Multi-Entity Dependence Learning with Rich Context via Conditional Variational Auto-encoder

Luming Tang[2*], Yexiang Xue[1*], Di Chen[1], Carla P. Gomes[1]
[1] Cornell University; [2] Tsinghua University; * indicates equal contribution

Introduction

Multi-Entity Dependence Learning (MEDL) explores conditional correlations among multiple entities. We propose MEDL_ CVAE, which encodes a conditional multivariate distribution as a generating process. The variational lower bound of the joint likelihood can be optimized via a conditional variational auto-encoder and trained end-to-end on GPUs.

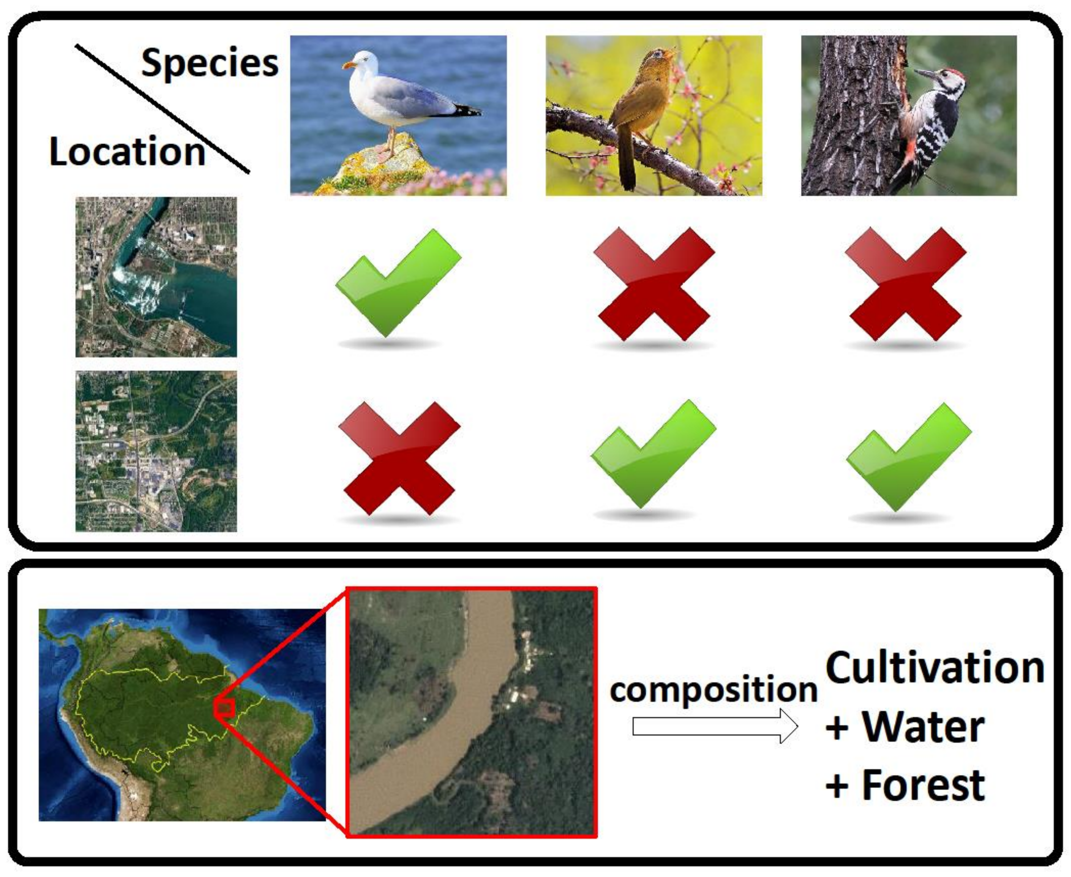


Figure 1: Two computational sustainability related applications for MEDL CVAE. 1) is to model species interactions using the *eBird* data. 2) is a landscape categorization using *Amazon* satellite images.

Method

Preliminaries

We consider modeling the dependencies among multiple entities on problems with rich contextual information. Our dataset consists of tuples:

$$D = \{(x_i, y_i) \mid i = 1, \dots, N\}$$

$x_i = (x_{i,1}, \dots, x_{i,k}) \in R^k$ is a high-dimensional contextual feature vector;
$y_i = (y_{i,1}, \dots, y_{i,l}) \in \{0,1\}^l$ is a sequence of l indicator variables, $y_{i,j}$ represents whether the j-th entity is observed in an environment characterized by x_i.
The problem is to learn a conditional joint distribution Pr(y|x) which maximizes the conditional joint log likelihood over N data points:

$$\sum_{i=1}^{N} \log \Pr(y_i \mid x_i)$$

The probability of each entity's existence are not mutually independent:

$$\Pr(y_i \mid x_i) \neq \prod_{j=1}^{l} \Pr(y_{i,j} \mid x_i)$$

Our Approach

We propose MEDL_CVAE to address following two challenges:

- Noisy and potentially multi-modal responses.
- Incorporation of rich contextual information such as satellite images.

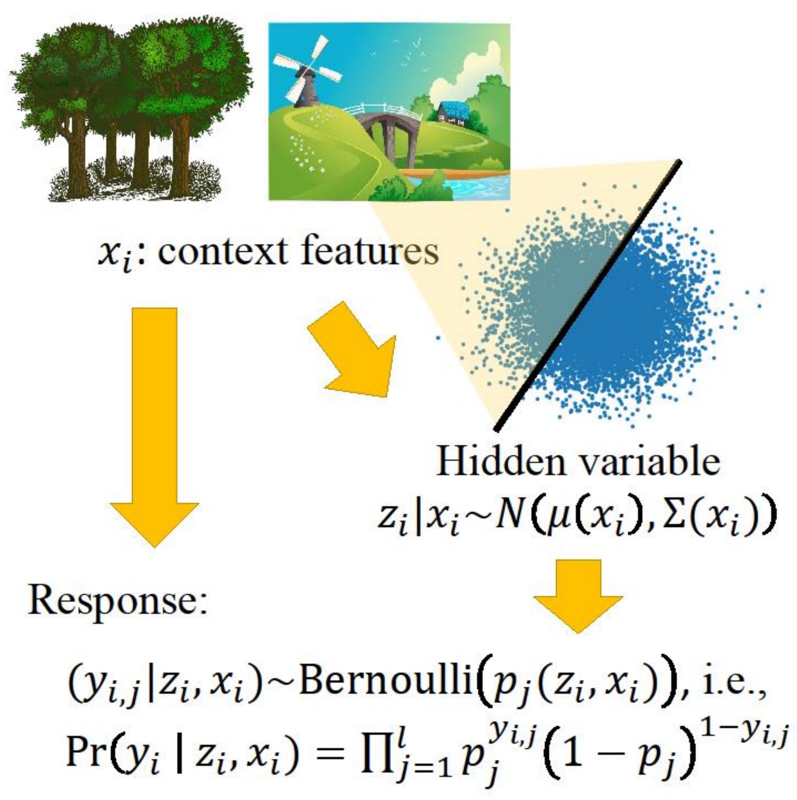


Figure2. Our proposed conditional generating process.
Given contextual features x_i such as satellite images, we use hidden variables z_i conditionally generated based on x_i to capture noisy and multi-modal response. The response y_i depends on both contextual information xi and hidden variables z_i.

The original optimization problem:

$$\max_{\theta_d} \sum_i \log Pr(y_i|x_i) = \sum_i \log \int Pr(y_i|x_i, z_i) Pr(z_i|x_i) \mathrm{d}z_i$$

It is intractable because of a hard integral inside the logarithmic function.
Based on the following variational equality:

$$\log Pr(y_i|x_i) - D\left[Q(z_i|x_i, y_i) || Pr(z_i|x_i, y_i)\right] \quad (6)$$
$$= \mathrm{E}_{z_i \sim Q(z_i|x_i,y_i)} \left[\log Pr(y_i|z_i, x_i)\right] - D\left[Q(z_i|x_i, y_i) || Pr(z_i|x_i)\right]$$

