

# Hierarchical multi-task learning with self-supervised auxiliary task

Seunghan Lee<sup>a</sup>, Taeyoung Park<sup>1,ab</sup>

<sup>a</sup>Department of Statistics and Data Science, Yonsei University;

<sup>b</sup>Department of Applied Statistics, Yonsei University

---

## Abstract

Multi-task learning is a popular approach in machine learning that aims to learn multiple related tasks simultaneously by sharing information across them. In this paper, we consider a hierarchical structure across multiple related tasks with a hierarchy of sub-tasks under the same main task, where representations used to solve the sub-tasks share more information through task-specific layers, globally shared layers, and locally shared layers. We thus propose the hierarchical multi-task learning with self-supervised auxiliary task (HiSS), which is a novel approach for hierarchical multi-task learning that incorporates self-supervised learning as an auxiliary task. The goal of the auxiliary task is to further extract latent information from the unlabeled data by predicting a cluster label directly derived from the data. The proposed approach is tested on the Hyodoll dataset, which consists of user information and activity logs of elderly individuals collected by AI companion robots, for predicting emergency calls based on the time of day and month. Our proposed algorithm is more efficient than other well-known machine learning algorithms as it requires only a single model regardless of the number of tasks, and demonstrates superior performance in classification tasks using various metrics. The source codes are available at: <https://github.com/seunghan96/HiSS>.

Keywords: classification, clustering, hierarchical multi-task learning, machine learning, self-supervised learning

---

## 1. Introduction

다중 작업 학습(multi-task learning; MTL) (Caruana, 1997)은 서로 연관성이 높은 여러 작업의 성능을 향상시키기 위해 모든 작업을 공동으로 학습시키는 기계 학습 방법론이다. 이를 통해 모델은 여러 작업 간의 공유된 표현을 학습하고, 작업들 간의 종속성을 활용하여 성능을 향상시키고 더 나은 일반화를 달성할 수 있다 (Ruder, 2017). MTL은 컴퓨터 비전 (Misra 등, 2016), 자연어 처리 (Liu 등, 2015), 음성 인식 (Huang 등, 2022) 등의 다양한 분야에서 적용되었으며, 최근에는 MTL과 메타 러닝(meta-learning)을 접목한 연구 (Munkhdalai와 Yu, 2017)와, 변화하는 환경에 적응하기 위해 온라인 방식으로 MTL을 학습하는 방법론 (Pentina 등, 2015) 등이 개발되었다.

---

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2020R1A2C1A01005949, RS-2023-00217705), the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) support program (RS-2023-00259934) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation), and the Son Jiho Research Grant of Yonsei University (2023-22-0006).

<sup>1</sup>Corresponding author: Department of Applied Statistics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, E-mail: [tpark@yonsei.ac.kr](mailto:tpark@yonsei.ac.kr)

계층적 다중 작업 학습(hierarchical multi-task learning; HMTL)은 작업들 간에 계층적 관계가 있는 특정 유형의 MTL이다. 따라서, 해당 프레임워크 하에서 개발된 모델 아키텍처의 경우에는 작업 간의 계층적 관계를 고려하여 설계한다. Chen 등 (2018)이 제안한 모델의 경우, 저수준 작업(low-level task)을 해결하기 위한 벡터는 모델의 하위 층에서, 고수준 작업(high-level task)을 해결하기 위한 벡터는 모델의 상위 층에서 추출한다. 본 논문에서 제안하는 HiSS (hierarchical multi-task learning with self-supervised auxiliary task)의 경우, 작업들 간의 수준(level)으로 계층을 나누는 앞선 방법과는 다르게, 여러 하위 작업(sub-task)이 하나의 주요 작업(main-task)에 속하는 작업 계층 구조가 존재하는 문제를 해결하는 새로운 HMTL 방법론이다.

최근 기계 학습에서는 지도 학습을 위한 높은 레이블링 비용 문제를 해결하고자 자기 지도 학습(self-supervised learning; SSL)이 주목받고 있다. SSL은 레이블된 데이터를 사용하는 대신, 레이블이 없는 입력 데이터만을 사용하여 획득한 가상 레이블(pseudo-label)을 학습 중에 지도 신호로 사용하여 모델을 학습한다. 최근 연구에서는 SimCLR (Chen 등, 2020), SwAV (Caron 등, 2020), BYOL (Grill 등, 2020) 등과 같은 새로운 SSL 알고리즘들이 제안되었으며, 해당 방법론들은 기존의 지도 학습(supervised learning; SL) 방법론들보다 더 뛰어난 성능을 보인다. SSL과 마찬가지로 레이블을 요구하지 않는 군집분석은 유사한 데이터들을 그들의 특징에 따라 군집화하는 대표적인 비지도 학습 방법론이다 (Hastie 등, 2001). 해당 방법론은 각 그룹 내의 유사성을 증가시키고 그룹 간의 유사성을 감소시키는 것 (Tan 등, 2016)을 목표로 하며, 데이터 간의 거리를 측정하는 기준에 따라 k-평균 군집분석 (MacQueen, 1967), 계층적 군집분석 (Murtagh, 2014), 밀도 기반 군집분석(DBSCAN) (Ester 등, 1996) 등의 다양한 알고리즘들이 존재한다. 딥러닝을 활용한 군집분석 방법론인 자기 조직화 지도(SOM) (Kohonen, 1982)의 경우, 높은 차원의 데이터를 저차원의 격자로 비선형 매핑하면서 그들의 위상 특징을 보존할 수 있고, 노이즈와 이상치에 대해 강인하다는 점에서 전통적인 기계 학습 기반의 군집분석 알고리즘보다 많은 장점을 가지고 있다.

