Visualization of 3D latent space of My CVAE (Conditional VAE)

Ayumu Manabe (215x120x)

1. Introduction

私は現在、生体認証の分野における CVAE(Conditional VAE)という深層生成モデルの応用を研究している。CVAE の中にも入力とラベル、潜在変数の依存関係によって複数のモデル構造が考えられるが、研究で用いているモデルは Fig 1 のような構造をとっている。

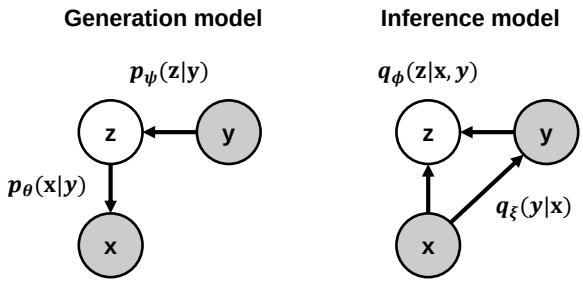


Fig 1　Graphical models of the proposal CVAE

Fig 1 のそれぞれの矢印はニューラルネットワークに対応しており、例えば$q_\xi(y|\mathrm{x})$は入力データxからラベルy(ここでのラベルはどの人物かを識別するもの)を推論するはたらきを持っている。そこで当初、学習済みのこの分類器を用いて入力データがどの人物のものかを予測することで、生体認証を行うことを考えた。実際にこの分類器によって、96%の精度で入力データから被験者を正しく分類することができた。

しかし認証と分類は似て非なるものである。本研究では N 人のデータが予め与えられた状態で、個人が特定されていないデータが入力された際に、その画像が N 人のうちどれであるか、またはどれでもないかを判定する 1:N 認証を行う必要がある。つまり、事前に学習した人物と、学習していない未知の人物を区別しなくてはならない。

そこで本リサーチペーパーでは、Fig 1 中の$q_\phi(\mathrm{z}|\mathrm{x},y)$によって得られる潜在変数zの分布を基に、入力データが学習済みの人物のものであるか否かを判別する方法を考察する。具体的には、学習時の潜在空間の分布と近い位置にプロットされたデータは本人、遠い位置にプロットされたデータは他人、という風に分類できないかどうかを検討する。

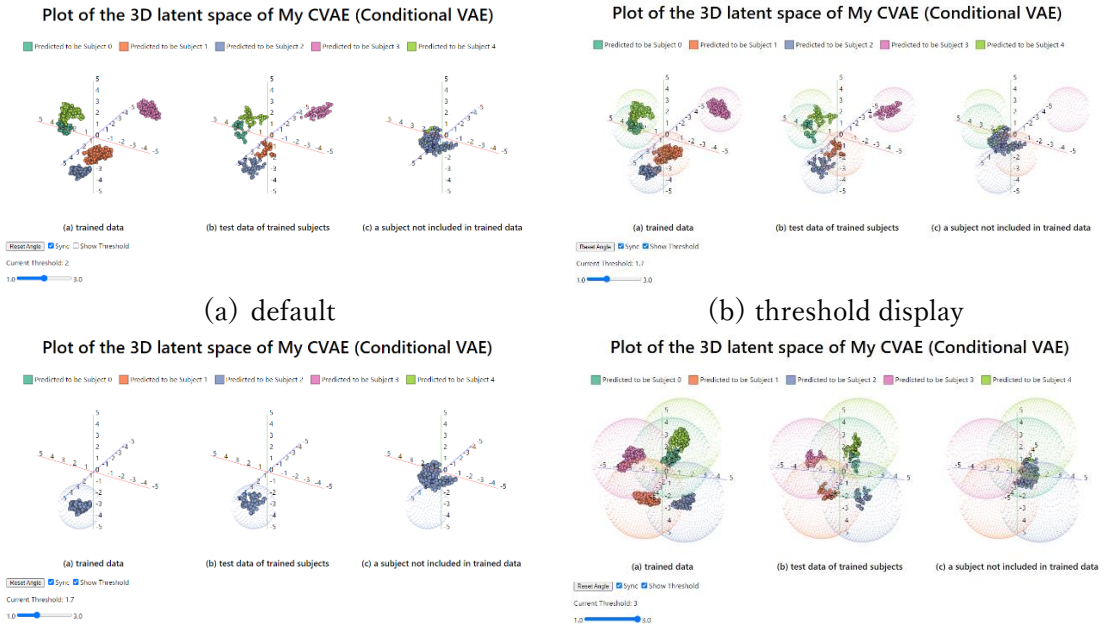
2. Method

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。



(a) default　(b) threshold display

(c) hovered to refine　(d) change angles and thresholds

Fig 2　Examples of interactive operations

3. 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。

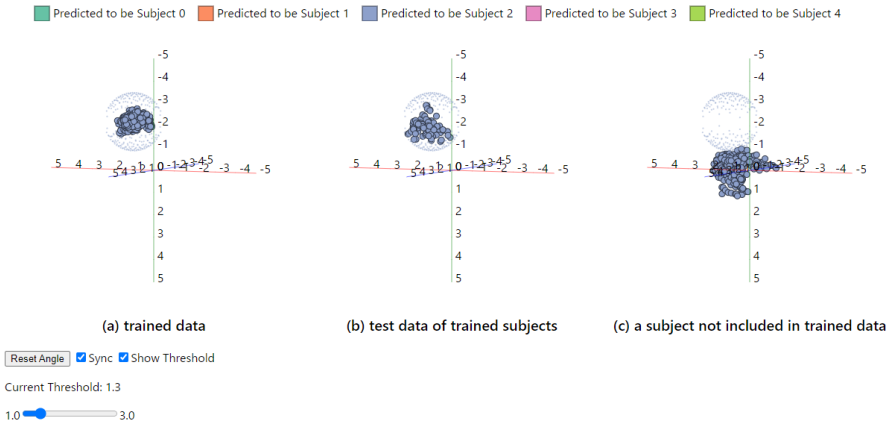


Fig 3　Data predicted to be subject 2 (threshold=1.3)

