

Generating Training Dataset of Machine Learning Model for Context-Awareness in a Health Status Notification Service

Jong Hyeok Mun[†] · Jong Sun Choi^{††} · Jae Young Choi^{†††}

ABSTRACT

In the context-aware system, rule-based AI technology has been used in the abstraction process for getting context information. However, the rules are complicated by the diversification of user requirements for the service and also data usage is increased. Therefore, there are some technical limitations to maintain rule-based models and to process unstructured data. To overcome these limitations, many studies have applied machine learning techniques to Context-aware systems. In order to utilize this machine learning-based model in the context-aware system, a management process of periodically injecting training data is required. In the previous study on the machine learning based context awareness system, a series of management processes such as the generation and provision of learning data for operating several machine learning models were considered, but the method was limited to the applied system. In this paper, we propose a training data generating method of a machine learning model to extend the machine learning based context-aware system. The proposed method define the training data generating model that can reflect the requirements of the machine learning models and generate the training data for each machine learning model. In the experiment, the training data generating model is defined based on the training data generating schema of the cardiac status analysis model for older in health status notification service, and the training data is generated by applying the model defined in the real environment of the software. In addition, it shows the process of comparing the accuracy by learning the training data generated in the machine learning model, and applied to verify the validity of the generated learning data.

Keywords : Context-Awareness, Machine Learning Model, Generating Training Dataset, Maintaining Accuracy

사용자 건강 상태알림 서비스의 상황인지를 위한 기계학습 모델의 학습 데이터 생성 방법

문종혁[†] · 최종선^{††} · 최재영^{†††}

요약

다양한 분야에서 활용되는 상황인지 시스템은 상황정보를 획득하기 위한 추상화 과정에서 규칙 기반의 인공지능 기술이 기존에 사용되었다. 그러나 서비스에 대한 사용자의 요구사항이 다양해지고 사용되는 데이터의 증대로 규칙이 복잡해지면서 규칙 기반 모델의 유지보수와 비정형 데이터를 처리하는데 어려움이 있다. 이러한 한계점을 극복하기 위해 많은 연구들에서는 상황인지 시스템에 기계학습 기술을 적용하였으며, 이러한 기계학습 기반의 모델을 상황인지 시스템에 사용하기 위해서는 주기적으로 학습 데이터를 제공해야 한다. 이에 기계학습 기반 상황인지 시스템에 대한 선행연구에서는 여러 개의 기계학습 모델을 적용하기 위한 학습 데이터 생성, 제공 등의 과정을 보였으나 제한된 종류의 기계학습 모델만을 적용 가능하여 확장성이 고려되어야 한다. 본 논문은 기계학습 기반의 상황인지 시스템의 확장성을 고려한 기계학습 모델의 학습 데이터 생성 방법을 제안한다. 제안하는 방법은 시스템의 확장성을 고려하여 기계학습 모델의 요구사항을 반영할 수 있는 학습 데이터 생성 모델을 정의하고 학습 데이터 생성 모듈을 바탕으로 각각의 기계학습 모델의 학습 데이터를 생성하는 것이다. 시스템의 확장성의 검증을 위해 실험에서는 노인의 건강 상태 알림 서비스를 위한 심박상태 분석 모델을 대상으로 한 학습데이터 생성 스키마를 기반으로 학습데이터 생성 모델을 정의하고 실험경에서 정의된 모델을 S/W에 적용하여 학습데이터를 생성한다. 또한 생성된 학습데이터의 유효성을 검증하기 위해 사용되는 기계학습 모델에 생성한 학습데이터를 학습시켜 정확도를 비교하는 과정을 보인다.

키워드 : 상황인지, 기계학습 모델, 학습데이터 생성, 정확도 유지

* 이 논문은 2019년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.2019R1A2C1007861).

[†] 준회원: 송실대학교 컴퓨터학과 박사과정

^{††} 정회원: 송실대학교 컴퓨터학부 조교수

^{†††} 종신회원: 송실대학교 컴퓨터학부 교수

Manuscript Received : October 11, 2019

First Revision : November 13, 2019

Accepted : December 13, 2019

* Corresponding Author : Jong Sun Choi(jongsun.choi@ssu.ac.kr)

1. 서론

최근 스마트 홈, u-헬스케어, 스마트 웨어하우스(Warehouse) 등 많은 분야에서 상황인지 시스템을 활용하기 위한 연구가 시도되고 있다[1-4]. Khozouie, Nasim[4]은 온톨로지 기반의 모바일 헬스케어 서비스를 제공하기 위해 상황인지 시스

템을 활용하였고 Gomes[12], Kabir[14] 등은 스마트 홈에서 상황인지 기반의 사용자 맞춤형 스마트 서비스를 위한 연구를 진행하였다. 이러한 연구들은 상황인지 시스템에서 상황정보를 획득하기 위해 인공지능 기술을 활용하였다. 상황정보는 주변 환경에서 발생된 정보를 토대로 상황을 표현한 정보이며 이는 상황인지 시스템에서 추상화 과정을 통해 얻어진다.

기존 인공지능 기반의 상황인지 시스템은 추상화 과정에서 휴리스틱(Heuristic) 로직 기반의 전문가 시스템(Expert System)을 적용하였다[20-22]. 전문가 시스템은 적용 환경에 따라 전문가가 지식에 근거하여 사전에 정의한 규칙을 바탕으로 수집된 환경정보로부터 상황정보를 생성한다. 상황인지를 위해 전문가 시스템을 적용하려면 규칙 설계자가 적용 분야에 대한 지식을 보유하고 있어야 하며, 이를 기반으로 명확한 규칙을 정의할 수 있어야 한다. 또한 이러한 규칙에는 적용되는 환경에 대한 현상이 반영될 수 있어야 한다[23].

