February 2021

INTELLIGENT MECHANISM FOR ROBOCALL PREVENTION USING ARTIFICIAL INTELLIGENCE-DRIVEN INTERACTIVE QUESTION AND ANSWER TECHNIQUES

Anupam Mukherjee

Faisal Siyavudeen

Sanjay Sinha

INTELLIGENT MECHANISM FOR ROBOCALL PREVENTION USING ARTIFICIAL INTELLIGENCE-DRIVEN INTERACTIVE QUESTION AND ANSWER TECHNIQUES

AUTHORS:
Anupam Mukherjee
Faisal Siyavudeen
Sanjay Sinha

ABSTRACT

A robocall is a phone call that uses a computerized auto-dialer to deliver a pre-recorded message, as if from a robot. Robocalls are often associated with political and telemarketing phone campaigns. Some robocalls use personalized audio messages to simulate an actual personal phone call. Criminals and dishonest robocallers often alter or spoof the calling number of their outbound telephone calls in order to deceive a called party. This deception increases the chance that the called party will answer a robocall. In other cases, the spam calls may be more malicious. This proposal provides an artificial intelligence (AI)-based interactive and intelligent technique that utilizes a voicebot to detect spam calls by analyzing various answer/response patterns (or contexts) of callers in order to add such callers to a deny list without any user intervention.

DETAILED DESCRIPTION

In 2019, people in the United States received, on average, 18 robocalls per month. Since this time, the number of such calls has risen by 35%, suggesting an increase in such robocall activity per month in U.S. households. As noted, robocalls are often associated with political and telemarketing phone campaigns. However, criminals and dishonest robocallers often alter or spoof the calling number of their outbound telephone calls in order to deceive a called party. In some instances, such alterations can merely include changing the calling number so that it appears that a known person is calling. This deception increases the chance that the called party will answer a robocall. In other cases, the spam calls may be more malicious.

Though blocklists and other regulations may stem some robocall activity, robocalls are still a problem in the U.S. In particular, spammers have become much more

sophisticated. Through the use of artificial speech resources, spammers can generate calls at a super-human scale without any live operator on duty.

To address these challenges, this proposal provides an AI-based unique, interactive, and intelligent technique through which a voicebot or interactive voice response (IVR) - based virtual assistant can be configured to detect spam calls by analyzing answer and response patterns (answer/response) of callers in order to add spam callers to a deny list without any user intervention.

As illustrated in Figure 1, the voicebot will be equipped with a list of predetermined Valid/Semi-Nonsensical/Nonsensical questions and weird or odd noises along with typical human and robotic answer-set/response-set data that can be input as training data to an AI-based model of supervised category.

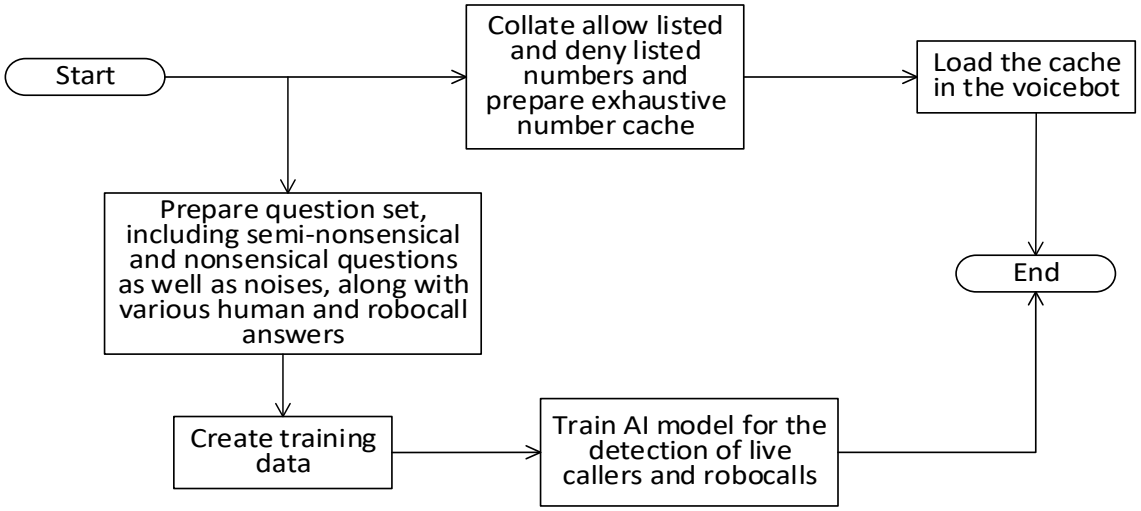

Figure 1: Data Preparation and Training

Once trained using the training data (context of the answer-set), the voicebot can then utilize the AI-based model to detect robocalls. Figure 2, below, illustrates example details associated with a workflow for detecting robocalls in accordance with the techniques of this proposal.

