PWS Cup 2019 競技ルール Ver1.4

PWS Cup 実行委員会

2019 年 9 月 20 日

1　はじめに

本ルールの記号や有用性・安全性指標の詳細，およびデータセットは，次の文献にて与えられる．

- CSS2019 に投稿された PWS Cup 2019 のルール論文 [1].
- 疑似人流データ [2] を基に新たに作成した PWSCup2019 用人工データ [3].

2　コンテストルール

2.1　概要

1. 各チームは，ルール論文 [1] の第 2.2 節に記載のコンテストの流れに沿って，データ（匿名加工フェーズでは加工トレース，ID 識別・トレース推定フェーズでは仮名表の推定値，元トレースの推定値）を提出すること．
2. 提出するデータのフォーマット，ファイル名については，それぞれ本資料の第 2.4 節，第 2.5 節に記載の事項を遵守すること．
3. 本資料の第 2.6 節に記載の禁止事項を実施しないこと．

2.2　コンテストの流れ

ルール論文 [1] の第 2.2 節を参照.

2.3　賞

1. 本コンテストでは，以下の賞を設ける．尚，ID 識別安全性とトレース推定安全性は，それぞれ予備選の値と本戦の値を 1:9 の割合で合計した上で，受賞チームを選定する．
2. **総合優勝・総合 2 位・総合 3 位**：ID 識別対策とトレース推定対策の両方で優れた成績を修めたチームに賞を与える．具体的には，ID 識別対策用の公開加工トレース $A'^{(i,1)}$ に対する ID 識別安全性 $s_{I,min}^{(i,1)}$ と，トレース推定対策用の公開加工トレース $A'^{(i,2)}$ に対するトレース推定安全性 $s_{T,min}^{(i,2)}$ の総和 $s_{I,min}^{(i,1)} + s_{T,min}^{(i,2)}$ が最も大きい上位 3 チームに賞を与える．尚，無効な加工トレースや未提出の加工トレースに対応する安全性は 0 と見なす．

