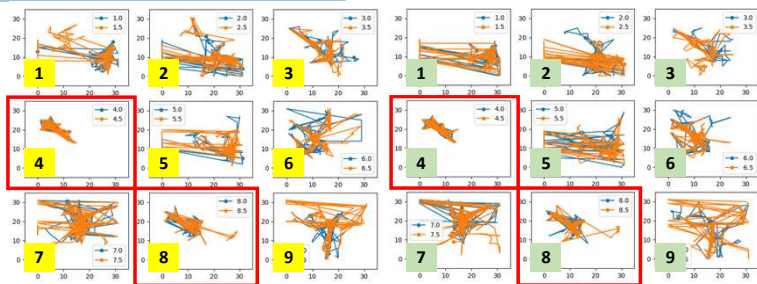


## 加工フェイズ戦略

元トレース(左)と参照トレース(右)の顧客の軌跡



元トレースと参照トレースで軌跡情報が似ている顧客がおり、こういった顧客は加工しないと識別される！

顧客ごとの各時刻の標準偏差(20日分)

	8:00	8:30	9:00	9:30	10:00	10:30	11:00	11:30	12:00	12:30	13:00	13:30	14:00	14:30	15:00	15:30	16:00	16:30	17:00	17:30
1	92	109	107	108	107	113	115	123	136	128	118	125	129	126	113	126	113	104	101	123
2	64	102	96	111	114	118	127	125	137	130	131	132	155	142	162	167	171	167	165	
3	8	28	64	57	91	78	101	118	119	117	137	107	118	110	147	148	145	137	141	153
4	25	42	51	58	78	76	87	71	80	93	102	101	99	110	75	111	95	80	90	67
5	19	38	55	73	75	97	93	104	103	101	113	100	106	111	104	115	115	122	118	113
6	27	86	95	117	124	143	137	124	135	161	154	166	169	174	151	148	202	194	193	201
7	60	137	142	140	141	161	161	140	136	98	120	137	122	137	137	134	130	156	167	152
8	55	66	87	101	99	122	122	104	107	113	129	136	138	128	110	93	118	123	107	117
9	46	80	123	133	153	170	170	173	160	164	177	161	151	161	164	152	138	157	157	176

元トレースで毎日特定の時刻に同じような場所にいる顧客は、参照トレースでも似たような移動をしていることが予想される！  
危険な(標準偏差が低い)時刻の履歴を加工する必要がある！

	8	8.5	9	9.5	10
加工されたトレース					
参照トレース					
標準偏差計算低いところ分析					
削除されている組み合わせを分析					
2001	1	1	0	0	0
2002	0	0	0	0	0
2003	0	0	0	1	1
2004	1	0	0	0	1

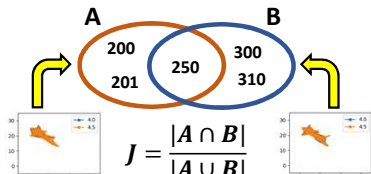
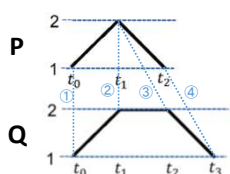
危険な時刻の地域IDを削除によって加工すると、加工トレースと参照トレースの削除パターンを比較することにより個人が識別される可能性がある！→**削除ではなくノイズ付加をする！**

## 識別フェイズ戦略

ID識別戦略：DTW距離やJaccard距離を用いて顧客の軌跡と仮名の軌跡の間の距離を求め、個人を識別する！

DTW距離(両部門で使用)  
2つの時系列データ間の距離を測る指標  
距離を最小化するように時間軸を伸長させるため、異なる長さのデータ間の距離も測れる

Jaccard距離(トレース推定部門のみで使用)  
2つの集合間の距離を測る指標  
顧客が訪れる全地域IDをその顧客の集合とし、仮名との距離を計算する(ノイズ付加には弱い)



