DTL-Based CSI Feedback Combined With Continual Learning in FDD Massive MIMO Systems

Mayuko Inoue
Graduate School of
Science and Technology
Keio University
Kanagawa 223-8522, Japan
Email: inoue@ohtsuki.ics.keio.ac.jp

Tomoaki Ohtsuki
Department of Information and
Computer Science
Keio University
Kanagawa 223-8522, Japan
Email: ohtsuki@ics.keio.ac.jp

Mondher Bouazizi
Department of Information and
Computer Science
Keio University
Kanagawa 223-8522, Japan
Email: bouazizi@ohtsuki.ics.keio.ac.jp

***Abstract*— The Deep Transfer Learning (DTL)-based Channel State Information (CSI) approach exploits the Deep Neural Network (DNN) model to provide low-cost CSI feedback for the target channel. It involves offline training on a given CSI dataset, followed by fine-tuning for a new environment using a small amount of collected CSI data, reducing feedback costs. However, in practical scenarios, the fine-tuned model performs worse on the source channel compared to the original source model. This leads to continued use of the poorly performing target model until it's fine-tuned again with newly acquired data. To address this, we propose combining DTL-based CSI feedback with continual learning. We introduce elastic weight consolidation (EWC) into the loss function during fine tuning. Simulations show that our method significantly reduces the degradation of the target model on the source channel, as measured by NMSE, compared to a method without continual learning.**

I. Introduction

Channel State Information (CSI) feedback is one of the most common approaches to acquiring downlink CSI at the base station (BS), wherein the user equipment (UE) estimates the CSI and sends it to the BS. While CSI feedback employing deep neural networks (DNN) has superior CSI reconstruction performance [1], the limited generalization capability of DNNs can pose challenges with varying wireless channels, necessitating retraining with updated CSI data. In the context of FDD Massive MIMO, the high-dimensionality of downlink CSI necessitates a substantial dataset for effective DNN training. To address this issue, the Deep Transfer Learning (DTL) based CSI feedback was proposed in [2]. DTL is a technique that leverages knowledge acquired from a specific environment and adapts it to a different environment. The DTL-based CSI feedback targets Clustered Delay Line (CDL) channel models representing various channel environments, which are categorized into five types from CDL-A to CDL-E. In the DTL-based CSI feedback, a deep learning model (source model) is initially trained using a large amount of CSI data (source data) collected from a source channel (CDL-A). Subsequently, the parameters of the source model are fine-tuned using a limited amount of CSI data (target data) acquired from a target channel (CDL-B/C/D/E). By doing so, it becomes possible to create a model tailored to the target environment (target model) using less data and in significantly less time. However, despite improving the CSI feedback in terms of Normalized Mean Square Error (NMSE) in the target channel, the fine-tuned model generated using this approach [2] no longer performs well in the source channel. In real-time-varying channel environments, it is quite possible for the channel environment to transition back from the target channel environment to the source channel environment. In such cases, the target model, which performs poorly on the source channel, continues to be used until it can be fine-tuned with the source data again. Continual learning enables DNN to leverage knowledge acquired from previous tasks (or in previous environments in our case) without the need to retrain from scratch even if the data used from these previous tasks is no longer available [3].

In this paper, to mitigate the degradation of the target model's reconstruction performance on the source channel, we propose a method that combines the DTL-based CSI feedback with continual learning. In wireless communications, continual learning has been applied to enhance robustness in areas of signal detection and channel estimation [4] [5]. However, as far as our investigation reveals, continual learning has not been previously employed in the context of DTL-based CSI feedback. In this paper, we apply one of the continual learning methods, a regularization-based technique known as Elastic Weight Consolidation (EWC), to the DTL-based CSI feedback approach. This incorporation of EWC aims to make the target model less susceptible to forgetting the knowledge acquired from the source data. Simulation results show that the target model generated by our method can alleviate the degradation of the NMSE on the source channel when compared to the target model without EWC. Furthermore, we illustrate that the target model produced by our method, using the CDL-ALL source data, which encompasses a mixture of CDL-A through CDL-E, consistently achieves favorable NMSE performance

