

패턴분류를 위한 지능형 에이전트 시스템

Intelligent Agent System for Pattern Classification

박지훈 · 김대수
Ji Hoon Park and Dae Su Kim

한신대학교 컴퓨터학과
jhpark@hanshin.ac.kr and daekim@hanshin.ac.kr

Abstract

최근 들어 개인용 컴퓨터의 광범위한 보급과 인터넷의 확산으로 인하여 이러한 시스템을 사용하는 사용자들은 보다 편리하고 사용자의 요구에 적절하게 대응할 수 있는 지능화 된 시스템을 필요로 하게 되었다. 이러한 배경을 바탕으로 하여 지능형 에이전트 이론을 활용한 연구와 개발은 다방면에 걸쳐 활성화되고 있다. 본 연구에서는 패턴분류에 있어서의 몇 가지 알고리즘을 통하여 공통된 데이터에 대한 패턴 클러스터링을 통한 패턴의 분류 방법을 고찰하고, 또한 지능형 에이전트 개념을 적용하여 패턴분류를 위한 지능형 에이전트 시스템을 모델링하고 구현하였다. 그 결과 4개의 알고리즘에 대한 300개의 3차원 데이터의 패턴분류가 정확하게 되는 것을 확인하였으며, 본 연구의 핵심분야인 지능형 에이전트 시스템의 다양한 에이전트들을 적용하여 기존의 시스템과는 차별화 된 인터페이스가 이루어질 수 있음을 보인다.

Key Words : 지능형 에이전트, 패턴분류, 관리 에이전트, 클러스터링, 지능시스템

I. 서론

최근 들어 편리한 사용자 인터페이스의 수준을 넘어서 보다 지능적이고 편리하게 사용자의 작업을 대행해주는 시스템이나 프로그램들이 나타나고 있다. 그럼에도 불구하고 아직까지도 시스템과 사용자간에는 상당히 많은 갭이 존재하며 이러한 문제점을 극복하기 위해 보다 편리한 인터페이스에 관한 연구가 1990년대 초반부터 진행되어 왔다.

이러한 연구는 지능적이며, 사용자에게 친숙한 인터페이스 시스템(User Friendly Interface System)으로 발전해오고 있으며, 컴퓨터 사용을 보다 편리하게 대행해 줄 수 있는 에이전트(agent)란 개념으로 적용되고 있다. 에이전트는 사용자를 대신하여 일을 처리해주는 프로그램이라고 볼 수 있는데, 이러한 에이전트라는 개념보다 한 걸음 앞서 자율적이고 지능적인 대행자(autonomous agent)[6]의 개념 즉, 지능형 에이전트(intelligent agent)가 태어났다.

지능형 에이전트는 일반적으로 인공지능, 신경망, 퍼지 등의 지능시스템 이론과 사용자 인터페이스 등의 실용적인 구현 기술과의 접합(fusion)을 통하여 시너지 효과를 얻을 수 있는 지능시스템이라고 할 수 있다. 이러한 지능형 에이전트는 전자상거래(E-commerce)[7], 지식 탐색(knowledge discovery) 그리고 의미 있는 데이터를 캐내는 데이터마이닝(data mining), 웹 상에서의 개인비서(personal assistant of Web)와 같은 다양한 분야에 걸쳐 활용되기 시작하였으며, 사용자에게 보다 편리하고 컴퓨터와 대화가 가능한 시스템으로까지 발전하고 있다.

그러나 지금까지의 지능형 에이전트는 대부분 사용자와의 인터페이스를 제공하는 정도에 불과했고 지능시스템의 이론적인 부분의 구현이나 실용적인 기

술과의 접합은 아직도 매우 어려운 실정이다.

그 뿐만 아니라 실제로 에이전트 상호간의 효율적인 연결, 에이전트간의 작업의 분배 그리고 에이전트 사이의 충돌(conflict)과 오류(error)가 발생했을 경우에 관한 문제 해결은 아직도 미비한 실정이다. 이러한 문제점은 특정한 지능형 에이전트만의 문제점이 아니며 지능형 에이전트의 개념을 바탕으로 한 어떤 응용에서도 동일하게 발생할 수 있는 문제점이다.

