김일성종합대학학보 주체109(2020)년
정보과학 제66권 제3호

불균형자료에 대한 초학습기계실현에서 확률비용함수를 리용하기 위한 한가지 방법

김훈, 주일령

경애하는 최고령도자 **김정은**동지께서는 다음과 같이 말씀하시였다.

《우리는 가까운 앞날에 전반적인 과학기술분야에서 세계를 디디고 올라설수 있다는 배심을 가지고 첨단돌파의 기적들을 련이어 창조하여야 합니다.》

초학습기계(ELM)[1, 2]는 일반화된 단층앞방향망(SLFNs)에 대한 새로운 학습알고리듬으로서 다른 인공신경망보다 실행이 간단하고 빠르므로 많은 실천에 응용된다. 그러나 자료의 분포를 정확하게 표현하지 못하므로 자료클라스들사이의 분류정확도가 높지 못하고 불균형자료에 대하여 소수클라스쪽으로 치우치므로 출입체계와 같은 인식과제에 적당하지 않다.[3, 4]

불균형자료문제에 대한 분류를 잘하지 못하는 초기ELM의 결함을 극복하기 위하여 무게붙은 ELM[3]이 제안되였다. 그 방법의 본질은 다수클라스의 영향을 약화시키고 동시에 소수클라스의 영향을 강화하도록 매 견본에 정확한 무게를 할당하는것이다.

무게붙은 ELM에서 매 훈련견본은 그것이 속하는 클라스의 견본개수에 따라 계산된 무게를 할당하였다. 그래서 무게붙은 ELM은 초기ELM의 우점을 유지하면서 불균형자료에 대한 분류성능을 개선하였다.

무게붙은 ELM에서는 모든 훈련견본 x_i와 관련된 1개 $N \times N$ 대각선무게행렬 W를 정의하고 소수클라스의 원소 x_{ii}에 대응하는 무게 W_{ii}는 다수클라스의 원소들에 대응하는 무게보다 상대적으로 더 크게 설정하였다. 결과 소수클라스들이 강조되고 동시에 다수클라스들이 약화되였다.

대각선무게행렬 W를 고려하면 무게붙은 ELM은 다음식으로 표시할수 있다.

$$L_{WELM} = \frac{1}{2}\|\beta\|^2 + C\frac{1}{2}\sum_{i=1}^{N}(W_{ii} \times \|\xi\|^2) \tag{1}$$

여기서

$$h(x_i)\beta = t_i^T - \xi_i^T,\ i = 1, \cdots, N$$

이고

$$\beta = H^* T = \begin{cases} H^{\mathrm{T}}\left(\dfrac{I}{C} + WHH^{\mathrm{T}}\right)^{-1} WT, & N < L\text{일 때} \\ \left(\dfrac{I}{C} + H^{\mathrm{T}}WH\right)^{-1} WH^{\mathrm{T}}T, & N \ge L\text{일 때} \end{cases} \tag{2}$$

이다. 그리고 N은 벡토르모임, T는 벡토르모임에 대응하는 클라스표식벡토르, t_i는 벡토르모임에 대응하는 클라스표식, H는 숨은층출력행렬, β는 숨은층출력무게행렬, $h(x_i)$는 숨은층출력행렬의 원소, ξ는 훈련견본 x_i에 대응한 N개의 출력마디의 훈련오유벡토르, C는 훈련오유최소화와 클라스들사이 거리최대화의 균형파라메터이다.

제안한 2개의 무게구성은 다음과 같다.

$$W_1: \qquad W_{ii} = \frac{1}{\# t_i}$$

$$W_2: \qquad W_{ii} = \begin{cases} \dfrac{0.618}{\# t_i}, & \# t_i > AVG(\# t_i)\text{일 때} \\ \dfrac{1}{\# t_i}, & \# t_i \le AVG(\# t_i)\text{일 때} \end{cases} \tag{3}$$

여기서 $\# t_i$ 는 i째 클라스에 속하는 견본들의 개수이며 $AVG(\# t_i)$ 는 모든 클라스들의 평균견본개수이다.

론문에서는 훈련견본자료가 매 클라스에 속할 확률을 고려하여 확률비용함수를 정의하고 이 확률비용함수를 리용하여 ELM문제의 최량풀이를 구하였다.

입력훈련자료를 주어진 훈련견본자료가 매 클라스에 속할 확률값으로 이루어진 1개 벡토르로 넘기고 매 훈련견본자료가 클라스에 정확히 속할 정도를 평가하기 위하여 확률비용함수를 정의하였다.

M클라스들로 이루어진 훈련모임을 $(x_i,\ t_i),\ i=1,\ \cdots,\ N$ 으로 표시하자.

이때 표준ELM알고리듬에 관한 이 훈련모임의 숨은층출력행렬 H는 다음과 같다.

$$H = \begin{bmatrix} F(a_1^{\mathrm{T}} \cdot x_1 + b_1) & \cdots & F(a_L^{\mathrm{T}} \cdot x_1 + b_L) \\ \vdots & \ddots & \vdots \\ F(a_1^{\mathrm{T}} \cdot x_N + b_1) & \cdots & F(a_L^{\mathrm{T}} \cdot x_N + b_L) \end{bmatrix}_{N \times L} \tag{4}$$

여기서 F는 숨은층의 적용함수이다.

편리하게 하기 위하여 $H_i (i=1,\ 2,\ \cdots,\ N)$ 를 다음과 같이 정의하자.

$$H_i = (F(a_1^{\mathrm{T}} \cdot x_i + b_1), F(a_2^{\mathrm{T}} \cdot x_i + b_2),\ \cdots,\ F(a_L^{\mathrm{T}} \cdot x_i + b_L)) \tag{5}$$

그러면

$$H = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_N \end{bmatrix}_{N \times L}$$

이다.

