

Markov Chain을 이용한 국내 폐차발생량 예측

이은아 · 최희련 · 이홍철[†]

고려대학교 산업경영공학부

A Study on the Forecasting of the Number of End of Life Vehicles in Korea using Markov Chain

Euna Lee · Hoeryeon Choi · Hongchul Lee

Department of Industrial and Management Engineering, Korea University

As the number of end-of-life vehicles (ELVs) has kept increasing, the management of ELV has also become one of the academic research focuses and European Union recently adopted the directive on ELVs. For the stakeholders has become a principle agent of dealing with all about ELVs, it is relevant investment decision to set up and to decide high-cost ELVs entity locations and to forecast future ELVs' amount in advance. In this paper, transition probability matrixes between months are made by using Markov Chain and the number of ELVs is predicted with them. This study will perform a great role as a fundamental material in Korea where just started having interests about recycling resources and studies related to the topic. Moreover, the forecasting method developed for this research can be adopted for other enhancements in different but comparable situations.

Keywords: ELVs(End of Life Vehicles), Markov Chain, Forecasting, Transition Probability Matrix, Time Series Analysis

1. 서론

1.1 연구의 배경 및 목적

산업화와 정보화로 인한 빠른 생활 패턴에 따라 자동차의 수요는 지속적으로 증가하여 <Figure 1>과 같이, 2012년 1월 현재 우리나라의 자동차 등록대수는 약 1,844만대를 기록하게 되었다. 또한, 자동차 보유대수의 증가에 비례하여 폐차발생대수도 증가하여 왔으며, <Figure 2>에서 볼 수 있듯이 전국에서 연간 약 85만대의 폐차가 발생하고 있다

이러한 흐름에, 이미 선진국에서는 폐차를 하나의 자원으로 인식하고 자동차 재활용 기술개발 및 폐차처리 관리체계에 대한 연구가 활발히 진행되어 왔으며, 자동차 선진국인 EU를 중심으로 폐차 재활용 법규도 확대되었다. <Table 1>과 같이 2006년까지는 자동차 총 중량의 85%, 2015년까지는 95%를 재사용

및 회수해야 한다는 기준을 정해 놓고, 자국의 자동차뿐만 아니라 수입차까지도 기준을 엄격히 적용하고자 하고 있다(Hong, 2004).

한편, 국내의 자동차 제조사에서도 이런 세계적인 동향에 무관할 수 없으며 이들 EU국가들에 대한 수출의 필요성에 따라, 재활용률을 높이기 위한 방안으로 전기·전자제품 및 자동차의 자원순환에 관한 법률안이 공포되었으며 최근에는 관련 연구도 활발히 진행되고 있다(Byeon, 2010; Kim *et al.*, 2007).

대표적인 국내 연구로는, 과학기술부의 자원재활용 기술개발 사업단이 시행하는 연구과제가 한국과학기술연구원(KIST)과 (주)이엔텍을 중심으로 폐자동차 해체시스템을 개발하고 그 시범공장을 2000년도부터 시작하여 경북 포항에 설립하였으며, 산업자원부의 청정기술 기반사업으로 자동차 부품연구원 주관하에서 과학기술부 과제와 동일한 내용으로 진행된 바 있다.

이 논문은 환경부 글로벌탑 환경기술개발사업의 '적용형 해체 plant 설계·운영지원기술 개발' 연구과제로 수행되었음.

[†] 연락처 : 이홍철 교수, 136-713 서울시 성북구 안암로 145 고려대학교 산업경영공학부, Tel : 02-3290-3389, Fax : 02-929-5888

E-mail : hlee@korea.ac.kr

2012년 3월 6일 접수; 2012년 4월 12일 수정본 접수; 2012년 4월 16일 게재 확정.

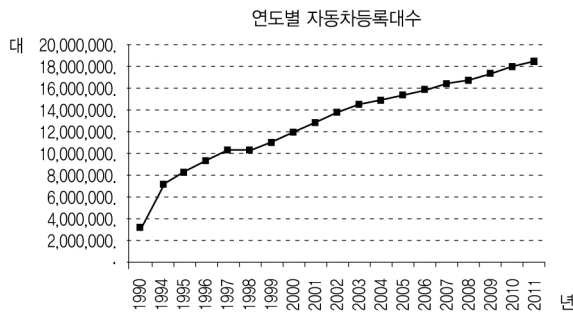


Figure 1. 연도별 자동차등록대수 추이

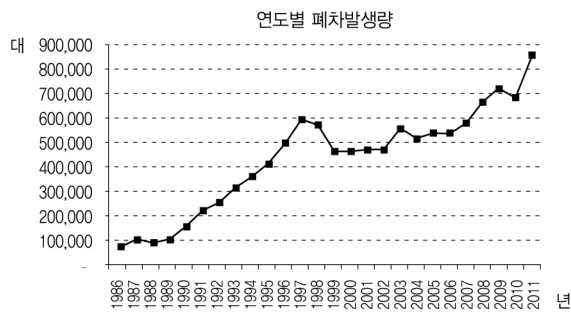


Figure 2. 연도별 전국의 폐차발생량 추이

또한, 2011년에는 환경부 산하의 글로벌탑 사업단이 출범되어 현재 자동차산업의 자원순환기술에 관한 몇 가지 연구들을 진행하고 있다. 한국과학기술연구원(KIST)을 주관으로 영남대학교에서는 자동차 해체 및 재활용 과정의 부품/자원/폐기물을 추적 및 모니터링하고 있으며, 실질적인 재활용률에 대한 연구 또한 함께 이루어지고 있다. 이와 더불어, 고려대학교에서는 글로벌 수요자별 환경에 최적화된 적응형 해체플랜트 시뮬레이터 개발 및 가상화 연구를 수행하고 있다. 이 외에도 현대자동차, 오토인프라, ARC 등의 기업이 참여하여 해체플랜트의 효율적 운영을 지원하기 위한 전산정보 시스템 개발 및 폐자동차 네트워크 개발에 관한 연구가 진행 중이다.

폴란드의 폐차관련시설 입지선정을 다룬 한 연구에서는 폐차 네트워크에서 시설비와 수송비가 차지하는 비중이 93%대 7%로 시설운영비가 월등히 많이 드는 것으로 나타났다(A Merkisz-Guranowska, 2011). 폐차처리 주체로 떠오른 자동차 제조업체로서는, 이러한 고가의 폐차관련시설의 설립을 계획하고 운영하는 것이 가장 중요한 투자결정이라 할 수 있기에 폐차발생량을 예측하는 일은 관련 정책을 수립하기 위한 중요한 전제가 된다. 따라서, 본 연구에서는 국내 폐자동차 발생량의 특징을 잘 반영하는 예측방법론을 제시하여, 그동안 활발하지 못했으나 이제 막 자원 재활용에 관심을 가지고 시작단계에 있는 우리나라의 폐자동차 관련 연구에 초석이 되고자 한다.

1.2 논문의 구성

본 연구는 모두 5장으로 구성된다. 제 1장 서론에 이어 제 2장에서는 기존연구에 대해 알아보고, 제 3장에서는 연구방법론

Table 1. EU 폐차 리사이클 법규

구분	규제항목	규제시기
폐차	무상회수 시스템 구축	
	• 02.7.1 이후 판매차량 • 02.7.1 이전 판매차량	02.7.1 02.7.1
재활용	친환경 처리 시설 구축	
	재사용/재활용/회수 • 재사용/재활용 : 80%, 재사용/회수 : 85% • 재사용/재활용 : 85%, 재사용/회수 : 95%	06.1.1 15.1.1

을 정리하였다. 제 4장은 본 연구의 실증분석에 해당되는 부분이다. 표본기간동안 마코브체인을 이용한 폐차발생량 예측모형을 만들어 마코브체인을 활용하지 않은 네 가지 시계열 예측모형과 비교하였으며, 실험의 편향성을 방지하고 객관성 확보를 위해 몇 가지 사항을 추가로 고려하여 추가실험을 구성하였다. 마지막으로, 제시된 모형으로 전국 16개 지역 향후 3년(36개월)의 폐차발생량을 예측하였으며, 제 5장에서 결론으로 마감하였다.

2. 기존연구

폐자동차의 발생량을 예측하고자 하는 본 논문과 관련한 문헌 범주로는 크게 폐차관련연구와 수요예측 방법론에 관한 연구로 나눌 수 있다.

우선적으로, 폐차관련 관한 연구로서 A Merkisz-Guranowska (2011)은 폴란드의 사례를 들어 폐차 네트워크를 정의한 후 기존의 폐차관련 시설의 수와 위치를 최적화하여 비용을 절감하는 방안을 제시하였으며, Reynaldo *et al.*(2008) 등은 멕시코의 폐자동차 중간집하장 최적입지선정에 관한 연구를 발표하였다. 하지만, 이러한 연구들에서 폐차 네트워크를 다룰 때, 입력값으로 전제된 폐자동차 발생량 예측치에 대한 부분은 몇 가지 시나리오의 형태로 제시해둔 채 의사결정자의 몫으로 남겨두었다. 또한, 이러한 각 시나리오에서 각 경우의 예측치는 전국의 폐차발생량에 대한 것이었으며, 실증연구에 필요한 발생지별 폐차발생량은 어떻게 정의되었는지에 대한 구체적 설명이 제시되어 있지 않았다.

수요예측에 관한 연구로, 최근에 국가별 GDP와 차량수명주기를 이용해 EU25국의 폐차발생량을 예측한 연구가 Anderson *et al.*(2007) 등에 의해 발표되었다. 하지만, 예상(projection)이라는 단어를 사용한 것처럼 거시적관점에서 EU25개국의 폐차발생경향을 개괄적으로 파악하는 데 그쳤다.

