Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Learning Large-Scale Dynamic Discrete Choice Models of Spatio-Temporal Preferences with Application to Migratory Pastoralism in East Africa

Stefano Ermon
Stanford University
ermon@cs.stanford.edu

Yexiang Xue
Cornell University
yexiang@cs.cornell.edu

Russell Toth
University of Sydney
russell.toth@sydney.edu.au

Bistra Dilkina
Georgia Tech
bdilkina@cc.gatech.edu

Richard Bernstein
Cornell University
rab38@cornell.edu

Theodoros Damoulas
NYU CUSP
td47@nyu.edu

Patrick Clark
USDA Research Service
pat.clark@ars.usda.gov

Steve DeGloria
Cornell University
sdd4@cornell.edu

Andrew Mude
International Livestock Research Institute
a.mude@cgiar.org

Christopher Barrett
Cornell University
cbb2@cornell.edu

Carla P. Gomes
Cornell University
gomes@cs.cornell.edu

Abstract

Understanding spatio-temporal resource preferences is paramount in the design of policies for sustainable development. Unfortunately, resource preferences are often unknown to policy-makers and have to be inferred from data. In this paper we consider the problem of inferring agents' preferences from observed movement trajectories, and formulate it as an Inverse Reinforcement Learning (IRL) problem . With the goal of informing policy-making, we take a probabilistic approach and consider generative models that can be used to simulate behavior under new circumstances such as changes in resource availability, access policies, or climate. We study the Dynamic Discrete Choice (DDC) models from econometrics and prove that they generalize the Max-Entropy IRL model, a widely used probabilistic approach from the machine learning literature. Furthermore, we develop SPL-GD, a new learning algorithm for DDC models that is considerably faster than the state of the art and scales to very large datasets.

We consider an application in the context of pastoralism in the arid and semi-arid regions of Africa, where migratory pastoralists face regular risks due to resource availability, droughts, and resource degradation from climate change and development. We show how our approach based on satellite and survey data can accurately model migratory pastoralism in East Africa and that it considerably outperforms other approaches on a large-scale real-world dataset of pastoralists' movements in Ethiopia collected over 3 years.

Introduction

A useful and important tool in developing sensible policies for productive land use and environmental conservation is a set of micro-behavioral models that accurately capture the choice process of agents in the system. This is particularly true when we wish to analyze behavioral responses under as-yet unobserved states of the world, such as under alternative policy regimes or climate change. However specifying and fitting suitable models in settings with complex spatio-temporal aspects raises important research challenges, including the "curse of dimensionality" associated with handling large state spaces, and capturing the agents' preferences.

We tackle these issues specifically in the context of migratory pastoralism in the Borena plateau, Ethiopia, which is an exemplar of both the technical aspects of the problem we have in mind, and a setting with crucial policy relevance. Migratory pastoralists manage and herd livestock as their primary occupation. They face uncertainty over shocks to resource availability from drought and climate change, conflict, and disease. During semi-annual dry seasons they must migrate from their home villages to remote pastures and waterpoints, which can be modeled as selection amongst a discrete set of camp sites.

In this movement choice problem the pastoralists face a key tradeoff: they want to locate at the most abundant resource points (measured by observable water and forage), but movement (measured by distance) carries energy costs. Scouting out resource abundance on an ongoing basis also carries effort costs. How they balance these factors affects their response to changes that affect resource abundance (climate change, resource degradation and renewal, waterpoint maintenance), and distance (creation of new waterpoints, land use restrictions). A suitably fitted model of individual preferences over movements yields a number of opportunities for policy-relevant simulation analyses. Policymakers regularly face decisions over land use controls (e.g., zoning for housing, farmland, parkland), waterpoint maintenance (many of the waterpoints in the system are man-made and require ongoing investment) and herd re-stocking after droughts, among others. With the parameters governing pastoralists' individual decisions over movement in hand, we will be able to modify exogenous characteristics of the system such as land access, waterpoint presence, or herd sizes, and simulate predictions about behavioral responses.

Our goal is to develop a model to capture the planning decisions made by the herders. The model must be structural, meaning that its parameters provide intuitive insights into the decision-making process, as well as generative, meaning that it can potentially be used to simulate behavior under new circumstances such as changes in resource availability, access policies, or climate. This raises key challenges due to the relatively large choice set and complex state space.

