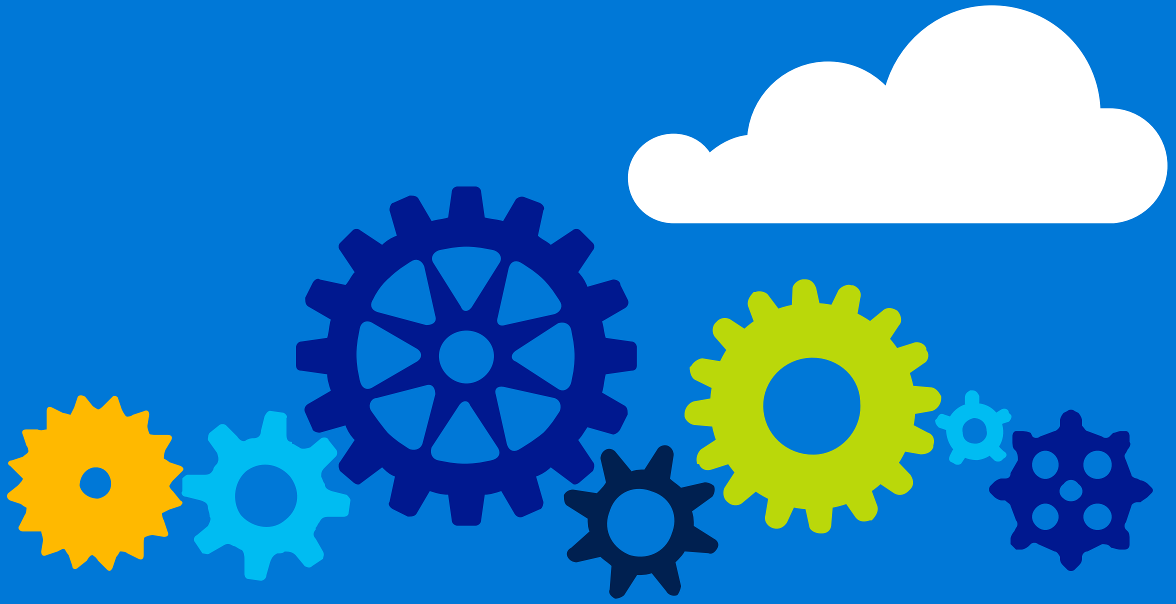
予兆保全のサンプル紹介
Modelling Guide for Predictive Maintenance

Version 1.0



本ドキュメントの目的

- 予兆保全のサンプルをご紹介いたします
  - 本サンプルの「解説」ではありません
  - 詳細は記載した URL のページを参考にしてください

- 予兆保全のサンプルは、以下を目的としたものです
  - データサイエンスの一般的なプロセスを理解します
  - 機械学習のアプローチによる故障予測の方法を示します

対象となる予兆保全のサンプル

- Predictive Maintenance Modelling Guide R Notebook
  - https://gallery.cortanaintelligence.com/Notebook/Predictive-Maintenance-Implementation-Guide-R-Notebook-2

- Predictive Maintenance Modelling Guide Python Notebook
  - https://gallery.cortanaintelligence.com/Notebook/Predictive-Maintenance-Modelling-Guide-Python-Notebook-1
  - 上記の R 版を Python 版に移植したもの

- 本書では Python 版でご説明をします。

[Step1]
Data preparation

サンプルで利用するデータ

- マイクロソフトによって作成されたサンプルデータ
  - 2015/01/01 06:00 ～ 2016/01/01 06:00 の時系列データ
    1. Telemetry - PdM_telemetry.csv
    2. Errors - PdM_errors.csv
    3. Maintenance - PdM_maint.csv
    4. Machines - PdM_machines.csv
    5. Failures - PdM_failures.csv

- 下記からダウンロード可能
  - Predictive Maintenance Modelling Guide Data Sets
    - https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Implementation-Guide-Data-Sets-1

Telemetry - PdM_telemetry.csv

- The telemetry time-series data which consists of voltage, rotation, pressure, and vibration measurements
- Collected from 100 machines in real time averaged over every hour collected during the year 2015

