

基於 SAC 深度強化式學習於 5G 網路高度動態流量條件下之資源允入控制最佳化

1. 摘要

第五代行動通訊網路 (5G) 提供更高頻寬和更低延遲，但隨之而來的是比前幾代更多樣且巨大的網路流量。為了充分利用 5G 網路資源，必須有精確的資源允入控制 (Resource Admission Control, RAC) 機制來有效分配網路資源，最大化整體網路效能。本研究提出一套基於深度強化學習的資源允入控制方法，結合 Soft Actor-Critic (SAC) 深度強化學習演算法的離散化版本，稱為 SAC-RAC 方法。該方法利用 SAC 演算法中的最大熵 (Maximum Entropy) 策略特性來提升在高度動態環境下的決策穩健性；同時引入優先經驗回放 (Prioritized Experience Replay) 機制，讓演算法能夠更有效地從重要經驗中學習，提升訓練效率和效益。此外，SAC-RAC 加入了一種有條件的學習更新機制，透過設定獎勵門檻來決定神經網路是否進行參數更新，以適度減少更新次數。實驗結果顯示，所提出的 SAC-RAC 方法在多種動態流量情境下均取得最佳表現。相較於傳統方法（如貪婪式分配、固定切片）以及深度 Q 網路 (Deep Q Network, DQN)，SAC-RAC 在相同環境下的平均獎勵提升約 25%，網路壅塞率降低約 84%。整體而言，本研究所提出的方法有效地在高度動態的 5G 網路環境中達成較高的資源利用效益與穩定的服務品質。

2. 研究背景與動機

隨著行動裝置和物聯網設備激增，現代網路面臨海量且瞬息萬變的流量需求挑戰。5G 網路雖提供了技術上的優勢，但其複雜的網路環境必須承載遠超過以往的流量負載。在這種高度動態流量條件下，如果缺乏有效的資源允入控制機制，網路可能出現資源不足、壅塞甚至服務中斷的情況，導致服務品質 (QoS) 嚴重下滑。傳統上，一些簡單的方法如貪婪式請求處理（盡可能接受

所有請求)或**固定切片**資源分配(預先將頻寬固定劃分給各類服務)被用於 RAC。然而,貪婪策略可能因無差別地允入請求而導致資源耗盡和網路壅塞;固定切片則缺乏彈性,無法適應流量的即時變化,可能造成某些切片資源閒置而另一些切片資源不足的情況。

近年來,深度強化學習 (Deep Reinforcement Learning, DRL) 為代表的人工智慧技術被廣泛應用於解決此類複雜環境下的資源管理問題。例如,深度 Q 網路 (DQN) 方法已被證明能透過神經網路學習環境狀態與動作的關係,在一定程度上做出資源分配決策。然而,DQN 等方法在高動態環境中面臨挑戰:其學習過程可能需要大量時間來探索狀態空間,且在環境快速變化時策略更新不夠穩健。因此,本研究的動機在於尋找一種**更高效且適應性更強**的強化學習方法來進行 5G 網路的資源允入控制,使其能在劇烈變動的流量條件下依然維持高效能和高品質的服務。Soft Actor-Critic (SAC) 演算法以其最大熵強化學習策略聞名,有潛力提供更穩定且具探索性的決策機制,適合應對高度不確定性的環境。因此,我們基於 SAC 發展出專門用於 RAC 問題的改進型演算法,希望克服傳統方法的不足,實現最佳的資源分配效益和 QoS 表現。

3. 研究方法

本研究提出的 **SAC-RAC** 方法核心是將軟演員-評論家演算法 (Soft Actor-Critic, SAC) 應用於 5G 網路資源允入控制,並針對問題特性進行三項改良:

- **SAC 演算法與最大熵策略**: SAC 是一種先進的離線 (off-policy) 演員-評論家型深度強化學習演算法。與傳統強化學習不同的是, SAC 在目標函數中引入了策略的熵值,透過最大化「報酬+熵」來訓練策略。這種**最大熵策略**鼓勵智能體在獲取高獎勵的同時保持行為的隨機性,從而增加探索多樣性。在 5G 高動態網路環境下,最大熵方法的優勢在於:即使面對快速變化的流量模式,智能體也能避免過早收斂到次佳策略,持續探索更優的資源分配方案,提升決策的穩健性和適應性。本研究使用**離散化的 SAC (SAC-Discrete, SACD)** 演算法來適應資源分配動作的離散特性。也就是說,將 SAC 原本定義於連續動作空間的策略擴展至可處理**離散動作空間**的版本,使其能決策如「接受或拒絕某類服務請求」這類離散的允入控制行動。

