

그래프 신경망을 이용한 최적 VNF 관리 시스템

김희곤, 유재형, 홍원기
포항공과대학교 컴퓨터공학과,

{sinjint, jihyoo78, jwkhong}@postech.ac.kr

The VNF Management System using Graph Neural Network

Hee Gon Kim, Jae Hyung Yoo, James Won Ki Hong

Department of Computer Science and Engineering, POSTECH

요약

가상 네트워크 환경은 유연한 네트워크 관리를 가능하게 하여 네트워크의 확장성을 증가시키고 유연한 변화가 가능하게 하였다. 본 논문은 가상 네트워크 환경에서 그래프 신경망 모델을 이용하여 최적의 Virtual Network Function (VNF) 관리 정책을 생성하는 모델을 제안하였다. 제안한 모델을 통하여 네트워크 관리자는 네트워크의 운영비용을 최소화하면서 사용자가 요구하는 네트워크 서비스를 만족하도록 VNF 를 적절하게 배치할 수 있다. 본 논문은 다양한 네트워크 환경에서 실험을 진행하였으며, 개발한 모델은 VNF 의 종류별로 최적의 설치 위치 및 필요 인스턴스 개수를 생성한다.

I. 서론

Software-Defined Networking (SDN) 과 Network Function Virtualization (NFV) 는 OPEX 와 CAPEX 를 감소시키며 네트워크를 동적이고 유연하게 만들어 준다. SDN 컨트롤러와 NFV 관리 시스템은 네트워크와 NFV 의 전반적인 뷰를 제공하며 네트워크의 Orchestration 을 가능하게 한다. 하지만 SDN/NFV 는 어디까지나 관리 기능을 제공할 뿐, 자체적인 네트워크 최적화 모델을 제공하고 있지는 않다.

기존의 네트워크 관리는 네트워크 관리자의 오랜 경험에 맡기거나 Integer Linear Programming 혹은 휴리스틱 알고리즘을 이용하여 최적화를 시도하였다. 하지만 현재 네트워크 관리 모델은 새로운 국면을 맞이하고 있다. SDN/NFV 의 도입으로 네트워크는 보다 유연한 관리가 가능해지게 되었으며, 동시에 거대하고 다양해진 차세대 네트워크의 등장으로 네트워크 관리의 복잡도는 기하급수적으로 증가되었다. 또한 5G 네트워크 등장은 네트워크 관리가 실시간으로 이루어져 네트워크의 변화에 대응하기를 요구하고 있으며 이러한 변화에 대해 기존의 방법들은 요구조건을 맞추고 있지 못하다. 네트워크 관리자가 현재의 복잡한 네트워크의 모든 요소를 고려하여 경험적으로 관리를 하는 것은 사실상 불가능에 가까우며, Integer Linear Programming 은 계산 시간에 따른 비용으로 네트워크 변화에 대응하기 어려우며 휴리스틱 방법은 관리 결과에 대한 무수한 경험 없이 성능을 보장할 수 없으며 동적으로 변화하는 네트워크에 대응하는 능력이 떨어진다[1].

최근 네트워크 관리에 기계 학습(Machine Learning)을 적용하려는 시도가 많이 이루어지고 있다[2]. 기계 학습을 이용하여 효율적인 네트워크의 자원 사용 및 서비스 품질 향상을 바라보고 있으며 자동화된 네트워크

관리의 구축 등 차세대 네트워크의 등장을 기대하고 있다. 하지만 여태까지 생명, 화학 등 많은 분야에서 기계 학습을 사용한 것과 달리 네트워크 분야에서 기계 학습의 적용은 볼모지와 다름이 없었는데, 이는 네트워크 데이터와 네트워크 관리의 특이성 때문이라고 말할 수 있다. 복잡하고 다양한 네트워크 데이터의 학습은 기계 학습의 수행에 있어 하나의 큰 장애물로 존재하며 기계 학습이 수행해야할 정교한 네트워크 관리 기능도 또다른 장애물로 존재하였다. 네트워크 학습에 대한 어려움으로 현재 기계 학습을 이용한 네트워크 관리 연구는 비교적 간단한 네트워크 데이터를 사용하여 간단한 네트워크 관리 기능을 제공하고 있다[3].

