공학석사학위논문

Image-based statistical learning of liquid pinch-off dynamics

유체의 얇아지는 거동에 대한 이미지 기반의 통계적 학습 분석

2023년 2월

서울대학교 대학원
화학생물공학부
장 준 형

공학석사학위논문

Image-based statistical learning of liquid pinch-off dynamics

유체의 얇아지는 거동에 대한 이미지 기반의 통계적 학습 분석

2023년 2월

서울대학교 대학원
화학생물공학부
장 준 형

Image-based statistical learning of liquid pinch-off dynamics

유체의 얇아지는 거동에 대한 이미지 기반의 통계적 학습 분석

지도교수 남 재 욱

이 논문을 공학석사 학위논문으로 제출함

2022년 12월

서울대학교 대학원

화학생물공학부

장 준 형

장준형의 공학석사 학위 논문을 인준함

2022년 12월

위 원 장	안경현	(인)
부위원장	남재욱	(인)
위 원	이원보	(인)

Abstract

Image-based statistical learning of liquid pinch-off dynamics

Junhyeong Jang
School of Chemical and Biological Engineering
The Graduate School
Seoul National University

Functional materials used in many industrial processes are made of various raw materials such as polymers, metals, and carbon particles to improve their functions. As a result, they show various rheological behaviors such as viscoelasticity and yield behavior. Frequently, such behaviors can be identified in the pinch-off dynamics of the fluids. The rheological behaviors of fluids appear visually characteristic in the thinning behaviors. So it is important to analyze the thinning behaviors of fluids visually.

This thesis describes research on analyzing the overall thinning behaviors of fluids using Machine Learning methods. First, fluid thinning behavior images were acquired using DoS-CaBER(Dripping onto Substrate - Capillary Break-up Extensional Rheometer). Then, image pre-processing methods were developed for effective Machine Learning. PCA(Principal Component Analysis) was used as a machine learning algorithm to obtain new rheological information from the images. As a result, PCs(Principal Components) and Eigen-thinning, which explain the fluid thinning behavior, were obtained. The accuracy of PCs was verified through a PCA-based classification model, and the rheological complexity of PCs was confirmed through a comparison between conventional rheological properties and PCs.

Also, we did application tests. The PCs were used for two prediction tests. The first test was to predict the flow ratio of fluids in complex micro-channels. It showed that PCs are useful predictive indicators when simple, fast, and rough predictions are required. The second test was to predict the mixing ratio of the mixture. The feasibility of the predictive model was confirmed when the rheological property dominant in the PCs had a linear change according to the mixing ratio. Finally, the advantages, limitations, and future development directions of the proposed machine learning method were explained.

keywords: Fluid thinning, Liquid pinch-off dynamics, Machine learning, PCA, DoS-CaBER, Rheological property

student number: 2021-26394

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Many materials used in the industrial process of producing electronics have a variety of rheological behaviors. The rheological behaviors of the materials have a great influence on process conditions and product quality. Therefore, it is important to understand the rheological properties of materials. Overall, in past studies, the rheological behaviors of materials have been mainly studied based on shear properties. However, depending on the material and process, not only the shear behaviors but also the extensional behaviors have a lot of influence. Extensional viscosity is a representative property explaining the extensional behavior. The extensional viscosity is relatively difficult to measure compared to the shear viscosity. DoS-CaBER(Dripping onto Substrate - Capillary Break-up Extensional Rheometer) is an instrument used for accurate extensional viscosity measurement [1].

In general, studies on extensional behavior using DoS-CaBER have been conducted based on rheological theories by users [2, 3, 4]. Or, they have been studied based on the user's visual information [5]. Previous studies have confirmed that various rheological behaviors are observed in the overall thinning behavior of fluids [6]. In this thesis, the overall thinning behaviors were analyzed using machine learning methods rather than the human-based methods used in previous studies. We used

PCA (Principal Component Analysis); a representative machine learning algorithm that extracts features(features are used interchangeably with variables here). PCA creates new variables by combining original variables, and the new variables are called PCs (Principal Components). PCA extracts PCs using a statistical method. The method has three major advantages: 1)Dimensional reduction, 2)Noise removal, and 3)Feature visualization. Because of these advantages, PCA is used in various ways in machine learning [7, 8, 9, 10, 11, 12, 13]. In Ch.2, We explained how to extract PCs using PCA for thinning behavior of several fluids. And we obtained the characteristic shape of thinning behavior through PCs, which was called Eigen-thinning. In Ch.3, the Validation of PCs with new variables were described. The accuracy of PCs was verified through a PCA-based classification model. In addition, the rheological complexity of PCs was confirmed by comparing the conventional rheological property values and score values of PCs. In Ch.4, two applications using PCs are described, and the advantages and limitations of utilizing PCs are described.

Chapter 2

Method

Figure2.1 is the overall flow of this study. The method of this study focused on 1) the Experiment part, and 2) the Machine learning method part.

2.1 Experimental method

In this study, we measured the rheological behavior of fluids using DoS-CaBER and a Rotational rheometer. Referring to the general measurement process of previous studies, extensional properties such as extensional viscosity and relaxation time were measured. Shear properties such as shear viscosity, storage modulus, and loss modulus were measured using a rotational rheometer [14]. In this study, a new method for analyzing rheological behavior was additionally implemented. We tried to analyze the thinning behavior of fluids using Machine Learning for thinning images measured using DoS-CaBER. Considering these analysis methods, we set up the DoS-CaBER measurement conditions and prepared the measurement fluids as follows.

Devices and Operating conditions

Figure 2.2 is a configuration of DoS-CaBER. The overall configuration of DoS-CaBER followed the configuration of Sharma group [1]. The measurement method is as follows. The test fluid is moved by the pump and a drop is formed at the end of the nozzle. The formed drop shows thinning behavior by capillary force from when it touches the substrate. Depending on the fluid, the thinning behavior occurs within a few seconds, so it is usually observed through a high-speed camera. Observed video images are stored, and we analyzed the stored thinning images through Machine Learning. The details of the main components in the measurement process are as follows. We used a PHD ULTRA Infusion Only model(Harvad Apparatus) as a pump and an N718 NDL 6 model(HAMILTON) as a needle. The thinning behavior of the fluid can also be affected by the state of the substrate. Therefore, in this study, the slide glass of MARIENFELD's HSU-1000412 model was consistently used. VEO710L 72G Mono Camera(Phantom) was used as a high-speed camera to record the thinning behavior, and an x5 magnification lens(EDMUNDOPTICS) was used as the lens. The following are the conditions for measurement. The infuse rate of the pump was set to 0.2mm/hr, and the operation of the pump stopped when the drop of the fluid made contact with the slide glass. The distance between the needle and the slide glass was set at $3D_i$, which is three times the inner diameter of the needle. The resolution of the high-speed camera was set to 960px * 600px, and the sample rate was set to 10000fps.

Materials

In this study, 14 fluids were used as shown in the table2.1. In this thesis, the index for each solution was set for the convenience of the notation of the solution. Based on the index in the table, F(1)~F(3) are solutions composed of a single material, and F(4)~F(14) are mixtures of the solutions. The CMC solution corresponding to F (1) was prepared as a 1wt% solution using M_w=250,000 g/mol CMC(Sigma-Aldrich) and DI water. The PEO solution corresponding to F(2) was prepared as a 1 wt% solution using M_v=2,000,000 g/mol PEO(Sigma-Aldrich) and DI water. Carbopol solution corresponding to F(3) was prepared as 0.14wt% solution using Carbomer941. F(4)$\sim$F(6) are mixtures of two solutions of F(1)$\sim$F(3), mixed at 1:1 based on mass ratio, and prepared by stirring at room temperature at 300 rpm for 24 hr. F(7) was prepared by mixing F(1)$\sim$F(3) solutions at a mass ratio of 1:1:1. Solutions corresponding to F(8)$\sim$F(14) were prepared by mixing according to the ratio specified in the fluid type in the table. For example, the PEO3+CMC1+Carbopol1 solution corresponding to F(8) was prepared by mixing PEO, CMC, and Carbopol at a mass ratio of 3:1:1.

2.2 Machine Learning method

In this study, the overall thinning image of the fluid was analyzed using a machine learning method. For machine learning, 1) Image preprocessing, and 2) Machine learning algorithm methods were applied.

2.2.1 Image preprocessing

Image preprocessing is a method to improve the quality of data used in machine learning. Through this method, we tried to increase the amount of information that can be obtained from data by removing unnecessary data and adding useful data. In this study, four major steps were performed: 1) Frame range selection, 2) Image whitening, 3) Image stacking, and 4) Image cropping and centering.

