Predictive Routing Deployment and Operations Guide

Routing Scenarios Using GPR

5/5/2025

Contents

Routing Scenarios Using GPR

To deploy the GPR subroutines, you modify your IRD strategies or Composer workflows o incorporate them. Rather than picking the Agent with required skills who has been available longest, or using simple Agent Group routing, Predictive Routing predicts the best results for a specific interaction, based on customer intent or other relevant information.

This topic presents the following information:

- A high-level view of a Predictive Routing interaction flow
- How the URS Strategy Subroutines work together to score agents and identify a routing target
- Routing scenarios using Predictive Routing, explaining how URS ranks agents by score

> **Important**
>
> - If you would like to evaluate Genesys Predictive Routing for use with schedule-based routing (using Genesys Workforce Management), service-level routing, or business-objective routing, contact Genesys Professional Services for a review of your routing environment.
> - If your environment uses multiple URS instances receiving interactions from a single T-Server, the only criterion used to select the next interaction for routing is priority.
> - This topic assumes that you are using a virtual agent group (VAG) as the target for your routing. If you route using a skill expression to identify your targets, convert it to a VAG string expression using the IRD functions MultiSkill or CreateSkillGroup before passing the resulting string as an argument to the ActivatePredictiveRouting_v3 subroutine. See Using Agent Skills for Ideal Agent Selection in the *Supplement to the Universal Routing 8.1 Reference Manual* for more information.

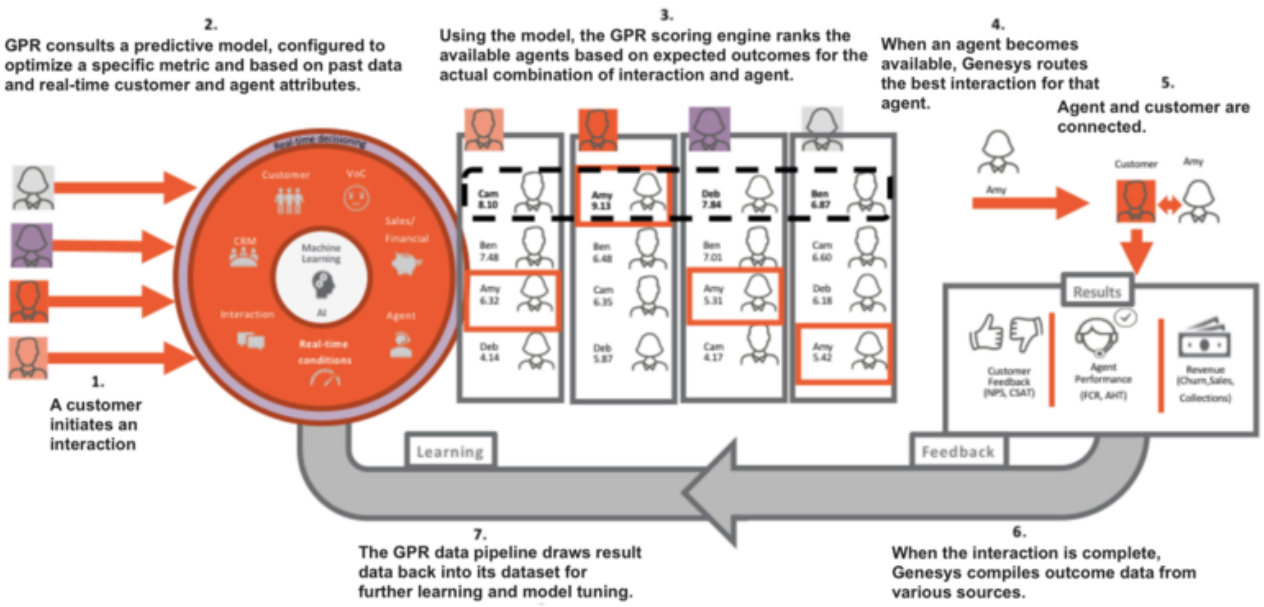
High-Level Predictive Routing Interaction Flow

The graphic below shows a very general interaction flow using Predictive Routing. Refinements to the flow depend greatly on details of your environment. Key aspects that differ in various environments:

- Your data - That is, the interaction types supported and the applications that might have relevant information. Genesys Info Mart is a key data source, but CRM systems and other applications in your environment can also provide important data. See Prepare Your Data for more information.
- Your pre-routing data flow - This depends on the interaction type and the exact architecture in your environment. For example, is this a chat interaction or a call? Do you use an IVR, and if so, what information do you attach?
- The Genesys routing solution you are using - Predictive Routing supports routing with IRD/URS and with

Composer/ORS/URS.

