

하이퍼네트워크에서 본 단어간 긴밀성과 다양성

(Affinity and Variety between Words in
the Framework of Hypernetwork)

김 준 식 [†] 박 찬 훈 [‡] 이 은 석 ^{***} 장 병 탁 ^{****}
(Joon-Shik Kim) (Chan-Hoon Park) (Eun-Seok Lee) (Byoung-Tak Zhang)

요약 전체문서(corpus)에서의 두 단어 간 연결 상태를 파악하여 앞 단어 다음에 오는 단어의 빈도 수를 기반으로 여러 형태의 그룹을 분류하여 단어 간 다양성과 긴밀성을 살펴보았다. 기존의 연구에서 Zipf's Power Law는 Chinese Restaurant Process로 설명되었고 Scale Free Network에서는 edged의 수에 따른 노드의 profile을 조사하여 hub들을 찾는 연구가 수행되었다. 본 연구에서는 단어 간 연결의 유일성과 다양성을 조사하여 Zipf's Power Law와 hub profile을 동시에 살펴보았다. 데이터 분석 결과 단어 간 연결의 긴밀성과 다양성 사이에서 대칭성으로 힘축되는 유의한 결과를 얻었으며 이는 소위 'exploitation'과 'exploration'의 관점에서 설명될 수 있다. 또한 분석 자료인 TIPSTER에서 관찰된 약간의 대칭성 깨짐(symmetry breaking)에 대해서도 논한다.

키워드 : 다양성, 긴밀성, 지프의 지수 법칙, 대칭성, 이용, 탐색

Abstract We studied the variety and affinity between the successive words in the text document. A number of groups were defined by the frequency of a following word in the whole text (corpus). In the previous studies, the Zipf's power law was explained by Chinese restaurant process and hub node was searched after by examining the edge number profile in scale free network. We have observed both a power law and a hub profile at the same time by studying the conditional frequency and degeneracy of a group. A symmetry between the affinity and the variety between words were found during the data analysis. And this phenomenon can be explained within a viewpoint of "exploitation and exploration." We also remark on a small symmetry breaking phenomenon in TIPSTER data.

Key words : variety, affinity, Zipf's power law, symmetry, exploitation, exploration

- TIPSTER data를 제공해주시고 좋은 논의를 해주신 엄재홍씨께 감사를 드립니다. 이 연구는 한국 과학기술부의 국가지정연구소(NRL) 프로젝트와 산업자원부의 분자진화연산(MEC) 프로젝트 그리고 교육인적자원부의 BK21-IT의 지원을 받았습니다.
- 이 논문은 2007 한국컴퓨터종합학술대회에서 '하이퍼네트워크 관점에서 본 문서에서의 단어간 긴밀성과 다양성의 대칭성'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 서울대학교 물리천문학과
jskim.ozmagi@gmail.com

‡ 학생회원 : 서울대학교 컴퓨터공학과
chpark@bi.snu.ac.kr

*** 학생회원 : 서울대학교 인지과학협동과정
eunseok.lee@gmail.com

**** 종신회원 : 서울대학교 컴퓨터 공학부 교수
btzhang@bi.snu.ac.kr

논문접수 : 2007년 9월 27일

심사완료 : 2008년 1월 23일

Copyright @2008 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복제본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문서와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용 행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 시스템 및 이론 제35권 제4호(2008.4)

1. 서 론

Zipf's law로 불리는 ranking profile의 power law 분포는 여러 물리 생물 시스템에서 나타난다. 그 예로 단어의 빈도수[1], 지진의 규모[2], 우주에서 발광체의 분포[3], 메타볼리즘[4], 유전자 발현[5] 등의 시스템을 들 수 있다. 여기서 관찰되는 power law를 설명하는 방법도 다양하여, scale free network[6], Chinese Restaurant Process[7], 세포 단위의 동력학 법칙[5], 그리고 Markov Chain Monte Carlo[8] 등 많은 방법들을 그 예로 들 수 있다. 이는 Zipf's law의 보편성을 증거하는 것에 다름 아닌바, 이에 본 연구는 이러한 성질을 바탕으로 하이퍼네트워크(hypernetwork)[9,10] 관점에서 단어들을 표상하는 노드들 간의 연결 관계를 설명하고자 한다. 즉 하이퍼네트워크 관점에서 정리된 기억의 기본 소자 양식을 이용하여 문서 속에서의 단어 간 연결

