

# DNA 코딩 방법을 이용한 사용자의 행위를 학습하는 에이전트 모델

윤효근\*, 이상용\*\*

\*공주대학교 컴퓨터공학과

\*\*공주대학교 정보통신공학부

e-mail: {kosher\*,sylee\*\*}@kongju.ac.kr

## The Model of an Agent to learn Users' Action using DNA Coding Method

Hyo-gun Yun\*, Sang-yong Lee\*\*,

\*Dept. of Computer Engineering, Kongju National University,

\*\*Division of Information & Communication Engineering,  
Kongju National University

### 요 약

현재 에이전트는 강화 학습 모델을 토대로 사용자의 간접 없이 사용자의 의도를 파악하며 능동적으로 행동하는 기술들이 발달되어 왔다. 하지만 인터넷을 기반으로 한 계획이나 학습 등을 위하여 보다 지적인 능력을 갖춘 에이전트의 기술이 요구된다.

따라서 본 논문에서는 DNA 코딩 기법을 이용하여 사용자의 프로파일을 학습하고, 사용자를 분류하는 AUA(Agent for learning Users' Action)를 제안하고자 한다. AUA는 사용자 학습 에이전트로 사용자의 행위를 관찰하고 행위서열을 생성하고 구분함으로써, 사용자의 관심정도를 보다 세밀하게 분석하고 계획할 수 있다. 또한 AUA는 에이전트간에 관계를 설정함으로써 사용자에게 보다 나은 정보 검색을 지원할 수 있다.

### 1. 서론

인터넷의 발달은 기하급수적으로 늘어난 많은 정보를 사용자가 직접 분석하는 부담을 안고 있다. 따라서 사용자의 부담을 덜어주기 위해서 자동화된 에이전트 기술이 등장하게 되었다. 현재 에이전트는 사용자의 행위를 바탕으로 사용자의 성향을 학습시키며 학습 에이전트를 중심으로 발전하였다. 이러한 에이전트의 학습 모델은 기호적 학습, 비기호적 학습, 확률적 학습, 강화학습 등으로 구분할 수 있다 [1].

특히 주어진 환경과 상호작용에 의한 강화학습은 학습 에이전트를 설계하는데 있어 다음과 같은 문제점을 가지고 있다. 첫째, 현재 사이트의 내용을 학습 에이전트가 사용할 수 있는 환경으로 재조정해야 한다. 둘째, 쿠키를 활용하거나 회원등록을 통해 사용자를 확인하는 과정이 필수적이다. 이러한 문제점들은 사용자들이 여러 번 접속하거나 긴 시간동안 사이트를 사용한 후에야 사용자들에게 의미있는 서비

스를 제공할 수 있기 때문이다[2][3].

따라서 본 논문에서는 이러한 문제점을 해결하기 위해 DNA 코딩 방법을 사용하여 짧은 시간동안 에이전트가 환경을 재조정하고 사용자 정보를 분석할 수 있는 학습모델인 AUA를 제안한다.

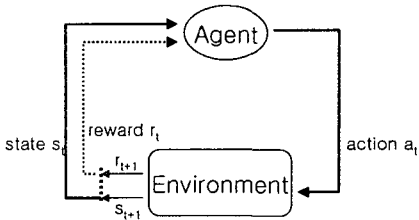
AUA는 강화학습의 문제를 보강하기 위해 DNA 코딩 방법을 사용하였으며, 에이전트가 사용자 행위를 관찰하고 행위서열을 생성하도록 했다[4]. 행위서열은 동적인 사용자의 행위를 쉽게 분석하고, 웹 환경을 쉽게 재조정할 수 있다. 또한 사용자에게 맞는 의미정보들을 빠르게 제공할 수 있다. 본 논문에서 사용한 DNA 코딩 방법은 긴 염기 서열을 짧은 시간 내에 분석할 수 있어 사용자의 행위를 빠르게 압축하고, 분석할 수 있다[4]. 또한 계속 축적된 사용자의 정보는 효과적으로 재구성되어, 사용자에게 알맞은 정보를 서비스할 수 있다.

