(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7548327号
(P7548327)

(45)発行日　令和6年9月10日(2024.9.10)

(24)登録日　令和6年9月2日(2024.9.2)

(51)国際特許分類

		FI		
G08G 1/00 (2006.01)		G08G	1/00	D
G08G 1/16 (2006.01)		G08G	1/16	A

請求項の数　10　（全18頁）

(21)出願番号　特願2022-558680(P2022-558680)
(86)(22)出願日　令和2年10月28日(2020.10.28)
(86)国際出願番号　PCT/JP2020/040484
(87)国際公開番号　WO2022/091261
(87)国際公開日　令和4年5月5日(2022.5.5)
審査請求日　令和5年4月13日(2023.4.13)

(73)特許権者　000004237
日本電気株式会社
東京都港区芝五丁目７番１号
(74)代理人　100103090
弁理士　岩壁 冬樹
(74)代理人　100124501
弁理士　塩川 誠人
(72)発明者　比嘉 亮太
東京都港区芝五丁目７番１号 日本電気株式会社内
審査官　武内 俊之

最終頁に続く

(54)【発明の名称】　評価システム、評価方法および評価プログラム

(57)【特許請求の範囲】

【請求項１】

ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、または当該値の累積和の期待値を最大化するように、当該行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習する学習手段と、

交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または前記対象資源を他エージェントから譲られた場合のミッションの効用を、前記計画評価関数を用いて算出される前記内部価値の差異を定義した効用関数を用いて評価する評価手段とを備え、

前記学習手段は、前記ミッション評価関数を報酬関数として用いて、方策関数および状態価値関数を学習し、

前記評価手段は、前記状態価値関数を前記計画評価関数とする効用関数を用いて効用を評価する

ことを特徴とする評価システム。

【請求項２】

ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、または当該値の累積和の期待値を最大化するように、当該行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習する学習手段と、

交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または前記対象資源を他エージェントから譲られた場合のミッションの効用を、前記計画評価関数を用

いて算出される前記内部価値の差異を定義した効用関数を用いて評価する評価手段とを備え、
前記学習手段は、前記ミッション評価関数を報酬関数として用いて、状態行動価値関数を学習し、学習された前記状態行動価値関数を用いて方策関数および状態価値関数を生成し、
前記評価手段は、生成された前記状態価値関数を前記計画評価関数とする効用関数を用いて効用を評価する
ことを特徴とする評価システム。
【請求項３】
評価手段は、他エージェントから要求される対象資源を使用しなかった場合のミッションの内部価値を計画評価関数を用いて算出し、当該対象資源を他エージェントに譲った場合の効用を効用関数を用いて評価する
請求項１または請求項２記載の評価システム。
【請求項４】
評価手段は、自エージェントが要求する対象資源を使用した場合のミッションの内部価値を計画評価関数を用いて算出し、当該対象資源を他エージェントから譲られた場合の効用を効用関数を用いて評価する
請求項１から請求項３のうちのいずれか１項に記載の評価システム。
【請求項５】
計画評価関数が評価する状態に、移動体の位置情報を含み、
効用の引数になる対象資源に、ある時刻における移動体の位置情報を含む
請求項１から請求項４のうちのいずれか１項に記載の評価システム。
【請求項６】
計画評価関数が評価する状態に、交渉に直接依存しない情報を含む
請求項１から請求項５のうちのいずれか１項に記載の評価システム。
【請求項７】
ミッション評価関数は、価値の算出に用いる項として行動に対する対価を含む
請求項１から請求項６のうちのいずれか１項に記載の評価システム。
【請求項８】
コンピュータが、ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、または当該値の累積和の期待値を最大化するように、当該行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習し、
前記コンピュータが、交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または前記対象資源を他エージェントから譲られた場合のミッションの効用を、前記計画評価関数を用いて算出される前記内部価値の差異を定義した効用関数を用いて評価し、
前記学習の際、前記コンピュータが、前記ミッション評価関数を報酬関数として用いて、方策関数および状態価値関数を学習し、
前記コンピュータが、前記状態価値関数を前記計画評価関数とする効用関数を用いて効用を評価する
ことを特徴とする評価方法。
【請求項９】
コンピュータが、ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、または当該値の累積和の期待値を最大化するように、当該行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習し、
前記コンピュータが、交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または前記対象資源を他エージェントから譲られた場合のミッションの効用を、前記計画評価関数を用いて算出される前記内部価値の差異を定義した効用関数を用いて評価し、
前記学習の際、前記コンピュータが、前記ミッション評価関数を報酬関数として用いて、状態行動価値関数を学習し、学習された前記状態行動価値関数を用いて方策関数および状

態価値関数を生成し、
前記コンピュータが、生成された前記状態価値関数を前記計画評価関数とする効用関数を用いて効用を評価する
ことを特徴とする評価方法。
【請求項１０】
コンピュータに、
ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、または当該値の累積和の期待値を最大化するように、当該行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習する学習処理、および、
交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または前記対象資源を他エージェントから譲られた場合のミッションの効用を、前記計画評価関数を用いて算出される前記内部価値の差異を定義した効用関数を用いて評価する評価処理を実行させ、
前記学習処理で、前記ミッション評価関数を報酬関数として用いて、方策関数および状態価値関数を学習させ、
前記評価処理で、前記状態価値関数を前記計画評価関数とする効用関数を用いて効用を評価させる
ための評価プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、自エージェントが行う計画を評価する評価システム、評価方法および評価プログラムに関する。
【背景技術】
【０００２】
近年、自動交渉の重要性がより高まっている。例えば、ドローンなどの無人航空機（Unmanned Aircraft Vehicles，ＵＡＶ）を活用する場面において、ＵＡＶが安全に飛行するための運航経路や運航領域を確保もしくは提供する仕組みとして、自動交渉の技術は有効である。
【０００３】
例えば、運航に使用される領域に対して他エージェントとの間でインタラクションがある場合、他エージェントとの間で移動に使用する領域を共有し取り合う必要がある場合にも、自身にとって不利にならない選択をするように交渉することが重要になる。
【０００４】
特許文献１には、移動体の運航に用いられる領域を評価する領域評価システムが記載されている。特許文献１に記載された領域評価システムでは、第１の領域を使用する第１のミッションと、第１の領域を使用しない第２のミッションとが与えられたとき、第１の領域の効用を、第１のミッションの効用と第２のミッションの効用の差異に基づいて評価する。
【０００５】
なお、非特許文献１には、自動交渉を行うプロトコルの一例であるＡＯＰ（Alternating Offers Protocols）について記載されている。
【先行技術文献】
【特許文献】
【０００６】
【文献】国際公開第２０１９／１６７１６１号
【非特許文献】
【０００７】
【文献】Reyhan A, et al., "Alternating Offers Protocols for Multilateral Negotiation", Modern Approaches to Agent-based Complex Automated Negotiation, p

