Towards Minimally Domain-Dependent and Privacy-Preserving Architecture and Algorithms for Digital Me Services: EdNet and MIMIC-III Experiments

Kyoung Jun Lee
Department of Big Data Analytics, Kyung Hee University, Korea
klee@khu.ac.kr

Baek Jeong
Department of Big Data Analytics, Kyung Hee University, Korea
ylbaek@khu.ac.kr

Youngchan Kim
Graduate School of Artificial Intelligence, POSTECH, Korea
kyc618@postech.ac.kr

Suhyeon Kim
Department of Big Data Analytics, Kyung Hee University, Korea
kimsuuuu_99@khu.ac.kr

Abstract

"Digital Me" refers to a service that mirrors an individual's goals, monitors and predicts their status, and provides recommendations to improve the status. This study investigates the architecture and algorithms designed to predict user status and indicate actions, relying solely on the user's data. The objective is to reduce dependency on domain-specific models while promoting the sustainable and privacy-preserving accumulation of data. To validate the proposed architecture and algorithms, we employed the EdNet dataset in the education sector and the MIMIC-III dataset in healthcare. We developed algorithms that recommend activities to optimize users' goal achievement. These algorithms follow a fundamental principle applicable to general Digital Me services: recommending actions most likely to improve the user's subsequent state based on their probability of success and expected outcome. Additionally, we demonstrate the viability of creating effective health prediction algorithms through personal federated learning. This enhances privacy by avoiding centralizing sensitive health data and storing it on individual devices or private clouds.

Keywords: Digital Me, Personalized Federated Learning, EdNet, MIMIC-III

1. Introduction

Digital Me is an AI-driven service that enables real-time management of various aspects of an individual's life, such as health, beauty, memory, and knowledge etc. using models such as Transformer to provide personalized recommendations. The system continuously learns from user interactions and optimize recommendations to align with user's goals. As shown in Figure 1, Digital Me assesses the user's condition through measurement and evaluation, continually improving the user's state while providing recommendations.

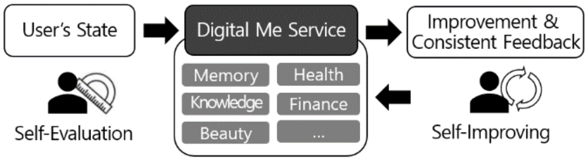


Figure 1. Structure of Digital Me

The development of the Digital Me service algorithm can be broadly divided into two main tasks: predicting states and proposing actions. State prediction involves accurately assessing a user's historical and current conditions and predicting their future state through quantitative analysis (Bai et al., 2021; Pham et al., 2017). Based on the user's goals and the desired state, Digital Me then recommends actions (Davtalab & Alesheikh, 2021; Kulev et al., 2013). To implement Digital Me services, we introduced the AMPER structure, as depicted in Figure 2 (Lee et al., 2022).

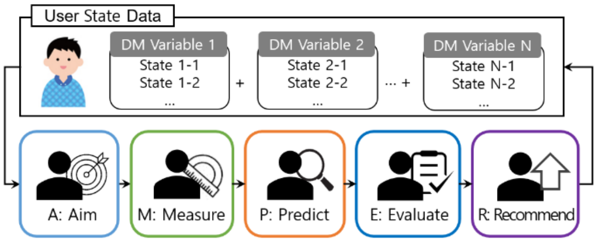


Figure 2. AMPER(Aim-Measure-Predict-Evaluate-Recommend)

The AMPER framework integrates key principles from personalized learning and AI ethics. Personalized learning involves adapting content and pathways based on individual learner needs and preferences, ensuring that each user receives a tailored experience that maximizes their potential (Dandachi, 2024). The framework extends this concept across various life domains. Moreover, AI ethics (Córdova & Vicari, 2022;

Nguyen et al., 2023), particularly concerning data privacy and algorithmic fairness, is central to our approach designed to safeguard user data by minimizing the need for centralized data storage.

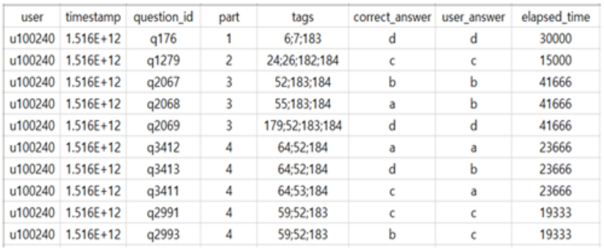
Despite the advances in personalized services, existing algorithms often face significant challenges when applied to across different domains due to their domain-specific dependencies. Cross-domain recommendation systems have attempted to address this issue but still struggle with data dependency and cold-start problems (Zhao et al., 2024; Zhu et al., 2022). To overcome these limitations, we try to find a minimally domain-dependent and privacy-preserving architecture and algorithms (Yan et al., 2022).