3. **匿名加工賞（ID 識別対策部門）**：ID 識別対策用の公開加工トレース $A'^{(i,1)}$ に対する ID 識別安全性 $s_{I,min}^{(i,1)}$ が最も大きい 1 チームに賞を与える．
4. **匿名加工賞（トレース推定対策部門）**：トレース推定対策用の公開加工トレース $A'^{(i,2)}$ に対するトレース推定安全性 $s_{T,min}^{(i,2)}$ が最も大きい 1 チームに賞を与える．
5. **リスク評価賞（ID 識別部門）**：ID 識別安全性を最も下げた 1 チームに賞を与える．但し，総合優勝，および匿名加工賞（ID 識別対策部門）を受賞したチームは除外する．
   具体的には，他のチームの公開加工トレース $A'^{(i,j)}$ に対する ID 識別安全性 $s_I^{(i,j)}$ と，**実行委員が別途用意した仮名化トレース（2.3.7 節で詳述）**に対する ID 識別安全性 s_I^* の総和が最も小さいチームに賞を与える．ここで，「トレース推定対策用」の公開加工トレースも含めて $s_I^{(i,j)}$ の総和をとることに注意する．即ち，$s_I^* + \sum_{i=1}^{z}\sum_{j=1}^{2} s_I^{(i,j)}$ が最も小さいチームに賞を与える．但し，無効な加工トレースや，自分のチームの加工トレースの ID 識別安全性は 0 と見なす（即ち，総和をとる対象から除外する）．従って，自チームへの ID 識別の結果を提出しても，評価されない．また，他チームの有効な公開加工トレースの中で，ID 識別を試みていないものに対しては，ID 識別安全性を 1 と見なす．
   ［参考］ID 識別用・トレース推定用含めて，なるべく多くの他チームの公開加工トレースに対して ID 識別を試みた方が本賞をとりやすい．
6. **リスク評価賞（トレース推定部門）**：トレース推定安全性を最も下げた 1 チームに賞を与える．但し，総合優勝，および匿名加工賞（トレース推定対策部門）を受賞したチームは除外する．
   具体的には，他のチームの公開加工トレース $A'^{(i,j)}$ に対するトレース推定安全性 $s_T^{(i,j)}$ と，**実行委員が別途用意した仮名化トレース（2.3.7 節で詳述）**に対するトレース推定安全性 s_T^* の総和が最も小さいチームに賞を与える．ここで，「ID 識別対策用」の公開加工トレースも含めて $s_T^{(i,j)}$ の総和をとることに注意する．即ち，$s_T^* + \sum_{i=1}^{z}\sum_{j=1}^{2} s_T^{(i,j)}$ が最も小さいチームに賞を与える．但し，無効な加工トレースや，自分のチームの加工トレースのトレース推定安全性は 0 と見なす（即ち，総和をとる対象から除外する）．従って，自チームへのトレース推定の結果を提出しても，評価されない．また，他チームの有効な公開加工トレースの中で，トレース推定を試みていないものに対しては，トレース推定安全性を 1 と見なす．
   ［参考］ID 識別用・トレース推定用含めて，なるべく多くの他チームの公開加工トレースに対してトレース推定を試みた方が本賞をとりやすい．
7. **実行委員の仮名化トレース**：実行委員が別途用意する仮名化トレースとは，（ユーザ集合がどのチームのトレースとも異なる）元トレースに対して，位置情報の加工をせずに仮名化のみを施した公開加工トレースのことである．ID 識別・トレース推定フェーズにおいて，実行委員はこの仮名化トレースと，対応する参照トレースを全チームに配布し，各チームは仮名化トレースに対して ID 識別・トレース推定を試みる．上述のとおり，リスク評価賞では，仮名化トレースに対する ID 識別安全性 s_I^* とトレース推定安全性 s_T^* も考慮してチームを選定する（尚，仮名化トレースに対して ID 識別，トレース推定をしなかった場合，それぞれ $s_I^* = 1$，$s_T^* = 1$ と見なす）．
8. **プレゼンテーション賞**：各チーム P_i は，どのような匿名加工，ID 識別，トレース推定を行ったかを明確にするため，CSS2019 の会場において，匿名加工，ID 識別，トレース推定のアルゴリズムの概要をオーラルで発表し，詳細をポスター形式で発表する．その後，審判 Q が指名した数名の審査委員による投票により，最も優れた発表をしたと判定された 1 チームに賞を与える．また，発表したオーラル・ポスター資料については，後日審判 Q に提出する（但し，諸事情により当日発表できない，或いは資料を提出できないチームについては，応相談とする）．

2.4 データフォーマット

1. **ユーザ ID**：本コンテストでは，チーム番号 i（$1 \leq i \leq z$）およびデータセット番号 j（$1 \leq j \leq 2$）ごとに異なる，2000 名のユーザ集合の参照トレース $R^{(i,j)}$ と元トレース $O^{(i,j)}$ を用いる．各元トレース $O^{(i,j)}$ において，ユーザ ID は 1 から 2000 までの自然数である．
2. **仮名 ID**：各公開加工トレース $A'^{(i,j)}$ において，仮名 ID は 2001 から 4000 までの自然数である．
3. **領域 ID**：東京中心部（緯度：35.65〜35.75，経度：139.68〜139.8）に対して，均等に $32 \times 32 = 1024$ 個の領域 $x_1, \cdots, x_{1024}$ に分割し，領域 ID を割り当てる．領域 ID は 1 から 1024 までの自然数である．
4. **時刻**：時刻については，8 時から 17 時 59 分までを 30 分おきに区切って離散化する．予備選では，1 日目と 2 日目のトレースを参照トレースに，3 日目と 4 日目のトレースを元トレースとして用いる．即ち，各トレースの長さは $t = 40$ である（時刻 1 は 1 日目の 8 時，時刻 40 は 2 日目の 17 時 30 分，時刻 41 は 3 日目の 8 時，時刻 80 は 4 日目の 17 時 30 分）．但し，本戦では参照トレースと元トレースの日数を（2 日分から）変更する可能性がある．
5. **参照トレース $R^{(i,j)}$・元トレース $O^{(i,j)}$**：ユーザ ID，時刻，領域 ID が書かれた csv ファイルである．1 行目に “user_id,time_id,reg_id” というヘッダを記し，2 行目以降には（ユーザ ID，時刻）の小さい順にユーザ ID，時刻，領域 ID を記す．例えば，図 1 では，以下のようにする．

```
user_id,time_id,reg_id
1,5,1
1,6,3
1,7,2
1,8,1
...
3,8,4
```

