



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

도시계획학 석사학위논문

심층 신경망을 이용한
수단선택모형 추정

Estimation of Travel Mode Choice
Models using Deep Neural Networks

2019년 02월

서울대학교 환경대학원
환경계획학과 교통학전공
방 대 환

국문초록

교통계획에서 교통수요를 예측하기 위해 전통적으로 4단계 교통수요 추정방법이 사용된다. 그 중 세 번째인 수단분담선택 단계는 최근 각 지자체의 신교통수단의 도입이나 산업의 발전에 따른 스마트 모빌리티, 카셰어링 등의 이용으로 복잡해진 통행자의 통행패턴 변화로 인해 그 중요성이 더욱 커지고 있다. 그동안 수단분담선택 단계에서는 주로 로짓 모형을 이용하여 통행자의 수단을 예측하여왔다. 그러나 로짓 모형은 통행자 개개인의 선호 다양성을 반영하기 어려운 문제와 비관련 대안의 독립성 문제를 가지고 있었다. 이에 이를 극복하고자 하는 연구가 다양하게 이루어짐을 선행연구를 통해 알 수 있었으나 여전히 비관련 대안의 문제, 자료의 희소성 문제 그리고 다양한 환경 조건과 통행자 개인의 특성을 온전히 반영하기에는 어려운 한계가 존재한다. 또한 통행자가 이용할 수 있는 교통수단이 매우 복잡해지고 다양해지면서 기존의 로짓 모형이 빠르게 변화하고 진화하는 교통수단에 대해 대응가능한지 확인하고, 이를 보완할 새로운 수단선택 모형의 제안이 필요함을 알 수 있다.

따라서 본 연구는 딥러닝의 한 종류인 심층 신경망을 이용하여 기존 로짓 모형의 한계를 극복할 수 있는 수단선택 모형을 추정한다. 본 연구에서 제시하는 심층 신경망은 수리통계학적 모형인 로짓 모형과는 다르게 상관관계가 높은 변수도 입력 자료로 사용할 수 있어 다양한 데이터의 적용과 많은 변수를 고려할 수 있는 장점이 있다. 또한 비관련 대안의 독립성 문제와 자료 희소성의 문제도 극복할 수 있다.

본 연구는 통행자의 사회경제적 특성과 통행특성을 동시에 고려하기에 용이한 2016년 가구통행실태조사 자료를 이용하여 서울시 통근목적 통행을 대상으로 수단선택 모형을 추정한다. 수단선택 모형의 변수는 통행자의 통행특성을 반영하는 통행시간(차내시간+차외시간), 통행비용과 사회경제적 특성을 나타내는 성별, 나이, 소득, 자동차보유여부로 설정하였다. 사회경제적 특성을 나타내는 변수 중 하나인 나이의 경우, 대중교

통 요금의 할인이 적용되는 청소년, 전철 요금 면제가 적용되는 노인, 그리고 일반인으로 3가지 범주로 나누었으며, 소득은 가구원별 평균소득으로 재구성하였다. 본 연구는 다양하고 성격이 비슷한 교통수단에도 정확성이 높은 모형을 추정하기 위해 종속변수로 단일수단 뿐만 아니라 복합수단도 포함하여 승용차, 버스, 전철, 버스+전철로 총 4가지 교통수단에 대해 추정한다.

심층 신경망 모형의 경우, 모형의 복잡도에 따라 성능이 달라지고, 특히 은닉층의 개수가 증가하며 층이 깊어질수록 다차원의 분류가 가능하다. 따라서 이러한 모형의 성능을 확인하기 위해 은닉층의 개수에 따라 유형을 3가지로 구분하여 정립하였다. 모형의 추정과 검증을 위해 가구통행실태조사 원시 자료를 전처리하여 구축한 총 34,013 통행의 표본을 가지고 모형을 추정한다. 전체 데이터 중 70%인 23,809개의 통행은 모형의 학습을 위한 데이터로 사용하며, 나머지 30%의 데이터는 모형의 검증을 위해 활용한다.

교통수단 선택모형의 추정결과, 은닉층을 3개 가진 심층 신경망 Type3 모형의 예측 정확도가 96.87%로 검증 데이터 10,204 통행 중 단 319 통행만을 오분류한 것으로 분석되어 가장 정확도가 높았다. 그 뒤를 이어 은닉층이 2개인 심층 신경망 Type2가 95.24%, 심층 신경망 Type1은 91.10%로 심층 신경망을 적용하여 추정한 수단선택 모형은 정확도가 90%를 상회하는 수준으로 매우 높은 성능을 보였다. 반면, 다항 로짓 모형은 정확도가 52.72%로 심층 신경망에 비해 낮은 분류 정확도를 보였다. 특히, 복합수단인 버스+전철 수단에 대해 7.84%의 낮은 정확성을 보여 분류가 어려운 복합수단에 대해 매우 취약한 것으로 나타났다. 따라서 모형의 분류 정확성은 심층 신경망 모형이 다항 로짓 모형보다 우수한 것으로 입증됐다.

모형의 직관적인 해석을 통한 일반화 가능성을 판단하기 위해 통행시간과 통행비용의 수요 탄력성을 산정하여 검증했다. 각 모형별 탄력성을 산정한 결과, 모든 모형의 탄력성 부호가 음수로 나타나 통행시간과 통행비용이 증가할수록 해당 수단에 대한 수요가 감소함을 알 수 있어 모

형의 신뢰성을 확인했다. 통행시간과 통행비용의 교통수단별 탄력성 비교결과, 통행비용에 비해 통행시간의 탄력성의 절대값이 높아 통행자가 통행비용보다 통행시간에 더 민감하게 반응함을 알 수 있었다. 모형의 정확성이 높은 심층 신경망 Type3를 기준으로 통행시간의 탄력성을 분석하면, 승용차의 자기 탄력성의 절대값이 대중교통에 비해 작게 나타났다. 이는 승용차를 이용하는 통행자는 통행시간이 조금 증가하더라도 편안함과 빠른 접근이 가능한 승용차 수단을 선호하기 때문으로 판단된다.

통행비용에 따른 탄력성의 절대값은 대중교통에 비해 승용차가 다소 높은 것으로 산정되었다. 이는 승용차를 이용하는 통행자가 대중교통에 비해 통행비용에 민감하게 반응한다는 것을 알 수 있었지만 그 값이 그리 크지 않아 통행자가 통행비용에 크게 의존하지는 않는 것으로 보인다.

본 연구는 딥러닝의 한 종류인 심층 신경망을 이용하여 기존의 수단선택 단계에서 주로 쓰였던 로짓 모형 한계를 극복하고, 심층 신경망 모형이 다양한 통행수단에 대해서 높은 정확도로 분류함을 확인하였다. 특히 분류가 어려운 복합수단에 대해서도 심층 신경망 모형이 다항 로짓에 비해 매우 높은 성능을 나타냄을 확인하였다.

산업의 발전으로 인해 통행자의 통행패턴이 더욱 다양해짐에 따라 기존에 수단선택 모형에서 주로 고려되어온 통행특성 외에 다양한 요소(날씨특성, 정보특성, 실시간 환경특성 등)의 고려는 필수적일 것이다. 따라서 심층 신경망을 이용하여 다양한 변수를 모형에 적용할 필요가 있으며, 심층 신경망 모형의 최적의 은닉층 깊이와 하이퍼 파라미터를 설정한다면 본 연구의 결과처럼 복잡하고 다양한 교통수단에 대해서도 높은 정확도를 갖는 수단선택 모형을 추정할 수 있을 것이다. 따라서 심층 신경망의 도입은 향후 교통계획에서 보다 지능화되고, 신뢰성 있는 교통수요 예측을 위한 하나의 대안이 될 수 있을 것으로 판단된다.

주요어 : 심층 신경망, 교통수단선택모형, 다항 로짓 모형, 딥러닝, 교통계획

학번 : 2017-26819

목 차

| | |
|------------------------------|----|
| 제 1 장 서론 | 1 |
| 제 1 절 연구의 배경 및 목적 | 1 |
| 제 2 절 연구의 범위 | 4 |
| 제 3 절 연구의 절차 | 4 |
| | |
| 제 2 장 이론 및 선행연구의 고찰 | 7 |
| 제 1 절 심층 신경망 | 7 |
| 1. 심층 신경망의 이론적 고찰 | 7 |
| 2. 심층 신경망의 연구 동향 | 12 |
| 제 2 절 교통 수단선택 모형 | 16 |
| 1. 로짓을 이용한 수단선택 모형 | 16 |
| 2. 로짓 이외의 수단선택 모형 | 19 |
| 제 3 절 시사점 | 23 |
| | |
| 제 3 장 연구 방법론의 정립 | 25 |
| 제 1 절 심층 신경망 모형의 정립 | 26 |
| 1. 변수설정 | 26 |
| 2. 심층 신경망 모형의 정립 | 29 |
| 제 2 절 비교 모형 : 다항 로짓 모형 | 33 |
| 제 3 절 모형의 검증 방법 | 34 |
| 1. 모형의 적합성 검증 지표 | 34 |
| 2. 모형의 탄력성 | 36 |

| | |
|---------------------------------------|----|
| 제 4 장 자료의 특성 및 구축 | 38 |
| 제 1 절 자료의 특성 | 38 |
| 제 2 절 자료의 구축 | 41 |
| 1. 교통수단 설정 | 41 |
| 2. 통행자의 사회경제적 특성 | 44 |
| 3. 통행시간 및 통행비용 | 45 |
| 제 5 장 모형의 추정 및 검증 | 50 |
| 제 1 절 수단선택 모형의 추정 | 50 |
| 1. 심층 신경망 모형 추정 결과 | 50 |
| 2. 다항 로짓 모형 추정 결과 | 53 |
| 제 2 절 수단선택 모형의 검증 | 55 |
| 1. 적합성 | 55 |
| 2. 탄력성 | 63 |
| 제 6 장 결론 | 65 |
| 제 1 절 결론 | 65 |
| 제 2 절 한계 및 향후 연구과제 | 68 |
| 참고문헌 | 70 |
| 부 록 | 74 |
| [부록 A] 심층 신경망 유형별 정확도와 오차 변화 | 74 |
| [부록 B] 수단선택 모형별 정오표 | 77 |
| [부록 C] 통행시간과 통행비용에 따른 교차 수요 탄력성 | 79 |
| Abstract | 80 |

표 목 차

| | |
|--|----|
| [표 1] 연구의 범위 | 4 |
| [표 2] 신경망 활성화 함수 | 9 |
| [표 3] AND와 XOR 게이트 | 11 |
| [표 4] 수단선택 모형 추정에 대한 선행 연구의 주요 현황 | 22 |
| [표 5] 국내 수단선택 모형 변수 | 26 |
| [표 6] 변수설정 및 설명 | 28 |
| [표 7] 본 연구의 심층신경망 하이퍼 파라미터 설정 | 32 |
| [표 8] 분류 정오행렬의 예 | 34 |
| [표 9] 단일수단과 복합수단 간 통행시간 설정 방법 | 37 |
| [표 10] 단일수단과 복합수단 간 통행비용 설정 방법 | 37 |
| [표 11] 2010년, 2016년 가구통행실태조사 자료 비교 | 40 |
| [표 12] 2016년 서울시 가구통행실태조사 목적별 통행량 | 40 |
| [표 13] 가구통행실태조사 수단 변경(21개 수단→ 9개 수단) | 42 |
| [표 14] 15개 수단을 주수단 8개(수도권 9개)로 변경 | 43 |
| [표 15] 본 연구의 선택 수단 구축 현황 | 43 |
| [표 16] 통행자의 사회경제적 특성 범주화 | 44 |
| [표 17] 구축자료의 수단별 평균 차외시간 | 46 |
| [표 18] 구축자료의 수단별 통행속도 | 46 |
| [표 19] 속도별 승용차 운영비용(수도권) | 47 |
| [표 20] 유류비 변화(수도권) | 47 |
| [표 21] 물가지수 변화(수도권) | 48 |
| [표 22] 대중교통 요금체계(수도권) | 48 |
| [표 23] 분석자료 기술통계량(표본 수 : 34,013) | 49 |
| [표 24] 심층 신경망 모형의 유형별 오차와 정확도 | 51 |

| | |
|----------------------------------|----|
| [표 25] 다항 로짓 모형 계수 추정 결과 | 54 |
| [표 26] 심층 신경망 Type1의 정오표 | 56 |
| [표 27] 심층 신경망 Type2의 정오표 | 57 |
| [표 28] 심층 신경망 Type3의 정오표 | 57 |
| [표 29] 다항 로짓 모형의 정오표 | 57 |
| [표 30] 심층 신경망 Type1의 검증결과 | 59 |
| [표 31] 심층 신경망 Type2의 검증결과 | 60 |
| [표 32] 심층 신경망 Type3의 검증결과 | 60 |
| [표 33] 다항 로짓 모형의 검증결과 | 60 |
| [표 34] 통행시간에 관한 모형별 수요 탄력성 | 64 |
| [표 35] 통행비용에 관한 모형별 수요 탄력성 | 64 |

그림 목 차

| | |
|---------------------------------------|----|
| [그림 1] 연구흐름도 | 6 |
| [그림 2] 심층 신경망 구성 | 8 |
| [그림 3] AND와 XOR분류 문제 | 11 |
| [그림 4] 심층 신경망 수단선택모형 | 30 |
| [그림 5] 심층 신경망 유형에 따른 정확도 변화 | 52 |
| [그림 6] 심층 신경망 유형에 따른 오차 변화 | 52 |
| [그림 7] 모형별 정확도 비교 | 58 |
| [그림 8] 모형별 검증 지표 비교(수단: 자동차) | 61 |
| [그림 9] 모형별 검증 지표 비교(수단: 버스) | 61 |
| [그림 10] 모형별 검증 지표 비교(수단: 전철) | 62 |
| [그림 11] 모형별 검증 지표 비교(수단: 버스+전철) | 62 |

제 1 장 서론

제 1 절 연구의 배경 및 목적

교통계획에서 교통수요를 예측하기 위해 전통적으로 사용되어 온 방법은 4단계 교통수요추정방법이다. 4단계 교통수요추정방법은 통행발생, 통행분포, 수단분담, 통행배분의 순차적인 단계로 구성되며, 각 단계에서 개별적인 과정을 거친 뒤 예측이 종결된다. 이 방법은 분석가가 각 단계에서 적합한 모형을 선택하여 분석할 수 있고 분석과정에서 생기는 오류를 수정하기에 용이하다는 이점이 있어 오랫동안 사용되어 왔다.

그 중 3단계인 수단분담 단계는 선행단계에서 예측한 통행 발생량을 토대로 통행자 개개인의 통행 특성 행태를 반영하여 교통수단분담을 예측하는 단계이다. 최근 각 지자체의 신도시 개발에 따른 신교통수단(경전철, 트램 등)의 도입으로 통행자가 이용할 교통수단에 대한 예측이 중요해지면서 수단분담 선택 단계의 중요성은 더욱 커지고 있다.

그동안 수단분담선택 단계에서는 주로 로짓 모형(Logit model)¹⁾을 이용하여 수단을 예측하였다. 그 이유는 로짓 모형을 이용하면 비록 교통수요자 개개인의 수단선택에 대한 선호의 다양성을 모두 고려하기는 어렵지만, 적어도 독립변수에 따른 개개인의 교통수단에 대한 상대적인 효용을 산정할 수 있어 직관적으로 이용자의 행태를 판단하기에 용이하기 때문이다. 그러나 일반적인 로짓 모형²⁾은 비관련 대안의 독립성(Independence of Irrelevant Alternatives, IIA)³⁾에 대한 한계를 가지고,

1) 로짓 모형은 계량경제학에서 파생된 이론으로 확률 효용을 최대화하는 대안을 선택하는 원칙에 기초하고 있다 (McFadden, 1973).

2) 대안의 수에 따라 이항로짓모형(Binary Logit Model, BL)과 다항로짓(Multinomial Logit Model, MNL)으로 나뉘며, 기본적으로 개인의 선호다양성을 고려하지 못하는 문제가 있다.

3) 관련이 없는 교통수단 대안의 독립성은 유지된다는 가정으로, 새로운 교통수단의 도

개인의 선호를 모두 반영하기 어려워 추정된 모형의 정확도가 떨어지는 문제를 가지고 있다.

최근 들어 교통수단분담과 관련된 선행연구는 대부분 기존에 쓰이던 일반적인 로짓 모형의 한계를 극복하기 위해 혼합 로짓 모형(Mixed Logit Model), 다항 프로빗 모형(Multinomial Probit Model), 중첩 로짓 모형(Nested Logit Model) 등 다양한 로짓 계열 수단선택모형을 도입한 연구가 진행되며, 개인의 선호다양성에 대한 고려와 모형의 정확도를 높으려는 시도가 계속 되었다. 그러나 로짓 계열 수단선택 모형은 수리통계학적 모형으로 독립변수 간 상관관계가 높은 변수들을 함께 고려하기 어려워 다양한 상황을 반영하는 데 여전히 한계를 가지고 있다⁴⁾.

또한 카풀이나 택시 앱 서비스, 스마트 환승교통 체계로 인해 개인이 선택할 수 있는 교통수단 대안과 패턴은 더욱 다양화되고 있어, 이를 모두 반영하고 높은 성능을 가질 수 있는 수단선택 모형이 필요하다.

따라서 본 연구는 로짓 모형의 한계를 극복하고, 다양한 상황을 반영할 수 있는 수단 선택 모형을 추정하기 위해 데이터 마이닝 모형⁵⁾을 적용하여 통행자의 수단선택을 추정한다. 본 연구에서 제시하는 데이터 마이닝 모형은 최근 4차 산업혁명과 더불어 알파고, 소피아등 인공지능이 대두되면서 활발히 연구가 진행되고 있는 딥러닝⁶⁾을 이용한 방법이다. 인공 신경망을 토대로 발전된 형태인 딥러닝은 기존 로짓 모형이 가지고 있었던 비관련대안에 대한 한계를 극복할 수 있으며, 통계적으로 상관관계가 높은 변수들도 입력 자료로 사용할 수 있어 데이터의 융합과 다양한 변수를 적용하기에 용이하다. 특히, 딥러닝은 은닉층이 많아질수록 다차원의 분류가 가능한 매우 유연한 모형으로 분류 문제에서 매우 좋은

입되더라도 기존에 이용되는 교통수단에 대한 선택확률 비는 일정하다는 가정이다.

4) 정보화로 인해 통행자가 모르는 버스 노선도 효과적으로 이용할 수 있게 되어 통행자의 교통수단 선택 패턴도 더욱 복잡화 되었다 (남대식, 2017).

5) 데이터 마이닝 모형은 기존의 로짓 기반 모형보다 대안과 개개인의 특성사이의 관계를 나타내는데 보다 유연한 구조를 가지고 있으며, 표본 데이터에서 추출한 패턴 특성을 학습하고 식별하는 능력이 뛰어나다 (Xie, 2003).

6) 딥러닝은 인공신경망에 다수의 은닉층을 가지는 심층 신경망(DNN)을 학습하기위해 고안된 알고리즘을 말한다 (Yamasita, 2017).

성능을 보여 음성인식, 통신, 전력수요예측, 공간인식 등 다양한 산업 분야에서 폭 넓게 사용되고 있다. 최근, 교통 분야에서도 버스도착시간, 대기길이, 주행경로, 소통 상황(교통류), 교통사고, 자전거 수요 예측 등 다양한 요소를 고려하고, 예측의 정확성을 높이기 위해 딥러닝을 적용한 연구가 활발히 이뤄지고 있다.

본 연구에서는 수단선택 모형으로 딥러닝의 한 방법인 심층 신경망(Deep Neural Network, DNN)을 적용하여 추정한다. 그리고 기존에 수단선택 모형으로 주로 쓰였던 다항 로짓 모형(MNL)과 비교분석하여 수단선택모형의 정확도를 확인하고, 수단선택 모형으로 딥러닝의 적용 가능성을 판단한다.

본 연구는 개인의 다양한 통행 패턴을 반영하기 위해 수단 선택으로 단일수단뿐만 아니라 복합수단(버스+전철)도 추가하여 분석하며, 통행시간과 통행비용에 따른 탄력성 분석을 통해 각 속성이 대안 선택에 미치는 영향을 직관적으로 확인하여, 추정된 모형의 적합성을 검증한다.

제 2 절 연구의 범위

본 연구의 공간적 범위는 서울특별시로 설정하였으며, 서울시내 424개의 행정동 간 가정기반 출근목적 통행을 대상으로 분석한다.

본 연구의 자료는 통행자의 사회경제적 특성과 통행 수단을 세밀하게 알 수 있는 가구통행실태조사 자료를 이용하여 분석한다. 가구통행실태조사 자료는 한국교통연구원에서 제공하는 원시자료를 재가공하여 활용한다. 본 연구는 가구통행실태조사 자료 중 구득 가능한 가장 최신 자료인 2016년 서울시를 대상으로 한 자료를 활용하기 때문에 본 연구의 시간적 범위는 2016년으로 한정한다.

본 연구에서는 2016년 가구통행실태조사 자료에 기입된 총 21개 유형의 교통수단을 한국교통연구원에서 발행한 「2017 전국 여객 O/D 전수화 및 장래수요예측」 보고서를 기반으로 교통수단을 분리하고, 출근목적 통행자에 한해 승용차, 버스, 지하철, 버스+지하철로 총 4개의 수단으로 구분하여, 이를 추정하는 수단선택 모형을 개발한다.

[표 1] 연구의 범위

| 연구의 범위 | 내용 |
|--------|----------|
| 공간적 범위 | 서울특별시 |
| 시간적 범위 | 2016년 |
| 분석 대상 | 출근목적 통행자 |

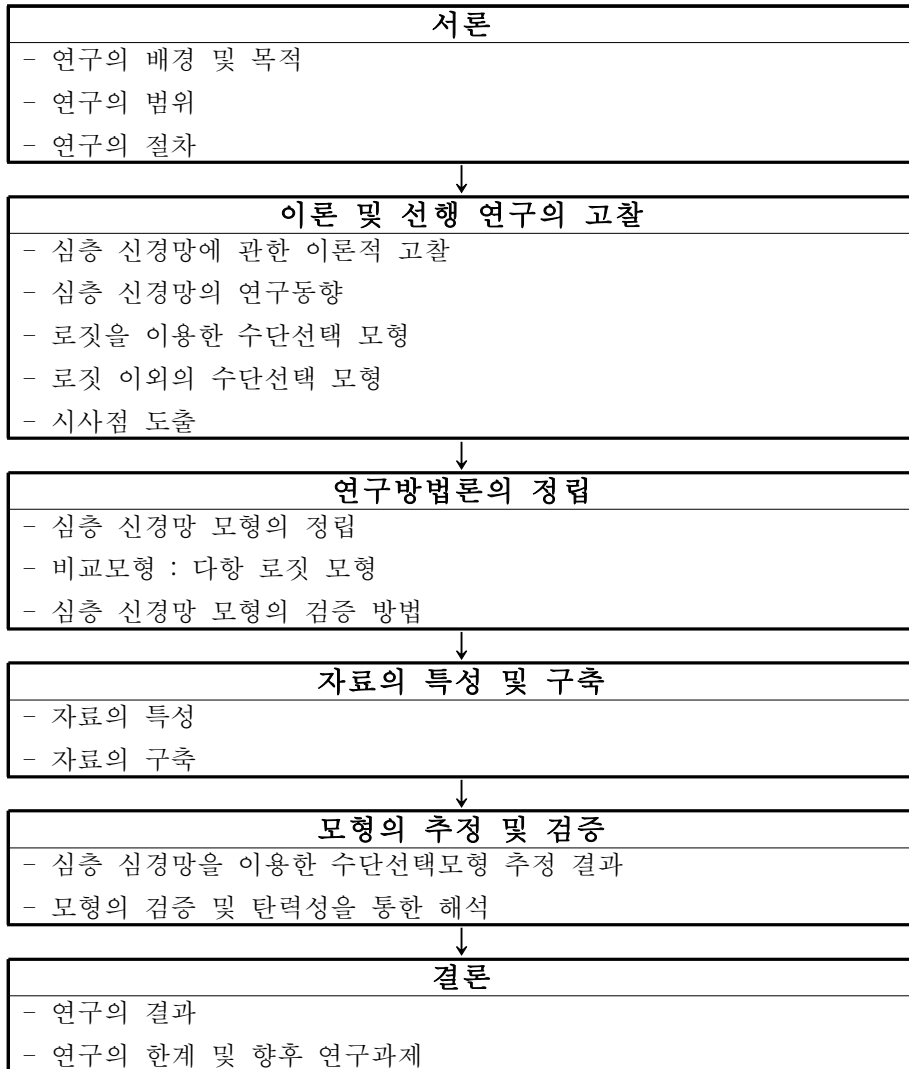
제 3 절 연구의 절차

본 연구는 기존에 수단선택 모형으로 주로 사용됐던 로짓 모형의 한계를 극복하고, 통행자의 복잡한 통행 선택 행태에도 높은 정확도를 가진 수단선택 모형을 추정하기 위해 심층 신경망을 이용 모형을 추정한

다. 심층 신경망 모형과 다항 로짓 모형을 이용한 수단선택 모형의 추정 결과를 비교하여 수단선택 모형으로 심층 신경망 모형의 적합성을 검증 하는데 목적이 있다.

따라서 본 연구는 딥러닝의 한 방법인 심층 신경망에 대한 이론적 개념에 대해 고찰하고, 심층 신경망을 이용한 연구 동향에 대해 살펴본다. 또한 로짓 모형을 이용한 수단선택 모형과 로짓 이외의 모형을 이용하여 수단선택 모형을 추정한 선행 연구들을 살펴봄으로서 본 연구의 시사점을 도출한다.

