신경망을 이용한 갑상선 암 METABOLIC PATHWAY 분석

연세대학교 대학원
의학전산통계협동과정
길 성 화

신경망을 이용한 갑상선 암 METABOLIC PATHWAY 분석

지도교수 김 동 기·윤 창 노

이 논문을 석사 학위논문으로 제출함

2004년 12월 일

연세대학교 대학원

의학전산통계협동과정

길 성 화

길성화의 석사 학위논문을 인준함

심사위원 _____ 인

심사위원 _____ 인

심사위원 _____ 인

연세대학교 대학원

2004년 12월 일

제 목 차 례

그 림 차 례

표 차 례

국 문 요 약

신경망을 이용한 갑상선 암 METABOLIC PATHWAY 분석

갑상선암 발병의 중요 호르몬을 추정하기 위해서 49명의 피실험자를 대상으로 Steroid Pathway 관련 일부 Hormone의 농도를 측정하였다. 이 자료를 토대로 KEGG(Kyoto Encyclopedia of Genes and Genomes) Metabolic Pathway와 문헌 연구를 통하여 Network Analysis Format에 맞게 재구성하였다.

측정된 26개의 Hormone을 대상으로 응용된 Neural Network에 분석변수로 사용하여 단위별 Network를 구성하였다. 구성된 Network에서 Weight값을 정상군과 환자군의 차를 나타내는 지표를 사용되는 DMSW(Differece Mean Square of Weight)를 계산하였다. 계산된 DMSW값을 Clustering(K-means Clustering) Algorithm을 사용하여 4개의 Cluster들을 추정하였다. 추정된 각 Cluster들을 Association Algorithm을 통해 Network Relation을 추정한 결과 갑상선암 질병 발명에 관련된 Steroid Hormone Pathway 상에서, 추정된 4개의 Cluster들의 몇 가지 규칙을 찾아내게 되었다.

핵심되는 말 : 갑상선암, Metabolic Pathway, Hormone, Neural Network, Weight, DMSW, Clustering, Association Algorithm

제 1 장 서 론

갑상선 질환은 대사 질환의 대표적인 질병으로 크게 2가지 형태로 나타난다. 먼저, 갑상선 기능의 이상에 의한 것으로 이는 갑상선 Hormone이 과다 분비 되거나 부족할 때 발생하는데 갑상선 기능항진증과 갑상선 기능 저하증으로 나뉘어진다. 다음으로, 갑상선 종양이 있는데 갑상선의 기능에 상관없이 갑상선에 혹이 생기는 질환이다. 갑상선 질환의 원인은 아직 확실하게 밝혀져 있지 않다. 하지만 여성 Hormone인 Estrogen 영향 때문이라는 주장이 존재한다. 갑상선질환은 사춘기 이전에는 남녀 차가 없다가 여자가 월경을 시작하는 나이부터 남녀 차이가 증가해서 30-40대에는 여자 환자가 월등히 많아진다. 그러다 여성이 폐경에 이르는 60대에 이르면 발병률이 확 준다. 여자의 Estrogen 분비 사이클과 맞물려 갑상선 질환 발생도 변화하는 것을 알 수 있다. 실험을 통해서도 갑상선질환이 여성 Hormone 영향 때문이라는 간접적 증거도 있다. 또, 갑상선 기능 항진증을 앓는 환자는 '임신기간 중 증세가 좋아졌다'거나 '출산 후 다시 몸이 나빠졌다'고 호소하기도 한다. 전문가들은 갑상선질환 발병에 유전과 환경요소가 각각 8대 2 비율로 영향을 미친다고 말한다.

이에 본 논문에서는 한국과학기술연구원의 도핑센터에서는 갑상선암 발병의 중요 Hormone을 추정하기 위해서 49명의 피 실험자를 대상으로 Androgen Hormone, Estrogen Hormone, Steroid Hormone 일부인 26개의 Hormone을 측정하였다. 이 자료를 토대로 Hormone Pathway 정보(KEGG DB, ExPASy DB, Metabolism 서적)를 바탕으로 Network Analysis Format에 맞게 재구성하였다.

본 논문에서의 분석방법론은 측정된 26개의 Hormone을 대상으로 다른 통계적 모형에 비해 유연성이 좋은 응용된 Neural Network에 분석변수로 사용하여 소단위 Network를 구성한다. 구성된 Network의 Weight값을 정상군과 환자군의

차를 나타내는 DMSW(Difference Mean Square of Weight)를 추정하여 각 Hormone을 Clustering(K-menas Clustering)한다. 추정된 각 Cluster들을 Association Algorithm을 통해 Network Relation을 추정하여 질병 발병에 몇 가지 규칙을 찾아내고자 한다.

제 1 절 Hormone 및 효소

1.1 Hormone

Hormone이란 내분비선으로부터 분비되는 체내 기관의 생리적 기능을 조절하는 물질을 총칭하며, 혈액 중에 분비되어 표적 세포로 운반되어 미량으로 특수한 영향을 미치는 물질로 신체의 성장, 분화, 및 대사에 관여한다. Hormone은 그 화학적 조성에 따라 크게 3가지로 분류된다. 즉 1) insulin이나 부신피질자극 Hormone 등의 펩타이드 Hormone 2) 티록신이나 아드레날린 등의 페놀 유도체 Hormone, 3) 코르티코이드, 안드로겐 및 에스트로겐과 같은 스테로이드 Hormone으로 나뉘어 진다. 이러한 Hormone은 뇌하수체 부신, 갑상선, 부갑상선, 정소, 난소 등의 여러 기관에서 형성되어 또 다른 여러 기관에 영향을 준다. 뿐만 아니라, 신체 전체적으로 서로 협동 또는 길항적으로 작용한다. 즉, Hormone은 체내의 상태를 일정하게 유지하는 작용(항상성) 메카니즘이 있고 내분비계, 자율신경계, 면역계가 중요한 역할을 한다. 그러므로, Hormone은 항상 일정한 농도를 유지하기 위하여, Hormone이 이용 및 배설 작용으로 감소하면 바로 내분비 기관이 신호를 받아 Hormone의 분비를 촉진시키고, 반대의 항진으로 인해 Hormone이 과다하게 분비되면 내분비 기관이 Hormone의 분비를 억제하게 된다.

스테로이드 Hormone은 주로 부신으로부터 분비되고 생체 내에서의 작용에 따라 크게 몇 가지로 나뉘어 진다. 1) 코르티코이드 : 글루 코르티코이드는 당질대사에 관여하여 기초 대사를 유지하고, 미네랄 코르티코이드는 무기 염류의 대사에 관여한다. 2) 안드로겐, 에스트로겐 : 단백질 동화 작용을 촉진하고 제2의 성

징 발현, 생식 기능을 유지한다.

스테로이드는 아세테이트로부터 시작되어 생합성되는 콜레스테롤 전구물질로 하여 대사가 진행된다. Desmolase 효소에 의해서 콜레스테롤은 프레그넬론으로서 전환을 시작으로 여러 가지 효소들의 영향을 받아 인체 내에서 대사과정을 거치게 되며, 최종적으로 코르티코이드, 안드로겐 및 에스트로겐을 형성한다. 한편, 이들의 일부는 그 자체로, 일부는 주로 간과 장관 등에서 비 활성화되고, 또한 클루쿠론산이나 황산에 포함되어 뇨 또는 대변으로 배설된다. 이러한 내분비계를 조절하는 다양한 스테로이드 Hormone은 효소의 결핍으로 인해 그 생체 질서가 어긋나면서 여러 가지 질환을 발생하게 된다. 즉, Hormone의 양이 적절하게 유지되지 못하고 과분비 혹은 부족하게 되면 기초 대사, 성징 및 신장 기능 등에 이상이 생기게 된다.

1.2 효소

효소는 생물학적 촉매로 화학반응을 촉진 일반적으로 촉매 하는 화학반응의 형식에 따라 분류된다. 국제생화학연합 효소위원회에서는 1978년에 2,132 종류의 효소에 대하여 새로운 효소번호(enzyme code number)에 의한 분류법을 제정하였다. 이 분류법은 효소번호, 권장명, 계통명 순으로 쓰며 alcohol dehydrogenase의 경우를 예를 들면 다음과 같이 표시된다.

[효소번호] [권장명] [계통명] EC 1.1.1.1.
alcohol dehydrogenase alcohol: NAD^+ oxidoreductase

여기서 첫번째 숫자는 효소가 촉매하는 화학반응의 형식을 나타내는 것으로 다음과 같이 6군으로 분류된다.

EC 1군 : 산화환원반응을 촉매하는 효소(oxidoreductase)

EC 2군 : 작용기의 전이를 촉매하는 효소(transferase)

EC 3군 : 가수분해반응을 촉매하는 효소(hydrolase)

EC 4군 : 이탈반응과 부가반응을 촉매하는 효소(lyase)

EC 5군 : 이성화반응을 촉매하는 효소(isomerase)

EC 6군 : ATP 등의 가수분해와 공역하여 2개의 저분자를 연결하는 합성 반응을 촉매하는 효소(ligase)

2, 3번째 숫자는 반응을 더욱 세분화하여 기질의 형식이나 종류에 의해 분류되고 4번째 숫자는 각 효소에 대해 효소위원회가 부여한 일련번호이다.

본 논문에서 표기의 편리성을 위해 재 명명하였다.

표1 효소 재 명명

ENZYME	NAME	ENZYME	NAME	ENZYME	NAME
1.1.1.145	e1	1.13.99.6	e10	1.3.1.3	e18
1.1.1.146	e2	1.14.13.-	e11	1.3.1.30	e19
1.1.1.50	e3	1.14.15.4	e12	1.3.99.5	e20
1.1.1.51	e4	1.14.15.5	e13	1.3.99.6	e21
1.1.1.53	e5	1.14.99.-	e14	4.1.2.30	e22
1.1.1.62	e6	1.14.99.10	e15	2.1.1.-	e23
1.1.1.63	e7	1.14.99.9	e16	2.8.2.15	e24
1.1.1.64	e8	1.3.1.23	e17	5.3.3.1	25
1.1.12.239	e9				

제 2 절 대사체학 및 PATHWAY DATABASE

2.1 대사체학

대사체학 (METABOLOMICS)이란 생물표현형 (PHENOTYPE)을 가장 잘 나타내는 정량 할 수 있는 소분자로 이들의 집단체를 대사체라 하며, 특정 생리 및 병리적 상태에서 대사체 변화를 분석한 후 이를 생리적 기능과 연관 지어 해석함으로써 그 변화 원인을 규명하는 새로운 기술을 말한다. 특히, 대표적인 대사 질

환의 하나인 갑상선에 관련된 질환들은 갑상선 hormone과 함께 다양한 내인성 물질 및 무기 물질과 긴밀한 관계를 형성하기 때문에 hormone 혹은 무기 물질의 투약에 의한 내인성 물질들의 변화 및 관계에 대한 연구가 활발히 진행되고 있으며, 최근 들어 갑상선 hormone 외에 갑상선에 관련된 다른 다양한 hormone들과 내인성 물질들과의 관계에 대한 연구들이 활발히 진행되고 있다.

Biological Pathway Consortium에서는 Bio Pathway를 생화학 신체 조직기관 안의 모든 형태의 분자적 상호작용들과 프로세스들을 포함하는 포괄적인 의미의 용어라 정의한다. 생화학적인 경로들은 서로 긴밀하고 유기적으로 얽혀있지만, 크게 화학물질들의 효소반응으로 일어나는 물질 수송과 에너지 변환에 관한 신진대사경로(metabolic pathways), 세포 주기와 특정 유전자의 정보 발현에 관한 신호 전달경로(signal transduction networks), 유전정보 전달과 발현을 위한 처리과정에 관한 유전자 조절 경로(gene regulation networks)로 분류할 수 있다. 이 중에 본 논문에서는 신진 대사경로를 바탕으로 분석을 하겠다.

2.2 PATHWAY DATABASE

2.2.1 KEGG DATABASE

KEGG(Kyoto Encyclopedia of Genes and Genomes)는 교토대학의 Institute for Chemical Research에서 Japanese Human Genome Program의 일부로 분자의 경로와 세포의 처리과정을 계산하기 위한 수단을 제공하기 위해 만들어졌다. 경로에 관련된 모든 정보들을 현재 급속도로 진행 중인 유전체 프로젝트의 결과물들과 잘 조합하여 제공하고 있으며, 자바 기술 등 최신 전산 기술들을 이용하여 사용자들에게 편리한 인터페이스를 제공하는 것을 목표로 하고 있다.

분자 생물학과 세포 생물학의 연구를 통해 밝혀진 경로에 관련된 정보들을 모아 구축한 신진대사 경로와 조절 경로에 관련된 정보들을 구분하여 제공하고 있는 경로 DATABASE를 운영하고 있으며, 경로에 관련된 유전자들 중에 서열이 밝혀진 유전자들의 유전자 카달로그를 만들어 제공할 수 있는 유전자 데이터 베

이스, 경로에 관련된 화합물들을 정리하여 경로 DATABASE와 결합시킨 리간드 DATABASE도 함께 제공한다.

2.2.2 ExPASy DATABASE

SIB(Swiss Institute of Bioinformatics)에서 단백질 서열과 구조를 분석하기 위해 구축한 단백체학 서버인 ExPASy(Expert Protein Analysis System)에서는 신진대사 경로 지도로 유명한 'Boheringer Mannheim'사의 'Biochemical Pathway Map'의 웹 버전을 제공한다. 이 경로 지도는 모든 유기체상에서 나타나는 신진대사 경로를 하나의 지도에 표현하고 있다. 효소 활동의 증감과 속도를 표현하고 있으며, 이화 관계와 동화 관계를 구분한다. 경로 안에서 나타나는 화합물들의 화학식도 경로지도 상에 나타내고 있다. 또한 가역반응에서 더 자주 일어나는 반응 방향 등을 Pathway 지도에 표현하여 한번에 종합적인 정보를 표현할 수는 있으나 그렇게 함으로 발생하는 이해하기가 복잡하다는 단점을 지니고 있다.

제 3 절 Neural Network Algorithm

3.1 Neural Network의 정의

Neural Network이란 인간의 두뇌 작용을 신경 세포들 간의 연결 관계로 모델링 한 것이다. Neural Network은 인간의 뇌에 대한 정보처리 과정을 단순히 모방해 보자는 취지에서 출발하였으며, 생물학적 Neuron의 구조 및 기능을 단순화하여 수학적 모델로 표시하고 이 Neuron 모델을 상호 연결시켜 망을 형성한 것이다. 어떤 컴퓨터 용어 사전에서는 Neural Network의 뜻을 사람 뇌의 동작에 가깝게 만든 프로그램이나 데이터 구조 시스템이라고 정의하고 있는데, 다른 말로 표현하면 뇌의 구조와 기능에 대한 이해를 바탕으로 뇌가 수행하는 연산 기능의 원리로부터 새로운 추출해 구현한 시스템이라고 할 수 있다.

그림1은 생물학적 신경 세포의 모습을 나타내주고 있다. Neuron의 각 부분들이 하는 일은 다음과 같다. 수상돌기(Dendrites)는 다른 Neuron으로부터 오는 신호를 시냅스를 통해 전달받는 기능을 한다. 신경절(Synapse)은 받아들인 자극을 강도에 따라서 증폭 또는 축소하는 기능을 수행하는 부분이다. 세포체는 활성적인 입력신호와 억제적인 입력신호를 더하는 기능을 수행한다. 축색돌기(axon)는 세포체의 점화에 의해 발생되는 전기적 에너지 의해 신호를 다른 Neuron으로 전달하는 임무를 맡고 있다. Neuron간의 정보교환은 모두 시냅스를 통하여 행하여지며 정보의 전달은 항상 한쪽 방향을 향한다.

