

基于连续学习损失正则化的网络分布迁移下的信道预测

穆罕默德·阿赫迈德·莫欣¹, 穆罕默德·乌梅尔¹, 阿桑·比拉勒², 穆罕默德·伊布萨姆·卡迪尔³,
穆罕默德·阿里·贾姆谢德⁴, 迪恩·F·休根², 约翰·M·奇奥菲¹

¹ Dept. of Electrical Engineering, Stanford University, Stanford, CA, USA

² School of Computer Science, University of Oklahoma, Norman, OK, USA

³ Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA

⁴ College of Science and Engineering, University of Glasgow, Glasgow, UK

ABSTRACT

现代无线网络在移动用户穿越具有不同天线布局、载波频率和散射统计的异构网络配置时面临关键挑战。传统预测器在分布变化下性能下降,交叉配置切换期间 NMSE 上升 37.5%。这项工作通过提出基于损失正则化的持续学习框架来解决信道预测中的灾难性遗忘问题。该方法通过添加选择性保留先前配置所需网络参数的惩罚项,增强标准训练目标,同时使网络能够适应新环境。研究了两种主要的正则化策略:弹性权重固化(EWC)和突触智能(SI)。在 3GPP 场景和多个架构中,SI 将高 SNR 下的 NMSE 基底降低了多达 1.8 dB ($\approx 32\text{--}34\%$),而 EWC 达到最多 1.4 dB ($\approx 17\text{--}28\%$)。值得注意的是,标准 EWC 会产生 $\mathcal{O}(MK)$ 复杂度(存储 MFisher 对角线条目和对应参数快照跨 K 任务),除非进行合并,而 SI 保持 $\mathcal{O}(M)$ 内存复杂度(存储 M 模型参数),与任务序列长度无关,使其适用于资源受限的无线基础设施¹。

Index Terms— 持续学习、通道预测、灾难性遗忘和损失正则化。

1. 介绍

新兴和即将推出的无线网络依赖于先进的物理层设计,例如大规模多输入多输出(MIMO),以满足不断增长的吞吐量和可靠性需求。这些技术的有效性在

根本上取决于准确及时的信道状态信息(CSI)。挑战随着 5G 新无线电(NR)规范而增加,该规范要求时分双工(TDD)模式下的上行链路探测间隔至少为 2 毫秒,然而移动用户在 28 GHz 的信道相干时间可能短至 0.3 毫秒 [1]。这种差异要求从单纯的信道估计转向主动的信道预测,其中对未来信道状态进行预测以抵消除内在 CSI 过时 $\rho(\Delta t) = J_0(2\pi f_D \Delta t)$,其中 $J_0(\cdot)$ 是零阶贝塞尔函数, f_D 是最大多普勒频移, Δt 是探测延迟。未能如此操作可能导致吞吐量大幅下降,研究表明,在中等移动性下 4 毫秒的反馈延迟可以将系统总速率降低大约 50% [2, 3, 4]。

基于学习的预测器捕捉时变信道的非线性动态,并且通常优于统计模型 [5]。然而,当用户穿越异构小区时,延展扩散、路径损耗和空间相关性的分布变化可能导致 NMSE 增加 37.5% (1.8 dB) [6]。根据领域适应理论, $\epsilon_T(f) \leq \epsilon_S(f) + d_{\mathcal{H}}(S, T) + \lambda$,其中增长的源-目标差异 $d_{\mathcal{H}}(S, T)$ 在切换过程中解释了这种增加,并促使持续学习。

处理这种分布变化的直接方法是对新数据进行微调模型,但这种方法会导致“灾难性遗忘”,即神经网络在学习新任务时突然丢失之前获得的知识 [7]。对于重新访问先前单元的用户而言,遗忘是非最优的。基于回放的持续学习可以缓解这一问题,但需要 $\mathcal{O}(NK)$ 内存,使得这种方法不适合资源受限的基础设施。因此,鲁棒的信道预测自然是一个持续学习任务 [8]。

