

Identification and classification of fresh lubricants and used engine oils by GC/MS and bayesian model

Nam Yee Kim[★], Geum Mun Nam, Yuna Kim, Dong-Kye Lee, Seh Youn Park,
Kyoungjae Lee¹ and Jaeyong Lee¹

Chemical Analysis Sec. Forensic Science Div. National Forensic Services, Seoul 158-707, Korea

¹*Department of Statistics, Seoul National University, Seoul 151-742, Korea*

(Received April 30, 2013; Revised August 13, 2013; Accepted January 29, 2014)

GC/MS 분석과 베이지안 분류 모형을 이용한 새 윤활유와 사용 엔진 오일의 동일성 추적과 분류

김남이[★] · 남금문 · 김유나 · 이동계 · 박세연 · 이경재¹ · 이재용¹

국립과학수사연구원 법과학부 화학분석과, ¹서울대학교 통계학과

(2013. 4. 30. 접수, 2013. 8. 13. 수정, 2014. 1. 29. 승인)

Abstract: The aims of this work were the identification and the classification of fresh lubricants and used engine oils of vehicles for the application in forensic science field-80 kinds of fresh lubricants were purchased and 86 kinds of used engine oils were sampled from 24 kinds of diesel and gasoline vehicles with different driving conditions. The sample of lubricants and used engine oils were analyzed by GC/MS. The Bayesian model technique was developed for classification or identification. Both the wavelet fitting and the principal component analysis (PCA) techniques as a data dimension reduction were applied. In fresh lubricants classification, the rates of matching by Bayesian model technique with wavelet fitting and PCA were 97.5% and 96.7%, respectively. The Bayesian model technique with wavelet fitting was better to classify lubricants than it with PCA based on dimension reduction. And we selected the Bayesian model technique with wavelet fitting for classification of lubricants. The other experiment was the analysis of used engine oils which were collected from vehicles with the several mileage up to 5,000 km after replacing engine oil. The eighty six kinds of used engine oil sample with the mileage were collected. In vehicle classification (total 24 classes), the rate of matching by Bayesian model with wavelet fitting was 86.4%. However, in the vehicle's fuel type classification (whether it is gasoline vehicle or diesel vehicle, only total 2 classes), the rate of matching was 99.6%. In the used engine oil brands classification (total 6 classes), the rate of matching was 97.3%.

요 약: 국내 시판제품으로 서울시내에서 구입한 산업용 윤활유, 이륜구동 윤활유, 선박용 윤활유, 자동차용 윤활유(엔진오일, 수동 변속기 기어유, 자동변속기 오일) 등 80종(기유 4종 포함)의 새 윤활유들(80 classes)과 8종의 경유 차량과 16종의 휘발유 차량에 각각 3종씩의 경유와 휘발유 전용 엔진 오일

[★] Corresponding author

Phone : +82-(0)2600-4960 Fax : +82-(0)2600-4792

E-mail : nykim60@korea.kr

로 교환하여 차량별 및 주행거리별로 각각 채취한 사용 엔진 오일 86 종을 GC/MS로 분석한 TIC로 데이터베이스를 만들고, 새 윤활유와 사용 엔진오일들의 동일성 추적과 차량별 분류를 위하여 차원 축소와 베이지안 방식의 분류 모형을 개발하였다. 새 윤활유의 분류는 웨이블렛 적합방법과 주성분 분석방법으로 차원 축소하여 베이지안 방식의 분류 모형을 적용한 결과 각각 97.5%와 96.7%의 정분류율을 보여 차원 축소는 웨이블렛 적합방법이 더 좋은 결과를 나타냈다. 그리고 새 윤활유의 분류에서 선택된 웨이블렛 적합방법의 차원 축소와 베이지안 방식의 분류 모형에 의한 사용 엔진 오일의 차량별 분류(총 24 classes)는 86.4%의 정분류율을 보였고, 경유 차량인지 휘발유 차량인지를 구분하는 차량 연료 타입별 분류(총 2 classes)는 99.6%의 정분류율을 나타내었고, 사용 엔진 오일 브랜드별 분류(총 6 classes)는 97.3%의 정분류율을 나타내었다.

Key words: forensic science, lubricants, engine oil, classification, Bayesian model, wavelet fitting, GC/MS

1. 서 론

법과학적 측면에서 윤활유는 뺑소니 교통사고, 살해, 보험관련 위장 교통사고, 범칙 현장에서 범인과 피해자 간 윤활유 전이 등으로 용의자 추적 및 사건 해결의 단서가 되는 중요한 증거물이며 최근에는 주행 중 차량 화재 또는 차량 고장 등의 원인으로 윤활유가 지목되는 경우도 증가하고 있다. 그래서 현장 증거물인 미지 시료에서 윤활유의 존재 여부, 윤활유의 종류, 용의 차량 또는 압수품과의 동일성 여부 등의 감정은 범인 검거, 사건 해결 및 사건 재구성 등에 중요한 단서를 제공한다.¹⁻⁹

윤활유는 맞닿아 움직이는 부위를 분리시켜 열과 표면 마모를 유발하는 마찰을 줄일 뿐만 아니라 열의 제거, 퇴적물 축적 방지, 물 또는 잔해 조각 등 외부 입자 제거, 녹과 부식 방지, 밀봉성 증가, 열에너지를 전달, 동력 전달 등의 목적으로 사용되므로 차량 뿐만 아니라 모든 움직이는 부위에 필요하여 각종 기계류, 차량 등 일상생활 주변에서 폭넓게 사용되고 있다.¹⁰⁻¹¹

윤활유는 기유에 여러 가지 첨가제를 혼합하여 제조하는 데, 기유의 종류로 광유계, 반합성유, 합성유 등이 있고, 첨가제로는 산화방지제, 청정분산제, 점도지수 향상제, 유동점 강하제, 극압제, 유성 향상제, 방청제, 소포제 등이 있다.¹² 용도에 따른 종류로 자동차용 윤활유, 산업용 윤활유, 선박용 윤활유 등이 있다. 그리고 자동차용 윤활유는 다시 가솔린 엔진 오일, 디젤 엔진 오일, 2-사이클 엔진 오일, 수동변속기 기어 오일, 자동변속기 오일 등이 있으며, 산업용 윤활유는 유압작동유, 기계유, 공업용 기어유, 압축기유, 열매체유, 전기절연유, 그리스 등이 있다.¹³

특히 윤활유는 작동 과정에서 높은 온도와 산소 등

에 노출되어 퇴화가 진행되므로 사용 정도에 따라 첨가제들의 분해산물들, 산화 생성물들, 먼지, 수분, wear metals 등의 오염물질이 함유되므로 사용 전 윤활유와는 다소 차이를 보일 수 있다.^{2,7}

본 연구는 두 가지 목적으로 진행하였는데, 첫째는 국내에서 시판되고 있는 윤활유(새 윤활유)들을 GC/MS로 분석하여 얻은 Total Ion Chromatogram(TIC) 데이터로 분류 규칙을 만들어 미지 윤활유 시료에 대한 종류 및 분류(classification)와 동일성 추적을 위한 베이지안^{14,15} 분류 모형을 개발하고자 하였다.

두 번째는 윤활유 증거물은 뺑소니 교통사고와 연관이 있는 경우가 많은데 주로 차량 하부에 묻은 윤활유가 피해자의 의류나 소지품 등에 묻거나 사고로 인해 윤활유가 새어 바닥에 유류되기도 한다. 이런 사건에서 자주 제시되는 증거물들인 사용 중인 차량 엔진 오일들에 대한 분석을 시도하였다. 엔진 오일의 종류별, 차량별로 엔진 오일 교환 후 주행 거리에 따라 채취한 사용 엔진 오일들을 새 윤활유 분류 모형으로 개발된 베이지안 통계 분류 모형을 적용하여 사용 엔진 오일들의 동일성 추적과 차량 및 엔진 오일 브랜드별 분류를 시도해 보고자 하였다.

2. 실험 및 방법

2.1. 새 윤활유 시료

실험에 사용된 윤활유는 2011년 3월 서울 시내 오일 총판점 및 온라인 대리점에서 시판하는 완제품을 구매하였고, 첨가제 혼합 전인 윤활 기유(base oil, BA)는 윤활유 전문회사((주) S-oil)에서 협조를 받았으며 그 목록을 Table 1에 나타내었다. 구입한 윤활유는 용도별로 다음과 같이 분류하였다. 윤활 기유 (base

Table 1. List of fresh lubricants used in this study

Sample No. (Class No.)	Usage	Brand name	Manufacturer
1		P-31	S-oil
2	Base Oil (BA)	P-96	S-oil
3		US-6	S-oil
4		PAO-4	S-oil
5		ZIC Supervis 46	SK
6		Gold Speed Tap	Golden Trade
7		Royco 782, hydraulic fluid	Royal Lubricants, US
8		SK, sewing machine oil	SK
9		Pipe machine oil	DS Lube & Chem
10	Industrial Lubricants (ID)	SKAF power steering oil	SK Chemicals
11		LPS 2 Lubricants	LPS
12		RANDO 46 HD CZ	GS Caltex
13		Speed Tap	Korean Golden Trade
14		KF-96-1, Silicone	Shin Etsu, Japan
15		Tri-Flow superior lubricants	KRYLON, US
16		Super Lube	Synco Chemical, US
17		Vacuum pump oil	Inland Vacuum Industries
18		GS Dexron DX - III	GS Caltex
19		Dragon Dexron III	S-oil
20		Daewoo Motor ATF Dexron III	Daewoo Moto Company
21		SHELL ATF III	Shell Korea
22		TERACAN ATF	Hyundai MOBIS
23	Automatic Transmission Fluid (AT)	Mobil MT ATF	Exxon Mobil
24		ATF	Hyundai MOBIS
25		GS ATF HD	GS Caltex
26		SUPER ATF (for SSANGYONG)	76 LUB/US
27		MobilI Synthetic ATF (old model)	Exxon Mobil Oil Corp.
28		MobilI Synthetic ATF (new model)	Exxon Mobil Oil Corp.
29		Mobil ATF 550	AMSOIL
30			FX Gear 75w85
31		GS HD Gear	GS Caltex
32	Gear Oil (GR)	Gear oil, genuine parts	Hyundai MOBIS
33		LSD Gear oil, genuine parts	Hyundai MOBIS
34		THUBAN GL-4 Gear oil	GS Caltex
35		Gear oil HD	GS-Caltex
36		Mobil Synthetic Gear Lubricant	Exxon Mobil Oil Corp.
37	Boat use (BT)	Gold Power 1000	Coastalunilube INC, US
38		2 Stroke multi purpose oil	Valvoline Korea
39		Power ACE 2T	L&C
40	2-Cycle use (CC)	ELF 2T	S-oil total lubricants
41		Super 2 Stroke oil	GS Caltex
42		Super Stroke oil	GS Caltex
43		DS-2-Cycle motor oil	DAESUNG Petrochemical

