

교통량 데이터의 실시간 보정 로직 - 국도 3호선을 중심으로 *

이지연¹⁾ 도명식²⁾ 김성현³⁾ 류승기⁴⁾

요약

교통량, 속도에 근거한 통행시간 추정은 국도교통관리체계(NHTMS)에서 운전자들에게 신속하고 정확한 교통상황에 대한 정보를 제공하기 위해 필요한 중요 작업이다. 이를 위해 기 구축된 교통시스템에서 수집된 교통량 데이터의 특징을 파악하고, 결측이 발생했을 때 원활한 교통정보 제공을 위한 실시간 보정 방법을 알아본다. 기존의 평균을 이용한 보정과 인근의 유사 검지기의 정상적인 데이터를 이용한 자기회귀 보정 등의 방법을 살펴보고 2000년도 국도 3호선에서 얻어진 실제 교통량 데이터에 적용하여 비교한다.

주요용어: 교통량, 국도교통관리체계, 평균 보정, 자기회귀 보정

1. 서론

국도 3호선은 남해~초산(철원)까지 총 연장 555km의 도로로서 이 중 수도권 남부 지역인 성남~이천간의 약 50km 구간에는 지능형 교통시스템(ITS : Intelligent Transport Systems)이 구축되어 있다(건설교통부(1999), 서울지방국토관리청(2000)). 지능형 교통시스템이란 첨단 정보통신기술을 활용한 교통관리시스템으로서, 도로상 교차로와 교차로 사이의 구간에 설치된 영상 및 루프 검지기를 통해 도로구간별 교통 정보를 자동파악하는 시스템이다. 이 시스템에서 수집된 정보는 운영센터에서 전산 처리되어 도로변의 도로전광표지를 통해 도로 이용자에게 제공된다. 또한, 교통사고 등의 유사시에는 도로 관리청, 경찰청 등에 통보하여 교통관리를 수행토록 한다(건설교통부(1999), 서울지방국토관리청(2000), Tarko and Roupail(1997)). 이를 위해 성남~이천간에는 34대의 영상 검지기(카메라), 1개소의 지역 운영센터, 9대의 도로 전광표지 등이 설치되어 있고, 연계구간으로 송파~성남 구간에 3개의 검지기, 42호선의 마장~이천구간에 2개의 검지기, 43호선의 하남~광주구간에 8개의 검지기가 설치되어 총 47개의 검지기가 설치되어 있다(2001년 기준). 이 시스템에 의해 국도 3호선의 성남~이천간 도로 이용자는 도로변에 설치된 도로전광표지를 통해 구간별

* 이 연구는 한국건설기술연구원의 지원으로 이루어졌음

1) (712-749) 경북 경산시 대동 214-1, 영남대학교 통계학과, 부교수.

E-mail : leejy@yu.ac.kr

2) (305-719) 대전시 유성구 덕명동 산 16-1, 한밭대학교 토목도시환경공학부, 전임강사.

E-mail : msdo@hanbat.ac.kr

3) (411-712) 경기도 고양시 일산구 대화동 2311, 한국건설기술연구원 토목연구부, 연구원.

E-mail : vocer@kict.re.kr

4) (411-712) 경기도 고양시 일산구 대화동 2311, 한국건설기술연구원 토목연구부, 선임연구원.

E-mail : skryu@kict.re.kr

교통소통 상황, 목적지까지 소요시간 등의 교통정보를 제공받을 뿐 아니라, 도로가 정체될 경우에는 인근의 우회도로를 이용하여 통행 시간을 단축시킬 수 있게 되었다. 이런 교통 정보를 제공하기 위해서 각 검지기에서는 교통량, 속도, 점유율 데이터가 실시간으로 수집되고 있다. 교통량이란 ‘도로의 한 지점, 또는 단면을 단위 시간 동안 통과하는 차량의 수’를 의미하는 것으로 1분, 5분, 1시간으로 누적하여 저장하고 있다(한국건설기술연구원(1995), 건설교통부(2001)). 속도, 점유율은 1분, 5분, 1시간의 평균을 계산하여 각각 저장하며 이 자료들로부터 통행시간을 실시간으로 추정하여 도로 전광표지를 통해 도로 이용자들에게 제공한다(김영찬 등(1997), 조중래, 김현주(1998)). 그러므로 검지기에서 수집된 자료 중에서 누락이나 불량으로 판단되는 자료는 통계 자료의 객관성을 유지하기 위해서 보정을 해야하며, 또한 검지기 미설치 구간 및 검지 신뢰도가 낮은 구간에서는 원활한 교통정보의 제공과 통행시간의 추정을 위해서 보정이 이루어져야 한다.

