

대리운전 시장의 지역별 수요 예측 모형의 성능 향상을 위한 방법론 연구

A Study on Methodology for Improving Demand Forecasting Models in the Designated Driver Service Market

김민섭 · 박기균 · 허재현 · 권재은 · 배혜림*

부산대학교 산업공학과 산업데이터공학융합전공

요약

대리운전 시장의 요금체계는 이용자의 대리운전 서비스 요청 시, 주위 대리운전 기사의 수, 대리운전 서비스 이용자 수, 날씨 등의 다양한 영향에 의해 실시간으로 변하는 Dynamic Pricing의 특징이 있다. 불확실한 변동성은 대리운전 서비스 요금을 상승시켜 고객의 이탈과 대리운전 기사의 배차거부를 유발하는 주된 원인이 되며, 이러한 문제를 해결하기 위해 적절한 수요를 예측하고 선제적으로 대응하기 위한 수요 예측 모형의 도입이 요구된다. 본 연구에서는 대리운전 서비스 이력 데이터를 활용하여 지역별, 시간대별 대리운전 서비스 수요를 예측하는 모형을 제시한다. 이후 실제 대리운전 서비스 이력 데이터를 활용하여 시간과 요일에 따른 조건부 확률을 구축하고 이를 예측 모형과 결합한 Time-Series with Conditional Probability 방법론을 제안하였으며 실험을 통해 SARIMA, Prophet의 기존 시계열 모형보다 성능이 우수함을 검증하였다. 본 연구는 제안된 방법론을 통해 구축된 수요 예측 모형을 활용하여 대리운전 서비스의 단기 전략 수립에 활용할 수 있다는 시사점이 있다.

■ 중심어 : 대리운전, 수요 예측, 조건부 확률, SARIMA, Prophet

Abstract

Nowadays, the Designated Driver Services employ dynamic pricing, which adapts in real-time based on nearby driver availability, service user volume, and current weather conditions during the user's request. The uncertain volatility is the main cause of price increases, leading to customer attrition and service refusal from driver. To make a good Designated Driver Services, development of a demand forecasting model is required. In this study, we propose developing a demand forecasting model using data from the Designated Driver Service by considering normal and peak periods, such as rush hour and rush day, as prior knowledge to enhance the model performance. We propose a new methodology called Time-Series with Conditional Probability(TSCP), which combines conditional probability and time-series models to enhance performance. Extensive experiments have been conducted with real Designated Driver Service data, and the result demonstrated that our method outperforms the existing time-series models such as SARIMA, Prophet. Therefore, our study can be considered for decision-making to facilitate proactive response in Designated Driver Services.

■ Keyword : Designated Driver Service, Demand Forecasting, Conditional Probability, SARIMA, Prophet

2023년 05월 19일 접수; 2023년 06월 12일 수정본 접수; 2023년 06월 16일 게재 확정.

* 이 논문(또는 특허 등)은(는) 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. RS-2023-00208999)

* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성(Grand ICT연구센터) 사업의 연구결과로 수행되었음(IITP-2023-2020-0-01791)

† 교신저자 (hrbae@pusan.ac.kr)

I. 서론

1.1 대리운전 시장의 규모

대리운전 서비스는 대한민국에서 가장 일반화된 생활 밀착형 교통서비스 중의 하나로 대리운전 업체가 기사와 손님을 중개해 주는 플랫폼이며, 연간 2~3조원 규모로 작지 않은 시장 규모를 가지고 있는 것으로 알려져 있다[1]. 현재 대한민국의 대리운전 서비스를 이용하는 이용자는 1일 약 50만 명, 대리운전 서비스를 제공하는 대리기사는 약 9만 명, 대리운전 업체는 약 3,851개로 추정하고 있다[1]. 최근에는 대기업의 대리운전 시장 진출 등의 계기로 신규 대리운전 기사들이 대거 유입되어, 현재 대리운전 기사 수는 2016년 11월 기준 최소 11만 명이 넘는 것으로 추정된다[2].

1.2 대리운전 서비스의 작동 방식

일반적인 대리운전 서비스 운영 방식은, 서비스를 이용하고자 하는 이용자가 대리운전 업체에 대리운전 서비스를 요청하면, 대리운전 업체가 출발지와 목적지, 요금 등의 대리운전 서비스 요청 내용을 프로그램에 입력한다. 해당 정보는 대리운전 기사의 스마트폰에 나타나게 되며, 각 지역에 대기하고 있는 대리운전 기사는 해당 정보를 보고 요청을 수락하면 대리운전 서비스가 성사된다.

대리운전 기사는 수요가 많은 지역, 주거지로부터 멀지 않은 지역 등 자신의 기준에 맞게 다양한 지역에 분포하며, 이는 대리운전 시장에서 공급에 해당한다. 반대로 대리운전 서비스를 이용하고자 하는 이용자는 수요에 해당한다.

1.3 대리운전 요금의 변동성

대리운전 시장의 요금체계는 이용자로부터

대리운전 서비스 이용 요청이 발생했을 때, 주위에서 이용자를 기다리고 있는 대리운전 기사 수(공급)와 대리운전 서비스를 원하는 고객의 수(수요), 출발지와 도착지에 따라 생성되는 상품의 형태에 따라 결정되며, 이 외에도 날씨, 해당 지역의 대규모 행사, 연휴 등의 영향을 받는 Dynamic Pricing의 특징을 가지고 있다. 대표적으로 공급 감소를 발생시키는 요인은 우천, 기온 등이 있으며, 수요를 증가시키는 요인은 연휴, 대규모 행사, 상권 발전 등이 있다.

