

데이터마이닝의 신경망과 C4.5를 이용한 엔진오일 교환 시기 예측

▣ 홍 유 식 / 상지대학교

▣ 김 기 환 / 세명대학교

1. 서 론

어떠한 기계라도 윤활유가 없으면 부드럽게 작동하지 않는다. 자동차의 엔진은 그 회전 속도가 빠르고 그 힘 또한 크기 때문에 적정 양의 윤활유가 필요하다. 자동차의 부품 중에 윤활유가 필요한 것은 많지만 자동차에서 가장 중요한 부분인 엔진에 사용되는 엔진오일은 그 중요성을 아무리 강조해도 지나치지 않는다.

일반 운전자는 엔진 오일의 교환 시기를 잘 알지 못한다. 우리가 정비센터에 가더라도 정비소에서는 우리가 처음 엔진 오일을 교환한 시기를 묻고 또 얼마나 운행했는가를 묻는다. 아니면 엔진오일이 어느 정도 소모되었는지를 판별하기 위해서 엔진 오일스틱을 통해서 엔진 오일이 현재 어느 정도인지를 추출 하고, 점도를 손의 느낌에 의존해서 점도가 낮으면 교환한다는 지극히 추상적인 방법에 의존한다.

이와 같이 비과학적인 방법을 개선하기 위해서 신경망과 C4.5 알고리즘을 이용해서 엔진 오일의 교환 시기를 자동으로 알려주는 시스템을 설계하고자 한다. 엔진오일의 역할은 엔진 베어링 부분에서 금속과 금속의 접촉부분을 원할 하게 하는데 있다. 이것을 엔진오일의 윤활작용이라고 하는데, 이것 이외에도 엔진오일이 하는 역할은 여러 가지가 있다. 그 대표적인 예가

냉각작용이다. 라디에이터등의 냉각계통이 엔진의 과열을 방지하지만 엔진오일도 엔진내부의 부품 각 부분을 순환하면서 열을 낮추고 순환함과 동시에 엔진내부의 이물질을 씻어낸다. 이것을 청정작용이라고 한다. 그리고 엔진오일은 기밀 작용도 하는데, 피스톤과 실린더의 틈을 메워서 실린더 안의 여러 가지 혼합 가스나 배기가 엔진 내에 들어가지 않도록 한다. 그리고 엔진내부의 금속 표면에 부착된 오일은 금속이 부딪히는 부분에서 표면에 부착된 오일이 완충제로 작용하여 금속의 악화를 방지한다. 이것이 엔진오일의 완충작용이다. 엔진오일은 엔진의 하단부위의 오일 팬에 모여 이곳에서부터 오일펌프에 의하여 엔진내부를 순환한다. 오일계통의 시발점인 오일 팬으로부터 오일이 올라가는 부분에는 오일 스트레이너라는 금속필터가 있어 큰 이물을 제거하고 오일계통의 중간에는 오일필터가 있어 미세한 이물을 제거하여, 끝 부분에 달한 오일은 중력에 의해 밑으로 떨어져 오일 팬으로 돌아온다.

본 논문에서는 엔진오일을 최적의 시기에 교환하고, 엔진오일 교환시기를 자동적으로 감지하기위해서 2장에서는 기존 엔진오일 교환시기예측 방법을 알아보고, 3장에서는 전처리 기법을 이용한 신경망 기법을 이용하여 엔진오일 교환시기를 예측하고, 4장에서는 지능형 조건을 고려한 엔진오일 교환 알고리즘을 소

개 하고자 한다.

2. 엔진오일의 상태 점검

엔진오일은 주행시에는 고온이 되었다가 주차시에는 공기의 온도로 되돌아온다. 이처럼 온도의 오르내림을 되풀이하는 동안 오일은 나빠지기 시작하며 공기와 접촉하는 것만으로도 오일은 산화되며 성능이 저하되고, 물론 금속끼리의 마찰로도 성능이 많이 나빠진다.

또한 오래 사용하면 금속가루나 슬러지등의 이물질이 모여 오일과 섞이면서 오일의 기밀 작용이나 윤활 작용이 떨어지게 된다. 엔진오일이 부족하거나 매우 나빠졌을 때 가장 큰 문제점은 엔진이 타버리는 것입니다. 이것은 윤활이 불충분하게 되어 엔진이 달아올랐을 때 금속끼리 붙어버리는 것으로 이렇게 되면 정비할 때 많은 비용이 들게 됩니다. 주행 중 갑자기 엔진이 정지하는 위험한 경우가 바로 이런 때이다.

