

SuperKEKB加速器真空システムにおける 機械学習を応用した圧力異常検知手法の検討

末次祐介 KEK

2023/8/30

内容

- はじめに：動機
- 検知手法の基本方針
 - 運転状態、基準データ・調査データを定義
 - 基準データの回帰曲線を導出
 - 2クラス分けの2層Feedforward Neural Network (FNN)を構築
 - 過去のデータから決定境界用のFNN重みパラメータを学習
 - 結果を表示
- 模擬試験
- まとめ

はじめに：動機

- **SuperKEKB加速器**：KEKつくばキャンパスにある電子・陽電子衝突型加速器。
 - 高電流、極細衝突ビームで高いルミノシティーを目指す。
- 主リングの真空システムは2016年の運転開始以来順調に稼働しているが、ビーム電流が大きいこと、ビームロス・再入射の頻度が高いこと等から、**熱サイクル(熱応力)**に起因するリークや、**ビームパイプ内の放電等**による**変則的な圧力上昇**が度々発生している。
- 時には、運転に支障をきたすような重大な問題に発展する場合もある。
- 上記のような圧力異常の兆候を早い段階で検知できれば、大きな問題となる前にあらかじめ対策とる、あるいは準備することが可能となろう。
- そこで、機械学習を応用し、運転中の圧力異常の兆候をつかみ、注意喚起する**異常検知手法**を提案、検討している。

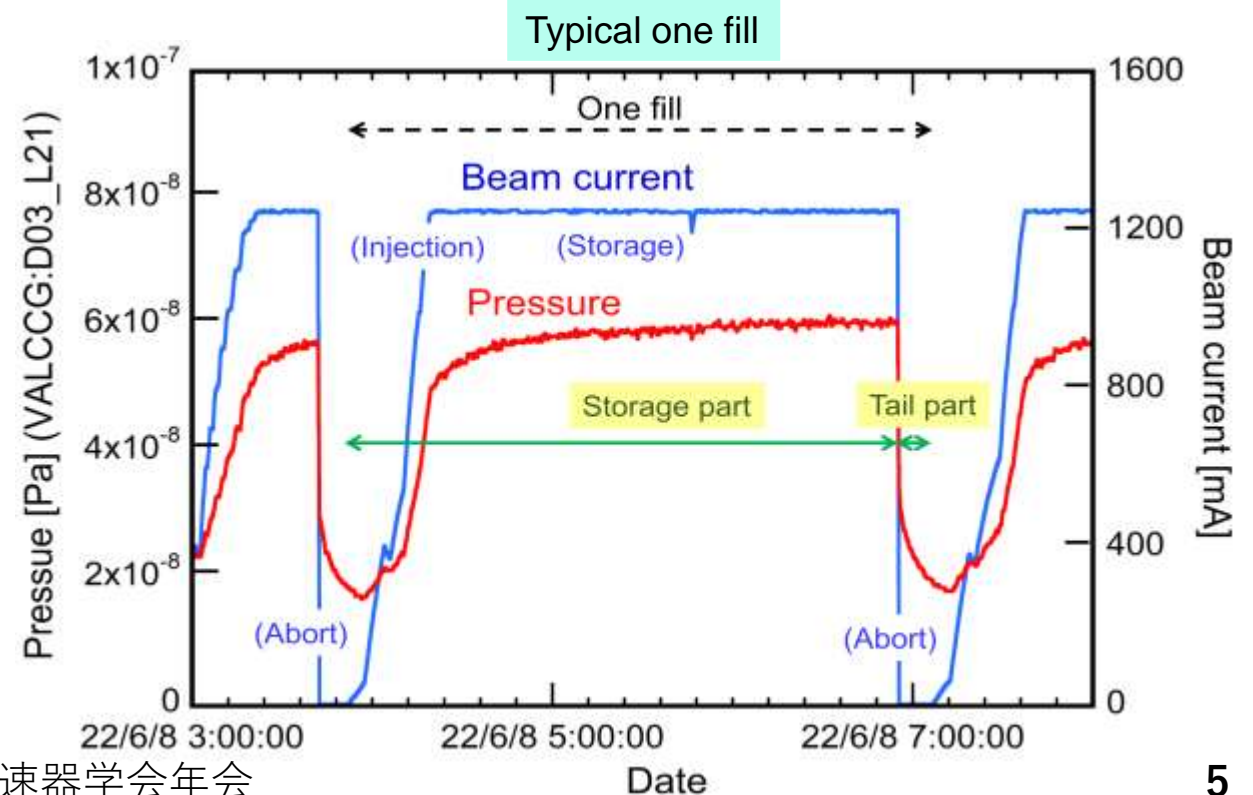
- **「異常」とは**：加速器の運転パラメータは日々変化するので、ここで言う「**圧力異常**」とは、直近の圧力のビーム電流や時間に対する振る舞いが、その少し(数日)前の振る舞いから予想されるものと違った場合、と定義する。

検知プログラムの基本方針

- **運転状態、データを定義**：運転状態を、ビーム入射からアボート直後までを”1フィル”とし、それを「ビーム(再)入射から蓄積(以後Storage部)」と「ビームアボート直後(以後Tail部)」の二つの部分に分け、それぞれの状態で、調べたい時(調査データ)の少し前の期間のデータを「基準データ」とする。
- **回帰曲線を求める**：基準データにおける圧力のビーム電流、あるいは時間に対する振る舞いの回帰曲線を求める。
- **FNNを構築**：求めた回帰曲線を用いて調査データの二乗平均平方根誤差(RMSE)等を計算し、それらの組み合わせを入力変数(特徴変数)として、「正常」と「異常」の2クラスに分類する2層フィードフォワードニューラルネットワーク (FNN) を構築する。
- **決定境界のための重みパラメータを学習**：分類の決定境界を与えるFNNの重みパラメータを過去の実際の真空トラブル発生時の圧力データから学習して決める。
- **検知結果を表示**：加速器制御システム上で異常検知プログラムを走らせておき、圧力異常が検知されたら当該真空計をGUI画面やアラームで加速器運転員に知らせる。プログラムは機械学習でよく使用されるPythonを使用する。