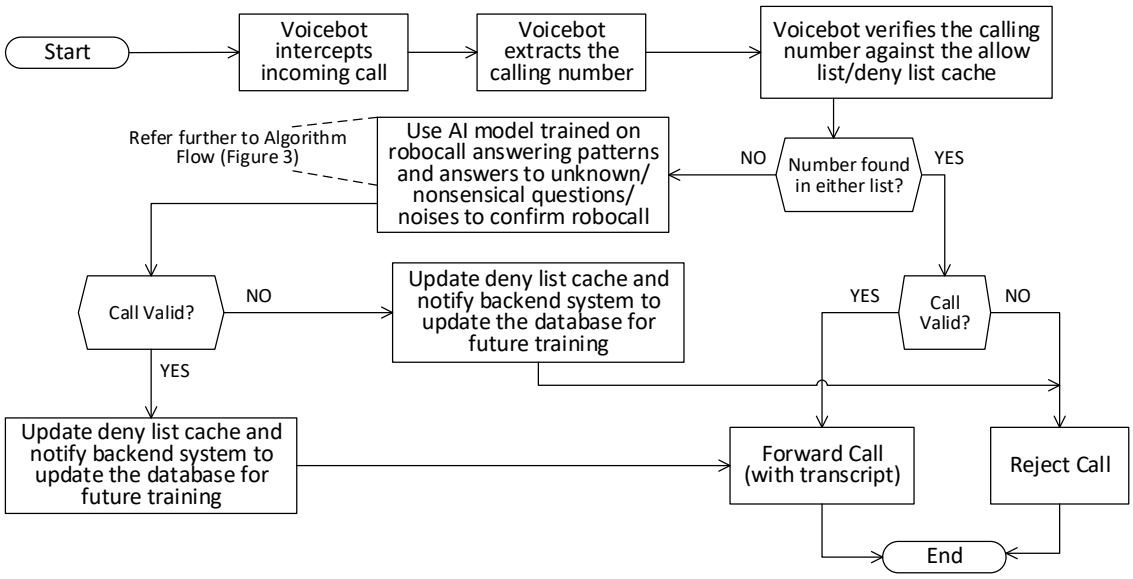

Figure 2: Voicebot Workflow

As part of its workflow, as illustrated in Figure 2, the voicebot can intercept an incoming call and verify the calling number against a precompiled caller-allow list and a caller-deny list. Here, in order to allow the emergency notifications or school related announcements, the caller IDs for such notifications/announcements should be added to the allow list as mentioned in the workflow. In this case, the calls, in spite of being robocalls, will be allowed to continue uninterrupted.

If the calling number is on neither list, the voicebot can ask one or more random (e.g., semi-nonsensical or nonsensical) questions and may even make odd noises (e.g., whistling, etc.) as a question. Reactions/responses given by the caller (including robocallers) will be captured by the voicebot and both syntactic and semantic aspects of the reactions/responses can be analyzed in real-time via the AI model within the context of the questions asked. The call will be classified as a valid or a spam call/robocall depending on the similarity between the caller's response context and the expected response context.

For the question inquiry portion of the workflow, the voicebot will try to determine whether an interaction is occurring with a robocall instead of an actual person; hence, the voicebot can attempt various 'tricks' such as making odd noises, etc. in combination and/or in lieu of posing unknown/nonsensical questions. For example, a live caller will undoubtedly reply in some manner to clarify (potentially with a tone of surprise) or ask a

follow-up question, whereas a robocaller may either reply in a stereotyped pattern with "I'm sorry, I didn't understand that" or keep on speaking based on the genre of the robot.

Here, a fundamental philosophy is that a human can detect both meaningful and nonsensical questions and respond differently based on the question asked. Although the choice of words during both scenarios can differ and may varies from person to person, answers/follow-up queries, nonetheless, will likely not deviate from the context of the questions. However, robots will likely respond in a stereotyped fashion, as they typically have a fixed dictionary of words that they use in all unknown scenarios/contexts. Accordingly, the solution provided herein will be trained to detect the same pattern and, thus, prevent robocalls.

