

A Framework for Internet of Things (IoT) Data Management

Kyung-Chang Kim*

Abstract

The collection and manipulation of Internet of Things (IoT) data is increasing at a fast pace and its importance is recognized in every sector of our society. For efficient utilization of IoT data, the vast and varied IoT data needs to be reliable and meaningful. In this paper, we propose an IoT framework to realize this need. The IoT framework is based on a four layer IoT architecture onto which context aware computing technology is applied. If the collected IoT data is unreliable it cannot be used for its intended purpose and the whole service using the data must be abandoned. In this paper, we include techniques to remove uncertainty in the early stage of IoT data capture and collection resulting in reliable data. Since the data coming out of the various IoT devices have different formats, it is important to convert them into a standard format before further processing. We propose the RDF format to be the standard format for all IoT data. In addition, it is not feasible to process all captured IoT data from the sensor devices. In order to decide which data to process and understand, we propose to use contexts and reasoning based on these contexts. For reasoning, we propose to use standard AI and statistical techniques. We also propose an experiment environment that can be used to develop an IoT application to realize the IoT framework.

▶ Keyword: IoT, RDF, Context, Framework, Context Aware Computing, Semantic Web

I. Introduction

사물인터넷(IoT)은 최근 학계와 산업 전반에서 급속히 발전하고 있는 유비쿼터스 기술을 통합하고 새롭게 확장하기 위한 개념이다. 이는 센싱 기술과 유무선 통신 기술을 활용하여 실제 세계에 존재하는 공간, 사물, 사람, 다양한 정보 등의 모든 사물들이 인터넷을 통해 서로 유기적으로 연결되어 실시간 정보를 서로 의사소통하는 기술 및 서비스를 말한다. IoT는 유/무선, LAN, Ethernet 등의 통신망 접근 방법 등에 비해 훨씬 많은 시설과 설비를 제공하기 때문에 현재의 컴퓨팅 시대에 주요 연구 관심 분야로 자리 잡고 있다. IoT는 우리 주위에 있는 모든 스마트 객체들이 서로 연결되고 인간의 간섭이 거의 없이 서로 즉흥적으로 통신할 수 있는 세상을 약속하고 있다. IoT의 최종 목적은 우리 인간들을 위해 보다 좋은 세상을 만들어 주고자 하는 것이다. IoT는 컴퓨팅의 새로운 떠오르는 모델로서 물리적인 객체와 가상적인 객체들을 즉흥적으로 서로 연결해 준다.

센서 기술의 발달로 인하여 센서들은 더욱 강력해지고, 가격이 점점 인하되며 크기가 작아지는 추세이다. 오늘날 수많은 다양한 센서 장치들이 인터넷에 연결되고 있으며 이들 감지 장치들에 의하여 엄청난 양의 빅 데이터가 발생되고 있다. 따라서 수집한 데이터들을 분석하고, 해석하고 이해하지 못하면 이들 데이터들은 가치 있고 의미 있는 정보를 제공해주지 못할 것이다. 상황 인식 컴퓨팅은 센서 데이터들과 연결된 상황 정보의 저장을 가능하게 하므로 센서 데이터들에 대한 해석을 좀 더 의미 있고 쉽게 할 수 있게 한다. 또한 상황은 IoT 환경의 핵심 요소인 기계 대 기계의 통신을 보다 쉽게 수행하게 한다.

IoT 응용은 네트워크를 통해 사물인터넷 장치들로부터 여러 종류의 데이터를 감지함으로써 사용자에게 유용한 서비스를 제공한다. IoT 장치는 원자력 발전소 혹은 지하 터널과 같은 사람이 접근하기 어려운 곳이거나 위험한 환경에서 유용하게 사용

*First Author: Kyung-Chang Kim, Corresponding Author: Kyung-Chang Kim
*Kyung-Chang Kim (kckim@hongik.ac.kr), Dept. of Computer Engineering, Hongik University
• Received: 2019. 02. 14, Revised: 2019. 03. 26, Accepted: 2019. 03. 26.
• This work was supported by Hongik Univ. Research Grant (2017).

되고 있다. IoT 장치들로부터 수집되는 데이터는 실시간성이며, 스트림 데이터의 가능성이 높고, 작은 크기의 중복 데이터가 지속적으로 저장될 수 있고, 반대로 기기들로부터 일정 시간에 집중적으로 수집되고, 일정시간 수집되지 않을 수도 있다 [6]. IoT 기기들은 각 기기들의 특성과 여러 환경요소에 따라 민감하게 영향을 받기 때문에 기기들로부터 수집된 데이터에 대한 신뢰성 문제는 선행적으로 처리되어야 한다. 수집된 데이터의 불확실성을 제거하여 신뢰도가 향상되면 정확한 분석과 품질 저하 등의 문제를 예방할 수 있고 효율적인 IoT 응용 개발이 가능하다.

시맨틱 웹 기술은 컴퓨터가 정보를 제공하기 위해 웹상의 정보를 탐색 및 수집하여 논리적으로 추론하는 정보처리 기능이다. 즉, 컴퓨터가 정보를 이해하고 다양한 정보들 간 의미요소를 연결함으로써 지능적 판단에 따라 정보를 추출, 가공하는 처리방식이다. 기존의 정보는 인간이 이용하기 위해서만 처리되었기 때문에 기계는 인식할 수 없었다. 시맨틱 웹은 이러한 정보들에 키워드를 표시해 기계도 이해 가능하도록 만든 것이다. IoT 데이터는 다양한 센서에서 발생하는 숫자 형 데이터뿐만 아니라 여러 기기에서 다양한 형태의 데이터가 발생한다. 물론, 현재의 사물인터넷 분야는 아직 작은 단위의 관리지만, 앞으로 점점 그 단위가 커질 것이고 각 데이터가 어떤 센서에서 무엇을 측정했는지 값을 명확히 표현할 필요가 있다. 따라서 앞으로 지속적으로 만들어질 사물인터넷 데이터의 공유, 개발, 공동 활용을 위해서는 시맨틱 기술을 이용하는 것이 효과적이다.