이러한 대리운전 시장의 요금체계의 경우 변동성이 적은 항공권 예매시장 등과는 다르게 수요와 공급이 매우 유동적이며 상품 또한 출발지와 도착지에 따라 무한대에 가깝게 만들어질 수 있어 Dynamic Pricing의 특성이 있는 다른 시장들에 비해서도 변동성이 큰 특징을 가지고 있다[28].

특히 대리운전 시장의 경우, COVID-19의 영향으로 집합 제한 등의 행정 명령이 있었을 당시, 집합 제한 시간을 기준으로 수요가 급격하게 증가하고 감소하는 등의 문제가 발생하기도 하였다.

〈표 1〉 대리운전과 항공권 예매시장의 수요 공급의 특징

	대리운전 시장	항공권 예매 시장
수요 특징	짧은 시간 주기가 여러 개	긴 시간 주기가 고정적
	- 일별 주기 - 주별 주기(일/월요일 콜 감소) - 월별 주기(연말 성수기)	- 일별 주기(황금 연휴, 여름 휴가 등)
공급 특징	공급이 유동적 (예측 어려움)	공급이 고정적
	- 우천시 대리운전 기사 감소 - 지역별 대리운전 기사의 이동으로 공급 변화	- 항공기의 공급 고정적
상품의 특징	상품이 무한함	상품이 유한함
	- 이용자의 위치와 목적지에 따라 실시간으로 상품이 생성	- 사전에 정해진 기종, 노선에 따라 상품이 생성

이러한 변동성은 대리운전 서비스가 이루어지기까지 이용자의 대기시간을 증가시키거나, 대리운전 서비스 요금의 인상, 대리운전 기사의 배차거부 등 많은 비효율의 직접적인 원인이 되며, 이는 곧 대리운전 업체의 이용자 감소로 이어지게 된다. 이에 따라 변동성을 미리 예측하고 선제적으로 대응하기 위해 대리운전 시장에서 수요 예측의 필요성이 강조되고 있다.

본 연구에서는 이러한 문제를 극복하기 위하여 대리운전 서비스 이력 데이터를 활용하여 행정동별 단기 수요 예측 모형을 개발하고, 성능 향상을 위해 시계열 모형과 조건부 확률을 결합한 Time-Series with Conditional Probability(이하 TSCP) 방법론을 제안한다. 마지막으로 기존의 시계열 모형과의 비교 검증을 통해 제안 방법론의 우수성을 검증한다.

II. 선행 연구

수요 예측과 관련된 연구는 이전부터 다양한 분야에서 활발하게 진행되어왔다. 국내에서는 분야에 따라서 항만 컨테이너 물동량 예측 연구 [18], 최대 전력 수요 예측 연구[19], 철도 여객 수요 예측 연구[22] 등의 연구가 진행되었다.

수요 예측과 관련된 연구들은 공통적으로 미래 시점에 대한 예측을 통해 적절한 대응 방안을 마련하고자 하는 것에 그 목표가 있다. [18]에 따르면 정확한 수요 예측은 적절한 시점에 적절한 규모의 증설 및 증축을 위해 필요하다고 언급하고 있으며, [19]는 수요 예측을 통해 주/월간 예방정비 계획과 예비전력 운용 체계화에 활용할 수 있다는 점을 제시한다. [22]의 경우, 수요 예측 결과를 기반으로 이윤 극대화, 자원의 효율적 활용을 위한 증장기 및 단기 전략 수립에 사용할 수 있다는 점을 강조함과 동시에, 이를 위해서는 정확한 수요 예측이 선행되어야 한다고 언급한다. 이렇듯 수요 예측은 다양한

분야에서 장/단기 계획 수립 및 미래에 대한 선제 대응을 수립하기 위해 활용되고 있으며 높은 성능의 수요 예측 모형은 이를 위한 필수적인 선행 과제라고 할 수 있다.

동종 업계인 택시 서비스의 경우도 수요 예측 연구가 활발하게 진행되고 있다. [8]을 비롯하여 택시 서비스에서의 수요 예측 관련 연구는 아래 <표 2>와 같다.

택시 시스템을 활용한 인간의 이동 패턴을 예측한 [3]의 연구에서는 택시 시스템을 활용하여 개선된 ARIMA 모형을 제안하며, 택시 서비스

<표 2> 택시 서비스의 수요 예측 관련 연구 및 예측 모형

No	논문명	주요 성과	Method
[3]	Prediction of urban human mobility using large-scale taxi traces and its applications	택시 시스템을 활용한 인간의 이동 패턴 예측	ARIMA
[4]	A framework for passengers demand prediction and recommendation	승객 여행 수요를 기반으로 택시의 추천 핫스팟 예측	EWA
[5]	Predicting taxi - passenger demand using streaming data	택시 서비스에서 Streaming 데이터를 활용하여 승객 수요 예측	ARIMA
[6]	Taxi booking mobile app order demand prediction based on short term traffic forecasting	택시 수요 단기 예측을 위한 Wave SVM 방법론 제안	Wave SVM
[7]	Spatial-Temporal Diffusion Convolutional Network: A Novel Framework for Taxi Demand Forecasting	택시 수요 예측을 위한 ST-DCN 프레임워크 제안	ST-DCN
[8]	A model for short-term taxi demand forecasting accounting for spatio-temporal correlations	단기 택시수요 예측을 위한 GCRF 모델 제안	GCRF