While methods for structural estimation of behavioral models have been known in economics since at least the 1980s (Rust 1987), methods for estimation in complex spatio-temporal settings are still in their infancy. In parallel, there is a growing literature in computer science under the name of inverse reinforcement learning (IRL), see e.g. (Ng and Russell 2000; Kolter and Ng 2009; Taylor and Stone 2009). While often motivated by a different set of modeling problems, IRL shares the objective of fitting the agent's choice function. In this paper, we study the Dynamic Discrete Choice models from Econometrics, which are widely used in economics (Aguirregabiria and Mira 2010) and engineering (Ben-Akiva and Lerman 1985) but have received little attention so far in the computer science literature. In fact, we show an interesting connection: under some conditions, Dynamic Discrete Choice models generalize the Max-Entropy IRL Model (Ziebart et al. 2008), a widely used approach from the Machine Learning literature. Despite the numerous applications in the economics literature there has been little effort on developing scalable algorithms for learning DDC models on very large datasets. In this paper we fill this gap by developing SPL-GD, a new learning algorithm for Dynamic Discrete Choice models that is considerably faster than the state of the art and scales to very large datasets. Our technique combines dynamic programming with stochastic gradient descent, which is often used to scale machine learning techniques to massive datasets (Bottou and Bousquet 2008).

Our method allows us to infer micro-behavioral models in complex spatio-temporal settings. We apply it in the context of migratory pastoralism in the Borena plateau, Ethiopia. The available data includes surveys from 500 households; static geospatial map layers including village and road locations, ecosystem types, elevation and other terrain features; a dynamic greenness index (NDVI) from satellite sensing; locations of wells, ponds, and other water points identified by interview, field exploration, and satellite imagery; and GPS collar traces of 60 cattle from 20 households in 5 villages, at 5-minute intervals collected over 3 years. The GPS traces are our primary source of information regarding behavior and resource use. We show that using our approach with this data we can accurately model pastoralist movements, and considerably outperform other approaches, including a Markov model and the Maximum Entropy IRL model.

Problem Definition

We consider planning problems represented as *finite* Markov Decision Processes (MDP). Formally, an MDP is a tuple (S, A, P, r, η) where S is a finite set of states, A is a finite set of actions, P is a finite set of transition probabilities and $r : S \mapsto \mathbb{R}$ is an (immediate) reward function (the more general case $r : S \times A \mapsto \mathbb{R}$ can also be handled), and $\eta \in [0, 1]$ is a *discount factor*. If an agent executes an action $a \in A$ while in a state $s \in S$, it receives an immediate reward $r(s)$ and it transitions to a new state $s' \in S$ with probability $P(s'|s, a)$.

Planning. Let the planning horizon T be the (finite) number of time steps that the agent plans for. A plan is represented by a *policy*, where a policy is a sequence of decision rules, one for each time step in the planning horizon. A policy π is called *Markovian* if, for each time step t in the planning horizon, the decision rule $\pi_t : S \to A$ depends only on the current state s_t. If the MDP is deterministic, i.e. $P(s'|s, a) \in \{0, 1\}$, a policy is equivalent to a sequence of T actions (or alternatively states) for each possible initial state. We define the *value* of a policy π from an initial state $s \in S$ as $v^\pi(s) = \mathbf{E}^{s,\pi}\left[\sum_{t=0}^{T-1} \eta^t r(s_t)\right]$, which is the expected value of the discounted total reward when the initial state is $s_0 = s$ and the action taken at time t is chosen according to π.

The goal in a probabilistic planning problem (also known as optimal control or reinforcement learning) is to compute a policy π that maximizes the value function $v^\pi(s)$ for a given MDP, a problem that is widely studied in the literature (Puterman 2009; Bertsekas 1995; Powell 2007).

Inverse Planning Problem

In an inverse planning problem, also known as inverse reinforcement learning (Ng and Russell 2000) or structural estimation of an MDP (Rust 1987), the goal is to identify an MDP that is consistent with observed planning choices made by a rational agent. Specifically, we assume that we are given a state space S, an action set A, transition probabilities P and we want to find a reward function r (intuitively, capturing preferences of the agent over states), which *rationalizes* the observed behavior of an agent. For finite state spaces, the reward function can be represented as vector of real numbers $\mathbf{r} \in \mathbb{R}^{|S|}$, where each component gives the reward for one state. For large state spaces, it is common (Ng and Russell 2000; Ziebart et al. 2008; Powell 2007; Kolter and Ng 2009) to assume linear function approximation of the reward function r, relying on state-based features:

$$r(s) = \theta \cdot \mathbf{f}_s,$$

where $\mathbf{f}_s$ is a *given* feature vector of size m for each state $s \in S$ and $\theta \in \mathbb{R}^m$ is an *unknown* parameter vector to be estimated.

