(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2013-252247
(P2013-252247A)
(43)公開日　平成25年12月19日(2013.12.19)

(51)Int.Cl.		F I		テーマコード（参考）
A63F 13/10	(2006.01)	A63F 13/10		2C001

審査請求 未請求　請求項の数 8　ＯＬ　（全 19 頁）

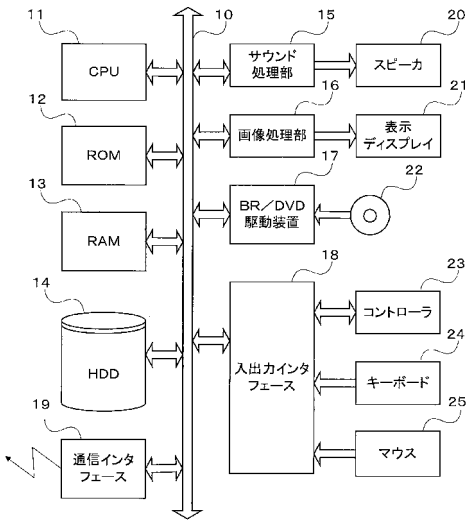
(21)出願番号　特願2012-128918 (P2012-128918)
(22)出願日　平成24年6月6日 (2012.6.6)

(71)出願人　504238806
国立大学法人北見工業大学
北海道北見市公園町１６５番地
(74)代理人　100081271
弁理士　吉田　芳春
(74)代理人　100159628
弁理士　吉田　雅比呂
(74)代理人　100162189
弁理士　堀越　真弓
(72)発明者　前田　康成
北海道北見市公園町１６５番地　国立大学法人北見工業大学内

Ｆターム(参考)　2C001 AA17 BA06 BB01 BC01 BC08 CB00 CB01 CB02 CB06 CC01 CC08

(54)【発明の名称】ロールプレイングゲームの攻略法算出装置、算出方法、算出プログラム及びこのプログラムを記録した記録媒体

(57)【要約】
【課題】被験者の体験データを取得することなく、ＲＰＧの攻略法を算出することができ、人間が楽しいと感じるゲーム要素を工学的に把握することができ、コンピュータによって操作されるキャラクタの行動選択の仕方を容易にプログラミングすることができるＲＰＧの攻略法算出装置、算出方法、算出プログラム及びこのプログラムを記録した記録媒体を提供する。
【解決手段】ＲＰＧの攻略法算出装置は、ＲＰＧのプレイヤーのゲーム開始時点における初期状態と制御期間の長さとが与えられた際に、この制御期間中に得られる総報酬に相当する期待総利得を最大にする政策を出力する最適政策算出部と、プレイヤーの状態及び時点が与えられるとその時点のその状態においてそれ以降の期待総利得を最大にする最適行動及び期待総利得の最大値を出力する行動決定部とを備えており、プレイヤーの初期状態及び制御期間の長さに対して制御期間における期待総利得を最大にすることが保証された最適政策を出力する。
【選択図】図１



【特許請求の範囲】
【請求項１】
ロールプレイングゲームのプレイヤーのゲーム開始時点における初期状態と制御期間の長さとが与えられた際に、該制御期間中に得られる総報酬に相当する期待総利得を最大にする政策を出力する最適政策算出部と、
前記プレイヤーの状態及び時点が与えられると当該時点の当該状態においてそれ以降の期待総利得を最大にする最適行動及び期待総利得の最大値を出力する行動決定部とを備えており、
前記プレイヤーの初期状態及び制御期間の長さに対して制御期間における期待総利得を最大にすることが保証された最適政策を出力することを特徴とするロールプレイングゲームの攻略法算出装置。
【請求項２】
前記最適政策算出部は、前記制御期間におけるマルコフ決定過程問題を動的計画法で求めるように構成されていることを特徴とする請求項１に記載の攻略法算出装置。
【請求項３】
ロールプレイングゲームのプレイヤーのゲーム開始時点における初期状態と制御期間の長さとが与えられた際に、該制御期間中に得られる総報酬に相当する期待総利得を最大にする政策を出力する最適政策算出工程と、
前記プレイヤーの状態及び時点が与えられると当該時点の当該状態においてそれ以降の期待総利得を最大にする最適行動及び期待総利得の最大値を出力する行動決定工程とを備えており、
前記プレイヤーの初期状態及び制御期間の長さに対して制御期間における期待総利得を最大にすることが保証された最適政策を出力することを特徴とするロールプレイングゲームの攻略法算出方法。
【請求項４】
前記最適政策算出工程は、前記制御期間におけるマルコフ決定過程問題を動的計画法で求めるように構成されていることを特徴とする請求項３に記載の攻略法算出方法。
【請求項５】
ロールプレイングゲームのプレイヤーのゲーム開始時点における初期状態と制御期間の長さとが与えられた際に、該制御期間中に得られる総報酬に相当する期待総利得を最大にする政策を出力する最適政策算出手順と、
前記プレイヤーの状態及び時点が与えられると当該時点の当該状態においてそれ以降の期待総利得を最大にする最適行動及び期待総利得の最大値を出力する行動決定手順とをコンピュータで実行させ、
前記プレイヤーの初期状態及び制御期間の長さに対して制御期間における期待総利得を最大にすることが保証された最適政策を出力することを特徴とするロールプレイングゲームの攻略法算出プログラム。
【請求項６】
前記最適政策算出手順は、前記制御期間におけるマルコフ決定過程問題を動的計画法で求めるように構成されていることを特徴とする請求項５に記載の攻略法算出プログラム。
【請求項７】
ロールプレイングゲームのプレイヤーのゲーム開始時点における初期状態と制御期間の長さとが与えられた際に、該制御期間中に得られる総報酬に相当する期待総利得を最大にする政策を出力する最適政策算出手順と、
前記プレイヤーの状態及び時点が与えられると当該時点の当該状態においてそれ以降の期待総利得を最大にする最適行動及び期待総利得の最大値を出力する行動決定手順とをコンピュータで実行させる攻略法算出プログラムを記録したコンピュータ読み取り可能な記録媒体であり、
前記プレイヤーの初期状態及び制御期間の長さに対して制御期間における期待総利得を最大にすることが保証された最適政策を出力することを特徴とするロールプレイングゲー

ムの攻略法算出プログラムを記録した記録媒体。
【請求項８】
前記最適政策算出手順は、前記制御期間におけるマルコフ決定過程問題を動的計画法で求めるように構成されていることを特徴とする請求項７に記載の記録媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ロールプレイングゲーム（ＲＰＧ）をモデル化してゲームの攻略法（行動選択の仕方）を算出するための攻略法算出装置、算出方法、算出プログラム及びこのプログラムを記録した記録媒体に関する。
【背景技術】
【０００２】
近年、コンピュータの低価格化に伴い、テレビゲーム機が広く普及し、ゲームの一分野として、ＲＰＧも広く普及している。このようなＲＰＧの開発を行う場合、従来は以下のような問題点を有していた。
（１）プレイヤーが遊ぶ際にどのようにプレイするかを把握するためには、多くの被験者に遊んでもらいプレイヤーの体験データを取得することが行われる。しかしながら、これは、多くの被験者を雇う必要があるため、コストが大幅に高くなるのみならず、データを得るのに多大な時間を要する。
（２）日々販売されるＲＰＧは多数あるが、その中でヒットするＲＰＧは非常に少数である。そのようにヒットしたＲＰＧには、人間が楽しいと感じる要素が多数含まれていると考えられるが、どのような要素が楽しいと感じる要素であるのか工学的には未だ把握されていない。
（３）最近は、プレイヤーの補佐を行うキャラクタ（お仲間キャラクタ）をコンピュータが操作するＲＰＧも多いが、このようなお仲間キャラクタの賢い行動（コマンド）選択の仕方をプログラミングすることはかなり難しい。
【０００３】
一方、ゲーム情報学等の工学分野においては、ＲＰＧをマルコフ決定過程（ＭＤＰ）等の確率モデルを用いてモデル化し、ゲームを攻略する戦略について数理工学的に扱う研究がなされ、報告されている（例えば、非特許文献１及び２）。
【先行技術文献】
【非特許文献】
【０００４】
【非特許文献１】高木幸一郎、雨宮真人、「ロールプレイングゲーム（ＲＰＧ）の戦闘におけるバランス自動調整システム開発のための基礎的考察」、情報処理学会研究報告、ＧＩ、２００１（２８）、ｐｐ．３１－３８（２００１）
【非特許文献２】高木幸一郎、雨宮真人、「ロールプレイングゲーム（ＲＰＧ）のバランスとは何か、分析およびその調整に関する提案」、情報処理学会研究報告、ＧＩ、２００１（５８）、ｐｐ．６７－７４（２００１）
【発明の概要】
【発明が解決しようとする課題】
【０００５】
非特許文献１及び２に開示されているごとき、従来のモデル化は、ＲＰＧの一部のみのモデル化であった。即ち、プレイヤーがマップ上を移動するマップモードと、マップモードにおいて敵と遭遇した際に開始される戦闘モードとからなる冒険型のＲＰＧにおいて、戦闘モードのみのモデル化であり、これでは、ＲＰＧ全体の攻略法（行動選択の仕方）を算出することができず、前述した（１）～（３）のごとき開発支援上の問題点を解決することができなかった。
【０００６】
従って本発明の目的は、被験者の体験データを取得することなく、ＲＰＧの攻略法（行

