JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 22, No. 5: 728~733, May 2018

단백질 이차 구조 예측을 위한 합성곱 신경망의 구조

지상문*

Architectures of Convolutional Neural Networks for the Prediction of Protein Secondary Structures

Sang-Mun Chi*

Department of Computer Science, Kyungsung University, Pusan, 48434, Korea

요 약

단백질을 구성하는 아미노산의 서열 정보만으로 단백질 이차 구조를 예측하기 위하여 심층 학습이 활발히 연구되고 있다. 본 논문에서는 단백질 이차 구조를 예측하기 위하여 다양한 구조의 합성곱 신경망의 성능을 비교하였다. 단백질 이차 구조의 예측에 적합한 신경망의 층의 깊이를 알아내기 위하여 층의 개수에 따른 성능을 조사하였다. 또한 이미지 분류 분야의 많은 방법들이 기반 하는 GoogLeNet과 ResNet의 구조를 적용하였는데, 이러한 방법은 입력 자료에서 다양한 특성을 추출하거나, 깊은 층을 사용하여도 학습과정에서 그래디언트 전달을 원활하게 한다. 합성곱 신경망의 여러 구조를 단백질 자료의 특성에 적합하게 변경하여 성능을 향상시켰다.

ABSTRACT

Deep learning has been actively studied for predicting protein secondary structure based only on the sequence information of the amino acids constituting the protein. In this paper, we compared the performances of the convolutional neural networks of various structures to predict the protein secondary structure. To investigate the optimal depth of the layer of neural network for the prediction of protein secondary structure, the performance according to the number of layers was investigated. We also applied the structure of GoogLeNet and ResNet which constitute building blocks of many image classification methods. These methods extract various features from input data, and smooth the gradient transmission in the learning process even using the deep layer. These architectures of convolutional neural networks were modified to suit the characteristics of protein data to improve performance.

키워드 : 단백질 이차 구조, 심층 학습, 합성곱 신경망, GoogLeNet, ResNet.

Keyword : Protein secondary structure, Deep learning, Convolutional neural networks, GoogLeNet, ResNet.

Received 28 January 2018, **Revised** 8 February 2018, **Accepted** 16 April 2018

*** Corresponding Author** Sang-Mun Chi (E-mail:smchiks@ks.ac.kr,Tel:+82-51-663-5146)
Department of Computer Science, Kyungsung University, Pusan 48434, Korea

Open Access http://dx.doi.org/10.6109/jkiice.2018.22.5.728 pISSN:2234-4772

Ⅰ. 서 론

생명현상에 필요한 대부분의 생화학 반응이 단백질에 의해 수행하므로, 그 기능과 구조에 대한 많은 연구가 수행되어 왔다. 단백질은 아미노산들이 일차원으로 연결된 구조로서, 아미노산 서열이 단백질의 구조와 기능을 결정한다[1]. 아미노산의 서열정보만으로 단백질의 구조와 기능을 알아내는 연구가 활발한데, 단백질 이차 구조 예측은 이러한 연구의 중요한 요소 기술이다. 단백질 일차 구조는 단백질을 구성하는 아미노산 서열을 의미하고, 이차 구조는 이들 아미노산의 카보닐기 산소 원자와 아민기 수소 원자 사이에 수소 결합이 형성되어 나타나는 국소적인 규칙적 구조를 말한다[2]. 이차 구조 예측에 사용된 방법은 용매접근도, 국소적 골격 각도 등의 단백질 구조를 예측하는 정보로 활용된다.

단백질 이차 구조를 예측하기 위하여 심층 신경망(deep neural networks)을 이용하는 많은 연구가 있다[3-7]. 본 논문에서는 가장 높은 성능을 보이는 논문 [7]에서 사용한 합성곱 신경망의 여러 구조를 적용하고, 합성곱 신경망이 커다란 성능향상을 거둔 이미지 분류 분야의 기술을 단백질 이차구조의 예측에 적용하고자 한다.

