



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2009-0077536
(43) 공개일자 2009년07월15일

(51) Int. Cl.

G06F 9/44 (2006.01)

(21) 출원번호 10-2008-0003544

(22) 출원일자 2008년01월11일

심사청구일자 2008년01월11일

(71) 출원인

인하대학교 산학협력단

인천 남구 용현동 253 인하대학교

(72) 발명자

정덕진

인천 남구 용현3동 인하대학교 하이테크 510호

김병수

인천 연수구 연수1동 동남아파트 102동 402호

(뒷면에 계속)

(74) 대리인

이원희

전체 청구항 수 : 총 7 항

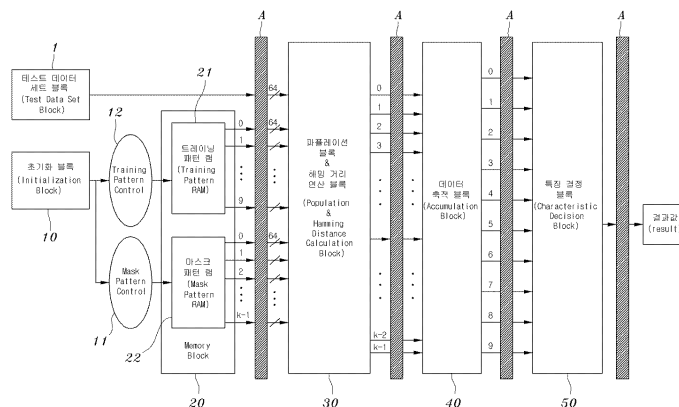
(54) DNA 컴퓨팅 기술에 기초한 패턴 인식 장치

(57) 요약

본 발명은 대규모 병렬 처리를 통해 인식 시간을 줄일 수 있는 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치에 관한 것으로, 본 발명에 따른 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치는,

트레이닝 데이터 세트와 랜덤 마스크 세트가 생성되는 초기화 블록; 상기 초기화 블록에서 생성된 트레이닝 데이터 세트를 수신하여 저장하는 트레이닝 패턴 램(RAM)과 상기 초기화 블록에서 생성된 랜덤 마스크 세트를 수신하여 저장하는 마스크 패턴 램을 구비한 메모리 블록; 상기 트레이닝 패턴 램과 마스크 패턴 램으로부터 트레이닝 패턴과 랜덤 마스크 패턴을 수신하여 로딩(load)하는 과플레이션 블록; 상기 랜덤 마스크 패턴을 사용하여 테스트 패턴과 상기 트레이닝 패턴을 비교하는 해밍 거리 연산 블록; 비교 후 매칭(matching)되는 패턴들의 수를 가산하는 데이터 축적 블록; 및, 2개의 클럭(clock) 동안 각각의 디지트(digits)에 대한 매칭된 수를 비교하고, 테스트 패턴과 매칭된 수에 대한 등급을 생성하는 특징 결정 블록을 포함하는 것을 특징으로 한다.

대표도



(72) 발명자

왕규열

인천시 남동구 구월1동 동아아파트 2동 410호

이승열

인천 남구 학익동 306-127 광림빌라 2동 302호

송재연

인천 남구 용현3동 210-56 영동하이츠 빌라 201호

이상설

인천 남구 용현3동 210-56 영동하이츠 빌라 201호

이중호

서울 강남구 압구정동 신현대아파트 124-603호

박재현

서울 강남구 압구정동 현대아파트 20-803호

특허청구의 범위

청구항 1

트레이닝 데이터 세트와 랜덤 마스크 세트가 생성되는 초기화 블록;

상기 초기화 블록에서 생성된 트레이닝 데이터 세트를 수신하여 저장하는 트레이닝 패턴 램(RAM)과 상기 초기화 블록에서 생성된 랜덤 마스크 세트를 수신하여 저장하는 마스크 패턴 램을 구비한 메모리 블록;

상기 트레이닝 패턴 램과 마스크 패턴 램으로부터 트레이닝 패턴과 랜덤 마스크 패턴을 수신하여 로딩(load)하는 파플레이션 블록;

상기 랜덤 마스크 패턴을 사용하여 테스트 패턴과 상기 트레이닝 패턴을 비교하는 해밍 거리 연산 블록;

비교 후 매칭(matching)되는 패턴들의 수를 가산하는 데이터 축적 블록; 및,

2개의 클럭(clock) 동안 각각의 디지트(digits)에 대한 매칭된 수를 비교하고, 테스트 패턴과 매칭된 수에 대한 등급을 생성하는 특징 결정 블록

을 포함하는 것을 특징으로 하는 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치.

