한국생물공학회지 제21권 제4호
Korean J. Biotechnol. Bioeng.
Vol. 21, No. 4, 231-235(2006)

확률적 그래프 모델을 이용한 세포 간 정보 네트워크 추론

[1]나 상 동 · [2]신 현 재 · [2†] 차 월 석
[1]조선대학교 전자정보공과대학 컴퓨터공학부, [2]조선대학교 공과대학 생명화학공학과
(접수 : 2006. 6. 9., 게재승인 : 2006. 8. 13.)

Informatics Network Representation Between Cells Using Probabilistic Graphical Models

Sang Dong Ra[1], Hyun-Jae Shin[2], and Wol-Suk Cha[2†]
[1]Department of Computer Engineering, College of Electronics and Information Engineering,
[2]Department of Chemical & Biochemical Engineering, Chosun University, Gwangju 501-759, Korea
(Received : 2006. 6. 9., Accepted : 2006. 8. 13.)

This study is a numerical representative modeling analysis for the application of the process that unravels networks between cells in genetics to web of informatics. Using the probabilistic graphical model, the insight from the data describing biological networks is used for making a probabilistic function. Rather than a complex network of cells, we reconstruct a simple lower-stage model and show a genetic representation level from the genetic based network logic. We made probabilistic graphical models from genetic data and extends them to genetic representation data in the method of network modeling in informatics

Key Words : Network search, genetics informatics, network motifs

서 론

IT와 BT가 접목되면서 분자생물학 연구에 많은 변화와 발전이 일어나고 있다. 생물학적 유전자 배열은 유전 스케일에서 세포를 입증할 수 있는 고 처리 분석의 개발로 가능하며, 이 분석법은 다양한 레벨에서 분자 네트워크와 그 요소들을 검출하여 DNA로부터 전령에 전사(轉寫)된 유전정보와 단백질-단백질과 단백질-DNA 상호관계, 염색체 구조, 단백질 양, 위치측정, 환경의 영향에 의한 비유전성의 일시적 변이를 포함한다(1). 이런 데이터를 통해 세포 처리 과정, 생물학적 네트워크 등 다양한 각도에서 정보공학 네트워크에 응용 한다.

세포 시스템을 수치해석적 분석법으로 얻은 데이터의 통합은 유전자 발현과 단백질 상호작용보다 통일성 있는 재구성을 가능하게 하여 잡음 효과를 감소시킨다. 그러나 이러한 통합을 얻기 위해서는 검측법과 짝을 이루는 생화학적 원리를 이해해야만 한다.

† Corresponding Author : Department of Chemical & Biochemical Engineering, Chosun University, Gwangju 501-759, Korea
Tel : +82-62-230-7218, Fax : +82-62-230-7226
E-mail: wscha@chosun.ac.kr

생물학적 시스템 모델은 추상적인 것을 단순화한 형태로 다양한 조건에서 시스템의 행동과정을 발생시켜 이 행동들에서 시스템 요소의 역할을 추론하고, 확률적 모델 검출 잡음과 생물학적 시스템의 측면을 설명하기 위한 확률함수를 사용한다. 데이터를 분석하는 모델접근법에서 모델의 공간을 정의하고 모델링하기 위해 논리적 추론을 통한 예측으로 데이터를 발견하는 학습절차법을 이용한다.

본 논문은 세포 간 네트워크의 모델분석법에 의해 확률적 그래프 모델로 알려진 수학적 그래프 모델에서(2, 3) 상호작용을 통계분야로 개발하고 확률적 결과와 모델을 연속적으로 관측하여 확률적 그래프 모델을 추론하는 수치해석적 절차에서 정보공학적 네트워크를 연구한다.

