

SSN-MLRG at Text to Picto 2024: A BERT-Based Approach for Mapping French Sentences to Pictogram Terms

Notebook for the ImageCLEF Lab at CLEF 2024

Bhavana Anand¹, Themozhi J¹, Shreyas Sai R¹, Charumathi P¹ and Mirnalinee TT¹

¹ Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu, India

Abstract

Language impairment arising from different factors such as genetic diseases or incidents such as a car accident or stroke can impair language development skills and may lead to a partial or complete loss of the ability to communicate in written or spoken language. Augmentative and Alternative Communication refers to the means of communication used to substitute or replace spoken or written language. This paper focuses on the semantic mapping of French sentences to corresponding AAC pictograms. For this task, a transformer model utilizing CamemBERT embeddings, a French BERT model fused with a contrastive learning technique, was implemented. The model has obtained a PictoER score of 141.909, a BLEU score of 3.419, and a METEOR score of 14.351.

Keywords

Augmentative and Alternative Communication, Semantic Mapping, Transformer Models, CamemBERT, Natural Language Processing, Multimodal Communication, Contrastive Learning

1. Introduction

Communication is the act of giving and receiving information about that person's needs, desires, perceptions, knowledge, or affective states of another person. Language is the structured conventional system that is used to communicate with one another. AAC, by definition, is a therapeutic approach [1] that employs manual signs, symbol-based communication boards, and speech-generating computerized devices, integrating a person's entire spectrum of communication abilities.

Pictograms are visual communication tools that adeptly convey meanings, particularly effective in disambiguating homophones and other linguistically confounding terms. They provide the advantage of enabling communication from a foundational level—suitable for individuals with low cognitive abilities or those in early developmental stages—to a rich and advanced level although not with the same completeness and flexibility as written language [2].

The motivation behind this research stems from the communication gap between AAC users and others in society. There is a lack of awareness about such alternative communication methods. This leads to the need for a tool that converts modalities like speech and text into a sequence of pictograms to bridge this divide.

Inspired by the successful application of transformer models in natural language processing, we participated in the ImageCLEF 2024 ToPicto task [3] under the ImageCLEF 2024 evaluation campaign [4]. This task involves the semantic mapping of French sentences to AAC pictograms. The data set for this task was constructed from the Traitement de Corpus Oraux en Français (TCOF) corpus [5], which features a wide range of conversations, including arguments, commonplace events, and medical advice among various demographics in French. The translation process involves transforming raw French source sentences into their corresponding target sentences, where each word is reduced to its root form and semantically mapped to the most suitable pictogram.

CLEF 2024: Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France

✉ bhavana2110584@ssn.edu.in (B. Anand); themozhi2110992@ssn.edu.in (T. J); shreyassai2110425@ssn.edu.in (S. S. R); charumathi2110213@ssn.edu.in (C. P); mirnalineett@ssn.edu.in (M. TT)

🌐 <https://www.ssn.edu.in/staff-members/dr-t-t-mirnalinee/> (M. TT)

🆔 0009-0009-8624-9408 (B. Anand); 0009-0009-1608-699X (T. J); 0009000553392153 (S.S. R); 0009-0003-2170-1586 (C. P); 0000-0001-6403-3520 (M. TT)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Section 1 provides an overview of the need for text-to-pictogram translation tools. Section 2 discusses existing research work in this area. Section 3 discusses the provided dataset. Section 4 presents the methodology of the proposed model. Section 5 showcases the results and includes a discussion summarizing the key findings and highlighting potential future research directions.

2. Related Works

In the domain of AAC (Augmentative and Alternative Communication), the translation of text into pictograms is essential to assist individuals with communication impairments. Historically, various technologies have been explored, ranging from rule-based systems to sophisticated neural network approaches, each aimed at enhancing the efficacy and accessibility of AAC solutions [6].

