Discriminative System Identification via the Principle of Maximum Causal Entropy

Xiangli Chen XCHEN40@UIC.EDU
University of Illinois at Chicago, 851 S. Morgan St., Chicago, IL 60607 USA

Brian D. Ziebart BZIEBART@UIC.EDU
University of Illinois at Chicago, 851 S. Morgan St., Chicago, IL 60607 USA

Extended Abstract

System identification addresses the important problem of estimating the answers to *what will happen if...* questions. In black-box system identification, minimal assumptions are made about the unknown system and its dynamics are estimated entirely from control input and observational data. Black-box methods for identifying linear systems are well-established (Ljung, 1999). The perspective of these methods is often taken in non-linear black-box system identification as well, with a regression-based focus on squared loss minimization (Sjöberg et al., 1995).

With the black-box motivation of making as few assumptions about the unknown system as possible, we frame the system identification problem as a robust statistical estimation task by following the principle of maximum causal entropy (Ziebart et al., 2010; 2013). It prescribes the least committed (or most uncertain) stochastic process possible that matches observed properties of the unknown system. This formulation provides a discriminative approach for estimating an unknown controlled dynamical system that robustly minimizes the predictive log loss for observation sequences when the system is driven by a known control policy.

Controlled Systems as Interacting Processes

A controlled dynamical system can be viewed as a sequence of system observations, $\mathbf{O}_{1:T}$, and a sequence of control actions, $\mathbf{A}_{1:T}$ (both random variable sequences). Estimates of future observations given previous observations and controls, $\mathcal{T}(o_t|\mathbf{o}_{1:t-1}, \mathbf{a}_{1:t-1})$, are needed to construct good control policies, $\pi(a_t|\mathbf{a}_{1:t-1}, \mathbf{o}_{1:t-1})$.

In this work, we employ the causally conditioned probability, denoted as $P(\mathbf{y}_{1:T}||\mathbf{x}_{1:T}) \triangleq \prod_{t=1}^{T} P(y_t|\mathbf{y}_{1:t-1}, \mathbf{x}_{1:t})$ or $P(\mathbf{x}_{1:T}||\mathbf{y}_{1:T-1}) \triangleq \prod_{t=1}^{T} P(x_t|\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1})$, from directed information theory (Marko, 1973; Massey, 1990; Kramer, 1998) to represent the interactions between these variable sequences. Using $\{P(\mathbf{y}_{1:T}||\mathbf{x}_{1:T})\}$, we denote a fully defined causally conditioned probability distribution (i.e., for all pairs of sequences). The joint probability distribution of observation and control sequences factors into two interacting processes,

$$P(\mathbf{a}_{1:T}, \mathbf{o}_{1:T}) = \pi(\mathbf{a}_{1:T}||\mathbf{o}_{1:T})\, \mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1}), \quad (1)$$

a known control policy, $\pi(\mathbf{a}_{1:T}||\mathbf{o}_{1:T}) \triangleq \prod_{t=1}^{T} \pi(a_t|\mathbf{a}_{1:t-1}, \mathbf{o}_{1:t})$, and the system's dynamical process governing generated observations, $\mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1}) \triangleq \prod_{t=1}^{T} \mathcal{T}(o_t|\mathbf{o}_{1:t-1}, \mathbf{a}_{1:t-1})$. The key characteristic of these causally conditioned probabilities are that future values (of states and controls) do not influence earlier conditioned variables (states and controls). This is an important contrast with conditional probability distributions, $P(\mathbf{y}_{1:T}|\mathbf{x}_{1:T}) = \prod_{t=1}^{T} P(y_t|\mathbf{y}_{1:t-1}, \mathbf{x}_{1:T})$, which are conditioned on the entire sequence, $\mathbf{x}_{1:T}$.

A Robust Estimation Formulation

Our solution to the task of estimating the unknown process, $\mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T})$, builds upon the principle of maximum entropy (Jaynes, 1957). This principle provides a robust statistical estimation framework (Topsøe, 1979; Grünwald & Dawid, 2004) for structured prediction tasks corresponding to unknown joint, marginal, or conditional probability distributions. Conditional random fields (Lafferty et al., 2001) are a very successful example of the approach in practice for estimating conditional distributions, $P(\mathbf{y}_{1:T}|\mathbf{x}_{1:T})$.

