

인지 IoT 컴퓨팅 기술동향

Cognitive IoT Computing Technology Trends

배명남 (M.N. Bae) IoT 연구본부 책임연구원
이강복 (K.B. Lee) IoT 연구본부 책임연구원
방효찬 (H.C. Bang) IoT 연구본부 책임연구원

* 본고는 2017년도 ETRI 주요사업의 연구비 지원으로 수행되었음[1711032842, Infra-less 보행방법 기반 증강인지 커넥티드 헬멧 시스템 기술 개발].

사물인터넷은 모든 사람과 사물이 인터넷을 통해 서로 소통하고 새로운 가치를 창출할 수 있는 기술이며, 정보의 확산, 연계, 활용을 가능하게 하는 중요한 연결고리이다. 인지 IoT는 이러한 사물인터넷 인프라와 함께 인공지능 기술을 활용하여, 사물이 스스로 생각하고 판단하며, 보다 잘 연결하고 더 똑똑해지도록 하는 사물지능 실현 기술이다. 본고는 인간 두뇌의 기능을 모방하여 인식, 행동, 인지 능력을 재현해내는 대표 인지 컴퓨팅 기술인 IBM 왓슨, 딥 러닝, 뉴로모픽칩 기술을 요약하며, 또한, 사물수준 지능 실현 사례인 IBM 퀵크, CISCO DMo와 D3의 개발 현황을 소개한다.

초연결 지능 인프라 특집

- I. 서론
- II. 인지 컴퓨팅
- III. 인지 IoT
- IV. 결론



본 저작물은 공공누리 제4유형
출치표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.

I. 서론

인지 컴퓨팅은 인간 두뇌의 기능을 모방하여 인식, 행동, 그리고 인지능력을 재현해내는 기술을 의미한다. 컴퓨터와 모바일, 웨어러블 기기 등이 사람의 목소리나 몸짓을 감지하여 명령을 자동으로 수행하고, 학습을 통한 상황파악과 사람의 두뇌가 처리하지 못하는 빅데이터도 분석할 수 있다. 즉, 인지 컴퓨팅의 핵심은 사용자를 이해하고 사용자와 교감을 나누고, 사용자의 특성이나 행동을 직관적으로 파악해 관련 정보와 서비스를 제공하는 것이다. 결국, 인지 컴퓨팅은 생활에서 인간의 기술을 향상시키고 휴먼증강을 통해 인간의 능력을 향상시켜 준다[1].

한편, 사물인터넷 기기는 2020년까지 54억개로 늘어날 것으로 예측되고 있고, 생산하는 데이터의 양도 기하급수적으로 늘어나, 더 이상 감당할 수 없는 수준의 정보가 쌓일 것으로 전망되고 있다. 따라서 중앙 서버 인프라와 분담하여 사물에 부여된 인지 컴퓨팅 기능을 통해 사물이 스스로 생각하고 판단함으로써, 자체적인 데이터 처리와 운영 방식을 통한 유연성뿐만 아니라 의사결정 속도를 개선하고 현장의 성능 최적화도 가능할 것으로 기대된다. 나아가 지능을 가진 사물 실현의 예로, 인간 삶의 동반자로서 인간의 감정을 공유하고 이해하는 연구들도 진행되고 있으며, 소프트뱅크의 페퍼(pepper), MS의 음성 서비스 코타나(Cortana), 아마존은 에코(Echo), SK텔레콤의 인공지능 스피커 ‘누구’를 통해 실현되어 가고 있다.

본고에서는 사물인터넷에서 인지능력을 갖는 사물의 실현과 이를 위한 관련 기술의 개발 단계와 방향에 대해 살펴보고자 한다.

II. 인지 컴퓨팅

인지 컴퓨팅 시스템은 인지 컴퓨팅의 실현을 통해 인

간의 능력을 확장하고 올바른 의사결정을 지원하는 시스템이다. 본고에서는 인지 컴퓨팅이라고 하는 새로운 컴퓨팅 환경을 두 가지 관점에서 살펴본다. 먼저, 현재 유일하게 실용화된 규칙기반의 IBM 왓슨과 기존 컴퓨팅 환경과 달리 생명체의 두뇌에서 신경세포들의 상호작용에 의해 스스로 학습하고 성장하는 신경망 컴퓨팅 환경에 대한 새로운 연구인 딥 러닝과 뉴로모픽칩에 대해 기술한다.

