Literature Review Matrix - Multi-Agent Inverse Reinforcement Learning

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Authors | Reference | Year | Title | Method | Pros | Cons | Comparison |
| R. Tian  and eg. al | Tian2021 | 2021 | Learning Human Rewards by Inferring Their Latent Intelligence Levels in Multi-Agent Games: A Theory-of-Mind Approach with Application to Driving Data | * Theory-of-Mind * Maximum Entropy Inverse Reinforcement Learning (MaxEnt IRL) * Human’s latent intelligence levels | * Better than (a) Equilibrium solution-based method; (b) Leader-follower roles * Better understanding the real driving data by assigning intelligence levels to multi-agents (Human) | * States are not continuous * Intelligence levels are constant | Dear Dr. Tian,  I just read your paper and want to discuss more details with you.  In your paper, you mentioned that in the real world, the infinite equilibrium solution-based method seems useless. But in our POSG model, we use the real data in Canada’s Burger Markets and find that the equilibrium solution-based method fits the data.  Moreover, I think in our model, the intelligence levels are just the same as the hidden states in our model and we can handle the situation when the levels are changeable over just one agent. |
| R. W. Thomas  And J. D. Larson | Thomas2021 | 2021 | Inverse Reinforcement Learning for Generalized Labeled Multi-Bernoulli Multi-Target Tracking | * GLMB filter as output * Deep RL * Deep IRL | * Work online * Predict agent states/motion * Useful for co-operative multi-agent teams | * Need more test cases * No difference between one agent and multi-agent (No relation about the number of agents) * Don’t try different objects | Dear Dr. Thomas,  In your paper, I only see the simulation results but not real world results. Is it mean the Deep RL or Deep IRL is not suitable for real world case with small scale data. How much data is enough for your method?  In your paper, you said you use Deep RL to recover dynamic part and Deep IRL to recover reward part, I think the name for this algorithm should be two-stage Nested Fixed method which is provided by Rust in 1987.  In your paper, you said your method based on MDP which is just for one agent case. For multi-agent case, it should be Markov Games. |
| T. Oda | Oda2021 | 2021 | Equilibrium Inverse Reinforcement Learning for Ride-hailing Vehicle Network | * SEIRL(spatial equilibrium inverse reinforcement learning) * MaxEnt IRL * Passenger-vehicle matching/ drivers competitive game | * First study to design multi-agent behavioral modeling on road network for a ride-hailing service * Stability of the algorithm * Obtain significant performance gains with unknown dynamics | * All agents are homogeneous * Consider more hidden states, like multiple garage locations or the effect of travel time (caused by congestion) | Dear Dr. Oda,  I just read your paper and want to discuss future details with you.  In your paper, you mentioned that the agents are homogeneous. In our model, we just assume they share same dynamics but with different reward structure which results in hetorgenous among different agents by different policy function.  In your paper, you mentioned that want to take into account the effect of travel time. I think, in our POSG model, we can handle this term as hidden states, which can’t be observed directly. Then introduce agents’ belief for the hidden states term. |
| J. Fu  And eg.al | Fu2021 | 2021 | Evaluating Strategic Structures in Multi-Agent Inverse Reinforcement Learning | * Inverse equilibrium single-agent reduction (IESAR) algorithm * \epsilion-Nash equilibrium * MaxEnt * Reduce the multi-agent problem to N single-agent problems | * Contributes to the computational benefits by reducing the problems to N single-agent problems * Doesn’t require restrictive assumptions | * Strong-rationality assumption * Concerned with Nash equilibria on auctions model. Not the best and suitable solution for the problem * Only policy shared but not utility * Don’t take into account the sampling error (complexity of sample) | Dear Dr. Fu,  After reading you paper, I have the following questions.  In your paper, you don’t mentioned the hidden states situation. Namely, imperfect information.  In your paper, you use \epsillion-Nash equilibrium for the Game’s solution to avoid more restrictions. But in our POSGs, perfect markov equilibrium can reach the same goal and achieve the same results. |