- Your reporting solution for Predictive Routing - Whether you are using GCXI, Genesys Pulse, or another solution to present the data stored in Genesys Info Mart.



How the Strategy Subroutines Work

The following sequence provides a basic overview of the way the various GPR subroutines work together to evaluate agent scores and determine the best match given the currently available agents and the currently waiting interactions.

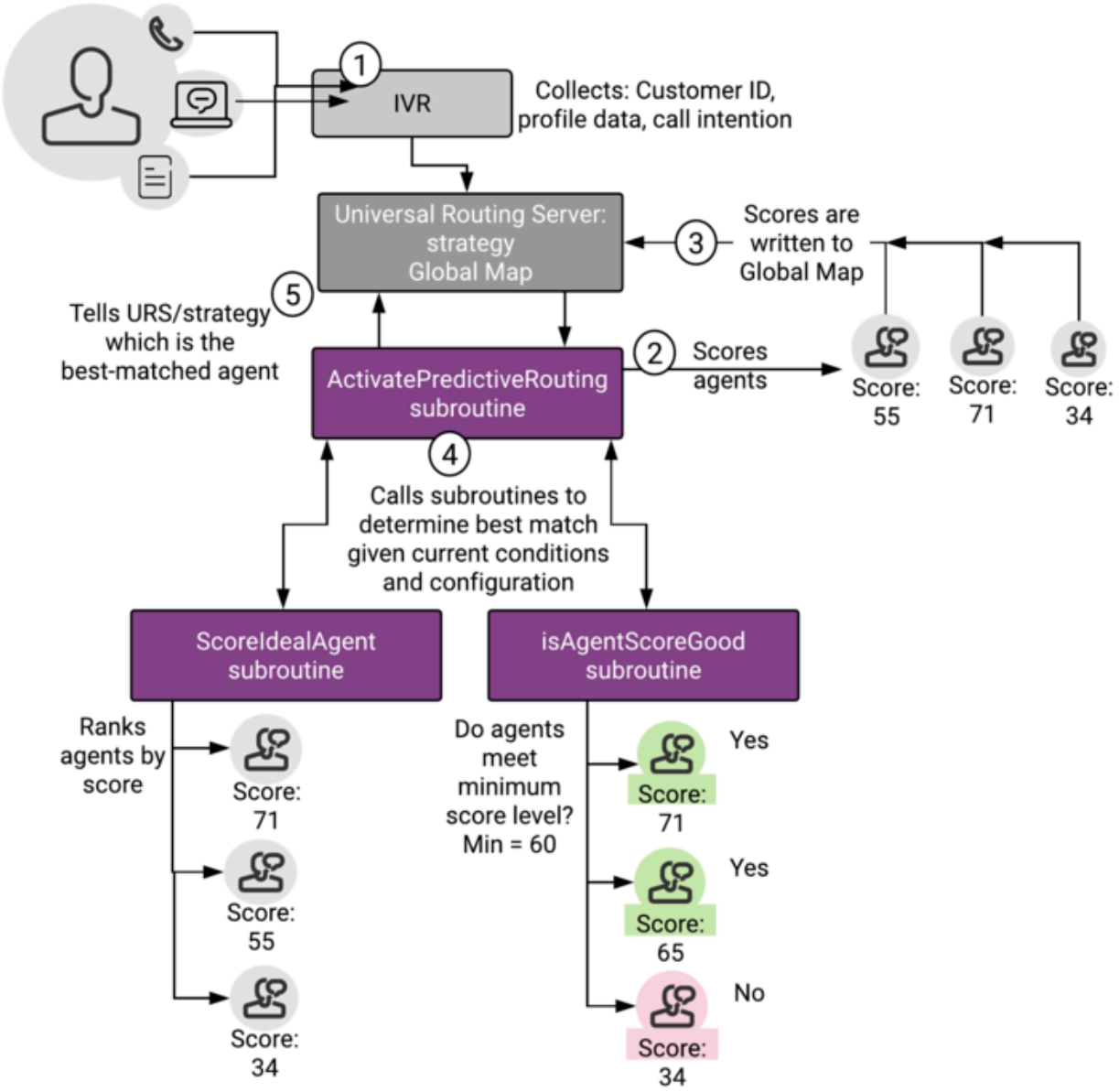
1. ActivatePredictiveRouting retrieves agent scores are retrieved from AI Core Services via REST API request and stores them in the global map in URS memory. The name of the map is the interaction ConnectionId (the original ID, if the interaction is a consult). This map contains pairs of agent employee IDs as the keys and their scores for the interaction as values. ActivatePredictiveRouting calls the SetIdealAndReadyCondition subroutine for further interaction processing.
2. SetIdealAndReadyCondition processes the different modes of Predictive Routing. It calls the SetIdealAgent IRD function to schedule the execution of the URS callback subroutines. It calls the ScoreIdealAgent subroutine to facilitate interaction queueing according to their scores, and calls SetReadyCondition (if enabled) to call the isAgentScoreGood subroutine.
   The parameters for these callback subroutines are retrieved and verified before any URS request is invoked to enable the callbacks.
3. After establishing a list of potential targets based on the target expression (Skill, Agent Group, and so on) SetIdealAndReadyCondition then executes the ScoreIdealAgent callback subroutine.
4. ScoreIdealAgent retrieves the scores for the potential target agents from the global map set in Step 1.
5. When an agent becomes ready, URS executes the isAgentScoreGood subroutine to determine whether

