00：00：00

小池：では、1時になりましたので、午後の部を始めたいと思います。よろしくお願いします。

　午後の最初は、午前中の最後に行ったトラベリングサブジェクトのデータを使った補正法ができない場合、Combat法という方法もあって、同じようにと言うわけではないですけれども、プログラムを使って行列データを補正できますので、そちらを解説いただければと思います。

　では、舞草さん、お願いします。

舞草：もしもし。舞草と申します。よろしくお願いいたします。これから画面の共有をしたいと思います。

　全画面で表示しているのですけれども、見えていますでしょうか。

　まず、Combat法を用いた複数プロトコルのデータセットということで、午後のチュートリアルを開始させていただきたいと思います。

　午前の朱さんのご説明にもありましたけれども、お昼も挟みましたし、復習の意味を込めて、もう一度整理させていただきたいと思います。

　MRI等を使った画像解析というのは、疾患特異的な構造や機能変化、灰白質の萎縮やFunctional connectivityの変化などを抽出することができ、これを用いることによって再現性、客観性、定量性に優れる疾患の代理バイオマーカーとなるというのは、皆さんもご存じのとおりだと思います。

　特に近年、疾患の進行動態解明や、有効なバイオマーカーを策定するためには、多数の疾患例を取得する必要がありまして、1施設ではやはり限界があるということで、多施設研究が広く行われているところです。

　これは一例として、アメリカのGAAINというプロジェクトがまとめているのですけれども、アルツハイマー関連の脳画像研究です。53個ぐらいの臨床研究があって、世界的には50万人ぐらいの被験者が登録されているのですけれども、このように多施設での画像研究が世界的な広がりを見せているという現状があります。

　ただし、MRIはかなり昔から、撮像装置の間にMeasurement Bias、つまり撮った装置によって画像が異なることが認められており、補正Harmonizationが必要ということが議論されています。この議論はかなり古くから行われており、補正方法はいろいろあるのですけれども、最初は、午前の田中先生や午後の林先生の発表でもあると思うのですけれども、まずは撮る装置での撮像プロトコルを統一しましょうという試みがあり、その後、前処理的に画像を処理し、機種間差のバイアスを取り除こうという試みが昔から行われています。例えばHuman Connectome Pipelineでも使われています、〓グラッターマップ〓というゆがみの補正、ゆがみを取り除く画像処理や、N3といわれる構造の信号の不均質性を取り除くような補正方法が使われています。それを施した後でも、やはり機種間差が残るということで、近年は後処理的なHarmonizationが必要ということが議論されています。

　機種間差があると何がいけないかというと、まずはやはり疾患の効果量自体が少ないもの、微細な構造変化を示すものに関しては、これがSite Effectより小さい場合には当然疾患を同定することができないので、問題があります。あとは近年多く行われていますが、機械学習を使って疾患を識別するようなモデルを作った場合、例えばスキャナーAで多数のデータを集めて、疾患モデルを識別するような機械学習のモデルを構築して、それをスキャナーBのデータに適用してみると、感度0％、特異度100％のような悲しい結果になるようなことがあります。こういった異種機種間での機械学習モデルが適用できないなど、こういったさまざまな問題が発生します。

　ですので、こういった問題を解決するためにHarmonizationという手法が必要で、本日はCombat法について実践を行っていきたいと思います。

　TSとの補正方法の違いですけれども、補正効果自体はTSのほうが非常に高いのですけれども、やはり1つの問題としてTS自体に撮像にエフォートを非常に有する方法ですので、そういったエフォートが割けないような場合、あるいはretrospectiveなデータセットは今からTSを取ろうにも装置がないということもあろうかと思います。そういった後ろ向きの研究につきましても、Combatであれば現行のデータセットでHarmonizationが必要ということで、すみ分けが可能かと考えています。

00：05：22

　Combatは、もともとはマイクロアレイの機種間差を取り除く方法として2007年に報告されており、それが2017年ごろにFAマップとMRIの構造の皮質厚の補正にも適用可能だという論文があり、それ以降、非常に広まってきた手法です。

　補正のモデルはこのような式で表されています。yが観測された皮質厚やFAマップ、Functional connectivityはiがSiteで、jがとある被験者のインデックスになっていて、vがどの特徴量かを示しますので、例えば皮質なら海馬の皮質厚を、Functional connectivityならどこのネットワークなるのかということを示すものです。