다음으로 연구방법론의 정립에서는 본 연구에서 제시하는 심층 신경망 모형과 모형의 검증 및 해석 방법을 정립한다. 그리고 본격적인 연구 분석에 앞서 본 연구의 분석 자료인 2016년 가구통행실태조사 자료의 특성을 살펴본 후 자료의 전처리 과정과 구축방법을 자세히 명시한다. 특히 가구통행실태조사 자료는 통행자가 선택하지 않은 비선택 대안에 대한 통행특성 자료를 구축하기 어려운 한계가 있기 때문에 이를 보완하기 위한 방법을 자료구축 단계에서 제시한다. 구축된 자료를 토대로 연구의 방법론에서 정립한 방법대로 심층신경망을 이용한 수단선택 모형과 다항 로짓을 이용한 수단선택 모형을 추정하고, 검증 지표를 통해 모형의 적합성을 판단한다. 이후 통행시간과 통행비용에 따른 수요 탄력성을 통해 블랙박스라고 여겨졌던 심층 신경망 모형을 직관적으로 이해하고 본 연구의 시사에 대해 논한다.



[그림 1] 연구흐름도

제 2 장 이론 및 선행연구의 고찰

제 1 절 심층 신경망(Deep Neural Network, DNN)

1) 심층 신경망의 이론적 고찰

심층 신경망(DNN)은 심층 학습이라 불리는 머신러닝의 한 방법으로, 입력된 데이터에 대한 유형 분류나 회귀를 수행하는 기법이다. 심층 신경망은 입력층(Input layer)과 출력층(Output layer) 사이에 2개 이상의 다중 은닉층(Hidden layer)을 포함하는 인공 신경망(Artificial Neural Network, ANN)을 의미하며, 심층 신경망을 학습하기 위해 고안된 특별한 알고리즘들을 딥러닝(Deep learning)이라 부른다. 깊은 신경망을 학습시킬 수 있는 새로운 알고리즘의 개발과 하드웨어의 성능 향상으로 딥러닝은 다양한 분야에 적용되고 있으며, 분야에 따라 신경망을 학습시키는 알고리즘이 변형되면서 음성인식, 영상인식, 자연어 처리, 예측 등 다양한 분야에서 뛰어난 성능을 나타내고 있다⁷⁾.

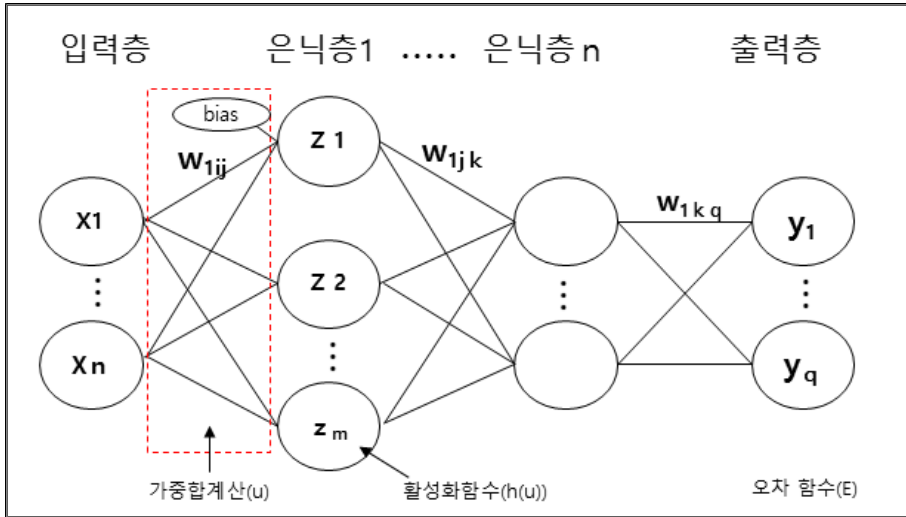
심층 신경망은 적용 알고리즘에 따라 이미지 처리에서 많이 쓰이고 있는 합성곱 신경망(Convolution Neural Network, CNN), 시계열자료의 처리에 주로 쓰이는 순환 신경망(Recurrent Neural Network, RNN) 등 다양한 형태로 발전되어 적용되고 있는 실정이다⁸⁾.

심층 신경망의 적용 원리와 구조는 [그림 2]와 같이 인공신경망과 매우 유사하다. 인공 신경망은 생물학적 신경망을 모방한 것으로 n 개의 입력데이터(X_1, X_2, \dots, X_n)가 들어오면 이에 대한 가중치(w_1, w_2, \dots, w_n)가 각각 곱해지고, 그 가중합이 바이어스(bias)와 더해져 뉴런의 활성화 함수

7) 양준영, and 장준혁. "딥러닝 모델 adaptation 기술의 연구 동향." 한국통신학회지 (정보와통신) 33.9 (2016): 3-7.)

8) Yamasita(2017), 그림과 수식으로 배우는 통통 딥러닝, 제이펍

를 통과해 다음 뉴런으로 전달될 출력이 결정된다.



[그림 2] 심층 신경망 구성

여기서, 하나의 뉴런 특성을 결정하는 것은 활성화 함수이며, 활성화 함수의 종류는 [표 2]와 같이 다양하게 발전하고 있다. 그 중 인공신경망 분석에서 주로 쓰인 함수는 sigmoid와 hyperbolic tangent이다⁹⁾. 그러나 이 두 함수는 두 개 이상의 은닉층을 갖는 심층 신경망에서는 활성화 함수로 사용하기에는 적합하지 않는 문제가 있다. 그 이유는 심층 신경망에서 다수의 은닉층이 존재하게 되면 층과 층을 연결하는 결합 가중치에 대한 오차 수정학습이 불가능하기 때문이다. 이는 오차 역전파 알고리즘의 한계로 출력층에서 오차를 입력층까지 전달하면 각 층의 가중치를 경사 하강법을 이용해 수정할 때, 은닉층의 개수가 늘어날수록, 출력층의 오차가 입력층까지 전달되지 않는 경사 사라짐 문제(Gradient Vanishing)가 있어 신경망의 학습을 원활히 수행할 수 없는 문제를 일으킨다. 오차 수정 학습은 입력 데이터와 정답 신호 그리고 출력 간의 관계를 이용하여 결합 가중치를 수정하지만, 층 단위를 넘어서 오차를 수정할 수는 없기 때문이다¹⁰⁾. 즉, 학습을 위한 과정이 길어질수록, 어느

9) 안홍섭(2016), 심층신경망의 효과적인 초기화를 위한 분할 훈련 및 병합 기법에 관한 연구, 인하대학교 대학원 박사학위 논문

시점에 가중치가 갱신되지 않는 이벤트가 발생한다¹¹⁾. 이 문제를 해결하기 위해 2006년 Hinton에 의해 ReLU(rectified linear unit) 함수가 개발되면서 다수의 layer에서 역전파법 적용이 가능해졌다. 이후 ReLU에서 파생된 활성화 함수인 Leaky ReLU, PReLU, RReLU, 맥스아웃(maxout) 등 심층 신경망 학습을 위해 다양한 활성화 함수가 개발되고 사용되는 추세이다¹²⁾.

[표 2] 신경망 활성화 함수

| Activation Function | Formula |
|---|--|
| 계단함수(step function) | $h(u) = \begin{cases} 1 & (u \geq 0) \\ 0 & (otherwise) \end{cases}, u = \sum_{i=1}^n w_i x_i$ |
| 시그모이드(sigmoid) | $S(u) = \frac{1}{1 + e^{-u}}$ |
| 하이퍼볼릭 탄젠트 (Hyperbolic tangent, tanh) | $\tanh(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}$ |
| 렐루(Rectified linear unit, ReLU) | $R(u) = \max(u, 0) = \begin{cases} u & (u \geq 0) \\ 0 & (otherwise) \end{cases}$ |
| 소프트맥스(softmax) | $p(y^k) = e^{u_{2k}} / \sum_{q=1}^Q (e^{u_{2k}})$ |

신경망 네트워크 중 중요한 단계는 뉴런을 잇는 시냅스간 연결강도를 조정하는 것이다. 이를 위해 신경망은 학습이란 단계를 수행한다. 학습은 신경망의 결합가중치(w)와 바이어스(bias)를 수정하는 과정으로, 학습 분류에 따라 지도학습, 비지도학습, 강화학습 등으로 분류할 수 있다. 학습을 위한 방법으로 대표적으로 퍼셉트론 학습과 델타학습 그리고 이를 발전시킨 오류 역전파 학습(Back propagation)이 있으며, 이는 오차함수를 이용하여 연산된다. 오차함수는 교차 엔트로피 함수나 제곱오차 함수 등

10) Hochreiter S et al(2001) "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," A Field Guide to Dynamical Recurrent Neural Networks. IEEE Press.

11) 이재성. (2016). 심층 신경망의 발전 과정과 이해. 한국통신학회지(정보와통신), 33(10), 40-48.

12) Yamasita(2017), 그림과 수식으로 배우는 통통 딥러닝, 제이펍

문제 유형에 따라 직접 정의한 함수를 사용해도 된다. 역전파법은 오차 함수의 미분을 통해 기울기를 구하고, 그에 따른 결합 가중치의 업데이트 양을 계산하여 가중치를 수정한다. 역전파법에서는 학습의 표본을 노출시키는 방식에 따라 배치학습과 순차적 학습 등으로 나뉘며, 학습 표본의 일부만을 사용해서 업데이트 양을 구하는 방법을 확률적 경사 하강법(stochastic gradient descent, SGD)이라고 한다. 즉, 오차함수인 식(1)을 미분하여 기울기를 구하고, 기울기가 양이면 왼쪽으로, 음이면 오른쪽으로 이동시키며 오차함수가 최소가 되는 계수를 반복하여 결정하는 것을 의미한다. 업데이트 양은 식(2)와 같이 나타낼 수 있다.

$$E^t = \sum_{n=1}^n E_n^t \quad \dots(1)$$

$$w^{t+1} = w^t - \eta \frac{\partial E^t}{\partial w^t} \quad \dots(2)$$

신경망의 마지막 단계인 출력은 문제 유형에 따라 선택한 우도함수(Likelihood function)에 의해 결정되는데, 입력을 그대로 출력하는 항등 함수와 멀티클래스 분류문제에서 주로 사용하는 소프트맥스 함수(Softmax function)로 구성된다. 소프트맥스 함수는 식(3)과 같이 나타낼 수 있으며, 출력층의 모든 유닛값을 누적값으로 나누어 정규화하기 때문에 각 유닛은 0과 1사이의 값을 갖게 된다. 이를 토대로 멀티클래스 분류가 가능하다.

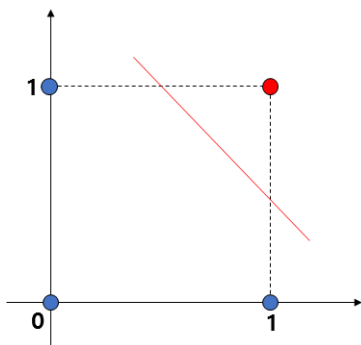
$$P(y^k) = \frac{e^{u_{2k}}}{\sum_{q=1}^Q (e^{u_{2q}})} \quad \dots(3)$$

본 연구에서 적용하고자 하는 심층 신경망은 다층 퍼셉트론(Multi-layer Perceptron, MLP)을 기반으로 한 신경망으로, 기존의 인공 신경망과는 다르게 은닉층이 두개 이상의 다수의 층으로 구성된 신경망이다. MLP는 단층 퍼셉트론과 유사한 구조를 지니고 있지만, 중간층과

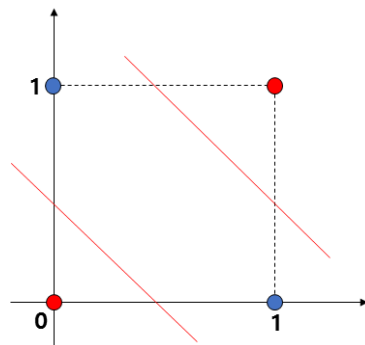
각 node의 입출력 특성을 비선형으로 변환시킴으로서 네트워크의 능력을 향상 시키는 장점이 있다. 예를 들면 은닉층이 하나인 경우는 로짓 모형과 마찬가지로 패턴 공간을 두 구역으로 나누어 AND게이트에 대한 문제는 해결할 수 있지만, XOR게이트는 해결할 수 없었다. 그러나 층이 2층인 경우, 불록한 구역을 형성하여 XOR게이트를 해결할 수 있다. 즉, 다수의 은닉층은 이론상 어떠한 형태의 구역도 형성하여 정교한 분석이 가능하다¹³⁾. 아래 [표 3]는 선형으로 분리할 수 있는 AND게이트 문제와 선형분리가 어려운 XOR게이트를 문제를 나타내며, [그림3]¹⁴⁾을 통해 XOR 게이트 문제를 해결하기 위해 여러 개의 퍼셉트론을 계층적인 구조를 갖도록 조합해야함을 알 수 있다.

[표3] AND와 XOR 게이트

| AND 게이트 | | | XOR 게이트 | | |
|---------|----|---|---------|----|---|
| X1 | X2 | y | X1 | X2 | y |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 | 1 |
| 0 | 1 | 0 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 0 |



(a) 하나의 양의 표본과 나머지를 나눔



(b) 두 개의 양의 표본과 나머지를 나눔

[그림 3] AND와 XOR분류 문제

13) Saito Goki(2017), 밑바닥부터 배우는 딥러닝, 한빛미디어

14) Yamasita(2017), 그림과 수식으로 배우는 통통 딥러닝, 제이펍

심층 신경망은 비선형 모형으로 매우 유연한 다차원 모형이지만 회귀 분석이나 의사결정나무와 같이 결과를 해석하기 유용한 모형은 아니다. 즉, 입력된 변수의 가중치에 대한 완벽한 해석은 불가능하며, 종속변수에 대한 결정요인을 분석하기 어렵다. 그러나 최근 2016년 Ribeiro에 의해 LIME(Local Interpretable Model-agnostic Explanations)¹⁵⁾이라는 해석 방법이 등장하고 있으며, 해석의 어려움에도 불구하고 다양한 머신러닝 기법 중에서 예측과 분류문제에 있어 매우 성능이 높은 것으로 입증되어 많은 연구자에게 선호되고 있는 실정이다.

2) 심층 신경망의 연구 동향

심층 신경망은 과거 인공신경망 모형을 기반으로 진화된 모형이다. 과거 1990년대 인공신경망 모형을 적용한 연구들이 발표되었으나, 통계학적 예측방법에 비해 효과가 월등히 뛰어나지 않았으며¹⁶⁾, 학습시간이 길고, 결과에 대한 해석의 문제가 제기되어 인공신경망의 활용이 떨어졌다. 그러나 2000년대에 들어 4차 산업혁명에 의한 빅데이터의 대두, 다양한 알고리즘의 발전과 컴퓨터의 성능향상으로 과거 인공신경망이 지녔던 문제들이 해결되며 신경망을 활용한 연구들이 다양한 분야에서 적용되며 계속 고도화되어, 딥러닝의 확산을 이끌고 있다. 특히, 은닉층의 수가 많은 심층 신경망은 분류와 회귀문제에 모두 적용될 수 있으며, 최근 로봇, 음성인식, 통신, 전력수요예측 등 매우 다양한 영역에서 적용되고 있다.

신진호 (2018)는 로봇의 인간 감정인식을 위한 분류문제를 해결하기

15) LIME은 모형이 어떠한 방식으로 작동하는 지 해석하고자 하는 방식으로 예측값의 근방에서 모형이 어떻게 작동하는지 분석한다. 즉, 입력값을 조금씩 다양한 방식으로 변화(perturbation)시키며, 모형의 예측값 변화를 통해 중요변수를 설명하는 방법이다.

16) Nijkamp et al. (1996), "Modelling inter-urban transport flows in Italy: A comparison between neural network analysis and logit analysis", Transportation research. Part C, Emerging technologies, 4(6), 323-338.

위해 심층 신경망을 적용하였으며, 정석환 (2018)은 오디오 신호 분류를 위해 전통적으로 사용되어 온 SVM(Support Vector Machine)대신 우수한 성능으로 각광받고 있는 심층 신경망을 적용한 연구를 진행하기도 했다. 또한 전력수요예측¹⁷⁾, 특히 수명 예측¹⁸⁾, 프로야구 일일 관중 수 예측¹⁹⁾ 등 미래현상에 대한 이해와 추정을 위해서도 널리 적용되고 있다. 특히 최근 교통 분야에서도 심층 신경망을 활용하여 통행과 소통상황을 예측하는 모형을 추정하는 연구가 진행되고 있다.

Wu (2018)는 심층 신경망을 이용하여 교통류를 예측하는 정확도를 개선하는 연구를 수행하여, 데이터의 양이 많을수록 성능이 높아지는 심층 신경망의 도입은 실시간으로 대용량의 자료가 수집되고 있는 현대에 매우 적합한 방식임을 주장했다.

Liu (2017)는 대중 교통 승객의 수요를 예측하는 모형을 은닉층이 3개인 심층 신경망을 이용하여 추정하였다. 실시간 승객의 흐름을 반영하여 SAE(Stacked Auto Encoders)와 심층 신경망을 이용한 하이브리드 모형으로 4개의 BRT역에 대해 승객의 흐름을 예측한 결과 보다 정확하고 보편적인 승객의 이동류를 예측할 수 있었다.

Treethidtapthath (2017)은 대중교통(버스)의 GPS 데이터를 이용하여 버스의 도착시간을 예측하는 모형을 추정하였으며, 추정 모형은 은닉층이 4개인 심층 신경망을 적용하여 분석하였다. 태국 방콕의 BMTA-8 버스를 대상으로 선형회귀 모형과 비교한 결과 버스 도착시간 예측 모형으로 심층신경망 모형이 높은 성능을 나타내었다.

Yi (2017)는 내비게이션을 장착한 차량을 통해 수집된 데이터를 이용하여 교통 소통 상황을 예측하는 모형을 개발하였으며, 교통 소통상황은 교통성능지수를 기준으로 소통과 원활의 2가지 범주에 대해 분별했다. 해당 연구에서 사용된 모형은 은닉층이 3개인 심층 신경망 모형이며, 적

17) 김경환, et al. "인공 지능 기법을 이용한 전력수요예측 동향." 대한전기학회 학술대회 논문집 (2017): 457-458

18) 정병기 et al. "딥 뉴럴 네트워크 기반의 특허수명 예측모형 개발." 대한산업공학회 추계학술대회 논문집 (2016): 981-985.)

19) 박동주, et al. "Deep Neural Network 기반 프로야구 일일 관중 수 예측: 광주-기아 챔피언스 필드를 중심으로." 스마트미디어저널 7 (2018): 1-8.

용 결과 99%의 정확도를 보였다.

이용주 (2018)는 심층 신경망(DNN)을 이용하여 차량의 대기길이 추정하는 모형을 개발했다. 기존의 차량대기길이 관련 선행 연구는 주로 통계학적 접근에 따른 확률분포모형을 개발하거나 검지기 데이터와 차량 궤적데이터를 이용한 충격파를 적용하는 연구가 주를 이루었으나 해당 연구는 기존 보다 차량대기길이의 정확도를 높여 교통운영기법에 적용하기 위해 심층 신경망을 도입하여 분석하였다. 자료는 VISSIM을 이용한 시뮬레이션 자료를 이용하였으며, 차량대기길이가 링크의 길이를 초과하는지의 여부에 따라 링크 미초과는 은닉층의 개수를 5, 10개, 링크 초과는 은닉층을 3, 5개로 나누어 분석했다. 모형 검증 결과, 검지기 데이터 기반 차량대기길이 산출법보다 심층 신경망을 이용한 산출방법이 높은 성능을 나타내었고, 특히 차량대기길이 링크 초과여부 분류모형은 98%의 정확도를 보였다.

박상민 (2018)은 심층 신경망을 이용하여 교통사고 건수를 예측하는 모형을 추정했다. 모형의 은닉층은 총 3개로 구성했다. 심층 신경망 모형을 교통사고건수 예측에 주로 사용되는 음이항 회귀모형과 비교한 결과, 심층 신경망을 이용한 모형이 다소 우수한 것으로 나타났다. 특히 사고 건수가 적은 구간에서의 예측성능이 높은 것으로 확인되었다.

정동기 (2018)는 심층 신경망을 이용하여 차량의 주행경로를 학습하고, 주행 경로에 대한 예측의 가능성을 확인하였다. V2I 통신으로 수집되는 차량의 주행정보를 이용하여 차량의 주행 상태 정보를 파악하고, 자율주행정보 획득의 가능성을 제시하였다.

철도기술연구원은 딥러닝을 이용하여 ‘트립스’²⁰⁾라는 대중교통운영관리시스템을 구축하여 더 많은 데이터의 융합과 학습을 통해 이용자의 다양한 선호와 행태를 반영한 대중교통수요를 예측하려는 시도를 하고 있다.

이처럼 딥러닝은 통계적으로 상관관계가 높은 변수들도 입력 자료로

20) 철도기술연구원에서 대중교통 빅데이터인 교통카드데이터를 이용한 첨단대중교통관리시스템 구축한 사업이다.

사용할 수 있다는 장점이 있어²¹⁾, 다양한 데이터의 융합과 많은 변수를 고려해야하는 현시대에 적합한 모형으로 추구된다. 또한 잘 학습된 딥러닝은 현실에 대한 재현율과 미래에 대한 예측에서도 높은 성능을 보이는 것으로 알려져 다양한 분야에서 적용되고 있다.

본 연구에서는 딥러닝의 한 종류인 심층 신경망(Deep Neural Network, DNN)을 기반으로 학습시키는 알고리즘을 선택하여 다양한 요소를 고려하고 복잡한 통행패턴에도 높은 정확성을 가진 교통수단 선택 모형을 추정한다. 교통수단 선택 모형 추정을 위해 고려되어야할 변수와 분석 방법, 그리고 기존에 사용되었던 로짓을 이용한 교통수단선택 모형과 로짓 이외의 수단선택모형에 대한 선행연구는 다음 절에서 살펴본다.

21) 남대식, 서원호 (2017) 'AI를 만나면 교통은 어떻게 똑똑해질까?', 「카카오 AI 리포트」 Vol.9 : 8-12

제 2 절 교통 수단선택 모형

1) 로짓을 이용한 수단선택 모형

교통수단 분담 단계에서 전통적으로 사용되어온 모형은 로짓 모형이다. 계량경제학에서 파생된 로짓 모형은 개개인이 교통수단에 대한 상대적인 효용을 바탕으로 각 대안에 대한 선택확률을 산정한 후 수단선택을 결정하기 때문에 오랫동안 사용되어 왔다. 최근에 개인의 선호다양성을 고려한 연구의 필요성이 제기되고, 정교한 수단선택 모형 추정의 요구와 함께 일반적인 로짓 모형에서 진화된 형태인 다양한 로짓 계열 수단선택 모형의 적합성을 비교하는 연구가 이루어지고 있다.

전명진 (1997)은 서울시 통근 통행수단을 중심으로 토지이용패턴과 통행수단 선택과의 관계를 분석했다. 해당 연구는 도시형태와 토지이용패턴이 수단선택에 미치는 영향을 분석하기 위해 도시의 주거 밀도와 직장 밀도를 변수로 사용하였고, 다항 로짓 모형(MNL)을 이용하여 교통수단 선택모형을 추정했다. 모형은 출발지 기준과 도착지 기준으로 총 2개의 모형을 개발하여 분석하였으며, 분석 결과, 밀도가 높을수록 버스와 전철 등 대중교통수단 이용률이 높았으며, 특히 직장밀도가 높을수록 버스를 선호할 확률이 철도를 선호할 확률보다 높게 나타났다. 고밀도 지역의 경우, 수송 부담율이 수단별로 크게 차이가 나지 않음을 알 수 있었다. 그리고 출발지 기준에 따른 도보 대 전철간 통행수단 선택은 주거 밀도와는 상관이 적은 반면 도착지 기준에서는 직장밀도에 따라 전철을 선택할 확률이 높은 것으로 분석되어 직주분리현상에 대한 분석이 필요함을 알 수 있었다.

김성희 외(2001)는 주거지와 대중교통과의 접근시간과 접근거리가 수단선택에 미치는 영향을 분석하였다. 이전까지의 선행연구가 통행시간과 통행비용을 주요 설명변수로 간주하고 분석을 진행한데 반해 해당연구는 통행수단 선택에 관련된 대부분의 연구가 개인의 효용극대화에 입각한

통행자 수단선택 이론에 기초를 두고 있다고 주장하며, 보다 정교한 수단선택 모형을 위해서는 세분화된 설명변수를 설정할 필요가 있음을 제시하였다. 가구와 개인특성과 관련된 변수를 세분화하여 분석하였지만 승용차와 대중교통의 수단선택에 있어 상대적인 비용과 시간이 여전히 수단선택에 큰 영향을 미치는 것으로 판단되었다. 특히 대중교통 선택 확률의 경우 접근거리와 환승시간이 감소할수록 높아짐을 알 수 있었다. 이 연구를 통해 보다 정교한 수단선택 모형을 위해 주수단의 통행시간과 통행거리 뿐만 아니라 수단으로의 접근성에 대한 변수가 고려될 필요가 있음을 알 수 있었다.