p.153-167, April 2017.
【発明の概要】
【発明が解決しようとする課題】
【０００８】
上述するように、例えば、他エージェントとの間で領域の排他制御を行う運航管理システムに運航計画の申請を行う場面や、他エージェントと交渉して領域の使用権を譲渡等する場面を考える。このような場面では、自身にとって不利にならないように、ミッション内容およびそれに含まれる各移動体の運航経路ひいてはその運航経路で使用する領域を適切に評価した上で、ミッション内容を選択することが必要になる。
【０００９】
また、計画済みのミッション内容で使用するある領域を評価した結果、その領域を使用すると自身にとって不利になるようであれば、ミッション計画に立ち戻ってより適切なミッションを決定する、すなわちミッションを再計画することが求められる。
【００１０】
このように、他エージェントとの間でインタラクションがある場合、自身にとって不利な選択をしないために、ミッションレベルでの効用、より具体的にはそのミッションに含まれる全てのタスクおよびそれに基づく効用を勘案して評価をする必要がある。
【００１１】
本発明では、ミッションを、「エージェントが管理する１つ以上のタスク（務め）に対して実行性を持たせたもの（具体的な計画群）」と定義する。例えば、上述する運行計画の場合、ミッションを「エージェントが管理する１つ以上のタスクの各々を実行可能なように資源（移動体資源、領域資源等）の割り当てを行って得られる運航計画群」と定義することが可能である。
【００１２】
また、上述する運行計画において、タスクを、移動体の運航を伴うタスクであって、空間と時間の指定を少なくとも含むものと定義できる。なお、あるタスクに対して、移動体資源の割り当ておよび割り当てられた各移動体資源に対する運航経路の決定により得られる運航計画群は、ミッションの一例である。
【００１３】
特許文献１に記載された領域評価システムを用いることで、例えば、運航に使用される領域に対して他エージェントとの間でインタラクションがある場合にも、自エージェントにとっての領域の評価を適切に行うことは可能である。
【００１４】
一方、特許文献１に記載された領域評価システムでは、ミッションの報酬およびコストを引数として予め定められた関数に基づいてミッションの効用を評価している。しかし、ミッションの効用は、各エージェントが計画する内容によって異なると考えられる。そのため、ミッションを既存の報酬関数に基づいて評価するだけではなく、各エージェント内部の計画等を反映した独自の評価を行えることが好ましい。
【００１５】
そこで、本発明は、計画するミッションについて、他エージェントとの間にインタラクションがある場合にも、自エージェントにとっての効用を適切に評価できる評価システム、評価方法および評価プログラムを提供することを目的とする。
【課題を解決するための手段】
【００１６】
本発明による評価システムは、ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、またはその累積和の期待値を最大化するように、上記行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習する学習手段と、交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または対象資源を他エージェントから譲られた場合のミッションの効用を、計画評価関数を用いて算出される内部価値の差異を定義した効用関数を用いて評価する評価手段

とを備え、学習手段が、ミッション評価関数を報酬関数として用いて、方策関数および状態価値関数を学習し、評価手段が、状態価値関数を計画評価関数とする効用関数を用いて効用を評価することを特徴とする。また、本発明による他の評価システムは、ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、またはその累積和の期待値を最大化するように、上記行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習する学習手段と、交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または対象資源を他エージェントから譲られた場合のミッションの効用を、計画評価関数を用いて算出される内部価値の差異を定義した効用関数を用いて評価する評価手段とを備え、学習手段が、ミッション評価関数を報酬関数として用いて、状態行動価値関数を学習し、学習された状態行動価値関数を用いて方策関数および状態価値関数を生成し、評価手段が、生成された状態価値関数を計画評価関数とする効用関数を用いて効用を評価することを特徴とする。

【００１７】

本発明による評価方法は、コンピュータが、ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、またはその値の累積和の期待値を最大化するように、上記行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習し、コンピュータが、交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または対象資源を他エージェントから譲られた場合のミッションの効用を、計画評価関数を用いて算出される内部価値の差異を定義した効用関数を用いて評価し、学習の際、コンピュータが、ミッション評価関数を報酬関数として用いて、方策関数および状態価値関数を学習し、コンピュータが、状態価値関数を計画評価関数とする効用関数を用いて効用を評価することを特徴とする。また、本発明による他の評価方法は、コンピュータが、ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、またはその値の累積和の期待値を最大化するように、上記行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習し、コンピュータが、交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または対象資源を他エージェントから譲られた場合のミッションの効用を、計画評価関数を用いて算出される内部価値の差異を定義した効用関数を用いて評価し、学習の際、コンピュータが、ミッション評価関数を報酬関数として用いて、状態行動価値関数を学習し、学習された状態行動価値関数を用いて方策関数および状態価値関数を生成し、コンピュータが、生成された状態価値関数を計画評価関数とする効用関数を用いて効用を評価することを特徴とする。

