Big Data and Bot Detection: BD + BD = BFF?

Esther Seyffarth

June 2, 2016

Abstract

This paper deals with the problems posed by bots pretending to be human in the context of video games, social media, and review websites. Since these services are made for human consumption and enjoyment, injecting bots into these situations is only profitable to their operators, and harmful to human users of the services. We present an overview of the most widespread problems caused by bots on the web and discuss a range of strategies for bot detection from recent literature. Specifically, we approach the subject of bot detection from a Big Data perspective, and explore how analyzing big data about the way humans normally use and interact with a given service can help identify non-human users.

1 Introduction

The analysis of big amounts of user-originated data allows insights into average behaviours or usage patterns, but can it tell us anything about individual users? After all, noticing trends in the mass of information that is being analysed does not always give useful insights about individual data points. However, there is one question that might be answered using knowledge about all users of a service: The question whether a single user account is controlled by a computer program, often referred to as a bot. The intuitive reasoning is that since the bulk of all usage data — it is difficult to estimate the exact percentage — is human-generated, bots can be recognized from their deviation from the norm in the given context.

Bots are not a new problem, but have been one for a long time as the internet developed into its current form. The most widely known type of bot is the spambot, which disrupts human communication by posting unsolicited messages in otherwise human-dominated areas (such as email, IRC or discussion forums), with the goal of making users click on links to boost ad revenue or even infect their devices with malware. In the early days of the internet, the costs of maintaining a network connection were so high that bots led to economic problems for some providers by simply posting these meaningless messages ([Seewald and Gansterer, 2010]).

This paper aims to explore the possibilities of detecting bots in the context of services and platforms that are made for human use and generate big amounts of data. Three of the most well-researched areas of this topic in recent years are video games, social media platforms and rating websites. In the following sections, we hope to answer the following research questions, based on current literature on bot detection in these application contexts:

1. How harmful are bots in the context of games, social media, and rating or review websites?
2. How can Big Data be utilized to identify harmful bots in these different contexts?
3. Which problems arise from these Big Data approaches and what are their limits in each specific area?

The remainder of this paper is structured as follows. Sections 2 to 4 will each elaborate on one of the contexts mentioned above and attempt to answer the research questions for the current context. Since bots can be harmful in different ways in each of the contexts, the possible detection techniques and pitfalls also differ, and will be commented on in each of these sections. Section 5 provides a summary of the current state of research in the area of bot detection and comments on the general usefulness of Big Data in this area.

In each of the three situations referred to throughout this paper, there is a threshold for the amount of bot interaction that is acceptable before the service or website becomes completely unusable. It makes no sense to play a difficult game when bots can gain levels and collect loot much more efficiently than a human can, and do so at such a high rate that humans have no chance of catching up. It does not seem wise to create an account for a dating platform when the odds of meeting another human being on it are infinitesimally small due to a high bot population. Rating systems that can be sabotaged lose their purpose of comparing products and taking other buyers' opinions into account, and become meaningless.

Therefore, each of these types of services is in need of a routine to detect bots and minimize the harm done by them. Since most bot detection techniques in any context can be tricked, it would be utopian to believe a bot population could be lowered to zero. In the relevant literature, the race between developers of bot detection techniques on the one hand and bot creators on the other hand is sometimes referred to as a "game of cat and mouse". Therefore, all approaches mentioned in the following sections, while employing current technologies and achieving considerable success, are subject to change as needed to keep up with improved, less detectable bots as they are being developed.

2 Game Bots

2.1 How Bots Harm Video Games and Their Human Players

Online games have become so popular that a whole industry has evolved not only around developing them, but also around playing them. This has led not only to international gaming tournaments and video-streamed events, but also to a global real-world market for virtual-world goods. Especially in Massively Multiplayer Online Role Playing Games (MMORPGs), a player's character becomes stronger over time, according to the skill points earned and the items equipped that influence the character's stats. Items that grant or boost certain abilities of the character are valuable to the players because they have a significant influence on their enjoyment of the game; and since items can be traded between players in most MMORPGs, they bring the added benefit of personal economic gain when they are traded for real-world money. The same applies for trades of game money for real money.

