再生型生命維持システムの運用管理における情報の取り扱いに関する考察

Considerations concerning use of Information in Operation Management of Regenerative Life Support System

○宮嶋宏行（東京女学館大）、広崎朋史（日大）、石川芳男（日大）

Hiroyuki Miyajima *, Tomofumi Hirosaki**, Yoshio Ishikawa**
*Tokyo Jogakkan College, 1105 Tsuruma, Machida-shi, Tokyo 194-0004, Japan
E-mail : miyajima@m.tjk.ac.jp
**Nihon University, 7-24-1 Narashinodai, Funabashi-shi, Chiba 274-8501, Japan

Abstract

A Regenerative Life Support System (RLSS) is a system that establishes self-sustained material recycling and circulation within a space bases on the Moon or Mars. It is a large-scale and complicated system comprising a lot of components such as human, plants and material circulation system. A RLSS has many factors with uncertainty, such as dynamics of plants and humans, and failure and performance deterioration of devices. An environment with uncertainty or a large-scale and complicated system may not be properly addressed by a centralized system. In particular, such system cannot always gather accurate information in one place in a frequently shifting environment, thus appropriate processing may be difficult. Therefore, we attempted autonomous decentralization of information or decision-making using a Multi-Agent System (MAS). This report discussed the use of the information in operation management of a RLSS when materials circulation control system is designed using a MAS and a method in which a MAS acquires cooperative action using the information in bottom-up by means of computer simulation. So far, we confirmed the effectiveness of this method for a RLSS material circulation control system. Hence it was proved to enable the automatic acquisition of a cooperation rule with the autonomous learning among agents.

Key words: Distributed Control System, Multi-Agent System, Reinforcement Learning, Q-Learning

１．はじめに

再生型生命維持システム(RLSS :Regenerative Life Support System)とは、人間の生活に必要な物質を再生しながら生命を維持するシステムで、人間・植物・物質循環システムなどからなる。植物は人間に食料を供給したり、光合成により水や気体を再生したりする役割を果たし、物質循環システムは人間や植物が使用した物質を物理化学的に再生して循環させる役割を果たす。このシステムは有人宇宙活動の内容が、これまでのような短期のものから宇宙ステーション、月面基地、火星基地などに代表される長期の活動に移ってきたために注目されるようになった。RLSS は、バッチ操作による処理系や生物を含むので、運転サイクルの違いや生物の代謝量の変動により複雑な挙動を示す[1)]。通常、バッチ処理を多く有するプラント等では自動で計測・制御する部分と人間により監視・操作する部分が存在する。しかし、宇宙基地での RLSS の利用を考えた場合、クルーが本来のミッションに十分な時間を使うことができるようにするためには、運用管理の一層の自動化が期待される。例えば、ミッションにあわせて事前に運用スケジュールを計画することも考えられるが、動植物の代謝量の変動や装置性能の変化などにすべて対応することはできない。そこで、環境が変化した場合にも柔軟に対応できるような運用管理システムが必要となる。実際には、不確実性がある環境を取り扱う場合、分散した大量の情報を一箇所に集め、それを基に処理する中央集権型システムでは限界がある。特に、頻繁に変化する環境では正確な情報をつねに一箇所に集めることはできず、適切な処理が難しくなる[2)]。本研究では、不確実性がある環境を取り扱うために、学習主体者(エージェント)が環境と相互作用し、情報の獲得と行動の選択をするメカニズムを持つ分散制御型システム(DCS :Distributed Control System)の RLSS 物質循環制御系への適用について考える。DCS を利用することで情報や意思決定の自律分散化を図ることができ、この DCS の有力な技術の１つがマルチエージェントシステム(MAS : Multi-Agent System)である。本報告では、MAS を利用して RLSS 物質循環制御系を設計する場合の情報の取り扱いについて検討し、さらに MAS がそれらの情報を利用してボトムアップ的に協調的行動を獲得する方法についてシミュレーションを利用して検討する。

