PNC TJ9604 98-002

公開資料

資源制約を考慮したプラント異常時の意思決定機構に関する研究（1）

―要件の整理と課題の抽出―

（動力炉・核燃料開発事業団研究委託内容報告書）

１９９８年３月

動力炉・核燃料開発事業団
大洗工学センター

Ⓐ

公　開　資　料
PNC TJ9604 98-002
1 9 9 8 年 3 月

資源制約を考慮したプラント異常時の意思決定機構に関する研究（1）
-要件の整理と課題の抽出-

椹木哲夫 [1)]

要旨

従来、各種工学プラントや航空機に代表される大規模・複雑システムで目指されてきた完全自律のシステム化に対して限界や問題点が広く認識されるに至っている。このような問題認識のもと近年では人間（オペレータ）を対象システムの監視・制御の系から排除するのではなく、適宜自動化システムから人間による判断に決定責任を委譲できるようなシステム設計（human-in-the-loop design）を目指す設計理念が模索されている。この理念は意思決定に関わる資源、特に時間制約が厳しい状況で非常に重要になる。本稿ではこの対策として、人間側からのアプローチ、システム側に具備すべき工学的アプローチ、そして両者の「関わりあい」の在り方そのものに対する再考、の大きく３つの方向性から人間-人工物の協調系の設計について考察する。

本報告書は京都大学が動力炉・核燃料開発事業団の委託により実施した研究内容結果である。
契約番号：090D0274
事業団担当部課室：本社　核燃料サイクル技術開発部
1) 京都大学大学院工学研究科精密工学専攻

PNC TJ9604 98-002
March, 1998

Decision Making Mechanism for Plant Anomalies Coping with Resource Boundedness (1)
-- Requirements Identification and Problem Extraction--

Tetsuo Sawaragi

Abstract

Completely autonomous systematization used to be a technical goal for large scale, complex engineering systems such as process plants and airplanes, but recently opposing attitude has been gradually spread as its limitations and problems are recognized. This recognition now leads to a research stream to look for a system design based on a new paradigm where human operators can take over decision making responsibility from automated system on any appropriate situations. This paradigm becomes eminently crucial under tight resource boundedness, especially under strict time limitations. This report discusses the design policy of human-artifact cooperation system through major three aspects: approach to be taken by human, engineering approach to be equipped with engineered systems, and reconsideration on what human artifact interaction should be.

Precise Machinery Engineering, Engineering Faculty, Garaduate School of Kyoto University

目次

図目次

1．緒言

現在の原子力プラントにおいて、炉心損傷に至らないという意味での安全性は、安全保護系の確実な作動による異常時の炉停止を前提とすれば既にほぼ達成されている。この安全保護系は、その起動条件に定められた各プラントパラメータの何れかが正常値から変位してその閾値を越えた場合に制御棒を一斉に装入することで炉心における熱発生を停止する。その起動条件は、プラント諸系統の熱輸送機能及びバウンダリ健全性の保持のための条件を基に、技術者が予め定める。

原子力プラントに診断・制御のための高度な情報処理機構を具備することの意義は、異常発生時に制御棒一斉装入よりも望ましいプラントの挙動を達成することにある。即ち、プラントを、その異常を除去するための行動が可能な状態まで、熱過渡等の健全性への影響を最小限に抑えつつ安全に移行させる機能の達成にある。

現在まで、診断・制御系に人工知能と称される分野の研究成果を反映した情報処理手法を応用することにより、従来は炉停止に至るある種の異常事象に対しては、プロセスパラメータが安全保護系の閾値を越えることなしにプラントを安定な部分負荷運転に移行させられることが示されている[1]が、実際にこのような技術を適用することを想定すれば、完全な炉停止よりも部分負荷運転への移行が望ましくまた可能であるという判断こそが重要である。この判断機能の実現のためには、それぞれの操作を想定したプラントの挙動予測は言うに及ばず、原因となる事象の拡大の可能性の判断、その事象回復が可能なプラント状態の範囲の同定、等高度な情報処理が必要である。しかもそれらの処理をプラント状態の変化に先だって行う必要がある。部分負荷運転への移行について、その可能性や炉停止に対する優位性の判断のための処理時間が一定量を越えるのであれば、炉停止を決定すべきである。つまり、情報処理機構自身が、解の精度や詳細さと処理速度とのトレードオフを判断する必要がある。

本研究では、このような情報処理機構の機能構成、構成要素となる要素技術、各要素技術の現状における達成度を明かにし、今後取り組むべき課題を抽出することを目的とする。

(問題の背景)

複雑でかつ安全性に対する社会の要求が厳しいプロセスプラントの建設においては、導入可能な限りの手段を講じた構成がとられる。しかしながら、不具合の発生に際してその進展を押し留めて影響を最小限に抑えるためには、運転員には最大限の努力を払うことが要求される。さらには、異常事象の進展を設

計上の配慮から防止するために設置された様々な機器や系統こそが、プラントを一層複雑なものとして、不具合が発生した際の運転員の認知的負荷を増大させる可能性も否定できない。

真に安全性の向上に寄与しうる意思決定機構に必要な機能には、極くおおまかに考えても以下のようなものが挙げられる。

a)対応操作決定以前に必要な機能

・同定された事象の拡大／進展可能性の判断

例えば、ポンプが停止した場合、ポンプの機能がそれ以上失われることは想定しがたいが、蒸気発生器の伝熱管が破損した場合は隣接細管の損傷に至り得るし、弁の固着の場合は再び固着が解除される場合がある。このような情報によって目的とするプラント状態や実行すべき対応操作は異なる。

・異常の回復が可能となるプラント状態の範囲の決定

ポンプの停止の場合、その上下流で他の主要機器を含まない範囲に十分な性能の弁が存在する場合はそのポンプを隔離して他の系統では冷却材の循環を継続できる場合がある。しかしポンプ交換のために雰囲気を置換する必要があれば、火災等の望ましくない反応の危険回避のための他の操作が必要な場合があり得る。

b)対応操作決定のために必要な機能

・可能な操作手順(群)の抽出、及び抽出された操作手順の難度、危険性の評価
・操作に要求される速度、定量的な精度、それらが満たされない場合の影響等
・対応操作手順の各候補を想定した場合のプラント挙動予測、および予測された挙動の評価
・予測されたプラント挙動に基づく熱過渡応力や圧力変動に対する構造強度の裕度等

c)処理全般に必要な機能

・資源制約を反映して意思決定戦略そのものを動的に制御する機能

最低限の安全が保たれる操作（含む制御棒一斉挿入）に対する、計算資源の投入による優れた操作手順獲得の可能性と意思決定遅延のリスクとのトレードオフの判断

運転員は通常、原子力プラントにおける中央制御室のような、プラント全体の状態推移を観察するために主要（と設計時に想定された）パラメータの計器が集中設置された場所で監視業務を行っていると想定するのが現実的である。一方、不具合の発生に対して、プラントが危険な状態へ推移する危険を回避する手段を講じ、また新たに不具合が再発する可能性を摘むために必要な情報は、プラントの複雑さから考えれば、集中監視が可能な範囲をはるかに超えている。

つまり、不具合の発生から集中監視用の計器にその影響が現れた時点では、運転員はその原因として複数の第一事象を候補として想定しなければならない。異常診断に関する研究の多くが真の原因以外の故障仮説を効率的に否定するアルゴリズム上の新手法開発に重点を置いているのに比べ、実際の異常診断には次のように様々な要因が加わる：

・追求する意味のある最善のシナリオ
・現実的な可能性を有するシナリオ群の中で最悪のもの
・最善、最悪の評価尺度
・故障仮説毎の進展防止のための猶予時間
・故障仮説毎の確認及び否定のために必要な時間と財産的コスト(情報獲得と推論)

等。

これらの要因は互いに密接に関係しており、複合的に派生する新たな要因としては、ある対応操作がその後の推論負荷（日常生活における「心配」とほぼ同義）をどれだけ減ずる効果があるか、というようなものがある。

これらの要因を総括的に反映した異常時対応用の意思決定機構を開発するためには、現在の人工知能ないしは情報処理技術は十分に発達していないと言わざるを得ない。一方、少なくとも現在では、訓練を受け、経験を積んだ人間の運転員は人工の意思決定機構よりもはるかに高い能力を有している。従って、運転員の認知負荷の低減と原子力プラントのような大規模プロセスプラントの一層の安全性の確立のためには、両者の協調による意思決定機構が形成されるような情報処理機構及びマンマシンインタラクション環境の開発を目指す方がはるかに有効であると考えられる。

本研究では、上述した様々な要因を含んだ運転員の意思決定の実例を参考にしながら、人間－自動化系の協調的意思決定機構に関する、人間からのアプローチ、自動化系において考慮すべきアプローチ、及び協調のあり方そのものについて、関連研究動向を踏まえて考察を加える。この報告書では、2 章でこの３つの問題のそれぞれについて論じた後、3 章においては資源制約とマンマシンインターフェースのありかたを更に詳細に検討する。なお、本研究にあたっては、人間が生来的に行う意思決定プロセスが、情報獲得を人工的に支援する手段がある時にどのように行われるがを把握しておくことが極めて重要であるが、その実例と観察及び評価を Appendix として追記する。

2．原子力プラントの運転のための意思決定における資源制約とマンマシンインタラクションの関係

2－1 はじめに

従来、各種工学プラントや航空機に代表される大規模・複雑システムでは、その操作への人間の介入を極力減らすことで安全性と信頼性の向上を目指す、いわば完全自律のシステム化を探るアプローチが目指されてきた。ここでの人間は、一般ユーザとしての人間ではなく、特定のシステムの監視や制御に当たる十分に訓練を重ねた熟練オペレータが対象であり、それが対峙すべきシステムは、ハード・ソフトの別なく人工物システム一般を指す。しかしこのような完全自律化の途は現在の技術レベルをもってしても未だその実現には限界があるのみならず、機械への代替による熟練技能の喪失も懸念されることから、最近では人間との自然な協調を実現できるシステム設計が求められている。さらに各種自動化ツールに代表されるようにそれ自身は高度に洗練されているはずの工学的成果の産物が、これを操る側の人間による使い方次第では必ずしもシステム全体としてのパフォーマンス向上に繋がらないだけでなく、時として取り返しのつかない大事故を誘発してしまうことも大きな懸念を生んでいる。これは単に使い方を誤ったヒューマンエラー（人的過誤）として片付けられる問題ではなく、このようなピットフォールを埋め込んでしまったシステム設計の問題、さらに限定するならばヒューマンシステムインタラクション設計の問題にほかならない。

