

predict film property

Ki Jun, Choi

2018-08-30

Introduction

In order to manufacture a PET film, first, a solid PET chip is melted, pressed in a DIE process, and then rapidly cooled to form an amorphous sheet. Thereafter, while heating the sheet, it is stretched in the longitudinal direction (TD) and the width direction (MD). The PET film manufactured in this way is for packaging, industrial, and optical purposes. Material) and solar power.

Heat shrinkage is a physical property value that refers to the rate at which heat is applied to a product, and varies depending on stretching, heat treatment, and extrusion conditions. 18' SKC Data CoE Pjt selected products produced by SKC's 7th unit as a target. In the case of No. 7, the film is stretched in the width direction (MD). The process (TSM) has a characteristic consisting of a total of 13 Tenters. In this Tenter, not only the transverse direction (MD) is stretched, but also heat is simultaneously Apply. Among the various factors that determine film heat shrinkage, this analysis includes crystallinity (CHIP properties: IV), draw ratio (MD/TD direction stretching & relaxation), A model was developed to predict the heat contraction rate through factors such as stretching speed and heat treatment (stretch zone, heat stabilization, etc).

```
# Library
library(data.table)
library(jsonlite)
library(dplyr)
library(ggplot2)

# source
source(file = "./src/data_loader.R")
source(file = "./src/feature_extraction.R")
source(file = "./src/model.R")
source(file = "./src/utils.R")

# config file
config <- jsonlite::fromJSON(txt = "config.json")

# define arguments from config file
data_dir <- config$data_loader$args$data_dir
filename <- config$data_loader$args$filename
doff      <- config$variable$args$doff
species   <- config$variable$args$species
outlier_doff      <- config$variable$args$outlier_doff
outlier_species <- config$variable$args$outlier_species
```

Load data

Data from 2015 to 2018 were used. The data from 2015 to 2017 were recorded in the working paper(Handwriting input), and 2018 data is collected through the sensor.

```
# read data
df <- data_loader( data_dir = data_dir, filename = filename
                  , doff = doff, species = species
                  , outlier_doff = outlier_doff
                  , outlier_species = outlier_species )
```

Loading DataThe dimension of data: 879 74

The description of the data is shown in the table below.

Columns	Description	Remark
Doff	Film identification number	The unit of analysis is the doff unit Analysis of specific types (eg. Excluding special types of film)
Species	Film type	
Thick	Film thickness (spec thickness not actual thickness)	Target property of analysis A value indicating the degree of crystallinity
MDHS	Heat shrinkage in MD(machine direction) direction	
IV	CHIP characteristic value	
Speed	Film production speed (spec speed)	
LSM_Speed_2	Speed before stretching in TD(tenter direction) direction	
LSM_Speed_3	Speed after stretching in TD(tenter direction) direction	
PV	Measuring sensor temperature	
RPM	Measuring sensor RPM	
Pattern	The length of the film in the MD direction measured between the tenters	

Derived Variable

Through the sensor values, variables that affect the actual heat shrinkage were created. Largely, there are variables related to the pattern of the film and variables related to the amount of heat.

1. Variable related to Pattern

- ETA1: TD direction draw ratio
- ETA1_SPEED: Speed of TD draw
- ETA2: MD direction draw ratio
- TD_SPEED: The speed of TD Stretch
- RELAX: The rate of relaxation of film

