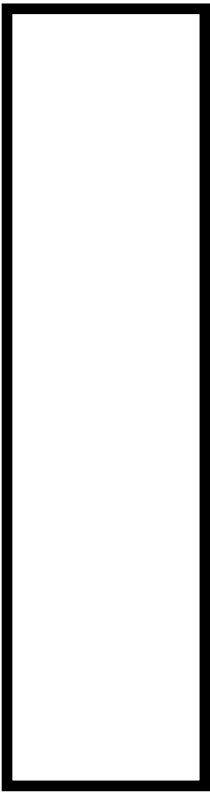


Related Work

Actor-Critic Based Methods

- Actor (Explainer): Pick features for a given instance
- Critic: Evaluate the goodness of current Actor













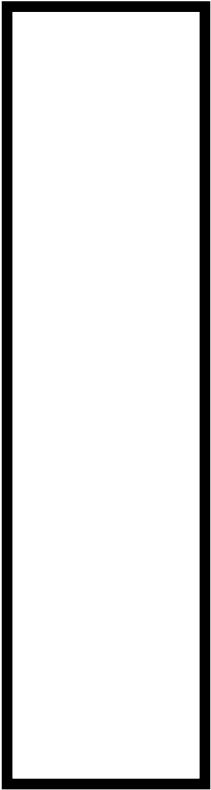


x_i^1

x_i^2

x_i^d











P_i^1

p_i^2

p_i^d



Explainer (Actor)







Critic





Feedback

✿ Advantages:

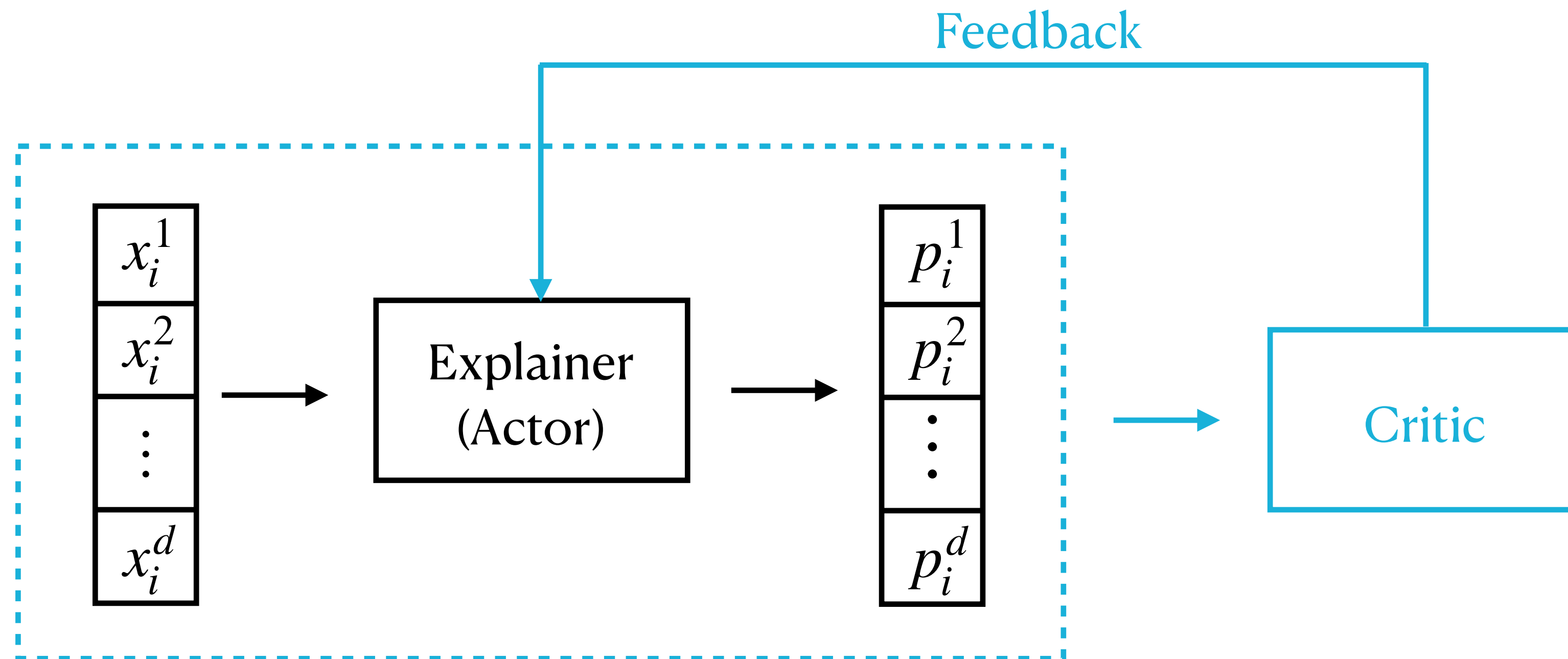
- Give an appropriate feedback (loss) function
- Totally data-driven