Table 1. Continued

Sample No. (Class No.)	Usage	Brand name	Manufacturer
44		Kixx G1	GS Caltex
45		ZIC A	SK Lubricants
46		ZIC XQ RV/SUV	SK Lubricants
47		PREMIUM Gasoline engine oil	S-oil
48		PREMIUM LS Diesel engine oil	SK
49		Kixx RV	GS Caltex
50		SSANGYONG MOTOR genuine parts engine oil for diesel/gasoline	Isu Chemical, Korea
51		RENAULT SAMSUNG MOTORS genuine parts engine oil for gasoline/LPG	SK Lubricants
52		ZIC XQ Gasoline LPG fully synthetic	SK Lubricants
53		GM Daewoo diesel engine oil	Isu Chemical, Korea
54	Engine oil (EG)	SSU GXO gasoline engine oil	S-oil
55		SSU DXO Turbo diesel engine oil	S-oil
56		SSU Euro XT	S-oil
57		Kixx D1 Turbo	GS Caltex
58		Kixx PAO	GS Caltex
59		Mobil1	Exxon Mobil Oil Corp.
60		Mobil 1 Extended Performance	Exxon Mobil Oil Corp.
61		Elf excellium full-tech	Elf
62		Ravenol SSL	Ravenol
63		BPvisco7000	BP Korea
64		Castrol EDGE Sport	BP Korea
65		Castrol Magnetic Professional	Castrol
66		REPSOLELITE Evolution Fuel Economy	Repsol YPF Lubricants
67		RAVENOL FO	Ravenol
68	2-Cycle use (CC)	New hypol 2T, high performance 2 Stroke for engine oil premium grade	S&T Motors Shell Korea
69		New hypol 2T, high performance motor cycle engine oil premium grade	S&T Motors Shell Korea
70		Shell, advance SX 2T, 2-cycle engine oil	S&T Motors Shell Korea
71		Zic, MAHA 2T	SK, Lubricant
72		Valvoline, Turbo 2T, advanced synthetic, motor cycle oil	Valvoline Korea
73	Engine oil (EG)	Mobil Super 3000	Mobil Korea
74		MOTUL 300V chrono	France
75		REPSOL ELITE Evolution	Repsol YPF Lubricants
76		RAVENOLVMO	Germany
77		Kendall GT-1 With Liquid Titanium	Kendall Motor Oil
78		Mobil 1	Exxon Mobil Oil Corp.
79		ZICXQ (10.4.30)	SK Lubricants
80		AMSOIL High Zn/phosphorus	US

oil) 4 종, 유압유, 작동유, 공업용 등에 사용하는 일반 산업용 윤활유(industrial use, ID) 13 종, 선박용 윤활유(boat oil, BT) 1 종, 오토바이, 스쿠터, 예초기 등에

사용하는 2-사이클(2-cycle, CC) 전용 윤활유 11 종 그리고 자동차용 윤활유로 자동변속기용 오일 (automatic transmission fluid, AT) 12 종, 수동변속기 및 액슬 등

에 사용하는 기어유(gear oil, GR) 7종, 엔진 오일(engine oil, EG) 32종으로 모두 80개의 새 윤활유를 대상으로 하였다.

새 윤활유는 각각의 윤활유를 class로 하여 윤활유 시료 번호와 class 번호를 동일하게 하였다(Table 1). 그리고 한 class당 3개씩 시료를 만들어 3개의 데이터를 얻었다. Class는 번호로, 3개의 시료는 대괄호 속의 1,2,3으로 표기하였다. 예를 들어 20번 윤활유(20 class)의 두 번째 시료는 “20[2]”로 표기하였으며,

이는 20 class에 속한다.

2.2. 사용 엔진 오일

2.2.1. 사용 엔진 오일 시료

실험에 사용된 엔진 오일(Table 2)은 2.1항의 시료와 동시에 구매하였다. ZIC는 SK의 모든 엔진 오일에 사용하는 패밀리브랜드이며, Kixx는 GS칼텍스의 모든 엔진 오일에 사용하는 패밀리브랜드이고, 현대순정부품은 GS칼텍스와 SK가 제조하여 현대모비스가 판매

Table 2. The 6 kinds of engine oil brand for mileage

Fuel type	Brand name	Products name	Manufacturer	SAE
Gasoline engine oil	ZIC A	ZIC A Gasoline synthetic VHVI, 4 L	SK Lubricants	5w30
	Kixx G1	Kixx G1, 4 L	GS Caltex	5w30
	Hyundai G	PREMIUM Gasoline engine oil, 4 L	S-oil	5w20
Diesel engine oil	ZIC XQ	ZIC XQ RV/SUV Diesel fully synthetic, 1 L	SK Lubricants	5w40
	Kixx RV	Kixx RV, 6L	GS Caltex	10w40
	Hyundai D	PREMIUM LS Diesel engine oil, 6 L	SK	5w30

Table 3. The information of vehicles for used engine oils

Class of vehicle	Fuel type	Brand name	Vehicle information			Vehicle's mileage (km)
			Brand	Model	Model year	
UD1	Diesel	Zic XQ	Kia	Spotage	2005	126,900
UD2		Kixx RV	Hyundai	Starex	2006	72,362
UD3		Kixx RV	Hyundai	Starex	2010	5,238
UD4		Kixx RV	Hyundai	Grand Starex	2011	2,450
UD5		Hyundai D	Hyundai	Starex	2002	210,000
UD6		Zic XQ	Hyundai	Grace	1999	146,542
UD7		Zic XQ	Hyundai	Trajet XG	2005	43,108
UD8		Hyundai D	Hyundai	Starex	2003	106,643
UG1	Gasoline	Zic A	Hyundai	Accent	1996	205,200
UG2		Zic A	Samsung	SM5	1998	160,000
UG3		Zic A	Kia	K7	2006	53,345
UG4		Kixx G1	Kia	Spectra	2000	219,722
UG5		Kixx G1	Kia	Sepia	1997	146,474
UG6		Kixx G1	Samsung	SM5	1998	17,800
UG7		Hyundai G	Hyundai	NF Sonata	2004	42,390
UG8		Hyundai G	Kia	Credos	1997	202,571
UG9		Hyundai G	Hyundai	EF Sonata (LPG)	2004	115,419
UG10		Kixx G1	Hyundai	Avante	2009	24,724
UG11		Zic A	Hyundai	Grandeur XQ	1999	104,838
UG12		Hyundai G	Hyundai	Avante	2005	73,200
UG13		Zic A	Kia	Morning	2009	25,070
UG14		Hyundai G	Hyundai	Avante (hybrid)	2010	10,266
UG15		Zic A	Hyundai	Sonata	1999	79,290
UG16		Hyundai G	Hyundai	Avante	2009	34,400

한다. SK와 GS칼텍스가 시장규모의 약 60%를 점유하고 있고 그 외 쌍용, 이수화학, 모빌, 한국캐스트롤, 한국셀석유 등이 한 자릿수 썩의 점유율을 나타내고 있다. 기유 특성으로 보면, 국내 유통되는 대부분의 엔진오일은 광유계 엔진 오일 또는 광유계 기유를 한 번 더 정제하여 점도 지수를 강화한 정제 광유계 오일(VHVI, Very High Viscosity Index 급)이다. 이와 같은 국내 차량 및 엔진 오일의 유통 상황을 감안하여 본 연구에서는 국내 정유 업체 제품을 대상으로 하였으며, 가장 대중적인 제품으로 Table 2와 같이 휘발유 차량용과 경유 차량용 각각 3종씩 브랜드 제품을 선정하여 사용하였다(사용 엔진 오일 브랜드 class는 총 6종). 이들 제품은 모두 정제 광유계 엔진 오일이고 제품별로 SAE (Society of Automotive Engineers) 점도가 다소 차이가 있다.

2.2.2. 차량 선정

실험에 사용된 차량은 생산년도, 차량 주행거리, 운전자 특성 등은 상관하지 않고 자발적 참여자로 한정하였으며 실험 전에 교환하였던 엔진 오일(브랜드명 또는 제조사 등)에 대한 정보는 대부분 알지 못하여 실험 전 차량에 사용했던 엔진 오일에 대한 정보는 고려하지 않았다. 한 브랜드 엔진 오일 제품을 3대 이상의 차량에 엔진 오일 필터와 함께 교환한 다음부터 새로 주행거리를 산정하였고, 실험에 사용된 차량 주행거리, 차량의 사양 및 실험에 사용한 엔진 오일은 Table 3과 같으며, 경유 차량 8종과 휘발유 차량(LPG 차량 1종 포함) 16종으로 모두 24종을 대상으로 하였다.

