

퍼지연상기억장치에 기반한 협력 추천 방법

(A Collaborative Recommendation Method based on Fuzzy Associative Memory)

이 동 섭[†] 고 일 주^{**} 김 계 영^{***}
 (Dong Seop Lee) (Il Ju Ko) (Gye Young Kim)

요약 최근 인터넷의 발전으로 정보의 접근이 용이할 뿐 아니라 그 양 또한 기하급수적으로 증가하고 있다. 정보의 홍수 속에서 원하는 정보만을 자동으로 추출할 수 있는 기술은 정보검색에 소요되는 시간과 노력을 절약할 수 있는 매우 중요한 연구이다.

본 논문에서는 관심 범위가 유사한 사용자에게 양질의 정보를 자동으로 추천하기 위하여 협력적 여과 방법에 관하여 제안한다. 제안하는 방법의 기본적인 배경은 사용자는 선택항목의 선호도를 입력하고, 여과 장치는 이 선호도에 근거하여 추천집합을 자동으로 생성하는 것이다. 선호도로부터 추천집합을 추천하기 위하여 본 논문에서 퍼지연상기억장치에 기반한 방법을 제안한다. 제안된 방법은 웹 서버상에서 기술문서 특히, 정보기술문서를 검색하는 분야에 대하여 구현하였으며 그 결과를 보인다.

키워드 : 협력 추천, 퍼지연상기억장치, 정보검색

Abstract At recent, people can easily access to information by Internet to be rapidly evolving. And also, the amount is rapidly increasing. So the techniques, to automatically extract the required information are very important to reduce the time and the effort for retrieving information.

In this paper, we describe a collaborative filtering system for automatically recommending high-quality information to users with similar interests on arbitrarily narrow information domains. It asks a user to rate a *gauge set* of items. It then evaluates the user's rates and suggests a *recommendation set* of items. We interpret the process of evaluation as an inference mechanism that maps a gauge set to a recommendation set. We accomplish the mapping with FAM (Fuzzy Associative Memory). We implemented the suggested system in a Web server and tested its performance in the domain of retrieval of technical papers, especially in the field of information technologies. The experimental results show that it may provide reliable recommendations.

Key words : Collaborative Recommendation, Fuzzy Associative Memory, Information Retrieval

1. 서론

인터넷의 발전으로 원하는 정보에 언제 어디서나 접근할 수 있을 뿐 아니라 그 양 또한 매우 풍부하며 증가의 속도는 기하급수적이라 할 수 있다. 통신망 상에 분산되어 있는 정보의 홍수 속에서 사용자가 원하는 정보를 포함하는 문서를 찾기 위하여 많은 노력과 인내가

요구된다. 따라서, 필요한 정보를 효과적으로 검색할 수 있는 방법에 관한 연구는 정보검색에 소요되는 노력과 시간을 절약하므로 대단히 중요하다.

대표적인 방법에는 내용에 기반한(content-based) 방법, 연관규칙에 의한 방법, 협력적 여과(collaborative filtering)에 의한 방법 등이 있다. 내용에 기반한 방법 [1]은 새로운 항목이 입력되었을 때 사용자의 특성을 고려하지 않고 과거에 좋아했던 항목의 내용을 분석하여 유사한 항목을 추천하는 기술이다. 이 방법의 문제는 사용자의 특성을 고려하지 않는 점과 입력항목과 추천항목 사이의 연관성을 고려하지 않는 점이다. 연관규칙에 의한 방법 [2]은 항목간의 연관성을 고려하여 관련 항목을 추천하는 기술이다. 이 방법은 입력항목과 추천항목 사이의 연관성은 고려하지만 사용자의 특성을 고려하지

· 본 연구는 첨단정보기술연구센터(AiTrc)를 통하여 과학재단의 지원을 받았다

† 비 회 원 : (주)클릭큐

leeds@clickq.com

** 비 회 원 : 송실대학교 미디어학부 교수

andy@ssu.ac.kr

*** 종신회원 : 송실대학교 컴퓨터학부 교수

gykim@computing.ssu.ac.kr

논문접수 : 2001년 8월 9일

심사완료 : 2004년 5월 29일

않는 문제가 있다. 그리고 협력적 여과에 의한 방법[3]은 사용자의 특성을 고려하지만 입력항목과 추천 항목 간의 연관성은 고려하지 않는다.

효과적인 정보검색을 위하여 사용하는 추천시스템에 관한 연구에 있어서 중요한 고려 사항으로 첫 번째는 사용자 그룹에 대한 것이고, 두 번째는 입력항목과 추천 항목 사이의 연관성이다. 사용자를 그룹화하지 않을 경우 추천시스템이 너무 범용적이므로 추천결과를 신뢰할 수 없을 뿐 아니라 추천시스템의 규모가 증가하여 처리 시간이 많이 소요되며 사용하는데 불편한 문제점이 있다. 이와 같은 문제들을 해결하는 방법으로 알려지지 않은 새로운 사용자의 선호도를 다른 사용자들의 알려진 선호도로부터 예측하여 추천할 수 있는 협력적 여과에 의한 추천 방법이 있으며 최근에 이와 관련된 많은 연구가 수행되고 있다[4].

