

SNP 조합 인자들의 진화적 학습 방법 기반 질병 관련 복합적 위험 요인 추출

(Identifying Compound Risk Factors of Disease by Evolutionary Learning of SNP Combinatorial Features)

이 제 근 [†]
(Je-Keun Rhee)

하 정 우 ^{††}
(Jung-Woo Ha)

배 설 희 ^{†††}
(Seol-Hui Bae)

김 수 진 ^{††††}
(Soo-Jin Kim)

이 민 수 ^{†††††}
(Min Su Lee)

박 근 준 ^{†††††}
(Keun-Joon Park)

장 병 탁 ^{†††††}
(Byoung-Tak Zhang)

요약 생체 내에서 질병 발생은 다양한 요인들의 복합적인 작용에 의해 발생한다. 하지만, 대부분의 질병 관련 원인을 발견하고자하는 연구들에서는 여러 요인들의 다양한 조합들을 복합적으로 고려하여 분석하기에는 한계가 있는 경우가 많다. 단 하나의 질병 관련 요인들을 찾는 것에 그치고 있다. 본 연구에서는 유전체 정보과 임상 정보를 이용하여 질병 분류 모델 기반 인자 조합들의 진화적 학습 방법을 제안한다. 이 방법을 이용하여 본 연구에서는 질병 관련 주요 인자를 찾고, 이를 시각화할 수 있는 시스템을 만드는 것을 목적으로 한다. 본 연구를 통해 정확도 높은 당뇨병 환자군 분류 모델을 만들고, 당뇨병 발생에 중요한 영향을 미치는 인자들의 조합을 찾을 수 있었다. 또한 생물학적인 분석을 통해 본 연구에서 찾은 인자들의 조합이 실제로 당뇨병 발생에 영향을 미치는 인자가 될 수 있음을 확인하고, 특히 각 인자들이 하나씩 존재할 때보다, 조합으로 존재할 경우 당뇨병 발생 가능성이 높아질 수 있음을 확인할 수 있었다.

키워드 : 기계학습, 진화연산, 생물정보학, 인자추출

Abstract Most diseases are caused by complex processes of various factors. Although previous researches have tried to identify the causes of the disease, there are still lots of limitations to clarify the complex factors. Here, we present a disease classification model based on an evolutionary learning approach of combinatorial features using the datasets from the genetics and cohort studies. We implemented a system for finding the combinatorial risk factors and visualizing the results. Our results show that the proposed method not only improves classification accuracy but also identifies biologically meaningful sets of risk factors.

Key words : Machine Learning, Evolutionary Learning, Bioinformatics, Feature Selection

1. 서론

인간유전체사업(HGP, Human Genome Project)[1]이 완료된 후 포스트 지놈 시대(post-genome era)가 시작됨에 따라 유전체 서열 데이터와 역학 데이터, 그리고 대용량 처리(high-throughput) 실험에 의한 다양한 데이터들이 산출되고 있다. 이러한 데이터들을 효과적으로 분석하기 위한 다양한 분석 기법들이 개발되고 있으며, 최근에는 개인 간의 염기서열 차이정보인 SNP(Single Nucleotide Polymorphism)에 근거하여 질병 예측, 진단 및 예방이 가능한 맞춤 의학(personalized medicine)을 제공하고자 하는 노력들이 이루어지고 있다[2].

기존의 많은 방법들을 활용하여 다양한 분석이 수행

- 본 연구는 질병관리본부 학술연구용역사업(2007-S6- J-002), 지식경제부 한국산업기술평가관리원 산업원천기술개발사업(MARS, KEIT-2009-A1100-0901-1639), 교육인적자원부 학술진흥재단(KRF-2008-314-D00377), 교육인적자원부 BK21-IT, 산업자원부 차세대기술개발사업 분자진화컴퓨팅(MEC) 및 컴퓨터연구소(ICT)에 의해 지원되었음.
- 이 논문은 2009 한국컴퓨터종합학술대회에서 'SNP 조합 인자들의 진화적 학습 방법 기반 질병 관련 복합적 위험 요인 추출'의 제목으로 발표된 논문을 확장한 것임