An innovative approach is Sevens et al.'s text-to-pictogram translation system [7] for emails, which leverages lexical-semantic databases to map Dutch text to pictograms using both direct and semantic translation routes to address linguistic challenges. The AraTraductor [8] system employs syntactic analysis to improve the accuracy of pictogram generation from text, showcasing the impact of syntactic parsing on the understandability of AAC content. Further, PictoBERT [9], a BERT derivative, predicts pictogram sequences on AAC boards by adapting transformer architecture to utilize word-sense data, moving beyond traditional n-gram models. The PrAACT [10] methodology adapts large transformer models for AAC, focusing on customization and adaptability, thus improving user-specific communication. Additionally, the BabelDr system [11] integrates the Unified Medical Language System (UMLS) with neural machine translation techniques to convert medical dialogue into pictograms, enhancing patient-doctor communication. This system streamlines complex medical interactions into pictographs by leveraging a neural architecture that parses and translates speech into UMLS-based semantic glosses, significantly improving understanding through intuitive visual representations.

3. Dataset

The dataset for the Text to Pictogram task was built from the Traitement de Corpus Oraux en Français (TCOF) Corpus [5]. Daily life conversations encompassing a wide range of categories between AAC users and their caregivers are present in this dataset.

In the train dataset, for each unique utterance, the dataset consists of a source sentence that is transcribed from speech. Each of these utterances is characterised by a unique ID. The source sentences are converted to target sentences that typically represent the base form of each word in the source corresponding to a sequence of pictogram terms. These words are then mapped to a list of pictogram identifiers linked to each pictogram term. The pictograms are taken from ARASAAC, a collection featuring over 25,000 pictograms. The test dataset contains a series of source sentences that are oral

Tag	Definition	Example
id	unique identifier of each utterance	cefc-toof-Acc_de1_07-1
src	source of the utterance - text from oral transcription	tu peux pas savoir
tgt	target of the utterance - sequence of pictogram terms (tokens)	toi pouvoir savoir non
pictos	a list of pictogram identifiers linked to each pictogram terms (the size is the same as the target output).*	[6625, 35949, 16885, 5526]




Figure 1: Sample Data

transcriptions of utterances and the unique ID corresponding to them. The proposed model is used to generate a hypothesis of target sentences obtained for each of these utterances.

4. Methodology

This paper introduces a unique application of contrastive learning and transformer-based language models to improve the semantic mapping of French sentences to AAC pictograms. This method enhances the capabilities of CamemBERT, improving its adaptability to various linguistic contexts and its learning ability from minimal examples. Our approach addresses the limitations of previous systems by effectively handling complex semantic relationships and adapting to diverse linguistic contexts without the need for extensive training data.

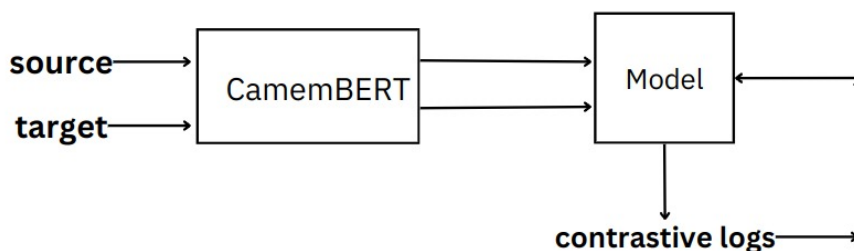


Figure 2: Proposed Model Architecture Overview

4.1. Embeddings

The proposed is anchored in the transformative capabilities of BERT (Bidirectional Encoder Representations from Transformers) [12], a groundbreaking model in natural language understanding. The dataset is tokenized and passed to the CamemBERT model [13] to generate fixed-size embedding vectors (768 for the proposed model). A compressed version of the CamemBERT model, specifically designed for French language processing tasks. Distilled CamemBERT [14] inherits its capabilities from CamemBERT, which is built upon the RoBERTa architecture [15] and trained on a large corpus of French text, achieving state-of-the-art performance in various NLP tasks. The key innovation of BERT lies in its ability to capture complex linguistic patterns and semantics in French text, utilizing bidirectional Transformers, self-attention mechanisms, and dynamic masking techniques that are utilized by the proposed model.