성현곤 외(2014)는 2010년 장거리 여객통행 실태조사 자료를 이용하여, 장거리 통행자를 대상으로 승용차와 고속철도를 수단으로 선택하는 결정요인에 대해 분석했다. 해당 연구는 도시특성에 대한 수단 선택을 반영하기 위해 인구 밀집도와 종사자 밀집도를 설명변수로 추가하여 이항 로짓을 이용하여 수단선택 모형을 추정했다. 분석결과, 통행거리가 길고, 오전에 출발, 그리고 숙박을 하지 않은 통행일수록 고속철도를 이용할 확률이 통계적으로 높았으며, 이용자 특성에서는 여성과 20대 혹은 60대에 가까울수록 고속철도를 이용할 확률이 높게 나타났다. 해당 연구에서 설정한 도시특성 변수의 경우, 인구밀도가 높은 도시에서 고속철도를 이용할 확률이 통계적으로 높음을 알 수 있었으며, 고용밀도는 여가 통행을 제외하고는 통계적으로 유의하지 않음이 분석됐다. 해당 연구는 장거리 여객통행자의 수단선택 확률을 이용하여 고속철도로의 수단전환 전략을 수립하는데 유용할 것으로 판단했다.

김희진 외(2015)는 수도권과 5대 광역시(대전, 대구, 광주, 부산, 울산)의 지역 간 통행을 대상으로 수단선택모형을 개발하였다. 다항 로짓 모형과 네스티드 로짓 모형을 적용하여 분석한 결과, 네스티드 로짓 모형의 설명력이 가장 높은 것으로 분석되었다. 이 연구는 단거리와 장거리로 구분하는 거리특성변수를 적용하여 수단선택모형을 정산한다. 즉, 통행거리에 따른 군집화를 이용하여 로짓 모형에 적용하였다. 분석결과, 거리특성변수(145km)를 적용한 네스티드 로짓 모형의 설명력이 가장 높은

것으로 분석되었다. 해당 연구는 단거리와 장거리 통행 간 효용 차이를 반영하여 보다 현실을 잘 묘사할 수 있는 모형을 구해낸 데 의의가 있지만, 접근성에 대한 변수는 고려되지 않은 한계가 있었다.

조신형 외(2017)의 연구는 일반철도와 고속철도의 수단 선택 모형을 추정했다. 통행자가 인지하고 있는 철도 수단에 대한 시간과 비용 개선에 따른 수단 전환의 특성을 분석하기 위해 잠재선호조사를 통한 자료를 이용하여 분석하였으며, 일반적인 로짓 모형이 개인의 선호다양성을 고려할 수 없다는 한계를 극복하기 위해 혼합 로짓 모형을 이용하여 분석했다. 대중교통이용자 510여명을 대상으로 진행한 연구결과에 따르면 통행목적별 통행시간 가치는 업무통행이 가장 높게 도출되었으며, 다항 로짓 모형의 경우, 7,953원/시로 도출된 반면 개인 선호다양성을 반영한 혼합 로짓 모형은 약 6,819원/시로 결과가 나왔다. 해당 연구에서 모형의 적합도는 정규분포를 가정한 혼합 로짓 모형이 가장 우수한 모형으로 분석되었으며 수단간 이질성을 반영할 수 있다고 판단했다.

김민주 (2018)는 로짓 계열 수단선택모형의 비교에 관한 연구를 진행했다. 해당 연구에서는 선호다양성을 고려한 수단선택모형을 비교하기 위해 다항 로짓 모형, 혼합 로짓 모형 그리고 잠재계층모형을 이용하여 수단선택 모형의 적합성에 대해 검증하였다. 가구통행실태자료를 이용하여 서울시 통근자를 대상으로 수행하였으며, 모형의 적합성을 확인한 결과 잠재계층모형이 가장 설명력이 높은 것으로 추정되었다. 해당연구는 개인의 선호다양성을 나타내기 위한 로짓 계열의 수단선택 모형을 비교하며 현재 교통수요분석에 적용되고 있는 로짓 모형의 개편을 주장하였다.

로짓 모형을 이용한 수단선택 모형과 관련한 선행연구의 고찰을 통해 모형의 정교화를 위해 개인 혹은 지역(토지)특성을 반영한 다양한 설명 변수를 고려하는 시도가 있음을 알 수 있었으며, 특히 진화된 형태의 로짓계열 수단선택 모형을 이용하여 모형의 적합성을 높이려는 시도가 이루어짐을 알 수 있었다.

2) 로짓 이외의 수단선택 모형

선행연구 고찰에서 살펴보았듯이 로짓 모형은 각 대안의 선택 확률을 통한 선택의 용이성과 추정된 계수를 통해 설명변수와 대안선택간의 관계를 설명할 수 있는 이점을 가져 오랫동안 사용되어 왔다. 그러나 로짓 모형은 비관련대안으로부터 독립성 문제를 가지고 있다. 즉, 하나의 대안이 다른 대안에 대한 선택확률로 선택 가능한 제 3의 대안의 존재여부에 영향을 받지 않는다는 것이다²²⁾. 이러한 모형의 문제점을 해결하기 위해 데이터 마이닝 기법에서 고안한 다양한 모형을 이용해 개개인의 특성을 반영할 수 있는 수단선택모형 추정 연구가 이루어졌다.

Nijkamp 외(1996)의 연구는 교통수단선택 모형에 인공신경망의 개념을 도입한 초기 연구이다. 해당 연구는 이탈리아의 철도와 도시대중교통 수단 선택에 대한 교통수단 선택 모형을 추정하기 위해 인공신경망과 로짓 모형을 비교분석했다. 수단선택에 영향을 미치는 설명변수는 각 수단에 따른 통행시간, 통행거리, 통행비용으로 설정하여 분석하였으며, 인공신경망 모형 구성을 위해 은닉층(hidden layer)의 뉴런 수에 따라 3가지 유형으로 분류²³⁾하고, 이항 로짓 모형의 결과와 비교했다. 분석결과, 4가지 인공신경망 모형 중 한 개의 모형만 로짓 모형에 비해 좋은 성능을 나타냈으며, 수단선택 모형에서 인공신경망을 적용하는 데 어려움과 적용가능성을 동시에 제시했다.

Rao 외(1998)는 수단선택 모형을 위한 효용 극대화의 개념을 기반으로 개발된 이산 선택 모델은 개인의 이동 행동을 모형화 하는데 널리 이용되지만, 이러한 모델의 예측 능력은 이용 가능한 대안들의 수가 증가함에 따라 급격하게 감소함을 문제로 제기했다. 문제의 해결을 위해 인공신경망의 도입을 주장하며 인공신경망이 많은 수의 설명변수로 이용자

22) 강호논문연구소, <https://blog.naver.com/edustudy21/220745514103>

23) 인공신경망 모형은 초기에 분석가 설정해야하는 하이퍼 파라미터에 따라 분석 결과가 매우 상이한 결과를 나타낸다. 보통 신경망에서 은닉층의 개수와 뉴런의 개수, 활성화함수, 오차함수의 설정은 분석가의 시행착오로 인해 설정되고 있으며, 최근에는 하이퍼 파라미터 최적화에 대한 연구가 활발히 진행되고 있다.

의 수단 선택 행동을 모형화 하는데 매우 유용할 수 있는 하나의 새로운 도구임을 주장하였다. 해당 연구는 뭍바이에서 대중교통(철도) 접근에 대한 수단 선택 분석을 인공신경망을 이용하여 추정했으며, 모형의 적합성 검증을 위해 다항 로짓 모형(MNL)과 비교 분석하였다. 분석결과, 다항 로짓 모형은 총 2,663개의 수단선택 중 482개를 오측한데 반해 인공신경망은 단지 13개의 오측 결과가 나타났다. 또한 해당연구는 블랙박스라고 여겨졌던 인공신경망 모형의 해석을 위해 수요탄력성의 개념을 도입하여 분석하였고, 다항 로짓 모형의 경우 변수의 범주와는 상관없이 일정한 탄력성을 가지는데 반해 인공신경망은 입력값의 범주에 따라 탄력성이 변화됨을 나타내었다. 해당연구는 수단선택모형에서 인공신경망 도입의 필요성과 기존에 인공신경망의 해석문제를 수요 탄력성 추정을 통해 극복한 연구이다.

Xie 외(2003)는 기존의 수단선택모형은 경제학 이론에 근거한 확률 효용 극대화를 이용한 방법을 선택하고 있는 것에 대한 한계를 언급하며 다양한 인간 행동패턴을 반영한 수단선택모형에 대한 고려가 필요함을 제기했다. 이를 위해 패턴인식 자료 분석 방법인 의사결정나무와 인공신경망의 도입을 제안하며 다항 로짓 모형(MNL)과 데이터 마이닝 기법을 통한 모형을 비교하여 적합성을 판단했다. 자료는 San Francisco Bay Area Travel Survey (BATS) 2000 자료를 사용하였으며, 분석결과 의사결정나무(76.8%)와 인공신경망(78.2%) 모형이 다항 로짓 모형(72.9%)보다 우수함을 입증했다. 해당 연구는 다수의 은닉층을 통한 모형 고도화의 필요성을 제기하였으며, 인공신경 모형의 우수한 성능을 입증함과 동시에 모형 해석의 어려움 대한 부분을 한계로 언급하였다.

Cantarella (2005)는 이탈리아의 통근 통학목적 통행 자료를 이용하여 다층퍼셉트론(MLP)을 적용하여 수단선택모형을 추정하였다. 독립변수로 이용자특성, 통행특성, 기상특성을 설정하였으며, 모형의 은닉층 수는 1~2개로 설정하였다. 해당 연구는 하이퍼 파라미터 변화에 따른 모형의 적합성을 비교하고 판단하였다. 다양한 데이터마이닝 모형과 심층 신경망 모형의 추정결과를 비교한 결과, 심층 신경망 모형이 효용함수를 이

용한 확률효용모형보다 나은 성능을 보였다.

Tortum 외(2009)는 인공신경망을 이용하여 도시간 화물운송에서 교통 수단선택 모형을 개발했다. 이 연구는 철도와 도로로 나누어 각 수단에 따라 거리, 시간, 비용을 설명변수로 설정하고 다중회귀분석모형, 로짓 모형, 인공신경망, 퍼지 모형(Neuro-fuzzy)을 이용한 수단선택모형을 추정하고 비교분석했다. 분석결과, 인공신경망과 퍼지모형이 도시간 화물교통 수단의 비선형적인 선택을 분류하는데 매우 효과적인 것으로 분석했다.

정진욱, 정진혁(2006)은 기존 로짓 모형이외의 데이터 마이닝 기법에서 사용되는 베이지안 망, 의사결정나무, 신경망 알고리즘을 이용하여 수단분담 모형을 추정했다. 해당연구는 수단을 승용차와 버스로 나누어 이진분류를 진행하였으며, 자료는 TransCAD에서 구득 가능한 가상 자료를 이용하여 분석했다. 해당 연구에서 구성한 인공신경망 모형은 은닉층이 단일층인 모형으로 활성화함수는 하이퍼볼릭 탄젠트함수를 이용하여 분석하였다. 분석결과, 인공신경망 모형이 81.89%로 가장 높았으며, 다음은 의사결정나무가 81.78%, 이항로짓모형이 80.75%, 베이지안 망이 75.97%의 예측률을 가지는 것으로 분석되었다. 해당 연구는 적용 데이터가 실증데이터가 아닌 점에서 연구의 실용가능성에 대해서는 의문을 제기할 수 있지만, 이 연구를 통해 분류문제에서 인공신경망의 높은 성능에 대해 확인할 수 있었다.

위 선행연구에서 살펴보았듯이 수단선택 단계에서 주로 사용되었던 로짓 모형의 한계를 극복하려 다양한 접근법을 제시하는 연구가 이루어졌다. 특히 기존 로짓 모형이 가지고 있는 다수의 대안에 대한 분석의 정확도 문제를 제기하며 인공신경망을 이용한 수단선택 모형의 새로운 접근법을 제시하는 연구가 다수 있었으며, 인공신경망 모형의 정확도가 로짓 모형에 비해 매우 우수함을 나타내었다. 하지만 몇몇의 연구는 인공신경망 모형의 해석이 로짓 모형에 비해 어렵다는 점을 한계로 언급했으나, 한 연구는 인공신경망 모형을 통해 추정된 결과를 수요탄력성을 이용해 설명변수가 종속변수에 미치는 영향을 분석한 연구를 진행하여 해석의 한계를 극복하려 시도한 것을 확인할 수 있었다.

[표 4] 수단선택 모형 추정에 대한 선행 연구의 주요 현황

| 구분 | 연구 | 모형 | 설명변수 | 종속변수 |
|-----------|-------------------|--|--|---------------------------|
| 로짓 모형 | 김민주(2018) | MNL, MXL, LCM | 이용자특성, 통행비용, 통행시간 | 승용승합, 승용동승, 시내버스, 지하철 |
| | 조신형(2017) | MNL, MXL | 이용자특성, 통행비용, 통행시간 | 일반철도, 고속철도 |
| | 김희진(2015) | MNL, NL | 이용자특성, 통행시간, 통행비용 | 승용차, 버스, 일반철도, 고속철도 |
| | 성현곤(2014) | BL | 이용자특성, 통행특성, 도시특성 | 승용차, 고속철도 |
| | 김성희(2001) | BL | 이용자특성, 통행비용, 통행시간, 접근거리, 대기시간, 이용가능한 노선수 | 승용차, 대중교통 |
| | 전명진(1997) | MNL | 통행특성, 도시특성(직장밀도, 인구밀도) | 전철, 버스, 자가용, 도보 및 자전거, 기타 |
| 로짓 이외의 모형 | 정진욱(2006) | BN, DT, ANN | 이용자특성, 통행비용, 통행시간 | 승용차, 버스 |
| | Tortum(2009) | MNL, Fuzzy, ANN | 통행비용, 통행시간, 통행거리 | 도로, 철도 |
| | Cantarella (2005) | MLP, MNL, NB, SVM, ANN, BOOST, RF, BAG | 이용자특성, 통행특성, 기상특성 | 승용차, 대중교통, 이륜차, 도보 |
| | Xie(2003) | MNL, DT, ANN | 이용자 특성, 통행시간, 통행비용 | 승용차, 카풀, 대중교통, 자전거, 도보 |
| | Rao(1998) | MNL, ANN | 이용자특성, 통행시간, 통행비용, 대기시간 | 도보, 버스, 이륜차, 승용차, 택시 |
| | Nijkamp(1996) | BL, ANN | 탑승 승객수, 통행시간, 통행비용 | 철도, 승용차 |

주 : MNL(Multinomial Logit Model), MXL(Mixed Logit model), LCM(Latent Class Model), BL(Binary Logit Model), DT(Decision Tree), ANN(Artificial Neural Network), MLP(Multi-Layer Perceptron), NB(Naive Bayes), SVM(Support Vector Machine), BOOST,BAG, RF : ensemble model

제 3 절 시사점

교통수단 선택모형에서 결과해석의 용이함으로 전통적으로 로짓 모형이 주로 사용되어 왔다. 그러나 일반적인 로짓 모형은 비관련대안의 독립성 문제와 통행자의 다양한 선호를 반영하기 어려운 문제가 있어 이를 보완하기 위한 연구가 다양하게 이루어졌다. 연구는 크게 기존의 로짓 모형을 발전시킨 형태인 네스티드 로짓이나 혼합 로짓 모형 등을 적용하는 방법과 로짓 이외의 모형(의사결정나무, 신경망, 베이지안망 등)을 적용하는 방법으로 나눌 수 있었으며, 로짓 모형의 한계를 극복하고, 수단 선택 모형의 정확성을 높이기 위한 시도가 계속되고 있음을 선행연구를 통해 확인할 수 있었다. 그러나 선행연구에서 제시된 방법인 로짓이나 베이지안망 등과 같은 기계학습 모형은 데이터 희소성(Data Sparsity)문제를 가지고 있어, 학습 데이터가 편중되거나 충분하지 않은 경우, 학습되지 않은 데이터의 정보에 대해서는 분류가 상당히 취약한 문제가 존재하며²⁴⁾, 로짓 모형의 경우, 선형분류 기법이기 때문에 XOR문제와 같은 복잡한 문제에 대해 분류가 취약한 한계를 가지고 있다. 또한 상관관계가 높은 변수들을 함께 고려하기 어려워 통행자의 다양한 행태를 반영하는데 여전히 한계를 가지고 있음을 알 수 있었다. 더욱이 산업의 발전으로 스마트 모빌리티(Smart Mobility), 트램(Tram), 카셰어링(Carpool) 등의 서비스가 활성화되며 교통수단은 나날이 복잡해지고 있어 수단 선택 모형의 정확도를 향상시키는 것은 점점 어려워지고 있는 실정이다.

이에 기존에 일반적으로 사용되어온 로짓 모형이 다양한 형태의 교통수단(복합수단)을 높은 성능으로 예측(분류)할 수 있는지를 확인하고, 선행연구에서 제기한 수단선택 모형의 한계를 보완할 새로운 수단선택 모형의 도입이 필요하다.

따라서 본 연구에서는 상관관계가 높은 변수도 입력 자료로 사용할 수 있어, 다양한 요소를 고려할 수 있으며 최근에 많은 연구에서 높은

24) Lai, S., Xu, L., Liu, K., and Zhao, J. (2015), "Recurrent Convolutional Neural Networks for Text Classification.", In AAAI (Vol. 333, pp. 2267-2273).

분류 정확도를 가지는 것으로 판단된 심층 신경망을 이용하여 기존의 수단선택 모형의 한계를 보완하고, 더욱 다양해진 교통수단에 대응 가능한 수단선택 모형을 제안한다.

본 연구는 심층 신경망을 이용한 수단선택 모형의 적합성을 검증하기 위해 기존에 수단선택모형에서 일반적으로 사용되어온 다항 로짓 모형과 비교하여 적합성을 검증한다. 또한 모형의 해석을 위해 계량 경제학에서 사용하는 탄력성을 도입하여, 설명변수가 대안 선택에 미치는 영향을 분석한다. 본 연구에서 추정하는 교통수단은 단일수단과 복합수단을 포함한 모형으로 승용차, 버스, 전철, 버스+전철 총 4가지 대안에 대해 추정한다. 본 연구는 가구통행실태조사 자료를 이용하여 이용자의 특성을 반영하고, 복합 교통수단(버스+지하철)을 추정하는 모형을 제시하여 다양해진 교통수단에 대응하는 모형을 추정한다는 점에서 다른 연구와 차별성을 가진다. 또한 오랫동안 수단선택 모형으로 사용된 로짓 모형의 한계를 보완하고, 대체가능한 모형을 제시한다는 점에서 의의 있는 연구이다.

제 3 장 연구 방법론의 정립

본 연구의 심층 신경망 수단 선택 모형은 통행자의 사회경제적 특성과 통행특성을 반영할 수 있는 2016년 가구통행실태조사 원시자료를 이용하여 서울특별시 가정기반 출근목적 통행을 대상으로 분석한다.

본 연구에서 수단선택 모형으로 제안하는 심층 신경망²⁵⁾은 모형의 복잡도에 따라 정확성과 일반화 가능성이 차이가 나기 때문에 본 연구에서는 은닉층의 개수를 기준으로 은닉층이 1개일 때, 2개일 때, 3개일 때로 총 3가지 유형으로 구분하여 분석하고 각 모형을 비교한다. 또한 교통수단선택 모형으로 심층 신경망의 적용 적합성을 검증하기 위해 기존에 교통계획의 수단선택 단계에서 일반적으로 사용되고 있는 다항 로짓 모형(Multinomial Logit Model)을 본 연구의 제안 모형과 비교하여 검증한다.

본 장의 제 1절에서는 수단선택 모형을 구축하기 위한 입력 자료인 설명변수를 정의한다. 그리고 본 연구에서 제안하는 심층 신경망 모형의 (하이퍼)파라미터²⁶⁾를 설정하여 수단선택모형으로서 심층 신경망의 구조를 정립한다. 제 2절에서는 심층 신경망 모형의 비교대상인 다항 로짓 모형의 구조를 정립한다. 마지막 제 3절에서는 모형의 검증을 위한 검증지표를 설정하고, 모형을 직관적으로 이해하기 위해 설명변수의 변화에 따른 모형의 수요예측 변화를 계량경제학의 수요 탄력성의 개념을 적용하여 확인한다.

25) 다층 신경망(Multi-layer Neural Network)은 여러 개의 단층 신경망을 쌓아서 만들어진 구조로 되어 있으며, 층을 여러 개 쌓을수록 깊어진다는 의미로 심층 신경망(Deep Neural Network)이라고 부른다 (이재성, 2016).

26) 하이퍼 파라미터는 신경망 모형에 의해 수정될 수 없는 파라미터를 말한다. 즉, 분석가가 직접 설정해줘야 하는 파라미터를 하이퍼파라미터라고 부르며, 은닉층의 개수와 뉴런의 수, 학습률, 최적화함수 설정 등이 대표적인 예이다.

제 1 절 심층 신경망 모형의 정립

1) 변수 설정

수도권 교통본부에서 발행한 『2017년 전국 여객 O/D 전수화 및 장래수요예측 공동사업(수도권)』 보고서에 따르면 국내 장래교통 수요예측을 위한 수단선택 모형의 변수는 [표 5]와 같이 차내 통행시간, 차외 통행시간, 총 통행비용, 환승횟수 그리고 도착 지역 구분 더미변수를 사용하여 분석한다.

[표 5] 국내 수단선택 모형 변수

| 변수 | 변수의 설명 |
|-------|--|
| IVT | 차내 통행시간(단위:시간) |
| OVT | 차외 통행시간(단위:시간) 출발지에서 주요 통행수단정류장까지의 도보접근시간 + 탑승대기시간+주요통행 수단정류장에서부터 최종목적지까지의 도보접근시간 |
| COST | 총 통행비용(단위:백원) 승용차 : 연료비+주차비+보험료, 택시 : 요금, 대중교통:요금+비편익비용 |
| TRNF | 버스환승횟수 |
| DGANG | 도착지가 강남일 때 1 아니면 0 |

그러나 국외의 사례를 보면 NCTCOG(North Central Texas Council of Government)의 수단선택 모형의 경우, 통행특성 뿐만 아니라 수단특성변수, 지역변수 그리고 가구변수를 사용한 총 31개의 설명변수로 수단선택모형을 추정했다. 해당 사례는 각 특성변수를 세분화하여 개인의 특성을 최대한 반영할 수 있도록 변수를 설정하는 것을 알 수 있었다.

본 연구는 앞장에서 살펴보았던 대다수의 선행연구와 마찬가지로 수단선택 모형으로 통행자의 통행특성 뿐만 아니라 통행자의 개인특성도 반영할 수 있는 모형을 개발한다. 따라서 통행자의 통행특성인 차내 통

행시간, 차외 통행시간, 통행비용과 가구통행실태조사 자료를 통해 구득 가능한 통행자의 소득, 나이, 성별, 자동차보유대수를 설명변수로 설정한다.

본 연구의 설명변수인 통행시간은 차내 통행시간(IVTT), 차외 통행시간(OVTT)을 합한 총 통행시간으로 분석한다. 가구통행 실태조사 자료는 응답자가 이용한 교통수단에 대한 통행시간정보만을 알 수 있으므로, 비선택 대안의 통행시간과 통행거리는 가구통행실태조사 자료를 통해 구득가능한 OD Pair를 이용하여 비선택 대안에 대한 총 통행시간을 정산하고, 구득하지 못한 나머지 정보에 대해서는 서울시 교통수단별 평균통행속도 통계를 이용하여 응답자가 이용한 주수단의 통행시간으로 이동한 통행거리를 산정한 후 비선택 대안에 대한 통행시간을 산정한다²⁷⁾.

통행비용의 경우, 본 연구의 시간적 범위가 2016년을 기준으로 분석하기 때문에 이를 기준으로 수단별 요금을 보정한다²⁸⁾. 비선택 대안에 해당하는 변수 자료를 구축하기 위한 자세한 방법은 제 4장에서 다루도록 하겠다. 본 연구에서 최종적으로 설정한 수단선택 모형의 변수는 [표 6]과 같으며, 이 중 범주형 변수인 통행자의 성별, 나이, 차량보유, 소득수준은 더미변수²⁹⁾로 변환하여 분석한다.

27) 제 4장 자료의 구축에서 자세히 설명한다.

28) 본 연구는 한국교통연구원에서 발행한 여객보고서를 기반으로 보정하였다.

29) 변수의 범주가 2개인 경우는 0과 1로 표기가 가능하지만 범주가 3개인 경우는 (0,0),(0,1),(1,0)으로 더미변수를 추가하여 분석한다. 즉, n개의 범주를 가지는 변수는 n-1개의 더미변수가 생성된다.

