# Study on Distributed Models Combining Method for Performance and Efficiency

1st TSUCHIYA Takeshi
*Institute for Data Science Education*
*Tokyo International University*
Tokyo, Japan
ttsuchi@tiu.ac.jp

2nd Ryuichi Mochizuki
*MR Systems*
Nagano, Japan
ryuichi_mochizuki@protonmail.com

3rd Hiroo Hirose
*dept. of Applied Information*
*Suwa University of Science*
Nagano, Japan
hirose@rs.sus.ac.jp

4th Hiroshi Ichikawa
*Otsuma Women's University*
Tokyo, Japan
ichikawa.h@otsuma.ac.jp

***Abstract*—The authors proposed a distributed machine learning method that generates and combines machine learning models created by group set of users in previous study. This paper discusses a method for selecting distributed models to efficiently construct machine learning models. First, a method for selecting models based on their similarity is discussed. The evaluation shows that selecting models only based on similarityis problematic. Next, to improve the performance of individual machine learning models for each set, we integrated pre-trained base learning models. However, we discover that its performance doesn't improve considerably when compared to the typical one without pre-training.**

***Index Terms*—distributed machine learning, model similarity, model combinations**

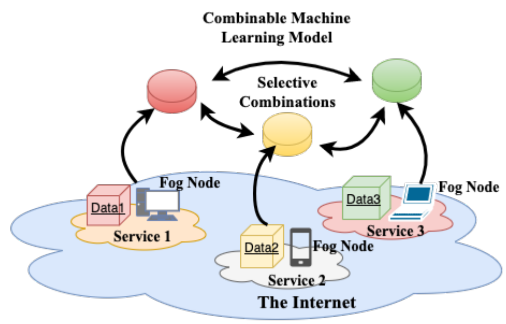


Fig. 1. Conmbinations of Machine Learning Models

## I. Introduction

In recent years, AI-powered services have been trequired to independently collect data, which is done for each service. However, it is extremely difficult to collect the essential data for tiny businesses with insufficient customers, or for new services that have just beguns operations.

In our previous studies, [1], [2], we proposed a storategy for constructing models with arbitrary performance by combining machine learning models. Machine learning models, as non-reversible objects formed from data, have the feature of safegurding data privacy even when the model itself is released. Therefore, machine learning models are created in small units such as a group of users. The models created by integrating these models have the same or better performance than the traditional way of accumulating data on a server. In the previous study, this machine learning model that can be combined was referred as a 'fog model', and will be referred to as such in this paper.

A approache uses the similarity of the fog models as a parameter to select the fog model to be combined in the study [2]. This storategy is thought to strike an advantageous balance between the model performance improvement and the processing load of the fog model combination. First, this study shows how combining improves fog model performance and the changes its similarity due to combining. Then, the criteria required to increase the performance of machine learning models by integrating them are explored based on these results.

This research is supported by Grant-in Aid for Scientific Research (C), 2021-2024 21K11850

## II. Distributed Machine Learning Model

This section provides an overview of our prior study [1] [2]. Due to data privacy and other constraints, data collected via web services cannot be shared with other services. In this situation, machine learning models, which are irreversible objects created from the data, are integrated to obtain their features. This enables for the same impact to be achieved without explicitly learning the data.

Fig. 1 shows a group of users, each of whom is a unit for sharing data. Each set independently creates a machine learning model. Users can choose a model to be combined from these generated models, resulting in a model tailored to their purpose.

### A. Data Management

Each group of users has their own shared data, which is controlled independently. The fog node is not a fixed node, but rather a role allocated to one of the nodes in the set. Each group of nodes collects shared data and uses it to create machine learning models. Fog nodes in each node set do not share or exchange the data they controll. However, the node is assigned an ID, and frequent synchronization is performed to ensure that the existence of the fog node.

### B. Combinations of Fog Models

The user combines fog models generated within a node set. However, the fog models developed separately for each set are not compatible as is. As a result, the dimensionality information of each created model is synchronized using the method showned in the II-A. Then, the dimensional information is synced, allowing for the construction of comparable models.

