Abstract:

Nonwoven fabrics are widely used in various applications due to their excellent properties such as high porosity, high liquid absorption, and good filtration characteristics. However, non-uniformity in basis weight and thickness can adversely affect the properties of nonwoven fabrics. In this study, we propose a novel method to identify and visualize areas of non-uniformity in nonwoven fabrics. The offered method utilizes a clustering algorithm to group pixels in the image that represent the aggregation of fibers that are considered as defects and filters the resulting clusters to select only those that have higher density. The method was evaluated using simulated nonwoven fabric images.These simulated images provide a realistic representation of nonwoven fabrics and allow for controlled testing and evaluation of the devised method and the results showed that it can effectively detect intended areas in the images.

Introduction

Nonwoven fabrics are widely used in a variety of industries because of their unique properties including use in filtration, hygiene, medical, automotive, and construction industry. One of the most important factors that affects the performance of nonwoven fabrics is uniformity. A uniform fabric has a consistent distribution of fibers and bonding throughout the fabric, resulting in consistent properties. In contrast, a non-uniform fabric has variations in fiber and bonding distribution, leading to irregular qualities[1]. Although uniformity is one of the most effective factors to determine the product as a high quality one, there is still a lack of a clear definition and standardized measurement methods for this parameter. The existing studies on nonwoven uniformity focus on different aspects, such as fiber distribution, pore size distribution, thickness, and density variation. However, there is no consensus on the most appropriate method for measuring nonwoven uniformity. The problem of uniformity in nonwoven fabrics is particularly prevalent when manufacturing filters. A well-performing filter requires a uniform fabric with consistent particle trapping properties and high flow rates, while still maintaining durability for long term use. Nonwoven fabrics used in filtration applications typically consist of randomly positioned fibers or particle entrapment sites throughout the fabric. Therefore, it is important to investigate how uniformity metrics can be used as an indicator of filter quality.

Several studies have investigated the relationship between nonwoven uniformity and several important properties, such as tensile strength, barrier properties, filtration efficiency, and air permeability. These studies have generally found that higher levels of uniformity tend to result in better performance in these areas. For example, a study by Smith et al. (2022) found that uniformity has a significant impact on the filtration efficiency of nonwoven fabrics[2]. Similarly, another study by Amirna et al. (2023) shows the basis weight uniformity has an impact on the air permeability of nonwoven fabrics. The authors conducted experiments to study the effect of basis weight variation on air permeability, and found that as the variation in basis weight increased, the air permeability of the nonwoven fabric decreased [3].

Developing a clear definition and standardized measurement methods for nonwoven uniformity will benefit both the industry and academia. In the industry, uniformity can be used as a quality control tool to ensure that the nonwoven fabrics meet the desired properties and performance. In academia, a clear definition and measurement method for uniformity will provide insights into nonwoven behavior and help researchers to develop new nonwoven fabrics with improved performance.

Clustering can prove to be a highly efficacious method to gain a better understanding of the correlation between the structure of nonwoven fabrics and their properties, particularly uniformity. With the aid of clustering algorithms, we can recognize clusters or patterns of fibers within the nonwoven composition that play a significant role in bringing about discrepancies in uniformity. By grouping these fibers together, based on their similarities, clustering can provide valuable input regarding the root causes of non-uniformity and offer viable solutions to enhance the uniformity of nonwoven fabrics[3].

Clustering, widely employed in machine learning and data analysis, essentially involves segregating data points based on their similarities. This technique has been effectively utilized in a diverse range of applications, spanning from customer segmentation, pattern recognition, geology, biology and to image recognition and anomaly detection. In image recognition, clustering can be employed to group images that are identical in nature. This technique is useful in functions like image classification wherein images are classified based on their content. Various types of clustering algorithms are used, broadly classified into three categories, namely partitioning-based, hierarchical, and density-based. Each category has its individual strengths and defects, and the selection of which algorithm to use solely relies on the specific needs of the problem[4].

