

데이터마이닝을 이용한 운행패턴 분석방법에 대한 연구

An Application of Data Mining Techniques in the Driving Pattern Analysis

김 현 속* 최 종 우** 김 대 우*** 박 호 성**** 노 성 기***** 박 정 희*****
(Hyun-Suk Kim) (Jong-Woo Choi) (Dae-Woo Kim) (Ho-Sung Park) (Sung-Kee Noh) (Cheong-Hee Park)

요 약

근래에는 경제운전에 대한 중요성이 점차 부각되고 있어 운전자의 운전 행태나 성향을 자동으로 분석한 후 경제운전을 위한 방법을 운전자에게 알려줄 수 있는 연구가 필요하다. 본 논문에서는 이를 위해 차량에 대한 운행일시, 운행거리, 운행시간, 주행속도, 공회전시간, 급가속/급감속 횟수, 연료소모량 등의 운행정보를 수집하였고, 데이터마이닝을 이용하여 운전자의 운행패턴이 경제운전에 어떤 영향을 미칠 수 있는지 분석하였다. 본 연구 결과는 주행 중 운전자에게 지속적으로 공회전과 과속 정보, 급가속/급감속 횟수를 차량 단말에 표현하여 제공하고, 공회전과 과속 비율이 일정 임계치를 초과할 경우 경고 정보를 제공함으로써 경제운전에 악영향을 미칠 수 있는 운전 습관을 미리 예방할 수 있는 방안에 활용할 수 있다.

Abstract

Recently, as the importance of Economical Driving has been gradually growing up, the needs for research on automatic analysis of driving patterns that will ultimately provide drivers the methods for Economical Driving have been increasingly risen. Based on this purpose, we have executed two things in this paper. First, we have collected overall driving information such as date, distance, driving time, speed, idle time, sudden acceleration/deceleration count, and the amount of fuel consumption. Second, we have analyzed the influences of driving patterns on economical driving by employing the data mining techniques. These results can be applied in preventing bad driving patterns which will have consequently bad effects on Economical Driving in two aspects: by presenting some information on the terminal of the vehicles such as idle time, over-speed time, sudden acceleration/deceleration count continuously and by providing the drivers with alert information when the idle time ratio and the over-speed time ratio are excessive.

Key words: Economical driving, driving patterns, data mining, idle, over-speed, acceleration/deceleration

† 본 연구는 국토해양부 건설기술혁신사업 스마트하이웨이사업(07기술혁신A01)의 전천후 도로-자동차 상황관리시스템 구축 연구과제 결과물로 수행하였습니다.

* 주저자 : 한국전자통신연구원 친환경차량IT연구팀 책임연구원

** 공저자 : 한국전자통신연구원 친환경차량IT연구팀 연구원

*** 공저자 : 충남대학교 박사과정

**** 공저자 : 충남대학교 석사과정

***** 공저자 : 한국전자통신연구원 유무선통합기술연구팀 책임연구원

***** 공저자 : 충남대학교 컴퓨터공학과 교수

† 논문접수일 : 2009년 7월 23일

† 논문심사일 : 2009년 11월 2일

† 게재확정일 : 2009년 11월 5일

I. 서 론

근래에 고유가 시대와 세계 경제 위기에 들어서면서, 경제운전에 대한 중요성이 부각됨에 따라 환경부, 에너지관리공단, 자동차시민연합 등의 정부기관과 민간단체에서는 친환경 운전 및 경제 운전 방법에 대한 지침이나 홍보물을 제작하여 운전자에게 배포하는 노력이 진행 중에 있다 [1-3]. 그러나, 경제운전을 위한 지침이나 홍보물은 주로 운전자의 운전 방법과 차량 관리 방법에 대한 가이드 정보를 제공하고 있으나 운전자가 이러한 가이드들을 모두 숙지하고 있을지라도 실제 운전 적용하기는 쉽지 않은 현실이다. 따라서, 개별 운전자의 운전 행태나 성향을 자동으로 분석한 후 운전패턴이 경제운전에 어떤 영향을 미치는지를 비교하여 운전자에게 알려주는 방법에 대한 연구가 필요하다.

민석기 등의 연구에서는 실험차량을 통해 운전자의 주행패턴을 결정하는 파라미터로 Time-gap과 TTC(Time-to-Collision)를 사용하였다. 이를 위해 평균 주행속도, 선행 차량과의 평균 유지 거리를 측정하였다. Time-gap은 선행 차량과 자기 차량의 상대거리를 자기 차량의 속도로 나누어 계산한 값이고, TTC는 선행차량과 자기 차량의 상대 속도와 상대거리의 비율을 통해 계산하였다. Time-gap 과 TTC는 각각 가속과 감속시의 운전자 특성을 나타내는 파라미터이므로 운전자의 운행 패턴을 급가속, 급감속에 관련된 성향만 분석하여 제공할 수 있다[4].

박용성 등의 연구에서는 경형, 소형, 중형 승용차별로 자동차의 주행속도에 따른 연비, 에어컨 작동에 의한 연료소모량, 공회전시 연료소모량, 급가속시 연료소모량, 신호등 정차에 의한 연료소모량을 실험을 통하여 분석한 결과를 보여주었다. 이것은 자동차의 특정한 운전조건에 따른 연료소모량과 연비를 시험한 결과만 제공하므로, 경제 운전을 위한 경제 속도정보와 바람직한 운전 습관에 대한 일반적인 정보를 알 수 있지만, 임의의 운전자의 운행패턴을 일정 기간 관측하고 매 운행마다의 연료소모 효율을 운전자의 주행패턴과 연관하여 자동으로 분석해 줄 수 있는 방법은 제공할 수 없다 [5].

미국에서는 차량의 이상운행을 실시간으로 모니터링하기 위해 VEDAS(Vehicle DAta Stream mining)시스템을 개발하였다. VEDAS에서는 PDA와 OBD-II를 연결하여 차량의 방향, 차량의 속도, 차량 상태정보 등을 수집하고, 각 차량별로 PDA 단말기를 통해 분산 데이터마이닝을 수행하여 음주운전, 졸음운전 등과 같은 위험한 운전패턴을 분석하는 연구를 수행하였다 [6].