두 분류 ELM분류기에서 출력무게행렬 β 는 $L \times M$ 행렬이다.

$$\beta = (B_1,\ B_2,\ \cdots,\ B_M)_{L \times M}$$

여기서 $B_i = [B_{1i},\ B_{2i},\ \cdots,\ B_{Li}]^{\mathrm{T}},\ i=1,\ \cdots,\ M$ 이다.

따라서 표준ELM의 출력은 다음과 같다.

$$H\beta = \begin{bmatrix} H_1 B_1 & H_1 B_2 & \cdots & H_1 B_M \\ H_2 B_1 & H_2 B_2 & \cdots & H_2 B_M \\ \vdots & \vdots & \ddots & \vdots \\ H_N B_1 & H_N B_2 & \cdots & H_N B_M \end{bmatrix}_{N \times M} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix}$$

여기서

$$f_i = (H_i B_1,\ H_i B_2,\ \cdots,\ H_i B_M),\ i=1,\ 2,\ \cdots,\ N$$

이다.

이때 표준ELM의 최량화공식은 다음과 같다.

$$L_{ELM} = \frac{1}{2}\|\beta\|^2 + C\frac{1}{2}\sum_{i=1}^{N}\|\xi_i\|^2 \tag{6}$$

여기서 첫항은 여분거리와 관련되며 둘째 항은 훈련오유와 관련된다.

우리는 식 (6)의 둘째 항을 주어진 훈련견본이 매 클라스에 속할 확률과 관련된 값으로 변화시켰다.

이를 위해 다음과 같은 확률함수를 리용한다.

$$p(x_i,\ j,\ \beta) = \frac{e^{H_i B_j}}{\sum_{j=1}^{M} e^{H_i B_j}},\ i = 1,\ \cdots,\ N,\ j = 1,\ \cdots,\ M \tag{7}$$

$$0 < p(x_i,\ j,\ \beta) < 1,\ \sum_{j=1}^{M} p(x_i,\ j,\ \beta) = 1,\ j = 1,\ \cdots,\ M$$

이 값들은 훈련견본 x_i가 매 클라스에 속할 확률로 볼수 있다. 다시말하여 매 훈련견본이 임의의 클라스에 속할 확률은 이 함수에 의하여 얻어질수 있다.

그러므로 확률비용함수를 다음과 같이 정의하였다.

$$L(\beta) = \frac{1}{2}\|\beta\|^2 - C\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} t_{ij} \log p(x_i,\ j,\ \beta) \tag{8}$$

여기서 훈련견본 x_i에 대응하는 목표벡토르

$$t_i = [t_{i1},\ t_{i2},\ \cdots,\ t_{iM}],\ i = 1,\ 2,\ \cdots,\ N$$

의 매 원소들은 클라스 k에 속하면 $t_{ik} = 1$로, 그렇지 않으면 $t_{ik} = 0$으로 설정된다.

$L(\beta)$의 최소값을 구하기 위하여 그라디엔트하강법, L-BFGS와 같은 반복최량화알고리듬을 리용한다. 이때 그것의 도함수는 다음과 같다.

$$\nabla_{B_j} L = B_j - \frac{C}{N}\sum_{i=1}^{N}[H_i(t_{ij} - p(x_i,\ j,\ \beta))],\ j = 1,\ \cdots,\ M \tag{9}$$

β에 관한 $L(\beta)$의 최소화를 진행하여 최량출력무게행렬을 얻는다.

확률비용함수에 기초한 초학습기계의 동작알고리듬은 다음과 같다.

입력: 훈련자료 $(x_i,\ t_i),\ i = 1,\ \cdots,\ N$, 숨은층의 능동함수 F

출력: 출력무게 β

걸음 1 우연적으로 입력파라메터 $a_i,\ b_i,\ i = 1,\ \cdots,\ L$ 들을 할당한다.

걸음 2 식 (4), (5)로 H_i를 계산한다.

걸음 3 식 (7)로 확률함수 $p(x_i,\ j,\ \beta),\ j = 1,\ \cdots,\ M$을 정의하고 이를 리용하여 최량화식 즉 확률비용함수식 (8)을 얻는다.

걸음 4 반복최량화알고리듬을 리용하여 확률비용함수식 (8)의 최량풀이 β를 얻는다.

걸음 5 β를 출력무게로 하여 출력한다.

맺 는 말

훈련견본자료가 매 클라스에 속할 확률을 고려한 확률비용함수를 정의하고 이 확률비용함수를 리용하여 ELM최량화문제를 정식화하였으며 이 방법으로 두 분류문제를 풀기 위한 학습알고리듬을 제안하였다.

참 고 문 헌

[1] G.-B. Huang et al.; Neurocomputing, 70, 1, 489, 2006.
[2] G. B. Huang et al.; IEEE Computational Intelligence Magazine, 10, 2, 18, 2015.
[3] A. Iosifidis et al.; IEEE Transactions on Cybernetics, 46, 1, 311, 2016.
[4] E. A. Garcia et al.; IEEE Trans. Knowl. Data Eng., 21, 9, 1263, 2009.

주체109(2020)년 5월 5일 원고접수

A Study on a Method for Using Probability Cost Function in Extreme Learning Machine Realization on Imbalance Data

Kim Hun, *Ju Il Ryong*

In this paper, we propose a modified ELM algorithm that defined by probability cost function where probability which training samples belong to each class is found.

Keywords: extreme learning machine, imbalance learning, probability cost function