

Detecting Aggressiveness in Mexican Spanish Social Media Content by Fine-Tuning Transformer-Based Models

Mircea-Adrian Tanase, George-Eduard Zaharia, Dumitru-Clementin Cercel and Mihai Dascalu

Computer Science Department, University Politehnica of Bucharest, Bucharest, Romania

Abstract

Aggressiveness and several other related problems, such as hate speech, offensive language, or harassment, are experiencing a growing online presence in the context of contemporary social media platforms. The research efforts towards detecting, isolating, and stopping these disturbing behaviors have intensified, in tight relation with the increasing performance of deep learning techniques applied in various Natural Language Processing (NLP) tasks. This study presents our NLP architectures for tackling the problem of aggressiveness detection in the context of the MEX-A3T@IberLEF2020 shared task. We experimented with several pre-trained Transformer-based models, fine-tuned on various combinations of task-specific datasets. Our best model on the MEX-A3T dataset achieves an offensive F1-score of 79.69% on the test dataset, the third in the competition; nevertheless, the difference between the winning solution versus our model is marginal, of only 0.29%. This result argues that Transformer-based models can be successfully used to detect aggressiveness in Mexican Spanish tweets.

Keywords

BETO, XLM-RoBERTa, social media, aggressiveness detection, mexican spanish

1. Introduction

Smart Insights¹ notes that nearly 60% of the world population is online, with more than one third using social media platform, out of which Facebook and Twitter are the most popular alternatives. The massive rise of social media technologies in both personal, business, as well as political communication raised a number of new concerns regarding their misuse because these platforms can also become channels for the proliferation of disturbing trends, such as aggressiveness, harassment, hate speech or cyberbullying².

Social media companies made various attempts towards detecting, removing, and stopping these behaviors, with both Twitter and Facebook rolling out several tools for flagging and

Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020)

EMAIL: mircea.tanase@stud.acs.upb.ro (M. Tanase); george.zaharia0806@stud.acs.upb.ro (G. Zaharia); dumitru.cercel@upb.ro (D. Cercel); mihai.dascalu@upb.ro (M. Dascalu)



© 2020 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/>. Accessed July 6th 2020

²<https://www.stopbullying.gov/cyberbullying/what-is-it>. Accessed July 6th 2020

reporting unwanted pieces of content^{3,4}; however, these efforts encountered several problems. First, a very small percentage of the victims even consider using these tools⁵; thus, they are unaware of the offences towards them. Second, the amount of data that needs to be flagged and analyzed by human moderators is enormous - for example, Internet Live Stats⁶ estimates that 6,000 tweets are posted online every second. These reasons powered the increasing research efforts towards automated processes, grounded in Natural Language Processing (NLP) techniques, of identification and removal of aggressive, offensive, or hateful content in online social media. Nevertheless, creating an annotated corpus of social media content suitable for this work proved to be a very challenging task due to the subjective and fluctuating definitions of the labels [1].

This work presents our approach for the detection track within the MEX-A3T@IberLEF2020 shared task [2], which required a binary classification of aggressiveness in tweets written in the Mexican Spanish dialect. State-of-the-art NLP models were considered and fine-tuned starting from pre-trained Transformer-based architectures on Spanish and English, together with multilingual entries.

The remainder of the paper is structured as follows. A brief analysis on state-of-the-art approaches is performed in section 2, followed by a description of the datasets and details on the methods employed for automated detection of aggressiveness in section 3. Section 4 outlines the evaluation process, while conclusions are drawn in section 5.

2. Related Work

The task of automated detection of online aggressiveness is a necessity for modern social media platforms. Early studies are based on classical machine learning algorithms - for example, Greevy and Smeaton [3] used the Support Vector Machines to detect racist texts in web pages. However, machine learning algorithms evolved in the last decade, with numerous NLP systems being developed and employed for such problems. Cambria et al. [4] used the *sentiment computing* paradigm to detect web trolls. Their approach aims to improve the performance for recognition and interpretation of sentiments and opinions in texts, by employing Semantic Web and Artificial Intelligence techniques. Davidson et al. [5] analyzed and proposed, for the first time, a hate speech detection corpus together with several machine learning experiments with multiple algorithms (e.g., logistic regression or random forest) and pre-processing techniques (e.g., TF-IDF scores, stemming). Malmasi and Zampieri [6] also analyzed this dataset and tried to improve the results using skip-gram features. Gambäck and Sikdar [7] improved the results on the same dataset by employing a Convolutional Neural Network (CNN) [8] as classifier. Also, Zhang et al. [9] analyzed both CNNs and Gated Recurrent Units [10], and achieved better results.