위와 같이 기존의 연구에서는 경제 운전 필요 운전습관에 대한 기본적인 방안을 제공하거나, 위험한 운행을 하는 차량 운전자를 검색할 수 있는 방안을 제공하는 연구로 구성되므로, 개별 운전자의 운전 패턴을 알 수 있는 정보를 지속적으로 수집하여 운전 성향을 자동으로 분석함으로써 보완되어야 할 운전패턴에 대한 정보를 해당 운전자에게 쉽게 알려주어 경제 운전을 유도할 수 있는 방법에 대한 연구가 필요하다.

이를 위해서는 다양한 운전자의 운행기록에 대한 정보가 대량으로 확보되어야 하며, 대량의 데이터의 연관성을 분석하기 위해서는 데이터 마이닝 방법을 활용하는 것이 필요하다. 데이터마이닝(Data Mining)이란 용어는 1995년부터 사용되기 시작하였으며, 광산에서 광물을 캐내는 것에 비유한 것으로 금광석에 극히 미량으로 포함된 금을 여러 단계를 걸쳐 추출하듯이 대량의 데이터로부터 유용한 정보를 자동적으로 탐색하는 과정을 의미한다.

이건학의 연구에서는 공간데이터마이닝을 통해 서울시 강남구를 사례로 하여 교통사고의 공간적 군집분석을 통해 교통사고의 특성을 유형별로 분류하였다 [7].

Marzi 등의 연구에서는 특정 대학의 신입생들을 대상으로 고등학교 성적 정보와 대학교 1학년 성적 정보를 데이터마이닝하여 출신 고교별로 학업성취능력을 분석하였다 [8].

Sahu 등의 연구에서는 대규모의 날씨 데이터를 데이터마이닝하여 기상과 관련한 파라미터사이의 연관성을 분석하고 다차원 연관 규칙을 추출하였다[9].

이에 본 논문에서는 운전자의 차량 운행 정보에 대해서도 대량의 정보를 축적할 수 있는 환경을 구축할 수 있으므로 경제적인 운행 패턴에 영향을 미칠 수 있는 공회전 시간, 급가속/급감속 회수, 과속 주행 비율 등의 정보를 수집한 후, 데이터마이닝 기법을 활용하여 운전자별 운행정보를 분석함으로써 경제운전을 유도할 수 있는 유용한 정보를 도출하고 이를 운전자에게 제공할 수 있는 방법을 제안하고자 한다.

2장에서는 데이터마이닝의 여러 방법 중에서 본 논문에서 사용된 방법에 대해 간략히 소개를 하고, 3장에서는 기본적인 운행 패턴을 분석할 수 있는 데이터 구축을 위해 차량의 운행 정보를 실시간으로 수집할 수 있는 차량 및 운전자 정보 수집 환경 구축에 대해 소개한다. 4장에서는 실제 수집된 운행 데이터를 이용하여 운전자별 운행 패턴 분석 과정과 경제 운전에 대한 정보 제공 방법에 대해 설명하고 5장에서 결론을 맺는다.

II. 데이터마이닝 개념

2장에서는 본 논문에서 사용한 통계 또는 데이터 마이닝 기법에 대한 기본 개념을 소개 한다.

1. 회귀 모형(Regression Model)

회귀 모형이란 변수들 간의 상호 관련성을 규명하고 어떤 변수들의 변화로부터 다른 변수들의 변화를 예측하는 통계기법이다. 즉, 변수들이 서로 어떤 관계이며, 얼마나 밀접한가와 관심 있는 하나의 변수가 다른 변수들에 의해 얼마나 잘 예측될 수 있는지를 알아보는 것이다. 회귀 모형은 공정변수가 하나일 때, 공정변수와 품질변수간의 선형관계에 대한 분석을 하는 단순 회귀 모형(simple regression model)과 품질변수에 영향을 미치는 변수가 여러 개일 때, 이들 공정변수들과 품질간의 선형관계에 대한 분석을 하는 다중 회귀 모형(multiple regression model)으로 구성된다.

2. 군집 분석(Clustering)

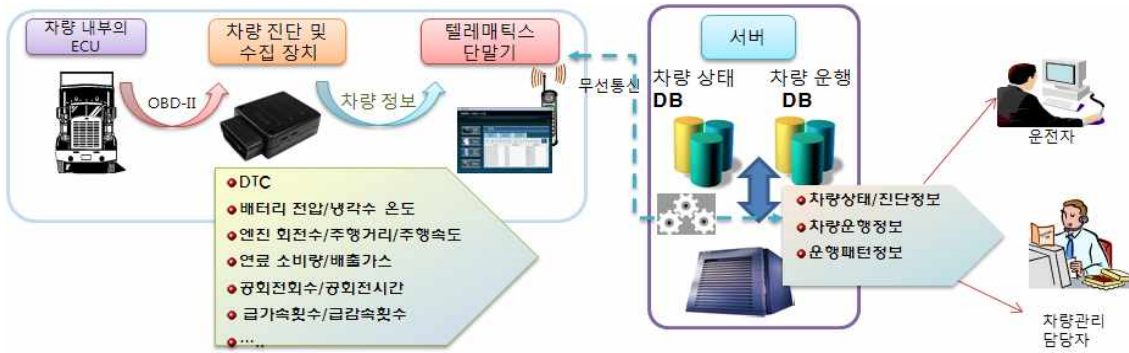
관찰 데이터들의 Grouping을 의미하며, 주어진 관찰 데이터 중에서 유사한 것들을 몇몇의 집단으로 그룹화하여 각 집단의 성격을 파악함으로써 전체 데이터에 대한 의미 있는 정보를 얻을 수 있는 분석 방법이다. 군집들의 전체 모음을 일반적으로 군집화(clustering)라고 부르고, 군집화의 유형은 계층적 또는 분할적으로 구분된다. 계층 군집화(Hierarchical clustering)는 군집이 다른 하위 군집들을 가지는 경우이고, 트리 구조로 내포된 군집들의 집합이다. 분할 군집화(Partitional clustering)는 데이터 객체들을 중복이 없는 부분집합으로 나누는 것이다. 분할 군집화 기법중에서 K-means 클러스터링이 많이 쓰인다. K-means 기법에서는 K개의 점들을 초기 중심점으로 선택하고 각 개체를 가장 가까운 중심점에 할당함으로써 K개의 군집을 형성한 후, 각 군집의 중심점을 다시 계산하여 중심점이 바뀌지 않을 때까지 반복하여 최종적으로 군집을 결정한다.