[†] 학생회원 : 서울대학교 생물정보학 협동과정
jkrhee@bi.snu.ac.kr

^{††} 학생회원 : 서울대학교 컴퓨터공학부
jwha@bi.snu.ac.kr

^{†††} 비 회원 : 질병관리본부 국립보건연구원 유전체센터 바이오과학정보과
shbae1981@gmail.com
park-kj@cdc.go.kr

^{††††} 비 회원 : 서울대학교 생물정보학 협동과정
sjkim@bi.snu.ac.kr

^{†††††} 종신회원 : 서울대학교 컴퓨터공학부
mslee@bi.snu.ac.kr

^{††††††} 종신회원 : 서울대학교 컴퓨터공학부 교수
btzhang@bi.snu.ac.kr

논문접수 : 2009년 8월 14일

심사완료 : 2009년 10월 14일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨팅의 실제 및 레터 제15권 제12호(2009.12)

되고 있지만 질병은 유전적, 환경적 요인 등이 복합적으로 작용하여 발생되므로 여러 가지 복합적인 요소들을 함께 고려하여 분석해야할 필요성이 있다. 하지만 기존 대용량의 생물학 데이터를 분석하는 도구들은 일반적으로는 기본적인 통계적 테스트를 이용하거나 특정 요소 하나와 질병과의 관련성 확인하는 정도의 분석에 그치고 있고[3], 복합 요인을 찾고자 하는 연구들도 최근들어 진행되고 있지만 이런 연구들 역시 한계를 가지고 있는 것이 사실이다[4]. 질병 진단 및 예측, 그리고 질병 관련 주요 인자를 찾는 데에 많이 이용되는 기계학습 알고리즘을 이용한 분석 도구들의 경우에도 데이터에 대한 이해를 돕기보다는 분류 성능 향상에 초점을 맞추고 있다[5]. 또한 특성 선정 작업과 분류 작업이 분리되어 있기 때문에 기존 도구를 이용하는 경우 낮은 데이터 분석 효율이라는 한계를 가지고 있으며, 기존 분석 도구들은 직관적인 인터페이스가 결여되어 있어 생물학자들이 손쉽게 사용하기 힘들다는 단점도 있다.

따라서 본 연구에서는 중요 인자들의 조합을 발견하여 분류작업을 수행하는 진화적 하이퍼망 알고리즘에 기반하여 다양한 종류의 대규모 생물학 데이터(SNP, 임상, 역학 데이터 등)를 분석하고 분석 결과를 시각화할 수 있는 시스템을 설계 및 구축하는 것을 목표로 한다. 진화적 하이퍼망은 패턴 분류[6], 문서 분류[7], 마이크로어레이 데이터 분석[8-10] 등 다양한 문제들에 적용되어 왔으며, 다양한 문제들에서 경쟁력 있는 분류 성능 및 결과를 보여주었다. 본 연구에서는 데이터 인자를 나타내는 노드(vertex)와 연관되어 있는 인자들의 조합을 표현하는 하이퍼에지(hyperedge), 그리고 하이퍼에지의 가중치(weight)로 구성된 그래프 계열 진화 연산 기반 분류 및 예측 알고리즘을 이용하여 당뇨병 관련 주요 SNP 인자 조합을 추출하였다. 본 연구에서 사용한 알고리즘을 데이터에 적용하면 3개 이상의 SNP 인자간의 연관관계와 그 강도를 표현하는 것이 가능하며 이 과정에서 데이터의 목표 항목을 예측하는데 있어 중요한 인자들의 조합들을 찾아냄으로써 특성 추출 작업을 수행한다. 이를 통해 유전체 데이터 분석을 위한 하이퍼망 모델링 기술을 개발하고 실제 SNP 데이터 분석에 적용하여 다른 분석 알고리즘에 비해 우수한 성능을 보임을 확인한다. 또한 하이퍼망 기반 진화적 모델링 방식은 특히 인자들의 조합에 의한 질병과의 관계를 밝히는 데에 적합함을 확인하였다.