을 count 수로 나타내는 방법을 사용한다. 이 방법은 하이퍼네트워크를 기억 인출이나[9], 문서 분류[10]의 기능을 수행하는 측면보다는 기억의 기본 소자로서 이용하여 전체문서 혹은 기억의 통계적인 분포를 알아보고자 하는 데 사용하고자 하는 것이다.

하이퍼네트워크 모델은 학습과 메모리의 random graph 모델로서 세포의 문자 네트워크 구조로부터 유래하였다. 한번, 하이퍼네트워크는 가중치가 부여된 하이퍼그래프(hypergraph[11])로서 설명될 수 있는데, 하이퍼그래프를 구성하는 k개의 vertice로 이루어진 k-order 하이퍼에지는 그 edge의 가중치를 통해 하이퍼네트워크를 구성한 패턴과 그 성분(vertice)간의 k-order correlation을 나타낼 수 있다. 이미 하이퍼네트워크의 이러한 특성을 활용하여 다양한 분야에서 패턴 인식 및 완성에 관련된 실험[9,10,12]이 이루어진 바 있다.

본 논문에서는 전체문서(corpus)와 그 일부를 복제한 부분문서(query)를 통해 단어 간의 연결에서의 긴밀성과 다양성 간에 대칭적인 성질이 있으며 이는 기계학습의 이용(exploitation)과 탐색(exploration)으로 설명할 수 있음을 논의 하고자 한다. 우리는 양자역학의 원자모델에서의 에너지 준위(energy level)를 유추(analogy-making)했고, 같은 에너지를 가지는 자유도의 개수인 축퇴도(degeneracy)라는 물리량을 사용하여 문서를 분석해보았다.

부분문서(query)의 단어들을 따라가며 전체문서에서 그 단어 다음에 올 수 있는 모든 인접 단어를 조사하여 이를 그 빈도수(frequency)에 따라 분류한 그룹으로 나누었다. 우리는 다시 Markov Chain Monte Carlo 모델을 적용하여 부분문서의 단어 다음에 오는 인접 단어가 속한 그룹의 빈도수(frequency)와 그 축퇴도(degeneracy)를 살펴보았다. 여기서 축퇴도(degeneracy)란 같은 빈도수(frequency)를 가지는 모든 단어들의 개수를 의미한다. 빈도수(frequency)는 다음 단어와의 연결의 긴밀함을 나타내고 축퇴도(degeneracy)는 다양성을 나타낸다. 실제로 이러한 열역학적이고 Markov Chain Monte Carlo(MCMC)적인 방법은 DNA computing의 열역학적 모델링의 패러다임을 따른다고 볼 수 있다[13]. 여기에서 DNA 간의 entropy와 degeneracy가 log 함수로 연관되어 있고 enthalpy와 frequency가 역시 log 함수로 연결되어 있으며 Gibbs free energy는 degeneracy와 frequency의 곱의 마이너스 로그 값임을 밝혀둔다.

우리는 부분문서의 단어들의 전개에 따른 group들은 degeneracy와 frequency의 정보를 분석하여 보았고 이 다양성과 긴밀성을 나타내는 척도의 분포는 상당히 대칭적(symmetrical)임을 관찰하였다. 또한 단어 다음에 오는 그룹의 상대적 순위(relative ranking, group rank-