본 연구는 기존의 연구에서 네트워크 데이터의 학습 및 관리 기능의 수행에 어려움을 겪는 이유를 학습 데이터를 이해할 수 있도록 알맞게 설계된 기계 학습 모델의 부재로 보고 새로운 학습 모델을 제안하였다. 본 논문에서는 네트워크 데이터 별로 Graph Neural Network(그래프 신경망 모델) 와 Feed Forward Neural Network 를 사용한 모델을 제안하였다. 제안한 모델은 그래프 신경망 모델을 이용하여 네트워크 토폴로지 구조를 학습하고, Feed Forward Neural Network 로 유저의 서비스를 학습하여 동적이고 변화하는 네트워크 서비스에 대해 고려를 하여 최적의 VNF 의 배치 정책을 생성하였다.

II. 본론

그림 1 은 학습 모델은 개념도이다. 제안한 모델은 학습 단계 (Training Step)와 평가 단계(Test Step)로 최적의 VNF 배치에 대해 학습하고 평가를 받는다. 모델이 학습을 하기 위해서는 학습에 사용되는 데이터 셋이 필요하다. 데이터 셋은 특성 데이터(Feature Data)와 라벨링 데이터 (Labeling Data)로 구성이

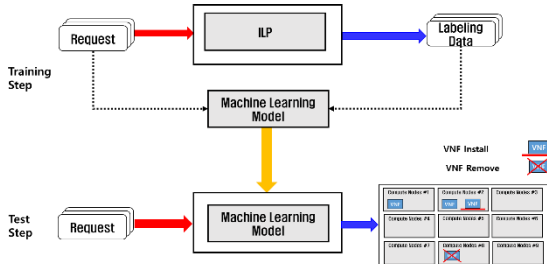


그림 1 학습 모델 개념도

되며 특성 데이터로 네트워크 토폴로지 및 자원 정보와 사용자의 서비스 요구 데이터를 사용하였다. 라벨링 데이터는 VNF의 배치 정책이 되며 해당 데이터는 특수한 ILP[4]식을 이용하여 생성하였다. 자세한 특성 데이터와 라벨링 데이터의 내용은 아래와 같다.

● 특성 데이터

▶ 서비스 정보

- 트래픽의 출발지, 도착지 노드
- 트래픽의 요구 대역폭
- 트래픽의 최대 허용 지연 시간
- 서비스 타입

▶ 네트워크 정보

- 토폴로지 구조
- 노드의 CPU Core 의 수
- 노드의 최대 허용 대역폭
- 링크의 최대 허용 대역폭
- 링크의 지연시간
- 노드의 Idle, Peak 전력소모

● 라벨링 데이터

- 최적의 VNF 배치 위치 및 인스턴스 수

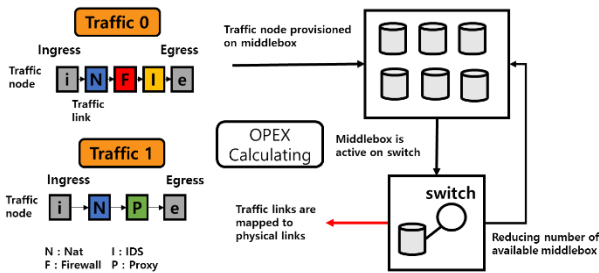


그림 2 Integer Linear Programming (ILP) 모델 개념도

그림 2 는 본 논문에서 사용한 ILP 모델의 개념도를 나타낸 것이다. ILP 는 서비스 정보와 네트워크 정보를 입력으로 받아 최적의 VNF 배치 위치 및 인스턴스 수를 생성한다. 그림 2 에서는 두개의 트래픽(Traffic1, Traffic 2)의 서비스 타입을 일련의 VNF 로 정의하였다. 서비스를 구성하는 VNF 들은 Traffic node 로 정의되며 OPEX 를 최소화하도록 특정 Switch 에 설치된다. 이때, 초기에는 Traffic Node 의 수에 대해 특별한 제한을 두지 않지만 계속해서 이진 탐색으로 가장 적은 수의 Traffic Node 가 Switch 에 설치될 수 있도록 정답을 찾는다. ILP 가 계산을 모든 계산을 완료하게 되면 가장 적은 수의 VNF 의 인스턴스를 가장 OPEX 가 최소화되도록 특정 Switch 에 설치하게 된다. 또한 이때, ILP 는 하나의 서비스에 대해서만 최적화를 하는 것이 아닌, 모든 서비스에 대한 최적의 정답을 생성한다.

