Development of a comprehensive model to check and correct the traffic detector data errors

Hsu, Tien-Pen
Institute of Civil Engineering
National Taiwan University
hsutp@ntu.edu.tw

ABSTRACT

This paper presents a comprehensive model developed for checking the detector data errors and correcting the data errors for freeway traffic flow practically. The detector data per minute are collected by the 606 detectors within a month on the freeway, which have more than 26million minute data taken as example to develop the model. The model found that about 39.06% of the detector data reported in traffic control center have errors. The model developed in this paper includes four stages for checking data errors and one ANN model to correct the error data and to impute the missing data. The four stages to check the data errors are missing data checking, data transmitting errors checking, detector errors checking based on reasonable fundamental relationship of traffic parameters, and data errors based on the relationship of the traffic flow fundamental diagram. Through application of this model, the detector errors can be found out and corrected adequately by programming the model into the traffic data base in traffic control center.

Keywords: Detector, Data error, Neural network, Data imputation

1. INTRODUCTION

Vehicle Detector is the essential equipment to collect the traffic data (FHWA 2006). The traffic data can be used in control and management of traffic facilities. In Taiwan, all the sections on freeway are installed of vehicle detectors. The data collected by detector are transmitted to the traffic control center per minute. The traffic information of congestion level of freeway sections and the control strategy for managing freeway traffic are generated according to the detector data. The efficiency of the data has the crucial effect on traffic management efficiency. To maintain the effective traffic data is then the major task of a traffic management center. In the past, there were a number of studies on imputing the missing data of permanent traffic count. They used the various model to imputing the missing data, such as moving mean, ARIMA autoregressive model, grey prediction model, ANN artificial neural network model (Zhong etc. 2004; Williams etc, 1998). Cheevarounothai etc, (2006) checked the sensitivity problem of detector. Vanajakshi made the diagnostics of detector data based on flow conservation principle (2004). In this paper, the data is not the permanent traffic count data for generating the AADT or hourly traffic volume. The traffic data is the data in minute for traffic control and traffic information service. Therefore, the checking of the missing data and the imputation of missing data should be more effective than other purposes. In this paper, the traffic data will be checked of missing or error first, and then using the imputation model to correct it and impute the data for real time application.

The works of the study include to check the missing data and the data errors of existing data base in traffic control center of all the detectors on freeway and then to find out the data error types, using video film to check the detector traffic data accuracy for determining the effectiveness of detector, and furthermore, using the traffic data to develop the fundamental diagram for checking the flow-speed-density relationship of the traffic data in order to find out the unreasonable values. Thus, in this paper a compressive procedure was structured and applied in practical

2. DETECTOR DATA ERRORS

Usually the detector data errors can be found by "by-eye" method. However, for enhancing the efficiency and for treating huge amount of the data, automatic checking errors procedure should be developed. Furthermore, if the data is not missing but its relationship between the traffic parameters is not reasonable, in such case can be identified only through using the traffic flow model. Therefore, for checking the traffic detector data errors, there are four various errors are identified.

- Data missing
  No data of the time interval, on the data base, the column of time interval shows null. It occurs sometimes also regularly and in several minutes continuously.
- Data errors with unreasonable value of traffic parameters
  The traffic parameters of detector data are traffic flow rate, speed and occupancy. If the occupancy is greater than 100, or one of the parameters is zero but the others are not zero, or same value of a parameter occurred continuously same, it means there are detector data errors.
- Unreasonable relationship between flow rate, speed and occupancy
  Sometimes the flow rate is too bigger than the possible estimated value using

speed and occupancy according to macroscopic traffic flow model. For identifying this type of error, the traffic flow model is obtained by regression of speed and occupancy equation, and if the value locates out of the possible range of statistical confidence interval, it will be seen as the outlier and indentified as possible detection errors. Furthermore, based on flow conservation, the data of two detectors next to each other can be compared in pairs according to estimated density between the pair of detector. If one pair of data is out of bound, the detector data is identified as errors, because the traffic density on a segment should not change exceeding a certain bound [Hsu etc, 2010].

Through using the real detector data of 606 detectors on freeway in Taiwan, the time period is one month and the data are record per minute. All together, the amount of traffic data is 27,051,840 min. data. Totally, the 39.06% of detector data errors are found with missing value or unreasonable value, as illustrated in Table 1. One of the data errors is illustrated in Table 2, in which the occupancy is greater than 100%. The other example is shown in Table 3, in which there are the regular same values.

