

헬스케어 정보 수집을 위한 딥 러닝 기반의 서브넷 구축 기법

정윤수

목원대학교 정보통신융합공학부

Subnet Generation Scheme based on Deep Learning for Healthcare Information Gathering

Yoon-Su Jeong

Dept. of Information Communication & Engineering, Mokwon University

요약 최근 IoT 기술이 발달함에 따라 헬스케어 서비스를 제공하는 많은 의료기관에서는 IoT 기술을 이용한 의료 서비스가 점점 증가하고 있는 추세이다. 그러나, 사용자 신체에 부착된 IoT 센서의 갯수가 증가하면서 서버에 전달되는 헬스케어 정보는 점점 복잡해져서 서버에서 사용자의 헬스케어 정보를 분석하는 시간이 증가하는 문제점이 발생하고 있다. 본 논문에서는 사용자에 부착된 IoT 장치를 통해 전달되는 다량의 헬스케어 정보를 서버에서 의료 목적에 맞게 헬스케어 정보를 수집 및 처리하기 위한 딥 러닝 기반의 헬스케어 정보 관리 기법을 제안한다. 제안 기법은 서버에 전달된 헬스케어 정보에 속성값을 부여하여 속성값에 따른 서브넷을 구축한 후 서브넷간 연계 정보를 시드로 추출하여 계층적 구조로 그룹화 한다. 제안 기법은 서버에서 그룹화된 헬스케어 정보를 딥 러닝에 적용하여 사용자의 치료 및 처방에 대한 관찰 속도 및 정확도를 향상시킬 수 있어 최적화된 정보를 추출할 수 있는 특징이 있다. 성능평가 결과, 제안기법은 헬스케어 서비스 모델에서 동작되는 의료 서비스의 처리속도가 기존기법에 비해 평균 14.1% 향상되었고, 서버의 오버헤드는 기존 기법보다 평균 6.7% 낮았다. 헬스케어 정보 추출에 대한 정확도는 기존 기법보다 평균 10.1% 높은 결과를 얻었다.

주제어 : 헬스케어, 딥 러닝, 의료 서비스, 서브넷, 정보 추출

Abstract With the recent development of IoT technology, medical services using IoT technology are increasing in many medical institutions providing health care services. However, as the number of IoT sensors attached to the user body increases, the healthcare information transmitted to the server becomes complicated, thereby increasing the time required for analyzing the user's healthcare information in the server. In this paper, we propose a deep learning based health care information management method to collect and process healthcare information in a server for a large amount of healthcare information delivered through a user - attached IoT device. The proposed scheme constructs a subnet according to the attribute value by assigning an attribute value to the healthcare information transmitted to the server, and extracts the association information between the subnets as a seed and groups them into a hierarchical structure. The server extracts optimized information that can improve the observation speed and accuracy of user's treatment and prescription by using deep learning of grouped healthcare information. As a result of the performance evaluation, the proposed method shows that the processing speed of the medical service operated in the healthcare service model is improved by 14.1% on average and the server overhead is 6.7% lower than the conventional technique. The accuracy of healthcare information extraction was 10.1% higher than the conventional method.

Key Words : Healthcare, Deep Learning, Hospital Service, Subnet, Data Extraction

* 본 연구는 산업통상자원부 지역혁신센터사업인 민군겸용보안공학연구센터 지원으로 수행되었음.

Received 21 December 2016, Revised 2 March 2017

Accepted 20 March 2017, Published 28 March 2017

Corresponding Author: Yoon-Su Jeong(Mokwon University)

Email: bukmunro@mokwon.ac.kr

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ISSN: 1738-1916

1. 서론

IT 기술이 발달함에 따라 정보 기술의 중요성은 점점 증가하고 있는 상황이다. 특히, 정보 기술을 통해 생성되는 데이터는 소셜미디어, IoT(Internet of Things), 멀티미디어, 센서 네트워크 등을 중심으로 수집된 후 헬스케어, 바이오인포매틱스, 공공 관리 분야 등에서 빅 데이터가 저장, 처리 및 검색되고 있다[1,2].

