Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Propaganda Detection Techniques for Social Media Texts using Lifelong Learning

Thrupthi N[1], Shreya Kishor[2], Shruti Karande[3], Vankadara Neha[4], and Bhaskarjyoti Das[5]
[1] *Department of Computer Science and Engineering, PES University, Bengaluru, Karnataka, India*
[2] *Department of Computer Science and Engineering in AI & ML, PES University, Bengaluru, Karnataka, India*
thrupthi2804@gmail.com,shreyakishor2002@gmail.com,shrutikarande27@gmail.com,
neha.vankadara@gmail.com
bhaskarjyoti01@gmail.com

ARTICLE INFO

Received: 24 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Most people misuse social media to influence the thoughts of large masses. The use of propaganda mainly targets to play with the emotions of people and communities and thus influence them easily. In today's society, where social media plays a crucial role in society, it becomes necessary to detect such propaganda in social media texts like tweets, online articles, Facebook posts, and advertisements to protect people from being easily misguided. Our work focuses on detecting the propaganda techniques employed in the social media texts of various domains. To achieve this, we use the approach of Lifelong machine learning. When the model is trained sequentially on different domains, most paradigm ML algorithms undergo catastrophic forgetting of previous domains when the model encounters a new domain. Our model overcomes this drawback by continuously learning the new propaganda domain without catastrophically forgetting the previous domain when a new domain is introduced. We use the Elastic Weight Consolidation(EWC) regularization technique to prevent catastrophic forgetting of previous domains when the model is trained on the new domain and to ensure good performances across all domains.

Keywords: *Propaganda, Propaganda techniques detection, Lifelong Ma- chine Learning, Elastic Weight Consolidation.*

1 INTRODUCTION

As the news articles were popularly used before, the research was focused on propaganda detection in the news articles initially. But with the introduction of social media, nowadays, most people seldom read news articles due to a shift to the online medium. Utilizing this opportunity, most social media influencers and political leaders spread propaganda through present-day popular platforms such as social media to influence huge masses. It is critical to detect propaganda in social media to protect the fundamental rights of the people and protect them from being influenced. Thus, the state-of-the-art research in propaganda mainly concentrates on social media texts[2].

Some papers achieved the binary classification of social media tweets or articles as propagandistic or non-propagandistic[3]. A few others identified propaganda techniques employed in the texts [4]. Identifying propaganda techniques makes the work more meaningful. Besides, multiple propaganda techniques may be used in the text to influence people. Therefore, our work focuses on identifying multiple propaganda techniques employed in social media texts.

There are 18 commonly used propaganda techniques which include Appeal to Authority, Appeal to Fear/Anxiety, Bandwagon, Causal Oversimplification, Exaggeration/Minimisation, Loaded Language, Name Calling, Strawman, Repetition, Slogans, Whataboutism, Doubt, Confusion, Red Herring, Black-And-White Fallacy, Thought-terminating Cliche, Flag Waving and Reductio-ad-Hitlerum [4]. Some of the published papers considered only 14 propaganda techniques[5] where they have combined some of the similar techniques into a single technique. Whereas 22 propaganda techniques are also used in some papers[6]. However, in this paper, we have considered 18 commonly used propaganda techniques.

Generally, when the paradigm machine learning model is trained sequentially on different domains, it undergoes catastrophic forgetting of domains trained previously. To overcome the forgetting[7], Lifelong learning techniques are essential[8]. Plasticity and Elasticity are associated with lifelong learning. The plasticity ensures the knowledge preservation of previous domains. While the elasticity determines the flexibility of the model to adapt to the new domain. Establishing a balance between plasticity and elasticity results in uniform performance across all the domains. To the best of our knowledge, lifelong machine learning has not yet been implemented in propaganda research. Thus, our work tries to fill this gap. Our work focuses on using Lifelong learning with Elastic Weight Consolidation(EWC), preserving the knowledge of previous and new domains, and ensuring good performances across different propaganda datasets. The EWC approach helped the model to overcome the forgetting of previous knowledge and achieve lifelong learning on multiple propaganda techniques detection used in the social media texts.

