MEDS Clinical Medicine (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/medsc.2024.050104
ISSN 2616-1907 Vol. 5 Num. 1

Research on the data acquisition and processing technology of spectral flow cytometry

Li Rui

Focused Photonics (Hangzhou) Inc, Hangzhou, Zhejiang, China

Keywords: Spectral flow cytometry, data acquisition, data processing, peak identification, data cleaning

Abstract: Spectral flow cytometry is an important biomedical instrument to be used for the quantification and analysis of cell surface markers. This paper studies the data acquisition and processing technology of spectral flow cytometry to improve the data quality and analysis effect. First, the principle and application fields of spectroscopic flow cytometry are presented. Then, key issues in the data acquisition process, including laser power regulation, instrument calibration, and sample preparation, are discussed. Then, common data processing methods, including peak identification, data cleaning and normalization. Finally, the effectiveness and feasibility of the proposed data acquisition and processing techniques are verified by experiments. This study provides useful guidance for the accurate acquisition and reliable analysis of spectral flow cytometric data.

1. Introduction

Spectral flow cytometry is a very widely used technique in biomedical research and clinical diagnostics that enables high-throughput detection and analysis of cell surface markers. With the continuous development of technology, the data acquisition and processing of spectroscopic flow cytometer has become a key link of research. Accurate acquisition and processing of flow cytometry data is essential to obtain reliable results and accurate analysis. However, there are some challenges and problems in data acquisition and processing due to instrument complexity and data diversity. Therefore, it is necessary to conduct in-depth research on spectroscopic flow cytometry data acquisition and processing techniques to improve the data quality and analysis results.

2. Data acquisition technology

2.1. Operating principle of the spectral flow cytometry

Spectral flow cytometry is a widely used instrument in biology, medicine and life science research, which enables the detection and analysis of cells and microparticle features in real time. The working principle of the spectroscopic flow cytometer is based on the interaction of the laser and the optical system. First, the laser emits a high-energy, monochromatic, one-way laser beam. The laser is focused through a series of lenses and lenses, and then passes through a grating, further dispersed into spectra of different wavelengths. Cells or microparticles in the sample to be tested were injected into a flow cytometer and passed through a fine flow tube under irradiation with a laser beam. When the samples to be tested pass through the laser beam, they interact with the laser, and scattering and absorption phenomena occur. The light signals generated by these interactions were collected and analyzed. Spectral flow cytometry usually uses two main optical signal detection methods: forward scatter and side scatter. Forward scatter is the scattered light signal of the cells or particles in the sample to the forward propagation direction of the laser beam. Lateral scattering is the scattered light signal of the cells or particles in the sample in the lateral propagation direction of the laser beam. In addition, spectroscopic flow cytometry can also use fluorescent labeling techniques to detect specific molecules or cellular structures in the samples. In this case, the sample is labeled with fluorescent dyes or fluorescent markers. When the laser beam is illuminated to a labeled sample, the fluorescent marker emits fluorescent light signals at a specific wavelength, which is collected and used for analysis.

2.2. Selection of data acquisition parameters

The wavelength and power of the laser are important parameters for data acquisition. The wavelengths should be chosen based on the excitation wavelength and fluorescence emission wavelength of the fluorescent dye in the sample to be tested. The choice of power should take into account the absorption and scattering of the laser to avoid sample damage and background signal interference. Forward and side scatter are commonly used scattering signals. The forward scatter signal reflects the size and shape characteristics of the cells or particles in the sample to be tested, while the side scatter signal reflects the complexity and granularity of the cells or particles. By selecting the appropriate scattering signal detector and scattering angle, the accurate scattering signal information can be obtained. If a fluorescent signal is required to be detected, the appropriate fluorescent dye and the corresponding wavelength detector need to be selected. The selection of fluorescent dyes should be based on the specificity and spectral properties of the molecules to be tested. Meanwhile, care is needed to avoid the stacking and cross-excitation of fluorescence signals. The sampling rate determines the speed of data acquisition, while the number of events determines the number of samples collected. Choosing the appropriate sampling rate and number of events requires balancing the quality of the data and the time of acquisition. Higher sampling rate and number of events can improve the statistical significance of the data, but also increase the complexity of data volume and processing complexity.