【００１８】

本発明による評価プログラムは、コンピュータに、ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、またはその値の累積和の期待値を最大化するように、上記行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習する学習処理、および、交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または対象資源を他エージェントから譲られた場合のミッションの効用を、計画評価関数を用いて算出される内部価値の差異を定義した効用関数を用いて評価する評価処理を実行させ、学習処理で、ミッション評価関数を報酬関数として用いて、方策関数および状態価値関数を学習させ、評価処理で、状態価値関数を計画評価関数とする効用関数を用いて効用を評価させることを特徴とする。

【発明の効果】

【００１９】

本発明によれば、計画するミッションについて、他エージェントとの間にインタラクションがある場合にも、自エージェントにとっての効用を適切に評価できる。

【図面の簡単な説明】

【００２０】

【図１】本発明による交渉システムの一実施形態の構成例を示すブロック図である。

【図２】計画部の動作例を示すフローチャートである。

【図３】交渉システムの動作例を示すフローチャートである。
【図４】交渉システムの他の動作例を示すフローチャートである。
【図５】本発明による評価システムの概要を示すブロック図である。
【図６】少なくとも１つの実施形態に係るコンピュータの構成を示す概略ブロック図である。
【発明を実施するための形態】
【００２１】
以下、本発明の実施形態を図面を参照して説明する。
【００２２】
図１は、本発明による交渉システムの一実施形態の構成例を示すブロック図である。本実施形態の交渉システム１００は、ミッション評価部１０と、計画部２０と、交渉部４０とを備えている。
【００２３】
ミッション評価部１０は、経済的または業務的な価値（以下、報酬（Ｒｅｗａｒｄ）と記すこともある。）の観点でミッションを評価する関数（以下、ミッション評価関数と記す。）を計画部２０に入力する。
【００２４】
ミッション評価関数は、ある状態ｓ（ｓｔａｔｅ）での自エージェントの行動ａ（Ａｃｔｉｏｎ）の（ミッションに対する）価値（すなわち、報酬）を算出する関数である。この行動は、自エージェントの意思決定（ポリシ）の結果であり、強化学習の場合、方策関数に基づく結果と言うことができる。言い換えると、ミッション評価関数は、ある状態ｓで行動ａをとり、次の局面ｓ´に移ったときに得られる報酬を導出する関数である。
【００２５】
より具体的には、ミッション評価関数は、数理最適化（ＯＲ：Operations Research）の目的関数に対応し、強化学習で用いられる場合には、報酬関数に対応する。本実施形態のミッション評価関数は、業界専門のオペレータ等により、予め設定される。ミッション評価関数の態様は任意であり、例えば、上述する運航計画の場合、経路のコストｃとタスクの報酬ｒを用いて、ミッション評価関数ｆが、ｆ＝ｒ－ｃと定義されてもよい。
【００２６】
また、ミッション評価関数に、価値の算出に用いる項として、行動に対する対価が含まれていてもよい。例えば、繰り返し交渉を行う状況において、他エージェントから行動に対する対価を得られる場合、ミッション評価関数に対価の項を含めることにより、この対価を価値関数へフィードバックすることが可能になる。
【００２７】
これは、自エージェントを想定した価値関数に、他エージェントの情報を一部含めて更新する機能に対応すると言える。そのため、環境情報に相手の情報を反映させて強化学習を行う、いわゆるマルチエージェント強化学習に対応させることも可能になる。
【００２８】
計画部２０は、エージェントごとに独自に計画を行う装置である。計画部２０は、入力部２２と、学習部２４と、計画評価関数格納部２６と、効用関数格納部２８と、評価部３０とを含む。
【００２９】
入力部２２は、報酬関数の入力を受け付ける。なお、入力部２２は、後述する計画評価関数の入力を受け付けて、計画評価関数格納部２６に記憶してもよい。また、入力部２２は、例えば、オペレータや他の装置から、交渉による報酬に関する情報の入力を受け付けてもよい。
【００３０】
学習部２４は、上記行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する関数（以下、計画評価関数と記す。）を学習する。本実施形態の学習部２４は、入力されたミッション評価関数の値、または、その値の累積和の将来に渡る期待

値を最大化するように計画評価関数を学習する。

【００３１】

学習部２４は、任意の機械学習により計画評価関数を学習してもよい。具体的には、学習部２４は、状態ｓと行動ａとを明確に分離したうえで、時系列に数理モデル化したＭＤＰ（マルコフ決定過程：Markov decision process ）を用いて強化学習を行うことにより、計画評価関数を生成してもよい。

【００３２】

学習部２４は、例えば、ミッション評価関数を報酬関数として用いた強化学習により、計画評価関数として、方策関数π（a_t｜s_t）および状態価値関数Ｖ（s_t）を生成してもよい。状態s_tの一例として、ある時刻ｔにおける移動体の地図上の位置や、天候、風向、ミッションのコストなどを示す情報が挙げられる。すなわち、状態s_tには、ある時刻における位置情報のような交渉に直接依存する情報だけでなく、天候やミッションのコストのような、交渉に直接依存しない情報が含まれていてもよい。また、行動a_tの一例として、ある時刻ｔにおいてどの方向に移動するかを示す情報が挙げられる。

【００３３】

また、学習部２４は、ミッション評価関数を報酬関数として用いた強化学習により、計画評価関数として、状態行動価値関数Ｑ（s_t，a_t）を生成し、生成された状態行動価値関数Ｑ（s_t，a_t）を用いて方策関数π（ａ｜ｓ）および状態価値関数Ｖ（ｓ）を生成してもよい。

【００３４】

以下、状態行動価値関数Ｑ（s_t，a_t）に基づいて方策関数π（ａ｜ｓ）および価値関数Ｖ（ｓ）を生成する方法の一例を説明する。例えば、０≦ε≦１とした場合、行動を生成する方策関数を、以下に例示する式１のように定義できる。

【００３５】

【数１】

$$\pi(a|s)=\begin{cases}1-\varepsilon+\frac{\varepsilon}{|\mathcal{A}|} & (a=\underset{a}{\operatorname{argmax}}\,Q(s,a))\\ \frac{\varepsilon}{|\mathcal{A}|} & (a\neq\underset{a}{\operatorname{argmax}}\,Q(s,a))\end{cases} \qquad (式１)$$

