Fuzzy aggressive behavior assessment of toxic players in multiplayer online battle games

Guilherme R. Andrigueto  
Computer Science Program  
Universidade Anhembi Morumbi (UAM),  
São Paulo, Brazil

Ernesto Araujo  
Biomedical Engineering Post–Graduate Program  
Universidade Anhembi Morumbi (UAM)  
Laureate International University (Laureate),  
Centro de Inovação, Tecnologia e Educação (CITE),  
Inteligência Artificial em Medicina e Saúde (IAMED),  
São José dos Campos, SP, Brazil  
ernesto.araujo.br@gmail.com

Abstract—A fuzzy system for assessing aggressive behavior of players in electronic games of Multiplayer Online Battle type is proposed in this paper. Such an approach employs the fuzzy logic and the fuzzy set theory for evaluating and classifying the subjective feeling of aggressiveness, in particular, concerning toxic players in this sort of competitive gaming. This paper aims at achieving a meta analysis of the aggressiveness according to the fundamental feelings within the psychology and psychiatric. Fear, happiness, anger and sadness are the input linguistic variables that compose the input premise space mapped into the aggressiveness outcome by using fuzzy IF–THEN rules. Such a fuzzy aggressiveness assessment system is designed to stratify and grade levels of emotional reactions that occur during the also so–called electronic sports. The proposed approach comes to be an alternative for identifying and measuring altered temper during, or after, the match.

Index Terms—Toxic Players, Aggressive Behavior, Fuzzy Logic, Classification, Assessment, Multiplayer Online Battle Games

I. INTRODUCTION

Computer games currently concern to the category of electronic sports that have been continuously growing. Simultaneously, it presents huge business potential, reflecting the growing interest in competitive gaming. A pastime for adolescents and adults worldwide, this market deals with an audience of approximately 2.1 billion players worldwide.

Such a billionaire market and the success achieved in electronic games have been prompting research on modeling the profile and behavior of the users. A prominent research area, in particular, regards the aggressiveness among players, their relationship, whether games transform people, or even whether people use games drive their feelings, and so far [1].

One style of game that became known worldwide for its very aggressive players is the Multiplayer Online Battle Arena (MOBA) [2]. This terminology is used to identify games where two teams compete for an objective within a delimited map, each one with individual group strategies. In order to accomplish complex team-based strategies, players take on very specific roles within a team [3]. The multiplayer online battle game developers, in general, bring about the aggressive attitude of players during the game [4]. The behavior known as to troll in the gaming community refers to aggressive players behaving and being categorized by using verbal offenses as well as to assign points within the games that induce to negative actions leading the opposite team to lose the match [5].

Nevertheless, aiming at the economic interest, MOBA developers are currently looking for a pleasant entertainment environment for their non–aggressive players. Due to that, various actions to change the image of the game and the attitude of aggressive players have been taking, simultaneously stimulating competitiveness and penalizing aggressive players. For instance, automatic conversion of aggressive into a non–aggressive phrase, temporary suspensions, loss of the account of the aggressive player illustrate the strategy of producers and developers of games to reduce this sort of behavior in the multiplayer online battle style. The question that comes up is if there would be a manner to measure such a subjective feeling of aggressiveness for classifying the player emotional behavior in the team-based multi-player game.

In this paper a fuzzy system is designed to assess the fundamental emotional reactions and, thus, the aggressive behavior that a person outward when playing multiplayer online battle electronic games. The system herein proposed is based on the feeling wheel (Fig. 1) derived from the psychology along with

Fig. 1. Feeling Wheel representation of emotional reaction and behavior.
the fuzzy set theory and the fuzzy logic to design a meta–analysis model for aggressiveness assessment and analysis taking into account the feelings of fear, anger, happiness and sadness [6]–[8].

The proposed approach aims at evaluating the influence of those seminal emotions that lead to aggressive reactions of such players. The output linguistic variable comprises the levels of aggressiveness determined by the aggregation of these emotional reactions through fuzzy IF–THEN rules. In so doing, the fuzzy player aggressiveness assessment system is investigated as an alternative to stratify and to grade the diverse levels of aggressive reaction concerning players in MOBA electronic games.

