

УДК 004.042

О пределе обнаружения ботов по временным меткам

Калабкин Е.С.¹, Мамедова Т.Ф.¹

Национальный исследовательский Мордовский государственный университет¹

Аннотация: В статье методами имитационного моделирования исследуется устойчивость алгоритмов обнаружения ботов в системах обмена сообщениями к усложнению их поведения. Пользователь описывается самовозбуждающимся точечным процессом Хоукса. На его основе построена иерархия из шести уровней маскировки бота. Численный эксперимент даёт кривую деградации детектора, выявляет парадоксальный рост обнаружимости при более полной имитации и устанавливает принципиальный предел качества индивидуальной классификации по временным меткам.

Ключевые слова: имитационное моделирование, точечный процесс Хоукса, обнаружение ботов, метод максимального правдоподобия, адверсарная адаптация.

В мессенджерах обнаружение ботов затруднено: содержание переписок и графы связей недоступны по соображениям приватности, поэтому единственным источником данных остаётся последовательность временных меток событий пользователя. Цель работы – методами имитационного моделирования оценить, как качество детектора зависит от сложности маскировки бота и где лежит предел обнаружимости [1, 2].

Определены шесть уровней маскировки по возрастанию сложности:

L_0 – детерминированный интервал;

L_1^{CTOX} – равномерный случайный интервал;

L_1 – пуассоновский процесс ($\alpha = 0$) с человеческой частотой;

L_2 – с циркадной модуляцией;

L_3 – полная имитация (Хоукс и циркадный ритм).

Параметры L_3 -бота берутся из распределений, характерных для людей, поэтому он статистически не отличим по одиночным признакам.

Траектории генерируются методом обратного преобразования. Параметры по временным меткам оцениваются методом максимального правдоподобия (рекурсия Озаки) и формируют вектор из 18 признаков: оценки μ , α , β статистики межсобытийных интервалов и характеристики циркадного ритма. Классификация осуществлена ансамблем решающих деревьев (Random Forest), качество F_1 – мерой при разбиении 70/30 (5 прогонов). Генератор и детектор реализованы на Go; эксперименты воспроизводимы.

Численно построена кривая деградации детектора (рис. 1): $F_1 = 1.00 (L_0, L_1^{CTOX}) \rightarrow 0.88(L_1) \rightarrow 0.74(L_2) \rightarrow 0.78(L_3)$. Самое резкое падение наблюдается при включении ботом циркадного ритма ($L_1 \rightarrow L_2$).

Обнаружен парадокс: более совершенная имитация (L_3) не снижает, а слегка повышает обнаружимость (0.78 против 0.74)

Парадокс объясняется не свойствами бота, а асимметрией распределения признака между классами: у L_3 -ботов самовозбуждение включено всегда, у людей – лишь у половины, и эта разница сама становится различимым признаком.

Таким образом, для индивидуального детектора по временным меткам существует потолок разделимости: F_1 выходит на плато 0.74–0.78 уже на L_2 и не растёт при

дальнейшей имитации.

Потолок принципиален – бот с человеческими параметрами неотличим в индивидуальном пространстве признаков, и его расширение (новые статистики, нейросети) потолок не поднимает.

Преодоление возможно лишь через признаки, зависящие от других пользователей, что обосновывает переход к каскадной архитектуре обнаружения.

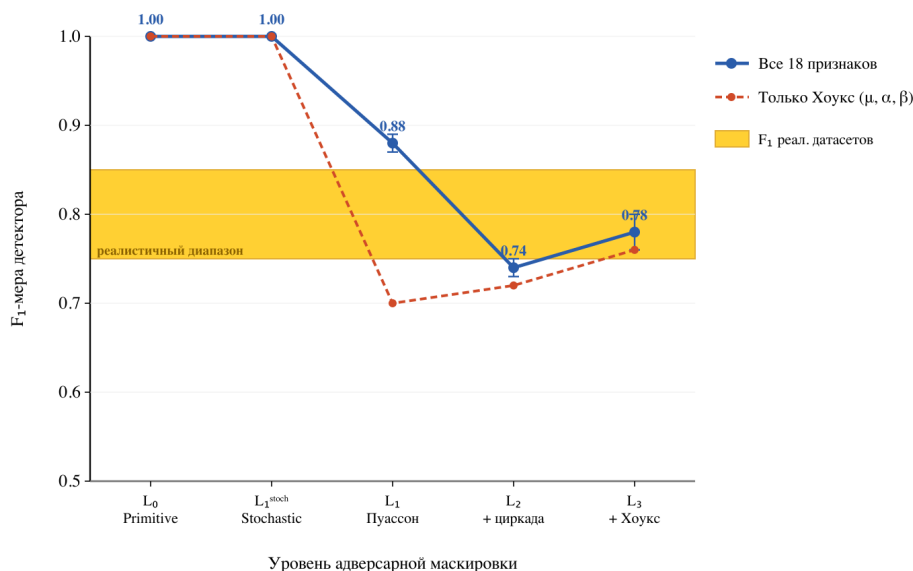


Рис. 1. Кривая деградации детектора по уровням адверсарной маскировки.

Полученная зависимость сведена воедино на рис. 1. График наглядно показывает, что качество детектора падает при усложнении маскировки и уже на уровне L_2 выходит на плато 0.74–0.78, не восстанавливаясь при дальнейшей имитации.

Литература

1. Cresci S. A decade of social bot detection // Communications of the ACM. 2020. Vol. 63, No. 10. P. 72-83. URL: <https://www.sci-hub.ru/10.1145/3409116>.
2. Cresci S., Di Pietro R., Petrocchi M. et al. The paradigm-shift of social spambots // Proc. 26th Int. Conf. on World Wide Web Companion. 2017. P. 963-972.

MSC C94A99

On the Detection Limit of Bots Based on Timestamps

E.S. Kalabkin¹, T.F. Mamedova¹

National Research Mordovia State University¹

Abstract: Simulation modeling is used to study the robustness of bot-detection algorithms in messaging systems against increasing behavioral sophistication of bots. A user is modeled as a self-exciting Hawkes point process, on top of which a hierarchy of six adversarial masking levels is constructed. A numerical experiment yields the detector degradation curve, reveals a paradoxical increase in detectability under more complete imitation, and establishes a fundamental ceiling on the quality of individual classification based on event timestamps alone.

Keywords: simulation modeling, Hawkes point process, bot detection, maximum likelihood estimation, adversarial adaptation.