The pursuit of minimally domain-dependent algorithms within the 'Digital Me' framework is driven by the need to create versatile, scalable solutions that can adapt across different data domains without relying heavily on domain-specific models. This approach streamlines the development process and enhances the generalizability and applicability of the algorithms across varied data types and user scenarios.

The Digital Me algorithm enhances a user's situation by recommending the most effective actions to achieve a desired state, using data from the user's current condition. The algorithm follows a user-centric approach by aligning with user’s specific aim ('A'). It establishes a measurement ('M') to assess the user's current state, primarily relying on this data and other relevant factors to predict('P') the user’s future states . After evaluating ('E') the user’s potential future states, the algorithm optimizes state improvement by providing recommendations offering a recommendation('R') for actions that lead to a desired outcome. This study addresses a significant need in the existing literature by introducing a domain-independent, privacy-preserving framework that can be effectively applied across multiple domains without requiring extensive adaptation. The AMPER framework generalizes across various sectors, especially education and healthcare.

2. Exploration with EdNet Data

This section illustrates the process of developing minimally domain-dependent prediction and recommendation algorithms using EdNet data, providing insight into the methodology for Digital Me. The EdNet dataset comprises user question-answer records and detailed question information (Choi et al., 2020). We randomly selected 1% of the EdNet-KT1 dataset, which includes 784,309 users studying for the Test of English for International Communication (TOEIC), an English proficiency test widely used to assess non-native speakers' English language skills. After preprocessing, this subset contained data from 7,843 users, 681,618 question-answer instances, and 11,276 unique question categories. The dataset includes the timestamp (indicating when a user began answering a specific question), question ID, the user's response, and elapsed time (duration taken to answer the question), as shown in Figure 3. During preprocessing, we also integrated data indicating answer accuracy ('O' for correct and 'X' for incorrect answers).



user	timestamp	question_id	part	tags	correct_answer	user_answer	elapsed_time
u100240	1.516E+12	q176	1	6;7;183	d	d	30000
u100240	1.516E+12	q1279	2	24;26;182;184	c	c	15000
u100240	1.516E+12	q2067	3	52;183;184	b	b	41666
u100240	1.516E+12	q2068	3	55;183;184	a	b	41666
u100240	1.516E+12	q2069	3	179;52;183;184	d	d	41666
u100240	1.516E+12	q3412	4	64;52;184	a	a	23666
u100240	1.516E+12	q3413	4	64;52;184	d	b	23666
u100240	1.516E+12	q3411	4	64;53;184	c	a	23666
u100240	1.516E+12	q2991	4	59;52;183	c	c	19333
u100240	1.516E+12	q2993	4	59;52;183	b	c	19333

Figure 3. Illustration of Preprocessed EdNet Data

We divided the user's data into clusters based on their chronological order of answering questions. Specifically, each cluster contained 21 units: the first 20 units were used for training the model, while the 21st unit was reserved for validation as a label. This method ensured that the model was trained on sequential data, reflecting the natural progression of the user's learning process. The clusters were created by simply segmenting the dataset into sequential blocks of 21 question-answer instances per user, ensuring that each cluster was representative of the user's ongoing performance. Users with fewer than 21 responses were excluded from the dataset to maintain consistency. As a result, the final dataset included 2,846 users, with 596,020 question-answer pairs for training and 29,801 for validation.

Utilizing the AMPER framework for a Digital Me service tailored to enhance users’ TOEIC test abilities can be summarized as follows:

(A) Aim: Users aspire to improve their English proficiency scores.

(M) Measure: The user's English proficiency is assessed by analyzing their correct responses to questions in the dataset.

(P) Predict: The system predicts the user's likelihood of answering the following question correctly by analyzing their responses to previous questions. For the prediction algorithm, we employed the transformer model (Vaswani et al., 2018), training it with two encoders and two decoders. We compared the correct answer from the validation data with the model’s predictions to determine prediction accuracy. By training on the first 20 questions the user answered, we could predict the response to the 21st question with an accuracy rate of 70.77%. The effectiveness of the transformer algorithm is substantial, especially when compared to a random algorithm, which would have an expected performance of approximately 25%, given that the questions are multiple-choice with four options.

(E) Evaluate: For each TOEIC question, the collective incorrect answer rate is used to score the question, influencing the user's overall question-answer score. The system calculates the user's English proficiency score by assigning a value based on the aggregate error rate for each question. Then, it suggests questions that can most effectively boost the user's score. For example, if Question 2 has an error rate of 70% among all users, it is assigned a value of 0.7 points. This scoring method operates on the premise that questions with higher error rates are more challenging, so a correct response carries more weight. For question No. 511, with an error rate of 45%, a score of 0.45 points is assigned. Similarly, for question No. 82, with a 30% error rate, a value of 0.3 points is assigned. Answering questions No. 511 and No. 82 correctly would accumulate 0.75 points. In this way, the Digital Me service evaluates each user's proficiency level. Table 1, for instance, shows the scores of users A, B, and C for the most recent 20 questions. All users' average +incorrect answer rate increments each correct answer's score for that question. As a result, users A, B, and C have scores of 5.0, 1.7, and 3.1, respectively.

