

교통 이력 분석을 통한 교통정보 우선순위 결정 시스템

이병우, 조현성, 이현정, 오병화, 양지훈
서강대학교 컴퓨터학과
e-mail : elva1212@naver.com

Prioritized Traffic Information Delivery Based on Historical Data Analysis

Byungwoo Lee, Hyunsung Jo, Hyunjung Lee, Byonghwa Oh, Jihoon Yang
Dept. of Computer Science, Sogang University

요 약

본 논문에서는 교통 이력 데이터 분석을 통해 운전자에게 유용한 정보를 식별하는 방법을 제안한다. 이를 위해 차량 속도 분석을 이용한 요일, 시간 별 도로 중요도, 도로속성을 이용한 도로 중요도 결정 시스템을 개발하였다. 또한, 돌발상황 발생 시에도 그 예측 영향범위에 따른 선별적 정보 제공이 가능한 시스템을 개발하였다.

1. 서론

기존의 교통 정보 시스템은 상습정체구간을 중심으로 고정적이고 반복적으로 이루어지는 방식이 대부분이다. 그러나 도로의 중요도가 교통량과 비례한다고 보았을 때, 기존의 방법은 시시각각 변화하는 교통의 흐름을 반영하기 힘들다는 단점이 있다. 본 논문에서는 교통 이력 데이터의 요일, 시간 별 분석을 통해 통행속도를 이용해 도로 우선순위를 결정하는 방법과 도로속성에 따른 우선순위 결정법을 제안한다. 마지막으로, 예기치 못한 돌발상황이 발생한 경우에도 그 영향의 범위를 예측하고 그에 따른 정보의 우선순위를 결정할 수 있는 시스템을 소개한다.

2. 교통 데이터

건설교통부에서는 지능형교통체계(ITS) 호환성과 상호연계 운용 효율성을 확보하고, 대국민서비스 정보제공 편의증진을 위하여 표준 노드(Node), 링크(Link)를 교통정보 실시간 교환에 기본이 되는 전국 단위로 사용한다. 노드란 차량이 도로를 주행함에 있어서 속도의 변화가 발생하는 곳을 표현한 것이다.

예를 들어, 교차로, 교량 시종점 등을 말한다. 링크란 속도변화 발생점이 노드와 노드를 연결한 선을 의미하며 실세계에서의 도로이다. 예를 들어 도로, 교량 등을 말한다[1].

각 링크는 권역정보, 통행제한차량, 최고제한속도 등과 같은 고유의 속성값을 가지고 있다. 이러한 링크 속성 데이터는 건설교통부 건설교통종합정보센터에서 제공하고 있다[1]. 또한 본 연구에서는 2006년 9월 11일부터 2006년 12월 10일간의 서울시 링크의 속도 데이터를 이용하였다.

3. 링크의 우선순위 분석

이 장에서는 운전자에게 우선적으로 제공되어야 하는 링크의 우선순위를 정하는 방법을 제안한다. 우선 통행속도기록과 도로속성을 분석하여 요일, 시간 별 도로 중요도를 결정하였고, 유고 데이터를 이용하여 돌발상황의 영향 범위를 예측할 수 있는 시스템을 개발하였다.

3.1. 차량 속도를 이용한 우선순위 분석

이 절에서는 통행 패턴을 분석하고 이를 현재 교통 상황과 비교함으로써 링크의 우선순위를 정하는 방법을 제안한다.

※ 본 연구는 한국과학재단 특정기초연구 (R01-2004-000-10689-0) 지원으로 수행되었음

3.1.1. 교통 이력 데이터의 통계적 분석

평일에는 아침 출근 시간, 퇴근 시간의 교통량이 가장 많다. 휴일에는 나들이 차량으로 항상 교통량이 많다. 이처럼 평일과 휴일의 교통 상황은 서로 다르다. 또한 평일이라도 요일마다 교통상황이 다르다. 따라서 통행 패턴을 분석하기 위해서는 요일 별로 나눠 분석해야 한다. 또한 평일이지만 공휴일인 경우는 해당 요일과는 별도로 분석해야 한다.

평균 속도는 속도의 중심성향, 즉 평소의 통행 패턴을 보여준다. 또한 표준편차는 평균 속도가 같더라도 평소 속도 변화가 심한 링크와 그렇지 않은 링크를 구별해 준다. 따라서 본 연구에서는 통행 패턴을 보여주는 수치로 평균과 표준편차를 사용했다. 그러나 평소의 통행 패턴을 보여주기 위한 것이므로 공사나 사고와 같은 돌발상황은 제외시켜야 한다. 따라서 3.3에서 설명한 이상데이터는 제외시켰다.