　計測されたデータは、全体の平均にXとβで示すように、〓ホバリエート〓による個人の効果があり、さらにγとβというSiteのEffectが乗った状態であるとモデル化されています。ですので、Site Effectであるγ、βを、Combatの場合は経験的ベイズ推定を行っているのですけれども、これを推定してやることによって、Site Effectを取り除くことが可能であるというモデルです。

　実際に補正後のCombat後の特徴量はこのような式で表され、まず個人の効果とSiteの効果を取り除いた後に、再度、個人の効果を足してあげるというモデルになっています。

　注意点ですけれども、Combat自体、特徴量は例えばVOIの測定値、どの領域の皮質厚かやFunctional connectivityか、解剖学的に標準化されたVoxel値、FAマップや●でというところのワーピングした後のVoxel値には対応可能で、生画像を補正するものではないということです。

　ただ、補正自体は位置ずれがないという仮定の下に行っていますので、標準化の際に位置ずれが起こったようなvoxelデータに関しては、補正効果が十分に得られないという注意点があります。

　Combatはgithubから入手可能で、内容としてはMATLABとPythonとR上で動作する関数となっています。ですので、FreeSurferのようにコマンド一発でできるようなものではなく、MATLAB、Python、Rなどの何らかのプラットフォームにのっとって、プログラムを書いて実装する必要があります。

　本日は、MATLABでデモを行いたいと思います。MATLABが有償ソフトですので、お持ちでない方は1カ月の試用版が発売元から出されますので、そちらをご利用いただいてもいいのですけれども、この間試しにやってみたら、試用した1カ月後に営業の電話が結構かかってきて大変だったので、今回、もしお持ちでない方は聴講のみでも結構です。また、MATLABは使わないという方でしたら、Python、Rでも基本的な流れは一緒ですので、今回は参考にしていただければと思います。

　では、実際に準備編で、事前資料を配布させていただきましたけれども、Combatをgithubのページからダウンロードしていただき、事前配布いたしました、午前の朱さんの発表と同じデータフォルダー上に解凍していただいている状態という前提で進めます。

　次に、ここからは実際にMATLABを動かしながらやっていきたいと思います。SPM等をされている方は慣れていると思うのですけれども、まず最初に、パスの設定が必要です。MATLABを開いていただいた後に、ホーム画面に「パスの設定」というところがありますので、こちらを開いていただいて、「サブフォルダーも追加」というところをクリックすると、フォルダーを選択する画面になります。

00：10：06

　チュートリアルのデータとCombatをダウンロードしたフォルダーを選択していただくと、選択したものがここに表示されますので、保存をして閉じていただくと設定が完了となります。

　今回はデモなのですけれども、事前配布した 「tutorial200904\_CombatDemo」というソースコードがありますので、こちらをダブルクリックしていただきますと、MATLAB上にデータが開き、ソースの中身を見ることができます。

　次に、現在のカレントディレクトリを変更しなければいけないのですけれども、上部にフォルダーが出ています。左にフォルダー選択のボタンがあるので、こちらをクリックしていただいて、ソースコードのある事前資料のデータを配布したフォルダーを選択していただきます。そうすると、左に現在のファイルが出てきますので、ここにtutorial、CombatDemo.mファイルがあることをご確認ください。

　まずデータの説明です。午前中の朱さんのチュートリアルと一緒なのですけれども、配布するデータにdata1、2のcsvというものがあります。これはdata1がSite1で、2がSite2で撮像した脳画像、MRIを基にFreeSurferの解析をかけたという体の模擬データになっています。Site1、2が縦断的な撮像という仮定ですので、tutorial\_demographicにはこのようなsexとAgeのデータがあり、性別と、データ1に対応するAge、データ2に対応するAgeが別の列として保存されています。ですので、このdata1、2とdemographicのデータを読み込んで補正、Harmonizationをかけるという処理をこれから行っていきたいと思います。

　デモですけれども、まずMATLAB上で「データ読み込み」というところをクリックすると色が変わるかと思います。その状態で上のエディターを選んでいただいて、セルを「実行して次に進む」というボタンがあるので、このボタンをクリックしていただくと、データの読み込み部分を実行することが可能になっています。そうしますと、左にdata1、2という変数が出るのですけれども、やっている内容としましては、MATLABのcsvreadという関数を使ってデータ1を読み込んで、データ1の変数に突っ込み、データ2のcsvをデータ2に突っ込みます。データ1、2に関しては最終列が頭蓋内容積になっていますので、今回は使わないので削除するという処理をしています。

