

논문 2012-49CI-2-15

뉴로 네트워크에서 코사인 모듈화 된 가우스함수의 다항식과 계단함수의 근사

(Approximation of Polynomials and Step function for cosine modulated Gaussian Function in Neural Network Architecture)

이 상 화*

(Sangwha Lee)

요 약

본 논문에서는 CosGauss라고 하는 코사인함수로 모듈화 된 가우시안 활성화함수가 뉴로 네트워크에서 다항식과 계단함수의 근사에 사용될 수 있음을 증명한다. CosGauss 함수는 시그모이드, 하이퍼볼릭 탄젠트, 가우시안 활성화 함수보다 더 많은 범프(bump)를 구성 할 수 있다. 이 함수를 캐스케이드 코릴레이션 뉴로 네트워크 학습에 사용하여 벤치마크 문제인 Tic-Tac-Toe 게임과 아이리스(iris) 식물 문제와 실험하고 여기에서 얻어진 결과를 다른 활성화 함수를 사용한 결과와 비교 분석한다.

Abstract

We present here a new class of activation functions for neural networks, which herein will be called CosGauss function. This function is a cosine-modulated gaussian function. In contrast to the sigmoidal-, hyperbolic tangent- and gaussian activation functions, more ridges can be obtained by the CosGauss function. It will be proven that this function can be used to approximate polynomials and step functions. The CosGauss function was tested with a Cascade-Correlation-Network of the multilayer structure on the Tic-Tac-Toe game and iris plants problems, and results are compared with those obtained with other activation functions.

Keywords : cosine-modulated gaussian function, Tic-Tac-Toe game problem, iris plants problem

I. 서 론

활성화 함수(activation function)가 등성이 형태(bump-like)를 이루어 패턴 인식을 하는 다층구조 뉴로 네트워크에서 코사인 모듈화 된 가우시안(CosGauss) 활성화 함수는 다항식과 계단함수의 근사에 사용될 수 있음을 증명한다.

그리고 이 활성화 함수를 다층 구조를 이루는 캐스케이드 코릴레이션 뉴로 네트워크에 적용하여 벤치마크 문제인 Tic-Tac-Toe 게임 문제와 아이리스(iris) 식물

문제에 대하여 실험하고 얻어진 결과를 다른 활성화 함수를 사용한 결과와 비교 분석한다.

CosGauss 함수는 다음과 같이 정의된다.

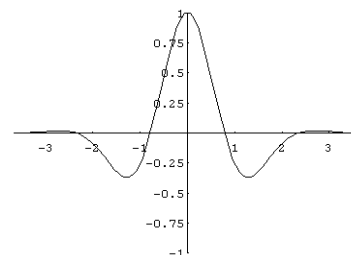


그림 1. $b=0.5$, $c=2$, $d=0$ 의 파라미터 값을 갖고 있는 CosGauss 함수

Fig. 1. CosGauss function with parameter $b=0.5$, $c=2$, $d=0$.

* 정회원, 서원대학교 정보통신공학과
(Department of Information and Communication,
Seowon University)
접수일자: 2011년10월27일, 수정완료일: 2012년3월5일

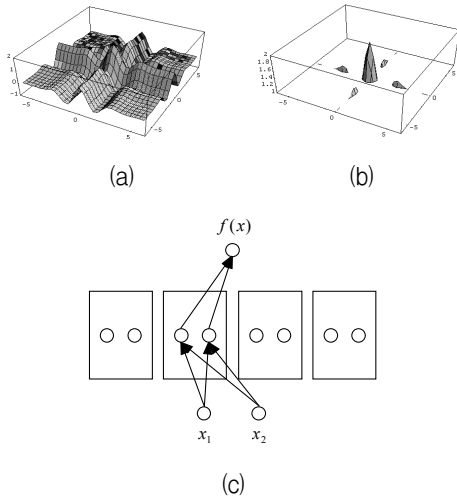


그림 2. (a) 한 쌍의 CosGauss 함수($b=0.1, c=2, d=0$)가 입력공간에서 형성하는 파형, (b) 얻어지는 로컬 범프(local bumps), (c) 그에 대응하는 함수근사(function approximation)를 위한 네트워크 구조

Fig. 2. (a) A pair of CosGauss function with parameter $b=0.1, c=2, d=0$, (b) local bumps, (c) network structure for the function approximation.

$$f_{act}(x) = e^{-b(x-d)^2} \cos(c(x-d))$$

여기에서 파라미터 b 는 함수가 가우시안함수에 의한 지수함수의 기울기를 결정하고 d 는 초평면(hyperplane)의 위치를 결정한다. c 는 함수 주기(period)의 길이를 결정한다.

II. CosGauss 함수의 동치(equivalence)

CosGauss 활성화함수가 뉴로 네트워크에서 스플라인(spline) 네트워크를 에뮬레이션(emulation)할 수 있음을 증명하기 위하여 다항식(polynomials)과 계단 함수(step function)를 근사(approximate)할 수 있음을 보여 주어야한다.^[1-2, 4] 이 Theorem들은 본 논문의 appendix에 기술하였다.