This study is based on textual data. Each user set retrieves words from the text information stored in the data. Using information synchronization, IDs are allocated to each word, which is then shared throughout the fog nodes. If a word does not appear in this information, a new ID is assigned, and the synced data is updated. This synchronized word information is then used as dimensionality information to create a machine learning model that can be integrated. Each fog node creates a machine learning model using the data handled by each set based on the synchronization information. Because each fog node is generated with different input, the values of each dimension of the resulting model will differ.

### C. Combination Methods for Fog models

Fog models with identical dimensional data are combined by calculating the total of each dimension and normalizing it. It is feasible to combine the models and retain the features. The user picks and combines the fog models.

”SeQuential Combining Selection” (SQ) combines fog models in the order of their fog node IDs, which manage the created models. This method combines all of the models without specifying the goal combination. This method is similar to the traditional way for collecting all data.

”Adaptive Selection” (AS) arranges fog models in the order of their fog node IDs. Then it does simple assessments to determine whether to keep or eliminate the combo. The assessment is based on extracting data from the target data to indicate the task’s characteristics. Because the fog nodes to be combined are chosen from models that demonstrate performance improvement, the performance gain from combining is expected to be high. On the other side, this strategy necessitates a join and review procedure for each join choice.

“ Similar Model Selection ”(SM) is a strategy that considers the similarity of fog models. The fog models are combined based on their resemblance. This method can be deemed practical because it just determines model similarity.

### D. Performance Comparison

A comparative evaluation of combining strategies is depicted in Fig.2. Wikipedia [5] information is utilized as the data. This evaluation employs 100 fog nodes (node set 100) to partition the data into 100 segments. Currently, each fog node receives more than 1000 contents. Each fog node creates a fog model with Word2Vec [6]. The association between the number of models combined or considered for combination and performance is assessed.

The evaluation results are displayed in Fig.2. In this graph, the $x$-axis represents the number of fog models combined or considered for combination, while the $y$-axis represents each fog model ’s accuracy.

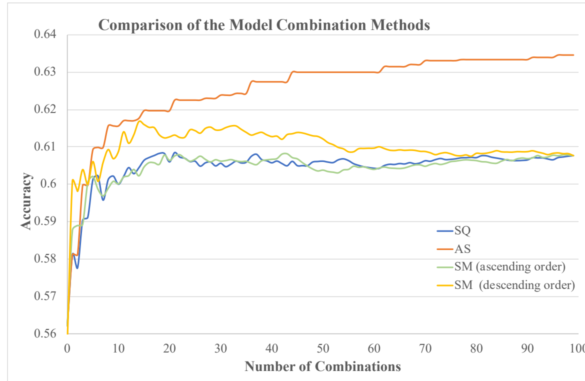


Fig. 2. Comparison Results of The Combinations

SQ is a simple combination method, but when around 10 of fog models are combined, the performance is practically constant. The performance of SQ is expected given that the accuracy of the typical method of data accumulation is 0.6.

AS clearly enhances accuracy when compared to other approaches. This is due to the fact that combining is only done in circumstances of increased precision. However, this is impractical because the evaluation must be performed each time to determine combining.

SM determines which fog models to combine based on similarity. SM (ascending order) depicts the scenario in which fog models with high similarity are integrated first; the performance is not considerably different from SQ, and there are no notable features. In contrast, SM (descending order) depicts the circumstance where fog models with low similarity are combined. The performance is comparable to AS when there are fewer than 15 combined fog models. The performance is superior to SQ for small number of combinations.

According to the above evaluation, AS is a superior combination method in the prior study. However, this method limits its availability. Furthermore, SM (descending order) offers comparable accuracy in ranges with a small number of choices.

## III. Fog Model Combinations

This section discusses how the fog model’s features change as a result of combining.