ing/number of total groups)의 분포가 상위 부분과 하위 부분이 대칭적으로 높게 나옴을 관찰했다. 이는 상대적 순위(relative ranking)의 시계열 분포로부터 알 수 있듯이 중간 ranking에서 상대적으로 상위와 하위 ranking으로 주도적으로 전이되는 현상에 기인한다. 마지막으로 Gibbs free energy와 frequency 그리고 Gibbs free energy와 degeneracy 그래프로부터 frequency와 degeneracy가 부분적으로 깨짐(partial symmetry breaking) 현상을 관찰했다. 이는 높은 frequency를 가지는 단어들은 그 degeneracy가 1의 값만 가지지만 높은 degeneracy를 가지는 단어들은 1, 2, 3, 4, 5 등의 다섯 가지 이상의 frequency를 가지기 때문으로 분석된다. 이 현상으로부터 가능한 해석은, 이용(exploitation)은 한 가지 단어만을 계속하여 연결하는 것이 유리하고 탐색(exploration)은 여러 가지 경우를 가지는 것이 유리하기 때문이라는 점이다.

2. 이론

양자역학에서, Potential 함수에 갇힌 입자는 discrete 한 에너지를 가지며 각 에너지 준위에는 정해진 자유도(degree of freedom)이 부여된다. 예를 들자면, 원자핵에 잡힌 전자들의 상태들이나 양성자와 중성자로 이루어진 원자핵의 상태가 그러하다[14]. 여기서 에너지가 음수로 큰 값을 가질 경우 자유도를 나타내는 degeneracy는 작은 값을 가지고 음수로 작은 값을 가지는 높은 에너지 준위에서 degeneracy는 큰 값을 가진다(그림 1). 이와 비슷한 현상이 Hypernetwork 관점에서 분석한 인접한 단어들 간에도 관찰되었다. 즉, 빈도수(frequency)가 높은 단어의 축퇴도는 대체로 1이었고 빈도수(frequency)가 낮은 단어의 축퇴도(degeneracy)는 큰 값을 가졌다. 그림 2에서 두 척도의 상관관계를 보인다.

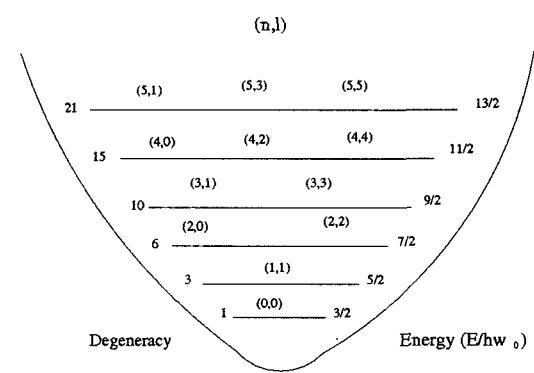


그림 1 삼차원 조화진동자(harmonic oscillator)의 degeneracy와 energy[14], 여기서 n 은 주양자수이고 l 은 궤도 양자수이다.

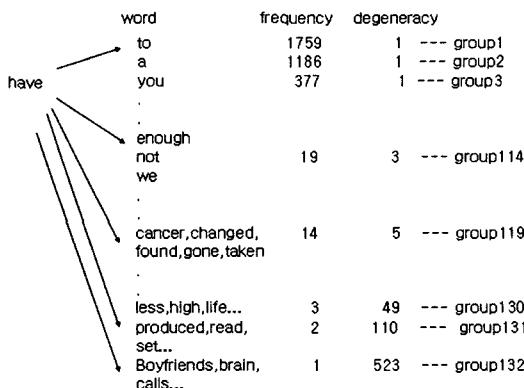


그림 2 단어의 frequency를 기준으로 분류한 group, frequency, degeneracy의 예

여기서 degeneracy = $\exp(\text{entropy}/k)$ 와 frequency = $\exp(-\text{enthalpy}/kT)$ 의 열역학적 관계가 도입된다. k 는 Boltzmann 상수이고 T 는 절대 온도이다. 이로 볼 때, Gibbs free energy = enthalpy - $T \cdot \text{entropy}$ 이며 이는 $-kT \cdot \ln (\text{degeneracy} \cdot \text{frequency})$ 로 주어진다. 위의 Gibbs free energy를 도입하여 부분문서(query)의 단어 간 진행을 Markov Chain Monte Carlo(MCMC) 모델로 설명할 수 있다.

