Automatic classification of player complaints in social games

Koray Balcı and Albert Ali Salah

Abstract

Artificial intelligence and machine learning techniques are not only useful for creating plausible behaviors for interactive game elements, but also for the analysis of the players to provide a better gaming environment. In this paper, we propose a novel framework for automatic classification of player complaints in a social gaming platform. We use features that describe both parties of the complaint (namely, the accuser and the suspect), as well as interaction features of the game itself. The proposed classification approach, based on Gradient Boosting Machines, is tested on the COPA Database of 100.000 unique users and 800.000 individual games. We advance the state-of-the-art in this challenging problem.

Index Terms

Online social games, Sociability, In-game aggression, Abusive behavior, Chat analysis, Machine learning

I. Introduction

Automatic analysis and classification of player behavior during online games is of great interest to gaming companies that wish to improve the gaming environment. The social environment created during gameplay has its own rules and etiquette, and it needs to be protected and nurtured, just like any other social environment. In this paper, we investigate the problem of dealing with user complaints in a social game, and propose a classification framework that helps game moderators to deal with complaints effectively and reliably.

K. Balcı and A. A. Salah are with the Department of Computer Engineering, Boğaziçi University, Istanbul, Turkey. This is an uncorrected author proof. Please cite this work as: Balc, K., A.A. Salah, “Automatic classification of player complaints in social games,” IEEE Transactions on Computational Intelligence and AI in Games, vol.9, no.1, pp.103-108, 2017.

Our proposed approach specifically deals with abusive behavior via in-game communication channels. Most social online games provide several such channels, including in-game chat, private messaging, gifting (i.e. sending a virtual gift to another player), message boards, friendship and alliance requests, and such. Rapid identification and resolution of offending acts over these channels is important for the gaming community.

We explore in this work a number of features that can be used for player profiling in social online games. In particular, we use a supervised machine learning approach to create models of abusive and aggressive verbal behavior from labeled instances of abuse in such an online game, based on actual player complaints. While mechanisms for handling player complaints (e.g. due to use of hate speech, insults, aggressive behavior, etc.) exist in social games, game moderators need to spend time and energy to analyze player complaints to resolve each case individually. Subsequently, labeled data are costly to obtain. We have previously introduced the labeled COPA corpus for this purpose [1], and proposed an approach to classify abusive players by evaluating player profiles and in-game data [2]. In this paper, instead of classifying players, we directly investigate and classify individual complaints, which may involve abusive or offensive behavior. We also take into account the accusing player's in-game data, as well as both accuser and suspected players' recent communication history with each other. This not only doubles the number of features we use compared to our previous study (19 features in [2] as opposed to 43 features in the present work), but also fundamentally changes the perspective taken. It is not only the suspect's data that should be investigated to judge cases of abuse, but also the accuser, as well as the interaction between the suspect and the accuser. We report in our experimental evaluations the clear improvement of the proposed approach in comparison with the approach that just looks at the suspect.

For classification, we have used the Gradient Boosting Machine formalism, which is an ensemble of weak classifiers [3], [4]. In addition to being flexible and powerful, this approach allows the designers to look at individual features and their contribution to prediction accuracy.

To evaluate our proposed methodology, we have used the CCSoft Okey Player Abuse (COPA) Database, which we have collected over six months of game play, with 100.000 unique users, and 800.000 individual games [2]. Our labeled complaint data comprises 1.066 player complaints, and is currently the most extensive database for this problem.

This paper is organised as follows. First, we give an overview of related work in social game analytics. We then introduce the game of Okey used in our experiments in Section III. We explain our proposed methodology in Section IV. In Section V, we present the COPA database, describe its annotation, and report experimental results. We conclude with a summary of findings in Section VI.

II. Background

Artificial intelligence (AI) and machine learning methods have their uses on various domains of gaming for a while now [5]. In a recent study, Yannakakis listed the major game-related domains in which AI techniques found various uses, as well as the interactions of these domains and their mutual influences [6]. While automatic analysis of game complaints is a novel area in game research, it can be positioned in this panorama somewhere close to player modeling [7].