그림 1 생물학적 신경세포

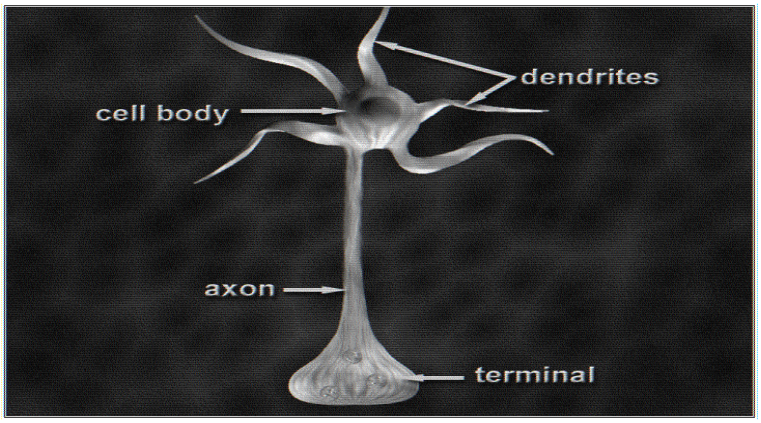
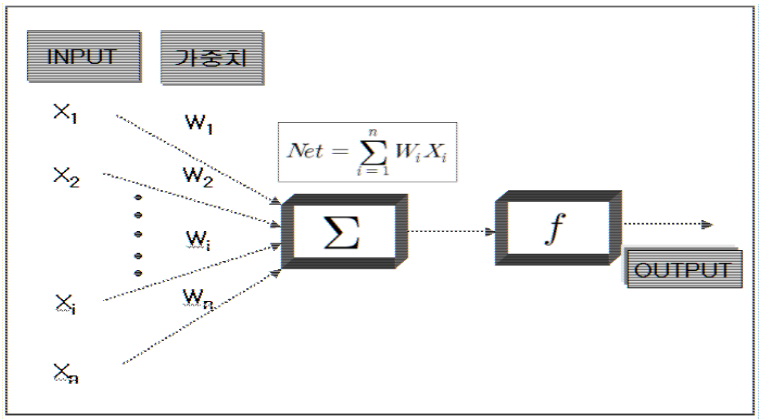


그림1에 나타난 생물학적 신경 세포를 본뜬 인공 신경 세포의 구성은 그림2와 같다.

인공 신경세포는 생물의 신경세포가 가지는 일차적인 특성들을 흉내 내도록 설계되었다. 요약하면, 다른 신경세포의 출력들인, 일단의 입력들이 주어진다. 각 입력들은, 시냅스의 연결강도에 해당하는, 각각의 Weight와 곱해지게 되고, 가중치가 적용된 모든 입력들이 신경세포의 활성도를 결정하기 위해 합해진다. Neural Network의 다양성에도 불구하고, 거의 대부분이 이러한 외형에 바탕을 두고 있다.

그림 2 인공 신경세포



신경 세포는 n개의 입력에 대해 입력 벡터 $X(=x_1,x_2,...,x_n)$와 가중치 벡터 $W(=w_1,w_2,...,w_n)$를 가지고 있다. 신경 세포 기능부위에서는 입력에 대해 입력·가중치 곱의 합을 구하는 일과 구해진 합을 활성함수에 전달해 최종 출력 값을 결정하는 일을 수행한다.

3.2 Neural Network의 역사

Neural Network 모델의 시초는 1943년 McCulloch와 Pitts의 논문에서 찾을 수 있다. 그들은 인간의 두뇌를 수많은 신경세포들로 이루어진 잘 정의된 컴퓨터라고 여겼다. 그들은 단순한 논리적 업무를 수행하는 모델을 보여주었고, 또한 패턴분류 문제가 인간의 지능적인 행위를 규명하는 이론에 매우 중요하다는 것을 인식하였다.

Hebb의 학습규칙은 두 Neuron 사이의 Weight를 조정할 수 있는 최초의 규칙이다. 이 규칙은 학습에 관한 연구를 발전시켰으며 Adaptive Neural Network 연구에 많은 영향을 끼쳤다.

Rosenblatt은 1957년 Perceptron이란 최초의 Neural Network Model을 발표하였는데, 여기서는 학습 프로세스에 알파강화 규칙을 사용하였다. Perceptron에 대한 관심의 주된 이유는 어떤 프로세스에 알파강화 규칙을 사용하였다.

Perceptron에 대한 관심의 주된 이유는 어떤 타입의 패턴이 입력층에 주어졌을 때 이 모델이 반응하게 하는 Weight 강도의 집합을 스스로 발견하는 자동적인 절차에 있다. 학습은 현재 주어진 입력 행렬에 대하여 현재의 각 Weight강도를 조절함으로써 얻어질 수 있다. 이러한 Perceptron 모델은 그 당시에는 매우 가능성이 큰 것으로 여겨졌으나, 그 후 XOR함수와 같이 단순한 비선형 분리 문제도 풀 수 없는 것으로 밝혀졌다.

Minsky와 Papert가 Perceptrons란 저서에서 Perceptron모델을 수학적으로 철저히 분석하여 그 모델의 단점들을 밝혀내고 난 후 Neural Network에 관련된 연구는 약 20년간 침체의 길을 걷게 되었다. 거의 같은 시대에 출발하였으며 Symbolic 처리를 하는 인공지능은 그 후 급속히 발전하였다. 그 동안 신경망은 인공지능에 비해 매우 낮은 정도의 관심과 미미한 연구자금 등으로 어려움을 겪었는데, 1970년대 말과 1980년 초반에 들어 Kohonen, Hopfield, Kirkpatrick, Hinton, Grossberg, Rumelhart등이 Neural Netowork을 다시 활성화시켰다.

Perceptron과 같이 하나의 Single-Adjustable Layer으로 구성된 모델들의 한계점들 때문에 입력층, 출력층, 그리고 한 개 또는 그 이상의 Hidden Layer을 쓰는 새로운 모델들이 1980년대 중반에 제안되었으며, 특히 PDP(Parallel Distributed Processing) 그룹에 의한 폭 넓은 연구가 있었다. 이 그룹에서 제안한 모델은 Hidden Layer와 Back Propagation 학습 Algorithm을 사용함으로써 선형 분리문제 뿐만 아니라 여러 가지 문제점들을 해결할 수 있는 계기를 마련하였다. Back Propagation 학습 Algorithm은 오차를 정정하는 규칙으로서, 입력에 대해 원하는 반응과 실제로 얻어진 것들에 대한 차이를 줄여 나가는 것이다. Error Propagation에 의한 내부표현 학습에서, 입력패턴은 충분한 Hidden Unit들만 있으면 항상 코드화될 수 있다. 이 과정은 Network의 Weight 강도를 반복적으로 조정하여 실제 Neural Network의 Vector와 원하는 출력간의 차이를 줄여 나간다.

현재 Neural Network에 대한 연구는 여러 가지 다양한 논제와 연구가 이루어지고 있으며 여러 분야들에 응용되고 있는데, Hopfield 모델과 Boltzmann Machine등도 역시 Neural Network에서 매우 중요한 모델들이다. 튜토리얼 수준

의 Neural Network에 대한 연구내용은 Lippman에서 찾을 수 있다.

Neural Network와 Symbolic한 처리를 하는 인공지능은 어떤 주어진 문제에 대하여 서로가 전혀 다른 접근 방법을 쓴다. Neural Network는 생물학적인 시스템에서 영감을 얻어 정보처리 시스템의 구조화에 관심이 있는 반면, 인공지능은 어려운 문제를 풀기 위해 여러 가지 형태의 지식을 표현하고 유추하는데 많은 관심을 가지고 있다. 1960년경 인공지능 연구의 초기 단계에서 Learning Function을 연구하는데 상당한 열정을 보였다. 그러나 초기 단계에서의 낙관적인 기대에도 불구하고 만족할 만한 결과를 얻지 못했다. 학습 기계에 대한 초기 단계에서의 모델링 접근은 상당히 좋았으나, 그 모델들에서 몇 가지 중요한 성질들이 제외되었다. 첫 번째는 Associative Memory이다. Associative Memory장치는 영상이나 신호 패턴들의 이부 또는 다른 단서들로부터 전체적인 표현을 추출할 수 있는 장치를 말한다. 두 번째는 처리장치들의 공간적 순서의 중요성이 완전하게 이해되지 않았다는 것이다. 생물학적 두뇌에는 상당수의 정보가 처리 요소들의 공간주소에 코드화되어 있다는 것을 간과한 점이다.

3.3 Neural Network의 구성요소

생물학적 Neural Network이 단순한 신경세포들의 대단위 병렬연결로 이루어져 있듯이, Neural Network도 단순한 기능을 수행할 수 있는 신경세포들의 수많은 병렬연결로 이루어져 있다. 기능면에 있어서도 생물학적 Neural Network과 마찬가지로 병렬 분산 처리를 할 수 있을 뿐만 아니라, 학습이나 훈련을 통해서 연결강도를 조정하여 정보를 추가하거나 변경할 수 있는 적응특성을 가지고 있다.

Neural Network의 기본요소는 Neural Network의 모델, node 특징, 학습규칙 등이 있는데 Neural Network의 모델은 신경세포 또는 처리 소자들 간의 연결 및 상호 작용을 정의한 것이며 node 특성은 각각의 node에서 특성을 의미한다. 학습규칙은 Neural Network이 적절한 결과를 만들어 내도록 하는 학습규칙 또는 훈련 알고리즘이다.

또한 Neural Network은 입력층, 출력층, 가중치, 임계치로 구성되어 있는데 입력

층은 외부의 입력을 받아들이고 출력층은 Neural Network의 처리결과를 보내는 역할을 수행하며 가중치는 각 층들을 연결하는 Weight이며 Cut-Off Value는 Neuron의 값을 결정하는 경계 값이다.

Activation Fun.는 입력 값들을 모아서 하나의 결과 값으로 출력하는데 Step fun., Sigmoid Fun., Hyperbolic Fun.등이 있다.

3.4 Neural Network 모델

Neural Network 모델은 인공 신경 세포, 처리 소자들 간의 상호 작용 및 연결 방법 등에 대한 정의를 말한다. 신경회로망에 대한 연구가 시작되고 여러 시행착오를 거치면서 다양한 Neural Network 모델들이 만들어지게 되었다. 아래의 표2 는 헥트(Hecht)-닐센(Nielsen)이 분류한 Neural Network 모델들을 나타내 주고 있다.

표 2 헥트-닐센이 분류한 Neural Network 모델

모델	연구 개발자	연도	주요 응용분야
Perceptron	F. Rosenblatt	1957	인쇄체 문자인식
Madaline	B. Widrow	1960~1962	적응적 변복조 장치
Avalanche	S. Grossberg	1967	연속적인 음성 인식, 로봇 팔에 대한 기계 명령어 교육
Brain State in a Box	J. Anderson	1977	DATABASE에서 지식 추출
Self-organizing Map	T. Kohonen	1980	서로 다른 기계학적 영역으로 매핑, 패턴 분류
Cerebellatron	D. Mar, J. Albus, A. Pellionez	1969~1982	로봇 팔에 대한 기계 작동 제어
Hopfield	J. Hopfield	1982	부분적인 자료로부터 완전한 데이터나 영상

			을 검색
Neocognitron	K. Fukushima	1978~1984	손으로 쓴 문자 인식
Back-Propagation	P. Werbos, D. Rumelhart	1974~1985	문자인식, 텍스트로부터 음성합성, 로봇팔의 적응적 제어
Adaptive Resonance Theory	G. Garpenter, S, Grossberg	1978~1985	레이더나 수중 전파탐지 등의 복잡한 패턴의식
Bidirectional Associative Memoy	B. kosko	1985	Content-addressable Associative Memory
Boltzman & Cauchey Machines	J. Hinton, H. Szu, T. Sejnowski	1985~1986	영상, 수중전파탐지, 레이더를 위한 패턴인식
Counter-Propagation	R. Hecht-Nielsen	1986	영상압축, 통계적인 분석, 은행의 대부 응용 프로그램

표2에서 분류한 Neural Network 모델들은 그 기능들과 작동 원리를 기준으로 분류되었다. 이와는 별개로 Neural Network 모델들은 입력 형식과 학습 방식에 따라서도 분류 될 수 있다. 입력 형식과 학습 방식에 따른 분류는 표3에 나타내었다.

표 3 입력 형식과 학습 방법에 따른 분류

입력형식	학습방식	Neural Network 모델
이진수	교사 학습	홉필드 네트워크
	교사 학습+비교사 학습	역전파 네트워크
	비교사 학습	ART모델
실수	교사 학습	퍼셉트론/다층 퍼셉트론
	비교사 학습	경쟁 학습/SOM

본 논문의 데이터는 입력 형식은 실수형, 학습 방식은 목표 값이 있는 교사

학습, 1개 이상의 Hidden Layer가 존재하는 Neural Network로 MLP(Multi-layer Perceptron)을 적용할 수 있다.

3.5 Perceptron

1957년 미국의 Frank Rosenblatt에 의해 발명된 Perceptron은 처음 소개되었을 때 상당한 센세이션을 불러일으켰다. Perceptron은 비교적 정확히 기술된, 계산에 의한 최초의 Neural Network 모델이었으며 여러 분야에 걸쳐 커다란 영향을 끼쳤다. Rosenblatt은 원래 심리학자였으며 Perceptron은 그러한 심리학적인 요구에 부응하는 것이었다. 또한 Perceptron이 잠재적으로 복잡한 적응행위를 할 수 있는 Learning Machine이라는 점은 엔지니어들에게는 매우 매력적인 것이었다.

그가 기술한 Perceptron 모델은 매우 복잡했다. Perceptron, 그리고 그것과 유사한 모델들은 분석하기가 매우 어려웠으나 Learning Machine의 능력과 제한점에 대한 통찰력을 제공해 주었다. Perceptron 이후의 연구 개발은 대부분 엔지니어와 물리학자들에 의해 진행되었다.

그러나 그의 중요한 논문은 읽기가 매우 어려웠다. 그는 Perceptron을 기술할 때 여러 개의 Version으로 기술했으며 각 Version마다 제 나름대로 이름을 붙였기 때문에 혼란을 가져 왔다. Perceptron에 대한 분석 또한 쉽지 않았다. 여러 가지 옵션과 변수와 학습 규칙들이 제대로 정리되지 않은 채 소개되어 이해에 상당한 혼란을 주었다.

Rosenblatt은 노이즈가 포함되어 있거나 완전하지 않은 연결이 있을 때의 Perceptron의 능력에 대해서도 기술하였는데, 메모리가 여러 곳에 분산되어 있어 손상에 대해 영향을 적게 받는다는 주장이었다.

몇 년이 지난 후 이 연구 논문은 논문으로서는 아이디어를 스케치한 것에 불과하다고 여겨졌다. Perceptron이 많은 분류를 학습할 수 있다는 그 유명한 Perceptron Convergence Theorem에 대한 증명이나 그것에 대한 인식조차 없었다. 단지 학습 가능성에 대한 몇 가지 통계적인 계산만이 포함되었고 나중에 밝

혀진 대로 학습 가능성에 대한 제한성을 간과하고 있었다.

최초의 Neural Network 장치인 마크 I Perceptron은 1957년에 제작되었으며 1958년에 성공적인 시범을 보였다.

3.6 MLP(Multi-Layer Perceptron)

3.6.1 MLP(Multi-Layer Perceptron)의 배경

Minsky와 Papert가 1969년 Perceptrons란 저서에서 Perceptron 모델을 수학적으로 철저히 분석하고 그 모델의 결정적인 단점들을 밝혀낸 이후 Neural Network에 관련된 연구는 약 20년간 침체의 길을 걷게 되었다.

