

표적 마케팅을 위한 CBR 시스템의 유사 임계치 및 커버리지의 동시 최적화 모형

안 현 철*

*한국국방연구원 군사기획연구센터

서울 동대문구 청량리2동 산5-7, 130-012

Tel: +82-2-961-1335, Fax: +82-2-961-1163, E-mail: hcahn@kida.re.kr

초록

사례기반추론(CBR)은 많은 장점으로 인해, 생산, 재무, 마케팅 등의 분야의 다양한 경영의사결정문제 해결에 적용되어 왔다. 그러나, 효과적인 CBR 시스템을 설계, 구축하기 위해서는 연구자가 직관적으로 설정해야 할 많은 변수들이 존재한다. 본 연구에서는 이러한 CBR의 여러 설계요소들 중, '결합할 유사사례의 선택' 과 관련해, CBR이 보다 개선된 형태로 경영문제 해결에 응용될 수 있는 모형을 제시하고 있다. 본 연구의 제안모형은 결합할 유사사례를 선택하는 기준으로 특정 사례수(k -NN)나 유사도의 상대적 비율을 사용하는 기존의 CBR과 달리 0에서 1사이의 값을 갖는 절대적 유사 임계치를 적용하고 있다. 다만, 절대적 유사 임계치를 사용할 때, 그 값이 작아질 경우 예측결과의 생성이 과도하게 이루어지지 않을 수 있는 문제를 해결하기 위해, 커버리지를 모형에 함께 반영하여 사용자가 원하는 수준의 커버리지는 유지한 상태에서 가장 효과적인 유사 사례를 찾아, 추론을 수행할 수 있도록 설계하였다. 제안모형을 검증하기 위해, 본 연구에서는 이 모형을 실제 인터넷 쇼핑물의 고객 발굴 사례에 적용해 보았다. 이를 통해, 제안모형의 적용가능성을 확인하고, 향후 추가연구가 요구되는 개선방향을 고찰해 보았다.

Keywords:

CBR, 유사 임계치, 커버리지, 유전자 알고리즘, 표적 마케팅

1. 서론

분석 고객관계관리(Analytic CRM)에 있어 중요한

이슈 중 하나는 '기업에서 팔고자 하는 상품을 구매할 가능성이 높은 잠재 구매자를 발굴'하는 고객 분류 모형(customer classification model)을 구축하는 것이다. 고객 분류 모형은 다양한 마케팅 기회 창출에 활용될 수 있는데, 예를 들어 일대일 마케팅이나 DM(direct mailing) 발송을 통한 마케팅, 전화나 이메일 등을 이용한 판매 촉진(sales promotion) 등에 있어 대상 고객을 선별하는데 유용하게 활용될 수 있다. 때문에 Ford와 같은 제조업체나 Allstate와 같은 금융회사, 그리고 1-800-flowers.com과 같은 온라인 기업들을 포함한 전세계 유수의 기업들이 고객의 프로필 및 구매 행태를 분석해 상품 구매 가능성을 예측하는 고객 분류 모형을 구축하는데 많은 노력을 기울이고 있다 (<http://www.sas.com/success/> 참고).

이러한 고객 분류 모형을 구축하는데에는 전통적으로 로지스틱 회귀분석(LR, Logistic Regression), 인공신경망(ANN, Artificial Neural Networks), 사례기반추론(CBR, Case-Based Reasoning) 등 다양한 방법들이 적용되어 왔다(Chiu, 2002; Kumar & Reinartz, 2006; Ahn et al., 2006a; Ahn et al., 2006b, Ahn et al., 2007). 그 중에서도 특히 CBR은 적용이 쉽고, 유지보수가 상대적으로 편리하며, 실시간으로 연속해서 학습이 이루어진다는 장점으로 인해, 고객 분류 모형 구축을 비롯한 여러 경영문제 해결에 널리 활용되어 왔다 (Shin & Han, 1999; Kim & Han, 2001; Chiu, 2002; Chiu et al., 2003 등 참고).