확률적 그래픽 모델

생물학적 시스템을 모델링할 때 유전자 시스템과 관련된 엔티티 (entity)와 표현 레벨은 확률적 모델의 임의의 변수로 취급한다(4). 임의의 변수에는 유전자 배열처럼 모델에 감추어진 특성과 실험에서 보여지는 유전자의 표현 레벨 등이 포함된다. 확률적 그래프 모델은 한계점의 산출물에 따라 다중 변량의 결합 확률 배분을 나타내어 소수의 변수에만 연관된 산출물의 구조를 그래프로 표현한다. 여

기서 그래프는 일반 한계점을 나타내는 변수와 관련되어 있으므로 이 그래프는 분배의 생산형태를 명시하고 산출물에 의해 나타나는 특성을 추론할 수 있는 도구가 되며 (5) 산재된 그래프에서의 표현도 효율적인 추론과 학습에 처리한다.

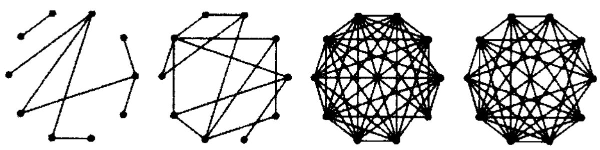

Figure 1. Network search graph.

Fig. 1에서 확률적 그래프는 객체의 확률적인 면과 구조적인 면을 모두 함께 다룰 수 있기 때문에 복잡한 객체가 되므로 다중 계층에서 확장되면서 네트워크 확률 탐색입증으로 객체 표현이 가능하다. 하부 계층에는 객체의 기본이 되는 정보를 표현하고 상위 계층에는 하부 계층의 조합으로 얻어질 수 있는 네트워크에서 확률적 정보를 추론함으로써 상위 계층으로 올라 갈수록 학습된 모델을 이용하여 탐색입증이 가능하다. 여기서 확률적 탐색입증 그래프의 특성과 정보 표현 탐색 인식 등이 확률적 방법에 의해 이루어지므로 생물학적 시스템을 모델링할 때 확률적 그래프에서 출력되는 결과를 네트워크 노드로 유도된 에지의 집합으로 이루어지므로 확률적 네트워크 그래프는 다음과 같이 정의한다(6).

정의 1 (network graph)

An network graph over $A_V \text{Y} A_E$ is a 4-tuple $G=(V(G),E(G),\phi,\varepsilon)$ such that

- $V(G)$ is a finite, non-empty set of vertices;
- $E(G) \subset V(G) \times V(G)$ is a set of ordered pairs of distinct elements in $V(G)$, called edges;
- A_V is a finite, non-empty set of vertex labels (primitive descriptions);
- A_E is a set of edge labels (relation descriptions);
- $\phi : V(G) \to A_V$ is a function, called a vertex interpreter;
- $\varepsilon : E(G) \to A_E$ is a function, called an edge interpreter.

위 정의에서 사용하는 네트워크 그래프는, 인식 대상 네트워크 자체의 특성을 표현하기 위해서 사용되는 확률적 그래프 조건 변수를 특정한 방법으로 조건 확률 파라메터로 정점과 에지를 구성하고, 각 정점과 에지에 확률적 성분 값을 연결시켜서 네트워크 확률적 그래프를 생성한다. 을 확률 그래프 에서 확률 함수 $U=(\upsilon,\delta)$ 에 의해 출력되는 확률적 그래프라고 하면, 에서 가 출력될 확률은 G의 에지 출력 확률 곱으로 표현되므로 다음 식 1과 같다.

$$P_R(G,U) = \prod_{\alpha_i \in W} P_r(\nu = \upsilon'(\alpha_i)) \prod_{B_i \in B} P_r(b = \delta'(B_i) | \nu = \upsilon'(\alpha_k)) \quad (1)$$

단, $\alpha \in N, \beta \in E$이고 μ'와 δ'는 각각 μ와 δ의 역함수이고 $U=(u,\delta)$는 $\mu : N \to R_N$와 $\delta : E \to R_E$로 변환하는 확률 함수이다. 식(1)은 확률적 그래프에 있는 모든 정점과 에지가 방향을 제외하고는 독립적이라는 가정에서 얻을 수 있다. 식(1)에서 출력 확률은 정점의 확률과 에지의 확률을 곱으로 얻는다.