動選択の仕方）を算出することができるＲＰＧの攻略法算出装置、算出方法、算出プログラム及びこのプログラムを記録した記録媒体を提供することにある。

【０００７】

本発明の他の目的は、人間が楽しいと感じるゲーム要素を工学的に把握することができるＲＰＧの攻略法算出装置、算出方法、算出プログラム及びこのプログラムを記録した記録媒体を提供することにある。

【０００８】

本発明のさらに目的は、コンピュータによって操作されるキャラクタの行動選択の仕方を容易にプログラミングすることができるＲＰＧの攻略法（行動選択の仕方）を算出することができるＲＰＧの攻略法算出装置、算出方法、算出プログラム及びこのプログラムを記録した記録媒体を提供することにある。

【課題を解決するための手段】

【０００９】

本発明によれば、ＲＰＧのプレイヤーのゲーム開始時点における初期状態と制御期間の長さとが与えられた際に、この制御期間中に得られる総報酬に相当する期待総利得を最大にする政策を出力する最適政策算出部と、プレイヤーの状態及び時点が与えられるとその時点のその状態においてそれ以降の期待総利得を最大にする最適行動及び期待総利得の最大値を出力する行動決定部とを備えており、プレイヤーの初期状態及び制御期間の長さに対して制御期間における期待総利得を最大にすることが保証された最適政策を出力するＲＰＧの攻略法算出装置が提供される。

【００１０】

ＲＰＧの自動開発に資する研究開発としてＭＤＰなる確率モデルを用いた定式化は従来より行われていたが、従来技術では、ＲＰＧの一部のイベントを定式化するのみの不充分なものであった。

【００１１】

本発明によれば、ＲＰＧ全体をＭＤＰで定式化している。これにより、
（Ａ）開発者のみが知っている情報を既知と仮定したもとでの攻略法の算出を行い、
（Ｂ）開発者のみが知っている情報を未知と仮定した（一般のプレイヤーと同様の立場を仮定した）もとでの攻略法の算出を行っている。

【００１２】

これは、ゲーム情報学分野における学際的な知見を与えるのみならず、ゲーム産業においても以下のごとく有用である。

【００１３】

第１に、本発明で算出される目的（お金の獲得等）のもとでの攻略法（行動選択の仕方）をコンピュータにシミュレーションさせることによって、その目的におけるプレイヤーのゲーム結果をシミュレートできる。このシミュレーションを利用することによって、マップ上に隠されたアイテムやイベントに遭遇する割合（仮にプレイヤーが１万人いた場合に何人が遭遇できるか）等を把握でき、その割合を見ながら適切な隠し場所を設定することができる。

【００１４】

第２に、本発明で算出される数理工学的に最適な攻略法と実際のプレイヤーによる攻略法との比較を行うことにより、人間が楽しいと感じるゲーム要素を工学的に把握できる可能性がある。人間が楽しいと感じるゲーム要素を確率モデル上のパラメータ設定のある種のパターンとして把握できれば、そのような要素を多く含むゲーム開発を行うことができる。

【００１５】

第３に、近年では、ゲーム中にプレイヤーに協力するコンピュータ操作のキャラクタの登場が多い。このようなコンピュータによって操作されるキャラクタの行動選択の仕方をプログラミングするのは難しかったが、本発明により種々の目的毎の攻略法を算出することによって、各目的に適した（プレイヤーに協力する）キャラクタの行動選択の仕方（攻

略法）をプログラミングすることが可能となる。
【００１６】
最適政策算出部は、制御期間におけるＭＤＰ問題を動的計画法（ＤＰ）で求めるように構成されていることが好ましい。
【００１７】
本発明によれば、さらに、ＲＰＧのプレイヤーのゲーム開始時点における初期状態と制御期間の長さとが与えられた際に、この制御期間中に得られる総報酬に相当する期待総利得を最大にする政策を出力する最適政策算出工程と、プレイヤーの状態及び時点が与えられるとその時点のその状態においてそれ以降の期待総利得を最大にする最適行動及び期待総利得の最大値を出力する行動決定工程とを備えており、プレイヤーの初期状態及び制御期間の長さに対して制御期間における期待総利得を最大にすることが保証された最適政策を出力するＲＰＧの攻略法算出方法が提供される。
【００１８】
最適政策算出工程は、制御期間におけるＭＤＰ問題をＤＰで求めるように構成されていることが好ましい。
【００１９】
本発明によれば、さらにまた、ＲＰＧのプレイヤーのゲーム開始時点における初期状態と制御期間の長さとが与えられた際に、この制御期間中に得られる総報酬に相当する期待総利得を最大にする政策を出力する最適政策算出手順と、プレイヤーの状態及び時点が与えられるとその時点のその状態においてそれ以降の期待総利得を最大にする最適行動及び期待総利得の最大値を出力する行動決定手順とをコンピュータで実行させ、プレイヤーの初期状態及び制御期間の長さに対して制御期間における期待総利得を最大にすることが保証された最適政策を出力するＲＰＧの攻略法算出プログラムが提供される。
【００２０】
最適政策算出手順は、制御期間におけるＭＤＰ問題をＤＰで求めるように構成されていることが好ましい。
【００２１】
本発明によれば、さらに、ＲＰＧのプレイヤーのゲーム開始時点における初期状態と制御期間の長さとが与えられた際に、この制御期間中に得られる総報酬に相当する期待総利得を最大にする政策を出力する最適政策算出手順と、プレイヤーの状態及び時点が与えられるとその時点のその状態においてそれ以降の期待総利得を最大にする最適行動及び期待総利得の最大値を出力する行動決定手順とをコンピュータで実行させる攻略法算出プログラムを記録したコンピュータ読み取り可能な記録媒体であり、プレイヤーの初期状態及び制御期間の長さに対して制御期間における期待総利得を最大にすることが保証された最適政策を出力するＲＰＧの攻略法算出プログラムを記録した記録媒体が提供される。
【００２２】
最適政策算出手順は、制御期間におけるＭＤＰ問題をＤＰで求めるように構成されていることが好ましい。
【発明の効果】
【００２３】
本発明によれば、制御期間における期待総利得を最大にすることが保証された政策が出力されるので、プレイヤーの初期状態と制御期間長とに対して制御期間における期待総利得を最大にする政策を出力することが可能となり、そのＲＰＧに関する攻略法（行動選択の仕方）を算出することができる。
【００２４】
即ち、本発明によれば、攻略法をコンピュータにシミュレーションさせることによって、被験者の体験データを取得することができ、算出される数理工学的に最適な攻略法と実際のプレイヤーによる攻略法との比較を行うことにより、人間が楽しいと感じるゲーム要素を工学的に把握することができ、さらに、人間が楽しいと感じるゲーム要素を確率モデル上のパラメータ設定のある種のパターンとして把握できれば、そのような要素を多く含

むゲーム開発を行うことができる。また、種々の目的毎の攻略法を算出することによって、各目的に適したプレイヤーに協力するキャラクタの攻略法を容易にプログラミングすることが可能となる。

