Journal of Theoretical and Applied Information Technology
15th December 2024. Vol.102. No. 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPROVING TEXT CLASSIFICATION IN FEDERATED LEARNING THROUGH TRANSFER LEARNING: A COMBINED APPROACH FOR ENHANCED CONVERGENCE AND PERSONALIZATION

Dr. MOHAMMED ABDUL WAJEED *[1], Dr. ANNAVARAPU CHANDRA SEKHARA RAO[2], Dr. T.K. SHAIK SHAVALI[3], Dr. ABDUL RASOOL MD[4]

[1]Professor, Lords Institute of Engineering and Technology, Dept. Computer Science & Engineering, Hyderabad, India

[2]Associate Professor, Indian Institute of Technology, Department of Computer science & Engineering, Jharkhand, India

[1]Professor and Head, Dean SA, Lords Institute of Engineering and Technology, Dept. Computer Science & Engineering, Hyderabad, India

[1]Associate Professor and Head, Lords Institute of Engineering and Technology, Dept. Computer Science & Engineering (AIML), Hyderabad, India

E-mail: [1]drwajeed1@gmail.com

ABSTRACT

Federated learning (FL) has become a prominent decentralized approach to machine learning, allowing collaborative model training across distributed clients without compromising sensitive data. However, the diversity of data among clients, especially in text classification tasks, often results in slower convergence and subpar performance. This study introduces a novel framework that combines transfer learning with federated learning to tackle the issues of slower convergence as the learning begins from scratch, due to which the parameters of the learning model are randomly initialized. By utilizing pre-trained language models, we propose a two-phase approach where clients first fine-tune a pre-trained model by taking consideration of the parameters already obtained, on their local text data before engaging in a federated averaging process. This integration facilitates quicker convergence, better generalization across varied client data, and improved model customization for individual clients. We assess our methodology on several text classification datasets with differing levels of data heterogeneity and demonstrate that our approach significantly enhances both overall accuracy and communication efficiency compared to conventional federated learning techniques. The findings highlight the potential of transfer learning in boosting the effectiveness and scalability of federated learning for real-world text classification applications.

Keywords: *Five*: *Federated Learning, Transfer Learning, Text Classification`, Federated Averaging, Heterogeneous data*

1. INTRODUCTION

This The proliferation of data from decentralized sources like smartphones, IoT devices, and distributed computing systems has necessitated novel machine learning (ML) approaches that can extract insights from dispersed data while safeguarding privacy. Conventional ML models typically employ centralized training, consolidating data from various sources into a single location for learning. Although effective, this method raises concerns about data privacy, security, and adherence to regulations such as GDPR and HIPAA. Additionally, in sectors like healthcare and finance, transferring sensitive information to centralized servers is often undesirable or legally prohibited.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Federated Learning (FL) has emerged as a robust solution to these challenges, enabling collaborative model training across multiple clients without exchanging private data. FL functions by allowing each client (e.g., a smartphone, hospital, or other data-generating entity) to train a local version of an ML model using their private data as discussed in [1][15]. Rather than transmitting raw data to a central server, clients only send model updates, such as gradients or parameters. These updates are combined at the server to enhance a global model, which is then redistributed to the clients. This cyclical process continues until the global model reaches convergence, thus facilitating learning from distributed data sources while maintaining privacy and complying with data governance regulations. Despite its potential, FL faces significant hurdles that impact its efficiency, particularly when applied to real-world tasks like text classification. Data heterogeneity presents a major challenge in federated learning. In many FL applications, the data available across different clients is often non-IID (Independent and Identically Distributed), meaning that data distributions vary among clients as given in [17][8]. For instance, in an FL system involving multiple users, one user's smartphone might primarily contain emails, while another's may consist mostly of text messages or social media posts. This variability can lead to slower convergence of the global model, suboptimal performance, and difficulties in generalizing across all clients. Moreover, training from scratch in such a decentralized manner often requires substantial communication between clients and the central server, which can be resource-intensive.