This is a powerful motivation for players to cheat, or for people to start playing an MMORPG as a means to make money from selling in-game items or gold. The process of playing with a focus on collecting items or gold is known as farming. This task is so repetitive in nature that bots can easily be programmed to do it. For the bot owner, this has the advantage that they have to put less effort into playing the game and can transfer the loot from the bot character to other characters or players from time to time, which ensures a continuous stream of either money from people who buy the virtual items, or upgrades to their own characters if they also play the game for their own enjoyment.

However, the actions of bots have a direct influence on other, human players' enjoyment of the games. Because players play on shared servers, they also share the virtual resources of the in-game environment; and when bots spend all day and night farming for items, this means that fewer items are left for the human players to collect. Even the collection of gold is harmful to human players: Gold is often an infinite resource, and when bots collect big amounts of it, they can swamp the in-game item market and trigger an in-game inflation that makes it impossible for non-cheating players to keep up and be successful in the game.

Thus, bots in online games are generally considered to be harmful to human players, and by extension to the providers of the game, since humans who stop enjoying the game will stop paying for a subscription. On the other hand, owners of game bots can make easy money with them, and people who pay for items that were farmed by bots also profit from that transaction, as they do not have to spend as much time improving their in-game characters as they would have to otherwise.

2.2 Detection of Game Bots

One difficulty of detecting bots in online video games is that the only trace they leave is the protocol of their interactions in the virtual world. There is no textual "footprint", only a stream of actions an individual player character takes in a specific order over the course of a playing session. However, these protocols can be used as a basis for bot detection algorithms with high accuracy.

Depending on the type of game, the bots will display different behaviours. Current literature about bot detection in video games focuses on two types of online games: Role-Playing Games (RPGs), and First-Person Shooter (FPS) games. While operating bots in RPGs is much more profitable to bot owners due to their complex economies, and therefore much more prevalent than bots in FPS games, they do occur in those games as well, and many of the bot detection methods described in the literature can be applied to both types of games. The references cited in this section deal predominantly with RPGs.

Role-Playing Games place a strong focus on the continued development of a single player character over months or even years of real-time. Characters can collect skill points, refine their abilities, collect items that support different aspects of their playing style, and generally become stronger and more powerful the more they are being played. The difference between newly-created characters and well-established ones with a wide choice of equippable items is so big that players around the world trade in-game items for real-world money, with the sole motivation of making their character stronger with a smaller effort.

The problem about this economy that transcends the virtual world is that it can be tricked. Bots are being programmed to automatically collect game items and virtual gold in a much more efficient and less tedious way than human players could. Since this disturbs the in-game economic balance and thus presents a danger of driving away real, human players, it makes sense to try and identify bot accounts and disable them.

In order to do so, the most important question to ask is the following: How do bots behave differently from humans *in order to be more efficient*? After all, the advantage of bots is that they do not need breaks to focus, take care of their bodily needs, or sleep. Authors of recent papers about this utilize that fact for the development of their bot detection theories and frameworks.

A characteristic aspect of bot behaviour in RPGs is the trace of their movements through the virtual world. In [van Kesteren et al., 2009], several characteristics of player characters' movement are taken into account, in addition to the raw positional information. The authors propose a measure that combines data about the ratio of standstill time to movement time of player characters, the angles of their movement when they change direction, the pace and smoothness of their movements, and analyses of detours taken by the characters. The reasoning behind this is that bots will display a lower degree of randomness in their movements than humans do. Additionally, humans often take detours from their planned paths to react to new developments in their surroundings (for example, a monster approaching them); bots are either pre-programmed to take specific paths, or have a programmed response to some changes in the game environment. Both of these implementations will inevitably lead to regularities in their behaviour, showing that they are "too deterministic to be real". Using a leave-one-out cross-validation and a test data set of 25 human players and 25 bots, the authors were able to achieve 100% accuracy in their bot detection.

