

[Dacon] 블랙 장난감 제조 공정 최적화 경진대회

_____ (팀명)

2020년 월 일 (제출날짜)

1. 본 코드는 대회 참가를 돕고자 단순 예시를 작성한 것으로 참고용으로 사용바랍니다.
2. 본 코드는 자유롭게 수정하여 사용 할 수 있습니다.
3. 추가 모듈 보러가기: <https://bit.ly/36MNs76>

▼ 1. 라이브러리 및 데이터

Library & Data

```
from google.colab import drive
drive.mount('/content/drive/')
import sys
sys.path.append('/content/drive/My Drive/Colab Notebooks/LG')
```

➡ Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803

```
Enter your authorization code:
. . . . .
Mounted at /content/drive/
```

```
import pandas as pd
import numpy as np
import multiprocessing
import warnings
from copy import deepcopy
from module.genome import Genome, genome_score
warnings.filterwarnings(action='ignore')
np.random.seed(777)
```

```
!python --version
print('Pandas : %s'%(pd.__version__))
print('Numpy : %s'%(np.__version__))
```

➡ Python 3.6.9
Pandas : 1.0.5
Numpy : 1.18.5

▼ 2. 데이터 전처리

Data Cleansing & Pre-Processing

입력하세요.

▼ 3. 탐색적 자료분석

Exploratory Data Analysis

입력하세요.

▼ 4. 변수 선택 및 모델 구축

Feature Engineering & Initial Modeling

```
CPU_CORE = multiprocessing.cpu_count() # 멀티프로세싱 CPU 사용 수
N_POPULATION = 100                      # 세대당 생성수
N_BEST = 10                             # 베스트 수
N_CHILDREN = 10                         # 자손 유전자 수
PROB_MUTATION = 0.15                   # 돌연변이
REVERSE = False                        # 배열 순서 (False: ascending order, True: descending order)

score_ini = 1e8                         # 초기 점수
input_length = 125                      # 입력 데이터 길이
output_length_1 = 5                    # Event (CHECK_1~4, PROCESS)
output_length_2 = 12                   # MOL(0~5.5, step:0.5)
h1 = 50                                # 히든레이어1 노드 수
h2 = 50                                # 히든레이어2 노드 수
h3 = 50                                # 히든레이어3 노드 수
EPOCHS = 30                            # 반복 횟수

genomes = []
for _ in range(N_POPULATION):
    genome = Genome(score_ini, input_length, output_length_1, output_length_2, h1, h2, h3)
    genomes.append(genome)
try:
    for i in range(N_BEST):
        genomes[i] = best_genomes[i]
except:
    best_genomes = []
    for _ in range(5):
        genome = Genome(score_ini, input_length, output_length_1, output_length_2, h1, h2, h3)
        best_genomes.append(genome)
```

▼ 5. 모델 학습 및 검증

Model Tuning & Evaluation

1. PRT는 고정값 사용
2. Event A, Event B (MOL_A, MOL_B) 를 같은 값으로 제한
3. Event는 CHECK와 PROCESS 만 사용함
4. 목적 함수로 수요 부족분만 고려함
5. Event와 MOL에 대해 인공신경망 모델을 만들어 유전 알고리즘으로 학습

```

n_gen = 1
score_history = []
high_score_history = []
mean_score_history = []
while n_gen <= EPOCHS:
    genomes = np.array(genomes)
    while len(genomes)%CPU_CORE != 0:
        genomes = np.append(genomes, Genome(score_ini, input_length, output_length_1, output_length)
        genomes = genomes.reshape((len(genomes)//CPU_CORE, CPU_CORE))

    for idx, _genomes in enumerate(genomes):
        if __name__ == '__main__':
            pool = multiprocessing.Pool(processes=CPU_CORE)
            genomes[idx] = pool.map(genome_score, _genomes)
            pool.close()
            pool.join()
    genomes = list(genomes.reshape(genomes.shape[0]*genomes.shape[1]))

    # score에 따라 정렬
    genomes.sort(key=lambda x: x.score, reverse=REVERSE)

    # 평균 점수
    s = 0
    for i in range(N_BEST):
        s += genomes[i].score
    s /= N_BEST

    # Best Score
    bs = genomes[0].score

    # Best Model 추가
    if best_genomes is not None:
        genomes.extend(best_genomes)

    # score에 따라 정렬
    genomes.sort(key=lambda x: x.score, reverse=REVERSE)

    score_history.append([n_gen, genomes[0].score])
    high_score_history.append([n_gen, bs])
    mean_score_history.append([n_gen, s])

    # 결과 출력
    print('EPOCH # %s WtHistory Best Score: %s WtBest Score: %s WtMean Score: %s' % (n_gen, genomes[0].

