

사례기반 상품추천시스템 개발을 위한 사례표현에 관한 연구

정대율* · 하동현**

요 약

사례기반추론기법을 통한 유사상품의 탐색과 사용자 요구에 적합한 상품추천을 위해서는 다양한 요구에 부응할 수 있는 사례베이스의 구축이 우선되어야 한다. 이에 본 연구에서는 인터넷 쇼핑몰의 상품추천시스템에서 번들상품구성문제를 표현하는데 적합한 사례표현기법을 개발하며, 유사사례를 추출하기 위한 유사도 척도의 개발에 연구의 첫 번째 주안점을 둔다. 본 논문에서는 번들상품추천을 위한 사례표현기법으로 속성-값(feature-value) 방식인 프레임(frame) 형식을 사용하고 있다. 또한 유사도 측정을 위하여 각 속성(행사, 예산, 참여자 수, 고객 거주지 등)에 대하여 유사도 테이블을 작성하고, 이들 속성들의 가중합계방법을 이용하고 있다.

Key words : CBR(Case-Based Reasoning), Case Representation, Bundle Productions Suggestion

1. 서론

최근, 전자 상거래 시장이 본격적으로 형성되고 있으며 이용 빈도 또한 높아지고 있는 추세이다. 하지만 이러한 추세에도 불구하고 현재의 많은 전자상거래시스템은 전통적인 상거래에 비해 제공되는 기능이 상대적으로 부족하여 사용자에게 많은 불편을 주고 있다. Vollrath et. al.(1998)의 연구에 의하면 전자 상거래가 상당히 유용하지만 실제 전자 상거래를 통한 비즈니스의 양은 얼마 많지 않은 이유로 ① 인터넷상에서 거래의 불안정한 보안 보증, ② 새로운 미디어(internet)에 대한 불안정한 수용, ③ 경험과 예술적 감각에 의존한 온라인 카탈로그의 인터페이스 디자인, ④ 원하는 상품을 직접 찾기가 어려움 등을 들고 있다.

본 연구에서는 인터넷 쇼핑몰에서 Vollrath 등(1998)이 제시한 네 번째 문제를 해결하기 위한 상품추천시스템의 개발에 관심을 두고 있다. 지금까지 상품추천을 위한 대부분의 시스템들이 사용자의 요구사항과 사용자 프로파일을 기반으로 하고 있다. 상품추천을 위해 추천되는 기법으로는 ① 최적화 모델중심의 기법, ② 휴리스틱 규칙중심의 기법, ③ 제약조건만족(CSP) 기법, ④ 사례기반추론기법 등을 들 수 있다.

근래에 들어, 과거 경험을 토대로 새로운 문제를 해결하는 방법으로 사례기반추론(case-based reasoning)에 대한 연구가 활발하다. 사례기반추론은 일상 생활에서 인간이 생각하는 것처럼 과거의 경험적 사실들이 저장되어 있는 사례 베이스(case base)의 유사 사례(analogical case)를 기반으로 새로운 문제를 해결하는 지식 처리 방법이다. 즉, 사례 베이스로부터 해결할 문제에 적용할 최적의 유사 사례를 검색하는 일은 근사해의 산출과 밀접한 관련이 있다. 사례기반추론을 중심으로 한 시스템은 이미 여러 응용 분야에서 성공적으로 운영되고

있으며, 상품화될 시스템 역시 여러 분야에서 다양하게 개발되고 있다. 특히, 사례기반추론시스템은 지속적으로 발생하는 과거 경험들을 이용하여 유사한 문제를 해결하려는 분야에 적합하다.

일반적으로 사용되는 관계형 데이터베이스시스템은 사례기반추론의 관점에서 볼 때, 실세계에 있었던 과거 사례들의 저장소로 볼 수 있다. 역으로 말해 과거에 있었던 사례 데이터를 관계형 데이터베이스의 릴레이션으로 표현할 수 있다는 것을 의미한다. 그러나 어떤 문제를 해결할 지식으로 관계형 데이터베이스 내에 저장되어 있는 사례를 활용하기 위해서는 해결할 문제의 특성과 완전히 일치하는 조건으로 관계 데이터베이스를 검색하는 방법만 가지고서는 문제를 해결할 사례를 찾는 것은 불가능하다. 왜냐하면, 해결할 문제와 똑같은 특성을 갖는 과거 사례가 실제로 존재할 가능성은 거의 없기 때문이다.

사례기반추론기법을 통한 유사상품의 탐색과 사용자 요구에 적합한 상품추천을 위해서는 다양한 요구에 부응할 수 있는 사례베이스의 구축이 우선되어야 한다. 이에 본 연구에서는 상품구성문제를 표현하는데 적합한 사례표현기법을 개발하며, 유사사례를 추출하기 위한 유사도 척도의 개발에 연구의 첫 번째 주안점을 둔다.

사례표현을 위한 도메인(domain)으로 수산물 인터넷 쇼핑몰의 번들상품추천(bundle products suggestion)분야를 선정하였다. 번들상품추천이란 특정 목적에 맞추어 관련성이 높은 상품들을 하나의 패키지(packet) 또는 번들로 묶어 추천하는 것을 말한다. 수산물 쇼핑몰의 경우 특정한 집안행사(제사, 생일, 돌, 회갑, 집들이 등)시 사용되는 수산물의 상품구성(예, 제사용 어물)이 요구된다. 본 연구에서는 이를 차림상(번들상품)으로 정의하고, 이의 추천을 위한 문제를 해결하기 위하여 사례기반추론기법을 적용하고자 한다.

* 경상대학교 경영대학 경영학부 부교수

** 경상대학교 대학원 경영정보학과 석사

따라서 본 논문에서는 인터넷 쇼핑몰에서 자료 저장소로 흔히 이용되는 관계형 데이터베이스시스템의 미비점을 해결하기 위하여 사례기반추론기법을 이용하여 고객이 원하는 상품을 적절하게 추천할 수 있는 시스템을 개발하는데 있어 가장 핵심이 되는 사례베이스를 표현하고 유사도 측정방법을 제시하는데 그 목적이 있다.

II. 상품추천시스템 관련연구

전자상거래와 관련하여 지금까지 연구된 지능형 에이전트 또는 정보 에이전트에 대한 종합적인 연구조사는 Klusch(2001)의 연구들 들 수 있다. Klusch(2001)는 현재도 그렇지만 앞으로의 추세는 역시 인터넷과 관련된 에이전트 개발이 많은 비중을 차지할 것으로 예측되며, 특히 초고속 정보망과의 연계를 통하여 전자상거래도 활성화될 것이라고 예측하였다.