The notion that bots move too predictably to be convincing is taken even further by [Mitterhofer et al., 2009]. The authors describe a strategy of bot "training" that involves running a pre-determined path using a human-controlled character and having the bot memorize that path. Later, the bot can go back and forth on that path and kill any monsters it encounters, with the goal of collecting as much gold and as many items as possible. The assumption of pre-determined paths that bots take makes it extremely easy to detect them, as humans never repeat paths exactly the same way while playing. This method is a successful bot detection algorithm and needs 12 to 60 minutes of gameplay to analyse in order to identify recurring paths.

[Pao et al., 2010] also rely on the paths taken by player characters to distinguish between bots and humans, here focusing on FPS games. They employ many of the same measures used by [van Kesteren et al., 2009], and additionally make some observations about the way bots navigate small, restricted spaces. For instance, they notice that bots tend to move in the middle of a hallway or spend much time in the middle of an open space, since their path-calculating methods tell them that those are strategically useful positions — starting from there, they can move in any direction with the same amount of effort. This differs from human players in that human players would spend as little time as possible in open spaces without cover. Using this and other movement measures for an entropy model, the authors achieve an accuracy of 100%, provided the character paths that are evaluated are equal or longer than 1000s.

In some papers, bots are found to be detectable by their social behaviour. For example, [Kang et al., 2013] focus on the differences between playing parties consisting of humans and those consisting of bots. They define party play as a group of player characters combining their efforts temporarily to go on in-game missions or quests together. For human players, the advantage of party play is that they can defeat more powerful monsters and therefore gain more experience points and collect more items than they could by playing on their own. However, bots that form parties rarely do so to complete big challenges, and instead follow a routine fighting smaller monsters and completing more easy quests. As in the literature mentioned above, repetition is key to recognizing bots; parties that play in the same area for a long time and repeat individual quests are very suspicious and highly likely to be bots. Additionally, the authors find that parties with a duration of 12 hours or more are almost certainly made up of bots, since humans need breaks from playing. By calculating weights for several different features centering around the analysis of party play in MMORPGs, the authors were able to achieve 95.92% accuracy in detecting bots.

A different approach that also analyses social behaviour is presented in [Lee et al., 2016]. They argue that bots have a more structured, hierarchical routine for in-game trading of virtual items than humans do, since this type of system allows the bots to generate the highest profit from their work. Humans often trade outside their preferred social group, while bots are more likely to keep items among their own small group. The authors also say that other features of the social network structure in MMORPGs would be useful for bot detection; however, they choose not to analyse those features because this would take more computing power than was available to them at the time. This study focuses on a calculation of player characters' self-similarity by evaluating the patterns found in their in-game actions, achieving an accuracy of 93.99 to 99.42% (depending on the game).

2.3 Problems With Bot Detection in Games

The results presented in the literature surveyed in section 2.2 suggest that it is very easy and straightforward to detect bots in online games. This may be so; but only over a certain period of time, as the protocols of player characters' movement need to be evaluated after the fact, and a certain amount of data is needed to calculate a reasonably accurate probability whether or not the player is, in fact, a bot. This means that the bots that cannot be identified yet are able to do some damage before they are detected, which is difficult to undo. One possible solution to this may take the shape of immediately annulling a bot character's achievements and emptying their accounts of all virtual loot collected during their unidentified time, in order to minimize the gain for the bot's owner. However, a bot can conceivably be programmed to continuously offer its loot or gold for sale even while playing, which might in some games be so successful that the owner does not need to run a bot for more than ten or twenty minutes at a time. In some cases, a human player with repetitive playing habits might be misidentified as a bot, and taking away all of their in-game achievements will not be met with much patience.