In addition, several shared tasks, surveys, and workshops were published in the last years on the previously mentioned topics. For example, Schmidt and Wiegand [11] presented a survey on non-neural network methods for detecting hate speech. Workshops and shared tasks on the

³<https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>. Accessed July 6th 2020

⁴https://www.facebook.com/communitystandards/hate_speech. Accessed July 6th 2020

⁵<https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/>. Accessed July 6th 2020

⁶<https://www.internetlivestats.com/>. Accessed July 6th 2020

subject include both editions of Abusive Language Online [1, 12], which targeted, among others, the cyberbullying issue. HASOC [13] also addressed the problem of hate speech. Moreover, both the SemEval 2019 Task 5 (OffensEval 2019) [14] and SemEval 2020 Task 12 (OffensEval 2020) [15] shared tasks address offensive language detection in social media. The TRAC-1 [16] and TRAC-2 [17] workshops proposed the problem of aggression detection on Twitter, focusing on the English, Bangla, and Hindi languages; the best performing models for TRAC-2 used Transformer-based architectures, such as BERT [18] or RoBERTa [19].

With a stronger focus on the identification of aggressiveness in the Mexican Spanish social media, we also considered the first two editions of the IberLEF MEX-A3T shared task [20, 21]. Reviewing the research of the teams that obtained the highest scores in the 2019 edition reveals that most approaches used deep learning architectures, such as CNNs or Long Short-term Memory networks [22], with varying degrees of success.

3. Method

3.1. Datasets

The MEX-A3T dataset proposed by the organizers for training consists of 7,332 Mexican Spanish tweets, out of which 2,110 are labeled as positive for aggressiveness. Ten percent of the data was held for validation purposes, while preserving the label distribution. The test set of 3,143 tweets was used to compare the submitted solutions.

Our aim was to automatically increase the size of the training data in order to improve the results of our models. At the same time, we wanted to analyze the impact on the model performance if various training datasets, having different labeling schemes, are combined. Each additional dataset has various particularities (see table 1), such as language specific structures and annotation schemes; thus, a standardization process (i.e., thresholding) was imposed to make them suitable for a binary classification task.

3.2. Baseline

The XGBoost [30] model was considered as baseline. In the text pre-processing phase, the tweets were converted to a lowercase representation and stopwords were removed. Tf-Idf scores were computed, alongside character n-grams with $n = 1,2,3$ as a feature extraction step. Moreover, a grid search was performed to select the best parameters for the XGBoost classifier.

3.3. BERT

Our models are based on the architecture proposed by Devlin et al. [18], namely the Bidirectional Encoder Representations from Transformers (BERT). BERT is a deep neural network designed for various NLP tasks, leveraging the power of both the WordPiece [31] embeddings and the Transformers. The architecture is pre-trained using large language corpora designed for a generic task, such as next sentence prediction, and is usually fine-tuned on a downstream task, using a specific corpus. A distinct feature and major advantage of BERT is the unified

architecture across tasks, as there are minimal differences (weights) between the pre-trained architecture and the final architecture used for a particular task.

Table 1

Considered additional datasets.

Dataset	Description
OLID [23]	13,240 Twitter samples binary labeled for offensive language detection, with a 33.23% positive label rate.
OffensEval [15]	<p>A collection of offensive language data for five different languages covering the following languages:</p> <ul style="list-style-type: none"> • English (SOLID) [24], consists of over nine million English tweets annotated in a semi-supervised manner, using OLID as the starting point. By applying the previously mention thresholding process, we obtained a 12.58% positive label rate. • Arabic [25], including 7,000 Twitter samples, binary labels, 19.58% positive rate. • Danish [26], with 3,000 Twitter, Reddit, and Facebook samples, binary labels, 12.80% positive rate. • Greek [27], containing 7,843 tweets, binary labels, 28.43% positive rate. • Turkish [28], having 31,277 tweets, binary labeled, 19.33% positive rate.
HASOC [13]	A binary annotated corpus for both hate speech and offensive language identification, including subsets for three languages (i.e., English, German, and Hindi), with 5,852 (38.63% positive rate), 3,819 (10.65% positive rate) and 4,665 (47.07% positive rate) tweets, respectively.
Davidson et al. [5]	24,783 tweets annotated using three classes: a) hate speech, b) profanity, but not hate speech, and c) none. We consider the former two positive and the latter as negative to standardize labels; 16.09% positive label rate was achieved after standardization.
HatEval [14]	Two subsets, one with 10,000 English tweets (42.36% positive) and the other with 5,000 Spanish tweets (41.58% positive), binary annotated for hate speech detection.
SIMAH [29]	6,374 binary labeled Twitter samples proposed for the SIMAH 2019 competition.