Table 2.1: Material properties

Index	fluid type	η_0 (Pa.s)	η_E^t (Pa.s)	G' (Pa)	G'' (Pa)
F(1)	CMC	0.055	1.8	0.34	0.03
F(2)	PEO	0.08	64	0.51	0.07
F(3)	Carbopol	52	19	5.01	20.83
F(4)	PEO+CMC	0.09	48	0.55	0.05
F(5)	PEO+Carbopol	32	16.6	7.7	17.1
F(6)	CMC+Carbopol	0.07	0.74	0.3	0.08
F(7)	PEO+CMC+Carbopol	0.2	31.4	0.81	0.24
F(8)	PEO3+CMC1+Carbopol1	0.18	42	0.86	0.19
F(9)	PEO1+CMC3+Carbopol1	0.04	10	0.3	0.05
F(10)	PEO1+CMC1+Carbopol3	0.05	12	1.0	0.46
F(11)	PEO3+CMC3+Carbopol1	0.045	21.8	0.28	0.04
F(12)	PEO3+CMC1+Carbopol3	0.55	22.8	1.23	0.57
F(13)	PEO1+CMC3+Carbopol3	0.043	6.6	0.24	0.05
F(14)	PEO5+CMC1+Carbopol5	1.88	12.4	2.13	1.61

Frame range selection

This is the process of selecting frames that are essential for analysis to remove data that does not correspond to the fluid characteristics and reduce data capacity and calculation. The frame where the angle formed by the drop and the needle is lower than 90 degrees is set as the starting point, t_s, and the frame where the thread formed by thinning is pinched-off is set as the last point, t_b.

Image whitening

The original image has a noise problem due to light transmission. To solve this problem, first, the image was polished through canny edge detection, and the area with fluid between the contours was converted to white color. Through this work, noise caused by light was removed. Through this process, the image was converted into a black-and-white image, and the color channel of the pixel consists of 255 or 0. (Figure2.3)

Image stacking

In previous studies, it was confirmed experimentally or through simulation that the pinch-off shape according to the fluid type was different. However, we judged that the characteristics of the fluid will be expressed not only in the pinch-off behavior but also in the behavior in which thinning occurs. Therefore, the amount of information about the image is increased through image stacking as shown in the figure2.4. At this time, the color values of the pixels of each frame image in the t_s to t_b range are divided by 8000 and then merged to form one frame image. Since the time from thinning to pinch-off is different for each fluid, the brightness of the thinning image in the merged frame is different as shown in figure2.6. We tried to indirectly add time information to the 2D image by creating the brightness difference between these images.

Image cropping and centering

In the original image measured by DoS-CaBER, the needle is exposed on the upper part of the image, and the slide glass is exposed on the lower part of the image. Image data that is not related to the properties of the fluid has a bad effect on the progress of machine learning. Therefore, the upper and lower parts of the image were cut out so that only the characteristics of the fluid were well defined as features.(Figure2.5)

We performed image preprocessing for 7 fluids in this study. As a result, a processed image like figure2.6 was obtained.

2.2.2 Machine Learning algorithm

In previous studies, rheological properties have been obtained based on rheological theories from the thinning behavior of the fluid, or pinch-off shapes for each fluid type have been identified using human visual information. In this study, what humans have been doing in the past was conducted through machine learning. We tried to extract useful variables from thinning image data through machine learning and use them. PCA (Principal Component Analysis) and K-NN (K-Nearest Neighbor) methods were used as machine learning algorithms in this study. PCA is one of the unsupervised machine learning methods and is a representative method for dimensionality reduction while maintaining variance as much as possible. PCA mainly has three functions. The first is dimensionality reduction, and the second is the removal of unnecessary information. The last is the extraction of new variables called PCs (Principal Components) [15, 16]. Also, in the case of image data, visualized information of PCs can be obtained [17, 18]. We aimed to obtain optimized new variables and their visualized information for the thinning behavior of fluids using PCA. K-NN is a simple supervised machine learning algorithm used to evaluate the accuracy of the PCA model through classification tests. The series of processes applying PCA and K-NN in this study are as follows.

Step1. Data preprocessing for PCA

First, vectorize the image data, and stack each sample into each row to create a data frame in the form of (s x p). Here, s is the number of samples, p is the number of original variables (or features), and the number of pixels in the case of image data. In this study, s=227, p=144000. Next, $\mathbf{X}_{mean}$ was obtained by subtracting the average image of the entire image data used from the image of each sample, and this was called mean subtracted data **B**. The data structure with the average subtracted to proceed with PCA is as follows:

$$\mathbf{X}_{mean} = \begin{bmatrix} X_{11}-\mu_1 & X_{12}-\mu_2 & \cdots & X_{1p}-\mu_p \\ X_{21}-\mu_1 & X_{22}-\mu_2 & \cdots & X_{2p}-\mu_p \\ \vdots & \vdots & \vdots & \vdots \\ X_{s1}-\mu_1 & X_{s1}-\mu_2 & \cdots & X_{sp}-\mu_p \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ B_{s1} & B_{s1} & \cdots & B_{sp} \end{bmatrix} = \mathbf{B} \tag{2.1}$$

Step2. Obtaining PCs(Eigen vectors)

This is the process of extracting PCs, which are new variables. Applying SVD(Singular Value Decomposition) to **B** results in the following equation. Here, non-zero singular values s are obtained:

$$\mathbf{B} = \mathbf{U}\Sigma\mathbf{V}^T = \begin{bmatrix} U_{11} & U_{12} & \cdots & U_{1s} \\ U_{21} & U_{22} & \cdots & U_{2s} \\ \vdots & \vdots & \vdots & \vdots \\ U_{s1} & U_{s2} & \cdots & U_{ss} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \sqrt{\lambda_s} & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} V_{11} & V_{21} & \cdots & V_{p1} \\ V_{12} & V_{22} & \cdots & V_{p2} \\ \vdots & \vdots & \vdots & \vdots \\ V_{1p} & V_{2p} & \cdots & V_{pp} \end{bmatrix} \tag{2.2}$$

PCA is a method of extracting features by finding directions with large variances. Therefore, to apply PCA, the covariance matrix **C** must be obtained as follows. When SVD is applied to the covariance matrix, the following equation is obtained:

$$\mathbf{C}_{cov} = \frac{1}{(N-1)}\mathbf{B}^T\mathbf{B} = \frac{1}{(N-1)}\mathbf{V}\Sigma^2\mathbf{V}^T \tag{2.3}$$

V is a matrix in which each column is an eigen vector of the covariance matrix. Each column of **V** corresponding to the eigen vector is a PC, and PCs with large variance sequentially from the left column. A large variance means that there is a lot of information about data. So, the PC with the most information is PC1, the eigen vector of the first column from the left. At this time, if the PC is reshaped to the original image size, a visualized feature of the thinning behavior of the fluid can be obtained. We called this visualized feature Eigen-thinning.(Figure 2.7) The number of PCs is a hyper-parameter that the user must decide. The number of PCs means how many dimensions to reduce and is determined between 1 and s through a hyper-parameter optimization process.

Step3. Calculating score and loading matrix

Using the eigenvector matrix $\mathbf{V}$, the Z score (or PCA score) can be obtained as follows. The obtained Z value is the value of the projected sample data of $\mathbf{X}$ on the PC axis. For each new variable (PC), as many Z values as the number of samples (s) are created. i is equal to the number of dimensions reduced by the number of selected PCs:

$$\mathbf{Z} = \mathbf{XV} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{s1} & X_{32} & \cdots & X_{sp} \end{bmatrix} \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1i} \\ V_{21} & V_{22} & \cdots & V_{2i} \\ \vdots & \vdots & \vdots & \vdots \\ V_{p1} & V_{p2} & \cdots & V_{pi} \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1i} \\ Z_{21} & Z_{22} & \cdots & Z_{2i} \\ \vdots & \vdots & \vdots & \vdots \\ Z_{s1} & Z_{2p} & \cdots & Z_{si} \end{bmatrix} \tag{2.4}$$

The Z value obtained according to equation(2.4) is plotted as figure2.8. Although the data is projected on i-dimension made of i new variables, the dimension that can be expressed visually is limited to 3-dimension, it is a graph visualized using only some PC axes in a 2-dimensional or 3-dimensional coordinate system.