그러나 기존 방식(규칙 기반 인공지능 추상화 시스템)은 최근 연구되는 분야들에서 다음과 같은 문제로 서비스 품질을 보장하기 어렵다. 첫 번째, 사용자의 서비스에 대한 요구사항이 높아지면서 더욱 비정형 데이터의 사용이 요구된다. 이러한 비정형 데이터를 규칙으로 정의하는 것은 많은 비용이 소요되어 이에 비정형 데이터를 처리하기 위한 방법이 필요하다[24]. 두 번째, 헬스케어, 스마트 홈 등의 분야에서는 서비스 대상의 맞춤형 서비스를 제공하기 위해 각각의 대상마다 규칙을 정의하게 된다. 이는 규칙이 증대되어 유지보수 문제를 야기한다[24, 25]. 이는 데이터를 기반으로 한 기계학습 기술을 적용함으로써 극복될 수 있다. 기계학습 기반 모델의 최적화를 위해서는 지속적인 학습이 이루어져야 하므로 기계학습 모델에 학습데이터를 주기적으로 제공할 수 있어야 한다. 따라서 하나 이상의 기계학습 모델이 사용되는 상황인지 시스템[14, 15, 17]은 각기 다른 기계학습 모델의 학습데이터를 생성하여 제공하는 과정이 수행되어야 하며, 이를 위한 관리 방법이 요구된다[13]. 따라서 시스템에서는 학습 데이터를 생성을 위한 모듈이 동작하여 학습데이터를 생성하여 제공할 수 있어야 한다. 이에 논문에서는 상황인지 시스템에 기계학습 모델을 관리하는데 필요한 학습데이터를 생성하는 방법을 제안한다. 제안하는 방법은 각기 다른 요구사항(입력, 전처리, 출력 등)이 다른 기계학습 모델을 위한 학습 데이터 생성 모델을 정의하고 이를 기반으로 학습 데이터 생성 모듈을 적용하여 각각의 기계학습 모델의 학습 데이터를 생성하는 것이다. 학습데이터 생성 모듈은 입력 데이터 수집, 전처리, 적재 단계를 거쳐 학습데이터를 생성하며, 필요에 따라 식별정보를 추가하여 적재한다. 적재된 학습데이터는 상황인지 시스템에서 관리되어 기계학습 모델에 반영될 수 있다.

2. 관련 연구

본 절에서는 상황인지 시스템의 추상화 단계에서 기계학습 기술을 적용한 연구 사례를 살펴보고 각 연구에서 수행된 기

계학습 모델의 관리 방법들을 기술한다.

기계학습은 데이터 자체에 대한 규칙 및 패턴을 학습한다. 이러한 방식은 전문가에 대한 의존성을 줄이고, 학습 데이터를 갱신하는 것으로 알고리즘을 지속적으로 최적화하여 변동이 잦은 서비스에 대한 유연한 대처가 가능하다. 이러한 기계학습 기술은 상황인지 기술과 접목되어 추천 시스템, 지능형 홈, 로봇 등에 다양한 분야에서 활용된다.

Shin[11], Gomes[12]의 연구에서는 각 적용 분야에 대한 상황인지 서비스 제공을 위한 상황인지 시스템을 개발하고 성능 향상과 서비스 품질 개선을 위해 기계학습 기술이 활용되었다. 특히 Gomes는 각 서비스별로 서브 시스템을 개발하고 이를 통해 분석을 수행하였다. 서브 시스템들은 통합 시스템을 통해 추가/변경 등의 관리가 이루어지며 서브 시스템의 하위 소프트웨어인 기계학습 기반 분석 모델 또한 통합 시스템을 통해 관리되어 분석 모델에 대한 확장성을 제공한다.

스마트홈은 대표적으로 상황인지 기술이 활용되는 분야로, 각 환경에 적합한 맞춤형 서비스를 제공하기 위해 서비스 환경에서의 상황을 판단하는데 상황인지 시스템을 사용한다. Bahramian[10], Shin[12], Kabir[14]의 연구에서는 스마트 홈 환경에서 변화의 빈도가 높은 사용자 습관에 따른 맞춤형 서비스를 제공하기 위해 각 사용자의 상황을 판단하는데 기계학습 기술을 사용하였다. 또한 Kabir[14, 15]은 스마트 홈에서 상황을 인식하기 위한 Context를 정의하고 적용하기 위한 방법을 연구하였다. Context는 사용자 식별정보, 날짜, 시간, 위치, 행동 등을 포함한다. 또한 추가 연구로, 정의한 Context를 바탕으로 스마트 홈 서비스 제공을 위한 기계학습 기반 서비스 제공 프로세스를 제안하였다[15]. 여기에 추가적으로 사용자의 패턴 변화를 유연하게 대응하기 위해 사용자로부터 서비스에 대한 피드백 기능을 포함한다. 최종적으로 사용자로부터 피드백을 받아, 이에 따라 학습데이터를 생성하고 기계학습 모델을 재학습함으로써 서비스의 품질을 보장한다.

헬스케어 분야에서 또한 상황인지 기술과 기계학습 기술을 접목하여 서비스 제공하기 위한 연구를 진행되었다. Banos[18]의 연구에서는 이기종 스마트 기기로부터 생체 데이터 수집, 분석, 제공 과정을 수행하는 프레임워크를 제안한다. 생체 데이터의 분석은 기계학습 모델을 기반으로 수행되며, 모델은 전처리 모듈에서 처리된 전처리 데이터를 입력받아 분석을 수행한다. 그러므로 기계학습 모델의 학습 과정에 사용되는 데이터는 전처리 과정에서 생성된다.