III. 차량 운행 정보 수집 환경

운전자의 운행패턴을 분석하기 위해서는 각 운전자별로 실제 운행한 정보를 수집하여 저장할 수 있는 시스템이 필요하다. 본 논문에서는 그림 1의 차량 및 운전자 정보 수집시스템 환경을 이용하여 데이터를 수집하였다.

1. 차량과 텔레매틱스 단말기간의 차량정보 수집 방법

차량정보를 실시간으로 수집하기 위해서는 차량의 내부정보를 차량 내에 설치된 텔레매틱스 단말기로 전달해 주어야 한다. 정보를 추출하는 방법으로는 차량 내부에 직접 선을 연결하는 와이어링 방법과 차량의 전장장치인 ECU (Electronic control unit)와 연결된 OBD-II (On Board Diagnostics) 인터페이스를 통해 차량정보를 읽어오는 방식이 있다.



<그림 1> 차량 및 운전자 정보 수집 시스템
 <Fig. 1> Vehicle and driver information collection system

<그림 1>의 수집 환경에서는 차량의 OBD-II 인터페이스에 수집장치를 연결하여 차량 정보를 수집하고 이를 텔레매틱스 단말기에 전달하는 방법을 사용하였다.

2. 수집되는 차량정보 특성

수집장치를 통한 차량 정보 수집주기는 편의에 따라 정할 수 있다. 본 논문에서는 1초마다 수집되는 정보와 10초마다 수집되는 정보로 구별하였다. 1초마다 수집장치를 통해 주기적으로 수집되는 정보의 종류로는 엔진 회전수(RPM), 주행거리, 주행속도, 연료 분사량, 배터리전압, 냉각수 온도, DTC(Diagnostic Trouble Code) 등의 정보가 있다. DTC 정보의 경우는 차량의 내부에 이상이 생길 경우에만 발행되며, 장애 발생시 텔레매틱스 단말기로 전송하게 된다. 10초마다 수집되는 정보에는 운행거리, 최고속도, 공회전시간, 급가속/급감속 횟수 등이 있다.

본 논문에서는 운전자가 시동을 켜고 주행한 후, 시동을 끌 때까지의 시간동안을 하나의 운행이라 정의하고 이에 대한 운행시작 및 종료시간, 운행거리, 공회전시간, 급가속/급감속 횟수, 연료소모량 등의 정보도 수집하였다.

3. 텔레매틱스 단말기와 서버간의 차량정보 전송 방법

텔레매틱스 단말기에 1초마다 수집된 정보와 10초마다 수집된 정보가 계속 저장이 되고, 저장된 정보는 주기적으로 이동통신망을 이용하여 서버로 전송된다. 또, 새로 시동이 켜질 때에는 바로 이전의 운행정보에 대해 서버로 송신한다. 단말기에 저장된 차량 정보가 이동통신망을 이용하여 전송되므로 서버에는 최신의 정보가 항상 유지될 수 있으며, IT를 잘 모르는 운전자도 불편하지 않게 시스템을 사용할 수 있다.

4. 차량 운행정보 데이터베이스

서버는 단말기로부터 차량 정보를 수신하여 데이터베이스의 차량 상태 DB와 차량 운행 DB에 저장한다. 본 논문에서는 1톤 2대, 2.5톤 2대, 5톤 2대, 11톤 2대 등의 4개 차종 8대에 대한 차량 정보를 수집하였다. 차량 정보 수집 기간은 2008년 11월부터 2009년 3월까지의 4개월간 진행하였다. 4개 차량에 대한 데이터 중에서 2.5톤 차량에 대해 수집한 운행정보에 대해 데이터 분석을 수행하였다.

<표 1>은 차량운행 정보 릴레이션 구성 요소들을 나타낸다. 1일 동안 운전자는 여러 번 운행할 수 있기 때문에 차량운행 정보 릴레이션에는 운행별로 1 tuple로 정보를 저장할 수 있어야 한다. CAR_ID는 차량의 번호이므로 운전자를 구별할 수 있으며, 한 번의 시동을 켜 시간과 시동을 끈 시간은 START_

DAY_TIME 과 END_DAY_TIME에 저장된다. 또, 한 번의 운행시간동안의 주행 거리는 DISTANCE에 저장하고, 운행하는 동안의 공회전 시간은 IDLE_TIME에 저장한다. 운행 동안의 급가속 횟수와 급감속 횟수는 HIGHSP_CNT와 LOWSP_CNT에 각각 저장한다. 운행시간동안 소모된 연료소모량은 TOT_FUEL에 저장하고 운행시간은 DRIVING_TIME에 저장한다.

<표 1> 차량운행 정보 릴레이션 설명
<Table 1> Vehicle driving relation description

변수명	변수 설명
CAR_ID	차량 ID
START_DAY_TIME	운행시작 일자 및 시간
END_DAY_TIME	운행종료 일자 및 시간
DISTANCE	운행거리(m)
IDLE_TIME	공회전시간(sec)
HIGHSP_CNT	급가속횟수
LOWSP_CNT	급감속횟수
TOT_FUEL	총연료소모량(ml)
DRIVING_TIME	운행시간(sec)

<표 2> 차량운행 정보 샘플 레코드
<Table 2> Vehicle driving sample data

CAR_ID	START_& END DAY_TIME	DISTANCE	IDLE_TIME	HIGH_CNT	LOW_CNT	TOT_FUEL	DRV_TIME
2691 T2P5	20081217084225 20081217084904	1938	72	8	0	258	399
2691 T2P5	20081217085310 20081217100651	54106	285	16	14	21548	4421
2691 T2P5	20081222202428 20081222211355	17152	2017	12	2	2324	2967
2691 T2P5	20081229224951 20081229232230	34613	383	26	2	5274	1959
2691 T2P5	20090105131502 20090105155046	33162	2298	55	12	23835	9344
2691 T2P5	20090107130957 20090107140826	50674	513	42	9	7397	3509
2691 T2P5	20090109105627 20090109113951	36821	581	33	6	5516	2604

본 논문에서는 2.5톤 트럭 운전자의 차량 운행 정보를 활용하여 경제운전과 공회전시간, 급가속

및 급감속 횟수가 어떤 영향을 미치는지에 대하여 분석한다.