We can also analyze the relationship between the original variables (pixels in the case of image) of data $\mathbf{X}$ and the new variable PCs. We used a method of obtaining the loading matrix $\mathbf{L}$ and visualizing it through the Circle of correlation method(figure2.9). In the case of image data, by analyzing the relationship between pixels and PCs, it is possible to find out the characteristic image shape highly related to each PC in the image. **L** is obtained through the following equation:

$$\mathbf{L} = \mathbf{V}\sqrt{\lambda} = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1i} \\ V_{21} & V_{22} & \cdots & V_{2i} \\ \vdots & \vdots & \vdots & \vdots \\ V_{p1} & V_{p2} & \cdots & V_{pi} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_i} \end{bmatrix} = \begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1i} \\ L_{21} & L_{22} & \cdots & L_{2i} \\ \vdots & \vdots & \vdots & \vdots \\ L_{p1} & L_{2p} & \cdots & L_{pi} \end{bmatrix} \tag{2.5}$$

Step4. Classification using K-NN

Because PCA is unsupervised learning, it cannot evaluate the model itself. Therefore, we used K-NN, which is commonly used for classification, as a simple supervised machine learning algorithm. K-NN is a distance-based classification algorithm like figure 2.10 and has a hyperparameter called k. This hyper-parameter is determined through an optimization process together with the number i of PCs. In this study, the K-NN method was used for the classification of data in the new coordinate system obtained through PCA.

Step5. Hyperparameter Decision

There were two parameters (Number of PCA components: i, Number of judgment criteria in K-NN: k) that we have to decide. The classification was possible through K-NN, and we calculated the accuracy of the model through this. To check the performance of the model, accuracy was used as a criterion in hyperparameter optimization. As shown in figure2.11, we set the i parameter at an interval of 5 from $1 \sim 46$, and set the k parameter at an interval of 1 from $1 \sim 20$ according to the i parameter to calculate the accuracy by iterative calculation.