본 논문에서는 HMTL 구조를 사용하여 모든 작업 간에 공유되는 공통 정보, 특정 주요 작업 하에 속한 하위 작업들 간의 공유되는 공통 정보, 하위 작업별 고유의 정보의 총 세 가지 정보를 사용하여 여러 작업을 수행하는 HiSS를 제안한다. 해당 방법론은 각각의 계층적 정보를 추출하기 위해 전역적 공유층(globally shared layers), 지역적 공유층(locally shared layers), 작업 별 특정층(task-specific layers)를 사용한다. 더 나아가서, 모델의 성능을 더욱 향상시키고 레이블이 없는 데이터도 학습에 사용하기 위하여, 우리는 HMTL 프레임워크 내에서 SSL을 보조 과제로써 결합한다. 보다 구체적으로, 보조 과제를 위해서, 레이블이 없는 데이터를 군집분석한 뒤 이로부터 얻어진 군집 레이블을 가상 레이블로 사용하여 다중 분류 문제를 수행한다. 제안한 HiSS는 하위 작업이 특정 상위 (주요) 작업에 속하는 계층적 구조가 존재하는 작업들을 수행하기 위해 특화된 HMTL 알고리즘으로, SSL을 HMTL에 추가적인 보조 과제로 결합한 점에서 기존의 방법론들과 차별성을 제공하며 새롭게 제안된다.

본 논문의 구성은 다음과 같다. 2절에서는 주요 과제를 수행하기 위한 계층적 다중 작업 학습(HMTL)과 이에 대한 보조 과제인 자기 지도 학습(SSL)를 결합한 HiSS 알고리즘에 대해 소개한다. 3절에서는 제안된 방법을 독거 노인에 대한 다양한 사용자 정보를 포함하는 효율 데이터에 적용하여 시간별 및 월별 패턴을 기반으로 응급 호출을 예측하는 분석을 수행하고, 효율 데이터를 이용하여 제안된 방법론과 응급 호출 예측을 위한 다양한 방법론을 비교하며, 제안된 방법론의 비교 실험을 수행한다. 마지막으로 4절에서는 결론 및 향후 연구 방향에 대해 논의한다.

## 2. HiSS: Hierarchical multi-task learning with self-supervised auxiliary task

HiSS는 계층적 다중 작업을 수행하는 HMTL을 주요 과제로 하고, 레이블이 없는 데이터의 정보를 추가적으로 활용하는 SSL을 보조 과제로 하는 학습을 동시에 진행한다. 이 두 가지 과제 모두 종단 간(end-to-end) 방식으로 학습되며, 제안된 방법론의 전체 아키텍처는 Figure 1에 설명되어 있다.

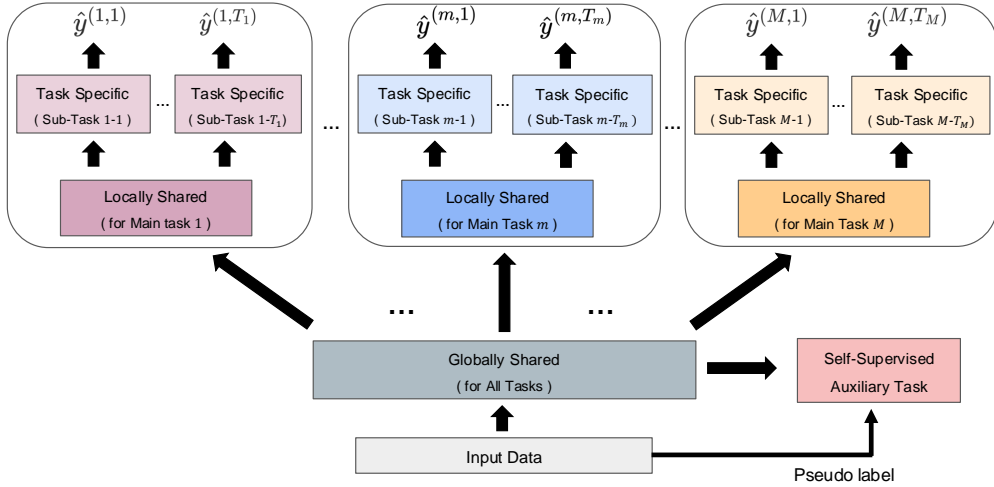


Figure 1: **Overall architecture of HiSS.** In the proposed methodology, globally shared layers are used for input data processing across all tasks, followed by locally shared layers and task-specific layers. The auxiliary task is performed using hidden representations obtained from the globally shared layers across all tasks.

## 2.1. 계층적 다중 작업 학습(HMTL)을 활용한 주요 과제(main task)

계층적 다중 작업 학습(HMTL)은 작업 간의 다양한 계층 구조가 존재할 때 사용될 수 있는 다중 작업 학습(MTL)의 한 종류이다. HMTL에서는 하위 작업 고유의 정보, 주요 작업에 대한 정보, 그리고 작업 무관한 공통의 정보와 같이 여러 하위 작업들이 상위 작업으로 묶이는 계층적 구조를 가지고, 이러한 정보를 계층적으로 포착할 수 있다. HiSS에서는 이러한 세 가지 정보를 모두 효과적으로 추출하기 위한 HMTL 구조를 채택한다. 즉, 최하단의 전역적 공유층(globally shared layer)을 통해 작업과 무관한 공통된 정보를, 그 위의 지역적 공유층(locally shared layer)을 통해 몇 개의 하위 작업을 묶은 하나의 주요 작업의 정보를, 그리고 최상단의 작업별 특정층(task-specific layer)을 통해 하위 작업 고유의 정보를 포착한다. 이러한 구조를 통해 하위 및 주요 작업들 간의 상호 작용을 단일 모델로써 효과적으로 포착하는 것이 가능하다.

