**AMLI v2: A More Robust and Stable Feature Data Augmentation Method**

# 3. Preliminaries

AMLI v2 has been improved based on AMLI [1], and it has better performance than AMLI. To illustrate the proposed method clearly, AMLI method are introduced briefly in this section. In addition, we also summarize the shortcomings of the AMLI method.

## 3.1 AMLI

The AMLI (adaptive multipath linear interpolation) method is mainly based on the linear interpolation method to expand the samples of the original data. The idea is to divide the original feature space into several subspaces with an equal number of samples, extract one sample from each subspace as a class, and then perform linear interpolation for the samples in the same class, which is k-path linear interpolation. This method requires two hyperparameters (k and η) in advance. The visual interpretation of parameter k is the number of samples existing in each feature subspace, while η is the number of samples interpolated per unit distance in the linear interpolation of the samples.

The data set is given, where and Consider the model:

where is the actual value of  after removing the observation noise;  is a continuous function;  is the set of independent and identically distributed observation noise; and is the model error. Let for , we have , and call  the sample minimum point.

Given a hyperparameter k, the feature space can be partitioned into n/k subspaces, each containing k samples of equal size. i.e., , , , . For two adjacent subspaces, can be approximated as a linear function of , and then (1) can be transformed into:

where is the linear fitting error term. When the distance between two subspaces continues to approach and the measurement tends to zero,

Using the AMLI method, we randomly select one sample from each subspace without replacement to form a class, and repeat this process k times to obtain ,. For all samples in , define:

Where is the sample minimum point, . Perform interpolation in sequentially according to the order of values. Given another hyperparameter η, the number of samples inserted using linear interpolation method in is:

taking and (, ) as example, is the set of inserted samples, the linear interpolation formula is defined as:

*,* (6)

*.*

Repeat k times, perform interpolation on each class in , can be seen as k-path linear interpolation. According to (1), some verification indicators, the proportion of samples with an error greater than and the mean square error (MSE) between the sample observation values and the actual value, are selected to test the optimization effect of the AMLI method on the original sample after processing, as follows:

It has been proofed that the *MSE* and *p* of samples processed through the AMLI decreases when the following assumptions are met:

1. is a continuous function;

2.;

3. and are continuous variables.

## 3.2 Shortcomings

In the AMLI method, the selection of parameter k is critical, which varies with different samples. By selecting appropriate hyperparameters, many valid samples can be augmented, and the proportion of samples in which the observed value deviates greatly from the actual value is reduced so that the composition structure of samples with error is adjusted and sample optimization is achieved. Consequently, the impact of observation noise on the prediction result is also greatly reduced. Regarding the above, we have provided proof in [1]. However, the AMLI method has two serious shortcomings, as follows:

1. The results obtained from processing data using AMLI method are not unique.

2. When interpolating neighboring subspaces, the matching of the interpolated samples is not optimal.

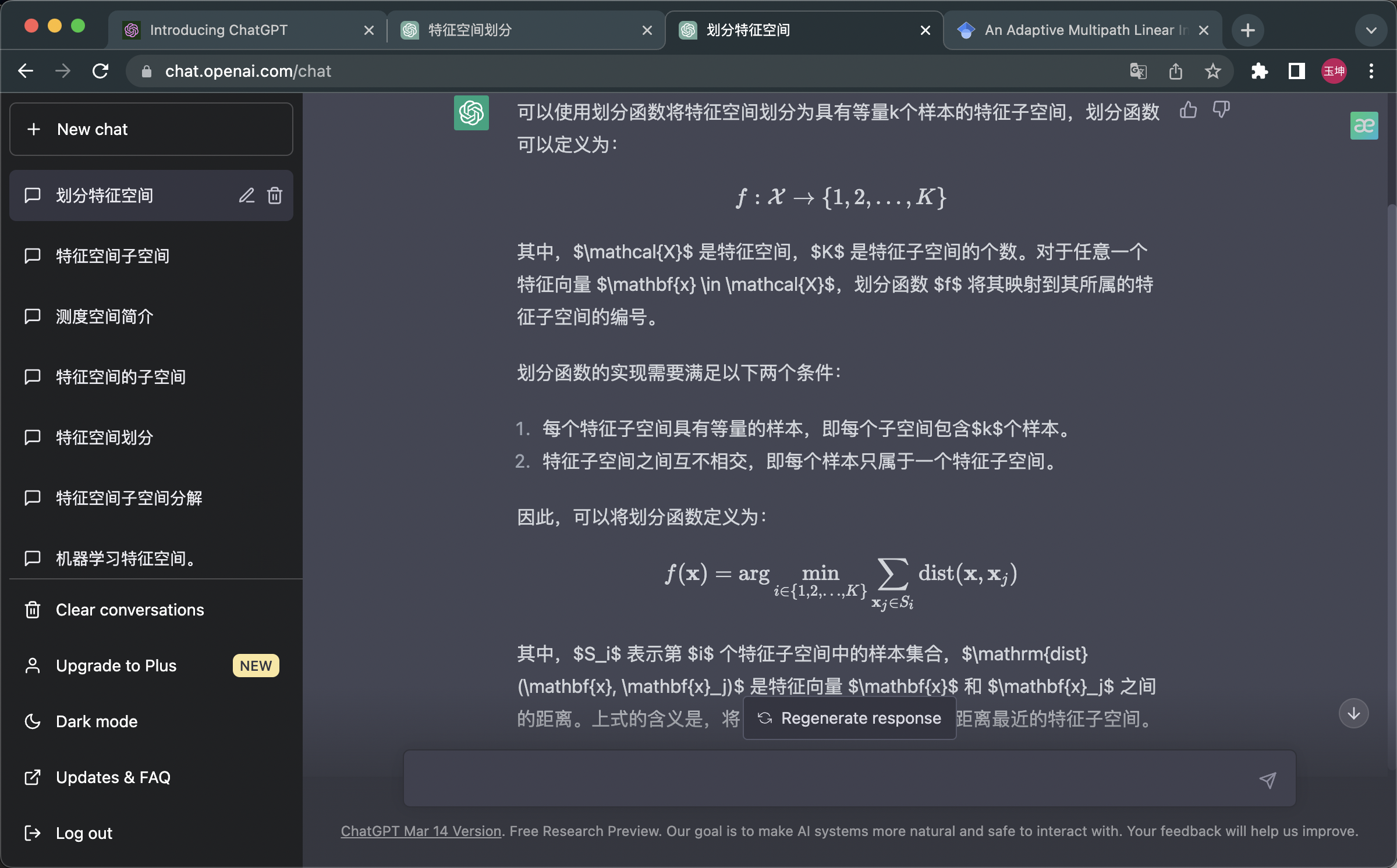
# 4. AMLI v2

In this section, we will explain the proposed AMLI v2 method in detail.

## 4.1 AMLI Cluster

We define:

这里定义一下子空间的距离公式以及子空间大小的公式.当趋于0时,公式2 的e趋于零.所以我们要有一个聚类方法…..



## 4.2 Subset Sample Matching

## 4.3 Application of Classification

# 5. Experiments

## 5.1 Artificial Data

## 5.2 Benchmark Data Sets

# 6. Conclusion

# References

[[1] Du Y, Jin X, Wang H, et al. An Adaptive Multipath Linear Interpolation Method for Sample Optimization[J]. Mathematics, 2023, 11(3): 768.](https://www.mdpi.com/2227-7390/11/3/768)