1 LITERATURE SURVEY

Propaganda detection: Initially, most of the propaganda research was focused on news articles [9]. The propaganda detection in articles is done at the sentence level, document level, and fragment level. Nowadays, this research is more focused on social media texts such as tweets[10] as social media is a huge platform that has become vulnerable to influence many people. Propaganda detection is done in various ways. It includes the classification of texts as propaganda or not[11], multi-label, multi-class propaganda techniques classification and span detection[12]. Span detection involves identifying the propaganda phrases used in the text. The paper [13] focuses on span identification and propaganda techniques detection in the SemEval articles. It throws light on the various results obtained during the competition. The paper [14] detects the propaganda in the online articles and also identifies the propagandistic phrases and sentences. It also provides an option to get to know the extent of propaganda employed online in the particular topic. The detection of propaganda techniques used in the texts would be more meaningful and provide better insights. Thus, our paper focuses on multi-label, multi-class propaganda techniques, and the classification of social media texts. A Few papers[15] achieved domain adaptation across news articles, tweets, and speeches. The paper focused on domain adaptation from news articles(source) to tweets(target) using a pivot-based language model. The paper [15] achieved cross-domain learning across news articles, tweets, and speeches using a pairwise propaganda model. It learned pairwise with propaganda and non-propaganda texts in parallel. Additionally, the research is also focused on network-based analysis. The network-based analysis adds more context to the text. Additionally, the analysis provides better insights into the spread of propaganda[16]. Recently, the papers also have achieved propaganda detection in various languages. The paper [17] has achieved the cross-lingual propaganda detection between English and Chinese texts. Whereas our work focuses on English texts alone.

Lifelong Learning: When the model is trained sequentially on different domains/tasks, it undergoes 'catastrophic forgetting' of previous domains . In order to alleviate forgetting, lifelong learning techniques are used. Sequential learning without forgetting can be achieved using memory-based approaches[18], regularization-based approaches [19], and architecture-based approaches [20]. The effectiveness of the Elastic Weight Consolidation(EWC) regularization technique to overcome catastrophic forgetting is evident in many of the previous works[21]. The paper [2] has achieved sequential domain adaptation on sentiment analysis of reviews of various domains using Elastic Weight Regularization. Generalization of the model on unseen target domain using EWC is also achieved . The paper [22] uses EWC along with prominent state-of-the-art models like large language models and improved generalization. Thus, it shows that EWC helps to achieve generalization. The paper [22] also compared the performances with lifelong learning methods such as EWC, Weight Initialization, Combined Training, Incremental Moment Matching (IMM), and Hard Attention to Task (HAT). IMM and HAT along with EWC are used to overcome catastrophic forgetting in lifelong learning. EWC's continual learning approach performed better than other lifelong learning techniques.

Therefore, we consider this approach of Elastic Weight Consolidation to balance the model's plasticity and elasticity to keep up good propaganda detection across all the domains. To the best of our knowledge, the lifelong learning has not yet been tested in the propaganda research. Our work fills this gap.

2 DATASET

In this paper, to achieve lifelong learning of propaganda techniques detection environment on social media texts of various domains, 3 propaganda datasets are used : (1) CAA-Protest [3], (2) Covid-19 Disinformation dataset [7] and (3) Presidential [1]. CAA-Protest dataset contains weakly supervised labels. Whereas we manually annotated Presidential and COVID-19 disinformation datasets. We down sampled an equal amount of 3000 instances from each dataset.

Each in- stance contains 19 labels as considered in [14]. It includes 18 propaganda techniques and '-1' label. The binary labels (0or1) for each propaganda technique are annotated for each instance. '-1' label indicates non-propagandistic text.

The CAA (Citizen-ship Amendment Act)-Protest dataset contained the tweets that were spread during the CAA-Protest which happened in India in 2020. It contained 68462 tweets. As the source dataset contains weakly supervised labels, we have manually corrected 3000 samples from the dataset. These manually corrected 3000 CAA-Protest tweets are used in our work.