本工作通过损失正则化解决了信道预测中的灾难

¹本工作是在与英特尔公司、三星电子和爱立信合作下完成的。

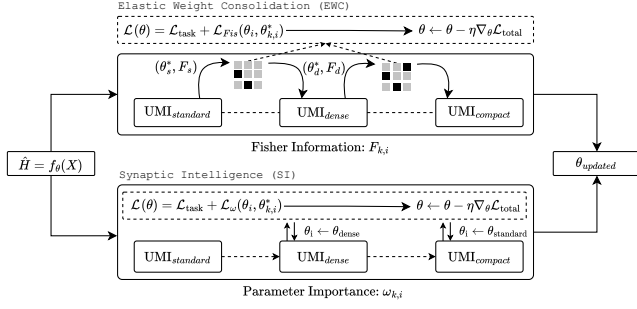


Fig. 1: EWC 和 SI 在连续通道预测中的比较。该框架展示了使用费舍尔信息 ($F_{k,i}$) 和参数重要性 ($\omega_{k,i}$) 的任务损失正则化，能够在适应分布变化的同时保留知识。

性遗忘问题：通过增加保留对先前环境至关重要的参数的惩罚项来增强任务损失。该方法平衡了可塑性（学习新条件）与稳定性（保持先验知识）。我们评估了两种关键策略，弹性权重固化 (EWC) 和突触智能 (SI) [9]，如算法 1 和图 1 所示。

贡献：我们将异构网络配置中的信道预测框架视为一个持续学习任务，并提出了一种损失正则化框架，该框架比较了 EWC 和 SI。我们明确地将 EWC 的二次惩罚与具有对角高斯先验（精度 $F_{k,i}$ ，方程 4 和 5）的最大后验估计联系起来，而 SI 通过在线权重累积（如方程 10 所示）得出数据驱动的精度。这澄清了为什么在典型的移动信道的小批量噪声和非平稳性下，SI 更加稳定。实验表明，EWC 可降低 0.8 至 1.4 dB ($\approx 17\text{--}28\%$) 的 NMSE，而 SI 则能达到高达 1.8 dB ($\approx 32\text{--}34\%$)，并且使用与任务无关的 $O(M)$ 内存，使得该方法在不同的深度学习框架和 3GPP 场景中都具有可行性。

2. 问题公式化

一系列信道预测任务 $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_K\}$ 被考虑，其中每个任务 \mathcal{T}_k 对应一个特定的网络配置和数据集 $\mathcal{D}_k = \{(\mathbf{X}_i^{(k)}, \mathbf{H}_i^{(k)})\}_{i=1}^{N_k}$ 。这里， $\mathbf{X}_i^{(k)} \in \mathbb{C}^{2 \times T \times N_{\text{tx}} \times N_{\text{rx}}}$ 表示一个包含 T 个过去信道实现的序列（实部和虚部分开），而 $\mathbf{H}_i^{(k)} \in \mathbb{C}^{2 \times N_{\text{tx}} \times N_{\text{rx}}}$ 是待预测的目标信道矩阵。

当具有参数 θ 的神经网络顺序训练任务 $\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_k$ 时，简单的微调会导致灾难性遗忘 [10, 11]，即模型

Algorithm 1 统一损失正则化连续通道预测 (EWC / SI)