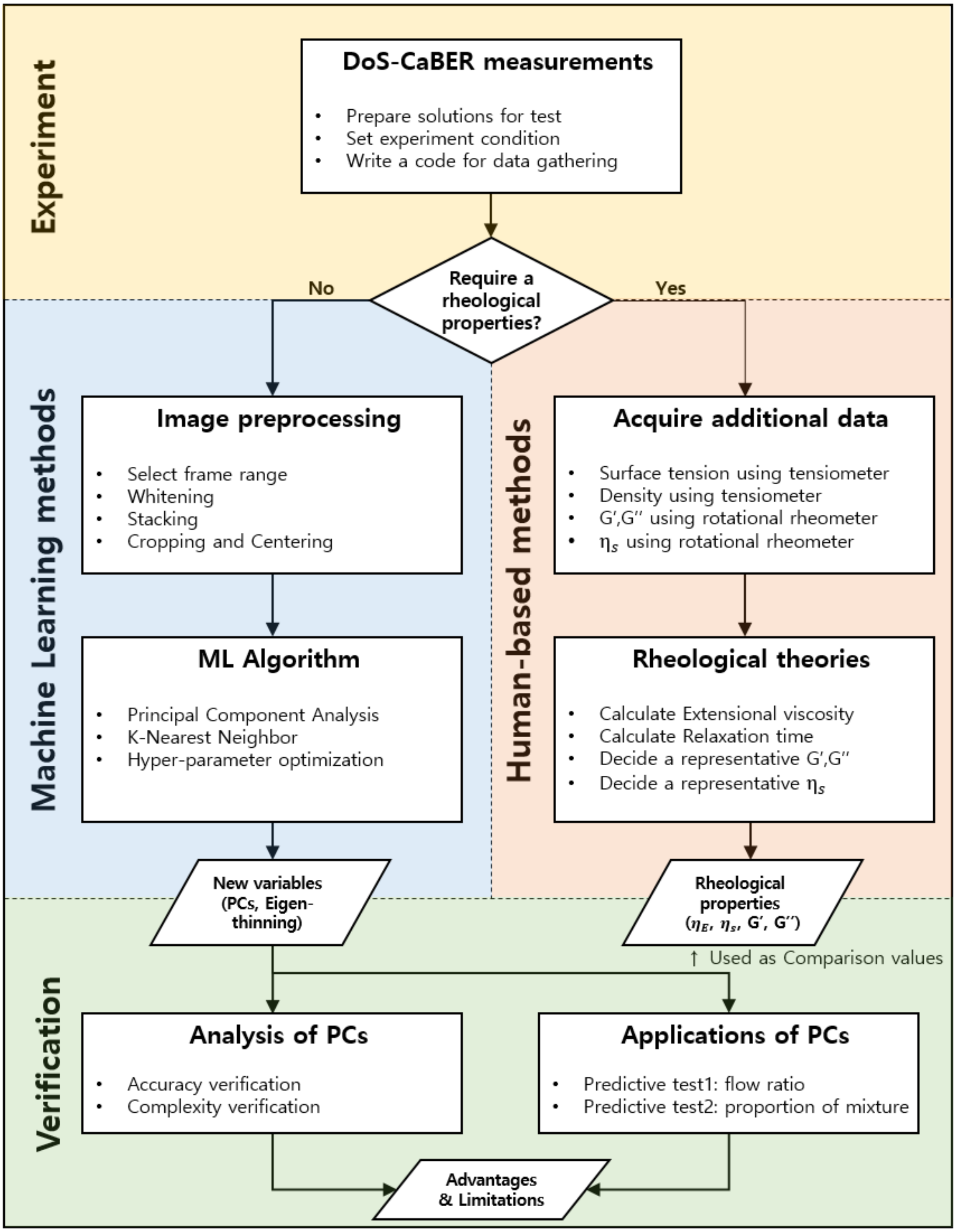


Figure 2.1: Flow chart

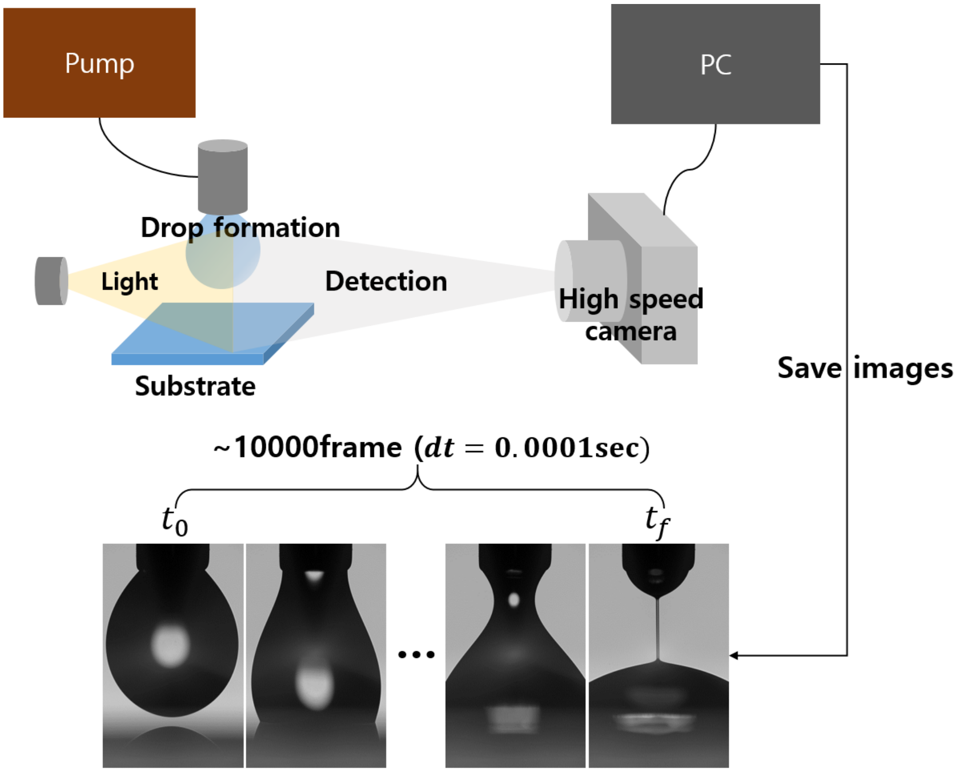


Figure 2.2: Configuration of DoS-CaBER

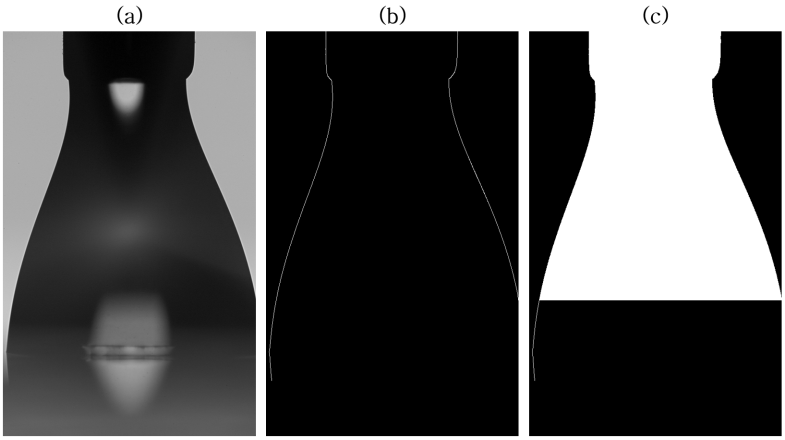


Figure 2.3: Image whitening. (a): Raw image, (b): Edge image, (c): Whiten image

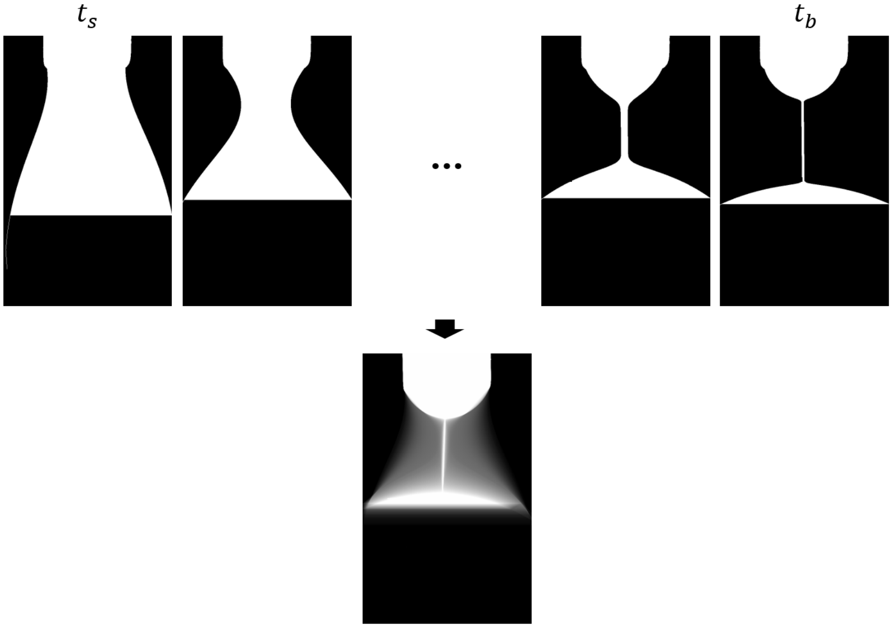


Figure 2.4: Image stacking. Images measured from t_s to t_b are stacked into one image.

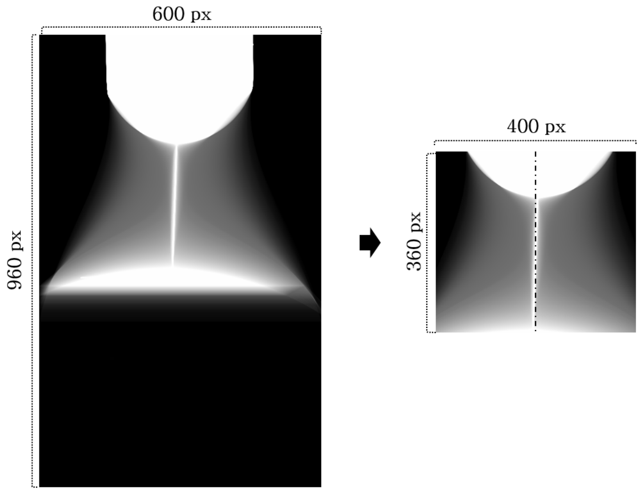


Figure 2.5: Image cropping and centering

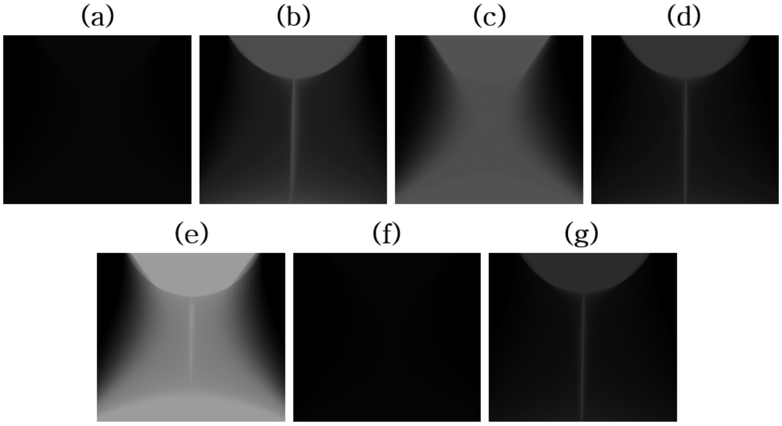


Figure 2.6: Images of 7 types of fluids. (a):CMC,(b):PEO,(c):Carbopol,(d):PEO+CMC,(e): PEO+Carbopol,(f): CMC+Carbopol,(g): PEO+CMC+Carbopol.

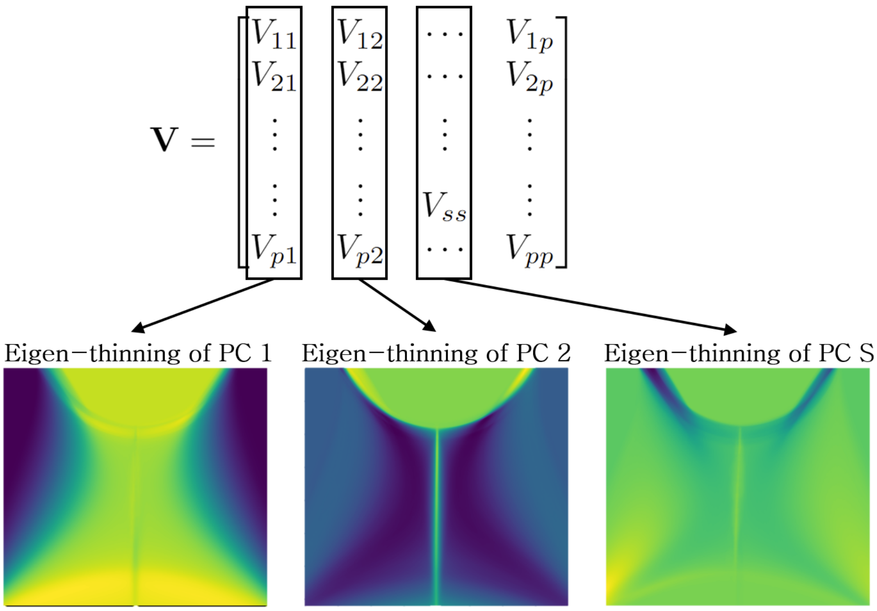


Figure 2.7: Eigenvector matrix V and Eigen-thinning of each PC.

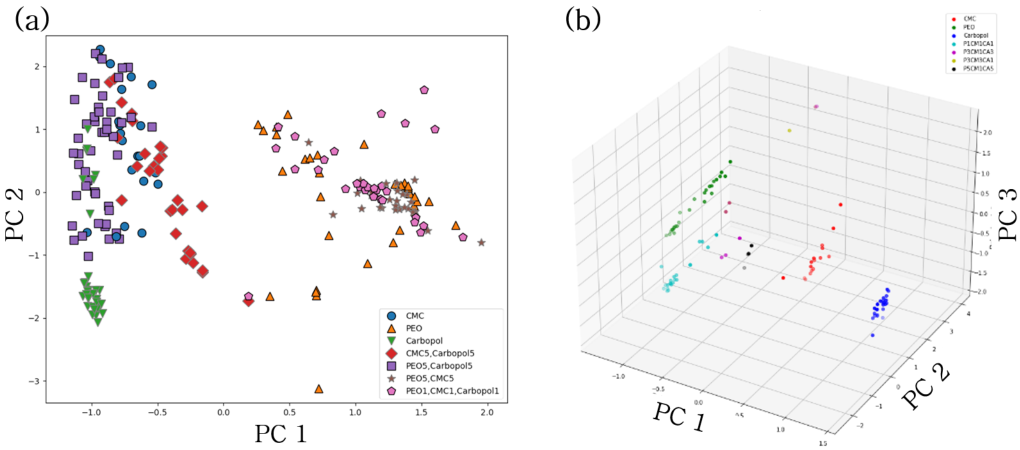


Figure 2.8: PCA score graph. (a) is a 2D projected graph, and (b) is a 3D projected graph

