

심층 신경망을 활용한 유동인구 예측 방법*

도유철, 권희창, 김강민, 송양의, 이용규
 동국대학교 컴퓨터공학과-서울
 e-mail : facee22av@naver.com

Floating Population Prediction Method Using Deep Neural Networks

Yoo-Cheol Doh, Hee-Chang Kwon, Gang-Min Kim, Yang-Eui Song, Yong Kyu Lee
 Department of Computer Science and Engineering, Dongguk University-Seoul

요 약

유동인구는 상권 분석에서 중요하게 사용되는 지표 중 하나로 유동인구 수와 속성에 따라 점포의 개점 가능성이 달라진다. 그중 유동인구 수는 상권의 이익과 직결되는 항목으로 그 수가 많을수록 기대되는 수익이 높다. 하지만 유동인구 조사는 사람이 직접 길목에 서서 세는 방식을 사용하고 있어 조사에 소모되는 비용이 매우 크다. 본 논문에서는 통계 데이터와 일반 근사 이론에 기반한 심층신경망을 통해서 유동인구를 예측하는 방법을 제시하고자 한다.

1. 서론

유동인구 수가 많다는 것은 좋은 상권이 되는 중요한 조건에 해당한다. 명동, 강남 등 많은 사람이 몰려 흔히 말하는 '좋은 목'은 비록 임차료가 비싸더라도 높은 매출이 기대되기 때문에 많은 예비창업자가 주목하는 지역이다. 하지만 모두가 명동, 강남 같은 명백히 유동인구가 많은 명당에 자리를 잡을 수 없다. 이미 유명한 장소 외에 유동인구가 높은 곳을 찾기 위해서 예비 창업자는 직접 발품을 팔거나 통계청에서 조사한 자료를 사용한다. 현재까지 유동인구 조사는 사람 손에 의존하고 있다. 직접 사람이 일정 거점마다 서서 온종일 시간대별로 지나가는 사람을 카운트해서 기록하는 방식이다. CCTV 를 통해서 유동인구를 기록하는 방법도 있지만, 서울시에서도 제한적인 지역만 실시되고 있다. 알고자 하는 지역마다 이 방법을 반복하는 것은 비효율적이다. 따라서 본 논문에서는 일반 근사 이론에 기반한 심층 신경망을 통해서 과거의 데이터를 바탕으로 직접 조사하지 않고 유동인구를 예측하는 방법을 제시하고자 한다.

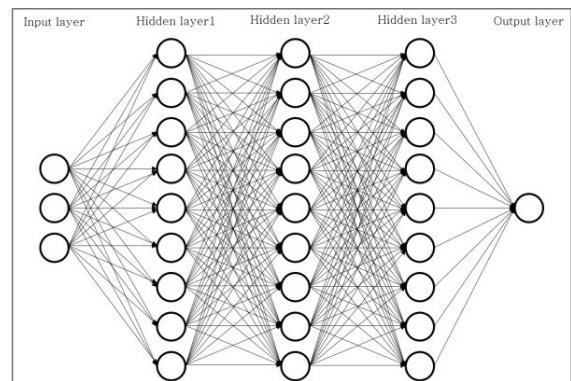
2. 관련 연구

2.1 심층신경망

딥러닝[2]이란 기계학습의 한 분야로 수치화 된 데이터뿐만 아니라 종전에 사람만이 구분할 수 있었던 다양한 데이터도 높은 수준의 분류가 가능한 인공지능의 이론 중 하나이다.

딥러닝에는 다양한 종류의 심층신경망이 존재한다. 심층신경망은 인공신경망과는 다르게 은닉층과 각 계층의 다수 뉴런이 연결되어있다. 입력 계층은 새

로운 은닉계층으로 연결되고 은닉계층의 수와 특징은 심층신경망의 목적에 맞게 다양한 형태로 구현된다. 마지막 은닉계층은 결과를 출력하는 출력계층을 통해 의미 있는 결과를 출력한다. 심층신경망의 은닉계층들은 계층별로 입력 값의 세부적인 특징을 잡아내는 역할을 하며 이런 심층신경망은 오류 역전과 알고리즘으로 학습된다. 다른 기계학습 기법들보다 상대적으로 우수하지만 계산복잡도가 높아 높은 컴퓨팅파워가 필요하다. 심층신경망 구조를 도식화하면 (그림 1)과 같다.