3. 실험 과정

입력된 텍스트를 하이퍼네트워크로 재구성하였다. 하이퍼네트워크는 order 2의 하이퍼에지(hyper edge)로 구성되어 있으며 이는 2개의 단어(feature)가 연결되었음을 표상한다. 하이퍼에지를 구성하는 2개의 단어는 문장을 기본 단위로 하여 연속되어 조개진다. 즉, "A B C"의 3개 단어로 이루어진 문장의 경우 "A B"와 "B C" 두개의 하이퍼에지를 생성한다. 하이퍼에지(또는 library elements, 동일하게 사용)는 동일한 하이퍼에지가 많이 존재할수록 더 큰 weight를 갖게 된다. 즉 문장을 조개어 생성된 각각의 하이퍼에지는 기존에 동일한 하이퍼에지가 존재할 경우 그 하이퍼에지의 count를 하나(1)씩 증가시켜주며, 그렇지 않을 경우 새로운 하이퍼에지가 된다.

프로그램 상에서는 각 하이퍼에지마다 counter를 달아서 동일한 하이퍼에지가 입력될 경우 그 counter의 값을 1 증가시키도록 구현하였다. 전체문서(corpus)에 대하여 모든 문장을 하이퍼에지로 전환하면, 하이퍼네트워크가 구성된다. 구성된 하이퍼네트워크에 대하여 연속된 부분문서(query)를 입력하였다. query는 마찬가지로 문장 수준으로 들어오며, 이 문장을 initialization과 마찬가지로 order 2의 하이퍼에지로 변환하여 matching을

수행한다.

각 query 문장으로부터 생성된 하이퍼에지 각각은 만 들어둔 하이퍼 네트워크와의 matching을 통해 corpus의 정보를 분석한다. query 하이퍼에지들은 자신의 첫 번째 vertex(단어)와 동일한 vertex를 포함하고 있는 하이퍼네트워크 상의 모든 하이퍼에지들을 검색하며, 그 하이퍼에지들의 다른 한쪽 vertex 정보를 각 query 별로 저장한다. 이때 저장되는 정보들은 각각의 query 하이퍼에지마다

a. query 하이퍼에지와 첫 번째 vertex가 match되는 모든 하이퍼에지의 수(total number of group)

b. query 하이퍼에지의 다른 한쪽 vertex를 포함하는 하이퍼에지가 속하는 group의 번호(appointed group ranking)

* 각 그룹은 동일한 frequency(= count)를 갖는 하이퍼 네트워크상의 하이퍼에지의 집합으로 구성된다.

c. b group의 원소의 개수(degeneracy)

d. b group의 frequency이다.

실험 문서로는 TIPSTER VOL1의 AP data를 썼으며 전체문서(corpus)는 2.2MByte이고 이중 처음의 56 KByte 분량을 복사하여 부분문서(query)로 사용하였다.

4. 실험 결과

우리는 앞으로 제시되는 그림 3-8, 그리고 9를 통하여 빈도수(frequency)와 축퇴도(degeneracy) 사이의 대칭성을 살펴보고 약간의 대칭성 깨짐 현상도 관찰할 것이다. 각 결과의 설명은 각 그림의 설명에 자세히 기술하였다.

Zipf law는 많은 시스템에서 관측되며, 본 논문에서는 Zipf law가 관찰되는 시스템의 하나인 문서를 Hyper-network 관점에서 분석해보았다. 이 때 열역학적 모델에 따른 MCMC 방법을 가정하였다. 실험 결과는 상대적 relative ranking의 상위와 하위 부분이 대칭적으로 큰 값을 가짐이 관측되었다(그림 7). 이로부터 frequency와 degeneracy 간의 대칭성(symmetry)을 볼 수 있었다. 이와는 약간 반대되는 성질인 frequency와 degeneracy에서의 약간의 대칭성 깨짐(symmetry breaking) 현상도 Gibbs free energy와 각각의 두 변수(frequency, degeneracy)의 그림에서 관찰할 수 있었다(그림 8, 9). 위의 관찰 결과들은 exploitation과 exploration의 효과적 사용으로 짜임새 있는 의미 전달 가능한 문서가 완성된다는 의견을 뒷받침한다. 즉, 빈도수가 높은 단어는 하나의 축퇴도만을 가짐으로써 그 이용이 극대화되고, 축퇴도가 높은 단어들은 다양성이 중요하므로 빈도수가 1, 2, 3, 4, 그리고 5인 경우를 가진다. 여기에서의 짜임새 있는 문서의 기본 단위는 문장이며