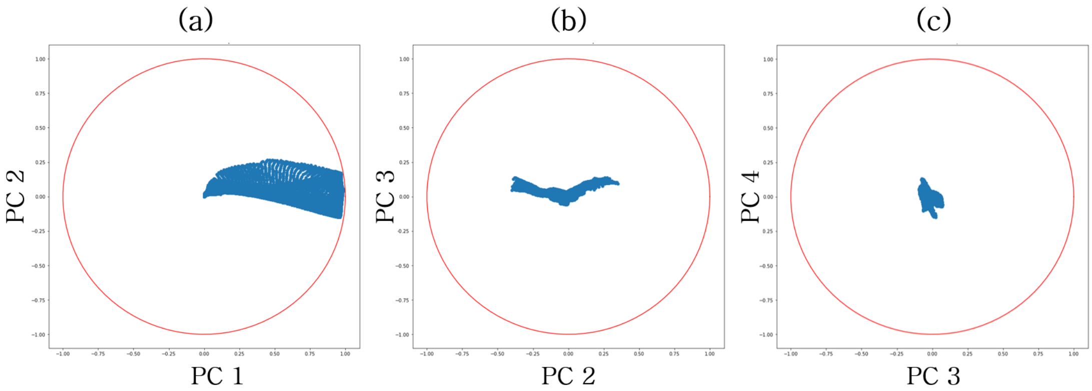
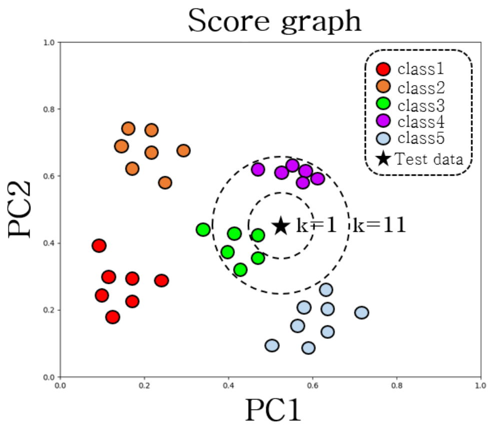


Figure 2.9: Circle of correlation graph.



If k = 1, ★ : Class 3

	Class1	Class2	Class3	Class4	Class5
Probability of K−NN	0	0	1/1	0	0

If k = 11, ★ : Class 4

	Class1	Class2	Class3	Class4	Class5
Probability of K−NN	0	0	5/11	6/11	0

Figure 2.10: K-NN algorithm. The distance is calculated based on the Euclidean distance. k is the number of adjacent data to be used for classification judgment and is a hyperparameter. The Knn probability for each class is calculated by dividing the number of data in each class by k, and the test data is classified into the class with the highest knn probability.

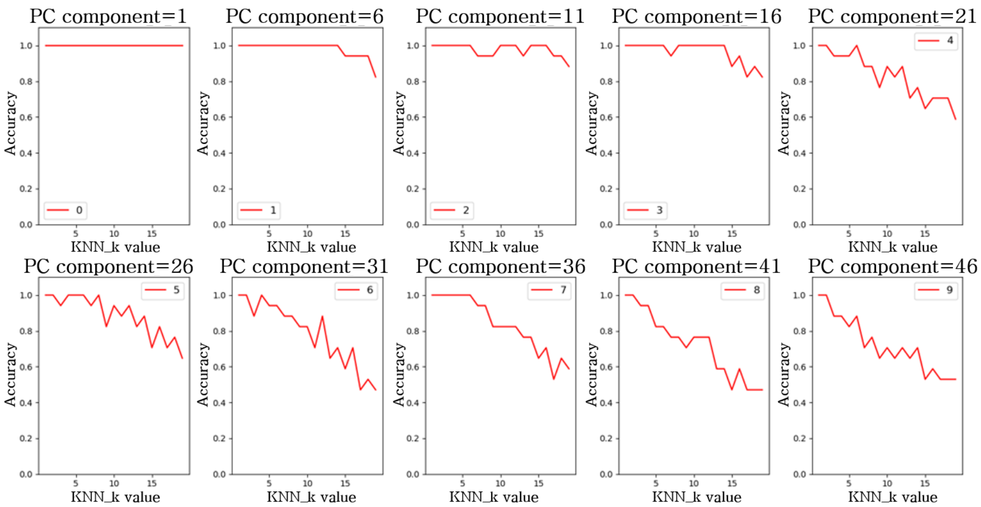


Figure 2.11: Hyperparameter optimization. Accuracy graphs were plotted according to PCA components (1-46,term:5), and K-NN parameters (1-20,term:1).

Chapter 3

Results and Discussion

3.1 Accuracy of PCs

We verified the effect of applying image preprocessing and PCA through the accuracy test of the classification model. The performance of classification means that the new variables (PCs) well classify thinning images of fluids, which means that PCs explain the thinning behavior of fluids well. We used 7 types of fluids corresponding to Index F(1) $\sim$ F(7) as the data set for the test. Each F(1) $\sim$ F(3) is a pure solution of CMC, PEO, or Carbopol, and F(4) $\sim$ F(7) are a mixture of the solutions. Figure3.1 is the accuracy test result in the range of hyper-parameter i = 6 of PCA components and hyper-parameter k = 1 $\sim$ 20 of K-NN. The four lines in the figure are the results according to the conditions of PCA and image preprocessing. As a result, when both images preprocessing and PCA were applied, the highest value was obtained with an accuracy of 93 $\sim$ 100%. In particular, the effect of PCA was greater for data that was image preprocessed. This is considered to be because useful data and noise increase together in the process of increasing the complexity of data by image stacking in image preprocessing. In conclusion, we verified extracted new variables (PCs) through the PCA classification model from image preprocessed data.

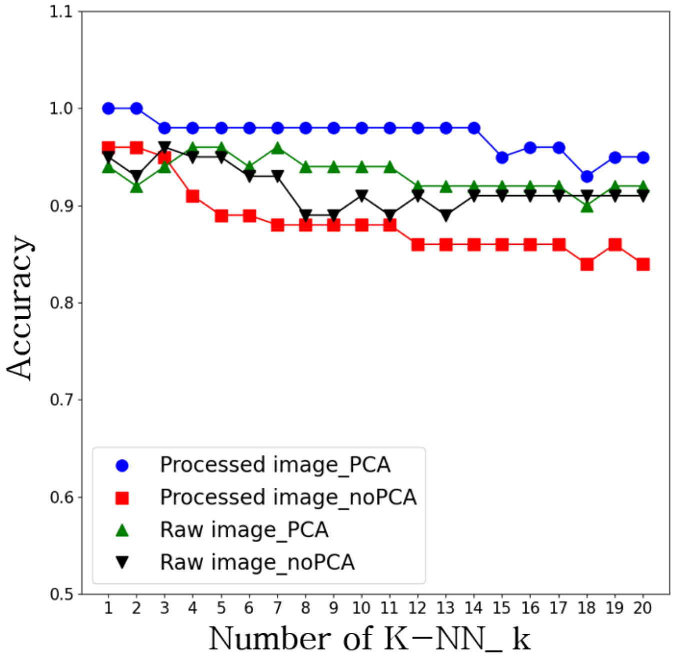


Figure 3.1: Accuracy test result. Blue line(●) is image-preprocessing:O, PCA:O condition, Red line(■) is image-preprocessing:O, PCA:X condition, Green line(▲) is image-preprocessing:X, PCA:O condition, Black line(▼) is image-preprocessing:X, PCA:X condition

3.2 Relationship between PCs and rheological properties

PCs of new variables obtained through PCA are created through linear combinations of original variables. In this study, the original variables are not rheological properties, but thinning images themselves that show rheological behavior. Since we extracted new variables from the original variables, We expected that PCs would complexly explain the dominant rheological behaviors existing in the thinning behavior. Since PCs are variables statistically calculated by machine learning, it is difficult to directly explain their connection with specific rheological properties. However, we tried to indirectly interpret the rheological meaning of PCs by comparing the conventional rheological property values and PC score values for the test fluid. As a result, in this study, it was confirmed that each of PC1 and PC2 showed a tendency similar to storage modulus(G'), extensional viscosity(η_E), as shown in figure3.2. In conclusion, we indirectly confirmed that PCs complexly explain the rheological behavior, and in Application part 1, we conducted an experiment on the complexity of these PCs.