- **優先經驗回放機制**：為了提高學習效率，SAC-RAC 引入了**優先經驗回放 (Prioritized Experience Replay)**。經驗回放機制讓強化學習代理可以將歷史交互經驗存入記憶體，並在訓練時隨機抽取一批經驗進行學習。然而，並非所有經驗對於策略提升都有相同價值；優先經驗回放透過為每筆經驗分配一個優先級（通常根據 TD 誤差大小），使得具有較高學習價值的經驗被抽取到的概率更高。於是，代理人會更頻繁地重新學習那些「意外性高」或「與預期差異大」的關鍵情境，有助於加速收斂並提升最終策略的品質。在 SAC-RAC 中，我們將優先級機制作為經驗回放的一部分，確保演算法將更多注意力放在可能導致資源分配策略顯著改進的經驗上，從而**增加此方法所帶來的效益**。該機制的實現包含維護一個優先級緩衝記憶體，以及依據經驗的重要度（例如獎勵誤差）來調整抽樣機率等技術細節。
- **有條件的學習更新機制**：針對 5G 網路環境中流量變化快速、訓練樣本數龐大的特性，本研究在 SAC-RAC 中設計了一種**有條件觸發的學習更新策略**。一般強化學習演算法在每一個時間步或每一回合都固定地對神經網路參數進行更新，然而在動態環境中，過於頻繁的更新可能導致不穩定的收斂，且增加大量計算時間。為此，我們設定一個**獎勵閾值 (reward threshold)** 作為學習觸發條件：在每一訓練回合結束時，代理根據該回合得到的累積獎勵與預先定義的閾值相比較。只有當回合獎勵**低於**此閾值且經驗回放緩衝中已儲存足夠的經驗時，才啟動一次網路參數的抽樣更新；反之，若當前回合獲得的獎勵已高於閾值，則認為策略表現良好，**暫停更新**一次，以節省訓練時間。如此一來，代理能夠將更多計算資源集中於表現不佳的情況進行學習修正，同時避免在狀態分布尚未有較大變化時的冗餘更新。該機制作法有效**穩定了策略收斂的波動範圍**，並因降低了更新頻率而**縮短了訓練時間**，而最終策略性能仍得以維持。

綜上所述，SAC-RAC 方法充分結合了最大熵強化學習的穩健探索優勢、優先經驗回放的高效學習特性，以及條件式更新策略的計算效率提升。在實現上，本研究首先利用排隊理論建立了一個模擬的 5G 網路環境，其中包含多類型（最多四種）不同優先等級的服務流量。系統將可用的資源（切片數量）以及新抵達的服務請求數作為環境狀態輸入，讓代理（SAC-RAC 或對比演算法）決定對哪些請求提供服務資源、對哪些請求拒絕允入。透過這樣的設計，智能體學會在不斷變化的流量條件下動態調配有限資源，以最大化長期獎勵（例如吞吐量或收益）並維持服務品質。

4. 實驗設計與結果分析

為驗證所提出 SAC-RAC 方法的效能，本研究搭建了模擬的 5G 網路環境並設計多組實驗情境進行比較分析。實驗環境基於排隊理論模型，模擬不同**流量類型與優先級**的服務請求抵達過程。透過調整參數，我們可以控制 1 至 4 種不同優先等級的流量輸入強度（例如使用 Poisson 過程的到達率 λ 參數，不同數值代表不同負載程度），其中 $\lambda=0$ 表示該類流量不出現。在每種情境下，我們設定總可用資源（切片）數量的不同取值，讓各方法在資源充足與匱乏的情況下都接受考驗。例如，在實驗中我們分別考慮了總切片資源為 30、40、50、60 的情境，以觀察資源多寡對各方法效能的影響。對於比較方法，本研究選取了傳統**貪婪法 (Greedy)**、經典**固定切片法 (Fixed Slice)**，以及引入優先經驗回放的**深度 Q 網路 (DQN)** 作為基準。貪婪法每回合盡可能接受所有請求，固定切片法則將切片資源事先均勻分配給各類服務（在不同情境下手動調整預分配比例）。DQN 方法我們同樣為其加入優先經驗回放機制以增強學習效率，使比較更加公平。

實驗設計方面，在進行訓練前，我們可為 SAC-RAC、DQN、Greedy 三種代理設定初始可用的資源切片數量，由代理自行學習如何分配；固定切片法則如前述需要人工配置資源比例。訓練過程中則對 SAC-RAC 和 DQN 代理設定獎勵閾值，以啟用其各自的條件更新策略來控制訓練長度。在模型收斂後，我們對每種方法進行多次獨立測試以減少隨機性的影響——每個情境下重複執行 50 次模擬，取其結果的平均值作為最終性能評估依據。評估指標包括**累積獎勵（收益）**以及**壅塞率**等。其中累積獎勵反映網路資源利用效益（越高越好），壅塞率則表示因資源不足而被拒絕服務的請求比例（越低越好），可視為網路擁堵情況的衡量。