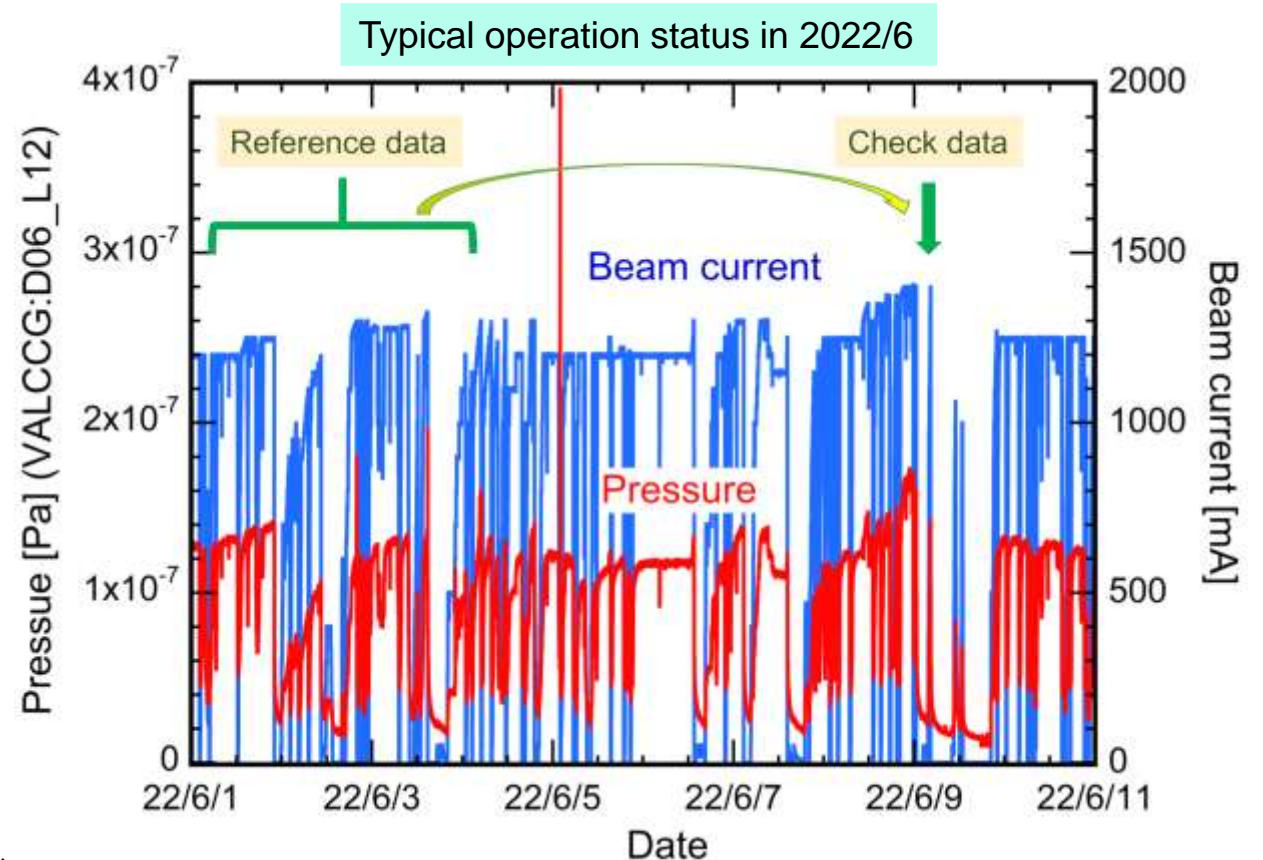
運転状態の定義

- SuperKEKB加速器では、平均すると1日に数回、何等かの原因でビームがアボートされる(捨てられる)。
- 圧力に関するトラブルは、熱負荷が大きくなる大電流蓄積中や、ビームがアボートされた直後の熱負荷の変化が大きく、熱応力が高まる時に発生しやすい。
- ここでは、運転状態を、ビーム入射からアボート直後までを”1フィル”とし、それを「ビーム(再)入射から蓄積(以後Storage部)」と「ビームアボート直後(以後Tail部)」の二つの部分に分ける。
- “1フィル”中の圧力異常上昇、ビームアボート直後の圧力バーストや異常な上昇の有無を調べる。
- アボート直後を調査するは、アボートの多いSuperKEKB特有か。



基準データ、調査データの定義

- 「異常」の定義：直近の圧力のビーム電流や時間に対する振る舞い(調査データ：Check data)が、その少し前の振る舞い(基準データ：Reference data)から予想されるものと違った場合。
- 異常の検知は、基本的にビームアポートする度に行う。→直近の1フィルを調査データとする。
- 基準データは、調査データの8日前から5日前までの3日間のデータとする。
- 右図は2022年6月の運転状況の例。
- ただし、蓄積時間が長くなると検知が遅れる可能性があるため、4時間以上アポートしないと、その4時間をStorage部の調査データとして調べる。(つまり少なくとも4時間に1度は調べる。)
- ビームが無い時(シャットダウン中とか)も同様。



回帰曲線の導出-1

- **Storage部**：各真空計で基準データの「**圧力 vs ビーム電流**」の回帰曲線を求める。
- 全圧力(P)はベース圧力(P_b)、放射光による**光刺激脱離**に起因する**圧力上昇(ΔP_s)**、コンポーネントの発熱による**熱脱離**に起因する**圧力上昇(ΔP_t)**の和で表せる。
- ΔP_s は**ビーム電流(I)**に比例する。一方、 ΔP_t は、

$$\Delta P_t \propto \exp\left(-\frac{E_d}{R(T + \Delta T)}\right) - \exp\left(-\frac{E_d}{RT}\right)$$