In recent years, web and mobile analytics tools (such as Flurry[1], Google Analytics[2], Game-Analytics[3], and Mixpanel[4]) have matured, and are frequently used to gather user data. Many games incorporate these third party tools to track user behavior, to monitor engagement, to measure retention and churn. These tools focus on game events, and typically do not incorporate any pattern recognition or machine learning modules that can be customized to benefit game behavior moderation. Nonetheless, there has been a significant amount of research on game data analysis [8]–[10]. For instance, Xie et al. used decision trees to predict the level of engagement of players by using past data from other players [11]. Bauckhage et al. recently provided an overview of recent applications of clustering on behavioral data in games [12]. Hadiji et al. used behavioral data like playtime, session length and intervals to predict churn (i.e. players that no longer play the game) in five different free-to-play games and reported a high accuracy [13]. Similarly, [14] studied clustering of behavioral patterns for generation of user profiles and evaluated several unsupervised techniques on a large dataset of *World of Warcraft*[5] players, spanning five years of gameplay interval. The question of how to apply AI techniques to improve game moderation (rather than for implementing revenue increasing game changes) is still open.

[1]http://www.flurry.com/

[2]http://www.google.com/analytics/

[3]http://www.gameanalytics.com/

[4]http://mixpanel.com/

[5]http://eu.battle.net/wow/en/

A recent survey on human behavior analysis for computer games illustrates that while game designers analyze player behavior intensively when designing their games, real time behavior analysis is rarely incorporated into the game [15]. Gaming companies that govern online games with many subscribers also use data analysis tools in monitoring player activity, for instance to detect cheating behaviors [16], or for the analysis of player performance in different dimensions like demographics, archetypes, classes, and sub-classes [17]. These tools, also called game analytics, have direct impact on game revenues, and therefore are receiving more and more interest [8], [18]. The system we propose in this paper can be seen as such an analysis tool to help the governance of an online social game.

The gaming behavior we study in this paper involves multiparty chat messaging among other variables. Multiparty chat refers to communications in microtext format where multiple participants converse asynchronously via text messages. Uthus and Aha provide a survey of artificial intelligence methods applied to the analysis of multiparty chats, and establish that while multiparty chat analysis has been the focus of substantial research in social and behavioral sciences, very few studies have been conducted for chat analysis in the gaming context [19]. Reynolds et al. previously used machine learning to detect language patterns that indicated cyberbullying [20]. While we do not analyze the actual chat content to a great depth in this study, our study contributes to the field through the inclusion of non-verbal signals and bad language usage in our analysis.

Our study has further motivational roots in the literature of personality computing in psychology [21]. Close scrutiny of our data shows that some of the conflicts between players arise from the misunderstanding of expressed social cues. The *Brunswik Lens* model has recently gained importance in multimedia computing, and provides a useful abstraction for social interactions [22]. This concept helps one to conceptualize complaints as composite constructs, including the social cues emitted by one party, and the percept created by the other party based on inference on these cues [23]. Subsequently, the Brunswik Lens suggests that the analysis of a complaint (as well as other social constructs) should include both the source and the receiver of the social message, in our case, the *accuser* and the *suspect*, respectively.

III. An Online Social Game: Okey

We evaluate the proposed complaints classification approach on game data collected from a gaming platform for the popular *Okey* game, which is one of the most widely played social games in Turkey. In this section we briefly describe the game and the platform.

The Okey game resembles Rummy, and digital versions of Okey attract a very large audience. According to the independent statistics service AppMtr, the five most popular Okey games on the Facebook platform have a total of 6.500.000 monthly active unique users as of December 2014[6].

The ordinary Okey game is played with four players, seated around a table (See Fig. 1). In the online version of the game, the leading player sets up a virtual table in order to start a game, and other players join in. Once the game is played out, the players may disband and join other tables, or they may continue playing together, with a bot replacing the resigned player. Each game has a single winner and betting with virtual currency is a common aspect of the game.

Online versions of Okey typically offer a chat window next to the main game window, as the social aspect of the game is very prominent. Often during the turn waiting period, players use the chat area to socialize and enjoy their time. Players can also establish opt-in friendships and send each other offline messages within the game. These communication channels sometimes see abusive or offending message exchange between parties. In such cases, players can submit complaints about other players.

If a player abuses or offends one or more players, these need to be dealt with quickly and decisively in order to keep the ambiance of the game intact. Possible actions taken by game moderators include sending warning messages to the abusive player, or banning the player permanently or for a certain a period of time, depending on the severity and recurrence of the abusive act. However, analyzing these cases and deciding on a verdict requires a thorough investigation of both parties' profiles, game histories and chat logs in game tables they shared. For game moderators, dealing with these incidents requires a lot of attention and it is a very time consuming task. The system we propose in this paper is a step towards robust tools for dealing with player complaints in social games, and can be generalized beyond the Okey game.