Another problem of the bot detection approaches discussed in the referenced literature is that most evaluations depend heavily on knowing what type of game the data comes from, how the architecture of the virtual world is shaped, and which interactions of players with the world are possible. This means that it is extremely difficult or even impossible to design an algorithm that can deal with bot detection in all types of games. Games are too different from each other, which means that the strategies employed by bots differ as well. Game server operators can only hope to minimize the damage done by bots by dealing with them quickly and efficiently as soon as they are detected, and by continuously adapting their detection systems to any changes in strategy that the bot owners come up with.

3 Social Media Bots

3.1 How Bots Harm Social Media Platforms and Their Human Users

The category of social media bots may be the least homogeneously harmful group of those discussed here. At the more dangerous end of the spectrum, there are bots that lure other users to click on their "homepage" links to infect the users' devices with malware or trick them into entering their credit card information or sensitive passwords; at the opposite end of the spectrum, there are assistant bots, or bots that are developed as art projects to entertain and amuse human users and have neither an intention nor even the ability to do any harm.

This makes social media a context in which bot detection is more difficult to implement than in the other contexts listed here. After all, it is up to human intuition to decide whether a bot that posts an automatically-generated poem every hour should be considered spam or not. Unfortunately, the question of how the difference between harmful and harmless social media bots can be recognized automatically is beyond the scope of this paper.

As an example of harmful social media bots, consider the case of dating platforms. These platforms often attract more male users than female ones, which, for heterosexual users, leads to an imbalance between "supply and demand". The owners of these platforms may decide to create a number of fake, bot-controlled accounts to boost the (perceived) number of female users. This leads to one of the few situations where the creators of possibly harmful bots are at the same time responsible for the platform itself on which the bots interact with humans.

The implications of bot-controlled dating platform profiles are obvious. If the bots outnumber the humans on the site, it will become less and less enjoyable for humans to spend their time there, resulting in financial losses for the companies that operate the dating service when the users choose to try a different platform that does a better job fulfilling their needs. The creation of fake profiles to lure in users is as structurally sound as a Potemkin village.

3.2 Detection of Social Media Bots

Bots on social media are traditionally seen as spam-posting pests. They populate IRC rooms, send unwanted emails, and over the past decade, they have also taken over a big portion of social networks such as twitter. Since spam emails are mostly sent with the help of botnets, most research in that area focuses on network analyses in order to find out whether an email is unwanted. In the area of IRC and twitter bots, the research uses more content- and metadata-based factors to decide if an account is human-controlled.

One possibility of preventing bots on social media is to limit their access to the platforms. CAPTCHAs (Completely Automated Public Turing Tests To Tell Computers and Humans Apart) have the goal of presenting a challenge that is easily solvable for human users, but very difficult to automate, before messages are posted. Several approaches exist for this, the most popular being one where distorted character sequences are displayed and the user has to enter the correct sequence. The idea behind this is that a computer will have more trouble recognizing the distorted letters or numbers than a human does. Some insights about good word-entering CAPTCHAs are presented in [Bursztein et al., 2014].

Other implementations of CAPTCHAs focus more on image recognition and interpretation. For example, Google presented a system in 2014 which has users select images depicting a certain category of items from a pool that also contains some irrelevant images (see [Shet, 2014]). In [Rui and Liu, 2004], the CAPTCHA takes the shape of distorted human faces, with the task of clicking specific parts of the face in order to prove one's humanity.

All image-based CAPTCHA systems are based on the idea that humans are better at evaluating or interpreting unknown pictures and making the appropriate decisions; however, this assumption is not without its dangers, as the skills necessary to solve any type of CAPTCHA are already being automated and existing tools to automate CAPTCHA solutions are continuously getting more refined. Some CAPTCHA providers even design the CAPTCHAs specifically to crowdsource labelling tasks whose output is then turned into a training set for new classifiers or image recognition software. In that way, CAPTCHAs need continuous improvement and thus get more complicated as the border between human abilities and the quality of automated CAPTCHA solvers is shifting.