한편, 국내의 폐차 관련 연구로서 Byeon(2008)와 Kang(2005)는 각각 울산지역의 폐자동차 재활용 과정에서의 배출오염물에 대한 연구와 국내 폐차산업의 현황분석 및 재활용 개선방안에 관한 연구에서 폐차 발생량에 대한 현황분석 결과를 주관적 견해를 바탕으로 제시하였다. 하지만, 폐차발생량에 대한 계량적 접근이 이루어지지 않아 객관성을 확보하기 어렵다는 단점

Table 2. 폐차발생량의 저하요인

자동차 기술적 요인	외부 환경적 요인
<ul style="list-style-type: none"> • 자동차 제조기술의 향상 • 자동차 제조사의 A/S 및 정비 기술의 발달 • 부품의 원활한 공급 • 부품 가격인하 	<ul style="list-style-type: none"> • 도로포장율의 향상 • 자동차 보험제도의 완비 • 사용년수에 따른 자동차세의 감면 • 자동차 소유자의 의식변화 • 경기침체로 인한 대체수요 감소 • 가구당 자동차 보유대수 증가 • 주행거리의 감소

이 있다. 이 외의 폐차 발생량 예측법에 관한 국내 연구는 전무한 실정이다.

Baek *et al.*(2011)은 선형회귀분석으로 열 수요를 예측하였고, Yim(2007)이 지수평활법과 ARIMA 모형으로 레스토랑 매출액을 예측을 하였으며, Yoon *et al.*(2010) 등은 시계열 모형을 이용하여 항공수요를 예측하는 등 다양한 형태의 데이터에 적합한 예측모형을 개발하기 위한 많은 연구가 진행되어 왔다. 하지만, 이러한 방법론들은 폐물류만의 특징인 불확실성(uncertainty)이 고려되지 않아 폐차발생량 예측에 적용하기에는 한계가 있다.

폐기물의 배출량예측에 관한 국내연구로서 Bae *et al.*(2006) 등은 음식물폐기물의 수요를, Song(1996), Min(2007)와 Kang *et al.*(2007) 등은 각각 폐플라스틱, 가전제품과 폐휴대폰의 배출량을, 그리고 Park *et al.*(2008) 등은 폐목재 발생량을 예측한 연구가 있다. 이러한 연구들은 폐기물에 영향을 주는, 예를 들어 폐목재의 발생량을 예측할 경우에 인구수, 산림면적, 벌채량, 1인당 GDP 등과 같은 원인변수들과의 상관관계를 이용하는 다중회귀분석법을 사용하여 예측하였다. 하지만, 이를 폐차발생량 예측에 적용하기 위해서는 <Table 2>와 같은 폐차발생량과 관련된 원인변수의 파악과 그 상관관계 정도에 대한 선행 연구가 필요하다(Kang, 2005). 하지만, 현재 우리나라는 폐자동차에 관한 연구가 이제 막 시작단계에 있으며, 기존연구가 미비한 상태로 다변량수요예측법을 사용하기에는 곤란한 부분이 있다. 위에서 언급한 이러한 연구들을 정리하면 다음과 같다.

- ① 구체적인 폐차 발생량 예측치의 미제시
- ② 계량적 접근이 이루어지지 않아 객관성 부족
- ③ 국내실정에 맞는 폐차관련 연구의 미비
- ④ 폐물류발생량의 특징인 불확실성(uncertainty)이 미반영
- ⑤ 원인변수에 크게 의존하는 예측법들

따라서, 본 연구에서는 이러한 한계를 극복하기 위하여 <Table 3>과 같은 부분에 중점을 두고 국내 폐차 발생량의 특성을 잘 반영한 예측법을 마련함으로써, 폐차관련 정책수립 및 관련 연구에 도움을 주고자 한다.

Table 3. 본 연구의 접근방향

현 연구의 과제	해결방안
계량적 접근으로 객관성 확보	구체적 예측모형제시
발생지별 예측치의 구체적 제시	전국대비 지역별 폐차발생분포 상태의 비교적 안정된 경향을
불확실성(uncertainty)을 고려한 수요예측법 제시	이용, 마코브체인 전이행렬표를 구성
원인변수를 고려하지 않는 예측법 적용	일변량 시계열 모형에 기초한 분석

3. Markov Chain의 개념

마코브 과정(Markovian Process)은 반복되는 상황에 대해 특정 상태(state)에서 다음의 상태로 어떻게 변화할 것인가를 확률적으로 예측하는 기법으로서 어떤 사건이나 현상이 바로 이전 사건이나 현상에 의해서만 결정되는 확률적 과정(Stochastic Process)에 관한 이론을 최초로 전개한 러시아의 수학자 Andrei Andreyevich Markov(1856년~1922년)의 이름을 따라 붙여졌다(Kim, 2005).

마코브 과정을 전이확률과 함께 마코브 체인이라고 하며 이를 수학적으로 정리하면 다음과 같다. 즉, t번째 기간에 i라는 상태에 있는 것을 $X_t = i$ 라 표기할 때, 다음의 관계를 만족하는 경우 마코브 체인이라 한다(Shelon, 2000).

$$P\{X_{t+1} = j | X_t = i, X_{t-1} = i-1, \dots, X_0 = i_0\} = P\{X_{t+1} = j | X_t = i\} = P_{ij}$$

3.1 전이확률행렬표

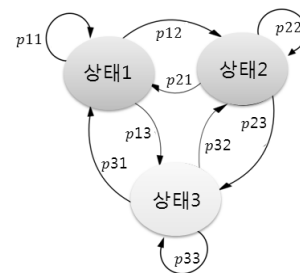


Figure 3. 마코브체인 전이확률 다이어그램

전이확률(Transition Probability)은 마코브 체인을 이해하기 위한 중요한 요소로서 어떤 반복되는 사건이나 현상이 현재 상태에서 다른 상태로 바뀌는 확률을 말하며, 이 확률을 행렬의 형태로 나타낸 것을 전이확률행렬표(Transition Probability Matrix)라고 한다.

전국을 16개 지역인 강원, 경기, 경남, 경북, 광주, 대구, 대전, 부산, 서울, 울산, 인천, 전남, 전북, 제주, 충남, 충북으로 나누어,

전국대비 이들 지역의 월간 폐차동차 발생분포상태를 적당한 수로 나누는 것이 가능하다. 예를 들어, 16개 지역의 월별 폐차 발생량 분포를 세 개의 상태로 나누었다고 가정하면, <Figure 3>과 같이 ‘상태 1’, ‘상태 2’, ‘상태 3’으로 표현이 가능하고 n월의 폐차발생패턴에 대하여 상태 i에 있다가 n+1월에서 상태 j에 있게 될 전이확률을 p_{ij} 로 정의한다면,

전국대비 16개 지역의 폐차발생분포 상태에 대한 전이확률행렬표는 다음과 같이 상태의 개수에 따라 3×3행렬로 나타낼 수 있다.

$$P = [P_{ij}] = \begin{pmatrix} p_{11} & \dots & p_{1m} \\ & p_{ij} & \\ p_{m1} & \dots & p_{mm} \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

$P = [P_{ij}]$: 전이확률행렬표

p_{11}, \dots, p_{mm} : 상태 i에서 상태 j로 바뀌는 전이확률($0 \leq p_{ij} \leq 1$)

m : 상태의 수

예를 들어, 위의 3×3행렬에서 p_{13} 은 이번 달에 상태 1와 같은 폐차발생분포를 보였을 때, 다음 달에 상태 3과 같은 분포를 나타내게 될 확률을 나타내며, p_{22} 는 이번달 폐차발생분포가 상태 2와 같을 때 다음 달의 분포가 다시 상태 2와 같은 형태를 보일 확률을 나타낸다.

3.2 Chapman-Kolmogorov 방정식

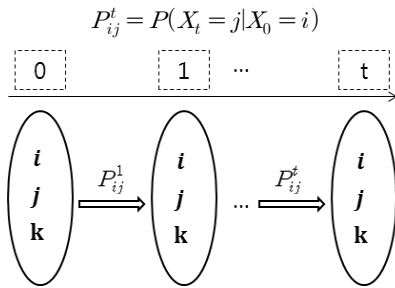


Figure 4. Chapman-Kolmogorov 방정식

마코브 체인 $\{X_t, t \geq 0\}$ 에서 x_0 의 분포를 초기분포라 하고 $p_e = p(X_0 = e)$ 로 할 때, x_0 가 나타나는 상태 e를 초기상태라 부른다. 또 상태 i에서 출발한 마코브 체인이 t번째에서 상태 j에 도달할 확률을 i에서 j로의 t-단계 전이확률이라고 부르며 P_{ij}^t 로 표기한다.

임의의 정수 l과 m($l, m \geq 0$)에 대하여 $\sum_{k=0}^{\infty} P_{ik}^l \cdot P_{kj}^m$ ($l \geq 0, m \geq 0$)이 성립한다. 이것이 Chapman-Kolmogorov 방정식이며, 이를 이용하여 t번째 전이 확률을 구할 수 있다. $P^{(t)}$ 을 t-단계 전이확률 $P_{ij}^{(t)}$ 의 행렬이라고 하면 다음과 같은 식이 성립한다.

$$P^{(l+m)} = P^{(l)} \cdot P^{(m)}$$

예를 들어, 2-단계에 대한 전이확률행렬은 다음과 같다.

$$P^{(2)} = P^{(1+1)} = P^{(1)} \cdot P^{(1)} = P^2$$

따라서, t-단계 전이확률행렬은 1-단계 전이확률행렬 P를 t-제곱 한 것과 같다.