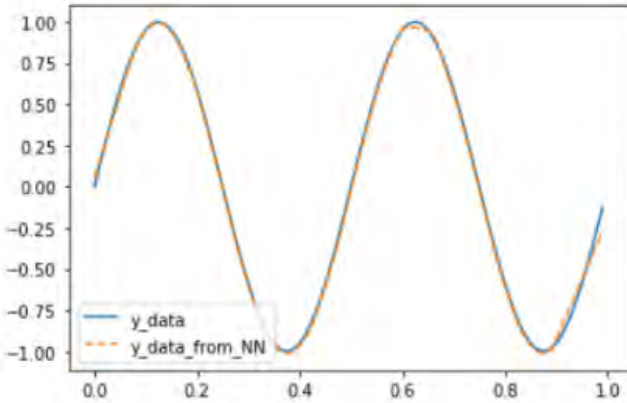
(그림 1) 심층신경망 구조

2.2 일반 근사 이론

1989 년 발표된 시벤코 정리는 뉴런 수만 충분하다면 은닉층 하나로 어떤 함수도 근사 할 수 있음을 밝

* 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW 중심대학지원사업의 연구결과로 수행되었음(2016-0-00017)

힌 이론이다. 다른 말로 Universal Approximation Theorem[1] 즉, 일반 근사 이론이라고도 한다. (그림 2)의 실선은 실제 sin 함수를 그리고 있다. 그리고 점선은 일반 근사 이론으로 1 개의 은닉층을 가진 뉴런 네트워크로 근사시킨 그래프이다. 두 그래프가 굉장히 유사하게 그려진 것을 확인할 수 있다. 이 이론을 바탕으로 유동인구 변화량을 함수형태로 근사시켜 값을 예측할 수 있다.



(그림 2) sin 그래프와 심층신경망이 학습한 sin 그래프

3. 심층신경망을 활용한 유동인구 예측 방법

3.1 데이터 처리

예측에 사용할 데이터를 심층신경망에 학습시키기 위해서 전처리 과정이 필요하다. 과거 유동인구 데이터는 조사지점별 요일 별 시간 순으로 정렬되어 있다. 지점별 측정 데이터는 학습목적에 따라 지역에 맞게 병합하는 과정이 필요하다.

```
X2013 <- inner_join(X2013, junggu_position, by = 'EXAMIN_SPOT_CD')

#동별로 축약
temp <- aggregate(POPLTN ~ SIG_DONG_CD
+EXAMIN_DATE +TMZON,X2013,sum)

#필동 추출 (코드번호 1102057)
temp <- temp %>% filter(SIG_DONG_CD == '1102057')
DATE <- c(3,2,0,4,1)
DATE <- rep(DATE, 14)
temp <- cbind(temp, DATE) # 요일 추가
temp <- temp %>% arrange(DATE) #요일순 정렬
TIME <- c(0:69)
temp <- cbind(temp, TIME) #시간추가
temp <- subset(temp, select=c(TIME, POPLTN))
temp #필동 시간별 유동인구
```

(그림 3) 필동 시간별 유동인구만 추출하는 R 구현 예시

본 논문에서는 동별로 유동인구 예측을 실시하였다. 데이터 처리는 R 언어를 사용할 수 있다. (그림 3)은 그 중 필동의 시간 별 유동인구를 추출하는 구현 예시이다.

3.2 예측 방법

미래의 유동인구를 예측하기 위해서는 과거의 데이터를 1 개 이상의 은닉층을 가지는 심층신경망에 학습시켜 결과 데이터를 얻을 수 있다. 이 데이터는 일반 근사 이론에 의하면 신뢰도가 높은 데이터이기 때문에 미래의 유동인구라고 예측 될 수 있다.

