

Fuzzy c-means 알고리즘을 이용한 TCS 데이터 주행특성 분류 방법 연구

박원식*, 김동근*, 양영규*
“경원대학교 전자계산학과”

e-mail: pws1981@paran.com, cocacu@paran.com, ykyang@kyungwon.ac.kr

Driving Characteristics Classification of TCS Data Based on Fuzzy c-means Clustering Algorithm

Won-Sik Park*, Dong-Keun Kim*, Young-Kyu Yang*
*Dept of Computer Science, Kyungwon University

요 약

현재 사용되고 있는 통행시간 분류방법은 하나의 통행시간을 대푯값으로 가지고 있다. 이에 문제점은 고속도로 특성으로 규정 속도 이상의 속도로 주행하는 차량, 규정 속도 및 휴게소 이용차량, 운전자의 운전 습성, 통행 목적, 피로의 정도, 운전자 성향과 도로상황에 따라 통행시간이 다르게 나타나는 점이다. TCS(Toll Collection System) 자료는 고속도로의 다양한 특성이 포함되어 있으며, 대상 구간의 거리가 멀수록 목적지에 도달하는 통행시간의 분산이 커지는 특성 또한 보인다. 따라서 이를 처리하기 위한 효율적인 통행시간 분류, 구간대표통행시간 추출 알고리즘이 필요하다. 기존의 방법은 전체 통행차량의 통행시간을 감안한 방법으로 통행시간 예측시 정확성이 저하된다. 본 연구에서는 TCS 자료를 Fuzzy c-means 알고리즘을 이용하여 일일 고속도로 통행시간의 시간별 주행특성을 고려한 대푯값을 추출하는 알고리즘을 제안하였다. 실제 서울-청주 구간을 운행한 TCS 자료를 가지고 실시한 실험으로, 주행특성 및 도로상황을 고려한 Fuzzy c-means를 이용한 통행시간 분류방법과 기존의 통행시간 분류 방법을 통한 통행시간을 PIFAB를 사용 TCS 자료의 실제 통행시간과 경로통행시간을 비교 평가하였다. 평가한 결과 본 연구에서 제안하는 Fuzzy c-means기법은 기존 방법인 MAD기법보다 75%, 신뢰구간(95%) 추출법 대비 81%의 정확성을 제고하였다.

1. 서론

최근 고속도로 통행시간 교통정보의 제공에 대한 운전자들의 관심이 주 5일 근무 실시 및 유류비용의 증가, 시간가치에 대한 관심의 증대 등으로 어느 때 보다 커지고 있는 실정이다.

고속도로 통행료수납시스템(TCS: Toll Collection System) 자료는 전국의 고속도로를 통행한 모든 차량의 주행경험이 담겨있는 전수자료로서 교통정보제공에 있어서 그 활용도와 정확도가 매우 크다 할수 있겠다. 이에 따라서 최근에는 TCS 데이터에 관한 연구가 활발히 진행되고있다. TCS 자료는 전국 262개 영업소에서 하루 평균 약 320만대의 고속도로를 이용하는 차량을 대상으로 수집하는 전수 자료로서 고속도로를 통행하는 모든 차량의 출발, 도착 톨게이트 간 통행시간 정보를 담고 있다. 즉 TCS 자료는 고속도로 이용차량이 경험한 실제통행시간이며 ‘구간정보’이다. 차량이 경험한 TCS 통행시간은 해당구간을 통과하기까지의 정체상황 변화를 반영한 결과로서 최적경로선택을 위한 가장 직접적인 정보가 되며 이용자 입장에서 유용한 교통정보 산출에 적합하다는 특징을 가지고 있다.

