Jan.19th, 2015/BRSystems Inc.

PCA 主成分分析　解説書

１．主成分分析の目的

平均偏差（変数と平均の差）の分散が最も大きくなるようにしたものによって、評価したい時。

例えば、

① 生理学的な systemic noise を低減し、activation pattern を抽出します。
② 他のバイオマーカー測定値と相関する成分を抽出し、波形を復元します。
③ 同一被験者の複数タスク間の相似性、非相似性を確認します。

２．分析手法（分散の最大化）

手法としては 3 種類有ります。特異値分解法(SVD,Singular Value Decomposition、文献１)、共分散行列の固有値分解法(EIG,Eigenvalue Decomposition of the covariance matrix、文献２)と交互最小 2 乗法(ALS,Alternating Least Squares）です。

EIG は、信号数 p よりサンプル数 n が大きい場合、SVD より処理速度が速くなりますが、精度は落ちると言われています。

本ツールでは、SVD を採用します。

３．SVD 法

「線形代数　基礎と応用、新井仁之、日本評論社」のページ 326 から抜粋しています。

x_{ij} は n 行 p 列の測定データ（n はサンプル数、p は信号チャネル数）とし、
w_{ij} は係数（方向比）とすると、
次の式であらわされる z を主成分とします。m は主成分の数（m≤p<n）。

$$z_1 = w_{11}x_{11} + w_{12}x_{12} + \cdots \quad + w_{1j}x_{1j} + \cdots + w_{1p}x_{1p}$$
$$z_2 = w_{21}x_{21} + w_{22}x_{22} + \cdots \quad + w_{2j}x_{2j} + \cdots + w_{2p}x_{2p}$$

$$z_i = w_{i1}x_{i1} + w_{i2}x_{i2} + \cdots \quad + w_{ij}x_{ij} + \cdots + w_{ip}x_{ip}$$

$$z_m = w_{m1}x_{m1} + w_{m2}x_{m2} + \cdots \quad + w_{mj}x_{mj} + \cdots + w_{mp}x_{mp}$$

z の分散が最大化する（データのばらつきを最も良く表す）係数w_{ij}を算出すれば、主成分が求まります。

次の平均偏差行列

$$\tilde{X} = \begin{bmatrix} x_{11} - \frac{\sum_{i=1}^{n} x_{i1}}{n} & \cdots & x_{1p} - \frac{\sum_{i=1}^{n} x_{ip}}{n} \\ \vdots & \ddots & \vdots \\ x_{n1} - \frac{\sum_{i=1}^{n} x_{i1}}{n} & \cdots & x_{np} - \frac{\sum_{i=1}^{n} x_{ip}}{n} \end{bmatrix}$$

の特異値分解をします。$\tilde{X}$ の特異値を

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_p \geq 0$$

とします。

この時、特異値分解

$$\tilde{X} = \sum_{j=1}^{p} \sigma_j \boldsymbol{u}_j \boldsymbol{w}_j^{\boldsymbol{T}}$$

を与えるような実数ＲのR^nの正規直交基底$\{u_1, \cdots, u_n\}$とR^pの正規直交基底$\{w, \cdots, w_p\}$が存在します。

