

부정 선택을 이용한 DNA의 패턴 분류

Classification of DNA Pattern Using Negative Selection

심귀보 · 이동욱

Kwee-Bo Sim and Dong-Wook Lee

중앙대학교 전자전기공학부

요 약

인간 및 여러 생물들의 DNA 서열이 밝혀짐에 따라 DNA 서열 정보를 이용할 수 있는 계산적 처리방식에 대한 요구가 늘어나고 있다. 본 논문에서는 DNA의 패턴을 분류할 수 있는 면역계 부정 선택에 기반한 알고리즘을 제안한다. 부정 선택은 면역세포 생성시 자신을 인식하지 않는 항원 인식부를 생성하기 위한 과정이다. 이 항원 인식부를 통해 자기와 비자기를 구별한다. 이것을 n 개의 자기 또는 비자기 집단으로 확장하고 n 개의 항원 집단을 구성하면 n 개의 패턴 분류가 가능하다. 본 논문에서는 부정 선택에 기반한 DNA 염기 레벨에서의 패턴 분류방법과 아미노산 레벨에서의 패턴분류 방법을 제안한다.

Abstract

According to revealing the DNA sequence of human and living things, it increases that a demand on a new computational processing method which utilizes DNA sequence information. In this paper we propose a classification algorithm based on negative selection of the immune system to classify DNA patterns. Negative selection is the process to determine an antigenic receptor that recognize antigens, nonself cells. The immune cells use this antigen receptor to judge whether a self or not. If one composes n group of antigenic receptor for n different patterns, they can classify into n patterns. In this paper we propose a pattern classification algorithm based on negative selection in nucleotide base level and amino acid level.

Key words : 부정 선택, DNA, 패턴분류, 면역시스템

1. 서 론

분자생물학의 발달과 게놈프로젝트에 의하여 인간 및 여러 생물들의 유전자 정보를 갖는 DNA의 서열을 밝혀낼 수 있게 되었다. 하지만 게놈의 DNA 서열을 모두 밝혀냈다고 하여 이중 어떤 부분이 유전자 이고, 어떤 유전자가 언제 어떻게 발현되는지는 알 수 있는 것은 아니다. 또한 30억 염기의 인간 유전자중 대략 10%정도만이 단백질을 합성하는 유전정보를 가지고 있는데, 이 영역도 매우 복잡하게 분포 되어 있다. 따라서 본격적인 게놈 프로젝트는 밝혀진 게놈정보를 이용해 유전자의 기능을 밝히는 포스트 게놈시대로 들어섰다. 또한 포스트 게놈시대에 들어서면서 유전정보를 정보학으로 다루는 바이오정보기술(bioinformatics)의 중요성이 점점 높아지고 있다[1, 2].

생체의 면역계[3, 4]는 외부에서 침입한 항원을 인식하고 제거하는 고도로 발달된 기관이다. 본 논문에서는 면역계의 면역세포 생성메커니즘 중 하나인 부정 선택(negative selection)을 이용한 DNA 패턴 분류 알고리즘을 제안한다. 생체의 면역계는 외부의 항원을 인식하기 위한 다양한 항체를 생성한다. 항체를 생성하는 대표적인 면역세포는 B세포로서, B세포는 자기 자신을 항원으로 인식하지 않기 위해 초기 생성시 부정 선택의 과정을 거친다. 이와 같이 부정 선택을 거

친 B세포들은 자기 자신과 외부물질을 분류하는 능력을 가진다.

면역세포를 생성하는 방법에는 부정 선택과 긍정 선택의 두 가지 방법이 있다. 부정 선택은 자기자신을 인식하지 않고 항원을 인식하는 올바른 항원 인식부를 가진 면역세포를 생성하는 과정이다. 한편 긍정 선택은 자기자신의 특징적 패턴인 MHC(주조직 적합성 복합체)를 올바르게 인식하는 인식부를 가진 면역세포를 생성하는 과정이다. Forest[5-7] 등은 부정 선택 방법을 컴퓨터 시스템의 비정상 탐색에 적용하였고 Kim과 Bently[8]는 부정 선택 방법을 네트워크의 침입탐지에 적용하였다. Sim[9] 등은 긍정 선택과 부정 선택을 모두 이용한 자기-비자기 인식 알고리즘을 제안하였고 Esponda[10] 등은 긍정 선택 및 부정 선택의 일반적인 틀을 제시하였다.

