

설명기반 유전알고리즘을 활용한 경영성과 데이터베이스의 데이터마이닝

조성훈* · 정민용**

Data-Mining in Business Performance Database Using Explanation-Based Genetic Algorithms

Seong-Hoon Cho* · Min-Yong Chung**

■ Abstract ■

In recent environment of dynamic management, there is growing recognition that information and knowledge management systems are essential for efficient/effective decision making by CEO. To cope with this situation, we suggest the Data-Mining scheme as a key component of integrated information and knowledge management system. The proposed system measures business performance by considering both VA (Value-Added), which represents stakeholder's point of view and EVA (Economic Value-Added), which represents shareholder's point of view. To mine the new information & knowledge discovery, we applied the improved genetic algorithms that consider predictability, understandability (lucidity) and reasonability factors simultaneously, we use a linear combination model for GAs learning structure. Although this model's predictability will be more decreased than non-linear model, this model can increase the knowledge's understandability that is meaning of induced values. Moreover, we introduce a random variable scheme based on normal distribution for initial chromosomes in GAs, so we can expect to increase the knowledge's reasonability that is degree of expert's acceptability. The random variable scheme based on normal distribution uses statistical correlation/determination coefficient that is calculated with training data. To demonstrate the performance of the system, we conducted a case study using financial data of Korean automobile industry over 16 years from 1981 to 1996, which is taken from database of KISFAS (Korea Investors Services Financial Analysis System).

* 삼성전자 경영혁신팀

** 건국대학교 산업공학과

1. 서론

기업의 최고이사결정자인 CEO는 급속하게 변화하는 경영환경에 대응하기 위해서, 기업의 전 조직이 정보·지식을 동시에 공유할 수 있는 효율적인 의사결정 체계를 요구하고 있다. 이러한 의미에서 기업은 정보기술을 이용하여 적절하게 통합된 정보·지식을 꼭 필요로 하는 부서에 적시에 제공할 수 있어야 하며, 정보·지식 공유를 통하여 모든 조직원의 역량을 기업 목표라는 초점에 집중시킬 수 있는 기반을 갖고 있어야 한다.

따라서 기업의 경영관리영역인 재무(Finance)·생산(Manufacturing)·판매(Marketing) 등에서 발생하는 기업의 현황을 자동화된 정보체제로 흡수·축적하여 조직 전체가 공유할 수 있는 정보·지식으로 변환하는 체계를 구축하는 것이 기업의 생존에 중요한 요소로 부각되고 있다.

이에 본 연구는 경영성과데이터의 지속적 관리차원에서 데이터웨어하우스(Data Warehouse)를 구축하고 경영성과의 변화를 주도하는 중요 관리변수를 추출하기 위하여 유전알고리즘(Genetic Algorithms)을 활용한 데이터마이닝(Data Mining) 체계를 구현하였다. 제시된 설명기반 유전알고리즘 체계는 적합도함수와 같은 예측력(Predictability)이 외에도 이해력(Understandability, Lucidity)과 타당성(Reasonability)을 동시에 고려한 구조를 갖는다.

본 연구에서 제시하는 체계는 경영성과지표를 관리하는 의사결정자의 관점에서, 정보기술의 효과적인 활용을 통하여 기업 전체의 조직이 공유할 수 있는 경영성과분석체계의 바탕이 될 것이라 판단한다.

2. 경영성과 데이터베이스의 구성

2.1 부가가치와 경제적부가가치를 결합한 경영 성과

현대의 기업경영관리는 부가가치 또는 시장점유

율의 극대화를 추구하는 일본·독일형 기업과 이윤극대화를 강조하는 미국형이 대조를 이루고 있다. 우리나라의 경우는 일반적으로 미국형의 수익을 강조하는 경영정책 및 원칙이 두드러져 보이니 장·단기적으로 기업의 유지·성장을 도모하기 위해서는 장기적 부가가치창출력과 단기적 이익을 동시에 고려할 수 있는 경영관리체계가 절실히 요청된다고 판단된다.

시대적 흐름이나 지역적 차이에 따라 부가가치와 이익, 최근에는 이익을 대신하여 경제적부가가치가 혼재하여 경영성과의 측정기준으로서 사용되고 있다. 전술한 두 가지 목표는 각국의 시장경제의 발달 및 사회·문화적 환경에 따라 크게 2가지 형태로 발달되어왔는데 하나는 미국·영국을 중심으로 자유경쟁의 원리하에서 주주의 부를 극대화하는 주주 자본주의와 다른 하나는 독일·일본을 중심으로 국가의 주도아래 기업을 둘러싼 여러 이해관계자의 이익을 중시하는 이해관계자 자본주의이다. 이러한 흐름에 대하여 김태성은 이해관계자 자본주의측면에서 [부가가치/종업원수]와 주주 자본주의측면에서 [이익/자기자본]이라는 기업 재무평가의 두 가지 목표를 동시에 고찰하여야 한다는 생산성관리네트워크를 제시하였다[4].