【００３６】

他にも、例えば、β≧０とした場合、方策関数を以下に例示する式２のようにも定義できる。

【００３７】

【数２】

$$\pi(a|s)=\frac{\exp(\beta Q(s,a))}{\sum_a \exp(\beta Q(s,a))} \qquad (式２)$$

【００３８】

また、上述する方策関数を用いて、価値関数を以下に例示する式３のように定義できる。

【００３９】

【数３】

$$V(s)=\sum_a \pi(a|s)Q(s,a) \qquad (式３)$$

【００４０】

さらに、全時間にわたる累積和（すなわち、報酬の総和）Ｇは、以下に例示する式４で

定義できる。なお、式４をＭＤＰにおける時系列モデルで期待値を算出したものが価値関数（Ｑ（ｓ，ａ）、Ｖ（ｓ））になる。

【００４１】

【数４】

$$G = \sum_{t=1}^{T} r(s_{t-1}, a_{t-1}) \qquad \text{（式４）}$$

【００４２】

なお、方策関数π（ａ｜ｓ）および価値関数Ｖ（ｓ）の定義方法は、上述する方法に限定されない。例えば、関数近似を用いて、方策関数および価値関数がニューラルネットに置き換えられも良い。また、他にも、Ｑ関数を用いて、方策関数および価値関数が定義されてもよい。すなわち、学習部２４がＱ関数を学習し、学習されたＱ関数を用いて、方策関数および価値関数が定義されてもよい。

【００４３】

また、学習部２４は、時間発展を表わす関数である状態遷移モデルを用いて、計画評価関数を学習してもよい。状態遷移モデルは、例えば、以下に例示する式５で表わすことができる。なお、学習部２４は、状態遷移モデルの他、シミュレータや、予め定められた状態遷移を表わす関数を用いて、計画評価関数を学習してもよい。

【００４４】

ｐ（ｓｔ＋１｜ｓｔ，ａｔ）　（式５）

【００４５】

状態遷移モデルを用いて計画評価関数を学習することで、一時的な価値だけでなく、全体としての予測を反映した価値を同時に評価できる。例えば、上述する運航計画の場合、ある地点の価値だけでなく、経路情報によって分かる価値と経路を同時に生成できる。すなわち、逐次予測に基づく意思決定がＭＤＰの特色であることから、現実の問題に対して長期予測を反映することにより経路交渉を最適化できる。なお、状態遷移モデルは、機械学習によって生成されてもよく、外部のシミュレータや、与えられた関数、データベース等が状態遷移モデルとして用いられてもよい。例えば、運航計画を行う場合、外部のエンジンから情報を参照することにより、シミュレータとしての機能を実現してもよい。

【００４６】

また、学習部２４は、後述する交渉部４０による交渉の結果を反映して学習を行ってもよい。学習部２４は、例えば、ミッション評価関数と交渉による報酬の和を再度報酬関数として定義し、定義された報酬関数を用いて学習を行ってもよい。交渉による報酬として、例えば、入力部２２に入力される情報が用いられてもよい。これにより、これまで独立に考えられていた、経路計画と交渉で得られる報酬とを連動させたモデルを生成することが可能になる。

【００４７】

例えば、同じような業務を繰り返す場合、予め交渉の起こりやすさを知っていることは、各エージェントにとって有効な情報になると考えられる。そのため、例えば、相手がオペレータ（人）の場合、交渉の結果をモデルに反映することで、よく調整する会社との取引を効率よく進めることも可能になる。

【００４８】

このように、学習部２４は、ミッション評価関数を報酬として、各エージェントの内部の計画を最大化するような計画評価関数を計算する。そして、学習部２４は、生成した計画評価関数（より具体的には、Ｑ、Ｖなどの価値関数、方策関数、状態遷移関数等）を計画評価関数格納部２６に記憶する。

【００４９】

計画評価関数格納部２６は、学習部２４による機械学習の結果生成された計画評価関数

および入力部２２により受け付けられた計画評価関数の少なくとも一方を記憶する。計画評価関数格納部２６は、例えば、磁気ディスク等により実現される。

【００５０】

効用関数格納部２８は、計画評価関数を用いてミッションの効用を定義した関数（以下、効用関数と記す。）を記憶する。例えば、時刻ｔにおけるエージェントの状態をs_tとした場合、効用関数Ｕ＝Ｖ（s_t）と定義されてもよい。

【００５１】

なお、本実施形態では、特に、交渉の対象候補になる資源（以下、対象資源と記す。）ωを他エージェントに譲った場合、または、対象資源を他エージェントから譲られた場合のミッションの効用を示す効用関数を、計画評価関数を用いて算出される内部価値の差異で定義する。

【００５２】

対象資源は、上述するように、交渉の対象とする資源であって、他のエージェントと排他的に使用される資源（例えば、時間、空間、等）である。時刻ｔにおける状態および対価を、それぞれs_t，r_tとした場合、対象資源ωをω：＝（s_t，r_t）と定義できる。状態s_tには、例えば、ある時刻ｔにおける移動体の位置情報が含まれる。

【００５３】

また、他エージェントとの交渉がない場合の内部価値Ｖ（ｓ）をオリジナルの内部価値と記す。このとき、オリジナルの内部価値Ｖ（ｓ）をベースラインとすると、効用関数Ｕ（ω）を、以下に例示する式６のように定義できる。

【００５４】

$$U(s_t, r_t) := V(s_t) - V(s) + r_t \quad \text{（式６）}$$

【００５５】

上述する式６は、交渉によって調整される内部価値（ω）とオリジナルの内部価値との差分に、支払い（ω）を加算した式と言うことができる。また、効用を内部価値の比で扱えるようにするため、効用関数Ｕを、以下に例示する式７のように定義してもよい。

【００５６】

$$\ln V(s_t) - \ln V(s) + \ln(r_t) = \ln(V(s_t)\, r_t / V(s)) \quad \text{（式７）}$$