２．マルチエージェントシステム

本研究では、RLSS 物質循環システムのシステム形態を以下のように捉え[3)]、MAS を利用して RLSS 物質循環システムの制御系を設計する。

- システムの構成要素を独立したエージェントと捉え、構成要素の状態と動作を情報と捉える。
- 分散配置された個々のエージェントは、他のエージェントと協調しながら問題を解決する。
- エージェントの協調によって、システム全体の物質循環制御が維持される。

MAS は学習能力を持って初めて、環境変化に柔軟に対応することができる。ここでいう学習能力とは、過去の経験を活かすことによって未来の行動の自己改善を行う能力である。また、大規模なシ

ステムの統合的な運用を計画するためには、独立して意思決定する個々のエージェントが、お互いに協調するためのアルゴリズムが必要である。しかし、MAS において、協調のメカニズムを設計することは容易ではない。MAS を協調させるには、情報を共有して協調させる方法(協調学習)と情報を共有しないで協調させる方法(分散学習)の２つがある。前者は、明示的な相互作用や通信を陽に用いて全体をコントロールする。後者は、明示的な相互作用や通信を用いないで全体をコントロールする。前者の方法は、グローバルに情報を共有するため状態の爆発を招く。この状態の爆発を防ぐためには関連するエージェント間で必要な情報のみに限定して共有すべきであるが、必要な情報が何であるのかを事前に判断することは非常に難しい。そこで情報を直接には共有しないで協調的行動を獲得する方法が必要となる。その有力な方法の１つである強化学習について検討する。強化学習を用いて協調的行動を実現するにはエージェント間で「目的」と「報酬」を共有する必要がある。エージェント間で報酬を共有することは、システムの目的と制約条件の間に複雑なトレードオフがあるので容易ではないといわれている[2]。エージェント間で協調を実現するための報酬が設計できれば有効であるが、設計者がトレードオフを事前に与えるのではなく、学習により適応的かつ柔軟に決められるのが望ましい。そこで、情報を危険レベルという指標を用いて隣り合うエージェント同士で共有し、その伝播を通してエージェント間の協調を図る[3]。

３．強化学習

強化学習とは、環境との相互作用を通じて、環境に適応する制御規則を獲得する学習法のことである[4]。教師信号がなくとも行動の結果から得られる評価を手掛かりとして、どの行動が最も高い報酬を得ることが出来るのかを試行錯誤を繰り返すことにより学習していく。つまり環境のモデルを直接必要としない。行動と評価のサイクルを繰り返しながら学習が進む。すなわち強化学習は動的な環境にもある程度対応できる可能性がある。確率的な行動規則の獲得が可能なため、ノイズの多い実環境にもある程度対応できるという特徴がある。Fig. 1 に強化学習機構の概念図を示す。

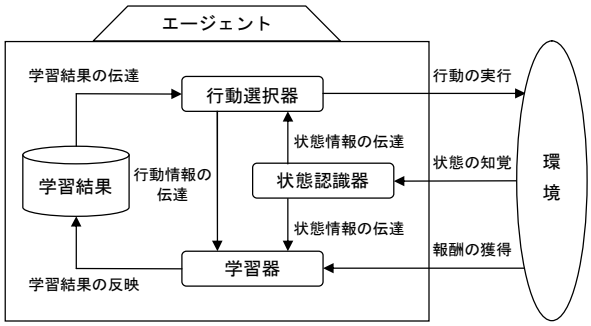

Fig. 1　強化学習機構の概念図

RLSS の物質循環制御に強化学習のアルゴリズムとして最も一般的な *Q*-Learning を適用する。*Q*-Learning では政策と呼ばれる状態と行動の組に対する行動価値関数が計算される。この関数の値を *Q* 値と呼ぶ。このアルゴリズムでは、ある状態における行動ごとの *Q* 値を学習することにより、その状態でとるべき最適行動を獲得することが可能となる。状態 x において行動 a をとった場合の Q 値は $Q(x,a)$と記述される。すべての状態 x と行動 a に対する Q 値を Q テーブルと呼ぶ。エージェントが時刻 t の状態 x_t において行動 a_t を選択した結果、状態 x_{t+1} となり、報酬 r_t が得られたとすると、この Q 値は以下のように更新される。

$$Q(x_t,a_t)=(1-\alpha)Q(x_t,a_t)+\alpha\left(r_t+\gamma\max_{a_k}Q(x_{t+1},a_k)\right) \quad (1)$$