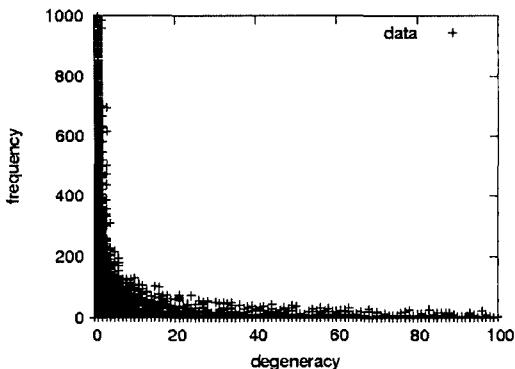


그림 3 TIPSTER data의 AP 문서를 Hypernetwork로 분석했을 때 이웃하는 단어 간의 전체문서에서의 빈도수(frequency) 다양성 혹은 축퇴도(degeneracy)의 분포. 우리는 이 그림에서 단어 간 연결 상태는 빈도수가 높거나 혹은 축퇴도가 높거나 하는 대칭적인 분포를 가짐을 알 수 있다. 즉 스케일을 무시하면 $y=x$ 직선에 대해서 대칭이다.

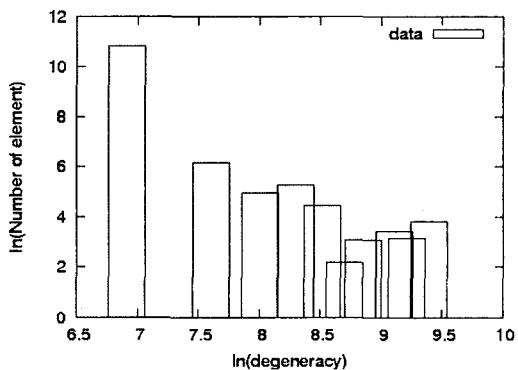


그림 4 그림 3의 data 들을 degeneracy 값들에 따라 도수 분포표를 그려보았다. x, y 값은 자연로그를 취한 값들이다. 데이터들이 선형적 감소를 보인다. 이 결과는 degeneracy profile의 power law를 따름을 보인다.

이 논문의 단어 간 통계는 의미가 있는 한 문장 내에서의 단어 간 통계이다. 또한 power law distribution은 self organized criticality(SOC)에서 기인한다는 연구를 통해[15] 잘 짜여진 문서는 뇌의 neuron들의 SOC 상태에서 나온 power law 성질을 가지는 결과물이라고 추론할 수 있다.

5. 결 론

문서에서의 분석 결과를 가지고 인적 네트워크에 적용하여 생각해볼 수도 있다. 우리가 학교 혹은 직장에서

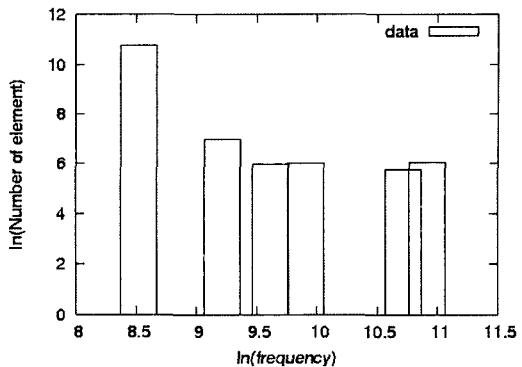


그림 5 그림 3의 data들을 frequency 값들에 따라 도수 분포표를 그려보았다. x, y 값들은 자연로그를 취한 값들이다. 처음 4개의 데이터에서 선형적 감소를 볼 수 있고 그 옆에 두 개의 독립된 data들을 관찰할 수 있다.

