ComBat法を用いた複数プロトコルデータセット のHarmonize法

東京大学進化認知科学研究センター 国立研究開発法人 国立精神・神経医療研究センター脳病態統合イメージングセンター

伯秀

舞草

背景

- MRIによる疾患特異的な構造・機能変化の抽出は、再現性・客 観性・定量性に優れる疾患の代理バイオマーカーになる
- ・疾患の進行動態解明のため、有効なMRIバイオマーカー策定のため、多数の疾患例を取得する必要であり、多施設研究が有効
- しかしMRIは撮像装置間のMeasurement Bias(Site Effect)が 認められており、補正(Harmonization)が必要



Ex.)



本日は ComBatによる構造MRIについてのHarmonizationの実践 を行います。

Combatについて

元々はマイクロアレイの機種間補正法として開発(Johnson et al., 2007) MRIのFAマップ(Fortin et al., 2017)、皮質厚(Frotin et al., 2018)の補正に適応

$$y(i,j,v) = a(v) + \mathbf{X}^{T}(i,j)\mathbf{\beta}(v) + \gamma(i,v) + \delta(i,v)\varepsilon(i,j,v)$$

a(v): 特徴量全平均 γ(*i*, v), δ(*i*, v): Site Effect

 γ^*_{iv} and δ^*_{iv} :経験的ベイズ推定により求める 補正された特徴量(皮質厚など)は次式

$$y_{(i,j,v)}^{combat} = \frac{y_{(i,j,v)} - \hat{a}_{(v)} - X_{(i,j)}\hat{\beta}_{(v)} - \gamma_{(i,v)}^*}{\delta_{iv}^*} + \hat{a}_{(v)} + X_{(i,j)}\hat{\beta}_{(v)}$$

注意点

特徴量は**VOI測定値**か解剖学的標準化されたVoxel値に対応 生画像の補正はできない ただしVoxel値の場合、標準化の際の位置ずれがない仮定

ComBat <u>https://github.com/Jfortin1/ComBatHarmonization</u> より入手可能

Jean-Philippe F	ortin committed 0e10dde 19 days ago 🛄 🗸	🕒 201 commits 🛛 🗜 1 branch 🛛 🕤 tags
📄 Matlab	Fixed typo combat code	19 days ago
Python	Updated neuroCombat	last month
R	Updated readme	25 days ago

MATLAB, Python, Rが準備されていますが、今回はMATLABでのデモを行います

MATLABは有償ソフトですので、お持ちでない方は1か月の試用版を入手されても良いのですが、試用後営業の電話がかかってきますので、聴講のみでも結構です

Python, Rでも基本的な流れは同じです

準備編1

● ComBat・チュートリアル用サンプルデータのダウンロード

☐ Jfortin1	/ ComBatHarmoniza	tion				() Wat
<> Code	() Issues 5 11 Pull r	equests 🕞 Actions	III Projects	() Security	🗠 Insights	
2º master	→ P1 branch ⊗0 tag	GitHub is home to	Joir over 50 million d manage projec	n GitHub t evelopers work ts, and build so Sign up	oday ing together to host and review oftware together. Go to file	w code,
Jean-I	Philippe Fortin Updated neuro	Combat to 0.2.7		Clone with Use Git or ch	HTTPS ③ eckout with SVN using the web	URL.
📄 Matla	0	Fixed typo combat coo	de	https://gi	thub.com/Jfortin1/ComBatHarm	C
Pytho	ı	Updated neuroComba	at to 0.2.7			
R		Started adding training	g/test functional	🛃 Open w	ith GitHub Desktop	
Testin	3	Updated neuroComba	at to 0.2.7	Downlo	ad ZIP	

"Code"をクリック後 "Download ZIP"をクリック "ComBatHarmonization-master.zip" をダウンロードして解凍 ここでは D:¥200927tutorial¥ComBatHarmonization-master に解凍

事前配布した5つのデータファイル(CSV形式)も D:¥200927tutorial¥ 20200927tutorial _ComBat_sample にコピー

(D:¥200927tutorialにCombatとサンプルデータが フォルダにある状態)

ComBatHarmonization-master	2020/09/27 0:01	ファイル フォルダー	
200904tutorial_demographics	2020/09/15 5:10	Microsoft Excel CSV	2 KB
APPLY-Tutorial2020.ipynb	2020/09/15 5:10	IPYNB ファイル	51 KB
🖾 Data01	2020/09/15 5:10	Microsoft Excel CSV	23 KB
🛛 Data02	2020/09/15 5:10	Microsoft Excel CSV	23 KB
GML_HMD-ROI-Tutorial2020.ipynb	2020/09/15 5:10	IPYNB ファイル	43 KB
Module_install.ipynb	2020/09/15 9:06	IPYNB ファイル	2 KB
🔁 python_install	2020/09/23 0:13	Adobe Acrobat Docu	1,399 KB
💌 TS01	2020/09/15 5:10	Microsoft Excel CSV	11 KB
TS02	2020/09/15 5:10	Microsoft Excel CSV	11 KB
慉 tutorial200904_ComBatDemo	2020/09/16 4:53	MATLAB Code	4 KB