청구항 2

청구항 1에 있어서,

상기 초기화 블록은 마스크 패턴의 선택 비트, 트레이닝 패턴에 대한 파플레이션(population) 수, 그리고 병렬 처리 프로세싱 유닛의 수를 설정하는 것을 특징으로 하는 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치.

청구항 3

청구항 2에 있어서,

상기 선택 비트는 31비트, 상기 파플레이션 수는 100, 상기 병렬 처리 프로세싱 유닛의 수는 1000으로 각각 설정되는 것을 특징으로 하는 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치.

청구항 4

청구항 1에 있어서,

상기 트레이닝 패턴 램은 3823개의 패턴을 가진 10진수 디지털 세트로 이루어지고, 상기 마스크 패턴 램은 10진수 디지털을 위한 4000개의 랜덤 마스크 패턴으로 이루어지는 것을 특징으로 하는 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치.

청구항 5

청구항 1에 있어서,

상기 파플레이션 블록은 4000개의 랜덤 마스크 패턴 중에서 100개의 랜덤 마스크 패턴을 로딩하고, 3760개의 패턴으로부터 0~9 디지트(digit)을 갖는 처음 10개의 패턴을 로딩하는 것을 특징으로 하는 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치.

청구항 6

청구항 1에 있어서,

상기 해밍 거리 연산 블록은 매칭 여부를 분류하기 위해, 논리 연산 프로세서와 병렬 처리 프로세서를 사용하는 것을 특징으로 하는 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치.

청구항 7

청구항 1에 있어서,

상기 데이터 축적 블록은 다중 클럭(multiple clock)을 사용하는 것을 특징으로 하는 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치.

명세서

발명의 상세한 설명

기술 분야

- <1> 본 발명은 패턴 인식 장치에 관한 것으로, 보다 상세하게는 대규모 병렬 처리를 통해 인식 시간을 줄일 수 있는 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치에 관한 것이다.

배경 기술

- <2> 최근, 분자 컴퓨터(molecular computing)의 포텐셜(potential)에 대한 관심이 증가되고 있다. 분자 컴퓨터는 대규모 병렬 처리 프로세싱을 지원하고, 이중 교배(hybridization)와 복제(duplication)와 같은 단순한 작업을 이용한다.
- <3> 장(Zhang)과 장(Jang)은, 패턴 분류와 관련된 연산 문제를 해결하기 위해 분자 컴퓨터의 포텐셜을 조사하기 위해 간단한 DNA-컴퓨팅 작업을 이용하는 분자 단위 전개 학습 모델을 제안하였다. (Zhang, B.-T. and Jang, H.-Y., "A Bayesian algorithm for in vitro molecular evolution of pattern classifiers", Lecture Notes in Computer Science(DNA11), 3384:458-467, 2005. 참조)
- <4> 본 발명의 발명자들은 소프트웨어 시뮬레이션으로 8 x 8이미지들의 3823개 패턴의 디지털 인식 데이터 세트에 대한 상기 분자 단위 전개 학습 모델의 수행 능력을 평가하였다.
- <5> 그 결과, 도 1에 도시된 바와 같이, 92.7%의 분류 정확도를 얻을 수 있었고, 각각 91.4%, 84.9%의 분류 정확도를 가진 SVM(Support Vector Machine)과 결정 트리(decision trees)와 같은 표준 기계 학습 알고리즘(standard machine learning algorithms)보다 성능이 우수하다는 것을 알게 되었다. 도 1에서 Pop는 과플레이션 수를 의미한다.

발명의 내용

해결 하고자하는 과제

- <6> 전술한 분자 단위 전개 학습 모델은 간단한 작업을 이용하면서도 상기 표준 기계 학습 알고리즘과 비교하여 우수한 수행 능력을 보여준다. 또한, 그것은 대규모 병렬 처리 작업을 필요로 한다. 분자 단위 전개 학습 모델의 모든 특징들은 하드웨어 임플러멘테이션(어떤 컴퓨터 언어를 특정 기종의 컴퓨터에 적합하게 하는 것)에 적합하기 때문에, 본 발명의 발명자들은 분자 단위 전개 알고리즘("DNA Hypernetworks for Information Storage and retrieval," Lecture Notes in Computer Science (DNA12), LNCS 4287, pp.298-307, 2006. 참조)에 기초한 대규모 병렬 처리를 통해 인식 시간을 줄일 수 있는 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치를 제공하고자 한다.