이 연구에 의하면 한 기업의 경제 전반의 성장가능성을 측정하는데 중요한 요소로서 종업원 1인당 부가가치를 사용하며, 경영자층에서 일반적으로 다루어지는 경영성과평가의 척도로서 자본투자가의 이익과 관련하여 자기자본에 대한 이익(자기자본이익율)을 사용하였다.

그러나 자기자본이익율에서 사용되는 이익은 회계적 이익개념으로서, 현행의 회계이익으로 경영성과를 평가하는 과정에서는 자기자본조달사용에 대한 대가로 지불되어야 할 자기자본비용을 비용으로 인식하지 않는다. 따라서 단순히 회계상의 법인세 차감후순이익을 기준으로 경영성과를 평가하는 것은 자본제공자인 주주의 입장에서 판단한다면 자본조달에 따른 자본비용을 충분히 감안하지 못한다고 인식할 수 있으며, 사전에 예상된 자기자본비용

보상에 대한 주주의 중요관심사를 외면하게 된다.

회계적 이익이 기업의 경영성과를 올바르게 반영하지 못하는 이유는 다음과 같이 정리할 수 있다.

첫째, 기업의 경영성과가 좋을수록 회계이익은 증가한다. 그러나 회계이익이 큰 경우에도 주주의 요구수익률이 높은 경우 자기자본 조달비용을 많이 지급해야하므로 총자본 비용 보상후의 경영성과는 회계이익과 반드시 비례하지는 않는다. 실제로 기업의 경영상태가 불안한 경우에는 자기자본을 투하하는 주주는 위험률에 비례하여 많은 분량의 자기자본비용을 기대하게 된다.

둘째, 기업의 경영성과에서 이자비용을 공제하여 추정되는 회계이익이 크더라도 기업의 자기자본사용 규모가 큰 경우에는 주주에게 충분한 보상을 지급할 수 없는 경우가 발생할 수 있다. 이 경우 기업의 부채비용이 상대적으로 낮아 이자비용의 지급액은 작아지므로 기업의 회계이익은 커지지만, 자기자본사용부분에 대한 이자비용을 고려한다면 실질적인 기업의 이익은 줄어들 수 밖에 없다. 따라서 타인자본과 자기자본을 동시에 고려한다면 총자본비용 보상후의 경영성과는 회계이익과 반드시 비례관계에 있다고 할 수는 없다.

국내의 연구 결과, 1987년부터 1992년까지 EPS (Earning Per Share)가 모두 양의 값을 갖는 표본 상장기업중 약 70~80%에 이르는 기업들의 「경제적부가가치/주(Share)」가 음(-)의 값을 갖는 것으로 나타났다[3]. 이는 상당수의 기업들이 영업활동을 통하여 얻는 수익성으로 자기자본조달비용을 충분히 보상하지 못하고 있음을 나타내는 결과라고 할 수 있다[20].

결국 회계이익의 경우에는 타인자본비용만을 고정적으로 보상한 후의 경영성과이지만, 경제적부가가치는 타인자본비용과 자기자본비용을 모두 고

려한 경영성과를 의미하기 때문에 자본제공자의 입장에서는 회계이익정보와 요구자기자본비용을 동시에 포함한 경제적부가가치의 개념이 더 유용한 정보원천으로 활용할 수 있다.

이와 같이 기업이 산출한 이익 개념으로서 회계상의 당기순이익보다는 영업관련 현금흐름에 입각한 경제적부가가치가 중시되는 최근의 경향을 중시하여, 본 연구에서는 기업의 경영성과를 평가하는 틀로서 이해관계자 입장에서의 1인당부가가치(부가가치/종업원수)와 자본제공자 입장에서의 투자자본대비경제적부가가치(EVA/평균투자본)를 동시에 사용하고자 한다.

2.2 경영성과 데이터베이스

본 연구에서는 원시데이터베이스로서 기업의 재무제표를 사용하고, 사용자의 요구에 따라 원시데이터로부터 2개의 경영성과지표(1인당부가가치, 투자자본대비경제적부가가치)와 85개의 재무비용을 추출하여 데이터베이스에 저장한다.

기업의 경영활동을 평가하기 위해서는 기업이 공시한 재무제표를 개별적으로 분리하여 판단할 수 없으며, 대차대조표, 손익계산서와 함께 제조원가명세서, 현금흐름 등도 동시에 고려해야한다고 할 수 있다. 이에 본 연구에서는 기업경영의 결과로 발생하는 운영적 재무결과를 5개의 테이블로 분류하여, 「대차대조표」, 「손익계산서」, 「제조원가명세서」, 「현금흐름」, 「기업제반」 테이블을 구성하고 이를 기업의 경영성과분석을 위한 원시데이터로서 사용한다.

