

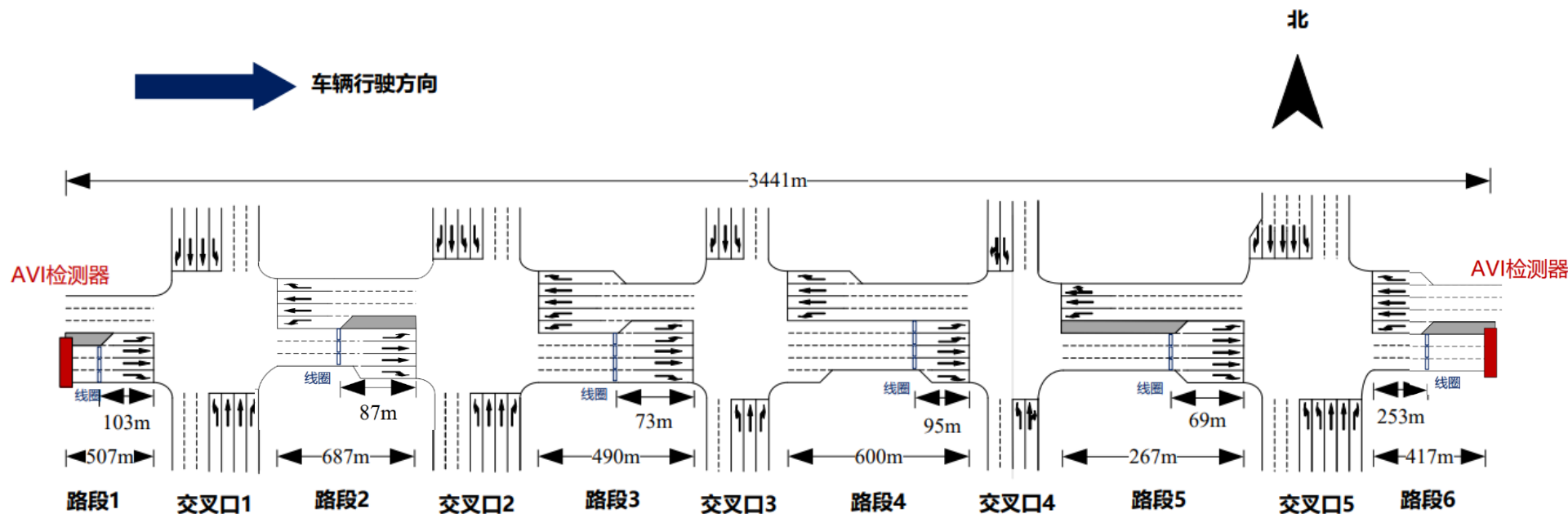
数据来源：连云港市某干道的仿真数据（仿真时长为14小时），数据采集的方向从西向东

2021年交通数据分析 期末汇报 ▶

2018级
赵冠华 综合交通信息与控制工程系
1852127

数据采集

5min 168条数据



基于采集数据、干道示意图、交叉口流量与信号配时方案完成任务

TASK 1

计算干线平均行程速度



计算权重

路段编号	路段1	路段2	路段3	路段4	路段5	路段6
长度/m	507	687	490	600	267	417
总长/m	2968					
权重	0.1708	0.2315	0.1651	0.2022	0.0900	0.1405



加权平均

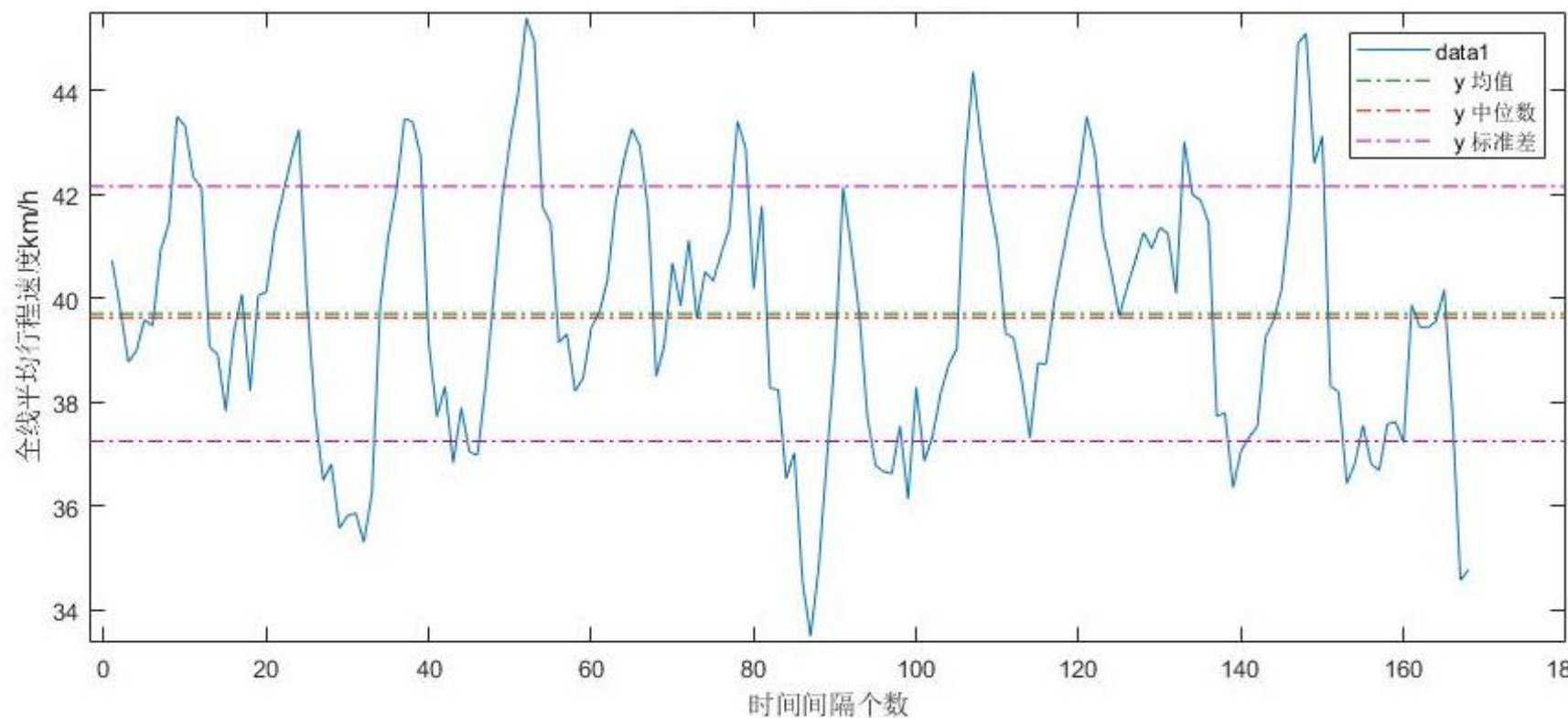
$$v_{\text{干线}} = \sum_{i=1}^6 v_i * w_i$$

- 借助MATLAB工具以路段长度加权的方计算168个5分钟时间间隔的整条干线的平均行程速度。

TASK 1

基本统计分析与可视化

样本数	最大值	最小值	平均值	中列数	中位数	标准差	变异系数
186	45.3881	33.4947	39.7015	39.4414	39.6295	2.4506	0.0617



利用histc与cumsum函数绘制直方图与累积分布频率曲线

TASK 2

剔除异常数据

① 独立判断法

② 三倍标准差法 (分路段)