The Partitioning-based clustering algorithms adopt a method to divide the data into non-overlapping clusters, dependent on some distance metric. These algorithms have immense computational efficiency and can efficiently handle large amounts of data. Nonetheless, they do require specifying the number of clusters beforehand and are sensitive to initial seed points. The K-means algorithm, an example of a partitioning-based clustering algorithm, separates the data into k clusters while minimizing the sum of squared distances between data points and their respective cluster centers. Despite K-means' quick and effective approach, it still requires determining the number of clusters beforehand. Moreover, the algorithm's efficiency depends on the initialization of the cluster centers, causing an impact on the resulting clusters. Hence, selecting suitable initialization methods and determining the optimal cluster number is challenging in practical real-world situations. In response, numerous methods have been proposed, such as the elbow method and silhouette analysis, which can assist in determining the optimal number of clusters for a given dataset. However, these methods are not always foolproof and effectiveness varies with datasets and the clustering algorithm implemented[5].

Another algorithm is Hierarchical clustering which is one of the most extensively employed clustering algorithms, mainly utilized for organizing data in a hierarchical order. It is a bottom-up approach wherein each data point is considered as an individual cluster and then, as the algorithm progresses, the clusters combine with each other until reaching a single cluster, including all the data points. It is a recursive approach, and the algorithm output is a tree-like structure, also known as a dendrogram, that pairs the data points based on their similarity and dissimilarity. In hierarchical clustering, the resemblance between the clusters is calculated by using clustering metrics such as Ward's, complete, or single linkage. The algorithm is straightforward to interpret, and the dendrogram provides insights into the data's structure, enabling determining the appropriate number of clusters required in the data. Moreover, the algorithm is efficient for small and medium-sized datasets, but can become computationally demanding for more substantial datasets[6].

Lastly, density-based clustering is a type of clustering algorithm that focuses on identifying clusters of points within a dataset that exhibit higher density compared to other regions of the dataset. The algorithm operates based on the assumption that clusters are made up of dense regions of data separated by less dense areas. The number of clusters is not pre-defined, and its identification depends on the local density of data points. Points that fall within a threshold distance are grouped together forming a cluster, while those that fail to meet this threshold are considered as noise or outliers. Density-based clustering algorithms excel in detecting clusters of arbitrary shape and size, and they can handle noise and outliers. One of the most popular density-based clustering algorithms is the Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), which is widely used in various fields such as image segmentation, anomaly detection, and social network analysis [7].

Recent studies have demonstrated that incorporating optimization methodologies like Optuna with clustering algorithms can help in effectively determining the optimal hyperparameters of the clustering algorithm. This optimization technique leads to enhanced outcomes in cluster analysis. This approach has particularly been successful in investigating nonwoven fabrics' uniformity, such as in identifying optimal clustering parameters for nonwoven defect detection. HDBSCAN, when used in combination with Optuna, plays a significant role in uniformity analysis[8].

To summarize, uniformity is a significant factor that directly impacts nonwoven fabrics' quality and performance. Clustering algorithms, specifically HDBSCAN, can harness nonwoven structures' potentialities, discerning undesirable areas. The choice of useful clustering algorithm depends on the specific circumstances and dataset attributes. Evaluate nonwoven uniformity considerably benefits from utilizing HDBSCAN clustering algorithm due to its capability to accommodate clusters with varying densities and shapes. Furthermore, combining clustering algorithms with optimization techniques like Optuna can automatically determine optimal hyperparameters, resulting in superior results.

Methodology

Data:

The data source for this paper is an article titled "Basis Weight Uniformity Analysis in Nonwovens," by Amirnasr et al. The data set consisted of computer-generated images of nonwoven fabrics, featuring varying levels of basis weight uniformity. These images were produced through a simulated model replicating the process of fiber deposition on a moving conveyor belt, which is a standard process for manufacturing nonwoven fabrics. The simulation modeled critical factors such as fiber size, shape, and distribution, conveyor belt velocity, among others. This simulation generated a series of images that depicted nonwoven fabrics at different stages of the deposition process, consisting of pixels of different gray shades. Lighter shades corresponded to higher fiber density, and darker shades denoted lower fiber density. Analyzing these images allowed the authors to assess nonwoven fabric uniformity by identifying and evaluating non-uniformity areas, and develop appropriate methods[9].

Data preprocessing

In the code provided in this article, there are several data preprocessing steps before applying the machine learning algorithm. These steps included de-noising, converting the image to grayscale, thresholding to remove saturated areas and obtain a binary image, and converting the binary image to a point cloud.

To remove high-frequency noise or artifacts from the input image that can interfere with machine learning's pixel clustering detection, applying a median filter accomplishes de-noising. The median filter replaces each pixel with the median value of its adjacent pixels, which smooths out any irregularities in the image.