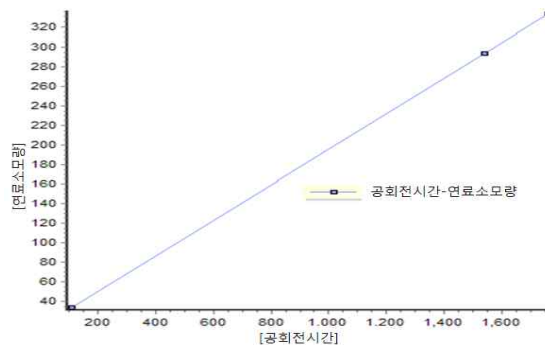
IV. 차량 운행정보 데이터 분석

본 장에서는 회귀분석을 이용하여 공회전, 급가속, 급감속 등과 연료소모량간의 관계를 분석하였고, 군집분석을 통해 경제운전에 크게 영향을 미치는 공회전 비율에 대한 분석을 수행한다. 본 논문에서는 ECMiner를 사용하여 데이터를 분석하였다..

1. 공회전시간과 연료소모량과의 관계에 대한 회귀분석 결과

일반 승용차 운전자와 다르게 상용 트럭 운전자들은 운행을 하기 전에 시동을 켜고 짐을 싣는 등의 준비를 하는 경우가 많이 있다. 이때, 공회전이 많이 발생하게 된다. 공회전으로 인해 불필요하게 소모되는 연료소모량에 대한 분석을 통해 공회전을 경감하는 운전안내를 제공하여 경제운전을 유도할 수 있다.

<그림 2>는 2.5톤 A-차량의 공회전시간과 연료소모량간의 관계이다. 차량운행 정보 데이터중에서 시동을 걸었으나 운행거리는 0이면서, 공회전만 시킨 운행인 경우이다. <그림 2>에서 볼 수 있듯이 공회전 시간이 증가할수록 연료 소모량도 증가함을 알 수 있으며, 회귀 분석을 수행한 결과 다음과 같



<그림 2> 2.5톤 A-차량의 공회전시간과 연료소모량과의 관계
<Fig. 2> Relationship between idle-time and fuel consumption of 2.5 ton A-vehicle

은 연관식을 추출할 수 있었고 회귀 분석 중 상수값과 IDLE_TIME의 p-value = 0으로 측정되어 아래 연관식이 유효한 정보임을 알 수 있었다.

$$TOT_FUEL = 14.56903 + 0.18092 IDLE_TIME \quad (1)$$

2. 운전패턴과 연료소모량과의 관계에 대한 회귀분석 결과

동일한 운전자일 경우에도 운전패턴에 따라 연료소모 비율이 다를 수 있다. 본 논문에서는 공회전 비율, 급가속 횟수, 급감속 횟수에 따라 구분되는 운전패턴에 따라 연료소모비율이 어떤 영향을 받는지에 대해 분석하기 위해, 공회전비율(IDLE_RATIO)과 연료소모비율 (MILEAGE)은 식(2)와 식(3)과 같이 계산하여 사용하였다.

$$\text{공회전비율} = IDLE_TIME/DRIVING_TIME \quad (2)$$

$$\text{연료소모비율} = DISTANCE/TOT_FUEL \quad (3)$$

2.5톤 A-차량은 <표 2>와 같은 형태의 운행 기록이 3개월간 약 317건 정도 수집되었다. 이 운행 기록에 대해 식(3)을 적용하여 연료소모비율을 계산하였다. 동일한 거리를 적은 연료를 이용하여 주행할 수록 연료소모비율의 수치가 크게 나오며, 좋은

운행기록으로 볼 수 있다. 다음은 연료소모비율 수치가 크게 나온 상위 10%의 운행 기록과 연료소모비율이 작게 나온 하위 10%의 운행 기록에 대한 회귀분석을 수행하였다.

1) 상위 10%의 운행기록에 대한 분석

<그림 3>은 317건의 운행 기록중에서 연료소모비율이 높아 상위 10%안에 포함된 운행 기록에 대해 공회전 비율과 연료소모비율과의 관계, 급가속/급감속 횟수와 연료소모비율과의 관계를 볼 수 있는데, 상위 10%의 연료소모비율로 갈수록 공회전 비율, 급가속/급감속 횟수가 적게 보임을 알 수 있다.

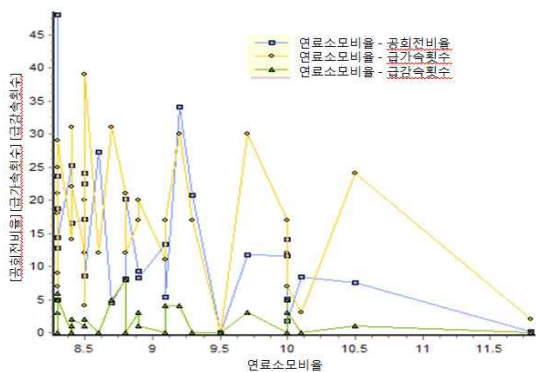
이 데이터에 대한 회귀 분석을 수행한 결과 식 (4)와 같은 연관식을 추출할 수 있었다. 식 (4)의 연관식에서 상수값, IDLE_RATIO, HIGHSP_CNT의 p-value는 모두 0으로 측정되어 MILEAGE와 크게 연관이 있음을 알 수 있었고, LOWSP_CNT의 p-value는 0.16으로 측정되어 급감속 횟수는 연료소모비율에 큰 영향을 미치지 않음을 알 수 있었다.

