

신경망(Neural Network)모델들의 비교를 통한 가맹사업 매출예측분석

Franchise Business Sales Forecasting by comparison of Neural Network models

김주영 · Kim, Juyoung

최근의 경영이슈는 기업경영에 있어서의 사회적 가치의 중요성이다. 그중에서 유통산업, 특히 가맹사업에서는 가맹본부와 가맹점 간의 관계에 있어서 사회적으로 상생할 수 있는 방법을 모색하려는 노력이 많다. 가맹사업은 경쟁력이 검증된 사업아이템을 일정한 비용을 받고 창업희망자들에게 제공하고 이들이 계속해서 사업이득을 얻을 수 있게 도와주는 것이다. 가맹본부가 가맹점주 들 혹은 미래의 가맹점주 들에게 올바른 의사결정을 위한 정보를 알려주지 않고 무리한 요구를 해왔다고 사회적 이슈가 되고 있다. 문제를 해결하기 위해서는 관련된 정보를 정확히 알아야 하는데, 본사의 경쟁력이나 가맹점의 경영능력 등을 알기가 어렵고, 그들이 처한 경제상황적인 변수도 제대로 고려하지 않은 정보들은 실제 의사결정에는 크게 도움이 되지 않는다.

본 연구는 간단하지만, 보다 정확한 의사결정을 위해서 가맹본부와 가맹점포의 매출액을 설명하는 신경망 모델들을 전통적인 회귀분석과 함께 비교 제안하였다. 공정거래위원회의 가맹사업 정보공개서에 있는 3개년간의 자료들을 이용해서 외식, 서비스, 도소매로 가맹사업을 구분하고, 매년 23개씩의 반복되는 변수들을 독립변수로, 가맹점평균매출과 가맹본부매출을 종속변수로 두었다. 총 2459개의 표본 중 1967개를 학습표본 492개를 검정 표본으로 삼았다. 2개년도의 자료로 3개년도차의 종속변수들을 예측하는 4개 모델들을 비교 제안하였다. 전통적인 선형회귀분석과 간단한 신경망, 은닉층이 4개있으며 산업 dummy가 처음과 중간 은닉층에 직접 영향을 미치는 깊은 은닉층, 앞의 년도의 은닉층이 뒤의 년도 은닉층에 영향을 미치는 Recurrent 신경망을 RMSE와 MAPE로 비교했는데, 깊은 신경망이 가장 좋은 결과를 보였다.

또, 신경망의 결과 중 변수중요도를 계산하는 새로운 방법을 두 개 제안하였다. 변수 중에는 가맹점매출에는 본사직원수, 신규개점이나 폐점 등 가맹점수, 과거가맹점매출평균 또 본사의 자산규모, 자본, 부채가 영향을 미치고 있다. 이외에도 기존의 가맹본부를 재무상태등을 기준으로 판단할 때 쓸 수 있는 방법과 새로 가맹점을 하려고 할 때 가맹본부들을 평가하는 방법들을 제시하였다.

핵심주제어: 신경망, 가맹사업, 가맹점매출, 모델비교, 변수중요도

김 주 영 | 서강대학교 경영대학 교수(jkimsg@sogang.ac.kr)

ABSTRACT

Social value in corporate management is getting attention. The retail industry, especially the franchise business, is eager to find ways to socially coexist in the relationship between franchisers and franchisees. The franchise business is to provide franchisee entrepreneurs with the proven business know-how at certain amount of fee and to help them continue to gain business benefits.

It is claimed that the franchiser has made unreasonable pushes without properly informing the franchisee or future franchisees for making the right decision. One needs to know the exact information in order to solve the problem. It is difficult to know the real competitiveness of the headquarters and the management ability of the franchiser.

In this study, neural network models that explain the sales of the franchiser and the average sales of franchisee stores are proposed and compared with traditional regression analysis in order to enable more accurate decision making. Using the three years' data in the Fair Trade Commission's Franchisee Information Disclosure Documents, we divide franchise into service, food service, and retail industries, and 23 recurring variables are used as independent variables.

Of the total of 2459 samples, 1967 were used as training samples and 492 samples as testing samples. The traditional linear regression analysis, and the simple neural network, and the deep neural network with four hidden layers, and the recurrent neural network were used to compare on RMSE/MAPE of their predictions. It was judged that the optimal model is somewhere between the simple neural network and the deep neural network.

In addition, two new methods were proposed to calculate the importance of independent variables of neural networks. Among the variables, the number of employees, asset size, capital and debt of the head office are affecting the sales of franchiser. In addition, it suggested methods to judge existing franchiser based on their status of independent variables, and methods to evaluate the franchiser when attempting to open a new merchant for entrepreneur.

Keywords: Neural Network, Franchise Business, Franchisee Sales, Model Comparison, Sensitivity Analysis

1. 서론

경영과 마케팅의 다양한 이슈들이 많지만 최근에는 효율성뿐만 아니라 사회적 가치에 대한 중요성이 많이 부각되고 있다(김인숙, 리상섭 2018). 그중에서도 유통산업의 가맹사업은 가맹본사와 가맹점을 포함한 2단계로 비즈니스모델이 형성되어있다. 가맹사업에서의 사회적가치의 문제는 전통적인 기업경영에서 과도한 이윤을 남겨서 소비자에게 올바른 가치를 제공하지 못하는 것에 대한 것이 아니라, 가맹사업이라는 비즈니스모델 내에서의 올바른 가치교환이 이루어져야 한다는 것이다. 이를 위해서 올바른 가치가 측정되어야 하는데, 이것이 매우 어려우며, 잘못된 가치측정은 가맹본부와 가맹점사업자의 의사결정에 많은 문제를 야기 시킨다.

교환 가치측정의 첫 번째 문제는 가맹점사업자가 가맹본부에 지불해야 하는 가맹비 혹은 로열티, 혹은 상품공급대금이나 가맹본부가 강제하는 인테리어비용, 광고비용 등 각종비용의 올바른 산정이라고 할 수 있다(임영균, 김주영 2018). 이러한 비용이 적정한지를 판단하기 위해서는 해당비용들을 다른 경쟁사들은 어떻게 책정하고 있으며, 또 실제원가에 적절한 이윤을 붙이고 있는지를 알아야 한다. 하지만, 가맹본부의 실제원가를 외부에서 알기도 힘들며 이는 기업의 고유경쟁력과 연관되어 있으며, 이를 경쟁사들과 직접적으로 비교하기도 어렵다. 또 가맹본부별로 각종비용들의 이름도 다르고, 비용들을 받는 방법들도 다양하기에 정확하게 파악하기는 매우 어렵다.

현실적으로 가능한 방법으로는, 위의 다양한 수익과 비용이 포함되어 있는 가맹본사의 매출이나 순이익을 비교하는 방법이다. 가맹본부가 올바르게 회계보고를 한다면, 외부에 공개된 자료들을 활용해서 가맹본부들의 실제 매출과 가맹본부의 경영상태 등 정보를 활용하여 예

측한 매출과 비교할 수 있다. 비슷한 수준의 가맹본부보다 매출이 높다면 가맹본부의 경쟁력이 높거나 혹은 과도한 비용을 받고 있다고 할 수 있을 것이다.

잘못된 가치측정의 두 번째 이슈는 미래의 가맹점사업자가 가맹본부를 선택할 때 정확하지 않은 가치교환의 약속이나 정보를 보고 선택하게 되는 것을 말한다. 즉 첫 번째 핵심이슈는 현재 가맹점주와 가맹본부의 문제였다면, 두 번째 핵심이슈는 가맹점 사업희망자가 올바른 가맹본부를 선택할 때 발생하는 문제이다. 가맹점 사업희망자는 대부분 경력이나 정보에 있어서 가맹본부보다 뒤떨어질 수밖에 없기에, 브랜드의 명성이나 제시된 창업비용만 보고 선택하는 경향도 있다(성대권 등 2012; 송병화 등 2006; 김성태 등 2014). 하지만, 선택후 실제 가맹점 운영결과가 기대보다 낮으면 이를 받아들이기가 어려운 것이다. 만약 올바른 회계정보를 알 수 있다면, 가맹본부의 매출을 가맹본부의 규모 등 경영수준요소들로 평가해보고, 또 가맹점의 매출을 가맹점의 규모 등 수준으로 평가한 후에, 가맹본부의 경영수준에 따른 가맹본부의 적정매출과 가맹점의 수준에 따른 적정 가맹점매출을 찾아낼 수 있다. 비슷한 수준의 경쟁 가맹본부 매출과 비교해서 특정 가맹본부의 실제매출이 이유 없이 높다거나, 혹은 가맹점예상매출이 유사가맹점들보다 현저히 낮다면 해당 가맹본부를 선택할 이유가 적어지게 될 것이다.

하지만, 올바른 가맹본부와 가맹점들의 매출정보를 알기 어렵고, 또 가맹본부와 가맹점들의 경영수준정보들도 얻기 어렵다. 공개된 정보를 통하지 않고 상황에 따른 설명과 해석 및 단기적인 여론몰이를 가지고는 올바른 산업정책을 세우기도 어렵다. 지금까지 공정거래위원회는 가맹사업 정보공개사업을 펼치고 있으며, 모든 가맹본부는 정해진 양식에 따라 공정거래위원회에 사업정보를 제공함으로써, 현재의 가맹점사업자들과 미래의

가맹점사업자들에게 올바른 정보를 제공하고, 분쟁이 생겼을 시에는 한국공정거래조정원, 가맹거래사, 분쟁조정협의회 등을 두어 해결하고자 하는 노력을 하고 있다 (<http://franchise.ftc.go.kr/main/index.do>).