that target is acceptable. If you enabled agent hold-out, URS executes the isAgentScoreGood subroutine when an agent becomes ready, which determines whether that agent reaches the specified threshold score. If not, URS waits for a configured timeout period, then checks whether any agent now satisfies the adjusted threshold value. See How Does GPR Score Agents? for a detailed discussion of how agent hold-out routing works.

6. Once the isAgentScoreGood subroutine locates an available agent who scores above the current threshold, it sends the target details to URS to initiate routing.
7. URS calls the GPRIxnCompleted subroutine as a custom step from a Routing Block in the strategy. It collects the Predictive Routing outcome for the successfully routed interaction (the DBID of the agent to whom it was distributed, the score of the agent, other interaction statistics relevant for the Predictive Routing performance) and prepares the user data for Predictive Routing reporting.
8. URS calls the GPRIxnCleanup subroutine from both the success and failure exits from the Routing block, or if the interaction is abandoned. The purpose of the subroutine is to publish the Predictive Routing reporting user data and to clean up the ScoreIdealAgent and isAgentScoreGood callback subroutines. GPRIxnCleanup publishes reporting data in two ways:
   - It sends a UserEvent containing the user data relevant for Predictive Routing to T-Server/SIP Server, from which is enters the Genesys historic reporting solution flow.
   - It can submit the same data AI Core Services via REST API request where it is stored in the score_log and can be retrieve using an API request.

Important

If your routing strategy uses the SelectDN and SuspendDN IRD functions instead of a Routing Block, consult with Genesys Professional services about how the ActivatePredictiveRouting, GPRIxnCompleted and GPRIxnCleanup subroutines can be integrated into your strategy.



Routing Scenarios Using Predictive Routing

When you are using Predictive Routing to route interactions, there are two main scenarios that affect how this matching plays out:

- **Agent Surplus** - There are relatively few interactions, which means there could be a number of high-score agents available. You can configure a minimum threshold so that, if the agents available are not

very highly ranked, the strategy keeps the interaction in queue until a better-scoring agent becomes available.