参照トレース			加工トレース			仮名推定表	
user_ID	time_ID	reg_ID	仮名	time_ID	reg_ID	仮名	user_ID
1	401	30	2001	401	40	2001	1
1	402	31	2001	402	42	2002	10
:	:	:	:	:	:	:	:
1	799	60	2001	799	50	3999	10
1	800	59	2001	800	50	4000	1

DTW距離や  
Jaccard距離で

最も近い顧客を  
正解と推定する

DTW距離やJaccard距離で最も近い顧客を正解と推定する

元トレース

user ID	時刻	地域 ID
1	7/1 08:00	100
1	7/1 09:00	110
1	7/1 10:00	110
1	7/2 08:00	100
1	7/2 09:00	100
1	7/2 10:00	110
2	7/1 08:00	200
2	7/1 09:00	200
2	7/1 10:00	210
2	7/2 08:00	210
2	7/2 09:00	210
2	7/2 10:00	210

標準偏差表

	user 1	user 2
08:00	0	b > 0
09:00	a > 0	c > 0
10:00	0	0

加工トレース

user ID	時刻 ID	地域 ID
1	7/1 08:00	115
1	7/1 09:00	110
1	7/1 10:00	200
1	7/2 08:00	120
1	7/2 09:00	100
1	7/2 10:00	150
2	7/1 08:00	200
2	7/1 09:00	200
2	7/1 10:00	100
2	7/2 08:00	210
2	7/2 09:00	210
2	7/2 10:00	300

標準偏差低い順ランキング

RANK	user ID	時刻
1	1	8:00
2	1	10:00
3	2	10:00
4	2	8:00
5	1	9:00
6	2	9:00

ID識別対策

顧客ID\*各時刻の標準偏差を求めてランキングを作り、上位x%の組み合わせの地域IDをランダムな地域に飛ばす

元トレース

user_id	time_id	reg_id
1	401	605
1	402	445
1	403	316
1	404	2
1	405	315
1	406	282
...	...	...
10	800	635

病院を含む地域IDがずらされたトレース

user_id	time_id	reg_id
1	401	605
1	402	445
1	403	316
1	404	34
1	405	315
1	406	282
...	...	...
10	800	635

ランダムに±1か±32 (1マスずらす)

993

994

...

1023

1024

961

962

...

991

992

...

...

...

...

...

33

34

...

63

64

1

2

...

31

32

トレース推定対策

上記のID識別対策加工を行った後、病院情報の含まれる地域IDをランダムに1マスずらす(10倍スコアを防ぐため)また、全時刻の履歴が危険だと判断された顧客の履歴は、全く削除されない顧客のトレースと同じにする。

加工トレースで地域IDが削除されている場合は1つ前の値か1025に置き換えて計算し、一般化されている場合には先頭の値か重心に置き換えて計算する。

reg_ID	reg_ID	reg_ID
42	42	42
*	42	1025
reg_ID	reg_ID	reg_ID
42	42	42
40 42 104 106	40	73

トレース推定戦略：DTW再識別・Jaccard再識別によって個人を識別し、加工トレースの地域IDを再送する。識別されなかった顧客や削除されている箇所に対しては参照トレースを再送する。

加工トレース

仮名	time_ID	reg_ID
2001	401	40
2001	402	45
2002	401	*
2002	402	100
2003	401	310
2003	402	320

加工トレース

ID	time_ID	reg_ID
1	1	100
1	2	110
2	1	30
2	2	20
3	1	400
3	2	400

仮名推定表

仮名	user_ID
2001	1
2002	2
2003	1

推定トレース

reg_ID
40
45
30
100
400
400

加工トレースと参照トレースの顧客間のJaccard距離平均が小さいチームに対してはJaccard再識別で攻撃し、それ以外のチームに対してはDTW再識別で攻撃する。また、病院周辺の地域IDはすべて病院のある地区が加工されたものであると推定する。

各加工データのJaccard距離平均表	
データ	Jaccard距離平均
Team001_TRP	0.555
Team002_TRP	0.768
:	:
Team019_TRP	0.672

993	994	...	1023	1024
961	962	...	991	992
...	...	...	...	...
33	34	...	63	64
1	2	...	31	32