협력적 여과시스템은 능동적인 여과시스템과 자동적 여과시스템으로 구분할 수 있다[5]. 능동적 여과시스템에서 사용자는 유사한 흥미를 가지는 다른 사용자를 능동적으로 식별한 후 서로 협력한다. 따라서, 이 방법은 반드시 협력 그룹을 정의하여야 하므로 사용자에게 요구하는 정보의 종류 및 수가 많다. 한편, 자동적 여과시스템에서는 협력그룹을 자동적으로 생성함으로써 수동적 협력에 의한 여과시스템에서 요구되는 정보를 경감한다. 그러나 '유사한 흥미를 가지는 그룹을 어떻게 형성하고 이 그룹들의 흥미를 어떻게 표현할 것인가' 하는 문제를 해결하여야 한다.

자동적 협력에 의한 여과시스템을 보다 세분하면 사용자가 선호하는 항목(user-selected item) 만을 선택하는 시스템과 전역 항목(universal item)에 대한 사용자의 선호도를 입력 받아 사용하는 시스템으로 분류할 수 있다[6]. 첫 번째 방법은 많은 항목들이 값을 가지지 않는 산포행렬(sparse matrix)을 생성한다. 전역항목을 사용하는 시스템은 모든 사용자에게 동일한 항목들의 집합을 나타낸 후 모든 항목들에 대하여 즉각적인 피드백을 얻는 방법이다.

본 논문에서는 적은 정보영역에서 임의의 유사한 흥미를 가지는 사용자들에게 양질의 정보를 추천할 수 있는 협력적 여과시스템에 관하여 기술한다. 제안하는 방법은 먼저 사용자에게 항목의 측정집합(gauge set)에 대한 선호도를 요구한 후, 이 선호도를 평가하여 항목들의 추천집합(recommendation set)을 제시하는 방법으로 기본적인 그 운영원리는 Eigentaste 시스템[6]과 같다. 그러나, 본 논문에서는 측정집합과 추천집합 사이의 사상(mapping)을 위한 추론 방법으로 퍼지연상기억장치(FAM : Fuzzy Associative Memory)[7]를 사용하여 보다 양질의 추천집합을 생성하는 방법을 제안한다.

FAM은 하나의 퍼지집합을 다른 퍼지집합으로 사상하는 기본적인 골격을 제공한다[8]. 이러한 사상은 입력 퍼지집합(측정집합)과 출력 퍼지집합(추천집합) 사이의 연관관계를 나타내는 퍼지규칙의 집합으로 볼 수 있다. 또한 FAM은 입력과 출력 사이의 연관성을 설립하기 위한 Hebbian 형태의 학습 방법을 제공[9]하는데, 이 학습 방법은 매우 단순하고 적은 계산시간을 소요하는 장점이 있다.

협력에 의한 여과시스템에서 또 다른 하나의 문제는 유사한 흥미를 가지는 사용자 그룹을 형성하는 방법이다. 이 문제는 새로운 사용자의 알려지지 않은 선호도를 예측하기 위하여 사용자 그룹의 알려진 선호도를 사용하기 위한 것으로 군집화(clustering)의 전형적인 형태라 할 수 있다. 그러나, 본 논문에서는 군집화 과정이 FAM의 노드들 사이의 연결가중치(connection weight)에 내포됨으로써 명시적인 군집화를 사용하지 않는다. 연결가중치는 학습단계에서 사용자의 감독에 의하여 이루어진다.

본 논문에서 제안하는 협력에 의한 여과시스템은 학습부와 추론부로 구성된다. 학습부에서는 퍼지집합을 형성하기 위하여 입력과 출력의 쌍으로 만들어진 학습자료를 분석하는 단계로써 오프-라인으로 이루어진다. 학습부의 입력자료는 측정집합에 있는 모든 항목에 대한 선호도의 값들의 집합이고, 출력자료는 추천집합에 대한 선호도의 값들의 집합이다. 추론부에서는 새로운 사용자에서 측정집합의 항목들을 제시한 후 사용자가 입력한 선호도 값들을 처리하여 학습단계에서 형성된 퍼지규칙에 따라 추천집합의 선호도 값들을 생성하는 단계로써 온-라인으로 이루어진다. 본 논문에서 제안하는 협력에 의한 여과시스템의 구조는 그림 1과 같다.

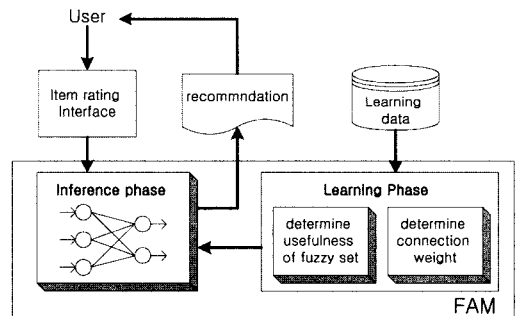


그림 1 전체적인 시스템 구성도

본 논문의 구성은 다음과 같다. 2장에서는 FAM에 기반한 추론시스템에 관하여 설명하고, 3장에서는 관찰에 근거한 학습모델에 관하여 기술한다. 4장에서 실험 결과를 보인 후 5장에서는 결론을 기술한다.