4. 실증연구

시계열 모형에 의한 수요예측은 예측될 변수 자체의 과거의 자료에서 어떠한 패턴을 발견하여 미래에도 그러한 패턴이 특성을 잃지 않고 반복될 것이라는 가정 하에, 모형을 확립하여 미래 수요를 예측하는 방법이다. 그러므로 확립된 시계열 모형은 특정한 자료의 집합에 모형이 얼마나 잘 적합한가에 따라 전적으로 타당성이 결정된다.

본 연구에서는 다중회귀분석 등 인과모형에 의한 수요예측과 달리 일변량 시계열 모형을 적용하여 미래 수요를 예측하고, 실제 폐차동차 발생량 자료와의 비교를 통해 예측모형의 타당성을 확인하고자 하였다. 인과모형에 의한 예측방법은 원인이 되는 변수에 대한 지표를 확보해야 하는 어려움이 있기 때문에 그동안 폐차량예측에 관한 국내 연구가 전무했던 점을 감안한다면 이 방법에 있어서의 예측결과의 타당성을 담보하기에는 어려움이 있다(Huh, 2010). 예를 들어, 자동차등록수가 폐차발생량에 높은 영향을 준다는 것은 자명한 사실이지만, 미래의 자동차등록수에 대한 예측이 선행되어야 하고, 이에 대한 타당성이 입증되어야 하는 문제를 안고 있다.

시계열 모형의 수요예측에 사용될 수 있는 방법론에는 회귀분석, 지수평활, ARMA(Autoregression Moving Average), 인공신경망기법 등이 있으며, 현재 폐차량만의 수요예측 모형에 대한 연구는 국내·외 모두 아직 미비한 실정이다. 본 연구에서는

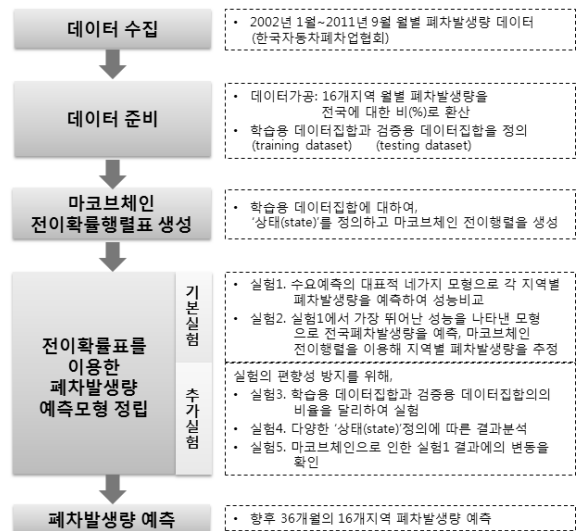


Figure 5. 연구의 흐름도

국내의 폐차발생량 예측을 위해 앞에서 언급한 네 가지 일변량 시계열 모형의 성능을 비교하였으며, 이 중 가장 성능이 좋은 모형을 더 개선하기 위해 마코브체인을 이용한 새로운 모형을 제안한다. 새로 제시되는 모형은 폐자동차 발생량의 월별 발생 경향은 비록 불규칙적이지만, 전국대비 지역별 발생분포형태는 비교적 일정함을 이용하여 마코브체인 전이행렬을 구성하고, 전국 16개 지역 폐차발생량을 예측하였다. <Figure 5>는 본 연구에서 수행된 연구의 흐름도를 나타낸다.

4.1 데이터 수집

본 연구를 수행하기 위한 데이터의 시간적 범위는 2002년 1월부터 2011년 9월까지 약 10년(117개월)간으로 월별 시계열 데이터를 이용하여 폐자동차 발생량 예측분석을 실시하였다. 분석에 사용된 자료의 원천은 한국자동차폐차협회에서 제공하는 폐차통계 중 승용차 통계를 이용하였다.

4.2 데이터 준비

마코브 체인을 이용하여 전국 16개 지역의 월별 폐차발생량을 예측하기 위해서는 먼저 전이확률행렬표 생성이 필요하며, 이 전이확률생성표를 생성하기 위해서는 다양한 값을 가지는 전국대비 지역별 폐차발생비율에 대한 정의가 필요하다(Lee et al., 2009). 본 연구에서는 먼저 16개 각 지역의 월별 폐차발생량을 전국에 대한 백분율로 산출하였다.

$$X_i^r = \frac{X_i}{K}$$

X_i^r : 전국에 대한 지역 i 의 폐차발생비율(%)

X_i : 지역 i 의 폐차발생량(대)

K : 전국의 폐차발생량(대)

i = 강원, 경기, 경남, 경북, 광주, 대구, 대전, 부산, 서울, 울산, 인천, 전남, 전북, 제주, 충남, 충북

예를 들어, 특정달의 강원지역 폐차발생량을 20대, 전국의 폐차발생량을 100대라 한다면 전국대비 강원지역의 폐차발생비율 $X_{강원}^r$ 은 $20/100 = 0.2(20\%)$ 와 같이 구할 수 있다.

이러한 방법으로 가공한 데이터에 대하여 2002년 1월부터 2008년 12월(84개월)까지의 기간에 대해 모형을 적합해 보고, 2009년 1월부터 2011년 9월(33개월)까지의 기간에 대해 예측을 해보도록 하겠다.

4.3 전이확률행렬표 생성

앞에서 나눈 표본기간 84개월 동안의 데이터를 바탕으로 전이확률행렬표를 생성하기 위하여, 가장 먼저 아래의 조건을 만족하도록 자료를 클러스터링하고 상태집합을 정의하였다.

조건 1: 모든 84개 관측치에 대하여, 각각을 포함하는 단 하나의 상태가 존재한다.

조건 2: 패턴집합의 원소인 각 상태에 대하여, 모든 지역의 구간길이가 4%를 넘지 않도록 한다.

클러스터링이란, 유사성 등의 개념에 기초하여 데이터를 몇 개의 그룹으로 분류하는 기법을 말하는데, 본 연구에서는 각 클러스터간 거리사이의 분산을 최소화시키는 k-means 클러스터링 기법이 X_i^r 의 최대 구간길이를 제한하는데 가장 적합하다고 생각하여 matlab R2009a으로 알고리즘을 구현하였다.

k-means기법을 이용하여 자료를 클러스터링한 결과 <Table 4>, <Figure 6>과 같은 ‘상태 1’, ‘상태 2’, ‘상태 3’, ‘상태 4’, ‘상태 5’로

Table 4. 표본기간 전국 16개 지역의 폐차발생분포에 대한 상태집합정의표

상태 1	$X_{강원}^r$	$X_{경기}^r$	$X_{경남}^r$	$X_{경북}^r$	$X_{광주}^r$	$X_{대구}^r$	$X_{대전}^r$	$X_{부산}^r$	$X_{서울}^r$	$X_{울산}^r$	$X_{인천}^r$	$X_{전남}^r$	$X_{전북}^r$	$X_{제주}^r$	$X_{충남}^r$	$X_{충북}^r$	
범위	max	5.2%	38.9%	8.7%	7.3%	2.8%	6.2%	3.4%	5.9%	0.0%	3.6%	7.5%	4.5%	4.5%	1.5%	5.3%	4.8%
	min	4.4%	36.4%	7.8%	6.1%	2.0%	4.2%	2.6%	3.8%	0.0%	2.0%	5.1%	3.8%	3.6%	0.9%	3.8%	3.5%
상태 2	$X_{강원}^r$	$X_{경기}^r$	$X_{경남}^r$	$X_{경북}^r$	$X_{광주}^r$	$X_{대구}^r$	$X_{대전}^r$	$X_{부산}^r$	$X_{서울}^r$	$X_{울산}^r$	$X_{인천}^r$	$X_{전남}^r$	$X_{전북}^r$	$X_{제주}^r$	$X_{충남}^r$	$X_{충북}^r$	
범위	max	7.2%	36.2%	8.9%	7.5%	2.9%	7.3%	3.7%	6.1%	0.0%	3.9%	7.8%	5.0%	5.2%	4.1%	6.3%	5.1%
	min	4.4%	34.2%	7.8%	6.4%	2.0%	4.0%	2.6%	3.9%	0.0%	2.1%	5.3%	3.7%	3.8%	1.0%	3.6%	1.3%
상태 3	$X_{강원}^r$	$X_{경기}^r$	$X_{경남}^r$	$X_{경북}^r$	$X_{광주}^r$	$X_{대구}^r$	$X_{대전}^r$	$X_{부산}^r$	$X_{서울}^r$	$X_{울산}^r$	$X_{인천}^r$	$X_{전남}^r$	$X_{전북}^r$	$X_{제주}^r$	$X_{충남}^r$	$X_{충북}^r$	
범위	max	11.1%	33.4%	8.5%	6.9%	1.9%	4.2%	2.6%	5.0%	0.0%	2.2%	6.7%	3.9%	3.8%	1.4%	4.8%	3.7%
	min	11.1%	33.4%	8.5%	6.9%	1.9%	4.2%	2.6%	5.0%	0.0%	2.2%	6.7%	3.9%	3.8%	1.4%	4.8%	3.7%
상태 4	$X_{강원}^r$	$X_{경기}^r$	$X_{경남}^r$	$X_{경북}^r$	$X_{광주}^r$	$X_{대구}^r$	$X_{대전}^r$	$X_{부산}^r$	$X_{서울}^r$	$X_{울산}^r$	$X_{인천}^r$	$X_{전남}^r$	$X_{전북}^r$	$X_{제주}^r$	$X_{충남}^r$	$X_{충북}^r$	
범위	max	5.3%	36.4%	12.1%	7.8%	2.7%	6.8%	3.7%	6.1%	0.0%	3.9%	7.7%	6.8%	5.0%	1.7%	6.0%	5.0%
	min	4.3%	33.8%	8.9%	4.0%	1.9%	4.1%	2.5%	3.1%	0.0%	2.3%	5.4%	3.7%	3.7%	1.1%	4.0%	3.3%
상태 5	$X_{강원}^r$	$X_{경기}^r$	$X_{경남}^r$	$X_{경북}^r$	$X_{광주}^r$	$X_{대구}^r$	$X_{대전}^r$	$X_{부산}^r$	$X_{서울}^r$	$X_{울산}^r$	$X_{인천}^r$	$X_{전남}^r$	$X_{전북}^r$	$X_{제주}^r$	$X_{충남}^r$	$X_{충북}^r$	
범위	max	4.2%	31.4%	17.0%	6.2%	1.7%	4.0%	2.6%	6.1%	0.0%	2.3%	6.2%	4.3%	4.5%	1.0%	4.8%	3.5%
	min	4.2%	31.4%	17.0%	6.2%	1.7%	4.0%	2.6%	6.1%	0.0%	2.3%	6.2%	4.3%	4.5%	1.0%	4.8%	3.5%