979-8-3503-7099-7/24/$31.00

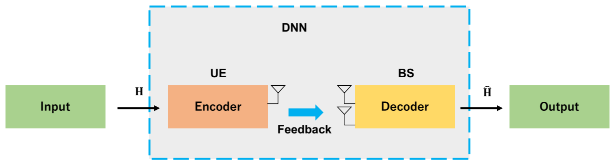

Fig. 1. A process of the DNN-based downlink CSI feedback approach

TABLE I
NMSE ACHIEVED BY THE SOURCE MODEL AND THE TARGET MODEL (SOURCE CHANNEL CDL-A, TARGET CHANNEL: CDL-B).

Evaluation channels	NMSE of each model in evaluation channels [dB]	
	source model	target model
CDL-A (source channel)	-28.48	-20.77
CDL-B (target channel)	-14.82	-25.22

across various channels.

The remainder of this paper is structured as follows. Section II highlights the key idea and concepts behind the DTL-based CSI feedback as described in [2]. Section III describes our method, which combines DTL-based CSI feedback and continual learning. Section IV describes our computer simulation parameters and presents the obtained results. Finally, Section V concludes this paper.

II. DTL-BASED CSI FEEDBACK

In this section, will first describe how DNNs are used for CSI feedback approach using DNN. Afterwards, we introduce the DTL-based CSI feedback approach [2].

A. The DNN-based downlink CSI feedback approach

The DNN-based downlink CSI feedback approach is illustreated in Fig. 1. Initially, the UE estimates the downlink CSI, represented as $\mathbf{H}$. This estimated CSI is then passed through the DNN's encoder, resulting in a low-dimensional code word represented as $\boldsymbol{s}$. Subsequently, the code word, $\boldsymbol{s}$ is fed back from the UE to the BS. When the BS receives $\boldsymbol{s}$, it is input to the decoder in the DNN and reconstructed into the original CSI $\hat{\mathbf{H}}$. To assess the accuracy of DNN recovery, we employ the NMSE as an evaluation metric.

$$NMSE = E\left[\left\|\mathbf{H}-\hat{\mathbf{H}}\right\|_2^2 / \left\|\hat{\mathbf{H}}\right\|_2^2\right]. \quad (1)$$

B. DTL-based CSI feedback approach

The wireless scenarios in this research are simulated using CDL channel model [6]. CDL channel model is specifically defined within the 5G standard. It includes five types: CDL-A, CDL-B, and CDL-C for Non Line-of-Sight (NLOS) channel environment, and CDL-D and CDL-E for Line-of-Sight (LOS) channel environment.

In [2], DTL is utilized to obtain the DNN for individual target channels. First, the DNN is trained as the source model using an extensive dataset of CDL-A source channel as the source data. Subsequently, the source model undergoes fine-tuning using a smaller dataset of CDL-B/C/D/E target channel, which serve as the target data respectively. This process enables the generation of a target model for each target channel, CDL-B/C/D/E, with fewer samples and shorter training time compared to approaches that do not employ DTL. However, the target model tends to forget the training results of the source task, resulting in a significant degradation in NMSE performance on source channel compared to that of the source model. Table I shows the NMSE that the source model and the target model achieve on the source channel (CDL-A) and the target channel (CDL-B). In general, when dealing with new tasks, neural networks tend to forget the knowledge acquired from previously learned tasks. In the case of Table I, the DNN has forgotten the source task, which involved training on the source data (CDL-A), while it was focused on the target task of fine tuning with the target data (CDL-B). As a result, the NMSE of the target model on CDL-A source channel is much worse than that of the source model. This is known as catastrophic forgetting, and it presents a challenge in real-time-varying channel environments. In real-time-varying channel environments, it becomes necessary to collect data from the changed environment and retrain the target model with this new data. There's a possibility that the environment may change and become similar to the source data. In such cases, the target model must continue to be used with degraded performance until retraining is finished.

Continual learning is a progressive learning mechanism to mitigate catastrophic forgetting and enable a model to learn multiple tasks. Methods of continual learning can be broadly categorized into three main categories: (i) regularization-based, (ii) replay, and (iii) parameter isolation methods [3].