In many practical settings, we do not know the agent's policy π, but can observe the agent's trajectories, i.e. sequences of states from S that are visited, from which we can try to infer the agent's rewards and policy (also known as imitation learning). Specifically, we observe K finite sequences $\mathcal{S} = \{\tau^1, \cdots, \tau^K\}$ of state-action pairs made by the agent. For simplicity of exposition, we assume all sequences have length T, $\tau^k = (s_0, a_0)^k, (s_1, a_1)^k, \ldots, (s_{T-1}, a_{T-1})^k$, where $s_t \in S$ and $a_t \in A$.

If we assume that the trajectories are obtained by following a policy π^*, i.e. for each trajectory k and for each time step t, $(s_t, a_t)^k$ is such that $a_t = \pi^*(s_t)$, then a further natural modeling assumption that captures the rationality of the agents is that the policy π^* is an optimal policy with respect to the (unknown) reward function r (existence of an optimal policy is guaranteed, see (Puterman 2009; Bertsekas 1995)). Formally, this means that the expected policy value is such that $v^{\pi^*}(s) \geq v^{\pi}(s), \quad \forall s \in S, \quad \forall \pi$. Unfortunately, this formulation is known to be ill-posed because it is clearly under-determined. For example, if the reward function $r(\cdot) \equiv 0$, the optimality equation is satisfied by any policy π^*. Additional modeling assumptions are needed to resolve this ambiguity. One option is to introduce a margin (Ng and Russell 2000; Ratliff, Bagnell, and Zinkevich 2006), maximizing the difference between the reward from the optimal policy and its alternatives (Ng and Russell 2000).

Since our final goal is that of informing policy-making, we take a probabilistic approach and focus on *generative models* that can be used to simulate behavior under new circumstances such as changes in resource availability and policies. In the probabilistic approach, we assume the data (i.e., the observed trajectories $\mathcal{S}$) are samples from a family of probability distributions, which depend on the unknown reward function, allowing for suboptimal behavior. Estimating the reward function becomes a statistical inference problem. Notable approaches include Maximum entropy IRL (Ziebart et al. 2008) from the AI literature and Dynamic Discrete Choice models (Rust 1987) from the econometrics literature. We start by reviewing Maximum Entropy IRL (MaxEnt-IRL), which is the most closely related to our approach, and then review logit Dynamic Discrete Choice models (logit DDC), which will be the foundation for the work developed in this paper.

Maximum Entropy Inverse Reinforcement Learning Instead of assuming that the given trajectories follow an optimal policy, Ziebart et al. (Ziebart et al. 2008) assume that each observed trajectory $\tau^k = (s_0, a_0)^k, (s_1, a_1)^k, \ldots, (s_{T-1}, a_{T-1})^k$ is an i.i.d. sample from a probability distribution [1]:

$$P_\theta^{ME}(\tau^k) = \frac{\exp(U_\theta(s_0^k, \cdots, s_{T-1}^k))}{\sum_{\mathbf{s}'} \exp(U_\theta(\mathbf{s}'))} \tag{1}$$

where $U_\theta(s_0, \cdots, s_{T-1}) = \sum_t r(s_t) = \sum_t \theta \cdot \mathbf{f}_{s_t}$ and the sum is over are all possible trajectories of length T starting from state s_0. In this way, trajectories with higher total reward U are more likely to be sampled, but it is possible to observe sub-optimal behavior, specifically trajectories that do not provide the highest possible total reward. We can then recover the reward function by solving $\theta^*_{ME} = \arg\max_\theta \prod_{k=1}^K P_\theta^{ME}(\tau^k)$ to find the maximum likelihood estimate of the model parameters. The optimization problem is convex for deterministic MDPs, but not in general for stochastic MDPs (Ziebart et al. 2008).

[1] For simplicity, we report the above equation for a deterministic MDP and refer the reader to (Ziebart et al. 2008) for the general stochastic MDP case.

Dynamic Discrete Choice Modeling We will again assume that the decision makers do not always take optimal actions. This can be motivated by thinking about additional features (beyond the vector $\mathbf{f}_s$ we consider) that are taken into account by the agent but are not included in our model, hence giving rise to a behavior that is apparently not rational based on the data. When this effect is modeled as random noise affecting the decision process with an extreme value distribution, it gives rise to another stochastic model for the observed trajectories called logit Dynamic Discrete Choice (Rust 1987).