수치 해석적 생물학 체인 그래프도 확률적 노드와 확률적 에지들의 집합으로 구성되는데 다중 대 일 확률이 가능하다는 점에서 입증 모델링에 적용 가능하다. 체인 그래프에 대한 정의는 생물학적 확률 그래프의 정의와 비슷하지만 확률적 에지에 몇 개의 특징들이 매칭되어 있어야 세포 간 네트워크가 형성하는지 추론할 수 있다. 생물학적 확률분포 인식에서는 많은 특징점들이 매칭되는 것을 나타내는 확률적 모델 분포가 되는 것이다. 결과적으로 생물학적 확률분포 체인 그래프의 에지는 확률 분포의 방향과 길이를 모델링하기 위해서 사용되고 노드는 네트워크 간의 확률적 연결 형태를 모델링하게 된다.

생물학적 확률분포 체인그래프가 통계적으로 처리하기 때문에 많은 수를 생산하게 되므로 정보 출력 확률은 모든 노드와 에지 확률의 곱으로 세포 간 네트워크 확률 $P(X|U)$은 식 (2)와 같다.

$$P(X|U) = \sum_{\gamma} P(X,\gamma|U) \quad (2)$$

위 식에서 X와 U는 확률적 모델이고 γ는 확률로부터 모든 노드와 에지로 체인되는 확률적 함수이므로 식 (2)는 다시 식 (3)으로 표현된다.

$$P(X,\gamma|U) = \prod_{\alpha \in X} P(\alpha = \Phi(\alpha)) \prod_{\nu \in X} P(B = k(\nu) | \forall \alpha \in A_B, \exists \alpha \in A_\nu, A = \Phi(\alpha)) \quad (3)$$

식3에서 φ는 생물학적 확률 에지 매핑 함수이고, κ는 노드 매핑 함수, a는 에지, ν는 노드, α는 확률에지 β는 확률노드, A_β는 확률노드 β에서 연결된 확률 에지의 집합으로 A_ν는 ν노드로 네트워크가 연결된 확률적 에지의 집합이다. 세포간 네트워크 확률 $P(X|U)$을 구하기 위해 입력에서 모델이 너무 복잡해지기 때문에 이를 현실적으로 접근하기 위해 $\max_\gamma P(X|U)$로 추론되어도 무방하다 사료된다. 이와 같은 확률적 근거는 식 (2)와 식 (3)에서 다시 확률적 추론을 식 (4)와 같이 나타낼 수 있다.

$$P(X|U) = \sum_{\gamma} P(X,\gamma|U) \cong \max_\gamma P(X,\gamma|U) \quad (4)$$

식 (4)에서 생물학적 확률 매칭 함수 와 이 매칭 확률은 동적 프로그래밍 법을 이용하여 확률적 입증을 표현하는데 생물학적 확률 그래프만을 사용할 경우 너무 많은 양의 그래프가 생성되기 때문에 계층적 표현으로 문제를 분할해서 유도한다.

유전자 공유 계층 분배를 바탕으로 한 확률적 계층의

요소를 표현하기 위해서 새로운 추론 모델을 제시하고, 추론 모델을 바탕으로 확률적 계층 표현에서 가장 상위계층에 생물학적 신호경로를 표현하는 확률적 모델이므로 신호경로 모델은 네트워크간의 위치관계에 대해 세포 확률로 표현한다.