ここで、αは学習率と呼ばれるパラメータであり、$0<\alpha\leq 1$ である。学習率が小さいほど今までの推定値を重視し、逆に大きいほど得られた結果を重視する。またγは割引率と呼ばれるパラメータであり、$0\leq\gamma\leq 1$ である。割引率は、将来獲得予定の報酬を現時点でどれだけ重要と考えるかの割合を示す。よって、(1)式の右辺第一項は過去の学習による経験値を表わし、第２項は今回の学習結果(報酬)を表わし、第３項は未来の最適行動をとった場合の推定価値を表わす。すなわち、次の状態において採用できる行動の価値を考慮に入れた学習が可能となる。ある状態において選択可能な複数の行動に対して Q 値が高いほど良い行動であり、低いほど悪い行動ということになる。この式において Q 値は逐次更新が可能であるのでエージェントが試行錯誤的に行動を行う中で Q 値を学習することができる。*Q*-Learning を利用した制御アルゴリズムを Fig. 2 に示す。

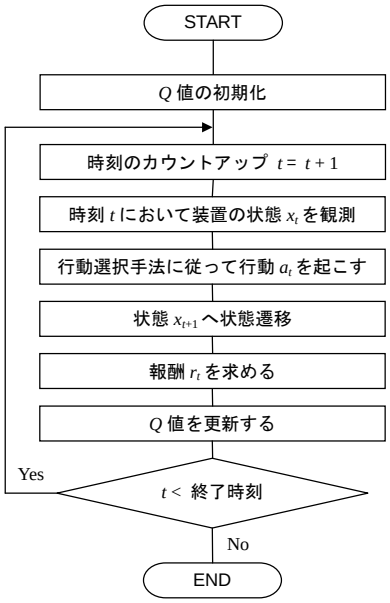

Fig. 2　*Q*-learning を利用した制御アルゴリズム

状態空間の設計

各装置のエージェントを設計する場合には、状態数を如何に抑えるかということが重要である。ここでは状態空間を設計するために

危険レベルという概念を導入する。危険レベルとは、モジュール内の物質量やタンク内の物質量をいくつかの領域に分割し、離散的に表現する指標である[3]。ここでは、危険レベルを Table 1 に示すように５段階で表現し、正の値をとると目標より物質量が多すぎることを意味し、負の値をとると目標より物質量が少なすぎることを表わす。すなわち危険レベルの絶対値が大きくなるほどタンクは危険な状態となる。エージェントとした装置が隣り合うモジュールやタンクの危険レベルの組合せを用いて状態が定義される。例えば、O_2 分離装置エージェントは、植物栽培モジュールの O_2 濃度と O_2 タンクの O_2 量の 2 つの物質量を計測しているので、状態数は 5×5=25 となる。

次に湿式酸化装置のエージェントを設計する場合には、その装置の動作回数が少ないため、状態数が多いと十分な学習ができない可能性がある。湿式酸化装置エージェントは O_2 タンクの O_2 量と CO_2 タンクの CO_2 量を計測して on を決定する(off になるのは 8 時間後である)。そこで、この装置の動作を決定するときに使われる O_2 タンクと CO_2 タンクの物質量に湿式酸化装置の特性に合わせて Table 2, 3 に示すような危険レベルの定義をすることで状態数を少なくする。さらに、O_2 が 1 バッチに必要な量以上あるときエージェントが判断を行うという条件をつけ、イベントドリブン型のエージェントとすることで状態の一部を使用しないようにして、実際に利用する状態数を削減した。つまり、状態数は 3×3=9 であるが、O_2 タンクの危険度が 0 のときの状態は使われないので実際の状態数は 3×2=6 となる。

Table 1 危険レベルの定義

物質量		危険レベル
以上	以下	
最大許容量	最大値	2
(目標値＋最大許容量)/2	最大許容量	1
(目標値＋最小許容量)/2	(目標値＋最大許容量)/2	0
最小許容量	(目標値＋最小許容量)/2	-1
最小量	最小許容量	-2