$$\begin{aligned} MILEAGE = & 10.03024 - 0.03763 IDLE_RATIO \\ & - 0.02141 HIGHSP_CNT - 0.02877 \\ & LOWSP_CNT \quad (4) \end{aligned}$$

2) 하위 10%의 운행기록에 대한 분석

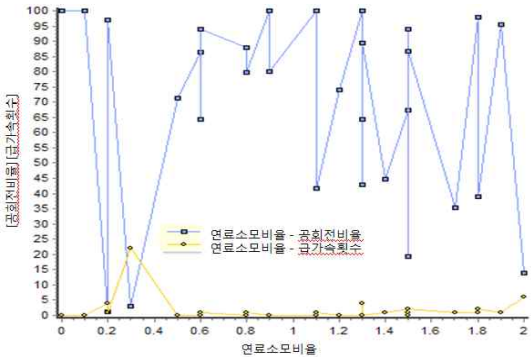
<그림 4>는 317건의 운행 기록중에서 연료소모비율이 낮아 하위 10%안에 포함된 운행 기록에 대해 회귀분석을 수행한 결과이고, 식(5)와 같은 연관식을 추출할 수 있었다. <그림 4>에서는 급감속 횟수는 0으로 기록되어 있어 의미가 없어 제외시켰다. 하위 10%의 연료소모비율로 갈수록 공회전 비율이 대부분 40%이상을 차지하기 때문에 급가속 횟수가 비교적 적게 나타나도 연료소모비율이 낮음을 알 수 있다. 이를 통해, 공회전 비율을 낮추는 운전 패턴이 경제 운전에 영향을 미침을 알 수 있었다.

회귀 분석을 수행한 결과 다음의 식(5)와 같은 연관식을 추출할 수 있었고 회귀 분석 중 상수값의 p-value는 0이므로 유효하고 IDLE_RATIO의 p-value



<그림 3> 2.5톤 A-차량의 상위10% 연료소모 비율을 보인 운행기록 분석

<Fig. 3> Regression analysis about upper 10% mileage record of 2.5t on A-vehicle



<그림 4> 2.5톤 A-차량의 하위10% 연료소모비율을 보인 운행기록 분석

<Fig. 4> Regression analysis about lower 10% mileage record of 2.5 ton A-vehicle

는 0.07, HIGHSP_CNT의 p-value는 0.16이므로 연관식이 유효하지는 않지만, 도표를 통해 공회전이 연료소모비율에 많은 영향을 주는 것을 알 수 있었다.

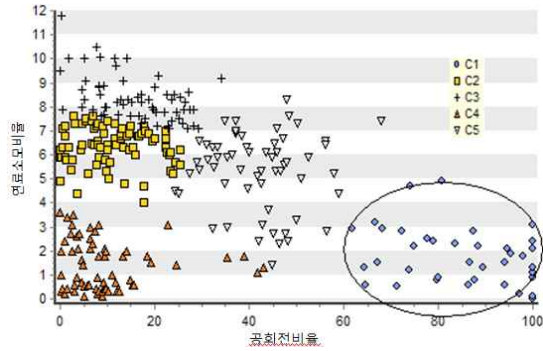
$$\text{MILEAGE} = 1.61746 - 0.00776 \text{ IDLE_RATIO} - 0.04702 \text{ HIGHSP_CNT (5)}$$

3. 공회전과 연료소모량과의 관계에 대한 군집분석 결과

2.5톤 A(2691)-차량의 운행정보는 317건이 있고, 이에 대해 KMeans 클러스터링을 5개의 군집으로 할 경우 <그림 5>와 같은 결과가 보여진다.

<그림 5>와 같이 공회전 비율이 60% 이상일 경우 연료소모비율은 0-3 이 대부분이고 최대 5 정도의 수치를 보인다. 연료소모비율이 9-12인 경우에는 공회전비율이 20% 이하인 경우에 해당된다. 따라서, 주행 중에 공회전비율을 감시하여 임계치를 넘어가지 않게 안내하여 경제운전을 유도할 수 있다.

<그림 5>의 C1~ C5 까지의 5개의 군집에 대해 각 군집의 빈도수, IDLE 중심값, MILEAGE 중심값은 <표 3>에 보여진다. <표 3>에서 2.5톤-A 차량은 C1 군집에서는 공회전 중심값이 높아 연료소모비율이 낮은 중심값을 보이며, C3 군집에서는 공회전 중심값이 낮아 연료소모비율이 높은 중심값을 보인다. 군집 C2와 군집 C4에서는 공회전 중심값은



<그림 5> 2.5톤 A-차량의 군집분석

<Fig. 5> Clustering analysis about 2.5 ton A-vehicle

각각 10.8과 9.5로 비슷하게 낮지만, 연료소모비율 중심값에서는 6.3과 1.3으로 많은 차이를 보이므로 공회전 비율 이외의 연료소모비율을 낮게 한 다른 요인에 대한 분석이 추가로 필요하다.

<표 3> 2.5톤 A-차량의 군집별 빈도수 및 백분율
<Table 3> Detail Information of 5 clustering of 2.5 ton A-vehicle

군집	빈도수	백분율 (순위)	IDLE 중심값	MILEAGE 중심값
C1	39	12.30%(5)	85.6	1.6
C2	85	26.81%(1)	10.8	6.3
C3	68	21.45%(2)	16.0	8.3
C4	66	20.82%(3)	9.5	1.3
C5	59	18.61%(4)	42.0	5.3

4. 급가속/급감속 횟수와 연료소모량과의 관계에 대한 회귀분석 결과

<표 3>의 군집 C2와 군집 C4에서 공회전 비율 이외의 연료소모비율을 낮게 한 다른 요인에 대한 분석을 위해 <그림 6> 과 <그림 7>에 각 군집에 대한 회귀분석을 수행한 결과를 나타내었다.