II. MULTIPLAYER ONLINE BATTLE AGGRESSIVENESS ASSESSMENT SYSTEM BASED ON FUZZY LOGIC

One of the most primitive and basic characteristics of all existing animals is to ensure preservation. In the studies of psychology and psychiatry, to maintain and affirm the existence of the individual, the self-preservation can, in general, be associated with aggressive impulses. Characterizing the feelings of an individual requires to assess the self-control, or lack of this condition, in social life. A feasible manner to contextualize the emotional condition of and individual is the feeling wheel, as shown in Fig. 1. In this sense, individuals’ stillness, calmness, or aggression condition can be related to the composition of the primary emotional conditions of the feelings of fear, happiness, sadness, and anger.

Constantly present while playing the game, the first input linguistic variable concerns fear that comes about at distinct instants of time during the match, in different degrees of intensity both for each player. In general, when the player is going to accomplish some new or different task, or is close to achieve a goal, fear can be considered quite certain to be present. In turn, anger is another emotion that can also be treated as part of aggressiveness as various actions take place within the game. For instance, the constant loss of matches, troll of players, losing a goal, or actions that are not pleasing or are unexpected condition concerning the player, all of them lead to frustration. In contrast, happiness is generally associated with the feeling of reward. In the game context, it is the most prestigious feeling since the happiness can regard the player carried out a good move, accomplished a goal, or obtained the victory in the match. Like other feelings, happiness can become harmful if there is no control. Thus, a state of euphoria can affect the individual in expressing aggressiveness, not being the objective to consider the broad expression of feelings. Due to that the linguistic terms partitioning the output variable refer to levels of classification according to aggressiveness during the match, and do not concern to psychological or psychiatric approach.

In this sense, the proposed approach addresses the input premise space encompassing the fear, \(x_{\text{Fear}}\), happiness–sadness, \(x_{\text{Happiness–Sadness}}\), which compose the input three–dimensional Cartesian product \(X_{\text{Fear}} \times X_{\text{Anger}} \times X_{\text{Happiness–Sadness}}\). These input linguistic variables are mapped into the aggressiveness outcome, \(y_{\text{Aggressiveness}}\), by using a nonlinear input–output set of fuzzy IF–THEN inference rules, as depicted in Fig. 2.

A. FUZZY INPUT–OUTPUT INFERENCE MAPPING

The fuzzy modeling employed in this paper uses the Mamdani type inference [9], prompt to imitate and to represent knowledge concerning the experience of healthcare professionals. Characterized as a set of IF–THEN rules:

\[ R_j : \text{IF } \langle x_1 \text{ is } M_{j1}(x_1) \rangle \text{ AND } \ldots \text{ AND } \langle x_n \text{ is } M_{jn}(x_n) \rangle \text{ THEN } \langle y \text{ is } N \rangle, \quad (1) \]

the antecedent part, IF \(\langle \text{proposition} \rangle\), defines the premise while the consequent part, THEN \(\langle \text{proposition} \rangle\), refers to conclusion, both described by linguistic expressions in propositional form, \(P = \langle x \text{ is } M \rangle\). The \(j\)-th rule, \(j = 1, 2, \ldots, m\), represents...
the amount of rules, \( \langle x_i \rangle \). The set of input fuzzy propositions, \( P_i \) \( \forall i = 1, \ldots, n \), where \( n \) is the number of input universe of discourse and represents the dimensionality of the premises; and \( y \in Y \), the inferred fuzzy proposition. The elements \( x_i \) and \( y \) refer, respectively, to the \( i \)-th input and the output concerning objects inserted in distinct classes (sets) named universe of discourse, \( x_i \in X_i \) and \( y \in Y \), also assigned linguistic variables. The input vector, \( x = [x_1, \ldots, x_n]^T \), is related to the premises (antecedent of the rule) while the output, \( y \), is associated to the conclusion (consequent of the rule). The linguistic expressions “AND” corresponds to the set operation, intersection, \( \cap \), logic operation, conjunction, \( \land \), and Triangular norm operation, \( T \)-norm, \( t(x, y) \). An operator \( T : [0, 1]^2 \to [0, 1] \) is called a \( T \)-norm if it is commutative, associative, monotonic and has 1 as neutral element. When using the Mamdani fuzzy system, the \( T \)-norm is carried out by the minimum operation. The defuzzification operation is herein carried out by employing the center of area. The elements \( M_i \subset X_i \) and \( N \subset Y \) are fuzzy sets and assigned linguistic terms, as well, partitioning the respective universes of discourse.