| S. Bergerson | Bergerson2021 | 2021 | Multi-Agent Inverse Reinforcement Learning: Suboptimal Demonstrations and Alternative Solution Concepts | * Theory of Mind * Nash Equilibrium * MaxEnt | * Extended the MaxEnt method from one agent to multi agent * Contributes to suboptimal decision making problems (with bias, noise and heuristics) * Introducing of ToM provides more flexible methods | * Real-world data * Benefits in more complicated enviroments | Dear Dr. Bergerson,  I’m wondering if the real world case is specific means the social environment cases. In your paper, you only mentioned nose, biases and heuristics of the decision makers. But I think, we can dealing this issue from another direction. Namely, instead of thinking the reationality of decision makers (the imperfect of decision makers), we can focus on the imperfect information the decision makers can get. Namely, use the partial observable stochastic games to handle the issue. |
| X. Yang  And eg.al | Yang2020 | 2020 | Student Subtyping via EM-Inverse Reinforcement Learning | * EM-IRL (expectation-maximization IRL) * Group student subtypes into 3: learning-oriented/efficient-oriented/no-learning | * Dynamics not static data in K-means * Robust performance over simulation data * Effectively group decision makers(student) over real dataset | * Find method to detect students’ strategy as early as possible * More empirical studies to figure out the effectivity of EM-IRL | Dear Dr. Yang,  In your paper, you don’t mention EM-IRL’s convergence and identification issue.  Besides, the meaning of states in your model is defined as learning environment. I think learning environment is difficult to define and is an imperfect information. If replace the MDP process in EM-IRL by POMDP, namely, introduce a hidden states to interpret learning environment, the results should be better.  Moreover, I think we can further induce the competition or coorperate relation between different group of students into the model. |
| W. Jeon  And eg.al | Jeon2020 | 2020 | Scalable Multi-Agent Inverse Reinforcement Learning via Actor-Attention-Critic | * MA-AIRL(Multi-agent adversarial inverse reinforcement learning) * Discriminators (min-max objective) * Multi-agent actor-attention-critic (MAAC), i.e, off-policy actor-critic | * More scalable and sample-efficient * Using various types of discriminators (decentralized and centralized, here decentralized means each agent with single agent function) | * Not obvious out perform in centralized discriminator (each agent with multi-head function) | Dear Dr. Jeon,  In your paper, you assume different agents have different observations for the same hidden states. But in our model, we think different agents should share the same observations for the same hidden states which is a common knowledge among all the agents.  The objective function in adversary game is different from ours. We use max likelihood method which can deal with both adversary game and cooperate game. |
| L. Yu  And eg.al | Yu2019 | 2019 | Multi-Agent Adversarial Inverse Reinforcement Learning | * MA-AIRL * Maximum pseudolikelihood estimation * Adversarial reward learning framework (min max problem) | * Effective and scalable for Markov games with high-dimensional and unknown dynamics * Tractability by maximum pseudolikelihood | * Regularization in reward * Overfitting and leverage prior knowledge | Dear Dr. Yu,  In your model, you use the Markov Game, which assume each agent process follow the Markov Decision Process. I think we can use POMDP to replace MDP. |
| R. Raileanu  And eg.al | Raileanu2018 | 2018 | Modeling Others using Oneself in Multi-Agent Reinforcement Learning | * Self Other-Modeling (SOM) * Imperfect Information * Neural network as value function to avoid parametrize reward function * Online * Deep multi-agent RL | * Better performance than considering the other agent as environment * Flexibility (due to online) to numerous other environments and tasks * Simplicity (all agents share same network, namely same parameter in network) | * The agents are identical resulting in being more suitable for symmetric games * Assume stationary strategies not non-stationary strategies * More complex enviroments | Dear Dr. Raileanu,  I don’t think your work can be multi-agent RL. In your work, you assume that the transition matrix for two agents are identically distributed. It means that there is no relation or interaction between the two agents. So multi-agent is just multiple one agent making decisions at the same time. There is no influence between the agents and your model is not even the symmetric game.  