International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 3139-3149

https://doi.org/10.70517/ijhsa463263

Research on the Application of Speech Recognition Technology in the Protection and Inheritance of Intangible Cultural Heritage of Chinese Languages and Words

Li Chen[1,*]
[1] Department of Chinese Language, Pingdingshan Vocational And Technical College, Pingdingshan, Henan, 467000, China
Corresponding authors: (e-mail: 19937516797@163.com).

Abstract The protection and inheritance of Chinese language and writing intangible cultural heritage is facing the urgent demand of digital transformation, this paper proposes a speech recognition algorithm that integrates meta-learning and transfer learning. The MetaTL algorithm captures the common patterns of cross-language speech features through meta-measure learning network, and combines with the dynamic gradient strategy to realize the fast adaptation of the target language. Relying on performance tests and functional tests, the effectiveness of the proposed algorithm is evaluated. IPA analysis is used to examine the practical application effect of the algorithm. The experiments show that the MetaTL algorithm converges after 111 rounds of training, reaching a correctness rate of 0.9 (90%), and the loss value stabilizes below 0.4 after 10 training sessions. The results of the survey show an overall mean importance value of 3.965 and a mean satisfaction value of 3.757. The interface dimension is located in the first quadrant “Advantageous Enhancement Zone”, the system functionality dimension is located in the second quadrant “Continuation Zone”, and the recommendation intention dimension is located in the third quadrant “Subsequent Opportunity Zone”. The dimension of system functions is located in the second quadrant “continue to maintain”, the dimension of willingness to recommend is located in the third quadrant “follow-up opportunities”, and the dimension of user experience is located in the fourth quadrant “urgent need to improve”. This study provides technical feasibility and practical paths for the digital preservation of linguistic non-heritage, and promotes the in-depth application of AI technology in the field of cultural heritage.

Index Terms linguistic text, meta-learning, transfer learning, speech recognition techniques, MetaTL algorithm

I. Introduction

Language and writing is the core and soul of a country's culture, which carries the cultural inheritance of the nation and the expression of the individual's thoughts, and is also the embodiment of a country's soft power [1], [2]. As a country with a long history, China's language and writing have a long history and unique charm, which not only have an important influence at home, but also receive wide attention internationally [3]-[5]. As China's intangible cultural heritage, the protection and inheritance of language and writing helps to enhance China's cultural self-confidence and national image, and is crucial for maintaining cultural diversity and promoting cultural inheritance [6]-[8].

In order to protect and pass on the language and writing, the Chinese government and all sectors of society have taken a series of measures such as popularizing universal language and writing education, strengthening international exchange and cooperation, and protecting laws, but the effect of these measures is not obvious [9]-[11]. And with the development and application in artificial intelligence, the application of speech recognition technology in the protection and inheritance of Chinese language and writing, an intangible cultural heritage, has received widespread attention [12], [13]. Speech recognition technology, as an important branch in the field of artificial intelligence, in the protection and inheritance of language and writing, it can be transcribed into text through speech recognition of language and writing, thus facilitating the protection and inheritance of future generations [14]-[16]. With the continuous progress of technology, the accuracy and efficiency of speech recognition technology have been significantly improved, which undoubtedly provides a more convenient and efficient means for the protection and inheritance of intangible cultural heritage [17]-[19].

In this paper, we first propose a digital preservation strategy based on the needs of Chinese language intangible cultural heritage preservation. A speech recognition algorithm based on meta-migration learning is designed by decoupling cross-linguistic features through metric learning network and optimizing the damage based on dynamic gradient update strategy. The non-legacy language database is selected for performance testing to explore the algorithm in terms of accuracy and damage optimization. The functionality of different meta-metric learning networks

on the digitization platform is examined to evaluate their end-to-end inference latency and throughput metrics. Reveal user requirement priorities through IPA analysis to confirm system strengths and discover key optimization directions.

II. Design of speech recognition algorithms based on meta-migration learning

II. A.Strategies for Digital Preservation of Linguistic Intangible Cultural Heritage

With the rapid development of science and technology, digital preservation and transmission has become an important means of intangible cultural heritage preservation and transmission. The use of modern scientific and technological means, such as digital recording, virtual reality (VR), artificial intelligence (AI), etc., can realize the comprehensive, accurate and vivid preservation and dissemination of language and cultural practices. Digital recording technology can digitize the text, images, audio, video and other materials of intangible cultural heritage to form a digital resource library for long-term preservation and retrieval. Virtual reality technology can simulate the original scenes and atmosphere of intangible cultural heritage, so that the audience can feel its charm immersively. The application of artificial intelligence technology can play a great role in the recognition, translation and creation of language and writing, and promote the innovation and development of intangible cultural heritage. Digital preservation and inheritance not only solves many problems existing in traditional preservation methods, but also greatly broadens the dissemination channels and audience scope of intangible cultural heritage.

II. B.Metric Learning Networks

II. B. 1) Basic structure

Metric learning algorithms are designed to allow the model to learn how to compare the similarity of samples from a large number of similar but disjoint tasks, and when a new learning task arises, the model can complete the corresponding classification or recognition task by simply comparing the sample features. A metric learning network is usually composed of 2 parts, the embedding module and the metric module, the former is mainly used to extract the deep features of the input data, which can be used as convolutional neural network or recurrent neural network according to the different types of data; and the metric module is mainly used to measure the similarity between the samples of the support set and the samples of the query set in different tasks. Its framework structure is shown in Figure 1.