엔진오일을 교환할 경우, 색이 검다고 해서 반드시 교환 시기는 아니니만큼 반드시 끈적거림도 함께 점검해야 한다. 색깔이 검고 끈적거림까지 많으면 엔진오일 교환 시기가 되었음을 의미하니, 이때는 엔진 보호를 위해서 신속히 교환해 준다.

엔진오일은 내부에서 어느 정도 소모되기 마련이지만 그 정도가 심하고 소음기 끝에서 흰 연기까지 나오는 경우에는 내부 연소가 많은 것이니 원인을 찾아 반드시 수리해야 한다.

엔진오일을 교환시기를 비교적 정확하게 예측하기 위해서는, 엔진오일을 교환한 차량의 점도값, 주행거리, 단거리 전용 자동차, 장거리 전용 자동차 등의 과거의 데이터(history data)를 이용하여, 경향성을 파악할 수 있는 모형을 수립하고, 이 모형을 이용하여 미래를 예측하는 방법이다. 다시 말해서 설명한다면, 현재 시점이라고 할 때, t시점까지의 가용한 과거 데이터를 이용해서, 미래(t+1 시점)를 예측하는 것이다. 수요예측 방법은 크게 '정성적(qualitative)인 방법'과 '정량적(quantitative) 방법'으로 나누어 볼 수 있다. 정성적인 방법은 과거 데이터가 없거나, 수리적 모델링이 불

단순시계열법

$y(n) = a + b * x$ (단, $y(n) = n$ 기의 예측치)

$$a = \sum \frac{n}{y} b * x / n$$

$$b = \frac{\{n * \sum xy - (\sum x) * (\sum y)\}}{\{n * \sum x^2 - (\sum x)^2\}x} = \text{예측 년도} \quad n : \text{데이터 수}$$

이동평균법 (t期 이동평균법)

$$y(n) = \frac{\{x(n-t) + x(n-t+1) + \dots + x(n-1)\}}{t}$$

단, $y(n) : n$ 기의 예측치, $x(n-t) : (n-t)$ 기의 실적치, $t : \text{예측 기간}$

가중평균법 (t期 가중평균법)

$$y(n) = \{a1 * x(n-t) + a2 * x(n-t+1) + \dots + at * x(n-1)\}$$

단, $y(n) : n$ 기의 예측치, $x(n-t) : n-t$ 기의 실적치, $a1 : n-t$ 기의 가중치 ($\sum a = 1$), $t : \text{예측 기간}$

지수평활법

$$y(n) = a * x(n-1) + (1-a) * y(n-1)$$

단, $y(n) : n$ 기의 예측치, $x(n) : n$ 기의 실적치, $a : \text{평활계수}$

그림 1 엔진오일 교환시기 예측 모델

가능한 상황에서 사용되는데 전문가들의 지식과 의견에 따라 예측하는 것으로, Delphi 방법이 가장 많이 알려져 있다.

정량적인 방법은 크게 시계열 분석방법과 인과 분석방법이 있다. 시계열 분석 방법이란 변수 하나를 선택한 후에, 해당 변수의 과거 데이터를 근거로 해당 변수의 미래 값을 예측하는 방법이다.

수요예측 방법은 모두 시계열분석에 속한다. 인과 분석은 어떤 변수의 값이, 다른 변수들에 의해 영향을 받아 결정될 때에 다른 변수들의 과거 값과 해당 변수의 관계를 모델링하여 원하는 변수의 미래 값을 추정하는 방법이다.

이방법은 관계를 규명하기 어려울 뿐만 아니라, 타 변수들이 어떠한 경향을 가지고 변하고 있는지를 별도의 시계열분석과 같은 방법으로 분석해야 하는 등 상당히 많은 노력과 지식을 필요로 하는 단점이 있다.

