

機械学習を用いたゲーム依存症診断システムの提案

○仰文龍 Ivan Tanev 下原勝憲 (同志社大学)

A Game Addiction Diagnosis System by Using Machine Learning

* Y. Wenlong, I. Tanev and K. Shimohara (University of Doshisha)

Abstract — The Internet is rapidly developing and becoming widespread in the world. So overplaying games has become a big problem all over the world. In this research, we proposed a system for diagnosing game addiction using machine learning techniques. The diagnostic criteria for game addiction were based on DSM-5 (Diagnostic and Statistical Manual of Mental Disorders) published by the American Psychiatric Association.

As a result, we know that the accuracy rate is 87.5%, the precision rate is 60%, the recall rate is 60%, and the F value is 60%. From the results, we can see that the most important features are question 7, question 4, question 5, question 8 and age 25~30.

Key Words: Game Addiction, decision tree, DSM-5

1. はじめに

世界において、インターネットは急速に発展、普及している。個人用のパソコンが普及し、持ち歩けるパソコンと言われているスマートフォンも中学生まで大勢の人持っている。また、駅、レストラン、学校など公衆場所に無線 LAN の設置により、人々は概ね場所や状況に関係なくインターネットに接続可能になってきた。それとともに、ゲームが近年段々盛んで来た。ファミ通により日本国内のゲーム市場規模は 2010 年に比べて、2019 年が倍以上増えてきた[1]。また、2020 年に新型コロナウイルス感染症の影響でこの現象が更に拡大している。

インターネットやスマートフォンの普及でゲームに過度に依存する問題が世界各地で指摘されている[2]。ゲーム依存とは、大まかに言えば、ゲームのプレイ状況が単なる好きの域を超えてしまい、プレイ時間を制限したり、実生活に支障を及ぼさないようにコントロールしたりすることが難しくなっている状態のことを指す。厚生労働省の調査によると、ネット依存が疑われる人は成人で推定約 400 万人、中高生で約 90 万人いると推定した。そのうちの 90% がゲーム依存症だった[3]。ゲーム依存症の患者に起きた問題は欠席、欠勤、引きこもり、昼夜逆転、退学、失職するなどがある。

ゲーム依存がどの程度広がっているのかについては部分的な報告しかなく、社会問題としての認知度は低い。本研究は、機械学習の手法を利用し、ゲーム依存症を診断するシステムを提案し、人々が簡単にゲーム依存症を診断できるようにすることが目的である。

2. 研究コンセプト

2.1. ゲーム依存症の診断基準

本研究で利用するゲーム依存症の診断基準は 2013 年

にアメリカ精神医学会 (American Psychiatric Association) から発表した精神疾患の診断と統計の手引き第 5 版 (Diagnostic and Statistical Manual of Mental Disorders) (DSM-5) である[4]。DSM-5 はアルコールや覚醒剤などの何らかの物質による「物質関連障害」と同じ、ギャンブルに対する依存として「ギャンブル障害」も精神疾患として採用された。また、インターネットゲームに対する依存としてインターネット障害が検討されるべき疾患と提案された[5]。DSM-5 の基準は 5 つまたはそれ以上が 1 年間以内に起こっていると依存と判定される。

Table. 1 DSM-5 診断基準

1	インターネットゲームに精通している。個人は以前のゲーム活動について考えるか、次のゲームをすることを予期する。インターネットゲームは日常生活において支配的な活動となる。
2	インターネットゲームが取り去られたときの離脱症状。これらの症状は、典型的には、過敏症、不安、または悲しみとして説明されるが、薬理学的離脱の物理的兆候はない。
3	耐性 - インターネットゲームに携わる時間を増やす必要がある。
4	インターネットゲームへの参加を制御する試みが失敗した。
5	インターネットゲームを除いた、以前の趣味や娯楽の利益の喪失。
6	心理的、社会的な問題があることが判明しても、インターネットゲームの過度の使用を継続した。
7	インターネットゲームの使用に関して、家族、セラピスト、または他の人を欺いた。
8	否定的な気分 (無力感、罪悪感、不安感) から逃げるためにインターネットゲームを使用する。
9	インターネットゲームへのために、重要な人間関係、職業、教育または就職の機会を危険にさらしたり、失ったりした。

2.2. 決定木 CART アルゴリズム

決定木は非常に一般的で優れた機械学習アルゴリズムである。利点は理解と解釈が容易で、分類モデルとして使用でき、回帰モデルにも使用できる。一般的に使用される決定木アルゴリズムは、ID3、C4.5、CART などがある。本研究はCART (Classification and Regression Tree) アルゴリズムを利用することにした。