하지만, 정보공개서의 내용은 모두 채워져 있지도 않고, 내용도 제3의 기관이 검증한 것도 아니며, 또 공개된 것도 최근 3년 정도 밖에는 없다. 그럼에도 불구하고, 본 연구는 정보공개서의 자료를 가지고, 가맹본부의 매출과 가맹점평균매출을 추정해보고자 한다. 가맹점 개별점포들의 정보는 없기에, 가맹점 수준에 따른 분석은 하지 못했다.

가장 간단한 분석은 전통적인 회귀분석을 통해서 가맹본부의 매출을 가맹본부의 재무 상태를 포함한 여러 정보들, 독립변수들을 가지고 예측하는 것이며, 또 가맹점 평균매출도 회귀분석을 통해서 동일한 가맹본부의 독립변수들을 가지고 예측하는 것이다. 하지만, 전통적인 회귀분석은 독립변수들 간의 상관관계가 적어야 하지만, 특히 여러 기간의 자료가 축적된 경우에는 유사한 독립변수들이 많이 생기게 되어서 효과적인 방법이라고 하기 어렵다. 따라서 보다 복잡한 시계열분석방법이 필요한데, 이러한 방법들은 자료에 대한 가정도 많고, 모델의 형태에 따라서 결과가 많이 달라질 수 있으며, 특히 추정된 계수들을 해석할 수 있는 하나 이상의 방법들이 존재할 수 있다.

비교적 최근에 주목받기 시작한 신경망(neural network) 방법은 인공지능(Artificial intelligence)시대에 대표적인 방법론이다. 신경망 연구는 오래전에 시작되었지만, back-propagation이라는 계수추정방법이 개발되면서 다양한 문제를 해결하는데 활용되고 있다. 특히 빅데이터(Big data)시대를 맞이하여 경영환경에서 얻어질 수 있는 다양한 문자정보와 그래픽정보, 소리정보들을 처리하는데 매우 많이 활용되고 있다. 하지만, 신경망방법은

이렇게 새롭게 처리되기 시작한 정보들 이외에도 지금까지 쌓아져왔던 경영의 많은 계량정보들을 처리하는데도 효과적일 수 있는 방법론이다. 본 연구에서는 신경망모델을 통하여, 정보공개서에서 얻을 수 있는 가맹본부의 경영정보들을 이용해서 가맹본부의 매출과 가맹점평균매출을 추정해보고자 한다. 간단한 신경망모델부터 깊은 신경망모델과 Recurrent 신경망모델을 제안할 것이며, 이것들과 전통적인 회귀분석모델을 예측력을 기준으로 비교할 것이다. 또한 신경망모델이 가지고 있는 설명력의 부족에 도움이 될 수 있는 새로운 변수중요도 측정방법을 제안하고자 한다. 마지막으로 보다 정확한 예측 결과들을 가지고 현재의 가맹본부를 평가하는 방법과 새로운 가맹점사업자의 입장에서 좋은 가맹본부를 선택할 때 활용할 수 있는 방법을 제안하고자 한다. 본 연구는 특정가맹본부에 대한 전략적인 판단을 하거나 해당 본부들을 평가하고자 함이 목적이 아니며, 전략 수립과 선택에 보다 도움이 되는 자료분석 방법을 제안하는 것이다.

II. 신경망 모델과 가맹점자료

1. 신경망모델

신경망의 시작을 어디에서부터 보는지는 학자마다 신경망이라는 개념을 정의하는 방법의 차이에 따라서 다르지만, McCulloch and Pitts (1943) 혹은 Rosenblatt (1958), 또 Hebbian learning (Hebb 1949) 등의 초기의 모델개발 시기를 거치면서 많은 관심을 받기도 했지만, Minsky등 (1969)에서 그 당시의 한계점을 발표한 이후에는 오랫동안 활용되지 못하던 방법론이었다 (Schmidhuber 2014; Agatonovic-Kustrin and

Beresford 2000). 하지만, Rumelhart 등(1986)이 미분의 chain-rule을 활용한 Backpropagation 방법을 사용하면서 더욱 복잡하고 여러 은닉층(hidden layer)이 있는 신경망이 개발되었다(Hinter 2007).

신경망분석은 인간의 신경망(neuron)이 자극을 받아서 뇌로 전달하는 것을 수리적으로 모방하여 만든 것이라고 알려져 있다(Agatonovic-Kustrin and Beresford 2000). 즉, 아래 <그림 1>의 우측 그림이나 <그림 2>처럼 피부에 있는 신경망에 자극이 오면 그것이 그 다음 신경망에 연결되고, 그것이 계속되어 뇌에 까지 이르는 것과 유사하게 되어있다. 인간의 신경망처럼 받은 자극(자료)가 어떤 경우에는 더욱 커져서 전달이 되기도 하고, 어떤 때는 제대로 전달이 되지 않기도 하기에 그런 것을 그림들에서 보이는 연결선에 있는 계수 혹은 가중치로 크기를 조절하고, 그다음 신경망으로 가기 전 활성화함수(activation function)이라는 곳에서 연결하면서 비선형적으로 변형시키기도 하고 아예 끊어버리기도 하게 만들어졌다. 전통적인 회귀분석의 관점에서 보면, 여러 회귀분석 혹은 로지스틱회귀분석이 연결되어 있는 것으로도 볼 수 있다.

보통 깊은 신경망모델(deep neural network, 혹은 더 넓은 개념으로 딥러닝, deep learning)은 연결된 신경망의 수가 2개 이상인 경우를 의미하는데, 자료를 받는 부분과 결과를 내보내는 부분을 제외하고 가운데 있는 층을 은닉층(hidden layer)라고 한다. 깊은 신경망에는 주로 그림정보를 처리하는 컨볼루션 신경망(Convolution Neural network), 시간적으로나 의미적으로나 공간적으로 연결되어있는 자료를 처리하는 순환 신경망(Recurrent Neural network), 자료안의 구조를 파악하는 Restricted Boltzmann Machine(RBM), 현재의 상태에 기초하여 최적의 상태를 찾아가는 길을 찾는 강화학습(Reinforcement Learning) 등 많은 모델

들이 있다(Sutton and Barto 2017; Schmidhuber 2014; Bengio 2009).

신경망은 지금까지의 비정형자료라고 불리는 문자정보, 그림정보, 소리정보들을 분석하는데 많이 활용되어 왔으며, 마케팅에서도 고객댓글이나 SNS에서 자주 언급되는 단어나 내용에 대한 분석, 또 우수고객 확인 등에 많이 쓰이고 있다. 크게 보면, 신경망은 빅데이터를 분석하기 위해 활용되고 있는 Data Analytics의 다양한 마이닝(mining) 방법론의 한 가지 방법이라고 생각할 수도 있겠다.

마케팅에서의 빅데이터는 비정형자료뿐만 아니라 숫자로 표현되는 많은 자료들도 회사의 내부의 거래기록이나 소비자들과의 접촉기록, 또 관련된 경기나 날씨를 포함한 외부의 다양한 자료들이 해를 거듭할수록 쌓이면서 만들어지게 된다. 한글을 포함한 문자자료나 비정형자료들도 많지만, 수량으로 되어있는 빅데이터를 통해서 보다 정교한 정보를 만들어낼 필요가 있다.

본 연구에서는 가맹사업 관련된 올바른 의사결정을 위해서 보다 정교한 가맹본부매출과 해당본부 산하의 가맹점평균매출을 찾는 모델을 제안하고 이들을 비교해보고자 한다. 서로 다른 모델들을 앙상블(ensemble)해서 보는 것도 또 다른 모델로 생각해볼 수 있지만, 본 연구는 가장 높은 예측력이 있는 것을 찾는 다기 보다는 각 모델에 대한 특징을 비교하고, 가능성들을 검토하기에 따로 앙상블모델을 검토하지는 않았다.

2. 회귀모델과 간단한 신경망 모델

먼저 가장 간단한 회귀분석을 통해서 분석해보고자 한다. 종속변수는 가맹본부의 매출과 가맹본부의 매출이고, 독립변수는 정보공개서에서 얻을 수 있는 자료들을 가능한 모두 활용하였다. 회귀분석은 다른 모델들의 평

가기준이 되며, 많이 활용하는 SPSS 통계패키지를 이용하였다.