최근에는 많은 수의 층을 사용하는 심층 신경망의 구조가 높은 성능을 보이고 있다. 이미지 인식 대회인 ILSVRC (Large Scale Visual Recognition Challenge [8])에서 좋은 성능을 보이는 최근의의 방법들은 수십-수백 개의 층을 사용하고 있다. 2014년 ILSVRC에서 1000개 클래스를 분류하는 분야에서 1위를 기록한 GoogLeNet[9]은 인셉션 구조의 층으로 구성되어 있는데, 인셉션 구조는 3개의 합성곱과 1개의 풀링을 결합하여 층을 구성하는데, 이것이 합성곱 하나만으로 층을 구성하는 기존의 방법과 차이점이다. 2015년 ILSVRC의 1위인 ResNet[10]은 층을 깊게 하였을 경우에도 학습을 효율적으로 할 수 있는 스킵 연결을 사용하는데, 스킵 연결은 입력 자료를 합성곱 계층을 건너뛰어 합성곱 계층의 출력과 더하는 구조이다. 최근의 이미지 분류에서 높은 성능을 보이는 대부분의 방법들도 GoogLeNet과 ResNet 구조에 기반한 방법들이다. 본 논문에서는 GoogLeNet과 ResNet 구조를 단백질 이차 구조의 예측에 적용하여 효과적인지를 조사한다. 또한, 많은 수와 넓이를 가진 층을 사용하는 심층 신경망은 학습이 어렵고, 과적응이 발생할 수 있으므로, 이를 해결하기 위하여 조기 중단과 드롭아웃을 적용한다.

Ⅱ. 합성곱 신경망의 다양한 구조

본 논문에서는 이미지 분류에 널리 사용되는 합성곱 신경망(Convolutional Neural Network, CNN)을 단백질 이차구조 예측에 사용한다. 여러 신경망의 근간을 이루는 다층 신경망은 여러 층으로 구성되어져 복잡한 자료를 모델링하기에 적합하다. *l*-번째 층의 *i*-번 유닛 o_i^l은 (*l*-*1*)-번째 층의 유닛들로부터 얻는다. 처음 층의 o_i^0는 입력 자료이다.

$$o_i^l = f(\sum_j o_j^{l-1} w_{j,i}^l + w_{0,i}^l) \quad (l = 1,2,...,L-1) \quad (1)$$

$w_{j,i}^l$은 (*l*-*1*)-번째 층의 *j*-번 유닛과 *l*-번 계층의 *i*-번 유닛을 연결하는 가중치, $w_{0,i}^l$는 *l*-번째 층의 *i*-번 유닛에 더해지는 편향 값이다. 변환함수 f로 자주 사용되는 함수는 $\mathrm{relu}(x) = \max(0, x)$, $\sigma(x) = 1/(1+\exp(-x))$, $\tanh(x) = (1-\exp(-2x))/(1+\exp(-2x))$이다. 분류를 수행하는 문제에서는 마지막 계층의 f를 없애고 얻은 마지막 층의 출력에 식 (2)의 소프트맥스(softmax) 함수를 적용하여 각 부류의 사후 확률 y_i을 얻는다.

$$y_i = \exp(o_i^L) / \sum_j \exp(o_j^L) \quad (2)$$

합성곱 신경망은 국부적 연결을 나타내는 파라미터를 이용하여 인접한 자료간의 상관관계를 모델링하기에 적합하며, 식 (1)의 완전연결 층에서는 무시되는 위치에 따른 자료의 형상을 유지한다. 본 논문의 입력 자료는 기존연구 [3-7]과 같은 단백질 프로파일 $p_{i,j}$로서, 첨자 i는 서열의 위치이고 j는 20차원으로 각 아미노산의 치환 빈도이다. 따라서 단백질 프로파일은 이미지와 같은 이차원 자료로서 위치별 특징이 반영되는 합성곱 신경망의 적용이 가능한다. 합성곱 층에서는 입출력 자료를 특징맵(feature map)이라 한다. $h_{i,j}^{m,k}$를 m층의 k번째 특징맵이라 하고, $m-1$층의 여러 특징맵을 이용하여 얻는다.