[표 6] 변수설정 및 설명

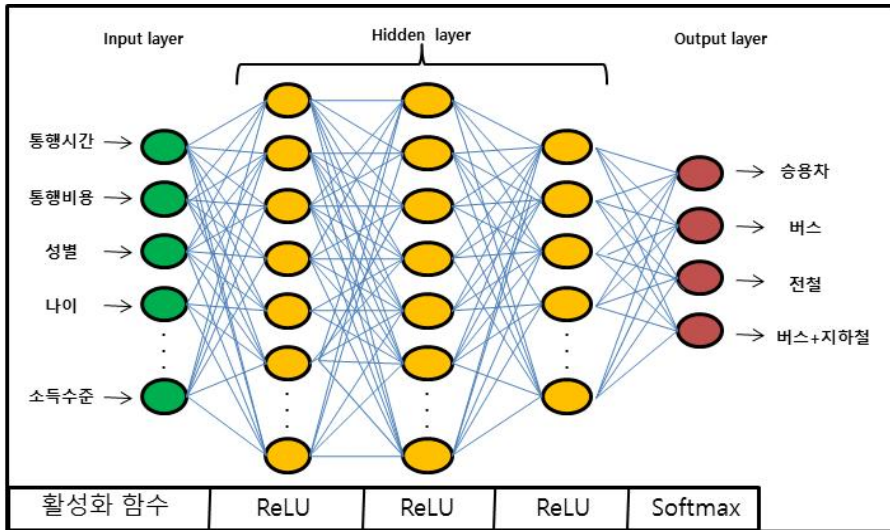
| 변수 | 구분 | 설명 | 표기 | |
|-------|-----------|----------------------------|---|----------------------------|
| 종속 변수 | 교통수단 | 0:승용차, 1:버스, 2:전철, 3:버스+전철 | m | |
| 설명 변수 | 통행 특성 | 통행 시간 (분) | 차내시간 + 차외시간 | TT_m |
| | | 통행 비용 (원) | 승용차 | 연료비+엔진오일비+타이어비+유지정비비+감가상각비 |
| | 버스 | | 기본요금(1,200원) | |
| | 전철 | | 기본요금(1,250원) +추가요금(10km이상은 5km마다 100원추가) | |
| | 버스+전철 | | 기본요금(1,200원) +추가요금(10km이상은 5km마다 100원추가) | |
| | 사회 경제적 특성 | 성별 | 0: 남, 1: 여 | X_{sex} |
| | | 나이 | 0 : 20세 미만, 1: 20~65세 미만, 2 : 65세 이상 | X_{old} |
| | | 소득 수준 (가구원 평균) | 0: 300만원 미만, 1: 300 만원 이상 | X_{inc} |
| | | 차량 보유 | 0 : 보유안함, 1 : 보유함 | X_{hc} |
| | 상수 | 대안특성상수 | | α_m |

2) 심층 신경망 모형의 정립

본 연구에서 정의하는 심층 신경망(DNN)은 일반적으로 은닉층이 두 개 이상인 인공신경망을 말한다. 심층 신경망은 구조의 변화와 하이퍼 파라미터 설정에 따라 모형의 적합성이 매우 다르게 나타난다. 본 연구에서는 은닉층의 개수에 따라 은닉층이 한 개인 경우 Type1, 은닉층이 두 개인 경우 Type2, 은닉층이 세 개인 경우를 Type3로 총 3가지 유형으로 나누고 각 모형의 적합성을 분석한다. 본 연구에서 설정한 심층 신경망 모형의 구조는 [그림 4]³⁰⁾와 같이 입력층, 은닉층, 출력층으로 구성된다.

입력층은 앞 절에서 살폈던 독립변수의 개수(더미변수의 개수 포함)인 총 14개의 뉴런이 입력되며, 출력층은 수단선택의 개수인 4개의 뉴런으로 구성된다. 은닉층의 뉴런 개수는 시행착오를 통해 각 층마다 256개의 뉴런으로 구성하였다. 기존 인공신경망의 역전파법의 문제를 해결하기 위해 입력층과 은닉층 뉴런의 활성화 함수는 렐루함수를 적용하였으며, 출력층은 분류문제에서 일반적으로 사용하는 소프트맥스 함수를 적용했다. 또한 과적합의 문제를 해결하기 위해 각 은닉층마다 드롭아웃을 적용하였다. 하이퍼 파라미터 설정과 관련한 부분은 다음 (2)하이퍼 파라미터 설정 챕터에서 명확히 제시하도록 하겠다.

30) 심층 신경망 모형 구성은 Saito Goki의 논문을 참고하여 재구성하였다.



[그림 4] 심층신경망 수단선택 모형

(1) 데이터 전처리

심층 신경망에서 딥러닝은 입력 데이터의 특징을 변환하지 않고 입력 데이터 자체를 입력하여 분석하는 패턴인식 기반의 분석 방법이다³¹⁾. 그러나 변동이 큰 변수가 있을 경우에는 데이터를 전처리하여 분석해야 한다. 본 연구에서는 통행비용과 통행시간 같은 변동이 큰 변수가 포함되기 때문에 식(4)를 이용하여 모든 표본의 평균과 분산이 정규분포를 따르도록 전처리하였다. 이를 통해 변수 간 단위에 따른 연산의 변동을 최소화했다. 범주형 변수의 경우, 더미변수를 생성하여 자료를 재구축하였다.

$$X_{norm} = \frac{X - \mu}{\sigma} \quad \dots(4)$$

여기서, μ : 속성 X의 평균

σ : 속성 X의 표준편차

31) Takayoshi Yamashita ,(2017), '그림과 수식으로 배우는 통통 딥러닝', 제이펍

(2) 하이퍼 파라미터(Hyper parameters) 설정

심층 신경망 모형을 학습시키기 위해 연구자가 초기에 모형의 기본값과 구조를 설정하는 파라미터를 하이퍼 파라미터라고 한다. 하이퍼 파라미터³²⁾는 오차함수와 역전파법을 이용하여 모형이 스스로 수정하는 weight와 bias와는 다르게 모형이 수정할 수 없는 고정 값을 의미한다.

본 연구에서 설정한 하이퍼 파라미터는 [표 7]과 같다. 기존의 인공신경망의 문제점은 학습 층의 깊이가 깊어질수록 과잉 학습문제가 발생하며, 오차 역전파 알고리즘이 동작하지 않는 문제(Gradient vanishing)가 있었다³³⁾. 이를 해결하기 위해 2006년에 Hinton에 의해 활성화 함수로 ReLU함수가 고안되었고, 본 연구에서는 다층의 신경망 학습을 위해 은닉층의 활성화 함수로 ReLU함수를 적용한다. 오차함수는 분류문제에서 일반적으로 사용되고 있는 교차엔트로피(식(5))를 이용하여 분석하며, 최적화 함수는 2015년 Kingma에 의해 고안된 Adam 최적화³⁴⁾를 적용한다. Adam 최적화는 momentum최적화 방법(식(6))과 RMSProp의 방법(식(7))이 결합된 형태로 식(8)과 같다. 기존 신경망 모형의 가장 큰 취약점은 모형이 복잡해질수록 과적합(Overfitting)문제가 발생한다는 것이다. 본 연구는 이를 해결하기 위해 2014년 Srivastava 에 의해 고안된 기법인 드롭아웃(DropOut)³⁵⁾ 방법을 적용하여 인공 신경망 모형의 한계를 보완한다.

32) 대표적인 하이퍼 파라미터는 은닉층, 학습률, 최적화함수 등 이다.

33) Hochreiter S. et al (2001) "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," A Field Guide to Dynamical Recurrent Neural Networks. IEEE Press,.

34) Momentum방식과 유사하게 기울기의 지수 평균과 RMSProp과 같이 기울기의 제곱값의 지수평균을 저장하는 방식으로, 일반적으로 β_1 은 0.9, β_2 는 0.999, ϵ 은 10^{-8} 으로 설정하여 분석한다.

35) 드롭아웃은 과적합 방지를 위한 방법으로 은닉층에 있는 뉴런을 일정비율로 삭제하여 weight가 과도하게 쏠리는 것을 방지한다. 과적합 방지를 위한 다른 방법으로는 Drop Connector가 있다.

$$E = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K r_{nk} \ln(y_{nk}) \quad \dots(5)$$

여기서,

y_{nk} : 모형이 n번째 데이터의 k번째 값을 예측한 값

r_{nk} : n번째 데이터의 k번째의 실제 값

$$\hat{m}^t = \frac{m^t}{(1-\beta_1^t)} \quad \dots(6) \quad , \quad \hat{v}^t = \frac{v^t}{(1-\beta_2^t)} \quad \dots(7)$$

$$w^{t+1} = w^t - \frac{\eta}{\sqrt{\hat{v}^t} + \epsilon} \hat{m}^t \quad \dots(8)$$

[표 7] 본 연구의 심층신경망 하이퍼 파라미터 설정

| 하이퍼 파라미터 | | 종류/값 |
|----------------------------------|-----|---------------|
| 활성화 함수 (activation function) | 은닉층 | ReLU |
| | 출력층 | Softmax |
| 오차함수(loss function) | | Cross Entropy |
| 최적화함수(optimizer) | | Adam |
| 드롭아웃(dropout) | | 0.3 |
| 학습률(Learning rate) | | 0.0002 |
| 학습횟수(epoch) ³⁶⁾ | | 150 |
| 배치사이즈(batch size) ³⁷⁾ | | 256 |

36) 학습 횟수는 데이터가 입력변수를 통해 입력되고 은닉층을 거쳐 출력층에 도착하고, 역전파법을 통해 다시 오차를 수정하는 하나의 Cycle 횟수를 의미한다.

37) 배치사이즈는 모형의 효율적인 학습을 위해 고안된 방법으로 모형이 한 번에 읽어들이는(학습하는) 데이터의 양을 의미한다. 즉 대량의 빅데이터를 처리할 때 다수의 데이터를 입력하고, 한 번에 결과값을 비교할 수 있어 매우 효율적인 방법이다.

제 2 절 비교 모형 : 다항 로짓 모형

본 연구에서 제안한 심층 신경망을 이용한 수단선택 모형의 적합성을 검증하기 위해 일반적으로 교통수단 선택모형으로 쓰이고 있는 다항 로짓 모형과 비교한다. 다항 로짓 모형의 변수설정은 심층 신경망 모형에서 설정한 변수와 동일하게 설정하여 적용한다. 위에서 설정한 설명변수를 토대로 각 수단별 효용함수를 나타내는 식은 식(9)과 같다.

$$U_n(m) = \alpha_m + \beta_1 TT_m + \beta_2 TC_m + \beta_{pcn} X_{pc} + \epsilon_m \quad \dots(9)$$

여기서,

| | |
|---------------|----------------------|
| $U_n(m)$ | : 개인 n이 이용한 수단 m의 효용 |
| TT_m | : m 수단의 총 통행시간 |
| TC_m | : m 수단의 통행비용 |
| X_{pc} | : 개인의 사회경제적 특성 |
| β_1 | : m 수단의 총 통행시간 계수 |
| β_2 | : m 수단의 통행비용 계수 |
| β_{pcn} | : 사회경제적 특성에 대한 계수 |
| α_m | : m 수단의 상수항 |
| ϵ_m | : m 수단의 확률적 효용 |

로짓 모형은 이용자의 효용을 극대화하는 대안을 선택한다는 기본가정을 가지고 있기 때문에, 개인 n이 수단m을 선택할 확률은 식(10)과 같이 나타낼 수 있다. 이용자는 각 수단별 효용을 비교하고, 가장 높은 효용을 가지는 수단을 선택하는 것으로 간주하여 수단선택 모형을 정립한다.

$$P_n(m) = \frac{e^{U_n(m)}}{\sum_{m=0}^M e^{U_n(m)}} \quad \dots(10)$$

제 3 절 모형의 검증 방법

본 연구에서 제안하는 심층 신경망 수단선택 모형의 적합성을 판단하기 위해 기존 교통수단 선택 단계에서 일반적으로 사용되고 있는 다항 로짓 모형과 비교하여 검증한다. 이 절에서는 기계학습 연구 분야에서 분류문제에서 검증 지표로 일반적으로 사용되고 있는 정오행렬 (Confusion matrix)을 이용하여 모형의 성능을 평가하고, 정확도 (Accuracy), 정밀도(Precision), 재현율(Recall), F1-score 지표를 통해 비교 분석한다.

1) 모형의 적합성 검증 지표

(1) 정오행렬(Confusion Matrix)

정오행렬은 분류된 행렬로부터 유도된다. 정오행렬은 모형이 분석 데이터 세트에 대해서 산출한 정확한 분류 내용과 부정확한 분류 내용을 한 표로 요약한다. 즉, 분류 대상의 실제값과 예측된 값을 행렬 형태로 비교하는 방식으로 나타내어, 대각행렬의 값은 올바른 분류의 수를 의미한다. 정오행렬은 [표 8]과 같이 나타낼 수 있으며, 분류 속성별 오측된 분류와 정분류를 확인할 수 있는 장점을 가지고 있다³⁸⁾.

[표 8] 분류 정오행렬의 예

| 분류정오행렬 | | | |
|----------|----------|----------|----------|
| | | 실제정답 | |
| | | Positive | Negative |
| 예측 정답 | Positive | TP | FP |
| | Negative | FN | TN |

주: TP : True Positive, FP: False Positive, FN : False Negative, TN: True Negative

38) Galit Shmueli, Peter C. Bruce, Nitin R. Patel (2017), '비즈니스 애널리틱스를 위한 데이터 마이닝', 이앤비플러스

(2) 정밀도(Precision), 재현율(Recall), F1-score

데이터 마이닝 분야에서 분류 정오행렬을 이용하여 모형의 적합성을 검증하는 대표적인 지표는 정밀도와 재현율이다.

[표 8]의 정오행렬을 기준으로 정밀도는 측정치의 오차정도를 가리키는 것으로, 모형이 예측한 개체 중 실제 값을 예측한 비율을 나타내며, 이를 수식으로 나타내면 식(11)과 같다. 재현율은 실제 참값의 개수 중에 모형이 참값으로 예측한 비율을 나타낸 것으로 식(12)와 같다. 정확도는 모형이 전체 개체 중에 정분류한 비율을 나타내는 것으로 식(13)과 같다.

$$Precision = \frac{TP}{TP+FP} \quad \dots(11)$$

$$Recall = \frac{TP}{TP+FN} \quad \dots(12)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad \dots(13)$$

F1-score는 정밀도와 재현율의 조화평균 값을 의미하며, 식(14)와 같이 나타낼 수 있다. 분석 대상 자료가 비대칭적인 경우, 이진 분류의 경우 True의 개체가 False보다 압도적으로 많을 때 모형이 제대로 학습이 되지 않아 모두 True로 예측하여도 정밀도가 높게 나오는 경우가 발생할 수 있다. 이러한 경우는 모형이 제대로 학습됐다고 보기 어렵다. 따라서 자료의 비대칭적 배치 문제로 인한 모형 검증의 한계를 보완하기 위해 F1-score를 이용하여 모형을 검증한다.

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad \dots(14)$$

2) 모형의 탄력성

신경망 모형은 층이 깊어질수록 인간이 이해하기 어려운 다차원의 공간에서 연산이 수행되기 때문에 모형이 어떠한 과정을 거쳐 결과를 도출하는지 해석하기 어려운 문제가 있다. 성능은 높지만 머신의 연산 과정을 알지 못하기 때문에 딥러닝을 블랙박스라고 불렀다³⁹⁾. 그러나 최근 딥러닝 모형의 해석을 위한 다양한 연구들이 진행되며, 심층 신경망 모형을 해석할 수 있는 다양한 방법론(Grad-Cam, LIME, Feature Visualization, Loss Landscape 등)이 등장하고 있다⁴⁰⁾.

본 연구에서는 미시경제학에서 가격에 변화에 따른 수요의 변화량을 나타내는 탄력성 개념을 도입하여 심층 신경망 모형을 해석하고 검증한다. 탄력성은 통행시간과 통행비용의 변화율에 따른 모형의 선택확률의 변화를 이용하여 분석한다. 심층 신경망의 경우, 다항 로짓 모형과 같이 정형화된 탄력성 산정방법이 없기 때문에 본 연구는 식(15)와 같이 해당 독립변수의 변화율에 따른 모형의 각 수단별 수요 예측량의 변화를 이용하여 심층 신경망의 탄력성을 산정한다. 다항 로짓 모형의 경우, 일반적으로 사용되는 탄력성 산정 방법에 따라 식(16)과 같이 독립 변수의 변화에 따른 로짓 모형의 선택확률 변화를 이용하여 탄력성을 산정한다. 본 연구에서 산정하는 탄력성은 표본의 평균값을 기준으로 해당 독립변수(통행시간, 통행비용)를 변화시켜 산정한다.

$$IE_{ijk} = \frac{D'_{ijk} - D_{ijk}}{D_{ijk}} / \frac{P_j}{100} \quad \dots(15)$$

$$E_j^i(j, k) = \beta^k z_{ji}^k (1 - P_{ji}) \quad \dots(16)$$

39) Nijkamp, P. et al (1996). Modelling inter-urban transport flows in Italy: A comparison between neural network analysis and logit analysis.

40) <http://www.shallowmind.co/jekyll/pixyll/2017/12/30/tree-regularization/>

여기서,

IE_{ijk} : 수단k를 이용하는 개인i의 변수j에 대한 수요 탄력성

D_{ijk} : 변수j에 대한 수단k의 수요 예측량

D'_{ijk} : 변수j에 대한 수단k의 수요 예측 변화량

$E_j^i(j,k)$: : 수단k를 이용하는 개인i의 변수j에 대한 다항 로짓 모형 탄력성

본 연구에서 추정하는 교통수단은 단일수단과 복합수단이 존재하기 때문에 단일수단과 복합수단간 통행시간과 통행비용의 변화는 서로에게 영향을 끼친다. 따라서 본 연구는 [표 9], [표 10]과 같이 수단간 통행특성변화를 고려하여 탄력성을 도출한다. 예를 들어 버스의 통행시간에 따른 탄력성을 예측할 때 해당 독립변수인 버스의 통행시간을 1%증가시킨다면, 이는 복합수단인 버스+전철 수단의 통행시간에도 영향을 미칠 것이므로 복합수단(버스+전철)의 통행시간도 0.5%로 같이 증가시켜주어 버스의 통행시간에 따른 수요 탄력성을 분석한다. 이를 통해 단일수단의 통행특성 변화에 따른 복합수단 유기적인 변화까지 고려하여 탄력성을 산정할 수 있다.

[표 9] 단일수단과 복합수단 간 통행시간 설정 방법

| 통행시간 | |
|--------|----------------|
| 버스 1%▲ | BS 0.5%▲ |
| 전철 1%▲ | BS 0.5%▲ |
| BS 1%▲ | 버스 1%▲, 전철 1%▲ |

[표 10] 단일수단과 복합수단 간 통행비용 설정 방법

| 통행비용 | |
|--------|----------------|
| 버스 1%▲ | BS 1%▲ |
| 전철 1%▲ | BS 1%▲ |
| BS 1%▲ | 버스 1%▲, 전철 1%▲ |

제 4 장 자료의 특성 및 구축

제 1 절 자료의 특성

본 연구는 출근목적 통행자의 수단선택 모형을 추정한다. 통행자의 사회경제적 특성과 통행특성을 반영한 수단선택 모형을 추정하기 위해 정부에서 실시하는 가구통행 실태조사 자료(이하 가통자료)를 기반으로 분석한다. 본 연구에서는 구득 가능한 가장 최신 자료인 2016년 가구통행 실태조사 자료를 사용한다. 2016년 가구통행 실태조사 자료는 2010년 자료에 비해 조사항목이 추가되어 개인의 통행을 보다 세부적으로 알 수 있다.

가구정보의 경우, 2016년 가통자료는 보유한 차량의 종류와 연식정보를 추가로 조사하였으며, 2010년 가통자료가 가구원수와 미취학 아동수를 기입한데 반해 2016년 가통자료는 총 가구원 수와 5세 이상 가구원수의 정보를 얻을 수 있다. 또한 2010년 가통자료에서는 통행자가 버스 정류장까지 접근 한 시간과 전철역까지 접근한 시간이 조사되어 있지만, 2016년 가통자료의 경우, 이용자가 이용한 수단에 대한 접근 시간만을 알 수 있다⁴¹⁾.

가구원 정보의 경우, 2016년과 2010년 자료는 큰 차이는 없었다.

통행정보의 경우, 통행목적과 통행수단의 분류에서 차이가 있었다. 2010년 가통자료의 통행목적은 총 10개로 구성되어 있는데 반해 2016년 가통자료는 2010년 자료에 있던 [여가/오락/외식/친지방문] 목적통행을 외식과 친지방문을 별도로 구분하여 12개의 목적으로 분류하였다. 통행수단의 경우, 2010년 자료에서는 총 18개의 수단으로 구분했지만 2016년

41) 통행자의 목적통행을 위한 출도착시각과 수단통행의 출도착시각이 각각 기재되어 있어 분석가가 이를 통해 산출해야한다.

(주 수단까지 접근 시간 = 수단통행 출발시간 - 목적통행 출발시간)

자료는 21개의 수단으로 세분화하여 분석하였다.

2010년 가동자료와 2016년 가동자료의 가장 큰 차이는 2010년 가동자료의 경우 통행자의 출발시간과 도착시간, 그리고 도착지 유형을 통해 이용자가 환승한 횟수와 이용한 교통수단(복합수단)을 알 수 있는데 반해, 2016년 가동자료는 도착지 유형에 환승을 별도로 기재하지 않고, 목적통행시간과 수단통행시간을 분리하여 자료를 구성한 것이다. 2010년 가동자료와 2016년 가동자료의 비교표는 [표 11]을 통해 확인할 수 있다.

2010년 가동자료의 경우, 오랜 기간 수정 작업으로 인해 자료의 신뢰성이 높고, 분석에 용이한 형태로 자료가 구성되어있다. 반면, 2016년 가동 자료의 경우, 도착종류에 환승여부가 기재되지 않아 통행자가 이용한 교통수단이 단일 혹은 복합인지 명확히 처리하는 과정⁴²⁾을 거쳐야만 유의미한 분석이 가능하다.

통행자의 단일/복합수단을 구분하고, 데이터를 가공한 결과, 2016년 서울시 가동자료의 통행량은 [표 12]과 같이 나타났다. 자료의 서울 시내 하루단위 수단 통행량⁴³⁾은 369,265 (통행/일)이며, 목적 통행량⁴⁴⁾은 196,784 (통행/일)이다. 이 중 출근목적인 수단 통행은 총 105,831 (통행/일)이며, 출근목적인 목적 통행량은 50,613 (통행/일)으로 전체 수단 통행의 약 28%를 차지한다. 본 연구의 분석에서 제외되는 수단인 도보, 장거리 통행수단, 화물차, 이륜차 등을 제외한 서울시 출근목적 목적 통행량은 총 38,468 (통행/일) 이다.

본 연구는 분석의 정확도를 위해 잘못 표기되거나 노이즈가 큰 아웃라이어 데이터를 제외하여 자료를 구축하여, 총 34,013 (통행/일)의 표본을 이용하여 수단선택 모형을 추정한다.

42) 본 연구는 가구정보 ID에 가구원별 통행정보를 담은 통행자 ID를 재구성하고, ID별 입력된 수단통행시간들의 도착시간과 출발시간의 차이가 0일 경우 환승을 한 것으로 처리하여 분석한다.

43) 통행자 n이 기점부터 종점까지 이동하는데 이용한 수단의 개수를 말한다.

44) 통행자 n이 최초출발지부터 최종목적지까지 통행한 한 통행을 말한다.

[표 11] 2010년, 2016년 가구통행실태조사 자료 비교

| 구분 | | 2010년 가구통행실태조사 | 2016년 가구통행실태조사 |
|--------|-------|--|---------------------------------|
| 가구 정보 | 차량 정보 | 차량유형별 보유대수 | 차량유형, 연식, 보유여부 |
| | 가구원 | 가구원수, 미취학 아동수 | 가구원수, 5세 이상 가구원 |
| | 접근 시간 | 버스정류장까지 접근시간 전철역까지 접근시간 | 이용한 수단에 대한 접근시간 |
| 가구원 정보 | 가구원 | 통행일자정보 미표기 | 통행일자정보 표기 |
| 통행 정보 | 통행 목적 | 총 10개의 통행목적 | 총 12개의 통행목적 |
| | 통행 수단 | 총 18개의 통행수단 | 총 21개의 통행수단 |
| | 도착 유형 | 1: 집, 2: 직장, 3: 학교, 4: 갈아타는 곳(환승), 5: 기타 | 1: 자택, 2: 직장, 3: 학교, 4: 기타장소 |
| | 통행 시각 | 출발/도착시각 | 목적별, 수단별 출발/도착시간 |

[표 12] 2016년 서울시 가구통행실태조사 목적별 통행 (단위: 통행/일)

| 구분 | 목적 | 목적 통행량 | 비율 | 수단 통행량 | 비율 |
|-----------|-----------|----------------|----------------|----------------|----------------|
| 1 | 배웅 | 859 | 0.44% | 1,219 | 0.33% |
| 2 | 귀사 | 2,399 | 1.22% | 2,779 | 0.75% |
| 3 | 귀가 | 88,585 | 45.02% | 172,833 | 46.80% |
| 4 | 출근 | 50,613 | 25.72% | 105,831 | 28.66% |
| 5 | 등교 | 14,826 | 7.53% | 27,547 | 7.46% |
| 6 | 학원 | 5,137 | 2.61% | 6,975 | 1.89% |
| 7 | 업무 | 2,008 | 1.02% | 3,058 | 0.83% |
| 8 | 쇼핑 | 7,956 | 4.04% | 10,903 | 2.95% |
| 9 | 여가 | 8,791 | 4.47% | 11,972 | 3.24% |
| 10 | 외식 | 3,532 | 1.79% | 4,657 | 1.26% |
| 11 | 친지방문 | 1,076 | 0.55% | 2,250 | 0.61% |
| 12 | 기타 | 11,002 | 5.59% | 19,241 | 5.21% |
| 합계 | | 196,784 | 100.00% | 369,265 | 100.00% |

제 2 절 자료의 구축

본 연구를 위한 기초자료는 2016년 가구통행실태조사 자료를 기반으로 한국교통연구원에서 발간한 『2017년 전국 여객 O/D 전수화 및 장래 수요예측』 보고서(이하 여객보고서)를 참조하여 구축하였다. 자료의 구축과 전처리를 위해 동적 타이핑 범용 프로그래밍 언어인 Python³⁴⁵⁾ 프로그램 언어와 Microsoft사에서 제공하는 Visual Basic for Application⁴⁶⁾을 이용하여 구축했다.