We turn to maximizing the variational lower bound of original log likelihood:

$$Q(z_i|x_i, y_i) = N(\mu_e(x_i, y_i), \Sigma_e(x_i, y_i)).$$

$$\max_{\theta_d, \theta_e} \sum_i \mathrm{E}_{z_i \sim Q(z_i|x_i,y_i)} \left[\log Pr(y_i|z_i, x_i)\right] - D\left[Q(z_i|x_i, y_i) || Pr(z_i|x_i)\right].$$

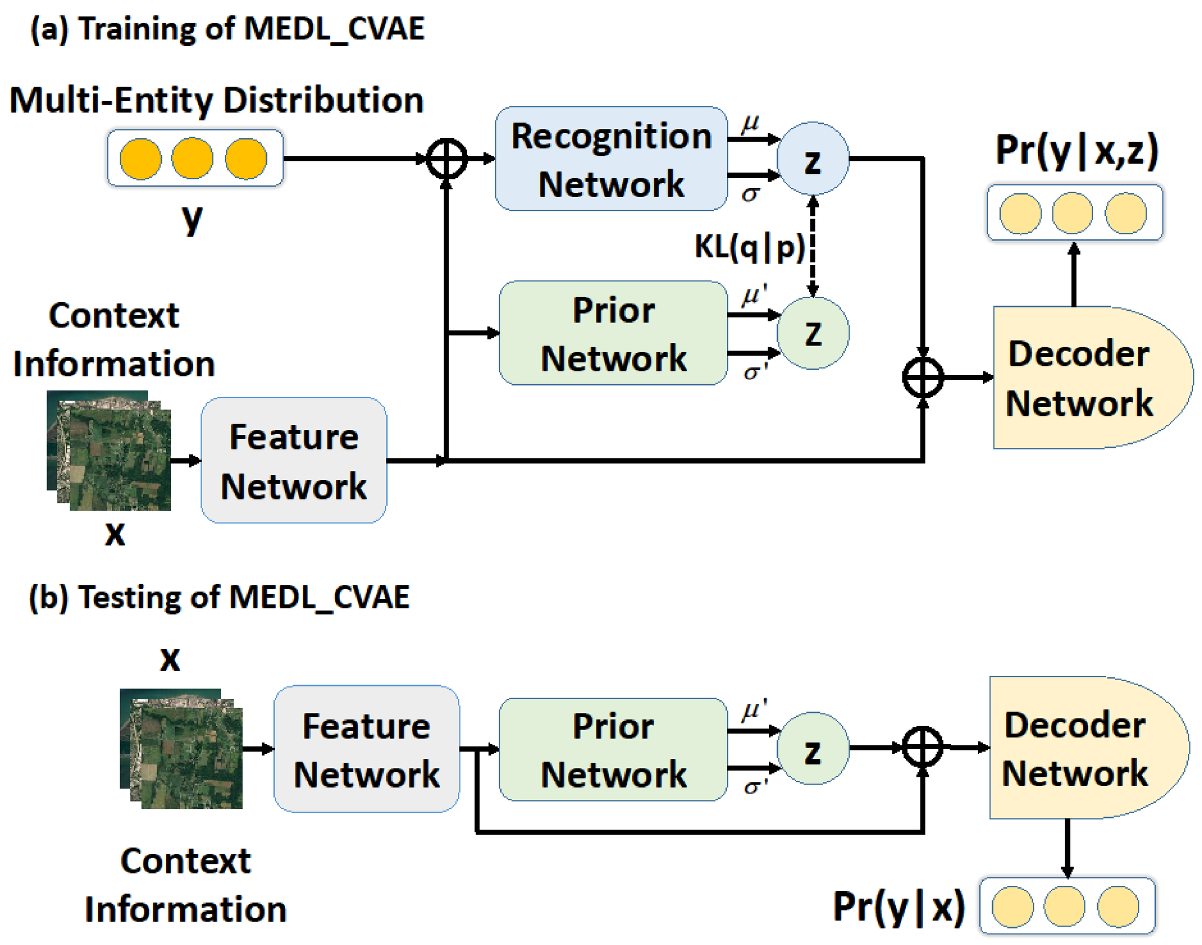


Figure 3. Overview of the neural network of MEDL_CVAE for training and testing.

