RADemics

Quantum Feature Selection Methods for Improved Machine Learning Models

Prasanth SP, Yukti Varshney
V.S.B. ENGINEERING COLLEGE, MORADABAD INSTITUTE OF TECHNOLOGY

Quantum Feature Selection Methods for Improved Machine Learning Models

Prasanth SP, Assistant Professor, Department of CSE, V.S.B. Engineering College, Karur, India, spmtechit@gmail.com

Yukti Varshney, Assistant Professor, Moradabad Institute of Technology Moradabad Uttar Pradesh India, yuktivarshney16@gmail.com

Abstract

Feature selection was a critical process in machine learning that enhances model performance by identifying the most relevant features from high-dimensional datasets. This book chapter delves into various feature selection methodologies, emphasizing the significance of hybrid approaches that leverage the strengths of different techniques, including filter, wrapper, and embedded methods. It critically examines the advantages and limitations of each method, providing insights into their applicability across diverse domains. The chapter also explores future directions in hybrid feature selection, including the integration of advanced algorithms, adaptation to varying data types, and the incorporation of domain knowledge. Emphasizing computational efficiency and real-time application potential, this work serves as a comprehensive guide for researchers and practitioners aiming to enhance machine learning models through effective feature selection. The findings and discussions presented herein contribute to the ongoing discourse in the field and provide a roadmap for future research initiatives.

Keywords:

Feature Selection, Hybrid Methods, Machine Learning, Computational Efficiency, Real-Time Applications, Domain Knowledge

Introduction

Feature selection plays a pivotal role in the machine learning landscape, directly impacting the accuracy and interpretability of predictive models [1]. In an era characterized by big data, the availability of high-dimensional datasets has surged, creating both opportunities and challenges for data scientists [2]. High-dimensional data can lead to the “curse of dimensionality,” where models become overly complex and prone to overfitting [3,4,5]. This phenomenon highlights the necessity for effective feature selection techniques, which aim to retain only the most relevant features while eliminating noise and redundancy [6]. By focusing on a subset of relevant variables, feature selection can significantly enhance the model's ability to generalize to unseen data, thus improving overall performance [7].

The diverse array of feature selection methods can be broadly categorized into three main types: filter, wrapper, and embedded methods [8,9,10]. Filter methods assess the relevance of features based on intrinsic properties of the data, such as correlation or statistical significance, independently of any learning algorithm [11]. These methods are typically computationally

efficient, making them suitable for large datasets [12]. Wrapper methods, on the other hand, evaluate feature subsets by employing a specific learning algorithm, thus providing a more tailored approach to feature selection [13,14]. Their computational demands can be substantial, particularly with high-dimensional data [15]. Embedded methods incorporate feature selection into the model training process itself, resulting in a hybrid approach that balances efficiency and accuracy [16,17,18].

Hybrid feature selection methods represent an innovative convergence of the aforementioned techniques, combining their strengths to produce superior outcomes [19]. By integrating filter and wrapper approaches, hybrid methods can optimize feature selection while maintaining computational efficiency [20]. This synergy enables practitioners to identify the most relevant features more effectively, leveraging the strengths of each technique to mitigate their respective limitations. As machine learning continues to evolve, hybrid methods are gaining traction as a preferred approach, especially in applications that require both high performance and interpretability [21].

The application of hybrid feature selection methods spans various domains, including healthcare, finance, and image processing [22]. In healthcare, for instance, accurate feature selection can lead to improved diagnostic models by identifying the most relevant clinical variables [23]. Similarly, in finance, effective feature selection can enhance predictive accuracy for stock market forecasting or credit risk assessment [24]. The significance of feature selection in these domains underscores the necessity for ongoing research and development of innovative methodologies [25]. As the complexity of datasets increases, the demand for efficient and effective feature selection techniques becomes ever more critical.