Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Review:

Domain adaptation in reinforcement learning: a comprehensive and systematic study[#]

Amirfarhad FARHADI[1], Mitra MIRZAREZAEE[1], Arash SHARIFI[†‡1], Mohammad TESHNEHLAB[2]
[1]*Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran 1477893855, Iran*
[2]*Department of Control Engineering, K.N. Toosi University of Technology, Tehran 1999143344, Iran*
[†]E-mail: a.sharifi@srbiau.ac.ir

Received Oct. 1, 2023; Revision accepted Mar. 18, 2024; Crosschecked Oct. 29, 2024

Abstract: Reinforcement learning (RL) has shown significant potential for dealing with complex decision-making problems. However, its performance relies heavily on the availability of a large amount of high-quality data. In many real-world situations, data distribution in the target domain may differ significantly from that in the source domain, leading to a significant drop in the performance of RL algorithms. Domain adaptation (DA) strategies have been proposed to address this issue by transferring knowledge from a source domain to a target domain. However, there have been no comprehensive and in-depth studies to evaluate these approaches. In this paper we present a comprehensive and systematic study of DA in RL. We first introduce the basic concepts and formulations of DA in RL and then review the existing DA methods used in RL. Our main objective is to fill the existing literature gap regarding DA in RL. To achieve this, we conduct a rigorous evaluation of state-of-the-art DA approaches. We aim to provide comprehensive insights into DA in RL and contribute to advancing knowledge in this field. The existing DA approaches are divided into seven categories based on application domains. The approaches in each category are discussed based on the important data adaptation metrics, and then their key characteristics are described. Finally, challenging issues and future research trends are highlighted to assist researchers in developing innovative improvements.

Key words: Reinforcement learning; Domain adaptation; Machine learning
https://doi.org/10.1631/FITEE.2300668 **CLC number:** TP391

1 Introduction

Machine learning is a subset of artificial intelligence (AI) that focuses on creating algorithms enabling systems to learn and make predictions or decisions from data without explicit programming (Abdul Samad et al., 2023; Zhang NJ et al., 2023). Within machine learning, neural networks simulate the human brain's interconnected neurons to process information (Farhadi et al., 2023). These networks, forming the basis of deep learning, consist of multiple layers allowing systems to learn patterns and representations from complex data (El Jery et al., 2023). Deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have revolutionized various fields, particularly computer vision, natural language processing (NLP), and pattern recognition, by enabling machines to comprehend, interpret, and generate data with remarkable accuracy and efficiency (Wang HN et al., 2020; Farhadi and Sharifi, 2024).

Reinforcement learning (RL) enables autonomous learning of complex behaviors from

[‡] Corresponding author

[#] Electronic supplementary materials: The online version of this article (https://doi.org/10.1631/FITEE.2300668) contains supplementary materials, which are available to authorized users

ORCID: Arash SHARIFI, https://orcid.org/0000-0002-2441-9477 low-level sensor data (Liu SP et al., 2022). RL is a branch of AI that provides useful tools for optimizing decision sequences for long-term results (Bu and Wang, 2019). It is a trial-and-error learning algorithm in which an autonomous agent uses automatic learning to identify the most optimal solution to a given problem (Khader and Yoon, 2021). Over the last decade, the value of RL has been proven in a series of artificial domains, especially in real-world scenarios (Voulgarelis et al., 2021). Besides, it has demonstrated remarkable effectiveness in resolving different sequential decision-making issues in machine learning (Boute et al., 2022). In most effective RL applications like robotics, autonomous vehicles, and the poker game, more than one single agent is involved (typically considered a multi-agent RL). Multi-agent RL has recently re-emerged as a result of advancements in single-agent RL methods (Li SD et al., 2020; Rajput et al., 2023).

Fundamentally, RL is characterized by interaction with the environment (Sutton and Barto, 2018). Moreover, it has a low speed and may not be effective in complicated situations, particularly those involving continuous state–action spaces in the actual world (Li XT and Sun, 2021). The issue becomes challenging when facing continuous domains where generalization is essential. In this regard, transfer learning (TL) solves such a problem by improving learning performance through offering generalization across different tasks. The key problem in TL involves integrating knowledge gained from learning into different but relevant situations. TL applies knowledge gained from a source task to a target task (Shoeleh and Asadpour, 2017). Domain adaptation (DA) is a form of TL in which knowledge is transferred from one domain to another with sparse or unavailable labels (Di et al., 2018). It learns domain-invariant features by adversarial training or moment matching. Some existing DA approaches assume that label spaces in the source and target domains are similar. Nevertheless, in real-world circumstances such as sentiment analysis, an approach to NLP, finding a source domain labeled similarly to a target domain is challenging (Finn et al., 2017; López et al., 2019; Singhal et al., 2023). In DA, in the context of RL, the Internet of Things (IoT) plays a pivotal role by providing a diverse array of real-world data streams, enabling the adaptation of RL algorithms to variable environments and facilitating the transfer of knowledge across domains for enhanced decision-making capabilities (Pourghebleh and Navimipour, 2017; Pourghebleh et al., 2020).

Various DA approaches suitable for continuous RL problems have been proposed recently. Nevertheless, to our knowledge, no comprehensive or detailed studies have been conducted to analyze the existing methods in this field. This gap motivates us to present the current study with the following major contributions:

1. Provide a comprehensive and detailed review of RL-related DA approaches with state-of-the-art methods and pointers to the most relevant.

2. Categorize the existing methods into three major classes (supervised, semi-supervised, and unsupervised), explore them, and highlight their main features.

3. Specify open issues and make suggestions for upcoming studies.

Table 1 summarizes the main abbreviations used in the paper, along with their definitions.

2 Related surveys

Several reviews have been published in the last 10 years that emphasize the importance of DA in various fields. Although these studies have contributed to a better understanding of several aspects of the DA problem, they have not discussed DA solutions for continuous RL issues. In this section, we review the related studies to highlight the need to conduct the current study to fill the research gaps in this area.

In a review of recent advances in DA for visual recognition presented by Patel VM et al. (2015), the advantages and weaknesses of existing DA methods were discussed, and some promising hints for research in this area were identified. Some theoretical results and existing algorithms for the multi-source DA problem were reviewed by Sun et al. (2015). They specified the datasets used in the proposed algorithms and then made some suggestions for future studies.

Wang M and Deng (2018) introduced TL and DA with a particular focus on visual applications. At first, they highlighted the role of DA in the more general TL issue. Then, they analyzed the recent techniques for various scenarios, described the historical shallow strategies, and addressed the homogeneous and heterogeneous DA approaches. They noted that

Table 1 Abbreviation table

Abbreviation	Definition	Abbreviation	Definition
AC	Actor–critic	GPRL	Gaussian processes reinforcement learning
AI	Artificial intelligence	ITM	Instantaneous topological map
ALE	Arcade learning environment	MDP	Markov decision process
CFD	Contrastive forward dynamics	NLP	Natural language processing
CODAS	Cross-modal domain adaptation with sequential structure	NN	Neural network
		PCA	Principal component analysis
DARLA	Disentangled representation learning agent	PPO	Proximal policy optimization
DDPG	Deep deterministic policy gradient	RL	Reinforcement learning
DQN	Deep Q-network	SDG	Selection distribution generator
DVE	Data value estimator	SVM	Support vector machine
EM	Expectation maximization	WBS	Work breakdown structure
GDANs	Generative domain-adaptive nets		

the development of deep convolutional architectures has resulted in the establishment of a new class of DA techniques, including adaptation inside the deep architecture. Moreover, they reviewed DA methods in various fields of image categorization.

A detailed analysis of DA techniques applied to computer vision was provided by Zhao SC et al. (2022). They proposed a classification of various schemes according to data properties that specify the difference between the two domains. Then, they summarized approaches into groups based on training loss and briefly analyzed the existing techniques under these groups. Furthermore, they provided an overview of computer vision applications beyond image classification. Finally, they highlighted some limitations of existing approaches and proposed several future directions.

Zhao SC et al. (2020) discussed various multi-source DA methods and summarized available datasets for evaluation. They compared the recent multi-source DA methods in the deep learning area, including intermediate domain generation and latent space transformation. Finally, they provided some research directions for the future. Madadi et al. (2020) reviewed DA techniques for classification tasks in computer vision and categorized the proposed unsupervised DA approaches into five classes (representation, reconstruction, adversarial, discrepancy, and attention-based).

Chu and Wang (2020) presented a thorough review of state-of-the-art DA approaches for machine translation. They categorized methods selected into two key groups, i.e., model-centric and data-centric. In the model-centric group, neural machine translation models are specialized for DA, from a training objective-centric, architecture-centric, or decoding-centric perspective. The data-centric group is more concerned with the data being used instead of domain-specific models. Data can be in-domain monolingual corpora, synthetic corpora, or parallel corpora.

Saunders (2022) presented a taxonomy of DA schemes corresponding to developing a neural machine translation system. The author made a distinction between DA and multi-DA, especially when the language domain of interest is not fixed or known, and focused on the twin problems of forgetting and overfitting applied to the existing methods. The strengths of DA framing for other lines of machine translation research have also been emphasized.