[6]AppMtr - Facebook App Usage Metrics Tracking, available at http://www.appmtr.com/search/?q=okey. [Online; accessed 12-December-2014].

Fig. 1. Online version of Okey. Chat area is in bottom left. Players have unique and customizable avatars. Player on the left side is a bot. Players and visitors in table are listed on the bottom middle area and can interact with chat and gifting.

IV. Proposed Method

We propose a system to automatically classify player complaints. Our game database contains abusive acts, as well as other types of complaints, such as cheating or game platform related issues, and false (bogus) complaints. In order to automatically filter such cases, player complaints are manually labeled as 'abusive' or 'offending' by human moderators using the short description submitted by the player. The proposed system learns the classification in a supervised way.

First, we select several features from the database related to these labelled complaints. We incorporate both accuser and suspected players' profile and a set of cues that help to describe their recent interactions in the game environment. The dataset and features we use are explained in Section IV-A. Next, we train a supervised binary classifier, where the input is an actual player complaint, and the two classes stand for *genuine complaint* and *not genuine*, as labelled by human moderators. We present the classifier of our choice in detail in Section IV-B.

A. Data and Features

In this study, we used the proprietary CCSoft Okey Player Abuse (COPA) Database, consisting of player demographics, statistics, game records, interactions and complaints, first introduced in [1]. The database is acquired from a commercial Okey game over a six months period, and incorporates roughly 100.000 unique players who at least played the game once. All the player identification information is deleted to protect players' privacy. In the mentioned period, a total of 800.000 Okey games were recorded along with the player interactions in the chat area. Within this time frame, 1.066 of the 5.000 submitted player complaints were manually labeled by game moderators to involve possible abusive acts. 933 of the abuse cases involved players who have played at least one game and 907 of these involved actual interaction between accuser and suspected players via the chat interface within the game table. We use these 907 complaints in our experiments. Note that game moderators initially only read the short reports of the complaints, but a full annotation requires checking further evidence. Later, these labeled complaints were thoroughly analyzed by inspecting table logs and private messages between involved parties (i.e. accuser and suspect players). As revealed by an investigation of the database annotations, in 398 of these complaints, 239 unique players were found out by game moderators to have actually harassed others. Note that some of the abusive players involved in more than one complaint. The remainder of the complaints are annotated as being not abusive. Our annotations include a subjective judgment of the severity of each abusive case, denoted on a Likert scale (1 being the least severe, and 5 being the most severe).

The ratio of genuine complaints is relatively low. Complaints are submitted sometimes by mistake (due to poor complaint submission interface). A more prevalent pattern is that of an abuser submitting a fake complaint about the abused player, before a genuine complaint is filed, to mislead the moderators. Sometimes players losing a game will submit complaints to get revenge on the opponents who beat them. Finally, the cultural diversity in the player background may lead to misunderstood signals. Our system is designed to learn the characteristics of genuine complaints by using moderator actions as labels, and to operate during the lifetime of the game as a tool that filters and prioritizes complaints for the moderators. Ultimately, a human moderator needs to decide on the moderation action for each case, even if the verdict is not given by the moderation team[7].

For each player, we extract information that constitutes a player's profile. In order to form a detailed profile that can help moderators during evaluation of player complaints, we analyse table chat logs and players' game performance and retrieve a number of features related to the player from the game database. In addition to accuser and suspected players' profile data, we augment the feature set with information related to communications between these two parties involved in the complaint. The list of features used and their brief descriptions are given in Table I.

B. Gradient Boosting Machine (GBM)

There are several supervised machine learning methods in the literature that can be used for classification purposes in this setting [24]. An ideal approach for this problem should produce interpretable information for the human moderators, as the decision will invariably rest with them. Subsequently, black-box approaches are not preferable. In this study, we use Gradient Boosting Machine (GBM) for classification. In this section, we give a brief overview of GBMs, and explain our motivation for their usage in this specific problem. We have contrasted GBMs with several other classifiers that produce interpretable results, including support vector machines, naive Bayes, Decision Forest and k-nearest neighbor classifiers. We report our comparisons in Section V.

GBM, first proposed by Friedman [3], is a popular machine learning technique that is based on ensembles of sequential weak learners. GBM minimizes an error term based on classification error using a gradient descent method iteratively.