Table 2 湿式酸化装置エージェントにおける O_2 タンクの危険レベルの定義

O_2 タンクの物質量	危険レベル
(最大許容量－湿式酸化に必要な O_2 量)以上	2
(最大許容量－湿式酸化に必要な O_2 量×2)以上	1
(最大許容量－湿式酸化に必要な O_2 量×2)未満	0

Table 3 湿式酸化装置エージェントにおける CO_2 タンクの危険レベルの定義

CO_2 タンクの物質量	危険レベル
(最大許容量－湿式酸化により発生する CO_2 量)以上	0
(最大許容量－湿式酸化により発生する CO_2 量×2)以上	1
(最大許容量－湿式酸化により発生する CO_2 量×2)未満	2

報酬の設計

報酬は危険レベル *risk* の改善度合いと定義する。つまり、接続されているモジュールやタンクの物質量が N 個計測される場合、時刻 t における報酬 r_t は式(2)のようになる。

$$r_t = \sum_{k=1}^{N} \left(\left| risk_{k(t)} \right| - \left| risk_{k(t+1)} \right| \right) \tag{2}$$

行動選択手法

Q-learning は有限マルコフ決定過程においては、行動選択手法によらず、全ての状態が十分にサンプルされるならば最適な Q 値に収束することが保証されている。マルチエージェントの問題は有限マルコフ決定過程ではないが、*Q*-Learning はマルチエージェントの問題においても有力な手法の１つである。マルチエージェントの問題においてランダム方策を用いることは、収束までにかかる時間ステップが非常に大きくなるという問題を生じさせる。そのため学習後だけではなく学習中も Q 値を利用する方法が一般的である。先ほど定義された状態と行動の組合せからなる Q テーブルの Q 値に従って、それぞれのルールの価値の比によって確率的に行動が選択される。比の計算には、ボルツマン分布が用いられる。

$$\pi(x,a) = \frac{e^{Q(x,a)/T}}{\sum_{a_i \in A} e^{Q(x,a_i)/T}} \tag{3}$$

ここで、$\pi(x,a)$は、状態 x において行動 a が選択される確率を表わす。T はボルツマン温度である。T が大きいほどランダムな行動が選択され、T が 0 に近いほどグリーディな行動が選択される。

４．RLSS モデルの説明

例題として取り上げる RLSS の物質循環システムを Fig. 3 に示す。ここでは、MAS を利用した RLSS 物質循環制御系の有効性を確かめることが目的であるので、O_2 と CO_2 の循環に関係した物質のみがモデル化されている。次に、植物、人間、O_2 分離装置、CO_2 分離装置、湿式酸化装置、O_2 供給装置、CO_2 供給装置のそれぞれのモデルや動作条件について示す。

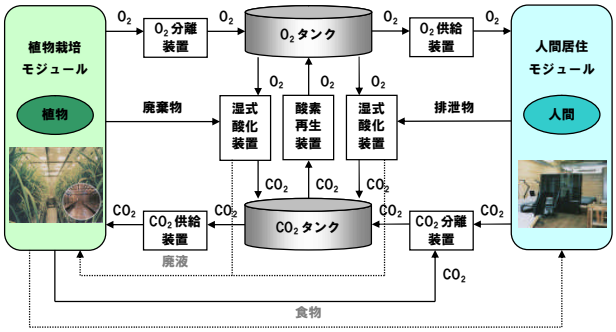

Fig. 3　RLSS の物質循環システム
(O_2 と CO_2 の循環に関係する物質のみに限定)