<그림 6>의 군집 C2보다는 <그림 7>의 군집 C4의 급가속 횟수 분포가 전체적으로 높은 회수를 보여주고 있다. 또, 아래의 <표 4>에서 군집 C2의 급가속 중심값은 12.65 회이고, 군집 C4의 급가속 중심값은 30회로 약 2.4 배의 차이를 보이고, 군집 C2

의 급감속 중심값은 6 회이고, 군집 C4의 급감속 중심값은 10.5 회로 약 1.8 배의 차이를 보인다. 따라서, <표 3>의 군집 C2와 군집 C4와 같이 공회전 비율이 비슷하더라도 급가속/급감속 횟수의 차이가 클 경우에는 연료소모비율에 많은 차이를 보이고 있음을 알 수 있다.

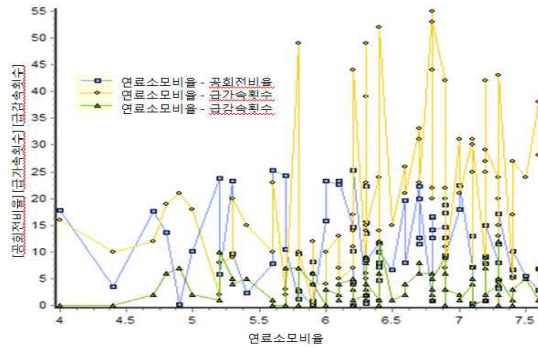
<표 4> 군집 C2와 군집 C4의 급가속/급감속 횟수 비교
<Table 4> Comparison with sudden acceleration/brake count of group-C2 and group-C4

군집	급가속중심값	급감속중심값
C2	12.65	6
C4	30	10.5

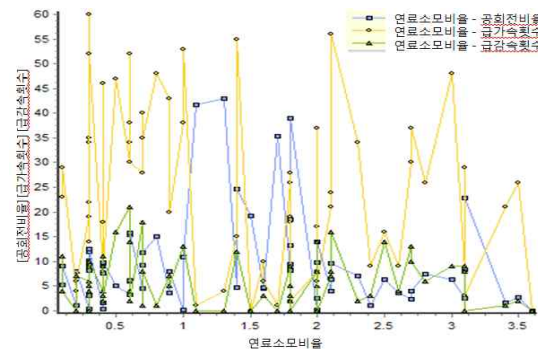
5. 과속 횟수와 연료소모량과의 관계에 대한 회귀분석 결과

과속 운행이 연료소모비율에 영향을 미치는 부분에 대해 알아보기 위해 317건의 운행중에서 비슷한 운행시간을 보이면서 연료소모비율이 차이가 나는 운행에 대해 <표 5>와 같이 크게 4개의 실험으로 구분하여 각 실험별 운행 특징을 분석하였다.

분석한 내용중 실험 2-1과 실험2-2에 대한 운행 특징을 <그림 8>과 <그림 9>에 나타내었다. <표 5>의 실험2-1은 실험2-2보다 공회전비율과 급가속 횟수가 낮은 경우에도 연료소모비율이 상대적으로 낮게 측정됨을 볼 수 있다. <그림 8>의 실험2-1에 대한 속도 구성을 보면 최고속도가 94까지 보이고,



<그림 6> 2.5톤 A-차량의 군집 C2 분석
<Fig. 6> Clustering analysis about group-C2 of 2.5 ton A-vehicle

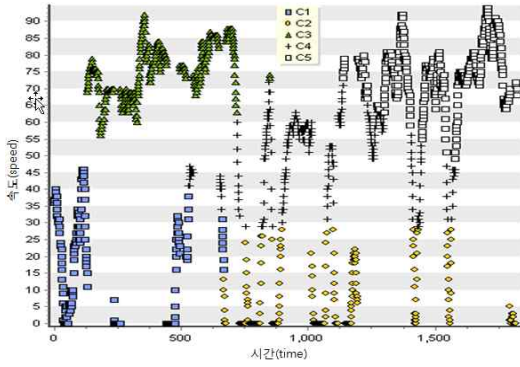


<그림 7> 2.5톤 A-차량의 군집 C4 분석
<Fig. 7> Clustering analysis about group-C4 of 2.5 ton A-vehicle

<그림 9>의 실험2-2에 대한 최고속도가 137까지 보여진다. 따라서, 비슷한 운행시간일지라도 과속 운행이 더 많은 경우 연료소모비율 수치가 낮게 측정됨을 보여준다.

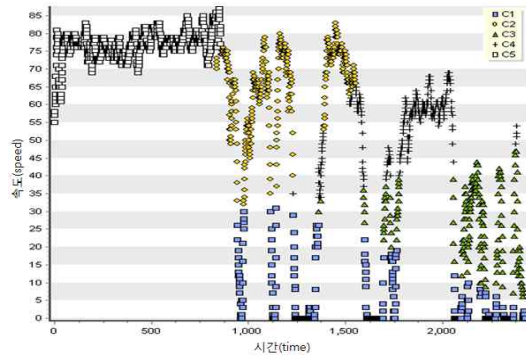
<표 5> 과속 운행패턴에 따른 연료소모비율 실험 레코드
<Table 5> Overspeed driving pattern sample data

실험구분	START_DAY_TIME END_DAY_TIME	DISTANCE (m)	IDLE_TIME (sec)	IDLE_RATIO (%)	HIGHSP_CNT (회)	LOWSP_CNT (회)	TOTFUEL (ml)	DRIVING_TIME (sec)	MILEAGE
실험 1	1-1 20090211 135817 20090211 141849	10417	348	28.2	17	2	1323	1232	7.9
	1-2 20090120 164530 20090120 170622	35442	0	0.0	0	4	6027	1252	5.9
실험 2	2-1 20090304 93352 20090304 100425	26407	276	15.1	23	1	3226	1833	8.2
	2-2 20081217 161850 20081217 164928	43883	149	8.1	12	6	7433	1838	5.9
실험 3	3-1 20090121 204623 20090121 212029	39018	162	7.9	21	4	5874	2046	6.6
	3-2 20081223 000112 20081223 003550	28603	264	12.7	21	0	3432	2078	8.3
실험 4	4-1 20090129 101048 20090129 105127	35807	342	14.0	17	5	3584	2439	10.0
	4-2 20090226 145640 20090226 153810	59067	144	5.8	8	10	11410	2490	5.2