2. FAM에 기반한 추론

FAM은 연상기억장치(Associative Memory)와 퍼지 논리(Fuzzy Logic)을 결합한 것으로 두 퍼지집합 사이의 관계를 정의할 수 있다. FAM의 기본적인 구조는 그림 2와 같다. 즉, 관계를 정의해야 하는 두 퍼지집합 (A_i, B_j)와 이들 사이를 연결하는 가중치들 w_{ij} 로 구성된다. 그림 2에서 A_i 는 입력층에 있는 노드들을 나타내고 B_j 는 출력층에서의 노드들을 나타낸다. 입력노드 A_i 와 출력노드 B_j 사이를 연결하는 가중치 w_{ij} 가 큰 값을 가지면 A_i 와 B_j 는 서로 연관성이 크다는 것을 의미한다.

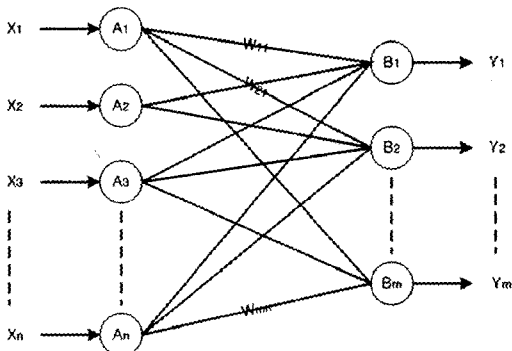


그림 2 Fuzzy associative memory

본 논문에서는 협력에 의한 여과시스템을 위하여 입력층에 있는 노드들은 측정항목에 대한 퍼지집합에 대응되고 출력층에 있는 노드들은 추천항목에 대한 퍼지집합이 된다. 연결된 가중치는 측정항목의 퍼지집합과 추천항목의 퍼지집합 사이의 연관되는 정도를 나타낸다. FAM에서 측정항목과 추천항목 사이의 관계를 설정하는 예는 다음과 같다. 즉, 측정항목 "high preference of computer"는 추천항목 "high preference of internet" 사이의 연결 가중치는 0.8과 같다. 여기서 가장 중요한 문제는 입력노드들과 출력노드들 사이를 연결하는 가중치를 생성하는 방법이다. 이 문제를 해결하기 위한 여러 가지 방법들이 있지만, 본 논문에서는 단순하고 적은 계산시간을 소요하는 장점을 가지는 Hebbian 학습[8] 방법을 FAM 학습에 이용한다. FAM을 학습시키는 방법에 대해서는 다음 절에서 상세히 설명한다. 이 절에서는 입력노드들과 출력노드들 사이의 가중치가 정의되어 있다는 전제하에서 결과를 추론하는 방법에 관하여 상세히 기술한다.

연결가중치가 결정되어 있는 FAM에서 결과를 추론하는 방법으로, 본 논문에서는 다음과 같이 MAX-MIN 연산을 통하여 추론한다[8]. 먼저 FAM의 구성요소가 가지는 값을 다음과 같이 가정하자. 입력노드 A_i 가 가

질 수 있는 값의 범위는 항목 x_i 가 가질 수 있는 값의 범위이고, 항목 x_i 의 입력값은 퍼지소속함수 μ_{A_i} 범위의 값으로 사상된다. 사상된 결과값을 $a_i = \mu_{A_i}(x_i)$ 라 한다. 출력노드 B_j 가 가질 수 있는 값의 범위는 항목 y_j 가 가질 수 있는 값의 범위이고, 항목 y_j 의 출력은 퍼지소속함수 μ_{B_j} 범위의 값으로 사상된다. 사상된 결과값을 $b_j = \mu_{B_j}(y_j)$ 라 한다. 이때 FAM에서 연산은 식 1과 같이 수행된다.

$$A \circ W = B \tag{1}$$

식 (1)에서 기호 \circ 는 MAX-MIN 합성 연산을 나타낸다. 즉, 출력값 b_j 는 식 (2)와 같이 a_i 와 W 의 j 번째 열과의 퍼지곱 연산을 수행하여 얻어진다.

$$b_j = \max_i \{ \min(a_i, w_{ij}) \} \tag{2}$$

이와 같이 FAM은 입력된 퍼지집합과 출력하는 퍼지집합을 연결하는 가중치를 통해 상호 연결성을 정의할 수 있으므로, 측정항목과 추천항목 사이의 관계로부터 유사한 성향을 갖는 사용자들을 그룹화 할 수 있다. 이를 위하여 본 논문에서는 퍼지규칙의 조건부를 측정항목으로, 결론부를 추천항목으로 표현한다. 그리고, 규칙의 신뢰 정도는 연결 가중치로 표현한다.