구체적으로, 입력 데이터  $x \in \mathbb{R}^d$ 에 대해 전역적 공유층  $f_G$ 을 사용하여 모든 작업에 대한 공통 표현(common representation)인  $h_G$ 을 추출한다. 여기서  $d$ 는 각 데이터의 입력 차원을 뜻한다. 이렇게 추출한  $h_G$ 을 이용하여,  $m$ 번째 주요 작업에 해당하는 주요 작업별 특정 표현(locally shared representation)인  $h_L^{(m)}$ 을  $m$ 번째 지역적 공유층  $f_L^{(m)}$ 을 사용하여 추출한다. 마지막으로,  $m$ 번째 주요 작업의  $s$ 번째 하위 작업에 해당하는 예측 값  $\hat{y}^{(m,s)}$ 은  $m$ 번째 하위 작업별 특정층  $f_S^{(m,s)}$ 에  $h_L^{(m)}$ 를 입력함으로써 얻는다. 활성화 함수로는 ReLU를 사용하며, 앞서 서술한 각 하위 작업에 대한 계층적 과정은 다음과 같이 표현될 수 있다.

$$h_G = \max(0, xW_G + b_G) \quad (2.1)$$

$$h_L^{(m)} = \max(0, h_G W_L^{(m)} + b_L^{(m)}), \quad m = 1, \dots, M \quad (2.2)$$

$$\hat{y}^{(m,s)} = g(h_L^{(m)} W_S^{(m,s)} + b_S^{(m,s)}), \quad m = 1, \dots, M, \quad s = 1, \dots, T_m, \quad (2.3)$$

여기서  $W_G \in \mathbb{R}^{d \times d'}$ ,  $b_G \in \mathbb{R}^{d'}$ ,  $W_L^{(m)} \in \mathbb{R}^{d' \times d''}$ ,  $b_L^{(m)} \in \mathbb{R}^{d''}$ ,  $W_S^{(m,s)} \in \mathbb{R}^{d'' \times 1}$ ,  $b_S^{(m,s)} \in \mathbb{R}$ 이고, 입력 데이터의 차원을  $d$ , 중간 층의 차원을  $d'$ , 그 이후 최상단 층의 차원을  $d''$ 로 표기하며,  $g$ 는 최종 하위 작업의 태스크에 따라 결정되는 함수를 의미한다. 이러한 과정은 모든 하위 작업에 대해 병렬로 수행된다.

HMTL의 손실 함수는 각 주요 작업에 속하는 모든 하위 작업들의 손실 함수의 가중합으로 구성된다. 모델 학습을 위한 학습 데이터셋  $D = \{x_i, \{y_i^{(m,s)}\}_{m=1,\dots,M,s=1,\dots,T_m}\}_{i=1,\dots,N}$ 에서  $x_i$ 는 입력을,  $y_i^{(m,s)}$ 는  $m$ 번째 주요 작업의  $s$ 번째 하위 작업의 레이블을 의미한다. 여기서  $N$ 은 데이터의 개수를,  $M$ 은 주요 작업의 총 개수를,  $T_m$ 은  $m$ 번째 주요 작업의 하위 작업의 총 개수를 나타내고,  $\hat{y}_i^{(m,s)}$ 는 입력  $x$ 에 대해 모델이 예측한  $m$ 번째 주요 작업의  $s$ 번째 하위 작업의 예측값을 나타낸다. 이를 활용하여, HMTL의 손실 함수  $\mathcal{L}_{\text{HMTL}}$ 는 다음과 같이 계산된다.

$$\mathcal{L}_{\text{HMTL}} = \sum_{i=1}^N \sum_{m=1}^M \sum_{s=1}^{T_m} \lambda_{m,s} L_{m,s}(\hat{y}_i^{(m,s)}, y_i^{(m,s)}; \theta), \quad (2.4)$$

여기서  $L_{m,s}$ 와  $\lambda_{m,s}$ 는 각각  $m$ 번째 주요 작업의  $s$ 번째 하위 작업의 손실 함수와 해당 작업의 손실 함수 최종 HMTL 손실 함수에 기여하는 가중치를 나타내고,  $\theta$ 는 모델의 파라미터를 나타낸다. 손실 함수  $L_{m,s}$ 와 가중치  $\lambda_{m,s}$ 는 풀고자하는 작업의 특징에 따라 결정되며, 제안된 방법론에서는 작업 별로 균일한  $\lambda_{m,s} = 1$ 을 기본값으로 설정한다.

## 2.2. 자기 지도 학습(SSL)을 활용한 보조 과제(sub task)

HiSS는 주요 과제인 HMTL과 더불어 SSL을 보조 과제로 활용함으로써 레이블이 없는 데이터 또한 학습에 동시에 활용한다. 이를 위해 모든 데이터에 대해 군집분석을 사전에 진행하고, 이로부터 얻어지는 군집 레이블을 보조 과제의 가상 레이블로 사용하여 멀티 레이블 분류(multi-class classification) 과제를 수행한다. 군집분석을 수행하기에 앞서서, 다양한 차원축소 기법을 사용하여 데이터를 보다 낮은 차원으로 축소하게 되면, 이러한 군집분석 과정을 다음과 같이 표현할 수 있다.

$$z = \text{DimReduce}(x), \quad (2.5)$$

$$y_{\text{SSL}} = \text{Clustering}(z), \quad (2.6)$$

여기서  $z$ 는 차원축소 방법인 DimReduce를 사용하여 얻은 잠재 벡터를 나타내고,  $y_{\text{SSL}}$ 는 SSL의 가상 레이블로 사용할 군집을 나타낸다. 3절에서 설명하듯, HiSS에서는 테이블 데이터를 다루기 위한 차원축소 방법론으로는 요인분석(factor analysis)을 사용하고, 군집분석 방법론으로는 SOM을 사용한다.