As for the offline part, online is used only because deep RL needs more parameter and large data set to train the network. In most cases, we assume the reward structure to be linearly or quaratic is not based only on the domain knowledge, but more on the structure is work in most situation and simple. Just like, fully connection layer is usually used in neural network.  As for the imperfect information part, you mentioned that you use belief term to handle the hidden state of other agents. But it should be like that. Hidden state is hidden to both agents and unknown environment information. |
| X. Lin  And eg.al | Lin2018 | 2018 | Multiagent Inverse Reinforcement Learning for Two-Person Zero-Sum Games | * Multiagent inverse reinforcement learning (MIRL) * Two-agent zero-sum game (soccer game) * Minimax bipolicy * Bayesian MIRL | * Extension IRL to MIRL on zero-sum stochastic games * Offer insights into agents’ behave or dynamic change | * Multiple feasible solutions resulting in difficulty of selecting reward function * More general scenaros, like bipolicy is unobservable * Transition matrix is known * N-player, general-sum case | Dear Dr. Lin,  I think your paper is a good motivation for multi-agent stochastic game. Although you only extend IRL to MIRL on two-agent zeros-sum game with perfect information, it still inspires to our work. Right now, our work can extend IRL to MRIL on multi-agent stochastic game with imperfect information by Markov Perfect Equilibrium Solution.  In your work, you use min max problem setting. To simplify the case, in our work, we assume each agents maximize their own reward separately to avoid the min max solution. Although it still occurs multiple feasible solutions, it’s more general and suitable for both competitive and coordinate game. |
| S. Natarajan  And et .al | Natarajian2010 | 2010 | Multi-Agent Inverse Reinforcement Learning |  |  |  |  |
| S. Russell | Russell1998 | 1998 | Learning agents for uncertain environments (extended abstract) |  |  |  |  |

Literature Review Matrix – Partially Observable Stochastic Games/ Markov Games

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Authors | Reference | Year | Title | Method | Pros | Cons | Comparison |
| V. Kovarik  And et.al | Kovarik2022 | 2022 | Rethinking formal models of partially observable multiagent decision making |  |  |  |  |
| R. Ceren  And et.al | Ceren2021 | 2021 | PALO bounds for reinforcement learning in partially observable stochastic games |  |  |  |  |
| K. Horak  And et.al | Horak2020 | 2020 | Solving Zero-Sum One-Sided Partially Observable Stochastic Games | * Theoretical analysis of one-sided POSGs * Convergency of value-iteration algorithm * Heuristic search value-iteration algorithm to approximate solution for scalability * Use approximate value functions to derive strategies * Can solve one-sided POSGs of non-trivial sizes | * The first approximate algorithm that generalizes the ideas behind point-based algorithms * Possible to translate selected results form the single-agent setting to zero-sum games | * Scalability of the algorithm in specific security games * More heuristics and methods which can be translated into game-theoretic setting * Generalization beyond the strictly adversarial setting (eg. Computing Stackelberg equilibrium) | Dear Dr. Horak,  In this paper, you generalized your previous work form theoretical part to application part. But still for the specific game, namely, the one-sided zero-sum POSGs in security area. |
| E. Wei  And eg.al | Wei2019 | 2019 | Multiagent Adversarial Inverse Reinforcement Learning | * Cooperative continuous stochastic games (multiagent reinforcement learning with continuous actions and identical reward for all agents) * Multiagent Acotr-Critic model in Maximum Entropy Reinforcement Learning (MERL) framework * Method 1: GAIL (Generative Adeversarial Imitation Learning) to cooperative game * Method2: AIRL (Adversarial Inverse Reinforcement Learning) to cooperative game | * Better coordination in multiagent cooperative tasks, especially than MADDPG method * Avoid relative overgeneralization | * Input space of the discriminator grow linearly with the number of agents | Dear Dr. Wei,  In your work, you mentioned you extend competitive game’s method like GAIL and AIRL into cooperative game. But the method is more rely on network and deep RL method. The drawback is the dimension of the parameter in the model. So I think traditional method by using Nash Equlibrium in the model will get better results. |