Transfer Learning (TL) offers an effective strategy to address these issues. TL is an ML technique that utilizes pre-trained models—those trained on large, generalized datasets—and adapts them to new tasks or domains where labeled data may be limited. The underlying principle of transfer learning is that models trained on one problem can retain knowledge applicable to other related problems. For example, a model pre-trained on a vast corpus of general English text can be fine-tuned to classify medical documents or categorize customer reviews with relatively little new data. Transfer learning generally decreases the necessary training data and time for new tasks, as the model can leverage features acquired from previous tasks. This approach is especially beneficial when resources are limited or when the target task lacks sufficient labeled data for training from the ground up.

Combining Transfer Learning with Federated Learning presents a promising hybrid method to tackle FL challenges, particularly in diverse data environments. By using a pre-trained model as an initial point for each client, transfer learning enables the federated learning process to start with a model that already possesses general knowledge, thus significantly reducing the need for extensive local training. Clients can adapt the pre-trained model to their specific data distribution through fine-tuning. Following local training, clients transmit their model updates to the central server, which consolidates the updates and enhances the global model. This process harnesses both the strengths of transfer learning—by customizing pre-trained models for specific local tasks—and the benefits of federated learning—by ensuring raw data remains private between clients. Consequently, integrating transfer learning can result in quicker convergence, improved generalization, and enhanced model personalization in federated learning systems.

In text classification tasks, which aim to categorize or label text documents into predetermined categories (such as spam detection, sentiment analysis, or topic classification), the fusion of federated learning and transfer learning can be particularly beneficial. Text data is often highly confidential, making centralized training challenging. Moreover, text data can vary significantly between clients, depending on factors like language, dialect, or domain. For instance, healthcare institutions might work with clinical notes, while corporate clients may handle customer support emails or legal documents. A pre-trained language model, such as one trained on a vast corpus like BERT or GPT, can serve as an excellent foundation for federated text classification tasks. Clients can fine-tune these models locally using their specific text data, and the model can be iteratively refined through the federated learning process.

The main contribution of the paper is to study and investigates the incorporation of transfer learning into federated learning for text classification tasks. Federated learning aims in privacy preservation of the individual data, exploring federated learning on text classification would results in delayed convergence of the learning model due to difference in the distribution of the data at the clients. To speed up the convergence of the learning model, we introduce transfer learning for TC. A novel framework that allows clients to employ pre-trained language models as a foundation and fine-tune them on local

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

text data within a federated learning environment. Our approach seeks to address the challenges of data heterogeneity and slow convergence in FL while enhancing the efficiency of the training process. We assess our method on various text classification tasks across multiple datasets with heterogeneous distributions to showcase its effectiveness and scalability. The findings indicate that merging transfer learning with federated learning significantly improves both model accuracy and communication efficiency, making it a viable solution for distributed machine learning tasks in privacy-sensitive settings.

2. RELATED WORK

In recent times, federated learning (FL) and transfer learning (TL) have become significant paradigms in machine learning, tackling various challenges in model training, particularly in scenarios involving sensitive data or limited resources. While these approaches have been studied separately in different fields, the combination of transfer learning with federated learning is a relatively new concept, especially for text classification (TC) tasks. This section reviews the relevant literature on federated learning, transfer learning, and the attempts to combine these paradigms.

Federated Learning, first introduced by [1], transformed machine learning by allowing decentralized training across multiple clients while safeguarding data privacy. This method has found applications in diverse sectors, including healthcare, finance, and mobile applications. In FL, individual clients train local models using their private data and only share model updates (such as gradients or parameters) with a central server for consolidation. This strategy ensures that sensitive information remains local, minimizing privacy risks and adhering to data protection regulations. Initial research on FL primarily addressed issues like communication efficiency and data heterogeneity. For example, studies by [1][7][14] investigated methods for secure aggregation and efficient communication in FL settings. However, managing non-IID (non-independent and identically distributed) data continues to be a major challenge. In the realm of natural language processing (NLP) and TC, FL has been utilized for distributed text data across various clients, such as mobile devices or organizations as in [9][12]. [5] investigated FL in text mining tasks, demonstrating its effectiveness in applications like sentiment analysis and spam detection. Nevertheless, the primary limitation in FL-based TC is the variability in data distribution among clients—different clients may exhibit significantly different language patterns, document types, and domains, making it challenging for a global model to generalize effectively across all clients as in [4].