Since CAPTCHAs are continuously becoming more complex and therefore less quickly solvable for human users, other methods of bot detection that are less invasive are more promising in the context of social media. On social platforms, users interact in various ways, leaving a trace of data that helps identify whether an individual user is, in fact, a human. For instance, in [Gianvecchio et al., 2011], a rich framework of rules that govern the distinction between bots and non-bots is presented. Among the criteria the authors employ are the rate of repetition of individual messages, the inclusion of (repeated) links in messages, text obfuscation methods such as “cens0r1ng w0rds”, and the general entropy among all messages of an individual user. The idea is that humans are more likely to react to new situations in idiosyncratic and spontaneous ways, while bots have a predetermined set of possible messages or recipes for messages. The above criteria are combined with a small number of metadata features, such as response time of an account when reacting to messages directed at it. Using an entropy measure based on the feature set, the authors achieve between 3% and 100% accuracy on several datasets and with various classifiers. The reason for this extreme range of results is that the authors evaluate many different types of bots using the same feature set. Some of the features, such as the entropy of the timestamps of all messages from one account, can be tricked by building a random delay into the bot before messages are posted. A critical evaluation of these results follows in section 3.3.

Other approaches, such as the one by [Chu et al., 2013], focus more on usage metadata than on the content of messages to detect bots. The focus of that study are bots on blogging services. The authors inserted a logger into the header template of the website of an unnamed “busy blog site consisting of over 65,000 members” and evaluated the data collected by that tool to identify bots on the platform. They evaluate the pattern of keystrokes and mouse actions detected from the account as the basis for their classifier, which, similar to the systems mentioned above, focuses on entropy and classifies an account as a bot if the entropy of the features is too low. This leads to an accuracy of over 99% for the test set used in this study.

Another approach that relies on metadata on social networks to identify bots is that by [Chu et al., 2012], which deals with twitter bots specifically. In addition to textual clues such as the inclusion of “spammy” links or repeated messages, the authors base the classification on features like the message timestamps and their entropy, geodata and device information. Another useful criterion they use is the social behaviour of accounts: Is the relation between followers and followees of an account similar to that of known human accounts? According to the authors, bots often display an aggressive following behaviour, which also leads to an ever-changing network of contacts, which is less likely to happen with human-operated accounts. The study aims to distinguish not only between humans and bots, but also takes into account a third category, called “cyborgs”, meaning that an account is partially automated, but also has humans involved in its curation. The authors report the average accuracy of this three-class classifier as 96%.

Finally, there are many approaches to bot detection that rely exclusively on the evaluation of the network behaviour of clients that login to the systems where they distribute their automated messages. One example of this type of approach is that described in [Seewald and Gansterer, 2010]. One advantage of the task of identifying botnets, as the authors of this publication do, is that botnets are usually very structured.

If one bot that is a part of a botnet is found on a DNS blacklist, all other bots from the same botnet can also be detected easily, provided they use IP addresses from the same range. If bots use dynamic IP addresses to obfuscate the fact that their traffic originates from the same operator, the fact that dynamic IP addresses are used can itself be a feature that is useful to the classification. The network access behaviour of a bot candidate can also be informative: As in many of the studies mentioned above, the entropy of the timestamps is an important clue, as well as TCP packet similarity, a measure that is partly related to the measure of message similarity. The authors report a high accuracy of their approach (99.15% for the network metadata classification); however, the accuracy may vary depending on the architecture of a botnet and the measures taken by the operator to obfuscate their operations.

3.3 Problems with bot detection on Social Media

Among the approaches described in section 3.2, the ones using CAPTCHAs are likely to be moderately successful, but require constant maintenance since computers are specifically being improved to be able to solve the tasks previously assumed to only be solvable for humans. The better the computers become at these tasks, the more invasive and time-consuming the CAPTCHAs are likely to be for human users. Thus, CAPTCHAs may be a quick way to slightly lower the percentage of bots on the given platform, but do not suffice to noticably reduce them.