Require: 学习率 η ; 方案 $r \in \{\text{EWC}, \text{SI}\}$; EWC 系数 α ; SI 系数 β ; 阻尼 ξ

- 1: 初始化参数 θ ; 如果 $r = \text{EWC}$ 然后 $\mathcal{B} \leftarrow \emptyset$ 否则 $\omega_i, \tilde{\omega}_i \leftarrow 0, \theta_i^{\text{ref}} \leftarrow \theta_i \forall i$
- 2: for each task \mathcal{D}_k do
- 3: for each mini-batch $(\mathbf{X}, \mathbf{H}) \in \mathcal{D}_k$ do
- 4: $\ell_{\text{task}} \leftarrow \ell_{\text{NMSE}}(\mathbf{H}, f_{\theta}(\mathbf{X}))$
- 5: $\ell_{\text{reg}} \leftarrow \begin{cases} \frac{\alpha}{2} \sum_{(\theta_j^*, F_j) \in \mathcal{B}} \sum_i F_{j,i} (\theta_i - \theta_{j,i}^*)^2, & r = \text{EWC} \\ \frac{\beta}{2} \sum_i \omega_i (\theta_i - \theta_i^{\text{ref}})^2, & r = \text{SI} \end{cases}$
- 6: $\theta \leftarrow \theta - \eta \nabla_{\theta} (\ell_{\text{task}} + \ell_{\text{reg}})$
- 7: if $r = \text{SI}$ then
- 8: $\tilde{\omega}_i \leftarrow \tilde{\omega}_i + (\partial_{\theta_i} \ell_{\text{task}})^2 \eta \forall i$
- 9: end if
- 10: end for
- 11: if $r = \text{EWC}$ then
- 12: $F_{k,i} \leftarrow \frac{1}{|\mathcal{D}_k|} \sum_{(\mathbf{X}, \mathbf{H}) \in \mathcal{D}_k} (\partial_{\theta_i} \ell_{\text{NMSE}})^2$; $\mathcal{B} \leftarrow \mathcal{B} \cup \{(\theta_k^* \leftarrow \theta, F_k)\}$
- 13: else
- 14: $\Delta \theta_i \leftarrow \theta_i - \theta_i^{\text{ref}}$; $\omega_i \leftarrow \omega_i + \frac{\tilde{\omega}_i}{(\Delta \theta_i)^2 + \xi}$; $\tilde{\omega}_i \leftarrow 0$;
 $\theta_i^{\text{ref}} \leftarrow \theta_i$
- 15: end if
- 16: end for

在早期任务 $\mathcal{T}_1, \dots, \mathcal{T}_{k-1}$ 上的性能严重下降。实证研究表明，在网络配置转换过程中标准化 NMSE 增加 37.5%（如第 4 节所述）。

令 $f_{\theta} : \mathbb{C}^{2 \times T \times N_{\text{tx}} \times N_{\text{rx}}} \rightarrow \mathbb{C}^{2 \times N_{\text{tx}} \times N_{\text{rx}}}$ 表示由 θ 参数化的信道预测函数。任务 k 的预测损失定义为：

$$\mathcal{L}_k(\theta) = \frac{1}{|\mathcal{D}_k|} \sum_{(\mathbf{X}, \mathbf{H}) \in \mathcal{D}_k} \frac{\|\mathbf{H} - f_{\theta}(\mathbf{X})\|_F^2}{\|\mathbf{H}\|_F^2}, \quad (1)$$

其中 $\|\cdot\|_F$ 表示弗罗贝尼乌斯范数。挑战是在保持之前任务性能的同时最小化当前任务的损失 $\mathcal{L}_k(\theta)$ ，即避免 $j < k$ 的 $\mathcal{L}_j(\theta)$ 显著增加。

损失正则化方法通过在训练目标中增加惩罚项来解决灾难性遗忘问题，这些惩罚项会根据参数对先前任务的重要性选择性地约束参数更新。与需要 $O(NK)$ 内存来存储跨 K 任务的 N 样本的回放方法不同，正则化技术保持了 $O(M)$ 的内存复杂度用于 M 模型参数，这使得在资源受限的无线基础设施上部署成为可能。这一基本见解利用了网络参数对不同传播环境中

预测准确性的异质贡献。一些参数编码与任务无关的时间依赖性，而另一些则捕获必须在网络转换过程中保留的特定配置信道统计信息。通过将连续学习公式化为一个约束优化问题，其中参数重要性权重 ω_i 调节正则化强度（如图 2 所示），这些方法在不阻碍适应的情况下保留知识 [12]，并且具有与网络配置无关的固定开销。

