

先端生命維持システムの運用計画立案の支援に関する検討

Study of operation planning support for Advanced Life Support Systems

○宮嶋宏行（東京女学館大）、広崎朋史（宇宙開発システム）、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

** Space Systems Development Corporation, 2-10-1-202 Minamihanazono, Hanamigawa-ku, Chiba-shi, Chiba 262-0022, Japan

***Nihon University, 7-24-1 Narashinodai, Funabashi-shi, Chiba 274-8501, Japan

Abstract

This paper discusses the integration of intelligence into the supervision and control system of an Advanced Life Support System (ALSS). An ALSS is a complex, large-scale system that should be maintained solely by a small crew. Since an ALSS is operated far from Earth, the propagation delay of radio waves makes remote support from Earth difficult to implement immediately. Accordingly, an autonomous supervision and control system is essential. A supervision and control system comprises an automatic controller for ALSS equipment and crew who supervise the equipment conditions and operate it. The application of intelligent control to such an automatic controller reduces the load on the crew, allowing them to concentrate on their major missions. The concept of intelligent control proposed here is based on the SRK model that expresses human behavior with three cognitive behaviors. This paper utilizes simulation to verify the effectiveness of the proposed methods for intelligent control.

Key words: Intelligent Control, SRK Model, Hierarchical Control, Planner, Fuzzy inference

1. はじめに

先端生命維持システム(ALSS : Advanced Life Support Systems)は、地球から遠く離れた場所での長期間のミッションに対応するための生命維持システム(LSS : Life Support Systems)である。国際宇宙ステーション(ISS : Intentional Space Station)のLSSとの違いは、食料生産を行うこと、物理化学的処理に加えて生物学的処理を利用すること、そして作物の非可食部や人間の排泄物を再生して資源を回収することである¹⁾。想定される代表的なミッションには、有人火星探査(行き150日+火星滞在619日+帰り110日=ミッション期間879日)²⁾がある。ALSSは、複雑かつ大規模なシステムであるため、これを少数の搭乗員だけで維持することは困難であり、搭乗者が主要なミッションに専念するためには監視制御システムの自律化が必要である。また、ALSSは地球低軌道(LEO : Low Earth Orbit)のISSのLSSに比べ地球から離れた場所で利用されるため地上からの支援には電波伝播遅れによる時間差が発生する。その理由からも監視制御システムの自律化が必要となる。

監視制御システムは、一般的にはFig.1のように定義される。それはプラント機器のための自動制御装置とその状況を監視し、操作する運転員からなる。ここでは自動制御装置の知能化を図ることによって運転員の負荷の軽減を図る。自動制御装置の知能化とは、現在の状況に基づいて機器の操作(機会主義的制御)を決定したり、あらかじめ決められた規則に基づいて機器の操作(戦術的制御)を決定したりするだけでなく、全体的状況を把握し、将来の

目標までを含めた長期的視野に立った操作(戦略的制御)を実現することである。このような制御は、人間の行動を3つの認知活動によって表現したSRKモデルに基づいて定型化できる。そこで本研究では、SRKモデルに基づいた知能制御をALSSの監視制御システムへ適用した運用計画立案の支援について論じる。

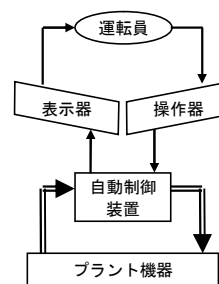


Fig.1 監視制御システム³⁾

2. 知能制御の概念

まず、本論文で考える知能制御の概念について定義する。一般的に知能とは、(1)経験から学習する能力、そこで得た知識を記憶する能力、(2)新しい状況に適応し、対処する能力、(3)概念レベルで思考し、推論により問題を解く能力、と考えられている。この能力は決して特別なものではなく、人間は日常の行動において利用している。その意味では、Rasmussenが人間の行動を表現したSRKモデルが知能制御を考えるうえで参考になる。Rasmussenは、人間の行動がFig.2に示すようにスキル(Skill)ベース、ルール(Rule)