植物

植物成長モデルは、ロジスティック関数によりモデル化されてい

る。植物の光合成、すなわちバイオマスの単位面積あたりの増加率 dx/dt は、

$$dx/dt = \alpha x\left(1 - x/x_m\right) \tag{4}$$

と記述できる。ここで x はバイオマス[g/m^2]、x_m は最大バイオマス[g/m^2]、αは光合成係数[1/day]である。x_m は植物種によって、αは光の強さ、温度、二酸化炭素濃度などの環境因子によって決まり、

$$\alpha = \alpha_{avg}\alpha_I\alpha_T\alpha_{co2} \tag{5}$$

と表せる。ここでα_{avg}は平均光合成速度である。α_I, α_T,α_{co2}は光・温度・二酸化炭素濃度が植物の成長に与える影響を無次元関数で表したものである。α_I, α_{co2}はそれぞれ式(6), (7)のようになる。α_Tは1と設定し常に変動なしと仮定する。

$$\alpha_I = a_I \cdot I \qquad (0 \le I \le 2000) \tag{6}$$

$$\alpha_{co2} = \begin{cases} a_{co2}co2 + c_{co2} & (100 \le co2 < 500) \\ 1 & (co2 \ge 500) \end{cases} \tag{7}$$

ここで I は光の強さ［μ mol/m^2/s］、$co2$ は二酸化炭素濃度［ppm］である。a_I, a_{co2}, c_{co2}はα_I, α_{co2}が線形関数で表わせると仮定した場合の係数である。このときシーケンシャル栽培(シーケンシャル栽培とは、栽培ベットの中で植付けを分割して定期的にずらしながら連続的に栽培と収穫をする方法である)を行っている植物 i の栽培フェーズ j の植物成長モデルは式(8)のように表わせる。

$$dx_{ij}/dt = \alpha_i x_i \left(1 - x_{ij}/x_{mi}\right) \tag{8}$$

ここで、i は植物の種類を表わし、j はその植物の栽培ステージを表わす。よって植物栽培モジュール全体の植物成長モデルは式(9)となる。

$$dx/dt = \sum_{i=1}^{M}\sum_{j=1}^{N_i} dx_{ij}/dt \tag{9}$$

ここで M は最大栽培植物数、N_iはある植物の最大栽培ステージ数を表わす。

人間

人間の活動の強さは、式(10)に示す活動指数関数 $z(t)$によってモデル化されている。

$$z(t) = \left(1 - A\cos(2\pi t/T)\right) \tag{10}$$

ここで、A は活動指数の振幅($0 < A < 1$)、t は時間(min)、T は活動の周期であり通常は一日の時間(min)である。

人間の物質入出力のうち O_2, CO_2 は $z(t)$によって変動し、食事や排泄はあらかじめ決められた時刻に、あらかじめ決められた量行われる。

O_2分離装置

O_2分離装置は、空気から O_2 を分離する装置である。この装置の分離能力は常に一定とする。この装置は on/off を 1 時間に 1 回決定できる。

CO_2供給装置

CO_2供給装置は、差圧により CO_2 タンクから植物栽培空間に CO_2 を供給する装置である。この装置は on/off を 3 分間に 1 回決定できる。

O_2供給装置

O_2供給装置は、差圧により O_2 タンクから人間居住空間に O_2 を供給する装置である。この装置は on/off を 3 分間に 1 回決定できる。

CO_2分離装置

CO_2 分離装置は、空気から CO_2 を分離する装置である。この装置は on/off を 1 時間に 1 回決定できる。この装置の分離能力 v_{co2} は式(11)のようにモデル化されている。

$$v_{co2} = \begin{cases} a_{co2} \cdot co2 & (0\,ppm \le co2 < 2000\,ppm) \\ c_{co2} & (co2 \ge 2000\,ppm) \end{cases} \tag{11}$$

ここで、$co2$ は二酸化炭素濃度、a_{co2}, c_{co2}は v_{co2}が 2000ppm で飽和する線形関数で表わせると仮定したときの係数である。

湿式酸化装置

湿式酸化装置は、廃棄物である有機物を物理化学的処理により、水、二酸化炭素、硝酸、アンモニアなどの無機物に分解する装置である。この装置は、処理に８時間を必要とし、最大でも１日３回の運転しか行うことができない。この装置は O_2 が 1 バッチに必要な量以上あるときのみ動作 on を 1 分ごとに判断できる。