본 절에서는 수단선택모형의 추정 대상인 종속변수(교통수단) 설정과 자료 구축 방법을 정립한다. 그리고 가구통행실태조사에 기입된 가구의 사회경제적 특성을 가구원별 사회경제적 특성으로 분류하여 재정립한다. 선행연구 고찰을 통해 분석된 수단선택 모형의 핵심 설명변수인 통행시간, 통행비용은 통행자가 이용한 주수단의 통행시간을 제외하고, 나머지 변수는 가구통행실태조사 자료를 통해 구득할 수 없다. 따라서 본 절에서 통행자의 비선택 대안에 대한 통행시간과 통행비용을 구축하는 방법을 정립하고, 구축한 자료의 특성을 살펴본다.

1) 교통수단 설정

2016년 가구통행실태조사에서 조사된 총 21개의 수단을 [표 13]과 같이 총 9개의 수단으로 구분한다. 그리고 [표 14]의 구분방법에 따라 통행자의 통행수단을 단독수단과 복합수단으로 구분하여, 최종적으로 9개의 수단을 정립한다. 본 연구에서는 구분된 9개 수단 중 승용차, 버스, 전철, 버스+전철의 총 4개의 수단을 통근자의 선택 대안으로 설정한다.

45) 1991년 네덜란드의 귀도 반 로섬에 의해 개발된 프로그래밍 언어로, 확장성과 문법이 간결하여 입문하기 쉬우며, 웹개발, 데이터 분석, 머신러닝, 그래픽 등 다양한 분야에서 활용되고 있다(위키백과).

46) 1990년대 개발된 Visual Basic을 마이크로소프트 오피스에 탑재한 형식으로 특수 함수 작성 및 업무 자동화 등에 사용되었으며, 자료를 가시적으로 쉽게 확인할 수 있는 장점이 있다(두산백과).

[표 13] 가구통행실태조사 수단 변경(21개 수단→ 9개 수단)

| 가구통행실태조사기준 | | 수단분담 모형 기준 | |
|--------------|----------|--------------------|------|
| 수단구분 | TYPE | 수단구분 | TYPE |
| 도보 | 1 | 비기관 교통수단 도보 | 1 |
| 자전거 | 17 | 비기관 교통수단 자전거 | 2 |
| 소형화물 | 15 | 수단선택 비대상 화물/기타 | 3 |
| 중대형화물 | 16 | | |
| 오토바이 | 18 | | |
| 기타(항공,선박,기타) | 19,20,21 | | |
| 시외버스 | 7 | 수단선택 비대상 기타버스 | 4 |
| 고속버스 | 8 | | |
| 기타버스 | 9 | | |
| 일반철도 | 13 | 수단선택 비대상 철도/KTX | 5 |
| 고속철도 | 12 | | |
| 승용승합 | 2 | 수단선택 대상 승용차 | 6 |
| 승용승합동승 | 3 | | |
| 택시 | 14 | 수단선택 비대상 택시 | 7 |
| 시내버스 | 4 | 수단선택 대상 버스 | 8 |
| 마을버스 | 5 | | |
| 광역버스 | 6 | | |
| 지하철, 경전철 | 10,11 | 수단선택 대상 지하철/전철 | 9 |

자료 : 2017 전국 여객 O/D 전수화 및 장래수요예측

[표 14] 15개 수단을 주수단 8개(수도권 9개)로 변경

| 단독/복합수단 정의에 따른 기준 | | 수단선택 대상 수단 기준 | |
|-------------------|------|--------------------------|------|
| 수단구분 | TYPE | 수단구분 | TYPE |
| 도보 | 1 | 비기관 교통수단 | 1 |
| 자전거 | 2 | | |
| 화물/기타 | 3 | 화물/기타 | 2 |
| 기타버스 | 4 | 기타버스 | 3 |
| 일반철도, KTX | 5 | 일반철도, KTX | 4 |
| 승용차 | 6 | 승용차 | 5 |
| 택시 | 7 | 택시 | 6 |
| 택시 + 승용차 | 8 | | |
| 버스 | 9 | 버스 | 7 |
| 버스 + 승용차 | 10 | | |
| 버스 + 택시 | 11 | | |
| 지하철 | 12 | 지하철 | 8 |
| 지하철 + 승용차 | 13 | | |
| 지하철 + 택시 | 14 | | |
| 버스 + 지하철 | 15 | 버스 + 지하철(수도권) | 9 |

자료 : 2017 전국 여객 O/D 전수화 및 장래수요예측

본 연구의 분석 대상인 교통수단의 수단 분담률 현황은 [표 15]과 같다. 본 연구에서 구축한 수단 분담률은 승용차가 약 34.6%, 버스 30.2%, 전철 22.5%, 버스+전철 12.7%이다. 복합수단을 제외한 단일수단 간 비율이 거의 비슷하여, 데이터양의 불균형 문제에 따른 모형의 오추정 문제를 극복할 수 있다. 즉, 모형의 학습이 잘못되어 모형이 데이터의 양이 많은 수단을 모두 선택할 때, 생기는 정확성 해석의 오류를 줄일 수 있다.

[표 15] 본 연구의 선택 수단 구축 현황

| | | 승용차 | 버스 | 전철 | 버스+전철 | 합계 |
|------|-----|-------|-------|-------|-------|-------|
| 본 연구 | 통행량 | 11755 | 10276 | 7667 | 4315 | 34013 |
| | 비율 | 34.6% | 30.2% | 22.5% | 12.7% | 100% |

2) 통행자의 사회경제적 특성

본 연구에서 설명변수로 적용하는 통행자의 사회경제적 특성은 차량 보유여부, 나이, 성별, 소득으로 총 4가지 변수를 고려한다. 이용자 특성 중 나이의 경우, 나이가 많아질수록 특정 교통수단을 이용할 확률이 늘어나는 것은 아니기 때문에 범주형으로 구분하였다. 가구통행실태조사에는 출생연도가 표기되기 때문에 분석의 시간적 범위인 2016년을 기준으로 재구성하였다. 소득의 경우, 가구통행 실태조사 자료는 가구 단위의 월평균 수입을 산정하기 때문에 본 연구의 분석단위인 가구원 단위의 정확한 소득을 파악하기 어렵다. 따라서 본 연구에서는 가구통행 실태자료에서 명시한 소득 범주를 기반으로 중앙값을 가구당 월평균 수입으로 간주하고, 이를 가구당 5세 이상의 가구원 수로 나누어 소득을 재분배하였다. 나이는 20세 이하의 경우, 대중교통 할인요금이 적용되고, 65세 이상은 지하철을 무료로 이용하는 특성을 반영하기 위해 [표 16] 와 같이 3가지 범주로 나누어 분석한다. 가구원 평균 소득 또한 분석의 편의를 위해 조사된 중앙값인 300만원을 기준으로 2가지 범주로 나누어 분석한다.

[표 16] 통행자의 사회경제적 특성 범주화

| 나이 | TYPE | TYPE | 가구원 평균 소득 | TYPE | TYPE |
|----------|------|------|--------------|------|------|
| 20세 이하 | 0 | 0 | 0~100만원 미만 | 0 | 0 |
| 20~30세미만 | 1 | 1 | 100~200만원 미만 | 1 | |
| 30~40세미만 | 2 | | 200~300만원 미만 | 2 | |
| 40~50세미만 | 3 | | 300~400만원 미만 | 3 | |
| 50~65세미만 | 4 | | 400~500만원 미만 | 4 | |
| 65세 이상 | 5 | 2 | 500만원 이상 | 5 | 1 |

3) 통행시간 및 통행비용

본 연구의 분석을 위한 주 자료인 가구통행 실태조사 자료는 통행자의 사회경제적 특성과 통행자가 통행한 출발지와 목적지 그리고 이용한 주수단에 대한 통행시간과 접근시간을 알 수 있다. 그러나 수단선택 모형을 추정하기 위해서는 통행자가 선택하지 않은 다른 대안(수단)에 대한 통행특성 자료가 필수적으로 구득되어야 한다. 다수의 선행연구는 비선택 대안의 통행특성을 구축하기 위한 방법으로 교통수요분석 패키지인 EMME를 이용하여 네트워크자료를 통해 구축했다⁴⁷⁾. 하지만 EMME를 통해 구축한 자료는 존간 통행거리가 센트로이드 거리로 동일하다는 가정 하에 분석되기 때문에 이용자가 실제 통행한 거리와 통행시간과는 상당한 괴리가 있다. 특히 행정동 내 통행의 경우, 통행거리가 0으로 산정되는 문제를 가지고 있다. 본 연구는 가구통행 실태조사 자료와 EMME 네트워크자료의 매쉬업 가능성을 확인하기 위해 가구통행 실태조사 자료의 출근 목적 통행자 중 기종점의 행정코드가 같고 동일한 수단을 이용한 통행을 비교한 결과, 기종점과 이용한 교통수단이 같음에도 불구하고 통행자에 따라 통행시간의 격차가 상당히 큰 케이스가 존재함을 발견하여 네트워크 자료의 활용은 배제했다.

따라서 본 연구에서는 가구통행 실태조사 자료를 기반으로 통행자가 이용한 주수단 자료를 이용하여 기종점별 수단별 평균 총 통행시간 자료를 구축하여 기종점이 동일한 이용자의 비선택 대안에 대한 자료를 구축한다. 그리고 OD pair 비교를 통해 구축하지 못한 비선택 대안 정보에 대해서는 서울시내 교통수단별 통행속도⁴⁸⁾를 이용하여 통행자가 이동한 통행거리를 산출한 후, 비선택 대안의 차내시간을 정산한다. 차외시간의

47) 김희진, (2015), 통행 거리특성변수를 도입한 지역 간 통행수단선택 모형의 개발, 한양대 석사학위 논문.

48) 가구통행실태조사에서 통행자가 기입한 통행일자와 동일한 날짜에 조사된 오전 첨두시 수단별 평균통행속도를 이용한다. 승용차와 버스의 평균통행속도는 서울특별시 차량통행속도 보고서를 참고하여 구축하였으며, 버스+전철인 복합수단의 경우, 계산의 편의를 위해 공간평균속도로 도출하였다.

경우, [표 17]과 같이 통행자의 주 수단별 평균 차외시간을 적용하여 비선택 대안에 대한 총 통행시간 자료를 구축한다. 각 통행자의 통행 거리를 구하는 산출식은 식(17)과 같으며, 통행시간은 각 수단별 평균통행속도인 [표 18]을 참고하여 식(18)에 따라 도출한다. 본 연구의 통행시간 변수는 수단별 차내시간(IVTT)과 차외시간(OVTT)을 합한 총 통행시간으로 적용하여 분석한다.

$$DST_n = V_m \cdot IVTT_m^n / 60 \quad \dots(17)$$

$$IVTT_m = DST_m \cdot V_m \quad \dots(18)$$

여기서,

- DST_n : 개인 n 이 통행한 통행 거리(km)
- V_m : 수단 m 의 통행속도(km/h)
- $IVTT_m^n$: 개인 n 이 주수단 m을 이용한 차내 시간(min)
- $IVTT_m$: m 수단의 차내 시간(min)

[표 17] 구축자료의 수단별 평균 차외시간

| 수단 | 차외시간 |
|----------|------|
| 승용차 | 0 분 |
| 버스 | 13 분 |
| 지하철 | 17 분 |
| 버스 + 지하철 | 14 분 |

[표 18] 구축자료의 수단별 통행속도

| 수단 | 평균통행속도 |
|-------------------|-----------|
| 승용차 ¹⁾ | 25.0 km/h |
| 버스 ²⁾ | 20.2 km/h |
| 지하철 ³⁾ | 33.1 km/h |
| 버스 + 지하철 | 25.1 km/h |

자료 : 1) 서울시 교통정보센터(www.topis.seoul.go.kr)

2) 2016 서울특별시 차량통행속도 보고서

3) 서울교통공사(www.seoulmetro.co.kr), 표정속도기준

통행비용 변수는 선행 단계에서 구축한 수단별 평균 통행속도와 통행자의 개별 통행거리를 이용하여 구축한다. 승용차 통행비용은 유류비, 엔진오일비, 유지보수비, 감가상각비를 더한 값으로 구하며, 이는 여객보고서를 기반으로 통행거리에 따른 승용차 운영비용을 곱한 값으로 설정했다. 본 연구에서 구축한 출근통행 목적의 승용차 평균 통행속도는 25.0km/h이기 때문에 [표 19]을 참고하여 통행거리를 단위 km당 300.17원을 곱하여 통행비용을 정산하였다. 참고한 여객보고서는 2010년 기준의 자료를 [표 20]에서 제시한 유류비 증가율과 [표 21]에 나타난 물가지수 변화(물가 상승률)를 이용하여 2016년에 적합하게 보정한 자료이다. 여객보고서의 구축된 자료는 본 연구의 시간적 범위와 보정 연도가 동일하기 때문에 별도의 보정 없이 적용하였다.

[표 19] 속도별 승용차 운영비용(수도권)

(단위: 원/km)

| 속도 | 유류비 ¹⁾ | 엔진 오일비 ²⁾ | 타이어비 ²⁾ | 유지 정비비 ²⁾ | 감가 상각비 ²⁾ | 합계 |
|-----------|-------------------|----------------------|--------------------|----------------------|----------------------|---------------|
| 10 | 170.95 | 8.37 | 0.62 | 10.71 | 260.82 | 451.48 |
| 20 | 103.08 | 6.98 | 1.15 | 12.66 | 222.3 | 346.17 |
| 30 | 87.68 | 6.04 | 1.77 | 14.99 | 189.68 | 300.17 |
| 40 | 79.19 | 5.11 | 2.58 | 15.58 | 160.04 | 262.51 |
| 50 | 80.88 | 5.11 | 3.28 | 17.52 | 136.34 | 243.14 |
| 60 | 83.97 | 5.11 | 4.18 | 18.5 | 123 | 234.76 |
| 70 | 88.59 | 5.11 | 5.15 | 19.47 | 112.63 | 230.96 |
| 80 | 90.79 | 4.65 | 6.22 | 21.41 | 100.78 | 223.85 |
| 90 | 98.76 | 4.18 | 7.53 | 22 | 93.36 | 225.83 |
| 100 | 106.45 | 5.11 | 8.95 | 23.38 | 86.84 | 230.73 |
| 110 | 115.18 | 6.51 | 10.91 | 22 | 79.43 | 234.04 |
| 120 | 129.14 | 9.76 | 12.95 | 28.23 | 70.24 | 250.32 |

주: 1) 유류비 증가율 -18.1%를 보정하여 사용함

2) 물가 상승률 10.9%를 보정하여 사용함

[표 20] 유류비 변화(수도권)

| 구분 | 2010년 10월 2주 | 2016년 10월 2주 | 증가율 |
|-------|--------------|--------------|---------|
| 일반휘발유 | 1,710원 | 1,402원 | -18.10% |

자료 : www.opinet.co.kr

[표 21] 물가지수 변화(수도권)

| 구분 | 2010년 10월 2주 | 2016년 10월 2주 | 증가율 |
|------|--------------|--------------|-------|
| 총 지수 | 91.05 | 100.97 | 10.9% |

자료 : www.kosis.kr

대중교통비용은 [표 22]의 수도권 대중교통 요금체계에 기반하여 구축한다. 단독수단 통행의 경우, 본 연구는 서울특별시 내 통행을 기본으로 분석하기 때문에 통행자는 이용한 버스종류는 지선버스와 간선버스를 이용한다고 간주한다. 따라서 단독수단인 버스의 요금은 1,200원으로 정산하여 분석한다. 지하철과 복합수단(버스+지하철)은 거리비례제 요금을 적용하여 정산한다. 거리비례제 요금은 기본요금과 추가요금으로 나누어 정산되며, 추가요금의 경우 총거리에서 기본거리를 뺀 거리에 요금단가를 적용하여 산출한다. 복합수단의 경우, 기본거리는 10km이며, 기본요금은 1,250원이다. 요금단가는 5km당 100원을 부과하여 자료를 구축하였다.

단, 나이가 20세 이하인 청소년의 경우, 대중교통 비용은 기본 운임에서 350원을 제하고 20%할인한 청소년 할인⁴⁹⁾ 요금을 적용하였다. 65세 이상인 노인의 경우, 전철 통행비용을 0원으로 산정하여 구축하였다.

본 방법을 통해 구축된 분석 자료의 기술통계량은 [표 23]과 같다.

[표 22] 대중교통 요금체계(수도권)

| 구분 | | 요금체계 | 비고 |
|----------|-----------|--|--------------------|
| 단독 수단 | 시내 버스 | 균일요금제(기본요금) | |
| | 수도권 전철 | 거리비례제(기본요금+추가요금) 추가요금 = (총거리-기본거리)*요금단가 | 기본거리 = 10km |
| 복합수단 | | 거리비례제(기본요금+추가요금) 추가요금 = (총거리-기본거리)*요금단가 | 요금단가 = 100원/5km |

49) 기존 만 13~18세 미만의 중고생 혹은 조금 늦게 진학한 19~24세 중고생은 기본요금에서 350원을 공제한 금액의 20%를 할인한 요금을 적용한다.

[표 23] 분석자료 기술통계량(표본 수 : 34,013)

| 구분 | 통행시간 | | | | 통행비용 | | | |
|------|-------|-------|-------|-------|---------|---------|---------|---------|
| | 승용차 | 버스 | 전철 | 버스+전철 | 승용차 | 버스 | 전철 | 버스+전철 |
| 평균 | 31.34 | 51.51 | 40.94 | 45.68 | 3944.58 | 1199.01 | 1077.26 | 1318.25 |
| 표준편차 | 16.43 | 21.22 | 12.70 | 16.48 | 2073.43 | 21.75 | 243.34 | 134.93 |
| 최소값 | 2 | 7 | 10 | 17 | 901 | 720 | 0 | 720 |
| 25% | 20 | 35 | 30 | 34 | 2,402 | 1,200 | 1,050 | 1,200 |
| 50% | 30 | 50 | 39 | 43 | 3,603 | 1,200 | 1,050 | 1,300 |
| 75% | 40 | 64 | 48 | 55 | 5,103 | 1,200 | 1,250 | 1,400 |
| 최대값 | 104 | 145 | 110 | 135 | 12,908 | 1,200 | 1,250 | 1,900 |

제 5 장 모형의 추정 및 검증

본 장에서는 제 4 장에서 구축한 서울특별시 출근목적 통행 34,013 (통행/가구원)의 표본을 가지고 심층 신경망 모형과 다항 로짓 모형을 통해 추정한다. 수단선택 모형의 학습과 검증을 위해 전체 데이터를 7:3의 비율로 나눠 총 23,809개의 학습 데이터(Train Set)와 10,204의 테스트 데이터(Test Set)를 구분하여 모형을 추정 한다. 심층 신경망의 경우, 모형의 확장성 비교를 위해 학습 데이터, 검증 데이터(Validation Set), 테스트 데이터를 6:1:3으로 세분화시켜 분석하였다. 각 모형별 수단선택 모형 추정이 완료되면, 심층 신경망 모형의 수단선택 모형으로서 적합성을 검증하기 위해 제 3장 연구 방법론에서 정립한 검증지표를 이용하여 다항 로짓 모형과 비교분석한다. 그리고 모형의 해석을 위하여 통행시간과 통행비용에 따른 수단선택의 변화를 탄력성 지표를 이용하여 해석한다.

제 1 절 수단선택 모형의 추정

1) 심층 신경망 모형 추정 결과

본 연구는 심층 신경망 모형의 추정을 위해 Google에서 공개한 오픈소스 소프트웨어인 TensorFlow⁵⁰와 Keras⁵¹를 이용하여 분석하였으며 컴퓨터 언어는 Python3을 이용했다. 은닉층 개수에 따라 심층 신경망 모형을 3가지로 구분하였고, 구분된 각 모형의 추정 결과를 비교한다.

심층 신경망 모형의 확장성(일반화)과 정확성을 비교하기 위해 학습 데이터의 오차(Etrain)와 검증 데이터(Etest)의 오차, 그리고 검증 데이터

50) 2015년 구글에서 공개한 딥러닝과 머신러닝 기술을 위한 오픈소스 소프트웨어로 딥러닝 분야에서 널리 사용되고 있는 프로그램이다.

51) API형 오픈소스 신경망 라이브러리로 Tensorflow를 기반으로 구현되어 있으며, 구현의 용이성이 있어 많이 이용되고 있는 라이브러리이다.

와 실제 값의 오차로 총 3개의 오차⁵²⁾를 비교한다.

$E_{test} \approx E_{train}$ 이면 모형의 오차 정도가 학습데이터와 검증 데이터 간 유사함을 나타내기 때문에 모형의 일반화(Generalization) 가능성이 높고, $E_{test} \approx 0$ 으로 테스트 데이터의 오차가 0에 가까우면, 모형의 예측값이 실제값과 거의 유사함을 나타내기 때문에 모형의 적합성이 높음을 의미한다⁵³⁾.

본 연구에서 심층 신경망을 이용하여 추정한 수단선택 모형의 결과는 정확도가 90%이상을 상회하는 수준으로 매우 높은 정확도를 나타냈다. [그림 5]를 보면 심층 신경망의 은닉층 개수가 증가할수록 정확도가 증가하는 것을 볼 수 있었고, 특히 은닉층이 3개인 Type3의 경우 약 97%로 매우 높은 성능을 나타냈다. 또한 [표 24]과 [그림 6]에서 심층 신경망 모형의 유형별 오차를 확인하면, 은닉층의 수가 적을수록 E_{train} 과 E_{test} 의 오차가 가장 작아 단조로운 모형인 Type1이 일반화하기에 적합한 것으로 나타났지만, 정확성이 가장 높은 Type3의 학습오차와 검증오차 간 차이가 그리 크지 않기 때문에 정확성이 높은 Type3를 수단선택 모형으로 적용하기에 적합한 것으로 판단된다.

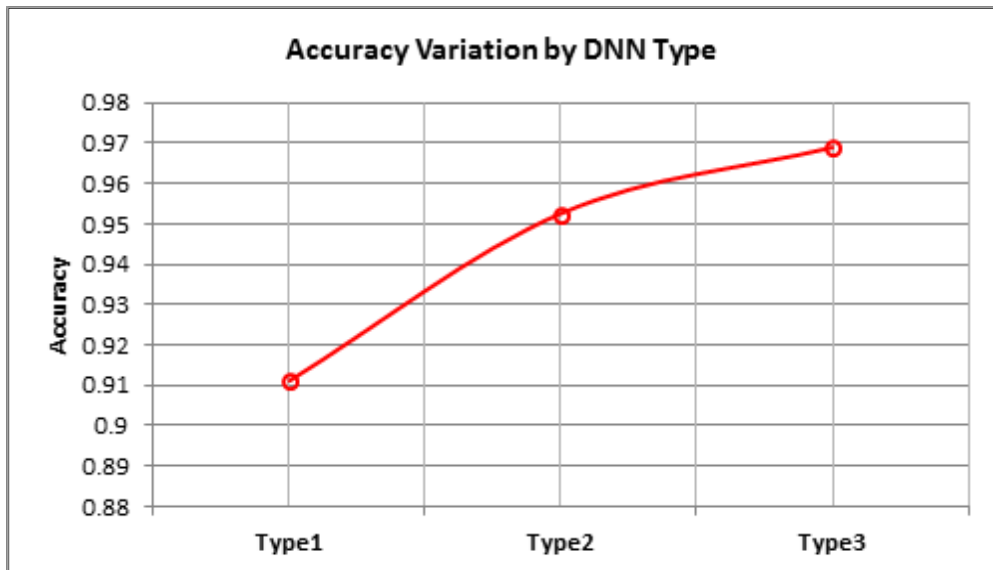
모형별 실제값과 예측값의 분류 결과는 다음 절인 모형의 검증에서 자세히 다루도록 하겠다. 심층신경망 유형별 학습 횟수에 따른 E_{train} 과 E_{test} 의 오차 변화율과 유형별 정확도 비교분석 그래프는 [부록A]에서 확인할 수 있다.

[표 24] 심층 신경망 모형의 유형별 오차와 정확도

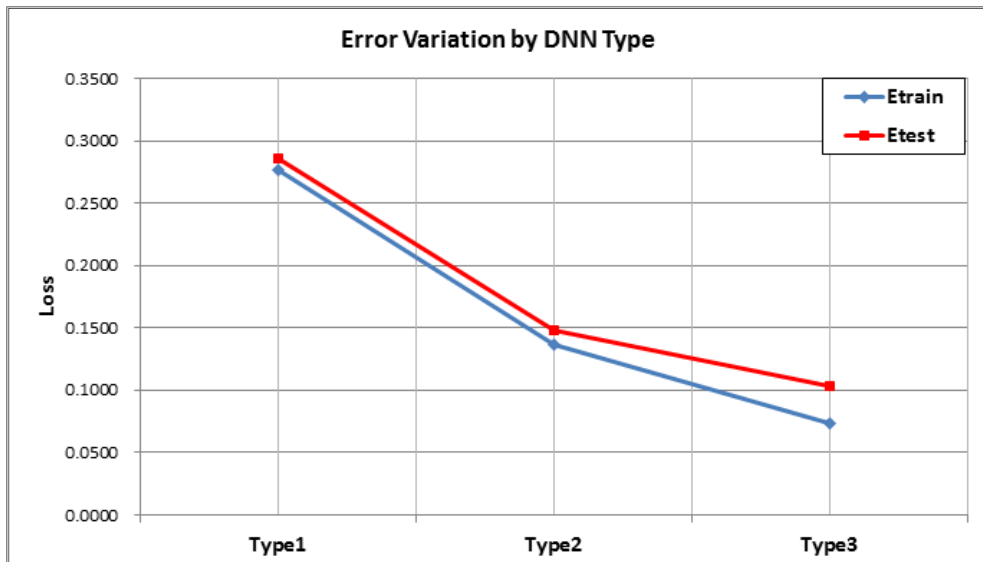
| | Etrain | Etest | Etest - Etrain | Accuracy |
|--------------------------|---------------|--------------|-----------------------|-----------------|
| Type1 (은닉층1개) | 0.2760 | 0.2860 | 0.0100 | 0.911 |
| Type2 (은닉층 2개) | 0.1364 | 0.1481 | 0.0117 | 0.9524 |
| Type3 (은닉층 3개) | 0.0733 | 0.1031 | 0.0298 | 0.9687 |

52) 교차엔트로피 함수(식(5))에 따른 오차값을 의미한다.