3. 신경망 기법을 이용한 엔진오일 교환

본 논문에서 사용된 수요예측과정은 다음과 같으며, X축에는 시간, Y축에는 변수의 값(과거 데이터 값)을 의미한다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \epsilon \quad (1)$$

단, Y : 엔진오일 교환시기

X_1 : 종속변수에 영향을 주는 요인1

X_2 : 종속변수에 영향을 주는 요인2

X_3 : 종속변수에 영향을 주는 요인3

⋮

X_{10} : 종속변수에 영향을 주는 요인10

본 논문에 사용된 학습 신경망 구조는 다음과 같다.

- ① offsets, weight를 초기화한다
- ② input, target의 패턴을 신경망에 제시
- ③ 출력 신경세포들의 에러와 델타를 구해서 은닉층으로 역 전파 한다.

$$e_j = t_j - a_j$$

$$\delta_j = a_j (1 - a_j) e_j$$

- ④ 역 전파된 델타로부터 은닉층 신경세포들의 에러와 델타를 구해서 역 전파한다.

$$e_j = \sum_k w_{jk} \delta_k$$

$$\delta_j = a_j (1 - a_j) e_j$$

- ⑤ 델타 규칙에 의해서 연결가중치를 조절한다.

$$W(\text{new})_{ij} = W(\text{old})_{ij} + \alpha \delta_i a_j + \beta \Delta w_{ij}(\text{old})$$

$$\text{bias}(\text{new})_{ij} = \text{bias}(\text{old})_{ij} + \alpha \delta_i \cdot 1 + \beta \Delta \text{bias}_{ij}(\text{old})$$

- ⑥ 1-5의 과정을 모든 입력패턴에 대해서 반복 한다.
- ⑦ 4 과정을 신경망이 완전히 학습 될 때 까지 반복 한다.

<표 1>은 엔진오일 교환 시기를 예측 하기위한 5가지 서로 다른 조건을 입력 하였을 때 최종 엔진오일 교체시기를 예측하는 과정을 나타내고 있다.

신경망 학습의 초기값을 설정하는 것은 중요한 문제다. 초기값을 적절하게 선택함으로써 학습오차가 작

고 학습과정이 빠르게 수렴될 수 있기 때문이다. 일반적으로 신경망의 학습은 특정 초기값에서 시작한다.

그리고 학습률은 모수 값들을 어떻게 선택하는냐에 따라서 학습오차가 작으면서 학습과정이 빠르게 수렴할 수도 있고 초기 포화점에 빠질 수도 있다. 그렇기 때문에 분석하고자 하는 자료에 적당한 모수를 선정하여 오차가 최소 값이면서 학습과정이 빠르게 수렴될 수 있게 학습하도록 하는 것은 매우 중요한 문제다.

그래서 제한적이지만 x, θ, ϕ, μ (kappa, theta, phi, mu)만을 가지고 각 범위 0.1, 0.3, 0.5, 0.7, 0.9에 따라 모든 경우를 고려해서 임의의 경우로 실험을 해보았다. 그리고 학습시간을 각각 500회로 제한하였다.

(그림 2)는 다음과 같은 처리를 수행한다.

- ① 신경망을 이용하여 10개의 서로 다른 조건 테스트 데이터를 학습시킨다.
- ② 10개의 테스트 데이터에 대하여 예측을 한 뒤 테스트 데이터와 예측 데이터의 오차를 계산한다.

$Z_1, Z_2, Z_3, \dots, Z_n$: 테스트 데이터

$\hat{Z}_1, \hat{Z}_2, \hat{Z}_3, \dots, \hat{Z}_n$: 예측값

$$e_i = Z_i - \hat{Z}_i \quad (1)$$

i 시점 시계열 테스트 자료와 예측값에 대한 차이

$$Z'_j = Z_j + W(Z_j) \quad (2)$$

여기서, Z'_j 는 j 번째 특이 값으로 식별된 테스트 테

표 1 신경망을 이용한 엔진오일 교환시기 전문가시스템 입력 데이터

입력조건	엔진오일 수명감소	엔진오일 수명증가
1. 과거 12개월치 엔진오일양	small	Big
2. 과거 12개월치 엔진오일 점도	small	Big
3. 과거 12개월치 주행거리	Big	Small
4. 과거 12개월치 도로 조건	Big	Small
5. 운전자 급발진 급제동 습관	Small	Big