고객의 상품구매활동을 도와주기 위해서는 다양한 에이전트들이 요구될 수 있다. Guttman et al.(1998)은 고객의 구매행위단계에 따라 6 가지 종류의 에이전트(① 소비자 요구 파악, ② 상품탐색 및 추천, ③ 판매자 추천, ④ 협상, ⑤ 구매 및 배달, ⑥ 제품서비스 및 평가)가 개발될 수 있음을 제시하였다. 이 중에서 본 연구와 관련된 상품탐색 및 추천 에이전트는 소비자의 요구가 파악된 후, 소비자의 요구를 가장 잘 충족시킬 수 있는 제품을 추천하는 에이전트이다. 이러한 에이전트는 소비자의 정보검색비용을 낮추어준다.

지금까지 개발된 대표적인 상품추천에이전트로는 PersonaLogic(www.personalogic.com), Firefly(ww.firefly.com), Techno Salesman (http://www.technoagent.co.kr) 등을 들 수 있다. PersonaLogic은 소비자의 구매 대상이 될 수 있는 다양한 제품군 중에서 소비자가 명시한 특정 조건(제품, 특성)을 충족시키는 제품만을 우선 순위와 함께 제공해주며, 유저의 프로필에 의해 상품을 좁힐 수 있도록 하는 에이전트이다. Firefly는 음반이나 서적과 같은 상품을 비교하여 소비자에게 추천해 주는 에이전트로 합작 필터링(collaborative filtering)기술을 이용하고 있다. Techno Salesman은 필터링(filtering)과 스코어링(scoring)을 통해 적절한 수의 상품을 추천하여 고객의 요구에 상응하는 제품이 없을 경우에는 릴렉세이션(relaxation)을 통하여 고객의 요구에 가장 적절한 제품을 추천해 주는 기능을 제공하고 있다.

본 연구와 관련하여 특히 관심을 끄는 것은 상품구성을 추천하는 지능형 에이전트에 관한 연구로서 Rahmer & Voß(1996), Vollrath, et al.(1998), Kim S.H. et al.(2000), Kim, Steven H. et al.(2000), 성백균 등(2000)의 연구이다. 이들 연구들을 요약하면 <표 1>과 같다.

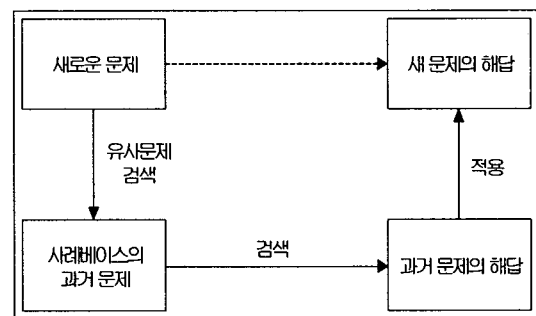
<표 1> 상품추천 에이전트에 대한 연구

연구자	연구내용
Rahmer & Voß (1996)	사례기반추론(CBR)기법을 이용한 Telecooperation Systems의 구성
Vollrath, et. al. (1998)	사례기반 추론기법을 이용하여 온라인 상품 카탈로그와 상품 데이터 베이스를 활용한 지능형 판매 에이전트 시스템 제안
Kim S.H. et. al. (2000)	계약만족(CSP)기법을 이용한 컴퓨터 구성품의 추천
Kim, Steven H. et. al. (2000)	Multistrategy filtering 접근법을 이용한 개인화된 상품추천
성백균 등 (2000)	사례기반 추론을 이용한 판매 지원 에이전트를 제안. 다중 에이전트 시스템 설계방안 제안하고 사용자의 취향을 알아내는 사례기반의 지능형 에이전트 설계

III. 사례기반추론과 사례표현

3.1 사례기반추론의 정의

사례기반 추론(Case-based Reasoning: CBR)은 현재 해결해야 할 새로운 문제와 과거에 해결했던 문제(사례)를 비교하여 현재의 문제를 해결하는 방법이다. 사례기반 추론의 기본적인 가정은 사례와 해결해야 할 문제가 유사(similar)하며, 해결방법 또한 유사하다는 데 있다. CBR의 일반적인 과정을 도식화하면 <그림 1>과 같다(Vollrath et. al., 1998, p. 5).



<그림 1> 사례기반추론의 기본 모형

CBR은 기억장치에서 현재의 문제와 유사한 이미 해결된 문제를 찾고, 과거의 문제와 현재의 문제간의 차이를 분석하여 이전의 해결책을 현재의 문제에 맞게 조정하는 과정을 거친다. CBR의 학문적 배경은 크게 심리적인 것과 컴퓨터 공학적인 것으로 대별된다.

심리학적 모델로서의 CBR은 Schank(1982)에 의해 개발된 동적 기억이론(dynamic memory theory)과 문제해결(problem solving) 및 학습(learning)에서의 회상(reminding)의 역할에 관한 연구를 기점으로 본격적으로 시작되었다. 동적 기억이론은 개별적인 경험이 어떻게 기억 속에 저장될 수 있는지, 그들이 어떻게 결합되고 추상화될 수 있는지, 그리고 그들이 어떻게 검색되고 사용될 수 있는지를 기술하기 위한 골격을 제공하였다. 이러한 CBR의 기본원리는 사람이 복잡한 문제를 해결하기 위하여 과거에 문제를 해결했던 경험을 되살려서 비슷한 사례를 추출하여 주어진 문제를 풀고자 하는 유사추론, 혹은 경험적 추론을 하는 것에 기인한다.

여기에서 CBR과 유사추론은 때때로 동의어로 사용되어지기도 하지만 엄밀히 말해서 두 방법은 서로 다른 것이다. CBR은 한 적용분야(intra-domain)안에서 사용되는 방법이고, 유사추론은 서로 다른 적용분야(across-domain)에서 사용되는 방법이다(Aamodt & Plaza, 1994. p. 15).

한편 컴퓨터 모델로서의 CBR은 규칙기반시스템의 한계를 극복하기 위한 방안으로 연구되었다. 1970년대부터 개발되기 시작한 대부분의 전문가 시스템에서는 규칙을 지식의 기본 단위로 표현하고, 규칙을 통해 문제해결을 위한 추론을 수행하였다. 그러나 규칙기반시스템의 개발에 대한 경험이 증가할수록 지식 획득의 어려움, 과거 결과에 대한 기억 불가능, 불확실한 자료의 표현과 처리가 불가능, 지식의 수정 및 추가의 어려움 등 규칙기반 시스템의 단점을 발견하게 되었다. CBR은 인간전문가의 지식 및 추론과정에 대한 인지과학(Cognitive Science)적 고찰을 토대로 제안되어 규칙기반시스템의 문제를 해결하기에 적절하기 때문에 빠르게 발전하였다.