### 2. 관련연구

2.1 강화학습의 환경 모델

학습자가 주어진 환경과 상호 작용을 할 때 상태 (state), 행동(action), 보상(reward)이라는 세 가지 기본 구조를 이용한다. 환경은 주로 상태를 표현하며 학습자는 적절한 정책에 따라 행동을 취하게 된다. 이 때, 환경은 행동에 대한 보상을 학습자에게 주게 된다.

[그림 1]과 같이 강화학습에서 학습 에이전트는 t 시간동안에 행동 $a_t$ 를 취하면 행동에 대한 보상  $r_t$ 가 환경으로부터 주어진다. 그리고 행동에 의해서 상태가  $s_t$ 가  $s_{t+1}$ 로 변화된다. 학습에서 환경 모델은 마코프 속성을 만족하는 MDP(Markov Decision Process)이다. 마코프 속성은 시간 t+1의 환경에서 반응은 오직 시간 t의 상태와 행동에만 의존하게 되는 속성이다[9].



[그림 1] 강화 학습에서의 환경 모델

1) 보상함수(Reward Function)

보상 함수는 학습자의 행동에 대한 환경으로부터 반응을 스칼라(scalar)값으로 표현한다. 학습에 주어지는 보상은 지시적인(instructive)특성을 갖는 것이 아니라 평가적인(evaluative)특성을 갖고 있으며 식1과 같이 표현한다.

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \dots \dots \dots (식1)$$

식1에서  $\gamma$ 는 할인 상수(discount factor)로 미래에 받게 될 보상이 현재의 상태의 가치나 상태-행동의 가치에 반영되는 값을 조절한다. 즉  $\gamma$ 의 값이 1에 가까울수록 t 시간 이후에 받게 될 보상을 그대로 반영하게 되는 것이다[10].

2) 가치 함수(Value Function)

학습이 계속되는 동안 누적되는 보상과 현재의 상태에 대한 가치를 평가하는 함수로 일반적인 가치함수는 식2와 같이 표현한다.

$$V(s) = V(s) + \alpha [V(s') - V(s)] \dots \dots (식2)$$

가치 함수는 현재 상태의 가치를 평가하여 다음 상태의 가치를 추정하여 학습하는 규칙이다. 여기에서  $\alpha$ 는 현재 상태의 변화량 즉, 학습률을 조절하는 상수이다. 학습 목표에 가깝게 추정된 값  $V(s_{t+1})$

와 현재 상태의 값의 차이를  $\alpha$ 만큼씩 현재의 상태를 변화시켜 가는 형식으로 학습 목표를 달성한다. 구체적으로 식3과 같이 상태에 대한 가치 함수식 또는 식4와 같은 상태-행동 쌍에 대한 가치 함수식으로 나뉘어지며, 이것은 t+1 이후에 시간에 대한 기대값으로 표현된다.

$$V^\pi(s) = E_\pi \left[ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s \right] \dots \dots (식3)$$

$$Q^\pi(s, a) = E_\pi \left[ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a \right] \dots (식4)$$

또한  $\pi$ 는 학습자의 행동 선택에 기준이 되는 정책을 말하며, 각 상태의 가치 또는 어떤 상태에서 어떤 행동을 선택했을 때 가치를 계산하는 것을 반복하게 되면 결국 최종 목적에 수렴하게 된다[11].

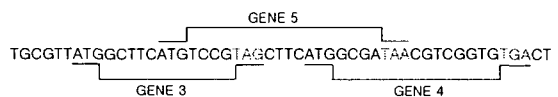
2.2 DNA 코딩 방법

DNA 코딩 방법은 1995년 Yoshikawa가 제시한 변형된 형태의 유전자 알고리즘으로 DNA를 이용한 선택, 재생, 교배, 돌연변이 연산자를 사용한다[12]. DNA는 아미노산 번역표에 따라 20개의 아미노산으로 해석된다. 하나의 아미노산으로 해석되기 위해서는 3개의 염기 서열이 필요하며, 이것을 생물학적 용어로 코돈(codon)이라 한다.