Table 1. User Scores at Current and Next Stages

State	Score	User		
		A	B	C
Current	Score of each User	5.0	1.7	3.1
Next	Score (solving randomly recommended questions)	8.3	4.4	6.0
	Score (solving well-recommended questions)	20.8	18.3	19.1

(R) Recommend: The system suggests questions to expedite the improvement of user's English scores. Of course, it is important to distinguish between mere score maximization and actual improvement in language proficiency. In Table 1, determining the optimal 20 questions to recommend to Users A, B, and C requires a tailored approach. Random selection may present users with questions that are either too difficult, hindering score improvement, or too easy, providing minimal score gains due to low error rates among users. To maximize each user's potential score in the next phase, it is essential to recommend questions that strike a balance: they should be challenging enough to substantially boost the user's score when answered correctly, yet within the user's ability to answer. Leveraging the transformer model, which predicts user responses with approximately 70% accuracy, the most effective strategy involves calculating the probability of a user's correct response for each available question. This approach encourages users to engage with more challenging material, promoting real progress in their English skills by focusing on questions slightly beyond their current ability. This probability is then multiplied by each question's overall incorrect answer rate. The top 20 questions with the highest values are recommended to the user. Although the system focuses on individual user data for recommendations, the underlying prediction model is trained on many users' collective data, utilizing similar users' performance patterns to enhance recommendation accuracy.

A minor challenge with this approach is that the Digital Me service must compute the probabilities for numerous unsolved questions. When scaled to a large user base, the computational demands increase substantially. However, this complexity remains relatively manageable since it grows linearly due to the number of users, questions, and the effort required to execute the transformer model. As illustrated in Table 1, the gains from this recommendation strategy are substantial, demonstrating a marked performance boost compared to random suggestions. Combining the prediction algorithm and the evaluation method created a recommendation algorithm that makes optimal actions possible. Promoting real learning outcomes by optimizing student's actions aligns with educational best practices.

However, this approach may not apply to all Digital Me application domains, particularly in the health sector. Nevertheless, the insights gained from the EdNet experiment are expected to be highly valuable in designing and developing Digital Me services across various domains. Suppose the probability of a user successfully acting can be calculated numerically, and the potential gain from that action can be reasonably estimated. In that case, the Digital Me service can suggest the most optimal action alternatives for the user.

3. Exploration with MIMIC-III Data

In this section, we utilize the MIMIC-III dataset (Johnson et al., 2016a; Johnson et al., 2016b). The Medical Information Mart for Intensive Care III (MIMIC-III) database, developed through a collaboration with MIT, is sourced from the intensive care units of Beth Israel Deaconess Medical Center. It includes de-identified health records for 61,532 patients admitted to intensive care between June 2001 and October 2012, comprising data for 53,432 adults and 8,100 infants. The dataset includes patient demographics, vital signs, laboratory results, medication details, caregiver notes, imaging data, and mortality statuses.

Within the broad scope of Digital Me services in healthcare, we have chosen to focus on hypertension management. Hypertension is a prevalent chronic condition, affecting over 25% of adults aged 30 and above (KOSIS, 2021). Among the global adult population, approximately one in three individuals has hypertension (WHO, 2024). Health-related data is classified as sensitive under the Personal Information Protection Act, making it a matter of significant public concern and requiring careful handling. We aim to demonstrate the capabilities of the personalized federated learning (PFL) approach using the MIMIC-III dataset. This approach is a foundational technology for creating a Digital Me service that leverages health data while maximizing privacy protection.

From the MIMIC-III dataset, we extracted details such as patient number, gender, age, body mass index (BMI), systolic blood pressure (SBP), and the date of blood pressure measurement. To derive the SBP values, we referenced the D_ITEMS table, using the ItemID to fetch the corresponding label (Wang et al., 2020). For patients with multiple blood pressure measurements within a day, we considered only the average (mean) value. During preprocessing, we eliminated blood pressure readings above 400 or below 0. Additionally, age values greater than 89 and below 0 were removed because all patients over 89 in the MIMIC-III database had their age values altered to 300 years. After these preprocessing steps, the dataset was streamlined to include 2,065,091 records spanning 9,908 individuals, averaging approximately 208 records per person.

The data was organized to predict future SBP values by segmenting it into sequential units for each patient. Each patient’s data was divided into five consecutive SBP measurements, with the first four serving as input features and the fifth as the target label. This segmentation allowed the model to learn the temporal patterns in SBP changes. We utilized Time2Vec to convert these time series data into vectors, and employed BERT, enhanced with MA (Moving Average) and ARIMA (Autoregressive Integrated Moving Average) models, to predict future SBP values. The preprocessing steps, including the handling of irregular readings and data segmentation, were crucial to ensuring the model could effectively capture and predict blood pressure trends over time.

