The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

電力ダム操作における強化学習型シンボルグラウンディングによる意思決定支援に関する検討

Toward Supporting Decision Making for Flood Control in a Hydroelectric Dam by Reinforcement Learning and Symbol Grounding

田中　友紀子 *1*2　平岡　拓也 *1*2　大西　貴士 *1*2
Yukiko TANAKA　Takuya HIRAOKA　Takashi ONISHI
鶴岡　慶雅 *2*3　滝野　晶平 *4　松本　陽介 *5
Yoshimasa TSURUOKA　Shohei TAKINO　Yosuke MATSUMOTO

*1NEC セキュリティ研究所
NEC Security Research Laboratories
*2産業技術総合研究所 人工知能研究センター
National Institute of Advanced Industrial Science and Technology, Artificial Intelligence Research Center
*3東京大学 The University of Tokyo
*4東京電力ホールディングス株式会社 Tokyo Electric Power Company Holdings, Inc.
*5東電設計株式会社 Tokyo Electric Power Services Co., Ltd.

We present a hierarchical planner for dynamic systems based on reinforcement learning and symbol grounding. In experiments, we evaluate our planner using a hydroelectric dam simulator and demonstrate the potential effectiveness of our approach in supporting decision making in dam flood control.

1. はじめに

近年、人間の技能を超える人工知能が登場しており、その発達が目覚ましい。囲碁の領域では、人間の知識を使うことなく人間の技能を超える人工知能、AlphaGo Zero が誕生している [Silver 17]。産業分野においては、プラント制御や生産管理などの現場で熟練者の減少に伴う知識伝承の課題があり、現場で働く人間の支援に向けた人工知能活用の期待が高まっている。

人工知能の活用が期待できる産業分野のひとつとして、ダムの放流操作が挙げられる。ダム操作の中でも特に、洪水時 *1 のダム放流操作は操作員にとって難しい操作である。

発電を目的とする電力ダムでは通常、ダムに貯留した水を取水し、下流の発電所に導水することで発電を実施する。電力ダムの操作では河川の安全確保を第一優先とする。電力ダムの上流河川でのゲリラ豪雨や台風通過によって雨量が増加し洪水に至る可能性がある場合、流域の安全を考慮し、発電を停止し、洪水に至る前までに予め貯留した水を放流することでダム水位の低下を完了することがダム操作規程で定められている。水位低下のためには放流量を増やす必要があるが、放流量の急激な増加はダム下流域の河川氾濫を引き起こすため、河川の状況を監視しながら徐々に放流量を増やす必要がある。

洪水時や洪水時前後のダム放流操作はダム操作規程に定められているが、ダム操作員はダムへの流入量を予測し、洪水時までに放流すべき水量や水位低下に必要な時間を見積り、ダムの放流計画を立案することが求められる。現状では、熟練のダム操作員が気象状況や河川状況に合わせて、ダムの放流計画を逐次見直しながらダム放流操作をおこなっているが、時間的な制約の中、流域の安全確保が至上命題であり、操作員への心理的・肉体的な負担は非常に大きい。そこで本研究では、ダム操作員の負担軽減を目的とし、人工知能を活用した意思決定支援の有用性を検討する。

連絡先: 田中 友紀子，NEC セキュリティ研究所，川崎市中原区下沼部 1753，y-tanaka@jz.jp.nec.com

*1　ダムの流入量がある閾値を超えることを洪水という

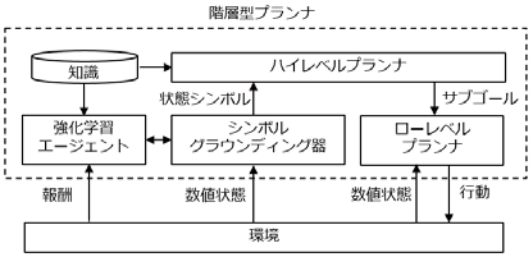

図 1: 強化学習と、シンボルを介したプランニングを融合した枠組み

ダム放流操作のような重要インフラ操作の場面では、責任所在の観点から最終的な意思決定は操作員が実施することが求められるため、人工知能が出した解を人間が理解できる必要がある。例えば、人工知能の分野で成果を挙げている人工知能 AlphaGo Zero [Silver 17] で用いられる手法は人間の技能を超えるという点で優れた手法だが、人工知能が出した解 (囲碁の場合は選んだ手) の根拠を人間が理解することが容易でない。