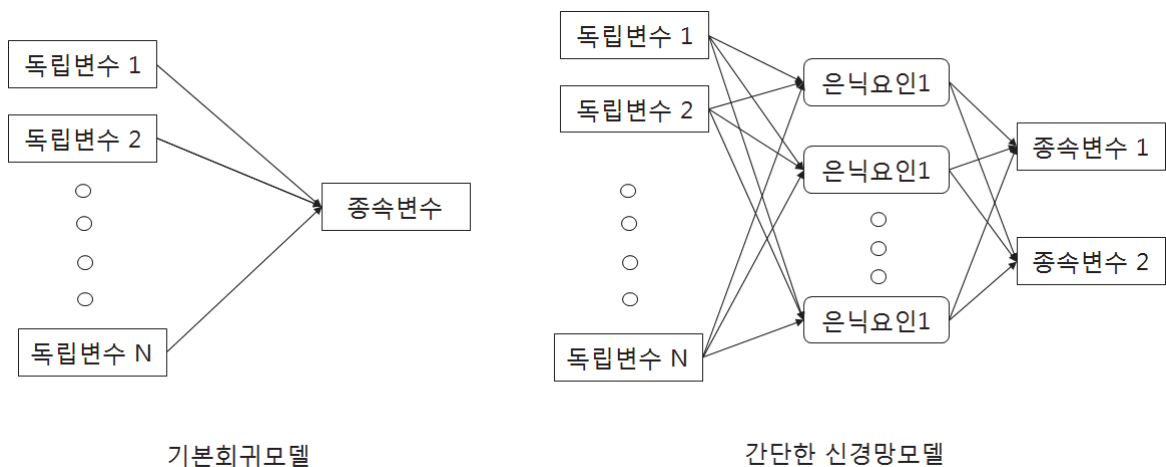
두 번째로는 간단한 신경망을 적용하고자 한다. 이것도 SPSS 통계패키지(IBM SPSS Statistics ver25.)가 제공하는 다층 퍼셉트론을 이용하였다. 아래 <그림 1>에서 보는 것처럼, 회귀모델에 비해서 간단한 신경망모델은 독립변수와 종속변수사이에 은닉층을 두어서 독립변수의 값들을 각 은닉요인마다 다른 가중치를 곱해서 합한 후에 이를 다양한 S 모양의 활성화함수를 통해 변형시킨 후 다시 이들 은닉요인들의 값을 가중치를 곱해서 합한 후 적절한 activation 함수를 통해 종속변수 예측 값을 만들어 낸다. 회귀식과 다른 점은 은닉층이 있다는 것과 활성화함수를 통해서 일반화선형모형(generalized linear model)과 같이 비선형효과를 만들어 낸다는 것과 종속변수가 한 개 이상으로 다양해도 된다는 것이다.

3. 깊은 신경망 모델과 순환(Recurrent) 신경망 모델

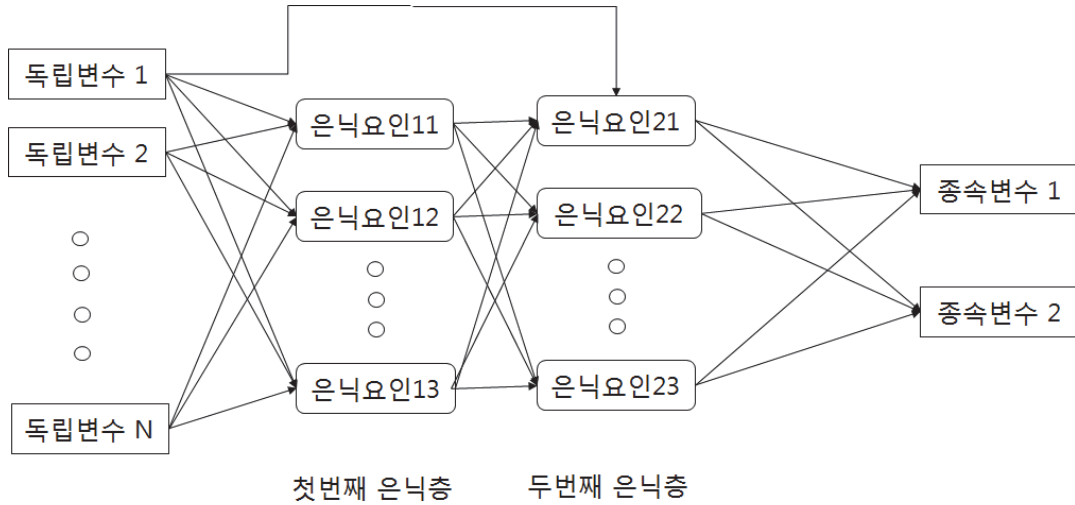
세 번째 깊은 신경망 모델은 <그림 2>와 같이 간단한 신

경망 모델보다 은닉층을 더 넣었으며, 조금 변형을 시켜서, 독립변수 중 일부의 영향력이 첫 번째 은닉층에도 미치지만, 두 번째 은닉층에도 직접 영향을 미치는 모델을 제안했다. 본연구에서는 독립변수들 중에서 특히 가맹사업 중 외식, 서비스, 도소매 등 산업별로 차이가 있으리라고 생각하고 그것이 중요하다고 생각했기 때문이다. 은닉층이 많은 딥러닝(Deep learning) 모델들은 모델의 계수추정에 필요한 chain rule에 따른 미분값들이 작아지는 기울기 값의 소실(vanishing gradient) 현상이 나타나서 종속변수의 추정 값을 실제 값과 일치시키기 위해서 계수추정값을 조금씩 변화해가는 학습의 효과를 없애기 때문에, 다양한 기법들을 필요로 하게 된다. 예전부터 많이 사용하던 최적화(optimization) 기법들 이외에 학습속도를 정교히 조정해주는 adam optimizer(Kingma and Ba 2015)라든가 계수추정치 초기 값을 작은 값으로 만들어주는 Xavier 초기 값 방식(Xavier and Bengio 2010). 이외에도 많은 은닉층사이의 연결선들(추정해야 할 계수들)이 과적합(overfitting)이 되도록 할 수 있기에 regularization이라는 이슈가 발생한다. 이를 위해서 계수추정치가 작

<그림 1> 회귀모델과 간단한 신경망모델



〈그림 2〉 깊은 신경망모델



은 것들을 0으로 만들어주는 방법들과 아예 추정을 하면서 이들 중 일부를 없는 것처럼 생각하고 나머지 것들만 추정하는 drop out 방법들도 활용되고 있다(Srivastava et.al. 2014).

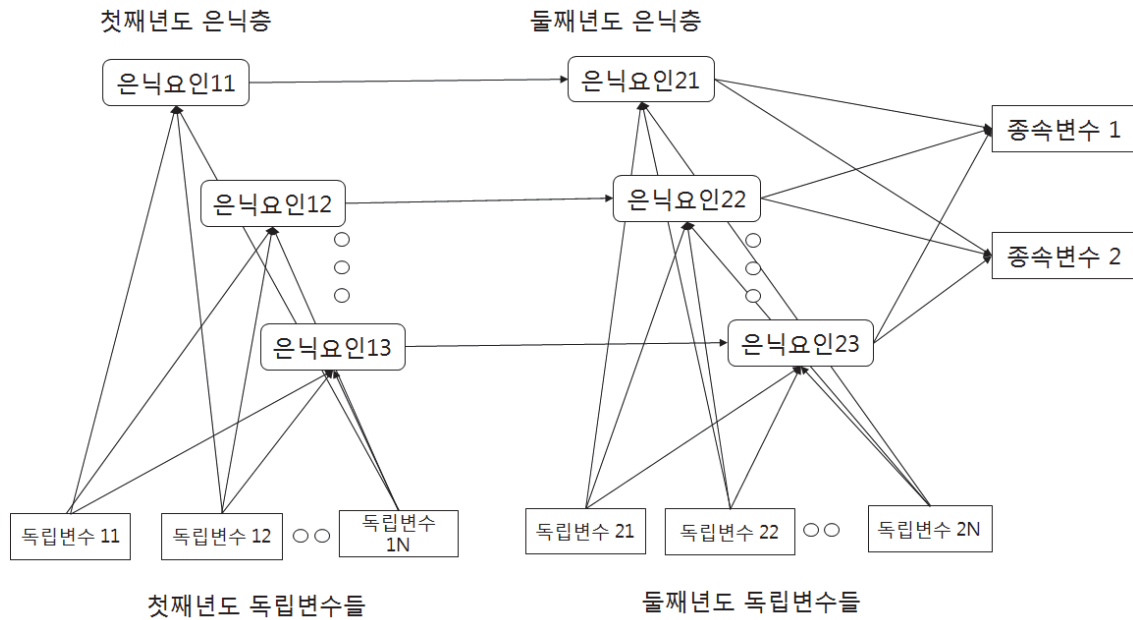
이런 방법들은 빅데이터를 위한 data analytics의 방법들에서도 적용되어왔고 또 이를 기반으로 발전된 것들이다. 본 연구에서는 Xavier초기값 설정법, drop out 방법, adam optimizer등을 활용하였으며, 은닉층도 4개까지 설정하였다.

네 번째 방법은 〈그림 3〉에 나타난 것과 같은 Recurrent 신경망 모델이다. Recurrent 신경망 모델은 동일한 자료나 문제가 한번만 발생하는 것이 아니라 계속해서 발생하며, 앞에서 발생한 것이 다음번에 발생하는 것에 영향을 미치는 경우에 활용 할 수 있는 방법이다. 경영학에서는 시계열(time series) 모델이나 vector autoregression (VAR) 모델과 유사하다고 할 수 있다. 하지만 순환 신경망모델은 컨볼루션 신경망모델과 마찬가지로 모든 변수와 은닉층간의 관계를 매번 새롭게 하는 것이 아니라 매번 같은 계수추정치를 사용함으로써 추정해야 할 계수의 숫자를 줄여주는 것이 기본적인 목적이다.

