韓國컴퓨터情報學會 論文誌
第14卷 第12號, 2009. 12.

2009-14-12-1-5

전략적 제한에 기초한 지식 및 전략 시스템

구 자 록*

Knowledge and Strategic Ability based on Strategic Constraints

Jarok, Koo *

요 약

다중에이전트 시스템의 분석에 중요한 시간, 지식, 그리고 전략에 관한 개념들을 구체화하기 위해 번역시스템(*Interpreted Systems*)과 *ATL* 및 *ATEL*을 살펴보고, 특히 *ATEL*의 문제점, 즉 하나의 에이전트는 그 자신의 상태에 대해 확실한 정보를 갖지 못하더라도 전략을 구성함에 있어서 전체 시스템의 현재 상태에 접근할 수 있으며, 또한 *ATEL*의 불명확한 행위들에 대한 표현은 일반적인 상황을 모델링하는 것을 어렵게 하는 문제점들을 해결하기 위한 방안으로 게임이론의 서브게임 완전한 나쉬평형(*subgame perfect Nash equilibrium*)에 기초한 전략적 제한(*strategic constraints*)을 그 문제의 해결책으로 제안한다. 또한, 전략적 제한에 기초한 번역시스템을 다중에이전트 시스템에서의 모델체킹(*model checking*)을 위한 하나의 방법으로 제안한다.

Abstract

We study Interpreted Systems, *ATL*, and *ATEL* to capture the notion of time, knowledge, and strategy which are important in the analysis of multi-agent systems and propose strategic constraints based on subgame perfect Nash equilibrium of game theory as one of the solutions for the issues of *ATEL* which an agent can access the current state of the whole system when making up his strategy even when he should be uncertain about the state, and no explicit representation of actions in *ATEL* models makes some natural situations harder to model. Also, we present strategic constraints-based Interpreted Systems for model checking of multi-agent systems.

▸ Keyword : 번역시스템(Interpreted Systems), ATL, ATEL, 게임이론(game theory), 전략적 제한(strategic constraints), 모델체킹(model checking)

• 제1저자 : 구자록
• 투고일 : 2009. 09. 08, 심사일 : 2009. 09. 21, 게재확정일 : 2009. 12. 24.
* 울산대학교 컴퓨터정보통신공학부 부교수
※ 본 논문은 2002년도 울산대학교 교내연구비 지원을 받아 수행되었음

I. 서 론

시간, 지식, 그리고 전략에 관한 추론은 다중에이전트 시스템의 분석에 있어서 중요하다. 이러한 개념들을 구체화하기 위해 다양한 논리적인 체제와 언어들이 개발되었다[4, 11, 13]. 번역시스템은 에이전트들의 시간적 특성과 지식에 관한 특성에 적합한 형식화로 증명되어 왔다. 다중에이전트 시스템[22, 23]을 모델링하고 에이전트들의 지적 및 시간적 특성에 관한 추론을 하기 위하여 번역시스템의 형식화를 설명하고, 지식, 시간, 그리고 전략에 관한 추론을 위한 논리언어인 *ATL*(*Alternating-time Temporal Logic*)과 *ATEL*(*Alternating-time Temporal Epistemic Logic*)을 소개한다. *ATL*논리는 다중에이전트 시스템의 전략적 특성을 표현할 수 있는데, 이 다중에이전트 시스템은 근본적으로 거의 완전한 정보를 갖는 포괄적 게임으로서 모델링이 된다. *ATL*논리는 다중에이전트 시스템의 형식적인 증명에서의 응용과 더불어 *CTL*(*Computation Tree Logic*)의 게임-이론적인 일반 형태이다[5, 9, 10]. *ATL*의 정의에 따르면 각각의 에이전트는 게임구조의 상태에 대하여 완전한 정보를 가진다. 즉, 모든 변수들은 모든 에이전트들에게 접근이 가능하다. 그러나 현실적인 모델은 불완전한 정보와 완전한 기억에 의해 제공된다[4, 15].

게임은 상호작용하는 의사결정자를 위한 모델이다. 이러한 모델은 각 선수들이 그들 자신의 행위뿐만 아니라, 모든 다른 선수들의 행위들에 의해서도 영향을 받도록 함으로써 선수들 사이의 상호작용의 의미를 분석한다. 일반적인 게임을 다루는 문제에 대해 1951년 John Nash가 설명하였는데, 그는 그러한 해법을 평형이라 불렀으며, 최근에 와서 나쉬평형(*Nash equilibrium*)이라 부른다[1, 2, 3].

Van der Hoek과 Wooldridge는 *ATL*에 지식연산자를 확장하여 매우 직관적인 번역시스템 의미론을 제공하였는데, 이로 인하여 에이전트들이 이러한 상태들을 구별할 수 없을지라도, 에이전트들은 항상 다른 상태에서 다른 선택을 할 수 있도록 가정하고 있다. *ATEL*의 원래 번역은 매력적인 계산의 특징을 지니고 있지만, 이러한 반직관적인 특성으로 인하여 여러 추론에 있어서 논리의 정교함을 필요로 한다[13, 14, 16, 17].