III. PROPOSED METHOD

To mitigate the substantial degradation of NMSE in the source channel of the target model, we introduce continual learning into the DTL-based CSI feedback approach [2]. As mentioned earlier, there are three main methods of continual learning: regularization-based, replay, and parameter isolation method. We opt for a regularization-based approach called EWC. It is the most straightforward to apply to the DTL-based downlink CSI feedback approach among the three methods because there is no need to modify the architecture of the neural network itself. In EWC, a regularization term is incorporated into the loss function, which signifies the importance of retaining knowledge from previous tasks.

In this paper, the regularization term is used in the loss function during fine tuning. Generally, the loss function during fine tuning denoted as $\mathcal{L}_{target}$ for the DNN is represented as follows.

$$\mathcal{L}_{target} = \frac{1}{N_s}\sum_{i=1}^{N_s}\|f(\boldsymbol{\theta}_{target}, \mathbf{H}_i) - \mathbf{H}_i\|^2 \quad (2)$$

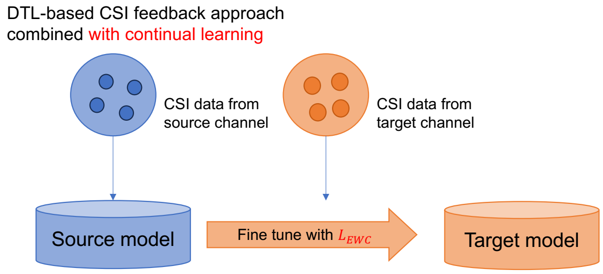

Fig. 2. A flow of proposed approach. (The conventional DTL-based CSI feedback approach uses L_{target} instead of L_{EWC})

TABLE II
SIX CASES TO BE CONSIDERED.

#	Source channel	Model to measure NMSE
1	CDL-A	Source model
2	CDL-A	Target model without EWC
3	CDL-A	Target model with EWC
4	CDL-ALL	Source model
5	CDL-ALL	Target model without EWC
6	CDL-ALL	Target model with EWC

where N_s is the number of samples used for fine tuning, $\mathbf{H}_i$ is the estimated CSI, $f(\boldsymbol{\theta}_{target}, \mathbf{H}_i)$ is the output of DNN, $\boldsymbol{\theta}_{target}$ are weights of the target model. This loss function will learn for the target task forgetting the source task. In this paper, we introduce L_{EWC}, which combines L_{target} with an EWC regularization term, as a loss function during fine tuning.

$$\mathcal{L}_{EWC} = \mathcal{L}_{target} + \sum_{i=1}^{L} \frac{\mu}{2} F_i(\boldsymbol{\theta}_{target,i} - \boldsymbol{\theta}_{source,i})^2 \qquad (3)$$

L represents the number of DNN layers, and μ is the coefficient, which indicates the importance of the source task with respect to the target task. F_i is the Fisher information matrix for layer i. $\boldsymbol{\theta}_{source,i}$ and $\boldsymbol{\theta}_{target,i}$ are the weight matrices of the source and target models of layer i, respectively. By adding the EWC regularization term, during the learning of a target task, any deviation from the optimal weights of the source task is penalized. This prevents the weights from becoming biased toward the target task, allowing the target model to retain knowledge from the source task. It is also important to highlight that this approach is agnostic to the DNN architecture and can be applied to various types of DNNs. The flow of this approach is presented in Fig. 2.

IV. SIMULATION RESULTS

A. Parameters Setting

Our experiment focused on the FDD massive MIMO scenario. The uplink channel frequency is 2.0 GHz, while the downlink channel frequency is 2.1 GHz. The Number of UE antenna is 2, and the BS has 32 antennas. Our experiment employed 72 subcarriers with a spacing of 15 KHz and 14 OFDM symbols. The estimation of the UE's CSI and the feedback CSI at the BS were assumed to be error-free. In the simulation, we used the DNN from [2], and the compression ratio was set at 1/8. 50000 source data samples are used for pre training while fine-tuning employs 1000 target data samples. As for the source channel, we utilized CDL-A and CDL-ALL, where CDL-ALL consists of 10 000 samples from each CDL-A/B/C/D/E.