의 경우 수요 예측 결과를 통해 수요와 공급을 완벽하게 일치시킬 수 있다면 비효율을 약 60% 까지 낮출 수 있다고 연구 결과를 제시하고 있다. 택시 핫스팟 추천 예측 진행한 [4]의 연구에서는 승객 여행 수요를 기반으로 EWA를 활용한 예측 모형을 제안하였다. ARIMA 모형을 활용하여 택시 수요 예측을 진행한 [5]의 연구에서는 Streaming 데이터를 기반으로 택시 수요 예측 모형을 제시하였고, [6]의 연구에서는 택시 서비스 이력 데이터를 활용하여 수요 예측을 위한 WaveSVM 모형을 제시하였다. 마찬가지로 택시 수요 예측을 진행한 [7]과 [8]의 연구에서는 동적으로 시공간의 상관관계를 고려하는 ST-DCN 방법론과 GCRF 방법론을 각각 제시하였다. 또한 [8]에서는, 택시의 수요를 예측하는 것은 택시를 적절한 곳에 배치하여 공급을 조절하는 등의 Dynamic Pricing의 정책을 결정하기 위해 필요하다고 언급하고 있다.

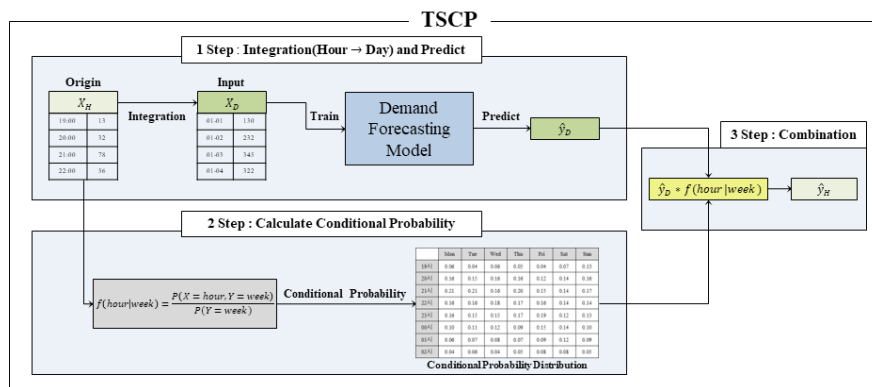
반면 대리운전 서비스와 관련하여 국내/국외를 막론하고 이러한 수요 예측 연구는 진행되지 않고 있다. 국내에서 대리운전과 관련된 연구는 대리운전 기사들의 근로조건 및 제도개선에 대한 연구[9], 음주운전 예방 대책에 대한 연구[10-12], 대리운전 시스템의 개선에 관한 연구[13,14], 대리운전의 마이리지 이용 특성에 관한 연구[15] 등의 연구만 제한적으로 진행되고 있는 실정이

다. 이러한 이유로는 첫 번째로 대리운전 서비스의 경우 전국 단위의 대리운전 서비스를 운영하고 있는 기업이 없고 지역 단위의 영세업체로 이루어져 있어 수요 예측 등의 데이터 기반의 분석을 수행하기 위해 필요한 데이터 인프라 구축의 미비, Digital Transformation 등의 영향이 미미하다는 점, 그리고 대한민국에서 주로 운영되는 형태의 대리운전 서비스는 택시 서비스와 같이 많은 국가에서 범용적으로 운영되는 서비스가 아니므로, 일반적으로 연구되기 힘든 주제인 점 등이 있다.

하지만 연간 2~3조원의 규모인 국내 대리운전 시장은, 2017년 기준 국내 택시운송업 규모인 약 8.5조원[27] 비교해보더라도 작지 않은 시장이며 택시 서비스에서 활발하게 진행되고 있는 수요 예측 연구가 대리운전 시장에서도 적용될 수 있음을 시사한다. 본 연구에서는 대리운전 서비스 이력 데이터를 활용하여 수요 예측 모형을 구축하고 성능 향상을 위한 방법론을 제시하였다.

III. 연구 방법론

본 연구에서 제안하는 방법론은 <그림 1>과 같다. 본 방법론의 목표는 지역별 1시간 뒤의 수요를 예측하는 예측 모형을 구축하고 성능을 향상시키는 것이며, 크게 3가지 단계로 구성된다.



<그림 1> Time-Series with Conditional Probability 프레임워크

첫 번째 단계는 시간 단위의 데이터를 일 단위의 데이터로 통합하여, 시계열 예측 모형을 구축하는 단계이다. 이 단계의 목적은 시간 단위의 데이터에서 나타날 수 있는 일반적이지 않은 패턴, 데이터상의 Noise 등을 예측 모형이 학습하지 않고, 강건한 형태의 예측을 진행하기 위한 단계이다.