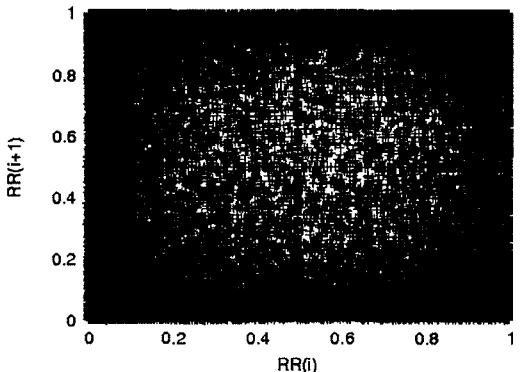


그림 6 이웃하는 단어의 그룹(group)의 상대 순위(relative ranking)의 전이 그림. 그룹은 빈도수(frequency)가 같은 모임을 뜻하며 앞 단어 다음에 올 수 있는 단어들을 그 빈도수 기준으로 그룹을 나누고 실제 부분문서(query)에서 선택되어지는 뒤 단어의 그룹 순위를 구한다. 위 그림에서 x 축은 i 번째 단어의 그룹순위를 전체 그룹의 개수로 나눈 상대 순위를 나타내며 y 축은 그 다음에 연결되는 $(i+1)$ 번째 단어의 상대 그룹 순위이다. 이 그림에서 x 값이 0.1 이하이거나 0.9 이상이면 단어 연결은 모든 상대 순위 그룹으로 전이하지만 그 사이에서는 상위 그룹 (y 값이 0.2 이하) 혹은 하위 그룹(y 값이 0.8 이상)으로 우세하게 전이됨을 볼 수 있다. 실제로 다음 그림 7은 이를 잘 보여준다.

빈도수가 많게 매일 보는 상사와 동료들이 있는가 하면 일 년에 두세 번 정도로 드물게 만나는 친구들과 고객

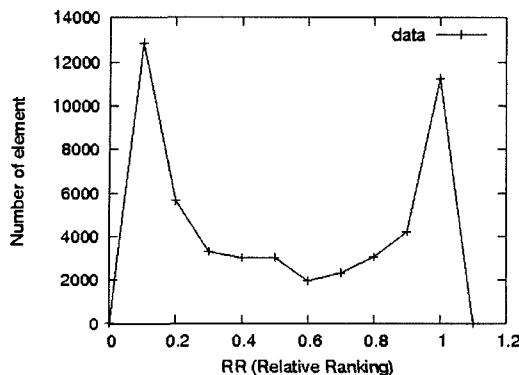


그림 7 그룹의 상대 순위의 도수 분포를 나타낸다. 0.2 이하의 상위 그룹과 0.8 이상의 하위 그룹에 데 이타가 몰려 있음을 볼 수 있다. 상위 그룹은 하위 그룹에 비해 빈도수(frequency)가 높고 축퇴도(degeneracy)가 낮으며 하위 그룹은 그 반대이다. 이 그림에서 우리는 빈도수와 축퇴도사이의 대칭성을 볼 수 있다. 즉 $x=0.5$ 직선을 중심으로 좌우 대칭이다.

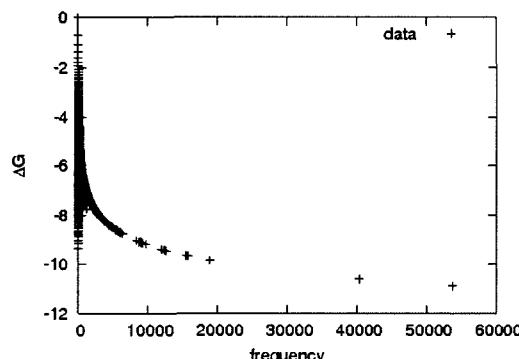


그림 8 x 축은 빈도수이며 y 축은 $kT=1$ 일 때의 Gibbs free energy 즉 $-\ln(\text{frequency} \cdot \text{degeneracy})$ 이다. 여기서 k 는 Boltzmann 상수이고 T 는 절대 온도이다. 하나의 branch가 보이며 이는 degeneracy가 1일 때의 data들이다.