이 논문의 나머지 부분은 다음과 같이 구성된다. 2장에서는 연구에서 수행된 실험방법을 설명하고 있으며, 3장에서는 실험결과를 제시한다. 마지막으로 4장에서 결론을 도출하고 향후 연구방향을 요약한다.

2. 실험 방법

2.1 하이퍼망 기반 데이터 모델링 방법

하이퍼망(hypernetwork)은 학습과정에 있어서 진화적 개념을 도입한 그래프 기반의 기계학습 알고리즘으로, 하이퍼망에서의 에지는 2개 이상의 벡터 연결관계를 동시에 표현할 수 있다. 하이퍼망 H 를 벡터 집합 V , 하이퍼에지의 집합 E , 에지 가중치 집합 W 로 표현하면

$$H = (V, E, W) \tag{1}$$

으로 정의된다. 그리고 하이퍼망에서 벡터는 인자(attribute)의 인덱스와 그 값의 쌍(pair)을 의미하고 하이퍼에지는 벡터들간의 연관관계를 표현하게 되며 가중치는 이들 연관관계의 중요도를 표현한다. 또한 하나의 하이퍼에지를 구성하는 벡터의 수를 차수(order)라 정의하며 차수가 k 일 때 이를 k -order 하이퍼에지라 부른다. 데이터의 집합 $D = \{x^{(n)}\}_{n=1}^N$ 이 주어졌을 때, 정규화 값(partition function) $Z(W)$, k -order 하이퍼에지의 모든 경우의 수 $C(k)$ 를 이용하여 하이퍼망은 W 와 D 의 조건부 확률로 식 (2)와 (3)의 형태로 표현된다[11].

$$P(D|W) = \prod_{n=1}^N P(x^{(n)}|W), \tag{2}$$

$$P(x^{(n)}|W) = \frac{\exp \left[\sum_{k=1}^K \frac{1}{\alpha(k)} \times \sum_{i_1, \dots, i_k} w_{i_1, \dots, i_k}^{(k)} x_{i_1}^{(n)} \dots x_{i_k}^{(n)} \right]}{Z(W)} \tag{3}$$

이 때 하이퍼에지의 차수(order)는 하나의 하이퍼에지가 동시에 연결한 벡터의 수를 의미한다. 하이퍼망은 수많은 다양한 차수의 하이퍼에지들로 구성되며 하이퍼에지는 벡터들의 조합(combination)으로 표현된다. 하이퍼망은 벡터들 조합의 멱집합으로 표현되는 매우 큰 조합의 공간을 모델링하려는 시도이며 이렇게 광대한 조합의 공간을 탐색하기 위하여 진화적 방법을 도입함으로써 합리적인 시간 내에 데이터의 패턴을 학습하게 된다[12].

2.2 SNP 데이터를 이용한 하이퍼망 기반 질병 분류 및 주요 인자 조합 추출 방법

2.2.1 데이터 전처리

질환 관련 데이터를 하이퍼망으로 모델링하기 위해서는 전처리 절차가 필요하다. 2.1절에서 설명한 바와 같이 하이퍼망 모델 학습을 위해서는 데이터들을 카테고리 형태로 변환할 필요가 있다. 따라서 나이 및 각종 측정수치와 같이 연속적인 실수값을 갖는 인자들에 대해 이산화 작업 및 병합 후 하이퍼망 모델 생성을 위한 입력데이터로 주어져야 한다.

2.2.2 SNP 데이터의 하이퍼망 모델링

전처리가 완료된 데이터는 알고리즘을 통하여 하이퍼망으로 생성되고 생성된 하이퍼망을 이용하여 당뇨병질환과 관련된 중요한 인자들의 조합을 검색한다. 즉 데이터

를 구성하는 임상정보 및 SNP 정보의 인자가 각각의 벡터를 구성하며 벡터 간의 연관관계가 하이퍼에지 로 표현된다. 이 때 생성된 하이퍼망은 당뇨질환 예측에 중요한 영향을 끼치는 하이퍼에지 즉 인자조합들로 구성되어 있으므로 하이퍼에지 분석을 통하여 중요인자 조합 검색이 가능하다.

