FonMTL: Towards Multitask Learning for the Fon Language

Bonaventure F. P. Dossou
McGill University - Mila Quebec AI Institute
Masakhane NLP, Lanfrica
Lelapa AI
bonaventure.dossou@mila.quebec

Iffanice Houndayi
Carnegie Mellon University Africa
ibh@alumni.cmu.edu

Pamely Zantou
Carnegie Mellon University Africa
pzantou@andrew.cmu.edu

Gilles Hacheme
Microsoft AI For Good Kenya
Masakhane NLP
ghacheme@microsoft.com

Abstract

This paper presents the first explorative approach to multitask learning, for model capabilities enhancement in Natural Language Processing for the Fon language. Specifically, we explore the tasks of Named Entity Recognition (NER) and Part of Speech Tagging (POS) for Fon. We leverage two language model heads as encoders to build shared representations for the inputs, and we use linear layers blocks for classification relative to each task. Our results on the NER task for Fon show a competitive performance compared to the performance of individual several multilingual models. Additionally, we perform a few ablation studies to leverage the efficiency of two different loss combination strategies and find out that the equal loss weighting approach works best in our case.

1 Introduction

Multitask learning (MTL) is a learning paradigm that aims to improve the generalization capacity of a model by sharing knowledge across different but related tasks. Learning one task at a time might be ineffective and prone to overfitting, low data efficiency or slow learning (Zhang et al., 2022), especially for large and complex problems. This might sometimes lead to the training of very dedicated models with generalization capacity only local to the tasks they have been trained on. MTL, precisely for that reason targets a more efficient way of learning, by sharing common representations of the inputs and by implicitly transferring information among the various tasks (Caruana, 1993, 1997). MTL efficiency and promises have been demonstrated in the literature (Gessler and Zeldes, 2022; Ruder, 2017) along with the benefit of the understanding of tasks (their similarity, relationship, hierarchy) for MTL.

For African low-resourced languages in general, building NLP models on specific tasks can be in general difficult for several reasons, including the availability of datasets (Joshi et al., 2020; Emezue and Dossou, 2021; Nekoto et al., 2020). This is even more true for the Fon language, a truly low-resourced African language. Therefore, MTL appears to be a promising approach to explore for downstream tasks on the Fon language.

Part-of-Speech (POS) assigns grammatical groups (i.e. whether it is a noun, an adjective, a verb, an adverb, and more) to words in a sentence using contextual cues and assigning corresponding tags. On the other hand, Named Entity Recognition (NER) tries to find out whether or not a word

5th Deep Learning Indaba Conference (DLI 2023).

is a named entity (persons, locations, organizations, time expressions, and more). This problem is twofold: detection of names and categorizing names. Consequently, whereas POS is more of a global problem since there can be relationships between the first and the last word of a sentence, NER is rather local, as named entities are not spread in a sentence and mostly consist of uni, bi, or trigrams. Albeit different, and adding the fact that both tasks are classification tasks, we then speculate that both tasks could benefit each other.

2 Experiments and Results

In this paper, we explore the first multitask learning model for the Fon language, particularly, focusing on the NER and POS tasks. We used a hard parameter sharing approach — the most popular MTL method used in the literature Caruana (1993); Ruder (2017) and falls into the joint training method described by Zhang et al. (2022) in their taxonomy of MTL approaches for NLP.

Experiments: For NER, we used the Fon dataset of MaskhaNER 2.0 (Adelani et al., 2022), a NER dataset of 20 African languages, including Fon. For the POS task, we used the MasakhaPOS dataset (Dione et al., 2023), the largest POS dataset for 20 typologically diverse African languages, including Fon. Due to the fact the test sets for MasakhaPOS are not public yet, we hence used the dataset solely in our training and validation process. In MassakhaNER 2.0, we have 4343/621/1240 as train/dev/test set sizes, with a total of 173099 tokens. In MasakhaPOS we have 798/159/637 as train/dev/test set sizes, with a total of 49460 tokens and 30.6 tokens on average per sentence.

