

교통 데이터에 대한 품질 평가 및 자료 처리 기법의 구현

정수정⁰¹, 송수경¹, 이민수¹, 남궁성²

¹이화여자대학교 컴퓨터정보통신공학과

²한국도로공사 도로교통기술원 교통연구그룹

bloom01@ewhain.net, happymint@ewhain.net

Implementation of Quality Evaluation, Error Filtering, Imputation for Traffic Missing Data

Sujeong Cheong⁰¹, Sookyung Song¹, Minsoo Lee¹, Sung Namgung²

¹Dept. of Computer Science and Engineering, Ewha Womans University

²Korea Expressway Corporation & Transportation Technology

요 약

대용량의 자료가 생산됨에 따라 데이터를 효율적으로 저장, 관리, 이용할 수 있는 데이터웨어하우스의 역할이 중요하게 되었고, 그에 따라 자료 처리 기법의 개발은 필수 과제가 되었다. 품질 평가와 오류 판단, 결측 보정의 자료 처리 과정은 자료의 신뢰도를 판단하고 활용도를 높일 수 있는 과정으로 매우 중요하다. 본 논문에서는 우리나라의 실제 교통상황을 반영하고 평가 기준의 오차를 줄이면서 더욱 간단 명료한 평가 계산식을 도입하여 효율적인 품질평가와 오류판단, 결측 보정의 자료 처리 기법을 제안한다. 또한 오류 판단 기준에 새로운 파라미터를 도입하여 교통 연구자의 요구 사항을 반영할 수 있게 하였다. 결측 보정 과정은 여러 기법을 연구하고 기존의 결측 보정 기법에 입력 변수를 추가하여 실제 대용량의 교통 자료에 적용하였다. 그리고 교통 자료가 저장되는 데이터베이스에 직접 접근하여 결측 보정 과정을 수행하도록 PL/SQL로 구현하였으며, 이를 통해 교통 연구자에게 쉽고 다양한 방법으로 결측 보정을 수행하고 그 결과를 이용하여 다양한 교통 정보를 가공할 수 있는 환경을 제공하였다.

1. 서 론

품질 평가는 자료의 완전성과 유효성을 체크하여 결측된 자료의 비율과 이용 가치가 있는 자료의 비율을 도출함으로써 저장된 자료에 대한 평가를 할 수 있다. 오류 판단은 자료 중 오류 데이터를 분별하고 확인 가능한 수로 도출함으로써 신뢰성 있는 자료를 교통 연구자에게 제공할 수 있다. 기존 한국도로공사의 검지기 자료처리 시스템인 FTMS의 자료처리 알고리즘은 품질평가와 오류판단에 있어서 전 차선을 적용하고 있지 않고, 결측 자료를 반영하지 않는 등 실제 교통 상황에 적합하지 않으며 기존의 무분별한 자료의 적재와 부정확한 품질 평가와 오류 판단으로 인한 자료의 비유효성은 교통 연구자에게 올바르게 못한 결과 값을 도출하게 하고 자료 활용에 신뢰성을 가할 수 없게 된다. 결측 보정은 원시 자료 중 유효하지 않는 자료의 값을 다른 유효 자료의 값을 이용하여 보정하는 과정으로 본 논문에서는 결측 보정의 여러 기법을 연구하고 대용량의 실제 교통 자료에 적용하였다.

본 논문의 구성은 다음과 같다. 2장의 관련연구를 통해서 캘리포니아 PeMS 자료처리 과정을 살펴보고 3장에서는 완전성, 유효성을 체크하는 품질 평가의 새로운 기법에 대해 자세히 설명한다. 4장에서는 품질 평가 후 이루어지는 오류 판단과 결측 보정과 같은

자료 처리 과정의 개선과 다양한 기법을 제공하며 마지막으로 5장에서는 결론으로 끝맺는다.