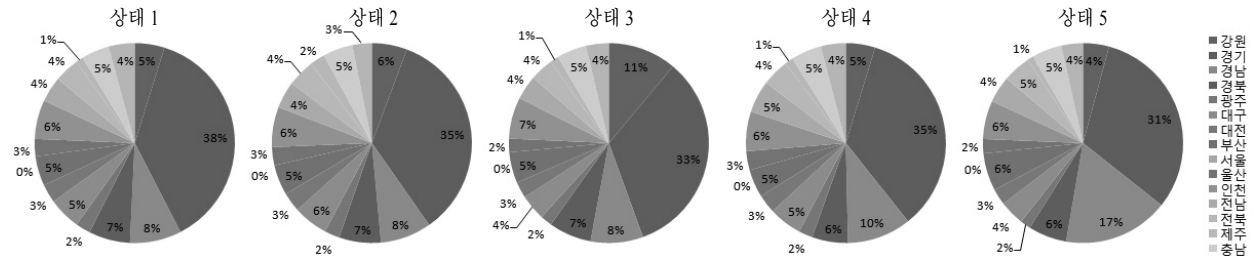


Figure 6. 표본기간 전국 16개 지역의 폐차발생분포에 대한 상태집합 도표

구성된 하나의 상태집합을 얻을 수 있었으며, 이는 위 두 가지 조건을 모두 만족한다. 즉, 강원 4.4~5.2%, 경기 36.4~38.9%, 경남 7.8~8.7%, 경북 6.1~7.3%, 광주 2.0~2.8%, 대구 4.2~6.2%, 대전 2.6~3.4%, 부산 3.8~5.9%, 서울 0.0%, 울산 2.0~3.6%, 인천 5.1~7.5%, 전남 3.8~4.5%, 전북 3.6~4.5%, 제주 0.9~1.5%, 충남 3.8~5.3%, 충북 3.5~4.8%의 구간에 속하는 모든 월 관측치들을 하나의 상태로 보고 그 상태를 ‘상태 1’이라는 이름으로 명명하겠다는 의미이다. 나머지 ‘상태 2’, ‘상태 3’, ‘상태 4’에 대해서도 같은 방식으로 생각할 수 있다.

한편, 이러한 조건에 부합하는 상태집합은 본 연구에서 제시한 것 외에도 다양한 기법에 의해 여러 가지로 제시될 수 있으나, 본 연구에서는 임의로 하나를 선정하여 사용하였고 상태집합에 따른 모형의 성능에 관해서는 추후 연구과제로 남겨두고 이번 연구에서는 다루지 않는다.

지금까지, 전이확률생성표를 생성하기 위한 상태집합을 정의하였다. 생성된 상태집합에 의한 각 상태별 실제데이터들의 분포는 <Table 5>, <Figure 7>에서 볼 수 있다. <Figure 7>을 살펴보면, 16개 지역의 월별 폐차발생량 분포는 주로 ‘상태 2’, ‘상태 4’와 같은 형태를 보이고 있음을 알 수 있다.

Table 5. 월별 폐차발생량분포표

구 분	월별데이터
상태 1	15
상태 2	35
상태 3	1
상태 4	32
상태 5	1
합계	84

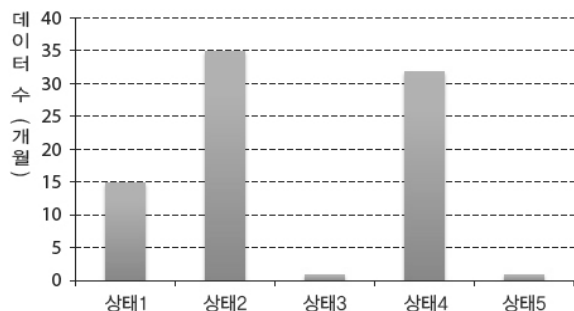


Figure 7. 월별 폐차발생량분포 그래프

5×5행렬에 대한 16개 지역의 폐차발생상태의 경우, 다섯 가지로 분류하였기 때문에 현재 상태에서 다음 상태로의 전이는 다섯 가지 경우인 ‘상태 1’, ‘상태 2’, ‘상태 3’, ‘상태 4’, ‘상태 5’이 될 수 있다. <Figure 8>은 월간 폐차발생상태의 전이에 대한 개념도를 나타낸다.

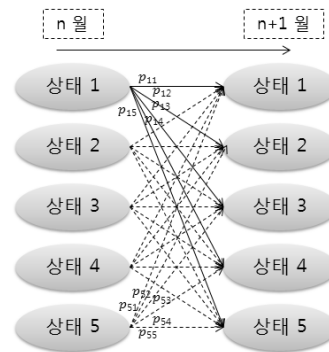


Figure 8. 월간 폐차발생량의 상태전이

만약, 2002년 1월부터 2008년 12월(84개월)까지의 월간 폐차발생량 데이터에 대하여 상태집합을 <Table 5>와 같이 정의한 경우, 상태 2에 속하는 자료는 총 35개가 존재한다. 그리고 상태 2에 속하는 35개의 데이터 중 그 다음 달에 상태 4의 형태를 보이는 데이터는 6개이고, 따라서 p_{24} 의 값은 $6/35 = 0.1714$ 로 정의할 수 있다. 이와 같은 방식을 이용하여 생성한 2002년 1월부터 2008년 12월(84개월)까지의 월간 폐차발생량 데이터에 대한 전이확률행렬표는 <Table 6>과 같다.

생성한 전이확률행렬표를 살펴보면 상태 1에서 그 다음달에 상태 3으로 전이될 확률과, 상태 3에서 상태 1로 전이될 확률 그리고 상태 3에서 상태 3으로 전이될 확률값 등 몇몇 전이 확률값들이 모두 ‘0’의 값을 나타냈다. 이는 본 연구의 분석대상 데이터에서는 상태 1에서 상태 3으로 전이될 확률과, 상태 3에서

Table 6. n월 → n+1월 전이확률행렬표

구 분	상태 1	상태 2	상태 3	상태 4	상태 5
상태 1	0.6667	0.2000	0	0.0667	0.0667
상태 2	0.1143	0.6857	0.0286	0.1714	0
상태 3	0	1	0	0	0
상태 4	0.0323	0.1935	0	0.7742	0
상태 5	0	0	0	1	0

상태 1로 전이될 확률, 또한 상태 3에서 상태 3의 형태로 전이될 확률 등이 발생하지 않는다는 것을 의미한다.

4.4 전이확률행렬표를 이용한 폐차발생량 예측모형 정립

4.4.1 기본실험

본 연구에서는 마코브 체인에서 n단계 전이 확률을 구하기 위하여 사용하는 Chapman-Kolmogorov 방정식을 사용하였으며, 이를 위해 월간 폐차발생량의 전이는 동질하다는 가정을 하였다.

$$P(1) = P(2) = P(3) = \dots = P(84)$$

따라서, Chapman-Kolmogorov 방정식을 이용하여 현재부터 n 번째 달의 전이확률을 구하면 $P(1) \times P(2) \times \dots \times P(n) = P(1)^n$ 과 같으며, 이를 이용하여 n 번째 달의 폐차발생량은 다음 식을 이용하여 예측할 수 있다.

$$F_n = P_0 \times \{P(1) \times P(2) \times \dots \times P(n)\} \times A_n \times T$$

$$= P_0 \times \{P(1)^n\} \times A_n \times T$$

- P_0 : 초기 폐차발생분포상태행렬
- $P(n)$: n 번째 달의 전이확률행렬
- A_n : n 번째 달의 전국 폐차발생량
- T : 상태행렬

즉, n 번째 달을 예측한 폐차발생량은 초기 폐차발생률 패턴 행렬(P_0), 전이확률행렬($P(n)$), n 번째 달의 전국 폐차발생량 (A_n), 그리고 상태행렬 (T)의 곱으로 나타낼 수 있다. 초기 폐차 발생분포상태행렬은 현재달의 폐차발생분포 상태를 의미하며, 5×5행렬의 경우 <Table 7>과 같이 표현할 수 있다.