6. **加工トレース $A^{(i,j)}$**：（ユーザ ID，時刻）の小さい順に加工後の領域 ID が書かれた csv ファイルである（ユーザ ID，時刻については記載しないことに注意）．1 行目に “reg_id” というヘッダを記し，2 行目以降には（ユーザ ID，時刻）の小さい順に加工後の領域 ID を記す．加工の方法としては，ノイズ付与，一般化，削除がある．一般化，削除はそれぞれ領域 ID のリスト（空白区切り），アスタリスク（*）で表す．例えば，図 1 では，

```
reg_id
2
3
2 4 5
*
...
1 2 3
```

とする．ここではユーザ ID 1 の時刻 5 の領域に対して $x_1 \to x_2$ とノイズを付与し，時刻 7 の領域を $x_2 \to \{x_2, x_4, x_5\}$ と一般化し，時刻 8 の領域を x_1 から削除している．

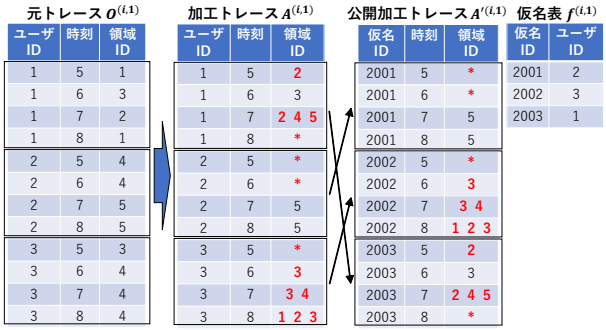

図 1　匿名加工の例．一般化，削除はそれぞれ領域 ID のリスト（空白区切り），アスタリスク（*）で表している．例えば，「2 4 5」は $\{x_2, x_4, x_5\}$ を意味する．

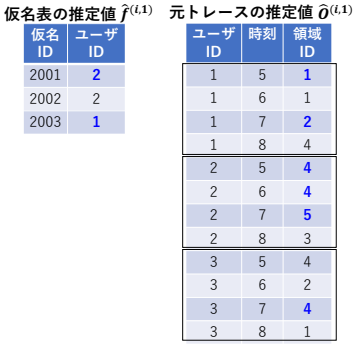

図 2　図 1 の公開加工トレース $A'^{(i,1)}$ に対する ID 識別・トレース推定結果の例（青字：元データと完全に一致しているユーザ ID ／領域 ID）．

7. **公開加工トレース** $A'^{(i,j)}$：仮名 ID，時刻，加工後の領域 ID が書かれた csv ファイルである．1 行目に “pse_id,time_id,reg_id” というヘッダを記し，2 行目以降には（仮名 ID，時刻）の小さい順に仮名 ID，時刻，加工後の領域 ID を記す．例えば，図 1 では，以下のようにする．

```
pse_id,time_id,reg_id
2001,5,*
2001,6,*
2001,7,5
2001,8,5
...
2003,8,*
```