E_d : 吸着エネルギー
 R : 気体定数
 T : 温度

空間中の気体分子数と表面に吸着している気体分子数が準平衡状態にあるとする。

と表されるが、温度が室温に近い場合には**近似的に(ΔT)²に比例**することが分かっている [1]。

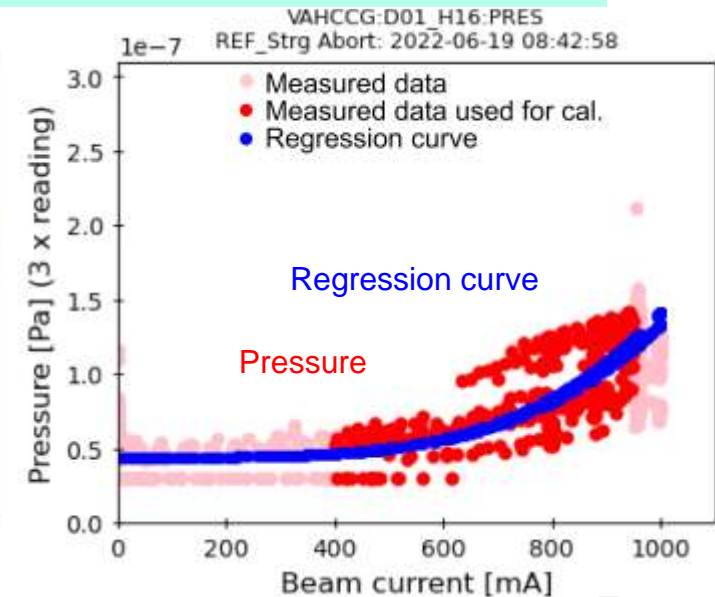
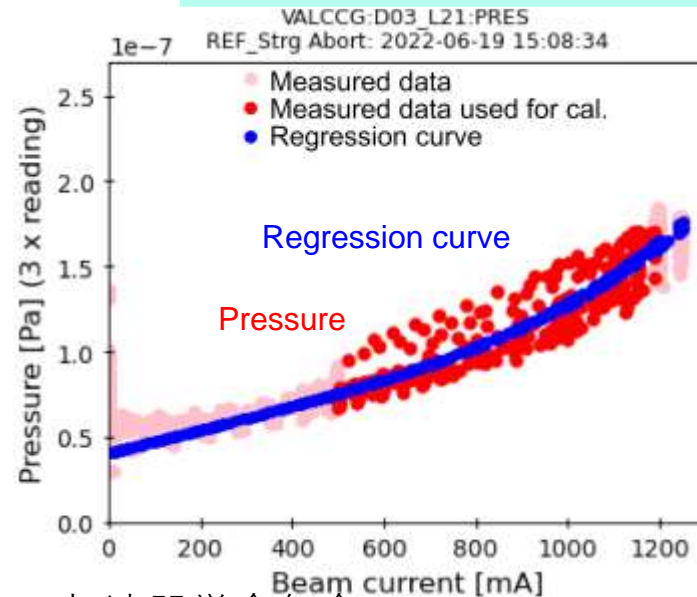
[1] Y. Suetsugu et al., PRAB (2023), doi: 10.1103/PhysRevAccelBeams.26.013201

- 熱入力が主にHOMパワーやジュール損だとすると、 $\Delta T \propto I^2/N_b$ なので、結局、

$$P(I) = P_b + \Delta P_s + \Delta P_t \\ = w_0 + w_1 I + w_2 (I^2/N_b)^2$$

- $w_0 \sim w_2$ は定数。 N_b はバンチ数。
- $w_0 \sim w_2$ は >0 とする(合理的仮定)。
- **平均二乗誤差(Mean Square Error, MSE)**が最小になるよう**勾配法**で決定。解析解あるいは数値解を使用。

Examples of pressures and regression curves for the Storage part



回帰曲線の導出-2

- **Tail部**：各真空計で基準データの「**圧力 vs 時間(アボート後)**」の回帰曲線を求める。
- ビームパイプ内の気体の排気は、**空間内の気体の排気**による圧力低下 ΔP_v と、**パイプ内表面に吸着していた気体の排気**による圧力低下 ΔP_w の2つに大きく分けることができる。
- ΔP_v の時間 t に対する変化は $\exp(-k_1 t)$ に比例する(k_1 は定数)。一方、 ΔP_w は $-1/t$ に比例することが経験的に知られている。つまり、

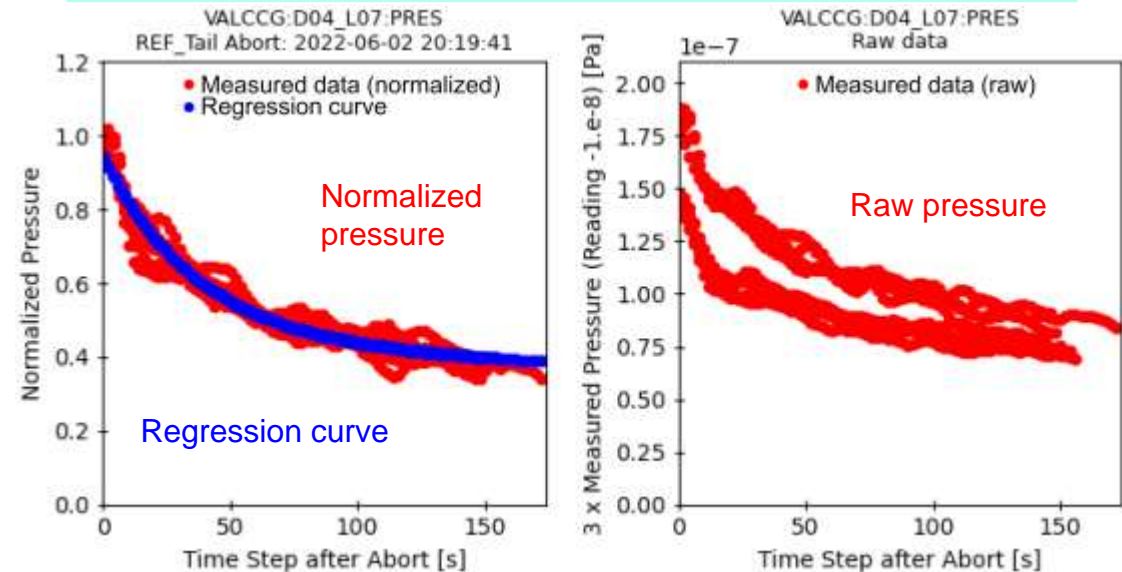
$$P = P_b + \Delta P_v + \Delta P_w = P_b + k_0 \exp(-k_1 t) + \frac{k_2}{t + k_3} \quad k_0 \sim k_3: \text{定数}$$