- **Interaction Surplus** - There are many interactions, so that most agents are busy and it might be more difficult to find an ideal agent for each interaction. In such a scenario, you can have agents matched to the interaction for which they have the highest probability of getting a positive result.

Agent Surplus Flow

In this case there are agents logged in and in the Ready state who can respond to interactions immediately. From a Virtual Agent Group that is defined by skill expression, URS first tries to route an interaction to an agent with the best score, using the following process to match agents and interactions:

1. An interaction arrives at the routing strategy, which has a target group of agents.
2. The ActivatePredictiveRouting subroutine sends a request to the Predictive Routing scoring server via HTTP request.
3. Predictive Routing returns scores for each agent in the target group based on the criteria you selected in the active model.
4. The ActivatePredictiveRouting subroutine updates a global cache in URS memory, which keeps agent scores for all interactions. When URS tries to route the current interaction to the agent group, it sorts the agents according to their scores, in descending order, and routes to the agents with the best score first.

When URS takes an interaction from the queue:

1. URS calls the ScoreIdealAgent subroutine, which reads the agent scores in the target group from global map and ranks the agents by score.
2. URS calls the IsAgentScoreGood subroutine, which selects the available agent with the highest score, assuming the agent has a score high enough to be selected for this interaction.
   In an agent-surplus scenario, it is typically not a problem to route to an agent with a good score. For scenarios where this is not the case, see **Interaction Surplus Flow**, below.
3. URS calls the PrrIxnCompleted subroutine, which updates user data with the scoring result for storage in Genesys Info Mart.
4. URS calls the PRRLog macro, which logs the result in the URS log file.

Interaction Surplus Flow

This scenario covers situations when all agents are already busy handling interactions and new interactions are queued. When one of the agents becomes ready, the system selects the interaction for which the agent has the best score. This is not necessarily the interaction that has been in the queue longest.

When interactions are waiting, URS uses a number of criteria to decide the order in which it directs the interactions to the best target. In general, URS uses the following hierarchy:

1. Interaction priority.

2. Best agent score.

> **Important**
>
> In scenarios where both scored and unscored interactions might have the same priority, scoring is disregarded for all the interactions and the selection is based on the next differentiating criterion, time.

3. Time in queue, which can be based on age of interaction or time in queue and can incorporate predicted wait time.
4. Interaction ID (URS selects the interaction with the lowest—oldest—ID). This is a rarely-used "tie-breaker" criterion.

Using gpmStatus and gpmSuitableAgentsCount to Monitor Your Routing

gpmStatus and gpmSuitableAgentsCount are KVP values written in the Genesys Info Mart database when an interaction is routed using the GPR subroutines. (You can also retrieve the values by using the GPR API to query the score log.)

- gpmStatus indicates whether there was an agent-surplus or an interaction-surplus condition when the interaction was routed.
- gpmSuitableAgentsCount indicates the number of agents who have scores returned from AICS greater than or equal to the initial threshold value when the scoring response is received. If gpmSuitableAgentsCount is `0`, then no agents have eligible scores compared with the threshold value, so the interaction must wait for a higher-scoring agent to become available or for the next threshold relaxation step

These KVP values, when analyzed for different interactions over a representative day or week period can help you understand your contact center traffic and GPR performance. The following table indicates certain scenarios and how to interpret them.

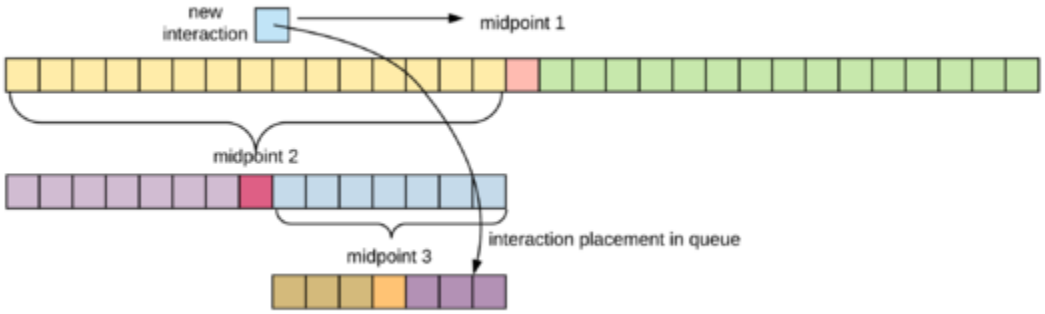
KVP Values	Inference
gpmStatus = caller-surplus gpmSuitableAgentsCount > 0	Your GPR Model is returning useful scores with relation to the configured routing threshold, but agent staffing is not adequate to produce satisfactory wait times.
gpmStatus = agent-surplus gpmSuitableAgentsCount > 0 or consistently a very small number	Analyze why the scores GPR returns are not meeting the configured threshold. You might need to retrain your Model, adjust the scoring expression, or reduce the threshold level.

