Journal of Research Administration
Society of Research Administrators International

AN ELDERLY CARE SERVICE DEMAND ANALYSIS MODEL BASED ON MLP-SELF-ATTENTION

Shuang Yang[1,2]
[1]Graduate School of Management, Post Graduate Centre, Management and Science University, University Drive, Off Persiaran Olahraga, Section 13, 40100 Shah Alam, Selangor, Malaysia
[2]School of Foreign Languages and Tourism, Wuxi Institute of Technology, 1600 Gaolang Western Road, Wuxi, Jiangsu 214121, China

Wan Nur Tasnim Wan Hussin[3]*
[3]Faculty of Business Management and Professional Studies, Management and Science University, University Drive, Off Persiaran Olahraga, Section 13, 40100 Shah Alam, Selangor, Malaysia, *Corresponding Author: Wan Nur Tasnim Wan Hussin

Abstract

Purpose: With the intensification of the aging population trend, the demand for elderly care services is steadily rising. Therefore, accurately understanding the actual needs of the elderly is a prerequisite for formulating effective elderly care policies and providing high-quality services. Through demand analysis, specific information about the elderly's health status, lifestyle, and mental state can be gathered, providing a basis for personalized services. Additionally, demand analysis can assist elderly care institutions or governments in planning suitable service offerings, enhancing service quality, and meeting the diverse needs of the elderly population. Moreover, timely understanding and addressing the actual needs of the elderly can improve their quality of life and sense of happiness, promoting social harmony and stability.

Theoretical framework: Demand analysis for elderly care services is a crucial component in the construction of the elderly care service system, holding significant importance for ensuring the well-being of the elderly and the sustainable development of society. This study proposes a multi-label model MAC for elderly care service demand analysis, incorporating a Multi-Layer Perceptron (MLP), Self-Attention mechanism, and Cosine distance.

Methods: Using survey results and feedback from individuals aged 65 and above in Jiangsu Province as the dataset, MAC is employed to derive demand analysis results for elderly care services.

Findings: The model utilizes the MLP as the foundation for linear fitting, effectively reducing individual differences between data. It employs the cosine value of the angle between two vectors in vector space as a metric for measuring the magnitude of differences between individuals, thereby enhancing the classification performance of elderly care services.

Practical implications: The model incorporates a Self-Attention mechanism to strengthen the fitting of the importance of all factors in the survey questionnaire to the final result. Experimental results demonstrate that the MAC model can provide more rational planning suggestions for elderly care service providers. As a fusion algorithm utilized in elderly care service analysis, MAC exhibits commendable performance.

Originality/Value The originality lies in its emphasis on the main approaches, analysis methods, and recently used models in the study of aging population trend, the demand for elderly care services is steadily rising.
Keywords: Elderly Care Service Analysis, MLP, Attention, Cosine, Multi-Label.

1. Introduction

Analysis of elderly care service needs aims to understand and predict the requirements of elderly individuals and the elderly population as a whole. This allows governments, social institutions, businesses, and individuals to better meet their needs. This analysis holds significant importance at the societal, economic, and policy levels. Firstly, by analyzing the factors influencing elderly care service needs, governments can gain a better understanding of the requirements of the elderly, enabling them to formulate more targeted social policies. This includes policies related to healthcare, long-term care, retirement benefits, and more. Governments can adjust resource allocation based on demand predictions to meet the needs of the elderly and enhance societal well-being. Secondly, based on the analysis of demand factors, social institutions and elderly care service providers can more effectively plan and provide services. They can determine which areas require more elderly care facilities, medical resources, social support, and more to meet the needs of the elderly. Furthermore, analyzing factors influencing elderly care service needs can help businesses identify commercial opportunities. They can develop new products and services to meet the needs of the elderly, such as smart health monitoring devices and age-friendly home renovations. With the increasing aging population, the elderly care service industry presents a potential employment opportunity, particularly in the fields of nursing, health management, and social services. Additionally, analyzing factors influencing elderly care service needs contributes to promoting social integration. Understanding the needs of the elderly can assist social organizations and volunteer groups in providing more social support and participation opportunities, reducing the elderly's sense of social isolation. Finally, individuals and families can use the analysis of demand factors to plan their finances. This allows for a more convenient understanding of needs in areas such as pensions, medical expenses, and long-term care costs, enabling wiser financial decision-making. Through the analysis of elderly care service needs, a deeper understanding of the structural factors influencing these needs and the importance of impression factors can be gained.