In our MTL model (Figure 1), we use two language model (LM) heads: AfroLM-Large (Dossou et al., 2022) and XLMR-Large (Conneau et al., 2020). AfroLM has been pretrained from scratch on 23 African languages in an active learning framework, while XLMR-Large has been pretrained on 100 languages (with ≤5 African). Each LM head is used to build representations of the inputs from each task, and both representations are then combined (in an additive way) to build a shared representation across models and tasks. Following the shared representation, we built two linear layers, serving as classification layers for each respective task.

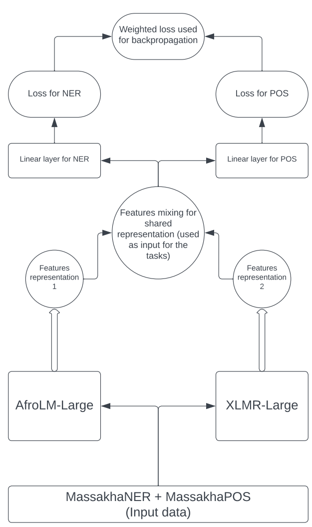

Figure 1: Our model architecture

We combined the cross-entropy losses of both tasks in two ways: (a) unweighted sum (Kurin et al., 2022) i.e. the unitary sum of each individual loss, and (b) equally weighted where losses are weighted by two parameters α and β then added up together. For simplicity, in our experiments, we set $\alpha = \beta = 0.5$. We used a small batch size of 4, the AdamW (Loshchilov and Hutter, 2017) as optimizer with a learning rate of 3e-5.

Table 1: Performance (F1-Score) of several multilingual pretrained language models (baselines) versus our MTL variants, on the Fon NER test set. MTL Sum is the MTL model with the unweighted sum of both losses, and MTL Weighted is the model with equally weighted sum of both losses.

Models	**F1-Score**
AfroLM-Large	80.48
AfriBerta-Large	79.90
XLMR-Base	81.90
XLMR-Large	81.60
AfroXLMR-Base	82.30
AfroXLMR-Large	82.70
MTL Sum (ours)	79.87
MTL Weighted (ours)	81.32

Results: Along with our MTL variants, we evaluated several multilingual pretrained language models (MPLMs) on the NER task: AfroXLMR (Alabi et al., 2022) (base and large), AfroLM-Large (Dossou et al., 2022), AfriBERTa-Large (Ogueji et al., 2021), and XLMR (Conneau et al., 2020) (base and large). AfriBerta is a multilingual model pretrained on 11 African languages, while AfroXLMR is an adapted version of XLMR to 17 African Languages via Multilingual Adaptive Fine-Tuning.

In Table 1, we can see that our MTL models are competitive, and in some cases outperformed some MPLMs. This means that as we speculated earlier, POS task-related data and representations provided some additional useful information for the NER downstream performance. Moreover, this is also expected and desired as several empirical results have demonstrated in literature (Liu et al., 2019; Crawshaw, 2020; Gong et al., 2019) on various tasks like Natural Language Understanding (NLU) and GLUE tasks. Furthermore, we can see that the equal-weighted sum of both losses worked better than the unweighted sum (Gong et al., 2019; Rengasamy et al., 2020), which makes sense as we treated both tasks *equally*.

3 Conclusion

In this paper, we presented the first effort to leverage MTL for NLP downstream tasks in Fon. Our results on Fon NER task showed a competitive (and sometimes better) performance of our MTL methods. We hope these results enable more exploration of MTL in NLP for Fon in particular, and African Languages in general. In future works, we want to explore dynamic weighted average loss (Rengasamy et al., 2020; Gong et al., 2019) which has been empirically demonstrated to be effective.

