Institute of Interactive Systems and Data Science ▪ Sandgasse 36/III, 8010 Graz, Austria ▪ office.isds@tugraz.at

Multilingual Label-Aware Contrastive Pre-Training of Transformers for Few- and Zero-shot Framing Detection

Markus Reiter-Haas, Alexander Ertl, Kevin Innerhofer, Elisabeth Lex

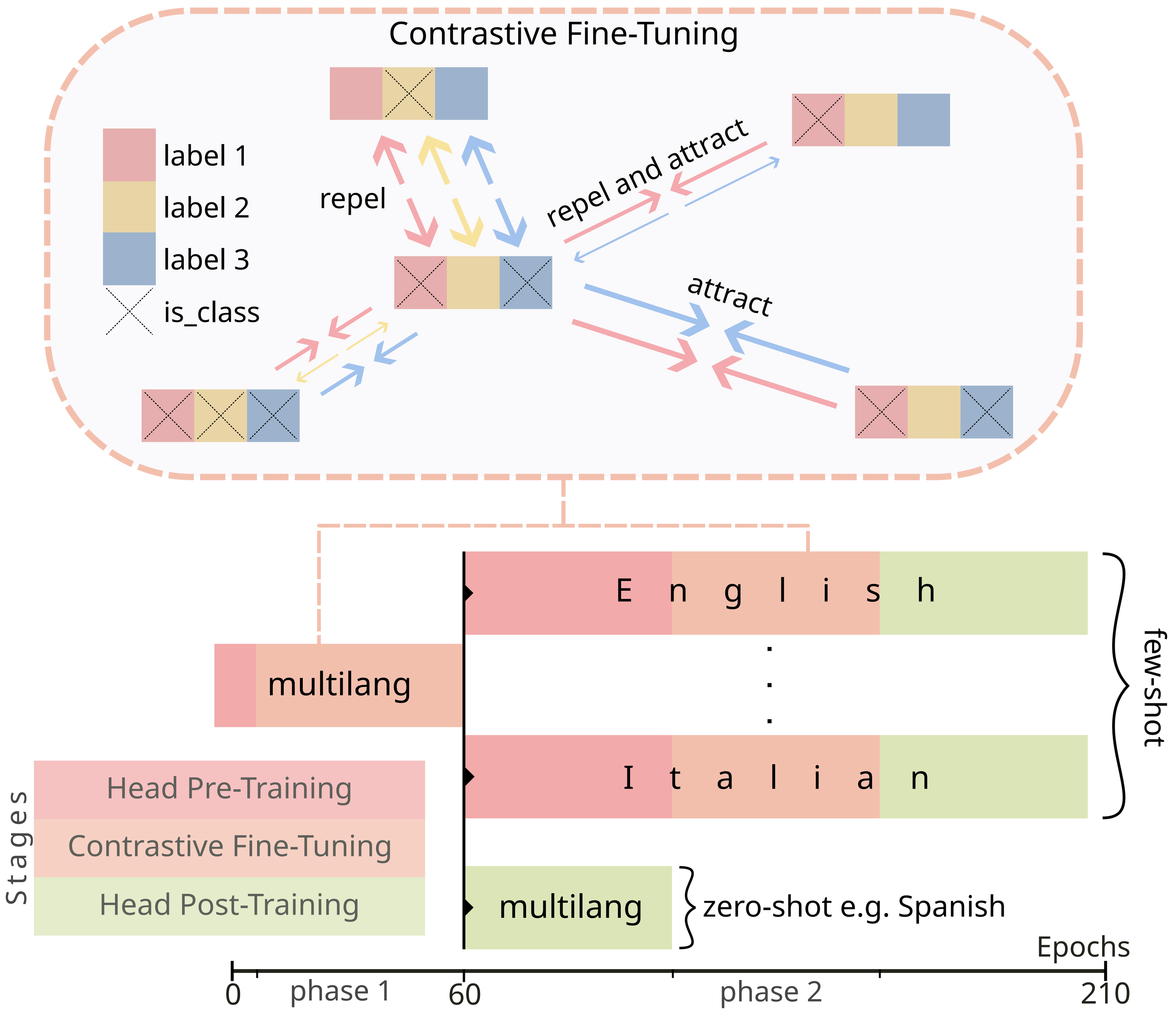
Institute for Interactive Systems and Data Science, Graz University of Technology
reiter-haas@tugraz.at, ertl@student.tugraz.at, innerebner@student.tugraz.at, elisabeth.lex@tugraz.at

Abstract

This paper presents the winning system for the zero-shot Spanish framing detection task, which also achieves competitive places in eight additional languages. The challenge of the framing detection task lies in identifying a set of 14 frames when only a few or zero samples are available, i.e., a multilingual multi-label few- or zero-shot setting. Our developed solution employs a pre-training procedure based on multilingual Transformers using a label-aware contrastive loss function. In addition to describing the system, we perform an embedding space analysis and ablation study to demonstrate how our pre-training procedure supports framing detection to advance computational framing analysis.

mCPT

- Multilingual Transformer body with a neural network classification head
- Two-phase, multi-stage pipeline. The first phase learns a multilingual model on all labeled data. Phase two is required to precisely learn the target distribution.
- The contrastive fine-tuning stage employs a contrastive learning loss function adopted from Zheng et al.[1] to improve representations



Our system performs label-aware contrastive fine-tuning (top) in a two-phase procedure (bottom). Embeddings of samples with similar labels are attracted, while they are repelled for dissimilar labels.