현재까지는 고속도로 구간의 통행시간 분류 선정 시 각 구간별로 하나의 대푯값으로 분류한다. 이의 문제점은 고속도로 특성으로 규정 속도 이상의 속도로 주행하는 차량, 규정 속도 및 휴게소 이용차량, 운전자의 운전 습성, 통행 목적, 피로의 정도, 도로상황에 따라 통행시간이 다르게 나타나는 점이다. TCS 자료는 고속도로의 다양한 특성이 포함되어 있으며, 대상 구간의 거리가 멀수록 목적지에 도달하는 통행시간의 분산이 커지는 특성을 보인다. 이는 운전자의 주행특성 및 고속도로 상황을 고려하지 않고 변화하는 도로상황 하에 기존의 통행시간 분류방법으로 선정된 대푯값이 TCS 데이터의 특성을 파악하지 못해 정확성이 떨어진다는 것이다. 또한 시간대 별 통행량과 고속도로 이용목적이 다른 관계로 각각 다른 방법으로 통행시간 분류를 하여야 하는데 지금까지는 단순히 통계학적인 분류방법으로 이러한 요소를 반영하지 못해 정확한 통행시간 분류를 하지 못하였다.

고속도로 통행시간 정보의 정확도와 신뢰성에 대한 요구가 높아져 가고 있는 현실에서 운전자가 효과적인 통행 의사결정을 하기 위한 진정한 가치를 지닌 ‘정보다운 정보’를 만들어내는 일이야 말로 매우 중요한 일이다.

이에 본 연구에서는 TCS 자료를 이용하여 보다 정확하고 신뢰성 있는 고속도로 통행시간 제공을 위해 Fuzzy c-means를 이용한 주행특성 분류 방법을 제안한다.

2. 관련연구

본 연구에서 적용한 Fuzzy c-means 알고리즘과 비교 평가를 위하여 사용된 현재 통행시간 분류 방법인 중위절대편차 그리고 경로를 구성하고 있는 링크 통행시간의 문제점을 해결하기 위한 전진반복 전후방탐색법에 대하여 자세히 설명한다.

2.1 중위절대편차(MAD: Median Absolute Deviation)

수집된 데이터의 중앙값(median)을 이용하여 비정상적으로 크거나 작은 값들을 걸러내는 방법으로써 중위절대편차를 계산하는 기본 식은 다음과 같다.

$$MAD = b \times \frac{1}{N} \times \sum_{i=1}^N |x_i - x_{med}| \quad (식 1)$$

위의 식을 통해 MAD를 선정하고, 이상치를 포함한 데이터에 대해 Robust Standard Units z_i^{MAD} 를 계산한 후 그 값을 제거변수(cut-off value) z_{out} 과 비교하여 이상치를 판별하게 된다. z_i^{MAD} 를 계산하는 식은 다음과 같다

$$z_i^{MAD} = \frac{x_i - x_{med}}{MAD} \quad (식 2)$$

z_i^{MAD} 이 z_{out} 보다 크면, x_i 는 이상치로 판별되어 제거 대상이 된다. 일반적으로 z_{out} 은 3으로 설정한다.

2.2 Fuzzy c-means 알고리즘

본 연구에서 적용한 Fuzzy c-means 알고리즘은 고속도로 해당 구간의 주행특성을 고려한 통행시간을 추출하기 위하여 고속도로 통행차량은 고속으로 주행하는 차량그룹, 중간속도로 주행하는 차량그룹, 휴게소에서 휴식한 차량을 포함한 매우 저속으로 달리는 주행그룹 3개 그룹으로 나누어 대푯값을 선정한다. Dunn in이 제안한 Fuzzy c-means 알고리즘은 다음과 같이 요약된다.

Step 1 : 클러스터의 개수 c ($2 \leq c \leq n$)을 정하고 지수의 가중(exponential weight) m ($1 < m < \infty$)을 선택한다. 초기 소속 함수를 초기화한다. 알고리즘 반복 횟수를 표시한다.

Step 2 : 퍼지 클러스터 중심 $\{V_i | i = 1, 2, \dots, c\}$ 을 계산한다.