The blood pressure management Digital Me service is structured around the AMPER framework, which includes the following components:

(A) Aim: Establish a clear target to improve the user's blood pressure health metrics.

(M) Measure: Evaluate the user's health status by considering gender, age, systolic blood pressure (SBP), and body mass index (BMI).

(P) Predict: Predict the next SBP reading based on the user's historical blood pressure data. The prediction model utilizes the patient's gender, age, body mass index, and previous SBP readings. The data is divided into five units to train the model to predict future SBP values. Each patient’s SBP readings are stored in the same row, with the final reading as the target value (label). For instance, if patient 36’s data in the real MIMIC-III database appears as shown in Table 2, the model is trained with the input data “<esp> M 69 34.44444 <esp> 113.1765 <esp> 117.3333 <esp> 127.7391 <esp> 141.0909 <esp>” and the output "<esp> 143.7857 <esp>" as the predicted SBP.

Table 2. The SBP Time Series of a Patient Over Five Days

Date	Gender	Age	BMI	SBP (mmHg)
21340512	Male	69	34.44	113.1765
21340513	Male	69	34.44	117.3333
21340514	Male	69	34.44	127.7391
21340515	Male	69	34.44	141.0909
21340516	Male	69	34.44	143.7857

Since our data is a time series, we utilized Time2Vec (Kazemi et al., 2019), which converts time series data into vectors. To predict SBP values, we employed bidirectional encoder representations from transformers (BERTs; Devlin et al., 2018), a natural language processing model. However, BERT has limitations in directly handling temporal patterns, especially given the irregular blood pressure readings in the MIMIC-III data. To overcome this, we integrated BERT with MA and ARIMA models. The combined model leverages the linear regularities captured by MA and ARIMA and the non-linear, contextual information learned by BERT, thereby enhancing prediction performance. The evaluation metric used was mean absolute error (MAE), calculated by averaging the absolute differences between the actual and predicted values. The datasets were split into training, validation, and test sets in a 70:15:15 ratio. The baseline BERT model achieved an MAE of 8.42 mmHg. When combined with MA, BERT's MAE slightly improved to 8.36 mmHg; with ARIMA, the MAE further improved to 7.15 mmHg.

(E) Evaluate: The system evaluates a user's blood pressure reading by comparing it against 'Reference Standard of Korean Blood Pressure' provided by the National Health Insurance Service. This standard provides a benchmark for different age groups to determine whether a blood pressure reading is within a healthy range. This comparison is translated into a score using simple code, as shown in Table 3.

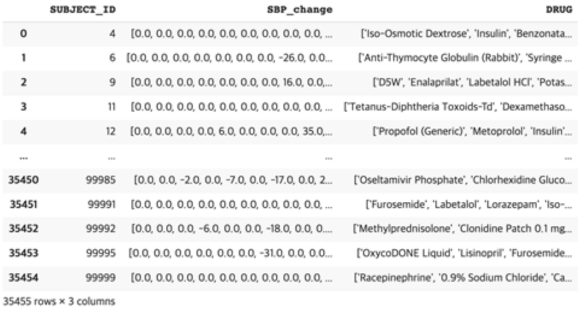
Table 3. Code for Calculating User's Score

Simple Code

```
1. Set bo_table based on gender:
    If gender is 'M', use the male bo_table
    If gender is 'F', use the female bo_table
2. Find the appropriate age_range in bo_table based on user's age
3. Calculate avg_bo using the found age_range
4. Calculate bo_diff as the difference between user's bo and avg_bo
5. If bo_diff is less than 0:
    score = 100 + bo_diff
    score_text = "Low" + score + "point"
  Else:
    score = 100 - bo_diff
    score_text = "High" + score + "point"
6. Return score_text
```

Readings above the standard are labeled "High," while those below the standard are labeled "Low." The final score is calculated by subtracting the deviation from the reference standard from 100. For example, if a 53-year-old man's predicted SBP is 150 mmHg, and the reference standard for his age group is 125 mmHg, the 25 mmHg difference above the standard results in a score of "High" 75 points.

(R) Recommendation Phase: In the context of healthcare's Digital Me, the recommendation phase involves identifying users with similar health trajectories, particularly concerning blood pressure, and suggesting appropriate medical treatments or lifestyle modifications, such as diet or exercise, to improve their health outcomes. The core concept for behavioral recommendations related to blood pressure is to identify other patients whose recent blood pressure changes closely resemble those of the user, utilizing cosine similarity as the matching metric. If these patients have experienced improvements in their blood pressure, the medication they received may be recommended to the user. To implement this approach, data was extracted from the MIMIC-III dataset, yielding 51,076,814 instances where SBP measurements were matched with corresponding medication prescriptions on the same dates, creating a robust database. Blood pressure variations and prescribed medications for 35,455 patients were identified (as shown in Figure 4). This data enabled the development of a strategy to identify patients with blood pressure changes similar to those of a given patient and to recommend medications that have proven effective in improving blood pressure in similar cases.



