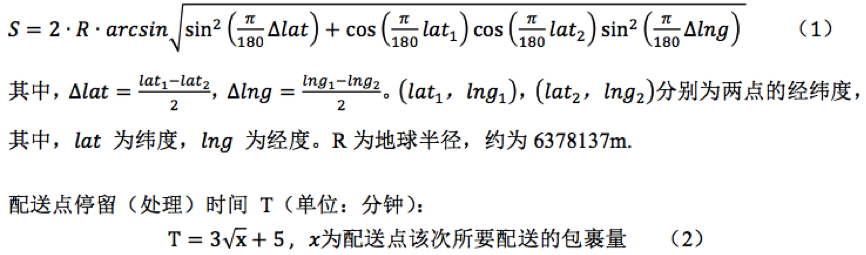
**赛题背景**：  
电商的蓬勃发展使得目前很大一部分的物流包裹均来源于线上电商订单。在中国，该比例超过了60%。这些包裹在配送的最后环节，是由快递员将包裹从网点送到消费者手中。另一方面，随着互联网逐渐向线下渗透，涌现出了越来越多的同城包裹配送需求，如外卖订单或鲜花蛋糕等等同城订单。这两类包裹的配送是目前中国最后一公里配送中最典型的场景。这次菜鸟算法大赛希望能够通过大数据对中国物流最后一公里提供智能的配送方案，通过全局优化来提升效率及降低成本。

**赛题内容：**  
最后一公里极速配送大赛主要针对赛题背景中提到的两类包裹提供最优的快递员配送方案。第一类是电商包裹，快递员需要从网点提取并配送至消费者，第二类是同城O2O包裹，快递员需要在指定时间去商户提取并在指定时间内配送至消费者。  
上海市有124个网点，这些网点负责全部上海电商包裹的配送,大概每天229000件。每个网点的配送范围两两不重合并完全覆盖了整个上海。每天8点前，所有上海的电商包裹都抵达网点。快递员在8点开始从网点进行派送，快递员需要在晚上8点前将所有的电商包裹送至消费者手中。在配送电商包裹的同时,快递员还要配送同城O2O包裹（如外卖订单等），对于这类包裹，快递员需要在指定时间之前到达商户领取包裹，并在指定时间之前送达消费者手中。参赛选手需要提供所有快递员的调度计划，即快递员在网点、配送点和商户的到达时间，离开时间，取/送订单及包裹量。  
为了简化模型，我们做了如下假设:  
1. 我们将上海所有要派送的包裹（含电商包裹和同城O2O包裹）地址汇聚9214个配送点上，每个配送点上都有若干包裹。快递员把包裹送至消费者手中可抽象为快递员把包裹送到离消费者最近的配送点，并在配送点处理完成。因此，我们将不会提供消费者的地址信息，只提供这9214个配送点的经纬度。  
2. 由于每个网点的配送范围两两不重合，所以每个配送点仅会被一个网点服务到。  
3. 每个快递员任何时刻携带的包裹量不得大于140件。  
4. 每个配送点的电商包裹只能一次配送完毕,不能分多次配送。  
5. 我们选取上海较大的O2O商户（含外卖等）598家。这些同城O2O包裹将由快递员到商户取走并送到我们的配送点给消费者。每笔同城O2O订单可能包含若干包裹，从消费者体验考虑，只可在满足总包裹量不大于140件的情况下一次取走，不可分批取走。  
6. 每笔同城O2O订单，有商户的领取时间限制，快递员不得晚于该时间到达商户，如果快递员早于该时间到达，则快递员需要等待至领取时间。同时，每笔订单还有消费者最晚收货时间，即快递员不得晚于该时间送达至配送点，如果快递员早于该时间送达至配送点，快递员不需等待，可直接进行配送。  
7. 我们提供所有地点的经纬度信息并假定两辆之间的距离符合计算公式（1），并且快递员的平均时速是15公里／小时.   
8. 快递员在配送点的停留(处理)时间，这里停留时间和配送地点该次所要配送的包裹量相关，并满足计算公式（2）。  
9. 快递员数量上限1000个，有配送任务的快递员数量可以少于1000，快递员可以服务多个网点，编号从D0001到D1000。  
10. 作为起始条件，参赛选手可以指定快递员8点从任何一个网点出发，快递员如果完成当天的任务，则在最后一个配送点结束。

参赛的选手需要提供所有快递员的派送计划，含快递员id，每个地点（网点、递送点或者是商户）的到达时间和离开时间，每个地点的取/送订单id及包裹量。

两点间距离计算公式



**赛题数据集：**  
1. 网点id及经纬度，供124个网点

|  |  |
| --- | --- |
| 字段 | 说明 |
| Site\_id | 网点id（e.g. A001） |
| Lng | 网点经度 |
| Lat | 网点纬度 |