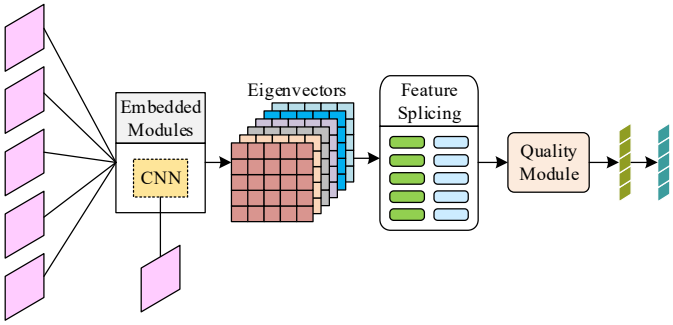

Figure 1: Meta-metric learning network framework structure

The current highly representative metric learning networks mainly include prototype network (P-Net), relational network (R-Net) and its variants.The composition structure of P-Net and R-Net is shown in Fig. 2, where S represents the support set, Q is the query set, and μ is the feature mean of each type of samples, i.e., the prototype of the class, in the support set.

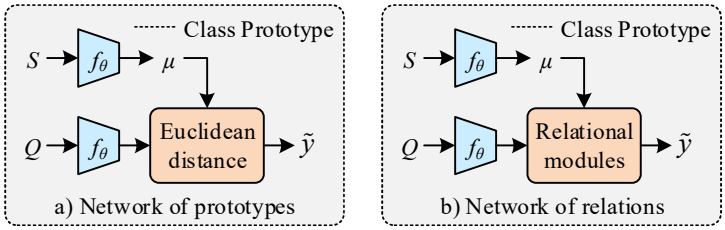

Figure 2: Composition of P-Net and R-Net frameworks

The relational network in Fig. 2 is an improvement of the prototype network, and its framework structure is very similar. f_θ The embedding module of P-Net and R-Net uses the same deep neural network to extract features from the input data, and both of them choose to average the extracted features of the support set samples by classes and represent them as class prototypes. μ The difference between them is that the prototype network chooses a single and fixed Euclidean distance function as the classifier, while the relational network uses a learnable shallow neural network (relational module) to measure the similarity of the samples instead. The difference between the two is that the prototype network chooses a single and fixed Euclidean distance function as the classifier, whereas the relational network instead uses a learnable shallow neural network (relational module) to measure the similarity of the samples.

II. B. 2) Optimization process

In the small sample learning task, the implementation process of different metric learning algorithms is more or less the same, in this paper, we will take the relational network as an example to introduce the optimization process of metric learning network in detail, and its implementation process is described as follows:

(1) First, a batch of tasks from a large number of tasks are randomly sampled and fed into the embedding module to extract the features $f_\theta(x_i)$ and $f_\theta(x_j)$ of the support set samples and query set samples.

(2) Subsequently, the support set sample features are averaged by class to obtain various types of prototypes μ_i, and the computational process is shown in Equation (1).

$$\mu_i = \frac{1}{|S_k|}\sum_{(x_i,y_i)\in S_k} f_\theta(x_i) \tag{1}$$

(3) Then the various types of prototypes are feature spliced with the query set samples respectively to form combined features as shown in equation (2).

$$Z = \mu_i \oplus f_\theta(x_j) \tag{2}$$

(4) Finally, the combined features spliced together are then input to the relationship module R_φ, which calculates the similarity $r_{i,j}$ between the support set and the query set samples through a learnable neural network, as shown in equation (3).

$$r_{i,j} = R_\varphi(Z) = R_\varphi\left(\mu_i + f_\theta(x_j)\right) \tag{3}$$

(5) During the training process of the relational network, the model is optimized by continuously minimizing the mean square error loss (MSE) function, whose optimization objective is shown in equation (4).

$$\theta,\varphi \leftarrow \arg\min \sum_{i=1}^{m}\sum_{j=1}^{n}\left(r_{i,j} - I(y_i = y_j)\right)^2 \tag{4}$$

where I is an indicator function that takes the value of 1 when $y_i = y_j$; 0 otherwise.

II. C.Multilingual speech recognition algorithm based on meta-migration learning

II. C. 1) Description of the algorithm for multilingual speech recognition based on meta-migration learning

Transfer learning approach is to transfer the knowledge of other languages to the learning of the target language, meta-learning approach is to learn a generalized model to be used for all language tasks, this paper proposes meta-transfer based algorithm for multilingual speech recognition.

In the multilingual speech recognition task, the audio sequence input by our model is $X = \{X_1,\ldots X_{n1}\}$, the text label sequence corresponding to the input audio sequence is $Y = \{Y_1,\ldots Y_{n2}\}$, the series of text data generated by the model is $\hat{Y} = \{\hat{Y}_1,\ldots \hat{Y}_{n2}\}$, the overall dataset is defined as $D = (X,Y)$, and the dataset can be divided into $D = \{D_1,\ldots D_k\}$ by task, including training task D_i^{tr}, validation task D_i^{dev}, and fine-tuning task $D_i^{finetune}\ (0 < i \le k)$. Training task D_i^{tr} consists of m source language subtasks $D_{ij}^{src}\ (0 < j \le m)$ and one target language subtask D_i^{tgt}. The overall framework of meta-transfer learning is shown in Figure 3. Taking local dialects as an example, firstly, in each iteration i, we extract D_i^{tr} training tasks from Beijing dialect (BJ), Chongqing dialect (CQ),

Guangzhou dialect (GZ) and other dialects (OT), including source language subtask $D_{ij}^{src}\left(0<j\le m\right)$ and target language subtask D_i^{tgt}. Next, the training data is imported into the speech recognition model, where we use the Transformer model consisting of an encoder module and a decoder module as the speech recognition model. Finally, the meta-transfer learning method changes the gradient update strategy, and then changes the direction of gradient update, so that the final trained model has better performance than the traditional backpropagation model.