- しかし、ビームアボート時($t=0$)の圧力は、その時のビーム電流に依存する。そこで、圧力上昇($P - P_b$)をアボート時の圧力上昇($P - P_b$) $|_{t=0}$ で規格化した圧力上昇、 ΔP_n をここでは考える。

$$\Delta P_n \equiv \frac{P - P_b}{(P - P_b)|_{t=0}} = w_0 \exp(-w_1 t) + \frac{w_2}{t + w_3}$$

- $w_0 \sim w_3$ は定数。簡単のため、ベース圧力 P_b は真空計の測定限界である 1×10^{-8} Paとした。
- **平均二乗誤差(MSE)**が最小になるよう**勾配法**で**決定**。Pythonのモジュール"**minimize**"を使用。

Example of pressure and regression curve for the Tail part



FNNの構築-1

- **Storage部** : ここでは、下記三つのパラメータを入力変数とした。

$$x_0 = \log(R_T/R_{REF})$$

$$x_1 = \log(R_I/R_{REF})$$

$$x_2 = \log(R_T/P_{AV})$$

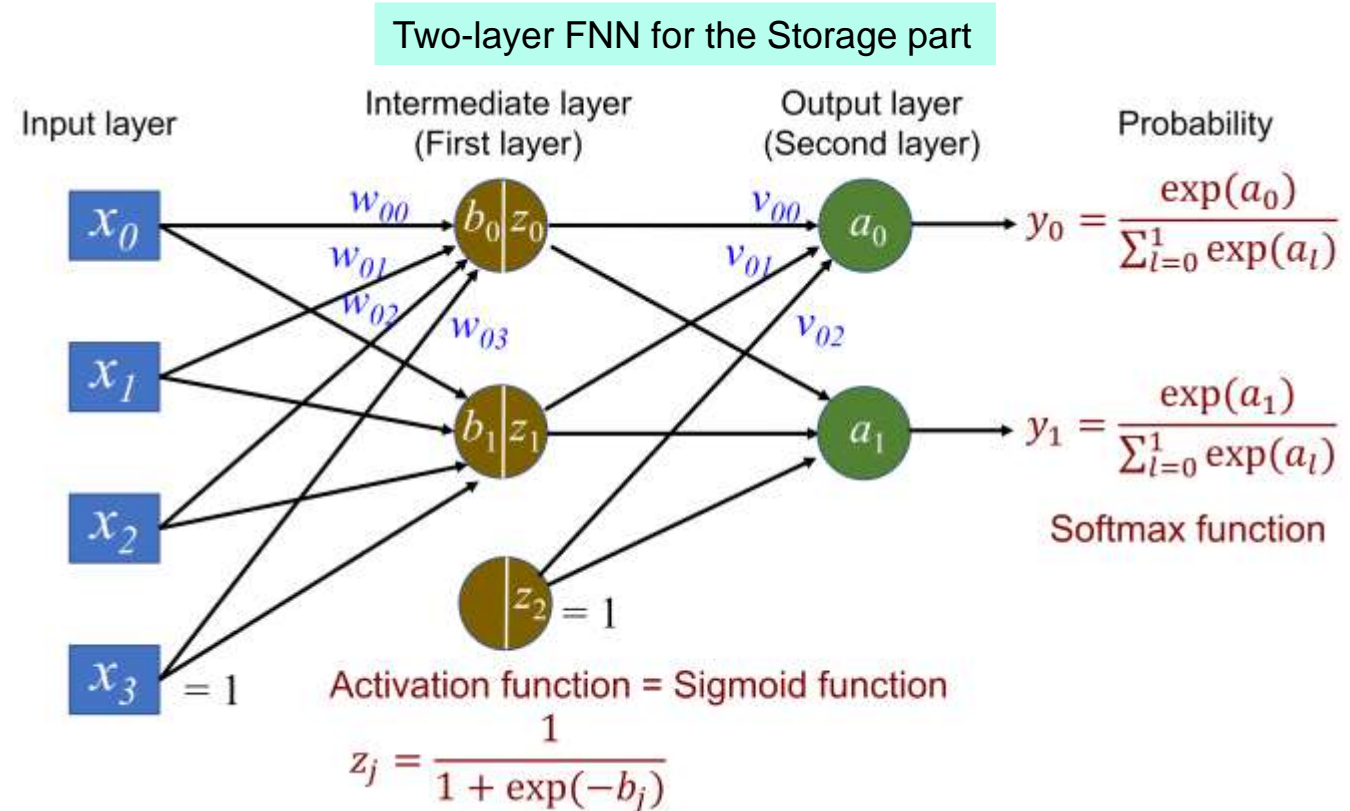
R_{REF} : 基準データの二乗平均平方根誤差 (Root Mean Square Error, RMSE)

R_T : 等時間間隔(30秒)の調査データのRMSE

R_I : 等電流間隔(50 mA毎)の調査データのRMSE

P_{AV} : 圧力の平均値 (圧力が高いと同じ割合の変化でもRMSEは大きくなるため)

- 調査データと基準データのRMSEの比は、データのばらつきが**正規分布**していると仮定した場合の「**異常度**」として通常用いられる。
- 圧力は桁で変わるので**対数**をとった。
- **3次元入力2層のFNN**を構築。 x_3 は常に1。
- w_{ji} 、 v_{kj} は重みパラメータ。
- 活性化関数を**シグモイド関数**、確率を出力する関数として**ソフトマックス関数**を用いた。



