

계층적 군집 패킷 분류기법을 이용한 Cloud 기반 IoT를 위한 큐 할당 기법

이민경⁰, 홍충선*
경희대학교 컴퓨터공학과
{immigrationk, cshong}@khu.ac.kr

The method of Queue allocation for Cloud based IoT using Hierarchical clustering

MinKyung Lee⁰, ChoongSeon Hong*
Department of Computer Science and Engineering, Kyung Hee University

요 약

최근 사물인터넷 패러다임과 함께 수많은 기기들이 인터넷을 기반으로 통신할 수 있는 환경이 발달하고 있다. 그러나 사물인터넷 환경에서의 기기들은 방대한 데이터들을 생성하므로 에너지 효율성 및 서비스의 질 저하 등과 같은 문제와 더불어 분산된 저장소, 프로세싱 수용량 등 데이터 처리과정 관련 문제점들을 지니고 있다. 이를 보완하기 위해 사물인터넷 환경에서 클라우드 컴퓨팅과의 결합이 떠오르는 추세이나, 이 역시 방대한 데이터가 서버로 한 번에 전달됨에 따라 서비스를 제공하는데 있어 요구되는 서비스의 질을 만족하기가 어렵다. 따라서 본 논문은 사물인터넷 센서 기기들에서 발생한 방대한 데이터들이 클라우드 서버로 저장될 시, 서버 상에서 추가된 패킷들을 계층적 군집 알고리즘 기반 특징 별 분류 과정을 거친 후, 형성된 군집을 기반으로 우선순위를 부여하여 높은 우선순위부터 큐에 할당하는 방안을 제안하고자 한다.

1. 서 론

사물인터넷(IoT, Internet of Things)이란, “표준 통신 프로토콜 기반 고유한 주소 지정이 가능한 상호 연결된 객체의 세계적 네트워크” 로[2], 우리 주변에 만연히 존재하는 사물들이 인터넷 기반으로 서로 상호작용 및 협력하며 본연의 역할을 충실히 수행할 수 있도록 하는 새로운 패러다임을 의미한다. 이와 같은 의미를 지닌 사물인터넷은 점차 잠재적 사용자들의 행동과 일상생활에 많은 영향을 끼칠 것으로 예상된다[1]. 그러나 사물인터넷 환경에서 사용 가능한 기기들은 서로 다른 목적을 지닌 다량의 데이터들을 생성함으로 사물인터넷 환경 내 에너지 효율성 측면과 서비스의 질(QoS, Quality of Service)과 관련된 문제들을 야기한다[1]. 특히, 사물인터넷의 경우, 분산되고 제한된 저장 장소의 수용량 및 방대한 데이터의 처리량 한계 등의 단점들을 지니고 있어 지나친 프로세싱 시간이 소모되고 작은 기기들에서 발생하는 네트워크 과부하와 같은 문제들을 지니고 있다[2]. 사물인터넷의 이와 같은 단점들을 보완해 줄 방안으로 클라우드 컴퓨팅(Cloud Computing)이 각광받고 있다.

클라우드 컴퓨팅이란, 가상 컴퓨팅 서버로, 저장소와 프로세싱 전력에 대해 제한 없는 수용량을 지니고 있다[2]. 이와 같은 클라우드 컴퓨팅의 특징은 사물인터넷의

단점을 보완해준다는 점에서 클라우드 기반 사물인터넷 환경이 급부상하는 추세이다. 그러나 발생하는 데이터량이 방대함에 따라 클라우드 컴퓨팅 기반 사물인터넷 환경에서도 데이터를 클라우드 서버에 저장할 시 다량의 데이터가 한 번에 서버로 추가되기 때문에 과도한 프로세싱 시간 및 네트워크 과부하의 원인으로 인해 서비스의 질(QoS) 저하가 발생하는 단점이 있다[2]. 이에 본 논문에서는 사물인터넷 센서 기기들로부터 발생하는 다량의 데이터들이 클라우드 서버로 저장될 시, 클라우드 서버 상에서 유입된 패킷들을 응집형 계층적 군집 알고리즘을 이용하여 패킷 특징에 알맞게 군집화한 후, 다량의 패킷을 포함하는 군집들에 높은 우선순위를 부과하여 큐에 저장하는, 패킷 우선순위 기반 큐 할당 방안을 제안하고자 한다.