$$h_{i,j}^{m,k} = f(\sum_{n=1}^{F_{m-1}} \sum_{u=-T/2}^{T/2} \sum_{v=-L/2}^{L/2} h_{u,v}^{m-1,n} w_{i-u,j-v}^{k,n} + w_m^k) \quad (3)$$

단, F_{m-1}는 $m-1$ 층의 특징맵들의 개수, T와 L은 계산에 사용되는 국부영역의 크기로서 학습할 필터 w의 크기를 결정한다. 단백질 프로파일은 합성곱 신경망의 입력으로 $h_{i,j}^{0,0}$의 역할을 하며, 입력은 하나의 특징맵으로 구성된다. 풀링(pooling) 층에서는 아래층 특징맵의 평균값 또는 최대값을 구하는데, 이미지 분류에서 위치의 변화에 무관한 특징을 추출하는데 유용하다. 마지막 계층은 출력을 일차원으로 변환하여 식 (1)의 신경망의 입력으로 사용하여 여러 개의 다층 신경망을 구성하고, 식(2)로 최종 부류별 확률을 구한다.

기존 심층망에서는 식 (3)의 합성곱을 하나만 사용하여 층을 구성하는데, GoogLeNet[9]은 여러 개의 합성곱을 결합하여 층을 구성한다. GoogLeNet이 사용하는 인셉션 구조는 3개의 합성곱(1×1, 3×3, 5×5 필터 크기)과 1개의 3×3 크기의 최대값 풀링을 결합하여 만든 층이다. 인셉션 구조를 이루는 3×3과 5×5 합성곱의 입력은 차원 축소를 위하여 1×1 합성곱을 앞단에 먼저 적용하여 출력을 얻고, 이 출력에 합성곱을 적용한다. 3×3 최대값 풀링은 풀링을 하여 얻은 출력에 다시 1×1 합성곱을 적용한다.

ResNet[10,11]은 많은 층으로 구성된 신경망이 원활히 학습이 되고 정확도가 향상되도록 다음의 스킵 구조를 사용한다.

$$y_l = h(x_l) + F(x_l, W_l), \quad x_{l+1} = f(y_l), \tag{4}$$

단, x_l, x_{l+1}은 l 층의 입력과 출력이고, $h(x)=x$인 항등함수 이고, f는 논문 [10]에서는 $relu$이었으나, 논문 [11]에서는 전방향과 역방향으로 모두 신호가 보다 직접적으로 전파할 수 있도록 항등함수를 사용하였다. 잔차함수 F는 여러 처리로 구성되어 있는데, ResNet [11]에서는 입력 값 정규화, $relu$, 가중치 행렬, 입력값 정규화, $relu$, 가중치 행렬로 구성되어 있다. 본 논문에서도 이를 기반으로 다양한 구조를 실험하였다.

Ⅲ. 실험 및 결과

3.1. 실험 자료

실험 자료는 CullPDB server[12]로부터 30%이하의 서열 동일성을 갖고 단백질 구조 해상도가 2.5 옹스트롬보다 높은 단백질 자료를 추출하였다. 논문 [5,7]과 같은 방법으로 길이가 50이상이고 700이하인 단백질 체인 9091개를 선택하였다.

예측하고자 하는 단백질의 이차구조는 기존의 여러 연구에 따라 DSSP[13] 프로그램을 사용하여 이차구조 8개 (G: 3-helix, H: alpha helix, I: 5-helix, B: residue in isolated beta-bridge, E: extended strand, participates in beta ladder, T: hydrogen-bonded turn, S: bend, ".": otherwise)로 분류하였고, 이 결과를 세 가지 (G, H, I -> H; B, E -> E; T, S, "." ->C)의 대분류인 나선형의 구조 (H), 병풍형태의 평평한 구조 (E)와 이들 이외의 불규칙한 형태인 코일(C)로 변환하여 이차 구조를 정의하였다.