Guan and Liu (2022) discussed the DA techniques and challenges associated with medical image examination and summarized the existing methods in a systematic manner according to their characteristics. The methods were categorized into two main groups, i.e., shallow models and deep models. Then, each group was subdivided into three classes, i.e., supervised, semi-supervised, and unsupervised.

3 Domain adaptation in the context of reinforcement learning

3.1 Domain adaptation

Most machine learning models assume that the test and training samples originate from equal distributions. However, there are numerous instances where the distributions of training and test data differ. In this study we concentrate on cases where a model is trained across multiple domains and applied to a different but related domain. Learning in this manner is defined as DA, which is fundamental to machine learning (Gardner et al., 2020). Over the last decade, DA has gained widespread research attention as a perennial problem in various realistic scenarios, including email filtering (Ge et al., 2013), sentiment analysis (Fang et al., 2014), NLP (Jiang and Zhai, 2007), and computer vision. DA is a subset of TL in which labeled input from related domains is used to perform target-domain tasks. The purpose of DA methods is to handle the domain shift or the distribution change (Fig. 1). The risk associated with the target, denoted as $\epsilon_{\rm t}(f) = E_{(x,y)\sim q}[f(x) \neq y]$, can be mitigated by leveraging the supervised data from the source domain. In most works on domain adaptation, the source domain is denoted as $D_{\rm S} = \{x_i^{\rm s}, y_i^{\rm s}\}_{i=1}^{n_{\rm s}}$ consisting of $n_{\rm s}$ labeled samples, where $y_i^{\rm s}$ represents the ground truth of the source data. Also, there is a target domain $D_{\rm T} = \{x_j^{\rm t}\}_{j=1}^{n_{\rm t}}$ with $n_{\rm t}$ samples. Both $D_{\rm S}$ and $D_{\rm T}$ share the same label space, i.e., $y \in \{0, 1, \ldots, M-1\}$, where M denotes the number of classes. In the context of unsupervised domain adaptation (UDA), the marginal distributions of the two domains are not identical, i.e., $P(x^{\rm s}) \neq Q(x^{\rm t})$ (Wei et al., 2021).

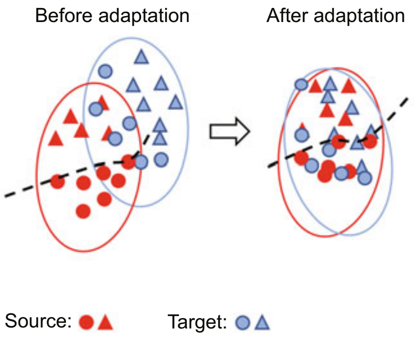

Fig. 1 Schematic of the domain adaptation process (References to color refer to the online version of this figure)

Because of unlabeled training samples and a statistical difference between supervised and unsupervised data, DA extends the semi-supervised learning problem. DA tasks can be carried out solely with source-domain data or with some target-domain samples. Furthermore, training can be accomplished by using only samples labeled in the source domain, thereby preventing the need for adaptation during training (Zhang H et al., 2022).

RL explores how agents make choices in a given environment to optimize their rewards. The environment is often represented as a Markov decision process (MDP) defined by the tuple (S, A, T, R), in which S represents the set of possible states, A represents the set of possible actions, T indicates the transition function, and R stands for the reward function. At each time step t in the MDP, the agent takes an action in the environment depending on the current state s_t. The agent then receives a reward r_{t+1} and transitions it to the next state s_{t+1}. The agent's goal is to create a policy $\pi(s)$ that chooses actions to maximize the sum of future rewards discounted by a factor γ, a value between 0 and 1. In the context of RL, DA entails establishing source and target domains, referred to as $D_{\rm S}$ and $D_{\rm T}$, respectively. Each domain is associated with an MDP, defined by the tuple (S, A, T, R). The MDPs in the source and target domains are denoted by $(S_{\rm S}, A_{\rm S}, T_{\rm S}, R_{\rm S})$ and $(S_{\rm T}, A_{\rm T}, T_{\rm T}, R_{\rm T})$, respectively. Although the source and target domains may have distinct state spaces S, it is important for their action spaces A to stay the same. Additionally, their transition functions T and reward functions R should show similarity due to shared internal dynamics. More precisely, the emphasis is on the transfer of policies when the source policy $(T_{\rm S})$ is approximately equal to the target policy $(T_{\rm T})$, the source resources $(R_{\rm S})$ are approximately equal to the target resources $(R_{\rm T})$, the source actors $(A_{\rm S})$ are equal to the target actors $(A_{\rm T})$, but the source strategy $(S_{\rm S})$ is not equal to the target strategy $(S_{\rm T})$.

Autonomous driving may be used as an example to demonstrate how different weather conditions correlate to different domains. For example, the source domain may correspond to driving under clear skies, whereas the target domain relates to driving in wet conditions. While the visual observations in the status space S may vary depending on weather conditions, the action space A, which includes throttle and steering, stays constant. The resemblance between the transition function T and reward function R is crucial since both domains rely on traffic circumstances and driving control to determine state transitions. Additionally, the movement of the vehicle plays a significant role in determining the reward function for both domains.

Within an unsupervised situation, we are dealing with partial domain adaptation. In this context, we have a labeled source domain $D_\mathrm{s} = \left\{(x_i^\mathrm{s}, y_i^\mathrm{s})\big|_{i=1}^{N_\mathrm{s}}\right\}$, which is sampled independently and identically from the source distribution $p(x)$, where y_i^s belongs to the set Y_s. At the same time, there is an unmarked target domain $D_\mathrm{t} = \left\{x_j^\mathrm{t}\big|_{j=1}^{N_\mathrm{t}}\right\}$ that is randomly picked from the target distribution $q(x)$. In this context, N_s and N_t denote the numbers of occurrences in the source and target domains, respectively. The target class label space Y_t is a subset of the source class label space Y_s, denoted as $Y_\mathrm{t} \subset Y_\mathrm{s}$. The classes in Y_s that are absent in Y_t are termed outlier classes, while those present in both Y_s and Y_t are referred to as shared classes. Crucially, the data distributions of the source and target domains are different, shown as $p(x) \neq q(x)$. The domain adversarial RL architecture (Fig. 2) seeks to choose source instances that have class labels $y_i^\mathrm{s} \in y_\mathrm{t}$. Consequently, it aims to acquire transferable characteristics from the selected source examples and target instances inside the same label space Y_t. This section categorizes the current studies on DA in the context of RL.

By examining 17 research studies in this area, a clear trend in DA development could be discerned. For this purpose, the innovations, merits, limitations, and distinctions of the proposed strategies are discussed. The approaches selected are classified based on different application domains. The most important qualitative parameters and requirements for a DA method are defined below:

1. Performance: it provides insights into the effectiveness of the approach (Zhao N et al., 2024).

2. Optimizability: the method is simple to train and needs little tuning of hyperparameters (Mou et al., 2023).

3. Data dependency: the method is trainable using small datasets (Liu Q et al., 2021).

4. Data scalability: the method is appropriate for large and complicated datasets with a wide range of data (Saeed et al., 2022).

5. Task scalability: the method is applicable for solving complex tasks such as object detection and semantic segmentation (Guo et al., 2024).

6. Efficiency: this highlights the computation cost of training and evaluating data adaptation methods (Li X et al., 2021).

3.2 Reinforcement learning and dialogue systems

RL is pivotal in developing NLP applications, particularly in dialogue systems. In this context, RL is the driving force behind agent-based conversations, allowing intelligent systems to learn optimal decision-making strategies through user interaction. These dialogue agents use RL algorithms to navigate the vast space of possible responses, selecting actions that maximize cumulative rewards while engaging in conversations. RL-based dialogue systems have shown promise in various domains, from virtual assistants and customer service chatbots to language tutoring platforms. By continuously learning and adapting their responses, RL-driven dialogue systems aim to provide more natural, context-aware, and effective interactions, ultimately enhancing user experiences in human–computer conversations.

3.2.1 Sample-efficient neural network methods

Su et al. (2017) introduced two sample-efficient neural network methods for accelerating dialogue policy optimization. These methods were designed to enhance the efficiency of learning in dialogue systems.

