You Are a Game Bot!: Uncovering Game Bots in MMORPGs via Self-similarity in the Wild

MAR 14TH, 2016

论文下载

这篇论文介绍了使用 self-similarity 来衡量一定时间内用户重复某一动作的频率，以此来检测出Game Bots.
作者及其团队从2013年2月就开始做这个系统，并在韩国三个比较有名的网络游戏上进行部署，检测出超过15000个 Game Bots。

概述

通过游戏记录，系统使用 Self-similarity 来检测Game Bots。
Self-similarity 是记录一段时间内用户动作相似度的特征。
贡献：

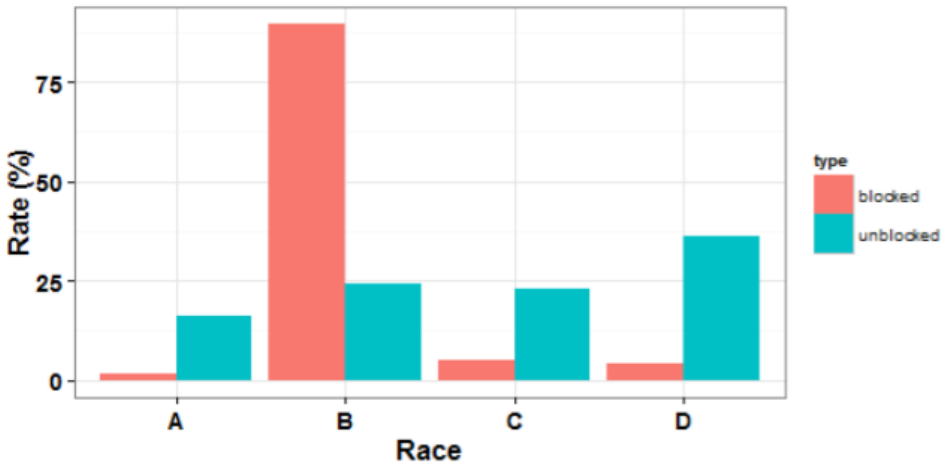
- 不同层面的数据分析
- 提出了 self-similarity 特征

- 建立一个检测系统，可以动态自动适应Game Bots的变化
- 在2015年2月完成了系统，并应用于“Lineage”。系统检测出了超过15000个Game Bots。系统不久之后应用于“Blade & Soul” 和 “Aion”，是第一个应用于多个游戏的bot检测系统。

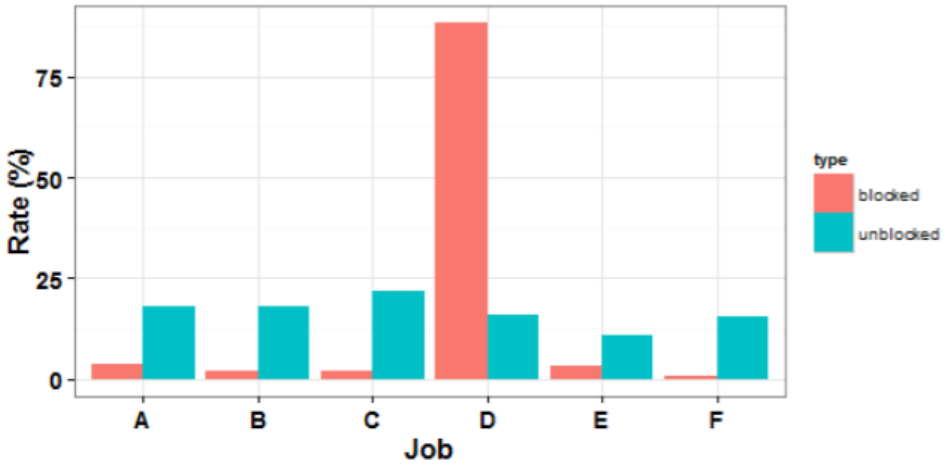
GAME BOT 特征

一些可以区分出Game Bots的特征，但并不在每个游戏中都适用

- 选择职业的倾向



(a) race



(b) job

Fig. 2. Comparison of demographic data between bots and humans' characters in B&S.