과제 해결수단

- <7> 상기과 같은 본 발명의 목적을 달성하기 위한 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치는,
- <8> 트레이닝 데이터 세트와 랜덤 마스크 세트가 생성되는 초기화 블록; 상기 초기화 블록에서 생성된 트레이닝 데이터 세트를 수신하여 저장하는 트레이닝 패턴 램(RAM)과 상기 초기화 블록에서 생성된 랜덤 마스크 세트를 수신하여 저장하는 마스크 패턴 램을 구비한 메모리 블록; 상기 트레이닝 패턴 램과 마스크 패턴 램으로부터 트레이닝 패턴과 랜덤 마스크 패턴을 수신하여 로딩(load)하는 과플레이션 블록; 상기 랜덤 마스크 패턴을 사용하여 테스트 패턴과 상기 트레이닝 패턴을 비교하는 해밍 거리 연산 블록; 비교 후 매칭(matching)되는 패턴들의 수를 가산하는 데이터 축적 블록; 및, 2개의 클럭(clock) 동안 각각의 디짓들(digits)에 대한 매칭된 수를 비교하고, 테스트 패턴과 매칭된 수에 대한 등급을 생성하는 특징 결정 블록을 포함하는 것을 특징으로 한다.
- <9> 또한, 상기 초기화 블록은 마스크 패턴의 선택 비트, 트레이닝 패턴에 대한 과플레이션(population) 수, 그리고 병렬 처리 프로세싱 유닛의 수를 설정하는 것을 특징으로 한다. 바람직하게는 상기 선택 비트는 31비트, 상기

파플레이션 수는 100, 상기 병렬 처리 프로세싱 유닛의 수는 1000으로 각각 설정될 수 있다.

- <10> 또한, 상기 트레이닝 패턴 램은 3823개의 패턴을 가진 10진수 디지털 세트로 이루어지고, 상기 랜덤 마스크 세트는 10진수 디지털을 위한 4000개의 랜덤 마스크 패턴으로 이루어지는 것을 특징으로 한다.
- <11> 또한, 상기 파플레이션 블록은 4000개의 랜덤 마스크 패턴 중에서 100개의 랜덤 마스크 패턴을 로딩하고, 3760개의 패턴으로부터 0~9 디짓(digit)을 갖는 처음 10개의 패턴을 로딩하는 것을 특징으로 한다.
- <12> 또한, 상기 해밍 거리 연산 블록은 매칭 여부를 분류하기 위해, 논리 연산 프로세서와 병렬 처리 프로세서를 사용하는 것을 특징으로 한다.
- <13> 또한, 상기 데이터 축적 블록은 다중 클럭(multiple clock)을 사용하는 것을 특징으로 한다.

효 과

- <14> 상기와 같은 본 발명에 따른 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치에 따르면, 대규모 병렬 처리를 통해 패턴 인식 시간을 줄일 수 있는 효과가 있다.