	SUBJECT_ID	SBP_change	DRUG
0	4	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	['Iso-Osmotic Dextrose', 'Insulin', 'Benzonata...
1	6	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, -26.0, 0.0...	['Anti-Thymocyte Globulin (Rabbit)', 'Syringe ...
2	9	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 16.0, 0.0,...	['D5W', 'Enalaprilat', 'Labetalol HCl', 'Potas...
3	11	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	['Tetanus-Diphtheria Toxoids-Td', 'Dexamethaso...
4	12	[0.0, 0.0, 0.0, 0.0, 6.0, 0.0, 0.0, 0.0, 35.0,...	['Propofol (Generic)', 'Metoprolol', 'Insulin'...
...	...	...	...
35450	99985	[0.0, 0.0, -2.0, 0.0, -7.0, 0.0, -17.0, 0.0, 2...	['Oseltamivir Phosphate', 'Chlorhexidine Gluco...
35451	99991	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	['Furosemide', 'Labetalol', 'Lorazepam', 'Iso-...
35452	99992	[0.0, 0.0, 0.0, -6.0, 0.0, 0.0, -18.0, 0.0, 0....	['Methylprednisolone', 'Clonidine Patch 0.1 mg...
35453	99995	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, -31.0, 0.0, 0.0...	['OxycoDONE Liquid', 'Lisinopril', 'Furosemide...
35454	99999	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	['Racepinephrine', '0.9% Sodium Chloride', 'Ca...

35455 rows × 3 columns

Figure 4. Blood Pressure Changes and Medication Mapping

We compared the medication lists of 3,545 patients, representing 10% of the randomly selected dataset, with our recommended medications. The recommendation was considered accurate if any suggested medications were on the list. This method resulted in an accuracy rate of 74.01%. However, this approach has a limitation: the algorithm may be biased by the data it was trained on. For instance, if the training data primarily involves medical treatments (e.g., medications), the recommendations may be skewed towards similar interventions, potentially overlooking non-medical solutions like lifestyle changes. This bias could inadvertently group patients with similar SBP profiles but different underlying conditions or health needs, leading to suboptimal recommendations. We acknowledge this as a limitation of our current research and suggest that future studies explore more diverse datasets to mitigate this bias.

4. Personalized Federated Learning Experiment

This section explores the feasibility of developing sufficiently performant prediction algorithms using PFL (Fallah et al., 2020; Tan et al., 2022) without consolidating individual data into a single database. Data is stored on each individual's device or in a private cloud. Ensuring user privacy is paramount for Digital Me services, particularly in the healthcare sector. We applied the prediction algorithm from the previous section to assess its performance within a PFL environment (Figure 5).

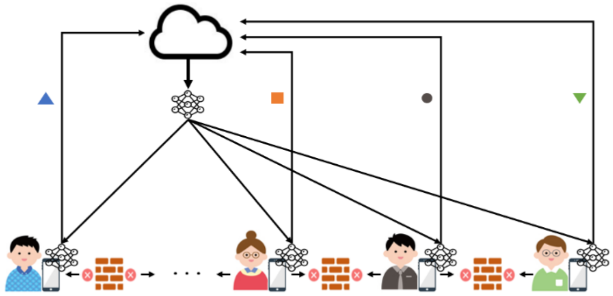


Figure 5. Personalized Federated Learning

The personalized federated learning experiment utilized the Per-FedAvg algorithm (Fallah et al., 2020), a method inspired by Model-agnostic Meta Learning (MAML) that seeks to identify a robust initialization to quickly adapt to varying client data distributions. The PFL experiment segmented data by patient ID and virtually allocated it to distinct devices. Each device corresponds to a virtual representation of a user's data, effectively simulating a real-world environment where data is distributed across multiple devices. Specifically, the dataset comprised 9,908 users, totaling 2,065,091 records, corresponding to an average of 208 days' data per user. The user with the most extensive dataset had data spanning 26,282 days, while the one with the least had data for only two days. To implement this, we first preprocessed the data by segmenting it chronologically for each patient. Each segment of data was then assigned to its respective "device" to simulate the federated learning environment.

We employed the BERT with ARIMA Model, which had demonstrated superior performance in previous tests, to train local models on each device using its respective data. The ARIMA+BERT model is a hybrid approach that combines the strengths of ARIMA for handling time series data with the transformer-based BERT model. ARIMA is particularly effective for capturing linear trends and patterns in time series data, while BERT, a powerful NLP model, is adept at learning complex, non-linear relationships and contextual information. By combining these models, the hybrid approach aims to leverage ARIMA’s capacity for modeling temporal dynamics with BERT’s ability to learn intricate patterns, leading to improved prediction accuracy.

