修士論文

マルチエージェント強化学習による
長期電力需給モデルの構築

Development of a Long-Term Energy Model
for Electricity Demand and Supply
using Multi-Agent Reinforcement Learning

平成 18 年 2 月 3 日提出

指導教官　山地憲治教授

電気工学専攻　　46376

渡邊裕美子

内容梗概

日本の電力産業を取り巻く情勢は、京都議定書の発効、政治力学の変化、自由化により大きく変化しており、それを分析するためのモデルのひとつにマルチエージェントシミュレーションがある。本研究では、多時点に渡る動的設備投資戦略に焦点を当てた電力需給モデルを構築し、そのモデル化手法の特徴課題を明らかにし、電力設備投資における競合の分析を行うことを目的とした。

電力需給モデルの構成要素として考えたのは、電力市場、電力需要、発電事業者、政府機関、燃料価格である。発電事業者と政府機関は強化学習を行うエージェントとしてモデル化した。

本論文は第 1 章から第 6 章で構成され、第 1 章が序論、第 2 章から第 5 章が本論、第 6 章が結論となっている。

初めに第 1 章では、研究の背景と目的について述べた。

次に、第 2 章の冒頭で、自由化後の電力供給における多様な利害関係者の存在を考慮した長期電力需給モデルの概要を示し、モデル化に必要な手法として、最適化、ゲーム理論、強化学習、マルチエージェントシミュレーションといった経済的相互作用を扱う方法について説明した。

第 3 章では構築したモデルの構成要素である電力市場、発電事業者、電力需要、政府機関、燃料価格のモデル化について詳細に説明した。

第 4 章では、解析の準備として、電源構成初期値の設定や、強化学習における適切なパラメータの探査を行った。

そして第 5 章では、発電事業者エージェント数の変化や異質エージェントである政府機関エージェントの導入による結果への影響の確認をおこなった。また、多数の利害関係者による電力需給の様子を定性的に観察した結果を示した。

最後に第 6 章で、結論として本研究で得られた知見をまとめ、今後の課題について述べた。

目次

第1章 序論

日本の電力産業を取り巻く情勢は、京都議定書の発効、政治力学の変化、自由化により大きく変化しており、それを分析するためのモデルのひとつにマルチエージェントシミュレーションがある。これを踏まえ、本研究では、多時点に渡る動的設備投資戦略に焦点を当てた電力需給モデルを構築し、そのモデル化手法の特徴課題を明らかにするとともに、自由化後の制度設計について知見を得ることを目的とする。

1-1 日本の電力需給を取り巻く情勢

従来、日本は国産エネルギー資源に乏しいため、いかに安定供給を行うかがエネルギー供給の第一の課題であったが、昨今はエネルギー安全保障に関する政治力学の変化、国際的な温室効果ガス削減目標である京都議定書の発効、規制緩和・自由化による新規事業者の参入など、従来の集権型エネルギー供給に新たな利害関係が加わりつつある。分けても、自由化により大きな変革が予想され、温室効果ガス排出も多い電力産業に、これらの情勢変化は大きな影響を与えている。

1-1-1 京都議定書の発効

国際連合気候変動枠組条約の数値目標を定めた京都議定書が、2005 年 2 月 16 日に発効した。日本ではマイナス 6%など、2008 年～2012 年の第一約束期間中の温室効果ガス排出を 1990 年比で削減する目標が設定されている。

日本に課された数値目標はマイナス 6%である。日本では温室効果ガス排出中における二酸化炭素分寄与は 9 割以上を占め、そのほとんどがエネルギー起源によるものである[1]。日本政府は数値目標をさらにセクターごとに細分化した「京都議定書目標達成計画」を閣議決定した。これには数十にも及ぶ分野・区分ごとに温室効果ガス排出抑制可能量の目標値を設定している。

このうち、産業部門やエネルギー供給部門での中心となっているものは、日本経済団体連合会（経団連）など各業界団体の自主行動計画のフォローアップである。経団連は、環境税は景気や国際競争力に悪影響を与え自主行動の基盤を阻害するものであり、効果にも疑問があるとして強固に反対している一方で、「環境自主行動計画」中で温暖化対策を掲げ、これが京都議定書目標達成計画として認められた。

自主行動計画では、オフィスにおける省エネ、物流の改善、従業員や消費者の意識改革、省エネ製品やサービスの提供、植林事業のほか、京都メカニズムで認められたクリーン開発メカニズム（CDM）の利用や共同実施によるクレジットの取得、環境情報の公開などが含まれている[2]。電力業界団体である電気事業連合会は、「電気事業における環境行動計画」において、「2010 年度における使用端 CO_2 排出原単位を 1990 年度実績から 20%程度低減するよう努める」との目標を掲げた[3]。

1-1-2 エネルギーセキュリティ

2006 年 1 月 1 日、ロシア大手の天然ガス企業が、隣国ウクライナへのガス供給を停止した。原因はロシア側が 2006 年のガス輸出価格を 4 倍に引き上げることをウクライナ側へ要求したが、ウクライナ側がこれを拒否したためと伝えられている。一方で、ロシアが欧米寄りの政策を採るウクライナのユシチェンコ政権への圧力であるとも指摘されている。

この例のように、エネルギー供給の意図的な不安定性の主な要因としては政治的意図と経済的意図がある。これは、主要な資源輸出国における武力紛争や輸出禁止を伴う政治的事件の発生と主要な生産国によるカルテル形成[4]ということもできる。

エネルギー供給の安定性、つまりエネルギーセキュリティ確保が重要であるのは、不安定なエネルギー供給は経済的な損失につながるからである。日本では 1970 年代に二度の石油危機を体験した。この後、エネルギー消費国では石油から原子力や天然ガスなどへの燃料転換や、第二次産業から第三次産業への産業構造の転換が行われ、石油供給途絶が経済へ及ぼす影響は少なくなっている。

一方で、近年、石油供給途絶への対応という従来のエネルギーセキュリティの概念を新たにするような事態も生じている。中国を中心とする東アジア地域でのエネルギー需要の著しい増加、石油の利権を巡

る中東での戦争、そして新たに天然ガス供給源として見込んでいたロシアによる戦略的行動などである。これらの事態が日本のエネルギー価格を高騰させる可能性は捨てきれない。

これらの状況を受けて、電力分野においては、資源が比較的遍在している石炭、燃料価格が発電原価に占める割合の低い原子力、東南アジアなどからの安定供給が可能な天然ガス、純国産である新エネルギーの利用が重要であるとされている。しかし、石炭火力の拡大には地球温暖化問題による制約がある。原子力や新エネルギーの利用拡大には技術や設備に対する長期的投資が不可欠である。

1-1-3 電力自由化

2005 年 4 月、すべての 6000V 以上高圧需要家に対する電力供給が自由化され、全面的な自由化には 2007 年予定の家庭向けを残すのみとなった。

日本においては、地域ごとに独占的な民間の電力会社が発送配電のすべての事業を担う体制が戦後からとられており、電力会社にはこの体制が電力の安定供給を達成してきたという自負があった。しかし、1990 年代の公益事業への規制緩和の流れから、電力供給においても、海外に比較して高い電力料金を引き下げるという目的で規制緩和が検討され、1995 年には電気事業法の改正により電力供給は自由化への道を進んでいた。

貯蔵が困難である電力の瞬時需給バランスをとるために、電力会社は発送配電一貫体制を維持することになった。これは短期的な安定供給の確保が目的である。一方で、電力供給が事業者の設備投資戦略に依存することにより、長期的な電力の安定供給を阻害する可能性が指摘されている。

従来、電力価格は総括原価方式により、すべての投資コストを回収した上で一定割合の報酬を得ることができるように設定されていた。この方式は発電事業者に強い投資インセンティブを与える一方、電力価格の内外格差を産み出す原因でもあった。電力市場が自由化されれば、発電事業者は将来の電力価格、それから得られる利益、そしてリスクを考慮した上で投資を行うことが必要になる。

設備投資による一つの問題は、投資不足により発電予備力の確保が困難になる可能性である。米国カリフォルニア州では、2000 年に大規模な停電が発生し、電力供給は危機的な状況に陥った。1990 年から、1999 年の間に大規模な発電設備の建設が行われなかった一方、需要が増加し予備力が減少していたのが原因である。カリフォルニア州では 1998 年に自由化制度を導入したが、その制度設計は議論開始の 1994 年頃から不確定な状況にあった。将来の不確実性が存在すれば、それは投資意欲を減速させる可能性が指摘されている。また、制度設計上の問題としても、発電能力の確保義務の欠如、長期契約や先物取引が未発達であることが新規投資へのインセンティブを損なわせたと指摘されている。[5]

また、投資の偏りがエネルギーセキュリティの確保を困難にさせる可能性もある。ガスや石油といった火力発電は一般に初期投資額が小さく、投資が集中しやすい。一方、エネルギーセキュリティ確保に有利とされている原子力は、初期投資額が大きいこと、将来の技術開発や政策動向が不確実であることから、投資を維持するためには国策による保護が必要であるという見方、一方で価格や供給の安定性から自由化市場でも十分に生き残れるとする見方もある。

1-2 エネルギーモデル

世界の主要なエネルギー源は石油･石炭･天然ガスを中心とした化石燃料であり、それらは偏在する燃料生産地から採掘され、電力や輸送･燃焼用の多様多種の化石燃料製品に変換されて供給されている。エネルギー需給に関わる問題を考えるためには、このようなエネルギーの生産、変換、輸送、消費など、関連する技術的･経済的活動を「システム」としてとらえる必要がある。エネルギーシステムをシミュレーション分析可能なように簡略化したものを総称して、エネルギーモデルと呼ぶ。

エネルギーモデルの目的は議論や意思決定に定量的な裏づけを与えることであり、対象となる問題にはエネルギー需要や価格の予想、エネルギー技術市場の推定、環境税や技術開発といった政策の評価などがある。対象とする問題により、モデルの地理的･時間的対象や要素、また定式化の技法も様々である。

前節で述べた地球温暖化問題、エネルギーセキュリティ、電力自由化という日本のエネルギー供給を取り巻く課題を検討するためのエネルギーモデルを構築するにあたり、「技術導入」「不確実性事象」「競合」の三つのキーワードから、これまで開発されてきた様々なエネルギーモデルを紹介する。

1-2-1 技術導入の分析

地球温暖化問題の解決と経済成長を共存させるためには、新技術の開発･導入が不可欠であるといわれている。エネルギー技術の分野においては、二酸化炭素回収貯留技術、水素利用技術、クリーンコールテクノロジーなどが注目されている。エネルギーモデルの役割の一つには、このような新技術がどれだけ普及するかを分析することがある。

「DNE21」モデルによる技術ミックスの分析[6]

1990 年代に開発された代表的エネルギーモデルのひとつ DNE21 では、全世界を 10 地域に分割し、2000 年から 2010 年まで 10 年ごと 11 時点のエネルギー需給を対象に、図 1.1 に示されるようにマクロ経済モデルと気候変動モデルを統合し、二酸化炭素濃度が与えた値に安定化する総コストが最小になるようなエネルギー技術の最適ミックスを計算している。

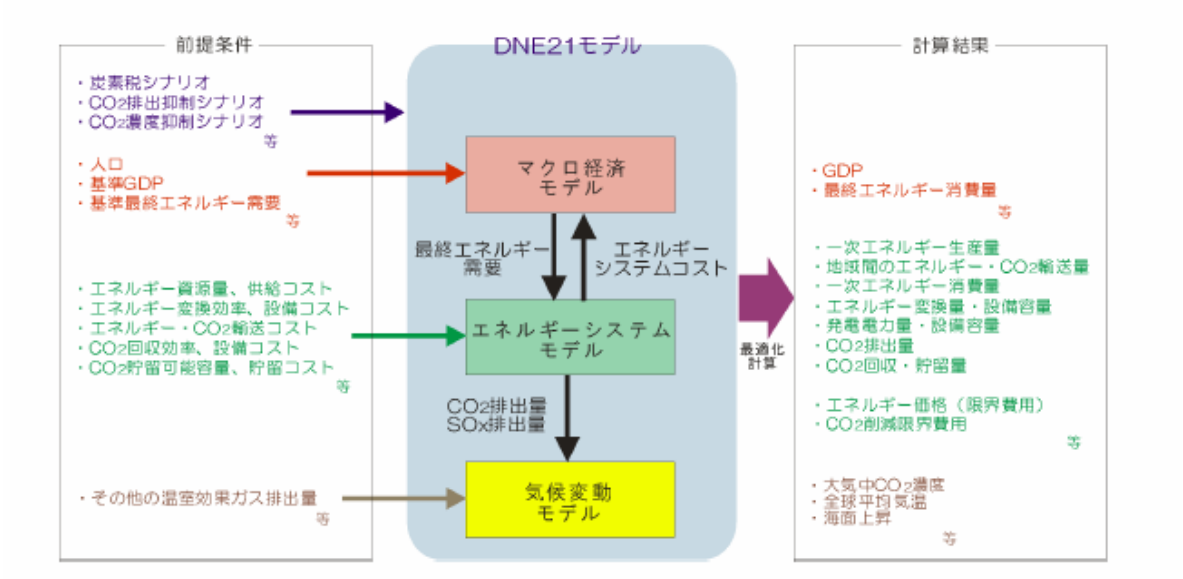


図 1.1　DNE21 モデル

「エネルギー･環境技術総合評価システム」による技術開発･導入の分析[7]

DNE21 を初めとする技術評価を中心としたモデルでは、どの技術が市場で選択されるかについてはコスト最小化によるモデリングが為されてきた。技術開発や導入をよりミクロな視点で見ていくためには、コスト最小化のモデル化では表現が困難な要因がある。半導体に代表される新技術は、その累積生産量が 2 倍になると単位当たり生産コストが 80%程度に減少するという、学習効果または習熟効果と呼ばれる経験則がある。太陽光発電パネルや風力発電機にも学習効果が働くとの報告がある[8]。日本の太陽光発電技術支援政策は、この学習効果を狙って行われたものである。また、技術開発完了の時期や程度には不確実性があり、要素技術の開発には波及効果もある。

技術開発を扱ったモデル「エネルギー･環境技術総合評価システム」[7]は図 1.2 のように、技術研究開発のプロセスを GERT 手法と呼ばれる手法により分析したサブモデルと、技術導入量を予測しその技術の重要度を評価するための静学的逐次最適化型サブモデルから成る。技術の性能･コスト･運用開始時期などを推定してその技術の導入量を予測する、またはその逆に、技術の開発のために必要な開発投入資金や計画を計算することが行える。しかしこれらの情報は一方方向のみに用いられており、技術がどの程度利用されるのか、その技術にどれだけ投資すればいいのかを同時に決定することはできない。

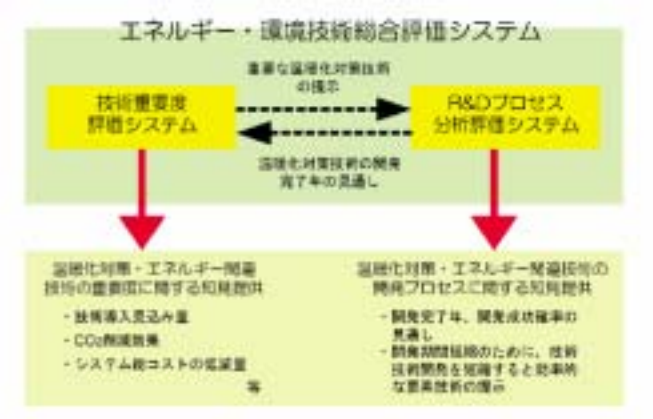


図 1.2　「エネルギー･環境技術総合評価システム」の概要[7]

技術導入の分析に関する課題

全時点を同時に最適化する動学的線型計画では、学習効果や技術開発の波及効果などの非線型な要因を表現することができない。最適化モデルのみでも混合整数計画ならばこのような要因を表現することは可能であるが、線型計画に比較し計算は困難である。また、各時点の逐次最適化を行うモデルと各時点の技術開発モデルを併用する方法では、各時点の効用に注目した結果となり、将来の効用を考えた投資行動を表現することができないという問題がある。そこで、各時点の逐次最適化を行いながら非線型要因を考慮しつつ、同時に意思決定は将来に渡る効用を考慮して行われる状況を表現するモデルが必要である。

1-2-2 不確実性事象の分析

事故や紛争によってエネルギー供給に影響が生じる可能性があるが、これらはいつ起こるのか、どの程度起こるかについては全く不確実であるということができだろう。

「石油備蓄経済効果モデル」におけるシナリオ分析[9]

エネルギー供給不安定をもたらす政治的事件を数値的にモデルで扱う試みも為されている。1993 年の段階だが、米国のコンサルタント W 社は、今後 5～7 年の中東石油供給中断の可能性を表 1.1 のように数値化している。1993 年における中東の石油生産量は日量 20 百万バレルであった。現在とは政情が異なるにしろ、石油供給に政治的不安定さがかなりの確率で存在するものとして指摘されている。日本の石油備蓄の経済効果を推計した「石油備蓄経済効果モデル」[9]では、これらのシナリオにおける経済影響をモデルで計算し、これらの確率に中断率に応じた経済効果を乗じた値を経済影響の期待値として比較評価し、どのシナリオに対しても比較的ロバストな石油備蓄計画を示している。

表 1.1　石油供給遮断の可能性 [9]（一部改変）

シナリオ	確率	中断量 [百万バレル/日]	中断期間
イランの威圧	100%	1～2	～数年
テロリストによる対サウジアラビア攻撃	25%	3～4	6ヶ月以内
イランによる対イラク攻撃	10%	3	---
イランによる対サウジアラビア攻撃	5%	8	6ヶ月～1 年
イランによるサウジ王族の内部転覆	25%	8	短期間
複合したシナリオ	5%	10	短期間

「拡張 MARIA モデル」における熱塩循環停止不確実性の政策への影響評価[10]

熱塩循環停止は地球温暖化の超長期的な影響として懸念されている問題である。熱塩循環とは、地球全体を 1000 年～2000 年かけて循環する海流で、これが停止すれば気候が重大な影響を与えるとされているが、その可能性や影響の程度には不確実性が大きい。

この超長期の不確実性が中長期の政策に与える影響を評価するため、「拡張 MARIA モデル」[10]では、動学的非線形最適化モデルである MARIA に、気候感度が 5 通りのシナリオを用意し、不確実性が解消するまではいずれの可能性に対しても同じ行動を取るとした意思決定アプローチ(Act Then Learn)を適用した。その結果、不確実性が解消するまでは最悪の事態に合わせた最小水準の政策が採られることを占めた。

不確実性事象の分析に関する課題

不確実な状況を確率やパラメータで定められるシナリオとして表現して各々の状況をシミュレーションし、またそれらの期待値をとることで、確率的な事象をモデルで扱うことは可能である

このシナリオに対する最適行動は、ともすれば最も厳しい条件に従った解のみが導かれやすい。確率計画法では、このような条件式を扱う方法に「リコースを持つ確率計画問題」と、「機会制約条件計画」という二つがある。リコースを持つ確率計画問題では、確率変動する制約条件に対して与える差異に対してペナルティ（リコース）を設け、それを目的関数に組み込む。一方、機会制約条件問題においては、条件式がある一定の確率で成り立てばよいとする。[11]

しかし、このように不確実性を扱ったとしても、シナリオは離散的に与えられるものであり、考えうる全ての状況の組み合わせを考えるとシナリオの数は膨大なものになる。これは空間的･時間的計算量の制約を受ける。また、確率計画を解くための計算アルゴリズムは完全に数学的な意味合いしか持たず、現実の意志決定者の意思決定機構として解釈するのが難しいという欠点もある。[12]

1-2-3 競合の分析

ゲーム理論によるカルテルの分析

電力自由化の自由化により発電事業者間の競争が激しくなれば電力価格は下がると考えられているが、市場に参入する発電事業者が暗黙のうちにカルテルを組んで電力価格高騰につながる可能性も否定できない。このように価格形成に大きな影響を与えるプレーヤーが存在する市場をモデル化する方法のひとつに、ゲーム理論を用いるものがある。

1976 年 12 月、OPEC は原油価格の値上げを巡って分裂し、カルテルの崩壊の可能性を示した事件は、有名なゲーム「囚人のジレンマ」と同じ状況であると指摘されている[13]。この指摘に従えば、カルテル崩壊は例えば次の利得行列で表すことができるだろう。2 体の石油生産国プレイヤーがあり、共に高価格をつけた場合はそれぞれが大きな利益を得ることができる。これがカルテルが形成されている状況である。しかし、一方が裏切りを起こして低価格戦略をとれば、そのプレイヤーの石油への需要が大きくなり、高価格戦略のプレイヤーは利益を奪われてしまう。これを避けようと両方が低価格戦略をとれば、共に利益は小さくなる。これは OPEC カルテルを非常に単純なゲームとして設計しているが、さらに石油需要に対する情報、供給関数の導入、プレイヤーの数の増加など複雑なモデル設定は可能である。

表 1.2 石油カルテルの囚人のジレンマゲーム

（石油生産国 1、石油生産国 2）の利得

石油生産国 1 / 石油生産国 2	高価格（協力）	低価格（裏切り）
高価格（協力）	6, 6	8, 2
低価格（裏切り）	2, 8	3, 3

マルチエージェントシミュレーションによる電力市場の分析[14]

自由化の拡大が進む電力市場のシミュレーションを行うために、マルチエージェントベースの手法が研究されている。RPS 制度のような電力関係制度設計の影響評価、複数の取引形態が存在する場合の市場の振る舞いの検討、地理的要因に焦点を当てた市場分析、分散電源の影響シミュレーションなどが行われている。これらは、同時同量の維持が必要、流通経路に制約が多いという電力市場の特殊性に焦点をあてたものであり、必然的に扱われる電力市場シミュレーションは短期のものとなる。

一方で、電力市場の特殊性には、新たな電源開発に極めて時間がかかるという点もある。このような設備投資戦略については、短期間シミュレーションから投資効果や利益を予測するという扱われ方はなされているものの、設備投資そのものに焦点をあてたシミュレーションの例は特に指摘されていない。

マルチエージェントシミュレーションによるエネルギー戦略の国際競合関係の分析[15]

世界各国がエネルギー需給戦略を決定する際にその利益追求が大きな要因を占めることを直接的にモデル化を試るために用いられた手法がマルチエージェントシミュレーションである。

当研究室における先行研究「エネルギー戦略の国際競合関係考慮のためのエージェントベース世界エネルギーモデル」[15]では、DNE21 の流れを汲むエネルギー供給コスト最小化モデル「世界地域細分化エネルギーモデル」に、マルチエージェント要素を組み込んだ。従来はエネルギー市場に完全競争を仮定し、生産や輸送に必要なコストがそのままエネルギー価格に一致するものとしていたが、2010 年の 1 期における価格戦略に焦点を当て、OPEC や EU といった地域におけるエネルギー輸出入戦略を導入したのである。すなわち、エネルギー輸出入時のプレミアム・関税を戦略変数として導入し、従来モデルのコスト値の代わりにこれらの戦略変数を用いて最適化計算を行うことでエネルギーシステムを計算する。それぞれの地域をエージェントとしてモデル化し、彼らが地域のエネルギーシステムコストが最小になるような戦略変数の値を、強化学習を用いて探索するという方法を用いている。計算の流れを図 1.3 に示す。

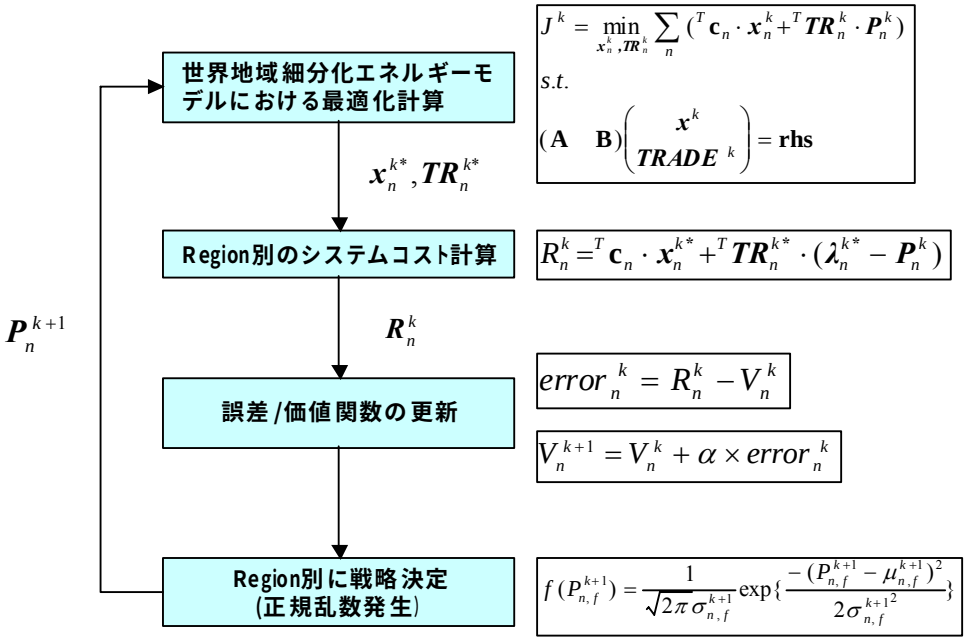


図 1.3　エネルギー戦略の国際競合関係考慮のためのエージェントベース世界エネルギーモデル[15]