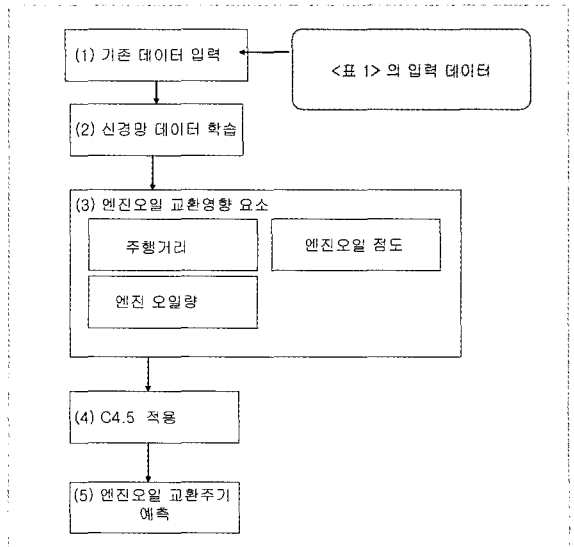


그림 2 엔진오일 교환 시기 예측과정

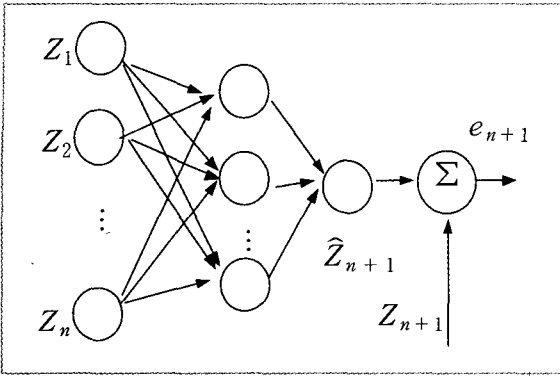


그림 3 신경망을 이용한 엔진오일 교환 시기 예측 모형

이터 Z_j 의 수정된 값을 의미한다.

- ③ 후처리로 데이터 마이닝의 C4.5 알고리즘을 이용한다.

4. C4.5 알고리즘을 이용한 엔진 오일교환

본 논문에서는 신경망 알고리즘을 이용해서 최적의 엔진오일 교환 주기를 추출하는 것이 목적이다. 신경망 알고리즘은 자료 분석 분야에서 복잡한 구조를 가지고 있는 자료에 대하여 예측 문제를 해결하기 위한 유연한 비선형 모형의 하나로 분류될 수 있다. 인간의 신경생리학과 유사성 때문에 일반적으로 다른 통계적 예측모형에 비해 보다 흥미롭게 연구 되어지고 있다. 특히, 예측 기법으로써 로지스틱 회귀분석 보다 신경망의 우수함을 비교한 연구들이 고려되고 있다.

그러나 신경망은 미래의 목표 값을 예측하는데 있어 입력벡터의 값의 수나 형태를 결정할 수 있는 체계적인 방법의 결여와 모델의 분류가 어떻게 이루어지는지 명확하게 이해 할 수 없는 단점이 제시되고 있다, 이러한 단점을 해결하기 위하여 신경망에서 상징적 분류 규칙을 찾거나 의사결정 나무를 통하여 이해 할 수 있는 해석을 얻고자 하는 연구 등이 이루어졌다.

C4.5의사결정 나무를 형성하기 위하여 처음 수행하는 작업이 분할정복이다. 입력되는 훈련 집합이 성공적으로 분할 되도록 모든 하부 집합에 하나의 클래스가 속하는 경우들로 구성될 때까지 나무를 형성한다.

노드를 분리하는 기준으로 정보이득비율(Informa-

tion gain ratio)이 사용된다. 주어진 예를 분류하기 위하여 요구되는 평균정보를 가장 감소시킬 수 있는 방법으로 현재의 혼련을 분리하기 위한 것이다.

나무형 분류/회귀의 많은 부분은 통계학자들에 의하여 개발되었지만 흔히 기계학습(machine learning)기법으로 간주되고 있다.

나무형 모형의 특징은

If A, Then B, Else C

형식의 규칙으로 표현된다는 것이다. 예컨대

If GENDER = 'm', then RESPONSE = 'yes',

Else RESPONSE = 'no'

와 같은 모습이다.