빅 데이터를 사용하는 분야 중 헬스케어 서비스 분야는 2013년 탈모 U-헬스케어 서비스를 시작으로 국가가 주도하는 건강 산업으로 성장하였으며, 헬스케어 서비스의 주된 목적은 질병의 치료보다는 예방과 관리를 중요시하는 추세로 변화하고 있다[3]. 특히, 헬스케어 서비스는 질병 발생 전에 환자의 건강을 효율적으로 관리하기 위해서 아파치 스파크 플랫폼, 클러스터링 및 딥 러닝(deep learning) 알고리즘 등을 활용한 빅 데이터 서비스가 확대 적용되고 있다[4].

딥 러닝 알고리즘은 여러 비선형 변환기법의 조합을 통해 높은 수준의 추상화 -다량의 데이터나 복잡한 자료들 속에서 핵심적인 내용 또는 기능을 요약하는 작업-를 시도하는 기계학습(machine learning) 알고리즘의 집합을 말한다[2]. 딥 러닝은 대량의 빅 데이터 정보를 컴퓨터가 이해할 수 있는 형태로 표현(representation)하고 이를 학습에 적용하기 위해 많은 연구가 진행되고 있다[16-19]. 이러한 노력의 결과로 deep neural networks, convolutional deep neural networks, deep belief networks와 같은 다양한 딥 러닝 기법들이 컴퓨터 비전, 음성인식, 자연어처리, 음성/신호처리 등의 분야에 적용되고 있다[10,20,21].

헬스케어 서비스 분야는 다른 여러 분야에 비해 다음과 같은 차별성을 가진다[10]. 첫째, 개인 진료 정보 분석에 소비되는 시간과 비용을 줄이면서 질병 정보를 분석 및 활용할 수 있다. 둘째, 질병을 유발하는 유전적인 요인을 분석하고 그에 맞는 치료를 손쉽게 찾을 수 있다. 셋째, 장소 및 시간에 상관없이 사용자의 생체정보를 IoT와 같은 장치를 통해 건강상태를 점검받을 수 있다.

헬스케어 서비스는 사용자의 질병 정보를 유·무선 환경을 통해 병원에 전달한 후 사용자의 건강 상태를 확인하여 질병 상태에 따라 진료 및 치료를 제공할 뿐만 아니라 질병 예방 서비스도 사용자에게 제공할 필요가 있다. 헬스케어 서비스가 다른 기술에 비해 사용자 편의

의 서비스 제공이 어려운 것은 질병 정보를 효율적으로 수집 및 관리가 원활하지 않기 때문이다. 헬스케어 서비스의 원활한 제공을 위해서도 사용자의 질병 정보를 효율적으로 수집하고 관리하는 방법이 필요하다[5,6].

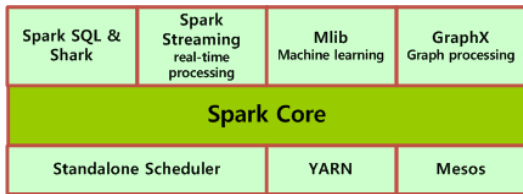
본 논문에서는 센서가 부착된 IoT 장비(e.g. 팔찌, 목거리 등)와 휴대폰을 이용하여 사용자의 건강정보를 효율적으로 관리할 수 있는 딥 러닝 기반의 헬스케어 서비스 관리 기법을 제안한다. 제안된 기법은 IoT 장비를 통해 전달되는 다량의 건강정보를 통해 효율적으로 분석할 수 있도록 딥 러닝 알고리즘을 헬스케어 서비스에 적용하였다. 제안기법에서는 딥 러닝을 헬스케어에 적용함으로써 사용자의 건강상태를 낮은 비용으로 정확하게 체크할 수 있는 것이 특징이다. 또한, 제안기법은 사용자의 헬스케어 정보를 IoT 장치별로 생성하여 계층적으로 통합 관리 하는데 유리하다. 특히, 제안 기법은 IoT 장치에서 사용자의 헬스케어 정보를 확인하고 처리할 수 있는 의료 서비스 절차가 간소화되었다.

이 논문의 구성은 다음과 같다. 2장에서는 제안 기법에서 사용되는 아파치 스파크와 K-Means을 알아보고, 제안 기법과 관련된 기존 연구에 대해서 알아본다. 3장에서는 딥 러닝 기반의 빅 데이터 수집 및 처리 기법을 제안하고, 4장에서는 기존 기법과 제안 기법을 비교 평가하고, 마지막으로 5장에서 결론을 맺는다.