그러나 많은 장점에도 불구하고, 효과적인 CBR 시스템을 설계, 구축하기 위해서는 해결해야 할 문제들이 상당수 존재한다. 특히, CBR은 적절한 유사도 측정방법이나 사례 인덱싱 방법, 유사사례간의 결합방법 등에 대해 명확한

방법론이나 원리를 제공해 주지 못하고 있다. 때문에 이러한 CBR의 각종 설계 요소들은 실험자나 사용자가 자신의 경험이나 직관에 의해 결정해야 하는 어려움이 있었다. 특히 이러한 어려움은 때때로 연구자를 만족시키지 못하는 예측성과를 산출하는 문제를 야기시켜, 다른 방법론에 비해 CBR의 활용을 저해하는 요소로 작용하기도 하였다. 이런 이유로 지난 오랜 기간 동안 최적 사례 유사도 측정방법이나 사례의 특징을 대표하는 최적 변수군의 선정방법, 혹은 유사사례 결합시 적용할 가중치의 최적화 방법 등이 많은 연구자들에 의해 연구되어왔다 (Shin & Han, 1999; Kim & Han, 2001; Chiu *et al.*, 2003 등 참고).

이러한 여러 설계 요소들 중에 본 연구에서 집중적으로 개선시키고자 하는 요소는 바로 CBR에서 ‘예측결과를 생성할 때 참조하는 적절한 유사사례의 선택’과 관련된 요소이다. CBR 기법에서 결합할 유사사례의 개수를 최적화 하는 것이 CBR 기법의 성과를 향상시키는데 있어서, 매우 중요한 요소가 될 수 있다는 연구는 기존 연구에서 이미 여러 차례 제시된 바 있다 (Lee & Park, 1999; Garrelli & Guio *et al.*, 1999; Ahn *et al.*, 2006b; 박윤주, 2006). 하지만, 기존 연구들에서는 모두 특정 사례수 (k -NN의 k)나 유사도의 상대적 비율에 의거해 유사 사례를 선정함으로써, 경우에 따라 절대적 관점에서 유사하지 않은 기존 사례를 참고해서 예측결과를 생성할 수도 있는 구조적 한계점을 갖고 있었다.

이에 본 연구에서는 0에서 1사이의 값을 갖는 절대적 유사 임계치(similarity threshold)에 기반한 새로운 CBR 모형을 제안하고자 한다. 그런데, 절대적 유사 임계치는 값이 너무 작게 설정되면, 예측결과의 생성이 과도하게 이루어지지 않을 수 있다. 이에, 본 연구에서는 커버리지(coverage) 변수도 모형에 함께 반영함으로써, 사용자가 원하는 수준의 커버리지는 유지한 상태에서 가장 최적의 절대적 유사 임계치를 탐색하도록 하였다. 상기 두 변수의 동시 최적화 방법으로는, 전통적으로 많이 적용되어 온 유전자 알고리즘(GA, Genetic Algorithms)을 적용하였다.

본 연구는 이 모형의 우수성을 검증하기 위해, 한국 인터넷 쇼핑몰의 특정 상품에 대한 표적 마케팅 대상 고객 발굴 모형에 적용하여 그 예측력을 살펴보고자 하였다.

본 논문의 뒷부분은 다음과 같이 구성된다. 우선 2장에서는 기존 문헌 연구를 간략히 살펴보고,

3장에서는 본 연구의 제안 모형인 유전자 알고리즘 기반의 유사 임계치 및 커버리지 동시 최적화 CBR 모형을 소개한다. 4장에서는 앞서 제시한 모형의 유용성을 검증하기 위한 실험 데이터 및 설계 내용을 설명하고, 5장에서는 실험 결과를 종합적으로 정리해 제시하도록 한다. 끝으로 마지막 절에서는 결론과 함께 향후 연구계획이 함께 제시된다.

