

근사질의응답기능을 이용한 협동적 사례기반추론

김진백*

<목 차>

I. 서론	3.1 AQA를 이용한 CBR의 추론과정
II. 협동적 응답기능으로서의 AQA	3.2 유사성개념을 이용한 선호 휴리스틱기법
2.1 엄밀질의응답기능과 협동적 응답기능으로서의 AQA 비교	3.3 적용예
2.2 유사성의 제측면	IV. 결 론
2.3 인접응답집합 결정을 위한 추상화 계층구조	참고문헌
III. 협동적 응답기능을 이용한 CBR의 추론과정	Abstract

I. 서론

사례기반추론(Case-Based Reasoning : CBR)은 새로운 문제가 주어지면 이를 분석해서 사례형식으로 저장되어 있는 과거의 경험중 새로운 문제와 가장 유사한 경험을 검색한 후, 검색된 사례를 직접 혹은 부분적인 수정을 통해 문제해결에 이용하려는 기법이다. 따라서 사례는 경험을 나타내는 상황적 지식이다(Kolodner, 1993). Aamodt 등(1994)에 의하면 CBR 과정은 검색, 재사용, 수정, 보관이라는 4 단계로 나누어진다. 검색이란 새로운 문제의 특성을 이용해서 유사한 특성을 가진 과거에 경험했던 사례를 찾는 과정이며, 재사용은 검색된 과거 사례를 문제해결에 이용하는 과정이다. 과거 사례의 재사용시 고려되어야 할 점은 새로운 문제와 과거 사례간의 문제 특성에 차이가 있을 수 있다는 것이다. 따라서 이들간의 차이를 고려해서 과거의 해결안을 새로운 문제에 적합한 해결안으로 변형하는 과정이 필요하며, 이는 수정단계에서 이루어진다. 그리고 새로운 문제해결경험은 미래에 재사용을 위해 보관이라는 과정을 거치게 된다. 사례의 보관시 고려할 점은 CBR의 문제해결능력은 저장된 사례의 수에만 의존하는 것이 아니라 다양성에도 의존한다는 것이다(Mahapatra and Sen, 1994; O'Leary, 1993). 따라서 이를 위해 보관단계에서는 기존에 저장된 사례와

* 부산여대 경영정보과

동일한 사례는 배제하고 유일한 사례와 상위 사례(superset case)만을 저장한다는 사례저장 휴리스틱을 따르게 된다(Liang, 1993).

이와 같이 CBR은 반복적인 경험을 통해 문제해결방법을 습득하는 기법으로서 일종의 귀납적 추론방법이라 할 수 있다. 따라서 CBR의 대상영역은 경험이 풍부한 영역이 되는 것이 일반적이므로 많은 사례가 있어야 유용한 기법이 될 수 있다. 이로 인해 추론시 관련 사례들이 여러 개가 추출될 수 있으며, 이 경우에는 새로운 문제와의 유사성에 따라 유용한 사례들을 선별해서 문제해결에 이용해야 한다.

그러나 CBR에서는 유용한 사례들을 결정해주는 수단이 명확하지 않아 이용자나 개발자의 자의적인 판단기준에 따라서 관련 사례들을 선정한다. 만약 경험이 풍부한 전문가라면 이러한 방법을 이용해도 무관하겠으나 비전문가의 경우에는 관련성 여부를 결정하기 어려울 뿐 아니라 문제해결 시점마다 다른 기준을 적용하므로 일관성이 결여될 수 있다. 본 논문에서는 이러한 문제를 해결하기 위해 근사질의응답기능을 도입해서 일관성 있게 관련 사례들을 결정할 수 있는 협동적 CBR 과정을 수립하기로 한다.

II. 협동적 응답기능으로서의 AQA

2.1 엄밀 질의응답기능과 협동적 응답기능으로서의 AQA 비교

대부분의 데이터베이스시스템이나 일반 정보시스템들은 사용자와 협동적으로 작업을 수행하지 않기 때문에 이용의 어려운 점이 많다. 사람과 달리 정보시스템은 입력된 내용 자체만을 가지고 질의에 응답하므로 유연성이 부족하다. 따라서 응답내용이 적절할 수도 있으나 경우에 따라서는 부적절한 응답이 주어지기도 하기 때문에 Kaplan(1982)이 지적한 방해현상(stonewalling)이 발생할 수도 있다.

정보시스템이 보다 지적이 되기 위해서는 협동적 응답(cooperative answering)기능을 가질 필요가 있다. 협동적 접근법은 문제에 대한 불명확한 정의를 허용한다(Forslund, 1995). 정보시스템에 협동적 기능을 부여하는 방법에는 질의와 관련된 정보를 일반화시키는 방법과 관련 주제에 관한 정보를 추가시키는 방법이 있다(Gaasterland, Godfrey, and Minker, 1992). 후자의 방법은 시스템개발시점에 지정된 정보를 관련 정보로 이용하게 하므로 문제해결시점에 문제의 상황적 특성을 고려하지 못한다. 그러나 전자의 방법은 일반화의 정도를 달리하므로 문제의 상황적 특성을 고려해서 관련 사례의 대상범위를 결정할 수 있다. 따라서 본 논문에서는 협동적 응답기능을 강화하기 위해 전자의 방법을 이용할 것이다.

전통적 방식의 질의처리는 질의에 명시된 조건과 정확히 일치하는 데이터만을 제공해 주는 엄밀질의응답(exact query answering)기능을 따른 방식이라 할 수 있다. 따라서 엄밀질의응답기능은 질의조건 ϕ 를 완전히 충족시켜 주는 객체에 대해서는 1,

그렇지 않은 객체에 대해서는 0의 값을 발생시켜 1의 값을 갖는 객체들에 대한 관련 정보만을 제공해주는 특성함수(characteristic function)로 볼 수 있다. 그러므로 엄밀 질의응답의 특성함수 M_ψ 를 이용해서 엄밀응답집합 A_ψ^0 를 정의하면 다음과 같다 (Chu and Chen, 1992).