2. 관련연구

2.1 아파치 스파크

아파치 스파크(Apache Spark)는 빠르고, 범용적 목적의 분산 고성능 클러스터링 플랫폼 시스템을 말한다. 아파치 스파크는 하둡의 Map & Reduce를 사용할 수 있을 뿐만 아니라 Storm의 스트리밍 처리도 가능하다[3,4]. 아파치 스파크는 Java, Scala, Python, 그리고 R언어의 API를 제공하기 때문에 쉽게 구현을 할 수 있는 장점을 가지고 있다[7]. 특히, 아파치 스파크는 여러 대의 노드에서 연산을 할 수 있는 범용 분산 클러스터링 플랫폼으로써 [Fig. 1]처럼 Map * Reduce나 스트리밍 처리 등의 모듈을 추가하여 기능을 수행한다.



[Fig. 1] Overall Stack Structure of Apache Spark

[Fig. 1]처럼 아파치 스파크는 다음과 같이 인프라 계층, 스파크 코어, 스파크 라이브러리 등 크게 3가지로 구성된다. 인프라 계층은 독립적으로 기동할 수 있는 Standalone Scheduler, 하둡 종합 플랫폼인 YARN, Docker 가상화 플랫폼인 Mesos가 각각 사용된다. 스파크 코어는 메모리 기반의 분산 클러스터 컴퓨팅 환경을 의미한다. 파크 라이브러리는 빅 데이터를 SQL로 핸들링할 수 있게 해주는 Spark SQL, 실시간으로 들어오는 데이터에 대한 리얼타임 스트리밍 처리를 해주는 Spark Streaming, 그리고 머신러닝을 위한 MLlib, 그래프 데이터 프로세싱이 가능한 GraphX가 있다.

2.2 K-평균

K-평균(K-Means)은 가장 유명하고 강력한 파티션 기반 클러스터링 알고리즘 중 하나이다[11,12,13,14]. K-평균은 임의로 선택된 클러스터 센터에 포인터를 할당하도록 반복적인 작업을 수행한다. 각 반복과정의 끝에는 클러스터 센터를 업데이트 하는 것으로 마무리한다. 이 과정은 센터의 변화가 없을 때까지 반복수에 무관하게 반복 과정을 계속 수행한다. <Table 1>은 K-Means 알고리즘의 Pseudo 코드를 보여주고 있다.

[Table 1] Pseudo Code of K-Means Algorithm

<p>Algorithm : K-Means Algorithm</p> <p>1: k centers C are selected randomly from the data set D.</p> <p>2: while m > itr do</p> <p style="padding-left: 20px;">for each x_i distance with all centers C_j is calculated</p> <p style="padding-left: 20px;">Assign x_i to the nearest C_j</p> <p style="padding-left: 20px;">calculate new centroid</p> $v_i = \left(\frac{1}{ c_i } \sum_{i=1}^{c_i} x_i \right)$ <p>3: end while</p> <p>4: Stop</p>

2.3 기존 연구

최근 빅 데이터를 위한 클러스터링 알고리즘이 분산 및 병행 계산에 기반을 두어서 제안되고 왔다. C. Xialoli et. al. 은 반복적인 카운터로 선언된 MapReduce을 사용하여 K-평균을 최적화한 빅 데이터를 제안하였다[11]. 이 기법은 3개의 MapReduce(MR)를 사용하여 K-평균을 최적화한다. 그러나, 이 기법은 첫 번째 MR 과 마지막 MR에서 샘플링 기술이 사용되었고, 데이터 셋은 Voronoi 다이어그램을 사용하여 centroid에 매핑하였다. 샘플링 기술은 빅 데이터에서 중요한 특징이기 때문에 샘플링 기술에서는 효율성을 향상시키기 위해서 데이터 셋을 사용한다[6]. W. Zhao et. al.은 MR에 기반을 둔 병렬 K-Means 알고리즘을 제안하였다[12]. 그러나 이 기법은 초기 시드 선택이 랜덤하게 선택되는 문제점이 있다. 특히, 이 알고리즘은 다중 반복에 의해서 영향을 받는 MR jobs을 사용한다. Bahmani et al.은 MR에 기반을 둔 K-Means 클러스터링 알고리즘을 제안하였다[13]. 그리고, Cai et. al.은 큰 크기의 이질적인 데이터 셋을 처리하기 위해서 K-Means을 수정하였다[14]. 기존 기법에서는 알고리즘을 시작하기 전에 클러스터 수를 명시해야 하며, 최적화된 클러스터를 얻기 위해서 다중적인 동작이 요구된다. [9,15]는 MR에서 DBSCAN 알고리즘을 사용하였지만, 일부 연구자들은 성능을 향상시키기 위해서 GPU 클러스터와 플랫폼을 사용하여 규모가 큰 데이터 셋을 클러스터 하였다[7,8].