ILP 식이 고려하는 네트워크 OPEX 는 VNF 설치 비용, 서버의 에너지 사용 비용(Idle & Peak), 패킷 전송 비용, SLA 위반 비용, 자원 단편화 (Resource

Fragmentation Cost)로 총 5 가지의 비용을 최소화한다. ILP 의 계산 시간은 네트워크의 복잡도에 따라 계산 시간이 비례하게 증가하며, 동시에 고려하는 서비스의 수에 따라 지수적으로 시간이 증가한다. ILP 의 계산 시간은 네트워크 관리에 있어 단점이 되지만 현재 제한하는 모델은 ILP 가 계산한 데이터를 오프라인 데이터로 학습하여 사용하기 때문에 해당 문제가 해결되었다.

특성 데이터에 대한 자세한 시나리오는 다음과 같이 구성하였다. 우리는 다양한 환경에서의 실험을 위해 총 2 개의 서비스 데이터 셋 A, 서비스 데이터 셋 B 를 구성하였다. A 데이터 셋의 서비스 데이터는 33Mbps 에서 38Mbps 의 대역폭을 가지며 B 데이터 셋의 서비스 데이터는 330Mbps 에서 380Mbps 의 대역폭을 가진다. 두 데이터 셋은 대역폭을 제외하고는 모든 설정을 공유한다. 네트워크가 한번에 고려해야하는 서비스의 수를 서비스 복잡도 $n(\pi)$ 라고 했을 때, π 의 비율은 $1:2:3:4=0.14:0.33:0.36:0.17$ 이다. 최대 허용 지연시간은 700ms 에서 750ms 사이의 값을 무작위 하게 배정받으며 서비스 타입과 타입의 비율은 표 1 와 같이 정의되었다.

표 1 서비스 카탈로그

서비스 ID	서비스 타입	비율
1	NAT-Firewall-IDS	0.3
2	NAT-Proxy	0.4
3	NAT-WANO	0.3

본 논문에서는 총 5 개의 VNF 를 고려하였으며 각 VNF 는 표 2 와 같이 정의되었다. 우리는 모든 노드가 동일한 16 개의 CPU Core 를 가진다고 가정하였으며 가용 CPU Core 수를 다르게 정의하는 것으로 VNF 의 설치에 제약을 주었다. 서버의 Idle 에너지는 80.5W 이며 Peak Energy 는 2735W 로 정의하였다.

표 2 VNF 카탈로그 [5]

VNF Type	Required CPU core	Processing Capacity	Processing Delay
Firewall	2	900 Mbps	45ms
Proxy	2	900 Mbps	40ms
IDS	4	600 Mbps	1ms
NAT	1	900 Mbps	10ms
WANO	2	400 Mbps	5ms

우리는 다양한 네트워크 토폴로지를 고려하기 위해서 총 4 개의 네트워크 토폴로지를 학습하였다. 각각의 토폴로지는 A, B, C, D 의 이름을 부여 받았으며 그림 3 에 제시되었다. 토폴로지 A 는 실험에서 가장 기본이 되는 토폴로지인 Internet2 토폴로지의 구성을 가져온 것이다. 해당 토폴로지는 15 개의 노드와 15 개의 링크로 구성이 되었다. 토폴로지 B 는 토폴로지 A 를 압축한 구조로 총 8 개의 노드와 10 개의 링크로 구성이 되어있다. 토폴로지 C 는 데이터 센터 네트워크 토폴로지인 토폴로지 A 와의 비교를 위해 동일한 15 개의 노드로 구성이 되었으며 링크의 수는 2 개 더 많은 17 개로 구성되었다. 토폴로지 D 는 Mesh 토폴로지인 네트워크 복잡도에 따른 모델의 성능을 평가하기 위해 토폴로지 B 와 같이 8 개의 노드로 구성이 되었으며 링크의 수는 17 개이다.