1	时间段(min)	平均速度(km/h)						流量(Veh)						占有率						
92	445	450	38.5	51.9	44.8	34.3	38	56.9	130	128	161	113	143	154	0.196	0.046	0.062	0.073	0.243	0.049
93	450	455	36	48.8	51.2	0000	52.7	56.8	141	124	130	208	116	117	0.186	0.23	0.038	0.111	0.043	0.038
94	455	460	29.2	47.5	53.3	37.8	47.5	56.6	140	151	126	128	158	149	0.311	0.062	0.033	0.07	0.07	0.048
95	460	465	26.8	53.7	51.2	36.8	48.5	57	140	130	144	134	118	146	0.408	0.048	0.043	0.083	0.053	0.048
96	465	470	34.6	46.5	45.2	36.4	43.9	56.5	134	127	144	137	125	113	0.228	0.054	0.099	0.088	0.058	0.035
97	470	475	39.6	47	43.1	39.1	35.8	0000	168	143	117	113	133	195	0.359	0.059	0.231	0.068	0.092	0.041
98	475	480	39.9	46.7	41.4	38.7	37.4	56.6	144	170	152	130	98	131	0.225	0.29	0.115	0.08	0.082	0.043
99	480	485	59.7	44.7	37	35.8	36.1	56.8	104	100	133	128	136	111	0.061	0.12	0.239	0.1	0.093	0.034
100	485	490	46.2	51.2	36.4	38.5		57.1	113	104	102	87	109	99	0.053	0.038	0.09	0.059	0.118	0.031
101	490	495	42.2	44.8	40.5	36.9	0000	56.7	114	128	120	114	188	141	0.064	0.055	0.121	0.077	0.113	0.045
102	495	500	47.7		39.6	39.2	0000		115	117	139	142	195	121	0.052	0.044	0.145	0.112	0.211	0.039
103	500	505	45	50.9	38.3	36.1	39.1		99	90	111	96	106	121	0.077	0.032	0.07	0.073	0.163	0.039
104	505	510	39.8	45.9	43.9	34	37.7	56.8	134	127	109	119	107	140	0.115	0.18	0.079	0.083	0.217	0.047
105	510	515	38.8	49.9	42	38.3	38.2	56.4	143	157	132	133	151	35	0.162	0.06	0.072	0.103	0.271	0.035
106	515	520	37.1	47.5	42.8	36.7	36.4	56.6	117	114	163	130	116	117	0.147	0.045	0.066	0.088	0.323	0.065
107	520	525	38.2	45.2	49.9	34.9	37.2	56.6	133	116	127	145	141	156	0.11	0.052	0.038	0.092	0.283	0.05
108	525	530	0000	53.2	54	42.8	40	57.1	198	136	105	104	128	127	0.089	0.048	0.027	0.044	0.198	0.042
109	530	535	25.8	50.7	55.8	42.7	48.8	57	125	114	145	113	97	99	0.08	0.041		0.047	0.039	0.031
110	535	540	39.6	51.5	57.8	0000	51.8	0000	120	110	125	205	119	130	0.091	0.04	0.042	0.047	0.043	0.042
111	540	545	44	51.1	42.3	40.8	51.8	0000	114	118	103	91	109	138	0.067	0.043	0.048	0.043	0.041	0.043
112	545	550	42.9	48.2	40.1	42.7	54.1	57.1	108	105	113	112	84	83	0.053	0.043	0.137	0.068	0.03	0.027
113	550	555	45.4	52.5	36.5	32.6	50.6	56.7	92	92	108	103	116	86	0.061	0.032	0.259	0.092	0.043	0.027
114	555	560	48.9	51.4	36.1	0000	44	56.4	96	80	78	70	75	110	0.038	0.032	0.134	0.035	0.035	0.035
115	560	565	44.7	46.6	41.5	34.7	40.8	55.9	108	113	89	93	77	76	0.066	0.046	0.093	0.057	0.046	0.025
116	565	570	43.3	60.8	36.6	36.2	36	56.9	108	95	116	98	112	88	0.093		0.177	0.072	0.085	0.027
117	570	575		51.8	37.6	38	37.9	56.6	93	96	105		110	125	0.05	0.034	0.082	0.05	0.09	0.041
118	575	580		49.2	36.7	32.9	38.2	56.9	91	104	77	116	101	111	0.036	0.039	0.056	0.118	0.128	0.034
119	580	585	50.6	50.4	39.6	36.8	33.2	57	96	89	105	78	101	78	0.043	0.033	0.053	0.049	0.188	0.025
120	585	590	49.1	49.8	0000	42.7	0000	56.5	98	73	100	88	78	111	0.039	0.028	0.036	0.041		0.067
121	590	595	44.7	45.4	0000	34.4	0000	56.5	103	91	86	97	96	91	0.05	0.04	0.022	0.053	0.165	0.029
122	595	600	52.3	48.5	55.9	40.8	31.7	57.4		50	98	78	87	70	0.03	0.04	0.028	0.035	0.211	0.023

路段 线圈ID 速度(km/h)

编写Matlab脚本，运行程序；最终得到25组异常数据

F: 修正系数，通常为1.3~1.5，这里取1.5

由于是多车道集成监测站，因此这里占有率的最大值为80%

TASK 2

25 + 90

补全缺失数据

✓ 加权平均法

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	时间段(min)		平均速度(km/h)						流量(Veh)						占有率					
23	100	105	42.6	45.3	46.2	38.1	43.5	56.8	120	110	155	101	106	145	0.112	0.047	0.053	0.055	0.06	0.045
24	105	110	41.7	49.5	52.9	36.2	40	56.5	124	103	130.8	131	115.86667	113	0.08	0.039	0.027	0.085	0.067	0.035
25	110	115	43.6	52.4	55.9	40.9	33.3	56.3	130	132	102	100	111	103	0.093	0.27	0.026	0.044	0.096	0.032
26	115	120	41.9	51.8	53.3	46.6	35.9	56.6	125	121	141	103	104	139	0.067	0.042	0.036	0.032	0.117	0.044
27	120	125	38.6	49	43.4	40.6	36.2	56.3	135	117	133	124	109	99	0.114	0.046	0.148	0.058	0.09	0.031
28	125	130	39.4	49.9	39.1	39	33.9	56.4	166	137	112	112	133	121	0.191	0.051	0.141	0.064	0.219	0.038
29	130	135	37.3	50.4	41.3	41.013333	35.24	56.6	137	159	151	134	124	137	0.172	0.061	0.16	0.09	0.172	0.045
30	135	140	38.4	51.3	38.1	36	35.033333	56.8	136	121	138	149	139	136	0.1462	0.045	0.249	0.112	0.236	0.043
31	140	145	39.9	50.9	40.7	37.5	36.6	56.8	143	133	134	115	125	130	0.129	0.048	0.183	0.07	0.141	0.041
32	145	150	39.7	49.5	44.8	37.5	39.3	56.6	141	164	161	128.06667	142	146	0.146	0.061	0.092	0.088	0.256	0.046
33	150	155	28.1	50.5	40.4	38.1	33.7	56.4	135	120	139	142	132	145	0.56	0.045	0.089	0.101	0.2076667	0.045
34	155	160	35.2	50.4	41.2	38.3	37.6	56.9	131	125	143	127	129	132	0.23	0.047	0.082	0.086	0.337	0.042
35	160	165	37.8	50.373333	43.9	35	37.3	56.5	159	192	139	145	137	140	0.304	0.054	0.062	0.1	0.285	0.044
36	165	170	37.4	49.6	43.5	37.1	44.4	56.5	136	174	148	149	153	162	0.18	0.065	0.24	0.09	0.17	0.053
37	170	175	50	51.2	46.1	35.7	48.2	56.7	144	122	169	130	114	125	0.109	0.045	0.058	0.08	0.048	0.039
38	175	180	38.4	51.2	49.5	35.7	48.7	56.9	152	121	123	149	132	133	0.167	0.044	0.038	0.089	0.056	0.042
39	180	185	44.2	48.2	51.4	41.7	52.8	57.2	136	161	126	107	130	140	0.106	0.065	0.035	0.049	0.049	0.046
40	185	190	41.9	51	53.1	39.9	48.4	56.4	113	119	147	125	101	112	0.105	0.045	0.038	0.051	0.04	0.036
41	190	195	42.593333	48.9	48.3	43.1	42.1	57.1	112	94	103	105	122	94	0.091	0.037	0.081	0.064	0.063	0.031
42	195	200	41.8	49.78	50.213333	38.5	38.7	57.2	141	130	104	96	101	109	0.48	0.058	0.116	0.054	0.063	0.034
43	200	205	39.9	47.2	50.391111	36.4	36.7	57.1	110	142	140	110	95	111	0.124	0.06	0.168	0.072	0.087	0.036
44	205	210	41.2	52	36	37.4	36.7	57.1	118	102	124	107	128	106	0.119	0.035	0.218	0.085	0.156	0.034
45	210	215	45.9	51	37.3	38.28	36.2	56.6	113	96	110	98	102	111	0.054	0.035	0.29	0.062	0.125	0.034