〈그림 3〉에서 보면 동일한 독립변수1, 2, N이 첫 번째 년도에도 나오고, 두 번째 년도에도 나온다. 따라서 첫 번째 독립변수들로 만든 은닉요인11, 12, , 1N을 만들고, 두 번째 년도에도 역시 두 번째 년도의 독립변수들로 은닉요인 21, 22, , 2N을 만드는데, 첫 번째 년도에 만든 은닉요인들이 두 번째 년도의 은닉요인을 계산할 때 영향을 미치게 되는 것이다. 만약 연도별 자료가 또 있다면, 세 번째 년도와 네 번째 년도까지 확장할 수 있을 것이다. 이때 각각의 계수들은 연도별로 동일하다고 가정하고 추정을 하게 된다. 본 연구를 위해서는 두 번째 년도만 있는 모델을 세웠으며, 종속변수도 세 번째 년도의 가맹본부매출과 가맹점평균매출을 추정하게 될 것이지만, 첫 번째 년도와 두 번째 년도의 가맹본부매출과 가맹점평균매출을 각각의 종속변수들로 추가해서 모델을 세울 수도 있다. 가장 많은 순환 신경망모델의 활용은 문자자료 분석이라고 할 수 있으며, 금융 쪽에서 주가 등을 예측하는 곳에서도 많이 활용되고 있다.

신경망모델은 다양하게 변화시킬 수 있고, 새로운 형태의 모델도 계속 개발되고 있으나 아직 SPSS와 같이 손쉽게 하는 방법보다는 code를 직접 프로그래밍하게

〈그림 3〉 Recurrent 신경망 모델



되어있다. 본 연구에서는 세 번째와 네 번째 신경망모델은 구글에서 개발되어 2015년부터 공개된 open source 소프트웨어인 tensorflow(<https://www.tensorflow.org/>)를 통해서 모델을 추정하였다.

4. 모델적용을 위한 가맹자료

자료는 공정거래위원회의 가맹사업 정보공개서를 활용하였다(<http://franchise.ftc.go.kr/user/extra/main/69/firHope/listIndus/jsp/LayOutPage.do>). 2008년 8월부터 모든 가맹사업을 하고 있는 사업자들은 자신들의 사업내용을 공정거래위원회가 지정한 양식에 따라 공개하도록 되어있다. 하지만, 과거자료는 현재 열람이 안되며, 본연구가 진행되던 때에는 3개년도 2015, 2016, 2017년도가 열람 가능하였다. 따라서 3개년도의 열람된 가맹 브랜드 중에서 3년간 모두 존재했던 가맹브랜드들만 선별하고, 다시 재무자료가 모두 없거나 입력된 자

료의 단위가 상식적이지 않은 자료들을 제외하고, 최종적으로 2459개의 가맹브랜드의 3년간 자료를 활용하였다. 참고로 2015(2017)년에는 4844(5700)개의 가맹브랜드가 등록되어 있다.

본 연구에 활용되는 변수들은 〈표 2〉에서 보듯이 매년 23개의 변수들의 값이다. 이들은 가맹브랜드별 점포수의 변동과 가맹본부임직원수, 가맹점을 개설할 때 요구하는 각종 비용들, 가맹본부의 여러 가지 재무상태 변수 및 범위반횟수 등을 포함하고 있다. 이외에 가맹브랜드의 사업개시 일에서부터 2017년까지의 사업개월 수를 환산해서 사용하였으며, 각 가맹브랜드들이 속해있는 산업을 크게 3가지로 구분하여 더미(dummy) 변수 화하여 활용하였다.

전통적인 통계모델들과 마찬가지로 신경망모델들은 결측치를 원래 처리할 수 없고, 아직 신경망모델들에 더 최적화된 결측치 대체방법이 개발되지 않아서, 본 연구에서는 산업별로 각 변수마다 중앙값(median)을 찾아

서 대체하였다. 그 이후에도 변수들의 분포형태가 heavy tail이고 편향이 강하여서 로그치환을 하였으며, 최종적으로 평균표준편차를 통한 표준화 값을 사용하였다.

연구에 활용된 변수는 연구의 목적이 가맹본부매출과 가맹점평균매출을 예측하는 것이기에 회귀분석에서는 2015년과 2016년 변수들 46개(해마다 23개변수)와 산업 dummy변수 3개들 간의 종속성 때문에 2개만을 활용해서 2017년 가맹본부매출과 가맹점매출을 예측하였다. 간단한 신경망모델과 깊은 신경망모델에서는 변수들 간의 종속성이 문제가 되지 않기 때문에 총 49개변수를 모두 활용하였다. 순환 신경망에서는 매 년도에 들어가는 변수의 수가 동일해야 하기에 2015년도에도 더미 3개와 23개 변수, 2016년도에도 동일한수의 변수가 들어가서 총 52개변수가 독립변수로 들어갔다.

III. 제안 모델들의 비교

1. 모델들의 결과 적합도 비교

모델들을 평가하는데 먼저 살펴보는 것은 자료를 모델이 얼마나 잘 설명하는지를 적합도 또 그에 따른 예측력관점에서 판단하는 것이다. 전통적인 통계모델들을 비교하는데 활용되었던 AIC기반의 통계치나 R2와 같은 설명력기반 통계치들, 또 모델 안에 활용된 파라미터들의 수를 기반으로 간명한 모델을 선호하게 한 통계치 들은 확률모형이 아닌 빅데이터를 통한 계량모델들에게는 큰 의미가 없어졌다. 지금도 많은 개발이 되고 있겠지만, 보편적으로 쓸 수 있는 것은 root mean square error (RMSE) 통계치이다. RMSE는 종속변수 예측치와 실제값을 차이에 자승을 한 후에 평균을 구한 후 제곱근을 한 값이다. 또 많이 활용하는 값이 mean absolute percent

error(MAPE)으로 이것은 예측치와 실제 값의 차이의 절대 값을 실제 값으로 나눈 것이다(Ryu et.al. 2017). MAPE는 실제 값이 0이 되면 쓸 수 없는 단점이 있지만, 많이 활용되고 있다.

이때 추정해야 할 계수(파라메타)의 숫자를 많이 하면, 과적합이 되며, 이러한 증상은 전체 자료의 수가 아무리 크다고 할지라도 발생할 수가 있다. 과적합이 되면 모델을 추정하는데 사용되었던 학습자료(training data)에서는 좋은 낮은 RMSE이나 MAPE를 보이지만, 학습에 활용하지 않은 검정자료(testing data)에서는 높은 RMSE/MAPE를 보이게 된다.

더 나아가서 data analytics의 많은 방법들에서는 모델의 기본구성 모습은 동일해도, 은닉요소수를 늘리거나 추정할 때 최적화에서 iteration 수를 늘려서 더 많은 학습을 하도록 하거나, 자료가 큰 경우에 학습 자료를 한 번에 사용하지 않고, 작게 batch로 구분해서 하는데, 이때의 batch size와 batch에 따른 iteration 수(전체 자료 / batch size), 또 학습 자료를 한번 뿐(epoch가 1)이 아니고 N번 활용(epoch가 N)할 때 몇 번을 할 것이냐 등 모델의 meta 파라메타들을 조정해서 검정자료에서도 가장 낮은 RMSE를 만드는 모델을 찾기 위해 노력하게 된다. 이럴 때에는 검증자료뿐만 아니라 여기에도 활용하지 않은 검증자료(validation data)의 RMSE/MAPE를 비교하기도 한다. 하지만 본 연구에서는 회귀 분석에서는 이것이 필요 없으며, 다른 신경망모델들에서도 meta파라메타를 통한 최적화를 시도하지 않았기에, 검정자료 RMSE/MAPE까지만 활용하기로 한다.

2459개 전체표본에서 20%인 492개를 검정표본으로 하였다. 결과는 다음 <표 1>에 나타나 있듯이, 가맹점평균매출에서는 전통회귀분석과 간단한 신경망이 검정표본 RMSE에서 좋은 결과를 보여주고 있으며, 학습표본 RMSE에서는 역시 깊은 신경망과 순환 신경망이 좋은

〈표 1〉 제안모델들의 RMSE와 MAPE 결과 비교

종속변수: 매출	전통회귀분석		간단한 신경망(SPSS)		깊은 신경망		Recurrent 신경망	
	가맹점	가맹본부	가맹점	가맹본부	가맹점	가맹본부	가맹점	가맹본부
학습표본 RMSE(train)	0.83	0.58	0.80	0.58	0.34	0.30	0.18	0.10
검정표본 RMSE(test)	0.81	0.69	0.81	0.71	0.82	0.51	1.29	1.00
검정표본 MAPE(test)	0.87	1.33	0.87	1.33	0.63	1.22	2.00	3.66

결과를 보여주고 있다. 가맹본부 매출은 깊은 신경망이 다른 모델들에 비해서 좋은 값을 보여주고 있다. 검정표본에서의 RMSE값들은 학습표본의 값들보다 조금 클 수 있다. 깊은 신경망에서도 조금 그런 현상이 보이는데, 순환 신경망에서는 차이가 커서 확실히 과적합이 되었다고 하겠다. 〈표 1〉에 나와 있는 검정표본으로 계산한 MAPE를 보면, 깊은 신경망 모델이 가장 좋은 값을 보여주고 있다. 같은 결과를 가지고, RMSE와 MAPE가 약간 다른 모습을 보이는 것은 계산방법이 달라 오류의 크기와 횟수에 따라 영향을 받기 때문이다. 굳이 어떤 측정값이 확실히 좋다고 이야기 하기는 어렵다.