CART アルゴリズムはジニ係数を使用し、最適な特徴を選択する[6]。ジニ係数はある集合の不純度を表している。不純度は簡単に説明すると、どれくらい乱雑であるかを数値化したものである。だから、ジニ係数の値が大きいほど、集合の不純度が大きい。逆に、ジニ係数小さい特徴は比較的強い分類能力を持つことがわかる。

3. ゲーム依存症の診断モデル

3.1. アンケート

データを取得するため、アンケートを実施した。アンケートはDSM-5の基準を利用し、各項目を「ない」「偶に」「時々」「よく」「いつも」などの5段階評価で行った。それと性別と年齢を加え、インターネットで126人の応募者が回答してもらった。

Table. 2 アンケート結果の一部

答卷#	Q1	Q2	您的	Q3	您会	Q4	如果	Q5	您希	Q6	您限	Q7	通过	Q8	尽管	Q9	您是	Q10	您更	Q11	您更	
1	男	25-30岁	时不时	偶尔	完全没有	不是	一般	完全不会	不是	一般	偶尔会	是	完全不是	完全没有								
2	男	25-30岁	时不时	偶尔	经常	不是	一般	偶尔会	是	完全不是	完全没有											
3	女	18-24岁	经常	完全没有	总是	不是	完全是	不会	是	是	没有											
4	女	18-24岁	完全没有	完全没有	偶尔	完全不是	完全不是	完全不会	完全不是	不是	完全没有											
5	男	25-30岁	总是	经常	经常	完全是	是	会	完全是	是	偶尔											
6	男	18岁以下	时不时	经常	时不时	是	是	会	不是	是	偶尔											
7	男	18-24岁	完全没有	偶尔	经常	不是	是	偶尔会	完全是	不是	没有											
8	女	25-30岁	经常	时不时	偶尔	是	是	会	不是	是	有											
9	女	18-24岁	完全没有	完全没有	偶尔	完全不是	一般	完全不会	不是	完全不是	完全没有											
10	男	25-30岁	时不时	时不时	时不时	是	是	总会	完全是	一般	没有											
11	男	18-24岁	偶尔	经常	经常	不是	一般	会	不是	是	偶尔											
12	男	31-40岁	总是	总是	总是	是	完全是	总会	完全是	一般	有											
13	女	18岁以下	偶尔	完全没有	总是	一般	是	不会	完全不是	是	没有											
14	女	18-24岁	经常	经常	总是	是	一般	总会	完全是	一般	有											
15	男	18-24岁	偶尔	完全没有	总是	是	是	会	完全是	不是	偶尔											
16	男	25-30岁	总是	经常	时不时	是	不是	会	一般	是	没有											
17	男	25-30岁	经常	总是	总是	一般	完全是	总会	完全是	不是	没有											
18	女	18-24岁	时不时	完全没有	总是	不是	不是	不会	是	不是	没有											
19	男	18-24岁	时不时	经常	总是	不是	完全是	不会	不是	完全是	有											
20	男	18-24岁	时不时	经常	偶尔	是	是	会	是	是	没有											

3.2. データ処理

アンケート結果を数値化する必要がある。性別と年齢は連続データとして扱うべきではないため、各種類を分けて表示した。今回のアンケートは40歳以上の人がいなかったため、40歳以上の項目を削除した。DSM-5の質問五段階回答を1から5まで点数をつけて数値化した。また、DSM-5の基準でサンプルがゲーム依存症がある

Table. 3 データセットの一部

seq	male	female	-18	18-24	25-30	31-40	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	lable
1	1	0	0	0	1	0	3	2	1	2	3	1	2	1	1	0
2	1	0	0	0	1	0	3	2	4	2	3	3	4	1	1	0
3	0	1	0	1	0	0	4	1	5	2	5	2	4	4	2	0
4	0	1	0	1	0	0	1	1	2	1	1	1	1	2	1	0
5	1	0	0	0	1	0	5	4	4	5	4	4	5	4	3	1

時、数値1で、ゲーム依存症がない時、数値0で表す。

3.3. 決定木モデル生成

データセットの四分の一をテストデータ、四分の三を訓練データで分割した。Pythonを使い、CARTアルゴリズムの決定木モデルを作成した。それから、graphviz関数を使い、生成された決定木を可視化した。