采用加权移动平均法与相邻检测器法补全缺失的115组数据

TASK 3

聚类分析

168
组数
据

三个
变量

划分
为3-5
类

异常
数据
已被
剔除

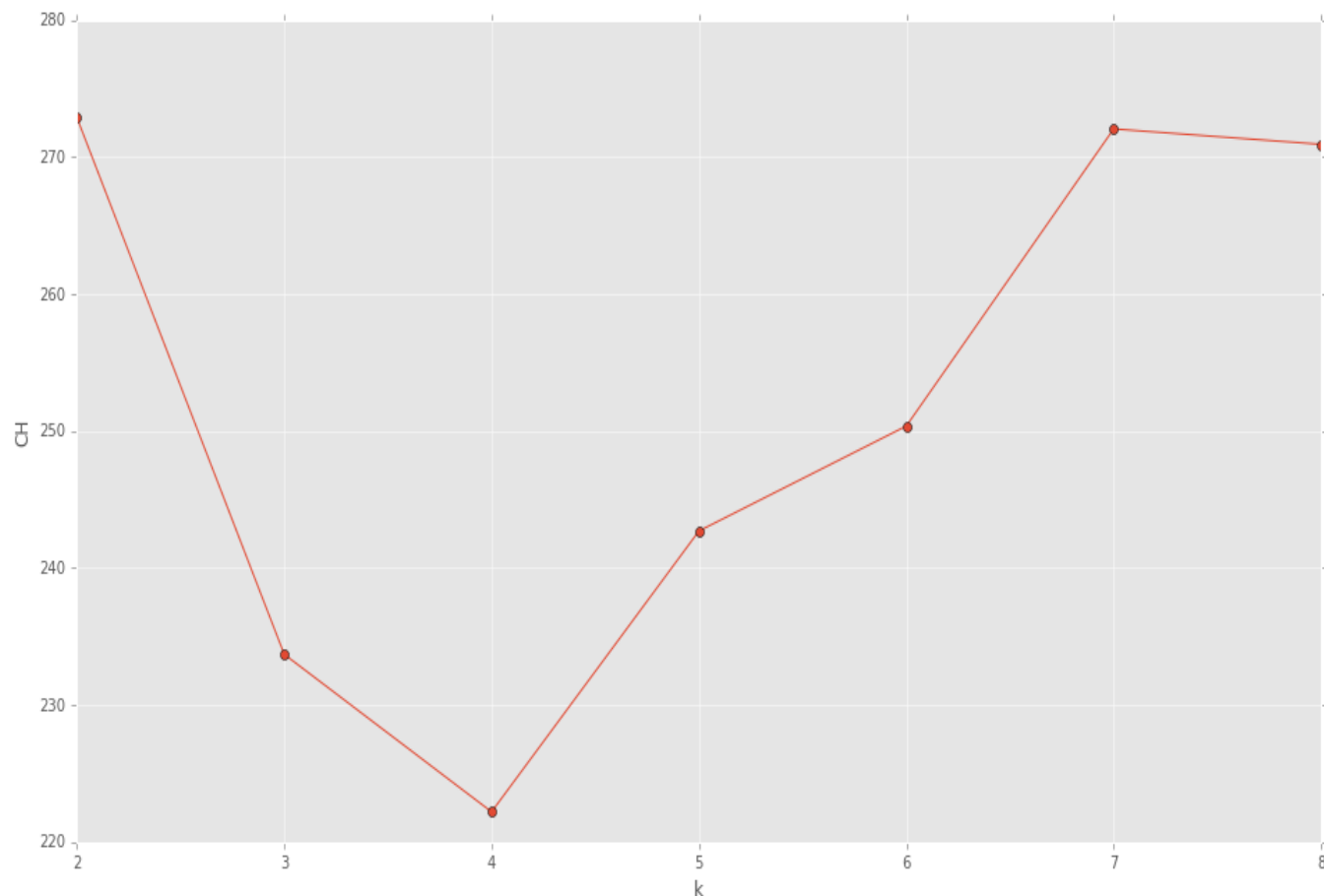
K-Means

选择K-means
算法进行聚类，
由C-H指数确
定聚类簇数为
2或7。

Calinski-Harabaz 通过评估类之间的方差与类内方差来计算得分。

TASK 3

聚类分析



Calinski-Harabaz 通过评估类之间的方差与类内方差来计算得分。

K-Means

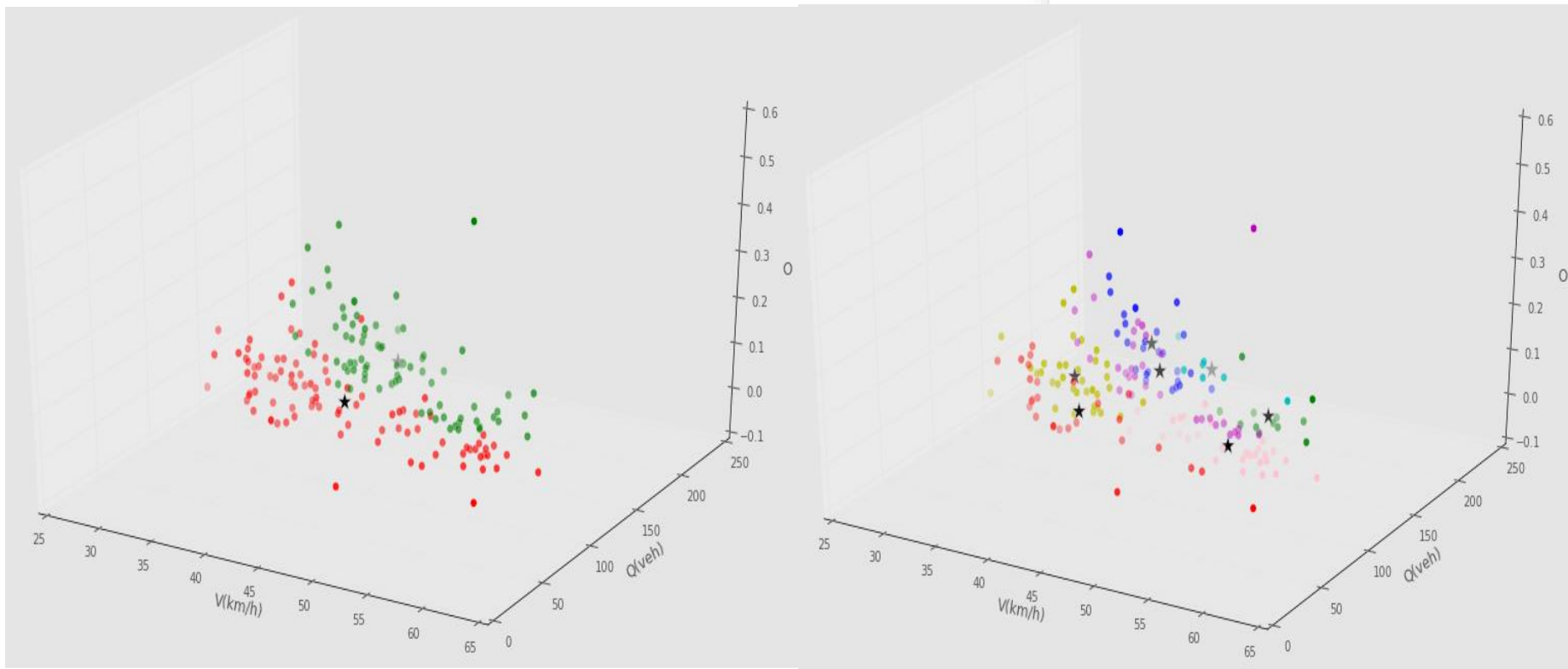
选择K-means
算法进行聚类，
由C-H指数确
定聚类簇数为
2或7。

TASK 3

聚类分析

K-Means

```
In [115]: print (sil_score)  
0.4143491628647458
```



类间间距/类内间距得分：2-2.32； 7-3.78



使用sk-learn计算7簇轮廓系数=0.414，聚类质量尚可

轮廓系数 $\in [-1, 1]$ ，越趋于1说明聚类质量越好

TASK 3

聚类分析

K-Means

簇内相似性

v:2~3(km/h)

q:10/20(veh)