このような問題認識のもと近年では人間中心の自動化システム（human-centered automation）のあり方が模索されている。これは人間（オペレータ）を対象システムの監視・制御の系から排除するのではなく、適宜自動化システムから人間による判断に決定責任を委譲できるようなシステム設計（human-in-the-loop design）を目指す設計理念であるが、人間-機械の間での最終的な意思決定の権限の所在、さらに両者間での機能的役割の分担については依然「絵に描いた餅」の域を出ず、その実現方法に疑問を投げ掛ける声も多い。本稿では上述のような問題に対する対策として、人間側からのアプローチ、システム側に具備すべき工学的アプローチ、そして両者の「関わりあい」の在り方そのものに対する再考、の大きく３つの方向性から人間-人工物の協調系のデザイン原理について概説する。

2－2 人間中心のシステム設計に向けての技術課題

従来までのインタラクション設計は、人間-人工物の間のインタラクションそのものを対象化し、観察主体と切り離すことによってあらたな実証性を確保

し、鳥瞰的な視点からこの相互作用をコントロールするアプローチがとられてきた（外在的アプローチ）。しかしこのように観察主体をインタラクションの外部に位置づけることで境界を定め、境界内の可制御性、最適化を追及してきた結果として、境界外の非制御性、混沌が現われてきている。実際このことはまさに今日の自動化システムの在り方が抱えている問題そのものであり、自動化の皮肉（ironies of automation）や自動化に誘引された混乱（automation-induced surprise）といった問題として指摘されている。環境外に観察者が立ちえる場合には、支援は外部の問題となり内部の最適化が可能となるが、現在我々が直面する制御対象の複雑さの拡大はその枠を超え、観察者は内部に位置せざるを得なくなってきている。すなわち自己から切り離された対象の最適化ではなく、最適化の主体を内在化させ、他者とのコミュニケーションを自ら設計していけるアプローチが求められる。

２－２－１　ヒューマンシステムコミュニケーション

ヒューマンエラーの根源は、コミュニケーション設計の主体と実際の行動主体の乖離にあると指摘されている。従来は人間-機械の間での役割分担はシステム設計者の手に委ねられ固定的なものであった。これまで領域固有知識を実装して意思決定や問題解決の支援に当たることを想定された各種の意思決定支援システムやエキスパートシステムは、未熟練オペレータへの訓練システムとしては機能しても、経験を重ねた熟練のオペレータのパートナーとなるまでには至らなかった。またシステム変数が異常値に変化した際に発せられる単純な警報装置ですら、各変数の異常変化に連動して一挙に次々と鳴り響くため、大局的な瞬時の状況認知が求められる緊急時等ではオペレータにより故意的に機能を停止されることが多いという。これらの問題は真に共感的な、そして協調を創出するためのコミュニケーションの在り方を再考する必要があることを物語っている。

上述のような人間-機械の間でのインタラクション設計におけるコミュニケーション原理の基本には「交信モデル」、すなわち「伝え手の側にまず伝えられるべき観念ないしは思考内容（伝達意図）がそれとして分節されてあり。それを伝え手がコード化して送信したものを受け手が受信してデコード化し伝え手の観念や思考を再構成する」という形に限定された通信である。真の協調のもつ意味は、「単独では成しえないが故に不足する部分を補うこと」（augumentive）が一義的であり、その意味では警告や助言として人間に情報を提示することに意義があるものの、これ以外にも「多様な知識や視点の融合」（integrative）の側面、そして真偽はさておき「批判や批評を交わし共有することを通しての繋がり」（debative）の意義が存在する。とくに意図が明確

に事前に定まらない段階での多声的な「関わり合い」を通してのコミュニケーションから産み出される共感的な領域があることが指摘されている[2-鯨丘 97、岡田 95]。Entine and Safaty らは単声的で系列的な情報提示が人間の意思決定に与えるバイアスについて議論している[2-Entine 97]。さらに Smith らは真の協調を実現するには一方の指示を拒否したり自らの意志を優先して実行できるような適宜オーバーライド機能が必須であることを実験から検証している[2-Smith97]。人間と機械の間でのインタラクションを考えるとき、必ずしも直示的な形で指示を出せる場合ばかりではない。自動車のハイウェイ交通管理や次世代航空管制システムとして提案されているフリーフライトにおいて議論されているように、明確な意図のもとに指示を発してそれに隷属させるような「交信モデル」に基づくコミュニケーションはもはや機能し得なくなることが予想される。これは交通管制のように指示を提供する側に交通量の増大に対してこれらを網羅的に把握することに限界があり、さらに指示を受ける側の価値や目標・状況が多様であるがために最適化基準を事前に設定し難いという現実に依拠するところが大きい。Murray and Liu らはハイウェイ交通管理を対象として、助言システムに代わる勧告システム (hortatory system)、すなわち人間オペレータが経験則として有する定性的な「勧告」を、分散したエージェントが各々の置かれている局所的な固有状況を参照して、解析的に自らの判断基準に照らし合わせて人間の発した勧告を受け入れたり拒絶したりするシステムの概念を提唱している[2-Murray 97]。

２－２－２ 創発的協調活動のモデリング

近年の複雑大規模プラントの監視制御、人工衛星の地上管制業務や ICU（集中治療室）での重病息者の手術中ならびに術後ケアなど、高度な信頼性と安全性が求められる多くの監視業務は、通常複数の人間からなるチームによる協調活動である。これらチーム成員の各々の責任や業務内容は予め明確に区分され各々のとるべき手続きも標準化されているのが通常であるが、それでもその間では部分的な重複がもたされこれによる冗長性によってチーム全体としての信頼性が確保されている。ここでの特徴は、チーム成員の流動性、厳しい実時間制約、緊急時の組織構成の可変性、が挙げられるが、このような協調活動のモデリングはヒューマンシステムインタラクション設計に際して今後ますます重要性を帯びてくる。これに対して従来までにも DAI (Distributed Artificial Intelligence)・CSCW(Computer Supported CooperativeWork)・EI(Enterprise Integration) の分野においてアプローチがなされてきた。そこでの議論の焦点はやはり個々の活動主体（エージェント）の情報処理プロセスや合理的設計規範に則った推論技法に集約されており、これら個々のエージェント設計は、

システムの全体目標の下での一貫性・効率性・正確性に相当する性能指標のもとに進められてきた。

これに対して、近年のサイバネティクス分野からの創発的協調活動設計に対する視点が興味深い。ここでは開放系としての組織設計の規範はすべてトップダウンに与えられるものではなく、実践（すなわち他者との関わりあい）を通して個々の行為主体がどのようにその行為の意味を見出していくかについてのプロセスに着目する。そして、社会的規範（social norm）や個々の行為主体（actor）の役割、行為主体を取り巻く数々の工学的所産が呈するアフォーダンス、の間で構成される相互限定性（reciprocity）の側面に焦点が当てられる[2-Contractor 93、Jones 97]。実際、人間の意思決定は、その多くが単独に他者から切り離されてなされるのではなく、その決定が他人に対して及ぼす影響・反響を考慮して下されるものであり、また決定それ自体も他者を通じて獲得したはずの「規範」や「価値」に影響される。すなわち、メタ認知として自省的・内省的過程を働かせること自体、自分の行動をどれだけ制御しているのかということに関連し、傍若無人でない社会的行動を行うために重要なことである。この意味において協調活動のモデリングには、社会的推論の枠組みが要請されるところとなる。すなわちここでの推論の対象は、自己、他者、状況、社会（社会を動かす決定因として自己が果たす役割の認知）であり、そこで準拠すべき規範は達成性の目標、すなわち何らかの最終的な対象や状態を獲得したり習得するような目標のみではなく、コンサマトリー性の目標、つまり特定の行動や状態を経験するプロセスそのものが目標の実現に直結するような行動原理の解明をも併せて進めていく必要がある[2-塩瀬 98]。

２－２－３ 資源制約下での意思決定のモデリング

ヒューマンシステムインタラクションを設計していくうえで、ヒューマンモデリング、すなわち人間がどのような情報からどのように判断して行動を出力しているのかについてのオペレータ行動モデルの設計は重要である。近年の人間機械系における人間の役割は従来の操作タスクから監視・状況認知・計画立案・診断と言った管理タスクへと移行してきている。ここでモデル化するべき人間は、環境や状況から切り離されて存在し知覚入力→認識→選択→出力実行の一連の系列を逐一的確に行うような情報処理的なものとは限らない。むしろ熟練したオペレータの行っている処理は、自身以外の周辺要素、環境や状況、文脈からの拘束条件を巧みに知覚・利用すると同時に自らの能動的・溝成的な理解を重ね合わせて文脈を創りだすような相互循環を繰り返すことで巧みに選択自由度の軽減を計っていく動的プロセスとして特徴付けられる[2-Bainbridge 97、Woods 95]。

従来までの意思決定モデルと言えば、規範的決定理論（normative decision theory)、すなわち「決定がどうあるべきか」を論じる古典的決定理論が主流であり、ベイズ推論や多属性効用理論、期待効用最大化の原理、などはその代表格と言える。これに対して「実際の決定」がこのような規範的なものから乖離することが古くから指摘されており Tversky and Kerneman らを代表として人間の意思決定に混入する各種のバイアスに関する研究も盛んに研究されてきた。いずれも意思決定活動を決定者の置かれた文脈や状況から隔離された中での選択行為として捉えるアプローチには変わりがない。これに対して、決定活動を選択行為のみならずその後の行為実行まで含めたより大きな時間的文脈の中で、さらに決定の要因を決定者の内部のみに求めるのではなくその周りの環境や状況の特性がどのように決定を方向づけるのかという観点からの意思決定モデリングの再考が進められている。これらの分野は自然派意思決定（naturalistic decision making）の理論と呼ばれており、熟練専門家が厳しい種々の制約下、とくに緊急時にみられるような大きなタイムストレス下で瞬時の判断を巧みにやってのける決定活動のモデリングを目的としている[2-Klein 93]。