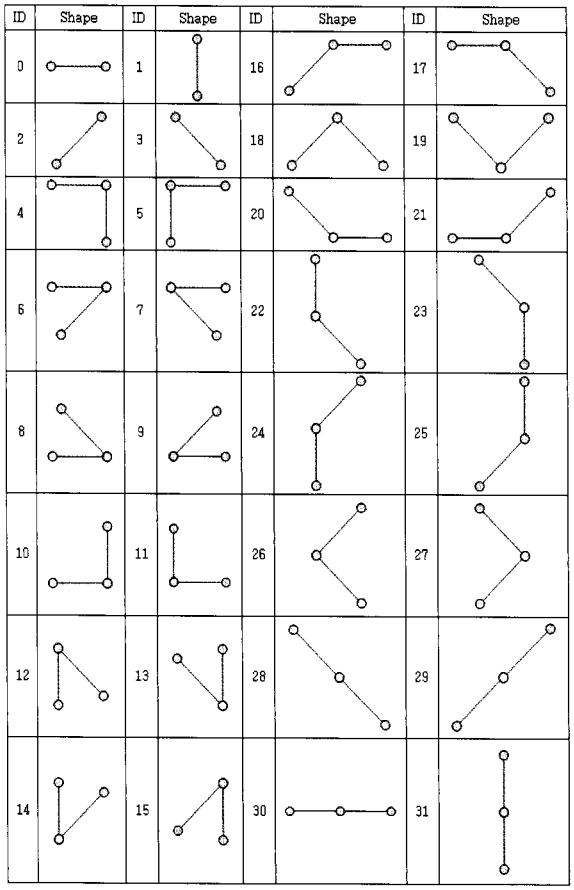

Figure 2. The probability distribution model which is combined with the gene.

Fig. 2는 유전자로 결합된 확률분포로 두 개의 기본 확률분포 사이의 결합을 이용한 32가지로 상호경로의 네트워크 노드로 유도된 것이다. 확률적 상호경로는 동그라미 확률 분포로 결합하였기 때문에 확률적 상호경로 모델의 에지와 노드의 파라메터를 구성하여 확률적 모델에서 각 에지의 생물학적 상호경로 방향에 대한 확률분포가 존재한다. 두 개의 에지가 연결된 부분의 노드에는 두 에지 간의 결합에서 세포 간 네트워크 확률분포 체인이 존재한다.

결합된 확률분포 모델의 매칭은 세포간 인식과정에서 가장 먼저 이루어져야 하는 매칭 과정으로, 세포 인식에서 각 세포와 세포의 접점과 꺽인점이 확률적 그래프의 체인이다. 이 체인 정점들 간의 생물학적 상호 경로 연결 관계를 나타내는 에지를 추출하여 확률적 그래프를 얻는 확률함수 응용이 부분 세포 확률적 모델 그래프에서 두 개의 확률을 가질 수 있도록 다음과 같이 추론할 수 있다.

Fig. 2에서 학습으로 추론된 모델이 관찰된 기저의 분포와 최대한 가까운 모델을 학습하는 것이기 때문에 파라미터 추정과 모델 선택은 모델 구조에 대한 조건적 확률의 파라미터가 된다. 모델 선택에서는 이산적 최적화 문제가 데이터를 잘 매칭되도록 최대화하였고(6), 매칭 분포를 계산하기 위해 임의의 변수 값을 예측하기 위해 선택하였다(7).

유전자 표현 프로파일 모티프

유전자 표현 데이터에서 그래프 모델링까지 확장시킬 수 있는 규칙을 포함해서 세포간 네트워크에 대해 고 대역 처리 데이터의 주된 근원은 유전 표현 프로파일 모티프 (Motifs)이다(8). 유전자표현에서 임의 변수 $X_{g,a}$ 값에서 g는 유전자에 대한 지표이고 a도 분석에 대한 지표이다.

