Evaluation of Lempel-Ziv Complexity for Brain Region Segmentation Using fMRI Data

Sijie Li, Yushan Wei, Xintong Shen and Yidan Yao
Souther University of Science and Technology
Correspondence: liuqy@sustech.edu.cn Contact

BMEB215
机器学习与医学工程应用
ML & MEA (2024 Spring)

Introduction

Functional Magnetic Resonance Imaging (fMRI) has become a pivotal tool in neuro- science, providing insights into the functional connectivity and activity patterns of the human brain. Traditional approaches to brain region segmentation using fMRI data often rely on features derived from functional connectivity, requiring large datasets and significant computational resources. In contrast, Lempel-Ziv Complexity (LZC) has shown promising bio-interpretability in other applications, but its use in fMRI data analysis remains underexplored. This study aims to fill this gap by employing LZC for brain region segmentation. By utilizing both hierarchical clustering and K-means clus- tering, we achieve accurate and efficient brain region segmentation with minimal data and computation. Our results demonstrate that LZC-based features significantly out- perform traditional time series data in clustering performance, as validated by metrics such as correlation, homogeneity, completeness, silhouette score, and neural network prediction accuracy. This study highlights the potential of LZC as a robust and inter- pretable feature for brain region segmentation, providing a novel approach that is both computationally efficient and highly accurate

Data and Methods

The framework of our work is illustrated in Figure 1.

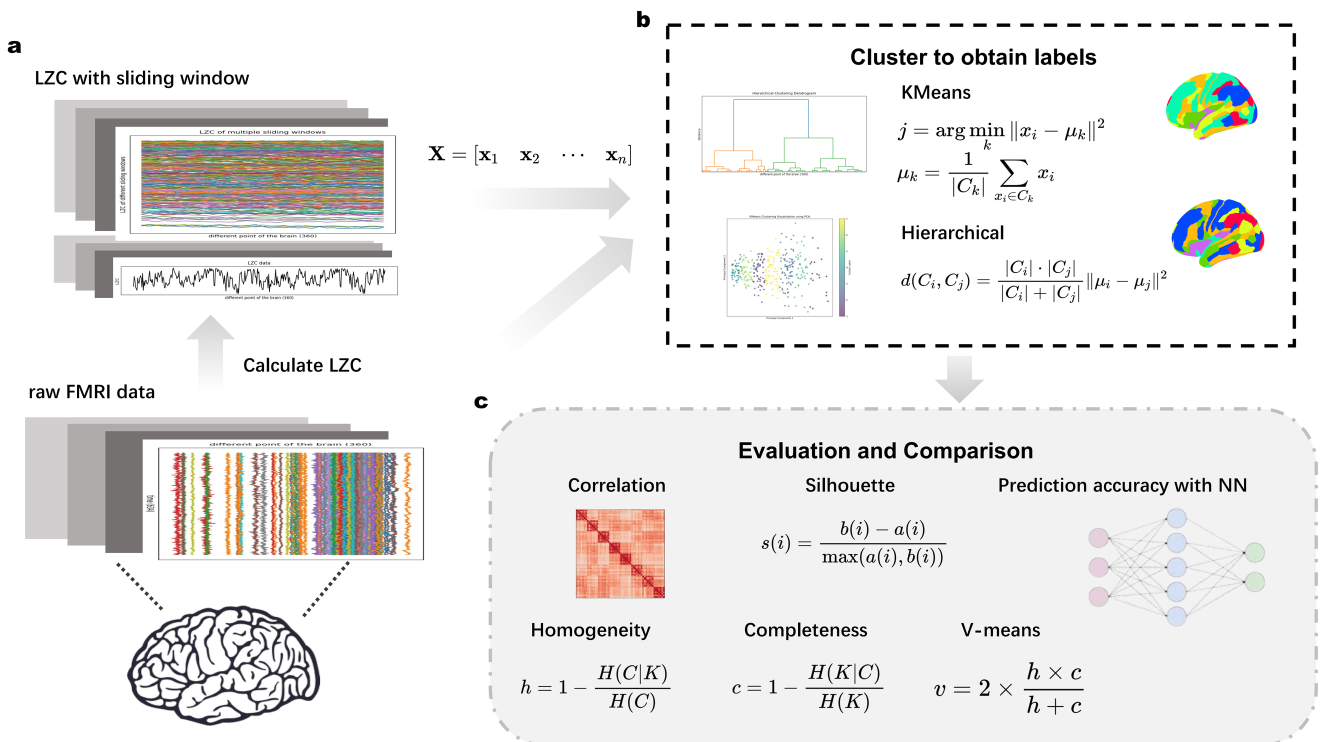

Figure 1: (a) Extract LZC features from the fMRI time series of the 50 Chinese participants' REST1 state data. Also, apply sliding window techniques with various window sizes and steps, ensuring the representation of dynamic brain activity. (b) Normalize the data to mitigate individual differences, followed by clustering using two methods: hierarchical clustering with the Ward method and K-means clustering, both set to identify seven clusters corresponding to distinct brain regions. (c) Validate the clustering results using correlation, homogeneity, completeness, silhouette score, and prediction accuracy with neural networks.