Currently, many deep learning algorithms have been applied in various fields. For instance, Ruma et al. developed a socialized healthcare service recommendation using deep learning, which predicts the ratings of healthcare services based on trust relationships among users who have social connections with active users[1]. Additionally, Meng et al. built a doctor recommendation system on healthcare consultation platforms, integrating knowledge graphs and deep learning to address research gaps in this area[2]. This study proposed an advanced doctor recommendation method utilizing health knowledge graphs to overcome data sparsity issues and generate accurate and interpretable recommendations using deep learning techniques. The widespread use and rapid development of deep learning have not only propelled industrial advancement but also fostered progress in the humanities. Tseng et al. applied deep learning and other technologies to extract critical information in elderly care, establishing an alert mechanism. This process aims to not only safeguard the physical well-being of the elderly but also attend to their mental health, establishing a model of well-being in intelligent community care[3]. Fahim et al. created a digital environment that replicates the homes of the elderly to inconspicuously monitor their daily activities. They successfully addressed learning challenges posed by different activities at home by introducing a deep meta-sequence model[4]. This method groups activities into meta-classes based on their nature, enabling abstract learning of feature spaces. This research provides a robust framework for digitally monitoring the behavior of the elderly, aiding healthcare professionals in understanding the level of support needed for their daily tasks and potential risks of emergencies at home. Pandia et al. developed an automated physiological signal monitoring system, leveraging emerging Internet of Things (IoT) technology to advance electronic healthcare systems[5]. They employed accurate signal prediction algorithms based on deep neural networks and achieved intelligent data collection using smart sensors and National Instrument myRIO. They also designed the Smart-Monitor sensor product. Through experiments, they verified that the system could offer reliable assistance and accurately predict signals, achieving an average accuracy of 97.2%. This demonstrates the reliability and accuracy of their automated system. Ghosh et al. proposed a comprehensive intelligent home healthcare system (FEEL) for the elderly, addressing major challenges in medical IoT systems, including sparse labeled data and diverse user needs[6]. The system introduced a novel federated learning framework for analyzing and recommending health data and contextual information based on small sample learning. It also modeled the impact of health parameters and recommendations on the environment using user and context-based knowledge graphs, conducted user activity monitoring and location estimation using deep learning architecture, and presented an edge-fog-IoT collaboration framework to collect, store, and share medical recommendations while safeguarding user privacy. Valentina et al. introduced an innovative Human Activity Recognition (HAR) system that harnesses the potential of wearable devices and deep learning to accurately identify the most common activities in the daily lives of family members. The designed wearable sensor embeds Inertial Measurement Units (IMUs) and a Wi-Fi component, capable of sending data to cloud services. Users can also directly connect to the internet through a regular home router for convenient installation and management. This sensor is coupled with a Convolutional Neural Network (CNN) to perform inference with minimal resources, maintaining flexibility and feasibility on low-cost or embedded devices. The system is specifically designed for daily activity monitoring and accurately identifies nine different activities, achieving a high accuracy rate of 97%[7].

Deep learning has achieved remarkable results in the application of elderly care services. It plays a crucial role in monitoring the activities of the elderly, assessing their health status, and diagnosing diseases. Through smart sensors and data analysis, real-time monitoring and personalized care for the elderly have been realized, thereby enhancing the quality and efficiency of elderly care services. However, there are also some pressing issues that need to be addressed. One of the most prominent concerns is data privacy and security. It is imperative to establish robust data protection mechanisms to safeguard the personal information of the elderly. In addition, the lack of interpretability of deep learning models is also a challenge, especially in critical areas like medical diagnosis. There is a need to further enhance the understandability of the models so that healthcare professionals can trust and effectively utilize these technological advancements. To address these challenges, this paper proposes the application of the MAC model in the analysis of elderly care service demands. Using elderly individuals aged 65 and above in Jiangsu Province as the research subjects, a dataset is constructed by collecting survey results and feedback. The MAC model is employed as the demand analysis model to derive the results of elderly care service demands. This model utilizes a multi-layer perceptron as the basic linear fitting model, which effectively reduces individual differences between data points. It uses the cosine value of the angle between two vectors in vector space as a measure of the magnitude of differences between individuals, thereby enhancing the classification performance of elderly care services. Additionally, the model incorporates a self-attention mechanism, which strengthens the fitting of the importance of all factors in the questionnaire to the final results. Experimental results demonstrate that the MAC model can provide more rational planning suggestions for elderly care service providers.

2. Related Information

2.*1 Multilayer Perceptron MLP*

MLP is a feedforward neural network model used to address tasks such as classification, regression, and other supervised and unsupervised learning tasks[8]. As shown in Fig. 1, MLP consists of multiple layers, including the input layer, multiple hidden layers, and the output layer. During forward propagation, input data propagates from the input layer to the hidden layers, and after passing through multiple hidden layers, it propagates to the output layer. In each layer, the input undergoes a transformation through a combination of weights and activation functions and is then passed to the next layer. There are many activation functions available, such as the Relu function, Sigmoid function, and Tanh function, all of which have a characteristic feature: they can facilitate easy backpropagation[9]. Backpropagation is accomplished by calculating the gradients of the model parameters (weights and biases) with respect to the loss function, and the entire model uses gradient descent or its variants to update the model parameters in order to minimize the loss function.

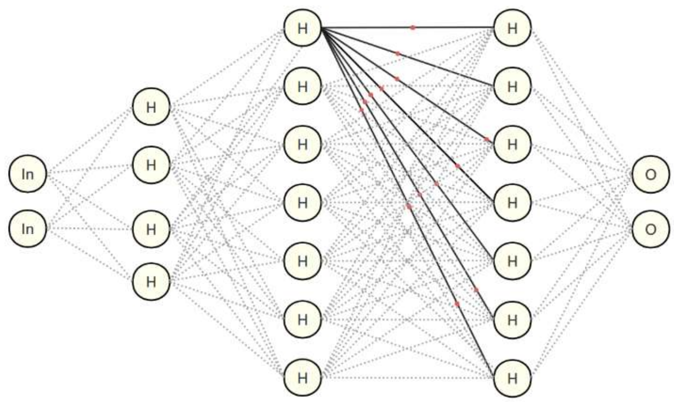

Fig. 1 Schematic diagram of MLP structure.

Source: Authors

The MLP is composed of multiple hidden layers, with each hidden layer containing numerous neurons, enabling it to learn and represent complex nonlinear relationships[10]. The network structure of MLP is also highly flexible, allowing adaptation to different types of data and tasks by adjusting the number and size of hidden layers, as well as the activation functions. Moreover, during both the forward propagation and backpropagation processes of the MLP, computations for each neuron can be carried out in parallel, resulting in rapid training speeds on graphics processing units (GPUs) and superior fitting performance compared to traditional machine learning algorithms.

2.*2 Commonly used distance functions*

In deep learning, distance functions are a method used to measure the similarity or dissimilarity between data points. These functions are commonly employed in various tasks such as clustering, dimensionality reduction, and classification. As shown in Table 1, this paper selects some commonly used distance functions and similarity metrics.