自然派意思決定では上記のような設定のもとでとられる熟練専門家の状況認知、すなわち直面する状況をどのような問題として認知し構成するかのフエーズで経験がものを言うのであって、さまざまな代替案の間の評価を網羅的に行うわけではないことを指摘する。むしろこのような競合代替案が生まれないような問題の構造化・フレーミングをできる能力に長けているのであって、そこでの処理に資源を優先させ、その後の選択のフェーズでは絞られた代替案の正当性や実現可能性の是非をメンタルシミュレーションで追認する程度にとどめることで計算資源を極力抑えている点を強調する。この点が、決定主体とは異なる実験者によって設計され外在的に与えられた決定問題の枠内での合理的決定を論じる古典的意思決定論と前提を違える点であり、決定主体の状況認知や問題設計・構成能力が選択に優先されるスタイルの意思決定である（recognition-primed decision making)。そして決定者の熟練度が上がるほど、また時間制約がより厳しく不確定性のより高い状況であるほど、この種の意思決定のスタイルが支配的になるとしている。

以上の意思決定研究の流れの根源は、Simon による従来の計算的合理性の限界を指摘した限定的合理性原理の概念の流れを汲むものである。知能の本質は、問題や探索空間を簡単に定義できない場合にも適切に行動できる点にある。問題空間の合理的探索は、空間そのものが生成されなければ不可能であり、形式的構造が状況にうまく対応している場合にのみ有効である。従って問題や探索空間そのものを定義するフエーズをも含めたモデリングが求められることにな

る。

２－３ 人工知能研究における資源制約下での合理性を保証する推論

前章では現状でのヒューマンシステムインタラクション設計が抱える問題について、そこでの技術課題を中心に概説を行った。本章ではこれらの技術課題の解決に向けての工学的アプローチ、とくにヒューマンシステムインタラクションの知的支援を目的としたシステムを構築する上で要求される技術のいくつかについて概説する。とくにここでは従来のエキスパートシステムの延長線上で、人間との実時間協調を目的とするものとしてアソシエイトシステム（associate system）をとりあげる。これはP. Maesによる、アプリケーションソフトウエアとこれのユーザとなる人間の問に介在するインタフェースエージェント（interface agent）の概念に近いもので、人間オペレータと複雑大規模システム（もしくはそれへの自動制御機器やより一般には Task Interactive Computer）との間に介在する。そして自身の領域固有知識を背景に自律的な問題解決能力を有するものの、人間の作業を代替することを究極の目的とするのではなく、オペレータ挙動を観察しながら文脈に依存したタイムリーな支援と、人間オペレータとの間で適宜状況に合わせて動的に役割を交替しながら協調を創出していくための人工システムである。Command and Controlの領域など実時間制約の厳しい状況下でしかも緊急時のような不測の事態の生起時にも適切に人間オペレータへの対等な立場での知的パートナーとして機能することを求められ、まさに人間中心の自動化の概念を現実のものとしていく上で興味深い[2-Rouse 87]。ここではこのような知的人工物としてのアソシエイトシステムを工学的に実現していくための要素技術のいくつかについて、前章までの議論との関連からまとめる。

２－３－１ 資源制約下での計算合理性の定式化

アソシエイトシステムには通常大規模な領域固有知識の実装が求められる。しかしプランニングや問題解決に費やされる時間に制限がある場合には、各々の計算に要する時間やその計算の結果どのようなものが得られ、それらに関する選好がどうであるかを考慮する必要がある。いわゆる計算資源の管理（マネジメント）であり、推論のストラテジー選択を柔軟に制御することで、資源の制約を有しながらもその範囲内で合理的な推論の進め方を見いだしていくことが求められる。これは同時にメタ認知の問題であり、エージェントは自らの行う意思決定行為そのものについて内省することができなければならない。すなわち複数の決定手続きの中から、その効果に対する評価や見積を考慮にいれた上で、次にいずれの決定手続きを実行していくかを決めるのがここでのメタレ

ベル推論の主たる仕事である。従来の古典的な意思決定論では、結果として得られる解のもつ効用を最大化することが主眼であり、この解の導出に要するコストや処理時間に対する制約というものは陽に考慮されていなかった。大規模・実時間知識ベースシステムとしてのアソシエイトシステムにとっては、行為実行の遅延というリスクを犯してでもさらなる計算処理を進めるか否かについての判断は決定的に重要になる。以下ではこのようなエージェントのメタ認知を内包した推論技法の設計法について概観する。

Russell and Wefald は大規模空間内での探索問題を例にエージェントの推論制御の問題、いわゆる“deliberate or act”（熟考か行動か）の定式化を行っている[2-Russell 91]。エージェントの目標は、結果の集合上で定義された効用の最大化である。エージェントにはいくつかの現実世界に実際に作用をもたらすような基底レベル行為（base-level action）の集合 A が用意されており、ある状態のもとで行為を実行することである結果が得られる。エージェントはいつの時点においてもその時点で最善と思われるデフォールトの行為 $\alpha \in A$ を実行することができ、これに加えていくつかの計算行為（computational action）{Si} が実行可能で、この後者の実行により現時点でのデフォールト行為 α を別のものに改訂できる。このときエージェントのとり得る代替案としては、α、S1、S2、... 、Sk のオプショシが存在し、計算行為はエージェントの内部状態のみを変更する。しかし計算時間の経過の間は行為実行の遅延による機会損失が生じることから、計算のネット価値は、計算を行った結果として得られる効用から現時点でデフォールト行為を実行することにより得られる効用を差し引いたものとして評価される。１ステップ先の計算を実行するのが合理的であるのは、この計算によって現在のデフォールト行為以外の行為に対する評価が現在のデフォールト行為に対する評価を上回るような場合に限られ、評価が不変である場合には計算を実行する価値はない。従って計算によって得られる期待ゲインは、現在のデフォールト行為に対する評価と、計算後に改訂される別の行為に対しての評価の期待値、の両者の差分として計られる。この期待ゲインが計算にともなう遅延コストを上回れば、さらなる熟考（すなわち計算あるいは推論）を続けて現時点での最善策である行為を実行に移す前により慎重にこれについての評価を続行すべきであるし、そうでない場合は熟考を中止し即応的に現時点での最善策である行為を実行に移すべきということになる。ここで熟考を続けると判断すべきケースは以下の２つの場合になる。第一のケースは、さらなる計算が別の基底レベル行為 β の効用に対する評価に影響を与える場合で、計算の結果、基底レベル行為 β の効用評価が現在の最善策である行為 α の効用を上回る場合である。第二のケースは、さらなる計算が現時点での基底レベル行為の最善策 α の効用評価に影響を与える場合そ、計算の結

果、行為αに対する評価が下落して、別の行為βの現時点での効用の方が上回る場合である。Russell and Wefald らはこれらの各々のケースについての計算価値（value of computation）の指標を走義し、この算出結果に基づいた推論制御の方策を提案している。また Ezioni はプランニングを対象として満足化探索手法[2-Simon 75]、すなわちいずれかの目標に到達するための期待探索コストを最小化する問題に対して、資源制約を有するエージェントの推論制御問題を３つの次元、すなわち、熟考コスト、実行コスト、目標の効用、の３つの次元から解析を試みている[2-Ezioni 89]。エージェントはいずれの行為を実行するかのみならず、どの目標にコミットするかについても選択を強いられる。ある手段（method）を実行した際の目標達成の成功確率（どのような状態で手段を適用すればどのような効用をもつ状態がどれくらいの確率で得られるか）と期待されるコストの評価は経験から同定されねばならないが、この同定のために、過去の履歴の保持とクラス形成の能力をもたせ、過去の経験を集積するにしたがってその評価を修正していく手法を提唱している。そして手間のかかる高い効用をもつ代替案に代わり、迅速な近似解をより選好するような限界期待効用を最大化するような欲張り合理性（greedy rationality）の原則についての定式化を行っている。さらに Etzioni は上述のような制御アルゴリズムで用いられる諸量の評価における信頼性と状態概念の構成・再構成に関連して、資源制約を有するエージェントにとって状態概念の再構成によりもたらされる利益とそれに要するコストの間でのトレードオフ解析について触れている[2-Barnett 84]。これらの知見は、資源制約を有するエージエントにとって計算や推論の継続やプランニングに関する熟考のプロセスのみならず、状態空間設計も含め問題をどのように表現して構成するのかについてのコストをも内省できる能力を有するエージェント設計が必要になることを示唆している。一方、アソシエイトシステムのように実時間に振る舞うことが要請されるエージェントの推論制御問題を定式化する試みが Fehling and Breese により行われている[2-Fehling88]。アソシエイトシステムのようなエージエントでは、自身の目標達成と不確実対象に対する信念管理の両者を併せて実行していかねばならない。これは適応システムの制御で論じられてきた二重制御の理論（dual control theory）の考え方に近い[2-Feldbaum 65]。すなわちエージェントの得るフィードバック情報は２つの役割、すなわちシステムを設定値に安定にもっていくための利用と、システムが構築している内部モデルの修正や改善に寄与する情報の収集を計画するための利用、の両者に用いられる。このとき後者の目的を遂行するためには前者を犠牲にする、すなわちある程度は不安定になることに目をつぶらねばならないが、これは目標達成からは一時的に乖離するもののモデル（信念）を改善することを優先することに対応する。Fehling and

Breese らはこのような二重制御下でのトレードオフについて意思決定論に沿った定式化を試みている。

その他 Heckerman らはベイジアンネットワーク (Bayesian Network) [2-Pearl 81] として知られる確率推論 (probabilistic reasoning) を用いた診断モデルについて、このようなモデルベースでの診断とその結果をコンパイルしたルール知識に基づく診断の２つの診断ストラテジー間でのトレードオフ分析を行っている[2-Heckerman 90]。すなわちエージェントの即応性を高めるためには、複雑でコストのかかるモデルベース推論よりも、前もってモデルを解法し計算された結果をコンパイルして保持しておき、これを単純な照合によって実時間で利用していくアプローチが有効と考えられる。計算にコストを要する確率推論モデルの解をコンパイルして利用することは、推論に要する時間コストを考えると実時間での制御が要求される分野では非常に効果的ではある。ただし、満足のいく挙動を生み出すという点においてはこれらのコンパイル知識は有効に機能するかもしれないが、規範的な原理に則って知識を設計していく立場から考えると。このコンパイル知識をやみくもに活用するわけにはいかず、コンパイル知識の功罪として、問題領域で得られる証拠の信頼性、とるべき行為のもたらす効用、実行時の遅延に伴う時間コスト、メモリのコスト、の諸点を考慮しておく必要がある。