8. **仮名表** $f^{(i,j)}$：仮名 ID，ユーザ ID が書かれた csv ファイルである．1 行目に “pse_id,user_id” というヘッダを記し，2 行目以降には仮名 ID の小さい順に仮名 ID，ユーザ ID を記す．例えば，図 1 では，以下のようにする．

```
pse_id,user_id
2001,2
2002,3
```

```
2003,1
```

9. **仮名表の推定値 $\hat{f}^{(i,j)}$**：仮名 ID の小さい順にユーザ ID の推定値が書かれた csv ファイルである（仮名 ID については記載しないことに注意）．1 行目に “user_id” というヘッダを記し，2 行目以降には仮名 ID の小さい順にユーザ ID の推定値を記す．例えば，図 1 では，以下のようにする．

```
user_id
2
2
1
```

10. **元トレースの推定値 $\hat{O}^{(i,j)}$**：（ユーザ ID，時刻）の小さい順に領域 ID の推定値が書かれた csv ファイルである（ユーザ ID，時刻については記載しないことに注意）．1 行目に “reg_id” というヘッダを記し，2 行目以降には（ユーザ ID，時刻）の小さい順に領域 ID の推定値を記す．例えば，図 1 では，以下のようにする．

```
reg_id
1
1
2
4
...
1
```

2.5 ファイル名

1. **全ファイル共通**：匿名加工を行うチーム番号 i（$1 \leq i \leq z$）を 3 桁の数字 XXX で，データセット番号 j（$1 \leq j \leq 2$）を 2 桁の数字 YY（= 01 or 02）で表す．また，攻撃（ID 識別またはトレース推定）を行うチーム番号 i'（$i' \neq i$）を 3 桁の数字 WWW で表す．さらに，元トレースに対して行われた加工が ID 識別対策用か，トレース推定対策用かを表すフラグ ZZZ を拡張子（.csv）の直前に付ける．データセット番号が $j = 1$ のときは ID 識別対策用に加工するため ZZZ=IDP とし（IDP は ID Protection の意味），$j = 2$ のときはトレース推定対策用に加工するため ZZZ=TRP とする（TRP は Trace Protection の意味）．
2. **参照トレース $R^{(i,j)}$ のファイル名：reftraces_teamXXX_dataYY_ZZZ.csv**
例えば，チーム番号 $i = 12$，データセット番号 $j = 1$ のときは reftraces_team012_data01_IDP.csv であり，$i = 12$，$j = 2$ のときは reftraces_team012_data02_TRP.csv である．
3. **元トレース $O^{(i,j)}$ のファイル名：orgtraces_teamXXX_dataYY_ZZZ.csv**
例えば，チーム番号 $i = 12$，データセット番号 $j = 1$ のときは orgtraces_team012_data01_IDP.csv であり，$i = 12$，$j = 2$ のときは orgtraces_team012_data02_TRP.csv である．
4. **加工トレース $A^{(i,j)}$ のファイル名：anotraces_teamXXX_dataYY_ZZZ.csv**
例えば，チーム番号 $i = 12$，データセット番号 $j = 1$ のときは anotraces_team012_data01_IDP.csv で