Further as illustrated in the workflow of Figure 3, answers provided by a caller can be captured by the voicebot and text can be generated from the speech in real-time using a speech detection engine. Next, the answers can be analyzed in real-time (both syntactic and semantic aspects of the uttered words) by the AI model within the context of the asked questions and the call can be classified as a valid call or a spam call depending on the answer/response patterns. Additional details regarding the analysis is discussed below with reference to Figure 3. If a caller is classified as valid, the calling number will be added to the allow list and the call will be forwarded to the called number (optionally with a transcript). Otherwise, the call will be rejected and the number will be added to the deny list.

Figure 3, below, illustrates example details associated with the AI-based algorithm that may be utilized by the voicebot disclosed herein.

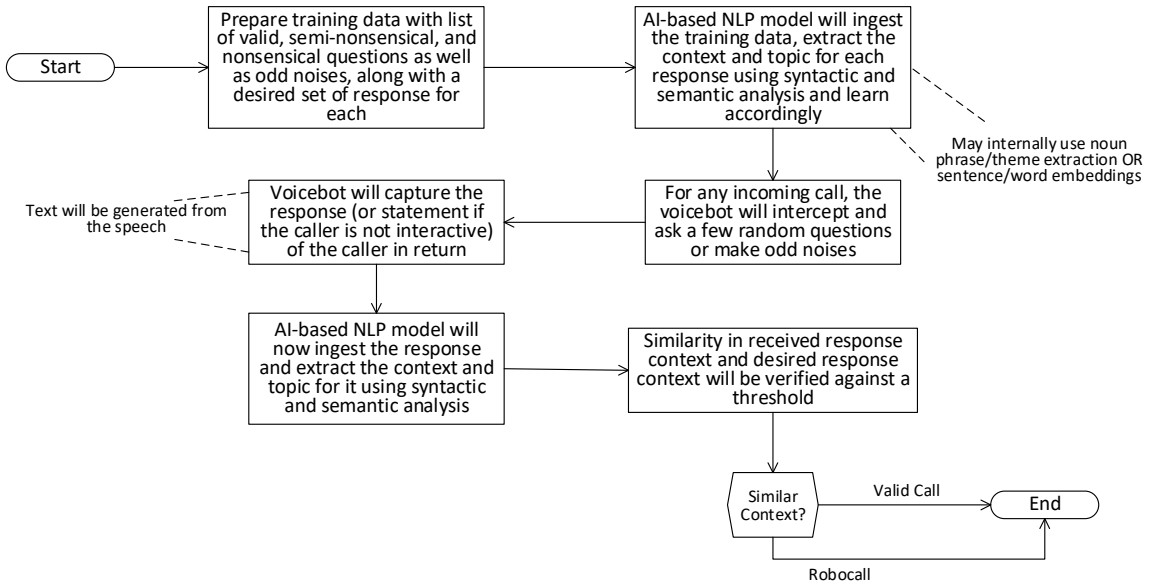

Figure 3: AI-based Voicebot Algorithm Workflow

As illustrated in Figure 3, the voicebot (e.g., IVR-based virtual assistant) can be equipped with the list of predetermined Valid/Semi-Nonsensical/Nonsensical questions and noises alongside their typical human and robotic answer-set/response-set as the training data. The voicebot will internally utilize an AI-based model (e.g., a Supervised model type) that will be trained on the training data.

As part of the training, the AI based model will learn the context (e.g., what is being discussed) for all the questions and their answers/responses using syntactic and semantic analysis. Internally, in one instance, the voicebot may use noun phrase extraction or theme extraction (with Relevancy Scoring using Lexical Chaining) to learn question and answer/response context. Alternatively, in one instance, the voicebot may create sentence embeddings (and word embeddings) from the sample answer-set/response-set of the questions and sounds.

During operation for any desired call, the voicebot can present to the caller a few random questions (e.g., semi-nonsensical or nonsensical) and/or may even make odd noise(s) (e.g., whistling) as a question utilizing the predetermined question-set/noise-set in order to determine whether the voicebot is interacting with a robocall instead or an actual person.

For instances involving a live caller, the live caller will undoubtedly reply in some manner to clarify (potentially, with a tone of surprise) or may ask a follow-up question in whereas a robocaller may either reply in a stereotyped pattern with "I'm sorry, I didn't understand that" or keep on speaking based on the genre of the robot. Here, the fundamental philosophy is that a human can detect both meaningful and stupid questions & respond differently. The choice of words during both scenarios can differ and/or vary from person to person. Nonetheless, the answers / follow-up queries will likely never deviate from the desired context of the ideal answers.