3. 딥 러닝 기반의 IoT 헬스케어 서브넷 구축 기법

이 절에서는 IoT 장치를 통해 전송되는 다량의 헬스케어 정보를 서버에서 효율적으로 수집 및 처리하기 위한 딥 러닝 기반의 IoT 헬스케어 정보 관리 기법을 제안한다. 제안 기법은 서버에 전달된 헬스케어 정보에 속성 값을 부여하여 계층적 서브넷을 구축할 때 서브넷의 시드들의 연계 정보를 계층적으로 그룹화한다. 계층화된 연계 정보는 딥 러닝 알고리즘을 통해 사용자의 치료 및 처방에 대한 관찰 속도 및 정확도를 향상시킨다.

3.1 개요

최근 IoT 기술이 발달함에 따라 IoT 기술을 활용한 의료서비스는 새로운 모습으로 변화하고 있다. 특히, 의료 서비스에 사용되는 IoT 센서는 사용자 신체에 손쉽게 탈부착이 가능하며, 시간과 장소에 상관없이 사용자의 헬스케어 정보를 실시간으로 센싱하여 서버에 전달한다. IoT 기반의 헬스케어 서비스는 여러 가지 요소 기술들이 통합되어 새로운 의료 서비스를 제공해야하기 때문에 IoT 장치에서 발생하는 헬스케어 정보의 수집 및 분석을 최소 시간 안에 정확하게 처리해야 하는 문제점이 발생한다.

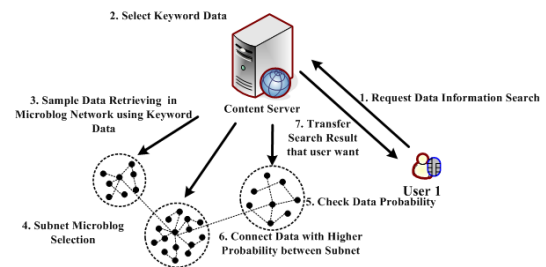
본 논문에서는 사용자의 신체에 부착된 IoT 장치를 통해 수집된 다량의 헬스케어 정보를 유·무선 장치를 통해 의료진에게 전달하며, 의료진은 수집된 다량의 사용자 헬스케어 정보를 빠른 시간 안에 정확한 사용자 건강상태를 확인하여 진료 및 예방 서비스를 효과적으로 진행할 수 있도록 딥 러닝 기반의 IoT 헬스케어 정보 관리 기법을 제안한다. 제안 기법은 사용자의 헬스케어 정보 중 치료와 관리를 위한 정보를 정확하게 하면서도 빠르게 찾기 위해서 속성 정보에 따른 확률정보를 서버넷을 구성하는데 사용하여 의료 서비스를 제공하기 위한 헬스케어 최종 정보를 선택한다.

제안 기법은 IoT 장비를 통해 수집된 헬스케어 정보의 종류, 기능 및 특징에 따라 유사 정보를 서버넷으로 묶어 계층적으로 관리함으로써 헬스케어 정보 수집 및 분석 속도를 향상시킨다. 계층적으로 묶어 있는 서버넷 정보는 딥 러닝 알고리즘을 서로 연계하여 대규모의 헬스케어 서비스 환경에서도 적용할 수 있다. 제안 기법에서는 딥 러닝 알고리즘을 헬스케어 정보의 다양한 속성 정보에 결합하여 계층적으로 관리된 정보 중 유사 정보를 클러스터링 함으로써 질병 정보에 대한 의료 서비스 처방 및 관리 방안을 정확하면서도 단시간에 결정할 수 있다.