Transfer Learning has gained widespread adoption as a technique for addressing machine learning problems in resource-limited environments, particularly when labeled data is scarce or when training a model from scratch is computationally intensive. TL involves using a pre-trained model (typically trained on a large, general dataset) and fine-tuning it for a specific task or domain with minimal additional training. Pre-trained models such as BERT [2][3][11][20] have shown remarkable performance in NLP tasks including text classification, sentiment analysis, and question answering.

TL has proven especially effective in TC tasks where large-scale, labeled datasets are not readily available for the target domain as in [3]. presented the Universal Language Model Fine-tuning (ULMFiT) approach, showing how transfer learning could achieve state-of-the-art results in various TC tasks by fine-tuning a pre-trained language model on domain-specific dataas in [16]. Similarly, the use of pre-trained transformers like BERT has been shown to significantly reduce training times and enhance model performance in various NLP tasks. However, most research on transfer learning has focused on centralized settings, where pre-trained models are fine-tuned using data aggregated in a single location, making it unsuitable for privacy-preserving applications like FL.

2.1 Integrating Transfer Learning with Federated Learning

Although FL tackles privacy concerns in distributed settings, it faces challenges with data diversity, potentially resulting in suboptimal model performance and slow convergence. Conversely, TL reduces the need for extensive labeled datasets by employing pre-trained models with existing domain knowledge. Combining TL with FL is a logical step, as it can improve FL by offering a more robust starting point for the global model. Current studies have investigated the combination of TL and FL to address their respective limitations. [10] introduced FedMeta, which merges meta-learning with FL, transferring models trained on one group of clients to new clients in different domains. Likewise, [10] examined federated transfer learning (FTL), utilizing transfer learning to handle heterogeneous data distributions in FL. FTL shares knowledge between similar clients by exchanging model updates, enabling better generalization in federated environments with non-IID data.

Customization has become crucial in federated learning, as clients often require models tailored to their specific data, particularly in non-IID settings. Studies on federated personalization have developed techniques such as federated meta-learning [6][17] and personalized federated learning (PFL) [15], allowing clients to adapt global models to their local data. Integrating these customization methods with transfer learning could further enhance performance, especially in tasks like text classification, where linguistic patterns and document types vary significantly across clients. Nevertheless, the pursuit of personalization in federated learning often conflicts with the aim of creating a unified global model. Incorporating transfer learning provides a compromise, enabling clients to fine-tune a pre-trained model containing general domain knowledge while still contributing to the global model through federated updates. This approach not only enhances the global model's performance but also ensures that local models are customized to individual client data.

Despite progress in both FL and TL, their integration for TC remains relatively unexplored. This paper introduces a novel approach that uses transfer learning to enhance the performance and efficiency of federated learning in text classification tasks. By employing pre-trained language models like BERT, clients can fine-tune the model locally on their specific text data, addressing data heterogeneity issues. We also propose a personalization mechanism, allowing clients to further adapt the global model to their local requirements. Our approach is tested on multiple TC datasets with varying degrees of data heterogeneity, demonstrating its effectiveness in improving both convergence speed and model accuracy compared to conventional FL methods.

3. PROPOSED WORK

In this section, we present an innovative framework that combines TL with FL to tackle the challenges of heterogeneous data and slow convergence in decentralized TC tasks. Our approach aims to harness the advantages of both paradigms. TL's ability to provide robust model initialization through pre-trained models, and FL's privacy-preserving, distributed nature. Our proposal involves a combined approach where users employ a pre-trained language model as a foundation, customize it locally with their own text data, and then engage in federated model consolidation. This integration improves both the effectiveness and adaptability of text classification models across various user groups.

The fundamental concept underlying this method is to merge the benefits of pre-trained models with federated learning, resulting in a more efficient and potent system for text classification.