O:0.5%~4%

类簇	0	1	2	3	4	5	6
样本数	26	29	35	13	13	26	26
平均速度 km/h							
类簇	0	1	2	3	4	5	6
最大值	61.30	44.70	43.90	57.60	52.70	59.90	57.80
最小值	43.90	37.00	29.50	43.50	39.15	32.80	36.00
均值	52.34	40.54	37.75	50.22	43.98	40.66	47.85
方差	20.11	3.49	8.34	23.25	10.93	45.23	40.52
流量 veh							
类簇	0	1	2	3	4	5	6
最大值	110.00	145.00	116.00	148.00	197.00	91.00	130.80
最小值	87.00	128.00	93.00	132.00	151.00	32.00	112.00
均值	99.87	136.50	105.06	141.38	161.69	78.92	121.84
方差	38.30	20.14	40.57	28.09	156.06	230.71	26.31
占有率							
类簇	0	1	2	3	4	5	6
最大值	0.116	0.38	0.29	0.24	0.163	0.186	0.48
最小值	0.021	0.052	0.029	0.035	0.049	0.021	0.027
均值	0.042	0.156	0.109	0.080	0.089	0.086	0.093
方差	0.001	0.007	0.003	0.004	0.001	0.003	0.010

使用sk-learn计算7簇轮廓系数=0.414，聚类质量尚可

轮廓系数 $\in [-1,1]$ ，越趋于1说明聚类质量越好

TASK 4

相异性分析

① 欧几里得距离

$$\sqrt{(X_{I1} - X_{J1})^2 + \dots + (X_{Ip} - X_{Jp})^2}$$

② DTW距离

$$DTW(Q, C) = \min \sqrt{\sum_{k=1}^K w_k / K}$$

路段	路段3		路段6	
数据来源	线圈数据	浮动车数据	线圈数据	浮动车数据
欧几里得距离均值	13.781	15.316	0.464	0.790
DTW距离	1406.861	564.300	29.767	59.700

- ① 路段6的估计速度比路段3更接近真实值;
- ② 路段6线圈比浮动车估计更为准确
- ③ 路段3两种方法结果不同。

TASK 4

相关性分析

① 协方差

$$COV(X, Y) = E(X, Y) - E(X) * E(Y)$$

② 相关系数

$$\rho(X, Y) = \frac{COV(X, Y)}{\sqrt{D(X) * D(Y)}}$$

路段	路段3		路段6	
数据来源	线圈数据	浮动车数据	线圈数据	浮动车数据
协方差	49.307	70.247	0.045	0.050
相关系数	0.755	0.830	0.623	0.330

- ① 路段3的同相度与相关性高于路段6;
- ② 路段3浮动车相关性更高;
- ③ 路段6线圈数据相关性更高。

TASK 5

行程速度估计误差计算

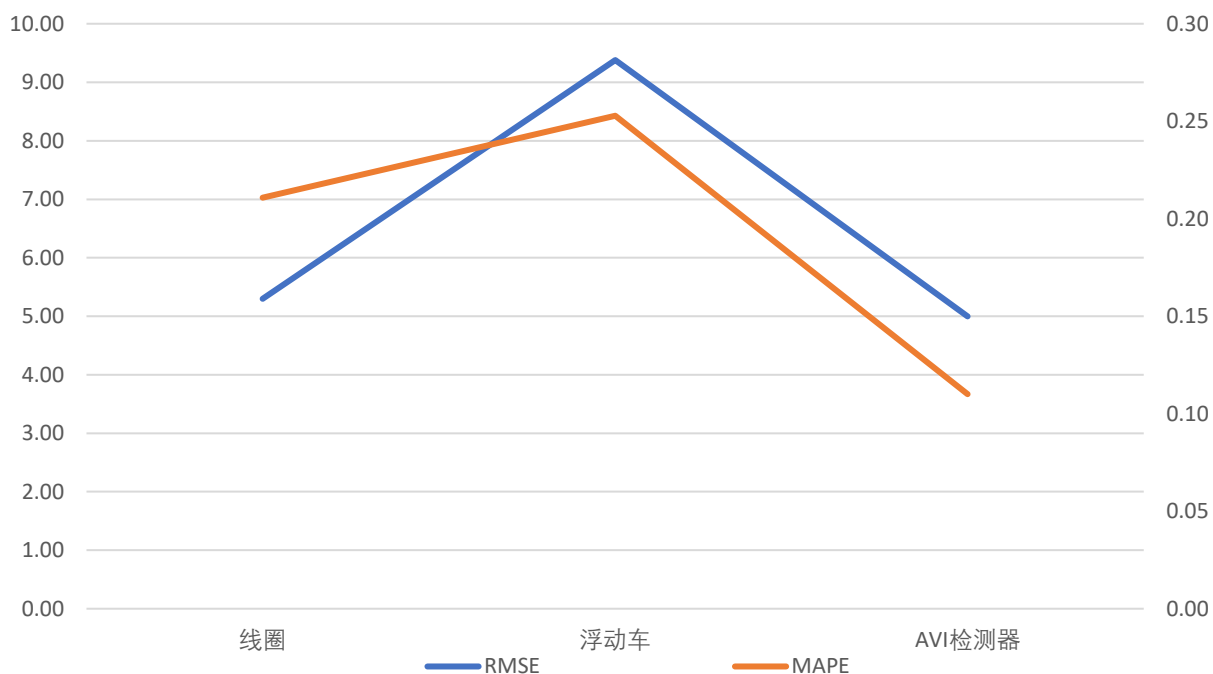
MAPE(平均绝对百分比误差)、RMSE(均方根误差)
误差越小代表预测结果越好