본 연구에서는 면역세포가 자기와 비자기를 구별하는 원리를 이용해 패턴을 분류하는 알고리즘을 개발하고 이를 DNA의 패턴분류에 적용하고자 한다. 생물정보학에서 패턴 분류가 이용되는 부분은 유전자 영역(gene region)과 비유전자 영역(intergenic region)의 구분, RNA의 구조 예측, 단백질 구조 예측, 단백질 군 분류, RNA 군분류, DNA 칩의 마이크로 어레이 해석, 유전자 발현정보 분류 등이 있다. 이와 같은 문제를 해결하기 위해 신경회로망, 진화연산, 확률 그래프 모델 등의 기계학습법이 이용된다[11-13]. 본 논문에서는 부정 선택에 기반한 DNA 염기 레벨에서의 패턴 분류 방법과 아미노산 레벨에서의 패턴 분류 알고리즘을 제안하고 시

접수일자 : 2004년 2월 21일
완료일자 : 2004년 7월 20일

물레이션을 통하여 그 유효성을 검토한다.

2. DNA의 구조

모든 생물체는 각자 고유의 DNA(deoxyribonucleic acid)를 가지고 있다. DNA는 개체의 특성을 발현시키는 유전코드로서, A(아데닌) T(티민, RNA에서는 U:우라실) G(구아닌) C(시토신)의 4개의 염기배열로 이루어져 있다. 그 구조는 길다란 나선형의 형태로 꼬여 있으며 A는 T, G는 C와 상보적으로 결합하고 있다. 또한 DNA는 염기 3개의 배열이 한 의미단위를 이루어 해석된다. 이 의미단위를 생물학적인 용어로 코돈(codon)이라 한다. 코돈의 가지 수는 $4 \times 4 \times 4 = 64$ 개이며 이것이 코드화하는 아미노산은 20가지이다. 코돈의 64가지 패턴에 대하여 생성하는 아미노산이 20가지인 이유는 다른 코돈이 같은 아미노산을 만들기도 하기 때문이다. 이것은 표 1에 나타나 있다[3].

DNA는 RNA로 전사되어 리보솜에서 단백질로 번역된다. 즉 아미노산을 암호화하는 DNA의 배열에 따라 아미노산의 합성순서를 결정하여 여러 종류의 단백질을 만들어낸다. RNA의 단백질로의 번역은 AUG에서 시작해서 UGA(UAA,UAG)에서 번역이 끝난다. 따라서 DNA 코드중 단백질로 번역되는 부분은 시작코돈인 AUG와 종료코돈인 UGA(UAA, UAG) 사이에 존재하는 염기들이다.

그림 1은 DNA의 번역 예를 나타낸다. 하나의 코돈이 하나의 아미노산을 생성하며 시작코돈의 위치에 따라 코돈 즉 3개의 염기가 짝지어지는 방법이 바뀐다. 따라서 같은 DNA도 코돈의 시작부위에 따라 세 가지 방식으로 해석될 수 있다. 이 세 가지로 묶는 방식을 reading frame이라 한다.

표 1. 유전 코드
Table 1. Genetic code

	U		C		A		G		
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	UCA	UAA		정지	UGA	정지	A	
	UUG	UCG	UAG		정지	UGG	Trp	G	
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA	CCA	CAA		Gln	CGA	A		
	CUG	CCG	CAG		CGG	G			
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA	ACA	AAA		Lys	AGA	Arg	A	
	AUG	Met	ACG		AAG	AGG	G		
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG	GGG	G		

아미노산 약어 알라닌-Ala, 아르기닌-Arg, 아스파라긴-Asn, 아스파르트산-Asp, 시스테인-Cys, 글루탐산-Glu, 글루타민-Gln, 글리신-Gly, 히스티딘-His, 이소류신-Ile, 류신-Leu, 리신-Lys, 메티오닌-Met, 페닐알라닌-Phe, 프롤린-Pro, 세린-Ser, 트레오닌-Thr, 트립토판-Trp, 티로신-Tyr, 발린-Val.