２－３－２ 実時間・大規模知識ベースシステムにおける動的モデル合成

大規模・実時間知識ベースシステムにとって、その両者の特性、すなわち大規模性と実時間性はとりもなおさず相反するものであり、これら２つの目的をどのように両立しえるかがその実現に向けての鍵となる。この場合、知識ベースとしては大規模で網羅的な知識を用意しておくにせよ、実際の知識利用時には、動的に変化する兆候データや文脈情報に基づき、当該の問題解決を行うに必要最小限の部分的知識のみを活性化してできるかぎりコンパクトな問題解決モデルを構築し、その解法に大きなコストを費やすことなく問題解決の即応性・適宜性を追及するという方策が考えられる。いわば知識のスケーリングと再利用性を重視するアプローチで、知識に基づく動的なモデル構築 (Knowledge-Based Model Construction : KBMC) の技法[2-Breese 91、Leng 91]と呼ばれる。Provan らは心臓病診断を対象として KBMC のシステムを試作している。ここでのプロセスは、背景情報の意味づけ、問題領域の文脈の生成、決定問題の定式化、決定モデルの構築、評価、の５段階が各々領域固有知識のガイド下で実行される。ただし、このプロセスを経て構築されたモデルもその妥当性がいつまでも保証されるものではない。診断では、動的 (dynamic) かつ時系列的 (sequential) な推論をいかに実現するかがその信頼性を大きく左右

する。多くの診断モデルはある時刻におけるスナップショット時の診断を扱うものが大部分であるが、変数間の（確率的）依存性が時間と共に変化するような対象の診断ではこのような手法ではうまくいかない。このような対象では、現在構築されているモデル自身が妥当なものか否かのメタ認知としての判断が、現時点で利用可能な情報に基づいてなされなければならない。Provan は感度分析（sensitivity analysis）の手法を使ってこの種のモデルの妥当性の検討を行なう手法を提案している[2-Provan 94]。すなわち感度分析の結呆、代替となるモデルの方がよりよい決定を推奨するならばモデルの更新を適宜行う。具体的には、時間と共に変数間の依存関係に関する確率と因果関係の構造ともに変化を許容する知識が与えられたもとで、どのようにモデルを更新していけばよいのかについて、特に結果の等価性（equivalence of outcomes）という基準を導入したモデル更新のアルゴリズムを提案している。

２－４　人間中心の自動化実現に向けたインタフェースエージエント設計

最後に以上の内容に関連するところで、現在我々のグループで進行中の研究の概要を述べる。我々はこれまでに大規模複雑な対象プラント（都市ガス製造プラント）と人間オペレータとの間に介在して、人間判断と自動化システムとの間の協調を創出する機能を有するインタフェースエージェントの設計を進めてきている。このようなエージェントとしては、オペレータシステム間のインタラクション系列をオペレータの肩越しに観察し、オペレータの概念構造を構築しながら、人間（オペレータ）にとっての環境（プラント）の「見え」を適宜構成していくことが必要になる[2-Sawaragi96]。その上で、エージエントは対象システムの置かれている状況、また人間オペレータの状況を観察しながら、オペレータ行動を監視し、適切な支援を与える一方で、状況によってはオペレータの入力する操作にオーバライドして自らの判断で適切な操作を決定し実行していかねばならない。

我々はまず観察からの概念学習手法[2-Fisher 87]に対して、入力事例を記述する属性群の間での取捨選択、遺伝的アルゴリズムによる複数の属性からの新たな属性の構成的（constructive）な合成、の両者の機能を付加することによって、雑多な対象のなかから的確に意味ある情報のみを選別し抽出できる環境認識と概念生成の能力を具備した学習アルゴリズムを開発している。このアルゴリズムをオペレータ対象システム間でのインタラクション系列の観察事例に対して適用し、熟練オペレータの認知構造を反映した概念木への洗練化を可能にしている[2-Sawaragi 98]。さらにこのように構築された階層型の状態概念の分類に基づいて、オペレータが対象システムの異常状態を同定し、それに基づいて適切な対応操作を決定する際のモデリングを行っている。ここでの状

況は異常の生起後、観測される異常兆候の観測に基づいて、その異常のタイプを同定し、それに対する適切な復旧操作を、やはりその異常タイプに固有に決まるデッドラインまでに決定して実施しなければならない。より多くの兆候の観測を続ければ続けるほどより確実な異常同定が可能になるものの、デッドラインまでに残される応答時間は短縮される。このような時間制約の存在する状況下では、対象システムの異常状態を唯一に確定した上で対応操作が決定できるわけではなく、あくまで不確実な状態同定のもとで、操作を施した結果もたらされる効用の期待値を最大化する操作決定を行なわざるを得ない。我々はこのようなシステム状態の同定と対応操作の決定のためのオペレータの診断・操作決定モデルを、ベイジアンネットを意思決定分析用に拡張したインフルエンスダイアグラム[2-Howard 83]により構築している。ここでのオペレータにとっての効用は、対象システムの異常状態（異常仮説）、採るべき復旧操作、時間（異常発生からの経過時間）、の三者に依存して決まる時間依存効用（time dependent utility）として定義している。

上記の診断モデルを構築するためには、可能性のあるすべての異常仮説を網羅的かつ排他的に分ける分類クラスの集合がその実現値として定義されなければならない。この変数はオペレータが対象世界を認識する際に用いているカテゴリに相当する。このようなオペレータの内面で捉えられているカテゴリの内容を特定することは一般には困難であるが、ここでは上述の概念学習により構築した概念階層（概念木）から全異常事象の集合を排他的に分ける分類クラスのいずれかを充てる。このような分類クラスの集合のとり方を概念被覆（conceptual cover）と呼ぶ。概念被覆としてはさまざまな抽象レベルでの状態概念のとり方が可能となり、これらの各々が異常仮説変数の実現値を表わすことになる。概念被覆の選び方によっては構築された意思決定モデルで出力される推奨案のもつ期待効用値ならびに各選択肢間の効用値の分布の性状が変化する。我々はこの定量情報をもとに、個々のオペレータの有する価値観（例えば、プラント運転の効率性と安全性）の違い、そして置かれている状況（例えば、緊急性）の違いに応じて適切な概念被覆を決定するためのアルゴリズムを開発している。ここでは診断モデルが出力する結果の効用のみならず、一旦同定した概念被覆の更新のためのコスト、すなわち自身の決定が根ざすべき診断モデルの妥当性とその修正・改訂をどこまで精緻に行う価値があるかについても、エージェント自身が熟慮するための資源制約下推論の定式化を行っている[2-Sawaragi 97]。

一方、人間オペレータを系に取り込んだ協調モードでのインタフェースエージェントに科せられるタスクは、観測可能な兆候データをフィルタリングしてその一部をオペレータに提示することによって、この提示情報からオペレータ

に異常事象を同走させ、今後の事象推移を予測して対応操作を決定させることになる。この支援モードでは、兆候データの提示からオペレータの解釈を経て解が特定されそれが実行に移されるまでのプロセスでより多くの不確定要因が介入することになる。その主たる要因としては、オペレータの習熟度の違いならびにオペレータの認知負荷の状況が挙げられる。人間が行なえる意思決定の質は提供される情報の複雑性と量、さらに緊急性の度合いに従って低下することは各種の実験により認められており、このような人的要因を考慮した情報提示の管理が求められることになる。我々は現在この提示情報の決定のためのアルゴリズムを開発中である。

エージェントにとって人間オペレータとの真にフレンドリィな協調関係を樹立するためには、上述のような自身のコミットする活動のみならず、他者である人間オペレータのモデルを自身に内包したインタラクション設計を行えることが必須となる。そのためには、エージエント自身が現在観測可能な兆候群に基づいて有している対象プラントの状態に対する視点と、その内の一部の兆候しか提示されない場合のオペレータの有する対象プラントの状態に対する視点、の両者の融合が必要となる。

◇参考文献◇

[2-Bainbridge 97] Bainbridge、 L.: The Change in Concepts Needed to Account for Human Behavior in Complex Dynamic Tasks、 Trans. of IEEE on SMC、 27-3、 pp. 351-359 (1997).

[2-Barnett 84] Barnett、 J.A.: I-low Much Is Control knowledge Worth? A Primitive Example、 Artificial Intelligence、 22、 pp. 77-89 (1984).

[2-Breese 91] Breese、 J.S.、 Goldman、 R. and Wellman、 M.P.: Knowledge-Based Construction of Probabilistic and Decision Models: An Overview、 Working Notes of AAAI Workshop on Knowledge-Based Construction of Decision Models (1991).

[2-Contractor 93] Contractor、 N.S. and Seibold、 D.R.: Theoretical Frameworks for the Study of Structuring Processes in Group Decision Support Systems、 Human Communication Research、 19、 pp. 528-564 (1993).

[2-Entine 97] Entine、E.E. and Serfaty、D.: Sequential Revision of Belief: An Application to Complex Decision Making Situation、 Trans. of IEEE on SMC、 27-3、 pp. 289-301 (1997).