2. 관련 연구

2.1 계층적 군집 알고리즘(Hierarchical clustering)

계층적 군집 알고리즘이란, 비슷한 군집끼리 군을 이뤄가며 최종적으로 하나의 군집을 이룰 때까지 묶어나가는 클러스터링 알고리즘을 의미한다. K-평균 알고리즘과는 다르게 군집 수를 사전에 정의하지 않아도 되며, 군집 간의 거리를 기반으로 군집화를 진행하는 알고리즘이다[3]

이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신 기술진흥센터의 지원을 받아 수행된 연구임 (No.2015-0-00557, IoT 기기의 물리적 속성, 관계, 역할 기반 Resilient/Fault-Tolerant 자율 네트워크 기술 연구) 본 논문은 산업 통상 자원부 산업 핵심 기술개발사업으로 지원된 연구결과입니다[10049079, 퍼스널 빅데이터를 활용한 마인딩 마인즈 핵심 기술 개발]

*Dr. CS Hong is the corresponding author

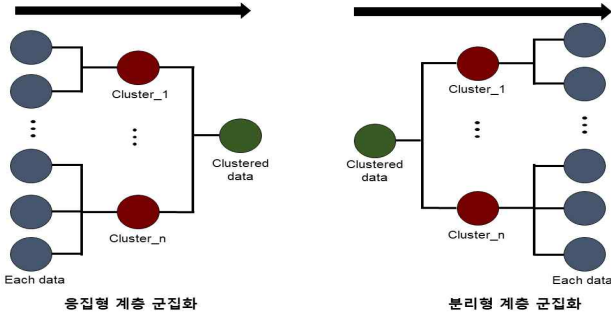


그림 1. 계층적 군집 알고리즘 모델

계층적 군집 알고리즘은 그림 1과 같이 응집형과 분리형 모델로 구분된다. 응집형 계층 군집화의 경우, 주어진 데이터 각각을 독립된 군집으로 가정하고, 병합하며 상위단계의 군집을 구성해 나간다. 최종적으로 데이터 전체를 멤버로 하는 하나의 군집을 구성하는 모델이다. 이와 반대로 분리형 계층적 군집은 데이터 전체를 멤버로 하는 하나의 군집에서 개별 데이터로 분리해가며 군집을 구성하는 모델이다[3]

2.2 적응형 우선순위 큐 기반 스케줄링 기법

스케줄링(Scheduling)이란, 멀티 프로세스 기반 운영체제에서 CPU를 프로세스에 골고루 할당해주는 기법이다. 이러한 기법에 대해서 [5]에서는 적응형 우선순위 큐 기반의 작업 할당 방법을 제안하였다.

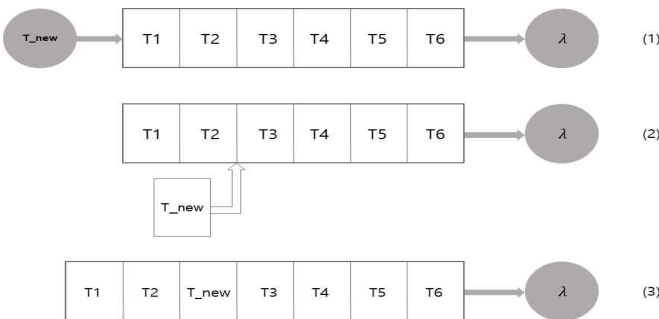


그림 2. 적응형 우선순위 큐 할당 방법

[5]에서 제안한 기법은 적응형 우선순위 큐에 존재하는 데이터들의 데드라인과 센서에서 새로 추가 발생된 데이터의 데드라인을 비교하여 데드라인이 짧을수록 우선순위가 높게 할당된다. 그림 2는 [5]에서 제안한 적응형 우선순위 큐 기반의 스케줄링 제안에서 적응형 우선순위 큐 기반으로 큐를 할당하는 방법을 보여준다.

3. 제안 사항

본 연구는 계층적 군집 알고리즘을 통해, 새로 유입되는 센서 데이터들을 특징별로 군집화한 후, 이를 기반으로 다량의 데이터들이 군집화 된 군집에 높은 우선순위를 부여함으로써 큐를 할당하는 방안을 제안한다.