| V. Tangkaratt  And et.al | Tangkaratt2019 | 2019 | VILD: Variational Imitation Learning with Diverse-quality Demonstrations | * VILD | * Learn reward function and the level of demonstrators’ expertise * Scalable and data-efficient | * Efficiently estimate parameters | Dear Dr. Tangkaratt,  In your paper, you mentioned using VILD to learn the reward structure and label the demonstrator’s level at the same time. But your model is based mainly on single agent. Maybe we can extend this area to multi agent with multiagent reinforcement learning method. |
| K. Horak  And et.al | Horak20191 | 2019 | Compact Representation of Value Function in Partially Observable Stochastic Games | * Projecting high-dimensional beliefs to characteristic vectors of significantly lower dimension * Novel compact representation of the belief (compared with PCA) * One-sided POSG (subclasses of POSGs) * Cybersecurity example | * Scalability improvement in large hidden states space dimension * Domain-independent * Practical aspects of algorithms for solving subclasses of POSGs |  | Dear Dr. Horak,  In your work, you construct a compact presenting (PCA like method) to shrink down the dimension of hidden states and improve the scalability in large dimension states space problem. But the method is still only suitable for one-side zero-sum POSGs. |
| K. Horak  And et.al | Horak20192 | 2019 | Optimizing honeypot strategies against dynamic lateral movement using partially observable stochastic games | * Real-world applications—cybersecurity * Honeypots for lateral movement * One-Sided POSGs for Lateral Movement POSG * Novel compact representation for scalability | * Dramatically improve the ability to solve Lateral Movement POSG. |  | Dear Dr. Horak,  In your work, you use the one-side POSG algorithm specifically to the real-world application in cybersecurity. Achieve the convergency and scalability of the problem. But the one-side POSG is only for two-agent defender and attacker problem which is too specific. |
| K. Horak  And B. Bosansky | Horak20193 | 2019 | Solving Partially Observable Stochastic Games with Public Observations | * Two-player zero-sum games with public observation * Both agents only get imperfect information of each other * PO-POSG (public observation-POSG) * An observation for each player is generated and made publicly known to both players | * Scalability in hundreds of states * First practical general algorithm to solve both players lack information about the game state | * Improve scalability furthermore * Application in security domains (e.g. cybersecurity) real-worl problems * Adapt different objective functions * Relax the factorization over the states | Dear Dr. Horak,  In your work, you construct a new zero-sum POSGs. In your game, each agents have their own hidden states which is kept only to themselves. But the agents can observe the hidden states of other agents and this observations are known to all the agents. In our work, we assume the hidden states is from the environment which is not known to both agents, instead of generating directly by the private information of agents. To handle the hidden states, we also introduce the observations which is the same and known to both agents. |
| K. Horak  And et. al | Horak2017 | 2017 | Heuristic Search Value Iteration for One-Sided Partially Observable Stochastic Games | * Two-player zero-sum one-sided POSG * Heuristic search value iteration (HSVI) | * Guarantees the value functions are convex * Convergency to optimal values * Approximates optimal strategies * Applicability and scalability | * Scalability is limited | Dear Dr. Horak,  In your work, you constructed a HSVI method for two-agent zero-sum POSGs with one-side information. Namely, one agent has perfect information but the another agent has imperfect information. I think the setting of this game is limited and can be future generalization. |