[2-Ezioni 89] Ezioni、 O.: Tractable Decision-Analytic Control、 in Brachman、 R.J. 4 et al. (Eds.)、 Proc. of the First Int. Conf. on Principles of KnowledgeRepresentation and Reasoning、 Morgan-Kaufmann、

Los Altos、 CA、 pp. 114-125 (1989).
[2-Fisher 87] Fisher、 D.: Knowledge Acquisition via Incremental Conceptual Clustering、 Machine Learning、 2、 pp. 139-172 (1987).
[2-Fehling 88] Fehling、 M. R. and Breese、 J. S.: A Computational Model for Decision-Theoretic Control of Problem Solving under Uncertainty、 Rockwell Int. Sci. Ctr.、 TM837-88-5 (1988)
[2-Feldbaum 65] Feldbaum、 A. A.: Dual-Control Theory IIV、 Optimal Control Systems、 New York、 Academic Press (1 965).
[2-Heekerman 90] Heekerman、 1). 1:.、 Breese、 . J. S.. and Horvitz、 E. J.: The Compilation of Decision Models Proc. of Int. Conf. of Uncertainties in AI、 pp. 162-173 (1 990) .
[2-Howard 83] Howard、 R. A. and Matheson、 J. E.: Influence Diagrams、 in Howard、 R. A. and Matheson、 J. E. (eds.)、 The Principles and Applications of Decision Analysis、 Strategic Decision Group、 Menlo Park、 CA (1983).
[2-Jones 97] Jones、 P. M. and Jasek、 C. A.: Intelligent Support for Activity Management (ISAM): An Architecture to Support Distributed Supervisory Control、 Trans. of IEEE on SMC、 27-3、 pp. 274-288 (1997).
[2-Klein 93] Klein、 G. A. et al.: Decision Making in Action: Models and Methods、 Ablex Pub. Corp.、 Norwood、 NJ (1993).
[2-鯨丘 97] 鯨岡:原初的コミュニケーションの緒相、ミネルヴァ書房 (1997).
[2-Leng 91] Leng、 T-Y.: Representation Requirements for Supporting Decision Model Formulation、 Proc. of Int. Conf. of Uncertainties in AI、 pp. 212-219 (1991).
[2-Murray 97] Murray、 J. and Liu、 Y.: Hortatory Operations in Highway Traffic Management、 Trans. of IEEE on SMC、 27-3、 pp. 340-350 (1997).
[2-岡田 95] 岡田:口ごもるコンピュータ、共立出版 (1995)
[2-Pearl 88] Pearl、 J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference、 San Mateo、 Calif.: Morgan Kaufmann Pub. (1988).
[2-Proven 94] Provan、 G. M.: Tradeoffs in Knowledge-Based Construction of Probabilistic Models、 IEEE Trans. of SMC、 24-11、 pp. 1580-1592 (1994).
[2-Rouse 87] Rouse、 W. B.、 Geddes、 N. D. and Curry、 R. : An Architecture for Intelligent Interfaces、 HumanComputer Interaction、 3-2 (1987).
[2-Russell 91] Russell、 S. J. and Wefald、 E.: Do the right things: Studies in Limited Rationality、 Cambridge MA、 MIT Press (1991).
[2-椹木 95] 椹木、片井、岩井、向井、福澤:意思決定分析手法に基づく制

御知識構造の導出法とし尿処理プラントの協調制御系への応用、システム制御情報学会論文誌、 8-10、 pp. 574-584 (1995) .
[2-Sawaragi 96] Sawaragi、 T.、 Takada、 Y.、 Katai、 O. and Iwai、 S.: Realtime Decision Support System for Plant Operators Using Concept Formation Method、 Preprints of International Federation of Automatic Control (IFAC) 13th World Congress、 Vol. L、 pp. 373-378、 San Francisco (1996).
[2-Sawaragi 97] Sawaragi、 T. and Katai、 O.: Resource-Bounded Reasoning for Interface Agent for Realizing Flexible Human-Machine Collaboration、 Proc. of the 6th IEEE International Workshop on Robot and Human Communication、 pp. 484-489、 Sendai (1997).
[2-Sawaragi 98] Sawaragi、 T.、 Tani、 N. and Katai、 O.: Evolutional Concept Learning from Observations through Adaptive Feature Selection and GA-Based Feature Discovery、 to be appeared in Journal of Intelligent and Fuzzy Systems (1998).
[2-Simon 75] Simon、 H. A. and Kadane、 J. B.: Optimal Problem-Solving Search: All-or-None Solutions、 Artificial Intelligence、 6、 pp. 235-247 (1975).
[2-塩瀬 98] 塩瀬、椹木、片井、岡田：双参照モデルにおける相互学習のダイナミクス：マルチエージェント環境下における個と集団の相互限定、計測自動制御学会第 22 回知能システムシンポジウム (1998).
[2-Smith 97] Smith、 P. J. et al.: Brittleness in the Design of Cooperative Problem Solving Systems: The Effects on user Performance、 Trans.. of IEEE on SMC、 27-3、 pp. 360-370 (1997).
[2-Woods 95] Woods、 D. D.: Towards Theoretical Base for Representation Design in the Computer Medium: Ecological Perception and Aiding Human Cognition、 in Flach、 J.、 Hancock、 P.、 Cairo、 J. and Vicente K. (Eds.): Global Perspectives on the Ecology of Human-Machine Systems、 Vol. 1、 Lawrence Erlbaum Associates、 pp. 157-188 (1995).

3．マンマシンインターフェースと時間制約

3－1 はじめに

本報告書の 2 章で述べた、人間中心の自動化設計を実用化するに当たって困難な点は主に次の 3 点に要約される：
1)共に自律的に活動する人間と機械がどのように協調してゆくのか
2)人間に対する「情報の洪水」へどう対応するか
3)人間の熟練度に応じた対応の多様性をどう受け入れるか

そこで本章では、インタフェースエージェントの導入、時間制約下におけるエージェントによる推論及び意思決定、エージェント構築のための方法論、定式化、等について議論する。

3－2 人間と協調するインターフェースエージェント

ここで言うインタフェースエージェントとは、オペレータの行動を観察して学習するとともに、オペレータに対して適切な支援を与えたり、代行することができる知的なコンピュータプログラムを指す（図３－１参照)。

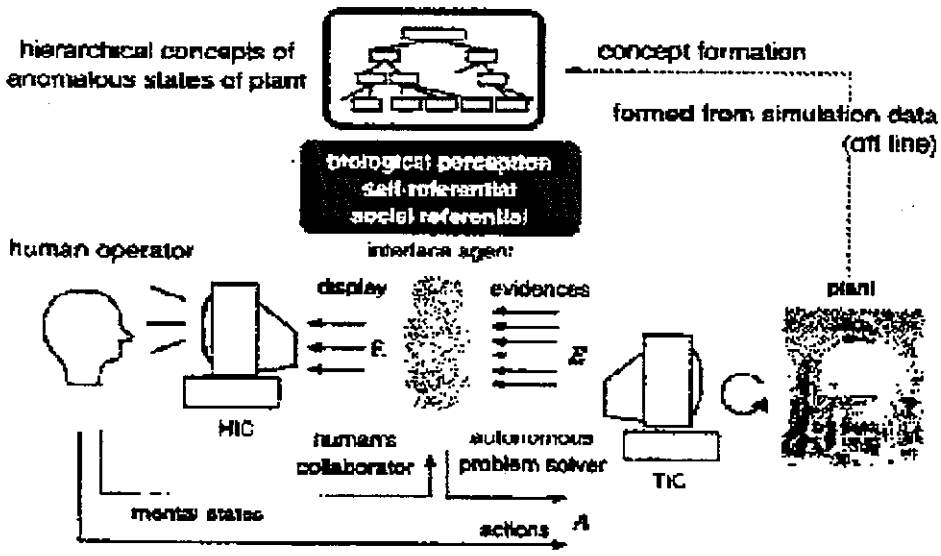


図３－１:人間とシステム間の対話環境に適合するインターフェースエージェント

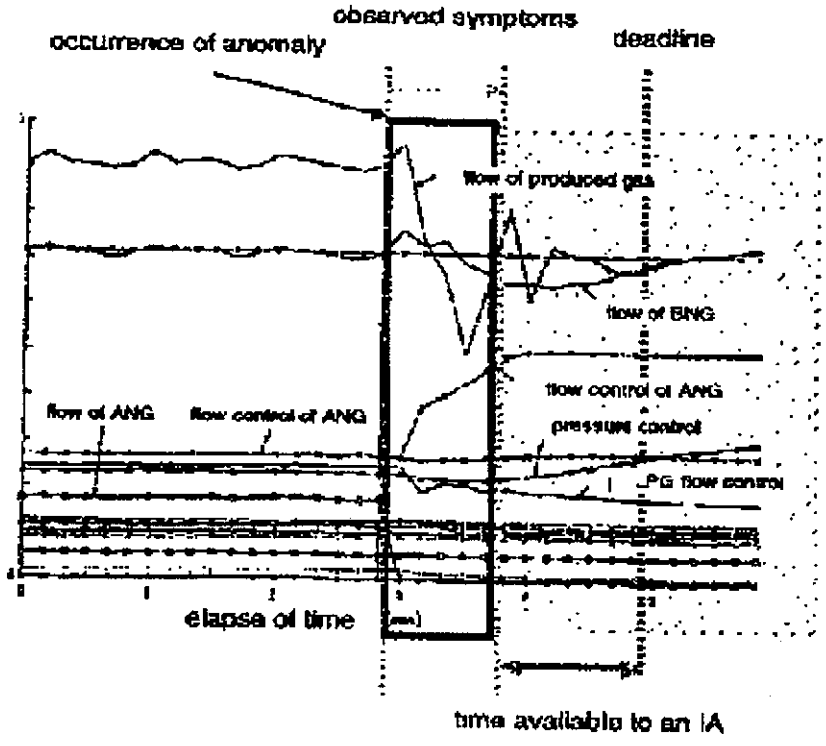
エージェントデザインの変化

エージェントデザインに当たって考察すべき人間の意思決定については、従来の規範的意思決定から、人間の本来的な性向を視野に入れた「自然な」意思決定をそのモデルとして採用すべきである。さらに、熟練専門家の状況評価能力、状況の見方、良質の知識ベースによる迅速な説明等を参考にした意思決定（recognition-primed decision(PRD) model）を前提とする必要がある。

3－3 プラント異常状態の動的分類(Dynamic categorization of plant anomalies)

インタフェースエージェントのポイント

対象となるシステムの「appearance（外観）」の構築に当たっては、対象の特徴を単純に集めて提示するのでは不適切であり、個々のオペレータの時間制約に応じて提示すべきである。この提示によって、人間と機械の協調のためにエージェントは、オペレータの「situation awareness（状況認識）」を助け、それによって適切な対応操作を気付かせるように情報を提示しなければならない。時間制約がある程度以上厳しく、またハイリスクな状況でのオペレータに多量の情報の提示は、オペレータが混乱することにより、時間浪費、判断エラーを誘発するのに対し、逆に価値の低い情報をふるい落として提示することによりオペレータのパフォーマンスは向上する。



プラントで異常が発生した場合は、その生起と同時にプロセス信号の挙動が変化するので、この時エージェントには、
・生起異常事象の種類の同定
・結末の予想
・適切な回復操作の決定・実行
におけるオペレータへの支援が求められる。