今回は、電力ダムのシミュレータを計算機上で構築し、人間が理解しやすい操作案を出すことが可能な枠組みである強化学習型のシンボルグラウンディングに基づく人工知能に、熟練者相当の操作を学ばせることを試みた。実験結果を評価した結果、有用性を確認することができたので報告する。

2. アプローチ

2.1 枠組み

本研究では、熟練者相当の操作を獲得でき、かつ人間が理解できる形で操作案を提示することを狙う枠組みを構築した (図 1 参照)。具体的には、強化学習とシンボルグラウンディングを含む階層型プランナの枠組みを構築した。

本研究での階層型プランナの大まかな処理の流れは、次の

通りである。まず、階層型プランナが、操作対象の電力ダムや上流河川などを含む環境から電力ダムへの流入量や水位などの数値状態を受け取る。次に、シンボルグラウンディング器が、各数値状態を状態シンボルに変換する。例えば、水位が 10m であるという数値状態を、「発電可能水位」という状態シンボルに変換する。その後、ハイレベルプランナが、状態シンボルの集合をもとに、人手で事前に設計した知識を用いたプランニングにより、現在達成すべき目標状態であるサブゴールを導く。本研究では目標水位の値がサブゴールに相当する。最後に、ローレベルプランナが、現在の数値状態とサブゴールをもとに現在取るべき行動を導き、環境に与える。本研究では放流量の値が行動に相当する。これらの処理を意思決定周期毎に繰り返す。

本研究の階層型プランナは、シンボルを介する。ダム放流操作では、最終的な操作決定と実行は操作員が担うものであるため、階層型プランナが導出した操作案を操作員が理解できることが求められる。シンボルを介する階層型プランナは、操作員が理解しやすい形で解を提示できる期待がある。

また、本枠組みでは、前述の階層型プランナの一連の処理に加え、シンボルグラウンディング器が担うシンボルグラウンディング、つまり数値状態と状態シンボルの対応づけに強化学習を用いる。強化学習を担う強化学習エージェントは、環境から得た報酬と数値状態をもとに、シンボルグラウンディングのふるまいを決めるグラウンディング関数のパラメータを学習する。強化学習は、熟練者の操作ログが豊富に用意されていない状況でも、報酬関数を定義することで、熟練者相当の操作を獲得できる期待がある。本研究では熟練者相当の操作を獲得するため、強化学習を用いる。

2.2 シンボルグラウンディング

階層型プランナ内のシンボルグラウンディング器が担うシンボルグラウンディングについて述べる。本研究でのシンボルグラウンディングは、階層型プランナが環境から得る複数の数値状態のうち各数値状態に対しそれぞれ 1 つの状態シンボルを対応づける。以降、1 つの数値状態から 1 つの状態シンボルへのグラウンディングを例に、シンボルグラウンディングの説明をする。

S をシンボルの集合を表す有限集合とし、数値状態を $x \in \mathbb{R}$、状態シンボルを $s \in S$ と定義する。本研究では、x から s への変換をシンボルグラウンディングとする。

シンボルグラウンディングの挙動を制御するパラメータ $\boldsymbol{\theta}$ を次のように定義する。

$$\boldsymbol{\theta} = (\mu_{s_1}, ..., \mu_{s_i}, ..., \mu_{s_{|S|}}, \sigma_{s_1}, ..., \sigma_{s_i}, ..., \sigma_{s_{|S|}}) \quad (1)$$

μ_{s_i}, σ_{s_i} はそれぞれ、$S = \{s_1, ..., s_i, ..., s_{|S|}\}$ としたときの、シンボル s_i に対応する平均と分散である。$\boldsymbol{\theta}$ を用いて、シンボルグラウンディングに使用するグラウンディング関数を、

$$\pi(s, \boldsymbol{\theta}|x) := \pi(s|x, \boldsymbol{\theta})P(\boldsymbol{\theta}) \quad (2)$$

