



データサイエンティスト実践セミナー Python/Sparkで始める分析の実際

ディープラーニングで名刺を解析する

Sansan株式会社

高際睦起



発表の内容と目的

- 開催中の画像認識コンペ紹介
 - 名刺に書かれた項目(名前、会社名、、、)を推定

- Deep Learningを使ったベンチマークコード公開中
 スコアボードのbenchmarkとは違います。おそらく0.015~0.02程度のスコア(未計測)
 https://github.com/takagiwa-ss/deepanalytics_compe26_benchmark
 最新の手法を出来るだけ平易に書いたつもりです
 - ※弊社Sansanで実際に使っている手法とは異なります

• ベンチマークコードを例に画像認識の方法を解説



> コンペ内容の紹介



〉データ

名刺画像











矩形情報・正解ラベルデータ

filename	left	top	right	bottom	company_	full_name	position_	address	phone_nu	fax	mobile	email	url
2842.png	491	455	796	485	0	0	0	0	0	0	1	0	0
182.png	24	858	311	886	0	0	0	0	0	0	1	0	0
95.png	320	498	865	521	0	0	C	0	0	1	1	0	0



> 訓練用のデータ = 例題

4	Α	В	С	D	Е	F	G	Н	1	J	K	L	M	N
1	filename	left	top	right	bottom	company	full name	position	address	phone nu	fax	mobile	email	url
2	2842.png	491	455	796	485	0	0	0	0	0	0	1	0	0
3	182.bng	24	858	311	886	0	0	0	0	0	0	1	0	0
+	95.png	320	498	865	521	0	0	0	0	0	1	1	0	0
5	2491.png	65	39	497	118	1	0	0	0	0	\ 0	0	0	0
6	3301.png	271	83	333	463	0	1	1	0	0	0	0	0	0
7	3240.prg	441	45	985	96	1	0	0	0	0	0	0	0	0
8	1102.png	338	197	637	259	0	1	0	0	0	0	0	0	0
9	2141.png	344	817	588	843	0	0	0	0	1	0	0	0	0
10	3595.png	354	262	674	326	0	1	0	0	0	þ	0	0	0
11	1787.png	156	224	389	249	0	0	0	0	0	1	0	0	0
12	2022.png	58	43	813	103	1	0	0	0	0	0	0	0	0
13	3577.png	517	266	789	329	0	1	0	0	0	0	0	0	0







> 評価

\angle	Α	В	С	D	Е	F	G	
1		filename	left	top	right	bottom		
2	0	1942.png	57	114	361	173		
3	1	1128.png	58	3/3	519	422		
4	2	2719.png	62	289	297	314		
5	3	641.png	58	668	416	747		
6	И	2529 nnø	12	212	202	244		
					〒150-000 東京都渋谷原 TEL 03-67 FAX 03-34 MOBILE 09 E-mail dun SITE http://	X神宮前 5-52- /58-0033 /409-3133 90-1234-567 nmy092@sai //jp.corp-sar	2青山オーバルビル 13F 78 nsan.co.jp	

評価用データは、画像と矩形座標のみ、項目ラベルは無しコンペ参加者には項目ラベルを予測して頂きます 9種の項目それぞれ0~1の実数



Deep Learningを使った名刺の画像認識の方法 と ベンチマークコードの解説



> ベンチマークコードの処理概要

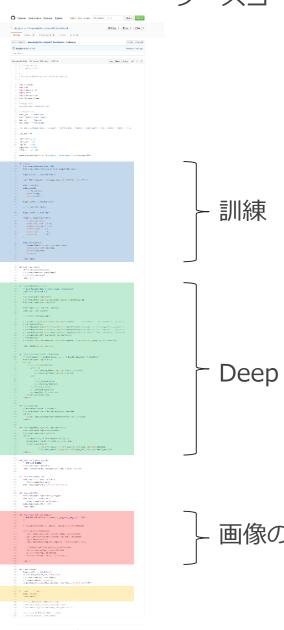
- 訓練用のcsv, png画像を読み込み、 訓練データとして格納
 - csvからpngファイル名と矩形座標、及び項目ラベルを得る
 - 訓練に使えるように画像を加工

- 2. 学習に使うDeep Learningモデルを定義し、 訓練データでモデルを訓練
 - ライブラリkerasを使う

3. 評価用のcsv, png画像を読み込み、予測を出力



ソースコード概観



Deep Learningモデルの定義

画像の読み込み加工



> ベンチマークコードの処理概要

- 訓練用のcsv, png画像を読み込み、 訓練データとして格納
 - csvからpngファイル名と矩形座標、及び項目ラベルを得る
 - 訓練に使えるように画像を加工

- 2. 学習に使うDeep Learningモデルを定義し、 訓練データでモデルを訓練
 - ライブラリkerasを使う

3. 評価用のcsv, png画像を読み込み、予測を出力



```
C Personal Commonater Backers: Eighers: 19675 Die School Philosophie, Sont Marie Park
                  Esdagina a Magandytia, compdé boxbook devan 1 dons 1 von 1
                            **** ***** despendy for compact benchmark: training
                                                                          Jackson - Tacchage
Jackson - Europe
Jackson - M
                                                                                                                to tract!

for trace-delice, law, IE

for trace-represent plays beet legitial-cons
                                                                                                                Egglyceld in bedge of a
in the real reng
Egglyceld in test by t
                                                                                                                         the sect
                                                                                                                                  The population of the control of the
                                                                                                                         The transference of the Control Section (Control of the Control of
                                                                                                                                                               a from broad raw, maps
                                                                                                                                  making Milan, gutoni, managinating 
making apper layer amountain 
for the making both of 
an extra 
managine of the first managinating by the 
managine of the first managine of the 
managine of the second of the 
managine of 
managine 
managine of 
managine of 
managine 
managine

    Allered Silver of new silve of note silve silver selections.
    Allered Silver of new silver of Ling Silver selections.

                                                                                             and places and deligned and places are replaced and another anothe
```

```
def _load_rawdata(df, dir_images):
166
         '''画像を読み込み、4Dテンソル (len(df), 1, _img_len, _img_len) として返す
168
169
         X = np.zeros((len(df), 1, _img_len, _img_len), dtype=np.float32)
170
171
         for i, row in df.iterrows():
172
             img = Image.open(os.path.join(dir_images, row.filename))
173
             img = img.crop((row.left, row.top, row.right, row.bottom))
174
             img = img.convert('L')
175
             img = img.resize((_img_len, _img_len), resample=Image.BICUBIC)
176
177
             # 白黒反転しつつ最大値1最小値0のfloat32に画素値を正規化
178
179
             img = np.asarray(img, dtype=np.float32)
180
             b, a = np.max(img), np.min(img)
181
             X[i, 0] = (b-img) / (b-a) if b > a else 0
182
183
         return X
```



> 画像を扱う定番の方法

1. 本コンペのように、画像中の一部の矩形領域を使う場合、 矩形内の画像を切り抜く

画像サイズを合わせる
 矩形のサイズは**ばらばら** リサイズして大きさをそろえ、扱いやすくする

3. 値のレンジをそろえる輝度値(0~255のレンジ)を浮動小数(0~1)に変換Deep Learningの計算を安定化

このコードではやってないが、平均値を引く、分散でわる、こともよくある



> 画像→訓練データの具体例





MOBILE 090-1234-5678











> 画像→訓練データ(1)画像切り抜き

ATT株式会社
企業情報本部
部長

久胡 アユカ

〒150-0001
東京都渋谷区神宮前 5-52-2
青山オーバルビル 13F
TEL 03-6758-0033
FAX 03-3409-3133
MOBILE 090-1234-5678
E-mail dummy142@sansan.co.jp
SITE http://jp.corp-sansan.com/

PILのcropで矩形座標を指定して 画像切り抜き



MOBILE 090-1234-5678











> 画像→訓練データ(2)サイズ変更

ATT株式会社

部長

久胡 アユカ

〒150-0001

東京都渋谷区神宮前 5-52-2

青山オーバルビル 13F

TEL 03-6758-0033

FAX 03-3409-3133

MOBILE 090-1234-5678

E-mail dummy142@sansan.co.jp

SITE http://jp.corp-sansan.com/



MOBILE 090-1234-5678

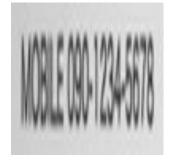


PILresizeで固定サイズに変換

長方形を無理やり正方形にする

Deep Learningでは正方形にする のはよくやる方法

歪み激しいが学習してみたら そこそこ精度出たのでまぁいいか









> 画像→訓練データ(3)レンジの正規化

ATT株式会社

企業情報本

久胡 アユカ

〒150-0001

東京都渋谷区神宮前 5-52-2

青山オーバルビル 13F

TEL 03-6758-0033

FAX 03-3409-3133

MOBILE 090-1234-5678

E-mail dummy142@sansan.co.jp

SITE http://jp.corp-sansan.com/

pixel値の範囲(0~255)を 0~1の浮動小数に変換、白黒反転 float32を使うのはGPUの都合

値のレンジをそろえないと 計算が発散して学習がうまく うごかないことがある

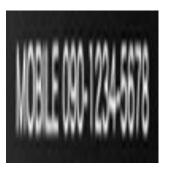


MOBILE 090-1234-5678





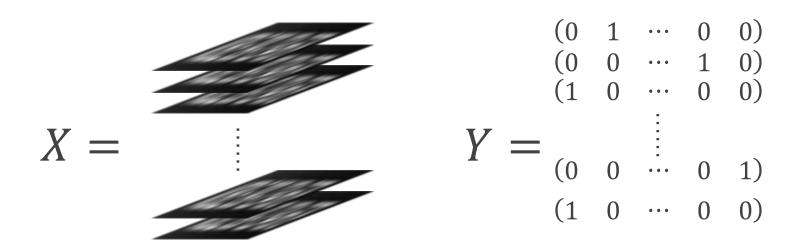






> 画像→訓練データ

全画像を読み込んで、全画像(X)と項目ラベル(Y)を積み重ねて 学習データとする





> ベンチマークコードの処理概要

- 訓練用のcsv, png画像を読み込み、 訓練データとして格納
 - csvからpngファイル名と矩形座標、及び項目ラベルを得る
 - 訓練に使えるように画像を加工

- 2. 学習に使うDeep Learningモデルを定義し、 訓練データでモデルを訓練
 - ライブラリkerasを使う

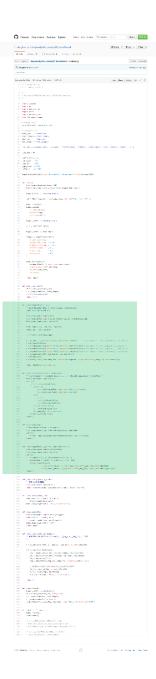
3. 評価用のcsv, png画像を読み込み、予測を出力



```
Spin form
O Female Communication Explanal Explanal Prints Day Science Philippiness Service
                  Esalugina is Edecparalytics, competit, broadmark
                                                                                        K-market broken bedre bedre farmer Section
                                                                                                           or t<sub>i</sub> maybe a description of the control of the c
                                                                           make - 1 may
```

```
def train():
38
        from keras.optimizers import SGD
        from keras.preprocessing.image import ImageDataGenerator
40
        logging.info('... building model')
41
42
        sgd = SGD(1r=_sgd_1r, decay=_sgd_decay, momentum=0.9, nesterov=True)
43
44
45
        model = resnet()
                                                Deep Learningのモデル
46
        model.compile(
47
           loss= objective,
                                                と最適化の方法を定義
48
           optimizer=sgd,
           metrics=['mae'])
49
50
        logging.info('... loading data')
51
52
        X, Y = load_train_data()
53
54
55
        logging.info('... training')
56
        datagen = ImageDataGenerator(
57
58
           # data augmentation
           width_shift_range = 1./8.,
           height_shift_range = 1./8.,
                                              "データ拡張"の機能を使う
           rotation_range
61
                             = 0.,
62
                             = 0.,
           shear_range
                             = 0.,
63
            zoom_range
64
65
        model.fit_generator(
67
            datagen.flow(X, Y, batch size= batch size),
                                                      訓練を行う
           samples_per_epoch=X.shape[0],
69
           nb_epoch=_nb_epoch,
70
           verbose=1)
71
72
        return model
```



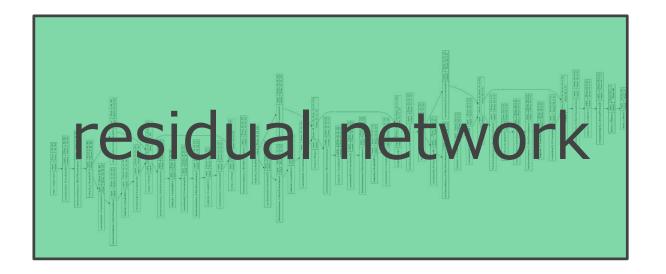




> Deep Learningのモデル定義

residual networkはDeep Learningの一種

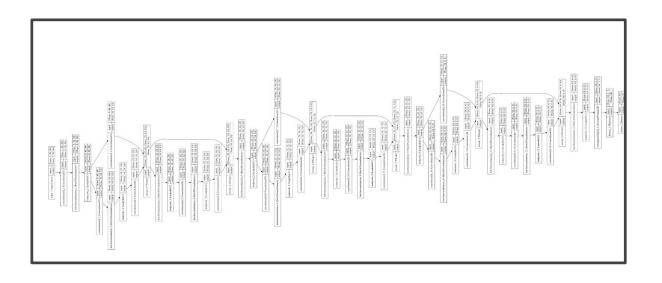
2016年時点、画像認識でトップの性能をもつ





ライブラリkerasを使って、Deep Learningモデルを定義

□と→の繋がり方を作る

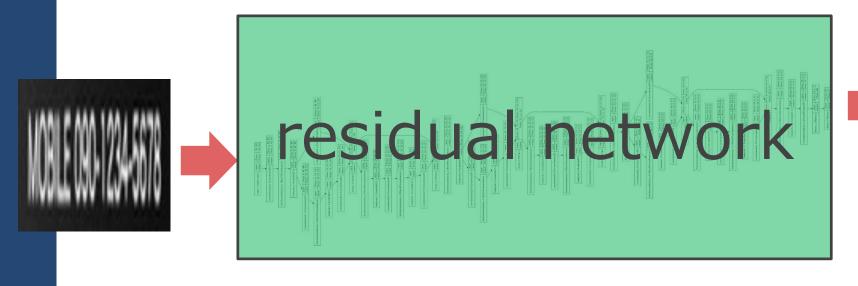


この発表では詳しい説明は省略



> モデルの定義

固定サイズの行列 = 画像(96x96, グレー画像)を入力すると



0.146 0.128 0.070 0.044 0.100 0.937 0.064

0.0687

項目(9種)ごと"らしさ"=予測を9次元ベクトル0~1で出力

と、するようにresidual networkのモデルを作る



```
Spin form
Commont Commonte Existent Eightes - Febru Rich Colons - Telegraphics - Colons
                        Esologina is Ederprindytis jumpid 6 jumbriok
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Gross I Was I Yes I
                                                                                 A soul - br

and - br

and - 64

and - 64
                                                                                         and place and other times are a program of the prog
                                                                                         ar Joseph art September (1994) (1994) (1994) (1994)
                                                                                                                  K-market broken bedre bedre farmer Section
                                                                                                                                          or t<sub>i</sub> maybe a description of the control of the c
                                                                                                  make - 1 may
```

```
def train():
38
        from keras.optimizers import SGD
        from keras.preprocessing.image import ImageDataGenerator
40
        logging.info('... building model')
41
42
        sgd = SGD(1r=_sgd_1r, decay=_sgd_decay, momentum=0.9, nesterov=True)
43
44
45
        model = resnet()
                                                Deep Learningのモデル
46
        model.compile(
47
            loss= objective,
                                                 と最適化の方法を定義
48
            optimizer=sgd,
            metrics=['mae'])
49
50
        logging.info('... loading data')
51
52
        X, Y = load_train_data()
53
54
55
        logging.info('... training')
56
        datagen = ImageDataGenerator(
57
58
            # data augmentation
            width_shift_range = 1./8.,
                                               kerasの"データ拡張"の
            height_shift_range = 1./8.,
            rotation range
61
                             = 0.,
62
                             = 0.,
            shear_range
                             = 0.,
63
            zoom_range
64
65
        model.fit generator(
67
            datagen.flow(X, Y, batch size= batch size),
                                                      訓練を行う
            samples_per_epoch=X.shape[0],
69
            nb_epoch=_nb_epoch,
70
            verbose=1)
71
72
        return model
```



〉学習

fit_generatorを呼んで、画像と項目ラベルの対応を覚えさせる = 学習

0.068 0.146 0.128 0.070 0.044 0.100 0.937 0.064

ـ0.231 ا

出力

出力と正解ラベルの差を縮めるように networkを更新する

正解ラベル

Copyright © 2014 Sansan, Inc. All rights reserved.