1. IBM 왓슨(Watson)

IBM 왓슨은 자연어 인터페이스를 통해 인간과 질의-응답을 통한 대화가 가능하며, 질의에 대한 가설 설정과 이에 대한 평가를 제공하고 사람의 선택과 조언을 학습하여 스스로 진화할 수 있는 규칙기반의 인공지능 시스템이다.

왓슨은 의료, 교통, 재정, 전력제어 등의 여러 분야에 직접 활용하기 위한 전문 지식을 지속적으로 습득하고 있으며, 최근 국내에서도 종양학 진료에 왓슨을 적용한 ‘왓슨 포 온콜로지(Watson for Oncology)’를 본격 도입하였다. 왓슨은 이미 종양학 분야의 300개 이상의 의학 학술지, 200개 이상의 의학 교과서를 포함해 거의 1,500만 페이지에 달하는 의료정보를 학습한 상태이다. 의사들은 왓슨을 통해 전문가 검토가 이뤄진 연구결과와 임상 가이드라인 및 전문가 소견을 확인할 수 있어, ‘근거에 기반한’ 진단 및 치료가 가능하다[2].

왓슨의 역할은 학습해야 하는 방대한 자료를 볼 때 명확하다. 현재 의학 문헌은 5년마다 2배씩 늘어나고 있고, 지난 해에만 전 세계적으로 약 4만 4천 건에 달하는 종양학 논문이 의료 학술지에 발표됐다. 매일 약 122개의 새로운 논문이 발표되고 있어, 이는 이미 인간의 인지능력을 넘어서는 양으로 의사들이 이러한 최신 정보들에 직접 접근하기는 어렵다. 진료행위가 ‘의사의 경험’에 의존하고 있다는 점에서, 환자의 상태를 이해하

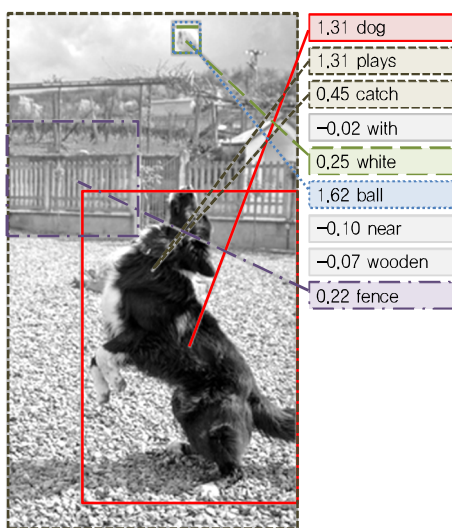
고, 치료법을 결정하며, 치료법이나 약물의 부작용 등의 확인에 정확한 의학적 근거의 부족으로 연결될 수 있다는 점이다. 왓슨은 의사와 서로 대화하고 자신의 가설에 대한 근거를 통한 신뢰도 부여를 통해, 이러한 상황을 타계할 새로운 조연자로 여겨지고 있다.

이외에도, 왓슨은 소프트뱅크의 인공지능 로봇 페퍼 (Pepper)나 나오미(Nao-mi)에도 탑재되어 있다. 국내에서는 IBM과의 협업을 통해 왓슨 기반 에이브릴 (Aibril)을 출시하고 생산공정이나 헬스 케어 분야에서의 감염병의 조기 정밀 진단과 환자 맞춤형 치료 등의 다양한 활용을 위해 노력하고 있다.

2. 딥러닝

2006년에 등장한 딥 러닝은 컴퓨터 상에서 인간 두뇌의 생물학적 모델링을 통해 인공지능을 구현한 혁신적인 기술로 주목받고 있다. 기본적으로 대량의 데이터 안에서 패턴을 발견하고 분류하여 예측하는 방법론이다.