본 연구에서는 국도 3호선과 연계구간의 총 47개 검지기에서 고장이나 전송 오류로 인해 결측이 발생했을 때, 원활한 교통 정보 제공을 위한 결측 교통량 데이터의 실시간 보정 방법을 알아본다. 일반적으로 1시간 누적 교통량 데이터의 보정은 실시간 통행 시간 추정과 교통소통 정보 제공의 입장에서 그 효과가 떨어지고, 1분 누적 교통량 데이터도 교통신호 대기 시간을 고려할 때 그 지점의 순수한 교통량 변화 패턴을 반영하는 데이터로는 보기 힘들다. 따라서 본 논문에서는 통행시간 추정을 위해 5분 교통량 데이터를 중심으로 교통량의 결측 데이터를 보정하는 방법에 대해서 살펴본다. 2장에서는 검지기에서 수집된 2000년도 교통량 데이터를 이용하여 검지기의 평균적인 변동 패턴의 유의성을 검정하고 이 결과를 이용하여 3장에서는 결측이 발생했을 때, 적용할 수 있는 보정방법들을 살펴보고 서로 비교한다.

2. 연평균 교통량 데이터의 특징

검지기별로 얻어진 2000년의 국도 3호선 상행 교통량 데이터로부터 하루 교통량 변동 패턴이 월별, 요일별로 유의한 차이가 있는지 분석한다.

2.1. 월별, 요일별 하루 교통량 변화 패턴의 유의성 검정

5분 교통량의 하루 변화 패턴이 월별, 요일별로 유의한지를 파악하기 위해 검지기별로 월과 요일을 개체-간 요인으로 설정하고, 개체-내 요인은 5분 교통량 대신에 1시간 교통량 데이터를 이용하여 교통량 변화 패턴을 분석하였다. 하루의 교통 패턴을 반영한다는 점에서는 1시간 교통량 데이터의 변동 패턴과 5분 교통량 데이터의 변동 패턴이 크게 다르지 않을 것이고, 288개의 5분 교통량 데이터를 개체-내 요인으로 설정하기에는 그 수가 많기 때문이다(노형진(1999), 원태연, 정성원(2001)). 그래서 24개의 1시간 교통량을 개체-내 요인으로 설정하여 반복이 있는 분산분석을 실시하였다. 예를 들어 4번 검지기의 교통량 데이터에 대해서는 다음 <표 2.1>이 얻어졌다.

대부분의 검지기들에서 <표 2.1>과 같이 1시간 교통량의 하루 변화 패턴이 월별, 요일별로는 유의한 차이가 나타나고, 월과 요일 사이의 교호작용은 약하게 나타났다. 이것은

표 2.1: 4번 검지기의 1시간 교통량의 월별, 요일별 유의성 분석

소스	제 III 유형 제곱합	자유도	평균제곱	F	유의확률
Intercept	4.213E+09	1	4.213E+09	17072.426	0.000
요일	5411254.0	7	773036.284	3.132	0.004
월	7669875.6	9	852208.404	3.453	0.001
요일*월	19287997.0	59	326915.205	1.325	0.084
오차	42694716.0	173	246790.264		

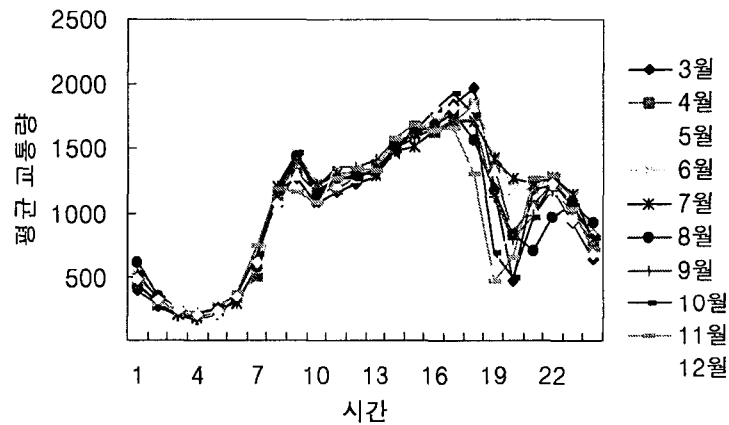


그림 2.1: 4번 검지기의 월별 하루 평균 교통량 변화 패턴

<그림 2.1>과 <그림 2.2>에서 7, 8월과 11, 12월의 교통 변화 패턴에 큰 차이가 있고, 일요일과 공휴일이 다른 요일과 비교하여 그 변화 패턴이 뚜렷하게 차이가 나는 것으로도 확인할 수 있다.

2.2. 하루 교통량 변화 패턴이 유사한 월과 요일에 대한 다중 비교

2.1절의 결과로 하루 교통량 변화 패턴이 월별, 요일별로 유의한 차이가 있음을 확인하였다. 이제는 그 변화의 패턴이 유사한 월과 요일을 그룹으로 형성하는 그룹핑 분석을 하고자 한다. 먼저 시점별로 얻어지는 교통량 데이터에 대해 월과 요일에 대한 다중비교를 실시하였다. <표 2.2>는 4번 검지기의 06:00~07:00 동안의 월과 요일에 대한 Tukey 다중비교의 결과이다. 오전 출근 시간이라 일요일, 공휴일은 주중요일과는 유의한 차를 보이고 교통량도 주중의 다른 요일과는 뚜렷한 차이를 보임을 알 수 있다.

