

ART1 클러스터링 성능 향상을 위한 초기 학습순서 조정

*최대훈, 임성길, 이현수
 경희대학교 컴퓨터공학과

e-mail : r6master@khu.ac.kr, skan72@khu.ac.kr, leehs@khu.ac.kr

Adjustment of initial learning order to improve clustering performance of ART1

*Tae-Hun Choi, Sung-Kil Lim, Hyon-Soo Lee

Department of Computer Engineering, Kyung Hee University

Abstract

This paper presents adjustment of input order to improve clustering performance of ART1. We propose new method for On-line clustering which adjusts initial input data using buffer. We demonstrate the clustering performance of the proposed algorithm by testing it on Zoo data set from UCI and created artificial data set for simulation. Experimental results show that preposed method increases 7.8% of clustering performance than ART1 model on the average.

I. 서론

ART1 (Adaptive Resonance Theory 1) 신경망 모델은 과거에 학습한 데이터를 유지하며 새로운 데이터를 지속적으로 학습 할 수 있는 On-line 비교사 클러스터링 알고리즘이다. ART1의 클러스터링 결과는 입력순서에 따라 달라지며, 또한 Vigilance Parameter (ρ)값에 따라 클러스터의 수가 결정된다.

ART1에서 입력 순서에 대한 문제점을 해결하기 위해서 같은 클러스터에 속하는 데이터를 우선적으로 학습하는 Re-Ordering [1] 방법과 서로 다른 클러스터에 속하는 데이터들을 순서대로 학습하는 Circular Ordering [2] 방법이 제안되었다. 그러나 이 방법들은 입력순서를 결정하기 위하여 클러스터 정보가 필요하다는 단점을 가지고 있다. 또한 전체 입력 데이터에 대하여 입력순서를 조정하기 때문에 Off-line 클러스터링에 적합하며 On-line 클러스터링에는 적용하기 어려운 방법이다.

본 논문에서는 ART1에서 사전 클러스터링 정보가 필요 없고 On-line 학습에 적합한 학습순서 조정 방법을 제안한다. 제안하는 방법은 On-line 학습을 위하여 학습초기의 데이터에 대해서만 학습순서를 조정하며, 같은 클러스터에 속한 데이터임을 판단하기 위하여 사

전 클러스터링 정보 대신 ρ 값을 이용한 유사도를 사용한다.

II. 제안한 학습순서 조정 방법

같은 클러스터에 속하는 데이터들을 먼저 학습하면 학습된 클러스터 중심이 원래의 클러스터 중심과 가까워지기 때문에 생성된 클러스터가 안정되어 클러스터링 성능을 향상시킬 수 있다 [1]. 그러나 사전 클러스터링 정보 없이 새로운 입력 데이터가 기존 클러스터에 속하는가를 판단하는 것은 매우 어려운 일이다. 따라서 본 논문에서는 ART1에서 ρ 값을 기준으로 새로운 클러스터 생성 여부를 결정 한다는 점을 이용하여 입력 데이터들이 같은 클러스터에 속하는가를 판단한다. 그리고 ART1 모델의 On-line 학습을 유지하며 학습순서를 조정하기 위해서 초기의 입력데이터를 버퍼에 저장 하였다가 한 번에 학습하는 방법을 사용한다.

ART1에서 유사도는 (식1)과 같다. 식 1에서 C_j 는 클러스터 중심, X_i 는 현재 입력 데이터, $\|\cdot\|$ 는 \cdot 의 Norm을 의미한다. 유사도의 범위는 [0, 1] 이며 현재 입력 데이터와 클러스터의 중심이 유사 할수록 1에 수렴한다.

$$S_{C_j}(X_i) = \frac{\|C_j \cap X_i\|}{\|X_i\|} \quad (식1)$$

본 논문에서 제안하는 ART1 모델의 학습순서 조정 방법에서는 전체 학습과정을 학습초기와 학습후기로 나눈다. 학습초기에는 입력 데이터를 버퍼에 저장한 후, 제안하는 학습순서 조정과정과 ART1 모델의 동작을 반복적으로 수행한다. 학습후기에는 ART1 모델과 동일하게 동작한다.

학습초기에 버퍼의 크기만큼 저장된 입력 데이터 중 하나를 첫 번째 클러스터 중심으로 결정하고, (식1)에 의해 버퍼안의 나머지 데이터와 첫 번째 클러스터 중심

과의 유사도를 계산한다. 계산된 유사도에 따라 버퍼안의 나머지 데이터를 내림차순으로 정렬한다. 기존에 학습된 모든 클러스터가 Inhibit 될 때 까지 정렬된 데이터를 차례로 ART1 모델에 입력하여 클러스터링을 수행한다. 모든 클러스터가 Inhibit 된 경우 새로운 클러스터를 생성한다. 새로운 클러스터 생성은 버퍼에 남아있는 데이터중 하나를 클러스터 중심으로 결정하고, 나머지 데이터와 유사도를 계산, 유사도에 따라 정렬하는 과정으로 구성된다. 버퍼에 데이터가 남아있는 동안 ART1 모델 동작과 새로운 클러스터 생성과정을 반복한다.