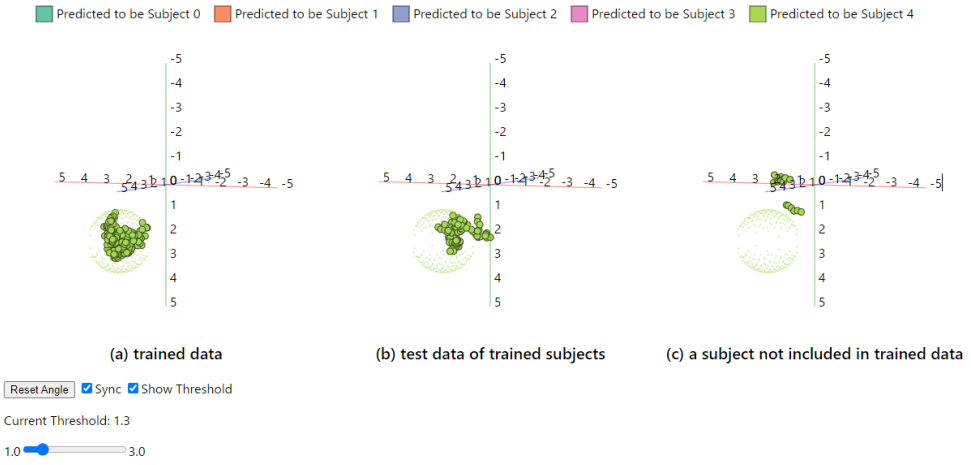


Fig 4　Data predicted to be subject 4 (threshold=1.3)

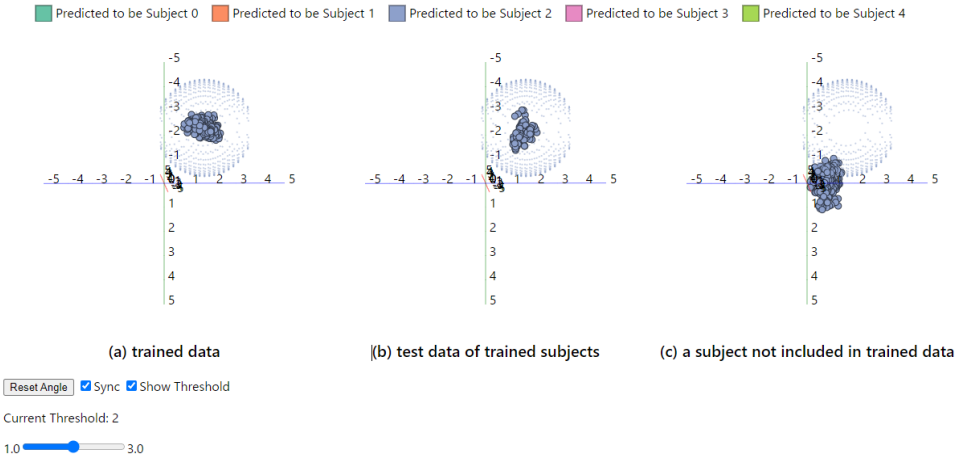


Fig 5　Data predicted to be subject 2 (threshold=2.0)

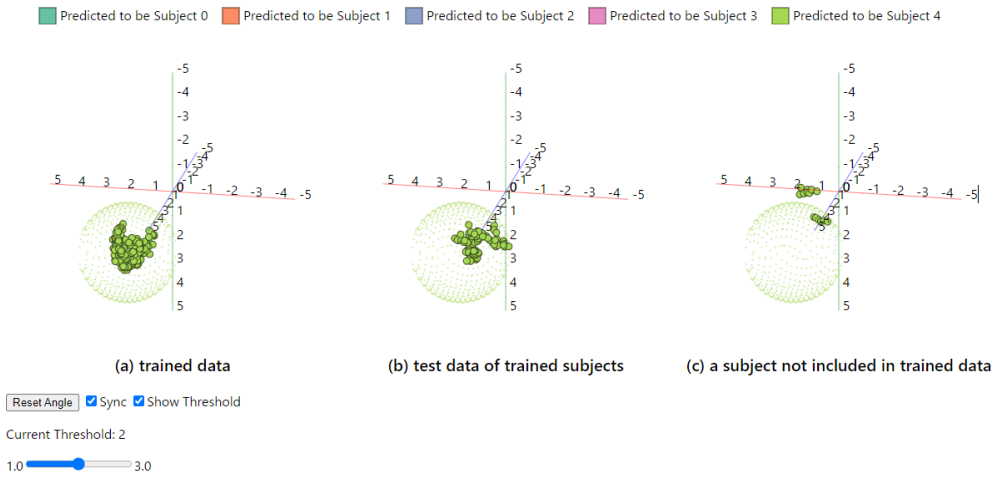


Fig 6　Data predicted to be subject 4 (threshold=2.0)

4. 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。

5. 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。

6. 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)