制限時間の存在

時間制約が厳しい状況では、オペレータの状況認識(situation awareness)を支援するように適切な外観(appearance)での情報提供が重要となる。「何が大切で、何がそうでないのか」を判断する機能がここで必須となる。本章では、適切な情報提供形態の決定において概念形成(concept formation)を採用したシステムを取り上げる。

概念形成(concept formation)とは、機械的学習法を用いて階層的にプラントの異常事象を分類する機能である（図３－３参照)。

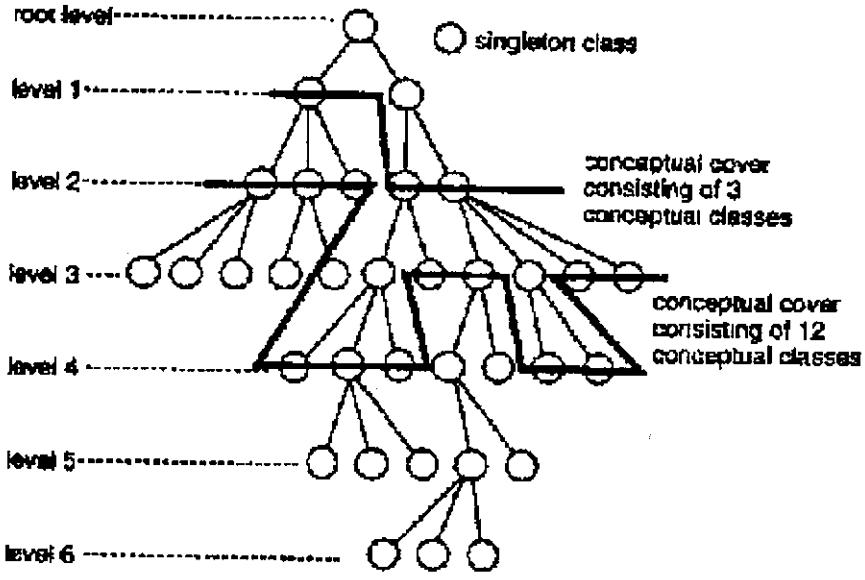


図３－３：ｲﾝﾀｰﾌｪｰｽｴｰｼﾞｪﾝﾄ毎の意思決定詳細度に適合した異常状態概念の「括り方」

図３－３に示した異常状態概念の階層構造には、"root node"：全ての異常事象をカバーする概念クラス、及び"leaf node"：個々の異常事象のケースがあり、この構造に基づけば、適切な外観(appearance)の決定とは、階層式分類法における異常事象の、適切な概念の括り(conceptual cover)の決定であるということができる。

３－４ 認識に基づいた意思決定モデル(Recognition-primed decision making model)

ここでは適切な conceptual cover 決定のための手法について述べる。これは、作用関係図(Influence diagram)に基づいて、部分的観測からもっともらしい異常事象を推論する手法である（図３－４）。

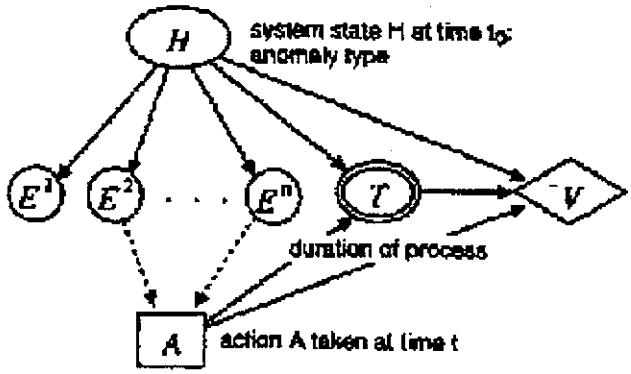


図３－４：作用関係図(influence diagram)で表された資源制約下のｲﾝﾀｰﾌｪｰｽｴｰｼﾞｪﾝﾄの推論

アルゴリズムの概略は次のように説明できる：

各仮説の事前確率・各仮説と各観測量間の確率的関係を知識として準備しておき、これと観測結果から Bayes 理論を用いて状況認識をアップデートする。その上で候補となる仮説・必要時間・対応操作の utility（有用性）を評価して、最大のものを選択する。

ここで、各ノードの意味は、

H（仮説）：プラントの異常事象の仮説

E（症候）：各センサにより獲得される

A（決定）：対応操作を決定

T（時間）：状況認識、行動決定に要する時間（遅れ時間）を表現

V（価値）：有用性

となる。

時間依存の utility の例を次に挙げる。

異常事象は H1、H2 の二種類、対応操作 A1、A2 の二種類を仮定する。

utility at time t=t0	H1	H2
A1	u(A1,H1,t0) time variant	u(A1,H2,t0) time invariant
A2	u(A2,H1,t0) time invariant	u(A2,H2,t0) time invariant

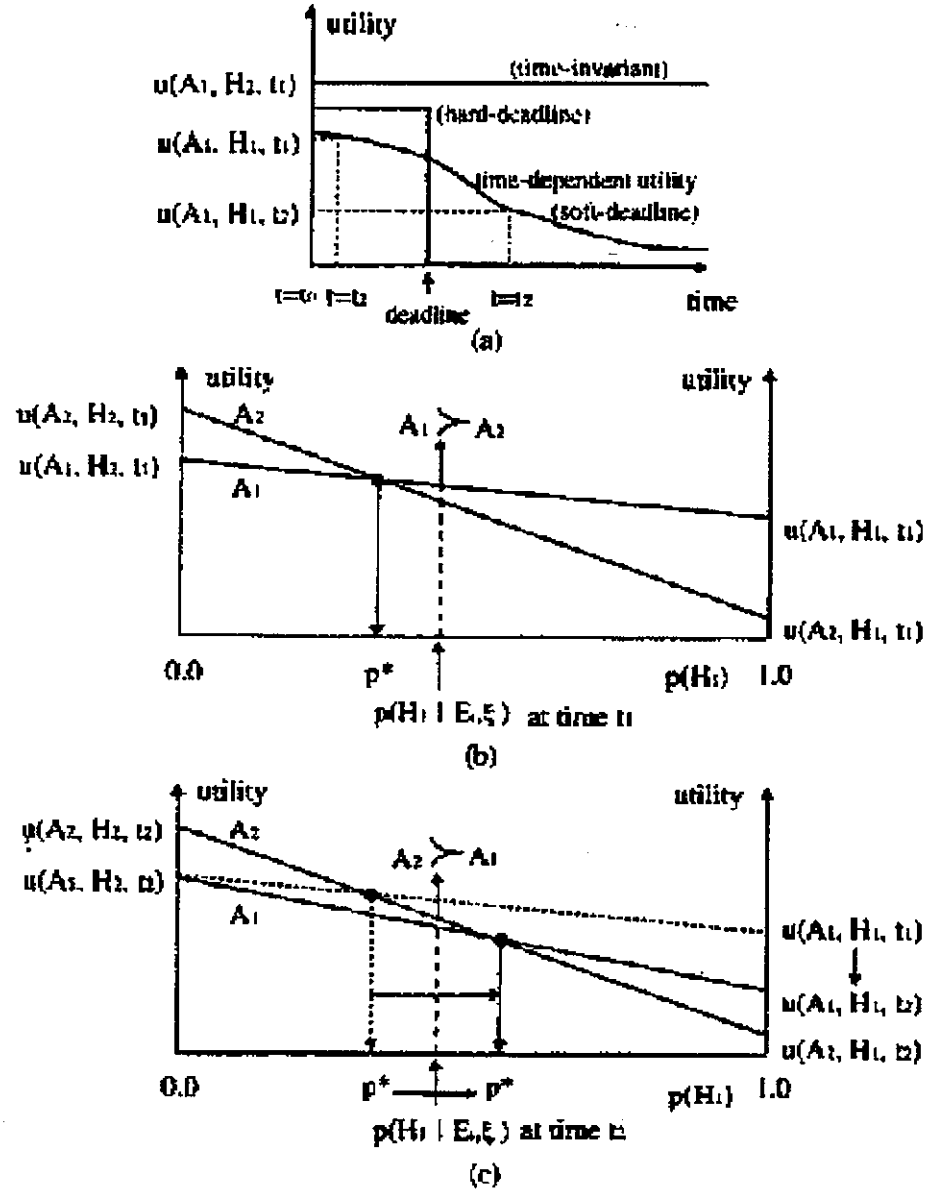


図３－５：(a)時間に依存する有用性、(b)時刻 t1 における２つの行動の有用性期待値、(c) 時刻 t2(>t1>t0)における２つの行動の有用性期待値

ここで、

・(a)utility の時間依存の種類：
　　時間依存（soft deadline）
　　時間依存（hard deadline）
　　時間独立
・(b)A1、A2 各々の Expected Utility
　　仮説 H1、H2 の確率に依存
・(c)u(A1,H1,t2)が時間依存
　　時関経過と共に Expected Utility も変化
方針：utility の最も高い行動を選択

認識に基づいた意思決定(recognition-primed decision making)

認識に基づいた意思決定とは、本研究においては、異常事象の分類の抽象度を反映した意思決定と言いかえることができる。ここで、EVC(Expected Value of Categorization)と称する定量的な尺度が重要となる。これは、先に述べた「概念の括り方："conceptual cover"」が与えられた場合(z)に推奨される決定・行動の有用性の期待値を表すもので、次の式で定義される。

$$A_z^* = \arg\max_A \sum_j u(A, H_j) p(H_j \mid E, \xi)$$

として、

$$EVC = \sum_j u(A_z^*, H_j) p(H_j \mid E, \xi)$$

但し、
A*：推奨される選択肢
とする。

3－5 時間制約下におけるｴｰｼﾞｪﾝﾄによる「複雑さ」の管理(Agent's managing complexity under time criticality)

人間と機械の知的な協調関係においては、エージェントはオペレータ対して推奨される行動を強いるのではなく、適切な異常の候補への注意の促進を支援することを目的とする。これによって初めて適切な conceptual cover の決定が行われる。

conceptual cover は、次の基準に基づいて決定される必要がある：
・ agent はオペレータが情報を獲得してから行動に移るまでに遅れ時間があることを考慮する必要がある
・対応操作による効果が取るに足らないような異常事象については、より抽象