FNNの構築-2

- **Tail部**：ここでは、下記2つのパラメータを入力変数とした。

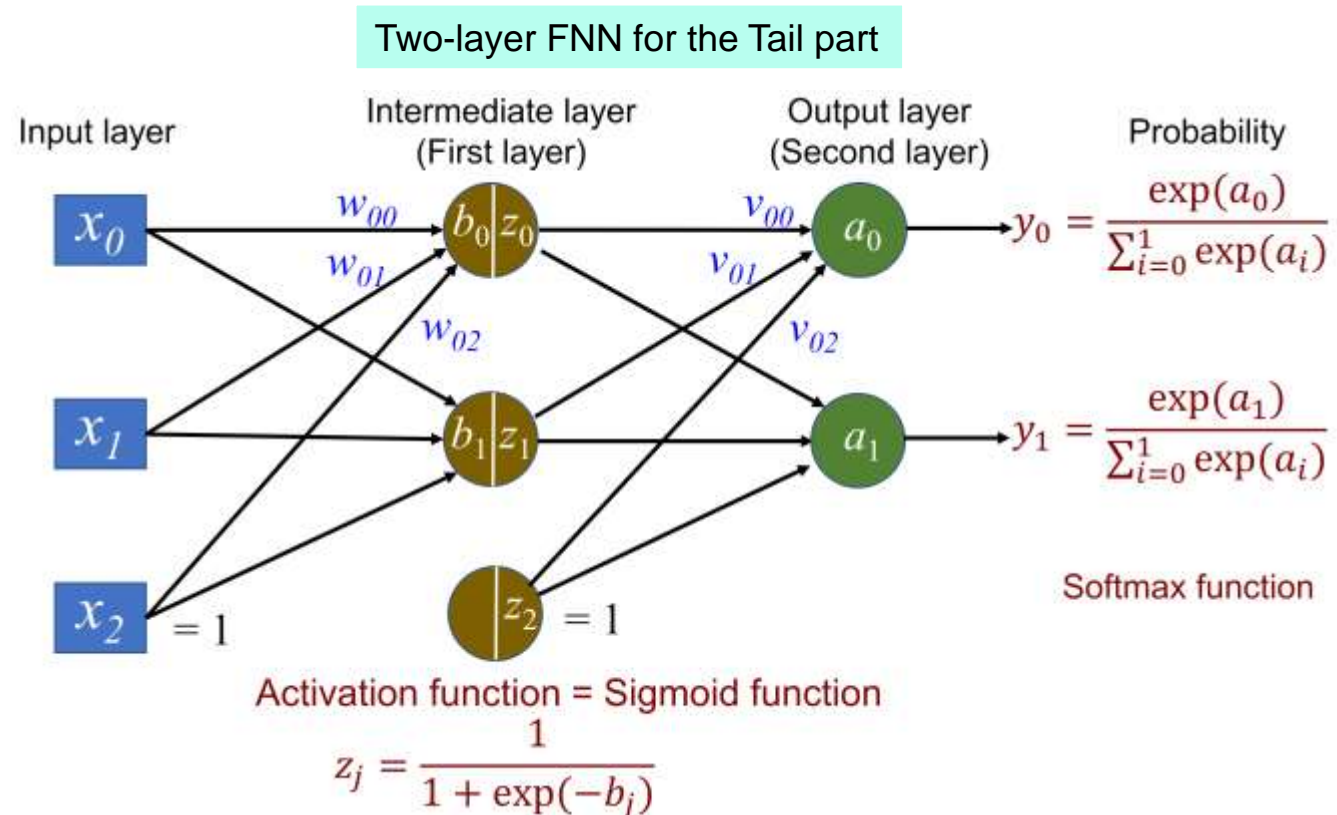
$$x_0 = \log(R_T/R_{REF})$$

$$x_1 = \log(R_T)$$

R_{REF} : 基準データの二乗平均平方根誤差 (Root Mean Square Error, RMSE)

R_T : 等時間間隔(2秒)の調査データのRMSE

- Storage部と基本的に同様。ただし、規格化した圧力を使っているので P_{AV} で割るパラメータは使わない。
- 2次元入力2層のFNNを構築。 x_2 は常に1。
- w_{ji} 、 v_{kj} は重みパラメータ。
- 活性化関数をシグモイド関数、確率を出力する関数としてソフトマックス関数を用いた。



FNN重みパラメータの学習

- 2016年から2022年までに**実際発生した圧力異常(リーク等)20例**(下表)について各時期のデータを使い、「正常」、「異常」を判断する(つまり決定境界用の)**重みパラメータ**を学習した。
- 電子リングと陽電子リングで使用するデータは区別しなかった。
- 全体(HER、LER)のStorage時、Tail部のデータをそれぞれシャッフルして二分し、**Training data(学習データ)**と**Test data(テストデータ)**に分けた。
- 全部で1052個のデータ(Storage 758, Tail 294)。

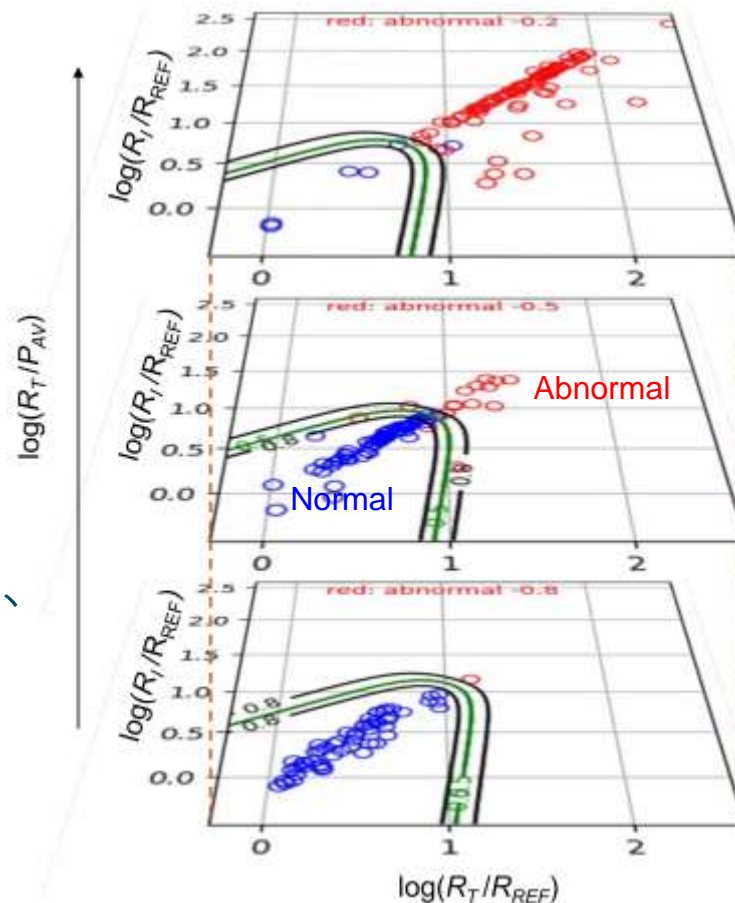
Vacuum troubles used for learning the weight parameters