3.1.2. 현재 속도를 이용한 우선순위 부여 방법

평균 속도는 과거 이력 데이터로부터 얻은 예측된 속도이다. 실제 속도가 평균 속도보다 느리다면 평소와는 달리 차량 소통을 방해하는 요인이 발생했다는 것이므로 운전자에게 우선적으로 제공되어야 한다. 그렇지만 단순히 속도 차만을 고려한다면 평소에도 평균 속도와 속도 차가 큰 링크와 작은 링크는 같은 우선순위를 갖게 된다. 실질적으로 후자는 상대적으로 전자에 비해 평소와 차이가 크다. 따라서 평균과 실제 데이터의 차이를 보여주는 표준편차도 고려해야 한다.

위와 같은 방법으로 우선순위를 정하면 또 다른 문제가 발생한다. 극단적인 예로, 평균 속도와 현재 속도의 차가 10 이고, 표준편차가 15 인 경우와 속도 차가 1 이고, 표준편차가 0.5 인 경우를 생각해 보자. 전자는 속도 차가 표준편차보다도 작기 때문에 후자보다 우선순위가 낮다. 그러나 후자는 속도 차가 겨우 1 밖에 되지 않는다. 따라서 이러한 문제점을 해결하기 위해 그 시간대 전체 링크의 표준편차의 평균을 고려해준다.

$$\frac{C_v - M_v}{\delta_v + A_\delta} \quad (1)$$

C_v 는 현재 속도, M_v 는 평균 속도, δ_v 는 표준편차, A_δ 는 전체 링크의 표준편차의 평균이다. 전체 링크의 표준편차를 사용함으로써 전체 링크 중에서 상대적으로 표준편차가 작은 링크가 속도 차가 작은데도 불구하고 우선순위가 높아지는 것을 방지한다.

3.2. 링크 속성을 이용한 우선순위 분석

고가차도와 일반도로의 차량 속도는 다를 것이다. 또한 이륜차의 통행이 제한되는 도로와 모든 차량이 통행 가능한 도로의 차량 속도도 다를 것이다. 이처럼 각 도로의 특징에 따라 차량의 속도가 다르다. 이

처럼 각 링크가 가지고 있는 속성값은 차량 속도에 영향을 준다. 그러나 그 영향력은 속성마다 다를 것이다. 이 장에서는 각기 다른 링크 속성의 영향력을 이용하여 링크의 우선순위를 정하는 방법을 제안한다.

3.2.1. 차량 속도와 링크 속성간의 연관성 분석

확률이론과 정보이론(Information Theory)에서 두 확률변수의 상호 정보라 함은 두 확률변수 사이의 상호 의존도의 정도를 수치적으로 표현한 것을 의미한다. 수학적으로 두 이산확률변수(Discrete Random Variables) X 와 Y 사이의 상호 정보는 아래와 같이 정의된다.

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (2)$$

여기서 $p(x,y)$ 는 X 와 Y 의 결합확률분포함수(Joint Probability Distribution Function)이고, $p(x)$ 와 $p(y)$ 는 각각 X 와 Y 의 주변확률분포함수(Marginal Probability Distribution Function)를 의미한다[2][3][4].

결과적으로 상호 정보는 X 와 Y 가 공유하는 정보의 양을 측정하는 값이다. 즉, 두 변수 중 하나의 값을 알았을 때, 그 정보가 다른 한 변수의 값을 파악함에 있어 얼마나 많은 정보를 제공하는지에 대한 측정값이 바로 상호 정보이다. 따라서 연관성이 많을수록 상호 정보가 높은 값을 가지게 된다.