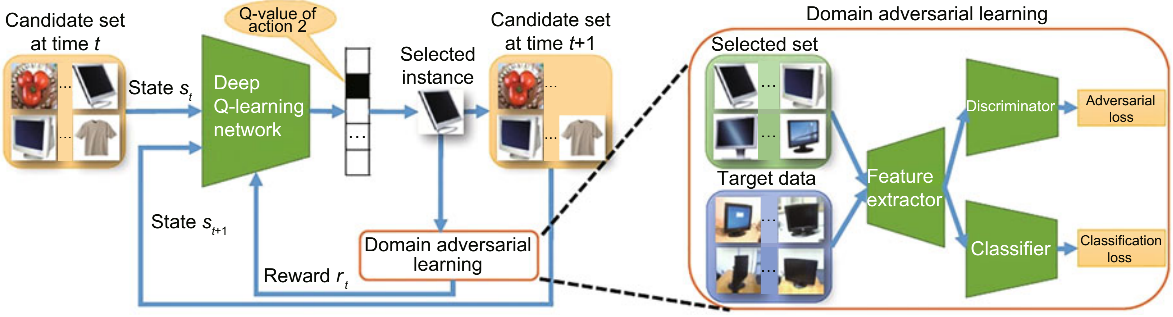

Fig. 2 Domain adversarial reinforcement learning (RL) architecture

The first method is trust area based learning step adjustment. This method relies on the trust area concept to control learning step size and prevent drastic model change. It aims to maintain model stability during learning. However, it may lead to slower learning progress due to its cautious nature.

The second method is natural gradients for rapid convergence. This method uses natural gradients to identify the optimal direction for fast convergence. It enhances sample efficiency by quickly converging to solutions. However, it may be less stable and prone to getting stuck in suboptimal solutions.

Both methods leverage off-policy learning and experience replay to improve sample efficiency. Nonetheless, their neural network architectures may struggle to handle uncertainty, particularly in noisy environments, compared to alternatives like Gaussian processes reinforcement learning (GPRL) (Gašić and Young, 2014). Overall, the proposed methods offer a practical way of learning dialogue policies based on deep RL and demonstrate their efficiency in task-oriented information seeking. However, there is room for improvement in handling uncertainty and noisy environments.

Modular spoken dialogue systems can benefit from the proposed framework, which focuses on the dialogue management component. The model takes the belief state b as input, representing a distribution over the possible user intents and the dialogue history. The primary objective of the model is to select the system action a at each turn, which can maximize the cumulative reward and lead to a successful dialogue outcome. The system action is then transformed into a semantic-level system reply, which is subsequently passed to the natural language generator for output to the user. The semantic reply comprises three components: the response intent (e.g., inform), the relevant slots to be discussed (e.g., area), and a corresponding value for each slot (e.g., east). To ensure tractability, the policy restricts the selection of action a from a limited action set, which identifies the intent and sometimes a slot. Any additional information required to complete the reply is extracted from the tracked belief state using heuristics.

3.2.2 Generative domain-adaptive nets (GDANs)

Yang ZL et al. (2017) investigated the issue of semi-supervised question answering. They developed a new neural framework, GDANs, to better use unlabeled text. The proposed framework can be summarized as follows:

1. Answer chunk extraction. Possible answer chunks are extracted from unlabeled text using linguistic tags.

2. Question generation. A sequence-to-sequence model generates questions based on answer chunks and their contexts.

3. Discriminative model training. Model-generated and human-generated question–answer pairs are merged to train a discriminative model. However, the data distribution produced by the model may differ from human-supplied data, leading to suboptimal models.

To solve this problem, Yang M et al. (2018) proposed two DA methods that consider the data distribution generated by the model as a distinct domain. First, an additional domain tag was used to denote whether a model or a human being produces a question–answer pair. They conditioned the discriminative model using the domain tags to train it to factor out domain-specific and domain-invariant representations. In addition, they used the RL method to fine-tune the generative model in an adversarial manner to diminish the loss of the discriminative model. Moreover, they provided a straightforward and efficient baseline technique for semi-supervised question answering. The baseline technique performs worse than the proposed GDANs strategy, but it is very simple to apply and can still lead to significant improvements when only restricted labeled data are available. The authors tested their innovation on the SQuAD dataset with different labeling rates and amounts of unlabeled data. Their findings indicated that the proposed GDANs framework significantly outperforms both the supervised learning setting and baseline techniques like adversarial DA and dual learning.

3.2.3 Deep Q-learning for semi-supervised domain adaptation

Patel Y et al. (2018) used deep Q-learning to tackle the issue of semi-supervised DA of classification algorithms. The central idea was to use a source domain network's prediction on target-domain data as noisy labels and train a sampling strategy that maximizes classification accuracy on a small, annotated reward partition of the target domain. In contrast to recent related techniques, the authors used fixed representations for both the source- and target-domain samples, limiting the performance of the suggested approach. To address this issue, they included a labeler in the Q-agent to be tuned in conjunction with the existing sampler to achieve better representations and fewer noisy labels for the target domain. Another strategy was to learn sampling policies from the representations generated by unsupervised DA using current feature alignment methods.

3.2.4 Skill-based transfer learning in continuous reinforcement learning

Shoeleh and Asadpour (2020) presented a new skill-based TL method that applies DA to continuous RL problems. The suggested method obtains high-level skills from the source task and then uses a DA strategy to assist the agent in discovering the mapping between states and actions as a link between the source and target tasks. Source skills can be adapted, and learning on a new target task can be accelerated by using such mapping. The suggested method consists of three main stages:

1. Source task learning. Abstract skills are derived from the source task using a graph-based skill learning framework.

2. Domain adaptation. A three-dimensional (3D) mapping between the source and target task's state–action spaces is established using DA techniques to reduce distribution differences.

3. State–action mapping. State–action mapping across tasks is established to integrate skills with associated value functions into the target task.

Two approaches are presented for state–action mapping, using multi-layer perceptrons for explicit mapping discovery and using Q-values for offline learning of the value function of target skills. Experiments demonstrated the method's ability to transfer abstract skills and enhance performance in target tasks through knowledge transfer.

3.3 Data valuation in machine learning and domain adaptation

Data valuation in machine learning and DA are closely interconnected fields that underscore the importance of data quality and relevance. In machine learning, understanding the value of data is pivotal for model training and decision-making. It involves assessing the impact of data on model performance, allowing organizations to prioritize data collection and curation efforts effectively (Bolhassani and Oksuz, 2021). On the other hand, DA deals with transferring knowledge from one domain to another, often with variations in data distribution. Here, data valuation plays a critical role in identifying which data from the source domain are most valuable for adaptation to the target domain. By combining these two concepts, organizations can optimize their data-driven strategies, ensuring that valuable data are leveraged for DA processes. This ultimately leads to more robust and accurate models in varying real-world scenarios. This synergy between data valuation and DA is essential for building adaptable, high-performance machine learning systems.

As depicted in Fig. 3, a neural network is used to estimate the value of data. By using data valuation in DA, it becomes possible to detect outlier and noise data that may lead to domain shift.

3.3.1 Data valuation using RL

Yoon et al. (2020) introduced data valuation using RL (DVRL) as a novel approach to learning data values in conjunction with predictive models, adapting to the task. DVRL leverages two learnable functions:

1. Target task predictor model. This model is trained on training data to minimize the weighted loss function.

2. The data value estimator (DVE) model. DVE plays a crucial role in DVRL by calculating the likelihood of each data point being used for training the prediction model. DVE is trained using a reinforcement signal that reflects performance on the target task.

DVRL's distinct advantage lies in its scalability. It has been shown to work efficiently on large-scale datasets like WideResNet-28-10, ResNet-32, and CIFAR-100. Evidence shows that DVRL outperforms existing techniques for estimating data values across diverse datasets and application scenarios.

3.3.2 Semi-supervised domain adaptation with selective pseudo-labeling

Liu BY et al. (2020) designed a technique for semi-supervised DA based on RL and selective

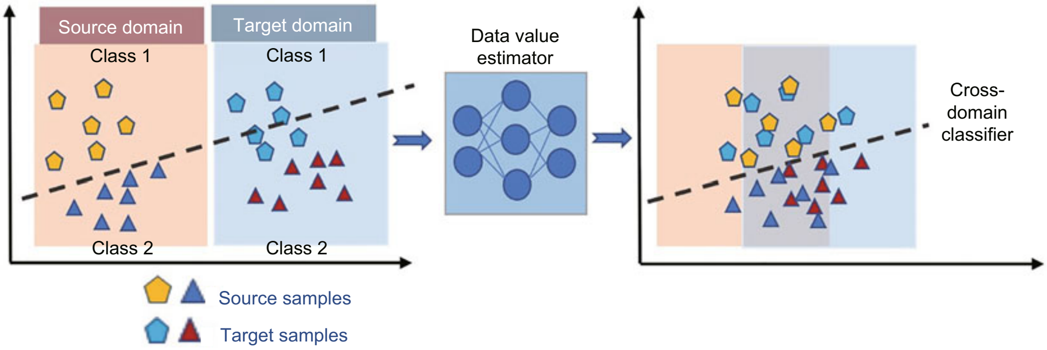

Fig. 3 Data valuation in domain adaptation (References to color refer to the online version of this figure)

pseudo-labeling. Traditional pseudo-labeling methods often struggle to balance the correctness and representativeness of pseudo-labeled data. To address this challenge, the authors introduced the following components:

1. Deep Q-learning model. This model is designed to select pseudo-labeled instances accurately and representatively. It leverages RL principles to make informed choices in pseudo-labeling, enhancing the quality of the labeled data.

2. Target margin loss. The authors proposed a new target margin loss during training to improve the base model's discriminative power. This loss function encourages the model to acquire discriminative features even with the minimum labeled data.