### A. Similarity of Fog Models

The SM used the similarity of the fog models as a signal to select the target for combining. According to our initial assumption, each set could not create a fog model with adequate accuracy solely using their own managed data. Combining fog models trained on data with similar properties could supplement missing data training. However, the data shown in Fig. 2 contradicted this notion.

Now, comparing the fog models generated in each set, Wikipedia ’s 100000 contents are divided into 100 fog nodes,

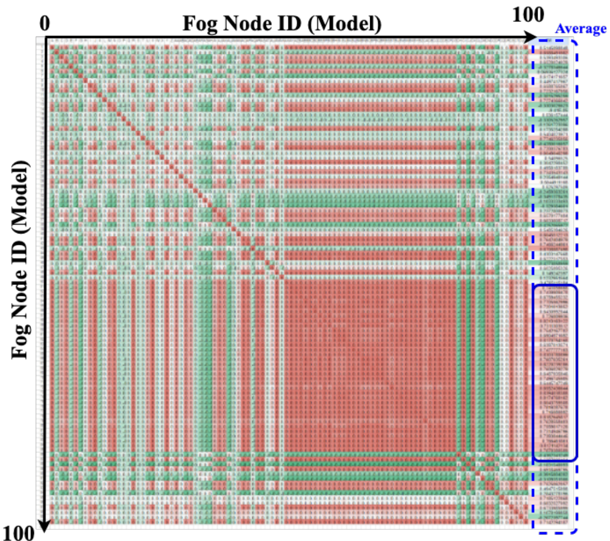


Fig. 3. Heat Map of Fog Model Similarity

with each fog model is generated. The similarity of the fog models is determined, and a heat map is depicted in Fig. 3.

Fig. 3 depicts the correlation matrix derived from cos similarity for each fog model in the order of fog node IDs. The cos similarity matrix is labeled red for high similarity, white for intermediate similarity, and green for low similarity. The findings indicate that while not all fog models are similar, they do share some characteristics. The similarity between fog models trained on completely different data is analogous to similar topics and expressions. On the other hand, fog models with low similarity can be thought of as having distinct topics. Our intuition led us to believe that, for fog models with insufficient training, combining fog models with related topics would improve their performance.

The right side of Fig. 3 (the wavy line labeled Average) displays a heat map of the similarity to the combined model of all fog models. The result is identical to the final model that combines all fog models in SQ. In this case, the color of the items with the highest similarity to the conventional fog model is lighter, indicating that the similarity is decreasing. In the area surrounded by the solid line, the overall similarity with the fog model appears to be lighter, indicating a decrease. The decrease in similarity for each item could be attributed to the use of fog models with different topics.

Consider the accuracy of the combined fog model. The combined fog model without selecting all of the fog models is equivalent to the combined 100 model in SQ shown in Fig. 2. It can be argued that the accuracy is at least superior to that of the uncombined fog model. Therefore, it is possible to conclude that the similarity of the fog models is insufficient as a selection criterion in the current situation.

### B. Fog Model Changes in AS

In the AS, combining is only used when the accuracy improves. Fig. 4 depicts how the similarity changes as the fog models are combined. It displays the similarity to the 100 fog models (horizontal axis) as shown in Fig. 3, as well as a heat map depicting the change in similarity during the

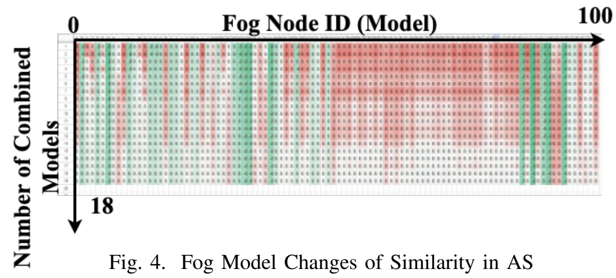


Fig. 4. Fog Model Changes of Similarity in AS

AS-combined fog model process. The vertical axis shows the number of combined fog models, which was 19 models in this case, as combining is only for accuracy improvement.