3.2 IoT 장비에서 헬스케어 정보 수집

제안 기법은 사용자에게 부착된 IoT 장치와 서버간 헬스케어 정보를 전송하는데 사용되는 채널을 제3자로부터 안전하다고 가정한다. 이 때, 헬스케어 정보는 IoT 장치에서 서버로 전송할 때 헬스케어 정보의 종류와 특성에 따라 헬스케어 정보를 IoT 장치에서 자동 분류한다. 제안 기법은 IoT 장치로부터 전달된 헬스케어 정보를 분

산 처리될 수 있도록 서버넷을 계층적으로 구성한다. 서버넷을 구성한 헬스케어 정보는 속성정보와 함께 연계 처리할 수 있도록 계층적 레벨에 따라 딥 러닝 알고리즘을 적용한다. 제안 기법에 적용된 딥 러닝 알고리즘은 기존 딥 러닝 알고리즘을 확장하여 다중 사용자의 정보를 분산 처리하여 사용자 건강상태와 일치하는 정보를 정확하게 추출한다. 추출된 정보는 계층적 분산 처리가 가능하여 사용자의 건강관리를 위한 치료 시간 및 헬스케어 의료 서비스에 대한 정확도를 향상시킬 수 있는 특징이 있다.

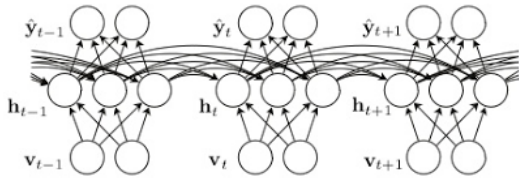


[Fig. 2] Overall Process of Proposed Scheme

[Fig. 2]은 제안 기법에서 동작되는 헬스케어 서비스의 전체 동작과정을 보여주고 있다. [Fig. 2]에서 제안 기법은 대규모의 헬스케어 정보를 속성정보와 함께 계층적으로 분산 처리하기 위해서 헬스케어 정보를 서버넷으로 구성한다. 구성된 서버넷의 크기는 헬스케어 정보의 종류, 기능 및 특징에 따라 연계정보를 (확률값, 서버넷 정보) 쌍으로 구성하여 클러스터링 되기 때문에 자동으로 결정된다. 제안 기법에서는 이 같은 과정을 거쳐 사용자의 헬스케어 정보 수집 및 분석 속도를 향상시킨다. 또한, 의료 서비스 처방 및 관리 방안을 정확하면서도 단시간에 결정할 수 있다.

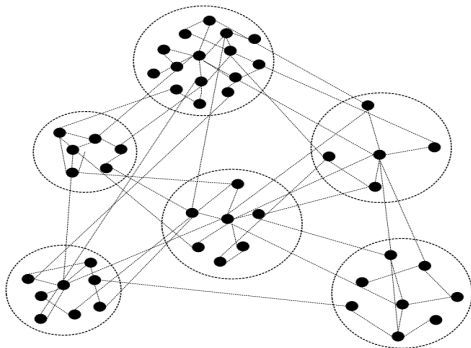
3.3 헬스케어 연계정보를 이용한 서버넷 구성

IoT 장치에서 수집된 헬스케어 정보는 [Fig. 3]처럼 딥 러닝을 이용하여 사용자별 의료 목적에 따라 속성정보 v_t 를 헬스케어 정보의 유사정도 h_t 에 따라 시드 \hat{y}_t 를 선택하여 서버넷을 구성한다. 이 때, 서버넷은 서버넷 간 연계 정보를 확률값의 우선 순위에 따라 시드를 선택적으로 선별하고 헬스케어 정보의 유사정보를 [Fig. 3]과 같이 해쉬체인으로 묶는다.



[Fig. 3] Subnet Component Selection using Deep Learning Algorithm

제안 기법은 [Fig. 3]에 의해 선택된 항목들에 따라 헬스케어 정보와 속성 정보를 해쉬 체인으로 묶어 서브넷으로 구축한 후 서브넷 간 연계 정보들을 [Fig. 4]처럼 계층적 구조로 그룹화 한다. 이 때, 서버에서는 데이터베이스에 저장되어 있던 다량의 헬스케어 정보를 기반으로 사용자의 치료 및 처방에 대한 관찰 속도 및 정확도를 향상시킬 수 있도록 최적화된 정보를 추출한다.