2.     配送点id及经纬度，共9214个配送点

|  |  |
| --- | --- |
| 字段 | 说明 |
| Spot\_id | 配送点id （e.g. B0001） |
| Lng | 配送点经度 |
| Lat | 配送点纬度 |

3.     商户id及经纬度，共598个商户

|  |  |
| --- | --- |
| 字段 | 说明 |
| Shop\_id | 商户id （e.g. S001） |
| Lng | 商户经度 |
| Lat | 商户纬度 |

4.     电商订单，共9214笔电商订单，总包裹量为229780

|  |  |
| --- | --- |
| 字段 | 说明 |
| Order\_id | 订单id（e.g. F0001） |
| Spot\_id | 配送点id |
| Site\_id | 网点id |
| Num | 网点需要送至改配送点的电商包裹量 |

5.     同城O2O订单,共3273笔O2O订单，总包裹量为8856

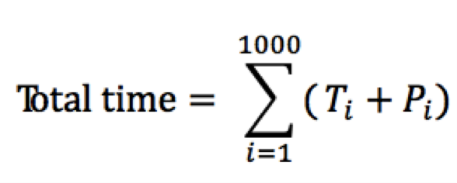
|  |  |
| --- | --- |
| 字段 | 说明 |
| Order\_id | 订单id（e.g. E0001） |
| Spot\_id | 配送点id |
| Shop\_id | 商户id |
| Pickup\_time | 到商户的领取时间（e.g. 11:00） |
| Delivery\_time | 送达至消费者的最晚时间（e.g. 20:00） |
| Num | 订单所含包裹量 |

6.     快递员id列表，最多1000位小件员

|  |  |
| --- | --- |
| 字段 | 说明 |
| Courier\_id | 快递员id（e.g. D0001） |

**赛题评估标准：**  
选手需提交所有快递员的调度计划：

|  |  |
| --- | --- |
| [字段](http://?spm=5176.100068.555.8.m54xpW) | 说明 |
| Courier\_id | 快递员id |
| Addr | 网点或配送点或商户id |
| Arrival\_time | 到达时长（距离08:00时长分钟数. e.g. 到达时刻为11:00，则到达时间为180） |
| Departure | 离开时长（距离08:00时长分钟数. e.g. 离开时刻为15:00，则离开时间为420） |
| Amount | 取/送货量（取为+，送为-） |
| Order\_id | 订单id |

所有快递员总耗时，最短的队伍获胜。所有快递员的总耗时包括所有快递员的行驶时间和所有配送点的停留处理时间：  
  
其中Ti为快递员i的所有行驶时间，Pi为快递员i在他所经过的所有配送点的停留处理时间。  
1. 请按arrival\_time 排序，注意同一个配送地点的到达时间和离开时间需满足假设8，前一个地点的离开时间和下一个地点的到达时间需满足假设7的行驶时间.如果同一个配送地点的到达时间和离开时间不满足假设8或前一个地点的离开时间和下一个地点的到达时间不满足假设7的行驶时间，则产生偏差的时间部分10倍计入总耗时。  
示例1. 假定快递员在一个配送点要处理36个包裹，如果他在9点到达，则根据假设8，他的处理时间是23分钟，因此他在9:23离开。如果参赛者提供的时间不为9:23（比如9:25），那么就产生了误差时间（|9:25-9:23|=2 minutes），10\*2=20分钟将计入总耗时。  
示例2. 假定快递员从配送点A至配送点B，如果两点距离为10km，根据假设7, 行驶时间为40分钟。如果快递员9:00离开A，则他到达B的时间为9:40。如果参赛者提供的到达时间不为9:40（比如9:35），那么就产生了误差时间（|9:40-9:35|=5 minutes），10\*5=50分钟将计入总耗时。  
2. 如果出现 O2O 到达商户的时间超过商户要求的领取时间，或送达用户的时间超过用户要求的送达时间，则超出的时间5倍计入总耗时。  
3. 所有时间精确到分钟