앞으로의 IoT 플랫폼 서비스들은 사물이 스스로 상황을 인식하여 행동을 하는 사물인터넷의 최종 목표로 향해 갈 것이다. 최종적으로 다양한 사용자들에게 수집된 IoT 데이터에 대해 콘텐츠 서비스를 제공하기 위해서는 IoT 데이터의 활용이 매우 중요하다. IoT 데이터를 활용하기 위해서는 시맨틱 웹 기술을 이용하여 고급 데이터(즉 상황 정보)로 변환하고 저장한다. 그 다음의 숙제는 수집된 엄청난 데이터들(즉 빅 데이터)을 이해하고 의미를 두기 위하여 추론하는 것이다. 위의 연구를 통해 제안한 기법들을 기반으로 신뢰성 높은 IoT 플랫폼을 구축하고 IoT 데이터의 활용을 증명해 보일 수 있는 실험 환경도 제안하고자 한다.

II. Preliminaries

1. Related works

불확실성 처리에 대한 연구는 과거 상황인식 컴퓨팅 환경에서 상황의 불확실성을 다루기 위한 몇 가지 접근방법이 연구되었다[3,4,5]. Qoc(Quality of Context) 기법은 상황의 품질에 대한 개념을 제안하였고, 상황정보가 실제계 현상을 표현하는 정도를 정밀도(precision)로 정의한 후에 어떤 상황이 결정되는 특정 시점에, 상황 인스턴스에 대응되는 실제계 현상을 정확하

게 표현하는 확률을 가지고 정확성을 정의하였다. 또, 정확성과 완전성이라는 두 가지 단위를 기반으로 상황의 품질을 측정하는 방법을 제안하였고, 이러한 상황요소들은 상황인식 어플리케이션에 의해 사용될 수 있다.

또 다른 불확실성을 다루는 계량기법은 베이지안 네트워크이다. 베이지안 네트워크를 사용하여 상황의 불확실성을 처리하는 사용자의 행위를 측정하였으며 확률 값과 관계링크를 통해 상황을 정의하기 위해 온톨로지를 이용한 연구들이 있었다. 비슷한 연구에서는 베이지안 네트워크와 온톨로지를 사용하였지만 상황 온톨로지 정의를 재사용하는데 초점을 두었다[2]. 하지만 위의 연구들에서 시스템 도메인에 적합한 온톨로지를 생성하기 위해서는 전문가의 도움이 필요할 뿐 아니라 시스템 개발자들이 해야 할 일이 늘어난다. 또한, 온톨로지를 베이지안 네트워크로 전환하는 일은 자동화되지 않아서 매우 힘든 작업이 될 수 있다.

최근 사물인터넷 환경에서 발생하는 센서 데이터의 가치와 데이터의 상호 운용성을 증진시키기 위해 시맨틱 웹 기술과의 접목에 대한 연구가 활발히 진행되고 있다[11]. 센서 데이터와 서비스 도메인 지식의 융합을 위한 센서 데이터의 시맨틱화에 대한 연구에서 실시간성, 대용량성의 특징이 있는 사물인터넷 환경에서는 기존과 다른 시맨틱 변환 기술이 필요하다는 것을 밝히고 변환 규칙 및 온톨로지 기반 센서 모델을 통해 실시간 병렬 시맨틱 변환에 대한 연구와 이를 빅데이터 프레임워크인 아파치 스톰을 이용하여 실험하여 발표되었다.

IoT는 지난 몇 년 동안 상당한 관심을 받았다[8,9,10]. 센서 하드웨어 기술의 발달과 저렴한 재료의 공급으로 센서들을 우리 주변에 있는 모든 객체에 붙이고 최소한의 인간의 개입으로 서로 통신할 수 있게 될 것이다. IoT가 당면한 가장 중요한 과제 중의 하나가 센서 데이터를 이해하는 것이다. 이 비전은 여러 정부, 이익 단체, 기업 그리고 연구소에서 지원하고 투자되고 있다. 예를 들면, 유럽 연합(EU)에서 지원하는 CERP-IoT 라는 IoT 연구 프로젝트를 관장한 유럽 연합 기구는 상황 인식을 IoT에서 중요한 연구라고 판명하였다. EU는 IoT에 초점을 둔 상황 인식 컴퓨팅의 연구 개발을 2015-2020 사이에 진행하는 시간대를 배정하였다.

상황 인식 컴퓨팅의 많은 도전과 과제를 해결하기 위하여 지금까지 연구자들이 제안한 시스템, 미들웨어, 응용, 기법 그리고 모델 등 많은 해결책들이 있다. 그 결과 IoT 패러다임에서 상황 인식의 중요성이 확실하게 들어났다.

III. The Proposed Scheme

향후 인터넷은 인간이 생성하고 가공한 데이터, 정보 및 지식뿐만 아니라, 우리 주변의 센싱과 구동체 기능을 갖는 일상 사물로부터 정보를 획득할 수 있다. 그리고 이들 사물은 자체

식별자를 갖고 컴퓨팅과 통신 기능을 통해 세상과 상호작용할 수 있을 것으로 예상된다. 특히 다양한 사물들에 센싱 기술과 통신기술을 접목하여 이용자들 간의 실제 생활과 밀접하게 관련된 정보를 수집하고, 수집된 정보를 사물 간 직접적인 정보 교환과 공유를 통해 기존의 독자적인 서비스들과 결합하여 새로운 형태의 서비스로 재창조하는 것이 사물인터넷이 추구하는 핵심 방향이다. 이러한 서비스들의 중요한 이슈는 신뢰성이다.

신뢰성이 높은 시스템을 만들기 위해서는 10개의 올바른 결과보다 하나의 문제가 되는 것을 없애는 것이 더 중요하다. 본 논문에서는 이러한 불확실성을 해결하여 신뢰성 높은 시스템을 개발하기 위한 내용을 제안한다. 불확실성에 대한 다양한 불확실성의 문제들을 사물인터넷 아키텍처의 일부 레이어 별 처리를 통해 최소화하여 신뢰성 높이고자 한다.