競合の分析に関する課題と解決案

ゲーム理論で扱われているモデルの環境設定を複雑化し、各プレイヤーに与える情報量を限定したものがマルチエージェントシミュレーションということができる。モデル設定の現実性、解の容易さ、表現の明確さはトレードオフの関係にもあるが[16]、なんらかの既存の適切なモデルをベースにマルチエージェント要素を組み込むことはそれなりに妥当性はあろう。

長期のエネルギーシステムを対象にすれば、価格戦略だけでなくエネルギー関連設備への投資戦略も検討することが必要になる。設備投資戦略を検討するに当たっては、設備は意思決定後長期に渡って残存することから、将来期間に渡る利益を考慮した意思決定、つまり動学的な意思決定機構が必要である。

1-3 研究の目的

日本の電力産業においては、エネルギー安全保障に関する政治力学の変化、国際的な温室効果ガス削減目標である京都議定書の発効、規制緩和･自由化による新規事業者の参入など、従来の集権型エネルギー供給に新たな利害関係が加わりつつある。

一方で、このような問題を分析するためのエネルギーモデルを「技術導入」「不確実性事象」「競合」の観点から見たとき、従来の全体最適化モデルでは扱えない問題要因があることを指摘した。その課題を解決するため、本研究ではマルチエージェント強化学習シミュレーションによる、利害関係者の動学的意思決定のモデル化を試みる。

・　技術導入

各時点の逐次最適化を行いながら非線型要因を考慮しつつ、同時に意思決定は将来に渡る効用を考慮して行われる状況を表現するモデルが必要である。強化学習では、エージェントと呼ばれる意思決定主体が、遷移する状態の中で最終的な総報酬を最大化させるような試行を繰り返される。

・　不確実性事象

シナリオは離散的に与えられるものであり、考えうる全ての状況の組み合わせを考えるとシナリオの数は膨大なものになり、これは空間的･時間的計算量の制約を受ける。マルチエージェント強化学習を用いることで、エージェントの学習機構に意思決定者の意思決定モデルを用いることが可能であるし、エージェントは存在環境に関して多くの不確実性が存在しても動作すべきであることが予め予想されている[17]。

・　競合

既存のマルチエージェント競合分析モデルは静的な価格戦略のみを考慮しており、多時点に渡る動的設備投資戦略への発展という課題があった。設備投資戦略を検討するに当たっては、設備は意思決定後長期に渡って残存することから、将来期間に渡る利益を考慮した意思決定、つまり動学的な意思決定機構が必要である。

以上を踏まえ、本研究の目的を以下のように設定する。

- エネルギーモデルの対象として一地域の電力需給を選び、多時点に渡る動的設備投資戦略に焦点を当てたエネルギーモデルを構築し、そのモデル化手法の特徴や課題を明らかにする。
- 同時に電力需給における多数の利害関係者を直接モデル化し、電力設備投資における競合の分析を行う。

第2章 研究に用いる手法

本章では、自由化後の電力供給における多様な利害関係者の存在を考慮した長期電力需給モデルの概要を示した。そしてモデル化に必要な手法として、最適化、ゲーム理論、強化学習、マルチエージェントシミュレーションといった経済的相互作用を扱う方法について説明した。

なお、モデルの定式化については第 3 章で説明する。

2-1 構築を行うモデルの概要

2000 年から 2030 年までの電力需給を対象とし、地域区分や送電については考えない一地域の電源構成計画をモデル化した。2000 年から 2030 年までを 5 年間隔に分割して 6 時点を設け、各時点を 1 季節･24 時間帯に分割した。

電力市場、電力需要、発電事業者、政府機関、燃料価格を電力供給における構成要素とし、これらの間に図 2.1 に示す関係を考えた。電力市場に対して、需要家は外生的な需要曲線を入札し、発電事業者は変動する燃料価格に基づいた供給曲線を入札し、電力市場はこれの決済を行って電力取引量を決定する。発電事業者はそれにより得られる利益をもとに、将来の時点に向けての設備増設を行う。これに、温暖化対策や電力安定供給を目的とする政府機関が、電力の供給状態に応じて発電事業者へ課税や補助といった経済的手段で介入する。

このうちの決定変数は発電事業者による設備増設量と政府機関による税率や補助率の値である。

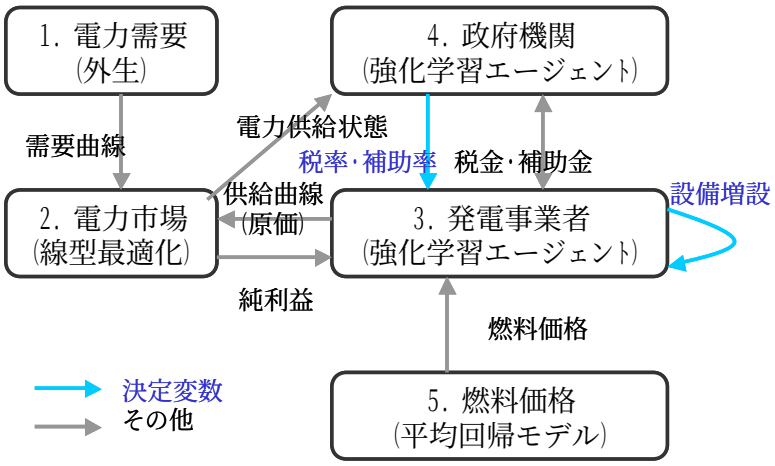


図 2.1 「マルチエージェント型一地域電力供給モデル」の構成要素

2-2 経済学における相互作用の扱い

ミクロ経済学では、生産や消費、分配などの経済活動に関わる者が個々の利益を最大化するという経済合理的な行動をとることを仮定し、それらの相互作用が財の価格や需給にどのような影響を及ぼすかを記述する。相互作用の形態を決める要因には、相互作用に関連する意思決定主体の種類、相互作用の強さ、取引する財の種類などがあるが、中でも、図 2.2 に示すように、主体の数と他者の選択が自己へ与える影響の大きさによって、最適なモデル化手法が異なってくる。[18]

このうち、最適化、強化学習、ゲーム理論、マルチエージェントシミュレーションについて以下の節で述べる。

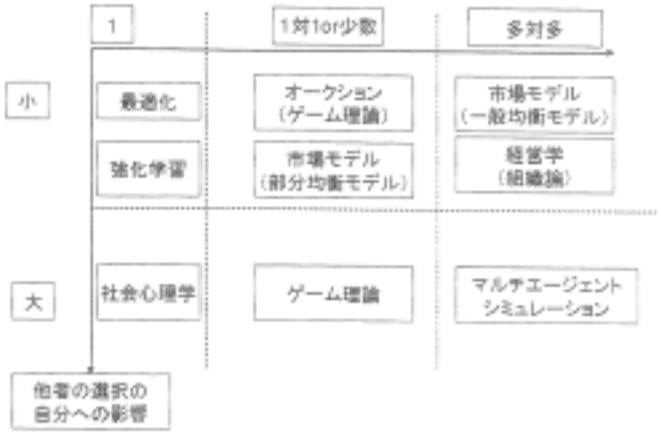


図 2.2　経済学での相互作用の扱われ方と関連領域[18]

2-3 最適化

2-3-1 完全競争市場のモデル

生産者がある財を供給するとき、供給する財の量に関わらず一定に発生する固定費と、供給する財の量に従って発生する可変費の費用が存在し、財の量とこれらの費用を対応づける関数を費用関数という。この関数を財の量で微分すれば、その生産者が財を一単位増産するのに必要とする費用を表す関数が得られ、これを限界費用関数という。限界費用は一般に、生産量の増大に従い逓増する。

消費者には、その財を消費することによって効用が発生する。消費者が財を一単位多く消費するときに増加する効用を限界効用と呼ぶ。限界効用は、財を一単位多く手に入れるために消費者が支払ってもよいと考える額（Willing to Pay）に相当する。限界効用は一般に消費量の増大に伴い逓減する。

ここで完全競争市場、つまり生産者や消費者の数が十分大きく、個々の生産者や消費者が市場価格を操作できない状況を仮定する。このとき、財の市場価格は個々の生産者の行動に依らず与えられる。
個々の生産者がその利益を最大にするためには、消費者の消費量が十分多いとすれば、限界費用が市場価格に一致する量まで生産を行えばよい。限界費用関数は市場価格に対応する合理的な供給量を表しているため、限界費用関数の逆関数は供給関数と呼ばれる。個々の消費者にとっては、その効用を最大にするためには、生産者による生産量が十分多いと仮定すれば、効用が限界費用に一致する量まで消費を行えばよい。限界効用関数は市場価格に対応する合理的な需要量を表しているため、限界効用関数の逆関数は需要関数と呼ばれる。

個々の生産者の供給関数、個々の消費者の需要関数が与えられたとき、それを足し合わせることによって、その市場における供給関数、需要関数を求めることができる。限界費用逓増則、限界効用逓減則より、供給曲線は価格に対する増加関数、需要関数は価格に対する減少関数となる。この市場における市場価格は、市場価格に対して供給量と需要量が均衡する点、つまり供給曲線と需要曲線の交点として求められ、この価格を均衡価格という。

生産者が財の生産によって得た利益から、財の生産に要した費用を引いたものを生産者余剰、消費者が得た財の消費によって得た効用から、財の消費に要した費用を引いたものを消費者余剰という。また、生産者余剰と消費者余剰の和を社会的厚生という。市場価格が均衡価格であるとき、社会的厚生は最大となる。

2-3-2 完全競争市場の最適化による定式化

完全競争においては、個々の生産者や消費者の限界費用曲線や限界効用曲線が数式として与えられれば、その市場における個々の生産者や消費者の合理的な行動は、社会的厚生を最大化するという目的関数を持つ最適化問題を解けば求められるよいことになる。

限界費用曲線を $c_i(s_i)$、限界効用曲線を $u_j(d_j)$とする。需給が均衡するという制約の下で、消費者効用と生産者費用の差を最大化する問題(2.1)に帰着する。

$$\sum_i s_i = \sum_j d_j = Q$$
$$PS = \sum_i \int \left(p - c_i(s_i)\right) ds_i$$
$$CS = \sum_j \int \left(u_i(d_i) - p\right) dd_i$$
$$PS + CS = -\sum_i \int c_i(s_i) ds_i + \sum_j \int u_i(d_i) dd_i \to \max \tag{2.1}$$

市場価格 p は、最後の消費 Q を行った消費者の限界効用であり、最後の生産 Q を行った生産者の限界費用である。すなわち、この最適化問題における市場価格 p は式(2.2)のように表せる。

$$p^* = \frac{\partial PS^*}{\partial Q^*} = \frac{\partial CS^*}{\partial Q^*} \tag{2.2}$$

ここで、財の需要に弾力性が無い、つまり価格によらず定数 D で表せると仮定する。このときの消費者効用は無限大で変化が無く、考慮する必要がなくなる。このため上記の最適化を図るためには生産者費用の項を最小化すればいいことになる。すると、問題(2.1)は問題(2.3)のように変換される。

$$\sum_i s_i = D$$
$$V = \sum_i \int c_i(s_i) ds_i \to \min \tag{2.3}$$

このときの市場価格は、式(2.4)で表せる。

$$p^* = \frac{\partial V^*}{\partial D} \tag{2.4}$$

この p のように、制約式右辺定数項の微小変化に対する目的関数の微小変化量は、一般に制約式のシャドウプライスと呼ばれる。

以上をまとめると、(1)完全競争、(2)需要の非弾力性、を仮定するとき、個々の生産者の経済合理的な行動は、全体の供給コストを最小化するという最適化問題を解くことで計算することができる。

2-4 ゲーム理論

2-4-1 ゲーム理論と戦略

ゲーム理論では、ある主体の行動選択戦略が他の体の行動選択戦略に強い影響を与えるような状況下をモデルとして表す。考慮する主体の数は少数であり、どの主体にとっても利得を最大とするような戦略の組、つまり均衡戦略が存在することがある。

各主体がどの戦略を選ぶかは、その戦略が他の戦略に対してどれだけ多くの利得を得られる可能性があるかによって定まる。

- 支配戦略　　他主体の行動選択戦略が何であっても、他のどの戦略より大きな利得を得られるような戦略。
- 弱支配戦略　　他主体の行動選択戦略が何であっても、他のどの戦略以上の利得を得られるような戦略。

・　マックスミニ戦略　　他主体が自分の利得をできる限り低くしようとする戦略を持っていると仮定し、その場合に得られる利得を最大にしようとする戦略。
・　ミニマックス戦略　　他主体が獲得しうる利得の最大値をできる限り低くしようとする戦略。

これらの戦略の組として、次のような均衡戦略が存在し得る。
・　支配戦略均衡　　どの主体にとっても支配戦略が存在するとき、その戦略の組。
・　反復支配戦略均衡　弱支配された戦略をとることは無いと仮定して得られる戦略の組。
・　ナッシュ均衡　　どの主体にとっても、その戦略の組から逸脱すれば利得が下がる、つまり逸脱するインセンティブを持たないような戦略の組。
・　マックスミニ均衡　　すべての主体のマックスミニ戦略の組。
・　ミニマックス均衡　　すべての主体のミニマックス戦略の組。

また、ある戦略の組による結果の利得が、すべてのどの戦略の組よりもすべての主体により多くの利得をもたらすとき、その戦略の組をパレート最適であるという。

マックスミニ戦略のような考え方は、ミニマックス原理と呼ぶ。状況危険回避的な主体にとっての最適戦略でもなく、相手の行動から得られる情報（信念）から整合的に判断した戦略でもないこと[16]が指摘されているが、一方、状況が不確定で、自分に対して敵対的のときは、ミニマックス原理による発想法が有用な考え方である[13]とも言われている。

2-4-2 ゲームと情報

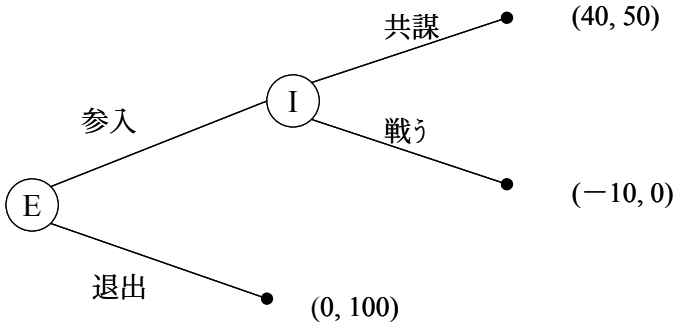
主体の行動が同時ではないゲームや繰り返し行われるゲームでは、これまで他主体がどう行動してきたかという情報から形成される、他主体の行動戦略についての「評判」が戦略決定に重要な役割を果たす。

評判の例として、参入阻止ゲーム[16]を考える。既存の市場に新規企業は参入か退出を選択し、新規企業が参入した場合は、既存企業は共謀するか戦うかを選択する。このゲームでは新規企業の選択が最初の手番となるため、ゲームは標準型（利得行列ともいう）で表 2.1、展開型で図 2.3 のように表現される。

表 2.1　参入阻止ゲーム（標準型）

（参入企業, 既存企業）の利得

参入企業＼既存企業	共謀	戦う
参入	40, 50	−10, 0
退出	0, 100	0, 100



（参入企業, 既存企業）の利得

図 2.3　参入阻止ゲーム（展開型）

表 2.1 に表された標準型では、ナッシュ均衡は（参入、共謀）と（退出、戦う）になるが、実際は図 2.3 に表されるように既存企業が退出した参入企業と戦う必要は無い。一方、一度参入を決意した既存企業に対しては、既存企業は共謀することが最適反応戦略である。このため、仮に既存企業が参入企業に対して参入に対しては迎え撃つという脅しを通告できたとしても、参入企業にそれを信用させるためには、既存企業は利益を 50 の犠牲にしてでも戦うという戦略をとるのだという評判を形成する必要がある。

2-5 強化学習[17]

2-5-1 強化学習の特徴

情報処理の分野では、環境の状態を知覚し、知的なメカニズムにより処理を行い、その結果に従って行動するものをエージェントと呼んでいる。強化学習は、エージェントの知的メカニズムのひとつとしてよく用いられる手法で、エージェントが過去に経験した行動とそれに対して得た報酬をもとに、より多くの報酬を得るための行動戦略を自ら獲得する学習方法である。これに対して、機械学習では、報酬をもたらす最適な行動パターンの知識が予め存在し、エージェントはそのような行動パターンを導出するための戦略を学習する。

エージェントは、環境がどのような状態にあるかを知覚し、その状態を報酬という数値で評価する。報酬は、その全期間に渡る総和を最大化することがエージェントの目的になるように設計されるものである。エージェントは試行を通じ、その状態に達した以後に得られるであろう総報酬の値を推定する。この総報酬の値の推定値は、状態の価値を表していることから価値関数と呼ばれる。そして、エージェントは現在の状態からより多くの報酬を得られる状態に移るよう、ある確率に従って行動を選択する。この選択確率は方策と呼ばれ、これを決定することが強化学習を用いる目的である。

強化学習は次のような特徴を持つ。

- 環境との相互作用を利用するため、環境の正確なモデルを必要としない。
- 報酬を最大化するという目標指向型の学習方法である。
- 何が最適な行動であるかについての知識を持たないため、試行錯誤的な探査が必要である。
- 非定常で不確実性の多い一般の環境においても動作する。

2-5-2 価値関数

状態 s 以降に方策πに従うときの総報酬の推定値を状態価値関数 $V^{\pi}(s)$と呼ぶ。また、ある状態 s で行動 a を選択し、その後は方策πに従ったときの総報酬の推定値 $Q^{\pi}(s, a)$を行動価値関数と呼ぶ。これらを最適に推定できれば、最適方策は用意に得られる。最適に推定された価値関数は、式(2.5)の Bellman 最適方程式を満たす。

$$V(s) = \max_a E\{r_{t+1} + \gamma V(s_{t+1}) \mid s_t = s, a_t = a\}$$
$$Q(s,a) = \max_a E\left\{r_{t+1} + \gamma \max_{a'} QV(s_{t+1}, a') \mid s_t = s, a_t = a\right\} \tag{2.5}$$

現在の方策πから価値関数 $V^{\pi}(s)$や $Q^{\pi}(s, a)$を推定することを方策評価という。また、価値関数 $V^{\pi}(s)$と $Q^{\pi}(s, a)$に基づき、$V^{\pi'}(s) > V^{\pi}(s)$を満たす新たな方策π'を探すことを方策改善という。方策反復と方策改善を連続的に組み合わせることで、最適な方策を見つける手法を方策反復という。

実際の強化学習の手法においては、方策反復は、方策評価と方策改善の過程を、一方を完全に終了させてから他方の過程を開始するやり方でなく、二つを細かく交互に実行するやり方で行われる。

2-5-3 方策改善

強化学習では、知識として得られた価値関数を利用する行動と、現在は低く評価されている行動の実際の価値を確認するための探査的行動とのバランスが重要である。例えばεグリーディ方策と呼ばれる

手法では、小さな確率 ε で、価値関数から導かれる最適（グリーディ）な行動ではなく探査的な行動を選択する。

ε グリーディ手法ではグリーディ行動以外の選択確率を等しくしているが、行動価値関数 $Q(s, a)$の代わりとして用いられる行動動優先度 $p(s, a)$と呼ばれる値を用いて行動選択の重みをつける方法があり、例えばソフトマックス行動選択規則と呼ばれる手法がある。

追跡手法では、行動価値推定 $Q(s, a)$と、グリーディな行動を「追いかける」目的で使われる行動優先度 $p(s, a)$の両方を使用する。行動優先度は、最も単純には行動確率 π (s, a)そのものが用いられる。この場合、追跡手法においては、式(2.6)のようにグリーディ行動の選択確率を 1 に向かって β の比率で増加させ、残りの行動の選択確率を 0 に向かって β の比率で減少させる。

$$\pi(s,a) \leftarrow \begin{cases} \pi(s,a) + \beta\big(1 - \pi(s,a)\big) & a = \arg\max Q(s,a') \\ \pi(s,a) + \beta\big(0 - \pi(s,a)\big) & otherwise \end{cases} \tag{2.6}$$

2-5-4 方策評価

モンテカルロ法では、最終状態で総報酬が得られるまで待ち、その値を価値関数の推定に用いる。状態 s で行動 a を選択した後、方策 π に従った場合の最終状態までの総報酬を $R(s, a)$と表すと、状態 s で行動 a を選択することに対する行動価値関数は式(2.7)で推定される。

$$Q(s,a) = average(R(s,a)) \tag{2.7}$$

TD (Temporal Difference: 時間的差分) 学習では、最終状態に達するまで待つのではなく、次状態での価値推定値を利用して現状態の価値推定値を更新する。TD 学習の種類として、Q 学習、Sarsa、アクタークリティック手法などがある。

TD 学習法が有力なのは、次状態が現状態と現在の行動のみに依存するときである。このような性質をマルコフ性という。環境がマルコフ性を持つとき、現状態と行動が与えられれば次の状態と今後期待される報酬を予測できるため、現状態をもとに行動を選択することは妥当である。実際には環境がマルコフ性を持たない場合でも、マルコフ性を仮定することは有用である。

Q 学習では、状態 s で行動 a を選択した即時報酬が $r(s, a)$ であり、次状態 s'におけるグリーディ行動を a'としたとき、状態 s で行動 a を選択することに対する行動価値関数は式群(2.8)で推定される。ここで α は学習率、γ は割引率を表す。Q 学習のように行動価値関数の推定に現在用いている方策を用いない方法を、方策オフ型であるという。ここで δは TD 誤差と呼ばれる値である。

$$\begin{aligned} \delta &= r(s,a) + \gamma Q(s',a') - Q(s,a) \\ a' &= \arg\max Q(s,a) \\ Q(s,a) &\leftarrow Q(s,a) + \alpha \times \delta \end{aligned} \tag{2.8}$$

また、Sarsa では、a'の決定にグリーディ行動ではなく、方策 π から導かれる行動を用いる。Sarsa のように行動価値関数の推定に現在用いている方策を利用する方法を、方策オン型であるという。

モンテカルロ法と TD 学習法を一般化し、n ステップ後までの実際の報酬と、それ以後の価値推定値を利用して価値関数を更新する方法を、n ステップ TD 予測という。ここでは状態価値の推定を考える。

状態 s_tでの即時報酬が r_tであり、その後 n ステップ後まで方策 π に従い、各状態で報酬 $r_{t+1}, \ldots r_{t+n}$が得られ、$t+n+1$ に状態 s_{t+n+1} に達したとする。状態価値関数は n ステップ収益 $R^{(n)}$を用いて式(2.9)のように更新される。

$$
\begin{aligned}
R_t^{(n)} &= \sum_{t'=1}^{n} \gamma^{t'} r_{t+t'} + \gamma^{n+1} V_t(s_{t+n+1}) \\
\Delta V_t(s_t) &= \alpha\left[R_t^{(n)} - V_t(s_t)\right] \\
V_t(s_t) &\leftarrow V_t(s_t) + \Delta V_t(s_t)
\end{aligned}
\tag{2.9}
$$

これまで挙げた方法により価値関数を推定する際、学習率に定数を用いる方法では、その初期値の影響が減衰してはゆくが、ある程度永続的に残る。このため、初期の設定は、エージェントにどの程度の報酬が期待できるかということについて事前知識を与えることになる。特にすべての行動･状態に対して楽観的な数値を与えることにより、価値推定量が収束する前にすべての行動が数回試みられることになる。この方法を、探査を促進するためのオプティミスティック初期値という。

2-5-5 アクタークリティック法

アクタークリティック手法は、方策改善･評価ともに独特の手法である。ここでは、方策は価値関数とは完全に独立して決定される。行動を選択する方策部分はアクターと呼ばれ、その行動を評価する部分はクリティックと呼ばれる。