Lemma 1: CosGauss 활성화함수를 사용하는 뉴로 네트워크는 한 유한범위에서 다항식을 ε -근사 할 수 있다.

Proof: CosGauss 활성화함수 $g(x) = e^{-x^2} \cos(x)$ 가 Theorem 1의 조건을 만족함을 보일 것이다.

두 함수 $g_1(x) = e^{-x^2}$ 와 $g_2(x) = \cos(x)$ 는 첫 번째 조건

을 만족시키기 위한 $g_1(x)$ 와 $g_2(x)$ 의 0에서의 테일러급수(taylor series) 전개는 다음과 같다.

$$g_1(x) = \sum_{k=0}^{\infty} \frac{(-x^2)^k}{k!}$$

$$g_2(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}$$

두 테일러급수는 모든 $x \in \mathbf{R}$ 에서 수렴하기 때문에 이 두 테일러급수의 곱도 $x \in \mathbf{R}$ 에서 수렴한다.

$$g(x) = e^{-x^2} \cos(x)$$

$$= \left(\sum_{k=0}^{\infty} \frac{(-x^2)^k}{k!} \right) \left(\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} \right)$$

$$= \left(\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{k!} \right) \left(\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} \right) \quad (1)$$

식(1)은 코시 곱(Cauchy-Product)에 의하여 다음과 같이 주어진다.^[7]

$$\sum_{n=0}^{\infty} \left(\sum_{k=0}^n (-1)^k (-1)^{n-k} \frac{1}{k!} \frac{1}{(2n-2k)!} \right) x^{2n}$$

$$= \sum_{n=0}^{\infty} (-1)^n \left(\sum_{k=0}^n \frac{1}{k!} \frac{1}{(2n-2k)!} \right) x^{2n}$$

$$= \sum_{n=0}^{\infty} (-1)^n \frac{1}{(2n)!} \left(\sum_{k=0}^n \frac{1}{k!} \cdot \prod_{i=2n-2k+1}^{2n} i \right) x^{2n}$$

$$= \sum_{n=0}^{\infty} (-1)^n \frac{1}{(2n)!} \left(\sum_{k=0}^n \prod_{j=k+1}^n j \cdot \prod_{i=2n-2k+1}^{2n} i \right) x^{2n}$$

$$= \sum_{n=0}^{\infty} (-1)^n \frac{1}{(2n)!} \frac{1}{n!} \left(\sum_{k=0}^n \prod_{j=k+1}^n j \cdot \prod_{i=2n-2k+1}^{2n} i \right) x^{2n}$$

여기에서 $g(x)$ 는 멱급수(power series)로 표현 할 수 있으므로 첫 번째 조건을 만족한다.

멱급수에서 x 는 지수가 짝수로 나타나기 때문에 $n \in \mathbf{N}$ 을 위한 a_n 의 멱급수는 다시 다음과 같이 나타낼 수 있다.

$$a_{2n+1} = 0$$

$$a_{2n} = (-1)^n \frac{1}{(2n)! \cdot n!} \left(\sum_{k=0}^n \prod_{j=k+1}^n j \cdot \prod_{i=2n-2k+1}^{2n} i \right)$$

더 이상 줄일 수 없는 분수를 $a_{2n} = \frac{p_{2n}}{q_{2n}}$ 으로 놓으면, 분모 q_{2n} 은 상수 $c \in \mathbb{N}$ 에서 $q_{2n} = (2n)! \cdot n! \leq (2n)^{2n} \cdot n^n = 2^{2n} \cdot 2^{2n \cdot \log n} \cdot 2^{n \cdot \log n} \leq 2^{c(2n)^2}$ 으로 추정 할 수 있다. 또한 분자 p_{2n} 은 상수 $d \in \mathbb{N}$ 에서 다음과 같이 추정할 수 있다.

$$|p_{2n}| \leq \sum_{k=0}^n \prod_{j=k+1}^n j \cdot \prod_{i=2n-2k+1}^{2n} i \leq (n+1) \cdot n^n \cdot (2n)^{2n} \leq (n+1) \cdot 2^{n \cdot \log n} \cdot 2^{2n} \cdot 2^{2n \cdot \log n} \leq 2^{d(2n)^2}$$

이것으로 DasGupta와 Schnitger의 Theorem 1의 두 번째 조건을 증명하였다. Theorem 1의 첫 번째와 두 번째 조건이 만족하므로 세 번째 조건도 당연히 성립하고 따라서 Theorem 1에 의해서 CosGauss 활성화함수를 사용한 뉴로 네트워크는 한 유한한 범위 $[-D, D]$ 에 있는 다항식을 ε - 근사 할 수 있다.