그리고 5개의 운영데이터(Operational Data) 테이블을 기반으로 「성장성」, 「규모」, 「수익성(수익/자본)」, 「수익성(수익/비용)」, 「수익성(비용구조)」, 「안정성」, 「활동성」, 「경제적부가가치」, 「생산성」의 9개 정보데이터(Information Data) 테이블을 생성·유지한다. 관리자이외의 분석목적으로 시스템에 접근하는 사용자들은 9개의 정보데이터 테이블에 접근하여 기업의 경영성과를 분석한다.

공신경망을 이용한 연구가 이루어져왔다. 기존의 연구결과를 통해서도 주식시장의 변화 예측, 도산기업 예측 등의 경영·재무분야에서 인공신경망은 기존의 복잡한 통계적 분석 기법보다 더욱 좋은 결과를 보여준다[10].

그러나 인공신경망의 내부배열(은닉층(Hidden Layer)·은닉마디(Hidden Node))은 확실한 논리적 이론보다는 경우에 따라 시행착오 또는 어렵짐작의 방법으로 결정되는 경우가 많으며, 이러한 이론적 토대의 부재가 인공신경망 모형의 응용연구에서 항상 걸림돌이 되어왔다. 또한 인공신경망 모형은 그 자체의 블랙박스(Black Box)적 특성 때문에 입력변수와 출력변수와의 상호관계에 대한 일관적인 이해가 쉽지 않다. 통계모델에서 계수와 변수들의 확률값은 관계의 강도와 유의도를 나타내는데 반하여, 인공신경망 모델에서 내부 가중치들은 입력층·은닉층·출력층내에서 모두 함께 얽혀 있어 변수들간의 상호관계를 분명하게 표출하지 못함으로 결과 도출의 이해력이 떨어진다[6, 12]. 또한 인공신경망의 가장 큰 단점으로 지적되는 것은 학습과정에서 해의 전역성을 확보하기가 어렵다는 것이다[22]. 최근에는 이러한 단점을 극복하고자 기존의 인공신경망 알고리즘을 수정하거나 전역성을 확보할 수 있는 다른 기계학습 기법과의 결합이 많이 시도되는 추세이다[21].

이에 반하여 유전알고리즘은 다차원적이고 노이즈가 포함된 대량의 자료에도 그 적용이 용이하다. 특히, 휴리스틱적 지식체계가 거의 없거나 불완전한 다그룹·다차원의 경우에도 경향을 효과적으로 추출할 수 있기 때문에, 원자료로부터 정보를 분석·요약하고 추출할 수 있는 유용한 데이터마이닝 수단으로 대두되고 있다[19].

국내·외에서도 기존의 최적해 탐색이라는 목적과는 달리, 유전알고리즘을 이용하여 데이터마이닝을 위한 판별 및 패턴인식에 대한 연구가 수행된바 있다[7, 11, 13, 18, 16, 15].

이에 본 연구에서는 기계학습의 한 종류인 유전알고리즘을 통하여 경영성과 데이터베이스의 데이

터마이닝을 수행하고자 한다. 다양한 기계학습 이론중에서 유전알고리즘을 활용하는 이유는 통계적 분석 기법과는 달리 오차분포의 선형성·정규성의 가정이 필요하지 않으며, 해공간이 넓고 자료의 노이즈가 심한 경우에도 비교적 전역해를 빠른 시간에 탐색할 수 있기 때문이다. 실제로 고전적인 통계기법은 데이터베이스와 같이 자료의 양이 방대한 경우 선형성이나 정규성을 만족하기가 더욱 어려우며, 이와 같은 측면에서 데이터베이스를 기반으로 한 지식발견의 영역에서는 유전알고리즘과 같은 기계학습 기법이 유용하게 사용될 수 있다[17].

3.2 설명기반의 유전알고리즘

Hart는 기계학습기법으로 도출된 지식을 전문가의 검증없이 그대로 사용하는 것은 때때로 매우 위험하며, 이러한 지식들을 최종사용자가 어떻게 받아들일지나 하는 것은 데이터마이닝 기법의 실제 활용측면에서 매우 중요한 문제라고 지적한바 있다[1, 2, 8].

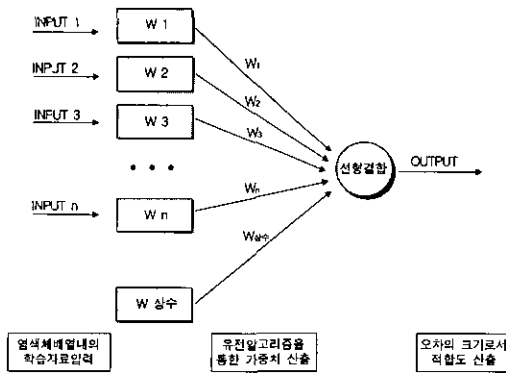
결국 데이터마이닝 모델의 성과는 단순히 예측력만으로는 평가될 수 없으며, 설명력, 이해력, 안정성, 단순성 등 고려해야 하는 많은 기준이 있으므로 전술한 바와 같이 대상문제의 구조화정도에 따라 분석 목적의 효율성을 결정하는 여러 기준을 적절히 고려하여 데이터마이닝 기법을 선정하는 것이 바람직하다. 따라서 경영성과 정보시스템내의 유전알고리즘을 활용한 데이터마이닝 체계는 기존의 기계학습과 같이 적합도 함수의 오차 최소화라는 예측력만으로 지식추출의 최적해 기준을 설정하기는 어려우며, 추출된 지식체계의 명제성 또는 당위성이 확보되어야 한다. 즉 예측력이외에도 이해력이나 타당성과 같은 여러 데이터마이닝 기준을 동시에 만족시키는 경우에만 기업의 경영성과분석이라는 반구조적·비구조적 대상을 적절하게 지식화하여 표현할 수 있다는 것이다.