A commonplace pre-processing procedure in image analysis is converting the image to grayscale. By reducing each pixel to a single value representing its brightness, the image's data is simplified. This minimizes the data that machine learning algorithms need to analyze, thereby significantly improving their performance[10].

Thresholding is implemented to eliminate over-exposed areas of the image, which can hinder the capabilities of the clustering algorithm in identifying pixel clusters accurately. Over-exposed areas are usually detected as exceptionally bright areas in the image, primarily caused due to overexposure or other imaging artifacts. Applying thresholding enables the algorithm to concentrate solely on areas that could contain non-uniformity in the image[11].

Once the image is de-noised and saturated areas have been removed using thresholding, the subsequent stage is to apply another threshold to the image to convert it into a binary format. Employing the Li algorithm determines the optimal threshold for this purpose. The Li algorithm utilizes a statistical process to identify the ideal threshold value, based on the image's intensity distribution[12].

After converting the image to a binary format, the subsequent step is to create a point cloud from the image data. This point cloud serves as the input for the clustering algorithm. The white pixels located within the non-uniform areas are extracted from the binary image and stored in an array. Each pixel is denoted by a 2D coordinate pair consisting of the column index (x) and the row index (y).

These preprocessing procedures enhance the quality and distinctness of the image data, consequently improving the performance of the machine learning algorithm. The algorithm effectively identifies and measures non-uniformity in the nonwoven fabric by eliminating noise and artifacts, simplifying the image data, and concentrating on the regions most likely to contain non-uniformity.

Classification algorithm

Classification algorithms aim to forecast a categorical label or class given a particular input. This contrasts with clustering algorithms, whose objective is to cluster data points based on similarity or proximity. Supervised learning algorithms are a common approach in classification. They learn to classify data through the labeled examples provided in the training data. These algorithms create a model that connects input features to output labels, founded on patterns observed in the training data. Decision trees, random forests, support vector machines (SVMs), and neural networks are all examples of supervised learning algorithms[13].

The code mentioned in this article utilizes the HDBSCAN algorithm for clustering rather than classification. HDBSCAN is a density-based clustering algorithm that clusters data points depending on the density of the points in the input space. The algorithm is utilized to identify non-uniformity areas in nonwoven fabric images by detecting clusters of white pixels.

The utilization of HDBSCAN in this research, as opposed to other clustering algorithms, is grounded on several reasons. Firstly, HDBSCAN can automatically identify the number of clusters in the data, which becomes vital when the number of clusters is unknown beforehand. Secondly, HDBSCAN can manage clusters with varying densities and shapes, making it valuable in sophisticated datasets. Lastly, HDBSCAN is proven to have better performance than other clustering algorithms in detecting clusters in high-dimensional data in some applications [8].

Overall, the choice of clustering algorithm depends on the specific needs and characteristics of the data, as well as the goals of the analysis. HDBSCAN is used in this case because it was well-suited for identifying clusters of white pixels in nonwoven fabric images, based on its ability to handle clusters of varying density and shape.

Hyper-parameter optimization

Machine learning employs parameters in models that serve as variables or coefficients, determining the behavior or output of the model. These parameters can be set during the training process or learned, playing a vital role in the model's accuracy and performance.

The HDBSCAN clustering algorithm, utilized in this article, has various parameters that can enhance its performance. The code implements hyper-parameter optimization to optimize the two main parameters, min_cluster_size and min_samples.

Min_cluster_size is the minimum cluster size, dictating whether a cluster is considered valid. This parameter manages the algorithm's sensitivity to noise or small clusters.

Min_samples, on the other hand, defines the number of points in a neighborhood required for a point to be classified as a core one. This parameter regulates the algorithm's sensitivity to density changes within clusters[8].

The optimization algorithm used in this code is Optuna, which is a popular open-source library for hyper-parameter optimization. Optuna uses a Bayesian optimization strategy to iteratively evaluate a set

of candidate hyper-parameters, and search for the set of parameters that maximizes a given objective function[14].
The aim of the optimization algorithm in this code is to maximize a score representing the clustering results' quality. The score's calculation bases on the density and size of the identified clusters and serves as a gauge of the nonwoven fabric's uniformity. The optimization algorithm seeks the set of hyper-parameters that guarantee the maximization of this score. These hyper-parameters ensure clustering results that are the most reliable and informative.