方策決定には行動価値関数を用いず、という定めてこの値を更新し、ソフトマックス手法などによって行動選択確率に変換する、もしくは行動選択確率を分布関数で与え、そのパラメータを直接更新するという方法をとる。

実際の行動 a による報酬 $r(s, a)$から見込まれる総報酬と、価値関数の値との差を、TD 誤差 δ といい、式(2.10)で表される。

$$
\delta = \left(r(s,a) + \gamma V(s')\right) - V(s)
\tag{2.10}
$$

クリティックは TD 誤差 δ を用いて、式(2.11)に従って行動優先度 $p(s, a)$の値を更新する。β は定数である。

$$
p(s,a) \leftarrow p(s,a) + \beta\delta
\tag{2.11}
$$

また、行動優先度の代わりに方策を分布関数で表す方法では、行動選択のための演算が容易であり、なおかつ連続値行動の表現が可能である。方策 $\pi(s, a)$が行動の平均値$\mu(s)$と標準偏差$\sigma(s)$で定まる正規分布で表されていたとすると、このときのアクターによる方策改善は、図 2.4 に示されるように δ、$a-\mu(s)$、$|a-\mu(s)|-\sigma(s)$の正負によって定まる。すなわち、TD 誤差が正であれば、平均μは、実行した行動 a に近づけるように更新し、標準偏差σは、行動が分布の外側なら大きく、分布の内側なら小さく更新すればよい。

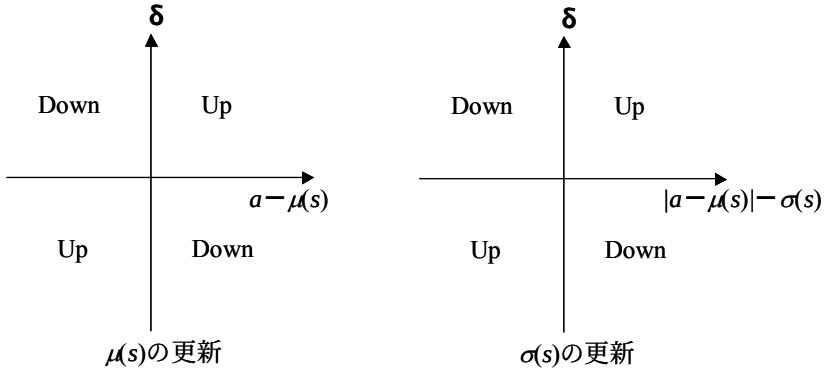


図 2.4 アクタークリティック法における方策改善

アクタークリティック法には次のような利点がある。

- 行動選択に最小限の計算量しか必要としない。
  一方、Q 学習などでは、行動選択のために a を走査して $Q(s, a)$の値を調べる必要がある。
- 確率的な方策を陽に学習することができ、非マルコフ 過程に対しても有用である。

2-5-6 強化学習の性能

前項では、知識利用と探査のバランスをとるための、ε グリーディ方策、ソフトマックス方策、追跡手法などについて述べた。問題によって、これらの方策のうちどれが最も良い性能を発揮するかは異なる。

例えば、「10 本腕バンディット問題」の例[17]を採り上げる。これは、エージェントはスロットマシンの賞金の出やすさの異なる 10 本のレバーから一本を選択する問題で、レバーa に対する賞金額は平均は $Q(a)$、分散を 1 とする正規分布で与えられる。$Q(a)$ も平均 0、分散 1 の正規分布に従って決定されている。エージェントは 1000 エピソードの試行を通じて最良のレバーを選択する。

明らかに最適なレバーは $\arg\max_a Q(a)$ であり、この値は 1.55 程度であるが、このような単純に見える問題でも、方策やパラメータが異なればその学習解は全く異なる。例えば、この問題を $Q(a)$ の異なる 2000 のスロットマシンで行い、各エピソードで何%のスロットマシンについて最適なレバーを発見できたか（行動最適度）を示している。

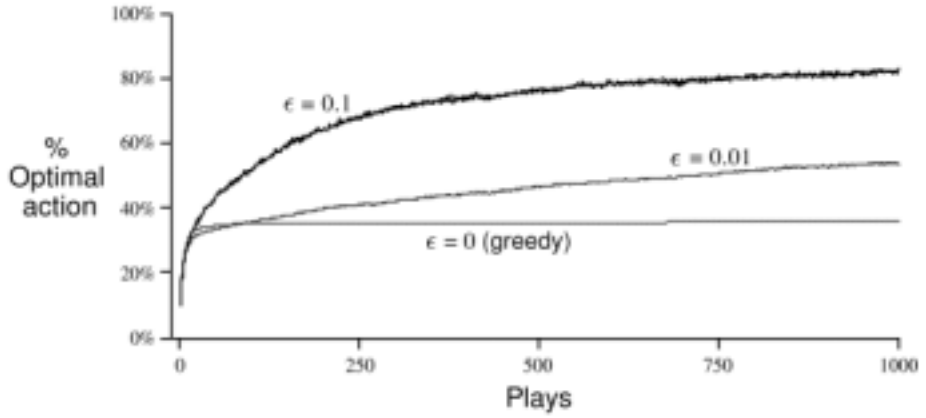


図 2.510 本腕テストによる行動最適度[17]

このように、強化学習によってエージェントは必ずしも最適な解に到達できるとは限らない。しかし、これは強化学習手法が途上段階にあるということではない。強化学習が有効なのは、非定常で不確実性の多い一般の環境においても動作するということである。その点ではこれまで述べたような簡潔な手法で十分であると指摘されている。

2-6 マルチエージェントシミュレーション

2-6-1 マルチエージェントシステム

マルチエージェントは、各エージェントそれぞれが分散問題を解決することによっ、問題全体を解決しようとするシステムであり、分散問題を効率的に処理するものとして注目されている。[19]

また、エージェントが複数存在して相互作用を通じつつ強化学習を行う図 2.6 のようなシステムを、マルチエージェント強化学習システムと呼ぶ。

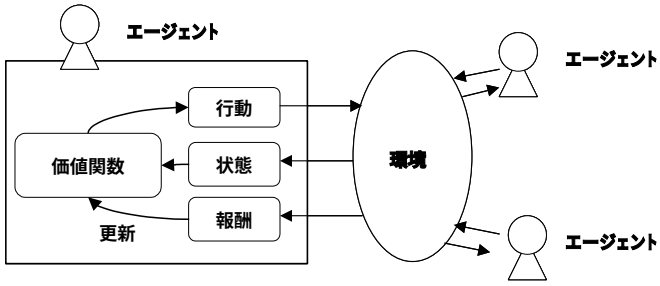


図 2.6 マルチエージェント強化学習

2-6-2 情報とマルチエージェントシステム設計における課題

マルチエージェント強化学習の枠組みは、主体の持ちうる情報が限定されているゲームであるということができる。

一般にゲームにおける情報構造には、次の四つの観点がある。[16]

- 完全情報　各情報集合が単一である。
- 確実情報　どのプレイヤーの手番の後にも自然の手番がない。
- 対称情報　自分の手番となるとき、あるいは終節において、他のプレイヤーと異なる情報をもつプレイヤーが存在しない。
- 完備情報　最初の手番が自然によるものの場合、それがすべてのプレイヤーによって観察される。あるいは、最初の手番が自然によるものではない。

「プレイヤー」と呼ばれているものは、マルチエージェント強化学習における「エージェント」と同義であり、「手番」とは戦略に従った行動を決定することにあたる。また、自然の手番とは、確率的に環境が変化することを示している。上記の観点を、マルチエージェント強化学習の言葉で書き直すと次のようになる。

- 完全情報　各エージェントが現在の環境を常に正確に知っている。また、複数のエージェントが同時に環境に作用することがない。
- 確実情報　エージェントの行動によって得られる報酬や環境の変化が確率的である。
- 対称情報　どのエージェントも有用な私的な情報をもっていない。
- 完備情報　学習開始以前の環境の前提をすべてのエージェントが知っている。

エージェント i が環境の状態 s と他のエージェントの行動 a_{-i} を知っていれば、$Q_i(s, a, a_{-i})$はゲーム理論の利得行列そのものになり、最適戦略や均衡戦略は 2-4-1 に挙げたような考え方で決まる。

しかし一般に、エージェント i が行動決定時に所有している情報は、環境の一部の状態 s と自分の行動 a から得られるべき報酬を推測した行動価値関数 $Q_i(s, a)$、環境の一部を知覚して得られた現在の状態 s_i、自分の行動 a_i のみである。すなわち、マルチエージェント強化学習においては、情報は非対称である。各エージェントは環境を通してのみ他のエージェントの影響を知るため、各エージェントの持っている情報 $Q_i(s, a)$、s_i、a_i はそのエージェント固有のものであるからである。また、エージェントが環境に対して持っている情報は s_i であるが、状態空間の爆発を避けるために環境の一部しか知覚していないならば、その情報は不完全である。

このため、マルチエージェントの状況下で学習を行わせる際には考慮すべき特有の課題がいくつかある。[19]

まず、複数のエージェントが独立に学習する場合、自分の学習した結果が自分の行動によるものなのか他のエージェントの行動によるものなのか判別できず、適切な学習が困難であるという問題がある。すべてのエージェントが環境の状態を正確に観察してはいない、すなわち情報の非完全性が原因である。これは同時学習問題として知られており、環境を通して相互作用を行うマルチエージェントの枠組みでは避けられない問題である。

次に、複数のエージェントが相互作用を行いながらひとつの問題解決を行うことから、環境がある状態に達した際にどのエージェントへどれだけの報酬を配分するかが自明ではないという問題がある。マルチエージェントではない一般の強化学習下においても、それ以前に到達した各状態にも報酬を配分するかどうかという問題があり、信用（信頼度）割り当て問題として知られている。

最後に、環境の状態を正確に知覚しようとすると、採りうる状態が膨大な組み合わせになり、実時間内での学習を行うことが不可能になってしまうという問題がある。これは状態空間の爆発として知られている。

2-6-3 ポリエージェントシステム[20]

エージェント間に相互作用が存在する場合、システム全体の目的と各エージェントの目的に不整合が生じる。一般に、システム全体の目的は全体の最適解（パレート最適解）を求めることであるが、各エージェントが目的を個別に達成しようとしたときに到達する解は個人合理性に基づく解（ナッシュ均衡解）になる。[21]

これを回避するひとつの方法は 2-6-2 項でも指摘したように適切に設計を報酬することであるが、個々のエージェント（ミクロレベル）とシステム全体の集計変量（マクロレベル）の関係を「間接制御」するエージェントを導入する方法もある。このエージェントは、何らかの組織目標を充足するように、エージェントの活動の境界条件を調整する。

このような機能や状態構造の階層性に注目したマルチエージェントシステムは、特にポリ（重合）エージェントシステム、または他主体複雑系と呼ばれている。すなわち、同質のエージェントが複数存在するだけのシステムでなく、構造上の上位のエージェントを形成するメカニズムや異質のエージェント間の機能的な階層関係を分析の対象とするシステムである。

多主体複雑系のモデル観を図 2.7 に示す。

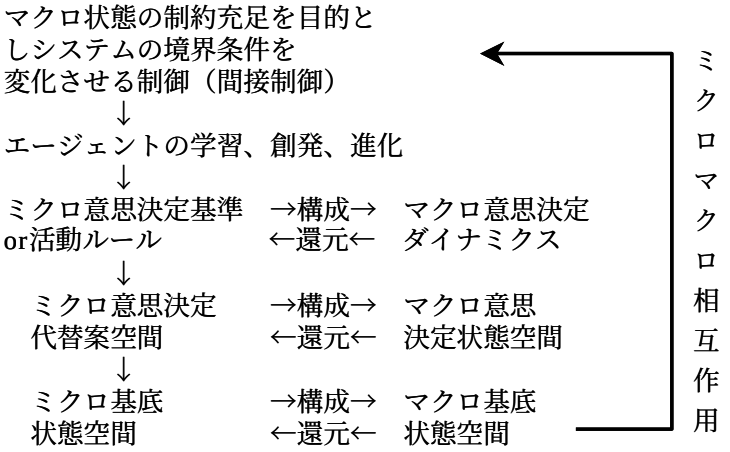


図 2.7　多主体複雑系のモデル観[19]

2-6-4 ポリエージェントシステム中の政府機関の役割

ポリエージェントシステムの枠組みは、社会的なネットワーク・システムのあり方の分析などにおいて概念的に用いられていることが多い。実社会をシステムとして考えると、人間という主体が存在し、それが集まって企業などの主体を構成し、それらの活動を政府という主体が制御している、つまりポリエージェントシステムとしての捉え方が可能である。そのため、システム中に「企業」と「政府」という異質のエージェントを導入することは、社会的な視点からは自然なことである。

このような「政府」を導入したポリエージェントの枠組みを学習と結びつけたものの例としては、「共有地の悲劇」を強化学習の一種であるクラシファイアシステムによってモデル化した研究[19]がある。共有地の悲劇とは、有限の共有地を放牧のような形で複数の主体が使うとき、個人が共有地を自己利益のために過剰に利用するために共有地が疲弊して、結果として資源の枯渇した状態になってしまうという状況を表

したものである。エージェントの社会活動を羊の売買、食料購入、消費として自己利益の最大化を行わせた結果では、明らかな一人勝ち構造の結果が得られ、多数のエージェントは破産した。そこに、税金と補助金、コミュニケーションの要素を加えることで、より安定した社会が実現されたという。

電力市場に多数の発電事業者が参入する状況でも、多数の発電事業者が参入することにより、電力価格が下落して一つの発電事業者あたりの利益が損なわれたり、発電事業者が参入せず、需要家にとっての利益が損なわれたりすることも考えられる。そのため、政府機関エージェントを導入し、税金や補助金といった手段によって電力の供給状態を監視し、より安定な電力供給を実現することを考えることができよう。

2-7 平均回帰過程

エネルギー価格や金利の変動をモデル化するために、平均回帰過程が用いられる。平均回帰とは、変動が大きいほど平均値に向かって引き戻される動きも大きくなり、長期的には一定の平均値に収斂するという性質である。

変動変数を x、その平均を $\bar{x}$、時間を t、変動の標準偏差を σ とすれば、その変動分 dx は式(2.12)で表せる。

$$dx = \eta(\bar{x} - x)dt + \sigma dz \tag{2.12}$$

第一項が時間の経過に伴う平均への回帰を示しており、η は回帰速度と呼ばれる。

第二項中の z はウィーナー過程に従う変数であり、連続だが微分不可能という特徴を持つ。ε を標準正規分布に従う変数とすると、その変動分 dz は式(2.13)で表せる。

$$dz = \sqrt{dt}\,\varepsilon \tag{2.13}$$

第3章 マルチエージェント電力供給モデルの構築

前章で述べた手法を用いて、電力供給における利害関係者に注目し、マルチエージェント強化学習型の電力需給モデルを構築した。本章ではモデルの構成要素である電力市場、発電事業者、電力需要、政府機関、燃料価格のモデル化について説明する。

3-1 モデルの概要

3-1-1 モデルの対象

2000 年から 2030 年までの電力供給を対象とし、地域区分や送電については考えない一地域の電源構成計画をモデル化した。2000 年から 2030 年までを 5 年間隔に分割して 6 時点を設け、各時点を 1 季節･24 時間帯に分割した。

割引率は 5%とした。詳細は 3-1-2 項で示す。

この電力供給モデルは図 3.1 に示すように、電力市場、電力需要、発電事業者、政府機関、燃料価格の各構成要素から成る。このうち決定変数を持つエージェントは発電事業者と政府機関である。発電事業者エージェントは、電力会社のように、単体だけでなくグループとして行動することも考えられる。このように異種のエージェントやエージェント集団が混在する状況は、2-6-1 で述べたポリエージェントシステムとなっている。

これらを用いた実際の計算の流れを図 3.2 に示す。2000 年から 2030 年までの 1 回の流れを「エピソード」と呼び、このエピソードを繰り返すことで発電事業者と政府機関が持つ決定変数を決定する。

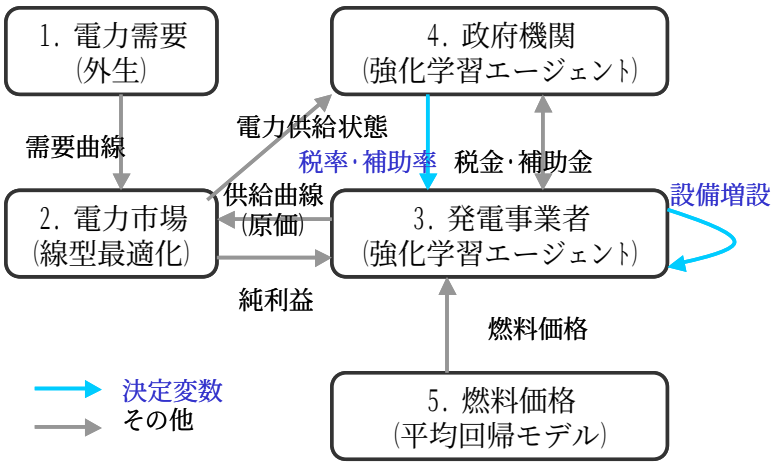


図 3.1 「マルチエージェント型一地域電力供給モデル」の構成要素
(図 2.1 の再掲)

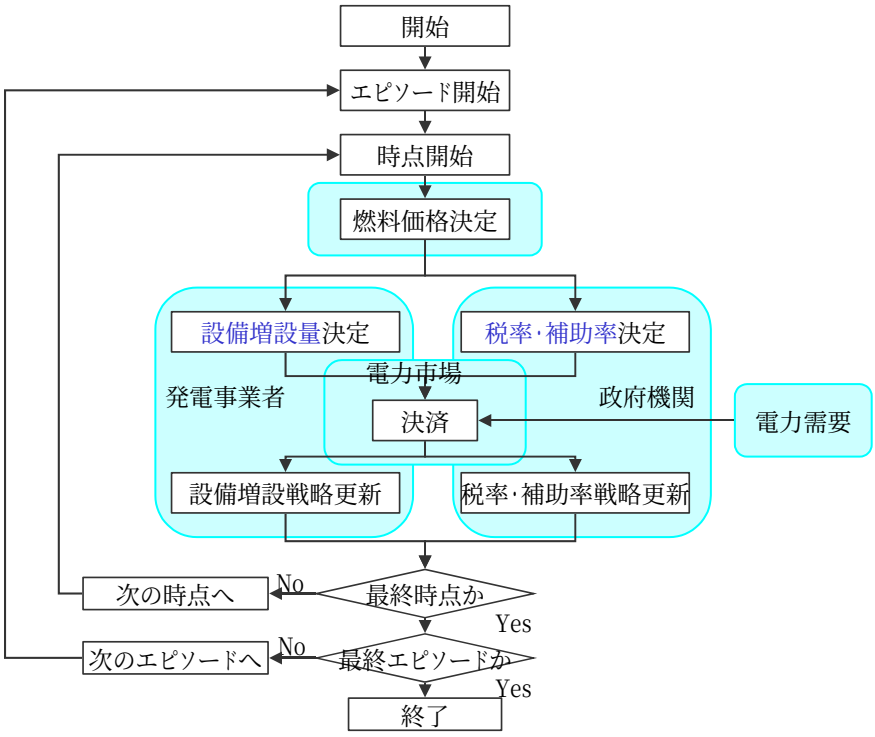


図 3.2　計算フロー

3-1-2 割引率

割引率は単位時間あたりの価値の低減度合いを示す値である。割引率を r とすれば、y 年目での 1 単位の価値 v は、式(3.1)に従って初年度（0 年目）の価値に換算される。

$$v(y) = \frac{1}{(1+i)^y} \tag{3.1}$$

名目利子率からインフレ分を取り除いた実質利子率を i としたとき、初年度で実質価値 1 によって y 年目には実質価値 $(1+i)^y$ を手に入れることができるので、これらの価値は等しいと考えれば、式(3.1)と同様の換算式が得られる。このため、割引率は実質利子率と同じ値に設定することができる。一般には、割引率は、現在と将来が同じ状況であったとしてもどちらにより多く関心を払うかという時間的選好の度合い（時間割引）、生活水準向上を前提に将来の世代への費用を転じるという考え方（成長割引）の概念を含むものである。[22]

2000 年から 2030 年まで 5 年間隔で時点を設け、各時点において 1 年分の電力需給をモデル化しているが、実際は各時点は図 3.3 に示すように三角パルス型の関数 $W_t(y)$で重み付けられた複数年の状況を代表しているものである。

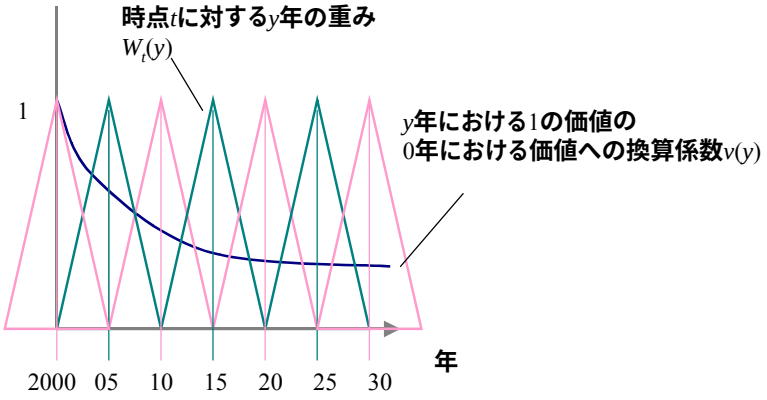


図 3.3　各時点に対する各年の重み付け

そのため、全期間に渡る価値の総和を考える際には、各時点の 1 年分の価値に、式(3.2)で示されるファクターdr_tを乗じて足し合わせる必要がある。今後、dr_tを「時点 t における割引率」と呼ぶことにする。

$$dr_t = \sum_y v(y) \times W_t(y) = \sum_y \frac{1}{(1+i)^y} \times W_t(y) \tag{3.2}$$

$i = 5\%$として設定した。時点 t における割引率を図示すると図 3.4 のようになる。
なお、本節中で使用している記号の説明を表 3.1 にまとめる。

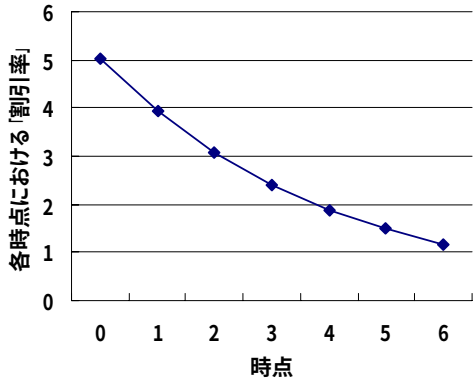


図 3.4　各時点における割引率

表 3.1　3-1-2 項中の記号

y：年	t：時点（5年間隔）
$i = 5\%$：利子率	dr_t：時点割引率
$v(y)$：価値換算係数	$W_t(y)$：時点重み付け係数

3-2 需要のモデル化

3-2-1 基準需要

簡略化のために需要の季節変動は考慮せず、2000 年の基準需要を図 3.5 のように 24 時間帯で与えた。この基準需要は年率 1.1%[23]で増加するものとした。この中には、技術の進歩や電力シフト、産業構造の変化によるエネルギー需要変化も含まれているものとする。

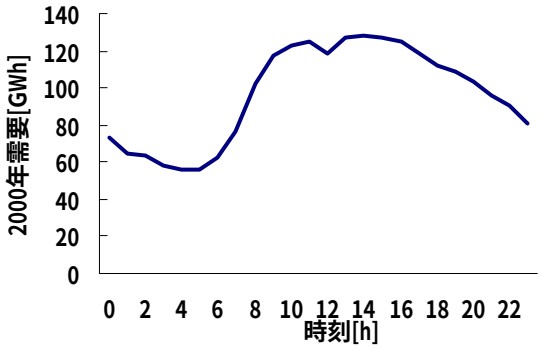


図 3.5　2000 年時間帯別基準需要

3-2-2 省エネルギー

技術進歩や産業構造の変化以外の原因で電力価格が高騰すると、需要の変化（省エネルギー）が起こりうる。ここでは需要の価格弾力性を利用して需要曲線を作成した。エネルギー需要の価格弾性値は、エネルギー価格が 1%変化したときにエネルギー需要が何%変化するかを表す値であり、その期内に表れる需要変化を考えたものを短期価格弾性値、製品購入や設備投資の変化まで含めた需要変化を考えたものを長期価格弾性値という。

