

상황인식 기반 클러스터링의 적응적 자율 학습 분할 알고리즘

전일규 · 이강환*

한국기술교육대학교

Context-awareness Clustering with Adaptive Learning Algorithm

Il-Kyu Jeon · Kang-whan Lee*

Korea University of Technology and Education

E-mail : stj@koreatech.ac.kr / kwlee@koreatech.ac.kr

요 약

본 논문은 이동 노드간 클러스터링을 함에 있어 보다 효율적인 클러스터링을 제공하고 유지하기 위한 딥러닝의 자율학습에 따른 군집적 알고리즘을 제안한다. 대부분 이동 노드의 클러스터링 군집데이터를 처리함에 있어 상호관계에 따른 분류체계가 제공된다. 이러한 경우 새롭게 입력되거나 변경된 데이터가 비교정보에서 오염된 정보로 분류될 경우 기존 분류된 클러스터링으로부터 오염된 정보로 이해되어 군집성을 저하시키는 요인으로 작용 할 수가 있다. 본 논문에서는 이러한 상황정보를 이해하고 클러스터링을 유지할 수 있는 자율학습기반의 학습 모델을 제시 한다.

ABSTRACT

This paper propose a clustering algorithm for mobile nodes that possible more efficient clustering using context-aware attribute information in adaptive learning. In typically, the data will be provided to classify interrelationships within cluster properties. If a new properties are treated as contaminated information in comparative clustering, it can be treated as contaminated properties in comparison clustering. In this paper, To solve this problems in this paper, we have new present a context-awareness learning based model that can analyzes the clustering attributed parameters from the node properties using accumulated information properties.

키워드

Clustering Algorithm, K-means Clustering, non-supervisor learning, Context-awareness

I. 서 론

일반적으로 클러스터링 기반의 데이터를 처리함에 있어 데이터의 속성간의 분해와 상관관계에 따른 분류는 클러스터링을 유지하고 분석함에 매우 중요한 요소이다. 특히 상황인식기반에서의 클러스터링 단위 데이터를 분석하고 이를 제공함에 있어 비지도학습에 따른 K-means 알고리즘은 가장 많이 적용되는 방법이기도 하다. 본 논문에서는 제공되는 노드간의 속성데이터를 분석하고 구분하기 위해 노드간의 속성간의 정보를 분석하고 이로부터 클러스터링을 유지하는 비지도 학습방법 중의 하나인 클러스터링 분할 알고리즘을 위한 딥러닝의

자율 학습방법을 제시하고자 한다.

II. 본 론

크기가 비교적 균일한 클러스터를 생성하는 클러스터링 데이터에서 유효한 데이터의 한계성을 극복하는 것은 매우 중요하다. 이 과정에서 일반적으로 K-means 클러스터링 알고리즘은 다음과 같은 5가지의 문제점에 대해 고려를 해야 한다[1][2].