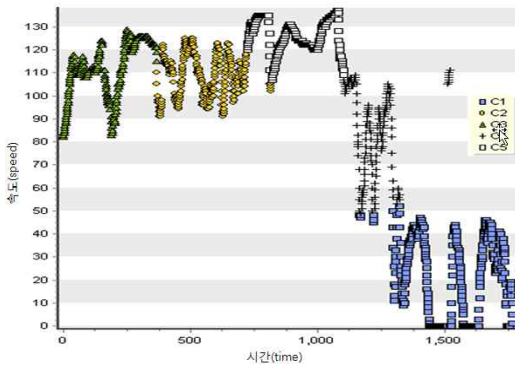
<그림 8> 실험2-1에 대한 군집분석

<Fig. 8> Clustering analysis for Experiment 2-1



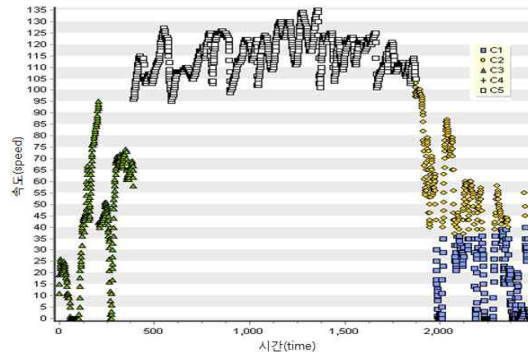
<그림 10> 실험4-1에 대한 군집분석

<Fig. 10> Clustering analysis for Experiment 4-1



<그림 9> 실험2-2에 대한 군집분석

<Fig. 9> Clustering analysis for Experiment 2-2



<그림 11> 실험4-2에 대한 군집분석

<Fig. 11> Clustering analysis for Experiment 4-2

<표 6> 실험2-1의 군집별 중심속도 및 빈도수
<Table 6> Central speed and frequency of Exp. 2-1

군집	중심Speed	빈도수	백분율
C1	16	224	12.29%
C2	4	326	17.88%
C3	75	460	25.23%
C4	51	337	18.49%
C5	75	476	26.11%

<표 7> 실험2-2의 군집별 중심속도 및 빈도수
<Table 7> Central speed and frequency of Exp. 2-2

군집	중심Speed	빈도수	백분율
C1	20	460	26.09%
C2	108	360	20.42%
C3	112	368	20.87%
C4	87	203	11.51%
C5	126	372	21.10%

<표 6>과 <표 7>을 비교해 보면 실험 2-1 운행 시간에는 C3, C4, C5의 중심 속도가 51~75로 보여지고, 실험 2-2 운행시간에는 C2, C3, C5의 중심 속도가 108~126으로 보여진다. 따라서, 실험 2-2의 운행시간에서는 실험 2-1 보다 과속의 비율이 더 많아서, 연료소모비율이 낮게 나타남을 알 수 있다.

<표 8> 실험4-1의 군집별 중심속도 및 비율 비교
<Table 8> Central speed and frequency of Exp. 4-1

군집	중심Speed	빈도수	백분율
C1	3	488	20.14%
C2	66	475	19.60%
C3	27	232	9.57%
C4	56	377	15.56%
C5	76	851	35.12%

<표 9> 실험4-2의 군집별 중심속도 및 비율 비교
<Table9> Central speed and frequency of Exp. 4-2

군집	중심Speed	빈도수	백분율
C1	12	298	12.03%
C2	62	309	12.47%
C3	42	390	15.74%
C4	-	0	0%
C5	115	1480	59.75%

<그림 10>과 <그림 11>은 각각 실험4-1과 실험 4-2에 대한 시간별 속도에 대한 변화를 나타낸다. 실험 4-1에서는 최고속도가 87까지 보이고, 실험4-2에서는 최고속도가 135까지 보여진다.

<표 8>과 <표 9>를 비교해 보면 실험 4-1 운행시간에는 C2, C4, C5의 중심 속도가 56~76으로 보여지고, 실험 4-2 운행시간에는 C5의 중심 속도가 115로 보여진다. 따라서, 실험 4-2의 운행시간에서는 실험 4-1 보다 과속의 비율이 더 많아서, 연료소모비율이 낮게 나타남을 알 수 있다.

V. 결 론

근래에 고유가 시대와 세계 경제 위기에 들어서면서, 경제운전에 대한 중요성이 부각됨에 운전자의 운전 행태나 성향을 자동으로 분석한 후 경제운전을 위한 방법을 운전자에게 알려주기 위한 방법이 필요하다. 본 논문에서는 운전자의 운행정보를 실시간으로 수집 및 저장하여 운행패턴을 자동으로 분석하는 방법에 대한 연구를 수행하였다.

운행정보를 수집하기 위해 차량의 속도, 공회전 시간, 급가속/급감속 횟수를 수집할 수 있는 진단장치를 차량에 설치하고, 이를 차량단말에서 수집한 후, CDMA망을 통해 차량정보서버로 송신하였다.

차량정보 서버에 저장된 운전자의 운행정보 데이터를 이용하여 공회전 시간, 급가속/급감속 횟수와 과속 주행시간이 연료소모비율에 미치는 영향에 대해 다음과 같이 분석해 볼 수 있었다.

공회전 시간 증가에 따라 연료소모량도 증가하

는 경우로 분석 결과가 나타났다. 또, 동일한 차량에 대해서도 운전자가 일정한 운행패턴을 보이는 것이 아니라 연료소모비율이 좋은 경우와 나쁜 경우가 공존하는 것으로 나타났다. 2.5톤-A차량에 대해 주행시간에 대한 공회전 시간을 비교하여 계산한 공회전 비율을 검토한 결과 상위 10%로 연료소모비율이 좋은 경우는 평균 공회전 비율이 14.5%, 하위 10%의 연료소모비율이 나쁜 경우는 평균 공회전 비율이 70%로 나타나 공회전 비율이 적게 분포될 수록 연료소모비율이 좋음을 알 수 있었다.

