PRECOG: PREdiction Conditioned On Goals In Visual Multi-Agent Settings

NICHOLAS RHINEHART, ROWAN MCALLISTER, KRIS KITANI, SERGEY LEVINE

PRESENTED BY: DEREK YANG
Objective & Motivation

Conditional forecasting in multi-agent environment:
- How other agents respond to different decisions
- How agent’s goal affects trajectory forecasting
Notations

Total time steps: $T$

Number of agents (vehicles): $A$

Robot’s position ($x,y$ coordinate) at time $t$: $S^r_t \in \mathbb{R}^2$

Other agents’ position at time $t$: $S^h_t \in \mathbb{R}^{A-1 \times 2}$

Complete state: $S \in \mathbb{R}^{T \times A \times 2}$

Observation (LIDAR): $\chi \in \mathbb{R}^{200 \times 200 \times 2}$

Agent perception: $\phi = \{S_{-T:0}, \chi\}$

* $\chi$ is robot-centric.
Methodology

Estimating Social-forecast Probability (ESP)

- Probabilistic modeling: $S \sim q(S|\phi)$
- $q$ is considered Gaussian
- Sampling $S$ from $q$:

$$Z \sim \mathcal{N}(0, I); \quad S = f(Z; \phi); \quad S, Z \in \mathbb{R}^{T \times A \times D}.$$

$$S_t^a = \mu^a_\theta(S_{1:t-1}, \phi) + \sigma^a_\theta(S_{1:t-1}, \phi) \cdot Z_t^a \in \mathbb{R}^D$$

- Probability:

$$q(S_t|S_{1:t-1}, \phi) = \prod_{a=1}^A \mathcal{N}(S_t^a; \mu_t^a, \Sigma_t^a), \quad \Sigma_t^a = \sigma_t^a \sigma_t^{aT}$$

$Z$: A latent ‘action’ made by agents.
Methodology

Estimating Social-forecast Probability (ESP)

- Implementation:

\[
S_t^a = \mu_t^a(S_{1:t-1}, \phi) + \sigma_t^a(S_{1:t-1}, \phi) \cdot Z_t^a \in \mathbb{R}^D
\]

\[
S_t^a = \frac{2S_{t-1}^a - S_{t-2}^a + m_t^a(S_{1:t-1}, \phi)}{\mu_t^a} + \frac{\sigma_t^a(S_{1:t-1}, \phi)}{\sigma_t^a} \cdot Z_t^a.
\]

- \(2S_{t-1}^a - S_{t-2}^a = S_{t-1}^a + (S_{t-1}^a - S_{t-2}^a)\), constant velocity
Methodology

PREdiction Conditioned On Goals (PRECOG)

- ESP: An estimation of multi-agent state, based on random control input $Z$: $q(S|\phi)$
- What if we can assign $z^r$, i.e. the control input of the robot, and we want to reach a specific goal?
- Objective: Given goal $G$ at time $T$, find trajectory $S$ that maximizes the likelihood of reaching goal.
- A lower bound of $p(S|G)$ based on ESP (first term):

$$
\log \mathbb{E}_{Z^h}[p(S|G, \phi)] \geq \mathbb{E}_{Z^h}[\log p(S|G, \phi)] \quad (6)
$$

$$
= \mathbb{E}_{Z^h}[\log (q(S|\phi)p(G|S, \phi))] - \log p(G|\phi) \quad (7)
$$

$$
\mathcal{L}(z^r, G) \doteq \mathbb{E}_{Z^h}[\log q(S|\phi) + \log p(G|S, \phi)] \quad (8)
$$

$$
= \mathbb{E}_{Z^h}[\log q(f(Z)|\phi) + \log p(G|f(Z), \phi)], \quad (9) \quad p(G|S, \phi) = \mathcal{N}(w; \hat{S}^r_T, \epsilon I).
$$

$$
z^{r*} = \arg\max_{z^r} \mathcal{L}(z^r, G).
$$
Analysis and results

Metric:

- Log-likelihood ($p', \eta$ are used to resolve numerical issues):

\[
H(p, q) = -\mathbb{E}_{S^* \sim p(S^* | \phi)} \log q(S^* | \phi)
\]
\[
\hat{e} = \left[ H(p', q) - H(\eta) \right] / (T_{AD})
\]