The tweets dataset developed by [1] is used as a Presidential dataset in our work. They combined the Twitter IRA corpus and twitter7 data from SNAP. The tweets dataset contained 8963 tweets related to Presidential elections.

Covid-19 disinformation dataset [7] contains the disinformation texts that were spread during the Covid-19 pandemic period. It contained 27049 texts in total. During this period, conspiracy theories were spread online and influenced people easily. The disinformation during this period had a huge impact on the health and beliefs of people. It misled people and created a distortion in their health and beliefs. Therefore, it is critical to determine such propaganda used in social media during a pandemic breakout where there are high chances of people getting misled.

It becomes vital to detect the presence of propaganda online in domains like protest, elections and health-related texts. Therefore, we have considered these datasets in this paper.

3 METHODOLOGY

The multilabel multiclass propaganda techniques classification model is trained sequentially on different tasks. When learning a new task, ML algorithms undergoes forgetting of previous knowledge. In order to overcome this forgetting, Elastic Weight Consolidation (EWC) is used here. With EWC, our model aims to achieve good performances of propaganda techniques detection across all tasks. In this section, our approach and methodologies used to achieve lifelong learning of various tasks are explained in detail.

3.1 Elastic Weight Consolidation

Elastic Weight Consolidation (EWC) is one of the regularization techniques of lifelong learning approaches, which is used to overcome catastrophic forgetting

[3] CAA-Protest source: https://doi.org/10.5281/zenodo.7797035

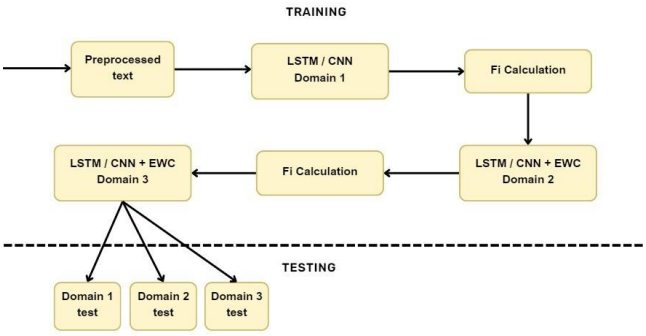

Fig. 1. Lifelong learning environment of propaganda

of old domain data when new domain is introduced. The regularization EWC term is added to the current domain loss to control the forgetting of previous domains. The EWC term for lesser important parameters of previous domains will be less. It does not alter the new loss to a greater extent and thus those parameters are flexible for the new data. On the other hand, the parameters which are important for the previous training data increases the total loss value. Hence these parameters cannot be changed easily for the new data because of high loss value generated when trying to modify those parameters which are important for previous tasks. Thus this prevents the catastrophic forgetting and preserves knowledge of previous domains. The following loss function helps to overcome catastrophic forgetting:

$$TotalLoss = CurrentDomainLoss + EWCterm \quad (1)$$

$$L'(\theta) = L_B(\theta) + \sum_i \frac{\lambda}{2} F_i(\theta_i - \theta_A^*, i)^2 \quad (2)$$

According to the stated loss equation 1,2, the EWC term in it considers the square of difference between the present value and previous value of the parameter. It is weighted by the fisher information value and penalty factor. Fisher information value (Fi) determines the importance of the particular parameter for the previous domain. Fi value for each parameter is calculated by averaging the gradient values for each parameter. If Fi value is higher, it gives a higher weight for that particular parameter.

Additionally, the value of the penalty factor λ determines the amount of importance given to the previous domain as a whole when the new domain is seen. The higher penalty factor makes the model perform well in the previous domain.

3.2 Lifelong Learning Environment with EWC

The Figure 1 illustrates the approach of lifelong learning of propaganda techniques classification of 3 domains considered in our work.