The framework integrates transfer learning with federated learning in text classification. It has distributed pre-trained model, local fine-tuning, model aggregation, and global model updates, along with optional personalization for each client. The framework consists of the following phases:

i. **Pre-trained Model Dissemination:**
The central server initially distributes a pre-trained language model (such as BERT or GPT) to all participating users. This model has undergone pre-training on an extensive corpus of general text data, enabling it to grasp basic language features like syntax, semantics, and word associations.

ii. **Local Customization:**
Each user tailors the pre-trained model to their local text dataset. The local datasets are often non-IID, indicating that different users may have varying data distributions. For instance, some users might have email data, while others possess customer reviews or legal documents. Customization involves modifying the pre-trained model's parameters to optimize its performance for each user's specific task, thereby allowing the model to learn user-specific patterns and enhance text classification accuracy locally.

iii. Federated Consolidation:
After local customization is finished, users transmit their model updates (e.g., model weights or gradients) to the central server. The central server consolidates these updates using a method like Federated Averaging (FedAvg), which calculates a weighted average of the local model updates to generate a new global model. This global model now incorporates knowledge from the

distributed users and is shared back with the users for further local customization.

iv. Iterative Training and Personalization:

This process of local customization and federated consolidation is repeated over multiple communication rounds until the global model converges. During each round, the model becomes increasingly specialized, taking into account the unique characteristics of each user's data while improving its generalization across all users. Additionally, a personalization step can be incorporated for each user to fine-tune the consolidated model based on the user's specific dataset, resulting in models that are better suited to individual user requirements.

3.1 Algorithm Design

Algorithm 1 outlines the detailed process for the suggested hybrid framework. The key steps are as follows:

Algorithm1: Federated Transfer Learning for Text Classification

Input: Pre-trained model MPre-trained, Client datasets Di (where i € {1,2,…N} N number of Clients)

Output: Global Model Mglobal and personalized models MiPersonalized

1. [Server Distribution of Pre-trained Model]

The Server disseminates MPre-trained to all clients.

2. [Local Refinement]

Each client Ci refines the pre-trained model MPre-trained using its local dataset Di, resulting in a locally refined model Milocal.

3. [Transmission of Model Updates]

Clients send their model updates (either model weights or gradients) to the server.

4. [Federated Consolidation]

The server consolidates local updates from all clients using a method such as Federated Average.

Mglobal = wi * Milocal i varying from 1 to N.

where wi represents the weight based on client dataset size.

5. [Model Distribution and Refinement]

The server sends the updated global model Mglobal back to clients. Each client further refines Mglobal on its local dataset, producing a personalized model MiPersonalized.

6. [Iterative Training]

Steps 3-5 are repeated for several rounds until model convergence.

7. [Optional Customization]

After the final round, clients may further customize the global model Mglobal by training it exclusively on their local data for additional epochs to obtain MiPersonalized..

3.2 Benefits of Proposed Framework

i. Enhanced Convergence and Generalization:

Initiating with a pre-trained model equips clients with a model already possessing general language structure knowledge. This minimizes the need for extensive local data training, resulting in quicker convergence compared to training models from scratch in a federated learning setup. Additionally, the incorporation of transfer learning ensures better generalization of the global model across diverse client datasets, as the pre-trained model's language representations help reconcile differences in data distributions.

ii. Addressing Data Heterogeneity

A significant challenge in federated learning is the non-IID nature of data across clients. In conventional federated learning, this heterogeneity can lead to slow convergence and a global model that performs poorly on individual client data. Our approach tackles this issue by allowing each client to refine a pre-trained model on its local dataset. This refinement enables the model to adapt to each client's unique data distribution before contributing to the global model.

iii. Decreased Communication Overhead

By refining pre-trained models locally, the proposed framework reduces the number of communication rounds needed for global model convergence. Clients begin with a robust base model, allowing them to focus on enhancing the model with their specific data rather than training from scratch. This efficiency not only shortens training time but also minimizes data exchange between clients and the server, crucial in bandwidth-limited environments.

iv. Customization for Client-Specific Requirements

The framework supports client-specific customization, allowing clients to further refine the global model to meet their local data needs.