The figure depicts how the fog models with high similarity (shown in red) become more similar to the white models with low similarity as the number of combined fog models increases. The fog models with low similarity and green color initially become lighter or remain unchanged as the number of combined fog models increases. Fig. 2 shows that combined fog models improve accuracy. The combination tends to shift the similarity between the fog models in the direction of divergence. And as the accuracy improves, it is discovered to differ significantly from the initial fog model similarity. This suggests that there is some relationship between accuracy and similarity. Although the correlation itself is unclear in 4.1, it is clear that there is some relationship between accuracy and similarity, albeit not as straightforward as one might expect.

### C. Considerations

Previous research has shown that combining fog models to improve accuracy. To accomplish this, fog models from various target topics with low similarities must be combined. This means that the difference in accuracy between the combined fog model with SM (descending order) and AS cannot be explained in this simple example. This means that multiple topics must be addressed in the fog model. Simultaneously, the similar fog models on related topics must be improved.

Each fog model has lower accuracy than the machine learning model created by data accumulation. This is because the data owned by each fog node is part of the training data. In this method, each set's training data is managed differently. Thus, each fog model is optimized for its specific content during the training phase. Even fog models with a high similarity will not match unless all of the content is the same. Combining fog models with high similarity effectively eliminates over-fitting, which is the optimized part of each set. Combining a number of very similar fog models is effective.

Based on our previous discussions, it appears to be effective to combine the fog models based on their similarity. However, selecting fog models based on similarity (i.e., nearness or distance) is insufficient as shown in Section II-D. It is necessary to correct for over-fitting to specific content by combining closely related fog models with homogeneous topics. Simultaneously, it must adapt to multiple topics by combining distant fog models (with different topics). The similarity of the fog models is an important criterion for selecting the fog models to combine.

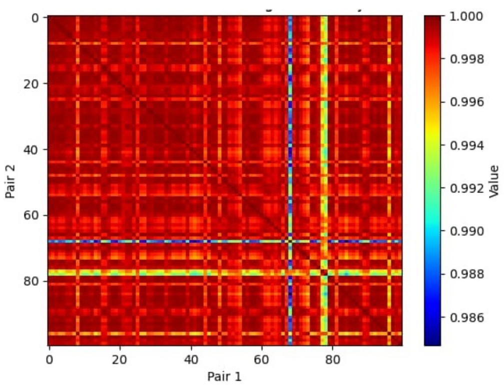

Fig. 5. Pre-trained Fog Model of Similarity

In a related study, BERT [8] assigns a Large Language Model (LLM) to each user as a basic model trained on a diverse and large amount of data, regardless of the target task. Then, the model is further trained on the user 's personal data to optimize it for the task at hand. In conjunction with the research presented in this paper, LLM is used to adapt to various topics. The optimization is performed to adapt to the user's content and task.

Though this study assumes combining the multiple fog models, each fog model represents the data of each set as a feature. The goal is to achieve performance without directly training the data. Previous research has identified issues such as uneven distribution of contents for each set and over-fitting of fog models to specific topics. However, because the fog models themselves are stable, this has no significant effect on performance. Furthermore, the basic fog model performance is expected to be stable performance even if the fog models are not properly combined. The contents of each set can be handled by combining of the fog models.

## IV. Use of Pre-Trained Model

This section compares the similarity of the fog models to pre-trained machine learning models.

### A. Generating Fog Model using Pre-Trained Model

This paper uses text8 [9], a compressed corpus of past Wikipedia articles, to construct a pre-training model. This corpus is sampled so that each word appears equally. Thus, no specific fields or words are distributed unevenly. Machine learning models that have been pre-trained on this corpus are distributed among the sets. Each set performs additional training with its own data. The fog models are synchronized, allowing them to be combined. The similarities between fog models are compared using pre-trained models.