군집분석은 특정 작업의 레이블과 무관하게 오직 입력 변수만을 대상으로 진행을 하므로, 군집 레이블을 예측하는 SSL 보조 과제는 전역적 공유층  $f_G$ 를 사용하여 추출한 공통 표현인  $h_G$ 를 사용하여 예측한다.  $h_G$ 를 이용하여 예측된 군집 레이블  $\hat{y}_{\text{SSL}}$ 는 다음과 같이 표현될 수 있다.

$$\hat{y}_{\text{SSL}} = \text{Softmax}(h_G \cdot W_{\text{SSL}} + b_{\text{SSL}}), \quad (2.7)$$

여기서  $W_{\text{SSL}} \in \mathbb{R}^{d' \times C}$ ,  $b_{\text{SSL}} \in \mathbb{R}^C$ 이며,  $C$ 는 고유한 군집의 수를 나타낸다. 따라서 SSL의 손실 함수  $\mathcal{L}_{\text{SSL}}$ 는 다음과 같이 정의된다.

$$\mathcal{L}_{\text{SSL}} = \sum_{i=1}^N \sum_{m=1}^M \sum_{s=1}^{T_m} L_{\text{SSL}}(\hat{y}_{i,\text{SSL}}^{(m,s)}, y_{i,\text{SSL}}^{(m,s)}; \theta), \quad (2.8)$$

HiSS에서는 군집 레이블을 예측하는 다중 클래스 분류(multi-class classification) 작업을 수행하므로,  $L_{\text{SSL}}$ 로 크로스 엔트로피(cross-entropy) 손실 함수를 사용한다.

## 2.3. HiSS의 목적 함수(objective function)

HiSS는 HMTL과 SSL의 두 과제를 동시에 활용하여 학습이 진행되므로, 전체 손실 함수  $\mathcal{L}$ 는 다음과 같이 이 두 종류의 작업의 가중합으로 구성된다.

$$\mathcal{L} = \mathcal{L}_{\text{HMTL}} + \alpha \cdot \mathcal{L}_{\text{SSL}}, \quad (2.9)$$

여기서  $\alpha$ 는 최종 손실 함수에 기여하는 보조 과제의 기여도를 나타내고, 본 논문에서는 기본값을 1로 설정한다. 앞서 설명한 HiSS에 대한 전체적인 의사코드는 Algorithm 1에 서술되어 있다.

---

**Algorithm 1** : HiSS 알고리즘
 

---

**Require:** 데이터
 $D = \{x_i, \{y_i^{(m,s)}\}_{m=1,\dots,M,s=1,\dots,T_m}\}_{i=1,\dots,N}$ : 학습 데이터셋

 $M$ : 주요 작업의 수

 $T_m$ :  $m$ 번째 주요 작업에 속한 하위 작업의 수
**Require:** 모델
 $f_G$ : 전역적 공유층 (파라미터:  $\theta_G$ )

 $f_L^{(m)}$ :  $m$ 번째 주요 작업의 지역적 공유층 (파라미터:  $\theta_L^{(m)}$ )

 $f_S^{(m,s)}$ :  $m$ 번째 주요 작업의  $s$ 번째 하위 작업의 작업 별 특징층 (파라미터:  $\theta_S^{(m,s)}$ )

 $g$ : 보조 과제의 다중 클래스 분류를 위한 선형층 (파라미터:  $\theta_g$ )

- 1: 모델 초기화  $\theta = \{\theta_G, \{\theta_L^{(m)}, \{\theta_S^{(m,s)}\}_{s=1,\dots,T_m}\}_{m=1,\dots,M}, \theta_g\}$ .
  - 2: **for all**  $n = 1$  to  $N$  **in parallel do**
  - 3:    $z_i \leftarrow \text{DimReduce}(x_i)$
  - 4:    $y_{i,\text{SSL}} \leftarrow \text{Clustering}(z_i)$
  - 5:    $h_{i,G} = f_G(x_i; \theta_G)$
  - 6:   **for all**  $m = 1$  to  $M$  **in parallel do**
  - 7:      $h_{i,L}^{(m)} = f_{i,L}^{(m)}(h_{i,G}; \theta_L^{(m)})$
  - 8:     **for all**  $s = 1$  to  $T_m$  **in parallel do**
  - 9:       $\hat{y}_i^{(m,s)} = f_S^{(m,s)}(h_{i,L}^{(m)}; \theta_S^{(m,s)})$
  - 10:     **end for**
  - 11:   **end for**
  - 12:    $\hat{y}_{i,\text{SSL}} = g(h_{i,G}; \theta_g)$
  - 13: **end for**
  - 14:  $\mathcal{L}_{\text{HMTL}} = \sum_{i=1}^N \sum_{m=1}^M \sum_{s=1}^{T_m} \lambda_{m,s} L_{m,s}(\hat{y}_i^{(m,s)}, y_i^{(m,s)}; \theta)$
  - 15:  $\mathcal{L}_{\text{SSL}} = \sum_{i=1}^N \sum_{m=1}^M \sum_{s=1}^{T_m} L_{\text{SSL}}(\hat{y}_i^{(m,s)}, y_i^{(m,s)}; \theta)$
  - 16:  $\mathcal{L} = \mathcal{L}_{\text{HMTL}} + \alpha \cdot \mathcal{L}_{\text{SSL}}$
  - 17:  $\mathcal{L}$ 에 대한 역전파를 통해  $\theta$  업데이트
- 

### 3. 효율 데이터에 대한 적용

#### 3.1. 데이터셋 및 작업 설명

본 논문에서는 독거 노인을 위한 AI 동반 로봇을 개발하는 회사인 효율의 데이터에 HiSS 알고리즘을 적용한다. 해당 데이터셋은 다음과 같은 세 가지 하위 데이터셋으로 구성되어 있다.