Different domains/tasks considered:

- TASK1: Propaganda detection in CAA protest dataset
- TASK2: Propaganda detection in Covid-19 dataset
- TASK3: Propaganda detection in Presidential dataset

According to the Figure 1, the classification model (LSTM or CNN) is trained on the preprocessed texts of first source domain S1 without EWC. When the model needs to be trained on new domain, Fisher information values for the previously trained model are calculated which are required to calculate the EWC term. Further, when the new domain data is introduced, the loss will also include the EWC term for the previous model's parameters such that catastrophic forgetting is alleviated, as explained in section 4.1.

Evaluation:

For the validation of the model, 5-folded cross-validation is employed. After the classification model is sequentially trained on 3 tasks, the metrics like micro- averaged Precision, Recall, F1-score values are calculated for test sets of each task. Average values across all tasks are also calculated indicating the generalized performance of the model across all tasks. Finally, these metrics are compared between sequential learning with EWC and sequential learning without EWC, showing the effectiveness of EWC in overcoming the catastrophic forgetting in a sequential learning environment and in achieving uniform performances across all domains.

4 RESULTS AND DISCUSSION

Our work aimed at addressing the challenge of propagandistic content detection and classification in social media texts through Lifelong learning, employing the Elastic Weight Consolidation(EWC) regularization technique. The two baseline models used are CNN and LSTM. The propaganda classification model is trained on the 3 domains sequentially. Finally, after training is done on the 3rd domain, we test the model on each of the trained domains. Different hyperparameters and various sequential training orders of tasks are experimented. The different training orders with suitable penalty factors λ does not result in significant differences in the results. However, in this section, the results of the sequential order of task1, task2, and task3 are discussed. After training sequentially on task1, task2, and task3, the best micro-averaged Precision, Recall, and F1-score values of the classification model across 3 tasks are compared between 2 environmental setups:

Table 1. Comparison of CNN with & without EWC

	Task	Avg Precision	Avg Recall	Avg F1-Score
Without	Task 1	0.5129	0.3324	0.4017
	Task 2	0.3829	0.4864	0.4271

EWC	Task 3	0.8140	0.6383	0.7068
	Total	0.5699	0.4857	0.5119
	Task 1	0.6833	0.5893	0.6307
With	Task 2	0.5823	0.7437	0.6488
EWC	Task 3	0.7862	0.5313	0.6311
	Total	0.6840	0.6214	0.6368

Table 2. Comparison of LSTM with & without EWC

	Task	Avg Precision	Avg Recall	Avg F1-Score
	Task 1	0.5334	0.3530	0.4247
Without	Task 2	0.5474	0.6905	0.6106
EWC	Task 3	0.8890	0.7715	0.8163
	Total	0.6566	0.6050	0.6172
	Task 1	0.7610	0.5302	0.6246
With	Task 2	0.5923	0.6677	0.6271
EWC	Task 3	0.8146	0.5261	0.6373
	Total	0.7226	0.5747	0.6297

4.1 Sequential learning without Elastic Weight Consolidation

– In the absence of EWC, LSTM & CNN demonstrated low performances on the previously trained domains i.e. task1(domain 1) and task2(domain 2).

– Whereas, sequential Learning without EWC resulted in high precision, Re- call, and F1-Score values on the new domain i.e. task3(domain 3)as shown in Tables 1, 2.

– The blue lines in graphs 2 and 3 indicate that the performances across all the tasks are not uniform after sequentially training the model without EWC. This can be explained by catastrophic forgetting, wherein, as task 3 is introduced, the model shows low performances on task 1 and task 2 that it was trained on initially.

4.2 Sequential learning with Elastic Weight Consolidation

– As shown in the Table 1, the CNN without EWC exhibited the average F1-score values of 0.4, 0.42, 0.7 for task1, task2, task3 respectively. With CNN+EWC, the F1-score values are 0.63, 0.648, 0.63.

– As shown in Table 2, the LSTM without EWC exhibited the average F1-score values of 0.42, 0.61, and 0.81 for task1, task2, and task3, respectively. With LSTM+EWC, the F1-score values are 0.624, 0.627, and 0.637.