그림 3에서는 본 논문에서 제안하는 추론모델의 구조를 보인 것이다. 추론모델은 4단계의 층(layer) 즉, 입력층(input layer), 퍼지화층(fuzzification layer), 선행층(antecedent layer), 그리고 결론층(consequent layer)으로 구성된다. 입력층은 n 개의 측정항목 $x_i (i=1, \dots, n)$ 의 노드로, 결론층에 m 개의 추천항목 $y_j (j=1, \dots, m)$ 의 노드로 구성된다. 퍼지화층의 노드 개수는 각각의 입력노드가 가지는 퍼지집합 P_i 의 합 즉, $\sum_{i=1}^n P_i$ 이다. 퍼지화층의 출력은 입력된 선택호의 정도를 퍼지소속함수의 값으로 변환한 값이다. 결론에 선행하는 선행층은 퍼지규칙의 조건들을 포함하고 있으며, 조건들 사이의 조합은 논리곱 연산에 의하여 이루어진다. 선행층의 입력은 퍼지화층의 모든 출력이며, 이들 사이에는 모든 조합이 가능하여야 하므로 노드의 수는 $N = \prod_{i=1}^n P_i$ 이다. 따라서, 각 노드는 n 개의 입력연결을 가지며 각 연결은 가중치를 가진다. 이 가중치는 연관된 퍼지집합의 유용도를 나타낸다. 즉, 퍼지화층의 어떤 노드로부터 나온 연결이 높은 가중치를 가지면, 이 노드에 포함된 퍼지집합은 요구되는 추천항목을 추론하는데 매우 유용한 퍼지집합을 포함하고 있는 것을 의미한다. 가중치를 결정하는 방법에 관해서는 다음 절에서 상세히 기술한다. 선행층의 각 노드는 가중화된 입력값들을 비교하여 최소값을 선택한다. 즉, 관련 있는 퍼지집합들의 실제값의 논리곱을 <

행하여 얻어진다. 결론층은 퍼지규칙의 결론부에 해당되며, m개의 소속함수를 가진다. 결론층은 퍼지규칙에 의하여 유도되는 결론부들을 포함하고 있으며, m개의 소속함수로 구성된다. 각 소속함수는 추천항목 각각의 선호도를 결정한다.

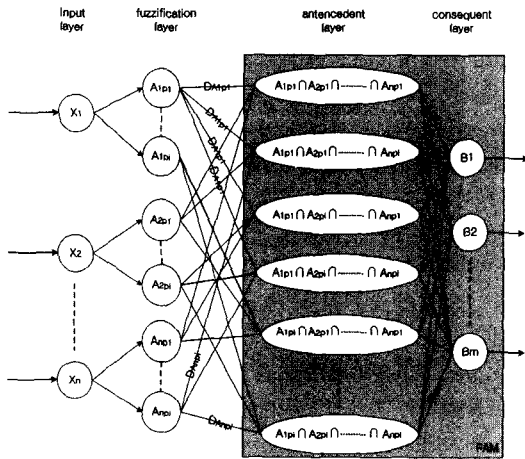


그림 3 FAM의 추론 시스템

본 논문에서의 FAM은 선행층과 결론층 사이가 완전 연결(full connection)된 구조이다. 즉, 조건부와 결론부 사이가 완전연결(full connection)된 구조이다. 모든 연결은 서로 다른 가중치를 가지고 있으며, 이 가중치는 연결의 신뢰성을 나타낸다. 결론을 추론하기 위한 연산은 기본적으로 최대-최소 조합규칙(max-min composition rule)을 사용한다[10]. 따라서, N개의 선행노드 A_1, \dots, A_N 가 가중치 w_{ij} 로 j번째 결론노드 B_j 에 연결되어 있을 때, 역퍼지화(defuzzification)을 통하여 B_j 로부터 출력되는 y_j 값은 식 (3)과 같다.

$$y_j = \mu^{-1}_{B_j}(\max_{1 \leq i \leq N}(\min(w_{ij}, output(A_i)))) \quad (3)$$

식 3에서 $\mu_{B_j}(y)$ 는 결론층에 있는 j번째 노드의 소속함수를 나타내고, $output(A_i)$ 는 선행층에 있는 i번째 노드의 출력 결과를 나타낸다. 즉, 결론부의 결과를 역퍼지화하여 얻어지는 값은 대응하는 항목의 추천 정도를 나타낸다.

그림 4에서는 FAM을 통한 추론과정을 보인 것이다. 두 개의 측정항목의 값을 x_1 와 x_2 라 할 때, 이들은 두 개의 퍼지항 A_{13} 와 A_{21} 와 관련되며 각각의 소속함수는 $\mu_{A_{13}}(x_1)$ 와 $\mu_{A_{21}}(x_2)$ 이다. 그리고 선행층과 연결될 때 유용도를 나타내는 가중치는 각각 $D_{A_{13}}$ 와 $D_{A_{21}}$ 이다. 결론층에 있는 노드의 소속함수를 $M_B(y)$ 라 하고 선행층과 연결되는 가중치는 w이다. 이와 같은 조건에서 결론

노드의 출력은 식 (4)와 같다.

$$y^* = \mu^{-1}_B[\min(\min(\mu_{A_{13}}(x_1) \cdot D_{A_{13}}, \mu_{A_{21}}(x_2) \cdot D_{A_{21}}), w)] \quad (4)$$