Lemma 2: CosGauss 활성화함수를 사용한 뉴로 네트워크에서 n 개의 은닉뉴런들이 $[-1, 1] - [-2^{-n}, 2^{-n}]$ 범위에서 최대 2^{-n} 의 에러를 가지고 계단함수를 근사할 수 있다.

Proof: CosGauss 활성화함수를 사용한 뉴로 네트워크는 Theorem 2의 세 조건을 만족함을 증명한다.

Theorem 2의 첫 번째 조건에서 $x \geq 1$ 와 $\varepsilon \geq 0$ 을 위하여 $|g(x) - g(x + \varepsilon)| = O(\varepsilon/x^2)$ 가 성립함을 보여야한다.

$$\begin{aligned} f(x, \varepsilon) &= \exp(-x^2) \cos(x) \\ &\quad - \exp(-(x + \varepsilon)^2) \cos(x + \varepsilon) \\ &= \frac{\cos(x)}{\exp(x^2)} - \frac{\cos(x + \varepsilon)}{\exp(x^2 + 2x\varepsilon + \varepsilon^2)} \\ &= \frac{1}{\exp(x^2)} \left[\cos(x) - \frac{\cos(x + \varepsilon)}{\exp(2x\varepsilon + \varepsilon^2)} \right] \end{aligned}$$

$|f(x, \varepsilon)|$ 을 계산하면 다음과 같다:

$$|f(x, \varepsilon)| \leq \frac{|\cos(x)| + \left| \frac{\cos(x + \varepsilon)}{\exp(2x\varepsilon + \varepsilon^2)} \right|}{\exp(x^2)}$$

$x \geq 1$ 와 $\varepsilon \geq 0$ 에서 $|\cos(x)| \leq 1 + \varepsilon$, $|\cos(x + \varepsilon)| \leq 1 + \varepsilon$, $|\exp(2x\varepsilon + \varepsilon^2)| \geq 1$ 가 성립하고

$\alpha \geq 0$ 에서 $\frac{2\alpha}{\exp(x^2)} \leq \frac{\alpha}{x^2}$ 가 성립하기 때문에 $|f(x, \varepsilon)|$ 를 다음과 같이 추정할 수 있다.

$$|f(x, \varepsilon)| \leq \frac{2(1 + \varepsilon)}{\exp(x^2)} \leq \frac{1}{x^2} + \frac{\varepsilon}{x^2} \leq 1 + \frac{\varepsilon}{x^2}$$

따라서 $|f(x, \varepsilon)| = O\left(\frac{\varepsilon}{x^2}\right)$ 이 성립하므로 Theorem 2의 첫 번째 조건을 만족시킨다.

Theorem 2의 두 번째 조건을 만족시키기 위해서 $0 < \int_1^\infty e^{-u^4} \cos(u^2) du < \infty$ 임을 보여야한다. 우선 모든 $k \in \mathbb{N}$ 에서 e^{-u^4} 함수는 스트릭트 모노톤 감소 함수 (strictly monotonic decreasing)이므로 다음과 같다.

$$\left| \int_{\sqrt{k \cdot \pi + \frac{\pi}{2}}}^{\sqrt{(k+1) \cdot \pi + \frac{\pi}{2}}} e^{-u^4} \cos(u^2) du \right| \geq \left| \int_{\sqrt{(k+1) \cdot \pi + \frac{\pi}{2}}}^{\sqrt{(k+2) \cdot \pi + \frac{\pi}{2}}} e^{-u^4} \cos(u^2) du \right| \quad (2)$$

$$\begin{aligned} \int_1^\infty e^{-u^4} \cos(u^2) du &= \int_1^{\sqrt{\frac{\pi}{2}}} e^{-u^4} \cos(u^2) du \\ &\quad + \sum_{k=0}^\infty \int_{\sqrt{k \cdot \pi + \frac{\pi}{2}}}^{\sqrt{(k+1) \cdot \pi + \frac{\pi}{2}}} e^{-u^4} \cos(u^2) du \end{aligned}$$

$\sum_{k=0}^\infty \int_{\sqrt{k \cdot \pi + \frac{\pi}{2}}}^{\sqrt{(k+1) \cdot \pi + \frac{\pi}{2}}} e^{-u^4} \cos(u^2) du$ 는 하나의 교대급수(alternating series)이기 때문에 (2)에 의해서 모노톤 감소하고 당연히 0으로 수렴(converge)한다. 따라서

$\int_1^\infty e^{-u^4} \cos(u^2) du < \infty$ 을 보이는 것은 $\int_1^{\sqrt{\frac{\pi}{2}}} e^{-u^4} \cos(u^2) du < \infty$ 을 보이는 것으로 충분하다. 그러므로 다음과 같은 관계가 성립된다.

$$\int_1^{\sqrt{\frac{\pi}{2}}} e^{-u^4} \cos(u^2) du \leq \int_1^{\sqrt{\frac{\pi}{2}}} e^{-u^4} du \leq \int_1^{\sqrt{\frac{\pi}{2}}} e^{-u} du < \infty$$