【００５７】

評価部３０は、効用関数格納部２８から効用関数を取得し、取得した効用関数を用いてミッションの効用を評価する。具体的には、評価部３０は、対象資源を他エージェントに譲った場合、または、対象資源を他エージェントから譲られた場合のミッションの効用を効用関数を用いて評価し、効用を得る。評価部３０は、例えば、上記の式３に示すような状態価値関数を計画評価関数として用いた効用関数を用いて効用を評価してもよい。

【００５８】

なお、評価部３０が評価を行う状況として、上述するように、対象資源を他エージェントに譲った場合、または、対象資源を他エージェントから譲られた場合の二通りが考えられる。どちらの状況かは、例えば、後述する交渉部４０の指示に応じて判断してもよい。

【００５９】

具体的には、対象資源を他エージェントに譲った場合、自エージェントからすれば、他エージェントから要求される対象資源を使用しない状況に対応する。そこで、評価部３０は、計画評価関数を用いて、他エージェントから要求される対象資源を使用しなかった場合のミッションの内部価値を算出する。ここで、算出される内部価値は、対象資源の使用が制限されることから、オリジナルの内部価値よりも低く算出されることが想定される。そして、評価部３０は、その対象資源を他エージェントに譲った場合の効用を、効用関数を用いて評価する。

【００６０】

一方、対象資源を他エージェントから譲られた場合、自エージェントからすれば、自エージェントが要求する対象資源を使用できる状況に対応する。そこで、評価部３０は、計

画評価関数を用いて、自エージェントが要求する対象資源を使用した場合のミッションの内部価値を算出する。ここで、算出される内部価値は、使用できる対象資源が増えることから、オリジナルの内部価値よりも高く算出されることが想定される。そして、評価部３０は、その対象資源を他エージェントから譲られた場合の効用を、効用関数を用いて評価する。

【００６１】

なお、評価部３０が、過去の実データやシミュレーションデータを蓄積するデータベース（図示せず）を保持している場合、このデータベースに、（s_t，a_t，r_t，π_t，V_t）のようなデータを問い合わせて情報を取得することで、効用値を算出してもよい。

【００６２】

なお、上記に示すように、本実施形態の計画部２０は、自エージェントにとっての効用を評価していることから、計画部２０のことを、評価システムと言うことができる。

【００６３】

入力部２２と、学習部２４と、評価部３０とは、プログラム（評価プログラム）に従って動作するコンピュータのプロセッサ（例えば、ＣＰＵ（Central Processing Unit ）、ＧＰＵ（Graphics Processing Unit））によって実現される。

【００６４】

例えば、プログラムは、計画部２０の記憶部（図示せず）に記憶され、プロセッサは、そのプログラムを読み込み、プログラムに従って、入力部２２、学習部２４および評価部３０として動作してもよい。また、計画部２０の機能がＳａａＳ（Software as a Service ）形式で提供されてもよい。

【００６５】

また、入力部２２と、学習部２４と、評価部３０とは、それぞれが専用のハードウェアで実現されていてもよい。また、各装置の各構成要素の一部又は全部は、汎用または専用の回路（circuitry ）、プロセッサ等やこれらの組合せによって実現されもよい。これらは、単一のチップによって構成されてもよいし、バスを介して接続される複数のチップによって構成されてもよい。各装置の各構成要素の一部又は全部は、上述した回路等とプログラムとの組合せによって実現されてもよい。

【００６６】

また、計画部２０の各構成要素の一部又は全部が複数の情報処理装置や回路等により実現される場合には、複数の情報処理装置や回路等は、集中配置されてもよいし、分散配置されてもよい。例えば、情報処理装置や回路等は、クライアントサーバシステム、クラウドコンピューティングシステム等、各々が通信ネットワークを介して接続される形態として実現されてもよい。

【００６７】

計画評価関数格納部２６および効用関数格納部２８は、例えば、磁気ディスク等により実現される。

【００６８】

交渉部４０は、計画部２０によって算出された効用に基づいて、他エージェントとの交渉を行う。具体的には、他エージェントから対象資源の要求を受けた場合、交渉部４０は、他エージェントからの要求を計画部２０に通知し、算出された効用を計画部２０から受信する。そして、算出された効用が予め定めた閾値を超える場合、交渉部４０は、他エージェントからの対象資源の要求を許容すると決定してもよい。

【００６９】

また、自エージェントが他エージェントに対して対象資源の使用を提案する場合、交渉部４０は、提案する要求を計画部２０に通知し、算出された効用を計画部２０から受信する。なお、提案する要求は、例えば、オペレータ等により交渉部４０に入力される。そして、算出された効用が予め定めた閾値を超える場合、交渉部４０は、他エージェントに対して対象資源の使用を提案すると決定してもよい。

【００７０】

なお、交渉部４０が交渉に用いるプロトコルは任意である。交渉部４０は、例えば、非特許文献１に記載されたＡＯＰ（Alternating Offers Protocols）を用いて他エージェントとの交渉を行ってもよい。

【００７１】

また、他エージェントになる交渉対象は、自動交渉を行うコンピュータ（情報処理装置）であってもよく、人手による判断が行われる装置であってもよい。人手による判断が行われる装置（パーソナルコンピュータ等）は、例えば、交渉部４０から送信される提案（諾否）に対する応答（受諾または拒否）や、カウンターオファーに用いられる値の入力が可能な装置（キーボード等）により実現される。

【００７２】

次に、本実施形態の交渉システムの動作を説明する。図２は、本実施形態の計画部２０の動作例を示すフローチャートである。学習部２４は、ミッション評価関数の値、またはその値の累積和の期待値を最大化するように、計画評価関数を学習する（ステップＳ１１）。そして、評価部３０は、計画評価関数を用いて定義される効用関数を用いてミッションの効用を評価する（ステップＳ１２）。

【００７３】

次に、交渉システム１００が他エージェントから対象資源の要求を受けた場合の動作例を説明する。図３は、本実施形態の交渉システム１００の動作例を示すフローチャートである。

