**Flowchat of Machine Learning Predict software**

**- Flowchat of software**

The software is largely divided into 2 parts, one is “Learning” part, and the other one is “Predict" part.

churn-training-aug-2016.csv

Attribute List.xlsx

Gen\_type\_list.py

Raw2real.py

Softmax\_bin.py

gen\_coef.py

Main\_predict\_without\_tensor.py

Training.csv

Model\_bin.ckpt

Coef.xls

Type.xlsx

Data\_prediction.xls

Arbitrary data

1. **“Learning” Part**

The learning part is outlined with a dashed line in the figure.

The goal of this part is get the learning coefficient by taking machine learning with source data.

* Gen\_type\_list.py

In this file, we get the “type.xls” file used in preprocessing step from sourc files.(“churn-training\_aug...”, “Attribute List.xls”)

“Attribute List.xls” is created manually to reflect what type of preprocessing should proceed according to the format of the every column data.

0: don’t use this column.

1: use it as it is.

2: rescale the numerical data(in our case, we don’t this item)

3: convert the string to integer data.

The result of this python code, “type.xls” contains reference material for preprocessing the source raw data in the next step.

* Raw2real.py

In this file, we convert the raw data to normalized data.

We convert the string data in raw data to numerical data in order to use it in machine learning stage.

As a result, “Training.csv” is created.

As you see this file, string data converted to numerical data.

* Softmax\_bin.py

Here, we use the normalized training data “Training.csv” to conduct machine learning and store the coefficient of model as a binart file “model\_bin.ckpt”.

In this step, we use the “TensorFlow”, machine learning library.

* Gen\_coef.py

In this file, we convert the binary model coefficient file to xls document “coef.xls”.

1. **“Predict” Part**

In this part, we predict any source data using model coefficient data which obtained in “learning” part.

* Main\_predict\_without\_tensor.py

This file uses arbitrary source data to be input and predicts the required value and output as an xls document.

In this step, we need only “type.xls”, “coef.xls” and predicted source data.

We don’t need Tensorflow and other files such as “training.csv”, “model\_bin.ckpt” and so on.