```

```

# 모델 업데이트
best_genomes = deepcopy(genomes[:N_BEST])

# CHILDREN 생성
for i in range(N_CHILDREN):
    new_genome = deepcopy(best_genomes[0])
    a_genome = np.random.choice(best_genomes)
    b_genome = np.random.choice(best_genomes)

    for j in range(input_length):
        cut = np.random.randint(new_genome.w1.shape[1])
        new_genome.w1[j, :cut] = a_genome.w1[j, :cut]
        new_genome.w1[j, cut:] = b_genome.w1[j, cut:]

    for j in range(h1):
        cut = np.random.randint(new_genome.w2.shape[1])
        new_genome.w2[j, :cut] = a_genome.w2[j, :cut]
        new_genome.w2[j, cut:] = b_genome.w2[j, cut:]

    for j in range(h2):
        cut = np.random.randint(new_genome.w3.shape[1])
        new_genome.w3[j, :cut] = a_genome.w3[j, :cut]
        new_genome.w3[j, cut:] = b_genome.w3[j, cut:]

    for j in range(h3):
        cut = np.random.randint(new_genome.w4.shape[1])
        new_genome.w4[j, :cut] = a_genome.w4[j, :cut]
        new_genome.w4[j, cut:] = b_genome.w4[j, cut:]

    for j in range(input_length):
        cut = np.random.randint(new_genome.w5.shape[1])
        new_genome.w5[j, :cut] = a_genome.w5[j, :cut]
        new_genome.w5[j, cut:] = b_genome.w5[j, cut:]

    for j in range(h1):
        cut = np.random.randint(new_genome.w6.shape[1])
        new_genome.w6[j, :cut] = a_genome.w6[j, :cut]
        new_genome.w6[j, cut:] = b_genome.w6[j, cut:]

    for j in range(h2):
        cut = np.random.randint(new_genome.w7.shape[1])
        new_genome.w7[j, :cut] = a_genome.w7[j, :cut]
        new_genome.w7[j, cut:] = b_genome.w7[j, cut:]

    for j in range(h3):
        cut = np.random.randint(new_genome.w8.shape[1])
        new_genome.w8[j, :cut] = a_genome.w8[j, :cut]
        new_genome.w8[j, cut:] = b_genome.w8[j, cut:]

    best_genomes.append(new_genome)

# 모델 초기화
genomes = []
for i in range(int(N_POPULATION / len(best_genomes))):

```

```

for bg in best_genomes:
    new_genome = deepcopy(bg)
    mean = 0
    stddev = 0.2
    # 50% 확률로 모델 변형
    if np.random.uniform(0, 1) < PROB_MUTATION:
        new_genome.w1 += new_genome.w1 * np.random.normal(mean, stddev, size=(input_length,
    if np.random.uniform(0, 1) < PROB_MUTATION:
        new_genome.w2 += new_genome.w2 * np.random.normal(mean, stddev, size=(h1, h2)) * np
    if np.random.uniform(0, 1) < PROB_MUTATION:
        new_genome.w3 += new_genome.w3 * np.random.normal(mean, stddev, size=(h2, h3)) * np
    if np.random.uniform(0, 1) < PROB_MUTATION:
        new_genome.w4 += new_genome.w4 * np.random.normal(mean, stddev, size=(h3, output_le
    if np.random.uniform(0, 1) < PROB_MUTATION:
        new_genome.w5 += new_genome.w5 * np.random.normal(mean, stddev, size=(input_length,
    if np.random.uniform(0, 1) < PROB_MUTATION:
        new_genome.w6 += new_genome.w6 * np.random.normal(mean, stddev, size=(h1, h2)) * np
    if np.random.uniform(0, 1) < PROB_MUTATION:
        new_genome.w7 += new_genome.w7 * np.random.normal(mean, stddev, size=(h2, h3)) * np
    if np.random.uniform(0, 1) < PROB_MUTATION:
        new_genome.w8 += new_genome.w8 * np.random.normal(mean, stddev, size=(h3, output_le
    genomes.append(new_genome)

if REVERSE:
    if bs < score_ini:
        genomes[len(genomes)//2:] = [Genome(score_ini, input_length, output_length_1, output_le
    else:
        if bs > score_ini:
            genomes[len(genomes)//2:] = [Genome(score_ini, input_length, output_length_1, output_le

n_gen += 1