【００７４】

まず、交渉部４０は、他エージェントから受信した対象資源の要求を計画部２０に通知する（ステップＳ２１）。計画部２０の評価部３０は、効用関数を用いて、要求された対象資源を他エージェントに譲った場合の効用を評価する（ステップＳ２２）。そして、評価部３０は、その効用を交渉部４０に送信する（ステップＳ２３）。

【００７５】

算出された効用が予め定めた閾値を超える場合（ステップＳ２４におけるＹｅｓ）、交渉部４０は、他エージェントからの対象資源の要求を許容すると決定し（ステップＳ２５）、許容する旨を他エージェントに通知する（ステップＳ２６）。一方、算出された効用が閾値以下の場合（ステップＳ２４におけるＮｏ）、交渉部４０は、他エージェントからの対象資源の要求を拒否すると決定し（ステップＳ２７）、要求を拒否する旨の交渉内容またはカウンターオファーを他エージェントに通知する（ステップＳ２８）。

【００７６】

次に、自エージェントが他エージェントに対して対象資源の使用を提案する場合の動作例を説明する。図４は、本実施形態の交渉システム１００の他の動作例を示すフローチャートである。

【００７７】

まず、交渉部４０は、他エージェントに提案する要求を計画部２０に通知する（ステップＳ３１）。計画部２０の評価部３０は、効用関数を用いて、対象資源を他エージェントに提案した場合の効用を評価する（ステップＳ３２）。そして、評価部３０は、その効用を交渉部４０に送信する（ステップＳ３３）。

【００７８】

算出された効用が予め定めた閾値を超える場合（ステップＳ３４におけるＹｅｓ）、交渉部４０は、他エージェントに対して対象資源の使用を提案すると決定し（ステップＳ３５）、提案内容を他エージェントに通知する（ステップＳ３６）。一方、算出された効用が閾値以下の場合（ステップＳ３４におけるＮｏ）、交渉部４０は、他エージェントに対して対象資源の使用を提案しないと決定する（ステップＳ３７）。

【００７９】

以上のように、本実施形態では、学習部２４が、ミッション評価関数の値、またはその値の累積和の期待値を最大化するように計画評価関数を学習し、評価部３０が、対象資源を他エージェントに譲った場合または対象資源を他エージェントから譲られた場合のミッ

ションの効用を、計画評価関数を用いて算出される内部価値の差異を定義した効用関数を用いて評価する。よって、計画するミッションについて、他エージェントとの間にインタラクションがある場合にも、自エージェントにとっての効用を適切に評価できる。

【００８０】

次に、本発明の概要を説明する。図５は、本発明による評価システムの概要を示すブロック図である。本発明による評価システム８０（例えば、計画部２０）は、ある状態（例えば、ｓ）での自エージェントの行動（例えば、ａ）の（ミッションに対する）価値を算出するミッション評価関数の値、またはその値の累積和（例えば、上記に示す式４におけるＧ）の期待値を最大化するように、その行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習する学習手段８１（例えば、学習部２４）と、交渉の対象候補になる資源である対象資源（例えば、ω）を他エージェントに譲った場合または対象資源を他エージェントから譲られた場合のミッションの効用を、計画評価関数を用いて算出される内部価値の差異を定義した効用関数（例えば、上記に示す式６）を用いて評価する評価手段８２（例えば、評価部３０）とを備えている。

【００８１】

そのような構成により、計画するミッションについて、他エージェントとの間にインタラクションがある場合にも、自エージェントにとっての効用を適切に評価できる。

【００８２】

また、学習手段８１は、ミッション評価関数を報酬関数として用いて、方策関数（例えば、π（ａ｜ｓ））および状態価値関数（例えば、Ｖ（ｓ））を学習し、評価手段８２は、状態価値関数を計画評価関数とする効用関数を用いて効用を評価してもよい。

【００８３】

他にも、学習手段８１は、ミッション評価関数を報酬関数として用いて、状態行動価値関数（例えば、Ｑ（ｓ，ａ））を学習し、学習された状態行動価値関数を用いて方策関数および状態価値関数を生成し、評価手段８２は、生成された状態価値関数を計画評価関数とする効用関数を用いて効用を評価してもよい。

【００８４】

また、評価手段８２は、他エージェントから要求される対象資源を使用しなかった場合のミッションの内部価値を計画評価関数を用いて算出し、その対象資源を他エージェントに譲った場合の効用を効用関数を用いて評価してもよい。

【００８５】

また、評価手段８２は、自エージェントが要求する対象資源を使用した場合のミッションの内部価値を計画評価関数を用いて算出し、その対象資源を他エージェントから譲られた場合の効用を効用関数を用いて評価してもよい。

【００８６】

また、計画評価関数が評価する状態に、移動体の位置情報を含んでいてもよく、効用の引数になる対象資源に、ある時刻における移動体の位置情報を含んでいてもよい。

【００８７】

また、計画評価関数が評価する状態に、交渉に直接依存しない情報（例えば、天候やミッションのコストなど）を含んでいてもよい。

【００８８】

また、ミッション評価関数は、価値の算出に用いる項として行動に対する対価を含んでいてもよい。

【００８９】

また、学習手段８１は、状態遷移モデル、シミュレータおよび予め定められた関数の少なくとも一つを用いて、計画評価関数を学習してもよい。

【００９０】

また、効用関数は、交渉によって調整される内部価値とオリジナルの内部価値との差分に対し対価を加算した関数（例えば、上記に示す式６）により定義されてもよい。

【００９１】

図６は、少なくとも１つの実施形態に係るコンピュータの構成を示す概略ブロック図である。コンピュータ１０００は、プロセッサ１００１、主記憶装置１００２、補助記憶装置１００３、インタフェース１００４を備える。

【００９２】

上述の評価システム８０は、コンピュータ１０００に実装される。そして、上述した各処理部の動作は、プログラム（評価プログラム）の形式で補助記憶装置１００３に記憶されている。プロセッサ１００１は、プログラムを補助記憶装置１００３から読み出して主記憶装置１００２に展開し、当該プログラムに従って上記処理を実行する。