전국의 폐차량 예측치(A_n)는 회귀분석법으로 추정하여 대입 하였다. 이 때, 회귀분석을 사용한 이유는, 본격적인 실증연구에 앞서서 국내 폐차발생량에 데이터에 대하여 선형회귀, 지수평활, 인공신경망, ARMA 네 가지 시계열 모형으로 전국 16개 지역의 폐차발생량을 예측해보고 그 결과를 비교해본 결과 회귀분석법이 가장 좋은 성능을 나타냈기 때문이다. <Table 8>은 앞에서 제시한 네 가지 시계열 모형의 성능을 비교·분석한 결과이며, 성능 비교를 위한 지표로는 절대편차(MAD : Mean Absolute Deviation)를 사용하였다. MAD는 ‘어떤 기대 값과 관찰된 값들의 편차 절대값들의 평균’이라고 정의되는데(APICS

Table 7. 초기 폐차발생률분포 상태행렬(P_0)

16개 지역의 폐차발생률 분포패턴	초기 폐차발생률 패턴행렬
패턴 1	(100 ... 0) _{1×m}
패턴 2	(010 ... 0) _{1×m}
⋮	⋮
패턴 m	(000 ... 1) _{1×m}

Table 8. 시계열 모형의 성능 비교

(단위 : 만 대)

구 분	회귀분석	ARMA	지수평활	인공신경망
16개 지역 MAD의 합	0.94	1.06	1.34	1.07

용어집), 편차는 수요 예측에 의해 발생된 오류 값들의 평균으로 오류의 경향은 고려하지 않고, 단지 오류의 평균값만을 고려하는 측도이다.

$$MAD = \sum | \text{실제 값} - \text{예측 값} | / \text{기간 수} \quad (1)$$

실험은 2002년 1월부터 2008년 12월(84개월)까지의 데이터에 대하여 16개 각 지역별 폐차발생량을 네 가지 모형으로 적합해 보고 2009년 1월부터 2011년 9월(33개월)까지의 기간에 대해 예측을 해 보는 형식으로 진행되었으며, <Table 9>의 수치는 16개 지역별로 도출한 MAD의 합을 나타낸다. MAD 통계량을 나타낸 식 (1)을 잘 살펴보면, 실제 폐차발생량과 모형으로 예측한 값이 근사할수록 통계량의 값은 더 작아지도록 계산이 된다. 그러므로 MAD 값이 작을수록 예측 능력이 더 좋은 모형이라고 생각할 수 있다.

상태행렬 (T)은 상태집합을 행렬로 표현한 것을 나타낸다. 상태행렬의 각 행과 열은 상태집합의 원소인 ‘상태 1’, ‘상태 2’, ‘상태 3’, ‘상태 4’, ‘상태 5’를 나타내는데, 이 때 범위로 정의된 상태 집합의 각 원소를 표현하기 위한 대푯값으로 ‘중앙값(median)’을 택하였다. 중앙값은 한 상태에 속하는 데이터들의 범위를 구간이라고 하였을 때 구간의 중앙의 값이라는 의미가 있다.

앞에서 설명한 방법을 이용하여, 현재부터 3개월 후 전국 16개 지역 폐차발생량은 식 (2)와 같이 구할 수 있다. 여기서의 현재 상태는 2008년 12월을 말하며, 이때의 전국 폐차발생량패턴은 ‘상태 4’와 같이 나타난다. 이 결과를 <Table 9>와 같이 정리해 보았다.

$$F_n = P_0 \times \{P(1) \times P(2) \times \dots \times P(n)\} \times A_n \times T$$

$$= (00010) \times \begin{pmatrix} 0.6667 & 0.2000 & 0 & 0.0667 & 0.0667 \\ 0.1143 & 0.6857 & 0.0286 & 0.1714 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0.0323 & 0.1935 & 0 & 0.7742 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}^3 \times 38642.5 \times \begin{pmatrix} 4.8\% & 37.6\% & 8.3\% & 6.7\% & 2.4\% & 5.2\% & 3.0\% & 4.8\% & 0.0\% & 2.8\% & 6.3\% & 4.1\% & 4.0\% & 1.2\% & 4.5\% & 4.2\% \\ 5.8\% & 35.2\% & 8.4\% & 6.9\% & 2.5\% & 5.7\% & 3.2\% & 5.0\% & 0.0\% & 3.0\% & 6.6\% & 4.4\% & 4.5\% & 2.5\% & 4.9\% & 3.2\% \\ 11.1\% & 33.4\% & 8.5\% & 6.9\% & 1.9\% & 4.2\% & 2.6\% & 5.0\% & 0.0\% & 2.2\% & 6.7\% & 3.9\% & 3.8\% & 1.4\% & 4.8\% & 3.7\% \\ 4.8\% & 35.1\% & 10.5\% & 5.9\% & 2.3\% & 5.4\% & 3.1\% & 4.6\% & 0.0\% & 3.1\% & 6.5\% & 5.3\% & 4.4\% & 1.4\% & 5.0\% & 4.1\% \\ 4.2\% & 31.4\% & 17.0\% & 6.2\% & 1.7\% & 4.0\% & 2.6\% & 6.1\% & 0.0\% & 2.3\% & 6.2\% & 4.3\% & 4.5\% & 1.0\% & 4.8\% & 3.5\% \end{pmatrix}$$

$$= (00010) \times \begin{pmatrix} 0.3518 & 0.3321 & 0.0081 & 0.2769 & 0.0313 \\ 0.1785 & 0.4820 & 0.0159 & 0.3130 & 0.0107 \\ 0.1601 & 0.5548 & 0.0196 & 0.2579 & 0.0076 \\ 0.0993 & 0.3402 & 0.0083 & 0.5476 & 0.0046 \\ 0.0687 & 0.2890 & 0.0055 & 0.6347 & 0.0022 \end{pmatrix} \times 38642.5 \times \begin{pmatrix} 0.048 & 0.376 & 0.083 & 0.067 & 0.024 & 0.052 & 0.030 & 0.048 & 0.000 & 0.028 & 0.063 & 0.041 & 0.040 & 0.012 & 0.045 & 0.420 \\ 0.058 & 0.352 & 0.084 & 0.069 & 0.025 & 0.057 & 0.032 & 0.050 & 0.000 & 0.030 & 0.066 & 0.044 & 0.045 & 0.025 & 0.049 & 0.032 \\ 0.111 & 0.334 & 0.085 & 0.069 & 0.019 & 0.042 & 0.026 & 0.050 & 0.000 & 0.022 & 0.067 & 0.039 & 0.038 & 0.014 & 0.048 & 0.037 \\ 0.048 & 0.351 & 0.105 & 0.059 & 0.023 & 0.054 & 0.031 & 0.046 & 0.000 & 0.031 & 0.065 & 0.053 & 0.044 & 0.014 & 0.050 & 0.041 \\ 0.042 & 0.314 & 0.170 & 0.062 & 0.017 & 0.040 & 0.026 & 0.610 & 0.000 & 0.023 & 0.062 & 0.043 & 0.045 & 0.010 & 0.048 & 0.035 \end{pmatrix} \quad (2)$$

$$= (2005 \cdot 4 \ 13664 \cdot 2 \ 3704 \cdot 2 \ 2449 \cdot 2 \ 919 \cdot 7 \ 2110 \cdot 9 \ 1206 \cdot 3 \ 1847 \cdot 4 \ 0 \ 1149 \cdot 5 \ 2519 \cdot 8 \ 1872 \cdot 1 \ 1691 \cdot 6 \ 688 \cdot 8 \ 1898 \cdot 1 \ 1475 \cdot 2)$$

Table 9. 마코브체인을 이용한 전국 16개 지역 폐차발생량 예측에

(단위 : 대)

	강원	경기	경남	경북	광주	대구	대전	부산	서울	울산	인천	전남	전북	제주	충남	충북
3개월 후	2005	13664	3704	2449	920	2111	1206	1847	0	1170	2520	1872	1692	689	1898	1475

Table 10. 마코브체인을 이용한 모형과 일변량 회귀분석모형의 성능 비교

(단위 : 대)

구 분	회귀분석 +Markov	지역별 회귀분석
16개 지역 MAD의 합	8884.4	9432.5

이와 같은 방법으로, 2009년 1월부터 2011년 9월(33개월)까지의 기간에 대해 마코브체인을 이용하여 폐차발생량을 예측해 보고, 이를 16개 각 지역을 일변량 회귀분석모형으로 적합한 결과와 비교하여 보았다. <Table 10>을 보면, 회귀분석과 마코브체인을 함께 활용한 모형의 16개 지역 MAD의 합은 8884.4(대)로 일변량 회귀분석 모형의 MAD인 9432.5(대)보다 작아 모형의 성능이 더 우수함을 알 수 있었다. 이러한 수치는, 2009년부터 2011년까지의 월평균 폐차발생량이 약 4.6만대임을 볼 때, 두 가지 예측 모형이 각각 평균적으로 약 8,900대와 9,400대의 오차를 나타냈다고 해석할 수 있다.

4.4.2 추가실험

이러한 결과가 특정 데이터집합을 사용한 했기 때문이라는 부분에 대한 객관성을 확보하고, 실험의 편향성을 방지하기 위해 세 가지 실험을 추가로 구성하였다.

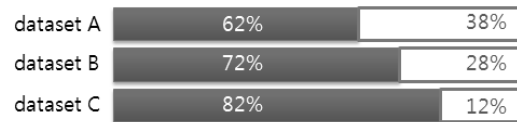
첫 번째로, 학습용 데이터집합(training dataset)과 검증용 데이터집합(testing dataset)의 비율을 달리하여 <Table 11>과 같은 세 종류의 데이터집합에 대해 같은 실험을 진행해보므로써, 마코브체인을 이용한 예측모형의 성능이 일변량 시계열 모형의 성능보다 우수함을 검증해 보았다.

다음으로, 회귀분석 외의 나머지 세 가지 지수평활, ARMA, 인공신경망 모형에 대해서도 마코브체인으로 인한 예측력 변화를 살펴보려 한다. 선행실험에서, 네 가지 예측모형 중 국내 폐차발생량의 특성을 가장 잘 반영하는 모형은 <Table 8>과 같이 회귀분석법으로 나타났기에, 가장 먼저 전국 폐차발생량(A_n)을 회귀분석으로 추정된 후 마코브체인을 이용해 새로운 모형

을 만들고 그 성능을 일변량 회귀분석모형과 비교해보았다.