		ボルト	回転	圧力	振動
datetime	machineID	volt	rotate	pressure	vibration
2015/1/1 6:00	1	176.217853	418.5040782	113.0779355	45.08768576
2015/1/1 7:00	1	162.8792229	402.7474896	95.46052538	43.41397268
2015/1/1 8:00	1	170.9899024	527.3498255	75.23790486	34.17884712
2015/1/1 9:00	1	162.4628333	346.149335	109.2485613	41.12214409
2015/1/1 10:00	1	157.6100212	435.376873	111.8866482	25.990511
2015/1/1 11:00	1	172.5048392	430.3233621	95.92704169	35.65501733

Errors - PdM_errors.csv

- These are non-breaking errors thrown while the machine is still operational and do not constitute as failures
- The error date and times are rounded to the closest hour since the telemetry data is collected at an hourly rate

datetime	machineID	errorID
2015/1/3 7:00	1	error1
2015/1/3 20:00	1	error3
2015/1/4 6:00	1	error5
2015/1/10 15:00	1	error4
2015/1/22 10:00	1	error4

マシンが停止する程ではないエラー。警告のようなもの。

Maintenance - PdM_maint.csv

- The scheduled and unscheduled maintenance record is generated if a component is replaced during the scheduled inspection or replaced due to a breakdown.
- The records that are created due to breakdowns will be called failures which is explained in the later sections.
- Maintenance data has both 2014 and 2015 records.

datetime	machineID	comp
2014/6/1 6:00	1	comp2
2014/6/1 6:00	6	comp2
2014/6/1 6:00	9	comp1
2014/6/1 6:00	9	comp2
2014/6/1 6:00	11	comp2

メンテナンス対象のコンポーネント

Machines - PdM_machines

- Some information about the machines which are model type and age which is years in service.

モデル
稼働期間

machineID	model	age
1	model3	18
2	model4	7
3	model3	8
4	model3	7
5	model3	2

Failures - PdM_failures

- The records of component replacements due to failures.
- Each record has a date and time, machine ID and failed component type.

故障したコンポーネント

datetime	machineID	failure
2015/1/5 6:00	1	comp4
2015/3/6 6:00	1	comp1
2015/4/20 6:00	1	comp2
2015/6/19 6:00	1	comp4
2015/9/2 6:00	1	comp4

[Step2]
Feature engineering

Feature Engineering

A) 予測に有効な特徴を作成する。

B) 複数のデータソースを1個のデータソースに編集する。

Telemetry

datetime	machineID	volt	rotate	pressure	vibration
2015/1/1 6:00	1	176.217853	418.5040782	113.0779355	45.08768576
2015/1/1 7:00	1	162.8792229	402.7474896	95.46052538	43.41397268
2015/1/1 8:00	1	170.9899024	527.3498255	75.23790486	34.17884712
2015/1/1 9:00	1	162.4628333	346.149335	109.2485613	41.12214409
2015/1/1 10:00	1	157.6100212	435.376873	111.8866482	25.990511
2015/1/1 11:00	1	172.5048392	430.3233621	95.92704169	35.65501733

Machines

machineID	model	age
1	model3	18
2	model4	7
3	model3	8
4	model3	7
5	model3	2

x1	…	xn	y

Errors

datetime	machineID	errorID
2015/1/3 7:00	1	error1
2015/1/3 20:00	1	error3
2015/1/4 6:00	1	error5
2015/1/10 15:00	1	error4
2015/1/22 10:00	1	error4

Maintenance

datetime	machineID	comp
2014/6/1 6:00	1	comp2
2014/6/1 6:00	6	comp2
2014/6/1 6:00	9	comp1
2014/6/1 6:00	9	comp2
2014/6/1 6:00	11	comp2

Failures

datetime	machineID	failure
2015/1/5 6:00	1	comp4
2015/3/6 6:00	1	comp1
2015/4/20 6:00	1	comp2
2015/6/19 6:00	1	comp4
2015/9/2 6:00	1	comp4