본 논문에서는 단백질 서열을 변환한 단백질 프로파일을 이차구조 예측을 위한 입력으로 사용하였다. 단백질 프로파일은 PSI-BLAST[14]를 사용하여 변환하려는 단백질 서열과 유사한 서열들을 단백질 데이터베이스에서 탐색하여 구성한 PSSM (position specific scoring matrix)이다. 이는 서열의 길이가 N일 때, N개의 위치에서 20개의 아미노산이 관측될 확률을 나타내는 $N\times20$ 차원의 행렬이다. PSI-BLAST를 사용한 탐색에는 단백질 서열 자료 UniRef50[15]를 이용하였고, 문턱치 E값은 1로 설정하였다.

3.2. 심층 신경망의 학습

심층 신경망은 다수의 층을 구성하고 있는 많은 수의 파라미터를 가지므로, 학습 자료에 나타난 특성에 지나치게 적응할 수 있다. 본 논문에서는 이러한 과적응을 방지하기 위하여 조기중단(early stopping)과 드롭아웃(DropOut)을 사용하였다. 조기중단은 학습 과정에서 학습 자료와 겹치지 않는 검증 자료를 예측하고 검증자료의 정확도가 현재의 파라미터 갱신 회수의 2배내에서 향상되지 않으면 학습을 중단하고, 가장 좋은 정확도를 보였던 모델을 최종 신경망으로 선택한다. 본 논문에서는 CullPDB server[12]에서 얻은 9091개의 단백질 서열을 균등한 개수로 5개로 나누고, 1개는 평가 자료로, 1개는 검증 자료로, 나머지 3개는 학습 자료로 사용하였다. 드롭아웃은 신경망을 규격화(regularization)하는 방법의 일종으로[16], 각 층의 출력은 확률 p를 가지고 그대로 유지되거나, $1-p$의 확률로 0으로 변환하는데, 본 논문에서는 $p=0.5$를 사용하였다.

심층신경망의 구현은 Theano[17]와 Lasagne[18] 패

키지를 사용하였다. 신경망의 파라미터를 갱신하는 학습 방법으로는 Adagrad[19]를 사용하였고, 학습률은 본 논문의 실험에서는 0.01때가 대부분 성능이 좋았으므로 이를 사용하였다.

3.3. 단백질 이차 구조 예측

Table. 1 Accuracy of protein secondary structure prediction for layer depth and filter size (%)

depth \ filter L	3	4	5	6	7	8	9
4	83.77	84.05	84.23	84.18	84.27	84.27	84.32
5	84.21	84.35	84.69	**84.82**	**84.80**	84.78	84.73
6	84.40	84.71	**84.84**	**84.83**	**84.80**	84.67	84.52
7	84.40	84.64	84.77	84.60	84.51	84.69	84.64
8	84.30	84.69	84.67	84.46	84.62	84.63	n.a
9	84.29	84.52	84.65	84.64	84.61	n.a	n.a
10	84.32	84.61	84.64	84.55	n.a	n.a	n.a
11	84.35	84.47	84.46	n.a	n.a	n.a	n.a

표 1에서 심층 신경망의 깊이에 따른 단백질 이차구조의 예측 정확도를 조사하였다. 입력은 예측하려는 이차구조의 아미노산 위치와 이 위치의 좌우로 10개씩으로 구성된 39×20 차원의 단백질 프로파일을 사용하였다. 첫 번째 층의 합성곱은 다양한 필터의 크기를 조사하였는데, 아미노산 한 개의 나타내는 20차원을 하나의 단위로 사용하는 1×20 필터가 성능이 높았고, 특징맵은 500을 사용하였다. 두 번째 층부터는 필터의 크기가 $L\times1$이고, 특징맵은 400을 사용하였다. 합성곱에 패딩은 사용하지 않는 것이 성능이 좋았고, 패딩을 사용하지 않았으므로 층이 깊어질수록 출력값의 길이는 감소한다. 따라서 층의 깊이에 따라 길이가 긴 필터는 적용할 수 없고, 표 1에는 n.a로 나타내었다. 마지막 층들은 앞 단계의 출력의 개수의 $1/10$의 출력 개수를 갖는 식 (1)을 사용한 완전 연결 층과 식 (2)의 분류를 수행하는 층으로 구성된다.