ATEL 의미론의 입장에서 이러한 문제를 다시 정의하면, 다음과 같다[17]: 첫째, 하나의 에이전트는 그 자신의 상태에 대해 확실한 정보를 갖지 못하더라도 전략을 구성함에 있어서 전체 시스템의 현재 상태에 접근할 수 있다. 둘째, *ATEL*의 불명확한 행위들에 대한 표현은 일반적인 상황을 모델링하는 것을 어렵게 한다. 따라서 본 논문은 이러한 *ATEL*의 문제들을 해결하기 위해 포괄적 게임에서 *ATEL*의 번역에 대한 하나의 대안인 전략적 제한을 그 해결책으로 제안한다.

II. 논리이론(*Logic Theory*)

1. 번역시스템(*Interpreted Systems*)

번역시스템의 형식주의는 다중에이전트 시스템에서 에이전트들의 지식 및 시간에 관한 특성을 추론하고 에이전트 시스템의 모형을 만들기 위해 도입되었다[7, 8]. 시스템에서 각각의 에이전트 $i(i \in \{1,\ldots,n\})$는 유한 지역 상태들의 집합 L_i와 유한 행위들의 집합 Act_i에 의해 특성을 나타낸다. 행위들은 프로토콜 $P_i : L_i \rightarrow 2^{Act_i}$에 따라 수행되는데, 이 프로토콜은 하나의 주어진 상태에서 어느 행위들을 수행할 것인가를 명시한다. 이 형식주의에서는, 에이전트가 활동하는 환경은 특별한 에이전트 E에 의해 모형이 만들어질 수 있다. E와 관련지어 지역 상태들의 집합 L_E, 행위들의 집합 Act_E, 그리고 프로토콜 P_E를 생각할 수 있다. 에이전트들의 지역 상태들의 데카르트 곱을 $S = L_1 \times \cdot \cdot \cdot \times L_n \times L_E$로 나타낸다. 각각의 i에 대한 $l_i \in L_i$와 $l_E \in L_E$에 대해 $g = (l_1,\ldots,l_n, l_E) \in S$를 하나의 전역상태라 부르는데, 이는 특정 순간에서의 시스템의 상태를 알려준다. 에이전트 i에 대한 전개함수는 기호 t_i 로 나타내는데, 함수 t_i: $L_i \times L_E \times Act_1 \times \cdot \cdot \cdot \times Act_n \times Act_E$로 모델을 만들 수 있다. 다중에이전트 시스템 기술을 완성하기 위해서는 초기전역상태 $I \subseteq S$와 평가관계 $V \subseteq AP \times S$와 더불어 원시명제들의 집합 AP가 등장한다. 따라서 에이전트들의 집합 $\Sigma = \{1,\ldots,n\}$이 주어지면, 하나의 번역시스템은 $IS = \langle (L_i, Act_i, P_i, t_i)_{i \in \Sigma}, (L_E, Act_E, P_E, t_E), I, V \rangle$이다. 번역시스템은 지식 및 시간에 관한 특성을 추론하기 위하여 다음과 같은 언어를 사용하여 의미론을 제공한다[6]:

$\phi ::= p \mid \neg\phi \mid \phi \vee \phi \mid EX\phi \mid EG\phi \mid E(\phi U \psi) \mid K_i\phi \mid E_\Gamma \phi \mid C_\Gamma\phi \mid D_\Gamma\phi$.

이 문법에서, $p \in P$는 원시명제이고, 연산자 EX, EG, 그리고 EU는 표준 CTL 연산자들이다. 하나의 번역시스템 IS가 주어지면, IS에 Kripke 모델 $MIS = (W, R_t, \sim_1,\ldots,\sim_n, V)$를 관련시킬 수 있는데, 이 모델은 위의 문법의 공식들을 번역하는데 사용될 수 있다. 위 언어의 형식적인 의미론은 다음과 같다:

- $MIS, w \models p$, $(p, w) \in V$일 경우,
- $MIS, w \models \neg\phi$, $MIS, w \not\models \phi$일 경우,
- $MIS, w \models \phi_1 \vee \phi_2$, $MIS, w \models \phi_1$ 또는 $MIS, w \models \phi_2$일 경우,

- $MIS, w \models EX\phi$, $\pi(0) = w$ 이고 $MIS, \pi(1) \models \phi$ 인 하나의 경로 π가 존재할 경우,
- $MIS, w \models EG\phi$, $\pi(0) = w$이고 모든 $i = 0$ 에 대해 $MIS, \pi(i) \models \phi$인 하나의 경로 π가 존재할 경우,
- $MIS, w \models E[\phi U \psi]$, $\pi(0) = w$이고, $MIS, \pi(k) \models \psi$이며, 모든 $0 < j < k$에 대하여 $MIS, \pi(j) \models \phi$인 $k \geq 0$ 에 대해 하나의 경로 π가 존재할 경우,
- $MIS, w \models K_i\phi$, 모든 $w' \in W$에 대하여 $w \sim_i w'$ 이 $MIS, w' \models \phi$을 의미할 경우.

2. ATL과 ATEL

ATL은 에이전트들의 집합 A가 $\langle\langle A \rangle\rangle$ 형태의 협동양식의 클래스로 CTL을 확장한 것이다. $\langle\langle A \rangle\rangle \Phi$의 직관적인 번역은 "에이전트 그룹 A는 시스템의 다른 에이전트들이 어떠한 행위를 하든 상관없이 Φ를 강요하는 집단전략을 갖는다"이다. 시간적 논리 ATL의 구문론은 명제들의 유한집합 Π와 선수들의 유한집합 $\Sigma = \{1, \cdots, n\}$에 대하여 정의된다. 하나의 ATL 공식은 다음과 같다[4]:

- p, 명제 $p \in \Pi$에 대하여,
- $\neg\phi$ or $\phi_1 \vee \phi_2$, ϕ, ϕ_1, and ϕ_2 는 ATL 공식들임,
- $\langle\langle A \rangle\rangle X\phi$, $\langle\langle A \rangle\rangle G\phi$, or $\langle\langle A \rangle\rangle \phi_1 U \phi_2$, 여기서 $A \subseteq \Sigma$는 선수들의 집합이고, ϕ, ϕ_1, 그리고 ϕ_2는 ATL 공식들임.