w_jを第ｊ主成分ベクトル、$\langle x_n - x,\ w_j \rangle$を個体ｎの第ｊ主成分得点

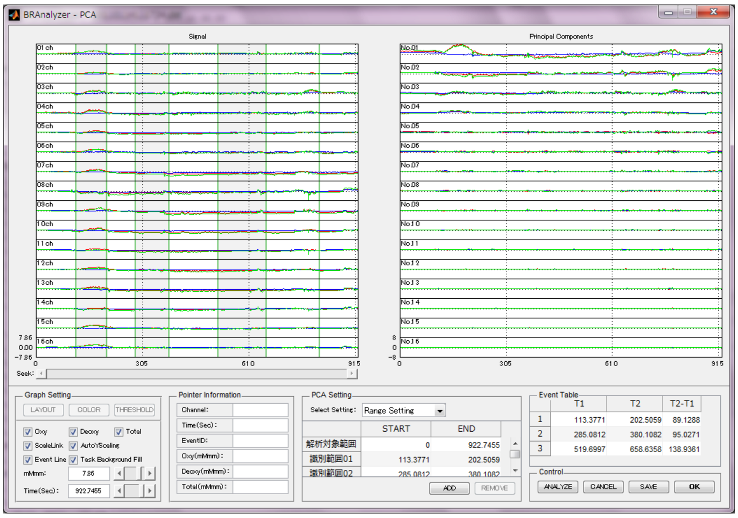
（(平均偏差行列）＊（固有ベクトル）＝主成分得点）と言います。

４．任意の主成分を除去した後の波形復元

任意の主成分を除去したい場合、主成分得点の図から除去したい成分を選択しますと、その主成分得点の列の値は０となります。

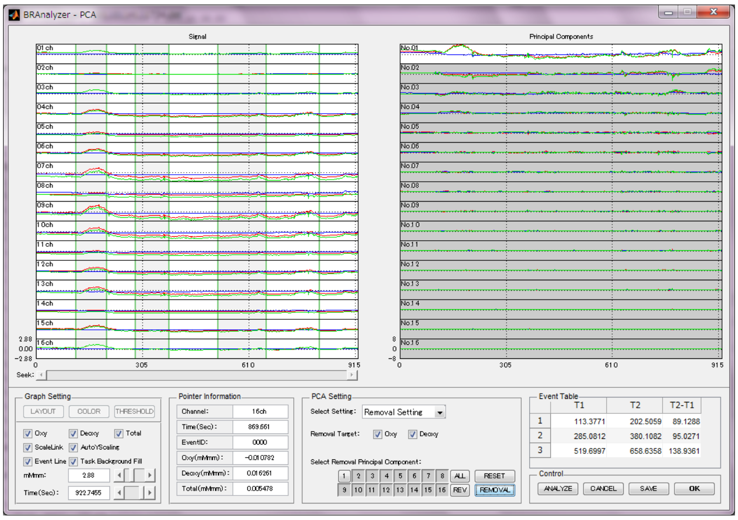
（修正後の波形）＝（修正後の主成分得点）＊（固有ベクトル）の逆行列

を計算することに拠り、任意の主成分を除去した後の波形復元が可能となります。

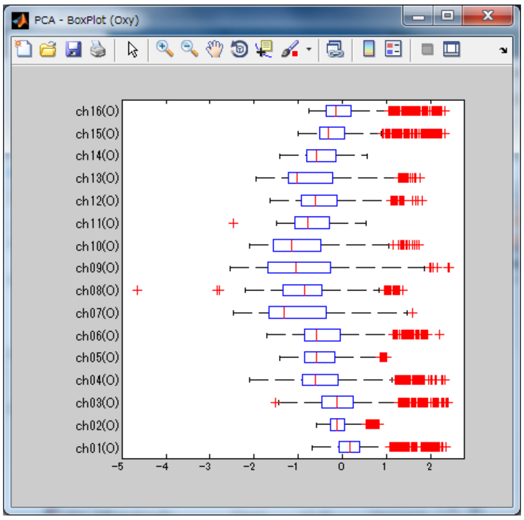


左図は原波形で、右図は主成分得点の図です。

例えば、１次の主成分得点のみ残し、他成分を削除しまして波形を復元しますと、下図の波形を得ます。

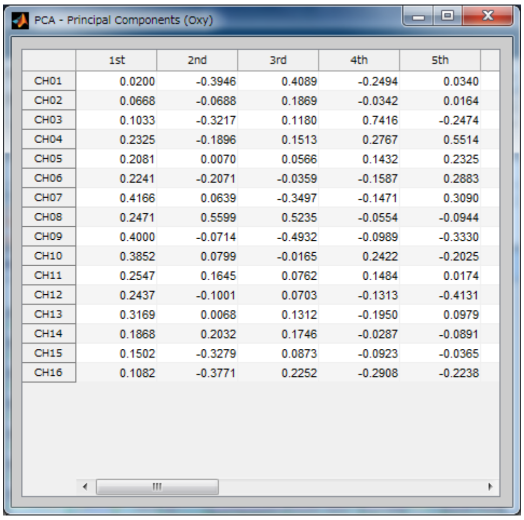


5．PCA 出力の説明



（1）ボックスプロット図

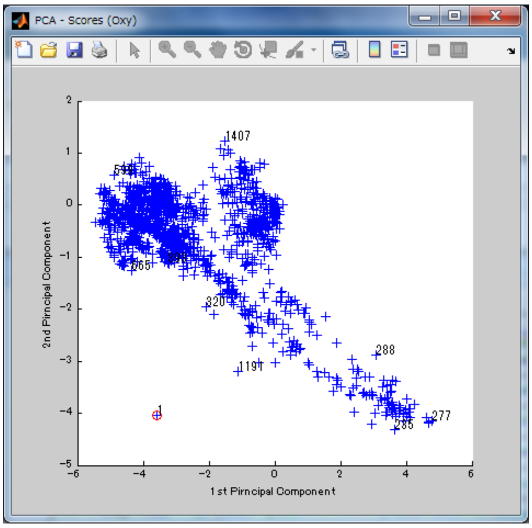
関数 boxplot(X) は、X のデータの箱ひげ図を作成します。X が行列の場合は、列ごとに 1 つのボックスがあり、X がベクトルの場合は、ボックスが 1 つだけあります。各ボックスにおいて、中心の印は中央値で、ボックスのエッジは 25 と 75 の百分位数です。ひげはアルゴリズムが外れ値でないと考慮する最極のデータ点に伸びます。


PCA - Principal Components (Oxy)

	1st	2nd	3rd	4th	5th
CH01	0.0200	-0.3946	0.4089	-0.2494	0.0340