2.2.3 SNP 데이터의 하이퍼망 모델링과 진화연산 기반 학습을 이용한 질병예측 및 중요인자 추출

본 연구에서 데이터로부터 당뇨여부를 판단하고 중요 인자 및 인자 조합을 도출해내 과정은 표 1 및 그림 1의 순서에 따라 진행된다.

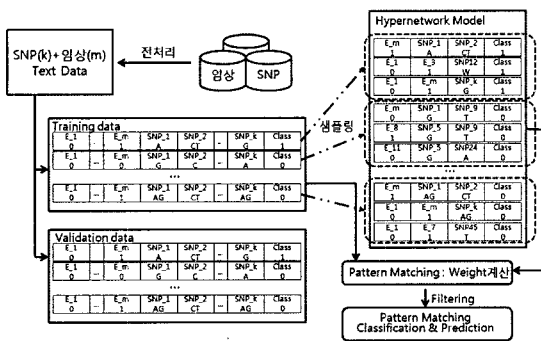


그림 1 하이퍼망 생성 및 학습 과정

표 1 하이퍼망을 이용한 학습 알고리즘

- 1) 분석대상 질병관련 데이터를 학습 데이터(Training data)와 검증 데이터(Validation data)로 분리
- 2) 학습 데이터에 속한 전체 데이터에 대하여 다음과 같은 작업을 수행함으로써 하이퍼망을 생성
 - 2-1) 하나의 샘플로부터 선택될 인자의 수를 랜덤하게 선택
 - 2-2) 선택된 인자의 수만큼 샘플로부터 랜덤하게 인자를 샘플링하여 조합하고 샘플의 클래스(class)값을 추가하여 하나의 하이퍼에지 구성
 - 2-3) 정해진 수만큼 하나의 샘플에 대하여 위 과정을 반복
- 3) 생성된 하이퍼망에 대하여 아래의 작업을 정해진 횟수만큼 반복 수행
 - 3-1) 하이퍼망을 구성하는 하이퍼에지와 학습 데이터 간의 패턴 매칭(pattern matching)을 통하여 하이퍼에지의 가중치를 계산
 - 3-2) 가중치를 기반으로 하여 가중치 값이 상대적으로 낮은 하이퍼에지들을 하이퍼망으로부터 제거
 - 3-3) 학습 데이터와 테스트 데이터에 대하여 패턴 매칭을 통한 분류(classification)를 수행하고 성능 측정

2.3 데이터 및 실험 설정

SNP 데이터는 안성, 안산 지역에서 얻어진 당뇨 관련 유전자에 대한 SNP 정보이다. 전처리 작업 이후에 본 연구에서 이용된 데이터는 표 2와 같다.

표 2 실험에 사용된 SNP 데이터 기본 정보

개요		SNP 1기 데이터		
SNP수	84			
샘플수	전체	정상	환자	
	1454	908	546	

임상정보는 안성안산 코호트 1기 역학정보 공개데이터를 이용하여 추출하였다. 역학정보는 21개의 테이블에 총 1,444개의 변수로 구성되어있다. 본 연구의 1차적인 초점은 SNP 데이터 분석에 두었으므로, 역학 정보는 SNP 분석을 위한 보조적인 자료로만 이용된다. 따라서 본 연구에서는 당뇨병과 직접적으로 연관되어있을 것으로 여겨지는 테이블을 사전에 선택한 후, 각 테이블 별로 상관관계 기반 인자 선택 방법[11]을 이용하여 각 테이블별로 당뇨병과 강한 상관관계를 가지는 특성들을 추출하였다. 실제 분석에 사용된 테이블과 변수들은 표 3에 보인다.