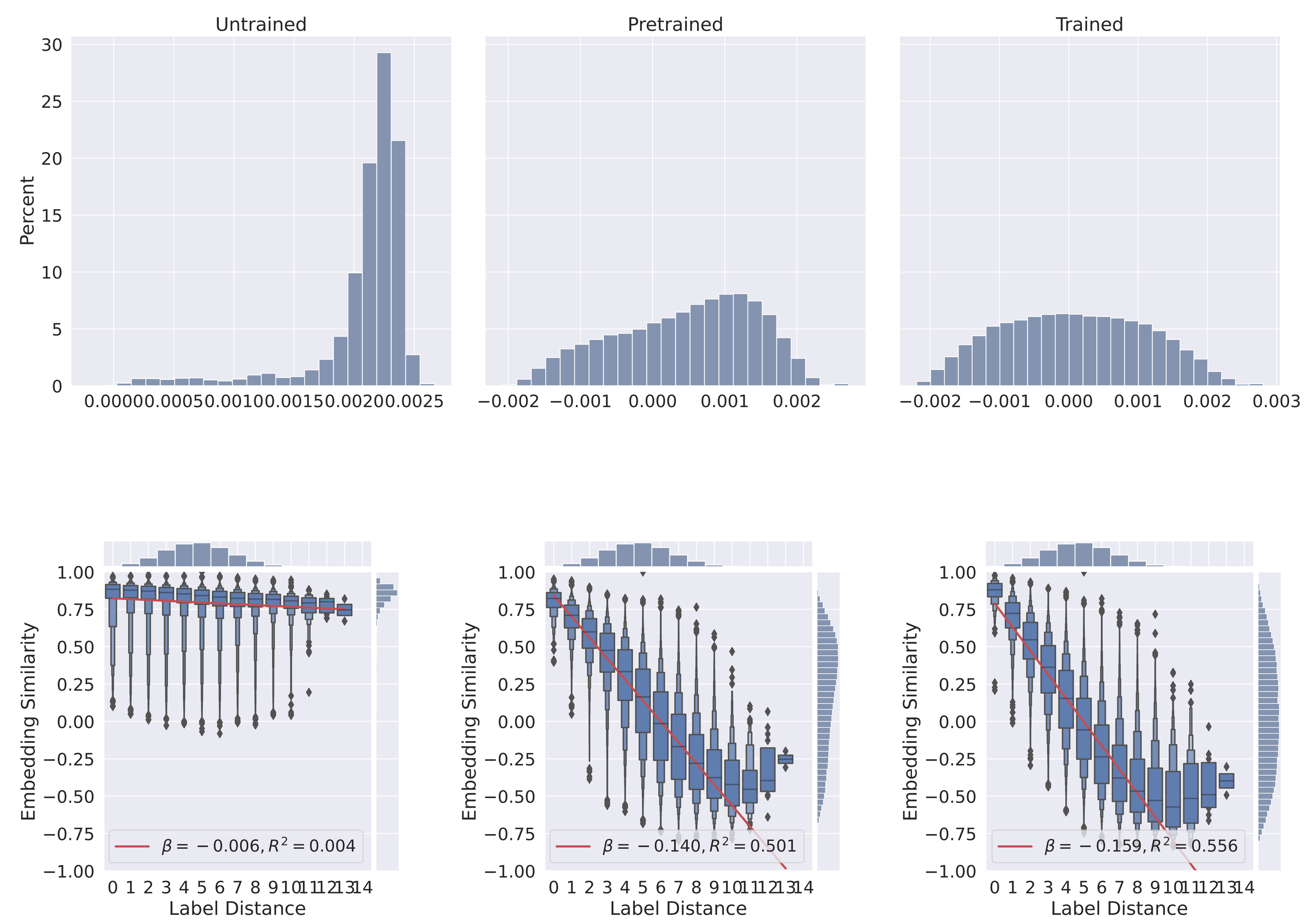
Results

The highlighted row shows the full model with contrast sampling enabled. Contrast sampling ensures that we find at least one negative pair per batch to improve the consistency of training. We then iteratively remove components of the model: (-CS) contrast sampling, (-PT) multilingual pre-training, (-$\mathcal{L}_{CON}$) the contrastive term of the loss function, and (-E2E) end-to-end training.

Model	en	it	ru	fr	ge	po	Δ
mCPT+CS	**.688**	**.590**	.519	**.575**	**.591**	.638	
- CS	.682	.585	**.520**	.570	.561	.636	-.008
- PT	.681	.545	.475	.563	.583	.616	-.015
- $\mathcal{L}_{CON}$	.657	.521	.436	.524	.570	**.645**	-.018
- E2E	.629	.519	.500	.535	.586	.633	.008

The generalization ability of our system is demonstrated by providing the winning contribution for the Spanish framing detection subtask where no training samples were available.

Training optimizes for uniformity[2], as shown by the shift of average similarities to zero (first row). It also preserves alignment[2] between samples with similar label vectors while pushing apart samples with dissimilar label vectors (second row).



(a) Without any training, pairs of embeddings are similar regardless of their label distance.

(b) After pre-training, the embedding cosine similarity reflects the Hamming distance of the labels.

(c) Fine-tuning on the target language further increases uniformity and alignment.

Conclusion

- Multilingual pre-training effectively increases the available amount of training data but may require target-distribution specific training
- Contrastive learning acts as a regularizer optimizing for uniformity and alignment in multi-label settings

References

[1] Zheng, L., Xiong, J., Zhu, Y., & He, J. (2022, August). Contrastive learning with complex heterogeneity. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 2594-2604).

[2] Wang, T. and Isola, P. (2020). Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. In International Conference on Machine Learning, pages 9929–9939.