A) Lag Features from **Telemetry**

- 1時間単位のレコードを3時間単位にする。
- 各センサーの値について、以下の計算結果を特徴量とする
  - 過去3時間分の平均値、標準偏差
    - [平均値] voltmean_3h、rotatemean_3h、pressuremean_3h、vibrationmean_3h
    - [標準偏差] voltsd_3h、rotatesd_3h、pressuresd_3h、vibrationsd_3h
  - 過去24時間分の平均値、標準偏差
    - [平均値] voltmean_24h、rotatemean_24h、pressuremean_24h、vibrationmean_24h
    - [標準偏差] voltsd_24h、rotatesd_24h、pressuresd_24h、vibrationsd_24h

machineID	datetime	voltmean_3h	rotatemean_3h	・・・	voltsd_3h	rotatesd_3h	・・・	voltmean_24h	rotatemean_24h	・・・	voltsd_24h	rotatesd_24h	・・・
1	2015/1/2 6:00	180.133784	440.6083201		21.32273479	48.77051197		169.7338089	445.1798646		11.23312028	48.71739478	
1	2015/1/2 9:00	176.3642932	439.349655		18.95221004	51.32963577		170.6148619	446.3648591		12.51940225	48.38507588	
1	2015/1/2 12:00	160.3845679	424.3853157		13.04707951	13.70249554		169.893965	447.0094075		13.3703566	42.43231716	
1	2015/1/2 15:00	170.4724608	442.9339969		16.64235413	56.29044728		171.2434437	444.2335635		13.2992806	41.34612144	
1	2015/1/2 18:00	163.2638057	468.9375583		17.42468821	38.6803798		170.792486	448.4404372		13.95451751	43.49023354	

A) Lag Features from **Telemetry**

datetime	machineID	volt	voltmean_3h	voltsd_3h	voltmean_24h	voltstd_24h
2015/1/1 6:00	1	176.217853				
2015/1/1 7:00	1	162.8792229				
2015/1/1 8:00	1	170.9899024				
2015/1/1 9:00	1	162.4628333	170.0289928	6.721032201		
2015/1/1 10:00	1	157.6100212	165.4439862	4.807414596		
2015/1/1 11:00	1	172.5048392	163.6875856	6.773501241		
2015/1/1 12:00	1	156.5560306	164.1925646	7.596570185		
2015/1/1 13:00	1	172.5227808	162.2236303	8.919370259		
2015/1/1 14:00	1	175.3245239	167.1945502	9.213232599		
2015/1/1 15:00	1	169.2184232	168.1344451	10.12458409		
2015/1/1 16:00	1	167.0609807	172.3552427	3.056496052		
2015/1/1 17:00	1	160.2639537	170.5346426	4.286124313		
2015/1/1 18:00	1	153.3534915	165.5144526	4.673268995		
2015/1/1 19:00	1	182.739113	160.226142	6.853822822		
2015/1/1 20:00	1	170.3354379	165.4521861	15.36447214		
2015/1/1 21:00	1	182.4671093	168.8093475	14.75213211		
2015/1/1 22:00	1	151.3356822	178.5138867	7.084050083		
2015/1/1 23:00	1	172.5353962	168.0460765	15.69147281		
2015/1/2 0:00	1	180.0974946	168.7793959	15.90195243		
2015/1/2 1:00	1	169.6058544	167.9895243	14.91003608		
2015/1/2 2:00	1	167.1291175	174.0795817	5.413594967		
2015/1/2 3:00	1	158.2714003	[illegible]	[illegible]		
2015/1/2 4:00	1	200.8724298	[illegible]	[illegible]		
2015/1/2 5:00	1	181.2575218	175.4243159	22.47931641		
2015/1/2 6:00	1	197.3631245	180.133784	21.32273479	169.7338089	11.23312028
2015/1/2 7:00	1	171.2008946	193.1643587	10.45984608	170.6148619	12.51940225

- voltmean_3h
  ⇒ 過去 3 時間分の平均値
- voltsd_3h
  ⇒ 過去 3 時間分の標準偏差

- voltmean_24h
  ⇒ 過去 24 時間分の平均値
- voltsd_24h
  ⇒ 過去 24 時間分の標準偏差

A) Lag Features from **Errors**

- The number of errors of each type in a lagging window.
  - The error IDs are categorical values and should not be averaged over time intervals like the telemetry measurements.