① 平均绝对百分比误差

$$MAPE = \frac{\sum_{i=1}^N \frac{|x_i - x_i^-|}{x_i}}{N}$$

② 均方根误差

$$RMSE = \sqrt{\frac{\sum_{i=1}^N |x_i - x_i^-|^2}{N}}$$

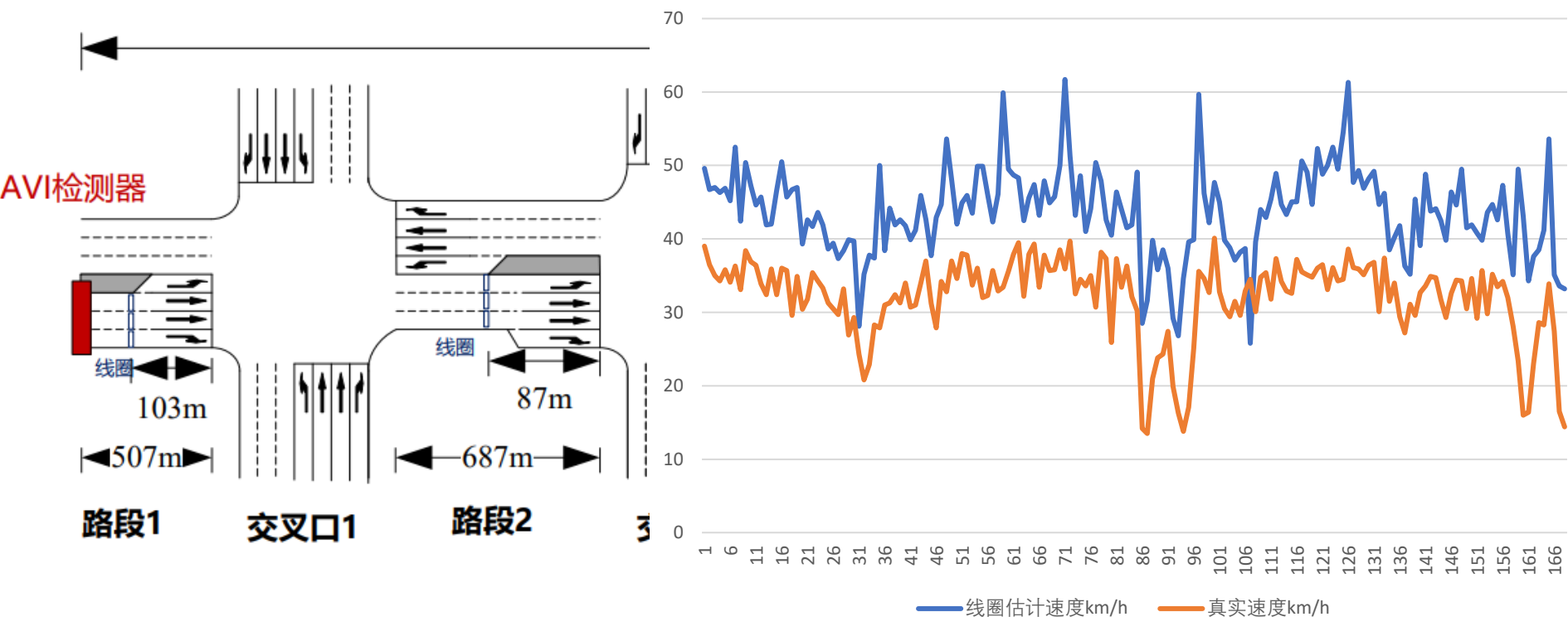


估计准确性排序：AVI > 线圈 > 浮动车

TASK 6

交通数据融合

线圈埋设位置距离交叉口仍有一定距离



线圈估计速度偏大，需要结合交叉口信控延误优化处理

TASK 6

交叉口信控延误

数据采集的行车方向为从西到东，所以这里的延误只计算西进口道车道的延误

交叉口	1	2	3	4	5
直行车辆流量	1185	1214	1368	1315	1361
直行车辆延误	29.56	28.00	29.01	29.06	29.59
左转车辆流量	175	139	292	215	147
左转车辆延误	58.10	58.13	56.44	56.41	57.46
右转车辆流量	121	50	103	76	57
右转车辆延误	0	0	0	0	0
西进口道总流量	1481	1403	1763	1606	1565
西进口道总延误	45195.60	42074.61	56172.44	50347.25	48721.49
西进口道平均延误	30.52	29.99	31.86	31.35	31.13

流量单位: pcu/h; 延误单位: s/pcu



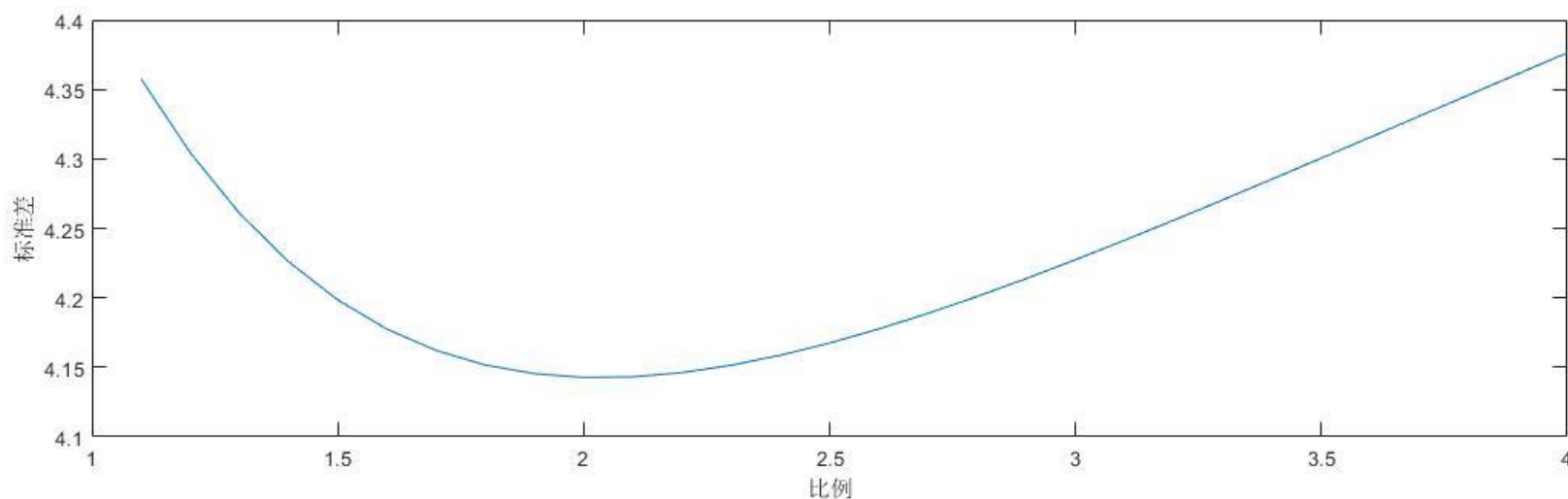
对每个交叉口各个方向车道延误加权平均，得到最终延误

TASK 6

信控延误影响

只采用部分信控延误

$$\cdot \text{新线圈速度} = \frac{\text{该路段全长}}{\frac{\text{对应交叉口信控延误}}{\text{某个大于1的值}} + \frac{\text{线圈前路段长度}}{\text{原线圈速度}}}$$