However, for instances involving robots, robots will typically respond in a stereotyped fashion as they typically have a fixed dictionary of words, which they are likely to use in all unknown scenarios/contexts/topics given that they are interactive. For example, if a robot is an announcer model robot, then it will not answer and keep on announcing. Otherwise, if the robot is an interactive model robot, then it will answer something like "Sorry, I could not understand. Please listen to the options carefully" and so on. The solution of this proposal will be trained to detect the same pattern to prevent robocalls.

Continuing with the flow illustrated in Figure 3, answers provided by the caller will be captured by the voicebot and text will be generated from the speech in real-time using a recurrent neural network (RNN), a speech to text mechanism (STT), automated speech recognition engine, or the like. Next, the answers will be parsed using natural language processing (NLP) (as part of syntactic and semantic analysis) and the respective context will be detected from the answers using noun phrase extraction or theme extraction (with relevancy scoring using lexical chaining), similar to the techniques as discussed above for the training. Alternatively, the NLP-based AI model can create sentence embeddings (and word embeddings) from the answer of the caller like, similar to the techniques as discussed above for the training.

Finally, similarity between the context extracted from the answer/response of the caller and the desired context of the answer/response will be verified in real-time using any standard semantic-based similarity detection algorithm, such as, for example, using cosine similarity of the vector Space, Bidirectional Encoder Representations from Transformers (BERT) -based sentence similarity, Jaccard similarity, etc. Based on the detected similarity

between both contexts (e.g., the caller's response and expected response), the call will be classified as a Valid Call or a Spam Call.

Robocalls have continued to become cleverer and can sometimes utilize AI to try to hide the fact that a machine is talking. According to recent studies, these machines are quite adept in their ability to provide answers to stereotyped questions (e.g., based on the type of call) in a matter of seconds. In some cases, these robocalls may try to sound like actual human beings by pausing in order to cause people to believe that they are, indeed, talking to a live agent. Hence, it is difficult to detect modern robocalls using standard mechanisms, such as a deny list number cache (as originating numbers can be spoofed), voice captcha/simple queries, or voice biometrics.

However, the solution provided herein overcomes limitations of these standard mechanism. For example, the model for the solution provided herein utilizes an interactive approach in which random questions (e.g., semi-nonsensical/nonsensical) can be presented and/or odd noises can be made and analysis of the answer/response patterns in the context of these questions can be performed to detect robocalls. Here, the fundamental philosophy is that humans can detect both meaningful and nonsensical/unknown/random questions and respond differently. Although the choice of words for both contexts can differ and/or vary from person to person, answers/follow-up queries will likely not deviate from the context of the questions.

However, robots will typically respond in a stereotyped fashion as they typically have a fixed dictionary of words, which they are likely to use in all unknown scenarios/contexts/topics given that they are interactive. For example, if a robot is an announcer model robot, then it will not answer and keep on announcing. Otherwise, if the robot is an interactive model robot, then it will answer something like "Sorry, I could not understand. Please listen to the options carefully" and so on.

The solution of this proposal will be trained to detect the same pattern to prevent robocalls, which will help to identify and segregate a robocall and filter such calls. It will be tough to predict the questions by the robocallers and, further, the questions and noises in our model will change with model updates.

For example, if any robo/spam calls can get through, the pattern of the answers will be identified and included in the training data to retrain the models (very low learning rate) for future spam call detection purposes.

In some instances, as illustrated in Figure 4, below models can be trained/retrained/tuned (hyperparameters) in a cloud architecture and the updated models can be pushed to the desired call control deployment from the same cloud using a cloud managed infrastructure.