2.2.3. 엔진오일 시료의 채취 및 보관

주행거리에 따른 엔진 오일의 채취는 차량별로 엔진 오일 교환 후 주행 거리 약 1,000~1,500 km 마다 채취하였으며, 사용 엔진 오일 채취는 PVC 재질의 수액세트를 엔진 오일 룸에 넣은 후 필터를 사용하여 약 3 mL 썩을 뽑아내고 이를 질소 가스로 밀어 내어 유리 재질의 갈색 캡 바이알에 담아 뚜껑을 닫은 후 실험 전까지 상온에서 보관하였다. 우리나라 운전자들의 개략적인 엔진 오일 교환 시점인 약 5,000~6,000 km를 종료 시점으로 하여 차량 당 1~5번의 엔진 오일을 채취하였다 그러나 차량 사용 중 예기치 못한 변수가 발생하여 약 5,000 km까지 주행하지 못한 경우에도 그 차량에서 채취된 것까지만 실험에 임하여 24개 차량에서 86종의 사용 엔진 오일 시료를 채취

하였다.

사용 엔진 오일은 차량에 따라 경유 차량 엔진 오일은 기호 UD1~UD8의 8개로, 그리고 휘발유 차량 엔진 오일은 UG1~UG16의 16개로 모두 24개의 차량별 class를 만들었다(Table 3). 또 각 class에서 주행 거리 약 1,000 km마다 채취한 1~5개의 시료로 subclass(차레로 A, B, C 등 표기)를 만들어 주행 거리에 따른 86개의 subclass를 만들었다. 그리고 새 윤활유와 마찬가지로 한 subclass의 엔진 오일 당 3개씩의 시료를 만들어 3개의 데이터를 얻었고 이는 대괄호로 표기하였다. 즉 UD4 차량에서 2,500 km 주행 후 채취한 엔진 오일에서 1번째 시료는 “UD4-B[1]”으로 표기하였다.

2.3. 가스크로마토그래피 / 질량분석법 (GC/MS)에 의한 분석

2.3.1. 기기

새 윤활유 및 사용 엔진 오일 분석에 사용한 분석 장치는 mass selective detector가 부착되고 자동주입기인 7683 autosampler가 부착된 7890A/5975C gas chromatography/mass spectrometry (Hewlett Packard, USA)를 사용하였다. 머무름 시간을 보정하기 위하여 RT_Lock method를 적용하여 사용하였다.

2.3.2. 시료 분석

기기 분석을 위한 시료는 노르말-헥산으로 묶혀 약 1%로 만든 후 1 μ L를 GC/MS에 주입시켜 Table 4에 나타난 조건으로 분석하였다. GC/MS 분석은 solvent delay time을 2분으로 하고 총 40분간 분석하여 데이터를 얻었다.

2.3.3. 정밀도 분석

동일 시료를 같은 날 3회 반복하여 측정 한 분석 결과와 동일 시료를 3회씩 매일 약 3주까지 반복 측정한 분석 결과를 비교하여 각각 within day와 between days에 대한 정밀도를 알아보았다.

2.4. 베이지안 방식의 분류 모형

2.4.1. 데이터 요약

TIC 데이터를 요약하지 않고 그대로 이용하여 분류 모형을 만들 경우, 하나의 TIC 데이터에 너무 많은 변수가 포함되어 있기 때문에 웨이블릿 적합, 주성분 분석 등의 차원 축소 기법을 활용하여 데이터 요약을 시도하였다. 먼저 모든 윤활유의 데이터를 같은 척도에서 비교하기 위해 주어진 데이터를 표준화시킨다

을 통계적 방법을 사용하여 주어진 데이터를 요약하고 요약된 값에 따라 분류를 실시하였다. 데이터 요약은 다음의 세 가지 방법으로 각각 수행하였다.

1) 웨이블릿 적합^{14,16,17}

웨이블릿은 일반적으로 제곱의 적분이 유한한(L²-함수) 함수를 추정하기 위한 방법 중 하나이다. 웨이블릿은 자료의 큰 피크를 잘 적합할 수 있다는 장점이 있는데, TIC 자료 분석에서는 작고 큰 피크가 중요한 의미를 가지므로 웨이블릿을 사용하는 것이 적절하다고 생각되었다. 웨이블릿 적합에서는 기본 함수로 두 가지의 함수를 사용하는데, 이 함수들을 웨이블릿이라고 부르고, 그 중 하나를 father 웨이블릿- ϕ 이라고 하고, 다른 하나를 mother 웨이블릿- ψ 이라고 하면, 이 두 함수는 다음의 조건을 만족시켜야 한다.

$$\int \phi(t)dt = 1, \int \psi(t)dt = 0$$

위의 두 함수를 이용하여 다음과 같은 함수열을 만들고, 그것들을 이용하여 원하는 함수를 추정하는 것이 웨이블릿의 함수 추정 방법으로 함수열은 다음과 같다.

$$\left\{ \phi_{jk}(t) = 2^{-\frac{j}{2}} \phi(2^{-j}t - k), j, k \in \text{자연수} \right\},$$

$$\left\{ \psi_{jk}(t) = 2^{-\frac{j}{2}} \psi(2^{-j}t - k), j, k \in \text{자연수} \right\}$$

만약 어떤 함수 f에서 $p = 2^J$ 개의 관측값을 얻었다고 하면, 웨이블릿의 방법을 이용하여 함수 f를 다음과 같이 추정할 수 있다.

$$f(t) \approx \hat{f}(t) = c_{j_0} \phi_{j_0}(t) + \sum_{j=1}^J \sum_{k=0}^{2^j-1} d_{jk} \psi_{jk}(t)$$

본 연구에서는, 관측된 하나의 TIC 데이터를 하나의 함수에서의 관측값으로 보고 이에 대한 웨이블릿 적합을 실시하였다. 웨이블릿은 Daubechies least asymmetric with 10 vanishing moments 웨이블릿, 즉 $\int t^l \psi(t)dt = 0, l = 0, \dots, 9$ 를 만족하는 웨이블릿을 사용하였고, 적합 결과로 나오는 웨이블릿의 계수들로 자료를 요약하였다. 이 때, 하나의 TIC 데이터에 관측값 (변수)이 6219 개밖에 없기 때문에 이를 2의 승수로 만들기 위해 자료의 끝에 1973 개의 0을 집어넣어서 자료의 개수가 8192, 즉 2^{13} 이 되도록 하였다.

웨이블릿의 적합 결과로는 항상 자료의 개수와 같은 개수의 웨이블릿 계수가 나오게 된다. 하지만 이

계수를 모두 사용하는 것은 자료를 모두 사용하는 것과 다름없어서 차원을 줄이고 자료를 효과적으로 요약하기 위해 thresholding이라는 것을 적용하였다. 일반적으로 웨이블릿에서의 thresholding은 하나의 함수에 대한 웨이블릿 계수들 중 큰 절대값을 가지는 것들만으로 함수를 추정하는 방법을 말하는 데 여기서는 총 240 개의 TIC 데이터 각각을 웨이블릿으로 적합 시키기 때문에 240 개 함수에 대한 계수들을 공통적으로 thresholding 해야 한다.

TIC 데이터 하나에 대한 웨이블릿 계수들을 $y_i = (y_{i1}, y_{i2}, \dots, y_{ip}), i = 1, \dots, 240$ 이라 하면 여기서 p는 자료의 개수인 8192 개 이다. y_{ij} 들의 절대값의 95% 백분위수를 λ_i 라 하고, 다음의 식 $\sum_{i=1}^{240} I(y_{ij} \geq \lambda_i)$ 은 j번째 웨이블릿 계수의 절대값이 λ_i 보다 큰 값을 가지는 TIC 데이터의 개수를 나타낸다. 즉, j번째 웨이블릿 계수가 정해진 threshold 보다 큰 값을 가지는 GC 자료의 개수이다. 적절한 수 m에 대하여

$$\sum_{i=1}^{240} I(y_{ij}(|\geq \lambda_i) \geq m$$

이 되도록 하는 계수들만을 이용하여 데이터를 요약하는 것이 여러 함수의 웨이블릿 계수를 공통적으로 thresholding 하기 위하여 제안된 방법 중의 하나이다.¹⁴ 본 연구에서는 95% 백분위수로 threshold값을 정한 뒤 $m=10$ 으로 하여 계수를 선택하였다. 그 결과 8192 개의 웨이블릿 계수 중 896 개만이 선택되었고 이 계수들로 각각의 TIC 데이터를 대표하여 요약하였다.

20번 시료 (20[1])의 경우로 보면 Fig. 1에서 보듯이 원 TIC에서 표준화시켜 동일 척도로 만든 후 웨이블릿 계수와 threshold 이후 선택된 웨이블릿 계수는 896 개 이었다.

2) 주성분분석(Principal Component Analysis - PCA)¹⁸

관측한 자료의 개수에 비해 변수의 개수가 많거나 자료의 손쉬운 해석을 위해 설명하는 변수의 개수를 줄이고 싶을 때 사용하는 기법 중의 하나가 바로 주성분분석이며 이런 자료를 설명하는 변수의 개수를 줄이는 것을 차원축소(dimension reduction)라 한다. 주성분분석은 기존의 변수들의 선형결합(linear combination)으로 새로운 변수를 만들어 그것들로 자료의 변동을 설명하는 방법이라고 할 수 있다.