2. 문헌연구

본 연구에서 제안하는 모형은 기본적으로 CBR과 GA가 결합된 형태로 구성되어 있다. 이에 기존 문헌을 검토하게 될 본 절에서는 우선 CBR의 기본적인 개념과 원리에 대해 먼저 살펴보고, 이어 CBR과 GA를 결합하고자 시도한 다른 기존 연구들을 살펴보고자 한다. 그리고 끝으로, CBR 시스템에서 결합할 유사사례의 개수나 구성을 최적화 하고자 시도했던 기존 연구들을 살펴보고, 그 한계점을 살펴보고자 한다.

2.1. CBR

CBR은 과거 사례나 경험을 이용해, 주어진 문제에 대한 해답을 찾아내는 문제해결 방법론이다. 일반적으로 다른 주요 인공지능기법들은 문제와 해법 사이의 일반적인 관계를 도출하여 이를 기반으로 추론을 하는 원리로 이루어져 있어, 비교적 정형화된 문제 해결에만 적합하고, 지식도 지속적으로 갱신되기 어려운 구조적인 한계를 가지고 있다. 하지만, CBR은 과거에 축적된 정보만 있으면, 어떤 문제든 해결이 가능하므로 복잡하거나 비구조화된 문제를 해결하는데 유리하며, 지식기반을 지속적으로 업데이트 할 수 있다는 측면에서 상대적으로 우수하다고 할 수 있다 (Shin & Han, 1999).

CBR은 다음 <그림 1>에 제시되어 있는 것과 같이, 이른바 4R이라 불리는 크게 4단계의 절차에 의해 이루어진다 (Aamodt & Plaza, 1994). 이 중에서, CBR 시스템의 효과를 결정짓는 가장 중요한 단계는 바로 1단계인 RETRIEVE 이다. 이 단계에서 시스템이 주어진 문제의 해결에 도움이 될 것으로 추정되는 사례들을 선택하게 되는데, ‘어떤 원리로 유사 사례들을 선별해서, 이들을 어떻게 조합해, 추천 결과를 만들어낼 것인가?’ 하는 것에 따라 CBR 시스템의 성능이 크게 변화하기 때문이다. 때문에 사례간 유사도를 어떻게 측정할 것인가,

추천 결과를 도출할 때 유사 사례는 얼마나 결합할 것인가 하는 등의 문제는 전통적으로 주요 CBR의 연구주제로 자리매김하여 왔다 (Chiu, 2002).

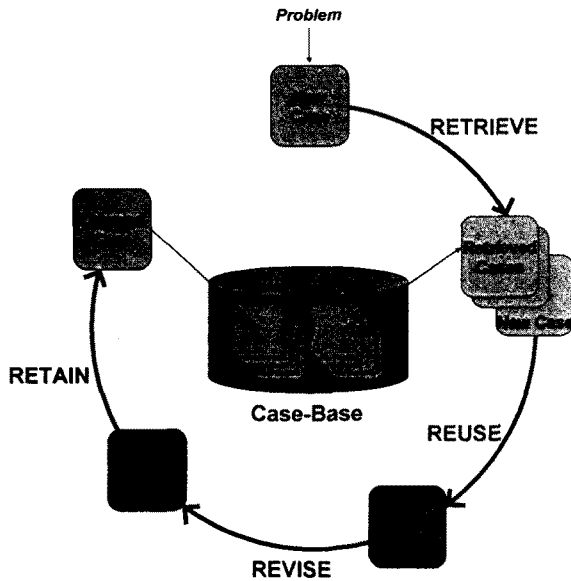


그림 1 - 일반적인 CBR 프로세스

2.2. GA를 이용한 CBR의 최적화

앞서 언급한 CBR의 최적화와 관련한 연구들에서 최적화 수단으로 가장 많이 활용되고 있는 알고리즘 중 하나가 바로 유전자 알고리즘(GA)이다. GA는 유전자, 선택, 교배, 돌연변이 등 생물학의 진화이론에 그 근간을 두고 있다. GA는 방대하고 복잡한 공간을 탐색하면서, 최적 혹은 최적에 가장 가까운 결과를 찾아주는 확률적 검색방법을 이용하는데, 이러한 특징 때문에 다양한 제약식을 포함한 상황에서 목적 함수(objective function)를 최적화 하는 ‘파라미터(parameter)’ 추정에도 널리 적용되고 있다 (Shin & Han, 1999).