2. 관련 연구

2.1. 캘리포니아 PeMS 자료처리 과정

PeMS (PERformance Measurement System)는 캘리포니아의 교통부에서 운영하는 교통 관리 시스템으로 캘리포니아 전역에서 수집된 루프 검지기의 자료를 수집, 가공, 저장하여 관리하는 시스템이다. UC버클리 대학과 CalTrans가 공동으로 연구, 개발하였으며, 교통 관리 시스템의 운영 효과를 어떻게 측정할 것인가에 관한 쟁점으로부터 시작하여, 1997년부터 기초연구가 시작되었다. 2003년에 PeMS 4.0 버전을 발표하면서 지금과 같은 형태의 모습을 갖추게 되었으며, 현재 6.2 버전이 발표된 상태이다. 웹 서비스로 제공되는 이 서비스는 검지기의 상태 및 자료의 질에 대한 정보는 물론이고, 이용자가 원하는 자료를 인터넷을 통해 쉽게 취득할 수 있다는 장점을 가지고 있다. 사용자가 원하는 자료에 대한 구간과 시간을 설정하면 그 결과를 텍스트 파일이나 엑셀 파일로 다운로드가 가능하며 특히 다양한 형태로 시각화되어 이용자의 편의성을 높였을 뿐만 아니라 자료 해석 수준을 향상시켜 ADMS 시스템의 좋은 예라

할 수 있다.

캘리포니아주 고속도로는 단일 루프 검지기와 이중 루프 검지기가 혼재해 있다. 단일 루프 검지기를 통해 수집된 자료로 속도를 추정하기 위해 g-factor를 개발하여 자료 처리 과정에 적용하고 있다. 5분 단위 자료로 집계하여 속도를 계산하고, outlier를 필터링하고 결측 자료(missing data)는 과거 자료를 이용하여 생성하며 이러한 절차로 처리된 자료는 1시간 단위, 1일 단위로 다시 집계된다. PeMS에서는 이력 자료 뿐만 아니라 실시간 자료를 쉽게 이해할 수 있도록 도표 및 관련 통계 측정값들을 제공하며 선택한 지역의 과거 30일 동안의 자료를 9가지 항목으로 간략히 보여주는 기능도 갖추고 있다. PeMS는 신속하고 정확한 자료 처리의 과정을 거쳐 사용자에게 이해하기 쉬운 시각화된 정보를 제공함으로써 우리나라 자료처리 시스템에도 부분 반영할 필요가 있다.

3. 품질 평가

3.1. 품질 평가의 개념

품질 평가란 자료 처리 과정 중 전처리 과정으로 자료의 완전성과 유효성을 체크한다. 완전성이란, 루프/지점별 수집된 전체 원시자료(교통량(vol)/점유율(occ)/속도(speed)) 중에서 결측되지 않은 자료의 비율을 말하며, 이때의 자료를 이용 가능한 자료라고 한다. 결측 자료는 검지기의 고장이나 차량의 부재로 인하여 누락된 데이터를 말하며 본 논문의 입력 자료에서의 결측 자료는 -111로 표시한다. 이용 가능한 자료의 비율이 높으면 완전성이 높다고 하고, 정량화 방법은 다음과 같다.

$$\text{완전성}(\%) = \frac{x}{X} \times 100$$

위의 식에서 x는 루프/지점별 관측된 원시자료(교통량/ 점유율/속도) 중 결측되지 않은 이용 가능한 자료(교통량/점유율/속도)이며, X는 루프/지점별 전체 원시자료(교통량 /점유율 /속도)이다.

유효성이란, 루프/지점별 수집된 이용 가능한 자료(교통량(vol)/점유율(occ)/속도(speed)) 중에서 오류판단 기준에 의거하여 오류가 없는 자료의 비율을 말하며, 이때의 자료를 유효한 자료라고 한다. 따라서 정량화 방법은 다음과 같다.

$$\text{유효성}(\%) = \frac{x}{X} \times 100$$

위의 식에서 x는 루프/지점별 관측된 이용 가능한 원시 자료 (교통량/점유율/속도) 중에서 오류가 없는 유효한 원시 자료 (교통량/점유율/속도)이며, X는 루프/지점별 관측된 이용 가능한 원시 자료(교통량/점유율/속도)이다.

3.2. 기존 품질평가 알고리즘

FTMS의 자료 처리 과정은 다양한 가공 프로세스의 지원이 필요하다고 보여 연구자들이 신뢰할 수 있는 데이터 처리가 가능하도록 다양한 매개변수 등을 설정하여 실험할 수 있도록 지원해야 하나 현재의 FTMS 자료처리 과정은 신뢰도를 판별하기 어려운 상황으로 오류가 포함될 가능성이 매우 높다.