3.4. Multilingual BERT (mBERT)

mBERT [32] is a model pre-trained by Google using a multilingual corpus⁷. Pre-training considered sample texts belonging to over 100 languages from different Wikipedia entries. This leads to better performance for highly represented languages (such as English, as well as the Spanish). The mBERT model was finetuned using the following combined datasets:

- The MEX-A3T training set only.
- The MEX-A3T training set along with all the OffenseEval data, for observing the influence on the performance of the test set by adding all data from OffenseEval (labeled using the same annotation techniques).
- All the available non-English data - this allows us to assess the impact of adding other non-English data to the training set.
- All the available data - this is a separate experiment from the previous one because the English set is significantly larger and English is the most represented language in the embedding layers of mBERT.

3.5. XLM-RoBERTa

Liu et al. [19] proved that BERT is under-trained and proposed a robust BERT pretraining approach which exceeds the results obtained by Devlin et al. [18] on several NLP tasks. Conneau et al. [33] designed XLM-RoBERTa, a cross-lingual masked language model, and pretrained it on a large corpus containing over 100 languages by applying the same technique as for mBERT. The resulting model significantly outperformed mBERT, while also obtaining good results for low-represented languages. We repeated all the experiments presented in Section 3.4 using the XLM-RoBERTa model instead of mBERT.

3.6. BETO

Canete et al. [34] introduced BETO, a pre-trained BERT model for Spanish. The pre-training corpus includes, aside from the entire Spanish Wikipedia content, all sources from the Spanish OPUS project [35]. Two experiments were performed using this model:

- Fine-tuning using the MEX-A3T training set only.
- Fine-tuning using the MEX-A3T training set merged with the HatEval Spanish subset to observe the effect on the performances of the model by adding a labeled dataset for slightly different tasks.

⁷<https://github.com/google-research/bert/blob/master/multilingual.md>

3.7. Dataset Translation

We also experimented with automatically translating the entire MEX-A3T training dataset in English through the Yandex translation API⁸. We chose English by considering the multitude of available texts and corpora used for the pre-training Transformer-based models. We then used the translated training set for fine-tuning the English pre-trained BERT. Our intuition is that the pre-trained BERT model should perform better, given that the semantics and the syntax of the translated entries are properly preserved. However, the preservation task is non-trivial, as the two languages (i.e., English and Spanish) have noticeable structural differences, and a translation process can seriously alter the original idea included in the source entry. Most translation engines suffer from this issue, inasmuch as that they are not capable to accurately transfer the particularities of the source language into the target one.

4. Results

Our aim was to integrate as much as possible the data presented in section 3.1, although most datasets contain other languages, and are labeled for related tasks in the offensive language detection field. All experiments were performed using an Ubuntu server machine with 4 Intel CPU Cores, one Nvidia RTX2080 Ti GPU and 64 GB RAM. The implementations from the Transformer Python package [36] was used for the BERT-based models. The Adam optimizer [37] with a learning rate of $2e-5$ was used for all the BERT and XLM-RoBERTa experiments.

The results (i.e., accuracy, precision, recall, and offensive F1-score) of each fine-tuned model presented in Section 3 and evaluated on the MEX-A3T validation set are summarized in Table 2.

Table 2

Results obtained on the MEX-A3T aggressiveness validation set.

Model Architecture	Pre-training Language	Pre-processing	Fine-Tuning Dataset	Acc (%)	P (%)	R (%)	F1 (%)
XGBoost	-	n-gram TF-IDF	MEX-A3T train set	58.61	42.12	34.82	40.61
BERT	English	Translation	All English data	68.42	54.18	48.41	51.07
BETO	Spanish	-	MEX-A3T train set	84.32	82.02	80.26	81.13
	Spanish	-	MEX-A3T train set + HatEval Spanish	85.24	81.90	81.22	81.55
BERT	Multi	-	MEX-A3T train set	72.44	69.20	64.81	66.93
	Multi	-	All non-English data	74.31	69.14	68.35	68.74
	Multi	-	All OffensEval data	76.88	72.14	69.83	70.96
	Multi	-	All data	76.27	71.92	70.18	71.03
XLM-RoBERTa	Multi	-	MEX-A3T train set	73.80	70.34	65.45	67.80
	Multi	-	All non-English data	77.22	72.62	68.91	70.71
	Multi	-	All OffensEval data	83.48	80.47	78.57	79.50
	Multi	-	All data	83.42	80.42	78.96	79.68

As expected, fine-tuning the Transformer-based models surpasses the classical machine

⁸<https://tech.yandex.com/translate/>

learning baseline, whereas XLM-RoBERTa outperforms mBERT. Furthermore, adding more data to the fine-tuning dataset improves the results, even if the added data is labeled for a different, but related task, and contains other languages.