하나의 관측치를 $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ 라 하고 표준화 되어있다고 가정하고, 다음의 행렬

$$X = (x_1, x_2, \dots, x_n)^T$$

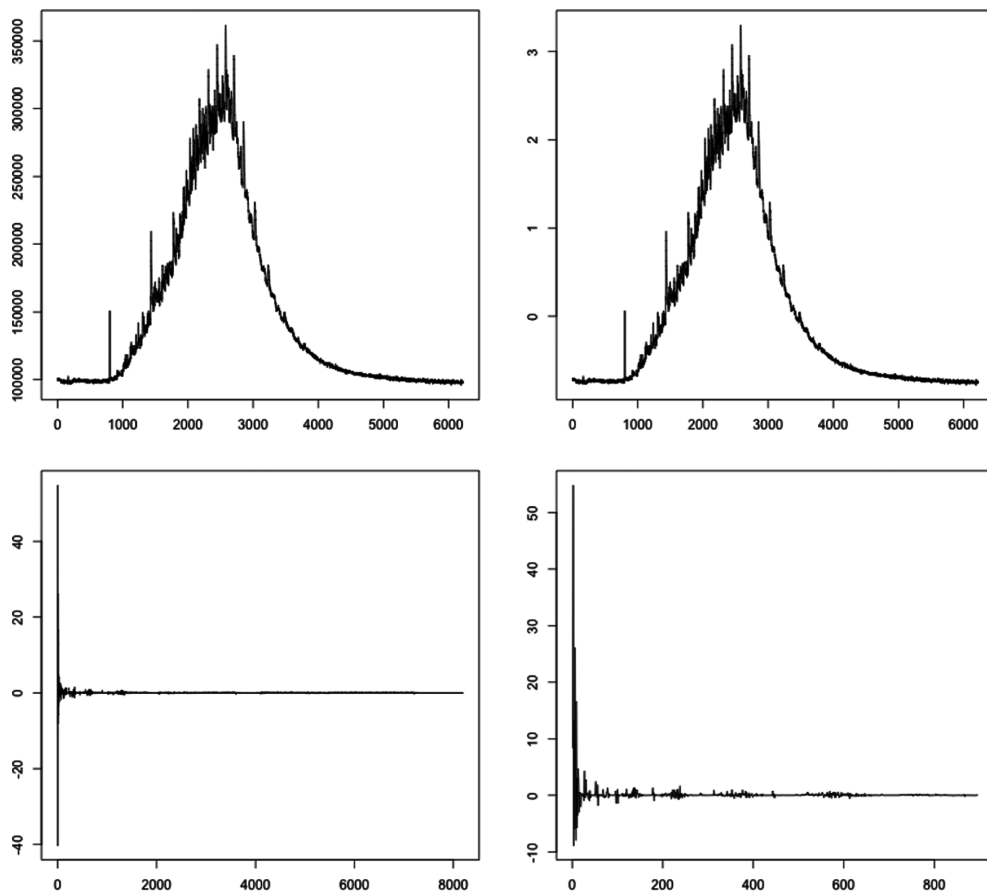


Fig. 1. TIC and chromatogram with normalization (upper from left), and wavelet coefficient after wavelet fitting and selected wavelet coefficient after thresholding (bottom, from left) of sample 20[1].

는 전체 자료의 행렬을 나타낸다. 이 때 X의 공분산 행렬은

$$S \equiv \frac{1}{n-1} \sum_{i=1}^n x_i x_i^T \equiv \frac{1}{n-1} X^T X$$

이 된다. 주성분분석은 자료의 변동을 가장 잘 설명할 수 있는 새로운 변수들을 찾아내는 방법으로 새로운 변수들은 기존 변수들의 선형 결합으로 만들며 변동을 가장 많이 설명하는 변수들을 순차로 구한다. 이러한 과정으로 공분산 행렬 S(표본상관계수 행렬로도 가능)의 고유벡터를 이용해 기존 변수들의 선형 결합하여 새로운 변수를 만든다. 각 변수들이 가지는 변동에 대한 설명력은 대응되는 고유값의 크기에 비례한다. 일반적으로 주성분분석 결과로 나오는 변수들 중 자료의 변동을 충분히 설명한다고 생각되는 개수의 주성분만을 선택하여 차원축소에 이용한다.

본 연구의 자료는 관측한 TIC 데이터의 개수가 240개, 변수의 개수(관측값 개수)는 6219개이므로, 주성분분석을 이용하여 차원축소를 시키기 적절하며 변수의 개수가 관측치의 개수보다 많아서 상관관계수에 대한 주성분분석을 바로 적용하기 어려우므로, 자료 자체에 대한 특이값분해(Singular Value Decomposition-SVD)를 이용하여 고유값과 고유벡터를 계산후 주성분분석을 적용하였다. 주성분분석 적용 결과로 얻어진 240개의 주성분 중 자료의 변동을 99% 정도 설명하는 27개의 주성분만을 선택하여 그 주성분 점수로 자료를 요약하였다. 자동차 자동 변속기 오일인 20[1] 시료의 TIC 데이터를 27개의 주성분으로 요약한 결과를 Fig. 2에 나타냈다.

3) 원 자료

주어진 데이터를 손실 없이 모두 이용하였다. 위의 두 방법에 비해 정보의 손실이 없다는 장점이 있으나,

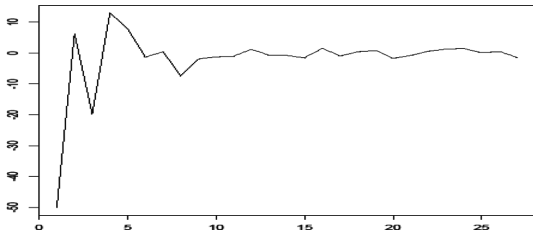


Fig. 2. The PCA score of 27 in sample 20[1].

작은 오차(noise)에 상대적으로 민감하고 변수의 개수가 너무 많아 계산이 오래 걸린다는 단점이 있었다. 또한 자료의 작은 변화에도 가장 민감하게 반응할 수 있는 방법이기 때문에, 위 두 방법(웨이블렛, 주성분 분석)보다 안정적이지 못한 단점도 있다.

2.4.2. 분류 방법

위에서 구한 자료의 요약값들을 이용하여 분류를 실시하였다. 이미 각 윤활유들이 어떤 종류인지 알고 있는 상황이므로, 주어진 정보를 이용하여 같은 종류의 윤활유로 3 개씩의 시료를 만들어 하나의 class로 구분하였다. 이러한 경우 베이지안 방법을 이용한 분류 문제는 다음과 같다.

자료의 요약값들을 $X_{11}, X_{12}, X_{13}, X_{21}, \dots, X_{80,1}, X_{80,2}, X_{80,3}$ 이라 하면, X_{ji} 는 j번째 윤활유 종류에서 i 번째로 관측한 자료에 대한 요약값을 나타낸다. 이것이 혼합 정규분포(mixture of normal distribution)에서 나왔다고 가정하고, 즉,

$$X_{j1}, X_{j2}, X_{j3} \sim \pi_j \equiv N_d(\mu_j, \sigma_j^2 I_d), j = 1, 2, \dots, K$$

$$P(\pi_j) = p_j$$

의 모형을 가정한다. 여기서 $N_d(\mu, \Sigma)$ 는 평균이 μ , 분산 행렬이 Σ 인 d 차원 다변량 정규분포이며 d 는 요약값 X_{ji} 들의 차원을 나타낸다. 또한 $P(\pi_j)$ 는 새로운 윤활유가 j번째 윤활유에 속할 확률을 나타낸다. 시료의 각 윤활유 종류를 편의상 class라 명하면, 윤활유 종류가 80 개 이므로, 우리의 자료에서 class의 개수인 K 는 80이다. 각 시점에서 관측된 자료들이 서로 독립이라는 의미로 정규분포의 분산행렬을 대각행렬로 가정하였다. 요약하자면 위 모형은 관측치 X 가 각 $j=1, \dots, 80$ 에 대하여 p_j 의 확률로 평균이 μ_j , 분산 행렬이 $\sigma_j^2 I_d$ 인 다변량 정규분포를 따름을 나타낸다. 미지의 값인 p, μ, Σ 는 추정치를 구하여 대입하는 plug-in 방식으로 하였다. 먼저 p 는 각 class에 들어갈 확률을 나

타내는데, 아무런 정보도 주어지지 않은 경우이므로 이것을 모든 class j 에 대하여 $\frac{1}{80}$ 이라고 가정하여, 즉 동등한 사전확률을 주어 편향이 일어나지 않도록 하였다. 다음으로 μ 와 σ^2 은 표본을 이용하여

$$\hat{\mu}_j \equiv \frac{1}{3} \sum_{i=1}^3 X_{ji}, \hat{\sigma}_j^2 \equiv \frac{1}{(3-1)d} \sum_{i=1}^3 \sum_{k=1}^d (X_{ji}^k - \hat{\mu}_j^k)^2,$$

$$X_{ji} = (X_{ji}^1, \dots, X_{ji}^d)^T$$

와 같이 불편추정량(unbiased estimator)을 구하여 사용하였다. 베이지안 방식의 분류 모형을 이용하면, 새로운 자료 x_{new} 에 대하여 각 class로 분류될 확률은

$$P(x_{new} \in j^{th} \text{ class} | x_{11}, \dots, x_{80,3}, \hat{\mu}_j, \hat{\sigma}_j^2 I_d, \hat{p})$$

$$\propto N_d(x_{new}; \hat{\mu}_j, \hat{\sigma}_j^2 I_d), j = 1, \dots, 80$$

이고, 이를 최대로 하는 class j 로 새로운 자료를 분류한다. 여기서 $N(x; \mu, \Sigma)$ 는 평균이 μ , 분산이 Σ 인 정규분포의 확률밀도함수에 x 를 대입한 값을 의미한다.

3. 결과 및 고찰

3.1. 새 윤활유들의 분류 및 동일성 추정

3.1.1. 정밀도

윤활유 분석의 재현성을 알아보기 위하여 2.3항에 의하여 분석한 GC/MS의 TIC는 Fig. 3에서 보듯이 붉은색의 영역이 정밀도를 나타내는 것으로 양호한 정밀도를 보였다.

3.1.2 베이지안 방식의 분류

2.4 항에서 제시한 세 가지 방법으로 자료를 요약한 뒤 주어진 240 개 시료 데이터로 분류 규칙을 만들었다. 이 때 생성된 분류 규칙의 신뢰도를 보기 위하여 cross validation 방법을 사용하여 분류를 정확하게 해내는지를 살펴보았다. 즉 첫 번째 자료 x_{11} 에 대해서 그 자료를 제외한 나머지 239 개의 데이터만으로 분류 규칙을 만든 뒤, x_{11} 이 첫 번째 class에 들어가게 분류하는지를 보았다. 다음은 각각의 요약 방법에 따른 분류 결과이다.