Step1：one-hot encoding

machineID	datetime	error1	error2	error3	error4	error5
1	2015/1/3 7:00	1	0	0	0	0
1	2015/1/3 20:00	0	0	1	0	0
1	2015/1/4 6:00	0	0	0	0	1
1	2015/1/10 15:00	0	0	0	1	0
1	2015/1/22 10:00	0	0	0	1	0

Step2：3 時間毎の時系列データに変換
※24時間以内に何回発生したか

machineID	datetime	error1count	error2count	error3count	error4count	error5count
1	2015/1/2 6:00	0	0	0	0	0
1	2015/1/2 9:00	0	0	0	0	0
1	2015/1/2 12:00	0	0	0	0	0
1	2015/1/2 15:00	0	0	0	0	0
1	2015/1/2 18:00	0	0	0	0	0
1	2015/1/2 21:00	0	0	0	0	0
1	2015/1/3 0:00	0	0	0	0	0
1	2015/1/3 3:00	0	0	0	0	0
1	2015/1/3 6:00	0	0	0	0	0
1	2015/1/3 9:00	0	0	0	0	0
1	2015/1/3 12:00	1	0	0	0	0
1	2015/1/3 15:00	1	0	0	0	0
1	2015/1/3 18:00	1	0	0	0	0
1	2015/1/3 21:00	1	0	0	0	0

A) Days Since Last Replacement from **Maintenance**

- How long it has been since a component is last replaced

datetime	machineID	comp
2014/6/1 6:00	1	comp2
2014/7/16 6:00	1	comp4
2014/7/31 6:00	1	comp3
2014/12/13 6:00	1	comp1
2015/1/5 6:00	1	comp4
2015/1/5 6:00	1	comp1

何日前に取り換えたか？

datetime	machineID	comp1	comp2	comp3	comp4
2015/1/1 6:00	1	19.000000	214.000000	154.000000	169.000000
2015/1/1 7:00	1	19.041667	214.041667	154.041667	169.041667
2015/1/1 8:00	1	19.083333	214.083333	154.083333	169.083333
2015/1/1 9:00	1	19.125000	214.125000	154.125000	169.125000
2015/1/1 10:00	1	19.166667	214.166667	154.166667	169.166667
2015/1/1 11:00	1	19.208333	214.208333	154.208333	169.208333
2015/1/1 12:00	1	19.250000	214.250000	154.250000	169.250000
2015/1/1 13:00	1	19.291667	214.291667	154.291667	169.291667
2015/1/1 14:00	1	19.333333	214.333333	154.333333	169.333333
2015/1/1 15:00	1	19.375000	214.375000	154.375000	169.375000
2015/1/1 16:00	1	19.416667	214.416667	154.416667	169.416667
2015/1/1 17:00	1	19.458333	214.458333	154.458333	169.458333
2015/1/1 18:00	1	19.500000	214.500000	154.500000	169.500000
2015/1/1 19:00	1	19.541667	214.541667	154.541667	169.541667

B) 複数のデータソースを1個のデータソースに編集

- The **machine** features can be used without further modification.

machineID	datetime	voltmean_3h	・・・	voltsd_3h	・・・	voltmean_24h	・・・	voltsd_24h	・・・	error1count	・・・	comp1	・・・	model	age
1	2015/1/2 6:00	180.133784		21.32273479		169.7338089		11.23312028		0		20		model3	18
1	2015/1/2 9:00	176.3642932		18.95221004		170.6148619		12.51940225		0		20.125		model3	18
1	2015/1/2 12:00	160.3845679		13.04707951		169.893965		13.3703566		0		20.25		model3	18
1	2015/1/2 15:00	170.4724608		16.64235413		171.2434437		13.2992806		0		20.375		model3	18
1	2015/1/2 18:00	163.2638057		17.42468821		170.792486		13.95451751		0		20.5		model3	18
1	2015/1/2 21:00	163.2784657		21.58049205		170.5566735		14.40274037		0		20.625		model3	18
1	2015/1/3 0:00	172.1911982		16.36983597		168.4605251		15.51381946		0		20.75		model3	18
1	2015/1/3 3:00	175.2100275		5.991920644		169.7729506		15.72696995		0		20.875		model3	18
1	2015/1/3 6:00	181.6901083		11.51444993		170.9005619		15.6350828		0		21		model3	18
1	2015/1/3 9:00	172.3829348		7.065150316		169.5331561		13.99546545		1		21.125		model3	18
1	2015/1/3 12:00	174.3038582		19.01719562		170.8660134		13.10036393		1		21.25		model3	18
1	2015/1/3 15:00	176.2463479		12.57250423		171.0416509		13.80848947		1		21.375		model3	18
1	2015/1/3 18:00	158.4335328		5.136951888		171.2445326		14.18798486		1		21.5		model3	18
1	2015/1/3 21:00	162.3879541		4.563331475		171.3850393		13.7077939		1		21.625		model3	18

Label Construction

- 業務要件を「24時間以内に故障を予測する」と仮定
  - 24時間前までのレコードを“故障”としてラベリング

machineID	datetime	voltmean_3h	…	voltsd_3h	…	voltmean_24h	…	voltsd_24h	…	error1count	…	comp1	…	model	age	failure
1	2015/1/2 6:00	180.133784		21.32273479		169.7338089		11.23312028		0		20		model3	18	none
1	2015/1/2 9:00	176.3642932		18.95221004		170.6148619		12.51940225		0		20.125		model3	18	none
1	2015/1/2 12:00	160.3845679		13.04707951		169.893965		13.3703566		0		20.25		model3	18	none
1							…									
1	2015/1/4 0:00	174.2431921		6.268729646		171.8806332		11.81860338		1		21.75		model3	18	none
1	2015/1/4 3:00	176.4433609		16.33028488		172.5132015		12.06939085		1		21.875		model3	18	none
1	2015/1/4 6:00	186.0928958		13.48908994		172.6862448		12.75523445		1		22		model3	18	none
1	2015/1/4 9:00	166.2818483		24.27622836		172.0424283		12.84864599		1		22.125		model3	18	comp4
1	2015/1/4 12:00	175.4121033		34.91868731		171.2196226		14.96835137		0		22.25		model3	18	comp4
1	2015/1/4 15:00	157.3477159		24.61773924		172.0134427		17.05821739		0		22.375		model3	18	comp4
1	2015/1/4 18:00	176.4505503		8.07140042		170.1763205		18.40576279		0		22.5		model3	18	comp4
1	2015/1/4 21:00	190.3258144		8.390777199		172.9322477		18.24983081		0		22.625		model3	18	comp4
1	2015/1/5 0:00	169.9851344		9.451482895		175.1211312		19.14128711		0		22.75		model3	18	comp4
1	2015/1/5 3:00	149.0826185		19.07595234		173.4072551		18.88703274		0		22.875		model3	18	comp4
1	2015/1/5 6:00	185.7827089		14.49566382		170.7578407		20.83799306		0		0		model3	18	comp4
1	2015/1/5 9:00	169.0848085		12.24554373		171.929104		21.29832165		0		0.125		model3	18	none
1	2015/1/5 12:00	165.5187903		23.17063844		170.9085223		21.2001834		0		0.25		model3	18	none
1	2015/1/5 15:00	175.9896423		4.028327262		170.4163255		18.81467929		0		0.375		model3	18	none
1	2015/1/5 18:00	188.5764441		8.278604591		173.3151666		16.76246944		0		0.5		model3	18	none