The purpose of distillation is to drastically reduce the complexity of the model while preserving the performance. The training objective of the student model, Distilled CamemBERT, is to closely approximate the behavior of the teacher, CamemBERT model. The training objective is composed of 3 parts, DistillLoss, Cosine Loss, and MLM Loss. The distillLoss measures the similarity between the output probabilities of the student and teacher models using cross-entropy loss on the Masked Language Modeling (MLM) task while the Cosine Loss Ensures alignment between the last hidden layers of the student and teacher models by computing the cosine similarity. The MLM loss implements the MLM task loss to train the student model according to the original task of the teacher model. The objective function is thus a combination of the 3 losses.

$$Loss = 0.5 \times DistilLoss + 0.3 \times CosineLoss + 0.2 \times MLMLoss$$

4.2. Contrastive Learning

These embeddings are then passed on to a Neural Network, which reduces the embedding size to 384. The objective of this Neural Network is to learn a new representation such that it minimizes the difference between similar representations, which in our case are the source and target representations. Contrastive learning techniques are used for this task. Contrastive learning is a paradigm that focuses on learning representations by contrasting positive pairs of similar instances against negative pairs of dissimilar instances. It is based on the principle that semantically similar instances should be closer

together in the learned representation space compared to dissimilar ones. By optimizing a contrastive loss function, the model learns to distinguish between positive and negative pairs, effectively pulling similar instances closer while pushing dissimilar ones apart [16]. This approach has gained significant traction due to its effectiveness in learning robust representations from unlabeled data, particularly for vision and language tasks. Here, a contrastive loss function is used which penalizes larger distances between similar vector embeddings and smaller distances between dissimilar vector embeddings. The loss function is given as

$$Loss = (1 - y) \times d^2 + y \times \max(0, m - d)^2$$

where y is the label for similar and dissimilar pairs, d is the Euclidean distance between the 2 embeddings and m is the margin, which is a constant defined for the minimum distance between dissimilar embeddings. The overall Contrastive Loss is typically calculated as the mean of the individual losses across all pairs in a batch or a dataset. During inference, the generated embeddings are compared with similar ones in the same representation space using cosine similarity to generate the pictograms.

5. Results and Discussion

To evaluate the effectiveness of our translation of text into pictograms, three metrics were utilized: PictoER [17], BLEU [18], and METEOR [19]. Each of these metrics offers a distinct perspective on the quality of the translation, providing a comprehensive assessment. Our system achieved a PictoER score of 141.909, a BLEU score of 3.419, and a METEOR score of 14.351. The breakdown of these individual scores provides valuable insights into the strengths and areas for improvement in our translation approach.

Table 1
Results

pictoer_score	bleu_score	meteor_score
141.909187	3.419165	14.350552

6. Conclusion

In this paper, we present a novel approach for mapping French sentences to corresponding AAC pictograms using a BERT-based model, specifically utilizing CamemBERT embeddings combined with contrastive learning techniques. Our methodology shows significant improvements in handling complex semantic relationships and adapting to diverse linguistic contexts. The effective integration of these techniques has demonstrated the potential to greatly enhance the communication capabilities of AAC users, bridging the gap between text and pictographic representation.

Future research on our Text-to-Pictogram conversion model could take several promising directions to enhance its utility and performance. First, exploring model optimization techniques such as pruning and quantization could improve the efficiency of our transformer-based model without impacting performance, making it more suitable for resource-limited environments. Furthermore, expanding the model to support multiple languages would increase its utility, particularly in diverse linguistic settings. Further development of semantic analysis techniques could improve the model's ability to handle complex sentence structures and ambiguities, enhancing translation accuracy. These improvements would not only extend the model's applicability but also boost its effectiveness in real-world situations.

References

- [1] M. Rowski, R. A. Sevcik, Augmentative communication and early intervention: Myths and realities, *Infants & Young Children* 18 (2005) 174–185.