3. 方法论

本工作提出了一种用于持续通道预测的损失正则化框架，该框架结合了两种互补的方法：EWC 和 SI，如图 1 所示。这两种方法都通过增加保持之前网络配置知识的正则项来增强标准的通道预测损失 [13]。损失正则化框架通过对重要的先前任务参数的变化进行惩罚项的添加，从而修改了标准训练目标。正则化损失函数的一般形式为：

$$\mathcal{L}_{\text{total}}(\boldsymbol{\theta}) = \mathcal{L}_{\text{task}}(\boldsymbol{\theta}) + \mathcal{L}_{\text{reg}}(\boldsymbol{\theta}), \quad (2)$$

其中， $\mathcal{L}_{\text{task}}(\boldsymbol{\theta})$ 是当前任务的标准 NMSE 损失， $\mathcal{L}_{\text{reg}}(\boldsymbol{\theta})$ 是从先前任务中保留知识的正则项。

EWC 通过使用费雪信息矩阵 (FIM) 估计参数的重要性来识别关键参数，并对偏离为先前任务找到的最佳参数值进行惩罚 [7, 14]。对于每个完成的任务 \mathcal{T}_k 和最优参数 $\boldsymbol{\theta}_k^*$ ，对角 FIM F_k 被经验性地计算为：

$$F_{k,i} = \frac{1}{|\mathcal{D}_k|} \sum_{(\mathbf{X}, \mathbf{H}) \in \mathcal{D}_k} \left(\frac{\partial \mathcal{L}_{\text{NMSE}}(\boldsymbol{\theta}_k^*; \mathbf{X}, \mathbf{H})}{\partial \theta_i} \right)^2. \quad (3)$$

这种基于 NMSE 的近似可以被表述为基于似然性的费雪信息：

$$F_{k,i} \approx \mathbb{E}_{(\mathbf{X}, \mathbf{H}) \sim \mathcal{D}_k} \left[\left(\frac{\partial}{\partial \theta_i} \log p(\mathbf{H} | \mathbf{X}; \boldsymbol{\theta}_k^*) \right)^2 \right]. \quad (4)$$

具有高费雪信息的参数被认为是维持该任务性能的关键。单个先前任务的 EWC 正则化项为：

$$\mathcal{L}_{\text{EWC}}^{(k)}(\boldsymbol{\theta}) = \frac{\alpha}{2} \sum_i F_{k,i} (\theta_i - \theta_{k,i}^*)^2, \quad (5)$$

其中， $\alpha > 0$ 是一个稳定系数，控制正则化的强度，而 $\theta_{k,i}^*$ 表示在完成的任务 \mathcal{T}_k 训练后参数 θ_i 的值。此二次惩罚可解释为最大后验估计 (MAP)，其中第一项对应

于当前任务数据的负对数似然，第二项编码了一个精度为 $F_{k,i}$ 的高斯先验：

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \arg \min_{\boldsymbol{\theta}} \left[-\log p(\mathcal{D}_k | \boldsymbol{\theta}) + \frac{\alpha}{2} \sum_i F_{k,i} (\theta_i - \theta_{k,i}^*)^2 \right]. \quad (6)$$

这里， $F_{k,i}$ 是费雪信息（方程.4），将对数似然的曲率与高斯先验强度联系起来。对于多个先前任务，EWC 损失变为：

$$\mathcal{L}_{\text{EWC}}(\boldsymbol{\theta}) = \frac{\alpha}{2} \sum_i \sum_{j=1}^k F_{j,i} (\theta_i - \theta_{j,i}^*)^2. \quad (7)$$

EWC 算法首先初始化模型参数和一个参数库。对于每个网络配置，它使用梯度下降处理结合的任务和正则化损失的小批量数据。随后，它计算费雪信息矩阵并存储具有其重要性权重的最佳参数以供未来正则化使用。

输入界面 通过在训练过程中在线跟踪参数的重要性解决了 EWC 的计算和内存限制问题，而无需显式地进行费雪信息的计算 [9, 15, 16]。在任务 \mathcal{D}_k 上的训练过程中，SI 维护每个参数 θ_i 的参数重要性 $\tilde{\omega}_i$ 的运行估计：

$$\tilde{\omega}_i \leftarrow \tilde{\omega}_i + \left(\frac{\partial \mathcal{L}_{\text{NMSE}}}{\partial \theta_i} \right)^2 \eta, \quad (8)$$

