UDC 004.81

APPLICATION OF SORTED METRICS BASED ON INSOLE FEATURE MINING

H. LI, V.Yu. TSVIATKOU

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received April 04, 2025

Abstract. To improve the accuracy and real time performance of fall risk assessment, this study proposes a feature selection method based on sorted metrics. The method evaluates previously generated spatiotemporal features using correlation coefficient (CCF), bayes factor (BF), and self-information index (SII) to identify the most informative features. The results show that these metrics play complementary roles in the feature selection process. CCF measures the linear correlation between features and helps eliminate redundancy. BF emphasizes the statistical significance between high-risk and low-risk groups. SII captures distributional differences from the perspective of information entropy. Comparative analysis across various metrics demonstrates that the combined method achieves high accuracy across multiple performance indicators while significantly reducing computational complexity.

Keywords: sorted metrics, fall risk, self-information index, bayes factor.

Introduction

Falls are a common and serious health issue among the elderly and individuals with neurological disorders. They not only lead to physical injuries but may also trigger psychological fear, a decline in quality of life, and increased medical costs. Therefore, developing an efficient, real time, and reliable method for fall risk assessment has become a crucial research focus in the field of intelligent health monitoring [1].

In recent years, with the advancement of wearable technology, gait analysis based on plantar pressure sensors has emerged as an important approach for evaluating individual stability and motor function. By analyzing spatiotemporal features associated with fall risk, researchers can identify potentially high-risk individuals at an early stage, enabling timely intervention and prevention.

However, the large number of gait features generated by nonparametric models can lead to the curse of dimensionality and may degrade model performance due to redundant or low-quality information. Therefore, efficiently selecting the most discriminative key features from massive datasets has become a core challenge in fall risk modeling.

To address this issue, this paper proposes a feature selection method based on sorted metrics, which applies multiple evaluation criteria to progressively screen features, thereby optimizing model performance and improving prediction accuracy. Compared with traditional feature selection methods, this approach not only significantly improves accuracy but also enhances real time responsiveness and model simplicity, making it highly suitable for practical applications.

Feature mining based on sorted metrics

Feature selection is a critical step in pattern recognition and machine learning. Its primary goal is to reduce the number of input variables as much as possible without compromising predictive performance, thereby lowering computational complexity and enhancing model generalization. Traditional feature selection methods perform well when dealing with linear features or small datasets, but they often suffer from information loss or overfitting when faced with high-dimensional, redundant, or nonlinear feature structures.

To address these challenges, this study introduces a sorted metrics method based on statistical and information theoretic principles. This method offers strong interpretability and adaptability, allowing flexible combination of various metrics suitable for different data distributions and feature types. Specifically, the proposed method ranks and filters features through multiple stages using three core metrics.

First, the correlation coefficient (CCF) [2] measures the linear correlation between features and removes redundant ones based on a defined threshold. Common forms include Pearson coefficient, Spearman rank coefficient, and Kendall Tau coefficient.

Second, the bayes factor (BF) [3], as a Bayesian hypothesis testing tool, evaluates the significance of features between high-risk and low-risk groups. A smaller BF value indicates stronger discriminative power.

Third, the self-information index (SII) [4] characterizes the variation in feature information based on entropy, capturing the degree of surprise to reveal key differences between risk groups.

Based on these metrics, a three-stage selection strategy is implemented. The first stage uses CCF to eliminate low correlation or duplicate features. The second stage employs BF to retain features with strong group discrimination. The third stage applies SII to select features with significant information differences, resulting in the final feature subset.

This strategy optimizes the selection process from correlation analysis to statistical significance evaluation and information density comparison, ensuring that the final features are representative, independent, and highly predictive.

Experimental results and analysis

This section presents a detailed evaluation of the proposed sorted-metrics-based feature selection approach by analyzing the performance of various metric-based methods. The comparison focuses on three categories of metrics used in feature screening: correlation-based, statistical hypothesis-based, and information-theoretic methods. The results are summarized in Figure 1 and evaluated using multiple performance indicators, including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and execution time.