1) 웨이블렛 적합법

웨이블렛 적합법을 이용한 베이지안 모델로 분류하는 과정은 먼저 TIC 원 데이터를 표준화하고 웨이블렛 적합법을 사용하여 웨이블렛 계수로 요약하여 분류를 진행한다. 20[1] 시료는 전체에서 20 class에 속

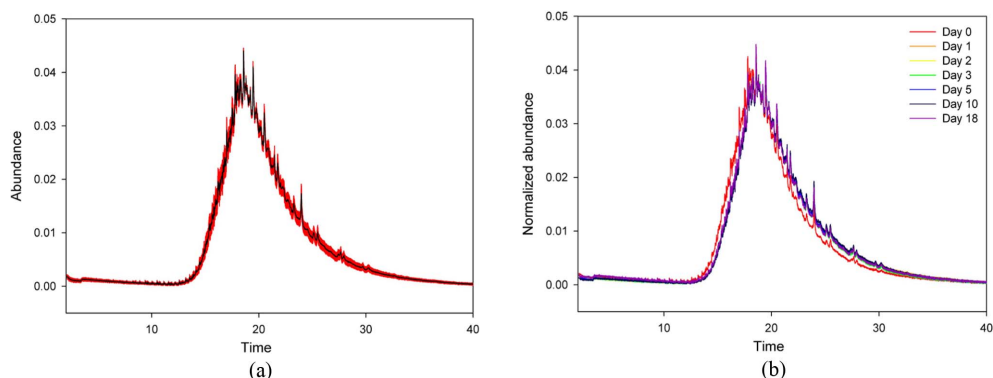


Fig. 3. TICs of lubricant sample (a) within day and (b) between days.

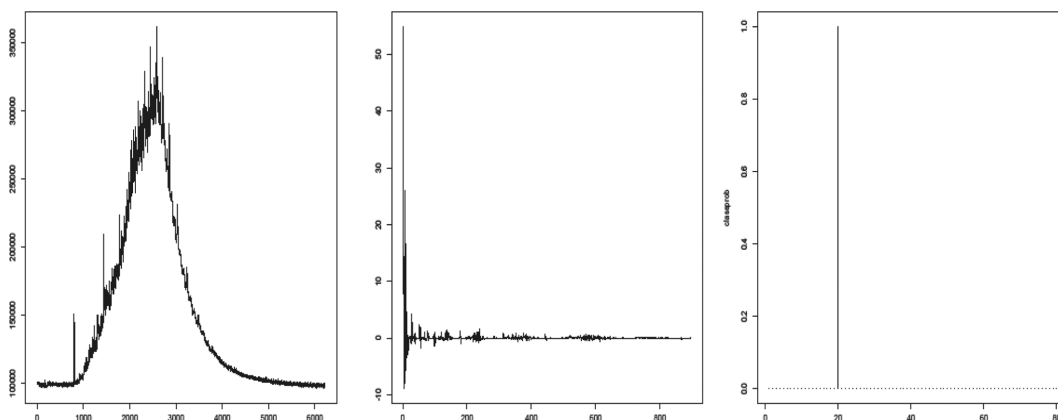


Fig. 4. The procedure of Bayesian model with wavelet fitting, that is chromatogram with normalization, wavelet coefficients, and probability of classification for sample 20[1] (from left).

하는 윤활유인데 이 시료를 예로 들어 분류를 진행해 보면 Fig. 4처럼 표준화한 후 웨이블릿 적합을 하고 마지막으로 오른쪽 그림처럼 20[1] 시료가 20번째 class로 분류될 확률이 1에 가까운 것으로 나타나게 된다. 이렇게 웨이블릿 계수로 자료를 요약한 다음 베이즈안 분류 모형을 적용하여 분류하면 총 240 개의 시료 중 6 개를 오분류하였다. 이는 오분류율이 2.5%로 매우 좋은 결과를 보였다. 특히 전체 계수를 다 사용하지 않고 896 개라는 상대적으로 적은 수의 계수만을 사용하여 매우 좋은 결과를 얻을 수 있었으며, 그 결과를 Table 5에 나타내었다. 특히 TIC 데이터에서 각 윤활유들의 특성을 구분하는 중요한 특징인 좁고 큰 피크를 웨이블릿이 잘 인식하기 때문에 이러한 결과를 얻을 수 있었다고 판단된다.

그 예로 1[1](base oil, P-31) 시료의 데이터와 45[3](Zic A engine oil) 시료의 데이터를 보면 Fig. 5에서와 같이 표준화시킨 후의 두 크로마토그램은 전반적

Table 4. Gas chromatography/Mass spectrometry condition for lubricants analysis

Parameter	Condition
Column	DB-5 MS (30 m×0.25 mm×0.25 μm)
Oven Temp.	120 °C (2 min)-10 °C/min-280 °C (30 min)
Injection Temp.	350 °C
Spilt ratio	30:1
Carrier gas flow	1 mL/min, He
ionization energy	70 eV
Injection Vol.	1 mL
Analysis time	2~40 min.

인 종 모양의 형태는 비슷하지만 세부적으로는 서로 차이를 나타내고 있음을 쉽게 알 수 있다. 이들 각각의 웨이블릿 계수 요약값은 Fig. 6에 나타냈듯이 크로마토그램의 차이점이 계수에 반영되어 웨이블릿 계수의 값이 서로 다른 것을 확인할 수 있었다.

그리고 오분류된 6 개를 살펴보면 (Table 6) 1번

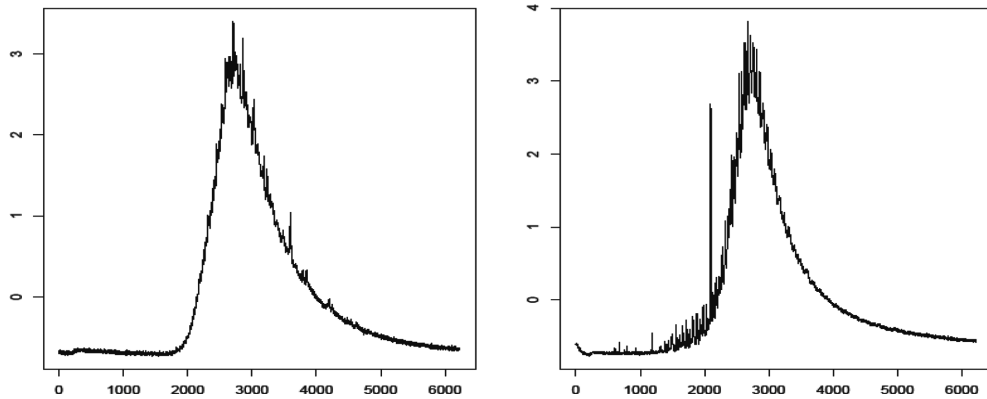


Fig. 5. Chromatograms with normalization of sample 1[1] and 45[3] (from left).

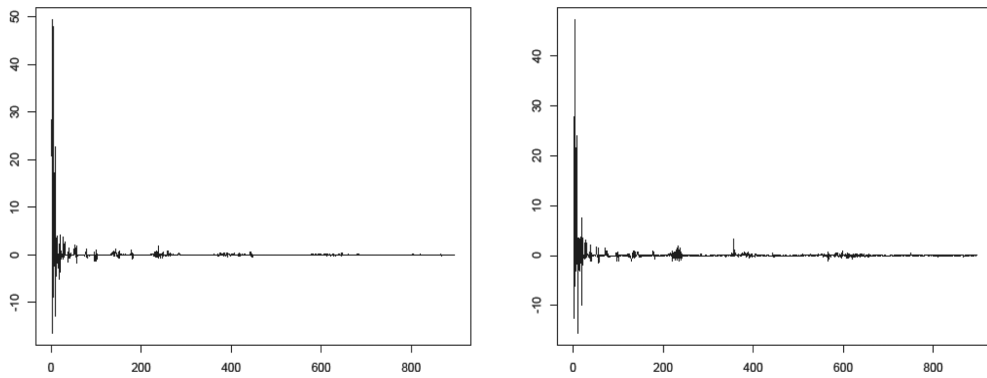


Fig. 6. The wavelet coefficients of sample 1[1] and 45[3] (from left).

윤활유의 경우 3 개의 시료에서 두 개는 자기 class인 1 class로 정분류되었고 나머지 하나인 1[2] 시료만 50 class로 오분류되었으며, 41번 윤활유의 세 개 시료는 모두가 오분류되었는데, 41[1] 시료는 40 class로, 41[2] 시료와 41[3] 시료는 모두 42 class로 오분류되었고, 50번 윤활유 중 50[2]만 1 class로 오분류되었으며 이들의 오분류 확률은 모두 1에 가까웠다.

1[2] 시료와 50[1] 시료를 보면 Fig. 7에서 알 수 있듯이 크로마토그램들이 서로 매우 유사 하여 전반적인 경향은 물론 작은 피크의 모양까지 매우 흡사한 모습을 보였다. 또한 TIC 데이터를 하나의 벡터(vector)로 보고 유클리디안 거리를 계산 해 보면 1[2] 시료로부터 50[1] 시료까지의 거리는 16.3431(표준화시킨 TIC 자료를 이용)정도로 매우 가까웠는데, 이는 시료 1[1], 1[3]까지의 거리인 93.1079, 63.0150에 비해 매우 작은 값을 보였다. 따라서 오분류가 되는 것은 웨이블릿 계수를 이용한 분류 규칙의 문제라기보다 크로마토그램 자체가 매우 유사하여 구분이 되지 않는 것이라고 판단

되었다. 실제로 1번 시료는 첨가제를 첨가하지 않은 상태의 광유계 기유(base oil)이므로 이수화학 엔진 오일인 50번 윤활유는 1번 시료를 기유로 사용하고, 주로 무기물 첨가제를 사용한 것으로 사료된다.

또한 41번과 42번 시료는 동일회사의 2-사이클용의 유사 브랜드 윤활유로 실제 크로마토그램에서도 매우 유사함을 보이고 있다(Fig. 8). 아마도 거의 동일한 조성을 가지고 다른 브랜드명으로 시판되고 있는 것으로 사료된다.