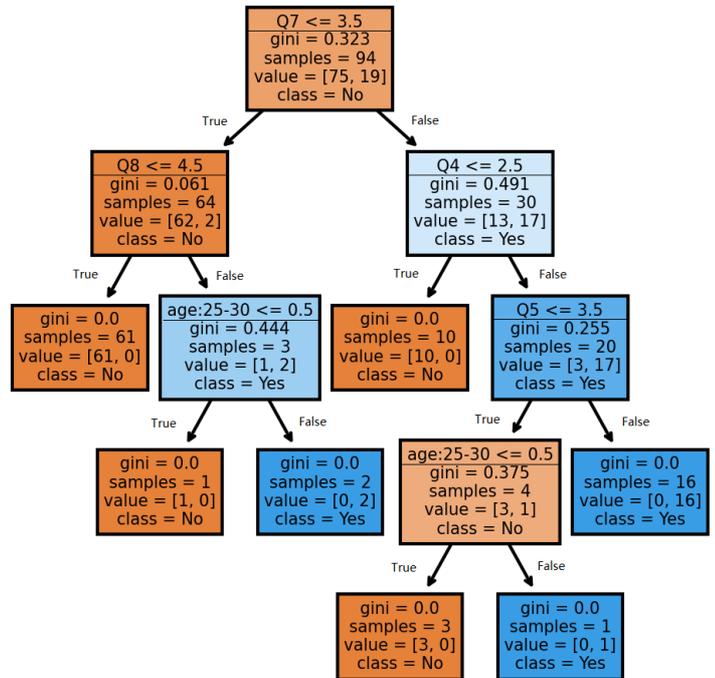


Fig. 1 決定木モデル

ノードの意味はFig.2に示す。一行目は特徴の判断条件、二行目はジニ係数の値である。三行目はサンプルの数、四行目はサンプルの分類数である。最後の5行目はこのノードから考えるとどちらかいうとゲーム依存症がないクラスに属する意味である。

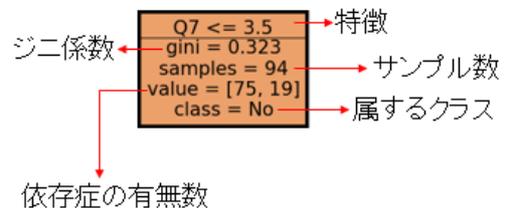


Fig. 2 ノードの意味

4. モデル分析と評価

4.1. モデル分析

この分類木モデルは、根ノードが質問7「インターネットゲームを除いた、以前の趣味や娯楽の利益の喪失。」である。得点が3点及び3点以下であれば、左側に分類される。逆に、3点以上であれば、右側の子ノードに分類される。左側の子ノードは質問8「心理的、社会的な問題があることが判明しても、インターネットゲームの過度の使用を継続した。」により分類される。5点「非常にそう思う」以外の回答が全部クラス0あるいはゲーム依存症がないクラスに分類される。5点と回答した3人のうちに、25-30歳に属する一人が、ゲーム依存症がないと判定し、ほかの2人が、ゲーム依存症があると判定される。右側の子ノードは質問4「インターネットゲームが取り去られたときの離脱症状。これらの症状は、典型的には、過敏症、不安、または悲しみとして説明されるが、薬理的離脱の物理的兆候はない。」により分類される。3点以下と答えた10人が、ゲーム依存症がないと判定される。それ以外のサンプルは質問5「インターネットゲームに携わる時間を増やす必要がある。」により分類される。3点以上答えた人16人が、ゲーム依存症があると判定された。残りの4人は年齢25-30歳である人3人が、ゲーム依存症がないと判定され、25-30歳ではない人1人が、ゲーム依存症があると判定された。

訓練データ94個のうちに、75個のデータはゲーム依存症がない、19個のデータはゲーム依存症があると判定した。75個データのうちに61個のデータは質問7が4点以下、質問8が5点以下のクラスに属することが分かった。19個のデータのうちに17個のデータは質問7が4点と5点で、質問4が2点以上のクラスに属することが分かった。

4.2. モデル評価

二分類問題において、機械学習モデル性能を評価するとき、混同行列(Confusion matrix)がよく使用されている[7]。混同行列は、あるデータを分類したとき、その正解と不正解の数を整理しておく表のことである。混同行列の定義はTable.4に示す。

Table. 5 混同行列 (Confusion matrix)

		Actual class	
		positive class	negative class
Predicted class	positive class	True Positive	False Positive
	negative class	False Negative	True Negative

- TP: ゲーム依存症があるデータを正しく推定した数
- TN: ゲーム依存症がないデータを正しく推定した数
- FP: ゲーム依存症がないデータを間違っゲーム依存症があると推定した数
- FN: ゲーム依存症があるデータを間違っゲーム依存症がないと推定した数

混同行列から正確率、適合率、再現率、F値4種類の評価指標が計算できる。

- 正確率 (Accuracy): すべてのデータのうち、正解したデータ数の割合
- 適合率 (Precision): ゲーム依存症があると予測した数のうち、正解に予測した数の割合
- 再現率 (Recall): 実際にゲーム依存症があるデータ数のうちに正確に予測できた数の割合
- F値: 再現率と適合率を偏らせずに均等に評価する