MATLABを起動します。(Versionによって画面が異なりますので、適宜ご対応ください。



ホームタグから、"パスの設定"アイコンをクリックします。

"サブフォルダーも追加"ボタンをクリックし



📣 サブフォルダーもパスに追加 → · ↑ → PC > Applications (D:) > (5 V 整理 -新しいフォルダー 名前 更新日時 種類 PC 3D オブジェクト 200927tutorial 2020/09/08 13:40 ➡ ダウンロード Program Files (x86) 2020/08/19 16:07 ファイル フォルダー ファイル フォルダー spm12 2020/02/10 15:45 = デスクトップ フォルダー: 200927tutorial フォルダーの選択 キャンセル

"**D:¥200927tutorial**"を選択して、フォルダー選択ボタンをクリックします。

←のようになっていればOKです。 "保存"ボタンを押した後、"閉じる"でウィンドウを閉じます。

準備編3 サンプルデータの説明 (午前のTraveling subjectで使用したデータと同一です)

- Data1.csv: FS風の皮質厚模擬データ(Site 1)
- Data2.csv: 皮質厚模擬データ(Site 2, Data1と同一被験者・縦断データ)
- 200904tutorial_demographics.csv: Data1,2の性別、年齢(共変量データ)
- tutorial200904_ComBatDemo.m: 今回のMATLABデモを実行するソース コード

	А	В	С	D	Е	F
1	SubjID	L_banksst	L_caudala	L_caudalm	L_cuneus_	L_entorhir
2	sub-data01-1	2.81652	2.865	2.647	2.0961	3.138
3	sub-data01-2	3.02428	2.93412	2.82	2.12366	3.61488
4	sub-data01-3	2.7324	3.133	2.67751	2.0196	2.57348
5	sub-data01-4	2.96901	2.86008	2.76114	1.96317	3.97782

	А	В	С	D	
1		sex	Age01	Age02	
2	sub-data01-1	2	10.5	15.9	
3	sub-data01-2	2	10.6	15.6	
4	sub-data01-3	2	10.8	15.8	
5	sub-data01-4	1	10.9	15.6	

Data1,2.csv

200904tutorial_demographics.csv

ComBatはTraveling Subjectsのデータセットを用いずにHarmonizeできるため、 TS01.csvとTS02.csvは使いません



- MATLABを立ち上げます。
- MATLABの"現在のフォルダ"を変更します。



↑のアイコンをクリックすると、フォルダ選択がでるので、 "D:¥200927tutorial¥20200927tutorial_ComBat_sample" を選択して、"フォルダーの選択"ボタンをクリックします。

解凍した配布ファイル"tutorial200904 ComBatDemo.m"をダブルクリックすると、MATLABエディタが立ち上がります。



エディタ内の"データ読み込み"より下の箇所をクリックして色を反転させます。

📣 MATLAB R2018b - academic use

ホーム	プロット	アプリ	エディター	パブリッシュ	表示				
☆ ☆	日本 10 mm →	の検索 ◆ ジ 移動 ◆ ◆ Q 検索 ◆ ナビゲート	挿入 🔜 f コメント % 🕺 インデント 🗊 💀 編集	* 3 ② 3 ③ 1 □ - クボイント ・ フレークボイント ・	▶ 実行 ▼ 次に	 セクションの実行 して、次に進む 実行 	_テ 実行および 時間の計測		
< 🔶 🖬 🍋	🛚 📙 🕨 D: 🕨 20	0927tutorial 🕨 202	00927tutorial_Com	Bat_sample 🕨					
現在のフォルダー) 📝 בדרא 🚺	C:¥Users¥85038¥Dr	ropbox¥東大¥チュー	トリアル¥tutorial200904	_ComBatDemo.m				
□ 名前▲	tutorial20	0904_ComBatDemo	o.m × +						
	1	<pre>xtion tutorial20 %% データ読み; close all; datal=csvread(' data2=csvread(' data1(:,end)=[] data2(:,end)=[] demogTemp=csvre demogTemp(:,1)= demogTemp(:,1)= demogTemp2=demogTem siteID1=repmat(siteID2=repmat() clear('demogTem</pre>	00904_ComBatDemo 込み ./data/Data01.c ./data/Data02.c ; %dataの最終列 ; %dataの最終列 ; %dataの最終列 edemogTemp(:,1)- p(:,[1,2]); %DemogTemp(:,1)- p(:,[1,3]); %DemogTemp(:,1)- p(:,[1,3]); %DemogAll()	() sv',1,1);%1行目、 sv',1,1);%1行目、 は頭蓋内容積で今回 は頭蓋内容積で今回 04tutorial_demogra 1; %Sexが1,2なのて mog Allからdata1(2 mog Allからdata2(2 1));%SiteIndexを付 1)); D変数削除	列目はHeaderと 列目はHeaderと]は使わないのて]は使わないのて]は使わないのて aphics.csv',1. [*] 対応する列を拍 =成	Indexなので読み飛 Indexなので読み飛 削除 削除);‰emographics 1 1)"(はすべての行.1 出 1列目sex 2列目 出 1列目sex 2列目	ぬずす。 ぬずす。 別目:Sex 2列 別目の意味 :Age :Age	目data1のAge, 3歹	IJ∃data2のAge