Results

Qualitative analysis (Figure 2. and 3.) is carried out to initially validate the effect of LZC feature in brain region segmentation.

Figure 2: Cluster a small data batch and visualize the result by PCA.

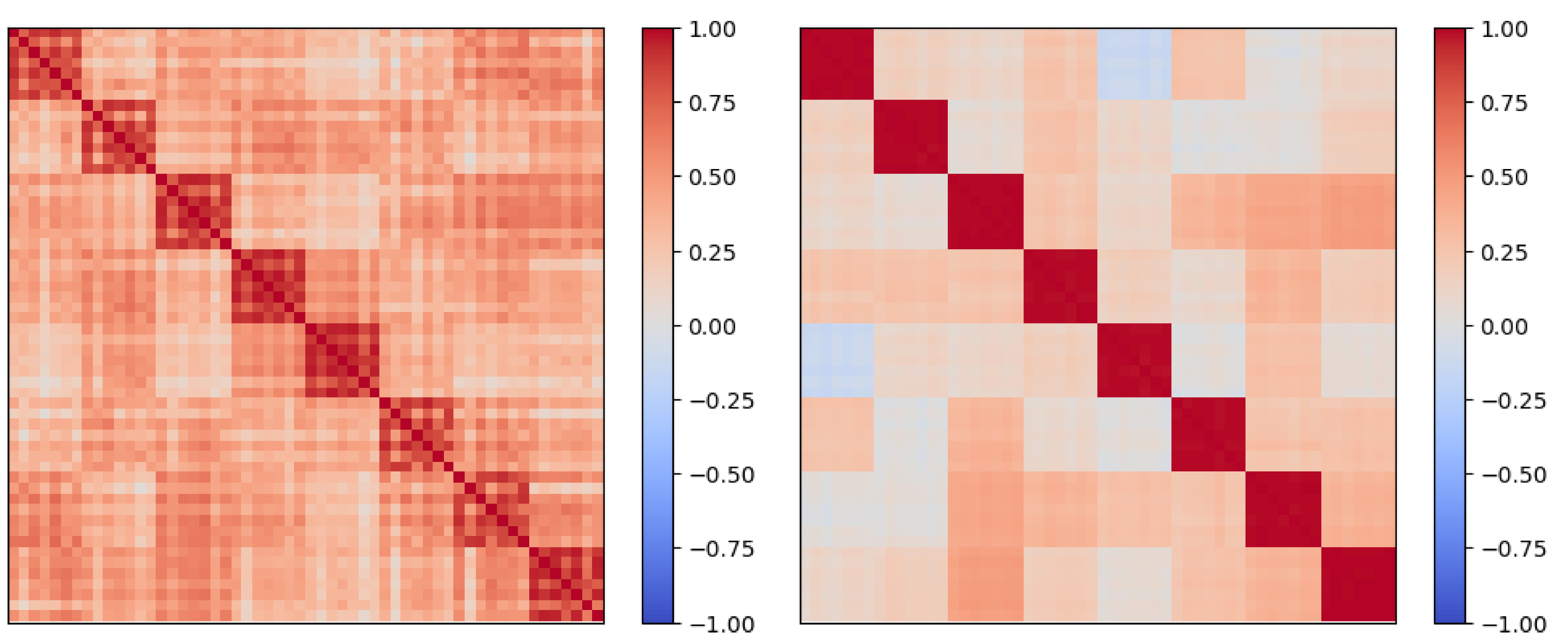

Figure 3: Comparison of correlation matrices: LZC-based clustering labels (left) versus random labels (right), representing relationships between brain regions across eight states.

As quantitative analysis, the homogeneity, completeness, V-measure, silhouette score, and neural network prediction accuracy are presented in Table 1, 2, 3, and 4.