| A. Kumar  And S. Zilberstein | Kumar2009 | 2009 | Dynamic Programming Approximations for Partially Observable Stochastic Games | * Epsilon pruning based approximation technique * MBDP-AS which is an any-space algorithm | * Provides better error bound * scalability |  | Dear Dr. Kumar,  In you work, you provide a method to estimate approximation solution for POSGs. But no more theoretical proof inside. |
| A. Guo  And V. Lesser | Guo2005 | 2005 | Planning for Weakly-Coupled Partially Observable Stochastic Game | * Computing Nash equilibrium and performing strategy elimination * NP-hard in theoretical | * Able to solve much larger problems * Reduce the size of the policy space by several orders of magnitude | * Only based on simulation 2-agent game | Dear Dr. Guo,  In your work, you mentioned that your work provided a algorithm and solution for POSGs both in theoretical part and large data practical part. But I don’t find more theoretical support in more general situation. It seems that it can only be applied to 2 agent specific game. |
| E. A. Hansen  And et.al | Hansen2004 | 2004 | Dynamic Programming for Partially Observable Stochastic Games | * Dynamic Programming * Iterated elimination of dominated strategies (finite-horizon POSGs) * Multi-agent dynamic programming operator | * Can be applied to any number of players * First exact algorithm for general POSGs * Can find optimal solutions for cooperative POSG | * Improving efficiency (prune policy trees) * Extension to infinite-horizon POSGs (infinite trees) | Dear Dr. Hansen,  In your work, you use dynamic programming to get the solution for POSGs, which is only suitable for small game. The reason is the limitation of efficiency and memory. I think we can use the combination of Game theory and IRL method to deal with this problem. |
| R. Emery-Montemerlo  And et.al | Emery2004 | 2004 | Approximate Solutions For Partially Observable Stochastic Games with Common Payoffs | * Decentralized (agents sharing same reward) * Heuristics Algorithm (transforming POSGs intos a series of smaller Bayesian games) | * Computationally efficient to real world * Allow to find solutions to much larger problem | * Locally optimal * Performance of approximations * Centralized POMDPs * Improve efficiency | Dear Dr. Emery-Montemerlo,  In your work, you try to divide the POSGs into small Baysianse game and get one-step policy for each small game to achieve efficiency in larger dataset. But I think, instead of trading off the optimal, using the univariant infinite policy and get estimation for the policy can be better. |
| L. S. Shapley | Shapley1953 | 1953 | STOCHASTIC GAMES |  |  |  |  |

Literature Review Matrix – Economy

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Authors | Reference | Year | Title | Method | Pros | Cons | Comparison |
| Y. Chang  And et al. |  | 2022 | Structural Estimation of Partially Observable Markov Decision Processes |  |  |  |  |
| Y. Chang  And et al. | Chang2020 | 2020 | Dynamic Discrete Choice Estimation with Partially Observable States and Hidden Dynamics |  |  |  |  |
| V. Aguirregabiria and P. Mira | Aguirregabiria20191 | 2019 | Identification of games of incomplete information with multiple equilibria and unobserved heterogenetiy |  |  |  |  |
| V. Aguirregabiria and A. Megesan | Aguirregabiria20192 | 2019 | Identification and Estimation of Dynamic Games When Players’ Beliefs Are Not in Equilibrium |  |  |  |  |
| W. Shi and et al | Shi2019 | 2019 | Soft policy gradient method for maximum entropy deep reinforcement learning |  |  |  |  |
| T. Haarnoja et. al. | Haarnoja2018 | 2018 | Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor |  |  |  |  |
| T. Haarnoja and et.al | Haarnoja2017 | 2017 | Reinforcement learning with deep energy-based policies |  |  |  |  |
| M. IGAMI and N. Yang | Igami2016 | 2016 | Unobserved heterogeneity in dynamic games: Cannibalization and preemptive entry of hamburger chains in Canada |  |  |  |  |
| C. L. Su and K. L. Judd | Su2012 | 2012 | Constrained optimization approaches to estimation of structural models. |  |  |  |  |
| P. Bajari and et al. | Bajari2009 | 2009 | Nonparametric and Semiparametric Analysis of a Dynamic Discrete Game |  |  |  |  |
| V. Aguirregabiria and P. Mira | Aguirregabiria2007 | 2007 | Sequential estimation of dynamic discrete games |  |  |  |  |
| K. E. Train | Train2009 | 2009 | Discrete choice methods with simulation |  |  |  |  |
| V. Aguirregabiria and P. Mira | Aguirregabiria2002 | 2002 | Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models |  |  |  |  |