【図面の簡単な説明】

【００２５】

【図１】本発明の第１の実施形態として、ＲＰＧの開発支援に用いる攻略法算出装置の全体構成を概略的に示すブロック図である。

【図２】図１の実施形態における攻略法算出装置の主要部の動作を説明するフローチャートである。

【図３】図１の実施形態における攻略法算出装置の主要部の構成を概略的に示すブロック図である。

【図４】図１の実施形態におけるＤＰグラフの一例を示す図である。

【図５】図１の実施形態における攻略法算出装置の行動決定部の動作を説明するフローチャートである。

【図６】第１の実施形態及び第２の実施形態の実施例におけるマップの構成例を示す図である。

【発明を実施するための形態】

【００２６】

図１は本発明の第１の実施形態としてＲＰＧの開発支援に用いる攻略法算出装置の全体構成を概略的に示しており、図２は本実施形態における攻略法算出装置の主要部の動作を説明しており、図３は本実施形態における攻略法算出装置の主要部の構成を概略的に示しており、図４は本実施形態におけるＤＰグラフの一例を示しており、図５は本実施形態における攻略法算出装置の行動決定部の動作を説明している。この第１の実施形態は、各種確率分布を支配する真のパラメータθ＊が既知の場合である。

【００２７】

図１に示すように、本実施形態における攻略法算出装置は、バス１０を介して互いに接続された中央処理装置（ＣＰＵ）１１と、リードオンリメモリ（ＲＯＭ）１２と、ランダムアクセスメモリ（ＲＡＭ）１３と、ハードディスク駆動装置（ＨＤＤ）１４と、サウンド処理部１５と、画像処理部１６と、ブルーレイディスク／デジタルバーサタイルディスク（ＢＲ／ＤＶＤ）駆動装置１７と、入出力インタフェース１８と、通信インタフェース１９とを備えたコンピュータ及びこれを作動させるプログラムから構成される。

【００２８】

サウンド処理部１５はスピーカ２０に接続されており、画像処理部１６は表示ディスプレイ２１に接続されている。ＢＲ／ＤＶＤ駆動装置１７はブルーレイディスク／デジタルバーサタイルディスク／コンパクトディスク（ＢＲ／ＤＶＤ／ＣＤ）２２が装着可能となっており、入出力インタフェース１８にはコントローラ２３、キーボード２４及びマウス２５が接続されている。

【００２９】

ＣＰＵ１１は、ＲＯＭ１２に記憶されているオペレーションシステム（ＯＳ）やブートプログラム等の基本プログラムに従ってＲＡＭ１３に記憶されているプログラムを実行して本実施形態の処理を行う。また、ＣＰＵ１１は、ＲＡＭ１３、ＨＤＤ１４、音声処理部１５、画像処理部１６、ＢＲ／ＤＶＤ駆動装置１７、入出力インタフェース１８、及び通信インタフェース１９の動作を制御する。

【００３０】

ＲＡＭ１３は攻略法算出装置のメインメモリとして使用され、ＨＤＤ１４やＢＲ／ＤＶＤ駆動装置１７から転送されたプログラムやデータを記憶する。また、ＲＡＭ１３は、プログラム実行時の各種データが一時的に記憶されるワークエリアとしても使用される。

【００３１】

ＨＤＤ１４は、プログラム及びデータがあらかじめ記憶されているか、又は通信インタフェース１９を介して外部のネットワーク取り込んだプログラム及びデータが記憶される

。

【００３２】

サウンド処理部１５は、ＣＰＵ１１の指示に従ってゲームの背景音や効果音等のサウンドデータを再生するための処理を行い、スピーカ２０へその音声信号を出力する。

【００３３】

画像処理部１６は、ＣＰＵ１１の指示に従って２次元又は３次元グラフィック処理を行い、画像データを生成する。生成された画像データは、表示ディスプレイ２１に出力される。表示ディスプレイ２１ではなく図示しないＴＶの画面に表示する場合には、同期信号を付加したビデオ信号を出力する。

【００３４】

ＢＲ／ＤＶＤ駆動装置１７は、ＣＰＵ１１の指示に従って、セットされたＢＲ／ＤＶＤ／ＣＤ２２からゲームに関連するプログラムやデータを読出し、ＲＡＭ１３へ転送する。また、セットされたＢＲ／ＤＶＤ／ＣＤ２２へプログラムやデータの書き込みをすることも可能である。

【００３５】

入出力インタフェース１８は、コントローラ２３、キーボード２４及びマウス２５とＣＰＵ１１又はＲＡＭ１３との間のデータのやり取りを制御する。コントローラ２３には、ゲームを行う際に操作される方向キーやボタン等を備えている。

【００３６】

通信インタフェース１９は、通信回線を介して外部ネットワークに接続されており、ＣＰＵ１１の指示に従って、外部ネットワークとの間でプログラムやデータのやり取りが可能となっている。

【００３７】

このような構成の攻略法算出装置において、ＣＰＵ１１は、作動時は、まず、ＲＡＭ１３内にプログラム記憶領域、データ記憶領域及びワークエリアを確保し、ＨＤＤ１４又は外部からプログラム及びデータを取り込んで、プログラム記憶領域及びデータ記憶領域に格納する。次いで、このプログラム記憶領域に格納されたプログラムに基づいて、図２に示す処理を実行する。ＣＰＵ１１がプログラムを実行することによって、図３に概略的に示すごとき攻略法算出装置が構築される。

【００３８】

即ち、図３に示すように、本実施形態の攻略法算出装置は、最適政策算出部３０と、行動決定部３１とを備えるように構築される。ここで、最適政策算出部３０はＤＰグラフ作成器３０ａと、ＤＰ実施器３０ｂとを備え、行動決定部３１は行動決定器３１ａと、遷移確率テーブル３１ｂと、利得テーブル３１ｃとを備えるように構築される。

【００３９】

この攻略法算出装置の各部説明を行う前に、本実施形態で用いるマルコフ決定過程（ＭＤＰ）を利用したロールプレイイングゲーム（ＲＰＧ）の数理モデルと、その数理モデルで表現されるＲＰＧの仕様とについて説明する。

【００４０】

まず、本実施形態におけるＲＰＧの仕様について説明する。

【００４１】

プレイヤーはヒットポイント（ＨＰ)と呼ばれる数値を持ち、ＨＰが０となると、次の期にマップ上のスタート位置から再開する。再開時には、ＨＰは、スタート時と同じ最大値M_{hp}まで回復する。

【００４２】

sm_iはマップ上の位置を示し、ＳＭ、$SM=\{sm_1, sm_2, \cdots, sm_{|SM|}\}$は、マップ上の位置の集合である。ここで、ゲーム開始時のスタート位置をsm_1とする。なお、本実施形態においては、スタート位置及び現在のプレイヤーの位置は、プレイヤーによって既知であるとする。f_iは、マップ上の地形の種類を示し、Ｆ、$F=\{f_1, f_2, \cdots, f_{|F|}\}$はマップ上の地形の種類の集合である。マップ上の各位置

がどの地形に該当するかは、関数Ｆ（ｓｍ$_{i}$）∈Ｆで分かる。

【００４３】

ｅ$_{i}$は、敵の種類を示し、Ｅ、Ｅ＝｛ｅ$_{1}$，ｅ$_{2}$，・・・，ｅ$_{|E|}$｝は敵の種類の集合である。Ｍ(ｅ$_{i}$）は、敵ｅ$_{i}$の出現時のこの敵ｅ$_{i}$のＨＰを示す。プレイヤーは、敵を攻撃することによって敵のＨＰを０以下にすると、その敵を倒し、その敵に該当する報酬Ｇ（ｅ$_{i}$）を得る。

【００４４】

プレイヤーが選択できる行動（コマンド）は、マップモードと戦闘モードとでは異なり、マップモードではａ$_{1}$からａ$_{4}$が選択可能であり、戦闘モードではａ$_{5}$及びａ$_{6}$が選択可能である。ａ$_{1}$，ａ$_{2}$，ａ$_{3}$，ａ$_{4}$は、マップ上でそれぞれ右、左、上、下に移動するための行動である。ｍｖ（ｓｍ$_{i}$，ａ$_{j}$）はプレイヤーが位置ｓｍ$_{i}$で行動ａ$_{j}$を選択した際の移動先位置である。プレイヤーの移動に際して、確率ｐ（ｅ$_{k}$｜Ｆ（ｍｖ（ｓｍ$_{i}$，ａ$_{j}$）），θ＊）で移動先ｍｖ（ｓｍ$_{i}$，ａ$_{j}$）に敵ｅ$_{k}$が出現し、戦闘モードになる。敵は同時に複数出現することはなく、確率１－Σ$_{ek∈E}$ｐ（ｅ$_{k}$｜Ｆ（ｍｖ（ｓｍ$_{i}$，ａ$_{j}$）），θ＊）で何も出現せずにマップモードが続く。