따라서, 두 번째 추가실험에서는 나머지 세 가지 모형(지수평활, ARMA, 인공신경망)에 대해서도 마코브체인을 이용하여 모형을 정립해보고, 마코브체인을 이용하지 않은 각각의 일변량 시계열 모형보다 모형의 예측력이 개선되는지 그 성능변화를 여부를 관찰해보았다. 여기서, 편의상 마코브체인을 이용해 새롭게 만든 모형들을 ‘회귀분석+Markov’, ‘지수평활+Markov’, ‘ARMA+Markov’, ‘인공신경망+Markov’로 명명하기로 한다.

또한, 이러한 마코브체인으로 인한 효과가 나머지 세 가지 모형에서도 긍정적으로 나타난다면, 그것이 회귀분석+Markov 모형의 성능도 뛰어넘는 정도인지 그 시너지 효과도 살펴보았다.



마코브체인을 이용한 예측모형을 만들기 위해서는 그 선행과정으로 전국 16개 지역의 월별 폐차발생량 데이터에 대해 상태집합을 정의해야 했다. 이 때, 상태집합의 원소인 각 상태의 16개지역 구간범위를 4%를 넘지 않도록 제한을 두었었는데, 추가실험의 마지막으로 이러한 제한을 2%, 3%, 4%, 5% 네 가지로 일반화하여 모형의 예측률을 관찰하였다.

<Table 12>와 <Figure 9>에 추가로 진행된 세 가지 실험의 결과를 정리하였다. <Figure 9>의 짙은색 막대그래프는 마코브체인을 이용하지 않고 전국 16개 지역을 각 지역별로 회귀분석, 지수평활, ARMA, 인공신경망 네 가지 예측모형을 이용해 적합한 모형의 MAD 통계량을 나타낸 것이고(편의상 모형을 ‘지역별 회귀’, ‘지역별 지수평활’, ‘지역별 ARMA’, ‘지역별 인공신경망’로 명명), 나머지 옅은색의 막대 그래프들은 마코브체인을 이용해 새롭게 만든 모형들의 성능을 나타낸다.

또한, 상태집합의 정의방법에 따른 실험결과 역시 데이터집합별 모형별로 옅은색 막대에 나누어 표시하였으며, 이 중 가장 좋은 성능을 나타낸 모형에 ☆표시를 해두었다(편의상, ‘2% 모형’, ‘3% 모형’, ‘4% 모형’, ‘5% 모형’이라 명명).

Table 11. 학습용 데이터집합과 검증용 데이터집합의 비율에 따른 데이터집합 A, B, C

	A	B	C
학습용 데이터집합 (training data)	2002년 1월~2007년 12월 (72개월)	2002년 1월~2008년 12월 (84개월)	2002년 1월~2009년 12월 (96개월)
검증용 데이터집합 (testing data)	2008년 1월~2011년 9월 (45개월)	2009년 1월~2011년 9월 (33개월)	2010년 1월~2011년 9월 (21개월)
학습용 : 검증용	62% : 38%	72% : 28%	82% : 18%

Table 12. 데이터집합별 시계열 모형별 마코브체인의 효과

	데이터집합 A					데이터집합 B					데이터집합 C				
	2%	3%	4%	5%	지역별 회귀	2%	3%	4%	5%	지역별 회귀	2%	3%	4%	5%	지역별 회귀
회귀분석 +Markov 16개 지역 MAD의 합	9749.1	9680.4	9561.2	12291.3	10444.6	8811.3	8484.1	8884.4	10616.6	9432.5	8824.6	6222.5	6947.9	7111.9	6459.8
지수평활 +Markov 16개 지역 MAD의 합	10157.4	10080.8	9980.3	12635.1	10294.4	10720.1	10108.3	10412.4	11594.9	10584.6	11223.5	14247.6	13352.2	13935.9	11822.8
ARMA +Markov 16개 지역 MAD의 합	9939.4	9868.4	9773.0	12463.0	10189.5	10721.0	9992.1	10280.0	11975.8	10728.8	9336.7	11240.9	10627.2	11175.2	10060.0
인공신경망 +Markov 16개 지역 MAD의 합	12104.4	12002.7	11874.7	14115.0	12189.9	10862.8	10249.0	10553.7	12206.5	13388.3	9702.3	7526.6	8429.4	8476.4	8820.8

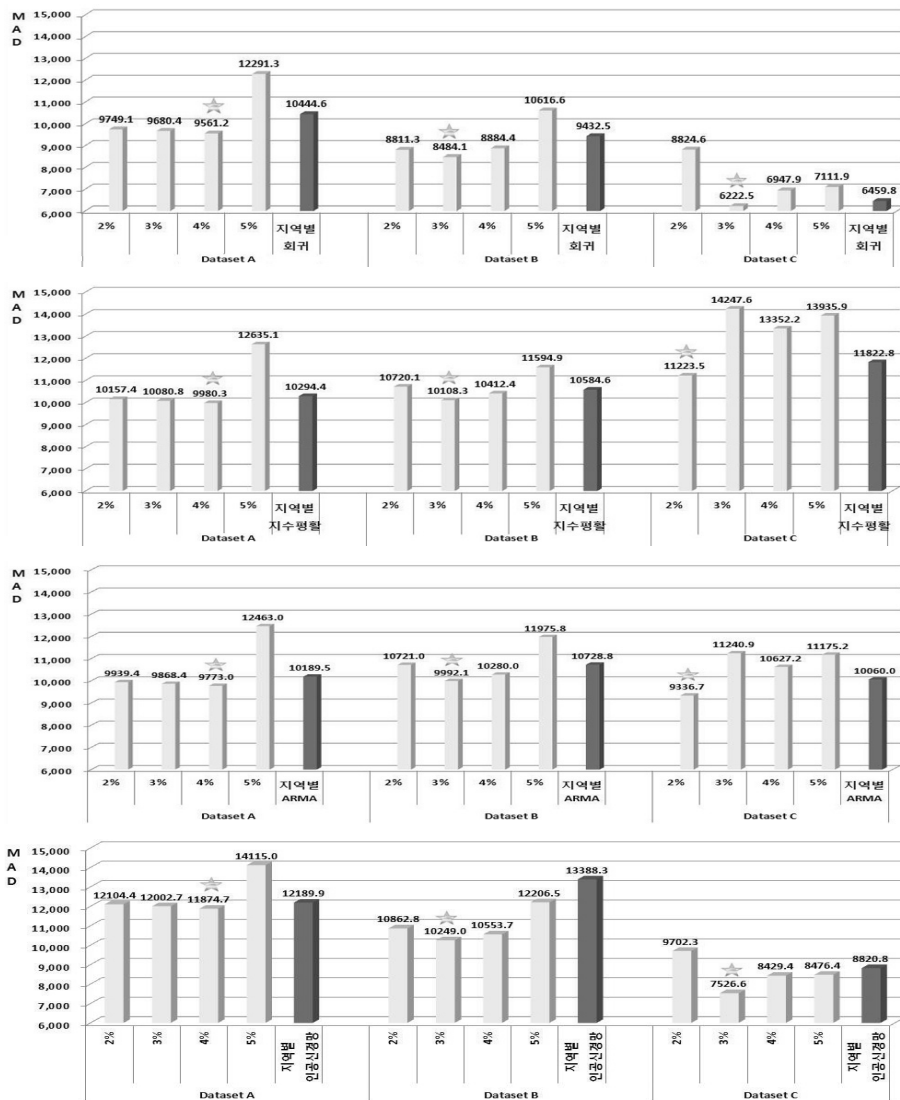


Figure 9. 데이터집합별 시계열 모형별 마코브체인의 효과

한편, 막대그래프의 높이는 MAD 통계량을 뜻하며 MAD의 정의에 따라, 그 값이 작을수록 예측력이 더 좋은 모형이라고 생각할 수 있다.

<Figure 9>와 함께 실험결과를 살펴보면, 첫째로 데이터집합과 시계열 모형의 종류에 관계없이 대부분의 경우에서 옅은색 막대들의 높이가 짙은색 막대의 높이보다 낮은 것을 확인할 수 있다. 이는, 첫 번째 실험으로 제안했던 데이터집합의 종류에 따른 마코브체인 모형의 효과를 나타내는데, 결과적으로 데이터집합에 독립적으로 마코브체인을 이용한 모형이 그렇지 않은 모형보다 좋은 예측력을 가짐을 뜻한다.

또한, ☆표시된 막대는 데이터집합별 모형별로 구분한 열 두가지 각각의 경우 모두에서 짙은색의 막대보다 낮은 높이에 위치해 있음을 볼 수 있다. 이는, 데이터집합과 시계열 모형의 종류에 독립적으로 마코브체인을 이용해 만든 모형이 그렇지 않은 시계열 모형보다 더 좋은 성능을 낼 수 있게 하는 상태집합이 항상 존재함을 의미한다.

한편, 각 데이터집합별로 살펴봤을 때 네 가지 시계열 모형 중 회귀분석에서 항상 ☆표시막대의 MAD 통계량이 가장 작게 나타났다. 이는, 앞선 결과에서 마코브체인을 이용하면 회귀분석 이외의 나머지 세 가지 시계열 모형(지수평활, ARMA, 인공신경망)도 역시 그 예측성능이 개선되는 것을 확인할 수 있었지만, 이러한 긍정적인 효과가 회귀분석을 뛰어넘지는 못하는 것으로 보아 그 시너지 효과는 존재하지 않는 것으로 보인다.