Perceptron과 같이 하나의 Single-Adjustable Layer로 구성되는 모델들의 한계점들 때문에 입력층, 출력층 그리고 한 개이상의 Hidden Layer을 쓰는 새로운 모델들이 1980년대 중반에 제안되었으며, 특히 PDP(Parallel Distributed Procession) 그룹에 의한 폭 넓은 연구가 진행되었다. Rumelhart등은 1980년대 후반에 출판된 "Parallel Distributed Processing(PDP)"란 저서를 통해 Back Propagation Algorithm을 널리 유행시켰다. 이 그룹에서 제안한 모델은 Hidden Layer을 가지 MLP(Multi-Layer Perceptron)에 Back Propagation 학습 Algorithm을 사용함으로써 선형 분리 문제뿐만 아니라 여러 가지 문제점들을 해결할 수 있었으며 이로 인하여 십여년 간 침체했던 Neural Network 연구가 새롭게 활기를 띠게 되었다.

3.6.2 MLP(Multi-Layer Perceptron)의 정의

MLP(Multi-Layer Perceptron)은 입력층과 출력층 사이에 하나 이상의 중간층이 존재하는 신경회로망의 계층구조를 갖는다. 이 때의 입력층과 출력층 사이의 중간층을 Hidden Layer이라 부른다. Network는 입력층, Hidden Layer, 출력층 방향으로 연결되어 있으며, 각 층 내의 연결과 출력층에서 입력층으로서 직접

적인 연결은 존재하지 않는 전방향 Network이다.

다층퍼셉트론은 단층퍼셉트론과 유사한 구조를 가지고 있지만, 중간층과 각 유닛의 입출력 특성을 비선형으로 함으로써 Network의 능력을 향상시켜 단층퍼셉트론의 여러 가지 단점들을 극복했다. 다층퍼셉트론은 층의 개수가 증가할수록 퍼셉트론이 형성하는 결정 구역의 특성은 더욱 고급화된다. 즉 단층일 경우 패턴 공간을 두 구역으로 나누어주고, 2층인 경우 오목한 개구역 또는 오목한 폐구역을 형성하며, 3층인 경우에는 이론상 어떠한 형태의 구역도 형성할 수 있다.

일반적인 MLP의 학습방법은 다음과 같다. 입력층의 각 Unit에 입력 데이터를 제시하면 이 신호는 각 Unit에서 변환되어 중간층에 전달되고 최종적으로 출력층으로 나오게 된다. 이 출력 값과 원하는 출력 값을 비교하여 그 차이를 감소시키는 방향으로 Weight강도를 조정하는 것이다. 그러나 중간층이 있으면 학습은 어려워진다. 왜냐하면 어떤 Weight가 오차를 유발시키는지 알 수 없기 때문이다.

3.6.3 MLP(Multi-Layer Perceptron)의 구성요소

Input Layer : 각 입력변수에 대응되는 마디들로 구성되어있다. Nominal Var.에 대해서는 각 수준에 대응하는 입력마디를 가지게 되는데, 이는 통계적 선형모형에서 Dummy Var.를 사용하는 것과 같다.

Hidden Layer : 여러 개의 Neuron로 구성되어 있다. 각 은닉마다 Input Layer로부터 전달되는 변수 값들의 선형결합을 비선형함수로 처리하여 Output Layer 또는 다른 Hidden Layer에 전달한다.

Output Layer : Target Var.에 대응하는 마디들을 갖는다. 여러 개의 Target Var. 또는 세 개 이상의 수준을 가지는 Nominal Target Var.이 있을 때에는 여러 개의 출력마디들이 존재한다.

Combination Function : Combination Fun.는 Input Layer 또는 Hidden

Layer의 Neuron들을 결합하는 형태를 의미한다. 대부분의 Neural Network에서는 Combination Fun.로 선형함수를 사용하지만, 다른 형태의 Combination Fun.를 사용하는 Neural Network들도 있다. 예를 들어, RBF(Radial Basis Function) Neural Network는 원형기준함수를 사용한다.

Activation Function : Input Var. 또는 Hidden Neuron들의 결합을 변환하는 함수를 의미한다. 보통 S-자형의 비선형곡선의 형태를 가지게 되어 Squashing Function이라고 불리기도 한다. Activation Fun.는 통계적 선형모형에서, Link Fun.의 역함수와 유사한 의미를 가지며, 가장 보편적으로 사용하는 Activation Fun. 의 Logistic Fun.과 Hyperbolic Tangent이다. 한편, Target Var.이 제한된 범위를 가지지 않는 연속형 변수인 경우에는, Activation Fun.을 Identity Fun.을 사용한다.

3.6 Hidden Layer와 Neuron의 개수

Neural Network는 다양한 모형을 포함하는 매우 유연한 모형이다. 그러나 데이터로부터 계수를 추정해야 하기 때문에 실제로 MLP(Multi-Layer Perceptron)는 이론과 같이 유연하지 못한다. 또한 주어진 함수를 근사화하기 위해 매우 많은 수의 Neuron이 필요할 수도 있다. Hidden Layer와 Neuron의 개수가 많으면 많을수록 Neural Network는 더욱 복잡해지며, 추정해야할 계수의 수가 급격히 증가하기 때문에 최적화가 훨씬 어렵게 된다. 사실, 적절한 Hidden Layer와 Neuron의 수를 결정하기 위해서 시행착오(Trial-and-Error)적인 방법이외의 별 다른 대안이 없는 경우가 많다.

3.7 Neural Network의 해석의 어려움

Regression Analysis나 Decision Tree Analysis는 분석의 결과를 비교적 쉽게 해석하고 이로부터 유용한 정보를 얻을 수 있다. 반면에 Neural Network는 매

우 유연하기는 하지만 결과를 해석하는 것이 매우 어렵다. 즉, Logistic Regression이 제공하는 것과 같은 계수들에 대한 간편한 해석이 불가능하여, 어떤 Input Var.가 중요한지 또는 그것들이 어떻게 상호 작용하는지를 결정하기 어렵다.

Data Mining에서 해석의 용이함이 언제나 예측모형의 중요한 특성이 되는 것은 아니다. 더 많은 해석적 용이함을 가지고 있으면서도 예측에 덜 효과적인 모형보다는, 매우 정확한 예측을 생산해 내는 Neural Network가 더 선호되는 경우가 많이 때문이다. 그럼에도 불구하고 해석적 어려움은 실제적인 단점이 될 수 있다. 변수의 중요성과 그것들 간의 상호작용을 이해하는 것은 향후 데이터 수집 작업을 향상시킬 수도 있다.

제 4 절 Clustering Analysis

Clustering Analysis는 어떤 개체나 대상들을 밀접한 Similarity 또는 Distance에 의하여 유사한 특성을 지닌 개체들을 몇 개의 군집으로 집단화하는 다변량 기법이다. Discrimination Analysis에서는 Group 미리 정해져 있고, 분석 목적은 각 Group을 구분할 수 있는 함수를 추정하고, 또한 각 개체를 어느 한 Group에 분류하는 것이다. 반면에 Clustering Analysis에서는 Group의 수 혹은 Group의 구조에 대한 가정이 없으며, 오직 개체들 사이의 Similarity 또는 Distance에 의하여 Cluster를 형성하고, 형성된 Cluster의 특성을 파악하여 Cluster들 사이의 관계를 분석하는 기법이다. 따라서 Clustering Analysis는 어떤 개체나 대상들이 가지고 있는 다양한 특성에 의하여 동질성을 지닌 Cluster로 Group화하는 방법이다. Clustering Analysis는 분명한 분류기준이 없거나 알려져 있지 않은 상태에서 활용될 수 있는 기법이다.

제 5 절 Association Rules

Association Analysis는 Association Rules를 통해서 사건에 포함되어 있는 둘 이상의 변수들의 상호 관련성을 발견하는 것이다. 일반적으로 Association Analysis는 수학과 통계학의 확률과 기대치에 대한 개념을 기반으로 하고 있는데, 이러한 Association Rules를 해석하는데 있어 원인과 결과의 직접적인과 관계로 생각해서는 곤란하고 두 개 또는 그 이상 변수들 사이의 상호의 관련성으로 해석해야 한다.

전체 Pathway상에서 갑상선 암 발병에 중요한 Hormone을 판단하기 위해서 Hormone들 사이의 확률 Pr(Hormone A∩Hormone B)에 대해 살펴봐야 할 것이다. 이러한 확률을 지지도(Support)라고 하고 rule Hormone A -> Hormone B(Hormone A가 나타나면 Hormone B도 나타난다)의 지지도는 다음 수식1과 같이 정의 내려질 수 있다.

수식 1 지지도(Support)

$$\text{지지도}(Support) = \frac{Hormone\ A\text{와 } Hormone B\text{를 동시에 포함하는 } Network\text{ 수}}{\text{전체 } Network\text{ 수}}$$

이러한 지지도는 상호 대칭적으로 rule A->B의 지지도는 rule B->A 의 지지도와 같다. 지지도는 두 Hormone이 동시에 Network상에서 얼마나 자주 나타나는가를 측정하는 것이지만, 이보다 관심 있는 것은 Hormone A가 나타났을 경우 Hormone B가 나타날 가능성일 것이다. 이러한 개념에서 신뢰도(Confidence) 수식2는 Hormone A가 나타났을 때 Hormone B가 나탈 확률인 조건부 확률 Pr(B | A)로서 Pr(A∩B)/P(A)와 같고, 다음과 같이 계산되는데, 이 신뢰도는 상호 대칭적이지는 않다.(rule A->B의 신뢰도와 rule B->A의 신뢰도는 같지 않다.)

수식 2 신뢰도(Confidence)

$$\text{신뢰도}(Confidence) = \frac{Hormone\ A\text{와 } Hormone\ B\text{가 동시에 포함하는 } Network\text{ 수}}{Hormone\ A\text{를 포함하는 } Network\text{수}}$$

이렇게 정의된 지지도와 신뢰도는 확률의 개념으로써 0과 1사이의 값을 갖게 되며 1에 가까울수록 Hormone사이의 관계가 높다고 볼 수 있다.

제 2 장 연구방법

제 1 절 DATA 및 PATHWAY의 구성

1.1 DATA의 구성

갑상선암을 앓고 있는 환자 20명과 정상인 환자 29명을 대상으로 Androgen Hormone, Estrogen Hormone, Steroid Hormone들 중 26개의 Hormone을 측정하였다. 측정한 호르몬은 표4와 같으며 Network 구성의 용이함을 위해 임의의 Numbering을 하였다.

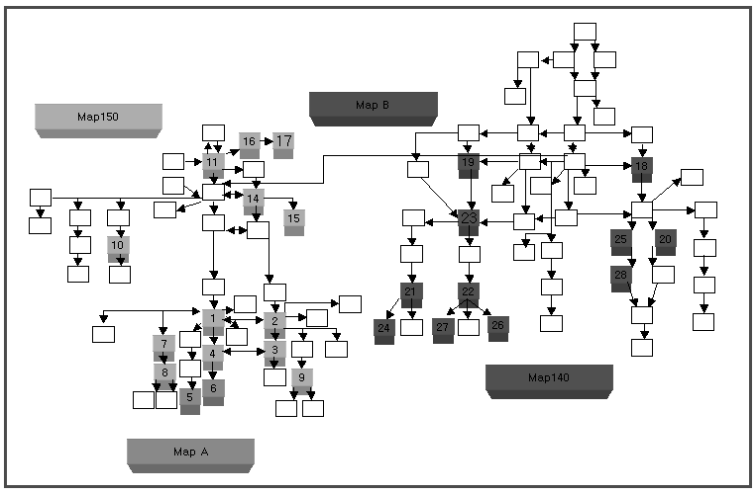
표 4 Hormone List

Hormone	Number	Hormone	Number	Hormone	Number	Hormone	Number
Estrone	1	2_Meo_E1	8	X_5AT	17	5a_THF	23
17b_Estradiol	2	2_3DiMeo_E2	9	THDOC	18	a_Cortolone	24
Estriol	3	Androsterone	10	THS	19	THB	25
16a_OH_E1	4	DHEA	11	THA	20	b_Cortol	26
16_Keto_E2	5	Testo	14	THE	21	a_Cortol	27
17_Epi_E3	6	DHT	15	THF	22	5a_THB	28
2_OH_E1	7	16OH_DHEA	16				

1.2 재구성한 PATHWAY

KEGG DATABASE C21-Steroid hormone metabolism(map140)과 Androgen and Estrogen Metabolism Pathway(Map140, Map150), ExPASy DATABASE의 Metabolic Pathway(Map A), 대사체 서적(Map B)을 바탕으로 그림3과 같이 구성하였다.

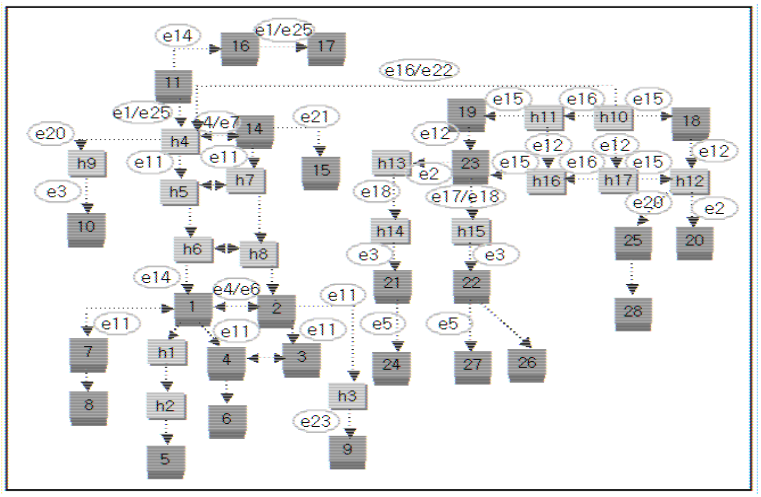
그림 3 재구성한 PATHWAY



1.3 축소한 PATHWAY

그림3의 Pathway를 바탕으로 실험을 통해 측정된 26개의 Hormone을 중심으로 재구성하였다.

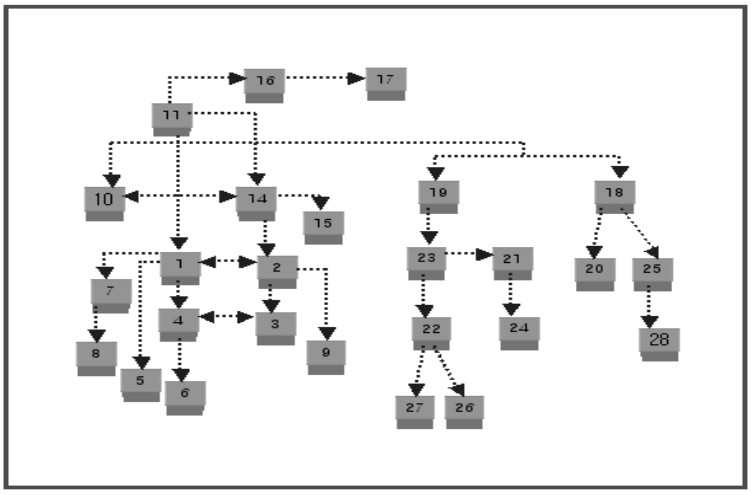
그림 4 축소한 Pathway



1.4 측정한 데이터로만 구성한 PATHWAY

측정되지 않은 hormone을 제외하고 측정된 데이터만으로 pathway를 구성하였다.

