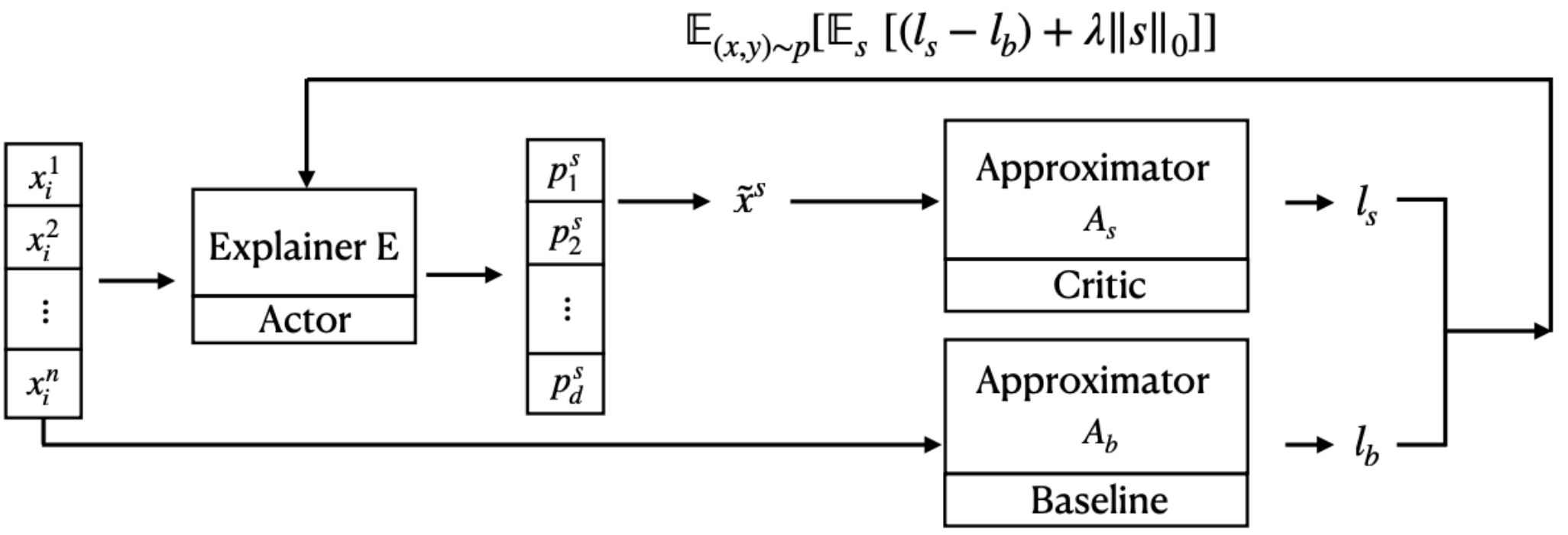
Related Work


$\mathbb{E}_{(x,y)\sim p}[\mathbb{E}_s\ [(l_s - l_b) + \lambda\|s\|_0]]$
x_i^1
x_i^2
$\vdots$
x_i^n
Explainer E
Actor
p_1^s
p_2^s
$\vdots$
p_d^s
$\tilde{x}^s$
Approximator
A_s
Critic
Approximator
A_b
Baseline
l_s
l_b

✤ Training Time

- Does longer training lead to better performance of the explainer?

✤ Sensitivity of Hyper-parameter

- How sensitive is the explainer's performance to the choice of hyperparameter λ?

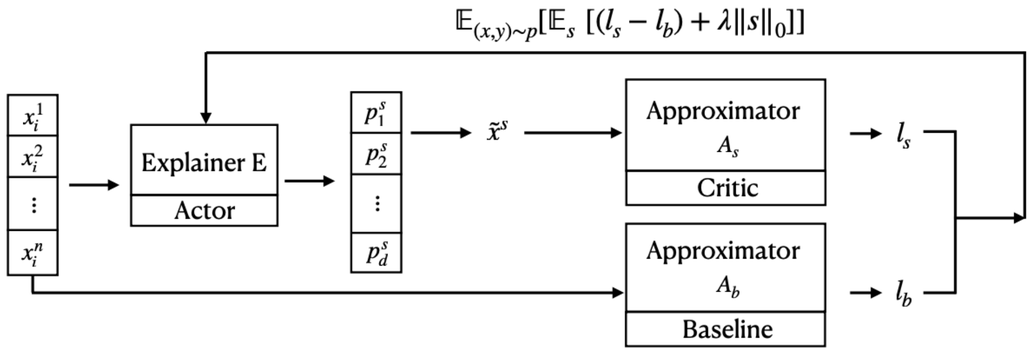
✤ Necessity of a Baseline Model

- Does the presence of a baseline model always enhance the explainer's performance?

INVASE (Actor-Critic Based)

Related Work

INVASE (Actor-Critic Based)



- **Training Time**
  - Does longer training lead to better performance of the explainer?
- **Sensitivity of Hyper-parameter**
  - How sensitive is the explainer's performance to the choice of hyperparameter λ?
- **Necessity of a Baseline Model**
  - Does the presence of a baseline model always enhance the explainer's performance?

Contents

1. Backgrounds & Related Work
2. **Research Questions**
3. Preliminaries
4. Analysis of Performance Influencing Factors in INVASE
5. Exploration of Potential Improvements to INVASE
6. D-ACES Framework Proposal and Analysis
7. Conclusion