５．結果

シミュレーション設定

人間１人が居住し、植物を利用した食糧生産により人間に食物が供給される。食糧生産のための作物には大豆が使われた。今回のシミュレーションでは、MAS を利用した RLSS 物質循環制御系の有効性を確かめることが目的であるので栽培作物は 1 種類のみとした。植物の栽培量は乾燥質量で 735 g /day、栽培面積は 120m^2 である。植物は 8 ステージに分けられ、明期 12 時間、暗期 12 時間のシーケンシャル栽培で栽培される。O_2 分離装置、CO_2 供給装置、O_2 供給装置、CO_2 分離装置、湿式酸化装置の定常処理能力は、それぞれ 4.35g/min, 4.86g/min, 1.21g/min, 3.38g/min, 1012g(dry mass)/8h である。O_2 タンク容量、CO_2 タンク容量はそれぞれ 6258g , 6998g である。人間居住空間の容積は 150m^3、CO_2 濃度の設定は 400ppm、O_2 濃度の設定は 20.96%、および植物栽培空間の容積は 269m^3、CO_2 濃度の設定は 700ppm、O_2 濃度の設定は 20.93%である。

このシミュレーションでは Fig. 3 に示した装置のうち O_2 分離装置、CO_2 供給装置、O_2 供給装置、CO_2 分離装置、植物栽培モジュールの湿式酸化装置をそれぞれエージェントとした。その他の O_2 再生装置と人間居住モジュールの湿式酸化装置は事前に決められたスケジュールで運転される。エージェントの学習は植物のシーケン

シャル栽培が 80 日で完成した後の閉鎖系モードで開始され、200 日までの 120 日間行われる。ここで 5 つのエージェントがそれぞれ 2 つの行動を持つので、物質循環制御系全体としては 32(2×2×2×2×2)の行動パターンを持つことになる。学習を進めることによってこの行動の中から最も良い行動が選ばれるようになる。

シミュレーション結果

シミュレーション結果では、まず強化学習の各種パラメータの影響について制約逸脱率の観点から解析する。制約逸脱率とはシミュレーション中の全学習回数に対して物質量がタンクやモジュールの上限や下限を超えた回数の割合をいう。次に、環境が変化したとき学習によってエージェントが新しい環境に適応できたかどうかを on/off 切替頻度と報酬獲得頻度の平均値の観点から解析する。

① 各種パラメータの影響

パラメータ α, γ, T が強化学習に与える影響を制約逸脱率を基に示す。αを 0.1 から 0.9 まで変化させた場合の各装置の制約逸脱率を Fig. 4 に示す。ただし、すべての場合において制約逸脱率が 0%であった O_2 分離装置、O_2 供給装置、湿式酸化装置を除く。αが 0.9 のときのみ CO_2 供給装置の制約逸脱率が 11.4%、CO_2 分離装置の制約逸脱率が 3.4%であった。このようにほとんどのαにおいて制約逸脱率が 0%になったのは危険度が-2, -1, 1, 2 の場合の Q 値の初期値にフィードバック制御となるように on の行動に初期値-1 や 1 を与え、すでにある程度学習された状態から計算を始めたためである。仮に Q 値の初期値をすべて 0 とした場合には、αは非常に感度の高いパラメータとなることが事前のシミュレーションによって確かめられている。

γを 0.1 から 0.99 まで変化させた場合の各装置の制約逸脱率を Fig. 5 に示す。ただし、すべての場合において制約逸脱率が 0%であった O_2 供給装置、湿式酸化装置を除く。γが 0.99 の場合を除いて制約逸脱率は 0%とはなっていない。特に CO_2 供給装置と CO_2 分離装置の制約逸脱率はγが 0.99 の場合を除いて非常に高い値を示している。

T を 0.1 から 0.9 まで変化させた場合の各装置の制約逸脱率を Fig. 6 に示す。ただし、すべての場合において制約逸脱率が 0%であった O_2 分離装置、O_2 供給装置、湿式酸化装置を除く。T が 0.1 と 0.2 の場合を除いて制約逸脱が発生している。特に T が 1 に近づくほど高い値を示している。