Specifically, in the dynamic discrete choice model it is assumed that at each step, the decision maker will not take the action with the largest future discounted utility, but instead will sample an action based on the following recursion:

$$V_\theta(s, a, T-1) = \theta \cdot \mathbf{f}_s, \quad \forall s \in S, \quad \forall a \in A \tag{2}$$

$$V_\theta(s, a, t) = \theta \cdot \mathbf{f}_s + \tag{3}$$

$$\eta \sum_{s' \in S} P(s'|s, a) \cdot \log \left(\sum_{a' \in A} \exp(V_\theta(s', a', t+1)) \right)$$

The probability of choosing action a in state s at time t is defined as

$$p_\theta^{DC}(s, a, t) = \frac{\exp\left(V_\theta(s, a, t)\right)}{\sum_{a'} \exp\left(V_\theta(s, a', t)\right)} \tag{4}$$

$$P_\theta^{DC}(\tau^k) = \prod_{t=0}^{T-1} p_\theta^{DC}(s_t^k, a_t^k, t)$$

The model is then fitted to the data $\mathcal{S}$ by setting θ as to maximize the likelihood of the observed transitions:

$$\theta^*_{DC} = \arg\max_\theta \prod_{k=1}^{K} \prod_{t=0}^{T-1} p_\theta^{DC}(s_t^k, a_t^k, t)$$

$$= \arg\max_\theta \log \left(\prod_{k=1}^{K} \prod_{t=0}^{T-1} p_\theta^{DC}(s_t^k, a_t^k, t) \right) \tag{5}$$

We will use the following notation for the log-likelihood function: $L_\theta^{DC} = \log \left(\prod_{k=1}^K \prod_{t=0}^{T-1} p_\theta^{DC}(s_t^k, a_t^k, t) \right) = \sum_{k=1}^K \sum_{t=0}^{T-1} \log p_\theta^{DC}(s_t^k, a_t^k, t)$ and $L_\theta^{DC}(s, a, t) = \log p_\theta^{DC}(s, a, t)$.

The objective function is optimized using gradient descent (Rust method, (Rust 1987)). The exact gradient can be computed by differentiating the likelihood expression (5) with respect to θ. The objective is generally not convex/concave (Rust 1987).

An Equivalence Relationship

Although on the surface the Maximum Entropy IRL model (1) and the Dynamic Discrete Choice model (4) appear to be very different, we prove the models are equivalent under some conditions:

Theorem 1. *For finite horizon deterministic MDPs, under the MaxEnt-IRL and logit DDC with $\eta = 1$, for any trajectory $\tau = (s_0, a_0), (s_1, a_1), \ldots, (s_{T-1}, a_{T-1})$ we have:*

$$P_\theta^{ME}(\tau) = P_\theta^{DC}(\tau).$$

Proof. See (Ermon et al. 2014). □

Since the log likelihood (1) for the Maximum Entropy IRL model is concave, we also have the following Corollary:

Corollary 1. *For deterministic MDPs and $\eta = 1$, the log likelihood (5) for logit DDC is concave.*

If we allow the discount factor η to be a free parameter, the class of DDC models are therefore strictly more general than MaxEnt-IRL models for *deterministic MDPs*. Note that using a discounted total reward (with $\eta < 1$) to score paths in the MaxEnt-IRL model (1) is not very meaningful, because the effect is that of putting more "weight" on the transitions occurring at the beginning of the trajectories. In the extreme case $\eta = 0$, only the first action taken matters with respect to scoring paths. On the other hand, DDC models are meaningful even for $\eta = 0$, and they simply become "static" discrete choice models where *at each step in the trajectory* agents are only considering the reward collected at the next time step.

Learning Discrete Choice Models at Scale

The standard method for learning Discrete Choice Models (solving the optimization problem (5)), is to use gradient descent as in (Rust 1987). This approach starts with a random initial θ, and iteratively updates θ following the gradient direction, until convergence. Intuitively, one has to iteratively solve a planning problem with the current estimate of the reward function (current θ), compare the results with the data (actual trajectories $\mathcal{S}$), and update the parameters as to make the predictions match the empirical observations. Convergence can be improved using (truncated) quasi-Newton techniques such as the BFGS algorithm (Liu and Nocedal 1989), which is considered one of the best algorithms for unconstrained optimization. However, since the objective (5) is generally not convex, the method might get trapped in local optima.