53) 모형의 정교화는 정확도를 올리지만, 일반화 측면에선 떨어질 수 있다.



[그림 5] 심층 신경망 유형에 따른 정확도 변화



[그림 6] 심층 신경망 유형에 따른 오차 변화

2) 다항 로짓 모형 추정 결과

서울시 출근목적 통행자의 수단선택 로짓 모형을 추정하기 위해 본 연구에서는 LIMDEP 9.0 NLOGIT4.0 프로그램을 이용하여 분석했다. 수단선택 모형 추정 시 다항 로짓에서 사용되는 변수인 통행시간과 통행비용은 모든 통행수단이 동일한 계수로 추정한다. 이용자 특성 변수는 수단마다 반영되기에 적합한 특성을 적용하였다. 먼저, 승용차 효용함수의 경우, 모형의 적합성을 판단하기 위해 성별을 이용자 특성으로 적용하였다. 버스의 효용함수는 청장년층에 해당되는 나이인지를 판단하는 나이 더미 변수(old1)와 소득 더미변수를 적용하였다. 전철의 효용함수는 무료승차가 가능한 나이(65세 이상)의 여부를 묻는 나이 더미 변수(old2)와 차량보유여부 더미변수를 적용하여 분석했다. 그리고 버스+전철 효용함수는 청소년 할인이 가능한 나이(20세 미만)의 여부를 묻는 나이더미 변수(old0)를 적용하여 분석했다. 위의 통행자의 사회경제적 특성을 반영한 각 수단별 효용함수는 식(19)~식(22)과 같이 나타낼 수 있으며, 이 효용함수를 반영하여 추정된 모형의 계수 값은 [표 25]와 같다.

$$U_n(Car) = \beta_1 TT_{car} + \beta_2 TC_{car} + \gamma_1 X_{sex} \quad \dots(19)$$

$$U_n(Bus) = \alpha_{bus} + \beta_1 TT_{bus} + \beta_2 TC_{bus} + \gamma_2 X_{inc} + \gamma_3 X_{old1} \quad \dots(20)$$

$$U_n(Subw) = \alpha_{subw} + \beta_1 TT_{subw} + \beta_2 TC_{subw} + \gamma_4 X_{old2} + \gamma_5 X_{hvc} \quad \dots(21)$$

$$U_n(BS) = \alpha_{BS} + \beta_1 TT_{BS} + \beta_2 TC_{BS} + \gamma_6 X_{old0} \quad \dots(22)$$

총 23,809개의 관찰치를 통해 추정한 결과, 통행시간과 통행비용의 추정 계수는 음수로 통행시간과 통행비용이 커질수록 해당 수단의 효용은 떨어지는 결과를 얻어 직관적으로 예상한 결과와 동일하게 나타났다. 또한 추정된 수단별 대안특성 상수항을 통해 모든 수단의 통행시간과 통행비용이 동일할 때, 버스가 가장 높은 효용을 가지며, 다음으로 승용차, 전철, 복합수단(버스+전철) 순으로 나타났다. 승용차는 버스에 비해 효용이 낮지만 전철과 복합수단에 비해 높은 효용을 가지는 것으로 분석되었다.

추정된 계수를 통해 이용자 특성을 살펴보면, 승용차의 경우, 성별의 계수 값이 음수로 나타나 여성보다 남성이 승용차를 선호하고, 효용이 높은 것으로 나타났다. 이는 서울시 자동차 보유의 성별 비율을 고려했을 때 납득 가능한 결과로 보인다. 청소년 할인을 받는 청소년일 경우, 복합수단(버스+전철)에 대한 효용이 높게 나타났으며, 노인의 경우, 전철의 무임승차 정책으로 인해 전철을 이용하는 효용이 높음을 알 수 있었다. 전철은 여성일수록 이용할 효용이 높아짐을 알 수 있었다. 모형의 추정된 계수의 부호를 확인한 결과, 직관적으로 이해하기에 적합한 결과가 추정됨을 알 수 있었다.

다항 로짓 모형에 추정된 계수를 토대로 Test set을 적용한 결과, 모형의 정확도는 약 52.72%로 나타내었다. 모형의 적합성에 대한 부분은 다음절에서 자세히 다루도록 하겠다.

[표 25] 다항 로짓 모형 계수 추정 결과

| Number of observations | | 23,809 | |
|-------------------------|--------|-------------|----------------|
| Iterations completed | | 20 | |
| Log likelihood function | | -27312.83 | |
| Variable | Mode | Coefficient | Standard Error |
| TT _m | All | -0.1133*** | 0.0017 |
| TC _m | All | -0.0003*** | 0.0001 |
| X _{inc} | Bus | 0.0891 | 0.0730 |
| X _{hc} | Subway | -0.6271*** | 0.0386 |
| X _{old0} | BS | 0.7892** | 0.3380 |
| X _{old1} | Bus | -0.5276*** | 0.0780 |
| X _{old2} | Subway | 0.0516 | 0.0936 |
| X _{sex} | Car | -1.1898*** | 0.0330 |
| α_m | Bus | 1.2798*** | 0.1108 |
| | Subway | -0.3274*** | 0.0460 |
| | BS | -0.5678*** | 0.0393 |

주 : 신뢰수준 99% : ***, 신뢰수준 95% : **, 신뢰수준 90% : * 에서 유의함
모든 값은 소수점 5자리에서 반올림하여 계산함

제 2 절 수단선택 모형의 검증

1) 적합성

본 절에선 모형의 적합성을 검증하기 위해 제 3절에서 정립한 지표인 정밀도, 재현율, F1 score를 이용하여 각 모형의 적합성을 검증한다. 검증지표를 산정하기 위해 검증 데이터를 이용하여 각 모형의 추정 결과를 정오표로 구축했다. 심층 신경망은 유형별로 [표 26], [표 27], [표 28]과 같이 정오표가 구축되었고, 다항 로짓 모형의 정오표는 [표 29]와 같다.

분석 결과, 전체 4개 수단을 모두 고려했을 때, 정확도는 [그림 7]과 같이 심층 신경망 모형이 다항 로짓 모형에 비해 월등히 높은 정확도를 나타내었다. 특히 심층 신경망 Type3 모형의 경우는 정확도가 약 97%로 가장 좋은 성능을 나타냈다.

각 교통수단별 모형의 정확도를 비교하면, 심층 신경망 모형의 경우, 4개 수단 모두 비교적 고르게 높은 정확도를 가지고 있었고, 특히 승용차에 대한 분류 정확도가 95%이상으로 4개 수단 중 가장 높은 것으로 나타났다. 반면, 버스+전철인 복합수단의 경우는 심층 신경망 유형별로 정확도에 차이가 남을 확인할 수 있었다. 은닉층이 한 개인 심층 신경망 Type1의 경우, 복합수단을 분류하는 비율이 70%로 다른 단일수단에 대한 분류정확도가 90%이상임을 비교해보았을 때 비교적 낮은 정확도를 나타내었다. 그러나 은닉층을 늘린 심층 신경망 Type2, Type3로 갈수록 분류가 어려웠던 복합수단의 정확도가 80%, 90%이상으로 증가하며 다른 단일수단과 마찬가지로 복합수단에 대한 분류 정확성이 올라갔다. 이는 심층 신경망의 은닉층이 증가함에 따라 모형이 차원을 증가시키며 세분화된 분류가 가능해지기 때문에 나타난 결과로 해석된다. 즉, 심층 신경망의 모형이 고도화될수록 성격이 비슷한 단일수단과 복합수단에 대한 분류정확도가 높아짐을 확인할 수 있었다. 심층 신경망 모형이 단일수단과 복합수단을 높은 정확도로 분류하는데 반해 다항 로짓 모형은 편중된

정확성을 나타냈다.

다항 로짓 모형의 경우, 4개의 수단 중 승용차에 대한 분류 정확성이 약 77%로 비교적 높았으나 대중교통인 버스와 전철에 대해서는 각각 약 50%, 45%로 낮은 정확성을 가졌고, 특히, 버스+전철인 복합수단에 대해서는 약 8%의 정확성을 나타내 단일수단과 복합수단을 동시에 분류하는데 매우 낮은 성능을 보였다.

정오표를 통한 모형의 정확성 분석을 통해 고도화되고 복잡한 교통수단, 특히 성격이 비슷한 교통수단에 대해서는 로짓 모형이 취약한 것을 알 수 있었다. 반면 심층 신경망 모형은 은닉층이 증가하고, 모형이 고도화될수록 단일수단뿐만 아니라 복합수단에 대한 분류 정확성도 높아 향후 보다 다양해질 교통수단에 대한 수단선택모형으로 적합한 것으로 보인다.

[표 26] 심층 신경망 Type1의 정오표

| DNN Type1 | | Predicted | | | |
|--|-------|------------------|------------------|------------------|-----------------|
| | | 승용차 | 버스 | 전철 | 버스+전철 |
| True | 승용차 | 3329 (94.79%) | 173 | 0 | 10 |
| | 버스 | 92 | 2965 (94.58%) | 42 | 36 |
| | 전철 | 21 | 82 | 2109 (92.42%) | 70 |
| | 버스+전철 | 52 | 178 | 152 | 893 (70.04%) |
| Misclassified : 908, Total : 10,204, Accuracy : 0.9131 | | | | | |

[표 27] 심층 신경망 Type2의 정오표

| DNN Type2 | | Predicted | | | |
|---|-------|------------------|------------------|------------------|------------------|
| | | 승용차 | 버스 | 전철 | 버스+전철 |
| True | 승용차 | 3413 (97.18%) | 90 | 0 | 9 |
| | 버스 | 28 | 3065 (97.76%) | 18 | 24 |
| | 전철 | 11 | 56 | 2186 (95.79%) | 29 |
| | 버스+전철 | 32 | 128 | 61 | 1054 (82.67%) |
| Misclassified : 486 Total : 10,204, Accuracy : 0.9524 | | | | | |

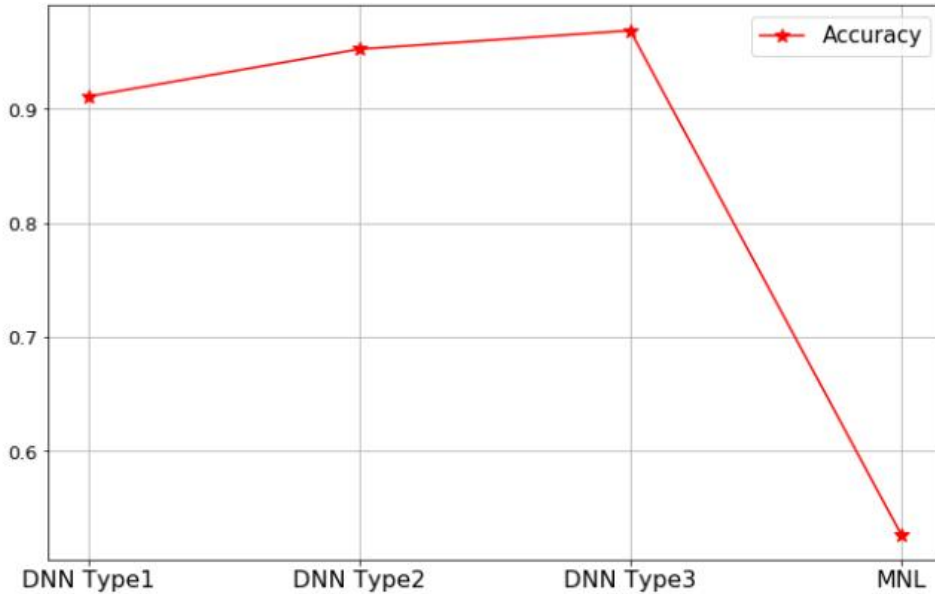
[표 28] 심층 신경망 Type3의 정오표

| DNN Type3 | | Predicted | | | |
|--|-------|------------------|------------------|------------------|------------------|
| | | 승용차 | 버스 | 전철 | 버스+전철 |
| True | 승용차 | 3470 (98.80%) | 36 | 1 | 5 |
| | 버스 | 46 | 3033 (96.75%) | 18 | 38 |
| | 전철 | 9 | 35 | 2216 (97.11%) | 22 |
| | 버스+전철 | 9 | 46 | 54 | 1166 (91.45%) |
| Misclassified : 248, Total : 10,204, Accuracy : 0.9687 | | | | | |

[표 29] 다항 로짓 모형의 정오표

| MNL | | Predicted | | | |
|--|-------|------------------|------------------|------------------|----------------|
| | | 승용차 | 버스 | 전철 | 버스+전철 |
| True | 승용차 | 2698 (76.52%) | 717 | 97 | 0 |
| | 버스 | 1312 | 1562 (49.82%) | 257 | 4 |
| | 전철 | 731 | 518 | 1020 (44.70%) | 13 |
| | 버스+전철 | 394 | 206 | 575 | 100 (7.84%) |
| Misclassified : 4,824, Total : 11,292, Accuracy : 0.5272 | | | | | |

Accuracy comparison by models



[그림 7] 모형별 정확도 비교

본 연구의 수단선택모형으로 추정된 모형별 추정결과를 세부적으로 판단하기 위한 모형별 정밀도, 재현율, F1 score의 결과는 [표 30]~[표 33]와 같다. 심층 신경망 모형의 은닉층 개수를 증가시켜 모형이 복잡화 될수록 정밀도, 재현율은 상승하는 것을 알 수 있다. 심층 신경망 Type1 모형은 다른 심층 신경망 유형에 비해 버스+전철(복합수단)수단에서의 정밀도와 재현율이 낮게 나타났다. 심층 신경망 Type1의 복합수단 정밀도는 0.89이고, 재현율은 0.70으로, 모형이 버스+전철(복합수단)이라고 예측하는 값 중 참 일 확률이 약 89%로 예측한 수단에 대한 정확성은 높으나, 전체 버스+전철(복합수단)수단 중 복합수단으로 예측하는 비율은 70%로 낮게 나타났다. 이는 분류한 값의 적중률은 높지만 분류 자체의 정확성은 떨어짐을 의미한다. 반면, 보다 고도화된 모형인 심층 신경망 Type3의 경우, 정밀도와 재현율 모두 모든 수단에서 고르게 높은 수치를 나타내고 있어, 분류가 어려운 복합수단도 효과적으로 분류하는 것으

로 분석됐다.

다항 로짓 모형의 복합수단(전철+버스)에 대한 정밀도는 0.85, 재현율은 0.08로 모형이 복합수단을 분류하는 전체적인 성능(예측률)은 낮으나, 모형이 복합수단이라고 예측한 값 중에서 그 값이 참일 확률(적중률)은 높은 것으로 나타났다. 전철에 대한 정밀도와 재현율은 각각 0.52, 0.45으로 모형이 해당 수단에 대한 예측률과 예측한 값의 적중률 모두 낮아 다항 로짓 모형은 성격이 비슷한 단일수단과 복합수단에 대해 분류가 매우 취약한 것으로 나타났다.

각 교통수단에 따른 모형별 검증지표의 비교결과는 [그림 8]~[그림 11]과 같다. 복합수단인 [그림 11]에 따르면 심층 신경망 Type1과 다항 로짓 모형의 정밀도가 같은 수준으로 높은 것을 알 수 있으나, 재현율을 비교해보았을 때 다항 로짓 모형의 성능이 현저히 낮아 복합수단에 대해 분류 정확성이 심층 신경망에 비해 현저히 떨어짐을 확인할 수 있었다.

본 연구의 결과를 통해 교통수단이 복잡해지고 다양화 될수록 수단선택 모형으로 심층 신경망을 적용하여 깊은 은닉층의 구성이나 적절한 하이퍼 파라미터 설정으로 모형을 고도화 하면 어려운 수단선택 예측의 문제에서도 높은 정확성을 갖는 수단 선택 모형을 추정할 수 있을 것으로 판단된다.

[표 30] 심층 신경망 Type1의 검증결과

| 모형 | 수단 | 정밀도 | 재현율 | F1 score |
|--------------|----------------------|------|------|----------|
| DNN Type1 | 승용차 | 0.95 | 0.95 | 0.95 |
| | 버스 | 0.87 | 0.95 | 0.91 |
| | 전철 | 0.92 | 0.92 | 0.92 |
| | 버스+전철 | 0.89 | 0.70 | 0.78 |
| | 가중 평균 ⁵⁴⁾ | 0.91 | 0.91 | 0.91 |

54) Machine Learning 라이브러리인 Scikit-learn에서 제공하는 가중평균에 의해 산정되었으며, 수단별 불균형한 샘플의 양을 고려하여 평균을 산정함

[표 31] 심층 신경망 Type2의 검증결과

| 모형 | 수단 | 정밀도 | 재현율 | F1 score |
|--------------|-------|-------------|-------------|-------------|
| DNN Type2 | 승용차 | 0.98 | 0.97 | 0.97 |
| | 버스 | 0.92 | 0.98 | 0.95 |
| | 전철 | 0.97 | 0.96 | 0.96 |
| | 버스+전철 | 0.94 | 0.83 | 0.88 |
| | 가중 평균 | 0.95 | 0.95 | 0.95 |

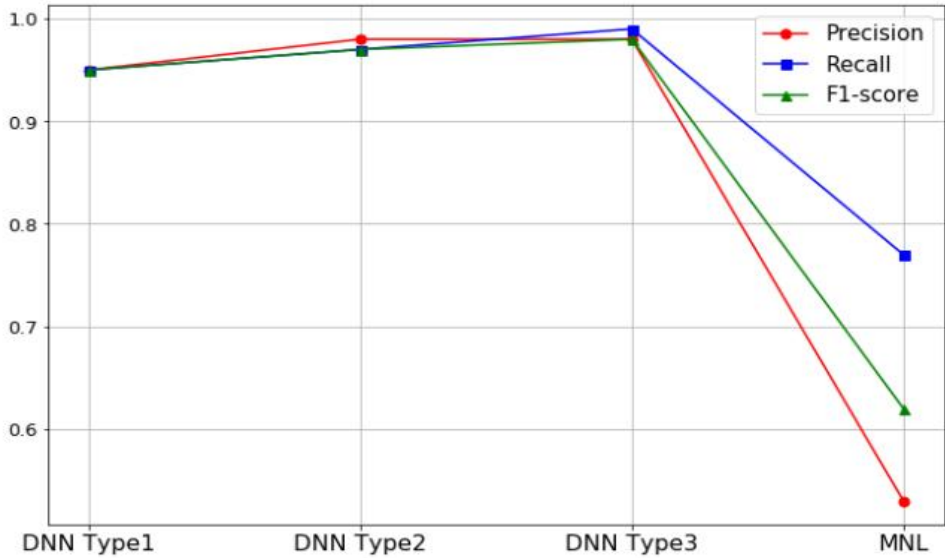
[표 32] 심층 신경망 Type3의 검증결과

| 모형 | 수단 | 정밀도 | 재현율 | F1 score |
|--------------|-------|-------------|-------------|-------------|
| DNN Type3 | 승용차 | 0.98 | 0.99 | 0.98 |
| | 버스 | 0.96 | 0.97 | 0.97 |
| | 전철 | 0.97 | 0.97 | 0.97 |
| | 버스+전철 | 0.95 | 0.91 | 0.93 |
| | 가중 평균 | 0.97 | 0.97 | 0.97 |

[표 33] 다항 로짓 모형의 검증결과

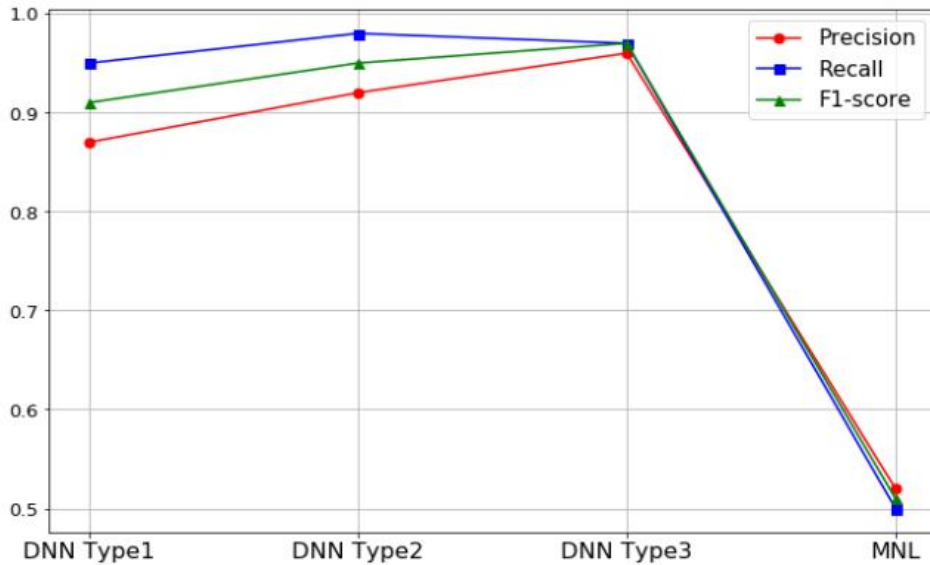
| 모형 | 수단 | 정밀도 | 재현율 | F1 score |
|-----|-------|-------------|-------------|-------------|
| MNL | 승용차 | 0.53 | 0.77 | 0.62 |
| | 버스 | 0.52 | 0.50 | 0.51 |
| | 전철 | 0.52 | 0.45 | 0.48 |
| | 버스+전철 | 0.85 | 0.08 | 0.14 |
| | 가중 평균 | 0.56 | 0.53 | 0.50 |

Validation index comparison by models (Mode: Car)



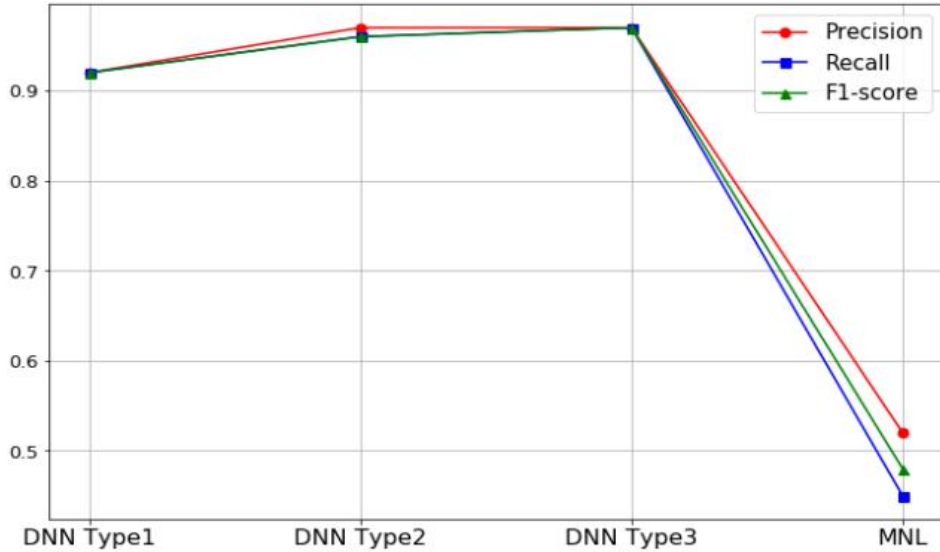
[그림 8] 모형별 검증 지표 비교(수단: 자동차)

Validation index comparison by models (Mode: Bus)



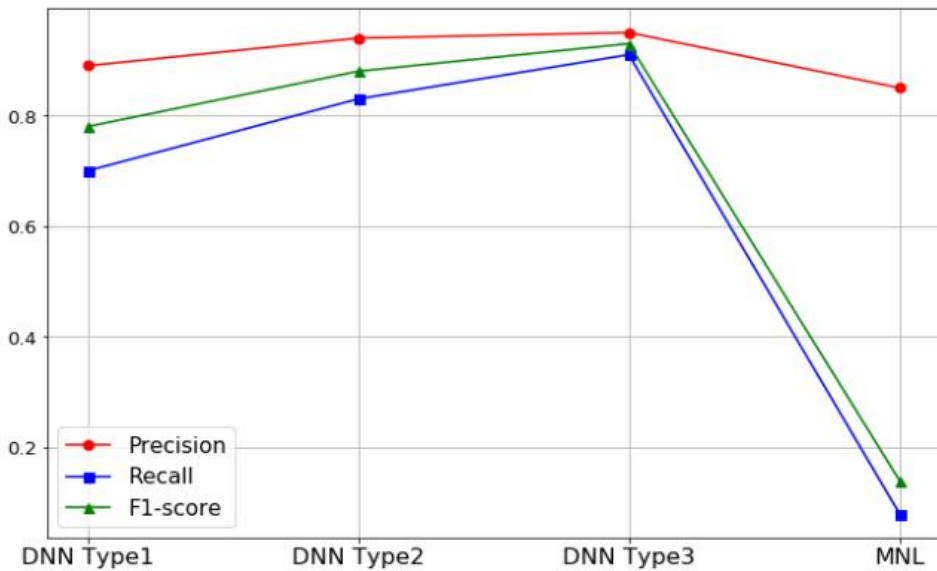
[그림 9] 모형별 검증 지표 비교(수단: 버스)

Validation index comparison by models (Mode: Subway)



[그림 10] 모형별 검증 지표 비교(수단: 전철)

Validation index comparison by models (Mode: BS)



[그림 11] 모형별 검증 지표 비교(수단: 버스+전철)

2) 탄력성

본 절에선 각 모형별로 설명변수가 종속변수에 미치는 영향을 해석하기 위해 탄력성의 개념을 도입하여, 모형을 해석한다. 탄력성은 3절 연구의 방법론에서 소개한 모형별로 교통 분야에서 일반적으로 사용되고 있는 수요 탄력성을 이용하여 분석하였다. 본 연구에서는 수요 탄력성을 산정하기 위해 통행시간과 통행비용의 변화율 대비 모형이 예측하는 교통수단별 수요 예측량의 변화율을 이용하여 산정한다. 본 연구의 교통수단은 단일수단과 복합수단이 함께 고려된다. 즉, 단일수단인 버스와 전철에 대한 통행시간, 통행비용의 변화는 복합수단인 버스+전철 수단에 영향을 미치기 때문에 제 3장 연구 방법론에서 언급한 [표 9]와 [표 10]의 내용을 적용하여 분석했다.

