Inverse Reinforcement Learning for Healthcare Applications: A Survey

Mohamed-Amine Chadi and Hajar Mousannif
LISI Laboratory, Department of Computer Sciences, Cadi Ayyad University, Marrakech, Morocco

Keywords: Inverse Reinforcement Learning, Healthcare Applications, Survey.

Abstract: Reinforcement learning (RL) is a category of algorithms in machine learning that deals mainly with learning optimal sequential decision-making. And because the medical treatment process can be represented as a series of interactions between doctors and patients, RL offers promising techniques for solving complex problems in healthcare domains. However, to ensure a good performance of such applications, a reward function should be explicitly provided beforehand, which can be either too expensive to obtain, unavailable, or non-representative enough of the real-world situation. Inverse reinforcement learning (IRL) is the problem of deriving the reward function of an agent, given its history of behaviour or policy. In this survey, we will discuss the theoretical foundations of IRL techniques and the problem it solves. Then, we will provide the state-of-the-art of current applications of IRL in healthcare specifically. Following that, we will summarize the challenges and what makes IRL in healthcare domains so limited despite its progress in other research areas. Finally, we shall suggest some prospective study directions for the future.

1 INTRODUCTION

In recent years, reinforcement learning (RL) (Richard S. & Andrew G., 2017) has been very successful at solving complex sequential decision-making problems in different areas like video games (K. Shao et al., 2019), financial market (Halperin, 2017), robotics (Kober et al., 2013), healthcare domains (Yu, Liu, & Nemati, 2019), including pathologies like cancer, diabetes, anaemia, schizophrenia, epilepsy, anaesthesia, and drug discovery, to mention a few. However, for RL applications to work correctly, an explicit reward function should be provided to specify the objective of the treatment process that clinicians have in mind. Manually specifying a reward function may require prior domain knowledge. It also depends on the clinician's personal experience, which undoubtedly differs across all health professionals, thus leading to a non-representative enough reward function of a real-world scenario or resulting in inconsistent learning performance.

The problem of inferring a reward function from an observed policy is known as *inverse reinforcement learning* (IRL) (Russell, 1998), (Ng & Russel, 2000) and (Pieter & Ng, 2004).

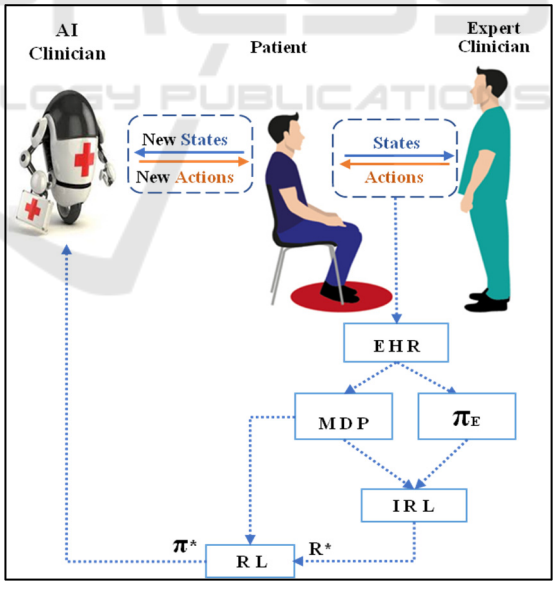

Figure 1: Global view of the RL - IRL framework in healthcare

During this past decade or so, IRL has been gaining a lot of attention among researchers in the artificial intelligence communities, psychology, control theory, and other domains. For a much broader discussion of IRL's progress, its theoretical

Chadi, M. and Mousannif, H.
Inverse Reinforcement Learning for Healthcare Applications: A Survey.
DOI: 10.5220/0010729200003101
In *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning (BML 2021)*, pages 97-102
ISBN: 978-989-758-559-3

foundations, different algorithms, application in various domains and its inherent challenges, one might be referred to the work done by (Z. Shao & Er, 2012) and (Arora & Doshi, 2021).

