2002年日本オペレーションズ・リサーチ学会
春季研究発表会

強化学習によるDynamic Power Management Systemの実装

岡村寛之 (01013754)†, 石倉武‡, 土肥正 (01307065)†
† 広島大学大学院工学研究科情報工学専攻
‡ 広島大学工学部第二類（電気系）

1. はじめに

近年，ハードウェアの小型化に伴いノートパソコンなどを携帯する機会が多くなり，バッテリ等の限られた電力容量の中でいかに最大の性能を維持しながら，長時間稼働を実現するかが重要な関心事となっている．このような状況の中，DPMS (Dynamic Power Management System) と呼ばれる省電力技術が注目されている．DPMS とはコンピュータシステムやその他の電子的なデバイスにおける省電力技術の総称であり，具体的な機能としては，CPU に処理率可変機能を持たせるLongRun, SpeedStep や自動スリープ機能などがある．本稿では特に自動スリープ機能について考える．

自動スリープ機能とは待機状態が長時間経過した場合に，電力消費を抑えるためにシステムを電力消費の少ないスリープ状態に移行させる機能である．自動スリープ機能では，待機状態からスリープ状態に移行するタイミングが非常に重要である．何故ならば，スリープ状態の間にジョブが到着するとシステムはジョブを処理するため稼動状態に移行する．しかしながら，その切り換えに要する電力は非常に大きいため，待機状態（システムがジョブ処理を行っていない状態）において常にスリープを実行する方策では，スリープから起動するための電力を余分に消費してしまう可能性がある．

上記の問題に対して，文献 [1,2] ではスリープ機能を持つシステムを確率モデルによって表現することで最適なスリープへの移行タイミングが存在するための条件を与えている．しかしながら，文献 [1,2] に基づいて最適なスリープへ移行するタイミングを決定しようとする場合，ジョブの到着間隔や処理時間に関するデータを用いて予めモデル内で用いる確率分布の特定およびそのパラメータ推定を行う必要がある．特に確率分布の特定はデータが持つ統計的な特徴を考慮した上で経験的にいくつかの候補を選ぶため，様々な環境で利用されるシステムにおいては非効率的な手法であると考えられる．

そこで，本稿では日々刻々と変化するジョブの到着率や処理負荷をどのように表現するかという問題に対して，実際のデータから確率分布等を考慮せずにスリープ状態に移行する最適なタイミングを導出する手法について考察を行う．より具体的には，SMDP（セミマルコフ決定過程）によってスリープ機能を確率的な環境の下でモデル化する．その後，強化学習 [3] とよばれる手法を用いて，DPMS の実装を行う．

2. SMDP によるモデル化

システムを以下の 3 つの状態に分類する．

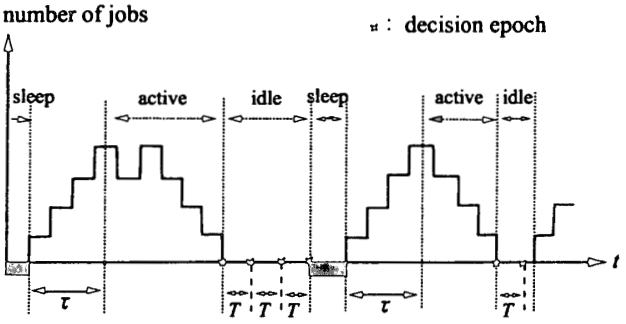

図 1: システムの振る舞い

スリープ状態 (sleep): システムの休止状態であり，ジョブが到着すると処理に対する準備期間 τ を経て起動状態へ移行する．準備期間においては単位時間当たり P_2 の電力を消費し，スリープ状態においての電力消費は 0 とする．

待機状態 (idle): ジョブの到着に対して待機している状態であり，ジョブが到着すると直ちに稼動状態へ移行する．待機状態においては単位時間当たり P_1 の電力消費とする．

稼動状態 (active): システムが実際にジョブに対する処理を行っている状態であり，システム内のジョブをすべて処理するまで継続する．単位時間当たりの電力消費は待機状態と同じである．

ジョブは一般の再生過程に従って到着し，到着した各々のジョブに対する処理時間は独立で同一の分布に従うものとする．このとき，システム内のジョブの振る舞いは $GI/GI/1$ 待ち行列を形成する (図 1 参照)．一般的な待ち行列の理論から，稼動時間と待機時間は直前の準備期間に依存することが知られている．そのため本稿では，準備期間を伴った場合の稼動時間及び待機時間に関する確率分布を $G_\tau(t)$ 及び $F_\tau(t)$，準備期間を伴わない場合の稼働時間及び待機時間に関する確率分布を $G(t)$ 及び $F(t)$ として表す．

PM(パワーマネージャー) は，待機状態において T 時間経過する毎に消費電力をおさえるためにスリープモードに移行するか否かの決定を行う．前述したように待機状態の準備期間を伴った場合と伴わない場合とで異なるため，準備期間を伴なう場合の待機時間を T 毎に区切った状態を $I'_0, I'_1, ..., I'_k$，伴わない場合を $I_0, I_1, ..., I_k$ によって表す．PM による決定が

可能なシステムの集合を $s=\{I_0,I_1,...,I_k,I_0',I_1',...I_k'\}$, PM が選択可能な決定の集合を $a=\{S,I\}$ とする．ここで，記号 S はスリープ状態に移行する決定で，記号 I は待機状態を継続する決定を意味する．

省電力を実現するための評価の規範として無限計画期間における総期待割引消費電力を用いる．無限計画期間における総期待割引消費電力は，割引率を α とするとき，次の式で定義される評価規範である．

$$V=E\left[\int_0^\infty e^{-\alpha t}P(t)dt\right].$$

ここで V は無限計画期間における総期待割引消費電力であり，$P(t)$ は時刻 t における単位時間当たりの消費電力を表す確率過程である．

いま，$V(s)$ を状態 s での総期待割引消費電力とし，$Q(s,a)$ を状態 s で行動 a を選択後，最適方策をとり続ける時の期待消費電力とする．このとき，以下の DP アルゴリズムに基ずいて最適方策を導出することができる．

Step 0: $s=I_0$, $n=0$, $V^0(s)=0$.