More formally, for an estimation problem with input feature vector $\mathbf{x} = \{x_1, \ldots, x_n\}$ and an output variable y, a supervised learning method uses a training set of N samples of known $\{y_i, \mathbf{x}_i\}_1^N$ values. The algorithm seeks a function $F^*(\mathbf{x})$ which maps an input $\mathbf{x}$ vector to output

[7]The *League of Legends* game has a Tribunal system that empowers players in collectively judging abusive actions.

Category	**Feature**	**Description**
Player Profile	Games Played	Number of games played by player.
	Wins	Number of wins of player.
	Incomplete Games	Number of games player left before game ends.
	Rating	ELO-like rating of the player, calculated in a multiplayer setting.
	Gender	Gender of the player as they declared during player creation initially.
	Credit Purchases	Player's total number of virtual currency purchases using real money.
	Daily Logins	Number of logged in days to the game.
	Tables Entered	Number of tables joined (for both playing and watching others play).
	Friends	Number of in-game friendships made. The game offers invitation based friendship mechanism.
	Friendship Requests	Number of in-game friendship requests sent. Some may have been unaccepted.
	Private Messages (PM)	Number of private (offline) messages sent to in-game friends, delivered at log in.
	Social Rewards	Number of social rewards earned by actions such as Facebook shares and likes.
	Gifts Purchased	Number of virtual in-game gifts (visuals placed next to player seats) purchased in game tables.
	Avatar Items	Number of items purchased to customize avatar used throughout game tables and rooms.
	Chat Entries	Number of chat utterances made in all tables entered.
	Words in Chat	Number of all words uttered in all chat sessions.
	Silence Before Chat	Average time passed in table before each chat utterance of player.
	Silence After Chat	Average time passed in table after each chat utterance of player.
	Bad Language Attempts	Number of utterances that are detected and prevented by the system as foul language.
Communications	Accuser Chats	Number of chat utterances by the accuser in last 3 shared tables with suspect.
	Suspect Chats	Number of chat utterances by the suspect in last 3 shared tables with accuser.
	Other Chats	Number of chat utterances by other players (excluding accuser and suspect) in last 3 shared table
	Accuser Words	Number of words uttered by the accuser in last 3 shared tables with suspect.
	Suspect Words	Number of words uttered by the suspect in last 3 shared tables with accuser.

TABLE I

FEATURES IN THE DATA SET AND THEIR DESCRIPTIONS.

y, such that for a distribution of $(y, \mathbf{x})$ pairs, the expected value of a loss function $\Psi(y, F(\mathbf{x}))$ is minimized:

$$F^*(\mathbf{x}) = \underset{F(\mathbf{x})}{\operatorname{argmin}}\, E_{y,\mathbf{x}} \Psi(y, F(\mathbf{x})) \tag{1}$$

GBM approximates $F^*(\mathbf{x})$ with a set of additive ensemble of weak learners of the form $h(\mathbf{x}; \mathbf{a})$ which are simple functions of $\mathbf{x}$ with parameters $\mathbf{a} = \{a_1, \ldots, a_n\}$:

$$F(\mathbf{x}) = \sum_{m=0}^{M} \beta_m h(\mathbf{x}; \mathbf{a}_m) \tag{2}$$

The expansion coefficients β_m are estimated with training data along with the parameters $\mathbf{a}_m$ in an iterative manner. The algorithm is initialized with randomly assigned coefficients, and the initial estimate is refined in M steps:

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \beta_m h(\mathbf{x}; \mathbf{a}_m) \tag{3}$$

Only a subset of the training samples are used in each intermediate training step to inject randomness and prevent overfitting. Each weak learner in step m is trained by using to the total pseudo residuals from the previous step:

$$r_{im} = -\left[\frac{\partial \Psi(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)}\right]_{F(\mathbf{x})=F_{m-1}(\mathbf{x})} \tag{4}$$

GBM is highly configurable, with several options for weak learners h. For a recent overview of GBMs, see [25]. In our approach, we use decision trees, as they can be easily interpreted. For tuning the internal parameters without overlearning the training set, a 5-fold cross validation scheme was employed.

One of the powerful features of GBM is the possibility of inspection of features upon training. The algorithm allows quantifying the contribution of each feature to the overall prediction accuracy. As noted earlier, such information may help human moderators when they start investigating complaint cases manually.

The proposed system was developed with *caret machine learning R library* [26] that includes a GBM package[8]. The code is made available as an Open Source GitHub public repository[9].