と定義する。ここで、$\pi(s|x, \boldsymbol{\theta})$ は、各 s に対応づけたガウス分布 $N(x|\mu_s, \sigma_s)$ を用いて以下のように定義される。

$$\pi(s|x, \boldsymbol{\theta}) := \frac{N(x|\mu_s, \sigma_s)}{\sum_{s' \in S} N(x|\mu_{s'}, \sigma_{s'})} \quad (3)$$

$P(\boldsymbol{\theta})$ は、人間が理解しやすい形で x と s を対応づけるための $\boldsymbol{\theta}$ に関する事前分布 $P(\boldsymbol{\theta}) := N(\mu_s|\mu_m, 1)N(\sigma_s|\sigma_m, 1)$ である。$\pi(s, \boldsymbol{\theta}|x)$ は、人間が経験的に得る知識や操作手順書から学ぶ知識から大きく外れないようにすることが望ましい。電力ダムの操作はダム操作規程に従うため、本研究では予め、ダム操作規程の記述を参考に、$P(\boldsymbol{\theta})$ に用いる μ_m, σ_m を人手で作成した。人手で作成した μ_m, σ_m の例を表 1 に示す。人間がもつ前記知識を $P(\boldsymbol{\theta})$ に埋め込み、学習初期値と学習時の制約として考慮する。

表 1: 人手で作成した事前分布パラメータ μ_m, σ_m の例

No.	s	知識	(μ_m, σ_m)
1	通常時	流入量が 300m^3/s 未満ならば通常時	(150, 116.3)
2	洪水警戒時	流入量が 300m^3/s 以上 400m^3/s 未満ならば洪水警戒時	(350, 38.8)

なお、本研究では学習対象に含めないルールベースのグラウンディング関数も用意する。例えば、「発電可能水位」というシンボルの真偽は、電力ダムの諸元と水位の値によって一意に決まる。このように x から s への対応づけが一意に決まるものは、ルールベースのグラウンディング関数を使用した。

2.3 強化学習

本研究では、2.2 節で述べた $\pi(s, \boldsymbol{\theta}|x)$ を改善するために強化学習をおこなう。強化学習アルゴリズムは REINFORCE [Williams 92] を使用する。式 (4) を最大化する $\boldsymbol{\theta}$ を、式 (6) を用いた更新で求める。

$$\arg\max_{\boldsymbol{\theta}} E_{\pi\boldsymbol{\theta}}[\sum_{t=0} r_t] \quad (4)$$

$$\Delta\boldsymbol{\theta} = \alpha \left(\sum_{m=1}^{M} \sum_{t=1}^{T} r_t^m \nabla_{\boldsymbol{\theta}} \log \pi(s_t^m, \boldsymbol{\theta}|x_t^m) \right) \quad (5)$$

$$= \alpha \left(\sum_{m=1}^{M} \sum_{t=1}^{T} r_t^m \nabla_{\boldsymbol{\theta}} \log \pi(s_t^m|x_t^m, \boldsymbol{\theta}) + \sum_{m=1}^{m} \sum_{t=1}^{T} r_t^m \nabla_{\boldsymbol{\theta}} \log P(\boldsymbol{\theta}) \right) \quad (6)$$

ここで α, r_t^m, s_t^m と x_t^m はそれぞれ学習率、エピソード m 中の時点 t における報酬、状態シンボル、数値状態である。

ある時点 t において環境が強化学習エージェントに与える報酬 r_t は、単位時間当たりの発電放流量が多い程 、また規定違反が無い程 r_t が大きくなるよう、以下のように設計する。

$$r_t = Q_e/10 + RulePoint \quad (7)$$

ただし、

$$RulePoint = \begin{cases} 5.0 & (Flood \cap GateFullOpen) \\ -100(l - l_{min}) & (Flood \cap \neg GateFullOpen) \\ -l & (l > H_{HWL}) \\ 0.0 & (otherwise) \end{cases}$$