첫 문자 (5 염기)	둘째 문자				셋째 문자 (3 염기)
Y	U	C	A	G	Y
U	UUU Phe	UCU Leu	UAU Tyr	UGU Cys	U C A G
	UUC Phe	UCC Leu	UAC Tyr	UGC Cys	
	UUA Leu	UCA Leu	UAA Stop	UGA Stop	
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U C A G
	CUC Leu	CCC Pro	CAC His	CGC Arg	
	CUA Leu	CCA Pro	CAA His	CCG Arg	
A	AUU Ile	AAC Asn	AUA Ile	AUG Met	U C A G
	AUC Ile	AAC Asn	AUA Ile	AUG Met	
	AUA Ile	AAC Asn	AUA Ile	AUG Met	
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U C A G
	GUC Val	GCC Ala	GAC Asp	GGC Gly	
	GUA Val	GCA Ala	GAA Asp	GGG Gly	

[그림 2] DNA 코드표

[그림 2]의 DNA 코드표를 사용해서 염기서열을 Start 코돈인 ATG에서 시작하여 Stop 코돈인 TGA(TAA, TAG)에서 번역이 끝나며, 이처럼 염기 서열을 짧은 아미노산 코드로 변환하기 때문에 많은 정보를 얻을 수 있다.



[그림 3] 유전자의 중복의 예

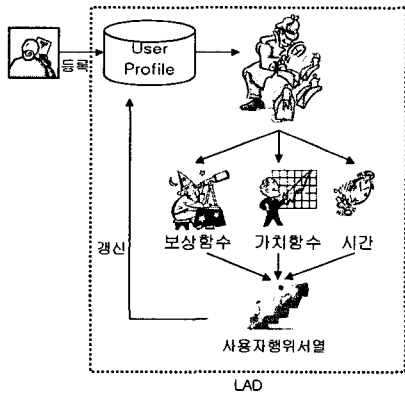
DNA 코딩 방법의 특징을 살펴보면 첫째, [그림 3]에서

보는 것처럼 염색체의 중복을 효율적으로 표현할 수 있다. 둘째로 하나의 아미노산을 만드는 코돈이 여러 개이므로 지식 표현이 쉽다. 셋째로 교차점이 임의로 주어지기 때문에 염색체의 길이가 가변적이다.

이러한 특징들로 인해 긴 길이의 염기 서열이 아닌 적은 수의 아미노산 서열을 사용할 수 있고, 0과 1을 사용하는 유전자 알고리즘에 비하여 DNA 코딩 방법은 4가지 염기를 사용하여 코딩하기 때문에 해의 표현이 다양하다.

3. AUA

웹 상의 사용자 행동 중에서 중요한 것은 특정한 페이지를 보는 시간과 인쇄한 페이지, 전자 상거래 사이트에서의 경우 구매한 상품과 쇼핑 카트에 넣은 상품 등을 들 수가 있다. 이러한 사용자의 행동을 관찰하기 전에 선행되어야 할 것은 사용자에 대한 정보가 일정한 기준에 따라 분류되어 있어야 한다는 것이다. 즉 사용자에 대한 정보(meta-data)가 제공될 내용과 함께 데이터베이스화되어 있어야 한다는 것이다.



[그림 4] AUA의 처리 순서도

따라서 본 논문에서는 [그림 4]에서 같이 AUA를 통하여 웹상에서의 사용자 활동을 관찰하고, 사용자의 관심정도를 평가하여 지속적으로 데이터베이스를 갱신시킨다.

AUA의 처리순서를 살펴보면, 초기 사용자 프로파일은 사용자가 로그인을 하기 위한 간단한 정보만이 요구된다. 로그인된 사용자의 행위를 관찰하여 보상함수, 가치함수, 관심도를 보인 시간을 체크하여 사용자 행위 서열을 생성하고, 서열에 맞춰 사용자의 프로파일을 지속적으로 갱신하도록 한다.