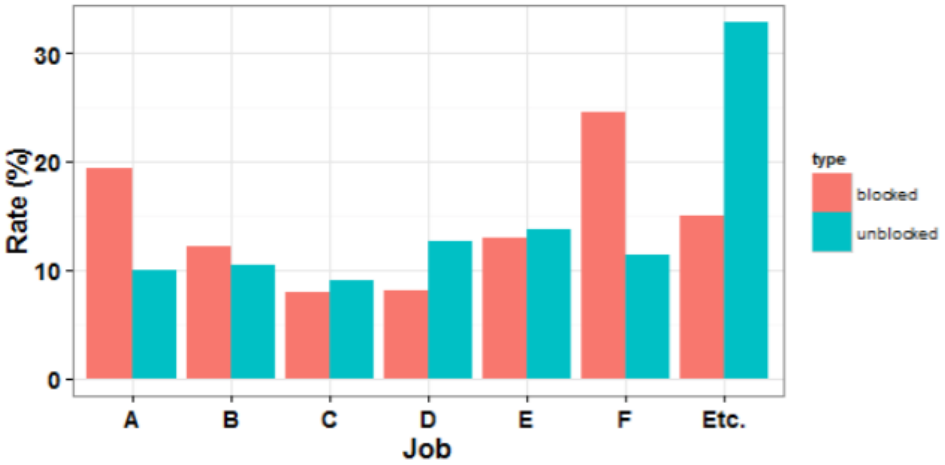
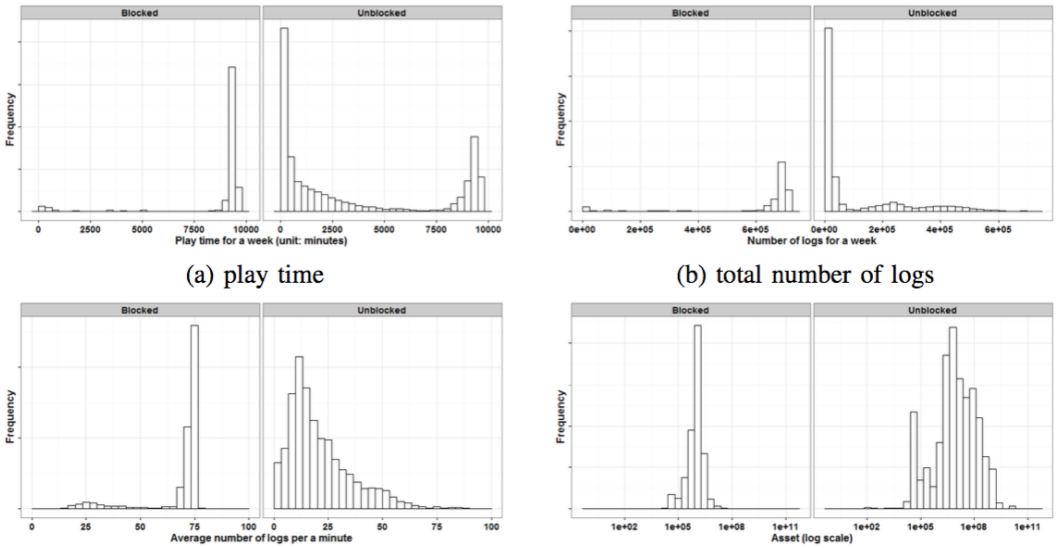


Fig. 3. Comparison of job distribution in Aion.

- 游戏时间、生成的log数量

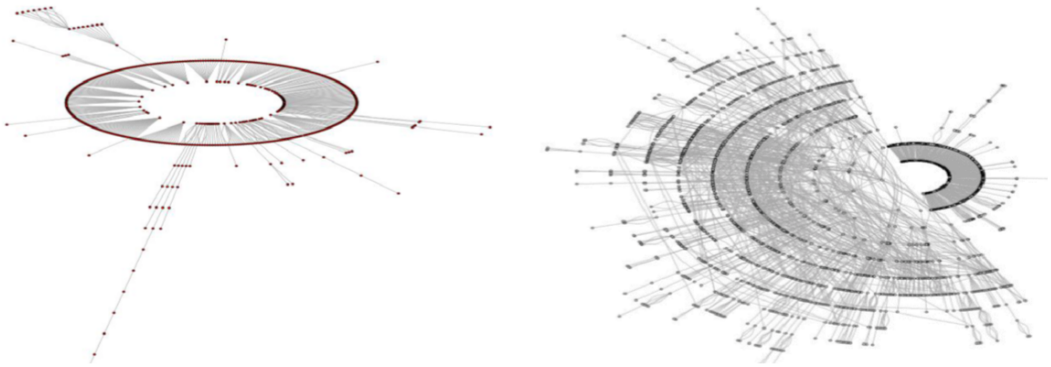


(a) play time (b) total number of logs

(c) average number of logs per unit time (d) asset

ig. 4. Comparison of activity statistics between blocked and unblocked users: left box is blocked (bots), right is unblocked (normal users).