Step 3 : 다음과 같이 새로운 소속 함수 $U^{(r+1)}$ 을 계산한다.

$$u_{ik}^{(r+1)} = \frac{1}{\sum_{j=1}^c \frac{d_{ijk}^m}{d_{jk}^m}} \quad \text{for } I_k = \emptyset \quad (식 3)$$

또는

$$u_{ik}^{(r+1)} = 0 \quad \text{for all classes } i, \text{ 여기서 } i \in \tilde{I}_k$$

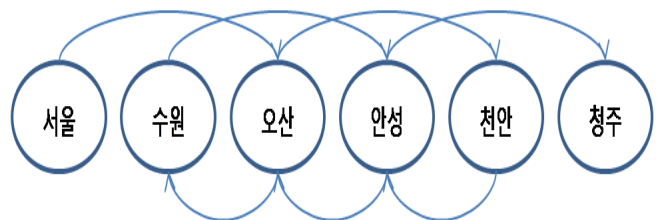
여기서, $I_k = \{i | 2 \leq c < n; d_{ik}^{(r)} = 0\}$ 이고 $\tilde{I}_k = \{1, 2, \dots, c\} - I_k$ 이다. 그리고 $\sum_{i \in \tilde{I}_k} u_{ik}^{(r+1)} = 1$

Step 4 : 다음 식을 계산해서 만일 $\Delta > \epsilon$ 이면 $r = r + 1$ 로 정하고 <Step 2>로 가서 다시 알고리즘을 반복 수행하고 그렇지 않고 $\Delta > \epsilon$ 이면 알고리즘을 종료한다. 여기서 ϵ 는 임계값.

$$\Delta = \| U^{(r+1)} - U^{(r)} \| = \max_{i,k} |u_{ik}^{(r+1)} - u_{ik}^{(r)}| \quad (식 4)$$

2.3 전진반복 전후방탐색법(PIFAB: Progressive Iterative Forward And Backward search)

전진반복 전후방탐색법은 출발시각 S_0 이 주어지면, 그것을 a-c구간에서의 출발시각으로 하는 도착시각 d_{ac} 를 탐색한다. 다음으로 b-c구간에서 d_{ac} 를 도착시각 d_{ac} 로 갖는 출발시각 S_{bc} 로 하는 도착시각 d_{bd} 를 탐색한다. 여기서 d_{bd} 는 결국 S_0 를 출발시각으로 할 때, a-d구간의 도착시각 d_{ad} 가 되며, 통행시간은 도착시각과 출발시각의 차인 $d_{bd} - S_0$ 가 된다. 전진반복 전후방 탐색법은 연이은 3개단위 링크를 하나의 단위로 하여 도착 톨게이트에 도달할 때까지 반복된다.

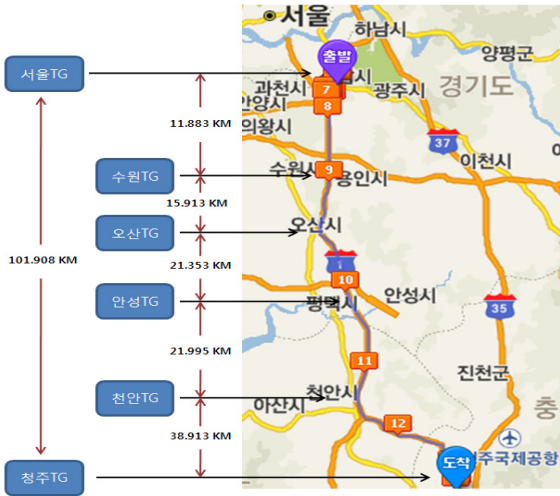


<그림 1> 서울-청주 전진반복 전후방 탐색법

3. 테스트자료

본 연구에서는 2007년 7월 9~16일 사이에 경부고속도로 서울-청주 구간 내에 위치한 수원, 오산, 안성, 천안, 청주 구간을 대상으로 통행료수납시스템(TCS)에 의해 수집된 자료를 이용하였다. 실험 자료는 고속도로 이용차량 약 100만대 이상의 TCS 자료로 한국도로공사 산하 도로

교통기술원에서 제공받았다. 실험 구간 선정을 서울-청주로 채택한 이유는 차량 통행이 많아 자료의 수집도가 높은 지역이라는 점이다. 실제 테스트에서는 13시~16시(총 3시간)의 자료만을 처리하였다.