본 연구에서는 이러한 측면을 고려하여 유전알고리즘을 선형결합형태로 설계함으로써, 복잡

한 비선형구조에 비하여 예측정도는 떨어질 수 있으나 입력변수와 결과변수간의 선형가중치를 변수의 중요도로 그대로 사용하여 도출된 지식의 이해력을 증진시킨다. 또한 유전알고리즘의 초기 유전자값을 일양분포 난수에 의하지 않고, 분석자료입력시 산출되는 입력변수와 결과변수의 상관계수(Correlation Coefficient)와 결정계수(Determination Coefficient)를 정규분포 난수의 Seed로 활용함으로써 추출된 지식의 타당성을 높인다.

3.2.1 선형결합구조의 설계

중요 관리변수 추출을 위한 유전알고리즘에서는 선형결합구조가 사용되며, 선형결합내의 가중치는 관리변수의 중요도로 사용될 수 있다. 선형결합구조를 도시하면 <그림 2>와 같다.



<그림 2> 유전알고리즘내의 선형결합구조

해당 유전자의 선형결합값은 다음의 식 (1)에 의하여 결정된다.

$$W_1[INPUT 1] + W_2[INPUT 2] + \dots + W_n[INPUT n] + W_{상수} \quad (1)$$

이와 같은 체계를 통하여 선형계수를 그대로 변수의 가중치로 사용할 수 있으며, 인공지능망 등의 비선형모형이 갖는 Black-Box적인 특성을 지양하고 그 결과를 쉽게 분석할 수 있으므로 중요 관리변수로서 추출된 변수의 이해력을 높일 수

있다.

3.2.2 정규분포를 이루는 난수구조의 설계

상관분석은 두 변수간의 선형관계의 방향과 그 정도를 나타내는 통계적 분석기법으로서, 상관계수를 사용하면 독립변수로서의 관리변수와 종속변수로서의 목표변수간의 관계를 쉽게 이해할 수 있다. 또한 결정계수는 관리변수가 목표변수의 변동을 설명하는 비율이므로, 상대적인 중요성의 판단 기준으로 사용할 수 있다.

따라서 본 연구에서 제시하는 유전알고리즘은 상관계수로서 초기 유전값의 방향을 결정하고, 결정계수로서 유전값의 크기를 결정한다. 또한 이 결정계수값을 정규분포의 평균값으로 활용함으로써, 유전알고리즘의 무작위 발생 특성은 그대로 살리면서 결정계수 근처의 값이 많이 발생하도록 연산자를 설계한다.

정규분포 난수 발생을 위해서는 반드시 수치해석법(Numerical Method)을 사용하여야 한다. 이는 정규분포의 누적분포나 역함수를 정확하게 표현하는 함수식 구성이 불가능하기 때문이다[9]. 따라서 본 연구에서는 수치해석법으로서 Beasley · Springer가 제시한 방법을 사용한다[14].

이 방법에서는 일양난수 p 에 대하여 $ABS(p-0.5)$ ($p-0.5$)가 0.42보다 적은 경우에는 식 (2)를 사용하고, 반대로 $ABS(p-0.5)$ 가 0.42보다 큰 경우에는 식 (3)을 사용하여 정규분포 난수를 발생시킨다.

$$Random(p) = (p-0.5) \times \frac{A(p)}{B(p)} \quad (2)$$

$$Random(r) = \pm \frac{C(r)}{D(r)}, \quad r = \sqrt{\ln(0.5) - ABS(p-0.5)} \quad (3)$$

여기서 $A(p), B(p), C(r), D(r)$ 은 각각 p 의 3차 다항식, p 의 4차 다항식, r 의 3차 다항식, r 의 2차 다항식이다.