2) 주성분 분석

주성분 분석 후 27 개의 주성분만을 선택하여 자료를 요약한 다음 베이지안 분류 모형을 적용하여 분류하면 총 240 개의 시료 중 8 개를 오분류(Table 7)하였다. 오분류율이 3.3%로 양호한 결과를 얻었고 분류 결과는 Table 5에 나타내었다. 웨이블릿과 마찬가지로, 6219 개의 변수를 27 개의 주성분으로 요약하여 분류했음에도 분류가 잘 이루어졌다. 변수의 개수를 줄임으로써 계산이 빨라지는 이점을 가지면서도 분류의

Table 5. The classification result of fresh lubricants by Bayesian model with various data preprocessing techniques

Sample number	Wavelet fitting		PCA		TIC data	
	Classification	Misclassification	Classification	Misclassification	Classification	Misclassification
1	2	1	1	2	2	1
2	3	0	3	0	3	0
3	3	0	3	0	3	0
4	3	0	3	0	3	0
5	3	0	3	0	3	0
6	3	0	3	0	3	0
7	3	0	3	0	3	0
8	3	0	3	0	3	0
9	3	0	3	0	3	0
10	3	0	3	0	3	0
11	3	0	3	0	3	0
12	3	0	3	0	3	0
13	3	0	3	0	3	0
14	3	0	3	0	3	0
15	3	0	3	0	3	0
16	3	0	3	0	3	0
17	3	0	3	0	3	0
18	3	0	3	0	3	0
19	3	0	3	0	3	0
20	3	0	2	1	3	0
21	3	0	3	0	3	0
22	3	0	3	0	3	0
23	3	0	3	0	3	0
24	3	0	3	0	3	0
25	3	0	3	0	3	0
26	3	0	3	0	3	0
27	3	0	3	0	3	0
28	3	0	3	0	3	0
29	3	0	3	0	3	0
30	3	0	3	0	3	0
31	3	0	3	0	3	0
32	3	0	3	0	3	0
33	3	0	3	0	3	0
34	3	0	3	0	3	0
35	3	0	3	0	3	0
36	3	0	3	0	3	0
37	3	0	3	0	3	0
38	3	0	3	0	3	0
39	3	0	3	0	3	0
40	3	0	3	0	3	0
41	0	3	0	3	1	2
42	2	1	2	1	2	1
43	3	0	3	0	3	0
44	3	0	3	0	3	0
45	3	0	3	0	3	0
46	3	0	3	0	3	0
47	3	0	3	0	3	0
48	3	0	3	0	3	0
49	3	0	3	0	3	0
50	2	1	2	1	2	1

Table 5. Continued

Sample number	Wavelet fitting		PCA		TIC data	
	Classification	Misclassification	Classification	Misclassification	Classification	Misclassification
51	3	0	3	0	3	0
52	3	0	3	0	3	0
53	3	0	3	0	3	0
54	3	0	3	0	3	0
55	3	0	3	0	3	0
56	3	0	3	0	3	0
57	3	0	3	0	3	0
58	3	0	3	0	3	0
59	3	0	3	0	3	0
60	3	0	3	0	3	0
61	3	0	3	0	3	0
62	3	0	3	0	3	0
63	3	0	3	0	3	0
64	3	0	3	0	3	0
65	3	0	3	0	3	0
66	3	0	3	0	3	0
67	3	0	3	0	3	0
68	3	0	3	0	3	0
69	3	0	3	0	3	0
70	3	0	3	0	3	0
71	3	0	3	0	3	0
72	3	0	3	0	3	0
73	3	0	3	0	3	0
74	3	0	3	0	3	0
75	3	0	3	0	3	0
76	3	0	3	0	3	0
77	3	0	3	0	3	0
78	3	0	3	0	3	0
79	3	0	3	0	3	0
80	3	0	3	0	3	0
Total	234	6	232	8	235	5
Rate of misclassification	2.5%		3.3%		2.08%	

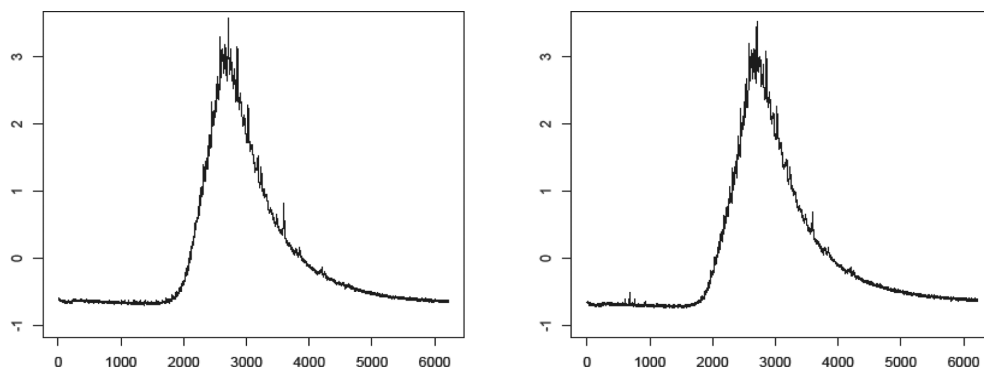


Fig. 7. Chromatograms with normalization of sample 1[2] and 50[1] (from left).

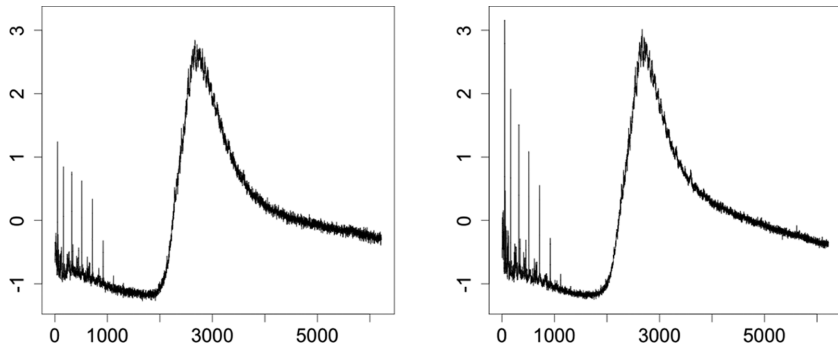


Fig. 8. Chromatograms with normalization of sample 41[2] and 42[1] (from left).

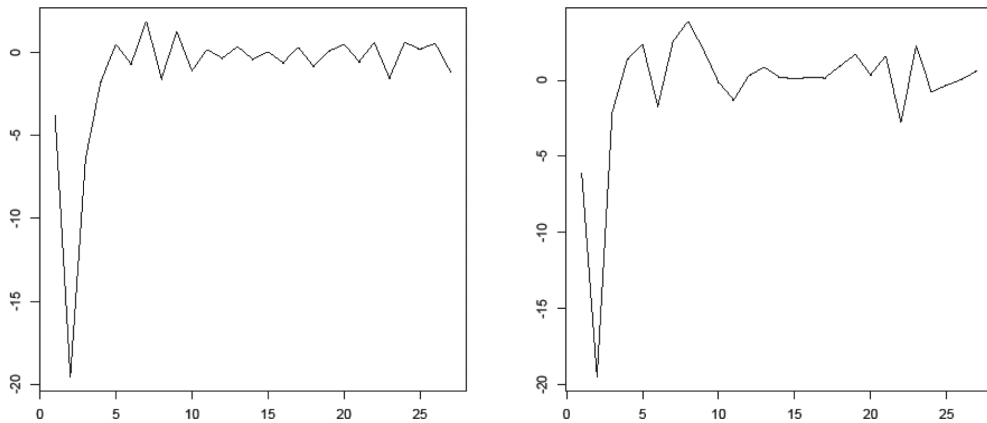


Fig. 9. The scores of principal component of sample 1[2] and 45[1] (from left).

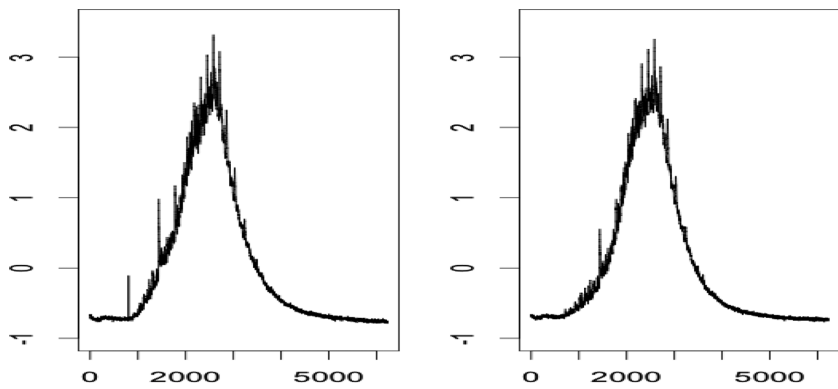


Fig. 10. Chromatograms with normalization of sample 20[3] and 26[1] (from left).

정확도를 유지하였는데, 이는 27 개 주성분이 전체 자료 변동의 99% 이상을 설명함으로써 원 자료의 특징을 잘 나타냈기 때문인 것으로 보인다.

주성분 분석이 자료를 잘 구분하는 지를 보기하기 위하여 1[1] 시료 데이터와 45[3] 시료 데이터에 대한 주성분 점수를 Fig. 9에 비교해 놓았는데 두 시료의

주성분 점수 차이가 확연히 나타나고 있음을 알 수 있었다. 그래서 주성분 분석으로 자료가 잘 구분된다고 사료되며 주성분 분석을 이용한 분류의 결과가 좋다는 것을 확인 할 수 있었다.