뉴럴 네트워크라고 불리는 두뇌의 구조를 공학적으로 재현한 기술은 있었지만, 오랫동안 실용적인 수준에 도달하지 못했다. 2012년에는 구글과 스탠퍼드 대학의 공동연구에서 딥 러닝을 이용하여 유튜브의 1,000만 개



(그림 1) 뉴럴 네트워크를 이용한 이미지 인식[3]

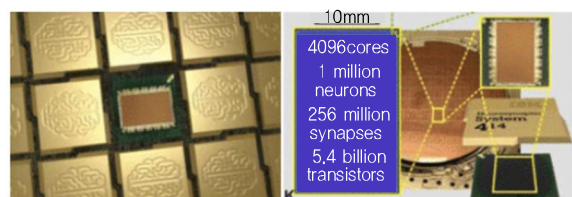
에 달하는 동영상에서 무작위로 추출한 이미지를 읽어 들인 후 인공지능에 학습시킨 결과, 자력으로 ‘고양이’라는 개념을 만들어냈다는 점에서 혁신적이었다.

2014년 11월에는 구글과 스탠퍼드 대학에서 새로운 인공지능 연구 성과를 발표했다. 리커런트 뉴럴 네트워크(RNN) 기술을 사용한 이 연구는 이미지나 동영상의 내용을 인식하고 여기에 있는 인물이나 물체를 식별하고 사건을 자연어로 설명할 수 있다. (그림 1)의 이미지에 대해, ‘dog plays catch with white ball near wooden fence’로 설명하는 수준까지 인식될 수 있음을 보인다 [4]. 시간 제약 내에 보다 정밀도 높은 이미지 인식이 가능해진다면, 상황에 따른 행동도 가능해질 것으로 기대된다. 이를 위해, 구글 알파벳이 최근에 발표한 하드웨어 유닛인 텐서처리장치(TPU), 퀄컴의 새로운 뉴럴 처리 장치인 제로스(Zeroth), 엔비디아의 딥 러닝 칩 (Parker), 또는 IBM의 뉴로모픽칩 트로노스(TrueNorth) 등의 하드웨어가 출시되었다. 그 외에도 더 다양한 빠른 칩이 개발되고 활성화되어 더 강력한 인지 컴퓨팅이 예상된다.

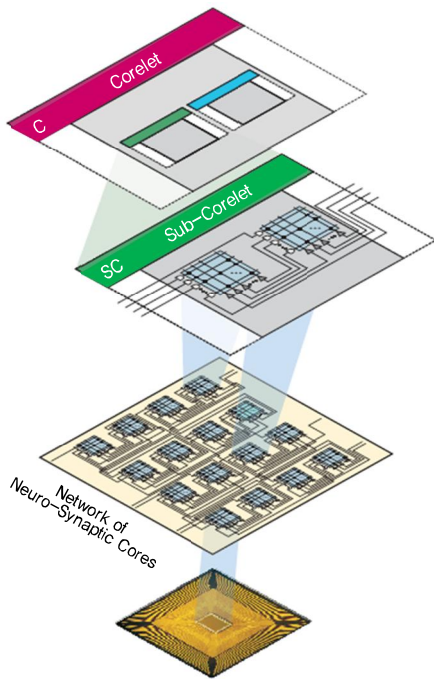
3. 뉴로모픽칩

소프트웨어에 의한 인공지능을 만들어내는 시도가 계속되고 있는 한편, 하드웨어의 구조도 뇌의 구조에 가깝도록 모사하여 만들어내려는 연구가 이뤄지고 있다.

(그림 2)의 뉴로모픽칩은 인간 두뇌의 구조를 모방한 컴퓨터 칩으로 2008년 DARPA의 SyNAPS 프로그램을 비롯하여 여러 기업에 의해 진행되고 있고, 이 칩이 실용화됨으로써 보다 인간의 사고에 가까운 인지 기능의



(그림 2) IBM의 TrueNorth[4]



(그림 3) TrueNorth 프로그램 구성-Corelet[7]

개발이 기대되고 있다. 2012년에는 인텔사가 독자 설계를 공개했고 2013년에는 퀄컴도 뉴로모픽칩 개발에 동참했다.

SyNAPS 프로그램의 지원을 받아, 2014년 8월에 IBM이 발표한 트루노스는 4,096개의 코어를 갖고 있고 인간 뇌의 100만 뇌신경과 2억 5,600만 개 시냅스를 재현하고 있다. 1와트로 초당 460억회 시냅틱 작동을 할 수 있고, 이미지와 소리 등 인식 데이터 분석에 활용될 수 있으며, 코어에는 100kb메모리가 구현되어 뉴런의 신호가 송수신 정보로 저장된다. IBM은 트루노스로 사람, 자전거, 자동차 등의 물체를 식별하는 데 성공한 바 있으며, 400×400화소 화질로 초당 30프레임씩 움직이는 동영상 처리에 사용한 전력이 63mW밖에 소비되지 않아 절전 성능도 갖추고 있다[5].