"実行して次に進む"ボタンをクリックするか、"Ctrl+Shift+Enter"キーを同時押します。



• "ComBat前のグラフ"の箇所の色が反転していることを確認して、"実行して次に進む"ボタンをクリックするか、"Ctrl+Shift+Enter"キーを同時押します。

<mark>%% ComBat前のグラフ</mark> %後で定義したグラフ作成の自作関数を実行 DrawGraph(demog1(:,2),mean(data1,2),demog2(:,2),mean(data2,2),'Raw data');%X軸Age,Y軸 平均皮質厚でplot(data1)



2 siteの散布図と回帰直線が図示されます。 見かけ上、年齢と皮質厚に正の相関がみられるが・・・



• "ComBat"部分の色が反転しているのを確認して確認して" 実行して次に進む"ボタンをクリックするか、"Ctrl+Shift+Enter"キーを同時押します。

%% ComBat

% ComBat入力用(こ2施設のデータを縦に結合 data_All=[data1;data2];%ComBat用(こデータ行列を縦結合 siteID_All=[siteID1,siteID2];%ComBat用(こSiteIDを結合(Vectorなので横結合) demog_All=[demog1;demog2];%ComBat用(こ共変量の行列を縦結合

%ComBatの実行とデータ整理 data_combat=combat(data_AII',siteID_AII,demog_AII,1); %ComBatの入力行列のフォーマットはdata[variable × n], siteID[1 × n],demog[n × covariate]なのでdata_AIIは転置 data_combat=data_combat'; %combatの出力も転置されているので戻す(「'」は転置) data1_combat=data_combat(siteID_AII==1,:);%data1に相当するデータを抽出 data2_combat=data_combat(siteID_AII==2,:);%data2に相当するデータを抽出

ComBatの引数 第一引数:Dataの格納された行列 第二引数:Site or Scanner IDが格納されたベクトル 第三引数:共変量の格納された行列 第四引数:経験的ベイズ推定をパラメトリックでやるかノンパラか?のフラグ(1がパラメトリック)

注意点

Dataと共変量行列の向きが逆なので間違わないように



"ComBat後のグラフ"の色が反転していることを確認して"実行して次に進む"ボタンをクリックするか、"Ctrl+Shift+Enter"
 キーを同時押します

%% ComBat後のグラフ DrawGraph(demog1(:,2),mean(data1_combat,2),demog2(:,2),mean(data2_combat,2),'ComBat (parametric)');



先行研究と同様に負の相関がみられた

実践編 ComBatの共変量を省略するとどうなるか? 例:Ageを抜いてみる

"ComBatの注意点 必要な共変量を入れないとどうなるか?"
 の色が反転していることを確認して、"実行して次に進む"ボタンをクリックするか、"Ctrl+Shift+Enter"キーを同時押します

% ComBat後のグラフ

DrawGraph(demog1(:,2),mean(data1_combat,2),demog2(:,2),mean(data2_combat,2),'ComBat (Age is not considered.)');

皮質厚が年齢に相関しており、かつ年齢に施設間差(Sampling Bias)があったため、年齢をSite Effectとして補正してしまった。

- 予備検証では、ComBat法はTS法よりも補正効果が低いが、 チュートリアルでしめした通り、簡便に実行できる。 (ComBatとしては一行で終わる)
- •TS法は多くのエフォートを必要とするため、また後方視的な研究ではComBat法は有用
- ただし、Sampling Bias等を考慮した共変量のデータがないと 実施できない。