결과적으로 현재 사용하는 자료가 일반적으로 깊은 신경망이 대상으로 하고있는 자료들의 크기보다는 작아서 그 장점이 잘 보이지 않는 한계점도 있지만, 현재 적합(fit) 측정치도 좋은 깊은 신경망을 기본모델로 추가분석을 하도록 하겠다.

2. 모델들의 변수중요도 비교

일반적으로 적합도와 같이 중요한 모델의 평가요소는 자료의 성격이나 종속변수의 결정요인이 무엇인지를 알아내는 것이다. 종속변수가 없는 경우의 무감독(unsupervised) 학습을 하는 신경망은 상대적으로 이를 자체적으로 평가하기가 어렵지만, 종속변수가 있는 감독(supervised) 학습을 하는 신경망은 전통적인 회귀모델보다는 명확하지 않지만, 독립변수가 종속변수의 결정에 미치는 영향

력을 측정하고 있다. 실질적으로 변수의 중요도라는 이름으로 보다는 응용할 수 있는 유사영역으로 민감도분석(sensitivity analysis)이 훨씬 더 많은 연구가 진행되었다. 민감도분석은 기본적으로 모델이 얼마나 계산알고리즘과 서버의 정확도와 입력변수의 입력오류 등에도 불구하고 안정적인 결과를 내는지를 위한 개념이다 (Yeung, Cloete, and Shi 2010).

이를 위해서는 크게 두 가지로 나뉘는데, 첫 번째는 함수의 미분 값을 취하는 것으로, analytic 하게 도함수를 계산하는 방법이다(Yang et.al. 2013). 기본개념은 아래 수식 1)과 같은 Jacobian 함수를 통해서 입력값에 따른 결과의 변화를 찾아내는 것이다.

$$J_{ij(x)} = \frac{\delta}{\delta x_j} f_i(x) \quad 1)$$

where f(x)는 neural network 모델

x: 입력변수

하지만 여러 가지 변수가 있고, 변수들 간의 독립이 보장되지 않는 상황에서 하나이상의 은닉층이 있고 활성화 함수가 비선형일 때 도함수(derivative)를 활용한 결과의 해석은 조심해야 한다는 주장도 있다(Sarle 2000).

두 번째는 통계적으로 찾아내는 방법인데, 입력 값을 변화시키면서 결과 값의 변화가 얼마나 되는지를 찾아내는 방법들이다(Choi and Choi 1992). 변수중요도를 알아내기 위한 좀더 구체적으로 살펴보면, 먼저, X를 변

화시키면서 Y의 값의 변화를 찾아보는 것이 있고, 그 다음으로는 특정 X를 빼고 Y값을 얼마나 잘 맞추는지 (RMSE)를 비교해 보는 것이 있다(Sarle 2000). 도합수에 의존하는 방법이 문제가 있는 것처럼, X를 얼마나 변화시키는지 또, 자료의 수가 추정계수들보다 적을 때는 안정적인 값을 찾기 어렵고, 수많은 계산을 해야 한다는 어려움도 있다.

IBM SPSS22.0의 다층 퍼셉트론(multi-layer perceptron)에서는 민감도 측정치를 제공하고 있는데, 이것은 원하는 입력변수에 크기변화를 준 몇 개의 값을 주고, 이들 값을 활용한 예측값의 차이들의 norm들 중에서 가장 큰 차이를 보여주는 norm을 아래 수식 2)와 같이 찾아서, 다시 수식 3)과 같이 이를 표준화하여 활용하고 있다(IBM Knowledge Center 2018).

$$d_{pm} = \max_{x_{p1}, x_{p2} \in S_p} \left\| \hat{Y}_{p1}^{(m)} - \hat{Y}_{p2}^{(m)} \right\| \quad 2)$$

where $\hat{Y}_{pk}^{(m)}$: case m의 p변수를 활용했을 때의 예측값 (표준화 한 상태)
 m: case의 수

$$d_p = \frac{1}{M} \sum_{m=1}^M d_{pm} \quad 3)$$

3. 새로운 변수중요도 측정법의 제안

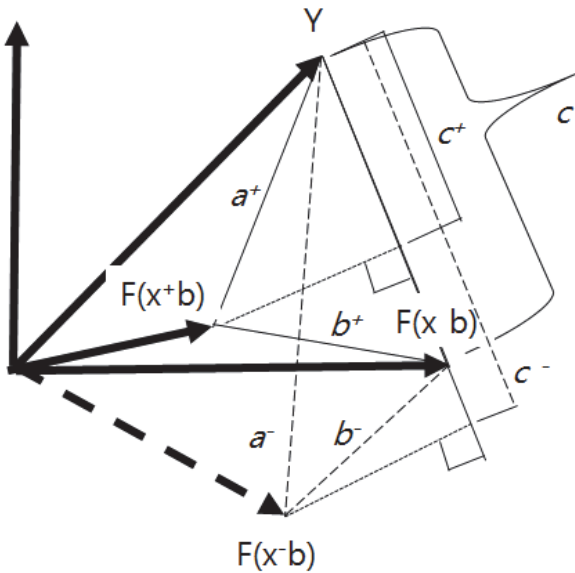
본 연구에서는 중요도를 계산하기 위한 새로운 방법을 제안하고자 한다. 먼저, 두 개의 통계량을 만들고자 한다. 첫 번째는 독립변수들의 값을 원래 값에서 +/-10% 변화시켜서 그것이 주는 종속변수의 차이를 활용하는 것이고, 두 번째는 해당 독립변수를 뺀 때의 종속변수의 변화를 보는 것이다. 제안방법들과 기존방법들과의 가장

큰 차이는 기존방법들이 종속변수와 변형된 독립변수들로 만든 추정치들의 차이에 근거한 방법이었다면, 본 방법은 종속변수와 원래의 독립변수들로 만든 추정치사이의 차이를 기본 축으로 하고, 새롭게 변형된 독립변수들로 만든 추정치가 기본축에서 볼 때 어느 정도의 차이를 만들고 있는 지로 평가하는 것이다. 즉, 변형된 독립변수들의 추정치를 원래의 독립변수들의 추정치의 관점에서 평가하는 것이라고 할 수 있다. 이 방법은 그냥 독립변수 하나씩 종속변수와의 관계를 보고 변수중요도를 평가하는 것보다는 독립변수들을 선택한 연구자의 모델링 의도 내에서의 해당독립변수의 변화를 보는 것에 더욱 충실한 방법이라고 할 수 있다.

첫 번째 통계량은 원자료인 종속변수, Y와 3가지의 Y 추정값: (1)모든 독립변수를 원래대로 가지고 추정한 Y 값, F(x b), (2) 특정한 한 개의 독립변수만을 원래 값에서 10% 더해서 추정한 Y값, F(x+ b), (3) 해당 독립변수만을 10% 줄여서 추정한 Y값, F(x- b)을 활용한다.

다음 <그림 4>과 같이 3개의 추정치는 주어진 X변수들을 가지고 모두 Y값과 가장 가깝게 되려고 노력한 결과이다. 목적은 각각의 독립변수들이 Y를 추정하는데 얼마나 중요한지를 보는 것이기에, 기본은 실제값 Y와 원래의 독립변수들로 추정한 F(x b)를 연결한 선을 기준으로 잡는다. 만약 일반 선형회귀식인 경우에는 이선은 Y에서부터 X의 space에 수직으로 연결한 선이 되지만, NN에서는 그렇게 되지는 않는다. 하지만 X변수의 중요도를 계산하기 위해서 F(x+ b)와 F(x- b)에서 Y와 F(x b)간의 직선에 수직이 되는 점을 찾아서 수식 4)와 5)와 같이 c+와 c-를 구하게 된다. 그래서 각 X변수가 Y를 추정하는데 있어서의 중요도는 X가 Y간의 거리 c에 있어서 X를 +/- 10% 했을 때 거리 c중에 얼마나 차지하는지를 가지고 평가하는 것이다.

〈그림 4〉 중요도 계산을 위한 3가지 Y추정 값과 Y실제 값의 관계



이때, 원래 모든 X를 자료 그대로 계산한 c 와 c^+ , c^- 의 크기에 따라서 X를 변화시켜서 만드는 중요도, *imp-rate* 공식은 아래 수식 6)와 같이 조금 달라진다(변수별 index는 생략하였다).

$$c^+ = \sqrt{\frac{|a^{+2} - b^{+2}| + c^2}{2c}} \quad 4)$$

$$c^- = \sqrt{\frac{|a^{-2} - b^{-2}| + c^2}{2c}} \quad 5)$$

$$\text{if } (c^+ - c) \times (c^- - c) > 0 \quad 6)$$

$$\text{imp-rate} = |c^+ - c^-|$$

else

$$\text{imp-rate} = |c^+ - c| + |c^- - c|$$

또 다른 중요도인 해당변수를 삭제한 후 차이를 보는 것(*imp-삭제*)은 수식 7)과 8)과 같이 위의 방식에서 X를 변형한 것을 한 개만 해서 거리를 구한 것과 동일하다.

$$c^0 = \sqrt{\frac{|a^{02} - b^{02}| + c^2}{2c}} \quad 7)$$

$$\text{imp-삭제} = |c^0 - c| \quad 8)$$

추후에 좀 더 연구자들에게 친숙한 p-value와 같은 통계치를 만들 수도 있을 것이다. 지금까지 변수중요도에 대한 통계치는 SAS나 SPSS와 같은 통계패키지회사에서 개발을 주도하고 있지만, 좀 더 신경망이 많이 활용된다면, 더 많은 연구가 계속될 것이라고 생각된다.