두 번째 단계는 시간대별 데이터를 활용하여 실험을 진행할 지역의 조건부 확률(Conditional Probability)을 산출하는 단계이다. 조건부 확률은 지역별로 구축되며, 본 연구에서 제안하는 방법론에서는 요일과 시간대에 따른 조건부 확률 분포를 산출하였다. 이러한 조건부 확률은 지역별 특성을 반영하고 있으며 식 (1)과 같다. 식 (1)의 x_1 은 요일 변수로 월요일~일요일을 나타내며, x_2 는 시간 변수로 본 연구에서 실험을 진행한 19시~02시까지의 시간대를 나타낸다. y 는 두 조건 변수에 따른 대리운전 서비스 요청 수를 나타낸다.

$$\begin{aligned} x_1 &\in \{Mon, Tue, Wed, Thu, Fri, Sat, Sun\} \\ x_2 &\in \{19, 20, 21, 22, 23, 0, 1, 2\} \\ y &= \text{대리운전서비스요청수} \end{aligned} \quad (1)$$

마지막으로 세 번째 단계는 앞서 두 단계에서 산출된 일별 수요 예측 결과와, 지역별 조건부 확률을 결합하여 시간대별 수요를 예측하는 단계이다. 본 연구에서 제안하는 프레임워크를 활용하여 각 행정동의 특성을 반영한 시간대별 예측을 진행할 수 있다.

IV. 실험

4.1 실험 데이터 설명

본 실험에서는 대리운전 서비스를 운영 중인 A 회사의 2019년 1~12월까지의 대리운전 서비스 이력 데이터를 활용하였다. 데이터는 행정동

단위로 구분되어 있으며, 대리운전 서비스 이용자의 출발지와 도착지, 대리운전 서비스 요청 이력과 시간, 대리운전 기사 수락 여부 등이 기록되어 있다.

<표 3> 대리운전 서비스 이력 데이터 예시

No.	Variable	Description
1	대리운전 서비스 요청 시간	이용자의 서비스 요청 시간
2	출발지	출발 지역
3	도착지	도착 지역
4	요금	서비스 요금
5	완료 여부	서비스 완료 여부(완료/취소)

위의 <표 3>을 활용하여 구축한 실험 데이터의 형태는 <표 4>와 같고 시간대별/지역별 수요가 기록되어 있다. 본 실험에서는 1~10월의 데이터로 학습을 진행하고 11~12월의 데이터를 검증 데이터로 활용하여 기존의 시계열 예측 모형과 제안 방법론을 적용한 결과를 비교 평가한다.

<표 4> 행정동 A의 시간대별 대리운전 서비스 요청 건수 예시

TimeStamp	행정동 (A~G)	대리운전 서비스 요청 건수
2019-01-01 19:00:00	A	13
2019-01-01 19:00:00	B	32
2019-01-01 19:00:00	C	24
2019-01-01 19:00:00	D	25
2019-01-01 19:00:00	E	32
2019-01-01 19:00:00	F	23
2019-01-01 19:00:00	G	65
2019-01-01 20:00:00	A	23
2019-01-01 20:00:00	B	34
2019-01-01 20:00:00	C	45
2019-01-01 20:00:00	D	57
2019-01-01 20:00:00	E	34
2019-01-01 20:00:00	F	55
2019-01-01 20:00:00	G	78

4.2 실험 대상 정의

4.2.1 실험 지역 정의

본 실험에서는 수요가 높은 상위 10%의 행정동을 대상으로 실험을 진행하였다. 본 연구에서 실험을 진행하는 행정동은 총 7곳으로 지역명은 A~G로 비식별화하였다.

4.2.2 수요 정의

정확한 수요를 예측하기 위해 다음과 같은 세 가지의 경우는 모두 수요로 간주하였다. 1) 실제 대리운전 서비스가 이루어진 경우, 2) 이용자가 대리운전 서비스를 요청하였지만 대리운전 기사와의 중개가 이루어지지 않은 경우, 3) 이용자가 대리운전 서비스를 요청하였고 중개가 되었지만 부득이한 이유로 완료되지 않은 경우, 위 세 가지 경우를 모두 포함하여 출발지를 기준으로 대리운전 서비스 수요로 정의하였다.

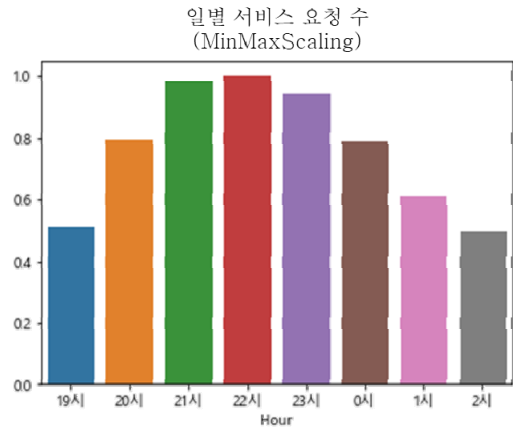
4.2.3 예측 시간대 정의

대리운전 시장은 늦은 오후~새벽 시간대에 주로 이루어지는 특징이 있다. 그러므로 대리운전 서비스가 거의 발생하지 않는 오전 시간대에 대한 예측은 미래의 수요를 예측하고 대비하기 위한 목표와 부합하지 않는다. 따라서 본 실험에서는 하루 전체에서 약 95%를 차지하고 있는 19~02시의 시간대에서 발생한 대리운전 서비스 이력 데이터만을 대상으로 예측 시간 범위를 정의하고 실험을 진행하였다.

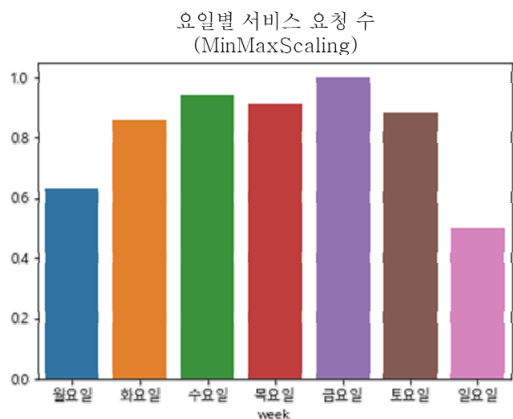
4.3 예측 모형 선정

대리운전 서비스는 <그림 2>와 같이 시간대별 주기와 <그림 3>과 같은 주별 주기 있다. 본 연구에서는 이러한 계절적 특징을 반영할 수 있는 모형들을 선정하고 실험을 진행하였다.