The suggested approach outperforms previous methods in various benchmark datasets for semi-supervised DA, showcasing its effectiveness in leveraging unlabeled data while maintaining model accuracy.

3.4 Sim2real transfer

Sim2real transfer is a pivotal concept in robotics and AI. It represents an attempt to bridge the gap between simulated environments and the complexities of the real world. In sim2real transfer, models and algorithms trained in simulated environments are adapted and fine-tuned for deployment in real-world scenarios. This process involves addressing challenges such as the reality gap, where the virtual world does not perfectly mimic real-world conditions. Achieving successful sim2real transfer ensures that AI systems, particularly robots, can function effectively in unpredictable and dynamic environments, such as factory floors or disaster response situations, or when used in autonomous vehicles. Sim2real transfer leverages the benefits of controlled, cost-effective simulations for initial training while acknowledging the need for adaptation to handle real-world nuances and uncertainties. It represents a critical step toward realizing the full potential of AI and robotics in practical applications.

3.4.1 Cross-modal domain adaptation with sequential structure (CODAS)

Chen HX et al. (2021) introduced CODAS, a novel technique that addresses the sim2real challenge. The primary goal is to learn a mapping function that bridges images in the target domain with states in the source domain. This learned mapping function enables policies trained on states in the source domain to be directly applied in the target domain, which consists of images. Key contributions of CODAS include:

1. Sequential variational inference problem. CODAS formulates the problem as a sequential variational inference challenge, capitalizing on the sequential nature of RL problems.

2. Optimization objectives. A set of practical optimization objectives is developed in CODAS to facilitate the mapping process. These objectives are designed to adapt policies across domains effectively. CODAS differs from previous approaches that rely on image input to learn state embeddings. Instead, it focuses on mapping images to specific subspaces within a low-dimensional vector space.

3.4.2 Bi-directional domain adaptation

Truong et al. (2021) introduced a bi-directional DA strategy that leverages simulated and real-world data to expedite learning and enhance the generalization of RL policies. To mitigate visual and dynamic domain gaps, they used DA techniques in two directions:

1. Real2sim. In this direction, the aim is to reduce domain differences between real-world and simulated data. Image translation techniques, specifically CycleGAN, transfer images from the real environment to the simulator. This approach effectively bridges visual domain gaps.

2. Sim2real. In the opposite direction, sim2real adaptation is used to learn residual errors in dynamics from simulation and apply them to the real world. CycleGAN, with a cycle-consistency loss function using unpaired images, is instrumental in achieving this goal.

The real2sim strategy, in particular, outperforms sim2real for visual DA. This approach separates the sensor adaptation module from RL policy training, reducing additional bottlenecks in the RL policy learning process. The authors focused mainly on enabling efficient sim2real transfer for point goal navigation. In this scenario, a robot was placed in an unfamiliar environment and tasked with navigating to a specified destination using only egocentric RGB-D observations within a limited time frame.

3.5 Visual control and robotics

Visual control is a fundamental aspect of robotics, enabling machines to interact with and navigate the physical world based on graphical sensory input. In robotics, visual control systems rely on cameras or other vision sensors to perceive their surroundings, interpret visual data, and make decisions or take actions accordingly. This technology is pivotal in various robotic applications, such as autonomous vehicles, industrial automation, and household robots. Visual control allows robots to perform tasks with exceptional precision, such as object recognition, obstacle avoidance, path planning, and fine-grained manipulation. It empowers robots to adapt to dynamic environments, respond to changing conditions, and collaborate with humans more effectively. As visual control systems continue to advance, robotics is poised to make even greater strides in enhancing efficiency, safety, and autonomy across various industries and everyday life.

3.5.1 Addressing the reality gap through image domain conversion

Zhang JW et al. (2019) introduced an innovative strategy to tackle the reality gap. Rather than attempting to enhance the visual quality of synthetic images generated by simulators during training, their approach concentrates on converting real-world image streams back into the synthetic domain during deployment, creating a seamless integration between simulated and real-world domains. Key features of this approach include:

1. Efficiency and flexibility. This technique offers an efficient, flexible, and lightweight solution for visual control, eliminating the need for additional transfer steps during the costly RL agent training phase in the simulation.

2. Environment independence. Trained RL agents are not limited to deployment in a specific real-world environment. The separation of policy training and transfer processes allows for concurrent execution.

3. Shift loss for consistency. The authors introduced a simple yet effective shift loss that operates independently of downstream tasks. This loss enforces consistency across consecutive frames, ensuring stable policy output. It is particularly valuable for tasks like artistic style transmission in videos and DA.

The proposed visual control method was rigorously validated through indoor and outdoor robotics tests, affirming its practicality and effectiveness.

3.5.2 Unsupervised domain adaptation for visual navigation

While significant progress has been made in learning-based navigation strategies within simulated environments, the real-world implementation of these policies presents a practical challenge. Li SD et al. (2020) developed an approach for visual navigation based on unsupervised DA to bridge this gap. Key aspects of their technique include:

1. Consistent domain transformation. The strategy involves transforming images from the target domain to the source domain in a consistent manner, aligning with the representations learned by the navigation policy.

2. Transfer performance. Experimental results demonstrated that this approach outperforms

baseline methods in transferring navigation policies across various tasks. These transfers were evaluated between two simulation domains and from simulation to the real world.

These methods mark significant advancements in overcoming the reality gap, enabling RL policies developed in simulations to operate effectively in the real-world settings, particularly visual control and navigation.

3.6 Text and language processing

Text and language processing is a transformative field of AI that focuses on understanding and manipulating human language. It encompasses a wide range of applications, from natural language understanding and sentiment analysis to machine translation and chatbots. Text and language processing systems use NLP and machine learning to extract valuable insights from textual data, enabling automated language understanding and generation. These technologies have revolutionized how we interact with computers, making machines able to comprehend, generate, and respond to human language meaningfully. As textual data grow exponentially in the digital age, text and language processing plays an increasingly crucial role in information retrieval, content summarization, and even medical diagnosis. It promises to shape the future of human–computer interaction and information management, making it a vital area of research and development in AI.

3.6.1 Personalized response generation model with dual learning

Yang M et al. (2018) proposed a personalized response generation model that combines dual learning and DA theories. The method uses a dual learning mechanism to capture human response styles from generic data and fine-tunes the model using a smaller dataset for personalized conversations. Key features of this approach include:

1. Simultaneous optimization. Dual learning enables the concurrent optimization of both response generation and post-generation models, recognizing their interdependence and providing performance benefits, especially in cross-domain scenarios.

2. Efficient target-domain data usage. The approach efficiently harnesses target-domain data, enhancing model customization.

3. Reward-based optimization. Four rewards are introduced to evaluate response quality, and policy gradient techniques are used to generate highly rewarded responses. This approach combines the power of seq2seq models for semantic understanding with RL capabilities to optimize for superior responses.

Extensive testing on real-world datasets showed that the proposed model surpasses previous techniques in generating customized responses for diverse users. It also outperforms state-of-the-art conversational systems in terms of bilingual evaluation understudy (BLEU) scores, perplexity, and human assessments.

3.6.2 Self-supervised domain adaptation for robot manipulation

Jeong et al. (2020) introduced a self-supervised DA approach that leverages unlabeled real robot data to enhance sim2real transfer learning for cube stacking tasks. Key components of this approach include:

1. Contrastive forward dynamics (CFD). It combines dynamic model learning with time-contrastive strategies to use the structure in unlabeled robot data for DA effectively. Results indicated that using the CFD objective for adaptation outperforms other methods, such as domain randomization and adversarial techniques.

2. Enhancing visual layers. Successful sim2real transfer for robotic manipulation involves improving the initial visual layers of the policy network while optimizing RL.

3. Sequence-based self-supervised loss. It yields the best DA results for manipulation tasks by exploiting the dynamic nature of the robotic system.

3.6.3 Reinforcement learning based data selection and representation

Liu MF et al. (2019) presented an RL framework for data selection and representation learning. Key features of this approach include:

1. Selection distribution generator (SDG). It conducts data selection and is updated based on rewards from the chosen data instances. A predictor within the framework ensures that a task-specific model can be trained on the selected data and provides feedback for reward computation.

2. Efficiency and generalizability. Experimental results across various NLP tasks, including sentiment analysis, dependency parsing, and part-of-speech tagging, demonstrated the efficiency of data selection and representation learning. The approach shows generalizability across different NLP tasks.

3.6.4 Domain adaptation with source instance selection

Chen J et al. (2022) proposed a DA technique that automatically selects source instances in shared classes to avoid negative transfer and minimize domain shift. Key components of this approach include:

1. Deep Q-learning. It is used to learn policies for selecting source instances by approximating the action–value function.

2. Adversarial domain learning. The agent learns domain-invariant characteristics of selected source and target instances and determines rewards based on the relevance of the chosen source instances to the target domain.

These methods represent significant advancements in personalized response generation, self-supervised DA, data selection, and transferable feature learning, contributing to machine learning and NLP.