In this experiment, we observed a significant enhancement in model performance even with a single round of iteration. The average MAE (Mean Absolute Error) of the local model was 11.70 mmHg, whereas the PFL (Personalized Federated Learning) model achieved an MAE of 8.03 mmHg, representing an improvement of over 30%. Notably, the data-shared model, which integrates all available data, achieved the lowest MAE at 7.15 mmHg, demonstrating the highest accuracy. However, the performance of the PFL model was very close to that of the data-shared model, indicating that even a single iteration was sufficient to explore the potential for significant performance improvements. Typically, multiple rounds are necessary to fully optimize the model, but the results from this limited iteration still underscored the effectiveness of our approach.

To further validate the effectiveness of the AMPER framework in a federated learning environment, we compared it with state-of-the-art methods in personalized federated recommendation systems. Luo et al. (2022) demonstrated the use of graph neural networks in federated learning to cluster users and adapt models for personalized recommendations across diverse data sources. Mmeta-learning-based approaches also have been effective in handling Non-IID data, a big challenge in federated learning (Jeong & Hwang, 2022) . Our approach aligns with these advanced methods by incorporating user clustering and model adaptation to improve performance while ensuring privacy.

Additionally, PFL offers significant advantages especially in handling sensitive data. In centralized learning, user data must be aggregated into a single database, raising concerns about data privacy and security breaches. In contrast, PFL keeps data localized on individual devices, reducing the risk of data exposure while still achieving near-centralized performance. This approach is particularly advantageous in healthcare, where maintaining patient confidentiality is critical. The reduction in MAE observed in our PFL model underscores its capability to deliver high-performance predictions while preserving user privacy, making it a robust alternative to traditional centralized methods.

The local model exhibited a standard deviation of 10.06 mmHg, whereas the PFL model demonstrated a reduced standard deviation of 7.65 mmHg. This reduction is significant as it indicates that PFL narrows the performance gap among individual local models. Furthermore, when comparing PFL performance across 9,908 patients versus 1,000 patients, there was a noticeable increase in the average MAE for the 1,000 patients, reaching 15.33 mmHg. This finding reinforces the understanding that the performance of PFL improves with a larger number of subjects (Kim et al., 2023).

However, a limitation of PFL is that each client′s data can have different distributions, which may negatively impact model performance. For instance, when the data is highly skewed or limited, the model may not generalize well across different devices, leading to potential overfitting. Addressing these challenges will require refining the PFL methodology, potentially by integrating techniques that can better handle diverse data distributions or by developing more robust model validation strategies.

5. Discussion

A phased approach using the AMPER framework is proposed to develop Digital Me services. Data is generated through the Aim and Measure stages, after which universal AI models, such as transformers and BERT, are employed in the Prediction stage. These models do not require domain-specific knowledge, thereby substantially reducing the need for such expertise and allowing the application of Digital Me services across various domains. Domain knowledge is necessary only in the Evaluation stage. In the final stage Recommendation, it is critical to identify actions that will most effectively enhance the user's score.

In the EdNet case, the system evaluated all questions by leveraging the question-answer records from other users' data. The MIMIC-III case demonstrated that identifying individuals with similar health data and recommending actions that improved blood pressure could effectively address hypertension management challenges.

In both cases, recommendations were derived by leveraging data from users with similar characteristics or performance patterns, enhancing accuracy and relevance. These two case studies showed the feasibility of implementing Digital Me services using only data without requiring extensive domain knowledge.

Given relevant domain-specific evaluation criteria, universal predictive models can offer personalized recommendations tailored to users' needs. As a result, Digital Me services can be effectively implemented across various domains, such as education and healthcare, providing tailored user experiences that address the unique needs of each sector.

However, implementing the proposed approach in real-world settings presents several practical challenges. One significant challenge is ensuring scalability, particularly in environments with vast and diverse datasets. As the number of users and the variety of data increase, the computational and communication overhead may become substantial, potentially affecting the system's responsiveness and performance. Additionally, the balance between privacy and performance is a critical consideration. While PFL offers significant privacy advantages by keeping data localized, it may also introduce challenges such as model divergence and slower convergence, especially when dealing with highly heterogeneous data distributions across clients. These challenges necessitate careful consideration and optimization to ensure the robust and efficient operation of Digital Me services in practical applications.