- 인형/사용자에 대한 통합 정보 데이터
- 고객 행동에 기반한 로그 데이터로 프로그램 로그 데이터와 동작 로그 데이터
- 응급 호출 데이터

여기서 인형/사용자에 대한 통합 정보는 사용자와 인형이 매칭된 2,285개의 고유 값으로 구성되며, 사용자의

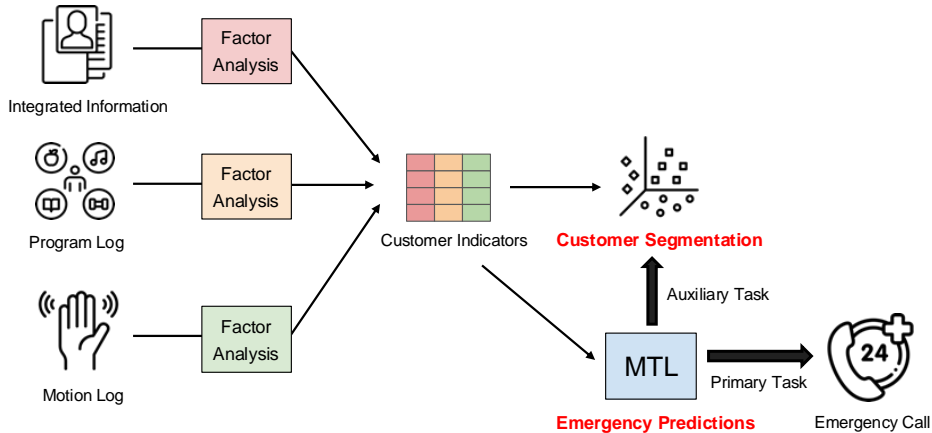


Figure 2: **Diagram of the HiSS applied to Hyodoll dataset.** HiSS generates customer metrics based on demographic and behavioral variables, which are used as input features for the HMTL model and clustering. The main tasks involve predicting whether customers will make emergency calls on a monthly/hourly basis, while the auxiliary task is to predict customer clusters.

성별, 생일, 종교, 기상 시간, 식사 시간, 수면 시간, 약 복용 알림 설정 시간, 걷기 및 환기, 그리고 질병 상태에 관한 정보가 포함되어 있다. 프로그램 로그 데이터는 사용자가 인형을 사용하여 수행할 수 있는 여러 프로그램에서 생성된 2차 로그 데이터로, 종교 텍스트 읽기, 트로트 음악 재생, 영어 수업, 기억력 훈련, 퀴즈 및 운동 등이 포함되어 있다. 이 데이터는 시간별로 그룹화하여 약 30만 개의 데이터 포인트로 구성된다. 동작 로그 데이터는 사용자가 인형과 상호작용할 때 생성되며, 인형의 머리 쓰다듬기, 손 만지기, 탭하기 및 인간 몸 감지 등의 행동을 포함한다. 이 데이터는 시간별로 그룹화되어 약 53만 개의 데이터 포인트로 구성된다. 응급 호출 데이터는 사용자가 긴급 상황 시 인형의 손바닥을 누를 때 생성되며, 시간별로 그룹화된 약 2.6만개의 데이터 포인트로 구성된다.

해당 데이터를 사용하여 수행하는 하위 작업은 월별 및 시간별 긴급 호출을 예측하는 이진 분류 작업이다. 해당 작업은 월별 호출에는 12개의 하위 작업이, 시간별 호출에는 5개의 하위 작업이 있는 계층적 구조를 가지고 있으며, 우리는 이를 두 개의 주요 작업과 17개의 하위 작업을 가진 계층적 구조로써 분석한다. 위와 같은 계층적 구조 하에서, HiSS 알고리즘이 포착하는 계층적 정보는 아래와 같다: 1) 모든 작업들과 하위 작업들 간에 공유되는 공통 정보는 월과 시간대에 무관하게 공통으로 나타나는 정보이고, 2) 특정 작업에 속하는 하위 작업들 간에 공유되는 공통 정보는 월에 대한 전체 정보와 시간대에 대한 전체 정보이며, 3) 각 하위 작업별 고유의 정보는 특정 월과 특정 시간대에 대한 고유한 정보를 의미한다.

### 3.2. 실험 결과

Figure 2은 HiSS에서 효돌 데이터를 사용하여 군집분석 알고리즘을 통한 고객 군집화(customer segmentation)와 다중 작업 학습을 통한 긴급 호출 예측(emergency prediction) 과정을 보여준다. 먼저, 통합 정보와 두 가지 유형의 로그 데이터(프로그램 로그와 동작 로그)를 요인분석(factor analysis)을 통해 차원축소하여 잠재 변수를 얻고, 이를 각 고객의 고객 지표(customer indicators)로써 사용한다. 해당 지표를 SOM 알고리즘에 사용하여 군집분석을 수행한다. 마지막으로, 해당 지표들을 사용하여 군집 레이블을 예측하는 SSL 보조 과제와 월별 및 시간별 긴급 호출 상태를 예측하기 위한 HMTL 주요 과제를 동시에 수행한다.