LER	2016/06/18 10:02:00	大穂ウィグラー部フランジからの漏れ	HER	2020/5/16 6:30:00	富士D08H3GV付近圧力異常
LER	2020/12/5 17:23	日光冷却水ポンプ故障	HER	2022/6/3 9:38:00	D01H16、D08H05圧力バースト
LER	2020/12/7 6:14	日光冷却水ポンプ故障	HER	2022/6/14 3:09:56	D09V1近くの圧力上昇
LER	2022/3/18 11:40	大穂ウィグラー部下流冷却水管からの漏れ	HER	2022/6/14 14:34:06	D09V1近くの圧力上昇
LER	2022/4/26 10:29	大穂ウィグラー部下流冷却水管からの漏れ	HER	2022/6/19 2:49:42	D01_H3GVの発熱
LER	2022/5/17 14:38	D06H1近くの圧力異常	HER	2022/6/19 8:42:58	D01_H3GVの発熱
LER	2022/6/2 20:19	大穂ウィグラー部フランジからの漏れ	HER	2016/5/27 2:32:01	D08_H24リーク
LER	2022/6/2 21:31	大穂ウィグラー部フランジからの漏れ	HER	2018/6/29 12:01:55	D02_H22, H21Aリーク
LER	2022/6/3 14:48	大穂ウィグラー部フランジからの漏れ、QCSクウェンチも	HER	2018/7/14 14:17:37	D02_H22リーク
LER	2022/6/5 1:59	富士D08_L17圧力計異常か			
LER	2022/6/8 17:31	D06H1近くの圧力異常、D06V1バースト			

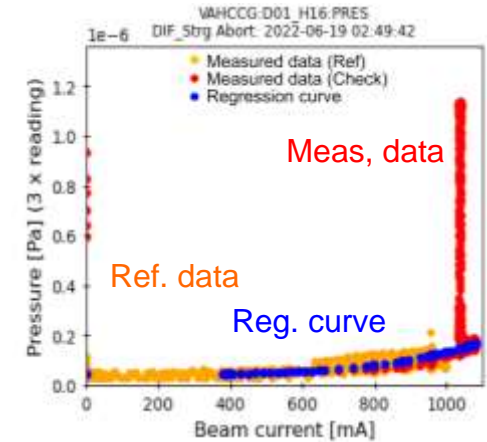
FNN重みパラメータの学習-1

- 平均交差エントロピー誤差(Mean Cross Entropy Error, MCEE)を最小にするように最適化し(勾配法)、重みパラメータ w_{ji} 、 v_{kj} を求める。
- **Storage部**の結果
- 右図青い丸、赤い丸がそれぞれ「正常」、「異常」時の入力パラメータ。
- 緑の線は、出力 y_0 が0.5となる線で、学習した「正常」と「異常」の境界線(決定境界)を意味する。
- 訓練用データとテストデータとも誤差が小さくなるように、すなわち、**過学習**とならないように注意した。

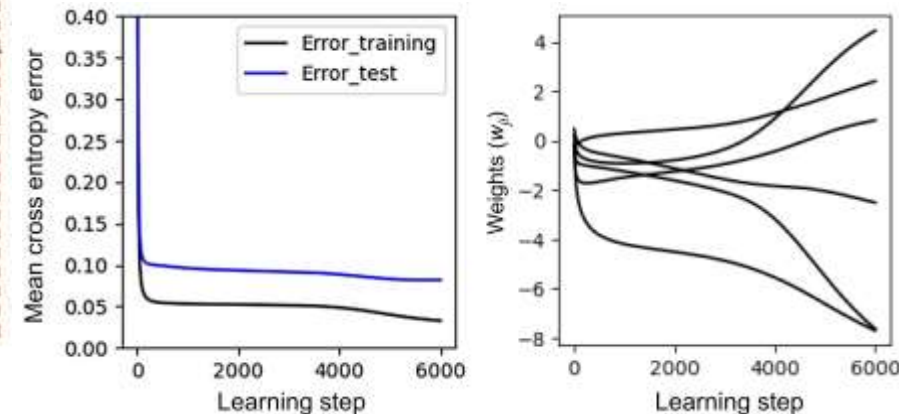
Results of learning



Example of abnormal behavior



MCEE and weight parameters vs learning step



FNN重みパラメータの学習-2

- MCEEを最小にするように最適化し(勾配法)、重みパラメータ w_{ji} 、 v_{kj} を求める。

- **Tail部**の結果

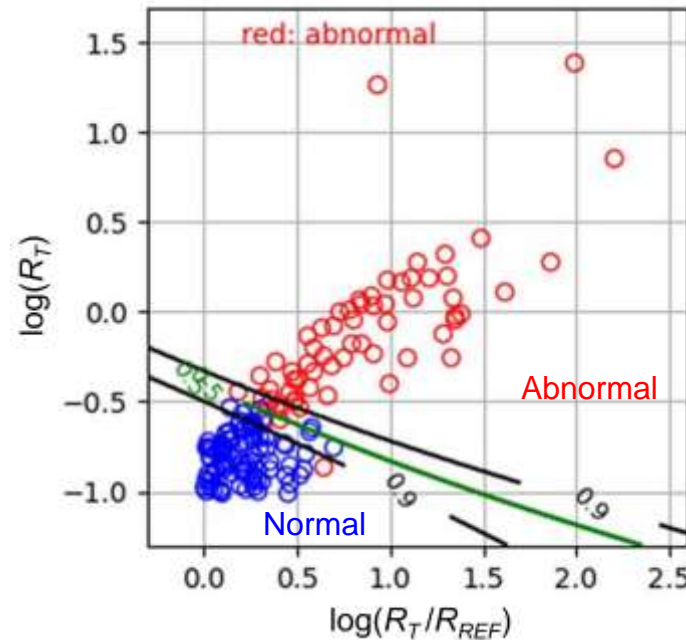
- 右図、青い丸、赤い丸がそれぞれ「正常」、「異常」時の入力パラメータである。

- 緑の線は、出力 y_0 が0.5となる線(決定境界)。

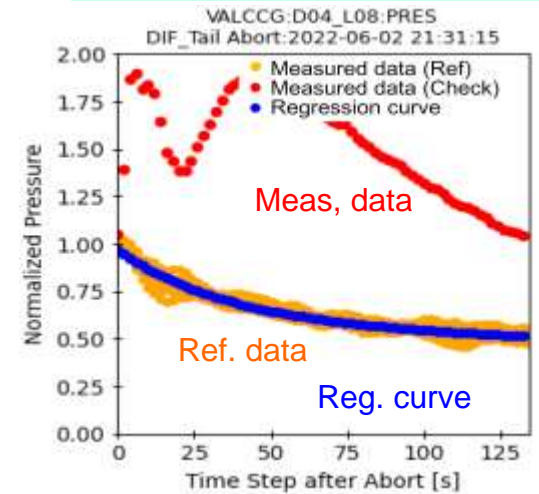
- 同様に、訓練用データとテストデータとも誤差が小さくなるように、すなわち、過学習とならないように注意した。