대리운전 서비스 이력 데이터와 같이 단변량



<그림 2> A 지역 시간대별 대리운전 서비스 수요 추이



<그림 3> A 지역 일별 대리운전 서비스 수요 추이

데이터를 활용하여 수요 예측을 진행하는 연구에서는 대표적으로 ARIMA 모형, SARIMA 모형, Prophet 모형 등이 활용된다. SARIMA 모형을 이용한 연구는 발전량 예측 모형을 구축한 연구[16,17,20], 컨테이너 물동량을 예측한 연구[18], 최대 전력수요 예측 연구[19], 김해국제공항의 여객 수요예측 연구[21] 등이 있다. Prophet 모형을 이용한 연구는 부산광역시 대기 PM₁₀ 농도 예측 연구[23], 국내 중소형 컨테이너항만 물동량 예측에 관한 연구[24]와 포항항 물동량 예측 연구[25], 인천공항 이용객 수요 예측 연구[26] 등이 있다. 이러한 모형들은 차분을 통해 정상성을 만

족할 수 있다는 점과, 시계열 데이터에서 나타나 는 다양한 계절성을 고려할 수 있다는 점으로 수 요 예측 등의 분야에서 주로 사용되고 있다.

본 연구에서는 다양한 주기와 단변량의 특징 을 가지고 있는 대리운전 서비스 이력 데이터의 특성을 반영하기 위해, SARIMA 모형과 Prophet 모형을 선정하여 실험을 진행하였다.

$$\varnothing_p(B)^d_p(B^s)(1-B)^d(1-B^s)^d y_t = \varnothing_q(B)\theta_q(B^s)a_t \quad (2)$$

SARIMA 모형은 위 <식 1>로 구성되며 통계 모형인 Auto Correlation 모형과 Moving Average 모형에 차분이 더해진 Autocorrelation Integrated Moving Average Model에 계절적 성분인 S (Seasonal)이 추가된 모형이다.

$$y(t) = g(t) + s(t) + h(t) + e_t \quad (3)$$

Prophet 모형은 2017년 Facebook에서 공개된 시계열 예측 모형으로 <식 2>로 구성되며, Multi-Seasonality와 Holiday의 특성을 반영하기 위해 개발되었다. <식 2>는 Prophet 모형을 나타내며 g(t)는 시계열 값의 비정상적인 변화를, s(t)는 주기적인 변화를 나타내고 h(t)는 불규칙하게 발생 하는 휴일의 영향을 나타낸다. 오차항은 모형으 로 설명할 수 없는 오차를 나타내고 정규분포를 따른다는 모수적 가정을 한다.

4.3.1 수요 예측 모형 평가지표 선정

검증 데이터로 사용한 11~12월은 대리운전 시장의 연말 특수가 반영되어, 평균적으로 수요 가 증가하는 경향이 있다.

<표 5> 연말에 따른 일별 평균 대리운전 서비스 이용 수 차이

그룹	N	t-value	p-value
1~10월 일별 평균 대리운전 서비스 요청 수	304	-2.563	0.010
11~12월 일별 평균 대리운전 서비스 요청 수	61		

이러한 경향에 강건하게 반응하는 수요 예측 모형의 평가지표를 선정하기 위해 본 연구에서 는 평균 제곱 오차(Mean Squared Error, MSE)가 아닌 평균 제곱근 오차(Root Mean Squared Error, RMSE)를 사용하여 수요 예측 모형의 성능을 평가한다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y - \hat{y})^2}{n}} \quad (4)$$

4.4 실험

4.4.1 선정된 모형에 대한 시간대별/일별 예측

본 실험은 제안 방법론을 적용하기 전, 선정된 모형을 사용하여 시간대별/일별 예측을 진행한 결과는 <표 6>과 같다.

<표 6>의 RMSE(1h)는 시간 단위의 대리운전 서비스 이력 데이터를 예측하는 시계열 모형의 성능을 나타내고 RMSE(1h)*8은 시간 단위의

<표 6> SARIMA, Prophet 모형의 시간대별 예측 결과

Model	Prov	RMSE(1h)	RMSE(1h)*8	RMSE(1d)
SARIMA	A	20.481	163.848	49.543
	B	14.451	115.608	40.556
	C	20.759	166.072	64.864
	D	16.106	128.848	50.549
	E	13.882	111.056	42.091
	F	18.701	149.608	47.556
	G	11.605	92.84	26.735
	Avg	16.57	132.55	45.98
Prophet	A	12.282	98.256	49.808
	B	9.499	75.992	40.435
	C	14.014	112.112	63.915
	D	10.605	84.84	51.052
	E	9.889	79.112	43.531
	F	11.597	92.776	49.238
	G	7.748	61.984	26.659
	Avg	10.80	86.44	46.38

Error에 8시간을 곱하여 하루 단위의 Error로 변환한 결과이다. 이를 통해 시간대별로 예측을 진행한 모형은, 일별 예측에 비해 Error가 3배 이상 높은 것을 확인할 수 있다. 또한 SARIMA 모형은 Prophet 모형과 달리, 일별 주기와 주별 주기를 동시에 고려할 수 없어 Error가 높게 나타난다.