Table 1: Customer clustering result

Cluster label	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)	총합
# Customers	291	221	320	417	141	339	1,729
Ratio (%)	16.8	12.8	18.5	24.1	8.2	19.6	100.0

The table represents the proportion of customers in each cluster.

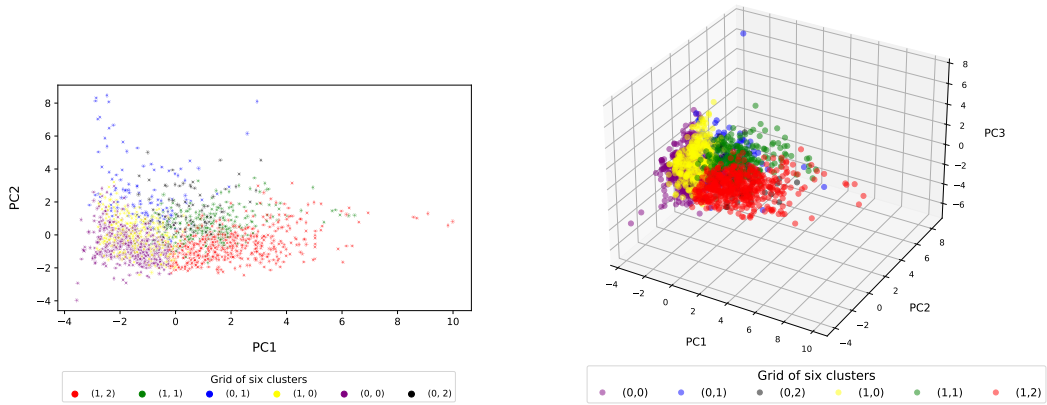


Figure 3: Visualization of clustering and factor analysis results.

### 3.2.1. SSL: 보조 과제를 위한 군집분석

HiSS에서는 세 가지 유형의 데이터셋(통합 사용자/인형 정보, 프로그램 수행 로그, 인형 접근 로그)을 요인 분석하여 각 데이터셋에 대한 여러 잠재 변수(고객 지표)를 생성하였다. 각 데이터셋의 특성을 반영하여, 세 그룹의 잠재 변수를 각각 종합 요인 점수, 활동 요인 점수, 사용 요인 점수로 명명하였다.

요인분석에서는 분산 설명 비율이 60%를 넘는 요인을 기준으로 요인 점수를 선택하고, 이를 표준화하여 세 그룹의 요인 점수를 구성하였으며, 이 세 그룹의 요인 점수를 모두 통틀어 고객 지표로 정의하였다. 그 결과, 수면 패턴, 규칙적인 일상 생활 정보, 약물 관련 정보를 포함한 6개의 종합 요인 점수, 사용 빈도, 종류 및 주요 사용 계절을 기반으로 한 7개의 활동 요인 점수, 효율 인형의 사용 패턴을 기반으로 한 4개의 사용 요인 점수가 생성되었다. 이러한 17개의 고객 지표를 변수로 사용하여 SOM 군집분석을 수행한 결과, 2×3 그리드의 6개의 군집이 형성되었는데, 이는 군집 개수 후보군인(4, 6, 8, 10) 중 실루엣 점수가 가장 높은 군집의 개수가 선택된 것이다. 각 군집 내 고객 비율은 Table 1에 나와 있으며, 요인분석을 통해 얻은 잠재 변수에 대한 시각화는 Figure 3에 제시되어 있다.

Figure 4는 각 군집의 17개 고객 지표 평균을 나타낸다. 이 그림을 통해 각 고객 그룹이 서로 다른 특성을 가지고 있으며, 그리드에서 인접한 고객 그룹이 멀리 떨어진 그룹보다 더 유사하다는 것을 확인할 수 있다. 이는 SOM이 다른 군집분석 알고리즘과 달리 군집 레이블의 위상 구조를 유지하는 장점을 가진다는 것을 보여주며, 이를 통해 얻어진 군집 레이블을 보조 과제의 가상 레이블로써 사용하는 것이 적절함을 보여준다.

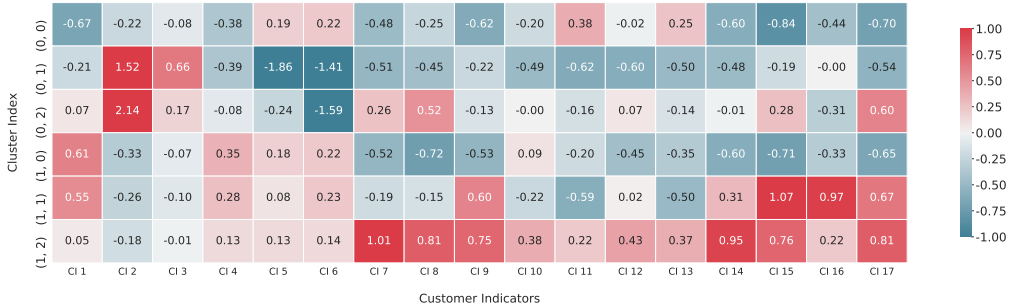


Figure 4: Average customer metrics for each cluster.

Table 2: Comparison of various models for emergency call prediction

Method	# Models	Metric		
		Accuracy	F1	AUROC
Gaussian Naive Bayes	17	0.799	0.354	0.785
Logistic Regression	17	0.824	0.310	0.806
Random Forest	17	0.831	0.368	0.794
HiSS	<b>1</b>	<b>0.835</b>	<b>0.444</b>	<b>0.820</b>

The evaluation metrics represent the average values across 17 tasks.