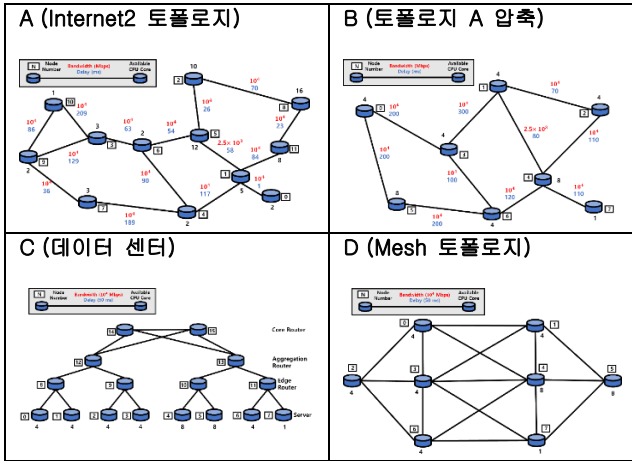


그림 2 토폴로지 A, B, C, D

본 논문에서는 GNN 모델 중 하나인 Edge-conditioned Filters in Convolutional Neural Network(ECGNN)[6]을 사용하여 네트워크 데이터를 학습하였다. 네트워크 데이터를 기계 학습 모델에 학습시킬 때, 가장 고려해야 할 부분은 네트워크 구조에 대한 모델의 이해도 향상이다. FNN 과 같은 보통의 모델로 그래프 데이터를 학습 시킬 경우, 노드와 링크의 특성 데이터는 학습시킬 수 있지만 노드와 노드 간의 연결 관계는 학습시키기 어렵다. 이는 노드와 노드 간의 연결 관계를 데이터로 표현하기 어렵기때문인데, 기계 학습이 이러한 데이터를 이해하기 위해서는 그래프를 Adjacency 매트릭스로 표현을 한 뒤, 매트릭스를 1 차원으로 펼쳐 준비를 해야한다. 이렇게 표현된 데이터는 굉장히 Sparse 한 구조를 가져 메모리적으로 비효율적이며 학습적인 측면에서도 제대로 된 학습결과를 얻기 어렵다. 하지만 ECGNN 과 같이 그래프 신경망 네트워크를 사용할 경우 노드의 특성 데이터에 전체 그래프 구조가 이미 학습되어 표현되어 있기 때문에, 노드와 링크의 특성데이터만으로 기계 학습 모델이 온전히 네트워크 구조를 이해할 수 있게 된다. 즉, GNN 을 이용하여 그래프 데이터를 기계 학습이 쉽게 이해할 수 있는 데이터로 표현하는 과정을 한번 거치고 추가로 네트워크를 최적화하는 학습 모델을 구성하는 것으로 그래프에 대한 모델의 이해도 향상은 물론 최적의 네트워크 관리 모델을 기계 학습을 통해 구현할 수 있게 되었다.

그림 4는 논문에서 제안한 모델로 GNN과 FNN 을 이용하여 네트워크 정보 및 서비스 정보를 학습하고 최적의 VNF 배치 결과를 생성하는 모습을 표현하고 있다. 서비스 데이터와 네트워크 데이터는 페어로 이루어져 있으며 유저의 요구 서비스가 변경되는 시점마다 서비스 데이터와 네트워크 데이터가 생성된다. 서비스 데이터는 유저가 요구하는 여러 서비스 정보가 존재하며 각 서비스는 FNN 모델을 통해 학습하게 되며, FNN 은 학습 파라미터인 Weight 을 공유한다. 네트워크 데이터는 노드, 링크, 토폴로지 구조로 분리된 다음 그래프 형태의 데이터로 변환이 되어 GNN 에 모델에 입력된다. 이후 FNN 과 GNN 모델은 합쳐지고 2 개의 추가 FNN 모델에서 학습을 진행하고 Fully Connected Layer (FCN)을 지나 결과값을 생성한다. 그림 5 는 모델이 생성한 라벨링 데이터를 표현한 것이다. 모델은 서비스 데이터 및 네트워크 데이터를 입력으로 받아

모든 서버들이 특정한 타입의 VNF 를 몇 개 설치해야 하는 지에 대한 정보를 제공한다.