이와 같은 뉴로모픽 컴퓨팅 환경은 분산처리, 매우 높은 수준의 상호 연결, 비동기성, 병렬, 대규모의 컴퓨팅 모델을 만들어내야 하는 것이고, 이에 상응하는 복잡한 알고리즘과 프로그래밍 언어를 개발해야 한다. 동시

〈표 1〉 프로세서 및 개발환경

칩(chip)	개발언어	라이브러리	개발/시험환경
TrueNorth	Corelet PL	Corelet library	COMPASS
CPU/GPU/TPU	Python	package	Tensor board
FPGA	VHDL	IP Core	Chipscope

에 소프트웨어와 하드웨어 상호 관계가 매우 관련성이 높기 때문에 효율적인 프로그래밍을 위한 새로운 패러다임이 필요하다[6]. 이를 위해, IBM은 코어릿 프로그래밍 언어와 개발 환경인 COMPASS를 제공한다. 초기에는 적용할 코어를 정밀하게 프로그래밍해야 했지만, 나중에 이들을 조합하고 재사용이 가능한 코어릿(Corelet)이라는 블록으로 변환된다(그림 3) 참조. 하나의 코어릿은 특정한 기능을 가질 수 있는 뉴로시냅틱 코어들의 네트워크 추상화로 볼 수 있는데, 예를 들어 소리를 인지하는 모든 개별 코어를 포함할 수 있다. 개발된 코어릿은 실제 뉴로모픽 하드웨어로 구현하기 전에, 트루노스 시뮬레이션 환경인 COMPASS를 통해 사전 시험과 구성이 가능하며, 동시에 컴파일러로서 코어릿의 설계, 개발, 배포에 사용할 수 있다[〈표 1〉 참조].

III. 인지 IoT(Cognitive IoT)

사물인터넷(IoT)은 모든 사람과 사물이 인터넷을 통해서 서로 소통하고 새로운 가치를 창출할 수 있는 기술이며, 정보의 확산, 연결, 활용을 가능하게 하는 중요한 연결고리이다. 인지 IoT는 이러한 사물인터넷 인프라에 더해 인공지능 기술을 활용하여, 사물을 서로 연결하고 더욱 똑똑해진 사물이 스스로 생각하고 판단하는 사물중심의 지능사회 실현을 위한 기술이다.

흔히 사물인터넷 환경은 모든 데이터를 중앙(클라우드 컴퓨팅 인프라)에 올려놓고 분산 처리만으로 충분하다고 여기지만, 결국 개인화나 개별 상황을 고려하지 않아 발생하는 통계적 왜곡으로 인한 분석 편준화 혹은 단지 수집만 될 뿐 분석에 활용하지 않는 데이터 버림 등이 발생한다. 또한, 모든 디바이스들이 중

양에 모두 연결하고, 데이터를 주고받다 보니 이제 너무 혼잡하고 속도도 느려진다는 것이다.

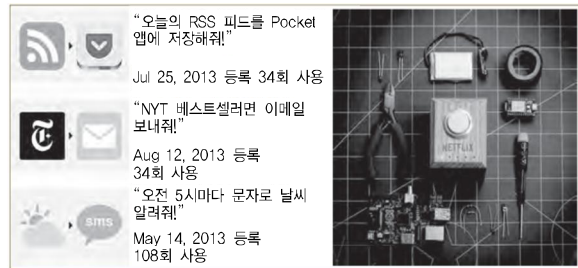
이러한 문제의 극복 방안으로, 사물이 지능을 가지고 스스로 학습과 판단함으로써, 방대한 양의 데이터를 굳이 먼 곳에 있는 대용량 데이터 서버에 저장하지 않고 데이터가 발생된 인근 지점에서 보다 빠르게 판단하고 반응하는 것이 바람직할 수 있다. 예를 들어, 도로에서 구급차가 감지되면 신호등을 즉각적으로 초록불로 바뀌는 스마트 교통 신호 등에 사용할 수 있을 것이다.