- Storage部、Tail部とも、うまくクラス分けできている。
- この重みパラメータを使って新たな調査データを判定していく。

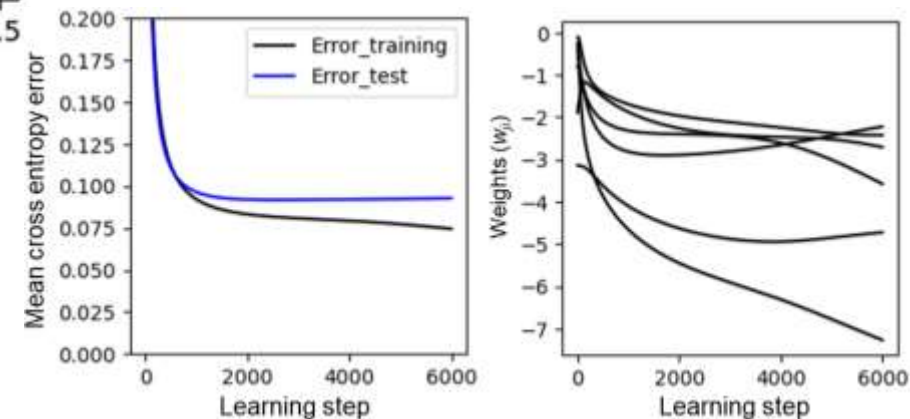
Results of learning



Example of abnormal behavior



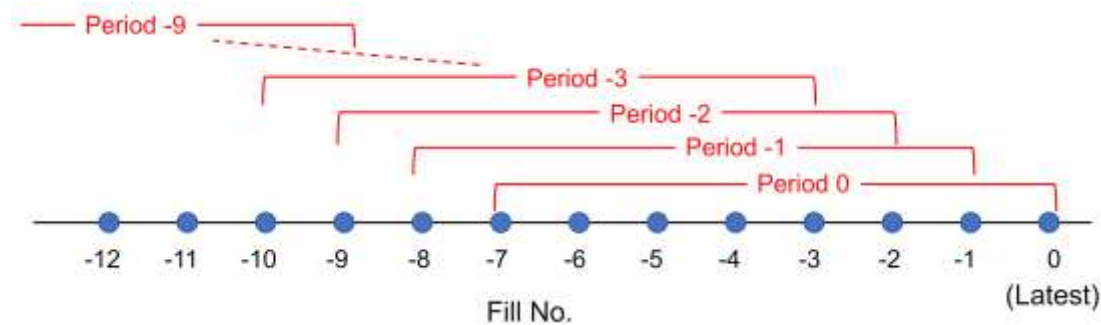
MCEE and weight parameters vs learning step



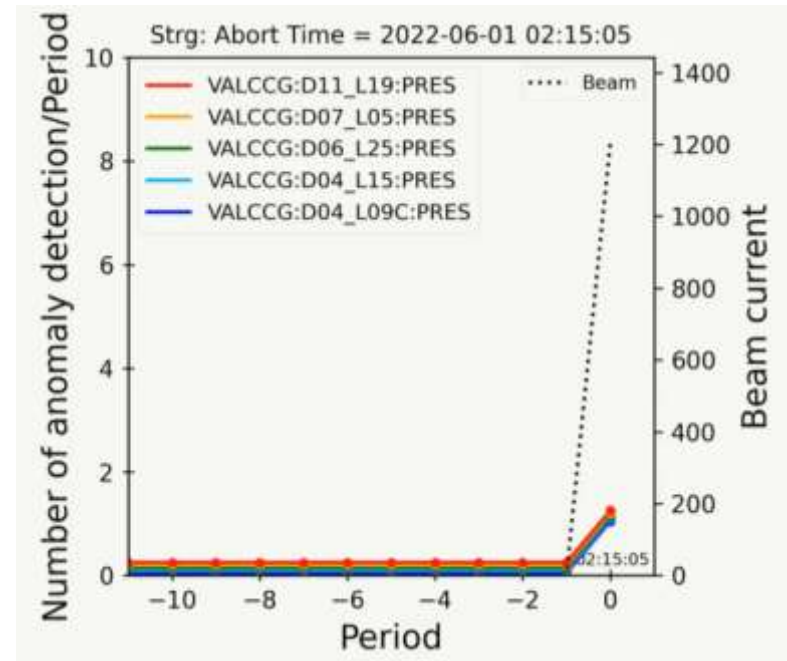
検知と結果の表示

- アボート毎、あるいは4時間毎の各調査データで「異常」と判断された真空計をGUI画面に表示する。
- しかし、1回のみ調査ではその真空計の異常の履歴や傾向がわからないため、8回前の調査データ(フィル)までを1調査期間(Period)とし、この期間内に該当する真空計が異常と判断された回数をカウントする(右上図)。
- この作業を繰り返すことで、異常状態が続けばその真空計のカウント数が上がり、異常が発生し始めたことを示す。
- 表示プログラムでは、最新のPeriod (Period 0)で異常と判断された回数の多いものから最大5個の真空計を選び、10回前のPeriodからそのカウント数をプロットする(右下図)。同時に真空計名(EPICS Record)も表示する。
- 例えば、5回以上になると、目立よう工夫する。