よってこれ以降のシミュレーションは、α=0.1、γ=0.99、T=0.1 で行われる。

② CO_2 分離装置の能力が低下した場合の影響

Table 4 に CO_2 分離装置の能力が 141 日目以降に 50%低下した場合の能力低下前と能力低下後の CO_2 分離装置の on/off 切替頻度と報酬獲得頻度の平均値を示す。能力の低下が起こった場合、切替頻度の平均値が 7.50 回/day から 2.87 回/day に下がっている。これは、能力の低下を補うために切替頻度を減らして連続運転を続けたためである。次に報酬獲得頻度は、能力の低下が起こらなかった場合には、すでに学習が進んでいるため 0.40 回/day から 0.28 回/day に低下しているが、能力の低下が起こった場合には、新たな環境に適応するために学習が促進されて 0.42 回/day から 0.83 回/day に上昇している。つまり、環境の変化に対してエージェントの学習が適切に行われているといえる。

Fig. 7 に同様の場合の CO_2 分離装置の切替頻度の変動を示す。141 日目以降に CO_2 分離装置の能力が 50%低下したため CO_2 分離装置の切替頻度が突然下がっている。先ほど述べたように on/off の切替頻度を減らして連続運転に移行した様子がわかる。エージェントが学習により新しい環境に適応した結果といえる。

③ 湿式酸化装置の運用方法を変更した場合の影響

湿式酸化装置を毎日運転した場合と 141 日目以降に隔日運転した場合の O_2 分離装置、CO_2 供給装置、O_2 供給装置、CO_2 分離装置の切替頻度と報酬獲得頻度の平均値を Table 5 に示す。CO_2 分離装置の能力低下とは違って、すべての装置に影響が波及する事例であるが、O_2 供給装置で報酬獲得頻度が上がっていることを除けば大きな変化は起こっていない。学習の結果、運転方法を大きく変更する必要はなかったといえる。