結果分析顯示，所提 SAC-RAC 方法在所有測試場景中均取得最優績效。在不同流量組合和不同資源配給的實驗環境下，SAC-RAC 都能穩定地達到最高的平均獎勵，同時保持最低的壅塞率。特別是在流量高度複雜且資源相對有限的情境中，SAC-RAC 相較其他方法的優勢更加明顯：由於其策略考慮了長期回報並進行隨機探索，能動態調整資源分配以應對突發的流量高峰，因而有效降低了網路壅塞發生的頻率，維持了整體 QoS 的穩定。相比之下，**貪婪法**雖確保盡可能服務更多請求，但容易在高流量時造成資源超載，導致壅塞率飆高；**固定切**

片法則因缺乏彈性，當某類服務流量暴增時無法及時重新分配資源，亦出現較高的壅塞和較低的獎勵。至於引入優先回放的 DQN 方法，其表現雖優於傳統策略，但仍不及 SAC-RAC。根據實驗數據，**SAC-RAC 相較 DQN 在平均獎勵上提升約 25%，壅塞率降低約 84%**，呈現出顯著的性能差距。這證明了最大熵策略配合優先回放機制在此問題上的有效性：智能體能更全面地探索策略空間並更快速地學習關鍵經驗，因而做出更佳的資源允入決策。

值得注意的是，在資源**極為充裕**的特殊情況下，各方法的表現差距會縮小。例如當我們將總切片數提高到原來的兩倍（如從 30 提升至 60）時，所有方法最終都能將壅塞率控制在很低的範圍，且貪婪法與 DQN 的最終收益表現接近。這說明當資源遠大於需求時，即使不採取複雜的分配策略，網路效能也不致於太差。然而在現實網路中，不可能無限制地增加資源來滿足需求。因此，在**資源有限且需求波動**的主流情境下，像 SAC-RAC 這樣能有效適應環境變化並提高資源利用效率的方法才顯得尤為重要。總體而言，實驗結果充分驗證了 SAC-RAC 方法在 5G 高度動態流量條件下的優越性，其能在保障 QoS 的前提下，顯著提升資源分配的收益表現。

5. 結論與未來展望

結論

本研究針對 5G 網路中高度動態的流量環境，提出了一套基於深度強化學習的資源允入控制方案。透過將 Soft Actor-Critic 演算法離散化並融入優先經驗回放與有條件更新等改良策略，所提出的 **SAC-RAC 方法**能有效地在變化劇烈的環境中維持高水平的資源分配效益。同時，SAC-RAC 兼顧了網路服務品質，顯著降低了壅塞發生率，避免網路因過載而癱瘓，達到穩定提供服務的目標。實驗結果證實，在各種模擬的 5G 流量場景中，本研究方法均展現出**最優的平均獎勵和最低的阻塞率**，相較傳統貪婪策略、固定資源切片配置以及深度 Q 網路等方法帶來了更高效且穩定的網路表現。這些成果凸顯了最大熵強化學習策略結合優先經驗回放的威力，說明透過提高演算法對環境變化的適應性和學習關鍵事件的效率，可以顯著優化 5G 網路下的資源管理決策。本論文的主要貢獻

在於提出了 SAC 在離散網路資源分配問題上的創新應用，並證明了該方法在動態環境中優於既有方案的效能表現。

未來展望

隨著未來 5G 及後 5G 時代網路持續演進，預期將出現**更多元且更複雜**的流量模式和服務類型輸入網路。如何讓強化學習代理能夠更迅速、精準地處理激增的各類請求，將是必須面對的課題。未來的研究方向可以從多方面著手優化本研究的方法：（1）**強化模型結構**：針對更複雜的流量特性調整神經網路架構，例如增加網路層數或引入注意力機制，以提升代理對不同流量模式的辨識與決策能力。（2）**特徵工程與編碼**：對輸入狀態資訊採用更有效的表示方法，例如利用 One-Hot 編碼或其他特徵嵌入技術來增強神經網路對不同服務類別的敏感度，幫助代理更好地區分各類流量的特性。（3）**自適應參數調整**：探索讓演算法的重要超參數（如最大熵的權重、優先回放中的參數等）能隨著環境狀況動態調整，減少人工調參的負擔並提高在各種情境下的適用性。

此外，在資源允入控制決策中，**拒絕服務請求**雖可在短期內避免超載，但從長遠看需要慎重考量其對網路**阻斷率**和使用者體驗的影響。未來研究應特別關注如何在保留網路穩定性的同時，將必要的拒絕率降到最低。例如，可以在獎勵函數中引入對拒絕次數的更高懲罰，或者結合預測模型提前預估流量高峰以提前調配資源，從而降低因臨時資源不足而必須拒絕請求的情況。總而言之，未來的工作將致力於讓**強化式學習的 RAC 系統更具泛化能力與智能性**，能適應不斷演變的網路環境並平衡效能與 QoS，為 5G 乃至 6G 時代的網路資源管理提供堅實可靠的技術支撐。