### 3.2.2. HMTL: 고객의 긴급 호출 여부 예측

계층적 멀티태스킹 학습 구조로서 전체 모델에 대해 256개의 유닛을 가진 단일 공유 레이어를 사용하고, 월별/시간별 공유 레이어에는 각각 128개의 유닛을 가진 단일 레이어를 사용하였다. 17개의 태스크별 레이어 각각에는 128개의 유닛을 가진 두 개의 레이어를 사용하였다. 손실 함수로는 이진 크로스 엔트로피를 사용하였고, Adam 옵티마이저를 학습률 0.001로 설정하여 모델을 학습시켰다. 각 태스크를 동등하게 취급하기 위해 최종 손실 함수에서 각 태스크의 기여도에 동일한 가중치를 부여하였다.

베이스라인 방법으로는 가우시안 나이브 베이즈, 로지스틱 회귀, 랜덤 포레스트를 선택하였다. 이는 전통적인 머신 러닝 기반 방법들이 일반적으로 딥러닝 방법에 비해 표 형식 데이터에서 여전히 좋은 성능을 보이기 때문이다 (Shwartz-Ziv와 Armon, 2022). 추가적으로, 자기 지도 학습 알고리즘을 사용한 사전 학습은 대량의 비레이블 데이터가 필요하기 때문에 데이터가 불충분한 경우에는 적합하지 않다.

반응 변수의 클래스 불균형 때문에, 정확도뿐만 아니라 F1-점수와 AUROC도 평가 지표로 사용하였다. 데이터의 80%를 학습 데이터로, 나머지 20%를 테스트 데이터로 사용하여 평가를 진행하였다. Table 2는 5번의 랜덤 시드로 평균낸 값을 바탕으로 세 가지 베이스라인과 비교한 응급 호출 예측 결과를 보여준다. 제안한 알고리즘은 다른 알고리즘과 비교하여 모든 지표에서 우수한 성능을 보일 뿐만 아니라, 모든 작업에 걸쳐서 단일 모델을 사용한다는 점에서도 보다 효율적임을 알 수 있다. Figure 5는 군집별 월 평균 응급 호출에 대한 예측값을 나타내며, 이는 서로 다른 계절/달에 각 그룹별로 다른 관리가 필요함을 시사한다.

### 3.3. 비교 실험(ablation study) 및 효율성 분석(eficiency analysis)

HiSS 아키텍처의 세 가지 층(전역적 공유층, 지역적 공유층, 작업 별 특정층)의 영향을 조사하기 위해 이들에 대한 비교 연구를 수행했다. 비교 연구의 결과는 Table 3에 제시되어 있으며, 모든 층을 사용하는 것이 가장 좋은 성능을 가진다는 것을 확인할 수 있다. 또한, SSL의 보조 작업을 수행했을 경우에는 추가적인 향상을



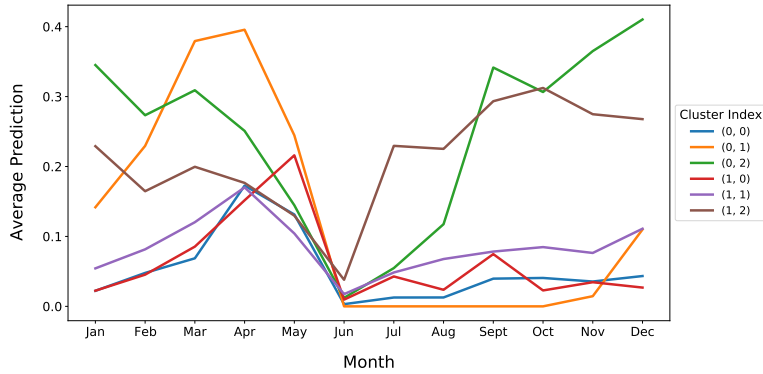


Figure 5: Monthly average emergency call predictions.

Table 3: Results of ablation studies

SSL	Shared layers		Task-specific layers	Accuracy	F1	AUROC
	Global	Local				
			✓	0.817	0.400	0.805
		✓	✓	0.825	0.425	0.815
	✓		✓	0.821	0.422	0.810
	✓	✓	✓	0.828	0.427	0.819
✓	✓	✓	✓	<b>0.835</b>	<b>0.444</b>	<b>0.820</b>

Ablation studies are conducted on using globally shared layers, locally shared layers, and task-specific layers for each task.

Table 4: Training time according to the number of tasks

	Number of tasks				
	1	2	4	8	16
Training time (sec.)	32	34	38	44	60

가진다는 것을 확인할 수 있다.

Table 4는 수행하는 다중 작업의 수에 따른 학습 시간을 보여주고 있고, 더 많은 작업을 사용해도 추가적인 시간 소요가 크지 않음을 확인할 수 있다. 이는 각 하위 작업 고유의 특징을 추출하는 작업 별 특정층을 단일 선형층으로 사용함으로써 계산량을 최소화하였기 때문이다.

#### 4. 결론

본 논문에서는 여러 작업들의 계층적 구조를 고려한 다중 작업 학습(HMTL)을 주요 과제로 하고 라벨이 없는 데이터를 활용하는 위한 자기 지도 학습(SSL)을 보조 과제로 결합한 HiSS를 제안한다. HiSS에서 추출된 표현 벡터는 작업 무관한 공통의 정보, 주요 작업에 대한 공통 정보, 그리고 하위 작업에 대한 고유 정보를 전부 담을 수 있다. HiSS의 우수성을 검증하기 위해 독거 노인을 위한 AI 동반 로봇에서 수집된 효율 데이터를 사용하였고, 이를 통해 HiSS의 성능 뿐만 아니라 복수의 작업을 수행하는 단일 모델의 효율성을 확인하였다. 제안된 알고리즘의 한계점은 여러 작업들 사이에 명시적인 계층 구조를 가져야 한다는 점이다. 향후 연구 주제로는 명시적인 구조가 없는 내재적 구조의 데이터를 처리할 수 있는 알고리즘을 개발하는 것이 포함될 수 있다.