마지막으로 상태집합의 정의방법에 따른 마코브체인의 효과를 살펴보면, 상태집합의 원소인 각 상태들의 최대구간범위를 5%로 정하였을 때에는 데이터집합별 시계열 모형별로 구분한 열 두 가지 대부분의 경우에서 마코브체인을 이용하지 않은 모형보다 좋은 성능을 나타내지 못하였다. 이는, 상태집합이 너무 적은 수의 원소들로 구성되면 상태의 전이확률이 제대로 반영되지 못해 예측성능이 낮게 나타난 것으로 보인다.

그러나, 가장 좋은 성능을 가진 모형을 의미하는 ☆표는 상태집합의 원소가 가장 세분화된 2% 모형이 아닌 3% 혹은 4%

모형에서 가장 많이 나타났다. 이는 상태집합의 원소인 각 상태의 최대구간범위를 너무 좁게 정의하여도 오히려 모형의 성능에 역효과를 나타낸다는 사실을 알려준다.

4.5 폐차발생량 예측

지금까지, 마코브체인을 이용하여 국내폐차발생량 예측모형을 만들고, 몇 가지 실험을 통해 객관성을 확보하였다. 이 중 가장 좋은 성능을 나타낸 ‘회귀분석+Markov 4%모형’을 이용하여 2002년 1월부터 2011년 12월까지의 월간 폐차발생량데이터로 예측모형을 만들고, 향후 3년(36개월) 전국 16개 지역의 폐차발생량을 예측해 보았다. 그 결과는 <Table 13>과 같으며 <Figure 10>은 그에 따른 그래프를 나타낸다.

5. 결론 및 향후 연구과제

본 연구에서는 기존 폐차발생량 관련 원인변수에 관한 연구가 미비했던 점을 감안하여, 국내폐차발생량을 일변량 시계열 모형들로 적합해보고 그 중 가장 뛰어난 예측력을 나타낸 모형을 마코브체인을 이용하여 그 성능을 개선시켜보았다. 또한, 실험의 객관성을 확보하고 편향성을 방지하기 위해 데이터집합 세 가지로 달리하여 모든 실험을 진행하였다.

실험결과, 본 연구에서 사용한 회귀분석, 지수평활, ARMA, 인공신경망 네 가지 시계열 예측모형 중 국내폐차량을 가장 잘 설명하는 모형은 회귀분석 모형으로 밝혀졌으며, 네 가지 일변량 시계열 모형 모두에서 마코브체인을 이용해 모형을 개량했을 때 데이터집합에 관계없이 모두 훨씬 더 좋은 성능을 나타냈다. 또한, 마코브체인을 이용해 모형을 개선해도 회귀분석에 마코브체인을 이용해 새롭게 만든 모형이 여전히 가장 좋은 성능을 나타내었다.

한편, 마코브체인을 이용해 모형을 만들기 위해서는 전이확

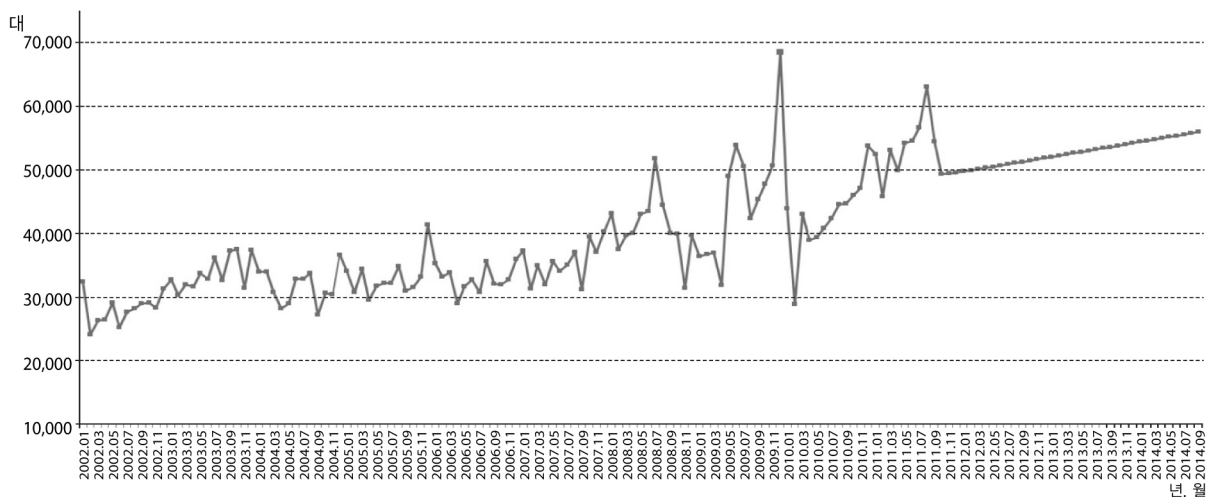


Figure 10. 2012년 1월부터 2014년 12월까지 전국 16개지역 폐차발생량 예측결과 그래프

Table 13. 2012년 1월부터 2014년 12월까지 전국 16개지역 폐차발생량 예측결과표

(단위 : 대)

	강원	경기	경남	경북	광주	대구	대전	부산	서울	울산	인천	전남	전북	제주	충남	충북	전국
1	2287.6	18822.6	4321.2	3236.9	1104.9	2650.7	1523.7	2412.4	0.0	1118.4	2781.0	2200.7	2120.6	1151.6	2145.2	1430.3	49307.8
2	2318.2	18651.2	4075.7	3236.5	1176.0	2624.8	1489.0	2414.3	0.0	1266.7	2772.1	2191.2	2119.3	1282.3	2392.1	1470.6	49480.2
3	2342.2	18591.8	4163.9	3266.2	1165.6	2639.0	1502.2	2418.8	0.0	1258.5	2821.1	2206.2	2127.8	1195.9	2359.5	1545.1	49603.8
4	2366.1	18543.6	4227.9	3292.4	1162.1	2649.4	1512.8	2424.5	0.0	1263.3	2856.1	2221.5	2135.7	1140.7	2345.4	1600.9	49742.5
5	2388.7	18510.5	4255.9	3313.2	1167.9	2657.5	1518.9	2429.3	0.0	1284.7	2883.9	2234.1	2144.0	1116.3	2362.4	1646.3	49913.7
6	2409.4	18494.3	4289.9	3333.8	1172.2	2667.4	1526.1	2434.2	0.0	1302.0	2911.9	2246.9	2153.1	1091.7	2374.1	1688.5	50095.7
7	2428.8	18489.5	4325.1	3353.8	1176.1	2678.0	1533.8	2439.7	0.0	1317.2	2938.6	2259.7	2162.4	1069.2	2383.7	1726.8	50282.3
8	2447.1	18493.3	4357.4	3373.0	1180.5	2688.6	1541.1	2445.5	0.0	1332.0	2963.4	2272.1	2171.7	1050.7	2394.4	1761.5	50472.1
9	2464.5	18504.9	4388.2	3391.5	1184.9	2699.2	1548.3	2451.7	0.0	1345.9	2986.8	2284.1	2181.0	1034.8	2405.0	1793.1	50663.8
10	2480.9	18523.1	4417.7	3409.4	1189.4	2709.9	1555.5	2458.2	0.0	1358.9	3009.0	2295.9	2190.2	1021.0	2415.5	1822.0	50856.6
11	2496.7	18547.1	4445.9	3426.9	1193.8	2720.5	1562.5	2465.0	0.0	1371.1	3030.1	2307.4	2199.4	1009.2	2425.9	1848.6	51049.9
12	2511.8	18576.2	4472.9	3443.8	1198.3	2731.2	1569.4	2472.1	0.0	1382.5	3050.2	2318.6	2208.5	999.0	2436.2	1873.1	51243.8
13	2526.3	18609.8	4498.8	3460.4	1202.8	2741.9	1576.2	2479.5	0.0	1393.3	3069.3	2329.5	2217.5	990.4	2446.4	1895.8	51438.0
14	2540.2	18647.5	4523.9	3476.6	1207.4	2752.5	1582.9	2487.0	0.0	1403.6	3087.7	2340.3	2226.5	983.1	2456.5	1916.8	51632.5
15	2553.7	18688.7	4548.1	3492.5	1211.9	2763.1	1589.6	2494.8	0.0	1413.3	3105.4	2350.8	2235.4	977.0	2466.5	1936.3	51827.2
16	2566.9	18733.1	4571.5	3508.2	1216.4	2773.7	1596.2	2502.8	0.0	1422.5	3122.4	2361.2	2244.3	972.0	2476.4	1954.5	52022.2
17	2579.6	18780.4	4594.3	3523.5	1221.0	2784.3	1602.7	2510.9	0.0	1431.3	3138.9	2371.4	2253.2	967.8	2486.3	1971.6	52217.3
18	2592.0	18830.1	4616.6	3538.7	1225.5	2794.9	1609.2	2519.2	0.0	1439.8	3154.9	2381.5	2262.0	964.5	2496.2	1987.6	52412.6
19	2604.1	18882.1	4638.2	3553.6	1230.1	2805.4	1615.6	2527.6	0.0	1447.9	3170.5	2391.4	2270.7	962.0	2505.9	2002.8	52608.1
20	2616.0	18936.1	4659.5	3568.4	1234.6	2816.0	1622.0	2536.1	0.0	1455.7	3185.6	2401.3	2279.5	960.1	2515.7	2017.1	52803.6
21	2627.7	18991.8	4680.3	3583.0	1239.2	2826.6	1628.4	2544.7	0.0	1463.3	3200.4	2411.0	2288.2	958.8	2525.4	2030.6	52999.3
22	2639.1	19049.2	4700.7	3597.5	1243.8	2837.1	1634.7	2553.4	0.0	1470.7	3214.9	2420.7	2296.9	958.0	2535.0	2043.6	53195.1
23	2650.4	19107.9	4720.8	3611.8	1248.3	2847.6	1641.0	2562.2	0.0	1477.8	3229.1	2430.3	2305.5	957.7	2544.7	2055.9	53391.0
24	2661.5	19167.9	4740.6	3626.0	1252.9	2858.2	1647.2	2571.0	0.0	1484.8	3243.0	2439.8	2314.2	957.8	2554.3	2067.7	53586.9
25	2672.4	19229.0	4760.2	3640.2	1257.5	2868.7	1653.5	2579.9	0.0	1491.6	3256.8	2449.3	2322.8	958.2	2563.9	2079.1	53783.0
26	2683.3	19291.0	4779.5	3654.2	1262.1	2879.2	1659.7	2588.9	0.0	1498.2	3270.3	2458.7	2331.4	959.0	2573.4	2090.1	53979.1
27	2694.0	19354.0	4798.6	3668.2	1266.6	2889.8	1665.9	2597.9	0.0	1504.8	3283.7	2468.0	2340.0	960.1	2583.0	2100.7	54175.2
28	2704.6	19417.7	4817.6	3682.1	1271.2	2900.3	1672.1	2607.0	0.0	1511.2	3296.9	2477.3	2348.6	961.4	2592.5	2111.0	54371.4
29	2715.1	19482.2	4836.4	3695.9	1275.8	2910.8	1678.3	2616.1	0.0	1517.5	3310.0	2486.6	2357.2	963.0	2602.0	2121.0	54567.7
30	2725.6	19547.2	4855.0	3709.7	1280.4	2921.3	1684.4	2625.2	0.0	1523.7	3322.9	2495.9	2365.8	964.7	2611.5	2130.8	54764.0
31	2735.9	19612.8	4873.5	3723.4	1285.0	2931.8	1690.6	2634.4	0.0	1529.8	3335.8	2505.1	2374.3	966.7	2621.0	2140.3	54960.4
32	2746.3	19678.9	4891.9	3737.1	1289.6	2942.3	1696.7	2643.6	0.0	1535.9	3348.5	2514.2	2382.9	968.8	2630.4	2149.6	55156.7
33	2756.5	19745.5	4910.2	3750.7	1294.2	2952.8	1702.8	2652.8	0.0	1541.9	3361.2	2523.4	2391.4	971.1	2639.9	2158.8	55353.2
34	2766.7	19812.4	4928.4	3764.3	1298.7	2963.3	1708.9	2662.0	0.0	1547.9	3373.7	2532.6	2400.0	973.4	2649.3	2167.8	55549.6
35	2776.9	19879.7	4946.5	3777.9	1303.3	2973.8	1715.1	2671.3	0.0	1553.8	3386.3	2541.7	2408.5	975.9	2658.8	2176.6	55746.1
36	2787.0	19947.2	4964.5	3791.5	1307.9	2984.3	1721.2	2680.6	0.0	1559.6	3398.7	2550.8	2417.0	978.5	2668.2	2185.4	55942.6