【００４５】

戦闘モードの行動ａ$_{5}$はプレイヤーが戦うための行動であり、確率ｐ（Ｃ（ｅ$_{i}$）｜ａ$_{5}$，ｅ$_{i}$，θ＊）で敵ｅ$_{i}$への攻撃に成功し、その場合、敵ｅ$_{i}$のＨＰがＣ（ｅ$_{i}$）だけ減少する。プレイヤーは、確率１－ｐ（Ｃ（ｅ$_{i}$）｜ａ$_{5}$，ｅ$_{i}$，θ＊）で敵ｅ$_{i}$への攻撃に失敗する。また、行動ａ$_{5}$の選択とは直接的に関係しないが、戦闘モードでは敵もプレイヤーに対して攻撃し、確率ｐ（Ｂ（ｅ$_{i}$）｜ｅ$_{i}$，θ＊）で敵ｅ$_{i}$がプレイヤーへの攻撃に成功し、その場合、プレイヤーのＨＰがＢ（ｅ$_{i}$）だけ減少する。攻撃は、プレイヤーが常に先攻すると仮定する。敵ｅ$_{i}$は、確率１－ｐ（Ｂ（ｅ$_{i}$）｜ｅ$_{i}$，θ＊）でプレイヤーへの攻撃に失敗する。行動ａ$_{6}$は、プレイヤーが敵から逃げるための行動であり、確率ｐ（ｍａｐ｜ａ$_{6}$，θ＊）でプレイヤーは次の期にマップモードで移動し、確率１－ｐ（ｍａｐ｜ａ$_{6}$，θ＊）で戦闘モードが続く。行動ａ$_{6}$を選択した場合にも、敵は攻撃してくる。よって、プレイヤーが逃げることに失敗し、かつ敵が攻撃に成功すると、プレイヤーはダメージを受ける。θ＊は、上述の各確率分布を支配する真のパラメータであり、本実施形態では既知であるとする。

【００４６】

次に、確率システムの動的な最適化問題を定式化する優れた能力を有する数理モデルであるＭＤＰについてその概要を説明する。

【００４７】

ＭＤＰについては、例えば、金子哲夫、「マルコフ決定理論入門」、槙書店（１９７３）や森村英典、高橋幸雄、「マルコフ解析」、日科技連、東京（１９７９）等に記載されている。

【００４８】

ＭＤＰは、状態ｓ$_{i}$、ｓ$_{i}$∈Ｓ、Ｓ＝｛ｓ$_{1}$，ｓ$_{2}$，・・・，ｓ$_{|S|}$｝（｜Ｓ｜は有限）、各状態で選択できる行動ａ$_{i}$、ａ$_{i}$∈Ａ、Ａ＝｛ａ$_{1}$，ａ$_{2}$，・・・，ａ$_{|A|}$｝（｜Ａ｜は有限）、状態ｓ$_{i}$で行動ａ$_{j}$を選択したもとで、状態ｓ$_{k}$へ遷移する遷移確率ｐ（ｓ$_{k}$｜ｓ$_{i}$，ａ$_{j}$，ξ＊）（ξ＊は遷移確率分布を支配する真のパラメータ）、遷移に伴って発生する利得ｒ（ｓ$_{i}$，ａ$_{j}$，ｓ$_{k}$）で構成される。ＭＤＰの目的は、行動を選び、状態が遷移し、利得を得るという一連のプロセスを繰り返しながら総利得を最大化することである。プロセスの繰り返し回数が有限の場合には、総利得の期待値（期待総利得）を最大化する最適な決定関数を動的計画法（ＤＰ）によって求めることができる。具体的には、真のパラメータξ＊が既知の場合であれば、下記の式（１）を用いて、ｔ期の状態がｓ$_{i}$という条件下におけるｔ期以降の期待総利得の最大値Ｖ（ｓ$_{i}$，ｔ）を逐次的に計算できる。決定関数は、状態と期とを受け取って、その期で選ぶべき行動を返す関数である。

【００４９】

【数１】

$$V(s_i,t)=\max_{a_j\in A}\sum_{s_k\in S}p(s_k\mid s_i,a_j,\xi^*)(r(s_i,a_j,s_k)+V(s_k,t+1)). \qquad (1)$$

【００５０】

次に、ＭＤＰと本実施形態におけるＲＰＧとの対応について説明する。

【００５１】

x_tは、ＭＤＰにおけるｔ期の状態を示す変数であり、式（２）のように構成される。

【００５２】

$$x_t=(x_{t,1},x_{t,2},x_{t,3},x_{t,4}) \qquad (2)$$

ただし、$x_{t,1}$はｔ期におけるプレイヤーのＨＰ、$x_{t,2}$はｔ期におけるプレイヤーのマップ上での位置、$x_{t,3}$はｔ期における敵の種類、$x_{t,4}$はｔ期における敵のＨＰをそれぞれ示し、マップモードの場合には敵は存在せず、$x_{t,3}=x_{t,4}=0$とする。

【００５３】

Ａ（x_t）は、状態x_tにおいて選択可能なＭＤＰの行動集合を示す。y_tはＭＤＰにおけるｔ期に選択した行動を示す変数である。

【００５４】

次に、マップモードのｔ期の状態x_tで行動y_tを選択したときの状態遷移について説明する。ｔ＋１期には、確率ｐ（e_i｜Ｆ（ｍｖ（$x_{t,2}$，y_t）），θ＊）で敵e_iが出現し、戦闘モードの状態x_{t+1}、

$$x_{t+1}=(x_{t+1,1},x_{t+1,2},x_{t+1,3},x_{t+1,4})=(x_{t,1},mv(x_{t,2},y_t),e_i,M(e_i)) \qquad (3)$$

に遷移する。ただし、ゲームのスタート位置であるsm_1が、移動先ｍｖ（$x_{t,2}$，y_t）の場合には敵は出現しない（ｐ（e_i｜Ｆ（ｍｖ（$x_{t,2}$，y_t）），θ＊）＝０）とする。また、確率$1-\Sigma_{e_i\in E}p(e_i\mid F(mv(x_{t,2},y_t)),\theta^*)$で敵が出現せずにマップモードの状態$x_{t+1}$に遷移する。このときの状態$x_{t+1}$は、移動先ｍｖ（$x_{t,2}$，$y_t$）が$sm_1$の場合には、

$$x_{t+1}=(x_{t+1,1},x_{t+1,2},x_{t+1,3},x_{t+1,4})=(M_{hp},sm_1,x_{t,3},x_{t,4}) \qquad (4)$$

移動先ｍｖ（$x_{t,2}$，y_t）がsm_1以外の場合には、

$$x_{t+1}=(x_{t+1,1},x_{t+1,2},x_{t+1,3},x_{t+1,4})=(x_{t,1},mv(x_{t,2},y_t),x_{t,3},x_{t,4}) \qquad (5)$$

である。式（４）の場合は、プレイヤーがスタート位置sm_1に戻り、ＨＰを最大値M_{hp}まで回復した状態である。

【００５５】

次に、戦闘モードのｔ期の状態x_tで行動y_tを選択したときの状態遷移について、行動y_tが行動a_5（戦う）の場合と行動a_6（逃げる）の場合とに分けて説明する。

【００５６】

まず、行動a_5（戦う）の場合について説明する。確率１－ｐ（Ｃ（$x_{t,3}$）｜a_5，$x_{t,3}$，θ＊）（１－ｐ（Ｂ（$x_{t,3}$）｜$x_{t,3}$，θ＊）でプレイヤーと敵との両方が攻撃に失敗し、状態x_{t+1}、

$$x_{t+1}=(x_{t+1,1},x_{t+1,2},x_{t+1,3},x_{t+1,4})=(x_{t,1},x_{t,2},x_{t,3},x_{t,4}) \qquad (6)$$

に遷移する。確率（１－ｐ（Ｃ（$x_{t,3}$）｜a_5，$x_{t,3}$，θ＊））ｐ（Ｂ（$x_{t,3}$）｜$x_{t,3}$，θ＊）でプレイヤーは攻撃に失敗し、敵は攻撃に成功し、状態x_{t+1}へ遷移する。

このときの状態x_{t+1}は、$x_{t,1}>B(x_{t,3})$の場合には、

$$x_{t+1}=(x_{t+1,1},x_{t+1,2},x_{t+1,3},x_{t+1,4})=(x_{t,1}-B(x_{t,3}),x_{t,2},x_{t,3},x_{t,4}) \qquad (7)$$