Despite the excellent theoretical progress, one can easily remark that there is a small amount of work regarding IRL applications to healthcare domains, which might be referred to many reasons, as we will explore in the corresponding sections of this article. In this respect, this article addresses the following points:

- An introduction to the theoretical foundations of IRL and the problem that it solves
- A state-of-the-art application of IRL in healthcare domains
- The challenges that face the applicability of IRL in healthcare
- Some potential directions for future research

Obviously, a conference article, will not cover everything. However, we tried to make this survey as comprehensive as possible.

2 THEORETICAL FOUNDATIONS OF IRL

To get a proper understanding of IRL, we will consider the most standard model, "the Markov Decision Process (MDP)". There are other models, such as the partially observable Markov Decision Process, the hidden-parameter Markov Decision Process, which we will not go through in this article.

Unlike RL, which gets as input **states of the environment** and a well-defined **reward function** so that it produces the optimal **behaviour** that maximizes the reward function, IRL is the exact opposite. IRL takes in states of the environment and information about the behaviour exercised by the **expert** to output a reward function.

2.1 Definitions

A **Markov Decision Process** of {S, A, T, R, γ} represents an agent's history of interactions, where:

- A is the action space, = *{a1, a2 . . . ak}*.
- S is the state space.
- **T** is the transition function, i.e., given state S, and executing action A. The agent will follow T to get to another state S'.
- **γ** is the discount factor. It is mainly used for two goals, first is to prove the convergence of the algorithm, second is to model the uncertainty of the agent about the successive decision instants.
- **R** is the reward received when action A is performed at state S.
- $\boldsymbol{\pi_E}$ is an assumedly optimal policy. In the case of IRL, it is the behaviour performed (or trajectories followed) by the expert.

2.2 Formulation of the IRL problem

- Given rollouts of the expert's policy $\boldsymbol{\pi_E}$, i.e., history of states and actions from the expert's interaction process with the MDP.
- Identify a set of possible reward functions (*R)* so that the policy π recovered from the R is (or as near as possible as) the policy performed by the expert $\boldsymbol{\pi_E}$ for the given MDP.

Many algorithms were developed to solve this problem since the first steps taken in 2000 by (Ng & Russel, 2000) and (Pieter & Ng, 2004) in 2004, such as Maximum Entropy IRL (Ziebart et al., 2008), nonlinear representations of the reward function using Gaussian processes (Levine et al., 2011), Bayesian IRL (Ramachandran & Amir, 2007), and many others as discussed in detail in (Z. Shao & Er, 2012) and (Arora & Doshi, 2021).

3 IRL FOR HEALTHCARE: STATE-OF-THE-ART APPLICATIONS

In most cases, IRL is just a step towards making a more robust approach to solve a healthcare-related decision-making problem using RL, where providing the reward function to the RL algorithm is not trivial, exploiting IRL techniques might be of massive benefit in solving many issues that RL faces when used in clinical settings, for a more detailed discussion on the use of RL in healthcare, the survey conducted in (Yu, Liu, & Nemati, 2019) is a good resource.

Despite being widely used, nearly every RL application assumes that a reward function is provided or easy to program manually, which is often not the case, calling for new robust ways such as IRL.

After working on modelling medical records of diabetes as an MDP in this first article (Asoh, Shiro, Akaho, Kamishima, et al., 2013), and utilizing the doctors' opinions for defining the reward for each treatment, *Asoh et al*. then developed a new method based on Bayesian-IRL to infer the reward function that doctors were considering for the treatment

process (Asoh, Shiro, Akaho, & Kamishima, 2013), in the conclusion of their work, *Asoh et al*. reported that the results they have achieved thus far are somewhat preliminary and still some difficulties that must be addressed, such as introducing the heterogeneity of doctors and patients using hierarchical modelling, applying the multitasks inverse reinforcement learning algorithm by (Dimitrakakis & Rothkopf, 2012), extending the MDP to POMDP which implicates a proper design of the state and the action spaces, and finally taking into account longer examinations histories and treatments.