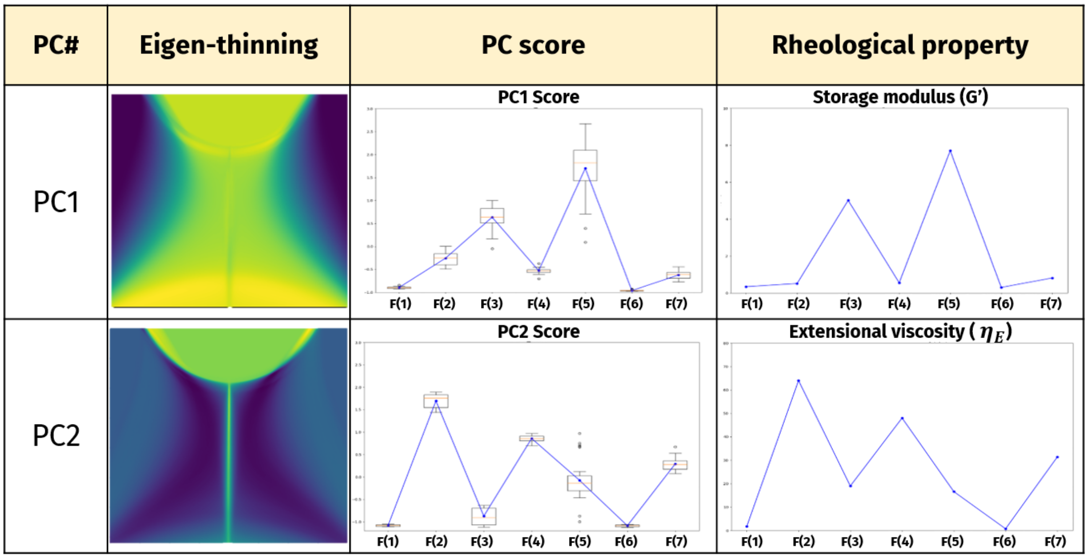


Figure 3.2: Relationship between PCA results and rheological properties.

Chapter 4

Applications

4.1 Prediction of flowability in micro-channel

4.1.1 Concept

We conducted a feasibility test to use PCs as an index for processes affected by rheological behavior. In summary, we tried to predict the flow of certain fluids in a complex micro-channel using several rheological properties and PCs and compared each prediction result. A channel was made in the form shown in the figure4.1. Some fluid flowed into the channel. The amount of fluid flowing into the wide area and the narrow area was measured and the ratio was obtained. We called that ratio the flow ratio. Seven types of fluids corresponding to Index F(1) to F(7) were used as classes to obtain PCs, and solutions corresponding to Index F(8) to F(14), which are mixtures of the solutions, were used as test fluids to be predicted. For prediction, the flow ratio of the fluid corresponding to the class was obtained experimentally in advance, and each flow ratio was normalized to the flow ratio of water.

$$Flow\ ratio \equiv mass_{wide\ pattern}/mass_{narrow\ pattern} \tag{4.1}$$

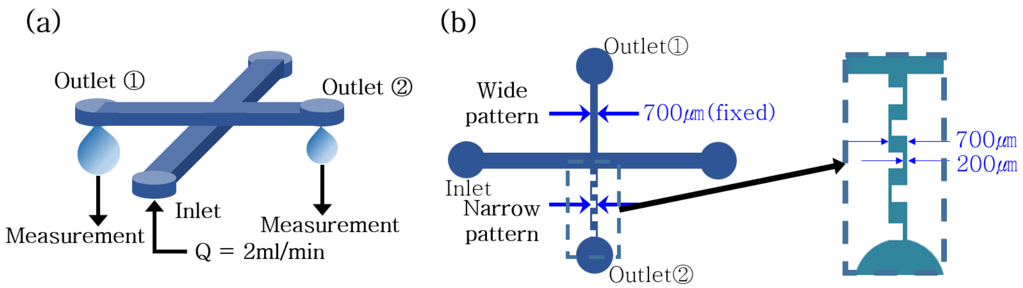


Figure 4.1: Configuration of Micro Channel experimernt. (a): summary of experimental conditions, (b): design of micro-channel.

4.1.2 Method

Hyperparameter optimization was performed to predict the flow ratio. The decision criterion was the minimum value of the RMSE between the predicted flow ratio of the validation data and the experimental flow ratio. As a result of optimization, it was determined that PCA components i = 1 and K-NN k = 1 as shown in table4.2. The predicted value of the flow ratio is obtained by the following equation. The KNN probabilities were calculated by the 2.10 method, and the flow ratio of the class was experimentally obtained (Table4.1). Here, n is each class number :

$$Flow\ ratio_{prediction} = \sum_{1}^{n}(knn\ probability_{class(n)} * Flow\ ratio_{class(n)}) \quad (4.2)$$

Table 4.1: Flow ratio of class

Fluid type	Flow ratio	Fluid type	Flow ratio
CMC	1.11	PEO+Carbopol	0.60
PEO	0.77	CMC+Carbopol	1.19
Carbopol	0.86	PEO+CMC+Carbopol	0.77
PEO+CMC	0.90		

4.1.3 Results and Discussion

This is the result of predicting the flow ratio of the cross channel through the model trained with 7 classes. The results were compared with the predicted results based on shear viscosity and extensional viscosity. For the prediction using viscosity, the simple polynomial regression equation obtained through the property information of class fluid and the flow ratio value was used. As a result, RMSE = 0.12 for prediction using shear viscosity, RMSE = 0.19 for prediction based on extensional viscosity, and

Table 4.2: hyperparameter optimization

Class	Class type	PCA components	KNN-k	RMSE
7	(CMC) (PEO) (Carbopol) (PEO+CMC) (PEO+Carbopol) (CMC+Carbopol) (PEO+CMC+Carbopol)	1	1	0.087

RMSE = 0.09 for prediction using the 7class PCA model.(Figure 4.2) When compared based on RMSE, it was confirmed that the error of the 7-class PCA model was the smallest. We confirmed that the flow ratio in the channel can be roughly predicted simply through this PCA model. However, it is difficult to predict the exact value. To improve predictability, in the case of existing rheological properties, simulations can be performed using rheological theories and models. On the other hand, since PCs are not variables based on rheological theories, it is necessary to consider how to use various PCs in combination. We are considering the application of machine learning methods such as Neural Network as a rough solution to these problems.

4.2 Prediction of the proportion of mixtures

4.2.1 Concept

We tried to predict the proportion of the mixture as well as the classification using the new variable PCs. In this experiment, a mixture was prepared using three types of solutions: PEO, CMC, and Carbopol. Two experiments were carried out, and one was an experiment to predict the ratio of a mixture(Test fluid: F(4)∼F(7)) with 3 classes

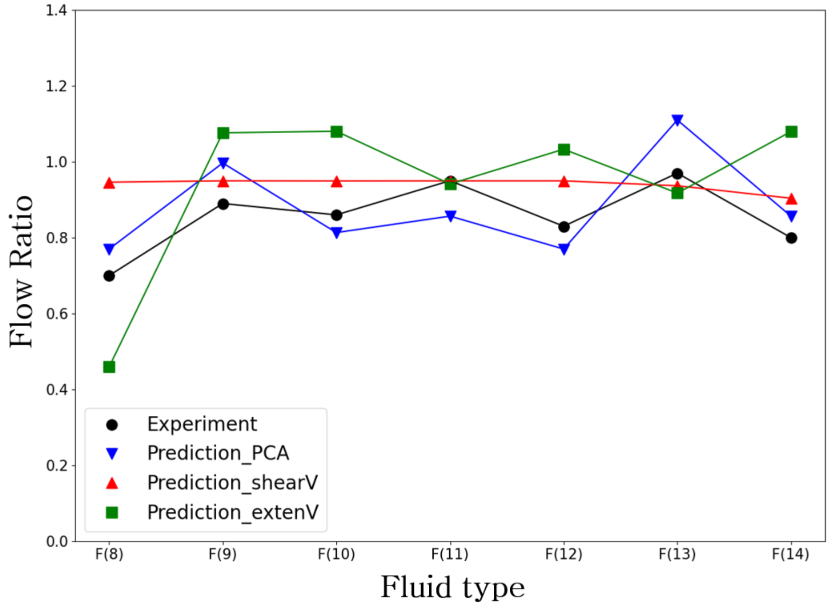


Figure 4.2: Flow Ratio Result. Black line(●) is experimental value, Blue line(▼) is prediction result based on PCA model, Red line(▲) is prediction result based on shear viscosity, Green line(■) is prediction result based on extensional viscosity

(Train fluid: F(1)∼F(3)) PCA model. The other was an experiment to predict the ratio of a mixture(Test fluid: F(8)∼F(13)) in which three solutions were mixed in a specific ratio using 3class(Train fluid: F(1)∼F(3)) PCA model and 7class(Train fluid: F(1)∼F(7)) PCA model. The fluid types corresponding to the F(1) to F(13) index are shown in table2.1.