The wide range of results reported in [Gianvecchio et al., 2011] is likely due to the fact that bots on social media can take many different forms. A bot that regularly posts links to malware sites can quickly be recognized by a detection algorithm that uses links as suspicious features; but a bot that, for instance, aggressively follows other accounts after they mention certain keywords can only be identified with a different feature, and if malware links are the most important feature used by the detection algorithm, the latter type of bot might not be recognized as such. Therefore, while the approach described in Gianvecchio et al. employs many different aspects of different types of bots, it is probable that some bots that show some of the ”less suspicious” behaviour fall through the grid and stay undetected. In this case, the same thing applies that was argued in section 2.3 about game bot detection: The bots can take so many shapes that all algorithms must weigh the features carefully to catch as many bots as possible.

Little research has been done on bot detection on dating platforms. This is probably due to the fact that these bots are actively introduced by the operators of the websites, who therefore are not especially interested in conducting any studies on the bot population of their service. We propose that measures that evaluate metadata, especially the duration of profile page views and the number of profiles viewed by an account over a certain amount of time, will be successful in these and several other social media contexts.

Finally, the approach implemented by [Chu et al., 2013], while academically interesting, may be seen as too invasive regarding the privacy of human users. A tool that logs every keystroke and mouse movement can easily be abused for harmful activities, and the risk is too high to safely use a method like this.

4 Rating Bots

4.1 How Bots Harm Rating Websites and Their Human Voters

Human opinions matter. In applications that employ algorithms of collaborative filtering, such as web shops, video streaming services or music databases, the ratings that users give individual items directly influence the content suggestions that the service then calculates for other users. However, there are some motivations to try to take advantage of this type of rating or voting system: to give one's own work more exposure or to discredit the quality of one's competitor's work. In order to do so, one simply has to create a number of bots that log in to the specified website, look for individual products or pieces of content and give ratings that skew the overall results in the desired direction.

This type of bot can be harmful in different ways. Firstly, for human users: The recommender system algorithm will start its calculations with faulty inputs, rendering the attempt to predict which items will be interesting to a user hopeless. Secondly, and more threateningly: The sellers of the products in the web shop, or creators of the music or video content on streaming sites, may suffer unpredictable disadvantages due to a manipulation of their ratings.

4.2 Detection of Rating Bots

The detection of bots on rating platforms such as Yelp or TripAdvisor can benefit from some of the same strategies that are also useful in social media bot detection, as detailed above in section 3.2. Most studies in this field focus on text-based ratings, which are similar to messages posted on social media. However, some additional criteria offer themselves in this context.

For example, in [Heydari et al., 2015], the authors report that spam reviews of a product or service capitalize the corresponding brand name more often than human-authored reviews. Another measure that uses the textual review content is employed in [Banerjee and Chua, 2014] (cited in [Heydari et al., 2015]): The authors build a statistical model of POS tag distribution in authentic reviews and compare possible fake reviews to their model. It is important to note that such a text-based approach is unlikely to be successful on its own; however, when combined with extralinguistic features such as metadata of the individual messages, an analysis of the text may improve the overall accuracy of the classification.

Among the extralinguistic features of reviews that are reported to be useful in [Heydari et al., 2015] are: Timestamps of review texts, star ratings (where available), geo-location from which a review was posted, MAC and IP address of the device from which reviews originate, and the duration of writing the review. The reason for including the geo-location is the idea that hotel reviews, for example, should be posted from a location far away from the hotel, because if a review originates from a location close to the hotel, it seems more likely that the author (who can be either a human or a bot) is affiliated with the hotel in some way. The reason for using the IP address as a feature is the same given above in section 3.2, namely, that a high number of reviews that are posted from the same IP address are likely to not be authentic and therefore bot-authored.

Another area that gets some attention in the study by Heydari et al. is one that could be summarized as an account's "habits" on a given rating website. For example, an account might review many hotels in a short period of time; if this is observed, it is unlikely that all reviews are authentic, especially if the review text is similar as well. Another observation that may mean an account is bot-controlled is the relation between positive reviews for one product or brand in combination with many negative reviews for products or brands that compete with the first one. Finally, an account that often posts reviews that deviate from the average rating of the product or service can also be suspicious, as review-spamming strategies are often employed to counter an existing majority of negative reviews or ratings.