4. 변수중요도 결과

회귀분석과 간단한 신경망, 그리고 깊은 신경망 모델들의 중요도 결과를 〈표 2〉에 요약하였다. 회귀분석은 종속변수별로 분석을 해서 두 개의 회귀식의 결과를 보여주고 있으며, 신경망들은 한 번에 분석한 것이다. 간단한 신경망(SPSS)에서는 종속변수 개수에 상관없이 한 세트의 중요도 값을 보여주고 있다. 먼저 산업 dummy는 회귀분석과 깊은 신경망에서 한 개씩만 유의하게 나타나고 있는데, 그게 어떤 산업이든, 산업의 효과가 있다고만 평가해도 될듯하다. 깊은 신경망은 본 연구에서 제안한 두 가지 중요도 통계치, 즉 각 변수값을 조금씩 조정했을 때의 변화크기와 삭제했을 때의 변화크기를 모두 보여주고 있다. 깊은 신경망의 값은 수식 6)과 8)의 값들을 가장 큰 값을 100으로 해서 변환한 값을 보여주고 있다.

회귀분석은 계수의 크기와 방향성, 통계적인 유의까지 나오지만, 신경망은 예측 값을 만드는 데의 중요도가 방향성이나 통계적 판단기준이 없이 나타나고 있다. 하지만, 기존의 통계유의성이 가지는 한계점과 빅데이터 시대의 해석방법은 다른 점이 있음을 감안해야 할 것이다. 깊은 신경망에서 변수 값을 변화시킨 변수중요도와 변수

〈표 2〉 제안모델들의 변수중요도 비교분석

	회귀분석				간단한 신경망(SPSS)		깊은 신경망			
	가맹점매출		본사매출		모든변수		가맹점매출		본사매출	
	계수	유의확률	계수	유의확률	중요도	표준화 중요도	+/-차이	삭제시	+/-차이	삭제시
(상수)	0.01	0.86	(0.00)	0.90						
도소매	0.00	0.99	0.15	0.00	0.012	19%	7.04	7.70	4.01	6.01
외식	(0.01)	0.83	(0.01)	0.78	0.005	8%	83.21	94.57	100.00	27.24
서비스					0.007	11%	16.48	21.98	17.16	11.45
가맹사업개월16	(0.35)	0.07	0.06	0.67	0.021	32%	23.97	40.01	16.84	17.28
가맹점수16	0.05	0.46	(0.07)	0.15	0.016	25%	43.19	40.77	23.95	15.43
가맹본부임직원수16	0.06	0.04	0.15	0.00	0.024	37%	62.52	31.95	44.92	42.76
신규개점16	0.01	0.78	0.03	0.17	0.009	14%	48.10	60.75	36.99	26.02
계약종료16	0.03	0.22	(0.03)	0.05	0.008	12%	20.34	30.65	13.16	13.72
계약해지16	(0.01)	0.59	(0.00)	0.82	0.006	9%	44.79	63.51	26.83	17.15
명의변경16	0.05	0.06	(0.02)	0.33	0.007	11%	31.70	43.74	17.34	16.52
가맹평균매출16	0.44	0.00	0.01	0.76	0.035	53%	63.46	57.19	41.66	15.81
가맹비16	0.05	0.27	(0.02)	0.43	0.010	15%	10.98	14.22	5.33	10.93
교육비16	0.04	0.47	(0.02)	0.61	0.005	8%	15.38	24.88	6.85	9.15
보충금16	0.02	0.74	0.01	0.90	0.005	8%	28.14	41.60	16.32	13.60
기타인테리어16	(0.02)	0.65	(0.04)	0.11	0.015	23%	8.82	8.62	4.99	10.39
창업합계16	0.04	0.29	0.07	0.03	0.017	26%	19.05	25.33	8.14	11.50
자산16	0.33	0.00	0.15	0.00	0.050	77%	58.79	60.26	65.77	88.38
매출액16	(0.12)	0.00	0.09	0.00	0.031	47%	15.67	26.56	35.27	24.05
영업이익16	(0.02)	0.23	(0.00)	0.72	0.031	48%	0.52	2.10	1.22	0.04
당기순이익16	(0.01)	0.63	(0.01)	0.59	0.052	80%	2.53	1.47	7.91	10.00
총자산증가율16	(0.07)	0.00	0.07	0.00	0.024	36%	26.74	39.22	23.05	22.43
매출액증가율16	0.00	0.97	0.00	0.87	0.016	25%	21.94	27.36	6.12	12.30
영업이익증가율16	0.00	0.85	0.01	0.40	0.019	28%	24.61	26.91	31.26	17.38
자본16	(0.05)	0.04	0.02	0.15	0.065	99%	3.38	1.30	2.41	24.64
부채16	0.09	0.00	(0.06)	0.00	0.019	30%	38.25	65.07	36.16	33.19
법위반횟수16	(0.03)	0.28	0.03	0.15	0.006	10%	12.39	18.29	2.42	12.11
가맹사업개월15	0.29	0.13	0.01	0.93	0.018	28%	30.90	36.51	16.59	15.44
가맹점수15	(0.01)	0.90	0.16	0.04	0.008	12%	43.41	53.48	16.54	20.84
가맹본부임직원수15	(0.07)	0.01	0.17	0.00	0.031	47%	46.15	66.74	67.98	57.46
신규개점15	0.02	0.63	0.01	0.76	0.007	10%	38.96	46.72	28.09	21.72
계약종료15	0.01	0.67	0.00	0.91	0.006	10%	30.87	46.34	18.57	16.24
계약해지15	0.00	0.87	(0.04)	0.03	0.008	13%	40.05	55.08	29.24	20.31
명의변경15	0.02	0.41	0.00	0.95	0.008	12%	33.46	45.42	28.84	19.95
가맹평균매출15	(0.02)	0.74	(0.12)	0.02	0.005	7%	100.00	100.00	43.72	21.80
가맹비15	(0.02)	0.63	(0.01)	0.80	0.014	21%	9.96	13.05	5.05	12.25
교육비15	(0.04)	0.54	(0.01)	0.89	0.007	10%	19.52	24.28	7.57	14.00
보충금15	(0.02)	0.74	0.02	0.65	0.006	9%	34.02	50.19	19.51	15.47
기타인테리어15	0.03	0.44	(0.00)	0.99	0.012	19%	5.64	5.49	5.72	11.10
창업합계15	(0.03)	0.51	0.01	0.62	0.018	27%	24.62	38.48	10.32	13.60
자산15	(0.21)	0.00	0.46	0.00	0.047	72%	59.97	68.80	87.38	100.00
매출액15	(0.01)	0.61	(0.01)	0.72	0.020	30%	12.83	12.57	36.50	28.77
영업이익15	(0.01)	0.52	(0.00)	0.76	0.046	70%	0.29	0.06	0.79	9.06
당기순이익15	0.03	0.06	0.05	0.00	0.025	39%	0.25	0.51	0.73	10.19
총자산증가율15	0.05	0.03	0.03	0.04	0.022	34%	22.12	30.62	30.34	20.58
매출액증가율15	0.07	0.00	(0.03)	0.03	0.029	44%	18.40	20.92	10.13	13.35
영업이익증가율15	(0.01)	0.47	0.01	0.26	0.015	24%	27.78	26.42	6.36	12.18
자본15	(0.03)	0.24	0.07	0.00	0.065	100%	3.50	4.35	2.22	32.21
부채15	0.15	0.00	(0.14)	0.00	0.032	49%	37.62	65.81	24.37	31.85
법위반횟수15	0.02	0.36	(0.00)	0.86	0.010	15%	15.15	18.70	4.01	13.11

를 뺀 때의 변수중요도 중에서 어떤 것들이 좋은지는 현재의 자료만 가지고 판단하기는 어렵고, 유의미한 정보를 제공하는 서로 다른 두 개의 변수중요도라고 생각해야 할 것이다.

결과적으로 보면, 회귀분석과 깊은 신경망이 조금 더 유사한 모습을 보이고, 간단한 신경망은 약간 다른 모습을 보여주고 있다. 깊은 신경망의 두 가지 중요도 통계치들이 많은 부분 유사한 결과를 보여주고 있으며, 이것들과 회귀분석과도 유사한 경우들을 보여준다. 물론 다른 결과를 보여주는 것들에 대한 추가적인 검증도 필요하다고 생각하지만, 기존 통계검증에서와 같이 한 번의 자료를 가지고 모델적용으로 해서 의사결정을 하기 보다는 하나의 모델로 계속 들어오는 자료들의 결과추이를 보면서 의사결정을 해야 하니 앞으로의 중요도들의 추세치들을 보는 것도 바람직할 것이다.