For Euclidean distance, it is one of the most commonly used distance metrics[11]. It has a straightforward and intuitive calculation method, making it suitable for continuous feature data spaces, such as images and numerical data. However, in high-dimensional spaces, computing Euclidean distance may lead to the "curse of dimensionality," and it may not be suitable for sparse data as it tends to overemphasize missing features. Additionally, Euclidean distance does not take into account directional information between features, which may not be ideal for certain tasks like text classification.

Manhattan distance is similar to calculating the walking distance between two points in a city, and like Euclidean distance, it is easy to understand and calculate[12]. However, unlike Euclidean distance, Manhattan distance is less sensitive to outliers. It only considers displacement along the coordinate axes and does not account for diagonal paths, which may lead to inaccurate distance measurements in certain cases.

In contrast to Euclidean distance, which may struggle in high-dimensional sparse domains, cosine similarity is not affected by the "curse of dimensionality" in high-dimensional spaces, making it highly effective for high-dimensional sparse data like text data[13]. Cosine similarity also considers the directional information between feature vectors, allowing it to better capture similarity in certain tasks. However, cosine similarity only takes into account the direction of feature vectors, without considering the distance information between them, which may make it less accurate than Euclidean or Manhattan distance in certain situations.

Table 1 Introduction to commonly used distance functions

Distance function	**Expression**
Euclidean distance	$d_{Eld} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$
manhattan distance	$d_{Mhd} = \sum_{k=1}^{n}\lvert x_{1k} - x_{2k}\rvert$
cosine distance	$d_{cos} = \frac{\sum_{k=1}^{n} x_{1k}x_{2k}}{\sqrt{\sum_{k=1}^{n} {x_{1k}}^2}\sqrt{\sum_{k=1}^{n} {x_{2k}}^2}}$

Source: Authors

As shown in Fig. 2, cosine similarity measures the similarity between two vectors[14]. A value closer to 1 indicates higher similarity, a value closer to -1 indicates higher dissimilarity, and a value close to 0 suggests relative neutrality. Cosine similarity takes into account not only the magnitude (length) of the vectors but also their directions[15]. In the demand analysis of this paper, the degree of demand from the elderly for different elderly care services can be represented as a vector, where each dimension corresponds to a factor or feature. Cosine similarity is employed to compare the similarity between demands.

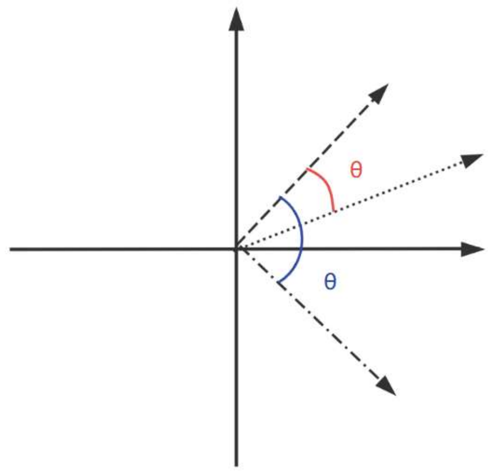

Fig. 2 Cosine distance function

Source: Authors

2.3 Commonly used activation functions

Activation functions in deep learning serve the purpose of providing non-linear transformations to the data, enabling neural networks to become more complex and possess stronger representational capabilities. Commonly used activation functions include the sigmoid function, hyperbolic tangent (Tanh) function, Rectified Linear Unit (ReLU), and Leaky ReLU, as described in Table 2. The sigmoid function maps inputs to the range of (0,1), making it suitable for binary classification tasks where the output can be interpreted as probabilities. In some cases, sigmoid can mitigate the issue of gradient explosion, though its effectiveness is typically modest. However, sigmoid suffers from the problem of vanishing gradients[16], especially in deep networks, meaning that during backpropagation, gradients may become very small, leading to slow training. Sigmoid outputs are not zero-centered, which can lead to issues like neuron deadness. In contrast to the sigmoid function, the Tanh function is more zero-centered and is better suited for binary classification and regression problems. Similarly, ReLU and Leaky ReLU functions help alleviate the problem of vanishing gradients, as they have a gradient of 1 or x in the positive range, preventing significant compression of gradients. Leaky ReLU addresses the issue of "dead neurons" in the ReLU function by introducing a small negative slope. However, the parameter (negative slope) of Leaky ReLU needs to be adjusted, which is a hyperparameter that requires manual tuning.

Table 2 Introduction to commonly used activation functions

Activation function	**Equation**	**Range**

Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$	$(0,1)$
Tanh	$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$(-1,1)$
ReLu	$f(x) = max\{0, x\} = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases}$	$(0, +\infty)$
Leaky ReLu	$f(x) = \begin{cases} x & x \leq 0 \\ kx & x > 0 \end{cases}$	$(-\infty, +\infty)$

Source: Authors

2.4Self-Attention mechanism

The core idea of the Self-Attention mechanism is that the model can allocate different weights at different time steps or positions in order to focus on the most relevant parts when processing input data. This can be seen as a kind of weighted pooling operation[17], where the weights represent the importance of different positions or features. As shown in Fig. 3, the neural network receives many vectors of varying sizes, and there are certain relationships between different vectors. However, during actual training, it is challenging to fully exploit these relationships among inputs, leading to poor model training results. To address the issue of conventional fully connected neural networks struggling to establish correlations among multiple related inputs, the Self-Attention mechanism is introduced. In practice, the Self-Attention mechanism aims to make the machine pay attention to the correlations between different parts of the entire input.

