



Xiangyu Sun, Jack Davis, Oliver Schulte, Guiliang Liu Simon Fraser University



# **Distilling Deep Sports Analytics**

#### The Motivation

- The problems:
  - The trade-off between accuracy and transparency
  - Provide insights to sports analytics
- Deep learning
- Linear model tree
- Propose and compare model tree learning heuristics







### The Neural Network Model

- Deep Reinforcement Learning model  $\bullet$ 
  - Consists of five layers:
    - Input layer
    - LSTM hidden layer
    - Two fully connected hidden layers
    - Output layer
  - SARSA is applied to estimate a Q-function (with discount factor=1)

[Liu, G., & Schulte, O. (2018)]

#### $\mathcal{L}_{t}(\theta_{t}) = \mathbb{E}[(R_{t} + \hat{Q}(S_{t+1}, A_{t+1}; \theta_{t}) - \hat{Q}(S_{t}, A_{t}; \theta_{t}))^{2}]$

$$\theta_{t+1} = \theta_t - \alpha \cdot \nabla_\theta \mathcal{L}_t(\theta_t)$$



#### **The Mimic Model**

- A linear model tree:
  - Regression tree with linear models on leaves

$$\hat{y} = \left(\sum_{i} w_i \cdot x_i\right) + b$$

- Interpretable
- High fidelity





#### **Action Replacement**

- A data augmentation technique in counterfactual strategic settings
- Replace observed action A by A' and ask the neural network model to predict  $Q(S, A') \rightarrow$  Expect lower Q-value





## **Heuristics for Computing Split Points**

- To address computational challenge on large datasets
- Our proposed heuristics are:
  - Iterative Segmented Regression
  - Sorting with Variance Reduction
  - Sorting with T-test

### **RMSE, Correlation and Running Time**

	Ice Hockey				Soccer							Ice Hockey				Soccer		
	Shots		Passes		Shots		Passes				Shots		Pas	Passes		Shots		
Split methods	action- values	impacts	action- values	impacts	action- values	impacts	action- values	impacts		Split methods	action- values	impacts	action- values	impacts	action- values	impacts	action- values	
Gaussian Mixture	0.05483	0.01990	0.04276	0.00687	0.00698	0.01312	0.01000	0.00577		Gaussian Mixture	0.91498	0.93709	0.94737	0.91687	0.99458	0.99001	0.98386	
Iterative Segmented Regression	0.01441	0.01999	0.00964	0.00691	0.00508	0.01275	0.00997	0.00575		Iterative Segmented Regression	0.99436	0.93620	0.99601	0.92018	0.99650	0.99422	0.98695	
Sorting + Variance Reduction	0.01219	0.01627	0.01012	0.00686	0.00646	0.01235	0.01092	0.00603		Sorting + Variance Reduction	0.99593	0.95834	0.99561	0.92137	0.99480	0.99459	0.98438	
Sorting + T-test	0.05709	0.02487	0.06695	0.00935	0.01223	0.01377	0.01796	0.00597		Sorting + T-test	0.91036	0.89935	0.79761	0.85017	0.98943	0.98705	0.95690	
Null Model	0.13924	0.05688	0.10808	0.01756	0.13648	0.11890	0.06151	0.00961	_	Null Model	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	

Action-Value of Shots in Ice Hockey









Impact of Shots in Soccer



Action-Value of Passes in Ice Hockey













### **Debugging Deep Neural Networks**



#### The tree can highlight the potential problems of the neural network model

Thank You