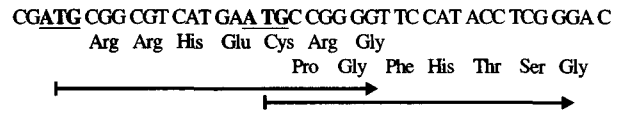


그림 1. DNA 번역 예

Fig. 1. An example of DNA translation

3. 면역계의 부정 선택 알고리즘

3.1 면역 시스템

생체의 방어체계인 면역계는 바이러스, 기생균, 병원균, 독소 등과 같은 항원이라고 통칭하는 매우 다양한 외부 유기체나 단백질에 대하여 생체를 방어할 수 있는 매우 정교하고 복잡한 시스템이다. 면역계를 구성하는 기본 요소는 두 가지 형태의 림프구이다. 이는 B세포(B 림프구)와 T세포(T 림프구)로써, B세포는 항체를 분비하는 체액성 반응을 하며, T세포는 면역에 관련된 세포를 자극 또는 억제하거나 감염된 세포를 죽이는 세포성 반응을 주로 담당한다[3, 4].

개체에는 각각 개인적인 특징을 나타내는 단백질이 존재한다. 이를 주조직 적합성 복합체(major histo compatibility complex, MHC) 단백질이라 한다. T세포에는 MHC 단백질을 인식하는 부분이 존재하며 이를 이용해 자신의 세포여부를 판단하게 된다. 한편 B세포나 T세포는 특정 항원을 인식할 수 있는 인식부를 가지고 있으며 이를 항원 수용체(antigen receptor)라 한다[4].

여러 가지 면역세포 중 자기를 판별해주는 MHC 단백질 인식부와 항원을 인식하는 항원 수용체를 가지고 있는 세포는 T세포이다. T세포는 이 두 가지 인식부를 가지고 항원에 의해 감염된 자기 세포를 인식한다. 따라서 면역계는 T세포 생성시 MHC 단백질 인식부와 항원 수용체의 정상적인 동작 여부를 확인하면서 T세포를 생성한다. 이때 수용체의 정상적인 동작여부를 가리는 방법으로 사용되는 것이 긍정 선택(positive selection)과 부정 선택(negative selection)이다. 또한 B세포도 항원 수용체를 생성하기 위해 부정 선택을 이용한다.

긍정 선택은 각 면역세포의 MHC 단백질 인식기능을 확인하는 선택방법이다. 자기세포에서 존재하는 MHC 단백질을 정확히 인지할 수 있는 면역세포만이 사용가능하기 때문에 갖 생성된 면역세포에 MHC 단백질을 결합시켜 긍정적인 선택이 되는 세포들로부터 면역 세포를 구성한다. 이때 선택되지 않은 면역 세포들은 제거 된다.

부정 선택은 항원의 인식에 있어서 자기 세포를 항원으로 인식하는 것을 배제하기 위한 방법이다. 면역세포에 MHC 단백질을 결합시켰을 때 항원수용체가 MHC 단백질을 인식하지 못하는 세포들만 선택된다. 이때 긍정적인 선택을 하는 면역세포는 MHC 단백질을 항원으로 인식하는 세포들이므로 제거 된다.

그림 2는 생체 면역계에서 정상적인 면역 세포의 형성과정을 나타낸다. 그림의 면역세포는 긍정 선택과정을 거치면서 MHC 인식부가 결정되고 부정 선택과정을 거치면서 자신을 인식하지 않는 항원인식부가 결정된다. 따라서 두 가지 선택과정을 거친 T세포는 자신과 항원을 인식할 수 있는 두 개의 인식부를 가진다.