2. Variable related to Heat (Refer to `feature_extraction.R`)

- Calculation of the amount of heat received by the actual film for each Tenter (1~13T)
- The calorie calculation method is developed under the leadership of the engineer

```
df <- df %>%
  mutate(
    # -- ETA
    ETA1_SPEED = (LSM_Speed3 / LSM_Speed2) * (LSM_Speed3 - LSM_Speed2)
    , ETA1 = LSM_Speed3 / LSM_Speed2
    , ETA2 = PTN_11 / PTN_4

    # -- Speed of Tenter Direction
    , TD_SPEED = ((PTN_8 - PTN_5) * .001) / ((3.0 * 3) / (SPEED/60))

    # -- The ratio of Relaxation
    , RELAX = ((PTN_13 - PTN_15) / PTN_13) * 100
    , RELAX_1 = ((PTN_14 - PTN_17) / PTN_14) * 100
    , RELAX_2 = ((PTN_13 - PTN_17) / PTN_13) * 100
  ) %>%
  arrange( DOFF_NO )

# calculate actual temperature ( using Chain calculation method )
temperature <- local(expr = {
  tmp <- list()
  for(i in 1:13){

    # Initial value is set to 30 when calculating for the first time
    if( i == 1 ){
      tmp_s <- calculate_heat_calories( df = df, config = config$variable$args
                                       , direction = "S", zone = 1, init = 30)
      tmp_n <- calculate_heat_calories( df = df, config = config$variable$args
                                       , direction = "N", zone = 1, init = tmp_s)

      tmp[[i]] <- data.table(tmp_s, tmp_n)
      names(x = tmp[[i]]) <- paste("temp", c("S", "N"), i, sep = "_")

      # Set previous calculated value as initial value
    }else{
      tmp_s <- calculate_heat_calories( df = df, config = config$variable$args
                                       , direction = "S", zone = i, init = tmp_n)
      tmp_n <- calculate_heat_calories( df = df, config = config$variable$args
                                       , direction = "N", zone = i, init = tmp_s)

      tmp[[i]] <- data.table(tmp_s, tmp_n)
      names(x = tmp[[i]]) <- paste("temp", c("S", "N"), i, sep = "_")
    }
  }
  tmp <- do.call(what = "cbind", args = tmp)
  return( tmp)
})

df <- cbind(df, temperature)
```

Modeling

The model for the MD direction heat shrinkage was estimated through the factors based on knowledge in the field. The estimated model was developed as a prediction trend chart screen and a simulation service screen. Both services are due to the needs of the field, and for this, the model was proposed as Regression. Regression may be somewhat insufficient in predictive power than other machine learning models, but there are parts that are easy to interpret. Due to the nature of a conservative factory, a model that can be understood well in the field, such as regression, is more appropriate than the Black-Box model(And it is also a good model to judge the degree of agreement between the knowledge of the field and the analyzed model). In addition, regression is also a way to show good predictive power as long as the relationship between actual factors is well discovered and applied to data well. Therefore, I decided that there was no big problem.

1. Validation

Before fitting the model, the data were partitioned for more correct model selection.

```
idx <- 1:nrow(x = df)
train_idx <- 1:(nrow(x = df) * .8)
valid_idx <- idx[-train_idx]

train_df <- df[train_idx, ]
```

2. Regression

There are a total of 6 variables used in Regression. More details are shown in the table below.

Columns	Description	Relationship	Remark
IV	CHIP characteristic value	Positive correlation	
ETA1	TD direction draw ratio	Positive correlation	
RELAX	Film type	Negative correlation	Insignificant effect
Temp_7_S	Temperature in Stretching Zone	Negative correlation	Molecular structure is most active
Temp_10_N	Temperature in Heat Fixation zone	Negative correlation	The most influential
Temp_13_N	Temperature in cooling zone	Negative correlation	No big influence

The table above shows the relationship between each factor and the heat shrinkage rate created based on engineer knowledge. When fitting the model, attention was paid to the residual test and multicollinearity. However, since the multicollinearity cannot be accurately confirmed, the standard error was estimated more robustly through the sandwich formula.

```
#---- Fit model
out <- model( formula = MDHS ~ IV + ETA1 + RELAX
              + poly(x = temp_S_7, degree = 3)
              + poly(x = temp_N_10, degree = 2)
              + temp_N_13
              , data = train_df)
```

Table 3: The Coefficients

	Estimate	Std(Robust)	t value	Pr(> t)	C.I.(Upper)	C.I.(lower)
Intercept	-0.360	0.207	-1.736	0.083	0.046	-0.765
IV	1.035	0.232	4.461	0.000	1.489	0.580
ETA1	0.424	0.033	12.957	0.000	0.488	0.360
RELAX	-0.039	0.009	-4.476	0.000	-0.022	-0.057
TEMP_S_7	-0.776	0.379	-2.045	0.041	-0.032	-1.520
(TEMP_S_7)^2	-0.663	0.160	-4.149	0.000	-0.350	-0.976
(TEMP_S_7)^3	-0.474	0.123	-3.867	0.000	-0.234	-0.714
TEMP_N_10	-3.291	0.230	-14.293	0.000	-2.839	-3.742
(TEMP_N_10)^2	-1.003	0.137	-7.294	0.000	-0.733	-1.272
TEMP_N_13	-0.008	0.001	-9.672	0.000	-0.006	-0.010