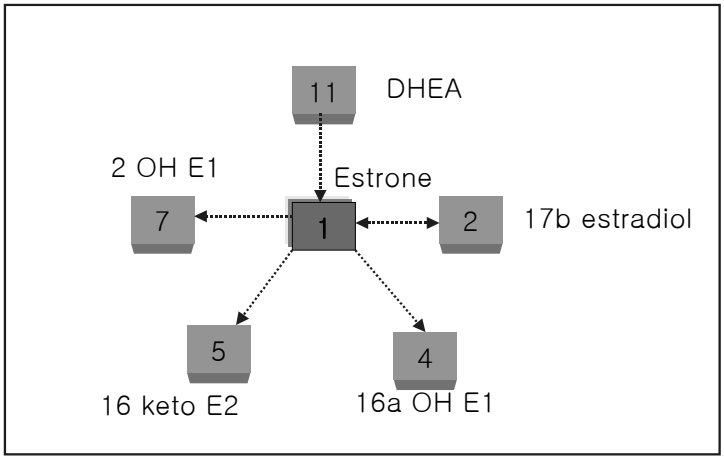
그림 5 측정한 데이터만으로 구성한 PATHWAY



제 2 절 Network의 구성

소규모 단위별 Network 상에서 Hidden Layer를 1개로 고정 후 Neuron의 수를 변화시키면서 SD값을 살펴보았다. 각 단위별 Neural Network의 구성은 26개 Hormone들 각각을 Target Var.로 하였으며 Pathway 상에서의 주변 Hormone을 Input으로 구성하였으며 SD(Standard Difference)결과 비교의 효율성과 타당성의 문제점을 해결하고자 Simple한 형태의 1 Hidden Layer의 MLP(Multi-Layer Perceptron)의 Network를 구성하였다. Hormone 1번의 Network 구성은 그림 6과 같다.

그림 6 Hormone 1번의 Network



또한 Activity Fn.은 다양한 연구를 통해 Hyper Tangent를 사용하였으며 Combination Fn.은 Linear Model을 사용하였다.

제 3 절 SD(Standard Difference)

각 Neuron의 수를 고려하여 예측 값과 실제 값 간의 차를 표준화된 지표인

SD(Standard Difference)로 정의를 하였으며 수식3과 같다.

수식 3 SD

$$SD = \sum_{i=1}^{N} \frac{| R_i - P_i |}{R_i} / N$$

(R_i : *Real Value* , P_i : *Predict Value* , N: *Number of Group*)

제 4 절 DMSW(Difference Mean Square of Weight)

SD값의 순위를 매겨 상위 순위에서 환자군과 정상군의 동일한 Neuron 수를 결정하였다. 결정된 Neuron에서 환자군과 정상군의 Weight 차이의 평균을 DMSW(Difference Mean Square of Weight)라고 정의하였으며 수식5와 같다. DMSW는 갑상선암 발병에 중요 호르몬 추정의 지표로 활용되어 진다.

수식 5 DMSW

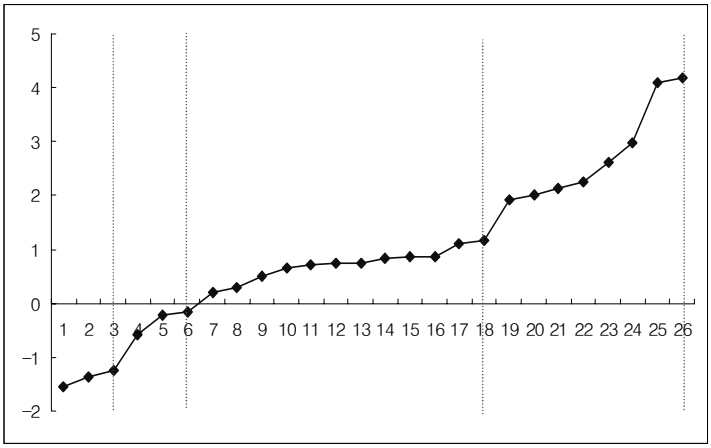
$$DMSW = \sum_{i=1}^{N} (T_i - N_i)^2 / N$$

(T_i : *Treatment Group* , N_i : *Normal Group* , N: *Number of Weight*)

제 5 절 K-means Clustering

DMSW을 Log Transformation하여 K-means Clustering Algorithm을 통해 질병 발병의 중요도에 따른 군집을 형성하였다. 군집의 수는 4개로 정하였으며 그림7에 따라서 4개 정도의 지점에서 Log Transformation값의 그래프 기울기의 변화가 있다고 추정되어졌기 때문이다.

그림 7 Log DMSW 값의 변화



제 6 절 Discovery Pattern

각 군집들의 Network 구성 Pattern을 살펴보고자 Association Analysis를 통해 Confidence Value를 구하였다. 의미 있는 Rule만을 추출하여 갑상선암 발병을 구성하는 Network를 추정하였다.

제 3 장 연구결과

제 1 절 DMSW의 Cluster

Hormone별 소규모 단위 네트워크의 응용된 Neural Network를 통해 DMSW를 구하여 표5와 같은 군집을 구하였다. 그림8에서와 같이 사전분석을 통해 4개의 군집을 추정하였으며 이를 K-means Clustering을 통해 각 구성 Hormone을 구한 결과 표5와 같이 나타났다. 이를 그림9에서 살펴본 결과 Pathway의 흐름과는 상이한 결과를 얻게 되었다. 즉 Pathway상에서 화살표 방향으로 진행 되어질 것이라고 예측된 갑상선 암 발병의 추정 경로에 대한 새로운 경로가 형성된 것이다.

그림 8 Cluster들의 분포

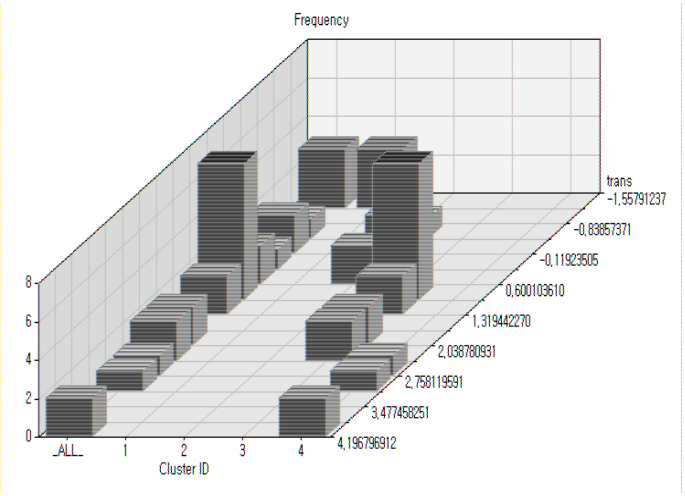
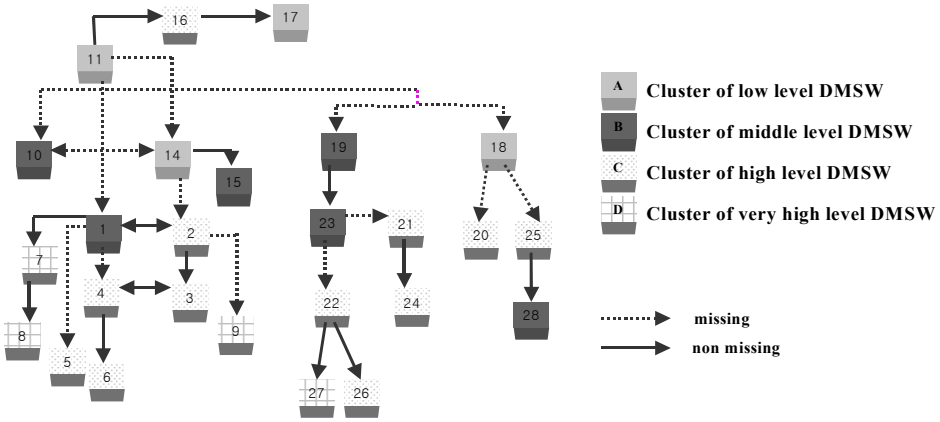


표 5 Cluster에 따른 Hormone의 구성

Cluster	node	DMSW	log_trans	Cluster	node	DMSW	log_trans
A	node18	0.027675	-1.55791237	C	node6	6.786775	0.831663451
	node14	0.043698	-1.35953844		node2	7.264366	0.861197746
	node11	0.054962	-1.25993747		node4	7.430303	0.871006524
B	node19	0.270815	-0.56732728		node21	12.54892	1.098606351
	node28	0.620542	-0.20722882		node5	14.35952	1.157139923
	node10	0.695346	-0.15779904		node3	83.23811	1.920322211
	node1	1.639499	0.214711087		node17	104.3033	2.018298049
	node23	1.938406	0.287444745		node24	132.1024	2.120910708
	node15	3.280999	0.516006098		node25	184.399	2.265758562
C	node16	4.673506	0.669642805	D	node9	420.6932	2.623965492
	node22	5.230416	0.718536232		node8	964.6566	2.98437274
	node26	5.588912	0.747327272		node7	12792.45	4.106953728
	node20	5.617905	0.749574391		node27	15732.47	4.196796912

그림 9 Pathway상에서 Cluster의 분포도



제 2 절 Cluster별 소규모 단위 Network의 관찰

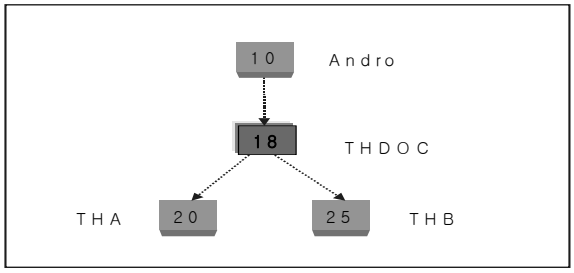
각 Network에서 Input Variable을 Elimination Algorithm을 통해 단계별 제거 후 DMSW 비교를 통해 각 Cluster의 Network 구성 요소를 추정하였다. 추정한 결과는 다음과 같다.

2.1 A Cluster의 Hormone별 소규모 단위 Network

2.1.1 18번 THDOC Hormone Network

THDOC를 Target Var.로 Androsterone, THA, THB를 Input Var. Network는 그림 10과 같다.

그림 10 THDOC Hormone Network



Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표6과 같다.

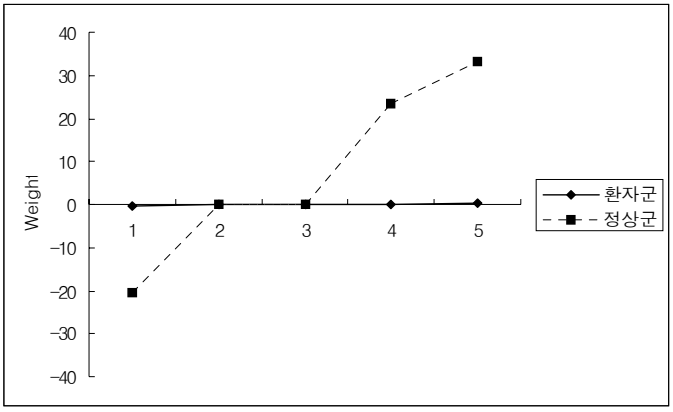
표 6 THDOC Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	21	0.027975	4	_20	5	0.089882	1
_10	5	0.679628	1	_25	5	417.3193	1

Hormone 18번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는

DMSW 값은 0.027675 이다. 앞의 결과에 근거하여 25번 Hormone인 THB 값이 다른 근접 Hormone 보다 많은 영향을 주는 것으로 추정되어진다. Neuron의 개수는 1개로 결정되었다. 영향을 주는 것으로 추정되는 THB Hormone은 Target을 기준으로 나가는 방향성을 지니고 있으며 Target과 THB Hormone 사이에는 하나의 Missing Hormone이 존재한다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포를 살펴본 결과 환자군과 정상군의 가중치의 분포는 서로 다른 형태로 차이가 있음을 분포 그래프 그림11을 통해 알 수 있다. Hormone 18번에 영향을 주는 THB Hormone과 관련되어 있는 효소는 E12(1.14.15.4), E20(1.3.99.5)이다.

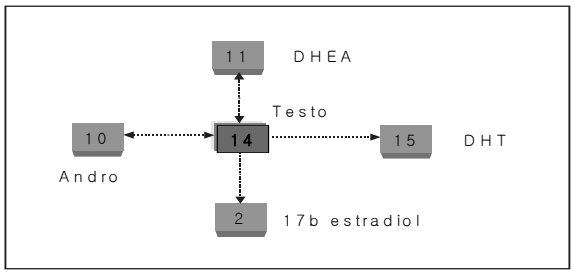
그림 11 THB 제거 후 Weight 그래프



2.1.2 14번 Testosterone Hormone Network

Testosterone을 Target Var.로 17b Estradiol, Androsterone, DHEA, DHT를 Input Var. Network는 그림 12와 같다.

그림 12 Testosterone Hormone Network



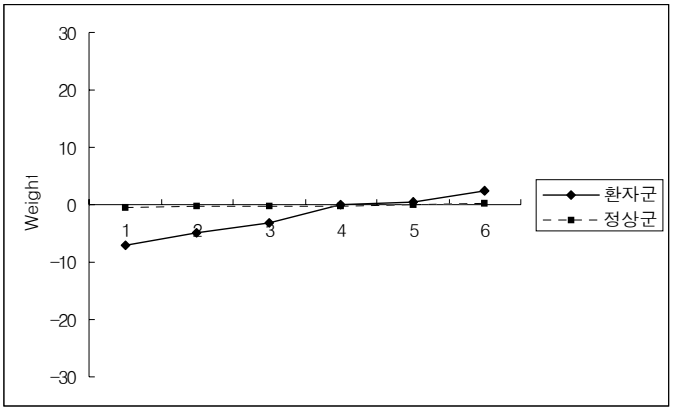
Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표7과 같다.

표 7 Testosterone Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	31	0.043698	5	_11	6	14.77007	1
_2	16	0.544924	3	_15	6	2.327752	1
_10	16	0.619634	3				

Hormone 14번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는 DMSW값은 0.043698 이다. 11번 Hormone인 DHEA값이 다른 nearest neighbor Hormone 보다 많은 영향을 주는 것으로 추정되어진다. Neuron의 개수는 1개로 결정되었다. 영향을 주는 것으로 추정되는 DHEA Hormone은 Target을 기준으로 양방향성을 지니고 있으며 Target과 DHEA Hormone 사이에는 하나의 Missing Hormone이 존재한다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포를 살펴본 결과 환자군과 정상군의 가중치의 분포는 서로 다른 형태로 차이가 있음을 분포 그래프 그림13을 통해 알 수 있다. Hormone 14번에 영향을 미친다고 생각되어지는 DHEA와 관련되어 있는 효소는 E1(1.1.1.145), E25(5.3.3.1), E4(1.1.1.51), E7(1.1.1.63)이다.

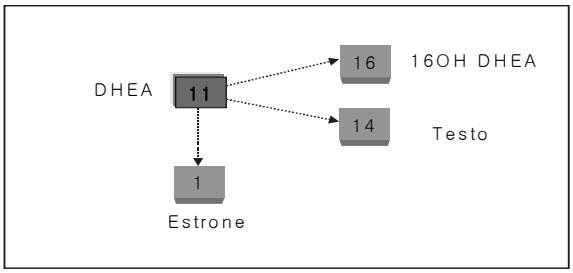
그림 13 DHEA 제거 후 Weight 그래프



2.1.3 11번 DHEA Hormone Network

DHEA를 Target Var.로 Estrone, Testosterone, 16OH DHEA를 Input Var. Network는 그림14와 같다.

그림 14 DHEA Hormone Network



Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표8과 같다.