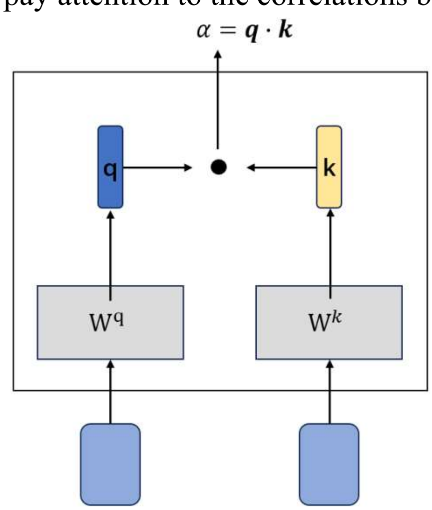

Fig. 3 Self-Attention correlation degree
Source: Authors

The Self-Attention mechanism allows the model to capture global dependencies when processing discrete sequential data, not just local information. Additionally, the Self-Attention mechanism can easily handle sequences of variable lengths[18]. This is because its computations dynamically adjust based on the length of the input sequence, without the need for padding or truncating the sequence. The Self-Attention mechanism automatically learns to extract features from discrete data, eliminating the need for manual feature engineering. This reduces the burden of feature engineering and makes the model more flexible, suitable for various types of discrete data.

3 Elderly care service demand analysis model based on MAC model.

3.1 Elderly care service demand analysis process

The process of elderly care service demand analysis based on the MAC model is illustrated in the diagram. First, data is read from the CSV file and undergoes preprocessing. Next, the data is augmented and fed into the input module of the model. The processed data is then passed to the fully connected layers for label classification, ultimately yielding the categorized results of elderly care service demands. While an MLP alone can function as an end-to-end classifier model, requiring only the use of the Softmax algorithm for classification decisions at the output layer, this study employs a composite model for data classification and prediction. The Self-Attention mechanism allows the model to consider information from different positions in the sequence when processing sequential data, rather than being confined to a fixed window or convolutional kernel. This enables the model to better capture contextual information and understand the relationships between different positions in the input data. Additionally, the Self-Attention mechanism provides a way to visualize the model's decision-making process, as it can indicate which parts of the input it focused on when generating the output[19]. In subsequent models, this will be further analyzed by examining the specific points of focus provided by the Self-Attention model's sub-modules. The analysis process of elderly care service demand based on MAC model is shown in Fig. 4:

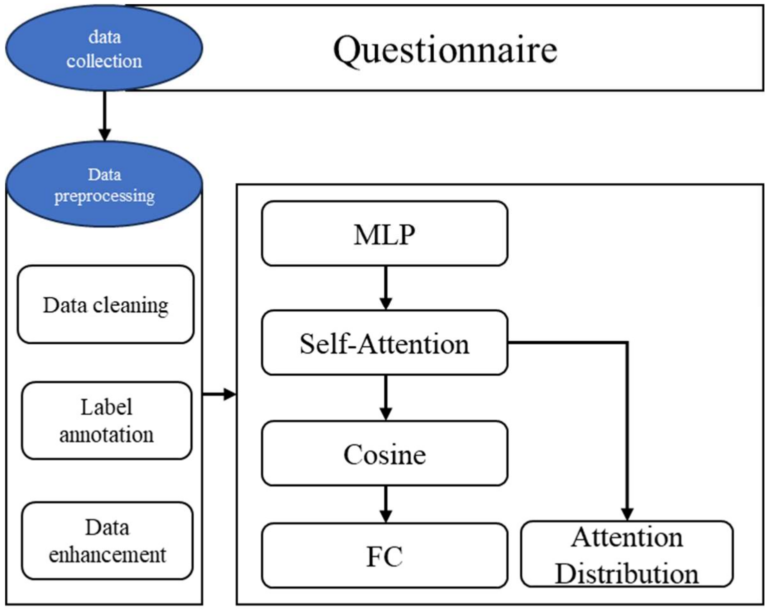

Fig. 4 Elderly care service demand analysis process based on MAC model.

Source: Authors

In the task of elderly care service demand analysis, the collected survey questionnaires are formatted, and then divided into three parts in proportion: the training set, validation set, and test set[20]. Typically, during the training process, the model is used to make predictions on the training set. The predicted labels are then compared with the actual labels to calculate the loss, which is used to update the model parameters. Additionally, during training, the validation set is employed to validate the accuracy of the model at different stages. The model that performs best on the validation set is selected and retained for further testing. Finally, the model's generalization ability and robustness are evaluated on the test set. The specific training process is illustrated in Fig. 5.

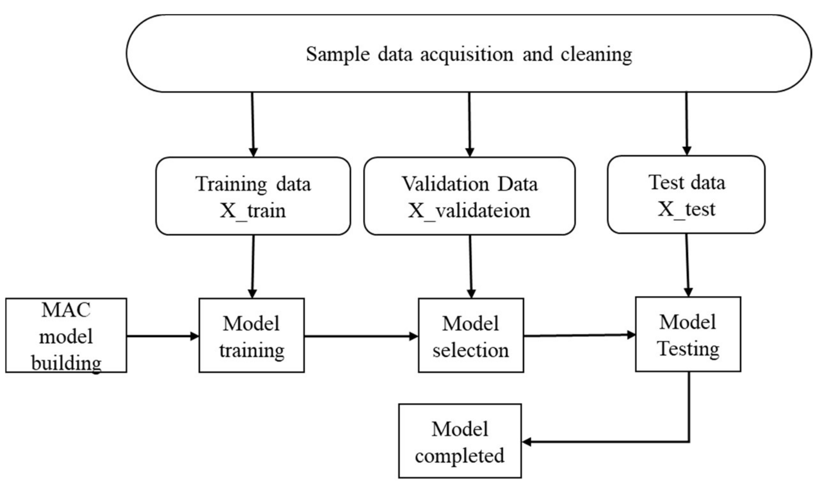

Fig. 5 Training process of MAC model

Source: Authors

3.2 MAC model pseudocode analysis

The MAC model used in this study primarily consists of three modules: the MLP module, the Self-Attention mechanism module, and the Cosine Distance module. As shown in the code snippet below, the pre-training data is first loaded into a sequence X. This data will undergo data augmentation, which will be explained in the following text on how it transforms the original data. Next, an MLP is created, and the X sequence data is loaded into this module to generate preliminary fitted data. Then, a Self-Attention model is created to compute the attention distribution for the output data from the MLP. Two fully connected layers are created, and the data is finally fitted, and the attention distribution results are returned. Finally, a cosine distance formula and a fully connected classification layer are created to estimate labels for the generated data. During training, the estimated labels are compared with the true labels to calculate the loss, and this loss value is backpropagated to the model to update it.