3.3 심층신경망 구현

동 별로 유동인구 변화 특징이 다르기 때문에 동 별로 개별적인 모델을 학습시켰다. 학습을 위한 신경망을 구성하는 데는 tensorflow 를 사용한다. 신경망을 구성하기 위해서 은닉층의 수, 은닉 노드의 수, 학습률, 활성화 함수, 입력 값, 출력 값, 을 지정해 주어야 하는데, 은닉층이 하나 이상이고 은닉 노드가 충분하기만 하면 되는 일반 근사 이론에 따라서 은닉층의 수와 은닉 노드의 수를 변화시켜가면서 여러 동 의 주간 유동인구 변화를 학습시켜보고 가장 적합한 신경망을 찾았다. 은닉층이 추가될수록 노드 수가 많을수록 높은 일치율을 보여주었다. 활성화 함수로는 Hyperbolic tangent, RELU, Maxout 등이 있는데 은닉층이 2 개 이상인 심층신경망에서는 기존의 Sigmoid 같은 활성화 함수보다 학습이 빠른 RELU 함수를 사용한다. RELU 함수는 기존의 Sigmoid 함수를 개선한 함수로 0 보다 작은 값에 대해서는 0 을 출력하고 기존 값은 그대로 전달한다. RELU 함수의 식은 식(1)과 같다.

$$f(a) = \max(0, a) \quad \text{식(1)}$$

입력 값은 일요일부터 월요일까지 시간 단위로 나열한 연속된 수가 들어가고 출력 값은 유동인구 수가 들어간다.

(그림 4)는 3 개의 은닉층과 은닉층별 300 개의 노드를 가진 심층신경망 구현 예시이다.

```
x_in=tf.placeholder(tf.float32, shape=[None,1], name='x-in')
y_tg=tf.placeholder(tf.float32, shape=[None,1], name='y-in')

rand_uni=tf.random_uniform_initializer(-1e-1, 1e-1)
fc1_output=tf.layers.dense(x_input_pc, 300, tf.nn.relu, use_bias=True,
kernel_initializer=rand_uni, bias_initializer=rand_uni)
fc2_output=tf.layers.dense(fc1_output, 300, tf.nn.relu, use_bias=True,
kernel_initializer=rand_uni, bias_initializer=rand_uni)
fc3_output=tf.layers.dense(fc2_output, 300, tf.nn.relu, use_bias=True,
kernel_initializer=rand_uni, bias_initializer=rand_uni)
y_output=tf.layers.dense(fc3_output, 1, None, use_bias=True,
kernel_initializer=rand_uni, bias_initializer=rand_uni)
error=(y_target_pc-y_output)
cost=tf.reduce_mean(tf.multiply(error, error), name="cost")

# node of tensorflow
train_step=tf.train.GradientDescentOptimizer(0.4).minimize(cost)
init=tf.global_variables_initializer()
sess=tf.Session()

sess.run(init)
```

(그림 4) 신경망 구성 tensorflow 코드 예시

3.4 정규화

학습데이터의 x 값 분포는 0~70 까지이고 학습데이터의 y 값 분포는 100~ 3000 까지 다양하다. 이 값들을 그대로 사용해서 학습시키면 계산 과정에서 오버플로가 발생하는 문제가 있기 때문에 일정 수치로 정규화해서 사용한다. 본 논문에서 x 값은 0~1 사이의 수로 y 값은 -1~1 사이의 수로 정규화해서 사용한다. (그림 5)는 학습데이터를 정규화하는 코드 예시이다.

```

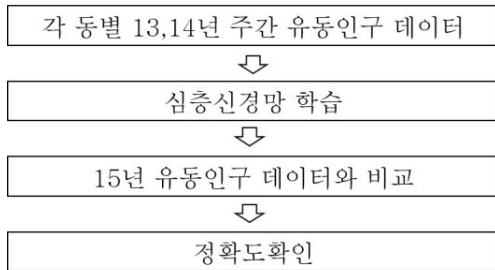
y_min = min(y_data)
y_max = max(y_data)
y_median = np.median(y_data)
for i in range(70) :
    x_data[i] = float(i) / 70.0
    y_data[i] = (y_data[i]-y_median)/(y_max-y_median)

x_data = [[i]for i in x_data]
y_data = [[i]for i in y_data]
x_data = np.array(x_data)
y_data = np.array(y_data)
    
```

(그림 5) 학습데이터를 정규화하는 코드 예시

4. 실험 결과

(그림 6)은 본 논문에서 제안하는 실험의 흐름을 도식화한 것이다.



(그림 6) 실험 진행 흐름

4.1 데이터 수집

서울시에서는 2012 년부터 2015 년까지 서울 지점마다 유동인구를 조사한 통계자료를 공개하고 있다. [3] 논문에서는 2013 년부터 2015 년까지의 조사지점별 유동인구 데이터와 조사지점코드 위치 데이터를 사용한다.