본 논문에서 제안하는 IoT 프레임워크에서 고려하는 시스템 구조는 그림 1과 같이 기존 연구에서 이미 사용하고 있는 4개 층(layer)으로 구성된다[10]. 4개 layer는 Object Sensing Layer, Data Exchange Layer, Information Integration Layer 그리고 Application Service Layer로 구성된다. Object Sensing Layer는 물리적 개체에 대한 감지와 데이터를 수집하는 레이어 이고, Data Exchange Layer는 데이터의 투명한 전송을 담당한다. Information Integration Layer는 네트워크로부터 얻어진 데이터(불확실한 데이터 포함)의 재결합, 정제 및 합성을 담당한다. 또한 불확실한 정보를 사용 가능한 정보로 통합한다. 마지막으로 Application Service Layer는 다양한 사용자들을 위해 콘텐츠 서비스를 제공한다.

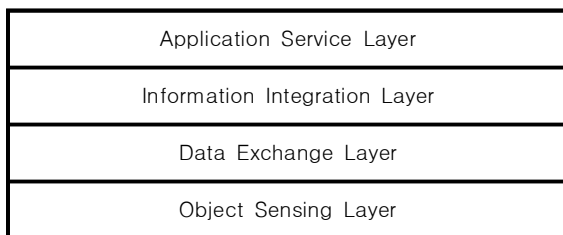


Fig. 1. Four-layer architecture of IoT

IoT 프레임워크에서 상황 정보 컴퓨팅의 적용과 다양한 연구들을 제안하기에 앞서 상황에 대한 생명주기를 살펴볼 필요가 있다. 상황 생명주기는 네 가지 단계로 구성된다. 첫 번째 단계는 상황 수집(context acquisition) 단계로 다양한 물리적인 센서들과 가상적인 센서 소스들로부터 상황을 얻을 필요가 있다. 두 번째 단계는 상황 모델링(context modelling) 단계로서 수집된 데이터를 모델링하고 의미 있는 형태로 표현한다. 세 번째 단계는 상황 추론(context reasoning) 단계로 주어진 모델링한 데이터를 가공하여 저급의 센서 데이터에서 고급의 상황 정보를 얻는다. 마지막 네 번째 단계는 상황 보급(context dissemination) 단계로 상황에 관심 있는 소비자에게 저급 및 고급 상황을 보급할 필요가 있다.

따라서 상황 인식 컴퓨팅을 IoT 프레임워크에 적용하기 위

해서는 상황 수집 단계를 object sensing layer와, 상황 모델링 단계를 data exchange layer, 상황 추론 단계를 information integration layer, 그리고 상황 보급 단계를 application service layer와 mapping 하면 될 것이다.

1. Data Capture in Object Sensing Layer

1.1. Context Collection

상황 수집에 사용되는 기법들은 다양하게 책임, 빈도수, 상황 소스, 센서 타입 그리고 수집 절차에 기반 한다.

(1) 책임에 기반

상황(예 센서 데이터) 수집은 push 와 pull 방법으로 이행한다. Pull 방식은 센서들로부터 센서 데이터를 수집하는 소프트웨어 컴포넌트가 센서 하드웨어로부터 즉시 혹은 주기적으로 데이터를 요청한다. Push 방식은 물리적 혹은 가상적 센서에서 데이터를 센서 컴포넌트로 보낸다.

(2) 빈도수에 기반

상황은 즉시(instant) 이벤트와 간격(interval) 이벤트에 기반 해서 수집될 수 있다. 즉시 이벤트는 즉시 이루어진다. 즉시 이벤트의 종류로는 문 열기, 불 켜기 등이다. 즉시 이벤트를 검출하기 위해서는 이벤트가 발생할 때 센서 데이터를 수집하는데 이때 push 및 pull 방법을 사용할 수 있다. 간격 이벤트는 일정 시간 격차를 두는 이벤트이다. 간격 이벤트의 종류로는 비가 내림, 동물이 식물을 먹다, 겨울 등이다. 이러한 이벤트를 검출하기 위해서는 센서 데이터를 주기적으로(예를 들면, 매 20 초마다 센싱하고 센서 데이터를 소프트웨어로 보내기) 수집할 필요가 있다. 이때에도 push 와 pull 방법을 사용할 수 있다.

(3) 소스에 기반

또한 상황 수집 방법은 상황이 어떤 소스에서 왔는지에 기반 하여 분류할 수 있다. 소스가 센서 하드웨어이면 상황은 센서 하드웨어 및 관련 API와 통신을 통해 센서로부터 직접 수집할 수 있다. 소스가 미들웨어 인프라이면 상황(즉 센서 데이터)은 센서 하드웨어가 아닌 GSN과 같은 미들웨어 솔루션에 의해 수집될 수 있다. 소스가 상황 서버이면 상황은 웹 서비스 호출과 같은 다른 메커니즘을 통해 여러 상황 스토리지(예를 들면 데이터베이스, RSS, 웹 서비스)로부터 수집될 수 있다.

(4) 센서 타입에 기반

상황을 수집하는데 여러 종류의 센서들을 사용할 수 있다. 센서들은 물리적, 가상적 그리고 논리적인 센서로 나눌 수 있다. 물리적 센서는 가장 많이 사용되는 센서로서 그들 스스로 센서 데이터를 생성할 수 있다. 물리적 센서로부터 검색된 데이터를 저급 상황이라 부르는데 의미가 거의 없고 단순하며 조금의 변화에도 취약하다. 가상적 센서는 그들 스스로 센서 데이터를 생성하지 않고 여러 소스로부터 데이터를 검색하고 그들을 센서 데이터라고 공개한다.

이들은 데이터를 주고받고 하는데 일반적으로 웹 서비스 기술을 이용한다. 논리적 센서는 보다 의미있는 정보를 생성하기 위하여 물리적 센서와 가상적 센서를 혼합한다. 기상 정보를 제공하는 웹 서비스를 논리적 센서라 할 수 있는데 기상 정보를 수집하기 위하여 몇 천개의 물리적 센서를 사용하고 또한 지도, 달력, 역사적 데이터와 같은 가상적 센서로부터 정보를 수집한다.

