

JAWS DAYS 2015

IoTデバイスとクラウドを使って センサーハック～機械学習～

株式会社プリファードインフラストラクチャー

自己紹介

- 名前：河原 一哉（37歳：おっさん）
- 所属：株式会社Preferred Infrastructure
- 好きなOS：Solaris
- 好きなマイコン：ATMEGA3238P-AU
- 好きな護衛艦：ごくりゅう



Preferred Infrastructure.

よろしくおねがいします！



ハンズオンのシナリオ

- 今日の目標
 - 加速度センサーの値を使ってPCのキーボード入力を監視しよう
- 今日の作業
 - シナリオ0：センサーからデータを取得してみる
 - シナリオ1：学習データを作成して精度測定する その1
 - シナリオ2：学習データを作成して精度測定する その2
 - シナリオ3：実際に分析を実施してみる
 - センサーデータをkinesisに送る
 - センサーデータをkinesisから受け取る



ハンズオンのシナリオ

- 今日使うセンサー
 - 3軸加速度センサー ADXL345
 - I2C通信対応
 - 電源電圧 : 3.0V ~ 5V DC
 - 待機時消費電力 : 0.1µA



 Preferred Infrastructure.

ハンズオンのシナリオ

- 今日使う環境
 - Intel : Edison
 - Amazon : EC2, Kinesis
 - PFI : Sedue for BigData, Sedue Predictor




Sedue
for BigData


Predictor



Amazon EC2

 Preferred Infrastructure.

準備

- 今日使うプログラムとかのダウンロード
 - ここからダウンロード：
 - https://github.com/kawaii/jaws2015_hands_on
- AWS環境の準備
 - **JAWS DAYS 2015 HackDay AWS側構築ハンズオンテキスト**
 - HackDay-Track3-Script-Rev0.1.pdf
 - 動作確認：<http://hostname>



今日使うPFI製品

- Sedue for BigData
 - リアルタイム大規模データ分析基盤
 - 高速で漏れの無い全文検索が得意
 - 今日はリアルタイムな部分だけ使います。
- Sedue Predictor
 - オンライン機械学習プラットフォーム
 - 特徴抽出、学習、分類、精度評価を簡単に
 - 今日はこっちを主に使います。



機械学習とは？

- 機械学習とは
 - 「明示的にプログラミングすることなく、コンピュータに行動させようとする科学」
 - 学習により、自動的にルールを見つけ出し、分類や予測などを行なえるようになる機能
- 学習の種類
 - 教師あり学習：人間が明示的に正解を教える
 - 教師なし学習：明示的な正解は与えない
 - 半教師あり学習、強化学習とかも。。。。



機械学習とは？

- 機械学習を使うために必要な事
 - データを作成する
 - 学習データを作成する
 - 特徴抽出の方法を考える
 - 学習をする
 - 精度を測定する
 - 機械学習を使用する



機械学習を使ってなにをするか？

- 機械学習はいろんな事ができる
 - 分類：カテゴリに分ける
 - 回帰：数値を予測する
 - クラスタリング：いくつかに分ける
 - 異常値検知：異常値を見つける
- 今回の問題設定
 - 加速度センサーの値を読み取って、以下の状態に分類してみる
 - キー入力している状態
 - キー入力していない状態



シナリオ0：センサーからデータを取得してみる

- センサーを調べる
 - ADXL345というチップを搭載
 - ライブラリを探す
 - <https://software.intel.com/en-us/iot/sensors>
- センサーを繋げる
 - 今回のセンサーはI2C接続なので、D I2Cポートに繋げる
- センサーからデータを取得するプログラムを書く
 - サンプルプログラムを見れば大体わかる。
 - ライブラリのおかげで超簡単



シナリオ0：センサーからデータを取得してみる

- Edisonへログイン
 - ssh root@Edisonのホスト名
- 色々インストール（入ってる人は不要）
 - # echo "src intel-iotdk http://iotdk.intel.com/repos/1.1/intelgalactic" > /etc/opkg/intel-iotdk.conf
 - # opkg update
 - # opkg upgrade
 - # node npm install aws-sdk



シナリオ0：センサーからデータを取得してみる

- プログラム作成
 - 以下のURLのサンプルを参考に作ってみよう
 - <https://software.intel.com/en-us/node/541816>
- プログラムの実行
 - # node プログラム名
 - 値が取れた！はず？



シナリオ0：センサーからデータを取得してみるまとめ

- シナリオ0の結論
 - センサーデータから値を取るのは簡単
 - ただしライブラリがあれば。無いと結構大変です。
→ ライブラリのあるセンサーを使用しましょうw
(ライセンスには注意してください。)
 - 取れたセンサーの値を見ても何も判らない
 - 可視化しないとなんのこっちゃ判らない。
 - 可視化しても何も判らない
 - 値が上下してるぐらいしか判りません。。。
 - 熟練すると判るらしいけど普通は無理。

→ 機械学習を使ってデータの分析をやってみよう！



シナリオ1：学習データ作成、精度測定 その1 Predictorの準備

- Sedue Predictorの準備
 - EC2インスタンスの80番ポートにHTTPアクセス
 - Sedue Predictorをクリック
 - ログイン
 - 使用アカウント : demo@demo.local / jaws2015
 - モデルの作成
 - new modelボタンをクリックしモデルを作成
 - 一緒に1個コンフィグが作られる