其中 η 是学习率。这种累积捕获了当前任务中每个参数对损失减少的总贡献。等效地，可以将此更新解释为沿优化轨迹由参数 i 执行的工作的路径积分。令 $g_i^t = \partial_{\theta_i} \mathcal{L}_{\text{NMSE}}(\theta^t)$ 和 $\Delta \theta_i^t = \theta_i^{t+1} - \theta_i^t$ 。则总功为

$$\mathcal{W}_i^{(k)} \approx \sum_{t \in \mathcal{T}_k} g_i^t \Delta \theta_i^t \approx \sum_{t \in \mathcal{T}_k} (g_i^t)^2 \eta, \quad (9)$$

这表明方程 (8) 中的 $\tilde{\omega}_i$ 是优化路径积分的一个离散近似。完成任务 k 后，SI 通过每个参数的平方位移来归一化这个累积工作：

$$\omega_i \leftarrow \omega_i + \frac{\tilde{\omega}_i}{(\Delta \theta_i)^2 + \xi}, \quad (10)$$

其中 $\Delta \theta_i = \theta_i - \theta_i^{(0)}$ 是在任务 k 期间参数 i 的总变化量，而 $\xi > 0$ 是一个小的阻尼项以防止除零错误。这给出了一个最终的重要性得分 ω_i ，该分数捕捉了每个参数相对于其位移执行了多少“有用的工作”。SI 正则化项为：

$$\mathcal{L}_{\text{SI}}(\boldsymbol{\theta}) = \frac{\beta}{2} \sum_i \omega_i (\theta_i - \theta_i^{(0)})^2, \quad (11)$$

其中, $\beta > 0$ 是一个类似于 EWC 中 α 的稳定性系数, 而 $\theta_i^{(0)}$ 是在任务开始时的参考参数。训练通过梯度下降在组合目标上进行:

$$\theta_i^{t+1} = \theta_i^t - \eta \left(\frac{\partial \mathcal{L}_{\text{task}}}{\partial \theta_i} + \lambda \frac{\partial \mathcal{L}_{\text{SI}}}{\partial \theta_i} \right), \quad (12)$$

其中, $\lambda = \beta$ 缩放 SI 惩罚项。这将 SI 与其他损失正则化方法统一为一个约束关键参数方向更新的带惩罚优化过程。

EWC 和 SI 表现出互补的权衡: EWC 需要 $O(MK)$ 内存, 并通过 Fisher 信息计算使训练时间翻倍, 而 SI 则通过在线重要性跟踪维持 $O(M)$ 内存并具有可忽略的计算开销。稳定性—塑形权衡可以表示为一个约束优化问题:

$$\min_{\theta} \mathbb{E}_{\mathcal{T}_k} [\mathcal{L}_{\text{task}}(\theta)] \quad \text{s.t.} \quad \Delta \mathcal{L}_{\text{prev}} \leq \epsilon, \quad (13)$$

其中 $\Delta \mathcal{L}_{\text{prev}}$ 是之前任务上的损失增加量。实际上, 通过二次惩罚 (EWC) 或在线重要性重新加权 (SI), 这个约束被放宽了, 强制过去任务的最大允许损失增加量为 ϵ 。

SI 更加稳健, 因为它在整个训练过程中累积参数的重要性, 而 EWC 的 Fisher 近似在高梯度随机性下可能不可靠。该框架通过参数正则化与长短期记忆 (LSTM)、门控循环单元 (GRU) 和 Transformer 集成: $\alpha, \beta \in [0.1, 1.0]$ 平衡可塑性和稳定性, $\xi \in [10^{-6}, 10^{-3}]$ 稳定 SI。范围来自对保留的 Urban Microcell (UMi) 验证集进行网格搜索, 跨所有主干网络, 其中 ξ 防止在 $(\Delta\theta)^2 + \xi$ 中出现接近零的分母。