Table 1 Detector data errors analysis

Error type		Night (24~06)	Peak hour am (06~09)	Non peak hour (09~16)	Peak hour pm (16~19)	Non peak hour (19~24)	total
1	missing value	1,213,338	674,526	1,508,092	757,175	1,392,611	5,545,742
		4.49%	2.49%	5.57%	2.80%	5.15%	20.50%
2	Null value	221,658	109,217	240,819	102,741	170,953	845,388
		0.82%	0.40%	0.89%	0.38%	0.63%	3.13%
3	Q>60	1,412	943	1,462	1,049	1,378	6,244
		0.01%	0.00%	0.01%	0.00%	0.01%	0.02%
4	V>255	0	0	0	0	0	0
		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
5	OCC>100	0	0	0	0	0	0
		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
6	Q=0,OCC≠0 or Q=0,V≠0	5,963	618	483	24	337	7,425
		0.02%	0.00%	0.00%	0.00%	0.00%	0.03%
7	Q≠0, OCC=0 or V≠0，OCC=0	1,238,614	134,951	170,458	30,220	186,004	1,760,247
		4.58%	0.50%	0.63%	0.11%	0.69%	6.51%
8	Same value in several columns	889,223	418,116	464,419	268,175	360,588	2,400,521
		3.29%	1.55%	1.72%	0.99%	1.33%	8.87%
Total		3,568,796	1,337,428	2,384,271	1,158,335	2,110,493	10,559,323
		13.19%	4.94%	8.81%	4.28%	7.80%	39.06%

Table 2 Data error of occupancy L1_OCC value greater than 100%

1	VDID	timeT	L1_S_Q	L1_S_V	L1_S_L	L1_B_Q	L1_B_V	L1_B_L	L1_T_Q	L1_T_V	L1_T_L	L1_Q	L1_V	L1_OCC	
13	VD-N3-S-409.592	2010/2/19 11:57	9	91	43	0	0	0	0	0	85	9	91	245	-23647
20	VD-N3-S-409.592	2010/2/18 05:51	2	100	44	0	0	0	0	0	128	2	100	160	-21488
28	VD-N3-N-371.395	2010/3/5 20:16	5	103	46	0	0	0	0	0	0	5	103	165	165
29	VD-N3-S-409.592	2010/2/18 06:03	2	100	44	0	0	0	0	0	128	2	100	160	-21488
30	VD-N3-S-409.592	2010/2/18 23:07	1	111	47	0	0	0	0	0	0	1	111	243	-5424
35	VD-N3-S-344.610	2010/3/10 08:17	1	113	41	0	0	0	0	0	32	1	113	166	9728
36	VD-N3-S-344.610	2010/3/10 08:18	1	113	41	0	0	0	0	0	32	1	113	166	9728
45	VD-N3-N-318.090	2010/3/10 18:20	3	2	2	123	46	1	117	57	0	243	50	236	0
46	VD-N3-N-318.090	2010/3/10 18:21	3	2	2	123	46	1	117	57	0	243	50	236	0
47	VD-N3-N-371.395	2010/3/5 20:19	5	103	46	0	0	0	0	0	0	5	103	165	165
48	VD-N3-N-371.395	2010/3/5 20:20	5	103	46	0	0	0	0	0	0	5	103	165	165
49	VD-N3-N-318.090	2010/3/10 18:26	3	2	2	123	46	1	117	57	0	243	50	236	0
56	VD-N1-S-349.620	2010/3/10 18:40	26	93	47	2	93	57	0	0	0	28	93	171	33
57	VD-N3-N-318.090	2010/3/10 18:40	3	2	2	123	46	1	117	57	0	243	50	236	0
58	VD-N1-S-349.620	2010/3/10 18:41	26	93	47	2	93	57	0	0	0	28	93	171	33
59	VD-N3-N-318.090	2010/3/10 18:41	3	2	2	123	46	1	117	57	0	243	50	236	0
60	VD-N1-S-349.620	2010/3/10 18:42	26	93	47	2	93	57	0	0	0	28	93	171	33
61	VD-N3-N-318.090	2010/3/10 18:42	3	2	2	123	46	1	117	57	0	243	50	236	0
62	VD-N1-S-349.620	2010/3/10 18:43	26	93	47	2	93	57	0	0	0	28	93	171	33
63	VD-N3-N-318.090	2010/3/10 18:43	3	2	2	123	46	1	117	57	0	243	50	236	0
64	VD-N1-S-349.620	2010/3/10 18:44	26	93	47	2	93	57	0	0	0	28	93	171	33

Table 3 Data errors of same value of several minutes continuously.