B. Simulation Results

1) Evaluating the impact of EWC coefficient μ: First, we evaluated the effect of the EWC coefficient μ. Table III illustrates the NMSE of the target model as $\mu = 2^x$ is varied. The target channel is CDL-B and the source channel is CDL-A. The bottom columns of Tables III display the average NMSE for the five evaluation channels (CDL-A, CDL-B, CDL-C, CDL-D, and CDL-E). For instance, in the first row and first column of Table III, the value of -22.22 dB represents the NMSE in CDL-A of the target model fine-tuned in CDL-B with $\mu = 2^{-9}$, using the source model trained in CDL-A. As μ increases, the NMSE achieved by the target model in the CDL-B target channel tends to degrade, while the NMSE achieved in the source channel (CDL-A) tends to improve. This is because with increasing μ, the penalty for large deviations from the source model parameters also increases. Figures 3 and 4 depict the average NMSE across the five evaluation channels of the target model as μ is varied. Fig. 3 displays the NMSE for models fine-tuned with CDL-B/C/D/E target channels, using the CDL-A source channel. On the other hand, Fig. 4 illustrates the NMSE for the model fine-tuned with CDL-B/C/D/E target channels, using the CDL-ALL source channel. For example, the value of -23.75 dB in Fig. 3 corresponds to the entry in the first column of the sixth row of Table III (the average NMSE across the five evaluation channels of the target model fine-tuned in CDL-B with $\mu = 2^{-9}$, using the source model trained in CDL-A). Based on this, we aimed to identify the value of μ that yields good NMSE performance, irrespective of the channel model. For the CDL-ALL source channel, it was observed that the average NMSE tends to be smaller when $\mu = 2^{-6}$. In the case of the CDL-A source channel, the lowest average NMSE is achieved when $\mu = 2^{-9}$. However, Table III indicates that the NMSE in the CDL-A source channel at $\mu = 2^{-9}$ experiences a significant degradation compared to that at $\mu = 2^{-6}$. Considering the above results, we selected $\mu = 2^{-6}$, and this value will be used for subsequent simulations.

C. Comparison of the case with and without EWC

We conducted a comparison between cases with and without the EWC regularization term, i.e., using $\mathcal{L}_{EWC}$ versus using $\mathcal{L}_{target}$ as the loss function during fine tuning. Table V shows the NMSE achieved by the source model, the target model without the EWC regularization term, and the target model with the EWC regularization term on both the source channel

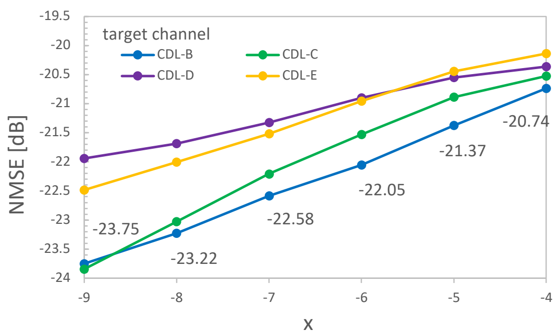

Fig. 3. Average NMSE across the five evaluation channels achieved by the target model when varying $\mu = 2^x$ (source channel: CDL-A, target channel: CDL-B/C/D/E)

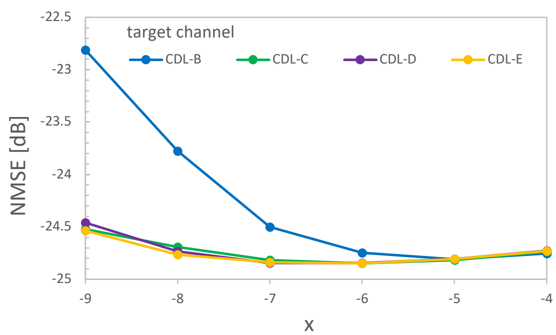

Fig. 4. Average NMSE across the five evaluation channels achieved by the target model when varying $\mu = 2^x$ (source channel: CDL-ALL, target channel: CDL-B/C/D/E)

(CDL-A) and the target channels (CDL-B/C/D/E). The NMSE of the target model without the EWC regularization term on the source channel significantly degraded compared to that of the source model. On the other hand, the target model with the EWC regularization term mitigated the degradation of NMSE achieved on the source channel, although it led to a degradation in NMSE on the target channel compared to the target model without the EWC regularization term. This result shows that in real-time-varying channel environments, our method can reduce the relearning cost of the target model when the environment switches from the target channel back to the source channel.