•K-means에 의한 군집화 결과의 분포는 데이터의 중심속성에 의한 유지성의 확보

* speaker

- 클러스터링의 중심점은 상호 인접한 다른 클러스터의 중심점과 충분한 마진의 확보

- 클러스터링의 구성은 데이터 학습에 따른 이동 예측이 가능

- 클러스터내 데이터 속성의 분포가 K-means 군집화의 성능에 어떠한 영향을 미칠 수 있는지를 고려

- K-means 클러스터링에서 데이터의 유사성과 비 유사성의 독립적 척도가 제공되는지의 여부

따라서 위와 같은 문제점을 해결하고 고려함이 필요하다. 이를 위해 제시된 노드간의 속성 분류의 그룹은 노드 속성에 대한 특징에 따라 상당히 다른 모집단의 변동을 비교할 수 있는 속성 가중치 인자가 요구된다. 즉, 주어진 속성가중치 인자의 평균값에 의해 모집단의 데이터 변화에 따른 오염도를 특정할 필요가 있다. 즉 노드간의 속성정보에 대한 모집단 데이터의 유사성과 비유사성의 척도를 판별할 수 있는 독립적 척도를 제공하여 클러스터링내에서 이동노드간의 속성 데이터 유지 및 이탈을 결정할 필요가 있게 된다. 본 연구에서는 데이터의 유효성을 위한 편향값(Biased effect) 내지는 클러스터의 분기된 클러스터에서 가중치가 적용된 속성데이터의 독립 속성척도값(Gini Index Property(GIP))을 제공하여 클러스터링내의 이동 노드간의 속성 데이터를 분석하고자 한다. 이를 위해 본 연구에서는 각 이동 노드 클러스터내에서의 데이터의 유효성에 대한 상대빈도값으로 이동 노드 클러스터(C_i)에서의 데이터 속성(i)의 상대빈도값 $P(C_i)$ 을 Gini Index(GI)로 표현하여 다음과 같이 나타낼 수 있다.

$$GI(C_i) = 1 - \sum_i P(C_i)^2$$

이는 임의의 이동 노드 클러스터(C_i)에서 데이터 속성(i)에 대한 상대빈도율을 의미하는 것으로 GI는 적을수록 데이터의 유효성은 높게 평가된다. 따라서 이를 노드의 속성 데이터에 대한 유효성 척도로 사용할 수 있을 것이다. 만일 어떤 k 개의 임의의 이동 노드 클러스터에서 해당 클러스터내에서 특정 속성의 노드 데이터가 변화되어 분기된 어떤 이동 노드 클러스터에서 각 속성데이터에 대해 가중치가 적용된 속성데이터의 독립적 속성척도값(Gini Index Property(GIP))은 다음과 같이 나타낼 수 있을 것이다.

$$GIP(C_i) = \sum_{i=1}^k \frac{n_i}{n} GI(C_i)$$

n_i =이동 클러스터내 전체데이터의 갯수,

n_i =이동클러스터(C_i)에서의 변화된 데이터갯수

따라서, 임의의 이동 노드 클러스터(C_i)에서 상관 관계를 가지는 이동 노드의 데이터 변화에 대한 속성데이터의 독립적 속성척도값(Gini Index Property(GIP))은 클러스터내에서 중심값을 유지해야 하는 K-means에 의한 군집 유지성을 의미하게 된다 적용하고자 하는 K-means 클러스터링 알고리즘은 비지도 학습방법 알고리즘으로 제공되는 이러한 속성정보로부터 일반적으로 주어진 데이터를 k 개의 클러스터로 묶는 알고리즘으로 각 클러스터와 거리 차이의 분산을 최소화하는 방식으로 동작한다. 이 알고리즘은 비지도 자율 학습의 일종으로, 레이블이 달려 있지 않은 입력 데이터에 레이블을 달아주는 역할을 수행한다. 본 연구에서는 누적된 이동노드 데이터의 속성을 사용하여 이동 노드 클러스터(C_i)를 구분을 하려고 한다. 주어진 데이터 집합으로부터 이러한 중심점 μ_k 의 값을 결정하고 이때 독립적 속성척도값(Gini Index Property(GIP))을 제공하여 클러스터내에서 중심값을 유지하는 것이다. 위와 같은 문제는 데이터 집합을 특정 클러스터에 할당하는 과정으로 우선 변수 $r_{nk} \in 0, 1$ 을 정의한다.