또한 (2)로부터 다음과 같이 유도된다.

$$\int_1^{\infty} e^{-u^4} \cos(u^2) du \geq \int_1^{\sqrt{\frac{\pi}{2}}} e^{-u^4} \cos(u^2) du + \int_{\sqrt{\frac{\pi}{2}}}^{\sqrt{\frac{3\pi}{2}}} e^{-u^4} \cos(u^2) du$$

따라서 다음의 관계식이 성립하고

$$\int_1^{\sqrt{\frac{\pi}{2}}} e^{-u^4} \cos(u^2) du > \left| \int_{\sqrt{\frac{\pi}{2}}}^{\sqrt{\frac{3\pi}{2}}} e^{-u^4} \cos(u^2) du \right| \geq 0 \quad (3)$$

$\int_1^{\infty} e^{-u^4} \cos(u^2) du > 0$ 이 성립한다.

이제는 Theorem 2의 두 번째 조건의 증명을 위해서 식 (3)을 증명해야 한다.

식 (3)의 증명:

$f(u) := e^{-u^4} \cos(u^2)$ 라 놓으면 함수 $f(u)$ 는 그림3과 같다.

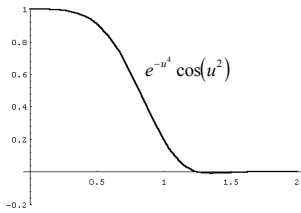


그림 3. 함수 $f(u)$
Fig. 3. Function $f(u)$.

다음에 $\int_1^{\sqrt{\frac{\pi}{2}}} e^{-u^4} \cos(u^2) du$ 의 계산을 위하여 그림4에서 는 구간 $[1, \sqrt{\frac{\pi}{2}}]$ 에서의 함수 $f(u)$ 를 보여준다.

그림 4의 점 $(1.1, f(1.1))$ 에서 함수 $f(u)$ 의 탄젠트 값에

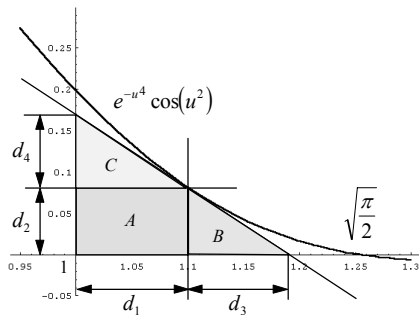


그림 4. 함수 $f(u)$
Fig. 4. Interval $[1, \sqrt{\frac{\pi}{2}}]$ of function $f(u)$.

의해서 그려지는 사선 아래에 만들어지는 직사각형 A, 직각 삼각형 B와 C가 있고 그에 대응하는 각각의 면적을 a, b, c라 한다.

$$\int_1^{\sqrt{\frac{\pi}{2}}} e^{-u^4} \cos(u^2) du > a + b + c \text{ 이 성립함을 보이기 위해}$$

서 $u \in [1, \sqrt{\frac{\pi}{2}}]$ 에서 $f''(u) > 0$ 임을 보이면 된다.

함수 $f(u)$ 의 2차 미분은 다음과 같이 계산되어진다.

$$f''(u) = \frac{-16u^2 \cos(u^2)}{e^{u^4}} + \frac{16u^6 \cos(u^2)}{e^{u^4}} - \frac{2 \sin(u^2)}{e^{u^4}} + \frac{16u^4 \sin(u^2)}{e^{u^4}}$$

여기에서 u^2 을 v 로 치환하면 v 는 구간 $[1, \frac{\pi}{2}]$ 에서 만 계산된다.

$$f''(\sqrt{v}) = \frac{-16v \cos(v) + 16v^3 \cos(v)}{e^{v^2}} - \frac{2 \sin(v) + 16v^2 \sin(v)}{e^{v^2}}$$

$$f''(\sqrt{v}) = \frac{v(16v^2 - 16) \cos(v) + (16v^2 - 2) \sin(v)}{e^{v^2}}$$

$v \in [1, \frac{\pi}{2}]$ 에서 $v > 0$, $(16v^2 - 16) \geq 0$, $\cos(v) \geq 0$, $(16v^2 - 2) > 0$, $\sin(v) > 0$ 가 성립하기 때문에 2차 미분 $f''(\sqrt{v})$ 는 항상 0보다 크다. 따라서 함수 $f(u)$ 는 두 점 $(1, f(1))$ 와 $(\sqrt{\frac{\pi}{2}}, 0)$ 사이에서 볼록(convex)하고 구간 $[1, \sqrt{\frac{\pi}{2}}]$ 에서 전환점(turning point)을 보유하지 않는다.