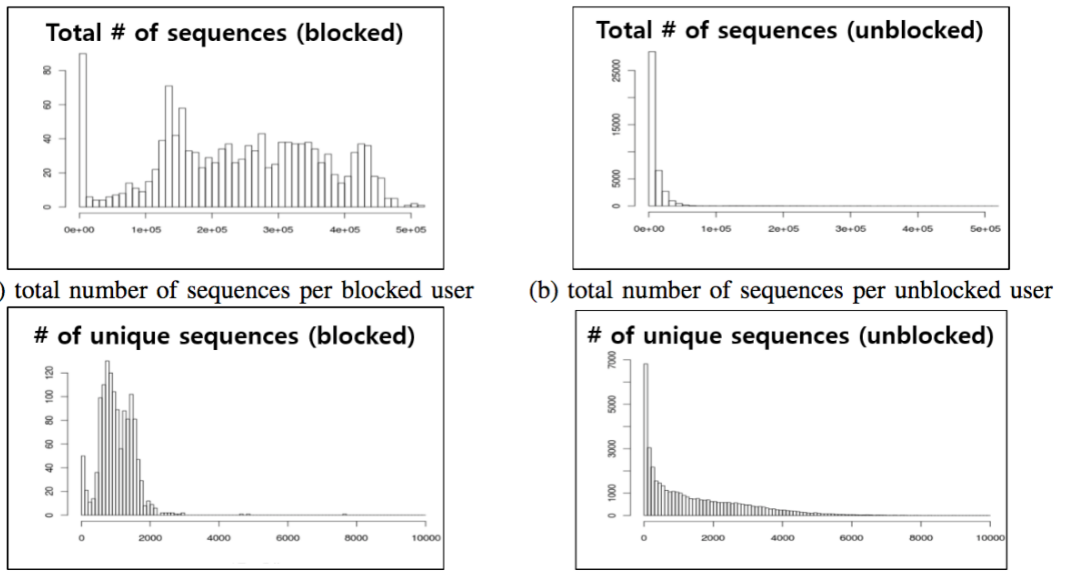
- 人际关系



(a) blocked users (b) unblocked users

Fig. 5. Network structures of item trade: (a) trade with blocked users, (b) trade with unblocked users.

- sequence data



(a) total number of sequences per blocked user (b) total number of sequences per unblocked user

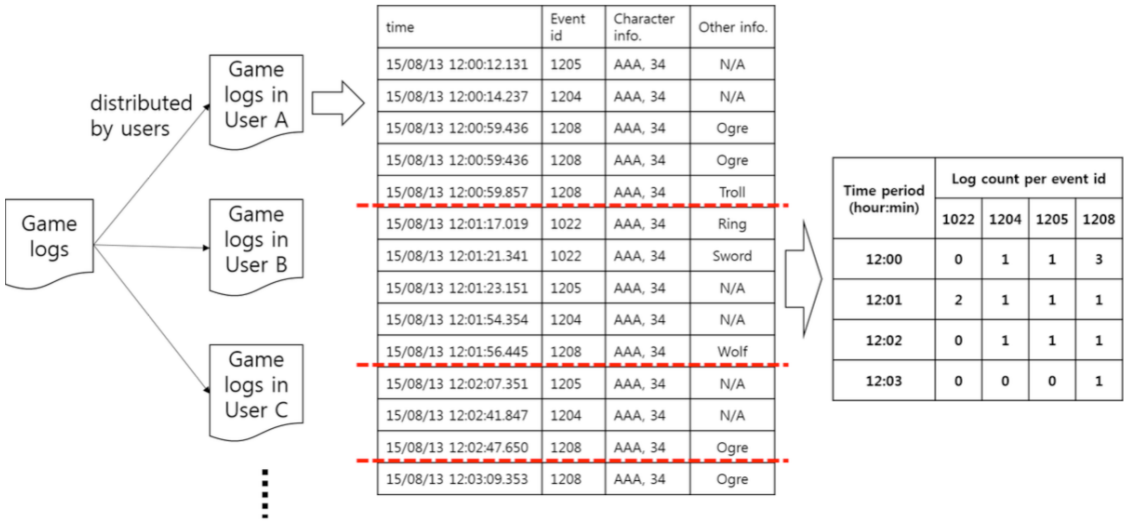
(c) number of unique sequences per blocked user (d) number of unique sequences per unblocked user

Fig. 6. Histogram of sequence data.