Fig. 5 depicts the similarity of the fog models distributed across 100 nodes. As in Fig. 3, both the vertical and horizontal axes represent the fog node IDs (1˜100). A heat map is used to illustrate the similarity in the figure. In this case, most of the figure is red, with only a few fog models showing different colors. The red areas on the scale indicate the

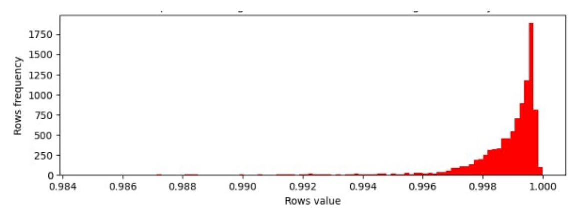

Fig. 6. Pre-trained Fog Model Distribution of Similarity

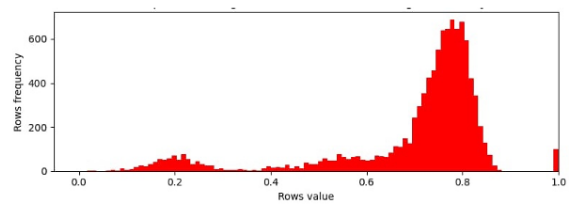

Fig. 7. Normal Fog Model Distribution of Similarity

similarity $0.996 \leqq similarity \leqq 1$, and the blue areas indicate the similarity $0.985 \leqq similarity$, indicating that all fog models are highly similar. This is the result of using the same pre-trained model for each set. Although additional training is performed for each set, the effect is minimal, and the similarity is extremely high due to the use of the same pre-trained fog model as the base.

The similarity of the fog models is depicted as a frequency distribution in Fig. 5. For comparison, the results from Fig. 3 are presented with a frequency distribution.

The distribution of the fog model is mostly around 0.996 or higher. The authors 'proposed method without prior learning in Fig. 7, shows that the similarity of the fog models is distributed in the range of most $0.4 \leqq similarity \leqq 0.9$ and in the wide range of $0.1 \leqq similarity \leqq 0.3$. Based on the foregoing, it is reasonable to conclude that fog model selection based on similarity indices has no meaning for fog models using pre-training.

### B. Performance Change by Pre-Trained Model

The performance of a fog model with pre-training is compared to that of the traditional method. This evaluation compares the performance of three fog models; the common pre-trained model (with no additional training), a fog model that combines the pre-trained model with additional training, and a fog model generated using the conventional method and combined by SQ. The accuracy of these fog models is compared when they are used for various tasks. The evaluation tasks included the word similarity evaluation tasks SL [10], WR, WS [7], and the word analogy task QW [11]. The comparative evaluation of multiple tasks confirmed that performance is not biased towards a specific task.

The accuracy for each task is depicted in Fig. 8. Each fog model lacks significant characteristics. This is simply a comparison of fog model performance, as there is no direct relationship between the evaluation tasks and the data managed

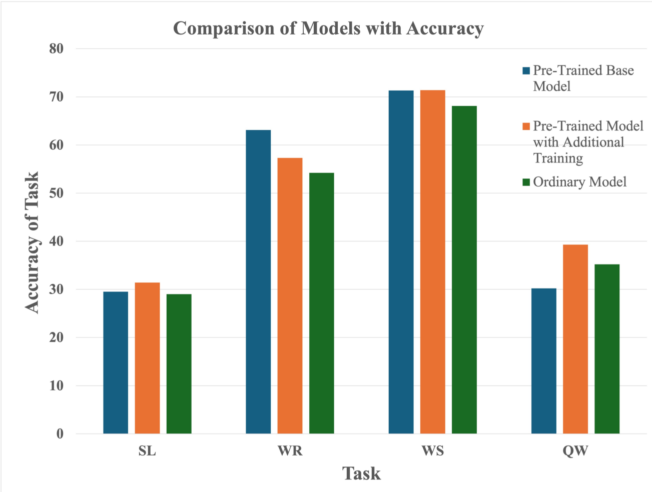


Fig. 8. Comparison of Model Accuracy for Tasks

by the fog nodes. In this case, the evaluation results vary depending on the task. There are no significant differences between the fog models with and without pre-training, and no significant differences result from additional training. The SQs in Fig. 2 perform similarly, so the results of this evaluation can be assumed to be generalizable. Therefore, the performance of the fog model based on the pre-trained model remains unchanged. On certain types of tasks, the use of pre-trained models improves performance. However, the improvement is not significant, at around 3˜4%, which is within the margins of each task.

