Review of the Diploma Thesis

The Faculty of Mathematics and Physics, Charles University

Thesis author	Jacobo Del Valle
Title	Analyzing Modular Cross-Lingual Transfer Learning
The year of completion	2023
Study programe	Computer Science - Language Technologies and Computational Linguistics
Study area	Computer Science - Language Technologies and Computational Linguistics

Review author	Tomasz Limisiewicz	**Role**	oponent
Workplace	Institue of Formal and Applied Linguistics		

Text of the review:

Content. The thesis uses the previously introduced method of Cross-lingual Modular training (X-MOD Pfeiffer et al.) for model adaptation to cover low-resource languages. The first chapter introduces multilingual language modeling and the methods for efficient training. The second is focused on the X-MOD method. It lists the claims of the original work, which are then tested in the experimental part.

The first experimental chapter reproduces the X-MOD method of post-hoc language addition. The results are close but consistently lower than the reported in the paper. The drop in results is explained by the differences in the hyperparameter choices, especially the number of training steps which was not reported by Pfeiffer et al. Subsequently, the author investigates the vulnerability of the method to three aspects of model adaptation: 1) the number of training steps; 2) vocabulary overlap between pre-trained and added language vocabularies; 3) the amount of data available for the added language. All these aspects are shown to be significant. Importantly, the analysis identifies the ranges of the examined parameters required for X-MOD's effectiveness.

The second experimental chapter applies the method to an NLI dataset of American low-resource language. The results show that X-MOD underperforms a language-adapted XLM-R trained by Ebrahimi et al. The author explains worse performance due to the scarcity of unannotated data for chosen American languages. This finding is supported by the analysis from the previous chapter and qualifies the claim of the original paper that X-MOD can be applied to any language. Finally, the thesis evaluates the possibility of merged adaptation, yet it does not bring improvement to the results.

Main strengths. The thesis focuses on an important problem of multilingual NLP. Namely, efficient model adaptation to low-resource languages.

The author diligently reproduces the results of one established method, i.e., X-MOD. Importantly, he critically addresses the statements made by the authors of the original paper. The thesis contains a multifactor analysis of the aspects influencing the method's effectiveness. The experiments in this area are well-designed, and their results highlight the constraints of the X-MOD method. Subsequent experiments show that these constraints limit the method's applicability in practice.

Limitations. The main drawback of the thesis is its limited scope. As mentioned before, the critical reproduction of X-MOD results deserves due recognition. However, it addresses only the claim about coverage, leaving a comparison with traditional pre-training unaddressed. Also, the author evaluates adaptation methods on just one downstream task (NLI). Noticeably, the experimental part spans only 23 pages. It is preceded by the background description of a comparable size. In my opinion, the first background chapter is too broad and at times chaotic. For instance, the sections about the course of dimensionality or next sentence prediction are redundant.

Because of the limited scope and occasional writing insufficiencies, it appears to me that the thesis was written under high time pressure.

Questions to the author.

1. How exactly the overlap coefficient was computed? Is it a portion of unique vocabulary units shared across two vocabularies? What would be the overlap factor when weighted by frequencies of the vocabulary units in a monolingual corpus?

2. How "perplexity divergence" phenomenon relate to the data sizes across languages? Specifically, is the minimum perplexity checkpoint obtained after the same number of epochs?

3. How would you explain that for NLI, it's better to pick the last checkpoint instead of the one achieving minimum perplexity?

Formal and technical aspects. Occasional typos and language mistakes, which at times make the text hard to understand. Table 2.4 contains factual mistakes: Pashto, Mongolian, and Sindhi are incorrectly assigned to language families.

Conclusion. The thesis addresses a crucial topic of model adaptation to low-resource languages. Although limited in scope, the thesis contains well-designed experiments and a critical reproduction of past results. These contributions fulfill the requirements for thesis defense.

I recommend the thesis to be accepted.

I do not recommend the thesis for the best thesis competition.

Prague on 25. 08. 2023

Signature: