Learning Contextualized Player Representations with A Variational Hierarchical Encoder
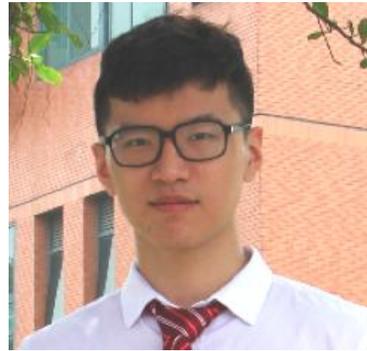

Guiliang Liu

Oliver Schulte

Compute the Estimated Goal (XG): (Scoring Opportunity):

- Perform the same action under the same game environment (all the spatial and temporal info are known).

- The XG, however, should be different between an all-start player v.s. a random player.

- To model the difference between players, we should add the player information to estimation.

Game Environment

Action: Shot
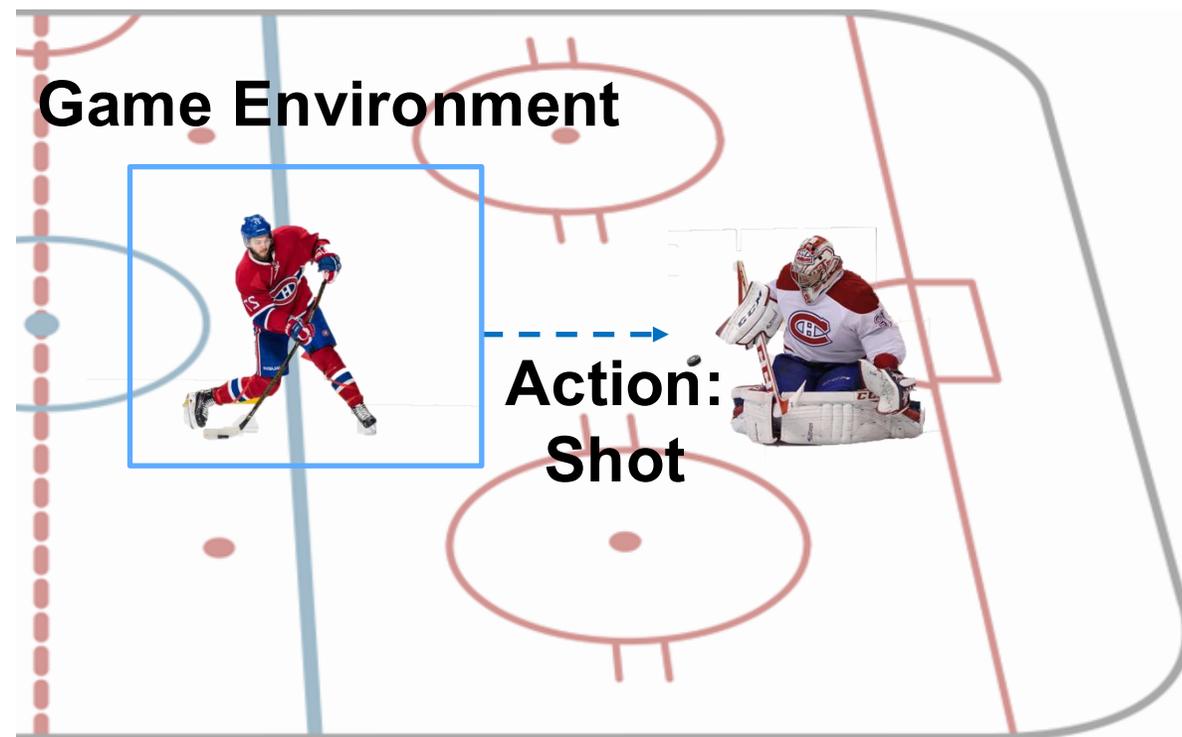
Add the player information into Modeling

- Under the same environment, the XG should be different between an all-start player v.s. a random player.

- We should add the player information to estimation.

Apply one-hot Player Id (Pid).

- Pros:
  1. Easy to add (insert to input).
  2. Pids are commonly available.

- Cons
  1. Not informative enough.
  2. Can't correlate individuals to their success.

Compute Player Embedding

- Pros:
  1. Contain abundant information.
  2. Applicable across different domains. (commonly investigated under NLP)

- Cons
  1. Sensitive to the design of embedding model.
  2. A related work trained a neural network encoder to perform a secondary prediction task[1]: given the current game context, predict the pids of all on-court players.

[1] Ganguly, Sujoy, and Nathan Frank. "The problem with win probability." Proc. of the 12th MIT Sloan Sports Analytics Conf. 2018.

Contextual Variables for NHL Players:

- The action a_t records the action of the on-puck player.

- The environment variables x_t describe the environment. We represent it as a feature vector specifying a value of the features in the following table:

| Type	Name	Range
Spatial Features	X Coordinate of Puck	[-100, 100]
	Y Coordinate of Puck	[-42.5, 42.5]
	Velocity of Puck	$(-\infty, +\infty)$
	Angle between the puck and the goal	$[-3.14, 3.14]$
Temporal Features	Game Time Left	[0, 3,600]
	Event Duration	$(0, +\infty)$
In-Game Features	Score Differential	$(-\infty, +\infty)$
	Manpower Situation	{ES, SH, PP}
	Home or Away Team	{Home, Away}
	Action Outcome	{successful, failure}

Contextual Variables for NHL Players:

- To alleviate the partial observability in the dataset, a game state s_t is applied to include the game history: $s_t = x_t, pl_{t-1}, a_{t-1}, x_{t-1}, pl_{t-1}, a_{t-2} \ldots \ldots x_0$ (pl_t records the current acting player).

- Our player representation $q\big(z_{pl,t}\big|pl_t, s_t, a_t\big)$ conditions on game environment.
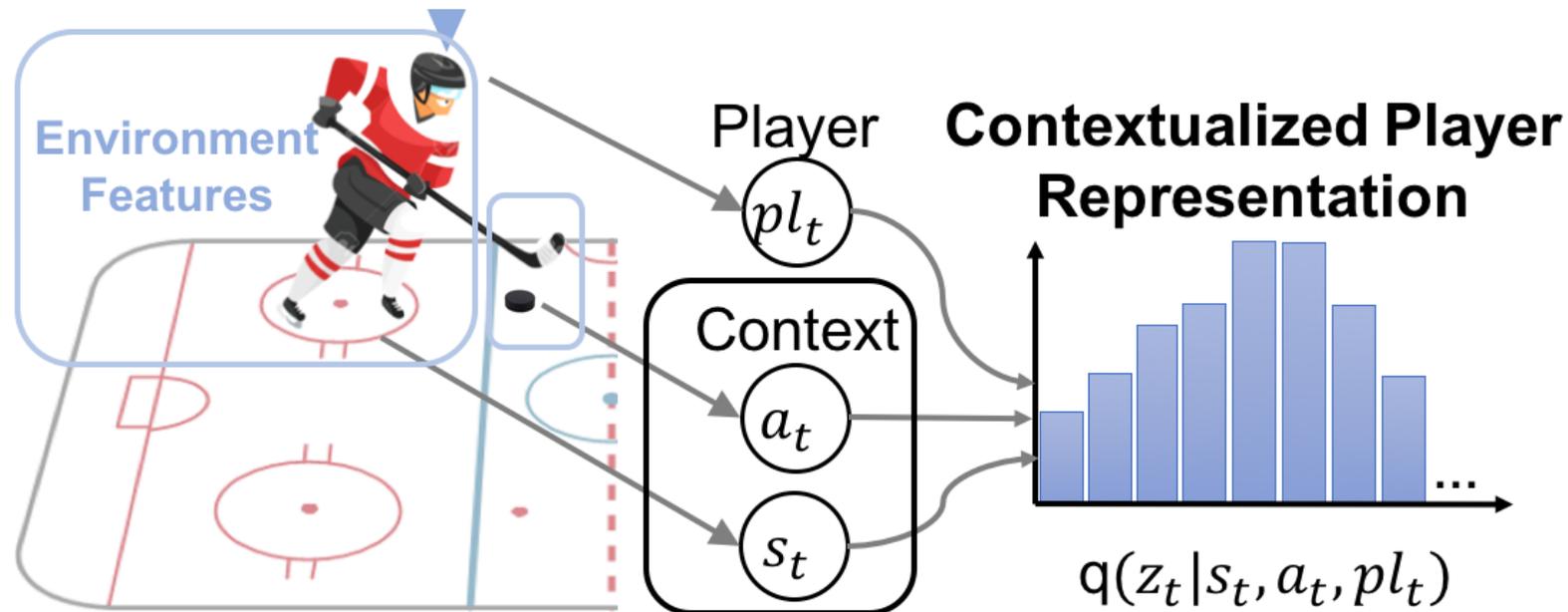
Contextualized Player Embedding

Motivation:

- The behavior of professional players, is highly sensitive to context.

- It is difficult to learn a fixed representation that can adequately describe a player's tendencies under every game context.

- Our Variational Hierarchical Encoder with Recurrence (VHER) model
  1. Learn a representation for game context by an neural hierarchical model.
  2. asymptotically adjusts the posterior representation for individual players to each context with Variational inference.
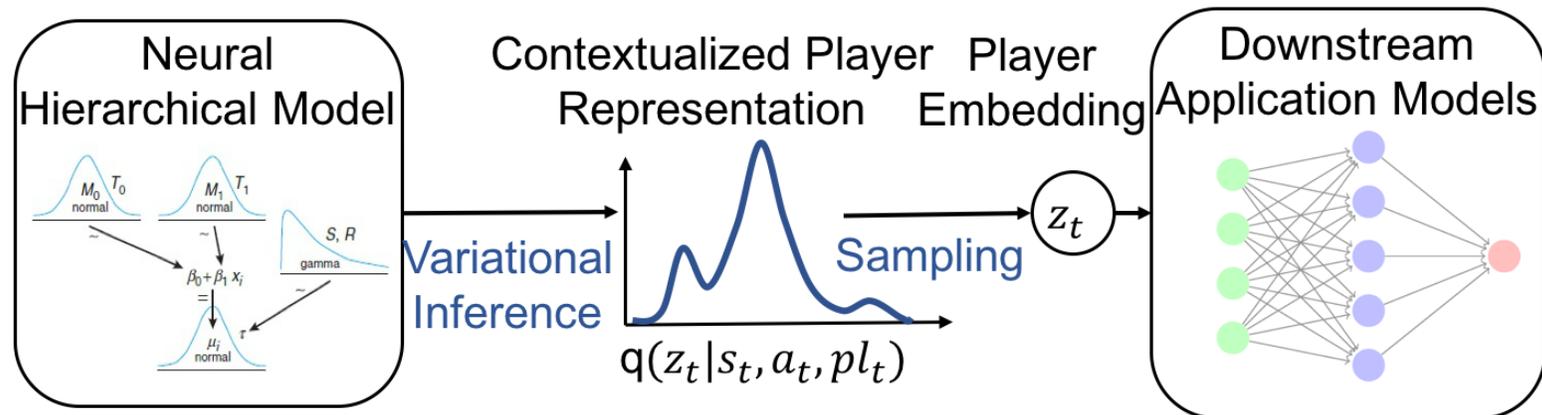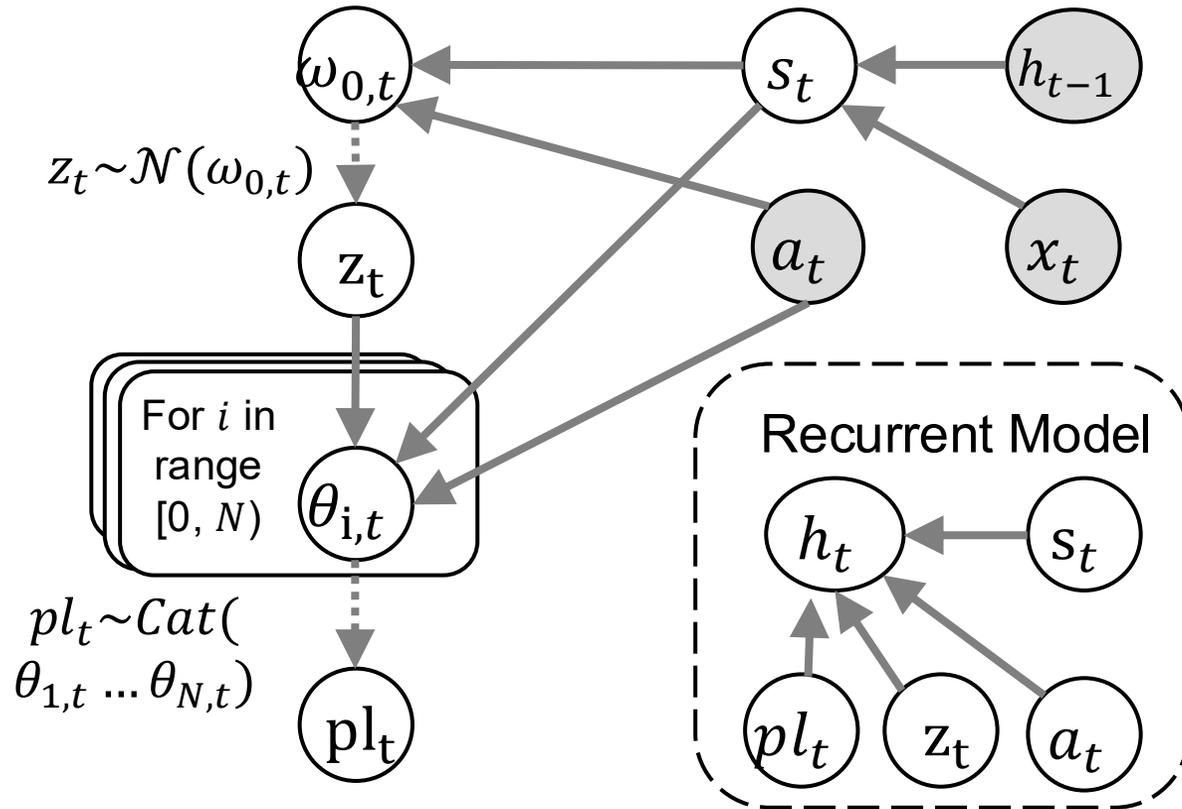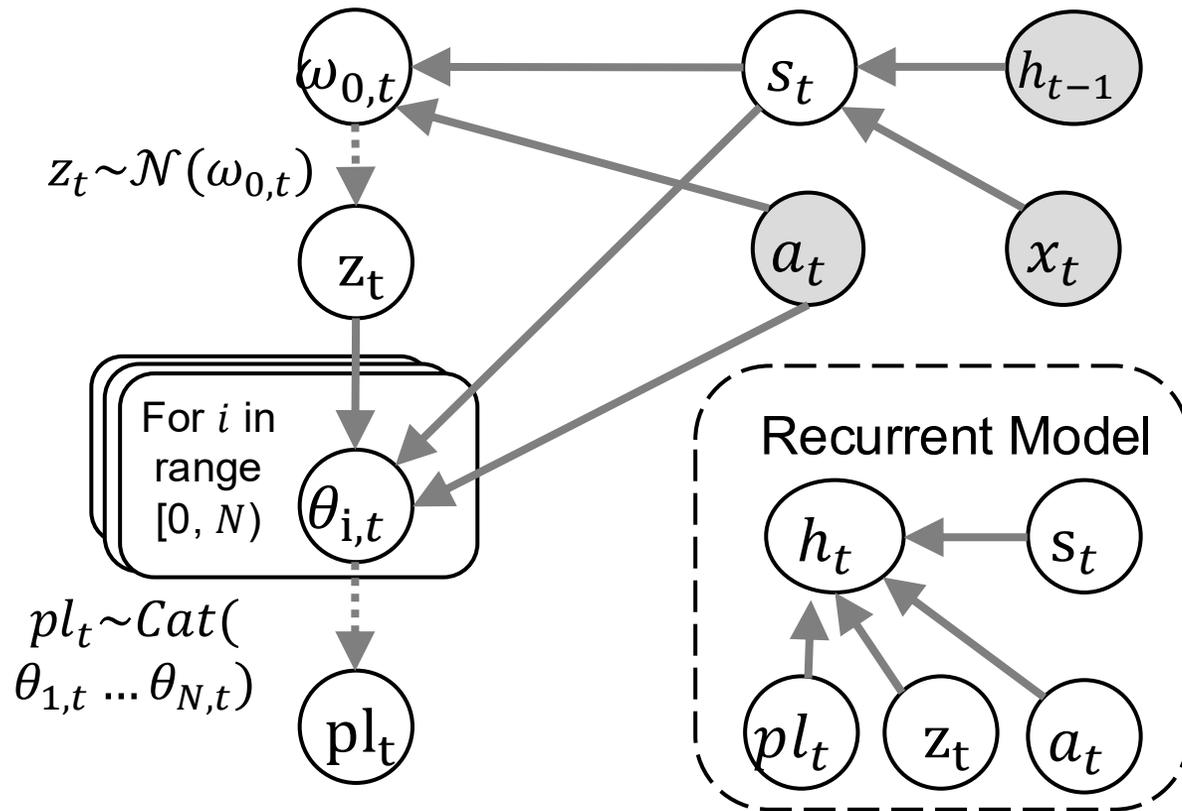
The neural Hierarchical Model:

- Thick (bold) line indicates logical function while dashed line denotes stochastic dependence. The shaded (observation) nodes are given during generation.

- $\omega_{0,t} \equiv (\mu_{0,t}, \sigma_{0,t})$ defines the parameters of Gaussian distribution. $\omega_{0,t}$ is conditioned on s_t, a_t.

- $s_t \equiv (a_t, x_t, h_{t-1})$ where x_t is the current observation and h_{t-1} is the hidden states from a recurrent model. $h_t = f(x_t, a_t, z_t, h_{t-1})$.

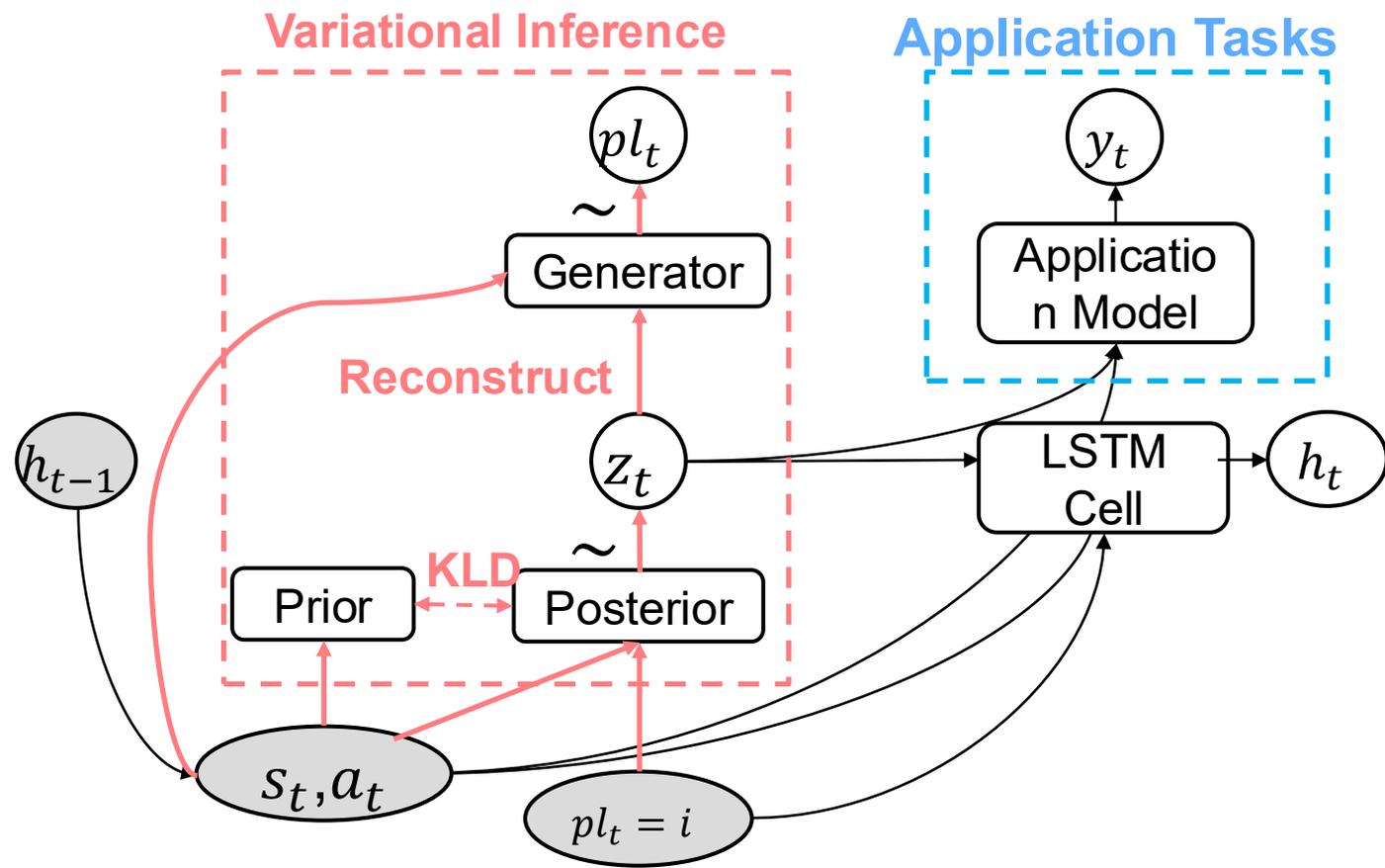**The neural Hierarchical Model:**

- Latent variable $z_{i,t}$ represents the embedding of the ith player.

- $\theta_{i,t}$ defines the parameter of ith Bernoulli.

- We aggregate the Bernoulli parameters and normalize them to be the parameters of a categorical where we sample player pl_t.

Variational Inference

Evidence Lower Bound(ELBo)

$$KLD\big(q(z_{pl,t}|pl_t, s_t, a_t)||p(z_t|s_t, a_t)\big) - \mathrm{E}_{z_{pl,t}}[\log p(pl_t|z_{pl,t}, s_t, a_t)]$$

- The KL-divergence term shrinks each player representation towards a local shrinkage prior conditioning on the game context.

- The reconstruction loss encourages the sparsity of player embedding by projecting the embeddings to the id of current on-the-ball player pl_t.

Variational Inference
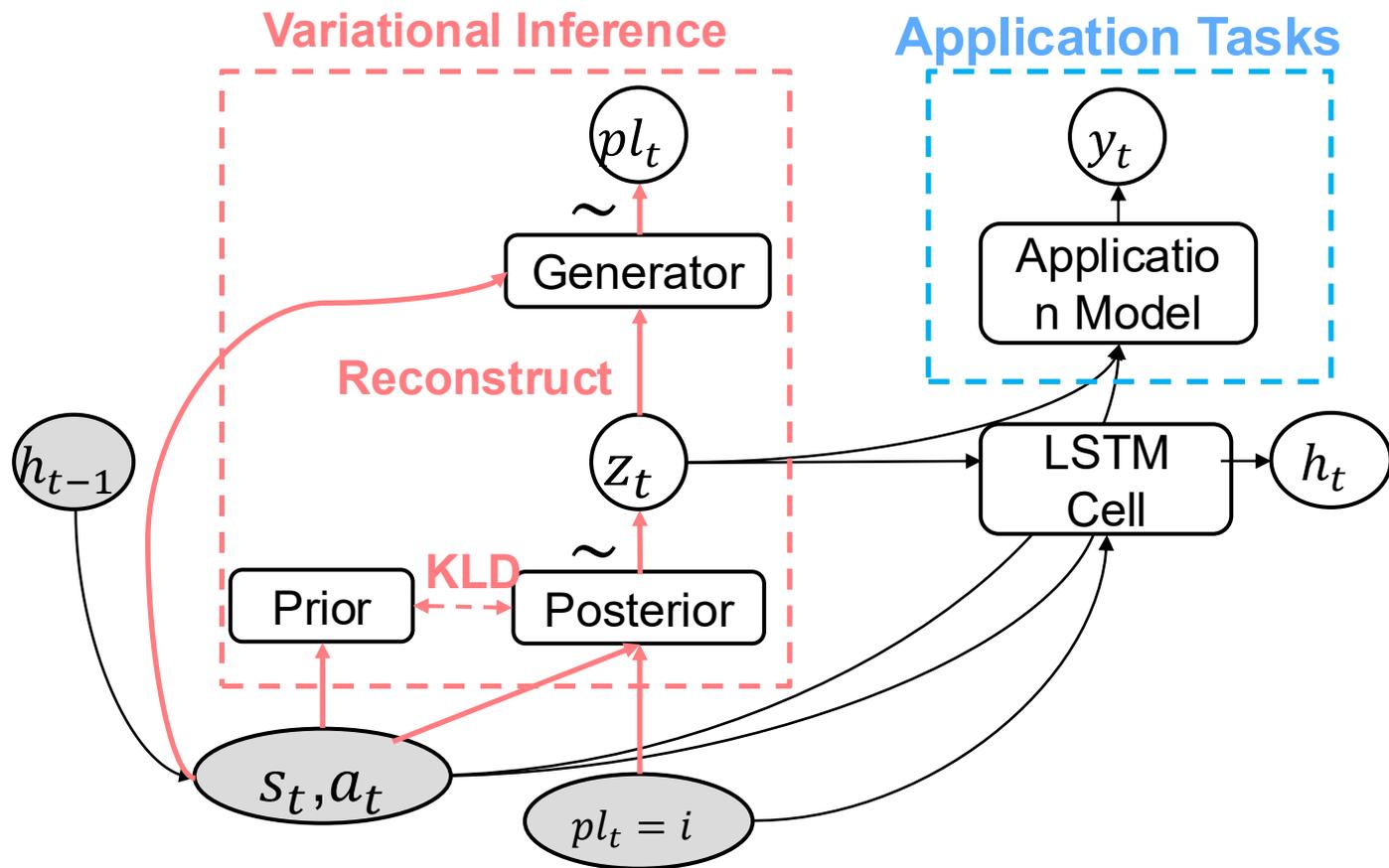
Shrinkage Effect:

- Shrinkage estimators allow information to be transferred between the observations of different individuals.

- The shrinkage effect becomes stronger for players who share many statistical similarities under a game context.

- This formalizes our intuition that *statistically similar players are assigned similar representations under similar game context.*

Comparison methods:

- **Deterministic Encoder (DE):** It is trained as a regressor to identify the acting player and implements a *deterministic projection* from the game context to player embedding *without modeling the prediction uncertainty (or variance).*

- **Conditional Variational Auto-Encoder (CVAE):** Compared to our VHER, CVAE conditions the player representation on current game observation, which does not incorporate the play history into embedding computation.

- **LSTM**: To study the influence of player embedding, we also include an LSTM as our third baseline model. LSTM directly finishes the experimented tasks without including any player information.