또한 그림 4에서 함수 $\int_1^{\sqrt{\frac{\pi}{2}}} e^{-u^4} \cos(u^2) du$ 는 사각형 A, 삼각형 B와 C에 대응하는 면적 a, b, c의 합을 반내림하여 계산한다. 사각형 A의 가로 d_1 , 세로 d_2 는 다음과 같다.

$$d_1 = 0.1$$

$$d_2 = f(1.1) \geq 0.081648$$

따라서 사각형 A의 면적 a를 추정하면 다음과 같다.

$$a = d_1 \times d_2 \geq 0.1 \times 0.081648 = 0.0081648$$

$f(u)$ 의 1차 미분함수는 (d_1, d_2) 에서 다음과 같이 계산된다.

$$f'(u) = \frac{-4u^3 \cos(u^2)}{e^{u^4}} - \frac{2u \sin(u^2)}{e^{u^4}}$$

$$f'(1.1) = \frac{-4 \cdot (1.1)^3 \cdot \cos((1.1)^2)}{e^{(1.1)^4}} - \frac{2 \cdot (1.1) \cdot \sin((1.1)^2)}{e^{(1.1)^4}} \leq -0.9107652$$

d_3 의 계산으로 면적 b를 구할 수 있다.

$$\frac{d_2}{d_3} \geq \frac{0.081648}{d_3} \geq \tan(0.9107652)$$

$$d_3 \geq 0.0633714$$

$$b = \frac{d_2 \times d_3}{2} \geq \frac{0.081648 \times 0.0633714}{2} \geq 0.002587$$

같은 방법으로 면적 c를 구하면 다음과 같다.

$$\frac{d_4}{d_1} = \frac{d_4}{0.1} \geq \tan(0.9107652)$$

$$d_4 \geq 0.1288403$$

$$c = \frac{d_1 \times d_4}{2} \geq \frac{0.1 \times 0.1288403}{2} \geq 0.006442$$

따라서 면적 $a+b+c$ 를 구하면 다음과 같다:

$$a+b+c \geq 0.0081648 + 0.002587 + 0.006442 = 0.0171938$$

다음 식을 계산하기 위하여

$$\left| \int_{\sqrt{\frac{\pi}{2}}}^{\sqrt{\frac{3\pi}{2}}} e^{-u^4} \cos(u^2) du \right|$$

함수 $h(u)$ 를 $h(u) := -e^{-u^4}$ 라 놓고 그림5에서 함수 $f(u)$ 와 $h(u)$ 를 구간 $\left[\sqrt{\frac{\pi}{2}}, \sqrt{\frac{3\pi}{2}}\right]$ 에서 나타낸다.

$|h(u)| \geq |f(u)|$ 의 관계가 성립하기 때문에 함수 $h(u)$ 는

$$\left| \int_{\sqrt{\frac{\pi}{2}}}^{\sqrt{\frac{3\pi}{2}}} e^{-u^4} \cos(u^2) du \right| \text{에서 계산되어진다.}$$

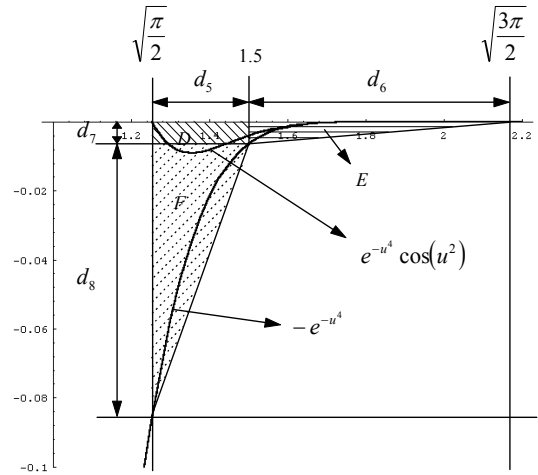


그림 5. 구간 $\left[\sqrt{\frac{\pi}{2}}, \sqrt{\frac{3\pi}{2}}\right]$ 에서 함수 $f(u)$ 와 함수 $h(u)$

Fig. 5. Interval $\left[\sqrt{\frac{\pi}{2}}, \sqrt{\frac{3\pi}{2}}\right]$ of functions $f(u)$ and $h(u)$.

그림 5에서 사각형 D와 삼각형 E, F에 각각 대응하는 면적 d, e, f의 합을 통해서 추정한다.

구간 $\left[\sqrt{\frac{\pi}{2}}, \sqrt{\frac{3\pi}{2}}\right]$ 에서 우선 임의의 한 점(1.5, 0)을 선택하면 d_5, d_6 을 알 수 있고 이어서 d_7, d_8 을 계산할 수 있다.

$$d_5 = 1.5 - \sqrt{\frac{\pi}{2}} \leq 0.2466859$$

$$d_6 = \sqrt{\frac{3\pi}{2}} - 1.5 \leq 0.67081$$

$$d_7 = h(1.5) \geq -0.0063298$$

$$d_8 = h\left(\sqrt{\frac{\pi}{2}}\right) - d_7 \geq -0.0784752$$

면적 d, e, f를 계산하기 전에 $\left| \int_{\sqrt{\frac{\pi}{2}}}^{1.5} -e^{-u^4} du \right| < d+f$ 와

$$\left| \int_{1.5}^{\sqrt{\frac{3\pi}{2}}} -e^{-u^4} du \right| < e \text{ 가 성립함을 보인다.}$$