ここで、Q_e は発電放流量 (m^3/s) である。また、Q_{in} をダム流入量 (m^3/s)、Q_{flood} を洪水流量 (m^3/s) とし、$Flood := (Q_{in} > Q_{flood})$ である。l は電力ダム水位 (m)、l_{min} を Q_{in} に対応するゲート全開時 *2 の水位 (m) とし、$GateFullOpen := (l = l_{min})$ である。H_{HWL} は発電可能最大水位 (m) である。

*2 ゲート全開は水位低下完了と同義である

表 2: ハイレベルプランナで用いたルール例

No.	状態シンボルとサブゴール
1	「洪水警戒時」かつ「発電可能水位」ならば水位を「発電可能水位+1m」に設定
2	「洪水時」ならば水位を「最低水位」に設定

3. 実験

3.1 環境構築

本研究では、図 1 で示した枠組み内の環境を数値シミュレータで構築する。強化学習には多くの試行数が必要であり、実在の電力ダムを操作し、実施することは不可能である。そのため本研究では、電力ダムのシミュレータを計算機上で構築し、シミュレータ上で強化学習を高速に試行することを可能とした。

本シミュレータは、電力ダムの初期水位、単位時間当たりの流入量と放流量 *3 をもとに、数値計算で電力ダムの水位を算出することで水位変化をシミュレーションする。電力ダムのシミュレーションモデルは、水文シミュレータ CommonMP [CommonMP] を参考に自作した。電力ダムのシミュレーションモデルに用いるパラメータは、日本国内にある電力ダムの諸元を参照し構築した。

本シミュレータは電力ダムのシミュレーションモデルに加え、階層型プランナによる操作の試行錯誤を可能にするインターフェースを含む。具体的には、階層型プランナの操作決定周期である 10 分毎にシミュレータが階層型プランナへ、数値状態 (本研究では電力ダムへの流入量と流入量予測値、ダム水位) を渡し、階層型プランナがシミュレータへ、行動 (本研究では放流量) を渡し、シミュレータが 10 分間の数値シミュレーションを実行するという手順を繰り返す。

シミュレーションで使用する電力ダムの単位時間当たりの流入量は、実在の電力ダム流域で発生した複数の洪水パターン *4 を含む期間内の雨量データを用いて構築した 36 期間分の人工データを使用した。

3.2 ハイレベルプランナ

数値状態 $\boldsymbol{x} = (x_1, x_2, ...)$ からシンボルグラウンディングで得られた状態シンボルを $S_x = \{s_{x_1}, s_{x_2}, ...\}$ と定義する。また、G をシンボルの集合を表す有限集合とし、サブゴールを $g \in G$ と定義する。ハイレベルプランナは、S_x を受け取り、g を出力する。本研究のハイレベルプランナは、実在の電力ダムのダム操作規程に倣い人手で構築したルールに従い g を決定する。

本研究で構築したルールの例を表 2 に示す。本研究では 2 種類のルールセット、ruleA と ruleB を用意した。ruleA は現在の流入量と水位のシンボルグラウンディング用のルールセット、ruleB は現在の流入量と流入量予測値、水位のシンボルグラウンディング用のルールセットである。表 2 には ruleA のルールの一例を示す。

3.3 ローレベルプランナ

ローレベルプランナは、g と x を受け取り、行動 $\boldsymbol{a} \in \mathbb{R}^n$ を出力する。本研究ではゲート放流量 Q_g と発電放流量 Q_e の 2 つの値を行動とする。つまり、$\boldsymbol{a} = (Q_g, Q_e)$ である。

*3 本研究が対象とする電力ダムの放流量は、ゲート放流量 (発電に使用しない放流量) と発電放流量 (発電に使用する放流量) の 2 つがある。ここでの放流量は、ゲート放流量と発電放流量の合計値を表す

*4 前線性、台風性など

表 3: 評価結果

ルール	比較対象	平均発電時間 [h]
ruleA	handmade	53.39
ruleA	learned	68.15
ruleB	handmade	61.13
ruleB	learned	67.63