The results of the evaluation revealed that the performance of each fog model was stable by using a common model that was pre-trained. At the same time, each fog model exhibited a high level of similarity. However, the effect of combining the fog models could not be seen in SQ, which combined all of the fog models. Thus, the similarities do not clearly indicate which fog models should be combined. It is necessary to specify which fog models are used to improve accuracy.

In this evaluation, there is no significant difference between the conventional method and the method with additional learning to the pre-training model, even when only the pre-training model is used. This means that the authors' method for combining the fog models has few advantages. Therefore, an approach using a Large Language Model (LLM) to improve

### C. Future Works

Clarify the issues that will be considered in the future. In previous studies, the combined fog model has been assumed. In future studies, we plan to investigate the impact of each combined dimension of the fog model on fog model performance, based on the changes in AS similarity considered in this section.

the performance of the pre-training model itself can be considered. LLMs do not require protocols like node synchronization or a data training phase. Furthermore, the combination of the fog models discussed in this paper is unnecessary. Therefore, it is one of the most effective approaches to using LLM for all nodes without creating a fog models.

## V. Conclusion

In this paper, we investigated the similarity of the fog models as a predictor for the method of selecting fog models to combine. The discussion focuses on the change in similarity after combining the models. It is clear using real data that fog models with accuracy before combining have low similarity. However, there is a risk of over-fitting for each set of data. This makes it difficult to generate a fog model with adequate performance using only fog model similarity as a selection criteria. Therefore, we used a common base learning model pre-trained previously. Each set undergoes additional learning optimizations. The fog models are combined. However, the performance of the combined model is not significantly different from that of the conventional method, as expected.

## References


[1] Tsuchiya, T., Mochizuki, R., Hirose, H., Yamada.T, and Quang Tran Minh, " Research on Selective Combination of Distributed Machine Learning Models ", SN COMPUT. SCI. 3, 438 (2022), Aug. 2022

[2] T.Tsuchiya, R.Mochizuki, H.Hirose, T.Yamada, K.Koyanagi, Q.T.Minh, Tran, " Selective Combination and Management of Distributed Machine Learning Models ", Int'l Conf. on Future Data and Security Engineering (FDSE 2021), Springer Lecture Notes in Computer Science, vol. 13076, pp. 113124, Nov. 2021, Vietnum

[3] TSUCHIYA, T., MOCHIZUKI R., "Study on Management Method for Distributed Machine Learning Models and their Similarity Based Model Combining Method",IEICE Technical Reports, Information Network, July.2024 (witten in Japanese) (to be printed)

[4] Flavio Bonomi, Rodolfo Milito, Preethi Natarajan and Jiang Zhu, N. Bessis and C. Dobre,"Fog Computing: A Platform for Internet of Things and Analytics", Big Data and Internet of Things: A Roadmap for Smart Environments,pp. 169-186,Springer International Publishing,2014

[5] https://dumps.wikimedia.org/

[6] word2vecTomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient estimation of word representations in vector space" CoRR, Vol. abs/1301.3781, 2013

[7] Agirre E, Alfonseca E, Hall K, Kravalova J, Pasca M, Soroa A, "A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches",Proc. of Human Language Technologies: the North American Chapter of the Association for Computational Linguistics, pp. 19–27 2009, Boulder, 2009

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", *CoRR,abs*/1810.04805,2018.Oct.

[9] text8, https://mattmahoney.net/dc/textdata.html(checked on 20th June, 2024)

[10] simlex999, https://fh295.github.io/simlex.html,(checked on 20th June, 2024)

[11] question-words, https://code.google.com/archive/p/word2vec/,(checked on 20th June, 2024)