3.7 Representation learning and disentangled representations

Representation learning is fundamental in AI and machine learning, emphasizing extracting meaningful and abstract features from raw data. It revolves around learning a suitable data representation that can facilitate various downstream tasks, such as classification, clustering, and generative modeling (Monjezi et al., 2023). Disentangled representations, a specific branch of representation learning, take this idea further by aiming to tease apart the underlying factors of variation in data. In other words, they strive to represent data in a way that separates independent and interpretable factors, like object identity, pose, or lighting conditions. Disentangled representations have gained significant attention due to their potential to enhance model understanding, generalization, and transferability. They find applications in diverse fields, from computer vision to NLP, where the ability to separate meaningful factors can lead to more robust and controllable AI systems. Pursuing disentangled representations remains a prominent research direction, offering promising prospects for advancing machine learning and AI capabilities.

3.7.1 Disentangled representation learning agent (DARLA)

Higgins et al. (2017) introduced DARLA, a multi-stage RL agent that excels in DA scenarios, outperforming various baselines. DARLA's vision is centered on acquiring a disentangled understanding of the observed environment, enabling it to capture shared representations between source and target domains without relying on target-domain data. The DARLA approach consists of three key steps:

1. Learning to see. In the initial phase, DARLA learns to perceive the environment, breaking it down into fundamental visual concepts such as objects, positions, colors, and more, creating a disentangled visual representation.

2. Learning to act. In the second step, the agent builds a robust source policy using the disentangled visual representation.

3. Transfer. Armed with the DARLA source policy, the agent exhibits enhanced adaptability to domain shifts, significantly reducing performance degradation in the target domain.

DARLA's innovative approach to disentangled representation learning contributes to its superior performance in DA scenarios, setting it apart from traditional methods relying on target-domain data.

3.7.2 Adversarial initialization and domain adaptation

Carr et al. (2019) devised an algorithm for initializing hidden feature representations in the target task. Their DA method facilitates the transfer of state representations across domains, tasks, and action spaces. The key components of their approach include:

1. Adversarial DA. The authors used adversarial DA concepts, coupled with an adversarial autoencoder architecture, to match the representation space of new policies with a pre-trained source policy using target task data generated by a random policy.

2. Initialization enhancement. The initialization phase significantly enhances the process of learning

a new RL task, emphasizing the broad applicability of adversarial adaptation techniques in improving transfer learning.

These methods represent significant advancements in DA for RL scenarios, offering efficient and effective approaches to bridging domain gaps and facilitating knowledge transfer across diverse environments and tasks.

3.8 Semantic representation learning

Semantic representation learning is a critical area of study within AI and NLP that focuses on capturing the meaning of words, phrases, or entire sentences in a way that machines can understand and manipulate. At its core, it seeks to bridge the gap between the rich, nuanced semantics of human language and the numerical representations required for computational tasks (Khodayari et al., 2019).

Techniques in semantic representation learning include word embeddings, sentence embeddings, and contextual embeddings like bidirectional encoder representations from Transformers (BERTs) and generative pre-training Transformers (GPTs). These methods enable machines to recognize words and sentences and understand their relationships, contexts, and associations. Semantic representation learning has revolutionized various NLP applications, including sentiment analysis, machine translation, and question–answering systems, making it a foundational pillar of modern AI that continues to advance our ability to process and generate human language (Bagheri, 2021).

3.8.1 Critical semantic-consistent learning model

Existing unsupervised DA techniques overlook that not all semantic representations are transferrable between domains. Therefore, non-transferable knowledge severely impedes domain-wise transfer. Also, due to class-agnostic feature alignment, these approaches fail to limit category-wise distribution shifts. To resolve these concerns, Dong et al. (2020) introduced a novel critical semantic-consistent learning model that minimizes the difference between category-wise and domain-wise distributions. Key components of their approach include:

1. Critical semantic-consistent learning. Existing unsupervised DA techniques often overlook that not all semantic representations are readily transferable between domains. To address this limitation, the critical semantic-consistent learning model was introduced. It aims to minimize the differences between category-wise and domain-wise distributions, effectively aligning semantic representations in a more domain-aware manner.

2. Transfer-based adversarial paradigm. The authors proposed a transfer-based adversarial paradigm that emphasizes the transferability of domain-specific knowledge while disregarding non-transferable knowledge. Despite the possibility of negative transfer due to non-transferable knowledge, the approach leverages a transferability-quantizer to maximize positive transfer gains under RL settings.

3. Symmetric soft divergence loss. A symmetric soft divergence loss was introduced to explore inter-class connections and facilitate category-wise distribution alignment. This loss mechanism uses a confidence-guided pseudo-label generator for target samples, enhancing the alignment of category-wise distributions.

This innovative approach aims to enhance unsupervised DA techniques by considering the selective transfer of knowledge, improving domain-wise transfer, and mitigating negative transfer effects. By addressing the challenge of non-transferable knowledge and category-wise distribution shifts, the proposed model offers promising prospects for improved DA.

4 Discussion

As previously mentioned, the existing DA methods applied to RL can be reviewed based on their different application domains. Here, we compare them in detail with the aim of helping develop effective schemes in this field by guiding the directions of future studies. Table 2 summarizes the application criteria, strengths, drawbacks, and resources consumed of methods in each category. Table 3 presents a comparison of the methods based on important DA metrics. The comparisons of the main characteristics of the discussed DA methods, including the application, experimental platform, dataset, function, and algorithm used by each method are detailed in Table S1 in the supplementary materials. Overall, the following information can be presented regarding DA techniques:

Table 2 An overview of supervised, semi-supervised, and unsupervised learning

Category	Application criteria	Strengths and drawbacks of the application	Resource consumption
Supervised	It is useful only when training set labels are provided.	Its accuracy can be confirmed effectively, but it is challenging to realize new knowledge.	Significant time consumption, high memory use, label cost, and CPU requirements.
Semi-supervised	The use of a supervised learning algorithm is not effective when labels are not available. If two criteria are fulfilled, it is preferable to use semi-supervised learning, in which unlabeled data outnumber the labeled data.	It is a kind of compromise algorithm that is capable of slightly validating the analytical results and discovering some new knowledge.	Human intervention could consume more resources and time compared to supervised learning, depending on the methods.
Unsupervised	It operates without labels on simple or small datasets but rapidly and without previous training.	It can discover more new knowledge, but it is not easy to verify its accuracy effectively.	No additional time or resources are needed during the analysis.

Table 3 A side-by-side comparison of domain adaptation (DA) approaches

Category	Reference	Performance	Optimizability	Data dependency	Data scalability	Task scalability	Efficiency
Supervised	Su et al. (2017)	✓	✓	✓	✓		
	Liu MF et al. (2019)	✓	✓		✓	✓	✓
	Yoon et al. (2020)	✓	✓	✓	✓	✓	✓
Semi-supervised	Yang ZL et al. (2017)	✓		✓	✓		
	Patel VM et al. (2015)		✓		✓	✓	
	Shoeleh and Asadpour (2020)	✓			✓		✓
	Liu BY et al. (2020)	✓		✓	✓		
	Truong et al. (2021)	✓	✓	✓	✓		
Unsupervised	Higgins et al. (2017)	✓				✓	
	Carr et al. (2019)	✓				✓	
	Yang M et al. (2018)	✓	✓		✓		
	Jeong et al. (2020)	✓	✓				
	Chen HX et al. (2021)	✓			✓		
	Zhang H et al. (2022)	✓			✓	✓	
	Zhang JW et al. (2019)	✓			✓	✓	
	Dong et al. (2020)	✓			✓	✓	✓
	Li SD et al. (2020)	✓				✓	

1. The datasets used to assess designed algorithms and approaches;

2. Experimental platforms and simulators applied to implement and investigate the suggested approaches;

3. Evaluation factors considered to assess the performance of approaches;

4. Different algorithms and functions used in designing the approaches.

In supervised learning, the system first learns from previously labeled data to generate an algorithmic model. Then, it uses the model to infer and make decisions (e.g., rate, score, modify learning supply for students, or give feedback). We investigate three related works (Su et al., 2017; Yoon et al., 2020; Chen HX et al., 2021) that used the supervised method for DA in the context of RL. This type of learning involves more human efforts than other techniques because the training data are typically labeled (e.g., scored/classified) by humans, which is time-consuming. The system uses the labeled data to obtain a collection of characteristics, which are then used to produce the algorithm. These characteristics are generally shared by the training and test data, thereby enabling the algorithm created on the basis of the training data to subsequently assess the test data.

Semi-supervised machine learning describes a situation in which a portion of the training data are labeled and can be used to examine the structure of the input data or detect patterns in the unlabeled data. This approach has significant potential for enhancing the accuracy of the unsupervised approach while minimizing the burden of time and expense associated with data preparation for the supervised method.