故障が発生しない場合は none をラベリング

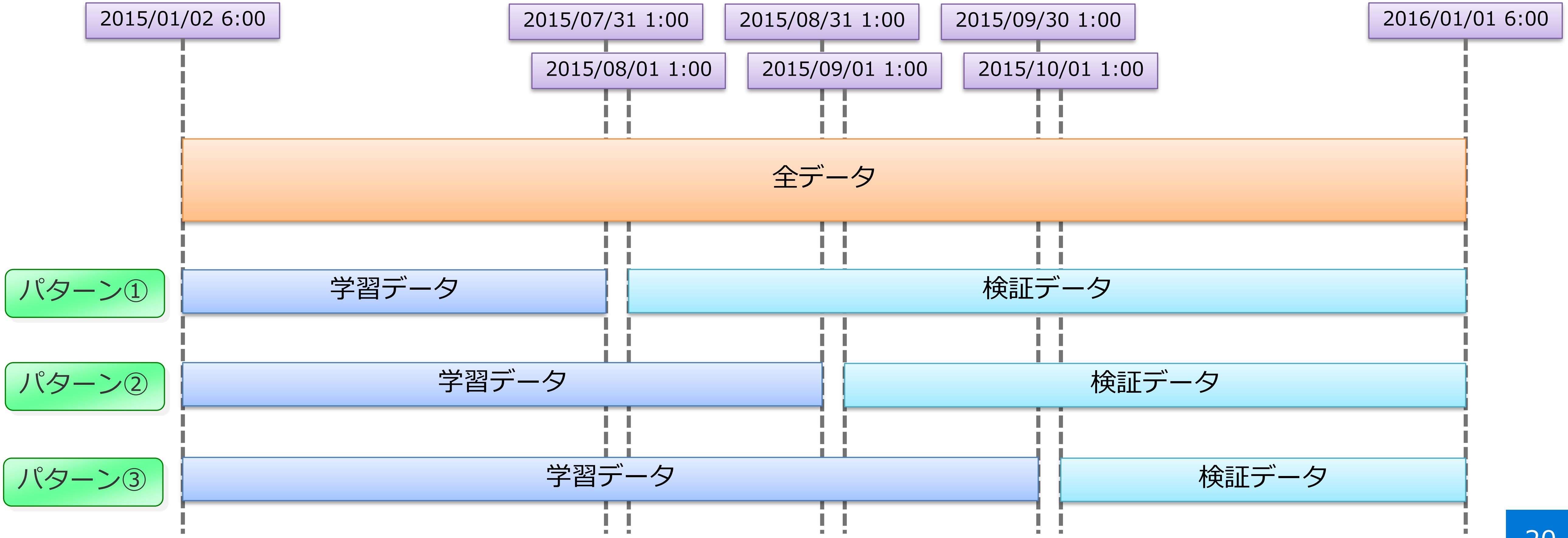
24時間前までのレコードも故障とみなす

comp4 が故障した時刻

[Step3]
Modelling

予測モデルの作成

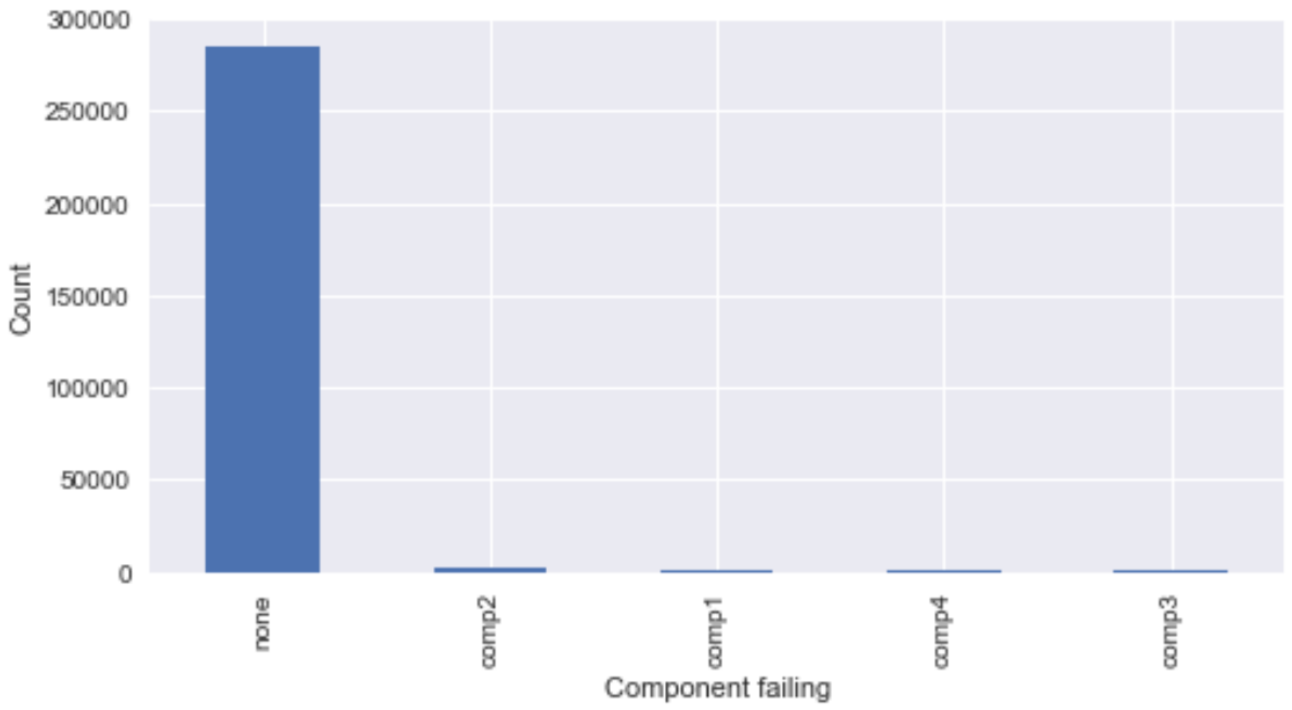
- Gradient Boosting Decision Tree を利用
- 以下の3パターンで学習し、3個の予測モデルを作成



[Step4]
Evaluation

不均衡データについて

- 正常クラスと異常クラスの標本数に著しい偏りがある場合、予測精度は高くなるが、それだけでは意味のないことが多い。
  - 正常標本が異常標本の10000倍であれば、何も工夫せずとも、すべてを正常と判定する「モデル」を作成すれば、予測精度は 99.99%となる。
- 異常クラスを正常クラスと誤判断しないモデルを作成する必要がある。



不均衡データへの対応 本サンプルでは省略されています。

- 重みづけ
  - 正常標本が異常標本の10倍であれば、正常標本に0.1の重みを付けて学習

- 間引き（ダウンサンプリング）
  - 正常標本が1000個、異常標本が100個であれば、正常標本から100個をランダムで取得して学習

- 水増し（アップサンプリング）
  - 上の例で言えば、逆に、異常標本を復元抽出（ブートストラップサンプリング）により1000個に水増し

Confusion matrix（混同行列）

パターン①

		予測結果				
		none	comp1	comp2	comp3	comp4
実際の分類結果	none	120285	21	0	4	3
	comp1	18	515	2	5	2
	comp2	0	1	867	0	1
	comp3	12	0	2	373	1
	comp4	2	1	6	0	498

パターン②

		予測結果				
		none	comp1	comp2	comp3	comp4
実際の分類結果	none	95957	13	0	4	3
	comp1	19	400	1	1	1
	comp2	0	1	707	0	0
	comp3	12	0	2	291	1
	comp4	2	2	4	0	392

パターン③

		予測結果				
		none	comp1	comp2	comp3	comp4
実際の分類結果	none	72415	7	0	4	3
	comp1	17	299	1	1	2
	comp2	0	1	555	0	0
	comp3	11	0	0	212	1
	comp4	2	0	3	0	275

評価指標

- 多クラス分類において、様々な評価指標が存在
  - 参考）Computing Classification Evaluation Metrics in R
    - http://blog.revolutionanalytics.com/2016/03/com_class_eval_metrics_r.html

Computing Classification Evaluation Metrics in R

by Said Bleik, Shaheen Gauher, Data Scientists at Microsoft

Evaluation metrics are the key to understanding how your classification model performs when applied to a test dataset. In what follows, we present a tutorial on how to compute common metrics that are often used in evaluation, in addition to metrics generated from random classifiers, which help in justifying the value added by your predictive model, especially in cases where the common metrics suggest otherwise.

Creating the Confusion Matrix

We will start by creating a **confusion matrix** from simulated classification results. The confusion matrix provides a tabular summary of the actual class labels vs. the predicted ones. The test set we are evaluating on contains 100 instances which are assigned to one of 3 classes a, b or c.

```
set.seed(0)
```