표 1에서 [depth]가 5, 6이고 [filter size]가 5, 6, 7 일 때 최적의 성능을 보였다. 같은 조건에서 특징맵의 개수를 작게 하거나 풀링 계층을 사용하면 성능이 하락하였다. 이는 분류하려는 분야의 특성에 따라 최적인 층의 깊이, 필터의 크기, 특징맵의 개수가 다르기 때문이다[20-23]. 문자인식 방법[20, 21]에서는 풀링층을 제외하면 4개의 층, 4×4와 5×5의 합성곱 필터크기, 수십 개의 특징맵 개수를 가진 것이 좋은 성능을 보이는 반면에, 이미지 인식 방법[9-11]에서는 백개 이상의 많은 층, 주로 3×3 작은 크기의 필터, 수백개의 특징맵을 가진 구조가 성능이 우수하다. 표 1에 보듯이 단백질 이차구조의 예측 문제는 문자 인식과 이미지 인식의 중간적 특성을 갖는다.

Table. 2 Accuracy of protein secondary structure prediction using GoogLeNet Inception structures(%)

inception depth:(number of feature maps)	accuracy
1,(100,50,50,150,150)	83.77
1,(150,50,100,100,100)	83.74
1,(50,100,50,150,150)	83.73
2,(100,100,150,50,100)	83.94
2,(50,150,150,50,100)	83.94
2,(150,100,100,100,50)	83.93
3,(150,100,150,50,50)	83.76
3,(100,50,100,100,150)	83.72
3,(150,50,50,150,100)	83.71

표 2는 GoogLeNet[9]의 인셉션 구조를 사용하여 단백질 이차구조를 예측한 결과이다. GoogLeNet의 구조를 그대로 사용하여 예측하면 정확도가 80% 이하이다. 이는 GoogLeNet은 입력 자료가 224x224 크기이고 이미지이고, 첫 번째 필터의 크기가 7×7이었는데, 단백질 프로파일에 적합하지 않기 때문이다. 따라서 표 1과 같이 첫 번째 필터는 1×20 합성곱을 사용하였고, 인셉션 구조에는 1×1, 3×1, 5×1, 7×1 합성곱 필터와 3×1 최대값 풀링을 사용하였다. 인셉션 구조에서 차원을 줄이기 위한 1×1 합성곱의 특징맵의 개수는 각 필터의 특징맵 개수의 1/2로 하였다. 신경망의 마지막 두층은 표 1과 동일한 것을 사용하였다.

인셉션 구조를 구성하는 각각의 필터의 최적인 특징맵을 조사하기 위하여 각 필터는 특징맵의 개수를 (50,100,150) 중의 하나의 값을 가지면서, 전체적으로 5개 필터의 특징맵의 합이 500이 되도록 고정하여 예측 정확도를 조사하였다. 표 2에는 인셉션 구조의 깊이 1, 2, 3에 대하여 높은 정확도를 보이는 상위 3개의 특징맵의 조합을 나타내었다.

예측 실험을 수행한 결과 인셉션 구조를 구성하는 필

터의 크기에 따라 큰 차이가 없었고, 인셉션구조를 사용한 것이 표 1의 단일한 합성곱 보다 성능이 낮았다. 단백질 이차구조의 예측에는 합성곱 필터의 크기를 달리 하여 결합하는 인셉션 구조의 효과가 크지가 않았다고 판단된다.