Unsupervised machine learning is another effective strategy for establishing automaticity in scientific evaluation. Unlike supervised machine learning, this type of learning requires no training with labeled data and therefore it reduces the amount of human effort. The machine is assumed to recognize the latent structure, distribution, or patterns existing in the dataset based on similarities and variations in the characteristics of individual instances in designing an algorithm. The generated algorithm can evaluate the test data when the training and test data originate from the same sample and share a latent structure or pattern. So, unsupervised learning here means that the system attempts to determine patterns in the training data without external labels.

4.1 Adopted algorithms and functions

To develop the algorithm, researchers must first select and modify the algorithms. Depending on the particular aim and circumstances, some algorithms may perform better than others (e.g., training data and sample size). More than 10 algorithms were used in the reviewed studies to develop DA approaches (Table S1). The platforms or programs used to design and implement the algorithms are also essential for attaining automation. The intended functions of the programs, their availability, and the amount of human input needed to use them are the main differences among the programs. According to the experimental platform column in Table S1 and as specified in Fig. 4, researchers have implemented their innovations in different environments like Python, MuJoCo, Habitat, FCN-8s, Gazebo, and ALE. As indicated in the function column in Table S1, different functions have been used to design DA methods, including prediction, mapping, auto encoding, and sequence-to-sequence (Fig. 5). Among the existing functions, it is clear that the prediction function used in 53% of cases was the most often applied.

4.2 Application scenarios and datasets

The discussed DA methods address different application tasks such as dialogue management, robotic grasping, question–answering, visual recognition, Atari games, NLP, and visual navigation.

1. Dialogue management. Speech-based human–computer interaction needs to solve many issues to acquire widespread acceptance. Controlling the dialogue flow effectively and naturally is one of the dialogue management challenges. In contrast to human design, which is error-prone, labor-intensive, and non-portable, automatic design brings about numerous benefits. Automatic learning offers an appealing alternative to the more time-consuming and costly learning dialogue strategies of real users (Liu X et al., 2023a).

2. Robotic grasping. The objective of robotic grasping is to determine the configurations of a robot and its end-effector based on sensor data such that its end-effector grasps a defined target. Many factors affect robotic grasping, including the gripper mechanism, object representation, and tasks (Liu X et al., 2023b). So, it is more rational and persuasive that the robot identifies the dataset by grasping the item by itself.

3. Question–answering. The question–answering system refers to the technology that enables a computer to interpret user-input queries and provide appropriate answers. The system retrieves short-text extracts or phrases from a large number of documents, including the answer itself, to answer any question. To address this issue, a number of datasets have been introduced related to open-domain textual question answering, including Quasar-T (Dhingra et al., 2017), TriviaQA (Joshi et al., 2017), SearchQA (Dunn et al., 2017), and SQuAD-open (Chen DQ et al., 2017). Table 4 provides details about these datasets.

4. Visual recognition. It refers to the ability to recognize and localize visual categories such as gestures, actions, emotions, human expressions, attributes, places, objects, and persons, as well as object interactions and relations in videos or images (Jannat et al., 2023). It plays a significant role in establishing the fundamental knowledge needed to build strategies for interacting with the environment and making decisions about possible actions to achieve objectives. Since training and test data

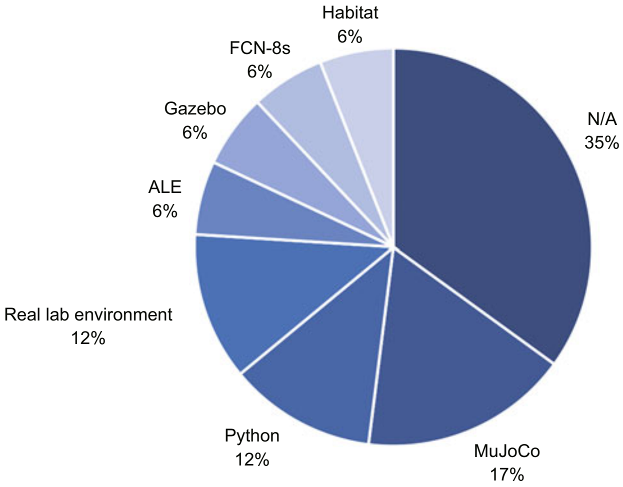

Fig. 4 Experimental platforms and environments for the domain adaptation process

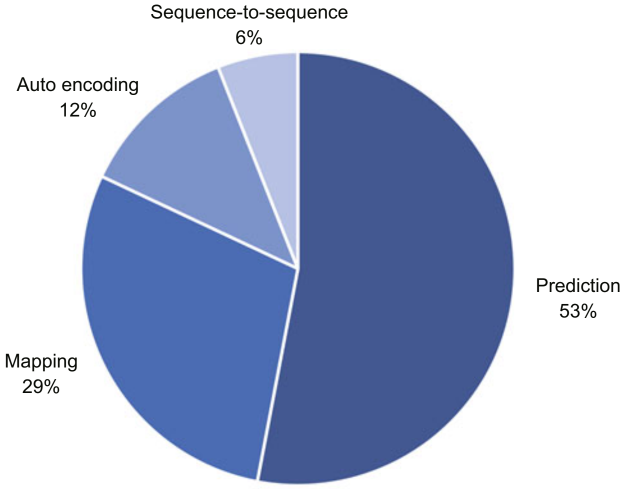

Fig. 5 Functions used to design domain adaptation methods

Table 4 Datasets used in question–answering

Dataset	Feature	Context source	Question source	Query form	Answer form	Granularity
Quasar-T	Contain 4.3×10^4 open-domain trivia questions and answers derived from a variety of online sources.	CluWeb09	Free database	Free database	Span	Paragraph level
TriviaQA	Contain 6.5×10^5 triples of context–question–answers divided into three domains: unfiltered, Wikipedia, and web.	Wikipedia and Bing search results	Trivia websites	Full question	Span	Document level
SearchQA	Comprise over 1.4×10^5 question–answering pairs, each containing 49.6 clips.	Google search results	Jeopardy	Phrase/ keyword	Span	Paragraph level
SQuAD-open	The open-domain version of SQuAD. Only question–answering pairs are provided, while evidence documents are taken from Wikipedia.	Wikipedia	Crowd-sourced	Full question	Span	Document level

distributions often differ, it is essential to adapt the DA to visual recognition problems.

5. Atari games. They are a typical benchmark for state-of-the-art RL algorithms and to test explanation approaches for such algorithms. The intrinsic mechanisms of Atari games vary immensely, so it is difficult to find a single algorithm and hyperparameter setting for all games, regardless of their video frame type.

6. NLP. This refers to a collection of methods for making the human language accessible to computers. NLP initially enables the interaction of AI and human–computer. The main focus of NLP is technology-driven machine translation and language recognition and their incorporation into real-world applications. The main related research areas include social media data and dialogue systems. Nevertheless, it is difficult to train deep frames, and shallow learning approaches successfully used in the past cannot be carried over to deep learning methods with the same effectiveness.

7. Visual navigation. The application of visual navigation in robotics plays a crucial role in performing various tasks, including manipulation, mobile robotics, and automated driving. Visual navigation involves moving toward specific objects or regions in an environment, with the main challenge of generalizing a scene not seen during the training because the scene's structure and the objects' appearance are unfamiliar.

5 Future directions

Existing DA techniques in the context of RL have shown outstanding performance in various fields. Nonetheless, the performance gap between these techniques and the upper bound is still challenging. In this section we aim to provide some research directions to help resolve the remaining problems. In this regard, more novel perspectives, applications, and practical settings of DA are presented.

Current DA techniques attempt to learn a generic representation that can be shared between different domains. However, discovering the knowledge behind visual tasks has been neglected by most researchers. It has been proven that humans have superior range generalization capacity due to their ability to acquire knowledge of underlying tasks and predict outcomes in various domains. The learning of common sense for DA, which can be used for mimicking the human ability to generalize domains, is an interesting direction for future studies. Models can be adapted more effectively through common sense. Most DA studies have concentrated mainly on improving target-domain performance without considering model robustness. Exploring the most effective way to conduct DA while improving model robustness in the domain of interest is another open issue for research.

Nowadays, edge devices, such as security cameras, autonomous cars, and mobile phones, play an important role in various vision-based perception models. They are typically deployed in a variety of environments requiring extensive DA. Distinct networks need to be customized based on user-provided personal data. It would be too costly to send all of the user data to the server and train numerous networks. On the other hand, training networks on edge devices can reduce computational complexity while maintaining privacy since the gathered data remain local. However, even though edge devices have restricted computation and power capacities, most data adaptation approaches, including adversarial generative models, require high-end graphics processing units (GPUs). In this regard, some techniques like software hardware co-design, neural architecture search, pruning, and quantization can be used for efficient on-device training. Using effective deep learning methods to perform data adaptation on edge devices is a useful and interesting research topic.