B. Input and Output Fuzzy Sets

The input fuzzy sets \( M^{\text{Fear}} \) and \( M^{\text{Happiness–Sadness}} \), for \( j_{\text{Fear}} = j_{\text{Happiness–Sadness}} = 1, \ldots, 2 \), meanwhile the input fuzzy set \( M ^{\text{Anger}} \) and output fuzzy set \( M^{\text{Agressiveness}} \), for \( j_{\text{Anger}} = j_{\text{Agressiveness}} = 1, \ldots, 4 \), have their membership functions defined according to the general description as follows. Consider a membership function, \( \mu_M : X_i \to [0, 1] \), defined upon an universe of discourse, \( X_i \), to which is associated a set of terms \( T = \{M_1, M_2, M_3\} \); a linguistic term \( M_j \in T \), where \( c(M_j) = \{x_0 \in X_i | \mu_M(x_0) = 1\} \) and \( s(M_j) = \{x_0 \in X_i | \mu_M(x_0) > 0\} \), respectively, denote the core and support of \( M_j \). In this paper, each linguistic term \( M_j \in T \) is shaped according to a Bell membership function:

\[
\mu_{M_j}(x_i; a, b, c) = \frac{1}{1 + \left(\frac{x_i - c}{a}\right)^{2b}}
\]

represented by \( \langle a, b, c \rangle \), where the slope is given by \( b/2a \); \( a \) defines the width of the membership function, where a larger value creates a wider membership function, \( b \) defines the shape of the curve on either side of the central plateau, where a larger value creates a more steep transition; and \( c \) defines the center of the membership function. The system is designed by employing the Ruspini partitions.

C. Input Linguistic Variables and Linguistic Terms

1) Fear: The Fear input variable presents two partitions corresponding to the set of linguistic terms \( T^{\text{Fear}} \) = \{Rational, Irrational\}, distributed into the universe of discourse in the range \( X^{\text{Fear}} = [0, 10] \). The set of terms for \( X^{\text{Fear}} \) is \( M^{\text{Fear}}_{\text{Rational}} = (5, 2.5, -1.11e -16) \) and \( M^{\text{Fear}}_{\text{Irrational}} = (5, 2.5, 10) \) (Fig. 3(a)).

2) Anger: The second input linguistic variable refers to the anger. The set of linguistic terms \( T^{\text{Anger}} \) = \{Annoyed, Furious\} and their associated membership functions are distributed in the universe of discourse with a range of \( X^{\text{Anger}} = [0, 10] \) (Fig. 3(c)). In so doing, the membership functions become \( M^{\text{Anger}} = (5, 2.5, -1.11e -16) \) and \( M^{\text{Furious}} = (5, 2.5, 10) \).

3) Happiness–Sadness: The third input linguistic variable corresponds to the compounding feelings of happiness and sadness, as illustrated in Fig. 3(b). Distinct of previous input
linguistic variables, such a universe of discourse is partitioned by four membership functions whose linguistic terms are $\mathcal{T}_{\text{Happiness-Sadness}} = \{\text{Dejected, Moody, Pleasant, Ecstatic}\}$. Their associated membership functions $\mathcal{M}^{\text{Dejected}}_{\text{Happiness-Sadness}} = (3.32.5 - 10), \quad \mathcal{M}^{\text{Moody}}_{\text{Happiness-Sadness}} = (3.32.5 - 3.3), \quad \mathcal{M}^{\text{Pleasant}}_{\text{Happiness-Sadness}} = (3.32.53.3), \quad \mathcal{M}^{\text{Ecstatic}}_{\text{Happiness-Sadness}} = (3.32.510)$, are distributed in the universe of discourse with a range of $X_{\text{Happiness-Sadness}} = [0, 10]$.