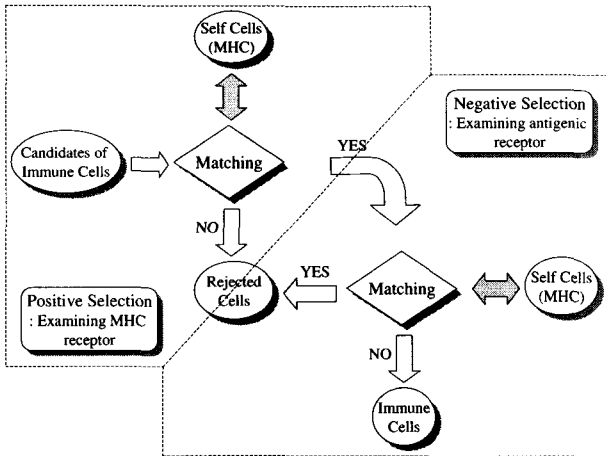


그림 2. T 세포의 형성과정
Fig. 2. Developing process of T cell

3.2 부정 선택 알고리즘

부정 선택에 기반한 변형 검지 알고리즘은 Forrest 등에 의해 제안된 자기-비자기 인식 알고리즘의 하나이다[5-7]. 이는 자기 공간에 대해서 부정 선택을 거쳐 인식부 세트를 구성하고 이를 이용해 비자기 인식에 사용하였다. 이때 항원 인식부를 변형 인식부라 명명하였다. 변형 인식부(anomaly detector)를 이용한 비자기 인식 알고리즘은 자기 공간에 부분적으로 변경된 부분과 추가된 부분을 인식부를 통해 찾아낸다.

이 알고리즘은 크게 두 부분으로 구성된다. 하나는 자기 공간을 검사하기 위한 변형 인식부를 구성하는 부분이며 다른 하나는 구성된 변형 인식부를 이용하여 자기 공간을 모니터링하며 변화의 발생을 검사하는 부분이다.

그림 3은 부정 선택에 의해 변형 인식부를 생성하는 방법을 나타낸다. 변형 인식부는 자기파일과 일치하지 않는 스트링(string)을 이용하여 구성한다. 우선 보호되어야 할 자기 공간 S 를 설정한다. 다음으로 일정한 길이 l 의 랜덤 스트링을 생성한다. 랜덤 스트링의 집합을 R_0 로 한다. R_0 의 모든 스트링과 S 의 모든 스트링을 사이의 r -인접 매칭 규칙을 적용한다. R_0 의 스트링 중 S 와 매치되지 않은 스트링으로 인식부 R 을 구성한다. 매칭된 스트링 집합 E 는 버려진다.

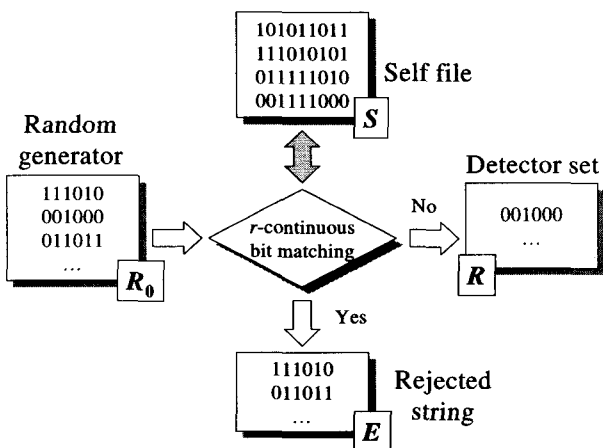


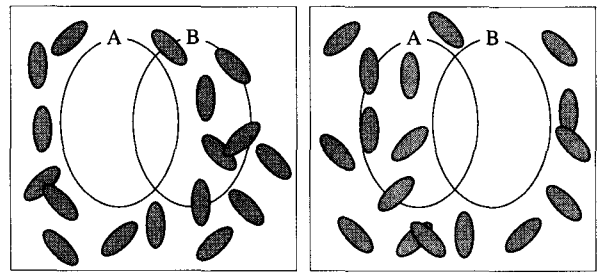
그림 3. 부정선택 방법
Fig. 3. Negative selection

이와 같은 과정을 통해 만들어진 변형 인식부들을 이용해 자기 파일이 아닌 부분을 인식한다. 이 알고리즘은 충분한 개수의 변형 인식부를 준비해 둬서 다양한 종류의 항원에 대해 인식할 수 있는 장점을 가지고 있다.