1	timeT	L1_Q	L1_V	L1_OCC	L2_Q	L2_V	L2_OCC	L3_Q	L3_V	L3_OCC	L4_Q	L4_V	L4_OCC	Total_V	Total_Q	Total_OC
59963	2010/3/28 11:20	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59964	2010/3/28 11:21	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59965	2010/3/28 11:22	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59966	2010/3/28 11:23	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59967	2010/3/28 11:24	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59968	2010/3/28 11:25	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59969	2010/3/28 11:26	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59970	2010/3/28 11:27	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59971	2010/3/28 11:28	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59972	2010/3/28 11:29	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59973	2010/3/28 11:30	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59974	2010/3/28 11:31	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59975	2010/3/28 11:32	30	80	25	27	73	24	31	65	29	25	67	22	113	71	25
59976	2010/3/28 11:33	16	86	12	20	83	15	20	81	15	20	76	16	76	81	14
59977	2010/3/28 11:34	16	86	12	20	83	15	20	81	15	20	76	16	76	81	14
59978	2010/3/28 11:35	16	86	12	20	83	15	20	81	15	20	76	16	76	81	14
59979	2010/3/28 11:36	16	86	12	20	83	15	20	81	15	20	76	16	76	81	14
59980	2010/3/28 11:37	16	86	12	20	83	15	20	81	15	20	76	16	76	81	14
59981	2010/3/28 11:38	16	86	12	20	83	15	20	81	15	20	76	16	76	81	14
59982	2010/3/28 11:39	16	86	12	20	83	15	20	81	15	20	76	16	76	81	14
59983	2010/3/28 11:40	16	86	12	20	83	15	20	81	15	20	76	16	76	81	14

Comparative analysis of the detectors at different locations, the main line segments have the most of detector error, even based on the odds ratio, as illustrated in Table 4.

Table 4 Detector data errors at different locations on freeway

	Errors type	Main line (358 detectors, 59.1%)	On ramp (140 detectors, 23.1%)	Off ramp (108 detectors, 17.8%)	Total
1	Missing value	3,404,423	1,264,860	876,459	5,545,742
		12.58%	4.68%	3.24%	20.50%
2	Null value	624,204	30,324	190,860	845,388
		2.31%	0.11%	0.71%	3.13%
3	Flow rate per min. Q>60	5648	376	220	6,244
		0.02%	0.00%	0.00%	0.02%
4	Speed V>255	0	0	0	0
		0.00%	0.00%	0.00%	0.00%
5	OCC>100	0	0	0	0
		0.00%	0.00%	0.00%	0.00%
6	Q=0,OCC≠0 or Q=0,V≠0	7,422	0	3	7,425
		0.03%	0.00%	0.00%	0.03%
7	Q≠0, OCC=0 or V≠0，OCC=0	1,428,049	154,880	177,318	1,760,247
		5.28%	0.57%	0.66%	6.51%
8	Same value in several columns	1,435,021	529,652	435,848	2,400,521
		5.30%	1.96%	1.61%	8.87%
Total		6,904,767	1,980,092	1,680,708	10,565,567
		25.52%	7.32%	6.21%	39.06%
Percentage of correct data		65.35%	18.74%	15.91%	100%
Odds ratio of location		65.35/59.1=1.11	18.74/23.1=0.81	15.91/17.8=0.89	

3. UNREASONABLE DETECTOR DATA

Even without the data errors like the types abovementioned, the relationship among flow rate, speed and occupancy, which seem to be reasonable, also could not match with the fundamental relationship of macroscopic traffic flow model. Therefore, in the research, a built-in mechanism to check the detector data is modeled. If the relationship among the data at one location is out of the range of traffic flow model, then, the relation of data e.g. flow-rate and occupancy is unreasonable. It needs to develop a traffic flow model to be the rule for checking suck kind of errors. Using the historical data of a detector, the fundamental relationship is established using regression. An example is shown in Fig 1. The 99% of the prediction interval of regression line is drawn as the upper- and lower-bound of the detector data. If some of the intervals of detector data locate outside the range, the detector data will be fault. Then, the data should be corrected using the data imputation model.