**实例：**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| D0545 | A083 | 0 | 0 | 34 | F6344 |
| D0545 | A083 | 0 | 0 | 11 | F6360 |
| D0545 | A083 | 0 | 0 | 19 | F6358 |
| D0545 | A083 | 0 | 0 | 12 | F6353 |
| D0545 | A083 | 0 | 0 | 63 | F6354 |
| D0545 | B5800 | 7 | 29 | -34 | F6344 |
| D0545 | B7555 | 30 | 59 | -63 | F6354 |
| D0545 | B7182 | 62 | 77 | -12 | F6353 |
| D0545 | B8307 | 79 | 97 | -19 | F6358 |
| D0545 | B8461 | 102 | 117 | -11 | F6360 |
| D0545 | A083 | 124 | 124 | 46 | F6349 |
| D0545 | A083 | 124 | 124 | 53 | F6325 |
| D0545 | A083 | 124 | 124 | 39 | F6314 |
| D0545 | B6528 | 132 | 157 | -46 | F6349 |
| D0545 | S245 | 160 | 257 | 1 | E0895 |
| D0545 | B3266 | 259 | 267 | -1 | E0895 |
| D0545 | B3266 | 267 | 294 | -53 | F6325 |
| D0545 | B2337 | 296 | 320 | -39 | F6314 |
| D0545 | A083 | 324 | 324 | 36 | F6366 |
| D0545 | A083 | 324 | 324 | 27 | F6345 |
| D0545 | A083 | 324 | 324 | 36 | F6346 |
| D0545 | A083 | 324 | 324 | 33 | F6308 |
| D0545 | S294 | 340 | 508 | 1 | E1088 |
| D0545 | B1940 | 525 | 547 | -33 | F6308 |
| D0545 | B6104 | 550 | 573 | -36 | F6346 |
| D0545 | B8926 | 577 | 585 | -1 | E1088 |
| D0545 | B9072 | 587 | 610 | -36 | F6366 |
| D0545 | B6103 | 612 | 633 | -27 | F6345 |

第545号快递员调度计划如下：该快递员08:00（到达时长=08:00-08:00=0，服务时长=0，离开时长=到达时长+服务时长=0+0=0）从网点A083开始取货，分别取订单编号为F6344, F6360, F6358, F6353, F6354的包裹（此时快递员携带包裹量为139<140），然后分别去订单编号为F6344, F6354, F6353, F6358, F6360的配送点B5800, B7555, B7182, B8307, B8461送货，服务时长（满足公式（2））分别为22（3sqrt(34)+5）, 29 (3sqrt(63)+5), 15, 18, 15，到达时长=上一地点的离开时长+上一地点与该地点的距离/速度，分别为7（08:07-08:00）, 30（08:30-08:00）, 62, 79, 102，由于配送点配送的都是电商订单，所以，离开时长=到达时长+服务时长，分别为29（7+22）, 59（29+30）, 77, 97, 117（目前，快递员携带包裹量为0）。类似，快递员去网点A083取订单编号为F6349, F6325, F6314的包裹（快递员携带包裹量138<140，服务时长=0），然后去配送点B6528配送订单编号为F6349的包裹，到达时间与离开时间计算方式同上，而后，快递员去店铺S245取O2O订单编号为E0895的包裹，到达时长=上一地点离开时长+上一地点与该地点的距离/速度=157+3=160，离开时长=max（到达时长，到商户的领取时长=到商户的领取时间-08:00）=max（160, 257=12:17-08:00）=257，接着快递员去配送点B3266配送O2O单E0895，到达时长=257+2=259，离开时长=到达时长+服务时长=259+8=267.后续配送计划计算逻辑如上对照，不做详述。

参赛选手提交的结果需要校验的所有逻辑如下，只要满足任一条件，即被判定为无效解，无效解将不在排行榜展示：  
1. 解文件读写异常（正常保存为csv文件即可）。  
2. 每个节点的到达时间早于上一节点的离开时间。  
3. 每个节点的离开时间早于该节点的到达时间。  
4. 同一笔订单，被取/送货多次（必须一次性取/送完）。  
5. 任一笔订单取/送货数量、地点与给定数据集不一致。  
6. 快递员达到取/送货点的离开时间早于订单要求的最早取货时间。  
7. 取货订单数量、订单编号与送货订单数量、订单编号不匹配（有取必有送，有送必有取）。  
8. 配送的订单集合必须与给定数据集中要配送的订单不完全匹配，遗漏订单。  
9. 作弊次数超过10次。作弊的判定：服务时间（排除O2O订单取货节点，因为需要等待至最早取货时间）和行驶时间不满足假设7、8。  
10. 快递员任意时刻携带包裹量>140。  
11. 任一笔订单取货时间晚于送货时间。  
注：对于首单为O2O单的快递员，其到达首单店铺时间应>=距离该店铺最近的网点距离/快递员速度,否则，差异时间会计入10倍惩罚