현재 검지기 자료 처리 시스템인 FTMS의 품질평가의 알고리즘을 자세히 살펴보도록 한다. 기존의 자료처리 알고리즘은 현 우리나라 교통상황을 제대로 반영하고 있지 않다. 다음은 기존 품질평가 알고리즘 중 12차선을 모두 적용하고 있지 않은 부분이다. 알고리즘에 함수를 삽입하여 차선을 반영하고 있지만 12차선에 설치된 24개의 루프를 적용하고 있지 않고 고정 함수의 도입으로 인해 지역별 차선 변경에 대해 융통적이지 못하다.

```
BEGIN
    delete quality where data_code = 'DAT001' and
    index_code = 'IDX002';
    delete quality where data_code = 'DAT001' and
    index_code = 'IDX004';
    select count(*) into total_lds from lds30sec;

    select
    to_char(min(sample_time)),to_char(max(sample_time))
    into st_time,ed_time from lds30sec;

    sum_lds30 := f_lds_q('vol',p_s_time,p_e_time);
    v_1 := (sum_lds30/(total_lds*24))*100;
    dbms_output.put_line('total1== '||v_1);

    sum_lds30 := f_lds_q('occ',p_s_time,p_e_time);
    v_2 := (sum_lds30/(total_lds*24))*100;

    sum_lds30 := f_lds_q('speed',p_s_time,p_e_time);
    v_3 := (sum_lds30/(total_lds*12))*100;
```

(그림 1) 기존 품질평가 알고리즘 중 잘못된 차선의 반영과 고정 함수의 도입 부분

3.3. 품질평가 알고리즘의 개선

동적 PL/SQL을 이용하여 입력 테이블을 'v_from_table'로 파라미터화 함으로써 사용자의 요구에 따라 다양한 테이블에 대한 품질평가가 이루어질 수 있게 한다. 출력은 화면상의 수치로 나타나며 사용자가 알아보기 쉽도록 교통량(v), 점유율(o), 속도(s)의 각 값에 대한 완전성과 유효성이 퍼센트화되어 출력되게 한다. 또한 입력 데이터의 결측 자료를 반영하여 결측 자료를 제외한 데이터에 한하여 품질평가의 완전성, 유효성과 오류판단이 이루어 지도록 하였다. 결측

자료는 3장에서 설명하였듯이 검지기 시스템의 고장이나 차량의 부재로 인하여 데이터가 없는 경우를 말한다. 결국 자료를 제외한 데이터에 대하여 완전성과 유효성을 측정할 수 있도록 하며 실제 데이터 환경을 반영하기 위하여 모든 차선 개수를 적용함으로써 실 교통 상황에 적합하도록 개선하였다.

또한 지역별 차선 변경에 대해 융통적인 적용이 가능하도록 차선 개수를 차선 id에 의하여 반영하는 FIND_COUNT_OF_LOOP1 함수를 도입 하였다.

기존의 완전성과 유효성 계산식과는 다르게 개선 완전성, 유효성 계산식은 간단 명료하여 연구자의 알고리즘 분석에도 용이하다. 다음 (그림 3)은 품질평가 중 오류가 없는 자료의 비율을 나타내는 유효성을 계산하는 부분이다. 결국 자료인 -111을 제외한 자료 중 오류 판단 기준인 임계값 검사(Data Threshold)와 관계 검사(Data Relation)를 거쳐 정상인 데이터의 개수를 퍼센트화하고 있다. 기존 유효성 계산식과는 다르게 개선 유효성 계산식은 간단하고 흐름(flow)의 이해를 쉽게 한다. 또한 과거 유효성 판단에서 유효하지 않은 자료의 비율을 화면상 결과값으로 출력했던 것을 개선하여 순수 유효한 데이터의 비율을 결과값으로 출력하도록 하였다.

```
CREATE OR REPLACE PROCEDURE quality_1(v_input
VARCHAR2) IS
...
v_from_table VARCHAR2(32767);
...
BEGIN
v_from_table := substr(v_input, 1);
...
OPEN cur_input
FOR 'SELECT lds_id'
|| ' FROM '
|| v_from_table
;
LOOP
FETCH cur_input INTO v_cur_input;
EXIT WHEN cur_input%NOTFOUND;
v_sum := v_sum +
FIND_COUNT_OF_LOOP1(v_cur_input);
...
dbms_output.put_line('완전성 v : '||to_char(v_q1_v));
dbms_output.put_line('완전성 o : '||to_char(v_q1_o));
dbms_output.put_line('완전성 s : '||to_char(v_q1_s));
...
dbms_output.put_line('유효성 v : '||to_char(v_q2_v));
dbms_output.put_line('유효성 o : '||to_char(v_q2_o));
dbms_output.put_line('유효성 s : '||to_char(v_q2_s));
```