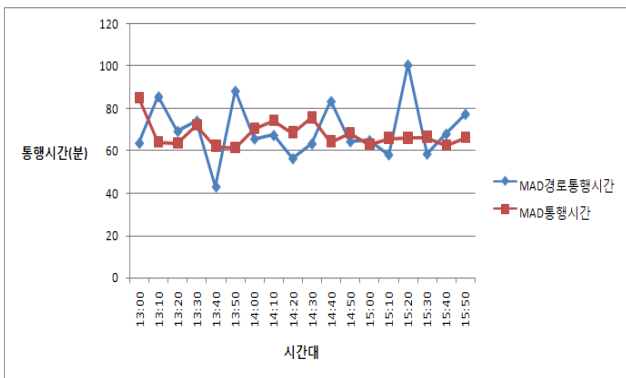


<그림 2> 실험구간

4. 비교평가

본 연구에서 제안하는 Fuzzy c-means의 방법으로 분류한 대푯값과 추출된 대푯값의 정확성을 증명하기 위해 전진반복 전후방탐색법을 이용하여 구간 실측치를 구한다. Fuzzy c-means를 이용한 TCS 데이터 주행특성분류 기법과 MAD 통행시간 분류 알고리즘을 비교한다.

2007년 7월 10일 화요일 13:00~16:00 시간대의 MAD(절대중위편차)를 이용한 서울-청주 통행시간과 전진반복 전후방탐색법을 사용하여 구한 경로통행시간과의 차이를 비교하였다. 그림 3은 7월 10일 서울-청주 MAD통행시간과 전진반복 전후방탐색법을 사용하여 얻어진 서울-청주 MAD경로통행시간을 비교해 본 결과이며 그래프에서 보는 바와 같이 오차가 많이 발생하는 것을 볼 수 있다.



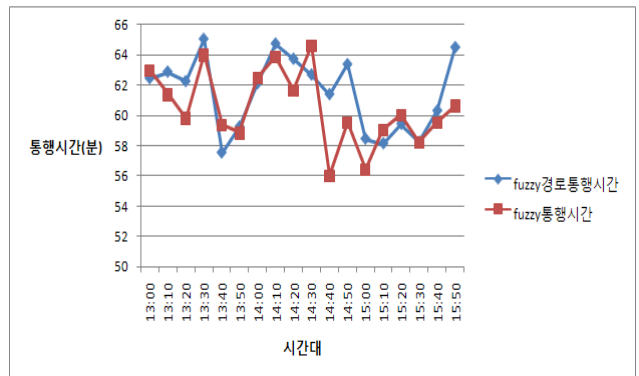
<그림 3> 서울-청주 MAD경로통행시간과 MAD통행시간 비교

표 1은 서울-청주 MAD경로통행시간과 통행시간, 오차(분), 오차비율(%)을 나타낸 결과이다. 오차비율은 경로통행시간과 통행시간에 오차를 나타낸 비율이다. 오차 평균은 12.4분, 오차비율의 평균 범위는 18.6%를 나타내었고, 오차, 오차 비율의 범위가 크면 클수록 경로통행시간과 통행시간의 차이가 커져 효과적인 분류방법이라 할 수 없다.