이에 따라, CISCO의 포그 컴퓨팅 및 IBM의 퀴크(Quarks) 기술은 중앙 서버로 운영되는 클라우드와 지역적으로 분산된 사물인터넷 기기들을 중간에서 연결하고 매개하는 또 하나의 네트워크 구성을 지향하는 구조로 제안되었으며, 기존 방식과 달리 주변(edge, 개별기기)에 더 주목한다는 점이 주요한 특징이다[8].

1. IFTTT

조건부 제어(IFTTT)는 일정한 조건이 발생하거나 사용자가 특정 명령을 내릴 때, 자주 사용하는 동작을 실행시켜 주는 체계이며, ‘이 조건에서는 저 동작을 해라’라는 뜻이 있고, 필요하다면 두 가지 이상의 동작을 조합하여 새로운 동작을 만들 수 있는 자동화 도구이다. 이를 통해, 사용자가 임의로 기존 서비스나 동작을 선택한 후 조건이나 동작에 대응하는 사물의 어떤 기능과 연결하여 설정하는 것이다. IFTTT는 서비스나 동작뿐만 아니라 하드웨어와도 연결될 수 있다. 예를 들어, 넷플릭스의 ‘더 스위치’, 필립스 휴의 LED 전구, 벨킨 위모 등의 전원 제어 하드웨어도 특정 조건에서 명령을 받아 작동될 수 있다. (그림 4)의 ‘더 스위치’는 버튼 하나로 조명을 켜거나 끄고, 매일 매일의 날씨 자동 조회, 특정 시간에 스마트 폰을 매너 모드로 변경, 혹은 미리 입력해 놓은 음식을 주문할 수도 있는 제품이다[9].

이 기술의 주요 특징은 불특정 다수의 일반 사용자가

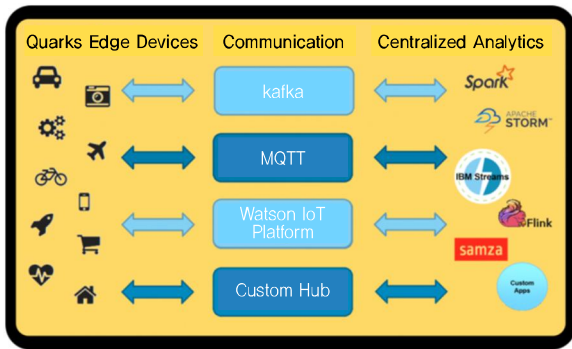


(그림 4) 자동화 앱 IFTTT의 조건/명령 예, 넷플릭스의 ‘더 스위치’[9]

자신에게 특화된 독창적인 서비스 혹은 특정 주변 환경에 적합한 수없이 많은 명령과 조합을 자신만의 요구에 맞춰 구성할 수 있다는 점이다. 이는 사물이 자신만의 판단과 경향을 가질 수 있다는 것을 의미한다. 물론, 조건부 제어의 한계로 단순하고 정교함이 떨어져, 작고 간단한 지능으로 만들어지지만 일반 사용자가 선호하는 하드웨어 인터페이스나 사용 시나리오들의 발굴에는 충분한 시작점이 될 수 있다. 또한, 사물에 가장 쉽게 적용할 수 있고, 컴퓨팅 자원이 가용한 범위 내에 인공지능 기술에 기반한 서비스나 행위가 수용될 수 있다면 더욱 더 다양한 확장도 가능하다.

2. IBM Quarks

IBM은 사물인터넷 환경에서 데이터 분석에 활용할 수 있는 분석도구인 퀴크를 오픈 소스로 공개하였다. 사물인터넷의 중앙 서버 인프라는 다양한 사물들에 산재되어 있는 모든 데이터를 수집하고 분석한다. 이 과정에서, 데이터를 송수신하는 과정에서 비용과 시간이 많이 소요된다. 이러한 문제를 해결하기 위해, 퀴크는 수집한 데이터를 일괄적으로 중앙의 클라우드 시스템에 전송하거나 이로부터 받지 않고, 사물 내부에서 데이터를 즉시 분석하며, 이상 징후 및 의미 있는 사건을 감지한 경우에만 중앙 서버에 데이터를 전송한다. 이러한 방식은 분석 없이 버려지거나 이상징후와 무관한 데이터 송신에 대한 무분별한 자원 활용을 최소화하여, 분석 성능과 효율을 높일 수 있다. 즉, 사물이 퀴크를 탑재하고 이를 활