Unfortunately, this technique is also not very scalable. In fact, evaluating the likelihood L_θ^{DC} of the data (and computing its gradient with respect to θ), is required in every iteration of the procedure, and this requires the computation of $V_\theta(s, a, t)$ for every $\forall t \in \{0, \ldots, T-1\}$, $\forall s \in S$ and $\forall a \in A$. Following a Dynamic Programming approach, computing $V_\theta(s, a, t)$ from the end of the horizon towards the beginning, results in complexity $O(T(|A| + |S|^2|A|^2))$ *per iteration*. This approach can be very expensive as a subroutine even for moderately sized MDPs.

Simultaneous Planning and Learning

Fitting the model using the gradient is expensive for large datasets and complex MDPs because at every iteration we have to: 1) go through the entire dataset, and 2) fill a Dynamic Programming (DP) table containing the $V_\theta(s, a, t)$ values. The first problem is ubiquitous in large scale machine learning, and a very popular and successful solution is to use stochastic gradient methods (Bottou and Bousquet 2008; Duchi, Hazan, and Singer 2011; Roux, Schmidt, and Bach 2012). The key idea is to trade off computational cost and accuracy in the computation of the gradient, which is approximated looking only at a (randomly chosen) subset of the training data. Unfortunately, in our case we still need to fill the entire Dynamic Programming table even to compute the gradient for a small subset of the training data. To overcome both scalability issues at the same time, we introduce a new scalable learning algorithm, called SPL-GD (**S**imultaneous **P**lanning And **L**earning - **G**radient **D**escent). Our technique uses approximate gradient estimates which can be efficiently computed exploiting the dynamic structure of the problem. We report the pseudocode of SPL-GD as Algorithm 1.

The key observation is that the log-likelihood from (5) can be decomposed according to time as $L_\theta^{DC} = \sum_{t=0}^{T-1} L_\theta^{DC}(t)$, where $L_\theta^{DC}(t) = \sum_{k=1}^{K} L_\theta^{DC}(s_t^k, a_t^k, t)$ represents the contribution from all the transitions from time t. Further notice from (4) and (3) that $L_\theta^{DC}(t_0)$ and its gradient $\nabla L_\theta^{DC}(t_0)$ depend only on $V_\theta(s, a, t')$ for $t' \geq t_0$.

Rather than updating θ using the full gradient $\nabla L_\theta^{DC} = \sum_{t=0}^{T-1} \nabla L_\theta^{DC}(t)$ (which requires the computation of the entire DP table), in SPL-GD we simultaneously update the current parameter estimate θ (Learning) *while we iterate* over time steps t to fill columns of the DP table (Planning). Specifically, while we iterate over time steps t from the end of the time horizon, we use an approximation of $\nabla L_\theta^{DC}(t)$ to update the current parameter estimate θ. After each update, we continue filling the next column of the DP table *using the new estimate* of θ rather than discarding the DP table and restarting.This introduces error in the estimates of $\nabla L_\theta^{DC}(t)$ because we are slowly annealing θ through the recursive calculation. However, we observe that with a small learning rate λ_j, the gradient estimates are sufficiently accurate for convergence. Notice that for $\eta = 0$ (if the discount factor is zero, the MDP is static) the bias disappears and SPL-GD corresponds to fitting $K \cdot T$ logistic models using a variant of mini-batch stochastic gradient descent (with a fixed ordering) where training data is divided into mini batches according to the time stamps t.

Empirical results In Figure 1 we report a runtime comparison between Algorithm 1 and the state-of-the-art batch BFGS (with analytic gradient, and approximate Hessian). The comparison is done using a small subset of our Borena plateau dataset (one month of data, $T = 30$) and the MDP model described in detail in the next section. We use a learning rate schedule $\lambda_j = \frac{1}{\sqrt{j}}$. We see that our algorithm is about 20 times faster than BFGS, even though BFGS is using approximate second-order information on the objective function. The advantage is even more significant on datasets covering longer time periods, where more gradient estimate updates occur per iteration.