발명의 실시를 위한 구체적인 내용

- <15> 이하, 첨부된 도면을 참조하여 본 발명의 실시예를 상세히 설명한다. 본 발명을 설명함에 있어서 관련된 공지기능 혹은 구성에 대한 구체적인 설명은 본 발명의 요지를 모호하게 하지 않기 위해 생략한다.
- <16> 도 2는 본 발명에 따른 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치의 블록도, 도 3은 Verilog HDL을 사용하여 구현된 본 발명에 따른 패턴 인식 장치(프로세서)에 대해 ModelSim을 사용한 시간 시뮬레이션 결과를 도시한 도이다.
- <17> 본 발명의 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치는 하이퍼 네트워크 모델을 이용하여 수기(Handwritten) 데이터를 판단하는 장치이다. 본 발명에 따른 최기존의 3823개의 수기 데이터를 트레이닝 패턴으로, 4000개의 64bit의 랜덤 시퀀스를 랜덤 마스크 패턴으로 저장하여 가지고 있다. 수기 데이터가 입력되어 들어오게 되면 트레이닝 패턴과 마스크 패턴을 이용하여 파플레이션(Population)과 비교 연산(compare)을 수행하고 일치된 결과값을 누적시킨다. 이 과정을 모든 트레이닝 패턴을 사용할 때까지 반복한 후 일치된 결과값을 가지고 입력된 수기 데이터를 최종 판단하게 된다. 상기 하이퍼 네트워크 모델은 "'DNA Hypernetworks for Information Storage and retrieval,' Lecture Notes in Computer Science (DNA12), LNCS 4287, pp.298-307, 2006."에 기재된 하이퍼 네트워크 모델로서 본 명세서에서는 이를 원용한다.
- <18> 도 2에 도시된 바와 같이, 본 발명의 패턴 인식 장치는, 초기화 블록(10), 메모리 블록(20), 파플레이션 & 해밍 거리 연산 블록(30), 데이터 축적 블록(40), 특징 결정 블록(50)으로 구성된다. 도면 부호 A는 파이프라인 기법을 개념적으로 나타내는 것이며, 바람직하게는 레지스터이다.
- <19> 테스트 데이터 세트 블록(1)에서 나오는 화살표는 1개의 테스트 패턴을 의미한다.
- <20> 상기 초기화 블록(10)에서는 트레이닝 데이터 세트와 랜덤 마스크 세트가 생성되고, 상기 초기화 블록은 마스크 패턴(11)의 선택 비트, 트레이닝 패턴(12)에 대한 파플레이션(population) 수, 그리고 병렬 처리 프로세싱 유닛의 수를 설정한다.
- <21> 예를 들면, 상기 선택 비트는 31비트, 상기 파플레이션 수는 100, 상기 병렬 처리 프로세싱 유닛의 수는 1000으로 각각 설정한다.
- <22> 상기 메모리 블록(20)은 트레이닝 패턴 램(RAM)(21)과 마스크 패턴 램(22)을 구비한다. 상기 트레이닝 패턴 램(21)은 3823개의 패턴을 가진 10진수 디지털 세트로 이루어지는 것이 바람직하고, 상기 마스크 패턴 램(22)은 10진수 디지털을 위한 4000개의 랜덤 마스크 패턴으로 이루어지는 것이 바람직하다.
- <23> 도 2에서 트레이닝 패턴 램(21)에서 나오는 화살표는 0 ~ 9 디짓(digit)을 갖는 10개의 패턴, 마스크 패턴 램(22)에서 나오는 화살표는 100개의 랜덤 마스크 패턴을 의미한다.
- <24> 상기 파플레이션 블록과 해밍 거리 연산 블록은 하나의 블록(30)에 포함된다.
- <25> 상기 파플레이션 블록은 상기 트레이닝 패턴 램(21)과 마스크 패턴 램(22)으로부터 트레이닝 패턴과 랜덤 마스크 패턴을 로딩(load)한다.