또 다른 기계학습을 활용한 헬스케어 서비스 연구로, Zoppi[19]에서는 모바일 헬스케어 프레임워크를 통해 이는 스마트 기기로부터 데이터를 수집하고 데이터를 분석하여 헬스케어 서비스를 제공하는 과정을 보인다. 데이터 분석에는 생체 분석을 위한 기계학습 모델이 적용되었으며, 기계학습 모델의 학습을 위해 수집된 데이터를 분류하는 모델을 구축하고 이를 통해 학습데이터를 생성하여 피드백 과정을 수행하였다.

상황인지 시스템에 사용되는 기계학습 모델은 서비스 환경이 다양한 외부 요인으로 인해 변동됨에 따라 정확도가 감소할 수 있다. 특히 헬스케어 분야에서는 사용자 맞춤형 서비스의

적용되는 기계학습 모델의 경우, 사용자의 상태, 습관, 생활 패턴 등의 변동으로 잦은 규칙의 변동이 발생할 수 있다. 그러므로 기계학습 모델의 정확도를 유지하기 위해서는 기계학습 모델의 학습 데이터를 생성, 제공, 학습 등의 관리가 필요하다. 그러나, 선행연구들은 시스템에 적용되는 기계학습 모델을 위한 관리 방법을 제안하였고 이는 기계학습 모델의 사용성과 시스템의 확장성을 한정시킨다. 이는 각 특성 및 목적이 다른 기계학습 모델을 추가하여 사용하기엔 어려움이 있다.

3. 기계학습 모델의 학습 데이터 생성 방법

상황인지 시스템에 기계학습 기술을 적용하기 위해서는 기계학습 모델의 정확도 유지를 위한 관리가 수행되어야 한다. 따라서 시스템에서는 기계학습 모델이 추가 및 변경됨에 따라 이를 위한 학습데이터 생성 모듈이 적용되어야한다. 본 절에서는 인공지능 기반 상황인지 시스템의 기계학습 모델 관리를 위한 학습데이터 생성 모듈의 학습 데이터 생성 방법을 기술한다. 시스템의 확장성을 고려한 시스템의 요구사항은 다음과 같다.

- 1) 수집되는 환경 데이터의 품질 보장
- 2) 추상화 과정에서 기계학습 모델 사용
- 3) 두 개 이상의 기계학습 모델을 적용
- 4) 운용되는 기계학습 모델은 제안하는 학습데이터 생성 모듈로부터 학습데이터를 제공 받음

3.1 기계학습 모델 기반 상황인지 시스템 개요

기계학습 기반의 상황인지 시스템은 환경 데이터 수집, 추상화, 상황인지 과정을 수행하며 추상화 과정에서 기계학습 모델이 사용된다. 이때, 기계학습 모델은 시스템상에서 지속적인 관리가 필요하므로 학습데이터 생성 모듈을 통해 학습 데이터를 제공받는다. 기계학습 기반의 상황인지 시스템의 개요는 Fig. 1과 같다.

- 1) 환경정보 수집 모듈
환경정보 수집 모듈(Environment Data Collecting Module)

은 주변 센서, 카메라, 웨어러블 기기 등과 같은 IoT 기기로부터 상황 판단에 사용되는 주변 환경에서 발생하는 데이터인 환경 데이터를 수집한다. 수집된 데이터는 수집 데이터 검증 모듈(Validator)을 통해 유효성을 검사하여 수집 데이터의 품질을 보장한다. 환경정보 제공 모듈(Environment Data Provider)는 검증된 데이터를 빅데이터 분석 모듈(Big data Analysis Module)과 저장소에 전달한다.

2) 빅데이터 분석 모듈

기계학습 모델을 관리하고 추상화를 수행하는 빅데이터 분석 모듈(Big data Analysis Module, 이하 BDAM)은 상황을 인지하는데 사용되는 상황정보를 생성하는 역할을 수행한다. BDAM의 추상화의 핵심은 기계학습 모델을 통한 환경 데이터 분석으로, 데이터와 상황정보의 특성에 따라 여러 기계학습 모델이 사용된다. 이러한 모델의 분석 결과의 정확도를 유지하기 위해 주기적으로 학습 데이터셋 생성 모듈(Training Dataset Generating Module)로부터 최신의 학습 데이터셋을 제공받아 학습을 수행한다.

3) 상황인지 엔진

상황인지 엔진(Context-aware Engine)은 상황정보를 바탕으로 상황을 인지하고 적절한 상황인지 서비스를 제공한다. 상황인지에 사용되는 상황정보는 BDAM의 기계학습 모델에서 제공받은 분석 정보이다. 본 논문에서는 상황인지 시스템에서 분석 주체인 BDAM의 기계학습 모델을 위한 학습데이터를 학습 데이터셋 생성 모듈(Training Dataset Generating Module, 이하 TDGM)을 통해 생성하는 방법을 보인다.

3.2 학습데이터셋 생성 모듈

학습데이터 생성 모듈(Training Dataset Generating Module)은 BDAM에서 분석을 수행하는 기계학습 모델의 학습을 위한 학습데이터를 생성하며 구성요소는 Fig. 2와 같다. 기계학습 모델에서 요구하는 학습데이터의 구성 파라미터, 전처리, 조건은 각기 다르며, 이러한 요구사항을 반영하기 위해 학습데이터 생성 객체(Dataset Generating Object, 이