표 3 실험에서 사용된 임상 및 역학 정보

테이블명	변수명	변수 설명
AS1_Gen	AS1_Age	연령
AS1_Anthro	AS1_PLLie2	누운 자세 - 2회 pulse
	AS1_Sub2	견갑하부(subscapular) 2회 측정
	AS1_Weight	체중
	AS1_DZ	골밀도 측정 - Distal Radius Z
AS1_Biochem	AS1_MT	골밀도 측정 - Midshaft Tibia T
	AS1_HBAIC	HbA1C
	AS1_GLU120	Glucose (OGTT 120분 후)

본 연구에서의 실험에서 사용한 파라미터 값은 표 4에 보인다.

표 4 실험에 사용된 파라미터 값

cross-validation	Fold	5
파라미터 설정	Sampling Ratio	10
	Boost Ratio	1
	반복횟수	10
	최소오더	2
하이퍼에지 저장 기준 정확도	최대오더	10
	Training 정확도	90
	Validation 정확도	60
	당뇨 가중치	15

3. 실험 결과

3.1 하이퍼망 기반 유전체 분석 시스템

하이퍼망 기반 유전체 분석 시스템은 그림 2와 같이 구성되며, 데이터 전처리 모듈, 하이퍼망 모델링 및 분석 모듈, 시각화 및 결과 출력 모듈 등으로 구성된다.

시각화를 위한 모델에서는 그림 3과 같이 질병 관련 주요 인자들을 직관적으로 확인할 수 있도록 구성하였다.

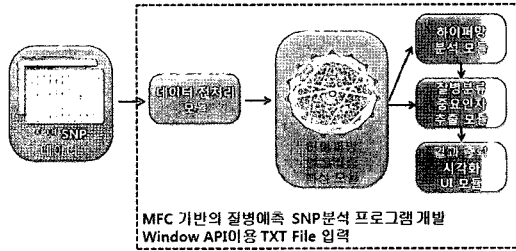


그림 2 유전체 분석 시스템 구성 개요도

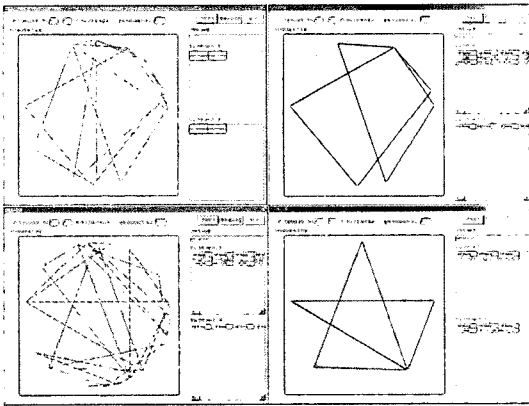


그림 3 시각화 화면

3.2 분류 성능 비교

본 연구에서의 분류 성능은 표 5와 같다. Accuracy는 전체적인 예측 정확도를 의미하며, Sensitivity는 실제 질병을 가진 사람을 정확하게 예측할 확률, Specificity는 정상을 정확히 예측할 확률을 의미하는 것으로 볼 수 있다. 일반적으로 많이 사용되는 다른 분류 알고리즘들에 비해 비슷하거나 우수한 성능을 보임을 알 수 있다.

표 5 각 방법론에 따른 분류 성능 비교

	하이퍼망	SVM	Naive Bayes	Decision Tree
Accuracy	77.72	74.90	73.66	76.34
Sensitivity	0.620	0.562	0.645	0.648
Specificity	0.871	0.861	0.792	0.833

3.3 당뇨 관련 주요 인자 추출

하이퍼망을 이용한 분류에 주요한 영향을 주는 인자들 중 질병으로 분류하는 데 주요한 인자들의 조합을 표 6에 보인다. 인자들의 조합은 2~10의 크기로 다양하게 만들어지지만, 컴퓨터를 이용한 계산 및 분석의 용이

표 6 질병 출현 인자 리스트

변수명			Number
IRS2_01	PRKCZ_02	PRKCZ_05	28
IRS2_02	RAPGEF1_03	PDK1_01	26
RAPGEF1_11	INSR_08	PTPN1_07	25
AKT1_06	PTPN1_02	CAP1_02	25
IRS2_01	IRS2_02	PRKCZ_02	24
AKT1_03	IRS2_03	RAPGEF1_03	24
IRS2_04	PTPN1_05	INSR_08	24
INSR_10	IRS1_02	CBL_02	24
IRS2_04	PRKCD_01	PRKCZ_02	24

함을 위해 인자 추출 시에는 이들을 3개 인자의 조합으로 분해하여 분석하고, 각 인자 조합의 출현 빈도를 표시하였다.

