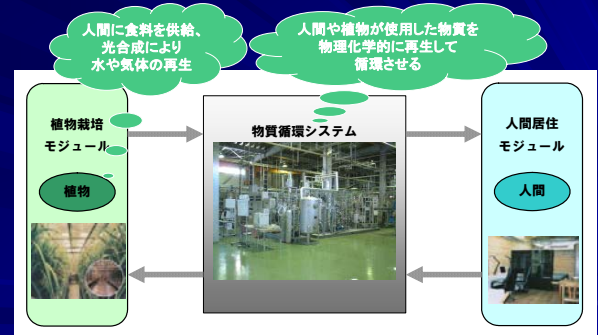


再生型生命維持システムの運用管理における情報の取り扱いに関する考察

宮嶋宏行 東京女学館大学
 広崎朋史 日本大学
 石川芳男 日本大学

再生型生命維持システム



RLSS : Regenerative Life Support System

研究の背景(1/3)

- 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。
- 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。
- 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専念するためには、運用管理の一層の**自動化**が期待される。

研究の背景(2/3)

自動化するとき何が問題か？

- 自動化を考えた場合、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**などの**環境の変化**に**すべて対応することは難しい**。
- 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求められる。

研究の背景(3/3) どのように解決するか？

集中管理系 運用管理	大域的な情報 を利用	効率的な動作の計画が可能 オフライン計画が一般的 大規模なシステムに向かない 変動や誤差に弱い
自立分散系 運用管理	局所的な情報 を利用	計算時間の短縮が可能 → オンライン計画に適用可能 → 変動や誤差に強い 効率的な動作の計画が困難 時には解にいきつかない

マルチエージェント
システム(MAS)

研究の目的

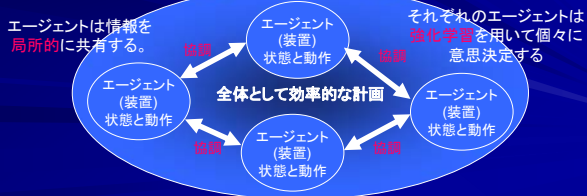
- MASを用いたRLSS物質循環システムの制御系の設計についてシミュレーションを利用して検討する。

環境の変化への対応
 解にいきつかない場合の回避
 効率的な動作の計画

MASを利用したRLSS物質循環システム制御系の設計

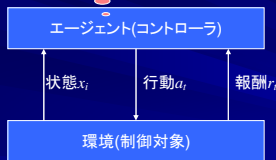
■ システム形態

- 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**とする。
- 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**効率的な計画**が立案される。



強化学習(Q-Learningを利用)

学習



Q-Table

	a_1	a_2	...	a_n
x_1	0	0.1		0
x_2	-0.1	0		0.2
x_3	0.1	0		0

Q-Learning

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

割引率 γ
 学習率 α

設計のポイント

■ 状態の設計

- オンラインでの利用を考えると、少ない学習回数で収束させる必要があり、そのためには**状態数を少なく**する必要がある。

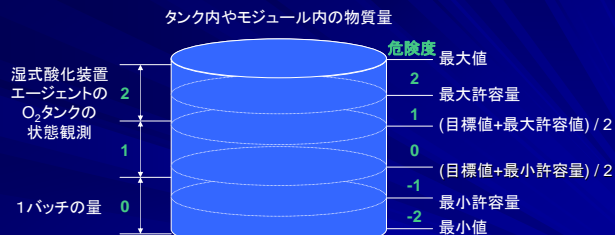
■ 報酬の設計

- システムの効率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**するような報酬を設計する必要がある。

■ 行動選択

- 収束までの時間が短い方が望ましいが、いったん収束した後でも**新たな環境に適応**できる必要がある。

状態の設計



状態は、エージェントが観測する要素(タンクやモジュール)と要素の**危険度の組**

$$\text{O}_2\text{タンク 危険度 5レベル} \times \text{CO}_2\text{タンク 危険度 5レベル} \rightarrow \text{状態数 25}$$

