PWSCUP2021

チーム14　大崎KG

匿名化フェーズ

加工内容

- 第一匿名化:
  1. ランダムに45%削除
  2. 特異なデータの削除
  3. Umark基準を満たすように削除
- 第二匿名化:
  1. Iloss基準を満たすようにデータのスワッピング

方針

- 危険度の要素の内、recall・precはランダムな予測でも0.5程度当たるため、topkを抑えたい

$$危険度 = \underline{recall \times prec} \times topk$$

ランダム予測の期待値:0.5

- topkの予測(攻撃)手法として、行ごとのIloss,L2距離の近傍k個を予測結果とする手法が考えられる

→L1距離・L2距離の予測が外れるように、C内の別データに置き換える

データスワッピング

Ilossが4以内のデータと入れ替える

※第一匿名化でUmarkを満たす様に削除しているため、データを入れ替えてもUmarkには無影響

※ペアを作れなかったデータについては元のままとする
(C:2278行中、34行)

C_T

id	0	1	2	3	…	11
1	Male	57.0	White	9th	…	0
2	Male	60.0	White	College	…	0
…	…	…	…		…	…

スワッピング

改良点

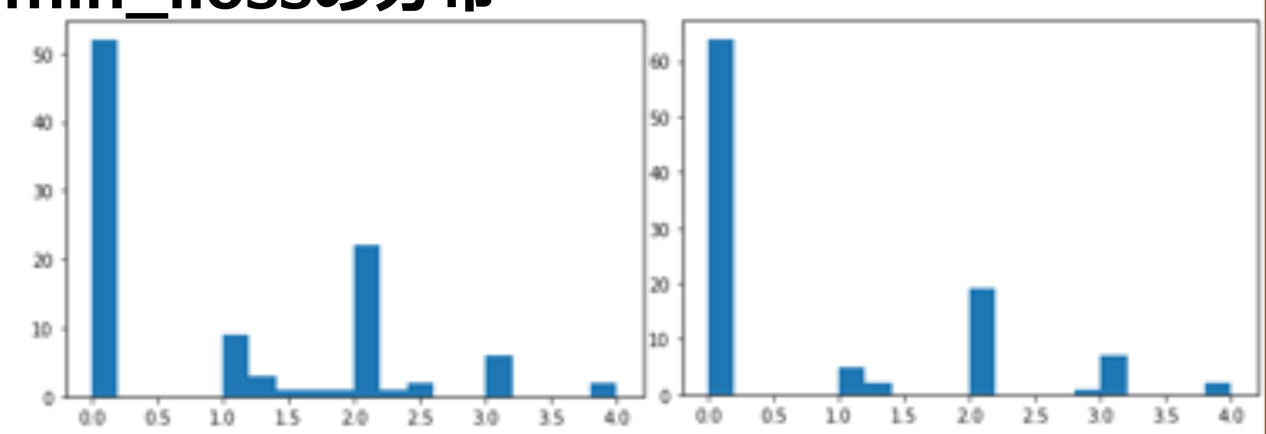
- スワッピングするペア数の最大化(組み合わせ)
- スワッピングするデータとのL1距離・L2距離ができるだけ大きくなる(Iloss<=4の条件下)ようなデータの選択
  (ペア数の最大化と距離を考慮したペア作成を同時に考慮できず断念)
- 第二匿名化時にランダムノイズを加える処理
  (Umark基準を満たせず断念)
- スワッピングではなく、k-匿名化によって同じデータを持つデータに加工する
  (Umark基準を満たせず断念)

攻撃フェーズ

攻撃手法

- D内のデータと完全一致するデータがC_T内にあれば、D内データと判断
  (完全一致するC_T内データ数:match_len)
- D内のデータとのL1距離の最小が4より大きいデータは、D外データと判断
- 以下の手順で予測
  1. C_Tの各データに対して、Dの各データとのIlossの最小値(min_iloss)と、最小となるDのindex(min_iloss_index)を計算
  2. C_Tの各データに対して、C_T内でIlossが4以下のデータの中からIlossが最大のデータを3つ選ぶ
  3. min_ilossの小さいmax(50,match_len)に対して、2で選んだ3つのデータのmin_iloss_indexを推定Eとする

min_ilossの分布



チーム10　　チーム8

min_iloss,min_iloss_indexの計算

C_T

id	0	1	…	11
1	Male	40.0	…	0
…	…		…	…

D

id	0	1	…	11	Iloss
11	Male	42.0	…	0	
12	Female	25.0	…	1	15.0
…	…		…	…	⋮
				min_iloss	2.0

id	0	1	…	11	min_iloss	min_iloss_index
1	Male	40.0	…	0	2.0	11

方針

- 模擬戦攻撃結果・本戦各チームC_Tデータ観察結果から、スワッピングに似たことをしているチームがあると仮定
- それらデータに対して、スワッピング元を特定するような手法を検討
  - Iloss・L2距離が小さいデータ同士でのスワッピングは考えにくいため、出来るだけ距離の遠いデータと入れ替えるのではないかという仮説の元、操作を決定

改良点

- チームごとへの攻撃手法の変更
  (スワッピングをしていなさそうなデータの攻撃手法(シンプルなL1距離・L2距離による予測)以外の方法が見つからず断念)