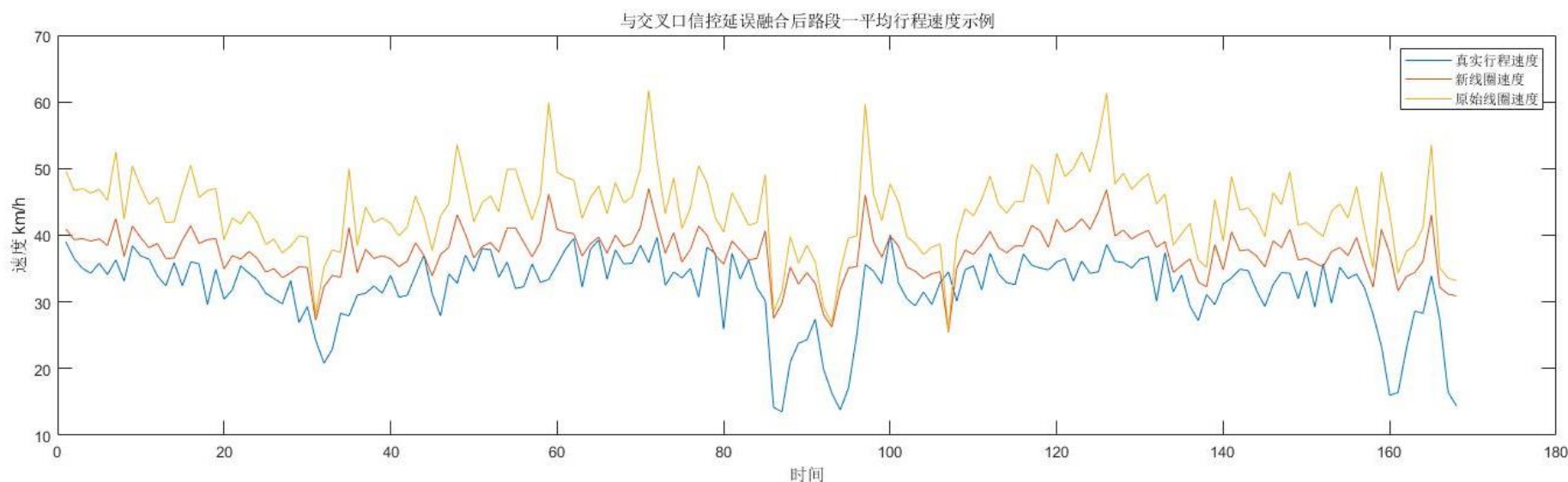
确定比例得到经信控延误优化后的新线圈速度

TASK 6

信控延误影响

只采用部分信控延误

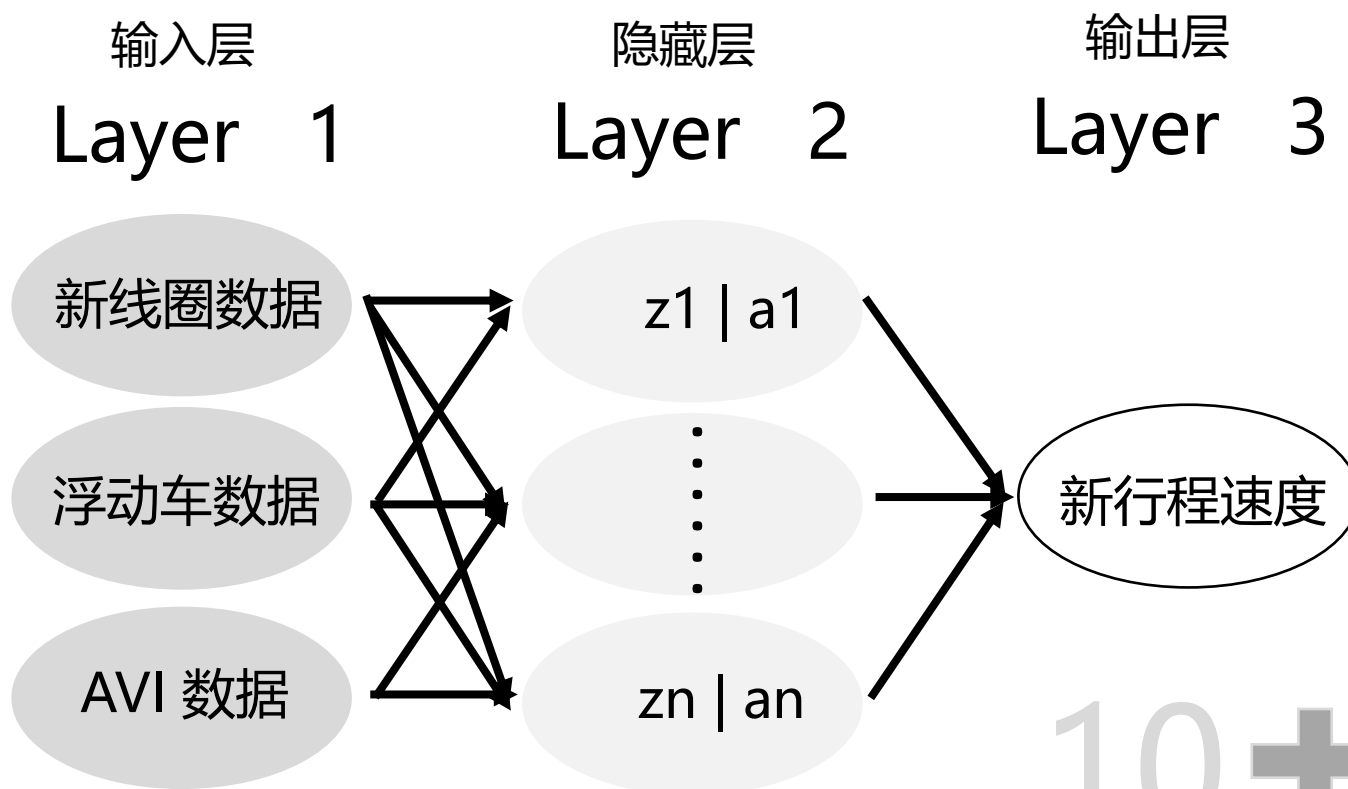
$$\cdot \text{新线圈速度} = \frac{\text{该路段全长}}{\frac{\text{对应交叉口信控延误}}{\text{某个大于1的值}} + \frac{\text{线圈前路段长度}}{\text{原线圈速度}}}$$



确定比例得到经信控延误优化后的新线圈速度

TASK 6

数据融合网络



采用BP反向传播算法进行数据融合

10 + 4

TASK 6

数据归一化处理

Z-Score

基于数据的均值与方差；
适用于高斯分布型的数据。

进行一次线性变换，映射数据到 $[0,1]$
适用于数据较为零散或者线性

MAX-MIN

采用MAX-MIN方式对数据进行归一化处理

TASK 6

确定隐含层个数

比较收敛速度与训练误差

$$n_1 = \sqrt{(n + m)} + a$$

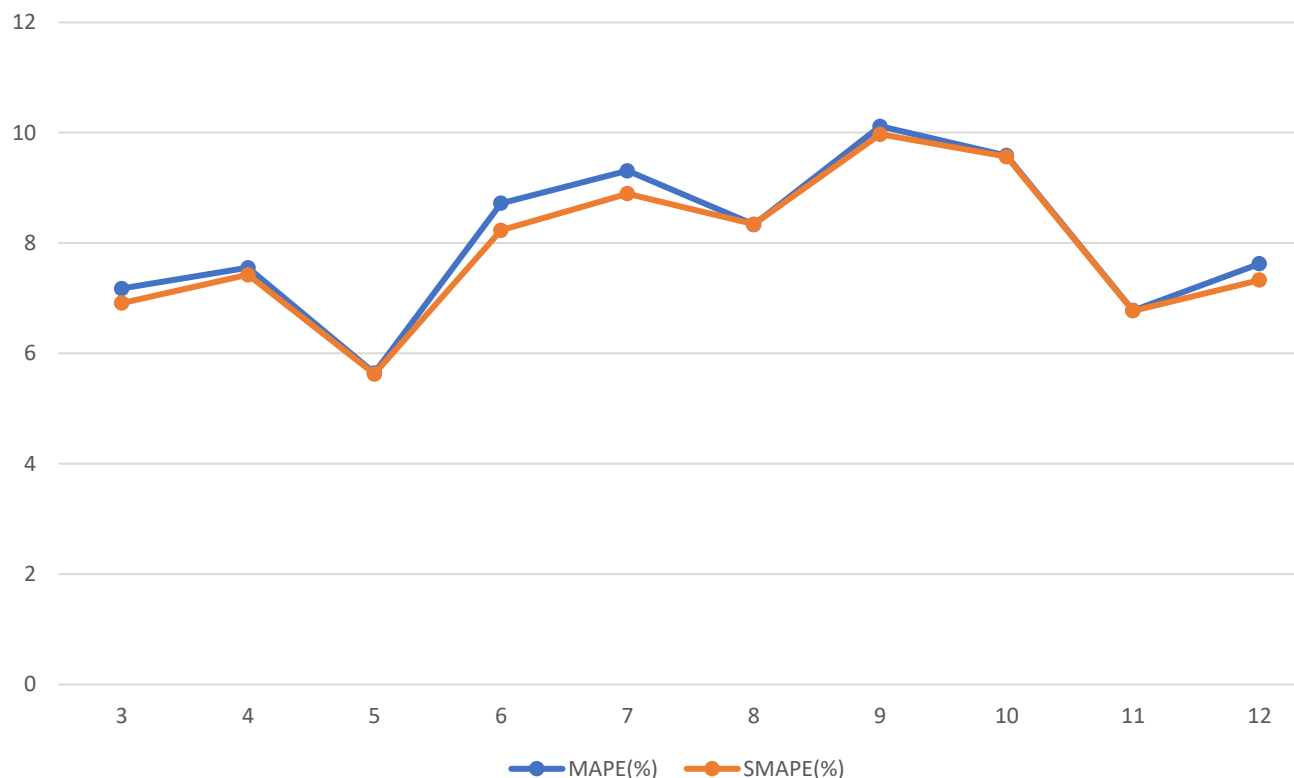
n : 输入单元数;

m : 输出单元数;

n_1 : 隐含层单元数;