연산자 $\langle\langle\ \rangle\rangle$는 경로한정자이고, X("next"), G("always"), 그리고 U("until")는 시간적 연산자들이다. ATL 공식의 의미론은 ATL과 같은 명제와 선수들(에이전트들)을 갖는 CGS(*Concurrent Game Structures*: 동시게임구조) S로 주어진다. 하나의 동시게임구조는 $S = \langle \Sigma, Q, \Pi, \pi, d, \delta \rangle$인데, Σ는 선수들의 집합이고, Q는 상태들의 집합이며, Π는 원시명제들의 집합이다. 각각의 상태 $q \in Q$에서 참인 명제들의 집합 $\pi(q) \subseteq \Pi$을 분류함수(또는 관찰함수)라고 부른다. d: $\Sigma \times Q \rightarrow N$는 어떤 상태에 있는 선수 i에게 가능한 동작의 자연수이며, δ는 시스템의 전개를 결정하는 전개(또는 전이)함수이다. 계산 S는 전개함수에 의하여 $i = 0$에 대해 각각의 q_i와 q_{i+1}를 연결하는 하나의 무한열 $\lambda = q_0, q_1, q_2, \ldots$이다. 선수 $i \in \Sigma$의 전략은 모든 비공집합 무한열 $\lambda \in Q^+$을 하나의 자연수에 대응하는 함수 f_i인데, λ의 마지막 상태가 q라면, $f_i(\lambda) = d_i(q)$이다. 하나의 상태 q와 선수들의 집합 $A \subseteq \{1, \ldots, n\}$, 그리고 전략들의 집합 $F_A = \{f_i | i \in A\}$이 주어지면, 집합 A에 있는 선수들이 F_A에 있는 전략들을 따를 때 강요하는 q-계산의 집합 $out(q, F_A)$는, $q_0 = q$이고, 모든 위치 $k = 0$에 대하여 $j_i = f_i(\lambda[0,k])$ 이고, $\delta(q_k, j_1, \ldots j_n) = q_{k+1}$인 동작벡터 $\langle j_1, \ldots j_n \rangle \in D(q_k)$이 존재한다면, 하나의 계산 $\lambda = q_0, q_1, q_2, \ldots$이며, $out(q, F_A)$에 존재한다. ATL의 형식 의미론은 다음과 같다:

- $q \models p$, 명제 $p \in \Pi$에 대하여, $p \in \pi(q)$일 경우,
- $q \models \neg\phi$, $q \not\models \phi$일 경우,
- $q \models \phi_1 \vee \phi_2$, $q \models \phi_1$ 또는 $q \models \phi_2$일 경우,
- $q \models \langle\langle A \rangle\rangle X\phi$, 모든 계산 $\lambda \in out(q, F_A)$, $\lambda[1] \models \phi$이고, A에 있는 각각의 선수에 대하여 전략들의 집합 F_A가 존재할 경우,
- $q \models \langle\langle A \rangle\rangle G\phi$, 모든 계산 $\lambda \in out(q, F_A)$이고 모든 지점 $i = 0$에 대하여, $\lambda[i] \models \phi$이고, A에 있는 각각의 선수에 대하여 전략들의 집합 F_A가 존재할 경우,
- $q \models \langle\langle A \rangle\rangle \phi_1 U \phi_2$, 모든 계산 $\lambda \in out(q, F_A)$이고 모든 지점 $i = 0$에 대하여, $\lambda[i] \models \phi_2$이고, $0 \leq j < i$에 대하여 $\lambda[j] \models \phi_1$이고, A에 있는 각각의 선수에 대하여 전략들의 집합 F_A가 존재할 경우.

$ATEL$은 ATL에 지식연산자를 추가하여 확장한 것이다. ATL 공식들이 동시게임구조에 기초하여 번역된 것처럼 $ATEL$ 공식들은 $AETS$(*Alternating Epistemic Transition System*)에 의해 번역된다. $AETS$ $S = \langle \Sigma, Q, \Pi, \sim_1, \ldots, \sim_n, \pi, d, \delta \rangle$는 ATL의 CGS에 지적접근관계, 즉 각각의 에이전트 $a \in \Sigma$에 대해 $\sim_a \subseteq Q \times Q$ 관계가 포함된다. $ATEL$의 공식은 $AETS$ S에 의해 다음 중 하나이다[14]:

- p, 여기서 $p \in \Pi$는 원시 명제들임,
- $\neg\phi$ 또는 $\phi_1 \vee \phi_2$, ϕ_1과 ϕ_2는 $ATEL$의 공식임,
- $\langle\langle A \rangle\rangle X\phi$, $\langle\langle A \rangle\rangle G\phi$, or $\langle\langle A \rangle\rangle \phi_1 U \phi_2$, 여기서 $A \subseteq \Sigma$는 에이전트들의 집합이고, ϕ, ϕ_1, 그리고 ϕ_2는 $ATEL$ 공식들임,
- $K_a\phi$, 여기서 $a \in \Sigma$는 하나의 에이전트이고 ϕ는 $ATEL$의 공식임,
- $E_A\phi$ 또는 $C_A\phi$, 여기서 $A \subseteq \Sigma$는 에이전트들의 집합이고, ϕ는 $ATEL$ 공식임.