V. Experimental Results and Discussion

We perform experiments on the COPA Database, as detailed in Section IV-A. In order to evaluate the results of the prediction phase, we use the number of true positives (TP), false

[8]http://cran.r-project.org/web/packages/gbm

[9]http://github.com/koraybalci/complaint-classification

positives (FP), true negatives (TN) and false negatives (FN) to calculate accuracy, sensitivity, and specificity.

$$accuracy = \frac{TP + TN}{TP + TN + FP + TP} \tag{5}$$

$$specificity = \frac{TN}{TN + FP} \tag{6}$$

$$sensitivity = \frac{TP}{TP + FN} \tag{7}$$

In general, one will try to set a confidence threshold that results in high accuracy and specificity, so that human moderators can prioritize such complaints. High accuracy means that the system catches truly abusive players and filters out questionable complaints, whereas high specificity means that false accusations are rejected with a high probability.

We report results with two different feature set configurations. The first configuration involves only suspected player's profile as input. This setting helps us to compare our results with our previous study [2], in which we also used the COPA dataset, but classified abusive players instead of classifying complaints. This configuration will establish a baseline to measure the impact of augmenting the dataset with accuser profile and features derived from the recent communication between accuser and suspected players.

In the second configuration, we present results with the entire feature set detailed in Section IV-A, including suspect and accuser profiles and their interactions.

As noted earlier, the whole data set contains 907 complaints. For the baseline experiment, which uses only the suspected player's profile, the feature vector has 19 dimensions. For the second experiment, we incorporate both accuser and suspect player profile data and their recent relevant interactions. Using 19 features per player, and 5 features related to inter-player communication (see Table I), we obtain a feature vector with 43 dimensions.

For each training and prediction configuration, we used 75% of the population for training a GBM and reserved the rest as a holdout (test) set. During training, we apply a 5-fold cross validation scheme for parameter estimation and fine tune the number of trees to be used for prediction to prevent overfitting. The holdout set is not used in this phase. We used a *Bernoulli* distribution to model the loss function, which is the suggested approach for binary classification.

We used half of the samples without replacement for subsampling in learning iterations in order to introduce some randomness and to prevent overfitting.

In Figure 2, we report accuracy, specificity and sensitivity for both configurations, using the holdout set. Our results show that the proposed system reliably classifies player complaints containing abusive behavior. Including accuser profile and the information related to the recent communication of suspects and accusers increases the success rate. In this setting, the approach proposed in [2] achieves 62.61% accuracy with only suspect profile features.

The system detects severe cases with higher reliability, which will help human moderators to focus on complaints that may require urgent attention. For only the highly severe cases (denoted with severity values 4 and 5 in the 5-point Likert scale), the classification accuracy is 85.3% in holdout set.