Experiment

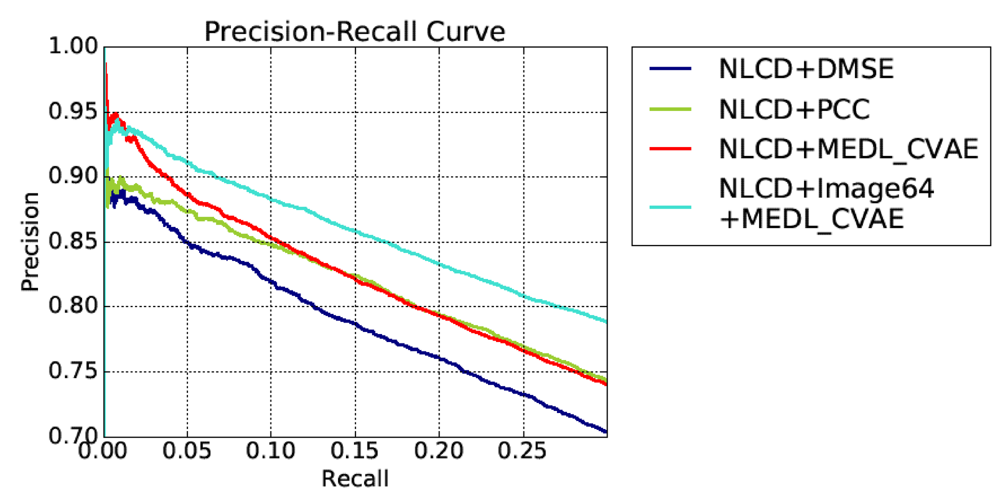
Dataset	Training Set Size	Test Set Size	# Entities
eBird	45855	5094	100
Amazon	30383	4048	17

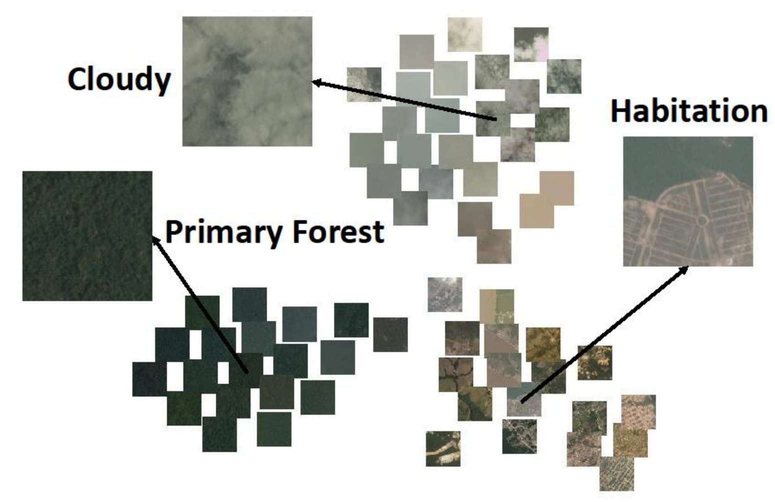
Table 1. the statics of the *eBird* and the *Amazon* dataset.

Method	Neg. JLL	Time (min)
NLCD+MLP	36.32	2
Image256+ResNet50	34.16	5.3 hrs
NLCD+Image256+ResNet50	34.48	5.7 hrs
NLCD+Hist64+MLP	34.97	3
NLCD+Hist128+MLP	34.62	4
NLCD+Image64+MLP	33.73	9
NLCD+MLP+PCC	35.99	21
NLCD+Hist128+MLP+PCC	35.07	33
NLCD+Image64+MLP+PCC	34.48	53
NLCD+DMSE	30.53	20 hrs
NLCD+MLP+MEDL_CVAE	30.86	9
NLCD+Hist64+MLP+MEDL_CVAE	28.86	20
NLCD+Hist128+MLP+MEDL_CVAE	28.71	22
NLCD+Image64+MLP+MEDL_CVAE	**28.24**	48

Method	Neg. JLL
Image64+MLP	2.83
Hist128+MLP	2.44
Image64+CNN	2.16
Image64+MLP+PCC	2.95
Hist128+MLP+PCC	2.60
Image64+CNN+PCC	2.45
Image64+MLP+MEDL_CVAE	2.37
Hist128+MLP+MEDL_CVAE	2.09
Image64+CNN+MEDL_CVAE	**2.03**

Table 2. Negative joint log-likelihood and training time of models assuming independence (first section), previous multi-entity dependence models (second section) and our MEDL_CVAE on the *eBird* (left) and *Amazon* (right) test set.