- [2] Croix-Rouge, Communication alternative améliorée (caa) : la croix-rouge française dévoile sa première étude d'impact social!, 2021.
- [3] C. Macaire, E. Esperança-Rodier, B. Lecouteux, D. Schwab, Overview of 2024 imageclef picto tasks – investigating the translation of natural language into pictograms, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction. CEUR Workshop Proceedings (CEUR-WS.org), Grenoble, France, 2024.
- [4] B. Ionescu, H. Müller, A.-M. Drăgulescu, J. Rückert, A. Ben Abacha, A. García Seco de Herrera, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, C. S. Schmidt, T. M. Pakull, H. Damm, B. Bracke, C. M. Friedrich, A.-G. Andrei, Y. Prokopchuk, D. Karpenka, A. Radzhabov, V. Kovalev, C. Macaire, D. Schwab, B. Lecouteux, E. Esperança-Rodier, W.-w. Yim, Y. Fu, Z. Sun, M. Yetisgen, F. Xia, S. A. Hicks, M. A. Riegler, V. Thambawita, A. Størås, P. Halvorsen, M. Heinrich, J. Kiesel, M. Potthast, B. Stein, Overview of imageclef 2024: Multimedia retrieval in medical applications, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the 15th International Conference of the CLEF Association (CLEF 2024), Springer Lecture Notes in Computer Science LNCS, Grenoble, France, 2024.
- [5] V. André, Canut, Mise à disposition de corpus oraux interactifs : le projet tcof (traitement de corpus oraux en français), *Pratiques. Linguistique, littérature, didactique* (2010) 35–51.
- [6] Cataix-Nègre, Communiquer autrement: Accompagner les personnes avec des troubles de la parole ou du langage : les communications alternatives, De Boeck Supérieur, 2017.
- [7] V. Vandeghinste, I. S. Sevens, F. Van Eynde, Translating text into pictographs, *Natural Language Engineering* 23 (2017) 217–244. doi:10.1017/S135132491500039X, [Online]. Available: <https://doi.org/10.1017/S135132491500039X>.
- [8] S. Bautista, R. Hervás, A. Hernández-Gil, C. Martínez-Díaz, S. Pascua, P. Gervás, Aratractor: text to pictogram translation using natural language processing techniques, in: Proceedings of the XVIII International Conference on Human Computer Interaction (Interacción '17), ACM, New York, NY, USA, 2017. doi:10.1145/3123818.3123825, [Online]. Available: <https://doi.org/10.1145/3123818.3123825>.
- [9] J. A. Pereira, D. Macêdo, C. Zanchettin, A. L. I. de Oliveira, R. d. N. Fidalgo, Pictobert: Transformers for next pictogram prediction, *Expert Systems with Applications* 202 (2022) 117231. doi:10.1016/j.eswa.2022.117231, [Online]. Available: <https://doi.org/10.1016/j.eswa.2022.117231>.
- [10] J. A. Pereira, C. Zanchettin, R. d. N. Fidalgo, Praact: Predictive augmentative and alternative communication with transformers, *Expert Systems with Applications* 240 (2024) 122417. doi:10.1016/j.eswa.2023.122417, [Online]. Available: <https://doi.org/10.1016/j.eswa.2023.122417>.
- [11] J. Mutal, P. Bouillon, M. Norré, J. Gerlach, L. O. Grijalba, A neural machine translation approach to translate text to pictographs in a medical speech translation system - the babeldr use case, in: Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (AMTA 2022), volume 1, Orlando, USA, 2022, pp. 252–263. [Online]. Available: Association for Machine Translation in the Americas.
- [12] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. arXiv:1810.04805.
- [13] L. Martin, B. Muller, P. J. Ortiz Suárez, Y. Dupont, L. Romary, Villemonte de la Clergerie, D. Seddah, B. Sagot, Camembert: a tasty french language model, *CoRR abs/1911.03894* (2019). URL: <http://arxiv.org/abs/1911.03894>. arXiv:1911.03894.
- [14] C. Delestre, A. Amar, Distilcamembert: a distillation of the french model camembert, 2022. arXiv:2205.11111.
- [15] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, 2019. arXiv:1907.11692.
- [16] T. Gao, X. Yao, D. Chen, Simcse: Simple contrastive learning of sentence embeddings, 2022. arXiv:2104.08821.
- [17] J. P. Woodard, J. T. Nelson, An information theoretic measure of speech recognition performance, in: Workshop on standardisation for speech I/O technology, Naval Air Development Center, Warminster, PA, 1982.

- [18] K. Papineni, S. Roukos, T. Ward, W. J. Zhu, Bleu: a method for automatic evaluation of machine translation, in: Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002, pp. 311–318.
- [19] S. Banerjee, A. Lavie, Meteor: An automatic metric for mt evaluation with improved correlation with human judgments, in: Proceedings of the ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization, 2005, pp. 65–72.