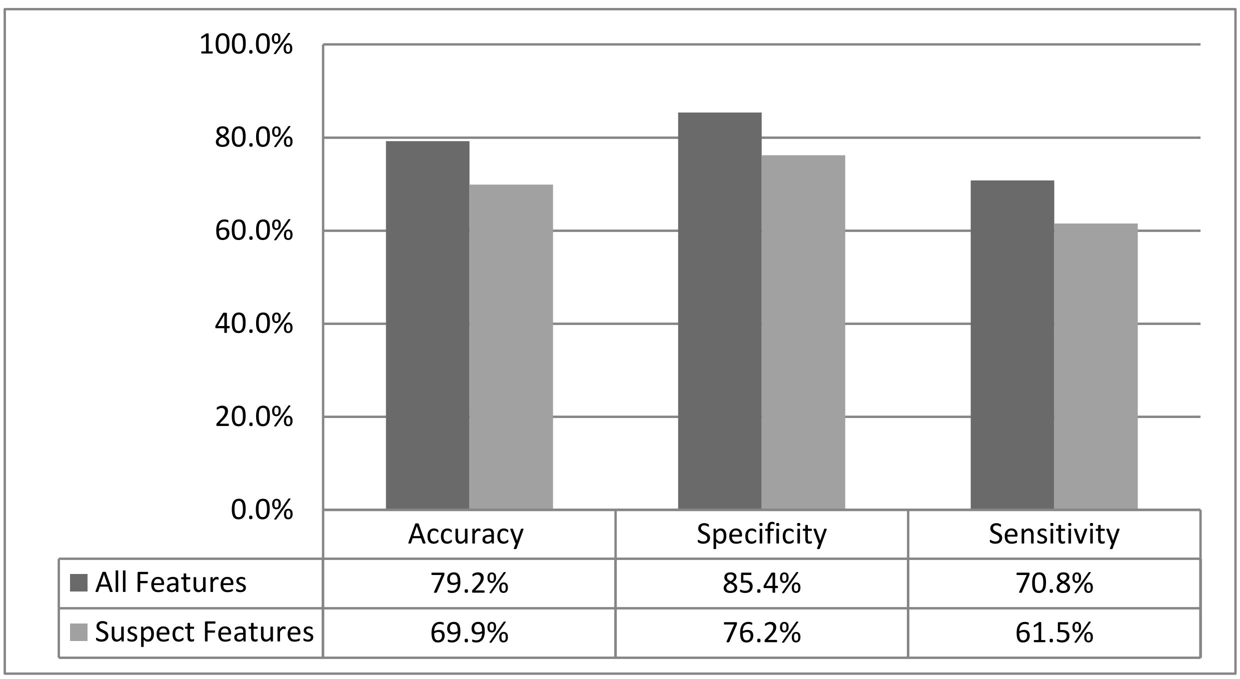

Fig. 2. Results of GBM on the holdout set using all features and only suspected player's profile features.

In addition to the system's performance, we can also inspect the trained system's internals and retrieve which features individually contribute most to the overall performance. In Figure 3, the top twenty most important features (from a total of 43 features) and their contributions are shown. Note that three of the five newly introduced inter-player communication features end up in this list. The communication features are potentially even more important, but they are difficult to analyze automatically (e.g. sarcasm is a potentially strong indicator, yet very difficult to detect). Bad language attempts, although caught by the system, are strong predictors of aggression. Suspect and accuser gender provide demographics, player profile features like games played, wins and rating show investment and involvement. Suspect words, silences and chat entries quantify social interaction and extraversion. Different games may have different features covering these factors.

Most of the important features belong to suspect profile, while accusers also play some important role in classification of complaints. Since suspect-only cues are very important, just by looking at the profile of the suspect, as proposed in [2], yields good results. Using all features results in a statistically significant increase in accuracy (paired t-test, p value $p < 0.005$).

We have experimented with several other machine learning methods. On the holdout set, GBM performs better (79.2%) when compared with other well known methods (Random Forest: 77.0%; SVM (Radial): 72.6%; k-Nearest Neighbour: 69.0%; naive Bayes: 68.1%).

VI. Conclusions

In this study, we have presented a methodology to semi-automatically identify genuine player complaints for verbal aggression and abusive behaviors in an online social game. Our method does not require human investigation and labeling of the complaints, nor annotation of chat messages during its operation, but since it is a supervised approach, the training of the system requires labeled data. Since these labels are already generated by the game moderators during their handling of day to day complaints, our approach does not require any labeling beyond what is normally performed in a gaming company.

Our previous study on player profiling suggested that using player data to spot abusive players is a valid approach [2]. With this study, we show that classifying complaints instead of players is better. In addition, we propose here the use of features from the accuser profile and communication features obtained from the interaction between both parties. An inspection of the individual contributions of the features confirmed our hypothesis. We have shown that the extended feature set outperforms the use of only suspected player profile. The data driven nature of our approach makes it applicable to other online games, provided that a rich game-specific feature set is employed to model player profiles and communication profiles. We expect that