가장 상위에 나온 SNP 데이터는 IRS2_01, PRKCZ_02, PRKCZ_05이다. 이들 SNP 정보는 두 유전자 Insulin receptor substrate와 Protein kinase C zeta의 서열로부터 얻어진 SNP 정보이며 이 두 유전자는 실제로 당뇨와 직접적으로 관련될 수 있음을 추정할 수 있다. 실제로 KEGG 데이터베이스에 있는 type II 당뇨 관련 대사경로(metabolic pathway)를 보면, IRS와 PKC zeta는 직접적으로 대사 경로 상에서 상호작용(interaction) 하며, 이를 통해 직접적으로 당뇨 발생과 연관될 수 있다[13]. 즉 대사경로 상에서 상호 작용하는 두 유전자의 변이가 당뇨 발생과 직접적인 연관이 있을 수 있음을 하이퍼망 모델 결과로도 확인할 수 있다.

다음으로 순위가 높은 인자 조합은 IRS2_02, PKD1_01, RAPGEF1_03이다. IRS2_02는 AA/AG/GG의 형태로 변이(variation)가 존재하며, PKD1_01은 AA/AT/TT, RAPGEF1_03은 AA/AG/GG의 형태로 변형이 일어난다. 하지만 인자들 각각과 당뇨 발생의 연관성을 확인해보면 유의한 수준의 연관 관계를 확인하기 어렵다. 하지만 각각의 조합에 대한 연관 관계를 살펴보면 확연한 차이를 보인다. 그림 4는 각 요인별 당뇨 발생 확률과 이들에 조합에 의한 확률을 비교한 것이다. 각각의 변이

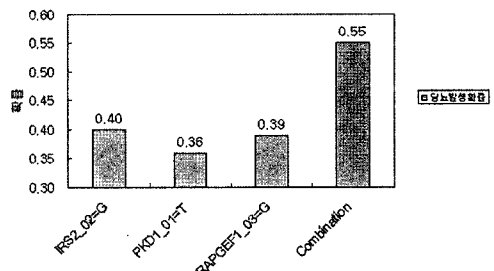


그림 4 각 인자 별 SNP 형태 및 이들의 조합에 따른 질병 발생 확률

하나만을 고려하였을 때는 당뇨 발생 확률이 낮으나, 이들의 조합을 비교하면 당뇨 발생 확률이 확연히 높아지는 것을 확인할 수 있다.

이와 같은 분석을 통해, 본 연구에서 수행한 하이퍼망을 이용한 방법은 질병 관련 요인들의 복합 효과를 추출해주며, 또한 질병 발생이 주요 요인들의 복잡한 복합 효과에 의해 일어난다는 것을 다시 확인할 수 있다.

4. 결론

본 연구에서는 대규모 데이터에 대한 분류 및 예측 모델을 구축하고 분류에 있어 영향을 주는 복합적인 인자들의 조합을 발견할 수 있는 하이퍼망 모델링 기술을 기반으로 데이터 분석 도구 및 시각화 도구를 개발하였다. 본 연구를 통해 개발된 하이퍼망 모델링 시스템은 생물학자들도 쉽게 데이터를 분석할 수 있는 인터페이스 환경을 제공하며, 발견된 인자들의 조합을 시각화하여 직접적이며 직관적인 분석이 가능하도록 하였다.