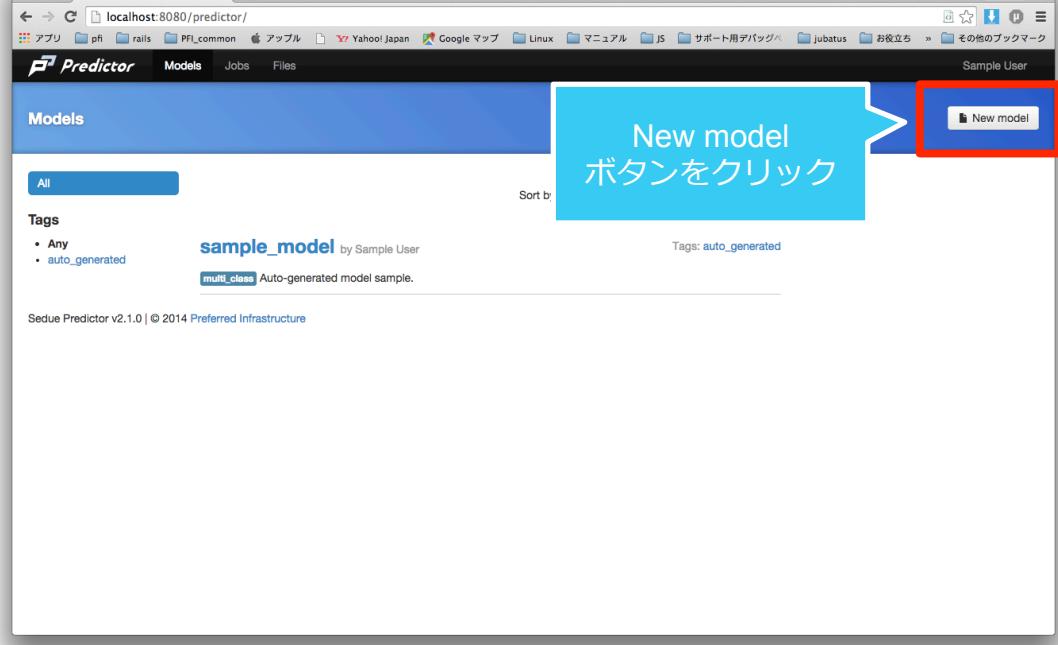
シナリオ1：学習データ作成、精度測定 その1 Predictorの選択

The screenshot shows the homepage of the Sedue Predictor service. The title is "Sedue製品デモ仮想環境" (Sedue Product Demo Virtual Environment). It features a "Learn More" button and a callout pointing to the "Predictor" icon, which is highlighted with a red border. The Predictor icon consists of a stylized blue 'P' logo with the word "Predictor" below it. To the right, there are other icons: "Sedue for BigData" (blue globe icon) and "Sedue Manager" (blue planet with rings icon). The Preferred Infrastructure logo is at the bottom right.

シナリオ1：学習データ作成、精度測定 その1 ログイン画面

The screenshot shows the login page for the Sedue Predictor service. The URL is 54.199.150.76/predictor/users/sign_in. The page features the Predictor logo and a form for entering Email and Password. A blue callout box contains the login credentials: ID : demo@demo.local and PASS : jaws2015. The Preferred Infrastructure logo is at the bottom right.

シナリオ1：学習データ作成、精度測定 その1 モデルの作成

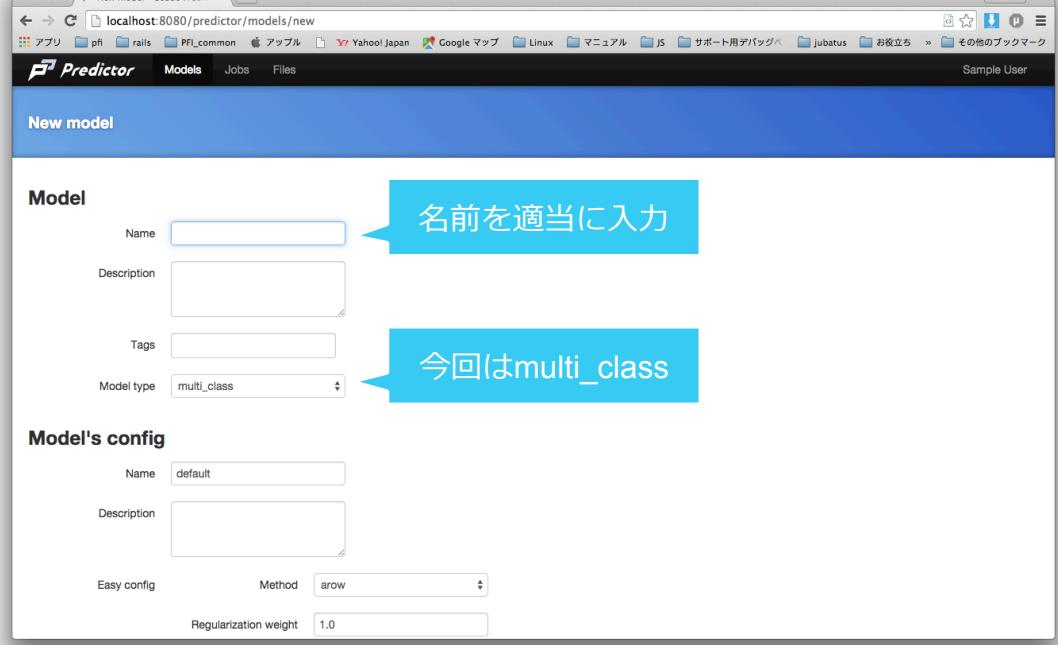


New model
ボタンをクリック

Sedue Predictor v2.1.0 | © 2014 Preferred Infrastructure

Preferred Infrastructure.

シナリオ1：学習データ作成、精度測定 その1 モデルの作成



Name

Description

Tags

Model type

Model's config

Name

Description

Easy config Method

Regularization weight

Preferred Infrastructure.

シナリオ1：学習データ作成、精度測定 その1 モデルの作成

Name: default

Description:

Method: arrow

Regularization weight: 1.0

String rule: space

Num rule: num

Fields: Use all fields Select fields

Batch train config:

- Train method: using_all_data
- Train times: 1

Create model

Sedue Predictor v2.1.0 | © 2014 Preferred Infrastructure

 Preferred Infrastructure.

シナリオ1：学習データ作成、精度測定 その1 モデルの作成

Configs

ID	Config name	Created at	Default config	Retrain	Last retrain status	Duplicate	Description
3	arow_num_all	15-03-21 18:59:50	<input checked="" type="button"/> Default	<input type="button"/> Retrain	Not Trained	<input type="button"/> Duplicate	
4	nherd_num_all	15-03-21 19:00:21	<input type="button"/> Set default	<input type="button"/> Retrain	Not Trained	<input type="button"/> Duplicate	

Model info

Model name (#id)	demo (#3)
Description	(No description)
Model type	multi_class
Owner	Sample User
Tags	

Download training data

With IDs	Status	Result	With ID	With Predictions	Encoding	File created at	
<input type="checkbox"/>	mint_documents-v...tar.gz	2015-03-17_1905.zip	<input type="checkbox"/>	2015-03-16_1633.zip	<input type="checkbox"/> noname	<input type="checkbox"/> put	<input type="checkbox"/> すべてを表示

 Preferred Infrastructure.