References

David Adelani, Graham Neubig, Sebastian Ruder, Shruti Rijhwani, Michael Beukman, Chester Palen-Michel, Constantine Lignos, Jesujoba Alabi, Shamsuddeen Muhammad, Peter Nabende, Cheikh M. Bamba Dione, Andiswa Bukula, Rooweither Mabuya, Bonaventure F. P. Dossou, Blessing Sibanda, Happy Buzaaba, Jonathan Mukiibi, Godson Kalipe, Derguene Mbaye, Amelia Taylor, Fatoumata Kabore, Chris Chinenye Emezue, Anuoluwapo Aremu, Perez Ogayo, Catherine Gitau, Edwin Munkoh-Buabeng, Victoire Memdjokam Koagne, Allahsera Auguste Tapo, Tebogo Macucwa, Vukosi Marivate, Mboning Tchiaze Elvis, Tajuddeen Gwadabe, Tosin Adewumi, Orevaoghene Ahia, Joyce Nakatumba-Nabende, Neo Lerato Mokono, Ignatius Ezeani, Chiamaka Chukwuneke, Mofetoluwa Oluwaseun Adeyemi, Gilles Quentin Hacheme, Idris Abdulmumin, Odunayo Ogundepo, Oreen Yousuf, Tatiana Moteu, and Dietrich Klakow. 2022. https://aclanthology.org/2022.emnlp-main.298 MasakhaNER 2.0: Africa-centric transfer learning for named entity recognition. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4488–4508, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Jesujoba O. Alabi, David Ifeoluwa Adelani, Marius Mosbach, and Dietrich Klakow. 2022. https://aclanthology.org/2022.coling-1.382 Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4336–4349, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

R Caruana. 1993. Multitask learning: A knowledge-based source of inductive bias1. In *Proceedings of the Tenth International Conference on Machine Learning*, pages 41–48. Citeseer.

Rich Caruana. 1997. https://doi.org/10.1023/a:1007379606734 Multitask learning. *Machine Learning*, 28(1):41–75.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. https://doi.org/10.18653/v1/2020.acl-main.747 Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Michael Crawshaw. 2020. Multi-task learning with deep neural networks: A survey. *arXiv preprint arXiv:2009.09796*.

Cheikh M. Bamba Dione, David Adelani, Peter Nabende, Jesujoba Alabi, Thapelo Sindane, Happy Buzaaba, Shamsuddeen Hassan Muhammad, Chris Chinenye Emezue, Perez Ogayo, Anuoluwapo Aremu, Catherine Gitau, Derguene Mbaye, Jonathan Mukiibi, Blessing Sibanda, Bonaventure F. P. Dossou, Andiswa Bukula, Rooweither Mabuya, Allahsera Auguste Tapo, Edwin Munkoh-Buabeng, victoire Memdjokam Koagne, Fatoumata Ouoba Kabore, Amelia Taylor, Godson Kalipe, Tebogo Macucwa, Vukosi Marivate, Tajuddeen Gwadabe, Mboning Tchiaze Elvis, Ikechukwu Onyenwe, Gratien Atindogbe, Tolulope Adelani, Idris Akinade, Olanrewaju Samuel, Marien Nahimana, Théogène Musabeyezu, Emile Niyomutabazi, Ester Chimhenga, Kudzai Gotosa, Patrick Mizha, Apelete Agbolo, Seydou Traore, Chinedu Uchechukwu, Aliyu Yusuf, Muhammad Abdullahi, and Dietrich Klakow. 2023. http://arxiv.org/abs/2305.13989 Masakhapos: Part-of-speech tagging for typologically diverse african languages.