このモデルにおいては、電力価格の高騰は電力設備量に直結した構造的な原因によるものであるから、それによる省エネルギーも長期的な需要変化の現れである。そのため、長期価格弾性値を用いる。価格弾性値として、ここでは－0.168[24]という値を採用した。

エネルギー需要 D、エネルギー価格を P とすると、D は P の関数となる。価格弾性値 $-\alpha$ の定義は

$$-\alpha = \frac{dD/D}{dP/P} \tag{3.3}$$

であり、一般に $\alpha > 0$ である。

基準需要が D_0、それに対応する基準となる電力価格が P_0、価格弾性値が $-\alpha$（$\alpha > 0$）であったとすると、価格弾性値の定義より、D と P の関係は、D と P がそれぞれ D_0 と P_0 の近辺において、

$$P(D) = P_0\left(\frac{D}{D_0}\right)^{-\frac{1}{\alpha}} \tag{3.4}$$

と表される。この曲線は需要曲線である。需要曲線と横軸の囲む面積は、消費者が払っても良いと考える価格の合計、つまり効用と呼ばれる量を表している。

ここで、価格上昇により省エネルギーが起こり、需要が $D(D < D_0)$ に減少したとする。供給量が D_0 のときに比較すると、供給量が D のときは、斜線部で表される面積の分だけ効用が減少している。これが、省エネルギーを行うためのコストであると考える。省エネルギー量を R という変数で表すと、

$$R = D_0 - D \tag{3.5}$$

である。省エネルギーコスト C は、R の関数として

$$C(R)=\int_D^{D0} P(D)dD=\int_0^R P_0\left(\frac{D_0-R}{D_0}\right)^{-\frac{1}{\alpha}}dR$$

$$=\begin{cases}\dfrac{\alpha}{1-\alpha}D_0P_0\left\{\left(1-\dfrac{R}{D_0}\right)^{\frac{\alpha-1}{\alpha}}-1\right\} & (\alpha\neq 1)\\ D_0P_0\log\left(\dfrac{D_0}{D_0-S}\right) & (\alpha=1)\end{cases}$$

と表せる。これより、単位省エネルギーあたりのコスト、つまりコスト係数 c は、式(3.7)で表せる。

$$c(R)=\frac{C(R)}{R} \tag{3.6}$$

また、省エネルギーとは逆に、価格低下に伴って需要は増加し、消費者の効用は増加する。これによるコスト低減分も同様に表すことができる。

線形計画へのモデル化の際には、この関数を線形化するためにステップ関数近似を行う。ある最終エネルギー需要のステップ番号 sp における対応する需要量を $(x(sp)-x(sp+1))\times D_0$ とする。$x(0)=1$ である。これまでに求めた式から、sp ステップ目の省エネルギーを行うのに必要なコスト $C'(sp)$ は次のように表される。

$$C'(sp)=\int_{x(sp+1)\times D_0}^{x(sp)\times D_0} P(D)dD=\frac{\alpha}{1-\alpha}D_0P_0\left\{x(sp)^{\frac{\alpha-1}{\alpha}}-x(sp+1)^{\frac{\alpha-1}{\alpha}}\right\}\quad(\alpha\neq 1) \tag{3.7}$$

これより、各ステップにおける省エネルギーコスト $c'(sp)$ は、

$$c'(sp)=\frac{C'(sp)}{(x(sp)-x(sp+1))\times D_0}=\frac{\alpha}{1-\alpha}\frac{x(sp)^{\frac{\alpha-1}{\alpha}}-x(sp+1)^{\frac{\alpha-1}{\alpha}}}{x(sp)-x(sp+1)}P_0\equiv a(sp)\times P_0 \tag{3.8}$$

と表すことができる。ここで、需要抑制コスト $c'(sp)$ のエネルギー価格 P_0 に対する割合を $a(sp)$ とおいた。

ステップの幅は、必要な近似に応じて任意に決定できる。ここではステップ幅を需要の 1%として、基準需要の 20%までの省エネルギーを許可し、20%の省エネルギーに対応する電力価格をプライスキャップとして設けた。また、価格低下に伴う需要増加分も基準需要の 20%までを許可し、同様にステップ関数近似を行った。基準電力価格 P_0 は、後に 4-1 節で示す「参照ケース」で計算したシャドウプライスを用いる。

このようにして作成する各時点の需要曲線のイメージを図 3.6 に示す。

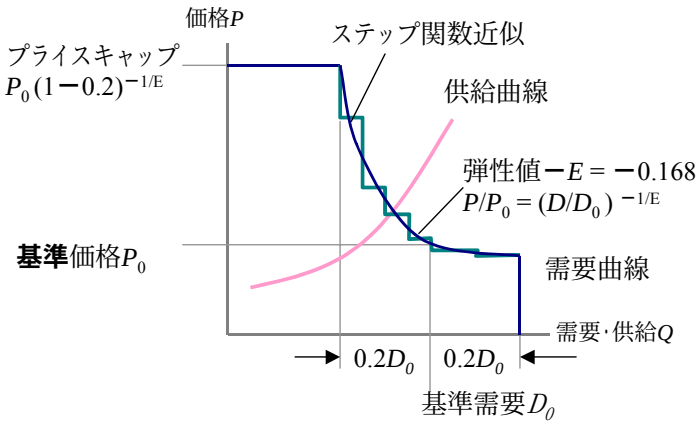


図 3.6　需要曲線

3-3 電力市場のモデル化

ここでは強制プール市場、つまり発電事業者側のシングルオークションにより電力市場をモデル化した。理想化された電力市場では、各時間帯において、発電事業者からの電力供給の入札を送電容量等を考慮しつつあるルールによって決済し、それぞれの発電事業者へ電力取引量とその価格を決定する。

1-1-3 項でも述べたように、実際の「電力市場」は、電力取引のための相対市場、スポット市場、リアルタイム市場、その他アンシラリーサービス市場や派生市場から成り立っている。ここでは、これを簡略化し、電力市場をスポット市場として模擬する。

電力市場の役割は、発電事業者から各時間帯に入札される価格と最大発電量をもとに、社会的厚生を最大化するように各発電事業者に発電量を割り当てることである。発電事業者数が十分多いときは完全競争が仮定でき、電力需要の価格弾力性が小さいことから、2-3 で述べたように、社会的厚生の最大化は供給コストの最小化として近似できる。

なお、この論文では発電事業者数が小さく完全競争が仮定できない状況も扱うが、簡単のために電力市場はすべて供給コストの最小化計算によって電力取引量が決定されるものとする。発電事業者と小売事業者の入札行動と市場決済をモデル化することについては、既に多数の研究が為されており、本研究では考慮の対象外とする。このときの発電事業者の入札価格は、燃料費に税金（負ならば補助金）を加算した価格である。

電力市場決済のための供給コスト最小化計算は、以下の式(3.9)の目的関数と制約式群(3.10)～(3.1)で表される線型計画として表している。数式中の記号の説明を表 3.2 に示す。

この線型計画問題は、市販の最適化パッケージである ILOG 社の CPLEX[25]を用いて解いている。なお、用いた最適化パッケージの性質で、問題が縮退しているとき、どの解が選択されるか、つまり、複数の発電事業者エージェントが同じコストで入札を行ったとき、どのエージェントに発電が割り当てられるかは不明である。そのため、一度最適化を行った後、割り当てられた発電量を同種の電源を持つ発電事業者エージェント間で合計し、それを各エージェントの所有設備量を用いて比例配分している。

目的関数

$$\sum_h \left(\sum_i (cy_{i,t} + ty_{i,t}) \times Y_{i,t,h} + \sum_k cc_{k,t,h} \times C_{k,t,h} + pc_{t,h} \times Z_{t,h} \right) \rightarrow \min \tag{3.9}$$

需給バランス

$$\sum_i Y_{i,t,h} + Z_{t,h} + \sum_k C_{k,t,h} = S_{t,h} + \left(1 + \frac{K\theta}{2}\right) \times D_{0t,h} \quad (\forall t,h) \tag{3.10}$$

各需要曲線ステップ上限　　$C_{k,t,h} < \theta \times D_{0_{t,h}} \quad (\forall k,t,h)$

(3.11)

発電出力制約　　$u_{i,h} \times F_{i,t} \geq Y_{i,t,h} \quad (\forall i,t,h)$

(3.12)

負荷追従制約　　$d_i^- \times Y_{i,t-1,h} \leq Y_{i,t,h} \leq d_i^+ \times Y_{i,t+1,h} \quad (\forall i,t,h)$

(3.13)

揚水入出力制約　　$u_{i,h} \times F_{i,t} \geq S_{i,t,h} \quad (i = storage, \forall t,h)$

(3.14)

揚水電力貯蔵バランス　　$\sum_h Y_{i,t,h} \leq Eff \times \sum_h S_{t,h} \quad (i = storage, \forall t)$

(3.15)

揚水貯蔵電力量上限制約　　$\sum_h S_{t,h} \leq M \times u_{i,h} \times F_{i,t} \quad (i = storage, \forall t)$

(3.16)

表 3.2　3-3 節中の記号

i : 発電所の種類　t : 時点　h : 時間帯　k : 需要曲線ステップ番号
K : 需要曲線ステップ数
変数:
$F_{i,t}$: 設備量　$Y_{i,t,h}$: 発電量　$Z_{t,h}$: 電力不足量　$S_{t,h}$: 揚水動力　$C_{k,t,h}$: 需要変化量
コスト係数:
$cy_{i,t,h}$: 可変費単価　$ty_{i,t}$: 発電量税率　$cc_{k,t,h}$: 省エネルギーコスト　$pc_{t,h}$: プライスキャップ
係数:
$r_{i,t',t}$: 設備残存率　$F_{initial\,i,t}$: 初期設備残存分
$u_{i,h}$: 設備利用率　Eff : 揚水電力貯蔵効率　M : 揚水容量係数
右辺定数項:
$D_{0_{t,h}}$: 基準需要　θ : 需要曲線ステップ幅

3-4 発電事業者のモデル化

3-4-1 発電設備に関する設定

発電設備の種類として、原子力、石炭火力、IGCC（二酸化炭素回収設備付）、ガス火力、石油火力、一般水力、揚水水力を対象とした。これらのプラントは建設決定後、建設年数を経た後から耐用年数を経るまで使用することができる。表 3.3 に発電プラントの主なパラメータを示す。発電効率の改善としては図 3.7 を仮定している。

表 3.3　発電プラントのパラメータ[26]他

	一般水力	原子力	石炭火力	IGCC	ガス火力	石油火力	風力	太陽光	揚水水力
固定費[万円/kW]	40.4	37.7	30.4	49.0	23.2	20.6	19.0	150	19.6
可変費[円/kWh] *1	0	1.65	1.23	1.34	3.6	4.52	0	0	0
耐用年数[年]	500	40	40	40	40	40	15	20	500
償却年数[年]	15	15	15	15	15	15	10	10	15
建設年数[年]	5	10	5	5	5	5	5	5	5
運転比率[%]*1	1.2	4.0	3.8	3.8	3.6	3.9	1.0	1.0	1.2
負荷追従率[%/h]*1	100	75~110	70~125	70~125	10~130	50~140	100	—	0~∞
設備量上限[GW]	40.4	50.0	∞	∞	∞	∞	2.7	42.0	40.0
設備利用率[%]	95	95	95	95	95	95	25	平均 12	95
CO2 排出係数 [kg-C/kWh] *2	0	0	0.237	0.024	0.124	0.183	0	0	0

*1: 初年度の値。時間（3-6 節の「燃料価格のモデル化」を参照）や発電効率改善（図 3.7 を参照））と共に変化する。
*2: 初年度の値。発電効率改善（図 3.7 を参照））と共に変化する。

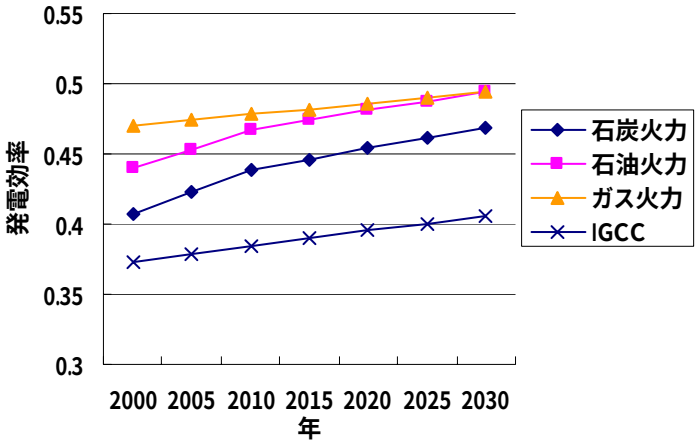


図 3.7　発電効率の改善

3-4-2 発電設備に係る費用

考慮期間は 2000 年～2030 年であるが、初年度における設備量は外生的に与えた。この設備は、それまでの期間に毎年等量ずつ建設された設備が残存している結果であると仮定し、時間の経過とともに同じ割合ずつで停止措置がとられるものとする。

発電設備増設者は、建設を決定した時点にそれ以後発生する固定費を全額支払い、運転期間中は発電量に比例する可変費のみを支払うものと仮定した。固定費は設備量に比例して決まる費用であり、可変費は発電量に比例して決まる費用である。

固定費は設備量によって定まる費用であり、減価償却費、利子、固定資産税、運転費からなる。償却期間中は減価償却費、利子、固定資産税、運転費が発生し、償却期間以後は運転費のみが発生する。

容量 X の設備を建設したときの設備投資額 I は、建設単価を pf とすれば式(3.17)で表せる。

$$I = pf \times X \tag{3.17}$$

運転費は設備の補修費や人件費などの総称であり、全設備投資額 I に比例する。その比例定数は運転比率 m である。

減価償却費は、償却期間中に減損していく固定資産の価値であり、これが決算の際の損失となる。償却期間 N[年]中毎年定額が償却されるものとし、償却期間後の固定資産の価値（残存価値）は 0 であると仮定すると、減価償却費 D は式(3.18)で表される。

$$D = \frac{I}{N} \tag{3.18}$$

利子と固定資産税は、減価償却により減損した資産額に比例して発生する。その比例定数である利子率と固定資産税率はそれぞれ i, f とする。設備稼働 y 年目初頭における償却後固定資産額 A_y は、式(3.19)で表せる。

$$A_y = I - D \times (y-1) = I \times \left(1 - \frac{y-1}{N}\right) \tag{3.19}$$

これらより、設備稼働 y 年目に発生する固定費は、L を耐用年数として、式(3.20)のように表せる。

$$FC_y = \begin{cases} D + (i+f) \times A_y + m \times I = \left\{\frac{1}{N} + (i+f) \times \left(1 - \frac{y-1}{N}\right) + m\right\} \times I & (0 \le y \le N) \\ m \times I & (N < y \le L) \end{cases} \tag{3.20}$$

以上挙げてきた固定費の概念について、図 3.8 にまとめる。

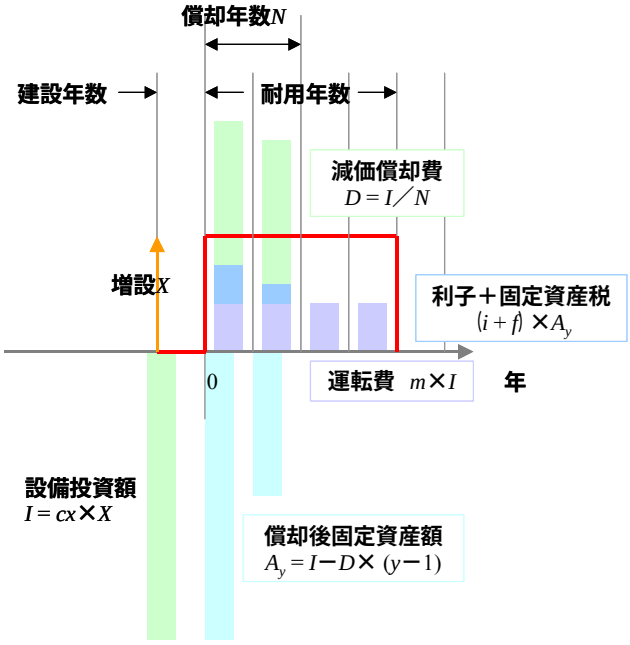


図 3.8　発電設備建設に伴う固定費の発生

さて、設備稼働 y 年目に発生する固定費のうち、設備投資額 I に比例する部分を年経費率 er_y と呼ぶ。すなわち、

$$er_y = \begin{cases} \frac{1}{N} + (i+f) \times \left(1 - \frac{y-1}{N}\right) + m & (0 \le y \le N) \\ m & (N < y \le L) \end{cases} \tag{3.21}$$

である。

年経費率を初年度における費用に換算し、設備が運転される期間に渡って足し合わせると、設備の建設によって発生する全固定費の、設備投資額 I に対する比例定数（ここでは累積年経費率と呼ぶ）ER が得られる。

$$ER = \sum_{y=0}^{L} I(y) \times er_y \tag{3.22}$$

このモデルにおける考慮期間は 2000 年から 2030 年に限定しており、5 年毎に時点を 0 から $T-1$ まで設定している。このため、累積年経費率を耐用年数 L と建設年数 C、考慮期間の残り$(T-t)\times 5$ を用いて補正する。それに建設単価 pf を乗じたものを時点 t における固定費単価 cx_t としている。固定単価 cx_t は(3.22)で表せる。

$$cx_t = \begin{cases} ER \times \dfrac{(T-t)\times 5}{L} \times pf & (C+L > (T-t)\times 5) \\ ER \times pf & (C+L < (T-t)\times 5) \end{cases} \tag{3.23}$$

なお、本項中で使用している記号の説明を表 3.4 にまとめる。

表 3.4　3-4-2 項中の記号

y：年	T：総時点数	
cx：建設単価	er：年経費率	ER：累積年経費率
m：運転比率	i：利子率	f：固定資産税率
I：設備投資額	D：減価償却費	A：償却後固定資産額
N：償却年数	L：耐用年数	C：建設年数

3-4-3 強化学習による動的意思決定の定式化

発電事業者は、時点と設備量からなる状態空間 (t, f) を探索し、各時点において発電設備の増設量を決定する。なお今後、設備量 f のみを指して「状態」と呼ぶことがある。図 3.9 にその概念図を示す。また、本項で用いる記号を表 3.5 に示す。

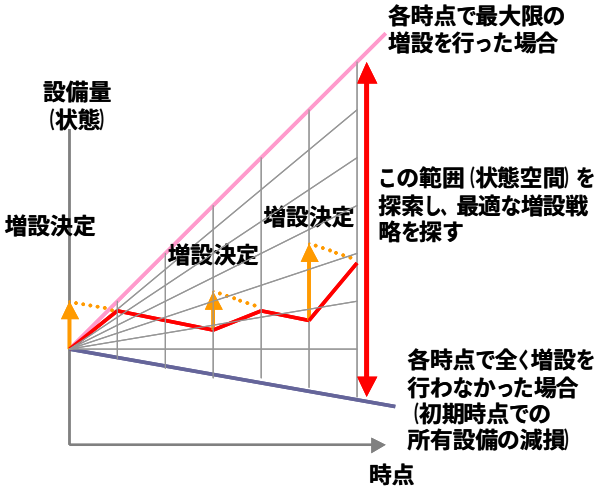


図 3.9　設備増設者の意思決定

報酬関数

各発電事業者エージェント i は、各時点での行動、つまり、$(X_{i,0}, X_{i,1}, X_{i,2}, \ldots, X_{i,T-1})$の設備増設量を表す非負変数の値を強化学習により決定する。エージェントには、式(3.25)で表される純利益 $R_{i,t,f}$ を報酬として与える。純利益は、設備建設固定費 cx、燃料費 cy、環境から与えられる電力価格 p、設備増設量 X、発電量 Y、そして割引率 dr などを用いて表現される。強化学習エージェントの目的は報酬の総和の最大化、つまり純利益総和の最大化である。

$$R_{i,t,f} \equiv dr_t \times \left\{ -cx_{i,t} \times X_{i,t} + \sum_h \left(p_{i,t,h} - cy_{i,t} \right) \times Y_{i,t,h} \right\} \tag{3.24}$$

なお、政府機関エージェントが介入した場合は、式(3.25)のように、その時点で保有している設備量 F、設備増設量 X、発電量 Y に比例する税金 tf、tx、ty が係数として加わる。

$$R_{i,t,f} \equiv dr_t \times \left\{ -tf_{i,t} \times F_{i,t} - \left(cx_{i,t} + tx_{i,f} \right) \times X_{i,t} + \sum_h \left(p_{i,t,h} - cy_{i,t} - ty_{i,t} \right) \times Y_{i,t,h} \right\} \tag{3.25}$$

方策評価

報酬関数(3.25)の特徴は、各エージェントにとって情報的に自明な項とそうでない項に分かれていることである。すなわち、各エージェントは、自分が増設した設備量 X、設備建設固定費 cx、燃料費 cy を知っているが、電力価格 p、発電量 Y は環境から与えられる値であり、ここには他のエージェントの行動も影響している。すなわち、関数 $r_{i,t}(X)$ と $U_{i,t,f}$ をそれぞれ式(3.27)、(3.27)のように定義すれば、純利益 $R_{i,t,f}$ は式(3.28)で表すことができる。エージェントが学習によって行うべきことは関数 $U_{i,t,f}$ を適切に推定することとなる。関数 $U_{i,t,f}$ は、設備費を除いた売電による利益を表している。

$$r_{i,t}(X) \equiv -dr_t \times cx_{i,t} \times X \tag{3.26}$$

$$U_{i,t,f} \equiv dr_t \times \sum_h \left(p_{i,t,h} - cy_{i,t} \right) \times Y_{i,t,h} \tag{3.27}$$

$$R_{i,t,f} = r_{i,t}(X_{i,t}) + U_{i,t,f} \tag{3.28}$$

状態価値関数 $V_{i,t,f}$ を、時点 t で所有する設備量が f のとき、以降の時点で獲得できる純利益総和の推定値を表すものとして導入する。$V_{i,t,f}$、$U_{i,t,f}$ が適切に推定されれば、設備建設期間が 1 期間であれば、これらの間には式(3.29)の関係が成り立つ。f' は次の時点での所用設備量であり、エージェントはこれを今の所有設備量とそれまでの設備建設量から正確に知ることができる。

$$V^*_{i,t,f} = \max_X \left\{ r_{i,t}(X) + U^*_{i,t,f} + V^*_{i,t+1,f'} \right\} \tag{3.29}$$

以下、$U_{i,t,f}$ を補助状態価値関数と呼ぶことにする。

状態価値関数 $V_{i,t,f}$、補助状態価値関数 $U_{i,t,f}$ の推定は、エージェントはエピソードの繰り返しによる経験の積み重ねを通じて行う。補助状態価値関数 $U_{i,t,f}$ の更新は(3.31)で表されるように、k 回目のエピソード中の売電利益(3.30)と、現在の推定値の差分を用いて行われる。

$$U^k_{i,t,f_t} = R^k_{i,t,f_t} - r_{i,t}(X^k_{i,t}) \tag{3.30}$$

$$U_{i,t,f_t} \leftarrow U_{i,t,f_t} + \alpha \times \left\{ U^k_{i,t,f_t} - U_{i,t,f_t} \right\} \tag{3.31}$$

状態価値関数$V_{i,t,f}$は補助状態価値関数$U_{i,t,f}$を用いて(3.32)のように計算できる。

$$V_{i,t,f_t} = \max_{X_{i,t}} \left\{ r_{i,t}(X_{i,t}) + U_{i,t,f_t} + V_{i,t+1,f_t'} \right\} \tag{3.32}$$

なお、本来のアクタークリティック法では、状態価値関数を計算するために(3.33)のようにXをスイープするようなことはしないで、得られた報酬を直接用いた更新を行っている。

$$V_{i,t,f_t} \leftarrow V_{i,t,f_t} + \alpha \times \left\{ R_{i,t}^k + V_{i,t+1,f'} - V_{i,t,f_t} \right\} \tag{3.33}$$

方策改善

最適な発電設備増設量は十分な経験を積み状態価値関数$V_{i,t,f}$を適切に推定できれば自ら決定されることになるが、ここでは価値関数とは独立に行動を決定する方法をとる。方策は平均μと標準偏差σをパラメータとする正規分布で表される行動選択確率で表現し、この平均と標準偏差を更新することにより、間接的に最適行動に近づける。

$$X_{i,t+1} = NormalRandom(\mu_{i,t,f_t}, \sigma_{i,t,f_t}) \tag{3.34}$$

ここでは、(3.32)のスイープによって得られた最適な行動に、現在の行動の平均値を近づけるような方策改善を行う。平均行動がもたらすであろう総報酬と、最適行動がもたらすであろう総報酬の差分をTD誤差と呼び、(3.35)で定義する。この値は正になる。

$$TDerror_{i,t}^k \equiv \left\{ r_{i,t}(\mu_{i,t}) + U_{i,t,f_t} + V_{i,t+1,f'} \right\} - V_{i,t,f_t} \tag{3.35}$$