수단별 수요 탄력성은 각 독립변수의 평균값을 기준으로 통행시간과 통행비용의 변화에 따라 값을 산정했다. 통행시간과 통행비용에 대한 수요 탄력성⁵⁵⁾은 분석 대상인 4개 수단 모두 자기 탄력성의 부호가 음수로 나타나 직관적으로 예상한 바와 동일하게 통행시간과 통행비용이 증가할수록 해당 수단에 대한 수요는 감소하는 것으로 나타났다. [표 34]에 따르면 대중교통의 경우, 통행시간에 따른 수요의 변화가 승용차에 비해 비교적 큰 것으로 나타났으며, 특히 모형의 학습이 잘 된(정확도가 높은) 심층 신경망 Type3는 승용차의 자기 탄력성의 절대값이 타 교통수단(대중교통)에 비해 작은 것으로 나타나 승용차를 이용하는 통행자는 통행시간에 대해 덜 민감하게 반응하는 것을 알 수 있었다. 즉, 승용차를 이용하는 통행자는 통행시간이 조금 증가하더라도 편안함과 빠른 접근이 가능한 교통수단을 선호한다고 판단할 수 있다.

[표 35]의 통행비용에 따른 탄력성을 살펴보면 심층 신경망 모형과 다항 로짓 모형 모두, 대중교통 수단에 비해 승용차가 비교적 큰 것으로 산정되었다. 이는 승용차를 이용하는 통행자가 대중교통에 비해 통행비용에 민감하게 반응한다는 것을 의미한다.

55) 통행시간과 통행비용에 따른 모형별 수단별 교차 탄력성은 [부록C]에 수록했다.

각 모형별 통행시간과 통행비용에 따른 자기 탄력성을 분석한 결과, 대중교통의 경우, 통행시간 탄력성의 절대값이 통행비용 탄력성 값보다 비교적 높게 나타나 대중교통을 이용하는 통행자는 통행비용보다 통행시간에 민감하게 반응하는 것으로 분석되며, 승용차를 이용하는 통행자의 경우, 대중교통에 비해 통행비용에 민감하게 반응하는 것을 알 수 있었다.

[표 34] 통행시간에 관한 모형별 수요 탄력성

| 통행시간 | 승용차 | 버스 | 전철 | 버스+전철 |
|-----------|---------|---------|---------|---------|
| DNN type1 | -0.1892 | -1.9113 | -3.4123 | -5.5412 |
| DNN type2 | -0.5412 | -0.2267 | -2.1245 | -1.5251 |
| DNN type3 | -0.0844 | -1.6640 | -1.9781 | -1.1154 |
| MNL | -0.0228 | -0.0365 | -0.0331 | -0.0176 |

주 : DNN의 탄력성 : 설명변수의 변화에 따른 모형의 수단별 수요 예측량의 변화량을 이용하여 탄력성을 산정함
MNL(다항 로짓 모형)의 탄력성: 표본의 평균치를 기준으로 산정한 점 탄력성을 의미함

[표 35] 통행비용에 관한 모형별 수요 탄력성

| 통행비용 | 승용차 | 버스 | 전철 | 버스+전철 |
|-----------|---------|---------|---------|---------|
| DNN type1 | -0.3731 | -0.0333 | -1.0929 | -0.0545 |
| DNN type2 | -1.1242 | -0.0084 | -0.1033 | -0.1164 |
| DNN type3 | -1.0213 | -0.0337 | -0.2741 | -0.0937 |
| MNL | -0.0076 | -0.0019 | -0.0023 | -0.0017 |

제 6 장 결론

제 1 절 결론

교통계획에서 교통수요를 예측하기 위해 전통적으로 4단계 교통수요 추정방법이 사용되었다. 그 중 3단계인 수단분담선택 단계는 최근 각 지자체의 신교통수단의 도입과 산업의 발전에 따른 스마트 모빌리티, 카셰어링 등과 같이 복잡해진 통행자의 통행패턴 변화로 인해 그 중요성이 더욱 커지고 있다. 그동안 수단분담선택 단계에서는 주로 로짓 모형을 이용하여 통행자의 수단을 예측하여왔다. 그러나 로짓 모형은 통행자 개인의 선호 다양성을 반영하기 어려운 문제와 비관련 대안의 독립성 문제를 가지고 있었다. 이에 이를 극복하고자 하는 연구가 다양하게 이루어짐을 선행연구를 통해 알 수 있었으나 여전히 비관련 대안의 문제, 자료의 희소성 문제 그리고 다양한 환경 조건과 통행자 개인의 특성을 온전히 반영하기에는 어려운 한계가 존재한다. 또한 통행자가 이용할 수 있는 교통수단이 매우 복잡해지고 다양해지면서 기존의 로짓 모형이 빠르게 변화하고 진화하는 교통수단에 대해 대응가능한지 확인하고, 이를 보완할 새로운 수단선택 모형의 제안이 필요함을 알 수 있다.

따라서 본 연구는 딥러닝의 한 종류인 심층 신경망을 이용하여 기존 로짓 모형의 한계를 극복할 수 있는 수단선택 모형을 추정한다. 본 연구에서 제시하는 심층 신경망은 수리통계학적 모형인 로짓 모형과는 다르게 상관관계가 높은 변수도 입력 자료로 사용할 수 있어 다양한 데이터의 적용과 많은 변수를 고려할 수 있는 장점이 있다. 또한 비관련 대안의 독립성 문제와 자료 희소성의 문제도 극복할 수 있다.

본 연구는 통행자의 사회경제적 특성과 통행특성을 동시에 고려하기에 용이한 2016년 가구통행실태조사 자료를 이용하여 서울시 통근목적 통행을 대상으로 수단선택 모형을 추정한다. 수단선택 모형의 변수는 통행자의 통행특성을 반영하는 통행시간(차내시간+차외시간), 통행비용과

사회경제적 특성을 나타내는 성별, 나이, 소득, 자동차보유여부로 설정하였다. 사회경제적 특성을 나타내는 변수 중 하나인 나이의 경우, 대중교통 요금의 할인이 적용되는 청소년, 전철 요금 면제가 적용되는 노인, 그리고 일반인으로 3가지 범주로 나누었으며, 소득은 가구원별 평균소득으로 재구성하였다. 본 연구는 다양하고 성격이 비슷한 교통수단에도 정확성이 높은 모형을 추정하기 위해 종속변수로 단일수단 뿐만 아니라 복합수단도 포함하여 승용차, 버스, 전철, 버스+전철로 총 4가지 교통수단에 대해 추정한다.

심층 신경망 모형의 경우, 모형의 복잡도에 따라 성능이 달라지고, 특히 은닉층의 개수가 증가하며 층이 깊어질수록 다차원의 분류가 가능하다. 이러한 모형의 성능을 확인하기 위해 은닉층의 개수에 따라 유형을 3가지로 구분하여 정립하였다. 모형의 추정과 검증을 위해 가구통행실태조사 원시 자료를 전처리하여 구축한 총 34,013 통행의 표본을 가지고 모형을 추정한다. 전체 데이터 중 70%인 23,809개의 통행은 모형의 학습을 위한 데이터로 사용하며, 나머지 30%의 데이터는 모형의 검증을 위해 활용한다.

교통수단 선택모형의 추정결과, 은닉층을 3개 가진 심층 신경망 Type3 모형의 예측 정확도가 96.87%로 검증 데이터 10,204 통행 중 단 319 통행만을 오분류한 것으로 분석되어 가장 정확도가 높았다. 그 뒤를 이어 은닉층이 2개인 심층 신경망 Type2가 95.24%, 심층 신경망 Type1은 91.10%로 심층 신경망을 적용하여 추정한 수단선택 모형은 정확도가 90%를 상회하는 수준으로 매우 높은 성능을 보였다. 반면, 다항 로짓 모형은 정확도가 52.72%로 심층 신경망에 비해 낮은 분류 정확도를 보였다. 특히, 복합수단인 버스+전철 수단에 대해 7.84%의 낮은 정확성을 보여 분류가 어려운 복합수단에 대해 매우 취약한 것으로 나타났다. 따라서 모형의 분류 정확성은 심층 신경망 모형이 다항 로짓 모형보다 우수한 것으로 입증됐다.

모형의 직관적인 해석을 통한 일반화 가능성을 판단하기 위해 통행시간과 통행비용의 수요 탄력성을 산정하여 검증했다. 각 모형별 탄력성을

산정한 결과, 모든 모형의 탄력성 부호가 음수로 나타나 통행시간과 통행비용이 증가할수록 해당 수단에 대한 수요가 감소함을 알 수 있어 모형의 신뢰성을 확인했다. 통행시간과 통행비용의 교통수단별 탄력성 비교결과, 통행비용에 비해 통행시간의 탄력성의 절대값이 높아 통행자가 통행비용보다 통행시간에 더 민감하게 반응함을 알 수 있었다. 모형의 정확성이 높은 심층 신경망 Type3를 기준으로 통행시간의 탄력성을 분석하면, 승용차의 자기 탄력성의 절대값이 타 교통수단(대중교통)에 비해 작음을 알 수 있다. 이는 승용차를 이용하는 통행자는 통행시간이 조금 증가하더라도 편안함과 빠른 접근이 가능한 승용차 수단을 선호하기 때문으로 판단된다.

통행비용에 따른 탄력성의 절대값은 대중교통 수단에 비해 승용차가 다소 높은 것으로 산정되었다. 이는 승용차를 이용하는 통행자가 대중교통에 비해 통행비용에 민감하게 반응한다는 것을 알 수 있었지만 그 값이 그리 크지 않아 통행자가 통행비용에 크게 의존하지는 않는 것으로 보인다.

본 연구는 딥러닝의 한 종류인 심층 신경망을 이용하여 기존의 수단 선택 단계에서 주로 쓰였던 로짓 모형 한계를 극복하고, 심층 신경망 모형이 다양화된 통행수단에 대해서 더 높은 정확도로 분류함을 확인하였다. 특히 분류가 어려운 복합수단에 대해서 다항 로짓 모형은 매우 취약함을 나타낸 반면 심층 신경망은 96% 이상의 높은 정확성을 가졌다. 이는 앞으로 통행자의 다양화된 통행수단에 대해서도 정확한 분류가 가능함을 확인하여 스마트 모빌리티, 카셰어링, 트랩 등 기존에 존재하는 교통수단과 성격이 비슷한 수단에 대해서도 높은 확률로 정분류가 가능할 것으로 보인다. 또한 다양하고 복잡해질 교통수단에 적극적으로 대응할 수 있는 수단선택 모형을 제시한 점에서 의의있는 연구로 판단된다.

산업의 발전으로 인해 통행자의 통행패턴이 더욱 다양해짐에 따라 기존에 수단선택 모형에서 주로 고려되어온 통행특성(통행시간, 통행비용) 외에 다양한 요소(날씨특성, 정보특성, 실시간 환경특성 등)의 고려도 더욱 중요해질 것이다.

따라서 심층 신경망을 이용하여 다양한 변수를 모형에 적용할 필요가 있으며, 심층 신경망 모형의 최적의 은닉층 깊이와 하이퍼 파라미터를 설정한다면 본 연구의 결과처럼 복잡하고 다양한 교통수단에 대해서도 높은 정확도를 갖는 수단선택 모형을 추정할 수 있을 것으로 보인다. 심층 신경망의 도입은 향후 교통계획에서 보다 지능화되고, 신뢰성 있는 교통수요 예측을 위한 하나의 대안이 될 수 있을 것으로 판단된다.

제 2 절 한계 및 향후 연구과제

심층 신경망은 4차 산업혁명으로 인한 빅데이터의 대두와 새로운 알고리즘의 개발, 하드웨어의 성능향상으로 깊은 은닉층을 효과적으로 학습시킬 수 있어 다양한 분야에서 적용되며 모형의 높은 성능이 입증되고 있다. 본 연구는 딥러닝의 한 방법인 심층 신경망을 이용하여 수단선택 모형을 추정하는 연구를 진행하였다. 심층 신경망을 이용하여 기존의 수단선택 로짓 모형이 가지고 있는 문제를 해결하고, 복잡한 교통수단에 대해서도 높은 정확성을 갖는 수단선택 모형을 추정할 수 있었다. 그러나 본 연구는 자료의 한계와 추정 모형에 대해 몇 가지 한계가 있다.

먼저, 본 연구의 자료로 활용된 2016년 가구통행실태 조사 자료는 배포시기가 얼마 되지 않은 최신 자료로 충분히 검수되지 못한 문제가 있을 수 있다. 또한 통행자의 비선택 대안(교통수단)에 대한 통행시간과 통행비용이 제공되지 않아 본 연구의 4장에서 제시한 자료구축방법을 통해 통행자의 비선택 대안에 대한 통행특성을 구축하였다. 특히 차외시간의 경우, 2010년 가구통행실태조사는 버스정류장과 전철역까지의 접근시간 정보가 있어 비선택 대안에 대한 차외시간 정보를 알 수 있었으나, 2016년 가구통행 실태조사 자료는 해당 정보가 기입되지 않았다. 본 연구는 OD pair를 통한 비선택 대안에 대한 총 통행시간 구축하고, 이를 통해 구축하지 못한 비선택 대안에 대해서는 주 수단별 평균 차외시간을 적용하여 비선택 대안에 대한 정보를 가공했다. 그러나 이와 같은 방법을 통해 구축한 자료는 통행자가 실제로 이용할 가능성이 있는 비선택 대안에

대한 통행특성 정보와는 약간의 괴리가 있을 수 있는 한계가 있다.

두 번째는 모형의 해석과 관련한 부분이다. 본 연구에서는 수요 탄력성의 개념을 도입하여 통행시간과 통행비용에 따른 각 수단별 수요의 변화량을 통해 모형을 직관적으로 해석했다. 로짓 모형의 경우, 입력 변수에 따른 계수가 추정되어 이를 활용한 시간가치 산정이 가능한 장점이 있다. 그러나 심층 신경망은 각 입력변수와 은닉층 그리고 출력층의 노드간 수많은 가중치와 바이어스로 이루어져 있어 입력 변수에 대한 대표적인 계수를 추정하기는 어려웠다.

본 연구가 비록 가구통행 실태조사 자료만을 사용하여 수단선택 모형을 추정하였지만, 통행자의 사회경제적 특성을 고려한 수단선택모형으로 적용하기에는 신빙성 있는 자료를 사용한 사실은 틀림없다.

향후 통행자의 행태와 환경을 반영한 양질의 자료의 구득이 가능해 진다면 다양한 변수의 적용이 가능한 딥러닝 모형은 수단선택 모형으로 널리 활용될 수 있을 것이다. 특히, 통행자의 정보특성이나 도시특성, 날씨특성과 같은 다양한 변수를 적용한 수단선택모형 연구를 진행한다면 통행자의 개인특성과 환경특성을 잘 반영한 모형을 추정할 수 있을 것으로 보인다.

본 연구의 결과를 통해 점점 복잡해지고 급변하는 사회의 현상에 효과적으로 대응하는 보다 지능화된 교통 수요예측 모형으로서 심층 신경망이 적용될 수 있을 것으로 판단되며, 통행자의 다양한 통행선택 요인을 분석하기에 적극적으로 활용될 수 있을 것으로 판단된다.

참 고 문 헌

- 교통연구원 (2017), 「전국 여객 O/D 전수화 및 장래수요예측」
- 김경환, 조세원, 박래준, 송경빈 (2017), "인공 지능 기법을 이용한 전력 수요예측 동향." 「대한전기학회 학술대회 논문집」 : 457-458.
- 김민주 (2018), 로짓계열 수단선택모형의 추정 및 적합성 비교분석, 서울대학교 대학원 석사학위논문
- 김성희, 이창무, 안건혁 (2001), "대중교통으로의 보행거리가 통행수단선택에 미치는 영향", 「국토계획」, 36(7) : 297-307.
- 김정기 (2016), 확률 심층 신경망에서 연산자 정밀도에 대한 연구, 서울대학교 대학원 석사학위논문
- 김희진 (2015), 통행 거리특성변수를 도입한 지역 간 통행수단선택 모형의 개발, 한양대학교 대학원 석사학위논문
- 남대식, 서원호 (2017), 'AI를 만나면 교통은 어떻게 똑똑해질까?', 「카카오 AI 리포트」 Vol.9 : 8-12.
- 박동주, 김병우, 정영선, 안창욱 (2018), "Deep Neural Network 기반 프로야구 일일 관중 수 예측: 광주-기아 챔피언스 필드를 중심으로." 「스마트미디어저널」, 7 : 1-8.
- 박상민, 이환필, 권철우, 박성호, 윤일수 (2018), "Deep Neural Network를 이용한 고속도로 교통사고 예측모형 개발 방법론에 관한 연구 (주요 노선을 대상으로)." 「한국 ITS 학회 학술대회」, 324-326.
- 박철웅, 이인범, 이병용, 박수봉 (1999), "다층신경망을 이용한 얼굴검색 알고리즘 (The Facial Image Searching Algorithm using Multi-Layer Neural Networks)", 「한국멀티미디어학회」, vol(2) : 672-677.
- 배윤경, 정진혁, 김형진 (2010), "잠재계층분석에 따른 수단선택모형비교 분석", 「대한교통학회지」, 28(3) : 99-107.
- 성현곤, 최막중, 이수기 (2014), "장거리 통행의 교통수단 선택 결정요인", 「국토계획」, 49(2) : 245-257.

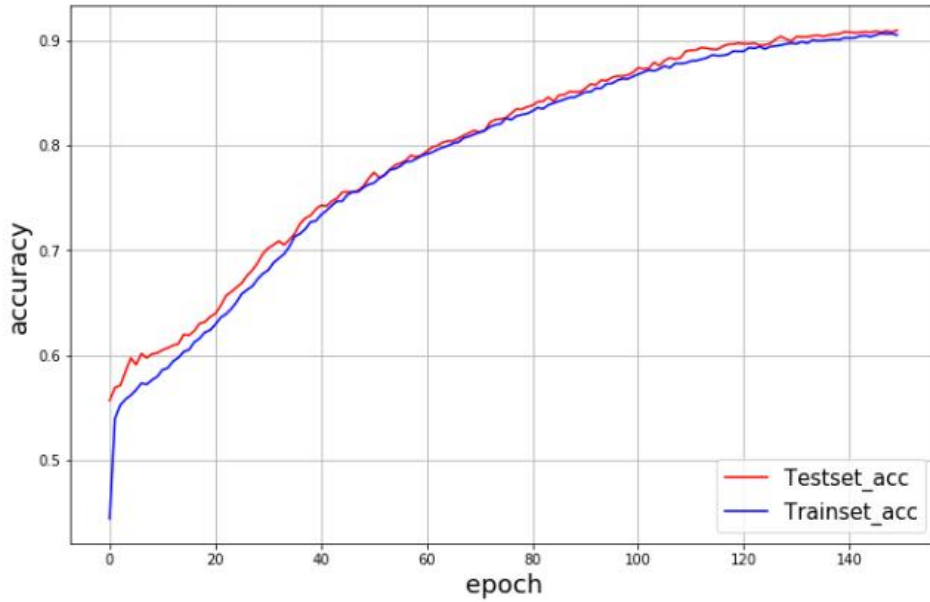
- 수도권 교통본부 (2017), 「2017년 전국 여객 O/D 전수화 및 장래수요에
 측 공동사업(수도권)」
- 신진호 (2018), SVM, DNN 분류기를 이용한 다중생체신호 기반 인간감
 정인식, 인하대학교 대학원 석사학위논문
- 안홍섭 (2016), 심층신경망의 효과적인 초기화를 위한 분할 훈련 및 병합
 기법에 관한 연구, 인하대학교 대학원 박사학위논문
- 양준영, 장준혁 (2016), "딥러닝 모델 adaptation 기술의 연구 동향." 「한
 국통신학회지 (정보와통신)」 33.9 : 3-7.
- 이용주, 황재성, 김수희, 이철기 (2018), "딥러닝을 활용한 차량대기길이
 추정모형 개발." 「한국 ITS 학회 논문지」, 17.2 : 39-57.
- 이재성 (2016), "심층 신경망의 발전 과정과 이해", 「한국통신학회지(정
 보와통신)」, 33(10) : 40-48.
- 원제무 (1999), 「도시교통론」, 서울 : 博英社
- 전명진 (1997), "토지이용패턴과 통행수단선택간의 관계", 「대한교통학
 회지」, 15: 39-49.
- 정병기, 고남욱, 윤장혁 (2016), "딥 뉴럴 네트워크 기반의 특허수명 예측
 모형 개발." 「대한산업공학회 추계학술대회 논문집」, : 981-985.
- 정석환, 정용주 (2018), "DNN을 이용한 오디오 이벤트 검출 성능 비교."
 「한국전자통신학회 논문지」, : 571-577.
- 정진욱, 정진혁 (2006), "베이지안 망, 의사결정나무, 신경망의 수단선택
 모형 적용에 관한 연구", 「대한교통학회 학술대회지」, 53 :
 158-164.
- 조신형, 서영현, 고승영, 이성모 (2017), "개인의 선호다양성을 고려한 지
 역간 수단선택 모형 구축 및 시간가치 추정 연구", 「한국철도학
 회」, 20(2) : 288-298.
- Saito Goki(2017), 「밑바닥부터 시작하는 딥러닝」, 서울 : 한빛미디어
 (주)
- Takayoshi Yamashita (2017), 「그림과 수식으로 배우는 통통 딥러닝」,
 경기 : 제이펍
- Galit Shmueli, Peter C.Bruce, Nitin R. Pate (2017), 「비즈니스 애널리
 틱스를 위한 데이터 마이닝」, 서울 : 이앤비플러스

- Cantarella, G. E., and de Luca, S. (2005), "Multilayer feedforward networks for transportation mode choice analysis: An analysis and a comparison with random utility models." *Transportation Research Part C: Emerging Technologies*, 13(2), 121-155.
- Celikoglu, H. B. (2006), "Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modelling", *Mathematical and Computer Modelling*, 44(7-8), 640-658.
- Hinton G. E., Osindero S. and Teh, Y. (2006), "A fast learning algorithm for deep belief nets," *Neural Comput.* Vol. 18, No. 7, pp. 1527 - 554,
- Hochreiter, S., Bengio, Y., Frasconi, P., and Schmidhuber, J. (2001), "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies", *A Field Guide to Dynamical Recurrent Neural Networks*, 237-244.
- Lai, S., Xu, L., Liu, K., and Zhao, J. (2015), "Recurrent Convolutional Neural Networks for Text Classification.", *In AAAI* (Vol. 333, pp. 2267-2273).
- Liu, and Chen. (2017), "A novel passenger flow prediction model using deep learning methods." *Transportation Research Part C: Emerging Technologies* 84 : 74-91.
- McFadden, D. (1973), "Conditional logit analysis of qualitative choice behavior". *Frontiers in Econometrics*, 105-142.
- Nijkamp, P., Reggiani, A., and Tritapepe, T. (1996), "Modelling inter-urban transport flows in Italy: A comparison between neural network analysis and logit analysis", *Transportation research. Part C, Emerging technologies*, 4(6), 323-338.
- Rao, P. S., Sikdar, P. K., Rao, K. K., and Dhingra, S. L. (1998), "Another insight into artificial neural networks through behavioural analysis of access mode choice", *Computers, Environment and Urban Systems*, 22(5), 485-496.
- Ribeiro, M. T., Singh, S., and Guestrin, C. (2016), "Why should i trust you?: Explaining the predictions of any classifier", *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM*, p. 1135-1144.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014), "Dropout: a simple way to prevent

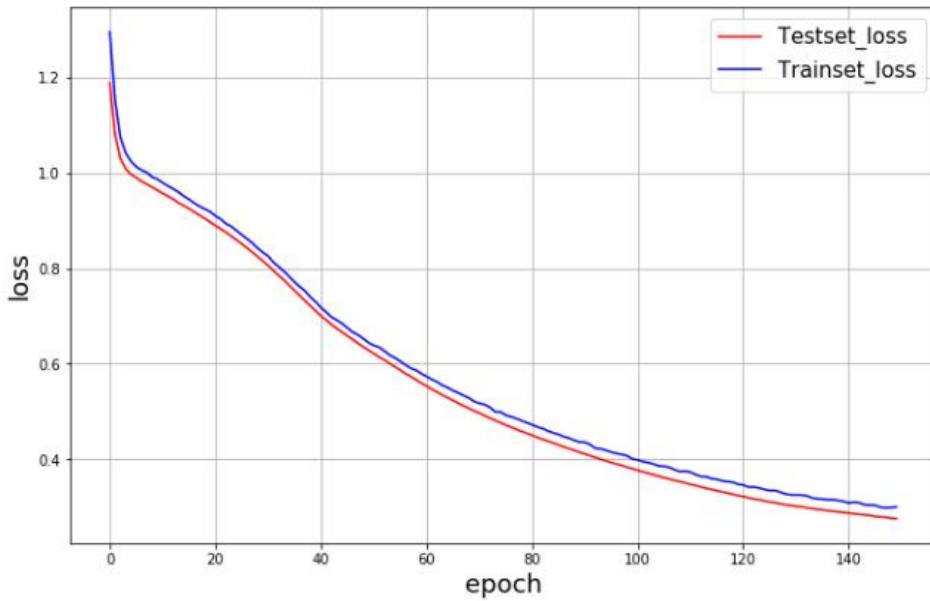
- neural networks from overfitting", *The Journal of Machine Learning Research*, 15(1), 1929–1958.
- Tirachini, A., Hurtubia, R., Dekker, T. and Daziano, R. A. (2017), "Estimation of Crowding Discomfort in Public Transport: Results from Santiago de Chile", *Transportation Research Part A: Policy and Practice*, 103: 311–326.
- Tortum, A., Yayla, N., and Gökdağ, M. (2009), "The modeling of mode choices of intercity freight transportation with the artificial neural networks and adaptive neuro-fuzzy inference system", *Expert Systems with Applications*, 36(3), 6199–6217.
- Treethidtapat, W., Pattara-Atikom, W., and Khaimook, S. (2017), "Bus arrival time prediction at any distance of bus route using deep neural network model." In *Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference on* (pp. 988–992). IEEE.
- Wu, Y., Tan, H., Qin, L., Ran, B., and Jiang, Z. (2018), "A hybrid deep learning based traffic flow prediction method and its understanding." *Transportation Research Part C: Emerging Technologies* 90 (2018): 166–180.
- Xie, C., Lu, J., and Parkany, E. (2003), "Work travel mode choice modeling with data mining: decision trees and neural networks", *Transportation Research Record: Journal of the Transportation Research Board*, (1854), 50–61.
- Yang, W., Chen, B. Y., Cao, X., Li, T., and Li, P. (2017), "The spatial characteristics and influencing factors of modal accessibility gaps: A case study for Guangzhou", *China Journal of Transport Geography*, 60, 21–32.
- Yi, H., Jung, H., and Bae, S. (2017), "Deep neural networks for traffic flow prediction." In *Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on* (pp. 328–331). IEEE.