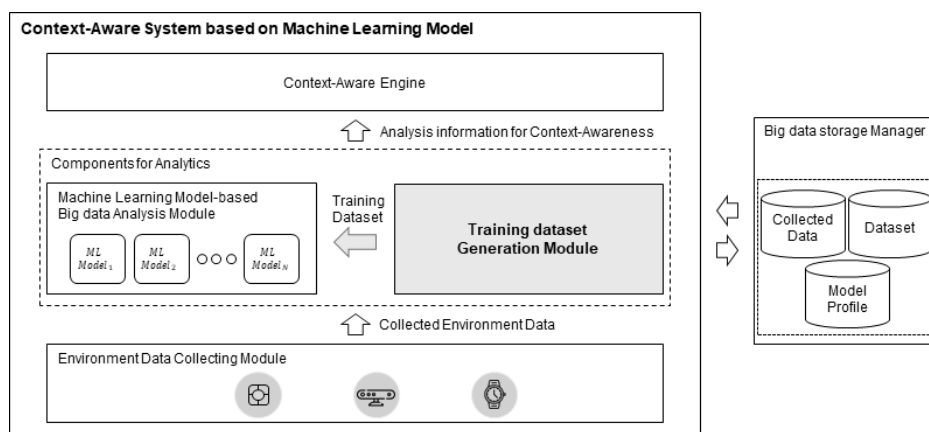


Fig. 1. Overview of Context-aware System Based on Machine Learning

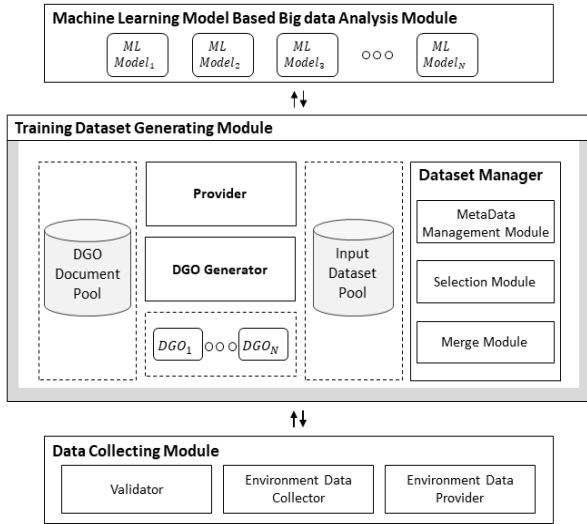


Fig. 2. Architecture of Training Dataset Generation Module

하 DGO)를 제어한다. DGO는 요구사항에 따른 학습데이터를 생성하며 이는 입력 데이터셋 풀(Input Dataset Pool)에 저장된다. 또한 학습데이터 생성 모델(DGM)을 기반으로 동작하며 이는 DGM 세부내용은 3.3절에서 다룬다.

DGO 관리자(DGO Manager)는 DGO를 운용하기 위한 자원을 관리한다. DGO 문서 풀(DGM Document Pool)에 새로운 DGM이 적재되면 DGO를 생성하여 동작하기 위해 자원을 할당하고 작업이 완료된 DGO의 할당 자원을 해제하는 역할을 수행한다. 입력 데이터셋 풀은 DGO가 생성한 학습데이터를 저장하며 데이터셋 관리자(Dataset Manager)는 이를 기계학습 모델을 학습하기 위한 형태로 변환한다. 데이터셋 관리자의 구성 모듈은 세 가지이다. 메타 데이터 관리 모듈(Meta Data Management Module)은 DGO의 기반인 DGM를 메타데이터로 학습데이터셋에 입력하는 모듈로, 학습데이터 생성 조건, 설명, 파라미터의 스토리지 접근 정보를 추가 기재한다. 선택 모듈(Selection Module)과 병합 모듈(Merge Module)은 생성된 학습데이터셋의 각 학습데이터의 분류 정보(Labeled Information)를 추가하기 위한 모듈로 선택 모듈을 통해 특정 학습데이터를 선택하고 이에 식별 정보를 각 학습데이터에 추가한다. 변환된 학습데이터셋은 Provider를 통해 기계학습 기반 분석 모듈에 제공된다.

3.3 학습데이터 생성 과정

기계학습 모델의 요구사항에 따라 학습데이터셋을 생성하기 위한 학습데이터 생성 모델(DGM)의 스키마는 Fig. 3와 같다. DGM의 스키마는 DGM 식별자, 데이터셋 정보, 생성 및 종료 조건, 전처리 등의 속성을 포함한다. 스키마의 각 속성에 대한 설명은 Table 1과 같다.

DGO는 DGM을 기반으로 하여 Fig. 3의 스키마를 따른다. 또한 DGM의 기술된 정보를 통해 학습데이터 생성부터 전처리, 객체의 실행·종료 등의 구동을 관리한다. 이러한 기능을 위한 DGO의 구성 모듈은 Fig. 4와 같다. DGO의 각

Table 1. Elements of Dataset Generation Model Schema

Element	Description
dgmId	DGM identifier
dataset	Generating dataset file name and description
observations	Configuration parameters for dataset and storage access information
observation	Observation contain index, term for storage access, it distribute parameters through using term while training dataset generate
label_option	Option training dataset whether it include labeled information
execution	It is information for executing while DGO is generated based on DGM, it contain control information such as set up number of dataset or interval
event	Event attribute contain interval, equation etc parameters for dataset generating
preprocess	It describes path of document for preprocessing. this document contain for preprocessing information when dataset generate