Our experiment with automated translation of Spanish samples to English and using them to fine-tune an English pre-trained BERT was a failed attempt. This is most likely caused by the poor quality of the automatic translation applied on short messages. After inspecting the translation, words are mainly translated correctly, but the syntax and semantics of the tweet were destroyed.

The best results were obtained using the BETO models, thus underlining that, when available, a language-specific pre-trained model scores better than a multilingual one. The best performing model was fine-tuned using both the MEX-A3T training dataset and the HatEval Spanish subset, proving once again that adding the hate speech corpus was beneficial. This last model was used for predicting the labels of the competition test set, obtaining scores of F1=79.69%, P=84.40%, R=86.68%, and Acc=87.59% on the dedicated test set for aggressiveness detection. The previous results ranked third out of 19 submissions, with the leader achieving an offensive F1-score of 79.98%.

5. Conclusions

This work presented our approach to the shared task of automated aggressiveness detection in Spanish social media samples organized at MEX-A3T 2020. We experimented with fine-tuning pre-trained Transformer-based models, solutions that achieved state-of-the-art results in multiple NLP tasks.

The obtained scores in the validation phase for the the MEX-A3T 2020 competition prove that the aforementioned method can be successfully applied for the current task, using both multilingual and Spanish pre-trained models. Furthermore, several combinations of datasets annotated for various related tasks (i.e., hate speech, offensive language, and harassment) were included for fine-tuning. We discovered that the performance of the models can be improved by adding more data, even if it is labeled for a slightly different task.

Future development directions include exploring other related datasets for offensive language, aggressiveness, and harassment detection fields. We will also consider potential techniques of pre-processing tweets, including the expansion of mentioned hashtags with corresponding details. In addition, advanced error analysis techniques, such as feature importance or model explainability, could be used to improve the model's performance.

References

- [1] Z. Waseem, T. Davidson, D. Warmusley, I. Weber, Understanding abuse: A typology of abusive language detection subtasks, in: Proceedings of the First Workshop on Abusive Language Online, 2017, pp. 78–84.
- [2] M. E. Aragón, H. Jarquín, M. Montes-y Gómez, H. J. Escalante, L. Villaseñor-Pineda, H. Gómez-Adorno, G. Bel-Enguix, J.-P. Posadas-Durán, Overview of mex-a3t at iberlefe 2020: Fake news and aggressiveness analysis in mexican spanish, in: Notebook Papers of

- 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain, September, 2020.
- [3] E. Greevy, A. F. Smeaton, Classifying racist texts using a support vector machine, in: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, 2004, pp. 468–469.
 - [4] E. Cambria, P. Chandra, A. Sharma, A. Hussain, Do not feel the trolls, ISWC, Shanghai (2010).
 - [5] T. Davidson, D. Warmesley, M. Macy, I. Weber, Automated hate speech detection and the problem of offensive language, in: Eleventh international aaai conference on web and social media, 2017.
 - [6] S. Malmasi, M. Zampieri, Challenges in discriminating profanity from hate speech, Journal of Experimental & Theoretical Artificial Intelligence 30 (2018) 187–202.
 - [7] B. Gambäck, U. K. Sikdar, Using convolutional neural networks to classify hate-speech, in: Proceedings of the first workshop on abusive language online, 2017, pp. 85–90.
 - [8] Y. Kim, Convolutional neural networks for sentence classification, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1746–1751. URL: <https://www.aclweb.org/anthology/D14-1181>. doi:10.3115/v1/D14-1181.
 - [9] Z. Zhang, D. Robinson, J. Tepper, Detecting hate speech on twitter using a convolution-gru based deep neural network, in: European semantic web conference, Springer, 2018, pp. 745–760.
 - [10] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv:1412.3555 (2014).
 - [11] A. Schmidt, M. Wiegand, A survey on hate speech detection using natural language processing, in: Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, 2017, pp. 1–10.
 - [12] D. Fišer, R. Huang, V. Prabhakaran, R. Voigt, Z. Waseem, J. Wernimont, Proceedings of the 2nd workshop on abusive language online (alw2), in: Proceedings of the 2nd Workshop on Abusive Language Online (ALW2), 2018.
 - [13] T. Mandl, S. Modha, P. Majumder, D. Patel, M. Dave, C. Mandlia, A. Patel, Overview of the hasoc track at fire 2019: Hate speech and offensive content identification in indo-european languages, in: Proceedings of the 11th Forum for Information Retrieval Evaluation, 2019, pp. 14–17.
 - [14] V. Basile, C. Bosco, E. Fersini, D. Nozza, V. Patti, F. M. R. Pardo, P. Rosso, M. Sanguinetti, Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter, in: Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 54–63.
 - [15] M. Zampieri, P. Nakov, S. Rosenthal, P. Atanasova, G. Karadzhov, H. Mubarak, L. Derczynski, Z. Pitenis, c. Çöltekin, SemEval-2020 Task 12: Multilingual Offensive Language Identification in Social Media (OffenseEval 2020), in: Proceedings of SemEval, 2020.
 - [16] R. Kumar, A. K. Ojha, S. Malmasi, M. Zampieri, Benchmarking aggression identification in social media, in: Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), 2018, pp. 1–11.
 - [17] S. Bhattacharya, S. Singh, R. Kumar, A. Bansal, A. Bhagat, Y. Dawer, B. Lahiri, A. K. Ojha,