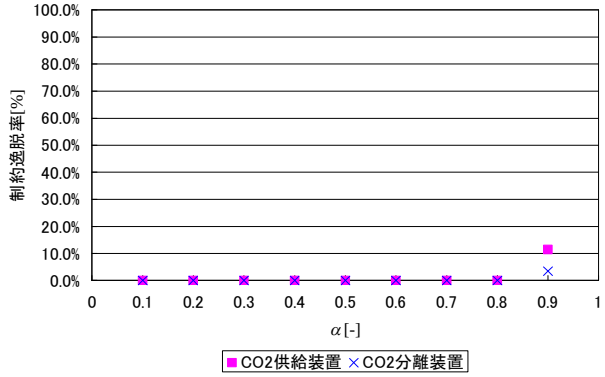

Fig. 4　学習率αを変化させた場合の制約逸脱率

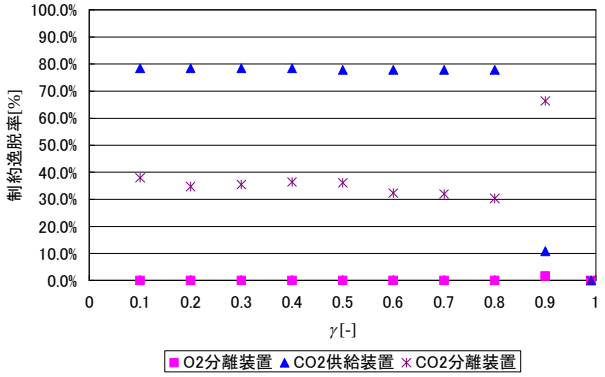

Fig. 5　割引率γを変化させた場合の制約逸脱率

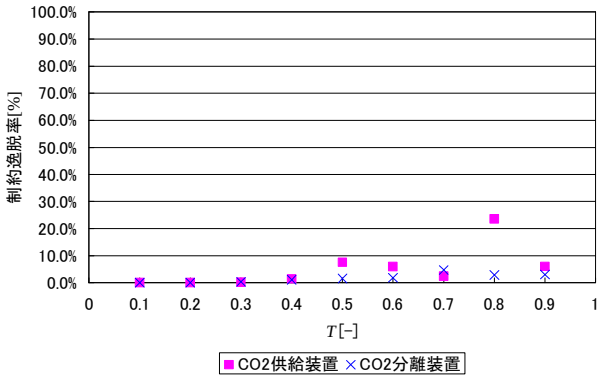

Fig. 6　温度パラメータ T を変化させた場合の制約逸脱率

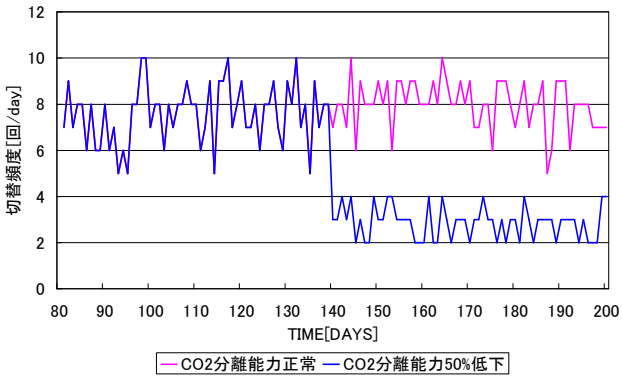

Fig. 7　CO_2 分離能力が 50%低下した場合の CO_2 分離装置の切替頻度の変化

Table 4　CO_2 分離能力が 50%低下した場合の CO_2 分離装置の切替頻度と報酬獲得頻度

期間[日]	CO2 分離能力正常		CO2 分離能力 50%低下	
	切替頻度	報酬獲得頻度	切替頻度	報酬獲得頻度
	[回/day]	[回/day]	[回/day]	[回/day]
81-140	7.57	0.40	7.50	0.42
141-200	8.02	0.28	2.87	0.83

Table 5　湿式酸化装置の運用方法を変更した場合の切替頻度と報酬獲得頻度

装置	期間[日]	毎日運転した場合		隔日運転した場合	
		切替頻度	報酬獲得頻度	切替頻度	報酬獲得頻度
		[回/day]	[回/day]	[回/day]	[回/day]
O_2 分離装置	81-140	6.27	0.10	6.27	0.10
	141-200	6.22	0.17	6.32	0.20
CO_2 供給装置	81-140	126.70	32.68	126.70	32.68
	141-200	126.77	33.98	126.60	36.20
O_2 供給装置	81-140	120.02	0.28	120.02	0.28
	141-200	121.03	1.05	119.80	5.28
CO_2 分離装置	81-140	7.57	0.40	7.57	0.40
	141-200	8.02	0.28	7.90	0.37

６．まとめ

本報告では、MAS を利用して RLSS 物質循環制御系を設計する場合の情報の取り扱いについて検討し、さらに MAS がそれらの情報を利用して協調行動を獲得する方法について検討した。ここでは、グローバルに情報を共有するのではなく、情報を危険レベルという指標を用いて隣り合うエージェント同士で共有し、その伝播を通してエージェント間の協調を実現できた。Q-Learning における α, γ, T のそれぞれのパラメータを適切に調整すれば個々のエージェントが協調して物質循環の制御を達成できた。また、CO_2 分離装置の能力が 50%低下した場合と湿式酸化装置の運用方法を変更した場合の結果から環境が変化した場合にも学習により、新たな運用方法が獲得できた。

参考文献

1. K. Abe, M. Endo, K. Nitta, H. Miyajima, Y. Ishikawa, S. Kibe, A Simulation Model for the CEEF Behavioral Prediction System, SAE 2003-01-2547, 2003
2. 高玉圭樹, マルチエージェント学習　ー相互作用の謎に迫るー, コロナ社 (2003)
3. T. Hirosaki, N. Yamauchi, H. Yoshida, Y. Ishikawa, H. Miyajima, Application on Multi-agent Reinforcement Learning to CELSS Material Circulation Control, PAIS2001 Conference Proceedings (2001)
4. Richard S. Sutton and Andrew G. Barto, Reinforcement Learning: Introduction, A Bradford Book, The MIT Press (1998)
5. H. Miyajima, K. Abe, Y. Ishikawa, A. Ashida and K. Nitta, Simulation to Support an Integration Test Project of CEEF, SAE 2001-01-2130, 2001