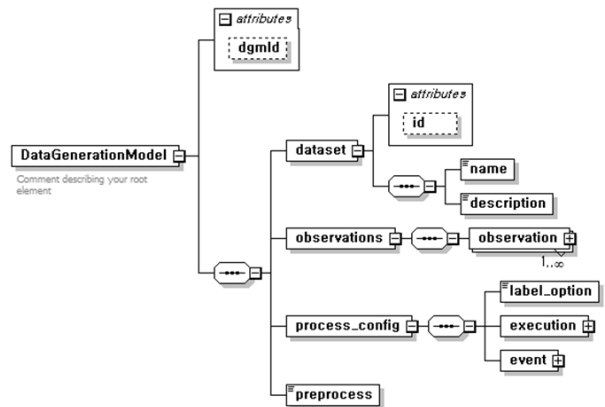


Fig. 3. Schema of Dataset Generation Model

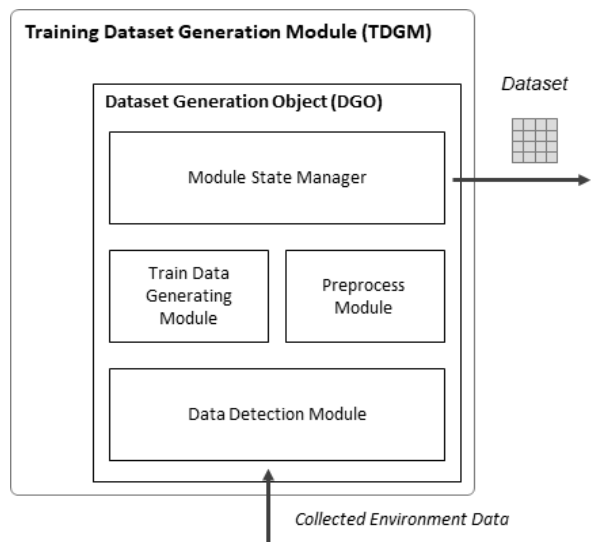


Fig. 4. Architecture of Dataset Generation Object

모듈은 Table 1에 기술된 정보를 바탕으로 동작한다. 데이터 감지 모듈(Data Detection Module)은 execution의 정보를 통해 조건에 부합하는 수집 정보를 획득한다.

학습 데이터 생성 모듈(Train Data Generating Module)은 Observations의 term를 바탕으로 테이블을 구성하고 수집 정보를 전처리 등의 처리과정을 거쳐 얻은 정보를 테이블에 적재하는 역할을 수행한다. 전처리 모듈(Preprocess Module)은 수집 정보를 전처리하는 모듈로써, preprocess에 있는 Document의 내용을 바탕으로 전처리 과정을 수행한다. 마지막으로 모듈 상태 관리자(Module State Manager)는 execution 정보를 토대로 DGO의 작업 완료 조건을 판단하여 기준에 따라 작업 완료시 프로세스 종료한다. DGO를 통한 학습데이터 생성 과정은 학습데이터를 구성할 데이터를 획득, 획득한 데이터를 전처리, 데이터셋에 적재, 저장하는 순서로 이루어지며 요구사항에 따라 식별정보를 추가하여 분석 모듈에 제공된다(Fig. 5).

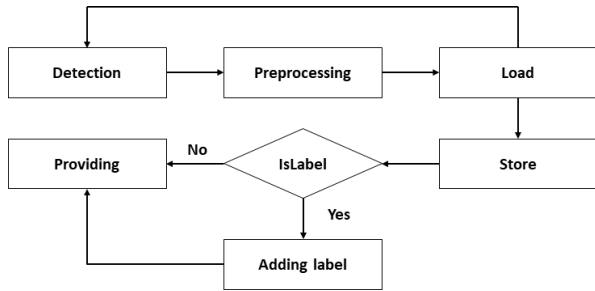


Fig. 5. Flow of Training Dataset Generation

Table 2. Dataset Generating Algorithm

Input : DGM
Output : Target Machine Learning Model Dataset of DGM
Procedure DGO-Execution(DGM)
1: L ← get(DGM.Label) // Label Option
2: DF ← File(DGM.Name) // Dataset File Stored
3: E ← get(DGM.Trigger) // Condition for Selecting parameters from observed data
4: S ← get(DGM.Execution) // Condition of DGO life-cycle
5: Pr ← get(DGM.Pr) // Preprocess Information
6: P[] ← Parameter of Machine Learning Model
7: O[] ← getObservations(DGM)
8: for i ← 0 to O.length -1 do
9: P[i] ← term(O[i])
10: Ds ← Create Matrix with that are column names P
11: if L ← True then
12: add column NL in Ds
13: while isExecutionTime(S) == True do
14: D ← detectEnvironmentData()
15: Line[] ← preprocess(Pr, D)
16: if L ← True then
17: Line[P.length] ← null
18: extend(Ds, Line)
19: write(DF, Ds)
20: store(DF)
end Procedure

Table 2는 DGO에서 학습데이터를 생성하기 위한 알고리즘으로 Fig. 5와 동일한 실행 순서를 갖는다. 라인 1-9는 DGM 문서로부터 학습데이터 생성 정보를 초기화하는 영역으로 데이터셋 프로파일, 파라미터, 구동 조건, 학습데이터 구성 조건, 전처리 등의 정보를 메모리로 적재한다.

라인 10-12에서는 생성된 학습데이터를 적재할 테이블 객체를 만든다. 테이블 객체는 DGM의 구성 파라미터를 바탕으로 생성되며 식별정보 입력 조건에 따라 식별정보 속성이 추가된다. 데이터 획득, 전처리, 적재 과정은 라인 13-17에서 수행되어 객체(Object) 형태로 메모리에 적재된다. 메모리에 적재된 객체는 라인 18-20에 따라 파일 형태의 학습데이터셋으로 생성된다.

4. 실험

본 실험에서는 시스템의 확장성을 고려한 학습데이터 생성 방법의 유효성 증명하기 위해, 노인을 위한 건강 상태 알림 서비스를 위한 심박상태 분석 모델 대상으로 한 예제를 보인다. 그리고 심박상태 분석 모델을 위한 학습데이터 생성 과정과 생성한 학습데이터를 적용한 결과를 보인다.

4.1 실험 환경

상황인지 시스템에서 기계학습 모델을 적용하기 위해서는 각 기계학습 모델의 학습데이터 생성 모듈을 구동하여 학습데이터를 제공할 수 있어야 한다. 특히 사용자의 습관, 생활 패턴, 생체 추이 등과 같이 규칙이 변동이 발생하는 맞춤형 서비스에서 필요하다. 이에 본 절에서는 노인의 건강 상태 알림 서비스를 위한 심박상태 분석 모델을 한정으로 실험을 보인다. 먼저 노인의 건강 상태 알림 서비스 내용은 다음과 같다.