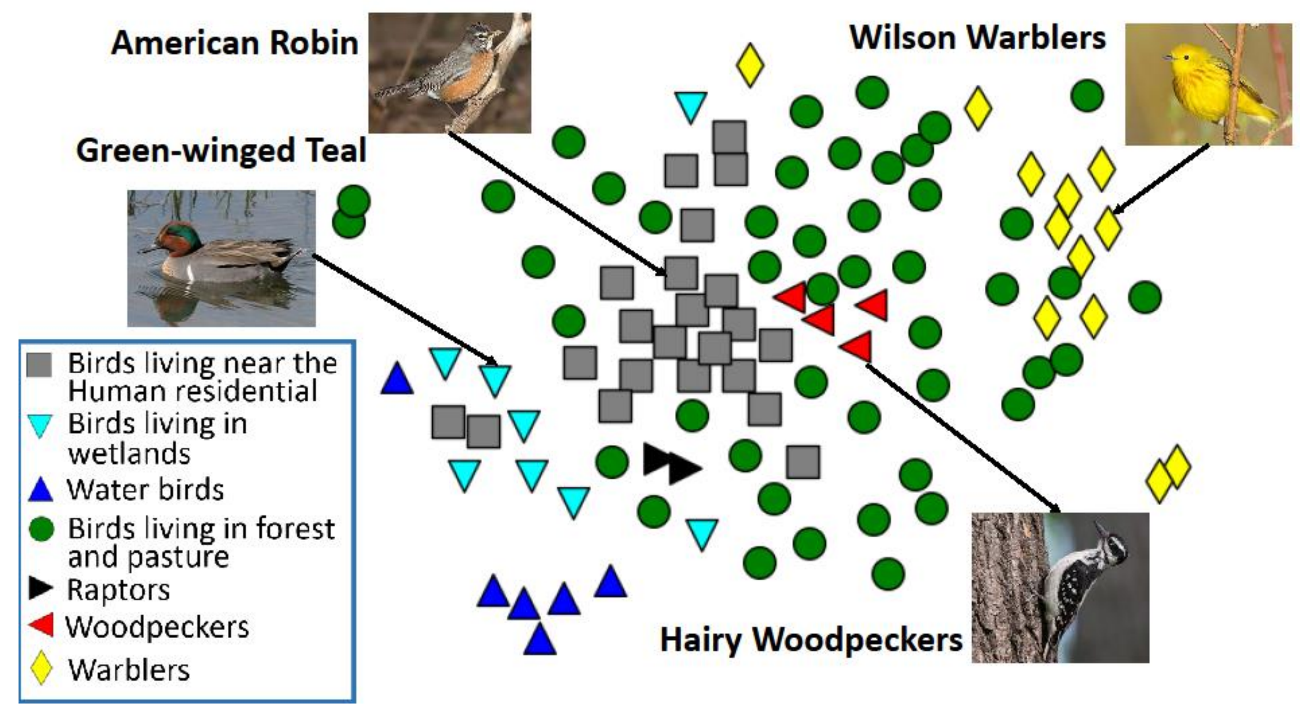


Figure 4 (top left). Performance of baseline models and MEDL_CVAE on the *Amazon* dataset.
Figure 5 (top right). Visualization of the posterior z ~ Q(z|x,y) on the *Amazon* dataset.
Figure 6 (bottom). Visualization of vectors inside decoder network's last fully connected layer.

Acknowledgement: This research was supported by NSF 0832782, 1522054, 1059284, 1356308, ARO grant W911-NF-14-1-0498, the Leon Levy Foundation and the Wolf Creek Foundation.