<표 1> 서울-청주 MAD 비교

시간	MAD 경로통행시간	MAD 통행시간	오차	오차 비율
13:00	63.71	84.81	21.1	24.9
13:10	85.64	64	21.64	33.8
13:20	69.31	63.56	5.75	9.0
13:30	74.20	72.13	2.07	2.9
13:40	43.03	62.21	19.18	30.8
13:50	88.33	61.56	26.77	43.5
14:00	65.72	70.56	4.84	6.9
14:10	67.41	74.35	6.94	9.3
14:20	56.45	68.59	12.14	17.7
14:30	63.41	75.67	12.26	16.2
14:40	83.46	64.38	19.08	29.6
14:50	64.42	68.53	4.11	6.0
15:00	64.97	63	1.97	3.1
15:10	58.21	66	7.79	11.8
15:20	100.69	66.07	34.62	52.4
15:30	58.58	66.57	7.99	12.0
15:40	68.03	62.63	5.4	8.6
15:50	77.41	66.13	11.28	17.1
평균			12.4	18.6

그림 4는 7월 10일 서울-청주 구간을 Fuzzy c-means를 적용하여 분류한 서울-청주 통행시간과 전진반복 전후방탐색법을 사용하여 각 구간을 Fuzzy c-means의 방법으로 분류한 통행시간의 합과 차를 구해 서울-청주 경로통행시간을 구하였다. 기존의 방법인 MAD보다 통행시간과 경로통행시간이 유사한 결과를 보여 본 연구에서 제안하는 Fuzzy c-means를 이용한 방법이 기존의 방법보다 서울-청주 통행시간을 추출하는데 더 정확한 결과를 나타내었다.



<그림 4> 서울-청주 Fuzzy c-means 경로통행시간과 Fuzzy c-means 통행시간 비교

표 2는 Fuzzy c-means를 이용한 경로통행시간과 통행시간 오차, 오차비율을 나타낸 결과이다. 오차분의 평균범위는 1.6분, 오차 평균 비율은 2.7%를 보였다. 기존 분류방법인 MAD, 신외구간 추출법에 의한 통행시간 오차 및 오차 비율 보다 훨씬 작은 값을 보였으며, 이는 전진반복 전후방탐색법을 사용하여 “경로통행시간과 통행시간은 같다”는 기존 연구에 부합하는 결과를 보여준다.

<표 2> 서울-청주 Fuzzy c-means 비교

시간	Fuzzy c-means 경로통행시간	Fuzzy c-means 통행시간	오차	오차비율	MAD 대비성능향상
13:00	62.45	62.98	0.53	0.8	97.49
13:10	62.87	61.38	1.49	2.4	93.11
13:20	62.25	59.80	2.45	4.1	57.39
13:30	65.04	64	1.04	1.6	49.76
13:40	57.55	59.36	1.81	3.0	90.56
13:50	59.26	58.83	0.43	0.7	98.39
14:00	62.12	62.50	0.38	0.6	92.15
14:10	66.53	63.90	2.63	4.1	62.10
14:20	63.73	61.67	2.06	3.3	83.03
14:30	62.28	64.60	2.32	3.6	81.08
14:40	61.39	56	5.39	9.6	71.75
14:50	63.37	59.50	3.87	6.5	5.84
15:00	58.46	56.40	2.06	3.7	4.57
15:10	58.16	59	0.84	1.4	89.22
15:20	59.44	60	0.56	0.9	98.38
15:30	57.75	58.21	0.46	0.8	94.24
15:40	59.28	59.52	0.24	0.4	95.56
15:50	60.11	60.63	0.52	0.9	95.39
평균			1.6	2.7	75.5