너무 단순한 것이 아닌가 싶겠지만 그렇지 않다. 앞의 일반적인 표현에서 B와 C가 다시 If-Then-Else의 구조를 가진 규칙일 수 있기 때문이다.

나무구조의 결정 규칙을 생성하기 위하여 각 단계에서 p 개의 설명변수 중 어느 것에 의하여 가지분리를 할 것인가를 선택해야 한다. 이 때 결정 규칙들은 각기 다른 기준을 쓰는데, C4.5는 엔트로피 기준을 사용한다.

엔트로피(entropy)는 열역학에서 쓰는 개념으로 무질서도에 대한 척도이다. 자료셋 T 가 Y 에 의하여 k 개의 범주로 분할되고 범주 비율이 p_1, \dots, p_k 라고 하자. T 의 엔트로피는

$$\text{Entropy}(T) = -\sum_{i=1}^k p_i \log p_i$$

로 정의된다. ($0 \log 0 = 0$ 으로 함, 편의상, 밑수가 $e = 2.7182$ 인 로그로 함).

예컨대, 4개의 범주가 (0.25, 0.25, 0.25, 0.25)의 비율로 섞여 있는 경우(T_0)는

$$\text{Entropy}(T_0) = -(0.25 \log 0.25) + 4 = 1.39$$

이지만, 4개의 범주가 (0.5, 0.25, 0.25, 0)의 비율로 섞여 있는 경우(T_1)

$$\begin{aligned} \text{Entropy}(T_1) &= -(0.5 \log 0.5) + 0.25 \log 0.25 \\ &\quad + 0.25 \log 0.25 = 1.04 \end{aligned}$$

로 T_1 의 엔트로피가 T_0 의 엔트로피보다 작다. 이것은 T_1 이 T_0 에 비해 덜 무질서 하여 예측이 쉽기 때문이다.

C4.5 모형은 엔트로피 기준에서 가장 엔트로피를 낮

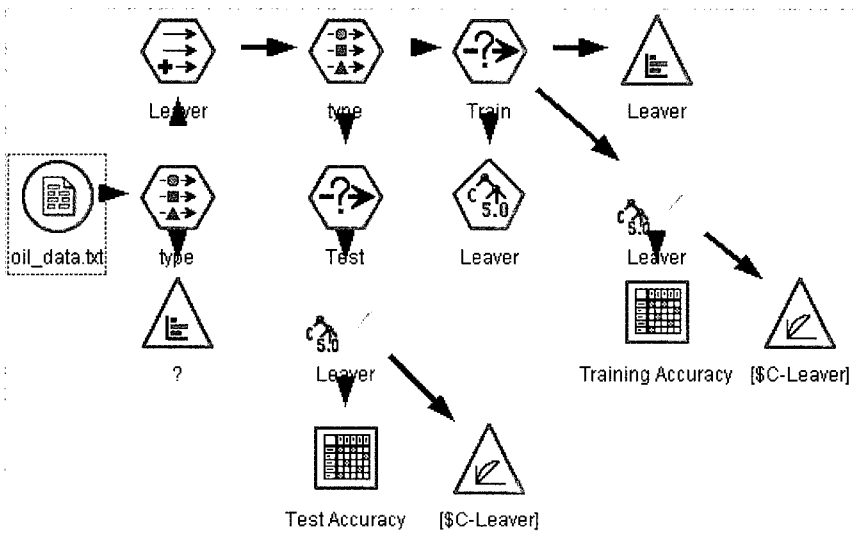


그림 4 나무형 분류 분석

추는 분리 변수를 찾고자 한다. 분리변수를 찾음으로서 가장 성취도가 좋은 변수 및 수준을 찾는 것이 나무규칙 생성 알고리즘이다.

본 논문에서는 클레멘타인 패키지를 이용하여 C4.5 알고리즘을 적용하여 신경망 알고리즘에서 설명할 수 없는 엔진 오일의 감소 및 교환에 영향을 주는 가장 큰 요소가 무엇인지를 찾아낼 수가 있다.

실험 데이터는 <표1>의 데이터를 이용하였고, 실험 결과 <표 2>와 같은 예측 결과를 얻었다.

C4.5 알고리즘을(그림 4) 실행한 결과 가장 큰 원인으로 주행거리, 장거리운행량이 가장 주요원인으로 확인하였다.