두점 $(1.5, h(1.5))$ 와 $\left(\sqrt{\frac{3\pi}{2}}, 0\right)$, 그리고 $\left(\sqrt{\frac{\pi}{2}}, h\left(\sqrt{\frac{\pi}{2}}\right)\right)$ 와 $(1.5, h(1.5))$ 사이에서 함수 $h(u)$ 는 오목(concave)하기 때문에 $-h''(u) > 0$ 이 성립한다.

$$-h(u) = e^{-u^4}$$

$$-h'(u) = \frac{-4u^3}{e^{u^4}}$$

$$-h''(u) = \frac{-12u^2}{e^{u^4}} + \frac{16u^6}{e^{u^4}} = \frac{u^2(16u^3 - 12)}{e^{u^4}}$$

$u^2 > 0$ 그리고 $(16u^3 - 12) > 0$ 이기 때문에 $u \in \left[\sqrt{\frac{\pi}{2}}, 1.5\right]$ 와 $u \in \left[1.5, \sqrt{\frac{3\pi}{2}}\right]$ 에서 항상 $-h''(u) > 0$ 이 성립 한다.

$$\text{그러므로 } \left| \int_{\sqrt{\frac{\pi}{2}}}^{1.5} -e^{-u^4} du \right| < d + f \quad \text{와} \quad \left| \int_{1.5}^{\sqrt{\frac{3\pi}{2}}} -e^{-u^4} du \right| < e \quad \text{이}$$

성립 한다. 따라서

$$\left| \int_{\sqrt{\frac{\pi}{2}}}^{\sqrt{\frac{3\pi}{2}}} e^{-u^4} \cos(u^2) du \right| \leq \left| \int_{\sqrt{\frac{\pi}{2}}}^{\sqrt{\frac{3\pi}{2}}} -e^{-u^4} du \right| < d + e + f$$

가 성립하기 때문에 면적 $d + e + f$ 를 추정 할 수 있다.

$$d = d_5 \times d_7 \leq 0.2466859 \times 0.0063298 \leq 0.00156148$$

$$e = \frac{d_6 \times d_7}{2} \leq \frac{0.67081 \times 0.0063298}{2} \leq 0.00212305$$

$$f = \frac{d_5 \times d_8}{2} \leq \frac{0.2466859 \times 0.0784752}{2} \leq 0.00967937$$

$$d + e + f \leq 0.00156148 + 0.00212305 + 0.00967937 = 0.0133639$$

면적 $a + b + c$ 에서 면적 $d + e + f$ 를 빼면 다음과 같다.

$$a + b + c - (d + e + f) \geq 0.0171938 - 0.0133639 = 0.0038299$$

따라서 $a + b + c - (d + e + f) > 0$ 이 성립하므로 Theorem 2의 두 번째 조건을 만족 시킨다.

이제 Theorem 2의 세 번째 조건을 만족시키기 위해서 다음식이 성립함을 보여야한다.

$$\left| \int_{2^N}^{\infty} g(u^2) du \right| = O(1/N) \quad (N \geq 1)$$

$$g(u^2) = e^{-u^4} \cos(u^2) \text{ 이므로}$$

$$|g(u^2)| \leq e^{-u^4} \leq e^{-u} \quad (u \geq 1)$$

$$\left| \int_a^{\infty} g(u^2) du \right| \leq \int_a^{\infty} e^{-z} dz = \left[-e^{-z} \right]_a^{\infty} = \frac{1}{e^a}$$

$$\left| \int_{2^N}^{\infty} g(u^2) du \right| \leq \frac{1}{e^{2^N}} \leq \frac{1}{(2^N)^n} \leq \frac{1}{N} \quad (n \in \mathbb{N})$$

$\left| \int_{2^N}^{\infty} g(u^2) du \right| = O\left(\frac{1}{N}\right)$ 가 성립하므로 Theorem 2의 세 번째 조건을 만족하고 따라서 CosGauss 활성화함수는 계단함수를 근사할 수 있다.

다음 Lemma 3에서는 CosGauss 활성화함수를 사용한 뉴로 네트워크가 스플라인 네트워크(Spline-Network)와 동치임을 보여준다.

Lemma 3: CosGauss 활성화함수를 사용한 뉴로 네트워크는 $e(n, l) = 2^{-n}$ 의 오류를 가지고 있는 스플라인 네트워크와 동치(equivalent)이다.

Proof: Lemma 1과 Lemma 2에서의 증명을 통해서 CosGauss 활성화함수를 사용한 뉴로 네트워크가 스플라인(spline) 네트워크를 에플레이션 할 수 있음을 확인 하였다. CosGauss 함수가 스플라인을 근사함을 통해서 스플라인 네트워크는 CosGauss 함수를 에플레이션 할 수 있다. 그러므로 CosGauss 활성화 함수를 사용한 뉴로 네트워크와 스플라인은 동치이다.

