# Introduction

Background:

In urban traffic, road users are often found moving in groups. These groups can be formed for different reasons. For instance, social connections between pedestrians (e.g. friends, couples, families); the mixed group formed by traffic regulations, like traffic participants who follow the same phase of traffic lights, etc. The members of the same group tend to keep similar speed and appropriate distance.

An obvious benefit that comes from grouping is safety. Being in a group creates a buddy system where people can look after one another on the streets. \cite{jacobsen2015safety} found that people walking and bicycling in larger groups are less likely to be injured by motorists because the motorists are more cautious with groups. Therefore, grouping can increase the dominance of vulnerable road users.

It will also have a beneficial effect on traffic planning: if groups are formed, this leads to an reduction in the number of road users that have to be included in computations, thus leading to a decrease in computational complexity for later applications, e.g. for traffic simulation and pedestrian navigation.

Motivation: ???

Inspired by the phenomenon above, the grouping of road users in shared spaces taken into consideration to increase safety and efficiency.

State-of-the-art: (可以与literature review合并) 5-7个就足够了

In a basic formulation, the facility location problem is the following: giving a set of demand points and a set of candidate facility sites with costs of building facilities at each of them, the goal is to select a subset of sites where facilities should be built. Each demand point is then assigned to the closest facility, incurring a service cost equal to the distance to its assigned facility. The objective is to minimize the sum of facility costs and the sum of the service costs for the demand points \cite{charikar1999improved}.

This problem has already been addressed by some authors. \cite{szkandera2017path} simply used a threshold based on the distance between the team leader and members to group the pedestrians with similar OD. However, this simple method is sensitive to the order of the input data because the algorithm is greedy - once the first possible solution is accepted, other solutions will never be reconsidered. \cite{he2016dynamic} clustered the original groups by the pairwise similarity metric defined over agents based on their starting positions and velocities. This works for the simulation application because the agents who are together at the beginning will keep coherent until the end of the experiment. However, the traffic scenario is more complicated. E.g. the road users who have the same origin and velocity at the beginning may split and reach different goals later.

Knowledge gaps

1. Current area traffic mainly consider the situation when there is lanes for different road users. However, in shared spaces, the traffic signs are removed from the road, therefore, the location pf clustering should be decided.
2. The area traffic regulation systems, like SCTOOT and GJ are not really online or dynamic. It learns regulations from historical data, and keep it as fixed periods. However, in shared spaces, the coming and leaving spots could be very different in different time periods, therefore, a dynamic algorithm is needed for road users here.

Objectives:

The objective of this paper is to find an dynamic clustering algorithm for the road users in shared spaces, considering their origins, destinations, and even the trajectory shapes.

(Definations): 非必须

Here, a group is a formation of road users moving in a coordinated manner. A group can split, merge, avoid collisions while moving (\cite{mihaylova2014overview}).

# Methodology

We concentrate on the following application scenario: Road users can appear from random locations around the shared space, then pass through, finally leave to their destinations. We are searching for a clustering algorithm to assign those road users to several groups according to their origins, destinations (OD data) and time.

In our application, where the incoming road users need to form a group, the group center/leader can be seen as the facility, and all road users are customers. In addition, the crossing behavior is relatively dynamic, which means the road users will appear when they arrive at the crossing and disappear when they finish crossing. Therefore, the temporal parameters should also be taken into consideration.

The new algorithm is based on the framework proposed by []. The framework will be shortly described here:

Therefore, our main algorithm works as follows: start with a solution given by Meyerson’s algorithm of cost Θ, use Lemma 2.3 to maintain a solution during Θ/4αf updates, and then recomputed from scratch. We call the intervals between consecutive recomputations periods, and note that they are random objects: the length of a period is determined by the cost of its initial solution, which is a random variable.

Assume the data is continuously come and ordered with coming moments. A window with size w is sliding from the start of data and each time insert a new point/delete the oldest point from the current data. Before the sliding window runs, w points are taken to calculate a course solution with repeated Meyerson algorithm, which is an approximate result of optimized solutions. Then, when each new point x is inserted, the transport cost from x to all current centers are calculated. If the minimum distance d is larger than the facility construction cost f, there has a possibility of d/f to build a new center, otherwise the new point will be assigned to the closet centers. The cost of each update is added in a round, until it reaches the threshold of update cost. Besides the cost threshold, to include the temporal info into consideration, we import a threshold called waiting time. In a cross, the road users has a tolerance waiting time, if they have to wait too long, they might go regards any rules. []. Therefore, there are two rules to control the update: cost update and waiting recomputation.

The pseudo code of current algorithms can be summarized as below:

Input: A list of road users, with their appearing time ot, appearing location (ox,oy), destination location (dx,dy), the sliding window size w, waiting time th\_waiting.

Output: A list of center locations Facils, and the facils update time mutation.

Let Facils = [] and take the first w points from data.

1: curse solution Meyerson():

1. generate initial solution by Meyerson(list of x0 to xw);
2. repeat Meyerson 5 times
3. Lowerest cost theta and initial centers

2: while taken the last user:

1. Waiting <- x\_w+t[‘of’] – x\_w[‘of’]
2. If waiting > th\_waiting:
3. Delete x0:xw <- previous center have gone
4. Recompute solution for ne period by Meyerson()
5. If waiting <= th\_waiting:
6. If t – lasttime > cost threshold: # reach the recompuation cost criteria
7. Recompute solution by Meyerson()
8. Else
9. If OD\_similarity < f:
10. Assign xt to nearest center
11. Else
12. Create new center at location xt
13. take a coming user, calculate its distance d <- OD\_similarity to the nearest center vi
14. if d

Algorithm Meyerson()

Distance metric OD\_similarity()

Return

data preprocessing

online facility location with OD data

explain the parameters

# Result

results on diff parameters

# Discussion

5.2 compare with state-of-art methods

# Conclusion

\section{Methodology}

In the following we present a definition suitable for shared space. Assume a set of road users J who need to be grouped during a finite time can be located and assume that several group centers $p\_{t}$ have to be activated in every period. The multi-period planning horizon, T. Let $I \subseteq J$ be the set of user locations where the group centers can be located and assume that $p\_{t}$ group centers have been activated in every period. The problem of deciding the best location for the group centers in each period, minimizing the total cost for reaching surrounding group members can be formulated as follows (\cite{nickel2019multi}):

In this model, group centers can be activated (deactivated) at the beginning (end) of a period; $m\_{t}$ is the maximum number of group centers that can be activated in each period $t \in T$. The binary variable $z\_{it}$ is equal to 1 if a center is activated at $i\in I$ in period $t\in T$ and 0 otherwise. The parameter $g\_{it}$ denotes the activation cost. The deactivation cost will not be considered in current case. The number of activated centers does not have to be the same in all periods, $p\_{t}$ denotes the number of centers to be activated in that period $t\in T$. \cite{galvao1992lagrangean} solved this kind of problem by Lagrangian relaxation based procedures.

Datasets provided by \cite{pascucci2017discrete} or \cite{robicquet2016learning} will be used. The paper will describe the approach and the experiment in detail. The source code and data will be made publicly available.

\section{Outlook}

In the paper, the basic grouping strategy of road users in shared space will be presented. Future work will investigate more interaction effects, e.g.: the correlation between a variety of road users (e.g. pedestrians, cyclists, vehicles, etc.), the different possibilities to form groups, as well as forming groups which are allowed to split and merge.