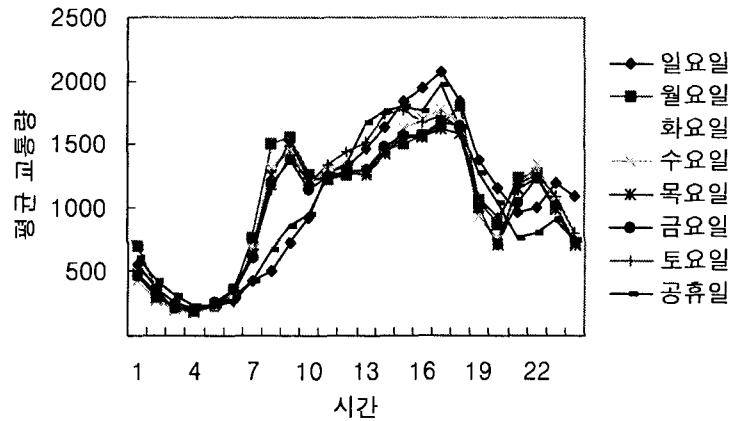


그림 2.2: 4번 검지기의 요일별 하루 평균 교통량 변화 패턴

2.3. 하루 평균 교통량 변화 패턴이 유사한 월과 요일의 군집분석

2.2절에서는 시점별로 교통량 데이터의 월별, 요일별 유의성을 살펴 보았고, 본 절에서는 하루 전체에 걸친 교통량의 평균적인 변동 패턴이 월별, 요일별로 유의한지를 알아본다. 즉, 검지기별로 각 시점별 월별, 요일별 연평균 교통량을 계산한 다음 그것을 결합하여 그 검지기의 평균적인 하루 교통량 변화 패턴으로 간주한다. 그러면 5분 교통량 데이터에서 계산된 288개 시점의 평균 교통량 데이터를 변수로 하여 월별, 요일별로 군집분석을 실시할 수 있다. 4번 검지기의 경우 2000년 1월과 2월 데이터가 모두 결측되었기 때문에 1월, 2월을 제외한 데이터를 가지고 월별 시점별 교통량 평균을 구하였다. 그런 후, 월평균 변화 패턴에 대한 제곱 유클리디안 거리 기준의 군집분석을 실시한 결과, 6월과 7월이 가장 먼저 동일 군집으로 형성되고 다음에 4월과 5월이 동일군집으로 형성되는 것이다. 두 개의 군집으로 분류할 경우 3월~10월이 하나의 군집을, 그리고 11월, 12월이 나머지 군집을 형성한다. 즉, 3월~10월이 11월, 12월과는 평균적인 교통패턴에 있어 유의한 차이가 있음을 알 수 있다.

실제 군집의 형태는 검지기마다 다르게 나왔지만 월별로 유의한 차이가 있다는 것은 공통적으로 얻어졌다. 따라서 3.1절의 과거 프로파일을 이용한 결측 보정에서 결측이 발생한 현 시점의 교통량 변동 패턴과 유사한 과거 프로파일을 이용할 때, 직전 5주의 데이터를 이용하는 것이 바람직할 것으로 보인다. 그 이유는 교통량 변동 패턴이 월별로 유의하기 때문에 5주보다 더 과거의 교통량 변동 패턴은 결측이 발생한 현 시점의 교통량 변동 패턴과 유의한 차이를 보일 수 있기 때문이다.

교통량 변화 패턴이 유사한 요일들을 그룹핑하기 위해서 검지기의 요일별 연평균 5분 교통량의 군집분석을 실시하였다. 요일 자료에서는 월별 자료와는 달리 대부분의 검지기

표 2.2: 4번 검지기의 06:00 ~ 07:00의 1시간 교통량에 대한 월과 요일 다중비교

	월	관측수	집단군			
			1	2	3	4
Tukey	4	27	519.6296			
	5	22	570.2273	570.2273		
	9	24	571.7500	571.7500		
	3	23	574.9565	574.9565		
	8	23		605.3043	605.3043	
	7	17		606.1176	606.1176	
	6	30		608.4667	608.4667	
	12	27		618.7037	618.7037	
	10	28			679.8214	679.8214
	11	28				756.3214
	유의확률			0.496	0.684	0.109

	요일	관측수	집단군			
			1	2	3	
Tukey	일요일	39	421.1795			
	공휴일	8	429.6250			
	금요일	38		609.1316		
	토요일	36		619.8056		
	목요일	32		647.0625		
	화요일	28		647.4643		
	수요일	35		675.9714		
	월요일	33			760.6364	
	유의확률			1.000	0.161	1.000