표2 엔진오일 교환 시기 예측효율 비교

(신경망 예측) / 군집(Clustering) 적용							엔진오일 교환시기 효율(%)	
주행 거리	급발진	엔진오일양	엔진오일	단거리	장거리	도로조건	지능 방식	기존 방식
정상:5000KM	급제동 조건		점도	운행차량	운행차량			
BIG	BIG	BIG	BIG	SMALL	SMALL	SMALL	48	58
BIG	SMALL	SMALL	SMALL	BIG	BIG	BIG	55	62
BIG	SMALL	SMALL	SMALL	SMALL	SMALL	SMALL	62	77
SMALL	SMALL	SMALL	SMALL	SMALL	SMALL	SMALL	87	79
BIG	medium	medium	medium	medium	medium	medium	92	84
SMALL	SMALL	medium	SMALL	SMALL	SMALL	SMALL	44	52
SMALL	SMALL	BIG	medium	BIG	medium	BIG	82	73
SMALL	BIG	SMALL	medium	BIG	BIG	BIG	92	66
SMALL	BIG	medium	BIG	SMALL	medium	BIG	73	62

5. 결론

엔진오일의 중요성은 누구나 말하지만 그 교환주기에 대해서는 의견이 많이 다르다. 어떤 사람은 5,000km당 교환을 해야 한다고 하고 어떤 사람은 10,000km, 혹은 믿을 수 없지만 20,000km라고 하는 사람도 있다. 차량과 운전 성향에 따라 약간의 차이가 있기 때문에 무엇이 정확하다고는 할 수 없지만 일반적인 경우 일년에 두

번 정도 교환하는 것이라고 알려져 있다.

그러나 자동차마다 차이가 있을 수도 있을 수 있으니 자신의 엔진오일 교환 시기는 스스로 점검을 하고 판단하시는 것이 좋다. 엔진의 무병장수를 위해서는 엔진오일은 필요조건이다. 시기적절하게 교환하지 않으면 엔진에 큰 무리가 가고 나쁜 경우에는 엔진이 망가질 수도 있다.

엔진오일 교환 시기 예측은 데이터 마이닝의 신경망 알고리즘과 C4.5 알고리즘을 이용하여 비교적 정확하게 예측이 가능하다.

본 연구에서는 엔진오일 교환시기의 예측을 보다 정확하게 보정하기위해서는 단거리주행차량, 장거리주

행차량, 엔진오일 점도, 엔진 슬러지, 도로 조건 등을 포함하여 신경망 알고리즘과 C4.5 알고리즘을 적용하였다.

본 논문에서 제안한 알고리즘은 이제까지 상업적인 제품으로 이용한 적이 없지만, 만일 완성차를 만드는 데 이러한 알고리즘을 이용할 수 있다면 차량의 유지보수에 도움이 될 것으로 기대한다.

참 고 문 헌

[1] 이영재외 3인, "주행거리 누적에 따른 엔진오일의 열화에 관한 연구", 한국자동차공학회 1999년도 춘계학술대회
 [2] 이래덕, 김한준, 세메노프, "자동차 엔진오일 열화상태 in situ 측정용 전기용량 센서", 2001년 7월 센서학회지 제10권 제4호
 [3] 강우, 김현철, "미래형자동차 기술개발 현황", Proceeding of the 6th PKATEC Workshop, 2001.

[4] Box, G. E. P. and Jenkins, (1976), G.M., "Time Series Analysis forecasting and control", SanFrancisco : Holden-Day.
 [5] Box, G. E. P. and G. C. Tiao, "Comparison of forecast and actuality," Appl. Statist., 25, 195-200, 1976.
 [6] 이상원, "학습하는 기계 신경망", Ohm사, p.412, 1995.
 [7] 장남식, 홍성완, 장재호, "데이터 마이닝", 대청, p202, 1999
 [8] 허명희, 이용구, 데이터마이닝 모델링과 사례, 아카데미 출판사, 2003.7.
 [9] Tian Zhang, Rahu Ramakrishnan, and Riron, "Data Mining and Knowledge Discovery," p141-182, 1997.
 [10] Tom M.Mitchell, MC Graw Hill "MACCHINE LEARNING," p414, 1997.