5. 결론

본 연구에서는 TCS 자료를 Fuzzy c-means Clustering 방법을 이용하여 고속도로 통행시간 주행특성을 분류하는 기법에 관한 연구를 수행하였다. 현재까지는 고속도로 구간의 통행시간 분류 선정시 각 구간별로 하나의 대푯값을 사용한다. 이의 문제점은 고속도로 특성으로 규정 속도 이상의 속도로 주행하는 차량, 규정 속도 및 휴게소 이용차량, 운전자의 운전 습성, 통행 목적, 피로의 정도, 운전자 성향과 도로상황에 따라 통행시간이 다르게 나타나는 점이다. 기존의 통행시간 분류 방법들은 주로 통계 기법에 의한 방법으로 TCS 자료의 특성을 반영 하지 못하였다. 이런 점을 착안하여, 본 연구에서는 TCS 자료의 도착시각 기준을 출발시각 기준으로 재구성하고, 자료의 분류 정도를 잘 나타내는 Fuzzy c-means Clustering 통해 고속도로 통행시간 주행특성분류에 맞게 구성하였다. 본 연구에서 제안하는 주행특성 및 도로상황을 고려한 Fuzzy c-means를 이용한 통행시간 분류방법과 기존의 통행시간 분류 방법을 통한 통행시간을 PIFAB를 사용 TCS 자료의 통행시간과 경로통행시간과 비교 평가하였다. 평가한 결과 본 연구에서 제안하는 Fuzzy c-means기법은 기존의 MAD방법보다 75%의 정확성을 보였다.

감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT 연구센터 지원사업의 연구결과로수행되었음.
(IITA-2009-C1090-0902-0040)

본 연구는 한국도로공사 OASIS(Operations Analysis and Supportive nformationstem)의 자료지원을 받아 수행되었습니다.

참고문헌

- [1] 남궁성, “고속도로 경로통행시간 산출을 위한 전진반복 전후방탐색법(PIFAB)의 개발,” 대한교통학회, 대한교통학회지, 제23권 제5호, pp. 147-155, 2005.
- [2] 강정규, 남궁성, “고속도로 통행료수납자료를 이용한 통행시간 예측모형 개발,” 대한교통학회, 대한교통학회지, 제20권 제4호, pp. 151-162, 2002.
- [3] 양영규, 박원식, 남궁성, “고속도로 통행시간 예측을 위한 TCS 자료 분석 기술 현황,” 한국통신학회, 한국통신학회지, 제25권 제7호, pp. 10-15, 2008.
- [4] 박원식, 최진우, 양영규, “고속도로 통행료수납자료를 이용한 통행시간 군집현상에 관한 연구,” 한국GIS학회, 한국GIS학회춘계학술대회연구집, pp. 195-210, 2008.
- [5] 남궁성, 윤일수, 연인성, 조범철, “ITS 기술개발 연구 (IV): 고속도로 통행시간 예측시스템 개발,” 제18회 도로기술 연구 성과 발표회 연구집, 한국도로공사, 2001.
- [6] 도명식, 이향미, 남궁성, “TCS 데이터를 이용한 이상치제거 및 결측보정 알고리즘 개발,” 대한교통학회, 대한교통학회연구지 제26권 제4호, pp. 241-250, 2008.
- [7] 김재진, 노정현, 박동주, 남궁성, “출발시각기준 링크통행시간 정보의 공간적 설계,” 대한교통학회, 대한교통학회지, 제25권 제2호, 2007.
- [8] 남궁성, 윤일수, 조범철, “TCS 자료를 이용한 고속도로 통행시간 예측,” 대한교통학회, 대한교통학회지, 제15호, pp. 489-494, 1999.
- [9] 김재진, 노정현, 박동주, 남궁성, “출발 및 도착시각기준 구간통행시간 차이의 원인규명에 관한 연구: 연속속류를 중심으로,” 국토연구, 제48권, pp. 71-86, 2006.
- [10] 강지혜, 김성주, “적응적인 초기치 설정을 이용한 Fast K-means 및 Fuzzy-c-means 알고리즘,” 한국정보과학회, 정보과학회연구지, 제 31권 제4호, pp. 516-524, 2004.
- [11] 김재현, 정진혁, 최민환, 장 훈, “TCS 자료 및 GIS를 이용한 한국의 통행패턴 분석,” 대한교통학회, 대한교통학회지 제26권 제3호, pp. 75-84, 2008.
- [12] 한학용, “패턴인식 개론”, 한빛미디어, 2005.