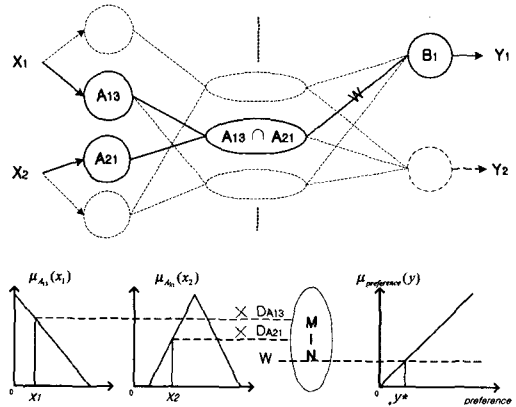


그림 4 추론의 예

3. 관찰에 의한 FAM 학습

앞 절에서 추론방법에 관하여 설명할 때, 노드의 소속함수와 이들 사이를 연결하는 가중치는 이미 결정되어 있다는 가정 하에서 기술했다. 이 절에서는 입력과 출력으로 구성된 학습자료로부터 필요한 정보를 유도하는 학습 방법에 관하여 기술한다. 학습자료는 사용자로부터 얻어지며 입력은 측정항목에 출력은 추천항목에 각각 대응한다.

FAM을 이용하여 추론시스템을 구축하는데 있어서 가장 중요하며 반드시 해결되어야 하는 문제는 각 항목에 대한 퍼지집합의 수와 이에 대응하는 소속함수를 결정하는 것이다. 이 문제는 입력과 출력의 각 항목이 가지는 값의 전체범위를 부 영역으로 나누고 각 영역에 소속함수를 연관시키는 작업이라 할 수 있으며, 효율성을 위하여 이를 체계적으로 정의할 수 있는 방법이 요구된다. 본 논문에서는 자료의 분포를 조사하여 조정하는 접근 방법을 사용한다[9]. 이 방법은 단순하여 적은 계산시간을 소요할 뿐 아니라 유용도를 측정하는데 충분한 장점이 있다.

소속함수의 기본적인 구조는 그림 5와 같다. 즉, 사용자는 어떤 하나의 항목에 대하여 자신의 선호도를 0에서 1 사이의 값으로 표현한다. 사용자에 의하여 표현된 값이 1에 가까우면 높은 선호도를, 0에 가까우면 낮은 선호도를 가지는 것을 의미한다. 측정항목에 대하여 값의 전체범위를 세 부분으로 나눈 다음 각 부분에 "low", "medium", "high"의 퍼지집합을 할당한다. 소속함수는

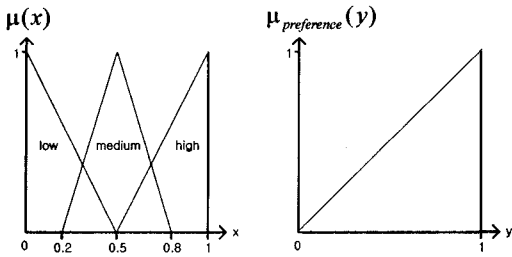


그림 5 소속 함수의 구조

식 (5)와 같다. 그리고 추천항목에 대해서는 값의 전체 범위를 “preference” 퍼지집합 하나를 사용하며 소속함수는 단수증가 즉, $\mu_{preference}(y) = y$ 이다. 이 방법은 추천모델에서 결론층의 적은 노드 수를 사용하는 장점이 있다. 만약, 특정항 추천항목에 대하여 보다 정확한 추천 값을 원하면 해당 추천항목을 위하여 여러 개의 퍼지집합을 사용할 수 있다.

$$\begin{aligned} \mu_{low}(x) &= \begin{cases} 1-2x, & 0 \leq x < 0.5 \\ 0, & 0.5 \leq x \leq 1 \end{cases} \\ \mu_{medium}(x) &= \begin{cases} (10x-2)/3, & 0.2 \leq x < 0.5 \\ (-10x+8)/3, & 0.5 \leq x \leq 0.8 \end{cases} \\ \mu_{high}(x) &= \begin{cases} 0, & 0 \leq x < 0.5 \\ 2x-1, & 0.5 \leq x \leq 1 \end{cases} \end{aligned} \quad (5)$$

다음으로 해결하여 하는 문제는 입력되는 퍼지집합의 유용도를 결정하는 것이다. 즉, ‘입력된 퍼지집합이 추천항목을 어떻게 구분할 것인가’ 하는 문제이다. 학습자료로 입력된 퍼지집합에 대응하는 추천항목의 선호도를 나타내는 값의 분포도가 평평한 형태이면, 이 퍼지집합은 특정한 항목을 추천하기에 유용하지 않다. 반면, 이 분포도가 뾰족한 형태이면, 이 퍼지집합은 유용하다. 이와 같은 기준은 구별이 분명한 퍼지집합을 강화하는 것과 동일하다. 이것은 불확실성을 의미하므로 입력자료가 가지는 불확실성을 측정하는 유용한 방법으로는 엔트로피(entropy)를 측정하는 방법이 있다. 엔트로피는 입력된 자료의 특성을 잘 나타내므로[10], 구분력 평가는 엔트로피(entropy)를 측정함에 의하여 이루어진다. 즉, 식 6과 같은 함수를 사용하여 입력된 퍼지집합의 유용도를 계산한다.