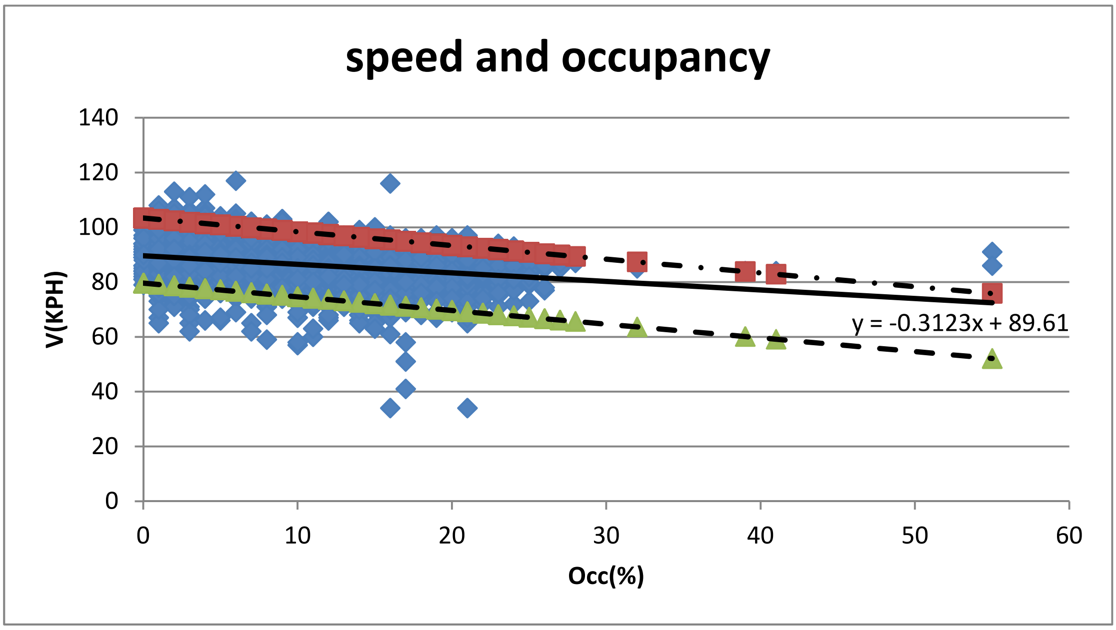

Figure 1 Macroscopic traffic flow model for checking reasonability of detector data

4. DATA IMPUTATION

To infill the missing value and/or to correct the data errors, it can be conducted real time by every minute for real time control, or be conducted off-line every day to maintain the data base of traffic count data. In this study, through a comparative investigation using Artificial Neural Network (ANN) and Rolling Grey Model (RGM)(Deng, 1999), the model of ANN is chosen for missing data imputation due to its better prediction result for missing data. The comparison results are illustrated in Table 5. The result of ANN is much better than the RGM model with the lower MAPE (Mean Absolute Percentage Data). Through the minimized the training errors by the fittest chromosome, the modeled neural network is with one hidden layers and input parameters of data of the time series (t, t-1,…,t-n) of upper bound detector (i-1) and lower bound detector (i+1) are taken for input to neural network for prediction the missing data of missing value of n intervals from time t to time t+n, as shown in Fig. 2. In the ANN model the input data are the data of the intervals which are same with number of the missing data intervals (n) will be imputed of upper bound detector and lower bound detector.

Table 5 Prediction errors of RGM and ANN model of data imputation (MAPE %)

		Imputation data interval							
		1 min	**2 min**	**3 min**	**4 min**	**5 min**	**10 min**	**15 min**	**30 min**
MAPE % (Morning Peak Hour)	RGM	14.0	14.5	16.0	16.8	17.7	21.1	26.2	48.1
	ANN	7.1	8.6	9.1	10.2	10.90	12.0	12.6	14.5
MAPE % (Afternoon peak hour	RGM	9.1	9.0	9.2	9.4	9.7	11.3	14.0	23.0
	ANN	3.3	6.3	6.3	5.8	6.7	8.0	8.1	9.2

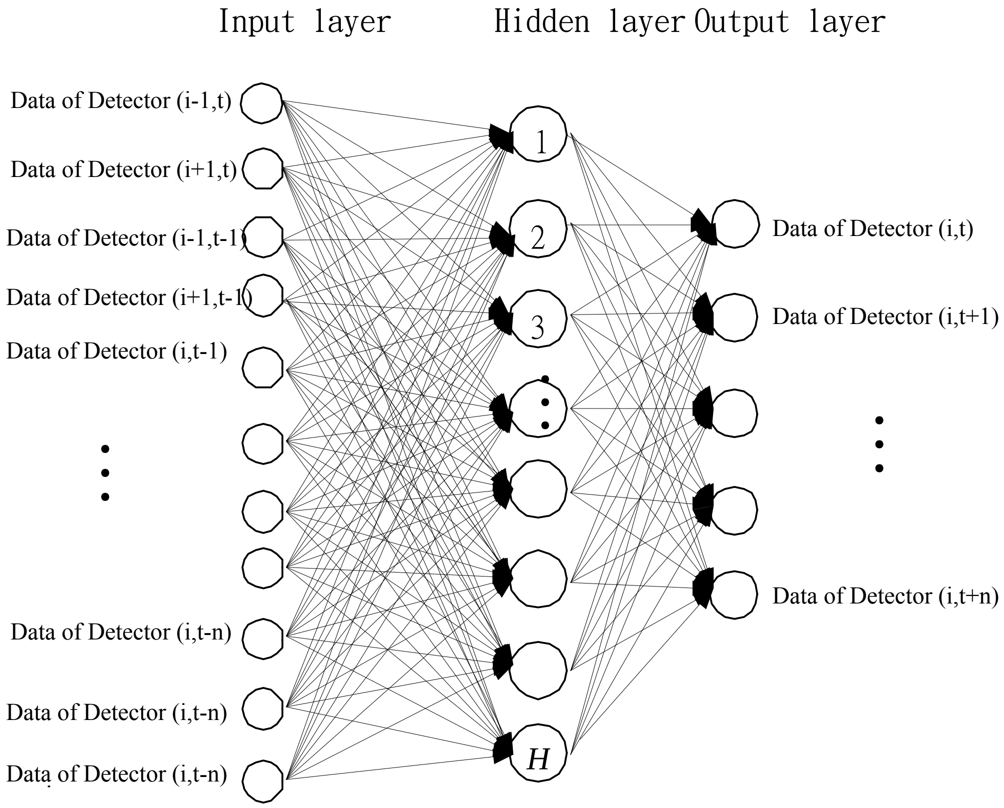

Figure 2 Neural Network model for data imputation

5. COMPREHENSIVE MODEL AND APPLICATION

By combining the abovementioned procedures, a comprehensive model is established to include the mission data checking, data error checking and the macroscopic traffic flow fundamental relationship checking combined with a Neural Network imputation model, as illustrated in Fig. 3. The model is now established into the traffic control center and run routinely every day for correcting and maintaining the detector data base.