RMSE(1d)는 시간 단위의 대리운전 서비스 이력 데이터를 일별 단위로 변환 후 학습한 시계열 모형의 성능을 나타내며, 본 연구의 제안 방법론의 Step 1에 해당한다. 이 경우, 선정된 두 모형 모두 주별 주기만을 고려하면 되어 모두 비슷한 성능을 가지고 있는 것을 확인할 수 있다.

4.4.2 TSCP 방법론 적용 및 예측

본 연구에서 제안하는 방법론을 적용하기 위해 Step 2에 해당하는 행정동별 조건부 확률 분포를 산출하는 과정을 진행하였다. 다음 <표 7>은 과거 3개월의 행정동별 대리운전 서비스 이력

<표 7> A 행정동의 조건부 확률 분포 예시

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
19시	0.06	0.04	0.06	0.05	0.04	0.07	0.13
20시	0.16	0.15	0.16	0.16	0.12	0.14	0.16
21시	0.21	0.21	0.16	0.20	0.15	0.14	0.17
22시	0.16	0.16	0.18	0.17	0.16	0.14	0.14
23시	0.16	0.15	0.15	0.17	0.19	0.12	0.13
00시	0.10	0.11	0.12	0.09	0.15	0.14	0.10
01시	0.06	0.07	0.08	0.07	0.09	0.12	0.09
02시	0.04	0.06	0.04	0.05	0.08	0.08	0.05

데이터를 활용하여 A동의 조건부 확률 분포를 산출한 결과이다.

본 연구에서 제안하는 방법론을 통해 시간대별 예측을 수행한 결과는 <표 8>을 통해 확인할 수 있고, 제안 방법론과 기존의 시계열 모형의 성능은 <표 9>를 통해 확인할 수 있다. <표 9>의 Original은 실험에 선정된 모형들을 사용한 시간대별 수요 예측 모형의 성능을 나타내며, Proposed

<표 8> 예측모형과 TSCP 방법론을 적용한 수요 예측 결과 시각화

Model	Prov	Original	TSCP
Prophet	A		
	B		
	C		
SARIMA	A		
	B		
	C		

〈표 9〉 시계열 모형에 대한 TSCP 방법론 적용 성능 비교

Model	Prov	RMSE	
		Original	Proposed Method
SARIMA	A	20.481	10.951
	B	14.451	8.923
	C	20.759	12.584
	D	16.106	9.987
	E	13.882	9.098
	F	18.701	10.929
	G	11.605	7.665
	Avg	16.57	10.020 (Improve 40%)
Prophet	A	12.282	10.973
	B	9.499	8.913
	C	14.014	12.492
	D	10.605	10.023
	E	9.889	9.208
	F	11.597	11.024
	G	7.748	7.663
	Avg	10.80	10.042 (Improve 7%)

Method는 제안 방법론의 시간대별 수요 예측 성능을 나타낸다.

제안 방법론을 적용한 경우, 두 모형 전부 유의미한 성능 향상을 확인할 수 있으며, 이러한 실험 결과의 이유는 다음과 같다.

첫 번째로, SARIMA 모형과 같이 하나의 주기만 고려할 수 있는 모형도 여러 주기를 고려하여 수요 예측을 진행할 수 있다는 점이다. 이는 제안 방법론의 Step 2의 조건부 확률을 구성하는 과정에서 다양한 주기들에 대한 정보를 반영할 수 있게 된 것이 그 이유이며, SARIMA 모형이 여러 주기를 고려할 수 있는 Prophet 모형에 비해 월등히 성능이 향상된 점을 통해 검증하였다.

두 번째로 수요 예측 모형이 학습 데이터에만 의존하지 않고 조건부 확률을 통해 최신 패턴을 반영하여 예측할 수 있다는 점이다. 기존의 모형의 경우 변동이 크지 않은 1~10월의 데이터만을 학습하게 되면 비교적 패턴이 다르게 나타나는 11~12월의 데이터에 대한 예측 성능이 감소할

수 있으나, 본 연구에서 제안하는 방법론을 적용하게 되면 조건부 확률을 통해 최신 경향의 데이터를 반영하여 수요 예측을 진행하게 되므로, 학습 데이터에서 볼 수 없는 형태의 데이터 패턴이 나오게 되더라도 이를 반영하여 예측을 진행할 수 있다는 장점이 있다.

세 번째로, 제안 방법론을 통해 Negative Prediction을 방지할 수 있게 된 점이다. Negative Prediction이란 수요 예측 중 특히 단기 수요 예측을 진행할 때 주로 발생하며, 예측 모형이 음수의 값을 예측하게 되는 문제이다. 이는 주로 데이터를 단기간의 형태로 변환하는 과정에서 0에 가까운 값이 많으면 발생한다. 본 연구에서는 특정 지역의 새벽 시간대에 대리운전 서비스 요청이 한 건도 발생하지 않은 경우가 대표적이며, 후처리 등을 통해 음수로 예측된 결과를 0으로 치환하는 등의 작업이 필요하다. 하지만 제안 방법론과 같이 일 단위의 예측 이후 조건부 확률을 결합하는 방식은 예측 결과가 음수로 발생하지 않으므로 이러한 문제를 해결할 수 있다.