앞의 ATL처럼, $AETS$에 대하여 $ATEL$의 공식들을 번역한다. 만족관계 $\models$를 정의하는 법칙들은 ATL의 형식 의미론에 다음을 포함한다:

- $q \models K_a\phi$, $q \sim_a q'$인 모든 q'에 대하여 $q' \models \phi$일 경우,
- $q \models E_A\phi$, $q \sim_A^E q'$인 모든 q'에 대하여 $q' \models \phi$일 경

우.

- $q \models C_A\phi$, $q \sim_A^C q'$인 모든 q'에 대하여 $q' \models \phi$일 경우.

III. 게임이론(*Game Theory*)

1. 전략적 게임(*Strategic Games*)

전략적 게임에서 각각의 에이전트는 게임의 초기에 사용 가능한 수많은 전략들을 가지는데, 각각의 에이전트는 독립적으로 하나의 전략을 선택하여, 행해진 행위들을 자세히 고려함이 없이 게임의 이익을 직접 계산할 수 있다. 하나의 전략적 게임 G는 하나의 $(\Sigma, \{S^X\}_{X\in\Sigma}, U)$인데, 여기서 $\Sigma = \{1, 2,..., n\}$은 에이전트들의 집합이고, 각각의 $X \in \Sigma$에 대하여 S^X는 에이전트 X를 위한 전략들의 집합이며, $U^X : (S^1 \times \cdot\cdot\cdot\cdot \times S^n) \rightarrow R$은 에이전트 X를 위한 효용함수이다. 표기법 $U^X(\vec{s})$는 벡터 $U(\vec{s})$의 X번째 요소를 나타내는데, 전략 프로파일 $\vec{s}$가 사용될 때, 에이전트 X의 효용을 나타낸다. 전략적 게임에서 각각의 에이전트는 그들 효용을 최대화하려고 한다. 나쉬평형의 정의를 위해, 주어진 함수를 최대화하는 모든 입력들을 출력하는 함수 $argmax$는 $argmax_x\ f(x) = \{x \mid \neg\exists y : f(x) < f(y)\}$이다. 함수 b^X는 주어진 게임과 전략 벡터에 대하여 에이전트 X를 위한 최상의 반응 전략을 출력한다. 하나의 게임 $(\Sigma, \{S^X\}_{X\in\Sigma}, U)$과 하나의 전략 프로파일 $\vec{s} \in (\Pi_X, S^X)$에 대해 최상의 반응 $b(\vec{s}) = b^1(\vec{s}) \times \cdot\cdot\cdot\cdot \times b^n(\vec{s})$는 $b^X(\vec{s}) = argmax_t U^X((s_{-X}, t))$로 정의된다. 최상의 반응 함수에서 고정점들을 찾을 수 있는데, 이를 나쉬평형(*Nash equilibrium*)이라 부른다[1, 2].

2. 포괄적 게임(*Extensive Games*)

포괄적 게임에서는 에이전트들은 최종적으로 결과에 도달하는 일련의 선택을 하게 되는데, 그 과정에 다중의 결정점이 있으며 각각의 결정점에서 에이전트들 중 하나는 다음에 무엇을 할 것인지를 결정해야 한다.

2.1 완전한 정보 게임(*Perfect Information Games*)

하나의 게임형식 F는 하나의 $F = (\Sigma, H, turn)$인데, Σ는 에이전트들의 유한집합이고, H는 유한열집합이며, $turn$은 함수 $turn : H \backslash Z(H) \rightarrow \Sigma$이다. 하나의 포괄적 게임 F는 하나의 $F = (\Sigma, H, turn, U)$인데, 여기서 $(\Sigma, H, turn)$는 하나의 게임형식이고, $U : Z(H) \times \Sigma \rightarrow R^{\Sigma}$이다. 게임형식 $(\Sigma, H, turn)$에서 에이전트 X의 순수전략 σ는 $\sigma(h) = A(H, h)$이고, 영역 $\{h \in H \mid turn(h) = X\}$인 함수 σ이다.

하나의 포괄적 게임의 나쉬평형은 전략적 게임에 해당하는 하나의 나쉬평형으로 정의될 수 있다. 예를 들면, 그림 1에 하나의 간단한 포괄적 게임이 있다. 이 게임에서 두 학생, 엘리스(A)와 밥(B)이 그들의 공동 사무실을 청소(행위 c)하거나 더러운 상태를 무시(행위 i)한다. 엘리스가 먼저 도착하기 때문에 그녀는 먼저 무엇을 할 것인가를 결정해야 한다. 만약 그녀가 사무실을 청소하지 않는다면, 밥이 똑 같은 선택을 하게 된다. 깨끗한 사무실은 2의 효용단위에 해당하지만, 사무실 청소는 에이전트에게 1의 효용단위를 비용으로 요구한다. 각각의 에이전트는 다음 표 1에 나타난 바와 같이, 각각 σ_c와 σ_i라 불리는 순수전략을 갖는다.

