University of Waterloo Pascal Poupart 教授演講：Inverse Constraint Learning and Risk Averse Reinforcement Learning for Safe AI

文／連云暄 資訊科學與工程研究所博士生

講者 Pascal Poupart 教授係加拿大 University of Waterloo David R. Cheriton 計算機科學學院的教授。他同時是加拿大 CIFAR AI 教授，在 Vector 研究院任職，並是 University of Waterloo AI 研究院的成員。他自 2022 年起擔任 Georgia Tech 的 NSF AI 優化進步研究院顧問委員會成員。他曾在加拿大皇家銀行的 Waterloo Borealis AI Research Lab 研究實驗室擔任研究主任和首席研究科學家（2018-2020 年）。他的研究重點是開發應用於自然語言處理和材料發現的機器學習算法。他在強化學習算法開發方面的貢獻尤為著名。他的研究團隊目前正在進行的重要項目包括逆向限制學習、平均場強化學習、強化學習基礎模型、貝葉斯聯邦學習、不確定性量化、機率深度學習、對話代理、轉寫錯誤糾正、體育分析、適應滿足性以及用於二氧化碳回收的材料發現。

在其演講中，Poupart 教授強調了強化學習（RL）和控制系統在實際應用中必須考慮現實生活限制的重要性，並提出了可行的演算法。這些限制條件有助於確保實施的可行性、安全性或關鍵性能指標的閾值。然而，某些限制條件難以具體定義，特別是在如自動駕駛這類複雜應用中，設定目標獎勵函數相對容易，但要明確定義專家駕駛員在確保安全、平穩及舒適駕駛中所遵循的隱性限制則更為困難。

Poupart 教授介紹了逆向限制學習（Inverse Constraint Learning, ICL）的概念。傳統上逆向強化學習（Inverse Reinforcement Learning, IRL）用於學習解釋專家行為的獎勵函數，但在許多實際應用場景中，僅知道獎勵函數並不足夠，還需要理解行為背後的限制條件。這些限制往往能提供比獎勵函數更直觀的行為解釋，例如在安全關鍵的應用中尤為重要。透過逆向工程反求限制條件，可以更深入地了解專家行為背後的隱性邏輯，從而設計出更符合人類行為模式的自動駕駛策略。

教授還探討了如何從專家軌跡中學習 soft constraints，這種方法假設已知獎勵函數並通過專家軌跡學習 soft constraints。在機器學習和強化學習的實際應用中，面對帶有噪聲的感測資料或不完美的專家示範是普遍存在的問題，這要求在資料的信賴度和模型的效能之間找到平衡。soft constraints 與傳統的 hard constraints（如能量使用上限）不同，可允許模型違反限制條件，在獎勵函數與限制條件中取得平衡，因此能使模型有更靈活的應對策略。

此外，Poupart 教授還介紹了一種基於吉尼偏差（Gini deviation）的風險規避強化學習方法。在現實生活中，我們有許多需要避免風險的場合，例如在自動駕駛中避免碰撞，在投資組合管理中則試圖避免巨大的財務損失。傳統的強化學習關注於最大化預期回報，而風險回避強化學習則同時考慮風險控制。吉尼偏差是對傳統基於變異數方法的一個替代方案，能更有效地評估策略執行過程中可能的風險，特別是在高風險的決策環境下。

本演講不僅提供多種新的研究工具，也對於如何在實際應用中實現安全人工智能提出了實用的見解，有助於提升未來應用於開發能自動適應複雜環境和嚴格安全要求的智能系統的可行性。這些見解和方法為機器學習和強化學習領域提供了寶貴的指導，特別是在處理不確定性和風險管理方面的應用，使這些技術更加貼近現實世界的需求和挑戰。

Speech by Dr. Pascal Poupart from the University of Waterloo: Inverse Constraint Learning and Risk Averse Reinforcement Learning for Safe AI

Dr. Pascal Poupart is a professor at the David R. Cheriton School of Computer Science at the University of Waterloo in Canada. He also serves as a CIFAR AI Chair at the Vector Institute and is a member of the AI Institute at Waterloo. Since 2022, he has been part of the advisory board for the NSF AI Research Institute at Georgia Tech. Previously, from 2018 to 2020, he held the roles of Research Director and Chief Research Scientist at the Waterloo Borealis AI Research Lab at the Royal Bank of Canada. His research primarily focuses on developing machine learning algorithms for natural language processing and materials discovery, with a particular emphasis on reinforcement learning. His team is currently engaged in several notable projects, including Inverse Constraint Learning, Mean-Field Reinforcement Learning, Foundational Models for Reinforcement Learning, Bayesian Federated Learning, Uncertainty Quantification and Calibration, Probabilistic Deep Learning, Conversational Agents, Transcription Error Correction, Sports Analytics, Adaptive Satisfaction, and materials to facilitate desirable chemical reactions for CO2 conversion and CO2 capture.

In his speech, Professor Poupart highlighted the crucial role of real-world constraints in the practical implementation of reinforcement learning (RL) and control systems, and proposed feasible algorithmic solutions. These constraints are essential for ensuring the feasibility of implementation, safety, and adherence to key performance indicators. However, some constraints are challenging to define precisely, particularly in complex applications such as autonomous driving. While establishing target reward functions is relatively straightforward, accurately articulating the implicit constraints that expert drivers adhere to for safe, smooth, and comfortable driving proves to be much more difficult.

Professor Poupart introduced the concept of Inverse Constraint Learning (ICL). While Inverse Reinforcement Learning (IRL) traditionally focuses on determining the reward functions that explain expert behavior, this approach is often insufficient in practical applications. Understanding the constraints underlying behavior is equally important, as these constraints frequently provide a more intuitive rationale for actions than reward functions, which is especially crucial in safety-critical contexts. By reverse-engineering these constraints, we can uncover the implicit logic behind expert behavior, enabling the design of autonomous driving strategies that more closely mimic human behavioral patterns.

He also examined methods for learning soft constraints from expert trajectories. This strategy relies on a known reward function and utilizes expert trajectories to derive soft constraints. In real-world applications of machine learning and reinforcement learning, challenges often arise from noise in sensor data or imperfect expert demonstrations. This necessitates a balance between the data's reliability and the model's performance. Unlike traditional hard constraints (like energy usage limits), soft constraints permit the model to occasionally breach certain restrictions, seeking for a balance between the reward function and constraints. As a result, the model can implement more adaptable strategies in response.

Moreover, Professor Poupart presented a risk-averse reinforcement learning approach that utilizes Gini deviation. In various real-life scenarios, such as avoiding collisions in autonomous driving or minimizing substantial financial losses in portfolio management, risk avoidance is crucial. While traditional reinforcement learning focuses on maximizing expected returns, risk-averse reinforcement learning also considers risk management. Gini deviation offers an alternative to conventional variance-based methods, allowing for a more effective assessment of potential risks during strategy implementation, especially in high-risk decision-making contexts.

This presentation introduces an array of new research tools and offers practical insights into implementing safe artificial intelligence in real-world applications. It enhances the feasibility of developing intelligent systems capable of autonomously adapting to complex environments and meeting rigorous safety standards. The insights and methodologies provide valuable guidance for the fields of machine learning and reinforcement learning, particularly in addressing uncertainty and managing risk. This alignment ensures that these technologies are better suited to meet the demands and challenges of real-world applications.