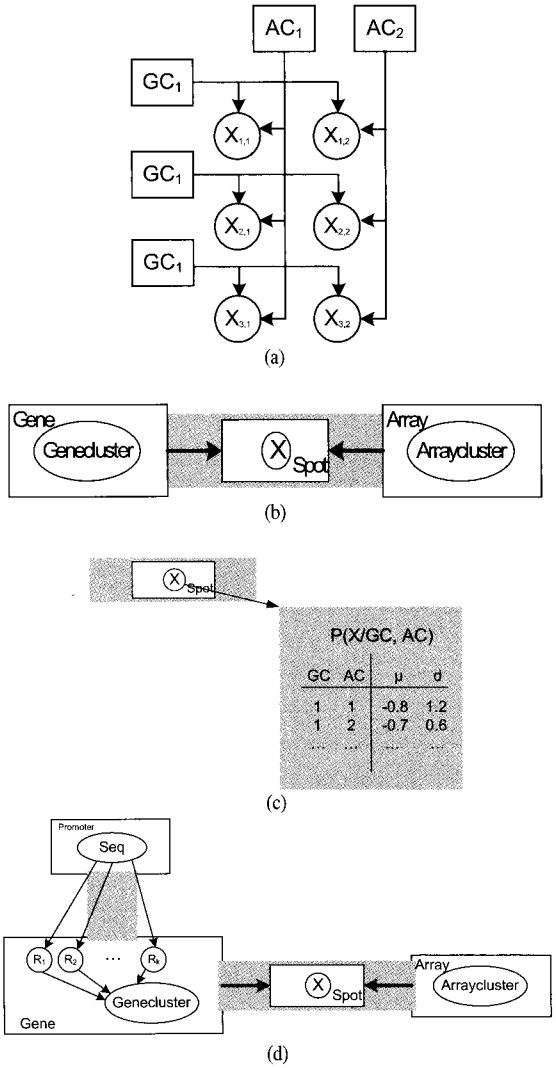

Figure 3. Base network group based on 3 genes and 2 alternatives ((a) Gene network structure, (b) Gene cluster/array model, (c) Gene annotation, (d) Characterization of the group of gene recombination).

모델링은 Fig. 3과 같이 유전자가 동일하게 표현된 유전자 클러스터로 분할된다는 것과 각 클러스터 내의 유전자가 각 배열에서 전형적으로 표현 레벨을 가지기 때문에 대열집단으로 분할된다고 가정하므로 유전자 표현은 같은

배열집단에 속한 마이크로 배열과 비슷하다. 그러므로 임의의 변수를 더하여 모델을 제시할 때 이 모델에서 Gene $Cluster_g$을 유전자 g의 집단으로 할당하면 Array $Cluster_g$는 대열 a의 집단 할당량을 나타낸다.

대열집단의 유전자 g는 유전표현 Gene$Cluster_g$와 Array $Cluster_g$에 달려있다. 그러므로 이 모델은 특정 유전자 집단과 대열집단에 대응하는 모든 검측을 같은 조건 분류에 의해 통제된다고 할 때 모델을 베이스의 네트워크로 각 표현 특성 $X_{g,a}$가 Gene$Cluster_g$와 Array$Cluster_g$의 근본이 된다. 실제의 네트워크 구조는 데이터에 있는 유전자의 수와 대열에 따라 달라지기 때문에 베이스의 네트워크로는 두 가지 중요한 측면을 명확하게 나타낼 수 있다. 첫째, 임의 변수는 유전자와 같은 각 실체의 특성을 나타낸다. 둘째 템플릿은 같은 타입의 모든 엔티티에 의해 공유되고 P($X_{g,a}$ IGene$Cluster_g$, Array$Cluster_g$)의 조건적 확률은 g와 a의 선택이 비슷하므로 더 간결한 모델을 구현한다. 유전자집단 문제에 대한 템플릿 모델을 보여주고 유전자와 대열만 주어지면 베이스의 네트워크를 발생시킨다.