관리변수 입력자료 (m개의 n차원 벡터)	Input_11	Input_12	...	Input_1j	...	Input_1n
	Input_21	Input_22	...	Input_2j	...	Input_2n
	...					
	Input_i1	Input_i2	...	Input_ij	...	Input_in
	...					
Input_m1	Input_m2	...	Input_m3	...	Input_mn	

목표변수 입력자료 (m개의 1차원 벡터)	(VA · EVA) ₁
	(VA · EVA) ₂
	...
	(VA · EVA) _i
	...
(VA · EVA) _m	

i = 입력자료벡터 구분기호(1 to m)
j = 입력자료내의 관리변수 구분기호(1 to n)

〈그림 3〉 설명기반 유전알고리즘을 위한 학습자료

염색체 배열 (k개의 n+1차원 벡터)	W ₁₁	W ₁₂	...	W _{1j}	...	W _{1n}	W _{상수}
	W ₂₁	W ₂₂	...	W _{2j}	...	W _{2n}	W _{상수}
	...						
	W _{i1}	W _{i2}	...	W _{ij}	...	W _{in}	W _{상수}
	...						
W _{k1}	W _{k2}	...	W _{k3}	...	W _{kn}	W _{상수}	

j = 염색체벡터내의 관리변수 가중치 구분기호(1 to n+1)

〈그림 4〉 설명기반 유전알고리즘내의 유전자형

3.2.3 학습자료 구조와 알고리즘 구조

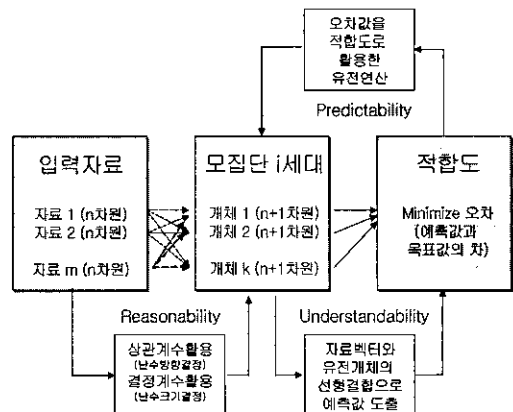
설명기반 유전알고리즘은 경영성과 데이터베이스에서 추출된 데이터를 입력자료로 사용하여 중요 관리변수추출을 위한 선형가중치를 도출한다. 입력자료로 사용되는 학습자료와 관리변수의 중요도를 나타내는 선형가중치를 산출하기 위한 유전자형 배열은 다음의 <그림 3>, <그림 4>와 같다.

3.2.4 학습구조 설계

유전알고리즘을 활용한 중요 관리변수 추출의 개요를 도시하면 다음의 <그림 5>와 같다.

<그림 5>의 학습구조에서는 비선형모형을 활용한 예측력의 증대를 추구하지 않고 선형구조를 채택함으로써 중요 관리변수 추출의 이해력을 도모하였다. 또한 정규분포를 활용한 초기 유전자 값 설계과정을 통하여 지나친 탐색공간의 확대를 방지하고 또한 상관계수와 결정계수를 초기 유전자 값의 Seed로 사용하여, 기계학습에 통계기법의 특성을 결합함으로써 도출된 결과를 더욱 합리적

으로 받아들이 수 있는 타당성이라는 측면을 고려하였다. 이와 같은 설명기반의 유전알고리즘 체계는 예측력(Predictability)이라는 요인이외에도 이해력(Understandability)과 타당성(Reasonability)이라는 지식의 명제적 성격을 동시에 추구한다고 할 수 있다.



〈그림 5〉 중요 관리변수 추출을 위한 유전알고리즘 학습구조

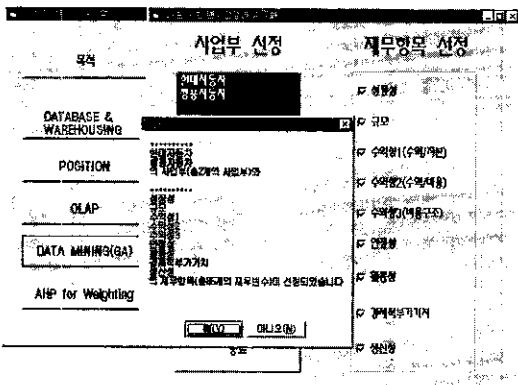
4. 유전알고리즘을 활용한 경영성과 데이터베이스의 데이터마이닝

4.1 경영성과 데이터베이스의 데이터마이닝

본 연구에서 구현한 경영성과분석 시스템은 관계형 데이터베이스를 기반으로 한 클라이언트-서버 구조에 바탕을 두고 있다. 서버의 운영시스템으로 윈도우 NT 4.0을 사용하고 데이터웨어하우스는 MS-SQL 7.0을 사용해 원시 데이터베이스와 같은 시스템상에서 구현하였다. Mining서버는 윈도우 NT 상에서 구현하며 MS-Visual Basic 6.0을 통하여 구축되었으며, Mining클라이언트는 윈도우 98상에서 구현하며 역시 MS-Visual Basic 6.0을 통하여 구현되었다.