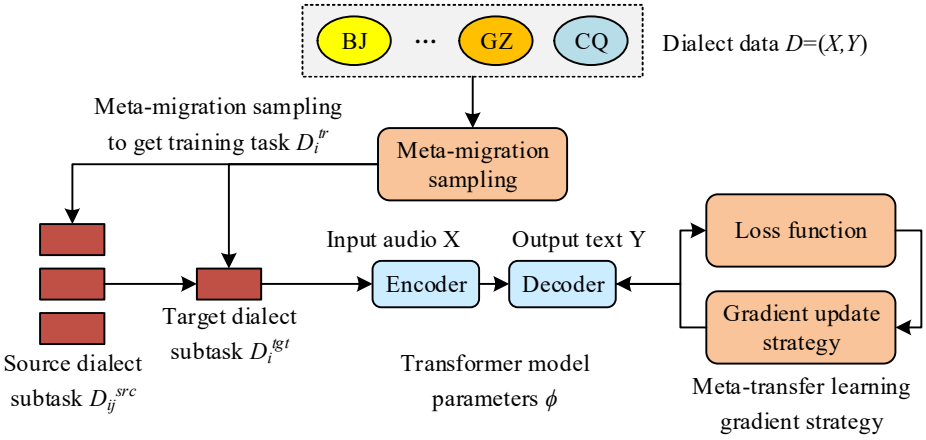

Figure 3: Overall Architecture of meta-transfer learning

The training data contains data in multiple languages.To match our gradient updating strategy, a special data sampling strategy called meta-migration sampling is designed.This method samples a series of source language sub-tasks at each iteration during the training phase D_{ij}^{src} and a target language sub-task D_i^{tgt} .After a period of training, the validation task D_i^{dev} is sampled to verify the effectiveness of the model learning.

End-to-end models using migration learning are prone to overfitting and lead to poor performance in low-resource languages. Therefore, we apply meta-migration learning to Transformer. The training process of meta-migration learning is shown in Figure 4.

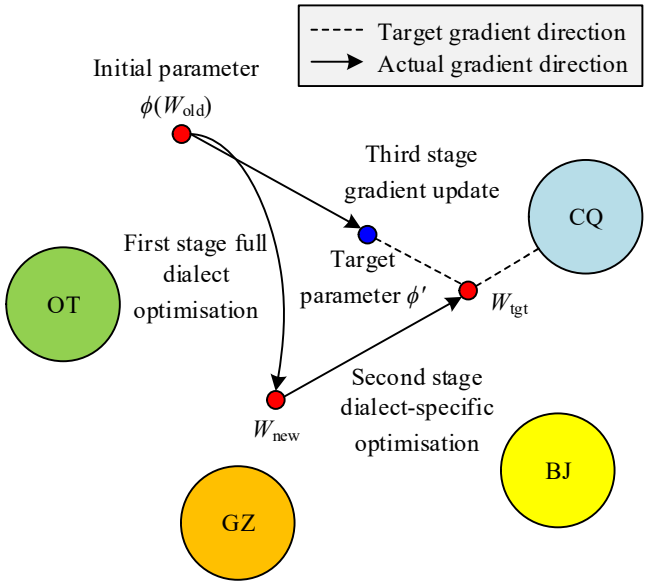

Figure 4: Meta-transfer Learning Training process

Define the weights of the parameters as W whose subscripts indicate the different training stages in which they are located. Define $Adam\left(L_{D_i^{tgt}},W_{new},k\right)$ as the optimization k step using D_i^{tgt} data and loss function L on the

parameters W_{new}. As shown in Eq. (5), in the first stage, full-language optimization is performed by first cloning the initial model parameters ϕ, generating the parameters W_{old} and using the source language subtask D_{ij}^{src} to compute the gradient so that the parameters are updated from W_{old} to W_{new}. As shown in Eq. (6), in the second stage, language-specific optimization is performed by updating the parameters using the target language subtask D_i^{tgt} to obtain W_{tgt} after k steps.

$$W_{new} = Learn(D_{ij}^{src}, W_{old})(0 < j \le m) \tag{5}$$

$$W_{tgt} = Adam(L_{D_i^{tgt}}, W_{new}, k) \tag{6}$$

Second, the goal of meta-migration learning is to derive the entire end-to-end model weights ϕ, which are used as initial weights for subsequent fine-tuning. In the third stage, we use the sum of the first and second stage gradients to update the initial model parameters ϕ with a learning rate of $\grave{o}$. The update formula is given in the following equation:

$$\phi \leftarrow \phi + \grave{o}(W_{tgt} - W_{old}) \tag{7}$$

After performing one round of training updates, we validate the model on validation task D_i^{dev}. After several rounds of training, we obtain the model parameters ϕ^* that can be quickly fitted and achieve the best results on the target language data.