CBR에 GA가 적용된 경우를 살펴보면, 우선 입력변수 선정에 GA를 적용한 Siedlecki & Sklansky (1989)의 연구와 GA를 입력변수의 범주화에 사용한 Kim & Han (2001)의 연구를 들 수 있다. GA를 CBR의 입력변수 가중치 선정에 적용한 기존 연구는 매우 많은데, Shin & Han (1999)의 연구나, Chiu *et al.* (2003) Chiu (2002)의 연구가 대표적인 기존 연구라고 할 수 있다. 그 외 *k*-NN의 *k*를 최적화한 Ahn *et al.* (2006b)의 연구와 여러 요소들을 GA로 동시에 최적화하고자 한 Kuncheva & Jain (1999), Ahn *et al.* (2006a, 2007)의 연구 역시 GA로 CBR을 최적화를 시도한 기존 연구들이다.

2.3. 결합 유사사례에 대한 최적화 연구

지금까지 살펴본 바와 같이, CBR 시스템에서 입력 변수의 선정 혹은 가중치 선정의 최적화와 관련해서는 지금까지 많은 연구가 이루어져 왔고, 그 결과 이미 다양한 방법들이 소개된 바 있다. 그러나, 결합 유사사례를 최적화 하기 위한 연구는 아직까지 그 수가 많지 않은 상태이다. 이 부분과 관련해 몇몇 기존 연구들을 살펴보면 다음과 같다.

우선, 이훈영과 박기남 (1999)의 연구는 *k*-NN의 *k*를 최적화하려고 시도하였는데, 유사도 분포에 따른 최적화 수리모형 기법을 제시하였다. 이 연구는 *k*-NN의 *k* 최적화를 시도한 첫번째 연구라는 측면에서는 의의가 있으나, 목표사례가 변화할 때마다 최적화 모형이 변화해 새로운 *k*값을 계속 계산해야 한다는 구조적 한계를 안고 있다.

Kim *et al.* (2002)은 유사사례의 수를 찾기 위해 교차검정방법 (cross validation method)를 사용하여 학습용 자료에서의 평균제곱오차를 최소화하는 방식을 이용하였는데, 이 연구에서 제안한 방식은 최적유사사례의 수를 탐색하는 공간이 한정적이므로 전역 최적화된 유사사례 수를 제시할 수 없다는 한계를 가지고 있다. 이에 Ahn *et al.* (2006b)은 상기 두 연구의 한계점을 극복하기 위해, GA를 활용한 유사사례 수 최적화 모형을 제안한 바 있다.

Sun & Hui (2006)는 유사사례 선정에 고정된 사례의 수(*k*-NN의 *k*)가 아닌, 상대적 비율값을 갖는 유사 임계치(similarity threshold)를 사용하는 방법을 제안하였다. 아래 식 (1), (2)가 이러한 방식을 설명하고 있는데, 대상 사례 u_0 의 예측결과 생성시 참고되는 유사 사례집단 U^* 가 비율값으로 주어지는 상대적 유사 임계치 p 에 의해 결정됨을 알 수 있다.

$$U^* = \{u_i | sim_{0,i} \geq T\} \text{ where } i^* = 1, 2, \dots, k \quad (1)$$

$$T = p \max(sim_{0,i}) \quad \forall i \quad (2)$$

박윤주(2006)는 앞서 살펴본 Sun & Hui (2006)과 유사하지만, 다소 차이가 있는 상대적 유사 임계치 사용을 제안하였다. 앞의 연구가 최대 유사도 거리 대비 상대적 비율을 유사 임계치로 사용했다면, 이 연구에서는 전체 학습 사례수 대비 상대적 비율을 유사 임계치로 제안하였다. 예를 들어, 이 연구모형에서는 유사 임계치를 5%로 설정할 경우, 전체 학습 사례 중 대상 사례와 가장 유사한 상위 5%의 결과를 참조해 예측결과를 생성하게 된다.