あり，$i = 12$，$j = 2$ のときは anotraces_team012_data02_TRP.csv である．

5. 公開加工トレース $A'^{(i,j)}$ のファイル名：**pubtraces_teamXXX_dataYY_ZZZ.csv**
   例えば，チーム番号 $i = 12$，データセット番号 $j = 1$ のときは pubtraces_team012_data01_IDP.csv であり，$i = 12$，$j = 2$ のときは pubtraces_team012_data02_TRP.csv である．
6. 仮名表 $f^{(i,j)}$ のファイル名：**ptable_teamXXX_dataYY_ZZZ.csv**
   例えば，チーム番号 $i = 12$，データセット番号 $j = 1$ のときは ptable_team012_data01_IDP.csv であり，$i = 12$，$j = 2$ のときは ptable_team012_data02_TRP.csv である（審判 Q のみが見れる）．
7. 仮名表の推定値 $\hat{f}^{(i,j)}$ のファイル名：**etable_teamWWW-XXX_dataYY_ZZZ.csv**
   チーム番号は，攻撃を行うチーム番号 i' と攻撃対象のチーム番号 i をハイフンを繋げて記載する．例えば，チーム番号 $i' = 20$ がチーム番号 $i = 12$，データセット番号 $j = 1$ の公開加工トレースを ID 識別するときは etable_team020-012_data01_IDP.csv であり，$i = 12$，$j = 2$ の公開加工トレースを ID 識別するときは etable_team020-012_data02_TRP.csv である．
8. 元トレースの推定値 $\hat{O}^{(i,j)}$ のファイル名：**etraces_teamWWW-XXX_dataYY_ZZZ.csv**
   チーム番号は，攻撃を行うチーム番号 i' と攻撃対象のチーム番号 i をハイフンで繋げて記載する．例えば，チーム番号 $i' = 20$ がチーム番号 $i = 12$，データセット番号 $j = 1$ の公開加工トレースをトレース推定するときは etraces_team020-012_data01_IDP.csv であり，$i = 12$，$j = 2$ の公開加工トレースをトレース推定するときは etraces_team020-012_data02_TRP.csv である．

2.6 禁止事項

2.6.1 匿名加工フェーズにおけるチームの禁止事項

以下の禁止事項を実施した場合は，最終順位が付かない場合があるので注意すること

1. 各チームで決められた上限を越えた数のデータを提出すること
2. 指定されたフォーマットに従わないデータを提出すること
3. 他のチームと結託すること．ただし, 自チームにのみ開示された情報が推定されない範囲で, プログラムやモジュール, アルゴリズムを共有することは, 結託にはあたらないとする

2.6.2 ID 識別・トレース推定フェーズにおけるチームの禁止事項

1. 指定されたフォーマットに従わないデータを提出すること
2. 他のチームと結託すること．ただし，プログラムやモジュール，アルゴリズムを共有することは結託に当たらないとする
3. 1 つの公開加工トレースに対して, 2 回以上仮名表の推定値を提出して, ID 識別を試みること
4. 1 つの公開加工トレースに対して, 2 回以上元トレースの推定値を提出して, トレース推定を試みること

2.6.3 実行委員の禁止事項

1. チームと結託すること (実行委員の特権により得た情報をチームに提供すること)
2. チームがそれを知ることで有利になるような情報を非公開にすること
3. コンテスト参加者として匿名加工や ID 識別・トレース推定を行う場合，データ提出受付期間中に実行

委員の特権を利用すること

2.7 公開情報

審判は以下の情報を全チームが閲覧できるように公開する．

2.7.1 サンプルプログラム

匿名加工と ID 識別・トレース推定のサンプルプログラム．詳細は文献 [4] を参照．

2.7.2 有用性評価プログラム

元トレース $O^{(i,j)}$ と加工トレース $A^{(i,j)}$ を入力として，有用性 $s_U^{(i,j)}$ を出力するプログラム．

2.7.3 ID 識別安全性評価プログラム

仮名表 $f^{(i,j)}$ と仮名表の推定値 $\hat{f}^{(i,j)}$ を入力として，ID 識別安全性 $s_I^{(i,j)}$ を出力するプログラム．

2.7.4 トレース推定安全性評価プログラム

元トレース $O^{(i,j)}$ と元トレースの推定値 $\hat{O}^{(i,j)}$ を入力として，トレース推定安全性 $s_T^{(i,j)}$ を出力するプログラム．

2.7.5 領域情報ファイル

各領域の情報が記されたファイル．「領域 ID，Y 軸（緯度）方向の ID，X 軸（経度) 方向の ID，領域中心部の緯度，領域中心部の経度，病院領域か否か（1:yes，0:no）を表すフラグ」からなる，以下のような csv ファイルである（1 行目はヘッダ）．

```
reg_id,y_id,x_id,y(center),x(center),hospital
1,1,1,34.6415625,135.441875,0
2,1,2,34.6415625,135.445625,1
3,1,3,34.6415625,135.449375,0
4,1,4,34.6415625,135.453125,0
...
1024,32,32,34.7384375,135.558125,0
```