(그림 2) 개선된 품질평가 중 입력 테이블의 파라미터화

출력 값의 도출, 차선 함수 반영 부분

```
BEGIN
v_statement := 'SELECT count(NULLIF(loop_o0, -111))+
count(NULLIF(loop_o1, -111))+... -
111))+count(NULLIF(loop_o22, -
111))+count(NULLIF(loop_o23, -111)) FROM '
|| v_from_table;
EXECUTE IMMEDIATE v_statement INTO v_loop_count_o;
EXCEPTION WHEN NO_DATA_FOUND THEN
v_loop_count_o := 0;
END;
...
BEGIN
v_statement := 'SELECT count(case when
((NULLIF(loop_v0, -111)) < 0 or (NULLIF(loop_v0, -111))
> 30) or... *(NULLIF(loop_o23, -
111))/12>18*(NULLIF(loop_v23, -111)))) then 1 else null
end) FROM '
|| v_from_table;
EXECUTE IMMEDIATE v_statement INTO
v_loop_count_invalid;
EXCEPTION WHEN NO_DATA_FOUND THEN
v_loop_count_invalid := 0;
END;
v_q2_v := 100-
((v_loop_count_invalid/v_loop_count_v)*100);
v_q2_o := 100-
((v_loop_count_invalid/v_loop_count_o)*100);
v_q2_s := 100-
((v_loop_count_invalid/v_loop_count_s)*100);
```

(그림 3) 개선된 품질평가 중 결국 자료의 반영과 모든 차선의 적용, 간단한 계산식의 도입 부분

4. 자료 처리 기법

자료 처리 과정은 일반적으로 오류 판단 과정, 결국 보정 과정, 평활화 과정으로 이루어지며, 원시 자료의 품질이나 사용 목적에 맞게 3가지 과정을 다양하게 조합하여 적용할 수도 있다. 본 논문에서는 오류 판단 과정과 결국 보정 과정에 중점을 두었다.

4.1. 오류 판단

오류 판단이란 오류 데이터를 -999로 변경하여 정상 데이터와 오류 데이터를 구별 가능하도록 하는 자료처리 과정으로 오류 자료(invalid data)는 교통량, 속도, 점유율 등의 속성이 논리적으로 정상적인 상태의 교통류 특성을 나타내는 범위를 벗어난 교통 자료를 의미한다. 오류 판단 기준으로는 임계값 검사와 관계 검사가 있다. 오류 판단 알고리즘에서는 입력

파라미터를 추가하여 사용자의 요구 사항에 따른 오류 판단 기준을 반영할 수 있도록 하였다. 입력 테이블과 출력 테이블을 'v_from_table'과 'v_temp_table'로 파라미터화 하였고 v_max_v는 최대 교통량을, v_max_o는 최대 점유율을, v_difference_v는 쌍루프 간 교통량 차를 나타내는 파라미터로 디폴트 값은 v_max_v가 30, v_max_o는 100, v_difference_v는 2이다. 또한 결측 자료 -111을 제외한 데이터에 대하여 오류 판단을 수행하도록 하여 -111 값을 유지하도록 함으로써 오류 데이터 -999 값과 차별화 한다. 오류 판단도 품질 평가와 같이 모든 차선에 대하여 임계값 검사와 관계 검사를 적용하도록 하며 교통량이 0일 때 점유율이 0이면 오류로 판단되는 등 기존의 잘못된 판단 기준을 올바르게 개선하고 오류 판단 계산식에 IF 함수를 효율적으로 적용하여 사용자의 알고리즘 분석을 쉽게 하고 알고리즘 수정을 용이하게 하였다. 본 논문에서는 '고속도로 차량검지기자료 조사·분석 및 활용기법 개발' 보고서에서 제안한 도로공사 FTMS 오류 판단 기준 개선안[1]을 적용하여 아래 <표 1>의 오류 판단 기준에 부합하는 교통 자료를 오류 자료로 정의하였다.