3.2 사례의 구조와 표현 형식

일반적으로 사례기반 시스템의 구축에서 가장 중요한 문제는 사례베이스의 구축을 위해 사례에 관한 정보를 표현하는 방법이며, 사례의 표현에서 고려해야 할 사항은 사례의 구조와 표현 형식이다. 즉, 사례를 어떻게 표현하여 좋은 사례베이스를 구축할 수 있는가 하는 것이다.

CBR에서 사례(case)란 어떤 영역(domain)의 문제에서 특정한 상황을 표현한 것으로 신경회로망에서의 형태(pattern)에 해당되는 것이다. 그리고 사례베이스(case base)는 이러한 사례들을 모아놓은 것으로 규칙기반 전문가 시스템에서 지식베이스(knowledge base)에 해당하는 것을 말한다. 사례기반추론을 하기 위하여서는 먼저 사례에 대한 색인 방법을 선택하여 사례기반추론의 구조를 결정해야 한다.

CBR은 일차적으로 저장된 사례에 기초해서 추론하므로 사례를 어떤 구조를 이용해서 저장하는가는 시스템의 전체적인 성능을 결정하는 데에 큰 영향을 끼친다. 또한 사례 베이스를 확장해나가는 과정에서 새로운 사례를 기존의 사례 베이스에 통합하는 것과도 사례의 저장 구조는 밀접한 관련이 있다. 사례의 저장 구조와 일반적인 문제 영역에서의

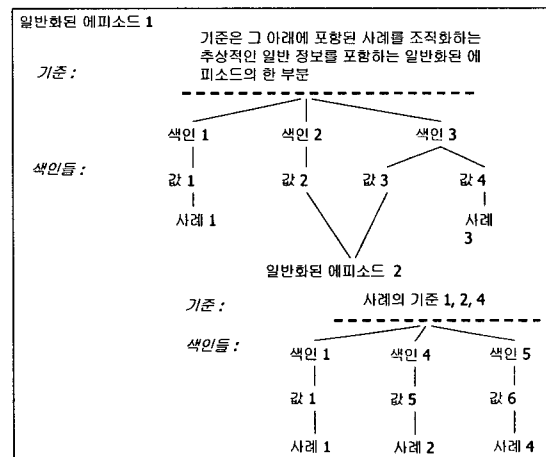
지식 모델의 통합 문제도 사례의 표현과 관련된 또 하나의 문제이다. <표 2>는 사례의 구조에 대한 내용이다.

<표 2> 사례의 구조

구성 항목	내용
문제/상황(Problem/Situation)	문제와 그 문제가 해결될 당시의 상황 묘사
해(Solution)	문제가 발생했을 당시의 전문가의 반응 및 추론된 해결 방법을 서술

사례의 표현 모델로서 많은 영향을 끼친 두 가지는 Schank(1982)의 동적 메모리 모델과 Kolodner(1988)의 범주-표본 모델을 들 수 있다.

동적 메모리 모델에서는 사례가 E-MOP(Episodic Memory Organization Packets)라는 요소들의 계층 구조의 형태로 저장된다. 이 모델의 기본 아이디어는 유사한 성질을 가지고 있는 구체적인 사례들을 일반화된 에피소드(Generalized Episode: GE)라고 불리는 보다 일반적인 구조하에 정리한다는 것이다. 이 모델에서 사례의 전체적인 저장 구조는 식별 네트워크(discrimination network)의 형태를 띠며, 새로운 사례가 저장될 때 이 사례의 특성 값이 기존의 사례의 특성 값과 일치하면 이들을 묶어 하나의 일반화된 에피소드를 형성하는 식으로 메모리 구조가 동적으로 변하게 된다. <그림 2>는 동적 메모리 모델에서 일반화된 에피소드의 한 예를 보여준다.

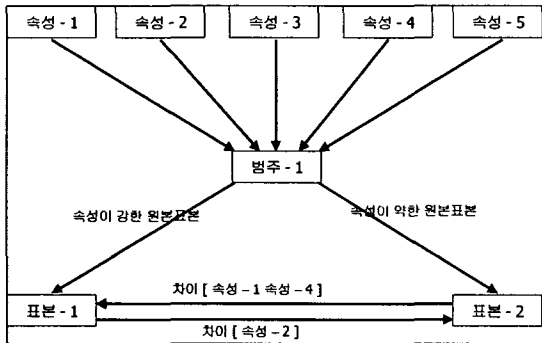


<그림 2> 동적 메모리 모델에서 사례와 일반화된 에피소드의 구조

범주-표본 모델에서는 사례를 표본(exemplar)이라고 부르는데 이 모델의 근저에 깔려있는 생각은 실세계의 개념들은 외연적으로 정의되어야 하고 일반화는 매우 조심스럽게 시도해야 하며 어떤 사례를 어떤 범주에 포함시킬 때 특성의 종류에 따라 중요도가 달라져야 한다는 것이다.

이 모델에서 사례 저장 구조는 범주, 사례와 색인 포인터의 네트워크 형태를 띤다. 색인 포인터는

문제의 특성 구분에서 범주나 사례로 연결되는 특성 고리(feature link), 범주에서 관련된 사례로 연결되는 표본 고리(exemplar link), 한 사례에서 약간의 특성 값만 차이를 보이는 이웃한 사례를 연결하는 차이 고리(difference link)의 세 종류로 분류된다. <그림 3>은 범주-표본모델의 한 예이다.



<그림 3> 범주-표본 모델의 구조

사례는 대개 특정한 결과를 이끌어내는 주요한 특성들(features)의 리스트(list)로서 표현된다. 사례를 표현하는 방법에는 사례의 특성을 정의하는 속성(feature)의 단순한 집합으로 사례를 표현하는 특징 표현(feature representation)방법, 속성의 값을 포함하는 슬롯(slot)의 집합으로 구성된 프레임(frame)이 하나의 노드(node)가 되어 계층적 구조(hierarchical structure)로 사례를 표현하는 프레임 방법, 그리고 프레임의 확장된 개념으로 사례를 표현하는데 있어서 규칙(rule), 프레임, 추상화(abstraction), 상속(inheritance) 등과 같은 개념을 혼합 적용하여 사례를 표현하는 방법으로 Schank에 의해 제시된 MOP(Memory Organization Package) 방법 등이 있다. 이들 방법 중에서 알맞은 사례 표현방법을 선택하기 위해서는 주어진 문제 영역의 특성을 고려하여 결정하여야 한다.