4. 부정 선택에 기반한 패턴 분류 알고리즘

4.1 염기에 기반한 방법

본 절에서는 부정 선택에 기반한 패턴 분류 알고리즘을 설명한다. 부정 선택은 자기를 인식하지 않는 항원 인식부를 결정하는 과정이다. 올바른 항원 인식부를 가진 면역세포들은 비자기로부터 자기를 분류해낼 수 있다. 같은 방법으로 특정한 자기의 부정 선택을 거친 면역세포들은 특정한 자기를 인식할 수 있다. 따라서 이러한 면역 세포들을 자기를 인식하는 인식부로 사용할 수 있다. 여기서 자기를 특정 패턴으로 간주하고 n 개의 패턴으로 확장하면 각각의 패턴에 대하여 n 개의 인식부 집합을 얻을 수 있다. 그림 4는 패턴 A와 패턴 B에 대한 인식부 집합을 나타내는 그림이다.



(a) 패턴 A에 대한 인식부 집합
(b) 패턴 B에 대한 인식부 집합
그림 4. 인식부 집합을 이용한 패턴 분류

(a) Detector set of pattern A
(b) Detector set of pattern B
Fig. 4. Pattern Classification using detector set

DNA는 4개의 염기로 구성되어있다. 따라서 변형 인식 알고리즘의 이진 스트링 대신 4진 DNA 스트링을 사용한다. 또한 매칭 방식은 r -인접 매칭 방법 대신 두 스트링의 해밍 거리(hamming distance)를 이용하여 임계값 M 을 이용해 인식부를 선택한다. 해밍 거리는 두 스트링의 특정위치의 문자가 같은 경우 1이 증가하고 다를 경우 0이 증가한다. 인식부는 인식부와 패턴의 해밍 거리가 먼 것을 선택한다. 선택에 사용되는 식은 (1)이다.

$$H(r, s_i) > M \quad (1)$$

단, $H()$ 는 해밍 거리 r 은 인식부, s_i 는 패턴 S 의 i 번째 위치부터 인식부 길이만큼의 염기배열이다. 그림 5는 부정 선택 방법에 의해 각 패턴에 대해 인식부 집합을 구성하는 방법을 나타낸다.

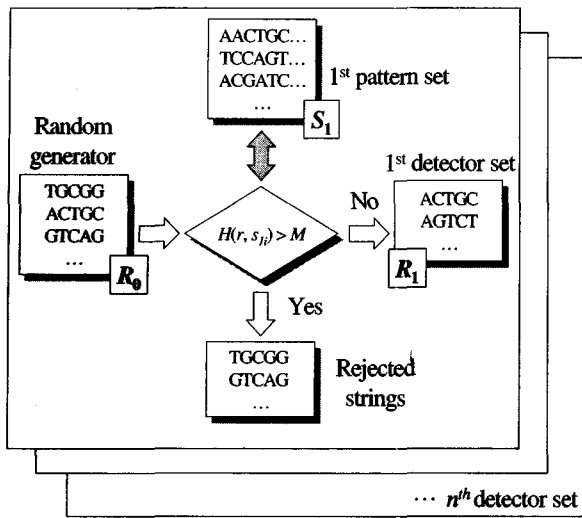


그림 5. 각 패턴에 따른 인식부 구성
Fig. 5. Composition of detector set for each pattern

이와 같은 과정에 의해 n 개 패턴에 대한 n 개의 인식부 집합이 생성된다. 이 인식부 집합을 이용해 새로운 입력 패턴에 대하여 어떤 패턴에 속하는지를 결정한다. 즉, 패턴인식 방법은 다음과 같다. 특정 패턴의 모든 인식부와 입력 패턴의 해밍 거리가 M 이하이면 입력 패턴은 특정 패턴에 속한 것으로 결정한다. 이 결정 함수를 식으로 표현하면 (2)와 같다.