먼저 관찰된 정보로부터 식5와 식6을 이용하여 각 독립된 보상함수(P)를 얻는다.

$$P(a \leq X \leq b, c \leq Y \leq d, e \leq Z \leq f) = P(a \leq X \leq b) \cdot P(c \leq Y \leq d) \cdot P(e \leq Z \leq f) \quad (식5)$$

X는 사용자의 행동, Y는 상태, Z는 보상으로 사용자에게 대한 환경조건을 나타낸다.

$$f(x, y, z) = f_x(x)f_y(y)f_z(z) = \begin{cases} \lambda_1 e^{-\lambda_1 x} \cdot \lambda_2 e^{-\lambda_2 y} \cdot \lambda_3 e^{-\lambda_3 z} & x > 0, y > 0, z > 0 \\ 0 & x \leq 0 \text{ 또는 } y \leq 0 \text{ 또는 } z \leq 0 \end{cases} \quad (식6)$$

가치함수(E)는 주어진 식7를 이용하여 얻는다.

$$E[h(X, Y, Z)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y, z) f(x, y, z) dx dy dz \quad \dots (식7)$$

시간 T는 사용자 웹페이지에서 머무른 시간을 최대 로 설정하여 사용자가 로그아웃할 때까지 지속적으로 관찰하여 벡터 형태의 서열로 표현한다.

$$V = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad \dots (식8)$$

이렇게 지속적인 관찰을 통하여 얻은 서열들을 가지고 사용자의 취향을 식를 통하여 얻을 수 있다.

4. 실험 및 분석

실험에서는 HMM(Hidden Markov Model)을 이용하여 사용자의 행위를 관찰하고 행위에 대한 보상함수와 가치함수, 시간을 측정하였다. 표1은 특정시간대의 사용자 행위를 관찰한 임시 사용자 프로파일이다.

[표 1] 사용자의 행위 관찰 예

시간	시버	경로	상태
10:15:30	ai.kongju.ac.kr	/zboard/./jeje09013xray/images/search1.gif	200
10:15:30	ai.kongju.ac.kr	/zboard/images/n_search.gif	200
10:15:30	ai.kongju.ac.kr	/zboard/images/n_information.gif	200
10:15:30	ai.kongju.ac.kr	/zboard/images/n_mail.gif	200
10:15:35	ai.kongju.ac.kr	/zboard/zboard.php	200
10:15:35	ai.kongju.ac.kr	/zboard/./jeje09013xray/images/i_delete.gif	200
10:15:54	ai.kongju.ac.kr	/board.htm	200
10:15:54	ai.kongju.ac.kr	/board_menu.htm	200
10:15:55	ai.kongju.ac.kr	/zboard/zboard.php	200
10:15:55	ai.kongju.ac.kr	/zboard/images/n_homepage.gif	200
10:15:59	ai.kongju.ac.kr	/people.htm	200
10:15:59	ai.kongju.ac.kr	/people_menu.htm	200
10:15:59	ai.kongju.ac.kr	/undefined	404
10:16:00	ai.kongju.ac.kr	/images/5_ov.gif	200

T={0.0, 0.05, 3, 10, ...}, P={3,4,4,6, ...}, E={2,2,5,4, ...}의 값을 갖는다. 이값을 DNA 코딩 방법에 적용하여 표현한 것이 [표2]이다.

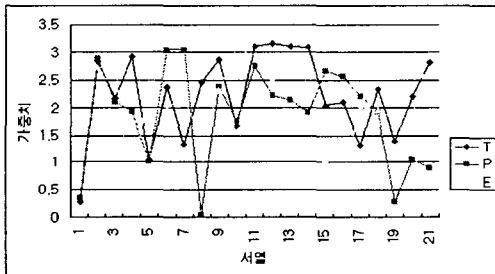
[표 2] 사용자 행위 가중치

		E				...
		2	2	5	4	
F	3	S	5	8	7	0.0
	4	6.005	6.005	9.005	8.005	0.005
	4	9	9	12	11	3
	6	18	18	21	20	10
...		...	...	...	...	...

이것을 사용자 서열로 표현하면,

S1={5, 5, 8, 7, ...}, S2={6.005, 6.005, 9.005, 8.005, ...}, S3={9,9,12,11, ...}, S4={18,18,21,20, ...}, ...