세 가지 모델들의 결과를 종합해보면, 가맹점매출에는 본사직원수, 신규개점이나 폐점 등 가맹점수, 과거가맹점매출평균 또 본사의 자산규모, 자본, 부채가 영향을 미치고 있다. 2017년 예측에 2016년도 자료와 2015년도 변수가 거의 유사하게 중요한 영향력을 보이고 있었다.

가맹본부매출 예측을 보면, 본사직원수, 계약종료 가맹점수, 본사의 자산, 부채, 과거매출 등이 영향력을 보이고 있었다. 하지만, 본사매출에 미치는 영향변수는 가맹점매출에 미치는 변수의 수보다 현저히 적었다. 본사매출은 보고되고 있는 변수들 외의 환경에 더 많이 영향을 받는다고도 볼 수 있겠다.

종합적으로 보면 본사의 직원 수와 자산 및 부채규모와 가맹점들의 수와 개점 폐점수가 공통적으로 중요한 변수라고 하겠다.

IV. 모델결과의 활용 및 토의

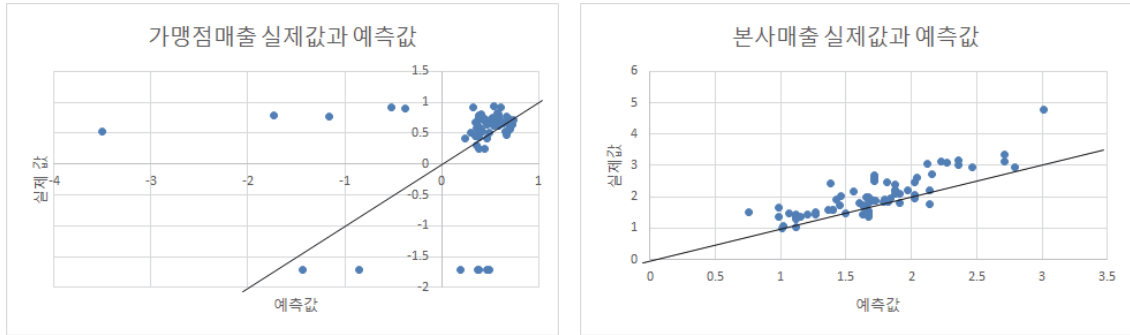
1. 모델결과의 활용

본 연구의 목적은 좀 더 정교한 예측을 통해서 첫째로, 가맹사업본부의 매출이 재무상태나 가맹점 개폐점 숫자 등 독립변수의 수준에 따라서 올려야 하는 매출수준을 제대로 올리고 있는지를 판단하고자 하는 것과 둘째로, 가맹본부매출과 해당본부의 가맹점평균매출을 경쟁사들과 비교하여 어떠한 가맹본부의 가맹점이 되는 것이 바람직할지를 결정할 정보를 제공하는 것이다. 두 번째 목적을 위해서는 현재의 매출상태를 보고도 생각해볼 수 있지만, 새롭게 가맹점을 해볼까 하는 경우에는 내년도 예측에 근거한 것이 더 바람직할 것이다.

첫 번째 목적을 위해서는 관심이 있는 가맹본부의 재정상태 등의 값을 추정되니 모델에 넣어서 그 예측치와 현재의 실제 값과 비교해보는 것이다. 일반적으로 가맹본부의 재정상태 변수를 중 한 두개를 넣는 것보다는 독립변수 값들을 모두 넣어서 비교하는 것이 좋으며, 하나의 유사한 여러 가맹본부사례들의 값을 넣어서 분포를 보는 것이 바람직할 것이다. 추후에 이들 추정치들이 적절하다는 평가를 받으면, 이를 근거로 한 다양한 통계치도 만들 수 있을 것이다.

본고에서는 예를 들어 외식산업에서 대형가맹본부라고 할 수 있는 가맹본부 임직원수, 신규개점수, 가맹점평균매출, 가맹본부매출, 자산, 자본변수들에서 1 표준편차 이상으로 큰 표본만을 대상으로 예측치와 실제 값을 비교해서 <그림 5>를 만들어보았다. <그림 5>를 보면 본사매출은 실제 값과 예측 값이 크게 차이가 없으나 가맹점매출은 실제 값보다 예측값, 즉 같은 수준의 가맹본부들이 가진 평균 가맹점매출보다 꽤 크거나 작은 점들이 존재한다. 따라서 이런 경우에는 주의해서 살펴보아야 하겠다.

〈그림 5〉 실제값과 예측값의 비교

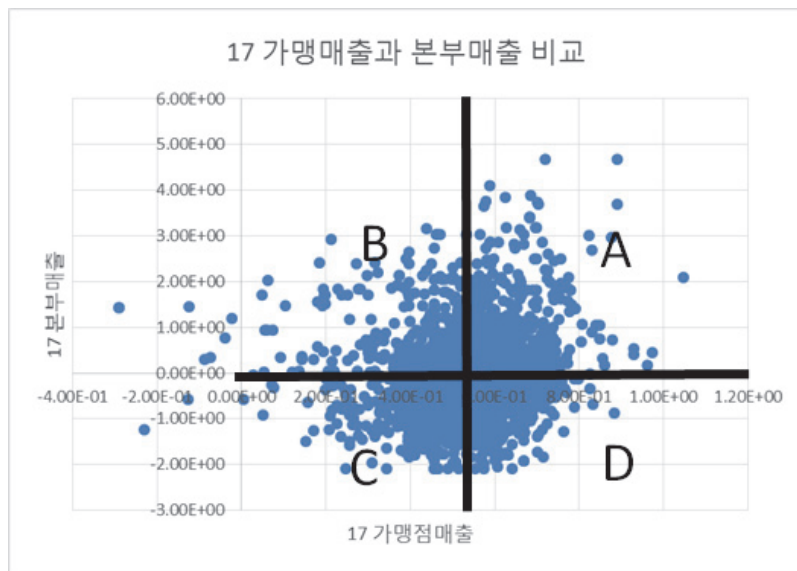


두 번째로 미래 가맹점을 개설하려는 미래 창업자의 경우, 먼저 가맹본부매출과 가맹점평균매출을 비교해 볼 수 있다. 〈그림 6〉에는 17년도 실제값을 비교해보았는데, 가맹계약을 맺고자 하는 가맹본부가 본부의 매출도 크고 가맹점평균매출도 크면 좋지만 (A부분), 가맹본부 매출은 큰데, 가맹점평균매출은 낮다면 (D부분) 조심해야 하는 것이다.

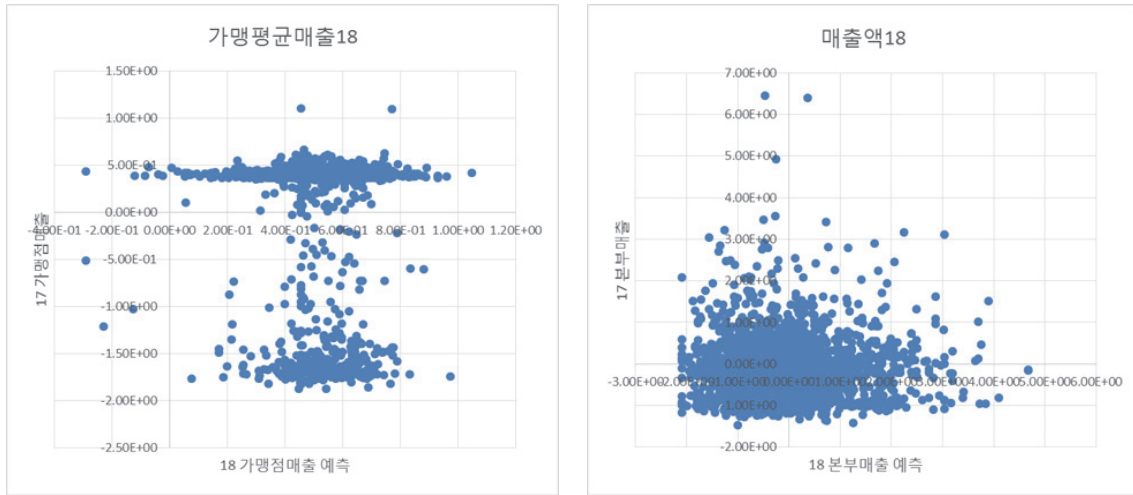
더 나아가서, 위 그림을 예측치로 바꾸어서 보면, 〈그림 7〉과 같은 모습을 볼 수 있다. 〈그림 7〉은 본 연구에서 실행한바와 같이 2015년과 2016년의 자료로 2017

년의 가맹점매출과 가맹본부매출을 예측하도록 학습시킨 계수추정치들을 통해, 2016년과 2017년 자료를 가지고 2018년 매출들을 추정해본 것이다. 아직 2018년 자료는 나오지 않아서 정확도는 판단할 수 없지만, 본사 매출액은 몇 개의 이상치(outlier)를 빼고는 명확한 구조는 보이지 않았는데, 가맹점평균을 보면, 18년도의 예측이 크게 두 집단으로 구분되는 것을 볼 수 있다. 모델의 정보를 따르면 아무래도 해당 가맹점평균매출이 증가할 것 같은 가맹본부를 선택하는 것이 바람직할 것이다.