– The F1-scores of previously trained domains improved from 0.4 of task1 and 0.42 of task2 with sequential learning using CNN alone to 0.63 of task1

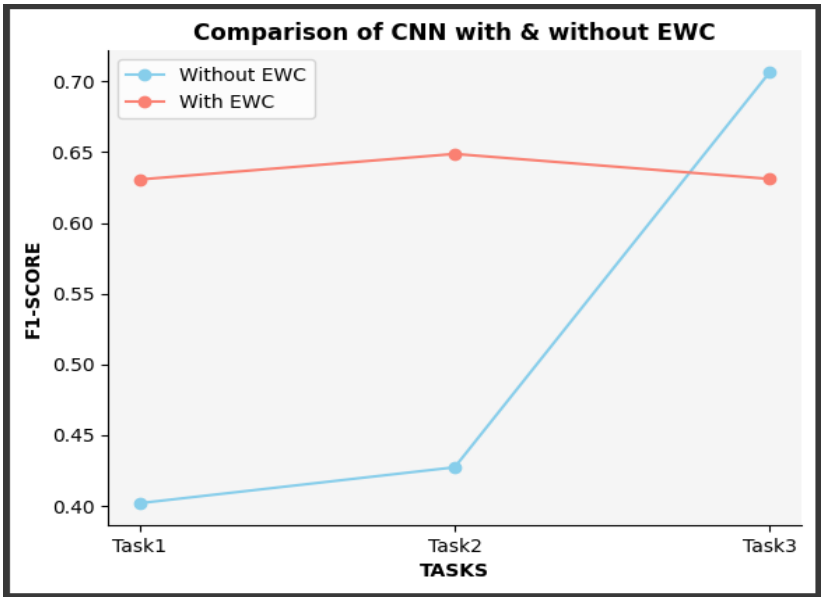

Fig. 2. Performance of CNN with & without EWC

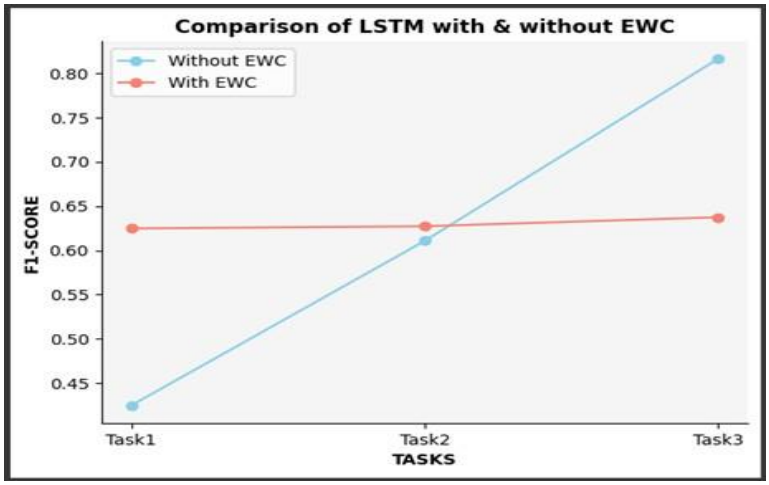

Fig. 3. Performance of LSTM with & without EWC

and 0.648 of task2 with sequential learning using CNN+EWC. Similarly,

0.42 of task1 and 0.61 of task2 with sequential learning using LSTM alone rises to 0.624 of task1 and 0.627 of task2 with sequential learning using LSTM+EWC. Thus, EWC played a vital role in overcoming catastrophic forgetting of previous domains i.e task1 and task2.

– Although it resulted in a decrease in the performance on the new domain

i.e. task3, all the precision, recall,F1-score values of each task are above 0.5. Also, the average F1-scores across each task are above 0.6.

– The orange lines in Figure 2,3 are nearly horizontal, unlike the blue lines. It shows that EWC helped to achieve uniform performances across all tasks and avoided overfitting on a single domain.

– The average F1-scores for each task is almost same for CNN+EWC and LSTM+EWC. There is no significant difference in the performance.