2.3. Optimization of the data acquisition process

Sample preparation is the prerequisite for data acquisition, and reasonable sample preparation and processing procedures can reduce sample variability and interference. This includes the steps of cell fixation, staining, washing, etc. Optimizing the sample preparation process can improve the sample consistency and reduce the background noise and non-specific signals. The setting of the flow rate is critical for the stability and accuracy of the data acquisition. Too high or too low a flow rate can lead to a distortion of the data. Too high flow rate may lead to too short residence time of cells or particles in the detection area and weakened signal. Too low flow rate may lead to aggregation and blockage of cells or microparticles, affecting the accuracy of the data. Therefore, the appropriate flow rate is selected according to the characteristics of the sample and the requirements of the instrument. Laser power is selected to balance the damage and the signal intensity of the sample. Excessive laser power may lead to the rupture or damage of the cells, affecting the accuracy of the data. Too low laser power may lead to weakening of the signal to detect features of cells. Therefore, experimental optimization is needed to determine the laser power range for the sample. Regular calibration of the instrument is an important step to ensure the accuracy of the data. Calibration corrects the drift and nonlinear response of the instrument. In addition, using standards for verification and quality control is also important to optimize the data acquisition process. Standards can be used for quantitative analysis, instrument performance assessment, and results verification.

3. Data preprocessing technology

3.1. Data correction and normalization

Data preprocessing is the process of processing and transforming the raw data prior to the data analysis. Among them, data correction and standardization are commonly used data preprocessing techniques used to eliminate noise, fluctuation and bias in data to improve the reliability and comparability of data. Data correction refers to the correction and adjustment of the raw data to eliminate the influence of factors such as instrument drift, background noise and systematic error on the data. Common data correction methods include zero-point correction and linear correction. Zero correction corrects the zero bias in the measurement results by measuring the background signal or blank samples without samples to ensure the accuracy and reliability of the data. Linear correction was performed by using a standard sample of known concentration, constructing a calibration curve between the sample concentration and the measured signal, and the measurements interpolated or extrapated to obtain an accurate sample concentration. Data standardization is the transformation of data at different scales and ranges into a uniform standard scale for easy comparison and analysis. Common methods of data normalization include max-minimum normalization, Z-score normalization, and decimal calibration normalization. Maximum-minimum normalization transforms the data linearly into the range of [0,1] by subtracting the minimum value and dividing by the difference between the maximum and the minimum value for each data point. Z-score normalization transforms the data into a standard normal distribution, such that the data has a mean of 0 and a standard deviation of 1, by subtracting the mean from each data point and dividing by the standard deviation. Decimal calibration normalization linearly transforms the data to the range of [-1,1] or [0,1], by dividing each data point by a fixed base, such as the maximum absolute value of the data or the range of the data.

3.2. Noise filtering and smoothing processing

Noise filtering and smoothing are common data pre-processing techniques used to reduce noise interference, fluctuations and outliers in data to improve data quality and reliability. Noise is a useless information or interference signal in the data, which may come from the error of the instrument, environmental interference, or random fluctuations during the data acquisition process. The goal of noise filtering is to extract the true signal from the raw data and reduce the effect of noise. Common noise filtering methods include moving average, median filtering, and Gaussian filtering. Moving averaging is a basic noise filtering method that smoothing the data by calculating the average of the data points within the sliding window. It is able to effectively smooth the high-frequency noise in the data, but may lead to a reduced smoothness of the signal. Median filtering is a nonlinear noise filtering method that replaces the current data point by calculating the median value of the data point in the sliding window. It is able to effectively remove isolated outliers and pulse noise, but may result in a relatively low smoothness of the signal. Gaussian filtering is a linear filtering method based on a Gaussian function to smooth the data by calculating a weighted average of the data points within the sliding window. It is able to effectively reduce the noise while retaining the signal features. Gaussian filtering works best when dealing with noise from a Gaussian distribution. Smoothness is a method to extracting signal trends by reducing fluctuations and noise in the data. In addition to noise filtering methods, there are other common smoothing methods, such as polynomial fitting, wavelet transform, and Loess smoothing. The polynomial fit estimates the trend of the signal by fitting a polynomial function of the data points. By selecting the appropriate polynomial order, the data can be smoothed and the trend information preserved.