값으로 생성된다. 이중 가중치가 가장 높은 상태순으로 정렬하여 처리하였다. 이렇게 100회 로그인을 평가한 후, [그림 5]와 같은 적합도 값을 얻었다. 따라서 실험결과, 보상함수, 가치함수, 시간의 최대치가 서로 일치하거나 근접한 결과를 얻고 있어 사용자 학습에 ±5%의 오차를 갖고 있다.



[그림 5] 적합도 평가

6. 결론

본 논문에서는 HMM(Hidden Markov Model)을 이용하는 학습 모델을 설정하고 사용자 프로파일을 학습시켜 사용자의 유형을 분석하였다.

AUA을 통해 학습한 에이전트는 주변 환경 변수 및 사용자와 에이전트간의 행동 패턴에 따라 유형을 쉽게 학습하였다. 또한 프로파일을 갱신하면서 사용자의 행위를 고려하여, 다음 행위에 필요한 가능성을 제시할 수 있었다. 따라서 사용자 패턴을 인식함으로써 보다 안정적이고, 정확한 서비스를 지원할 수 있었다.

향후 연구 과제로는 사용자 프로파일 서열을 보다 빠른 시간에 분석하고, 에이전트가 학습할 수 있는 환경을 제시하는 것이다. 또한 에이전트간의 통신 및 압축기능에 대한 지원이 필요하다.

참고문헌

[1] Lacher,M.S., and Koch, "An Agent-based

knowledge management framework.", In Proc. AAAI Spring Symposium 2000, pp 145-147, 2000.

[2] Koch, M. and Lacher, M.S., "Integrating Community Services A Common Infrastructure Proposal", *Proc. Knowledge-Based Intelligent Engineering Systems and Allied Technologies*, Brighton, UK, pp. 56-59, Sep. 2000.

[3] Michael Koch, Wolfgang Wörndl, "Community Support and Identity Management", *Proc. European Conf. on Computer Supported Cooperative Work (ECSCW 2001)*, Bonn, Germany, Sept. 2001.

[4] S. Henikoff and J. G. Henikoff, "Amino acid substitution matrices from protein blocks", *Proc. Natl. Acad. Sci. USA*, vol89, pp.10915-10919, 1992.

[5] Maron, M. E. and Kuhns, J. L., "On Relevance, Probabilistic Indexing and Information Retrieval", *Association for computing Machinery*, 7(3), pp. 216-244, 1960.

[6] Robertson, S. E. and Sparck Jones, K., "Relevance Weighting of Search Terms", *Journal of the American Society for Information Science*, 27, pp.129-146, 1976.

[7] Winona C. Barker, John S. Garavelli, Zhenglin Hou, Hongzhan Huang, Robert S. Ledley, Peter B. McGarvey, "Protein Information Resource: a community resource for expert annotation of protein data", *Nucleic acids Research*, Vol.29, No.1, pp. 29-32, 2000.

[8] Cathy H. Wu, Hongzhan Huang, Leslie Arminski, Jorge Castro-Alvear, Yongxing Chen, Zhang-Zhi Hu, Robert S., "The Protein Information Resource: an integrated public resource of functional annotation of proteins", *Nucleic Acids Research*, Vol.30, No.1, pp.35-37, 2002.

[9] Tom M. Mitchel, "Machine Learning", McGraw-Hill Com., 1997.

[10] Thomas G. Dietterich, "Hierarchical Reinforcement Learning with the MAXQ Value Function Decomposition", *Journal of artificial Intelligence Research*, 13, pp. 227-303, 2000.

[11] George J. Pappas, Member, IEEE, and Slobodan Simic, "Consistent Abstractions of Affine Control Systems",*IEEE Transactions on automatic control*, Vol. 47, No. 5, MAY 2002.

[12] T. Yoshikawa, T. Furuhashi, Y. Uchidawa. " The Effect of Combination of DNA Coding Method with Pseudo-Bacterial GA" *Proceeding of the 1997 IEEE International Intermag. 97 Magnetics Conference 1997.*