$$\phi^* = Learn(D_i^{dev}; \phi) \tag{8}$$

In the training phase, the initial model parameter vector of the model is defined as ϕ, and the training task D_i^{tr} is extracted in turn for computing the gradient of the model. First, the source language subtask D_{ij}^{src} is used to compute the gradient and update the model parameter W_{old} cloned from the initial model parameter ϕ to obtain the model parameter W_{new}, which enables the model to learn the multilingual task and thus has the ability to learn multiple languages quickly; second, the target language subtask D_i^{tgt} is used to compute the gradient and update it to obtain the model parameter W_{tgt}, and the model's learning ability from the previous stage is applied to the target language subtask to learn the target language domain; finally, we use $(W_{tgt} - W_{old})$ as the final direction for the gradient update of initial model parameter ϕ, which represents the optimal direction for the model to learn the target language, using the Adam optimizer. In the fine-tuning phase, we initialize the fine-tuned model with the parameters trained in the training phase, and take turns to extract the fine-tuning task $D_i^{finetune}$ for calculating the gradient of the fine-tuned model, to obtain the optimal model parameter $\phi_{finetune}$ after the fine-tuning, so that the model can obtain a best result on the corresponding language data.

II. C. 2) Gradient analysis of meta-transfer learning algorithms

The initial parameters of the model are defined as ϕ_0, the parameters of the source language subtask D_i^{src} after training are defined as ϕ_1, and the loss function at this stage is defined as $L_{D_i^{src}}$, and the parameters of the target language subtask D_i^{tgt} after training are defined as ϕ_2, and the loss function at this stage is defined as $L_{D_i^{tgt}}$. The first-order and second-order differential operators are defined as $L'(\phi) = \frac{\partial}{\partial \phi} L(\phi)$, $L''(\phi) = \frac{\partial^2}{\partial \phi^2} L(\phi)$.

$$\phi_1 = \phi_0 - \alpha L'_{D_i^{src}}(\phi_0) \tag{9}$$

$$\phi_2 = \phi_0 - \alpha L'_{D_i^{src}}(\phi_0) - \alpha L'_{D_i^{tgt}}(\phi_1) \tag{10}$$

The equivalent transformation of $L'_{D_i^{tgt}}(\phi_1)$ is shown in Eq. (11). Substituting Eq. (10) and Eq. (11) into Eq. (12), the gradient expression for meta-transfer learning is derived as shown in Eq. (12):

$$\begin{aligned} L'_{D_i^{tgt}}(\phi_1) &= L'_{D_i^{tgt}}(\phi_0) + L''_{D_i^{tgt}}(\phi_0)(\phi_1 - \phi_0) + O(\alpha^2) \\ &= L'_{D_i^{tgt}}(\phi_0) - \alpha L''_{D_i^{tgt}}(\phi_0) L'_{D_i^{src}}(\phi_0) + O(\alpha^2) \end{aligned} \tag{11}$$

$$\begin{aligned} g_{MTL} &= \frac{\phi_0 - \phi_2}{\alpha} = L'_{D_i^{src}}(\phi_0) + L'_{D_i^{tgt}}(\phi_1) \\ &= L'_{D_i^{src}}(\phi_0) + L'_{D_i^{tgt}}(\phi_0) - \alpha L''_{D_i^{tgt}}(\phi_0) L'_{D_i^{src}}(\phi_0) + O(\alpha^2) \end{aligned} \tag{12}$$

We define the expectation of the gradient to be $E_{D,D_i^{src},D_i^{tgt}}[\cdots]$ after a small batch of gradient updates on the entire task D for source language subtask D_i^{src} and target language subtask D_i^{tgt}, respectively, and next define Eq. (13) and Eqs. (3-10).

$$AvgGrad = E_{D,D_i^{src},D_i^{tgt}}\left[L'_{D_i^{src}}(\phi_0) + L'_{D_i^{tgt}}(\phi_0)\right] \tag{13}$$

$AvgGrad$ is the expectation of the gradient when training on all tasks D, and $(-AvgGrad)$ represents the direction of the gradient when performing regular gradient descent on all tasks D.

$$\begin{aligned} AvgGradInner &= E_{D,D_i^{src},D_i^{tgt}}\left[L'_{D_i^{src}}(\phi) L'_{D_i^{tgt}}(\phi)\right] \\ &= \frac{1}{2} E_{D,D_i^{src},D_i^{tgt}}\left[L'_{D_i^{src}}(\phi) L'_{D_i^{tgt}}(\phi) + L'_{D_i^{src}}(\phi) L'_{D_i^{tgt}}(\phi)\right] \\ &= \frac{1}{2} E_{D,D_i^{src},D_i^{tgt}}\left[\frac{\partial}{\partial \phi}(L'_{D_i^{src}}(\phi) L'_{D_i^{tgt}}(\phi))\right] \end{aligned} \tag{14}$$

$(-AvgGradInner)$ represents the direction in which the inner product of gradients increases at different stages, and updating along this direction can improve the generalization ability of the algorithm. Ultimately, the expectation of the meta-migration learning gradient can be expressed as equation (15).

$$E[g_{MTL}] = AvgGrad - \alpha(AvgGradInner) + O(\alpha^2) \tag{15}$$

$AvgGrad$ in the gradient expression minimizes the expected loss for all tasks D, while $AvgGradInner$ increases the inner product between the gradients computed for source language subtask D_i^{src} and target language subtask D_i^{tgt}. An increase in the inner product of the gradients indicates that the stronger the correlation between them, the more the source language subtasks can contribute to the optimization of the target language tasks. Compared with the conventional gradient descent direction $(-AvgGrad)$, the gradient descent direction of the meta-transfer learning algorithm includes $AvgGradInner$ terms, which can be modeled to pay more attention to the correlation between the source and target languages. Therefore, our algorithm can accelerate the loss optimization while optimizing the expected loss, and ultimately obtain fast language learning ability and high language recognition accuracy.