즉, 본 연구에서 제안하는 방법론은, 조건부 확률과의 결합을 통해 다양한 시계열성을 고려할 수 있고, Negative Prediction을 방지하며, 지역별 최신 패턴을 반영하여 이상치적인 특성에도 강건하게 예측을 할 수 있다는 장점이 있으며, 기존의 시계열 예측 모형보다 성능을 향상시킬 수 있다는 점을 검증하였다.

V. 결론

마지막으로 본 연구에서 제안하는 방법론의 기대효과, 대리운전 시장에서의 수요 예측 모형의 활용 방안 그리고 향후 연구 방향에 대해 논의를 진행한다.

5.1 기대효과

첫 번째로 단변량 데이터를 활용하여 시계열

예측 모형을 구축하는 경우, 유의미한 성능 향상을 기대할 수 있다. 본 연구에서 제안하는 방법론과 기존의 시계열 모형의 차이점은, 단변량 데이터만을 사용하지 않고 해당 데이터로 추가적인 조건부 확률에 대한 정보를 산출하고 이를 예측에 활용하는 것에 있으며, 이러한 과정을 통해 단변량 시계열 예측 모형의 예측 성능의 향상을 검증하였다. 노운승 및 도명식의 연구에서 언급한 바와 같이, 높은 성능의 수요 예측 모형은 장/단기의 예측을 통해 중장기 및 단기 전략 수립을 위해 필수적이며 이를 위해 정확한 수요 예측 모형이 요구되는 만큼, 본 연구의 필요성을 찾을 수 있다[22].

두 번째로는, 다양한 주기를 고려할 수 없는 모형도 본 연구에서 제안하는 방법론을 통해 활용할 수 있다는 점이다. 본 연구에서 진행했던 SARIMA 모형의 경우 여러 주기를 동시에 고려할 수 없다는 단점이 존재하였지만, 제안 방법론을 적용하였을 때 성능이 약 40% 상승한 점을 확인하였다. 이는 여러 주기에 대한 정보들이 반영된 조건부 확률 분포를 결합한 것이 다양한 주기를 고려할 수 있게 해주는 요인이 될 수 있다는 점을 시사한다.

5.2 수요 예측 모형의 활용 방안

대리운전 시장과 비슷한 형태를 가지고 있는 택시 시장의 경우 [8]과 같은 연구를 통해, 지역별 수요를 예측하여 수요가 증가할 것으로 예상되는 지역으로 택시를 배치하는 정책을 제안하였다. 이러한 정책을 통하여 수요가 증가하는 지역에 공급을 같이 늘려 비효율을 감소시키고 요금의 변동성을 감소시킬 수 있다고 언급하고 있다.

대리운전 시장의 경우 본 연구의 수요 예측 결과를 활용하여 대리운전 기사가 수요가 많은 지역으로 이동할 수 있게 배치하거나 유도함으

로써, 공급을 조절하여 요금의 변동성을 줄이고, 비효율을 감소시킬 수 있는 방향으로 활용될 수 있을 것이다. 다만 이동 수단이 같이 움직이는 택시 서비스의 특성상 공급의 이동이 자유롭지만, 대리운전 기사의 경우 자체적인 이동 수단이 부재한 경우가 많아 이동이 원활하지 않을 수 있다는 점은 한계로 남아 있다.

5.3 향후 연구 방향

대리운전은 행정동별로 주거지역, 상업지역, 유동인구 등 다양한 공급/수요 패턴이 나타나게 되며, 각각의 행정동의 특성을 다양한 방식으로 패턴화할 수 있다. 본 연구에서는 시간대와 요일에 대한 패턴을 고려하여 조건부 확률 분포를 구성하였지만, 이는 도메인 지식, 혹은 다양한 기준에 따라 새롭게 구성될 수 있으며, 추가적인 파라미터를 고려할 수 있다. 이를 통해 새로운 지역별 특성을 찾아내고 높은 성능의 수요 예측 모형을 구축하는데 기여할 수 있을 것이다.

참 고 문 헌

- [1] 국토교통부, 자가용자동차 대리운전 실태조사 및 정책 연구, 2014.
- [2] 산업연구원, 대리운전 서비스시장의 이슈와 과제, 2016.
- [3] X. Li et al., "Prediction of urban human mobility using large-scale taxi traces and its applications", *Frontiers Comput. Sci.*, vol. 6, no. 1, pp. 111-121, 2012.
- [4] K. Zhang, Z. Feng, S. Chen, K. Huang and G. Wang, "A framework for passengers demand prediction and recommendation", *Proc. IEEE Int. Conf. Services Comput.*, pp. 340-347, Jun. 2016.
- [5] L. Moreira-Matias, J. Gama, M. Ferreira, J.