4) Output Diagnosing Variable: The output linguistic variable concerning the toxicity of multiplayer aggressiveness is also partitioned by employing Bell membership functions whose membership functions are assigned the linguistic terms $\mathcal{T}_{\text{Aggressiveness}} = \{\text{Alert, Normal, Tolerable, Unacceptable}\}$ distributed in a range of $X_{\text{Aggressiveness}} = [0, 10]$. The membership functions partition the universe of discourse as $\mathcal{M}^{\text{Alert}}_{\text{Aggressiveness}} = (3.32.5 - 10), \quad \mathcal{M}^{\text{Normal}}_{\text{Aggressiveness}} = (3.32.5 - 3.3), \quad \mathcal{M}^{\text{Tolerable}}_{\text{Aggressiveness}} = (3.32.53.3), \quad \mathcal{M}^{\text{Unacceptable}}_{\text{Aggressiveness}} = (3.32.510)$ (Fig. 4).

The set of 2–4–2 linguistic terms that partition the input universes of discourse yields a set of 16 valid fuzzy regions in a two-dimensional input premise space, $x = [x_{\text{Fear}}, x_{\text{Happiness-Sadness}}, x_{\text{Anger}}]^T$. Each region is mapped into the linguistic terms that partition the output universe of discourse related to the degree of emotional state of aggressiveness, according to a set of fuzzy IF–THEN inference rules.

**D. Fuzzy Multiplayer Online Battle Game Aggressiveness Rules**

The resulting Mamdani–based fuzzy multiplayer online battle aggressiveness system\(^1\) is given as:

\[
\begin{align*}
R_1 : & \ \text{IF} \ (x_{\text{Fear}} \text{ is Light}) \ \text{AND} \ (x_{\text{Happiness-Sadness}} \text{ is Light}) \ \text{AND} \ (x_{\text{Anger}} \text{ is Light}) \\
& \ \text{THEN} \ (\text{Aggressiveness is Mild}) \\
R_2 : & \ \text{IF} \ (x_{\text{Fear}} \text{ is Light}) \ \text{AND} \ (x_{\text{Happiness-Sadness}} \text{ is Moderate}) \ \text{AND} \ (x_{\text{Anger}} \text{ is Light}) \\
& \ \text{THEN} \ (\text{Aggressiveness is Medium}) \\
\ldots
\end{align*}
\]

\[
\begin{align*}
R_{15}: & \ \text{IF} \ (x_{\text{Fear}} \text{ is Severe}) \ \text{AND} \ (x_{\text{Happiness-Sadness}} \text{ is Moderate}) \ \text{AND} \ (x_{\text{Anger}} \text{ is Light}) \\
& \ \text{THEN} \ (\text{Aggressiveness is Serious}) \\
R_{16}: & \ \text{IF} \ (x_{\text{Fear}} \text{ is Severe}) \ \text{AND} \ (x_{\text{Happiness-Sadness}} \text{ is Severe}) \ \text{AND} \ (x_{\text{Anger}} \text{ is Light}) \\
& \ \text{THEN} \ (\text{Aggressiveness is Serious})
\end{align*}
\]

The proposed system enables the gradual membership from an element to a class, yielding a smooth classification as shown in Fig. 5. It is worth mentioning that such a system may be employed with any shape of fuzzy sets available. Bell membership functions are employed to illustrate the proposed approach since it is taken into account that feelings present continuous changing in temper, even when there are bursts of emotional behavior.

**III. DISCUSSION AND ILLUSTRATIVE EXAMPLES**

At a glance, the Mamdani fuzzy multiplayer online battle game aggressiveness IF–THEN system results in a nonlinear input–output mapping, as depicted in Fig. 5. According to the resulting MOBA fear-happiness-sadness-anger-based aggressiveness surface, the toxic player behavior comes about when these feelings active certain firing levels, regardless of whether they are common sense considered good or not. Such an outcome surface is biased by the anger input linguistic variable since it prompts a seminal influence in aggressiveness.