[Fig. 4] Subnet Structure of Proposed Scheme

서브넷간 연계 정보들을 시드로 추출하여 계층적 구조로 그룹화하기 위해서는 식 (1)처럼 속성 값과 연결 정도 값을 쌍으로 하는 헬스케어 연계 정보를 생성한다.

$$CI_k = \begin{pmatrix} x_{00} & \dots & x_{0j} \\ \dots & \dots & \dots \\ x_{i0} & \dots & x_{ij} \end{pmatrix}, k=1,2,\dots, n \quad \text{식 (1)}$$

여기서, i 는 데이터 속성 정보를 나타내며, j 는 데이터간 연결 정도를 의미한다.

헬스케어 연결정보에 사용되는 연결 정도 값(DCI , Data Connection Information)은 식 (2)처럼 백분율로 계산된다.

$$DCI = \frac{P(IV)}{P(IV)P(C)} \quad \text{식 (2)}$$

여기서 IV 는 데이터의 중요도를 의미하고, C 는 데이터 사용 횟수를 의미한다.

식 (2)은 헬스케어 정보의 정확도에 영향을 미치며 사용자의 치료 및 건강관리에 필요한 정보 탐색 시간을 최소화하는데 사용된다.

4. 성능 평가

이 절에서는 헬스케어 정보 수집을 위한 서버의 의료 서비스 처리속도, 오버헤드, 헬스케어 정보 추출에 대한 정확도 등을 평가한다.

4.1 실험환경

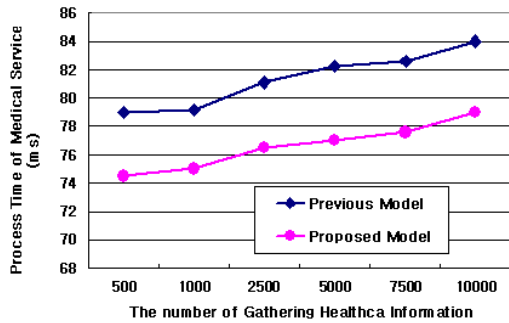
제안 기법의 성능 평가를 위해 <Table 2>와 같은 실험 환경을 적용하였다.

<Table 2> Simulation Setting

Parameter	Setting
Number of Subnet	$s = \{1, 3, 5, 10\}$
Number of Senced Healthcare Information	$d = \{500, 1,000, 2,500, 5,000, 7,500, 10,000\}$
Number of Property	$p = \{1, 2, 3, 4, 5\}$
Buffer	50 packet
Traffic	4 pkts/s

4.2 의료서비스 처리시간

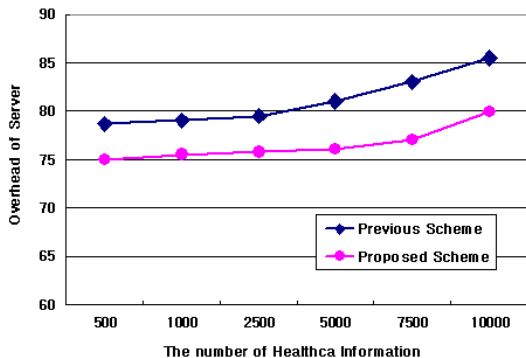
[Fig. 5]의 의료서비스 처리시간은 IoT 장치로부터 전달받은 사용자의 헬스케어 정보를 딥 러닝 알고리즘에 의해서 계층적 속성과 확률연계 값에 따라 서브넷을 구축하여 사용자에게 필요한 의료 정보를 추출하는데 필요한 처리시간을 나타내고 있다. [Fig. 5]의 실험결과, 서브넷간 서로 연계된 연계 정보로 딥 러닝 알고리즘을 적용한 제안 기법이 기존 기법보다 처리시간이 14.1% 향상되었다. 이 같은 결과는 제안 기법이 헬스케어 정보의 유사 정도에 따라 서브넷을 구성한 후 서브넷간 서로 연계된 정보를 확률 정보 값으로 나타내어 딥 러닝 알고리즘에 적용하였기 때문에 나타난 결과이다.



[Fig. 5] Process Time of Medical Service

4.3 오버헤드

[Fig. 6]의 오버헤드는 헬스케어 정보를 확률 연계정보로 클러스터링 함으로써 헬스케어 정보 증가에 따른 서버의 오버헤드 변화를 나타내고 있다.