그러나 오분류되는 경우로 20[3]과 26[1]의 두 크로마토그램들(Fig. 10)은 서로 매우 유사한데 전반적인

Table 6. The samples of misclassification by Bayesian model with wavelet fitting

Sample	1[2]	41[1]	41[2]	41[3]	42[1]	50[2]
Brand name	Base oil P-31	Super 2 Stroke oil	Super 2 Stroke oil	Super 2 Stroke oil	Super Stroke oil	SSANGY- ONG MOTOR
Result	50 class	40 class	42 class	42 class	41 class	1 class
Probability of misclassification	1	1	1	1	1	1

Table 7. The samples of misclassification by Bayesian model with PCA

Sample	1[2]	1[3]	20[3]	41[1]	41[2]	41[3]	42[1]	50[2]
Brand name	Base oil P-31	Base oil P-31	Daewoo Motor	Super 2 Stroke oil	Super 2 Stroke oil	Super 2 Stroke oil	Super Stroke oil	SSANGYONG MOTOR
Result	50 Class	50 Class	26 Class	40 Class	42 Class	42 Class	41 Class	1 Class
Probability of misclassification	1	0.5037	0.9990	1	0.9266	1	1	1

Table 8. The samples of misclassification by Bayesian model with data itself

Sample	1[2]	41[1]	41[3]	42[1]	50[2]
Brand name	Base oil P-31	Super 2 Stroke oil	Super 2 Stroke oil	Super Stroke oil	SSANGYONG MOTOR
Result	50 Class	40 Class	42 Class	41 Class	1 Class
Probability of misclassification	1	1	1	1	1

경향은 물론 작은 피크의 모양까지 매우 흡사하여 20[3] 윤활유는 26 class로 오분류되었다. 이 시료 각각을 하나의 벡터로 보고 유클리디안 거리를 재보면 20[3]에서 26[1]까지의 거리는 5.6314로, 같은 집단에 속한 20[1], 20[2]과의 거리인 4.6933, 4.7930보다 조금 더 크긴 하지만 근소한 차이라고 생각된다. 이로써 20[3] 시료가 오분류되는 것 또한 시료 자체의 구분이 어렵기 때문이라고 생각할 수 있다.

3) 원 자료

TIC 데이터 자체를 사용하여 베이지안 분류 모형을 적용하여 분류하면 총 240 개의 시료 중 5 개만 오분류(Table 8) 하였다. 즉 오분류율은 2.08%로 가장 좋은 결과를 보였다. 분류의 정확도는 웨이블릿보다 좋지만, 계산속도 면에서 가장 느리다는 문제점이 있었다. 분류 결과는 Table 5에 나타내었다.

3.2. 사용 엔진 오일들의 분류

2.2 항에 총 24 개 차량에서 86 개의 엔진 오일 시료를 채취하였고 각 엔진 오일 당 3 개의 시료를 만들어 총 258 개의 시료로부터 데이터를 얻었다. 또한 시

료 채취에 사용하였던 채혈용 수액 세트에서 DEHP (diethylhexylphthalate)가 머무름 시간 17.7분대에서 용출되어서 이를 제거하기 위하여 머무름 시간 17.740~17.868까지의 데이터는 삭제하고 통계 처리하였다.

시료의 통계처리는 새 윤활유의 분류에서 좋은 결과를 보여준 웨이블릿 적합법으로 요약하고 같은 방식으로 분류를 진행하였다. 분류 기준은 주행 거리별 subclass, 차량별 class, 연료 타입별 class와 브랜드별 class로 모두 4 가지로 시행하였다.

하나의 TIC 데이터는 DEHP의 피크를 제외하고 총 6197 개의 변수가 관측되었다. 이를 웨이블릿 계수를 이용한 방법으로 자료를 요약하고 분류 모형을 만들기 위하여 데이터를 표준화시키고 자료의 개수를 2의 승수로 만들기 위해서 1995 개의 0을 추가하였다. 차원을 축소하여 데이터를 효과적으로 요약하기 위해 thresholding을 적용하여 575 개의 웨이블릿 계수를 선택하였다.

1) 주행 거리별 분류

분류 모형을 주행 거리별 subclass 기준으로하면 즉 동일 엔진 오일을 같은 차량에 넣고 주행하여도 주행

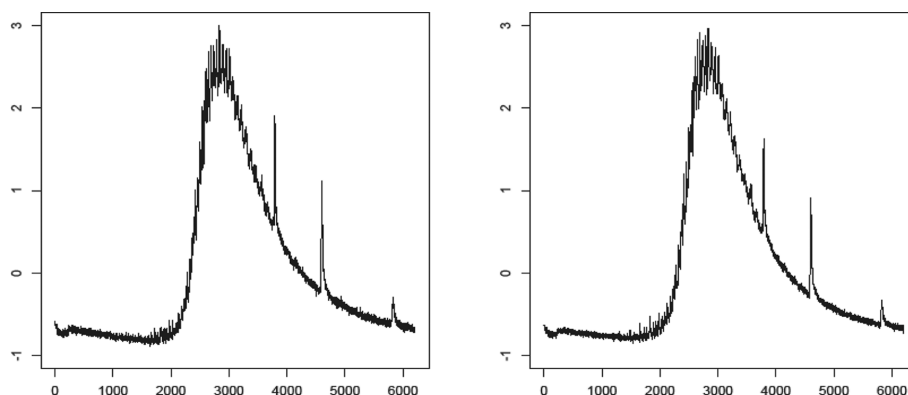


Fig. 11. Chromatograms with normalization of sample UD1-A[1] and UD1-E[1] (from left).

Table 9. The classification result of used engine oils by Bayesian model with wavelet fitting technique

Sample number	Symbol of vehicle	Brand name of engine oil	Mileage after replacing oil (km)	Symbol	Classification	Misclassification
1	UD1	Zic XQ	1,013	UD1-A	3	0
2			2,200	UD1-B	3	0
3			3,393	UD1-C	3	0
4			4,443	UD1-D	3	0
5			5,443	UD1-E	3	0
6	UD2	Kixx-RV	100	UD2-A	3	0
7			1,000	UD2-B	3	0
8			2,000	UD2-C	3	0
9			3,300	UD2-D	3	0
10			4,500	UD2-E	3	0
11			5,500	UD2-F	3	0
12	UD3	Kixx-RV	300	UD3-A	2	1
13			2,000	UD3-B	3	0
14			4,000	UD3-C	3	0
15			5,700	UD3-D	3	0
16	UD4	Kixx-RV	1,500	UD4-A	3	0
17			2,500	UD4-B	3	0
18			5,000	UD4-C	3	0
19	UD5	Hyundai D	1,500	UD5-A	3	0
20			2,500	UD5-B	2	1
21			4,000	UD5-C	3	0
22			5,500	UD5-D	3	0
23	UD6	Zic XQ	600	UD6-A	3	0
24			900	UD6-B	3	0
25	UD7	Zic XQ	90	UD7-A	3	0
26	UD8	Zic XQ	1,000	UD8-A	3	0
27	UG1	Zic A	1,500	UG1-A	1	2
28			3,000	UG1-B	3	0
29			6,000	UG1-C	3	0
30			7,500	UG1-D	2	1
31	UG2	Zic A	50	UG2-A	2	1
32			1,000	UG2-B	3	0
33			2,500	UG2-C	3	0
34			3,000	UG2-D	3	0
35			4,200	UG2-E	3	0

Table 9. Continued

Sample number	Symbol of vehicle	Brand name of engine oil	Mileage after replacing oil (km)	Symbol	Classification	Misclassification
36	UG3	Zic A	50	UG3-A	3	0
37			1,000	UG3-B	3	0
38			2,000	UG3-C	3	0
39	UG4	Kixx G1	1,000	UG4-A	3	0
40			2,000	UG4-B	0	3
41			3,000	UG4-C	1	2
42	UG5	Kixx G1	200	UG5-A	3	0
43			1,000	UG5-B	2	1
44			3,000	UG5-C	1	2
45			4,000	UG5-D	1	2
46			6,000	UG5-E	2	1
47	UG6	Kixx G1	50	UG6-A	0	3
48			1,000	UG6-B	3	0
49			2,500	UG6-C	3	0
50			3,100	UG6-D	2	1
51	UG7	Hyundai G	600	UG7-A	1	2
52			1,000	UG7-B	2	1
53			2,000	UG7-C	3	0
54			4,000	UG7-D	3	0
55			6,300	UG7-E	2	1
56	UG8	Hyundai G	1,000	UG8-A	3	0
57			3,000	UG8-B	3	0
58	UG9	Hyundai G	200	UG9-A	3	0
59			1,000	UG9-B	3	0
60			3,000	UG9-C	3	0
61			4,500	UG9-D	3	0
62			7,500	UG9-E	3	0
63	UG19	Hyundai G	1,500	UG16-A	3	0
64			2,800	UG16-B	3	0
65			3,800	UG16-C	3	0
66			6,500	UG16-D	3	0
67	UG10	Kixx G1	1,000	UG10-A	2	1
68			2,300	UG10-B	2	1
69			4,200	UG10-C	3	0
70			5,800	UG10-D	3	0
71	UG11	Zic A	1,300	UG11-A	3	0
72			2,000	UG11-B	3	0
73			3,100	UG11-C	2	1
74			4,500	UG11-D	3	0
75	UG12	Hyundai G	1,000	UG12-A	2	1
76			2,000	UG12-B	3	0
77			4,000	UG12-C	3	0
78			4,700	UG12-D	3	0
79	UG13	Zic A	1,200	UG13-A	0	3
80			3,900	UG13-B	3	0
81	UG14	Hyundai G	800	UG14-A	2	1
82			2,000	UG14-B	3	0
83			3,000	UG14-C	3	0
84			3,800	UG14-D	2	1
85			UG15	Zic A	1,300	UG15-A
86	2,200	UG15-B			3	0
Total					258	35
Rate of misclassification					13.6%	

거리에 따라 서로 다른 집단(subclass)으로 간주하는 분류를 진행하였다. 그러나 이런 경우 절반이 넘는 15 개의 시료가 오분류되었다. 이는 같은 차량일 경우 주행거리에 따라 TIC 데이터에 큰 차이가 없기 때문으로 생각되었다. 그 예로 Fig. 11을 보면 첫 번째 차량의 주행거리가 1,013 km 일 때의 첫 번째 시료의 (UD1-A[1]) TIC와 주행거리가 5,443 km 일 때의 첫 번째 시료의(UD1-E[1]) TIC가 서로 매우 유사하다. 이는 차량이 5,000 km 까지 주행하여도 엔진 오일의 성분변화가 크지 않기 때문인 것으로 사료된다.