How Interactions are Sorted within the Queue

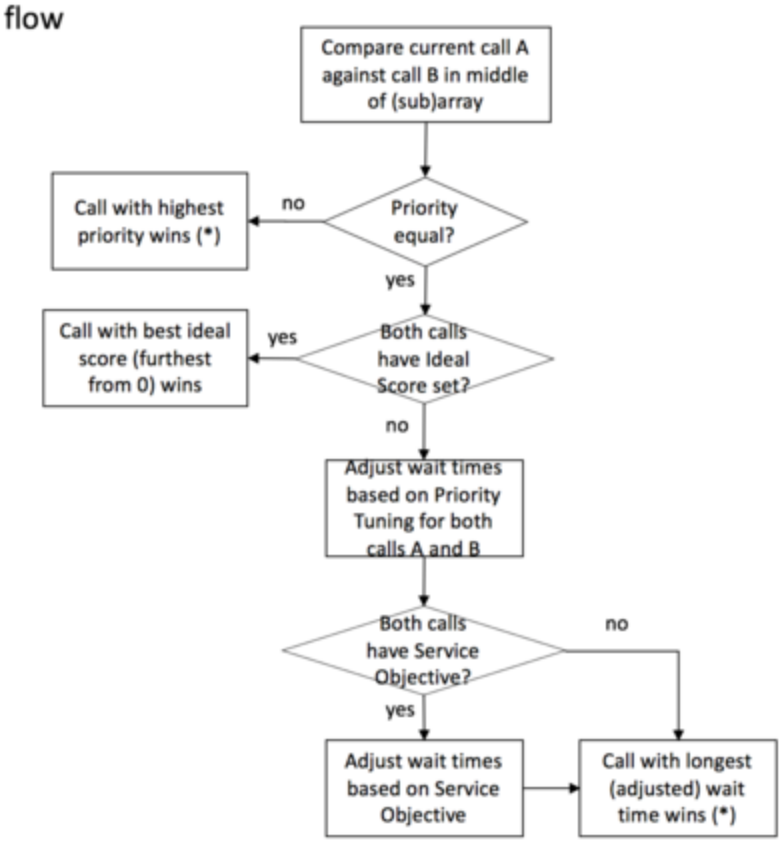
As each interaction comes in, it is scored, and then assessed relative to the interaction at the midpoint of the existing array of interactions. Does it come before or after the mid-point?

> **Important**
>
> The order in which interactions are prioritized is called an *array* here. This is not equivalent to a queue. These interactions might be from multiple queues, each of which is submitting interactions for URS sorting and routing.

After this decision, URS compares the new interaction against the midpoint within the selected region. Each time URS evaluates the interaction, it is assigned to a smaller region with the total array, always relative to the midpoint of the previous region.



The sorting decision tree:



Example 1

Three calls arrive at a contact center:

- C1 - priority = 1, agent score = 0.3, timestamp = 0:00, URS ID = 1
- C2 - priority = 1, agent score = n/a, timestamp = 0:05, URS ID = 2
- C3 - priority = 1, agent score = 0.6, timestamp = 0:10, URS ID = 3

1. C1 arrives first and is placed into the empty array.
2. C2 arrives. URS compares it with the middle (in this case, only) call in the array, C1.
   The priority is equal and, because C2 has no agent score, URS moves to the next decision criterion.
3. C2 has a shorter wait time, so is put behind C1.
   With only two calls in the array, no further comparison is needed.
4. C3 arrives. URS compares it against the "middle" entry of the array, C1.
   The priority is equal. C3 has better score (further from 0), so URS puts it in front of C1.