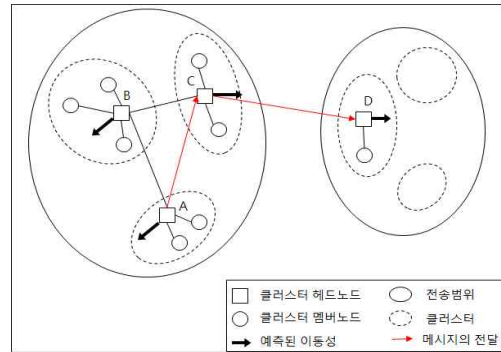


Fig. 1 The Selection of Clustering Center node by the estimation of context vector in mobile clustering nodes according to adaptive repeated learning

이때의 k 값은 $k = 1, \dots, k$ 이다. 이는 어떤 n 번째 샘플 X_n 이 k 번째 이동 노드 클러스터에 속하는 경우 $r_{nk} = 1$ 이고 아닌 경우 0이 된다. 이제 목적 함수를 정의한다. 이를 왜곡 측정(distortion measure) 함수로 식 다음과 같이 정의된다.

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|X_n - \mu_k\|^2$$

이에 따라 구성된 시스템의 흐름도는 그림 2와 같다. 흐름도에서 나타난 것과 같이 우선적으로 이동 노드 군집의 수를 지정하기 위해 시스템에서 사용할 노드의 수를 입력받고 이의 이동 노드에 대한 클러스터링과 중심 속성을 확보하고 유지한다.

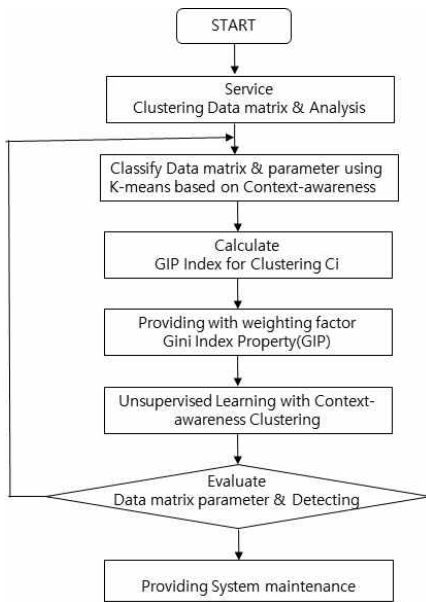


Fig. 2 Flow chart of proposed mobile node context-awareness unsupervised learning system in K-means clustering

Ⅲ. 결 론

논문에서는 상황인식정보를 이용한 이동 노드의 클러스터링내에서 효율적인 군집 알고리즘을 제안하였다. 일반적으로 헤드노드를 중심으로 관리되는 클러스터링은 종속된 그룹데이터의 유사도 측정이 필요하며 이로부터 실질적인 이동 노드의 클러스터링 유지관계가 성립된다. 즉, 입력된 속성 노드의 데이터 분석에 따라 클러스터링된 이동 노드는 어느 한 속성인자에 대해 의 정보가 급격하게 변화하더라도 주어진 GIP와 가중치 분석에 의해 같은 이동 노드 클러스터로 인식이 가능하여 정보의 수정이 필요가 없어 유연하게 반응 할 수 있었다. 본 연구에서는 이러한 연구의 결과를 이동통신망에서 클러스터링을 수행하고, 노드의 중심 속성정보를 이용한 클러스터링 알고리즘 연구결과를 적용한 이동 노드에서의 전달율을 향상시킬 수 있는 연구결과로 활용할 것을 제안한다[4][5].

References

[1] A. Likas, N. Vlassis and J.Verbeek, "The global k-means clustering algorithm," Department of Computer Science, University of Ioannina, 45110 Ioannina, Greece, pp. 452-456, 2003.

[2] L. Xue and W. Luan, "Improved K-means Algorithm in User Behavior Analysis," Dalian Maritime University, pp. 339-341, 2015.

[3] Lindgren, "Probabilistic Routing in Intermittently Connected Networks", ACM SIGMOBILE Mobile Computing and Communications Review, Vol. 7, no. 3, pp. 19-20. 2003.

[4] X. Meng, G. Xu, T. Guo, Y. Yang, W. Shen, and K. Zhao, "A Novel Routing Method for Social Delay-Tolerant Networks", Tsinghua Science and Technology, Volume 24, Number 1, February 2019.