## References

- Caron M, Misra I, Mairal J, Goyal P, and Bojanowski P (2020). Unsupervised learning of visual features by contrasting cluster assignments. In *European Conference on Computer Vision* (pp. 3–20), Springer, Online.
- Caruana R (1997). Multitask learning, *Machine Learning*, **28**, 41–75.
- Chen L, Song J, and Zhang Z (2018). Multi-faceted hierarchical multi-task learning for a large number of tasks with multi-dimensional relations, In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3404–3414.
- Chen T, Kornblith S, Norouzi M, and Hinton G (2020). A simple framework for contrastive learning of visual representations, In *International Conference on Machine Learning*, 1597–1607, PMLR.
- Ester M, Kriegel, Hans-Peter, Sander, Jörg, and Xu, Xiaowei (1996). A density-based algorithm for discovering clusters in large spatial databases with noise, *KDD*, **96**, 226–231.
- Grill J-B, Strub F, Alché F *et al.* (2020). Bootstrap your own latent—a new approach to self-supervised learning, *Advances in Neural Information Processing Systems*, **33**, 21271–21284.
- Hastie T, Friedman J, Tibshirani R, Hastie T, Friedman J, and Tibshirani R (2001). Unsupervised learning, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 437–508.
- Kohonen T (1982). Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, **43**, 59–69.
- Liu Y, Li W, Li H, Zheng Z-M, and Wang S (2015). Multi-task learning for natural language processing, In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2730–2736.
- MacQueen J (1967). Some methods for classification and analysis of multivariate observations, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 281–297.
- Huang, Z., Rao, M., Raju, A., Zhang, Z., Bui, B., and Lee, C. (2022). Multi-task learning for speaker-role adaptation in neural conversation models, In *Proceedings of the 4th Workshop on NLP for Conversational AI 2022*, 120–130.
- Misra I, Shrivastava A, Gupta A, and Hebert M (2016). Cross-stitch networks for multi-task learning, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3994–4003.
- Munkhdalai T and Yu H (2017). Meta networks, In *Proceedings of the 34th International Conference on Machine Learning—Volume 70*, 2554–2563. JMLR. org
- Murtagh F (2014). Ward’s hierarchical agglomerative clustering method: Which algorithms implement ward’s criterion?, *Journal of Classification*, **31**, 274–295.
- Pentina, Anastasia and Lampert, Christoph H (2015). Curriculum learning of multiple tasks, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5492–5500.
- Ruder S (2017). An overview of multi-task learning in deep neural networks, Available from: arXiv preprint arXiv:1706.05098
- Shwartz-Ziv R and Armon A (2022). Tabular data: Deep learning is not all you need, *Information Fusion*, **81**, 84–90.
- Tan P-N, Steinbach M, and Kumar V (2016). *Introduction to Data Mining*, Pearson, Upper Saddle River, New Jersey.

# HiSS: 자기 지도 보조 작업을 결합한 계층적 다중 작업 학습

이승한<sup>a</sup>, 박태영<sup>1,ab</sup>

<sup>a</sup>통계데이터사이언스학과 연세대학교; <sup>b</sup>응용통계학과 연세대학교

---

## 요약

다중 작업 학습은 여러 관련 작업들 사이에서 정보를 공유하며 동시에 학습하는 기계 학습에서 널리 사용되는 방법론이다. 본 논문에서는, 동일한 주요 작업(main task) 하에 속한 하위 작업(sub task)들의 계층적 구조를 고려하며 다중 작업 학습을 수행하기 위한 HiSS (hierarchical multi-task learning with self-supervised auxiliary task)라는 새로운 계층적 다중 작업 학습 방법론을 제안한다. 해당 방법론은 하위 작업을 해결하기 위한 표현 벡터를 학습하기 위해 전역적 공유층, 지역적 공유층, 작업 별 특정층을 활용하는 계층적 구조를 가진다. 또한, 제안한 방법론은 계층적 다중 작업 학습을 주요 과제로 하고, 자기 지도 학습을 보조 과제로 사용하여 학습을 동시에 진행한다. 이는 레이블 없이 입력 데이터만을 활용하여 획득한 군집 레이블을 보조 분류 태스크의 가상 레이블로 사용함으로써, 레이블이 없는 데이터로부터도 추가적인 정보를 획득하고자 함이다. 제안된 접근 방식은 AI 동반 로봇이 수집한 노인 개인의 사용자 정보와 활동 로그로 구성된 효율 데이터를 사용하여 검증되었으며, 시간대와 월을 기반으로 응급 호출을 예측한다. HiSS는 작업의 수에 관계없이 단일 모델만을 필요로 하여 작업에 따라 개별 모델을 사용하는 기존의 기계 학습 알고리즘보다 더 효율적이고, 다양한 메트릭을 사용하여 분류 작업에서 우수한 성능을 확인하였다. 해당 알고리즘에 대한 소스 코드는 다음 링크에서 확인할 수 있다: <https://github.com/seunghan96/HiSS>.

주요용어: 분류분석, 군집분석, 계층적 다중 작업 학습, 머신러닝, 자기 지도 학습

---

<sup>1</sup>교신저자: (03722) 서울시 서대문구 연세로 50, 연세대학교 응용통계학과. E-mail: tpark@yonsei.ac.kr