6. Summary and Conclusion

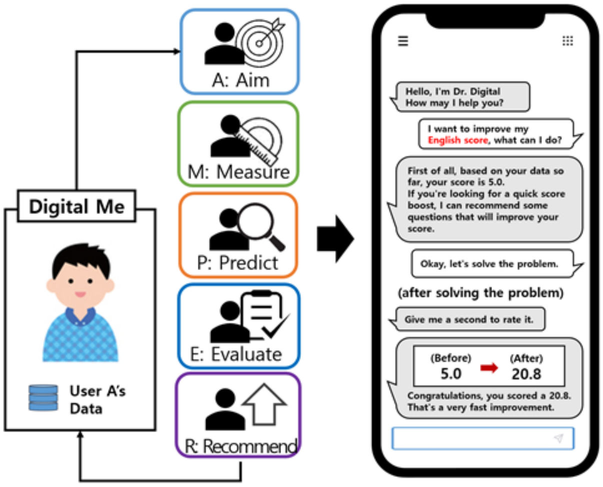
This study illustrates efforts to develop algorithms that predict a user's status and recommend desirable actions based solely on the user's data, minimizing reliance on domain-dependent models while sustainably accumulating user data in a privacy-preserving manner. We identified a principle for recommendation algorithms that can be broadly applied to Digital Me services: recommend actions that are most likely to improve the user's next state, considering both the likelihood of the user successfully acting and the score they will receive. Additionally, we demonstrate that it is possible to develop sufficiently performant health prediction algorithms using PFL without centralizing individual health information in a single database, instead storing it on each individual's device or in a private cloud.

To verify this concept, we utilized data from EdNet and MIMIC-III. Combining or sharing user data in healthcare is generally undesirable due to its sensitivity. Therefore, PFL was applied, resulting in a performance improvement of over 30% compared to the local model (Kim et al., 2023).

By keeping data localized on users' devices, PFL enhances user trust in the system. In addition, this decentralized approach not only enhances the safeguarding of personal information but also supports compliance with stringent data protection regulations, making it particularly suitable for sensitive domains like healthcare (Nori et al., 2024). However, full compliance also requires additional measures such as encryption and secure access controls to ensure that all aspects of data protection are addressed.

The AMPER framework's design inherently supports domain-independence by utilizing universal predictive models like transformers and BERT, which are not tied to specific domain knowledge. This allows the Digital Me service to be easily adapted across various fields such as education, healthcare, and finance, without the need for extensive customization for each domain. This capability significantly reduces the dependency on domain-specific models, thereby enhancing the scalability and applicability of the service across different sectors (Blodgett et al., 2023; Zhang et al., 2024).

To integrate the Digital Me algorithm into real-world services, we explore a conversational Digital Me service enhanced by a large language model (LLM) (Figure 6). This conversational interface aims to develop a Digital Me service that assesses a user's status in alignment with their personalized objectives based on their data. The service will naturally suggest optimal actions through interactive dialogues. Additionally, we are developing a small LLM (s-LLM), such as LLaMA2 (Touvron et al., 2023), capable of independently interpreting sensitive personal details, including health metrics, directly on the device (Qiu et al., 2023). This model will facilitate private conversations with users, offering insights derived from their data. By combining the strengths of PFL and the AMPER framework, a Digital Me service can be developed to offer a secure, adaptable, and domain-independent service that can meet the personalized needs of users across various domains. We believe incorporating the PFL approach and localized s-LLM is indispensable to ensure a secure, trustworthy, and user-friendly service that confidently oversees and enhances an individual's well-being (Kim, 2024; Qiu, 2024).



Hello, I'm Dr. Digital
How may I help you?

I want to improve my English score, what can I do?

First of all, based on your data so far, your score is 5.0.
If you're looking for a quick score boost, I can recommend some questions that will improve your score.

Okay, let's solve the problem.

(after solving the problem)

Give me a second to rate it.

(Before) 5.0 → (After) 20.8

Congratulations, you scored a 20.8.
That's a very fast improvement.

Figure 6. Conversational Digital Me Service Interface

7. References

Bai, X., Zhan, F., Li, J., Guo, T., Aziz, A., Jin, A., Xia, F. (2021). Educational big data: Predictions, applications and challenges. Big Data Research, 26, 100270.

Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2023). Bias and Fairness in Large Language Models: A Survey. Computational Linguistics, 1-83.

Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: Forecasting and control. Holden-Day.

Choi, Y., Lee, Y., Shin, D., Cho, J., Park, S., Lee, S., Baek, J., Bae, C., Kim, B., & Heo, J. (2020). EdNet: A large-scale hierarchical dataset in education. In Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, 2(21), 69-73.

Córdova, P. R., & Vicari, R. M. (2022). Practical ethical issues for artificial intelligence in education. In International Conference on Technology and Innovation in Learning, Teaching and Education, 437-445.

Dandachi, I. E. (2024). AI-powered personalized learning: Toward sustainable education. In Navigating the Intersection of Business, Sustainability and Technology, 109-118.

Davtalab, M., & Alesheikh, A. (2021). A multi-criteria point of interest recommendation using the dominance concept. Journal of Ambient Intelligence and Humanized Computing, 1-16.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Fallah, A., Mokhtari, A., & Ozdaglar, A. (2020). Personalized federated learning: A meta-learning approach. arXiv preprint arXiv:2002.07948.

Jeong, W., & Hwang, S. J. (2022). Factorized-fl: Personalized federated learning with parameter factorization & similarity matching. Advances in Neural Information Processing Systems, 35, 35684-35695.

Johnson, A., Pollard, T., & Mark, R. (2016a). MIMIC-III Clinical Database (version 1.4). PhysioNet.