4.1.1 데이터의 추출

데이터마이닝을 위해서는 <그림 6>과 같이 사업부와 재무항목을 선정함으로써 분석자료를 추출하여야 한다. 여기서 선정된 자료는 유전알고리즘 내의 학습자료로 활용되므로, <그림 6>와 같이 H 자동차와 S자동차 모두를 선정하고 9개의 재무항목을 모두 선정하는 경우에는 85개의 변수(9개 재무항목내의 재무비율)를 포함하는 32개 학습자료(16년 × 2개 사업부)가 생성된다.

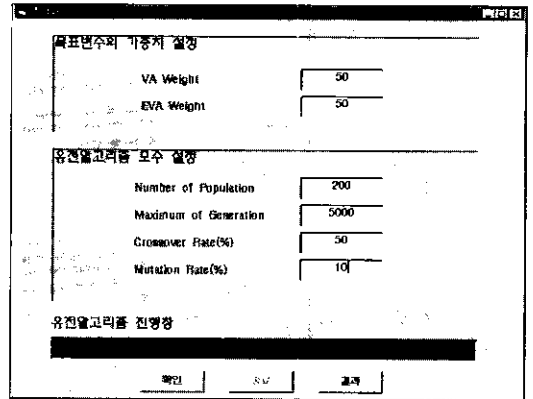


<그림 6> 데이터베이스에서 데이터 추출

4.1.2 유전알고리즘의 적용

이후에는 유전알고리즘에서 사용되는 목표변수

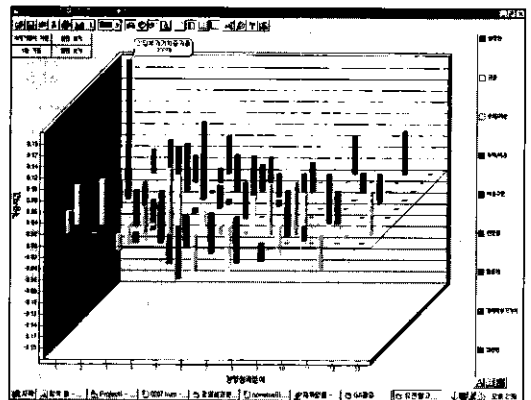
의 가중치와 유전연산자의 모수값을 부여한다. 목표변수의 가중값은 0과 100사이로서, 1인당부가가치만으로 목표변수를 설정하여 1인당부가가치에 상대적으로 민감하게 변동하는 관리변수를 추출하고자 하는 경우에는 VA Weight를 100으로 설정한다. <그림 7>에서는 동일하게 가중치를 부여함으로써 1인당부가가치와 투하자본대비경제적부가가치라는 두 목표변수에 대한 중요 관리변수를 추출하고자 하였다. 또한 모수값을 각각 200모집단, 5000세대, 50%의 교차변이율, 10%의 돌연변이율을 설정하였음을 나타낸다.



<그림 7> 유전알고리즘을 이용한 데이터마이닝

4.1.3 중요 관리변수의 시각화

유전알고리즘의 수행 결과를 도시하면 <그림 8>과 같다.



<그림 8> 중요 관리변수의 시각화

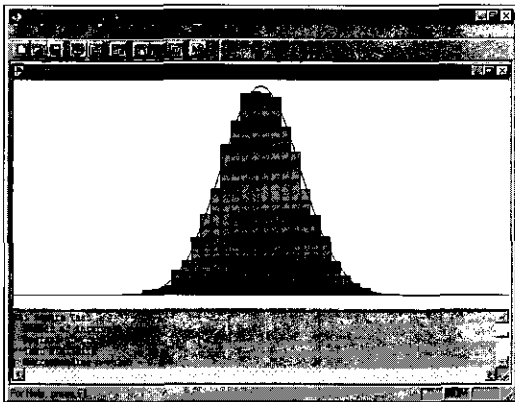
최고가중치를 가지는 1인당부가가치증가율의 경우 0.20658의 값이 가중치로서 산출되었으며, 가장 중요한 관리변수로 추출되었음을 알 수 있다.

4.2 데이터마이닝체계의 유용성

전술한 바와 같이 본 연구에서 제시하는 유전알고리즘은 설명기반의 특성을 지향하기 위하여 예측력과 함께 타당성과 이해력을 동시에 고려한다. 이해력은 선형 학습구조로서 설명될 수 있으므로, 본 절에서는 타당성과 예측력에 대하여 서술한다.

4.2.1 타당성 비교

본 연구에서는 정규분포형 난수를 사용하여 추출된 지식의 타당성을 높이고자 하였다. 유전알고리즘에서 사용된 정규분포형 난수는 다음의 <그림 9>와 같다.



<그림 9> 지식의 타당성 향상을 위한 정규분포형 난수

<그림 9>는 ARENA에서 제공하는 INPUT ANALYZER를 이용하여 유전알고리즘에서 발생된 3000개의 난수에 대한 정규성 검정을 실시한 것이다. 그 결과 Chi-Square 검정통계량의 p값이 0.508에 이르므로 발생된 난수가 정규분포를 따른다고 할 수 있다[5].