- 1) 노인은 스마트워치 착용하고 있으며, 스마트워치는 사용자의 심박수,스트레스, 시간 등의 정보를 제공한다.
- 2) 상황인지 시스템은 스마트워치로부터 건강정보를 수집하고 심박상태 분석 모델을 통해 심박 상태를 분석한다.
- 3) 심박상태 분석 결과는 노인의 상태가 정상/비정상인지 판별하여 모니터링 정보로 제공한다.

본 실험의 대상인 심박 상태 분석 모델은 노인의 심박수, 스트레스 등의 건강 정보를 입력받아 심박 상태에 대한 분석 결과(정상, 이상, 위험)를 제공한다. 이러한 생체를 대상으로 하는 분석의 경우, 주변 환경과 시간이 지남에 따라 생체 추이가 변하기 때문에 이를 대응할 수 있어야 한다. 이에 심박 상태 분석 모델의 학습을 위한 학습데이터를 실제 노인을 대상으로 건강 정보를 수집하여 학습데이터를 생성하고, 생성된 학습데이터를 대상 모델에 학습시켜 심박 상태 결과에 대한 정확도가 유지되는 것을 보인다.

4.2 학습데이터 생성 모델 기반의 학습데이터셋 생성 과정

첫 번째 실험에서는 심박 상태 분석 모델의 학습데이터셋을 생성하기 위한 모델을 정의하여 학습데이터를 생성한다.

Table 3은 심박 상태 분석 모델의 입력 파라미터, 환경 데이터 구성 조건, 전처리, 생성 기간 등의 정보를 포함한 학습 데이터 생성 모델의 정의이다. 본 실험에서는 총 4회의 걸쳐 학습데이터를 생성하였으며 Table 3는 이중 첫 번째 학습데이터를 생성하는데 사용된 모델이다. 각 회차별 수행 기간은 약 2주이며 총 2개월 학습데이터를 생성하였다. 학습데이터 생성을 위한 자세한 설정 내용은 Table 4와 같다.

Table 3. Training Dataset Generation Model Based Document

```

<DGM dgmId="CSA_DGM">
  <dataset id = "cardiac_state_analysis">
    <name> CSA_dataset201709 </name>
    <description>
      dataset for cardiac_state_analysis model
    </description>
  </dataset>
  <observations>
    <observation id="cardiac_value">
      <index> silver_health </index>
      <document_type> cardiac </document_type>
      <term> cardiac_value </term>
    </observation>
    <observation id="stress_value">
      <index> silver_health </index>
      <document_type> cardiac </document_type>
      <term> stress_value </term>
    </observation>
  </observations>
  <process_config>
    <label_option> true </label_option>
    <execution>
      <flag> date </flag>
      <value>
        2017-09-15T00:00:00E2017-09-23T00:00:00
      </value>
    </execution>
    <event>
      <flag> default </flag>
      <value> all </value>
      <max_col_num> inf </max_col_num>
    </event>
  </process_config>
  <preprocess> default </preprocess>
</DGM>
  
```

Table 4. Result of Generated Training Dataset

Dataset Name	Length	Parameter	Generating Date
CSA_dataset 201709 (D1)	73	cardiac_value, stress_value	2017.09.15 - 2017.09.23
CSA_dataset 201710 (D2)	87		2017.10.16 - 2017.10.27
CSA_dataset 201712 (D3)	105		2017.12.04 - 2017.12.16
CSA_dataset 201801 (D4)	51		2018.01.31 - 2018.02.10

4.3 기계학습 모델 피드백에 따른 정확도 비교

기계학습 모델의 재학습은 모델의 성능 유지를 위한 관리 과정이다. 심박 상태 분석 모델과 같이 사람의 생체를 입력 데이터로 사용하는 경우, 시간이 지남에 따라 생체 추이는 변동될 수 있으므로 학습데이터 생성 모델을 통해 생성한 학습데이터는 기계학습 모델에 반영되어, 기계학습 모델의 성능을 유지 또는 향상시킬 수 있어야 한다.

두 번째 실험에서는 제안하는 방법을 통해 생성한 학습데이터의 유효성을 검증과 최신의 데이터 학습의 필요성을 입증하기 위해, 기계학습 모델의 생성된 최신의 학습데이터를 제공한 모델의 분석 결과에 대한 정확도를 보인다. 학습에는 4.2에서 생성된 학습데이터인 Table 4를 사용한다.

Fig. 6은 4.1절에서 생성한 각 회차별 데이터셋(Table 4)을 테스트 데이터로 두 개의 모델에 적용하여 얻은 분석 결과에 대한 정확도이다. 실선은 피드백 모델, 점선은 기존 모델을 나타낸다. 실험은 총 4번 진행되었고 D1, D2, D3, D4 순으로 데이터를 입력하여 분석 결과를 획득하였으며 결과는 사람의 심박 상태이다. 정확도는 테스트 데이터의 분석 결과에 대한 정답률로, 예측된 심박 상태와 실제 심박 상태를 비교한 것으로 0에서 1사이의 값을 갖는다. 테스트에 사용된 데이터는 각 회차에 사용된 학습 데이터를 재생성한 것으로 실험을 위해 생성된 데이터이다. Table 5는 실험을 통해 얻은 분석 결과의 정확도를 수치로 나타낸 것으로 각 실험 단계에서 사용된 테스트 데이터셋과 두 모델을 적용하여 얻은 분석결과에 대한 정확도를 보인다.