シナリオ1：学習データ作成、精度測定 その1 学習データの作成（Edison上の作業）

- 学習データの作成
 - センサーの値を1回取得し、その際に、キー入力があったか無かったか正解データを付ける
 - PredictorはCSV形式でGUIから学習データを登録できる
 - # cd hands_on/edge/session_1
 - # node trainer.js > train.data
 - このプログラムはCSV形式でデータを作成します。
 - コンソール上でキー入力（強く叩きすぎないで。弱すぎもダメ）
 - 1000個学習データが出来ると終了。
 - ヘッダー情報を付ける
 - # cat header.txt train.data > train.csv
 - ローカルPCにファイルをコピー



シナリオ1：学習データ作成、精度測定 その1 Predictorの設定

- 学習データの登録
 - training dataを押す
 - upload fileボタンをクリック
 - ファイルを選択、アップロード
- コンフィグの作成
 - いろんなコンフィグを作ってみる
- 精度評価
 - Evaluationを選択
 - target configでall configを選択、他はデフォルトでOK
 - submitボタンを押す
 - Accuracyを確認



シナリオ1：学習データ作成、精度測定 その1 学習データの登録

The screenshot shows the Predictor web application interface. The top navigation bar includes links for 'Models', 'Jobs', and 'Files'. The main content area is titled 'demo (#3)'. A red box highlights the 'Training Data' tab, which is currently active. Below it, there's a section titled 'Config' with a table showing two rows of data. To the right of the table is a large blue callout box with the text 'Training Data をクリック'. Further down, there's a 'Model' configuration section with fields for 'Model name (#id)', 'Description', 'Model type', 'Owner', and 'Tags'. At the bottom, there's a 'Download training data' section with a file list table.

Preferred Infrastructure.

シナリオ1：学習データ作成、精度測定 その1 学習データの登録

The screenshot shows the Predictor web application interface, specifically the 'Training data' section for 'sample_model (#1)'. A red box highlights the 'Upload file' button in the top right corner of the main content area. Below it, there's a search bar and a dropdown menu for 'Using config: default'. The main table area is currently empty, showing 'Showing 0 - 0 of 0 results'. At the bottom, there's a footer note: 'Sedue Predictor v2.1.0 | © 2014 Preferred Infrastructure'.

Preferred Infrastructure.

シナリオ1：学習データ作成、精度測定 その1 学習データの登録

The screenshot shows the Predictor web application interface. The main title is "sample_model (#1) / Training data / Upload". Below it, there are tabs: Configs, Training Data (which is selected), Annotations, Fill Annotation, Evaluation, and Inspection. A large blue button labeled "Upload" is present. To its left, there's a "CSV file" input field with a placeholder "ファイルを選択" (Select file). A red box highlights this input field. Below the input field is a message: "ファイルを選択して Uploadボタンをクリック". To the right of the input field is a table showing upload history:

File	Created at	Started at	Finished at
hands_on5.csv	15-03-19 10:11:21	15-03-19 10:11:24	15-03-19 10:11:27
hands_on5.csv	15-03-16 11:07:44	15-03-16 11:07:46	15-03-16 11:07:50
hands_on2.csv	15-03-16 11:06:55	15-03-16 11:06:56	15-03-16 11:06:57

At the bottom, a success message says "200 data added and 0 data updated." The footer of the page includes "Sedue Predictor v2.1.0 | © 2014 Preferred Infrastructure".

シナリオ1：学習データ作成、精度測定 その1 学習データの登録

The screenshot shows the Predictor web application interface. The main title is "demo (#3) / Training data". Below it, there are tabs: Configs, Training Data (which is selected), Annotations, Fill Annotation, Evaluation, and Inspection. A search bar is present above the table. A dropdown menu shows "Using config: arrow_num_all". Below the search bar is a navigation bar with links for "Previous" and "Next" and a page number indicator "Showing 1 - 10 of 2002 results". A table displays data rows:

Annotation	Prediction	raw_1	raw_2	raw_3	x	y	z	Updated at
More	not_work	-230.0	-1.0	52.0	-0.87	-0.0	0.2	15-03-21 19:03:03
More	not_work	-232.0	-2.0	52.0	-0.88	-0.01	0.2	15-03-21 19:03:03
More	not_work	-230.0	-1.0	51.0	-0.87	-0.0	0.2	15-03-21 19:03:03
More	not_work	-231.0	-2.0	52.0	-0.87	-0.01	0.2	15-03-21 19:03:03
More	not_work	-231.0	-2.0	55.0	-0.87	-0.01	0.21	15-03-21 19:03:03
More	not_work	-231.0	-2.0	51.0	-0.87	-0.01	0.2	15-03-21 19:03:03

The footer of the page includes "Sedue Predictor v2.1.0 | © 2014 Preferred Infrastructure".

シナリオ1：学習データ作成、精度測定 その1 コンフィグの作成

_configs

Configs Training Data Annotations Fill Annotation Evaluation Inspection

Configs

ID	Config name	Created at	Default config	Retrain	Last retrain status	Duplicate	Description
3	arrow_num_all	15-03-21 18:59:50	★ Default	↻ Retrain	Not Trained	+ Duplicate	
4	hherd_num_all	15-03-21 19:00:21	Set default	↻ Retrain	Not Trained	+ Duplicate	

Model info

Model name (#id)	demo (#3)
Description	(No description)
Model type	multi_class
Owner	Sample User
Tags	

Download training data

With IDs	Status	Result	With ID	With Predictions	Encoding	File created at		
<input type="checkbox"/>	mint_documents-v...tar.gz	2015-03-17_1905.zip	<input type="checkbox"/>	2015-03-16_1633.zip	<input type="checkbox"/>	none	<input type="checkbox"/> put	<input type="checkbox"/> すべてを表示

PREFERRED INFRASTRUCTURE

シナリオ1：学習データ作成、精度測定 その1 コンフィグの作成

Name default

Description

Easy config

Method: arrow

Regularization weight: 1.0

String rule: space

Num rule: num

Fields: Use all fields Select fields

Batch train config

Train method: using_all_data

Train times: 1

Create model

シナリオ1：学習データ作成、精度測定 その1 コンフィグの作成

- コンフィグのパラメータ
 - Name : コンフィグの名前（なんでも良い）
 - Description : コンフィグの説明（設定内容のメモ書き）
 - Method : 機械学習のアルゴリズムの選択
 - arow, nherd両方作ってみましょう
 - Regularization weight : 正則化値
 - 例 : 0.01, 0.1, 1, 10, 100など（細かく変えてもあまり変わらない）
 - String rule : 文字列からの特徴抽出方法の選択
 - 文書を入れる時にいろんな方法を試します。
 - Number rule : 数値からの特徴抽出方法の選択
 - num,log両方作ってみましょう



シナリオ1：学習データ作成、精度測定 その1 コンフィグの作成

- コンフィグのパラメータ
 - Fields : 学習に使用するフィールドの選択
 - 全部、rawだけ、xだけ、yだけ、zだけなど作ってみましょう
 - Train method : 学習方法の選択
 - 今回はこのままで
 - Train times : 学習回数の選択
 - 何度も学習すると精度が上がる事がある
 - 過学習（over fit）に注意
 - 学習データに対して過度に最適化されてしまい、実際のデータで上手く行かなくなってしまう



シナリオ1：学習データ作成、精度測定 その1 精度評価の実施

All Configを選択

Preferred Infrastructure.