```



EPOCH	History Best Score	Best Score	Mean Score
EPOCH #1	67316015.0	67316015.0	76106259.
EPOCH #2	59235067.0	59235067.0	66200746.
EPOCH #3	59235067.0	59235067.0	62856576.
EPOCH #4	59235067.0	59235067.0	61320652.
EPOCH #5	55831136.0	55831136.0	58186653.
EPOCH #6	53882841.0	53882841.0	55976186.
EPOCH #7	53882841.0	53882841.0	55242928.
EPOCH #8	52558207.0	52558207.0	55019727.

6. 결과 및 결론

Conclusion & Discussion

EPOCH	History Best Score	Best Score	Mean Score
EPOCH #15	50444407.0	50444407.0	50474054.

결과 그래프

EPOCH	History Best Score	Best Score	Mean Score
EPOCH #10	52141427.0	52141427.0	52141427.

```
import matplotlib.pyplot as plt
```

```
# Score Graph
```

```
score_history = np.array(score_history)
```

```
high_score_history = np.array(high_score_history)
```

```
mean_score_history = np.array(mean_score_history)
```

```
plt.plot(score_history[:,0], score_history[:,1], '-o', label='BEST')
```

```
plt.plot(high_score_history[:,0], high_score_history[:,1], '-o', label='High')
```

```
plt.plot(mean_score_history[:,0], mean_score_history[:,1], '-o', label='Mean')
```

```
plt.legend()
```

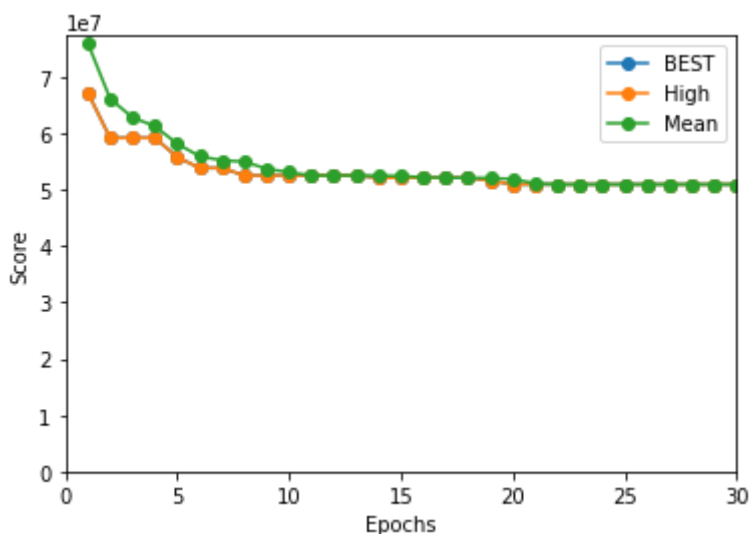
```
plt.xlim(0, EPOCHS)
```

```
plt.ylim(bottom=0)
```

```
plt.xlabel('Epochs')
```

```
plt.ylabel('Score')
```

```
plt.show()
```



Submission 파일 만들기

```

# 새고 계산
from module.simulator import Simulator
simulator = Simulator()
order = pd.read_csv('/content/drive/My Drive/Colab Notebooks/LG/module/order.csv')
submission = best_genomes[0].predict(order)
_, df_stock = simulator.get_score(submission)

# PRT 개수 계산
PRTs = df_stock[['PRT_1', 'PRT_2', 'PRT_3', 'PRT_4']].values
PRTs = (PRTs[:-1] - PRTs[1:])[24*23:]
PRTs = np.ceil(PRTs * 1.1)
PAD = np.zeros((24*23+1, 4))
PRTs = np.append(PRTs, PAD, axis=0).astype(int)

# Submission 파일에 PRT 입력
submission.loc[:, 'PRT_1':'PRT_4'] = PRTs
submission.to_csv('/content/drive/My Drive/Colab Notebooks/LG/Dacon_baseline.csv', index=False)

```

점수 향상 팁

해당 코드는 단순한 모델로 다음 방법으로 점수 향상을 꾀할 수 있습니다.

1. 성형 공정 2개 라인을 따로 모델링
2. CHANGE, STOP 이벤트 활용
3. 수요 초과분 외 다양한 양상을 반영하는 목적함수
4. 유전 알고리즘 외 효율적인 학습 기법