Model Predictions:

Propaganda detection across all tasks is tested with classification-model+EWC. The model's predictions on the text from each task are stated in Table 3. Table 3 compares the labels generated by our model with the expected labels of social media texts of various tasks.

Table 3. Model Prediction of Propaganda Techniques

Dataset	Online Text	Predicted Labels	Expected Labels
CAA protest	'Can you explain how is it dangerous? It is disruptive, absolutely, whole world got to know who are mindless, and what can happen to India if we give government to someone else than Modiji. CAA is making India strong by its core..'	'Appeal to Fear', 'Exaggeration Minimization', 'Flag Waving', 'Loaded Language'	'Flag Waving', 'Appeal to Fear', Loaded Language', 'Name-Calling', 'Exaggeration Minimisation", 'Loaded Language'
Covid-19	'Says a 2010 Rockefeller Foundation report shows the COVID-19 pandemic and the international response were meticulously planned at least 10 years ago'	'Exaggeration Minimization', 'Loaded Language'	'Exaggeration Minimization', 'Loaded Language', 'Appeal to Authority'
Presidential	'Rumors R Reporters Are Worried for America amp; THEIR Conscience May Over Take THEIR Orders to COVERUP for Clintons'	'Appeal to Fear', 'Flag Waving', 'Loaded Language', 'Name-Calling', 'Slogans'	'Appeal to Fear', 'Flag Waving', 'Loaded Language', 'Name-Calling', 'Slogans'

From the results stated in this section, our model with EWC thus helped to overcome catastrophic forgetting of the previous knowledge when the model is trained on new data and also achieved uniform performances across all tasks.

5 CONCLUSION AND FUTURE WORK

The lifelong learning of propaganda techniques detection in social media texts of various datasets is achieved by using Elastic Weight Consolidation regularization in our work. The Elastic Weight Consolidation helped to overcome catastrophic forgetting. The effectiveness of EWC in overcoming catastrophic forgetting is demonstrated by comparing the results between sequential learning with EWC and sequential learning without EWC. This work also uses EWC in such a way that the model achieves uniform performances across all domains. But in this process of overcoming catastrophic forgetting with the help of EWC, the performance is reduced on the new domain. Future work can be directed towards improving the performance of the new domain as well.

REFERENCES

[1] Wang, Liqiang, Xiaoyu Shen, Gerard de Melo, and Gerhard Weikum. 'Cross- domain learning for classifying propaganda in online contents.' arXiv preprint arXiv:2011.06844 (2020).

[2] Madasu, Avinash, and Anvesh Rao Vijjini. 'Sequential domain adaptation through elastic weight consolidation for sentiment analysis.' In 2020 25th International Con- ference on Pattern Recognition (ICPR), pp. 4879-4886. IEEE, 2021.

[3] Lv, Guangyi, Shuai Wang, Bing Liu, Enhong Chen, and Kun Zhang. 'Sentiment classification by leveraging the shared knowledge from a sequence of domains.' In Database Systems for Advanced Applications: 24th International Conference, DAS- FAA 2019, Chiang Mai, Thailand, April 22–25, 2019, Proceedings, Part I 24, pp. 795-811. Springer International Publishing, 2019.