(5) 수집 절차에 기반

상황을 수집하는 방법으로 센스(sense), 유도(derive), 그리고 수동 제공(manually provided)이 있다. 센스에서는 센서로부터 데이터를 센스 하는데 데이터베이스에 저장된 센스 데이터를 포함한다. 유도에서는 센서 데이터에 대해 컴퓨팅 연산을 수행하여 정보를 생성한다. 연산중에는 웹 서비스 호출이나 센스된 데이터에 대해 수학적 함수를 적용할 수 있다. 수동 제공에서는 사용자가 선호도와 같은 미리 정의된 세팅 옵션을 통해 상황 정보를 수동적으로 제공한다.

1.2. Uncertainty Processing

최근 IoT 기반의 다양한 스마트 환경에서 상황(context)에 적합한 맞춤형 서비스를 제공하기 위해서는 동적으로 변화하는 상황 정보가 필요하다. 특히 사물인터넷 기반의 지능적인 서비스를 제공하기 위해서는 사용자와 환경의 상황을 정확하게 파악할 필요가 있으며, 이를 위해서는 상황 인식 기술이 필수적이다. 여기서 상황은 공간에 설치된 센서로부터 가져올 수 있는 사용자 정보와 관련된 환경 정보 즉, 사용자가 처한 환경에 대한 정보를 말한다. 이러한 상황을 언어내는 과정을 상황 인식(context awareness)이라고 한다.

상황인식 컴퓨팅에서 불확실성이 발생하는 경우는 크게 두 가지로써 하나는 센서로부터, 다른 하나는 상황인지를 위한 모델 설계 및 추론에서 비롯된다.

상황정보는 센서 장치 또는 컴퓨팅 장치로부터 수집되는 하위 수준의 상황정보(Low-Level Context)와 상위 수준의 상황정보를 수집하여 추론 등의 다양한 프로세싱을 통하여 생성되는 상위 수준의 상황정보(High-Level Context)의 계층적 구조로 분류 된다. 하위 수준 상황정보는 센서로부터 직접 추론된다. 센서는 센싱하는 하드웨어를 포함한 상황정보를 제공하는 모든 데이터 소스를 의미한다.

사물인터넷의 Object Sensing Layer에서 일어나는 원천적 불확실성을 제거하기 위해, 첫 번째로 센서노드에서 사용되는 센서에서 오류가 발생할 수 있으므로 적용될 센서들은 실험과 분석을 통하여 성능이 우수하고 안정적인 센서를 선정하게 되며, 센서 노드 영역에서의 센서들의 활동을 살펴보면 센서노드는 일정시간 간격으로 센서를 통해 노드 주변의 상황 값을 획득한다. 획득한 센싱 값은 변환 및 필터링 과정을 거쳐 잡음과 오류 등을 제거한 후 특징점 추출을 통해 감지하고 그 상황 정보를 서버로 전달한다.

Object Sensing Layer의 원천적 불확실성을 제거하기 위한 방법으로 센서에서 Analog로 데이터를 수집하면 Digital로 우

선 변환하고 여기서 발생 하는 잔류 DC를 통해 최종 잡음 제거를 하게 된다. 이러한 방식은 데이터에 섞여있는 노이즈를 제거하는 방식으로 간단한 절차를 통해 잡음 제거가 가능하다.

2. Research in Data Exchange Layer

Data Exchange Layer에서의 연구는 데이터(상황) 모델링 연구와 표준 데이터 변환 연구가 있다.

2.1. Data Modeling Research

상황 모델링은 상황 표현이라고도 부른다. 상황을 표현하기 위하여 두 단계가 있는데 첫 번째 단계에서 새로운 상황 정보는 속성, 특성, 기존 상황과의 관계, 동기식 상황 요청 등을 통해 정의되어야 한다. 두 번째 단계에서 첫 번째 단계의 결과를 검증하고 새로운 상황 정보를 통합해서 기존 상황 정보 저장소에 추가해야 한다. 가장 많이 사용하는 여섯 가지 상황 모델링 기법들은 key-value, markup scheme, graphical, object based, logic based 그리고 온톨로지(ontology) 기반 모델링이다.

(1) Key-Value 모델링

상황 정보를 텍스트 파일과 binary 파일 같은 여러 서식의 key-value(키-값) 쌍으로 표현한다. 다른 기법들에 비해 가장 간단한 상황 표현 방법이고 데이터의 양이 작을 때는 관리하기가 쉽다. 하지만 이 모델링 기법은 확장성이 불가하며 복잡한 데이터 구조를 저장하는데 적합하지 못하다. 또한 메타데이터 첨부가 불가능하다.

(2) Markup Scheme 모델링

태그(tag)를 사용하여 데이터를 표현한다. 따라서 상황은 태그 내에 저장된다. 태그를 사용하는 장점은 효율적인 데이터 검색을 허용한다. 또한 스키마 정의를 통해 검증을 지원한다. XML과 같은 markup 기법에서는 복잡한 검증 도구를 가용할 수 있다. XML은 데이터를 임시로 저장하고 애플리케이션과 애플리케이션 컴포넌트 사이에 데이터를 전송할 수 있다. 하지만 XML 같은 markup 언어들은 추론을 허용하는 고급 표현 능력을 제공하지 않는다.

(3) Graphical 모델링

관계성을 갖는 상황을 표현한다. 이 모델링 기법의 몇 예는 UML과 ORM 이다. 표현력에 있어서 이 모델링 기법은 key-value나 markup scheme 모델링에 비해 월등하다. 그 이유는 상황 모델에 관계성을 표현할 수 있기 때문이다. 우리가 데이터베이스와 익숙하듯이 graphical 모델링은 잘 알려졌고, 배우기 쉬우며 사용하기가 쉽다. 데이터베이스는 많은 양의 데이터를 저장하고 간단한 데이터 검색 연산들을 제공한다. 반면에 상호 운용성이 떨어지고 SQL과 같은 데이터 검색 메커니즘에 한계가 있다. 복잡한 상황 검색의 요구에는 아주 복잡한 SQL 질의가 요구 되는데 결코 쉬운 문제가 아니다.