4.2 데이터 처리

<표 1>은 2013 년 조사지점별 시간대별 유동인구수를 기록한 데이터이다. 조사 지점은 코드로 분류가 되어 별도의 데이터에서 관리된다.

<표 1> 2013 년 조사지점별 유동인구 데이터

ID유동인구조사	조사지점코드	조사구분	조사요일	시간대	유동인구수	년도
ID_DYNMC_POPLTN	EXAMIN_SPOT_CD	EXAMIN_CLS	EXAMIN_DATE	TMZON	POPLTN	YEAR
1	01-029	본조사	금	07시-08시	387	2013
2	01-029	본조사	금	08시-09시	435	2013
3	01-029	본조사	금	09시-10시	924	2013
4	01-029	본조사	금	10시-11시	345	2013
5	01-029	본조사	금	11시-12시	1872	2013

<표 2> 조사지점코드 위치 데이터

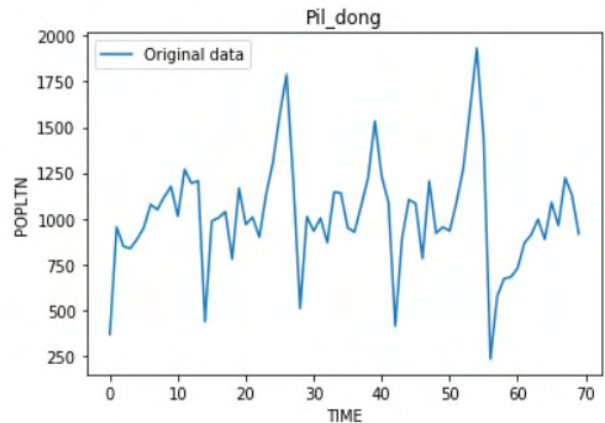
조사지점코드	조사지점명	X좌표	Y좌표	구코드	동코드	집계구코드
01-011	버스정류장에서 도로 바라보고 좌측 심홍상사.	200784.3716	452206.784	11020	1102059	1102059080001
01-012	뉴경신 속녀화.	201191.1455	452230.1139	11020	1102069	1102069020001
01-024	장안조명.	199682.2364	452120.0511	11020	1102060	1102060030001
01-076	버스정류장에서 도로 바라보고 좌측 20미터 장성풍	199156.7267	452071.4565	11020	1102055	1102055060001
01-226	동송동 LG텔레콤	200379.8543	451877.4618	11020	1102059	1102059080001

<표 2>는 조사지점코드별 위치정보가 기록된 데이터이다. 유동인구 변화를 시각화하기 위해서 조사지점코드별 위도 경도가 필요한데 원본 데이터에서는 TM 좌표계 [4]로 기록되어 있어 이를 위도 경도 좌표값인 WGS84 로 변환하였다. <표 3>은 변환이 완료된 새로운 데이터이다.

<표 3> 좌표계 변환된 데이터

조사지점코드	조사지점명	구코드	동코드	위도	경도
01-011	버스정류장에서 도로 바라보고 좌	11020	1102059	127.00887971	37.56935455
01-012	뉴경신 속녀화.	11020	1102069	127.01348475	37.56956434
01-024	장안조명	11020	1102060	126.99640270	37.56857329
01-076	버스정류장에서 도로 바라보고 좌	11020	1102055	126.99045363	37.56813508
01-226	동송동 LG텔레콤	11020	1102059	127.00430008	37.56638731

중구 데이터만 사용하기 위해서 <표 1> 데이터에서 (그림 4)의 코드를 활용해서 '필동' 시간별 유동인구를 추출한다. (그림 7)은 추출한 '필동' 시간별 유동인구 그래프이다. (그림 7)의 X 축은 일주일을 시간 순으로 나타낸 것이고 Y 축은 유동인구수를 나타낸다.