통계적 상관분석의 특성을 적절히 반영하는가를 입증하기 위하여 <표 1>에서는 정규분포형 난수

를 사용한 유전알고리즘과 일양분포형 난수를 사용한 유전알고리즘을 비교하였다. 물론 유전알고리즘의 유용성을 판단하는 기준으로 상관분석만을 사용할 수는 없지만, 전술한 바와 같이 통계적 상관분석은 전통적으로 변수간의 영향력을 파악하는 수단으로서 사용되었던 방법이므로 설명기반이라는 관점에서 기계학습으로 추출된 지식의 타당성을 입증할 수 있다고 판단하였다.

본 연구에서 의도한 바와 같이 정규분포형 난수체계를 이용한 결과가 일양분포형 난수체계를 이용한 경우보다 상관분석의 결과에 근접하였음을 알 수 있으며, 특히 정규분포형 난수를 사용한 3000세대와 5000세대의 경우에는 상관계수가 5% 수준에서 유의한 것으로 나타났다.

<표 1> 상관분석을 기준으로 한 정규난수와 일양난수 유전알고리즘의 타당성 비교

비교 항목 (세대수)	정규 난수 (1000)	정규 난수 (3000)	정규 난수 (5000)	일양 난수 (1000)	일양 난수 (3000)	일양 난수 (5000)
적률 상관	0.197	0.238*	0.238*	0.157	0.185	0.185

(* 5% 유의)

4.2.2 예측력 비교

타당성과 함께 유전알고리즘 예측력을 비교하기 위하여 유전알고리즘의 오차와 비선형모형을 사용하는 인공신경망의 오차를 비교한다.

인공신경망기법으로서 후방학습알고리즘(Back Propagation Algorithm)을 적용하는 BrainMaker for Professional V3.11을 사용하였으며, 입력노드는 85개, 출력노드는 1개, 은닉층 1개 그리고 은닉노드 85개를 사용하였다. 32개의 학습자료를 모두 훈련용으로 사용하였으며, 5000회의 반복훈련중에서 가장 작은 RMS(Root Mean Squares)를 기준으로 인공신경망의 예측력을 평가한다.

<표 2>에서 알 수 있는 바와 같이, 비선형모형을 갖고 있는 인공신경망의 경우 오차를 나타내는 RMS가 유전알고리즘보다 큰 것을 알 수 있다. 이는 85개의 관리변수에 대한 85개 차원의 최적해

탐색이 수행됨으로서 전역해를 보장하지 못하는 인공지능망의 특성이 두드러지게 나타난 것이라 해석할 수 있다. 또한 정규분포형 난수를 사용한 유전알고리즘의 RMS가 일양분포형 난수를 사용한 유전알고리즘의 RMS보다 상대적으로 적는데, 이는 초기해의 범위를 적절하게 축소하여 해의 탐색에서 발생하는 낭비를 방지함으로써 나타난 결과로 해석할 수 있다.

〈표 2〉 정규 난수와 일양 난수 유전알고리즘의 예측력 비교

비교 항목 (세대수)	신경망 (5000)	정규 난수 (1000)	정규 난수 (3000)	정규 난수 (5000)	일양 난수 (1000)	일양 난수 (3000)	일양 난수 (5000)
RMS	0.0367	0.0166	0.0143	0.0143	0.0171	0.0150	0.0150

5. 결 론

본 연구에서는 변화하는 기업환경에 부응하기 위하여, 경영성과지표를 관리하는 의사결정자의 관점에서 정보기술의 효과적인 활용을 통하여 기업 전체의 조직이 공유할 수 있는 경영성과분석 체계를 전개하였다. 이를 위하여 기업의 경영성과분석의 틀을 이해관계자 관점에서의 부가가치와 가치경영 관점에서의 경제적부가가치로서 제시하였으며, 제시된 경영성과분석 체계를 통하여 기업내부의 이해관계자는 물론 기업 외부의 이해관계자 역시 기업의 경영성과를 올바르게 평가할 수 있는 토대를 마련함으로써 기업의 경영성과라는 지식체계를 기업 내·외부의 전 조직이 공유할 수 있는 이론적 틀을 제시하였다.

그리고 이러한 이론 체계를 물리적으로 구현하기 위하여 경영성과 데이터웨어하우스를 구축하였고, 경영성과의 변화를 주도하는 중요 관리변수를 추출하기 위하여 유전알고리즘을 활용한 데이터마이닝체계를 구현하였다. 특히 경영성과 데이터베이스내의 데이터가 갖는 비정형적 특성을 고려하여 예측력·타당성·이해력을 동시에 고려한 설명 기반의 유전알고리즘체계를 제시하였다.