$$\max_i [H(r, s_i)] \leq M \quad (2)$$

4.2 아미노산에 기반한 방법

단백질의 군 분류나 구조 예측을 위해서는 DNA 염기 단위가 아닌 아미노산의 단위로 데이터를 처리하는 것이 필요하다. 따라서 DNA 염기 서열을 각 reading frame에 따라 아미노산으로 번역한 후 알고리즘을 적용한다. 이때 각 아미노산은 표 2와 같이 일련번호를 정하여 20개의 배열을 한 단위로 생각한다. 인식부 구성 및 패턴 인식 방법은 4.1절과 같다.

표 2. 아미노산 코드
Table 2. Amino acid code

번호	1	2	3	4	5	6	7	8	9	10
코드	A	R	N	D	C	Q	E	G	H	I
아미노산	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile
번호	11	12	13	14	15	16	17	18	19	20
코드	L	K	M	F	P	S	T	W	Y	V
아미노산	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val

6. 실험 결과 및 고찰

제안한 방법의 유효성을 증명하기 위하여 박테리아 (bacteria) 및 균류 (fungi)의 rRNA 데이터 분류 문제를 적용

하였다. rRNA는 리보솜 RNA로서 세포속에서 단백질을 합성하는 기관의 RNA를 말한다. rRNA의 패턴은 종에 따라 약간씩 다른데 원핵세포의 일종인 박테리아와 진핵세포의 일종인 균류의 RNA는 각각 특징적인 RNA 패턴을 가지고 있다. 본 실험에서는 박테리아의 한 종류인 actinobacteria와 균류의 하나인 basidiomycota의 데이터를 사용하였다. 본 데이터는 RNA 비교 웹 사이트[14]에서 얻었다. rRNA는 그 길이에 따라 3가지 형태로 나뉜다. 그것은 각각 5S(1S = 60 염기), 16S, 23S의 rRNA이다. 본 실험에서는 그중 가장 길이가 짧은 5S rRNA 데이터를 사용하였다. Actinobacteria의 rRNA 데이터는 총 60개가 알려져 있으며 basidiomycota 데이터는 41개가 알려져 있다. 동일한 실험을 위하여 이중 30개의 데이터를 학습패턴으로 사용하고 10개의 패턴을 테스트 패턴으로 사용하였다. 또한 FNR(False Negative Rate: 다른 패턴을 자신의 패턴으로 잘못 인식하는 비율)을 구하기 위하여 임의로 생성한 rRNA 패턴 100개를 두개의 패턴에 대하여 각각 분류 실험을 하였다. 각 실험은 일반적인 성능을 구하기 위하여 10회 실시하여 평균값을 구하였다. 이때 각 실행 때 마다 인식부를 새로 생성하였다.

표 3은 각 패턴을 학습하기 위한 파라미터값을 나타내며 표 4는 패턴분류 실험결과를 나타낸다. Actinobacteria 데이터는 93%의 인식률과 1.2%의 FNR을 나타내었고 basidiomycota 데이터는 87%의 인식률과 6.0%의 FNR을 나타내었다. 표 3의 파라미터는 휴리스틱한 방법으로 결정하였다. 본 실험에서 파라미터의 설정에 따라 패턴 인식 성능이 약간씩 변화를 보였다. 따라서 본 알고리즘의 성능을 개선하기 위하여 인식부의 개수, 인식부의 길이, 임계값 M 등의 최적화가 필요하다.