3.3. Detection and correction of outlier values

Outlier value detection and correction is an important step in data preprocessing, which is used to identify and process abnormalities in the data or values that do not conform to the normal rules. Outliers may be caused by measurement error, data entry error, equipment failure, or rare events. These outliers may have adverse effects on the results of data analysis and modeling, thus requiring efficient abnormal value detection and correction. Commonly used outlier detection methods include simple statistical methods, distance-based methods, model-based methods, and machine learning-based methods. Simple statistical methods can use standard deviations or boxplots to identify outliers that deviate significantly from other values. The distance-based approach identifies outliers far from other points by calculating the distance between the data points and their nearest neighbor data points. Model-based methods utilize probabilistic models or patterns of model learning data to detect outliers. Machine learning-based methods then use supervised or unsupervised learning algorithms to identify outliers. For correction of outliers, common methods include removing outliers, replacing outliers, smoothing and segmentation. Removing outliers is the simplest correction method, but may bias the data and loss of information. Alstitution outliers can be used in place of outliers, such as mean, median, or interpolation methods. Smoothness reduces the effect of outliers on the data by applying a smoothing method, bringing it closer to the trend of the surrounding data points. Segmentation processing segmented the data to analyze and correct the data within each segment to deal with the outliers more precisely[1].

4. Data analysis techniques

4.1. Data visualization and exploratory analysis

Data visualization and exploratory analysis are important technologies in data analysis, aiming to present and understand data through graphs, graphics, and visualization tools. It can help us reveal the patterns, trends and association relationships in the data, find the outliers and potential problems in the data, and provide intuitive ways to convey the information of the data. Data visualization is the process of transforming data into visual forms, using a variety of charts and graph types, such as line charts, bar charts, scatter plots, pie charts, heat maps, etc. Choosing the appropriate chart type and visualization method can visually show the data distribution, change trends and relative comparison. For example, line plots can be used to show changing trends in time series data, and scatter plots can be used to show the association relationship between two variables. Exploratory analysis is conducted on the basis of data visualization, extracting meaningful information and patterns by observing and analyzing the characteristics, distribution, and relationships of the data. Exploratory analysis usually involves the calculation of statistical indicators, grouping and clustering of data, and correlation analysis. These analytical methods can help us to discover patterns and trends in the data, explore the relationships between variables, and provide initial insights and hypotheses. The advantage of data visualization and exploratory analysis lies in its ability to transform complex data into intuitive and easy-to-understand forms, helping us to better understand the data and discover potential problems and opportunities. It can help us identify outliers in the data, check the data integrity and consistency, verify hypotheses and generate new research questions. Furthermore, data visualization also facilitates sharing data results with others, facilitating the process of communication and decision making. However, it should be noted that data visualization and exploratory analysis are only the first step in data analysis, and more in-depth analysis and modeling often require further data processing and statistical methods. Therefore, in addition to visualization and exploratory analysis, other analytical techniques and methods are needed to fully understand the data and draw accurate conclusions[2].