이와 같은 체계를 구축함으로써 기존과 같이 단발적인 방법으로 기업의 경영성과를 분석하는 것이 아니라, 경영성과의 변화를 연속적으로 추적하고 이에 영향을 미치는 여러 가지 변수를 실시간으로 분석함으로써, 기업 내·외부의 모든 이해관계자가 경영성과에 대한 정보·지식을 공유할 수 있는 체계를 구축할 수 있다고 판단된다. 그리고 구축된 경영성과에 대한 정보·지식의 발견·공유 체계는 급격한 변화에 능동적으로 대처할 수 있는 바람직한 의사결정의 토대가 될 수 있을 것이라 생각된다.

참 고 문 헌

- [1] 김광용, "여러 가지 Data Mining 기법으로부터 도출된 지식에 관한 전문가의 신뢰도에 대한 실증적 연구", *한국지능정보시스템학회논문지*, 제5권 제1호, (1999), pp.125-143.
- [2] 김현수·이창호, "데이터웨어하우스 환경에서의 설명기반 데이터마이닝", *한국지능정보시스템학회논문지*, 제5권 제2호, (1999), pp.15-27.
- [3] 김철중, "경영성과지표로서의 경제적부가가치의 유용성에 관한 연구", *재무관리논총*, 제2권 제1호, (1995), pp.101-126.
- [4] 김태성, 「생산성관리를 위한 네트워크모델 구축에 관한 연구」, 건국대학교 대학원 박사학위논문, 1997.
- [5] 문일영·윤원영·조규갑·최원준, 「ARE NA를 위한 시뮬레이션」, 교보문고, 1999.
- [6] 서영호·이정호, "인공신경망 모델을 이용한 주식시장에서의 투자전략에 대한 연구", *한국경영과학회지*, 제23권 제4호, (1998), pp.213-223.
- [7] 정원순, 「패턴인식을 위한 유전 알고리즘」, 고려대학교 대학원 박사학위논문, 1997.
- [8] Hart, A., *Knowledge Acquisition for expert systems (2nd Ed.)*, New York : McGraw-Hill, 1992.

- [9] *ARENA Variable Guide*, Systems Modeling Corporation, 1995.
- [10] Wong, B.K., T.A. Bodnovich and Y. Selvi, "Neural Network Applications in Business : A Review and Analysis of the Literature (1988-95)," *Decision Support System*, 19, 1997, pp. 301-320.
- [11] Pereira, C.M., N.A., R. Schirra and A.S. Martinez, "Learning an Optimized Classification System from a Database of Time Series Patterns using Genetic Algorithms," *Data Mining*, WIT Press, 1998, pp.21-34.
- [12] Deboeck, G.J., *Trading on the Edge (Neural, Genetic and Fuzzy Systems for Chaotic Financial Markets)*, John Wiley & Sons, Inc., 1994.
- [13] Flockhart, I.W. and N.J. Radcliffe, "A Genetic Algorithm-Based Approach to Data Mining," *Proceeding 2nd International Conference Knowledge Discovery and Data Mining*, (1996), pp.299-302.
- [14] Beasley, J.D. and S.G. Springer, "The Percentage Points of the Normal Distribution," *Applied Statistics*, Vol.26, (1977), pp.118-121.
- [15] Kelly Jr., J.D. and L. Davis, "A Hybrid Genetic Algorithm for Classification," *Proceeding 12th IJCAI-91*, (1991), pp.645-650.
- [16] Knight, L. and S. Sen, "PLEASE : a Prototype Learning System Using Genetic Algorithms," *Proceeding 6th International Conference Genetic Algorithms*, (1995), pp.429-435.
- [17] Martin, L., F. Moal and C. Vrain, "A Relational Data Mining Tool Based on Genetic Programming," *Research and Development in Knowledge Discovery and Data : Second Asia Pacific Conference ; Proceedings/PAKDD-98*, Springer, (1998), pp.130-138.
- [18] Yao, L., "Nonparametric Learning of Decision Regions via the Genetic Algorithm," *IEEE Transactions on Systems, Man and Cybernetics - Part B : Cybernetics*, Vol.26, No.2, (1996), pp.313-321.
- [19] Pei, M. and E.D. Goodman, "Pattern Discovery from data using genetic algorithm, KDD : Techniques and applications," *Proceedings of the First Pacific-Asia Conference on Knowledge Discovery and Data Mining*, World Scientific, (1997), pp.264-276.
- [20] Strassmann, P.A., *Information Productivity (Assessing the Information Management Costs of U.S. Industrial Corporations)*, The Information Economics Press, 1999.
- [21] Hashemi, R.P., B.A. Pearce, R.B. Arani, W.G. Hinson and M.G. Paule, "A Fusion of Rough Sets, Modified Rough Sets and Genetic Algorithms for Hybrid Diagnostic Systems," *Rough Sets and Data Mining*, Kluwer Academic Publishers, (1997), pp.149-175.
- [22] Tsaih, R., Y. Hsu and C.C. Lai, "Forecasting S&P 500 stock index futures with a hybrid AI system," *Decision Support Systems*, 23, (1998), pp.166-174.