들도 있다. 여기에서 빈도수가 많은 친구들은 비교적 수가 적으며 드물게 만나는 친구들은 다양하게 많다는 유추를 취해볼 수 있다.

마지막으로 왜 문서의 단어 간 연결 관계가 두 극단, 즉 빈도수와 축퇴도에 치우쳐 있는지를 설명하려면 우리는 Gibbs free energy minimization의 관점에서 이 문제를 볼 필요가 있다. 즉 Gibbs free energy = enthalpy $-T \cdot \text{entropy}$ 이며 빈도수가 큰 극단은 enthalpy가 음수로 작아져서 Gibbs free energy가 작아지는 경우이다. 또한 축퇴도가 큰 극단은 entropy가 양수로

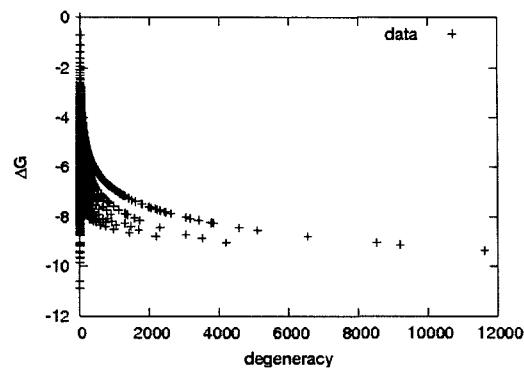


그림 9 위 그림의 x 축은 축퇴도이며 y 축은 Gibbs free energy이다. 다섯 가지의 branch를 볼 수 있으며 이는 각각 빈도수가 1, 2, 3, 4, 5일 때의 data 들이다. 그림 8과 비교하여 frequency와 degeneracy 간의 약간의 대칭성 깨짐(symmetry breaking) 현상을 관찰할 수 있다. 즉 frequency와 degeneracy를 바꾸어보았을 때 약간의 다른 차이를 발견할 수 있다. 여기서는 높은 frequency일 때 degeneracy는 1이지만 높은 degeneracy일 때 frequency는 1, 2, 3, 4, 5 등의 값을 가지는 것이 그 이유이다.

큰 경우이며 역시 Gibbs free energy가 최소화된다. 이러한 minimization principle와 문서의 의미 전달 최적화 사이에 구체적으로 어떠한 관계가 있는지는 좀 더 연구해 볼 필요가 있다. 또한 유전자(gene)를 단어로 볼 경우 DNA 상의 유전정보를 문장과 문서로 볼 수 있고 이 논문에서 수행한 방법들을 유전자 문서에 적용시켜 볼 수도 있을 것이다.

참 고 문 헌

- [1] Steyvers, M., Griffiths, T. L., and Dennis, S., "Probabilistic inference in human semantic memory," TRENDS in cognitive science Vol.10, No.7, pp. 327-334, 1998.
- [2] Bak, P., Christensen, K., Danon, L., and Scanlon, T., "Unified scaling law for earthquakes," Physical Review Letters Vol.88, No.17 p. 178501, 2002.
- [3] Bak, P and Chen, K., "Scale dependent dimension of luminous matter in the universe," Physical Review Letters Vol.86, No.19, pp. 4215-4218, 2001.
- [4] Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabasi, A.-L., "Hierarchical organization of modularity in metabolic networks," Science Vol.297, No.5586, pp. 1551-1555, 2002.
- [5] Furusawa, C., "Zipf's law in gene expression," Physical Review Letters Vol.90, No.8, p. 088102, 2003.