Next, is the work done in (Li & Burdick, 2017) and (Li et al., 2018) regarding clinical motion analysis of both physicians and patients. They worked on the problem of inverse reinforcement learning in large state space and solved it using function approximators approaches (e.g., a neural network) that do not necessarily need to go through the RL problems when learning the reward function. They reported that the proposed approach showed more accuracy and scalability when compared to traditional methods. A clinical application was also presented as a test for the proposed method, where it was applied to evaluate robot operators according to three surgical activities: tying knots, passing needles and suturing. And in their later work (Li & Burdick, 2020), the suggested technique was tested in two simulation settings. They used ground-truth data for comparing alternative configurations and extensions and then applied it to a clinician skill assessment and the analysis of a patient motion therapy.

Following the above, the authors in (Yu, Liu, & Zhao, 2019) applied the Bayesian IRL method and modelled the sequential decision-making problem of a ventilator weaning as an MDP, and used a batch RL approach, fitted-Q-iterations with a gradient boosting decision tree, to infer an appropriate strategy from actual trajectories in historical data in ICU. In their results, they concluded that the IRL approach could extract significant indications for prescribing extubation readiness and sedative dosage. This makes it clear that patient's physiological stability received greater importance by clinicians, rather than oxygenation criteria. In addition, new successful treatment procedures can be proposed when determining optimum weights. They also emphasized that although the results have confirmed the viability of inverse reinforcement learning techniques in complex medical environments, there are still many concerns that must be properly addressed before these approaches might be meaningfully applied.

In even more recent work, (Yu, Ren, & Liu, 2019) *Yu et al.* proposed a new model that incorporates the advantages of mini trees and Deep IRL. As reported in their conclusion, this approach can adequately address the problems of identifying factors that have to be taken into consideration when evaluating the decision-making performance, and the different roles these factors can play in treating sepsis.

4 IRL FOR HEALTHCARE: CHALLENGES

The previous section has summarized the state-of-the-art applications of IRL in healthcare domains over the past decade. While notable success has been obtained, IRL applications in healthcare still manifest some limitations. Most of these limitations are inherent to the RL framework in general and to the complexity of clinical data, which was exhaustively discussed in (Yu, Liu, & Nemati, 2019) and (Gottesman et al., 2018) and (Gottesman et al., 2019). In the present work, however, we will concentrate on the challenges encountered when using IRL in healthcare and what makes the literature as limited as we explored.

4.1 Data in Healthcare Settings

As shown in Figure 1, the primary sources of data for an IRL algorithm in healthcare are states of the patients and actions performed by clinicians accordingly. Thus, the main challenge here is how to collect useful data given that:

- Patients may fail to complete the treatment regime.
- Devices in clinical settings can be changed, and each device is subject to many inherent biases.
- For some complicated diseases, clinicians are still confronted with inconsistencies in selecting precise data as the state in each case (Vellido et al., 2018).
- States and actions -by credit assignment- should contain sufficient information for a clear distinction of different patients.
- States and actions must be causal, meaning they must have either direct or indirect effects on the task to learn or reward to achieve.

4.2 IRL for an Eventual RL Application

Let us consider that the IRL algorithm did well and gave us the reward function perfectly. The RL algorithm now have two choices, either: (**one)** learn using trial and error or (**two)** learn from another

policy, i.e., that of the expert clinician, also known as the *off-policy evaluation* problem. Learning via method one implicates executing the policy directly on the patients, this is impossible because of the large trial costs, uncontrolled treatment hazards, or just illegal and unethical humanistic considerations. Method two on the other hand (the *off-policy evaluation*), estimates the performance of the learned policies on retrospective data before testing them in clinical environments. Using sepsis management as an illustration, (Gottesman et al., 2019) discussed the reasons that make the assessment of policies on historical data a difficult problem, as any improper handling of the state representation, variance of importance-sampling-based statistical estimators, and confounders in more ad-hoc measurements, would result in inaccurate or even deceptive values of the treatment quality.