사례에 알맞은 표현 방법을 결정한 후에는 사례에 대한 색인(index) 방법을 선택하여야 한다. 색인은 어떤 특정한 사례가 주어졌을 때, 사례 베이스에서 비슷한 사례를 보다 효율적으로 찾기 위해서 사례 베이스의 사례들을 조직화하고 표지(label)을 붙이는 것이다.

IV. 상품추천시스템 사례표현과 유사도 측정

4.1 시스템 정의

본 논문에서 제안하는 번들상품추천시스템은 인터넷 쇼핑몰을 통하여 수산물을 구입하려는 고객에게 고객이 원하는 상품을 차립상이라는 패키지형 상품으로 추천하는 시스템이다. 차립상은 고객이 수산물을 이용하려는 행사의 내용 또는 고객의 취향에 알맞은 상품들로 구성된 번들상품단위를 말한다. 즉, 이 시스템에서는 단일 상품 추천뿐 아니라

수산물을 사용하려는 목적에 따라 다양한 수산물 구성을 사례베이스로부터 검색하여 고객이 가장 원하는 상품 구성을 추천해 주는 방식이다.

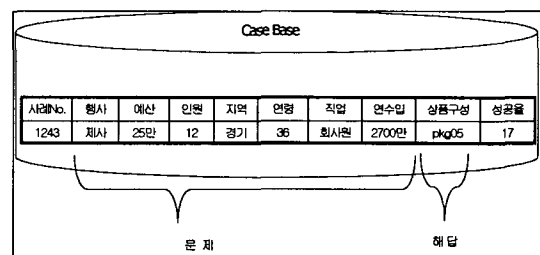
본 연구에서 지향하는 바는 수산물 인터넷 쇼핑몰에서 고객이 자신의 각종 집안행사(제사, 생일, 돌, 회갑 등)시 구입해야할 수많은 수산물들을 큰 고민 없이 최적의 상품구성을 선택할 수 있도록 지원해주는 시스템을 만드는 것이다. 이를 위해 쇼핑몰 개발자와 운영자는 요리연구가 등으로부터 고객의 상품선택과 관련된 각종 정보들을 수집하여 다양한 상품의 구성을 준비한다. 한편 고객은 자신의 요구사항(행사시 참여인원, 가능한 예산액, 지역, 요리종류 등)을 입력하면, 쇼핑몰 시스템은 최적의 상품을 구성하여 고객에게 제시한다.

4.2 사례 표현

사례기반추론기법에서 문제해결은 현재 해결해야 할 문제와 유사한 문제를 찾아 그 문제의 해결방법으로부터 해결해야 할 문제의 해결방법을 모색하는 것으로 과거 유사한 문제와 그 해결방법을 사례라고 한다.

본 논문에서 번들상품추천시스템의 사례는 다음 문제와 해결방법으로 구성된다. 먼저 문제는 특정 고객이 구입하려는 상황이며 그러한 상황에서 그 고객이 구입했던 상품들을 해결방법이 된다. 사례 베이스에 저장될 사례의 표현 방식으로 고객의 정보(직업, 나이, 지역, 연수입 등)와 행사에 대한 정보(행사의 종류, 인원수, 예산 등)를 숫자 또는 문자로 표시한다. 각 항목들은 유사도 측정에서 개별 속성으로 다루어진다.

본 연구에서는 <그림 4>와 같이 입력과 검색의 편의를 위해 기본적으로 속성-값(feature-value) 방식인 프레임(frame) 형식으로 사례를 표현하며, 일반적으로 자료의 축적을 위해 사용되는 관계형 데이터베이스와의 자료 공유를 위하여 관계형 데이터베이스의 자료구조인 테이블 형식으로 논리적인 사례 베이스를 설계하도록 하였다.



<그림 4> 사례 표현의 예

<그림 4>는 사례 베이스에 저장되는 사례의 표현 방식에 대한 예이다. 하나의 사례는 문제와 그에 대한 해답으로 이루어진다고 앞에서 밝힌 바 있다. 각각의 사례는 '사례No.'로 사례 베이스 속에서 유일하게 존재한다.

- 1) 사례No. : 사례 베이스 내에 존재하는 사례들을 구별하기 위한 식별자이다. 사례 베이스에서 사례No.는 유일한 값을 가지는 기본

키이다.

- 2) 행사 ~ 연수입 : 유사 사례의 검색이나 선택시 해결할 문제와 유사 사례간에 얼마나 유사한 지를 평가할 수 있는 항목들로 구성된다. 즉, 유사도 평가에 따라 유사 사례를 사례 베이스로부터 추출할 때 이용한다.
- 3) 상품 구성 : 사례 내에 존재하는 문제 해결 방법을 말하며 여기서는 행사 ~ 연수입의 속성의 내용에 따라 구입되는 상품의 종류 및 수량에 대한 내용이 담긴다. 상품 구성은 행사 ~ 연수입의 내용에 따라 하나의 상품이 되거나 또는 복수의 상품으로 구성된다.
- 4) 성공률 : 문제에 사례가 적용될 때 그 사례가 적용되었음을 누적 점수로 입력된다.

4.3 유사도 측정

4.3.1 유사도 측정방법

전통적 방식의 질의 처리는 명시된 조건과 정확히 일치하는 데이터만을 제공해 주는 엄밀 질의응답(exact query answering)기능을 따른 방식이라 할 수 있다. 따라서 엄밀 질의응답기능은 질의조건 q를 완전히 충족시켜 주는 객체의 경우 1, 그렇지 않은 객체에 대해서는 0의 값을 발생시켜 1의 값을 갖는 객체들에 대한 관련 정보만을 제공해주는 특성 함수(characteristic function)로 볼 수 있다.

그러므로 전통적 방식인 엄밀 질의응답기능에 따라 시스템을 이용하려면 사용자는 문제영역과 관련된 DB 스키마를 충분히 이해하고 있어야만 한다. 이는 검색 시스템의 이용률을 저하시키는 주요 원인 중 하나가 된다.