- Developing a multilingual annotated corpus of misogyny and aggression, in: Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying, European Language Resources Association (ELRA), Marseille, France, 2020, pp. 158–168.
- [18] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 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), 2019, pp. 4171–4186.
- [19] 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, arXiv preprint arXiv:1907.11692 (2019).
- [20] M. Á. Álvarez-Carmona, E. Guzmán-Falcón, M. Montes-y Gómez, H. J. Escalante, L. Villaseñor-Pineda, V. Reyes-Meza, A. Rico-Sulayes, Overview of mex-a3t at ibereval 2018: Authorship and aggressiveness analysis in mexican spanish tweets, in: Notebook Papers of 3rd SEPLN Workshop on Evaluation of Human Language Technologies for Iberian Languages (IBEREVAL), Seville, Spain, volume 6, 2018.
- [21] M. E. Aragón, M. Á. Á. Carmona, M. Montes-y Gómez, H. J. Escalante, L. V. Pineda, D. Moctezuma, Overview of mex-a3t at iberlef 2019: Authorship and aggressiveness analysis in mexican spanish tweets., in: IberLEF@ SEPLN, 2019, pp. 478–494.
- [22] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural computation* 9 (1997) 1735–1780.
- [23] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, R. Kumar, Predicting the type and target of offensive posts in social media, 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), 2019, pp. 1415–1420.
- [24] S. Rosenthal, P. Atanasova, G. Karadzhov, M. Zampieri, P. Nakov, A Large-Scale Semi-Supervised Dataset for Offensive Language Identification, in: arxiv, 2020.
- [25] H. Mubarak, A. Rashed, K. Darwish, Y. Samih, A. Abdelali, Arabic offensive language on twitter: Analysis and experiments, arXiv preprint arXiv:2004.02192 (2020).
- [26] G. I. Sigurbergsson, L. Derczynski, Offensive Language and Hate Speech Detection for Danish, in: Proceedings of the 12th Language Resources and Evaluation Conference, ELRA, 2020.
- [27] Z. Pitenis, M. Zampieri, T. Ranasinghe, Offensive Language Identification in Greek, in: Proceedings of the 12th Language Resources and Evaluation Conference, ELRA, 2020.
- [28] c. Çöltekin, A Corpus of Turkish Offensive Language on Social Media, in: Proceedings of the 12th International Conference on Language Resources and Evaluation, ELRA, 2020.
- [29] S. Sharifirad, S. Matwin, When a tweet is actually sexist. a more comprehensive classification of different online harassment categories and the challenges in nlp, arXiv preprint arXiv:1902.10584 (2019).
- [30] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.
- [31] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al., Google’s neural machine translation system: Bridging the gap between human and machine translation, arXiv preprint arXiv:1609.08144 (2016).

- [32] T. Pires, E. Schlinger, D. Garrette, How multilingual is multilingual bert?, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 4996–5001.
- [33] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, arXiv preprint arXiv:1911.02116 (2019).
- [34] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, in: Practical ML for Developing Countries Workshop@ ICLR 2020, 2020.
- [35] J. Tiedemann, Parallel data, tools and interfaces in opus., in: LREC, volume 2012, 2012, pp. 2214–2218.
- [36] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Transformers: State-of-the-art natural language processing, arXiv preprint arXiv:1910.03771 (2019).
- [37] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).