	Time series	LZC_0	LZC_200	LZC_300	LZC_400	LZC_500	LZC_600
Homogeneity	0.8624	1	0.888	0.8978	0.8617	0.8081	0.8485
Completeness	0.8619	1	0.8923	0.9100	0.8661	0.7927	0.8437
V-Measure	0.8622	1	0.8905	0.9039	0.8639	0.8003	0.8461

Table 1: Homogeneity, completeness, and V-measure scores for LZC and time series features across different K-means clustering seeds.LZC features consistently achieve better clustering results.

	Time series	LZC_0	LZC_200	LZC_300	LZC_400	LZC_500	LZC_600
Homogeneity	0.0830	0.8947	0.7775	0.7533	0.7397	0.6687	0.7784
Completeness	0.0701	0.8892	0.7933	0.7794	0.7661	0.6976	0.7966
V-Measure	0.0760	0.8919	0.7853	0.7661	0.7527	0.6828	0.7874

Table 2: Comparison of clustering methods (hierarchical vs. K-means) for LZC and time series features.LZC features show higher homogeneity, completeness, and V-measure scores regardless of the clustering method used.

	Time series	LZC_0	LZC_200	LZC_300	LZC_400	LZC_500	LZC_600
KMeans	0.5150	0.5753	0.4697	0.4977	0.5116	0.5284	0.5443
Hierarchical	-0.2462	0.5609	0.4406	0.4727	0.4820	0.5184	0.5087

Table 3: Silhouette scores for LZC and time series features.LZC features achieve significantly higher silhouette scores, indicating better clustering performance.

	Time series	LZC_0
KMeans	0.5722	0.6999
Hierarchical	0.5833	0.6278

Table 4: The LZC prediction results are better than the time series

All of the above results indicate that the LZC method demonstrates superior performance compared to time-series clustering across multiple evaluation metrics.

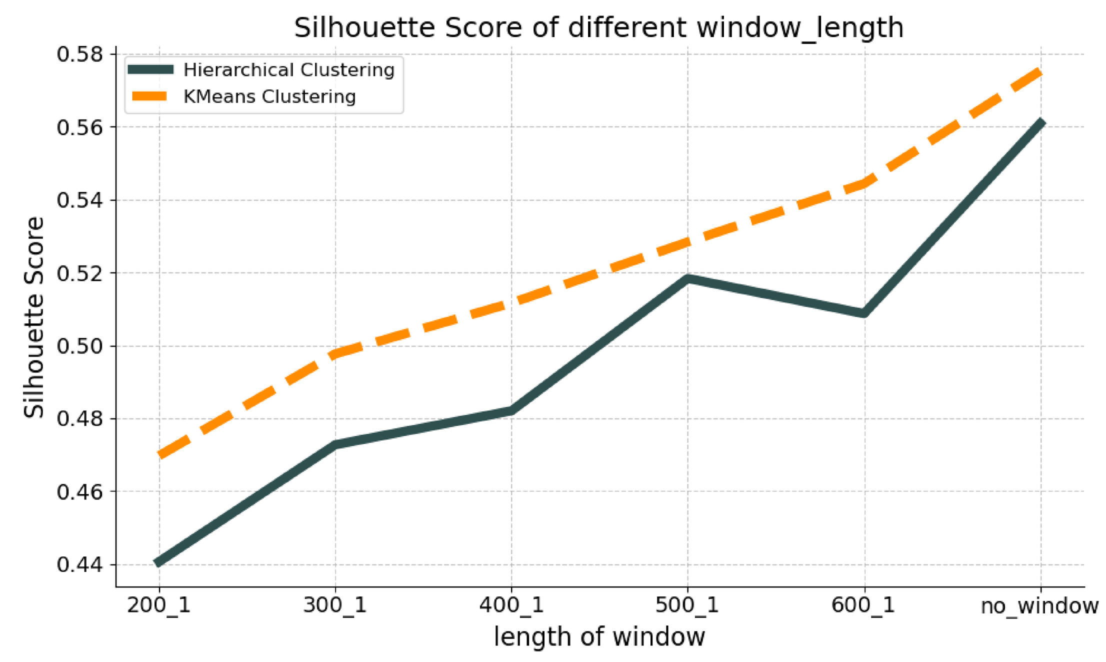

Figure 4: With the reduction of the number of Windows, the clustering effect becomes better

An interesting phenomenon occurs when the windowing is entirely removed, resulting in an overall LZC measure; the silhouette coefficient reaches its maximum.

Summary

In this study, we introduced an effective approach for brain region segmentation by leveraging LZC as a feature extraction method on fMRI data. Our methodology addressed the limitations of traditional time series-based clustering by capturing the dynamic nature of brain activity.

