

Traitement des problèmes liés au tracking dans les données

On observe trois types de problèmes liés aux ID :

1. *Saut d'ID* : lorsque l'ID d'un agent disparaît pendant une (ou plusieurs) frame(s).
2. *Échange d'ID* : lorsque deux (ou plus) agents intervertissent leurs ID lors d'une interaction.
3. *Apparition/disparition d'ID* : lorsqu'un ID apparaît ou disparaît hors des zones d'apparition/disparition (les bords gauche et droit du pont).

1 Traitement de chaque problème.

1.1 Saut d'ID

Le *saut d'ID* (*ID jump*) est le problème le plus simple à régler : supposons que l'ID i a pour position $X_n \in \mathbb{R}^2$ à la frame $n \in \mathbb{N}$ et disparaît à la frame $n + 1$ pour réapparaître à la frame $n + m$, avec $m \in \llbracket 2, +\infty \rrbracket$, ayant pour position X_{n+m} . On interpolera alors simplement linéairement :

$$\forall k \in \llbracket 1, m - 1 \rrbracket, \quad X_k := \left(1 - \frac{k}{m}\right)X_n + \frac{k}{m}X_{n+m}.$$

On trace sur la figure ci-dessous l'histogramme du **nombre** et de la **fréquence** de frames sautées par agent (c'est à dire $\frac{\text{\#frame sautees}}{\text{\#frames tot}}$). Pour calculer le nombre de frames sautées, on compte l'écart entre la dernière et la première frame d'existence, auquel on retranche le nombre de frames d'existence de l'agent.

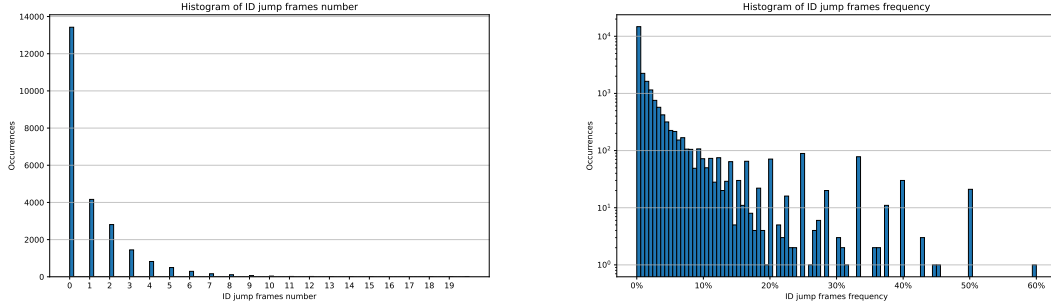


Figure 1. *Histogramme* du **nombre** (g) de frames sautées ainsi que de la **fréquence** (d) de frames sautées relativement à la trajectoire pour chaque agent. Gauche : échelle linéaire. Droite : échelle log en y et linéaire en x .

Notons qu'ici on observe le *nombre de frames sautées sur toute la trajectoire*. Ainsi, le *nombre de frame sautées par saut est forcément inférieur* à ces valeurs. Cet histogramme nous permet d'affirmer que le saut d'ID, bien que très fréquent (**44% des agents subissent un saut d'ID !**), ne dure que quelques frames* (*dans l'immense majorité des cas, il durera moins de 5 frames, soit 0.2 secondes!*), et qu'une interpolation linéaire est donc une approximation valable.

* Si on exclut les agents qui ne subissent pas de saut d'ID, le **nombre d'agents qui subissent $n \geq 1$ jump frames le long de leur traj. est proportionnel à $e^{-\frac{n}{2}}$.**

Attention : lorsqu'un jump d'ID survient, il est possible qu'il s'accompagne d'un ID switch. Nous traitons donc successivement ces problèmes.

1.2 Échange d'ID

L'échange d'ID (*ID switch*) est bien plus délicat à traiter. La difficulté réside en premier lieu sur la capacité à **déterminer quand il survient**. Cela se fera difficilement instantanément – on pourra utiliser les orientation – mais cela peut se repérer globalement : si un agent apparaît et disparaît (après une trajectoire de longueur minimale) à la même extrémité du pont, on peut dire avec confiance qu'il a subi un *ID switch*. Cependant cela ne nous permet que de fournir la caractérisation suivante :

- Un agent qui apparaît et disparaît à la **même** extrémité aura subi un nombre **pair** d'ID switch
- Un agent qui apparaît et disparaît aux **deux** extrémités aura subi un nombre **impair** d'ID switch

On ne peut donc pas même affirmer, via cette caractérisation, que les trajectoires totales sont exemptes d'ID switch.

Pour se représenter la quantité d'ID switch dans les données, nous partons du **postulat** suivant :

“La composante en x de la vitesse d'un agent sera globalement toujours dirigée dans la même direction (pas de demi-tour).” Ainsi, l'on peut chercher à étudier la **proportion de temps que l'agent passe dans la 'bonne' et la 'mauvaise' direction en x** (on ne s'intéresse pas ici à la composante en y). On détermine la 'bonne' direction de l'agent comme celle où il passe le plus de temps – la proportion de temps passé dans la mauvaise direction est donc par définition comprise entre 0 et 50%.