Data produced by IoT devices bring new issues for training machine learning models. Because of the rising storage and computing capacity of these devices, as well as data privacy issues, it becomes more desirable to store data on independent devices instead of storing them in centralized storage. Federated learning offers a privacy-preserving strategy for leveraging decentralized resources to train machine learning models. The basic principle underlying federated learning is that each node learns from locally available data without sharing either the model parameters or the data. This approach can enhance the efficiency and privacy of machine learning conducted across distributed networks. On the other hand, recent data adaptation techniques neglect the assumption that data on each node are gathered in an independent and identically distributed way, resulting in domain shift among the nodes. Hence, models trained by federated learning may still fail to adapt to new devices. Therefore, developing DA methods for federated learning is critical.

In most data adaptation methods, researchers have used a predefined dataset for training the algorithm, assuming that the data during the test phase have the same pattern as that in the training phase, but this is not always the case. Continuous learning and adaptation are required to ensure that the network performs efficiently all of the time. The network is, generally, supposed to be capable of continuously learning from empirical data, expanding on prior knowledge and adapting to new domains.

The research findings indicate that most of the techniques have concentrated on a single task with only a single-modal input. Nevertheless, in several contexts, multiple tasks must be performed on the same data simultaneously. In terms of computation, adapting each task individually might be redundant since the networks for multiple models may operate on the same collection of characteristics. Therefore, focusing on adapting multiple tasks simultaneously and effectively is recommended for upcoming studies.

6 Conclusions

The increase in the number of DA techniques for solving continuous RL issues demonstrates their perceived significance of this field. In this paper, we have presented an in-depth assessment of recent developments in the application of DA in the context of RL. We classified studies into three classes (supervised, semi-supervised, and unsupervised) and discussed them, considering vital metrics like performance, data dependency, optimizability, data scalability, task scalability, and efficiency. In the process of reviewing the methods, we determined the applications of each method, as well as the relevant experimental platforms used, datasets, functions, and algorithms. Finally, we suggested some potential directions for upcoming studies. We hope that this survey will contribute to a comprehensive knowledge of DA key principles and offer useful insights for future research.

Contributors

Amirfarhad FARHADI designed, drafted, and organized the paper. Mitra MIRZAREZAEE, Arash SHARIFI, and Mohammad TESHNEHLAB revised and finalized the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

References

Abdul Samad SR, Balasubaramanian S, Al-Kaabi AS, et al., 2023. Analysis of the performance impact of fine-tuned machine learning model for phishing URL detection. *Electronics*, 12(7):1642. https://doi.org/10.3390/electronics12071642

Bagheri M, 2021. Clustering Individual Entities Based on Common Features. PhD Dissemination, University of Houston, Houston, USA.

Bolhassani M, Oksuz I, 2021. Semi-supervised segmentation of multi-vendor and multi-center cardiac MRI. 29th Signal Processing and Communications Applications Conf, p.1-4. https://doi.org/10.1109/SIU53274.2021.9477818

Boute RN, Gijsbrechts J, van Jaarsveld W, et al., 2022. Deep reinforcement learning for inventory control: a roadmap. *Eur J Oper Res*, 298(2):401-412. https://doi.org/10.1016/j.ejor.2021.07.016

Bu FY, Wang X, 2019. A smart agriculture IoT system based on deep reinforcement learning. *Fut Gener Comput Syst*, 99:500-507. https://doi.org/10.1016/j.future.2019.04.041

Carr T, Chli M, Vogiatzis G, 2019. Domain adaptation for reinforcement learning on the Atari. 18th Int Conf on Autonomous Agents and Multiagent Systems, p.1859-1861.

Chen DQ, Fisch A, Weston J, et al., 2017. Reading Wikipedia to answer open-domain questions. 55th Annual Meeting of the Association for Computational Linguistics, p.1870-1879. https://doi.org/10.18653/v1/P17-1171

Chen J, Wu XX, Duan LX, et al., 2022. Domain adversarial reinforcement learning for partial domain adaptation. *IEEE Trans Neur Netw Learn Syst*, 33(2):539-553. https://doi.org/10.1109/TNNLS.2020.3028078

Chen XH, Jiang S, Xu F, et al., 2021. Cross-modal domain adaptation for cost-efficient visual reinforcement learning. 35th Conf on Neural Information Processing Systems, p.12520-12532.

Chu CH, Wang R, 2020. A survey of domain adaptation for machine translation. *J Inform Process*, 28:413-426. https://doi.org/10.2197/ipsjjip.28.413

Dhingra B, Mazaitis K, Cohen WW, 2017. Quasar: datasets for question answering by search and reading. https://doi.org/10.48550/arXiv.1707.03904

Di SM, Peng JS, Shen YY, et al., 2018. Transfer learning via feature isomorphism discovery. Proc 24th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining, p.1301-1309. https://doi.org/10.1145/3219819.3220029

Dong JH, Cong Y, Sun G, et al., 2020. CSCL: critical semantic-consistent learning for unsupervised domain adaptation. 16th European Conf on Computer Vision, p.745-762.
https://doi.org/10.1007/978-3-030-58598-3_44

Dunn M, Sagun L, Higgins M, et al., 2017. SearchQA: a new Q&A dataset augmented with context from a search engine. https://doi.org/10.48550/arXiv.1704.05179

El Jery A, Aldrdery M, Ghoudi N, et al., 2023. Experimental investigation and proposal of artificial neural network models of lead and cadmium heavy metal ion removal from water using porous nanomaterials. *Sustainability*, 15(19):14183. https://doi.org/10.3390/su151914183

Fang F, Dutta K, Datta A, 2014. Domain adaptation for sentiment classification in light of multiple sources. *Inform J Comput*, 26(3):586-598.
https://doi.org/10.1287/ijoc.2013.0585

Farhadi A, Sharifi A, 2024. Leveraging meta-learning to improve unsupervised domain adaptation. *Comput J*, 67(5):1838-1850.
https://doi.org/10.1093/comjnl/bxad104

Farhadi A, Mirzarezaee M, Sharifi A, et al., 2023. Unsupervised domain adaptation for image classification based on deep neural networks. *Intell Multim Process Commun Syst*, 4(1):27-37 (in Persian).

Finn C, Abbeel P, Levine S, 2017. Model-agnostic meta-learning for fast adaptation of deep networks. Proc 34th Int Conf on Machine Learning, p.1126-1135.

Gardner P, Liu X, Worden K, 2020. On the application of domain adaptation in structural health monitoring. *Mech Syst Signal Process*, 138:106550.
https://doi.org/10.1016/j.ymssp.2019.106550

Gašić M, Young S, 2014. Gaussian processes for POMDP-based dialogue manager optimization. *IEEE/ACM Trans Audio Speech Language Process*, 22(1):28-40.
https://doi.org/10.1109/TASL.2013.2282190

Ge L, Gao J, Zhang AD, 2013. OMS-TL: a framework of online multiple source transfer learning. Proc 22nd ACM Int Conf on Information & Knowledge Management, p.2423-2428.
https://doi.org/10.1145/2505515.2505603

Guan H, Liu MX, 2022. Domain adaptation for medical image analysis: a survey. *IEEE Trans Biomed Eng*, 69(3):1173-1185.
https://doi.org/10.1109/TBME.2021.3117407

Guo RY, Liu H, Liu D, 2024. When deep learning-based soft sensors encounter reliability challenges: a practical knowledge-guided adversarial attack and its defense. *IEEE Trans Industr Inform*, 20(2):2702-2714.
https://doi.org/10.1109/TII.2023.3297663

Higgins I, Pal A, Rusu A, et al., 2017. DARLA: improving zero-shot transfer in reinforcement learning. 34th Int Conf on Machine Learning, p.1480-1490.

Jannat MKA, Islam MS, Yang SH, et al., 2023. Efficient Wi-Fi-based human activity recognition using adaptive antenna elimination. *IEEE Access*, 11:105440-105454.
https://doi.org/10.1109/ACCESS.2023.3320069

Jeong R, Aytar Y, Khosid D, et al., 2020. Self-supervised sim-to-real adaptation for visual robotic manipulation. IEEE Int Conf on Robotics and Automation, p.2718-2724.
https://doi.org/10.1109/ICRA40945.2020.9197326

Jiang J, Zhai CX, 2007. Instance weighting for domain adaptation in NLP. 45th Annual Meeting of the Association of Computational Linguistics, p.264-271.

Joshi M, Choi E, Weld D, et al., 2017. TriviaQA: a large scale distantly supervised challenge dataset for reading comprehension. 55th Annual Meeting of the Association for Computational Linguistics, p.1601-1611.
https://doi.org/10.18653/v1/P17-1147

Khader N, Yoon SW, 2021. Adaptive optimal control of stencil printing process using reinforcement learning. *Robot Comput Integr Manuf*, 71:102132.
https://doi.org/10.1016/j.rcim.2021.102132

Khodayari M, Razmi J, Babazadeh R, 2019. An integrated fuzzy analytical network process for prioritisation of new technology-based firms in Iran. *Int J Ind Syst Eng*, 32(4):424-442.
https://doi.org/10.1504/IJISE.2019.101331

Li SD, Chaplot DS, Tsai YHH, et al., 2020. Unsupervised domain adaptation for visual navigation.
https://doi.org/10.48550/arXiv.2010.14543

Li X, Zhong JP, Kamruzzaman MM, 2021. Complicated robot activity recognition by quality-aware deep reinforcement learning. *Fut Gener Comput Syst*, 117:480-485. https://doi.org/10.1016/j.future.2020.11.017

Li XT, Sun Y, 2021. Application of RBF neural network optimal segmentation algorithm in credit rating. *Neur Comput Appl*, 33(14):8227-8235.
https://doi.org/10.1007/s00521-020-04958-9

Liu BY, Guo YH, Ye JP, et al., 2020. Selective pseudo-labeling with reinforcement learning for semi-supervised domain adaptation. 32nd British Machine Vision Conf, p.299.