4.2.2 Method

The finally determined hyperparameters (pca components, and KNN k) are shown in the table4.3. The hyper-parameter pca components i = 1 ∼ 50 were calculated at 5 intervals, and K-NN k = 1 ∼ 20 was calculated at 1 interval. The hyper-parameter was determined based on the minimum value of the actual mixture ratio and the expected mixture ratio RMSE. The method of predicting the pure solution proportion of the test mixture is as follows. Here, n is each class number:

$$CMC\,ratio_{prediction} = \sum_{1}^{n}(knn\,probability_{class(n)} * CMC\,proportion_{class(n)}) \tag{4.3}$$

$$PEO\,ratio_{prediction} = \sum_{1}^{n}(knn\,probability_{class(n)} * PEO\,proportion_{class(n)}) \tag{4.4}$$

$$Carbopol\,ratio_{prediction} = \sum_{1}^{n}(knn\,probability_{class(n)} * Carbopol\,proportion_{class(n)}) \tag{4.5}$$

Table 4.3: hyperparameter optimization of mixture prediction model

Class	Class type	PCA components	KNN-k	RMSE
3	(CMC)	21	2	0.08
(for test1)	(PEO)			
	(Carbopol)			
3	(CMC)	6	14	0.36
(for test2)	(PEO)			
	(Carbopol)			
7	(CMC)	6	15	0.17
(for test2)	(PEO)			
	(Carbopol)			
	(PEO+CMC)			
	(PEO+Carbopol)			
	(CMC+Carbopol)			
	(PEO+CMC+Carbopol)			

4.2.3 Results and Discussion

The result of predicting mixtures (F(4)∼F(7)) through the 3class (F(1)∼F(3)) PCA model was as shown in the figure4.3(a). Through the prediction results of PEO+CMC(F(4)), PEO+Carbopol(F(5)), CMC+Carbopol(F(6)), PEO+CMC+Carbopol(F(7)) When RMSE was obtained, each RMSE = 0.16, 0.00, 0.36. The predicted values of the PEO+Carbopol mixture and the PEO+CMC mixture showed a relatively small error, whereas the predicted values of the CMC+Carbopol mixture showed a large error. Figure4.3 (b) is a graph in which the score values of each fluid data are scattered on the axes of PC1 and PC2, which have the largest variance. Looking at the figure, it can be seen that the CMC+Carbopol solution is distributed in a region that is completely biased towards the data of the CMC solution, not the midpoint of the CMC and Carbopol pure solution. In other words, it can be seen that the thinning behavior of the CMC+Carbopol solution is similar to that of CMC from the perspective of PC1, and PC2, not the intermediate behavior of each pure solution. This is also shown in Figure 4.4. From graph(a), it can be seen that the shear viscosities of the PEO, Carbopol, and PEO+Carbopol mixture have relatively linear values. On the other hand, the shear viscosities of CMC, Carbopol, and CMC+Carbopol mixture show relatively more non-linear behavior. As a result of the score scattering graph, it can be seen that the CMC+Carbopol solution has a shear viscosity similar to that of CMC. We additionally built a classification model based on 7 classes of F(1) ∼ F(7) solutions to predict the proportions of a slightly more complex mixture. Through the model, we tried to predict the ratio of mixtures (F(8) to F(13)) in which all of F(1) to F(3) were mixed. Figure 4.5 (a),(b) are the results predicted by the 3class model and the 7class model. As a result, when using the 7class model, a lower error occurred compared to the 3class model (RMSE: 0.35 → reduced to 0.14). From this result, it can be inferred that the error will decrease if fluids mixed in various proportions are added as a new class. However, this is inefficient as it requires more experiments and data. The limitation of these predictions is the lack of information obtained through the model. The lack of information could have several causes. First, the thinning behavior itself may lack information to predict the proportion of the mixture, and in this case, data collection through other experiments is required. The second is that the PCA model in this study uses intensity-based data through image stacking. If thinning behavior occurs in a short time, feature extraction may be difficult due to low intensity in stacking images. The third is that it is difficult to extract features from the positional variable image data of PCA itself. Most of the thinning behavior data we used were collected under equal conditions, the shape was symmetrical, and the position was rarely changed. However, at the time of pinch-off, it was difficult to obtain a feature as a shape such as BOAS (Beads On A String) because the position or size was multivariable. In summary, the two problems estimated methodologically are as follows. 1) Limitation of Intensity-based data analysis, 2) Limitation of positional variable data analysis. To verify this, additional research applying machine learning algorithms such as CNN (with edge feature method and positional invariance characteristics) is needed.

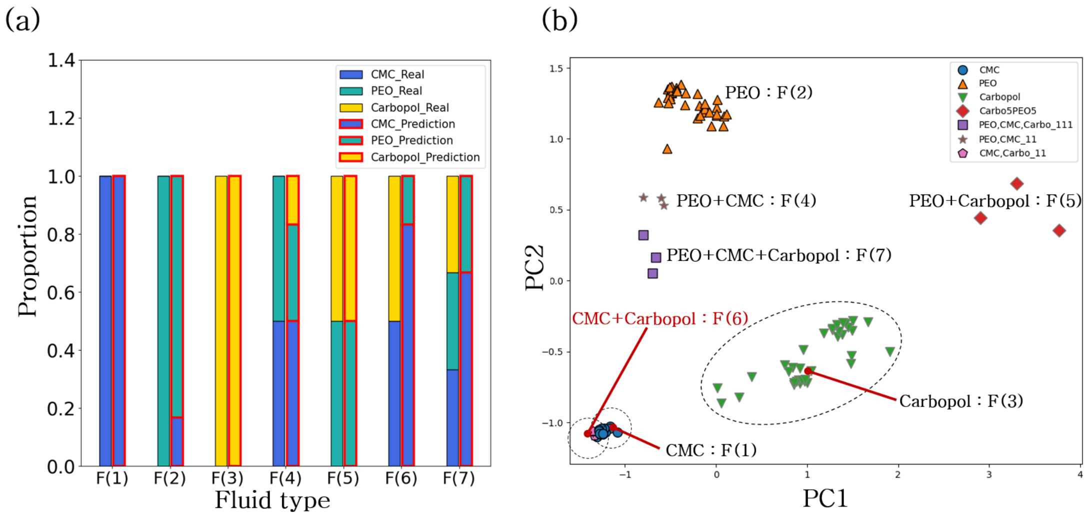


Figure 4.3: A result of prediction test of binary solutions. (a) A 3class-based proportion graph comparing real proportion with the predicted proportion of binary solutions. The black line bars mean the real proportion values, and the red line bars mean the predicted proportion values. (b) A score scattering graph of solutions on PC1, PC2 axis. The black dot line circle indicates CMC, Carbopol pure solutions, and the mixture of them.

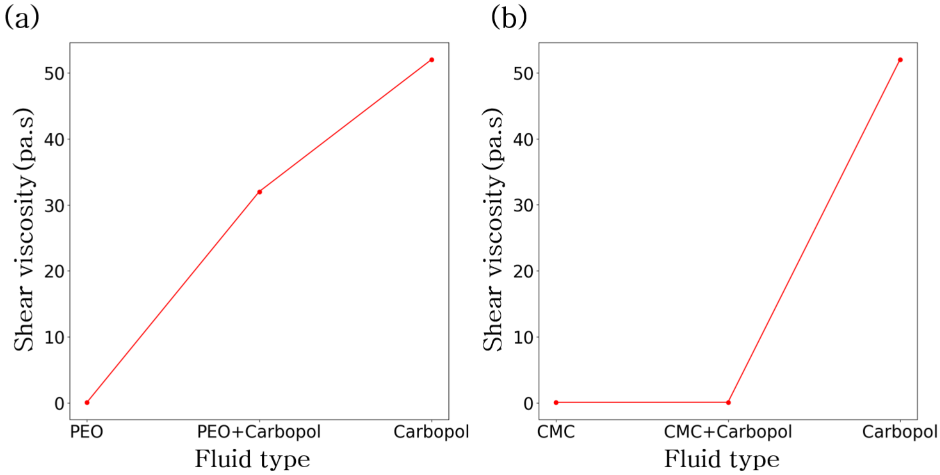


Figure 4.4: (a) A shear viscosity graph of PEO, Carbopol binary solution, (b) A shear viscosity graph of CMC, Carbopol binary solution.