또한 본 연구에서는 분류 모델만 만들어내는 기존의 기계학습 알고리즘과는 달리 특정 질병의 유전적, 환경적 요인들의 복합적 작용들에 대한 분석 및 요인들의 조합 발견을 가능하게 하였다. 구축된 시스템을 정상인과 당뇨 환자에 대한 SNP 데이터와 임상 및 역학 데이터 분석에 적용하여 당뇨 진단 및 예측을 위한 하이퍼망 모델링을 수행하였고, 전반적으로 우수한 성능을 보임을 확인하였다. 하이퍼망 모델링 과정에서 추출된 SNP 조합들은 type II 당뇨 발생 관련 대사경로에서 매우 직접적으로 연결되어 있었으며, 각 SNP 수준에서의 유전적 변이는 당뇨 발병과의 연관성이 떨어지지만 서로 특정 형태로 변이의 조합을 이루었을 때에는 당뇨 발병 가능성이 높아지는 경향을 확인할 수 있었다. 이처럼 하이퍼망 모델을 통해 발견된 높은 가중치를 가진 SNP들의 복합적 조합들은 당뇨를 진단하기 위한 후보 조합 마커(candidate compound marker)로 사용될 수 있을 것이다. 또한 이를 통해 유전자 치료에 사용될 수 있는 후보 유전적 변이들을 발견할 수 있는 가능성을 확인할 수 있었다.

본 시스템은 다른 질병 데이터를 비롯한, 다양한 다른 생물학에도 적용 가능하므로 SNP 데이터뿐만 아니라 다양한 데이터 분석에도 사용될 수 있는 확장성이 있어 대규모 다양한 바이오 데이터 분석 작업의 효율을 매우 높여주고, 실제로 질병 발생 원인의 이해를 높여줄 수 있을 것으로 기대된다.

참 고 문 헌

[1] F. S. Collins, et al. The Human Genome Project: Lessons from Large-Scale Biology, *Science*, vol.300,

- pp.286-290, 2003.
- [2] K. Bhasi, et al., Information-theoretic identification of predictive SNPs and supervised visualization of genome-wide association studies, *Nucleic Acids Res.*, vol.34, no.14, pp.e101, 2006.
- [3] R. Sladek, et al., A genome-wide association study identifies novel risk loci for type 2 diabetes, *Nature*, vol.445, pp.881-885, 2007.
- [4] J. Lee, et al., Analysis of asthma related SNP genotype data using normalized mutual information and support vector machines, *Journal of KIISE : Software and Applications*, vol.36, no.9, pp.691-696, Sept. 2009. (in Korean).
- [5] N. Tiffin, et al., Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes, *Nucleic Acids Res.*, vol.34, no.10, pp.3067-3081, 2006.
- [6] J.-K. Kim, B.-T. Zhang, Evolving hypernetworks for pattern classification, *IEEE Congress on Evolutionary Computation (CEC 2007)*, pp.1856-1862, 2007.
- [7] S. Kim, et al., Text classifiers evolved on a simulated DNA computer, *IEEE Congress on Evolutionary Computation (CEC 2006)*, pp.9196-9202, 2006.
- [8] C.-H. Park, et al., Finding cancer-related gene combinations using a molecular evolutionary algorithm, *IEEE 7th international conference on Bioinformatics & BioEngineering (BIBE 2007)*, pp. 158-163, 2007.
- [9] S. Kim, et al., Evolving hypernetwork classifiers for microRNA expression profile analysis, *IEEE Congress on Evolutionary Computation (CEC 2007)*, pp.313-319, 2007.
- [10] C.-H. Park, et al., Use of evolutionary hypernetworks for mining prostate cancer data, *8th International Symposium on advanced Intelligent Systems (ISIS 2007)*, pp.702-706, 2007.
- [11] B.-T. Zhang, "Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory," *IEEE Computational Intelligence Magazine*, vol.3, no.3, pp.49-63, 2008.
- [12] M. A. Hall, Correlation-based feature selection of discrete and numeric class machine learning. *International Conference on Machine Learning*, pp 359-366, 2000.
- [13] KEGG: Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.jp/kegg/>