| R. John | John1988 | 1988 | Maximum likelihood estimation of discrete control processes |  |  |  |  |
| J. Rust | Rust1987 | 1987 | Optimal replacement of GMC bus engines: and empirical model of Harold zurcher |  |  |  |  |

Literature Review Matrix – Belief Convergency

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Authors | Reference | Year | Title | Method | Pros | Cons | Comparison |
| A. M. Davis  And et.al | Davis2021 | 2021 | Consumer Learning from Own Experience and Social Information: An Experimental Study | * Divided social information into 4 class: no social information; shar-based social information; quality-based social information; full social information * Repeated two-armed Bernoulli bandit experiment * Hidden Markov Chain framework * Consumer beliefs | * Social information is important in complex optimal decision0making rules * Larger difference: quality-SI increases the higher quality firm’s market share, leads to the lowest demand uncertainty and fastest convergence to steady state * Share-SI reduces demand uncertainty both large and small gaps in service-quality levels * Studying different types of social information | * Experiments no real data set * In practice, how to define small- or large- gap * Theory about different social information * Enrich empirical and experimental evidence * Asymmetric strategies of firms with different social information | The big idea here, the author uses Hidden Markov Chain framework which can be proof in our previous work, is just the same as Partially Observable Markov Chain. The author divided the Social information into 4 type. Among which, quality social information and share social information are studied the most. Although they results in different types of conclusion between small gap firms and larger gap firms, it turns out the social information can lower down the demand uncertainty in both cases (different kinds of firms). |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Literature Review Matrix – Canada Business Cycle

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Authors | Reference | Year | Title | Method | Inputs | Outputs | Comparison |
| Garima Vasishtha, Philipp Maier |  | 2013 | The impact of the global business cycle on small open economics: a favar approach for Canada | * Large factor-augmented VAR(FAVAR) model | * Global developments * Commodity prices, foreign economic activity, foreign interest rates (higher relative) * Global interest rates and global inflation (lower relative) | Canadian economy | Previous studies have evaluated the transmission of  foreign shocks, mainly from the U.S., using small scale econometric models, such as vector autore-  gression models (VARs) or structural VARs (for instance, Burbidge & Harrison, 1985; Johnson &  Schembri, 1990; Kuszzak & Murray, 1987; Souki, 2008). Klyuev (2008) focuses on the impact of U.S.  financial conditions on financial conditions and real economic activity in Canada.  Souki (2008), using a SVAR including U.S. and Canadian real output  growth and inflation changes analyzes the importance of U.S. supply and demand shocks in Canadian  real GDP and inflation fluctuations |
| Marcel-Cristian Voia, J.Stephen Ferris |  | 2013 | Do business cycle peaks predict election calls in Canada? | * Not sure where the business cycle peak from |  |  |  |
| Hideki Ariizumi, Yaqin Hu, Tammy Schirle |  | 2015 | Stand together or alone? Family structure and the business cycle in Canada | * Linear regression to me | * Family formation and structure * Labour Force Survey * Marriages * Unemployment rate | * Business cycle fluctuations |  |
| J. Stephen Ferris |  | 2008 | Electoral politics and monetary policy: does the bank of Canada contribute to a political business cycle | * I don’t think really relavant | * Monetary policy * Election of a Liberal party government |  | work of Abrams and Iossifov (Public Choice 129:249– 262, 2006) to monetary policy in Canada |
| Harper, Stephen Joseph |  | 1991 | The Political business cycle and fiscal policy in Canada | * Another election, political goals, multivariate |  |  |  |
| Shyh-Wei Chen | Really Important | 2006 | Simultaneously modeling the volatility of the growth rate of real GDP and determining business cycle turning points: evidence from the U.S., Canada and the UK | * Identify business cycle turning points with GDP | * Growth rate of real GDP |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |