

D-15 maf: 機械学習・データマイニング実験ビルドシステム

得居誠也(Preferred Networks (PFN))、能地宏(総研大/NII)、大野健太(PFN)

```
def build(bld):
```

```
    NUM_FOLD = 10
```

```
    bld(source='data/svmguide3',
        target='data/train data/test',
        parameters=[{'fold': i} for i in range(NUM_FOLD)],
        rule=maflib.rules.segment_by_line(NUM_FOLD, 'fold'))
```

パラメータ作成

交差検定用に
データ分割

```
    bld(source='data/train',
        target='data/model log/train',
        parameters=maflib.util.product({
            'type': [1, 2, 3, 4, 5, 6, 7],
            'cost': [0.01, 0.1, 1, 10, 100]}),
        rule='liblinear-train -s ${type} -c ${cost} ${SRC} ${TGT[0].abspath()} > ${TGT[1].abspath()}.al')
```

パラメータ作成

訓練

```
    bld(source='data/test data/model',
        target='result/predict log/test',
        rule='liblinear-predict ${SRC} ${TGT[0].abspath()} > ${TGT[1].abspath()}')
```

テスト

```
    bld(source='log/test',
        target='result/accuracy',
        rule=maflib.rules.convert_libsvm_accuracy)
```

集計

```
    bld(source='result/accuracy',
        target='result/average_accuracy',
        aggregate by='fold',
        rule=maflib.rules.average)
```

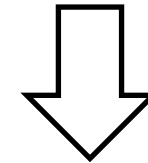
パラメータ集約

```
    bld(source='result/average_accuracy',
        target='result/plot.png',
        for each=[],
        rule=plot)
```

パラメータ集約

プロット

実験を簡潔に書いて
メンテナンス性も保ちたい



実験を短く書ける DSL と
その実行環境を提案

発表内容

- ◆ 仕組みの紹介
- ◆ 3つの活用例
 - ✓ LIBLINEAR
 - ✓ 自然言語処理 (パーズング)
 - ✓ 深層学習 (Caffe)

Get at: <https://github.com/pfi/maf>