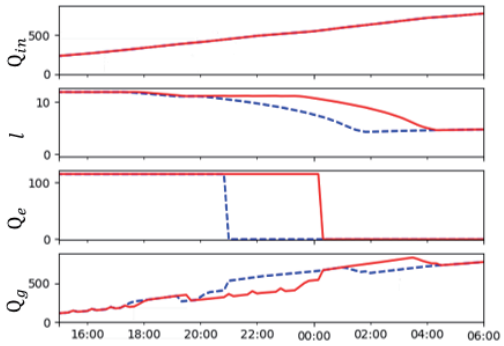

図 2: ダム放流操作案の例。点線が人手で作成したルールを用いて導出した操作案、実線が学習結果を用いて導出した操作案。縦軸はそれぞれ、Q_{in} はダム流入量 (m^3/s)、l はダム水位 (m)、Q_e は発電放流量 (m^3/s)、Q_g はゲート放流量 (m^3/s) である。

今回、g は目標水位と等価である。ローレベルプランナは目標水位 g と現在水位 x をもとに、水位上昇 (放流量減少)、水位低下 (放流量増加)、水位維持 (放流量維持)、の 3 つの操作から 1 つを選択する。選択した操作に従い、ルールベースの制御器でトータルの放流量、つまり Q_g と Q_e の合計値 $Q_g + Q_e$ を決定する。その後、Q_e の上限値を考慮した上で Q_e を最大化するように Q_e を設定し、残りの放流量を Q_g と設定する。本ルールベース制御器では、流域の安全を考慮し、急激な放流量の増減をしないための制約を含んでいる。

4. 評価

評価結果を表 3 に示す。今回は、学習データと別に用意した 9 期間分の評価データで検証し、各期間での発電時間を平均した値を評価結果とする。人手で作成したルールを使用した結果を handmade に示す。また、学習で得られたグラウンディング関数を用いた結果を learned に示す。

今回の評価データの範囲では、学習結果のほうが人手で作成したルールを用いる場合よりよい結果となった。具体的には、学習結果のほうが人手で作成したルールを用いる場合より発電時間を長くしていた。また、いずれの評価データの期間においても規定違反に該当する操作が認められなかった。今回、ハイレベルプランナで用いるルールセットを ruleA と ruleB の 2 種類用意したが、ruleA より ruleB のほうがよい結果となった。

評価の際に得られたダム放流操作の例を図 2 に示す。発電時間を長くし、かつ洪水時に至る前に水位低下を完了することができている。

5. 関連研究

5.1 シンボルグラウンディングと階層型プランナ研究

本研究で用いた枠組みはシンボルグラウンディングを含む階層型プランナの一種と見なすことができるが、階層型プランナに関連する従来研究で提案されてきた枠組みとは**人間がもつシンボルに関する事前知識を用いてグラウンディング関数を学習している**という点で大きく異なる。

階層型プランナに関する従来研究の多く [Nilsson 84, Malcolm 90, Cambon 09, Choi 09, Dornhege 09, Wolfe 10, Kaelbling 11] では、人間によってグラウンディング関数が設計され、グラウンディング関数の学習はおこなわれない。一方、本研究で提案した枠組みは、人間の設計者によるシンボルの定義に関する知識を基に、グラウンディング関数が強化学習される。そのため、人間の設計者があるシンボルの定義に関して不正確な（あるいは曖昧な）知識しか持っていない場合でも、階層型プランナが学習を通じてそのシンボルのより正確な定義を獲得することができる。

また、グラウンディング関数を自動で獲得する研究 [Konidaris 14, Konidaris 15, Konidaris 16] が存在するが、これらの研究では人間のシンボルに関する事前知識をグラウンディング関数の獲得に利用していない。そのため、獲得されたグラウンディング関数が出力するシンボルの意味を人間が解釈するためには、グラウンディング関数の分析が必要となる。一方、本研究で提案した枠組みは、人間のシンボルに関する知識となるべく異ならないようにグラウンディング関数を学習するため、学習後のグラウンディング関数の分析をせずとも、プランナが出力するシンボルを人間が解釈することができる。