で、$x_{t,1} \leqq B(x_{t,3})$の場合には、

$$x_{t+1} = (x_{t+1,1}, x_{t+1,2}, x_{t+1,3}, x_{t+1,4}) = (M_{hp}, sm_1, 0, 0) \quad (8)$$

である。式（８）の場合には、プレイヤーが敵に倒されて、ゲームのスタート位置sm_1からの再開である。確率$p(C(x_{t,3}) \mid a_5, x_{t,3}, \theta^*)(1-p(B(x_{t,3}) \mid x_{t,3}, \theta^*))$でプレイヤーは攻撃に成功し、敵は攻撃に失敗し、状態x_{t+1}へ遷移する。このときの状態x_{t+1}は、$x_{t,4} > C(x_{t,3})$の場合には、

$$x_{t+1} = (x_{t+1,1}, x_{t+1,2}, x_{t+1,3}, x_{t+1,4}) = (x_{t,1}, x_{t,2}, x_{t,3}, x_{t,4} - C(x_{t,3})) \quad (9)$$

で、$x_{t,4} \leqq C(x_{t,3})$の場合には、

$$x_{t+1} = (x_{t+1,1}, x_{t+1,2}, x_{t+1,3}, x_{t+1,4}) = (x_{t,1}, x_{t,2}, 0, 0) \quad (10)$$

である。式（１０）の場合には、敵$x_{t,3}$を倒すことに成功しているので、この状態遷移に伴い、利得$r(x_t, a_5, x_{t+1}) = G(x_{t,3})$を得る。確率$p(C(x_{t,3}) \mid a_5, x_{t,3}, \theta^*)\, p(B(x_{t,3}) \mid x_{t,3}, \theta^*)$でプレイヤーと敵の両方が攻撃に成功し、状態$x_{t+1}$へ遷移する。このときの状態$x_{t+1}$は、$x_{t,4} \leqq C(x_{t,3})$の場合には、

$$x_{t+1} = (x_{t+1,1}, x_{t+1,2}, x_{t+1,3}, x_{t+1,4}) = (x_{t,1}, x_{t,2}, 0, 0) \quad (11)$$

で、$x_{t,4} > C(x_{t,3})$かつ$x_{t,1} > B(x_{t,3})$の場合には、

$$x_{t+1} = (x_{t+1,1}, x_{t+1,2}, x_{t+1,3}, x_{t+1,4}) = (x_{t,1} - B(x_{t,3}), x_{t,2}, x_{t,3}, x_{t,4} - C(x_{t,3})) \quad (12)$$

で、$x_{t,4} > C(x_{t,3})$かつ$x_{t,1} \leqq B(x_{t,3})$の場合には、

$$x_{t+1} = (x_{t+1,1}, x_{t+1,2}, x_{t+1,3}, x_{t+1,4}) = (M_{hp}, sm_1, 0, 0) \quad (13)$$

である。式（１１）の場合には、敵$x_{t,3}$を倒すことに成功しているので、この状態遷移に伴い、利得$r(x_t, a_5, x_{t+1}) = G(x_{t,3})$を得る。式（１３）の場合はプレイヤーが敵に倒されて、ゲームのスタート位置sm_1からの再開である。

【００５７】

次に、戦闘モードのｔ期の状態x_tで行動a_6（逃げる）を選択したときの状態遷移について説明する。確率$p(map \mid a_6, \theta^*)$でプレイヤーが逃げることに成功し、状態$x_{t+1}$、

$$x_{t+1} = (x_{t+1,1}, x_{t+1,2}, x_{t+1,3}, x_{t+1,4}) = (x_{t,1}, x_{t,2}, 0, 0) \quad (14)$$

に遷移する。確率$1-p(map \mid a_6, \theta^*)(1-p(B(x_{t,3}) \mid x_{t,3}, \theta^*))$でプレイヤーが逃げることに失敗し、敵が攻撃に失敗し、状態x_{t+1}、

$$x_{t+1} = (x_{t+1,1}, x_{t+1,2}, x_{t+1,3}, x_{t+1,4}) = (x_{t,1}, x_{t,2}, x_{t,3}, x_{t,4}) \quad (15)$$

に遷移する。確率$(1-p(map \mid a_6, \theta^*))\, p(B(x_{t,3}) \mid x_{t,3}, \theta^*)$でプレイヤーが逃げることに失敗し、敵が攻撃に成功し、状態x_{t+1}へ遷移する。このときの状態x_{t+1}は、$x_{t,1} > B(x_{t,3})$の場合には、

$$x_{t+1} = (x_{t+1,1}, x_{t+1,2}, x_{t+1,3}, x_{t+1,4}) = (x_{t,1} - B(x_{t,3}), x_{t,2}, x_{t,3}, x_{t,4}) \quad (16)$$

で、$x_{t,1} \leqq B(x_{t,3})$の場合には、

$$x_{t+1} = (x_{t+1,1}, x_{t+1,2}, x_{t+1,3}, x_{t+1,4}) = (M_{hp}, sm_1, 0, 0) \quad (17)$$

である。式（１７）の場合には、プレイヤーが敵に倒されて、ゲームのスタート位置sm_1からの再開である。

【００５８】
前述した通り、プレイヤーが敵$x_{t,3}$を倒した状態遷移に伴う利得は、$r(x_t, a_5, x_{t+1}) = G(x_{t,3})$である。その他の状態遷移に伴う利得は、$r(x_t, a_5, x_{t+1}) = 0$である。本実施形態では、初期状態$x_1$が$x_1 = (Mhp, sm_1, 0, 0)$であり、各期の状態は観測可能である。また、プレイヤーや敵の攻撃力C（e_i），B（e_i）及び敵を倒したときの報酬G（e_i）等は全て既知であるとする。このもとで、T期間のプレイを行って総利得
【数２】
$$\sum_{t=1}^{T} r(x_t, y_t, x_{t+1})$$
の最大化を目的とする。
【００５９】
次に、図２に示されたフローチャート、図３に示されたブロック図、図４に示された図、及び図５に示されたフローチャートを参照して、本実施形態の攻略法算出装置の動作を説明する。
【００６０】
図２及び図３に示すように、まず、構築された最適政策算出部３０におけるＤＰグラフ作成器３０ａにプレイヤーの初期状態x_1と制御期間長Ｔとが入力される（ステップＳ１）。この初期状態x_1と制御期間長ＴとはＨＤＤ１４に格納されているデフォルト値又は前回のプレイ結果値をＲＡＭ１３のデータ記憶領域に格納したものであっても良いし、キーボード２４から入力した値であっても良い。
【００６１】
プレイヤーの初期状態x_1と制御期間長Ｔとが入力されると、ＤＰグラフ作成器３０ａは、Ｔ期間の期待総利得を最大化するための動的計画法（ＤＰ）の問題を解くためのＤＰグラフを作成する（ステップＳ２）。例えば、想定されるプレイヤーの全状態を要素とする状態集合Ｓが、$S = \{s_1, s_2, s_3, s_4\}$で$x_1 = s_1$の場合であれば、図４のようなＤＰグラフが作成される。これは、１時点目（１期）はプレイヤーの初期状態で表現され、２時点目からＴ時点目まではプレイヤーの想定される各状態で表現されたグラフにおいて、末端のＴ時点目（Ｔ期）のノードから遡りながらＤＰでＴ期間のＭＤＰ問題を解くことによって、Ｔ期間の期待総利得を最大化する最適政策を求めるための準備である。
【００６２】
次いで、ＤＰ実施器３０ｂがＤＰによってＴ期間のＭＤＰ問題を解くことによって、Ｔ期間の期待総利得を最大化する最適政策が求められる（ステップＳ３）。ＤＰ実施器３０ｂは、ＤＰグラフの末端の各ノードから順にそのノードでの最適な行動（ＲＰＧにおけるプレイヤーのコマンド選択）とそのノード以降の期待総利得の最大値を、行動決定部３１における行動決定器３１ａと連携して求める（ステップＳ４）。
【００６３】
即ち、各ノード毎にそのノードの時点ｔ（何時点目かを示す自然数）とプレイヤーの状態x_t（ｔ時点目のプレイヤーの状態）とを行動決定器３１ａへ送ると、そのノードにおける最適な行動とそのノード以降の期待総利得の最大値とが求められて行動決定器３１ａから送り返される。
【００６４】
その後、ＤＰグラフの１時点目のノードまで全て解き終わったかどうかが判断され（ステップＳ５）、解き終わっていればＤＰグラフの全ノードにおける最適な行動とそのノード以降の期待総利得の最大値が最適政策として出力される（ステップＳ６）。
【００６５】
次に、図３及び図５に示されたフローチャートを参照して行動決定部３１の動作、即ち図２におけるステップＳ４の動作を説明する。
【００６６】