표 8 DHEA Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	21	0.054962	4	_14	5	12.64616	1
_1	5	0.662594	1	_16	13	14.46373	3

Hormone 11번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는 DMSW값은 0.054962 이다. 14번 Hormone인 16OH DHEA값, 16번 Hormone인 Testosterone값이 다른 근접 Hormone 보다 많은 영향을 주는 것으로 추정되어진다. Neuron의 개수는 1개, 3개로 각각 결정되었다. 영향을 주는 것으로 추정되는 16OH DHEA Hormone, Testosterone Hormone은 각각 Target을 기준으로 나가는 방향성을 지니고 있으며 Target과 160H DHEA Hormone 사이에는 하나의 Missing Hormone이 존재하며 Target과 Testosterone Hormone 사이에는 Missing Hormone이 존재하지 않는다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포를 살펴본 결과 환자군과 정상군의 가중치 분포는 서로 다른 형태로 차이가 있음을 분포 그래프 그림15와 그림16을 통해 알 수 있다. Hormone 11번에 영향을 미친다고 생각되어지는 16OH DHEA, Testosterone과 관련되어 있는 효소는 E14(1.14.99.-) , E4/E7(1.1.1.51/1.1.1.63)이다.

그림 15 16OH DHEA 제거 후 Weight 그래프

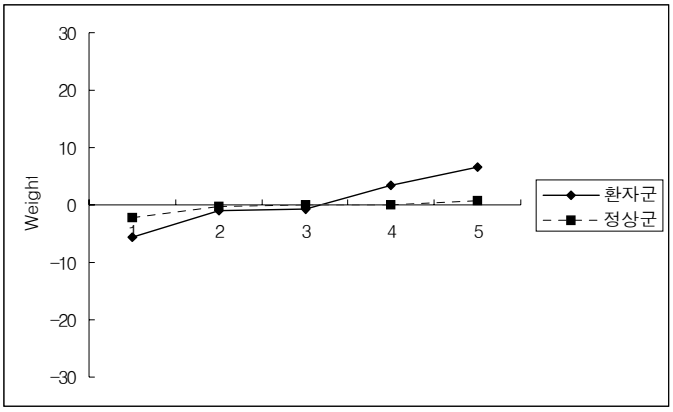
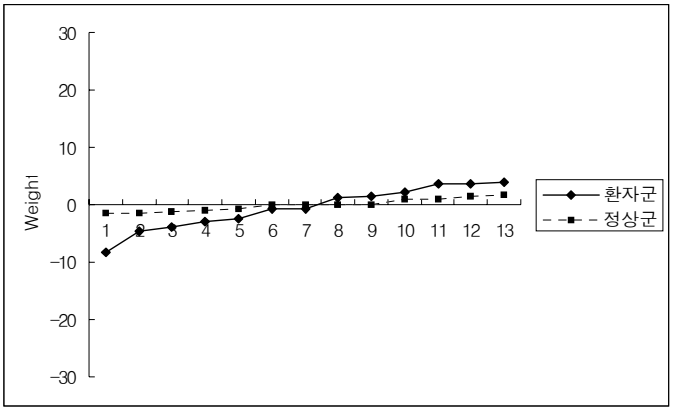


그림 16 Testosterone 제거 후 Weight 그래프

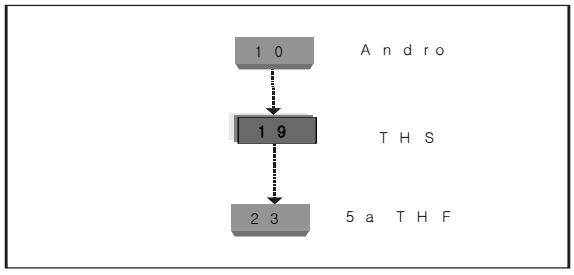


2.2 B Cluster의 Hormone별 소규모 단위 Network

2.2.1 19번 THS Hormone Network

THS를 Target Var.로 Androsterone, 5a THF를 Input Var. Network는 그림 17과 같다.

그림 17 THS Hormone Network



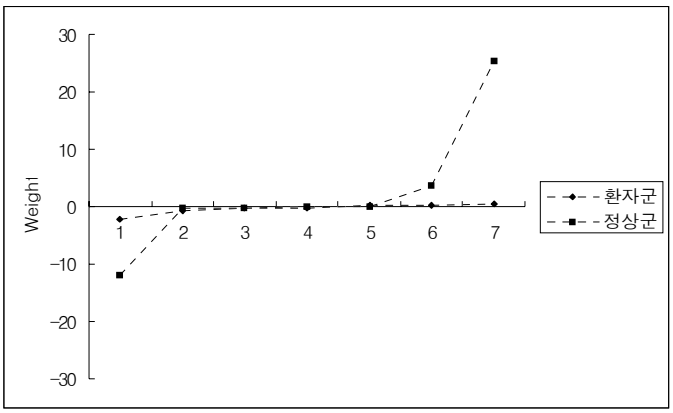
Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표9과 같다.

표 9 THS Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	5	0.270815	1	_23	4	0.091537	1

Hormone 19번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는 DMSW값은 0.270815 이다. 10번 Hormone인 Androsterone값이 다른 nearest neighbor Hormone 보다 많은 영향을 주는 것으로 추정되어진다. Neuron의 개수는 2개로 결정되었다. 영향을 주는 것으로 추정되는 Androsterone Hormone은 Target을 기준으로 양방향성을 지니고 있으며 Target과 Androsterone Hormone 사이에는 네 개의 Missing Hormone이 존재한다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포 그림18을 살펴본 결과 환자군과 정상군의 가중치 분포는 서로 다른 형태로 차이가 있음을 분포 그래프를 통해 알 수 있다. Hormone 19번에 영향을 미친다고 생각되어지는 Androsterone과 관련되어 있는 효소는 E3(1.1.1.50), E15(1.14.99.10), E16(1.14.99.9), E20(1.3.99.5), E16(1.14.99.9), E22(4.1.2.30)이다.

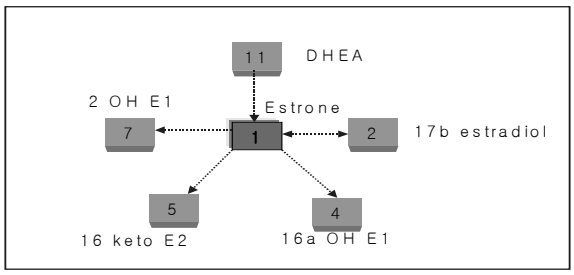
그림 18 Androsterone 제거 후 Weight 그래프



2.2.2 1번 Estrone Hormone Network

Estrone를 Target Var. 로 17b Estradiol, 16a OH E1, 16 Keto E2, 2 OH E1, DHEA를 Input Var. Network는 그림19와 같다.

그림 19 Estrone Hormone Network



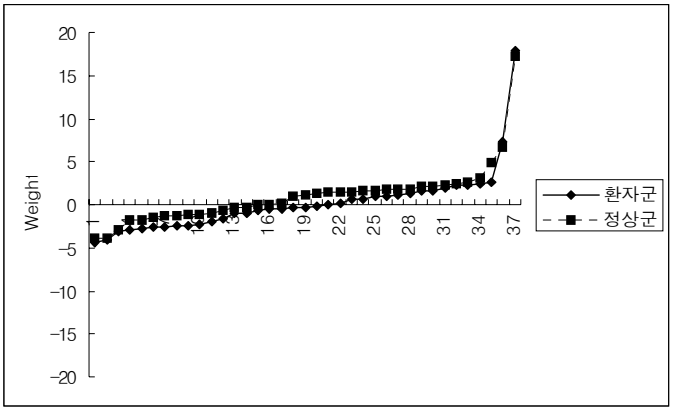
Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표10과 같다.

표 10 Estrone Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	64	1.639499	9	_5	49	3.224103	8
_2	49	2.535454	8	_7	37	2.563798	6
_4	49	3.009218	8	_11	37	8.649508	6

Hormone 1번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는 DMSW값은 1.639499 이다. 11번 Hormone인 DHEA값이 다른 nearest neighbor Hormone 보다 많은 영향을 주는 것으로 추정되어진다. 하지만, 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프 분포 그림18을 살펴본 결과 환자군과 정상군의 가중치 분포는 서로 비슷한 수치를 보이고 있다. 따라서 11번은 Hormone 영향에 의한 환자군과 정상군의 차이가 아닌 여러 개의 Hidden layer의 Neuron 수에 의한 영향으로 판단된다. 그러므로 이 Hormone는 비록 DMSW값의 차이는 있으나 DHEA 호르몬에 의한 영향이 아닌 것으로 판단된다.

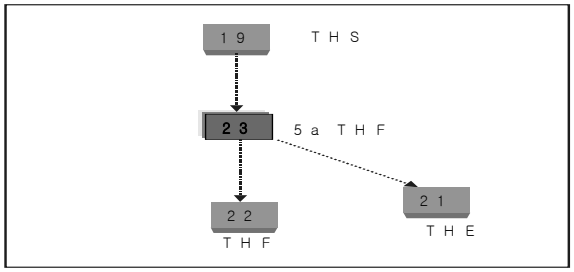
그림 20 DHEA 제거 후 Weight 그래프



2.2.3 23번 5a THF Hormone Network

5a THF를 Target Var.로 THS, THE, THF를 Input Var. Network는 그림 21과 같다.

그림 21 5a THF Hormone Network



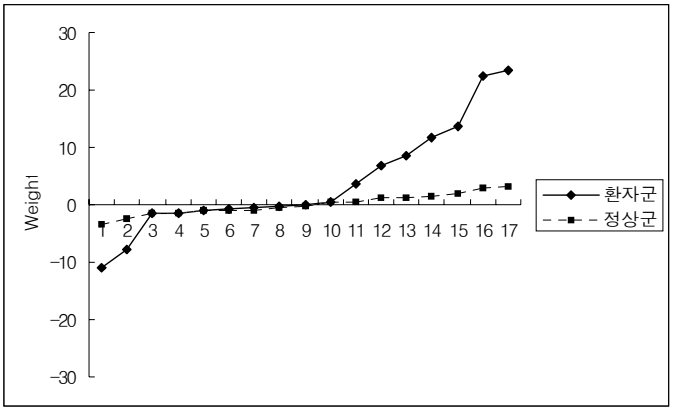
Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표11과 같다.

표 11 5a THF Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	26	1.938406	5	_21	17	80.87002	4
_19	17	1.281917	4	_22	9	5.400468	2

Hormone 23번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는 DMSW값은 1.938406 이다. 21번 Hormone인 THE값이 다른 nearest neighbor Hormone보다 많은 영향을 주는 것으로 추정되어진다. Neuron의 개수는 4개로 결정되었다. 영향을 주는 것으로 추정되는 THE Hormone은 Target을 기준으로 나가는 방향성을 지니고 있으며 Target과 THE Hormone 사이에는 두 개의 Missing Hormone이 존재한다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포 그림22을 살펴본 결과 환자군과 정상군의 Weight 분포는 서로 다른 형태로 차이가 있음을 분포 그래프를 통해 알 수 있다. Hormone 21번에 영향을 미친다고 생각되어지는 THE와 관련되어 있는 효소는 E2(1.1.1.146), E18(1.3.1.3)에 의한 영향으로 보인다.

그림 22 THE Hormone 제거 후 Weight 그래프

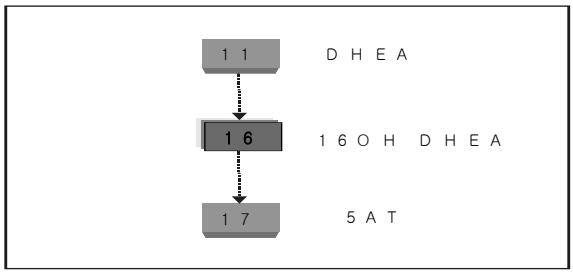


2.3 C Cluster의 Hormone별 소규모 단위 Network

2.3.1 16번 16OH DHEA Hormone Network

16OH DHEA를 Target Var.로 DHEA, 5AT를 Input Var. Network는 그림23과 같다.

그림 23 16OH DHEA Hormone Network



Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표12와 같다.

표 12 16OH DHEA Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	13	4.679506	3	_17a	4	4859.518	1
_11	7	53.54916	2	_17b	7	63960.14	2

Hormone 16번의 nearest neighbor 있는 Hormone의 전체를 Input으로 하는 DMSW값은 4.673506 이다. 11번 Hormone인 DHEA, 17번 Hormone인 5AT 많은 영향을 주는 것으로 추정되어진다. 그러나 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포 그림24, 그림25, 그림26을 살펴본 결과 환자군과 정상군의 Weight 분포에서 16_17A, 16_17B Weight 그래프의 분포가 동일 형태로 17번 Hormone 5AT에 의한 것이 아닌 이상치적인 값에 의한 것으로 규명되어진다. 반면 11번 Hormone을 제외했을 때 구해진 DMSW값은 17번 Hormone의 제외했을 때의 DMSW값은 작지만, Weight 분포 그래프 그림26을 살펴봤을 때 영향을 미친다고 판단되어 진다. Neuron의 개수는 2개로 결정되었다. DHEA Hormone은 Target을 기준으로 들어오는 방향성을 지니고 있으며 Target과 DHEA Hormone 사이에는 Missing Hormone이 존재하지 않는다. DHEA와 관련되어 있는 효소는 E14(1.14.99.-)이다.

그림 24 DHEA 제거 후 Weight그래프

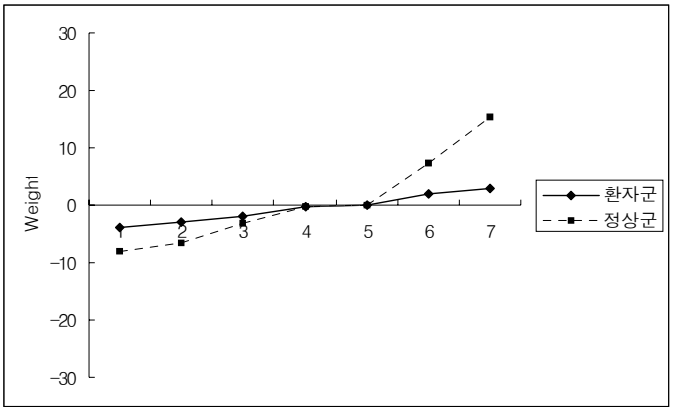


그림 25 5AT_a 제거 후 Weight그래프

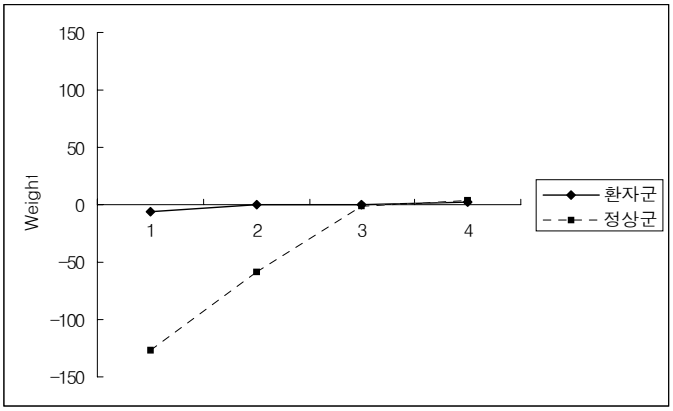
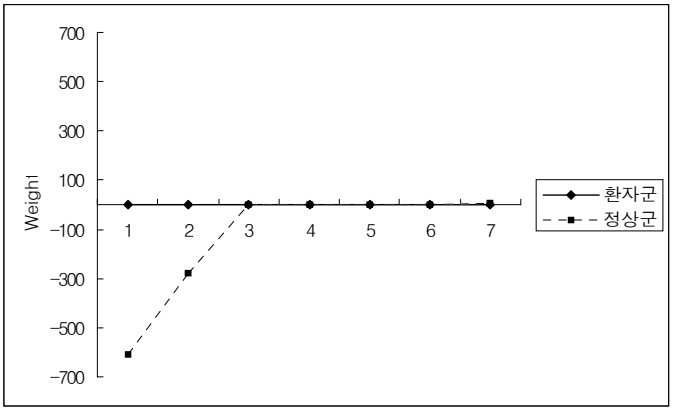


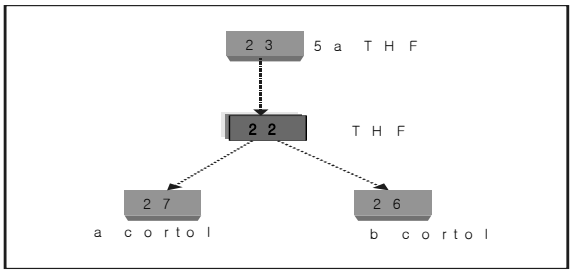
그림 26 5AT_b 제거 후 Weight그래프



2.3.2 22번 THF Hormone Network

THF를 Target Var.로 5a THF, a Cortol, b Cortol를 Input Var. Network는 그림27과 같다.