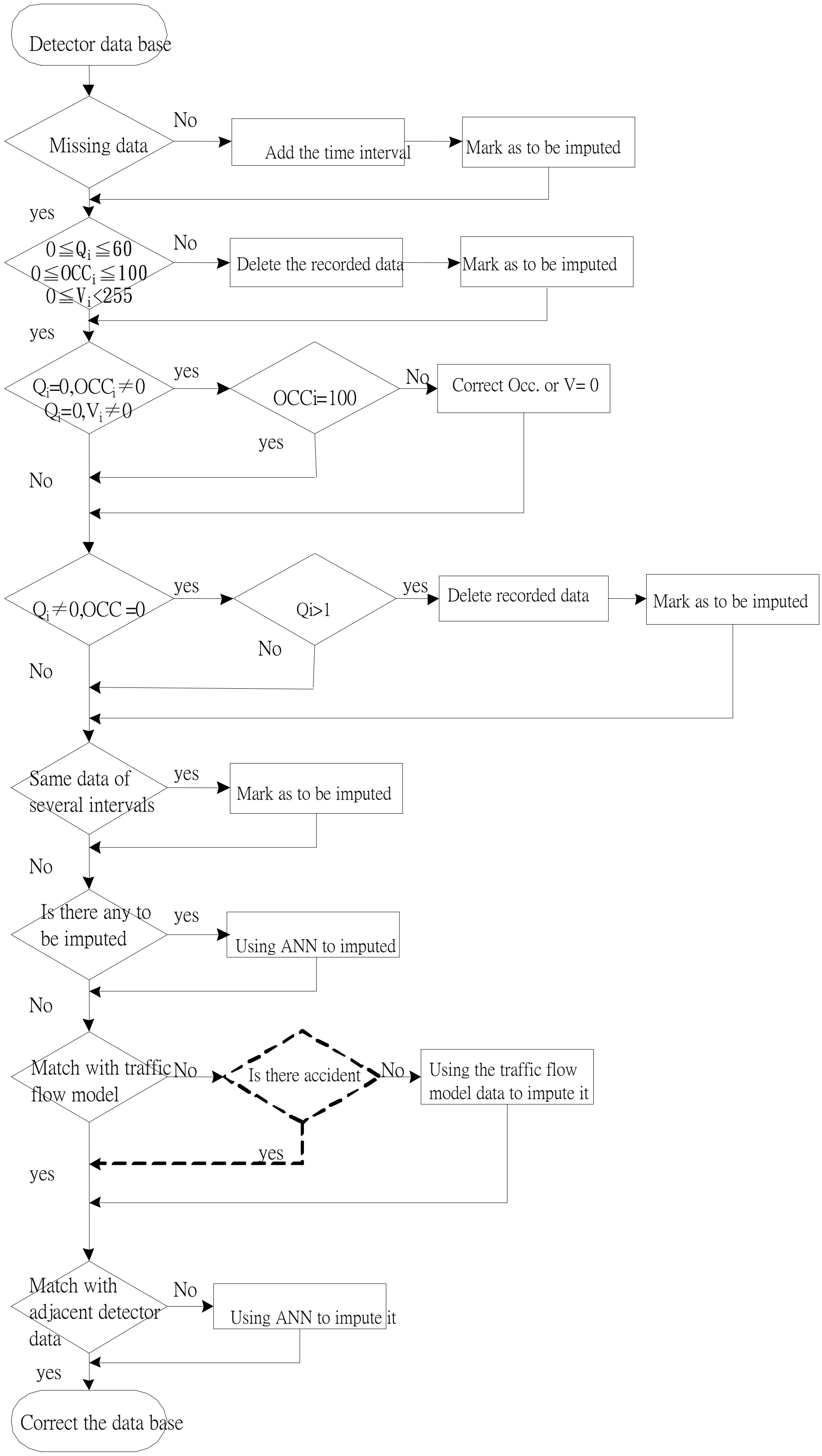

Figure 4 Comprehensive models of detector data errors checking and data imputation

6. CONCLUSION AND SUGGESTION

Errors of detector data will cause the traffic control strategy malfunctioned. To check the data errors of detector is a necessary work for maintaining the effective operation of traffic control center. This paper developed a comprehensive model to check the detector errors with various steps, including checking missing data, checking unreasonable data based on the traffic flow model etc. By combining with a neural network data imputation model, the comprehensive model then becomes an applicable model practically for off-line maintenance of detector data base. In the future, the model could be extended to apply for real time data correction model, and it needs to be developed furthermore.

REFERENCE

1. FHWA, Traffic Detector Handbook, 2006.
2. P. Cheevarunothai, Y. Wang, N.L. Nihan, "Identification and Correction of Dual-Loop Sensitivity Problems", Transportation Research Record, vol.1945 ,2006
3. L. Vanajakshi, L. R. Rilett, "Loop Detector Data Diagnostics Based on Conservation-of-Vehicles Principle", Transportation Research Record vol. 1870, Transportation Research Board of the National Academies, 2004
4. Zhong, M., Lingras, P., and Sharma, S., "Estimation of missing traffic counts using factor, genetic, neural and regression techniques", Transportation Research, Part C, Emerging Technologies, Vol.10(4), pp.139-166, 2004.
5. Williams, B.M. etc., Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models, Transportation Research Record 1644, 1998.
6. Deng, J.-L., The method and application of grey prediction model, 1999.
7. Hsu, T.-P., Chen Y.-H., A tentative check on detector error based on flow rate difference, Proceeding of 17^{th} ITS World Congress 2010.