3. GA를 활용한 유사 임계치와 커버리지의 동시 최적화 모형

앞서 언급한 기존 연구들에서 제시하고 있는 CBR 모형들은 모두 특정 사례수(k -NN의 k)나 유사도의 상대적 비율에 의거해 유사 사례를 선정하고 있다. 때문에, 절대적인 관점에서 주어진 문제를 해결하기 위해 참조하기에 적합한 유사한 사례가 아님에도 불구하고, 다른 사례들과 비교할 때 그나마 상대적으로 더 유사하다는 이유로 다소 억지스러운 예측결과를 생성할 수도 있는 내재적인 문제점을 안고 있다.

스팸 메일의 사례에서 볼 수 있듯이, 표적 마케팅의 경우에는 상품을 원치 않는 사람에게 마케팅을 시도할 경우, 오히려 고객의 불만만 가져올 수 있는 위험을 갖고 있다. 때문에 이 경우에는, 정말 확신할 수 있는 엄선된 표적에 대해서만 마케팅을 시도하는 것은 상당히 중요하다. 그런데, 기존의 CBR 시스템처럼 예측결과의 확신 정도에 관계없이 무조건 상대적으로 더 유사한 사례들을 참조해 결과를 생성하는 경우, 그 결과의 효과성은 떨어질 수 밖에 없을 것이다.

이에 본 연구에서는 0에서 1사이의 값을 갖는 절대적 유사 임계치(similarity threshold)에 기반한 새로운 CBR 모형을 제안하고자 한다. 즉, 절대적 유사 임계치를 기준으로 적용할 때, 유사한 사례가 하나도 나오지 않으면, 그 경우에는 예측결과를 생성하지 않고 '모름(don't know)'으로 결과를 회신할 수 있는 CBR 시스템을 제안하고자 하는 것이다.

이 같은 새로운 CBR 모형을 이진분류 문제에 적용할 경우, 기존 모형처럼 0 또는 1의 두가지 경우로만 결과를 예측하지 않고, 3가지 경우(0, 1, 모름)로 결과를 제공하기 때문에, 확실한 예측 대상을 발굴하는 것이 중요한 의미를 지니는 마케팅 분야나 의료 분야에 적용 시, 큰 가치를 지니게 될 수 있다.

그런데, 절대적 유사 임계치를 사용할 경우, 그 값이 너무 작아지면 예측결과의 생성이 과도하게 이루어지지 않을 수 있다. 이러한 한계를 보완하기 위해, 본 연구에서는 결과의 정밀도와 서로 상충(trade-off) 관계에 있는 커버리지(coverage) 변수를 모형에 함께 반영함으로써, 사용자가 원하는 수준의 커버리지는 유지한 상태에서 가장 효과적인 유사 사례를 찾아, 추론을 수행할 수 있도록 하였다. 그리고, 이러한 서로 다른 특성을 가진 두 변수인

절대적 유사 임계치와 커버리지를 동시에 최적화하기 위해서, 전통적으로 최적화에 많이 적용되어 온 GA를 적용하였다.

본 연구에서 제안하는 새로운 CBR 모형의 전체적인 프레임워크를 설명하면, 다음의 <그림 2>와 같다. <그림 2>에 제시된 바와 같이, 모형은 다음과 같이 크게 4단계 절차에 의해 구현된다.