[부록 A] 심층 신경망 유형별 정확도와 오차 변화

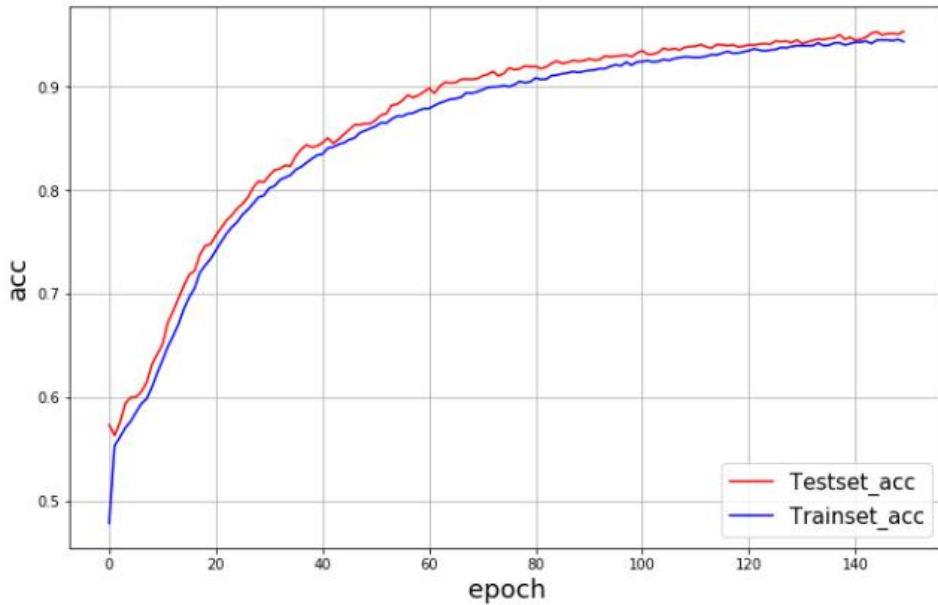
The changes of accuracy by learning on DNN type1



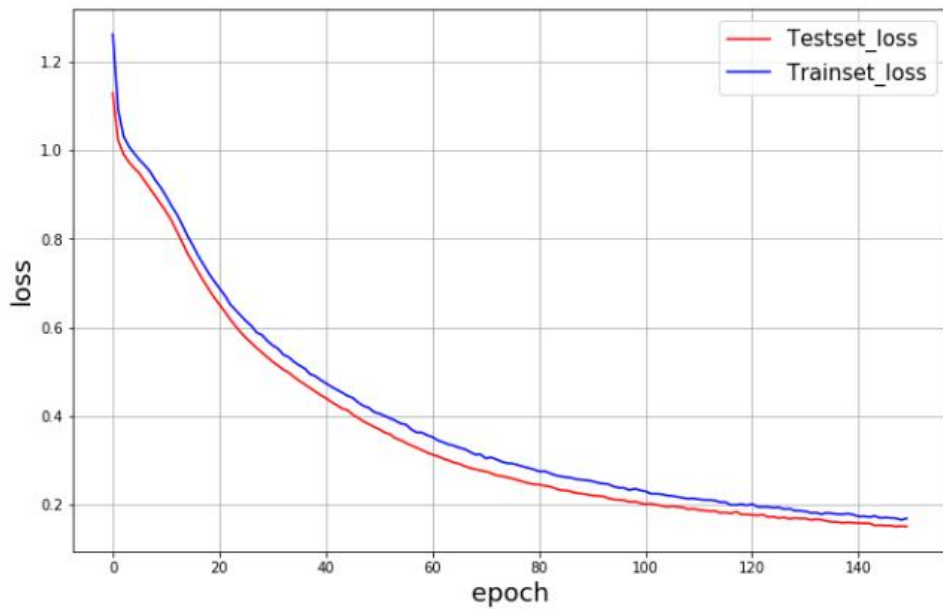
The changes of errors by learning on DNN type1



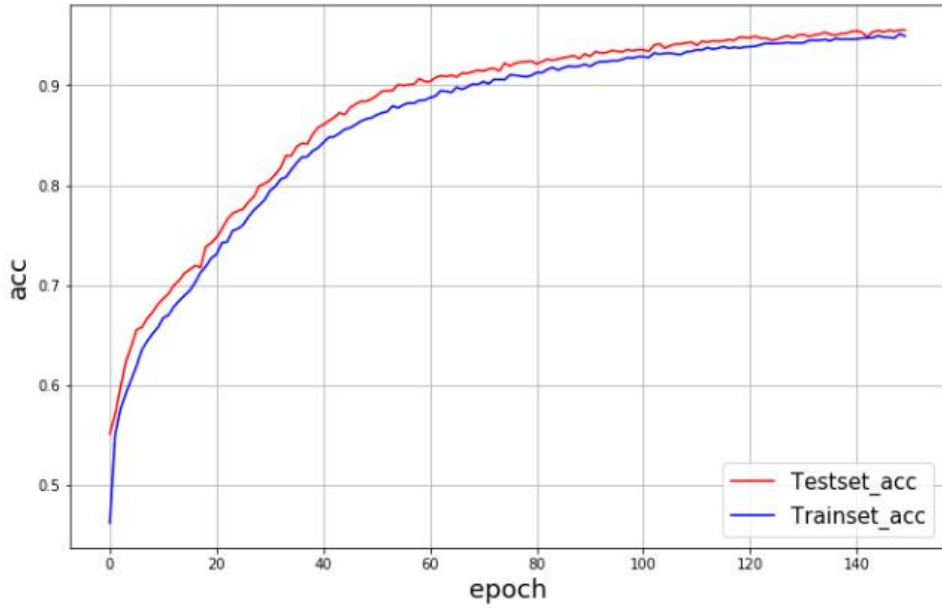
The changes of accuracy by learning on DNN type2



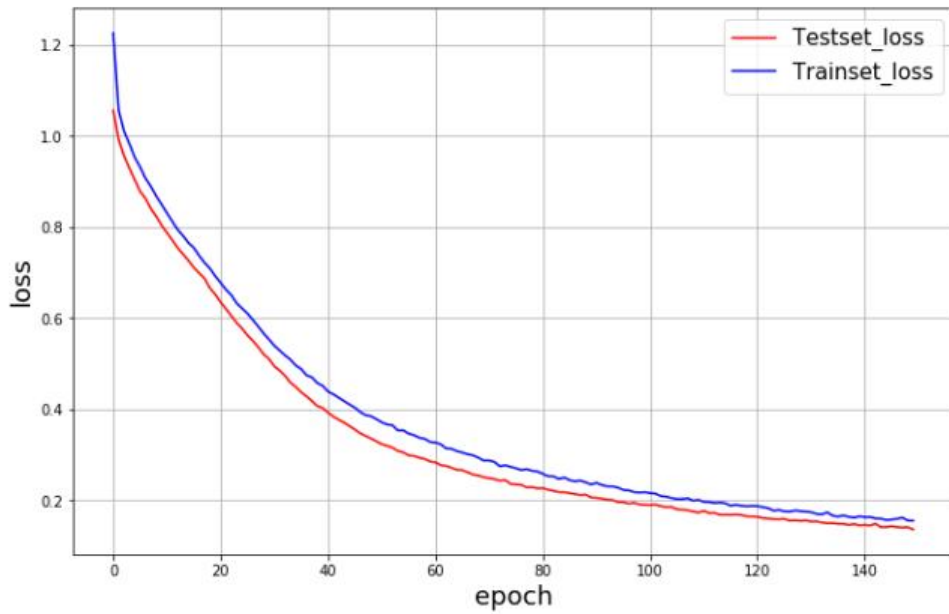
The changes of errors by learning on DNN type2



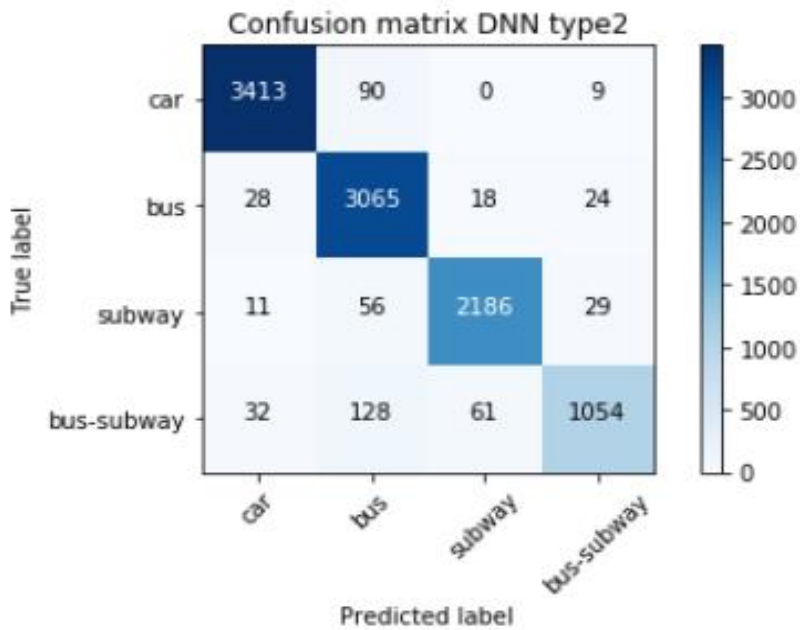
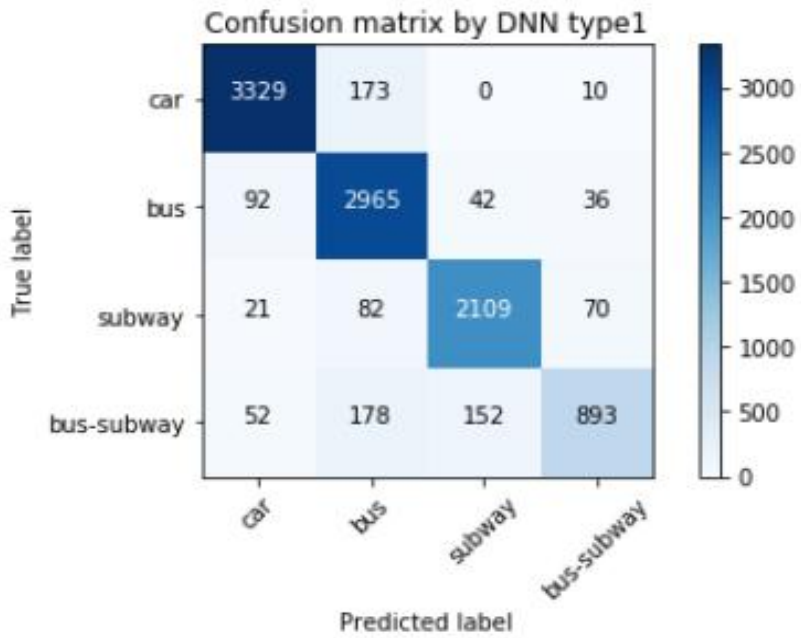
The changes of accuracy by learning on DNN type3

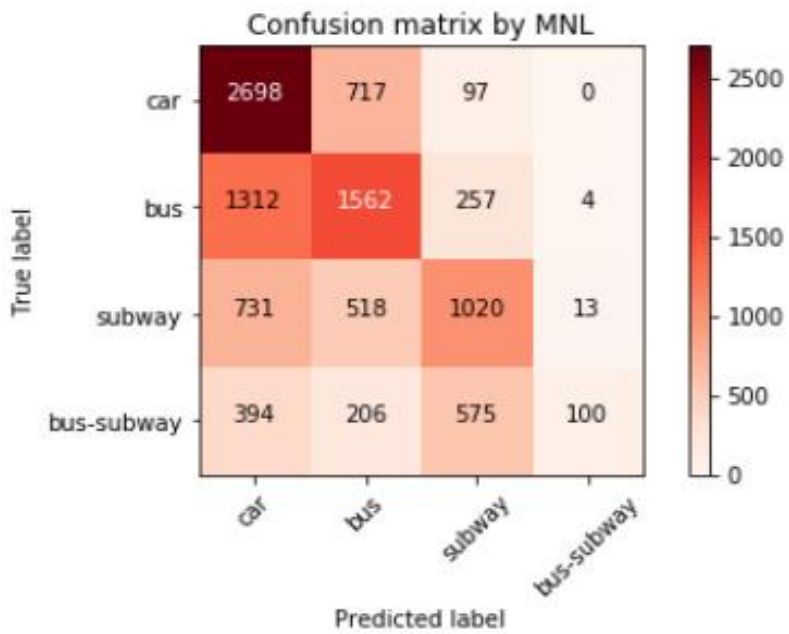
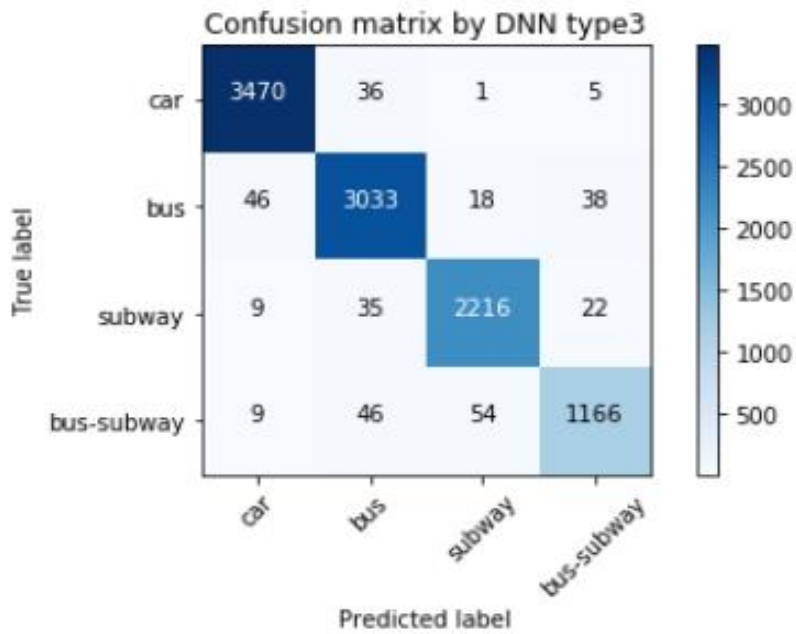


The changes of errors by learning on DNN type3



[부록B] 수단선택 모형별 정오표





[부록 C] 통행시간과 통행비용에 따른 교차 수요 탄력성

[부록 C-a] 통행시간에 따른 모형별 교차 수요 탄력성

| 통행시간에 따른 탄력성 | | 승용차 | 버스 | 전철 | 버스+전철 |
|--------------|-----------|----------------|----------------|----------------|----------------|
| 승용차 | DNN type1 | -0.1892 | 0.0887 | 0.0655 | 0.8831 |
| | DNN type2 | -0.5412 | 0.9264 | 0.04267 | 0.0698 |
| | DNN type3 | -0.0844 | 0.0097 | 0.2616 | 0.3484 |
| | MNL | -0.0228 | 0.0125 | 0.0125 | 0.0125 |
| 버스 | DNN type1 | 0.2638 | -1.9113 | 3.3107 | -1.2826 |
| | DNN type2 | 0.1938 | -0.2267 | 3.1106 | -0.5588 |
| | DNN type3 | 0.3396 | -1.6640 | 2.0011 | -0.418 |
| | MNL | 0.0214 | -0.0365 | 0.0214 | -0.0047 |
| 전철 | DNN type1 | 0.8982 | 2.1921 | -3.4123 | -1.3521 |
| | DNN type2 | 0.4966 | 2.9844 | -2.1245 | -5.6949 |
| | DNN type3 | 1.3328 | 1.7022 | -1.9781 | -0.8825 |
| | MNL | 0.0128 | 0.0128 | -0.0331 | -0.0131 |
| 버스+전철 | DNN type1 | 1.9415 | -0.3352 | -0.0426 | -5.5412 |
| | DNN type2 | 0.8658 | -0.0697 | -0.4384 | -1.5251 |
| | DNN type3 | 3.4529 | -0.8414 | -1.5466 | -1.1154 |
| | MNL | 0.0346 | -0.0241 | -0.0123 | -0.0176 |

[부록 C-b] 통행비용에 따른 모형별 교차 수요 탄력성

| 통행비용에 따른 탄력성 | | 승용차 | 버스 | 전철 | 버스+전철 |
|--------------|-----------|----------------|----------------|----------------|----------------|
| 승용차 | DNN type1 | -0.3731 | 0.0435 | 0.3033 | 1.8044 |
| | DNN type2 | -1.1242 | 0.4809 | 1.5507 | 2.0955 |
| | DNN type3 | -1.0213 | 0.075 | 1.3228 | 0.2189 |
| | MNL | -0.0076 | 0.0042 | 0.0042 | 0.0042 |
| 버스 | DNN type1 | 0.0082 | -0.0333 | 0.2923 | -0.5298 |
| | DNN type2 | 0.0544 | -0.0084 | 0.1207 | -0.3502 |
| | DNN type3 | 0.3879 | -0.0337 | 1.7867 | -1.5412 |
| | MNL | 0.0016 | -0.0019 | 0.0016 | -0.0023 |
| 전철 | DNN type1 | 0.1578 | 0.6649 | -1.0929 | -0.2261 |
| | DNN type2 | 0.0022 | 0.2481 | -0.1033 | -0.2156 |
| | DNN type3 | 0.1646 | 0.2627 | -0.2741 | -0.6279 |
| | MNL | 0.0009 | 0.0009 | -0.0023 | -0.0011 |
| 버스+전철 | DNN type1 | 0.3422 | -0.1316 | -0.0911 | -0.0545 |
| | DNN type2 | 0.0021 | -0.0003 | -0.05844 | -0.1164 |
| | DNN type3 | 0.0059 | -0.0006 | -0.3649 | -0.0937 |
| | MNL | 0.0023 | -0.0013 | -0.0009 | -0.0017 |

Abstract

Estimation of Travel Mode Choice Models using Deep Neural Networks

Daehwan Bang
Transportation Studies
Department of Environmental Planning
The Graduate School of Environmental Studies
Seoul National University

Traditionally, four-stage transportation demand estimation methods were used to predict traffic demand in the transportation plan. The third step has recently become more important due to the introduction of new travel choice modes induced by local governments such as smart mobility, car-sharing and complicated traveller's transit pattern changes following industrial development. Until now, the method used mainly logit models to predict the choice modes of traveller.

However, the Logit model had some problems that were difficult to reflect the diversity of individual preferences and limit of Independence of irrelevant alternatives(IIA). Although prior researches had shown that various studies are conducted to overcome this problem, there are still limitations that cannot fully reflect the

problems of IIA, data scarcity problem, and reflecting various environmental conditions and characteristics of individuals. It is also necessary to ensure whether the existing logit model is responsive to the complexity and diversity of transport choice modes. Also, a new suggestion of mode choice model is needed to complement problems mentioned above all.

Thus, This study suggests travel mode choice models using Deep Neural Networks, which is a type of deep learning, to estimate travel mode choice of traveller that can overcome the limitations of an existing Logit model. The Deep Neural Networks presented that highly correlated variables can be used as input data, unlike Logit model. so This model has some advantages which is possible to adapt convergence of various data and solves the problems of IIA and data scarcity.

This study estimates travel mode choice model for commuter traffic in Seoul using the 2016 household traffic survey data, which is easy to consider simultaneously the socioeconomic and transit characteristics of traveller. The variables in the travel mode choice model were travel time(in-vehicle travel time, out vehicle travel time), travel cost and traveller's socio-economic characteristic(gender, age, income and car possession or not). The age, one of socioeconomic characteristics, was divided into three categories; youth with a discount on transit fares, senior citizens with exemption from subway fares, and the others. The income was reconstructed as average income by the number of households. This study estimates four different choice modes of transportation, including passenger car, bus, subway, bus and subway.

For deep neural networks models, there were three types based on the number of hidden layers to identify the performance of the model

by the number of hidden layers. The models is to be multidimensional classified as the layer increases. This study uses 34,013 samples of passage through which the raw data of the household traffic survey pre-processed and constructed for estimation and verification of the model. Of the total data, 23,809 samples is used as data for learning the model, and remaining 30 percent is used for verification of the model.

As a result of the estimation of the travel mode choice model, the predicted accuracy of the Deep Neural Networks Type3 model was 96.87%, which is the highest of those models and next, Deep Neural Networks Type2 with two hidden layers was 95.24% and Deep Neural Networks Type1 has 91.10% accuracy. Totally, accuracy of the estimated models using deep neural networks was more than 90%.

On the other hand, the generalized logit model showed lower classification accuracy than deep neural networks with 52.72% accuracy. In particular, the results showed very low accuracy of 7.08% for the combined travel mode(bus and subway), indicating that they are vulnerable to complex measures that are difficult to classify.

Demand elasticity of travel time and travel cost was calculated and verified to determine the possibility of generalization through intuitive interpretation of the model. Based on the calculation of elasticity of each model, it was intuitively identified that elasticity of all models were negative, indicating that demand for the travel modes decreased as travel time and travel cost increased. Comparing elasticity of travel time and travel cost by transit modes, it was found that the absolute value of elasticity of the travel time is relatively higher than the travel cost, indicating that the traveller respond more sensitively to travel time. Analysis of the elasticity of the travel time based on

the deep neural networks Type3 with high accuracy in the model indicates that the absolute value of the magnetic elasticity of the passenger car is smaller than transit. This is because traveller taking passenger cars prefers comfort and quick access, even if travel time increases slightly.

The absolute value of elasticity due to travel cost was calculated to be somewhat higher for passenger cars than for public transport. This shows that passersby who use passenger cars are more sensitive to traffic costs than public transport, but the value is not so high that it seems that the person does not rely heavily on travel costs.

This study used an deep neural networks to overcome the logit models' limitations that were mainly used in the existing travel choice model, and confirmed that the deep neural networks model was classified with higher accuracy for various travel modes. In particular, generalized logit model is difficult to categorize complex transit mode, while deep neural networks have more than 96% accuracy. It is expected that the system will be able to be classified with a high probability for vehicles similar to the existing ones such as smart mobility, car-sharing, and trams, as it has confirmed that the diverse means of traffic can be classified. In addition, it is judged that a meaningful study was conducted in that it presented a model for selecting the means to actively respond to various and complex transportation modes.

Because of industrial development, the traffic patterns of passersby will be more diverse, and therefore more variables will be considered to predict traffic patterns of passersby, such as weather characteristics, information characteristics, and real-time environment characteristics.

Therefore, if the travel mode choice model using deep neural networks are applied, which is optimal hidden-layer depth and hyper-parameter settings, travel mode choice model will be able to get the high accuracy for complex and diverse modes of transportation like the result of this study. The Deep Neural Networks could be an alternative to predicting more intelligent and reliable transportation demand in the transportation planning area.

Key Words : Deep Neural Networks, Travel Mode Choice Model, Multinomial Logit Model, Deep Learning, Transportation Planning

Student Number : 2017-26819

사람이 마음으로 자기의 길을
계획할지라도 그의 걸음을 인도하시는
이는 여호와시니라

잠언 16장 9절

이 말씀을 붙잡고 행하여 무사히 졸업할 수 있었습니다.
모든 영광을 주님께 올려드립니다.
항상 응원해주시고 도움 주신 주변의 모든 분들께 감사드립니다.

2019년 2월
방대환