Intent classification for Polish language

Grzegorz Przybylski
grzegorz.przybylski@student.uj.edu.pl

Paweł Fornalik
pawel.fornalik@student.uj.edu.pl

Szymon Dziedzic
szymon.stanslaw.dziedzic@student.uj.edu.pl

Supervisor: Andrii Krutsylo
andrii.krutsylo@ipipan.waw.pl

1 Introduction

Natural Language Processing (NLP) has witnessed significant advancements in recent years, enabling machines to understand and process human language more effectively. One crucial NLP task is intent classification, which involves determining the underlying intent or purpose behind a given user query or statement. Intent classification has numerous practical applications, such as chatbots, virtual assistants, customer support systems, and information retrieval systems. While significant progress has been made in intent classification for English, there is a notable scarcity of resources and research in intent classification for the Polish language. This project aims to bridge this gap by developing an intent classification model specifically tailored for the Polish language.

The ability to accurately classify user intents in the Polish language is essential for building efficient and user-friendly NLP systems that can cater to Polish-speaking users. By understanding user intent, systems can provide relevant and appropriate responses, improving user satisfaction and overall system performance. Moreover, with the increasing demand for multilingual NLP applications, it is crucial to expand the scope of research beyond English and include languages such as Polish to ensure inclusivity and accessibility for a broader user base.

For this project, we will leverage the existing research and methodologies in intent classification and transfer learning while adapting them to the unique characteristics and linguistic nuances of the Polish language.

In particular, we will use well-researched models trained on the Polish language, like HerBERT [10], PolBERT [8], and a classifier fine-tuned for this task. For fine-tuning purposes, we will use the Polish section of the MASSIVE [5] dataset published in 2022, which is a multilingual dataset for training intent classification models. The performance results achieved by our models expressed as intent accuracy and slot F1 score will be compared against a common baseline. As our baseline, we will use mT5 [15] and XLM-R [2] multilingual models. Our experiments aim to compare the results obtained from fine-tuned models with the performance reported in the MASSIVE paper [5].

The main research questions we would like to answer in our paper are:

1. Can we fine-tune the existing intent classification model for the Polish language such that it performs better than known multilingual intent classification models?

2. How does the performance of the Polish intent classification model vary with different model architectures or training strategies?

2 Related work

In recent years several models were developed on the transformer architecture [13], achieving notable performance in several NLP tasks, for example, BERT [3] trained on slot filling or GPT [11].

While intent classification for English has been extensively studied, the research on intent classification for the Polish language is limited. However, some studies have explored related tasks in Polish NLP, such as sentiment analysis, named entity recognition, and part-of-speech tagging.

Regarding intent classification, a few notable works have focused on multilingual intent classification, which can be adapted to Polish. For example, [4] proposed Multilingual BERT (M-BERT), a transformer-based model pre-trained on multilingual text, which achieved state-of-the-art performance on various NLP tasks, including intent classification.

Another relevant work [7] introduced a Polish corpus called PolEmo 2.0, which includes sentiment labels for Polish sentences. Although their work is primarily centered around sentiment analysis, the dataset can be leveraged for intent classification, as understanding sentiment is often intertwined with determining the underlying intent.

However, for intent classification in languages other than English, there was a deficit of large multilingual datasets. Of those that were, notable examples are SLURP [1], NLU Evaluation Data [14], Airline Travel Information System (ATIS) [6], Multilingual

Task-Oriented Semantic Parsing (MTOP) [9] or Cross-lingual Multilingual Task-Oriented Dialog [12].

In 2022 Amazon presented the MASSIVE dataset [5] — Multilingual Amazon SLURP (spoken language understanding resource package) for Slot-filling, Intent classification, and Virtual assistant Evaluation. MASSIVE contains 1M realistic, parallel, labeled virtual assistant utterances spanning 51 languages. They also presented modeling results on XLM-R [2] and mT5 [15], including exact match accuracy, intent classification accuracy, and slot-filling F1 score. Our Polish-specific models will be compared with these results. Amazon has released this dataset, modeling code, and models publicly.

References

[1] E. Bastianelli, A. Vanzo, P. Swietojanski, and V. Rieser. SLURP: A Spoken Language Understanding Resource Package. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.

[2] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*, 2019.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Multilingual bert. https://github.com/google-research/bert/blob/master/multilingual.md, 2018.

[5] J. FitzGerald, C. Hench, C. Peris, S. Mackie, K. Rottmann, A. Sanchez, A. Nash, L. Urbach, V. Kakarala, R. Singh, S. Ranganath, L. Crist, M. Britan, W. Leeuwis, G. Tur, and P. Natarajan. Massive: A 1m-example multilingual natural language understanding dataset with 51 typologically-diverse languages, 2022.

[6] C. T. Hemphill, J. J. Godfrey, and G. R. Doddington. The atis spoken language systems pilot corpus. In *Human Language Technology - The Baltic Perspectiv*, 1990.

[7] J. Kocoń, P. Miłkowski, and M. Zaśko-Zielińska. Multi-level sentiment analysis of PolEmo 2.0: Extended corpus of multi-domain consumer reviews. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 980–991, Hong Kong, China, Nov. 2019. Association for Computational Linguistics.

[8] D. Kłeczek. Polbert: Attacking polish nlp tasks with transformers. In M. Ogrodniczuk and Łukasz Kobyliński, editors, *Proceedings of the PolEval 2020 Workshop*. Institute of Computer Science, Polish Academy of Sciences, 2020.

[9] H. Li, A. Arora, S. Chen, A. Gupta, S. Gupta, and Y. Mehdad. MTOP: A comprehensive multilingual task-oriented semantic parsing benchmark. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2950–2962, Online, Apr. 2021. Association for Computational Linguistics.

[10] R. Mroczkowski, P. Rybak, A. Wróblewska, and I. Gawlik. HerBERT: Efficiently pretrained transformer-based language model for Polish. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 1–10, Kiyv, Ukraine, Apr. 2021. Association for Computational Linguistics.

[11] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever. Improving language understanding by generative pre-training, 2018.

[12] S. Schuster, S. Gupta, R. Shah, and M. Lewis. Cross-lingual transfer learning for multilingual task oriented dialog. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3795–3805, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. 2017.

[14] P. S. Xingkun Liu, Arash Eshghi and V. Rieser. Benchmarking natural language understanding services for building conversational agents. In *Proceedings of the Tenth International Workshop on Spoken Dialogue Systems Technology (IWSDS)*, pages xxx–xxx, Ortigia, Siracusa (SR), Italy, April 2019. Springer.

[15] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, and C. Raffel. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online, June 2021. Association for Computational Linguistics.