

機械学習システムの動向と課題

山口 高平

(静岡大学工学部情報知識工学科)

本稿では、機械学習に関する筆者らの今までの研究である、知識コンパイル（故障診断）、ルール洗練化（株式市場分析）、事例ベース推論とモデル推論の統合（法律解釈、ソフトウェアプロセス）を通して、知識によるコンピューテーションとモデリングの関係について考察し、知識獲得ボトルネックを解消するためのアプローチとしての機械学習システムの見通しについて述べる。

Where Should Machine Learning Systems Be Going ?

Takahira YAMAGUCHI

Faculty of Engineering, Shizuoka University

Here is discussed the direction which machine learning systems should take in the future, from my experience about the research of machine learning including knowledge compiling, rule refinement, and integration of CBR and model inference. Finally this paper comes into the conclusion that machine learning systems should take two facilities: powerful computaion and knowledge modeling.

1. はじめに

近年、エキスパートシステム（ES）の産業界への浸透はかなり進み、数百程度の実用レベルのESが開発されるようになってきた。しかしながら、ESの成否の最大要因である質の高い専門家モデル（＝問題解決（タスク）構造＋知識構造）を構築する課題については、決め手となる技術がなく、専門家によるレビューなどのように、人手に任せられて、試行錯誤的に行われているのが現状である。

即ち、ナレッジエンジニア（KE）は専門家とのインタビューを通して、計算機世界をあまり意識しないで、問題領域世界における専門家モデルを忠実に抽出することを試みるわけであるが、専門家自身が知識を体系化していないこと、KEが問題領域に精通していないため専門家の話を完全に理解できないこと、専門知識を計算機世界にマップすること自体が困難であることなどから、質の高い専門家モデルを構築するには多大の時間を要してしまうのが現状であり、知識獲得ボトルネックと呼ばれている。

この知識獲得ボトルネックを解消するためのアプローチとしては、機械学習（それと関連した高次推論）と知識獲得支援機構がある。機械学習は、タスクを何らかの意味で（機能レベルであるいは効率レベルで）前回よりは旨く解決できるようになるシステム内の変化を実現する機構である。一方、知識獲得支援機構は、専門家と相互作用を取りながら、専門家が持つタスク構造と知識構造を同定することを支援する機構である。

しかしながら、両者の研究のアプローチには、大きな差異がある。すなわち、多くの機械学習システムでは、システムの自立性に重点が置かれるのに対し、知識獲得支援システムでは、有益なタスク構造や知識構造を専門家から引き出すための相互作用環境を整えることに重点が置かれる。機械学習と知識獲得のギャップはまさにこの点にあり、意見のズレ違いが生じる原因になっている。

特に、知識獲得支援サイドから機械学習サイドに対して、機械学習では、かなり整理された情報が事前に与えられなければならない事がよく批判される。例えば、ルールの構成単位である述語等は、機械学習においては、事前に揃っていることが前提になっている場合が多いが、それらの述語を完備することは、知識獲得において最も骨の折れる作業のひとつであり、それを前提にして議論が進むのは、非現実的と映るのである。以上の背景から、機械学習の研究に対して、全面的に、知識獲得ボトルネックの解消を期待することは、時期尚早であるという意見が現場レベルでは多い。

しかしながら、この研究の差異は、機械学習システムはコンピューテーション、知識獲得支援システムは知識によるモデリングを重視している事に起因しているともいえる。

幸い、平成4年度のICOTWG活動として、機械学習の応用ワーキンググループ（ICOTMLAWG）の活動に参加させて頂き、記号処理に基づく帰納的学習、演繹的学習、計算論的学習理論、ニューラルネットワーク、遺伝的アルゴリズム等、様々な機械学習の観点から、遺伝子情報処理や法律解釈等の実問題への適用可能性について、深く議論する機会を持った。そしてこの議論を通して、上述の批判を乗り越えるためには、「知識によるコンピューテーションとモデリングの統合」が、機械学習システムにおいて、検討されていかねばならないと強く感じた。

本稿では、機械学習に関する筆者らの今までの研究である、知識コンパイル（故障診断）、ルール洗練化（株式市場分析）、事例ベース推論とモデル推論の統合（法律解釈、ソフトウェアプロセス）を通して知識によるコンピューテーションとモデリングの関係について考察し、知識獲得ボトルネックを解消するためのアプローチとしての機械学習システムの見通しについて述べる。