Algorithm: MAC algorithm of the proposed method.

Input: **[x]** sequence of factors that affect the demand for pension services

1. Ioad pretrained params from Set-[X]
2. **Init Model:**
3. Create a Dense MLP for inputs.
4. Build a **Self-Attention** to calculate the **attention distribution** $\boldsymbol{\alpha_p}$ for each element in the input.And return attention distribution $\boldsymbol{\alpha_p}$
5. Build A fully connected layer A with input $\boldsymbol{\alpha_p}$ and output $\mathbf{E}_{\boldsymbol{a}}$
6. Build A fully connected layer B with input $\mathbf{E}_{\boldsymbol{a}}$ and output $\mathbf{E}_{\boldsymbol{b}}$
7. Build **Cosine** function to calculate $\boldsymbol{\theta_j}$
8. Build A fully connected layer Out with input $\boldsymbol{\theta_j}$ and output $\boldsymbol{\hat{y}_j}$
9. **While** *not converged* **do:**
10. Clculate **loss for** $\boldsymbol{\hat{y}_j}$ **and** $\boldsymbol{y_j}$
11. Using gradient descent algorithm to backpropagated the **loss**

Source: Authors

3.3 Self-Attention mechanism visualization

The Self-Attention mechanism can, to some extent, enhance the interpretability of results. Here, we achieve this by visualizing and analyzing the attention distribution $\boldsymbol{\alpha_p}$from the previous section. As shown in the pseudocode below, first, the attention distribution $\boldsymbol{\alpha_p}$from the previous section is averaged to obtain the attention mean $\boldsymbol{\alpha_v}$. Then, by summing $\boldsymbol{\alpha_v}$, we obtain $\boldsymbol{\alpha_s}$as the denominator for normalizing the attention distribution. Finally, the attention distribution $\boldsymbol{\alpha_p}$is used as the numerator to obtain the final attention distribution matrix D_norm. Normalizing the attention distribution matrix is mainly due to the fact that different features often have different units or dimensions, and their ranges of variation are also on different orders of magnitude. Without normalization, certain indicators may be overlooked, potentially affecting the results of data analysis.

Algorithm: Self-Attention Mechanism Visualization algorithm.

Input: [x] attention distribution $\boldsymbol{\alpha_p}$

1. Clculate the average $\boldsymbol{\alpha_v}$ of the matrix which axis=0
2. Clculate the sum $\boldsymbol{\alpha_s}$ of the matrix $\boldsymbol{\alpha_v}$ which axis=0
3. Return the normalized data $D_{norm} = \frac{\alpha_p}{\alpha_s}$
4. Return the D_{norm}

Source: Authors

4. Experimental design and result analysis

This study will use PyTorch as the deep learning framework for building and training neural networks. Compared to TensorFlow, PyTorch employs dynamic computation graphs, meaning the computation graph is constructed dynamically based on the order of operations[21]. This approach is more flexible, particularly suitable for models involving conditional branches and loops. In terms of API friendliness, PyTorch's API is considered more Pythonic and intuitive, making it easy to learn and use. The API design of PyTorch is closer to the Python programming style. However, in terms of portability, TensorFlow has certain advantages in deployment and portability in production environments, as it supports tools like TensorRT and TensorFlow Serving for high-performance inference and deployment[22]. This experiment is primarily intended for academic research and does not involve engineering extensions.

4.1 *Experimental data description*

The data used in this study are sourced from surveys collected by the authors. In machine learning and deep learning, dataset partitioning typically involves dividing the dataset into training, validation, and test sets. This process is undertaken to assess the model's performance, tune hyperparameters, and ultimately evaluate the model. As shown in Table 3, all 402 data points are divided with proportions of 70% for the training set, 20% for the validation set, and 10% for the test set. The training set is used to train the machine learning model and usually constitutes the majority of the total data (typically 70-80%). The model undergoes parameter updates and learns from these data. The validation set is employed for adjusting model hyperparameters and selecting the best-performing model. Typically, the validation set forms a small portion of the total data (usually 10-20%). It is not used for training the model but rather for evaluating the model's performance under different hyperparameter settings to choose the optimal model. The test set serves for the final evaluation of the model's performance. It should be independent of the training and validation sets to ensure the objectivity of the evaluation.

Table 3. Data set partitioning

Name	**Options**	**quantity**	**Percentage(%)**	**Accumulation percentage (%)**
Data set Partitioning	Training set	281	70%	70%
	Validation set	80	20%	90%
	test set	41	10%	100%
Total		402	100%	

Source: Authors

When training deep learning models, it is typically necessary to have a large amount of data to achieve good performance. Data augmentation techniques can increase the quantity of training data by generating variants, thereby reducing the risk of overfitting. Data augmentation introduces diversity in the data, allowing the model to generalize better to unseen data, which can enhance both the performance and stability of the model. By incorporating data augmentation, the model finds it more challenging to memorize the details of the training data, thus alleviating overfitting issues. Since the quantity of data for this experiment is relatively scarce for a deep learning model, data augmentation methods will be employed to generate new data and ensure the fitting effect and robustness of the model. First, to ensure the robustness of the data, a portion of it will undergo mean processing to ensure data smoothness and reduce the proportion of outliers. Second, some data will undergo handling of missing values, compelling the model to better understand the context and semantics. This approach ensures the model's robustness in handling special cases and simulates noise and incompleteness in the real world. Next, a small portion of the data will be recombined to generate new data that the model has never seen before, aiding the model in better understanding the structure and relationships within the data. Finally, linear interpolation will be applied to the data to increase diversity, using mathematical methods such as Newton's interpolation to generate reasonable multi-dimensional data.