그림 27 THF Hormone Network



Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표13과 같다.

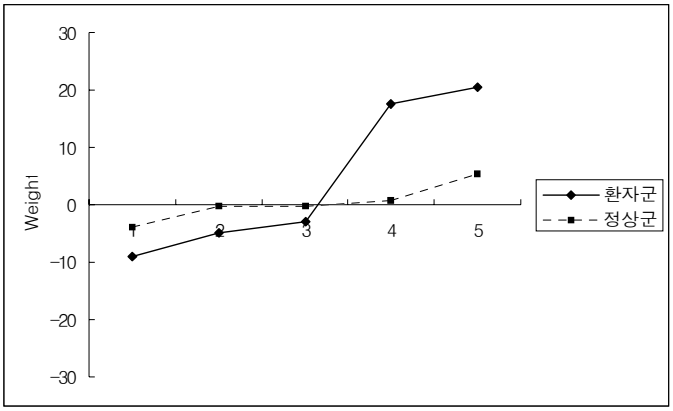
표 13 THF Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	11	5.230416	2	_26	13	6.386033	3
_23	13	11.72588	3	_27	5	119.4653	1

Hormone 22번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는

DMSW값은 5.230416 이다. 27번 Hormone인 a Cortol값이 다른 nearest neighbor Hormone 보다 많은 영향을 주는 것으로 추정되어진다. Neuron의 개수는 1개로 결정되었다. 영향을 주는 것으로 추정되는 a Cortol Hormone은 Target을 기준으로 나가는 방향성을 지니고 있으며 Target과 a Cortol Hormone 사이에는 Missing Hormone이 존재하지 않는다. 또한, 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포 그림28을 살펴본 결과 환자군과 정상군의 Weight 분포는 서로 다른 형태로 27번 Hormone인 a Cortol 이 갑상선 암 질병 발병에 영향을 주고 있는 것으로 판명되어 진다. Hormone 22번에 영향을 미친다고 생각되어지는 A CORTOL과 관련되어 있는 효소는 아직 실험에 의해 밝혀지지 않았다.

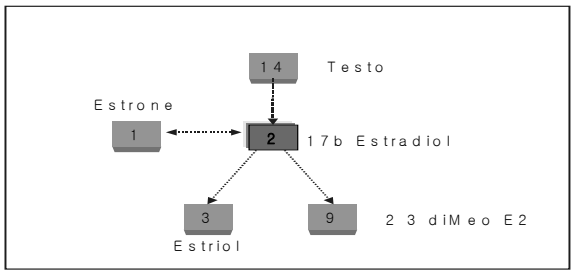
그림 28 a Cortol Hormone 제거 후 Weight 그래프



2.3.2 2번 17b Estradiol Hormone Network

17b Estradiol를 Target Var.로 Estrone, Estriol, 2 3 DiMeo E2, Testosterone를 Input Var. Network는 그림29와 같다.

그림 29 17b Estradiol Hormone Network



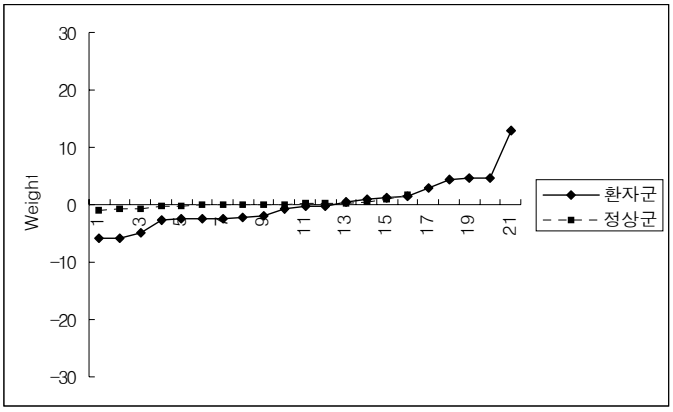
Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표14와 같다.

표 14 17b Estradiol Hormone Network 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	37	7.264366	6	_9	11	2.274559	2
_1	21	17.31041	4	_14	31	2.320778	6
_3	31	5.864667	6				

Hormone 2번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는 DMSW값은 7.264366 이다. 1번 Hormone인 Estrone값이 다른 nearest neighbor Hormone 보다 많은 영향을 주는 것으로 추정되어진다. Neuron의 개수는 4개로 결정되었다. 영향을 주는 것으로 추정되는 Estrone Hormone은 Target을 기준으로 양방향성을 지니고 있으며 Target과 Estrone Hormone 사이에는 Missing Hormone가 존재하지 않는다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포 그림30을 살펴본 결과 환자군과 정상군의 Weight 분포는 서로 다른 형태로 차이가 있음을 분포 그래프를 통해 알 수 있다. Hormone 1번에 영향을 미친다고 생각되어지는 Estrone과 관련되어 있는 효소인 E4(1.1.1.51), E6(1.1.1.62)이다.

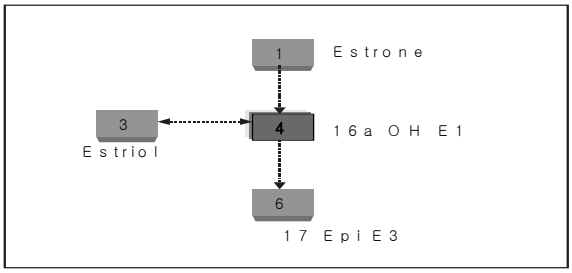
그림 30 Estrone Hormone 제거 후 Weight 그래프



2.3.3 4번 16a OH E1 Hormone Network

16a OH E1을 Target Var.로 Estrone, Estriol, 17 Epi E3을 Input Var. Network는 그림31과 같다.

그림 31 16a OH E1



Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표15와 같다.

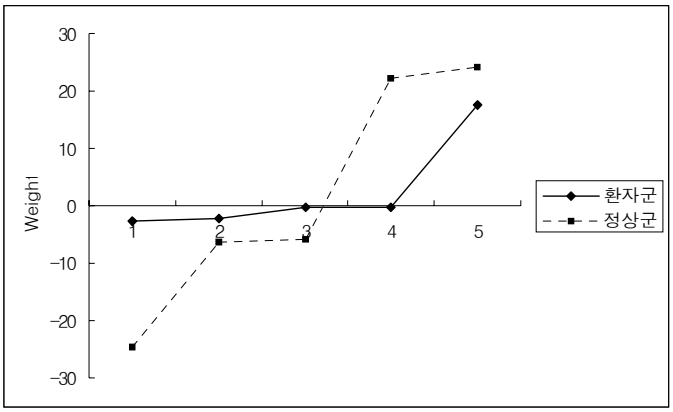
표 15 16a OH E1 Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	26	7.430303	5	_3	5	62.05483	1
_1	5	4.268426	1	_6	9	3.499198	2

Hormone 4번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는

DMSW값은 7.430303 이다. 3번 Hormone인 Estriol값이 다른 nearest neighbor Hormone 보다 많은 영향을 주는 것으로 추정되어진다. Neuron의 개수는 1개로 결정되었다. 영향을 주는 것으로 추정되는 Estriol Hormone은 Target을 기준으로 양방향성을 지니고 있으며 Target과 Estriol Hormone 사이에는 Missing Hormone가 존재하지 않는다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포 그림32에서 살펴본 결과 환자군과 정상군의 Weight 분포는 서로 다른 형태로 차이가 있음을 분포 그래프를 통해 알 수 있다. 즉 Hormone 4번의 3번 Hormone인 Estriol은 갑상선 암 질병 발병에 영향을 주는 Hormone임을 알 수 있으며 Hormone 4번에 영향을 미친다고 생각되어지는 Estriol과 관련되어 있는 효소는 아직 실험에 의해 밝혀지지 않았다.

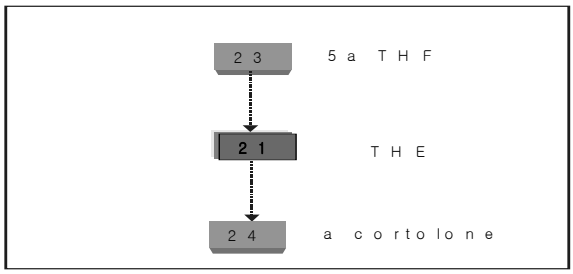
그림 32 Estriol Hormone 제거 후 Weight 그래프



2.3.4 21번 THE Hormone Network

THE를 Target Var.로 5a THF, a Cortolone를 Input Var. Network는 그림33과 같다.

그림 33 THE Hormone Network



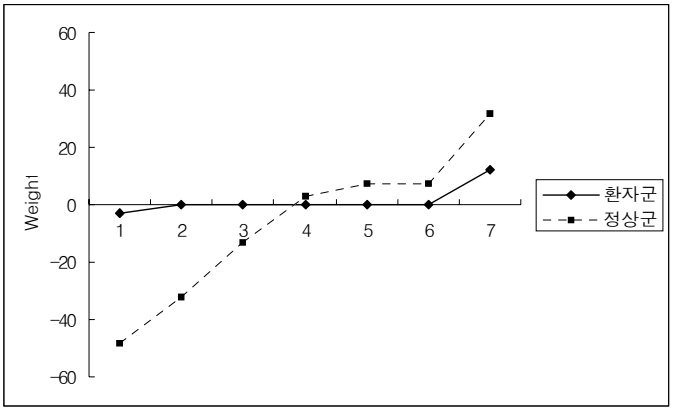
Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표16와 같다.

표 16 THE Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	17	12.54892	4	_23b	7	12.28219	2
_23a	4	9.067592	1	_24	7	634.3115	2

Hormone 21번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는 DMSW값은 12.54892이지만 24번 Hormone인 a Cortolone값이 다른 nearest neighbor Hormone 보다 많은 영향을 주는 것으로 추정되어진다. Neuron의 개수는 2개로 결정되었다. 영향을 주는 것으로 추정되는 a Cortolone Hormone은 Target을 기준으로 나가는 방향성을 지니고 있으며 Target과 a cortolone Hormone 사이에는 Missing Hormone가 존재하지 않는다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 가중치 그래프의 분포 그림34을 살펴본 결과 환자군과 정상군의 가중치 분포는 서로 다른 형태로 차이가 있음을 분포 그래프를 통해 알 수 있다. Hormone 21번에 영향을 미친다고 생각되어지는 a Cortolone와 관련되어 있는 효소는 E5(1.1.1.53)이다.

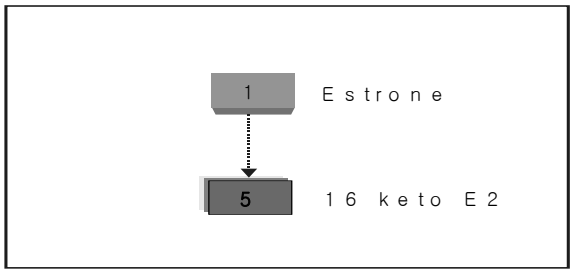
그림 34 a Cortolone Hormone 제거 후 Weight 그래프



2.3.5 5번 16 Keto E2 Hormone Network

16 Keto E2를 Target Var.로 Estrone를 Input Var. Network는 그림35와 같다.

그림 35 16 Keto E2 Hormone Network



Elimination Algorithm을 통해 DMSW 값은 표16과 같다.

표 17 16 Keto E2 Hormone Network의 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0a	4	14.35952	1	0b	7	59.51422	2

Hormone 5번의 nearest neighbor에 있는 Hormone의 하나를 Neuron 수에 따라 Hormone 전체를 Input으로 하는 DMSW값은 14.35952, 59.51422 이다. 따라서 1번 Hormone인 Estrone값이 많은 영향을 주는 것으로 추정되어진다. Neuron의 개수가 결정되지 않았다. 그러므로 Neuron 1개, 2개를 모두 살펴보았다. 영향을 주는 것으로 추정되는 Estrone Hormone은 Target을 기준으로 들어오는 방향성을 지니고 있으며 Target과 Estrone Hormone 사이에는 두 개의 Missing Hormone이 존재한다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포 그림36을 살펴본 결과 환자군과 정상군의 Weight 분포는 서로 다른 형태로 차이가 있음을 분포 그래프를 통해 알 수 있다. 즉 Hormone 5번의 1번 Hormone인 Estrone은 갑상선 암 질병 발병에 영향을 주는 Hormone임을 알 수 있다. Hormone 4번에 영향을 미친다고 생각되어지는 Estrone와 관련되어 있는 효소는 아직 실험에 의해 밝혀지지 않았다.

그림 36 Estrone_a Hormone Weight 그래프

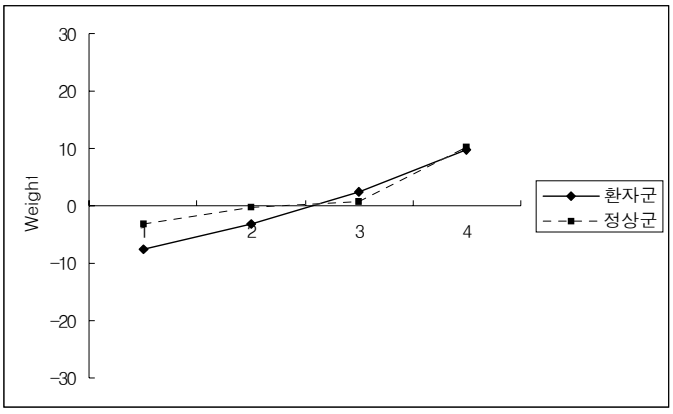
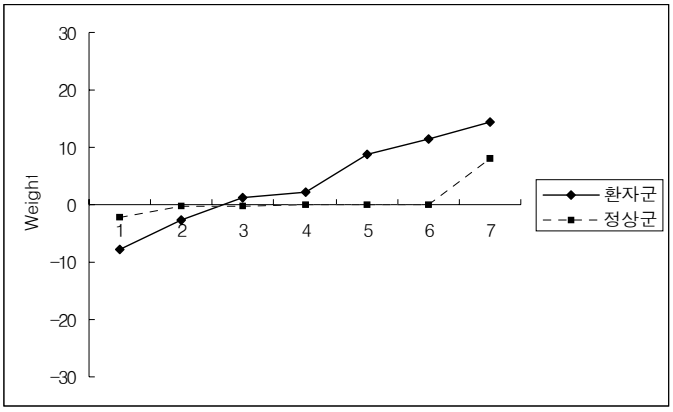


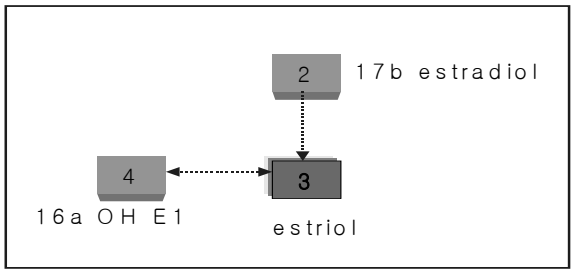
그림 37 Estrone_b Hormone 제거 후 Weight 그래프



2.3.6 3번 Estriol Hormone Network

Estriol를 Target Var.로 17b Estradiol, 16a OH E1을 Input Var. Network는 그림38과 같다.