シナリオ1：学習データ作成、精度測定 その1 精度評価は何をしているのか？

- 精度評価（交差検定）とは?
 - 登録してある学習データ（さっき登録したデータです。）の一部（今回は4/5）を使って学習を行ないます。
 - のこりのデータ（今回は1/5）の正解を隠して分類させます。
 - 分類結果と正解を照らし合わせてどれぐらい正解できているか確認
 - 上記を分割分（今回は5回）繰り返し、平均を取る。
- テストの種類
 - クローズテスト：学習データにあるデータを使ってテスト
 - オープンテスト：学習データに無いデータを使ってテスト
 - 交差検定はオープンテスト

Preferred Infrastructure.

シナリオ1：学習データ作成、精度測定 その1 精度評価は何をしているのか？

● 精度の考え方

- 適合率：機械が正解とした内、実際の正解が含まれる割合
 - どれだけ正確に正解を選べるか
- 再現率：実際の正解の内、機械が当てた割合
 - どれだけ抜け漏れが無いか
- F値：(適合率 × 再現率 × 2) / (適合率 + 再現率)
 - 適合率、再現率両方を加味した指標



シナリオ1：学習データ作成、精度測定 その1 精度の確認

The screenshot shows the Predictor web application interface. The top navigation bar includes links for Predictor, Models, Jobs, and Files, along with a Sample User dropdown. The main content area is titled "demo (#3) / Evaluations". Below this, there are tabs for Configs, Training Data, Annotations, Fill Annotation, Evaluation, and Inspection. The "Evaluation" tab is active.

The left sidebar contains sections for "New evaluation" (Target config: arrow_num_all, Evaluation type: Cross validation), "Cross Validation Options" (Folds: 5, Training times: 1), and "Data" (radio button selected for "All").

The right sidebar displays a table titled "Evaluation jobs" with two entries:

Status	Detail	Parameters	Accuracy	Duration	Finished at
success	View detail...	<ul style="list-style-type: none">Target config : nherd_num_allEvaluation type : Cross validationFolds : 5Training times : 1Data amount : All	50.00%	10s	15-03-21 19:04:08
success	View detail...	<ul style="list-style-type: none">Target config : arrow_num_allEvaluation type : Cross validationFolds : 5Training times : 1Data amount : All	53.60%	10s	15-03-21 19:04:08

The accuracy value "53.60%" in the second row is highlighted with a red box.



シナリオ1：学習データ作成、精度測定 その1 評価結果詳細の確認

New evaluation

Target config: `arow_num_all`

Evaluation type: Cross validation

Cross Validation Options

- Folds: 5
- Training times: 1

Data

- All
- # of training data: 100

Evaluation jobs

Status	Detail	Parameters	Accuracy	Duration	Finished at
success	View detail...	<ul style="list-style-type: none"> Target config : <code>nherd_num_all</code> Evaluation type : Cross validation Folds : 5 Training times : 1 Data amount : All 	50.00%	10s	15-03-21 19:04:08
success	View detail...	<ul style="list-style-type: none"> Target config : <code>arow_num_all</code> Evaluation type : Cross validation Folds : 5 Training times : 1 Data amount : All 	53.60%	10s	15-03-21 19:04:08

Submit

mint_documents-v...tar.gz | 2015-03-17_1905.zip | 2015-03-16_1633.zip | noname | put | 昨 すべてを表示

Preferred Infrastructure.

シナリオ1：学習データ作成、精度測定 その1 評価結果詳細 1

Cross validation results

Result	Annotation \ Prediction	not_work	work	Total	(Recall: $\sigma = 10.53$)
Fold #1	not_work	642	359	1001	(64.14 %)
Fold #2	work	570	431	1001	(43.06 %)
Fold #3	Total	(Precision: $\sigma = 0.79$) 1212 (52.97 %)	790 (54.56 %)	2002	(Accuracy: 53.60 %)
Fold #4					
Fold #5					

行 (→) : 機械が分類した内容

列 (↑) : 正解の内容

Evaluation information and parameters

Target config	<code>arow_num_all</code>
Evaluation type	Cross validation
Folds	5
Training times	1
Data amount	All

mint_documents-v...tar.gz | 2015-03-17_1905.zip | 2015-03-16_1633.zip | noname | put | 昨 すべてを表示

Preferred Infrastructure.

シナリオ1：学習データ作成、精度測定 その1 精度評価詳細の見方

• Confusion Matrixの見方

- 行（→）：機械が分類した内容
- 列（↓）：正解の内容
- 対角成分（青）：正解した件数
- それ以外（赤）：間違えた件数
- 色の意味：濃いと多い、薄いと少ない
- Accuracy：正答率（いわゆる精度）
 - 全体のF値は計算ができないため、正解を当てられた率を計算している。



シナリオ1：学習データ作成、精度測定 その1 精度評価詳細 2

The screenshot shows a web-based machine learning tool named "Predictor". The main page displays "Cross validation results" for an evaluation ID of #15. The results table is as follows:

Result	Annotation \ Prediction	not_work	work	Total	(Recall: σ = 0.0)
Fold #1	not_work	0	1001	1001	(0.00 %)
Fold #2	work	0	1001	1001	(100.00 %)
Fold #3	Total	(Precision: σ = 25.0) 0 (0.00 %)	2002 (50.00 %)	2002	(Accuracy: 50.00 %)
Fold #4					
Fold #5					

Below the results, there is a section titled "Evaluation informations and parameters" containing the following information:

Target config	nherd_num_all
Evaluation type	Cross validation
Folds	5
Training times	1
Data amount	All



シナリオ1：学習データ作成、精度測定 その1 まとめ

- シナリオ1の結論
 - これではキーを入力してあるかどうかを識別できない。
 - さて、どうするか。。。
- 作戦
 - 収集するデータの量を増やしてみる
 - 学習データを増やしてみる
 - 特徴抽出方法を変えてみる

→ 今回は、収集量を20回に増やしてみる。
意外な方法として文字列として扱ってみる。



シナリオ2：学習データ作成、精度測定 その2 学習データの作成（Edison上の作業）

- 学習データの作成
 - センサーの値を20回取得し、その際に、キー入力があったか無かったか正解データを付ける
 - PredictorはCSV形式でGUIから学習データを登録できる
 - # cd hands_on/edge/session_2
 - # node trainer2.js > train2.data
 - このプログラムはCSV形式でデータを作成します。
 - コンソール上でキー入力（強く叩きすぎないで。弱すぎもダメ）
 - 1000個学習データが出来ると終了。
 - ヘッダー情報を付ける
 - # cat header.txt train2.data > train2.csv
 - ローカルPCにファイルをコピー



シナリオ2：学習データ作成、精度測定 その2 Predictor上の作業

- モデルの作成
 - new modelボタンをクリックしモデルを作成
 - 一緒に1個コンフィグが作られる
- 学習データの登録
 - training dataを押す
 - upload fileボタンをクリック
 - ファイルを選択、アップロード



シナリオ2：学習データ作成、精度測定 その2 Predictor上の作業

- コンフィグの作成
 - 色々なコンフィグを作成してみよう
- 精度評価
 - Evaluationを選択
 - target configでall configを選択、他はデフォルトでOK
 - submitボタンを押す
 - Accuracyを確認
- 分類理由の確認
 - Configを選択
 - retrainを実施
 - Training Dataを確認



シナリオ2：学習データ作成、精度測定 その2 評価結果の確認

New evaluation

Evaluation jobs

Status	Detail	Parameters	Accuracy	Duration	Finished at
success	View detail...	<ul style="list-style-type: none">Target config : defaultEvaluation type : Cross validationFolds : 5Training times : 1Data amount : All	81.20%	10s	15-03-20 18:00:24

Cross Validation Options

Folds: 5

Training times: 1

Data:
 All
 # of training data: 100

[Submit](#)

Preferred Infrastructure.

シナリオ2：学習データ作成、精度測定 その2 評価結果の詳細

Cross validation results

Result	Annotation \ Prediction	rest	work	Total	(Recall: σ = 4.39)
Fold #1	rest	428	72	500	(85.60 %)
Fold #2	work	116	384	500	(76.80 %)
Fold #3	Total	(Precision: σ = 2.76) 544 (78.68 %)	456 (84.21 %)	1000	(Accuracy: 81.20 %)
Fold #4					
Fold #5					

Evaluation informations and parameters

Target config	default
Evaluation type	Cross validation
Folds	5
Training times	1
Data amount	All

Preferred Infrastructure.

シナリオ2：学習データ作成、精度測定 その2 分類理由の確認

- 学習の実施
 - Configを選択
 - retrainを実施
- 分類理由の確認
 - Training Dataを確認
 - Inspectionの実施



シナリオ2：学習データ作成、精度測定 その2 学習の実施

The screenshot shows the Predictor web application interface. The top navigation bar includes links for 'Models', 'Jobs', and 'Files'. The main content area is titled 'sample_model (#1)'. Below this, there is a navigation bar with tabs: 'Configs' (which is highlighted with a red box), 'Training Data', 'Annotations', 'Fill Annotation', 'Evaluation', and 'Inspection'. A large blue callout box with white text points to the 'Retrain' button in the 'Configs' table. The 'Configs' table has columns: ID, Config name, Created at, Default config, Retrain, Last retrain status, Duplicate, and Description. One row is shown with ID 1, Config name 'default', Created at '15-03-10 03:25:20', Default config 'Default', Retrain button (highlighted with a red box), Last retrain status 'success' (with timestamp '15-03-19 10:11:53'), and Description 'Auto-generated co...'. Below the table is a section titled 'Model info' containing fields: Model name (#id), Description, Model type, Owner, and Tags. At the bottom, there is a section titled 'Download training data' with a table header: Status, Result, With ID, With Predictions, Encoding, and File created at. A file list is shown below, including 'HackDay-Track3-SampleData.docx', 'mint_documents-v1.0.tar.gz', '2015-03-17_1905.zip', '2015-03-16_1633.zip', and 'noname'. A small 'すべてを表示' (Show all) link is at the bottom right.



シナリオ2：学習データ作成、精度測定 その2 分類理由の確認

sample_model (#1) / Training data

Configs Training Data Annotations Fill Annotation Evaluation Inspection

Search...

Using config: default

Annotation Prediction

More rest Set

Prediction: rest 1.442947 work -0.589727

raw x y z Updated at

raw	x	y	z	Updated at
-233	-0.88	-0.05	-0.46	15-03-19 11:11:11
-12	-0.88	-0.04	-0.47	
-118	-0.87	-0.05	-0.47	
-233	-0.87	-0.05	-0.46	
-11	-0.88	-0.05	-0.46	
-120	-0.88	-0.05	-0.46	
-231	-0.88	-0.05	-0.46	
-12	-0.88	-0.05	-0.46	
-12...				
More rest Set				

Preferred Infrastructure.

シナリオ2：学習データ作成、精度測定 その2 分類理由の確認

sample_model (#1) / Training data

Configs Training Data Annotations Fill Annotation Evaluation Inspection

Search...

Using config: default

Annotation Prediction

More rest Set

Prediction: rest 1.442947 work -0.589727

raw x y z Updated at

raw	x	y	z	Updated at
-233	-0.88	-0.05	-0.46	15-03-19 11:11:11
-12	-0.88	-0.04	-0.47	
-118	-0.87	-0.05	-0.47	
-233	-0.87	-0.05	-0.46	
-11	-0.88	-0.05	-0.46	
-120	-0.88	-0.05	-0.46	
-231	-0.88	-0.05	-0.46	
-12	-0.88	-0.05	-0.46	
-12...				
More rest Set				

Preferred Infrastructure.

シナリオ2：学習データ作成、精度測定 その2 分類理由の確認

Annotation

Label	rest
Prediction Score	rest 1.442947 work -0.589727

青が分類にPositiveに効いている特徴

Prediction: rest

Field	Content
raw	-233 -12 -118 -233 -11 -120 -231 -12 -120 -231 -13 -118 -232 -12 -117 -232 -12 -117 -231 -11 -118 -232 -12 -119 -232 -13 -120 -231 -13 -121 -231 -12 -117 -232 -12 -118 -233 -11 -119 -232 -12 -117 -231 -11 -119 -232 -11 -121 -232 -11 -120 -232 -11 -119 -232 -11 -117 -232 -13 -118 -232 -12 -119 -233 -11 -119 -232 -12 -120 -231 -12 -119 -233 -11 -119 -232 -13 -117 -231 -13 -118 -232 -12 -119 -231 -13 -118 -232 -12 -119 -233 -11 -119 -232 -12 -120 -231 -12 -119 -232 -13 -119
x	-0.88 -0.88 -0.87 -0.87 -0.88 -0.88 -0.87 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.88 -0.87 -0.88 -0.88 -0.87 -0.88
y	-0.05 -0.04 -0.05 -0.05 -0.05 -0.05 -0.04 -0.04 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.04 -0.04 -0.04 -0.04 -0.04 -0.04 -0.05 -0.05 -0.04 -0.05 -0.05 -0.04 -0.05 -0.05 -0.05
z	-0.46 -0.47 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.47 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46

rest

Field	Content
raw	-233 -12 -118 -233 -11 -120 -231 -12 -120 -231 -13 -118 -232 -12 -117 -232 -12 -117 -231 -11 -118 -232 -12 -119 -232 -13 -120 -117 -231 -11 -118 -232 -12 -119 -233 -13 -120 -231 -13 -118 -232 -12 -117 -231 -13 -119 -232 -11 -121 -231 -12 -117 -232 -12 -233 -12 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -119 -232 -11 -117 -233 -12 -117 -232 -13 -118 -232 -12 -119 -233 -11 -119 -232 -12 -119 -233 -11 -119 -232 -11 -119 -119 -232 -12 -120 -231 -12 -119 -233 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119
x	-0.88 -0.88 -0.87 -0.87 -0.88 -0.88 -0.87 -0.88 -0.88 -0.88 -0.87 -0.88
y	-0.05 -0.04 -0.05 -0.05 -0.05 -0.05 -0.04 -0.04 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.04 -0.04 -0.04 -0.04 -0.04 -0.04 -0.04 -0.04 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05
z	-0.46 -0.47 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.47 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.47 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.47 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.47

Preferred Infrastructure.

シナリオ2：学習データ作成、精度測定 その2 分類理由の確認

赤が分類にNegativeに効いている特徴

Prediction: work

Field	Content
raw	-233 -12 -118 -233 -11 -120 -231 -12 -120 -231 -13 -118 -232 -12 -117 -232 -12 -117 -231 -11 -118 -232 -12 -119 -232 -13 -120 -117 -231 -11 -118 -232 -12 -119 -233 -13 -120 -231 -13 -118 -232 -12 -117 -231 -13 -119 -232 -11 -121 -231 -12 -117 -232 -12 -233 -12 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -12 -118 -233 -12 -119 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -121 -119 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -121 -232 -11 -121 -117 -232 -13 -118 -232 -12 -119 -233 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -117 -232 -13 -118 -232 -12 -119 -233 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -232 -11 -119 -119 -233 -11 -119 -232 -13 -117 -231 -13 -117 -231 -13 -117 -231 -13 -117 -231 -13 -117 -231 -13 -117 -231 -13 -117
x	-0.88 -0.88 -0.87 -0.87 -0.88 -0.88 -0.87 -0.88 -0.88 -0.88 -0.87 -0.88
y	0.05 -0.04 -0.05 -0.05 -0.05 -0.05 -0.04 -0.04 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.04 -0.04 -0.04 -0.04 -0.04 -0.04 -0.04 -0.04 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05 -0.05
z	-0.46 -0.47 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.47 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.47 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.47 -0.47 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.46 -0.47

Preferred Infrastructure.

シナリオ2：学習データ作成、精度測定 その2 分類理由の確認

The screenshot shows the Predictor web interface with the 'sample_model (#1) / Inspections' page. The 'Inspection' tab is selected. On the left, there's a 'New inspection' form with a 'Target config' dropdown set to 'default', which is highlighted with a red box. Below it are 'Inspection type' (Feature ranking) and 'Sample size' (30). On the right, there's a table titled 'Inspection jobs' with columns: Status, Detail, Parameters, Duration, and Finished at. A blue callout box points to the 'Detail' column of the first row with the text 'All Configを選択'. At the bottom of the form is a 'Submit' button, which is also highlighted with a red box. A blue callout box points to it with the text 'Submitボタンをクリック'. The status bar at the bottom shows several file icons.

Preferred Infrastructure.

シナリオ2：学習データ作成、精度測定 その2 分類理由の確認

The screenshot shows the Predictor web interface with the 'jaws_demo (#2) / Inspections' page. The 'Inspection' tab is selected. On the left, there's a 'New inspection' form with a 'Target config' dropdown set to 'default'. On the right, there's a table titled 'Inspection jobs' with columns: Status, Detail, Parameters, Duration, and Finished at. In the 'Detail' column of the first row, there's a 'View detail...' button, which is highlighted with a red box. A blue callout box points to it with the text '詳細を確認'. The status bar at the bottom shows several file icons.

Preferred Infrastructure.

シナリオ2：学習データ作成、精度測定 その2 分類理由の確認

The screenshot shows a web browser window with the URL `localhost:8080/predictor/models/2/inspections/19`. The page title is "jaws_demo (#2) / Inspections / #19". The navigation bar includes "Configs", "Training Data", "Annotations", "Fill Annotation", "Evaluation", and "Inspection". The "Inspection" tab is active. Below it, there are two tabs: "Summary" (which is selected) and "rest work". A large blue callout box with white text "効いている特徴のランキングを表示" (Display ranking of effective features) points to a table titled "Summary". The table has three columns: "Field", "Feature", and two grouped columns under "rest" and "work". The "rest" group contains values like 0.231528, 0.208109, etc., while the "work" group contains values like -0.231528, -0.200911, etc. The table lists various fields such as raw, y, x, and their corresponding feature values and scores.

Preferred Infrastructure.

シナリオ2：学習データ作成、精度測定 その2 まとめ

- シナリオ2の結論
 - 精度を出すためには色々な工夫が必要
 - データの取り方、特徴の取り方、パラメータを変えてみたり
 - 普通はプログラムを書き直して行なう
 - 工夫をするたびに精度測定が必要
 - 普通は上記の組み合わせごとに学習させて分類して集計して。。。
 - 分類した理由を確認
 - 普通は出来ない。（モデルをダンプして解析する必要がある）
- Sedue Predictorは上記の作業を協力に支援！（宣伝）

では実際に分析をしてみましょう！

Preferred Infrastructure.

シナリオ3：実際に分析を実施してみる データの流れ



シナリオ3：実際に分析を実施してみる

- Sedue Predictorの準備
 - sample_modelを選択
 - 登録済みの学習データを削除
 - シナリオ2で作成した学習データを登録
 - defaultコンフィグを修正する
 - シナリオ2で精度の良かったコンフィグと同じ設定にする
 - コンフィグをretrainする

今回はsample modelを使って分類します。
(事前にしてある設定を使うため)

シナリオ3：実際に分析を実施してみる

The screenshot shows the Predictor web application interface. At the top, there's a navigation bar with links for Predictor, Models, Jobs, and Files. Below it, a sub-navigation bar for 'sample_model (#1)' includes tabs for Configs, Training Data, Annotations, Fill Annotation, Evaluation, and Inspection. The main content area is titled 'Configs' and displays a table with one row:

ID	Config name	Created at	Default config	Retrain	Last retrain status	Duplicate	Description
1	default	15-02-25 25:20	★ Default	Retrain	success 15-03-19 10:11:53	+ Duplicate	Auto-generated co...

Below this is a 'Model info' section containing fields like Model name, Description, Model type, Owner, and Tags. A large blue arrow points downwards from the 'Configs' table towards the 'Model info' section, with a callout bubble containing the Japanese text '下にスクロールしてください' (Please scroll down). Further down is a 'Download training data' section with various download options and a file list.

Preferred Infrastructure.

シナリオ3：実際に分析を実施してみる

This screenshot shows the Predictor interface again, focusing on the 'Dump state of the model and config' section. It includes dropdown menus for Target config (set to 'default') and Encoding (set to 'SJIS'), and a 'Execute' button. Below this is a 'Clear training data' section with a red 'CLEAR TRAINING DATA' button, which has a callout bubble with the Japanese text '学習データをクリア'. Further down is a 'Delete everything in this model and the model' section with a red 'DELETE EVERYTHING IN THIS MODEL AND THE MODEL' button.

Preferred Infrastructure.

シナリオ3：実際に分析を実施してみる 学習データの登録

The screenshot shows the Predictor web application interface. The URL is `localhost:8080/predictor/models/1/training_data`. The top navigation bar includes links for Predictor, Models, Jobs, and Files, along with a Sample User link. The main content area has tabs for Configs, Training Data (which is selected and highlighted with a red box), Annotations, Fill Annotation, Evaluation, and Inspection. A search bar is present below the tabs. On the right side, there is a large blue callout box with the text "Upload file ボタンをクリック" (Click the Upload file button) and an arrow pointing to a red box around the "Upload file" button in the top right corner of the main content area. Below the tabs, there is a dropdown menu set to "Using config: default". A message at the bottom states "Showing 0 - 0 of 0 results". The footer of the browser window shows several tabs open, including "mint_documents-v...tar.gz", "2015-03-17_1905.zip", "2015-03-16_1633.zip", "noname", and "put". The Preferred Infrastructure logo is visible in the bottom right corner.

シナリオ3：実際に分析を実施してみる 学習データの登録

The screenshot shows the Predictor web application interface, specifically the "Upload" step of the training data process. The URL is `localhost:8080/predictor/models/1/training_data/upload?config_id=1`. The top navigation bar and tabs are identical to the previous screenshot. The main content area now shows a "CSV file" section with a red box around the "ファイルを選択" (Select file) button. A blue callout box with the text "ファイルを選択して Uploadボタンをクリック" (Select the file and click the Upload button) points to the "Upload" button below the file selection field. To the right, a table displays upload history: "hands_on5.csv" was uploaded on 15-03-19 10:11:21, started at 15-03-19 10:11:24, and finished at 15-03-19 10:11:27. Another entry for "hands_on5.csv" was uploaded on 15-03-16 11:07:44, started at 15-03-16 11:07:46, and finished at 15-03-16 11:07:50. A third entry for "hands_on5.csv" was uploaded on 15-03-16 11:06:55, started at 15-03-16 11:06:56, and finished at 15-03-16 11:06:57. A success message at the bottom states "200 data added and 0 data updated." The footer and Preferred Infrastructure logo are also present.

シナリオ3：実際に分析を実施してみる

Configs

ID	Config name	Created at	Default config	Retrain	Last retrain status	Duplicate	Description
1	default			<input type="button" value="Retrain"/>	success 15-03-19 10:11:53	<input type="button" value="Duplicate"/>	Auto-generated co...

Model info

Model name (#id)	sample_model (#1)
Description	Auto-generated model sample.
Model type	multi_class
Owner	Sample User
Tags	auto_generated

Download training data

With IDs Status Result With ID With Predictions Encoding File created at

HackDay-Track3-Sc...docx mint_documents-v...tar.gz 2015-03-17_1905.zip 2015-03-16_1633.zip noname

Preferred Infrastructure.

シナリオ3：実際に分析を実施してみる

sample_model (#1) / default (#1)

Configs Training Data Annotations Fill Annotation Evaluation Inspection

Config info

Config name (#id)	default (#1)
Description	Auto-generated config.
Method	arow
Regularization weight	1.0
String rule	space
Num rule	num
Fields	All fields
Train method	using_all_data (1 times)
# of train queries	1000
# of features	241
# of labels	2
Created at	15-03-10 03:25:20
Updated at	15-03-10 03:25:20

Preferred Infrastructure.

シナリオ3：実際に分析を実施してみる

localhost:8080/predictor/models/1/configs/1/edit

Configs Training Data Annotations Fill Annotation Evaluation Inspection

Name: default
Description: Auto-generated config.

Easy config Method: arrow
Regularization weight: 1.0
String rule: space
Num rule: num
Fields: Use all fields Select fields

Batch train config Train method: using_all_data
Train times: 1

Update config

シナリオ2でベストだった
コンフィグを設定

HackDay-Track3-Sc...docx mint_documents-v...tar.gz 2015-03-17_1905.zip 2015-03-16_1633.zip noname

Preferred Infrastructure.

シナリオ3：実際に分析を実施してみる

localhost:8080/predictor

Models Jobs Files Sample User

sample_model (#1) Edit model

Configs Training Data Annotations Fill Annotation Evaluation Inspection

ID	Config name	Created at	Default config	Retrain	Last retrain status	Duplicate	Description
1	default	15-03-10 03:25:20	* Default	Retrain	success 15-03-19 10:11:53	+ Duplicate	Auto-generated co...

Model info

Model name (#id)	sample_model (#1)
Description	Auto-generated model sample.
Model type	multi_class
Owner	Sample User
Tags	auto_generated

Download training data

With IDs	Status	Result	With ID	With Predictions	Encoding	File created at
With predictions using config						

HackDay-Track3-Sc...docx mint_documents-v...tar.gz 2015-03-17_1905.zip 2015-03-16_1633.zip noname

Preferred Infrastructure.

シナリオ3：実際に分析を実施してみる データをkinesisに送る（Edison上の作業）

- プログラムを修正する

- session_3/put.jsを修正

```
var cognitoParams = {  
    AccountId: "ここをうめる",  
    RoleArn: "arn:aws:iam:: * * ", // こうめる  
    IdentityPoolId: "us-east-1: * * " // こうめる  
};  
...  
var params = { Data: JSON.stringify(data), PartitionKey: "demo",  
    StreamName: "ここを埋める" };
```

- プログラムを実行する

- # node put.js



シナリオ3：実際に分析を実施してみる データを受け取る（EC2上の作業）

- EC2にログイン

- \$ ssh demo@IPアドレス
 - パスワード : jaws2015

- プログラムを実行

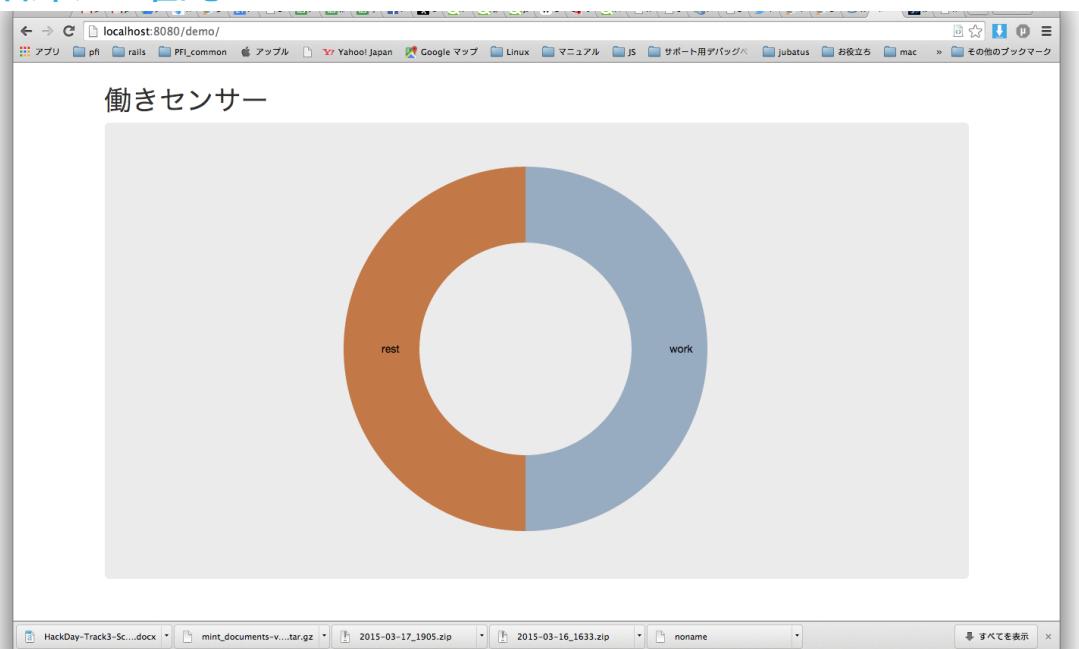
- \$ ruby get.rb --stream <stream名> --tag sfbd.query.add

シナリオ3：実際に分析を実施してみる 確認する

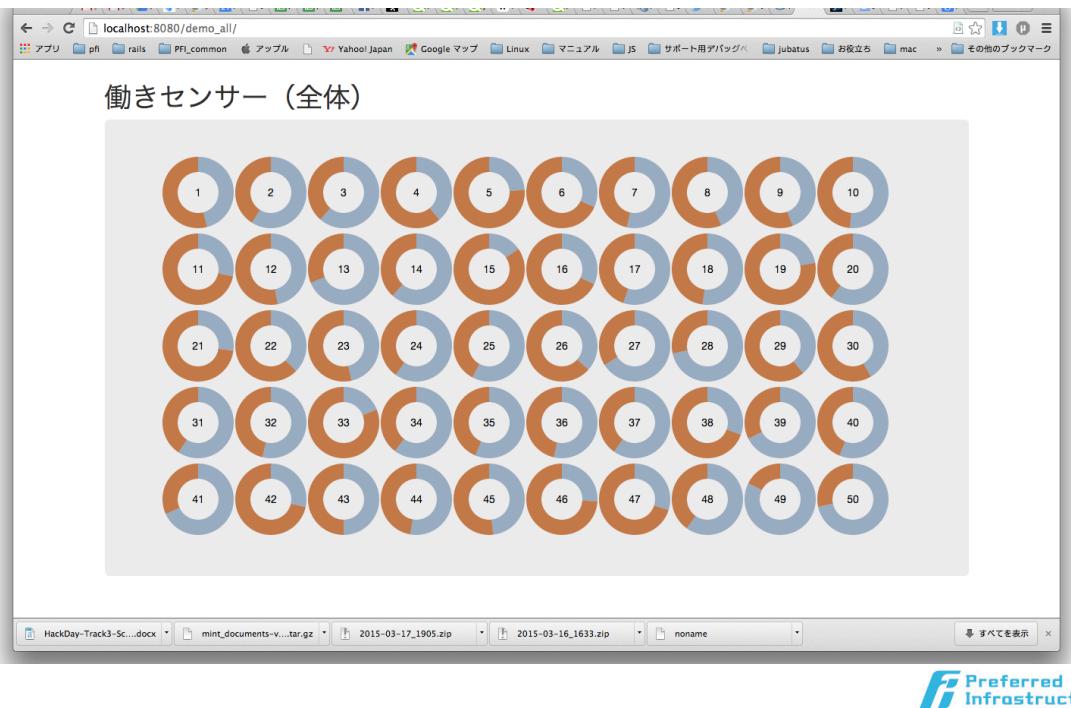
- 登録結果を確認する
 - ブラウザで確認
 - <http://ホスト名/demo>



シナリオ3：実際に分析を実施してみる 結果の確認



シナリオ3：実際に分析を実施してみる 結果の確認



Preferred Infrastructure.

まとめ

- 今日は機械学習を使ってデータの分析をしてみました。
- 機械学習の使い方の流れ分かりましたか？
 - データを収集する
 - 学習データを作成する
 - 特徴を抽出する
 - 学習をする
 - 精度評価を行なう
- センサーデータはただ収集して可視化するだけでは意味がありません。機械学習を使って分析しましょう！

Preferred Infrastructure.

最後に

- 環境を削除するのを忘れずに。
- ご質問とか有ればお気軽に！
- Sedue製品にご興味のある方はぜひお声がけください！
 - 体験版VMなど、ご用意しております！



Copyright © 2006-2015
[Preferred Infrastructure All Right Reserved.](#)