4.2 Experimental design description

This experiment involves testing the proposed algorithm on a private dataset. Initially, we will assess the basic classification performance using pure machine learning algorithms: Support

Vector Machine (SVM) and pure Multi-Layer Perceptron (MLP). Subsequently, the focus of this study is to investigate the importance of influencing factors in elderly care services. This entails not only fitting the data with the model but also interpreting the impression factors significantly affecting the classification results through the attention distribution output by the model's Self-Attention mechanism. The evaluation metrics used in the experiment include classification accuracy and accuracy variance. Table 4 provides the parameter settings for the MAC model. The learning rate is a hyperparameter that controls the magnitude of weight and bias updates in each iteration (a step-in batch gradient descent). A higher learning rate leads to larger weight updates but may result in an unstable training process, while a lower learning rate provides a more stable training process but may require more iterations to converge [23]. The Tanh function is zero-centered, meaning that, under the condition of zero mean input, the output's mean is also close to zero. This aids in accelerating the model's convergence. The output range of the Tanh function is useful for certain tasks (such as recurrent neural networks) because it can produce both positive and negative outputs. In this paper, the Cosine function is used, with a range of [-1, 1]. Therefore, the Rectified Linear Unit (Relu) function, which has a range of [0, +∞], is not chosen as the activation function in this paper. The number of categories corresponds to the degree of demand for community-based elderly care services and institutional elderly care services, each divided into 5 intervals. Finally, they are categorized into 10 types of elderly care service demand.

Table 4 Model parameter settings

parameter name	**Parameter size**
Learning rate	0.001
Number of categories	10
Activation function	Tanh

Source: Authors

4.3 Model performance verification experiment

This paper conducts three experiments to assess the impact of each module on the overall model. In this experiment, a controlled variable approach is employed to perform ablation studies on specific components, namely Self-Attention and the Cosine distance function. As shown in Table 5, when only the Cosine distance module is added, the accuracy of the model is not higher than that of the pure Multi-Layer Perceptron model; instead, it decreases by 0.8 percentage points. This may be due to the lower loss function during the fitting process with the Cosine distance module compared to the model with direct output, resulting in slower convergence. Consequently, within the same fixed training time, the accuracy does not surpass the baseline model. On the other hand, for the model with only the Self-Attention module added, there is a 1.4% improvement compared to the baseline model. This indicates that the Self-Attention module plays a positive role in extracting key information. For the purposes of this study, not only is an enhancement in classification performance needed, but also an analysis of the impact of different factors. Finally, for the ultimate MAC model in this research, with the simultaneous addition of both the Self-Attention module and the Cosine distance function, the model's accuracy improves by 4% compared to the baseline model, demonstrating that this model exhibits significant enhancement on the dataset.

Table 5 Ablation experiment

Self-Attention	**Cosine**	**Acc**	**Std**
		0.811	0.043
√		0.825	0.041
	√	0.803	0.058
√	√	0.851	0.039

Source: Authors

Based on the experimental results, the following four conclusions can be drawn:

(1) According to the ablation experiments, this study demonstrates that incorporating the Self-Attention module and Cosine distance into the Multi-Layer Perceptron leads to significant improvements in the model. The classification results are also optimal, and the standard deviation of this model is controlled at 0.039, lower than that of other models. This indicates that the discriminative ability and stability of the model used in this paper are relatively good. In this study, various data augmentation methods were applied to the input data, including operations like setting data to null or replacing it. These methods increase the richness and diversity of the data, while simultaneously providing the model with more training data and stronger robustness.

(2) Unlike other machine learning algorithms, deep learning algorithms can automatically learn to extract features from raw data without the need for manual feature engineering. This is highly advantageous in many tasks. In addition, this study introduced data augmentation methods in data preprocessing. On one hand, these data augmentation algorithms provide deep learning algorithms with more diverse data, enabling the model to have a stronger fitting ability. On the other hand, methods like setting data to null or replacing it simulate noise and incompleteness in the real world, enhancing the model's robustness in handling special cases. Deep learning models can capture

complex nonlinear relationships in data, making them perform better than traditional machine learning algorithms in certain tasks. For instance, the activation functions used in deep learning bring about a greater ability to fit complex nonlinear models. Deep learning models also offer high scalability and flexibility, effectively handling large amounts of data through parallel computing and distributed training. Deep learning can be accelerated using GPU cards, significantly reducing model fitting time [24].

(3) The Self-Attention model is particularly adept at globally capturing dependency information. The Self-Attention mechanism allows the model to dynamically allocate attention weights to adapt to the context at different positions in the input sequence. In the process of this study, the model leverages the Self-Attention mechanism to extract dependency information from the entire input data. This efficient computational approach ensures that the model can focus on crucial information while learning complex features.

(4) The MAC model used in this paper is an end-to-end neural network architecture. In contrast to traditional machine learning algorithms that require separate preprocessing and feature extraction, often necessitating manual design and selection of features in practical machine learning model design, which may require domain-specific expertise. By using deep neural network models, the issue of poor performance of traditional machine learning algorithms in handling highly complex nonlinear data relationships has been addressed. Moreover, the Self-Attention module used in this study eliminates the need for laborious data analysis using software like IBM SPSS, etc. After training, fitting, and testing the model, the required conclusions for this study can be obtained through a specialized model analysis report.

In summary, compared to traditional machine learning methods, the MAC model boasts higher accuracy, stability, and robustness. Additionally, the MAC model, as a deep learning model, eliminates the need for manual feature extraction, feature engineering, and thereby reduces the workload of this study [25]. The MAC model, as a deep learning model, not only achieves end-to-end learning but also directly analyzes the important parameter information needed for this study through the results returned by the introduced Self-Attention module.