CH02	0.0668	-0.0688	0.1869	-0.0342	0.0164
CH03	0.1033	-0.3217	0.1180	0.7416	-0.2474
CH04	0.2325	-0.1896	0.1513	0.2767	0.5514
CH05	0.2081	0.0070	0.0566	0.1432	0.2325
CH06	0.2241	-0.2071	-0.0359	-0.1587	0.2883
CH07	0.4166	0.0639	-0.3497	-0.1471	0.3090
CH08	0.2471	0.5599	0.5235	-0.0554	-0.0944
CH09	0.4000	-0.0714	-0.4932	-0.0989	-0.3330
CH10	0.3852	0.0799	-0.0165	0.2422	-0.2025
CH11	0.2547	0.1645	0.0762	0.1484	0.0174
CH12	0.2437	-0.1001	0.0703	-0.1313	-0.4131
CH13	0.3169	0.0068	0.1312	-0.1950	0.0979
CH14	0.1868	0.2032	0.1746	-0.0287	-0.0891
CH15	0.1502	-0.3279	0.0873	-0.0923	-0.0365
CH16	0.1082	-0.3771	0.2252	-0.2908	-0.2238

（2）主成分係数ベクトル

主成分係数ベクトル（固有ベクトル、直交行列）の表です。



（3）主成分得点図

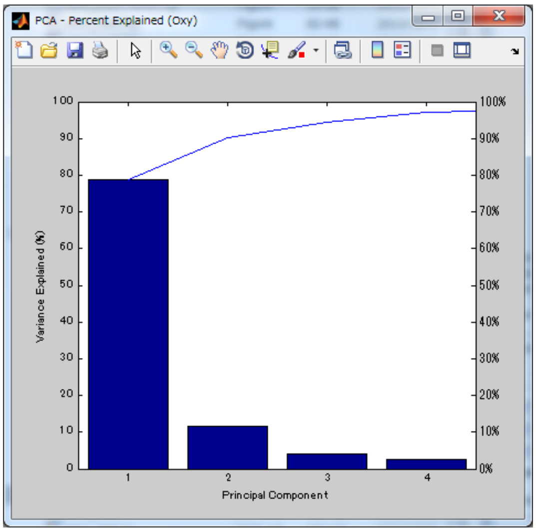
サンプル数 n の最初の点から最後の点までの、各サンプルの 1 次主成分得点と 2 次主成分得点の点を図示します。カーソルを持っていきますと、何番目かの番号が表示されます。赤丸は最初の点を表します。

PCA - Components Scores (Oxy)