This system provides both a comprehensive global model that generalizes effectively across various clients and customized models tailored to specific client tasks. This approach is especially advantageous in TC, where language patterns can differ greatly among clients.

4. EXPERIMENTATION AND EVALUATION

In this framework, we utilize pre-trained models like BERT or GPT as the foundational models distributed to clients. These models are renowned for their exceptional performance in NLP tasks, including text classification. Clients refine these models using conventional back propagation methods and their local datasets. The federated aggregation is executed using the Federated Averaging (FedAvg) algorithm, which effectively combines local model updates into a unified global model. The framework is built using PyTorch and TensorFlow Federated, which supply the necessary tools for both model refinement and federated learning.

We assess the proposed method on several real-world text classification datasets, encompassing datasets from sentiment analysis, spam detection, and topic classification. Each dataset is divided among clients in a non-IID manner to replicate real-world data heterogeneity. We contrast the performance of our framework with traditional federated learning approaches (training from the ground up) and centralized transfer learning (where data is consolidated). The primary metrics evaluated include classification accuracy, communication efficiency (quantified by the number of communication rounds), and convergence rate.

To gauge the efficacy of the proposed approach, we perform experiments using several widely-recognized benchmark datasets for text classification. The main goal of these experiments is to compare our federated transfer learning framework's performance with conventional federated learning (FL) methods and centralized transfer learning (TL) techniques. Specifically, we concentrate on the following key metrics:

Classification Accuracy: The model's capacity to accurately classify text across clients.

Communication Efficiency: The quantity of communication rounds required for model convergence as in [17].

Convergence Speed: The rate at which the global model attains stable performance.

Model Personalization: The ability of the local models to adjust to specific client data.

Data Sets: We select the following popular datasets that encompass a broad range of text classification tasks, including sentiment analysis, topic classification, and spam detection. These datasets are divided in a non-IID fashion to simulate real-world heterogeneous data distributions across clients.

AG News: This dataset comprises over 120,000 news articles categorized into one of four classes: World, Sports, Business, and Science/Technology. This dataset is frequently used for topic classification.

IMDB Movie Reviews: A dataset of 50,000 movie reviews for sentiment analysis, with each review labeled as either positive or negative. Sentiment analysis is particularly valuable in federated learning applications where client data may vary considerably in content.

Spam Assassin: A well-known dataset for spam detection, consisting of email messages labeled as either spam or not spam. This is highly relevant for applications such as mobile messaging or email categorization in federated learning.

Reuters-21578: A dataset containing over 21,000 news documents classified into various topics. It is often employed for multi-class text classification tasks, making it a suitable candidate for federated learning experiments.

These datasets represent diverse text classification tasks with varying degrees of data heterogeneity, allowing us to examine the generalizability of the proposed framework.

In our experiment, we replicate a federated learning setting with multiple participants, each holding a segment of one dataset. To reflect real-world diversity, we allocate non-IID data to each participant, resulting in varied data distributions. For instance, one participant might have predominantly positive IMDB reviews, while another has mostly negative ones. Likewise, for the AG News dataset, some participants may have a higher proportion of Sports articles, while others have more Business-related content.

Baseline Approaches

i. Standard Federated Learning (SFL): In this setup, participants train their models from the ground up without pre-trained models. The FedAvg algorithm is employed to combine the local models as in[13]. It is also termed as traditional FL.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ii. Centralized Transfer Learning (CTL): For comparison, we train a pre-trained BERT model on the complete dataset in a centralized fashion.

Suggested Method

iii. Federated Transfer Learning (FTL): In our proposed framework, each participant receives a pre-trained BERT model, fine-tunes it using their local data, and transmits the model updates to the central server for federated aggregation. It is the suggested proposed method.

Performance Metrics

i. Accuracy: We assess the overall accuracy of the global model on a separate test set from each dataset.

ii. Communication Efficiency: We monitor the number of communication rounds needed for the global model to reach convergence as in [19].

iii. Personalization Accuracy: After the global model converges, we evaluate the accuracy of personalized models that are fine-tuned locally for each participant.

