**人口密度と産業別就業割合の相関**

Ryo Takamatsu(229x112x)

# **Introduction**

戦後から現在において都市圏では、特に三大都市圏において人口の流入が続き、北海道、東北などからは人口転出が起こっている。それぞれで人口集中・過疎化がそれぞれ問題となっており、人口密集地では騒音・環境問題、過疎地域では限界集落の問題などがある。本リサーチペーパーでは各都道府県における産業別就業割合と人口密度の相関

# **Method**

3つの潜在空間をプロットする。1つ目は5人分の訓練データを用いて学習を行ったモデルの潜在空間である。2つ目は、同5人のテストデータを入力した場合の潜在空間であり、これは実際の認証において正規ユーザーが認証を試みた場合に対応する。3つ目は学習データにない6人目のデータを入力した場合の潜在空間であり、攻撃者が認証を試みた場合に対応している。

また、プロットには下記の機能を搭載している：

* ドラッグによる回転（3画面同期/独立）
* マウスオーバーによる強調表示（注目したい被験者以外を消去）
* 訓練データを入力した場合の潜在空間の、各人物の重心座標を中心とした球の表示  
  （球は基準となる座標からの距離を表し、本人か未知の人物かを分離できそうか確認できる）

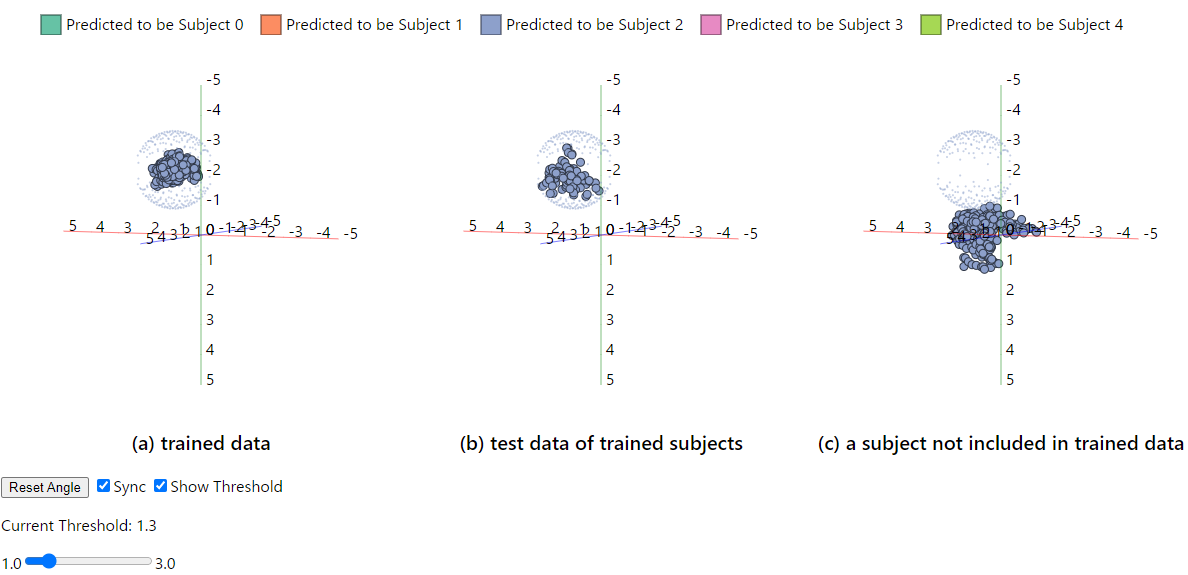
各機能の例をFig 2に示す。

|  |  |
| --- | --- |
| C:\Users\manabe ayumu\AppData\Local\Microsoft\Windows\INetCache\Content.Word\default.png   1. default | C:\Users\manabe ayumu\AppData\Local\Microsoft\Windows\INetCache\Content.Word\threshold.png   1. threshold display |
| C:\Users\manabe ayumu\AppData\Local\Microsoft\Windows\INetCache\Content.Word\hover.png   1. hovered to refine | C:\Users\manabe ayumu\AppData\Local\Microsoft\Windows\INetCache\Content.Word\angle.png   1. change angles and thresholds |