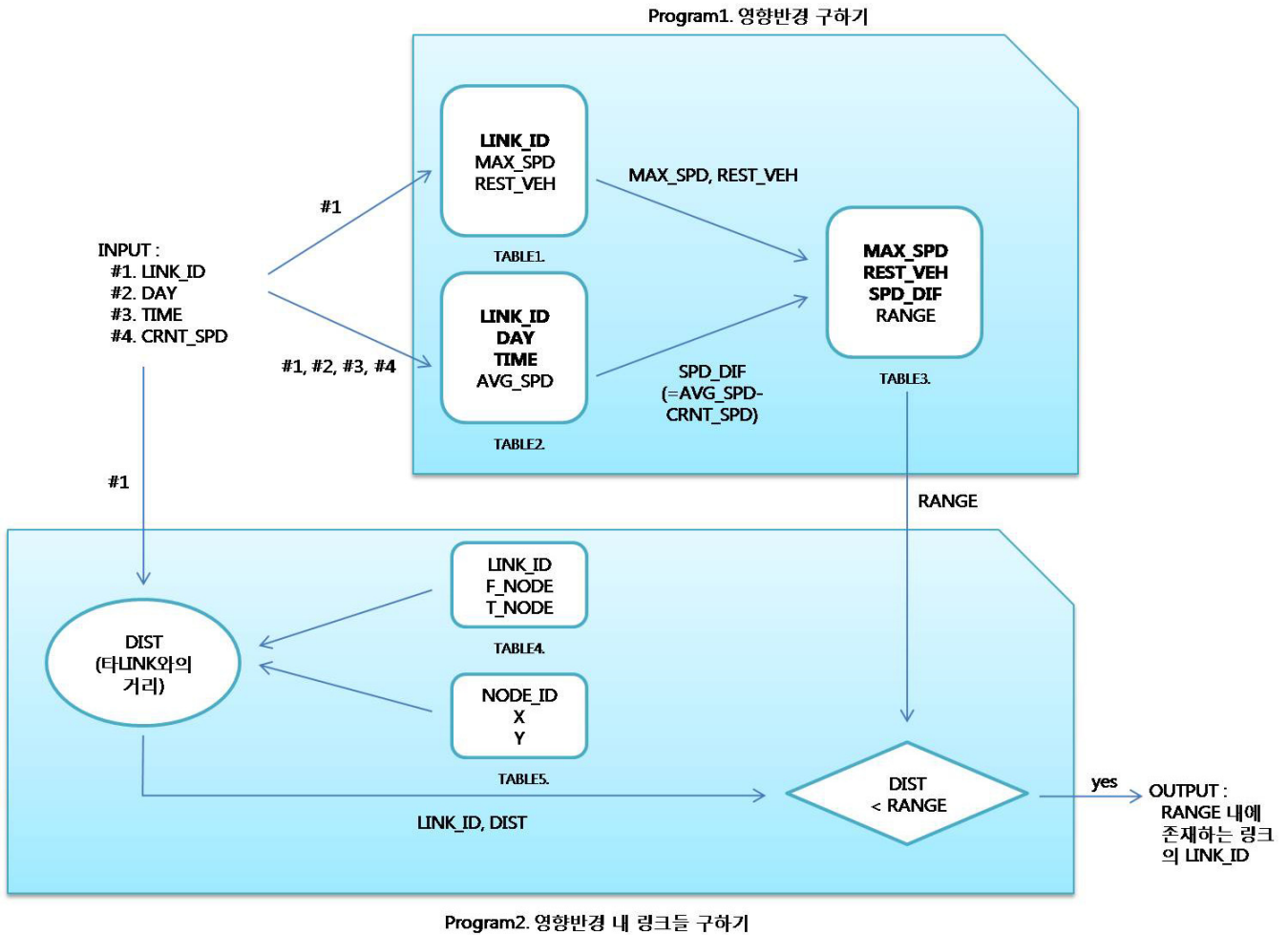
본 연구에서는 차량 속도와 링크 속성간의 연관성을 분석하기 위해 상호 정보(Mutual Information)을 사용한다.

3.2.2. 링크 속성을 이용한 우선순위 부여 방법

상호정보가 높은 속성을 가지고 있는 링크는 상호 정보가 낮은 속성을 가지고 있는 링크보다 속도 변화에 민감하므로 우선순위를 높여주어야 한다. 앞에서 설명했듯이 각 링크는 여러 개의 속성을 가지고 있다. 따라서 링크 속성들의 상호정보의 합을 구하고 이 값이 클수록 높은 우선순위를 부여한다.

3.3. 유고 데이터를 이용한 우선순위 분석

교통사고나 천재지변 등과 같이 사전에 예측하기 힘든 돌발상황의 경우 대응전략을 구축하기가 까다로울 뿐만 아니라 그 비용도 막대하다. 따라서, 본 절에서는 유고지점으로부터의 교통영향 범위에 대한 분석을 통해 돌발상황 발생 시 적절하고 신속하게 대응할 수 있는 교통영향반경 예측시스템 개발의 토대를 마련하였다. 우선, 교통 이력 데이터로부터 유고지점 및 그 영향권으로 추정되는 링크들을 수집하여 유고의 영향 범위에서 보이는 패턴을 분석하였다. 그리고 혼잡의 반경에 가장 큰 영향을 끼친다고 판단되는 도로 속성들을 추출하였다. 이를 바탕으로 임의의 링크에 돌발상황이 발생하였을 때, 그 영향권 내의 링크를 식별하는 시스템을 개발하였다.



(그림 1) 돌발상황 교통영향반경 예측 시스템

*LINK_ID: 링크 ID, DAY: 요일, TIME: 시간, MAX_SPD: 최대속도, RANGE: 영향반경,
REST_VEH: 통행제한차량, AVG_SPD: 평균속도, CRNT_SPD: 현재속도
** TABLE 1,2,3 에서 굵은 글씨는 Key 값

3.3.1. 유고 데이터의 수집 분석

본 항에서는 가상의 유고지점과 그 영향범위를 조사하여 영향반경 예측시스템 구축의 토대를 마련하였다. 그리고 영향범위와 도로속성의 관계를 추출하였다. 이로부터 얻은 다양한 정보를 바탕으로 이전의 돌발상황에 대한 정보가 없더라도 영향 범위를 예측할 수 있게 하였다.