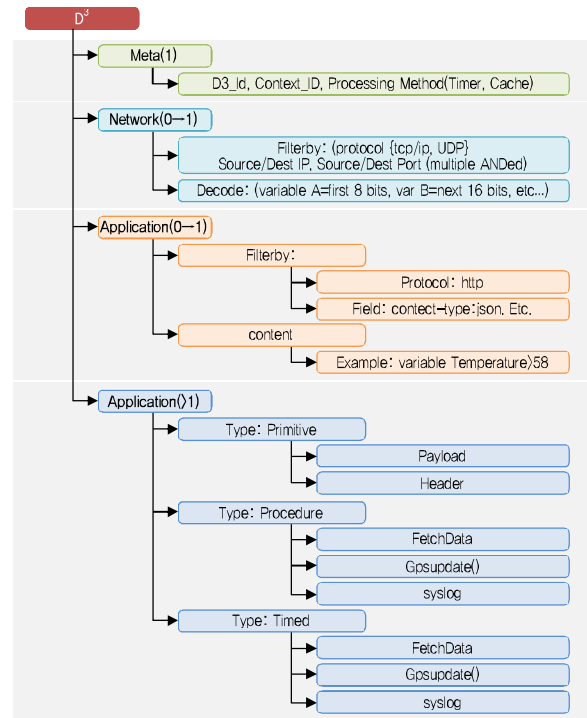
(그림 5) Quarks 활용 예[10]

용하면, 굳이 모든 센서 값을 네트워크를 통해 보낼 필요가 없으며, 평범한 99% 데이터는 그대로 두고, 1%의 독특한 데이터를 발견해 전송할 수 있다. 혹은, 사물 자체를 더 똑똑하게 만들 때도 쿼크를 활용할 수 있다. 예를 들어, 쿼크가 설치된 자동차는 화물 크기나 탑승 승객 수에 따라 구동력을 자동으로 조절하여 최적화할 수 있다[10].

쿼크는 기본적으로 데이터의 필터링과 aggregation, rule trigger 정의가 가능하며, 오픈소스 분석 플랫폼인 카프카, 스파크, 스톱 등과 통합해 쓸 수 있고, 상용 분석 서비스나 사물인터넷 플랫폼과 결합해 스트리밍 데이터 분석이나 엣지 컴퓨팅 분석에 적용될 수 있다.

3. CISCO DMO and D3

CISCO의 포그 컴퓨팅은 모든 센서와 단말 데이터를 중앙으로 보내는 대신, 즉시 활용할 데이터를 가려내 가까운 곳에서 분석하고 활용하는 인프라를 지칭한다. 여기서 CISCO 'IOx' 플랫폼을 포함한 라우터가 데이터 분석과 처리를 담당한다. CISCO는 데이터 처리 모델인 '데이터 인 모션(Data in Motion)'과 '동적 데이터 정의(Dynamic Data Definition, D3)'를 제공하고 있고, 사용자는 실제 데이터를 받아들이고 처리하는 일련의 동작 유형을 정의하는 규칙(룰셋)을 직접 프로그래밍할 수 있다. 프로그래밍된 규칙들은 라우터와 스위치에 탑재되고 모든 네트워크 데이터에 적용된다. CISCO가 제공하



(그림 6) D3 활용 예[11]

는 작은 분산형 엔진은 네트워크 장비에 탑재된 규칙을 해석하여 네트워크 상태(Layer4~7)를 색인화해 최적화된 검색이 가능하고, 개발자는 이를 사용하여 데이터 수집과 분석이 가능한 'IoT엣지 디바이스'를 쉽게 프로그래밍할 수 있다.

동적 데이터 정의는 패턴과 조건, 행위로 구성된 규칙으로, (그림 6)의 예는 특정 주소를 사용하여 오고 가는 TCP/UDP 패킷에 대해, 처음 8bit는 A변수에 다음 16bit는 B변수에 할당하며(Network primitive), 이때, 응용 프로토콜이 http인 경우에는 패킷의 body가 json 형식으로 확인된 경우(Application primitive)에 해당 속성 중 Temperature 값이 56보다 큰지 검사하며(Content primitive), 특정 트리거가 발생하였을 때 실행하는 행위(Action primitive)를 정의하고 있다.