로부터 동일한 다음의 결과를 얻었다. 분류할 군집의 수에 따라

- 5개 그룹 - {일요일}, {월요일}, {화, 수, 목, 금요일}, {토요일}, {공휴일}
- 4개 그룹 - {일요일}, {월, 화, 수, 목, 금요일}, {토요일}, {공휴일}
- 3개 그룹 - {일요일}, {월, 화, 수, 목, 금, 토요일}, {공휴일}
- 2개 그룹 - {일요일, 공휴일}, {월, 화, 수, 목, 금, 토요일}

로 얻어진다. 따라서 평균 교통량 변화 패턴은 주중 요일, 토요일, 일요일, 공휴일로 유의한 차이가 있고, 특히 오전시간대에는, 2.2절에서 살펴보았듯이 월요일을 나머지 주중과 구별해야함을 알 수 있다. 2000년도에서는 관측된 공휴일 데이터가 일년 중 9일에 불과하며 9일의 평균 교통량 패턴이 일요일의 그것과 유의한 차를 보이므로 공휴일의 결측 데이터는 별도 관리를 한다. 하지만 넓은 의미로는 일요일의 교통량 패턴 변화에 준하여 분석할 수 있다.

3. 결측 교통량 데이터의 보정로직

3.1. 프로파일을 이용한 평균 보정

2장에서 교통량 하루 변화 패턴에 월별, 주중토일별로 유의한 차이가 있음이 확인되었으므로 겹치기에서 결측 혹은 손실이 발생하면 발생한 시점으로부터 과거 5주의 주중, 토, 일요일별 동일시점의 프로파일을 이용해 결측 데이터를 보정한다. 특히, 주중의 경우는 과거 5주간의 동일 시점에서 얻어지는 25개의 데이터의 평균과 신뢰구간으로 결측 데이터를 보정할 수 있다. 다만, 월요일의 경우, 2.2절에서처럼 오전 시간 동안의 교통량이 다른 주중 요일과는 유의한 차이를 보이므로 결측이 발생한 시점이 오전 시간이라면, 주중요일 중에서도 월요일은 구별하여, 동일시점 과거 5주의 월요일 프로파일만 이용한 평균으로 보정하는 것이 효과적이다.

주중, 토, 일요일별 교통량 데이터의 5분 시점별로 평균 교통량과 표준오차를 통해 신뢰도에 따른 구간 추정도 할 수 있다. 시점 $i, (i = 1, 2, \dots, 288)$ 에서 직전 5주 주중, 토, 일별로 얻어진 데이터의 개수가 n_i 이고 평균이 \bar{x}_i , 표준편차가 s_i 라면 신뢰도 $(1 - \alpha)100\%$ 에 의한 평균 교통량의 신뢰구간에 대한 하한과 상한은 각각

$$\left(\bar{x}_i - t_{\alpha/2}^{n_i-1} \frac{s_i}{\sqrt{n_i}}, \quad \bar{x}_i + t_{\alpha/2}^{n_i-1} \frac{s_i}{\sqrt{n_i}} \right)$$

로 얻어진다. 이때 $t_{\alpha/2}^{n_i-1}$ 는 자유도 $n_i - 1$ 의 t 분포의 상위 $\alpha/2$ 에 해당하는 값이다.

주중의 5분 교통량 데이터가 결측 혹은 손실되었을 때 비교하고자 하는 프로파일을 이용한 보정로직은 다음과 같다.

- 보정 1: 직전 5주의 주중(월, 화, 수, 목, 금) 자료를 모두 포함한 산술평균
- 보정 2: 직전 5주의 자료 중에서 오전시간(09:00 이전)에는 월요일 자료를 주중자료에서 제외시키고 나머지 주중(화, 수, 목, 금) 자료를 모두 포함한 산술평균 그리고 월요일 오전시간은 직전 5주의 월요일 자료만 이용한 산술평균
- 보정 3: 직전 5주의 동요일 자료만 이용한 산술평균
- 보정 4: 직전 5주 동요일 자료에 대해 직전부터 5, 4, 3, 2, 1의 가중을 준 가중평균(기존 보정로직)

토요일, 일요일 교통량 데이터가 결측되었을 때는 보정 1, 보정 2, 보정 3은 직전 5주의 동요일 자료의 산술평균으로 보정하는 동일한 방법이고 보정 4만 가중평균으로 보정하는 다른 방법이 된다.

보정의 정확성을 상호평가하기 위하여 관측값(y_i)과 보정값(\hat{y}_i)의 차이를 나타내는 RMSE(Root Mean Square Error)와 MARE(Mean Absolute Relative Error)를 척도로 사용하였으

표 3.1: 4번 검지기의 교통량 프로파일 보정로직 비교

		보정1	보정2	보정3	보정4
12월 10일(일)	RMSE	44.7724	44.7724	44.7724	43.6114
	MARE	0.4716	0.4716	0.4716	0.4623
12월 11일(월)	RMSE	32.3086	36.6224	35.3432	36.7937
	MARE	1.1072	1.1018	1.1520	1.1325
12월 12일(화)	RMSE	25.4921	26.0958	27.3039	24.4659
	MARE	0.5482	0.5593	0.5685	0.4766
12월 13일(수)	RMSE	32.0194	32.1310	33.9459	33.8297
	MARE	0.5176	0.5197	0.5347	0.5236
12월 14일(목)	RMSE	24.7222	24.0279	30.5745	28.9600
	MARE	0.2427	0.2376	0.2531	0.2488
12월 15일(금)	RMSE	32.1975	32.7718	35.6792	35.9884
	MARE	0.4909	0.5044	0.5574	0.5579
12월 16일(토)	RMSE	38.1959	38.1959	38.1959	40.0242
	MARE	0.5161	0.5161	0.5161	0.5677