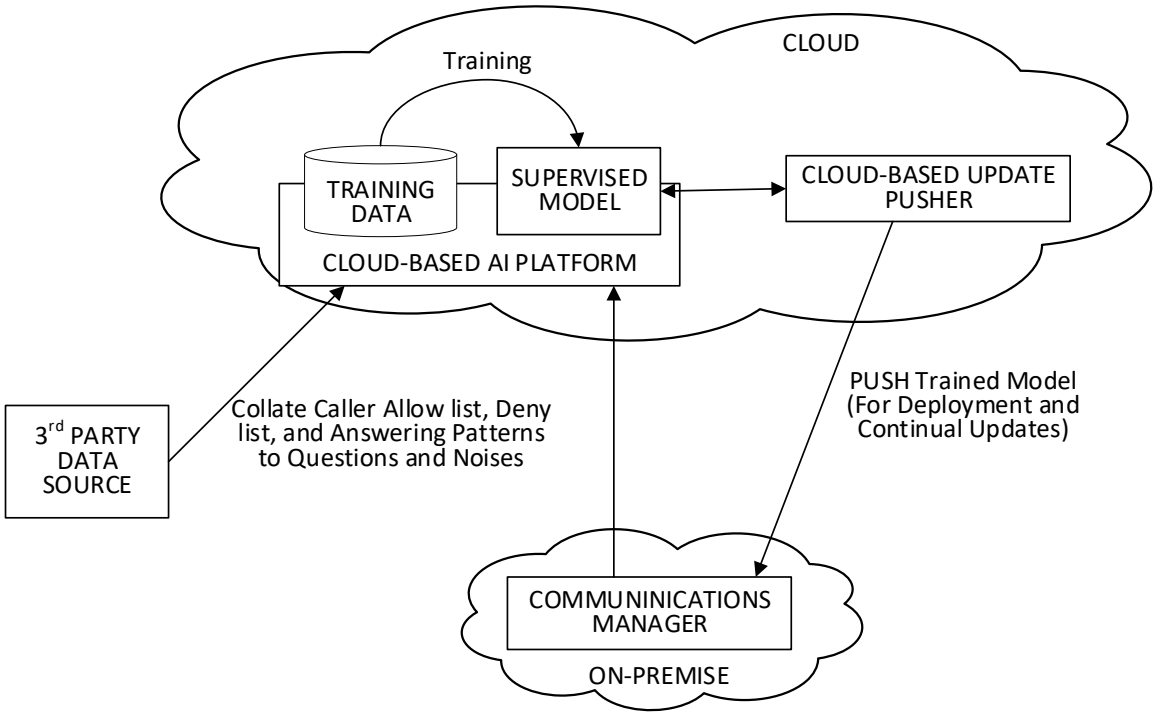

Figure 4: Example Cloud Architecture

With reference to Figure 4, query, noise, and answer-based training data can be refreshed/enriched regularly (using continual learning of the AI model) for maximum efficacy of the system. The proposed model can be seamlessly used in both Cloud based and On-Premise communications manager/management scenarios.

For example, one sample use-case as illustrated in Figure 4 involves the robocall detection model being trained in a cloud and being pushed to an On-Premise communications manager. In some instances, the model can be installed/upgraded as part of a container on an inbuilt docker farm within the communications manager. For any

incoming call, the voicebot will be activated to intercept the call (optionally and/or at random) and will use the aforesaid AI model to filter robocalls as explained above.

Accordingly, the solution provided herein may support a robust algorithm to facilitate any type of robocall detection using different question and answer patterns in which AI model driven voicebot will ask random valid/nonsensical/unknown questions and/or make odd noise(s) as a question and observe the answering/response pattern of callers in lieu of the accuracy of the answers. The voicebot will analyze both syntactic and semantic aspect of the uttered words and identify the caller as human or robot without any user intervention.

For scenarios in which robots may potentially be trained to overcome the features of this proposal, the dynamic/random query-set and noise set can be diversified to be continually updated. However, it will likely be difficult to train a robot to handle different variations of valid/semi-nonsensical/nonsensical questions and/or odd noises as robots have to be trained not only on answers but also on contexts, because the solution of this proposal does not detect the correctness of an answer. Rather, the solution detects the contextual similarity of answers, which is difficult to imitate for training a robot caller. Nonetheless, the query, noise, and answer-based training data can be refreshed/enriched regularly to facilitate maximum efficacy of the solution provided herein.

In summary, the solution herein provides a unique, interactive, and intelligent AI-based mechanism in which a voicebot can be utilized to detect robo/spam calls by analyzing the answer/response patterns (or contexts) of callers and add the callers to a deny list without any user intervention. The voicebot will be equipped with a list of predetermined valid/semi-nonsensical/nonsensical questions and odd noises alongside typical human and robotic answer-set/response-set as the training data. The voicebot will utilize an AI-based model that will be trained on the training data (e.g., context of the answer-set).

As part of its workflow, the voicebot will intercept an incoming call and verify the call against a precompiled caller-allow list and a caller-deny list. If the calling number is on neither list, the voicebot can ask one or more random (e.g., semi-nonsensical or nonsensical) questions and/or can make odd noises (e.g., whistling) as a question. The reactions/responses given by the caller (including robocallers) will be captured by the

voicebot and analyzed (for both syntactic and semantic aspects) via the AI-based model in real-time within the context of the asked questions and the call will be classified as a valid or a spam call depending on the similarity between the caller's response context and expected response context.