<표 1> 한국도로공사 FTMS 오류 판단 기준 개선안

구분	오류 판단 기준
임계값 검사	교통량 < 0 또는 교통량 > 30
	인접 검지기와의 교통량 차 > 2
	점유율 < 0 or 점유율 > 100
관계 검사	교통량 = 0 and 속도 != 0
	$2.7V \leq \frac{S \times O}{12} \leq 18V$ (v 교통량, o 점유율, s 속도)

4.2. 결측보정

결측 보정 과정은 원시 자료 중 결손 자료와 오류 자료를 다른 시공간 유효 자료를 참조하여 새로운 값으로 보정하는 것으로 자료 처리 과정 중 가장 중요한 과정이다. 특히 동일한 시간 정보와 인접한 공간 정보를 가지는 교통 자료 또는 인접한 시간 정보와 동일한 공간 정보를 가지는 교통 자료는 비슷한 교통량, 속도, 점유율을 가지는 특성은 결측 보정 기법에 효율적으로 적용할 수 있다[1, 2].

결측 보정 기법은 크게 두 가지로 살펴 볼 수 있다. 동일한 시간 정보와 인접한 공간 정보를 가지는 교통 자료를 참조하는 인접 지정 참조 기법과 인접한 시간 정보와 동일한 공간 정보를 가지는 교통 자료를 참조하는 이력 자료 활용 기법이다.

다음 <표 2>은 결측 보정 기법을 참조 공간과 참조 시간에 따라 분류한 것이다.

<표 2> 결측 보정 기법

분류	결측 보정 기법	참조 공간	참조 시간
인접 지정 참조	전후 지정 동일 주기 적용	이전 지정, 이후 지정	결측 시점 (동일 날짜, 동일 시간)
	이전 지정 동일 주기 적용	이전 지정	결측 시점
	지정간 이동 소요 주기 적용	이전 지정	결측 시점 - 이동 소요 주기
이력 자료 활용	이력 자료 동일 주기 적용	결측 지정	동일 요일, 동일 시간
	차로별 이용률 적용 (교통량만 보정)	결측 지정	동일 요일, 동일 시간

다음 <표 3>는 결측 보정 기법을 사용자가 다양하게 교통 자료에 적용시킬 수 있도록 추가한 입력 변수의 목록을 보여준다.

<표 3> 결측 보정 기법의 입력 변수

분류	결측 보정 기법	입력 변수
인접 지정 참조	전후 지정 동일 주기 적용	최대 x 번째 전후 지정 참조
	이전 지정 동일 주기 적용	최대 x 번째 이전 지정 참조
	지정간 이동 소요 주기 적용	최대 x 번째 이전 지정 참조
이력 자료 활용	이력 자료 동일 주기 적용	결측 시점 기준, 과거 x주 동안 이력 참조
	차로별 이용률 적용 (교통량만 보정)	과거 x주 동안 지정 교통량 참조, 최근 y주기 동안 차로 이용률 참조

4.2.1. 인접 지정 참조

인접 지정 참조 기법은 차량의 유입과 출력이 없는 폐쇄된 구간 내의 교통 자료는 서로 유사성을 가지는 특성을 이용하여 결측 지정과 인접한 지점의 유효 자료를 참조하는 기법으로 전후 지정 동일 주기 적용, 이전 지정 동일 주기 적용, 지정간 이동 소요 주기 적용 기법이 있다. 결측 지점을 통과한 차량은 반드시 참조할 인접 지점을 통과했다는 가정이 포함된 기법으로, 결측 지정과 참조할 인접 지정 사이에 차량의 유입이나 출력이 있으면 적용할 수 없는 기법이다.

아래 (그림 4)은 고속도로 왕복 6차선의 인접한 지정 3개에 설치된 검지기의 예를 보여준다.

전인 2007년 8월 17일 13시 00분 00초가 된다.

고속도로					
3차로	5	4		5	4
2차로	3	2		3	2
1차로	1	0		1	0
1차로	7	6		7	6
2차로	9	8		9	8
3차로	11	10		11	10
	지점1			지점2	
				지점3	

(그림 4) 고속도로 왕복 6차선의 인접한 지점 3개에 설치된 검지기 예시

고속도로 지점2, 2007년 08년 17일 13시 00분 00초					
3차로	5	4		5	4
2차로	3	2		3	2
1차로	1	0		1	0
1차로	7	6		7	6
2차로	9	8	13:00:00	9	8
3차로	11	10		11	10
	지점1			지점2	
				지점3	

(그림 5) 지점간 이동 소요 주기 적용 예시

<표 4>, <표 5>, <표 6>은 (그림 4)의 고속도로 지점2에서 2007년 8월 17일 13시 01분 00초에 0번 검지기와 9번 검지에서 결측 자료가 발생하였을 때, 각각의 인접 지점 참조 기법을 적용하면, 어느 지점의 몇 번 검지기를 참조하는지 보여준다. 예시에서 사용한 결측 보정 기법의 입력 변수는 모두 1이다.

<표 4> 전후 지점 동일 주기 적용 예시

결측 지점	결측 검지기	참조 지점	참조 검지기
2	0	1	0
		3	0
2	9	1	9
		3	9

<표 5> 이전 지점 동일 주기 적용 예시

결측 지점	결측 검지기	참조 지점	참조 검지기
2	0	3	0
2	9	1	9

<표 6> 지점간 이동 소요 주기 적용 예시

결측 지점	결측 검지기	이동 소요 주기	참조 지점	참조 검지기
2	0	1	3	0
2	9	2	1	9

(그림 5)는 <표6>를 그림으로 표현한 것이다. 지점2의 0번 검지기는 이동 소요 주기가 1주기로 계산되어 참조시점은 결측시점-1주기, 즉 2007년 8월 17일 13시 01분 00초보다 30초 전인 2007년 8월 17일 13시 00분 30초가 된다. 지점2의 9번 검지기는 이동 소요 주기가 2주기로 계산되어 참조시점은 결측시점-1주기, 즉 2007년 8월 17일 13시 01분 00초보다 1분

4.2.2. 이력자료 활용

이력 자료 활용 기법은 주로 인접 지점 참조 기법으로 결측 보정을 할 수 없는 경우 사용하는 기법으로, 동일 시간대의 교통 자료는 서로 유사성을 가지는 특성을 이용하여 이력 자료(즉 결측 시간 이전의 교통 자료)의 유효 자료를 참조하는 기법으로 이력 자료 동일 주기 적용, 차로별 이용률 적용 기법이 있다.

이력 자료 동일 주기 적용 기법은 결측 지점의 과거 교통 자료 중 결측 시점과 동일 요일, 동일 시간의 유효 자료를 참조하여 결측 자료를 보정하는 기법이다.

아래 <표 7>은 2007년 8월 17일 13시 00분 30초에 결손된 교통 자료를 이력 자료 동일 주기 적용 기법을 이용하여 보정한 결과와 보정에 참고한 이력 자료이다. 아래 예시는 이력 자료의 기간을 3주로 한 것이다.

<표 7> 이력 자료 동일 주기 적용 예시

구분	시간 정보 (년/월/일/ 시/분/초)	0번 검지기		
		교통량 (대)	점유율 (%)	속도 (km/h)
이력 자료	2007/07/27/ 13/00/30	1	1	100
이력 자료	2007/08/03/ 13/00/30	1	1	90
이력 자료	2007/08/10/ 13/00/00	1	1	110
보정 전	2007/08/17/ 13/00/30	-999	-999	-999
보정 후	2007/08/17/ 13/00/30	1	1	100

차로별 이용률 적용 기법은 교통량만 대상으로 하는

결측 보정 기법으로 결측 시점 이전의 최근 차로 이용률과, 이력 자료를 활용하여 얻은 지점 교통량을 이용하여 결측 자료를 보정하는 기법이다.

아래 <표 8>, <표 9>, <표 10>는 2007년 8월 17일 13시 00분 30초에 결손된 교통 자료를 차로별 이용률 적용 기법을 이용하는데 참고한 이력 자료 및 최근 자료와 교통량을 보정한 결과이다. 아래 예시는 지점 교통량을 계산하는데 이력 자료의 기간을 3주로 하였으며, 차로 이용률을 계산하는데 필요한 최근 자료의 기간을 4주기(2분)로 한 것이다.