III. 실험

첫 번째 실험은 Tic-Tac-Toe 벤치마크문제로서 Tic-Tac-Toe 게임에서 유래되었다.^[6] 이 게임은 두 명의 게임 참가자가 3x3 보드에서 번갈아 가며 ‘o’또는 ‘x’를 표시하여 가로, 세로 대각선 방향으로 연속적으로 3개를 표시한 자가 이긴다. 학습을 위한 데이터는

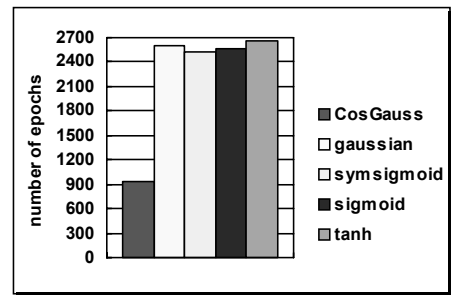


그림 6. 학습을 위해서 필요한 활성화 함수에 따른 에폭수의 비교
Fig. 6. Compare of the epochs with activation functions for learning.

표 1. 활성화 함수에 따른 학습 완료 실행 시간과 생성된 은닉 뉴런 수

Table 1. Completion learning time and the number of created hidden neuron.

활성화함수	트레이닝 데이터	실행시간	은닉뉴런 생성수
symsigmoid	958	0 hrs 22 min 38 sec	16.2
sigmoid	958	0 hrs 22 min 41 sec	16.5
gaussian	958	0 hrs 26 min 05 sec	20.1
tanh	958	0 hrs 29 min 38 sec	18.2
CosGauss	958	0 hrs 08 min 52 sec	7.3

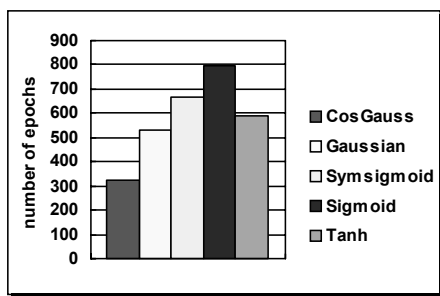


그림 7. 학습을 위해서 필요한 활성화 함수에 따른 에폭수의 비교

Fig. 7. Compare of the epochs with activation functions for learning.

표 2. 활성화 함수에 따른 학습 완료 실행 시간과 생성된 은닉 뉴런 수

Table 2. Completion learning time and the number of created hidden neuron.

활성화함수	트레이닝 데이터	실행시간	은닉뉴런 생성수
symsigmoid	150	0 hrs 0 min 56 sec	6.62
sigmoid	150	0 hrs 1 min 30 sec	7.9
gaussian	150	0 hrs 0 min 41 sec	5.3
tanh	150	0 hrs 0 min 47 sec	5.9
CosGauss	150	0 hrs 0 min 31 sec	3.5

Tic-Tac-Toe 보드의 마지막에 구성될 수 있는 958개로 얻어진다. 실험은 같은 조건하에 이루어 졌고 10번의 반복 실험을 통한 평균값을 결과 값으로 취하였다.

두 번째 실험은 아이리스(iris) 식물 문제로서 이 식물의 특징을 학습을 통하여 예견한다.^[6] 실험데이터는 3 클래스로 구성되고 각 클래스마다 50개의 경우에 따른 데이터로 구성된다. 줄기의 굵기와 길이, 잎의 길이와 넓이에 따라서 Iris-Setosa, Iris-Versicolour,

Iris-Virginica의 클래스로 구분되는 패턴을 인식한다.

두 실험에 의하면 CosGauss 활성화 함수가 기존의 다른 활성화 함수에 비해서 상당히 적은 에폭(epoch) 수를 기록하였고, 또한 적은 수의 은닉뉴런을 생성하였으며, 따라서 패턴인식의 실행시간도 적게 기록하였음을 확인 할 수 있었다. 이러한 결과로 CosGauss 함수가 다항식과 계단함수를 빠르게 근사하여 스플라인 네트워크를 에뮬레이션 할 수 있음을 보여주었다.

IV. 결 론

코사인 모듈화된 가우스함수인 CosGauss 함수가 뉴로 네트워크에서 다항식과 계단함수를 근사함을 보여줌으로써 CosGauss 활성화함수를 사용한 뉴로 네트워크가 스플라인네트워크(Spline-Netzwerk)와 동치임을 증명하였다. CosGauss 함수를 활성화함수로 사용한 실험에서는 뉴로 네트워크를 빠른 속도로 근사함으로써 기존의 다른 활성화함수에 비해서 적은 에폭수를 기록 하였고 또한 학습완료 실행 시간도 적게 소요되었음을 보여주었다. 따라서 상당히 빠른 속도의 패턴 인식을 갖고 있음을 알 수 있다. 앞으로는 실험을 통하여 패턴 인식 능력을 더욱더 최적화 시킬 수 있는 함수의 파라미터 값을 찾아내는 연구가 필요할 것이다.