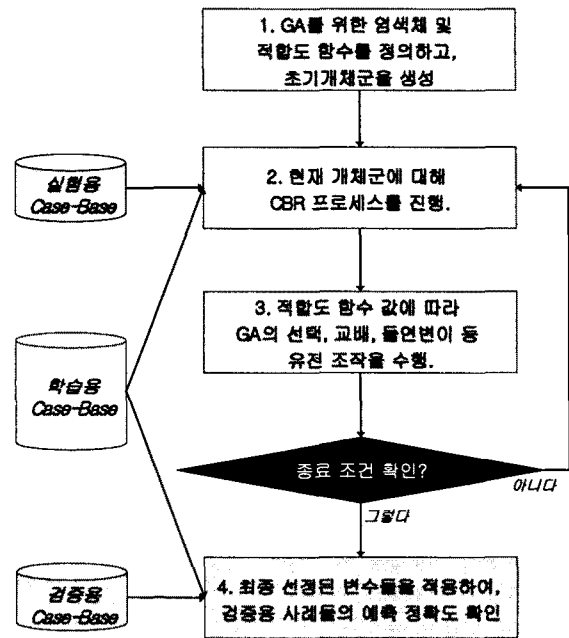


그림 2- 제안 모형의 프레임워크

1단계. 1단계에서는 우선 절대적 유사 임계치 값이 가질 수 있는 전체 탐색 공간 중에서, 가장 최적 혹은 최적에 가까울 것으로 추정되는 유사 임계치 값들을 임시로 설정하게 된다. 개체군 (최적의 유사 임계치를 찾아내기 위한 초기 탐색 위치들의 집합)은 본격적인 탐색 프로세스에 앞서, 무작위값으로 초기화되고, 탐색의 대상이 되는 유사 임계치는 GA가 인식할 수 있도록 염색체 (chromosome) 형태로 코드화 된다. 이렇게 코드화된 염색체는 특정 적합도 함수 (fitness function)를 최대화 하는 방향으로 진화해 나가게 된다. 본 연구에서 염색체는 유사 임계치가 1/10,000의 정확도를 갖는 0에서 1사이의 값을 가질 수 있도록, 14비트의 이진수로 모델링하였다. 적합도 함수는 CBR연구에서 일반적으로 가장 많이 적용되는 실험용 데이터 (test data)에 대한 평균 예측 정확도로 설정하였다.

2단계. 1단계 과정을 통해 개체군이 도출되면, 여기서는 개체군에 속한 개별 염색체들을 대상으로 CBR을 작동하게 된다. 본 연구에서 절대적 유사

임계치를 적용하기 위해서는, 유사도 값이 어떤 경우에도 0에서 1사이의 절대값을 갖도록 산출하는 것이 기본적으로 요구된다. 이에, 본 연구에서는 모든 입력변수들(a_k)에 최소-최대 정규화(Min-Max Normalization)를 적용한 뒤, 아래 식 (3)과 같이 유클리드 거리(Euclidean distance) 기반의 유사도를 계산하였다.

$$sim(u_o, u_i) = \sqrt{\sum_{k=1}^m (a_k^o - a_k^i)^2} / m \quad (3)$$

where $u_i = (a_1^i, a_2^i, \dots, a_m^i)$, $k=1, \dots, m$

이 때, 모든 a_k 는 $0 \leq a_k \leq 1$ 을 만족하므로, 어떤 경우에도 유사도 값은 0에서 1사이의 값을 갖게 된다. 본 연구에서는 최적의 결합 유사사례에만 관심을 갖고 있는 상황이므로, 입력변수의 가중치 등과 같은 다른 CBR의 설계요소들은 대부분의 CBR 시스템에서 적용하는 것과 같이 동일 가중치를 적용하였다 (Chiu et al., 2003).

3단계. 3단계에서는 2단계의 결과들, 즉 각 후보 유사 임계치 적용시 도출되는 실험용 데이터의 평균 예측 정확도에 의거해, 우수한 유사 임계치들을 선별하고 이들에게 각종 유전 조작을 수행해 새로운 개체군을 생성하는 작업이 이루어지게 된다. 이 단계에서 기설정된 목표 커버리지를 달성하지 못하는 유사 임계치들 역시 선택에서 제외되게 된다. 여러 유전 조작 기법 중 본 연구에서는 선택(selection), 교배(crossover), 돌연변이(mutation)의 3가지를 적용하며, 사전에 설정한 중지 조건이 만족될 때까지 2단계, 3단계 작업을 계속 반복한다.

4단계. 앞의 단계들이 모두 끝나면, GA의 중지 조건이 만족되는 시점에서 최적 혹은 최적에 근접한 유사 임계치가 도출되게 된다. 4단계에서는 이렇게 도출된 유사 임계치를 이용, CBR 모형 구축에 사용하지 않은 검증용 데이터(hold-out data)에 적용함으로써, 그 결과를 살펴보게 된다. 본 단계를 통해, 연구의 제안모형의 실제적인 적용가능성을 확인해 볼 수 있다.