<표 8> 차로별 이용률 적용 예시 - 이력 자료

구분	시간 정보	교통량(대)			
		1차로	2차로	3차로	지점
이력 자료	2007/07/27/ 13/00/30	1	1	1	3
이력 자료	2007/08/03/ 13/00/30	1	1	1	3
이력 자료	2007/08/10/ 13/00/00	2	1	0	3

<표 9> 차로별 이용률 적용 예시 - 차로 이용률

구분	시간 정보	1차로	2차로	3차로	차로 이용률
4주기 전	2007/08/17/ 12/58/30	3	1	1	0.6
3주기 전	2007/08/17/ 12/59/00	2	1	0	0.7
2주기 전	2007/08/17/ 12/59/30	2	1	0	0.7
1주기 전	2007/08/17/ 13/00/00	1	1	1	0.3

<표 10> 차로별 이용률 적용 예시

구분	시간 정보	0번 검지기		
		교통량(대)	정유율(%)	속도(km/h)
보정 전	2007/08/17/ 13/00/30	-999	-999	-999
보정 후	2007/08/17/ 13/00/30	2	-999	-999

5. 결론

방대한 교통 자료에 대하여 신뢰도를 판단하고 유효한 자료를 선별하여 활용도를 높이는 것이 중요하다. 기존의 품질 평가와 오류 판단 알고리즘을 개선하여 간단 명료한 계산식을 도입하고 실제 교통 상황을 반영하여 실질적인 알고리즘의 수행을 가능하게

하였으며 잘못된 오류 판단 기준을 수정하여 정확한 결과값을 도출하도록 하였다. 또한 입력 교통 데이터의 결측 자료를 반영하여 품질 평가의 완전성과 유효성이 올바르게 계산되도록 하였고 오류 판단을 적용할 때, 결측 자료와 별도로 오류 데이터와 정상 데이터를 구별하여 연구자에게 다양한 정보를 정확하게 제공할 수 있도록 하였다. 향후에는 변화하는 교통 환경을 반영하여 올바른 품질 평가와 오류 판단이 이루어질 수 있도록 지속적인 수정 작업이 요구된다.

기존에 제안되었던 결측 보정 기법에 입력 변수를 추가하여 사용자가 원시 자료의 사용 목적이나 품질에 맞게 동일한 기법이라도 다른 입력 변수를 이용하여 여러 가지 보정 결과를 얻을 수 있도록 구현하였다. 또한 교통 자료가 저장되는 데이터베이스로부터 원시 자료를 꺼내와 다른 프로그램을 이용하여 보정하는 것이 아니라, PL/SQL을 이용하여 교통 자료가 저장되는 데이터베이스를 직접 접근하여 보정하도록 구현하였다. 원시 자료를 데이터베이스로부터 꺼내와 MATLAB과 같은 다른 프로그램을 이용하여 보정하면, 그 결과를 관리하거나, 다른 교통 정보로 가공하고, 그 정보를 다른 연구자와 공유하는데 어려움이 따른다. 하지만 교통 자료가 저장되는 데이터베이스에서 직접 자료 처리 과정을 수행할 수 있게 된다면, 그 데이터베이스를 이용하는 교통 자료 관리 시스템과 연계하여 연구자 누구나 원시 자료를 이용하여 다양한 자료 처리 과정을 수행할 수 있으면서, 그 처리 결과를 공유함은 물론이며, 구간별 교통상황과 같은 다양한 교통 정보 제공 서비스를 받을 수 있다.

향후 연구로는 PL/SQL로 구현된 코드를 최적화하여 수행 시간을 단축시키는 것과 교통 자료를 위한 자료 처리 시스템으로 확장 구현하는 것이 있다. 대용량 데이터에 대한 SQL 수행 방법을 연구하여 코드에 반영하면 수행 시간을 단축시킬 수 있을 것이다. 또한 다른 분야의 자료 처리 방법을 분석하여 교통 자료에 적용시킬 만한 것이 있는지 살펴 보는 것도 좋은 연구 방향이라 할 수 있다.

참고문헌

- [1] 한국도로공사, '고속도로 차량검지기자료 조사·분석 및 활용기법 개발', 한국ITS학회, 2006.
- [2] 김정연, 이영인, 백승걸, 남궁성, "차량 검지기자료 결측 보정처리에 관한 연구(이력자료 활용방안을 중심으로)", 대한교통학회지, 제24권 제7호, pp.27-40, 2006.
- [3] ITS사업실, 도로교통기술원 교통연구그룹, 'ITS 구축·운영 편람', 한국도로공사, 2005.