표 1. 2인 선수 게임의 예
Table 1. An example of two-player game

$A \setminus B$	σ_c	σ_i
σ_c	(1, 2)	**(1, 2)**
σ_i	**(2, 1)**	(0, 0)

이 게임에서 두 개의 나쉬평형이 굵은 숫자로 나타나 있다. 하나의 포괄적 게임의 각각의 결정노드는 하나의 작은 포괄적 게임의 출발점으로 볼 수 있다. 그러한 하나의 게임은 원래 게임의 서브게임(*subgame*)이라 부른다.

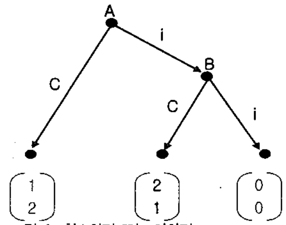

그림 1. 청소하기 또는 안하기 : 포괄적 게임
Fig. 1. To Clean or not to clean : An extensive game

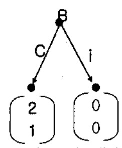

그림 2. 서브게임 : 완전한 나쉬평형
Fig. 2. A subgame : Perfect equilibrium

각각의 에이전트는 각각의 서브게임에서 합리적으로 행동하는데, 이러한 평형의 특성을 서브게임 완전한 나쉬평형이라 부른다. 그림 2에서 하나의 서브게임 완전한 나쉬평형의 예를 볼 수 있다. $G = (\Sigma, H, turn, U)$이 하나의 포괄적 게임이고, f는 포괄적 게임을 전략적 게임으로 축소하는 하나의 함수일 때, G의 모든 서브게임들 G'에 대해, $\vec{s}$가 $f(G')$의 나쉬평형이면, 전략 프로파일 $\vec{s}$는 하나의 서브게임 완전한 평형이다. 그림 1의 예제 게임은 에이전트 B가 사무실을 청소하는 하나의 서브게임 완전한 나쉬평형을 갖는다.

2.2 불완전한 정보 게임(*Imperfect Information Games*)

불완전한 정보게임은 하나의 에이전트 X가 내력(*history*) h와 h' 사이의 차이를 볼 수 없는 경우를 말한다. 그래서 에이전트 X가 h상황에 있을 때, 에이전트 X는 h'상황에도 있을 수 있다고 생각한다. 이러한 정보를 저장하기 위해 동치관계 $\sim_X$를 사용하며, 이러한 정보의 부족을 나타내기 위해 $h \sim_X h'$를 사용한다. 예를 들면, 포커게임의 경우, 에이전트 X는 상대 Y가 갖고 있는 카드들을 모른다. 만약 h와 h'사이의 유일한 차이는 Y의 손에 들어 있는 카드들이라면, $h \sim_X h'$이다. 그림 3은 불완전한 정보 게임형식의 하나의 예이다. 이 그림에서 정보의 부족은 점선으로 표시되어 있다. 따라서 에이전트 B는 내력 y와 n을 구별할 수 없으며, 에이전트 A가 선택한 행위에 대한 정보를 갖지 못 한다: $y \sim_B n$. 에이전트 X가 구별할 수 없는, 즉 $h \sim_X h'$인 상황 h와 h'에서 결정을 내려야 할 경우가 종종 발생하는데, X가 이러한 상황을 구별할 수 없기 때문에, X에 대한 전략들은 두 경우 모두에 대해 똑 같은 행위를 내려야만 한다. 만약 모든 $h \sim_X h'$에 대해 $\sigma(h) = \sigma(h')$을 만족한다면, 게임형식 $F = (\Sigma, H, turn, \sim, \Pi)$에서 X에 대한 전략은 일정하다(*uniform*)〔2, 12〕

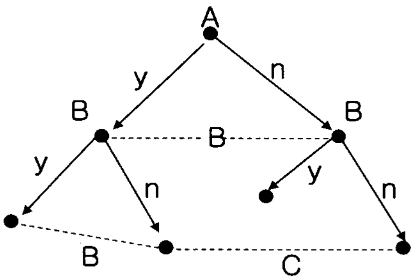

그림 3. 불완전한 정보 게임형식
Fig. 3. An imperfect information game form

IV. *ATEL*의 문제점 해결을 위한 에이전트들의 전략적 제한(*Strategic Constraints: SC*)

I장의 *ATEL*의 문제점 해결을 위한 방법의 하나인 Γ-일정한 번역시스템이 제안되었지만, 가능한 경우의 수가 최대 $\Pi_{i \in \Gamma} |Act_i|^{|Li|}$ 개까지 존재할 수 있어서 상태의 개수를 줄일 수 없다〔9〕. 따라서 이 장에서는 *ATEL*의 문제점들의 해결책으로 포괄적 게임과 번역시스템과의 관계를 살펴보고, 에이전트들에 대한 전략적 제한을 그 해결책으로 제시하고, 또한 전략적 제한에 기초한 번역시스템을 모델체킹의 한 가지 방법으로 제시한다.