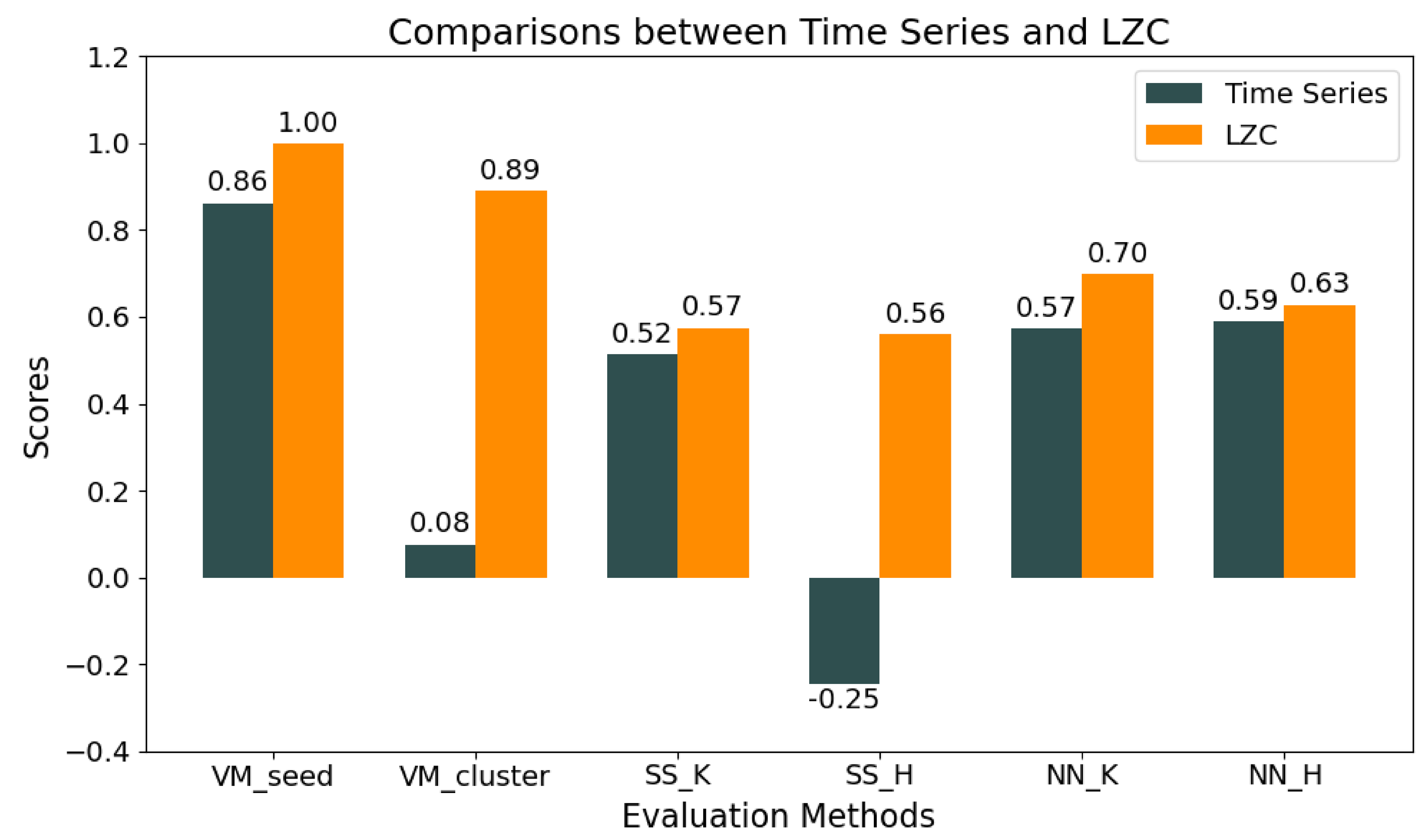

Figure 5: Clustering outcomes using LZC features consistently outperform those based on time series data across various metrics

Our approach is computationally efficient, requiring minimal data and simple clustering algorithms for accurate brain region segmentation.

In conclusion, our findings highlight LZC as an effective feature for brain region segmentation, providing a robust and interpretable alternative to traditional time series analysis. Future research could explore integrating LZC with advanced machine learning techniques to further enhance brain region segmentation and understanding of brain connectivity.

Acknowledgments

This study was supervised by Quanying Liu. The authors would like to thank Quanying Liu, Junjie Yu and Kaining Peng for their patient and inspiring tutorial.