응집형 계층적 군집 알고리즘을 통해, 사물인터넷 환경 내 다양한 센서 기기들로부터 방대한 데이터들이 유입할 시, 각각의 데이터들을 군집으로 간주하며, 점차적으로 넓은 범위의 군집화를 진행한다. 최종적으로 하나

의 군집을 형성하는 응집형 계층적 군집 알고리즘과는 다르게 군집이 총 3개 정도 최종적으로 형성되도록 제한을 둔다.

$$d(u, v) = \| c_u - c_v \|_2 \quad (1)$$

이때 형성되는 클러스터 간의 유사도를 측정하기 위해 수식 (1)에 명시되어 있는 클러스터 간의 중심점을 정의한 후, 각각 클러스터간 중심점의 거리로 유사도 값을 구한다.

표 1. 클러스터 간의 유사도 측정 변수 및 의미

변수	의미
u, v	형성된 클러스터
c_u	클러스터 u의 중심점
c_v	클러스터 v의 중심점
$d(u, v)$	클러스터 u와 v의 거리

알고리즘 Process for each packet ω

- 1: D(i) ← flowed data from IoT sensor devices
- 2: C(j) ← Cluster from D(i)
- 3: d ← distances between clusters
- 4: FC(k) ← number(k) of Final clusters
- 5: Q(n) ← nth Queues
- 6: F ← Features of each sensor data sets
- 7: P ← Parameter
- 8: IF C(j) is made by D(i) from IoT sensor devices
- 9: calculate d(distances) between them using Centroid
- 10: IF d is calculated
- 11: Making clusters again using features F
- 12: FOR clustering each data set UNTIL 3 clusters
- 13: IF finish 3 clusters
- 14: use priority-based queue allocation
- 15: ELSE IF lake of number of clusters
- 16: use centroid method to calculate each clusters d

위에 명시된 알고리즘을 토대로 사물인터넷 센서 디바이스에서 클라우드 서버로 유입된 데이터들은 데이터가 지닌 특징들을 기반으로 클러스터를 형성한다. 이 때, 형성된 클러스터 간 중심거리 기반 유사도 측정 수식(1)을 사용하여 각각의 특징별 거리를 구한다. 총 3개의 클러스터가 형성될 때까지 계층적 군집 알고리즘을 반복하여 수행한다.

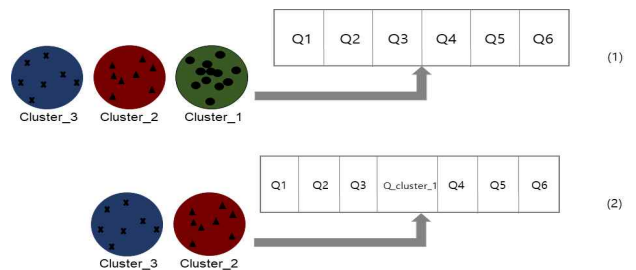


그림 3. 클러스터 군집 패킷량 우선순위 큐 할당

알고리즘 과정이 완료하여 군집 3개가 형성된 후, 각각의 군집의 크기를 비교한다. 그 중 가장 큰 규모의 분포도를 형성하고 있는 군집을 큐에 가장 높은 우선순위로 배치하는 과정을 수행한다.

그림 3은 사물인터넷 센서 기기들에서 발생한 데이터들 각각의 특징별로 계층적 군집 알고리즘을 수행하여 포함된 패킷 분포도가 가장 넓은 군집에 높은 우선순위를 부과하고, 이를 토대로 순차적으로 군집들을 클라우드 서버 큐에 할당하게 된다.

4. 성능 평가

4.1 시뮬레이션 환경

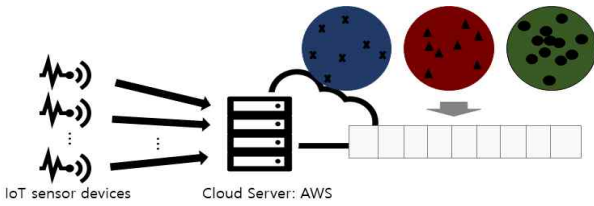


그림 4. 시스템 모델

본 논문의 연구를 수행하기 위해 crowdad 사이트에서 darthmouth/cenceme 센서 데이터 셋을 가상의 클라우드 서버에 전송하는 환경을 구축하였다[4] 그림 4는 본 논문이 제안한 사항들의 전반적인 시스템 모델을 나타낸다.