的な階層にまとめてしまう
・生起確率が高い、対応操作が異なる異常事象が集まっているクラスに関しては、これをより詳細に分類する
・「正解」を多くのケースと一緒になったクラスに埋もれさせたり、逆に重要でないケースを目立たせることを避ける なお、適切な conceptual cover 決定においては、より詳細な分類を行うことによる、
・決定の expected utility の増大
・時間・計算機などの資源の消費
のトレードオフがあることに注意する必要がある。

3－6 数の文脈がある場合の資源制約下のｴｰｼﾞｪﾝﾄ意思決定の定式化 (Formulating an agents resource-bounded reasoning under multiple contexts)

適切な conceptual cover 決定の公式化

ここで、オペレータ、エージェント双方の decision making に要する時間に着目して、
Cd(z)：オペレータの反応の遅れ時間のコスト
Cc(z)：エージェントが decision model を解くのに必要な遅れ時間のコスト
Ct(z)：トータルのコスト (=Cc(z)+Cd(z))
とする。EVC を修正した NEVC(net expected value of categorization)は全 conceputual cover に関して計算するのが理想だが、計算コストが非常に高価となる。

ここで、二種類の操作を繰り返すことにより、理想的な conceputual cover を探索することを考える。図3－6において、
specialization：あるクラスを複数のサブクラスに分割すること、
generalization：複数のクラスを一つのクラスに統合すること、
とする。

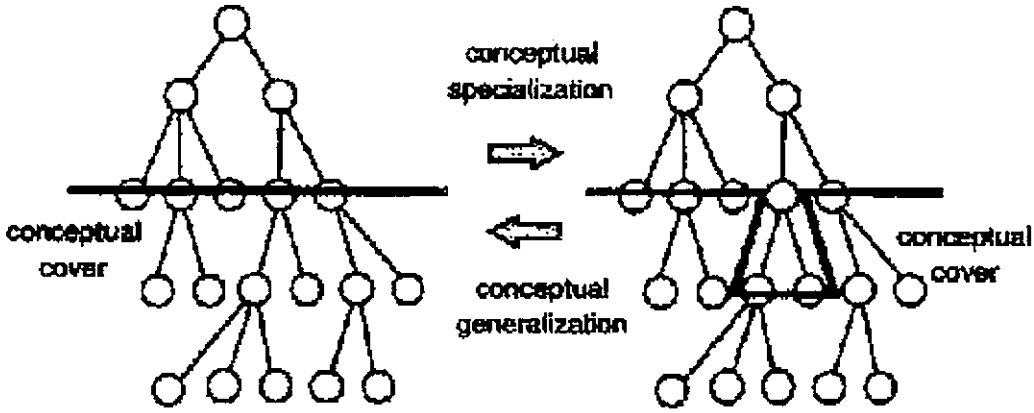


図３－６：概念の分割(specialization)と統合(generalization)

conceptual cover z0 が分割・統合操作 si により zi に変形した時、つまり zi=si(z0) としたときの有用性の変化を求めて EVMI(si)とする。その上で conceptual cover 変形によるコスト Ct、モデル修正によるコスト Cg を用いてこの変形による計算コストを算出した結果を NEVMI(si)とすると、全ての分割・統合操作に関して NEVMI<=0 となる conceptual cover が最適解である。

時間制約が厳しく、ハイリスクなシステムにおける human-centered デザインにおいては、オペレータには理解のための十分な時間を用意する必要があり、また多量のデータを提示することは混乱を引き起こすことを常に考慮する必要がある。

４．結論

本報告書では、複雑・大規模なシステムのプロセスコントロールにおいて、人間中心の自動化を進める場合問題となる、インタフェースエージェントが資源制約下、特に時間制約が厳しい状況においてオペレータに情報を提示する際の推論についての定式化を中心に考察と一部システム試作を通じた評価を行った。

最後に、あるベテランパイロットのコメントを引用する［桑野 97］。『無数のハードウェアとソフトウェア、そして人間が有形無形に複雑に組み合わされて巨大化の進むシステムは、デザインした技術者ですら動かしてみないとわからないという事態を引き起こしている。これに対して人間と機械の関わりあいを極力排除することによる信頼性確保が目指されているのが現状である。そして皮肉にもシステムの信頼性が向上すればするほどその運用に関わる人の信頼性は低下する傾向にある。この悪循環を断ち切るには、人間が冗長なシステムとして機能することを前提とするシステム設計に切り替えていかねばならない。その上で、ヒューマンエラーをあってはいけないものとするのではなく、人の自然なありさまとして認め、それを事故に結び付けないようにバックアップしていくことがとるべき最善の方向性のように考えられる。』

◇参考文献◇

［桑野 97］　桑野　他：機長の危機管理、講談社 (1997).

Appendix:原子力プラント緊急時における意思決定の具体的シナリオ

以下に記する運転員意思決定のシナリオは、ウェスティングハウス科学技術センターの Emilie M. Roth によって分析されたもので、人工的な意思決定支援環境がある程度ある場合の例である。

「生来的な意思決定:Naturalistic Decision Making」に顕著な特徴は、現実的な環境では、意思決定を行う人間が、その対象とする系に完全に孤立した状態で対処させられることは滅多にない、ということである。意思決定の過程を支援する何らかの補助が与えられることが多い。そのような補助には、意思決定の全ての局面を方向付けようとするかなり規定的な方向へ意思決定を拘束する方針や指針が含まれる。そこで現れた一つのやり方は、予め計画された診断や対処の戦略を提供して、運転員に逐一従わせるものであった。例としては、チェックリスト、印刷された手順書、そして従来型のエキスパートシステムがあり、これは航空機の離陸に始って、保守作業や緊急時管理に至る分野で人間の行動を「案内」するために用いられている。

ここでは、「生来的な意思決定」に関する最近の研究で、高度に規定的な手順書が意思決定に与える影響を調べたものを総括する。ここでは、応用分野を原子力プラントの運転に設定する。発電所で緊急事態が発生した時、運転員は手順書に逐一従うことを要求される。手順書には、どのパラメータをチェックするか、観測した兆候をどのように解釈すべきか、どの制御操作を行うべきかについて、詳細な手引きが書かれている。運転員がそのような極めて規定的な手順を用いるとして、緊急事態に適切に対処するために運転員に要求される認知的活動の性質と程度について疑問が生じる。一方の見方は、運転員に必要なのは、手順の個々のステップを理解して実行することだけだ、というものである。これに代わるもう一つの見方は、運転員の仕事が成功裏に行われるためには、状況認識、対応操作計画、システムの機能や構造に関するメンタルモデル等の、より高い認知活動がやはり重要である、というものである。この相対する２つの見方は、訓練、手順書、制御盤ディスプレイ、提供すべき支援等に大きく異なる意味をもたらす。

この章で我々は、例証によって、状況認識や対応操作計画が、詳細な手順書が提供される場合にあってもやはり「生来的な意思決定」の重要な要素であることを示そうと試みる。我々は、運転員が、プラント状態が運転員にとってどのような意味があるかを能動的に頭に描き、それを用いてマルファンクションの同定、将来起こる問題の予想、状況に応じた手順の適切さの評価、及び必要な場合には一連の手順の見直しをしている、という証拠を示す。

運転員による意思決定の調査

原子力プラント運転員の行動を調査する研究のため、手順書では十分に説明されていない点を含む、人間にとってわかりにくい２種類の異常をシミュレータで模擬して実験が行われた。運転員の行動は訓練用シミュレータによって２個所のプラントサイトで採取された。それぞれのプラントから計１１班が、シミュレータ上の緊急事態に当たり、合計３８ケースが分析された。運転班の行動は録画され、運転員間の会話記録が作成された。この記録には運転員の観察行動、検討した仮説、状況の評価、実行した操作が記載された。運転員が彼らの状況評価を総括した時点、手順書の目的や目標を参照した時点、手順書外の操作を行った時点については特に注意して記載した。データの採取と分析の方法に関する詳細はRoth他(1994)にある。

分析によって、並行して行われる２つの流れが明らかになった。１つは手順書によって統制されたものである。運転員は手順書に従うことで主な安全機能を確実に維持しようとし、またそのための行動を確実に行おうとする。同時に、もう一つの、自発的で認知の働きによる流れがある。運転員は、義務付けられた行動に加えて、監視、状況評価、そして対応操作計画の見直しを行っている。運転員らは警報を確認し、その意味を評価し、その状況の成り行きを評価・監視し、そして一連の手順に埋め込まれた戦略がその状況にとって適切か否かを注意して見ている。

運転員が状況認識や対応操作計画を行って、そのために手順書がカバーしきれていない状況への対処が可能になっているいくつかの例が分析によって突き止められた。以下に、観測された認知的行動をいくつか挙げる。この結果についてのより充実した記述はRoth他(1994)にある。

状況認識

いくつかの分野で見出された「生来的な意思決定」の一つの側面は、個々の人間が能動的に、状況が自分にとって意味するものを頭に描く、ということである。これは従来、人間が色々な観察を一つの因果的な説明へと合成する際の「物語構築」、状況の「頭の中のモデル」の発達、状況評価、または状況認識、等、様々な言葉で表現されてきた。

我々は、プラント状態に影響を与える因子について、知り得たものでも想定上のものでも、運転員はいかなる時でも能動的にその意味するところを頭に描くということの証拠を見出した。彼らはこの「頭に描いたもの」を使って、観

察したものを説明し、予測外に発見したことを同定・追跡し、将来の問題を予見する。このことを例証するために、我々が調査した事象の中から一つの例をここに描くことにする。それは、「系統間界面冷却材喪失事故(正式な日本語名称ではない、英語では Interfacing System Loss of Coolant Accident、以後ISLOCA と略す)」を模擬した実験で得られた。

ISLOCA は、格納容器内の炉容器冷却系(RCS)から、格納容器外の崩壊熱除去(RHR)系への漏洩である。この事象を複雑にする一つの要因は、初期の兆候の多く（例：格納容器内放射線の兆候）が、格納容器内での漏洩を強く示唆することである。さらに、このシナリオの特有の動きにより、緊急時手順書によって、運転員は格納容器内漏洩に対処するための手順を行うはめになる。ISLOCAに対処する手順も書かれてはいるが、手順変更の流れ図ではそこにたどり着くことができない。結局、崩壊熱除去系への漏洩であることを検知してこれに対応することができるかどうかは運転員が状況をうまく認識できるかどうかにかかっている。