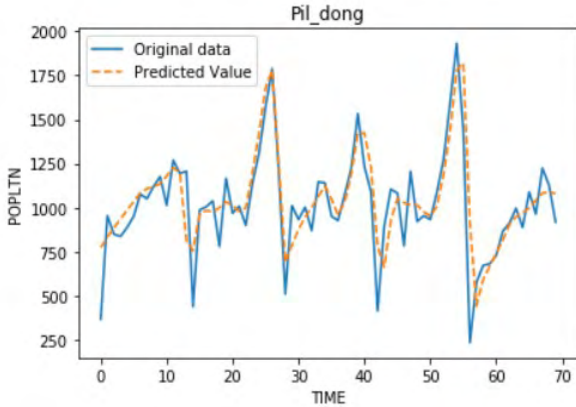


(그림 7) 필동 주간 유동인구 그래프

4.3 심층신경망 학습

필동의 주간 유동인구 수를 학습데이터로 (그림 4)와 같이 심층신경망을 학습시키면 다음과 같이 높은 일치율을 보이는 그래프를 얻을 수 있다. 유동인구를 구할 때는 예측된 값을 역 정규화 한다. (그림 8)은 심층신경망으로 예측된 필동 유동인구 그래프이다.

같은 방법으로 동별로 심층신경망을 각각 학습시켜서 특정 시기의 중구의 유동인구를 예측할 수 있다.



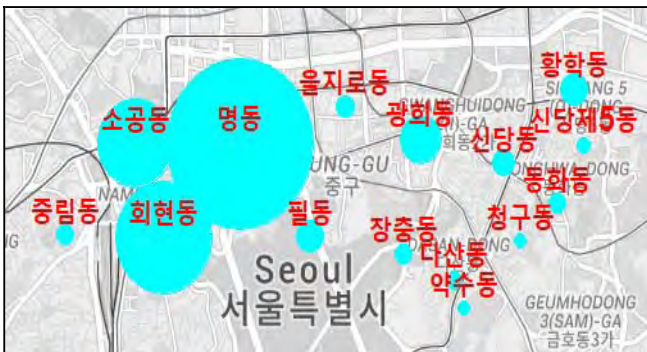
(그림 8) 심층신경망으로 예측된 필동 유동인구 그래프

4.4 실험 결과

예측한 결과가 실제 값과 일치하는지 확인하기 위해서 지도에 시각화하는 작업을 진행한다. 동별 학습된 심층신경망으로부터 화요일 11 시-12 시 유동인구를 예측값과 2015 년 유동인구 데이터와 지도 위 산점도로 비교한다. (그림 9)는 중구 화요일 11 시-12 시 유동인구 예측을 지도위에 산포도로 시각화한 그림이고 (그림 10)은 실제 2015 년 중구 화요일 11 시-12 시의 유동인구 지도이다.



(그림 9) 중구 화요일 11 시-12 시 유동인구 예측 지도



(그림 10) 2015년 중구 화요일 11 시-12 시 유동인구 지도

일부 을지로동, 광희동, 신당동, 청구동처럼 약간의 차이를 보이는 경우도 있고 명동, 소공동, 회현동처럼 비슷하게 결과가 나온 곳도 있다.

5. 결론

본 논문에서는 일반 근사 이론을 근거로 심층신경망을 사용해 동별 유동인구를 예측해 보았다. 유동인구는 시간대별로 유사한 흐름을 보이기 때문에 높은 일치율을 보이는 학습된 모델을 통해서 예측을 할 수 있었다.

본 논문의 결과를 바탕으로 직접 사람이 손으로 조사하지 않고 유동인구를 예측하는데 심층신경망을 활용할 수 있음을 확인할 수 있었다. 하지만 날씨 같은 유동인구에 큰 영향을 주는 외부변수에 대한 데이터가 반영되지 않았고, 또한 동별로 예측값과 실제값의 차이를 볼 때 유동인구 수가 단시간에 요동치는 경우 잘 예측하지 못하는 결과를 보여주었다.

따라서 추후 연구 사항으로 날씨, 경제 등 외부 변수를 획득해서 입력변수가 다양한 심층 신경망을 학습시키고 그 결과를 예측해 볼 예정이다.

참고문헌

- [1] Ronald R. Yager, Vladik Kreinovich - Universal Approximation Theorem for UninormBased Fuzzy Systems Modeling -University of Texas at El Paso 2002.10.1
- [2] Yann LeCun, Yoshua Bengio, Geoffrey Hinton - Deep learning - vol 521 NATURE 436-444 2015.5.28
- [3] 서울열린데이터광장, 년도별 유동인구, <https://data.seoul.go.kr>
- [4] 국토교통부, 국토지리정보원, <http://www.ngii.go.kr/>