우리는 본 연구에서 지금까지는 없었던 사용자에게 편리하게 동작할 수 있는 패턴분류를 위한 시스템을 지능형 에이전트의 개념을 통하여 디자인하고, 보다 지능적인 시스템이기 위해서 지능형 에이전트가 갖추어야 할 전반적인 기능과 지능이 부여 되어야 할 부분을 명시했다.

현재 컴퓨터를 통해 인간의 시각적 인식을 대신할 수 있는 기법인 패턴인식에 대한 연구는 매우 광범위하게 진행되고 있다.

문자, 숫자, 간단한 도형, 그림 등의 2차원의 데이터를 인식할 수 있는 기법들은 이미 개발되어 구현 단계에 있으며, 현재는 음성, 입체영상 등의 3차원 인식에 관한 연구가 진행되고 있다. 패턴을 인식하는 기법 중에서 시간 정보를 포함하는 음성인식의 경우에는 은닉 마르코프(Hidden Markov) 모델을 적용하지만 대부분의 패턴인식에는 패턴분류(pattern classification) 기술을 적용하여 패턴분류가 이루어지고 있다. 이들 기법 중 대표적인 방법으로는 k-means 알고리즘, Fuzzy c-means 알고리즘[2], 그리고 SONN 알고리즘[3, 4] 등이 있다.

그러나 이러한 기법들은 패턴분류를 위한 클러스터링(clustering) 정보와 그에 대한 소속척도(membership)들을 제공해 줄 뿐 사용자를 위한 보

다 친근한 사용자 인터페이스(user interface)까지는 제공하지 않는다. 따라서 3차원 그래픽을 통한 패턴분류의 시각화는 매우 중요한 부분을 차지하며, 현재까지는 거의 시도되지 않은 새로운 방법론으로 볼 수 있다.

또한 본 연구에서는 최근 들어 관심이 집중되고 연구의 필요성이 점차 커지고 있는 사이버 공간에서의 지능형 에이전트에 관한 시스템을 구현하여 이것을 패턴분류를 위한 에이전트로 활용하고자 한다. 지능형 에이전트에 관한 국내외 연구는 아직까지는 모델링 수준이지만 본 연구에서는 에이전트 개념을 적용하고, 이를 실제로 패턴분류에 적용한 지능적인 패턴분류 에이전트 시스템을 개발하고자 하는 것이 본 연구의 목적이다.

패턴(pattern)이란 넓은 의미에서 어떤 현상의 특징을 개념적으로 정립하는 것으로, 문자나 음성 등과의 같은 구체적인 것에서부터 투자 계획의 수립 등과 같이 머릿속에서 논리를 전개하여 보는 개념적인 것에 이르기까지 광범위하다. 그러나 우리가 관심을 갖는 좁은 의미에서의 패턴인식(pattern recognition)[2, 9]은 신문을 읽거나 음악을 들을 때 우리가 무의식적으로 하게되는 인식의 처리를 컴퓨터로 하여금 할 수 있도록 하는 것이다. 즉, 임의로 주어진 패턴을 적당한 클래스에 대응시키는 기능으로 정의하고 클래스(class) 부류 또는 카테고리(category)라는 인식의 결과 같다고 보여지는 패턴의 집합(set) 개념에 해당한다.

컴퓨터 패턴인식을 위한 핵심 기술인 패턴분류에 관한 국내외 연구는 매우 활발하게 진행되고 있다. 그러나 다양한 방법론에 대한 적용이 위주이며 그 결과에 대한 타당성(validity)에 대한 검증에는 상당한 추가적인 연구가 필수적이다. 그 주된 이유는 패턴분류 문제가 명확한 결정(decision)이 아닌 상당히 애매한 소속척도(membership) 문제를 적용해야 한다는 점이다. 이를 위한 국내외 동향으로는 퍼지 이론을 적용한 퍼지 척도를 사용한다는 점인데, 지금까지의 연구는 클러스터링(clustering)의 중심값(center)과 소속 척도만을 알려준다는 점이 제한적인 것이다[2].

따라서 본 연구에서는 소속척도 외에도 중심값과 주어진 데이터를 그래픽으로 표현함으로써 보다 정확하고 합리적인 패턴분류의 새로운 방법론을 제시하여 구현한다는 점이 다른 연구와 두드러진 차이점이라고 할 수 있다.

