August 30, 2018

Distinguishing Bots from Human Callers

Ori Kabeli

Nadav Bar

Benny Schlesinger

Distinguishing bots from human callers

ABSTRACT

Telephone users frequently receive unwanted calls from sales, advertising, or spam callers. Such calls may be made by automated agents or bots of such sophistication that they are often difficult to distinguish from human callers.

This disclosure presents machine-learning techniques that enable differentiation of bots from human callers. Machine-learning models are trained to recognize artifacts that distinguish bot callers. Users can report callers as bots or humans, thereby enabling federated learning of differences between human callers and bots. A suspected bot caller is challenged with audio or visual captchas to further filter out bots. An incoming call that is confirmed as bot-initiated is either not delivered to the user, or the call recipient is alerted that the caller is likely a bot.

KEYWORDS

- robocaller
- bot caller
- spam call
- audio artifact
- visual artifact
- captcha
- spam prevention
- bot detection

BACKGROUND

Telephone users frequently receive unwanted calls from sales, advertising, or spam callers. Such calls may be made by automated agents or bots of such sophistication that they are often difficult to distinguish from human callers. Bot callers have expanded their presence from traditional telephone or mobile phone calls to VoIP, video telephony, messaging services, etc. Identifying callers as bots enables such calls to be filtered out, or for call recipients to be informed that that are likely communicating with a bot.

DESCRIPTION

Per techniques of this disclosure, detection of bots within a voice or video communication channel is divided into passive or active techniques, each of which is described in detail below.

Passive techniques for bot detection

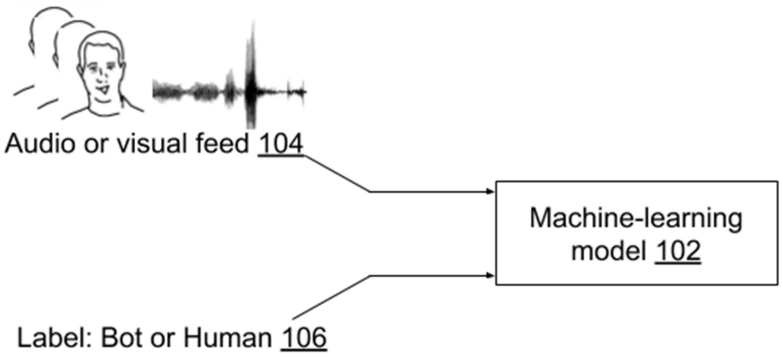

Fig. 1: Training of machine-learning model to distinguish bots versus humans

Bot-generated speech or video is known to possess certain features or artifacts, e.g., repetition of words, repetition of frames, visual artifacts between frames, unusual timing of filler phrases (e.g., “umm,” “ah,” etc.), non-idiomatic constructions, etc. As illustrated in Fig. 1,

a machine-learning model (102) is trained on labeled samples (106) of bot-generated and human-generated speech and/or video (104). Passive techniques include applying the trained machine-learning model to an ongoing conversation to identify synthetically generated voice or video.

The machine-learning model can include, e.g., regression learning models, neural networks, etc. Example types of neural networks that can be used for the classifiers include long short-term memory (LSTM) neural networks, recurrent neural networks, convolutional neural networks, etc. Other machine learning models, e.g., support vector machines, random forests, boosted decision trees, etc., can also be used.

In addition, the machine-learning technique of federated learning can be used advantageously to distinguish bots from human callers. In federated learning, a machine-learning model resides locally on the user's device. Learnings from several devices are transmitted, with the permission of respective users, to a server-based machine learner. The transmitted data does not include actual speech, video or call data, or features thereof. Rather only changes to the local machine-learning model, e.g., changes to neural network weights, are transmitted to the server.

Federated learning thus enables devices to collaboratively learn a shared prediction model while keeping training data on device. The collaborative learning process is carried out using distributed storage of training data. Under federated learning, whenever human users mark callers as bots, training occurs on their device and the local machine-learning model is updated. Then the deltas from the models on different user devices are sent to the server for aggregation. After the aggregation, the new server model includes the learning from all the users and is sent to individual user devices. In a similar manner, a human user can reclassify

callers initially misclassified as bots. In this manner, the techniques of this disclosure fight spam or bot-callers in a scalable manner.

Active techniques for bot detection

Active techniques for bot detection include challenges that are sent to a caller in order to verify the caller as a human or a bot. Such challenges include, for example, asking the caller to dial something on the keypad; asking the caller to say a specific word or sentence; asking the caller to answer a specific question about the call recipient to prove that caller knows call recipient; asking the caller to solve a problem that is easily solved by a human but not by a bot; etc. Challenges for a video call can include, for example, asking the caller to move in a certain way; asking the caller to show their hand or other body part; presenting the caller with a standard visual captcha over the video call; etc.

In both active and passive techniques for bot detection, identification of the caller as bot or not is automatic, without intervention from the human call recipient. Analysis of the incoming phone/video call, captchas issued to the caller, etc., is automatically performed by receiving device, e.g., a mobile phone or other device of the call recipient. An incoming call confirmed as bot-initiated is either filtered out, or the call recipient is alerted that the caller is a bot.

The techniques of this disclosure make the cost of operating a bot to place calls higher than the value derived from operating it, thereby reducing the motivation of bot operators to deploy bots.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network,

social actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure presents machine-learning techniques that enable differentiation of bots from human callers. Machine-learning models are trained to recognize artifacts that distinguish bot callers. Users can report callers as bots or humans, thereby enabling federated learning of differences between human callers and bots. A suspected bot caller is challenged with audio or visual captchas to further filter out bots. An incoming call that is confirmed as bot-initiated is either not delivered to the user, or the call recipient is alerted that the caller is likely a bot.

REFERENCES

1. Harry Pettit, "Wearable anti-AI AI' detects fake voices," *Daily Mail Online*, 29 May 2017, www.dailymail.co.uk/sciencetech/article-4551862/Wearable-Anti-AI-AI-detects-fake-voices.html
2. Twilio, "Answering machine detection." https://www.twilio.com/docs/voice/answering-machine-detection