집단과 유전자 할당량이 원래의 검측에서 동일한 유전자 표현을 가진 각 블록으로 분할할 가능성이 매우 높다 (9). 여기서 E-step과 M-step 사이에서 반복되는 기대의 최댓값을 사용하여 분배를 찾을 수 있다. E-step은 유전자 대열의 확률적인 집단배분을 찾기 위해 현 파라미터를 사용하며 M-step은 이 할당량을 토대로 각 유전자/배열집단 조합 내의 분포를 재 추정한다. 이 모델에서 생물학적 메카니즘에 대한 통찰력을 얻는 데에도 확장될 수 있으며, 유전자의 동일표현이 동일규칙에 반영되므로 핵심적인 조절 메카니즘에는 유전자 촉매 인자 부분에 전사 인자가 결합되는 것이 포함된다. 따라서 유전자의 촉매 인자 지역에 있는 전사인자 결합 지점을 찾아낸다. 이와 같이 유전자 집단 내에 있는 유전자들은 유사한 표현방식을 나타내고 각 집단을 조절하는 전사 인자와 연관시키므로 생물학적 메카니즘을 과도하게 단순화한 면도 있지만 이를 통해 유전자의 표현과 조절 사이의 여러 가지 차이점에서 최우선 신호를 얻을 수 있다.

직관적 통찰을 수학적 모델로 전환시키는 방법에서 하나는 특징적인 결합지점을 가지고 있는 촉매에 할당(annotate)하고 그 후 이것을 유전자 실체의 새로운 특성으로 사용하는 것이다. 임의적인 이원변수 (binary varialble) $R_{g,j}$는 유전자 g가 전사인자 j와 결합지점을 가지고 있고 각 유전자의 집단 할당량이 관련 결합지점과 표현 특성 (10)에 직접 영향을 주는 모델이다. 실제 촉매 시퀀스가 주어질 때 결합지점의 확률을 모델링 함으로써 모델을 증대시킬 수 있고 촉매 지역과 모델 촉매 시퀀스 Seq_g에 따라서 $R_{g,j}$을 나타내는 새 객체가 된다. 이 조건적 확률 파라미터는 전사인자가 인식하는 특정한 모티프를 규정한다.

수학적 모델 구성은 Gene$Cluster_g$와 연관된 조건적 분포의 표현으로 촉매에 포함하고 결합지점이 존재하는 것이 유전자가 어떤 집단(11)에 속하는지를 예측할 수 있다. 그러나 생물학적 조절에서 어느 하나가 기저의 논리에 매칭되는지는 명확하지 않다고 본다. 이런 종류의 모델은 유전자의 유사성과 촉매 측면에서 결합시킬 수 있기 때문에 모델 집단에 할당된 유전자를 설명하는 새로운 결합지점을 찾는 단계가 있고, 또 그들의 표현 프로파일과 촉매 지역을 토대로 한 집단에 유전자를 재배분하는 단계로 본다. 각 집단 내 표현의 배분을 재평가하는 단계는 학습의 두 종류의 데이터 사이 정보의 흐름을 포함하고 약간의 결함도 허용한다. 이러한 정보는 유전자 집단 변수에 의해 전해지기 때문에 유전자 집단 변수는 통일성 있는 유전자 표현 프로파일과 유사한 촉매들을 다 가진 유전자 집단을 나타내게 된다. 동시에 학습절차는 이러한 집단 배분이 일어날 결합지점 모티프를 확인한다.

조절 네트워크 재구성

유전자 표현 분석은 조절 네트워크 재구성으로 유전자 A의 표현이 단백질 B와 C에 의해 조절된다면 A의 표현 레벨은 B와 C를 접합해 활동 레벨이 구성된다. 생물학에서 검측 잡음의 변화가능성 때문에, A의 표현이 확률적인 기능으로 다루어졌다. 생물학 데이터부분에서 유전자 표현 레벨을 각 배열 $X_{g,a}$가 그 배열에서 조절자의 표현 레벨에 의존하도록 각 유전자 g을 위해 베이스 네트워크를 찾고자 한다. 또 베이스 네트워크에서 구조학습을 할 때 다양한 네트워크 구조를 검출하여 평가되므로 조절네트워크 재구성은 두 가지로 나타낸다.