그림 38 Estriol Hormone Network



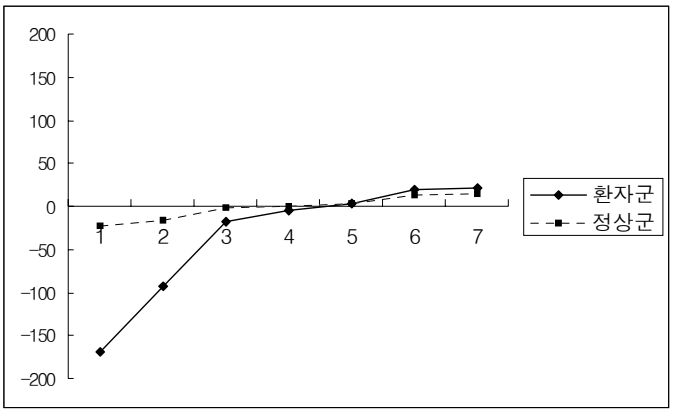
Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표18과 같다.

표 18 Estriol Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	9	83.23811	2	_2b	7	4.513822	2
_2a	4	0.927111	1	_4	4	8297.400	2

Hormone 3번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는 DMSW값은 83.23811 이다. 4번 Hormone인 16A OH E1값이 다른 nearest neighbor Hormone 보다 많은 영향을 주는 것으로 추정되어진다. Neuron의 개수는 1개로 결정되었다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포를 그림39에서 살펴본 결과 환자군과 정상군의 Weight 분포는 서로 동일 형태로 4번 Hormone인 16A OH E1이 갑상선 암 질병 발병에 영향을 주고 있지는 않으며 알려지지 않은 외부적인 환경에 의한 이상치적인 값이 존재하고 있는 것으로 판명되어 진다.

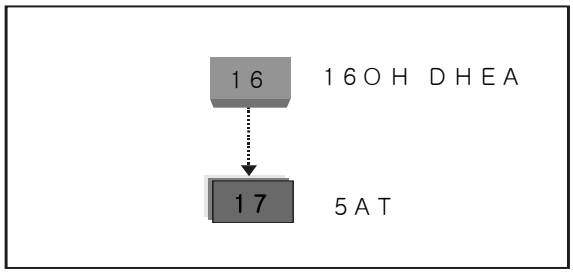
그림 39 16A OH E1 Hormone 제거 후 Weight 그래프



2.3.7 17번 5AT Hormone Network

5AT를 Target Var.로 16OH DHEA을 Input Var. Network는 그림38과 같다.

그림 40 5AT Hormone Network



Elimination Algorithm을 통해 DMSW 값은 표18과 같다.

표 19 5AT Hormone Network의 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0a	4	104.3033	1	0b	7	6.257544	2

Hormone 17번의 nearest neighbor에 있는 하나의 Hormone을 Neuron 수의 차이 1, 2를 Input으로 하는 DMSW값은 104.3033, 6.257544이다. Hormone인 16OH DHEA값이 Neuron 수가 1개의 경우 Neuron 수 2개 보다 많은 영향을 주는 것으로 추정되어진다. 하지만 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포 그림41,그림42를 살펴본 결과 환자군과 정상군의 Weight 분포는 서로 동일 형태로 16번 Hormone인 16OH DHEA는 갑상선암 질병 발병에 영향을 주고 있지는 않으며 외부적인 환경에 의한 이상치적인 값이 존재하고 있는 것으로 판명되어 진다.

그림 41 16OH DHEA_a Hormone Weight 그래프

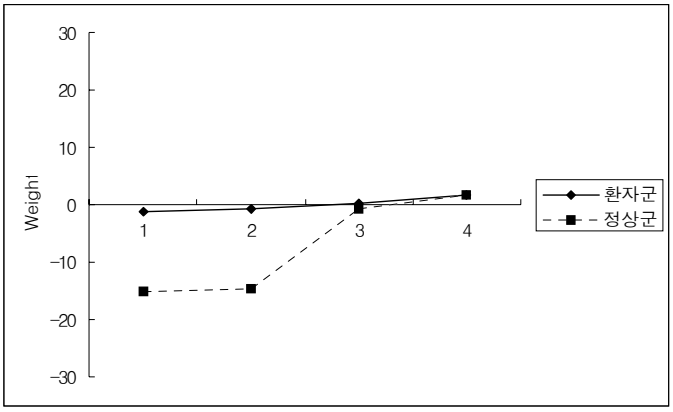
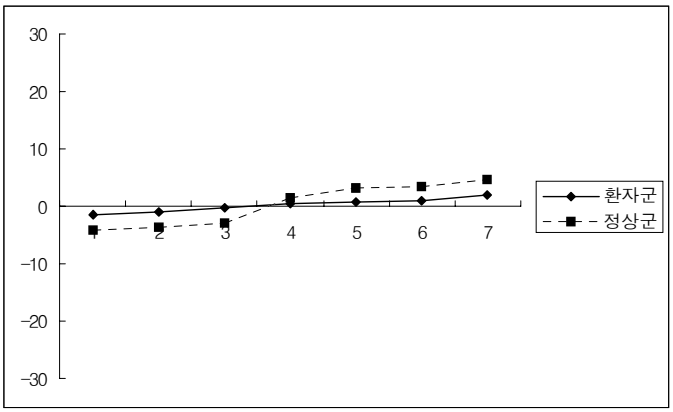


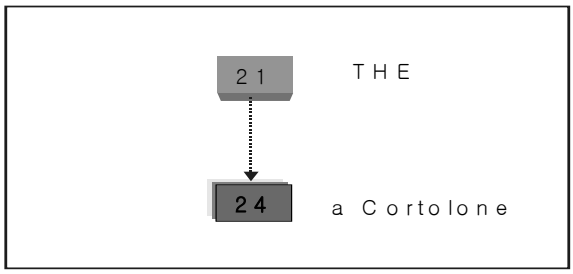
그림 42 16OH DHEA_a Hormone Weight 그래프



2.3.7 24번 a Cortolone Hormone Network

a Cortolone를 Target Var.로 THE을 Input Var. Network는 그림43과 같다.

그림 43 a Cortolone Hormone Network



Elimination Algorithm을 통해 DMSW 값은 표20와 같다.

표 20 a Cortolone Hormone Network의 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0a	4	132.1024	1	0b	7	2.489055	2

Hormone 24번의 nearest neighbor에 있는 하나의 Hormone을 Neuron 수의 차이 1, 2를 Input으로 하는 DMSW값은 132.1024, 2.489055이다. Hormone인 THE 값이 Neuron수 1의 경우 Neuron수 2보다 많은 영향을 주는 것으로 추정되어진다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포 그림44, 그림45를 살펴본 결과 환자군과 정상군의 Weight 분포는 Neuron수가 1개인 경우 Weight의 차이가 보인다. 21번 Hormone인 THE는 갑상선 암 질병 발병에 영향을 주는 것으로 규명되어 진다. Hormone 24번에 영향을 미친다고 생각되어지는 THE와 관련되어 있는 효소 E5(1.1.1.53)이다.

그림 44 THE_a Hormone Weight 그래프

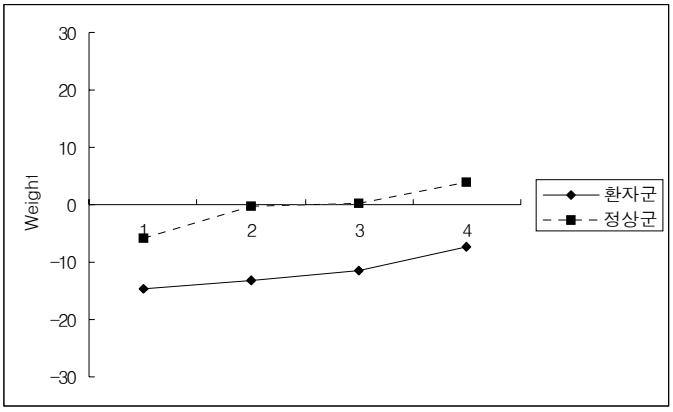
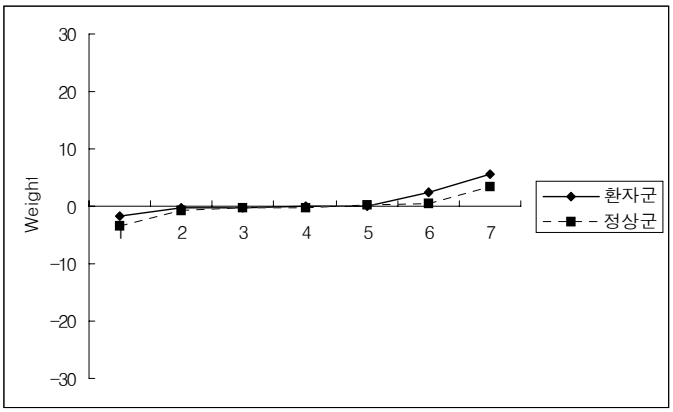


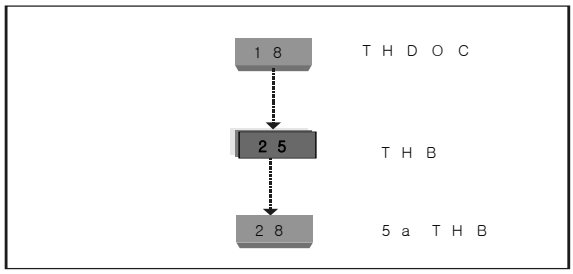
그림 45 THE_b Hormone Weight 그래프



2.3.8 25번 THB Hormone Network

THB를 Target Var.로 THDOC, 5a THB을 Input Var. Network는 그림46과 같다.

그림 46 THB Hormone Network



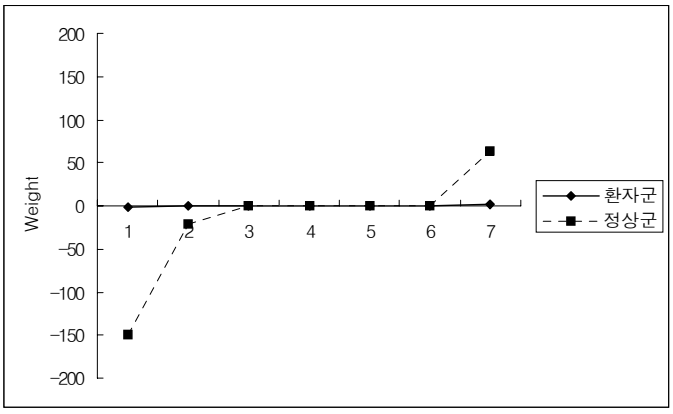
Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표21과 같다.

표 21 THB Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	13	184.3990	3	_28a	4	1.947614	1
_18a	4	5.127279	1	_28b	7	0.716455	2
_18b	7	3912.819	2				

Hormone 25번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는 DMSW값은 184.399 이다. 18번 Neuron 수 2인 Hormone THDOC값이 다른 nearest neighbor Hormone 보다 많은 영향을 주는 것으로 추정되어진다. 영향을 주는 것으로 추정되는 THDOC Hormone은 Target을 기준으로 들어오는 방향성을 지니고 있으며 Target과 THDOC Hormone 사이에는 하나의 Missing Hormone이 존재한다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포 그림47을 살펴본 결과 환자군과 정상군의 Weight 분포는 서로 다른 형태로 차이가 있음을 분포 그래프를 통해 알 수 있다. Hormone 25번에 영향을 미친다고 생각되어지는 THDOC와 관련되어 있는 효소는 E12(1.14.15.4), E20(1.3.99.5)에 의한 영향으로 보인다.

그림 47 THDOC Hormone 제거 후 Weight 그래프

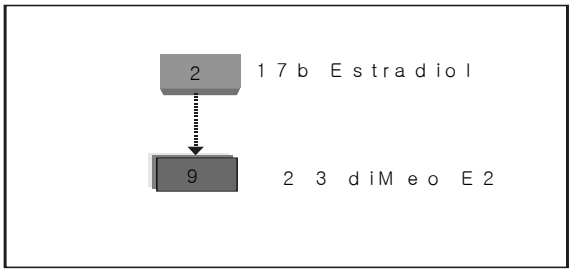


2.4 D Cluster의 Hormone별 소규모 단위 Network

2.4.1 9번 a Cortol Hormone Network

2 3 diMeo E2를 Target Var.로 17b Estradiol을 Input Var. Network는 그림48과 같다.

그림 48 a Cortol Hormone Network



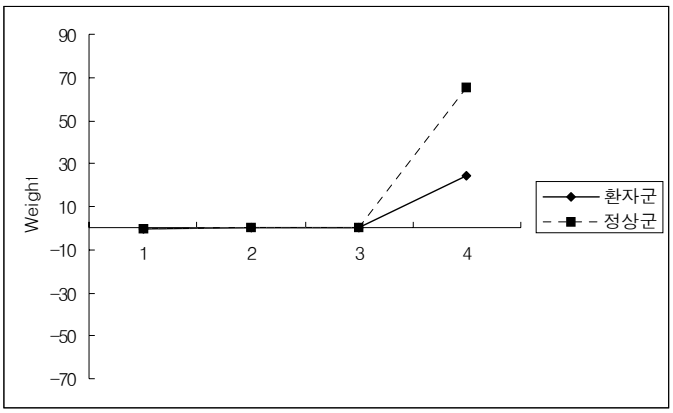
Elimination Algorithm을 통해 DMSW 값은 표22와 같다.

표 22 a Cortol Hormone Network의 DMSW값

TYPE	FREQ	DMSW	NEURON
0	4	420.6932	1

Hormone 9번의 nearest neighbor에 있는 하나의 Hormone 17b Estradiol를 Input으로 하는 DMSW값은 420.6932이다. Hormone 17b Estradiol은 영향을 주는 것으로 추정되어진다. 그러나, 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 Weight 그래프의 분포 그림을 살펴본 결과 환자군과 정상군의 Weight 분포는 서로 차이가 없다. 이것은 갑상선 암 질병 발병에 영향을 주고 있지는 않으며 외부적인 환경에 의한 이상치적인 값이 존재하고 있는 것으로 추정되어진다.

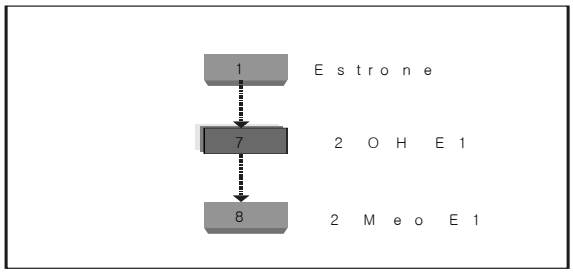
그림 49 17b Estradiol Hormone Wegight 그래프



2.4.2 7번 2 OH E1 Hormone Network

2 OH E1을 Target Var.로 Estrone, 2 Meo E1을 Input Var. Network는 그림50과 같다.

그림 50 2 OH E1 Hormone Network



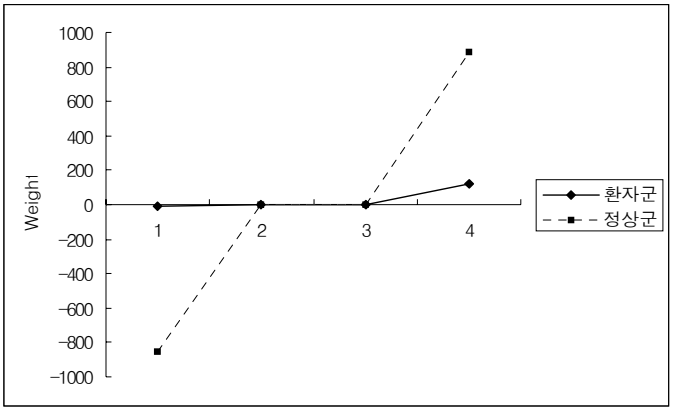
Elimination Algorithm을 통해 단계별 제거 후 DMSW 값은 표23와 같다.