지능형 에이전트에 관한 국내외 연구 동향은 아직까지는 시작단계로 여겨지고 있다. 미국에서는 지능형 에이전트를 지능적인 사용자 인터페이스로 구현한 사례가 몇 가지 있으며, 국내에서는 전자수표 시스템 및 전자상거래에서 일부 적용시킨 사례가 있으나 패턴인식 등의 핵심기술에 적용한 예는 아직까지는 없는 실정이다.

따라서 패턴분류를 위한 퍼지 결정 모듈과 지능형 에이전트를 결합한 이 시스템은 국내외를 막론하고 새로운 연구로 여겨지며, 그 결과는 상당한 파급효과를 가져올 것으로 예상된다.

앞서 설명한 지능형 에이전트와 최근 관심이 커지고 있는 패턴분류 이론의 접목을 통하여 사용자에게 편리하고 지능적인 패턴분류 에이전트를 제안한다.

본 연구에서는 지능형 패턴분류 에이전트를 이용하여 동일한 어떤 임의의 데이터를 동시에 분석하고 비교하며 또한 사용자가 결과를 쉬게 파악할 수 있

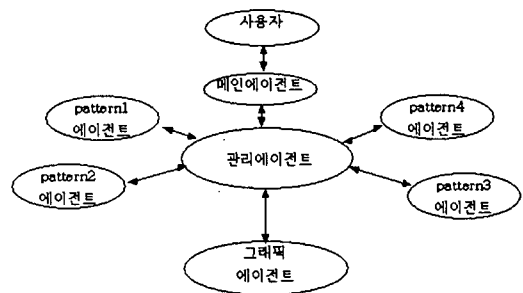
도록 하였다. 즉, 지능형 패턴분류 에이전트 시스템에서는 한가지만의 패턴분류 알고리즘을 이용하지 않고 여러 가지의 패턴분류 알고리즘을 이용하여 데이터를 분석하고 사용자가 결과를 명확하고 쉽게 파악할 수 있도록 그래픽 환경(GUI)에서 그래프를 통하여 보여준다.

II. 지능형 패턴분류 에이전트 시스템

1. 지능형 패턴분류 에이전트 시스템의 설계

패턴분류를 위한 지능형 에이전트의 개념모델은 아래와 같다. 지능적인 에이전트의 역할을 위해서 지능형 패턴분류 에이전트 시스템은 사용자와 직접 인터페이스 하는 메인 에이전트와 그밖에 여러 가지 에이전트들을 총 관리하는 관리 에이전트가 존재한다. 또한 패턴분류를 위한 3가지의 패턴분류 에이전트와 사용자에게 결과를 명확하고 이해하기 쉽게 그래픽 환경에서 보여주는 역할을 하는 그래픽 에이전트를 포함하고 있다. 각각의 에이전트는 저마다 고유한 역할이 있고 전체적인 시스템은 관리 에이전트에 의해서 조정되고 관리된다.

(1) 지능형 패턴분류 에이전트 시스템의 모델링



<그림 1> 지능형 패턴분류 에이전트 시스템

(1) 클러스터링 기법

패턴분류 에이전트는 한가지의 패턴분류 알고리즘만을 사용하지 않는다. <그림 1>에서 보여지는 바와 같이 패턴인식 에이전트는 여러 가지가 존재할 수 있고 각각의 패턴인식 에이전트마다 다른 알고리즘을 가지고 있다. 한가지의 예로 패턴인식 에이전트 중에 k-means 알고리즘을 이용한 패턴인식 에이전트가 있다.

(2) 패턴분류를 위한 지능형 에이전트들의 역할

주요 에이전트의 역할과 기능은 다음과 같다. 패턴분류를 위한 지능형 에이전트 시스템의 구현을 위해서 각각의 특징을 갖는 여러 가지의 에이전트들이 있다. 본 논문에서는 <그림 1>에서처럼 메인 에이전트와 관리 에이전트, 패턴분류 에이전트, 그래픽 에이전트의 역할을 정의하고 구현한다.