1. 포괄적 게임과 번역시스템과의 관계

먼저 불완전한 정보를 갖는 포괄적 게임과 번역시스템과의 관계를 살펴본다. 불완전한 정보를 갖는 포괄적 게임에서 서브게임 완전한 평형은 일정한 전략(*uniform strategies*)을 제공함은 물론, 해당 서브게임 내에 존재하는 에이전트들의 역할은 그 서브게임 내로 한정되고, 그 서브게임 밖의 어떠한 에이전트들도 해당 서브게임 내의 에이전트들에게 영향을 미칠 수 없다. 이러한 특성은 번역시스템에서 하나의 에이전트는 다만 자신의 지역 상태만을 인지하고 있음과 매우 유사함을 알 수 있다. 또 다른 하나의 유사한 관계는 전략에 관한 정의이다. 포괄적 게임에서의 전략은 에이전트 i가 게임을 어떻게 진행할 것인가, 즉 각각의 결정점에서 다음 행위를 어떻게 취할 것인가에 관한 계획이다〔18〕. 이는 *ATEL*에서의 하나의 전략은 상태들의 열에 대한 하나의 행위로의 함수로 정의됨과 유사하다. 또한 번역시스템에서 전략은 프로토콜로 등장하는데 이는 하나의 지역 상태에서 에이전트 i에 대한 하나의 행위로의 함수로 정의된다. 이러한 관점에서 전략과 프로코콜은 행위와 상태를 연결한다는 점에서 매우 유사하다〔9〕.

2. 전략적 제한(*Strategic Constraints: SC*)

따라서 포괄적 게임과 번역시스템과의 유사한 관계는 불완전한 정보를 갖는 포괄적 게임에서 서브게임 완전한 나쉬평형이 반직관적인 특성으로 인한 추론에서의 논리의 정교함을 필요로 하는 *ATEL*의 문제점들을 해결할 수 있는 일정한 전략을 제공한다. 즉, 다양한 응용분야에서 양호한 행위들에 약간의 제한요소(전략적 제한)를 부가하는데, 예를 들면, 불완전한 정보를 갖는 경우의 구별할 수 없는 상태에서 항상 같은 선택을 명시하는 일정한 전략을 고려하거나 또는 에이전트들의 나쉬평형을 발전시킨 서브게임 완전한 나쉬평형의 전략 프로파일을 진행한다(그림 4. 참조)〔1, 19〕.

예를 들면, 그림 4와 그림 5는 참가게임을 나타내고 있는데, 현직자는 도전자의 참가 가능성에 대비한다. 도전자는 참가를 하거나 안할 수 있다. 만약 참가하게 되면, 현직자는 동의를 하거나 싸움을 해야 한다. 도전자는 세 가지 선택을 할 수 있는데, 밖에 머물거나, 싸움을 준비하여 참가하거나, 준비함이 없이 참가할 수 있다. 현직자는 도전자가 준비 안 된 상태에서 참가할 때 싸우는 것이 비용이 적게 든다. 현직자는 도전자가 참가하는 것은 알 수 있지만, 도전자가 싸움을 준비했는지는 알 수 없다. 다시 말해, 불완전한 정보를 갖는 포괄

적 게임이 이 상황을 모델링할 수 있다.

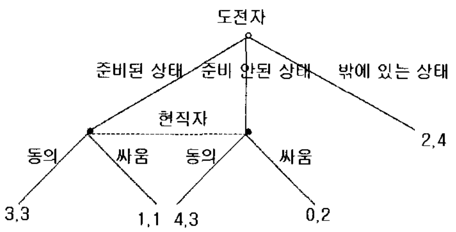

그림 4. 참가게임의 포괄적 게임형식
Fig. 4. The extensive game of an entry game

	동의	싸움
준비된 상태	3, 3	1, 1
준비 안된 상태	4, 3	0, 2
밖에 있는 상태	2, 4	2, 4

그림 5. 참가게임의 전략적 게임형식
Fig. 5. The strategic game form of an entry game

그림 5에서 이 게임은 두 개의 나쉬평형, 즉 (준비 안 된 상태, 동의)와 (밖에 있는 상태, 싸움)을 가지는데, 서브게임 완전한 나쉬평형에 의해, (밖에 있는 상태, 싸움)은 나쉬평형에 해당되지 않음을 확인할 수 있다. 따라서 (준비 안 된 상태, 동의)만이 나쉬평형에 해당되며, 현직자는 이 전략을 최적의 전략으로 선택하게 된다.

이러한 에이전트들의 전략적 제한을 에이전트들의 전략적 프로파일로 구현하는데, 이들은 나쉬평형의 특성을 가지며, 에이전트들의 행위에 대한 요구사항은 전략적 프로파일의 집합으로 모델링이 된다. 따라서 다음과 같이 에이전트들에 대한 전략적 제한을 정의할 수 있다:

하나의 전략적 제한, $\eta = \langle \vec{s}, A \rangle$,

$\vec{s} \subseteq \Pi_X S^X$: 서브게임 완전한 나쉬평형의 특성을 갖는 전략적 파일의 비공집합,

$A \subseteq \Sigma$: 전략적 제한이 적용될 에이전트들의 집합.

즉, 전략적 제한 η에 대해, 주어진 지역 상태에서의 행위들을 선택함에 있어서 적어도 일정한 번역시스템에 대한 추론을 하는 것이 유용하다.

