Machine Learning Applications in Biomass Supply Chain Management and Optimization

Jingxin Wang

Forest and biomass resource utilization for bioenergy and bioproducts is crucial for mitigating climate change and promoting a sustainable bioeconomy. Given that the biomass supply chain is a complex system, one of the most concerning issues is selecting and using appropriate modeling and analytical technologies to optimize the advantages of multi-feedstock biomass supply chains. Machine learning (ML) can enhance biomass supply chain management (BSCM) efficiency and sustainability. It can address the complexities in cultivation, harvesting, preprocessing, storage, transportation, and conversion. ML workflows involve data collection, preprocessing, model training, and optimization, using algorithms for prediction and decision-making. Accurate supply and demand forecasting via ML improves production planning and inventory management. Despite its potential, ML applications in BSCM need to deal with challenges such as data availability and quality, interpretability of models, and their generalization capabilities. Overcoming such challenges requires interdisciplinary efforts in data management and model development to fully leverage ML's applicability.

DOI: 10.15376/biores.19.4.6961-6963

Keywords: Machine learning; Biomass supply chain management; Predictive modeling; Data-driven optimization; Bioenergy; Bioproducts

Contact information: Professor and Head, Department of Forest Biomaterials, North Carolina State University, Raleigh, NC, USA, 27695; jingxin_wang@ncsu.edu

An Emerging Tool for Biomass Management

In the context of Industry 4.0, the potential of data-driven applications has promoted machine learning (ML) to stand out as a transformative force across various fields. ML, a subset of artificial intelligence, enables systems to learn from data and mimic the decision-making processes of the human brain. One promising yet underdeveloped application area is ML for biomass supply chain management (BSCM). Biomass includes organic materials such as wood, forest and agricultural residues, and energy crops. These renewable energy sources can be used for biopower, biofuels, and bioproducts. The development of biomass feedstocks not only provides a viable pathway to mitigating climate change but also enhances energy security, stimulates the bioeconomy, and drives technological innovation. However, the biomass supply chain is a complex system that encompasses the cultivation, harvesting, preprocessing, storage, transportation, and conversion of biomass feedstocks. Variations in supply and demand, large-scale planning and scheduling, storage and processing requirements, technological demands, and sustainability considerations pose significant challenges to BSCM. Therefore, integrating advanced ML technologies into BSCM is essential in wood and biomass supply chain optimization.

The Computer as Student

The essence of ML lies in enabling computer systems to automatically learn and improve through data-driven methods, thereby accomplishing specific tasks without explicit programming. This process relies on algorithms and statistical models to discover patterns and rules in data, as well as to make predictions, classifications, and decisions. A typical workflow of ML includes data collection, data preprocessing, feature extraction, model selection, model training, model evaluation, and model optimization. ML is fundamentally based on data. By handling and analyzing large amounts of multi-source data, algorithms can extract useful information. ML depends on various algorithms to train models for performing prediction and classification tasks. There are several types of ML, including supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. In supervised learning, algorithms are trained on labeled data, meaning each item of input data has a corresponding output. Unsupervised learning deals with unlabeled data, where algorithms need to find inherent patterns and structures within the data. Semi-supervised learning falls between supervised and unsupervised learning, using a small amount of labeled data along with a large amount of unlabeled data for training. Reinforcement learning, on the other hand, learns through interaction with the environment, where algorithms adjust their strategies based on feedback (rewards or punishments) to maximize cumulative rewards.

Forecasting

Accurate supply and demand forecasting is the cornerstone of efficient BSCM. Due to the seasonality and variability of biomass, traditional empirical models often struggle to achieve the desired results. ML models offer more robust solutions. These models analyze historical data and real-time inputs to predict the availability and demand of biomass more accurately. The improved prediction accuracy enables biomass facilities or factories to better adjust production plans, thereby reducing the risks of supply shortages or overproduction. Additionally, ML models can integrate various environmental, economic, and social factors, such as weather conditions, market trends, and population density, that significantly influence the supply and demand. By continuously learning from new data, these models can adapt to changing conditions, providing dynamic and reliable forecasts. This adaptability is crucial for the bioenergy and bio-based products industries. Improved supply and demand forecasting not only ensures a stable supply of raw materials but also helps maintain optimal inventory levels, thereby reducing storage costs.

Big data from various stages of the biomass supply chain provides inputs for training ML models. Traditionally, mathematical programming methods have been used to solve decision-making problems in BSCM, especially mixed-integer linear programming, through single-objective optimization. However, as the number of variables, objectives, and constraints increases, mathematical programming struggles to deal with the complexity of the model and the exponential growth of alternative solutions, making it difficult to solve large-scale and real-time problems. Currently, reinforcement learning has been proposed to address real-time online scheduling problems. Reinforcement learning can handle many constraints through penalty functions, demonstrating strength superior to traditional methods. In addition, as ML focuses on classification and regression, it can also simplify optimization models through variable classification and parameter prediction. For example, using decision tree algorithms to classify biomass feedstocks, employing support vector machines to evaluate suppliers and analyze uncertainty, and using decision trees and logistic regression to determine the optimal locations of biomass and bioproducts facilities.

ML can also offer technical support in wood and biomass conversion processes, including pyrolysis, co-pyrolysis, torrefaction, hydrothermal liquefaction, and gasification. By monitoring and analyzing various parameters throughout production, ML algorithms can optimize operational conditions, forecast product properties, and examine parameter impacts. For instance, algorithms such as random forest, extreme gradient boosting, and neural networks have been used to predict biochar and bio-oil yield and properties in the aforementioned conversions, as well as the metal adsorption efficiency of biochar. In the pulp and paper industry, ML technology proves highly advantageous. Through the analysis of biomass raw material characteristics and papermaking process parameters, ML can optimize manufacturing processes, reduce energy and water consumption, and elevate paper quality and production efficiency. For instance, algorithms can adjust the process parameters of pulping and papermaking based on the characteristics of biomass fibers to produce high-quality and more efficient paper products.

Challenges

While ML has significant potential for application in BSCM, it also faces several challenges, primarily related to data availability and quality. The biomass supply chain encompasses multiple stages and stakeholders, leading to diverse and uneven data sources. Raw data may be incomplete, erroneous, or inconsistent, which can result in inaccurate training and prediction outcomes. Data integration and consolidation are also complex issues because data are often dispersed across different systems and platforms. In BSCM, decisions often involve multiple factors and stakeholders. Some deep learning models may lack interpretability, affecting understanding and trust in decision-making. Additionally, BSCM involves various scenarios across different regions, times, and environments. ML models must have strong generalization capabilities to adapt to new data and scenarios. However, if a model overfits the data, it may lack the necessary generalization ability and fail to accurately apply to new situations. Addressing these challenges requires interdisciplinary collaboration and innovation, including efforts in data management, quality control, and model development.