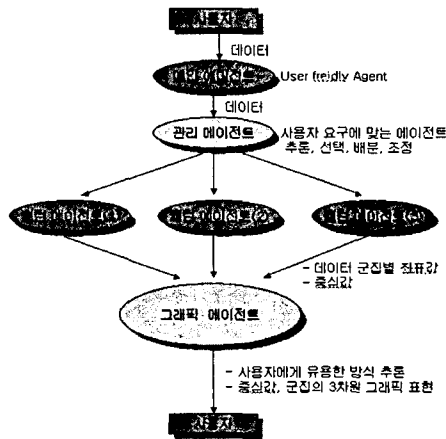
- 메인 에이전트 : 우선 메인 에이전트는 사용자에게 시스템을 원활하게 사용할 수 있도록 도와주는 역할을 한다. 즉, 사용자 인터페이스 에이전트이며 사용자의 요구사항을 받아들이고 사용자가 시스템을 이용하는데 있어서 보다 편리하고 사용하기 쉽도록 하는 에이전트이다. 메인 에이전트는 사용자로부터 요구사항과 패턴분류하기 위한 임의의 데이터를 입력받고 관리 에이전트에게 메시지를 보낸다.

- 관리 에이전트 : 관리 에이전트는 사용자가 원

하는 패턴분류의 방법이 무엇인지를 추론하고 어떤 패턴인식 에이전트를 동작시킬지 결정한다. 즉, 임의의 어떤 데이터를 패턴분류하기 위해 에이전트를 하나만 사용할 것인지 또는 복수개의 에이전트를 이용하여 패턴분류를 할 것인지를 추론하여 결정한다. 또한 관리 에이전트는 에이전트간의 모든 메시지를 감시하고 각 에이전트의 실행순서를 조정한다. 그리고 에이전트간의 오류가 발생하면 통제하는 역할도 함께 가진다. 그러므로 어떻게 작업을 할 것인가를 결정하면 패턴분류 에이전트를 호출한다.

- 패턴분류 에이전트 : 패턴분류 에이전트는 메인 에이전트로부터 패턴분류를 위하여 임의의 데이터를 입력받으면 입력받은 데이터를 각각의 패턴분류 알고리즘을 통해 데이터를 분류하고 분리된 데이터들의 군집 중심값을 찾아낸다. 군집 중심값이 나오면 다시 관리 에이전트에게 메시지를 보내고 관리 에이전트는 결과값을 보여주기 위해서 군집 중심값과 입력된 데이터를 그래픽 에이전트에게 넘긴다.
- 그래픽 에이전트 : <그림 2>에서처럼 그래픽 에이전트는 데이터를 적절한 차원의 그래프로 보여주는데 2차원의 데이터라면 2차원의 그래프로 보여주고 군집 중심을 보여준다. 만일 데이터가 3차원이라면 그래픽 에이전트는 데이터의 차원이 3차원임을 인식하고 3차원의 공간 그래프를 통해서 사용자가 명확하게 파악할 수 있도록 디스플레이 한다. 또한 한가지 데이터를 여러 가지 패턴분류 방법으로 시행했다면 동시에 각각 다른 군집 중심을 가진 데이터를 그래프로 디스플레이 한다.

(3) 지능형 패턴분류 에이전트 시스템 구조도



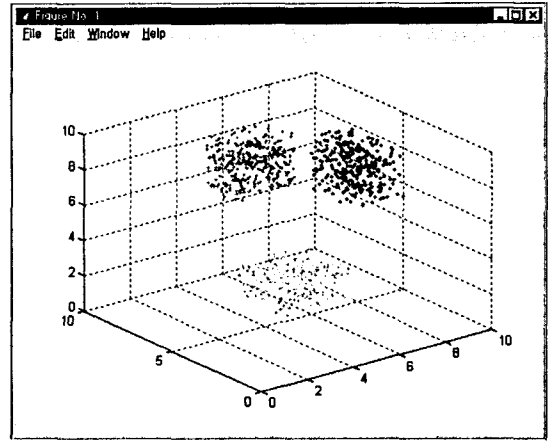
<그림2> 지능형 패턴분류 에이전트 시스템 구조

2. 에이전트 시나리오

우선 패턴분류 에이전트 시스템에 들어오는 데이터는 어떤 형태의 데이터인지 알 수 없는 임의(미지의)의 데이터이다. 즉, 입력값이 3차원 데이터인지 4차원 데이터인지 알 수가 없다. 메인 에이전트는 입력되는 데이터의 형태를 어떤 형식의 데이터인지를 감지한다. 3차원인지 4차원인지(차원 감지), 또한 데이터의 형태가 실수형인지, 지수형인지, 정수형인지, 문자형인지 등을 감지하고 데이터를 선별한다. 즉 메인 에이전트는 사용자의 데이터가 어떤 형태인지 파악하고 추론하여 그 형태를 관리 에이전트에게 보고한다. 관리 에이전트는 추론한 결과를 가지고 어떤 패턴분류 에이전트를 사용할 것인지 또는 분류가 불가

능한지 가능한지를 판단하고 결정(decision)한다.