Recently, the principle of maximum causal entropy (Ziebart et al., 2010; 2013) has extended the applicability of the maximum entropy approach to settings of interacting processes. This approach has been previously employed for predictive inverse optimal control tasks of estimating control policies in decision processes with known dynamics. For example, predicting the motion trajectories of mouse cursors in pointing tasks (Ziebart et al., 2012). System identification investigates the opposite problem: estimating system dynamics under a known control policy.

In the system identification setting, a natural desire is to minimize the predictive loss of each observation variable in the sequence: $\text{loss}(\{\hat{\mathcal{T}}(o_t|\text{history})\}, \{\mathcal{T}(o_t|\text{history})\})$, where the first probability distribution terms, $\hat{\mathcal{T}}(o_t|\text{history})$, are from the estimated distribution and the second, $P(o_t|\text{hist.})$, is the true distribution (each given the history of previous observations and control actions, $\mathbf{o}_{1:t-1}, \mathbf{a}_{1:t-1}$). When employing the conditional log loss, which expresses the number of bits of information needed to describe a sample from the true distribution, $P(o_t, \text{hist.})$, using an encoding scheme based on the estimate $\{\hat{\mathcal{T}}(o_t|\text{hist.})\}$, the sum of all of these losses reduces to[1]:

$$-\sum_{t=1}^{T}\sum_{\text{hist.},o_t} P(o_t, \text{hist}) \log_2 \hat{\mathcal{T}}(o_t|\text{hist.}) \tag{2}$$
$$= -\sum_{\mathbf{o}_{1:T},\mathbf{a}_{1:T}} \mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\pi(\mathbf{a}_{1:T}||\mathbf{o}_{1:T}) \log_2 \hat{\mathcal{T}}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})$$

This sequence loss, known as the causal log loss (Ziebart et al., 2013), corresponds to the expected number of bits of information needed to describe a sample from the sequence using an encoding based on the estimate $\hat{\mathcal{T}}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T})$.

Following the adversarial log-loss minimization argument of Topsøe (1979) and Grünwald & Dawid (2004), Theorem 1 establishes the relationship between robust causal log-loss minimization and maximizing causal entropy.

Theorem 1. *When Lagrangian duality holds (i.e., mild feasibility requirements), the process estimate that robustly minimizes the causal log-loss of Equation 2, $\{\hat{\mathcal{T}}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\}$, is equivalent to maximizing a causally conditioned entropy (Ziebart et al., 2010; 2013),*

$$H(\mathbf{O}_{1:T}||\mathbf{A}_{1:T-1}) \triangleq \sum_{t=1}^{T} H(O_t|\mathbf{O}_{1:t-1}, \mathbf{A}_{1:t-1}) = \tag{3}$$
$$-\sum_{\mathbf{o}_{1:T},\mathbf{a}_{1:T}} \mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\pi(\mathbf{a}_{1:T}||\mathbf{o}_{1:T}) \log_2 \mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1}),$$

which is a measure of the uncertainty of the process $\mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})$.

Proof. Robust log loss minimization is formulated as a sequential game in which an observation dynamics estimate is first chosen, $\{\hat{\mathcal{T}}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T})\}$, followed by an adversarially chosen evaluation dynamics, $\{\mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T})\}$:

$$\min_{\{\hat{\mathcal{T}}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\}} \max_{\{\mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\}\in\Xi} \tag{4}$$
$$-\sum_{\mathbf{o}_{1:T},\mathbf{a}_{1:T}} \mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\pi(\mathbf{a}_{1:T}||\mathbf{o}_{1:T}) \log_2 \hat{\mathcal{T}}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})$$
$$= \max_{\{\mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\}\in\Xi} \min_{\{\hat{\mathcal{T}}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\}} \tag{5}$$
$$-\sum_{\mathbf{o}_{1:T},\mathbf{a}_{1:T}} \mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\pi(\mathbf{a}_{1:T}||\mathbf{o}_{1:T}) \log_2 \hat{\mathcal{T}}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})$$
$$= \max_{\{\mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\}\in\Xi} H(\mathbf{O}_{1:T}||\mathbf{A}_{1:T-1}), \tag{6}$$

in which the first equality follows from Lagrangian duality, and the second from solving the internal optimization, $\hat{\mathcal{T}}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1}) = \mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})$. □

Note that in this adversarial formulation of the dynamics estimation task, the adversarially chosen dynamics are constrained to match certain known properties, denoted by the set Ξ.