Algorithm 1 SPL-GD ($\mathcal{S} = \{\tau^1, \cdots, \tau^K\}, \{\lambda_j\}$)

Initialize θ at random
for $j = 0, \cdots, M$ **do**
  for $t = T-1, \cdots, 0$ **do**
    if t = T-1 **then**
      $V(s, a, T-1) = \theta \cdot \mathbf{f}_s, \quad \forall s \in S, \quad \forall a \in A$
      $\nabla V(s, a, T-1) = \mathbf{f}_s, \quad \forall s \in S, \quad \forall a \in A$
    else
      $V(s, a, t) = \theta \cdot \mathbf{f}_s + \eta \sum_{s' \in S} \left(\log\left(\sum_{a' \in A} \exp(V(s', a', t+1))\right)\right) P(s'|s, a), \quad \forall s \in S, \quad \forall a \in A$
      $\nabla V(s, a, t) = \mathbf{f}_s + \eta \sum_{s' \in S} \left(\frac{\sum_{a' \in A} \exp(V(s', a', t+1)) \nabla V(s', a', t+1)}{\sum_{a' \in A} \exp(V(s', a', t+1))}\right) P(s'|s, a), \quad \forall s \in S, \quad \forall a \in A$
    end if
    for $k = 1, \cdots, K$ **do**
      $\nabla L^{DC}(s_k^t, a_k^t, t) = \nabla V(s_k^t, a_k^t, t) - \frac{\sum_{a'} \exp\left(V(s_k^t, a', t)\right) \nabla V(s_k^t, a', t)}{\sum_{a'} \exp\left(V(s_k^t, a', t)\right)}$
      $\theta \leftarrow \theta + \lambda_j \nabla L^{DC}(s_k^t, a_k^t, t)$
    end for
  end for
end for
return θ

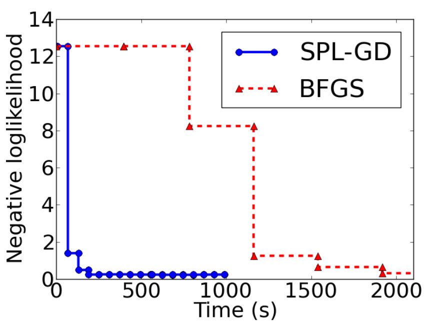

Figure 1: Runtime comparison between SPL-GD and BFGS. SPL-GD converges much faster than BFGS.

Modeling Pastoral Movements in Ethiopia

Our work is motivated by the study of spatio-temporal resource preferences of pastoralists and their cattle herds in the arid and semi-arid regions of Africa. Our overall goal is to develop a model for the planning decisions made by the herders, which is the focus of this paper, as well as the individual movements and consumption patterns of the cattle. This model must be *structural*, meaning that its parameters provide intuitive insight into the decision-making process, as well as *generative*, meaning that it can potentially be used to simulate behavior under new circumstances such as changes in resource availability, access policies, or climate.

Available Data: The available data includes survey data from individual households in the Borena plateau, Ethiopia; static geospatial map layers including village and road locations, ecosystem types, elevation and other terrain features; a dynamic greenness index (NDVI) at 250m $\times$ 250m (NASA LP DAAC 2014); locations of wells, ponds, and other water points identified by interview, field exploration, and satellite imagery; and GPS collar traces of 60 cattle from 20 households in 5 villages, at 5-minute intervals over sub-periods spread over 3 years. The GPS traces are our primary source of information regarding behavior and resource use.

State Space: The first modeling choice is the time scale of interest. Behaviorally, cattle could change movement patterns over minutes, while herding plans are likely to be made on a daily basis, though these might require multiple segments due to travel, sleep, etc. At the top level, the pastoralists migrate to remote camps as required to maintain access to nearby resources, as conditions change seasonally. While the end goal is a coupled model that incorporates these three scales (minutes, days, seasons), we have begun by focusing primarily on the migration decisions, which we represent as a decision whether to move to an alternate camp location each day. We extracted a list of observed camp locations by clustering the average GPS locations of the herds during the nighttime hours, across the entire time horizon for which we have collected data. There were nearly 200 camps that exhibited migration. We denote $C = \{c_1, c_2, \ldots, c_m\}$ the set of identified camping sites.

Features: We model the suitability of each campsite $c_i \in C$ as a function of a number of time-dependent features, which are generally selected data items listed above in their raw form and meaningful functions of those data. The features we considered are: distance from home village (a closer campsite might be more desirable than one far away), distance from major road, 8 variants of distance from closest water-point (based on different estimates of the seasonal availability of different classes of water points), and 2 representations of the greenness/vegetation index each intended to capture different spatio-temporal characteristics (normalized spatially over the Borena plateau region and temporally over 13 years of data).