결정이 되면 데이터를 우선 사용자가 어떤 형태의 데이터인지를 먼저 경험할 수 있도록 <그림 3>에서 보이는 것처럼 그래픽 에이전트를 이용하여 데이터의 군집형태를 보여준다. 2차원의 데이터일 경우 2차원의 그래프로 디스플레이하고 3차원의 데이터의 경우 3차원 그래픽으로 보여준다.



<그림 3> 3차원 데이터 형태를 표현한 그래픽 에이전트

3. 패턴분류 방법론

본 논문에서의 패턴분류에 있어서 적용된 방법론은 식별함수 분류기를 이용하는 방법과 LGB 알고리즘을 이용하여 표준패턴을 선정한 후 NN 규칙을 통해 패턴분류하는 방법, 그리고 K-means 알고리즘을 이용하여 1-NN 규칙에 의하여 패턴분류하는 방법이 있다.

이들 방법론들은 각각 다른 방법으로 데이터의 패턴분류를 시행한다. 분류를 하면서 각각의 군집 데이터들이 어떤 군집에 속하는지를 보여주고 또한 군집의 중심값을 결과로 내놓는다.

① k-means 알고리즘

전체 표본패턴의 군집화를 어느 수렴상태가 될 때까지 반복하는 k-means 알고리즘의 수행과정은 다음과 같다. 여기서 k는 알고리즘의 반복수행 횟수를 나타내는 변수로써 $S_j(k)$ 는 군집 중심이 $z_j(k)$ 인 j 번째 표본 패턴의 집합이고, N_j 는 표본 패턴의 집합 $S_j(k)$ 에 속하는 패턴의 수이다. 또한 K는 군집의 개수이다.

1. K 개의 초기 군집 중심

$z_1(1), z_2(1), \dots, z_k(1)$ 을 임의로 선정한다;

2. $i=1, 2, \dots, K, i \neq j$ 에 대하여 $d(x- z_j(k)) < d(x- z_i(k)) \rightarrow x \in S_j(k)$ 를 수행한다;

3. $j=1, 2, \dots, K$ 에 대하여

$$z_j(k+1) = \frac{1}{N_j} \sum_{x \in S_j(k)} x$$

로 새 군집의 중심 $z_j(k+1)$ 을 계산한다;

4. 모든 $j=1, 2, \dots, K$ 에 대하여

$$z_j(k+1) = z_j(k)$$

를 만족하면 처리를 종료하고, 그렇지 않으면 2단계로 간다.

② NN 규칙을 통해 패턴분류하는 방법

NN 규칙이란 패턴 $x \in V$ 에 대하여 전체 표준패턴

1. $T_1 \cup T_2 \cup \dots \cup T_k$ 에서 유사함의 척도를 나타내는 거리함수 $d(x, t)$ 를 최소로 하는 패턴 t 를 t_0 라 할 때, $t_0 \in T_j$ 이면 인식 함수 $\delta(x)$ 를

2. $\delta(x) = w_j$ 로 하는 규칙이다. 즉, x 를 x 에 가장 가까이 이웃하고 있는 패턴이 속하는 클래스로 분류하는 규칙이다. 단, t_0 가 다수 존재하던가 $t_0 \in T_j$ 이 되는 j 를 일의적으로 결정할 수 없을 때에는 가장 작은 j 를 선정하는 것으로 한다.

나머지 사용될 수 있는 패턴분류 알고리즘으로는 Fuzzy c-means 알고리즘, SONN 알고리즘 등이 있다.

4. 시뮬레이션

본 논문에서 우리는 패턴분류 에이전트를 이용하여 지능형 패턴분류 에이전트 시스템을 모델링 하고 시뮬레이션 해 보았다. 먼저 300개의 데이터에 대한 패턴 클러스터링과 패턴분류를 시뮬레이션 해 본 결과 각각의 패턴분류 에이전트마다 군집 중심값은 약간씩 다르게 나왔으나 큰 차이를 보이지 않았다.

