific algorithms with respect to some languages would help a lot in translating fairly across multiple language outputs from models.

Bibliographie:

1. **Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I.** (2017). *Attention is all you need*. In Advances in neural information processing systems (pp. 5998-6008).
2. This paper introduces the Transformer model, which revolutionized NLP tasks such as machine translation by eliminating recurrence and relying solely on attention mechanisms.
3. **Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K.** (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv preprint arXiv:1810.04805. This paper presents BERT, a Transformer-based model that demonstrated the value of pre-training on large text corpora for improved language comprehension.
4. **Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I.** (2019). *Language Models are Unsupervised Multitask Learners*. OpenAI Blog. This work explores GPT, which uses a similar architecture to the Transformer and shows how language models can generalize well across a variety of tasks.
5. **Zoph, B., & Knight, K.** (2016). *Multilingual Neural Machine Translation*. arXiv preprint arXiv:1601.01073.
   This paper explores neural machine translation techniques for multiple languages, highlighting the benefits of using shared models for low-resource language pairs.
6. **Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., ... & Dean, J.** (2017). *Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation*. Transactions of the Association for Computational Linguistics, 5, 339-351. This work discusses Google's multilingual neural machine translation system, which enables translation between languages without direct pairwise examples.
7. **Bahdanau, D., Cho, K., & Bengio, Y.** (2014). *Neural Machine Translation by Jointly Learning to Align and Translate*. arXiv preprint arXiv:1409.0473. Introduces the concept of attention in neural machine translation, which laid the groundwork for Transformer-based architectures.
8. **Sennrich, R., Haddow, B., & Birch, A.** (2016). *Neural Machine Translation of Rare Words with Subword Units*. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers, pp. 1715-1725). This paper introduces subword tokenization techniques, which are crucial for handling languages with limited data and large vocabularies.
9. **Koehn, P.** (2009). *Statistical Machine Translation*. Cambridge University Press. Provides an in-depth overview of the methods and challenges in statistical machine translation, offering a comparison with newer neural approaches.
10. **Arivazhagan, N., Bapna, A., Firat, O., Aharoni, R., Johnson, M., & Macherey, W.** (2019). *Massively Multilingual Neural Machine Translation in the Wild: Findings and Challenges*. arXiv

preprint arXiv:1907.05019.
Discusses multilingual neural machine translation in real-world settings and the challenges faced in low-resource languages.
11. **Tiedemann, J.** (2012). *Parallel Data, Tools and Interfaces in OPUS*. In LREC (Vol. 2012, pp. 2214-2218).
Describes the OPUS project, which provides a large-scale parallel corpus useful for training machine translation systems, particularly for low-resource languages.