5.2 ダム操作支援研究

ダム操作支援を目的とした研究の一例として、エキスパートシステムの研究 [Inoue 99] が挙げられる。従来研究でも、シンボルを介するプランニングと同様の仕組みを用いていた点で本研究と共通する部分がある。

従来研究 [Inoue 99] では、熟練操作員がもつ知識 (放流操作の意思決定に影響する判断基準) を数値的な重みとして設計し、ルールに付与していた。これらは熟練者相当の判断を再現するために決めた値であるが、学習による値の更新はしないため、人手で決めたルールよりよい操作を導出できない。本研究ではシンボルに関連付けたグラウンディング関数のパラメータを学習するため、人手で決めたルールよりよい結果を得ることができる。

6. まとめ

本研究では、電力ダムのシミュレータを構築し、人間が理解しやすい操作案を出すことが可能な枠組みである強化学習とシンボルグラウンディングを含む階層型プランナに、熟練者相当の操作を学ばせることを試みた。評価の結果、有用性を確認することができた。

本研究で提案した枠組みは従来研究と比較し、グラウンディング関数を強化学習する際に人間がもつ知識となるべく異ならないように制約を加える点が新しい。また、ダム操作員の負担軽減を目的とした意思決定支援をする上で、予め人間が決めたルールだけでプランニングをおこなう従来研究とは異なり、強化学習でグラウンディング関数を改善し熟練者相当のダム放流操作案を獲得する点も新しい試みである。

今回、過去事例を基に構築した流入量データを用いて学習したため、仮に過去の事例を著しく超える流入量が電力ダムに到達した場合、本研究の枠組みだけでは対応ができない。そのため、過去の経験を逸脱する事象への対処は大きな課題である。

また、現実環境での評価も今後の課題であり、今回構築した数値シミュレータの環境と現実環境はふるまいが異なることが予想されるため、ふるまいの差異を考慮した枠組みを導入することが必要となる。

参考文献

[Silver 17] Silver, D. et al.: Mastering the game of Go without human knowledge, Nature, 500, (2017).

[Inoue 99] 井上ら: 出水予測・ダム管理支援エキスパートシステム, 大ダム, 166, (1999).

[CommonMP] http://framework.nilim.go.jp/

[Williams 92] Williams, Ronald J: Simple statistical gradient-following algorithms for connectionist reinforcement learning, Reinforcement Learning, 5–32 (1992)

[Nilsson 84] Nilsson, Nils J: Shakey the robot, SRI INTERNATIONAL MENLO PARK CA, (1984).

[Malcolm 90] Malcolm, Chris, and Tim Smithers: Symbol grounding via a hybrid architecture in an autonomous assembly system, Robotics and Autonomous Systems 6.1-2 (1990).

[Cambon 09] Cambon, Stephane, Rachid Alami, and Fabien Gravot: A hybrid approach to intricate motion, manipulation and task planning, The International Journal of Robotics Research 28.1 (2009).

[Choi 09] Choi, Jaesik, and Eyal Amir: Combining planning and motion planning, Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, (2009).

[Dornhege 09] Dornhege, Christian, et al.: Integrating symbolic and geometric planning for mobile manipulation, Safety, Security Rescue Robotics (SSRR), 2009 IEEE International Workshop on. IEEE, (2009).

[Wolfe 10] Wolfe, Jason Andrew, Bhaskara Marthi, and Stuart J. Russell: Combined Task and Motion Planning for Mobile Manipulation, ICAPS, (2010).

[Kaelbling 11] Kaelbling, Leslie Pack, and Tomas Lozano-Perez: Hierarchical task and motion planning in the now, Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, (2011).

[Konidaris 14] Konidaris, George, Leslie Pack Kaelbling, and Tomas Lozano-Perez: Constructing Symbolic Representations for High-Level Planning, AAAI, (2014).

[Konidaris 15] Konidaris, George, Leslie Pack Kaelbling, and Tomas Lozano-Perez: Symbol acquisition for probabilistic high-level planning. IJCAI, (2015).

[Konidaris 16] Konidaris, George: Constructing abstraction hierarchies using a skill-symbol loop, IJCAI, (2016).