　同様にdemographicのデータを読み込んで、demographic添付というところに保存した後に、今、性別が1、2の変数で定義されているのですけれども、これはバイナリーデータとして扱うので、1を引くという処理で0、1に変換しています。

　あとは、2列目がAgeでdata1に対応する年齢なので、demog1に保存し、demog2はデータ2列目です。今回のデータに関してはSiteのインデックスの番号がないので、プログラム上で作っているのですけれども、csv上で1番目の施設が1、2番目の施設が2と各行を作っていただいても問題ありません。

　では、次に、ここまで読み込みますと、次の「Combat前のグラフ」というところをチェック、色を変えていただいて、「実行して次に進む」というボタンを押していただきます。そうしますと、少し時間がかかるのですけれども、グラフが表示され、これが補正前の年齢と平均皮質厚のプロットになります。午前の朱さんのチュートリアルと同じですけれども、data1、Site1のデータよりもSite2のほうが有意に皮質厚が増加しています。つまり年齢に依存して皮質厚が増加しているように見えるのですけれども、これは先行研究等を見ましても間違いで、本来は思春期を境に皮質厚は下がらなければいけませんが、皮質厚が上がっています。つまり、Site Effectがあるために見かけ上、皮質厚が上がっていることを示すグラフになります。

　次に、実際にCombatの補正処理を行っていきたいと思います。

00：15：00

　実際の処理は1行、Combatの行だけで済み、データを全て入力するだけです。こうして「実行して次に進む」をクリックしていただきますと、すぐにここのCombatの処理があり、ログのようなものが出るのですけれども、まずFound 2batches、つまり2Site分のデータを読み込んだというものがあり、あとはStandardizing Date、featuresやAdjusting the Dateが終わったというログが発生しています。

　Combatの使い方ですが、まず最初にデータ自体を今、data1と2に分けていたのですけれども、1次元のベクトルに合わせなければいけないので、これは縦に結合して、data\_Allというのを作っています。どういうものかというと、こういった形で、ヘッダーがないので見にくいかもしれないですけれども、途中までがSite1のデータで、後半部分がSite2のデータに全体になっています。

　同様に、Combat AllとSiteのIDも横に結合していまして、demogも縦に結合しています。Combatの入力ですけれども、データとしては、各行が被験者のデータになっていて、右に解剖学的領域が各列になっています。Combatの場合、入力が逆になっていて、各列が被験者で、各行が valuableにならなければいけないので、ここでダッシュを付けることによってMATLABで転置をしています。

　あとは、Site IDは1行N列の1次元配列でなければいけないので、横に結合しています。逆にdemographicはデータの並びとは逆で、最初のように各行が被験者で、各列がvaluableにならなければいけないので、ここの入力データの向きが違うというのが、Combatのプログラムの癖があるところで、そこは少し注意点になります。

　これを実行しますと、「data\_combat」というところにCombatで補正されたデータが入り、これも入力と同じように各列が被験者になっていますので、行列の配置を戻すために再度転置をします。全てのデータが入っていますので、Site IDでdata1とdata2に分離するという処理はしています。

　これを同じようにグラフ化しますと、このように年齢に相関して皮質厚が下がるという、先行研究と一致した結果になるという結果が得られるかと思います。午前の朱さんのトラベリングサブジェクトでも同じような結果が得られていますので、トラベリングサブジェクトと同様の結果がCombat、トラベリングサブジェクトを使わないデータでも実行可能ということが示せたかと思います。

　Combatの一番の注意点ですけれども、今、年齢と性別を共変量として入れているのですけれども、取りあえずやってみようということで、例えば年齢を除いてみた場合、どういうことになるかということを次にお示ししたいと思います。

　ここの入力で無理やりdemographicの性別だけを入れて、年齢は入れないようなCombatの処理を行っています。後半の処理は全く一緒で、年齢を入れないでCombatの補正をかけてグラフを表示するというプログラムになっています。

　これを同じようにセルを選択して実行してみますと、このようにSite1とSite2の皮質厚と年齢の回帰が全くフラットになってしまっています。これはどういうことかというと、今回、データとしてはSite1とSite2で有意に年齢が異なるデータセットになっていますので、しかもそれを共変量として入れていないので、年齢の差による皮質厚の変化を、Combat自体がSiteのeffectであるとみなして補正してしまったために、年齢の分布がフラットになってしまうという結果になっています。