4.3 the Black Box Problem

This problem, which is the lack of clear interpretability (Lipton, 2018) is inherent to the RL eventual application after an IRL usage. As illustrated in figure 1, most IRL applications are just a bridge (extracting the reward function) for an eventual RL application, and RL algorithms take-in some input data and directly output a policy, that is hard to interpret. Despite the tremendous success achieved in solving challenging problems like learning games such as Atari and Go, autonomous driving, etc. Adopting RL policies in medical applications might get strong resistance given the fact that clinicians are not expected to try any new treatment without laborious validation accuracy, safety, and robustness.

5 POTENTIAL DIRECTIONS FOR FUTUR RESEARCH

Since the very first introduction of IRL, a good number of remarkable improvements have enabled it to be integrated in more practical applications. However, IRL applications in healthcare are still very limited given the challenges manifested in current IRL applications in healthcare domains discussed above. Addressing these challenges would certainly be of enormous benefit in improving clinical decision making.

In this section, we will discuss some future perspectives that we believe are among the most important ones, mainly focusing on the following four axes:

5.1 The Missing Data

As concluded in (Yu, Liu, & Zhao, 2019), the learning accuracy will undoubtedly be affected by the errors brought in by the data collection and preprocessing phase. They proposed that IRL methods must be able to directly work on the raw, noisy, and incomplete data.

On the other hand, Asoh et al. believe that next time they should consider longer histories of examinations and treatments to introduce complex decision-making processes of doctors (Asoh, Shiro, Akaho, & Kamishima, 2013).

While enough available training samples are currently scarce in many healthcare domains (e.g., few retrospective data for new or rare diseases). Excellent solutions were proposed in other fields of research that might mitigate the effect of missing data immensely if exploited in healthcare applications, for example, using data augmentation techniques (Salamon & Bello, 2017) or GANs (Goodfellow et al., 2020) to increase the number of samples. Another solution is the application of knowledge distillation (Hinton et al., 2015), or meta-learning (Lake et al., 2017), or even knowledge transfer (Killian et al., 2019) from a well-known patient to another relatively similar patient to overcome the problem of insufficient data.

Although it is still early, significant progress has been made in the "small sample learning" research in recent years (Carden & Livsey, 2017). Building on these achievements and address the issue of IRL with little data in healthcare domains, is a calling area for future research.

5.2 Modelling Complex Clinical Settings

Most of the current IRL applications in healthcare assume that the agent operates in tiny domains with a discrete state space, which in contrast to the healthcare domains, that often involve continuous multi-dimensional states and actions. Learning and planning across such large-scale continuous models is a great challenge.

The authors in (Yu, Liu, & Zhao, 2019) reported that their future investigations would focus on applying IRL methods that can estimate the reward and the model-dynamics simultaneously (Herman et al., 2016). Because usually, IRL methods rely on the availability of an a precise MDP model, which is either **given** or **estimated from data.** Asoh et al. also faced a similar issue (Asoh, Shiro, Akaho, & Kamishima, 2013) regarding the over simplicity of

MDPs in a healthcare setting. Thus, they decided that the following works will consider some extensions of the MDP to POMDP and introducing the heterogeneity of doctors and patients by hierarchical modelling.

5.3 IRL or Apprenticeship Learning

Apprenticeship learning via IRL (Pieter & Ng, 2004) is learning a reward function using IRL from observed behaviour and using the learned reward function in reinforcement learning. While most studies use IRL as a bridge for an eventual RL application, just like in apprenticeship learning, in the case of healthcare where safety is of paramount importance, it might be helpful to consider using IRL alone for the sake of extracting the reward function. Because all it does is inferring the reward function of presumably optimal treatment policy and extracting information about the essential variables to consider and recommend for clinicians. Thus, unless we can provide a sophisticated and realistic simulation of the healthcare setting for RL algorithms to be trained and create interpretable RL solutions to improve the safety, robustness, and the accuracy of learnt policies in healthcare-domains which is currently an unresolved topic that necessitates more research, IRL can safely help improve clinician's decision-making.