아마존 클라우드 서버인 AWS 시스템을 토대로 클라우드 서버를 구축한 후, crowdad 사이트로 받은 사물 인터넷 기반 데이터셋을 구축된 서버로 넘기면, 서버에서 본 논문이 제안한 응집형 계층적 군집 알고리즘 및 이를 기반으로 구축한 서버 내 큐 할당을 수행한다.

4.2 성능 평가

본 장에서는 4.1에서 제시한 그림 4의 시스템 모델을 바탕으로 구축된 환경 내, 본 논문이 제안한 깊은 학습 알고리즘을 통하여 군집화 한 결과 및 큐 할당 여부를 평가하였다.

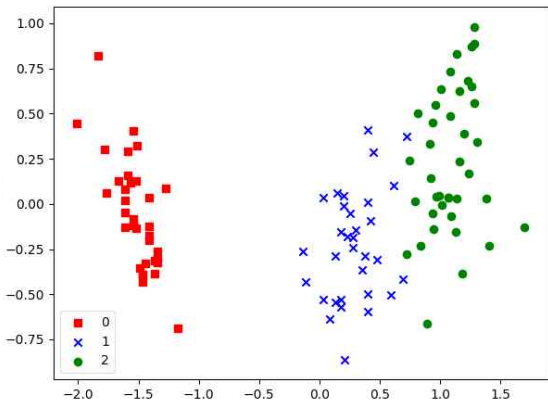


그림 5. 계층적 군집화 결과

그림 5는 구축한 AWS 클라우드 서버 상에서 사물인터넷 센서 기기들을 통해 유입된 데이터 셋을 본 논문이 제안한 계층적 군집 알고리즘을 토대로 분류한 결과 값을 보여준다. 이 결과 값을 토대로 우선순위를 부여하여 클라우드 서버 내 큐를 할당하였다. 그러나 군집 알고리

즘 결과 상 각각 군집 내 분포하고 있는 패킷들의 수가 비슷함에 따라 군집 분포도 역시 비슷함을 알 수 있었다. 따라서 큐 할당을 위한 우선순위를 부여할 시, 그 차이가 매우 미세하다는 것을 확인할 수 있었다.

5. 결론 및 향후 연구

본 논문에서는 클라우드 기반의 사물인터넷 환경이 구축되었으나, 사물인터넷 센서 기기들에서 클라우드 서버로 유입되는 방대한 데이터 량을 각각의 특징별로 군집화 한 후, 군집 분포도를 토대로 높은 우선순위를 부여하여 서버의 큐를 할당하는 방안을 제안하였다. 4장에서는 성능평가를 통해 본 논문이 제안한 사항 중 데이터 군집화 정도와 큐 할당의 여부를 평가하였다. 그러나 본 논문에서 사용한 The dartmouth/cenceme 데이터셋이 최근이 아닌 2008년도 데이터 셋이었다는 한계와 제안한 군집 알고리즘을 사용했을 시, 명확한 분류평면의 결과 값을 도출할 수 없었다는 한계가 존재한다.

따라서 향후 연구에는 2015년도 이후 형성된 데이터 셋 구축 혹은 직접 연구실 혹은 가정에 존재하는 인터넷 기반 센서 데이터들로부터 데이터 셋을 구축하여 각각 특징별로 군집화를 진행하고자 한다. 더불어, 군집화시 계층적 군집 알고리즘과 밀도 기반 군집 알고리즘을 비교 분석하여 더 높은 정확도를 나타내는 알고리즘이 무엇인지 연구하며, 이에 알맞은 최적화 방안 부분 연구를 진행하고자 한다.

참고 문헌

- [1] Luigi Atzori et al, "The Internet of Things: A survey", 2010 Computer Networks, Elsevier journal
- [2] Alessio Botta et al, "Integration of Cloud computing and Internet of Things: A survey", 2016 Future Generation Computer Systems, Elsevier journal
- [3] Stephen C. Johnson, "Hierarchical clustering schemes", September 1967, Psychometrika Volume 32, Issue 3, pp241-254
- [4] Mirco Musolesi et al, "Dataset of sensor data collected by the CenceMe system", The dartmouth/cenceme dataset(v. 2008-08-13), available on: <https://crowdad.org/dartmouth/cenceme/20080813/>
- [5] 이미진 외, "IoT 환경에서 센서데이터의 처리를 위한 적응형 우선순위 큐 기반의 작업 할당 방법", 한국경영과학회 학술대회 논문집, 2016.4, 6276-6278