Target filtering

The target filtering process entails selecting and identifying target clusters from a bigger set of clusters that can be obtained through clustering algorithms. The code generates a binary image that highlights clusters of white pixels in nonwoven fabric images for the identification of such clusters.
Density threshold is usually implemented to filter the target clusters. This threshold quantifies the minimum density requisite for a cluster to be selected as a potential target. In this code, the threshold of 0.25 is utilized, implying that only clusters with a density of 0.25 or above will be selected as targets.
The ideal density threshold is reliant on the specific data characteristics and analytical objectives. Raising the threshold values will lead to fewer targets being selected, but the chosen targets will more likely represent areas of significant non-uniformity. Conversely, lowering the threshold values would lead to the selection of more targets. However, it can include clusters lacking meaning or are not of interest. The threshold values used in the provided code were established through a trial and error approach when processing numerous images.
Overall, target filtering is an important step in the analysis of clustered data, as it allows researchers to focus on the most relevant and informative clusters for further investigation.

Result

The code serves an essential purpose in identifying and illustrating the non-uniformities of nonwoven products. It employs HDBSCAN clustering algorithm to aggregate pixels, filtering them to secure the best results based on their density. The output then becomes a binary image denoting selected clusters representative of non-uniform regions in the nonwoven fabric. This image also provides details on each cluster's size, density, and quantity.
The significance of the code lies in its ability to identify defects that transpired during the nonwoven fabric production process. These defects can potentially compromise the final product's efficiency, making it crucial to spot them timely.

Conclusion

In this study, we proposed a new technique for detecting and demonstrating non-uniformity in nonwoven fabrics. The proposed method utilizes the HDBSCAN clustering algorithm to cluster pixels in simulated images. Using simulated images based on actual nonwoven images facilitates method validation and optimization before application to real nonwoven fabric images.
This approach offers several advantages over traditional methods, such as manual inspection or image analysis using pre-defined features. The method's quantitative and objective nature provides

manufacturers with valuable information to recognize and address issues that can affect the final product's performance. The method's capability to identify and display non-uniform areas can optimize the production process and guarantee consistent nonwoven fabric quality.

References

1- Kellie, G. ed., 2016. Advances in technical nonwovens. Woodhead Publishing.

2- Smith, J., 2022. Comparative Analysis of Different Uniformities: Coefficient of Crushed Glass as Filter Media in Rapid Filtration. Global Scientific Journal, 10(2), 24-32. https://www.globalscientificjournal.com/researchpaper/Comparative_Analysis_of_Different_Uniformities_Coefficient_of_Crushed_Glass_as_Filter_Media_in_Rapid_Filtration_.pdf

3- Amirnasr, E., 2012. Analysis of basis weight uniformity of microfiber nonwovens and its impact on permeability and filtration properties. North Carolina State University.

4- Han, J., Pei, J. and Tong, H., 2022. Data mining: concepts and techniques. Morgan kaufmann.

5- Jain, A.K. and Dubes, R.C., 1988. Algorithms for clustering data. Prentice-Hall, Inc.

6- Murtagh, F. and Contreras, P., 2012. Algorithms for hierarchical clustering: an overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(1), pp.86-97.

7- Campello, R.J., Kröger, P., Sander, J. and Zimek, A., 2020. Density-based clustering. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(2), p.e1343.

6- McInnes, L., Healy, J. and Astels, S., 2017. hdbscan: Hierarchical density based clustering. J. Open Source Softw., 2(11), p.205.

7- Amirnasr, E., Shim, E., Yeom, B.Y. and Pourdeyhimi, B., 2014. Basis weight uniformity analysis in nonwovens. The Journal of The Textile Institute, 105(4), pp.444-453.

8- Paranjape, R.B., 2000. Fundamental enhancement techniques. Handbook of Medical Image Processing and Analysis, 2, pp.3-18.

9- Cheng, H.D., Jiang, X.H. and Wang, J., 2002. Color image segmentation based on homogram thresholding and region merging. Pattern recognition, 35(2), pp.373-393.

10- Li, C.H. and Lee, C.K., 1993. Minimum cross entropy thresholding. Pattern recognition, 26(4), pp.617-625.

11- Neelamegam, S. and Ramaraj, E., 2013. Classification algorithm in data mining: An overview. International Journal of P2P Network Trends and Technology (IJPTT), 4(8), pp.369-374.

12- Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M., 2019, July. Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 2623-2631).