4. 性能评估

实验评估表明, 所提出的损失正则化框架在三种具有不同传播复杂性的代表性的 UMi 场景中有效, 利用 LSTM、GRU 和 Transformer 架构进行信道预测, 并采用如图 2 所示的受限回放缓存。全面评估显示, SI 通过利用在线重要性跟踪始终超越 EWC, 这在小批量随机性下提供了更可靠的参数估计。在 UMi- 紧凑的场景中, SI 实现了 1.7–1.8 分贝 NMSE 改进 ($\approx 34\%$ 减少), 而在密集和标准设置中, 它保持了 0.8–1.4 分贝增益, 如图 2c 所示。

所提出的框架在信噪比与均方误差特性的表现中展现出卓越的鲁棒性, 在关键的 8–12 dB 信噪比区间内保持了显著的误差减少, 同时避免了传统方法在高信噪比下出现的饱和伪影 (图 2c)。在 20 dB 信噪比下的性能分析显示, SI 具备更强的能力维持低误差地板以应对各种信道条件变化, 这是实际无线基础设施部署的关键需求。在线积累参数重要性相较于事后 FIM 更准确地保持了环境间的信道预测。

架构无关的性能分析显示, 在 LSTM、Transformer 和 GRU 实现中都有一致的改进, 如图 2a 所示, LSTM 架构在两种正则化方法下均表现出最优性能。从紧凑到标准场景, 预测复杂度增加, 但 SI 维持的性能差距比基线更小, 在密集和标准 UMi 设置中持续保持 0.8 – 1.4 dB 的增益 (图 2c)。使用 $O(M)$ 内存且开销可忽略不计的情况下, SI 在分布变化下保留了先前的知识, 实现了无需重放或特定任务存储的可靠持续信道预测。

5. 结论与未来工作

持续信道预测依赖于平衡可塑性和稳定性。简单的微调会破坏这种平衡, 而损失正则化则保留了稳定适应所需的关键参数。SI 的在线重要性累积比基于 Fisher 的 EWC 更为可靠, 这强调了记忆编码设计与模型容量一样, 对于无线系统中的鲁棒持续学习至关重要。

未来的工作包括自适应正则化, 以调整在移动性和相干时间变化下的稳定性和可塑性之间的权衡, 并将框架扩展到大规模 MIMO、多小区协作和联邦持续学习以提高可扩展性。进一步的方向涉及在 3GPP CSI 反馈管道中量化参数的重要性, 并在配备可重构智能表面的超密集部署中进行验证, 同时保持轻量级内存和计算需求。

6. REFERENCES

- [1] Alejandro Villena-Rodriguez, Francisco J Martín-Vega, Gerardo Gómez, Mari Carmen Aguayo-Torres, and Georges Kaddoum, “Aging-resistant wideband precoding in 5g and beyond