IV. 결론

사물인터넷은 모든 사물과 사물이 인터넷을 통해 서

로 소통하고 새로운 가치 창출을 가능하게 하는 중요한 연결고리이다. 하지만, 사물인터넷 기기는 2020년까지 54억개로 늘어날 것으로 전망되고, 이에 따라, 생산하는 데이터의 양도 기하급수적으로 늘어나고 있어, 현재의 데이터 저장, 통신 장치로도 더 이상 감당할 수 없는 수준의 정보가 쌓이고 있다.

이에 따라, 사물은 핵심 정보만을 중앙 서버에 전송하고, 나머지 데이터는 별도로 관리하는 시대가 곧 다가올 것이며, 2020년 즈음에는 생산되는 데이터의 3분의 1은 사물 수준에서 처리될 것이라는 전망도 있다.

따라서, 사물이 지능을 가지고 스스로 학습하고 판단함으로써, 방대한 양의 데이터를 굳이 먼 곳으로 전송하지 않고, 데이터와 수요가 있는 인근 지점에서 보다 빠르게 판단하고 대응하는 것이 바람직할 것이다. 또한, 인간에게 편의성을 제공하기 위해 지능 처리가 반드시 필요하므로, 인공지능을 기반으로 하는 사물에서의 인지 컴퓨팅은 향후 그 중요성을 더해 갈 것이다.

용어해설

뉴로모픽칩(neuromorphic chip) 사람의 뇌 신경을 모방한 것으로 사람의 사고 과정과 비슷하게 정보를 처리하는 새로운 반도체 소자

약어 정리

AI	Artificial Intelligence
D3	Dynamic Data Definition
DMo	Data in Motion
HTTP	Hypertext Transfer Protocol
IFTTT	If This Then That
IoT	Internet of Things
JSON	JavaScript Object Notation
RNN	Recurrent Neural Network
SyNAPS	System of Neuromorphic Adaptive Plastic Scalable Electronics

TCP	Transmission Control Protocol
TPU	Tensor Processing Unit
UDP	User Datagram Protocol

참고문헌

- [1] Healthcare Business, “IBM의 인공지능 시스템 왓슨이 열어갈 미래 의료,” <http://www.chiweon.com/ibm-watson/>, 2016.
- [2] IT Daily, “알파고 부터 엑소브레인까지, ... AI 시대 본격 개막,” <http://www.itdaily.kr/news/articleView.html?idxno=81439>, 2016.
- [3] Andrej Karpathy and Li Fei-Fei, “Deep Visual-Semantic Alignments for Generating Image Descriptions,” CVPR 2015 Conference, 2015, pp. 3128-3137.
- [4] 월간 App, “인공지능의 미래 연구,” <http://m.postnaver.com/viewer/postView.nhn?volumeNo=3734444&memberNo=1844324&vType=VERTICAL>, 2016.
- [5] F. Akopyan, et. al., “TrueNorth: Design and Tool Flow of a 65mW 1Million Neuron Programmable Neurosynaptic Chip,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* vol. 34, no. 10, 2015, pp. 1537-1557.
- [6] A. Amir et al., “Cognitive Computing Programming Paradigm: A Corelet Language for Composing Networks of Neurosynaptic Cores,” *Int'l Joint Conf. on Neural Networks*, 2013, pp. 1-10.
- [7] A. Knapp, “IBM Developers Programming Language Inspired by The Human Brain,” <http://www.forbes.com/sites/alexknapp/2013/08/08/ibm-develops-programming-language-inspired-by-the-human-brain/#52c869a9d9b7>, 2013.
- [8] 방성수, “시스코의 야심작 ‘포그 컴퓨팅’ ... 사물인터넷과 박테이터 시대의 새 패러다임,” http://biz.chosun.com/site/data/html_dir/2016/10/21/2016102102501.html, 2016.
- [9] 넷플렉스, “더 스위치,” <https://youtube.com/TTIIGdlbTy4>, 2015.
- [10] 이지현, “Streaming analytics from the center to the edge,” <http://www.bloter.net/archives/250015>, 2016.
- [11] D.A. Maluf, “Let’s Build Tomorrow Today,” <http://www.slideshare.net/CiscoDevNet/devnet-1163-data-in-motion-ap-is>, 2015