- Mendes-Moreira and L. Damas, "Predicting taxi - passenger demand using streaming data", IEEE Trans. Intell. Transp. Syst., vol. 14, no. 3, pp. 1393-1402, Sep. 2013.
- [6] Li, Yunxuan, et al. "Taxi booking mobile app order demand prediction based on short term traffic forecasting." Transportation Research Record 2634.1 (2017): 57-68.
- [7] Luo, Aling, et al. "Spatial-Temporal Diffusion Convolutional Network: A Novel Framework for Taxi Demand Forecasting." ISPRS International Journal of GeoInformation 11.3 (2022): 193.
- [8] Qian, Xinwu, et al. A model for short-term taxi demand forecasting accounting for spatio-temporal correlations. No. 17-02470. 2017.
- [9] 김난희, 특수형태근로종사자의 근로자성 판단기준에 관한 연구, 충남대학교 대학원, 석사학위청구논문, 2012.
- [10] 하삼중, "경찰의 음주운전 예방대책에 관한 연구 : 음주운전 자체사고 유발요인분석을 중심으로", 중앙대학교 대학원, 석사학위청구논문, 2009.
- [11] 김상구, "음주운전 예방정책에 관한 연구: 외국 사례와 유발 요인분석을 중심으로", 중앙대학교 대학원, 박사학위청구논문, 2008.
- [12] 정대용, "자동차보험에 있어서 무면허 및 음주운전면책약관", 경북대학교 대학원, 석사학위청구논문, 2008.
- [13] 석한음, "O2O 방식이 적용된 대리운전 서비스 소비자 후생효과," 정보통신정책연구, 제23권, 제4호, pp.1-28, 2016.
- [14] 이철호, "대리운전기사 중심의 대리운전 시스템 제언", 한국민간경비학회지, 제12권, 제4호, 2013.
- [15] 안세홍, "서비스기업의 고객 마일리지이용특성이 브랜드도, 브랜드충성도, 재이용의도에 미치는 영향 연구-대리운전 이용자를 중심으로" 한국콘텐츠학회논문지 20.1, pp.202-216, 2020.
- [16] 이동현, 정아현, 김진영, 김창기, 김현구, & 이영섭. SARIMA 모형을 이용한 태양광 발전량 예보 모형 구축. 한국태양에너지학회 논문집, 제39권 제3호, pp.59-66, 2019.
- [17] 정하영, 홍석훈, 전재성, 임수창, 김종찬, 박형욱, & 박철영. "SARIMA 모델을 이용한 태양광 발전량 예측 연구", 멀티미디어학회논문지, 제25권 제1호, pp.82-91, 2022.
- [18] 민경창, 하현구. "SARIMA 모형을 이용한 우리나라 항만 컨테이너 물동량 예측", 대한교통학회지, 32.6: pp.600-614, 2014.
- [19] 김시연, 정현우, 박정도, 백승묵, 김우선, 전경희, & 송경빈. "계절 ARIMA 모형을 이용한 104 주 주간 최대 전력수요예측", 조명·전기설비학회논문지, 제28권 제1호, pp.50-56, 2014.
- [20] Jeong, H. Y., Hong, S. H., Jeon, J. S., Lim, S. C., Kim, J. C., Park, H. W., & Park, C. Y. "A Research of Prediction of Photovoltaic Power using SARIMA Model" Journal of Korea Multimedia Society, 제25권, 제1호, pp.82-91, 2022.
- [21] 이지영. "SARIMA 모형을 이용한 김해국제공항의 여객 수요예측" (Doctoral dissertation, 부경대학교), 2023.
- [22] 노운승, 도명식, "SARIMA 모형을 이용한 철도 여객 단기수송수요 예측", 한국 ITS 학회 논문지, 제14권 제4호, pp.18-26, 2015.
- [23] 박선엽, "Prophet 모형에 기반한 부산광역시 대기 중 PM₁₀ 농도의 시계열 예측 기법 평가", 국토지리학회지, 제56권 제4호, pp.333-351. 2022.
- [24] 김준기, 류동근, 남형식, "Prophet 모형을 활용한 국내 중소형 컨테이너항만 물동량 예측에 관한 연구: 인천, 평택·당진, 울산항을 중심으로", 인문사회 21, 제13권, 제1호, pp.561-576,

2022.

- [25] 오진호, 최정원, 강태현, 서영준, 곽동욱, “시계열 데이터를 활용한 포항항 물동량 예측: SARIMA, Prophet, Neural Prophet 의 적용”, 무역학회지, 제47권 제6호, pp.291-305, 2022.
- [26] 이지훈, 한혜림, 윤상후. “시계열 모형을 이용한 인천공항 이용객 수요 예측”, 디지털융복합연구, 제18권, 제12호, pp.87-95, 2020.
- [27] 한국소비자원 시장조사국 온라인거래조사팀, 택시 플랫폼 이용 실태조사, 2022.
- [28] 카카오 모빌리티, 카카오 모빌리티 리포트, 2020.



권재은(Jae-Eun Kwon)

- 2022년 : 동국대학교(경주) 빅데이터 · 응용통계학과 (이학사)
- 2022년~현재 : 부산대학교 산업공학과 (석사과정)
- 관심분야 : 머신러닝



배혜림(Hye-Rim Bae)

- 1996년 : 서울대학교 산업공학과 (공학학사)
- 1998년 : 서울대학교 산업공학과 (석사)
- 2002년 : 서울대학교 산업공학과 (박사)
- 2004년~현재 : 부산대학교 산업공학과 (교수)
- 관심분야 : 프로세스 마이닝, 빅데이터, BPM (Business Process Mining), 인공지능 등

저자 소개



김민섭(Min-Seop Kim)

- 2020년 : 창원대학교 일어일문학과 (인문학사)
- 2022년~현재 : 부산대학교 산업공학과 (석사과정)
- 관심분야 : 데이터마이닝



박기군(Ki-Kun Park)

- 2017년 : 가천대학교 응용통계학과 (경영학사)
- 2019년~현재 : 부산대학교 산업공학과 (박사과정)
- 관심분야 : 항만물류 최적화



허재현(Jae-Hyeon Heo)

- 2021년 : 동아대학교 경영학과 (경영학사)
- 2022년~현재 : 부산대학교 산업공학과 (석사과정)
- 관심분야 : 머신러닝