따라서 사용자의 부담을 줄여 주고 검색률을 높이기 위해서는 질의조건을 완화시켜 부분적으로 질의를 만족시키는 객체에 대한 정보도 제공하여야 한다. 이러한 질의응답방식을 근사 질의응답(approximate query answering)이라고 한다. 본 논문에서는 유사도를 평가하는 함수로서 근사 질의응답기능을 사용하였다.

사례 베이스의 사례와 고객이 질의한 문제의 유사도를 평가하기 위해서 먼저 필요한 작업이 유사도를 측정할 알고리즘을 작성하는 것이 필요하다. 유사도는 사례와 해결해야 할 질의의 비슷한 정도를 나타내는 것으로, 전체 유사도를 추론해 내기 위해 각 속성별 유사도 측정을 하여야 한다. 유사도를 표현하는 방식으로 0과 1사이의 숫자로 표현한다. 값 0이 뜻하는 것은 질의를 전혀 만족시키지 못함을 말하며, 1은 질의한 것에 일치한다는 의미이다(Vollrath, et. al., p. 7). 그리고 0과 1사이의 값은 부분적으로 질의를 충족한다는 의미이다.

예를 들어, 질의의 내용 중 수산물을 사용하려는 행사가 제사일 경우 사례의 속성 중 행사가 '제사'인 사례의 유사도는 1이 된다. 하지만 실제 모든 속성값을 숫자로 표현하는 데는 한계가 있다. 이런 경우에는 모든 가능한 속성값의 쌍에 대한 유사도를 정의한 유사도 테이블을 작성하여 사용할 수가 있다.

이와 같이 개별 속성에 대한 유사도 측정이 끝

나게 되면 사례와 질의 전체에 대한 유사도를 추론할 수 있다. 전체 유사도를 추론하는 일반적인 방법으로 다음과 같은 예를 보자. 속성 q_1, q_2, \dots, q_n 으로 표현되는 질의와 c_1, c_2, \dots, c_n 으로 표현되는 사례 c 가 있다고 고려해 볼 때, q 와 c 사이의 유사도 σ 는 각 개별 속성의 유사도인 σ_i 로부터 계산된다. 다음 식은 각 항목간의 유사한 정도를 나타내는 유사성매칭함수이며 해결해야 할 문제의 속성과 과거 사례의 속성이 가지는 값에 따라 다른 두 가지 형태의 함수를 적용한다.

$$\sigma(q, c) = \sum_{i=0}^n w_i \sigma_i(q_i, c_i)$$

여기에서 $\sum_{i=1}^n w_i = 1$, 모든 i 에 대해 $w_i \geq 0$

이 식은 사례 검색을 위해 최근이웃 추출방법을 사용한 것으로서 최근이웃 추출방법은 가장 가까운 거리에 위치하는 이전 사례를 발견하는 알고리즘으로 이미 여러 인공지능 분야에서 그 유용성이 검증되었으며, 기계 학습분야에도 널리 이용되고 있다.

(1) 점비교함수(point matching function)

주어진 두 값, 여기서는 해결해야 할 문제의 속성값과 과거 사례의 속성값을 말하며 이 두 값이 얼마나 유사한가를 차이 값의 범위에 따라 0과 1사이의 값으로 평가하는 함수이다.

(2) 구간비교함수(interval matching function)

속성의 값이 일정한 구간을 가지는 값일 때 이용하는 것으로, 문제의 속성값과 사례의 속성값이 겹치는 부분(r)이 과거 사례의 속성값의 범위(R) 중 유사한 정도는 r/R 이 된다.

위의 식에서 w_i 는 가중치로써, 일반적으로 전문가에 의해 그 값이 주어지지만 개인적 선호도에 의해 조정될 수도 있다.

4.3.2 본고에서 제안하는 유사도 측정 방법

본 시스템에서 사용할 유사도 측정은 고객의 정보(직업, 나이, 지역, 연수입 등)와 행사에 대한 정보(행사의 종류, 인원수, 예산 등)의 속성에 대한 개별 유사도를 정의한 뒤, 개별 유사도의 가중합으로써 사례의 유사도가 결정된다. 사례의 유사도가 가장 높은 것을 우선적으로 고객에게 제시되며, 이를 고객이 수용할 경우 그 사례의 성공률은 1포인트 증가한다. 구체적인 과정은 다음과 같다.

개별 속성별 유사도의 값에 대한 정의가 필요하다. 이 과정은 사례 베이스의 사례와 고객의 질의를 비교하여 유사도를 측정하기 이전에 필요한 작업으로 유사도 테이블을 통해 표현한다. 예를 들어 행사별 유사도의 값을 정의하기 위한 준비 과정은 다음과 같다.

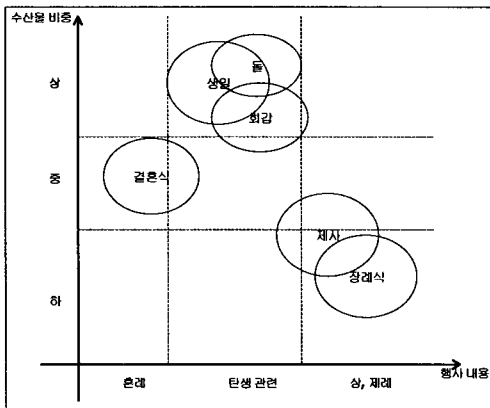
① 행사별 수산물의 선호 관계 조사

<표 3>을 바탕으로 각 행사별 수산물 소비 경향의 밀접한 정도를 다음 <그림 5>와 같이 나타낼 수 있다. <표 3>은 실제 수산물 소비 경향을 가상으로 표현한 것이므로 실제는 이와 다를 수 있다.

<표 3> 행사와 수산물 선호 관계

행사 종류 \ 수산물 종류	수산물 종류																
	돌	광어	참치	멍게	대구	연어	고등어	도다리	오징어	문어	새우	계	조개	굴	김	미역	통조림
돌	○	○	○	△	△	×	△	○	△	×	○	△	×	○	○	○	×
생일	○	○	○	△	△	△	○	○	△	△	○	×	×	○	△	○	△
결혼식	△	△	○	×	○	×	△	×	○	△	△	○	○	△	×	×	○
회갑	○	○	△	○	△	×	△	○	△	×	○	△	×	○	△	○	×
장례식	×	×	○	×	○	△	○	×	△	○	△	×	△	○	△	×	○
제사	×	×	○	×	○	○	△	×	○	○	△	×	○	○	×	×	△

※ ○: 많이 쓰임, △:중간 ×:쓰이지 않음



<그림 5> 수산물을 기준으로 한 행사별 밀집도

<표 3>과 <그림 5>를 볼 때, 생일과 회갑, 돌은 서로 비슷한 종류의 수산물로 구성되며, 제사와 장례식 또한 그러하다. 결혼식의 경우 쓰이는 수산물이 다른 행사와 차이가 있음을 알 수 있다. 이런 관계를 바탕으로 행사에 대한 유사도 테이블을 작성할 수 있다.