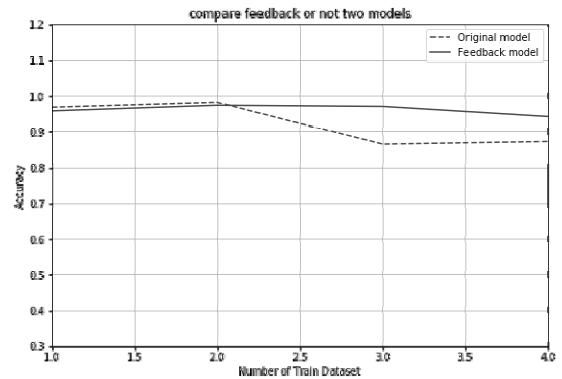


Fig. 6. Graph For Accuracy Comparison of Two Models(Original Model, Feedback Model)

Table 5. Accuracy Comparison of Two Models

Time	Test Dataset	Accuracy (%)	
		Original Model	Feedback Model
1	D1	0.96	0.95
2	D2	0.98	0.97
3	D3	0.86	0.97
4	D4	0.87	0.94

Table 5에서 볼 수 있듯이, 두 모델 모두 4회차동안의 85% 이상의 높은 분석 결과의 정확도를 보인다. 기존 모델은 1-2회차의 데이터에 대해서는 90% 이상의 높은 수준의 성능을 보이나, 3-4회차에 대한 정확도는 90% 이하로 떨어지는 것을 볼 수 있다. 피드백 모델은 1-2회차의 데이터의 분석 결과는 기존 모델보다 성능이 조금 낮지만, 3-4회차에서 높은 정확도를 가지고, 평균적으로 피드백 모델은 90% 이상의 정확도를 보인다.

기존의 정확도가 유지 또는 증가하는 부분은 실험에 사용된 데이터의 수집 기간이 생체 리듬의 변화가 이루어지기에는 짧은 기간이면서 수집 대상의 신체에 큰 변화가 없었기 때문으로 판단된다. 반대로 일부 변화한 부분에 대해서는 모델이 분석할 수 없는 범주의 데이터이므로 신뢰성이 떨어지는 부정확한 분석 결과를 제공하게 된다. 이와 다르게 피드백 모델은 최신의 학습데이터셋을 학습함으로써 1-2회차의 데이터를 포함한 생체 리듬을 포함하여 전체적으로 높은 수준의 정확도를 보인다.

5. 결 론

본 논문에서는 상황인지 시스템에서 기계학습 기반 모델의 적용할 때, 기계학습 모델의 분석결과에 대한 정확도를 유지하기 위한 관리에 필요한 학습데이터 생성 방법을 제안하였다. 기계학습 모델 학습데이터 생성에 따른 요구사항을 반영하기 위해 학습데이터 생성 모델 및 학습데이터 생성 알고리즘을 설계하였으며, 이는 학습데이터를 생성하는 모듈의 기반으로 사용되었다.

실험에서는 건강 상태 알림 서비스를 위한 심박상태 분석 모델을 대상으로 하여 설계한 학습데이터 생성 모델을 기반으로 학습데이터를 생성하고, 추가 학습이 이루어지지 않은 모델과 이를 반영한 모델의 분석결과에 대한 정확도를 비교하는 것으로 연구의 필요성을 입증하였다.

본 연구를 통해 상황인지 시스템에 적용되는 기계학습 모델을 관리할 수 있으며 이를 통해 기계학습 모델의 확장성을 제공할 수 있다. 추가적으로 텍스트와 이미지와 같은 입력 데이터를 고려하여 모델을 개선한다면 보다 범용적인 사용이 기대된다.

References

- [1] G. Manogaran and D. Lopez, "Health data analytics using scalable logistic regression with stochastic gradient descent," *International Journal of Advanced Intelligence Paradigms*, Vol.10, No.1-2, pp.118-132, 2018.
- [2] J.-W. Lee, H.-S. Lim, D.-W. Kim, S.-A. Shin, J. Kim, B. Yoo, and K.-H. Cho, "The development and implementation of stroke risk prediction model in National Health Insurance Service's personal health record," *Computer Methods and Programs in Biomedicine*, Vol.153, pp.253-257, 2018.
- [3] N. Bouri and S. Ravi, "Going mobile: how mobile personal health records can improve health care during emergencies," *JMIR mHealth uHealth*, Vol.2, No.1, e8, 2014.
- [4] N. Khozouie, F. Fotouhi-Ghazvini, and B. Minaei-Bidgoli, "Ontological MobiHealth system," *Indonesian J. Elect. Eng. Comput. Sci.*, Vol.10, No.1, pp.309-319, 2018.
- [5] S. Huang, L. Li, H. Cai, B. Xu, G. Li, and L. Jiang, "A Configurable WoT Application Platform Based on Spatiotemporal Semantic Scenarios," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol.49, No.1, pp.123-135, 2017.
- [6] L. Mainetti, V. Mighali, L. Patrono, and P. Rametta, "A novel Rule-based Semantic Architecture for IoT Building Automation Systems," *Software, Telecommunications and Computer Networks (SoftCOM)*, pp.124-131, 2015.
- [7] L. Pessoa, P. Fernandes, T. Castro, V. Alves, G. N. Rodrigues, and H. Carvalho, "Building reliable and maintainable Dynamic Software Product Lines: An investigation in the Body Sensor Network domain," *Information and Software Technology*, Vol.86, pp.54-70, 2017.
- [8] J. G. Nalepa and S. Bobek, "Rule-based solution for context-aware reasoning on mobile devices," *Computer Science and Information Systems*, Vol.11, No.1, pp.171-193, 2014.
- [9] N. Kolbe, A. Zaslavsky, S. Kubler, J. Robert, and Y. Le Traon, "Enriching a Situation Awareness Framework for IoT with Knowledge Base and Reasoning Components," In *International and Interdisciplinary Conference on Modeling and Using Context*, Springer, Cham, pp.41-54, 2017.
- [10] Z. Bahramian, R. Ali Abbaspour, and C. Claramunt, "A cold start context-aware recommender system for tour planning using artificial neural network and case based reasoning," *Mobile Information Systems*, 2017.
- [11] M. Shin, W. Paik, B. Kim, and S. Hwang, "An IoT Platform with Monitoring Robot Applying CNN-Based Context-Aware Learning," *Sensors*, Vol.19, No.11, pp.1-13, 2019.
- [12] L. Gomes, C., Ramos, A., Jozi, B., Serra, L., Paiva, and Z. Vale, "IoH: A Platform for the Intelligence of Home with a Context Awareness and Ambient Intelligence Approach," *Future Internet*, Vol.11, No.3, pp.58, 2019.
- [13] N. Polyzotis, S. Roy, S. E. Whang, and M. Zinkevich, "Data Management Challenges in Production Machine Learning," *Proceedings of the 2017 ACM International Conference on Management of Data*, ACM, pp.1723-1726, 2017.