첫번째는 통계치의 신뢰성 때문에 표현 레벨의 수십 가지 유전자를 연관시키는 네트워크는 문제가 될 수 있으므로 유전자의 확률적 기능을 제한하는 생물학 원리에 기반할 수도 있다.

두번째 생물학적 해석 방법(11)은 직·간접적인 조절과 발견에서 재구성 전사에 의해 관계를 찾아낸다.

Fig. 4(a)는 유전자 표현 프로파일 데이터에 대해 부트스트랩 절차를 적용하여 어떤 유전자가 조절 인지에 대한 사전 지식을 가지지 않고 네트워크 구조에서 유전자 A와 B가 진행되는 것을 나타낸다. Fig 4(b)는 학습된 네트워크와의 사이에서 높은 신뢰로 하부 네트워크를 유전자의 동의 하에 다양한 관계에 의해 유전자 모델 할당을 조절 비교하여 어떤 중간 단계에서 관련된 조절이 함께 반응하는지를 나타낸다.

Fig. 4(c)는 두 가지 연구에서 통찰력을 갖춘 신호 경로와 수용체 등 전사가 포함되었으며 다른 유전자 표현은 Fig. 4에서 네트워크 구성에 대한 제약의 학습절차가 데이터 중에서 가장 두드러진 재구성 표현 유전자로 나타났다. 모듈 네트워크에서 또 하나의 중요한 측면은 트리 묘사가 더 쉬운 해석을 가지므로 유전자의 조절 프로그램에서 적은 숫자의 모듈이 집중되었다는 것이다.

유전자 모델에서 프로파일 학습 방법은 타겟 유전자 레벨에서 변화 값을 명확히 탐지하여 조절 메카니즘에서 피드포워드 루프, 피드백 등을 수반해 단백질이 변화되면 유전자 표현 레벨도 변화시킨다. 그 결과 타깃의 표현 레벨에 있어서 단백질-DNA와 단백질-단백질 상호작용으로 업데이트된 변화를 탐지할 수 있다. 결과에서 하나는 Fig. 4(c)의 규정과 같은 동일표현을 구분하는 것으로 유전자

베이스 네트워크 논리를 철저한 유전자 조절의 간접적 방법으로 추론하였다.

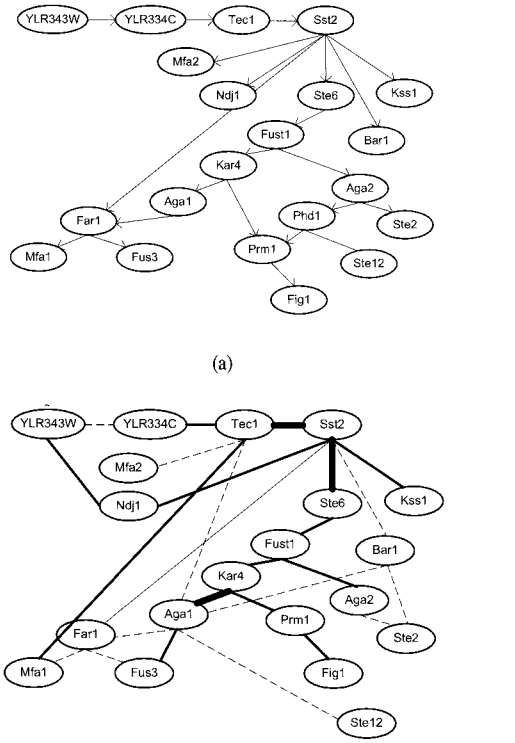

Figure 4. Different network structure models ((a) Gene expression structure, (b) Gene network control, (c) Acceptance rule of signalling pathway).