III. Exploring the application of speech recognition based on meta-transfer learning

III. A. Performance testing

The speech dataset used in this study is the Speech Database of Intangible Cultural Heritage of Chinese Languages, which covers endangered languages and minority languages, with a total of 12 non-heritage related language variants. There are a total of 15,000 valid recordings in the dataset (about 1,200 recordings per language), and the duration of a single audio is 3-10 seconds.

The MetaTL algorithm and the meta-learning (Meta) algorithm were subjected to controlled experiments, and the correctness of the two algorithms in 1000 rounds of testing is shown in Fig. 5. The MetaTL algorithm converged after 111 rounds of training and achieved a correctness rate of 0.9 (90%), while the Meta algorithm converged only after 197 rounds of training and achieved an accuracy rate of 0.85 (85%). The gap between the two is evident at the early stage of training, reflecting that MetaTL's meta-learning strategy adapts quickly to non-legacy speech features.

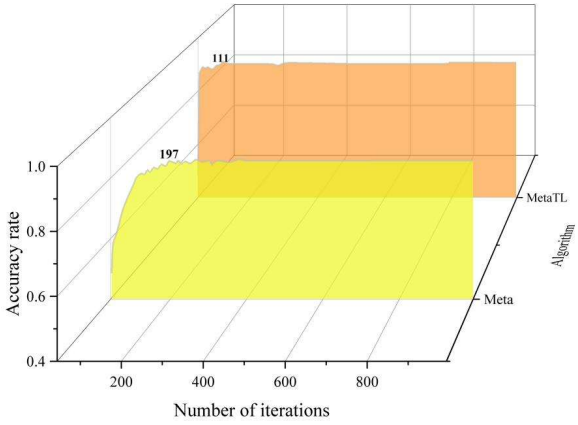

Figure 5: Changes in accuracy rate

The change in algorithm loss is shown in Fig. 6, MetaTL stabilized at a loss value below 0.4 after 10 training rounds, with much less loss and no overfitting occurred in the experiment. While Meta algorithm performs worse and starts overfitting around 250 rounds.

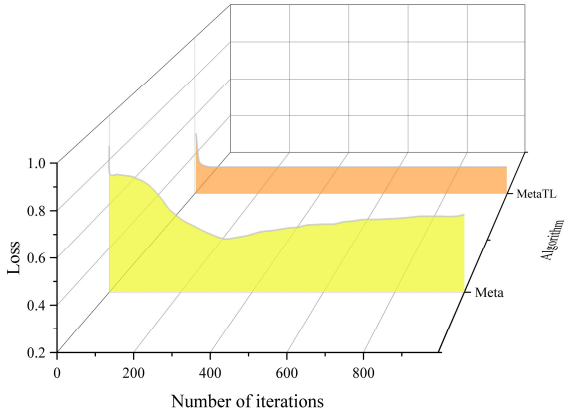

Figure 6: Loss changes

III. B. Functional Testing

In functional testing, the functionality of different metric learning networks on the digitization platform needs to be examined with reference to the ICH information demanded by users. In this study, the optimization of learning network without metrics is selected as the control group, the optimization of prototype network and the optimization of relational network are selected as the experimental group 1 and experimental group 2, and the end-to-end reasoning latency and throughput metrics of the three are compared in the same hardware environment.

When different users access the system at the same time, the response time comparison results of the 3 groups are shown in Fig. 7. According to Fig. 7, it can be seen that the system response time is getting longer and longer as the number of user requests increases. The response time of the system without metric learning network optimization is above 0.5s, and the response time with P-Net system is above 0.35s. When the number of user requests is less than 500, the response time with the R-Net system is within 0.1s, and as the number of user requests continues to increase, the R-Net response time can also be controlled within 0.2s.

Non-legacy digitization platform needs to store massive language and speech data and model parameters for a long period of time, and the limitation of storage resources directly affects the sustainability of the system. This paper analyzes the model compression potential and storage optimization effect of metric learning network. Taking the number of user requests as the independent variable, we test the storage space change of the system when different metric learning networks are selected for optimization, and the comparison results are shown in Figure 8. The results in Fig. 8 show that the storage space of the system is getting larger as the number of user requests increases. The average storage space is lower than 65MB when not optimized with metric learning network. The average storage space is lower than 70MB when optimized with P-Net, while R-Net can increase the average storage space of ICH information to more than 90MB, which meets the user's requirements.