2.7.6 時刻情報ファイル

各時刻（自然数）の情報が記されたファイル．「参照トレース（ref）と元トレース（org）のどちらか，時刻（自然数），日，時，分」からなる，以下のような csv ファイルである（1 行目はヘッダ）．

```
ref/org,time_id,day,hour,min
ref,1,1,8,0
ref,2,1,8,30
ref,3,1,9,0
```

```
ref,4,1,9,30
...
org,80,4,17,30
```

2.7.7 PWSCup2019 用人工データ（大阪データセット）

疑似人流データ [2] から大阪（緯度：34.64〜34.74，経度：135.44〜135.56）に対して，（東京と同じように）32×32 の領域に分割して生成モデルを学習し，2 チーム分の参照トレース $R^{(i,j)}$，元トレース $O^{(i,j)}$ を生成したもの（$1 \leq i \leq 2$，$1 \leq j \leq 2$）．参照トレースと元トレースがそれぞれどのようなものか，に関する理解が深まるよう全チームに公開する．

尚，このデータセットに対してサンプルプログラムを用いた評価実験を行っている．詳細は文献 [4] を参照．

2.7.8 PWSCup2019 用人工データ（東京データセット）

匿名加工フェーズにおいて，各チームが自身のデータと他のデータを比較できるように，（ユーザ集合がどのチームのトレースとも異なる）2 チーム分の参照トレース・元トレースを公開する．即ち，チーム数を z としたときに，$z+1$ 番目のチームと $z+2$ 番目のチーム用の参照トレース $R^{(i,j)}$，元トレース $O^{(i,j)}$（$z+1 \leq i \leq z+2$，$1 \leq j \leq 2$）を生成し，これを公開する．

2.8 予備戦

以下，予備戦に関する決定事項を記載する．

1. 有用性の要求値 s_{req} を，$s_{req} = 0.7$ とする．
2. 予備戦終了後，予備戦における各チームの元トレース，加工トレース，仮名表，および他チームに対する仮名表の推定値，元トレースの推定値を全チームに公開する．

2.9 本戦に追加するルール

以下，予備選終了後に，本戦に追加的に適用する可能性のある追加ルールを記載する．これらの追加ルールの適用が決定された場合，本戦前に全チームにアナウンスする．

1. 有用性の要求値 s_{req} を，予備戦の値から変更する可能性がある．
2. 参照トレース・元トレースの長さを，予備戦の長さ（それぞれ 2 日分）から変更する可能性がある．

[決定事項] 本戦では有用性の要求値を $s_{req} = 0.7$ とし，参照トレース・元トレースの長さを，それぞれ 20 日分ずつとする．尚，ここでの参照トレース・元トレースは，（平日 5 日の後に土日が続くような）連続した日々から構成されたものではなく，散発的にサンプリングされた日々から構成されたものである．

参考文献

[1] 村上隆夫，荒井ひろみ，井口誠，小栗秀暢，菊池浩明，黒政敦史，中川裕志，中村優一，西山賢志郎，野島良，波多野卓磨，濱田浩気，山岡裕司，山口高康，山田明，渡辺知恵美，“PWS Cup 2019: ID 識別・

トレース推定に強い位置情報の匿名加工技術を競う”，CSS2019.
[2] ナイトレイ，東京大学空間情報科学研究センター (CSIS)，疑似人流データ: https://nightley.jp/archives/1954/
[3] PWS Cup 2019 データセットについて：https://www.iwsec.org/pws/2019/cup19-dataset.pdf
[4] PWS Cup 2019 サンプルプログラムについて：https://www.iwsec.org/pws/2019/cup19-sample.pdf