며 각 식은 다음과 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$MARE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}$$

<표 3.1>은 4번 검지기의 12월 10일부터 12월 16일까지의 자료에 대한 로직들의 비교 결과이다. 대체로 보정 2의 결과가 타 보정로직보다 약간 우수함을 볼 수 있다.

과거 프로파일의 평균에 의한 보정은 현 시점의 결측 교통량을 그 검지기의 과거 자료에 의해 보정하기 때문에 현재 발생하는 특이 상황은 전혀 반영하지 못한다는 단점이 있다. 하지만 각 지점의 교통 성향이 과거와 유사한 평균적인 패턴을 보인다면 본 보정로직도 유효하다. 그리고 3.2절과 3.3절에서처럼 결측 검지기와 교통 패턴이 유사한 인근 검지기를 찾을 수 없다면 과거 프로파일을 이용한 평균 보정이 유일한 방법이다. 이 때, 동요일 직전 5주의 동시점 교통량의 평균에 의한 보정(보정3)보다는 결측이 발생한 요일이 주중요일이라면 직전 5주의 전체 주중요일을 모두 포함한 동시점 교통량의 산술평균으로 보정하는 것이 약간 효과적이다. 이는 교통량 변동 패턴이 주중 요일에서는 유사함을 확인했고, 보다 많은 샘플 수에서 추정하는 것이 그 추정오차가 작아지기 때문이다.

3.2. 유사 검지기의 시점별 데이터를 이용한 회귀 보정

교통량 패턴이 주중, 토, 일요일별로 유의함을 확인하였으므로 2000년도 자료를 이용하여 주중토일별, 시점별로 검지기들의 교통량 상관성을 분석하였다. 특정 시점에서 교통량

변동의 상관성이 높은 인근의 검지기들을 그 시점의 유사 검지기로 선별하는 것이다. 이 유사 검지기를 이용하여 결측이 발생한 시점에서 가장 교통량 상관성이 높은 인근 검지기의 정상적인 교통량 데이터를 이용하여 회귀 추정을 실시한다. 결측이 발생한 검지기의 주중, 토, 일요일별 5분 교통량 데이터를 Y_i 라고 하고 동일 시점의 유사 검지기에서 얻어진 교통량 데이터를 X_i 라고 하면 다음과 같은 단순선형 회귀모형을 고려할 수 있다.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

이 때, $\epsilon_i \sim N(0, \sigma^2)$ 을 따른다. 그러면 β_0 와 β_1 은

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

으로 추정할 수 있다. 여기서 \bar{y} 는 결측 검지기의 과거 5주의 교통량 데이터의 평균이고 \bar{x} 는 유사 검지기의 과거 5주의 교통량 데이터의 평균이다.

일반적으로 교통량 변동이 작은 새벽시간대나 저녁 늦은 시간대인 경우는 검지기들 사이의 상관성이 비교적 높게 얻어지는 반면에, 교통량 변동이 잦은 낮 시간대에는 시점별 교통량 상관계수가 회귀추정의 효과를 볼 수 있을 정도로 높은 유사 검지기를 가지고 있는 검지기가 매우 드물기 때문에 본절의 보정 로직은 적용의 한계가 있다.

3.3. 유사 검지기의 하루 전체 교통량 데이터를 이용한 자기회귀 보정

2.3절에서 하루 전체의 연평균적인 교통량 패턴은 주중, 토, 일요일별로 유의함을 확인하였다. 결측이 발생했을 때, 주중, 토, 일요일별의 해당 요일에 결측 검지기와 하루 전체의 평균적인 교통량 패턴이 유사한 인근 검지기가 있다면 그것을 보정에 이용할 유사 검지기로 선택한다. 일반적으로 새벽 시간대와 늦은 밤 시간대는 교통 변동이 극히 작아서 대부분의 검지기에서 그 시간대의 교통량 데이터의 상관성은 매우 높게 나타난다. 그러므로 하루 중에서도 새벽 시간과 늦은 밤 시간을 제외한 07:00~20:00의 교통량 하루 변화 패턴이 유사한 인근 검지기로 유사 검지기를 선택한다. 이 시간대에 상관계수가 높은 검지기가 있다면 교통량 변화 폭이 상대적으로 작은 새벽 시간대의 교통량 변화패턴은 아주 유사할 것으로 기대되기 때문에 결과적으로 하루 전체의 교통량 패턴이 유사한 검지기로 볼 수 있다. 결측이 발생한 시점의 직전 5주의 주중, 토요일별 5분 프로파일 자료를 이용하여 결측 검지기의 교통량을 종속변수로, 유사 검지기의 교통량을 독립변수로 하는 단순선형 회귀모형을 적용하고자 한다. 그 전에 선형모형이 보다 잘 적합하게 되도록 적당한 변수변환을 할 수도 있다.