このときの方策改善の方向は、図3.10に示されるように$X-\mu$、$|X-\mu|-\sigma$の正負によって定まる。すなわち、平均μは、実行した行動Xに近づけるように更新し、標準偏差σは、行動が分布の外側なら大きく、分布の内側なら小さくなるように更新する。

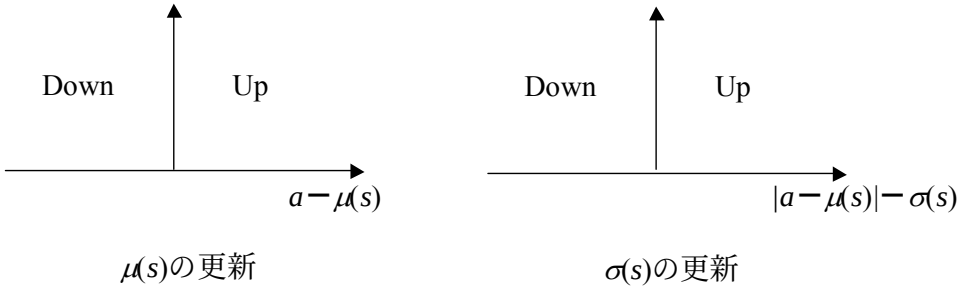


図3.10　エージェントの方策改善

この方策改善の方向に従って、式(3.36)と式(3.39)式の複号のうちいずれかを選択して平均と標準偏差を更新する。δ_μとδ_σは更新に用いる定数である。

$$\mu_{i,t,f_t} = \left(1 \pm \delta_\mu\right) \times \mu_{i,t,f_t} \tag{3.36}$$

$$\sigma_{i,t,f_t} = \left(1 \pm \delta_\mu\right) \times \sigma_{i,t,f_t} \tag{3.37}$$

この学習手法は、アクタークリティック手法とQ学習手法を組み合わせたような形になっている。行動価値関数を用いて行動を決定するのではないという点でQ学習と異なる。また、価値関数の更新に実際に

とった行動のみを用いるのではない（式 (3.32)）という点で、方策オン型である本来のアクタークリティック手法とも異なる。

表 3.5　3-4-3 項中の記号

i : エージェント番号　t : 時点　h : 時間帯　f : 離散化した設備量　k : エピソード数
報酬関数：
$X_{i,t}$: 設備増設量　$F_{i,t}$: 設備量　$Y_{i,t,h}$: 発電量
$dr_{i,t}$: 割引率　$cx_{i,t}$: 固定費単価
$tf_{i,t}$: 設備量税率　$tx_{i,t}$: 増設量税率　$ty_{i,t,h}$: 発電量税率
$p_{t,h}$: 電力価格　$cy_{i,t,h}$: 可変費単価
価値関数の更新：
$V_{i,t,f}$: 状態価値関数　$U_{i,t,f}$: 補助状態価値関数
$R_{i,t,f}$: 利益　$TDerror_{i,t}$: TD誤差
$\alpha = 0.1$: 学習率
方策の更新：
$\mu_{i,t,f}$: 行動選択確率の平均　$\sigma_{i,t,f}$: 行動選択確率の標準偏差
δ_{μ} : 平均更新率　δ_{σ} : 標準偏差更新率

3-4-4 電力会社のモデル化

電力会社のように、一社で複数の電源を持って連結決算を行い、その合計利益を最大化する目的を持つ主体は、報酬関数をその集団に所属しているすべてのエージェントの利益和として式(3.38)のように与えたエージェント集団としで表現することができる。ここで、$I(i)$はエージェントが所属している集団を表す。

$$R'_{i,t,f} \equiv \sum_{i' \in I(i)} dr_t \times \left\{ -cx_{i',t} \times X_{i',t} + \sum_h \left(p_{i',t,h} - cy_{i',t} \right) \times Y_{i',t,h} \right\} \tag{3.38}$$

ここでは、集団に属するエージェントもそうでないエージェントも、保有している情報は自らの所有設備量、増設量とコスト係数のみであり、他のエージェントの情報は一切用いていない。そのため、2-6-2 項で挙げたような同時学習による影響がより強まることは必須である。しかし、状態空間をできるだけ少なくすることを考え、状態としては自エージェントの時点と所有設備量のみを用いた。また、集団全体と個人の利益をある係数で重み付けするような報酬設計も有効であるかもしれない。

3-5 政府機関のモデル化

二酸化炭素排出原単位を一定値以下に抑える、電力予備力を確保する、電源のミックスを考慮して安定供給を確保するなどの制約は、発電事業者間にまたがるものであり、局所的な利益を最大化しようとする発電事業者エージェントのみの分散活動では充足できない。そのため、ある特定の目的を持ち、政府機関エージェントは、電力供給状況を大局的に見て、充足すべき制約を税金や補助金により間接的に制御する主体を導入し、これを「政府機関エージェント」と呼ぶことにする。

政府機関エージェントは、前時点の電力供給状態指標を参考に、今時点の発電事業者へ課す税率を決定し、発電事業者の行動に対して税を徴収する。そして、その税率の効果を次時点の電力供給状態指標から評価する。

3-5-1 強化学習による意思決定

政府機関エージェント一般について、強化学習によるその意思決定方法の定式化を説明する。ここで用いた方法はアクタークリティック手法である。表 3.6 に本項中で用いる記号の一覧を示す。

報酬関数

政府機関エージェント j は、各時点 t の電力供給状況指標 $index_{j,t}$ を目標値 $tar_{j,t}$ 以下にすることを目的とする。税率 $taxrate_{j,t}$（負ならば補助率）を設定することで発電発電事業者エージェントの行動に介入し、間接的に $index_{j,t}$ を変化させる。ただし、税収を上げることは目的ではなく、補助率は小さいほうが政府機関にとっては良いが、同様に同じ効果が得られるのならば税率は小さいほうが良い。

この政府機関エージェントは、式(3.39)で表される報酬 $R_{j,t}$ を持つ。

$$R_{j,t} \equiv -\max\left(index_{j,t}, tar_{j,t}\right) - \kappa \times \left| \frac{taxrate_{j.t}}{\overline{taxrate_{j,t}} - \underline{taxrate_{j,t}}} \right| \quad (\forall t) \tag{3.39}$$

実際に発電事業者から徴収する税額を計算するのに用いられる税率 $taxrate'$ は、式(3.40)で示すように、この政府機関が定めた税率 $taxrate$ に、今時点での電力供給状況指標の目標達成率を乗じた値とする。

$$taxrate'_{j,t} = \begin{cases} 0 & (index_{j,t} \le tar_{j,t}) \\ taxrate_{j.t} \times \dfrac{index_{j,t} - tar_{j,t}}{tar_{j,t}} & (index_{j,t} > tar_{j,t}) \end{cases} \tag{3.40}$$

方策評価

政府機関エージェントは、設備増設エージェントとは異なり、各時点での目標達成を行おうとするのみで将来については考慮しないものとした。

状態価値関数 $V_{j,t,s}$ を、時点が t、前時点 $t-1$ での電力供給状態指標が $index_{t-2} \equiv s_t$ のときに、時点 $t+1$ に達成できた電力供給状態指標（正確には得た報酬 $R_{j,t}$）の推定値を表すものとして定義した。すなわち、前時点の電力供給状態指標を参考に、今時点の発電事業者へ課す税率を決定し、発電事業者の行動に対して税を徴収する。そして、その税率の効果を次時点の電力供給状態指標から評価するということである。この概念図を図 3.11 に示す。

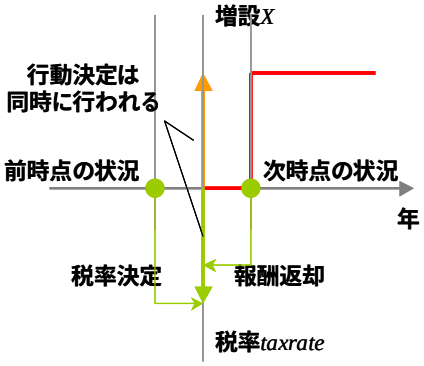


図 3.11　政府機関エージェントの意思決定の概念

この状態価値関数は、式(3.41)に従って更新する。

$$V_{j,t,s_t} \leftarrow V_{j,t,s_t} + \alpha \times (R^k_{j,t} - V_{j,t,s_t}) \tag{3.41}$$

方策改善

試行から得られた報酬 $R_{j,t}$ と状態価値関数 $V_{j,t,s}$ の差分を用いて、方策の改善を行う。この差を TD 誤差と呼び、(3.35)で定義する。

$$TDerror_{j,t}^{k} \equiv R_{j,t}^{k} - V_{j,t,s_t} \tag{3.42}$$

最適な発電設備増設量は十分な経験を積み状態価値関数 $V_{j,t,f}$ を適切に推定できれば自ら決定されることになるが、ここでは価値関数とは独立に行動を決定する方法をとる。方策は平均μと標準偏差σをパラメータとする正規分布で表される行動選択確率で表現し、この平均と標準偏差を更新することにより、間接的に最適行動に近づける。

$$taxrate_{j,t+1} = NormalRandom(\mu_{j,t,s_t}, \sigma_{j,t,s_t}) \tag{3.43}$$

このときの方策改善の方向は、図 3.10 に示されるように *TDerror*、*taxrate*－μ、|*taxrate*－μ|－σの正負によって定まる。すなわち、TD 誤差が正であれば、平均μは、実行した行動 *X* に近づけるように更新し、標準偏差σは、行動が分布の外側なら大きく、分布の内側なら小さくなるように更新する。

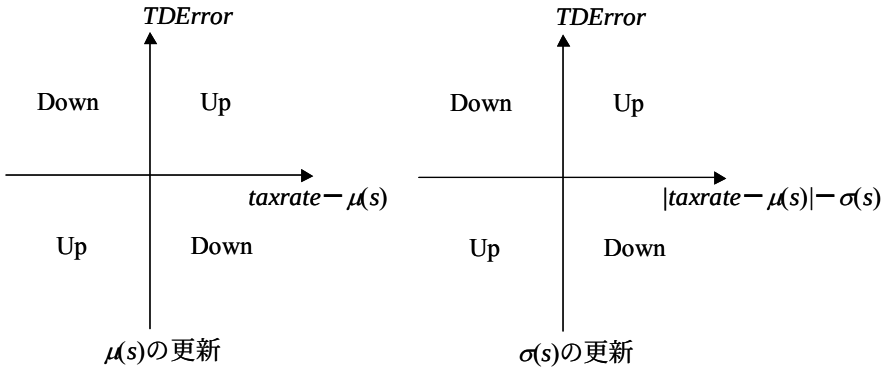


図 3.12　エージェントの方策改善

この方策改善の方向に従って、式(3.44)と式(3.45)の複号のうちいずれかを選択して平均と標準偏差を更新する。δ'_{μ} と δ'_{σ} は更新に用いる定数である。

$$\mu_{j,t,s_t} = \left(1 \pm \delta'_{\mu}\right) \times \mu_{j,t,s_t} \tag{3.44}$$

$$\sigma_{j,t,s_t} = \left(1 \pm \delta'_{\mu}\right) \times \sigma_{j,t,s_t} \tag{3.45}$$

表 3.6　3-5-1 項中の記号

j : 政府機関エージェント番号
$index_{j,t}$: 電力供給状況指標
$tar_{j,t}$: 排出原単位目標　$taxrate_{j,t}$: 炭素税率
$\overline{taxrate_{j,t}}$: 炭素税率上限　$\underline{taxrate_{j,t}}$: 炭素税率下限
$\kappa = 0.01$: 重み付け定数

3-5-2 設備確保エージェント

各時点の予備率を目標値以上にすることで電力供給不足による価格上昇や効用の低下の阻止を目的とする政府機関エージェントを導入し、設備確保エージェントと呼ぶ。すなわち、監視する電力供給状況は予備力であり、式(3.47)で表される。負号は、この指標は大きいほど目的に適うことを示している。

$$index_{予備力確保,t} \equiv -\left(1 - \frac{L_{t,peakh}}{\sum_i F_{i,t}}\right) \quad (\forall t) \tag{3.46}$$

i:設備所有エージェント番号　t:時点　$pealh$:ピーク時間帯
$F_{i,t}$:設備量　$L_{t,peakh}$:ピーク需要

この電源計画モデルでは、需要は中間期のものしか与えていないため、実際のピーク需要の鋭さとの対応を考えれば予備力は 20%程度の目標値が妥当であると思われる。

この予備力単位目標を達成するため、設備確保エージェントは設備補助金 [円/kW]を発電事業者エージェントの設備量に従って支払う。表 3.3 に示したように、設備建設には kW あたり数十万円のコストが必要である。そのため、設備建設のインセンティブを与えるためには、kW あたり十万円弱オーダーの補助率が必要なのではないかと考え、仮にその上限を 100,000[円/kW]と設定した。

3-5-3 温暖化対策エージェント

地球温暖化対策として、二酸化炭素排出原単位目標が達成されるように発電事業者の戦略に介入する政府機関エージェントを導入した。すなわち、監視する電力供給状況は、二酸化炭素排出原単位であり、式(3.47)で表される。

$$index_{温暖化対策,t} \equiv \frac{\sum_i \left(effi_i \times \sum_h Y_{i,t,h} \right)}{\sum_h Y_{i,t,h}} \quad (\forall t) \tag{3.47}$$

t:時点　h:時間帯
$Y_{i,t,h}$:発電量　$effi_i$:二酸化炭素排出係数

1-1-1 項で述べたように、電力業界団体である電気事業連合会は、「電気事業における環境行動計画」において、「2010 年度における使用端 CO_2 排出原単位を 1990 年度実績から 20%程度低減するよう努める」との目標を掲げている。[3]この目標値は二酸化炭素排出原単位を 0.093[kg-C/kWh]程度にすることに対応している。

そこで、2010 年の排出原単位目標値を 0.09[kg-C/kWh]とし、これが 2000 年度より 10%低減になるように 2000 年の目標値は 0.10[kg-C/kWh]と設定した。この二点を内挿･外挿することで、図 3.13 に示す排出原単位目標を設定した。

2030 年の排出原単位は 0.07[kg-C/kWh]で、2000 年よりも 30%減である。需要の増加は年率 1.1%を仮定しているため、この目標値は 2030 年の二酸化炭素排出量を 2000 年レベルに抑制することにも相当する。

この二酸化炭素排出原単位目標を達成するため、温暖化対策エージェントは炭素税 [円/kg-C]を発電事業者エージェントの発電量に従って課税する。

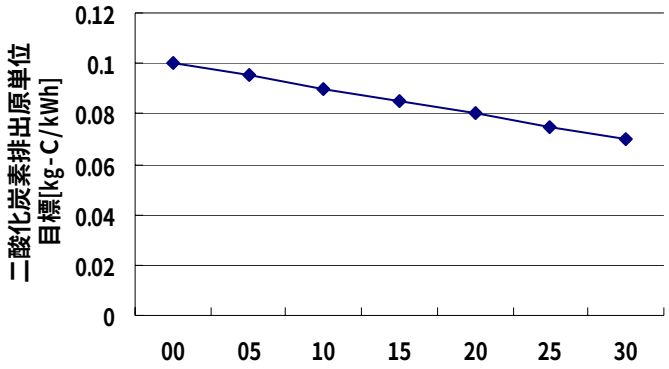


図 3.13　二酸化炭素排出原単位目標

3-6 燃料価格のモデル化

さて、石油、ガス、石炭、原子力（ウラン）の各燃料価格は、平均回帰過程を用いてモデル化した。1 年刻みの平均回帰過程として、5 年ごとの価格をその時点での燃料価格であるとする。

$$cx_{f,\tau+1} = cx_{f,\tau} + \eta\left(\overline{cx_{f,\tau}} - cx_{f,\tau}\right) + cov_{oil,f}\,\sigma \times \varepsilon \tag{3.48}$$

f : 燃料　τ : 年（1年間隔）

$\overline{cx_{f,\tau}}$: 燃料価格平均　$cx_{f,\tau}$: 燃料価格

$\eta = 0.4$: 平均回帰速度　$\varepsilon = NR(0,1)$: 標準正規乱数（共通）

$cov_{oil,f}$: 石油価格に対する相関係数　σ : ボラティリティ

石油価格を基本にして、それと各燃料との価格相関係数より石炭･ガス･原子力（ウラン）価格を決定した。ここでは価格の相関係数 *cov* は、表 3.7 のリスク分散･共分散行列から計算した表 3.8 の値を用いた。

表 3.7　リスク分散･共分散行列[27]

	石油	ガス	石炭	原子力	その他
石油	1.00	0.142	-0.010	-0.055	0
ガス	0.142	0.993	0.384	0.078	0
石炭	-0.010	0.384	0.134	-0.002	0
原子力	-0.055	0.078	-0.002	0.396	0
その他	0	0	0	0	0

表 3.8　価格相関係数

	石油	ガス	石炭	原子力
石油	1.00	0.142	-0.024	-0.087

燃料の平均価格として、2000 年の燃料価格実績と、2025 年まで燃料価格予測値を米国エネルギー省の統計[28] に基づいて設定し、それ以後は外挿として設定した。また、2005 年の燃料価格実績はトレンドよりも飛び抜けて高い値を示しているため、この値は用いず 2000 年と 2010 年の内挿として設定した。

このモデルによる石油価格変動の例を図 3.14 に示す。

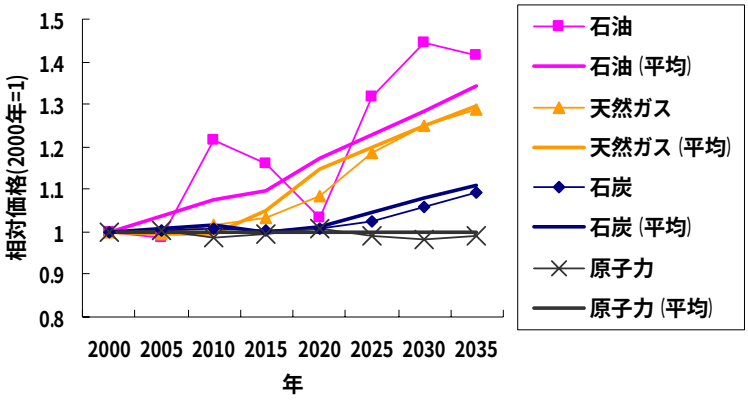


図 3.14　平均回帰過程による燃料価格変動例

第4章 モデルによる予備解析

本章では、解析の準備として、電源構成初期値の設定や、強化学習における適切なパラメータの探査を行った。

4-1 非エージェントモデルによる最適電源計画

作成したエージェント学習モデルでは、初期時点である 2000 年以前の既設設備を与えていない。実績値を元に設定することも可能であるが、需要やコスト設定の際においた仮定と整合性のとれない可能性があるため、ここでは類似の最適電源計画問題を設定することで、初期値の設定を行った。

4-1-1 参照ケースのモデル設定

電力の自由化が行われておらず、「電力会社」が、基準需要を満たすための電力供給コストを最小にするように電源計画を立てるケースを計算した。

参照ケースの線型計画を式群(4.1)～(4.9)に示す。これは 3-3 節に挙げた電力市場決済の定式化と類似しているが、目的関数(4.1)に設備建設量を表す項を含むことと、基準需要を満たすことを制約とするため需要変化量に関する部分が無いこと、そして前電源にまたがる制約である設備確保制約(4.8)と二酸化炭素排出原単位制約(4.9)を含むことが異なる。

設備確保目標は 20%、排出原単位目標は図 3.13 の通りとした。

本項中で用いた記号を表 4.1 にまとめる。

目的関数

$$\sum_t dr_t \times \left(\sum_i \left(cx_{i,t} \times X_{i,t} + \sum_h cy_{i,t} \times Y_{i,t,h} \right) \right) \rightarrow \min \tag{4.1}$$

需給バランス

$$\sum_i Y_{i,t,h} + Z_{t,h} = S_{t,h} + D_{0_{t,h}} \quad (\forall t,h) \tag{4.2}$$

発電出力制約

$$u_{i,h} \times F_{i,t} \geq Y_{i,t,h} \quad (\forall i,t,h) \tag{4.3}$$

負荷追従制約

$$d_i^- \times Y_{i,t-1,h} \leq Y_{i,t,h} \leq d_i^+ \times Y_{i,t+1,h} \quad (\forall i,t,h) \tag{4.4}$$

揚水入出力制約

$$u_{i,h} \times F_{i,t} \geq S_{i,t,h} \quad (i = storage, \forall t,h) \tag{4.5}$$

揚水電力貯蔵バランス

$$\sum_h Y_{i,t,h} \leq Eff \times \sum_h S_{t,h} \quad (i = storage, \forall t) \tag{4.6}$$

揚水貯蔵電力量上限制約

$$\sum_h S_{t,h} \leq M \times u_{i,h} \times F_{i,t} \quad (i = storage, \forall t) \tag{4.7}$$

設備確保制約

$$\sum_i u_{i,h} \times F_{i,t} \geq (1+\delta) \times L_{t,h} \quad (\forall t,h) \tag{4.8}$$

二酸化炭素排出原単位制約

$$\sum_i \left(effi_i \times \sum_h Y_{i,t,h} \right) - tar_t \times \sum_i \sum_h Y_{i,t,h} \leq 0 \quad (\forall t) \tag{4.9}$$

表 4.1　4-1-1 項中の記号

i : 発電所の種類　t : 時点　h : 時間帯
変数:
$X_{i,t}$: 増設量　$F_{i,t}$: 設備量　$Y_{i,t,h}$: 発電量　$S_{t,h}$: 揚水動力
コスト係数:
$cx_{i,t}$: 固定費単価　$cy_{i,t,h}$: 可変費単価
係数:
$r_{i,t',t}$: 設備残存率　$F_{initial\,i,t}$: 初期設備残存分
$u_{i,h}$: 設備利用率　Eff : 揚水電力貯蔵効率　M : 揚水容量係数
$effi_i$: 二酸化炭素排出係数　tar_t : 二酸化炭素排出原単位目標
右辺定数項:
$D_{0t,h}$: 負荷　δ : 供給予備力

4-1-2 参照ケースの結果

参照ケースの結果として得られた設備増設量と、その累積である設備構成を図 4.1 に示す。

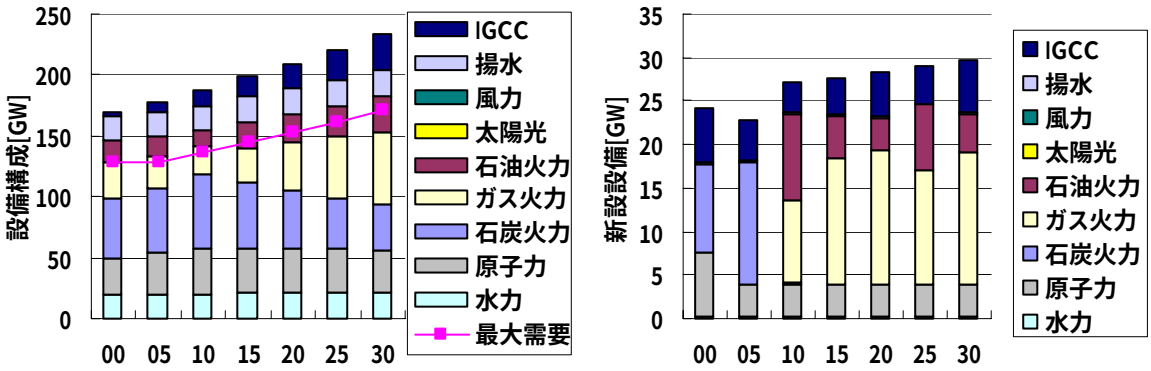


図 4.1　参照ケース　設備構成と増設分

需給バランス制約(4.2)のシャドウプライスは、その時点･時間帯における電力価格を表している。この価格は設備増設に要するコストを含んでおり、「電力会社」が設備投資分を完全に回収することのできる価格である。これは電力自由化前の電力価格に相当すると言える。以下のマルチエージェントシステムによる計算では、このケースで計算された電力価格を、電力需要曲線を作成する際の基準価格とした。電力基準価格を図 4.2 に示す。