III. 실험결과 및 고찰

본 논문에서는 클래스별 학습이 임의의 순서로 학습하는 것 보다 클러스터링 성능을 향상 할 수 있음을 보이는 실험과 버퍼의 크기와 클러스터링 결과의 관계를 알아보기 위한 실험을 수행하였다. 비교대상으로는 데이터 입력순서대로 학습하는 일반적인 ART1 모델을 선택하였다.

실험에는 Zoo 데이터[3]와 클러스터링 성능평가를 위해 임의로 생성한 6 Cluster 데이터를 사용하였다. 실험에 사용한 데이터는 모두 이진데이터 이며, 각 데이터에 대한 특성은 표1과 같다. 표1에서 nPtrn은 전체 패턴의 수, nClst는 클러스터의 수, Dim은 패턴의 차원 수, cDstn는 클러스터 중심간의 유클리드 거리의 평균, pDstn는 클러스터내 패턴간의 유클리드 거리의 평균을 나타낸다.

표 1. 실험 데이터의 특성

	nPtrn	nClst	Dim	cDstn	pDstn
Zoo Data	101	7	21	6.27	2.36
6 Cluster	150	6	100	20.13	9.12

성능 평가를 위한 파라미터는 각 입력 데이터들간의 오리지널 클러스터 인덱스와 클러스터링 결과 인덱스에 대한 정합률(Correct Clustering Rate)을 사용하였다. 클러스터링 결과 인덱스는 동일 클러스터 내에서 오리지널 인덱스가 가장 많은 것으로 선택 하였다. CCR은 (식2)와 같이 정의한다. (식2)에서 x 는 입력 데이터, C_i 는 오리지널 클러스터, P_j 는 클러스터링 결과, $\kappa(S)$ 는 집합 S 의 원소의 수를 의미한다.

$$CCR = \sum_i p(x \in P_i | x \in C_i) + \sum_j p(x \in C_j | x \in P_j) \quad (식2)$$

$$\text{where } P(x \in S) = \frac{\kappa(S)}{\sum_i \kappa(C_i) + \sum_j \kappa(P_j)}$$

그림 1은 임의의 순서로 학습한 경우와 제안한 클러스터링 방법, 클러스터별로 학습한 경우의 성능을 나타낸 것이다. 실험결과 Zoo 데이터의 경우에는 클래스별로 학습한 경우가 임의의 순서로 학습한 경우보다 클러스터링 성능이 11.3% 향상되었고, 제안한 방법은 7.8% 향상되었다. 6 Cluster 데이터의 경우에는 각각 9.8%

와 1.2% 향상되었다. 실험결과 클러스터별로 학습하는 것이 클러스터링 성능을 향상 시키는 것을 볼 수 있다.

그림 2는 버퍼 크기와 클러스터링 결과의 관계를 나타낸 것이다. 그래프에서 굵은 실선은 서로 다른 입력순서로 학습시킨 10회의 실험결과에 대한 평균값이며, 최상과 최저 클러스터링 성능을 함께 표시 하였다. 실험결과 입력버퍼의 크기가 증가 할수록 클러스터링 성능이 향상됨을 확인 할 수 있었다.

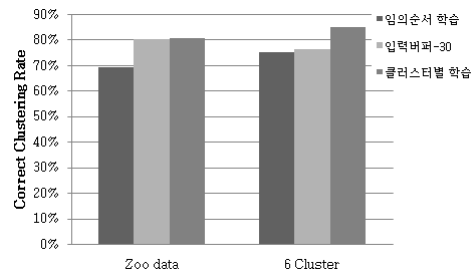
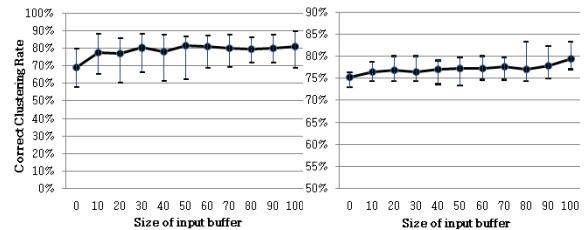


그림 1. 클러스터링 방법별 성능 (입력버퍼 크기 : 30)



(a) Zoo Data ($\rho=0.55$) (b) 6 Cluster Data ($\rho=0.3$)
그림 2. 버퍼 크기별 클러스터링 결과

IV. 결론 및 향후 연구방향

본 논문에서 ART1 모델의 클러스터링 성능 향상을 위해 학습순서를 조정하는 방법을 제안하였다. 제안된 방법은 학습 초기 데이터의 순서만을 조정하고 사전 클러스터 정보를 필요로 하지 않기 때문에 On-line 클러스터링에 적합하다. 실험결과 임의의 순서로 학습한 경우에 비해 클러스터링 성능이 향상 됨을 확인하였다.

문제에 따라 적정한 입력 버퍼의 크기와 ρ 값을 결정하는 것은 향후 연구해야 할 과제이다.

V. 참고문헌

[1] Dagher I et al, "An ordering algorithm for pattern presentation in fuzzy ARTMAP that tends to improve generalization performance", Neural Networks, IEEE Transactions on Volume 10, Issue 4, July 1999 Page(s) 768 - 778
 [2] Youngtae Park, "An ART2 trained by two stage learning on circularly ordered data sequence, 1994 IEEE International Conference on Volume 5, June 1994 Page : 2928 - 2933
 [3] University of California Irvine Machine Learning Repository, <http://archive.ics.uci.edu/ml/>