

SQUAD: SWQC事例検索システム*

6Q-6

北野宏明

柴田晃宏

梶原壽一郎

佐藤亜津美†

日本電気(株)‡

1 はじめに

本論文では、SQUAD Software Quality Control Advisor Systemの開発思想とその第一フェーズ・システムについて述べる。SQUADシステムが、対象とする領域は、SWQC (Software Quality Control)でありソフトウェア品質向上・生産性向上に関するドメインである。SQUADシステムは、事例ベース推論 (Case-Based Reasoning) を基本的パラダイムとして採用した。なぜなら、ソフトウェアの品質管理・生産性向上に関するドメイン・モデルの定義が困難であり、従来のエキスパート・システムでは、対応できないが、SWQC活動によって大量の事例が獲得できる可能性があるからである。また、SWQCによって蓄積された事例は、経験の蓄積であり、それを全社的規模で共有するメカニズムが必要であるという認識がある。従来、事例ベース推論システムは、ルール・ベースのエキスパート・システムと同列に論じられてきた。つまり、どちらも問題解決の為に推論機構という側面が強調されてきた。我々は、事例ベース推論を、経験共有のためのメディア実現の一つのメカニズム、そして、統計情報獲得のベース・メカニズムとしてとらえる。つまり、SQUADの開発思想は、「事例ベース・システムを戦略情報システムの中核に置く」というものである。

この思想を、実現させるためには、全社的な知識獲得プロセスを確立し、事例ベース・システムが常に有用な情報を持ち、柔軟なアクセスが可能である状態にする必要がある。我々は、これを、組織的アルゴリズムとシステム・デザインの統合をもって達成しようと試みる。もちろん、これは大きな課題であり、現段階では完全に達成されてはいない。しかしながら、実用的な、大規模事例ベース・システムの構築には不可欠のプロセスである。本論文では、この様な開発思想をもったシステムの例としてSQUADシステムの現状と、組織的アルゴリズムである知識工学化されたSWQCについて述べる。

2 全社的知識獲得プロセスとしてのSWQC

SWQC (Software Quality Control) は、ソフトウェアの品質向上を目的としたQC活動である [4]。日本電気では、1981年より導入され、現在では、全社的活動として、年間3,000件を越すSWQC事例が報告されている。従来、QC活動は、小集団活動による品質向上、モラルアップ等、労務管理的側面が強調されていた。しかしながら、我々は、SWQCを全社的知識獲得プロセスととらえ、組織的な事例収集プロセスとして再定義した。

知識獲得の困難さは、Knowledge Acquisition Bottleneckと呼ばれ、エキスパート・システム構築の際の最大の問題点であった。事例ベース推論では、ルールを定義せず、事例を直接利用し推論することによってこの問題点を回避できる可能性が指摘されていた。しかしながら、事例ベース推論に於いても、事例を収集しなければならない事には変わりなく、本格

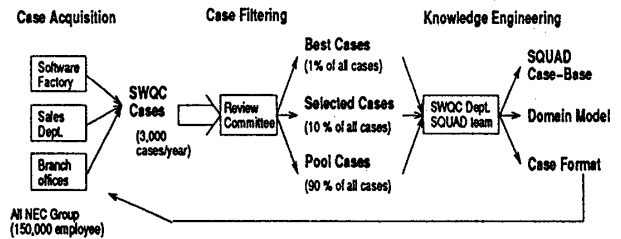


図1: SWQCでの知識工学的プロセス

的システムを構築する際には、事例獲得の困難さ、Case Acquisition Bottleneckが予想される。

我々は、この問題を既に定着しているSWQC活動の知識工学化により、事例獲得を行う事とし、解決しようとしている。図1に、SWQC活動における事例獲得プロセスを示した。

このプロセスで重要な点は、事例の選別、ドメイン・モデル (概念階層) の作成、そしてドメイン・モデルに基づく事例形式の詳細化・規格化である。SWQCの扱う領域は、ドメイン・モデルが、明確に理解されていないため、初期の段階から、事例の構造やfeatureセットやその値を定義する事が困難であった。このため、事例の形式、featureセット、属性値を段階的に詳細化する戦略を採用した。過去10年間のSWQC活動を通じ、我々は、累計約2万件の事例を蓄積し、毎年3,000件の新事例を獲得している。SQUADは、完成時には、世界最大級の事例ベース・システムとなると思われる。

3 SQUADシステム

SQUADシステムは、SWQC活動で収集された事例ベースに基づいてソフトウェア品質向上・生産性向上のアドバイスをを行うシステムである。第一フェーズの開発では、事例の検索部分のみを実装した。なぜなら、(1)アダプテーション・フェーズは、Adaptation Ruleを定義する必要があり、適切な設計がなされない場合に、その部分でCBRの良さが半減する危険がある、(2)適用事例の収集によってAdaptation Caseを利用するには、Adaptationの導入は第二フェーズとする必要がある、(3)類似事例の提示で多くの場合十分である、等の理由による。さらに、Memory-Based Reasoning的に展開するなら [7]、ルールによるAdaptationではなく、統計的推論を主に用いる事になる。第一フェーズ初期のSQUADシステムの画面を図2に示す。事例検索指定ウィンドウと検索結果表示ウィンドウがある。

SQUADシステムは、ドメイン・モデルが不明確な領域を対象にすると共に、全社の展開を前提としている。その為の、要求仕様として基本的な項目として、以下のものがある：

速度：事例検索の処理速度は、できるだけ高速である事が要求される。かなり大きな事例ベースに於いても、即時検索できることが必要である。

事例表現の柔軟性：大規模事例ベースを構築する場合には、事例の作成者が、多人数になるとともに、事例の記述レベルも一定ではない。このため、違ったレベルの事例表現、欠

* SQUAD: Case Retrieval System for Software Quality Control

†Hiroaki KITANO, Akihiro SHIBATA, Juichirou KAJIHARA and Atsumi SATO

‡NEC Corp.

図 2: SQUAD System

Case No.	F1	F11	F12	F121	F122
0001	1.00				
0002	1.00				
0003	0.00				

Case No.	F1	F11	F12	F121	F122
0001	1.00	0.50	0.50		
0002	1.00	0.50	0.50		
0003	0.00	0.00	0.00		
0004	1.00	1.00	0.00		
0005	1.00	0.00	1.00		
0006	1.00	0.00	1.00	0.00	1.00

図 3: 事例表現の変更

落情報を持つ事例への対応など、柔軟な事例表現力が必須である。

検索指定の柔軟性: 事例の表現が柔軟であると共に、検索の指定も柔軟である必要がある。SQUAD システムは、多数の色々な知識レベルのユーザーに使用されるので、あるユーザーは、詳細な条件指定を行うが、他のユーザーは、非常に抽象的な指定のみを行うなどと言う事が起こる。この様な、場合にも的確な検索が可能である事が望ましい。

事例構造の変更コスト: ドメイン知識が時間が経つに連れて、詳細になると共に、事例の構造も変化する。よって、事例ベースの変更の際のコストが低い事が重要である。

これらの、要求を満たすために、まず事例ベース検索の高速化と高精度化を計った。我々は、無関係な属性を無視して、探索次元を縮退させる、動的次元縮退法を採用した[8]。これによって、数千事例までは十分な反応速度が維持できる事が実測の結果明らかになった。さらに、事例が増えた場合には、事例ベースの分散処理、並列探索等の導入が必要になる。

次に、柔軟な事例表現・検索のためのメカニズムを開発した。まず、事例はフラットな表形式で表現され、この表の項目数の増加によって事例の構造が変化する。図 3に例を示した。まず、属性 F1 のみ、の表から開始し、新たな属性が、SWQC の知識工学プロセスの結果発見されたとする。新たな属性は、それが現在定義されている属性のサブクラスならば、F11、F12 として表に加えられる。最初からある、事例は F1 と指定されていた属性が、さらに詳細なレベルで、F11 か F12 かの情報が無いので、等確率で値が分配される。さらに、下位構造が発見されたときも同様である。

このメカニズムは、柔軟な検索も可能にする。図 4に、簡単な例を示した。Query は、各々の属性に関して色々な抽象度

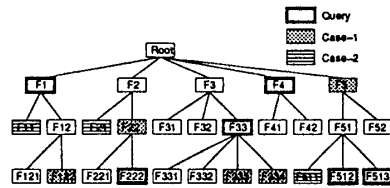


図 4: 検索指定と事例表現

で指定するが、指定された属性と、事例との上位・下位関係と確率配分で類似度を計算する。各属性のウェイトが同じだと仮定すると、この例では、Query に対して Case-1 は、0.305、Case-2 は、0.167 の類似度を持つ。

これらの、メカニズムによって、SQUAD システムは、Deployed System が対応すべき諸要求に対応している。

4 おわりに

本論文で、概略を示した SQUAD システムは、SWQC の組織的活動の結果、発生する事例の変更、多人数での事例獲得に伴う事例記述レベルのばらつき、情報欠落、検索指定のばらつき等の実用システムが直面する問題に対応する事を第一義に設計されている。これによって、組織的アルゴリズムである SWQC の知識工学のプロセスとシステムの統合が計られる。その結果、フラットな事例表現、動的次元縮退法等を採用する事になった。ここで、明らかにして置くべき事は、SWQC ドメインにおいては、概念階層が予め確定していない、さらに各々の検索時でその目的が違うため、indexing によって検索をガイドするという CBR で一般に使用されている手法 [3] が使えないと言う事である。また、問題解決のルール等が確立していないため、CBR+EBL[1] や推論の過程を保持する手法 [2] も利用できない。この様な状況で、有効な事例検索を可能にする事が重要であった。我々の、結論は、それらの複雑な手法より、ここで示した簡単な手法の方が、我々が直面しているタスクには有効であり、開発・保守効率が高いであろうと言うものである。

参考文献

- [1] Cain, T., Pazzani, M., and Silverstein, G., "Using Domain Knowledge to Influence Similarity Judgement," *Proc. of Case-Based Reasoning Workshop*, 1991.
- [2] Veloso, M. and Carbonell, J., "Variable-Precision Case Retrieval in Analogical Problem Solving," *Proc. of Case-Based Reasoning Workshop*, 1991.
- [3] Hammond, K., *Case-Based Planning: An Integrated Theory of Planning, Learning, and Memory*, Ph.D. Thesis, Yale University, 1986.
- [4] 水野 (監修) "ソフトウェアの総合的品質管理", 日科技連, 1990.
- [5] Riesbeck, C. and Schank, R., "Inside Case-Based Reasoning", Lawrence Erlbaum Associates, 1989.
- [6] 島津ほか, "事例ベース推論シェルの試作", 情処 43 全大, 1991.
- [7] Stanfil, C. and Waltz, D., "Toward Memory-Based Reasoning", *Communication of the ACM*, 1986.
- [8] 柴田ほか, "大規模事例ベース検索の効率的かつ高精度な手法", 情処 44 全大, 1992.