이러한 전략적 제한에서의 *ATEL*의 형식 의미론은 다음과 같이 정의될 수 있다:

- $q, \eta \models p$, 명제 $p \in \Pi$에 대하여, $p \in \pi(q)$일 경우,
- $q, \eta \models \neg\phi$, $q, \eta \not\models \phi$일 경우,
- $q, \eta \models \phi_1 \vee \phi_2$, $q, \eta \models \phi_1$ 또는 $q, \eta \models \phi_2$일 경우,
- $q, \eta \models \langle\langle A \rangle\rangle X\phi$, 모든 계산 $\lambda \in out(q, F_A)$, $\lambda[1], \eta \models \phi$이고, A에 있는 각각의 선수에 대하여, 전략적 제한 η에 따르는 전략들의 집합 F_A가 존재할 경우,
- $q, \eta \models \langle\langle A \rangle\rangle G\phi$, 모든 계산 $\lambda \in out(q, F_A)$이고 모든 지점 $i = 0$에 대하여, $\lambda[i], \eta \models \phi$이고, A에 있는 각각의 선수에 대하여, 전략적 제한 η에 따르는 전략들의 집합 F_A가 존재할 경우,
- $q, \eta \models \langle\langle A \rangle\rangle \phi_1 U \phi_2$, 모든 계산 $\lambda \in out(q, F_A)$이고 모든 지점 $i = 0$에 대하여, $\lambda[i], \eta \models \phi_2$이고, $0 \le j < i$에 대하여 $\lambda[j], \eta \models \phi_1$이고, A에 있는 각각의 선수에 대하여, 전략적 제한 η에 따르는 전략들의 집합 F_A가 존재할 경우,
- $q, \eta \models K_a\phi$, $q \sim_a q'$인 모든 q'에 대하여 $q', \eta \models \phi$일 경우,
- $q, \eta \models E_A\phi$, $q \sim_A^E q'$인 모든 q'에 대하여 $q', \eta \models \phi$일 경우,
- $q, \eta \models C_A\phi$, $q \sim_A^C q'$인 모든 q'에 대하여 $q', \eta \models \phi$일 경우.

3. 전략적 제한에 기초한 번역시스템의 모델체킹

전략적 제한에 기초한 번역시스템을 IS_η라고 표기할 때, 전략적 제한에 기초한 번역시스템 IS_η이 기존의 번역시스템과 양립하기 위해서는, 먼저 같은 지역 상태들의 집합, 행위 및 전개함수를 가지는 같은 에이전트들이 포함되어야 하고, IS_η에 있는 A 에이전트들의 프로토콜 P_i^A는 IS의 프로토콜 P_i에 포함되어야 하는데, 이는 모든 지역 상태 $l_i \in L_i$에 대하여 $P_i^A \in P_i$, 즉 하나의 행위만이 P_i^A에서 선택되어야 하고, 또한 그 행위는 P_i에 명시된 행위들 중 하나이다. 주어진 번역시스템 IS와 양립하는 전략적 제한에 기초한 번역시스템을 IS_η들의 집합을 $\{IS\}_\eta$로 표기한다. 일반적인 경우와 마찬가지로, $\{IS\}_\eta$에 있는 전략적 제한에 기초한 번역시스템에 Kripke 모델을 연결하여, 하나의 공식 ϕ가 전략적 제한에 기초한 번역시스템의 하나의 클래스에서 참이고, 하나의 공식 ϕ이 $\{IS\}_\eta$에 있는 전략적 제한에 기초한 번역시스템을 IS_η과 연결되는 하나의 모델이라면 $IS \models_\eta \phi$과 같이 표기할 수 있다. 다음 그림 6은 다중에이전트 시스템을 위한 모델체킹 도구 중의 하나인 *MCMAS(Model Checking Multi-Agent Systems)*에서 $S \models_\eta \phi$이 성립하는지를 증명하기 위한 모델체킹 프로시저이다[20, 21].

```
MC-SC(IS, A, φ) {
  전략적 제한 번역시스템 집합 {IS}_n을 계산한다
  for each (X ∈ {IS}_n) {
   if (MCCTLK(φ, X) == G) return true;
  }
  return false;
}
```

그림 6. 전략적 제한에 기초한 번역시스템의 모델체킹 프로시저
Fig. 6. Procedure for model checking of Interpreted Systems based on strategic constraints

*ATEL*의 문제점 해결을 위한 Γ-일정한 번역시스템에서는, 하나의 주어진 번역시스템과 하나의 그룹 에이전트들 Γ에 대하여, 번역시스템과 호환이 되는 여러 개의 Γ-일정한 번역시스템이 최대 $\Pi_{i\in\Gamma}|Act_i|^{|L_i|}$ 개까지 존재할 수 있다[6, 9]. 이에 비하여, 본 논문에서 제안한 전략적 제한에 기초한 번역시스템은, 서브게임 완전한 나쉬평형 알고리즘에 따라, 총 $m=\Sigma_A|Act_i|$개의 행위들을 갖고 길이가 2인 그래프 구조를 갖는 게임을 생각하면, 최대 $(m/3)*2^{m/3}$개의 비교가 이뤄진다. 따라서 본 논문에서 제안한 전략적 제한에 기초한 번역시스템이 Γ-일정한 번역시스템보다 효율적이다. 하나의 예를 들면, [9]의 E에 대한 행위를 $(a^1_E,\ a^2_E)$라고 가정하면 표 2와 같다.

표 2. 2인 선수 게임에 대한 전략적 제한에 기초한 번역시스템과 Γ-일정한 번역시스템의 비교
Table 2. Comparison between the *IS* based on strategic constraints and Γ-uniform *IS* for 2-player games

번역시스템	상태의 개수($\|L_i\|=2$)	상태의 개수($\|L_i\|=3$)
IS_n	8	8
$IS_{\Gamma\text{-uniform}}$	32	64

한편, 서브게임 완전한 나쉬평형의 개념은 다중에이전트 시스템의 중요한 성질 중 하나인 에이전트들의 합리성(*rationality*)을 구현하기 위한 하나의 방법으로도 사용될 수 있다. 게임이론의 관점에서 수많은 게임들은 모든 가능한 결과들의 경우의 수는 무한하며, 그 중 일부는 적당하다. 서브게임 완전한 나쉬평형과 같은 합리성의 이론은 다소 부적당한 결과들을 제거하고, 이상적인 선수들이 게임을 진행하는 경우, 무엇을 해야 할지를 결정하게 한다.

