

K-means 기반 사물인터넷 서비스 분류 기법

양찬우, 조정훈, 이대원
서경대학교 컴퓨터공학과
e-mail : {ycw0522, jojeong3766, daelee}@skuniv.ac.kr

An Internet of Things (IoT) Service Clustering Method based on K-means Algorithm *

Chanwoo Yang, Jeonghoon Jo, Daewon Lee
Dept. of Computer Engineering, SeoKyeong University

요 약

4 차산업 혁명을 맞이하여 다양한 사물 인터넷(IoT) 서비스가 폭발적으로 등장하고 있다. 현재의 IoT 서비스는 독립 서비스로 제공되는 상황이지만 향후 IoT 서비스는 기존 IoT 서비스의 활용과 결합을 목표로 개발되고 있다. IoT 서비스 간 결합 시 발생할 수 모듈의 중복성 문제를 해결하고 새로운 IoT 서비스의 이식성을 높이기 위해 본 연구에서는 K-means 알고리즘을 활용하여 IoT 서비스 간 유사도를 고려한 IoT 서비스 분류 알고리즘을 제안한다. 실험 및 분석을 통하여 K=8,9 인 경우 37 개의 상용 IoT 서비스가 효율적이고 적합하게 클러스터됨을 증명하였다.

* 이 논문은 2016 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2016R1C1B1008330)

1. 서론

최근 센서 장치 및 통신 기술의 발달에 기반하여 다양한 사물 인터넷(IoT) 서비스가 등장하였다. 폭발적으로 증가하는 IoT 서비스는 다양한 센서 장치간의 자율적 통신 및 분석을 통하여 사용자에게 적합한 서비스를 제공함을 목표로 한다. 현재의 IoT 서비스는 단일 서비스로 제공되고 있지만 기존 IoT 서비스들을 기반 및 결합한 다양한 IoT 서비스들이 개발되고 있다. IoT 서비스들의 결합과 기기종 센서 기기간의 데이터 통합 등에서 발생하는 문제점들을 해결하고 효율적인 관리를 위한 통합 IoT 서비스 플랫폼의 필요성이 증대되었고 그에 관한 다양한 연구가 진행되었다[1,2]. [3]에서는 통합 IoT 플랫폼 환경에서 IoT 서비스 모듈의 재사용성 및 이식성 향상을 위하여 IoT 서비스의 동작 기반 분류 알고리즘을 제시하였다.

본 연구는 [3]의 후속 연구로 동작 기반 분류 기법에 K-means 알고리즘을 적용하여 IoT 서비스 간 유사도를 고려한 분류 알고리즘으로 확장한다. IoT 서비스 간 유사도를 고려한 클러스터링을 통하여 모듈의 재사용성 향상과 새로운 IoT 서비스의 이식성을 향상시킨다.

2. 관련 연구

클러스터링이란 데이터 셋이 되는 데이터 그룹(클러스터)들의 대표점을 찾고 데이터들의 특성을 고려해 클러스터를 정의하는 데이터 마이닝의 한 방법이다. 대표적인 클러스터링 알고리즘으로는 유사도가 높은 두개의 그룹들을 합쳐 하나의 그룹으로 만들어 그 수를 줄여 나가는 응집 계층 알고리즘, 인간의 두뇌 구조를 모델링 한 신경망 기반 SOM 알고리즘 확률을 사용해 적합도를 평가해 군집화 하는 EM 알고리즘 등이 있다[4,5].

본 연구에서 사용할 K-means 알고리즘은 각 인스턴스를 유사도가 높은 K 개 그룹으로 군집화 하는 방법이다. 각 군집의 인스턴스 들의 평균값을 중심으로 주변의 인스턴스들과 유클리드 거리(Euclidean distance)를 계산하여 군집화 한다. <표 1>은 K-means 알고리즘이다.

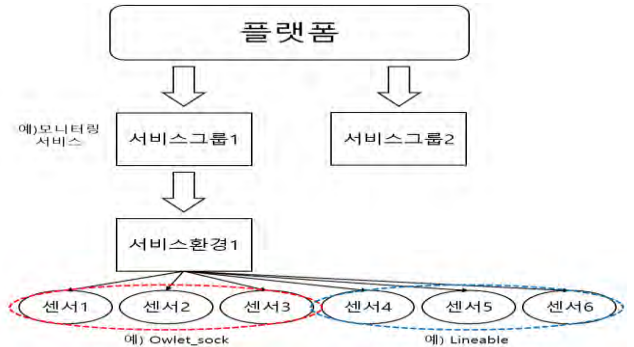
<표 1> K-means 알고리즘

```

Algorithm K-means :
Choose random k data points as initial Clusters Mean (cluster centers)
Repeat
  for each data point x from D
    compute the distance between x and each cluster mean(centroid)
    assign x to the nearest cluster
  endfor
Re-compute the mean for current cluster collections
until reaching stable clusters(current clusters means equals last clusters means)
    
```

3. 시스템 환경

본 연구의 시스템 환경은 데이터 수집 장치인 센서, 센서와 동작 장치들로 구성되어 단일 서비스를 제공하는 서비스 환경, 동일 서비스 환경에서 다양한 서비스로 구성된 서비스 그룹, 다양한 서비스 그룹을 관리하는 플랫폼으로 구성된다. 본 연구에서 기반하는 시스템 환경은 (그림 1)과 같다.



(그림 1) 시스템 환경

본 연구에서는 현재 상용으로 제공되는 IoT 서비스 37 개의 정보를 직접 수집하였고 초기 클러스터의 중심 값 결정을 위하여 [3]의 동작 기반 분류 알고리즘을 이용하여 17 개의 클러스터로 클러스터링 했다. 이를 기반으로 중심 값과 인스턴스 간의 유사도를 평가하여 클러스터의 개수를 결정한다. <표 2>는 동작 기반 분류 알고리즘의 클러스터 중 한 예인 헬스케어 서비스 클러스터이다.

<표 2> 헬스케어 서비스 클러스터

Group ID	0x88481080
Sensing	Self-power, Non-IP
Data management	Data preprocessing(server), Non volatile
Processing_type	Average, Threshold

Service_name	Explanation	Sensor specs	Operation Steps
Fitbit Charge HR	Activity Wristband	PurePulse, All Day Activity, SmartTrack, Auto Sleep, Bluetooth	sensing -> analysis -> report
Dexcom Sevenplus	Glucose monitoring	Glucose, 360 degrees of sensor	sensing -> analysis -> report
Lively	Medical Alert Watch	Lively passive activity Sensors	sensing -> analysis -> report

4. K-means 기반 서비스 분류 알고리즘

본 연구에서는 다양한 IoT 서비스에 대하여 유사도 기반 클러스터링을 하기 위하여 알고리즘에 K-means 알고리즘을 적용하여 [3]의 동작 기반 분류 기법을 확장한다. 동작 기반 분류 알고리즘 적용 결과 37 개의 서비스는 17 개의 클러스터로 분류되었다. 본 연구에서는 이 분류 결과를 기반으로 1 이하의 인스턴스를 갖는 클러스터를 배제하고 유사도를 기반으로 적합한 클러스터의 수를 결정하고 클러스터를 구성한다. <표 3>은 본 연구에서 제안하는 K-means 기반 서비스 분류 알고리즘이다.

<표 3> K-means 기반 서비스 분류 알고리즘