[4] Thota, Mamatha, Dewei Yi, and Georgios Leontidis. 'LLEDA—Lifelong Self- Supervised Domain Adaptation.' Knowledge-Based Systems 279 (2023): 110959.
[5] Rajmohan, Malavikka, Rohan Kamath, Akanksha P. Reddy, and Bhaskarjyoti Das. 'Emotion Enhanced Domain Adaptation for Propaganda Detection in Indian Social Media.' In Innovations in Computational Intelligence and Computer Vision: Pro- ceedings of ICICV 2021, pp. 273-282. Singapore: Springer Nature Singapore, 2022.
[6] Martino, G., Alberto Barr´on-Cedeno, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 'SemEval-2020 task 11: Detection of propaganda techniques in news articles.' arXiv preprint arXiv:2009.02696 (2020).
[7] Machova, Krist´ına, Marian Mach, and Michal Porezany´. 'Deep Learning in the Detection of Disinformation about COVID-19 in Online Space.' Sensors 22, no. 23 (2022): 9319.
[8] Lu, Pengyuan, Seungwon Lee, Amanda Watson, David Kent, Insup Lee, ERIC EATON, and James Weimer. 'Mako: Semi-supervised continual learning with min- imal labeled data via data programming.' (2021).
[9] Zhang, Wenshan, and Xi Zhang. 'Cross-Lingual Propaganda Detection.' In 2022 IEEE International Conference on Big Data (Big Data), pp. 4330-4336. IEEE, 2022.
[10] Vorakitphan, Vorakit, Elena Cabrio, and Serena Villata. 'PROTECT: A Pipeline for Propaganda Detection and Classification.' In CLiC-it 2021-Italian Conference on Computational Linguistics. 2022.
[11] Khanday, Akib Mohi Ud Din, Bharat Bhushan, Rutvij H. Jhaveri, Qamar Ray- ees Khan, Roshani Raut, and Syed Tanzeel Rabani. 'Nnpcov19: Artificial neural network-based propaganda identification on social media in covid-19 era.' Mobile Information Systems 2022 (2022): 1-10.
[12] Dash, Saloni, Arshia Arya, Sukhnidh Kaur, and Joyojeet Pal. 'Narrative Building in Propaganda Networks on Indian Twitter.' In Proceedings of the 14th ACM Web Science Conference 2022, pp. 239-244. 2022.
[13] Dimitrov, Dimitar, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. 'Detecting propa- ganda techniques in memes.' arXiv preprint arXiv:2109.08013 (2021).
[14] Da San Martino, Giovanni, Shaden Shaar, Yifan Zhang, Seunghak Yu, Alberto Barron-Cedeno, and Preslav Nakov. 'Prta: A system to support the analysis of propaganda techniques in the news.' In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 287-293. 2020.
[15] Sekaran, R., Munnangi, A. K., Ramachandran, M., & Gandomi, A. H. (2022). 3D brain slice classification and feature extraction using Deformable Hierarchical Heuristic Model. *Computers in Biology and Medicine*, *149*, 105990.
[16] McCloskey, Michael, and Neal J. Cohen. 'Catastrophic interference in connectionist networks: The sequential learning problem.' In Psychology of learning and motiva- tion, vol. 24, pp. 109-165. Academic Press, 1989.
[17] French, Robert M. 'Catastrophic forgetting in connectionist networks.' Trends in cognitive sciences 3, no. 4 (1999): 128-135.
[18] Zhang, Lei, Shupeng Wang, Fajie Yuan, Binzong Geng, and Min Yang. 'Lifelong language learning with adaptive uncertainty regularization.' Information Sciences 622 (2023): 794-807.
[19] Kanwatchara, Kasidis, Thanapapas Horsuwan, Piyawat Lertvittayakumjorn, Boonserm Kijsirikul, and Peerapon Vateekul. 'Rational LAMOL: A rationale-based lifelong learning framework.' In Proceedings of the 59th Annual Meeting of the As- sociation for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 2942-2953. 2021.
[20] Widmer, Gerhard, and Miroslav Kubat. 'Effective learning in dynamic environ- ments by explicit context tracking.' In Machine Learning: ECML-93: European Conference on Machine Learning Vienna, Austria, April 5–7, 1993 Proceedings 6, pp. 227-243. Springer Berlin Heidelberg, 1993.
[21] Huang, Yufan, Yanzhe Zhang, Jiaao Chen, Xuezhi Wang, and Diyi Yang. 'Contin- ual learning for text classification with information disentanglement based regular- ization.' arXiv preprint arXiv:2104.05489 (2021).
[22] Xiang, Jiannan, Tianhua Tao, Yi Gu, Tianmin Shu, Zirui Wang, Zichao Yang, and Zhiting Hu. 'Language Models Meet World Models: Embodied Experiences Enhance Language Models.' arXiv preprint arXiv:2305.10626 (2023).