$$A_\psi^0 = \{t \in R \mid M_\psi(t) = 1\} \quad \text{여기서 } t = \text{객체}$$

$$R = \text{객체집합}$$

그러므로 전통적 방식인 엄밀질의응답기능에 따라 정보시스템을 이용하려면 사용자는 문제영역과 관련 DB 스키마를 충분히 이해하고 있어야만 가능하다. 이는 정보시스템의 이용률을 저하시키는 주요 원인중 하나가 된다. 또한 정보검색에 있어서 주요 문제점중 하나는 용어의 변동성이다. 즉, 시스템의 모든 사용자가 동일한 용어로서 객체를 생성, 표현, 식별하지 않는다. 이로 인해서 정보검색률은 매우 낮게 나타나고 있으며, 정보검색률이 낮은 또다른 이유는 사용자가 관련 주제에 대한 모든 가능한 용어를 명시하도록 하는 부담에 있다(Srinivasan, 1991). 따라서 정보시스템 사용에 따른 사용자의 부담을 경감시켜 주고 검색률을 높이기 위해서는 질의조건을 완화(relaxation)시키므로써 부분적으로 질의조건을 만족하는 객체에 대한 정보도 제공해주어야 한다. 이러한 질의응답방식을 근사질의응답(approximate query answering : AQA)기능이라고 한다. 근사질의응답기능에서는 특성함수를 일반화한 만족함수 μ_ψ 를 이용해서 질의조건 ψ 에 대한 인접응답집합(neighborhood answer set)인 A_ψ^N 을 제공해주며, 만족함수 μ_ψ 를 정의하면 다음과 같다.

$$\mu_\psi : t \rightarrow [0, 1] \quad \text{여기서 } 1 : \text{완전 만족}$$

$$0 : \text{불만족}$$

$$0 \text{ 혹은 } 1 \text{ 이외의 값} : \text{부분 만족}$$

그러나 응답집합중 만족함수 $\mu_\psi = 0$ 의 값을 갖는 객체는 관련이 없는 응답이므로 이를 배제시킨 것이 사용자가 원하는 응답이므로, A_ψ^N 은 다음과 같이 정의될 수 있다.

$$A_\psi^N = \{t \in R \mid \mu_\psi(t) > 0\}$$

따라서 엄밀응답집합과 인접응답집합의 관계는 다음과 같이 나타낼 수 있다.

$$A_\psi^0 \subseteq A_\psi^N$$

그러나 사용자에게 따라서는 인접응답집합중 일부만을 원할 수도 있다. 즉, 인접응답집합의 요소중 만족함수의 값이 아주 낮은 객체에 대한 정보는 사용자의 입장에서 보면 유용성이 없는 정보일 가능성이 높다. 따라서 이러한 점을 고려해서 객체가 질의조건을 불만족시키는 크기를 나타내는 d 값에 따라 인접응답집합의 크기를 조정해야 한다. 만약 불만족의 크기가 d 이내에 있는 인접응답집합을 A_ψ^d 라고 하면, A_ψ^d 의 응답들은 다음의 조건을 충족시키는 것들이 된다.

$$A_\psi^d = \{ t \in A_\psi^N \mid 1 - \mu_\psi(t) \leq d \} \quad \text{여기서 } 0 \leq d \leq 1$$

엄밀질의응답을 불만족 크기의 개념을 통해 정의하면

$$A_\psi^0 = \{ t \in A_\psi^N \mid 1 - \mu_\psi(t) = 0 \}$$

이다. 따라서 엄밀질의응답 A_ψ^0 , d 거리 응답집합 A_ψ^d 와 인접응답집합 A_ψ^N 의 관계는 다음과 같이 나타낼 수 있다.

$$A_\psi^0 \subseteq A_\psi^d \subseteq A_\psi^N$$

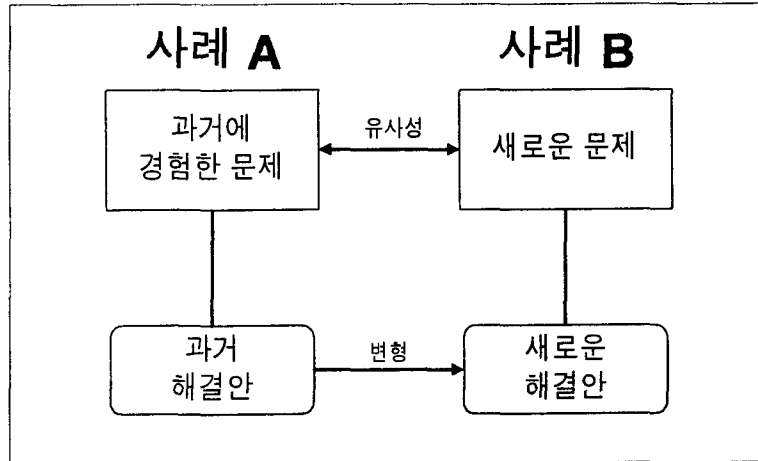
정보시스템이 협동적 응답기능을 갖더라도 사용자가 원하는 응답을 제공하기 위해서는 d 개념을 측정하기 위한 수단이 필요하다. 본 논문에서는 이를 위해 검색의 다양성을 제공해주면서 일정거리내의 인접응답집합을 제공할 수 있는 추상화 계층구조를 도입할 것이다.

2.2 유사성의 제측면

CBR의 문제해결방법은 <그림 1>과 같이 새로운 문제(사례 B)를 과거에 경험했던 문제(사례 A)를 통해 해결하려는 방법이다. 따라서 CBR 과정중 검색의 주요 목적은 새로운 문제와 유사한 기존의 사례를 사례베이스(casebase : CB)에서 추출하므로써 문제해결에 필요한 정보를 제공하려는 것이다. 즉, CBR은 새로운 문제와 과거에 해결했던 문제가 유사할 경우에 과거의 해결안을 이용해서 새로운 문제의 해결안을 생성하려는 방법이다.

만약 CBR이 과거에 해결했던 문제와 동일한 문제만을 해결할 수 있다면 이는 단순히 캐싱기능을 수행하는 것과 동일하다. 그러나 CBR은 과거의 해결안을 수정하므로써 새로운 문제의 해결을 가능하게 한다. CBR에 의한 문제해결방식의 기본가정은 두 문제간의 유사성 결정수단이 존재하며, 스크래치 방식에 의한 평균 문제해결시간보다 사례기억장소의 검색시간이 작으며, CBR의 문제해결비용은 두 문제간의 차이 크기에 비례한다는 것이다(Koton, 1989). 따라서 CBR의 적용가능범위는 유사한 문

제가 재발하고, 변동은 있으나 안정적이며, 상호작용이 제한적인 영역이 된다고 할 수 있다. 또한 CBR이 유효하기 위해서는 위의 3가지 기준을 충족시키면서 기존 사례와 새로운 문제간의 유사성을 평가할 수 있어야 한다.



<그림 1> CBR의 문제해결방법

Liang과 Konsynski(1993)는 유사성을 개체 유사성(entity similarity), 구조적 유사성(structural similarity) 및 기능적 유사성(functional similarity)이라는 3가지 측면에서 보았다. 개체 유사성이란 두 개체가 지닌 표면적 속성들을 비교해서 측정하는 것으로, 이는 의미론적인 측면에서의 유사성 개념이라 할 수 있다. 구조적 유사성은 대응하는 두 개체의 내부관계를 비교해서 측정하는 것으로 구조적 동질성 측면에서의 유사성 개념이라 할 수 있다. 또한 기능적 유사성이란 두 개체가 유사한 기능을 발휘하거나 유사한 목표를 달성하는가에 따라 측정되는 유사성 개념이다. 따라서 이는 실용적인 측면에서의 유사성 개념이라 할 수 있다.

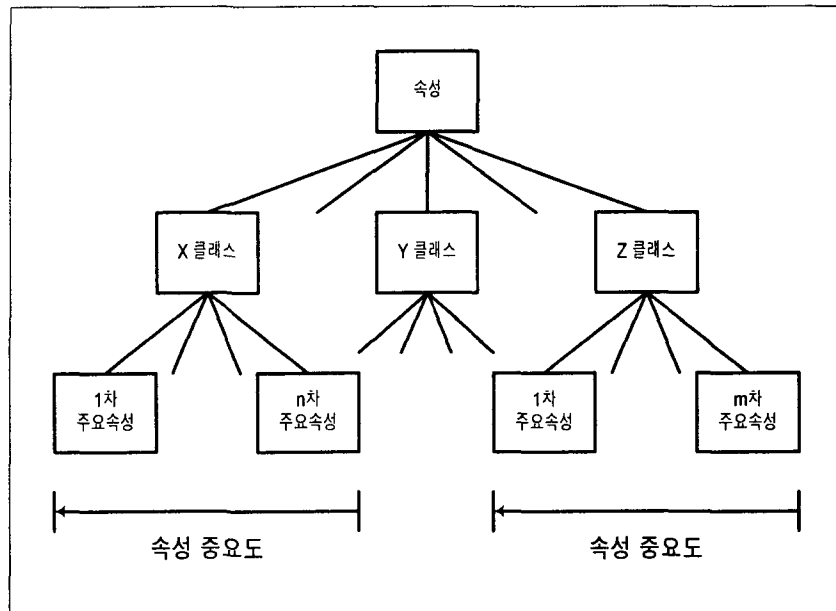
이러한 유사성의 제측면들을 모두 반영할 수 있어야 사례표현스킴에 대한 표현의 유연성 요구사항을 충족시켜 줄 수도 있다. 그러나 사례표현스킴이 개체의 모든 유사성 측면을 지원해주려면 CB의 구조가 너무 복잡해져 구현이 어려우므로, CB 구조를 단순화시켜 주면서 표현의 유연성 요구사항을 충족시켜 주기 위해서는 유사성개념을 추상화 계층구조에서 지원해주어야 한다.

2.3 인접응답집합 결정을 위한 추상화 계층구조

정보시스템의 이용률을 높이고 의사결정 지원능력을 향상시키기 위해서는 전형적인 검색절차와 달리 질의완화개념을 이용한 새로운 검색절차가 이용되어야 한다. 그러나 질의완화개념을 이용하려면 완화의 수단이 필요하므로 인접응답집합을 결정하기 위해 추상화 계층구조를 도입하기로 한다. 추상화 계층구조에는 유사성의 3가지 측면에 따라 속성추상화 계층구조(attribute abstraction hierarchy), 구조추상화 계층구

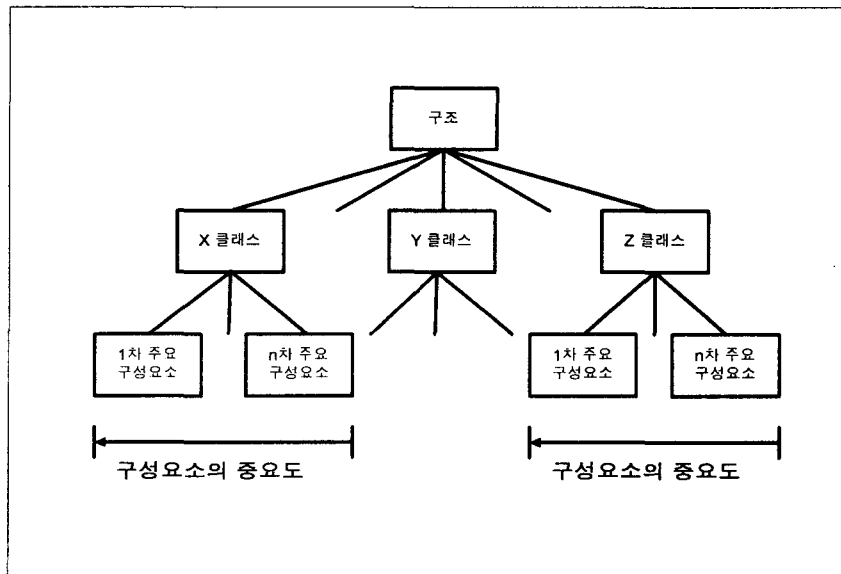
조(structure abstraction hierarchy), 목표추상화 계층구조(goal abstraction hierarchy) 등 3가지가 있다.

속성추상화 계층구조는 개체 유사성에 따른 체계이다. 개체 유사성이란 의미 유사성(semantic similarity)을 말하는 것으로, 의미론적 연결(semantic link)은 기본적인 검색경로를 제공해준다(Thagard and Holyoak, 1989). 따라서 속성추상화 계층구조는 기본적인 검색경로에 따라 CB를 추상화 관점에서 단순화시킨 것이라 할 수 있다. 따라서 검색과정에서 선택한 완화의 방향이 개체 유사성이면 <그림 2>와 같은 개체의 속성추상화 계층구조를 이용해서 원질의보다 추상화 정도를 높이기 위해 주요 속성만으로 질의내용을 표현하게 된다.



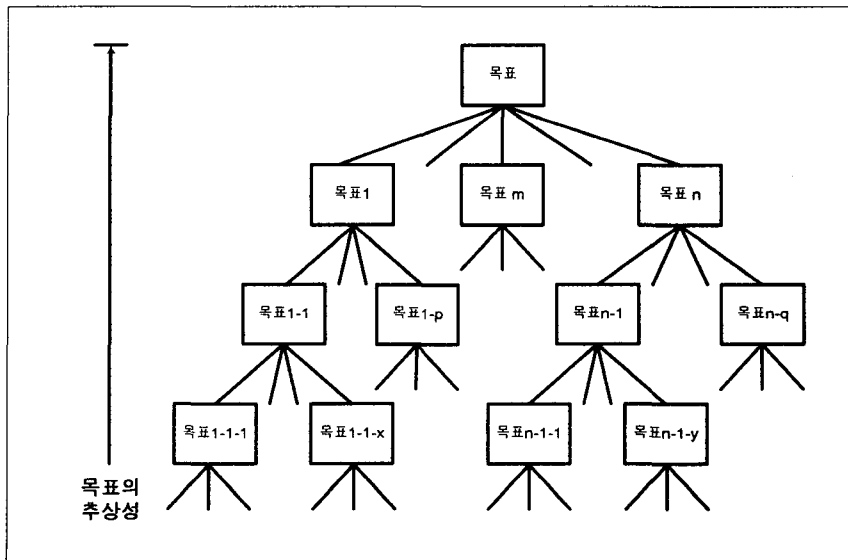
<그림 2> 속성추상화 계층구조

구조추상화 계층구조는 구조적 동질성에 따라 개체를 분류한 후, 이들의 주요 부분들을 중요도에 따라 계층화한 것이다. 만약 검색과정에서 선택한 완화의 방향이 구조적 유사성이면 <그림 3>과 같은 구조추상화 계층구조를 이용해서 검색대상의 구조적 특성을 이용해서 사례를 검색할 수 있도록 주요 부분에 대한 속성만을 남기고 나머지 속성들은 모두 삭제하여 질의내용을 수정한다.



<그림 3> 구조추상화 계층구조

그리고 목표추상화 계층구조는 개체의 이용목적에 기능적 동질성에 따라 분류한 체계이다. 그러므로 완화의 방향이 기능적 유사성으로 선택되면 <그림 4>와 같은 목표추상화 계층구조를 이용해서 질의내용을 수정하게 된다. 질의완화의 방향을 하나로 고정시키지 않고 다수로 하는 이유는 문제의 상황에 따라 다양한 측면에서 유용한 사례들을 검색가능하게 하고, 사용자의 선호도를 보다 잘 반영하기 위해서 이다.



<그림 4> 목표추상화 계층구조

III. 협동적 응답기능을 이용한 CBR의 추론과정

3.1 AQA를 이용한 CBR의 추론과정

전통적인 CBR의 추론과정에 AQA 기법을 도입할 경우 개선된 추론절차는 <표 1>과 같다. 제 1 단계에서는 사용자가 원하는 사례들로 구성된 엄밀응답집합 요소들이 CB에 대한 탐색결과로서 주어진다. 1 차 검색된 사례들에 대해 사용자가 만족하면 이를 이용해서 해결안을 도출하기 위해 제 4 단계로 나아가고 그렇지 않으면 제 2 단계로 이동하여 인접응답집합을 검색하게 된다. 제 2 단계는 엄밀응답집합에 해당하는 검색결과를 사용자가 받아들일 수 없을 경우이거나 엄밀응답집합이 공집합인 경우에 실행되는 단계이다. 따라서 제 2 단계는 인접응답집합에 속하는 사례들을 추출하기 위해 질의내용을 완화시킬 방향을 선택하는 단계로서 개체 유사성, 구조적 유사성, 기능적 유사성 등 3가지 측면에서 완화시킬 방향을 선택할 수 있다. 따라서 사용자의 선호기준에 따라 완화방향이 선택되면 필요한 조건만을 남기고 나머지 조건들을 모두 삭제하여 질의내용을 수정하게 된다. 그후 다시 CB를 검색하기 위해 제 3 단계로 가게 된다. 제 3 단계에서는 제 2 단계에서 지정한 부분적 문제속성을 이용해서 관련 사례들을 검색하며, 모든 부분에 대한 관련 사례들이 검색되면 제 5 단계로 이동한다.

만약 질의완화에 의한 검색결과도 만족스럽지 못하다면 이는 CB가 문제해결에 필요한 사례를 제공할 수 없음을 의미한다. 따라서 이때는 과거에 유사한 경험이 없는 경우이므로 CBR을 이용한 문제해결이 어려운 것으로 보고 종료하게 된다. 제 4 단계는 검색된 사례를 중심으로 최적의 사례를 기저사례(base case)로 선택하는 단계로서 다수의 검색된 사례가 제시되었을 경우에는 선호규칙에 따라 기저사례를 선택하게 된다. 질의완화를 통해 CB에서 검색된 사례들은 대부분의 경우 문제에 대한 부분적인 해결안만을 제공해준다. 따라서 제 5 단계는 문제의 일부분에 대한 관련 사례들을 결합해서 기저사례를 제공해주기 위한 단계이다. 그리고 제 6 단계는 제 4 혹은 5 단계에서 선택된 기저사례를 이용해서 과거의 해결안을 새로운 해결안으로 제시하는 단계로서 경우에 따라서는 과거 해결안에 대한 수정이 필요하기도 하다.

이상의 여러 단계들 중에서 제 1 단계, 제 4 단계, 제 6 단계는 기존 CBR의 과정과 동일하지만 제 2-3 단계는 질의완화로 인해 새로이 추가된 단계이다.

3.2 유사성개념을 이용한 선호휴리스틱기법

검색결과는 3가지로 요약될 수 있으며, 첫째는 둘 이상의 유사한 사례가 검색되는 경우, 둘째는 유일한 유사 사례가 검색되는 경우, 셋째는 유사한 사례가 검색되지 않는 경우이다(Mahapatra and Sen, 1994). 만약 세 번째 경우와 같이 응답집합이 공집합이면 이는 유사한 과거 사례가 존재하지 않음을 의미하는 것으로 CBR이 문제해결

<표 1> AQA 기법을 이용한 CBR의 추론과정

단 계	절 차
제 1 단계	사용자 요구사항에 따른 속성집합을 이용해서 CB 검색 IF 검색결과가 만족스러우면 THEN GOTO 제 4 단계 ELSE GOTO 제 2 단계
제 2 단계	SELECT(질의완화방향?) CASE-1 완화방향=개체 유사성 => 속성값을 추상화해서 문제표현, GOTO 제 3 단계 CASE-2 완화방향=구조적 유사성 => 구성요소를 단순화해서 문제표현, GOTO 제 3 단계 CASE-3 완화방향=기능적 유사성 => 목표값을 추상화해서 문제표현, GOTO 제 3 단계 OTHERWISE 검색실패로 종료
제 3 단계	추상화된 부분적 문제속성으로 CB를 검색 IF 관련 사례의 검색완료 THEN GOTO 제 5 단계 ELSE GOTO 제 2 단계
제 4 단계	검색된 사례중 최적의 사례 선택, GOTO 제 6 단계
제 5 단계	검색된 사례를 통합
제 6 단계	새로운 해결안 제시후 종료

에 도움을 주지 못함을 의미한다. 그리고 두 번째 경우와 같이 응답집합이 단위집합(unit set)에 해당되면 검색된 유일한 사례가 기저사례가 된다. 그러나 첫 번째와 같이 두개 이상의 사례가 검색된 경우에는 이들중 문제해결에 가장 적합한 사례를 기저 사례로 선택하는 과정이 필요하게 된다. 기저사례선택을 위한 사례들간의 유사성 비교는 수치적 접근법(numerical approach)과 휴리스틱 접근법(heuristic approach)이 주로 이용된다.

수치적 접근법에서는 인접일치함수(nearest-neighbor match function)를 이용해서 가장 높은 점수를 받은 후보사례를 기저사례로 선택하게 된다. 인접일치함수는 속성들간의 상호독립성, 가중치의 정확한 인식가능성, 양(positive)의 가중치만을 채택, 후보사례들간의 독립성 등을 가정하고 있다(Gupta and Montazemi, 1997). 이러한 가정으로 인해 인접일치함수는 유사성의 계산은 용이하나 비구조적 문제의 특성은 수용할 수 없다는 문제점이 있다. 즉, 비구조적 문제의 속성들은 상호 비독립적이며, 문제영역에 대한 지식이 부족하기 때문에 속성들의 가중치 역시 사전에 파악되기 어렵다. 그리고 사례의 유용성은 다른 후보사례의 존재에 종속적이며, 속성들중에는 비유사성을 의미하는 것들도 있으므로 음의 가중치개념이 필요하다. 따라서 이러한 문제점들로 인해 기저사례선택을 위한 휴리스틱기법이 등장하였다.

비구조적 의사결정환경하에서 의사결정자들의 문제해결방법은 다양하게 나타난다. 따라서 CBR이 개별 의사결정자들의 특성을 수용할 수 있으면 보다 유용한 기법이 될 수 있을 것이다. 또한 문제해결과정에서 사용자가 원하는 사례는 새로운 문제와 가장 유사한 것이 아니라 새로운 문제에 가장 유용한 것이다. 여러 사례들중 가장 유용한 사례를 찾기 위해 PARADYME에서는 6가지 유형의 선호기준에 의한 휴리스틱기법을 이용해서 상대적으로 우수한 사례를 기저사례로 선택하는 방법을 채택하고 있다(Kolodner, 1989). PARADYME에서는 6가지 선호기준을 3단계로 나누어서 적용하고 있으며 첫 번째 단계에서는 목표지향 선호도(goal-directed preference)와 주요 속성 선호도(salient-feature preference)를 이용해서 새로운 문제의 목표를 가장 잘 나타내는 사례를 추출하게 된다. 두 번째 단계에서는 상세 선호도(specificity preference)와 수정용이 선호도(ease-of-adaptation)를 이용해서 보다 우수한 해결안을 제시할 수 있는 사례를 선별한 후, 세 번째 단계에서는 빈도 선호도(frequency preference)와 최신 선호도(recency preference)를 이용해서 검색된 사례들의 우선순위를 부여하게 된다. PARADYME의 선호휴리스틱절차는 표면적 속성만을 이용해서 관련 사례들을 추출한다. 따라서 PARADYME의 휴리스틱절차는 유사성의 다양한 측면을 고려하지 않았다는 문제점이 있다.

AQA를 이용한 추론과정에서 질의완화과정을 거치지 않을 경우에는 PARADYME과 같이 표면적 속성만을 이용해서 관련 사례들을 추출하나, 질의완화과정을 거칠 경우에는 개체 유사성, 구조적 유사성, 기능적 유사성 등의 측면에서 관련 사례들을 추출하므로 이들 유사성측면을 기저사례선택단계에서 반영해줄 수 있는 휴리스틱절차가 필요하다. 따라서 본 논문에서는 6가지 선호기준을 개체 유사성, 구조적 유사성, 기능적 유사성과 결합해서 <표 2>와 같은 선호휴리스틱절차(김진백, 1999)를 이용한다.

<표 2>의 선호휴리스틱절차는 추론과정중 제 4 혹은 5 단계에서 이용된다. 선호휴리스틱절차의 제 1 단계에서는 비교대상이 되는 사례들이 질의완화과정을 거치지 않고 CB에서 검색된 사례들이라면 사용자의 판단에 따라 목표지향 선호도, 주요 속성 선호도, 상세 선호도 기준중에서 하나를 이용해서 관련 사례들을 평가하게 된다.

질의완화과정을 거친 사례집합들이 비교대상이 될 경우에는 완화의 방향에 따라서 특정 선호기준이 적용되게 된다. 만약 개체 유사성에 따라 질의조건을 완화하였다면 상세 선호도기준에 따라 검색된 사례들을 선별한다. 완화의 방향이 구조적 유사성이었으면, 주요 속성 선호도기준에 따라 상대적으로 중요한 구성부분의 속성들만을 이용해서 사례를 선별한다. 그리고 완화의 방향이 기능적 유사성이었으면 목표지향 선호도기준에 따라 추상화되지 않은 목적, 즉 원질의에서의 목적을 가지고 있는 사례를 선택한다.

선호휴리스틱절차의 제 2 단계에서는 앞단계에서의 평가결과 남은 사례의 수가 하나이면 더 이상의 휴리스틱절차를 무시하고 해당 사례를 기저사례로 선택한다. 만약 남은 사례가 둘 이상이면 다시 수정용이 선호도기준을 이용해서 보다 수정이 편리한 사례를 선별하게 된다.

<표 2> 유사성개념을 이용한 선호휴리스틱절차

단 계	절 차
제 1 단계	IF 평가될 사례들이 원질의에 의한 것들이면 THEN 목표지향 선호도, 주요 속성 선호도, 상세 선호도 기준 중에서 하나를 임의로 선택해서 평가 ELSE CASE-1 완화방향 = 개체 유사성 => 상세 선호도를 이용한 평가 CASE-2 완화방향 = 구조적 유사성 => 주요 속성 선호도를 이용한 평가 CASE-3 완화방향 = 기능적 유사성 => 목표지향 선호도를 이용한 평가
제 2 단계	IF 남은 사례의 수 ≥ 2 THEN 수정용이 선호도를 이용한 평가 ELSE 남은 1개의 사례를 기저사례로 선택한 후 종료
제 3 단계	IF 남은 사례의 수 ≥ 2 THEN 빈도 선호도와 최신 선호도중 하나를 선택해서 평가 ELSE 남은 1개의 사례를 기저사례로 선택한 후 종료
제 4 단계	남은 사례들중 속성수가 최다인 사례를 기저사례로 선택

선호휴리스틱절차의 제 3 단계에서는 제 2 단계 평가결과 남은 사례가 1개일 경우에는 해당 사례를 기저사례로 선택하고 휴리스틱절차를 종료한다. 그러나 남은 사례가 둘 이상이면 이들중에서 빈도 선호도나 최신 선호도를 이용해서 최적의 사례를 선택하게 된다.

제 3 단계의 평가를 마치면 대부분의 경우는 1개의 사례만이 남을 것이므로 휴리스틱절차가 종료되게 될 것이다. 만약 둘 이상의 사례가 남으면 제 4 단계에서는 수립될 해결안의 타당성을 높이기 위해 보다 많은 정보를 지녔다고 할 수 있는 최다 속성 사례를 기저사례로 선택하게 되며, 이때 이용되는 유사성 평가함수(S)는 다음과 같이 나타낼 수 있다.

$$S = \text{Max}(n_1, \dots, n_k) \quad \text{여기서 } n_i : i \text{ 사례의 속성수}$$

3.3 적용예

컴퓨터지원 공정계획(Computer Aided Process Planning : CAPP)을 적용영역으로 채택하기로 한다. 공정계획이란 공학적 도면의 정의에 따라 부품을 초기 형태에서 최종 형태로 전환하는데 필요한 제조공정과 매개변수의 정의에 관련된 것이다 (Rogers, Farrington, Schroer, and Hubbard, 1994). CAPP란 컴퓨터의 지원을 받아

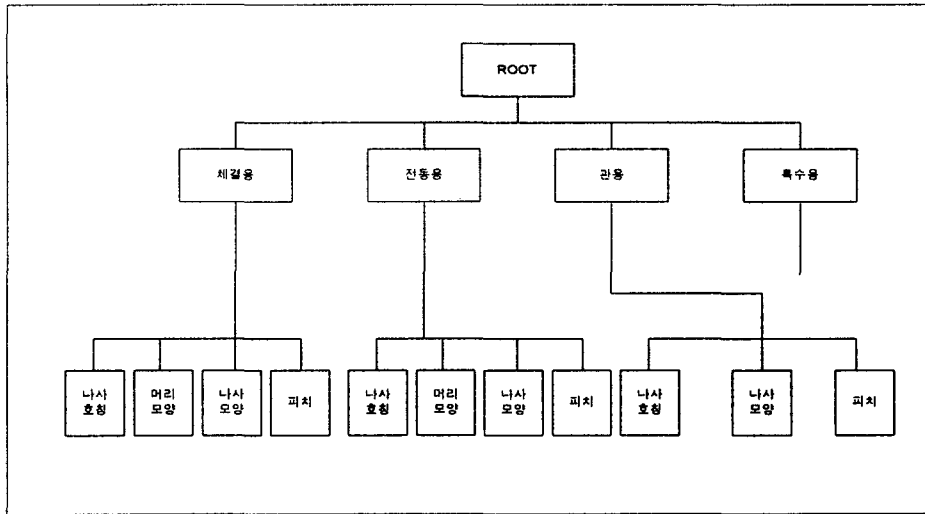
공정계획수립을 자동화하는 분야로서 CAD(Computer Aided Design)와 CAM(Computer Aided Manufacturing)의 교량 역할을 한다. CAPP를 위한 접근법으로는 변성형 접근법(variant approach)과 생성형 접근법(generative approach)이 있다 (Irani, Koo, and Raman, 1995).

변성형 접근법은 새로운 부품에 대한 공정계획이 어떤 부품계열의 표준계획을 수정함으로써 수립될 수 있다는 가정을 하고 있다. 따라서 새로운 부품에 대한 공정계획은 유사한 부품에 대한 기존 계획을 검색한 후 새로운 부품에 적합하게 필요한 수정을 하므로써 생성된다. 생성형 접근법은 CAD와 부품 청사진으로부터 획득 가능한 정보를 입력 자료로 해서 부품에 대한 공정계획을 합성하려는 접근법으로서 제조규칙과 설비능력을 컴퓨터 시스템에 저장해 두고 공정계획자의 개입없이 특정 부품에 대한 공정계획을 생성하게 된다.

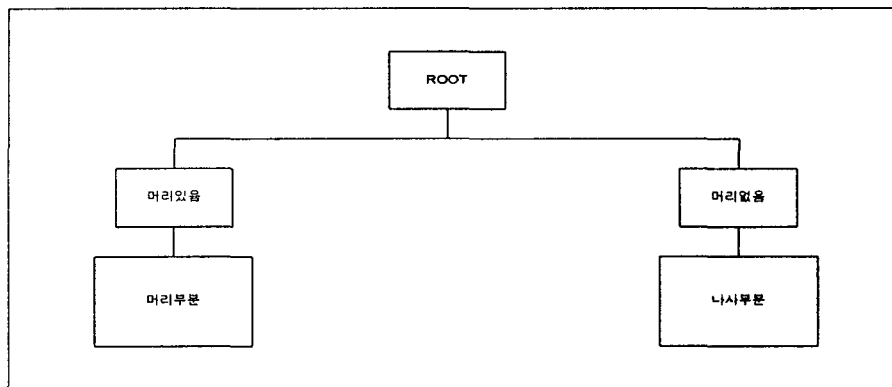
CBR은 과거의 문제해결경험을 새로운 문제해결에 이용하려는 방법이다. 따라서 CAPP에 대한 접근법 관점에서 보면 CBR은 변성형 접근법에 속한다. 제조분야는 지적시스템이 형식화된 지식(formalized knowledge)을 필요로 하는 방법들에 의존할 수 없는 영역이다(Tsatsoulis and Kashyap, 1993). 즉, 제조분야는 명시적이며 형식적이고 엄격한 지식과 규칙이나 절차로 나타낼 수 있는 지식이 부족한 영역이라 할 수 있다. 제조분야의 전문가들은 심도있는 도메인 지식뿐만 아니라 장기간의 경험을 통해 획득된 경험적 지식을 문제해결에 이용한다. 그러므로 추론시 경험을 이용하는 CBR은 제조분야의 문제해결지원을 위한 유용한 접근법이 될 수 있다.

본 논문에서 이용하고 있는 추론과정에는 질의완화여부에 따라 일반적인 검색과정과 질의완화에 의한 검색과정이 있다. 제안된 CBR 과정이 얼마나 실제문제의 해결에 도움이 되는가를 알아보기 위해 특수용을 제외한 체결용, 전동용, 관용에 속하는 모든 볼트형태에 대한 총 135개의 CAPP 사례중 불완전한 사례 20개를 제외한 115개의 사례를 가지고 사례검색을 시도해보았다. 질의조건을 완화하기 위해 이용된 추상화 계층구조는 <그림 5 ~ 7>과 같다. 일반적 검색방법으로 검색이 실패하여 질의완화를 할 경우, 개체의 표면적 속성측면에서 질의조건을 완화하게 되면 특정집단의 개체들이 갖는 주요 속성들만을 이용해서 질의조건을 단순화시킬 수 있도록 중요도에 따라 일부 속성들의 순서화된 리스트를 제공해주어야 한다. 따라서 속성추상화 계층구조가 문제해결에 유용한 사례를 제공해주기 위해서는 1차적으로 용도에 따라 개체들을 집단별로 분류한 후 각 용도에 해당하는 개체들의 일부 속성만을 중요도에 따라 순서적으로 나타내야 하며, 그 형태는 <그림 5>와 같다.

만약 개체의 구조적 측면에서 질의조건을 완화하게 되면 개체의 구성요소중 주요 부분에 대한 정보만을 이용해서 사례들을 검색할 수 있도록 개체의 주요 부분을 지정 해주어야 하며, 이는 구조추상화 계층구조를 통해 이루어진다. 볼트는 구조적 측면에서 볼 때 먼저 머리의 유무에 따라 분류될 수 있다. 머리의 유무에 따라 주요 부분은 머리부분 혹은 나사부분이 되며, 그 형태는 <그림 6>과 같다.



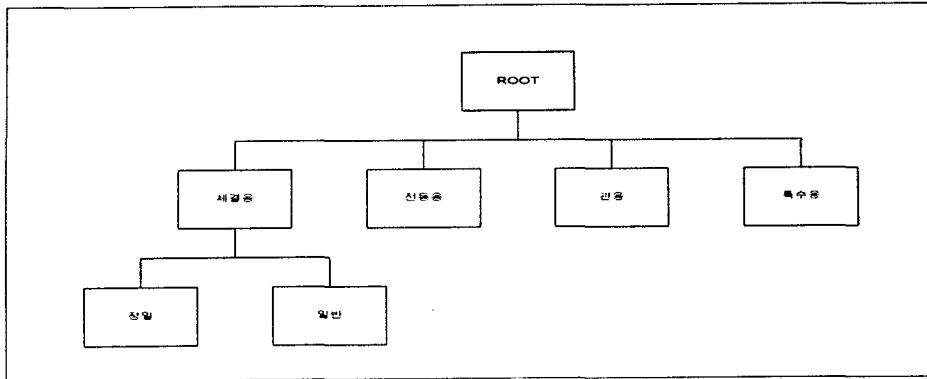
<그림 5> 볼트의 속성추상화 계층구조



<그림 6> 볼트의 구조추상화 계층구조

그리고 개체의 기능적 측면에서 질의조건을 완화하게 되면, 개체의 이용목적에 따른 속성만을 이용해서 관련 사례들을 검색할 수 있도록 용도에 대한 정보를 제공해 주어야 하며, 이는 목표추상화 계층구조를 통해 이루어진다. 볼트의 용도는 체결용, 전동용, 관용, 특수용 등으로 분류될 수 있으며, 체결용은 정밀용과 일반용으로 재분류될 수 있다. 따라서 볼트에 대한 목표추상화 계층구조는 <그림 7>과 같다.

<표 3>은 검색결과를 나타낸 것이다. 총 23개의 질의문에 대해 일반적 검색과정에 의한 계획수립이 가능한 횟수는 16회였으며, 계획수립이 불가능한 횟수는 7회였다. 일반적 검색과정에 의해 계획수립이 불가능한 7회의 질의문에 대해 질의완화에 의한 검색과정을 적용해본 결과 모두 관련 사례들의 검색이 가능하였다. 이러한 결과는 일반적 검색과정에 의해 지원될 수 없는 모든 문제들에 대해 관련 사례들이 존재하는 경우에는 질의완화기법을 이용함으로써 문제해결의 지원이 가능함을 나타내 준다.



<그림 7> 볼트의 목표추상화 계층구조

<표 3> CAPP 문제에 대한 적용결과

검 색 기 준	
일반적 검색	결의안회의 의한 검색
16회	7회

IV. 결론

의사결정자가 직면하는 문제들은 과거의 문제해결경험과 모든 측면에서 항상 새로운 것은 아니나 완전히 동일하지도 않을 것이다. 즉, 대부분의 문제들은 과거에 해결한 문제들과 많은 측면에서 공통적인 요소를 지니고 있을 것이다. 따라서 과거의 문제해결경험은 새로운 문제해결에 많은 도움을 줄 수 있다. CBR은 과거의 경험을 이용해서 문제를 해결하려는 기법으로서 주요 연구대상이 되는 부분은 정적인 측면에서는 사례의 표현과 구조화문제가 있으며, 동적인 측면에서는 사례의 검색절차와 수정이라는 해결안 생성과정이 있다. 본 논문에서는 동적 측면중 검색과정을 주요 연구대상으로 삼고 있다.

과거의 문제해결경험을 효과적으로 활용하기 위해서는 관련 사례들을 식별하기 위한 강력한 검색절차가 필요하다. 이를 위해 본 논문에서는 AQA 기법을 도입해서 검색능력을 강화시킨 협동적 사례검색절차를 제시하였다. 기존의 검색절차들은 주로 수치적 평가함수를 이용해서 관련 사례들을 추출하기 때문에 비계량적 속성이나 결측치가 있을 경우에는 사례검색이 어려웠다. 그러나 협동적 검색과정에서는 추상화 계층구조를 도입하여 비계량적 속성이나 결측치가 있는 사례들의 검색이 가능하게 하였을 뿐만 아니라 초기 검색에서 주요 속성들의 지정이 잘못되어도 다양한 측면에서 관련 사례들을 검색할 수 있게 하였다.

또한 검색된 결과들중 새로운 문제에 적합한 기저사례를 선택하기 위해 유사성개념의 3가지 차원인 개체 유사성, 구조적 유사성, 기능적 유사성을 이용한 선호휴리스틱절차를 이용하고 있다. 제시한 선호휴리스틱절차의 이용으로 인해 얻을 수 있는

이점은 문제해결시 사용자의 개인적인 선호도를 보다 잘 반영하므로써 해결안의 수용도가 높아진다는 것이다.

연구결과들의 유용성을 평가하기 위하여 CAPP 문제를 대상으로 문제해결능력에 대한 테스트를 실시하였다. 테스트를 실시한 결과, AQA를 이용한 협동적 응답기능으로 인해 문제해결능력이 향상되었다는 것을 알 수 있었다.

향후 지속적으로 연구가 계속되어야 할 부분은 동적 측면의 기존 해결안의 수정 부분이다. 이는 기존의 사례가 항상 새로운 문제와 일치하지는 않기 때문이다. 따라서 기존의 사례가 새로운 문제와 부분적인 차이가 있을 경우, CBR은 기존의 사례가 지닌 해결안을 부분적으로 수정하므로써 문제를 해결하려는 방법이므로 위의 검색과정에 대한 연구결과가 보다 유용하게 적용되기 위해서는 해결안의 수정에 관한 연구가 뒤따라야 할 것이다.

참 고 문 헌

- [1] 김진백, "볼트의 자동공정계획수립을 위한 CBR시스템의 개발," *경영정보학연구*, 제 9권, 제 2호, 1999, pp.19-38.
- [2] 편집부, *기계설계규격 및 데이터북*, 일진사, 1997.
- [3] Aamodt, A. and E. Plaza, "Case Based Reasoning : Foundational Issues, Methodological Variations, and System Approaches," *Artificial Intelligence Communications*, Vol.7, No.1, 1994, pp.39-59.
- [4] Alting, L. and H. Zhang, "Computer Aided Process Planning : the State-of-the-Art Survey," *International Journal of Production Research*, Vol.27, No.4, 1989, pp.553-585.
- [5] Chu, W. W. and Q. Chen, "Neighborhood and Associative Query Answering," *Journal of Intelligent Information Systems*, Vol.1, 1992, pp.355-382.
- [6] Forslund, G., "Toward Cooperative Advice-Giving Systems - A Case Study in Knowledge-Based Decision Support," *IEEE Expert*, August, 1995, pp.56-62.
- [7] Gaasterland, T., P. Godfrey, and J. Minker, "Relaxation as a Platform for Cooperative Answering," *Journal of Intelligent Information Systems*, Vol.1, 1992, pp.293-321.
- [8] Gupta, K. M. and A. R. Montazemi, "A Connectionist Approach for Similarity Assessment in Case-Based Reasoning Systems," *Decision Support Systems*, Vol.19, 1997, pp.237-253.
- [9] Iran, S. A., H.-Y. Koo, and S. Raman, "Feature-Based Operation Sequence Generation in CAPP," *International Journal of Production Research*, Vol.33, No.1, 1995, pp.17-39.
- [10] Kaplan, S. J., "Cooperative Responses from a Portable Natural Language Query System," *Artificial Intelligence*, Vol.19, 1982, pp.165-187.
- [11] Kolodner, J., "Judging Which Is the "Best" Case for a Case-Based Reasoner," *Proceedings : Case-Based Reasoning Workshop*, May, 1989, pp.77-81.
- [12] Kolodner, J., *Case-Based Reasoning*, Morgan Kaufmann, San Mateo, CA, 1993.
- [13] Koton, P., "Evaluating Case-Based Problem Solving," *Proceedings : Case-Based Reasoning Workshop*, May, 1989, pp.173-175.
- [14] Liang, T., "Analogical Reasoning and Case-Based Learning in Model Management Systems," *Decision Support Systems*, Vol.10, 1993, pp.137-160.
- [15] Liang, T. and B. R. Konsynski, "Modeling by Analogy - Use of Analogical Reasoning in Model Management Systems," *Decision Support Systems*, Vol.9, 1993, pp.113-125.

- [16] Mahapatra, R. and A. Sen, "Case Base Management Systems : Providing Database Support to Case-Based Reasoners," *Journal of Database Management*, Vol.5, No.2, 1994, pp.19-29.
- [17] O'Leary, D. E., "Verification and Validation of Case-Based Systems," *Expert Systems with Applications*, Vol.6, 1993, pp.57-66.
- [18] Rogers, J. S., P. A. Farrington, B. J. Schroer, and R. G. Hubbard, "Automated Process Planning System for Turned Parts," *Integrated Manufacturing Systems*, Vol.5, No.4/5, 1994, pp.41-47.
- [19] Srinivasan, P., "The Importance of Rough Approximations for Information Retrieval," *International Journal of Man-Machine Studies*, Vol.34, 1991, pp.657-671.
- [20] Thagard, P. and K. Holyoak, "Why Indexing Is the Wrong Way to Think about Analog Retrieval," *Proceedings : Case-Based Reasoning Workshop*, May, 1989, pp.36-40.
- [21] Tsatsoulis, C. and R. L. Kashyap, "Case-Based Reasoning and Learning in Manufacturing with the TOLTEC Planner," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.23, No.4, 1993, pp.1010-1022.

<Abstract>

Cooperative Case-based Reasoning Using Approximate Query Answering

Jin-Baek Kim

Case-Based Reasoning(CBR) offers a new approach for developing knowledge based systems. CBR has several research issues which can be divided into two categories : (1) static issues and (2) dynamic issues. The static issues are related to case representation scheme and case data model, that is, focus on casebase which is a repository of cases. The dynamic issues, on the other hand, are related to case retrieval procedure and problem solving process, i.e. case adaptation phase. This research is focused on retrieval procedure

Traditional query processing accepts precisely specified queries and only provides exact answers, thus requiring users to fully understand the problem domain and the casebase schema, but returning limited or even null information if the exact answer is not available. To remedy such a restriction, extending the classical notion of query answering to approximate query answering(AQA) has been explored. AQA can be achieved by neighborhood query answering or associative query answering. In this paper, neighborhood query answering technique is used for AQA. To reinforce the CBR process, a new retrieval procedure(cooperative CBR) using neighborhood query answering is proposed. An neighborhood query answering relaxes a query scope to enlarge the search range, or relaxes an answer scope to include additional information.

Computer Aided Process Planning(CAPP) is selected as cooperative CBR application domain for test. CAPP is an essential key for achieving CIM. It is the bridge between CAD and CAM and translates the design information into manufacturing instructions. As a result of the test, it is approved that the problem solving ability of cooperative CBR is improved by relaxation technique.