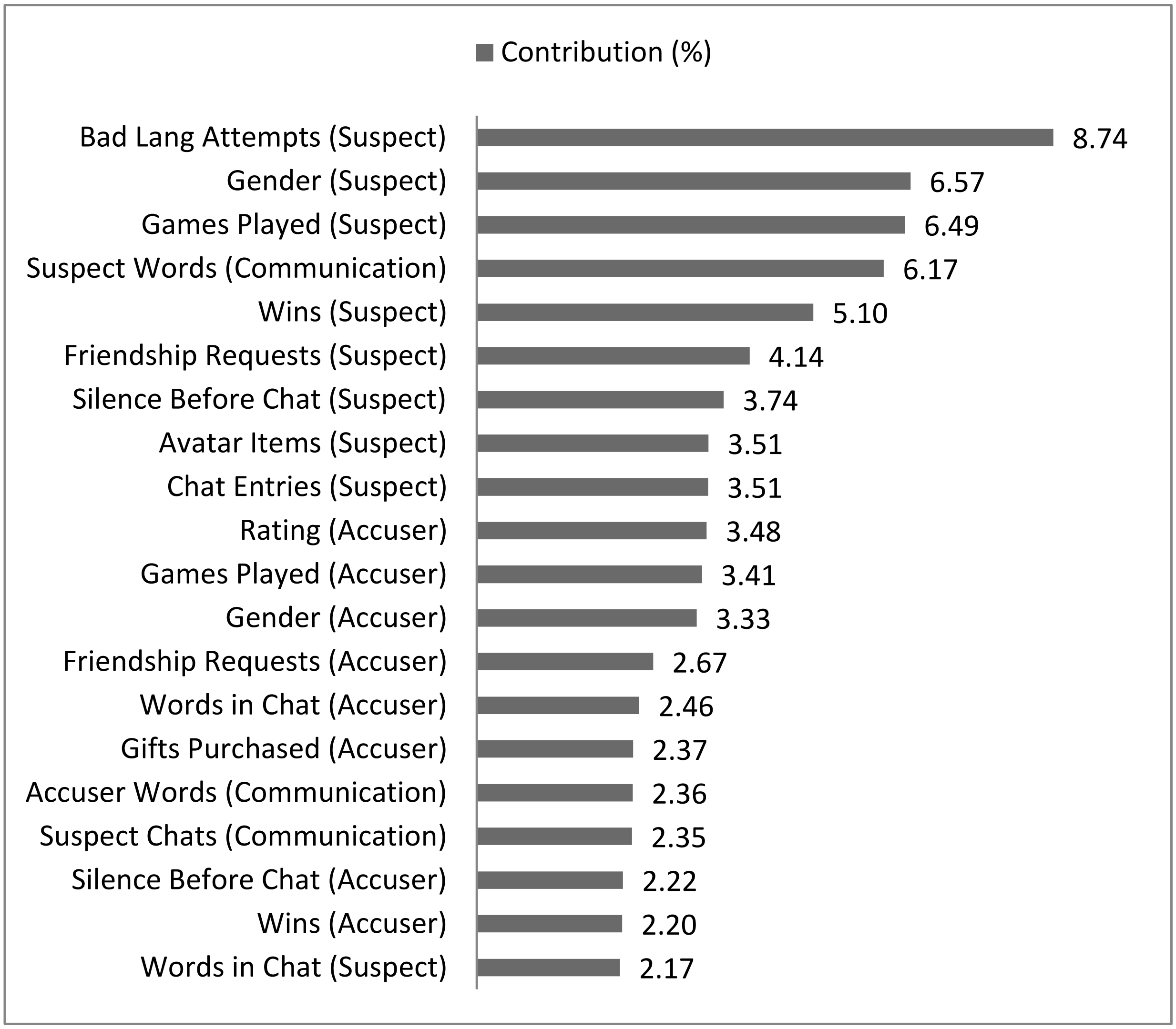

Fig. 3. Top 20 most contributing features to prediction of genuine complaints.

the game implements a complaint mechanism, which will trigger the analysis. We also expect that a human moderator will be the final judge to select the appropriate course of action. The classification itself takes a short amount of time (less than a second), provided that the historical player data can be quickly accessed. Since the moderator response is on the order of minutes, if not hours, the approach is entirely scalable to the biggest online social games currently in existence. We believe that online games which offer their players some means to socialize can benefit from our player profiling and interaction analysis in order to resolve conflicts among the players.

ACKNOWLEDGMENT

This work was supported in part by the Scientific and Technological Research Council of Turkey (TUBITAK) under grant number 114E481.

REFERENCES

[1] K. Balci and A. A. Salah, "Player profiling and offender classification from player complaints in online social games," in *Workshop on Design and Evaluation of Sociability in Online Games at CHI, Paris*, 2013.

[2] ——, "Automatic analysis and identification of verbal aggression and abusive behaviors for online social games," *Computers in Human Behavior*, vol. 53, pp. 517–526, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563214005482

[3] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, pp. 1189–1232, 2001.

[4] ——, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[5] S. M. Lucas, M. Mateas, M. Preuss, P. Spronck, and J. Togelius, "Artificial and computational intelligence in games (Dagstuhl seminar 12191)," *Dagstuhl Reports*, vol. 2, no. 5, 2012.

[6] G. Yannakakis and J. Togelius, "A panorama of artificial and computational intelligence in games," *IEEE Trans. on Computational Intelligence and AI in Games*, 2015.

[7] G. N. Yannakakis, P. Spronck, D. Loiacono, and E. André, "Player modeling," *Dagstuhl Follow-Ups*, vol. 6, 2013.