(4) Object based 모델링

Object 기반(혹은 객체 지향) 개념은 클래스 계층(class hierarchy)과 관계(relationship)를 사용하여 데이터를 표현한다. 객체 지향 패러다임은 캡슐화와 재사용을 도모한다. 대부분의 고급 프로그래밍 언어들이 객체지향 개념을 지원하기 때문에 모델링이 쉽게 상황 인식 시스템에 통합될 수 있다. 하지만 객체 기반 모델링은 내재된 추론 기능을 제공하지 않으며 객체 지향 설계의 검증이 또한 어렵다.

(5) Logic based 모델링

사실(fact), 수식(expressions), 그리고 규칙(rule)들이 상황에 대한 정보를 표현하는데 사용된다. 규칙들은 온톨로지(ontology) 같은 다른 모델링 기법에서 사용된다. 규칙들은 다른 모델과 비교하여 표현력이 훨씬 강하다. 따라서 추론은 어떤 단계까지는 가능하다. 하지만 표준화의 결여 때문에 재사용성과 적용성은 떨어진다. 기술자가 아닌 일반 사용자도 시스템의 실행 시간에 규칙과 논리를 추가할 수 있다. 논리 기반 모델링은 저급 상황을 사용하여 새로운 고급 상황 정보를 추출할 수 있다. 따라서 이 모델은 보충 역할을 하면서 다른 상황 모델링 기법들을 개선할 수 있다.

(6) Ontology Based 모델링

상황은 시맨틱 기술을 사용하여 온톨로지(ontology)로 구성한다. 요구에 따라 몇 개의 표준(RDF, RDFS, OWL)과 추론 기능을 가용할 수 있다. 여러 문헌에 의하면 상황 인식 컴퓨팅과 센서 데이터 관리에 있어 여러 약점에도 불구하고 상황 정보를 관리하고 모델링하는데 온톨로지가 선호되는 메카니즘(mechanism)이다. 온톨로지는 데이터의 표현을 허용하는 시맨틱 기술의 주요 컴포넌트이다. 온톨로지는 관계와 상황을 표현하는 표현적 언어와 종합적인 추론 메카니즘을 제공한다. 또한 온톨로지는 지식의 공유를 허용하고 애플리케이션이나 프로그램 코드로부터 지식을 분리시킨다. 다른 모델링 기법에 비해 온톨로지를 사용하는 가장 큰 이유는 사람과 소프트웨어 에이전트 사이의 정보 구조를 공통으로 이해하고, 도메인 지식을 분석하고, 운영 지식으로부터 도메인 지식을 분리하고, 도메인 지식의 재사용을 가능하게하고, 고급 지식을 추론하기 때문이다.

온톨로지는 공통적으로 개인(individual), 클래스, 속성, 릴레이션(relation), 함수 용어, 제약(restriction), 공리(axiom) 그리고 이벤트와 같은 여러 개의 주요 컴포넌트로 구성된다. 온톨로지 개발은 두 단계가 있다. 첫 번째, 도메인과 범위가 명확하게 정의되어야 한다. 그 다음은 기존의 온톨로지를 검토하여 새로운 온톨로지의 개발을 위한 수단으로 사용한다. 온톨로지의 주요 목적중의 하나가 공유한 지식을 재사용하는 것이다. 온톨로지 개발에 사용될 수 있는 몇 개의 시맨틱 웹 온톨로지 언어들은 RDF, RDFS, OWL 등이 있다.

2.2. Transformation to Standard Data

현재 대부분의 사물인터넷 데이터 표기는 7개의 속성을 이용하여 <object_id, instance_id, attribute, value, time, probability, reliability>와 같이 표현된다. 이 경우 모두 필요한 정보로 구성되어 있지만 네트워크 종류도 다양하고 큰 사물인터넷 생태계에서 7개의 속성을 가진 정보를 주고받는데 있어서 많은 데이터 손실이 발생할 확률이 높다. 위의 문제를 해결하기 위해 본 논문에서는 사물인터넷 데이터를 RDF와 같이 Triple구조로 변환하는 방법을 제안한다[7]. RDF는 자원(Resource)을 기술하기 위한 W3C에서 지정한 시맨틱 웹 표준으로 RDF는 사람이 쉽게 읽고 이해할 수 있을 뿐만 아니라 정보를 표현하기 위해 공통 프레임워크 제공으로 의미 손실이 없어 기계적인 처리가 용이하여 응용프로그램들이 표현된 정보들을 처리하기가 용이하다. 따라서 상황정보를 RDF 구조로의 변환을 통해 위의 두 가지 문제점을 해결 할 수 있을 것으로 보인다.

RDF 데이터를 저장하는 기존의 여러 기법들이 있다. 본 논문에서 제시하는 방식은 Vertical Partitioning 방식을 반영한 HBase 스키마이다.

Vertically Partitioning은 RDF 데이터의 <Subject, Property, Object> 3가지 속성 중 Property 별로 테이블을 생성하고 Subject와 Object를 칼럼으로 갖는다. 각 Table 별로 데이터를 분할하였기 때문에 실제로 접근해야 하는 데이터의 양이 적어지게 된다. 이는 조인 연산 수행 시 속도의 상승을 가져올 수 있다. 또한 이 방법을 사용하면 원하는 속성에만 접근하기 때문에 I/O Cost를 절약 할 수 있다.