Liu MF, Song Y, Zou HB, et al., 2019. Reinforced training data selection for domain adaptation. Proc 57th Annual Meeting of the Association for Computational Linguistics, p.1957-1968.
https://doi.org/10.18653/v1/P19-1189

Liu Q, Yuan H, Hamzaoui R, et al., 2021. Reduced reference perceptual quality model with application to rate control for video-based point cloud compression. *IEEE Trans Image Process*, 30:6623-6636.
https://doi.org/10.1109/TIP.2021.3096060

Liu SP, Tian GH, Cui YC, et al., 2022. A deep Q-learning network based active object detection model with a novel training algorithm for service robots. *Front Inform Technol Electron Eng*, 23(11):1673-1683.
https://doi.org/10.1631/FITEE.2200109

Liu X, Zhou GH, Kong MH, et al., 2023a. Developing multi-labelled corpus of Twitter short texts: a semi-automatic method. *Systems*, 11(8):390.
https://doi.org/10.3390/systems11080390

Liu X, Wang S, Lu SY, et al., 2023b. Adapting feature selection algorithms for the classification of Chinese texts. *Systems*, 11(9):483. https://doi.org/10.3390/systems11090483

López M, Valdivia A, Martínez-Cámara E, et al., 2019. E^{2}SAM: evolutionary ensemble of sentiment analysis methods for domain adaptation *Inform Sci*, 480:273-286. https://doi.org/10.1016/j.ins.2018.12.038

Madadi Y, Seydi V, Nasrollahi K, et al., 2020. Deep visual unsupervised domain adaptation for classification tasks: a survey. *IET Image Process*, 14(14):3283-3299. https://doi.org/10.1049/iet-ipr.2020.0087

Monjezi V, Trivedi A, Tan G, et al., 2023. Information-theoretic testing and debugging of fairness defects in deep neural networks. IEEE/ACM 45th Int Conf on Software Engineering, p.1571-1582. https://doi.org/10.1109/ICSE48619.2023.00136

Mou JH, Gao KZ, Duan PY, et al., 2023. A machine learning approach for energy-efficient intelligent transportation scheduling problem in a real-world dynamic circumstances. *IEEE Trans Intell Trans Syst*, 24(12):15527-15539. https://doi.org/10.1109/TITS.2022.3183215

Patel VM, Gopalan R, Li RN, et al., 2015. Visual domain adaptation: a survey of recent advances. *IEEE Signal Process Mag*, 32(3):53-69. https://doi.org/10.1109/MSP.2014.2347059

Patel Y, Chitta K, Jasani B, 2018. Learning sampling policies for domain adaptation. https://doi.org/10.48550/arXiv.1805.07641

Pourghebleh B, Navimipour NJ, 2017. Data aggregation mechanisms in the Internet of Things: a systematic review of the literature and recommendations for future research. *J Netw Comput Appl*, 97:23-34. https://doi.org/10.1016/j.jnca.2017.08.006

Pourghebleh B, Hayyolalam V, Aghaei Anvigh A, 2020. Service discovery in the Internet of Things: review of current trends and research challenges. *Wirel Netw*, 26(7):5371-5391. https://doi.org/10.1007/s11276-020-02405-0

Rajput SPS, Webber JL, Bostani A, et al., 2023. Using machine learning architecture to optimize and model the treatment process for saline water level analysis. *Water Reuse*, 13(1):51-67. https://doi.org/10.2166/wrd.2022.069

Saeed R, Feng HH, Wang X, et al., 2022. Fish quality evaluation by sensor and machine learning: a mechanistic review. *Food Contr*, 137:108902. https://doi.org/10.1016/j.foodcont.2022.108902

Saunders D, 2022. Domain adaptation and multi-domain adaptation for neural machine translation: a survey. *J Artif Intell Res*, 75:351-424. https://doi.org/10.1613/jair.1.13566

Shoeleh F, Asadpour M, 2017. Graph based skill acquisition and transfer learning for continuous reinforcement learning domains. *Patt Recognit Lett*, 87:104-116. https://doi.org/10.1016/j.patrec.2016.08.009

Shoeleh F, Asadpour M, 2020. Skill based transfer learning with domain adaptation for continuous reinforcement learning domains. *Appl Intell*, 50(2):502-518. https://doi.org/10.1007/s10489-019-01527-z

Singhal P, Walambe R, Ramanna S, et al., 2023. Domain adaptation: challenges, methods, datasets, and applications. *IEEE Access*, 11:6973-7020. https://doi.org/10.1109/ACCESS.2023.3237025

Su PH, Budzianowski P, Ultes S, et al., 2017. Sample-efficient actor-critic reinforcement learning with supervised data for dialogue management. 18th Annual SIGDIAL Meeting on Discourse and Dialogue, p.147-157. https://doi.org/10.18653/v1/W17-5518

Sun SL, Shi HL, Wu YB, 2015. A survey of multi-source domain adaptation. *Inform Fusion*, 24:84-92. https://doi.org/10.1016/j.inffus.2014.12.003

Sutton RS, Barto AG, 2018. Reinforcement Learning: an Introduction (2nd Ed.). Cambridge, UK.

Truong J, Chernova S, Batra D, 2021. Bi-directional domain adaptation for sim2real transfer of embodied navigation agents. *IEEE Robot Autom Lett*, 6(2):2634-2641. https://doi.org/10.1109/LRA.2021.3062303

Voulgarelis S, Fathi F, Stucke AG, et al., 2021. Evaluation of visible diffuse reflectance spectroscopy in liver tissue: validation of tissue saturations using extracorporeal circulation. *J Biomed Opt*, 26(5):055002. https://doi.org/10.1117/1.jbo.26.5.055002

Wang HN, Liu N, Zhang YY, et al., 2020. Deep reinforcement learning: a survey. *Front Inform Technol Electron Eng*, 21(12):1726-1744. https://doi.org/10.1631/FITEE.1900533

Wang M, Deng WH, 2018. Deep visual domain adaptation: a survey. *Neurocomputing*, 312:135-153. https://doi.org/10.1016/j.neucom.2018.05.083

Wei GQ, Wei ZQ, Huang L, et al., 2021. Center-aligned domain adaptation network for image classification. *Expert Syst Appl*, 168:114381. https://doi.org/10.1016/j.eswa.2020.114381

Yang M, Tu WT, Qu Q, et al., 2018. Personalized response generation by dual-learning based domain adaptation. *Neur Netw*, 103:72-82. https://doi.org/10.1016/j.neunet.2018.03.009

Yang ZL, Hu JJ, Salakhutdinov R, et al., 2017. Semi-supervised QA with generative domain-adaptive nets. 55th Annual Meeting of the Association for Computational Linguistic, p.1040-1050. https://doi.org/10.18653/v1/P17-1096

Yoon J, Arik S, Pfister T, 2020. Data valuation using reinforcement learning. 37th Int Conf on Machine Learning, p.10842-10851.

Zhang H, Luo GY, Li JL, et al., 2022. C2FDA: coarse-to-fine domain adaptation for traffic object detection. *IEEE Trans Intell Transp Syst*, 23(8):12633-12647. https://doi.org/10.1109/TITS.2021.3115823

Zhang JW, Tai L, Yun P, et al., 2019. VR-goggles for robots: real-to-sim domain adaptation for visual control. *IEEE Robot Autom Lett*, 4(2):1148-1155. https://doi.org/10.1109/LRA.2019.2894216

Zhang NJ, Fan KX, Ji HW, et al., 2023. Identification of risk factors for infection after mitral valve surgery through machine learning approaches. *Front Cardiovasc Med*, 10:1050698. https://doi.org/10.3389/fcvm.2023.1050698

Zhao N, Li DQ, Gu SX, et al., 2024. Analytical fragility relation for buried cast iron pipelines with lead-caulked joints based on machine learning algorithms. *Earthq Spectra*, 40(1):566-583. https://doi.org/10.1177/87552930231209195

Zhao SC, Li B, Reed C, et al., 2020. Multi-source domain adaptation in the deep learning era: a systematic survey. https://doi.org/10.48550/arXiv.2002.12169

Zhao SC, Yue XY, Zhang SH, et al., 2022. A review of single-source deep unsupervised visual domain adaptation. *IEEE Trans Neur Netw Learn Syst*, 33(2):473-493. https://doi.org/10.1109/TNNLS.2020.3028503

List of supplementary materials