Constraints for Generalized Predictions

Choosing appropriate properties of observed action-observation sequences for the estimated process to imitate or match is very important. The most salient properties of the unknown process should be employed (Jaynes, 1957). Using a large set of properties to narrowly define the constraint set Ξ (of Equation 6) will lead to overfitting when limited amounts of data are available. However, constraints that define the set Ξ too loosely will not provide good predictive performance either.

In this work, we consider constraints that require the observation dynamics process to be similar to statistics of observed trajectory sequences, $f(\mathbf{o}_{1:T}, \mathbf{a}_{1:T}) \in \mathbb{R}^K$, in expectation:

$$\{\mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\} \in \Xi \text{ iff} \tag{7}$$
$$|\mathbb{E}_{\mathcal{T},\pi_i}[f(\mathbf{O}_{1:T}, \mathbf{A}_{1:T})] - \mathbf{c}_i|_{(1|2)} \leq \epsilon_i \qquad \forall i,$$

where: each π_i is a different employed policy;

$$\mathbf{c}_i = \mathbb{E}_{\tilde{\mathcal{T}},\tilde{\pi}}\left[f(\mathbf{O}_{1:T}, \mathbf{A}_{1:T})\right] \tag{8}$$
$$= \sum_{\mathbf{o}_{1:T},\mathbf{a}_{1:T}} \tilde{\mathcal{T}}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\tilde{\pi}(\mathbf{a}_{1:T}||\mathbf{o}_{1:T}) f(\mathbf{o}_{1:T}, \mathbf{a}_{1:T})$$

[1] Our notation assumes univariate discrete-valued observation and action spaces in our notation, but the approach is also applicable to multivariate and/or continuous-valued spaces.

is a vector of statistics collected from empirical observation of that policy's interactions with the dynamical system; and ϵ_i is a bound on the L_1 or L_2 norm of the difference from past observations based on the sample size (and expert knowledge of the similarity between policies).

Maximum Causal Entropy System Identification

Combining the maximum causal entropy objective (Equation 3)—motivated as a robust log-loss minimizer—and the similarity constraints between observed action-observation sequences (Equation 7), yields a constrained optimization problem:

$$\max_{\{\mathcal{T}(\mathbf{o}_{1:T}||\mathbf{a}_{1:T-1})\}} H(\mathbf{O}_{1:T}||\mathbf{A}_{1:T-1}) \tag{9}$$

$$\text{such that:}$$

$$|\mathbb{E}_{\mathcal{T},\pi_i}[f(\mathbf{O}_{1:T},\mathbf{A}_{1:T})] - \mathbf{c}_i|_{(1|2)} \leq \epsilon_i \qquad \forall i,$$

where either the L_1 or L_2 norm is employed.

The objective function of Equation 9 is concave and its set of constraints define a convex set. Thus, standard convex optimization techniques can be employed to obtain the maximum causal entropy distribution. However, the primal optimization is defined over a large set of variables: the causally conditioned probabilities, $P(\mathbf{o}_{1:T}||\mathbf{a}_{1:T})$. As a more efficient alternative, we instead investigate the dual optimization problem.