ベース、知識(Knowledge)ベースの3つのレベルの認知活動によって表現できると考えた。

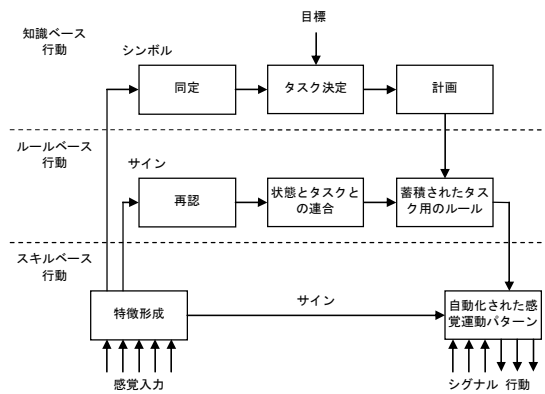


Fig. 2 SRK モデル⁴⁾

スキルベース行動は、自動化された感覚運動パターンによる行動である。すなわち人間の非常に熟練した定型作業における無意識でなめらかな行動を表す。感覚情報は、システムの振舞いを示す定量的“シグナル”として扱われ、これと意図した状態との誤差信号に応じた行動が出力される。

ルールベース行動は、過去の経験や教育によって獲得された規則や手順を意識的に適用する行動である。このような経験則は、過去の成功の経験に基づいて活性化されるように構造化されており、感覚情報のパターンとの照合によって選択される。ルールベース行動における感覚情報は“サイン”として知覚される。サインとは経験によってある行動と対応づけられたシステムの状態を示す情報のパターンである。

知識ベース行動は、未知あるいは不慣れな状況において、経験から獲得された手順や規則が適用できない場合に、より高次の概念レベルにおいて目標駆動的に決定される行動である。この行動では、メンタルモデルによって表された知識を用いて、システムの機能的特性からその状態を認識する。次に、認識された状態と心理的目標との手段-目標解析によって目標達成に必要な行動が計画、実行される。これはルールベース行動のいまだ学習されていないタスク用のルールを書き換える。計画は、その実行結果からフィードバックに基づいて修正されることもある。知識ベースの行動における感覚情報は、システムの機能的特性を表す選択的抽象概念(selective abstraction)である“シンボル”として知覚され、システムの状態の認識や予測のための推論や計算に用いられる。

3. 知能制御の方法

本論文で提案する知能制御は、SRK モデルに基づいてスキルベ

ース、ルールベース、知識ベースの3階層からなる。同様に、3階層からなる階層制御をALSSの水再生システムへ適用した例には、3T知能制御⁵⁾がある。次に我々の考える知能制御について示す。

スキルベース制御

スキルベース制御の適用の基本単位は1つ1つのセンサーとアクチュエータである。スキルベース制御では、従来の制御理論で用いられているフィードバック制御を利用する。

ルールベース制御

ルールベース制御の適用の基本単位はシステムを構成する機能である。ルールベース制御は機能ごとに意思決定を行う分散制御である。ルールベース制御ではそれぞれの学習主体(以下ではこれをエージェントと呼ぶ)が協力して独立に意思決定を行う。それぞれのエージェントは協調しながら強化学習を利用して on-line で自律的に制御規則を獲得する。つまりルールベース制御はマルチエージェントシステムである。その制御規則はサイン(状態と行動の組み合わせ)に対応して整理される。これがSRKモデルのサインに相当する。さらに、それぞれのエージェントが知識ベース制御のシンボルに相当する。

強化学習のアルゴリズムには、Q-Learning を利用する。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、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、すべての状態 x と行動 a に対するQ値の集まりを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値は次式のよう更新される。

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

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。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。

状態空間の設計

Hirosaki は、状態 x_t の集まりである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した⁶⁾。危険レベルとは、モジュール内の物質質量やタンク内の物質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。ここでは、危険レベルを Table 1 に示すように 7 段階で表現し、正の値をとると目標より物質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質量が少なすぎることを意味する。すなわち危険レベルの絶対値が大きくなるほどその状態は危険な状態となる。状態は、エージェントとした装置が計測するモジュールやタンクの危険レベルの組合せを用いて定義される。

Table 1 危険レベルの定義

物質質量		危険レベル
以上	以下	
最大値		3
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
	最小値	-3

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、エージェントが観測しているモジュールやタンクが N ある場合、時刻 t における報酬 r_t は次式のようになる。

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

行動選択手法

強化学習において、ランダムな行動選択手法を用いると、収束までにかかる時間ステップが非常に大きくなる。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。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を選択する。比の計算には、次式に示すようなボルツマン分布を用いる。