Johnson, A., Pollard, T., Shen, L., Lehman, L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L., & Mark, R. (2016b). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035.

Kazemi, S. M., Goel, R., Eghbali, S., Ramanan, J., Sahota, J., Thakur, S., Wu, S., Smyth, C., Poupart, P., & Brubaker, M. (2019). Time2vec: Learning a vector representation of time. arXiv preprint arXiv:1907.05321.

Kim, S. (2024). Personalized federated learning for healthcare digital me services. Master's thesis, Kyung Hee University.

Kim, S., Jeong, B., & Lee, K. J. (2023). Blood pressure management system with personalized federated learning: Based on the Digital Me algorithm. Paper presented at the 25th Convergence Conference of the Korea Academic Society for Business Administration. Busan, Korea, August 16-18.

Kim, S., Qiu, Y., Jeong, B., Ok, K., Piao, Y., & Lee, K. J. (2023). A study on the structure of personalized federated learning using a small/standalone large language model: Application of healthcare services. 2023 Joint Spring Conference of the Korea Society of Management Information Systems.

KOSIS. (2021). Hypertension prevalence trend. Retrieved from https://kosis.kr/statHtml/statHtml.do?orgId=177&tblId=DT_11702_N105&vw_cd=MT_ZTITLE&list_id=&scrId=&seqNo=&lang_mode=ko&obj_var_id=&itm_id=&conn_path=E1&docId=01482&markType=S&itmNm=%EC%A0%84%EA%B5%AD. Accessed: 2024-06-14.

Kulev, I., Vlahu-Gjorgievska, E., Trajkovik, V., & Koceski, S. (2013). Recommendation algorithm based on collaborative filtering and its application in health care. The 10th Conference for Informatics and Information Technology, CIIT 2013.

Lee, K. J., Jeong, B., Hwangbo, Y., Kim, Y., Bae, S., & Baek, T. (2022). AMPER (Aim-Measure-Predict-Evaluate-Recommend): The paradigm of Digital Me. The 23rd International Conference on Electronic Commerce. Daegu, Korea, June 22-23.

Luo, S., Xiao, Y., & Song, L. (2022). Personalized federated recommendation via joint representation learning, user clustering, and model adaptation. In Proceedings of the 31st ACM international conference on information & knowledge management, 4289-4293.

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. Education and Information Technologies, 28(4), 4221-4241.

Nori, H., King, N., McKinney, S. M., Carignan, D., Horvitz, E., & Liu, Y. (2024). Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nature Medicine, 30(8), 1577-1587.

Pham, T., Tran, T., Phung, D., & Venkatesh, S. (2017). Predicting healthcare trajectories from medical records: A deep learning approach. Journal of Biomedical Informatics, 69, 218-229.

Qiu, Y. (2024). Developing conversational service pilot by fine-tuning sLLMs with real and synthetic data. Master's thesis, Kyung Hee University.

Qiu, Y., Kim, S., Lee, Y., Ok, K., Jeong, B., & Lee, K. J. (2023). A Study on Healthcare sLLM Fine-tuning Based on Federated Learning: Utilizing Medical Consultation Data. 2023 Fall Conference of the Korean Institute of Information Systems.

Tan, A. Z., Yu, H., Cui, L., & Yang, Q. (2022). Towards personalized federated learning. IEEE Transactions on Neural Networks and Learning Systems, 23(1), 1-17.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

Wang, S., McDermott, M. B., Chauhan, G., Ghassemi, M., Hughes, M. C., & Naumann, T. (2020). MIMIC-Extract: A data extraction, preprocessing, and representation pipeline for MIMIC-III. In Proceedings of the ACM Conference on Health, Inference, and Learning, 222-235.

WHO. (2024). World Hypertension Day 2024: Measure Your Blood Pressure Accurately, Control It, Live Longer. Retrieved from https://www.who.int/srilanka/news/detail/17-05-2024-world-hypertension-day-2024--measure-your-blood-pressure-accurately--control-it--live-longer. Accessed: 2024-08-25.

Yan, D., Zhao, Y., Yang, Z., Jin, Y., & Zhang, Y. (2022). FedCDR: Privacy-preserving federated cross-domain recommendation. Digital Communications and Networks, 8(4), 552-560.

Zhang, Z., Carlini, N., Recht, B., Shankar, V., & Schmidt, L. (2024). AI models collapse when trained on recursively generated data. Nature, 616(7968), 508-512.

Zhao, P., Jin, Y., Ren, X., & Li, Y. (2024). A personalized cross-domain recommendation with federated meta learning. Multimedia Tools and Applications, 1-16.

Zhu, Y., Tang, Z., Liu, Y., Zhuang, F., Xie, R., Zhang, X., Lin, L., & He, Q. (2022). Personalized transfer of user preferences for cross-domain recommendation. In Proceedings of the fifteenth ACM international conference on web search and data mining, 1507-1515.