② 행사 속성에 대한 유사도 테이블 작성

<표 4>와 같은 행사 속성에 유사도 테이블을 작성한다. 유사도 테이블은 속성값의 쌍으로 나타내었다.

<표 4> 행사 속성에 대한 유사도 테이블

질의 \ 사례	돌	생일	결혼식	회갑	장례식	제사
돌	1	0.8	0	0.7	0.2	0.2
생일	0.8	1	0.2	0.8	0.1	0.2
결혼식	0	0.2	1	0.3	0.3	0.2
회갑	0.7	0.8	0.3	1	0.3	0.3
장례식	0.2	0.1	0.3	0.3	1	0.9
제사	0.2	0.2	0.2	0.3	0.9	1

③ 예산 및 행사 참여자수에 대한 유사도 테이블 작성
예산 속성과 참여자수 속성을 이산형으로 표현

하여 <표 5>와 <표 6>과 같은 유사도 테이블을 작성한다

<표 5> 예산에 대한 유사도 테이블

질의 \ 사례	5(만원) 이상 ~10미만	10~20	20~30	30~40	40~50	50~60
5이상~10미만	1	0.5	0	0	0	0
10~20	0.5	1	0.5	0	0	0
20~30	0	0.5	1	0.5	0	0
30~40	0	0	0.5	1	0.5	0
40~50	0	0	0	0.5	1	0.5
50~60	0	0	0	0	0.5	1

<표 6> 참여자수에 대한 유사도 테이블

질의 \ 사례	5명 이하	5초과 ~10이하	10~15	15~20	20~25	25~30
5명 이하	1	0.5	0	0	0	0
5초과 ~10이하	0.5	1	0.5	0	0	0
10~15	0	0.5	1	0.5	0	0
15~20	0	0	0.5	1	0.5	0
20~25	0	0	0	0.5	1	0.5
25~30	0	0	0	0	0.5	1

④ 고객 정보에 대한 유사도 테이블의 작성

고객 정보에 대한 유사도 테이블을 작성하기 위해, 고객 정보가 담겨 있는 테이블에서 수산물 선호와 관련 있는 속성들을 선택한 후, 각 속성에 대한 유사도 테이블을 작성한다. 예를 들어, 고객의 연령, 직업, 출신 지역, 연 수입 등은 수산물을 구입하는 데 어느 정도 영향을 끼치므로 이러한 속성에 대해 유사도 테이블을 작성하게 된다.

<표 7>은 지역속성에 대한 유사도 테이블이며, 고객의 연령과 직업 및 기타 상품 구입과 관련 있는 속성들을 위와 같은 방식으로 유사도를 정의 할

수 있다.

<표 7> 지역 속성에 대한 유사도 테이블

질의 사례	서울	경기권	충청권	전라권	강원권	경상권
서울	1	0.8	0.6	0.3	0.3	0.1
경기권	0.8	1	0.7	0.4	0.5	0.1
충청권	0.6	0.7	1	0.6	0.4	0.1
전라권	0.3	0.4	0.6	1	0.2	0.4
강원권	0.3	0.5	0.4	0.2	1	0.3
경상권	0.1	0.1	0.1	0.4	0.3	1

4.4 사례의 검색

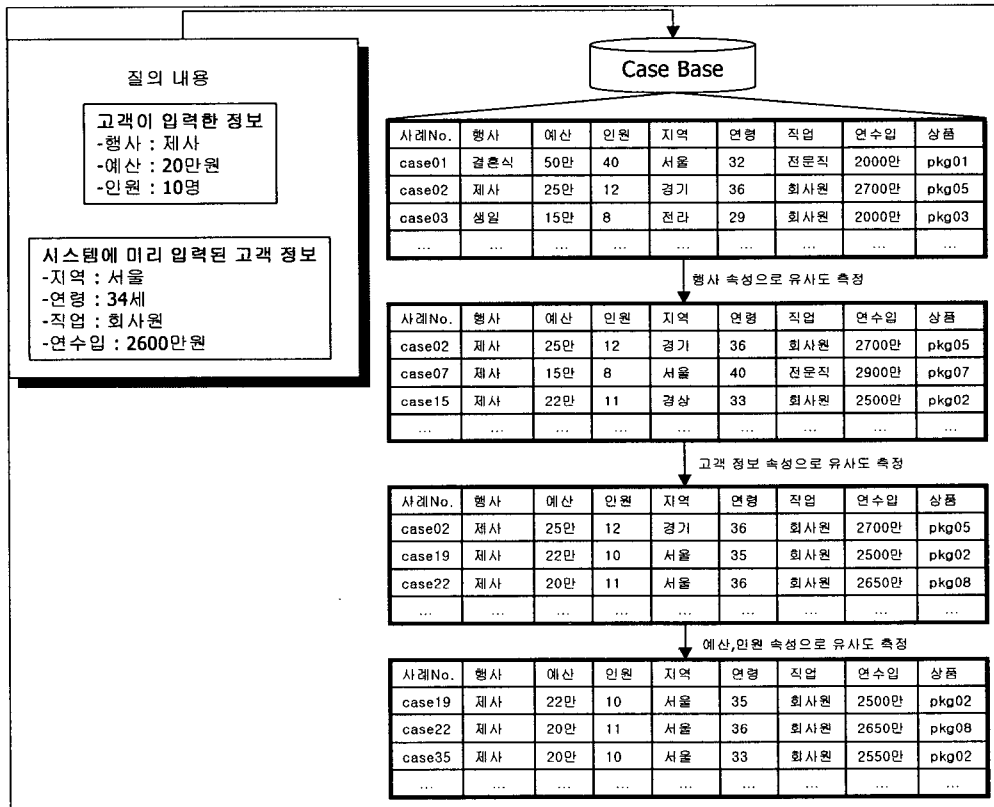
4.4.1 사례의 검색방법

사례기반시스템은 사례베이스를 효율적으로 활용하여 관련 있는 사례들을 검색할 수 있는 기능들을 제공해야 한다. 따라서 규모가 큰 사례 베이스로부터 정확한 사례가 검색될 수 있는 적절한 색인 과 검색 절차가 필요하다. 숙련된 사용자들은 데이터베이스로부터 데이터를 검색할 때처럼 사례베이스로부터 특정의 사례를 검색하기 위하여 표준 질 의어를 사용할 수 있다.