Running setting:

- We divide the dataset containing 1,196 games into a training set (80%), a validation set (10%) and a testing set (10%).

Identify the Acting Player:

- Study the performance of VHER as a predictive model.

- We also experiment with the options of including players' pre-game cumulative box score (e.g. The total number of goals, assists, points, etc.,) into game context.

Results:

- We show the results in the following table:

Method	No Box Score		With Box Score	
	ACC	LL	ACC	LL
DE	10.91 %	-19.482	14.85 %	-18.590
CVAE	7.42 %	-4.294	17.21 %	-4.850
LSTM	12.41%	-3.131	64.47%	-1.718
VHER	48.00 %	-2.228	82.13%	-1.402

Predicting the Expected Goal (EG):

- EG weights each shot by the chance of it leading to a goal.

- To see if the embedding will improve the prediction accuracy of EG, we 1) generate the player embedding for player pl_t. 2) input $z_{pl,t}, s_t, a_t$ to a validation model

Results:

Model	Metric			
Player Info	P	R	F1	LL
N/A	0.144	0.808	0.245	-0.641
Pids	0.103	0.691	0.179	-0.573
DE	0.206	0.903	0.335	-2.756
CVAE	0.252	0.939	0.397	-2.589
VHER	0.624	0.846	0.718	-0.281

THANK YOU!