The more intense and out of control each of these emotional input variables are, the worse is the aggressiveness, as it is possible to observe in those graphics in Fig. 5. In this sense, as in any competition, the team should focus on the goal and not on any other element that affects the emotional state and drives the attention. In contrary, as the technical perspective herein exposed, the analysis of those feelings detaches that when they assume high values, the aggressiveness of the players can reach unacceptable levels.

It is worth detaching the influence of the happiness-sadness input variable in the surfaces, taking into account the space premise of $X_{\text{Happiness-Sadness}} \times X_{\text{Fear}} \times X_{\text{Happiness-Sadness}} \times X_{\text{Anger}}$, respectively, illustrated in Fig. 5(a), and Fig. 5(b).

In a glimpse, it is possible to observe the fading of the aggressiveness when the emotional reaction concerning happiness and sadness is in equilibrium, i.e., when the player is not influenced by these feelings. The happiness-sadness variable imposes a notch in the response, contrary to most of fuzzy (diagnosing, decision-making, assessment) systems where the surface does not present changes of this nature.

Since the sadness is related to the negative signal and happiness is associated to positive signal, the emotional equilibrium condition is around zero. Hence, achieving high negative and

---

\(^1\)Disclaimer: The fuzzy rules listed here should not be used in clinical diagnosis without consulting experienced physicians or psychologists.
well, that the happiness when out of control (high positive values) – represented by the linguistic term Ecstatic – presents a worse aggressiveness than high negative values regarding intense sadness – represented by the linguistic term Depressed – prompting a non symmetrical surface outcome, as can be observed in Fig. 5(a), and Fig. 5(b).

Identifying the aggressiveness condition of the toxic players by using the proposed fuzzy MOBA fear-happiness-sadness-anger system can be exemplified in Table I. Consider a player, $P_1$ presenting a yet rational fear, $x_{Fear} = 4$ meanwhile feeling pure ecstasy, $x_{Happiness-Sadness} = 7.8$, and a controlled anger feeling mostly annoyed, $x_{Anger} = 2$. In this context, the system outcomes that the player scores $y_{Aggressiveness} = 4.3$, achieving predominantly the status of Alert from the monitoring perspective. The Normal and Tolerate classifications are also achieved, but firing with a lower intensity when compared to the higher degree of activation obtained with the Alert membership function. Taking into account the influence of sadness, contrary to the previous example, when happiness was assigned a high value, assume that a second player, $P_2$ achieves the same intensity, $x_{Happiness-Sadness} = -7.8$, keeping the other input variables unchanged. In this scenery, the aggressive reaction scores $y_{Aggressiveness} = 2.9$. It is worth detaching in these examples that happiness and sadness differentiated herein only by the positive or negative signal of the measure, respectively. Further, the same intensity of happiness and sadness presents distinct results, $y_{P_1} < y_{P_2}$, coherent with the surface analysis as previously carried out. Such a condition reflects the fact that sadness induce individuals to a state of passiveness meanwhile the happiness can lead an individual to activeness and, thus, being able to cause more conflict in terms of harmful behavior. The proposed system also presents different stratifications. Although firing the same membership functions in the output universe of discourse the degree of activation are characterized as $\mu_{\text{Alert}} > \mu_{\text{Tolerate}} > \mu_{\text{Normal}} > \mu_{\text{Unacceptable}}$ for the second example, meanwhile the first one there is $\mu_{\text{Alert}} < \mu_{\text{Tolerate}} < \mu_{\text{Normal}} < \mu_{\text{Unacceptable}}$.