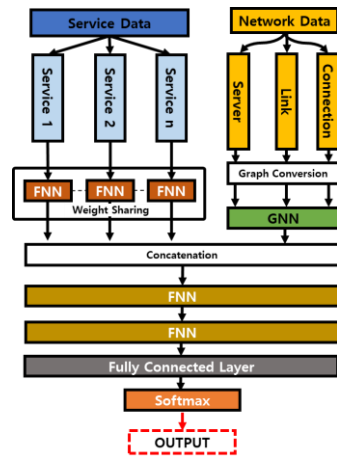


그림 4 GNN-FNN VNF Deployment 모델

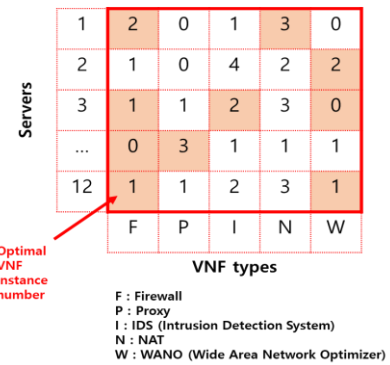


그림 5 라벨링 데이터

III. 실험

본 논문에서는 네 가지의 서로 다른 실험을 진행하였다. 첫번째 실험은 학습 데이터 수 및 서비스 복잡도에 따른 모델의 학습 성능 평가이다. 우리는 모델을 구현하는 도중 학습에 사용하는 데이터의 수 및 네트워크에 요구되는 서비스의 수에 따라서 모델의 성능이 급격하게 변화하는 것을 확인하였다. 표 3 은 토폴로지 A 와 서비스 데이터 셋 A 를 사용한 실험에 대한 결과이다.

표 3 학습 데이터 수 및 서비스 복잡도에 따른 성능 평가

서비스 복잡도	학습 데이터 수			
	1 주	2 주	5 주	10 주
$\Pi=1$	100%			
$\Pi \leq 2$	95.0%	98.3%	98.8%	100%
$\Pi \leq 3$	90.9%	91.2%	91.8%	92.3%
$\Pi \leq 4$	88.7%	88.6%	88.6%	88.6%

학습 데이터 수가 증가할수록 모델의 성능이 증가하였으며 서비스 복잡도가 증가할수록 학습 데이터의 성능이 감소하였다. 학습 데이터의 수와 모델의 성능은 자명한 결과이므로 서술은 넘어가겠다. 서비스 복잡도와 모델의 성능에 대한 관계는 다음과 같다. 서비스 복잡도가 증가할수록 모델은 더 많은 데이터를 입력 받기 때문에 모델의 복잡도는 증가하게 된다. 특히 입력 데이터의 선형적인 증가는 모델의 복잡도는 지수만큼 증가시키게 되는데, 이러한 결과로 1 주에서 10 주로 선형적인 학습 데이터 수의 증가는 서비스

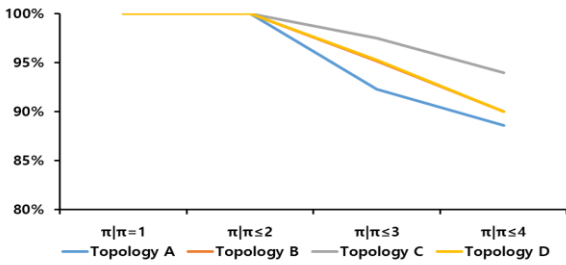


그림 6 토폴로지와 서비스 복잡도에 따른 모델의 성능

복잡도가 2 이하인 경우에는 5% 증가라는 유의미한 결과를 만들었지만 서비스 복잡도가 4 이하인 경우에는 유의미한 결과를 만들지 못하였다.

두 번째 실험은 토폴로지와 서비스 복잡도에 따른 모델의 성능 비교 실험이다. 우리는 서비스 데이터 셋 A 를 이용하여 앞서 정의한 4 개의 네트워크 토폴로지에서 실험하였다. 실험 결과는 그림 6 으로 표현되며 우리는 서비스 복잡도가 높을 경우, 네트워크 토폴로지 구조에 따라 모델의 성능이 다를 수 있음을 파악하였다. 서비스 복잡도가 1 일때, 모델은 토폴로지와 관계 없이 100%의 성능을 보였지만 서비스 복잡도가 증가할수록 토폴로지의 성능 차이가 발생하였다. 서비스 복잡도가 4 이하일 때, 토폴로지 A 네트워크가 성능이 가장 많이 감소하였고 B 와 D 가 같은 성능의 감소를 보였으며 C 가 성능 저하가 가장 적었다. 흥미롭게도 토폴로지 C 는 데이터 센터 네트워크 구조를 가지고 있으며 A 는 Mesh 네트워크 구조를 나타내고 있다.