$$\begin{aligned} D_{A_i} &= 1 + \sum_{j=1}^M P_{A_i}(j) \cdot \ln(P_{A_i}(j)) \quad (6) \\ P_{A_i}(j) &= \frac{\sum_{i=1}^K \mu_{A_i}(x_i) \cdot r_{B_j}(x_i)}{\sum_{i=1}^M \sum_{i=1}^K \mu_{A_i}(x_i) \cdot r_{B_j}(x_i)} \end{aligned}$$

x_i : i th input-output learning data,
 $\mu_{A_i}(x_i)$: a fit value of x_i to an input fuzzy set A_i ,
 $r_{B_j}(x_i)$: the rate of x_i on a recommendation item B_j ,
 M : number of recommendation items

K : total number of learning data

$P_{A_i}(j)$: probability density of rates on the j th recommendation item,

본 논문에서 사용하는 추론모델은 입력된 퍼지집합과 출력하는 퍼지집합 사이의 관련성을 표현하기 위해 사전에 값이 결정되어 있는 관계행렬(correlation matrix)을 요구한다. 이 관계행렬의 값을 결정하기 위하여 본 논문에서는 Hebbian 학습을 사용한다. Hebbian 학습은 비 감독에 의한 방법으로 입력과 출력 모두가 활성화되었을 때 이들을 연결하는 가중치를 증가하는 방법이다. 따라서, 습관이나 학습을 반복하면 할수록 결과는 점점 더 명확해진다[11]. 관계행렬의 원소들은 입력과 출력 사이의 가중치를 나타낸다. 이 가중치는 입력값과 출력값이 소속함수의 값으로 사상되므로 최소연산과 최대연산을 수행하여 얻어진다. 즉, $a_i(n)$ 는 학습 자료의 입력값이고, $b_j(n)$ 는 출력값이라 하면, 연결가중치의 수정은 식 (7)과 같이 이루어진다.

$$w_{ij}(n) = (1 - \eta) \cdot w_{ij}(n-1) \oplus \eta \cdot (a_i(n) \otimes b_j(n)) \quad (7)$$

η : a positive learning rate that is not greater than 1

\otimes : a minimum operator

\oplus : a maximum operator

퍼지규칙의 조건부에 대응하는 선행층은 측정집합에 있는 모든 항목에 대한 선호도의 값들의 집합으로 $N = \prod_{i=1}^M P_i$ 개의 노드로 구성되고, 결론층은 추천집합에 있는 항목에 대한 선호도의 값들의 집합으로 m 개의 노드로 구성된다. 이때 선행층의 출력은 $X=(a_1, a_2, \dots, a_n)$ 이고 결론층의 출력은 $Y=(b_1, b_2, \dots, b_m)$ 이다. 예를 들면, 측정집합이 논문 10편이고 추천집합이 논문 20편인 경우에 $N=10$ 과 $m=20$ 이 되어 $X=(a_1, a_2, \dots, a_{10})$ 와 $Y=(b_1, b_2, \dots, b_{20})$ 이 된다. 선행층의 출력과 결론층의 출력이 결정되면 관계행렬은 식 (8)과 같은 반복 학습에 의하여 얻어진다.

$$\begin{aligned} W(n) &= (1 - \eta) \cdot W(n-1) \oplus \eta \cdot W(n) \quad (8) \\ &= (1 - \eta) \cdot W(n-1) \oplus \eta \cdot (X^T \otimes Y) \end{aligned}$$

본 논문에서의 추론모델은 선행층과 결론층 사이에 완전연결을 허용하기 때문에 많은 규칙들을 생성하고 계산시간도 많이 소요한다. 그러나, 연결가중치의 값은 관련된 퍼지규칙의 신뢰성을 나타내는 척도로 고려할 수 있으므로 가중치 값이 적은 규칙은 유용한 결론을 추론하는데 중요하지 않음을 의미한다. 따라서, 구현할 때는 사전에 결정된 임계값 보다 적은 가중치를 가지는 연결은 제거함으로써 퍼지규칙의 수를 줄일 수 있다. 계산시간을 절약할 수 있으며 또한 본 논문에서 사용하는 Hebbian 학습은 적은 계산시간을 소요하는 장점이 있다.

4. 실험 결과

본 논문에서 제안한 추천모델의 성능을 평가하기 위하여 웹 서버상에서 기술논문 특히, 정보기술논문을 검색하는 분야에 적용하였다. 학습과 추천모델 구현은 visual C++를, 인터넷에서 사용자의 선호도를 취득하기 위한 인터페이스는 java를 사용하였다. 인터페이스는 검색한 결과논문을 나타내고, 그 논문에 대한 선호도는 마우스를 사용하여 값을 선택하여 입력할 수 있도록 구성했다. 학습단계에서 사용자는 제시된 논문에 대하여 선호도를 입력하여야 한다. 이 입력된 값들은 소속함수와 FAM의 연결가중치를 생성하는데 사용된다.