MDP Modeling: We model each household as a self-interested agent who is rationally taking decisions as to optimize an (unknown) utility function over time. Intuitively, this utility function represents the net income from the economic activity undertaken, including intangibles. In our model, each household is assumed to plan on a daily basis the next campsite to use, so as to optimize their utility

Method	Fold 1 LogLik.	Fold 1 Moves (191)	Fold 2 LogLik.	Fold 2 Moves (85)	Fold 3 LogLik.	Fold 3 Moves (78)	Fold 4 LogLik.	Fold 4 Moves (116)
Markov	-8864.5	2209.8	-1807.8	372.2	-7265.7	1756.0	-4570.2	1214.1
MaxEnt IRL	-1524.4	424.6	-787.7	293.8	-796.7	339.7	-1004.2	299.4
Discrete Choice	**-1422.1**	**102.5**	**-657.8**	**104.9**	**-643.4**	**115.9**	**-911.3**	**94.9**

Table 1: Data log-likelihood and predicted total number of movements along all trajectories, evaluated in cross-validation on the held-out data, for each model. The actual number of camp movements is given in parentheses.

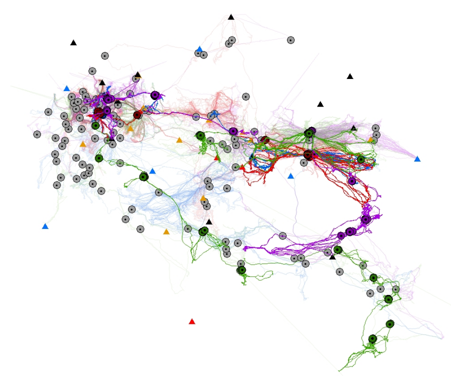

Figure 2: Trajectories (color denotes household), camps (circles), and waterpoints (triangles) for one village. Heavier trajectory lines illustrate movements in a one-month period during the wet season; faded lines denote movements during other times. Labels and other details omitted for privacy.

function looking ahead over the entire time horizon T. Formally, we model the problem as a Markov Decision Process as follows. Let $D = \{0, \cdots, T-1\}$. The action set is $A = C$, where each action corresponds to the next campsite to visit. We use an augmented state space $S = C \times C \times D$, where visiting a state $s = (c, c', t)$ means moving from camp c to c' at time t. This allows us to model the variability of the features over time and to incorporate information such as the distance between two campsites as state-based features. The MDP is deterministic, with transition probabilities $P((c_1', c_2', t)|(c_1, c_2, k), a) = 1$ iff $t = k + 1$, $c_1' = c_2$, $c_2' = a$, and 0 otherwise. This means that if the agent transitioned from c_1 to c_2 at time k, and then takes action $a = c \in C$, it will transition from c_2 to c at time step $k + 1$. We furthermore assume a utility function that is *linearly dependent* through θ in the features available to our model, and possibly additional information not available to the model. At present, we do not model competition or interactions between different households.

We then extract sequences of camping locations from the GPS collar data. These can be interpreted as K finite sequences of state-action pairs $\mathcal{S} = \{\tau^1, \cdots, \tau^K\}$ in our MDP model. A static illustration of the movements, camps and water points is shown in Figure 2. Our goal is to infer θ, i.e. to understand which factors drive the decision-making and what are the spatio-temporal preferences of the herders.

Results

We consider the dynamic discrete choice model, the maximum entropy IRL model and, as a baseline, a simple Markovian model that ignores the geo-spatial nature of the problem. For the Markov model, the assumption is that pastoralists at camp c_i will transition to camp c_j with probability p_{ij}. Equivalently, trajectories $\mathcal{S}$ are samples from a Markov Chain over C with transition probabilities p_{ij}, where the maximum likelihood estimate of the transition probabilities $\widehat{p}_{ij}$ is given by the empirical transition frequencies in the data, with Laplace smoothing for unobserved transitions.

We fit and evaluate the models in 4-fold cross-validation, in order to keep data from each household together, and stratify the folds by village. Training using SPL-GD on the entire 3-year dataset takes about 5 hours (depending on the initial condition and value of η), as opposed to several days using BFGS. We choose η in cross-validation with a grid search, selecting the value with the best likelihood on the training set.

We report results in Table 1. In addition to evaluating the likelihood of the trained model on the held-out test set, we also report the predicted number of transitions; although none of the models are explicitly trying to fit for this, it gives a sense of the accuracy and was used for example in (Kennan and Walker 2011).