Additional insights can be gained from observing the social behaviour of an account on a rating website, as reported by [Rahman et al., 2015]. The authors argue that accounts that participate in discussions, get good feedback on the reviews they post, and have a well-filled-out personal profile are very likely to be humans, while bot-controlled accounts rarely go beyond posting reviews on the platform, often with "google-plagiarized" profile pictures or no avatar at all. However, the authors also state that depending on these criteria may lead to many false positives in the bot detection process.

A purely extralinguistic approach is described in [Savage et al., 2015]. In this study, the classification of accounts as bots or humans is based solely on the distribution of star ratings or comparable measures, as the authors argue that evaluating the text of reviews is very expensive, especially as labelled datasets for training and testing are difficult to obtain. This approach relies on the identification of outliers in the star rating distribution in order to identify inauthentic reviews.

4.3 Problems With Bot Detection on Rating Platforms

Some of the approaches described for rating bot detection rely on statistically identifying "anomalies" and classifying those data points as bot-created. The problem that comes with this strategy is that it is based on the assumption that humans normally agree, which is a significant oversimplification. For hotels, there might be some objective criteria that lead most human reviewers to leave very negative ratings — but for products like books, music or movies, the ratings depend on many different aspects that will be evaluated differently according to each human's specific tastes. Therefore, approaches that confuse outliers with suspicious activity are likely to have a high false positive rate.

Another problem concerning the detection systems' precision arises with approaches that evaluate the social behaviour associated with an account on a rating platform. As described in the referenced literature, some algorithms use the absence of a (non-plagiarized) profile picture, for instance, as a strong indicator that the account is not authentic. However, many people sign up for rating platforms specifically to leave very negative or very positive reviews for products or services they particularly hated or enjoyed — not necessarily taking care to fill out their personal profiles before they write the review. This behaviour is plausible and makes so much sense from a human perspective that using it as a feature for the bot detection process is highly problematic.

Since human behaviour regarding ratings and reviews can follow many different patterns, any bot detection approach that relies purely on these patterns is unlikely to be accurate. For this context, approaches that focus on metadata, such as IP addresses of authors or number of reviews posted by an account in a certain period of time, seem more promising.

5 Conclusion

The identification of bots in online video games, social media and rating website contexts is an important topic as bots continue to threaten human users' enjoyment of these types of services. In particular, algorithms that rely on evaluating big data, such as recommender systems, need to be able to distinguish between authentic data and data that should be excluded from consideration because it originates from automated user accounts.

The current bot detection methods described in recent literature generally rely on classification systems that compare an individual account's behaviour with the usage patterns of known human users and known bot users. The most popular feature for the classifiers in each of the situation contexts mentioned here is entropy, as automated accounts always show some degree of predictability in their interactions with a service. However, this measure can be circumvented with simple randomization techniques if the bot operator is aware that entropy is a relevant feature, leading to a lower chance of detection. All methods of bot detection are in danger of becoming outdated as bot operators adapt to the currently-used detection systems of a service and camouflage their bots in specific ways to blend in with humans. Therefore, research into bot detection is an ongoing process that can always benefit from more work that takes state-of-the-art bots and their behaviour into account. We suggest further studies on the differences between harmless and harmful social media bots, as current literature on that field reports a wide range of results — between 3% and 100%. We assume that this range may become smaller if the classifiers used are trained on additional, more informative features.

Unfortunately, the topic of crowdsourcing bot detection is beyond the scope of this paper. Crowdsourcing could take the shape of users reporting suspicious accounts based on intuitive and observable criteria, possibly with an additional classifier using the information given by human reporters as features. Further research should be done to explore how this could be implemented and whether it improves the overall accuracy of the bot detection system.