Parametric Distribution Form

Following the previously developed theory of maximum causal entropy (Ziebart et al., 2010; 2013), we consider the dual optimization problem for $\epsilon = \mathbf{0}$ (for simplicity). Via Lagrangian duality, the decomposed conditional probabilities of the transition dynamics process in the optimization of Equation 9 has a form that satisfies the following set of equations:

$$\forall \mathbf{o}_{1:T}, \mathbf{a}_{1:T}, \quad -\sum_{t=1}^{T} \log \mathcal{T}(o_t|\mathbf{o}_{1:t-1},\mathbf{a}_{1:t-1}) \tag{10}$$

$$+ \sum_i \frac{\pi_i(\mathbf{a}_{1:T}||\mathbf{o}_{1:T})}{\pi(\mathbf{a}_{1:T}||\mathbf{o}_{1:T})} \theta_i^{\mathrm{T}} f(\mathbf{o}_{1:T},\mathbf{a}_{1:T}) = Z(\mathbf{a}_{1:T}),$$

in addition to constraints requiring the $\mathcal{T}(o_t|\mathbf{o}_{1:t-1},\mathbf{a}_{1:t-1})$ terms to define a valid conditional probability distribution (normalization, non-negativity).

The exact form of the distribution depends on the specific sequence statistic function, $f(\mathbf{o}_{1:T},\mathbf{a}_{1:T})$. When this function additively factors over timesteps, $f(\mathbf{o}_{1:T},\mathbf{a}_{1:T}) = \sum_{t=1}^{T-1} g(o_t, o_{t+1}, a_t)$, the resulting distribution will be first-order Markovian, recursively defined as follows:

$$\mathcal{T}(o_t|o_{t-1},a_{t-1}) \triangleq e^{Q(o_t,o_{t-1},a_{t-1})-V(o_{t-1},a_{t-1})}, \tag{11}$$

where:

$$\begin{aligned} Q(o_t,o_{t-1},a_{t-1}) &\triangleq \sum_{a_t} \pi(a_t|o_t)V(o_t,a_t) \\ &\quad + \sum_i \frac{\pi_i(a_t|o_t)}{\pi(a_t|o_t)} \theta_i^{\mathrm{T}} g(o_t,o_{t-1},a_t) \\ V(o_{t-1},a_{t-1}) &\triangleq \operatorname*{softmax}_{o_t} Q(o_t,o_{t-1},a_{t-1}), \end{aligned}$$

and $\text{softmax}_x\, f(x) = \log \sum e^{f(x)}$.

We note that this recursive definition is closely related to the value iteration algorithm (Bellman, 1957) for optimal control. Thus, this approach to system identification will in general non-myopically make predictions so that the future interactions between the control policy and the unknown system dynamics will be similar to previously observed behavior.

The sets of Lagrange multipliers, $\theta_i \in \mathbb{R}^K$, are chosen to satisfy the constraints of Equation 9. This is equivalent to a maximum likelihood estimation problem, where the observation-action trajectories from other policies, π_i, are probabilistically re-weighted by the ratios of policy probabilities. Note that the singularities of this reweighting ($\pi(\mathbf{a}_{1:T}||\mathbf{o}_{1:T}) = 0$) can be interpreted as allowing the system to behave deterministically in a way that most satisfies the optimization constraints.

Further, if these local functions are quadratic (with real vector-valued observations and controls), and the control policy is conditional Gaussian,

$$g(o_t,o_{t+1},a_t) = \text{vec}\left(\begin{bmatrix} o_t \\ o_{t+1} \\ a_t \end{bmatrix}\begin{bmatrix} o_t \\ o_{t+1} \\ a_t \end{bmatrix}^{\mathrm{T}}\right),$$

the observation transition dynamics, $\mathcal{T}(o_t|o_{t-1},a_{t-1})$, will be conditional normal probability distributions. The can then be efficiently computed in closed form, but are trained discriminatively rather than generatively. For inverse optimal control, this type of model has been recently investigated (Ziebart et al., 2012; Levine & Koltun, 2012).

The relaxation to softer constraints ($\epsilon > \mathbf{0}$) introduces L_1 or L_2 regularization terms for the distribution parameters in the dual optimization problem (Dudík & Schapire, 2006).

Related Work

There has been an ongoing debate between the use of generative and discriminative techniques in statistical machine learning (Ng & Jordan, 2002; Jebara, 2004; Ulusoy & Bishop, 2005; Liang & Jordan, 2008). Discrimative techniques for estimating dynamical systems have primarily focused on the problem of estimating a state sequence given a noisy observation sequence (Abbeel et al., 2005; Kim & Pavlovic, 2009). Exponential family predictive state representations (Wingate, 2007) have similar motives to our maximum causal entropy system identification approach, but are not trained by conditioning on the (stochastic) control policy and do not minimize the log loss over the entire sequence of outputs.