4. 실험 및 결과

4.1. 적용 사례 및 실험 설계

제안된 연구모형의 유용성을 검증하기 위해, 본 연구에서는 국내 한 다이어트 인터넷 쇼핑몰의 고객

분류 모형 구축 사례에 제안 모형을 적용해 보았다. 이 쇼핑몰에서는 여러 상품을 취급하고 있지만, 그 중에서 마진폭이 큰 ‘다이어트 보조식’의 판매에 큰 관심을 갖고 있다. 이에, 회원가입시 입력되는 고객 관련 정보를 활용해, ‘다이어트 보조식’의 구매여부를 예측하는 고객 분류 모형의 구축에 많은 관심을 갖고 있다. 이에 본 연구에서는 ‘표적 마케팅’을 위한 대상 고객을 산출할 수 있도록, 본 연구의 제안모형을 적용해 보고자 하였다.

활용한 데이터는 총 980건이었고, t-검정 등을 통해 대상 상품 구매여부와 유의한 관계가 있는 것으로 나타난 총 14개의 변수를 활용하였다. 데이터는 용도에 따라 학습용, 실험용, 검증용의 세 집단으로 구분했는데, 각각 60%, 20%, 20%를 차지하도록 구분하였다.

GA 탐색을 위한 제어 파라미터들과 관련해서는 개체군의 규모를 50개체 (organisms)로 설정하였으며, 교배 및 돌연변이 비율에 대해서는 각각 0.7, 0.1로 설정하였다. 아울러 중지 조건으로는 1000회 반복, 즉 20세대만큼 탐색을 반복하도록 설정하였다. CBR은 Microsoft Excel VBA (Visual Basic for Applications)를 이용해 구현하였으며, GA 최적화는 Palisade Software사의 Evolver Version 4.08를 결합하여 수행하였다.

4.2. 실험 결과

현재 본 연구는 실험이 완전히 마무리된 상태가 아니며, 미완성의 실험결과만 도출되어 있는 상태이다. 다소 미흡하지만, 다음의 표 1에 현재까지 확인된 전체적인 실험결과가 종합적으로 제시되어 있다. 표 1에서 볼 수 있듯이, 본 연구의 제안 모형은 결합할 유사사례와 관련해 아무런 설정을 하지 않는 전통적인 CBR 모형들에 비해 약간 더 우수한 성과를 나타내고 있다

표 1 - 실험 결과

구분		기존CBR		
		100%	90%	20%
목표 커버리지	실험용	100%	90.60%	21.37%
	검증용	100%	94.92%	23.73%
예측 정확도	실험용	64.96%	69.81%	73.68%
	검증용	65.25%	69.64%	73.91%

하지만 이와 같은 성과는 기대수준에 다소 못 미치는 결과로서, 그 원인 분석 및 대안에 대해서는 다음 5장 결론부분에서 추가로 논의하고자 한다.

5. 결론

본 논문에서는 이른바 ‘절대적 유사 임계치’와 ‘커버리지’ 개념을 도입하여, 사용자가 요구하는 정확도 수준에 따라 예측결과를 생성하는 새로운 개념의 CBR 시스템을 제안하였다. 본 연구의 제안모형은 이진분류 문제에서, 확실한 예측결과를 생성하기 어려운 경우, ‘모름’으로 회신함으로서 보다 확실한 경우에 대해서만 결과를 생성하는 것이 가능하도록 설계되어 있다. 이러한 본 연구의 제안모형은 확실한 예측 대상을 발굴하는 것이 중요한 의미를 지니는 표적 마케팅 분야나 의료 분야에 특히 유의할 것으로 전망된다.