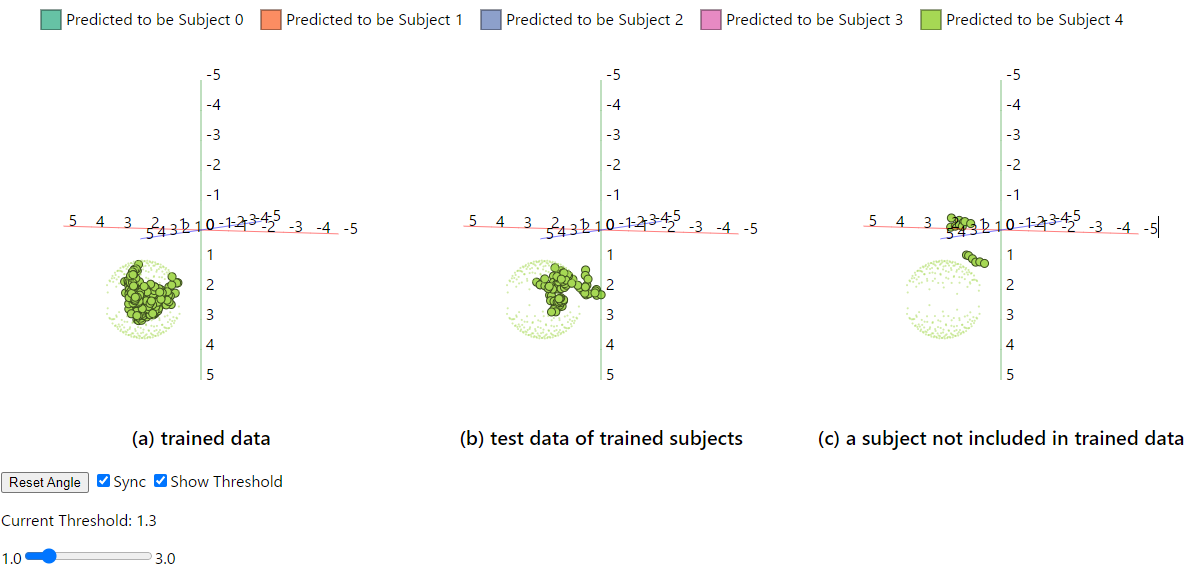
**Fig 2 Examples of interactive operations**

# **Result**

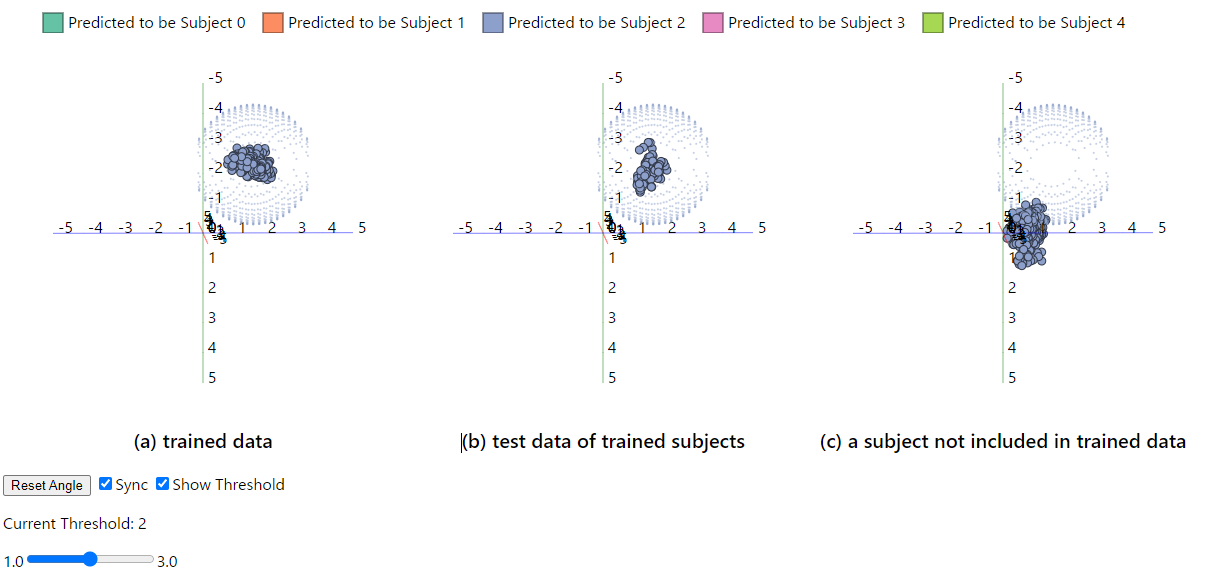
未知の人物は、すべて被験者2と被験者4として分類器によって分類された。しきい値（図中の球の半径に対応）を1.3としたときと2.0としたときの、被験者２と被験者４に分類されたデータのプロットをFig 3 - Fig 6に示す。