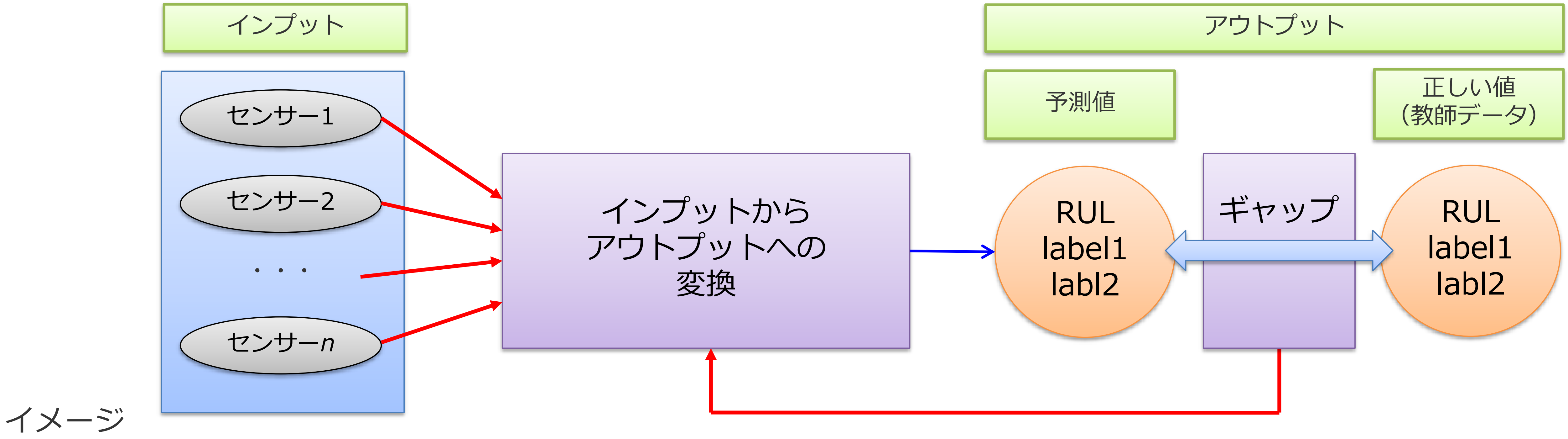
評価の例

- Recall で評価する場合を考える
  - 実際に comp*N*／none であるデータのうち、正しく comp*N*／none と予測されたデータの割合

		none	comp1	comp2	comp3	comp4
パターン①	recall for first split	0.999767	0.950185	0.997699	0.961340	0.982249
パターン②	recall for second split	0.999792	0.947867	0.998588	0.950980	0.980000
パターン③	recall for third split	0.999807	0.934375	0.998201	0.946429	0.982143

Appendix

Appendix 1.
機械学習の用語の整理



用語			
	特徴量、特徴ベクトル	モデル	識別クラス、クラスラベル、ラベル
	説明変数	機械学習アルゴリズム	目的変数、ターゲット変数

Appendix 2.
Microsoft Azure Machine Learning Studio の起動

- 各ページの「Open in Studio」をクリック
  - 任意のマイクロソフトアカウントでサインインして利用することが可能

EXPERIMENT

Predictive Maintenance: Step 1 of 3, data preparation and feature engineering

By AzureML Team for Microsoft • April 27, 2015

7 likes

Open in Studio

ここをクリック

+ Add to Collection

Summary

This experiment demonstrates the steps in building a predictive maintenance solution.

Description

Predictive Maintenance Template

Predictive maintenance encompasses a variety of topics, including but not limited to: failure prediction, failure diagnosis (root cause analysis), failure detection, failure type classification, and recommendation of mitigation or maintenance actions after failure. As part of the Azure Machine Learning offering, Microsoft provides a template that helps data scientists easily build and deploy a predictive maintenance solution.

Appendix 3.
予兆保全に関する、その他の参考資料一覧

- Cortana Intelligence Solution Template Playbook for predictive maintenance in aerospace and other businesses
  - https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/cortana-analytics-playbook-predictive-maintenance
- Deep Learning Basics for Predictive Maintenance
  - https://gallery.cortanaintelligence.com/Tutorial/Deep-Learning-Basics-for-Predictive-Maintenance

Appendix 3.
予兆保全に関する、その他の参考資料一覧

- Predictive maintenance
  - Step 1 of 3 - data preparation and feature engineering
    - https://gallery.cortanaintelligence.com/Experiment/df7c518dcba7407fb855377339d6589f
  - Step 2A of 3 - train and evaluate regression models
    - https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-2A-of-3-train-and-evaluate-regression-models-2
  - Step 2B of 3 - train and evaluate binary classification models
    - https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-2B-of-3-train-and-evaluate-binary-classification-models-2
  - Step 2C of 3, train and evaluation multi-class classification models
    - https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-2C-of-3-train-and-evaluation-multi-class-classification-models-2
  - Step 3A of 3, deploy web service with a regression model
    - https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-3A-of-3-deploy-web-service-with-a-regression-model-2
  - Step 3B of 3, deploy web service with a binary classification model
    - https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-3B-of-3-deploy-web-service-with-a-binary-classification-model-2
  - Step 3C of 3, deploy web service with a muiti-class classification model
    - https://gallery.cortanaintelligence.com/Experiment/Predictive-Maintenance-Step-3C-of-3-deploy-web-service-with-a-muiti-class-classification-model-2

Appendix 4.
機械学習に関する参考情報

- ブースティング入門
  - https://www.slideshare.net/Retrieva_jp/ss-80724064
- Computing Classification Evaluation Metrics in R
  - http://blog.revolutionanalytics.com/2016/03/com_class_eval_metrics_r.html