D. Consideration of the case when a good NMSE is achieved on any channel

We considered the case when a good NMSE is achieved on any channel. The six cases shown in Table II were considered. For source channels, we examined both CDL-A and CDL-ALL. CDL-ALL was considered because it is advantageous to incorporate a diverse set of channel models in the source data to attain good NMSE across various channels. Table IV

TABLE III
NMSE ACHIEVED BY THE TARGET MODEL WHEN VARYING μ (SOURCE CHANNEL: CDL-A, TARGET CHANNEL: CDL-B)

Evaluation channels	NMSE of the target model when varying $\mu = 2^x$ in evaluation channels [dB]					
	$x=-9$	$x=-8$	$x=-7$	$x=-6$	$x=-5$	$x=-4$
CDL-A	-22.22	-23.08	-24.09	-25.42	-27.12	-28.24
CDL-B	-23.14	-21.62	-19.9	-18.13	-16.42	-15.33
CDL-C	-23.82	-23.12	-22.28	-21.9	-21.91	-21.72
CDL-D	-25.01	-24.42	-24.32	-25.15	-26.33	-27.23
CDL-E	-25.34	-24.56	-24.17	-24.67	-25.32	-25.62
AVG	-23.75	-23.22	-22.58	-22.05	-21.37	-20.74

illustrates the NMSE achieved by the models generated in the six cases outlined in Table II across different channels (CDL-A/B/C/D/E). The target channel varies from the CDL-B to CDL-E samples. The target model created using CDL-ALL as the source channel and fine-tuned with the EWC regularization term (case 6 in Table II) exhibits superior NMSE compared to the other cases, regardless of the channel. These results indicate that the cases using CDL-ALL as the source channel and incorporating the EWC regularization term tend to achieve a good NMSE for any channel.

V. CONCLUSIONS

In this paper, we propose a DTL-based CSI feedback combined with continual learning to mitigate the degradation of the NMSE achieved by the target model on the source channel. Specifically, we introduced the EWC regularization term into the loss function during fine tuning. The target model generated by our method showed less degradation in the NMSE achieved on the source channel than the model generated without the EWC. Furthermore, when using the CDL-ALL source channel, our method's target model consistently achieved favorable NMSE across different channels. These results suggest the our approach can adapt effectively to real-time-varying channel environments, maintaining good performance without the need for extensive relearning.

REFERENCES

[1] C. Wen, W. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett,* vol. 7, no. 5, pp. 748-751, Oct. 2018.
[2] J. Zeng, J. Sun, G. Gui, B. Adebisi, T. Ohtsuki, H. Gacanin, and H. Sari, "Downlink CSI Feedback Algorithm with Deep Transfer Learning for FDD Massive MIMO System," *IEEE Trans. Cogn. Commun. Netw,* vol. 7, no. 4, pp. 1253-1265, Dec. 2021.
[3] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, and T. Tuytelaars, "Continual learning: A comparative Study on How to Defy Forgetting in Classification Tasks," *arXiv preprint arXiv:1909.08383,* 2019.
[4] S. Kumar, S. K. Vankayala, B. S. Sahoo, and S. Yoon, "Continual Learning-Based Channel Estimation for 5G Millimeter-Wave Systems," *Annu. Consum. Commun. Netw. Conf,* pp. 1-6, 2021.
[5] X. Yang, F. Li, T. Li, W. Ji, and Y. Liang, "Elastic Weight Consolidation Continual Learning Based Signal Detection in Multiple Channel MIMO System," *IEEE/CIC International. Conf. Commun. China,*July, 2021.
[6] 3GPP TR 38.901, "Study on channel model for frequencies from 0.5 to 100 GHz," *3rd Generation Partnership Project; Technical Specification Group Radio Access Network*, 2017.

