Types of Unwanted Calls

Within unwanted (or “spam”) calls, there are two categories of calls:

- **nuisance**: Likely to be annoying to some, but not all recipients. Generally originate from legitimate businesses with a real presence.
- **fraud**: Calls that are dangerous to answer, because they seek to perpetrate fraud. Generally originate from illegitimate parties whose real identity is difficult to track down.

A good call analytics service will seek to minimize user exposure to these calls, but not at the expense of blocking legitimate calls.

Creating a Data Sample

It is critically important to use a data sample for evaluation that consists of a complete log of actual phone calls placed to a set of phone subscribers that reflects the average subscriber. If two vendors are compared, Hiya recommends a sample consisting of at least 500,000 phone calls. Furthermore, the sample should be no older than 1 hour, given how often nuisance callers switch the originating phone number.

Metrics to Measure

In order to evaluate the effectiveness of the unwanted call detection solution by each competing provider, both of the following criteria need to be measured:

- **Identification Rate**: How often is an unwanted call correctly detected? Higher is better.
- **Error Rate**: How often is a call incorrectly detected as unwanted? (a.k.a. “false positives”) Lower is better.

Why both? These metrics move in opposing directions. The easiest way to achieve a high Identification Rate is to lower the threshold for what is considered nuisance or fraud. Hiya can easily do this, but this is bad practice. A lower threshold leads to a higher Error Rate, as innocent numbers become falsely labeled, and legitimate calls go unanswered. A high Identification Rate is only meaningful if the Error Rate is held sufficiently low.
Overall Success Metric

It is widely accepted that the cost of one false positive (i.e. an incorrectly detected nuisance/fraud call) outweighs the benefit of catching several unwanted calls. The reason is that false positives lead to expensive customer service calls and complaints from legitimate business callers - and especially so, if the calling business is inaccurately labeled as “fraud” or “scam”. The stakes are much higher than undelivered pizzas, and rise all the way up to personal safety issues (e.g. emergency alert calls to a neighborhood or school parents), and may carry legal and brand risk to the carrier.

These types of complaints typically do not occur with “false negatives” (i.e. failed detection of nuisance/fraud calls). In fact, Hiya has observed that customer complaints about false positives (“Why did you call me scam/nuisance?”) exceed complaints about false negatives (“Why didn’t you catch that scam caller?”) by a ratio of well over 10x.

In other words, it is far more disruptive to both the end customer and the carrier when the Error Rate is 1% too high than when the Identification Rate is 1% too low. Hiya currently advises that 1 false positive does as much harm as missing 5 identifications of unwanted calls.

Therefore, Hiya advises the following measurement for overall success:

\[
\text{Overall Detection Score} = \text{Identification Rate} - (5.0 \times \text{Error Rate})
\]
Analyzing a Live Call Log Sample

Step 1: Run the test
Gather a fresh sample of call tags from each of the competitive vendors within a tight time window (ideally within one hour), and then submit immediately for subsequent analysis.

Step 2: Assess the Agreement/Disagreement of Unwanted Status
It would be easier to evaluate an unwanted call detection service if the carrier had a “truth reference” as a reference for whether a call is unwanted or not. Of course, such a file does not exist, so you need to determine the truth yourself.

Let’s assume that you are evaluating two different providers. First, you will need to consider the following for each calling phone number:
● For which numbers did provider A and B both agree that a call was unwanted? (assume both providers were correct)
● For which numbers did provider A and B disagree? (only one provider can be correct)

For a sample of 100,000 unwanted phone calls detected by either provider, this can be visualized as follows:

<table>
<thead>
<tr>
<th>Provider A “Spam” Set</th>
<th>Provider B “Spam” Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-only Spam: 40,000</td>
<td>Both Spam: 30,000</td>
</tr>
<tr>
<td></td>
<td>B-only Spam: 30,000</td>
</tr>
</tbody>
</table>

Image 1: Example of unwanted call detection results.

Step 3: Determine the “truth reference” (which provider was correct)
Where the providers disagree, to determine who was right, a call-down or online research needs to be conducted. This will contribute to your “truth reference” file. This should be completed within 3-4 hours of when the sample was gathered.

Note: Establishing the truth reference is the single most critical step in the evaluation of the call detection solution. Without doing this, the rest of the measured metrics are meaningless.
From this process, each number can then be classified as likely spam, not spam, or uncertain. Because it is not reasonable to confirm every case of disagreement, we recommend taking samples of at least 500 random numbers from each side.

With these results, it is then possible to make assumptions about the entire sample. For example, if for Provider A the call-down found 20% were not spam and 50% were spam, those percentages can be applied to the entire set of Provider A-only unwanted results (20% x 40,000 not spam, 50% x 40,000 spam). The same process is applied to Provider B.

With this on hand, a final spam Identification Rate and Error Rate calculation can be made.

**Step 4: Calculate the Identification Rate**

For provider A, we now have a complete count of unwanted calls from the test. An Identification Rate can be calculated by comparing the spam calls that were correctly identified by a provider against the total count of all known unwanted calls from the test.

In this example, Provider A has 77% Identification Rate (50,000 / 65,000) and Provider B has 69% Identification Rate (45,000 / 65,000).

**Step 5: Calculate the Error Rate**

In order to determine the Error Rate, divide the not spam call-down results (aka “false positives”) by the total amount.
In our example, Provider A has a 14% Error Rate (8,000 / 58,000) and Provider B has a 6% Error Rate (3,000 / 48,000).

**Step 6: Calculate the Overall Detection Score**

In order to calculate the overall net benefit of the service, we utilize this formula:

\[
\text{Overall Detection Score} = \text{Identification Rate} - (5.0 \times \text{Error Rate})
\]

In this example, although Provider A has a higher Identification Rate (77% vs. 69%), due to the impact of a higher Error Rate (14% vs. 6%), Provider B is the more effective solution overall.
Appendix: Adding a Third Provider to the Competitive Evaluation

The methodology for evaluating three providers in a bake-off scenario remains the same.

To determine the “truth reference” set of spam numbers, we require that Provider A, B, and C all agree that a call was unwanted.

And when providers A, B, or C disagree, we still need to conduct research to determine who was right.

The calculation of the Identification Rate remains the same in principle, except that the denominator now also includes the correctly identified spam calls for Provider C. The calculation of the Error Rate for each provider remains unchanged.