$a \in [1, 10]$

$$n_1 \in [3, 12]$$



当隐含层个数为5时，BP网络效果最佳

TASK 6

确定BP训练参数

迭代次数

Epochs

```
net.trainParam.epochs=1000;  
net.trainParam.goal=1e-3;  
net.trainParam.lr=0.1;
```

学习率

Lr

目标值

Goal

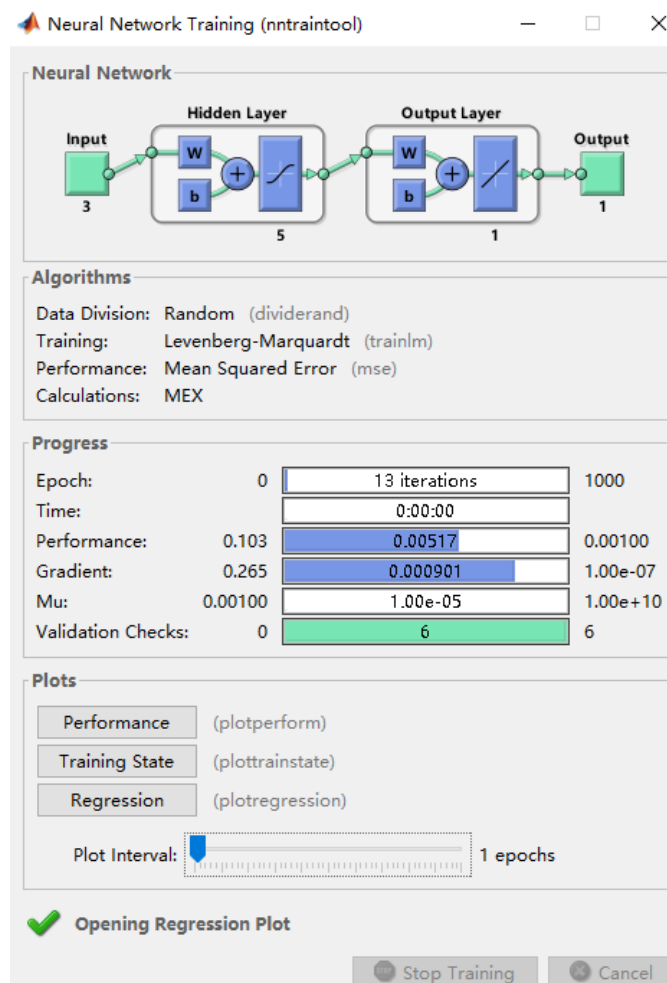
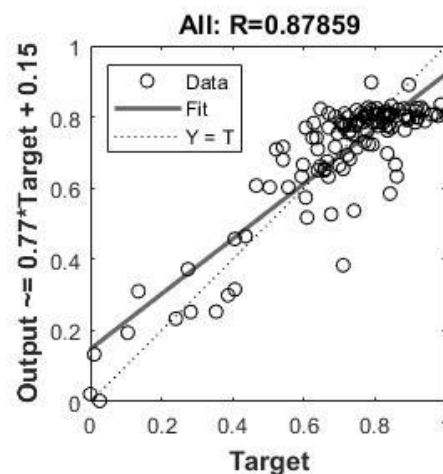
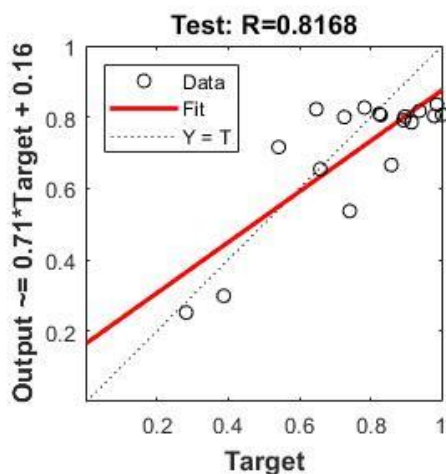
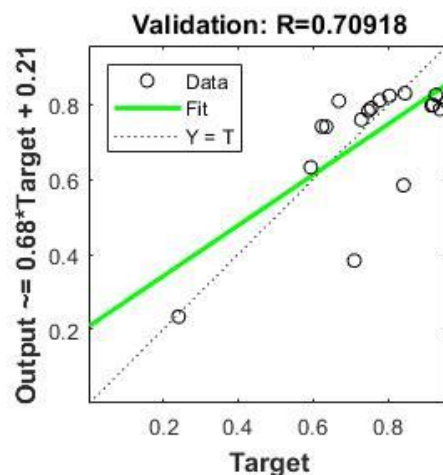
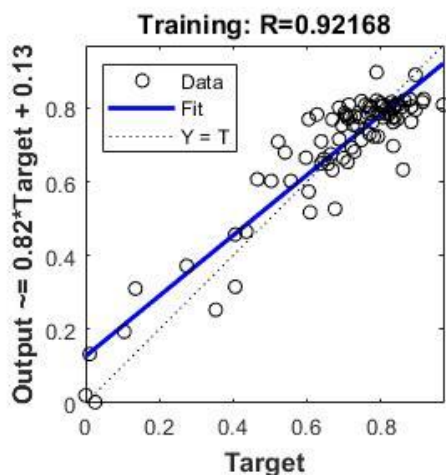
学习率越大：
输出误差对参数的影响越大；
参数更新越快；
受异常值影响越大，
很容易发散