4.2. Data Dimension reduction and feature extraction

Data dimension reduction and feature extraction are important techniques in data analysis, aiming to reduce data dimensions and extract key features to better understand data, reduce computational complexity and improve model performance. Data dimension reduction is mainly for the case of high-dimensional data, with more features. High-dimensional data can lead to dimensional disasters and computational difficulties, and may also contain redundant or unrelated features. The goal of data dimensionality reduction is to reduce the number of features by converting high-dimensional data into low-dimensional representations by preserving important information about the data. Common data dimension reduction methods include principal component analysis (PCA), linear discriminant analysis (LDA) and factor analysis. These methods project the data onto a new low-dimensional space by linear or non-linear transformations to retain the largest data variance or the largest category difference. Feature extraction is the process of transforming raw data into new features that are more informative and expressive. The goal of feature extraction is to find the features that best describe the data in order to better distinguish between different categories or patterns. Feature extraction can be achieved by various methods, including statistical methods, information theory methods, model-based methods, and machine learning methods. Common feature extraction methods include statistical feature extraction (such as mean, standard deviation), frequency domain feature extraction (such as Fourier transform), wavelet transform, and deep learning methods (such as convolutional neural network). The advantages of data dimension reduction and feature extraction are to reduce data complexity, reduce computational cost, remove redundant and irrelevant features, and extract more informative features. This helps to improve the effect of data analysis and modeling, and improve the performance and generalization ability of the models. Moreover, dimension reduction and feature extraction can also help us understand key factors and patterns in the data to support decision making and problem solving[3] .

4.3. Data classification and cluster analysis

Data dimension reduction and feature extraction are important techniques in data analysis, aiming to reduce data dimensions and extract key features to better understand data, reduce computational complexity and improve model performance. Data dimension reduction is mainly for the case of high-dimensional data, with more features. High-dimensional data can lead to dimensional disasters and computational difficulties, and may also contain redundant or unrelated features. The goal of data dimensionality reduction is to reduce the number of features by converting high-dimensional data into low-dimensional representations by preserving important information about the data. Common data dimension reduction methods include principal component analysis (PCA), linear discriminant analysis (LDA) and factor analysis. These methods project the data onto a new low-dimensional space by linear or non-linear transformations to retain the largest data variance or the largest category difference. Feature extraction is the process of transforming raw data into new features that are more informative and expressive. The goal of feature extraction is to find the features that best describe the data in order to better distinguish between different categories or patterns. Feature extraction can be achieved by various methods, including statistical methods, information theory methods, model-based methods, and machine learning methods. Common feature extraction methods include statistical feature extraction (such as mean, standard deviation), frequency domain feature extraction (such as Fourier transform), wavelet transform, and deep learning methods (such as convolutional neural network). The advantages of data dimension reduction and feature extraction are to reduce data complexity, reduce computational cost, remove redundant and irrelevant features, and extract more informative features. This helps to improve the effect of data analysis and modeling, and improve the performance and generalization ability of the models. Moreover, dimension reduction and feature extraction can also help us understand key factors and patterns in the data to support decision making and problem solving.

5. Conclusion

In this study, a series of effective methods and strategies were proposed by systematically studying the spectral flow-cytometry data acquisition and processing techniques. For data acquisition, we optimized the laser power regulation and instrument calibration methods to ensure the stability and accuracy of the spectral flow cytometer during acquisition. For data processing, we propose peak identification algorithms, data cleaning, and normalization methods to improve the reliability and comparability of the data. Through experimental validation, we demonstrate the effectiveness and feasibility of the proposed method in the acquisition and processing of data for spectral flow cytometry. The results of this study are very instructive for the accurate acquisition and reliable analysis of spectroscopic flow cytometry data, and can be widely used in biomedical research and clinical diagnosis.

Acknowledgements

Development of an intelligent fully automatic medical flow cytometer, Hangzhou Puyu Technology Development Co., Ltd, 2022YFC2406300.

References

[1] Tan Guozhen. Research on key technologies of single laser multiccolor flow cytometry [D]. Changchun University of Science and Technology, 2023.
[2] Li Xing. Spectral flow cytometry data acquisition and processing technology study [D]. Xiamen University, 2021.
[3] Fang Xiaoling. Development of a spectral flow cytometry fluorescence acquisition system [D]. Xiamen University, 2018.