Afterward, three individuals presenting the same average measures of fear, $x_{Fear} = 6.0$, and happiness-sadness, $x_{Happiness-Sadness} = 4.3$, but with distinct anger intensity are described to represent the influence of this latter variable upon the aggressiveness assessment. The first individual, $P_3$, presents the anger intensity of $x_{Anger} = 2$; the second one, $P_4$, is characterized as $x_{Anger} = 5.5$; and the third, associated to $x_{Anger} = 9$. The resulting outcomes are, respectively, $y_{P_3} = 3.6$, $y_{P_4} = 5.2$, and $y_{P_5} = 5.9$. These individuals fire all the membership functions of the output universe of discourse, but differ by the degree of activation that each class assume. For instance, for $P_3$ there is $\mu_{\text{Alert}} > \mu_{\text{Tolerate}} > \mu_{\text{Normal}} > \mu_{\text{Unacceptable}}$; meanwhile for $P_4$, $\mu_{\text{Alert}} = \mu_{\text{Tolerate}} = \mu_{\text{Normal}}$, $\mu_{\text{Unacceptable}} > \mu_{\text{Alert}} > \mu_{\text{Tolerate}} > \mu_{\text{Normal}}$, and, finally, for $P_5$, there is $\mu_{\text{Alert}} = \mu_{\text{Tolerate}} = \mu_{\text{Normal}} = \mu_{\text{Unacceptable}}$. 

The aggressiveness surfaces for multiplayer online battle game emotional analysis: Relationship between the happiness-sadness and the fear (5(a)); the anger and the happiness-sadness (5(b)); and the anger and the fear (5(c)) for aggressiveness assessment.

![Fig. 5. Aggressiveness surfaces for multiplayer online battle game emotional analysis: Relationship between the happiness-sadness and the fear (5(a)); the anger and the happiness-sadness (5(b)); and the anger and the fear (5(c)) for aggressiveness assessment.](image)


\[ \mu_{M}^{Aggressiveness} > \mu_{M}^{Aggressiveness} > \mu_{M}^{Aggressiveness} \]

The fact that the same membership functions are activated for these examples simultaneously that present different scores means that the proposed system enables to capture the approximate reasoning by flexing the perception of the subjectiveness of such a complex variable.

The influence of fear in the aggressiveness assessment is illustrated onwards. Consider two individuals who are characterized as being furious, \( x_{Anger} = 9.4 \), and in ecstasy (happiness), \( x_{Happiness-Sadness} = 8.9 \). Although the rational fear, \( x_{Fear} = 4.5 \), of one individual, \( P_6 \), is close to the irrational fear, \( x_{Fear} = 6.5 \), of the player \( P_7 \), the quantitative aggressive measures assume distinct values, \( y_{Aggressiveness} = 6.5 \), \( y_{Aggressiveness} = 7.3 \), respectively. Likewise previous example, in which all the membership functions are active, the qualitative aggressive measures are distinct by the degree of activation of the classes. In this sense, while \( P_6 \) is described by \( \mu_{M}^{Aggressiveness} > \mu_{M}^{Aggressiveness} > \mu_{M}^{Aggressiveness} \), the individual \( P_7 \) presents \( \mu_{M}^{Aggressiveness} > \mu_{M}^{Aggressiveness} > \mu_{M}^{Aggressiveness} \). In this sense, when the fear change from rational to irrational, concerning an increasing variation of \( \Delta x_{Fear} = 2 \), inflicts changes in the evaluation of the aggressive behavior from Tolerate to Unacceptable.

As it is possible to notice, the proposed approach enables capturing the subjective influence of more than one emotional behavior and reaction of fear, happiness, sadness, and anger related to the feeling wheel that, in turn, comes to be a single measure, since the fuzzy system also scores this emotional condition.

IV. CONCLUSION

In this paper, assessing the aggressiveness is carried out by employing the primary feelings of fear, happiness, anger and sadness that compose the feeling wheel to represent emotions. A set of fuzzy IF-THEN rules grade altered temper to support stratifying emotional reactions simultaneously that suits to capture the approximate reasoning by flexing the perception of the subjectiveness of such a complex variable. The resulting fuzzy aggressiveness assessment system addresses the behavior of very aggressive – i.e., toxic – players herein used for multiplayer online battle arena game evaluation. Future work extends this fuzzy fear-happiness-sadness-anger system to deal with experimental data after being approved by the human ethical committee to carry out practical assessment, also analyzing exogenous confounding or modulation (scheduling) factors that can interfere in this aggressiveness emotional analysis. The resulting fuzzy fear-happiness-sadness-anger aggressiveness assessment can be an alternative for MOBA producers and developers to support identification toxic players as well as aggressive behavior during, or after, the match.

REFERENCES