효율적인 사례의 검색을 위하여 사용되는 대표적

인 색인 방법으로 다음과 같이 3가지를 들 수 있다.

- ① 가중합계 방법(weight-sum approach) : 이 방법은 입력 사례의 특성에 대한 가중합계에 기초하여 사례를 측정하는 방법이다. 주어진 유사도 측정방법에 따라 가장 근접한 사례를 매칭하게 된다. 이 방법은 부동산 관련 데이터베이스로부터 유사한 문제들의 목록을 추출하는 것과 같이 사례의 추출요구가 특정 문제의 해결에 엄격하게 맞춰져 있지 않는 경우에 적합하다.
- ② 유도적 추출방법 : 이 방법은 체권 등급 평가문제와 같이 추출목적이 명확하게 정의된 경우 가장 적합하다. 이 방법에서는 사례들이 데이터 자체로부터 과생되는 결과에 영향을 미치는 가장 중요한 특성에 기초하여 색인이 부여된다. 이 결과 생성되는 의사결정 트리는 사례의 추출 시간을 상당히 단축시킨다. 사례의 색인에 사용되어지는 대표적인 유도기법은 ID3나 CART 등이 있다.
- ③ 지식기반 추출방법 : 이 방법은 문제 해결에 관련있는 사례의 추출을 위하여 이미 존재하는 영역지식을 적용한다. 이 방법은 규칙기반 전문가 시스템에서처럼 전문가가 사례를 분류하는데 사용되어지는 특성들을 결정하게 된다. 이 경우 시스템은 영역지식이 완전하지 않기 때문에 정확한 해결 방법을 찾기 위하여 다른 색인 방법을 활용하여 그 영역의 부분적인 모델을 자주 결합시키는 작업이 필



<그림 6> 사례 검색 과정

요하다.

4.4.2 본고에서 제안하는 사례검색방법

본 논문에서는 가중합계 방법과 유도적 추출방법을 응용하여 색인을 부여하고 사례를 추출하는 방식을 사용한다.

가중합계 방법은 사례의 속성 중에서 문제 해결에 영향을 끼치는 정도에 따라 속성별 가중치를 달리하나 본 논문에서는 문제 해결에 중요한 속성을 우선으로 검색하여 추출하는 다중 추출을 적용하였다. 즉, 유사 차립상을 검색하는 데 가장 큰 영향을 끼칠 것으로 판단되는 행사 속성에 대한 검색을 우선적으로 실시하고 이로부터 추출된 사례에 대해 다시 기타 속성에 대해 검색을 하는 방식을 취하였다. 전체적인 사례검색과정을 나타내면 <표 8>과 같다.

<표 8> 사례 검색 과정

단계	내용
1단계	사용자 요구사항 입력 내용을 속성별로 분류 IF 요구사항 불충분 THEN 사용자에게 요구사항 재입력요구 ELSE GOTO 2단계 (인원수 - 예산간의 함수를 이용하여 Non-sense 방지)
2단계	행사 속성에 대한 검색 : 행사 속성의 유사도값이 높은 사례 추출
3단계	고객 정보 속성에 대한 검색 : 고객 정보의 유사도값이 높은 사례 추출 IF 고객정보가 없는 경우 THEN 3단계 실행하지 않고 4단계 진행
4단계	예산, 인원 속성에 대한 검색 : 각 속성에 대한 유사도값이 높은 사례 추출
5단계	4단계까지 진행하여 남은 사례의 유사도를 합산하여 가장 높은 순으로 고객에게 추천

먼저 고객이 수산물 인터넷 쇼핑몰에 접속하여 HTML 폼을 통해 행사의 종류, 인원수, 예산 등을 입력하면 사례 베이스를 검색하게 된다. 이 때, 사용자의 입력이 올바른지 체크하기 위한 각종 함수가 사용된다. 예를 들어 인원-예산에 대한 함수가 사용될 수 있는데, 이 함수는 인원에 비해 지나치게 예산이 적은 경우와 같이 예산과 인원 항목의 값에 대한 입력을 체크한다.

다음으로 사례 베이스에서 선택되는 사례는 고객이 질의한 속성과 사례의 각 속성과 비교하여 각 속성별 유사도 값을 결정하고, 각 속성별 유사도에 가중합을 하여 가장 높은 값을 가지는 사례가 검색된다. 본 논문에서는 다음과 같은 순서로 각 속성의 유사도를 구하였다.

- ① 행사의 종류에 대한 유사도 측정
- ② 고객 정보에 대한 유사도 측정
- ③ 인원수·예산에 대한 유사도 측정

이 같은 순서로 결정된 이유는 행사의 중요성이 가장 크기 때문이며, 인원수 및 예산의 경우는 행

사를 중심으로 결정된 사례를 수정하여 새로운 사례로 만들기 쉽기 때문이다. 위의 과정을 도식화하면 <그림 6>과 같다. <그림 6>에서 보듯이 고객의 질의 내용을 행사→고객정보→예산·인원수 속성 순으로 검색한다. 만약 등록되지 않은 고객, 즉 비회원인 경우에는 고객 정보 검색 과정을 거치지 않고 예산·인원수 검색 과정으로 가게 된다. 사례의 검색이 완료되면 앞장에서 논의한 유사도 테이블로부터 개별 속성의 유사도를 측정 후 각 사례별 유사도를 가중합을 통해 얻어내게 된다. 사례별 유사도가 계산되면 유사도가 높은 사례 순으로 고객에게 제시된다.

V. 결론

인터넷 쇼핑몰에서 상품을 판매하기 위해서는 상당한 노력이 필요하다. 특히 수산물과 같이 신선도나 상품의 질이 일정하지 않은 상품인 경우 더욱더 인터넷을 통한 판매가 쉽지 않은 것이 현재의 모습이다. 더구나 각종 가족 행사에서 사용될 수산물을 제대로 알고 있지 않은 상황에서 기존의 구매 경험이나 시장 등에서 상인의 조언으로 수산물을 구입하는 사람인 경우 단순한 상품 사진만을 제공하는 인터넷 쇼핑몰에 신뢰가 가지 않을 것이다.