- MiniMSD:
  - Instead of MSE, allows multimodal estimation
  - Does not penalize other plausible samples

\[
\hat{m}_K^a = \mathbb{E}_{S^* \sim p(S^* | \phi)} \| S^* - S^{a, (k^\dagger)} \|^2 / T,
\]
\[
k^\dagger = \arg\min_{k \in \{1..K\}} \| S^* - S^{(k)} \|^2.
\]
Analysis and results

Dataset:
- **nuScenes:**
  - 850 episodes of 20 seconds of driving
  - 2 seconds of past and 4 seconds of future positions at 5H
  - LIDAR map

- **CARLA:**
  - Town_01, over 900 episodes of 100 seconds
  - 2~5 agents
  - 60,701 train, 7586 validation, and 7567 test examples
  - 2 seconds of past and 2 seconds of future positions at 10H
## Analysis and results

### ESP Results:

<table>
<thead>
<tr>
<th>Approach</th>
<th>CARLA Town02 Test</th>
<th>nuScenes Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 agents</td>
<td>3 agents</td>
</tr>
<tr>
<td></td>
<td>Test $\tilde{m}_{K=12}$</td>
<td>Test $\tilde{c}$</td>
</tr>
<tr>
<td>KDE</td>
<td>4.488 ± 0.145</td>
<td>8.179 ± 1.523</td>
</tr>
<tr>
<td>DESIREE [19]</td>
<td>1.159 ± 0.027</td>
<td>-</td>
</tr>
<tr>
<td>SocialGAN [14]</td>
<td>0.902 ± 0.022</td>
<td>-</td>
</tr>
<tr>
<td>R2P2-MA [30]</td>
<td>0.454 ± 0.014</td>
<td>0.577 ± 0.004</td>
</tr>
<tr>
<td>Ours: ESP, no LIDAR</td>
<td>0.633 ± 0.017</td>
<td>0.579 ± 0.006</td>
</tr>
<tr>
<td>Ours: ESP</td>
<td>0.393 ± 0.014</td>
<td>0.550 ± 0.004</td>
</tr>
<tr>
<td>Ours: ESP, flex. count</td>
<td>0.488 ± 0.017</td>
<td><strong>0.537 ± 0.002</strong></td>
</tr>
<tr>
<td>KDE</td>
<td>19.375 ± 0.798</td>
<td>3.760 ± 0.015</td>
</tr>
<tr>
<td>DESIREE [19]</td>
<td>3.473 ± 0.102</td>
<td>-</td>
</tr>
<tr>
<td>SocialGAN [14]</td>
<td>2.119 ± 0.087</td>
<td>-</td>
</tr>
<tr>
<td>R2P2-MA [30]</td>
<td>1.336 ± 0.062</td>
<td>0.951 ± 0.007</td>
</tr>
<tr>
<td>Ours: ESP, no LIDAR</td>
<td>1.496 ± 0.069</td>
<td><strong>0.920 ± 0.008</strong></td>
</tr>
<tr>
<td>Ours: ESP</td>
<td>1.325 ± 0.065</td>
<td>0.903 ± 0.008</td>
</tr>
<tr>
<td>Ours: ESP, Road</td>
<td>1.081 ± 0.053</td>
<td>0.929 ± 0.008</td>
</tr>
<tr>
<td>Ours: ESP, flex. count</td>
<td>1.464 ± 0.067</td>
<td>0.980 ± 0.003</td>
</tr>
</tbody>
</table>
Analysis and results

PRECOG Results:

- Still forecasting, NOT planning
- Find $z^*$ on trained ESP:

\[ z^{r*} = \arg \max_{z^r} \mathcal{L}(z^r, \mathcal{G}). \]