✿ Drawbacks:

- High difficulty to find a nice feedback function

Related Work

Actor-Critic Based IFS Methods



✦ Advantages:

- Give an appropriate feedback (loss) function
- Totally data-driven

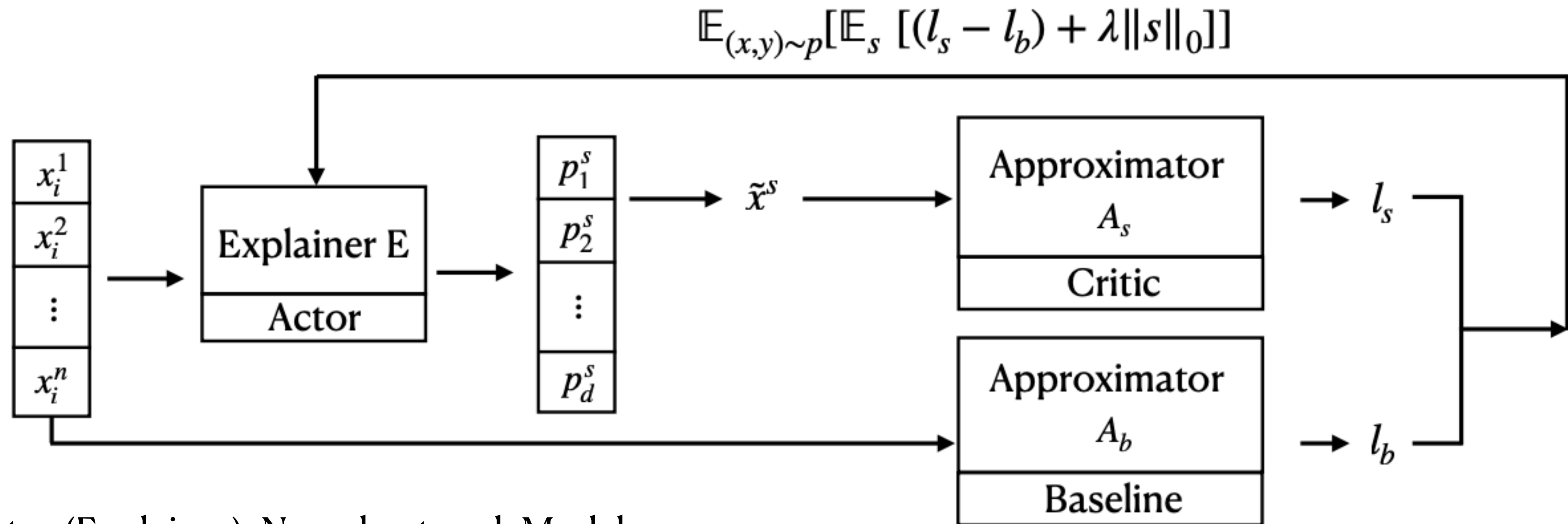
✦ Drawbacks:

- High difficulty to find a nice feedback function

- Actor (Explainer): Pick features for a given instance
- Critic: Evaluate the goodness of current Actor

Related Work

INVASE (Actor-Critic Based)



- Actor (Explainer): Neural network Model
- Critic: Neural network Model that use partial feature values to approximate original black-box model
- ✓ Achieve nice performance on synthetic datasets in terms of TPR & FDR