Step 1: $V^n(s)\leftarrow\min\{Q^n(s,S),Q^n(s,I)\}$.

Step 2: $n\leftarrow n+1$, Step 1 へ.

また，$Q^n(s,a)$ は以下のように定義される．

$$\begin{aligned}Q^n(s,S) &= \int_0^\infty e^{-\alpha t}dF_s(t)\Big\{\frac{P_2}{\alpha}(1-e^{-\alpha\tau})\\ &\quad+\frac{P_1}{\alpha}e^{-\alpha\tau}(1-G_\tau^*(\alpha))\\ &\quad+e^{-\alpha\tau}V^{n-1}(I_0')e^{-\alpha\tau}G_\tau^*(\alpha)\Big\}, \qquad (1)\end{aligned}$$

$$\begin{aligned}Q^n(s,I) &= \int_0^T\Big\{\frac{P_1}{\alpha}(1-e^{-\alpha t})\\ &\quad+\frac{P_1}{\alpha}e^{-\alpha t}(1-G^*(\alpha))\\ &\quad+e^{-\alpha t}G^*(\alpha)V^{n-1}(I_0)\Big\}dF_s(t)\\ &\quad+\int_T^\infty\Big\{\frac{P_1}{\alpha}(1-e^{-\alpha T})\\ &\quad+e^{-\alpha T}V^{n-1}(I_{k+1})\Big\}dF_s(t), \qquad (2)\end{aligned}$$

ここで

$$G_\tau^*(\alpha)=\int_0^\infty e^{-\alpha x}dG_\tau(x),$$

$$G^*(\alpha)=\int_0^\infty e^{-\alpha x}dG_(x),$$

$$F_s(t)=\frac{F_{s-1}(T+t)-F_{s-1}(T)}{1-F_{s-1}(T)}.$$

$F_s(t)$ の添え字 s は状態を表し，$s-1$ は一つ前の状態を表す．例えば $s=I_2$ の時 $s-1=I_1$ であり，$s=I_2'$ の時は $s-1=I_1'$ である．また $F_{I_0}(t)=F(t)$, $F_{I_0'}(t)=F_\tau(t)$ とする．

実際に本稿で示した DP アルゴリズムを用いるためには，待機時間や稼働時間に対する確率分布関数の詳しい表現が必要とされる．しかしながら，一般にこれらの分布を陽に表現することはジョブの到着や処理に関する分布関数が特定できたとしても非常に困難である．そこで，次節では強化学習を用いた最適な決定手法を示す．

3. 強化学習による DPMS の実装

強化学習とは，試行錯誤を通じて環境に適応する学習制御の枠組みである．教師付き学習と異なり，状態入力に対する正しい行動出力を明示的に示す教師が存在しない．その代わりに報酬というスカラーの情報を手がかりにして，環境との試行錯誤的な相互作用の繰り返しを通じて $Q(s,a)$ を推定するアルゴリズムである．2 節で導出した最適性方程式に対する強化学習のアルゴリズムを示す．

Step 0: PM は待機状態 s_n を観測する．

Step 1: PM は任意の行動選択方法に従って行動 a を実行する．

Step 2: 環境から期待割引消費電力 $\int_0^t e^{-\alpha t}P(t)dt$ を受ける．

Step 3: 以下の更新式により Q 値を更新する．

$$\begin{aligned}Q(s_n,a)\longleftarrow Q(s_n,a)+\beta\Big[&\int_0^t e^{-\alpha t}P(t)dt+\\ &e^{-\alpha t}\max_a Q(s_{n+1},a)-Q(s_n,a)\Big].\end{aligned}$$

ただし β は学習率である．

Step 4: $n\leftarrow n+1$, Step 1 へ.

4. 今後の課題

本稿では強化学習により DPMS を実装するための SMDP に基ずいたモデル化を行った．また，導出した最適性方程式に対する強化学習のアルゴリズムを示した．今後は実データを基に強化学習による DPMS の実装及びその性能評価を行う予定である．

参考文献

[1] 岡村寛之, 土肥正, 尾崎俊治, コンピュータシステムの自動スリープ機能による省電力効果 1 –再生過程によるモデル化, 情報処理学会論文誌, Vol. 39, No. 6, pp. 1855–1869 (1998).

[2] 岡村寛之, 土肥正, 尾崎俊治, コンピュータシステムの自動スリープ機能による省電力効果 2 –待ち行列モデル, 情報処理学会論文誌, Vol. 40, No. 3, pp. 1027–1040 (1999).

[3] R. S. Sutton and A. G. Barto (三上貞芳, 皆川雅章 (共訳)), 強化学習 , 森北出版 (2000).