그러나 이와 같은 방법은 기존 RDBMS 기반의 저장이나 싱글머신에서의 동작에 제한되어 있어 최근 각 분야에서 실질적으로 생산되고 있는 테라바이트(Terabyte) 혹은 페타바이트(Petabyte) 단위의 RDF 데이터 처리에는 한계가 있다. 데이터 처리에 대한 패러다임의 변화에 발맞추어 RDF 데이터를 빅 데이터 시스템에서 효율적으로 처리하기 위한 새로운 저장방식과 조인방식의 연구가 필요한 상황이다. 이에 맞추어 새로운 처리 시스템인 MapReduce에 RDF 데이터를 적용하고, 성능 향상을 얻기 위하여 연구가 필요한 상황이다.

3. Data Inference in Information Integration Layer

Information Integration Layer에서의 연구는 IoT 데이터의 추론 및 이해에 대한 연구와 IoT 프레임워크를 검증할 수 있는 실험 환경이다.

3.1. Context Inference

상황 추론은 이용 가능한 상황을 기반으로 새로운 지식을 추정하고 보다 더 이해하는 방법이라고 정의할 수 있다. 주어진 상황으로부터 고급 상황을 추론하는 절차라고 설명할 수 있다. 추론의 요구는 원시 상황의 불완전성과 불확실성 때문이다. 상황의 추론은 크게 상황 전처리(context pre-processing), 센서 데이터 융합(fusion), 그리고 상황 추론(inference)의 세 가지 단계로 나눌 수 있다.

상황 전처리 단계에서는 수집된 센서 데이터를 정제한다. 센

서 하드웨어 및 통신망의 비효율성으로 수집된 데이터가 정확하지 않거나 손실될 수 있다. 따라서 손실된 값을 보강하거나, 이상한 부분을 제거하거나, 다중 소스로 부터 상황을 검증 하는 등의 방법을 이용하여 데이터를 정제할 필요가 있다. 센서 데이터 융합 단계에서는 다중 센서로 부터 센서 데이터를 혼합하여 보다 정확하고, 완전하고 그리고 신뢰성 있는 정보를 생성한다. 상황 추론 단계에서는 저급 상황을 사용하여 고급 상황 정보를 생성한다. 추론은 한 번의 상호 작용 아니면 여러 번의 상호 작용에서 할 수 있다.

의사 결정 트리, naive Bayes, hidden Markov, K-nearest neighbor, 인공 신경 망, Dempster-Shafer, 온토로지 기반, 논리 기반, fuzzy reasoning 등 많은 상황 추론 모델들이 있다. 이들 모델들은 상황 추론에 한정되어 있지 않고 인공 지능과 기계 학습 분야에서 이미 사용되고 있는 모델들이다.

IoT 패러다임에서는 상황 정보를 감지하고 생성하는 센서들의 개수가 몇 억 혹은 몇 십억을 넘을 수 있기 때문에 수집된 정보가 엄청날 수 있다. 따라서 처리 시간, 전력, 저장 등 여러 이유 때문에 모든 상황 정보를 대상으로 추론하기에는 현실적으로 불가능하다. 높은 정확성을 갖는 고급 상황을 추론하기 위해서는 적절한 원시 상황을 선택하는 것이 핵심이다.

상황 추론 기법들을 다음 여섯 개의 항목으로 분류하여 설명한다: 지도 학습(supervised learning), 비지도 학습(unsupervised learning), 규칙(rules), 퍼지 논리(fuzzy logic), 온토로지 추론(ontology reasoning) 그리고 확률 추론(probabilistic reasoning).

(1) 지도 학습

이 기법에서는 먼저 훈련하는 예제를 수집한다. 다음은 기대하는 결과대로 그들을 표시한다. 그 다음은 훈련 데이터를 사용하여 기대하는 결과를 생성할 수 있는 함수를 얻는다. 이 기법은 모바일 폰 센싱 이나 activity 인지에 널리 사용된다. 의사결정 트리(decision tree)는 지도 학습 방법으로 데이터 집합에서 트리를 구축하고 그 트리를 데이터를 분류하는데 사용한다. Bayesian network는 확률 추론 개념에 기반하는 기법인데 directed acyclic graph(DAG)를 이용하여 이벤트와 그들 사이의 관계를 표현한다. 이 기법은 통계 추론에 널리 사용하는데 많은 소스로부터 불확실성 데이터를 혼합하여 고급 상황을 추론한다. 인공 신경 망(neural network)은 생물학적 신경 시스템을 흉내 내는 기법이다. 그들은 입력(input)과 출력(output) 사이의 복잡한 관계를 표현하거나 데이터에 패턴을 찾는 데 일반적으로 사용된다.

(2) 비지도 학습

이 기법들은 표시가 안 된 데이터에서 숨겨진 구조를 찾는다. 훈련 데이터를 사용하지 않기 때문에 잠재 해결책을 평가하는 오류 혹은 보상 신호가 없다. K-Nearest Neighbor (K-NN) 기법은 상황 인식 추론에 자주 사용되는데 라우팅 같

은 저급 센서 네트워크 연산자나 실내/외 위치 같은 고급 태스크에 clustering 기법을 사용한다. Kohonen Self-Organizing Map (KSOM) 같은 비지도 신경망 기법은 실시간으로 들어오는 센서 데이터를 분류하는데 사용된다.

(3) 규칙

이 추론 방법은 가장 간단하고 단순한 방법이다. 규칙은 보통 IF-THEN-ELSE 포맷의 구조로 되어있다. 규칙은 저급 상황에서 고급 상황 정보를 생성한다. 최근에는 규칙이 온토로지 추론과 혼합해서 널리 사용되고 있다. 대부분의 사용자 선호도가 규칙을 사용하여 표현되고 규칙은 또한 이벤트 검출에 사용된다. 규칙은 IoT에 상당한 역할을 하리라 기대 되는데 기계에 인간의 생각과 추론을 표현하는 가장 쉽고 간단한 방법이기 때문이다.