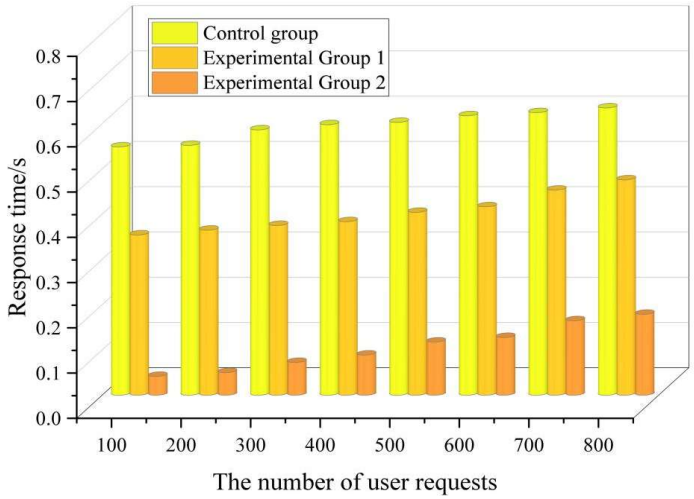

Figure 7: Comparison results of system response time

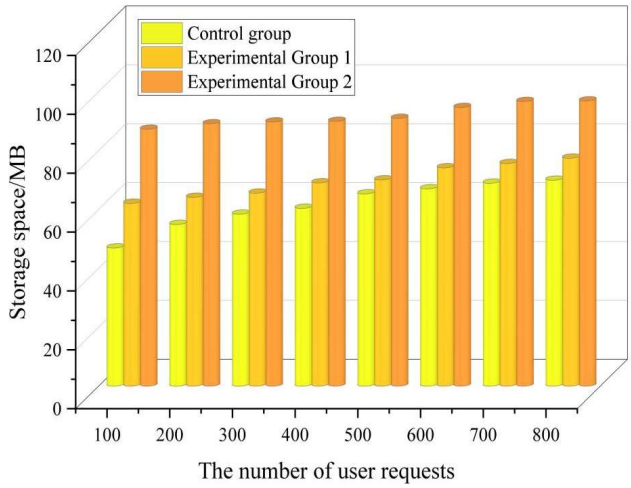

Figure 8: Changes in system storage space

III. C. Analysis of application effects

The IPA method is used to analyze the user's perception of importance and satisfaction, and to construct a coordinate matrix of the perceived elements of the speech recognition system in the protection and inheritance of Chinese language and writing intangible cultural heritage from multiple dimensions and specific indicators. The questionnaire design is based on the system characteristics verified in the previous experiments, combined with linguistic experts' interviews and focus group discussions of non-heritage inheritors, to refine the four core dimensions of the operating interface, system functions, usage experience, and recommendation willingness of users' concerns. The questionnaire adopts a mixed scale design, and through two-dimensional scoring, combined with qualitative feedback, it reveals the matching deviation between users' demand priority and technology landing. The survey targets cover non-genetic inheritors (40%), linguistic researchers (30%), personnel of cultural protection organizations (20%) and ordinary users (10%), ensuring the coverage of both cultural protection subjects and technology users.

III. C. 1) Overall Dimension IPA Quadrant Analysis

Based on the perceptual survey data of the survey respondents, with the mean importance value of 3.965 as the X-axis and the mean satisfaction value of 3.757 as the Y-axis, it can be divided into four quadrants, corresponding to the area of advantageous enhancement, the area of continued maintenance, the area of subsequent opportunities, and the area of urgently needing to be improved, so as to conduct the IPA analysis of the four dimensions of the meta-migratory learning-based speech recognition system.

As shown in Figure 9, the operation interface dimension is located in the first quadrant "Advantage Improvement Zone", the system function dimension is located in the second quadrant "Continuation Maintenance Zone", the

recommendation willingness dimension is located in the third quadrant "Follow-up Opportunity Zone", and the user experience dimension is located in the fourth quadrant "Urgent Need for Improvement Zone". The survey respondents gave a high importance and satisfaction evaluation to the operation interface design of the speech recognition system based on meta-transfer learning, indicating that it can better meet the user's experience needs, but it still needs to be further optimized and innovated, consolidate and give full play to the existing advantages, and continuously improve user satisfaction. The importance of the functional dimension of the system is relatively low, but the satisfaction evaluation is high, indicating that the function of the designed speech recognition system is generally good, and the existing level should be maintained in the future, and moderate improvement and promotion should be carried out. Users have a high evaluation of the importance of the two dimensions of recommendation willingness and user experience of the proposed speech recognition system, but the current satisfaction score is low, and these two aspects are the key areas that need to be improved urgently.

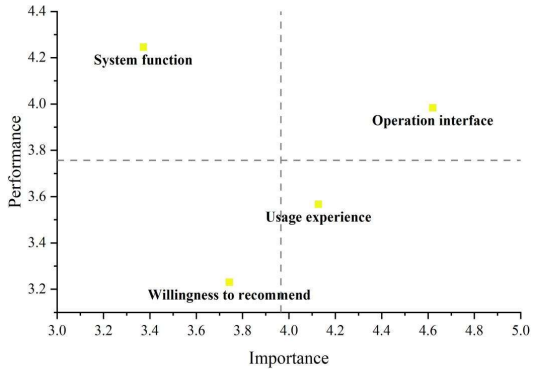

Figure 9: Analysis results of IPA in the overall dimension

III. C. 2) Segmentation Indicator IPA Quadrant Analysis

Further, the importance and satisfaction scores for each segment of the dimension that is in the second, third, and fourth quadrants can be calculated, and using the mean values as the axes, it can be determined where the IPA quadrants are for each type of segment within the different dimensions.