일반적으로 단순선형 회귀모형에서는 오차항들을 서로 독립이고 평균이 0, 분산이 σ^2 인 정규확률변수로 가정한다. 만약 다른 시차의 오차항들이 서로 독립이 아니고 상관되어 있다면, 자기 상관 혹은 계열 상관이 된다고 말한다(이우리, 오광우(1996), 최병선(1992)). 자기상관은 시계열 자료에서 흔히 나타나는데, 발생의 주된 이유는 모형속에 포함되어야 할 독립변수가 누락되어, 누락된 독립변수의 효과가 오차항에 포함되기 때문에 자기상관이 나

타나는 것이다. 그렇게 되면 추정된 회귀계수는 불편 추정량은 되지만 최소 분산을 가지지 않으므로 유효 추정량이 될 수 없고, 오차의 분산 σ^2 의 추정량인 MSE은 매우 과소 추정될 수 있다. 따라서 회귀계수의 표준오차도 과소 추정되는 등 t 분포나 F 분포를 사용하는 가설 검정을 더 이상 적용할 수 없게 된다(이우리, 오광우(1996), 최병선(1992)).

오차항의 독립성을 검정하는 방법으로 더빈-와트슨(Durbin and Watson)의 D 통계량이 있다. 자기상관에 대한 더빈-와트슨 검정방법은 오차항의 일차 자기상관만 검정할 수 있고, 또 표본의 크기가 어느 정도 이상 많아야 할 뿐만 아니라 검정결과 판정 오류의 영역이 나타나는 단점은 있으나 가장 많이 사용되는 방법이다. 더빈-와트슨의 통계량은 오차항의 자기상관 모수 ρ 가 0인가를 검정하는 것으로, 만약 $\rho = 0$ 이면, 오차항은 서로 독립이 된다. 예를 들어, 27번 검지기의 교통량과 유사 검지기로 판명된 26번 검지기의 교통량의 제공근에 대한 주중요일의 하루 5분 연평균 교통량에 대한 단순회귀분석의 결과를 살펴보면, 더빈-와트슨 검정통계량 값이 0.990으로 하한값 1.65이하의 값이 얻어졌다. 즉, 양의 자기 상관이 존재함을 알 수 있다. 일반적으로 자기상관에 대한 검정결과 오차항에 자기상관이 존재하는 것으로 판정되는 경우의 대체방안으로 다음 두 가지를 고려할 수 있다. 첫째 방법은 앞에서 설명한 것처럼, 오차항의 자기상관이 설명변수의 누락에 기인하므로 누락된 변수를 찾아서 모형 속에 포함시키는 것이다. 그러나 실제적인 응용에서는 그러한 변수를 찾아내는 것이 쉽지는 않다. 둘째 방법은 변수들의 변화를 통하여 새로운 모형을 설정하는 것이다. 여기서는 일반적으로 많이 사용되는 코크란-오커트(Cochrane-Orcutt)의 방법을 이용하여 오차항들간에 일차 자기상관이 존재하는 일차 자기회귀 오차모형(First-Order Auto-regressive Error Model)을 고려해서 결측 데이터를 보정하고자 한다.

시점 t 에서의 종속변수를 Y_t , 독립변수를 X_t 라고 했을 때,

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t, \quad \epsilon_t = \rho\epsilon_{t-1} + u_t$$

이 만족된다고 하자. 여기서, ρ 는 자기 상관모수로서 $|\rho| < 1$ 이고, u_t 는 새로운 오차항으로 $u_t \sim N(0, 1)$ 로 가정한다. 여기서, $Y'_t = Y_t - \rho Y_{t-1}$, $X'_t = X_t - \rho X_{t-1}$, $u_t = \epsilon_t - \rho\epsilon_{t-1}$, $\beta'_0 = \beta_0(1 - \rho)$ 로 나타내면 이 모형은 다음과 같이 된다.

$$Y'_t = \beta'_0 + \beta_1 X'_t + u_t$$

이 때 오차항 u_t 는 독립 오차항이 되므로 변형된 모형은 오차항의 독립성 가정이 성립되는 모형이 된다. 따라서 최소제곱법으로 β'_0 과 β_1 을 추정할 수 있게 되고, 이렇게 추정된 계수들은 좋은 통계적 특성을 지니게 된다.