【００９３】

なお、少なくとも１つの実施形態において、補助記憶装置１００３は、一時的でない有形の媒体の一例である。一時的でない有形の媒体の他の例としては、インタフェース１００４を介して接続される磁気ディスク、光磁気ディスク、ＣＤ－ＲＯＭ（Compact Disc Read-only memory ）、ＤＶＤ－ＲＯＭ（Read-only memory）、半導体メモリ等が挙げられる。また、このプログラムが通信回線によってコンピュータ１０００に配信される場合、配信を受けたコンピュータ１０００が当該プログラムを主記憶装置１００２に展開し、上記処理を実行してもよい。

【００９４】

また、当該プログラムは、前述した機能の一部を実現するためのものであっても良い。さらに、当該プログラムは、前述した機能を補助記憶装置１００３に既に記憶されている他のプログラムとの組み合わせで実現するもの、いわゆる差分ファイル（差分プログラム）であってもよい。

【００９５】

上記の実施形態の一部又は全部は、以下の付記のようにも記載されうるが、以下には限られない。

【００９６】

（付記１）ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、または当該値の累積和の期待値を最大化するように、当該行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習する学習手段と、

交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または前記対象資源を他エージェントから譲られた場合のミッションの効用を、前記計画評価関数を用いて算出される前記内部価値の差異を定義した効用関数を用いて評価する評価手段とを備えた

ことを特徴とする評価システム。

【００９７】

（付記２）学習手段は、ミッション評価関数を報酬関数として用いて、方策関数および状態価値関数を学習し、

評価手段は、前記状態価値関数を計画評価関数とする効用関数を用いて効用を評価する

付記１記載の評価システム。

【００９８】

（付記３）学習手段は、ミッション評価関数を報酬関数として用いて、状態行動価値関数を学習し、学習された状態行動価値関数を用いて方策関数および状態価値関数を生成し、

評価手段は、生成された状態価値関数を計画評価関数とする効用関数を用いて効用を評価する

付記１記載の評価システム。

【００９９】

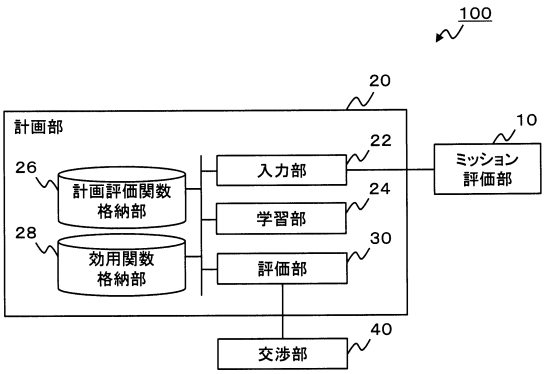
（付記４）評価手段は、他エージェントから要求される対象資源を使用しなかった場合のミッションの内部価値を計画評価関数を用いて算出し、当該対象資源を他エージェントに譲った場合の効用を効用関数を用いて評価する

付記１から付記３のうちのいずれか１つに記載の評価システム。

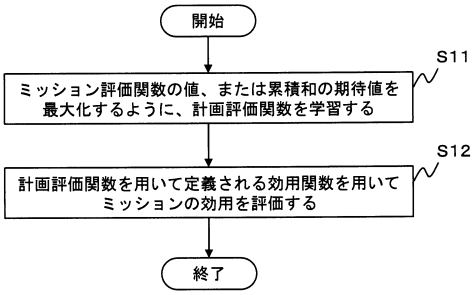
【０１００】
（付記５）評価手段は、自エージェントが要求する対象資源を使用した場合のミッションの内部価値を計画評価関数を用いて算出し、当該対象資源を他エージェントから譲られた場合の効用を効用関数を用いて評価する
付記１から付記４のうちのいずれか１つに記載の評価システム。
【０１０１】
（付記６）計画評価関数が評価する状態に、移動体の位置情報を含み、
効用の引数になる対象資源に、ある時刻における移動体の位置情報を含む
付記１から付記５のうちのいずれか１つに記載の評価システム。
【０１０２】
（付記７）計画評価関数が評価する状態に、交渉に直接依存しない情報を含む
付記１から付記６のうちのいずれか１つに記載の評価システム。
【０１０３】
（付記８）ミッション評価関数は、価値の算出に用いる項として行動に対する対価を含む
付記１から付記７のうちのいずれか１つに記載の評価システム。
【０１０４】
（付記９）学習手段は、状態遷移モデル、シミュレータおよび予め定められた関数の少なくとも一つを用いて、計画評価関数を学習する
付記１から付記８のうちのいずれか１つに記載の評価システム。
【０１０５】
（付記１０）効用関数は、交渉によって調整される内部価値とオリジナルの内部価値との差分に対し対価を加算した関数により定義される
付記１から付記９のうちのいずれか１つに記載の評価システム。
【０１０６】
（付記１１）コンピュータが、ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、または当該値の累積和の期待値を最大化するように、当該行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習し、
前記コンピュータが、交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または前記対象資源を他エージェントから譲られた場合のミッションの効用を、前記計画評価関数を用いて算出される前記内部価値の差異を定義した効用関数を用いて評価する
ことを特徴とする評価方法。
【０１０７】
（付記１２）ミッション評価関数を報酬関数として用いて、方策関数および状態価値関数を学習し、
前記状態価値関数を計画評価関数とする効用関数を用いて効用を評価する
付記１１記載の評価方法。
【０１０８】
（付記１３）ミッション評価関数を報酬関数として用いて、状態行動価値関数を学習し、学習された状態行動価値関数を用いて方策関数および状態価値関数を生成し、
生成された状態価値関数を計画評価関数とする効用関数を用いて効用を評価する
付記１１記載の評価方法。
【０１０９】
（付記１４）コンピュータに、
ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、または当該値の累積和の期待値を最大化するように、当該行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習する学習処理、および、
交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または前記対象資源を他エージェントから譲られた場合のミッションの効用を、前記計画評価関数を用