まず、最適政策算出部３０のＤＰ実施器３０ｂから行動決定部３１の行動決定器３１ａへ、時点ｔ（何時点目かを示す自然数）とプレイヤーの状態x_t（ｔ時点目のプレイヤーの状態）とが入力される（ステップＳ４１）。

【００６７】

次いで、入力された時点ｔとプレイヤーの状態x_tとに応じてそのノード以降の期待総利得の最大値とそのノードにおける最適な行動とが算出される（ステップＳ４２）。ｔ＝Ｔの場合には、式（１８）で、そのノード以降の期待総利得の最大値が求められる。

【００６８】

【数３】

$$V(x_T,T)=\max_{y_T\in A(x_T)}\sum_{x_{T+1}}p(x_{T+1}\mid x_T,y_T,\theta^*)r(x_T,y_T,x_{T+1}),\qquad(18)$$

ただし、Ｖ（x_T，Ｔ）はＴ時点目の状態x_Tから時点Ｔ＋１への状態遷移に伴う最後の１期間の期待総利得の最大値である。ｐ（x_{T+1}｜x_T，y_T，θ＊）は遷移確率テーブル３１ｂから読み取ったものである。ｒ（x_T，y_T，x_{T+1}）は利得テーブル３１ｃから読み取ったものである。１≦ｔ≦Ｔ－１の場合には次の式（１９）でそのノード以降の期待総利得の最大値が求められる。

【００６９】

【数４】

$$V(x_t,t)=\max_{y_t\in A(x_t)}\sum_{x_{t+1}}p(x_{t+1}\mid x_t,y_t,\theta^*)(r(x_t,y_t,x_{t+1})+V(x_{t+1},t+1)),\ (19)$$

ただし、Ｖ（x_T，Ｔ）はｔ時点目の状態がx_tという条件のもとでの、ｔ時点以降の期待総利得の最大値である。本実施形態ではＤＰを利用しているので、このように部分最適解を再利用している。ｔ＝Ｔの場合には次の式（２０）でそのノードにおける最適な行動が求められる。

【００７０】

【数５】

$$d^*(x_T,T)=\arg\max_{y_T\in A(x_T)}\sum_{x_{T+1}}p(x_{T+1}\mid x_T,y_T,\theta^*)r(x_T,y_T,x_{T+1}),\quad(20)$$

ただし、ｄ＊（x_T，Ｔ）はＴ時点目の状態x_Tにおいて選択すべき最適な行動である。１≦ｔ≦Ｔ－１の場合には次の式（２１）でそのノードにおける最適な行動が求められる。

【００７１】

【数６】

$$d^*(x_t,t)=\arg\max_{y_t\in A(x_t)}\sum_{x_{t+1}}p(x_{t+1}\mid x_t,y_t,\theta^*)(r(x_t,y_t,x_{t+1})+V(x_{t+1},t+1)),\ (21)$$

ただし、ｄ＊（x_t，ｔ）はｔ時点目の状態x_tにおいて選択すべき最適な行動である。

【００７２】

その後、そのノードにおける最適な行動とそのノード以降の期待総利得の最大値が最適政策算出部３０のＤＰ実施器３０ｂへ出力され（ステップＳ４３）、前述の最適政策が出力されるのである。

【００７３】

以上説明したように、第１の実施形態によれば、ＤＰを用いて各時点のプレイヤーの各状態において、その時点以降の期待総利得を最大化し、最終的に制御期間における期待総利得を最大にすることが保証された政策が出力されるので、プレイヤーの初期状態と制御期間長とに対して制御期間における期待総利得を最大にする政策を出力することが可能となり、そのＲＰＧに関する攻略法（行動選択の仕方）を算出することができる。

【００７４】
即ち、本実施形態によれば、攻略法をコンピュータにシミュレーションさせることによって、被験者の体験データを取得することができる。例えば、プレイヤーのゲーム結果をシミュレーションすることによって、マップ上に隠されたアイテムやイベントに遭遇する割合等を把握でき、その割合を見ながら適切な隠し場所を設定する等のゲーム開発支援を行うことができる。また、算出される数理工学的に最適な攻略法と実際のプレイヤーによる攻略法との比較を行うことにより、人間が楽しいと感じるゲーム要素を工学的に把握することができる。人間が楽しいと感じるゲーム要素を確率モデル上のパラメータ設定のある種のパターンとして把握できれば、そのような要素を多く含むゲーム開発を行うことができる。また、種々の目的毎の攻略法を算出することによって、各目的に適したプレイヤーに協力するキャラクタの攻略法を容易にプログラミングすることが可能となる。
【００７５】
次に、各種確率分布を支配する真のパラメータθ＊が未知である、本発明を拡張した第２の実施形態について説明する。
【００７６】
本実施形態における攻略法算出装置の構成は基本的には、第１の実施形態の場合と同様であり、従ってその構成の説明は省略する。
【００７７】
真のパラメータ未知の場合を説明するために、いくつかの新たな定義を行う。ｐ（θ）はパラメータθの事前分布であり、既知であるとする。Θはパラメータ空間であり、θ＊∈Θ、θ∈Θである。$x^t y^{t-1}$はｔ期目の状態x_tに至るまでの遷移系列であり、$x^t y^{t-1} = x_1 y_1 \cdots x_t$である。
【００７８】
真のパラメータ既知の場合には、ＤＰでＴ時点から遡りながら各時点の各状態に対して行動選択を行うが、真のパラメータ未知の場合には、ＤＰでＴ時点から遡りながら各時点の各状態と１時点からその時点に至るまでの各遷移系列の組に対して行動選択を行う。
【００７９】
Ｔ時点目の状態x_T（全ての状態の候補）と遷移系列$x^T y^{T-1}$（全ての遷移系列の候補）の組に対する処理は以下の通りである。
【００８０】
【数７】

$$d_B^*(x_T, x^T y^{T-1}, T) = \arg \max_{y_T \in A(x_T)} \sum_{x_{T+1}} \int_\Theta p(\theta \mid x^T y^{T-1}) p(x_{T+1} \mid x_T, y_T, \theta) d\theta \; r(x_T, y_T, x_{T+1}), \quad (22)$$

$$V_B(x_T, x^T y^{T-1}, T) = \max_{y_T \in A(x_T)} \sum_{x_{T+1}} \int_\Theta p(\theta \mid x^T y^{T-1}) p(x_{T+1} \mid x_T, y_T, \theta) d\theta \; r(x_T, y_T, x_{T+1}), \quad (23)$$

ただし、ｐ（θ｜$x^T y^{T-1}$）は１時点からＴ時点に遷移系列$x^T y^{T-1}$のように遷移した場合の事後分布である。
【００８１】
ｔ時点目（１≦ｔ≦Ｔ－１）の状態x_t（全ての状態の候補）と遷移系列$x^t y^{t-1}$（全ての遷移系列の候補）の組に対する処理は以下の通りである。
【００８２】

【数８】

$$d_B^*(x_t, x^t y^{t-1}, t) = \arg \max_{y_t \in A(x_t)} \sum_{x_{t+1}} \int_\Theta p(\theta \mid x^t y^{t-1}) p(x_{t+1} \mid x_t, y_t, \theta) d\theta$$
$$(r(x_t, y_t, x_{t+1}) + V_B(x_{t+1}, x^{t+1} y^t, t+1)). \quad (24)$$

$$V_B(x_t, x^t y^{t-1}, t) = \max_{y_t \in A(x_t)} \sum_{x_{t+1}} \int_\Theta p(\theta \mid x^t y^{t-1}) p(x_{t+1} \mid x_t, y_t, \theta) d\theta$$
$$(r(x_t, y_t, x_{t+1}) + V_B(x_{t+1}, x^{t+1} y^t, t+1)). \quad (25)$$

式（２２）から式（２５）を用いてｄ＊Ｂ（ｘ１，ｘ１，１）まで求めることによって、１時点目からＴ時点目までの全ての状態と遷移系列の組とに対して、ベイズ基準のもとで総利得を最大にするという点で最適な行動選択の仕方を求めることができる。