Table 1: AG News Dataset(Topic Classification)

Model	Accuracy	Communication Rounds	Convergence Rate
SFL	82.1	120	110
CTL	89.3	-	10
FTL	87.2	35	30

In table-1, the accuracy, communication rounds and convergence rates are given fot the dataset AG News, the FTL approach achieves 87.2% accuracy, considerably surpassing tradition FL (SFL) at 82.1%. Furthermore, FTL converges in only 35 communication rounds, compared to 120 for traditional FL. The centralized TL (CTL) model attains the highest accuracy (89.3%) as it's trained on the entire dataset, but it doesn't maintain data privacy.

. Table 2: IMDB Mvie Reviews (Sentiment Analysis)

Model	Accuracy	Communication Rounds	Convergence Rate
SFL	79.6	150	140
CTL	88.5	-	8
FTL	86.1	40	35

As given in table 2, in sentiment analysis using IMDb, FTL reaches 86.1% accuracy, which is nearer to the CTL model (88.5%) than traditional (standard) SFL (79.6%). FTL converges substantially faster, requiring just 40 communication rounds, versus 150 for traditional FL.

. Table 3: SpamAssassin Dataset(Spam Detection)

Model	Accuracy	Communication Rounds	Convergence Rate
SFL	91	130	120
CTL	95.5	-	7
FTL	93.8	30	25

As given in the table 3, for spam detection with the Spam Assassin dataset, FTL attains 93.8% accuracy, outperforming traditional SFL (91.0%) and closely matching the centralized CTL model (95.5%). Communication efficiency is notably improved, with FTL needing only 30 communication rounds compared to 130 for traditional FL.

Table 4: Reuters-21578 Dataset (Topic Classification)

Model	Accuracy	Communication Rounds	Convergence Rate
SFL	75.3	180	170
CTL	85.2	-	15
FTL	83.5	50	40

As given in table 4, with the Reuters-21578 dataset, FTL achieves 83.5% accuracy, markedly improving upon traditional SFL's 75.3%. Moreover, FTL requires only 50 communication rounds to converge, compared to 180 rounds in the traditional FL approach.

5. DISCUSSION

When The empirical findings across all four datasets indicate that the suggested federated transfer learning (FTL) approach surpasses conventional federated learning in both precision and communication efficacy. Utilizing a pre-trained model enables clients to begin with robust initial language representations, thus reducing the need for extensive local training and minimizing the required communication rounds for convergence. Although the centralized transfer learning method yields the highest accuracy, it fails to meet privacy requirements, making FTL a more appropriate choice for privacy-conscious applications.

Moreover, the FTL framework exhibits notable enhancements in heterogeneous settings where clients possess non-IID data distributions. By locally fine-tuning the pre-trained model, clients can adapt it to their specific data while contributing to the global model's effectiveness. This blend of personalization and global aggregation renders FTL a robust solution for distributed text classification tasks. Individual client's data could be from a different distribution from other clients data distribution, but due to the federated learning mechanism, taking into consideration the difference in the distribution, the model parameters are tuned to distribution and thus achieves personalization.

6. CONCLUSION AND FUTUER WORK

This study introduced an innovative framework that combines transfer learning (TL) with federated learning (FL) to tackle key challenges in decentralized text classification tasks. Our method capitalizes on the benefits of pre-trained language models, such as BERT, to enhance local model performance in federated learning environments, particularly in scenarios with non-IID data distributions. By refining pre-trained models on local client data and consolidating model updates at a central server, our approach significantly improves convergence speed, communication efficiency, and overall classification accuracy compared to traditional federated learning models trained from the ground up.

The experimental outcomes on various benchmark datasets, including AG News, IMDb, Spam Assassin, and Reuters-21578, showcase the efficacy of our proposed framework. The federated transfer learning method consistently outperforms traditional FL in terms of accuracy and necessitates considerably fewer communication rounds to achieve convergence. Additionally, the framework's capacity to tailor models for individual clients ensures that the system remains adaptable to the unique data distributions present in real-world federated learning scenarios.

