

データマイニング技術によるゲームシステムの改善

naegawa(@onuxy)

全解析結果：<http://nlab.flnet.org/ttt/>

概要

ロボット技術研究会では以前からゲーム開発が活発で、今年の工大祭でも数多くのゲームが並んでいました。筆者も今まで何度かゲームの展示をしてきました。筆者はゲームがどのようにプレイされているかを知ること、次の発表にそれを生かすことができると考えます。したがって、ゲームの展示というのは製作者にとっても非常に有益なことだと考えます。しかし、展示中はゲームの説明等を行う必要があるため、どのようにプレイされているかを見続けることはできません。そこで、今回初めてゲームの全プレイデータのログを取るということを試みました。その結果、集まったデータは142件にもなりました。この報告書では、実際に展示したゲームのプレイデータに対して行った分析とその結果を示したいと思います。

1 データマイニング

データマイニングとは、大量のデータからデータの特徴を発見することである。この分野は近年、扱うデータサイズが大きくなるとともに活発になった分野の一つである。データマイニング技術の発達と広まりによって今まで捨てていたデータから有用なデータが発見できるようになった。買い物をした客のデータからどういった商品が一緒に買われているのかを分析するマーケットバスケット分析では、得られた情報から商品の配置を工夫して売上をのばすというデータを生かすことができる例の一つである。

ところで、ゲーム展示をしたプレイデータは実際にお客さんにプレイしてもらった貴重なデータである。これらのデータの中にはゲームに関する有益な情報が入っている可能性が多分にある。そこで筆者は、買い物をした客のデータから客の行動を分析することで有用な情報が得られるのと同様に、ゲームのプレイデータからプレイヤーがどのような行動をとるのかを分析することで有用なデータが得られると考えた。この報告書では、実際に展示したゲームのプレイデータに対して行った分析手法とその結果を示す。

2 展示したゲームについて

筆者は工大祭で簡易戦略シミュレーションRPG「テックと鳥とタクティクス」というゲームを展示した(図1)。戦略シミュレーションRPGとはウォーゲームとも呼ばれ戦略を駆使した多対多の戦闘をシミュレーションしたゲームである。戦略シミュレーションRPGにはさまざまなものがあるが、「テックと鳥とタクティクス」は比較的一般的なターン制のゲームである。相

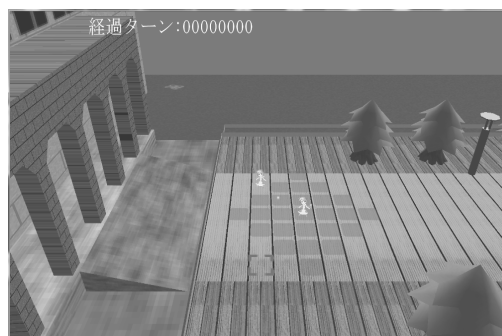


図1 テックと鳥とタクティクス：スクリーンショット

手とこちらのターンを交互に繰り返しながら各陣営のキャラクターは移動、攻撃、待機のいずれかを自分のターンに選択する。また、プレイヤーの陣営のキャラクター数は4で相手は10キャラクター以上いる。ただし、勝利条件は相手のキャラクターのうちのボスを倒すことで、残りの敵キャラクターは生き残っていても勝利となる。一方、プレイヤー側は味方キャラクターの全滅で敗北となる。今回は、非常に簡単な部分のみの実装となっており、操作も簡単にできるように工夫した。ただし、ゲーム自体は難しく、クリアをしたのは2日合わせてもたったの2人だった。

3 ゲームのモデル化について

コンピュータで計算やシミュレーションを行うに当たり、現実世界等の複雑な事象を全て入力することはできない。したがって、当面の目的を果たすために単純化した状況を考える。このことをモデル化という。

今回のゲームは簡単なルールなので分析は比較的容易

である。とはいっても、データをすべて考えることは計算量やパラメータの推定の点で現実的でない。特にゲームのプレイデータの場合、テスト用のデータ量に限界がある。実際、モデル化を行っていたにも関わらず工大祭1日目ではデータ量が不足していて、まともなデータは得られなかった。しばしば、複雑なモデルほど多くのデータが必要になるので、計算量と手に入るデータ量を考えてモデルの複雑さを考える必要がある。

4 今回用いた分析手法

今回はゲームの分析にクラスタリング分析をもちいた。クラスタリングとはデータの集合を部分集合（クラスタ）に切り分けて、それぞれのクラスタに含まれるデータが何らかの意味で「似ている」ようなデータに分けることである。このクラスタリングという手法を用いてデータを分けることにより、全てのデータを逐一チェックすることなくクラスタごとにチェックをするだけで全体を把握できるという利点がある。

やはりクラスタリングを行う場合でも何らかの方法でモデル化が必要である。まずモデル化の第1ステップとして特徴量の選択がある。つまり、これはいらぬ量を捨てるという事であり、例えば、物理でよく見られる「摩擦力を無視する」といった仮定は摩擦力という量を考えないようにしているため特徴量の選択を行っているともいえる。

今回、用いた特徴は味方キャラクタそれぞれの「HPの量」と「周囲に何人敵がいるか」である。これはつまり、ゲーム中のそれ以外の要素はすべて排除して考えるという意味であり、今味方がフィールド上のどの場所にいるかといった情報は一切無視して考えるという意味である。もちろん、これが最善かどうかというのには議論の余地があるが、データマイニングの文脈ではとにかく面白い（価値のある）情報が得られることの方が重要視される。