The simple Markov model dramatically overfits, failing to generalize to unobserved camp transitions, and performs extremely poorly on the test set. The other models based on an underlying MDP formulation perform much better. We see that Dynamic Discrete Choice outperforms MaxEnt IRL: allowing the extra flexibility of choosing the discount factor does not lead to overfitting, and leads to improvements on all the folds. These results suggest that the features we consider are informative, and that considering discount factors other than $\eta = 1$ (as in the MaxEnt IRL model) is important to capture temporal discounting in the herders' decisions. The trained model recovers facts that are consistent with our intuition, e.g. herders prefer short travel distances, and allows us to quantitatively estimate these (relative) resource preferences. This provides exciting opportunities for simulation analysis by varying the exogenous characteristics of the system.

Conclusions

Motivated by the study of migratory pastoralism in the Borena plateau (Ethiopia) we study the general problem of inferring spatio-temporal resource preferences of agents from data. This is a very important problem in computational sustainability, as micro-behavioral models that capture the choice process of the agents in the system are crucial for policy-making concerning sustainable development.

We presented the Dynamic Discrete Choice model and showed a connection with Maximum Entropy IRL, a well known model from the machine learning community. To overcome some of the limitations of existing techniques to learn Discrete Choice models, we introduced SPL-GD, a novel learning algorithm that combines dynamic programming with stochastic gradient descent. Thanks to the improved scalability, we were able to train a model on a large dataset of GPS traces, surveys, and satellite information and other geospatial data for the Borena plateau area. The model obtained is generative and predictive, and outperforms competing approaches. As a next step, we plan to start using the model for for policy-relevant simulation analyses, as well as to couple it with optimization frameworks to allocate limited resources under budgetary constraints.

Acknowledgments

We gratefully acknowledge funding support from NSF Expeditions in Computing grant on Computational Sustainability (Award Number 0832782), the Computing research infrastructure for constraint optimization, machine learning, and dynamical models for Computational Sustainability grant (Award Number 1059284), Department of Foreign Affairs and Trade of Australia grant 2012 ADRAS 66138, and David R. Atkinson Center for a Sustainable Future grant #2011-RRF-sdd4.

References

Aguirregabiria, V., and Mira, P. 2010. Dynamic discrete choice structural models: A survey. *Journal of Econometrics* 156(1):38–67.

Ben-Akiva, M. E., and Lerman, S. R. 1985. *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press.

Bertsekas, D. 1995. *Dynamic programming and optimal control*, volume 1. Athena Scientific Belmont, MA.

Bottou, L., and Bousquet, O. 2008. The tradeoffs of large scale learning. In *Advances in Neural Information Processing Systems*, volume 20, 161–168.

Duchi, J.; Hazan, E.; and Singer, Y. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research* 12:2121–2159.

Ermon, S.; Xue, Y.; Toth, R.; Dilkina, B.; Bernstein, R.; Damoulas, T.; Clark, P.; DeGloria, S.; Mude, A.; Barrett, C.; and Gomes, C. 2014. Learning large-scale dynamic discrete choice models of spatio-temporal preferences with application to migratory pastoralism in East Africa. Technical report, Department of Computer Science, Cornell University.

Kennan, J., and Walker, J. R. 2011. The effect of expected income on individual migration decisions. *Econometrica* 79(1):211–251.

Kolter, J. Z., and Ng, A. Y. 2009. Regularization and feature selection in least-squares temporal difference learning. In *Proceedings of the 26th annual international conference on machine learning*, 521–528. ACM.

Liu, D. C., and Nocedal, J. 1989. On the limited memory bfgs method for large scale optimization. *Mathematical programming* 45(1-3):503–528.

NASA LP DAAC. 2014. MOD13Q1. Vegetation Indices 16-Day L3 Global 250m.

Ng, A. Y., and Russell, S. J. 2000. Algorithms for inverse reinforcement learning. In *ICML*, 663–670.

Powell, W. B. 2007. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons.

Puterman, M. L. 2009. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Ratliff, N. D.; Bagnell, J. A.; and Zinkevich, M. A. 2006. Maximum margin planning. In *Proceedings of the 23rd international conference on Machine learning*, 729–736. ACM.

Roux, N. L.; Schmidt, M.; and Bach, F. R. 2012. A stochastic gradient method with an exponential convergence _rate for finite training sets. In *Advances in Neural Information Processing Systems*, 2663–2671.

Rust, J. 1987. Optimal replacement of GMC bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society* 999–1033.

Taylor, M. E., and Stone, P. 2009. Transfer learning for reinforcement learning domains: A survey. *The Journal of Machine Learning Research* 10:1633–1685.

Ziebart, B. D.; Maas, A. L.; Bagnell, J. A.; and Dey, A. K. 2008. Maximum entropy inverse reinforcement learning. In *AAAI*, 1433–1438.