- [6] Barabasi A. -L., and Albert, R., "Emergence of scaling in random networks," *Science* Vol.286, No.5439, pp. 509~512, 1999.
- [7] Goldwater, S., Griffiths, T.L., and Johnson, M., "Interpolating between types and tokens by estimating power-law generators," *Advances in Neural Information Processing Systems* Vol.18, pp. 459~466, 2006.
- [8] Kechedzhi, K.E., Usatenko, O. V., and Yampolskii V. A., "Rank distribution of words in correlated symbolic systems and the Zipf law," *Physical Review E* Vol.72, p. 046138, 2005.
- [9] Zhang, B.-T., and Kim, J.-K., "DNA hypernetworks for information storage and retrieval," *Lecture Notes in Computer Science, DNA12*, Vol.4287, pp. 298~307, 2006.
- [10] Kim, S., Heo, M.-O., and Zhang, B.-T., "Text classifier evolved on a simulated DNA computer," *IEEE Congress on Evolutionary Computation (CEC 2006)*, pp. 9196~9202, 2006.
- [11] Berge, C., *Graphs and Hypergraphs*, p.389, North-Holland Publishing, Amsterdam, 1973.
- [12] Ha, J.-W., Eom, J.-H., Kim, S.-C., and Zhang, B.-T., "Evolutionary hypernetwork models for aptamer-based cardiovascular disease diagnosis," *The Genetic and Evolutionary Computation Conference (GECCO 2007)*, Vol.4, pp. 2709~2716, 2007.
- [13] 김준식, 김종찬, 노영균, 이동윤, 장병탁, "DNA 컴퓨팅 연산 과정의 통계 물리적 예측," *한국컴퓨터종합학술 대회 2005 논문집*, 제32권 제1(B)호, pp. 253~355, 2005.
- [14] Krane, K. S., *Introductory Nuclear Physics*, p.33, John Wiley & Sons, Inc, 1988.
- [15] Maslov, S., Paczuski, M., and Bak, P., "Avalanches and 1/f noise in evolution and growth models," *Physical Review Letters* Vol.73, No.16, pp. 2162~2165, 1994.



박 찬 훈

2006년 8월 아주대학교 정보 및 컴퓨터 공학부 학사. 2006년 8월~현재 서울대학교 컴퓨터 공학부 석사과정. 관심분야는 텍스트마이닝, 정보추출, 기계학습



이 은 석

1990년 3월~1998년 2월 고려대학교 영어영문학과(학사). 2001년 3월~2003년 2월 서울대학교 대학원 협동과정 인지과학(석사). 2003년 3월~2006년 2월 서울대학교 협동과정 인지과학(박사과정 수료). 관심분야는 언어 인지 처리, 시지각 처리, 인지 처리 모델링



장 병 탁

1986년 2월 서울대학교 컴퓨터공학 학사
1988년 2월 서울대학교 컴퓨터공학 석사
1992년 7월 독일 Bonn 대학교 컴퓨터과학 박사. 1992년 8월~1995년 8월 독일 국립정보기술연구소(GMD) 연구원. 1995년 9월~1997년 2월 건국대학교 컴퓨터공학과 조교수. 1997년 3월~현재 서울대학교 컴퓨터공학부 교수, 생물정보학, 뇌과학, 인지과학 협동과정 겸임 교수
2001년 1월~현재 바이오정보기술연구센터(CBIT) 센터장
2002년 6월~현재 과학기술부 바이오지능 국가지정연구실 실장. 2003년 8월~2004년 8월 MIT Computer Science and Artificial Intelligence Laboratory(CSAIL) 방문교수
2005년 12월~2006년 2월 Bernstein Center Berlin 과학재단 방문교수. 관심분야는 Biointelligence, Probabilistic Models of Learning and Evolution, Molecular/DNA Computation



김 준 식

1991년 3월~1996년 2월 서울대학교 물리학과(학사). 1996년 3월~1998년 2월 서울대학교 물리학과(석사). 1998년 3월~2008년 2월 서울대학교 물리천문학부(박사). 현재 서울대학교 병원 신경정신과 연구원. 2001년 7월~2008년 1월 서울대학교 컴퓨터 공학과 바이오지능(bio intelligence) 연구실 방문 연구원. 관심분야는 문자 컴퓨팅, 뉴럴 네트워크, fMRI 분석 등