訓練データで予測した時、混同行列がTable.5に示す。訓練データを全部正確に予測したから正確率、適合率、再現率、F値全部100%である。

Table. 6 訓練データの混同行列

		Actual class	
		positive class	negative class
Predicted class	positive class	19 (True Positive)	0 (False Positive)
	negative class	0 (False Negative)	75 (True Negative)

テストデータで予測した時、混同行列がTable.6に示す。計算すると正確率は87.5%、適合率は60%、再現率は60%、F値も60%である。

Table. 4 テストデータの混同行列

		Actual class	
		positive class	negative class
Predicted class	positive class	3 (True Positive)	2 (False Positive)
	negative class	2 (False Negative)	25 (True Negative)

5. 考察と課題

5.1. 考察

各特徴の重要度を調べたとき、Fig.3のような結果が出た。

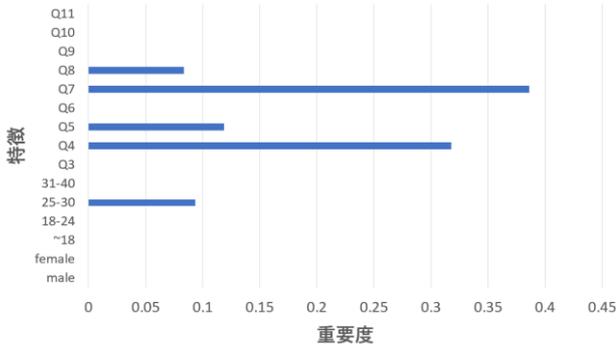


Fig. 3 各特徴の重要度

今回のモデルは五個以外の特徴と無関係のことが分かった。これはデータセットの数が少ないと考えられる。

また、訓練データで予測した時、正確率、適合率、再現率、F 値全部 100%という結果になった。逆に、テストデータで予測した時、正確率、適合率、再現率、F 値全部訓練データより低いことが分かった。これは学習過ぎるオーバーフィッティング (Overfitting) のためと考えられる。

5.2. 課題

今回のデータセットはランダムに訓練データとテストデータを分割し、1回のみ使用した。テストデータと訓練データを違うサンプルで分割し、複数の結果から特徴の偏りを確認することが課題である。また、本研究の診断結果はゲーム依存症があるとゲーム依存症がない二種類しか設定してなかった。発展としては、多段階、例えば重度、中度、軽度などのように診断結果を分けることも課題に残っている。

6. おわりに

本研究は、機械学習の手法を利用し、ゲーム依存症を診断するシステムを提案した。まず、ゲーム依存症の診断基準はアメリカ精神医学会が発表した DSM-5 を参考にした。DSM-5 の診断基準により、11 問の質問アンケートを作成した。アンケートはネットで 126 個のサンプルができた。126 個のサンプルを機械学習に使うため、すべて数値化を行った。本研究は決定木の CART アルゴリズムを利用した。Python プログラミング言語で決定木モデルを実装した。生成された決定木を可視化し、混同行列で評価した。モデルをテストデータで推定したとき、正確率は 87.5%、適合率は 60%、再現率は 60%、F 値

も 60%だった。そこから、モデルに対して、最も重要な特徴は質問 7、次に質問 4、順番で質問 5、質問 8、年齢 25~30 であることが分かった。

本研究は決定木の CART アルゴリズムを使った。二分類問題を解決するため、K 近傍法 (k-NN)、ナイーブベイズ、サポートベクターマシン (SVM)、ニューラルネットワーク (NN) など様々な機械学習アルゴリズムがある。将来の展望として、これらの方法を実装し、CART アルゴリズムと比較することができる。

参考文献

- 1) 株式会社 KADOKAWA Game Linkage : ファミ通ゲーム白書 2020 (2020)
- 2) 大嶋啓太郎, 小田哲久 : インターネット依存に関する研究 — 大学生への質問紙調査を中心に, 社会情報学会 (SSI) 学会大会研究発表論文集 (2017)
- 3) 厚生労働省 : 依存症対策
https://www.mhlw.go.jp/stf/shingi2/0000202961_00004.html
- 4) 北海道立精神保健福祉センター : インターネット依存症について
<http://www.pref.hokkaido.lg.jp/hf/sfc/grp/Internet.pdf>
- 5) American Psychiatric Association : Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, pp.795-796 (2013)
- 6) 奥喜正, 内桶誠二 : 決定木による判別と予測, 流通経済大学論集, Vol. 39, No. 4, pp. 33-43 (2005)
- 7) Visa S, Ramsay B, Ralescu A.L., Van Der Knaap E. : Confusion matrix-based feature selection, pp.120-127 (2011)