> ベンチマークコードの処理概要

- 訓練用のcsv, png画像を読み込み、 訓練データとして格納
 - csvからpngファイル名と矩形座標、及び項目ラベルを得る
 - 訓練に使えるように画像を加工

- 2. 学習に使うDeep Learningモデルを定義し、 訓練データでモデルを訓練
 - ライブラリkerasを使う

3. 評価用のcsv, png画像を読み込み、予測を出力



```
Comment Communication Explorer (News Day Scient Philippelling Section Majories Parks
                                  Esdegina a Megandytia compilé juntimak 690m 1 | $100 1 | $100 1
                                                                                                                                                                                                                               100,000

    managements and a management of the management of the
                                                                                                                                                                                                                               The transference of the Control Section (Control of the Control of

    Transit States Care, modulose.
    Transitada o resident discussion o facilità ministra o transferi
della transferia della facilità della companya dell
                                                                                                                                                                                                             and manufacturing against processing and accommodate to the contract against figure and a processing and a contract against the contract accommodate a

    All and Allien is now that it is to use all and a second control of the control of 
                                                                                                                                                                                                                                                                                                   or types for a farmously 
and a deep control and condition research to the control 
and a deceasable of the condition resolution resoluted? 
by a deposition (201) and only a control and 
only a deposition (201), and proof, consequence and 
only a deposition (201), and only a consequence and 
and only and an extension of the control of the control of the 
control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control of the control o
                                                                                     10 of the 17 cm;
10 mater (select)
```

```
def predict(model):
    logging.info('... predicting')
    df = pd.read_csv(test_csv, index_col=0)

    X = _load_rawdata(df, test_images)
    P = model.predict(X, verbose=1)
    pd.DataFrame(P).to_csv(fn_prediction, header=False, float_format='%.6f')

192
```

- 1. 訓練の時と同じように評価用の画像を加工
- 2. 学習したモデルで predict を呼ぶと全予測を行列として 出力
- 3. 行列をcsvファイルに書き出し



> ベンチマークコードの改良ポイント

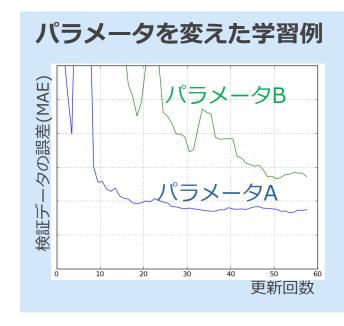
※公開したベンチマークコードコードで用いているパラメータ類は最適値ではありません※

- 訓練データを分割(80%:20%)し、 訓練用(80%)と検証用(20%)として使えるようにする
 - 訓練に使っていない検証データで精度を評価する
 - 次ページ以降のようなパラメータをチューニングするときに、手元で 精度評価して、パラメータの良し悪しを判断



> ベンチマークコードの改良ポイント

- 最大更新回数(_nb_epoch) を増やす
 - モデルを更新するループの回数
 - ベンチマークでは20回の更新で終了 最適値で無い
 - ベストの値は他のパラメータや 最適化方法にも依存



- 正規化画像サイズ(_img_len)を大きく
 - ベンチマークの96x96の画像では文字が小さい
 - 辺長128~512程度がよさそうかな? コードの都合上 16 の倍数



> ベンチマークコードの改良ポイント

- Deep Leaningのモデルを深く、幅広く
 - 構造を定義するパラメータ2個(trainer.pyの45行目: model = resnet())
 - > 深さパラメータ(repetition=2) □の数
 - > 幅広さパラメータ(k=1) □の大きさ
 - repetition=2~5, k=2~8程度がよさそうかな?
 - 過適合を抑える
 - > Dropout
 - > データ拡張(trainer.pyの59~63行目: ImageDataGenerator) みかけの訓練データ数を増やす
 - Deep Learning職人とは、Deep Learningの気持ちを理解してパラ メータチューニングがうまくできる人



開催期間:2016年8月8日~9月30日

懸賞金: 1位 **30万** 2位 **20万** 3位 **10万**



日本アイ・ビー・エム株式会社



ご参加お待ちしております



> Bibliography

- Keras https://keras.io/
 - <u>https://github.com/fchollet/keras/tree/master/examples</u> にある畳み込みニューラルネットワーク例
 - > mnist cnn.py 手書き数字0~9の画像認識
 - > <u>cifar10_cnn.py</u> 一般物体10種(犬、猫、船、…)の画像認識
 - > <u>resnet_50.py</u> 一般物体認識の学習済みモデルを使う例
- Deep Learning(Residual Network)関連
 - Deep Residual Learning for Image Recognition (2015)
 Identity Mappings in Deep Residual Networks (2016)
 Wide Residual Networks (2016)
 Weighted Residuals for Very Deep Networks (2016)
 - Residual Network作者Kaiming HeのICML2016での解説スライド http://icml.cc/2016/tutorials/icml2016 tutorial deep residual networks kaiminghe.pdf
 - 産総研片岡さんの解説スライド http://www.slideshare.net/HirokatsuKataoka/deep-residual-learning-ilsvrc2015-winner