While our proposed approach successfully enhances federated learning for text classification, several areas warrant further investigation:

Advanced Personalization Techniques: Although we incorporated client-specific fine-tuning for personalization, more sophisticated methods, such as meta-learning or multi-task learning, could further improve the personalization of federated models for clients with significantly different data distributions.

Dynamic Model Adaptation: The current framework relies on a fixed pre-trained model for all clients. Future research could explore adaptive transfer learning, where different pre-trained models or model layers are dynamically selected based on the client's data characteristics to further enhance local model performance.

Privacy-Preserving Techniques: While federated learning itself is privacy-preserving, future work could integrate differential privacy or homomorphic encryption techniques to further protect client data during model updates and aggregation.

Resource-Efficient FL: As federated learning environments are often limited by communication bandwidth and computational resources, optimizing the communication and computation costs of the proposed framework remains a crucial area for future research. Techniques such as model pruning or quantization can help reduce model size and communication overhead without compromising performance.

REFERENCES:

[1] McIlroy, M. Douglas, et al. "Mass-produced software components." Proceedings of the 1st International Conference on Software Engineering, GarmischPattenkirchen, Germany. sn, 1968.

[2] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS).

[3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). 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 (NAACL).

[4] Ruder, S. (2019). Neural Transfer Learning for Natural Language Processing. Ph.D. Dissertation, National University of Ireland, Galway.

[5] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., & Zhao, H. (2021). Advances and Open Problems in Federated Learning. Foundations and Trends® in Machine Learning, 14(1–2), 1-210.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[6] Chen, X., Liu, C., Li, B., & Lin, Q. (2020). Federated Learning of N-grams for Secure Text Classification. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL).

[7] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated Learning: Challenges, Methods, and Future Directions. IEEE Signal Processing Magazine, 37(3), 50-60.

[8] Liu, Y., Liu, J., Chen, J., & Xu, M. (2021). Federated Learning Meets Natural Language Processing: A Survey. arXiv preprint arXiv:2107.12603.

[9] Zhu, H., Zhou, Z., Jin, M., Li, C., Zheng, V. W., & Cui, P. (2021). Data-Free Knowledge Distillation for Heterogeneous Federated Learning. In Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD).

[10] Shokri, R., & Shmatikov, V. (2015). Privacy-Preserving Deep Learning. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS).

[11] Yu, H., Yang, X., Zhu, Y., Wu, W., & Ma, R. (2020). A Federated Transfer Learning Framework for Privacy-Preserving Text Classification. Information Sciences, 527, 64-78.

[12] Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to Fine-Tune BERT for Text Classification?. In Proceedings of the China National Conference on Chinese Computational Linguistics (pp. 194-206). Springer.

[13] Hard, A., Rao, K., Mathews, R., Beaufays, F., Augenstein, S., Eichner, H., & Ramage, D. (2018). Federated Learning for Mobile Keyboard Prediction. arXiv preprint arXiv:1811.03604.

[14] Li, Q., He, B., & Song, D. (2020). Practical Privacy-Preserving Federated Recommender Systems. In Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM) (pp. 2617-2624)

[15] Loshchilov, I., & Hutter, F. (2017). Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations (ICLR).

[16] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated Machine Learning: Concept and Applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1-19.

[17] Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification (ULMFiT). In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 328-339).

[18] Smith, V., Chiang, C. K., Sanjabi, M., & Talwalkar, A. (2017). Federated Multi-Task Learning. In Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS).

[19] Zhang, C., Xiong, Z., Lou, Y., Zhang, J., Wu, J., & Guizani, M. (2020). Privacy-Preserving Federated Learning for Safer Medical Diagnosis. IEEE Access, 8, 115832-115841.

[20] Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated Learning: Strategies for Improving Communication Efficiency. In Proceedings of the 30th Conference on Neural Information Processing Systems (NeurIPS) Workshop on Private Multi-Party Machine Learning.