세 번째 실험은 네트워크의 복잡도에 따른 노드(서버) 별 VNF 배치 정확도 실험이다. 우리는 토폴로지의 각 노드가 연결하고 있는 이웃 노드의 수와 VNF의 배치 정확도의 관계성을 파악하고자 노력하였다. 노드 별 VNF 배치 정확도는 표 4 와 같이 나타났다. 실험 결과를 보면 흥미롭게도 연결된 이웃 노드의 수가 많은 노드 일수록 VNF 배치 정확도가 떨어지는 것으로 나타났다. 이는 실험 2 에서 Mesh 네트워크 구조 토폴로지 A 가 데이터센터 네트워크 토폴로지 C 보다 성능이 낮은 것과 일맥상통하는 부분이 있다. 노드들의 연결이 더 많아질수록 네트워크는 Mesh 네트워크 형태가 되며 노드들의 VNF 배치 정확도는 감소하여 전체 네트워크에 대한 모델의 성능이 저하되었다.

표 4 토폴로지의 노드 별 VNF 배치 정확도 (이웃 노드의 수)

노드 번호	토폴로지			
	A	B	C	D
0	91% (1)	93% (1)	98% (2)	91% (4)
1	83% (4)	94% (1)	97% (3)	95% (4)
2	91% (2)	96% (1)	98% (2)	97% (3)
3	90% (3)	95% (1)	95% (3)	94% (6)
4	89% (3)	95% (1)	98% (4)	95% (6)
5	83% (3)	95% (1)	97% (2)	97% (3)
6	89% (3)	96% (1)	96% (3)	93% (4)
7	94% (2)	99% (1)	100% (1)	99% (1)
8	95% (2)			
9	95% (3)			
10	99% (2)			
11	90% (2)			

IV. 결론

본 논문은 기계 학습을 이용한 네트워크 관리 모델을 제시하고 있다. 해당 모델은 GNN 과 FNN 을 사용하여

네트워크 데이터와 서비스 데이터를 학습하고 최적의 위치에 적절한 타입의 VNF 를 알맞은 수만큼 설치하여, VNF 설치 비용, 서버의 에너지 사용 비용, 패킷 전송 비용, SLA 위반 비용, 자원 단편화 등의 네트워크 관리 비용을 감소하였다. 본 논문에서는 동적으로 변화하는 네트워크 환경을 고려하였으며 4 개의 서로 다른 네트워크 토폴로지를 사용하여 다양한 환경에서 실험을 진행하고 모델의 성능을 평가하였다. 향후 연구로 동시에 고려할 수 있는 서비스의 수를 증가시켜 실제 네트워크 환경에서 바로 적용이 가능한 모델을 만드는 것을 계획하고 있다.

ACKNOWLEDGMENT

본 연구는 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2018-0-00749, 인공지능 기반 가상 네트워크 관리기술 개발, IITP-2021-2017-0-01633* 대학 ICT 연구센터지원사업)

참 고 문 헌

- [1] Lange, Stanislav, et al. "Predicting VNF Deployment Decisions under Dynamically Changing Network Conditions." *2019 15th International Conference on Network and Service Management (CNSM)*. IEEE, 2019.
- [2] Cheng, Yang, et al. "Bridging machine learning and computer network research: a survey." *CCF Transactions on Networking* 1.1 (2019): 1-15.
- [3] Boutaba, Raouf, et al. "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities." *Journal of Internet Services and Applications* 9.1 (2018): 1-99.
- [4] Bari, Md Faizul, et al. "On orchestrating virtual network functions." *2015 11th International Conference on Network and Service Management (CNSM)*. IEEE, 2015.
- [5] Lange, Stanislav, et al. "Predicting vnf deployment decisions under dynamically changing network conditions." *2019 15th International Conference on Network and Service Management (CNSM)*. IEEE, 2019.
- [6] Gori, Marco, Gabriele Monfardini, and Franco Scarselli. "A new model for learning in graph domains." *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005..* Vol. 2. IEEE, 2005.