報酬の設計

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

エージェントが観測する要素の**危険度の改善度**

隣接する2つの要素の危険度を見ている

行動選択

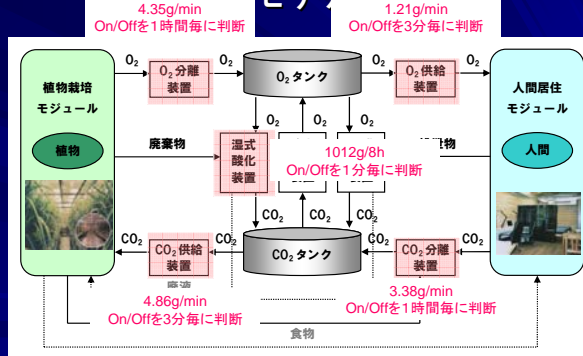
$$\pi(x, a) = \frac{e^{Q(x,a)/T}}{\sum_{a_t \in A} e^{Q(x,a_t)/T}}$$

ボルツマン分布

Tが大きいほどランダム行動、Tが0に近いほど最適な行動が選択される

行動は装置のon/offの2通り

RLSSモデル(1/2)



O₂とCO₂の循環に関する基本的要素に限定したモデル

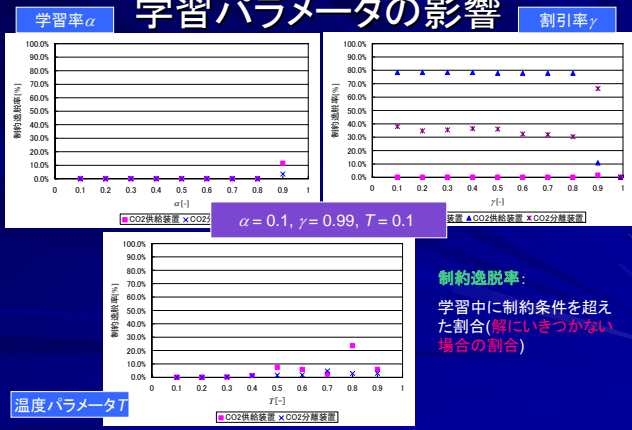
RLSSモデル(2/2)

- 植物: 植物の成長はロジステック曲線でモデル化
- 人間: 人間の代謝は三角関数でモデル化
- CO₂分離装置: 分離性能は線形近似でモデル化
- O₂分離装置: //
- CO₂供給装置: 差圧によって移送するモデル
- O₂供給装置: //
- 湿式酸化装置: O₂量が1バッチに必要な量あるときのみ動作onを1分毎に判断(いったんonになると8時間後にoff)

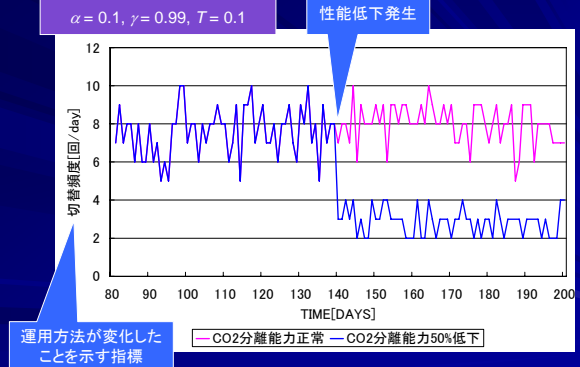
シミュレーション設定

- 人間 1人
- 栽培植物 大豆1種類のみ
 - 栽培量は735g/day(乾燥質量)、栽培は8ステージに分けられ、明期12時間、暗期12時間でシーケンシャル栽培される。
- 環境設定
 - 人間居住空間の設定: CO₂濃度: 400ppm
O₂濃度: 20.96%
 - 植物栽培空間の設定: CO₂濃度: 700ppm
O₂濃度: 20.93%
- エージェントの学習
 - 学習は、植物のシーケンシャル栽培が完成した後、81日から200日までの120日間行われる。

学習パラメータの影響



CO₂分離装置の能力が低下した場合



CO₂分離装置の能力が低下した場合

$\alpha = 0.1, \gamma = 0.99, T = 0.1$

運用方法が変化したことを示す指標

どれぐらい学習しているかの指標

期間	CO ₂ 分離能力正常		CO ₂ 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
日	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

141日目に性能低下発生

まとめ

- 学習パラメータを適切に調整することにより個々のエージェントを協調させて物質循環の制御を達成できた。
 - 解にいきつかない場合の回避
- CO₂分離装置の能力が50%低下した場合、学習により新たな運用方法を獲得できた。これはいったん収束した後でも、行動選択で用いたボルツマン分布が未知の環境に対して有効に機能していることを示している。
 - 学習により環境の変化に対応

今回想定した範囲では、
状態の設計、報酬の設計、行動選択が適切であったといえる。

報酬の設計については、効率的な動作の計画のために
さらに検討が必要である。