성능평가는 퍼지소속함수의 유용도로부터 유도된 선호도의 정확도에 미치는 영향과 FAM에서 노드들 사이의 연결 수로부터 유도된 선호도에 미치는 영향을 측정함으로써 이루어진다. 실험에서 식 7의 학습을 η 은 0.1로 설정하였고, 초기 가중치의 값들은 0.0으로 설정하였다. 사용자의 선호도를 입력 받기 위한 입력층의 노드는 10개로 하였으며, 출력층의 노드를 20개로 함으로써 20개의 추천집합을 갖도록 하였다. 결과적으로, 사용자는 시스템에서 제시하는 10개 범주의 대표적인 논문의 선호도를 지정하게 되고, 시스템은 유사한 흥미를 가지는 협력그룹에서 추천하는 검색결과를 사용자에게 제시한다.

실험을 위해, 100명의 사용자가 학습단계에 참여하였으며 시험단계에는 다른 사용자 100명이 참여하였다. 측정집합은 입력층의 노드를 10개로 설계하였으므로 10편, 출력층의 노드를 20개로 하였으므로 추천집합에는 20편의 정보기술논문을 사용하였다. 선택된 논문들은 각각의 전문분야에 중복되지 않도록 키워드의 중복을 피하고, 각 정보기술분야의 대표적인 논문을 선정하였다.

학습단계에서는 각 100명의 사용자에게 10편의 측정집합에 속한 논문과 20편의 추천집합에 속한 논문을 제시하고 논문들에 대한 선호도를 각각 입력 받아 학습을 수행하였다. 선호도는 0에서 1값을 갖게 하였다. 학습이 완료되면 다음 단계인 시험단계에서는, 다른 100명의 사용자에게 측정집합에 속한 논문 10편을 제시하고 각 논문의 선호도를 입력 받아 전 단계에서 학습된 FAM을 이용하여 추천집합을 추천하게 된다. 추천집합은 사용자에게 다시 제시되어 추천된 결과가 만족하는지를 검증하게 된다. 시험단계에서 추출된 측정집합과 추천집합과의 선호도를 이용하여 다시 한번 FAM을 학습함으로써 학습단계에서 학습된 FAM을 시험단계에서 학습의 결과를 강화시킨다.

성능의 정확도를 평가하기 위하여 본 논문에서는 식 (9)와 같은 OMAE(Overall Mean Absolute Error) 평가척도를 사용했다[4].

$$OMA E = \frac{1}{M} \times \sum_{j=0}^M MAE(j) \tag{9}$$

$$MAE(j) = \frac{1}{N} \times \sum_{i=0}^N |r_{ij} - P_{ij}|$$

P_{ij} : prediction for how user i will rate item j

r_{ij} : actual rate given by user i for item j

M : number of items user i has rated

N : total number of users involved in the test

식 9에서 $MAE(j)$ 는 j 번째 추천항목에 대하여 전체 사용자들의 평균오차를 나타내고, $OMA E$ 는 모든 추천항목과 모든 사용자들에 대한 평균오차를 나타낸다.

표 1에서는 본 논문에서 제안한 방법을 유용도에 대한 척도를 사용한 경우와 사용하지 않은 경우에 대한 성능을 평가한 결과를 보인 것이다. 표 1의 실험결과 유용도를 사용한 경우가 평균 0.07정도의 적은 $OMA E$ 를 가지는 것을 알 수 있다.

표 1 각각의 recommendation item의 $OMA E$

Recommendation item	OMA E without usefulness measure	OMA E with usefulness measure	Recommendation item	OMA E without usefulness measure	OMA E with usefulness measure
Paper01	0.234	0.215	Paper11	0.154	0.142
Paper02	0.164	0.156	Paper12	0.189	0.185
Paper03	0.209	0.201	Paper13	0.220	0.215
Paper04	0.186	0.178	Paper14	0.237	0.231
Paper05	0.160	0.149	Paper15	0.205	0.196
Paper06	0.207	0.203	Paper16	0.156	0.154
Paper07	0.242	0.234	Paper17	0.151	0.147
Paper08	0.219	0.216	Paper18	0.214	0.206
Paper09	0.182	0.179	Paper19	0.178	0.171
Paper10	0.170	0.168	Paper20	0.203	0.194
			Total	3.88	3.74
			Average	0.194	0.187

다음은 노드들 사이를 연결하는 연결의 수가 성능에 미치는 영향에 관하여 기술한다. 연결선의 수를 줄이고 시스템을 단순화하기 위한 이 평가는 임계값을 사용하여 이루어진다. 즉, 임계값 Th 보다 적은 가중치를 가지는 연결선을 삭제한 후 결과 OMAE를 측정한다. 실험에 의하여 임계값에 따라 제거되는 연결선의 수는 0.1인 경우에는 25% 정도 그리고 0.2인 경우에는 35% 정도이다. 표 2에서는 임계값 변화에 따른 성능 그리고 POP [12]과 Eigentaste[6]의 성능을 나타낸 것이다. 표 3에 따르면 임계값이 0.1일 때의 성능이 POP[12]와 Eigentaste[6]에 비하여 좋음을 알 수 있다. 또한, 임계값이 증가함에 따라 성능이 급격히 나빠짐을 알 수 있다.