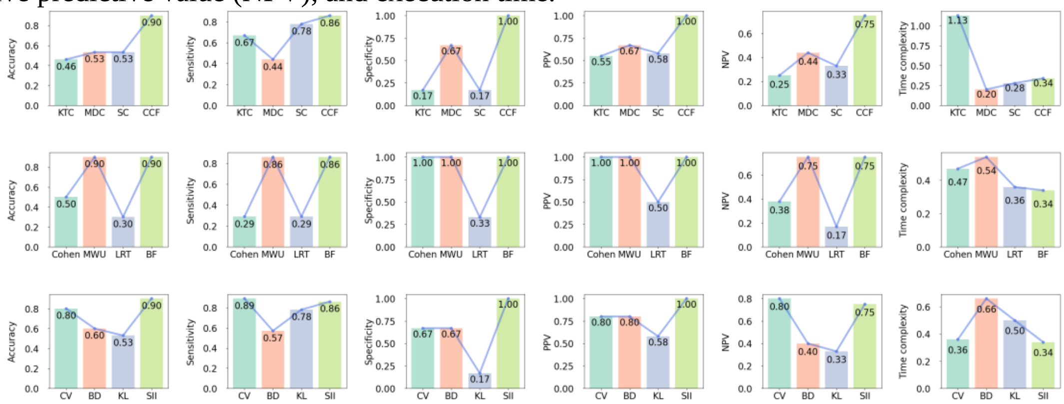

Figure 1. Performance assessment of multiple sorted-metrics based methods

The first category includes methods based on the correlation coefficient, such as Kendall tau correlation (KTC), multivariate distance correlation (MDC), and spearman correlation (SC). These metrics quantify the level of association between features, making them useful for identifying and removing redundant features during the initial screening stage. Among these, CCF demonstrates the most consistent and robust performance. It not only achieves high classification Accuracy but also maintains low computational cost, making it especially suitable for real time fall risk detection systems.

The second category consists of methods based on the bayes factor, including Cohen’s effect size, the Mann Whitney U test (MWU), and the likelihood ratio test (LRT). These metrics evaluate the difference in feature distributions between high risk and low risk groups using hypothesis testing principles. Both BF and MWU show strong predictive power, with BF slightly outperforming others in terms of Accuracy. However, MWU has longer Execution Time, which may hinder its practical use in applications requiring fast responses.

The third category includes information theoretic methods, such as the coefficient of variation (CV), Bhattacharyya distance (BD), and Kullback divergence (KL). These metrics assess the differences in feature distributions based on entropy, divergence, and variability. SII performs competitively in this category. Although it slightly lags behind CV in Sensitivity and NPV, SII surpasses other methods in Accuracy and shows a well-balanced performance across most criteria.

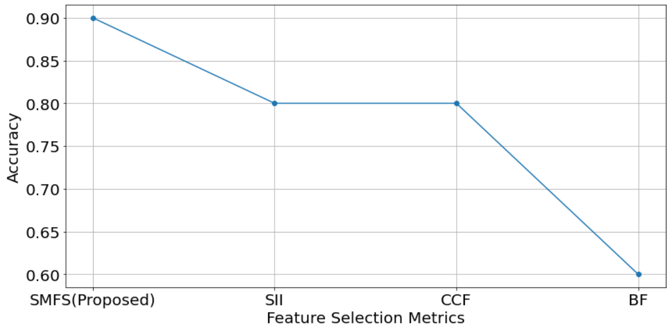

Figure 2. Performance assessment of hybrid and nonhybrid sorted-metrics based methods

In Figure 2 illustrates the relationship between fall risk prediction accuracy and the use of different sorted metrics, including the proposed SMFS method and individual metric based on methods derived from the three metrics: SII, CCF, and BF. The graph shows that both SII and CCF achieve the same level of accuracy. However, CCF selects 38 features, whereas SII requires only 4, highlighting the superior efficiency of SII in reducing feature dimensionality. The proposed SMFS method achieves the highest accuracy of 0,9, outperforming all single-metric approaches and clearly demonstrating the necessity and effectiveness of SMFS.

In summary, the experimental results indicate that CCF, BF, and SII consistently outperform other metrics in selecting the most relevant and discriminative features for fall risk prediction. The combination of these three metrics supports a multistep feature selection process that moves from redundancy reduction to statistical separation and finally to information density enhancement. This approach not only improves classification Accuracy but also ensures the selected features are efficient for real-time implementation in wearable systems using plantar pressure data.

Conclusion

This paper proposes a feature selection method based on sorted metrics for fall risk assessment. Experimental results demonstrate that the method achieves high prediction accuracy while reducing the number of selected features and satisfying real-time computational constraints. Future work will focus on enhancing multi-sensor data fusion and refining decision rules to develop more comprehensive solutions for both clinical and home-based monitoring applications.

References

1. Gardner H. // Harvard Graduate School of Education. 2011. P. 9.
2. Kass R. E. // Journal of the American Statistical Association. 1995. Vol. 90(430). P. 773–795.
3. Asuero A. G., Sayago A., González A. G. // Critical Reviews in Analytical Chemistry. 2006. Vol. 36(1). P. 41–59.
4. Li H. // IEEE Sensors Journal. 2025. Vol. 25(5). P. 8721–8736.