이를 위해 우선, 날짜 별, 링크 별로 10분마다 측정한 속도 데이터 중 요일-링크 평균 속도에 비해 30km/h 이상 느린 구간들을 ‘이상 데이터’라고 규정하였다. 여기서 기준이 되는 30km/h 라는 값은 ‘의미 있는 분석이 가능하면서도 운전자들이 일반적으로 정체를 인식할 수 있는 평소와의 속도 차’를 표현한 수치이다. 추출된 53,655 개의 이상데이터 중 연속된 시간 단위에서 반복적으로 나타나는 링크들을 뽑아서 직, 간접적으로 연결 돼 있는 링크의 집합들을 구하였다. 이 중에서 최하류의 링크를 유고지점으로 지정하고 나머지를 영향링크로 규정하였다.

이와 같은 유고 데이터는 영향 범위에 대한 분석의

기반이 되었다. 우선 각 유고지점의 영향반경을 수집하였고, 이 반경에 가장 밀접하게 연관된 속성값으로서 통행제한차량, 제한속도, 평균-현재 속도 차를 추출했다. 이 분석의 결과는 최종 시스템에서 각 속성값에 따른 영향범위를 구하는 데에 이용되었다.

3.3.2. 돌발상황 교통 영향반경 예측시스템

(그림 1)는 ‘돌발상황 교통영향반경 예측 시스템’의 전체적인 알고리즘을 표현한 그림이다. 도로상에서 돌발상황이 벌어지면 발생 시점의 링크 ID, 요일, 시간, 현재 속도에 대한 정보가 실시간으로 시스템에 입력된다. 시스템의 흐름은 크게 Program1 과 Program2 의 순차적인 두 부분으로 나뉘어지는데, Program1 은 입력으로부터 유고가 발생한 링크 ID, 요일, 시간, 현재속도 정보를 받아서 그에 따른 속성값을 추출하는 작업과 그 속성값들로부터 돌발상황의 영향반경을 계산하는 작업을 수행한다. 유고지점의 제한속도와 통행금지차량에 대한 정보는 링크 ID 와 TABLE1 을 이용하여 구하며, 평균속도와 현재속도의 차이는 TABLE2 에서

해당 링크, 요일, 시간의 평균 통행속도를 구하여 그 값에 현재속도를 감산하여 구한다. TABLE3 는 이상 데이터를 토대로 작성한 속성-영향반경 표이다. Program1 은 구해진 제한속도, 통행금지차량, 평균-현재속도 차의 값을 이용하여 해당하는 위치의 영향반경 값을 TABLE3 상에서 찾아서 출력한다.

Program2 는 유고지점 링크로부터 Program1 에서 출력된 예상 영향반경 이내에 있는 링크들을 찾아서 그 링크 ID 들을 사용자에게 되돌려준다. 구체적으로 말하면, 사용자로부터 유고지점의 링크가 입력되면 TABLE4 의 노드 정보와 TABLE5 의 위도, 경도 정보를 이용하여 해당 링크와 다른 모든 링크 사이의 유클리드 거리가 계산된다. 그 결과들 중 Program1 의 출력 예상 영향반경보다 짧은 거리에 있는 링크들이 사용자에게 출력된다.

4. 결론

효과적인 교통정보 서비스 제공 시스템은 우리나라의 도로 실정에 부합해야 하며 시시각각 변화하는 도로 상황에 가장 효과적인 대책을 제공할 수 있어야 한다.

본 논문에서는 시간 별, 요일 별로 교통 흐름 패턴을 파악하여 보다 실용적인 결과를 얻었다. 이를 현재 교통 흐름과 비교·분석하여 운전자에게 유용한 정보를 식별하였다. 또한 링크 속성과 속도와의 연관성 분석을 통해 링크의 우선순위를 정하였다. 마지막으로 돌발상황이 후방교통상황에 미치는 영향을 파악함으로써 적절한 대응 수준을 결정할 수 있게 하였다.

이를 바탕으로 변화하는 교통상황에 맞춰 운전자가 필요로 하는 유용한 정보를 식별할 수 있다.

참고문헌

- [1] 건설교통부 건설교통종합정보센터, "http://road.moct.go.kr/STD_NodeLink_Intro.aspx".
- [2] Guiasu, Silviu. "Information Theory with Applications". McGraw-Hill, 1997.
- [3] Athanasios Papoulis. "Probability, Random Variables, and Stochastic Processes" 2nd Ed. McGraw-Hill, 1984.
- [4] Duda R., Hart P., and Stork D. "Pattern Classification" 2nd Ed. Wiley-interscience, 2000.