결 론

데이터로부터 세포간 네트워크를 학습하기 위한 몇 가지 모델 접근법에 의해 유전자 표현 레벨을 논의하여 생물학 지식과 독립적인 분석에 의한 검증을 수행하였다.

생물학적 통찰력을 제공하는 모델을 얻기 위해 세포 간 및 다양한 레벨에서의 유전자 신호경로 수용을 조합하는 통합된 모델에 초점을 두어 데이터를 통합시켰다. 생물학적 원리에서 모델 디자인 통합을 학습하여 실제 세부사항으로 포착하여 네트워크 모델을 제시하였다. 고처리 데이터의 양과 다양성이 증가하는 것과 같이 3개의 유전자와 2개의 배열을 단일화한 베이스 네트워크 집단에서 서로 다른 네트워크구조 모델로 새로운 분석, 새로운 디자인을 포함한 단일 세포, 합성 기관, 전체적 유기체, 그리고 확률적 그래프 모델 수준에서의 시스템으로 나타냈다. 수치 해석적 분석법으로 데이터로부터 생물학적 신호경로 통찰력을 얻는데 표현했고, 모든 분석적 필요를 충족시키는데 추론하였다.

앞으로 다양한 세포 간 요소에서 모델링 선택과 분석방식의 적합성으로 모델의 추론과 학습을 위한 방법들을 확장시켜서 정보공학 네트워크에 더 연구한다.

요 약

유전자 생물학 분야에서 적용가능 한 세포간 네트워크를 입증하는 고처리 정보공학에 응용하려는 수치학적인 표현 모델을 분석 연구한다. 확률적 그래프 모델을 사용하여 데이터 네트워크로부터 생물학적 통찰력을 확률적 함수적으로 응용해 복잡한 세포간 네트워크보다 단순한 하부모델로 구성하여 유전자 베이스네트워크 논리를 유전자 표현 레벨로 나타낸다. 유전자 데이터로부터 확률적 그래프 모델들을 분석하여 유전자 표현 데이터를 정보공학 네트워크 모델의 방법으로 확장 추론한다.

REFERENCES

1. Lander, E. (1999), Array of hope, *Nature Genet.* **21**, 3-4.
2. Pilpel, Y., P. Sudarsanam, and G. Church (2001), Identifying regulatory networks by combinatorial analysis of promoter elements, *Nature Genet.* **29**, 153-159.
3. Pearl, J. (2005), Probabilistic reasoning in intelligent systems: networks of plausible inference, Morgan Kaufmann, San Francisco.
4. Buntine, W. (1994), Operations for learning with graphical models. *J. Artif. Intel. Res.* **2**, 159 - 225.
5. Getoor, L., N. Friedman, D. Koller, and A. Pfeffer (2001), In *Relational Data Mining, Learning Probabilistic Relational Models*, S. Dzeroski and N. Lavrac, Eds., pp.307-337. Springer-Verlag, Berlin.
6. Segal, E., B. Taskar, A. Gasch, N. Friedman, and D. Koller (2001), Rich probabilistic models for gene expression, *Bioinformatics* **17**(suppl. 1), S243.
7. Barash, Y. and N. Friedman (2002), Context-specific Bayesian clustering for gene expression data, *J. Comp. Biol.* **9**, 169-191.
8. Holmes, I. and W. Bruno (2000), Finding regulatory elements using joint likelihoods for sequence and expression profile data, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **8**, 202-210.
9. Segal, E., R. Yelensky, and D. Koller (2003), Genome-wide discovery of transcriptional modules from DNA sequence and gene expression, *Bioinformatics* **19**(suppl. 1), I273- I282.
10. Gasch, A. P. et al. (2000), Genomic expression programs in the response of yeast cells to environmental changes, *Mol. Biol. Cell* **11**, 4241-57.
11. Lee, T. et al. (2002), Transcriptional regulatory networks in *Saccharomyces cerevisiae*, *Science* **298**, 799-804.