Bonaventure F. P. Dossou, Atnafu Lambebo Tonja, Oreen Yousuf, Salomey Osei, Abigail Oppong, Iyanuoluwa Shode, Oluwabusayo Olufunke Awoyomi, and Chris Emezue. 2022. https://aclanthology.org/2022.sustainlp-1.11 AfroLM: A self-active learning-based multilingual pretrained language model for 23 African languages. In *Proceedings of The Third Workshop on Simple and Efficient Natural Language Processing (SustaiNLP)*, pages 52–64, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Chris Chinenye Emezue and Bonaventure F. P. Dossou. 2021. https://aclanthology.org/2021.wmt-1.48 MMTAfrica: Multilingual machine translation for African languages. In *Proceedings of the Sixth Conference on Machine Translation*, pages 398–411, Online. Association for Computational Linguistics.

Luke Gessler and Amir Zeldes. 2022. https://aclanthology.org/2022.mrl-1.9 MicroBERT: Effective training of low-resource monolingual BERTs through parameter reduction and multitask learning. In *Proceedings of the The 2nd Workshop on Multi-lingual Representation Learning (MRL)*, pages 86–99, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Ting Gong, Tyler Lee, Cory Stephenson, Venkata Renduchintala, Suchismita Padhy, Anthony Ndirango, Gokce Keskin, and Oguz H. Elibol. 2019. https://doi.org/10.1109/ACCESS.2019.2943604 A comparison of loss weighting strategies for multi task learning in deep neural networks. *IEEE Access*, 7:141627–141632.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. https://doi.org/10.18653/v1/2020.acl-main.560 The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

Vitaly Kurin, Alessandro De Palma, Ilya Kostrikov, Shimon Whiteson, and Pawan K Mudigonda. 2022. https://proceedings.neurips.cc/paper$_{f}iles/paper/2022/file/4f301ae934f396086bfefd1139039dbd-Paper-Conference.pdf In defense of the unitary scalarization for deep multi-task learning. In$ *Advances in Neural* $-12183. Curran Associates, Inc.$

Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019. Multi-task deep neural networks for natural language understanding. *arXiv preprint arXiv:1901.11504*.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Muhammad, Salomon Kabongo Kabenamualu, Salomey Osei, Freshia Sackey, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene Ahia, Musie Meressa Berhe, Mofetoluwa Adeyemi, Masabata Mokgesi-Selinga, Lawrence Okegbemi, Laura Martinus, Kolawole Tajudeen, Kevin Degila, Kelechi Ogueji, Kathleen Siminyu, Julia Kreutzer, Jason Webster, Jamiil Toure Ali, Jade Abbott, Iroro Orife, Ignatius Ezeani, Idris Abdulkadir Dangana, Herman Kamper, Hady Elsahar, Goodness Duru, Ghollah Kioko, Murhabazi Espoir, Elan van Biljon, Daniel Whitenack, Christopher Onyefuluchi, Chris Chinenye Emezue, Bonaventure F. P. Dossou, Blessing Sibanda, Blessing Bassey, Ayodele Olabiyi, Arshath Ramkilowan, Alp Öktem, Adewale Akinfaderin, and Abdallah Bashir. 2020. https://doi.org/10.18653/v1/2020.findings-emnlp.195 Participatory research for low-resourced machine translation: A case study in African languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2144–2160, Online. Association for Computational Linguistics.

Kelechi Ogueji, Yuxin Zhu, and Jimmy Lin. 2021. https://aclanthology.org/2021.mrl-1.11 Small data? no problem! exploring the viability of pretrained multilingual language models for low-resourced languages. In *Proceedings of the 1st Workshop on Multilingual Representation Learning*, pages 116–126, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Divish Rengasamy, Mina Jafari, Benjamin Rothwell, Xin Chen, and Grazziela P. Figueredo. 2020. https://www.mdpi.com/1424-8220/20/3/723 Deep learning with dynamically weighted loss function for sensor-based prognostics and health management. *Sensors*, 20(3).

Sebastian Ruder. 2017. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Zhihan Zhang, Wenhao Yu, Mengxia Yu, Zhichun Guo, and Meng Jiang. 2022. A survey of multi-task learning in natural language processing: Regarding task relatedness and training methods. *arXiv preprint arXiv:2204.03508*.