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

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。

知識ベース制御

知識ベース制御の適用の基本単位はシンボル(ルールベース制御

のエージェント)の因果関係で表わされたメンタルモデルである。つまり知識ベース制御はメンタルモデルを基にしたシングルエージェントシステムである。 O_2 と CO_2 の循環に限った ALSS の機能構造は Fig. 3 のように表せる。この sub-function の属性の因果関係を表したメンタルモデルが Fig. 4 である。これを用いてファジィ推論により目標達成に必要な行動計画が概念レベルで立案される。知識ベース制御は、ルールベース制御のいまだ学習されていないタスク用のルールを書き換える。

さて、植物栽培スケジュールが変更になった場合の廃棄物処理装置の運用計画の立案を例にして知識ベース制御について示す。Fig.4 のメンタルモデルを基に Fig.3 の ALSS の機能構造を利用して目標駆動的に計画を立案する。Fig. 3 において有機物の生産量が低下すると、食料の生産量が下がる(①)。食料の生産量が下がると資源の再生量が減る(②)。つまり O_2 の供給量、 CO_2 の再生量、廃棄物の発生量が減る(③)。さらに廃棄物量が減ると廃棄物処理量が減って、 O_2 の使用量と CO_2 の発生量が減る(④)。この現象には、栽培期間という時差が伴う。このような状況に対処するための廃棄物処理装置の運用計画を人間の代謝量とバランスをとりながら立案する必要がある。このとき式(4)に示すファジィ制御規則を用いて廃棄物処理装置の運用を決める。式(4)のファジィ制御規則を Table 2 に示す。これは設計者により与えられる。このとき O_2 タンクと CO_2 タンクのそれぞれの危険度 $risk$ と廃棄物処理装置のシーケンス開始確率 p のファジィ変数はそれぞれ Fig. 5 と Fig. 6 のように設定された。これらの設定は Fig.4 から推論される定性的な規則を与えたものである。 p の非ファジィ化(ファジィ量から実際の値に戻すこと)には重心法を用いる。これによりスケジュールはある時刻 t のシーケンス開始確率として表現される。

$$R_i: \text{If } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \text{ and } x_3 > W3_W \text{ then } p \text{ is } B_i, \quad (4)$$

$i = 1, 2, \dots, 13$

x_1 : O_2 タンク内の O_2 量

x_2 : CO_2 タンク内の CO_2 量,

x_3 : 廃棄物タンク内の廃棄物量

p : 廃棄物処理装置のシーケンス開始確率

$W3_W$: 1 バッチ当たりの廃棄物量

A_{i1} : $X_1(x_1$ の値の集合)で定義されるファジィ集合

A_{i2} : $X_2(x_2$ の値の集合)で定義されるファジィ集合

B_i : $P(p$ の値の集合)で定義されるファジィ集合

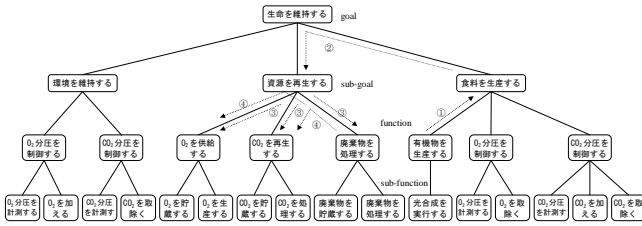


Fig. 3 ALSSの機能構造
(O₂とCO₂の循環に関する機能のみに限定)