6 CONCLUSIONS

Inverse reinforcement learning (IRL) presents a theoretical, and in lots of cases, a practical solution to infer the reward function or the objective behind a given policy. Usually, in healthcare domains, the policy is performed by a clinician (i.e., the expert).

This paper aims to provide a brief comprehensive survey of state-of-the-art applications of IRL techniques in healthcare, the challenges faced, and some potential directions for future research.

Although tremendous progress has been made in recent years in the field of IRL in a lot of other areas, clinical settings are uniquely critical and high risk-sensitive, thus the limited literature regarding these IRL applications in healthcare.

Nevertheless, IRL can be safely and efficiently exploited to extract meaningful indicators associated with the learned reward function. This can help to recommend new effective treatment protocols and therefore improving clinician's decision making.

However, the risks of IRL in healthcare applications is manifested highly when it is used as a bridge for an eventual RL application where the patient's health is becoming between the hands of an algorithm usually functioning as a non-interpretable black-box making clinicians unlikely to trust it. Also, most RL algorithms in healthcare learn either by trial and error, which is obviously unfeasible/unethical, or through another policy given as retrospective treatment data, which -as we discussed- is still not reliable enough and needs more improvement.

Finally, the hope is that more researchers from different disciplines exploit their domain-expertise and collaborate to produce more reliable solutions to improve the decision-making in the healthcare domains.

REFERENCES

Arora, S., & Doshi, P. (2021). A survey of inverse reinforcement learning: Challenges, methods and progress. *Artificial Intelligence*, *297*, 103500. https://doi.org/10.1016/j.artint.2021.103500

Asoh, H., Shiro, M., Akaho, S., & Kamishima, T. (2013). An Application of Inverse Reinforcement Learning to Medical Records of Diabetes. *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD '13)*, 1–8.

Asoh, H., Shiro, M., Akaho, S., Kamishima, T., Hasida, K., Aramaki, E., & Kohro, T. (2013). Modelling Medical Records of Diabetes. *Proceedings of the ICML2013 Workshop on Role of Ma- Chine Learning in Transforming Healthcare*, 6.

Carden, S. W., & Livsey, J. (2017). Small-sample reinforcement learning: Improving policies using synthetic data. *Intelligent Decision Technologies*, *11*(2), 167–175. https://doi.org/10.3233/IDT-170285

Dimitrakakis, C., & Rothkopf, C. A. (2012). Bayesian multitasks inverse reinforcement learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7188 LNAI*, 273–284. https://doi.org/10.1007/978-3-642-29946-9_27

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622

Gottesman, O., Johansson, F., Komorowski, M., Faisal, A., Sontag, D., Doshi-Velez, F., & Celi, L. A. (2019). Guidelines for reinforcement learning in healthcare. *Nature Medicine*, *25*(1), 16–18. https://doi.org/10.1038/s41591-018-0310-5

Gottesman, O., Johansson, F., Meier, J., Dent, J., Lee, D., Srinivasan, S., Zhang, L., Ding, Y., Wihl, D., Peng, X., Yao, J., Lage, I., Mosch, C., Lehman, L. H., Komorowski, M., Komorowski, M., Faisal, A., Celi, L. A., Sontag, D., & Doshi-Velez, F. (2018). *Evaluating*

Reinforcement Learning Algorithms in Observational Health Settings. 1–16. http://arxiv.org/abs/1805.12298

Halperin, I. (2017). Inverse Reinforcement Learning for Marketing. *SSRN Electronic Journal*, 1–18. https://doi.org/10.2139/ssrn.3087057

Herman, M., Gindele, T., Wagner, J., Schmitt, F., & Burgard, W. (2016). Inverse reinforcement learning with simultaneous estimation of rewards and dynamics. *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016*, *51*, 102–110.

Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the Knowledge in a Neural Network*. 1–9. http://arxiv.org/abs/1503.02531

Killian, T., Daulton, S., Konidaris, G., & Doshi-Velez, F. (2019). Robust and Efficient Transfer Learning with Hidden Parameter Markov Decision Processes. *Physiology & Behavior*, *176*(3), 139–148.

Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. *International Journal of Robotics Research*, *32*(11), 1238–1274. https://doi.org/10.1177/0278364913495721

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioural and Brain Sciences*, *40*(2012), 1–58. https://doi.org/10.1017/S0140525X16001837

Levine, S., Popović, Z., & Koltun, V. (2011). Nonlinear inverse reinforcement learning with Gaussian processes. *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*, *May*.

Li, K., & Burdick, J. W. (2017). *A Function Approximation Method for Model-based High Dimensional Inverse Reinforcement Learning*. http://arxiv.org/abs/1708.07738

Li, K., & Burdick, J. W. (2020). Human motion analysis in medical robotics via high dimensional inverse reinforcement learning. *International Journal of Robotics Research*, *39*(5), 568–585. https://doi.org/10.1177/0278364920903104

Li, K., Rath, M., & Burdick, J. W. (2018). Inverse Reinforcement Learning via Function Approximation for Clinical Motion Analysis. *Proceedings - IEEE International Conference on Robotics and Automation*, *December*, 610–617. https://doi.org/10.1109/ICRA.2018.8460563

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, *61*(10), 35–43. https://doi.org/10.1145/3233231

Ng, A. Y., & Russel, S. (2000). *Algorithms for Inverse Reinforcement Learning*. *0*.

Pieter, A., & Ng, A. Y. (2004). Apprenticeship Learning via Inverse Reinforcement Learning. *21 St International Confer- Ence on Machine Learning, Banff, Canada, 2004*, *346*, 1–2. https://ai.stanford.edu/~ang/papers/icml04-apprentice.pdf

Ramachandran, D., & Amir, E. (2007). Bayesian inverse reinforcement learning. *IJCAI International Joint Conference on Artificial Intelligence*, 2586–2591.

Richard S., S., & Andrew G., B. (2017). Reinforcement Learning: An Introduction. In *The MIT Press* (Vol. 3, Issue 9).

Russell, S. (1998). Learning agents for uncertain environments (extended abstract). *Conference on Computational Learning Theory (COLT)*, 1–3.

Salamon, J., & Bello, J. P. (2017). Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Processing Letters*, *24*(3), 279–283. https://doi.org/10.1109/LSP.2017.2657381

Shao, K., Tang, Z., Zhu, Y., Li, N., & Zhao, D. (2019). *A Survey of Deep Reinforcement Learning in Video Games*. *61573353*, 1–13. http://arxiv.org/abs/1912.10944

Shao, Z., & Er, M. J. (2012). A review of inverse reinforcement learning theory and recent advances. *2012 IEEE Congress on Evolutionary Computation, CEC 2012*, 10–15. https://doi.org/10.1109/CEC.2012.6256507

Vellido, A., Ribas, V., Morales, C., Ruiz Sanmartín, A., & Ruiz Rodríguez, J. C. (2018). Machine learning in critical care: State-of-the-art and a sepsis case study. *BioMedical Engineering Online*, *17*(S1), 1–18. https://doi.org/10.1186/s12938-018-0569-2

Yu, C., Liu, J., & Nemati, S. (2019). *Reinforcement Learning in Healthcare: A Survey*. http://arxiv.org/abs/1908.08796

Yu, C., Liu, J., & Zhao, H. (2019). Inverse reinforcement learning for intelligent mechanical ventilation and sedative dosing in intensive care units. *BMC Medical Informatics and Decision Making*, *19*(Suppl 2). https://doi.org/10.1186/s12911-019-0763-6

Yu, C., Ren, G., & Liu, J. (2019). Deep inverse reinforcement learning for sepsis treatment. *2019 IEEE International Conference on Healthcare Informatics, ICHI 2019*, 1–3. https://doi.org/10.1109/ICHI.2019.8904645

Ziebart, B. D., Maas, A., Bagnell, J. A., & Dey, A. K. (2008). Maximum entropy inverse reinforcement learning. *Proceedings of the National Conference on Artificial Intelligence*, *3*(January), 1433–1438.