27번 검지기의 교통량 데이터를 종속변수 Y_t , 26번 교통량의 제공근을 독립변수 X_t 로 했을 때 얻어지는 코크란-오커트 방법으로 다음과 같은 자기 회귀모형이 얻어졌다.

$$Y_t = -29.6346 + 13.7055X_t + \epsilon_t, \quad \epsilon_t = 0.5054\epsilon_{t-1} + u_t$$

$$(R^2 = 0.956, \text{ Durbin-Watson 통계량} = 2.12)$$

<표 3.2>는 27번 검지기의 12월 10일~16일의 각 시점의 5분 교통량이 결측되었을 때, 유사 검지기 26번을 이용하여 자기 회귀모형으로 보정한 것과 3.1절의 과거 프로파일을 이

표 3.2: 27번 검지기의 교통량 보정로직 비교

		자기회귀보정	보정1	보정2	보정3	보정4
12월 10일(일)	RMSE	23.0092	30.3604	30.3604	30.3604	32.4816
	MARE	0.3140	0.3113	0.3113	0.3113	0.3160
12월 11일(월)	RMSE	18.5712	23.7093	22.8819	22.0334	21.7573
	MARE	0.3479	0.3721	0.3601	0.3599	0.3540
12월 12일(화)	RMSE	18.4146	17.1986	17.1571	16.9234	17.1486
	MARE	0.3306	0.2432	0.2436	0.2452	0.2523
12월 13일(수)	RMSE	15.7335	13.8235	13.8261	15.1019	16.7374
	MARE	0.3513	0.2167	0.2182	0.2374	0.2516
12월 14일(목)	RMSE	16.4094	23.2099	23.6024	24.4153	24.3070
	MARE	0.3767	0.6572	0.6628	0.6787	0.6801
12월 15일(금)	RMSE	15.5001	18.7672	18.9752	18.9305	19.5162
	MARE	0.3055	0.3330	0.3367	0.3269	0.3407
12월 16일(토)	RMSE	23.2965	29.5983	29.5983	29.5983	27.8582
	MARE	0.4971	0.7485	0.7485	0.7485	0.7050

용한 보정의 결과이다. <표 3.3>은 11번 검지기로서, 토요일과 일요일의 경우 12번 검지기와 높은 하루 상관관계를 보이므로 12번 검지기의 교통량 데이터를 이용하여 보정한다. 이 경우에는 12번 검지기의 교통량 데이터보다는 교통량 데이터의 제공근에 대한 상관성이 더 높게 얻어져서 독립변수로는 12번 검지기의 교통량 데이터의 제공근을 사용하였다. 11번 검지기의 주중자료의 경우에는, 12번 검지기와 상관성이 낮아서 자기회귀보정에 의한 결과가 프로파일 평균을 이용하는 결과보다 나쁘게 얻어졌다. 그러므로 자기회귀보정은 결측이 발생한 검지기와 연평균 하루 5분 교통량의 변화 패턴이 주중토일별로 유사한 인근의 유사 검지기가 있는 경우에만 그 보정 효과를 얻을 수 있다.

4. 결론

국도 3호선에 설치되어 있는 검지기로부터 수집되는 교통량 데이터에서 결측이 발생했을 때의 보정은 먼저, 교통량 데이터가 월별 그리고 주중, 토, 일요일별로 유의한 차이가 있음을 확인하였으므로 이 결과를 이용한다. 즉, 결측이 발생했을 때 주중토일별로 구별하여 보정하고, 과거 프로파일을 이용하는 경우에도 직전 5주의 과거 데이터를 이용한 보정이 타당하다는 것이다.

결측 시점에서 결측 검지기와 과거 자료를 근거로 교통량 상관성이 높은 인근의 검지기를 찾는다. 그러나 실제 국도 3호선의 2000년도 자료로 분석한 결과, 결측 시점에서 유사 검지기의 정상 자료를 이용하여 회귀추정으로 보정할 수 있을 만큼 그 상관성이 높은 검지기는 소수에 불과했다. 따라서 특정 시점이 아니라 하루의 평균적인 교통 변화 패턴이 유사한 인근 검지기를 주중, 토, 일요일별로 선택하여, 이 유사 검지기의 정상 데이터를 이용한 보정이 제안된다. 만약 그런 유사 검지기가 없는 검지기에서 결측이 발생하면, 직전 5주

표 3.3: 11번 검지기의 교통량 보정로직 비교

		자기회귀보정	보정1	보정2	보정3	보정4
12월 10일(일)	RMSE	21.8701	24.9878	24.9878	24.9878	24.8032
	MARE	0.4117	0.3750	0.3750	0.3750	0.3754
12월 11일(월)	RMSE	19.9257	17.8161	17.9790	17.2308	16.9670
	MARE	0.4337	0.2928	0.2732	0.3961	0.3124
12월 12일(화)	RMSE	19.6660	16.3623	16.2259	18.8245	19.6007
	MARE	0.3657	0.2929	0.2644	0.2784	0.2763
12월 13일(수)	RMSE	20.9514	20.9277	20.7462	22.8899	22.7921
	MARE	0.3909	0.3424	0.3129	0.3279	0.3375
12월 14일(목)	RMSE	22.9303	17.9695	17.6943	19.1923	19.1611
	MARE	0.5283	0.3795	0.3425	0.3483	0.3435
12월 15일(금)	RMSE	19.5100	20.6125	20.3698	21.0045	21.1775
	MARE	0.3934	0.3024	0.2931	0.3095	0.3138
12월 16일(토)	RMSE	20.4508	24.8253	24.8253	24.8253	25.0117
	MARE	0.3535	0.3220	0.3220	0.3220	0.3217

의 프로파일을 이용하여 주중, 토, 일요일별로 평균으로 보정하는 것이 효과적이다. 만약 유사 검지기가 있다면, 유사 검지기의 정상적인 교통량 데이터를 독립변수로 하는 자기회귀모형을 적합시켜 보정하는 것이 효과적임을 확인하였다.