표 23 2 OH E1 Hormone Network의 단계별 제거 후 DMSW값

TYPE	FREQ	DMSW	NEURON	TYPE	FREQ	DMSW	NEURON
0	13	12792.45	3	_8a	4	42413.75	1
_1a	4	54675.97	1	_8b	7	24220.12	2
_1b	7	23434.35	2				

Hormone 7번의 nearest neighbor에 있는 Hormone의 전체를 Input으로 하는 DMSW값은 12792.45 이다. 1번 뉴론 수가 1인 Hormone Estrone값이 다른 nearest neighbor Hormone 보다 많은 영향을 주는 것으로 추정되어진다. 영향을 주는 것으로 추정되는 Estrone Hormone은 Target을 기준으로 들어오는 방향성을 지니고 있으며 Target과 Estrone Hormone 사이에는 Missing Hormone이 존재하지 않는다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 가중치 그래프의 분포 그림51을 살펴본 결과 환자군과 정상군의 가중치 분포는 서로 다른 형태로 차이가 있음을 분포 그래프를 통해 알 수 있다. Hormone 7번에 영향을 준다고 생각되어지는 Estrone와 관련되어 있는 효소는 E11(1.14.13.-)이다.

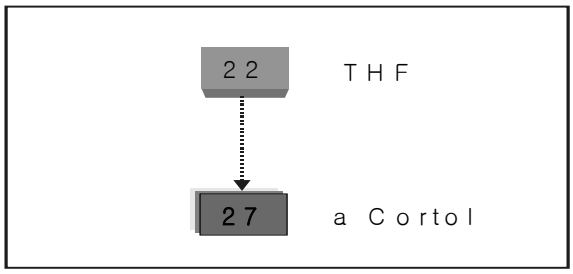
그림 51 Estrone Hormone 제거 후 Weight 그래프



2.4.3 27번 a Cortol Hormone Network

a Cortol을 Target Var.로 THF을 Input Var. Network는 그림52과 같다.

그림 52 a Cortol Hormone Network



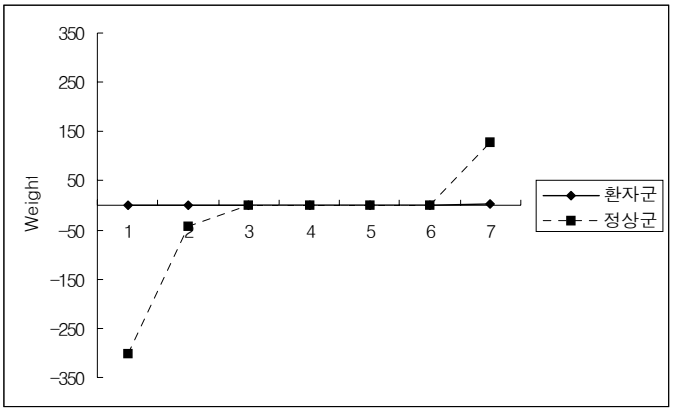
Elimination Algorithm을 통해 DMSW 값은 표24와 같다.

표 24 a Cortol Hormone Network DMSW값

TYPE	FREQ	DMSW	NEURON
0	7	15732.47	2

Hormone 27번의 nearest neighbor에 있는 하나 Hormone를 Input으로 하는 DMSW값은 15732.47 이지만 22번 Hormone THF값이 많은 영향을 주는 것으로 추정되어진다. Neuron의 수는 2개로 결정되었다. 영향을 주는 것으로 추정되는 THF Hormone은 Target을 기준으로 들어오는 방향성을 지니고 있으며 TARGET과 Estrone Hormone 사이에는 Missing Hormone이 존재하지 않는다. 또한 추정된 Hormone에 대해 환자군과 정상군의 Neural Network로 구한 가중치 그래프의 분포 그림을 살펴본 결과 환자군과 정상군의 가중치 분포는 서로 다른 형태로 차이가 있음을 분포 그래프를 통해 알 수 있다. Hormone 27번에 영향을 미친다고 생각되어지는 THF와 관련되어 있는 효소 E5(1.1.1.53)이다.

그림 53 THF Hormone Weight 그래프



제 3 절 갑상선 암 진행과정의 Network 구성의 규칙

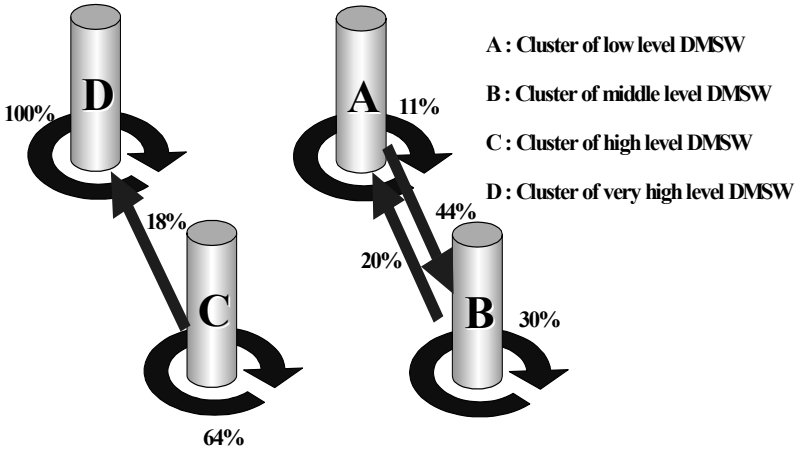
3.1 Pathway를 바탕으로 한 Association Rule

Association Analysis를 통해 재구성한 Pathway를 바탕으로 Cluster의 경향을 표25를 통해 그림54와 같은 결론을 얻게 되었다. 첫째, 갑상선 암 발병에 가장 높은 DMSW를 가지는 Cluster D는 자기 자신의 Network만을 형성한다. 두 번째로 Cluster C는 대부분 자기 자신의 Network를 형성하거나, Cluster D의 구성요소로 사용되어진다. 세 번째 Cluster B 자기 자신의 Network를 형성하거나 Cluster A의 구성요소로 사용되어진다. 네 번째 Cluster D는 자기 자신의 Network를 구성하기 보다는 Cluster D의 Network 구성요소로 사용되어지는 것을 볼 수 있었다.

표 25 Pathway를 바탕으로 한 Association Rule

CONF	RULE	CONF	RULE	CONF	RULE
100	D ==> D	30	B ==> B	18.18182	C ==> D
63.63636	C ==> C	20	B ==> A	11.11111	A ==> A
44.44444	A ==> B				

그림 54 각 Cluster에 대한 relation(Pathway)



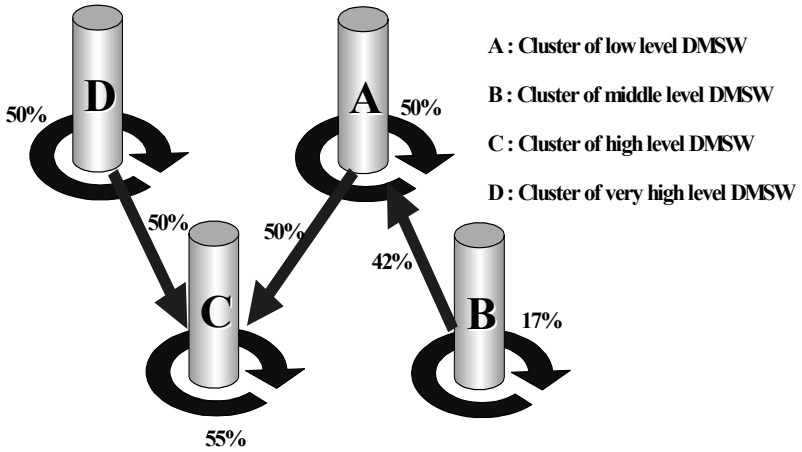
3.2 DMSW를 바탕으로 한 Association Rule

추정된 Network의 구성요소를 Association Analysis를 표26을 통해 그림 55과 같은 결론을 얻게 되었다. 첫째, 갑상선 암 발병에 가장 높은 DMSW를 가지는 Cluster D는 자기 자신의 Network를 형성하거나 또는 높은 DMSW를 가지는 Cluster C를 구성하는 구성원으로 사용되어졌다. 두 번째로 Cluster C는 대부분 자기 자신의 Network를 형성한다고 보여 진다. 반면 Cluster B와 같은 경우는 자기 자신의 Network를 형성하기 보다는 Cluster D의 구성요소로 사용되어지는 것을 볼 수 있었다. 마지막으로 Cluster A는 높은 DMSW를 가지는 Cluster C의 구성요소로 사용되어지거나, 자기 자신의 Network를 형성한다고 보여 진다.

표 26 DMSW 바탕으로 한 Association Rule

CONF	RULE	CONF	RULE	CONF	RULE
54.54545	C ==> CC	50	A ==> AA	41.66667	B ==> AA
50	D ==> DD	50	D ==> CC	16.66667	B ==> BB
50	A ==> CC				

그림 55 각 Cluster에 대한 relation(DMSW)

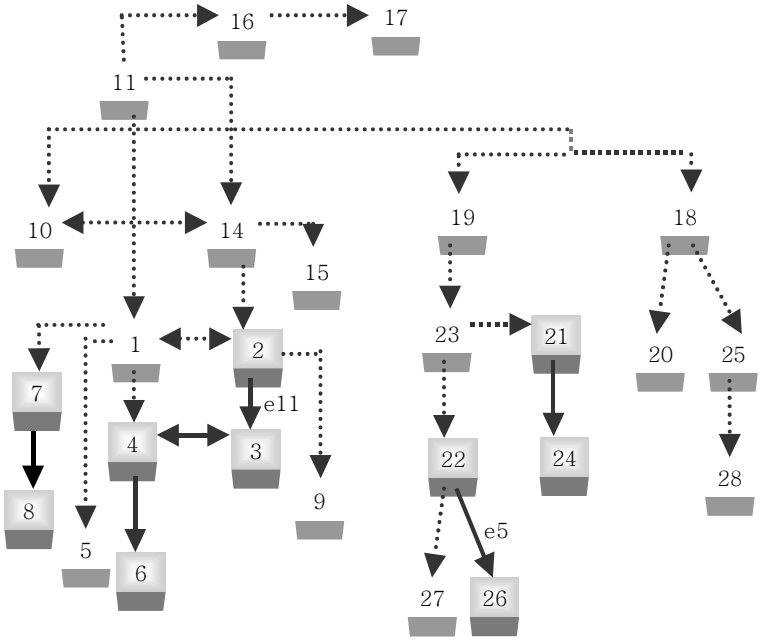


제 4 장 결과 및 토의

갑상선 암 발병의 중요 호르몬을 추정하기 위해서 49명의 피실험자를 대상으로 Steroid Pathway 관련 일부 Hormone의 농도를 측정하였다. 측정된 26개의 Hormone을 대상으로 응용된 Neural Network에 분석변수로 사용하여 소단위 Network를 구성하였다. 구성된 Network의 Weight값을 정상군과 환자군의 차를 나타내는 DMSW를 추정하여 Cluster Algorithm을 통해 4개의 Cluster로 나누었다. 또한, 소단위 Network마다 Elimination Algorithm을 통해 단계별 제거 후 DMSW값과 DMSW의 분포 그래프를 통해 중요 Hormone을 판별해 낼 수 있었다. 판별한 여러 개의 Hormone들을 Association Analysis를 통해, Cluster의 relation을 살펴봤을 때, 다음 그림56과 같다. D Cluster와 C Cluster의 자기 자신으로 가는 Network가 갑상선 암 발병에 중요 Hormone이며 중요 단계로 추정되어진다. 17b Estradiol, Estradiol, 16a OH E1, 17 Epi E3, 2 OH E1, 2 Meo E1, THE, THF, a Cortolone, b Cortol Hormone이 중요 Hormone으로 추정되어지고, 산화환원반응을 촉매 하는 1군 효소인 1.1.1.53, 1.14.13.- Enzyme이 중요 단계로 판단되어진다. 앞으로 중요 Hormone으로 추정된 각 단계의 Enzyme이 실험을 통해 밝혀져야 할 것이며, 갑상선 암 진단에 위에 언급한 Hormone의 탐색이 필요할 것으로 판단되어진다.

본 논문에서는 Regression Analysis나 Decision Tree Analysis에 비해 해석의 어려움을 지닌 응용된 Neural Network Algorithm을 사용하였다. 그러나, Hormone의 중요도와 상호작용을 판단하기 위해, Association Rule Algorithm을 사용하여 응용된 Neural Network의 단점을 극복하였다. 또한, Neural Network의 계수들에 대한 해석을 위해 DMSW라는 값을 추정하여 해석의 용이함을 더 하였다.

그림 56 갑상선 암 발병 관련된 중요 Hormone



참 고 문 헌

Ralph I. Dorfman and Frank Ungar, Metabolism of steroid hormones, New York, Academic Press, 1965

김대수, 신경망 이론과 응용1, 하이테크정보, 1992

김대수, 신경망 이론과 응용2, 하이테크정보, 1993

Young Sun Kim, So Young Sohn, Dong Kee Kim, Dog gen Kim, Yong Han Paik, Ho Shik Shim, Screening test data analysis for liver disease prediction model using growth curve, Biomedicine and Pharmacotherapy, 2003

http://www.genome.jp/kegg/pathway.html

http://www.expasy.org/cgi-bin/show_thumbnails.pl

http://www.chem.qmul.ac.uk/iubmb/enzyme/

http://www.webpages.uidaho.edu/~stevel/nn.html

http://user.chollian.net/~jazzy67/Thyroid.htm

http://blog.naver.com/whitecotton.do?Redirect=Log&logNo=2547272

Philip D. Wasserman, Neural Computing Theory and Practice, ANZA Reaserch, Inc.

성웅현, 응용 다변량 분석, 탐진출판사, 1997

최종후, 한상태, 강현철, 김은석, 김미경 데이터마이닝 기능과 사용법, 2000

강현철, 한상태, 최종후, 김차용, 김은석, 김미경 데이터마이닝 방법론 및 활용, 자유아카데미, 1999

Hiroyuki Ogata, Susumu Goro, Kazushige Sato, Wataru Fujibuchi, Hidemasa Bono and Minoru Kanehisa, KEGG : Kyoto Encyclopedia of Genes and Genome, 1999

배은진, 신진 대사 경로를 위한 XML 스키마 설계 및 경로표현, 2002

ABSTRACT

Metabolic Pathway Analysis for Thyroid Cancer using Neural Network

Kil, Seong Hwa
Dept. of Biostatistics and Computing
The Graduate School
Yonsei University

To estimate the important hormones causing thyroid cancer, the concentration of some hormones related to Steroid pathway toward 49 test subjects. These data were re-composed in compliance with Network Analysis Format through KEGG(Kyoto Encyclopedia of Genes and Genomes) Metabolic Pathway and study of literature.

To the measured 26 hormones, Network by unit was composed using analysis variables to applied Neural Network. At the composed Network, Weight Matrix were calculated to DMSW(Difference Mean Square of Weight) used as an indicators showing the differences of normal group and treatment group.

4 clusters were estimated using Clustering(K-means Clustering) Algorithm to the calculated DWSW values. As a result of estimating network relations of each cluster through Association Algorithm, some rules of the estimated 4 clusters could be found in the Steroid Hormone Pathway related to thyroid cancer.

Key Words : Thyroid Cancer, Metabolic Pathway, Hormone, Neural Network, Weight, DMSW, Clustering Association Algorithm