00：19：55

　デモとしては以上になりますが、総括、まとめますと、今回、われわれの送った予備検証、あとは山下先生の先行研究でも、Combat法はTS法よりも補正効果自体は低いのですけれども、チュートリアルで示しましたとおり、簡便に実行できます。多少プログラムの知識が必要ですけれども、やり方としては、データを読み込んで、Combatの関数にデータを全て入力することによって補正が得られるという、Combat自体の処理としては1行で終わるような簡便な方法になっています。また、TS法は最初に言いましたように、多くのエフォートを必要としますので、こういう後方視的な研究やエフォートが十分確保できないような研究では、Combat法が有効ではないかと思っています。

　ただし、非常に注意しなければいけない点は、Sampling Bias、今回のように年齢が違うなど、そういったものを考慮した共変量が出ないと、実施はできるのですけれども、正当な結果が得られないというような問題点があります。Combatはその点だけを注意しなければいけないということになっています。

　では、以上です。

小池：ありがとうございました。ご質問がある方はいらっしゃいますか。違う質問が飛んでいます。それは放置しておきましょう（笑）。別に全然構わないのですけれども、トラベリングサブジェクトではなく、Combatのほうで何かご質問がある人はいますか。

　こちらも、データが欠損値に対応できないのではないかと思うのですが、どうでしたか。

舞草：欠損は……。

小池：欠損値をどこに対応させるかによると思うのですけれども、もともと欠損のデータを埋めることは多分できないと思うのです。そういう欠損値の対応はできないと思います。Combatに欠損値がある場合に、Combatがそのまま走るかというと、走るのでしょうか。多分valuableごとに計算しているので、走るのだと思うのです。

舞草：そうです。そこは多分nullになるか、要はSite Effectを推定する際に、そこの部分は除かれる処理まであったかどうか分からないですけれども、原理的には推定部分で除かれるだけです。

小池：そうです。次は、Combatの解析Pipelineの違いに対応できますか。

舞草：実は今回お示ししたサンプルデータ自体が、施設間差もあり、解析方法も少し違うやり方で取っているので、差が非常に大きくなっているのですけれども、一応解析Pipelineの違いについても、線形な変化であれば補正できるということにはなるかと思います。ただ、Siteの違いプラス解析Pipelineの違いになると、相互作用などが出てくると思うので、その辺までは少し難しいかと思います。単純に線形な効果であれば補正できると思います。

　そうですね。SPMのGLMでSite Effect regulationを入れてはいけないでしょうかということです。以前、〓ネモト〓先生が同じようなことをされていて、SPMの最後の〓statistics〓のところで、Site Effectを入れることによって補正するということをされていたかと思うのですけれども、原理的には可能です。

　ただ、今回、GLMよりもCombatのほうがSite Effectの効果が1つ多いのです。GLMの場合はSite Effectが1施設に対して1になると思うのですけれども、Combatの場合は加法的な項と、乗数的な項の2つで表現しているので、よりCombatのほうが推定モデルが複雑なので、GLMに比べて推定精度が高いという報告が、先ほど最初に示した〓フォルティー〓の論文に出ていますので、そちらのほうがいいかと思います。

　Niftiはやり方としては同じで、先ほどcsvを読み込んだのですけれども、ここの部分を、MATLABやPythonのNifti読み込みライブラリーでIMG部分を読み込んで、補正すれば可能ですけれども、生データは駄目で、標準化されたデータ、つまりvoxelのインデックスが同じであれば、同じところを指しているNiftiデータでないと駄目です。

00：25：08

　実際〓VM〓で昔やったのですけれども、基本的なやり方は一緒で、IMGを読み込んでCombatにかけて、出力するときにIMGで出力してあげるという方法で可能です。

小池：他はありますか。GLMのSite Regulationを入れて、10施設ぐらいになってくると多分ずらーっと並んで、非常に自由度が下がっていくような気がします。nが十分であればあまり気にしなくていいような気がしますけれども。

　あとは相関や〓ポストボック〓に見たいときに、いちいちそれをやるのかという話になるような気がします。そういう大変さがあります。

舞草：●のSite Effectなのです。そうですね。Combatの場合や、別建てGMLのプログラムを組んだ場合は、voxelごとに違うものだと、その分をやるので確かにそのとおりです。

小池：そうですね。はい。他はありますか。大丈夫〓ですか〓。

00：26：16