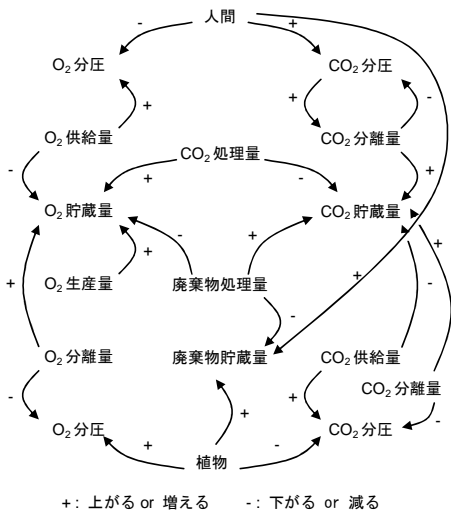


Fig. 4 因果関係のメンタルモデル

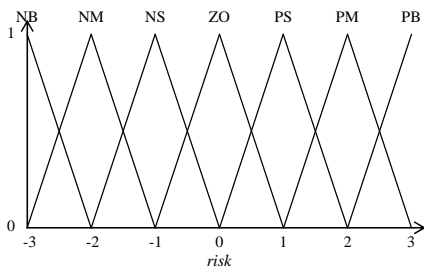


Fig. 5 危険度のファジイ変数

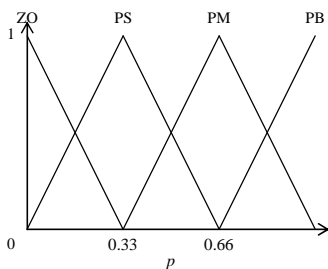


Fig. 6 廃棄物処理装置のシーケンス開始確率のファジイ変数

Table 2 ファジイ制御規則

p のファジイラベル	CO ₂ タンク A ₂						
	NB(-3)	NM(-2)	NS(-1)	ZO(0)	PS(1)	PM(2)	PB(3)
O ₂ タンク	NB(-3)						
	NM(-2)						
	NS(-1)						
	ZO(0)	PS	ZO				
A ₁	PS(1)	PM	PS	ZO			
	PM(2)	PB	PM	PS	ZO		
	PB(3)	PB	PB	PM	PS		

NB : Negative Big, NM : Negative Medium, NS : Negative Small, ZO : Zero, PS : Positive Small, PM : Positive Medium, PB : Positive Big
ファジイ集合のラベルは()内の危険度と対応している

4. 結果

シミュレーションモデル

例題として取り上げるALSSの物質循環システムをFig. 7に示す。ここでは、O₂とCO₂の循環に関係した物質のみがモデル化されている。

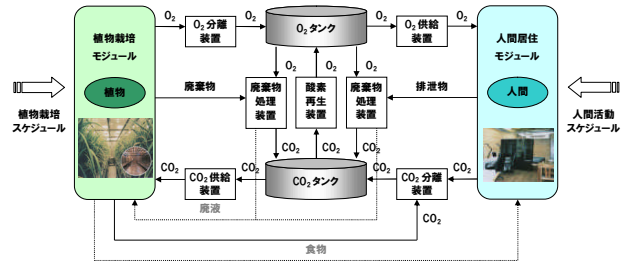


Fig. 7 ALSSの物質循環システム

(O₂とCO₂の循環に関する物質のみに限定)

シミュレーション設定

人間1人が居住し、植物を利用した食料生産により人間に食料が供給される。食料生産のための作物には大豆が使われる。今回のシミュレーションでは、知能制御の有効性を示す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735g/day、栽培面積は120m²である。植物は16ステージに分けられ、明期12時間、暗期12時間のシーケンス栽培で栽培された。食料は予備として5日分が余分に備蓄されている。O₂分離装置、CO₂供給装置、O₂供給装置、CO₂分離装置、廃棄物処理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。O₂タンク容量、CO₂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150m³、CO₂濃度の設定は400ppm、O₂濃度の設定は20.96%、および植物栽培空間の容積は269m³、CO₂濃度の設定は700ppm、O₂濃度の設定は20.93%である。

このシミュレーションではFig. 7に示した装置のうちO₂分離装置、CO₂供給装置、O₂供給装置、CO₂分離装置がルールベース制御層で、植物栽培モジュールの廃棄物処理装置が知識ベース制御層で、植物栽培モジュールの廃棄物処理装置が知識ベース制御

層でその運用が判断される。その他の O₂ 再生装置と人間居住モジュールの廃棄物処理装置は事前に決められたスケジュールで運用される。知能制御の適用は植物のシークエンス栽培が 80 日で完成した後の閉鎖系モードで開始され、第 200 日までの 120 日間行われた。このときのエージェントの強化学習のパラメータの値は $\alpha=0.1$ 、 $\gamma=0.99$ 、 $T=0.1$ である。