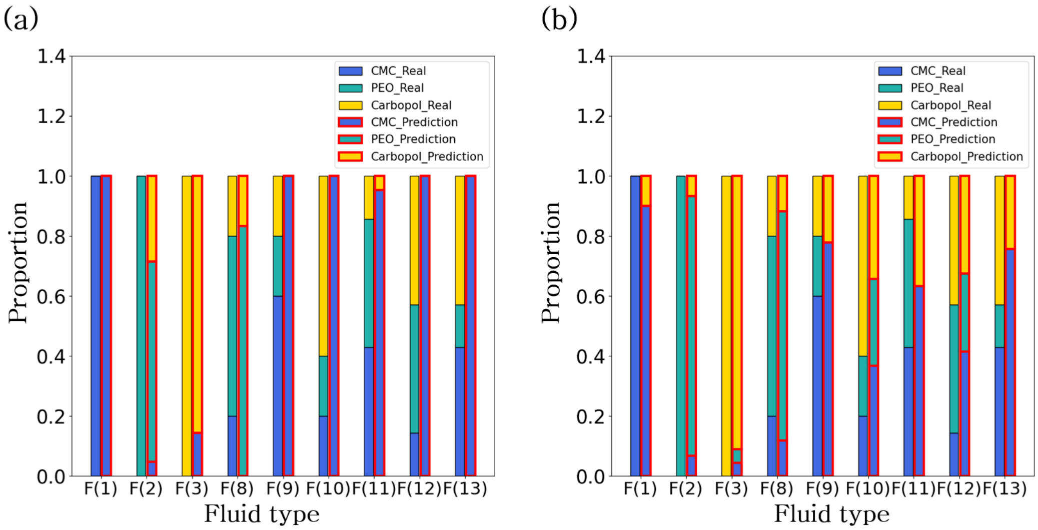


Figure 4.5: (a) A proportion graph of mixtures which are predicted with 3class model, (b) A proportion graph of mixtures which are predicted with 7class model.

Chapter 5

Final remarks

This thesis focused on analyzing fluid thinning behavior using machine learning. Our goal was to obtain new rheological information(PCs, Eigen-thinning) explaining the overall thinning behavior using the PCA machine learning algorithm.

Two methods have been described for the goal. The first was an experimental method, which is the acquisition of fluid thinning behavior images using DoS-CaBER. The second was the machine learning method. We first implemented the preprocessing method of fluid thinning images for machine learning. Next, new variables (PCs) were acquired from the image data using the PCA method. The effect and characteristics of the PCs were verified through the accuracy test of the classification model and comparison with conventional rheological properties.

We conducted prediction tests using PCs. The first test was an experiment to predict the flow ratio in the micro-channel, and the second test was to predict the ratio of the mixture. As a result, we confirmed the proposed method's advantages and limitations. The first limitation of the method in this study was the problem of the combination of new variables (PCs). For relatively accurate prediction, it is necessary to develop a method that can properly combine multiple PCs. The second was that some features of image data may be lost. The PCA model in this study may have problems due to the aspects of the image pixel intensity-based method, positional variance, and statistical variance-based analysis method. Considering these limitations comprehensively, it seems necessary to develop using additional machine learning methods such as CNN.

Bibliography

[1] Jelena Dinic and Vivek Sharma. Macromolecular relaxation, strain, and extensibility determine elastocapillary thinning and extensional viscosity of polymer solutions. *Proceedings of the National Academy of Sciences*, 116(18):8766–8774, 2019.

[2] Jelena Dinic, Yiran Zhang, Leidy Nallely Jimenez, and Vivek Sharma. Extensional relaxation times of dilute, aqueous polymer solutions. *ACS Macro Letters*, 4(7):804–808, 2015.

[3] Gareth H McKinley and Tamarapu Sridhar. Filament-stretching rheometry of complex fluids. *Annual Review of Fluid Mechanics*, 34(1):375–415, 2002.

[4] Jens Eggers, Miguel Angel Herrada, and JH Snoeijer. Self-similar breakup of polymeric threads as described by the oldroyd-b model. *Journal of fluid mechanics*, 887, 2020.

[5] Jelena Dinic, Leidy Nallely Jimenez, and Vivek Sharma. Pinch-off dynamics and dripping-onto-substrate (dos) rheometry of complex fluids. *Lab on a Chip*, 17(3):460–473, 2017.

[6] Gareth H McKinley. Visco-elasto-capillary thinning and break-up of complex fluids. 2005.

[7] Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng, and Yi Ma. Pcanet: A simple deep learning baseline for image classification? *IEEE transactions on image processing*, 24(12):5017–5032, 2015.

[8] Jakub Nalepa, Michal Myller, and Michal Kawulok. Hyperspectral data augmentation. *arXiv preprint arXiv:1903.05580*, 2019.

[9] Luke Taylor and Geoff Nitschke. Improving deep learning with generic data augmentation. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1542–1547. IEEE, 2018.

[10] Angelo Genovese, Vincenzo Piuri, Konstantinos N Plataniotis, and Fabio Scotti. Palmnet: Gabor-pca convolutional networks for touchless palmprint recognition. *IEEE Transactions on Information Forensics and Security*, 14(12):3160–3174, 2019.

[11] Juan José Rodriguez, Ludmila I Kuncheva, and Carlos J Alonso. Rotation forest: A new classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence*, 28(10):1619–1630, 2006.

[12] Isha Garg, Priyadarshini Panda, and Kaushik Roy. A low effort approach to structured cnn design using pca. *IEEE Access*, 8:1347–1360, 2019.

[13] Ujwala Patil and Uma Mudengudi. Image fusion using hierarchical pca. In *2011 international conference on image information processing*, pages 1–6. IEEE, 2011.

[14] Gebhard Schramm et al. *A practical approach to rheology and rheometry*. Haake Karlsruhe, 1994.

[15] Steven L Brunton and J Nathan Kutz. *Data-driven science and engineer-*

ing: Machine learning, dynamical systems, and control. Cambridge University Press, 2022.

[16] Jake Lever, Martin Krzywinski, and Naomi Altman. Points of significance: Principal component analysis. *Nature methods*, 14(7):641–643, 2017.

[17] Lawrence Sirovich and Michael Kirby. Low-dimensional procedure for the characterization of human faces. *Josa a*, 4(3):519–524, 1987.

[18] Michael Kirby and Lawrence Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern analysis and Machine intelligence*, 12(1):103–108, 1990.

국문 초록

많은 산업공정에서 사용되는 기능성 소재들은 기능의 향상을 위해 고분자,금속,탄소 입자들과 같은 다양한 원재료로 만들어진다. 이에따라 점탄성 및 항복 거동 등 다양한 유변학적 거동을 보인다. 이러한 유변학적 거동들이 유체가 얇아지는 거동에서 특징적인 형태로 나타남이 선행연구에서 확인이 되었다. 따라서 유체의 얇아지는 거동을 시각적으로 분석하는 것은 중요하다.

본 학위논문은 간단한 기계학습 방식들을 사용하여 유체의 얇아지는 거동을 분석하는 연구를 다루고있다. 우선, 유체의 얇아지는 거동의 이미지들을 DoS-CaBER (Dripping onto Substarte-Capillary Break-up Extensional Rheometer)를 사용하여 획득하였다. 다음으로, 효과적인 기계학습을 위해 이미지들의 전처리 과정을 개발하였다. 새로운 유변학적 정보들을 얻기위해 주성분분석(PCA)을 기계학습 알고리즘으로 사용하였다. 결과적으로 유체의 거동특성을 설명하는 주성분들과 고유의 얇아짐 이미지(Eigen-thinning)을 확보하였다. 주성분들을 활용한 분류모델을 통해 주성분들의 정확도를 검증하였으며, 전통적인 유변물성들과 주성분들의 비교를 통해 주성분들의 유변학적 복합성을 확인하였다.

우리는 또한 활용평가들을 진행하였다. 주성분들은 두가지 예측 평가에 사용되었다. 첫 번째는 복잡 미세관을 흐르는 유체의 흐름정도를 예측하는 평가였다. 이 평가를 통해 주성분들이 간단하고 빠르면서 대략적인 예측이 요구되는 경우에 유용한 예측 지표가 될 수 있음을 확인하였다. 두 번째는 혼합유체의 혼합비율을 예측하는 실험이었으며, 주성분에서 지배적으로 나타나는 유변물성이 혼합비에 따라 선형적인 변화를 가질 경우 혼합비율 예측의 가능성을 확인하였다. 마지막으로 제안한 기계학습 방법의 이점과 한계점 도출을 통해 향후 연구방향의 일부를 제시하였다.

주요어: 유체 얇아짐 거동, 액체의 끊어짐 거동, 기계학습, 주성분 분석, 신장유변 물성 측정 레오미터, 유변 물성
학번: 2021-26394