The results of the IPA analysis of the system function dimensions are shown in Figure 10. Among the system function dimensions, the accuracy of tone recognition, the ability of multi-language expansion, and the perception of storage resource consumption are located in the first quadrant, and the survey respondents give high importance and satisfaction to the system's accuracy of tone recognition, ability of multi-language expansion, and the consumption of storage resources, and think that the performance of these aspects is relatively outstanding at present, but still needs to be optimized and improved to consolidate the existing development advantages. However, further optimization and enhancement are still needed to consolidate the existing development advantages. System Adaptability is located in Quadrant 2, where respondents rated the robustness of different language accents as high, but recognized its importance as relatively low. Cultural identity retention is located in the third quadrant, while error feedback mechanisms are located in the fourth quadrant. The recognition coverage of the system's linguistic and cultural loaded words should be further improved, and the language recognition error feedback mechanism should be refined.

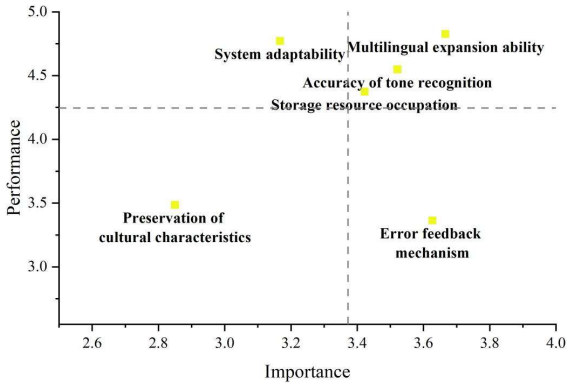

Figure 10: Results of IPA analysis of the system functional dimension

The results of the IPA analysis of the recommendation willingness dimension are shown in Fig. 11. Cultural communication utility and technological substitutability are located in the first quadrant, i.e., the satisfaction and importance scores are high, and the survey respondents generally agree with the extent of the contribution of the language recognition system to the living heritage of endangered languages, and many users are willing to take the initiative to recommend this paper's system to their friends and relatives. In addition, privacy and data security are located in the second quadrant, i.e., visitors are generally satisfied with the compliance of localized storage of speech data, but the importance rating of this indicator is low. Cost-effectiveness is located in the third quadrant, i.e., tourists' importance and satisfaction ratings are relatively low, suggesting that the current system's balance between hardware deployment costs and performance benefits still needs to be strengthened, which can be used as an opportunity point for subsequent development. Social impact is located in the fourth quadrant, survey respondents believe that the public perception of the digital conservation results of the enhancement is still insufficient, the future will be further through the technology to influence people's perceptions, and better stimulate the enthusiasm of users to participate in the enthusiasm and dissemination of motivation.

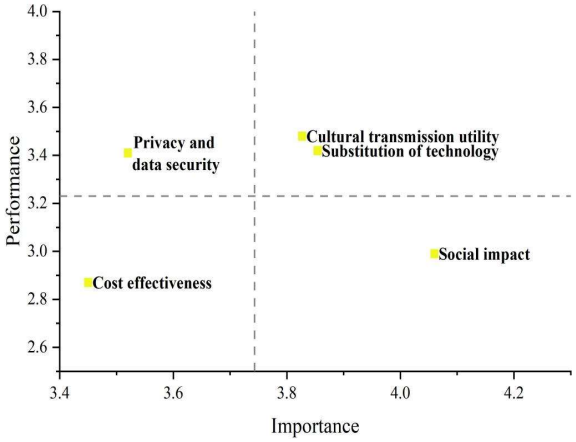

Figure 11: IPA analysis results of recommendation intention dimension

The results of the IPA analysis of the usage experience dimensions are shown in Figure 12. Real-time feedback efficiency and system stability are located in the first quadrant, indicating that the system performance can basically meet the experience needs of different users. Personalized adaptation is located in the second quadrant, i.e., the importance of the two is lower, but the performance is more satisfactory. In addition, survey respondents place relatively low importance and satisfaction ratings on multimodal interaction adaptability, placing it in the “follow-up opportunity zone”. Finally, error tolerance and error correction are in the fourth quadrant, where respondents generally believe that the interaction cost of manual correction is too high, and are in the “urgent need for improvement zone”.

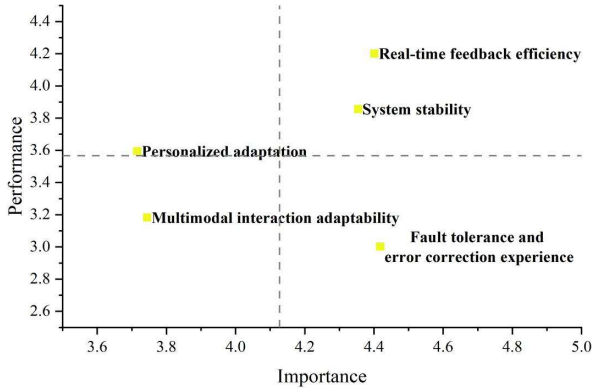

Figure 12: Results of IPA analysis using the experience dimension

IV. Conclusion

In this paper, we design a meta-transfer learning speech recognition algorithm, design performance tests and functional tests to explore its effectiveness, and analyze the effect of practical applications through questionnaires.