[8] T. V. Fields, "Game industry metrics terminology and analytics case study," in *Game Analytics*, M. Seif El-Nasr, A. Drachen, and A. Canossa, Eds. Springer London, 2013, pp. 53–71.

[9] S. Dumais, R. Jeffries, D. M. Russell, D. Tang, and J. Teevan, "Understanding user behavior through log data and analysis," in *Ways of Knowing in HCI*, J. S. Olson and W. A. Kellogg, Eds. Springer, 2014, pp. 349–372.

[10] C. S. Loh and Y. Sheng, "Measuring expert performance for serious games analytics: From data to insights," in *Serious Games Analytics*, ser. Advances in Game-Based Learning, C. S. Loh, Y. Sheng, and D. Ifenthaler, Eds. Springer, 2015, pp. 101–134.

[11] H. Xie, D. Kudenko, S. Devlin, and P. Cowling, "Predicting player disengagement in online games," in *Computer Games*, T. Cazenave, M. Winands, and Y. Bjrnsson, Eds. Springer, 2014, pp. 133–149.

[12] C. Bauckhage, A. Drachen, and R. Sifa, "Clustering game behavior data," *IEEE Trans. on Computational Intelligence and AI in Games*, 2014.

[13] F. Hadiji, R. Sifa, A. Drachen, C. Thurau, K. Kersting, and C. Bauckhage, "Predicting player churn in the wild," in *Proc. IEEE CIG*, 2014.

[14] A. Drachen, C. Thurau, R. Sifa, and C. Bauckhage, "A comparison of methods for player clustering via behavioral telemetry," *arXiv preprint arXiv:1407.3950*, 2014.

[15] B. Schouten, R. Tieben, A. van de Ven, and D. Schouten, "Human behavior analysis in ambient gaming and playful interaction," in *Computer Analysis of Human Behavior*, A. A. Salah and T. Gevers, Eds. Springer, 2011, pp. 387–403.

[16] J. Yan and H. Choi, "Security issues in online games," *The Electronic Library*, vol. 20, no. 2, pp. 125–133, 2002.

[17] K. Shim, R. Sharan, and J. Srivastava, "Player performance prediction in massively multiplayer online role-playing games (MMORPGS)," *Advances in Knowledge Discovery and Data Mining*, pp. 71–80, 2010.

[18] L. Mellon, "Applying metrics driven development to MMO costs and risks," *Versant Corporation, Technical Report*, pp. 1–9, 2009.

[19] D. C. Uthus and D. W. Aha, "Multiparticipant chat analysis: A survey," *Artificial Intelligence*, vol. 199-200, pp. 106–121, 2013.

[20] K. Reynolds, A. Kontostathis, and L. Edwards, "Using machine learning to detect cyberbullying," in *Proc. IEEE ICMLA*, 2011, pp. 241–244.

[21] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," *IEEE Trans. on Affective Computing*, vol. 5, no. 3, pp. 273–291, 2014.

[22] M. Cristani, A. Vinciarelli, C. Segalin, and A. Perina, "Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis," in *Proc. ACM MM*. ACM, 2013, pp. 213–222.

[23] E. Brunswik, *Perception and the representative design of psychological experiments*. Univ of California Press, 1956.

[24] C. M. Bishop, *Pattern Recognition and Machine Learning*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[25] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neurorobotics*, vol. 7, no. 21, pp. 1–21, 2013.

[26] M. Kuhn, "Building predictive models in R using the caret package," *Journal of Statistical Software*, vol. 28, no. 5, pp. 1–26, 2008.

Koray Balcı received his MSc degree from the Cognitive Sciences Department of Middle East Technical University, Ankara, Turkey in 2002. He worked in industry and research institutes for several years, mainly focusing on image processing, pattern recognition and computer graphics. He is the founder of CC Software Ltd., and a PhD student at the Dept. of Computer Engineering of Boğaziçi University. His current research interests are player profiling and analysis of in-game behavior in online video games using machine learning approaches.

Albert Ali Salah (M08) received the Ph.D. degree from the Computer Engineering Department of Boğaziçi University, where he is currently an Assistant Professor. For his work on facial feature localization, he received the inaugural EBF European Biometrics Research Award in 2006. He served as a Guest Editor for IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, and IEEE Pervasive Computing, and as an Associate Editor of IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT. He is a member of the IEEE AMD Technical Committee taskforce on Action & Perception.