를행렬을 구성해야하고 그 선행에 있어 상태집합의 정의가 필요한데, 상태집합의 세분화정도에 따라 모형의 성능도 달라지는 것을 알 수 있었다. 여기서 주목할 점은, 실험에 사용된 데이터집합과 시계열 모형의 종류에 독립적으로 마코브체인을 이용했을 때 이용하지 않은 모형보다 더 좋은 성능을 가져오게 하는 패턴집합이 항상 존재했다는 점이다.

본 연구는 한국폐차협회에서 제공하는 2002년부터 117개월만의 월별 폐차발생량 자료만을 이용하여 실험데이터가 부족했던 부분과 일별량 시계열 데이터를 이용하여 갑작스런 증감에 영향을 주는 원인변수들을 반영하지 못한 한계점이 있다.

하지만, 갑작스런 증감을 가져오는 모든 이상치들에 대한 원인변수를 파악하여 폐물류의 발생량을 정확하게 예측한다는

것은 불가능하며, 과거 시계열자료만을 이용하여 패턴을 파악하고 그것만으로 예측모형을 구축했다는 데 의미가 있다. 또한, 이러한 불규칙한 자료에 있어 패턴을 읽어내 평균적인 경향을 예상하겠다는 부분이 오히려 기간범위를 대상으로 하는 대부분의 정책수립에 큰 역할을 할 것으로 기대한다. 기존의 연구에서 전문가의 주관적 견해를 바탕으로 폐차발생량을 예측했던 점과는 달리, 구체적 모형을 만들어 계량적 접근을 했다는 점에 의의를 둔다. 이와 더불어, 전국 혹은 유럽 몇 개국의 광범위한 폐차발생량을 예측했던 점과는 다르게 실제 폐차 네트워크를 고려하여 관련시설입지선정문제 등을 다룰 때 입력값으로 충분히 반영이 가능한 발생지별 구체적 예측치를 제시했다는 점에서 기존연구의 한계점을 극복하였다.

본 연구에서 제시된 마코브체인을 이용한 예측모형이 폐차뿐만 아니라 원인변수의 파악이 어려운 특성을 가진 품목들이나 특히, 불규칙적 발생경향을 가진 다른 폐물류 발생량의 예측에도 활용할 수 있을 것으로 기대한다.

이번 연구가 이제 막 시작단계에 있는 국내의 폐차관련연구에 있어 초석이 되길 기대하며, 향후 연구로는 폐차발생량과 관련된 원인변수들에 대한 연구를 추가로 진행하여 다변량 예측 기법과 본 연구에서 제시된 모형과의 성능을 비교해 봄으로써 현실적인 부분을 더욱 보완할 수 있는 연구를 진행하고자 한다.

참고문헌

- A Merksiz-Guranowska (2011), WIT Transactions on The Built Environment : The optimization of vehicles recycling facility location, URBAN TRANSPORT XVII, 116, 65-76, WIT Press.
- Bae, J.-G., Kim, Y.-G., and Choi, H.-G. (2006), Studies on circulation status and prospects of Food Waste Compost, *Journal of the Korea Organic Resource Recycling Association*, 14(1), 31-49.
- Baek, J.-G. and Han, J.-H. (2011), Forecasting of Heat Demand in Winter Using Linear Regresson Models for Korea District Heating Corporation, *Journal of the Korea Academia-Industrial Cooperation Society*, 12(3), 1488-1494.
- Byeon, S.-H. (2010), A study on analysis of environmental burden and reduction strategies in end of life vehicle(ELV) dismantling company, Master's Thesis of University of Ulsan,
- Frits M. Andersen, Helge V. Larsen, Mette Skovgaard (2007), A European model for the number of End-of-Life Vehicles, *International Journal of Automotive Technology and Management*, 7(4), 343-355.
- Hong, J.-H. (2004), The End-of-Life Vehicle(ELV) Directive and the Trends of Technology Development in Each County, 81-88, *Environments and Energy*, Seoul, Korea.
- Huh, N.-G. (2010), A Case Study of Air Demand Forecasting using Time-Series Models, Doctor's Thesis of Korea Aerospace University.
- Kang, H.-S., Yoon, H.-M., Jang, Y.-C., Kim, M.-C., Lee, W.-Y., and Song, H.-T. (2007), Generation, Collection and Management System of Used and End-of-Life Mobile Phones in Korea, *Conf. on Korean Society for Atmospheric Environment*, 2007(1).
- Kang, J.-H. (2005), A study on status analysis and improvement on recycling of End-of-Life vehicles industry in Korea, Doctor's Thesis of Myongji University.
- Kim, H.-J. (2005), Calculation of LTV and management of customer relationship by using Markov Chain model, Master's Thesis of Seoul National University.
- Kim, Y'-J., Park, D'-H., and, Nam, H. (2007), Recycling status and manufacturer activity under EU ELV Directive, *Journal of the Korean Society of Automotive Engineers*, 9(6), 1667-1672.
- Lee, D.-H. and Dong, M. (2009), Dynamic network design for reverse logistics operations under uncertainty, *Research Part E : Logistics and Transportation Review*, 45(1), 61-71.
- Lee, S.-H., Moon, B.-S., and Park, B.-J. (2009), "The Bus Delay Time Prediction Using Markov Chain," *Journal of the Intelligent Transportation System*, 8(3), 1-10.
- Min, D.-K. (2007), A Study on the Resource Recycling of the 3-Major Electric Equipment, *Journal of the Korea Society of Environmental Administration*, 13(1-36), 9-15.
- Shelon, M. R. (2000), Introduction to Probability Models, *A Harcourt Science and Technology Company*, 186.
- Park, J.-S., Kim, J.-D., Hong, S.-Y., Do, I.-H., and Bae, J.-G. (2008), Prediction of Generation Quantity and Recycling Value of Wood Waste in Korea, *Journal of the Korea Society of Waste Management*, 25(4), 363-369.
- Reynaldo Cruz-Rivera, and Jürgen Ertel (2008), Reverse logistics network design for the collection of End-of-Life Vehicles in Mexico, *European Journal of Operational Research*, 196(3), 930-939.
- Song, B.-S. (1996), A Study on the Generation of Commingled Plastic Wastes in Korea, Master's Thesis of Konkuk University.
- Yim, E.-S. (2007), A forecasting restaurant sales : Exponential smoothing models and ARIMA model, *Journal of the Korean Academic Society of Hospitality Administration*, 16(2), 139-154.
- Yoon, J.-S., Huh, N.-K., Kim, S.-Y., and Hur, H.-Y. (2010), A Study on International Passenger and Freight Forecasting Using the Seasonal Multivariate Time Series Models, *Journal of the Korean Statistical Society*, 17(3), 473-481.