Definition of "Period" to count the number of anomalies



Histories of numbers of detected anomalies

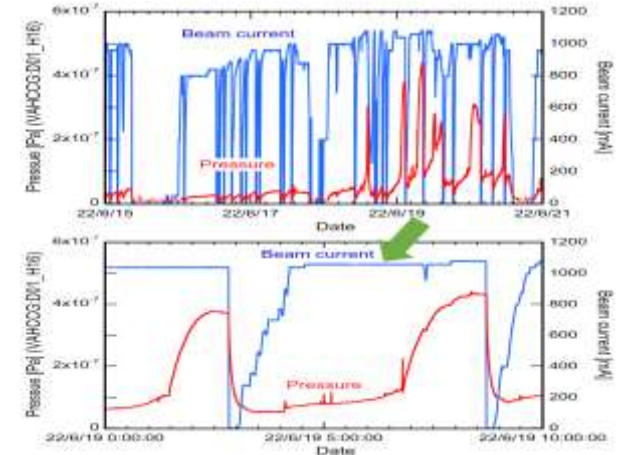
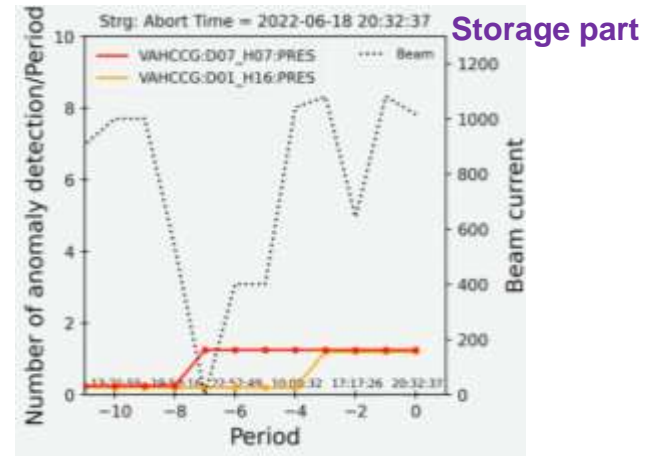


模擬試験

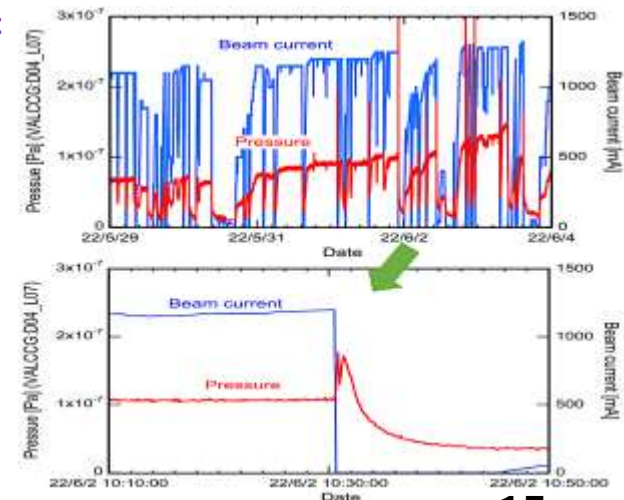
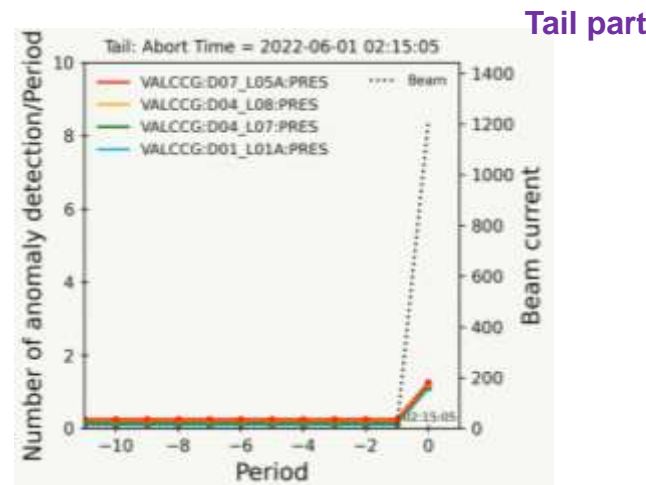
- 2022年の運転時のデータや2023年のシャットダウン時のデータを用いて模擬試験を行った。

- 右上図は運転中のStorage部の例で、左は異常と判断された真空計とそのカウント数の履歴を示し、真空計'VAHCCG:D01_H16'が10 Period前(Period -10)から「異常」を示し始めていることを示す。右はその時の圧力(表示値)の実際の振る舞いで、6/18頃からほぼ同じ蓄積ビーム電流でも圧力が急に上昇している。圧力は 10^{-7} Pa領域で、未だアラーム発報には至らない。

Histories of numbers of detected anomalies (left), and the time trends of pressure at the corresponding times (right),



- 右下図は、同様に、Tail部で真空計'VALCCG:D04_L07'が「異常」を示し始めた場合の例で、圧力の突出が頻繁に見え始めている。



- 異常を検知できそう。

まとめ

- 基本的な機械学習を応用してSuperKEKB主リングの圧力データを調査し、真空トラブルの予兆を捉える**圧力異常検知手法**を検討。
- 運転時の**Storage部**、**Tail部**のビーム電流、時間に対する変化について、基準データを用いて**合理的モデルに基づく回帰曲線**を導出。
- 調査データでのRMSE等を入力パラメータとして**2層のFNNでの2クラス分け**を行い、「正常」「異常」を判断。
- **決定境界**はこれまでの**実際の圧力トラブル時のデータ**で学習した。
- 2022年6月の運転時の圧力データや2023年シャットダウン中の圧力データを使って**模擬試験**を実施し、真空計の**異常検知件数の頻度を調べ**表示させることで、**異常が発生し始めている真空計を特定できる**ことが分かった。
- 異常検知手法として使える感触が得られたので、**2023年の運転再開時から試用する予定**である。
- 今後の課題としては以下の点が挙げられる。(1) **異常判断基準の精度向上**。今は明らかな異常状態のデータを使って学習している。(2) **異常な期間が長く継続した場合の判断**。(3) **プログラムの高速化**。(4) **「異常」の種類(リーク、放電など)の推定**。(5) **温度の異常検知などへの応用**。引き続きこれらについて検討を進める。

謝辞

- 本研究にあたり、KEKB真空グループの皆様、KEKBモニターグループの三塚岳氏にはいろいろ相談に乗って頂きました。ここに深く感謝いたします。

ご清聴ありがとうございました。