표 3. 실험 파라미터
Table 3. Experimental Parameters

패턴 이름	인식부 개수	인식부 길이(l)	임계값(M)
Actinobacteria	10	6	3
Basidiomycota	8	8	4

표 4. 패턴 분류 결과
Table 4. Experimental result of pattern classification

패턴 이름	학습 패턴 수	테스트 패턴 수	인식률 (%)	FNR (%)
Actinobacteria	30	10	93	1.2
Basidiomycota	30	10	87	6.0
Random string	-	100	-	-

제안한 방식은 면역세포가 자신의 세포와 외부세포를 구별하는 방식을 패턴분류에 응용한 것이다. 즉 자신의 패턴을 인식하지 못하는 항원 인식부 집합을 이용한 패턴 인식 방법이다. 본 방법은 패턴의 크기와 양에 구애 받지 않는 알고리즘이다. 특히 특징추출이 어렵고, 패턴이 크며, 크기가 일정하지 않은 패턴 분류 문제에 대하여 매우 유용할 것으로 생각된다.

하지만 제안한 방법을 DNA의 패턴 분류에 적용할 때 다음과 같은 3가지 고려할 점이 있다.

첫째, 자기 패턴을 구성할 때 패턴의 시작위치가 결정하는 것이 매우 중요하다. 같은 패턴이라도 시작위치가 한 염기만 달라져도 다른 패턴으로 인식할 가능성이 있기 때문이다. 따라서 패턴의 시작위치를 결정하는 것이 매우 중요한 문제가 된다. 본 문제를 해결하는 방법으로 특정코돈(예를 들면 시작코돈)을 패턴으로 시작점으로 설정하거나 시작점을 바꾸어 여러 번 인식부로 인식시키는 방법을 적용할 수 있다.

둘째, DNA의 특징상 진핵세포 생물에게는 유전자 사이사이에 아무런 의미를 가지고 있는 부분 즉 인트론(Intron)이라는 부분이 삽입되어 있다는 점이다. 인트론은 유전자 해석에 영향을 주지 않는 부분이므로 동일한 유전적 기능을 같은 부분도 인트론의 패턴은 다를 수 있다. 이러한 문제는 다른 패턴 분류 알고리즘에도 공통적으로 적용되는 문제점이다.

셋째, 패턴의 복잡도가 낮을수록 그 패턴에 대한 부정 인식부의 공간이 늘어나게 되어 많은 개수의 인식부가 필요하게 된다. 단적인 예로 아주 단순한 패턴은 대부분의 인식부와 매칭이 되지 않음으로서 임의의 패턴으로 잘못 분류될 가능성이 많다. 따라서 부정 선택 방법은 최소한의 복잡도 이상을 가지는 문제에 대하여 적합하다고 할 수 있다. 특히 복잡도가 높은 패턴일수록 기존의 방법보다 좋은 성능을 나타낼 수 있다.

5. 결 론

본 논문에서는 번역계의 부정 선택에 기반한 패턴 분류 알고리즘을 제안하였다. 부정 선택은 자기와 항원을 구별할 수 있는 번역세포의 항체를 생성하는 방법이다. 본 연구에서는 항체에 의한 자기 비자기 구별방법을 확장하여 n 개의 패턴에 대하여 n 개의 항원 인식부 셋을 구성함으로써 패턴 분류 하는 방법을 제안하였다. 기존의 패턴 분류 방법은 패턴 인식을 위한 입력을 결정하기 위해 특징 추출을 하거나 패턴의 크기를 일정하게 유지시켜야 한다. 하지만 제안한 방식은 특징추출 과정이 필요 없으며 패턴의 크기가 매우 크고 크기가 일정하지 않을 때 매우 효과적이다. 박테리아와 균류의 rRNA 패턴 분류 실험을 통하여 제안한 방식의 유효성을 알 수 있었다. 차후로 다른 패턴 분류 방식과 제안한 방식에 대한 비교 검토를 수행할 예정이다.