シミュレーション結果

シミュレーション結果では、ルールベース制御のみで対処できる場合と知識ベース制御を利用して対処する必要がある場合の例を示す。ルールベース制御のみで対処できる場合として、CO₂ 分離装置の能力が低下した場合の例を挙げる。環境が変化するとき学習によってエージェントが新しい環境に適応したかどうかを on/off 切替頻度と報酬獲得頻度の平均値から解析する。知識ベース制御を利用する必要がある場合として植物栽培スケジュールが変更された場合の例を挙げる。ここで知識ベース制御を用いない場合の廃棄物処理装置は事前に決められたスケジュールで 1 日 1 回運転される。植物栽培ステージが 1 ステージずれることにより O₂、CO₂、廃棄物のバランスがくずれ、それらを考慮した廃棄物処理装置の運用計画の立案が知識ベース制御を加えることで可能になるかどうかを解析する。

CO₂ 分離装置の能力が低下した場合

Table 3 に CO₂ 分離装置の能力が第 141 日以降に 50% 低下した場合の能力低下前と能力低下後の CO₂ 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. 8 に同様の場合の CO₂ 分離装置の切替頻度の変動を示す。第 141 日以降に CO₂ 分離装置の能力が 50% 低下したため CO₂ 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 1 日(24step に相当)以内に新しい環境に適応した結果といえる。

植物栽培スケジュールが変更された場合

栽培ステージ 5(第 101 日から第 180 日の栽培)の作物が栽培されず栽培ステージ 6(第 106 日から第 185 日の栽培)で 2 ステージ分が同時に栽培された場合のバイオマスの変化を Fig.9 に示す。比較のために植物栽培スケジュールの変更がない場合も示す。栽培スケジュールの変更があった場合は第 101 日から第 180 日においてバイオマスが多少減少している。さらに 185 日には 2 ステージ分が収穫されている。栽培ステージ 5 の作物が 5 日遅れで栽培されたためである。これにより光合成による O₂ の生産量と CO₂ の消費量が減る。そのため Fig.10 では知識ベース制御なしの場合、O₂ 量がタンク下限を下回っている。知識ベース制御ありの場合、O₂ タンク容量の範囲内で運用されている。これは O₂ タンクの O₂ 量、CO₂ タンクの CO₂ 量、廃棄物タンクの廃棄物量の長期的な変動が考慮されて廃棄物処理装置の運用計画が立案されたためである。CO₂ タンクの場合は、Fig.11 より知識ベース制御なし、ありともに CO₂ タンク容量の範囲内で運用されている。このときのタンク内の物質変動の振幅比較を Table 4 に示す。知能ベース制御の利用により振幅を O₂ タンクの場合には 5439g から 4568g に、CO₂ タンクの場合には 5062g から 4565g に抑えることができています。

5. まとめ

CO₂ 分離装置の能力が 50% 低下した場合の結果から、強化学習を利用したルールベース制御により環境が変化した場合に新たな運用方法を獲得できたと言える。植物栽培スケジュールが変更された場合の結果から、ファジィ推論を利用した知識ベース制御をルールベース制御に加えることにより、未知の環境における廃棄物処理装置の運用計画の立案が目標駆動型で行われた効果が確認されたと言える。データ駆動型のルールベース制御と目標駆動型の知識ベース制御を階層的に組み合わせることで、短期的な環境の変化に対応するだけでなく、未知の環境へ対応した長期的な戦略の立案もできるようになった一例を示した。今後、水循環への本手法の適用やその他の様々な事例についても検討していきたい。

引用文献

1. Crew and Thermal Systems Division : Requirements Definition and Design Conditions, CTSD-ADV-245 REV A, 1998.
2. Stephen J. Hoffman and David I. Kaplan : Human Exploration of Mars : The Reference Mission of the NASA Mars Exploration Study

Term, NASA Special Publication 6107, 1997.

3. 古田一雄, プロセス認知工学, 海文堂, 1998.
4. Jens Rasmussen, Skill, Rule, and Knowledge; Signals, Sifns, and Symbols, and Other Distinctions in Human Performance Model, IEEE Transactions on system, man, and cybernetics, Vol. SMC-13, 1983.
5. Pete Bonasso, David Kortenkamp and Carroll Thronesbery, Intelligent Control of Water Recovery System : Three years in the Trenches, AI Magazine Vol. 24, No. 1, 2003.
6. Tomofumi Hirosaki et al. : Application on Multi-agent Reinforcement Learning to CELSS Material Circulation Control, PAIS2001, 2001.
7. Hiroyuki Miyajima et al. : Application of Multi-Agent Reinforcement Learning to RLSS Material Circulation Control System, SAE Technical Paper Series 2004-01-2437, 2004.