Table. 3 Accuracy of protein secondary structure prediction using ResNet residual structures(%)

inc, L, s, nMap	accuracy
0, 5, 5, 400	83.12
0, 5, 6, 400	83.10
0, 3, 6, 600	83.08
1, 5, 2, 400	83.13
1, 3, 2, 400	83.09
1, 5, 3, 600	83.02
2, 3, 2, 400	82.97
2, 5, 2, 200	82.95
2, 3, 3, 400	82.93

표 3에서는 ResNet [11]구조를 사용하여 단백질 이차구조를 예측하는 실험을 수행하였다. 표 2와 마찬가지로 ResNet에서 7×7크기의 첫 번째 필터 대신에 1×20 을 사용하였다. ResNet [10, 11]에서는 식 (4)의 스킵구조를 $s=2,3,4,5,6$회 수행한 후에, 스트라이드 2를 사용한 합성곱 필터를 적용하면서 특징맵의 개수를 2배 증가시키는데, 본 논문에서는 특징맵의 증가회수 $inc=0,1,2$과 시작 특징맵의 수 $nMap=200,400,600$에 대하여 실험하였다. 또한, 식 (4)를 구성하는 합성곱 필터는 $L\times1\,(L=3,5)$크기에 대하여 실험하였다. 각 *inc*에서 상위 3개의 높은 성능을 갖는 s와 L, *nMap*에서의 정확도를 나타내었다.

식 (4)의 스킵구조에 2개의 합성곱이 수행되므로 s회의 스킵구조는 $2\times s$번의 합성곱이 수행된다. 각 *inc*에서 *inc+1*번의 $2\times s$번의 합성곱과 특징맵을 증가시키는 *inc*번의 합성곱이 수행되므로 $inc=1, s=5$인 경우에는 21개의 합성곱 층으로 구성된다. 예표 3에서 보듯이 *inc*가 3인 경우의 많은 층을 사용하여도 성능이 저하되지 않는다. 이는 ResNet구조로 인하여 그래디언트 값이 용이하게 역전파되는 특징을 갖기 때문이다[11]. 하지만 층의 개수를 증가시킴에 따라 전체적인 성능 향상은 발생하지 않았다.

Ⅳ. 결 론

본 논문에서는 아미노산 서열 정보만을 사용하여 단백질 이차 구조를 예측하는데 효과적인 합성곱 신경망의 구조를 조사하였다. 특히, 심층 신경망을 사용하여 높은 정확도를 얻고 있는 이미지 분류 분야의 기술을 단백질 이차구조의 예측에 적용하였다.

이미지 분류 분야에서는 많은 층을 가진 합성곱 신경망을 적용하여 좋은 결과를 얻는데, 단백질 이차 구조의 예측 문제는 문자 인식과 이미지 인식에서 각각 효과적인 층의 개수의 중간부분에서 높은 정확도를 보였다. 또한, 이미지 인식 분야에서 좋은 성능을 보이는 신경망 구조를 적용하였는데, 예측 정확도의 향상이 적었다. 즉, 단백질 프로파일처럼 고정된 구조의 입력은 GoogLeNet처럼 신경망의 층을 다양한 크기의 합성곱으로 구성하여 여러 크기의 특징을 추출하는 방법이 효과가 적었고, ResNet 처럼 심층구조의 신경망을 원활하게 학습시키기 위한 방법은 단백질 이차구조의 예측에는 아주 많은 층이 필요하지 않으므로 효과가 적었다.

향후에는 단백질 이차 구조의 예측에 적합하도록 신경망 구조의 구조를 최적화할 필요성이 있다. 또한 심층 신경망을 구성하는 파라미터를 효과적으로 학습하기 위하여 여러 가지 알고리즘을 적용할 필요성 있다. 본 논문에서는 단일한 구조의 분류기를 사용하였는데, 대부분의 단백질 이차 구조 예측 방법들은 여러 구조의 신경망들의 앙상블을 사용하여 성능을 향상시키므로, 여러 방법을 효과적으로 조합하는 방법을 사용하면 예측 정확도를 더욱 향상시킬 수 있을 것이다.

ACKNOWLEDGEMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2016R1D1A3B03935290).