급가속/급감속 횟수와 연료소모비율과의 비교를 위해 공회전비율이 비슷한 5개의 군집으로 나누어 군집분석을 수행하였다. 5개의 군집 중에서 공회전 비율이 비슷한 운행 중 연료소모비율이 가장 많이 차이나는 군집을 2개 선택하여 공회전 비율이 비슷한 경우에는 급가속 횟수가 높은 경우 연료소모비율이 낮아짐을 알 수 있었다. 연료소모비율이 높은 군집에서는 급가속 횟수가 평균 25.5회, 급감속 횟수는 평균 6.5회로 나타났고, 연료소모비율이 낮은 군집에서는 급가속 횟수가 평균 52회, 급감속 횟수는 11.5회로 나타났다. 과속 주행 비율과 연료소모비율과의 비교를 위해서는 일정한 운행시간동안 공회전 비율이 낮지만, 연료소모비율이 높은 운행에 대한 군집분석을 수행하였다. 시속 100km 이상의 속도로 주행하는 경우에는 공회전 비율과 급가속 횟수가 낮은 경우에도 연료소모비율이 낮음을 알 수 있었다. 본 논문의 결과는 경제운전에 대한 공회전 비율과 급가속/급감속 횟수, 과속주행 비율에 대한 임계치를 산정하는 정보로 활용할 수 있을 것이다. 예를 들어, 운전자가 차량을 운행할 때, 차량단말에서는 운전자의 공회전비율과 급가속/급감속 횟수, 과속주행 비율을 모니터링하여 각각 임계치 이상을 초과하는 경우에 대한 경고를 단계화하여 제공할 수 있다. 이를 통해 무의식중에 이루어지는 낭비 운행패턴을 경제 운행패턴으로 안내할 수 있어, 연료소모량을 절감할 수 있다.

추후, 운전자의 주요 운행시간에 대한 분석을 통해 유사한 거리에 대해 효율적인 운행시간대를 안내하는 방법, GPS와 연계된 운행 경로에 대한 분석

을 통해 동일 목적지에 대해서는 효율적인 라우팅 경로를 안내하는 방법 등에 대한 연구가 연계되어 진행되는 것이 필요하다. 연료소모량은 배출가스와 밀접한 관계를 가지므로 연료소모를 감소함으로써 배출가스도 감소될 수 있으므로 실시간으로 운전자에게 안내해주는 경제운전 방안은 그린 환경 구축을 위해서도 필요하다.

참 고 문 헌

- [1] 친환경 운전 10계명, 환경부, <http://www.me.go.kr>
- [2] 경제운전, 에너지관리공단, <http://bpm.kemco.or.kr/transport/>
- [3] 친환경 차계부, 자동차시민연합, <http://www.carten.or.kr>
- [4] 민석기, 문일기, 하정수, 이경수, “주행패턴 분석을 통한 운전자 모델 파라미터 추정,” *한국자동차공학회 춘계학술대회논문집*, pp. 1109~1114, 2002. 5.
- [5] 박용성, 최선모, 권해봉, 김종수, 엄성복, 조성우, “자동차 운전조건에 따른 연료소모특성에 관한 실험적 연구,” *한국자동차공학회 춘계학술대회논문집*, pp. 3~8, 2003. 5.
- [6] H. Kargupta, R. Bhargava, K. Liu, M. Powers, P. Blair, S. Bushra, J. Dull, K. Sarkar, M. Klein, M. Vasa, and D. Handy, “VEDAS: A mobile and distributed data stream mining system for real-time vehicle monitoring,” *Proc. SIAM Int. Data Mining Conf.*, pp. 300-311, April. 2004.
- [7] 이진학, “GIS와 공간 데이터 마이닝을 이용한 교통사고의 공간적 패턴 분석,” *대한지리학회지*, 제39권, 제3호, pp. 457~472, 2004. 6.
- [8] H. Marzi, C. MacNeil, and E. Marzi, “Assessing the effectiveness of data mining techniques,” *Proc. Int. Conf. Data Mining*, July. 2008.
- [9] R. Sahu, R. Anand, R. Agrawal, A. Vaid, and S. Kanojia, “Multidimensional association rules from large weather data set: A proposed methodology,” *Proc. Int. Conf. Data Mining*, pp. 298-303, July. 2008.

저자소개



김 현 숙 (Kim, Hyun-Suk)

2009년 3월~현재 : 충남대학교 박사과정(데이터마이닝전공)
2008년 3월~현재 : 한국전자통신연구원 책임연구원
1998년 9월~2008년 2월 : 한국전자통신연구원 선임연구원
1992년 8월~1998년 8월 : 한국전자통신연구원 연구원
1992년 8월 인하대학교 공학석사 (데이터베이스전공)



최 중 우 (Choi, Jong-Woo)

2004년 8월~현재 : 한국전자통신연구원 연구원
2004년 8월 : 서울대학교 석사과정 (실시간 운영체제 전공)



김 대 우 (Kim, Dae-Woo)

2007년 9월~현재 : 충남대학교 박사과정(컴퓨터통신 및 보안)
1993년 3월 ~ 현재 : KT 네트워크연구소 차장
1992년 8월 국민대학교 공학석사 (전자공학전공)



박 호 성 (Park, Ho-Sung)

2008년 3월~현재 : 충남대학교 석사과정(컴퓨터통신및보안전공)
2001년 3월~2008년 2월 : 충남대학교 학사(컴퓨터학과)



노 성 기 (Noh, Sung-Kee)

2006년 3월 충남대학교 박사졸업(컴퓨터네트워크 전공)
2007년 3월~현재 : 한국전자통신연구원 책임연구원
1998년 3월~2007년 2월 : 한국전자통신연구원 선임연구원
1992년 1월~1998년 2월 : 한국전자통신연구원 연구원
1992년 2월 포항공과대학 공학석사 (산업공학전공)



박 정 희 (Park, Cheong-Hee)

2005년 4월~ 현재 : 충남대학교 컴퓨터공학과 교수
2004년 8월 Univ. of Minnesota 컴퓨터공학과 공학박사 (데이터마이닝 전공)
2002년 8월 Univ. of Minnesota 컴퓨터공학과 공학석사 (컴퓨터그래픽스 전공)
1998년 2월 연세대학교 수학과 이학박사