また、2000 年の設備容量として得られた量を、その電源の 2000 年既存設備 F_0 として設定した。これらの既存設備は、2000 年より前に毎年同じ容量だけ建設された設備が残存しているものとし、そのため 2000 年以後に一定の割合で耐用期間経過により停止措置がとられていく。これにより、もし全く新しい設備投資が為されなければ、2015 年の時点において基準需要が満たせなくなる。また、2000 年既存設備容量とその低減の様子を図 4.3 に示す。なお、初期 2000 年時点の予備力は 24%である。

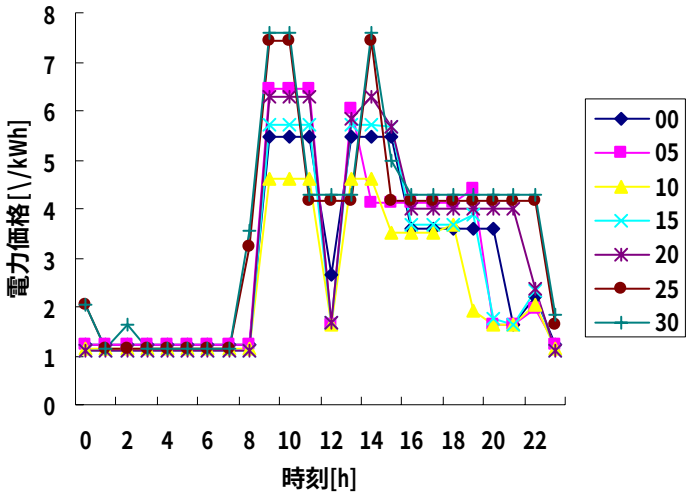


図 4.2　参照ケース　基準電力価格

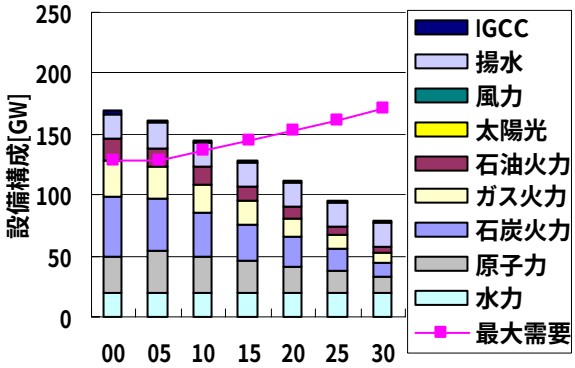


図 4.3　参照ケース　既存設備とその低減

4-2 エージェント学習モデルの性質

作成したエージェント学習モデルの性質を調べるために、強化学習を行うエージェントを 1 体のみ導入した理論解計算が可能な簡単なケース「シングルエージェントケース」を設定し、学習解との比較を行うことにより、強化学習における各種パラメータの設定について考察を行った。

4-2-1 シングルエージェントケースの設定

シングルエージェントケースに導入したエージェントの種類を表 4.2 に示す。ここには、初期設備を保有しているものの全く設備増設を行わない各電源 1 ずつ（初期設備所有・設備増設無）と、増設戦略を学習する唯一のガス火力エージェントが存在する。このエージェントの名前を「ガス火力 1」とする。

表 4.2 予備検討ケース エージェント設定

エージェント名	発電所の種類	2000 年所有設備	設備増設
初期設備所有・設備増設無	水力	$F_{0\,水力}$	無し
	原子力	$F_{0\,原子力}$	無し
	石炭火力	$F_{0\,石炭火力}$	無し
	ガス火力	$F_{0\,ガス火力}$	無し
	石油火力	$F_{0\,石油火力}$	無し
	太陽光	$F_{0\,太陽光}$	無し
	風力	$F_{0\,風力}$	無し
	揚水	$F_{0\,揚水}$	無し
	IGCC	F_{0IGCC}	無し
ガス火力 1	ガス火力	0	学習

4-2-2 動的計画による理論最適解

動的計画法による解法

ガス火力エージェントの設備建設期間は 1 期(5 年)であるため、次状態が現状態と現在の行動にのみで決まるというマルコフ性を満たしている。しかも、次状態への遷移は確定的であり、学習を行うエージェントが 1 体のみであることから、得られる報酬も確定的である。そのため、理論解を動的学習法によって簡単に計算することができる。

最終時点では、増設を決定したどのような設備も考慮期間中に建設が終わらないため、増設をしないことが明らかに最適解である。そのため、最終時点 $T-1$ の行動価値関数 $Q_{T-1,f,a}$ は式(4.10)で定義できる。$r_{t,f}$ は、このケースでガス火力エージェントが、時点 t で設備 f を保有しているときに得られる確定的な純利益を表す。

$$Q_{T-1,f,a} = r_{T-1,f} \tag{4.10}$$

その後、Bellman 方程式に従い、時点を一つずつさかのぼりながら行動価値関数を更新する。

$$Q_{t,f_t,a_t} = r_{t,f_t} + \max_{a'} Q_{t+1,f_{t+1},a'} \tag{4.11}$$

なお、次時点の設備量 f_{t+1} は、現時点の設備量に今時点の増設量を加えたものである（ガス火力発電所の耐用期間は考慮期間よりも長い）から、状態間の関係を示す式(4.12)が成り立つ。

$$f_{t+1} = f_t + a_t \tag{4.12}$$

これにより最適行動価値関数 $Q^*_{t,f,a}$ が定まるので、後は時点の最初からグリーディ方策に従って行動を選択すればよい。なお、行動価値関数と状態価値関数は(4.13)の関係にある。

$$V^*_{t,f} = \max_{a'} Q^*_{t,f,a} \tag{4.13}$$

所有設備量 f と設備増設量 a の離散化が必要であるが、ここではそれぞれの離散化単位を 1[GW]とし、最大値をそれぞれ 60, 40 として計算を行った。上記の動的計画法では、2400 回のエピソード計算が必要であることになる。

動的計画法による結果

シングルエージェントケースを動的計画法により計算して得られた結果について、図 4.4 は、状態価値関数 $V_{t,f}$ を(4.13)式に従って行動価値関数から求めた値を示したものである。

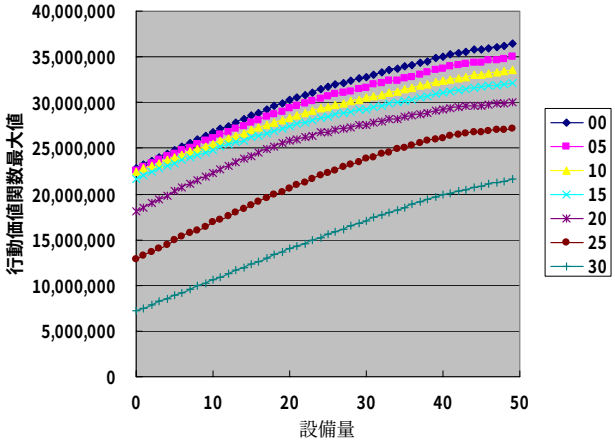


図 4.4　シングルエージェントケース（動的計画法）　行動価値関数最大値（状態価値関数）

横軸はその時点で所有している設備量 f で、1[GW]ごとに離散化をしてある。縦軸は状態価値関数の値 $V_{t,f}$（2000 年価値換算）である。t は 2000 年~2030 年の 5 年間隔 7 時点に離散化している。

また、図 4.5 には、行動価値関数から導かれる各状態における最適方策を示す。所有設備量が小さく時点が大きいときほど多くの設備を増設することが最適であることを示している。これは時点が大きくなれば、図 4.3 に示したように既存設備が減少し、新たに増設した電源へ多くの利益がもたらされるようになるからである。

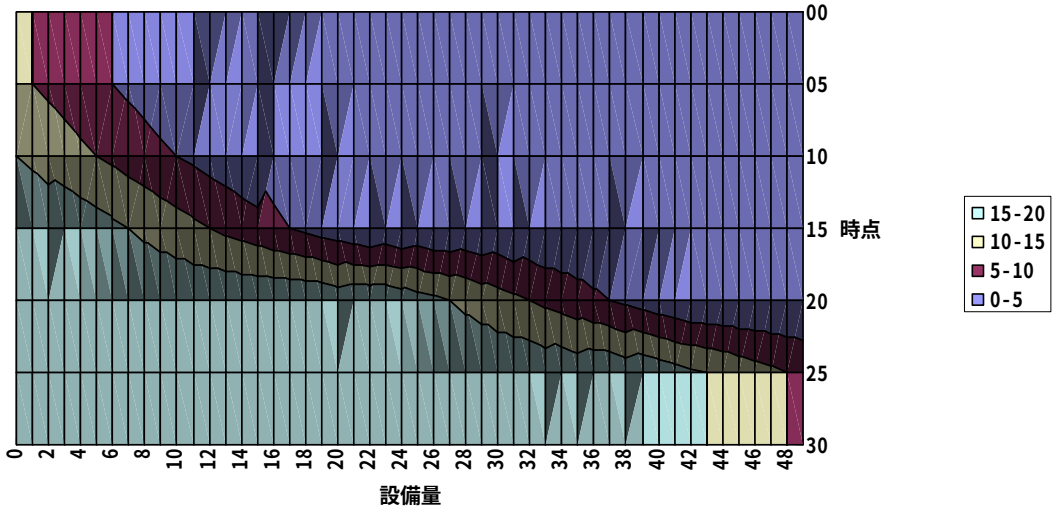


図 4.5　シングルエージェントケース（動的計画法）　最適方策

縦軸は 2000 年~2030 年の 5 年間隔 7 時点に離散化している。横軸はその時点で所有している設備量 f で、1[GW]ごとに離散化をしてある。その状態であれば、どれだけの設備容量[GW]を増設することが最適方策であるかを示している。

4-2-3 強化学習による学習解

強化学習におけるパラメータの役割の考察

3-4-3 で述べた強化学習の手法を用いると、シングルエージェントケースでは、状態の離散化による影響を無視すれば十分に各状態を訪問すれば補助状態価値関数$U_{i,t,f}$が一意に定まる。そうなれば状態価値関数$V_{i,t,f}$も一意に定まり、これより最適行動は自ら定まる。

これより、最適解を得るための課題は、状態価値関数の正しい推定のために十分に各状態を訪問できるか、またそれから導かれる最適行動に行動が収束できるか、の二点である。前者は探査と知識利用の釣り合いをとることが必要な強化学習一般の問題であり、後者は、価値関数とは独立して行動を選択する手法特有の問題である。

このため、方策改善の平均と標準偏差の更新に用いる定数δ_μとδ_σ、また価値関数の更新に用いる学習率αとのバランスや、方策平均と標準偏差の初期値$\mu_{0i,t,f}$と$\sigma_{0i,t,f}$、価値関数の初期値$U_{0i,t,f}$と$V_{0i,t,f}$の値が、結果に影響を与えうると考えられる。

例えば、補助状態価値関数$U_{i,t,f}$の更新は式(3.31)で行っているが、これによれば、補助状態価値関数の初期値がU^0であったとすれば、ある状態では常に売電利益として同じ値Uが得られるとすると、k回後のエピソードでの補助状態価値関数の値U^kは、(4.[19])で表される。

$$U^k = (1-\alpha)^k \times U^0 + U \tag{4.[19]}$$

一般はαは 0.1 ぐらいの値が用いられる**エラー! ブックマークが定義されていません。**。このとき、初期値の影響が 0.1%以下になるには、66 回程度の状態到達が必要である。

表 4.3 に示した 3 種類のパラメータにより、5000 エピソードで強化学習を行わせた。

動的計画ケースにより、初期時点の状態価値関数は 10,000,000[百万円]のオーダの数値になることがわかったので、設定 2 で、それを基準にして各状態でこれよりも 20 倍程度の値になるように設定した。設定 3 では、方策平均と標準偏差の更新率を半分の 0.004 に設定した。

方策平均と標準偏差については、どの設定においても同じ値を用いた。方策平均の初期値については、学習の初めは「事業に参入しない」という仮定をおくとして、0 に近い値に設定し、方策標準偏差は、図 3.9 にも示したように状態空間中の多くの状態に到達することができるよう大きな値設定した。そのため、学習中に負の増設を行うこともありうるが、エピソード数の経過に伴い増加するペナルティを課すことによりこのような行動は次第にはずれていくようにしている。また、学習の初期には、到達状態の前後数状態についても、その状態にも到達したものとして価値関数に同じ更新を施すことで、状態到達の粗さを補完する工夫をした。

表 4.3 予備検討ケース パラメータ設定

	設定 1	設定 2	設定 3	設定 4
方策平均更新率 δ_μ	0.008	0.008	0.004	0.004
方策標準偏差更新率 δ_σ	0.008	0.008	0.004	0.004
学習率（価値関数更新率） α	0.1			
方策平均初期値[GW] $\mu_{0i,t,f}$	0.01			
方策標準偏差初期値[GW] $\sigma_{0i,t,f}$ ※1	40×(t+1)			
価値関数初期値[百万円] $V_{0i,t,f}$ ※2	0	$10{,}000{,}000 \times \sum_{t'=0}^{t} dr_{t'}$	0	$10{,}000{,}000 \times \sum_{t'=0}^{t} dr_{t'}$
補助価値関数初期値[百万円] $U_{0i,t,f}$	0	$10{,}000{,}000 \times dr_t$	0	$10{,}000{,}000 \times dr_t$

※1 40 は各時点における最大増設量である。
※2 dr_t は、3-1-2 項で導出について説明した時点割引率である。ただし、明らかに到達できない状態についての初期値は 0 とし、価値関数の更新の対象外ともしている。

パラメータの差異が価値関数へ与える影響

価値関数初期値の違いによる、5000 エピソード後の価値関数 $V_{t,f}$ の違いを図 4.6 に示す。なお、2000 年所有設備は 0、各時点における最大増設量は 40 としているため、明らかに到達できない状態についてはその価値関数の初期値は 0 とし、更新の対象外にしている。

まず、価値関数初期値の違いに注目するために、設定 1 と設定 2、設定 3 と設定 4 の結果を比較する。
本来は図 4.4 に最適価値関数を示したように、時点 00 から時点 15 はほぼ同じ曲線に乗るはずであるが、時点が小さく設備量が大きい状態で、状態価値関数初期値の違いに起因する偏差が見られる。すなわち初期値が 0 であったパラメータ設定 1 や設定 3 では過小推定、初期値が大きかったパラメータ設定 2 や設定 4 では過大推定を行っている。

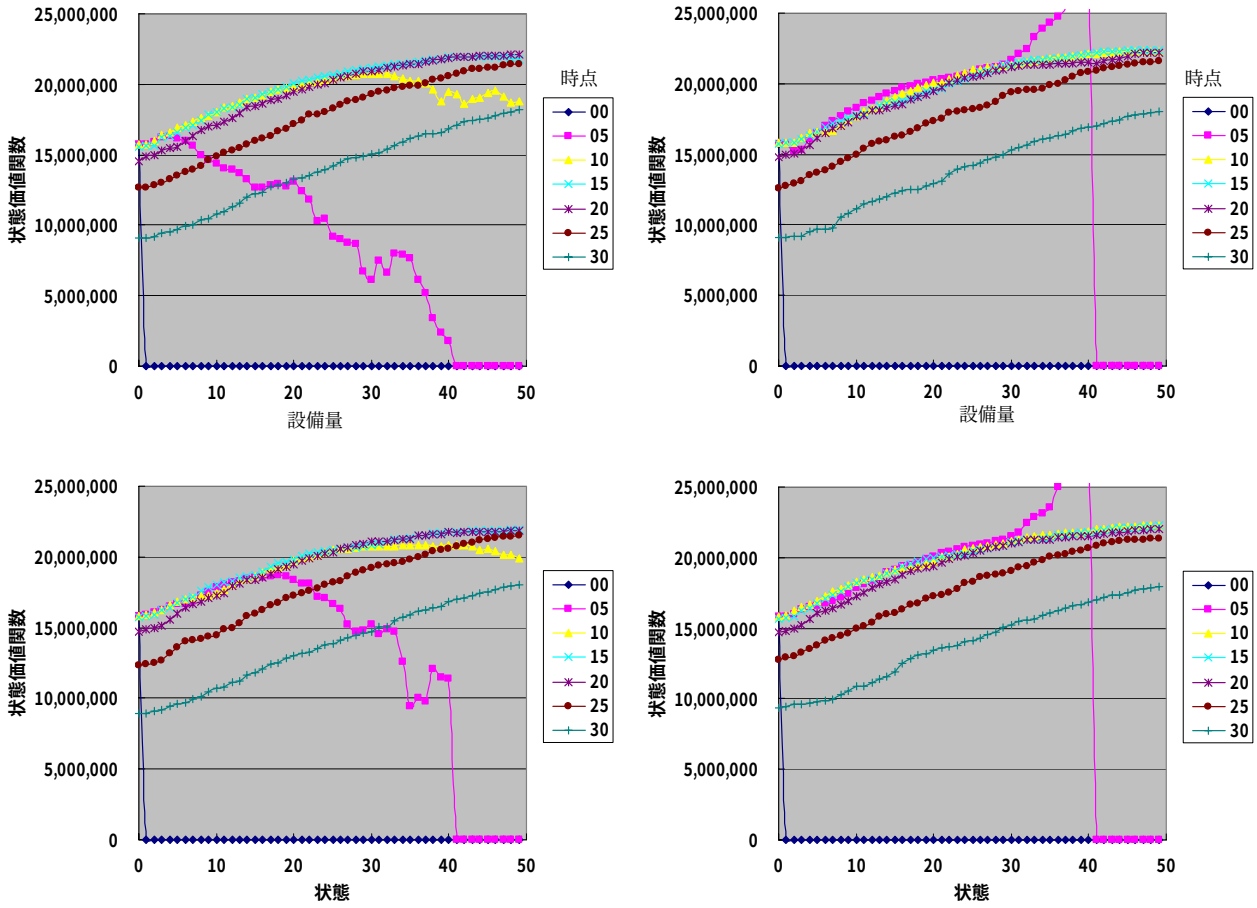


図 4.6　シングルエージェントケース 「ガス火力 1」の状態価値関数
価値関数初期値･方策更新率の違いによる比較
（パラメータ設定 1（左上）、設定 2（右上）、
設定 3（左下）、設定 4（右下）

横軸はその時点で所有している設備量 f で、1[GW]ごとに離散化をしてある。
t は 2000 年~2030 年の 5 年間隔 7 時点に離散化している。
縦軸は状態価値関数の値（2000 年価値換算）である。

各状態に十分な回数到達していれば、状態価値関数初期値の影響は小さくなるはずである。各状態への到達回数の比較を図 4.7 に示す。状態価値関数の初期値が小さい設定 1 においては、時点 t が小さく設備 f が大きい状態への到達回数が十分ではないことがわかる。逆に、状態価値関数の初期値が大きい設定 2 においては、5000 エピソード各状態に最低 100 回程度は到達している。(4.[19])に従えば、100 回の到達では初期値の影響は 0.003%程度と十分小さい。これは、楽観的な初期値（オプティミスティック初期値）を与えることにより、到達回数の小さい状態に対する状態価値の値が大きくなるため、エージェントはこの行動を取りやすくなるためである。

これより、設定 1 で得られた状態価値関数よりも、価値関数初期値を高く設定した設定 2 で得られた状態価値関数の方が適切な値に近いということが言える。

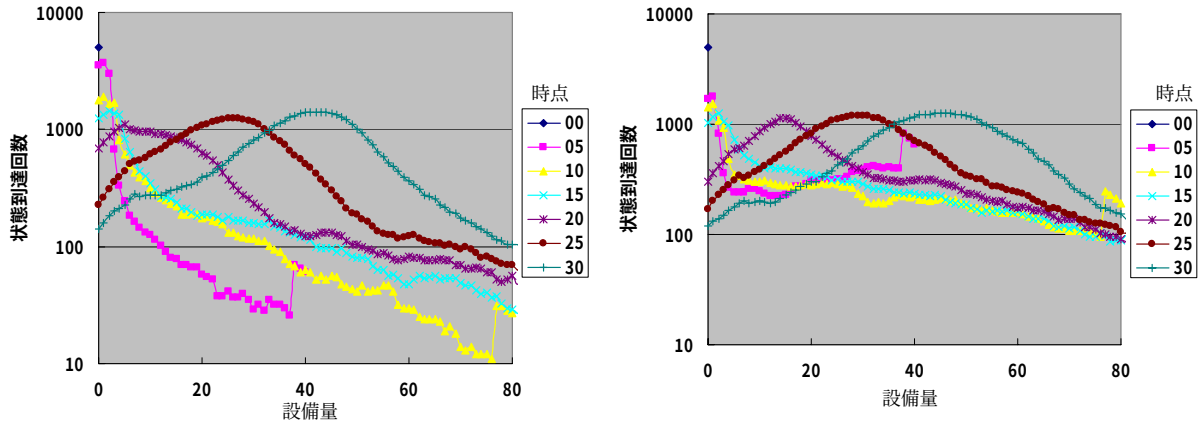


図 4.7　シングルエージェントケース　「ガス火力 1」の状態達成回数
価値関数初期値の違いによる比較（パラメータ設定 1（左）、設定 2（右））

次に、図 4.6 において、価値関数初期値が同じで方策更新率が異なる設定 1 と設定 3 の違いに注目すると、設定 3 の価値関数のほうが、05 時点や 10 時点においても比較的広い範囲で適切な推定が行われていることがわかる。これは、方策更新率が大きいと学習から得られた知識が示す最適行動を取りやすくなるため、それ以外の行動の探索機会が減少し、状態価値関数を適切に推定できなくなることを示している。ただし、設定 2 と設定 4 の違いがほとんど見られないことから、方策更新率の違いよりも価値関数初期値の方が結果へ与える影響が大きいことが分かる。

また、こうして強化学習によって得られた価値関数を、動的計画法によって得られた最適価値関数と比較したときの推定精度を図 4.8 に示す。ここでは価値関数初期値が 0 で方策更新率が小さめの設定 2 を例にした。時点 00、設備量 0 の状態に対する推定精度は 70%程度であり、これは強化学習によって得られる総利益が最適値の 70%程度であることを示している。

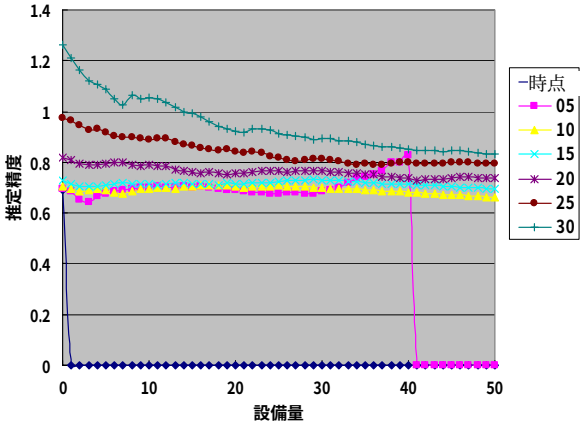


図 4.8　シングルエージェントケース　「ガス火力 1」の状態価値関数推定精度

各状態において、強化学習設定 2（図 4.6（右上））における価値関数の値の、
動的計画法による最適価値関数の値（図 4.4）に対する比を示している。

また、設定 2 の結果として得られた方策平均 $\mu_{0i,t,f}$ を図 4.9 に示す。所有設備量が小さく時点が大きいときほど多くの設備を増設することが最適であるという傾向は強化学習による最適方策（図 4.5）に似て