(1) 각 패턴분류 에이전트에 동일한 데이터를 입력한다. 먼저 식별 함수를 이용한 분류 방법과 NN 규칙 분류 방법은 먼저 분류를 하기 위해서 학습이 되어있다고 가정한다. 다시 말해 이 두 가지 방법은 패턴분류를 위해서 먼저 학습 데이터를 통하여 training되어 있고 입력값을 받아들여야한다. 나머지 k-means 군집 분류 방법은 학습이 필요하지 않다.

(2) 입력된 데이터는 패턴분류 되고 각각 군집 중심값을 찾아낸다. 입력된 데이터는 3개의 군집에 속하는 데이터이고 또한 3가지의 패턴을 가지고있는 3차원의 데이터이다.

5.100000 3.500000 1.400000
 4.900000 3.000000 1.400000
 7.100000 3.000000 5.900000
 7.000000 3.200000 4.700000
 : : :

(3) 각 방법론간의 결과 비교
 입력된 데이터의 결과값은 아래의 표와 같은 군집의 중심값이다.

<표1> 식별함수 분류 방법을 통한 군집 중심값

centroid of Class 1	4.671545	4.732413	4.678920
centroid of Class 2	7.352539	7.336982	7.313677
centroid of Class 3	7.258228	7.355703	7.392234

<표2> NN 규칙 분류 방법을 통한 군집 중심값

centroid of Class 1	4.671545	4.732413	4.678920
centroid of Class 2	7.352539	7.336982	7.313677
centroid of Class 3	7.258228	7.355703	7.392234

<표3> k-means 군집 분류 방법을 통한 군집 중심값

centroid of Class 1	0.642100	0.690100	0.969040
centroid of Class 2	7.119900	7.678933	7.039833
centroid of Class 3	6.055129	5.663086	6.235771

III. 결론

본 연구에서 우리는 지능형 에이전트를 이용하여 패턴분류 시스템을 디자인하고 구현 해 보았다. 현재 컴퓨터를 통한 패턴인식의 중요성이 커지고 있는 이 시점에서 패턴분류 방법에 있어서 지능형 에이전트의 적용은 보다 정밀한 패턴분류를 가능케 할 수 있음을 알 수 있었고, 사용자에게 보다 편한 환경에서 결과를 비교하고 각 알고리즘에 따라서 어떻게 차이가 나는지 명확하고 가시적으로 보여줄 수 있었다.

또한 퍼지 판단기(fuzzy decision) 모듈을 적용할 경우에는 패턴인식 기능에 있어서 상당한 기술적 발전을 할 수 있는 계기가 될 것으로 여겨진다. 이와 같은 지능형 에이전트를 활용한 패턴분류 시스템을 소프트웨어 패키지로 만들 경우에 보다 많은 연구자들이 이 프로그램을 활용하여 다양한 분야에 활용할 수 있을 것으로 기대된다. 본 연구를 통하여 지능형 에이전트를 활용한 패턴분류 문제를 보다 편리하게 수행하고 서로의 알고리즘의 효용성을 비교해 볼 수 있는 시스템 체계의 바탕을 마련하였다.

참고문헌

- [1] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York, 1981.
- [2] R. L. Cannon, J. V. Dave, and J. C. Bezdek, "Efficient Implementation of the Fuzzy c-means Clustering Algorithm", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 2, pp. 248-255, 1986.
- [3] Dae Su Kim and T. L. Huntsberger, "Self-organizing Neural Networks for Unsupervised Color Image Recognition", 23rd Southeastern Symposium in System Theory, March 1991.
- [4] Dae Su Kim and T. L. Huntsberger, "Self-organizing Neural Networks for Unsupervised Pattern Recognition", 10th Annual IEEE International Phoenix Conf. on Computers and Communications, March 1991.
- [5] J. T. Tou and R. C. Gonzalez, "Pattern recognition principles", Addison-Wesley, 1977.
- [6] Goodwin R., "Formalizing properties of agents", Technical Report CMU-CS-93-159, School of Computer Science, Carnegie Mellon University, 1993.
- [7] Bird S., "Toward a taxonomy of multi-agent systems", Int. J. of Man-Machine Studies, Vol. 39, pp 689-704, 1993.

감사의 글 : 본 연구는 한국과학재단 목적 기초 연구(2000-2-30300-003-3) 지원으로 수행되었음.