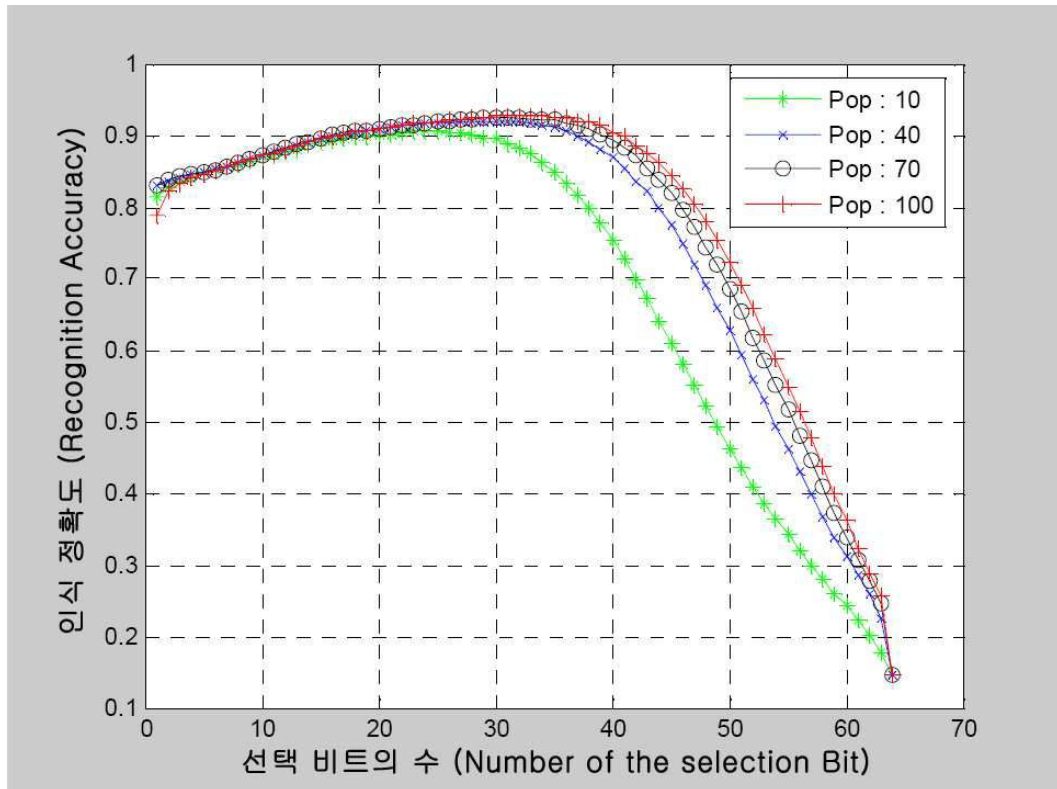
- <26> 예를 들면, 상기 과플레이션 블록은 4000개의 랜덤 마스크 패턴 중에서 100개의 랜덤 마스크 패턴을 로딩하고, 3760개의 패턴으로부터 0~9 디짓(digit)을 갖는 처음 10개의 패턴을 로딩한다.
- <27> 상기 해밍 거리 연산 블록(Hamming Distance Calculation block)은 상기 랜덤 마스크 패턴을 사용하여 테스트 패턴과 트레이닝 패턴을 비교한다. 상기 해밍 거리 연산 블록에서 매칭 여부를 분류하기 위해, 간단한 논리 연산(simple logical operation)과 병렬 처리 프로세서가 사용된다.
- <28> 도 2를 참조하면, 상기 과플레이션 블록에서 10개의 트레이닝 패턴과 100개의 랜덤 마스크 패턴을 이용하여 1000개의 비교 패턴을 만든 후, 상기 해밍 거리 연산 블록에서 1개의 테스트 패턴과 1000개의 비교 패턴을 비교하여 매칭되는 수를 추출한다.
- <29> 상기 데이터 축적 블록(40)은, 비교 후에 매칭된 패턴들의 수를 클래스(class) 별로 저장하고 누적한다. 예를 들면, 1에 매칭된 패턴은 1클래스를 증가시키게 된다. 상기 해밍 거리 연산 블록으로부터 나오는 화살표는 매칭된 패턴을 지칭한다.
- <30> 상기 데이터 축적 블록(40)은 매칭되거나 또는 매칭되지 않은 것을 지칭하는 1000개의 신호를 가산하는 데, 이것은 임계 경로(critical path)를 만든다. 여기서, 임계 경로라 함은, 여러 단계의 과정을 거치는 작업에서 그것을 완성하려면 여러 과정의 경로가 동시에 수행되어야 한다고 할 때 그 중 가장 긴 경로, 즉 전체 공정 중 시간이 가장 많이 걸리는 경로를 뜻한다. 이러한 문제를 해결하기 위해, 상기 데이터 축적 블록은 다중 클럭(multiple clock)을 사용한다.
- <31> 상기 특징 결정 블록(50)은 2개의 클럭(clock) 동안 각각의 디짓들에 대한 매칭된 수를 비교하고, 테스트 패턴과 매칭된 수를 최종적으로 판단하여 해당 클래스를 결정하여 패턴을 인식하고 판단하기 위한 데이터로 사용하게 된다.
- <32> 본 발명의 패턴 인식 장치의 유효성을 보여 주기 위해, C 시뮬레이션을 사용하여 소프트웨어 모델과 하드웨어 모델(본 발명의 패턴 인식 장치)을 비교한다. 본 발명에서 제안된 패턴 인식 장치는 Xilinx사의 xc4vlx100에 의해 입증되었다.
- <33> Verilog HDL을 사용하여 본 발명의 패턴 인식 장치를 구현하였다. 그리고, ModelSim을 사용한 시간 시뮬레이션 결과가 도 3에 도시되어 있다.
- <34> 도 3을 참조하면, Xilinx Virtex-4LX100 FPGA(Xilinx Corp., Virtex-4 User Guide, www.xilinx.com, Jan. 2007. 참조)를 사용하여 100MHz에서 약 4 μ s에 입력 패턴이 인식된다. 소프트웨어 시뮬레이션은 2.4GHz에서 약 12.5s의 시간이 소요된다.
- <35> 같은 주파수에서, 본 발명에 따른 패턴 인식 장치는, 단순 메모리 접근법, 단순 논리 연산법, 파이프라인 구조 및 대규모 병렬 처리 프로세서를 이용함으로써, 소프트웨어 시뮬레이션과 대비하여 정확도 손실없이 약 8 x 10⁷ 배로 인식 속도가 향상된 점을 알 수 있다.
- <36> 이상과 같이 본 발명에 따른 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치를 예시한 도면을 참조로 하여 설명하였으나, 본 명세서에 개시된 실시예와 도면에 의해 본 발명이 한정되는 것은 아니며, 본 발명의 기술사상 범위내에서 당업자에 의해 다양한 변형이 이루어질 수 있음은 물론이다.