2) 차량별 분류

다음으로 분류모형을 차량별 class로 하면 즉 주행 거리에 상관없이 동일 차량의 엔진 오일class에 맞추면 올바르게 분류한 것으로 간주하는 즉 첫 번째 차량(UD1)의 주행거리 1,000~5,000 km까지에서 매 1,000 km 마다 채취한 시료의 TIC 데이터(총 15 개)가 모두 UD1 Class에 속하는 것으로 보는 차량별 분류(총 24 classes)를 실시하였다. 실제 사건 해결에서도 주행거리에 따라 구분하기 보다는 차량을 정확히 맞추는 것이 중요하기 때문에 이 기준을 선택하였다. 그 결과, 전체 258 개 시료 중 35 개를 오분류하였다. 즉 오분류율은 13.6%로, 86.4%의 분류 일치율을 보였고 분류 결과는 Table 9에 나타내었다.

오분류된 결과를 보면 잘못 분류하는 오분류 확률이 거의 1에 가깝게 나타났는데 그 이유는 동일한 엔진 오일을 각각 다른 차량에 넣고 약 5,000 km까지 주행하더라도 차량 또는 운행 조건 등에 따라서 엔진 오일의 유기물 성분변화가 크지 않기 때문으로 사료되는 데 그 예로 Fig. 12을 보면 모두 Zic A인 엔진 오일을 각기 다른 차량에 넣고 주행거리 1,000 km에서 채취한 두 시료의 크로마토그램이 서로 매우 유사하게 나타났음을 알 수 있다. 이 경우는 UG1-A의 3 개 시료 중 2 개 시료가 오분류되고, UG13-A의 3 개 시료는 모두 오분류되는 결과를 보였는데(Table 9), 특히 주행거리가 짧아 엔진 오일의 유기물 성분 변화는 크지 않을 것으로 쉽게 예상 가능하여 서로 오분류될 확률이 높을 것으로 사료되지만 주행거리를 더 늘려 약 5,000 km 정도까지 주행하더라도 유기물의 성분변화는 크지 않아 오분류율이 높은 것으로 사료된다. 그래서 사용 엔진 오일의 GC/MS 분석만으로 차량을 추적하기는 다소 무리가 있음을 알 수 있었다.

3) 차량 연료 타입별 분류

다음으로 분류 모형을 차량 연료타입 즉 휘발유 차량(UG)인지 경유 차량(UD)인지를 맞추는 연료 타입별 분류(총 2 classes)는 99.6%의 일치율을 보여 휘발유차량 엔진오일인지 경유차량 엔진오일의 구별은 좋은 결과를 보여주었다. 이는 엔진 오일 시료가 휘발유 차량과 경유 차량의 각각 3 종씩으로 엔진 오일의 종류가 작아 표본 집단이 작아서 실제 사건에 적용하기 위해서는 더 많은 데이터 보강이 필요한 것으로 사료되지만 우리나라에서 주로 많이 사용하는 것들이므로 의미있는 결과로 사료된다.

4) 엔진 오일 브랜드별 분류

마지막으로 분류모형을 엔진 오일 브랜드명을 기준으로 하는 사용 엔진 오일 브랜드명별 분류(총 6 classes)는 97.3%의 일치율을 보였다. 그리고 사용 엔진 오일의 분류에는 첨가제 metal들과 주행거리에 따라 차량에 의한 wear metal들의 분석 데이터를 함께 고려해 주면 훨씬 효과적으로 분류할 수 있을 것으로 사료된다.¹⁹ 이 역시 표본이 작기는 하지만 의미있는 결과이며 더 많은 엔진 오일로 데이터 축적이 필요한 것으로 사료된다.

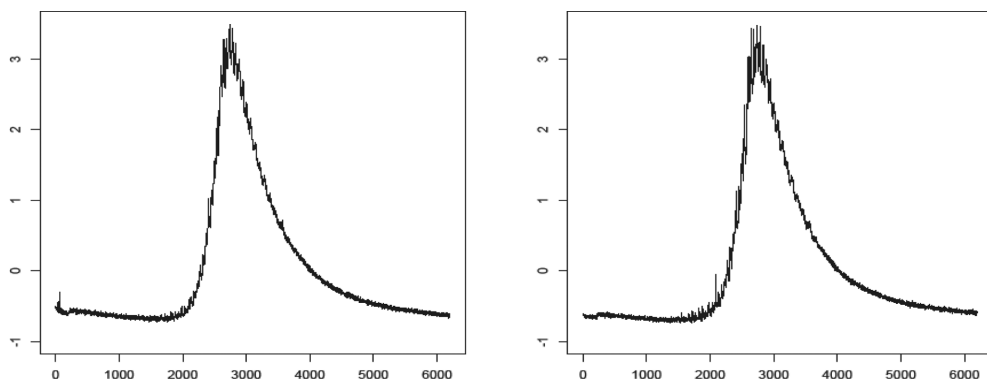


Fig. 12. Chromatograms with normalization of sample UG1-A[1] and UG13-A[1] (from left).

4. 결 론

본 연구에서는 윤활유의 법과학적 증거 가치를 높이기 위하여 윤활유 시료를 GC/MS로 분석하여 얻은 TIC 데이터에 베이지안 방식의 분류 모형을 적용하여 윤활유의 동일성 및 분류를 시도하였다. 이 때 TIC 데이터의 차원을 줄여서 간단하고 효율적인 분류 규칙을 만들고자 웨이블렛 계수를 이용한 요약과 주성분 분석을 이용한 요약 등 차원축소의 기법을 적용해 보았다.

새 윤활유의 경우 GC/MS 데이터로 웨이블렛 계수와 주성분 분석으로 차원 축소하여 베이지안 방식의 분류 모형을 적용하여 분류를 진행한 결과 정분류율 각각 97.5%와 96.7%로 모두 우수하였으나 웨이블렛 계수로 차원 축소하는 방법이 약간 더 우수하여 윤활유 분류 방법 중 차원 축소법으로 선택하였다.

사용 엔진 오일의 분류는 웨이블렛 계수로 차원을 축소하는 방법으로 진행하였다. 차량별 분류의 정분류율은 86.4%로 정확성이 떨어졌으며, 이는 동일한 엔진 오일을 비록 다른 차량에 사용하더라도 5,000 km까지의 주행거리만으로는 성분 변화가 크지 않아 크로마토그램들이 매우 유사하여 생길 수 있는 결과로 사료되었다. 분류 모형을 차량 연료 타입으로 하면 99.6%의 일치율을 보여 휘발유차량 엔진오일인지 경유 차량 엔진오일의 구별은 좋은 결과를 보여주었고, 엔진 오일 브랜드별 분류에서도 97.3%의 우수한 분류 일치율을 보였다. 이로서 사건 현장에 유류된 사용 엔진 오일의 GC/MS 분석으로 차량 추적에는 무리가 있으나 차량의 종류(휘발유 차량인지 경유 차량인지)와 엔진 오일의 브랜드명 추적에 베이지안 분류 모형의 활용은 적합할 것으로 사료되었다.

감사의 글

본 연구는 국립과학수사연구원 과학수사 감정기법 연구개발 사업비에 의하여 수행되었으며 이에 감사드립니다.

참고문헌

- J. Zieba-Palus and P. Koscielniak, *J. of Mol. Struct.*, **482-483**, 533-538 (1999).
- N. Gracia, S. Thomas, P. Bazin, L. Duponchel, F. Thibault-Starzyk and O. Lerasle, *Catal Today*, **155**, 255-260 (2010).
- J. Zieba-Palus, P. Koscielniak and M. Lacki, *Forensic Sci. Int.*, **122**, 35-42 (2001)
- K.-S. Yoo, J.-Y. Jyoung, S.-Y. Jeoung and S.-B. Woo, *Anal. Sci. Technol.*, **16**(5), 339-348 (2003).
- H.-S. Lim, G.-C. Han and S.-G. Lee, *Anal. Sci. Technol.*, **18**(2), 168-172 (2005).
- J.-H. Choi, H. Jeong and H.-D. Kim, *Anal. Sci. Technol.*, **10**(2), 139-145 (1997).
- M. A. Al-Ghouti and L. Al-Atoum, *J. Environ. Manage.*, **90**, 187-195 (2009).
- V. G. van Mispelaar, A. K. Smilde, O. E. de Noord, J. Blomberg and P. J. Schoenmakers, *J Chromatogr A.*, **1096**, 156-164 (2005).
- R. Hibbard, J. V. Goodpaster and M. R. Evans, *J. Forensic Sci.*, **56**(3), 741-753 (2011).
- T. Mang and W. Dresel, "Lubricants and Lubrication". 2nd Ed., Wiley-VCH, New-York, 2007.
- 최주환, "합성 윤활유 윤활유 첨가제", 푸른길, 2000.
- L. R. Rudnick, "Lubricant Additives Chemistry and Applications", 2nd Ed., CRC Press, Delaware, 2009.
- 대한석유회, 석유회보, **7**, 33-52, 1226-1254 (1982).
- Y. Qu, B.-L. Adam, M. Thornquist, J. D. Potter, M. L. Thompson, Y. Yasui, J. Davis, P. F. Schellhammer, L. Cazares, M. Clements, Jr. Wright and Z. Feng. *Biometrics*, **59**(1), 143-151 (2003).
- A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin. "Bayesian Data Analysis", Chapman and Hall/CRC, 2003.
- S. Mallat, "A Wavelet Tour of Signal Processing", 3rd Ed., The Sparse Way, Academic Press, 2008.
- B. Vidakovic, "Statistical Modeling by Wavelets", Wiley Series in Probability and Statistics, Wiley, 2009.
- I. T. Jolliffe, "Principal Component Analysis", 2nd Ed., Springer, 2002.
- Y. N. Kim, N. Y. Kim, S. Y. Park, D. K. Lee and J. H. Lee, *Forensic Sci. Int.*, **230**, 58-67 (2013).