TASK 6

BP网络训练结果

数据集返回13次后达到最优

以路段一数据融合为例



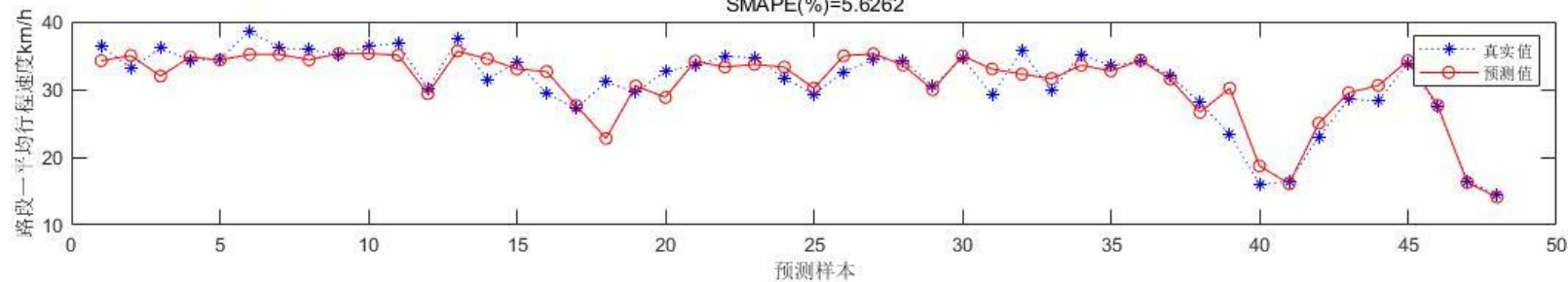
TASK 6

模型拟合度评价

测试集路段一平均行程速度预测结果对比

MAPE(%)=5.643

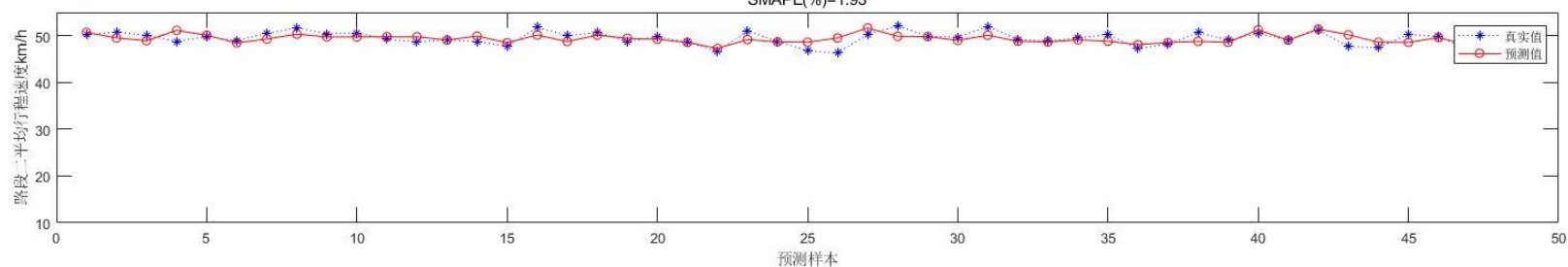
SMAPE(%)=5.6262



测试集路段二平均行程速度预测结果对比

MAPE(%)=1.9322

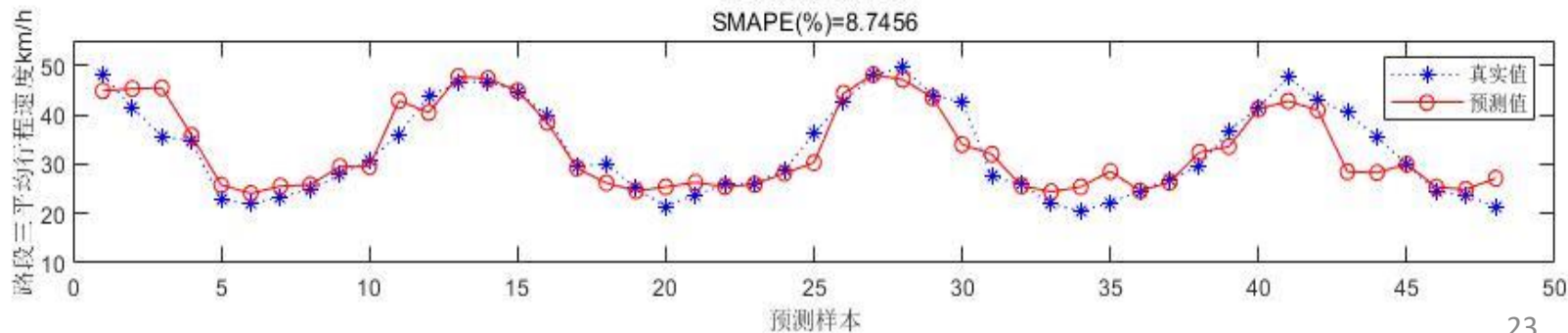
SMAPE(%)=1.93



测试集路段三平均行程速度预测结果对比

MAPE(%)=8.8956

SMAPE(%)=8.7456



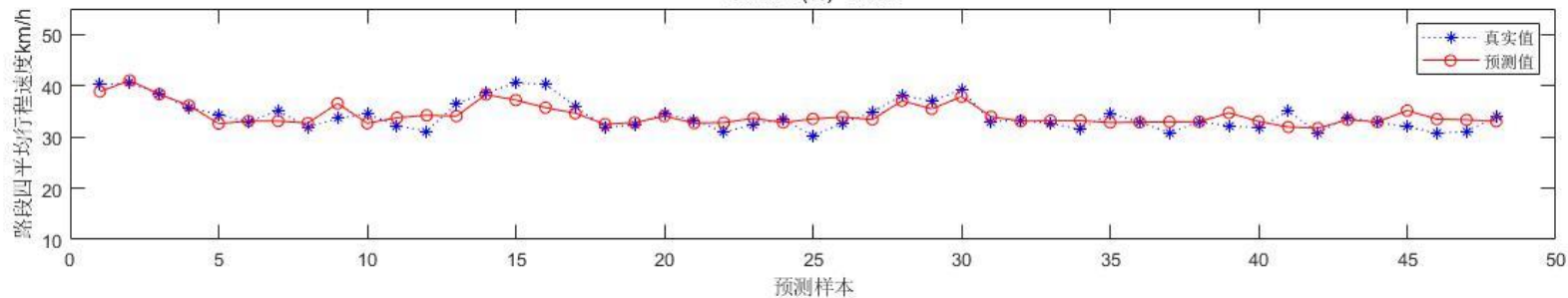
TASK 6

模型拟合度评价

测试集路段四平均行程速度预测结果对比

MAPE(%)=4.2566

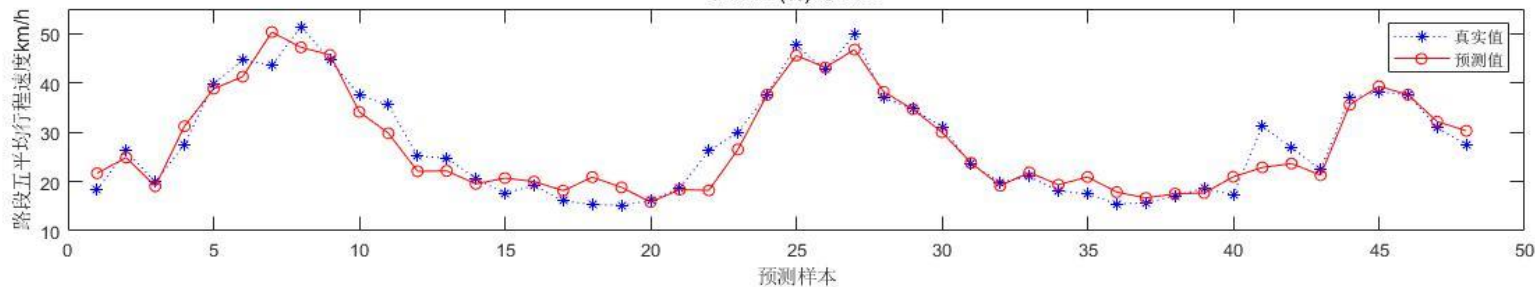
SMAPE(%)=4.234



测试集路段五平均行程速度预测结果对比

MAPE(%)=9.1903

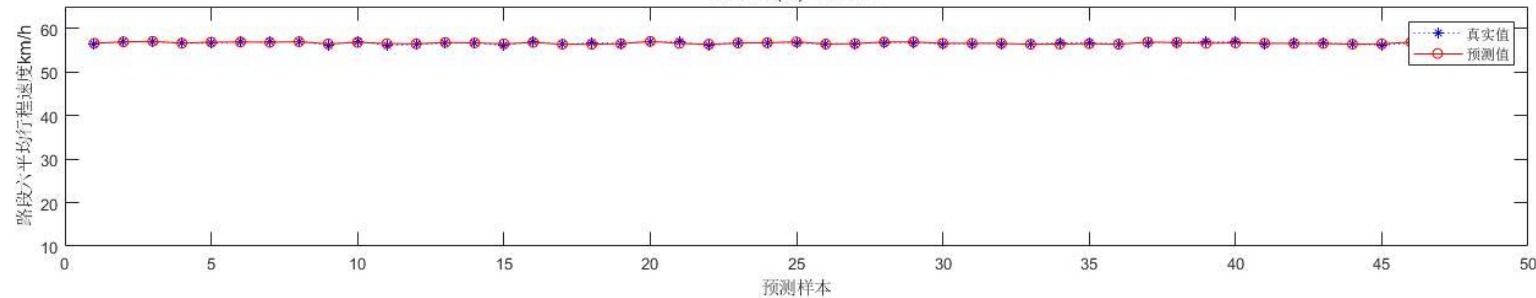
SMAPE(%)=9.1077



测试集路段六平均行程速度预测结果对比

MAPE(%)=0.2598

SMAPE(%)=0.25973



TASK 6

模型拟合度评价

平均MAPE值=5%，拟合程度较理想

路段	1	2	3	4	5	6	平均
MAPE(%)	5.6430	1.9322	8.8956	4.2566	9.1903	0.2598	5.0296
SMAPE(%)	5.6262	1.9300	8.7456	4.2340	9.1077	0.2597	4.9839

