Announcing Apprentice Mode (Preview)
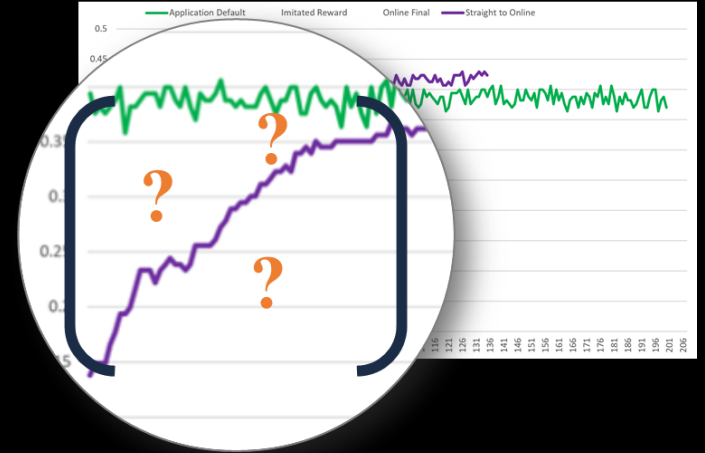
Dwaipayan Mukhopadhyay
Azure AI Platform
@dwaijam

Apprentice Mode : Try First , Buy Later

RL model can only train in production and new deployments take time to train. Apprentice mode is a learning behavior that addresses this constraint when deploying a new use case.

Mission:

Help customers adopt personalizer, by de-risking an assessment of "if" and "how long it will take" for Personalizer to match the business performance (rewards) of the current business logic.

Apprentice Mode

Use Apprentice Mode

Monitor apprentice mode metrics, available from the Azure portal and the API

Current learning behavior performance

Average rewards received over all events ⓘ

Application default
Personalizer
Rolling average reward

Application default	Personalizer	Rolling average reward	All events
Average reward	Average reward	Over 1000 events	
0.0071	0.0032	45%	875,270

Personalizer configuration | **Learning behavior**

Select the behavior mode for Personalizer:
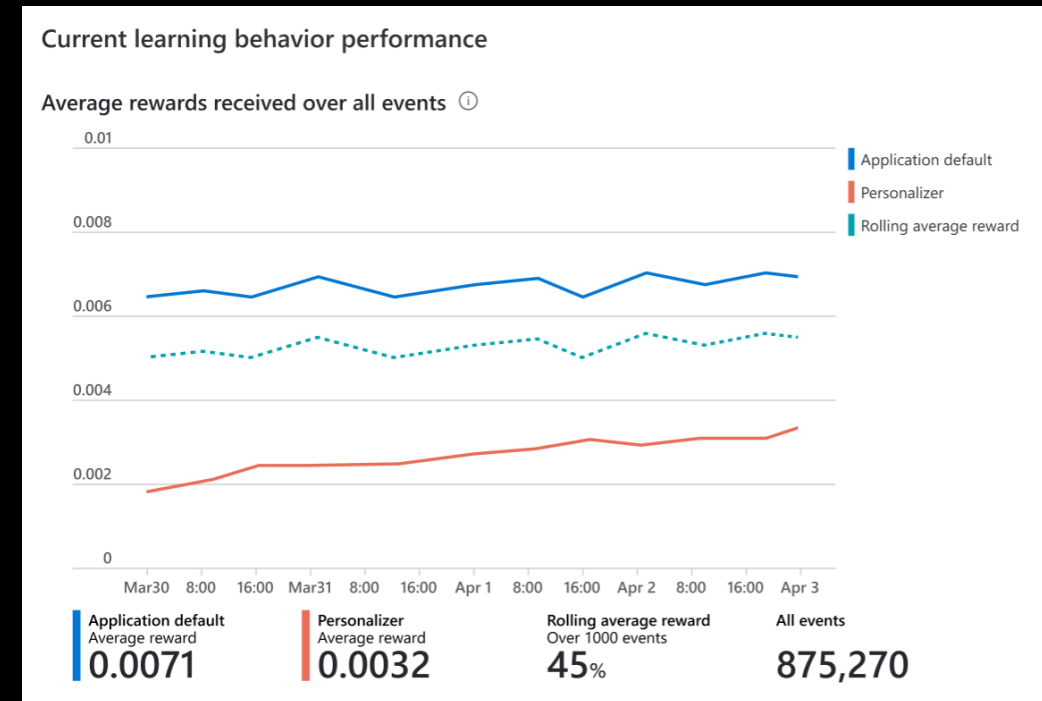
○ Return the best action, learn online

Your model will respond to Rank calls with the best action and will use Reward calls to learn and improve its selections over time.

◉ Return baseline action, learn as an apprentice

Your model will learn by observing the behavior of your current application. Rank calls will always return the first action (baseline).

Once Personalizer closely imitates baseline logic , switch over easily from Apprentice to Online mode on Azure portal

Apprentice Mode Learning

Baseline policy action order = {Apple, Banana, Orange}

	#1 (breakfast)	#2 (lunch)	#3 (dinner)
Baseline recommended action - A_0	Apple	Apple	Apple
Reward Obtained by A_0 - R_0	1	0.5	0
Personalizer recommended action - A_P	Banana	Apple	Apple
Reward Used for Personalizer - R_P R_P = if(A_P==A_0) then R_0 else 0	0	0.5	0

Imitation Ratio = $Count(Events^{when\ A_P\ ==\ A_0})/Count(Events)$

$\approx$ 60-80%

Reward Attainment Ratio = $\sum R_0{}^{when\ A_P\ ==\ A_0} / \sum R_0$

Why use Apprentice mode?

- **Developers** - To make sure the Rank and Reward APIs are being used correctly in the application, and that features being sent to Personalizer from the application are noise-free.

- **Data scientists** - To validate that the features are effective to train the Personalizer models, that the reward wait times aren't too long or short.

- **Business Decision Makers** - To gain confidence in Personalizer to improve results (i.e. rewards) compared to existing business logic and to mitigate cold start problems.

Apprentice mode (in Public Preview)

- **F0 tier** - Free to use for low traffic.

- **E0 tier –** For use at higher scale.

Thank You and have fun using Personalizer!!