표 2 임계값 변화에 따른 성능 및 비교

System	OMAE
Our system with $Th = 0.0$	0.187
Our system with $Th = 0.1$	0.192
Our system with $Th = 0.2$	0.245
POP[12]	0.302
Eigentaste[9]	0.284

5. 결론

본 논문에서는 퍼지연상기억장치를 이용한 협력 추천 방법을 제안하였다. 측정집합과 추천집합 사이의 사상을 위한 추천방법으로 퍼지연상기억장치를 사용하는 것이 보다 양질의 추천집합을 생성하는 것을 실험을 통하여 확인하였다. 그러나 퍼지연상기억장치의 특성상 학습할 데이터의 양의 증가하면 할수록 학습 속도가 현저하게 증가하는 현상을 피할 수 없다. 즉, 사용자의 관심분야인 측정집합의 수가 증가하면서 발생하는 학습 속도의 문제는 피할 수 없다. 이러한 단점은 여러 분야가 복잡하게 존재하는 상황에 적용한다면 커다란 약점으로 작용하게 될 수 있다. 하지만 제안된 모델은 독립된 사용자가 자신의 관심분야만으로 한정된 영역을 대상으로 하였기 때문에, 실제로 상용 검색 시스템에 적용하였을 경우라든 학습속도나 성능이 문제가 되지 않을 것이다. 향후 과제로는 퍼지연상기억장치가 학습 데이터의 양이 많아지더라도 학습속도나 성능에 많은 영향을 받지 않는 방법에 대한 연구가 있다.

참고 문헌

[1] Raymond J. Mooney, Loriene Roy, "Content-based Book Recommending Using Learning for test categorization," Proceedings of the 5th ACM conference on Digital Libraies, San Antonio, TX, pp.195-204, June, 2000.

[2] Weiyang Lin, Sergio A. Alvarez, Carolina Ruiz, "Collaborative Recommendation via Adaptive Association Rule Mining," WEBKDD 2000(Workshop on Web Mining for E-commerce - Challenges and opportunities), Boston, MA, August, 2000.

[3] Nathaniel Good, J.Ben Schafer, Joseph A. Konstan, Al Borchers, Badrul Sarwar, Jon Herlocker and John Riedl, "Collaborative Filtering with Personal Agents for Better Recommendation," Proceedings of the 1999 conference of the American Association of Artificial Intelligence(AAAI-99), July, 1999.

[4] D.M. Pennock and E. Horvitz, "Collaborative filtering by personality diagnosis : A hybrid memory-and model-based approach," In IJCAI Workshop on Machine Learning for Information Filtering, International Joint Conference on Artificial Intelligence, Stockholm, Sweden, August, 1999.

[5] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. "GroupLens: An open architecture for collaborative filtering of netnews," In Proceedings of the ACM Conference on Computer Supported Cooperative Work, 1994.

[6] Ken Goldberg and Theresa Roeder and Dhruv Gupta and Chris Perkins, "Eigentaste: A constant time Collaborative Filtering Algorithm," University of California, Berkeley, Electronics Research Laboratory Technical Report M00/41, 2000.

[7] Dae-Sik Jang, Hyung-Il Choi, "Fuzzy Inference system based on fuzzy associative memory," Journal of Intelligent and fuzzy systems, Vol.5, pp. 271-284, 1997.

[8] Kosko B, Neural Networks and Fuzzy Systems, Prentice-Hall International, 1994.

[9] Hideyuki T, Isao H, "NN-Driven fuzzy reasoning," International Journal of Approximate Reasoning, pp. 191-212. 1991.

[10] Freeman JA, Skapura DM, Neural Networks: Algorithms, Applications and Programming Techniques, Addison Wesley Publishing Company, 1991.

[11] Zimmermann HJ, Fuzzy Set Theory and Its Applications, KALA, 1987.

[12] Jonathan Herlocker, Joseph Konstan, Al Borchers, and John Riedl, "An algorithmic framework for performing collaborative filtering," In Proceedings of the SIGIR, ACM, August, 1999.



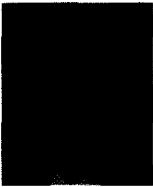
이 동 섭

1997년 군산대학교 학사 졸업(공학사)
 1999년 군산대학교 석사 졸업(공학석사)
 2002년 송실대학교 박사 수료(박사수료)
 2002년~현재 클리쿠 재직중. 관심분야는 퍼지시스템, 패턴인식, 인터페이스 에이전트, 정보추출 등



고 일 주

1992년 송실대학교 전산학과(학사). 1994년 송실대학교 전산학과(석사). 1997년 송실대학교 전산학과(박사). 2003년~현재 송실대학교 미디어학부 전임강사. 관심분야는 HCI, 정보검색, 컴퓨터비전, 감성공학, 컨텐츠평학



김 계 영

1990년 2월 송실대학교 전자계산학과 졸업(공학사). 1992년 2월 송실대학교대학원 컴퓨터학과 졸업(공학석사). 1996년 2월 송실대학교대학원 컴퓨터학과 졸업(공학박사). 1996년 3월~1997년 11월 한국전자통신연구원(Post Doc.). 1997년 12월~2001년 2월 한국전력공사 전력연구원(선임연구원). 2001년 3월~현재 송실대학교 컴퓨터학부(조교수). 관심분야는 컴퓨터비전, 형태인식, 생체인식, 증강현실, 영상 및 신호처리 등임