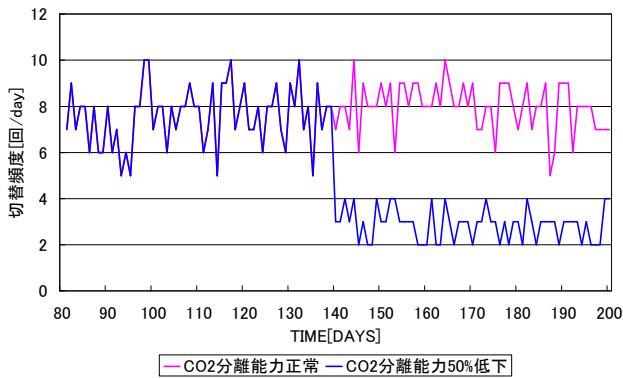


Fig. 8 CO₂分離能力 50%低下した場合の CO₂分離装置の切替頻度の変化

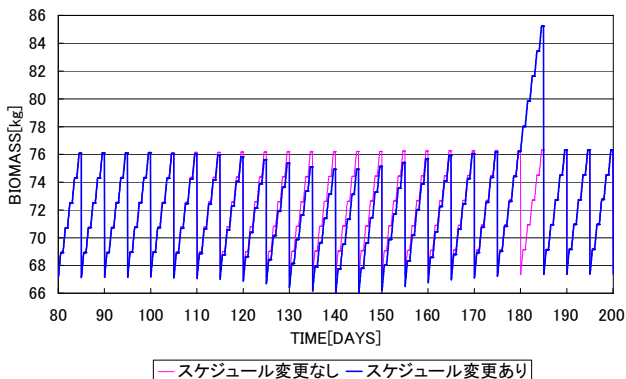


Fig. 9 バイオマスの変化

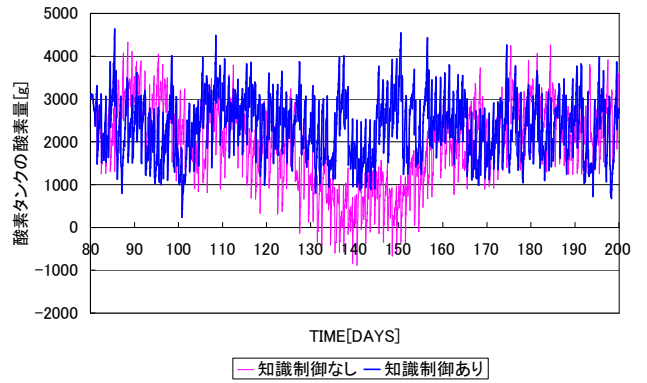


Fig. 10 酸素タンクの酸素量の変化

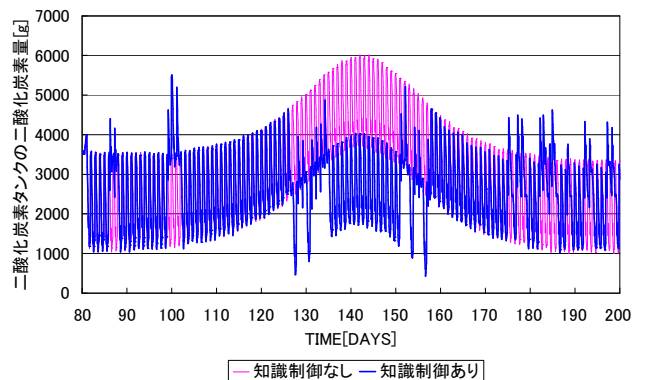


Fig.11 二酸化炭素タンクの二酸化炭素量の変化

Table 3 CO₂分離能力が50%低下した場合の CO₂分離装置の平均切替頻度と平均報酬獲得頻度

期間	CO ₂ 分離能力正常		CO ₂ 分離能力低下	
	切替頻度	報酬獲得頻度	切替頻度	報酬獲得頻度
[day]	[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 4 タンク内の物質質量変動の振幅比較

知能ベース制御の利用	O ₂ タンク			CO ₂ タンク		
	最大値[g]	最小値[g]	振幅[g]	最大値[g]	最小値[g]	振幅[g]
なし	4557	-883	5439	6009	947	5062
あり	4641	73	4568	5510	945	4565