4.4 Elderly care service demand analysis experiment

In this study, it is necessary to extract the direct outputs of the Self-Attention module in the analysis model of elderly care service demand factors. The weights of these outputs indicate the importance of different positions or features. As shown in Table 6, specific factors such as age, occupation, disposable income, number of children, lifestyle, number of chronic diseases, and sleep quality receive higher attention in the output values of the Self-Attention module. Among these, disposable income carries the highest weight. This suggests that in elderly care services, the level of disposable income often determines the choices made by elderly individuals regarding their care. Additionally, age, the number of chronic diseases, and lifestyle are all significant influencing factors in the elderly's demand for care. On one hand, as individuals age, the number of chronic diseases tends to increase. On the other hand, the lifestyle of the elderly often leads to various difficulties in their daily lives. These factors result in the elderly needing care services that provide good medical support and a quiet environment. When selecting care services, disposable income also determines the type of care service chosen by the elderly.

Table 6 Self-Attention module output values

Factor name	**Attention weight**
Age	4.4632071e-01
Profession	1.1016900e-01
Disposable income	9.0246118e-07
Number of children	1.2674949e-02
Lifestyle	3.4136746e-02
Number of chronic Diseases	3.9669773e-01
Sleep quality	1.0548121e-01

Source: Authors

The proposed MAC model in this paper performs classification for the analysis of elderly care service demands. This model innovatively incorporates the Self-Attention module and Cosine distance function as core components. By building upon the Multi-Layer Perceptron (MLP), the model focuses on global factors, integrating the strong generalization ability of the Cosine module to enhance the fitting of similar types. The data augmentation operations reduce individual differences between data points, generating a greater quantity and diversity of training data. This strengthens the model's fitting capability and robustness, especially in specific scenarios. However, the model used in this paper has certain limitations. First, in terms of comparing the fitting capabilities of the model, we only validated our own algorithm without conducting a horizontal comparison with pure machine learning methods. This is partly because the multitude of influencing factors used in our study exceeds the capacity of existing machine learning models to handle in large quantities. Additionally, our primary focus was on assessing the impact of the modules used on the final model's fitting ability, while the secondary focus was on studying factors with high attention in this model. Second, our model was unable to determine the positive or negative attention levels between influencing factors. Third, our model couldn't achieve the adjustment of the relationship between factors and influencing elements. Further research and development are needed in this area within the scope of this study for a more comprehensive understanding.

5. Conclusions

The MAC model proposed in this article classifies the analysis of elderly care service demands. This model innovatively utilizes a Self-Attention module and Cosine distance function as core components. It focuses on fitting global factors based on the use of a multi-layer perceptron (MLP) and integrates a highly generalizable Cosine module to enhance the model's fitting for similar types. Data augmentation operations on the data can reduce individual differences between data, generating a greater quantity and diversity of training data, thereby enhancing the model's fitting capability and robustness in special cases. The model used in this article also has certain limitations. Firstly, in terms of comparing the fitting capabilities of the model, it only validates its own algorithm without conducting a horizontal comparison using pure machine learning. On one hand, this is because the influencing factors used in this study are too numerous, and existing machine learning models cannot handle large quantities of data. On the other hand, the primary focus of this research is to verify the impact of the modules used on the final model's fitting capability, with a secondary focus on studying factors with high attention in this model. Second, this model is unable to determine the positive or negative attention between influencing factors. Third, this model is unable to achieve the adjustment of the relationship between adjusting factors and influencing factors. More time and development are needed for further research in this area.

Ethical Compliance

Research experiments conducted in this article with animals or humans were approved by the Ethical Committee and responsible authorities of our research organization(s) following all guidelines, regulations, legal, and ethical standards as required for humans or animals.

Not applicable

Conflicts of Interest

There are no conflicts to declare.

Acknowledgement

This study was supported by the Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province: Research on Construction Path of Aged Friendly Intelligent Community (2021SJA0938).

References

1. Ruma, Jannatul Ferdous; Anjum, Fayezah; Siddiqua, Rokeya; Rahman, Md. Abidur; Rohan, Abir Hossain; Rahman, Rashedur M., Outdoor patient classification in hospitals based on symptoms in Bengali language , Journal of Information and Telecommunication 7: 336-358, 2023.

2. Meng, Qiuqing ; Xiong, Huixiang, A doctor recommendation based on graph computing and lda topic model , International Journal of Computational Intelligence Systems 14:808-817, 2021..
3. Tseng, Hsiao-Ting; Huang, Hsieh-Hong; Hsieh, Chen-Chiung, Active Aging AI Community Care Ecosystem Design, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12208 LNCS:196-208, 2020.
4. Fahim, Muhammad ; Sharma, Vishal; Hunter, Ruth; Duong, Trung Q. Healthy Aging: A Deep Meta-Class Sequence Model to Integrate Intelligence in Digital Twin, IEEE Journal of Translational Engineering in Health and Medicine, 11:330-340, 2023.
5. Rajan Jeyaraj, Pandia; Nadar, Edward Rajan Samuel, Smart-Monitor: Patient Monitoring System for IoT-Based Healthcare System Using Deep Learning, IETE Journal of Research, 68:1435-1442,2022,
6. Ghosh, Shreya; Ghosh, Soumya K.,FEEL: FEderated LEarning Framework for ELderly Healthcare Using Edge-IoMT,IEEE Transactions on Computational Social Systems , 10:1800-1809, 2023.
7. Bianchi, Valentina; Bassoli, Marco; Lombardo, Gianfranco; Fornacciari, Paolo; Mordonini, Monica; De Munari, Ilaria,IoT Wearable Sensor and Deep Learning: An Integrated Approach for Personalized Human Activity Recognition in a Smart Home Environment, IEEE Internet of Things Journal 6: 8553-8562, 2019.
8. Sarraf, Aydin; Khalili, Saeed,An upper bound on the variance of scalar multilayer perceptrons for log-concave distributions, Neurocomputing 488: 540-546, 2022.
9. [4] Sebayang I S D , Fahmia M .Dependable flow modeling in upper basin Citarum using multilayer perceptron backpropagation[J].International Journal of Artificial Intelligence Research, 2021, 4(2):75.DOI:10.29099/ijair.v4i2.174.
10. Kim, Sung-Phil; Sanchez, Justin C.; Erdogmus, Deniz; Rao, Yadunandana N.; Wessberg, Johan; Principe, Jose C.; Nicolelis, Miguel, Divide-and-conquer approach for brain machine interfaces: Nonlinear mixture of competitive linear models, Neural Networks 16: 865-871, 2003.
11. Qureshi, Shahnawaz; Karrila, Seppo; Vanichayobon, Sirirut ,Human sleep scoring based on k-nearest neighbors, Turkish Journal of Electrical Engineering and Computer Sciences 26: 2802-2818, 2018.
12. Salmam, Fatima Zahra ; Madani, Abdellah; Kissi, Mohamed,Facial Expression Recognition Using Decision Trees, Proceedings - Computer Graphics, Imaging and Visualization: New Techniques and Trends, CGiV 2016, 125-130, May 10, 2016.
13. Wang, Xidong; Yang, Shaochun; Zhao, Yongfu; Wang, Ya,Lithology identification using an optimized KNN clustering method based on entropy-weighed cosine distance in Mesozoic strata of Gaoqing field, Jiyang depression, Journal of Petroleum Science and Engineering, 166:157-174, July 2018.
14. Wang, Shiann-Jang; Fuh, Chiou-Shann,RCV for adaptive dominant point detection,