(4) 퍼지 논리

이 방법은 확정되고 명쾌한 추론 대신 근사치 추론을 허용한다. 퍼지 논리에서는 부분 참 값도 허용된다. 대부분의 실세계 사실들이 명확하지 않기 때문에 실세계 시나리오를 보다 자연스럽게 표현할 수 있다. 또한 이 방법은 정확한 수치 값(예, 온도: 10도) 대신 자연어(예, 온도: 약간 덥다, 상당히 춥다) 정의를 사용할 수 있다. 다시 말해 상황 정보 처리에 매우 중요한 부정확한 개념들인 크다, 작다, 어둡다, 믿을만하다 등을 정의할 수 있다. 대부분의 경우 퍼지 논리는 독자적인 추론 기법으로 사용될 수 없고 규칙 기반, 확률 혹은 온토로지 추론 등의 다른 기법들을 보완하는데 사용한다.

(5) 온토로지 기반

온토로지 추론은 두 개의 시멘틱 웹 언어인 RDF 와 OWL 에서 주로 지원된다. 온토로지 추론의 장점은 온토로지 모델링과 잘 통합될 수 있다. 반대로 온토로지 추론의 단점은 애매한 정보나 결측값(missing value)을 찾아주지 못하는데 통계 추론 기법은 이를 가능하게 한다. 이 단점을 최소화하기 위하여 규칙을 사용하여 저급 상황에서 새로운 상황 정보를 생성한다. 결측값을 해결하기 위하여 규칙을 사용해서 결측값을 적당한 미리 정의한 값으로 대체할 수 있다. 온토로지 추론은 activity 인식, 하이브리드 추론 그리고 이벤트 검출 등에 널리 사용된다.

(6) 확률 추론

이 기법은 문제와 관련된 사실에 첨부된 확률에 기반하여 결정을 내린다. 두 개의 다른 소스로부터 센서 데이터를 혼합하는데 사용될 수 있고 상황 사이에 일어나는 충돌을 해결하는데 사용될 수도 있다. 이 기법은 대부분 이벤트의 발생을 이해하는데 사용된다. Dempster-Shafer는 확률 논리에 기반하는데 이벤트의 확률 계산을 위해 다른 증거들을 혼합한다. Dempster-Shafer는 activity 인식을 위한 센서 데이터 융합에 일반적으로 사용된다. Hidden Markov Model은 상태(state)를

직접 읽지 않고 관찰 가능한 증거들을 이용하여 상태를 표현하는 확률 기법이다.

지금까지 몇 개의 상황 모델링과 추론 기법에 대해 살펴보았다. 하지만 각각의 기법은 분명하게 단점과 장점이 존재한다. 완벽한 결과를 위하여 어떤 단일 기법을 사용할 수 없다. 따라서 IoT 프레임워크에서 상황 인식 문제를 해결하기 위한 가장 좋은 방법은 여러 개의 기법들을 혼합해서 서로 보완함으로써 단점을 줄이는 것이다. 예를 들면, 상황 정보에 불확실성이 존재할 때 불확실성 환경에서 추론할 수 있는 전략을 제시할 수 있는 Bayesian network, Dempster-Shafer 혹은 퍼지 논리가 그러한 상황에 적합하다. 따라서 다중 기법들을 어떻게 혼합하는지에 대한 절차도 아주 중요한 문제이고 이 문제를 해결하기 위한 방법론을 찾아야 한다.

3.2. Experiment Environment

위에서 제안한 기법들을 기반으로 신뢰성 높은 IoT 플랫폼을 구축하고 실험을 통해 검증할 수 있어야 한다.

본 논문에서 제안하는 실험 환경을 통해 상황 인식 컴퓨팅 기법들을 IoT 프레임워크에 적용하고 IoT 데이터의 활용을 증명해 보일 수 있다. 본 연구에서 제안하는 실험 환경은 먼저 정부 3.0이란 공공정보를 적극 개방·공유하고, 부처(→부처조직) 간 칸막이를 없애며 소통(→커뮤니케이션)·협력함으로써 국민 개인에 대해 맞춤형 서비스를 제공하는 노력에 최대 역점을 두는 새로운 정부 운영의 패러다임이다. 정부 3.0으로 인해 오픈된 여러 공공데이터를 획득 할 수 있다 [그림 2].

기관명	홈페이지명	홈페이지 주소	홈페이지 내용
건강보험심사평가원	보건의료빅데이터개방시스템	opendata.hira.or.kr	의료데이터, 의료통계 공공데이터 정보
경기도	경기도 공공데이터	data.gg.go.kr	경기도 공공데이터 개방 요청 창구
국토교통부	빅월드	map.vworld.kr	국가공간정보 제공
기상청	기상자료개발포털	data.kma.go.kr	기상자료와 설명정보 제공
농림수산식품기술기획평가원	농림수산식품연구개발사업통합정보서비스	www.rhis.go.kr	농림수산식품 R&D사업정보 제공
서울특별시	서울 열린데이터 광장	data.seoul.go.kr	서울특별시 공공데이터 개방정책, 분야별 검색 등
행정자치부	공공데이터 포털	www.data.go.kr	data 및 open api 공유지원 검색 및 활용
특허청	특허정보 검색서비스	www.kipris.or.kr	국내외 특허정보 검색, 특허서비스 안내
한국농수산식품유통공사	식품산업통계정보	www.afis.or.kr	식품 관련 통계 및 조사 정보 제공
한국정보화진흥원	공공정보 품질관리지원센터	www.gooddata.kr	공공기관 ai 품질관리
전라북도	공공데이터 개방 시스템	opendata.go.kr	전라북도 공공데이터
조달청	은통조달	http://opstat.go.kr	공공조달 통계정보 제공

Fig. 2. Public Data

Open Signal의 Weather Signal이라는 애플리케이션을 내려 받으면, 현재 이용자가 있는 곳의 온도를 보여준다. 이는 지역에 설치된 온도 센서를 통해 온도 정보를 제공하는 것이 아니다. 에어플리케이션 사용자로부터 동의를 받아 휴대폰의 배터

리 온도를 수집-분석해 실제 실내/외 온도를 산출한다. 클라우드 빅데이터 서비스로 정확한 온도 정보를 제공한다.