**小小攻略**

天池大多数的比赛都是ml类问题，但菜鸟最后一公里配送却是OR类问题，甚至是规模不小的OR问题，这可能对目前多数游走在天池的ML选手入门产生一定的困难，极有可能让多数选手无从下手。楼主在此作为一个非专业的入门（可能都算不上）级选手，给无从下手的亲们贡献点idea，起码做到可以下手，可以挤进排行榜哈~

**解题思路**  
首先，需要定位这是什么问题？显然是典型的VRP（Vehicle Routing Problem）问题，维基上关于VRP问题的介绍，https://en.wikipedia.org/wiki/Vehicle\_routing\_problem，在网上还可以搜到很多比较成熟的参考资料以及常用求解工具。

下一步，就需要对实际问题进行建模，搞清楚决策函数及约束条件。题目要求快递员总体配送时间最短，即为决策函数。对于约束条件，VRP中常见的约束无外乎最大容量约束（CVRP）、时间窗口约束（VRPTW）、先pickup后delivery约束（VRPPD），该题目全包含了。题目中描述所有电商单的配送需要到网点取后才能送，所有的O2O单要到店铺取后才能送，这是先pickup后delivery约束；题目中要求O2O单的取送货时间要满足要求的最早取货时间和最晚配送时间，这是时间窗口约束；题目中要求快递员任意时刻携带的包裹量不能超过140，这是最大容量约束。至此，该问题就建模完成，宝宝就可以很开心的调用求解工具或者自己coding采用贪婪的方式去求解问题了。

但是，比较popular的工具对问题规模都是有限制的，超过限制性能就会出现问题，比如求解时间会成倍增加。楼主了解到，目前上百个点的规划基本是不成问题的，速度很快，但最大规模也仅限于几百个点，像最后一公里中上万点的规模是相当大的，一般的工具应该是hold不住。我们只能退而求其次，由最优转为次优，将问题规模缩小，分治求解，采用什么样的分治方法，大家可以多尝试，比如配送点聚类后分给各小件员，问题规模可以一下缩很小，可行解很快就可以出来。但有时分配不合理就会出现找不到可行解的情况，仔细阅读一下题目，时间窗口约束较弹性，可以不满足，加惩罚就ok了，所以，加上松弛变量就大致ok了，甚至可以将约束去掉~，大不了cost大点，也是有效的解。

至此，一个还过得去的解就出炉啦~

**提交成绩**  
根据目前为止楼主较长一段时间的测评发现，大家提交的无效解存在的问题都很类似，楼主在此总结下，还望大家在发现成绩无的时候不至于惊慌失措。存在问题：  
一，快递员到同一地点取/送多笔订单时，未分开计算到达时间与离开时间。大家往往会忽视一点，快递员到达配送点，取/送多笔订单时，不同的订单还是要分开处理的。举例说明，快递员D0001在11点到达点B0001，配送2笔订单F0001,F0002，包裹量分别是10,20，则配送计划需要写成：

D0001,B0001,180,194,-10,F0001

D0001,B0001,194,212,-20,F0002

或者

D0001,B0001,180,198,-20,F0002

D0001,B0001,198,212,-10,F0001

二，两点间行驶时间（当前点到达时间-上一点离开时间）计算不正确。奇怪的是，出现这种情况往往不是整体出现，而是局部，所以，楼主也百思不得其解，为什么一些算对了，一些又不对，还望亲们好好检查下，不行就写个代码，将时间整体调整下。记得一定要采用四舍五入的方式哦~，是round不是int！突然想起来了，可能大家的行驶时间中还加入了等待时间，请大家谨记，等待时间只有可能发生在店铺取订单时，其他任何时候的等待都是非法等待，是要计入作弊次数的哈（累计10次以上就无效成绩了哈！）

三，订单的服务时间（离开时间-到达时间）计算不正确。对于网点来说，服务时间一定是0的哈，电商单的取货点是没有服务时间的，否则成绩直接无效哈；对于店铺来说，服务时间的有无决定于到达店铺的时间与商家要求的最早时间，如果早于要求的最早时间到达，是需要等待至最早取货时间才可以出发的哈，否则成绩直接无效，而不是惩罚哈；对于配送点，服务时间是严格遵守公式(2)的哈，记得同样是四舍五入（任何时候精确方式都采用四舍五入），计算方式应该是round(3\*sqrt(x)+5)。

四，其他直接判定无效的条件还请大家仔细阅读赛题底部说明哈，实在不行就自己coding下评测程序，不过后续我们评测程序会开发出来，大家稍安勿躁哈~

结语：大家可以在群里或论坛里多交流下解题思路，是骡是马都拿出来溜溜，谁跑的快还不一定呢哈~，毕竟这样排行榜的迭代速度会快点，高出不胜寒啊，前三甲坐久了也是会寂寞的哈~