Proceedings of the National Science Council, Republic of China, Part A: Physical Science and Engineering, 22: 61-77, Jan 1998.

15. Cravino, Nuno; Devezas, José; Figueira, Álvaro ,Using the overlapping community structure of a network of tags to improve text clustering, HT'12 - Proceedings of 23rd ACM Conference on Hypertext and social media, Milwaukee, WI, United states, 239-243, June 2012.
16. [Roodschild, Matías; Gotay Sardiñas, Jorge; Will, Adrián, A new approach for the vanishing gradient problem on sigmoid activation, Progress in Artificial Intelligence, 9:351-360, 2020.
17. Tliba, Marouane; Kerkouri, Mohamed A.; Ghariba, Bashir; Chetouani, Aladine; Coltekin, Arzu; Shehata, Mohamed Sami; Bruno, Alessandro,SATSal: A Multi-Level Self-Attention Based Architecture for Visual Saliency Prediction, IEEE Access 10:20701-20713, 2022.
18. Kumar, Sandeep; Solanki, Arun, an abstractive text summarization technique using transformer model with self-attention mechanism, Neural Computing and Applications 35:18603-18622,2023.
19. Ranjan, Ashish; Fahad, Md Shah; Fernandez-Baca, David; Tripathi, Sudhakar; Deepak, Akshay, IEEE/ACM Transactions on Computational Biology and Bioinformatics 20:1188-1199,2023.
20. Simon, Philomina; Vijayasundaram, Uma,WaveTexNeT: Ensemble Based Wavelet-Xception Deep Neural Network Architecture for Color Texture Classification,Traitement du Signal 39:1917-1927, 2022.
21. Jamali Golzar, Saleh; Karimian, Ghader; Shoaran, Maryam; Fattahi Sani, Mohammad,DGCNN on FPGA: Acceleration of the Point Cloud Classifier Using FPGAs, Circuits, Systems, and Signal Processing 42:748-779, 2023.
22. Le, Minh-Tri; Lien, Jenn-Jier James,Robot arm grasping using learning-based template matching and self-rotation learning network, International Journal of Advanced Manufacturing Technology 121:1915-1926.
23. Safari, Mir Jafar Sadegh; Rahimzadeh Arashloo, Shervin, Kernel ridge regression model for sediment transport in open channel flow, Neural Computing and Applications 33:11255-11271, 2021.
24. Osawa, Kazuki; Tsuji, Yohei; Ueno, Yuichiro; Naruse, Akira; Foo, Chuan-Sheng; Yokota, Rio, Scalable and Practical Natural Gradient for Large-Scale Deep Learning, IEEE Transactions on Pattern Analysis and Machine Intelligence 44:404-415,2022.
25. De Carvalho Filho, Antonio Oseas; De Sampaio, Wener Borges; Silva, Aristófanes Corrêa; de Paiva, Anselmo Cardoso; Nunes, Rodolfo Acatauassú; Gattass, Marcelo, Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index, Artificial Intelligence in Medicine 60: 165-177, 2014.
26. Nikoliuk, O., Sedikova, I., Savenko, I., Kalaman, O., & Korotych, O. (2023). Model of the State Policy for the Development of Hospitality Industry Enterprises in Crisis Situations and Military Operations. Journal of Law and Sustainable Development, 11(3), e812. https://doi.org/10.55908/sdgs.v11i3.812

27. Chanthes, S., Sawangloke, W., Pattiya, A., & Srithep, Y. (2023). Towards Sustainable Downstream Production: Upstream Innovation and Business Model Development for Low-Graded Ribbed Smoked Sheet Natural Rubber in Thailand. Journal of Law and Sustainable Development, 11(7), e907. https://doi.org/10.55908/sdgs.v11i7.907
28. Jian, Z., Rahman, N. F. A., & Ong, J. (2023). Assessing the Impacts of Climate Change on Water Resources Carrying Capacity Using Venism. Journal of Law and Sustainable Development, 11(8), e1426. https://doi.org/10.55908/sdgs.v11i8.1426
29. Abbas, M. A. Y., Setyadi, D., Paminto, A., & Azis, M. (2023). The antecedents of financial performance and their implications for firm value in mining sector companies listed on the Indonesia Stock Exchange. Journal of Law and Sustainable Development, 11(9), e991. https://doi.org/10.55908/sdgs.v11i9.991