いて算出される前記内部価値の差異を定義した効用関数を用いて評価する評価処理
を実行させるための評価プログラムを記憶するプログラム記憶媒体。
【０１１０】
（付記１５）コンピュータに、
学習処理で、ミッション評価関数を報酬関数として用いて、方策関数および状態価値関数を学習させ、
評価処理で、前記状態価値関数を計画評価関数とする効用関数を用いて効用を評価させる
評価プログラムを記憶する付記１４記載のプログラム記憶媒体。
【０１１１】
（付記１６）コンピュータに、学習処理で、ミッション評価関数を報酬関数として用いて、状態行動価値関数を学習させ、学習された状態行動価値関数を用いて方策関数および状態価値関数を生成させ、
評価処理で、生成された状態価値関数を計画評価関数とする効用関数を用いて効用を評価させる
評価プログラムを記憶する付記１４記載のプログラム記憶媒体。
【０１１２】
（付記１７）コンピュータに、
ある状態での自エージェントの行動の価値を算出するミッション評価関数の値、または当該値の累積和の期待値を最大化するように、当該行動を含むミッションを計画した場合の自エージェントにおける内部価値を評価する計画評価関数を学習する学習処理、および、
交渉の対象候補になる資源である対象資源を他エージェントに譲った場合または前記対象資源を他エージェントから譲られた場合のミッションの効用を、前記計画評価関数を用いて算出される前記内部価値の差異を定義した効用関数を用いて評価する評価処理
を実行させるための評価プログラム。
【０１１３】
（付記１８）コンピュータに、
学習処理で、ミッション評価関数を報酬関数として用いて、方策関数および状態価値関数を学習させ、
評価処理で、前記状態価値関数を計画評価関数とする効用関数を用いて効用を評価させる
付記１７記載の評価プログラム。
【０１１４】
（付記１９）コンピュータに、学習処理で、ミッション評価関数を報酬関数として用いて、状態行動価値関数を学習させ、学習された状態行動価値関数を用いて方策関数および状態価値関数を生成させ、
評価処理で、生成された状態価値関数を計画評価関数とする効用関数を用いて効用を評価させる
付記１７記載の評価プログラム。
【０１１５】
以上、実施形態を参照して本願発明を説明したが、本願発明は上記実施形態に限定されるものではない。本願発明の構成や詳細には、本願発明のスコープ内で当業者が理解し得る様々な変更をすることができる。
【符号の説明】
【０１１６】
１０　ミッション評価部
２０　計画部
２２　入力部
２４　学習部
２６　計画評価関数格納部
２８　効用関数格納部
３０　評価部

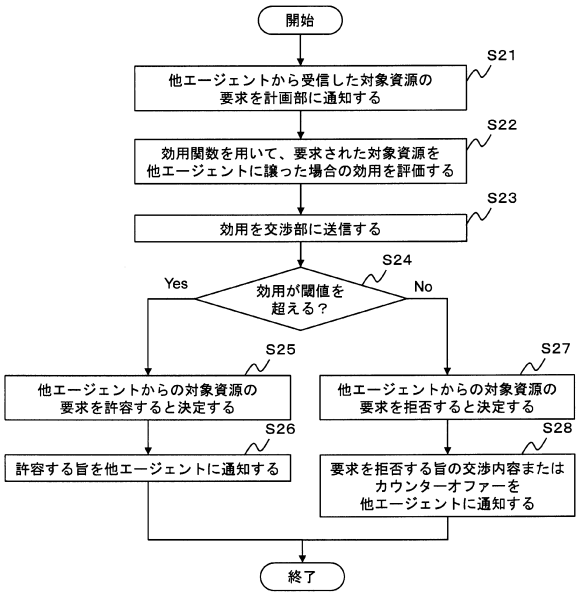
４０　交渉部
１００　交渉システム
【図面】
【図１】



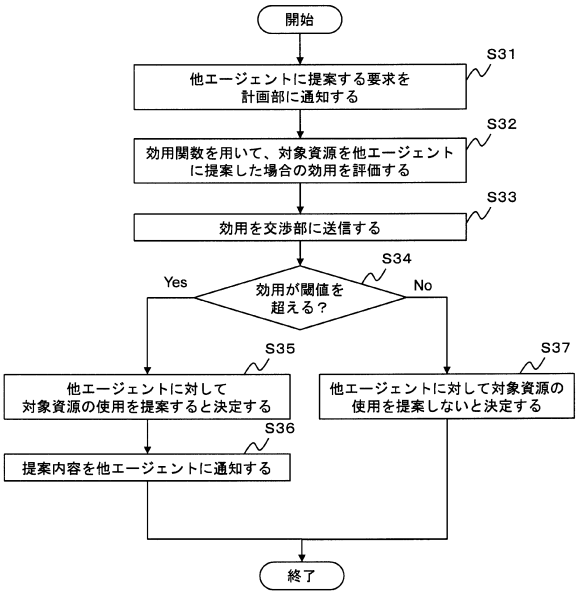
【図２】



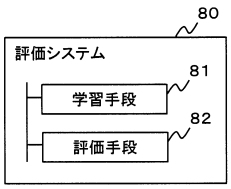
【図３】



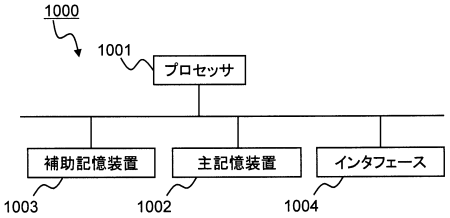
【図４】



【図５】


80
評価システム
81
学習手段
82
評価手段

【図６】


1000
1001
プロセッサ
補助記憶装置
主記憶装置
インタフェース
1003
1002
1004

フロントページの続き

(56)参考文献　　国際公開第２０１９／１６７１６１（WO，Ａ１）
(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０８Ｇ　１／００
Ｇ０８Ｇ　１／１６