To sum up, Big Data approaches can be very useful in the context of bot detection, and the operators of big services such as World of Warcraft, Twitter, or Yelp have the ability to record as much data as is needed to reach near-perfect results for bot detection on their platforms. This shows that big data, which is already commonly agreed to bring many advantages to the Web 2.0, can also significantly improve the aspect of online life discussed here.

References

[Banerjee and Chua, 2014] Banerjee, S. and Chua, A. Y. K. (2014). Applauses in hotel reviews: Genuine or deceptive? In *Science and Information Conference (SAI), 2014*, pages 938–942.

[Bursztein et al., 2014] Bursztein, E., Moscicki, A., Fabry, C., Bethard, S., Mitchell, J. C., and Jurafsky, D. (2014). Easy does it: more usable captchas. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2637–2646. ACM.

[Chu et al., 2013] Chu, Z., Gianvecchio, S., Koehl, A., Wang, H., and Jajodia, S. (2013). Blog or block: Detecting blog bots through behavioral biometrics. *Computer Networks*, 57(3):634–646.

[Chu et al., 2012] Chu, Z., Gianvecchio, S., Wang, H., and Jajodia, S. (2012). Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *Dependable and Secure Computing, IEEE Transactions on*, 9(6):811–824.

[Gianvecchio et al., 2011] Gianvecchio, S., Xie, M., Wu, Z., and Wang, H. (2011). Humans and bots in internet chat: measurement, analysis, and automated classification. *IEEE/ACM Transactions on Networking (TON)*, 19(5):1557–1571.

[Heydari et al., 2015] Heydari, A., ali Tavakoli, M., Salim, N., and Heydari, Z. (2015). Detection of review spam: A survey. *Expert Systems with Applications*, 42(7):3634–3642.

[Kang et al., 2013] Kang, A. R., Woo, J., Park, J., and Kim, H. K. (2013). Online game bot detection based on party-play log analysis. *Computers & Mathematics with Applications*, 65(9):1384–1395.

[Lee et al., 2016] Lee, E., Woo, J., Kim, H., Mohaisen, A., and Kim, H. K. (2016). You are a game bot!: Uncovering game bots in mmorpgs via self-similarity in the wild.

[Mitterhofer et al., 2009] Mitterhofer, S., Kruegel, C., Kirda, E., and Platzer, C. (2009). Server-side bot detection in massively multiplayer online games. *IEEE Security & Privacy*, (3):29–36.

[Pao et al., 2010] Pao, H.-K., Chen, K.-T., and Chang, H.-C. (2010). Game bot detection via avatar trajectory analysis. *Computational Intelligence and AI in Games, IEEE Transactions on*, 2(3):162–175.

[Rahman et al., 2015] Rahman, M., Carbunar, B., Ballesteros, J., and Chau, D. H. P. (2015). To catch a fake: Curbing deceptive yelp ratings and venues. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 8(3):147–161.

[Rui and Liu, 2004] Rui, Y. and Liu, Z. (2004). Artifacial: Automated reverse turing test using facial features. *Multimedia Systems*, 9(6):493–502.

[Savage et al., 2015] Savage, D., Zhang, X., Yu, X., Chou, P., and Wang, Q. (2015). Detection of opinion spam based on anomalous rating deviation. *Expert Systems with Applications*, 42(22):8650–8657.

[Seewald and Gansterer, 2010] Seewald, A. K. and Gansterer, W. N. (2010). On the detection and identification of botnets. *computers & security*, 29(1):45–58.

[Shet, 2014] Shet, V. (2014). Are you a robot? introducing "no captcha recaptcha". https://security.googleblog.com/2014/12/are-you-robot-introducing-no-captcha.html. Accessed on 2016-05-12.

[van Kesteren et al., 2009] van Kesteren, M., Langevoort, J., and Grootjen, F. (2009). A step in the right direction: Botdetection in mmorpgs using movement analysis. In *Proceedings of the 21st Belgian-Dutch Conference on Artificial Intelligence*.