方法

系统使用 self-similarity 作为特征，并使用机器学习的方法进行training

- 从 log 中提取有用的信息，进行整理，记录相应的动作次数



g. 11. Process for transforming game logs to vectors. In the above, four vectors are generated: $(0,1,1,3),(2,1,1,1),(0,1,1,1),(0,0,0,1)$.

- 通过实践总结出的有价值的特征

TABLE II. LIST OF FEATURES USED IN THE PROPOSED DETECTION METHOD.

No.	Feature	Description
1	self sim.	self-similarity
2	vector count	count of a set of log vectors
3	uniq. vector count	unique count of a set of log vectors
4	cosim. zero count	count of data in which cosine similarity is zero in a set of log vectors
5	vector mode	count of data that appears most often in a set of log vectors
6	total log count	total count of logs generated by user
7	char. level	character level
8	play time	play time during certain period per user
9	npc kill count	NPC kill count
10	trade take count	count of trade in which user takes items from another
11	trade give count	count of trade in which user gives items to another
12	retrieve count	count of activity in which user retrieves items from warehouse
13	deposit count	count of activity in which user deposits items to warehouse
14	log count per min.	average count of logs are generated per minute

- 进性training，若一段时间内bot发生变化，则重新进行training

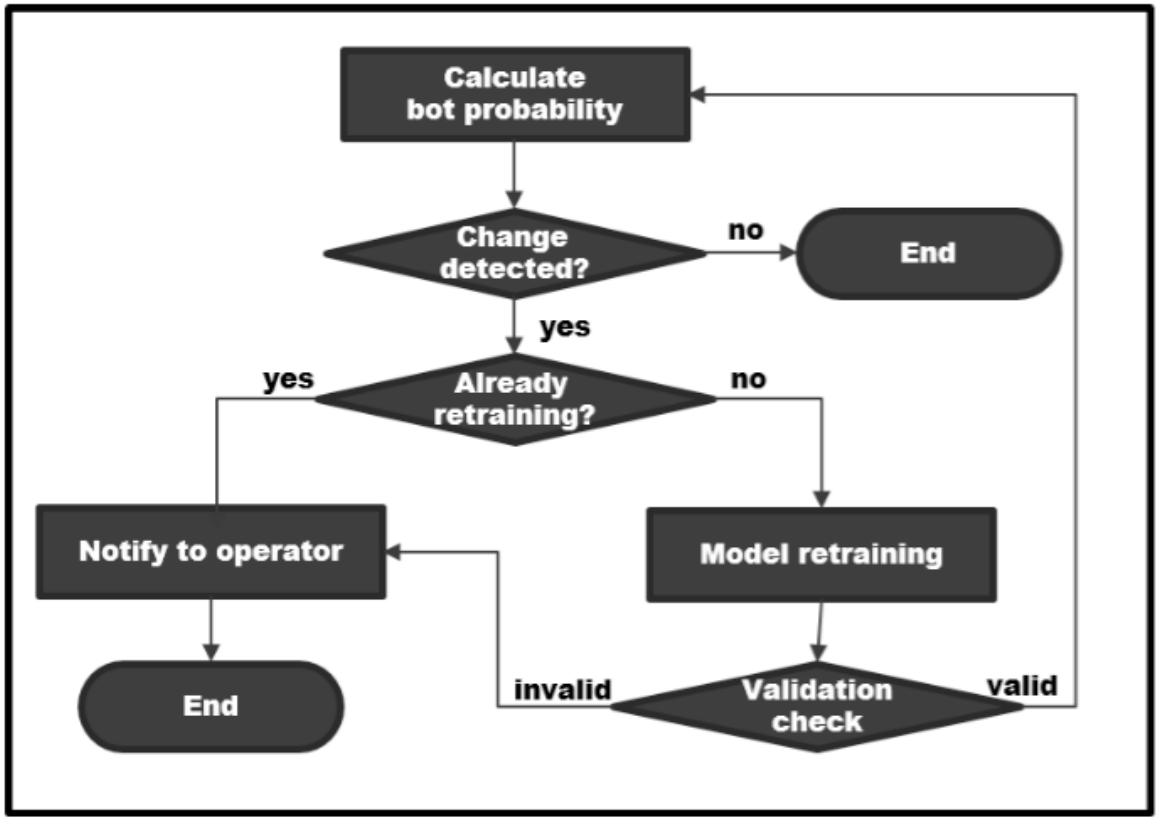


Fig. 12. Automated maintenance process.

实验

机器学习的各种参数都是经过大量实验而确定下来的。

系统应用于实际的网络游戏中，检测出了超过15000个Game Bots。