따라서 본 논문에서는 인터넷을 통한 수산물 구입을 원활하게 이루어지게 하기 위해 고객이 쓰고자 하는 목적에 알맞은 수산물 구성을 미리 알려주고 고객의 의사결정을 돕기 위해 사례기반추론을 이용한 번들상품추천시스템을 제안하였다. 이 시스템은 차립상이라고 하는 팩키지형 상품을 고객이 선택할 경우 이것을 사례라는 형태로 사례 베이스에 등록하여 고객이 원하는 상품 구성을 찾아 제시하는 형식을 취한다. 사례기반추론기법은 규칙기반추론(rule-based reasoning)에 비해 지식의 전이와 설명이 용이하며, 항상 주어질 문제에 대한 근사해를 제공한다는 장점을 가지고 있다. 이로써 고객이 직접 상품을 고르고 비교해야 하는 불편함을 해소할 것으로 기대된다. 그리고 다양한 사례를 이용함으로써 다른 고객의 구매 정보를 손쉽게 얻어 상품 구매에 반영할 수 있는 장점이 있다.

하지만 사례기반추론이 제대로 활용되기 위해서는 충분한 사례가 미리 사례 베이스에 들어 있어야 하며, 기계학습기간도 상당한 시간이 필요하다. 사례 베이스 내에 방대한 량의 사례를 체계적으로 저장하고 관리하기 위해서는 사례의 표현기법이 중요하며, 이를 통해 표현된 사례베이스로부터 질의내용에 적합한 유사사례를 검색하기 위해서는 유사도 측정의 개발이 중요하다. 본 연구에서는 번들상품추천을 위한 사례표현기법으로 속성-값 방식인 프레임(frame) 형식을 사용하고 있다. 또한 유사도 측정을 위하여 각 속성(행사, 예산, 참여자 수, 고객 거주지 등)에 대하여 유사도 테이블을 작성하고, 이들 속성들의 가중합계방법과 유도적 추출방법을 혼용하여 이용하고 있다.

본 논문에서와 같이 사례표현에 있어 프레임방식의 채택을 채택하고, 유사도 테이블의 이용하여 유사도를 계산할 경우, 관계형 데이터베이스를 통

해 쉽게 구현할 수 있는 잇점을 준다. 그러나 지능형 에이전트를 이용한 상품추천은 갖 유아기를 벗어난 상태라고 할 수 있다. 다양한 알고리즘을 활용하여 상품추천에 대한 연구가 진행되었거나 현재 진행되고 있는 상태이지만 상업적으로 활용되기에는 아직도 많은 문제가 있으며 이에 대한 지속적인 연구가 필요할 것으로 보인다. 수산물뿐만 아니라 다른 상품에도 사례기반추론이 적용되기 위해서는 무엇보다 사례와 사용자 질의의 유사정도를 비교할 수 있는 유사도 측정 방법이 각 상품별로 연구되어야 할 것이다.

참 고 문 헌

- 남기범, 이진명, "전자상거래 에이전트," 정보과학회지, 제18권 제5호, 2000. 5.
- 성백균, 김상희, 박덕원, "전자상거래를 위한 사례기반추론의 판매지원 에이전트," 한국정보처리학회 논문지, 제7권, 제5호, 2000.
- 이강재, "유사 정보 추출을 위한 관계 데이터베이스 관리 시스템의 사례 기반 추론 메카니즘," 동국대학교 대학원 박사학위 논문, 1996.
- 이상기, 이재규, "인터넷상의 Cyber 판매전문가시스템 : CYBERSES," KESS'97 추계학술대회, 1977.
- 이재호, "에이전트 시스템의 연구 및 개발 동향", 정보과학회지, 제18권 제5호, 2000. 5
- 이지은, "사례 기반 시스템의 이론적 고찰 및 구현 현황," 경영학연구 제6호, 1998. 12.
- 이창목, 양보석, "회전기계의 이상진동진단을 위한 사례기반 추론 시스템의 개발," 한국소음진동공학회. 2000.
- 최중민, "에이전트기술 및 전자상거래," 한국공학기술학회지, 7권 2호, 2000.
- 황하진, "사례기반 추론을 이용한 컴퓨터지원설계 시스템의 개발," 정보시스템연구 제7권 제2호, 1998. 12.
- Aamodt, A. and E. Plaza, "Case-based Reasoning : Foundational Issues, Methodological Variations, and System Approaches," *Artificial Intelligence Communications*, Vol. 7, No. 1, 1994.
- Guttman, R., A. Moukas, and P. Maes, "Agent-mediated Electronic Commerce : A Survey," *Knowledge Engineering Review*, June 1998.
- Kim, S. H., J. E. Jung, and G. S. Jo, "Recommender System with Configuration in Electronic Commerce," *Proceeding of International Conference on Electronic Commerce 2000*, ICEC.
- Kim, S. H., S. W. Shin, and J. H. Kim, "Personalized Recommendations for Retailing in Internet Commerce: A Multistrategy Filtering Approach," *Proceeding of International Conference on Electronic Commerce 2000*, ICEC.
- Klusch, M., "Information Agent Technology for the Internet: A Survey," *Data & Knowledge Engineering*, No. 36, 2001.
- Kolodner, J., *Case-based Reasoning*, Morgan Kaufmann, 1993.
- Kolodner, J., "Retrieving Events from Case Memory : A Parallel Implementation," *Proceedings from the Case-based Reasoning Workshop*, DARPA, Clearwater Beach, 1988
- Rahmer, J., and A. Voß, "Case-Based Reasoning in the Configuration of Telecooperation Systems," *AAAI 1996 Fall Symposium Workshop*, 1996.
- Roger S., D. Leake, "Creativity and Learning in a Case-based Explainer," *Artificial Intelligence*, Vol. 40, No. 1-3, 1989.
- Schank R., *Dynamic Memory: A Theory of Reminding and Learning Incomputers and Peoples*, Cambridge University Press. 1982.
- Schrooten, R., "Agent-based Electronic Consumer Catalog," *The First International Conference and Exhibition on the Practical Application of Intelligent Agent and Multi-Agent Technology '96*, 1996.
- Slade, S., "Case-based Reasoning : A Research Paradigm," *AI Magazine*, Spring 1991.
- Vollrath I., W. Wilke, and R. Bergmann "Intelligent Sales Support on the Web," *IEEE Internet Computing*, Vol. 2, No. 4, 1998.
- Wilke W., R. Bergmann, "Techniques and Knowledge Used for Adaptation During Case-based Problem Solving," *On the 11th International conference on Industrial and Engeer Application of Artificial Intelligence and Expert Systems*, 1998.