Autrement dit, on va regarder le ratio du nombre d'occurrences de l'événement “ v_x est positif” par le nombre d'occurrences d'existence de v_x (le nombre de frames d'existence de l'agent moins 1). Ce ratio est toujours compris entre 0 et 1. S'il est compris entre 0.5 et 1, cela signifie que la bonne direction de l'agent est négative, on remplace donc ce ratio par 1 – lui-même. On obtient l'histogramme suivant :

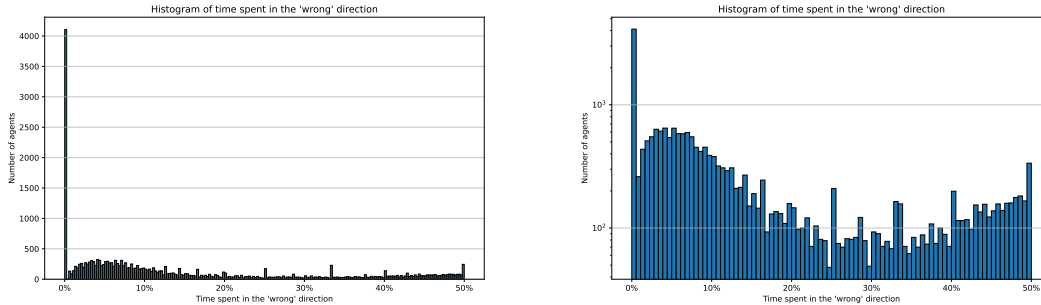


Figure 2. Histogrammes de la proportion de temps passé dans la “mauvaise” direction. Gauche : échelle linéaire, 200 bins. Droite : échelle log en y et linéaire en x , 90 bins.

On constate qu'environ 4100 agents ne passent pas de temps dans la 'mauvaise' direction... Mais souvenons nous qu'il y a 23884 agents ! On peut même constater que 9866 agents passent plus de 10% de leur temps dans la 'mauvaise' direction !

Ce n'est pas une preuve formelle d'échange d'ID, mais si le **postulat** tient (avec une tolérance du ratio de 2%), on s'attend alors à ce que dans nos données, au moins 75% des agents aient subi au moins un *ID switch* ! C'est conséquent !

1.3 Apparition/disparition d'ID

L'apparition/disparition d'ID (*ID appearance/vanish*) est très certainement le point le plus délicat à régler, même s'il est déjà bien plus simple à identifier.

On parlera d'*apparition/disparition d'ID* lorsque l'agent naît ou meurt hors des zones d'entrée ou de sortie du pont. Pour être plus précis, si la section du pont étudiée est $[0, L_x] \times [0, L_y]$, la zone d'entrée/sortie sera $[0, \varepsilon] \times [0, L_y] \cup [L_x - \varepsilon, L_x] \times [0, L_y]$, pour un certain petit paramètre $\varepsilon > 0$. Ainsi, il est très facile d'identifier la survenue d'une apparition/disparition.

Cet événement peut survenir pour deux raisons :

1. L'agent sort du pont par le bord (haut ou bas)
2. Misfit du tracking dû à un overcrowding (tout en anglais!)

Pour le premier point, l'apparition/disparition est naturelle et peut (doit) tout à fait être gardée comme telle. En revanche, pour le second, cela pose des problèmes si l'on souhaite faire une étude des interactions.

On caractérisera alors assez directement une *apparition/disparition d'ID*

1. due à une sortie/entrée par le bord (h ou b) si l'agent naît/meurt dans $[\varepsilon_x, L_x - \varepsilon_x] \times [0, \varepsilon_y] \cup [\varepsilon_x, L_x - \varepsilon_x] \times [L_y - \varepsilon_y, L_y]$
2. due à un misfit du tracking si l'agent naît/meurt dans $[\varepsilon_x, L_x - \varepsilon_x] \times [\varepsilon_y, L_y - \varepsilon_y]$.

On trouve que pour 23884 agents, avec $L_x = 100$, $L_y = 20$, $\varepsilon_x = \varepsilon_y = 2$, 18164 agents subissent une apparition/disparition d'ID, dont 13381 à cause du bord et 4783 via les misfits de tracking. En d'autres termes,

1. **56%** des agents naissent ou disparaissent au bord
2. **20%** des agents naissent ou disparaissent lors d'une interaction

Sur le graphique suivant, on représente l'influence du bord sur le temps de vie des agents. On représente le temps de vie moyen des agents qui apparaissent/disparaissent dans le domaine $[\varepsilon_x, L_x - \varepsilon_x] \times [\varepsilon_y, L_y - \varepsilon_y]$ en fonction de ε_x et ε_y (c'est donc une heatmap).

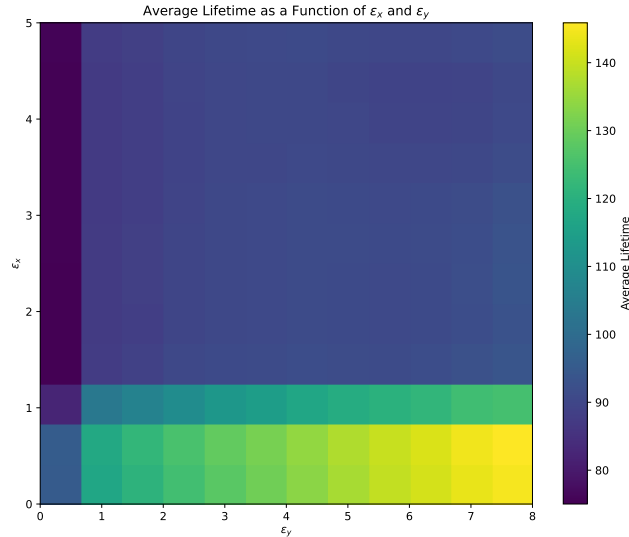


Figure 3. Heatmap du temps de vie moyen en fonction de l'apparition/disparition dans les domaines données par les paramètres ε_x et ε_y .

On observe sur la figure ci-dessus l'influence du bord sur le temps de vie : plus un agent entre ou sort du pont loin du bord, plus son temps de vie moyen sera élevé (voir les faibles valeurs de ε_x). Le temps de vie des agents qui apparaissent/disparaissent proche du bord est très affaibli (voir les faibles valeurs de ε_y).