Discussion on 17.07.2018

1. To calculate the quality of prediction:
   Example:
   Current_hour = 5
   Next_hour = 6

   Step 1:
   Calculate all stay points with IDs and their corresponding probabilities till next_hour.
   $g(l_t)$ = {($id1_t$,0.6), ($id2_t$,.0.2), ($id3_t$,0.1), ($id4_t$,0.1)}

   Step 2:
   Now predict based on training data for all these ids.
   $Id1_t$ = {(id1,0.07), (id2,0.2), (id3,0.7), (id4,0.03)...)}
   $Id2_t$ = {(id1,0.8), (id2,0.03), (id3,0.01), (id4,0.01)...)}
   $Id3_t$ = {(id1,0.7), (id2,0.01), (id3,0.2), (id4,0.01)...)}
   $Id4_t$ = {(id1,0.1), (id2,0.3), (id3,0.05), (id4,0.006)...)}

   Weight all the ID predictions with their own probability.
   $Id1_t$ = {(id1,0.07), (id2,0.2), (id3,0.7), (id4,0.03)...)} * 0.6
   $Id2_t$ = {(id1,0.8), (id2,0.03), (id3,0.01), (id4,0.01)...)} * 0.2
   $Id3_t$ = {(id1,0.7), (id2,0.01), (id3,0.2), (id4,0.01)...)} * 0.1
   $Id4_t$ = {(id1,0.1), (id2,0.3), (id3,0.05), (id4,0.006)...)} * 0.1
   And sum them up to get a single vector. Arrange these predictions in descending order and remove cumulative probabilities after a threshold θ (example 0.8).

   $g(t_{t+1})$' = {(id1, 0.282), (id2,0.157), (id3,0.447), (id4,0216) ..} This is the overall prediction.

   Step 3:
   Calculate the stay points with IDs and their corresponding time of stay for next hour.
   $g(l_{t+1})$ = {($id1_{t+1}$,0.9), ($id2_{t+1}$,.0.1)}

   Step 4:
   Now there are two cases:
   A. User visits one of the predicted places
   B. Calculate similarity for our predictions versus the ground truth

   For Case A: It is enough if $g(l_{t+1})$ has any location in $g(l_{t+1})$'.
   For case B:
   Dot product of $g(l_{t+1})$ and $g(l_{t+1})$' together will all the missing locations in both the vectors.
   Similarity = a(dot product) / b(magnitude of $g(l_{t+1})$)

2. Try clustering on raw points to see if we have anything new. Each new cluster should be treated separately in each iteration.