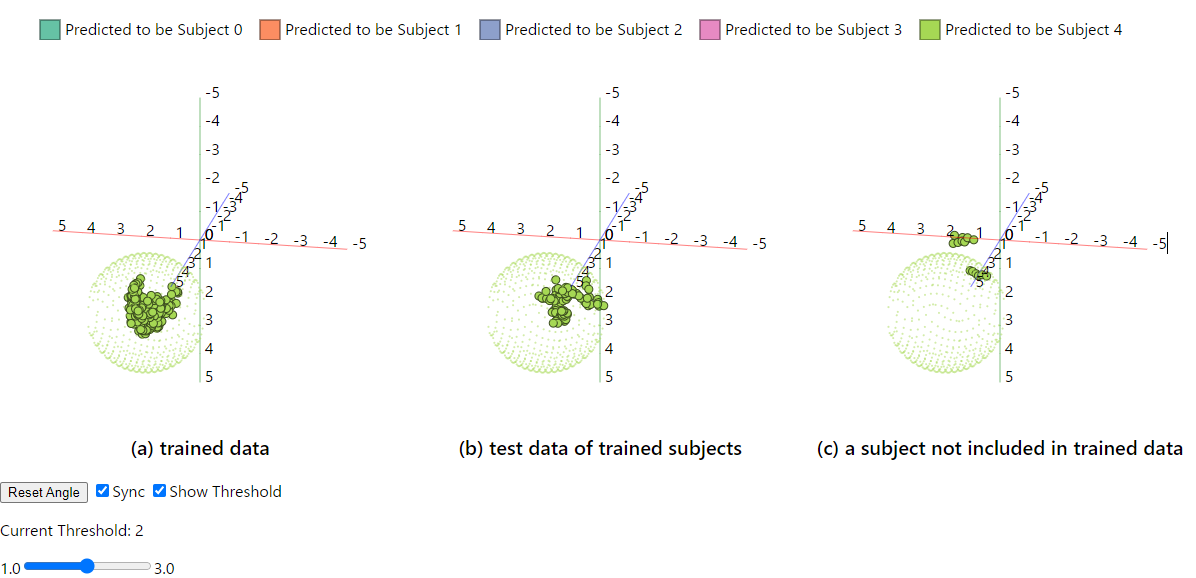
**Fig 3 Data predicted to be subject 2 (threshold=1.3)**



**Fig 4 Data predicted to be subject 4 (threshold=1.3)**



**Fig 5 Data predicted to be subject 2 (threshold=2.0)**



**Fig 6 Data predicted to be subject 4 (threshold=2.0)**

# **Discussion**

Fig 3 – Fig. 6より、(a)訓練データ、(b)テストデータ、(c)未知の人物のデータの順に、潜在変数が原点付近に偏って分布する傾向が確認できた。これは、訓練データとの差異が大きいデータの潜在変数ほど原点付近に分布するという性質があることを示している。つまり、訓練データの潜在変数の重心からの距離によって、本人か否かを判別することができる可能性は存在する。

次に、100%分離できるようなしきい値が存在するかどうかを確認する。Fig 3より、基準座標から1.3の距離をしきい値とすると、実際に被験者2であるデータ(図中(a)と(b))は球の内部に収まり、他人である図中(c)のデータは球の外部にプロットされている。よって、しきい値を1.3と設定すれば被験者2については本人か他人化を分類することが可能になる。しかしFig 4 (b)を見ると、本人のデータであるにも関わらず球の外部にプロットされているものが複数存在する。

そこでしきい値を2に変更した場合のプロットがFig 5とFig 6である。どちらの被験者についても、本人のデータは球の内部に分布しているが、他人のデータも一部球の内部に分布してしまう。つまり、本実験において完全に本人か否かを分離できるようなしきい値は存在しない。

# **Conclusion**

本リサーチペーパーでは、私が研究で用いているCVAEの潜在空間を生体認証に応用する方法の検討を行った。結果、潜在変数の分布によってデータが本人か否かを一定の割合で分離することは可能だが、100%の精度では難しいということが確認できた。

今後の研究では、複数のしきい値で本人を他人だと判断してしまう「本人拒否率」と他人を本人だと判断してしまう「他人受入率」を算出し、これらのトレードオフの関係にある指標のバランスが最適になるしきい値を検討したい。また、なぜ訓練データとの差異の大小によって潜在変数の分布に影響が出るのかを数式的に理解する必要もありそうだ。

# **Refference**

* d3-3d(<https://github.com/Niekes/d3-3d>)
* 3D scatter plot(<https://bl.ocks.org/Niekes/1c15016ae5b5f11508f92852057136b5>)
* Interactive grouped scatterplot in d3.js(マウスオーバーで強調表示)  
  (<https://www.d3-graph-gallery.com/graph/scatter_grouped_highlight.html>)
* Categorical legend: square  
  (<https://www.d3-graph-gallery.com/graph/custom_legend.html#cat3>)