3. Prediction

Through the test results obtained through the Sandwich Formula, it was confirmed that there was no problem with the model. After this, I checked the prediction results. The predictive power was evaluated quantitatively and qualitatively. Quantitative evaluation was performed through MAPE and RMSE, and qualitative evaluation was performed through graphs.

```
predicted_value <- predict(object = out$model, newdata = df)
df$predicted_value <- predicted_value

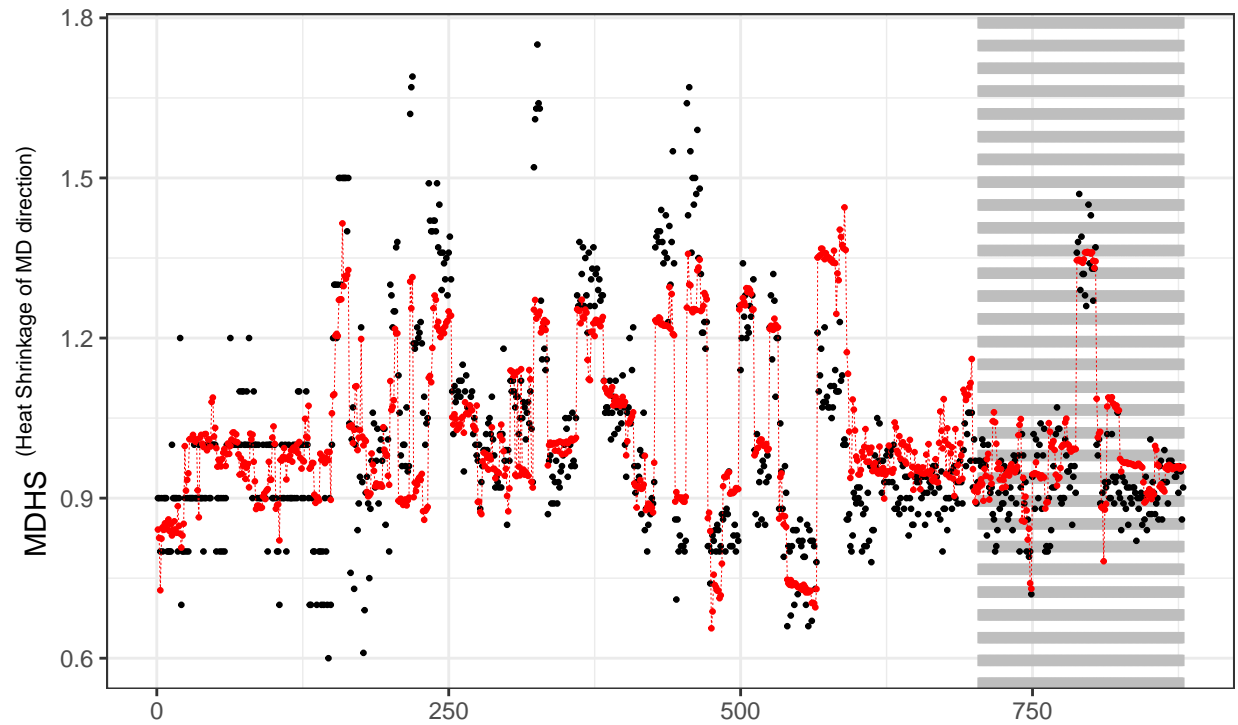
# Quantitative evaluation
eval_quantitative(data = df[valid_idx, ], actual = "MDHS", predict = "predicted_value")
```

MAPE: 9.07 & RMSE: 0.121

```
# Qualitative evaluation
df[, idx := 1:nrow(x = df)]
df %>%
  ggplot() +
  geom_vline( xintercept = valid_idx, linetype = "dashed", color = "grey", size = .5 ) +
  geom_point( mapping = aes( x = idx, y = MDHS), size = .5 ) +
  geom_point( mapping = aes( x = idx, y = predicted_value), color = "red", size = .5 ) +
  geom_line( mapping = aes( x = idx, y = predicted_value), linetype = "dashed"
    , color = "red", size = .1 ) +
  labs( title = "The Result of Regression"
    , subtitle = "Trend Chart of Actual value vs. Predicted value"
    , x = "", y = expression("MDHS " ^ (Heat Shrinkage of MD direction)) ) +
  theme_bw()
```

The Result of Regression

Trend Chart of Actual value vs. Predicted value



Reference

- Jung Gyu Lee. (2010). Investigation of Properties of the PET Film Dependent on the Biaxial Stretching, Polymer (Korea) v.34 no.6 , 2010, pp.579 - 587
- Cribari-Neto F. (2004). Asymptotic Inference under Heteroskedasticity of Unknown Form. Computational Statistics & Data Analysis 45, 215–233