2. 知識コンパイル^[1]

知識コンパイルは、ESにおいては、対象領域を基本的に理解するための深い知識から、タスクを効率よく処理するための浅い知識に変換する技術である。

文献[1]の知識コンパイラIIは、深い知識の構築容易性と汎用性、並びに、専門家が見落としやすい点を指摘できる能力の実現に留意し、機械の故障診断にタスクドメインを限定して、5種類の深い知識から故障診断木（故障診断ルール）を自動生成することを実現している。

5種類の深い知識とは、Device World（対象装置の構造、部品の機能などを表す知識）、Physical World（適用条件と共に記述された物理式）、Control World（物理パラメータの操作容易性や部品の耐久性によって、深い推論を制御する知識）、Interpretation World（機械の内部物理状態を故障概念に対応づける知識）、Failure Mechanism World（故障メカニズム用汎化知識）であり、図1のように、これらの知識は予め知識コンパイラII内に蓄えられ、異常兆候が与えられると、DW/PW/CW/IW（対象モデル）を用いた定性値伝播による推論とFMW（故障モデル）による推論を交互に適用することによって、故障原因まで究明する故障木を自動的に生成する。

この知識コンパイラIIを実規模に近い空調装置に適用し、「圧縮機の回転数減少」という異常徴候を与えると、図2のような故障木が生成され、「熱膨張率の差により冷媒が漏れ、このために部品Aがオーバーヒートを起こし、部品の一部が融解して経路に流れ出し、経路を流れている冷媒に冷却されて凝固して詰まりを起こし、この詰まりが原因となって圧縮機の回転数が減少した」（図3参照）という専門家にとっても複雑な故障メカニズムが生成され、専門家が見落としやすい点を指摘できる能力を部分的に実現したといえる。

この成果は、故障メカニズム汎化知識から生成される故障モデルが意外に専門家に意識され

ていないこと、並びに、主にDWとPWから生成される対象モデルと故障モデルとの相互作用から構成される故障探索空間が大きいことに起因するといえる。即ち、この成果に関する限り、故障診断における深い知識のモデリングが大きく寄与しており、定性値伝播等のコンピュータシミュレーションが寄与する所は小さい。しかしながら、生成された2000個程度のノード群から構成される故障木には、現実には起こり得ない故障シーケンス（パス）も含まれており、コンピュータシミュレーションによる故障木の評価機構が望まれる。

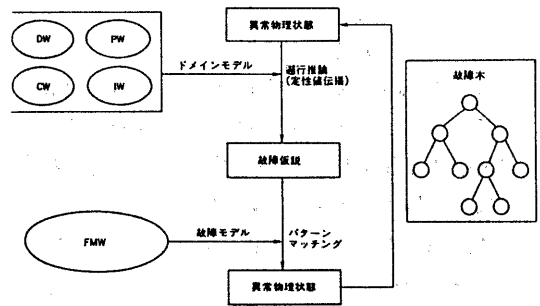


図1 知識コンパイラIIの概観

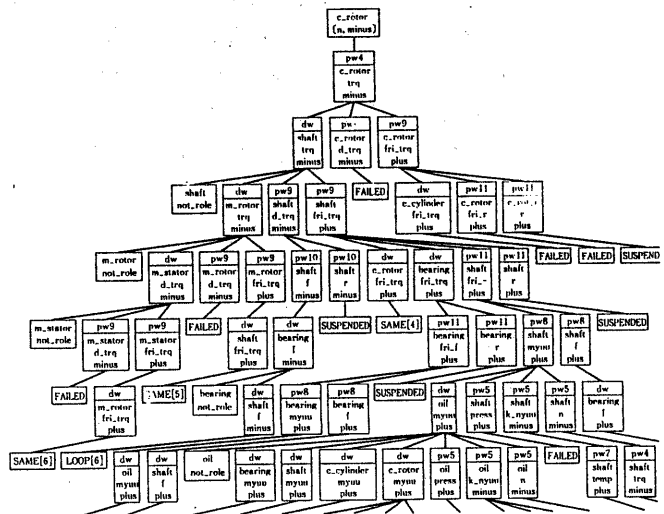


図2 生成された故障木の一部

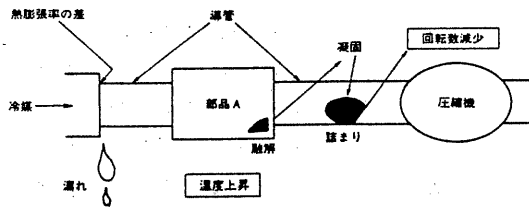


図3 生成された故障シーケンス

3. ルール洗練化^[2]