도면의 간단한 설명

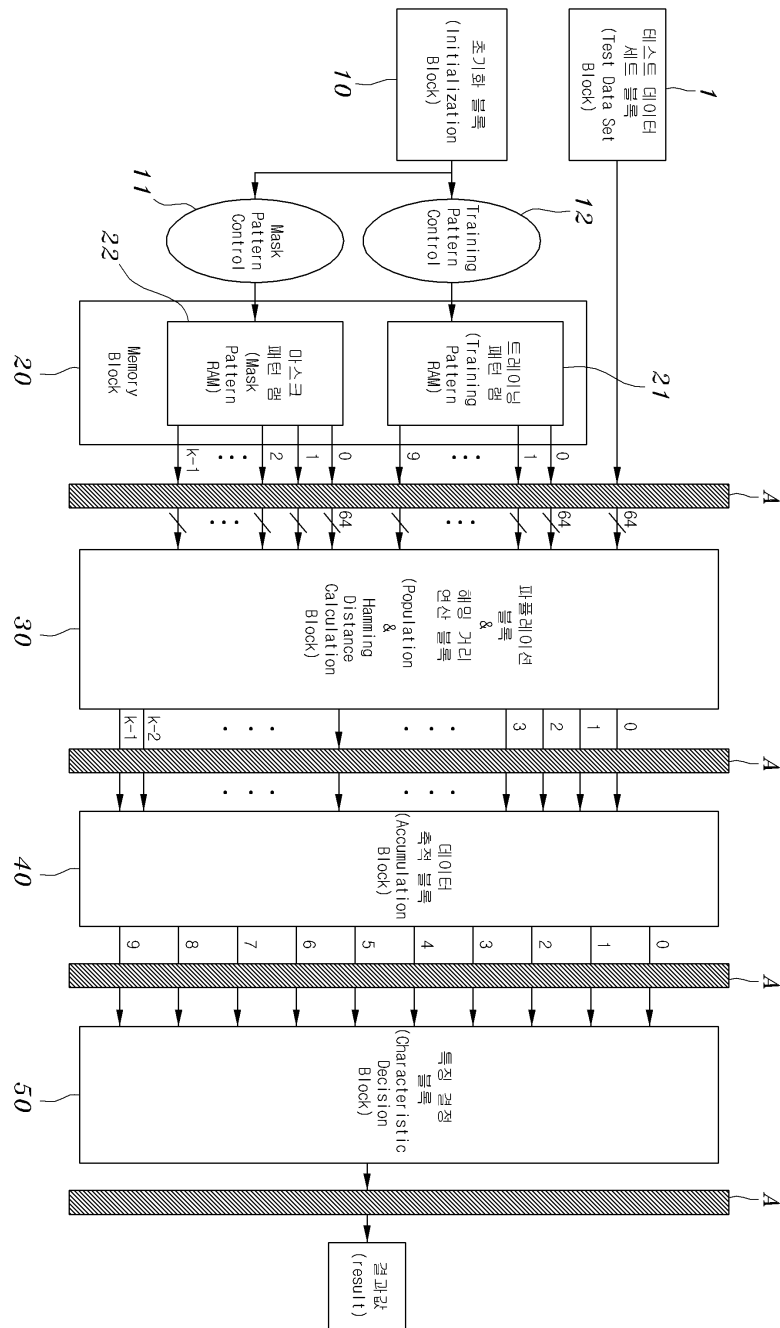
- <37> 도 1은 과플레이션(population) 수에 따라 시뮬레이션된 분자 단위 전개 학습 모델의 인식 정확도를 도시한 도,
- <38> 도 2는 본 발명에 따른 DNA 컴퓨팅 기술에 기초한 패턴 인식 장치의 블록도,
- <39> 도 3은 Verilog HDL을 사용하여 구현된 본 발명에 따른 패턴 인식 장치(프로세서)에 대해 ModelSim을 사용한 시간 시뮬레이션 결과를 도시한 도이다.

도면

도면1



도면2



도면3