- Plug it back to trained ESP and sample trajectories.
## Analysis and results

**PRECOG Results:**

<table>
<thead>
<tr>
<th>Data</th>
<th>Approach</th>
<th>Test $\hat{m}_{K=12}$</th>
<th>Test $\hat{m}_{K=12}$</th>
<th>Test $\hat{m}_{K=12}$</th>
<th>Test $\hat{m}_{K=12}$</th>
<th>Test $\hat{m}_{K=12}$</th>
<th>Test $\hat{m}_{K=12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CARLA</td>
<td>DESIRE [19]</td>
<td>1.837 ± 0.048</td>
<td>1.991 ± 0.066</td>
<td>1.683 ± 0.050</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>DESIRE-plan</td>
<td>1.858 ± 0.046</td>
<td>0.918 ± 0.044</td>
<td>2.798 ± 0.073</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>ESP</td>
<td>0.337 ± 0.013</td>
<td>0.196 ± 0.009</td>
<td>0.478 ± 0.024</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>PRECOG</td>
<td><strong>0.241 ± 0.012</strong></td>
<td><strong>0.055 ± 0.003</strong></td>
<td><strong>0.426 ± 0.024</strong></td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CARLA</td>
<td>DESIRE [19]</td>
<td>2.622 ± 0.030</td>
<td>2.621 ± 0.045</td>
<td>2.422 ± 0.048</td>
<td>2.710 ± 0.066</td>
<td>2.969 ± 0.057</td>
<td>2.391 ± 0.049</td>
</tr>
<tr>
<td></td>
<td>DESIRE-plan</td>
<td>2.329 ± 0.038</td>
<td>0.194 ± 0.004</td>
<td>2.239 ± 0.057</td>
<td>3.119 ± 0.098</td>
<td>3.332 ± 0.090</td>
<td>2.758 ± 0.083</td>
</tr>
<tr>
<td></td>
<td>ESP</td>
<td>0.718 ± 0.012</td>
<td>0.340 ± 0.011</td>
<td>0.759 ± 0.024</td>
<td>0.809 ± 0.025</td>
<td>0.851 ± 0.023</td>
<td>0.828 ± 0.024</td>
</tr>
<tr>
<td></td>
<td>PRECOG</td>
<td><strong>0.640 ± 0.011</strong></td>
<td><strong>0.066 ± 0.003</strong></td>
<td><strong>0.741 ± 0.024</strong></td>
<td><strong>0.790 ± 0.024</strong></td>
<td><strong>0.804 ± 0.022</strong></td>
<td><strong>0.801 ± 0.024</strong></td>
</tr>
<tr>
<td>nuScenes</td>
<td>DESIRE [19]</td>
<td>3.307 ± 0.093</td>
<td>3.002 ± 0.088</td>
<td>3.613 ± 0.140</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>DESIRE-plan</td>
<td>4.528 ± 0.151</td>
<td>0.456 ± 0.015</td>
<td>8.600 ± 0.298</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>ESP</td>
<td>1.094 ± 0.053</td>
<td>0.955 ± 0.057</td>
<td>1.233 ± 0.078</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>PRECOG</td>
<td><strong>0.514 ± 0.037</strong></td>
<td><strong>0.158 ± 0.016</strong></td>
<td><strong>0.871 ± 0.070</strong></td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>nuScenes</td>
<td>DESIRE [19]</td>
<td>6.830 ± 0.204</td>
<td>4.999 ± 0.219</td>
<td>6.415 ± 0.294</td>
<td>7.027 ± 0.360</td>
<td>7.418 ± 0.324</td>
<td>8.290 ± 0.532</td>
</tr>
<tr>
<td></td>
<td>DESIRE-plan</td>
<td>6.562 ± 0.207</td>
<td>2.261 ± 0.100</td>
<td>6.644 ± 0.314</td>
<td>6.184 ± 0.325</td>
<td>9.203 ± 0.448</td>
<td>8.520 ± 0.514</td>
</tr>
<tr>
<td></td>
<td>ESP</td>
<td>2.921 ± 0.175</td>
<td>1.861 ± 0.109</td>
<td>2.369 ± 0.188</td>
<td>2.812 ± 0.188</td>
<td>3.201 ± 0.254</td>
<td>4.363 ± 0.652</td>
</tr>
<tr>
<td></td>
<td>PRECOG</td>
<td><strong>2.508 ± 0.152</strong></td>
<td><strong>0.149 ± 0.021</strong></td>
<td><strong>2.324 ± 0.187</strong></td>
<td><strong>2.654 ± 0.190</strong></td>
<td><strong>3.157 ± 0.273</strong></td>
<td><strong>4.254 ± 0.586</strong></td>
</tr>
</tbody>
</table>
Personal Comments

Novelties:
- Exact likelihood inference (Gaussian assumption)
- Latent space planning
- Goal-conditioned generative multi-agent forecasting

Shortcomings:
- Assumption of symmetric Information
- Although PRECOG is modelled as a planning algorithm, the control is done in latent space.
- CARLA evaluation is done only on datasets, no actual simulation of the planning.
- If we can somewhat link $Z'$, control input in latent space, with actual control inputs (e.g. steer, brake, gas, etc.), we might result in an actual planning algorithm.
Takeaway

Goal-conditioned, multi-agent trajectory forecasting.

- Multimodal using latent control input.
- Exact likelihood inference with Gaussian assumption.
Quiz

Give an appropriate loss, or objective, for ESP.
Thanks