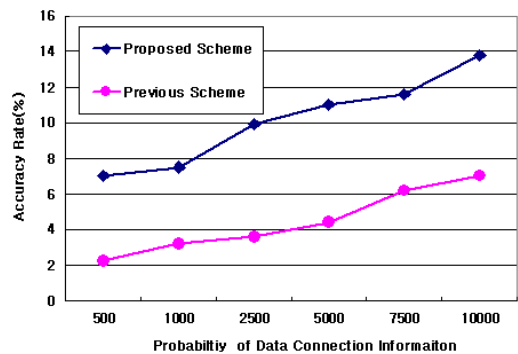
[Fig. 6] Overhead of Server through the Number of Healthcare Information

[Fig. 6]의 결과, 헬스케어 정보를 확률 연계 정보로 딥 러닝 알고리즘에 적용한 서버의 오버헤드가 기존 기법보다 평균 6.7% 낮게 나타났다. 이 같은 결과는 제안 기법이 헬스케어 정보의 종류, 기능 및 특징에 따라 연계정보를 서버에서 (확률값, 서버넷 정보) 쌍으로 구성하여 클러스터링 하도록 설정한 것과 확률정보가 높은 시드를 중심으로 딥 러닝 알고리즘에 헬스케어 정보를 적용하였기 때문에 나타난 결과이다.

4.4 헬스케어 정보 추출에 대한 정확도

[Fig. 7]의 헬스케어 정보 추출에 대한 정확도는 서버

에서 수집된 사용자의 헬스케어 정보를 데이터베이스에 구축된 헬스케어 정보와 비교하여 사용자의 헬스케어에 필요한 치료 및 예방관리에 적합한 정보 추출에 대한 정확도를 보여주는 결과이다. [Fig. 7]의 실험결과, 확률적 연계도를 딥 러닝에 적용하여 사용자의 헬스케어 치료 정보의 정확도를 평가한 제안기법이 기존기법보다 정보 추출에 대한 정확도가 평균 10.1% 높게 나타났다. 이 같은 결과는 헬스케어 정보와 속성 정보의 확률 값을 계층적으로 구성하여 서버넷의 시드 값을 유추하여 딥 러닝 알고리즘에 적용하였기 때문에 나타난 결과이다.



[Fig. 7] Extraction Accuracy of Healthcare Information

5. 결과

IT 기술이 발달함에 따라 헬스케어 분야에서 빅 데이터와 관련된 여러 기술들에 대한 관심이 증가하고 있다. 특히, 질병 발생 전에 환자의 건강을 효율적으로 관리하기 위해서 아파치 스파크 플랫폼 기반의 빅 데이터 처리 기술들이 확대 적용하고 있다. 본 논문에서는 사용자 신체에 센서가 부착된 IoT 장비(e.g. 팔찌, 목걸이 등)와 휴대폰을 이용하여 사용자의 건강정보를 효율적으로 관리할 수 있는 딥 러닝 기반의 헬스케어 서비스 관리 기법을 제안하였다. 성능평가 결과, 제안기법은 헬스케어 서비스 모델에서 동작되는 의료 서비스의 처리속도가 기존기법에 비해 평균 14.1% 향상되었고, 서버의 오버헤드는 기존 기법보다 평균 6.7% 낮았다. 헬스케어 정보 추출에 대한 정확도는 기존기법보다 평균 10.1% 높은 결과를 얻었다. 향후 연구에서는 본 연구의 결과를 기반으로 헬스케

어 정보를 가상현실 기술과 접목하여 응급 구조 서비스 기술에 적용할 계획이다.

ACKNOWLEDGMENTS

This work was supported by the Security Engineering Research Center granted by the Ministry of Trade, Industry and Energy.