【００８３】

式（２２）から式（２５）には積分計算が含まれており、一般的に、積分計算は計算量が多いが、二項分布（敵の出現以外の確率分布）の事前分布としてベータ分布を、多項分布（敵の出現の確率分布）の事前分布としてディリクレ分布をそれぞれ仮定すると、積分計算は四則演算で実施することができる（Ｍａｔｓｕｓｈｉｍａ，Ｔ．，Ｈｉｒａｓａｗａ，Ｓ．，Ａ Ｂａｙｅｓ ｃｏｄｉｎｇ ａｌｇｏｒｉｔｈｍ ｆｏｒ Ｍａｒｋｏｖ ｍｏｄｅｌｓ，ＴＥＣＨＮＩＣＡＬ ＲＥＰＯＲＴ ＯＦ ＩＥＩＣＥ，ＩＴ９５－１，ｐｐ．１－６（１９９５））。四則演算の一例として、マップモードのｔ期の状態ｘｔにおいて行動ｙｔを選択したもとで、敵ｅｉが出現し、戦闘モードの状態ｘｔ＋１に遷移する場合の

【数９】

$$\int_\Theta p(\theta \mid x^t y^{t-1}) p(x_{t+1} \mid x_t, y_t, \theta) d\theta$$

の計算を以下に示す。

【００８４】

【数１０】

$$\int_\Theta p(\theta \mid x^t y^{t-1}) p(x_{t+1} \mid x_t, y_t, \theta) d\theta = \int_\Theta p(\theta \mid x^t y^{t-1}) p(e_i | F(mv(x_{t,2}, y_t)), \theta) d\theta$$
$$= \frac{N(F(mv(x_{t,2}, y_t)), e_i \mid x^t y^{t-1}) + \alpha_1}{N(F(mv(x_{t,2}, y_t)) \mid x^t y^{t-1}) + \alpha_2}, \quad (26)$$

ただし，

$$\alpha_1 = \alpha(e_i \mid F(mv(x_{t,2}, y_t))), \quad (27)$$

$$\alpha_2 = \sum_{e_j \in E} \alpha(e_j \mid F(mv(x_{t,2}, y_t))) + \alpha(mv(x_{t,2}, y_t) \mid F(mv(x_{t,2}, y_t))), \quad (28)$$

ここで、Ｎ（Ｆ（ｍｖ（ｘｔ，２，ｙｔ）），ｅｉ｜ｘｔｙｔ－１）は系列ｘｔｙｔ－１中で地形の種類がＦ（ｍｖ（ｘｔ，２，ｙｔ））の位置で敵ｅｉが出現した回数、α（ｅｉ｜Ｆ（ｍｖ（ｘｔ，２，ｙｔ））はｐ（ｅｉ｜Ｆ（ｍｖ（ｘｔ，２，ｙｔ）），θ）に対するディリクレ分布（事前分布）のパラメータ、α（ｍｖ（ｘｔ，２，ｙｔ）｜Ｆ（ｍｖ（ｘｔ，２，ｙｔ））は

【数１１】

$$1 - \sum_{e_i \in E} p(e_i \mid F(mv(x_{t,2}, y_t)), \theta)$$

に対するディリクレ分布（事前分布）のパラメータを示す。このように、事前分布としてディリクレ分布やベータ分布を採用することにより、積分計算を四則演算で置き換えるこ

とができる。ディリクレ分布やベータ分布のパラメータの設定が事前分布の設定に相当するが、事前に何も情報が無い場合の設定の仕方についてはベイズ統計学やその応用分野で種々の方法が研究されている。多くの分野で良好な性質が報告されているジェフリーズの事前分布が有名であり、本実施形態に適用する場合は、各パラメータを０．５に設定することに相当する。例えば、Ｂｅｒｇｅｒ，Ｊ．Ｏ．，Ｓｔａｔｉｓｔｉｃａｌ Ｄｅｃｉｓｉｏｎ Ｔｈｅｏｒｙ ａｎｄ Ｂａｙｅｓｉａｎ Ａｎａｌｙｓｉｓ，Ｓｐｒｉｎｇｅｒ－Ｖｅｒｌａｇ，Ｎｅｗ Ｙｏｒｋ（１９８０）、繁桝算男，ベイズ統計入門，東京大学出版会（１９８５）、Ｍａｔｓｕｓｈｉｍａ，Ｔ．，Ｈｉｒａｓａｗａ，Ｓ．，Ａ Ｂａｙｅｓ ｃｏｄｉｎｇ ａｌｇｏｒｉｔｈｍ ｆｏｒ Ｍａｒｋｏｖ ｍｏｄｅｌｓ，ＴＥＣＨＮＩＣＡＬ ＲＥＰＯＲＴ ＯＦ ＩＥＩＣＥ，ＩＴ９５－１，ｐｐ．１－６（１９９５）、鈴木譲，ベイシアネットワーク入門，培風館，東京（２００９）を参照。事前分布にディリクレ分布やベータ分布を採用してジェフリーズの事前分布に設定し、式（２２）から式（２５）で処理することにより、真のパラメータ未知の場合にベイズ基準のもとで総利得を最大化することができる。

【００８５】

以上の説明では、真のパラメータ未知の場合のベイズ最適な行動選択の仕方を求めるアルゴリズムについて述べた。事前分布としてディリクレ分布やベータ分布を採用することにより、積分計算を四則演算に置き換えた。しかしながら、ベイズ最適な行動選択の仕方を求めるためには、多大な計算量が必要となる。真のパラメータ既知の場合には、ＤＰの各時点毎に式（１９）の処理を状態数分だけ実施すればよい。一方、真のパラメータ未知のベイズ最適の場合には、ＤＰの各時点毎に式（２５）の処理を状態数と遷移系列の個数との積分だけ実施する必要があり、処理の回数は時点の数（ｔ期のｔ）に対する指数オーダとなる。

【００８６】

そこで、本実施形態の望ましい態様として、真のパラメータ未知の場合に近似を行う例を説明する。ここで、学習データＬを新たに導入する。学習データは過去のゲームのプレイデータである遷移系列の集合であったり、敵の出現確率などの個々の確率分布について真のパラメータの分布から発生させたサンプルデータであったり、種々の形態が適用可能である。

【００８７】

前述したベイズ最適な方法では、各時点毎にその時点までの遷移系列$x^t y^{t-1}$に対する事後分布によって、

【数１２】

$$\int_\Theta p(\theta \mid x^t y^{t-1}) p(x_{t+1} \mid x_t, y_t, \theta) d\theta$$

を計算したが、近似アルゴリズムでは、時点に関係なく学習データＬによる事後分布を用いて

【数１３】

$$\int_\Theta p(\theta \mid L) p(x_{t+1} \mid x_t, y_t, \theta) d\theta$$

を計算する。具体的には、

【数１４】

$$\int_\Theta p(\theta \mid L) p(x_{t+1} \mid x_t, y_t, \theta) d\theta$$

を真のパラメータ既知の場合の式（１８）から式（２１）に代入して行動選択の仕方を求める。

【００８８】

近似アルゴリズムにより、ＤＰの処理の回数は真のパラメータ既知の場合と同じ回数に軽減することができる。有限の学習データに対する理論的な精度保証はないが、漸近的には学習データによる事後分布を用いた推定値が真のパラメータに収束するので、求める行

動選択の仕方も真のパラメータ既知の場合に収束する。

【実施例】

【００８９】

真のパラメータθ＊が既知の場合の行動選択の仕方を求めるアルゴリズムに関する第１の実施形態について実験を行った実施例を説明する。図６にこの実施例におけるマップを示す。実験結果が理解しやすいように９マスからなる小規模のマップとした。以下の表１～表５は地形の設定、確率の設定及びその他の設定を示している。

【００９０】

【表１】

表１　地形の設定

$F(sm_1)$	$F(sm_2)$	$F(sm_3)$	$F(sm_4)$	$F(sm_5)$
f_1	f_2	f_3	f_2	f_2

$F(sm_6)$	$F(sm_7)$	$F(sm_8)$	$F(sm_9)$
f_3	f_3	f_3	f_3

【００９１】

【表２】

表２　確率の設定（その１）

$p(e_1 \mid f_2, \theta^*)$	$p(e_2 \mid f_2, \theta^*)$	$p(e_1 \mid f_3, \theta^*)$	$p(e_2 \mid f_3, \theta^*)$
０．３	０．０	０．０	０．８