그런데 앞의 실험결과 부분에 제시된 바와 같이, 본 연구의 제안모형을 실제 사례에 적용했을 때 예측성능을 크게 개선하지 못하고 있는 것으로 파악된다. 따라서, 현재 제안된 연구모형은 개선이 요구되는 상황이다. 저자는 ‘참조되는 사례기반(즉, 학습사례)에 대한 정제 과정의 부재’가 현재 제안모형의 가장 큰 문제점인 것으로 예상하고 있다. 유사 임계치가 의미를 갖기 위해서는 기본적으로 모든 참조사례(학습사례)가 대표성을 갖는 정제된 사례라는 가정이 선행되어야 하는데, 현실세계 참조사례들은 소위 이상치(outlier)들을 상당수 포함하고 있기 때문이다. 때문에 앞으로 현 모형에 이와 같은 결점을 보완할 수 있는 요소를 결합하여, 모형을 보완하는 연구가 추가로 이루어져야 할 것으로 예상된다.

참고문헌

- [1] 이훈영, 박기남. (1999). “사례기반예측시스템의 정확한 예측을 위한 최적 결합 사례개수 결정방법에 관한 연구,” *경영학연구*, Vol. 27, pp. 1239-1252.
- [2] 박윤주. (2006). “통계적 분석 기법을 기반으로 한 사례기반추론에 대한 연구,” 박사학위논문, 경영공학전공, 한국과학기술원.
- [3] Aamodt, A. and Plaza, E. (1994). “Case-based reasoning; Foundational issues, methodological variations, and system approaches,” *AI Communications*, Vol. 7(1), 39-59.
- [4] Ahn, H., Kim, K.-j. and Han, I. (2006a). “Hybrid Genetic Algorithms and Case-based Reasoning Systems for Customer Classification,” *Expert Systems*, Vol. 23(3), pp. 127-144.
- [5] Ahn, H., Kim, K.-j. and Han, I. (2006b). “Global optimization of feature weights and the number of neighbors that combine in a CBR system,” *Expert Systems*, Vol. 23(5), pp. 290-301.
- [6] Ahn, H., Kim, K.-j. and Han, I. (2007). “A Case-based Reasoning System with the Two-Dimensional Reduction Technique for Customer Classification,” *Expert Systems with Applications*, Vol. 32(4), pp. 1011-1019.
- [7] Chiu, C. (2002). “A case-based customer classification approach for direct marketing,” *Expert Systems with Applications*, Vol. 22, pp. 163-168.
- [8] Chiu, C., Chang, P. C., and Chiu, N. H. (2003). “A case-based expert support system for due-date assignment in a water fabrication factory,” *Journal of Intelligent Manufacturing*, Vol. 14, pp. 287-296.
- [9] Gifford, E. (1995). *Investor's guide to technical analysis: Predicting price action in the markets*. Pitman Publishing: London.
- [10] Kim, K. and Han, I. (2001). “Maintaining case-based reasoning systems using a genetic algorithms approach,” *Expert Systems with Applications*, Vol. 21, pp. 139-145.
- [11] Kim, T.S., Yoon, J.H., and Lee, H.K. (2002). “Performance of a nonparametric multivariate nearest neighbor model in the prediction of stock index returns,” *Asia Pacific Management Review*, Vol. 7, pp. 107-118.
- [12] Kumar, V. and Reinartz, W. J. (2006). *Customer Relationship Management: A Databased Approach*. John Wiley & Sons: NJ.
- [13] Kuncheva, L. I. and Jain, L. C. (1999). “Nearest neighbor classifier: Simultaneous editing and feature selection,” *Pattern Recognition Letters*, Vol. 20, pp. 1149-1156.
- [14] Shin, K. S. and Han, I. (1999). “Case-based reasoning supported by genetic algorithms for corporate bond rating,” *Expert Systems with Applications*, Vol. 16, pp. 85-95.
- [15] Siedlecki, W. and Sklanski, J. (1989). “A note on genetic algorithms for large-scale feature selection,” *Pattern Recognition Letters*, Vol. 10, pp. 335-347.
- [16] Sun, J. and Hui, X.-F. (2006). “Financial Distress Prediction Based on Similarity Weighted Voting CBR,” *Lecture Notes in Artificial Intelligence*, Vol. 4093, pp. 947-958.