The MetaTL algorithm converged after 111 rounds of training and reached a correctness of 0.9 (90%), while the Meta algorithm converged only after 197 rounds of training and reached an accuracy of 0.85 (85%).The MetaTL stabilized at a value of less than 0.4 after 10 rounds of training, with much smaller losses, and overfitting did not occur in the experiment. The Meta algorithm, on the other hand, performed worse and started overfitting at 250 rounds. During functional testing, the system response time gets longer as the number of user requests increases. The system response time is above 0.5s without Meta Metric Learning Network Optimization and above 0.35s with P-Net system. When the number of user requests is less than 500, the response time of the system with R-Net is within 0.1s, and as the number of user requests continues to increase, the R-Net response time can also be controlled within 0.2s. The average storage space is lower than 65MB when metric learning network is not used for optimization. The average storage space is lower than 70MB when P-Net is used for optimization, and R-Net can increase the average storage space of intangible cultural heritage information to more than 90MB, which meets the user's requirements.

The survey results showed that the overall mean importance was 3.965 and the average satisfaction was 3.757. The operation interface dimension is located in the first quadrant "Advantage Improvement Zone", the system function dimension is located in the second quadrant "Continuation Maintenance Zone", the recommendation willingness dimension is located in the third quadrant "Follow-up Opportunity Zone", and the user experience dimension is located in the fourth quadrant "Urgent Improvement Zone". In terms of experience, the respondents' attention to multimodal interaction adaptability and satisfaction evaluation were relatively low, and they were in the "follow-up opportunity zone". Survey respondents generally believe that the interaction cost of manual correction by users is too high, and the fault tolerance and error correction experience is in the "urgent need for improvement".

References

[1] Fan, H. (2021). Exploring the Value of Chinese Language and Literature Research in Cultural Inheritance. Frontiers in Educational Research, 4(16).

[2] Yang, R., & Wang, W. S. Y. (2018). Categorical perception of Chinese characters by simplified and traditional Chinese readers. Reading and Writing, 31, 1133-1154.

[3] Xiao, W., & Treiman, R. (2012). Iconicity of simple Chinese characters. Behavior research methods, 44, 954-960.

[4] Kuo, L. J., Li, Y., Sadoski, M., & Kim, T. J. (2014). Acquisition of Chinese characters: The effects of character properties and individual differences among learners. Contemporary Educational Psychology, 39(4), 287-300.

[5] Chang, Y. N., Hsu, C. H., Tsai, J. L., Chen, C. L., & Lee, C. Y. (2016). A psycholinguistic database for traditional Chinese character naming. Behavior Research Methods, 48, 112-122.

[6] Xu, J., Zhou, C., & Liu, H. (2024). Cultural heritage as a key motivation for sustainable language protection: a case study of the Suzhou dialect protection project. Journal of Multilingual and Multicultural Development, 1-18.

[7] Yang, R. (2020). Research on the Protection and Inheritance of Samei Language. Open Journal of Social Sciences, 8(08), 285.

[8] Wang, H. (2020, February). Inheritance and development of Chinese dialect culture. In 6th International Conference on Education, Language, Art and Inter-cultural Communication (ICELAIC 2019) (pp. 924-926). Atlantis Press.

[9] Gelman, S. A., & Roberts, S. O. (2017). How language shapes the cultural inheritance of categories. Proceedings of the National Academy of Sciences, 114(30), 7900-7907.

[10] ZHANG, Y. (2023). The Function of Inheriting Traditional Culture in Chinese Language and Literature. DEStech Transactions on Social Science, Education and Human Science, (isss).

[11] Cao, W. (2024). Chinese Characters and Cultural Inheritance: Exploring the Important Role of Chinese Characters in the Inheritance of Chinese Culture. Cultura: International Journal of Philosophy of Culture and Axiology, 21(1).

[12] Zhao, M., & Cui, J. (2022, December). Artificial Intelligence Boosting Protection and Communication of Intangible Cultural Heritage. In 2022 International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID 2022) (pp. 1126-1133). Atlantis Press.

[13] Giannini, E., & Makri, E. (2023, March). Cultural Heritage Protection and Artificial Intelligence; The Future of Our Historical Past. In International Conference on Transdisciplinary Multispectral Modeling and Cooperation for the Preservation of Cultural Heritage (pp. 375-400). Cham: Springer Nature Switzerland.

[14] Yu, Y. (2012, March). Research on speech recognition technology and its application. In 2012 international conference on computer science and electronics engineering (Vol. 1, pp. 306-309). IEEE.

[15] Basak, S., Agrawal, H., Jena, S., Gite, S., Bachute, M., Pradhan, B., & Assiri, M. (2023). Challenges and limitations in speech recognition technology: a critical review of speech signal processing algorithms, tools and systems. CMES-Computer Modeling in Engineering and Sciences.

[16] Shadiev, R., & Liu, J. (2023). Review of research on applications of speech recognition technology to assist language learning. ReCALL, 35(1), 74-88.

[17] Wen, H. (2023). Research on Intelligent Experience and Living Inheritance of Cultural Heritage Against the Background of Artificial Intelligence. Lecture Notes in Education Psychology and Public Media, 10, 78-90.

[18] Trichopoulos, G. (2023, September). Large language models for cultural heritage. In Proceedings of the 2nd International Conference of the ACM Greek SIGCHI Chapter (pp. 1-5).

[19] Kadyanan, I. G. A. G. A., Er, N. A. S., Karyawati, A. A. I. N. E., Putra, I. G. N. A. W., Gunawan, I. M. S., Budiantari, N. M. J., & Octavia, H. C. (2025). Balinese text-to-speech dataset as digital cultural heritage. Data in Brief, 111528.