[14] M. H. Kabir, M. R. Hoque, H. Seo, and S. H. Yang, "Machine learning based adaptive context-aware system for smart home environment," *International Journal of Smart Home*, Vol.9, No.11, pp.55-62, 2015.

[15] M. H., Kabir, M. R., Hoque, H., Seo, and S. H. Yang, "Boolean Control Network Based Modeling for Context-Aware System in Smart Home," *International Journal of Smart Home*, Vol.10, No.4, pp.65-76, 2016.

[16] B. Ospan, N. Khan, J. Augusto, M. Quinde, and K. Nurgaliyev, "Context aware virtual assistant with case-based conflict resolution in multi-user smart home environment," *2018 International Conference on Computing and Network Communications (CoCoNet)*, IEEE, pp.36-44, 2018.

[17] P. Jiang, J. Winkley, C. Zhao, R. Munnoch, G. Min, and L. T. Yang, "An intelligent information forwarder for healthcare big data systems with distributed wearable sensors," *IEEE systems journal*, Vol.10, No.3, pp.1147-1159, 2014.

[18] O. Banos, R. Garcia, J. A. Holgado-Terriza, M. Damas, H. Pomares, I. Rojas, A. Saez, and C. Villalonga, "mHealthDroid: a novel framework for agile development of mobile health applications," *International Workshop on Ambient Assisted Living*, Springer, Cham, pp.91-98, 2014.

[19] T. Zoppi, A. Ceccarelli, and A. Bondavalli, "Context-awareness to improve anomaly detection in dynamic service oriented architectures," *International Conference on Computer Safety, Reliability, and Security*, Springer, Cham, pp.145-158, 2016.

[20] Y. Bai, H. Ji, Q. Han, J. Huang, and D. Qian, "MidCASE: a service oriented middleware enabling context awareness for smart environment," *2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*, IEEE, pp.946-951, 2007.

[21] L. O. Colombo-Mendoza, R. Valencia-Garcia, A. Rodriguez-Gonzalez, G. Alor-Hernandez, and J. J. Samper-Zapater, "RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes," *Expert Systems with Applications*, Vol.42, No.3, pp.1202-1222, 2015.

[22] P. H. Wu, G. J. Hwang, and W. H. Tsai, "An expert system-based context-aware ubiquitous learning approach for conducting science learning activities," *Journal of Educational Technology & Society*, Vol.16, No. 4, pp.217-230, 2013.

[23] D. Galar, A. Thaduri, M. Catelani, and L. Ciani, "Context awareness for maintenance decision making: A diagnosis and prognosis approach," *Measurement*, pp.137-150, 2015.

[24] A. I. Wang, and Q. K. Ahmad, "Camf-context-aware

machine learning framework for android," *Proceedings of the International Conference on Software Engineering and Applications (SEA 2010)*, CA, USA, 2010.

[25] H. Eldardiry, K. Sricharan, J. Liu, J. Hanley, B. Price, O. Brdiczka, and E. Bart, "Multi-source fusion for anomaly detection: using across-domain and across-time peer-group consistency checks," *JoWUA*, Vol.5, No.2, pp.39-58, 2014.



문 종 혁

<https://orcid.org/0000-0002-4919-0058>

e-mail : jonghyeokmun@soongsil.ac.kr

2017년 한경대학교 컴퓨터공학과(학사)

2019년 숭실대학교 컴퓨터학과(석사)

2019년 ~ 현 재 숭실대학교 컴퓨터학과 박사과정

관심분야 : 로봇 미들웨어, 빅데이터 처리, 인공지능 모델링



최 종 선

<https://orcid.org/0000-0001-9648-0667>

e-mail : jongsun.choi@ssu.ac.kr

2000년 숭실대학교 컴퓨터학부(학사)

2002년 숭실대학교 컴퓨터학부(석사)

2008년 ~ 2010년 유한대학교

e-비즈니스과 전임교원

2010년 숭실대학교 컴퓨터학부(박사)

2011년 ~ 2012년 숭실대학교 지능형로봇연구소 연구원

2012년 ~ 2013년 서일대학교 인터넷정보과 전임교원

2013년 ~ 현 재 숭실대학교 컴퓨터학부 조교수

관심분야 : 로봇 소프트웨어 플랫폼, 클라우드 컴퓨팅



최 재 영

<https://orcid.org/0000-0002-7321-9682>

e-mail : choi@ssu.ac.kr

1984년 서울대학교 제어계측공학과(학사)

1986년 미국 남가주대학교 전기공학과

(컴퓨터공학)(석사)

1991년 미국 코넬대학교 전기공학부

(컴퓨터공학)(박사)

1992년 ~ 1994년 미국 국립오크리지연구소 연구원

1994년 ~ 1995년 미국 테네시주립대학교 연구교수

1995년 ~ 현 재 숭실대학교 컴퓨터학부 교수

관심분야 : 로봇 미들웨어, 시스템소프트웨어, 병렬/분산처리,

고성능컴퓨팅