참 고 문 헌

- [1] P. Baldi, S. Brunak, *Bioinformatics: The Machine Learning Approach*, MIT Press, Cambridge, Mass., 2001.
- [2] 김성동, 장병탁, "바이오 데이터 마이닝을 위한 기계 학습 기법", *정보과학회지*, 제18권, 제8호, pp. 63-72, 2000.
- [3] R. A. Wallace, G. P. Sanders, and R. J. Ferl, *BIOLOGY: The Science of Life*, 3rd edition, HarperCollins Publishers Inc., 1991.
- [4] I. Roitt, J. Brostoff, D. Male, *Immunology*, 4th edition, Mosby, 1996.
- [5] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri "Self-nonsel self discrimination in a computer," *Proc. IEEE Symposium on Research in Security and Privacy*, pp. 202-212, 1994.
- [6] D. Dasgupta, S. Forrest, "An anomaly detection algorithm inspired by the immune system," in *Artificial Immune Systems and Their Applications*, D. Dasgupta, Ed., Springer, pp. 262-276, 1999.
- [7] P. D'haeseleer, S. Forrest, and P. Helman, "An immunological approach to change detection: algorithms, analysis, and implications," *Proc. IEEE Symp. Computer Security Privacy*, pp. 110-119, 1996.
- [8] J. Kim, and P. J. Bentley, "Towards an artificial immune system for network intrusion detection: An investigation of clonal selection with a negative selection operator," *Proc. Congr. Evolutionary Computation*, pp. 1244-1252, 2001.
- [9] K. B. Sim and D. W. Lee, "Self-nonsel self recognition algorithm based on positive and negative selection," *IEICE Trans. on Info. & Syst.*, vol. E87-D, no. 2, Feb. 2004.
- [10] F. Esponda, S. Forrest, and P. Helman, "Formal framework for positive and negative detection scheme," *IEEE Trans. Syst., Man, and Cybern.*, Part B, vol. 34, no. 1, pp.357-373, Feb. 2004.
- [11] M. Gelfand, "Prediction of function in DNA sequence analysis," *J. Computational Biology*, vol. 1, pp. 87-115, 1995.
- [12] K. B. Hwang, D. Y. Cho, S. W. Park, S. D. Kim, and B. T. Zhang, "Applying machine learning techniques to analysis of gene expression data: Cancer diagnosis," in *Methods of Microarray Data Analysis*, Kluwer Academic Publishers, pp. 167-182, 2002.
- [13] G. B. Fogel, K. Chellapilla, and D. B. Fogel, "Identification of coding regions in DNA sequences using evolved neural networks," in *Evolutionary Computation in Bioinformatics*, G. B. Fogel and D. W. Corne, Eds. San Francisco, CA: Morgan Kaufmann, pp. 196-218, 2003.
- [14] Comparative RNA Web Site, URL:<http://www.rna.icmb.utexas.edu/>

저 자 소 개

심귀보(Kwee-Bo Sim)

1984년 : 중앙대학교 전자공학과 공학사
1986년 : 동대학원 전자공학과 공학석사
1990년 : The University of Tokyo 전자
공학과 공학박사

1991년 ~ 현재 : 중앙대학교 전자전기공학부
교수

2000년 ~ 현재 : 제어자동화시스템공학회
이사 및 지능시스템연구회 회장

2002년 ~ 현재 : 중앙대학교 산학연컨소시엄센터 센터장

2003년 ~ 현재 : 일본계측자동제어학회(SICE) 이사

2003년 ~ 현재 : 한국퍼지 및 지능시스템학회 부회장

관심분야 : 인공지능, 지능로봇, 지능시스템, 다개체시스템,
학습 및 적응알고리즘, 소프트 컴퓨팅(신경망,
퍼지, 진화연산), 인공지능시스템, 침입탐지시스
템, 진화하드웨어, 인공두뇌, 지능형 홈 및 홈네
트워킹, 유비쿼터스 컴퓨팅 등

Phone : +82-2-820-5319

Fax : +82-2-817-0553

E-mail : kbsim@cau.ac.kr

Homepage URL : <http://alife.cau.ac.kr>

이동욱(Dong-Wook Lee)

1996년 : 중앙대학교 제어계측공학과
공학사

1998년 : 동 대학원 제어계측학과
공학석사

2000년 : 동 대학원 제어계측학과
공학박사

2002년~현재 : 중앙대학교 정보통신연구원
연구전담교수

관심분야 : 인공생명, 진화연산, 인공면역계, 인공두뇌 등

E-mail : dwlee@wm.cau.ac.kr