TABLE IV
NMSE THAT THE SOURCE MODEL, THE TARGET MODEL WITHOUT EWC, AND THE TARGET MODEL WITH EWC, ACHIEVE ON THE SOURCE AND THE TARGET CHANNELS (SOURCE CHANNEL: CDL-A).

(a) Target channel: CDL-B.

Evaluation channels	NMSE of each model in evaluation channels [dB]		
	source model	target model w/o EWC	target model w/ EWC
CDL-A	-28.48	-20.77	-25.42
CDL-B	-14.82	-25.22	-18.13

(b) Target channel: CDL-C.

Evaluation channels	NMSE of each model in evaluation channels [dB]		
	source model	target model w/o EWC	target model w/ EWC
CDL-A	-28.48	-25.55	-26.89
CDL-C	-21.37	-28.09	-23.98

(c) Target channel: CDL-D.

Evaluation channels	NMSE of each model in evaluation channels [dB]		
	source model	target model w/o EWC	target model w/ EWC
CDL-A	-28.48	-25.50	-28.04
CDL-D	-27.27	-32.54	-29.30

(d) Target channel: CDL-E.

Evaluation channels	NMSE of each model in evaluation channels [dB]		
	source model	target model w/o EWC	target model w/ EWC
CDL-A	-28.48	-24.31	-28.19
CDL-E	-25.27	-32.26	-27.05

TABLE V
NMSE THAT THE MODELS OBTAINED BY SIX CASES IN TABLE II.

(a) Target channel: CDL-B.

Evaluation channels	NMSE of each model obtained by six cases in evaluation channels [dB]					
	#1	#2	#3	#4	#5	#6
CDL-A	-28.48	-20.77	-25.42	-21.94	-16.97	-21.75
CDL-B	-14.82	-25.22	-18.13	-22.04	-24.27	-22.38
CDL-C	-21.37	-24.58	-21.90	-25.57	-22.52	-25.59
CDL-D	-27.27	-25.72	-25.15	-30.91	-26.07	-30.94
CDL-E	-25.27	-26.32	-24.67	-31.25	-27.52	-31.35

(b) Target channel: CDL-C.

Evaluation channels	NMSE of each model obtained by six cases in evaluation channels [dB]					
	#1	#2	#3	#4	#5	#6
CDL-A	-28.48	-25.55	-26.89	-21.94	-21.63	-22.02
CDL-B	-14.82	-20.27	-16.06	-22.04	-20.85	-22.13
CDL-C	-21.37	-28.09	-23.98	-25.57	-26.68	-25.82
CDL-D	-27.27	-29.28	-27.23	-30.91	-30.27	-31.49
CDL-E	-25.27	-28.94	-26.13	-31.25	-30.46	-31.92

(c) Target channel: CDL-D.

Evaluation channels	NMSE of each model obtained by six cases in evaluation channels [dB]					
	#1	#2	#3	#4	#5	#6
CDL-A	-28.48	-25.50	-28.04	-21.94	-21.59	-22.03
CDL-B	-14.82	-16.88	-15.69	-22.04	-20.06	-22.11
CDL-C	-21.37	-23.72	-22.64	-25.57	-24.97	-25.76
CDL-D	-27.27	-32.54	-29.30	-30.91	-33.03	-31.61
CDL-E	-25.27	-31.43	-27.72	-31.25	-32.86	-32.02

(d) Target channel: CDL-E.

Evaluation channels	NMSE of each model obtained by six cases in evaluation channels [dB]					
	#1	#2	#3	#4	#5	#6
CDL-A	-28.48	-24.31	-28.19	-21.94	-20.92	-22.03
CDL-B	-14.82	-17.93	-15.61	-22.04	-21.06	-22.12
CDL-C	-21.37	-24.07	-22.50	-25.57	-24.96	-25.76
CDL-D	-27.27	-31.92	-28.52	-30.91	-32.46	-31.59
CDL-E	-25.27	-32.26	-27.05	-31.25	-33.52	-32.06