【００９２】

【表３】

表３　確率の設定（その２）

$p(C(e_1) \mid a_5, e_1, \theta^*)$	$p(C(e_2) \mid a_5, e_2, \theta^*)$	$p(B(e_1) \mid e_1, \theta^*)$
０．９	０．９	０．６

$p(B(e_2) \mid e_2, \theta^*)$	$p(map \mid a_6, \theta^*)$
０．６	０．６

【００９３】

【表４】

表４　その他の設定（その１）

M_{hp}	E	$M(e_1)$	$M(e_2)$	$G(e_1)$	$G(e_2)$
１０	$\{e_1, e_2\}$	２	４	１	５

【００９４】

【表５】

表５　その他の設定（その２）

$C(e_1)$	$C(e_2)$	$B(e_1)$	$B(e_2)$	T
３	３	１	３	１０

【００９５】

以上の設定で、真のパラメータ既知の場合のアルゴリズムを適用して、１０期間の期待総利得を最大化するための各期における行動選択の仕方を求めた。

【００９６】

結果の一部について述べると、例えば、時点３（３期）でプレイヤーのＨＰが６であり位置sm_4にいるマップモードの状態では、最適の行動選択はa_3という上のマスsm_7への移動であった。これは、プレイヤーのＨＰにまだ余裕があるので、弱くて報酬の小さい敵が出現する右（a_1）やＨＰを回復する下（a_4）ではなく、強くて報酬の大きい敵が出現する上（a_3）への移動を選択しているのである。他方、同じ時点３（３期）のマップモードであってもＨＰが１であり位置sm_4にいる状態では、行動a_4を選択して下のＨＰを回復してくれるスタート位置sm_1へ移動した。これは、プレイヤーのＨＰに余裕がないので回復するための行動選択である。また、時点９（９期）にＨＰが１であり位置sm_4にいるマップモードの状態では、行動a_1を選択して、弱くて報酬の小さい敵が出現する右のsm_5へ移動した。これは、プレイヤーのＨＰに余裕はないが、残りの期間にも余裕がないため、弱い敵が出現するsm_5への移動を選択しているのである。

【００９７】

このように、真のパラメータ既知の場合のアルゴリズムを適用することにより、開発者であれば知っている真のパラメータの情報を利用して対象期間の期待総利得を最大にする行動選択の仕方を求めることができる。

【００９８】

次に、真のパラメータが未知の場合の行動選択の仕方を求めるアルゴリズムに関する第２の実施形態について実験を行った実施例を説明する。真のパラメータ既知の場合の実施例と同じ設定のもとで、真のパラメータ未知の場合の近似アルゴリズムを適用した。学習データとして、各確率分布毎にサンプルデータを発生させた。そのもとで、近似アルゴリズムを適用し、各時点の各状態における行動選択が真のパラメータ既知の場合と比較して一致するかどうか調べた。各確率分布の学習データ数を１０、１００、１０００と変化させ、それぞれの学習データ数に対して１００パターンの学習データを発生させて適用実験を行った。真のパラメータ既知の場合の行動選択との一致率は１００パターンの平均で、学習データ数が１０の場合で約８８％、学習データ数が１００の場合で約９４％、学習データ数が１０００の場合で約９６％であった。

【００９９】

このように、学習データによる事後分布を利用する近似アルゴリズムを用いることにより、真のパラメータが未知の場合でも、ＤＰに必要な計算量を、真のパラメータ既知の場合の計算量と同程度とすることができる。また、少ない実験例ではあるが、学習データの増加に伴い真のパラメータ既知の場合との行動選択の一致率が高くなることが確認できた。

【０１００】

以上述べた実施形態は全て本発明を例示的に示すものであって限定的に示すものではなく、本発明は他の種々の変形態様及び変更態様で実施することができる。従って本発明の範囲は特許請求の範囲及びその均等範囲によってのみ規定されるものである。

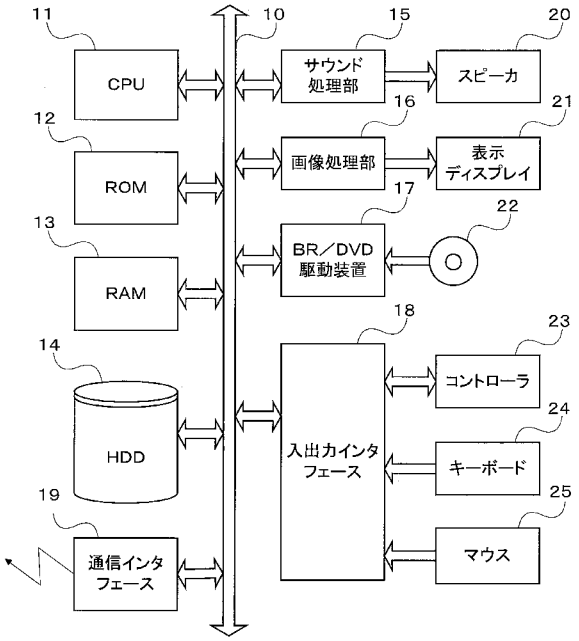
【符号の説明】

【０１０１】

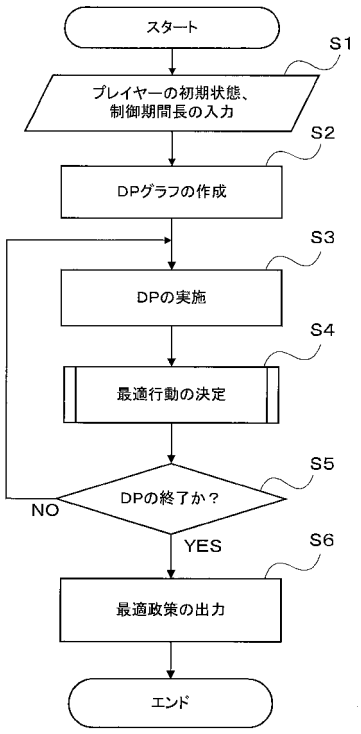
１０　バス
１１　ＣＰＵ
１２　ＲＯＭ
１３　ＲＡＭ
１４　ＨＤＤ
１５　サウンド処理部
１６　画像処理部
１７　ＢＲ／ＤＶＤ駆動装置
１８　入出力インタフェース
１９　通信インタフェース
２０　スピーカ
２１　表示ディスプレイ

２２　ＢＲ／ＤＶＤ／ＣＤ
２３　コントローラ
２４　キーボード
２５　マウス
３０　最適政策算出部
３０ａ　ＤＰグラフ作成器
３０ｂ　ＤＰ実施器
３１　行動決定部
３１ａ　行動決定器
３１ｂ　遷移確率テーブル
３１ｃ　利得テーブル

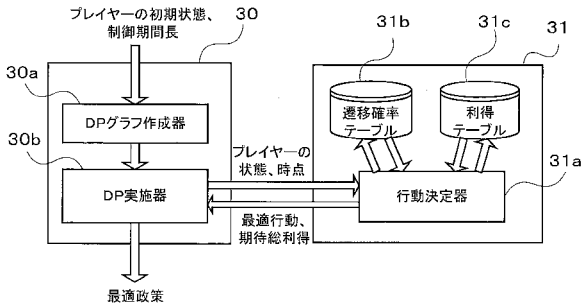
【図１】



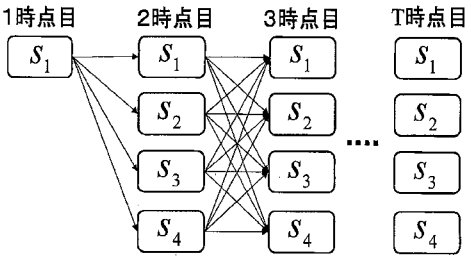
【図２】



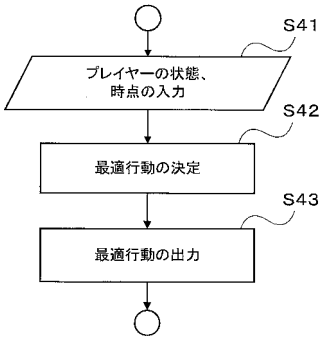
【図３】



【図４】



【図５】



【図６】

sm_7	sm_8	sm_9
sm_4	sm_5	sm_6
sm_1	sm_2	sm_3