V. 결 론

다중에이전트 시스템의 분석에 중요한 시간, 지식, 그리고 전략에 관한 개념들을 구체화하기 위해 번역시스템과 *ATL* 및 *ATEL*을 살펴보고, 특히 *ATEL*의 문제점을 해결하기 위한 방안으로 게임이론의 나쉬평형에 기초한 전략적 제한을 문제의 해결책으로 제안하였다. 뿐만 아니라, 전략적 제한에 기초한 번역시스템을 다중에이전트 시스템에서의 모델체킹을 위한 하나의 방법으로 제안하였다.

참고문헌

[1] Martin J. Osborne, "*An Introduction to Game Theory,*" Oxford University Press, 2004.

[2] Sieuwert van Otterloo, "*A Strategic Analysis of Multi-agent Protocols,*" Ph.D. thesis, University of Liverpool, 2005.

[3] Stuart Russel, Peter Norvig, "*Artificial Intelligence-A Modern Approach,*" Prentice Hall, 2003.

[4] Rajeev Alur, Thomas A. Henzinger, and Orna Kupferman, "Alternating-Time Temporal Logic," Journal of the ACM, Vol. 49, No.5, pp.672-713, Sept. 2002.

[5] V. Goranko, W. Jamroga, "Comparing Semantics of Logics for Multi-agent Systems," Synthese 00: pp.1-41, 2004.

[6] Franco Raimondi, "*Model Checking Multi-Agent Systems,*" Ph.D. thesis, University of London, 2006.

[7] R. Fagin, J.Y. Halpern, Y. Moses, M.Y. Vardi, "*Reasoning about Knowledge,*" MIT Press, Cambridge, MA,1995.

[8] Alessio Lomuscio, Marek Sergot, "Deontic Interpreted Systems," Studia Logica, 75(1), pp.63-92, 2003.

[9] Alessio Lomuscio, Franco Raimondi, "Model Checking Knowledge, Strategies, and Games in Multi-Agent Systems," AAMAS, '06.

[10] Wojciech Jamroga, Wiebe van der Hoek, and Michael Wooldridge, "On Obligations and Abilities," DEON 2004, LNAI 3065 , pp.165-

181, 2004.

[11] Clark, Grumberg, and Peled, "*Model Checking*," The MIT Press, 1999.

[12] P. Blackburn, J. Van Benthem, and F. Wolter, "*Handbook of Modal Logic*," Elsevier, 2007.

[13] Sieuwert van Otterloo, Geert Jonker, "On Epistemic Temporal Strategic Logic," Electronic Notes in Theoretical Computer Science 126, pp.77-92, 2005.

[14] Wiebe van der Hoek, Michael Wooldridge, "Tractable Multiagent Planning for Epistemic Goals," Proceedings of the 1st International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2002), pp.1167-1174, 2002.

[15] Pierre-Yves Schobbens, "Alternating-time logic with imperfect recall," Electronic Notes in Theoretical Computer Science 85, pp.1-12, 2004.

[16] G. Jonker, " Feasible strategies in alternating-time temporal epistemiclogic," Universiteit Utrecht Master Thesis, 2003.

[17] W. Jamroga, W. van der Hoek, "Some remarks on alternating-time temporal epistemic logic," Proceedings of the International Workshop on Formal Approaches to Multi-Agent Systems (FAMAS '03), pp.133-140, 2004.

[18] Paul Harrenstein, Wiebe van der Hoek, John-Jules Meyer, and Cees Witteveen, "On Modal Logic Interpretations of Games," Proceedings of ECAI, 2002.

[19] Wojciech Jamroga, "Knowledge and Strategic Ability for Model Checknig: A Refined Approach," IfI Technical Report Series IfI-07-11, 2008

[20] A. Lomuscio, H. Qu, F. Raimondi, "MCMAS: A model checker for the verification of multi-agent systems," Proceedings of the 21th International Conference on Computer Aided Verification (CAV 2009). Grenoble, France. Lecture Notes in Computer Science. Vol 5643 pp. 682-688. Springer.

[21] F. Raimondi and A. Lomuscio. "MCMAS - A tool for verification of multi-agent systems," 2009. *http://www.cs.ucl.ac.uk/staff/f.raimondi/MCMAS/*.

[22] 이근수, 윤선희, "멀티 에이전트를 이용한 인터넷 채용 협상 시스템의 구현," 한국 컴퓨터정보학회 논문지, 제11권, 제2호, 341-349쪽, 2006년 5월.

[23] 우종우, 김대령, "멀티 에이전트 기반의 지능형 시뮬레이션 도구의 개발," 한국 컴퓨터정보학회 논문지, 제12권, 제6호, 21-30쪽, 2007년 12월.

저 자 소 개

구 자 록

1989년 7월 : 서울대학교 전산과학과 (박사과정수료)

1989년 3월 ~ 현재 : 울산대학교 컴퓨터정보통신공학부 교수

관심분야 : 다중에이전트 시스템, 전자상거래, 산업현장자동화시스템