ルール洗練化は、学習（評価）用DBを利用して、与えられた初期RBをより高性能のRBに変換する機構であり、帰納的学習（における一般化/特殊化オペレータ）と深く関わる。

文献[2]では、株式市場テクニカル（Tと略記）分析の一手法である Granvilleの法則を表現したルールを洗練化して、原因理由が明らかでない株価変化の規則性（アノマリー）を発見することを試みている。

本システムでは、T分析と関連が深いファンダメンタル（Fと略記）分析の要因（以下属性と呼ぶ）をルールの適用環境要素とし、それらの属性をどのように組み合わせてスクリーニングすれば、ルールが旨く処理できるデータセット（DSと略記）を選定できるかを、ルールインダクションにより学習させ、そのDSにおいて、ルールの洗練化を試みる。

システムの概観を図4に示すが、まず、DS群と初期のオリジナルルールベース（IRB）をルール実行部に与えて実行し、DS群のある評価基準により成功DSと失敗DSに分類する。ただし、評価基準は、「株購入時の運用資金を100（以下、運用指数と呼ぶ）とし、3ヶ月後の運用指数の増加率が、日経平均株価の増加率と等しいか上回れば成功、下回れば失敗」と

している。次に、分類されたDS群に、表1のような属性に関する値を付加し、ルールインダクションにより決定木を作成する。決定木の根節点から葉節点までのパスが、葉節点に含まれるデータ群を選択するためのひとつのルール（以下、DS選定ルールと呼ぶ）となる。

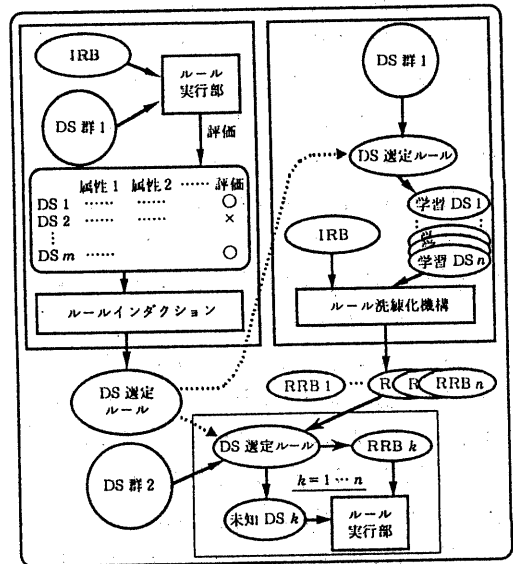


図4 ルール洗練化システムの概観

表1 データセットの分類属性

銘柄群（業種別）
日経平均株価の変化状態 （傾きとその持続期間）
金利（コールレート）
金利（長期国債）
為替相場（円/ドル）
消費者物価指数
卸売物価指数
マネーサプライ

表2 ルールの洗練化実験

	適用数	1月後	2月後	3月後
① IRB	517	101	104	108
	385	102	105	112
	385	102	105	112
② RRB	677	101	104	108
	543	104	105	108
	543	104	105	108
③DS RB +IRB	96	102	106	109
	17	116	105	108
	73	99	97	110
④DS RB +RRB	16	112	121	134
	16	141	183	176
	53	100	97	115
⑤ 日経	—	102	104	106

①～④の学習DS群： 上段 中段 下段

1984年1月～1987年9月の 1週 3週 同左

①～④の未知DS群：

1984年1月～1987年9月の 2-4週1,2,4週同左

さらに、得られたDS選定ルールベースを用いて、DS群からIRBを洗練化するための学習DSを選択し、IRBとその学習DSをルール洗練化機構に与え、4種類の洗練化戦略を多段適用しIRBを洗練化する。最終的に、DS選定ルールベースにより未知DSをスクリーニングさせて、洗練化ルールベース(RRB)を実行し評価する。

本システムを評価するために、以下の3項目をルールインダクションに与えて、DS選定ルールベース(決定木)を学習させた。

- (1) 買いのGranvilleの法則を表現した5個のルール(IRB)
- (2) 銘柄群別に分類された東証一部上場全銘柄の週足データ群から適当に選定したDS群
- (3) 表1に示した属性の値を0-99に変換した値次に、以下の2項目をルール洗練化機構に与えて、洗練化ルールベース(RRB1～RRBk)を学

習させた。

- (4) DS選定RBによりスクリーニングしたDS(学習DS1～学習DSn)

- (5) 買いのGranvilleの法則を表現した5個のルール

最終的に、洗練化に利用したDS選定ルールが付随した洗練化ルールを、(2)以外の未知DSにより評価した。その結果を表2に示す。

IRB(①)とRRB(②)を比較すると、いずれの場合も洗練化の効果が現れていないのに対し、DS選定ルールを付随させれば、RRBがIRBの性能を安定的に上回る(①と④の比較)ことが判る。この結果は、株式市場環境があまり変化していないデータ群に対しては、同様に成り立つことを確認している。

しかしながら、平成2年以降のように、株式市場環境が大きく変化したデータにより、本システムを評価すると結果は芳しくない。株式市場分析の専門家から、2年間連続運用して儲け続けられるシステムはほとんどないことを聞いたことがあるが、その理由はやはり、株式市場環境の変化にある。特に、平成元年までのデータを利用して何らかの学習をさせ、その学習結果により平成2年後半以降の予測を行うと、ほとんど結果は悲惨であったと聞いている。

通常の帰納的学習システム(本システム)は、与えられた属性によるモデリングの範囲で、問題(株式市場)分析に現れるBehaviour(市場データ)を分析して、部分的な規則性(アノマリー)を見つけるためのコンピューテーションに力点を置いており、そのBehaviourを生み出したStructure and Function(株式市場モデル)の構築には目を向けない。モデルが有効である場合にはよいが、一旦、モデルが使えない状況になれば、学習結果はほとんど役に立たない。

システムに、正確な株式市場モデルを持たせることは到底不可能であるが、不完全なモデルでもよいからシステムに持たせて、コンピューテーションの結果をフィードバックさせるような仕組みが必要であろう。

4. モデル推論とCBRの統合^{[3][4]}

前節では、コンピューテーションとモデリングの相互作用の必要性を示唆したが、本節では、事例に基づく推論（CBR）の事例構造を問題領域のモデリング、論理上の帰納推論であるモデル推論をコンピューテーションと位置づけて、その相互（協調）作用について考察する。

CBRとモデル推論の協調動作に基づく知識獲得支援システムの概観を図5に示す。以下、図中の番号に沿って説明する。

- (1) CBRでは、初期事例構造に基づいて、断片的事例群（事例ベース）を所持している。CBRの主要動作は、問題に対する類似事例（ベストマッチ）の検索、検索事例の適合／評価、適合事例の修正、主に、事例の再利用性による事例群のクラスタリングである。
- (2) CBRからの出力は、1つのクラスタであ

る。このクラスタは、再利用性の観点から分類された事例集合であり、節集合に変換され、モデル推論の入力となる。

(3) モデル推論では、入力節集合に対して、項及び節の最小汎化を試みると共に、Intra-Constructionにより、新規述語の考案を図り、Absorptionにより、有効なルールを見出すことを試みる。節の最小汎化が成功すれば、不用スロット、また、Intra-Constructionが成功すれば、追加すべき新規スロットをCBRにフィードバックすることになり、CBRの事例構造の更新を事例ベース開発者に促す。また、モデル推論が旨く動作しない場合は、入力が不適切として入力を再要求し、CBRでは、不適切な出力の生成に関連するクラスタリング基準を無効にして、新たなクラスタをモデル推論の入力として与える。

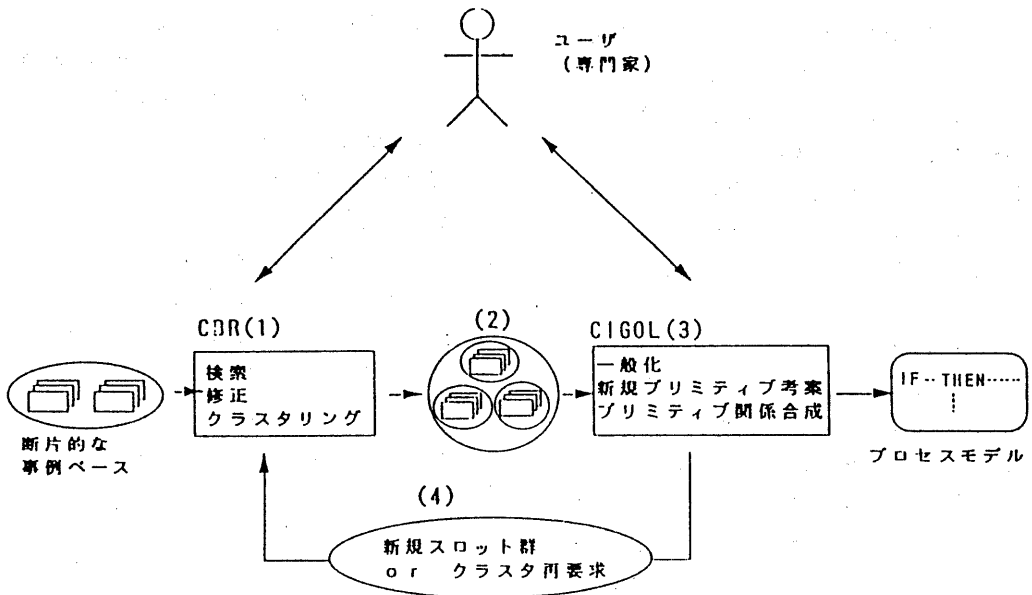


図5 CBRとモデル推論の協調動作に基づく知識獲得支援環境の概観

(4) モデル推論からCBRへの出力は、上述の事例構造を更新するためのスロットに関する情報、並びに、入力（クラスタ）再要求である。

以上のように、CBRとモデル推論の協調動作は、「事例構造の逐次的更新」と「その時点での事例構造を記述する概念の分解と合成」をスパイラル状に繰り返すことによって、知識を漸進的に獲得していく。

以下、法律解釈問題とソフトウェア開発プロセス記述問題における、本システムの適用結果について述べる。

まず、意思表示に関連する民法94条の適用時によく議論される、類推適用に関する知識獲得に関する実験結果と評価を以下に述べる。

類推適用は、法令に直接記述されていない事象に対して、判例や学説などから導びかれる規則を利用して、その事象を法令に当てはめる試みである。従って、本システムを適用した結果、判例や学説などから考案された記述子を含む法律解釈ルールが得られれば、それが類推適用時に利用されるルールとして評価される可能性がある。紙面の都合上、実験経緯は省略するが、Intra-constructionによって、仮装概観（X）と仮装概観に対する相手方の行為（X）という新規述語が考案され、それがCBRの事例構造に反映された結果、CBRの検索効率が向上した。これは、初期に与えた判例構造の持つ不備な箇所を同定できたといえる。またAbsorptionによって、民法94条を記述するルールのボディ部に、上記の2つの述語を付加したルールが得られ、このルールが民法94条の類推適用時のルールとして評価できる。また、「黙示の承認の限界線の下限を単なる放置とする」という学説に相当するルールも生成された。

一方、SA/SD法に関連するソフトウェアプロセスの知識獲得実験と評価は、以下の通りである。

SA/SD法では、要求仕様文を分析してデータフローダイアグラム（DFD）で表現し、変換分析によりDFDから（モジュール階層）

構造図を作成し、その構造図を様々な品質基準により評価して改善し、最終的に得られた構造図をコードに変換するという、ソフトウェア開発手法である。本実験では、様々な品質基準により構造図を評価して改善するフェーズに着目し、改善前の構造図、改善後の構造図などを蓄積した事例ベースから、

IF 構造図の特徴
THEN 構造図の改善行為

というようなルールの獲得を目標とした。

生成された構造図改善ルール群において、比較的有效と思われるルールを図6に示す。図6において、上部は「逐次的凝結モジュールは削除せよ」、下部は「削除される逐次的凝結モジュールの箇所には、その逐次的凝結モジュールの上位モジュールを割り当てよ」というルールを意味している。後者のルールは、一般性のあるルールではないかもしれないが、ある場面でかなり有効に働く可能性がある。

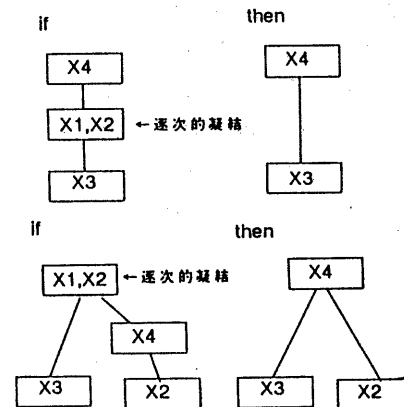


図6 生成された構造図変換ルール

上記の2つの題材において、有効なルールはいくつか生成されたものの、不要なルールがかなり生成され、コンピュータシミュレーションに課題が残った。しかしながら、コンピュータシミュレーションを通して、事例構造というモデルを逐次的に改良する仕組みができたことは価値のあることだと考えている。

5. おわりに

機械学習の研究に、全面的に知識獲得ボトルネックの解消を期待する時期は到来しておらず、課題は山積している。しかしながら、本稿で考察してきたように、「知識によるコンピューションとモデリング」の各々を軽視することなく、相互作用を通して、各々の不備を段階的に改良していくような枠組みの考察は、実問題上で有用な機械学習システムを構築するための一つのアプローチであり、機械学習システムの一つの方向を示していると考えている。

現在、知識工学の分野において、Ontology (KBを構成する基本単位)を整備し、知識の共有化を図り、再利用性を向上させる研究が活発化しようとしている。そのようなOntologyが与えられれば、知識によるモデリングの研究は加速的に進むことが予想され、さらには、コンピューションとの関連もより具体的なレベルで検討されていくであろう。

また、図7に示すように、第3世代ES構築用ツールは、知識によるモデリングを重視したツール群 (Task Specific Tool, Knowledge Acquisition Tool, Domain Specific Tool) と、コンピューション (機械学習) を重視したツール群 (Knowledge Refinement Tool, Rule Induction Tool, Neural Network(or GA) Tool)

に分化が進んでいるが^[5]、本考察から予想すれば、分化が進んだ第3世代ES構築用ツール群は、知識獲得機能をより向上させるために、統合化が図られ、第4世代ツールとして発展していくであろう。

参考文献

- [1]山口, 溝口他: 対象モデルと故障モデルに基づく知識コンパイラIIの構築と評価, 人工知能学会誌, 7巻, 4号, pp.663-674 (1992)
- [2]T. Yamaguchi and Y. Tachibana: Unifying Rule Induction and Rule Refinement-Towards Discovering Anomaly form Granville's Law in a Stock Market Technical Analysis (selected paper), Proc.of 1993 Korea/Japan Joint Conf. on ES, pp.209-226 (1993)
- [3]T.Yamaguchi and M.Kurematsu:A Framework for Knowledge Acquisition through Cooperation between CBR and Model Inference, Proc of the Second PRICAI, 2, 659-665 (1992)
- [4]山口, 落水: ソフトウェア設計プロセス知識獲得支援環境の構成法, 第10回設計シンポジウム予稿集, 51-58 (1992)
- [5]山口: AIツールの比較評価, コンピュータソフトウェア, 7巻, 3号, pp.60-70 (1990)

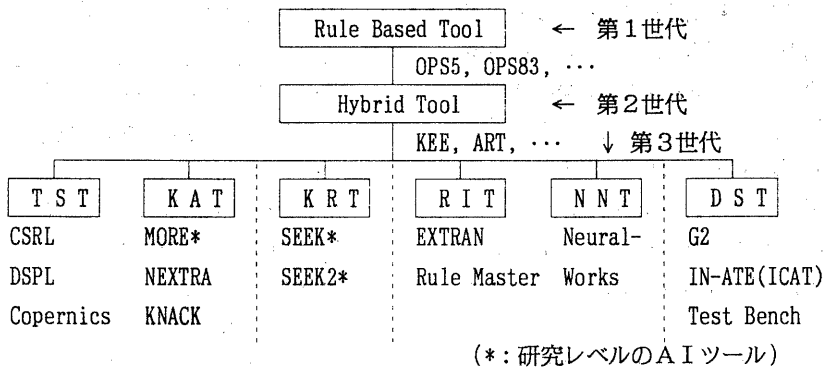


図7 ES構築用ツールの世代交代