	1st	2nd	3rd	4th	5th
1	-0.0012	-0.0041	0.0953	0.1065	0.1791
2	0.0505	-0.0430	-0.0516	0.0600	0.0868
3	0.1127	-0.0498	0.0172	0.0855	0.0135
4	0.2407	0.0036	-0.0174	-0.0618	-0.0799
5	0.1355	-0.1094	0.1285	0.0584	0.1169
6	0.0797	-0.0268	0.0746	0.0570	0.0248
7	-0.2382	0.0874	-0.0519	-0.0399	-0.3148
8	-0.1133	0.0760	-0.1551	-0.1446	-0.1302
9	-0.1387	0.1013	-0.0377	-0.0370	0.0435
10	-0.1147	-0.0441	0.0029	-0.0775	0.0684
11	-0.1475	0.0288	0.0245	-0.1864	-0.1743
12	0.0635	-0.0322	-0.0447	-0.0729	0.0341
13	-0.0258	-0.1021	0.0331	0.1618	0.2004
14	-0.0359	0.0135	-0.0684	-0.0559	-0.0304
15	0.0708	-0.0431	-0.0043	-0.0901	0.1973
16	-0.0774	0.0183	-0.0041	-0.1251	0.0777
17	-0.0848	0.0670	-0.1284	-0.0451	-0.0622
18	-0.1664	-0.0448	-0.0376	-0.0572	0.0035
19	-0.1250	0.0313	-0.0951	-0.0573	0.2760
20	-0.3752	0.1312	-0.0974	-0.0641	-0.1468
21	-0.2721	0.0673	-0.0441	-0.1036	0.0508

（4）主成分得点の表

１次から 16 次までの主成分得点の表です。横方向が次数で、縦方向がサンプル数です。



（5）主成分の割合図

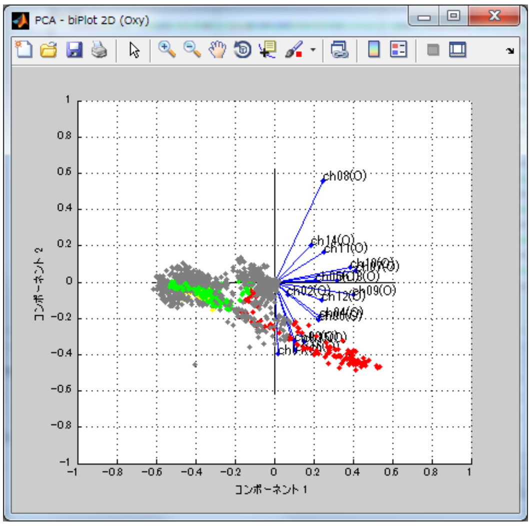
各主成分の寄与率の図です。

PCA - Components Variances (Oxy)

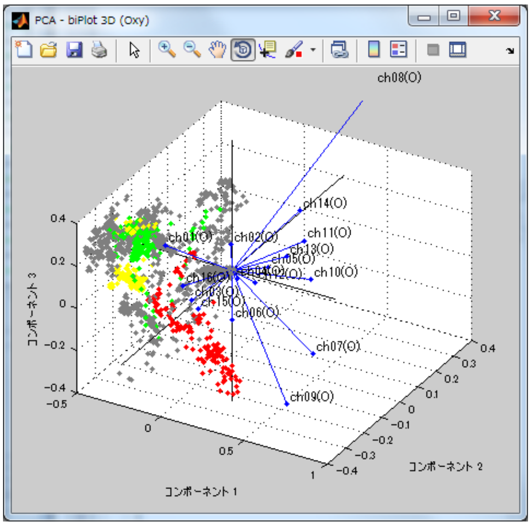
	1
1st	9.7655
2nd	1.4194
3rd	0.5194
4th	0.3161
5th	0.1154
6th	0.0811
7th	0.0436
8th	0.0348
9th	0.0219
10th	0.0204
11th	0.0138
12th	0.0131
13th	0.0097
14th	0.0055
15th	0.0053
16th	0.0037

（6）各主成分の分散表

各主成分によって説明される分散の表です。



（７）２ 次元座標（１次、２ 次）表示の各信号(ch)の主成分係数と全測定値の主成分得点を表示



（８）３ 次元座標（１次、２ 次、３ 次）表示の各信号(ch)の主成分係数と全測定値の主成分得点を表示
各タスクで色を変えていますので、タスク間の比較が可能となります。

以上

[参考文献]

１．「線形代数　基礎と応用」、新井仁之、日本評論社（2006）
２．「入門はじめての多変量解析」石村貞夫、東京図書（2007）
３．「主成分分析」MATLAB （2013a）