References

[1] D. Baker and A. Sali., "Protein structure prediction and structural genomics," *Science*, vol. 294 no. 5, pp. 93-96, Oct.

2001.
[2] H. Lodish, *et al.*, *Molecular Cell Biology*, 6th ed. New York, NY: W.H. Freeman and Company, 2007
[3] H. W. Buchan, *et al.*, "Scalable web services for the PSIPRED protein analysis workbench," *Nucleic Acids Research*, vol. 41, W72-W76, Jul. 2013.
[4] C. N. Magnan and P. Baldi, "SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity," *Bioinformatics*, vol. 30, no. 18, pp, 2592- 2597, Sep. 2014.
[5] J. Zhou, and O. Troyanskaya, "Deep supervised convolutional generative stochastic network for protein secondary structure prediction," *Proceedings of Machine Learning Research*, vol. 32, no. 1, pp. 745-753, Jun. 2014.
[6] M. Spencer, J. Eickholt, and J. Cheng, "A deep learning network approach to ab initio protein secondary structure prediction," *IEEE/ACM Transactions on Computational Biology Bioinformatics*, vol. 12, no. 1, pp. 103-112, Jan/Feb. 2015.
[7] S. Wang, *et al.*, "Protein secondary structure prediction using deep convolutional neural fields," *Scientific Reports 6*, Article number: 18962, Jan. 2016.
[8] Olga Russakovsky, *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211-252, 2015.
[9] C. Szegedy, *et al.*, "Going deeper with convolution," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9, Jun. 2015.
[10] K. He, *et al.*, "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, Jun. 2016.
[11] K. He, et al., "Identity mapping in deep residual networks," *European Conference on Computer Vision*, pp. 630-645, Sep. 2016.
[12] G. Wang and R.L. Dunbrack "PISCES: a protein sequence culling server," *Bioinformatics*, vol. 19, no. 12, pp. 1589-1591, 2003.
[13] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: pattern recognition of hydrohen-bonded and geometrical features," *Biopolymers*, vol. 22, no. 12, pp. 2577-2637, Dec. 1983.
[14] S. F. Altschul, *et al.*, "Gapped blast and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389-3402, Sep. 1997.
[15] B. E. Suzek, *et al.*, "Uniref: comprehensive and non-reduncant uniprot reference clusters," *Bioinformatics*, vol. 23, no. 10, pp. 1282-1288, May. 2007.
[16] G. E. Hinton, *et al.*, "Improving neural networks by preventing co-adaptation of feature detectors," [Online]. arXiv:1207.0580, Jul. 2012.
[17] Theano Development Team. "Theano: A Python framework for fast computation of mathematical expressions," [Online]. arXiv:1605.02688, May. 2016.
[18] S.. Dieleman, et al., "Lasagne: First release," [Internet]. Available: http://dx.doi.org/10.5281/zenodo.27878.
[19] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121-2159, Jul. 2011.
[20] W. Li, *et al.*, "Regularization of neural networks using dropconnect," *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, USA, vol. 28, no. 3, pp. 1058-1066, Jun. 2013.
[21] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3642-3649, Washington, DC, USA, Jun. 2012.
[22] Shin-Hye, *et al.*, "A Comparison of Predicting Movie Success between Artificial Neural Network and Decision Tree", *Asia-pacific Journal of Multimedia*, vol.7, no.4, pp. 593-602, 2017.
[23] S. Chi, "A Performance Comparison of Protein Profiles for the Prediction of Protein Secondary Structures," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 22, no. 1, pp. 26-32 Jan. 2018.

지상문(Sang-Mun Chi)

1991년 서울대학교 수학교육학과 졸업(이학사)
1993년 한국과학기술원 수학과 졸업(이학사)
1998년 한국과학기술원 전산학과 졸업(공학박사)
1993년 ~ 2000년 삼성전자 무선사업부 선임연구원
2001년 ~ 현재 경성대학교 소프트웨어학과 교수
※관심분야 : 생물정보학, 기계학습