이렇게 스마트폰의 센서 정보를 활용해 곳곳에 온도 센서를 설치하지 않고도 여러 곳의 날씨를 정확하게 유추해 낼 수 있다. 기상청과 같은 측정 기관에서 정확한 온도를 측정하기 위해 많은 비용을 투자해 데이터를 수집하는 것을 대신한 것이다. 이런 분석을 하기 위해 15만 개의 스마트폰으로부터 취합한 빅데이터를 분석해 서비스하고 있다.

추가적으로 사용할 수 있는 실험용 벤치마크에는 LUBM (Lehigh University Benchmark)가 있다[12]. LUBM은 여러 IoT 시스템을 평가하는데 사용할 수 있으며 많은 RDF/OWL 시스템에서 평가하는 벤치마크로 사용 되었다.

IV. Conclusions

IoT 데이터의 효율적인 활용을 위해서는 IoT 데이터의 신뢰가 필요하고 또한 이해가 필요하다. 본 논문에서는 이를 실현하기 위한 IoT 프레임워크를 제안하였다. IoT 프레임워크의 핵심은 4 layer IoT 구조에 상황 인식 컴퓨팅 기술을 적용하였다.

IoT 데이터를 신뢰하고 믿을 수 없다면 의도한 용도에 사용할 수 없으며, 해당 데이터와 관련된 전체 서비스는 해체 될 것이다. 본 논문에서는 IoT 환경에서 데이터의 불확실성을 낮추어 신뢰성 높은 정보를 생성하는 내용을 기술하였다.

다양한 IoT 기기들은 다양한 포맷의 데이터를 내보기 때문에 데이터의 관리 및 처리의 어려움이 있고 크기도 다양해 데이터를 교환할 때 안정성에 대한 문제도 발생할 수 있다. 따라서 다양한 포맷의 데이터들을 하나의 포맷으로 정형화시켜주는 연구가 필요하다. 본 논문에서는 IoT 기기들에서 나오는 다양한 포맷의 데이터들을 시맨틱 기술을 이용해 RDF형식으로 정형화하여 데이터의 불확실성을 낮추고, 관리 및 처리의 안정성을 높이는 연구를 제안하였다.

IoT에서 수집된 엄청난 양의 센서 데이터를 모두 처리하는 것은 현실적으로 불가능하다. 따라서 어떤 데이터를 처리할 필요가 있는지를 결정해야 하는데 이때 상황 인식 컴퓨팅 기술이 적용될 수 있다. 또한 윈시 센서 정보에 가치를 부여하기 위해서는 먼저 그 정보들을 이해 하여야한다. 이를 위하여 센서 데이터와 관련된 상황(context)의 추론이 매우 중요하다. 기존의 여러 인공지능 및 통계적 추론 기법들이 있는데 이들을 데이터 추론에 적용할 수 있다. 본 논문에서는 각각의 추론 기법의 장점과 단점을 기술하고 효율성을 위해 혼합 추론 기법을 제안하였다.

마지막으로 IoT 프레임워크를 사용하여 IoT 응용을 구축하고 실효성을 증명하기 위한 실험 환경을 제시하였다.

REFERENCES

- [1] Dey, A.K., Abowd, G.D. The Context Toolkit: Aiding the Development of Context-Aware Applications. In Proc. of Workshop on Software Engineering for Wearable and Pervasive Computing, pp. 1-3, 2000
- [2] Chen, H., Finin, T., Joshi, A. An Ontology for Context-Aware Pervasive Computing Environments. In Knowledge Engineering Review, Volume 18, No. 3, 2003
- [3] Satyanarayanan, M. Coping with uncertainty. IEEE Pervasive Computing, 2(3), 2, 2003.
- [4] Gu, T., Pung, H. K., & Zhang, D. Q. A bayesian approach for dealing with uncertain contexts. In 2nd International Conference on Pervasive Computing (Pervasive 2004). 2004.
- [5] Ranganathan, A., Al-Muhtadi, J., & Campbell, R. H. Reasoning about uncertain contexts in Pervasive computing environments. IEEE PERVASIVE Computing Journal, 3(2), 62-70. 2004.
- [6] J. Misić, V. B. Misić, and F. Banaie. Reliable and scalable data acquisition from IoT domains. In Proceedings of IEEE GlobeCom, Singapore, 2017.
- [7] KHADILKAR, Vaibhav, et al. Jena-HBase: A distributed, scalable and efficient RDF triple store. In: Proceedings of the 11th International Semantic Web Conference Posters & Demonstrations Track, ISWC-PD. 2012. p. 85-88.
- [8] Perera, C, Zaslavsky, Christen, A, Georgakopoulos, D. Context Aware Computing for The Internet of Things: A Survey. IEEE Communications Surveys &Tutorials, Volume: 16, Issue 1, 2014
- [9] Qin, Y, Sheng, Q.Z, Falkner, N.J, Dustdar, S, Wang, H, Vasilakos, A.V. When things matter: A survey on data-centric internet of things. Journal of Netw. Comput. Appl. 2016, 64, 137-153.
- [10] Hua-Dong Ma. Internet of Things: Objectives and Scientific Challenges. Journal of Computer Science and Technology, 26(6) 919-924, Nov. 2011
- [11] Chaoqun Ji, Jin Liu, Xiaofeng Wang. A Review for Semantic Sensor Web Research and Applications. Advanced Science and Technology Letters, Vol. 48 (ISA 2014), pp. 31-36
- [12] Yuanbo Guo, Zhengxiang Pan, Jeff Heflin. LUBM: A benchmark for OWL knowledge base systems. Journal of Web Semantics, Vol. 3, Issues 2-3, pp. 158-182, Oct. 2005

Authors



Kyung-Chang Kim received the B.S. in Computer Science from Hongik University in 1978, M.S. in Computer Science from KAIST in 1980 and Ph.D in Computer Science from University of Texas at Austin in 1990. Dr. Kim joined the faculty of the

Department of Computer Engineering at Hongik University, Seoul, Korea, in 1991. He is currently a Professor in the Department of Computer Engineering, Hongik University. He is interested in main memory databases, sensor databases, web databases, Internet of Things and big data processing.