참 고 문 헌

- [1] B. DasGupta und G. Schnitger, "The Power of Approximating: a Comparison of Activation Functions", Advances in Neural Information Processing Systems⁵, Edited by R. P. Lippmann, J. E. Moody and D. S. Touretzky, Morgan Kaufmann, pp. 615-622, 1993.
- [2] B. DasGupta und G. Schnitger, "Efficient approximation with neural networks: A comparison of gate functions", Pennsylvania State University, 1993.
- [3] S. E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture," Advances in Neural Information Processing Systems 2, Morgan Kaufmann, 1990.
- [4] G. W. Flake, "Nonmonotonic activation functions in multilayer perceptrons", Dissertation, Institute for Advance Computer Studies department of Computer Science University of Maryland, 1993.
- [5] B. Hammer, A. Micheli and A. Sperduti,

“Universal approximation capability of cascade correlation for structures,” Neural Computation, Vol.17, No.5, pp.1109-1159, 2005.

[6] L. Prechelt, “PROBEN1-A Set of Neural Network Benchmark Problems and Benchmarking Rules,” Technical Report 21/94, Department of Computer Science, University of Karlsruhe, 1999.

[7] W. Walter, “AnalysisI”, zweite Auflage, Springer- Verlag, pp.103, 1990.

Appendix

Definition 1: (DasGupta and Schnitger as mentioned in [1][2][4]) Let $e: N^2 \rightarrow R_+$ be an error measure whose arguments are the number of nodes of the network, n, and the number of the layers l.

Let Γ_1 and Γ_2 be two neural network architectures, and let N_1 and N_2 represent networks from Γ_1 and Γ_2 , respectively.

- Γ_1 emulates Γ_2 if and only if for all functions $f: [-1, 1]^l \rightarrow R$, if N_1 approximates f with n hidden units, l layers, and $e(n, l)$ error, then there exists an N_2 which approximates f with $(n+1)^k$ hidden units, $k(l+1)$ layers, and $e(n, l)$ error, for some constant k.
- Γ_1 and Γ_2 are equivalent if and only if they emulate each other with respect to e.

In the next theorem, neural networks with activation functions that satisfy the specified conditions are capable of approximating polynomials.

Theorem 1: (DasGupta and Schnitger as mentioned in [1][2][4]) A neural network with an activation function $g: R \rightarrow R$ can e-approximate any polynomial over some finite domain $[-D, D]$ if and only if there exist real numbers a, b, ($\alpha > 0$) and an integer k such that

- g can be represented by the power series $\sum_{i=0}^{\infty} a_i(x-\beta)^i$ for all $x \in [-\alpha, \alpha]$.
- The coefficients are rational numbers of the form

$$a_i = \frac{p_i}{q_i} \text{ with } |p_i|, |q_i| \leq 2^{\text{poly}(i)} \text{ (for } i > 1).$$

- For each $i > 2$ there exists j with $i \leq j \leq i^k$ and $a_j \neq 0$.

Theorem 2: (DasGupta and Schnitger as mentioned in [1][2][4]) A neural network with an activation function $g: R \rightarrow R$ and n hidden units can approximate the step function over the domain $[-1, 1] - [-2^{-n}, 2^{-n}]$ with error at most 2^{-n} when

- $|g(x) - g(x+\epsilon)| = O(\epsilon/x^2)$, for $x \geq 1, \epsilon \geq 0$.
- $0 < \int_0^{\infty} |g(1+u^2)| du < \infty$.
- $\left| \int_{2^N}^{\infty} g(u^2) du \right| = O(1/N)$, for all $N \geq 1$.

Theorem 3: (DasGupta and Schnitger as mentioned in [1][2][4]) Assume that Γ is a neural network architecture which can approximate polynomials in the sense of Theorem 1 and can approximate the step function in the sense of Theorem 2

- Γ can emulate spline networks.
- If each activation function in a neural network can be approximated over the domain $[-2^{-n}, 2^{-n}]$ with a spline network with n units, constant number of layers, and error 2^{-n} , then Γ is equivalent to splines.

— 저 자 소 개 —



이 상 화(정회원)
 1988년 도르트문트대학교(독일) 컴퓨터공학과 Vordiplom 졸업
 1994년 도르트문트대학교(독일) 컴퓨터공학과 Diplom 졸업
 1998년 도르트문트대학교(독일) 컴퓨터공학과 박사 졸업
 2000년~2002년 한국전자통신연구원 선임연구원
 2002년~현재 서원대학교 정보통신공학과 교수
 <주관심분야 : 컴퓨터 통신, 지능형 시스템, RFID, 뉴로컴퓨팅>