REFERENCES

- [1] V. Gadepally, J. Kepner. "Big data dimensional analysis", 2014 IEEE High Performance Extreme Computing Conference(HPEC) pp. 1-6, Sep. 2014.
- [2] Y. Demchenko, C. De Laat, P. Membrey, "Defining architecture components of the Big data Ecosystem", 2014 International conference on Collaboration Technologies and Systems(CTS), pp.104-112, May, 2014.
- [3] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker and I. Stoica, "Spark: cluster computing with working sets," Proceedings of the 2nd USENIX conference on Hot topics in cloud computing (HotCloud'10), pp. 10-10, Jun. 2010.
- [4] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker and I. Stoica, "Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing", Proceeding of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI'12), pp. 2-2, April. 2012.
- [5] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. H. Byers, "Big Data: The Next Frontier for Innovation, Competition and Productivity", Mckinsey Global Institute, pp. 1-137. 2011.
- [6] W. Yan, U. Brahmakshatriya, Y. Xue, M. Gilder and B. Wise, "p-Pic: Parallel Power iteration clustering for big data", Journal of Parallel and Distributed computing, Vol. 73, No. 3, pp. 353-359, Mar. 2013.
- [7] X. Cui, J. S. Charles and T. Potok, "GPU enhanced parallel computing for large scale data clustering", Journal of Future Generation Computer Systems, Vol. 29, No. 7, pp. 1736-1741, Sep. 2013.
- [8] G. Andrade, G. Ramos, D. Madeira, R. Sachetto, R. Ferreira and L. Rocha, "G-dbscan: A gpu accelerated algorithm for density based clustering", Journal of Procedia Computer Science, Vol. 18, pp. 369-378, Aug. 2013.
- [9] H. Karau, A. Konwinski, P. Wendell and M. Zaharia, "Learning Spark: Lightning-Fast Big Data Analysis", O'Reilly Media, Inc. pp. 1-276, Jan. 2015.
- [10] A. Katal, M. Wazid, R. H. Goudar, "Big data: Issues, challenges, tools and Good practices ", 2013 Sixth International Conference on Contemporary Computing(IC3), pp. 404-409, Aug. 2013.
- [11] X. Cui, P. Zhu, X. Yang, K. Li and C. Ji, "Optimized big data K-means clustering using MapReduce", Journal of Supercomputing, Vol. 70, Issue 3, pp. 1249-1259, Dec. 2014.
- [12] W. Zhao, H. Ma and Q. He, "Parallel k-means clustering based on mapreduce", Proceedings of the 1st International Conference on Cloud Computing, pp. 674-679, Dec. 2009.
- [13] B. Bahmani, B. Moseley, A. Vattani, R. Kumar and S. Vassivitskii, "Scalable k-means++", Proceedings of the VLDB Endowment, Vol. 5, No. 7, pp. 622-633, 2012.
- [14] X. Cai, F. Nie and H. Huang, "Multi-view k-means clustering on big data", Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, AAAI Press, pp. 2598-2604, 2013.
- [15] Y. H. Kim, K. S. Shim, M. S. Kim and J. S. Lee, "DBCURE-MR: an efficient density-based clustering algorithm for large data using MapReduce", Journal of Information Systems, Vol. 42, pp. 15-35, Jun. 2014.
- [16] Y. S. Jeong, "An Efficiency Management Scheme using Big Data of Healthcare Patients using Puzzy AHP," Journal of Digital Convergence, Vol. 13, No. 4, pp. 227-234, Apr. 2010.

- [17] Y. S. Jeong, "Design of Prevention Model according to a Dysfunctional of Corporate Information," Journal of Convergence Society for SMB, Vol. 6, No. 2, pp. 11-17, Jun. 2016.
- [18] Y. S. Jeong, "Tracking Analysis of User Privacy Damage using Smartphone", Journal of Convergence Society for SMB, Vol. 4, No. 4, Dec. 2014.
- [19] Y. S. Jeong, "Design of Security Model for Service of Company Information," Journal of Convergence Society for SMB, Vol. 2, No. 2, pp. 43-49, Nov. 2012.
- [20] S. H. Lee and D. W. Lee, "A Study on u-Health Fusion Field based on Internet of Thing," Journal of the Korea Convergence Society, Vol. 7, No. 4, pp. 19-24, Aug. 2016.
- [21] G. W. Jin, "A Study on the Data Colection Methods based Hadoop Distributed Environment," Journal of the Korea Convergence Society, Vol. 7, No. 5, pp. 1-6, Oct. 2016.

정 윤 수(Jeong, Yoon Su)



- 2000년 2월 : 충북대학교 전자계산학과 이학석사
- 2008년 2월 : 충북대학교 전자계산학과 이학박사
- 2009년 8월 ~ 2012년 2월 : 한남대학교 산업기술연구소 전임연구원
- 2012년 3월 ~ 현재 : 목원대학교 정보통신융합공학부 조교수

- 관심분야 : 유·무선통신 보안, 정보보호, 센서 네트워크, IoT, 이동통신, 암호이론,
- E-Mail : bukmunro@mokwon.ac.kr