Outcome: C3, C1, C2 (the example assumes that interactions are taken from the left end of the array)

Example 2

Five calls arrive at a contact center and are placed in either the Predictive Routing queue or a

conventional queue:

- C1 - priority = 1, agent score = n/a, timestamp = 0:00, URS ID = 1
- C2 - priority = 1, agent score = 0.5, timestamp = 0:05, URS ID = 2
- C3 - priority = 1, agent score = n/a, timestamp = 0:10, URS ID = 3
- C4 - priority = 1, agent score = 0.75, timestamp = 0:15, URS ID = 4
- C5 - priority = 1, agent score = 0.95, timestamp = 0:20, URS ID = 5

1. C1 arrives first. URS places it into the empty array.
2. C2 arrives. URS compares it with the middle (in this case, only) call in the array, C1.
   The priority is equal and, because C1 has no agent score, URS moves to the next decision criterion.
3. C2 has a shorter wait time, so is put behind C1. (In this example,
   Current order: C1 C2 (the example assumes that interactions are taken from the left end of the array)

   With only two calls in the array, no further comparison is needed.
4. C3 arrives. URS compares it against the "middle" entry of the array, C2.
   The priority is equal and, because C3 has no agent score, URS moves to the next decision criterion.
5. C3 has a shorter wait time, so is put behind C2.
   Current order: C1 C2 C3
6. C4 arrives. URS compares it against the middle entry of the array, C2.
   The priority is equal. C4 has a better score (further from 0), so URS places it before C2.
7. Now URS must determine whether C4 should be before or after C1, which is also before C2.
   The priority is equal and, because C3 has no agent score, URS moves to the next decision criterion.
8. C4 has a shorter wait time (a more recent timestamp), so URS places it behind C1.
   Current order: C1 C4 C2 C3
9. C5 arrives. URS compares it against the "middle" entry of the array, C2.
   The priority is equal. C5 has a better score (further from 0), so URS places it before C2.
10. Now URS must determine whether C5 should be before or after C4, the "middle" call in the section of the array before C2.
    The priority is equal. C5 has a better score, so URS places it before C4.
11. Now URS must determine whether C5 should be before or after C1.
12. C5 has a shorter wait time, so URS places it behind C1.
    Final order: C1 C5 C4 C2 C3

Using Agent Hold-Out

Agent hold-out enables you to have an interaction wait a specified time, even when an agent has become available, if the available agent is has a low score for the interaction and there is a chance a better-matched agent might become available within the configured time window. The interaction flow is as follows:

1. URS calls the IsAgentScoreGood subroutine, which determines whether any of the available agents meet the threshold for handling the interaction.

2. If available agents have low scores for this interaction and the interaction spent only a short time in the queue, URS waits for a better agent to become ready.
3. The minimum acceptable score required for an agent for the interaction is gradually reduced, so if no higher-scored agent becomes available, the lower-scored agent might finally be given the interaction.

After that determination occurs, the remainder of the flow is the same as that given in the agent-surplus flow above. Use the relevant Predictive_Route_DataCfg Transaction List Object configuration options to set up the priority increments.

Dynamic Interaction Priority Increments

To avoid having interactions lingering in a queue for an excessive amount of time, URS can trigger an escalation in interaction priority after a time delay that you set. To speed up interaction handling, you can incrementally relax the minimum skill level required for agents to handle the interaction or expand the pool of agents to consider.

Each time a routing strategy tries to route an interactions, it calls the ActivatePredictiveRouting subroutine. After each failed routing attempt, the strategy checks how long the interaction has been waiting in the queue and, if the time in queue is above a certain threshold, it routes the interaction to the next available agent, no matter their score for the interaction.

Use the relevant Predictive_Route_DataCfg Transaction List Object configuration options to set up the priority increments.