いるが、それよりも深く鋭い山を形成している。また、価値関数の過大評価による影響が時点 05 の部分に見られる。

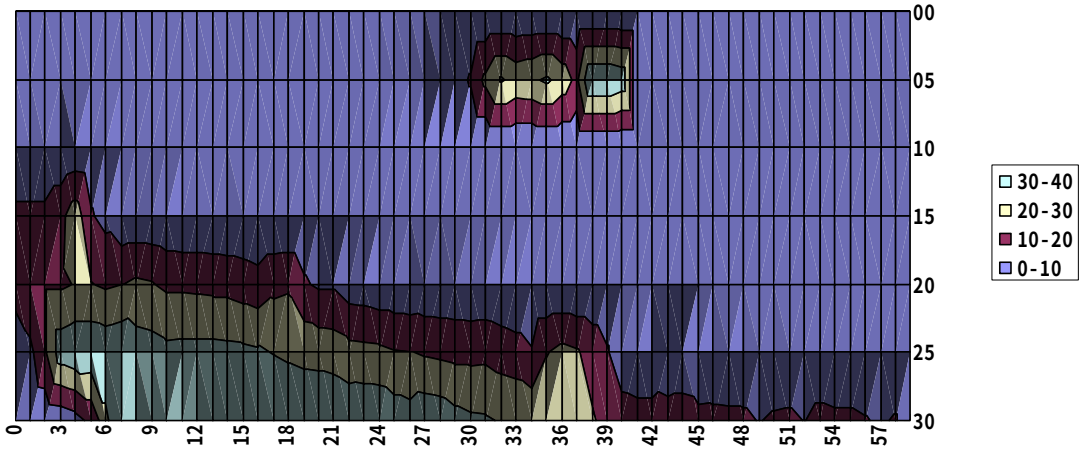


図 4.9　シングルエージェントケース 「ガス火力 1」の方策平均値（パラメータ設定 2）

標準パラメータ設定

状態価値関数の初期値を十分大きくとったり、方策更新率を小さくとったりすることにより、より適切な価値関数が推定できることがわかった。

これらのパラメータはマルチエージェント学習という非定常状態における学習にどれだけ有効であるかについては別の議論が必要ではあるが、今後は特に断らない場合、表 4.4 に示したパラメータ設定を標準として用いることにする。マルチエージェントにした場合の解の収束を促進するため、方策更新率は大きめの値を採用し、その代わりに価値関数の初期値は十分大きくとることとした。また、計算は 20000 エピソードを行った。

表 4.4　ケース標準パラメータ設定

方策平均更新率 δ_μ	0.008
方策標準偏差更新率 δ_σ	0.008
学習率（価値関数更新率） α	0.1
方策平均初期値[GW] $\mu_{0i,t,f}$	0.01
方策標準偏差初期値[GW] $\sigma_{0i,t,f}$	$40\times(t+1)$
価値関数初期値[百万円] $V_{0i,t,f}$	$10{,}000{,}000\times\sum_{t'=0}^{t} dr_{t'}$
補助価値関数初期値[百万円] $U_{0i,t,f}$	$10{,}000{,}000\times dr_t$
エピソード数	20000

第5章 マルチエージェントシステムとしての電力需給

本章では、発電事業者エージェント数の変化や異質エージェントである政府機関エージェントの導入による結果への影響の確認をおこなう。また、より現実的な問題をモデル化し、多数の利害関係者による電力需給の様子を定性的に観察する。

5-1 エージェントの競合の分析

4-2-1 では、戦略的に行動するのが 1 体のガス火力エージェントのみであり、このエージェントによる独占状態になっていた。ここではその数を増やすことにより、どのようにエージェント間の競争が起こるのかを観察する。

5-1-1 同種エージェント競合ケースの設定

同種エージェント競合ケースに導入したエージェントの種類を表 5.1 に示す。ここには、初期設備を保有しているものの全く設備増設を行わない各電源 1 ずつ（初期設備所有･設備増設無）と、増設戦略を学習するガス火力エージェントが存在する。このガス火力エージェントの数として 1,5,10,15,20 の 5 通りで計算した。強化学習におけるパラメータは表 4.4 に示した標準のものを用いた。

表 5.1 同種エージェント競合ケース
（ガス火力 1、ガス火力 5、ガス火力 10、ガス火力 15、ガス火力 20）

エージェント名	発電所の種類	2000 年所有設備	設備増設
初期設備所有･設備増設無	表 4.2 に示した「初期設備所有･設備増設無」設定と同じ		
ガス火力 1～1.5,10,15, 20	ガス火力	0	学習

5-1-2 同種エージェント競合ケースの結果

図 5.1 に、ガス火力エージェント数と、各ガス火力エージェントの全期間総利益の合計値との関係を示す。エージェント数が増えるに従い、総利益合計は減少している。

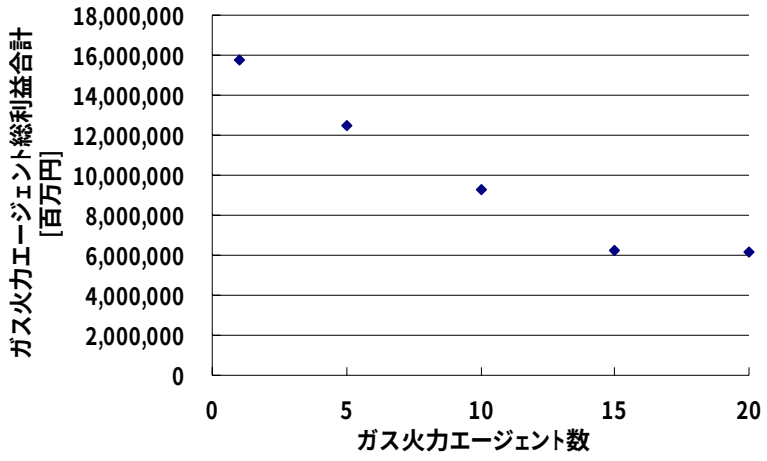


図 5.1 同種エージェントケース ガス火力エージェント総利益合計

図 5.2 には、ガス火力エージェント数と、電力供給に要する総コストを示す。電力供給コストの上限（ガス火力エージェントが全く設備を建設しなかった場合）、下限（ガス火力エージェントが電力供給コストが最小となるように設備を建設した場合）も示してある。

エージェント数が増加すれば、市場は競争的になり電力供給コストは低下するが、電力供給コストの最小値には及んでおらず、エージェント数 10 程度以上からほとんど横這いになっている。

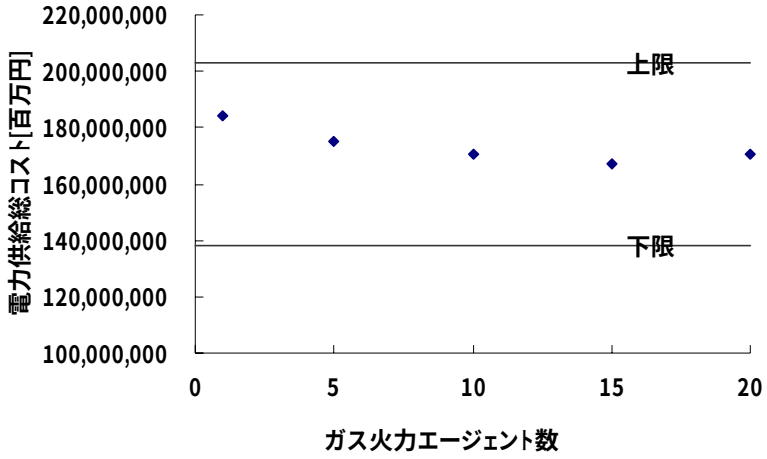


図 5.2　同種エージェントケース　電力供給コスト

図 5.3 に、同種エージェント競合ケースにおける戦略的ガス火力エージェントの、設備増設量合計の推移を示す。

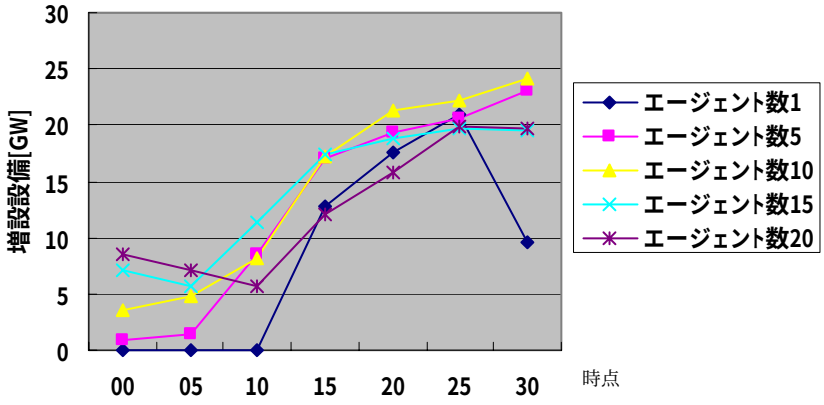


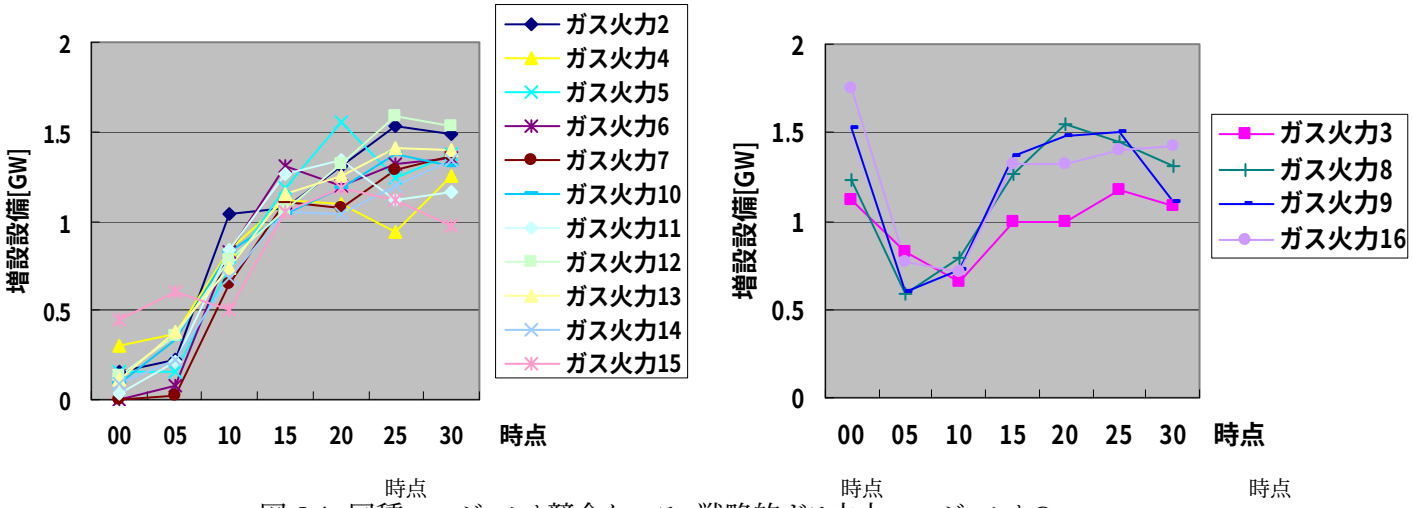
図 5.3　同種エージェント競合ケース　ガス火力増設設備量合計
（エージェント数 1、5、10、15）

エージェント数 5、10、15 の 3 設定を比較すれば、最終時点における設備量はほぼ変わらない。このため、エージェント数が 5 程度ならば寡占状態を脱してある程度競争原理が働いているのではないかと考えられる。

エージェント数が多いほど、増設を行う時点が均等化されていることがわかる。これは競合するエージェントが多い中で多くの利益を得るために各自の戦略が分散され、結果的に各時点の増設量が一定に近づいているのだと考察することができる。

例えば、エージェント数 15 のとき、それぞれの設備増設量から、異なる 2 つの特徴的行動をとるタイプのエージェントが存在している。ひとつは時点後期（25 年、30 年）に重点的に設備建設を行うエージェント、もうひとつは時点中期（15 年、20 年）に重点的に設備建設を行うエージェントであり、15 のエージェン

ト中それぞれ 5 体、4 体のエージェントが属している。また、それ以外の行動タイプのエージェントも存在する。



時点　　時点　　時点

図 5.4　同種エージェント競合ケース　戦略的ガス火力エージェントの
増設設備戦略のタイプ分け（エージェント数 15）

以後、意図しない独占状態による結果の歪みを避けるため、特に断らない限り同種のエージェント数は 5 とする。

5-2 政府機関エージェントの影響

このモデルでは、発電事業者エージェントとはまったく質の異なるエージェントである政府機関エージェントを導入している。ここでは、政府機関エージェントは発電事業者エージェントの行動に影響を及ぼしうるのか、またその影響はどのようなものかについて調べた。

5-2-1 設備確保エージェント導入ケースの設定

まず、設備確保エージェントの影響を知るため、表 5.2 に示すように戦略的電力事業者エージェントとしてガス火力 1 体、設備確保エージェントが 1 体というもっとも単純なケースを計算した。次に、ガス火力エージェントを 5 体にした表 5.3 のケースを計算した。それぞれ強化学習におけるパラメータは表 4.4 に示した標準のものを用いた。

表 5.2　設備確保エージェント導入ケース（ガス 1）

エージェント名	発電所の種類	2000 年所有設備	設備増設
初期設備所有・設備増設無	表 4.2 に示した「初期設備所有・設備増設無」設定と同じ		
ガス火力 1	ガス火力	0	学習
設備確保	—	—	—

表 5.3　設備確保エージェント導入ケース（ガス 5）

エージェント名	発電所の種類	2000 年所有設備	設備増設
初期設備所有・設備増設無	表 4.2 に示した「初期設備所有・設備増設無」設定と同じ		
ガス火力 1~5	ガス火力	0	学習
設備確保	—	—	—

5-2-2 設備確保エージェント導入ケースの結果

設備確保エージェントが発電事業者エージェントへ与える影響

表 5.2 に示した戦略的発電事業者エージェントがガス火力 1 体のみのケースで、設備確保エージェントを導入したときと導入しないときの状態価値関数の比較を図 5.5 に示す。時点 05 の設備量 30 以上では、状態到達回数が少ないことから価値関数の初期値の影響が残り、共に状態価値関数をやや過大推定している。

これらの比較すると、設備確保エージェントを導入することにより、状態価値関数が上方へ全体的にシフトしていることがわかる。その差分を図 5.6 に示す。

初期状態である時点 00、設備量 0 の状態では、この差分は 4,000,000[百万円]となっている。すなわち、最適行動により、政府機関エージェントからこれだけの補助金支給が見込めるという値である。

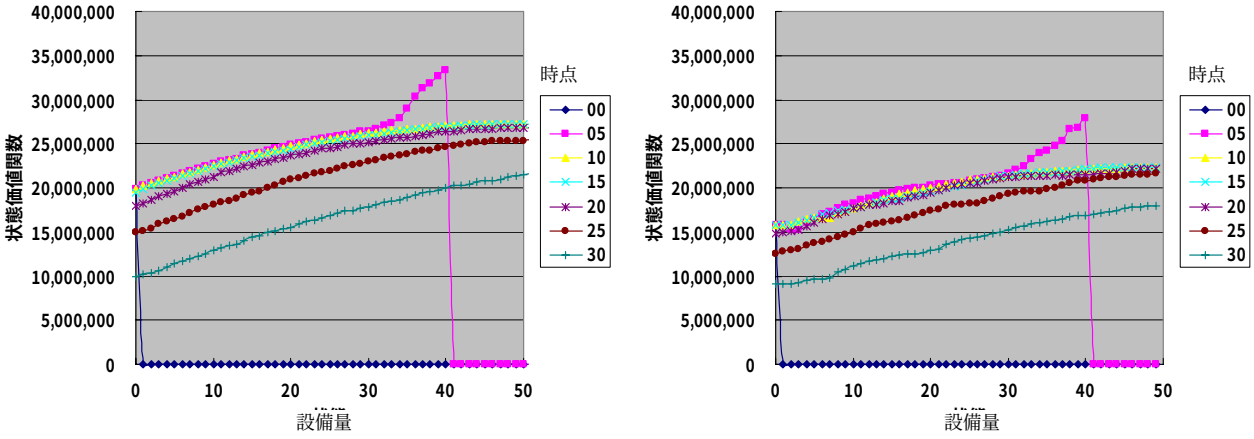


図 5.5 設備確保エージェント導入ケース（ガス 1）（左）と
設備確保エージェントの無いケース（右）の発電事業者エージェントの状態価値関数の比較

設備確保エージェント無しケースは、図 4.6（右）を再掲した。ただし、縦軸の最大値が異なる。

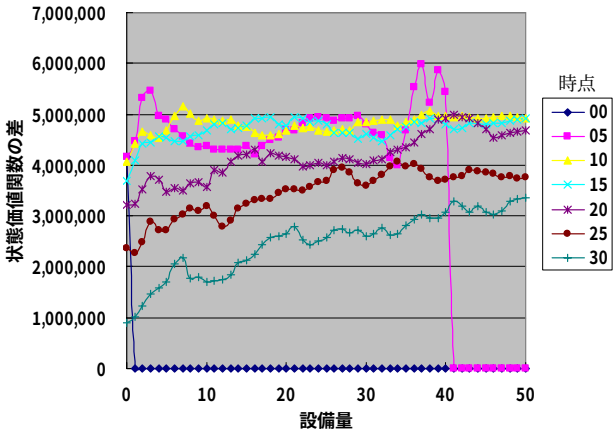


図 5.6 設備確保エージェント導入ケース（ガス 1）と
設備確保エージェントの無いケース（シングルエージェントケース）の状態価値関数の差分

発電事業者エージェント数の違いによる結果の相違

補助金による価値関数のシフトの結果として設備がより多く建設されるようになったかというとそうではない。図 5.7 はこの設備確保エージェント導入ケースと、同じ条件で設備確保エージェントの無いケース（4-2 節で取り扱ったシングルエージェントケース）の設備構成を比較しているが、まったく改善は見られない。

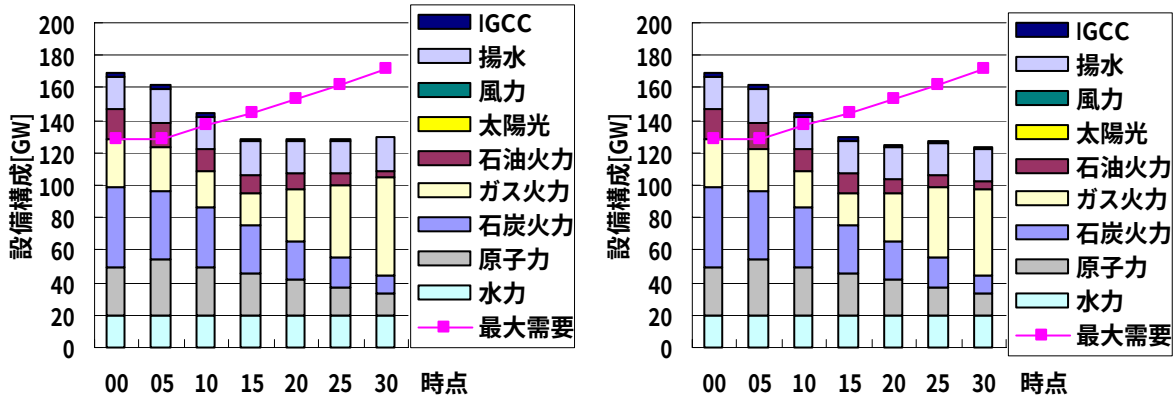


図 5.7 （左）設備確保エージェント導入ケース（ガス 1） と
（右）設備確保エージェントの無いケース（シングルエージェントケース）の設備構成

一方で、表 5.3 に示したガス火力 5 体のケースを計算し、設備増設エージェントの無いケース（5-1 節で取り扱った同種エージェント競合ケースガス火力 5 体）と比較したところ、十分では無いものの、図 5.8 に示すように明らかな設備増設量の増加が見られた。

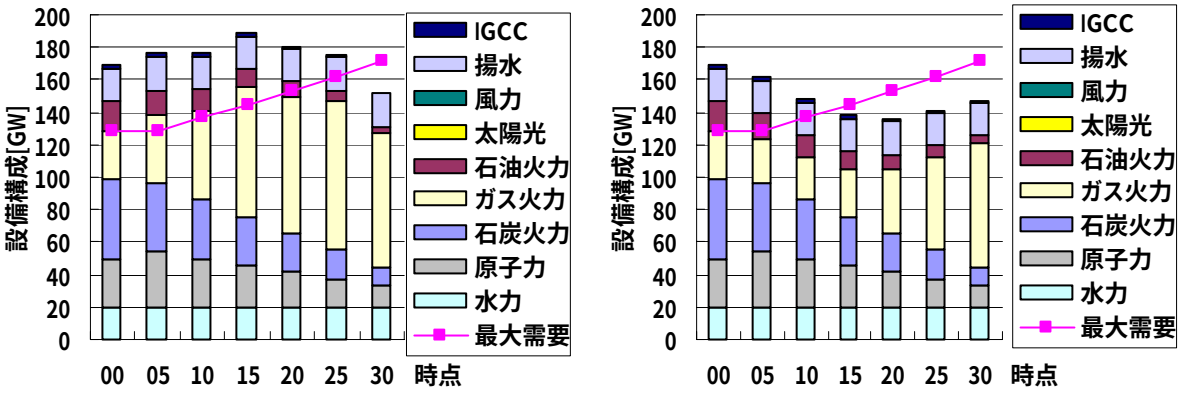


図 5.8 （左）設備確保エージェント導入ケース（ガス 5）と
（右）設備確保エージェントの無いケース（同種エージェント競合ケース（ガス 5））の設備構成

これより、このような枠組みにおいて政府機関による補助政策が機能するためには、市場が独占的でないことが条件であるという仮説が考えられる。以下では、この仮説をゲーム理論を用いて考察する。

ゲーム理論による政府機関の役割の考察

この政府機関という主体が発電事業者の行動にどのような影響を与え得るのかを、ゲーム理論の枠組みに沿って考察する。なお、これは新型旅客機の開発競争ゲーム[29]を参考にした。

ここでは、2 体のガス火力発電事業者と、設備確保担当政府機関によるゲームを考える。
まず、政府機関の介入が無い場合を考える。ガス火力発電事業者は、費用をかけてガス火力発電所を 1 単位建設する。建設費用は 5、どちらか 1 事業者のみ建設した場合は発電により 6 の利益が得られ、純利益は 1 である。しかし、両社が建設した場合は電力価格が下落し、それぞれ利益は 6 よりも小さい 4 しか得られず差し引き－1 の損失が発生するとする。

このゲームは表 5.4 で表現することができる。

表 5.4　2 体のガス火力発電事業者の設備建設ゲーム

（発電事業者 1、発電事業者 2）の利得

発電事業者 1 ＼ 発電事業者 2	建設する	建設しない
建設する	−1,−1	0, 1
建設しない	1, 0	0, 0

ガス火力発電事業者は双方がミニマックス原理に基づいて行動すると仮定する。そのマックスミニ均衡は（建設しない、建設しない）となる。また、ナッシュ均衡は（建設する、建設しない）、（建設しない、建設する）となる。

一方、設備確保担当政府機関としては、両方の事業者に発電設備を建設してもらいたい。この政府機関の利得は、両社が建設を行わなかったときは 0、一社が建設を行ったときは 5、2 社が行ったときの利得は 10 とする。設備確保担当政府機関は、建設を行う事業者には 1 単位あたり S の補助金を与えることを検討している。S の補助金につき、政府機関は κS（$\kappa < 1$）の損失を被る。

2 体のガス火力発電事業者の設備建設と、設備確保担当政府機関の補助金政策ゲームは表 5.5 で表現することができる。

表 5.5　2 体のガス火力発電事業者の設備建設と設備確保担当政府の補助金政策ゲーム

（発電事業者 1、発電事業者 2、政府機関）の利得

政府機関＝補助金なし		
発電事業者 1 ＼ 発電事業者 2	建設する	建設しない
建設する	−1,−1, 10	0, 1, 5
建設しない	1, 0, 5	0, 0, 0
政府機関＝補助金あり		
発電事業者 1 ＼ 発電事業者 2	建設する	建設しない
建設する	$-1+S$, $-1+S$, $10-2\kappa S$	0, $1+S$, $5-\kappa S$
建設しない	$1+S$, 0, $5-\kappa S$	0, 0, 0

政府が第一の手番であり、2 体の発電事業者は同時の手番をとるものとする。$S>1$ ならば、政府機関が補助金政策を行う場合、2 体の発電事業者にとっては「建設する」が支配戦略となり、支配戦略均衡が（建設する、建設する）となるため、政府機関は $10-2\kappa S$ の利得を得る。補助金政策を行わない場合、2 体の発電事業者のナッシュ均衡は（建設する、建設しない）と（建設しない、建設する）、マックスミニ均衡は（建設しない、建設しない）であったから、政府機関の利得は高々5 である。これより、政府機関にとっては $10-2\kappa S>5$ ならば「補助金あり」戦略が妥当である。特に κ が十分小さい、すなわち設備確保を行うということ自体への優先度が高い場合は、この式は成り立つと考えてよい。

ただし、政府が与えようとする補助金の額が十分でない場合、ここでは $S<1$ であれば、政府機関が補助金政策を行っても 2 体の発電事業者の均衡戦略は（建設する、建設する）にならない。また、補助金の額が多すぎる場合、ここでは $S>5/2\kappa$ であれば、政府機関は「補助金なし」戦略が妥当になる。