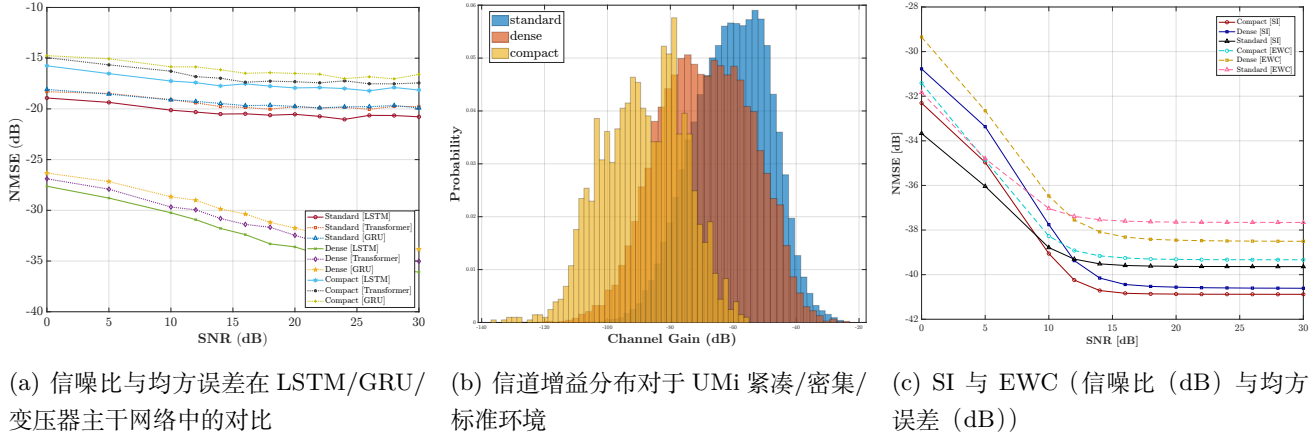


Fig. 2: EWC 与 SI 在持续通道预测中的比较

using 3d convolutional neural networks,” arXiv preprint arXiv:2407.07434, 2024.

- [2] Huaifu Li, Liqin Ding, Yang Wang, Peng Wu, and Zhenyong Wang, “Impact of channel aging on massive mimo vehicular networks in non-isotropic scattering scenarios,” in 2021 IEEE Global Communications Conference (GLOBECOM). IEEE, 2021, pp. 1–6.
- [3] Muhammad Umer, Muhammad Ahmed Mohsin, Ahsan Bilal, and John M. Cioffi, “Neural gaussian radio fields for channel estimation,” 2025.
- [4] Muhammad Ahmed Mohsin, Syed Muhammad Jameel, Hassan Rizwan, Muhammad Iqbal, Tabinda Ashraf, and Jen-Yi Pan, “Transformer-based distributed machine learning for downlink channel estimation in ris-aided networks,” in 2025 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), 2025, pp. 1–5.
- [5] Wei Jiang and Hans D Schotten, “Neural network-based fading channel prediction: A comprehensive overview,” IEEE Access, vol. 7, pp. 118112–118124, 2019.
- [6] Muhammad Ahmed Mohsin, Muhammad Umer, Ahsan Bilal, Muhammad Ali Jamshed, and

John M Cioffi, “Continual learning for wireless channel prediction,” arXiv preprint arXiv:2506.22471, 2025.

- [7] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al., “Overcoming catastrophic forgetting in neural networks,” Proceedings of the National Academy of Sciences, vol. 114, no. 13, pp. 3521–3526, 2017.
- [8] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars, “A continual learning survey: Defying forgetting in classification tasks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 7, pp. 3366–3385, 2021.
- [9] Friedemann Zenke, Ben Poole, and Surya Ganguli, “Continual learning through synaptic intelligence,” in International Conference on Machine Learning. PMLR, 2017, pp. 3987–3995.
- [10] Rajat Varshney, Chirag Gangal, Mohd Sharique, and Mohammad Samar Ansari, “Deep learning based wireless channel prediction: 5g sce-

nario,” *Procedia Computer Science*, vol. 218, pp. 2626–2635, 2023.

- [11] Sebin J Olickal and Renu Jose, “Lstm projected layer neural network-based signal estimation and channel state estimator for ofdm wireless communication systems,” *AIMS Electronics & Electrical Engineering*, vol. 7, no. 2, 2023.
- [12] Zhiyuan Chen and Bing Liu, “Continual learning and catastrophic forgetting,” in *Lifelong Machine Learning*, pp. 55–75. Springer, 2022.
- [13] Haoxuan Qu, Hossein Rahmani, Li Xu, Bryan Williams, and Jun Liu, “Recent advances of continual learning in computer vision: An overview,” *IET Computer Vision*, vol. 19, no. 1, pp. e70013, 2025.
- [14] Ferenc Huszár, “Note on the quadratic penalties in elastic weight consolidation,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 11, pp. E2496–E2497, 2018.
- [15] Constantin Octavian Puiu, “Rethinking exponential averaging of the fisher,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2022, pp. 327–343.
- [16] Gido M van de Ven, “On the computation of the fisher information in continual learning,” *arXiv preprint arXiv:2502.11756*, 2025.