**Introduction**

The rapid development in the field of data visualization and machine learning in the last decade has shown such promise in business analytics and artificial intelligence. Researchers have done tremendous improvements to overcome the challenges and optimize different techniques. To understand the design and implementation of these research works, a comprehensive survey of the research works proposed in a recent time can help to understand the overall trend. That’s why in this report, a paper titled “A survey of visual analytics techniques for machine learning” [1] is selected which reviews a total number of 259 papers in the area of data visualization and machine learning techniques. The literatures in the survey were selected by focusing on visual analytics techniques that help to develop explainable, trustworthy, and reliable machine learning applications. To comprehensively survey visual analytics techniques for machine learning, the paper presented a descriptive review of relevant top-tier sources in the past ten years (2010–2020). In this report, a brief outline of the selected paper is presented along with key findings of the techniques and critical analysis of the comprehensive survey. The report will also provide some recommendations for developing machine learning pipelines and point out key challenges and promising potential future research opportunities useful for visual analytics researchers.

**Body**

In the selected paper, the authors aimed to provide a comprehensive survey of visual analytics techniques for machine learning, which focuses on every phase of the machine learning pipeline. The paper comprehends a variety of visual analysis methods that had been proposed in the last ten years to make machine learning applications more explainable, trustworthy, and reliable. The literature survey is presented in three stages to present the overall pipeline along with supplemental tasks including data and feature quality and preparation. The authors divided the vital techniques from 259 papers based on the implementation methods such as process before model building (Data preparation & Feature extraction to improve data and feature quality), process during model building (Model Selection & Model Training for model understanding and diagnosis) and process after model building (Evaluation & Deployment to understand the performance and analyze). Several papers have been shown as examples to understand the key difference in such methods.

In this report, a descriptive analysis of all these three processes is presented and a set of key findings from the survey is outlined. Finally a few recommendations proposed for each of the processes.

**Techniques before model building:** The key motif for developing techniques before model building is to provide a firm guideline for the model developers about the data preparation. As the performance of the model will largely depend on the quality of the data, these techniques are really important. The quality of the data is mainly determined by the data itself and the features used. From the literatures, the paper outlined two important research directions before model building such as data quality improvement and feature engineering.

The insertion of incomplete data attributes and the correction of erroneous data marks are two examples of ways to increase data accuracy. Data consistency problems include missing values, outliers, and noise in instances and their names. Irrelevant features, redundancy between features, and other feature quality problems are examples of feature quality issues. Although manually handling these problems takes time, automated solutions may be ineffective. As a result, a variety of visual analytics approaches have been introduced to reduce experts' workload while also the the reliability of automated data and feature generation processes. These tasks were often completed by hand or with the assistance of computational approaches such as learning-from-crowds algorithms, which attempt to estimate ground-truth labels from noisy crowd-sourced labels. Visual analytics strategies are used in several works to interactively optimize data consistency in order to reduce effort or improve the results of automatic systems. The data includes instances and their names. Current attempts to improve data quality either concern instance-level improvement or label-level improvement, according to this viewpoint.

The best features for training the model are chosen using feature engineering. Instead of using raw image pixels, we might use HOG (Histogram of Oriented Gradient) features in computer vision. Interactive feature selection offers an interactive and iterative feature selection process in visual analytics. In recent years, in the age of deep learning, the majority of feature discovery and design has been done using neural networks. A typical feature selection technique is to pick a subset of features that has the least amount of redundancy and the most relevance to the targets (e.g., classes of instances). Along this line, several methods have been developed to interactively analyze the redundancy and relevance of features.

Weakly supervised learning generates models from data with consistency issues, such as faulty, incomplete, or inaccurate marks. Improving data quality can improve the performance of weakly supervised learning models. Noisy crowdsourced annotations and mark errors may also be issues with info. These can hamper performance of the model as well.

**Techniques during model building:** Building the model is the central stage for a successful machine learning application. Developing visual analytics methods to facilitate model building is also a growing research direction in visualization which described in the survey such as Model understanding, diagnosis, or steering. Model comprehension approaches attempt to visually illustrate how a model works, such as how changes in parameters affect the model and why the model produces a particular output for a given input. Model diagnosis approaches use immersive exploration of the training process to diagnose errors in model training. Model steering methods are mainly aimed at interactively improving model performance. To begin with, model developers must have a thorough understanding of models in order to free them from the time-consuming process of trial and error. It entails comprehending the effects of parameters as well as model actions. When the training process fails or the model does not work as anticipated, model developers must diagnose the problems that occurred during the training process. Finally, since too much time is expended during the model building process improving model performance, there is a need to assist in model steering. In addition to these demands, researchers have developed a number of visual analytics methods to enhance model recognition, diagnosis, and steering.

**Techniques after model building**: Following model construction, existing visual analytics projects seek to assist users in understanding and gaining knowledge from model outputs, such as high-dimensional data analysis. The corresponding methods were categorized according to the form of data analyzed since these methods are often data-driven. In graphic design, data's temporal property is crucial. As a consequence, approaches were divided into two categories: those that comprehend static data analysis results and those that comprehend dynamic data analysis results.

When analyzing static data analysis results, a visual analytics framework typically treats all model output as a broad array and analyzes the static structure. For complex data, the system focuses on demonstrating the progression of data over time, which is learned through the simulation model, in addition to understanding the analysis results at each time point.

It's important to explore and examine how latent themes in data shift over time, in addition to recognizing the implications of static data analysis. For example, a framework that offers an analysis of major public views on social media and how they evolve over time may assist policymakers in making timely decisions. The majority of current research focuses on deciphering the study findings of a data corpus in which each data object is linked to a time stamp. Current research on visual dynamic data processing may be classified as offline or online depending on whether the device supports the analysis of streaming data.

All data is accessible before processing in offline analysis, while streaming data is incoming during the analysis phase in online analysis. Subject analysis, case analysis, and trajectory analysis are three types of offline analysis study that can be categorized based on the analysis mission. Online analysis is especially necessary for streaming data, such as text streams. As a pioneering work for online analysis of text streams.

**Conclusion**

The paper proposed six research directions for future machine learning-related visual analytics research, including improving data quality for weakly supervised learning and explainable function engineering before model building, online testing diagnosis and intelligent model refining during model building, and multi modal data interpretation and idea drift analysis. The main goal of this paper was to identify data visualization and machine learning techniques based on the three schemes (Before, during and after model building). In this report, these three schemes have been summarized and critically analyzed based on the key findings from different literatures, also a few recommendations are presented to overcome a few challenges and to show promising future potentials. By critically analyzing the survey paper, recent progress and developments in visual analytics techniques for machine learning can be determined for future potential research opportunities.

**References**

[1] J. Yuan, C. Chen, W. Yang M. Liu, J. Xia and S. Liu, “A survey of visual analytics techniques for machine learning,” in *Comp. Visual Media* 7, 3–36 (2021), Nov. 2020, doi: 10.1007/s41095-020-0191-7