実際に、このシナリオでの実験に参加した中で９つの運転班が崩壊熱除去系への漏洩であることを突き止めて、その漏洩を食い止める操作を行おうとした。運転班がこの事象が単なる格納容器内の漏洩ではないことを認識できたきっかけと、彼らがそれを確認して崩壊熱除去系への漏洩を食い止めるために行った操作とが、運転員たちの何人かがプラント状態の意味するところを能動的に頭に描いていることの証拠である。９班の内５つの班は、この事象が単なる格納容器内の漏洩ではないことを、予期した兆候が見られないことによって、早くから気づいていた。加圧器の（圧力の？）減少速度から、運転員は、これが単純な漏洩なら現れるはずの格納容器での兆候に関して予測を立てていた。格納容器の圧力が彼らの予想のようには速く上昇しないと見た時に、彼らは格納容器外への漏洩を疑い始めた。

この例が示す点がいくつかある。第一に、運転員は能動的に状況を評価し、その評価結果は彼らの観察した一連の兆候に対する成立しうる説明になっている。この場合彼らは格納容器内の漏洩を仮想することから開始している。第二に、運転員はかれらの説明が正しければ見えるはずの付加的な兆候に関する予測をおこなっている。この場合彼らは格納容器内の急激な圧力上昇を見るであろうと予測している。第三に、彼らの予測が覆された時、彼らは観察した現象を説明できるような更なる（或いは別の）要因（影響）を探している。この場合は彼らは何か格納容器外での漏洩を疑い始めている。この作用によって彼らは漏洩箇所を熱心に探した。

運転員はまた、プラントシステム内部の物理的結びつきやプラントパラメータの量や時間的特性に関する知識を用いてプラント兆候に対して成立しうる説

明を生成したり絞り込んだりしている。例えば、この ISLOCA の場合、９班の内３つの班は、漏洩が崩壊熱除去系へのものであることを、系統間の物理的結びつきに関する知識を当てはめることによって突き止めた。炉容器冷却系からの漏洩が崩壊熱除去系での圧力上昇をもたらしたのである。崩壊熱除去系には圧力開放弁があり、これは格納容器内の加圧器開放タンクへと圧力を逃がすものである。加圧器開放タンクの液面と圧力が上昇して、それは実際に破裂した。運転員が加圧器開放タンクの兆候に気づいた時、彼らは加圧器開放タンクへと圧力を逃がす関係にある全てのプラント機器を検討して、それを崩壊熱除去系の開放弁に絞った。なぜなら加圧器開放タンクに急激な液面上昇を起こしてそれを破裂させるに十分な量の水を流入させるものは他になかったからである。運転員はまた複数の多様な兆候全てに整合する説明を能動的に探している証拠を示している。この ISLOCA の場合では、崩壊熱除去系での圧力上昇が実際に崩壊熱除去系と機器冷却水系との間の熱交換器内部での漏洩を起こした。これによって機器冷却水に冷却水水位と放射線の上昇が起こった。機器冷却水での兆候は炉容器冷却系から機器冷却水系への直接の漏洩という仮説に整合するものであったが、この説明は崩壊熱除去系で見られた兆候にも、加圧器開放タンクでの兆候にも当てはまらなかった。最も少ない数の故障でこれらの兆候全てに当てはまるものを探した結果、運転員は可能性を一つに絞って、崩壊熱除去系から機器冷却水への熱交換器の損傷しかない、と考えるに至った。このように運転員は崩壊熱除去系での兆候（つまり炉容器冷却系から崩壊熱除去系への漏洩）、加圧器開放タンクでの兆候（崩壊熱除去系から加圧器開放タンクへの圧力開放）、機器冷却水系での兆候（崩壊熱除去系での圧力上昇による機器冷却水系の熱交換器損傷）を同時に説明できる因果連鎖にたどり着くことができた。状況評価がプラント状態への複数の影響と将来の影響の予測を追跡する動的な活動であることの証拠もある。いくつかの場面で運転員は観察されたプラント挙動が、はたして手動や自動の操作と既知のマルファンクションのようなプラントへの既知の影響によるものか、或いは予想外のある(signaled)プラントマルファンクションがまだ彼らにはわかっていないのかを判断しなければならなかった。これはプラントが現に受けた影響から予測されるプラント挙動の規模と方向の評価に依存した。例としては、運転員が加圧器の液位と圧力に外乱を与えるような操作をしている時に、我々は加圧器開放弁の漏洩を入力したときのことがあげられる。漏れている開放弁を突き止められたのは、加圧器の挙動の変化の大きさがその時の操作の副次的効果としては説明できないものだということを認識できたからであった。

対処計画の生成／監視

緊急時手順書には緊急時に運転員が行うべき操作が規定されている。今回模擬した緊急事態では、運転員は事に当たっている間、常に手順書に従っていた。にもかかわらず、運転員はまた、プラントの状態を監視してその時々に従っている手順書の記載事項の適切さを彼らが理解するところのプラント状態に照らして評価していた。運転員が彼同士で、手順のあるステップの適切さについて、さらには安全目標を達成する行動をとるためには彼らが従っている手順でよいのか、を議論する行動が観察された。我々のデータには手順書に記載された対処計画を監視していることによって運転員が下記の能力を発揮する例が含まれている：

・誤りの発見とそこからの回復－－誤りには運転員の誤りと手順書の誤りの双方を含む
・運転員が辿っている手順の流れが正しいかどうか、あるいは他の手順の流れへの重要な移行を見失ったかどうかの評価
・手順書と実際の状況との差を埋めて適応すること
・手順書の手引きを超えた予期しない状況への対処

模擬した緊急事態で見られた挙動には、運転員が後の手順ステップを見越して一連の行動をとる、というものがあった。これらの行動はプラントの状態とその意味、そして一連の手順への理解に基づいていた。そのことによって運転員は手順書に記載された行動をより手際良く実行できた。一つの例として、このISLOCAの場合は、一度彼らがISLOCAの可能性を思い浮かべてから、運転班のいくつかはすぐに格納容器外での漏洩可能性を示す証拠を探すために補助建屋を電話で呼んだ。手順書にも補助建屋への電話連絡が記載されていたが、それはもっと後でするように記載されていた。このステップを予期して、運転員は早期に連絡をすることで、手順書によってそれが求められた時点で既に補助建屋の調査結果を手に入れることができた。

運転員はまた、手順書に込められた対処戦略全体の文脈のみならず、その時点での事象の文脈に照らして手順書の各ステップに記載された行動のどれに意味があるかを注意して見ている、ということも観察されている。これは運転員がエラーを検知してそこから回復するのを支援するに際して重要な要因である－ここで言うエラーには運転員のエラーと手順書のエラーの双方を含む。我々は、運転員が不注意から手順書の重要なステップを飛ばしたのを自分で気づくところを見た。彼らは自分達のエラーをその続きのステップを読み始めた時に

気づき、その行動がその時の状況に適していないことに気づいた。我々は、また、運転員が手順書のステップで指示された行動がそのままでは不適切だと気づくことで、誤ったステップの組み合わせを見出すケースも発見した。

運転員は自分が辿っている手順の流れの適切さを監視することも我々は観察した。我々が模擬した事象では、運転員が操作を中断して、今辿っている一連の手順で本当にプラントの安全な回復に重要と思っている操作ができるのかどうかについてグループで議論をするケースが再三見られた。緊急時の手順には手順から手順への一連の移行が含まれるので、与えられた手順が本当に、一連の移行を通じて、その事象に関連した手引きを含んだ緊急時手順に到達するものであることを確認するのは時に困難になる。この確認は、事象が刻々変化することで移行のステップで前提条件となる兆候がそのステップが終了後にやっと現れることがあり、或いは、他の兆候が先に現れてそれによって違う手順への移行が行われ、そこからは採るべき手順には行きつけないことがあるので、遥かに困難な作業となる。

結論とその意味

この研究は２つの事柄を示している。第一に、少なくともここで取り上げた分野では、状況の意味するところを能動的に頭に描くことが「生来的な意思決定」において中心的な役割を果たすことが立証されていることである。この研究では、ある分野で報告されているような、純粋に認識が先行する意思決定の根拠となるものは僅かしか見られなかった。運転員の意思決定は、どの時点においても、プラント状態に影響する既知の或いは想定上の要因のモデルを能動的に作ってそれに基づいて行われていた。これらの結果は、認識先行型意思決定に診断や「物語構築」の要素を含んだ詳細化モデルとは整合する。

第二に、この研究は、予め計画された手順というものの限界を指摘している。そのような手順は運転員がリアルタイムで診断や対処戦略を立てる負担を軽減しうるにしても、能動的に状況を評価し、対処計画を注意深く見る必要を無くするものではない。我々のシナリオでも、手順書でカバーしきれない事態がいくつか生じた。このような状況では運転員は：

・知識に基づいて監視した。
・プラント兆候と整合する説明を探した。
・状況評価に基づいて予測を立ててそれを意思決定に用いた。
・安全目標の達成への手順の有効性を評価した。
・手順書を（解釈して）実際の状況へ適応させた。

これらの結果は訓練や制御室での支援（手順書、ディスプレイ、意思決定支援）

に関連する意味を持つ。

我々の研究で、プラントの故障箇所を絞ってそれを緩和するための操作を決めるために、運転員が物理的なプラントモデルのメンタルモデルを必要とし、プラント状態に影響する様々な因子について効果を予測するために定性的に推論しなければならない状況があることの根拠が見出された。訓練、制御室のディスプレイ、及び意思決定支援ツールによって、正確なメンタルモデルを育てて支持する必要がある。

もう一つの必要な知識の種類は、プラントの主要な工学上の目標とそれを達成するための手段についてのものである。我々の研究で、運転員がプラントの目標について推論し、それを達成するための代替手段を評価しなければならないことが示された。運転員はまた、手順に関する知識が必要で、これには手順の個々のステップを如何に実行するかという知識だけでなく、手順の前提となっている仮定や論理の知識も含まれる。この（後者の）知識は、手順それぞれに固有のプラント目的間の優先度の知識、手順それぞれに組み込まれた対処計画とその理論的解釈、及び手順間移行ネットワークの知識から構成される。研究から導かれたこの結果は、訓練、基礎となる書類、及び手順書の構成や様式においてこれらの種類の知識に明示的に触れることの価値を示唆するものである。