실시간으로 수집이 되는 1분, 5분 교통량 데이터는 그 변동이 매우 심하여 그 추이를 예측하기에 오차가 아주 크다(Wright et al.(1997)). 하지만 과거 일년동안 누적된 데이터를 이용하여 각 검지기의 연 평균적인 교통 패턴에 대한 분석은 가능하다. 본 연구에서도 그런 연 평균적인 교통 패턴에 기인한 검지기들의 교통 패턴을 해석하여 인근의 교통 패턴이 유사한 검지기를 선택했다. 하지만 평균적인 성향과 특정한 날 나타나는 교통 패턴은 아주 상이할 수 있으며 특히, 돌발적으로 발생하는 상황에 대한 예측으로는 여전히 힘들다는 문제를 안고 있다. 또한 교통량 변동이 심한 첨두시간대의 보정 오차는 크게 나타난다. 다만 정상적인 패턴을 유지하는 일반적인 상황에서는 교통량 데이터가 결측되었을 때, 과거의 교통량 변화 패턴을 기준으로 현재의 정상적인 인근 유사 검지기의 데이터를 이용하여 결측을 보정할 수 있다고 여겨진다.

참고문헌

- [1] 도로 교통량 조사 운영체계 수립 및 관련 기술 개발(1995). 한국건설기술연구원.
- [2] 도로 교통량 조사 지침 (2001). 건설교통부.
- [3] 수도권 국도 교통관리체계 시범구간 확충 및 평가연구, 최종보고서 (1999). 건설교통부.

- [4] 수도권 국도 교통관리체계 상세설계 및 구축방안 연구, 최종보고서 (2000). 서울지방국토관리청.
- [5] 김영찬, 최기주, 김도경, 오기도 (1997). 단일루프검지기를 이용한 간선도로 실시간 통행속도 추정 방법론, <대한교통학회지> 제15권 4호, 53-71.
- [6] 노형진 (1999). <한글 SPSSWIN에 의한 다변량 데이터의 통계분석>, 도서출판 석정.
- [7] 원태연, 정성원 (2001). <한글 SPSS10K 통계조사분석>, SPSS 아카데미.
- [8] 이우리, 오광우 (1996). <회귀분석-입문 및 응용>, 탐진.
- [9] 조중래, 김현주(1998). 고속도로 실시간 교통량 예측을 위한 인공신경망 입출력 요소분석, <대한교통학회지> 제16권 4호, 21-31.
- [10] 최병선(1992). <단변량 시계열 분석 1>, 세경사.
- [11] A. P. Tarko and N. M. Rouphail (1997). Intelligent Traffic Data Processing for ITS Applications, *Journal of Transportation Engineering*, Vol. 123 No.4, 298-307
- [12] T. Wright, P. S. Hu, J. Young and A. Lu (1997). *Variability in Traffic Monitoring Data*, Final Summary Report.

[2002년 7월 접수, 2003년 3월 채택]

Real-time Adjustment of Traffic Volume - Based on the National Highway Route 3 *

Jiyeon Lee¹⁾ Myung Sik Do²⁾ Sung Hyun Kim³⁾ Seung Ki Ryu⁴⁾

ABSTRACT

In order to provide the drivers with more reliable transportation information in NHTMS(National Highway Transportation Management System), it is important to estimate the expected passage time by using the traffic volume and speed. In this study, we analyze the characteristics of the traffic volume in the national highway and we investigate two real-time adjustment methods: the average adjustment method and the auto-regressive adjustment method. In addition, we compare them using the real data collected at the National Highway Route 3 in 2000.

Keywords: Traffic volume; NHTMS; average adjustment; auto-regressive adjustment.

* This research was supported by the Korea Institute of Construction Technology.

1) Associate Professor, Department of Statistics, Yeungnam University, Kyeongsbuk, 712-749, Korea
E-mail : leejy@yu.ac.kr

2) Assistant Professor, Department of Urban Engineering, Hanbat National University, Daejeon,
305-719, Korea
E-mail : msdo@hanbat.ac.kr

3) Researcher, Division of Civil Engineering, Korea Institute of Construction Technology, Kyonggi,
411-712, Korea
E-mail : vocer@kict.re.kr

4) Senior Researcher, Division of Civil Engineering, Korea Institute of Construction Technology, Kyonggi,
411-712, Korea
E-mail : skryu@kict.re.kr