〈그림 6〉 가맹본부매출과 가맹점평균매출의 비교



〈그림 7〉 내년도 예측치와 현재 값의 비교



2. 토의

가맹사업은 창업을 통해서 사업을 하려는 사람들에게 경쟁력이 검증된 사업아이템을 일정한 비용을 받고 제공하고 계속해서 이득을 얻을 수 있게 도와주는 것을 의미한다. 하지만 가맹본부가 제대로 가맹점주 들 혹은 미래의 가맹점주 들에게 정보를 알려주지 않고 무리한 요구를 해왔다는 것이 사회적 이슈가 되고 있다. 문제를 해결할 수 있는 기본은 정확한 정보를 아는 것인데, 본사의 경쟁력이나 가맹점의 경영능력 등을 알기가 어렵고, 그들이 처한 경제상황적인 변수도 제대로 고려하지 않은 이슈성 정보들은 실제 의사결정에는 크게 도움이 되지 않는다.

본 연구는 간단하지만, 보다 정확한 의사결정을 위해서 가맹본부와 가맹점포의 매출액을 설명하는 신경망 모델들을 제안하였으며, 그 결과를 해석하는 새로운 변수중요도 방법도 제안하였고, 기존의 가맹본부를 재무상태 등을 기준으로 판단할 때 쓸 수 있는 방법과 새로 가맹점을 하려고 할 때 가맹본부들을 평가하는 방법들을 제시하였다.

깊은 신경망 모델의 예측력을 측정하는 RMSE/MAPE 측정치에서 우수한 모델로 평가되었지만, 더 복잡한 순환 신경망모델은 과적합이 되었다. 또 제안한 변수중요도 측정법들의 효과를 제대로 검증했다고 보기 어렵다. 좀 더 연구를 해서, 변수중요도들에 대한 통계적인 추론이 가능한 방법을 개발해야 할 것이다.

현재 사용한 자료가 3년 자료밖에 되지 못했지만, 더 많은 년도의 자료가 있고, 개개의 가맹점들의 매출정보나 구매정보, 재무상태 및 위치정보 등이 있어서 정형 및 비정형자료가 많아지고 또 나아가서 실시간으로 들어오게 된다면, 기존의 회귀분석은 사용하기가 불가능했을 것이며, 간단한 신경망도 제대로 예측을 못했을 것이다. data analytics를 포함해서 신경망모델들은 온라인 상태로 들어오는 자료들을 batch로 나누어서 그때그때 모델을 update시킬 수 있다. 전통적인 회귀분석은 매우 안정적인 결과를 보여주고, 수많은 연구가 진행되어서 많은 장단점과 특징들이 알려져 있으며, 다른 모든 모델들의 기준점을 제공한다. 하지만 자료의 양이 모수통계가 발달한 시대에서 자료가 무한에 가까운 시대로 발전해 새로운 방법론과 새로운 기준점이 필요한 시기라고

생각한다.

신경망모델은 독립변수들이 서로 독립이 아니고, 그들 간의 관계가 간단한 구조로 설명되기가 어려울 정도로 복잡한 관계가 되어있고, 자료들이 계속 발생하며, 원하는 종속변수도 다양할 때, 기존의 통계모델들에 비해서 활용하기가 뛰어난 방법론이다. 특히 antecoder나 RBM 과 같은 무감독 학습모델들을 포함하면(Bengio 2009; Schmidhuber 2014), 연구자가 임의로 가정하지 않은 가맹본부를 형성하고 있는 독립변수들 간의 관계를 확인 해주고, 이를 통해서 원하는 종속변수를 예측할 수 있는 모델을 만들 수도 있을 것이다. 온라인에서 찾을 수 있는 많은 비정형자료까지 포함해서 예측모델을 만들 수 있다면 마케팅의 의사결정을 유의하게 향상시키는 단계 까지 이를 수 있을 것으로 믿는다.

마케팅연구에 있어서 특히 학문적 연구를 신경망으로 하기 에는 적절한 자료를 얻기가 쉽지 않다. 대부분의 신경망연구들은 그림이나 문자를 활용하는 로보틱스나 자동차주행, 번역, 화상진료 등을 하는 다양한 공학이나 자연계에서 연구가 진행되고 경영에서는 금융이나 쇼핑 등에서의 자동추천에 활용된다. 하지만 이런 쪽의 연구 는 학계가 아닌 실무에서 이루어지고, 실사 학계에서 이 런 연구결과를 보고하려 해도, 전통적인 평가 잣대를 가 지고 연구 성과를 평가한다면, 마케팅영역에서 해당연구 를 진행하는 연구자는 늘어나지 못할 것이다. 신경망분 야도 핵심모델링 부분이 아니더라도 창의적인 응용으로 마케팅문제를 해결할 수 있는 많은 가능성이 있다. 마케팅에 관심 많은 후학들을 위해서라도 이 분야에 꾸준한 연구가 진행되길 기대한다.

〈최초투고일: 2018년 8월 7일〉

〈수정일: 1차: 2018년 8월 16일〉

〈게재확정일: 2018년 8월 16일〉

참고문헌

- Agatonovic-Kustrin, S. and R. Beresford(2000), "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," *Journal of Pharmaceutical and Biomedical Analysis*, 22(5), 717-727.
- Bengio, Yoshua(2009), "Learning Deep Architectures for AI," *Foundations and Trends® in Machine Learning*, 2(1), 1-127.
- Choi, J. Y. and C.-H. Choi(1992), "Sensitivity analysis of multilayer perceptron with differentiable activation functions," *IEEE Trans. Neural Networks*, 3, 101-107.
- Hebb, Donald(1949), *The Organization of Behavior*. New York: Wiley. ISBN 978-1-135-63190-1.
- Hinte, Geoffrey E.(2007), "Learning multiple layers of representation," *TRENDS in Cognitive Sciences*, 11(10), 428-434.
- IBM Knowledge Center (2018), *MLP algorithm output statistics*, [https://www.ibm.com/support/knowledgecenter/ko/SSLVMB_22.0.0/com.ibm.spss.statistics.algorithms/alg_mlp_statistics.htm].
- Jing Yang, Xiaoqin Zeng, and Shuiming Zhong(2013), "Computation of multilayer perceptron sensitivity to input perturbation," *Neurocomputing*, 99, 390-398.
- Kim, Sungtae, Myoungkyu Kim, and Byungin Im(2014), "Franchise, is it a right path for self-employed people?: Focused on restaurant industry," *Small to Medium Firm Research*, 36(3), 79-103
- Kim, Insook and Sangsum Rhee(2018), "A Model of Self-Generating Co-growth for Sustainable Management in Korea," *Franchise Management Research*, 9(1), 7-15
- Kingma, D., and Ba, J.(2015), Adam: A method for stochastic optimization. *presented at ICLR*.
- Lim, Young Kyun and Juyoung Kim(2018), "Improving the Franchisor Revenue Structure Through the Settlement of Royalty System," *Journal of Channel and Retailing*, 23(3), 49-75

- McCulloch, Warren and Walter Pitts(1943), "A Logical Calculus of Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics*, 5(4), 115-133.
- Minsky, Marvin and Seymour A. Papert(1969), *Perceptrons: An Introduction to Computational Geometry*. MIT Press. ISBN 0-262-63022-2.
- Rosenblatt, F.(1958), "The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain," *Psychological Review*. 65(6), 386-408.
- Ryu, Seunghyoung, Jaekoo Noh, and Hongseok Kim(2017), "Deep Neural Network Based Demand Side Short Term Load Forecasting," Ryu, Seunghyoung, Jaekoo Noh, and Hongseok Kim (2016), "Deep neural network based demand side short term load forecasting." *Energies* 10(3), 1-20.
- Rumelhart, D. E., Geoffrey E. Hinton, and Ronald J. Williams (1986), "Learning representations by backpropagating errors," *Nature*, 323(6088), 533-536.
- Sarle, Warren S.(2000), *How to measure importance of inputs?*, SAS Institute Inc., Cary, NC, USA.
- Schmidhuber, Jurgen(2014), "Deep learning in neural networks: An overview," *Neural Network*, 61, 85-117.
- Song Byunghwa, Sungshik An, and Juyeon Park(2006), "A Comparative Study on the Choice of Franchise Contractors," *Travel Research*, 20(3), 97-117
- Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov(2014), "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, 15(1), 1929-1958.
- Sung, Daekwon, Jongphil Woo, and Hyungkyun Lee(2012), "A Study on the Brand Choice Property of Preliminary Founders in the Opening of the Food and Vegetables franchise," *Franchise Management Research*, 3(1), 92-110
- Sutton, Richard S. and Andrew G. Barto, *Reinforcement Learning: an Introduction*, A Bradford Book, The MIT Press Cambridge, Massachusetts. London, England.
- Xavier, Glorot and Yoshua Bengio(2010), "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS) 2010*, Chia Laguna Resort, Sardinia, Italy. Volume 9 of JMLR: W&CP 9.
- Yeung, Daniel S., Ian Cloete, and Daming Shi(2010), *Sensitivity Analysis for Neural Networks*, Springer Berlin Heidelberg.