Conclusions

In this work, we have developed a framework for discriminative system identification using the principle of maximum causal entropy. It provides robust predictive guarantees (Theorem 1) and a recursive definition that can either be solved using dynamic programming in discretely-valued action-observation spaces, or in closed form for specific types of constraint functions and policies in continuous action-observation spaces.

References

Abbeel, Pieter, Coates, Adam, Montemerlo, Michael, Ng, Andrew, and Thrun, Sebastian. Discriminative training of Kalman filters. In *Proc. of Robotics: Science and Systems*, 2005.

Bellman, R. A Markovian decision process. *Journal of Mathematics and Mechanics*, 6:679–684, 1957.

Dudík, Miroslav and Schapire, Robert E. Maximum entropy distribution estimation with generalized regularization. In *Proc. Computational Learning Theory*, pp. 123–138, 2006.

Grünwald, P. D. and Dawid, A. P. Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. *Annals of Statistics*, 32: 1367–1433, 2004.

Jaynes, Edwin T. Information theory and statistical mechanics. *Physical Review*, 106:620–630, 1957.

Jebara, Tony. *Machine learning: discriminative and generative*, volume 755. Springer, 2004.

Kim, M. and Pavlovic, V. Discriminative learning for dynamic state prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(10): 1847–1861, 2009.

Kramer, G. *Directed Information for Channels with Feedback*. PhD thesis, Swiss Federal Institute of Technology (ETH) Zurich, 1998.

Lafferty, John, McCallum, Andrew, and Pereira, Fernando. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. International Conference on Machine Learning*, pp. 282–289, 2001.

Levine, Sergey and Koltun, Vladlen. Continuous inverse optimal control with locally optimal examples. In *International Conference on Machine Learning*, 2012.

Liang, Percy and Jordan, Michael I. An asymptotic analysis of generative, discriminative, and pseudolikelihood estimators. In *Proceedings of the 25th international conference on Machine learning*, pp. 584–591. ACM, 2008.

Ljung, Lennart. *System identification*. Wiley Online Library, 1999.

Marko, Hans. The bidirectional communication theory – a generalization of information theory. In *IEEE Transactions on Communications*, pp. 1345–1351, 1973.

Massey, James L. Causality, feedback and directed information. In *Proc. IEEE International Symposium on Information Theory and Its Applications*, pp. 27–30, 1990.

Ng, Andrew Y and Jordan, Michael I. On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In *Advances in neural information processing systems*, pp. 841–848, 2002.

Sjöberg, Jonas, Zhang, Qinghua, Ljung, Lennart, Benveniste, Albert, Delyon, Bernard, Glorennec, Pierre-Yves, Hjalmarsson, Håkan, and Juditsky, Anatoli. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12):1691–1724, 1995.

Topsøe, F. Information theoretical optimization techniques. *Kybernetika*, 15(1):8–27, 1979.

Ulusoy, Ilkay and Bishop, Christopher M. Generative versus discriminative methods for object recognition. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pp. 258–265. IEEE, 2005.

Wingate, David. Exponential family predictive representations of state. In *In Neural Information Processing Systems (NIPS*, 2007.

Ziebart, Brian D., Bagnell, J. Andrew, and Dey, Anind K. Modeling interaction via the principle of maximum causal entropy. In *Proc. International Conference on Machine Learning*, pp. 1255–1262, 2010.

Ziebart, Brian D., Dey, Anind K., and Bagnell, J. Andrew. Probabilistic pointing target prediction via inverse optimal control. In *Proceedings of the ACM International Conference on Intelligent User Interfaces*, pp. 1–10, 2012.

Ziebart, Brian D, Bagnell, J Andrew, and Dey, Anind K. The principle of maximum causal entropy for estimating interacting processes. *Information Theory, IEEE Transactions on*, 59(4):1966–1980, 2013.