また、戦略シミュレーションRPGの特徴として、ターン制ではあるものの時系列のデータ（何ターン目にどういった状態かといった情報）である。時系列データの取り扱いも問題になる場合があるが、ここにも何らかの仮定を置くことが多い。今回はマルコフ性を仮定することとした。マルコフ性とは「将来の状態が現在状態のみに依存し、過去のいかなる状態にも依存しない特性」と簡単に表現されるが、この戦略シミュレーションRPGに当てはめると、「次のターンに行くことは今のターンを見て決める」ということである。（つまり、あまり戦略性がない）

これらの仮定^{*1}の下でクラスタリングを行い各プレイヤーの行動を2種類のクラスタにデータを分離した。

5 分析結果

ここでは先ほど説明したクラスタ分析を工大祭展示期間中に得られたプレイデータ142回分から得られた結果を示す。まず、クラスタごとの平均的な経過ターン数について見る。このゲームではクリア者はほとんどいないので経過ターン数はキャラクタが一人でも生きていたターン数ということになる。図2のグラフは横軸が経過ターン数で縦軸がその人数の頻度である。

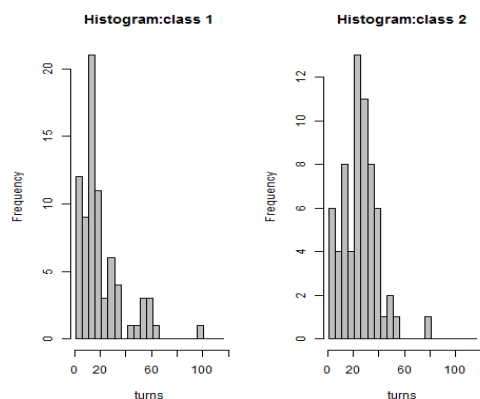


図2 経過ターン数のヒストグラム

このグラフより、クラスタ2の方が多く生き残る傾向があるように思われる^{*2}。

次に、得られたデータから「HPの量」と「周囲に何人敵がいるか」の変化をあるキャラクターについて見ていくこととする。まず、図3は「HPの量」の遷移を表した図である。縦軸は現在の状態（iniは初期状態）で横軸は次の状態、各点はそれらの遷移する確率を領域中の灰色の部分の割合で示している。例えば、この図から現在のHPが50%以上の時90%近くの確率で次のターンもHPが50%以上であると読み取れる。この図よりわかることは、より生き残っているターン数の短いクラスタ1のプレイヤーはクラスタ2のプレイヤー比べ、HPが半分を切ったからの低下が著しいことがわかる。したがって、よりゲームを盛り上げるためにはHPが一定値を下回った時にパワーアップするといった方法が考えられるだろう。

^{*1} 実際には「混合マルコフモデル」という確率モデルを用いており、混合マルコフモデルでは上で用いたマルコフ性の仮定以外にも特徴間の条件付き独立性を仮定している。

^{*2} ただし、検定の結果有意差はなかった

図4の縦軸は現在の周囲の敵の数 (*ini* は初期状態) で横軸は次のターンの周囲の敵の数、各点がそれらの遷移する確率を示している。この図より、クラスタ2のプレイヤーの方が積極的に敵に向かっていっているという事がわかる。したがって、よりゲームを盛り上げるためには周囲に敵がいらない時にはターンごとにHPが回復するといった方法が考えられる。

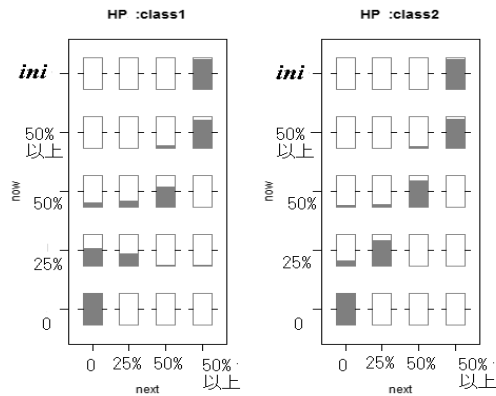


図3 HPの遷移確率
各ラベルの意味は残りHPが無し、25%未満、50%未満、50%以上

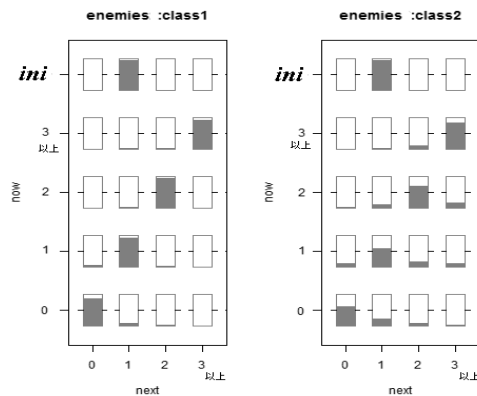


図4 周囲の敵の数の遷移確率
各ラベルの意味は周囲に敵がいらない、1人、2人、3人以上

スタイルは～です」という表示をした。このようにゲームの一部として分析を組み込んでしまうということも可能である。

今回、簡単な戦略シミュレーションRPGにおいてデータマイニング技術によってプレイヤーの行動を分析することができた。複雑なゲームのデータを適切に分析するのは時には難しい場合もある。しかし、適切な仮定(時には大胆な仮定)を置くことにより分析はかなり楽になることが多い。ぜひ、ゲーム製作者の方々はどうのように自分のゲームがプレイされているかを知り、それを次なるゲーム製作の糧としてほしい。

6 ゲームにおける分析の意味

クラスタリングのような分析では分析結果は未知データ(新しくプレイされたデータ)の予測として使える場合がある。今回の工大祭ではプレイした直後にどちらのクラスに「近い」かどうかを判断し「あなたのプレイ