一方で、1 体のガス火力発電事業者が、ガス火力発電設備を 0、1、2 単位のいずれを建設するかという 3 種類の戦略を持っているとする。このガス火力発電事業者と設備確保担当政府機関のゲームは表 5.6 で表現される。

表 5.6　1 体のガス火力発電事業者の設備建設と設備確保担当政府の補助金政策ゲーム

(発電事業者、政府) の利得

発電事業者 / 政府機関	2 単位建設する	1 単位建設する	建設しない
補助金なし	−2,10	1, 5	0, 0
補助金あり	$-2+2S$, $10-2\kappa S$	$1+S$, $5-\kappa S$	0, 0

政府機関にとっては「補助金なし」政策が支配戦略であり、手番の同時･非同時に関わらずこの戦略を選択するだろう。発電事業者にとっては $S<3$ なら「1 単位建設する」が弱支配戦略であるし、いずれにせよ政府機関が「補助金なし」政策に対しては「1 単位建設する」が最適反応戦略でもある。従って、このゲームの均衡は (1 単位建設する、補助金なし) となり、政府機関が意図した量の設備建設は為されない。

5-2-2 項で示したシミュレーションの結果も、これと同じことを表しているということができる。すなわち、電力事業者に対する政府機関による補助政策が機能するためには、1 つの電力事業者が市場を独占していないことが条件になることがわかる。

5-3 電源間の競合

石炭火力･IGCC･ガス火力･石油火力の化石燃料火力発電間で、自由化後に電源間競合が起こるのかを観察する。

5-3-1 電源間競合ケースの設定

異種エージェント競合ケースに導入したエージェントの種類を表 5.7 に示す。ここには、初期設備を保有しているものの全く設備増設を行わない各電源 1 ずつ (初期設備所有･設備増設無) と、増設戦略を学習する複数種のエージェントが存在する。

強化学習におけるパラメータは表 4.4 に示した標準のものを用いた。

表 5.7　電源間競合ケース (石炭 5:ガス 5:石油 5:IGCC5)

エージェント名	発電所の種類	2000 年所有設備	設備増設
初期設備所有･設備増設無	表 4.2 に示した「初期設備所有･設備増設無」設定と同じ		
石炭火力 1~5	石炭火力	0	学習
ガス火力 1~5	ガス火力	0	学習
石油火力 1~5	石油火力	0	学習
IGCC1~5	IGCC	0	学習

5-3-2 電源競合ケースの結果

電源間競合ケース (石炭 5:ガス 5:石油 5:IGCC5) ケースと、5-1 節で取り上げた同種エージェント競合ケース (ガス火力 5) の、電源構成を図 5.9 に比較する。設備量合計の傾向はほぼ同じで、それぞれ基準需要に対して基準需要に対して 10%程度の省エネルギーが行われている。ガス火力エージェントは、異種の電源が存在しないときに比較してシェアを奪われていることがわかる。

また、図 5.10 は、揚水を除いた電源間競合ケースの設備構成割合を、4-1 節で取り上げた最適電源計画による設備構成割合と比較している。最適電源計画では二酸化炭素排出原単位制約が課されているため、石炭火力は IGCC、石油火力はガス火力に置き換わってはいるが、水力と原子力からなるベース、石炭火力と IGCC からなるミドル、ガス火力と石油火力からなるピークの、各電源の割合はほとんど同じである。競争状態においても、これらの役割の異なる電源間での代替は行われにくいことを示している。

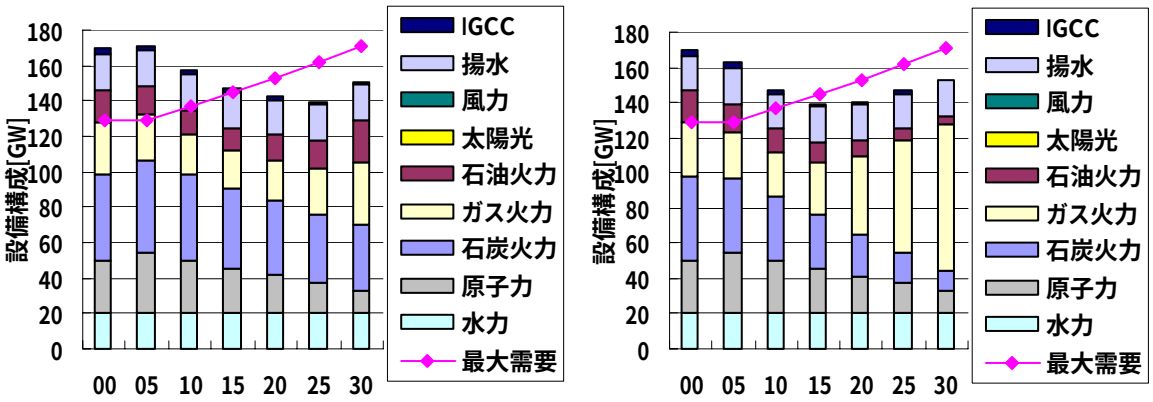


図 5.9　電源間競合ケース（石炭 5：ガス 5：石油 5：IGCC5（左））と
同種エージェント競合ケース（ガス 5（右））の設備構成の比較

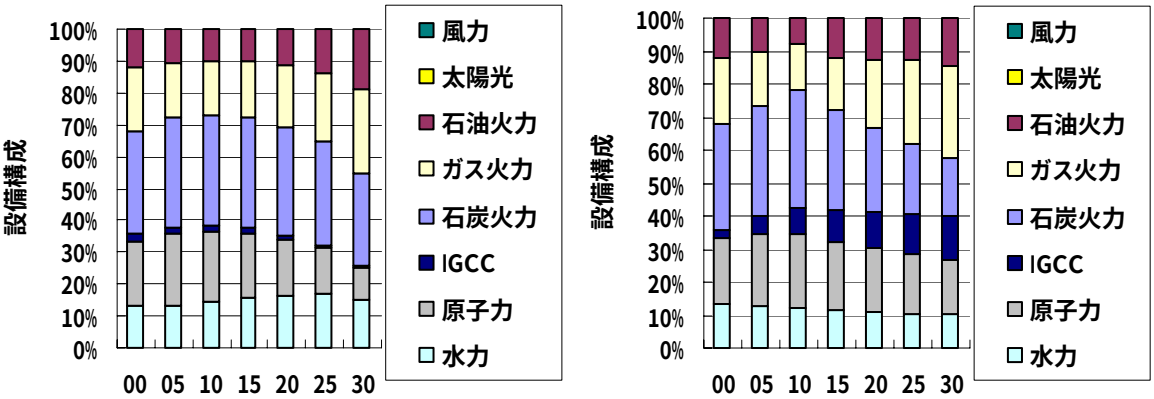


図 5.10　電源間競合ケース（石炭 5：ガス 5：石油 5：IGCC5）（左）と
最適電源計画（右）の設備構成（揚水を除く）

5-4 電力会社と新規参入発電事業者の競合

電力自由化後でも各電源（発電所）が独立して電力供給を行うようになるのではなく、電力会社のような複数種の電源を所有する主体もやはり存在する。このような複数種の電源を持つエージェント集団の戦略と、独立しているエージェント集団の戦略を比較した。

5-4-1 電力会社＋新規参入ケースの設定

設定した電力会社＋新規参入ケースについて表 5.8 に示す。2000 年既設電源はすべて電力会社所属のエージェントが所有しているものとし、新規参入事業者者として初期時点で所有設備を持たない石炭火力・IGCC・ガス火力・石油火力の化石燃料火力発電についてそれぞれ戦略的エージェントを導入する。ただし、電力会社所属のエージェントも、増設を行うのは石炭火力・IGCC・ガス火力・石油火力の 4 種の電源のみであるとする。

電力会社のような一社で複数の電源を持って連結決算を行う主体は、3-4-4 項で述べたように所属エージェントの利益合計を最大化するエージェント集団として表現されている。

強化学習におけるパラメータは表 4.4 に示した標準のものを用いた。

表 5.8　電力会社＋新規参入ケース

エージェント名	発電所の種類	2000 年所有設備	設備増設
電力会社	水力	$F_{0\,水力}$	無し
	原子力	$F_{0\,原子力}$	無し
	石炭火力	$F_{0\,石炭火力}$	学習
	ガス火力	$F_{0\,ガス火力}$	学習
	石油火力	$F_{0\,石油火力}$	学習
	太陽光	$F_{0\,太陽光}$	無し
	風力	$F_{0\,風力}$	無し
	揚水	$F_{0\,揚水}$	無し
	IGCC	F_{0IGCC}	学習
石炭火力 1~5	石炭火力	0	学習
ガス火力 1~5	ガス火力	0	学習
石油火力 1~5	石油火力	0	学習
IGCC1~5	IGCC	0	学習

5-4-2 電力会社＋新規参入ケースの結果

電力会社＋新規参入ケース（石炭 5：ガス 5：石油 5：IGCC5）と、5-1 節で取り上げた同種エージェント電源間競合ケース（石炭 5：ガス 5：石油 5：IGCC5）の、電源構成を図 5.11 に比較する。設備量合計の傾向や構成には大差は見られない。

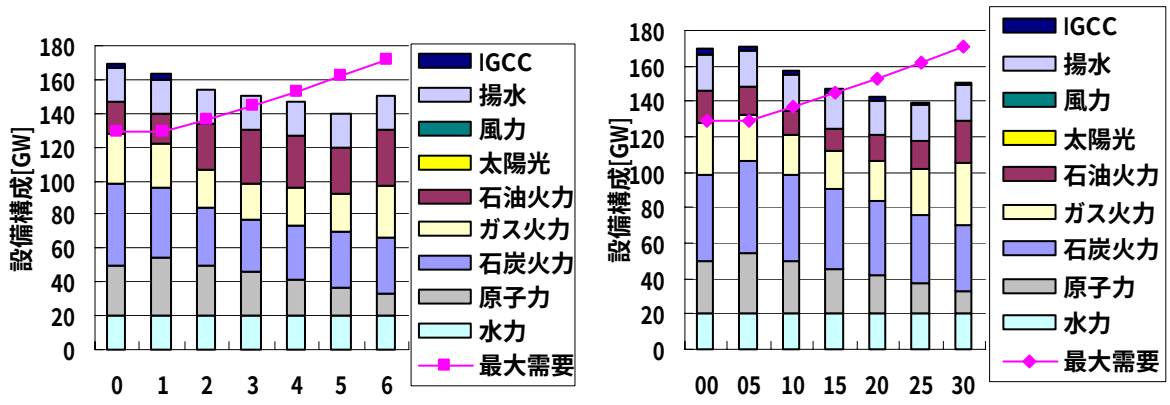


図 5.11　電力会社＋新規参入ケース（石炭 5：ガス 5：石油 5：IGCC5）の電源構成と
電源間競合ケース（石炭 5：ガス 5：石油 5：IGCC5）の電源構成の比較

設備増設を行っているエージェントの内訳を図 5.12 に示す。ほとんどが新規発電事業者に占められていることがわかる。ただし、これは同時学習を行うエージェント集団が適切に学習できていないために不利になっているという可能性がある。今後の検討事項である。

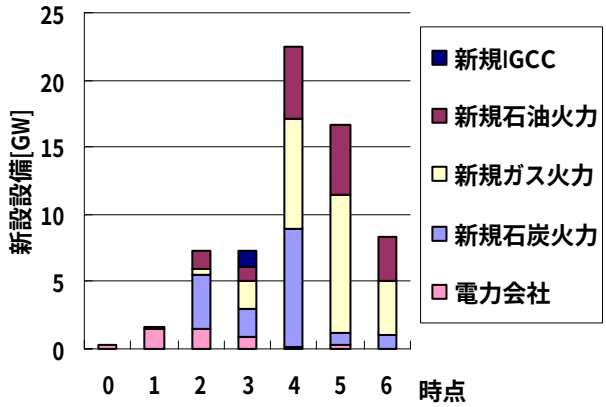


図 5.12　電力会社＋新規参入ケース（石炭 5：ガス 5：石油 5：IGCC5）の増設設備

第6章 結論

6-1 本研究の結論

本研究では、一地域の電力需給を対象とし、多時点に渡る動的設備投資戦略に焦点を当てたエネルギーモデルを構築した。このモデルでは電力自由化に伴う電力需給における多数の利害関係者を直接モデル化しており、発電事業者間の競合や、それへの政府機関の統治を表現した。

マルチエージェント強化学習によるエネルギー需給モデルの構築についての知見

マルチエージェント強化学習という手法を用いた、設備投資を伴う動学的なエネルギー需給モデルの構築ノウハウとして、以下のような知見を得た。今後同様のモデルを作成するときには留意すべき点である。

- 強化学習におけるパラメータ設定の重要性を確認した。
  特に動学的エネルギーモデルに適用する際には、初期時点での行動が後時点の状態到達を制限してしまうことがあるため、これを回避するための方策の初期値や更新の設定が必要である。また、オプティミスティックな初期値の設定もある程度有効である。
- 設備増設を行うエージェントが所有する情報として、現在の時点と所有設備量という最低限の状態を与えたが、これだけでもある程度に妥当な解が得られた。特に、この情報内では、行動により次にどの状態に達するかがエージェントにとって自明であり、行動価値関数を定義しなくても性能のよい学習を行うことができた。
- 発電事業者エージェント間にまたがる制約を間接的に課す方法として、政府機関という異質エージェントを導入した。これによる仮想的な補助率・税率の設定により、目的とする制約をある程度達成できることが確かめられた。
- 一方、設備増設を行うエージェントをそのまま集団化し、総報酬の合計値を最大化させるような単純なエージェント集団を設けたところ、同時学習問題が顕在化して十分な学習を行うことは難しかった。

電力自由化市場での競合についての知見

作成したモデルをもちいて、電源間の競合を設備投資という観点から分析し、以下のような知見を得た。

- 同種の事業者が 5〜15 程度参入すれば、十分な競争原理がはたらき、多数の事業者が参入した場合と類似の結果が得られることがわかった。
- 政府機関が設備確保という目的の元、補助金をもって発電事業者の設備増設戦略に介入しようとするとき、市場が独占的であればその政策は機能しないことが確かめられた。
- 発電事業者間の競争が行われても、ベース・ミドル・ピーク需要間での電源の代替は起こりにくいことが示された。

6-2 今後の課題

課題となる点を以下に列挙する。

設備確保以外の政府機関エージェントの導入

二酸化炭素排出原単位を目標値以下に抑える、燃料を分散してリスクに備えるなど、単一の発電事業者だけでは達成できない目標は設備確保以外にもある。このような制約に対応する政府機関エージェントの定式化は行ったが、シミュレーションを行い考察するには至らなかった。

燃料価格変動の影響の分析

燃料価格変動をモデル化したが、シミュレーションを行い考察するには至らなかった。燃料価格の変動が起これば、変動の大きい石油火力は他の電源に比較して不利になり、設備投資を抑えるのではないかと考えられる。

エージェントに与える情報

設備増設を行うエージェントが所有する情報として現在の時点と所有設備量を与え、これだけでもある程度に妥当な解が得られたが、電力市場におけるそのエージェントの状態を端的に表すものとして電力価格、発電量といった報酬そのものを用いることができるかもしれない。本来の強化学習においては、報酬は状態遷移を経て最終状態で与えられるもので、その報酬をそこに至る状態に適切に分配するのが価値関数の推定であるが、このエネルギーモデルの枠組みでは報酬を各状態における利益として明らかに設定することができるためである。

エージェント集団の同時学習問題の解決

設備増設を行うエージェントをそのまま集団化し、総報酬の合計値を最大化させるような単純なエージェント集団を設けたところ、同時学習問題が顕在化して十分な学習を行うことは難しかった。総報酬の合計値とそのエージェント自体への個体報酬を重み付けして足し合わせたものを報酬として与えるなどの工夫により、より適切な学習ができるようになるかもしれない。

強化学習手法の改善

今回用いた手法では、エージェント数が 1 体の場合の状態価値関数を精度約 70%で推定することができたが、何らかのパラメータの改良により、この推定精度を著しく向上させることは可能であると思われる。

また、価値関数の更新の際に次時点の状態空間をスイープする方法は、本来のアクタークリティック手法よりも価値関数推定能力は高いが、計算回数が多くなる。これはエージェントの増加に対し、計算効率を著しく下げる可能性があるため、価値関数がある程度推定できればアクタークリティック法に切り替えて学習を行うなどの方法により計算速度の向上が見込まれる。

電力市場決済の工夫

発電事業者の設備増設戦略に焦点を当てたため、電力市場への入札は燃料原価で行うものと仮定し、需給バランスの限界価格を電力価格とした。しかし、そのため、ピーク電源はもともと稼動割合が低い上に、電力供給可能量が基本需要を下回らない限り電力価格が原価を上回ることがなくなってしまう。

現実の電力市場では、需要家に容量確保義務を課すことにより、容量市場において稼動していない発電設備容量にも価格がつくような工夫も為されている。モデルにおいては、政府機関エージェントが設備に補助金を与えるという形でこれに代替しており、現実の問題を正確にとらえていない面がある。

そのほかの非線形要素の組み入れ

このモデルでは、逐次最適化と繰り返し計算を行っているため、技術コストの習熟効果、技術の波及効果などの時間に対し非線形な要素を組み入れることができる。これにより何か新しい解析ができるのではないかと期待される。

参考文献

[1] 環境省「2004 年度(平成 16 年度)の温室効果ガス排出量速報値について」 2005 年 10 月
http://www.env.go.jp/earth/ondanka/ghg/index.html
[2] 日本経済団体連合会「温暖化対策 環境自主行動計画 2005 年度フォローアップ結果 概要版」2005 年
[3] 電気事業連合会「電気事業における環境行動計画」2003 年
[4] LaCasse C., Plourde A., “on the Renewal of Concern for the Security of Oil Supply”, Energy Journal, VOl.16, No.2
[5] 経済産業省資源エネルギー庁 電力･ガス事業部「海外諸国の電力改革の現状と制度的課題」, 2001 年
[6] 地球環境産業技術研究機構「「地球再生計画」の実施計画作成に関する調査事業」 2001 年
[7] 地球環境産業技術研究機構「エネルギー使用評価システムに関する調査」 2003 年
[8] 槌屋治紀「技術革新の原動力は何か― ― ― 地球温暖化問題の解決のために」 中央環境審議会地球環境部会「京都議定書を巡る最近の状況に関する懇談会」提出資料
http://www.env.go.jp/council/06earth/y060-kyo/mat03.pdf
[9] 石油公団企画調査部「日本の石油備蓄の経済効果」 『石油の開発と備蓄』石油公団、1994 年 6 月
[10] 森谷友祐･森俊介･森本慎一郎「拡張 MARIA による超長期不確実事象の中短期政策への影響評価」 第 24 回エネルギー･資源学会研究発表会後援論文集、エネルギー･資源学会
[11] 伊理正夫･今野浩編『数理計画法の応用＜理論編＞』 産業図書,1982 年
[12] 渡邊裕美子･林武人･藤井康正･山地憲治「確率計画法によるエネルギーモデルへの不確実性の導入」 第 24 回エネルギー･資源学会研究発表会後援論文集、エネルギー･資源学会
[13] 松原望『意思決定の基礎』朝倉書店、1985 年
[14] 濱上知樹「知的情報処理を用いた電力市場のシミュレーション」 電気学会論文誌 C、Vol.126, No.2, 2005 年
[15] 篠原剛「エネルギー戦略の国際競合関係考慮のためのエージェントベース世界エネルギーモデルの構築」 東京大学修士論文、2005 年
[16] エリック･ラムスセン著、細江守紀･村田省三･有定愛展訳『ゲームと情報の経済分析 I』九州大学出版会、1990 年
[17] Richard S. Sutton and Andrew G. Barto 著、三上貞芳･皆川雅章訳『強化学習』 森北出版株式会社、2000 年
(Richard S. Sutton and Andrew G. Barto, “Reinforcement Learning: An Introduction”
http://www.cs.ualberta.ca/%7Esutton/book/ebook/the-book.html)
[18] 小山友介･塩瀬隆之「人間‐人間間相互作用」 『計測と制御』第 44 巻第 12 号、2005 年 12 月
[19] 高玉圭樹『マルチエージェント学習－相互作用の謎に迫る』コロナ社、2003 年
[20] 出口弘『複雑系としての経済学 自立的エージェント集団の科学としての経済学を目指して』 日科技連出版会、2000 年
[21] 生天目章･荒井幸代･服部聖彦「エージェント間の相互作用：望ましい関係性の創発」 『計測と制御』第 44 巻第 12 号、2005 年 12 月
[22] W. D. ノードハウス著、室田泰弘･山下ゆかり･高瀬香絵訳『地球温暖化の経済学』東洋経済新報社、2002 年
[23] 服部恒明他「2025 年までの経済･エネルギーの長期展望―持続的成長への途を求めて―」 電力中央研究所報告、2003 年 4 月
[24] 沈中元「日本におけるエネルギー需要の所得と価格の短･長期弾性値の計測」 第 19 回エネルギーシステム･経済･環境コンファレンス講演論文集、エネルギー･資源学会
[25]ILOG “ILOG Concert Technology 1.0 User’s Manual”, 2000

[26]西尾健一郎「我が国における再生可能エネルギー導入基準制度 RPS の評価」 東京大学修士論文、2002 年
[27]経済産業省資源エネルギー庁編「みつめよう！我が国のエネルギー―エネルギー環境制約を超えて―」 財団法人経済産業調査会、2001 年
[28]Official Energy Statistics from the U.S. Government “Annual Energy Outlook 2006” http://www.eia.doe.gov/oiaf/aeo/index.html
[29]石川城太「やさしい経済学－ゲーム理論で解く 通商政策と戦略 第 3 回 補助金の影響」 日本経済新聞 2005 年 7 月 14 日

発表実績

1 篠原剛･渡邊裕美子･林武人･藤井康正･山地憲治「エネルギー戦略の国際競合関係考慮のためのエージェントベース世界エネルギーモデルの構築」マルチエージェント型日本電力需給モデルの構築」 平成 17 年電気学会全国大会、2005 年 3 月

2 渡邊裕美子･林武人･藤井康正･山地憲治「確率計画法によるエネルギーモデルへの不確実性の導入」 第 24 回エネルギー･資源学会研究発表会、2005 年 6 月

3 Yumiko Watanabe, Takeshi Shinohara, Taketo Hayashi, Yasumasa Fujii, Kenji Yamaji “Analysis of the Energy System by Energy Model Formulated as Multi-agent Simulation” IEW2005, 2005 年 7 月

4 Yumiko Watanabe, Takeshi Shinohara, Taketo Hayashi, Yasumasa Fujii, Kenji Yamaji “Development of Agent-Based Global Energy Model for Considering International Competition of Energy Strategy” EIC2005, 2005 年 7 月

5 渡邊裕美子･林武人･藤井康正･山地憲治「マルチエージェント型日本電力需給モデルの構築」第 22 回エネルギーシステム･経済･環境コンファレンス、2006 年 1 月

謝辞

本研究を進めるにあたり、多くの方々にご指導、ご協力を頂きました。この場を借りて御礼申し上げます。

山地憲治教授には、指導教官としてご多忙にもかかわらず多くのご助言をいただきました。藤井康正助教授には、テーマ設定から研究の進め方に至るまで丁寧にご指導いただきました。また、お二方のご尽力のおかげで、国際学会発表の貴重な機会をいただくことができました。

林武人助手には、快適な計算機環境を整備して下さり、また研究室生活において様々なことを気にかけて下さいました。今年度で退官とのことで、これまでのご貢献に深く御礼申し上げるとともに、今後のご健勝をお祈りいたします。山本博巳客員助教授、竹下貴之助手、電力中央研究所の西尾健一郎様にも、研究室打ち合わせのたびに適切なご指摘をいただきました。

エネルギー総合工学研究所主催のエネルギーモデル検討委員会においては、当研究所の黒沢厚志様、東京理科大学の森俊介先生、電力中央研究所の長野浩司様をはじめ、多くの方々に研究への助言をいただきました。この機会を与えていただきました藤井康正助教授にも感謝申し上げます。

2004 年度修了の庭山亮一さん、2005 年修了の篠原剛さんには、モデルについて懇切丁寧にご指導いただき、また研究の成果を引き継がせていただきました。そのほかにもいろいろな形でご支援下さった修士課程の皆様、卒論生の皆様、諸先輩方、秘書の方々のおかげで非常に有益な研究室生活を送ることができました。

また、両親･姉妹には、長期に渡った学生生活を経済面･精神面から支えていただきました。

最後に、この 6 年間を共に過ごした友人への心からの謝意を表して、本研究の結びといたします。

2006 年 2 月
渡邊裕美子